A half-quadratic block-coordinate descent method for spectral estimation

Philippe Ciuciu, J. Idier

A half-quadratic block-coordinate descent method for spectral estimation

Philippe CIUCIU [a,*] and Jérôme IDIER [b]

[a] *Philippe CIUCIU is with the Commissariat à l'Énergie Atomique (DSV/DRM/SHFJ), 91406 Orsay, cedex France (email:ciuciu@shfj.cea.fr).*

[b] *Jérôme IDIER is with the Laboratoire des Signaux et Systèmes (CNRS–SUPÉLEC–UPS), 91192 Gif-sur-Yvette cedex, France (email:idier@lss.supelec.fr).*

Abstract

In short-time spectral estimation, Sacchi et al. (1998) and Ciuciu et al. (2001) derived new nonlinear spectral estimators defined as minimizers of penalized criteria. The first contributors have introduced separable penalizations for line spectra (LS) recovering, whereas the latter have proposed circular Gibbs-Markov functions for smooth spectra (SS) restoration, and combined both contributions for estimation of *"mixed" spectra* (MS) *i.e.,* frequency peaks superimposed on a homogeneous background (Ciuciu et al., 2001).

Sacchi et al. resorted to the *Iteratively Reweighted Least Squares* (IRLS) algorithm for the minimization stage. Here, we show that IRLS is a block-coordinate descent (BCD) method performing the minimization of a *half-quadratic* (HQ) energy. The latter, derived from Geman and Reynolds's construction, has the same minimizer as the initial criterion but depends on more variables. After proving that such a construction is not available for Gibbs-Markov penalizations, we extend the pioneering work of Geman and Yang (1995) that leads to a suitable HQ energy for any kind of penalization encountered in (Ciuciu et al., 2001). The BCD algorithm used for minimizing such HQ criteria is actually an original *Residual Steepest Descent* (RSD) procedure (Yarlagadda et al., 1985) and thus converges in any convex case. A comparison between RSD, IRLS when available, and a pseudo-conjugate gradient algorithm is addressed in any case.

Key words: Spectral estimation; half-quadratic regularization; Iteratively Reweighted Least Squares; Residual Steepest Descent; Legendre transform.

* Corresponding author. Tel: (33) 1 69 86 78 13, Fax: (33) 1 69 86 77 86.

1 Introduction

1.1 Penalized criteria

Nonparametric short-time spectral estimation consists in retrieving an estimate of the power spectrum from a short set of observations using the discrete Fourier transform (DFT) (Sacchi et al., 1998; Ciuciu et al., 2001). The goal is to estimate a large number of Fourier coefficients $\boldsymbol{x} \in \mathbb{C}^P$ of a time series, partially observed through the data $\boldsymbol{y} \in \mathbb{C}^N$:

$$\boldsymbol{y} = W_{NP}\, \boldsymbol{x}, \tag{1}$$

where $W_{NP} = [w_0^{np}]$ stands for the $N \times P$ inverse Fourier matrix, with $w_0 = \exp(2j\pi/P)$, $n \in \mathbb{N}_N$, $p \in \mathbb{N}_P$ and $\mathbb{N}_k = \{0, 1, \ldots, k-1\}$. Since $N \ll P$, system (1) is underdetermined, and there exists an infinite number of solutions for (1), *i.e.,* of minimizers of $\mathcal{Q}(\boldsymbol{x}) = \|\boldsymbol{y} - W_{NP}\boldsymbol{x}\|^2$. To cope with the ill-posedness of this problem, penalized approaches have been proposed (Ciuciu et al., 2001; Giovannelli and Idier, 2001; Sacchi et al., 1998; Cabrera and Parks, 1991). In particular, Ciuciu et al. (2001); Sacchi et al. (1998) have defined a nonlinear estimator of the spectral amplitudes as

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x} \in \mathbb{C}^P} \mathcal{J}(\boldsymbol{x}), \tag{2}$$

$$\text{where}, \quad \mathcal{J}(\boldsymbol{x}) = \mathcal{Q}(\boldsymbol{x}) + \lambda \mathcal{R}(\boldsymbol{x}). \tag{3}$$

The hyperparameter $\lambda > 0$ controls the trade-off between the closeness to data, measured by $\mathcal{Q}$, and the confidence in structural prior modeled by $\mathcal{R}$. The power spectrum estimator easily deduces as the vector of the squared modulus of the components of $\widehat{\boldsymbol{x}}$.

The reference (Sacchi et al., 1998) adopts the classical Bayesian interpretation of $\widehat{\boldsymbol{x}}$ as a maximum *a posteriori* estimate, derived from an independent and circular Cauchy prior model. The Cauchy density function is a *heavy-tailed* probability distribution. For this reason, it is well-suited for restoration of parcimonious frequency peaks. It is also suggested to choose $\lambda \searrow 0$, (at least in the *accurate data case*), in which case $\widehat{\boldsymbol{x}}$ is the constrained minimizer of $\mathcal{R}$ subject to (1).

In (Ciuciu et al., 2001), the methodology is generalized in order to encompass the smooth and "mixed" spectra (resp. SS and MS) problems. In any case, $\mathcal{R}$

is

- circular: $\mathcal{R}(\boldsymbol{x}) = \mathcal{R}(\boldsymbol{\rho})$ with $\rho_p = |x_p|$ and $\boldsymbol{\rho} \in \mathbb{R}_+^P$. (4a)
- strictly convex, (4b)
- continuously differentiable (C^1), (4c)
- "infinite at infinity", *i.e.,* $\lim_{\|\boldsymbol{x}\| \to \infty} \mathcal{R}(\boldsymbol{x}) = \infty$. (4d)

As a consequence, $\mathcal{J}$ is strictly convex as a sum of convex and strictly convex terms. Then, the minimizer $\widehat{\boldsymbol{x}}$ is unique and continuous w.r.t. the data (Bouman and Sauer, 1993); this guarantees the well-posedness of the regularized problem (Tikhonov and Arsenin, 1977). Constraints (4b)-(4d) make the computation of $\widehat{\boldsymbol{x}}$ feasible by many deterministic descent method (such as gradient-based methods, IRLS, etc).

The main contribution of this paper is to propose a special class of block-coordinate descent (BCD) methods and to show that it is competitive with a pseudo-conjugate gradient (PCG) algorithm in SS and MS cases, which is even more efficient for LS recovering.

1.2 Half-quadratic BCD methods

A BCD optimization algorithm is a multivariate extension of a coordinate descent method, *i.e.,* it minimizes a criterion w.r.t. blocks of variables (Bertsekas, 1995). BCD methods have recently become popular (Charbonnier et al., 1994; Vogel and Oman, 1996; Charbonnier et al., 1997; Vogel and Oman, 1998; Delaney and Bresler, 1998) in image restoration or reconstruction, in conjonction with the *half-quadratic* (HQ) formulation of regularized criteria (Geman and Reynolds, 1992; Geman and Yang, 1995).

On the one hand, to make the paper self-contained, we first recall the basic principles of HQ regularization. Then, we provide useful details that refer to convex duality (Rockafellar, 1970) (see Section 2). Starting from a non-quadratic criterion $\mathcal{J} = \mathcal{Q} + \lambda\mathcal{R}$ with $\mathcal{Q}$ quadratic in $\boldsymbol{x}$, HQ regularization amounts to deriving a new objective function $\mathcal{K}$, depending on additional variables $\boldsymbol{b}$, such that

$$\mathcal{K}(\boldsymbol{x}, \boldsymbol{b}) = \mathcal{Q}(\boldsymbol{x}) + \lambda\mathcal{S}(\boldsymbol{x}, \boldsymbol{b}), \tag{5a}$$

$$\text{with} \quad \inf_{\boldsymbol{b}} \mathcal{S}(\boldsymbol{x}, \boldsymbol{b}) = \mathcal{R}(\boldsymbol{x}). \tag{5b}$$

Hereafter, half-quadratic means that $\mathcal{S}$, and then $\mathcal{K}$, are quadratic in $\boldsymbol{x}$ when $\boldsymbol{b}$ is fixed and not jointly quadratic in $(\boldsymbol{x}, \boldsymbol{b})$. Since $\mathcal{K}$ is quadratic in $\boldsymbol{x}$, its minimization w.r.t. $\boldsymbol{x}$ only requires to solve a linear system. Moreover, explicit duality relations and separability of $\mathcal{S}$ in $\boldsymbol{b}$ (Rockafellar, 1970) allow to

straightforwardly perform the optimization step w.r.t. $\boldsymbol{b}$. Technical conditions have been proposed by (Charbonnier et al., 1997; Aubert and Vese, 1997; Idier, 2001) for proving that $\mathcal{K}$ and $\mathcal{J}$ have the same global minimizer. Then, a *HQ BCD algorithm i.e.,* a BCD method applied to $\mathcal{K}$, can be more attractive than a coordinate descent algorithm working on $\mathcal{J}$.

On the other hand, IRLS is a Reweighted Least Squares technique that has been recently applied to LS recovering (Sacchi et al., 1998). Following Idier (2001), it is shown in Section 2.1 that IRLS identifies with the so-called ARTUR algorithm (Charbonnier et al., 1997). The latter is a HQ BCD method derived from Geman and Reynolds's construction. This interpretation provides simple convergence criteria of IRLS given the existing results for ARTUR (Charbonnier et al., 1997; Idier, 2001).

In Section 3, it is established that IRLS/ARTUR has no natural extension to cope with SS and MS cases, in the sense that mathematical conditions for deriving $\mathcal{S}^{\text{GR}}$ are not fulfilled[1]. Consequently, the main contribution of this paper is devoted to propose another HQ development, adapted to these situations. More precisely, our contribution is a multivariate extension of Geman and Yang's work. The resulting HQ BCD method is nothing but a modified RSD algorithm, already used in seismic deconvolution (Yarlagadda et al., 1985), and also referred to as LEGEND in computed imaging (Charbonnier et al., 1994). For the LS case, the presentation of the HQ regularizing term $\mathcal{S}^{\text{GY}}$ is reported to Section 2.2. For SS and MS restoration, the augmented cost functions $\mathcal{S}^{\text{GY}}$ of the penalizations $\mathcal{R}$ encountered in (Ciuciu et al., 2001) are exhibited in Section 3 and Section 4, respectively. Then, the minimization of the augmented criterion $\mathcal{K}^{\text{GY}}$ is performed with an original RSD algorithm. Following Idier (2001), sufficient properties of $\mathcal{K}^{\text{GY}}$ are derived to guaranty convergence towards $\widehat{\boldsymbol{x}}$ of the RSD procedure.

Finally, the last concern addressed in Section 5 is to increase the speed of convergence of the proposed RSD method according to an over-relaxation scheme on $\boldsymbol{x}$ and $\boldsymbol{b}$. Then, RSD is compared to ARTUR/IRLS in the LS case, and to a PCG algorithm in all cases. Concluding remarks are drawn in Section 6.

