Regularized estimation of mixed spectra using a circular Gibbs-Markov model

Philippe Ciuciu, J. Idier, Jean-François Giovannelli

Regularized Estimation of Mixed Spectra using a Circular Gibbs-Markov Model

Philippe CIUCIU, Jérôme IDIER, and Jean–François GIOVANNELLI

Abstract—**Formulated as a linear inverse problem, spectral estimation is particularly underdetermined when only short data sets are available. Regularization by penalization is an appealing nonparametric approach to solve such ill-posed problems. Following Sacchi *et al.* [1], we first address line spectra recovering in this framework. Then, we extend the methodology to situations of increasing difficulty: the case of smooth spectra, and the case of *mixed* spectra, *i.e.,* peaks embedded in smooth spectral contributions. The practical stake of the latter case is very high since it encompasses many problems of target detection and localization from remote sensing.**

The stress is put on adequate choices of penalty functions: following [1], *separable* functions are retained to retrieve peaks, whereas Gibbs-Markov potential functions are introduced to encode spectral smoothness. Finally, mixed spectra are obtained from the conjunction of contributions, each one bringing its own penalty function.

Spectral estimates are defined as minimizers of strictly convex criteria. In the cases of smooth and mixed spectra, we obtain nondifferentiable criteria. We adopt a *graduated nondifferentiability* approach to compute an estimate. The performance of the proposed techniques is tested on the well-known Kay and Marple example [2].

Keywords—**Spectral estimation; regularization; high-resolution; spectral smoothness; mixed spectra.**

I. INTRODUCTION

THE PROBLEM of spectral estimation has been receiving considerable attention in the signal processing community since it arises in various fields of engineering and applied physics, such as spectrometry, geophysics [1], biomedical Doppler echography [3], radar, etc. In particular, our primary field of interest is short-time estimation of atmospheric sounding or wind profiling, possibly superimposed on a small set of targets, from radar Doppler data [4].

A survey of classical methods for spectral estimation can be found in [2]. When the problem at hand is the restoration of *smooth spectra* (SS), basic nonparametric methods based on discrete Fourier transform (DFT) such as periodograms are often taken up. Such techniques usually involve a windowing or an averaging step which requires a sufficiently large data set. By contrast, estimation of *line spectra* (LS) is more often dealt with parametric methods, such as Pisarenko's harmonic decomposition [5], Prony's approaches [6, 7], or autoregressive (AR) methods [2, 8, 9]. These techniques are known for their ability to separate close harmonics. Consequently, they are usually considered under the heading of *high-resolution* methods [2].

In the more difficult case of *mixed spectra* (MS), *i.e.,* small sets of harmonics embedded in smooth spectral components, no satisfying techniques exist according to [2, 9, 10]. The main aim of the present paper is to contribute to fill the gap, within a nonparametric framework related to a recent contribution due to Sacchi *et al.* [1]. One important conclusion drawn in the latter was that enhanced nonparametric methods can reach high-resolution, which somewhat contradicts the state of the art sketched in [2].

Following [1], Section II starts with modeling the unknown spectral amplitudes as the DFT of the available observations. In particular, the number of Fourier coefficients to be estimated is larger than the length of the data sequence. The current problem is therefore underdetermined. Then, we resort to regularization by penalization to balance the lack of information provided by data with an available prior knowledge, such as spikyness or spectral regularity. Since the main part of our construction is made in a deterministic framework, Section II is also devoted to a natural question, whether it is theoretically justified to resort to our approach to estimate power spectral densities (PSDs).

Philippe CIUCIU is with the Commissariat à l'Énergie Atomique (DSV/DRM/SHFJ), 91406 Orsay, cedex France (email: ciuciu@shfj.cea.fr). Tel: (33) 1 69 86 78 13, Fax: (33) 1 69 86 77 86. Jérôme IDIER and Jean–François GIOVANNELLI (emails: name@lss.supelec.fr) are with the Laboratoire des Signaux et Systèmes (CNRS–SUPÉLEC–UPS), 91192 Gif-sur-Yvette cedex, France.

Three penalty functions are designed for solving the LS, SS and MS issues, respectively (see Section III). Following [1], a *separable* function is retained for line spectra (Subsection III-B). To deal with smooth spectra estimation, our construction is inspired by Gibbs-Markov edge-preserving models for image restoration [11–13] (see Subsection III-C). Finally, mixed spectra are obtained from the conjunction of contributions, each one bringing its own penalty function (Subsection III-D).

In all cases, the spectral estimate is defined as the minimizer of a strictly convex criterion, which is chosen nonquadratic to avoid oversmoothing effects [1, 14]. Practical computation of spectral estimates is tackled in Section IV. In the cases of smooth and mixed spectra, we obtain a nondifferentiable criterion, and we adopt a *graduated nondifferentiability* approach to compute an estimate. The performances of our spectral estimates are tested in Section V on the well-known Kay and Marple example [2]. Finally, concluding remarks and perspectives are drawn in Section VI.

II. Problem statement

A. Deterministic framework

Following contributions such as [1, 15], we formulate spectral estimation as a linear underdetermined inverse problem in a deterministic framework. Given discrete time observations $\boldsymbol{y} = [y_0, y_1, \dots, y_{N-1}]^{\mathrm{t}}$, the goal is to recover the energy distribution of data between frequencies 0 and 1. In the general setting of the paper, complex discrete data are processed to estimate spectral coefficients for normalized frequencies between 0 and 1 (the real data case is specifically examined in Appendix D).

The harmonic frequency model is usually considered for this task. In such a model, the distribution of spectral amplitudes $X(\nu)$ is continuous with respect to (w.r.t.) frequencies ν. Then, the inverse discrete-time Fourier transform links the unknown spectral function $X(\nu) \in L^2_{\mathbb{C}}[0,1]$ to a complex time series $(x_n)_{n\in\mathbb{Z}}$ (of finite energy) according to

$$x_n = \int_0^1 X(\nu)\, e^{2j\pi\nu n}\, d\nu,\ n \in \mathbb{Z}, \tag{1}$$

The signal $(x_n)_{n\in\mathbb{Z}}$ is partially observed through the data:

$$x_n = y_n, \qquad n \in \mathbb{N}_N \triangleq \{0, 1, \dots, N-1\}.$$

Within this setting, our approach consists in extracting a deterministic extension $(x_n)_{n\in\mathbb{Z}}$ of the data $\boldsymbol{y}$. Since this extension is of finite energy, it cannot be interpreted in general as a sample path of a stationary random process (see Section II-B for details).

Estimating $X(\nu)$ from $\boldsymbol{y}$ is a discrete-time continuous-frequency problem. Akin to [1], we propose to solve a discrete frequency approximation. It corresponds to the juxtaposition of a large number of sinusoids, say $P \gg N$, at equally sampled frequencies $\nu_p = p/P$, $p \in \mathbb{N}_P$. The accuracy of the approximation depends strongly on P, since the discrete counterpart of (1) reads

$$y_n = \sum_{p=0}^{P-1} X_p\, e^{2j\pi\nu_p n},\ n \in \mathbb{N}_N, \tag{2}$$

where $X_p \in \mathbb{C}$ are unknown spectral amplitudes. In the case of line spectrum estimation, choosing a large P seems clear since the harmonic components do not necessarily coincide with any sample of the grid. In the case of a continuous background, P is selected for suitably balancing the trade-off between an efficient computation of the estimate and a more accurate result. If $P = N$ could be satisfactory for smooth spectra (*e.g.,* Gaussian spectra with variance $\sigma > 0.1$), it could be preferable to consider higher values for piecewise smooth spectra with sharp transitions, such as ARMA PSDs with zeros of the MA part close to the poles of the AR part [16].

Let $w_0 = \exp(2j\pi/P)$, so that $W_{NP} = [w_0^{np}]_{n\in\mathbb{N}_N}^{p\in\mathbb{N}_P}$ is a $N \times P$ Fourier matrix, and an equivalent formulation of (2) is

$$\boldsymbol{y} = W_{NP}\boldsymbol{X}, \tag{3}$$

where $\boldsymbol{X} = [X_0, X_1, \dots, X_{P-1}]^{\mathrm{t}}$. Since $N \ll P$, system (3) is underdetermined, and there exists an infinite number of solutions. The problem is to incorporate structural information to raise the underdeterminacy in an appropriate manner.

B. Random processes

Following [1], our spectral estimation approach is based on the ground of deterministic Fourier analysis. Hence, a natural question arises, whether it is theoretically justified to resort to our construction to estimate PSDs. In the present subsection, we put forward that our approach is not a natural tool as far as PSD estimation is concerned.

Let $(S_n)_{n\in\mathbb{Z}}$ be a complex-valued random time series defined by

$$S_n = \int_0^1 e^{2j\pi\nu n}\, dS(\nu), \tag{4}$$

where $dS(\nu)$ stands for the random spectral measure of S. In a discrete-frequency framework, (4) can be approximated by

$$S_n^P = \sum_{p=0}^{P-1} S([p/P,(p+1)/P[)\, e^{2j\pi np/P}.$$

Our approach consists in estimating the variables $S([p/P,(p+1)/P[)$ and then in evaluating a spectrum of S through the vector of squared modulus $|S([p/P,(p+1)/P[)|^2$ (see Section III). In the case of a regular random process, such quantities are random. Thus, they do not identify with a discretized version of the PSD.

Nonetheless, as shown in [17], it is possible to exhibit a family of singular random processes for which our approach allows to characterize the power spectral measure of such processes.

III. Methodology

A. General setting

Sacchi *et al.* [1] have proposed a penalized approach, where an estimator of spectral amplitudes is defined as

$$\widehat{\boldsymbol{X}} \text{ minimizes } \mathcal{J}(\boldsymbol{X}) \text{ in } \mathbb{C}^P, \tag{5}$$

with

$$\mathcal{J} = \mathcal{Q} + \lambda\mathcal{R}, \tag{6}$$

$$\mathcal{Q}(\boldsymbol{X}) = \|\boldsymbol{y} - W_{NP}\boldsymbol{X}\|^2, \tag{7}$$

and the power spectrum estimator easily deduces as the squared modulus of the components of $\widehat{\boldsymbol{X}}$.

The hyperparameter $\lambda > 0$ controls the trade-off between the closeness to data and the confidence in a structural prior embodied in $\mathcal{R}$. In particular, in the case of *accurate data* ($\lambda \to 0$ see [1, § 4.A]), Sacchi *et al.* resort to Lagrange multipliers to prove that $\widehat{\boldsymbol{X}}$ identifies with the constrained minimizer of $\mathcal{R}(\boldsymbol{X})$ subject to (3).