[1] In the following, the superscripts "GR" and "GY" stand for Geman and Reynolds and Geman and Yang, respectively.

2 HQ solutions to LS restoration

2.1 HQ interpretation of IRLS

In (Sacchi et al., 1998; Ciuciu et al., 2001), a *shift-invariant circular separable* penalization is considered for line spectra estimation:

$$\mathcal{R}_{\text{L}}(\boldsymbol{x}) = \sum_{p=0}^{P-1} R_0(\rho_p), \tag{6}$$

where $R_0 : \mathbb{R}_+ \mapsto \mathbb{R}_+$, and the subscript "L" stands for *Line*. Different potential functions have been investigated for choosing R_0. Sacchi et al. have selected a log-Cauchy function, $R_0(\rho) = \ln(1 + \rho^2/2\tau_0^2)$, whereas Ciuciu et al. (2001) have retained a component of the following set:

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

With $R_0 \in \mathcal{D}$, the global criterion $\mathcal{J}$ clearly fulfills (4b). 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 that would be oversmoothed by quadratic penalization.

In (Sacchi et al., 1998), IRLS is implemented to minimize $\mathcal{J}(\boldsymbol{x})$. Firstly, a *reweighting* diagonal matrix $\boldsymbol{Q}$ of size $P \times P$ is introduced. Its diagonal entries are defined by

$$\forall\, p \in \mathbb{N}_P, \quad Q_{pp} = 2\rho_p / R_0'(\rho_p). \tag{7}$$

Such a definition is extended by continuity for the case $\rho_p = 0$. Taking derivatives of $\mathcal{J}$ and equating to zero gives the implicit solution (see (Sacchi et al., 1998) for details):

$$\widehat{\boldsymbol{x}} = \left(W_{NP}^\dagger W_{NP} + \lambda \boldsymbol{Q}^{-1} \right)^{-1} W_{NP}^\dagger \boldsymbol{y}, = \boldsymbol{Q} W_{NP}^\dagger (\lambda I_N + W_{NP} \boldsymbol{Q} W_{NP}^\dagger)^{-1} \boldsymbol{y}, \tag{8}$$

where I_N stands for the $N \times N$ identity matrix. Since $\boldsymbol{Q}$ depends on $\boldsymbol{x}$, (8) is a nonlinear system, which can be solved iteratively using IRLS. The latter consists in repeating threefold iterations until convergence, after choosing $\boldsymbol{x}^{(0)}$:

- $IRLS_1$: Compute matrix $\boldsymbol{Q}^{(i)}$ from $\boldsymbol{x}^{(i)}$,

- $IRLS_2$: Solve the $N \times N$ Toeplitz system :

$$\left(\lambda I_N + W_{NP} \boldsymbol{Q}^{(i)} W_{NP}^{\dagger}\right) \boldsymbol{z}^{(i)} = \boldsymbol{y}, \tag{9}$$

- $IRLS_3$: Compute the DFT $\boldsymbol{x}^{(i+1)} = \boldsymbol{Q}^{(i)} W_{NP}^{\dagger} \boldsymbol{z}^{(i)}$,

where $IRLS_2$ can be implemented with a fast solver like Levinson's recursion. As it appears in (Yarlagadda et al., 1985), Byrd and Payne showed that the IRLS algorithm is globally convergent for convex functions R_0 that satisfy fairly weak conditions, *i.e.,* $R_0'(\rho)/\rho$ must be nonincreasing and bounded on $\mathbb{R}_+$. Since the log-Cauchy potential involved in (Sacchi et al., 1998) is not convex, IRLS is not ensured to converge to the global minimizer $\widehat{\boldsymbol{x}}$.

The purpose of the following is to identify the IRLS algorithm with a HQ BCD method. To this end, the HQ extension $\mathcal{S}_{\mathrm{L}}^{\mathrm{GR}}$ of the penalization $\mathcal{R}_{\mathrm{L}}$ is introduced.

Under the theoretical setting of (Idier, 2001), the stress is put on functions R_0 that satisfy the following hypotheses:

$$\begin{aligned} &\bullet\ R_0 \text{ is even, } C^0 \text{on } \mathbb{R} \text{ and } C^1 \text{on } \mathbb{R}^* = \mathbb{R} \backslash \{0\}, \\ &\bullet\ R_0(\sqrt{\cdot}) \text{ is strictly concave on } \mathbb{R}_+, \\ &\bullet\ \lim_{\rho \to \infty} R_0(\rho)/\rho^2 = 0. \end{aligned} \tag{10}$$

Remark that the log-Cauchy potential as well as the functions in $\mathcal{S}$ fulfill (10). Then, it can be shown from convex duality that R_0 reads

$$\begin{aligned} R_0(\rho) &= \inf_{b \in \mathbb{R}_+} \left(b\rho^2 + \psi(b)\right), \\ \text{where}, \quad \psi(b) &= \sup_{\rho \in \mathbb{R}_+} \left(R_0(\rho) - b\rho^2\right) \end{aligned} \tag{11}$$

is convex and C^1 on $\mathbb{R}_+^*$. Such a derivation of HQ energy was first introduced by Geman and Reynolds, without explicit reference to convex duality.

Let

$$\mathcal{S}_{\mathrm{L}}^{\mathrm{GR}}(\boldsymbol{x}, \boldsymbol{b}) = \sum_{p=0}^{P-1} \left(b_p \left|x_p\right|^2 + \psi(b_p)\right), \tag{12}$$

be the augmented regularizing term of (6) with $\boldsymbol{b} \in \mathbb{R}_+^P$. Then, (11) implies (5b) for $\mathcal{S} = \mathcal{S}_{\mathrm{L}}^{\mathrm{GR}}$, and the new objective function $\mathcal{K}_{\mathrm{L}}^{\mathrm{GR}}$, defined by (5a) and $\mathcal{S} = \mathcal{S}_{\mathrm{L}}^{\mathrm{GR}}$, also reads

$$\mathcal{K}_{\mathrm{L}}^{\mathrm{GR}}(\boldsymbol{x}, \boldsymbol{b}) = \boldsymbol{x}^{\dagger} \Lambda(\boldsymbol{b}) \boldsymbol{x} - 2\Re(\boldsymbol{x}^{\dagger} W_{NP}^{\dagger} \boldsymbol{y}) + \Psi(\boldsymbol{b}), \tag{13}$$

where $\Re$ is the real part operator and

$$\begin{cases} \Lambda(\boldsymbol{b}) = W_{NP}^{\dagger} W_{NP} + \lambda \, \mathrm{diag}\,[\boldsymbol{b}] \\ \Psi(\boldsymbol{b}) = \sum_{p=0}^{P-1} \psi(b_p). \end{cases}$$

The HQ BCD algorithm devoted to the minimization of $\mathcal{K}_{\mathrm{L}}^{\mathrm{GR}}$ is referenced to as BCD-GR in the following. Each iteration is composed of two steps. On the one hand, the auxiliary variables $\boldsymbol{b}$ are noninteracting, allowing then a parallelized calculation of the minimizer $\widehat{\boldsymbol{b}}(\boldsymbol{x})$ of $\mathcal{K}_{\mathrm{L}}^{\mathrm{GR}}$. According to (12), the updated value for each component $\widehat{b}_p$ is given by

$$\widehat{b}(x_p) = (\psi')^{-1}(-\rho_p^2) = \frac{R_0'(\rho_p)}{2\rho_p} = Q_{pp}^{-1}. \tag{14}$$

The last but one equality in (14) is obtained from convex duality (Rockafellar, 1970).

On the other hand, computing the minimizer $\widehat{\boldsymbol{x}}(\boldsymbol{b})$ of $\mathcal{K}_{\mathrm{L}}^{\mathrm{GR}}$ amounts to solving the $P \times P$ Toeplitz system

$$\widehat{\boldsymbol{x}}(\boldsymbol{b}) = \Lambda\Big(\boldsymbol{b}\Big)^{-1} W_{NP}^{\dagger} \boldsymbol{y},$$

which can be rewritten as (8) since $\boldsymbol{Q} = \mathrm{diag}\Big[\boldsymbol{b}\Big]^{-1}$ according to (14).

After setting $\boldsymbol{x}^{(0)}$, BCD-GR repeats the following iterative scheme until convergence:

- *BCD-GR*$_1$: Minimization of $\mathcal{K}_{\mathrm{L}}^{\mathrm{GR}}$ w.r.t. $\boldsymbol{b}$:

$$\boldsymbol{b}^{(i)} = \widehat{\boldsymbol{b}}(\boldsymbol{x}^{(i-1)}) = \Big[\ldots, \widehat{b}(x_k^{(i-1)}), \ldots\Big]^{\mathrm{t}}, \; k \in \mathbb{N}_P, \quad \text{(see (14))}.$$

- *BCD-GR*$_2$: Minimization of $\mathcal{K}_{\mathrm{L}}^{\mathrm{GR}}$ w.r.t. $\boldsymbol{x}$:

$$\boldsymbol{x}^{(i)} = \widehat{\boldsymbol{x}}(\boldsymbol{b}^{(i)}), \quad \text{(see (8))}.$$

Given the definition of $\boldsymbol{Q}$, *BCD-GR*$_1$ clearly corresponds to $IRLS_1$, whereas *BCD-GR*$_2$ may be implemented by $IRLS_2$–$IRLS_3$. Finally, both algorithms, IRLS and BCD-GR (known as ARTUR in (Charbonnier et al., 1997)), compute the same solution (2).

This result yields simple convergence criteria for IRLS using well-known results on convergence of BCD methods (Bertsekas, 1995; Ortega and Rheinboldt, 1970); indeed, provided that R_0 is strictly convex, Charbonnier et al. (1997); Idier (2001) have proved the convergence of ARTUR to the global minimizer

$\widehat{\boldsymbol{x}}$ of $\mathcal{J}$. Such a result is slightly less restrictive than convergence conditions of IRLS derived by Byrd and Payne.

Hereafter, another HQ development is shown off for a major reason. Geman and Reynolds's construction fails to provide an augmented HQ criterion $\mathcal{S}_{\mathrm{S}}^{\mathrm{GR}}$ coupled to a Gibbs-Markov energy $\mathcal{R}_{\mathrm{S}}$ for which (5b).

2.2 Generalization of Geman and Yang's construction

2.2.1 Principle

First, the *scalar* construction of HQ criteria introduced by Geman and Yang is reviewed (see also LEGEND in (Charbonnier et al., 1994)). For the restoration of a real-valued image $\boldsymbol{x}$, observed through $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + noise$, the following nonquadratic cost function is considered

$$\mathcal{J}(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|^2 + \lambda \sum_{c \in \mathcal{C}} \phi(\boldsymbol{d}_c^{\mathrm{t}} \boldsymbol{x}), \qquad \boldsymbol{x} \in \mathbb{R}^K,$$

where $\boldsymbol{d}_c \in \mathbb{R}^K$ are known vectors, such as finite differences, and $\mathcal{C}$ is a finite set ($|\mathcal{C}| = M$). Geman and Yang resort to the scalar convex conjugate (Rockafellar, 1970) of the function $x^2/2 - \phi(x)$ in order to get:

$$\phi(x) = \inf_{b \in \mathbb{R}} \left(\frac{1}{2}(x-b)^2 + \zeta(b) \right), \tag{15}$$

where

$$\zeta(b) = \sup_{x \in \mathbb{R}} \left(-\frac{1}{2}(x-b)^2 + \phi(x) \right).$$

From (15), it is straightforward to derive a new objective function $\mathcal{K}^{\mathrm{GY}}(\boldsymbol{x}, \boldsymbol{b})$ with $\boldsymbol{b} = (b_c) \in \mathbb{R}^M$, defined by:

$$\mathcal{K}^{\mathrm{GY}}(\boldsymbol{x}, \boldsymbol{b}) = \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|^2 + \lambda \sum_{c \in \mathcal{C}} \left(\frac{1}{2}(\boldsymbol{d}_c^{\mathrm{t}} \boldsymbol{x} - b_c)^2 + \zeta(b_c) \right).$$

$\mathcal{K}^{\mathrm{GY}}$ is HQ since the argument $\boldsymbol{d}_c^{\mathrm{t}} \boldsymbol{x}$ of each contribution $\phi(.)$ is a *linear* function of $\boldsymbol{x}$. Then, Geman and Yang proposed to minimize $\mathcal{K}^{\mathrm{GY}}$ rather than $\mathcal{J}$, since $\inf_{\boldsymbol{b} \in \mathbb{R}^M} \mathcal{K}^{\mathrm{GY}}(., \boldsymbol{b}) = \mathcal{J}(.)$.

In the spectral estimation framework, the penalization function $\mathcal{R}$ nonlinearly depends on the sought spectral amplitudes $\boldsymbol{x}$ since it is circular (see (4a)). In the particular case of LS restoration, the penalization $\mathcal{R}_{\mathrm{L}}$ is defined by $\phi(x) = R_0(\rho)$ (and $\boldsymbol{d}_c$ canonical). Then, (15) gives

$$R_0(\rho) = \inf_{b \in \mathbb{R}_+} \left(\frac{1}{2}(\rho - b)^2 + \zeta(b) \right). \tag{16}$$

Clearly, (16) shows that the quantity to be minimized is quadratic in $\rho = |x|$, but not in x, and the resulting criterion $\mathcal{K}_{\mathrm{L}}^{\mathrm{GY}}$ is not HQ.

Since $|x| = h(\Re(x), \Im(x))$, it is sufficient to couple the real and imaginary parts of each spectral amplitude x with a real-valued auxiliary variable, in order to get a satisfactory HQ extension of $\mathcal{R}_{\mathrm{L}}$. This amounts to linking x with a complex auxiliary variable b, provided that a *multivariate* extension of (15) is available. In the following, we turn to this mutivariate Geman and Yang's construction that will be also necssary for deriving the HQ criteria in the SS and MS cases.

2.2.2 Multivariate extension

For a complete overview on *multivariate* convex duality, (Rockafellar, 1970) is an essential reference. Only the necessary tools are reported hereafter.

Definition 1 *Let $f : \mathbb{C}^M \mapsto \mathbb{R}$ be a convex function. The* multivariate convex conjugate *of f is defined by*

$$\forall\, \boldsymbol{v} \in \mathbb{C}^M, \quad f^*(\boldsymbol{v}) = \sup_{\boldsymbol{u} \in \mathbb{C}^M} (\Re(\boldsymbol{v}^\dagger \boldsymbol{u}) - f(\boldsymbol{u})), \tag{17}$$

and it is a convex function on $\mathbb{C}^M$.

Definition 2 *Let (f, g) be a couple of positive real-valued functions on $\mathbb{C}^M$. If*

(a) f is strictly convex,
(b) f is continuous[2] and differentiable throughout $\mathbb{C}^M$,
(c) f and g are the multivariate convex conjugate to each other, i.e., $g = f^*$ and $f = g^*$,

then (f, g) is said a Legendre pair.

From basic results on convex duality (Rockafellar, 1970, § 26), the following proposition can be derived.

Proposition 3 *Let (f, g) be a Legendre pair on $\mathbb{C}^M$, then g is differentiable on $\mathbb{C}^M$ and its gradient mapping is given by $\nabla g = (\nabla f)^{-1}$, or equivalently: $\forall\, \boldsymbol{u}, \boldsymbol{v} \in \mathbb{C}^M$, such that $\boldsymbol{v} = \nabla g(\boldsymbol{u})$, then $\boldsymbol{u} = \nabla f(\boldsymbol{v})$.*

[2] Here, Rockafellar's closed-proper assumption (Rockafellar, 1970, pp. 52,253-254) is replaced by a stronger but simpler continuity condition on $\mathbb{C}^M$.

In the rest of the paper, the following function f_α will be considered for deriving HQ criteria:

$$\forall\, \boldsymbol{u} \in \mathbb{C}^M, \quad f_\alpha(\boldsymbol{u}) = \boldsymbol{u}^\dagger \boldsymbol{u}/2 - \phi_\alpha(\boldsymbol{u}), \tag{18}$$

$$\text{where} \quad \phi_\alpha(\boldsymbol{u}) = \alpha\phi(\boldsymbol{u}), \quad \alpha > 0,$$

$$\text{and} \quad \mathcal{R}(\boldsymbol{x}) = \sum_{p=0}^{P-1} \phi(\boldsymbol{u}_p). \tag{19}$$

Here, $\boldsymbol{u}_p \in \mathbb{C}^M$ is a subvector of $\boldsymbol{x} \in \mathbb{C}^P$. In the following, ϕ is assumed to be twice countinuously differentiable (C^2).

Let f_α^* be the multivariate convex conjugate of f_α and $\zeta_\alpha(\boldsymbol{v}) = f_\alpha^*(\boldsymbol{v}) - \boldsymbol{v}^\dagger \boldsymbol{v}/2$, then, (17) yields

$$\zeta_\alpha(\boldsymbol{v}) = \sup_{\boldsymbol{u} \in \mathbb{C}^M} \left\{ -\frac{\|\boldsymbol{u} - \boldsymbol{v}\|^2}{2} + \phi_\alpha(\boldsymbol{u}) \right\}. \tag{20}$$

Since $\mathcal{R}$ is circular, so are ϕ and f_α, *i.e.*, $f_\alpha(\boldsymbol{u}) = f_\alpha(|\boldsymbol{u}|)$, where $|\boldsymbol{u}| \in \mathbb{R}_+^M$ stands for the vector of the magnitudes of $\boldsymbol{u}$. Then, the following proposition states that ζ_α is also circular.

Proposition 4 *Let $\phi : \mathbb{C}^M \mapsto \mathbb{R}$ be a circular function involved in (18). Then, function ζ_α, defined by (20), is circular.*

Proof See Appendix A.

Given Proposition 4, if ϕ is circular and (f_α, f_α^*) is a Legendre pair, then ϕ_α reads (using Definition 2(c))

$$\phi_\alpha(|\boldsymbol{u}|) = \inf_{\boldsymbol{v} \in \mathbb{C}^M} \left(\frac{\|\boldsymbol{u} - \boldsymbol{v}\|^2}{2} + \zeta_\alpha(|\boldsymbol{v}|) \right), \tag{21}$$

where $|\boldsymbol{v}| \in \mathbb{R}_+^M$ stands for the vector of moduli of $\boldsymbol{v}$. Without strict convexity of f_α, expression (21) does not hold, *i.e.*, ϕ_α is not the infimum of an HQ local energy.

Following Definition 2, the circular function f_α has to fulfill hypotheses (a)-(b). The latter holds given that ϕ is C^2. For proving (a) *i.e.*, strict convexity of f_α, we resort to a result stated in (Ciuciu et al., 2001) that characterizes convex circular functions. For this purpose, *coordinatewise nondecreasing* functions have to be defined.

Definition 5 *A function* $f : \mathbb{R}_+^M \mapsto \mathbb{R}$ *is said* coordinatewise nondecreasing *if and only if* $\forall\, i \in \{1, \ldots, M\}$*:*

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

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

Proposition 6 *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*

(i) *(resp. strictly) convex*
(ii) *and coordinatewise (resp. increasing) nondecreasing.*

Proposition 6 is proved in (Ciuciu et al., 2001, Appendix A). Let us apply it to f_α defined by (18). The resulting convexity conditions of f_α are summarized in the following corollary, where (22) and (23) respectively correspond to conditions (i) and (ii) of Proposition 6.

Corollary 7 *Let* f_α *be defined by (18). Suppose that* ϕ *is circular,* C^2 *and convex on* $\mathbb{C}^M$*. Then,* f_α *is strictly convex if and only if*

$$\forall\, \boldsymbol{m} \in \mathbb{R}_+^M, \quad \forall\, i \in \mathbb{N}_M, \quad \alpha < m_i \left[\partial\phi/\partial m_i(\boldsymbol{m})\right]^{-1}, \tag{22}$$

$$\forall\, \boldsymbol{m} \in \mathbb{R}_+^M, \quad I_M - \alpha \boldsymbol{H}_\phi(\boldsymbol{m}) > 0, \tag{23}$$

where $\boldsymbol{H}_\phi(\boldsymbol{m})$ *stands for the Hessian matrix of* ϕ *at point* $\boldsymbol{m}$*.*

Let $\boldsymbol{b} = [\boldsymbol{v}_0, \ldots, \boldsymbol{v}_{P-1}]^{\mathrm{t}}$. If (22)-(23) are ensured, the HQ extension of $\mathcal{R}$ follows from (19) and (21):

$$\mathcal{S}^{\mathrm{GY}}(\boldsymbol{x}, \boldsymbol{b}) = \frac{1}{2\alpha} \sum_{p=0}^{P-1} \left(\|\boldsymbol{u}_p - \boldsymbol{v}_p\|^2 + 2\, \zeta_\alpha(|\boldsymbol{v}_p|) \right). \tag{24}$$

To complete this part, there remains to formulate two propositions pertaining to global convergence of the proposed RSD method minimizing $\mathcal{K}^{\mathrm{GY}} = \mathcal{Q} + \Lambda\mathcal{S}^{\mathrm{GY}}$. They constitute straightforward multivariate extensions of (Idier, 2001, Theorem 1) and (Idier, 2001, Corollary 1), respectively.