In [1], the chosen penalty function reads

$$\mathcal{R}(\boldsymbol{X}) = \sum_{p=0}^{P-1} \log(1 + |X_p|^2/2\tau^2), \tag{8}$$

where $\tau > 0$ is a tunable scaling parameter, that controls the amount of sparseness in the solution. In [18, 19], the absolute norm $\mathcal{R}(\boldsymbol{X}) = \sum_{p=0}^{P-1} |X_p|$ is rather used because of its convexity, even if it is nonsmooth at zero. In both cases, let us remark that $\mathcal{R}$ is

- separable, *i.e.,* it is a sum of scalar functions, (9a)
- shift-invariant:
$$\mathcal{R}(X_0, X_1, \cdots, X_{P-1}) = \mathcal{R}(X_1, \cdots, X_{P-1}, X_0), \tag{9b}$$
- symmetry-invariant:
$$\mathcal{R}(X_0, X_1, \cdots, X_{P-1}) = \mathcal{R}(X_{P-1}, \cdots, X_1, X_0), \tag{9c}$$
- circular:
$$\mathcal{R}(X_0, \cdots, X_{P-1}) = \mathcal{R}(|X_0|, \cdots, |X_{P-1}|). \tag{9d}$$

The reference [1] adopts the classical Bayesian interpretation of $\widehat{\boldsymbol{X}}$ as a maximum *a Posteriori* estimate. As a random vector, $\boldsymbol{X}$ is given a prior neglog-density proportional to $\mathcal{R}(\boldsymbol{X})$, which amounts to choosing a product of circular Cauchy density functions as the *a priori* model. In such a probabilistic framework, properties of $\mathcal{R}$ can be restated as properties of the complex random vector $\boldsymbol{X}$: it is white according to (9a), stationary according to (9b), reversible according to (9c), and phases are uniformly distributed according to (9d).

Considering a circular model is rather natural, since no phase information is available. Stationarity and reversibility are also fair assumptions, unless some specific frequency domain shape information is known *a priori* (see [15] and references therein). Finally, choosing an independent prior seems justified as far as line spectra estimation is concerned. In the present paper, this framework is generalized to other kinds of spectra. More specifically, a stationary Gibbs-Markov model in the frequency domain will be introduced to incorporate spectral smoothness (see Subsection III-C).

From the computational viewpoint, (8) may not be the better choice, since $\log(1+x^2)$ is not a convex function on $\mathbb{R}_+$: $\widehat{\boldsymbol{X}}$ is not necessarily unique, and minimizing (6) using a local method such as the *Iterative Reweighted Least Squares* (IRLS) algorithm used in [1] may provide a local minimizer instead of a global solution. The absolute norm is also a possible choice [18, 19]. However, because it is nondifferentiable at zero, its optimization requires more sophisticated numerical tools, such as quadratic programming methods [20]. In the present paper, we restrict the choice to *strictly convex* penalty functions $\mathcal{R}$, in order to ensure that $\mathcal{J}$ is also strictly convex. As a consequence, $\mathcal{J}$ admits no local minima. Moreover, the minimizer $\widehat{\boldsymbol{X}}$ is unique and continuous w.r.t. the data [21]; this guarantees the well-posedness of the regularized problem [22]. Finally, many deterministic descent methods (such as gradient-based methods, but also the IRLS algorithm [23, 24]) will be ensured to converge toward

$\widehat{\boldsymbol{X}}$ if $\mathcal{R}$ is

- continuously differentiable (C^1), (10a)
- strictly convex, (10b)
- "infinite at infinity": $\lim_{\|\boldsymbol{X}\|\to\infty} \mathcal{R}(\boldsymbol{X}) = \infty$. (10c)

The construction of penalty functions that fulfill (10) forms the guideline of the next three subsections, in the LS, SS and MS cases, respectively.

B. Line Spectra

We are naturally led to penalty functions $\mathcal{R}_{\text{L}}$ that satisfy (9)-(10) (the subscript "L" stands for *line*). It is not difficult to see that (9) imposes the following form for $\mathcal{R}_{\text{L}}$:

$$\mathcal{R}_{\text{L}}(\boldsymbol{X}) = \sum_{p=0}^{P-1} R_0(\rho_p), \tag{11}$$

with $\rho_p = |X_p|$ and $R_0 : \mathbb{R}_+ \mapsto \mathbb{R}$. Then, the following proposition characterizes those functions R_0 that ensure the convexity of $\mathcal{R}_{\text{L}}$.

Proposition 1: Let $f : \mathbb{C} \mapsto \mathbb{R}$ be a circular function. Then, f is (resp. strictly) convex *if and only if* its restriction on $\mathbb{R}_+$ is a (resp. strictly) convex, nondecreasing (resp. increasing) function.

Proof: This property corresponds to the scalar case ($m = 1$) of Theorem 2 (Subsection III-C), which is proved in Appendix B. ■

From Proposition 1, it is apparent that $\mathcal{R}_{\text{L}}(\boldsymbol{X})$ is not convex if $R_0(\rho) = \log(1 + \rho^2/2\tau^2)$. Moreover, it can be then proved that $\mathcal{J}$ is not convex either. Thus, we prefer an alternate *convex* function R_0 that would enhance spectral peaks like the Cauchy prior does. We have borrowed such penalty functions from the field of *edge-preserving* image restoration [11–13, 25–27]. More precisely, we propose to resort to the following set of functions:

$$\mathcal{S} = \Big\{ f : \mathbb{R}_+ \mapsto \mathbb{R} \text{ convex, increasing, } C^1, \\ f'(0^+) = 0, 0 < \lim_{x\to 0^+} \frac{f'(x)}{x} < \infty, \lim_{x\to\infty} f'(x) < \infty \Big\}.$$

If $R_0 \in \mathcal{S}$, the global criterion $\mathcal{J}$ clearly fulfills (10). On the other hand, functions in $\mathcal{S}$ behave quadratically around zero and linearly at infinite:

$$0 < \lim_{x\to 0^+} f(x)/x^2 < \infty, \quad 0 < \lim_{x\to\infty} f(x)/x < \infty.$$

This is a relevant behavior for erasing small variations, and also for preserving large peaks and discontinuities that would be oversmoothed by quadratic penalization.

Some functions of $\mathcal{S}$, such as the *fair* function $R_0(\rho) = \rho/\tau_0 - \ln(1 + \rho/\tau_0)$ [12, 28] or Huber's function $R_0(\rho) = \rho^2/2\tau_0 + \tau_0/2$ if $\rho < \tau_0$, ρ otherwise [29], are also long since known in the field of robust statistics [28, 29]. They behave quadratically under the threshold τ_0 and linearly above. In practical simulations (see Section V-B.2), we have selected the *hyperbolic* potential $R_0(\rho) = \sqrt{\tau_0^2 + \rho^2}$ in $\mathcal{S}$.

C. Smooth spectra

C.1 Complex Gibbs-Markov regularization

In the field of signal and image restoration, Gibbs-Markov potential functions are often used as roughness penalty functions [11–13, 21, 26, 27, 30]. Adopting this approach in the case of spectral regularity, one might think of simply penalizing differences between complex coefficients, using

$$\mathcal{R}^1_{\text{S}}(\boldsymbol{X}) = \sum_{p=0}^{P-1} R_1(|X_{p+1} - X_p|), \tag{12}$$

where $X_P = X_0$ because of the circularity constraint. In (12), the subscript "S" stands for *smooth*. Then, provided that R_1 is convex and nondecreasing on $\mathbb{R}_+$, it is not difficult to deduce that $\mathcal{R}^1_{\text{S}}$ is convex from Proposition 1. When R_1 is quadratic, the estimated spectrum is a windowed periodogram, *i.e.,* a low-resolution solution [14]. In Section V-B.3, we have performed simulations using the hyperbolic function $R_1(\rho) = \sqrt{\tau_1^2 + \rho^2}$ in order to obtain solutions of higher resolution. The corresponding results are actually disappointing (*e.g.,* Fig. 3). Empirically, we observe that the penalty term (12) corresponds to spectral smoothness only roughly, while it produces hardly controlable artefacts. In fact, (12) is not a circular function of $\boldsymbol{X}$: $\mathcal{R}^1_{\text{S}}$ does not satisfy (9d). The regularization function $R_1(|X_{p+1} - X_p|)$ also introduces a smoothness constraint on the phases of the sinusoids, which does not coincide with some available prior knowledge. For this reason, let us examine the consequences of restricting to circular penalty terms.

C.2 Circular Gibbs-Markov regularization

The simplest circular energy coding spectral continuity is clearly

$$\mathcal{R}^2_{\text{S}}(\boldsymbol{X}) = \sum_{p=0}^{P-1} R_1(\rho_{p+1} - \rho_p), \tag{13}$$

since only two magnitudes ρ_p and ρ_{p+1} are involved. As an extension, one could consider higher

order smoothness terms such as $R_1(\rho_{p+1} - 2\rho_p + \rho_{p-1})$, which would be better adapted to restore piecewise linear unknown functions.

It is readily seen that (13) satisfies all conditions (9), save separability. Unfortunately, $\mathcal{R}_{\text{S}}^2$ is not convex if R_1 is an even, convex function. This negative result is a solidforward consequence of Corollary 1, stated below. Therefore, we propose to retain a slightly more general circular expression

$$\mathcal{R}_{\text{S}}(\boldsymbol{X}) = \sum_{p=0}^{P-1} \left(\mu R_1(\rho_{p+1} - \rho_p) + R_2(\rho_p)\right), \quad (14)$$

where parameter $\mu \geqslant 0$ tunes the amount of spectral smoothness and $R_2 : \mathbb{R}_+ \mapsto \mathbb{R}$. Expression (14) still satisfies conditions (9b)-(9d).

In the following, a necessary and sufficient condition for the convexity of $\mathcal{R}_{\text{S}}$ is given. For this purpose, the definition of *coordinatewise nondecreasing* function is a prerequisite. We also provide a useful theorem regarding the composition of convex functions.

Definition 1: A function $f : \mathbb{R}_+^m \mapsto \mathbb{R}$ is said *coordinatewise nondecreasing if and only if* $\forall\, i \in \{1, \ldots, m\}$:

$$\forall\, \boldsymbol{x} \in \mathbb{R}_+^m,\ \forall\, t \geqslant 0, \qquad f(\boldsymbol{x}) \leqslant f(\boldsymbol{x} + t\mathbf{1}_i),$$

where $\mathbf{1}_i$ is the ith canonical vector. The function f si said *coordinatewise increasing* if the latter inequalities are strict.