Proposition 8 *Let* $\phi : \mathbb{C}^M \mapsto \mathbb{R}$ *be* C^1 *and convex. Then,* ζ_α *given by (20), is convex if conditions (22)-(23) hold and*

$$\alpha \leqslant \lim_{\|\boldsymbol{x}\| \to \infty} \|\boldsymbol{x}\|^2 / 2\phi(\boldsymbol{x}). \tag{25}$$

Strict convexity of ζ_α requires that ϕ is strictly convex and that inequality (25) is strict.

Proposition 9 *Assume that ϕ meets the conditions of Proposition 8. Then,*

$$\begin{aligned} \phi \textit{ strictly convex} &\implies \mathcal{S}^{GY} \textit{ strictly convex,} \\ f_\alpha \textit{ strictly convex} &\implies \mathcal{S}^{GY}\ C^1. \end{aligned}$$

Proposition 9 shows that $\mathcal{K}^{\mathrm{GY}}$ possesses properties (4b)-(4c) if strictly convex functions are considered for ϕ and f_α. The resulting HQ BCD algorithm converges towards the unique global minimizer $(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{b}})$ (Idier, 2001; Bertsekas, 1995).

2.2.3 Application to the separable case

Here, our aim is to show that a *bivariate* application ($M = 1$) of the proposed multivariate construction provides a HQ extension of $\mathcal{R}_{\mathrm{L}}$, the circular separable penalization encountered for LS recovering. In such a situation, ϕ is defined on $\mathbb{C}$ ($M = 1, \boldsymbol{u}_p = x_p$) by

$$\phi(\boldsymbol{u}_p) = R_0(\rho_p), \tag{26}$$

so that (19) holds.

For $R_0 \in \mathcal{D}$, the global criterion $\mathcal{J}$ satisfies constraints (4b)-(4c). Apply Proposition 6 with $M = 1$, then $\phi = R_0$ is (resp. strictly) convex on $\mathbb{C}$ *if and only if* it is (resp. increasing) nondecreasing on $\mathbb{R}_+$ and (resp. strictly) convex on $\mathbb{R}$. Since $R_0 \in \mathcal{D}$, it is a nondecreasing convex function so that ϕ is convex on $\mathbb{C}$. Since ϕ is defined on $\mathbb{C}$, the present convex conjugacy operation is bivariate and involves a single complex auxiliary variable $\boldsymbol{v}_p = b_p$. In order that (21) holds, f_α has to be strictly convex that can be shown as follows. From (26), ϕ and then f_α are circular. Thus, Corollary 7 is applicable and (22)-(23) take the following form:

$$\begin{cases} \alpha < \min\limits_{\rho \geqslant 0} \left[\rho / R_0'(\rho)\right], \\ \alpha < 1/\max\limits_{\rho \geqslant 0} R_0''(\rho) = 1/R_0''(0), \end{cases} \tag{27}$$

where the last equality deduces from the definition of $\mathcal{D}$.

Example 10 *For LS restoration, the* hyperbolic *potential $R_0(\rho) = \sqrt{\tau_0^2 + \rho^2}$ has been used in (Ciuciu et al., 2001). Then, from (27) we obtain that f_α is strictly convex if and only if $\alpha < \tau_0$.*

Since ϕ is circular, so is ζ_α according to Proposition 4. Then, given $|\boldsymbol{v}_p| = |b_p| = \beta_p \in \mathbb{R}_+$, (21)-(24) read

$$\phi_\alpha(\rho_p) = \inf_{b_p \in \mathbb{C}} \left(\frac{1}{2} |x_p - b_p|^2 + \zeta_\alpha(\beta_p) \right), \tag{28}$$

$$\mathcal{S}_{\text{L}}^{\text{GY}}(\boldsymbol{x}, \boldsymbol{b}) = \frac{1}{2\alpha} \Big(\|\boldsymbol{x} - \boldsymbol{b}\|^2 + 2 \sum_{p=0}^{P-1} \zeta_\alpha(\beta_p) \Big), \tag{29}$$

with $\boldsymbol{b} = [b_0, \ldots, b_{P-1}]^{\text{t}} \in \mathbb{C}^P$. Since a complex auxiliary variable b_p is coupled to any spectral amplitude x_p, $\mathcal{S}_{\text{L}}^{\text{GY}}$ depends on twice more real auxiliary variables than $\mathcal{S}_{\text{L}}^{\text{GR}}$.

When $R_0 \in \mathcal{D}$ and (27) are satisfied, ϕ is linear at infinity, then (25) is automatically ensured. Therefore, Propositions 8-9 apply and both energies $\mathcal{S}_{\text{L}}^{\text{GY}}$ and $\mathcal{K}_{\text{L}}^{\text{GY}} = \mathcal{Q} + \lambda' \mathcal{S}_{\text{L}}^{\text{GY}}$, with $\lambda' = \lambda/\alpha$, are strictly convex and C^1. In particular, for the hyperbolic potential of Example 10, if $\alpha < \tau_0$, $\mathcal{K}_{\text{L}}^{\text{GY}}$ fulfills (4b)-(4c). By contrast, with the log-Cauchy potential used in (Sacchi et al., 1998), no convexity result of $\mathcal{K}_{\text{L}}^{\text{GY}}$ is available, even if the function f_α is circular, C^2 and strictly convex for $\alpha < \tau_0^2$, according to (27).

2.2.4 The RSD algorithm for LS restoration

The different steps of the RSD (or BCD-GY) algorithm for computing line spectra are now detailed.

From (29), $\mathcal{K}_{\text{L}}^{\text{GY}}$ admits the following expression

$$\mathcal{K}_{\text{L}}^{\text{GY}}(\boldsymbol{x}, \boldsymbol{b}) = \boldsymbol{x}^\dagger \Lambda_{\text{L}} \boldsymbol{x} - 2\Re\Big(\boldsymbol{x}^\dagger \xi_{\text{L}}(\boldsymbol{b}) \Big) + \Psi_{\text{L}}(\boldsymbol{b}), \tag{30}$$

where

$$\begin{aligned} \Lambda_{\text{L}} &= W_{NP}^\dagger W_{NP} + \lambda'/2 I_P, \\ \xi_{\text{L}}(\boldsymbol{b}) &= W_{NP}^\dagger \boldsymbol{y} + \lambda' \boldsymbol{b}/2, \\ \Psi_{\text{L}}(\boldsymbol{b}) &= \lambda' \Big(\|\boldsymbol{b}\|^2/2 + \sum_{p=0}^{P-1} \zeta_\alpha(\beta_p) \Big). \end{aligned} \tag{31}$$

On the one hand, the auxiliary variables are updated jointly, since they do not interact. Thanks to Proposition 3, no closed form of ζ_α is necessary to calculate the minimizer $\widehat{\boldsymbol{b}}(\boldsymbol{x})$ of $\mathcal{K}_{\text{L}}^{\text{GY}}$. From the current expression of f_α, each component $\widehat{b}_p$, for $p \in \mathbb{N}_P$, is given by:

$$\widehat{b}(x_p) = f'_\alpha(x_p) = x_p - \alpha \phi'(x_p) = x_p - \alpha\, R'_0(\rho_p)\, x_p/\rho_p. \tag{32}$$

On the other hand, it is shown that the minimizer $\widehat{\boldsymbol{x}}(\boldsymbol{b})$ of $\mathcal{K}_{\mathrm{L}}^{\mathrm{GY}}$ can be computed in the Fourier domain thanks to circularity of Λ_{L}. To this end, remark that Λ_{L} is independent of $\boldsymbol{b}$. Moreover, $W_{NP}^{\dagger}W_{NP}$ is circulant as shown in (Giovannelli and Idier, 2001), which allows to decompose it in the Fourier basis $W_{PP}^{\dagger}$ ($W_{PP}W_{PP}^{\dagger} = W_{PP}^{\dagger}W_{PP} = PI_P$). More precisely, we have $W_{NP}^{\dagger}W_{NP} = W_{PP}^{\dagger}\Sigma W_{PP}$, where the diagonal matrix Σ is only composed of two different eigenvalues, 1 and 0, of respective order N and $P-N$. Therefore, Λ_{L} is circulant, and we get $\Lambda_{\mathrm{L}} = W_{PP}^{\dagger}\Delta_{\mathrm{L}}W_{PP}/P$, with

$$\Delta_{\mathrm{L}} = \left(\begin{array}{c|c} (P+\lambda'/2)\ I_N & 0_{N,P-N} \\ \hline 0_{P-N,N} & \lambda'/2\ I_{P-N} \end{array}\right). \tag{33}$$

Hence, Λ_{L} is invertible and $\Lambda_{\mathrm{L}}^{-1}$ reads

$$\Lambda_{\mathrm{L}}^{-1} = \frac{1}{P}W_{PP}^{\dagger}\Delta_{\mathrm{L}}^{-1}W_{PP}, \tag{34}$$

so that $\widehat{\boldsymbol{x}}(\boldsymbol{b})$ is given by:

$$\begin{aligned}\widehat{\boldsymbol{x}}(\boldsymbol{b}) &= \Lambda_{\mathrm{L}}^{-1}\xi_{\mathrm{L}}(\boldsymbol{b}) = \frac{1}{P}W_{PP}^{\dagger}\Delta_{\mathrm{L}}^{-1}W_{PP}\xi_{\mathrm{L}}(\boldsymbol{b}) \\ &= \frac{1}{P}W_{PP}^{\dagger}\Delta_{\mathrm{L}}^{-1}\left(P\widetilde{\boldsymbol{y}}_P + \frac{\lambda'}{2}W_{PP}\boldsymbol{b}\right),\end{aligned} \tag{35}$$

since $W_{NP}^{+}\boldsymbol{y}$ corresponds to the canonical projection from $\mathbb{C}^N$ onto $\mathbb{C}^P$:

$$W_{NP}^{+}\boldsymbol{y} = W_{PP}^{\dagger}\left[\begin{array}{c} I_N \\ \mathbf{0}_{P-N,N} \end{array}\right]\boldsymbol{y} = W_{PP}^{\dagger}\widetilde{\boldsymbol{y}}_P.$$

The computation of $\widehat{\boldsymbol{x}}(\boldsymbol{b})$ by IRLS (or BCD-GR) required to solve a $N \times N$ Toeplitz system, and the associated normal matrix $\Lambda(\boldsymbol{b}^{(i)})$ was modified during the iterations. By contrast, in the present HQ construction, $\widehat{\boldsymbol{x}}(\boldsymbol{b})$ is obtained after solving a $P \times P$ *circulant* system in the Fourier domain, whose normal matrix is constant in the course of the run. Consequently, the BCD-GY algorithm allow savings of numerical cost at each iteration.

After setting an initial value $\boldsymbol{x}^{(0)}$, the present iterative RSD method works as follows.