Theorem 1: Let $f : \mathbb{R}_+^m \mapsto \mathbb{R}$ be a convex, coordinatewise nondecreasing (resp. increasing) function, and let $\boldsymbol{g} : \mathbb{R}^n \mapsto \mathbb{R}_+^m$ a function such that each component $g_k : \mathbb{R}^n \mapsto \mathbb{R}_+$ is (resp. strictly) convex. Then, $f \circ \boldsymbol{g}$ is (resp. strictly) convex on $\mathbb{R}^n$.

Proof: see Appendix A. ■

Theorem 2: Let $f : \mathbb{C}^m \mapsto \mathbb{R}$ be a circular function. Then, f is (resp. strictly) convex *if and only if* its restriction on $\mathbb{R}_+^m$ is a (resp. strictly) convex coordinatewise nondecreasing (resp. increasing) function.

Proof: see Appendix B. ■

Because $R_1(\rho_{p+1} - \rho_p)$ is not a coordinatewise nondecreasing function of $\boldsymbol{\rho} = [\rho_0, \ldots, \rho_{P-1}]^{\text{t}}$, (13) is not convex, according to Theorem 2. In the case of (14), application of Theorem 2 yields the following result.

Corollary 1: Let $R_1 : \mathbb{R} \mapsto \mathbb{R}$ and $R_2 : \mathbb{R}_+ \mapsto \mathbb{R}$ be C^1 functions that satisfy the following assumptions:

- R_1 is even and convex, (15a)
- R_2 is (resp. strictly) convex and nondecreasing (resp. increasing), (15b)
- $\mu \leqslant \mu_{\text{sup}} = R_2'(0^+)/2R_1'(\infty)$. (15c)

Then, function $\mathcal{R}_{\text{S}}$ defined by (14) is (resp. strictly) convex.

Proof: See Appendix C. ■

Inequality (15c) gives an upper bound on the smoothness level that can be introduced while maintaining convexity of $\mathcal{R}_{\text{S}}$. It is important to notice that $\mu_{\text{sup}} > 0$ imposes $R_2'(0^+) > 0$. In the rest of the paper, we have selected the simplest potential R_2 that satisfies $R_2'(0^+) > 0$, *i.e.,* $R_2(\rho) = \rho$. Combined with the hyperbolic function $R_1(\rho) = \sqrt{\tau_1^2 + \rho^2}$, such a choice yields that $\mathcal{R}_{\text{S}}$ is convex if $\mu \leqslant 1/2$.

The condition $R_2'(0^+) > 0$ means that $R_2(|\cdot|)$ is not differentiable on $\mathbb{C}$ at zero, so $\mathcal{R}_{\text{S}}$ is nondifferentiable. Although conditions (15) are only sufficient, we have the intuition that convexity and differentiability are actually incompatible properties of $\mathcal{R}_{\text{S}}$ as defined by (14). In Section IV, we propose to minimize a close approximation of $\mathcal{R}_{\text{S}}$ that conciliates convexity and differentiability, so that a converging approximation of $\widehat{\boldsymbol{X}}$ can be easily computed.

D. Mixed spectra

A *mixed* spectrum consists of both frequency peaks and smooth spectral components, so we propose to split vector $\boldsymbol{X}$ into two sets of unknown variables: $\boldsymbol{X}_{\text{L}}$ for the frequency peaks, and $\boldsymbol{X}_{\text{S}}$ for the smoother components. The resulting fidelity to data term $\mathcal{Q}_{\text{M}}$ reads:

$$\begin{aligned} \mathcal{Q}_{\text{M}}(\boldsymbol{X}) &= \|\boldsymbol{y} - W_{NP}(\boldsymbol{X}_{\text{L}} + \boldsymbol{X}_{\text{S}})\|^2 \\ &= \left\|\boldsymbol{y} - W_{NP}\boldsymbol{X}\,[1,1]^{\text{t}}\right\|^2, \end{aligned}$$

where $\boldsymbol{X} = [\boldsymbol{X}_{\text{L}} \,|\, \boldsymbol{X}_{\text{S}}]$ is a $P \times 2$ complex matrix. The subscript "M" stands for *mixed.*

Then, it is only natural to introduce $\mathcal{R}_{\text{L}}$ (defined by (11)) and $\mathcal{R}_{\text{S}}$ (defined by (14)) as specific penalty terms for $\boldsymbol{X}_{\text{L}}$ and $\boldsymbol{X}_{\text{S}}$, respectively. The resulting criterion $\mathcal{J}_{\text{M}}$ reads

$$\mathcal{J}_{\text{M}}(\boldsymbol{X}) = \mathcal{Q}_{\text{M}}(\boldsymbol{X}) + \lambda_{\text{L}}\mathcal{R}_{\text{L}}(\boldsymbol{X}_{\text{L}}) + \lambda_{\text{S}}\mathcal{R}_{\text{S}}(\boldsymbol{X}_{\text{S}}), \quad (16)$$

which is a nondifferentiable function w.r.t. vanishing components of $\boldsymbol{X}_{\mathrm{S}}$, if $R_2'(0^+) > 0$. On the other hand, $\mathcal{J}_{\mathrm{M}}$ is (resp. strictly) convex w.r.t. $\boldsymbol{X}$ if $\mathcal{R}_{\mathrm{L}}$ and $\mathcal{R}_{\mathrm{S}}$ are (resp. strictly) convex. Then, the global minimizer is uniquely defined by

$$\widehat{\boldsymbol{X}} = \left[\widehat{\boldsymbol{X}}_{\mathrm{L}} \,|\, \widehat{\boldsymbol{X}}_{\mathrm{S}}\right] = \arg\min_{\boldsymbol{X}} \mathcal{J}_{\mathrm{M}}(\boldsymbol{X}),$$

In the Bayesian framework adopted in [1], it is not difficult to see that $\left(\widehat{\boldsymbol{X}}_{\mathrm{L}}, \widehat{\boldsymbol{X}}_{\mathrm{S}}\right)$ corresponds to the joint MAP solution obtained from a prior neglog-density proportional to $\lambda_{\mathrm{L}}\mathcal{R}_{\mathrm{L}}(\boldsymbol{X}_{\mathrm{L}}) + \lambda_{\mathrm{S}}\mathcal{R}_{\mathrm{S}}(\boldsymbol{X}_{\mathrm{S}})$. Finally, the estimated frequency distribution is taken as the squared modulus of the components of $\widehat{\boldsymbol{X}}_{\mathrm{L}} + \widehat{\boldsymbol{X}}_{\mathrm{S}}$.

Among possible refinements, a shorter vector $\boldsymbol{X}_{\mathrm{S}}$ could be introduced to encode the smooth components of the spectrum, as far as they require less accuracy. Then the fidelity to data term would become

$$\mathcal{Q}_{\mathrm{M}}(\boldsymbol{X}) = \|\boldsymbol{y} - W_{NP}\boldsymbol{X}_{\mathrm{L}} - W_{QP}\boldsymbol{X}_{\mathrm{S}}\|^2,$$

with $Q < P$. Such a modification could provide a (probably slight) increase of overall convergence speed, at roughly constant quality of estimation.

IV. Optimization stage

A. Graduated nondifferentiability

Nondifferentiable (*i.e., nonsmooth*) convex criteria can neither be straightforwardly minimized by gradient-based algorithms, since the gradient is not defined everywhere, nor by coordinate descent methods [31, p.61]. Nonetheless, there exist several ways to efficiently minimize such criteria [31–34]. Here, we resort to the so-called *regularization method* [31, 32, 35, 36]. In the following, it is rather referred to as a *graduated nondifferentiability* (GND) approach, in order to avoid the possible confusion with the notion of regularization for ill-posed problems. The principle is to successively minimize a discrete sequence of convex differentiable approximations that converge toward the original nonsmooth criterion.

We have adopted the GND approach because it is flexible, easy to implement, and also mathematically convergent. Under suitable conditions, the series of minimizers converges to the solution of the initial nonsmmoth programming problem [31, 32, 35, 36]. More specifically, we have the following result, based on [31, pp. 21-22].

Proposition 2: Let $\mathcal{J} : \mathbb{C}^P \mapsto \mathbb{R}$ fulfill (10b)-(10c), but not (10a), and $\mathcal{J}_\varepsilon$ ($\varepsilon > 0$) be a series of approximations of $\mathcal{J}$ that fulfills the three conditions (10). If $\mathcal{J}_\varepsilon$ converges toward $\mathcal{J}$ in the following sense:

$$\left\{ \begin{array}{l} \forall \boldsymbol{X},\ \lim\limits_{\varepsilon\to 0} \mathcal{J}_\varepsilon(\boldsymbol{X}) = \mathcal{J}(\boldsymbol{X}), \\ \lim\limits_{\varepsilon\to 0} \mathcal{J}_\varepsilon(\widehat{\boldsymbol{X}}_\varepsilon) \geqslant \mathcal{J}(\widehat{\boldsymbol{X}}), \end{array} \right. \tag{17}$$

where

$$\widehat{\boldsymbol{X}} = \arg\min_{\boldsymbol{X}\in\mathbb{C}^P} \mathcal{J}(\boldsymbol{X}), \quad \widehat{\boldsymbol{X}}_\varepsilon = \arg\min_{\boldsymbol{X}\in\mathbb{C}^P} \mathcal{J}_\varepsilon(\boldsymbol{X}),$$

then

$$\lim_{\varepsilon\to 0} \widehat{\boldsymbol{X}}_\varepsilon = \widehat{\boldsymbol{X}}.$$

Remark 1: In more general settings, convergence results akin to Proposition 2 can be obtained using the theory of Γ-convergence, which is a powerful mathematical tool to study the limiting behavior of the minimizer of a series of functions [37].

The remaining part of the section is devoted to the case of smooth spectra, *i.e.,* to the minimization of $\mathcal{J}_{\mathrm{S}}$ defined by (6), (7) and (14). Extension to the minimization of $\mathcal{J}_{\mathrm{M}}$ is straightforward.

B. Differentiable approximation of convex Gibbs-Markov penalty function

Practically, it is a prerequisite to build a differentiable convex approximation $\mathcal{R}_{\mathrm{S},\varepsilon}$ of the penalty term $\mathcal{R}_{\mathrm{S}}$, such that the series

$$\mathcal{J}_\varepsilon = \mathcal{Q} + \lambda\mathcal{R}_{\mathrm{S},\varepsilon} \tag{18}$$

satisfies the conditions of Proposition 2. Our construction of $\mathcal{R}_{\mathrm{S},\varepsilon}$ is based on the hyperbolic differentiable approximation of the magnitude function $|\cdot|$:

$$\varphi_\varepsilon : \mathbb{C} \mapsto \mathbb{R}_+, \qquad \varphi_\varepsilon(x) = \sqrt{\varepsilon^2 + |x|^2}, \tag{19}$$

where $\varepsilon > 0$. Such an approximation is known to satisfy conditions (17) [31, pp. 21-22], and has been already used in the field of image restoration [26, 27]. It is also called the *standard mollifier procedure* [26].

Let $q_p = \varphi_\varepsilon(X_p) = \varphi_\varepsilon(\rho_p)$ denote the above differentiable approximation of ρ_p and $\boldsymbol{q} = [q_0, q_1, \ldots, q_{P-1}]^{\mathrm{t}}$. Then the resulting modified smoothness penalty term $\mathcal{R}_{\mathrm{S},\varepsilon}$ satisfies (10) whereas $\mathcal{R}_{\mathrm{S}}$ only satisfies (10b)-(10c), according to the following consequence of Theorem 1 and of Corollary 1.

Corollary 2: Let R_1 meet the weak form of conditions (15) in Corollary 1, along with $R_2(\rho) = \rho$.

Then, the modified penalty term

$$\mathcal{R}_{\mathrm{S},\varepsilon}(\boldsymbol{X}) = \sum_{p=0}^{P-1} \left(\mu R_1(q_{p+1} - q_p) + q_p\right) \qquad (20)$$

is a strictly convex function of $\boldsymbol{X}$.

Proof: Let us remark that $\mathcal{R}_{\mathrm{S},\varepsilon} = \mathcal{R}_{\mathrm{S}} \circ \boldsymbol{\varphi}$, where $\boldsymbol{\varphi} = (\varphi_\varepsilon, \ldots, \varphi_\varepsilon)$ and $\mathcal{R}_{\mathrm{S}}$ is defined by (14) with $R_2(\rho) = \rho$. Then, the proof is an application of Theorem 1, with $\boldsymbol{g} = \boldsymbol{\varphi}$ and $f = \mathcal{R}_{\mathrm{S}}$, given that (i) each φ_k is strictly convex, (ii) according to Corollary 1, the restriction of $\mathcal{R}_{\mathrm{S}}$ on $\mathbb{R}_+^m$ is convex and coordinatewise increasing[1]. ■

C. Minimization of $\mathcal{J}_\varepsilon$

According to the principle of GND, for a finite sequence $\varepsilon_1 > \varepsilon_2 > \cdots > \varepsilon_K > 0$, the minimizers $\widehat{\boldsymbol{X}}_{\varepsilon_k}$ are recursively computed. At the kth iteration, a standard iterative descent algorithm is used to compute $\widehat{\boldsymbol{X}}_{\varepsilon_k}$. At iteration $k+1$, $\widehat{\boldsymbol{X}}_{\varepsilon_k}$ is used as the initial solution, and the process is repeated until $k = K$. Practical considerations regarding the stopping criterion, the updating rule of ε_k and the number K of iterations are reported in Section V.

For any $\varepsilon > 0$, the computation of $\widehat{\boldsymbol{X}}_\varepsilon$ can be obtained with many mathematically converging descent algorithms, since $\mathcal{J}_\varepsilon$ fulfills (10). Practically, several numerical strategies are studied and compared in [38]:

- The Polak-Ribiere version of conjugate gradient (CG) algorithm is implemented with a one-dimensional search [39].
- It is shown that the IRLS method proposed in [1] does not extend beyond the case of separable penalty functions.
- An original *Residual Steepest Descent* (RSD) [23] method is developed. It can also be seen as a deterministic *half-quadratic* algorithm based on Geman and Yang's construction [24, 30].

For a small value of ε_K, GND coupled to CG is more efficient than a single run of CG at $\varepsilon = \varepsilon_K$. This point is illustrated in Section V. In [38], the same conclusion is drawn concerning GND coupled to RSD.

V. Experiments

We illustrate the performances of the proposed spectral estimators in the context of short-time estimation by processing the well-known Kay and Marple example [2]. Such data have been extracted from a realization of a second-order stationnary random process. Since our approach is not theoretically well-suited for dealing with such processes, the spectral estimates will not be consistent with the true spectrum. Nonetheless, the results presented in the following prove that consistency is not a crucial issue as short-time estimation is addressed. As a preliminary question, the next subsection addresses the problem of hyperparameters selection.

A. Hyperparameters selection

In the first set of simulation results (Subsection V-B), hyperparameter values have been empirically selected after several trials, as those that visually work "the best". An alternative way for solving this step could be automatic hyperparameter selection. More specifically, when the sample size of the observations is large enough (several hundreds of data), the maximum likelihood estimate (MLE) can provide a valuable solution. In the last ten years, efficient Monte Carlo Markov Chains methods have been proposed to compute the MLE, for instance in the context of unsupervised line spectrum estimation [40].

In the case of small data sets, the MLE would probably lack of reliability, and more realistic solutions must be found, depending on the application. Automatic or assisted calibration of hyperparameter based on a training data set is sometimes possible. For instance, in the context of Doppler radar imaging as addressed in [41, Chap. V], an initial data set is recorded as the radar points at a reference direction that corresponds to an identified scenario, such as atmospheric sounding and wind profiling. This step allows to calibrate the radar sensor, but it could also be used for choosing the hyperparameters for the whole recording.

B. Kay and Marple example

B.1 Practical considerations

Following [1], the performances of the proposed methods are tested using Kay and Marple reference data set [2], which allows easy comparison with preexistent approaches. The data sequence is real, of length $N = 64$, and consists of three sinusoids at fractional frequencies 0.1, 0.2 and 0.21, superimposed on an additive colored noise sequence. The SNR of each harmonic is 10, 30, and 30 dB, respectively, where the SNR is defined as the ratio of the sinusoid power to the total power in the pass-

[1] Rigorous application of Corollary 1 only provides that the restriction of $\mathcal{R}_{\mathrm{S}}$ on $\mathbb{R}_+^m$ is nondecreasing. A careful inspection of Appendix C is needed to check that the strict result actually holds.

band of the colored noise process. The passband of the noise is centered at 0.35. The true spectrum appears in Fig 1.

[Figure 1 about here.]

Given the real nature of data and the symmetry properties studied in Appendix D, the spectra are only plotted on a half period $[0, 0.5]$. The different estimates have been computed using $P = 512$. In practice, taking $P > 512$ does not markedly improve the resolution.

As regards numerical implementation of CG, the following conjunction has been selected as stopping criterion:

$$\left|\mathcal{J}(\boldsymbol{X}^i) - \mathcal{J}(\boldsymbol{X}^{i-1})\right| / \mathcal{J}(\boldsymbol{X}^i) < \alpha_1$$
$$\left\|\boldsymbol{X}^i - \boldsymbol{X}^{i-1}\right\|_* / \left\|\boldsymbol{X}^i\right\|_* < \alpha_2,$$
$$\left\|\nabla \boldsymbol{X}^i\right\|_* < \alpha_3,$$

where $\boldsymbol{X}^i$ denotes the solution at the ith iteration of the minimization stage, and $*$ is 1 or 2. Following Vogel and Oman [26], we have rather chosen the l_1 norm, and the thresholds have been set to $(\alpha_1, \alpha_2, \alpha_3) = (10^{-7}, 10^{-5}, 10^{-6})$.

The same stopping criterion has been adopted for RSD, except that the third condition has not been tested.

B.2 Estimation of LS

The spectrum estimates depicted in Fig. 2 minimize penalized criteria with a separable penalty function: Fig. 2(a) corresponds to the quadratic potential $R_0(\rho) = \rho^2$, and Fig. 2(b) corresponds to the hyperbolic potential $R_0(\rho) = \sqrt{\tau_0^2 + \rho^2}$ for $(\lambda, \tau_0) = (0.06, 0.002)$.

As shown in [1, 14], quadratic regularization yields the zero-padded periodogram of the data sequence, up to a multiplicative constant. Since the nominal resolution of a 64-point sequence is 0.015, close sinusoids at 0.2 and 0.21 are not resolved. Moreover, this estimate is dominated by sidelobes that mask important features of the signal. In the following, the DFT of the zero-padded data sequence has been used to initialize all iterative minimization procedures.

The line spectra estimate depicted in Fig 2(b) is very similar to the spectral estimate computed with the Cauchy-Gauss model [1, Fig. 6], and also to the result given by the Hildebrand-Prony method [2, Fig. 16(h)]: the sinusoids are retrieved at the exact frequencies but with powers different from the original ones. Nonetheless, the power ratio (20 dB) is preserved between the three harmonics. On the other hand, the broadband part of the spectrum is not recovered. It is replaced by several spectral lines. This problem is also encountered in [1, 15] and in high-resolution parametric methods discussed by Kay and Marple [2].

From a computational standpoint, the IRLS method of [1] has been used as minimization tool. It is known to be convergent in the present situation [23, 24]. The solution is reached in about five to ten seconds on a standard Pentium II PC.

[Figure 2 about here.]

B.3 Estimation of SS

B.3.a Complex regularization. Fig. 3 shows the spectrum estimate computed from a convex penalized criterion with the noncircular penalty function $\mathcal{R}_{\mathrm{S}}^1$ defined by (12). It has been obtained with $\tau_1 = 0.1$ and $\lambda = 0.6$. Although the latter value corresponds to a high level of regularization, there remain some artefacts, the reversal of the lowest sinusoid being the main defect. To our opinion, such results definitely disqualify noncircular penalty functions.

[Figure 3 about here.]

B.3.b Regularization of the power spectrum. The three spectrum estimates depicted in Fig. 4 are obtained with a penalty function $\mathcal{R}_{\mathrm{S},\varepsilon}$ defined by (20). Three hyperparameters $(\lambda, \mu, \tau_1) \in \mathbb{R}_+^3$ need to be adjusted, let alone the target value ε_K for the closest approximation $\mathcal{J}_{\varepsilon_K}$ of $\mathcal{J}$. The results of Fig. 4 have been computed with $(\lambda, \tau_1) = (0.05, 0.001)$.

First, let us begin with general comments on Fig. 4. Akin to Fig 2(b), the three results nearly produce no sidelobes, compared to the periodogram. None of the three result allow to separate the two close harmonics, although a narrow-band component around frequency 0.2 is clearly distinguished. Similarly, the lowest sinusoid at frequency 0.1 is recovered under a broaden format. This is not surprising, since smoothness has been incorporated through the penalty function.

[Figure 4 about here.]