- *BCD-GY*$_1^{\mathrm{L}}$: Minimization of $\mathcal{K}_{\mathrm{L}}^{\mathrm{GY}}$ w.r.t. $\boldsymbol{b}$:

$$\boldsymbol{b}^{(i)} = \widehat{\boldsymbol{b}}(\boldsymbol{x}^{(i-1)}) = \left[\ldots, \widehat{b}(x_k^{(i-1)}), \ldots\right]^{\mathrm{t}},\ k \in \mathbb{N}_P, \quad \text{(see (32))}.$$

- *BCD-GY*$_2^{\mathrm{L}}$: Minimization of $\mathcal{K}_{\mathrm{L}}^{\mathrm{GY}}$ w.r.t. $\boldsymbol{x}$:

$$\boldsymbol{x}^{(i)} = \widehat{\boldsymbol{x}}(\boldsymbol{b}^{(i)}), \quad \text{(see (35))}. \tag{36}$$

The main motivation of this part was to introduce multivariate HQ regularization based on Geman and Yang's construction, from which we propose a HQ BCD algorithm different from IRLS. Indeed, for SS restoration, IRLS cannot be implemented whereas this multivariate process gives access to convex HQ criteria, and thus to a BCD-GY convergent method.

3 HQ solution to SS restoration

3.1 Regularizing energy

Denote $\boldsymbol{d}_p$ the pth first-order difference vector: $\boldsymbol{d}_p = \mathbf{1}_{p+1} - \mathbf{1}_p$ for any $p > 0$ and $\boldsymbol{d}_{P-1} = \mathbf{1}_0 - \mathbf{1}_{P-1}$, where $\mathbf{1}_p$ is the pth canonical vector. To retrieve SS estimates, the following circular Gibbs-Markov penalization has been proposed in (Ciuciu et al., 2001)

$$\mathcal{R}_{\text{S}}(\boldsymbol{x}) = \frac{1}{2}\sum_{p=0}^{P-1} l(x_p, x_{p+1}), \tag{37}$$

$$l(x_p, x_{p+1}) = q_p + q_{p+1} + 2\mu R_1(\boldsymbol{d}_p^{\text{t}}\boldsymbol{q}) \tag{38}$$

where the subscript "s" stands for *smooth* and parameter $\mu > 0$ tunes the amount of spectral smoothness. Vector $\boldsymbol{q} = [q_0, q_1, \ldots, q_{P-1}]^{\text{t}} \in \mathbb{R}_+^P$ is a differentiable approximation of $\boldsymbol{\rho}$, $q_p = \varphi_\varepsilon(x_p)$, and φ_ε is the strictly convex potential defined by

$$\varphi_\varepsilon : \mathbb{C} \mapsto \mathbb{R}_+, \quad \varphi_\varepsilon(x) = \sqrt{\varepsilon^2 + |x|^2}, \tag{39}$$

As stated in (Ciuciu et al., 2001, Corol. 2), l and then $\mathcal{R}_{\text{S}}$ satisfy (4b)-(4c) provided that

$$\begin{cases} R_1 \text{ is even and convex,} \\ \mu \leqslant \mu_{\text{sup}} = 1/2R_1'(\infty). \end{cases} \tag{40}$$

Example 11 *In (Ciuciu et al., 2001), simulations for SS restoration have been performed with the hyperbolic funtion* $R_1(\rho) = \sqrt{\tau_1^2 + \rho^2}$*, such the amount of smoothness must not exceed* $\mu_{\text{sup}} = 1/2$ *for ensuring strict convexity of* $\mathcal{R}_S$*.*

In the following, R_1 is even and meets the properties of potentials belonging to $\mathcal{D}$. Then, we first show that Geman and Reynolds's construction is unable to provide a HQ development of the penalization $\mathcal{R}_{\text{S}}$, before exposing a solution based on a multivariate extension of Geman and Yang's HQ regularization.

From the HQ viewpoint, inadequacy of IRLS can be studied as follows. To obtain a HQ extension of $\mathcal{R}_{\mathrm{S}}$, the potential $R_1(\boldsymbol{d}_p^{\mathrm{t}}\boldsymbol{q})$ involved in (38) should read as the infimum of an augmented HQ function. Unfortunately, following the process exposed in Subsection 2.1, we find

$$R_1(\boldsymbol{d}_p^{\mathrm{t}}\boldsymbol{q}) = \inf_{b\in\mathbb{R}_+} \left(b\,(\boldsymbol{d}_p^{\mathrm{t}}\boldsymbol{q})^2 + \psi(b) \right), \tag{41}$$

since R_1 meets conditions (10). Clearly, the augmented energy involved in (41) is quadratic in $\boldsymbol{q}$, but not in $\boldsymbol{x}$. Actually, we have found no modified version of (41) to compute SS estimates with the IRLS algorithm. On the contrary, proper adaptation of RSD is possible as shown now.

3.3 Quadrivariate extension of Geman and Yang's process

Following Section 2.2.3, function ϕ has to be defined. As outlined by Proposition 9, strictly convex functions ϕ provide simple convergence criteria for HQ BCD methods. Assuming that (40) holds, l defined by (38), is convex and hereafter, we set $\phi = l$ since the latter meets the conditions of Corollary 7.

Then, the present function f_α is defined on $\mathbb{C}^2$, which implies that the conjugacy operation at hand is quadrivariate ($M = 2$). Hence, two complex auxiliary variables $\boldsymbol{v} = [b_p^+, b_{p+1}^-]^{\mathrm{t}}$ are coupled to $\boldsymbol{u} = [x_p, x_{p+1}]^{\mathrm{t}}$. This amounts to involving twice more real auxiliary variables in $\mathcal{K}_{\mathrm{S}}^{\mathrm{GY}}$ than in $\mathcal{K}_{\mathrm{L}}^{\mathrm{GY}}$.

The second step for deriving an HQ extension of ϕ is to guaranty strict convexity of f_α (Remember that this key property allows ϕ_α to be expressed as in (21)). According to Proposition 6, the restriction of f_α on $\mathbb{R}_+^2$ has to be strictly convex and coordinatewise increasing. The latter result is shown in the following proposition.

Proposition 12 *Let us denote* $|\boldsymbol{u}_p| = [\rho_p, \rho_{p+1}]^{\mathrm{t}}$, $\boldsymbol{m}_{\boldsymbol{u}_p} = [q_p, q_{p+1}]^{\mathrm{t}}$ *and introduce*

$$\boldsymbol{\varphi}(|\boldsymbol{u}_p|) = \boldsymbol{m}_{\boldsymbol{u}_p}, \quad \textit{and} \quad t_\alpha(\boldsymbol{m}_{\boldsymbol{u}_p}) = \frac{\boldsymbol{m}_{\boldsymbol{u}_p}^\dagger \boldsymbol{m}_{\boldsymbol{u}_p}}{2} - \alpha\,\phi(\boldsymbol{m}_{\boldsymbol{u}_p}),$$

then, given (18) the restriction of f_α *on* $\mathbb{R}_+^2$ *reads*

$$f_\alpha(\boldsymbol{u}_p) = f_\alpha(|\boldsymbol{u}_p|) = t_\alpha(\boldsymbol{m}_{\boldsymbol{u}_p}) + \frac{|\boldsymbol{u}_p|^\dagger |\boldsymbol{u}_p| - \boldsymbol{m}_{\boldsymbol{u}_p}^\dagger \boldsymbol{m}_{\boldsymbol{u}_p}}{2} = t_\alpha \circ \boldsymbol{\varphi}(|\boldsymbol{u}_p|) - \varepsilon^2, \tag{42}$$

and finally f_α is strictly convex on $\mathbb{C}^2$ if

$$\begin{cases} \alpha < \dfrac{\varepsilon}{1+2\mu \max\limits_{\rho \geqslant 0} R_1'(\rho)} = \dfrac{\varepsilon}{1+2\mu\, R_1'(\infty)}, \\ \alpha < \dfrac{1}{4\mu \max\limits_{\rho \geqslant 0} R_1''(\rho)} = \dfrac{1}{4\mu\, R_1''(0)}. \end{cases} \tag{43}$$

Proof see Appendix B.

Note 1 *For the hyperbolic potential R_1 of Example 11, f_α is strictly convex if $\alpha < \varepsilon/(1+2\mu)$ and $\alpha < \tau_1/4\mu$.*

Since ϕ is circular ($\phi(\boldsymbol{u}_p) = \phi(|\boldsymbol{u}_p|)$), so is ζ_α according to Proposition 4. Let us denote $|\boldsymbol{v}_p| = [|b_p^+|, |b_{p+1}^-|]^{\mathrm{t}} = [\beta_p^+, \beta_{p+1}^-]^{\mathrm{t}}$, (21) and (24) are given by

$$\phi_\alpha(q_p, q_{p+1}) = \inf_{(b_p^+, b_{p+1}^-) \in \mathbb{C}^2} \left(\frac{|x_p - b_p^+|^2 + |x_{p+1} - b_{p+1}^-|^2}{2} + \zeta_\alpha(\beta_p^+, \beta_{p+1}^-) \right). \tag{44}$$

$$\mathcal{S}_{\mathrm{S}}^{\mathrm{GY}}(\boldsymbol{x}, \boldsymbol{b}) = \frac{1}{2\alpha} \Big(\left\| \boldsymbol{x} - \boldsymbol{b}^+ \right\|^2 + \left\| \boldsymbol{x} - \boldsymbol{b}^- \right\|^2 + 2 \sum_{p=0}^{P-1} \zeta_\alpha(\beta_p^+, \beta_{p+1}^-) \Big) \tag{45}$$

with $\boldsymbol{b} = [\boldsymbol{b}^- \mid \boldsymbol{b}^+]$ the $P \times 2$ complex matrix of auxiliary variables, and $\boldsymbol{b}^\pm = \left[b_0^\pm, b_1^\pm, \ldots, b_{P-1}^\pm \right]^{\mathrm{t}} \in \mathbb{C}^P$ ($b_0^- = b_{P-1}^+$ because of the circularity constraint $x_P = x_0$).

In order to prove convergence of the RSD (or BCD-GY) algorithm working on $\mathcal{K}_{\mathrm{S}}^{\mathrm{GY}} = \mathcal{Q} + \lambda' \mathcal{S}_{\mathrm{S}}^{\mathrm{GY}}$, we resort to Proposition 9 with $M = 2$. Since ϕ and f_α are strictly convex if (40) and (43) hold, respectively, $\mathcal{K}_{\mathrm{S}}^{\mathrm{GY}}$ is strictly convex and C^1. Therefore, we conclude that any BCD method minimizing $\mathcal{K}_{\mathrm{S}}^{\mathrm{GY}}$ converges to the global minimizer $(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{b}})$. The main steps of the RSD algorithm devoted to SS restoration are now highlighten.