In Fig. 4(a)-(b), the value of μ has been chosen to correspond to the bound of convexity of $\mathcal{R}_{\mathrm{S},\varepsilon}$: $\mu = \mu_{\sup} = 0.5$, according to Subsection III-C.2, and different values of ε_K have been compared. A small parameter value $\varepsilon_K = 0.001$ yields a rather inadequate blocky result, as shown in Fig. 4(b). The discontinuities are due to the quasi-nondifferentiability of $\mathcal{R}_{\mathrm{S},\varepsilon}$. The rougher approximation depicted in Fig. 4(a) ($\varepsilon_K = 0.9$) provides a smoother estimate. However, it is not smooth

enough compared to the broadband part of the true spectrum. Increasing μ beyond the bound of convexity is necessary to get smoother results. The spectrum of Fig. 4(c) has been computed with $\mu = 5$ and $\varepsilon_K = 0.9$. It provides a more regular broadband response, quite close to the smooth part of the true spectrum. Among the estimators tested in [2], the maximum likelihood estimate (Capon method) shown in [2, Fig. 16(l)] provides a somewhat similar result. We retain such a tuning as a good candidate for the smooth part of the mixed model.

As regards practical aspects of minimization, the three results correspond to contrasted situations.

- In the case of Fig. 4(a), $\varepsilon_K = 0.9$ yields a criterion that is sufficiently far from nondifferentiability to be efficiently minimized in a single run of CG (*i.e.,* $K = 1$), spending about 25 seconds of CPU time.
- Fig. 4(b) has been obtained after three iterations of GND based on CG: $(\varepsilon_1, \varepsilon_2, \varepsilon_3) = (0.1, 0.01, 0.001)$, which globally took about 35 seconds of CPU time. In comparison, a single run at ε_3 takes about 60 seconds, as depicted in Fig. 5.

[Figure 5 about here.]

- The value $\mu = 5$ corresponding to Fig. 4(c) does not ensure that the criterion is convex. Hence, it is possibly multimodal. For this reason, we gradually increase the value of μ, following the *graduated non convexity* (GNC) approach [42, 43]. The principle is very similar to the GND technique described in Section IV. The empirically chosen law of evolution for μ is simply $\mu_k = k \times \mu_{\sup}$, so the initial criterion $\mathcal{J}_{\mu_1}$ is convex, as prescribed by the GNC approach.

B.4 Estimation of MS

The spectrum estimates depicted in Fig. 6(a)-(b) are obtained from the minimization of a differentiable approximation of the penalized criterion $\mathcal{J}_{\text{M}}$ defined by (16):

$$\mathcal{J}_{\text{M},\varepsilon}(\boldsymbol{X}) = \mathcal{Q}_{\text{M}}(\boldsymbol{X}) + \lambda_{\text{L}}\mathcal{R}_{\text{L}}(\boldsymbol{X}_{\text{L}}) + \lambda_{\text{S}}\mathcal{R}_{\text{S},\varepsilon}(\boldsymbol{X}_{\text{S}}). \tag{21}$$

The regularizing terms $\mathcal{R}_{\text{L}}$ (11) and $\mathcal{R}_{\text{S},\varepsilon}$ (20) depend on τ_0 and on $(\mu, \tau_1, \varepsilon_K)$, respectively. Given the results presented in the two previous subsections, we have retained $\tau_0 = 0.002, \tau_1 = 0.001, \varepsilon_K = 0.9$, and we have tested the two settings $\mu = \mu_{\sup} = 0.5$ and $\mu = 5$.

Two additional hyperparameters $(\lambda_{\text{L}}, \lambda_{\text{S}})$ appear in (21). It is *a priori* suited to choose the same order of magnitude for the values of λ_{L} and λ_{S}, otherwise the over-penalized term would yield a vanishing component. The values $\lambda_{\text{L}} = 0.005$ $\lambda_{\text{S}} = 0.0033$ have been retained.

Fig. 6(a) corresponds to $\mu = \mu_{\sup}$, so the minimized criterion is strictly convex. The result has been computed with CG. It clearly shows that the mixed model is able to resolve close sinusoids, while the broadband response is much closer from the SS estimate of Fig. 4(a) than from the LS estimate of Fig. 2(b). However, the broadband response is not smooth enough, and the small sinusoidal component is not as sharp as expected.

Fig. 6(b) corresponds to $\mu = 5$, so the minimized criterion is not convex and possibly multimodal. The result has been computed with GNC based on CG. The three spectral lines have sharp responses at the sinusoid frequencies and the power ratio between the different harmonics is preserved. Moreover, its smooth part is very close to the broadband component of the true spectrum. It is clearly the most satisfactory result among all estimates proposed in this paper. It also outperforms classical solutions computed on the same data set in [2].

Fig. 6(c) and (d) separately show $|\widehat{\boldsymbol{X}}_{\text{L}}|^2$ and $|\widehat{\boldsymbol{X}}_{\text{S}}|^2$, which are the components of the solution depicted in Fig. 6(b). As expected, the former is rather spiky while the latter is rather smooth. However, perfect separation was not the goal, since it would require that true decisions be taken regarding the presence of a line at each frequency sample, whereas our motivation was only to accurately estimate the whole spectrum. There is a somewhat similar difference between *image segmentation* and *edge-preserving restoration.*

[Figure 6 about here.]

VI. Concluding remarks

In the context of short-time estimation, we have proposed a new class of nonlinear spectral estimators, defined as minimizers of strictly convex energies. Firstly, we have addressed separable penalization introduced in [1, 18] for enhancing spectral lines.

Then, a substantial part of the paper has been devoted to smooth spectra restoration. We have introduced circular Gibbs-Markov penalty functions inspired from common models for signal and image restoration. However, the fact that penalization applies to moduli of complex quantities introduces specific difficulties. A rigourous mathemat-

ical study has been conducted, in order to build criteria gathering the expected properties such as differentiability, strict convexity, and the ability to discriminate spectra in favor of the smoothest.

Finally, since many practical spectral analysis problems involve both spectral lines and smooth components, we have proposed an original form of mixed criterion to superimpose the two kinds of components. We argue that this approach provides a very sharp tool for the detection of isolated objects embedded in broadband events. One possible application is the tracking of planes using a Dopler radar instrument, since the informative data is often embedded on meteorological clutter at low SNR. The proposed spectral estimators have then been extended to this framework to additionnaly take spatial or temporal continuity into account [41, Chap. V].

After the present study, some issues remain open. On the one hand, we observed in Section V that minimizing a convex criterion did not always yield a sufficiently smooth estimate. In practice, we resorted to graduated nonconvexity to overcome the limitation found in the convex analysis framework. By now, it is hard to tell whether the latter takes root in fundamental reasons, or if we simply failed in finding the "good" convex penalty function.

On the other hand, the proposed penalty functions are quite sophisticated. In practice, several hyperparameters have to be tuned, which is not always a simple task. In some situations, hyperparameter values can be selected using training data. Otherwise, depending on the size of the data set, automatic selection using a maximum likelihood approach may provide an alternative solution.

Finally, the question of asymptotic properties remains open. For instance, given the well-known properties of the averaged periodogram, it could be interesting to study the properties of averaged versions of our smooth spectra estimator.

Appendix

A. Proof of Theorem 1

The stated sufficient condition is acknowledged in the scalar case [44, Theorem 5.1].

Firstly, let us prove the implication in the large sense. For any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^m, \boldsymbol{x} \neq \boldsymbol{y}$, and any $\alpha \in (0,1)$, let $\boldsymbol{t} = \alpha\boldsymbol{x} + \bar{\alpha}\boldsymbol{y}$ and $\bar{\alpha} = 1 - \alpha$. Each g_k is convex:

$$g_k(\boldsymbol{t}) \leqslant \alpha g_k(\boldsymbol{x}) + \bar{\alpha} g_k(\boldsymbol{y}). \tag{22}$$

Then, using repeatedly the fact that f is a coordinatewise nondecreasing function, we deduce

$$\begin{aligned} f(\boldsymbol{g}(\boldsymbol{t})) &\leqslant f(\alpha\boldsymbol{g}(\boldsymbol{x}) + \bar{\alpha}\boldsymbol{g}(\boldsymbol{y})), & (23)\\ &\leqslant \alpha f(\boldsymbol{g}(\boldsymbol{x})) + \bar{\alpha} f(\boldsymbol{g}(\boldsymbol{y})), & (24) \end{aligned}$$

where the latter inequality holds because f is convex.

In order to prove the strict formulation, remark that there is at least one k such that $x_k \neq y_k$, so the corresponding inequality (22) becomes strict because g_k is strictly convex. Then, the strict counterpart of inequalities (23) and (24) also holds since f is coordinatewise increasing (remark that the strict convexity of f is unnecessary here).

B. Proof of Theorem 2

B.1 Sufficient condition

Let $f : \mathbb{R}_+^m \mapsto \mathbb{R}$ be a (resp. strictly) convex and coordinatewise nondecreasing (resp. increasing) function, and let $\boldsymbol{g} : \mathbb{C}^m \mapsto \mathbb{R}_+^m$ be the mapping of the moduli: $\forall\, \boldsymbol{x} \in \mathbb{C}^m$, $\boldsymbol{g}(\boldsymbol{x}) = (|x_1|, |x_2|, \ldots, |x_m|)$. We have to prove that $f \circ \boldsymbol{g}$ is (resp. strictly) convex.

In the large sense, this result is an immediate consequence of Theorem 1, for $n = 2m$. However, the strict counterpart of Theorem 1 does not apply, since $|\cdot|$ is not a *strictly* convex function. We need a more specific derivation, which is actually generalizable to any function $\boldsymbol{g}$ with *hemivariate* [45] convex components.

Let us consider the proof of Theorem 1. If f is strictly convex, (24) readily becomes strict provided that $\boldsymbol{g}(\boldsymbol{x}) \neq \boldsymbol{g}(\boldsymbol{y})$. Otherwise, assume $\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{g}(\boldsymbol{y})$, so that (24) reads $f(\boldsymbol{g}(\boldsymbol{t})) \leqslant f(\boldsymbol{g}(\boldsymbol{x}))$. Since $\boldsymbol{x} \neq \boldsymbol{y}$, there exists at least one k such that $x_k \neq y_k$. Then, $|x_k| = |y_k|$ implies $|t_k| < |x_k|$, since t_k belongs to the cord (x_k, y_k) of the centered circle of radius $|x_k|$. Since f is coordinatewise increasing, it follows that $f(\boldsymbol{g}(\boldsymbol{t})) < f(\boldsymbol{g}(\boldsymbol{x}))$, which is the expected strict counterpart of inequality (24).

B.2 Necessary condition

Let $f : \mathbb{C}^m \mapsto \mathbb{R}$ be a strictly convex, circular function. Its restriction on $\mathbb{R}_+^m$ is obviously strictly convex. We have to prove that it is also coordinatewise increasing.