3.4 The RSD algorithm for SS restoration

The criterion $\mathcal{K}_{\mathrm{S}}^{\mathrm{GY}}$ is written in the form (30) for the following set $(\Lambda_{\mathrm{S}}, \xi_{\mathrm{S}}, \Psi_{\mathrm{S}})$:

$$\begin{aligned} \Lambda_{\mathrm{S}} &= W_{NP}^\dagger W_{NP} + \lambda' I_P = W_{PP}^\dagger \Delta_{\mathrm{S}} W_{PP}/P, \\ \xi_{\mathrm{S}}(\boldsymbol{b}) &= W_{NP}^\dagger \boldsymbol{y} + \lambda' \Big(\boldsymbol{b}^+ + \boldsymbol{b}^- \Big)/2, \\ \Psi_{\mathrm{S}}(\boldsymbol{b}) &= \lambda' \Big(\left\| \boldsymbol{b}^+ \right\|^2/2 + \left\| \boldsymbol{b}^- \right\|^2/2 + \sum_{p=0}^{P-1} \zeta_\alpha(\beta_p^+, \beta_{p+1}^-) \Big). \end{aligned} \tag{46}$$

Whereas $\mathcal{K}_{\mathrm{L}}^{\mathrm{GR}}$ and $\mathcal{K}_{\mathrm{L}}^{\mathrm{GY}}$ are separable functions of auxiliary variables, b_{p+1}^{-} and b_p^{+} locally interact within $\mathcal{K}_{\mathrm{S}}^{\mathrm{GY}}$. As a consequence, searching for the minimizer $\widehat{\boldsymbol{b}}(\boldsymbol{x})$ of $\mathcal{K}_{\mathrm{S}}^{\mathrm{GY}}$ requires to jointly update b_{p+1}^{-} and b_p^{+} in the core algorithm, in order to preserve a fully parallel scheme. From successively (18) and (42), and given Proposition 3, $\widehat{b}_{p+1}^{-}$ and $\widehat{b}_p^{+}$ read

$$\left\{\begin{array}{rl} \widehat{b}^{-}(x_p, x_{p+1}) &= x_{p+1} - \alpha \left.\dfrac{\partial \phi}{\partial x_{p+1}}\right|_{(x_p, x_{p+1})} \\ &= x_{p+1} - \alpha\, \varphi_\varepsilon'(\rho_{p+1}) \times \left(1 - 2\mu R_1'(q_p - q_{p+1})\right)/2x_{p+1}, \\ \widehat{b}^{+}(x_p, x_{p+1}) &= x_p - \alpha \left.\dfrac{\partial \phi}{\partial x_p}\right|_{(x_p, x_{p+1})} \\ &= x_p - \alpha\, \varphi_\varepsilon'(\rho_p) \times \left(1 + 2\mu R_1'(q_p - q_{p+1})\right)/2x_p. \end{array}\right. \tag{47}$$

Matrix Λ_{S} is circulant and its diagonal representation Δ_{S} in the Fourier basis identifies with (33), where λ' is replaced by its double. Hence, Λ_{S} is invertible, and $\Lambda_{\mathrm{S}}^{-1}$ is given by (34) where Δ_{S}^{-1} is deduced from Δ_{S}. It follows that the minimizer $\widehat{\boldsymbol{x}}(\boldsymbol{b})$ of $\mathcal{K}_{\mathrm{S}}^{\mathrm{GY}}$ is given by

$$\widehat{\boldsymbol{x}}(\boldsymbol{b}) = \frac{1}{P} W_{PP}^{\dagger} \Delta_{\mathrm{S}}^{-1}\Big(P\widetilde{\boldsymbol{y}}_P + \lambda'\Big(\boldsymbol{b}^{+} + \boldsymbol{b}^{-}\Big)/2\Big). \tag{48}$$

After setting $\boldsymbol{x}^{(0)}$, the iterative RSD algorithm works as follows.

- *BCD-GY*$_1^{\mathrm{S}}$: Minimization of $\mathcal{K}_{\mathrm{S}}^{\mathrm{GY}}$ w.r.t. $\boldsymbol{b}$:

$$\begin{aligned} \boldsymbol{b}^{(i)} &= \left[\boldsymbol{b}^{-\,(i)} \mid \boldsymbol{b}^{+\,(i)}\right], \text{ with} \\ \boldsymbol{b}^{-\,(i)} &= \left[\ldots, \widehat{b}^{-}\left(x_{k[P]}^{(i-1)}, x_{(k+1)[P]}^{(i-1)}\right), \ldots\right],\ k = -1, \ldots, P-2, \\ \boldsymbol{b}^{+\,(i)} &= \left[\ldots, \widehat{b}^{+}\left(x_{k[P]}^{(i-1)}, x_{(k+1)[P]}^{(i-1)}\right), \ldots\right],\ k \in \mathbb{N}_P. \end{aligned}$$

  where $[.]$ stands for the modulo operator and $\widehat{b}^{-}(\cdot), \widehat{b}^{+}(\cdot)$ are provided by (47).
- *BCD-GY*$_2^{\mathrm{S}}$: $\boldsymbol{x}^{(i)}$ is still computable in the Fourier domain according to (48): $\boldsymbol{x}^{(i)} = \widehat{\boldsymbol{x}}(\boldsymbol{b}^{(i-1)})$.

Note 2 *For solving (48), only the sum $\boldsymbol{b}^{+} + \boldsymbol{b}^{-}$ is needed. Consequently, the storage of $\boldsymbol{b}^{+}$ may be saved.*

Given both HQ developments of separable and Gibbs-Markov penalty functions, the purpose of the next part is to show that extension to the MS case is straightforward.

4 HQ solution to MS restoration

4.1 The "mixed" model

To retrieve "mixed" spectral distributions, *i.e.,* a small set of frequency peaks embedded in a homogeneous background, Ciuciu et al. (2001) introduces a specific model relating data to unknowns, called the "mixed" model. It supposes that the unknowns vector $\boldsymbol{x} = [\boldsymbol{x}_{\text{L}}^{\text{t}}, \boldsymbol{x}_{\text{S}}^{\text{t}}]^{\text{t}} \in \mathbb{C}^{2P}$ consists of a line portion $\boldsymbol{x}_{\text{L}}$ and a smooth portion $\boldsymbol{x}_{\text{S}}$. The resulting fidelity to data term $\mathcal{Q}_{\text{M}}$ [3] reads:

$$\mathcal{Q}_{\text{M}}(\boldsymbol{x}) = \|\boldsymbol{y} - W_{NP}(\boldsymbol{x}_{\text{L}} + \boldsymbol{x}_{\text{S}})\|^2 = \|\boldsymbol{y} - W_{NP}\boldsymbol{C}\boldsymbol{x}\|^2,$$

where $\boldsymbol{C} = [I_P \,|\, I_P]$ is a $P \times 2P$ circulant matrix. Then, the global regularization function $\mathcal{R}_{\text{M}}$ derived in (Ciuciu et al., 2001) penalizes $\boldsymbol{x}_{\text{L}}$ as for LS estimation and $\boldsymbol{x}_{\text{S}}$ as for SS restoration:

$$\mathcal{R}_{\text{M}}(\boldsymbol{x}) = \lambda_{\text{L}}\,\mathcal{R}_{\text{L}}(\boldsymbol{x}_{\text{L}}) + \lambda_{\text{S}}\,\mathcal{R}_{\text{S}}(\boldsymbol{x}_{\text{S}}), \quad (\lambda_{\text{L}},\, \lambda_{\text{S}} > 0), \tag{49}$$

where $\mathcal{R}_{\text{L}}$ is given by (6) and $\mathcal{R}_{\text{S}}$ by (37). Choosing $\lambda_{\text{L}} \ll \lambda_{\text{S}}$ nullifies $\boldsymbol{x}_{\text{S}}$ since $\mathcal{R}_{\text{S}}$ is made up by a separable penalty term, such as $\mathcal{R}_{\text{L}}$, and a Gibbs-Markov one. Choosing $\lambda_{\text{L}} \gg \lambda_{\text{S}}$ induces the reverse effect. As shown in (Ciuciu et al., 2001), λ_{L} and λ_{S} vary on the same range.

As expected, $\mathcal{R}_{\text{M}}$ is circular *i.e.,* $\mathcal{R}_{\text{M}}(\boldsymbol{x}) = \mathcal{R}_{\text{M}}(\boldsymbol{\rho})$, where $\boldsymbol{\rho} = [\boldsymbol{\rho}_{\text{L}}^{\text{t}}, \boldsymbol{\rho}_{\text{S}}^{\text{t}}]^{\text{t}} \in \mathbb{R}_{+}^{2P}$. In addition, $\mathcal{R}_{\text{M}}$ fulfills (4b)-(4c) as a sum of two strictly convex and C^1 penalty functions, $\mathcal{R}_{\text{L}}$ and $\mathcal{R}_{\text{S}}$. The global criterion $\mathcal{J}_{\text{M}}$, given by

$$\mathcal{J}_{\text{M}}(\boldsymbol{x}) = \mathcal{Q}_{\text{M}}(\boldsymbol{x}) + \mathcal{R}_{\text{M}}(\boldsymbol{x}),$$

also satisfies these properties, and its global minimizer is defined by

$$\widehat{\boldsymbol{x}} = \left[\widehat{\boldsymbol{x}}_{\text{L}}^{\text{t}},\, \widehat{\boldsymbol{x}}_{\text{S}}^{\text{t}}\right] = \underset{\boldsymbol{x}_{\text{L}}, \boldsymbol{x}_{\text{S}}}{\arg\min}\, \mathcal{J}_{\text{M}}(\boldsymbol{x}),$$

Finally, the estimated power spectrum is taken as the vector of the squared moduli of the components of $\widehat{\boldsymbol{x}}_{\text{L}} + \widehat{\boldsymbol{x}}_{\text{S}}$. Hereafter, we examine the HQ extension of $\mathcal{R}_{\text{M}}$.

4.2 HQ mixed criterion

As shown by (49), $\boldsymbol{x}_{\text{L}}$ and $\boldsymbol{x}_{\text{S}}$ do not interact within the penalization $\mathcal{R}_{\text{M}}$, so that deriving its HQ extension is a direct application of Section 2.2.3 and

[3] The subscript "M" stands for *mixed.*

Section 3.3. Provided that $R_0 \in \mathcal{D}$ and (40) holds, functions $\phi_{\mathrm{L}} = R_0$ and $\phi_{\mathrm{S}} = l$ are strictly convex. If in addition conditions (27)-(43) are fulfilled by parameters α_{L} and α_{S}, then ϕ_{L} and ϕ_{S} reread as infima of HQ energies, given by (28) and (44), respectively. As a consequence, expressions (29) and (45) of HQ criteria $\mathcal{S}_{\mathrm{L}}^{\mathrm{GY}}$ and $\mathcal{S}_{\mathrm{S}}^{\mathrm{GY}}$ are available, so that $\mathcal{S}_{\mathrm{M}}^{\mathrm{GY}}$ is defined by

$$\mathcal{S}_{\mathrm{M}}^{\mathrm{GY}}(\boldsymbol{x}, \boldsymbol{b}) = \frac{1}{\alpha_{\mathrm{L}}} \mathcal{S}_{\mathrm{L}}^{\mathrm{GY}}(\boldsymbol{x}_{\mathrm{L}}, \boldsymbol{b}_{\mathrm{L}}) + \frac{1}{\alpha_{\mathrm{S}}} \mathcal{S}_{\mathrm{S}}^{\mathrm{GY}}(\boldsymbol{x}_{\mathrm{S}}, \boldsymbol{b}_{\mathrm{S}}^{\pm})$$

Let $\boldsymbol{b} = [\boldsymbol{b}_{\mathrm{L}} \,|\, \boldsymbol{b}_{\mathrm{S}}^{+} \,|\, \boldsymbol{b}_{\mathrm{S}}^{-}]$ denote the $P \times 3$ matrix of complex auxiliary variables. The HQ objective function $\mathcal{K}_{\mathrm{M}}$ reads

$$\mathcal{K}_{\mathrm{M}}^{\mathrm{GY}}(\boldsymbol{x}, \boldsymbol{b}) = \mathcal{Q}(\boldsymbol{x}) + \lambda'_{\mathrm{L}} \mathcal{S}_{\mathrm{L}}^{\mathrm{GY}}(\boldsymbol{x}_{\mathrm{L}}, \boldsymbol{b}_{\mathrm{L}}) + \lambda'_{\mathrm{S}} \mathcal{S}_{\mathrm{S}}^{\mathrm{GY}}(\boldsymbol{x}_{\mathrm{S}}, \boldsymbol{b}_{\mathrm{S}}^{\pm}) \tag{50}$$

where $\lambda'_{\mathrm{L}} = \lambda_{\mathrm{L}}/\alpha_{\mathrm{L}}$ and $\lambda'_{\mathrm{S}} = \lambda_{\mathrm{S}}/\alpha_{\mathrm{S}}$. From results stated on $\mathcal{S}_{\mathrm{L}}^{\mathrm{GY}}$ and $\mathcal{S}_{\mathrm{S}}^{\mathrm{GY}}$ in Section 2.2.3 and Section 3.3, it is obvious to conclude that $\mathcal{S}_{\mathrm{M}}^{\mathrm{GY}}$ and then $\mathcal{K}_{\mathrm{M}}^{\mathrm{GY}}$ are strictly convex and C^1. As a consequence, the proposed RSD algorithm converges to the global minimizer $(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{b}})$.

4.3 The RSD algorithm for MS restoration

$\mathcal{K}_{\mathrm{M}}^{\mathrm{GY}}$ rereads as (30) for the following set $(\Lambda_{\mathrm{M}}, \xi_{\mathrm{M}}, \Psi_{\mathrm{M}})$:

$$\boldsymbol{\Lambda}_{\mathrm{M}} = \left(\begin{array}{c|c} \Lambda_{\mathrm{L}} & W_{NP}^{\dagger} W_{NP} \\ \hline W_{NP}^{\dagger} W_{NP} & \Lambda_{\mathrm{S}} \end{array} \right), \tag{51}$$

$$\xi_{\mathrm{M}}(\boldsymbol{b}) = \begin{bmatrix} \xi_{\mathrm{L}}(\boldsymbol{b}_{\mathrm{L}}) \\ \xi_{\mathrm{S}}(\boldsymbol{b}_{\mathrm{S}}) \end{bmatrix} = \begin{bmatrix} \widetilde{\boldsymbol{y}}_P + \lambda'_{\mathrm{L}} \boldsymbol{b}_{\mathrm{L}}/2, \\ \widetilde{\boldsymbol{y}}_P + \lambda'_{\mathrm{S}} \left(\boldsymbol{b}_{\mathrm{S}}^{+} + \boldsymbol{b}_{\mathrm{S}}^{-} \right)/2 \end{bmatrix},$$

$$\Psi_{\mathrm{M}}(\boldsymbol{b}) = \lambda'_{\mathrm{L}} \left\| \boldsymbol{b}_{\mathrm{L}} \right\|^2 /2 + \lambda'_{\mathrm{S}}/2 \left(\left\| \boldsymbol{b}_{\mathrm{S}}^{-} \right\|^2 + \left\| \boldsymbol{b}_{\mathrm{S}}^{+} \right\|^2 \right) + \sum_{p=0}^{P-1} \left(\lambda'_{\mathrm{L}} \zeta_{\alpha_{\mathrm{L}}}(\beta_{\mathrm{L},\, p}) + \lambda'_{\mathrm{S}} \zeta_{\alpha_{\mathrm{S}}}(\beta_{\mathrm{S},\, p}^{+}, \beta_{\mathrm{S},\, p+1}^{-}) \right).$$

Matrices Λ_{L} and Λ_{S} are given by (31) and (46), respectively.

On the one hand, the variables $\widehat{\boldsymbol{b}}_{\mathrm{L}}(\boldsymbol{x}_{\mathrm{L}})$ and $\widehat{\boldsymbol{b}}_{\mathrm{S}}^{\pm}(\boldsymbol{x}_{\mathrm{S}})$ can be updated according to (32) and (47), respectively, since they do not interact together.

On the other hand, computing the minimizer $\widehat{\boldsymbol{x}}(\boldsymbol{b})$ of $\mathcal{K}_{\mathrm{M}}^{\mathrm{GY}}$ w.r.t. $\boldsymbol{x}$ is not as easier as in the previous cases since $\boldsymbol{\Lambda}_{\mathrm{M}}$ is not circulant. Nonetheless, $\boldsymbol{\Lambda}_{\mathrm{M}}$ is block symmetric and its diagonal blocks are circulant matrices. As a consequence,

$\boldsymbol{\Lambda}_{\mathrm{M}}\widehat{\boldsymbol{x}} = \xi_{\mathrm{M}}(\boldsymbol{b})$ can be solved in a efficient way, provided that $\boldsymbol{\Lambda}_{\mathrm{M}}^{-1}$ is well-defined. Following Golub and Van Loan (1989), $\boldsymbol{\Lambda}_{\mathrm{M}}$ is invertible *if and only if* Λ_{S} and

$$\boldsymbol{T} = \Lambda_{\mathrm{L}} - W_{NP}^{\dagger} W_{NP} \Lambda_{\mathrm{S}}^{-1} W_{NP}^{\dagger} W_{NP}, \tag{52}$$

are both invertible. First, according to (34) and (46), Λ_{L} and Λ_{S} are invertible and we get $\Lambda_{\mathrm{L}}^{-1} = W_{PP}^{\dagger} \Delta_{\mathrm{L}}^{-1} W_{PP}$ and $\Lambda_{\mathrm{S}}^{-1} = W_{PP}^{\dagger} \Delta_{\mathrm{S}}^{-1} W_{PP}$, respectively. Second, given the circulant structure of Λ_{L}, $\Lambda_{\mathrm{S}}^{-1}$ and $W_{NP}^{\dagger} W_{NP} = W_{PP}^{\dagger} \Sigma W_{PP}$, $\boldsymbol{T}$ is circulant as a sum of product of circulant matrices:

$$\boldsymbol{T} = \frac{1}{P} W_{PP}^{\dagger} \boldsymbol{D}_{1,1}^{-1} W_{PP}, \quad \text{with} \quad \boldsymbol{D}_{1,1}^{-1} \triangleq \Delta_{\mathrm{L}} - \frac{1}{P} \Sigma \Delta_{\mathrm{S}}^{-1} \Sigma.$$

Matrix $\boldsymbol{T}$ is of full rank equal to P, provided that $\lambda'_{\mathrm{L}} \neq 2((P+\lambda'_{\mathrm{S}})^{-1} - P)$. The latter condition is always satisfied since $\lambda'_{\mathrm{L}} > 0$ whereas $(P+\lambda'_{\mathrm{S}})^{-1} - P < 0$. Hence, $\boldsymbol{T}$ and $\boldsymbol{\Lambda}_{\mathrm{M}}$ are invertible. Then, a straight application of the inversion lemma for block matrices (Golub and Van Loan, 1989) provides $\boldsymbol{B}_{\mathrm{M}} \triangleq \boldsymbol{\Lambda}_{\mathrm{M}}^{-1}$: applying the inversion lemma for block matrices (Golub and Van Loan, 1989) to $\boldsymbol{\Lambda}_{\mathrm{M}}$, its inverse, say $\boldsymbol{B}_{\mathrm{M}}$ is given by

$$\begin{aligned}
\boldsymbol{B}_{1,1} &= \boldsymbol{T}^{-1} = \frac{1}{P} W_{PP}^{\dagger}\, \boldsymbol{D}_{1,1}\, W_{PP} \\
\boldsymbol{B}_{2,2} &= \frac{1}{P} W_{PP}\, \boldsymbol{D}_{2,2}\, W_{PP} \quad \text{with} \quad \boldsymbol{D}_{2,2} \triangleq \left(\Delta_{\mathrm{S}} - \frac{1}{P} \Sigma \Delta_{\mathrm{L}}^{-1} \Sigma \right)^{-1} \\
\boldsymbol{B}_{1,2} &= \boldsymbol{B}_{2,1} = -\frac{1}{P^2} W_{PP}^{\dagger}\, \Delta_{\mathrm{L}}^{-1} \Sigma\, \boldsymbol{D}_{2,2} W_{PP}.
\end{aligned}$$

Finally, the structure of $\boldsymbol{B}_{\mathrm{M}}$ suggest thqt $\widehat{\boldsymbol{x}} = \boldsymbol{B}_{\mathrm{M}} \xi_{\mathrm{M}}(\boldsymbol{b})$ is still computable in the Fourier basis.

The next part starts with algorithmic adaptations devoted to accelerate convergence of BCD methods, and continues with an experimental comparison between IRLS, RSD and PCG.