Let $\mathbf{1}_k$ be the kth canonical vector in $\mathbb{R}^m$ and let $\widetilde{f}_{\boldsymbol{x},k}(t) = f(\boldsymbol{x} + (t - x_k)\mathbf{1}_k)$ the restriction of f to the line $\{\boldsymbol{u}, u_n = x_n, \forall\, n \neq k\}$ for any $t \in \mathbb{R}, \boldsymbol{x} \in \mathbb{R}^m$. Firstly, let us prove that all such restrictions $\widetilde{f}_{\boldsymbol{x},k}$ are even functions, *i.e.*, that $\widetilde{f}_{\boldsymbol{x},k}(-t) = \widetilde{f}_{\boldsymbol{x},k}(t)$.

$$\forall\, n \in \mathbb{N}_m,\ |x_n + (t - x_k)(\mathbf{1}_k)_n| = \begin{cases} |x_n| & \text{if } n \neq k, \\ |t| & \text{if } n = k. \end{cases}$$

Consequently, $|x_n + (-t - x_k)(\mathbf{1}_k)_n| = |x_n + (t - x_k)(\mathbf{1}_k)_n|$, and hence $f(\boldsymbol{x}-(t-x_k)\mathbf{1}_k) = f(\boldsymbol{x} + (t - x_k)\mathbf{1}_k)$ since f is circular. Therefore, $\widetilde{f}_{\boldsymbol{x},k}$ is even.

Since $\widetilde{f}_{\boldsymbol{x},k}$ is even and strictly convex on $\mathbb{R}$, it is increasing on $\mathbb{R}_+$, as shown below: $\forall s, t,\ 0 < s < t$, let $\alpha = (s+t)/2t$, so that $s = \alpha t+(1-\alpha)(-t)$. Since $\alpha \in (0,1)$ and $\widetilde{f}_{\boldsymbol{x},k}$ is strictly convex, $\widetilde{f}_{\boldsymbol{x},k}(s) < \alpha\widetilde{f}_{\boldsymbol{x},k}(t) + (1-\alpha)\widetilde{f}_{\boldsymbol{x},k}(-t) = \widetilde{f}_{\boldsymbol{x},k}(t)$ because $\widetilde{f}_{\boldsymbol{x},k}$ is even.

As a conclusion, all restrictions $\widetilde{f}_{\boldsymbol{x},k}$ are increasing on $\mathbb{R}_+$, *i.e.,* f is coordinatewise increasing on $\mathbb{R}_+^m$.

C. Proof of Corollary 1

First, let us decompose $\mathcal{R}_{\text{S}}$ according to $\mathcal{R}_{\text{S}}(\boldsymbol{X}) = \frac{1}{2}\sum_{p=0}^{P-1} S(X_p, X_{p+1})$, with

$$S(X_1, X_2) = S(\rho_1, \rho_2) = R_2(\rho_1) + R_2(\rho_2) + 2\mu R_1(\rho_1 - \rho_2), \quad (25)$$

and let us prove that conditions (15) imply the convexity of S on $\mathbb{C}^2$, which is a sufficient condition for the convexity of $\mathcal{R}_{\text{S}}$ on $\mathbb{C}^P$. Apply Theorem 2 to S. On one hand, S is convex on $\mathbb{R}_+^2$ as a sum of convex functions of (ρ_1, ρ_2). It is even strictly convex if R_2 is strictly convex.

On the other hand, let us prove that S is coordinatewise nondecreasing or even increasing as a function of (ρ_1, ρ_2) if conditions (15) hold. Since R_1 is even, $S(\rho_1, \rho_2) = S(\rho_2, \rho_1)$, so we need only to study the behavior of S as a function of ρ_1, say. Since R_1 is even and convex on $\mathbb{R}$, it is nondecreasing on $\mathbb{R}_+$ (the strict counterpart of this result is shown at the end of Appendix B). As a sum of nondecreasing functions of ρ_1, it is obvious that S is nondecreasing if $\rho_1 \geqslant \rho_2$. If $\rho_1 < \rho_2$, the condition $\partial S/\partial\rho_1 \geqslant 0$ reads

$$\forall\, \rho_1, \rho_2 > 0,\ \rho_1 < \rho_2,\quad R_2'(\rho_1) \geqslant 2\mu R_1'(\rho_2 - \rho_1),$$

which is equivalent to (15c) since R_1' and R_2' are nondecreasing. Finally, if R_2 is strictly convex, S is shown to be coordinatewise increasing along the same lines.

D. The real data case

The purpose of this appendix is to show that the proposed spectral estimation method (in either versions, LS, SS and MS) automatically preserves the Hermitian structure of the spectrum when real data are processed, so that the estimated power spectrum is symmetric.

Let us denote $\widehat{\boldsymbol{X}} = \mathcal{H}(\widehat{\boldsymbol{X}})$ the expected Hermitian property of $\widehat{\boldsymbol{X}}$, with

$$\mathcal{H}(X_0, X_1, \ldots, X_{P-1}) \triangleq (X_0^*, X_{P-1}^*, \ldots, X_1^*).$$

Equivalently, $\boldsymbol{X} = \mathcal{H}(\boldsymbol{X})$ means that the inverse DFT $\boldsymbol{x} = \text{IDFT}^{-1}(\boldsymbol{X})$ is a real vector. Convexity of the minimized criterion plays a basic role in the fulfillment of the Hermitian property of $\widehat{\boldsymbol{X}}$, as stated in the following proposition.

Proposition 3: Consider a real data set $\boldsymbol{y} \in \mathbb{R}^N$, and a penalty function $\mathcal{R} : \mathbb{R}_+^P \mapsto \mathbb{R}$ that fulfills (9b)-(9d) and (10b)-(10c). Firstly, the criterion $\mathcal{J}$ defined by (6)-(7) possesses the Hermitian symmetry $\mathcal{J}(\mathcal{H}(\boldsymbol{X})) = \mathcal{J}(\boldsymbol{X}), \forall \boldsymbol{X} \in \mathbb{C}^P$. Secondly, the unique minimizer of $\mathcal{J}$ satisfies $\widehat{\boldsymbol{X}} = \mathcal{H}(\widehat{\boldsymbol{X}})$.

Proof: Let us consider a non-Hermitian complex vector $\boldsymbol{X} \in \mathbb{C}^P$, *i.e.,* $\boldsymbol{X} \neq \mathcal{H}(\boldsymbol{X})$. Introduce $\boldsymbol{x} = \text{IDFT}^{-1}(\boldsymbol{X})$, so that

$$\mathcal{Q}(\boldsymbol{X}) = \sum_{n=0}^{N-1} |y_n - x_n|^2,$$

$$\mathcal{Q}(\mathcal{H}(\boldsymbol{X})) = \sum_{n=0}^{N-1} |y_n - x_n^*|^2.$$

Obviously, $\mathcal{Q}(\mathcal{H}(\boldsymbol{X})) = \mathcal{Q}(\boldsymbol{X})$ since $|y - x| = |y - x^*|$, $\forall y \in \mathbb{R}$, $x \in \mathbb{C}$. On the other hand, the modulus of the components of $\mathcal{H}(\boldsymbol{X})$ reads $(|X_0^*|, |X_{P-1}^*|, \ldots, |X_1^*|) = (|X_0|, |X_{P-1}|, \ldots, |X_1|)$, which proves that $\mathcal{R}(\mathcal{H}(\boldsymbol{X})) = \mathcal{R}(\boldsymbol{X})$ since $\mathcal{R}$ is shift-invariant (9b), symmetry-invariant (9c) and circular (9d). Finally, the identity $\mathcal{J}(\mathcal{H}(\boldsymbol{X})) = \mathcal{J}(\boldsymbol{X})$ gathers the two results. The first part of the proof is completed.

Now, consider the middle point

$$\boldsymbol{Z} = \big(\boldsymbol{X} + \mathcal{H}(\boldsymbol{X})\big)/2, \quad (26)$$

which obviously satisfies $\mathcal{H}(\boldsymbol{Z}) = \boldsymbol{Z}$. Since $\mathcal{J}$ is strictly convex,

$$\mathcal{J}(\boldsymbol{Z}) < \big(\mathcal{J}(\boldsymbol{X}) + \mathcal{J}(\mathcal{H}(\boldsymbol{X}))\big)/2 = \mathcal{J}(\boldsymbol{X}).$$

As a consequence, $\widehat{\boldsymbol{X}} = \mathcal{H}(\widehat{\boldsymbol{X}})$. ■

Proposition 3 directly applies to the LS and SS cases (including differentiable approximations considered in Subsection IV-B), while a straightforward generalization is needed in the MS case: along the same lines, it can be proved that $\mathcal{J}_{\text{M}}(\boldsymbol{X}_{\text{L}}, \boldsymbol{X}_{\text{S}}) = \mathcal{J}_{\text{M}}(\mathcal{H}(\boldsymbol{X}_{\text{L}}), \mathcal{H}(\boldsymbol{X}_{\text{S}}))$ in $\mathbb{C}^P \times \mathbb{C}^P$ and that $(\mathcal{H}(\widehat{\boldsymbol{X}}_{\text{L}}), \mathcal{H}(\widehat{\boldsymbol{X}}_{\text{S}})) = (\widehat{\boldsymbol{X}}_{\text{L}}, \widehat{\boldsymbol{X}}_{\text{S}})$, if both penalty functions $\mathcal{R}_{\text{L}}$ and $\mathcal{R}_{\text{S}}$ fulfill (9b)-(9d) and (10b)-(10c).

The remaining question concerns the situation where the criterion is nonconvex, as encountered in [1] or in GNC experiments, reported in Section V. Then, it does not seem possible to show that all minimizers (global or local) are Hermitian. However, the Hermitian symmetry of the criterion itself still holds (the corresponding part of the proof of Proposition 3 remains valid). This property has two favorable consequences:

- If $\mathcal{J}$ is unimodal, *i.e.,* it has one global minimizer $\widehat{\boldsymbol{X}}$ and no local minimizer, then $\mathcal{H}(\widehat{\boldsymbol{X}}) = \widehat{\boldsymbol{X}}$. Since strict convexity implies unimodality, this is an alternate argument for the second part of the proof of Proposition 3.
- The gradient of $\mathcal{J}$ is Hermitian: $\mathcal{H}(\nabla\mathcal{J}(\boldsymbol{X})) = \nabla\mathcal{J}(\boldsymbol{X})$, so gradient-based algorithms can be expected to propagate Hermitian symmetry along iterations from a Hermitian initialization point. We have also checked the same property for the IRLS algorithm used in [1].

References

[1] M. D. Sacchi, T. J. Ulrych, and C. J. Walker, "Interpolation and extrapolation using a high-resolution discrete Fourier transform", *IEEE Trans. Signal Processing*, vol. 46, pp. 31–38, January 1998.

[2] S. M. Kay and S. L. Marple, "Spectrum analysis – a modern perpective", *Proc. IEEE*, vol. 69, pp. 1380–1419, November 1981.

[3] J.-F. Giovannelli, G. Demoment, and A. Herment, "A Bayesian method for long AR spectral estimation: a comparative study", *IEEE Trans. Ultrasonics Ferroelectrics and Frequency Control*, vol. 43, pp. 220–233, March 1996.

[4] H. Sauvageot, *Radar météorologie. Télédetection active de l'atmosphère*, Eyrolles, Paris, France, 1982.

[5] V. Pisarenko, "The retrieval of harmonics from a covariance function", *J. of the Royal Astronomical Society*, vol. 33, pp. 347–360, 1973.

[6] B. P. Hildebrand, *Introduction to numerical analysis*, McGraw-Hill, New York, NY, 1956.

[7] R. N. McDonough and W. H. Huggins, "Best least-squares representation of signals by exponentials", *IEEE Trans. Automat. Contr.*, vol. AC-13, pp. 408–412, August 1968.

[8] T. J. Ulrych and R. W. Clayton, "Time series modelling and maximum entropy", vol. 12, pp. 188–200, 1976.

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

[10] S. L. Marple, *Digital Spectral Analysis with Applications*, Prentice-Hall, Englewood Cliffs, NJ, 1987.

[11] H. R. Künsch, "Robust priors for smoothing and image restoration", *Ann. Inst. Stat. Math.*, vol. 46, pp. 1–19, 1994.

[12] S. Brette and J. Idier, "Optimized single site update algorithms for image deblurring", in *Proc. IEEE ICIP*, Lausanne, Switzerland, September 1996, pp. 65–68.

[13] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging", *IEEE Trans. Image Processing*, vol. 6, pp. 298–311, February 1997.

[14] J.-F. Giovannelli and J. Idier, "Bayesian interpretation of periodograms", submitted to *IEEE Trans. Signal Processing*, GPI–L2S, 1999.

[15] S. D. Cabrera and T. W. Parks, "Extrapolation and spectral estimation with iterative weighted norm modification", *IEEE Trans. Signal Processing*, vol. 39, pp. 842–851, April 1991.

[16] C. I. Byrnes, T. T. Georgiou, and L. Anders, "A new approach to spectral estimation: a tunable high-resolution spectral estimator", submitted to *IEEE Trans. Signal Processing*, 2000.

[17] P. Ciuciu and J. Idier, "Statistical interpretation of short-time spectral estimators: valid case and fundamental limit !", Technical Report, GPI–L2S, 2001.

[18] N. Moal and J.-J. Fuchs, "Sinusoids in white noise: a quadratic programming approach", in *Proc. IEEE ICASSP*, Seattle, WA, May 1998, pp. 2221–2224, IEEE.

[19] J.-J. Fuchs, "Multipath time-delay estimation", *IEEE Trans. Signal Processing*, vol. 47, pp. 237–243, June 1999.

[20] D. P. Bertsekas, *Nonlinear programming*, Athena Scientific, Belmont, MA, 1995.

[21] C. A. Bouman and K. D. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation", *IEEE Trans. Image Processing*, vol. 2, pp. 296–310, July 1993.

[22] A. Tikhonov and V. Arsenin, *Solutions of Ill-Posed Problems*, Winston, Washington, DC, 1977.

[23] R. Yarlagadda, J. B. Bednar, and T. L. Watt, "Fast algorithms for l_p deconvolution", *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-33, pp. 174–182, February 1985.

[24] J. Idier, "Convex half-quadratic criteria and interacting auxiliary variables for image restoration", to appear in *IEEE Trans. Image Processing*, GPI–L2S, 2001.

[25] P. J. Green, "Bayesian reconstructions from emission tomography data using a modified EM algorithm", *IEEE Trans. Medical Imaging*, vol. 9, pp. 84–93, March 1990.

[26] R. V. Vogel and M. E. Oman, "Iterative methods for total variation denoising", *SIAM J. Sci. Comput.*, vol. 17, pp. 227–238, January 1996.

[27] Y. Li and F. Santosa, "A computational algorithm for minimizing total variation in image restoration", *IEEE Trans. Image Processing*, vol. 5, pp. 987–995, 1996.

[28] W. J. Rey, *Introduction to robust and quasi-robust statistical methods*, Springer-Verlag, Berlin, 1983.

[29] P. J. Huber, *Robust Statistics*, John Wiley, New York, NY, 1981.

[30] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization", *IEEE Trans. Image Processing*, vol. 4, pp. 932–946, July 1995.

[31] R. Glowinski, J. L. Lions, and R. Trémolières, *Analyse numérique des inéquations variationnelles, tome 1 : Théorie générale, Méthodes mathématiques pour l'informatique*, Dunod, Paris, France, 1976.

[32] D. Bertsekas, "Nondifferentiable optimization approximation", in *Mathematical Programming Studies*, vol. 3, pp. 1–25. Balinski, M.L. and Wolfe, P., Amsterdam, The Netherlands, 1975.

[33] C. Lemaréchal, *Nondifferentiable optimization*, pp. 149–199, Dixon, L. C. W. and Spedicato, E. and Szeg:o,

G. P., Boston, MA, non linear optimization edition, 1980.

[34] K. C. Kiwiel, *Methods of descent for nondifferentiable optimization*, Lecture notes in Mathematics. Springer Verlag, New York, NY, 1986.

[35] R. Acar and C. R. Vogel, "Analysis of bounded variation penalty methods for ill-posed problems", *Inverse Problems*, vol. 10, pp. 1217–1229, 1994.

[36] M. Z. Nashed and O. Scherzer, "Stable approximation of nondifferentiable optimization problems with variational inequalities", *American Mathematical Society*, vol. 204, pp. 155–170, 1997.

[37] G. Alberti, *Variational Models for Phase Transitions, an Approach via Gamma-Convergence*, in Differential Equations and Calculus of Variations. Springer Verlag, G. Buttazzo *et al.* edition, 1999.

[38] P. Ciuciu and J. Idier, "A Half-Quadratic block-coordinate descent method for spectral estimation", Technical Report, GPI–L2S, 2000.

[39] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical recipes, the art of scientific computing*, Cambridge Univ. Press, Cambridge, MA, 1986.

[40] C. Andrieu and A. Doucet, "Joint Bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC", *IEEE Trans. Image Processing*, vol. 47, pp. 456–463, October 1999.

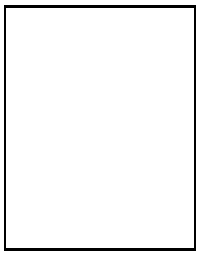
[41] P. Ciuciu, *Méthodes markoviennes en estimation spectrale non paramétrique. Applications en imagerie radar Doppler*, Phd thesis, Université de Paris–Sud, Orsay, France, October 2000.

[42] A. Blake and A. Zisserman, *Visual reconstruction*, The MIT Press, Cambridge, MA, 1987.

[43] M. Nikolova, J. Idier, and A. Mohammad-Djafari, "Inversion of large-support ill-posed linear operators using a piecewise Gaussian MRF", *IEEE Trans. Image Processing*, vol. 7, pp. 571–585, April 1998.

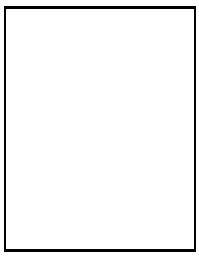
[44] R. T. Rockafellar, *Convex Analysis*, Princeton Univ. Press, 1970.

[45] J. Ortega and W. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, NY, 1970.



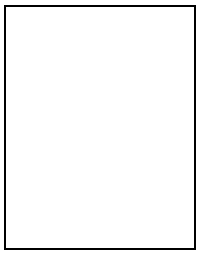
Philippe Ciuciu was born in France in 1973. He graduated from the École Supérieure d'Informatique Électronique Automatique, Paris, France, in 1996. He received also the DEA and Ph.D. degrees in signal processing from the Université de Paris-sud, Orsay, France, in 1996 and 2000, respectively.

Since November 2000, he has held a postdoctoral position with the Service Hospitalier Frédéric Joliot, Commissariat à l'Énergie Atomique. His research interests include spectral analysis, optimization, and presently he focuses on statistical methods and regularized approaches in signal and image processing for functional brain imaging.



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

Since 1991, he has been with the Laboratoire des Signaux et Systèmes, Centre National de la Recherche Scientifique. His major scientific interest are in probabilistic approaches to inverse problems for signal and image processing.



Jean–François Giovannelli was born in Béziers, France, in 1966. He graduated from the École Nationale Supérieure de l'Électronique et de ses Applications, Cergy, France, in 1990. He received the Doctorat degree in physics at the Laboratoire des Signaux et Systèmes, Université de Paris-Sud, Orsay, France, in 1995.

He is presently assistant professor in the Département de Physique at Université de Paris-Sud. He is interested in regularization method for inverse problems in signal and image processing, mainly in spectral characterization. Applications fields essentially concern radars and medical imaging.

FIGURES

Fig. 1 True spectrum

Fig. 2 Spectra reconstructed with separable regularization. (a): zero-padded periodogram, (b) line spectra reconstructed with the hyperbolic potential, $(\lambda, \tau_0) = (0.06, 0.002)$.

Fig. 3 Smooth spectrum reconstructed with a complex Gibbs-Markov penalty function. Parameters have been fixed to $(\lambda, \tau_1) = (0.6, 0.1)$.

Fig. 4 Smooth spectra reconstructed with a circular Gibbs-Markov penalty function, $(\lambda, \tau_1) = (0.05, 0.001)$; (a) convex case where $\mu = \mu_{\sup} = 0.5$, $\varepsilon_K = 0.9$, (b) convex case where $\mu = \mu_{\sup} = 0.5$, $\varepsilon_K = 0.001$, (c) nonconvex case where $\mu = 5$, $\varepsilon_K = 0.9$.

Fig. 5 Performance of the GND algorithm coupled with CG, in the SS case: the solid line corresponds to the minimization of $\mathcal{J}_{0.001}$ in a single run, dashed-dotted lines to the GND process coupled to CG.

Fig. 6 Mixed spectra. (a): convex case $\mu = 0.5$; (b) nonconvex extension $\mu = 5$; (c) and (d) correspond respectively to the line $(|\widehat{\boldsymbol{X}}_{\mathrm{L}}|^2)$ and smooth $(|\widehat{\boldsymbol{X}}_{\mathrm{S}}|^2)$ parts of $|\widehat{\boldsymbol{X}}|^2$ depicted in (b).

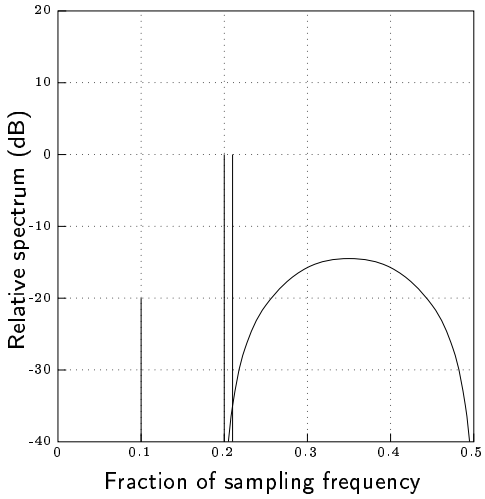


Fig. 1. True spectrum

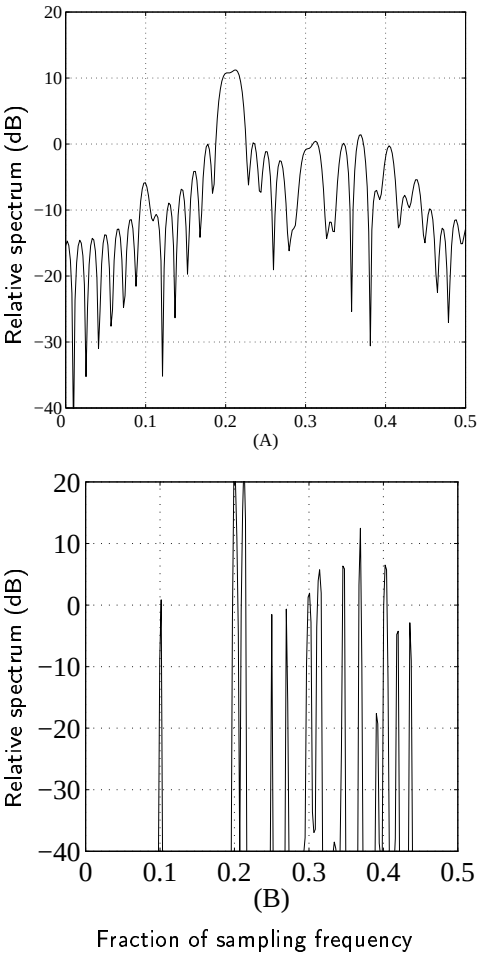


Fig. 2. Spectra reconstructed with separable regularization. (a): zero-padded periodogram, (b) line spectra reconstructed with the hyperbolic potential, $(\lambda, \tau_0) = (0.06, 0.002)$.

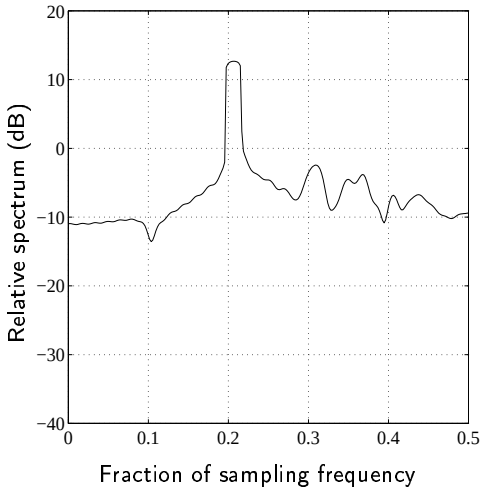


Fig. 3. Smooth spectrum reconstructed with a complex Gibbs-Markov penalty function. Parameters have been fixed to $(\lambda, \tau_1) = (0.6, 0.1)$.

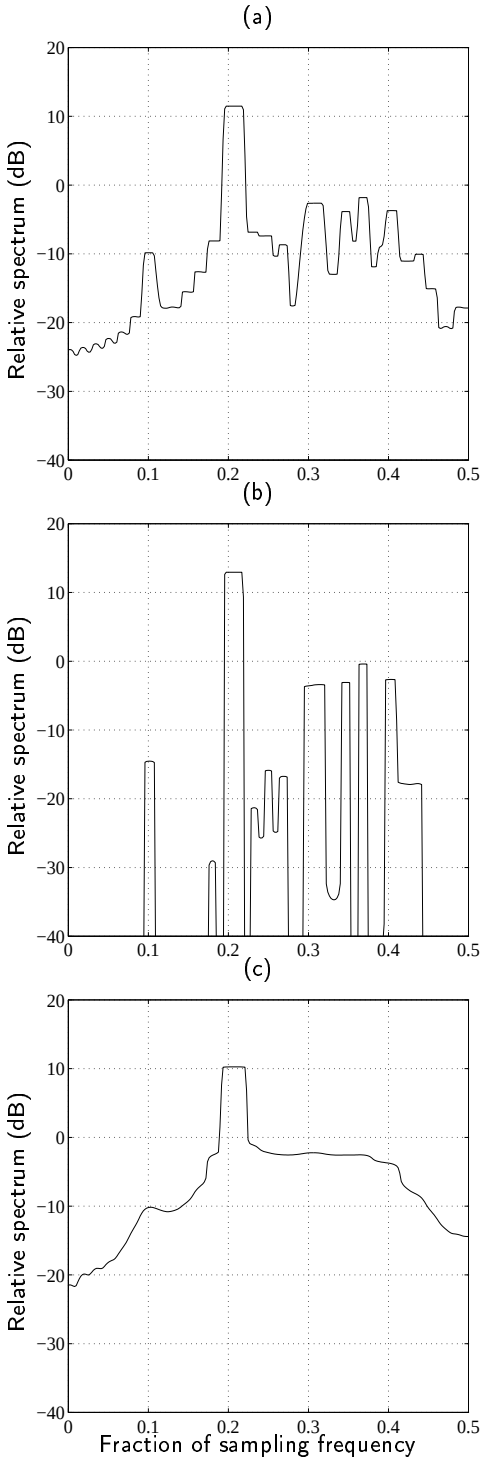


Fig. 4. Smooth spectra reconstructed with a circular Gibbs-Markov penalty function, $(\lambda, \tau_1) = (0.05, 0.001)$; (a) convex case where $\mu = \mu_{\sup} = 0.5,\ \varepsilon_K = 0.9$, (b) convex case where $\mu = \mu_{\sup} = 0.5,\ \varepsilon_K = 0.001$, (c) nonconvex case where $\mu = 5,\ \varepsilon_K = 0.9$.

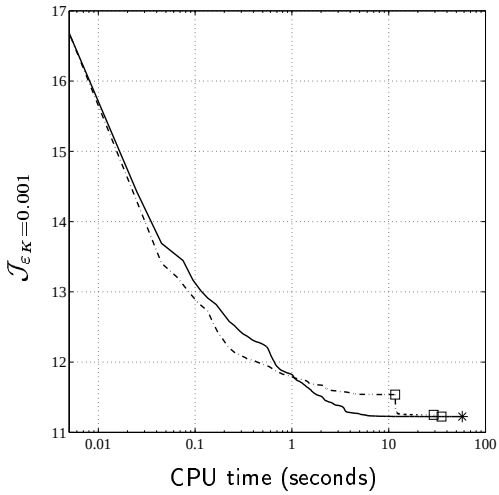


Fig. 5. Performance of the GND algorithm coupled with CG, in the SS case: the solid line corresponds to the minimization of $\mathcal{J}_{0.001}$ in a single run, dashed-dotted lines to the GND process coupled to CG.

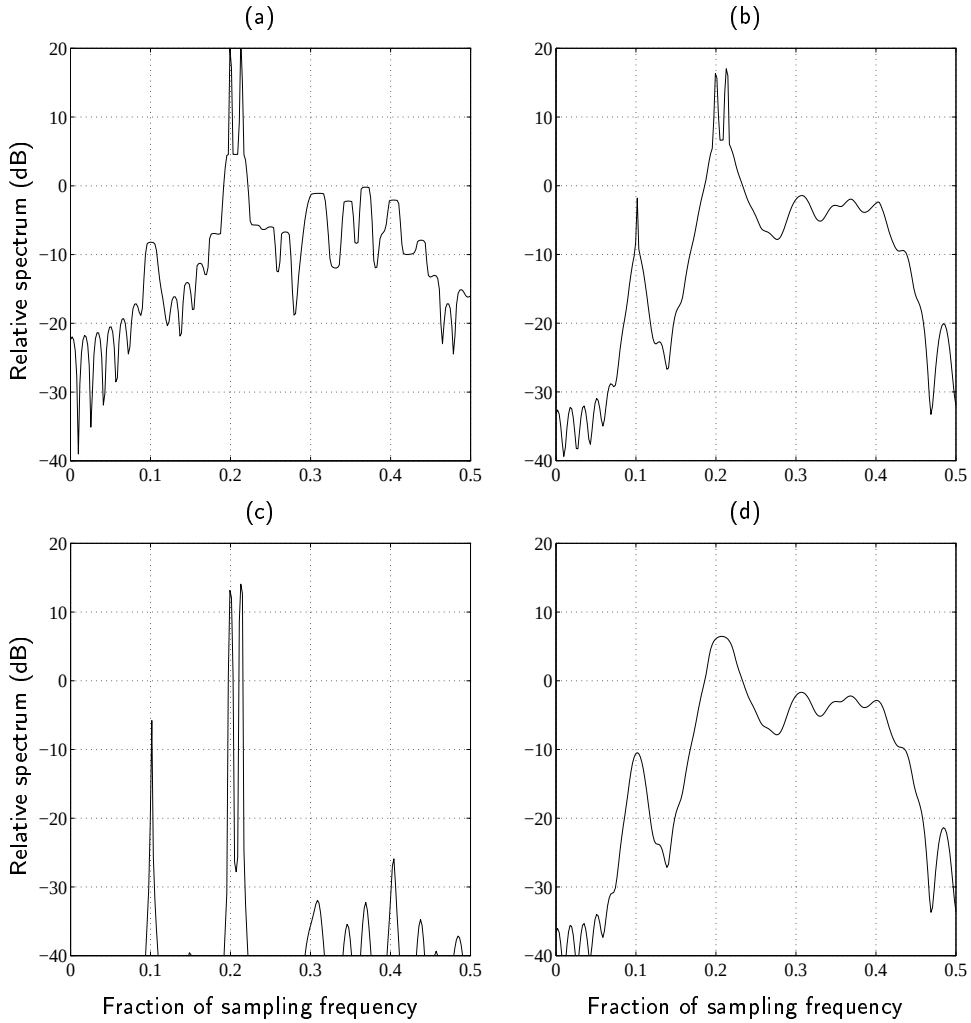


Fig. 6. Mixed spectra. (a): convex case $\mu = 0.5$; (b) nonconvex extension $\mu = 5$; (c) and (d) correspond respectively to the line ($|\widehat{\boldsymbol{X}}_{\mathrm{L}}|^2$) and smooth ($|\widehat{\boldsymbol{X}}_{\mathrm{S}}|^2$) parts of $|\widehat{\boldsymbol{X}}|^2$ depicted in (b).