5 Experimental comparisons

5.1 Over-relaxation of $\boldsymbol{x}$ and $\boldsymbol{b}$

As previously seen, IRLS or RSD minimize HQ criteria firstly w.r.t. $\boldsymbol{b}$ and secondly w.r.t. $\boldsymbol{x}$. The second step finds the solution of a linear system. Given the special structure of the normal matrix, either Toeplitz for LS estimation with IRLS, or circulant for LS and SS restoration with RSD, the solution $\widehat{\boldsymbol{x}}(\boldsymbol{b})$ of this linear system is efficiently computed without resorting to an iterative scheme such as Gauss Seidel (GS) algorithm. Normally, to accelerate the

numerical convergence of GS methods over-relaxation is proceeded. Here, we propose to introduce the same process in the following way. After computing $\widehat{\boldsymbol{x}}(\boldsymbol{b})$, over-relaxation consists in defining the new estimate as

$$\boldsymbol{x}^{(i)} = \omega\, \widehat{\boldsymbol{x}}(\boldsymbol{b}) + \bar{\omega}\, \boldsymbol{x}^{(i-1)},$$

where $\bar{\omega} = 1 - \omega$ and $\omega \in (1, 2)$. From our practical experience, $\omega \approx 1.9$ is a relevant choice for reducing the iterations number required for convergence. Practically, we have checked that efficiency of RSD is improved if over-relaxation is performed, not only on $\boldsymbol{x}$, but also on $\boldsymbol{b}$. By contrast, we have observed that overrelaxation on $\boldsymbol{b}$ does not speed up the IRLS algorithm.

In case of LS estimation, over-relaxation on $\boldsymbol{b}$ consists in appending to the updating equation (36) the following calculation in the core RSD algorithm, summarized in Appendix C:

$$b_p^{(i)} = \omega_p^{(i)}\, \widehat{b}(x_p^{(i-1)}) + \bar{\omega}_p^{(i)}\, b_p^{(i-1)}, \tag{53}$$

where $\widehat{b}(\cdot)$ is defined by (32).

For the SS counterpart, the above construction is generalized to

$$\begin{cases} b_p^{-(i)} = \omega_p^{(i)}\, \widehat{b}^{-}(x_{p-1}^{(i-1)}, x_p^{(i-1)}) + \bar{\omega}_p^{(i)} b_p^{-(i-1)}, \\ b_p^{+(i)} = \omega_p^{(i)}\, \widehat{b}^{+}(x_p^{(i-1)}, x_{p+1}^{(i-1)}) + \bar{\omega}_p^{(i)} b_p^{+(i-1)}, \end{cases} \tag{54}$$

where $\widehat{b}^{-}(\cdot)$ and $\widehat{b}^{+}(\cdot)$ are given by (47). The original part of MATLAB code for computing SS is also available for consultation in Appendix C. It is obtained by replacing equations (32)-(35) by (47)-(48).

Implementation of over-relaxation in the MS case mixes (53) and (54).

On the other hand, devising a theoretically converging *over*-relaxed scheme is not obvious in the *nonquadratic* case. In particular, $\omega_p^{(i)} \in (1, 2)$ does not ensure that the iterate (53)–(54) decrease $\mathcal{K}_{\text{L}}$ and $\mathcal{K}_{\text{S}}$, respectively. At stage i, it is possible to find a bound $\widehat{\omega}_p^{(i)}$ for each $\omega_p^{(i)},\ p \in \mathbb{N}_P$ such that $\mathcal{K}_{\text{L}}$ and $\mathcal{K}_{\text{S}}$ are decreased. This can be done analytically if ζ_α is not too complicated (as when the Huber potential is chosen for R_0 (Idier, 2001) for LS recovering), or numerically otherwise. In practice, the resulting schemes provide significantly less iterations to converge, compared to the basic scheme ($\omega_p^{(i)} = 1$). Unfortunately, the gain in CPU time is only marginal, because computing $\widehat{\omega}_p^{(i)}$ for each $(p,\ i)$ is too demanding. Finally, maintaining all $\omega_p^{(i)}$ at the same value empirically chosen in $(1, 2)$ reveals much more efficient. From a practical ground, an even more efficient scheme is as follows:

$$\forall p \in \mathbb{N}_P,\ \omega_p^{(i)} = \omega_0 + \omega_1(1 - \log(2)/\log(1 + i)).$$

At the beginning, the relaxation parameter $\omega_p^{(i)}$ is close to ω_0, and it progressively converges to ω_1. On one hand, we recommend to choose $\omega_0 \approx 1$ in order to avoid slow convergence. Indeed, if the new estimate, for instance $b_p^{(i)}$, is too far from $\widehat{b}(\cdot)$, the global HQ criterion $\mathcal{K}_{\mathrm{L}}$ may increase rather than decrease. On the other hand, ω_1 can be chosen close to 2.

The numerical descents reported in the following for IRLS and RSD correspond to over-relaxed versions since they are the most efficient.

5.2 Simulations results

We present the numerical performances of our RSD algorithms by processing the well-known Kay and Marple (1981) example. These 64-points data sequence constitute an important benchmark for evaluating most spectral estimators. The spectral estimates, computed in (Ciuciu et al., 2001) with $P = 512$ are not reported here.

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

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

where $\boldsymbol{x}^{(i)}$ denotes the solution at the ith iteration of the minimization stage, and $*$ is 1 or 2. Following Vogel and Oman (1996), we have rather chosen the l_1 norm, and the thresholds have been set to $(\alpha_1, \alpha_2, \alpha_3) = (10^{-7}, 10^{-6}, 10^{-6})$.

The same stopping criteria have been adopted for RSD, except that the third condition has not been tested. In all cases, $\boldsymbol{x}^{(0)}$ has been defined by the DFT of the zero-padded data sequence $\widetilde{\boldsymbol{y}}_P$.

5.3 Convergence speed of RSD, IRLS and PCG for LS restoration

Following Ciuciu et al. (2001), the hyperbolic potential $R_0(\rho) = \sqrt{\tau_0^2 + \rho^2}$ has been used to define the circular penalty function $\mathcal{R}_{\mathrm{L}}$ (see (6)). From practical experience, setting α to the upper bound of convexity of f_α, *i.e.,* $\alpha = \tau_0$ (see Example 10), allows to speed up numerical convergence of RSD.

Convergence of IRLS, RSD and PCG is illustrated through two different situations. The first one corresponds to $(\lambda, \tau_0) = (0.06, 0.2)$ and provides

an intermediate spectrum estimate, between the usual peridogram shown in Fig. C.1(a) and the LS estimate depicted in Fig. C.1(b). In such a case, the potential R_0 has two clearly separated areas, a quadratic one between zero and τ_0 and a linear one beyond τ_0. Fig. C.2(a) illustrates the efficiency of RSD since it takes about 4 seconds to compute $\widehat{\boldsymbol{x}}$ on a Pentium III 450 Mhz. The IRLS and PCG algorithms provide the solution after 7 seconds and so RSD is slightly faster on this example. Other simulations, not reported here, have confirmed this standpoint provided that τ_0 is not too small.

[Fig. 1 about here.]

The second situation corresponds to the LS estimate depicted in Fig. C.1(b). In this case, hyperparameters (λ, τ_0) are fixed to $(0.06, 0.002)$, so that R_0 is close to $|\cdot|$, which is nondifferentiable at zero. Clearly, as shown in Fig. C.2(b), a quasi nondifferentiability does not prevent IRLS to converge very quickly, in 11 seconds about. Such a result is not surprising given the well-known ability of IRLS for minimization of mixed L_1 (or L_p) and L_2 norms (Yarlagadda et al., 1985; Ruzinsky and Olsen, 1989), and can also be analyzed through properties of the HQ criteria: even for $R_0 = |\cdot|$ the HQ objective function $\mathcal{K}^{\mathrm{GR}}$ (see (10)) is differentiable. Then, ARTUR/IRLS is not penalized for minimizing $\mathcal{K}^{\mathrm{GR}}$.

By contrast, for minimizing the same energy, RSD and PCG require more than 200 seconds. On one hand, it is well-known that gradient-based algorithms require that $\mathcal{J}$ is C^1 to be convergent. On the other hand, as stated in Section II, handling HQ criteria $\mathcal{K}^{\mathrm{GY}}$ requires that $\mathcal{R}$ is C^1. In practice, the latter condition is almost unsatisfied, so that RSD and PCG converge to $\widehat{\boldsymbol{x}}$ very slowly. To speed up RSD and PCG, we have eventually resorted to the so-called "regularization method" (Bertsekas, 1975; Glowinski et al., 1976), also referenced to as GND (for *Graduated Non Differentiability*) in (Ciuciu et al., 2001). The basic principle of GND is to successively minimize a discrete sequence of convex differentiable approximations that converge toward the original nonsmooth criterion (see (Glowinski et al., 1976, p. 21-22)). In the present context, the original criterion is only nearly nonsmooth, and GND is a twofold iterative process. First, it consists in choosing an initial value of τ_0^0 not too small (*e.g.*, $\tau_0^0 = 0.2$). Second, $\mathcal{J}_{\tau_0^0}$ is minimized and the computed solution $\widehat{\boldsymbol{x}}_{\tau_0^0}$ serves as initialization for the next minimization of a closer approximation $\mathcal{J}_{\tau_0^1=0.02}$. This scheme is repeated until $\mathcal{J}_{\tau_0=0.002}$ is attained.

Simulations on GND are not reported since they do not allow to supplant IRLS. However, they show that coupled schemes GND-RSD and GND-PCG converge faster (60 seconds) than single runs of RSD and PCG (200 seconds).

[Fig. 2 about here.]

The hyperbolic potential $R_1(\rho) = \sqrt{\tau_1^2 + \rho^2}$ has also been chosen to define the smooth part of the penalization $\mathcal{R}_{\text{S}}$. Once again, setting α to the upper bound of convexity of f_α, *i.e.,* $\alpha = \min(\varepsilon/(1+2\mu), \tau_1/4\mu)$ (see Note 1), reveals much more efficient for accelerating numerical convergence.

Fig. C.3 illustrates the numerical descent of RSD and PCG for minimizing criteria $\mathcal{J}_{\text{S}} = \mathcal{Q} + \lambda\mathcal{R}_{\text{S}}$ and $\mathcal{J}_{\text{M}}$, versus the CPU time. The optimized criterion $\mathcal{J}_{\text{S}}$ in Fig. C.3(a) corresponds to the spectrum illustrated by Fig. C.4(a) for which the hyperparameters have been set to $(\lambda, \tau_1, \mu, \varepsilon) = (0.05, 0.001, \mu_{\text{sup}} = 0.5, 0.9)$, where μ_{sup} indicates the upper bound of convexity of $\mathcal{R}_{\text{S}}$ (see (40)). Keeping unchanged $(\tau_0, \tau_1, \varepsilon, \mu)$ and setting $(\lambda_{\text{L}}, \lambda_{\text{S}}) = (0.005, 0.004)$ leads to the mixed criterion $\mathcal{J}_{\text{M}}$ plotted in Fig. C.3(b), whose global minimizer yields the mixed spectrum shown in Fig. C.5(a). From a comptational point of view, it appears in both cases that RSD is competitive with the PCG algorithm, and thus more efficient than a standard steepest descent algorithm where the descent direction is only defined by the gradient. This is not surprising since, as pointed out in (Vogel and Oman, 1998), RSD (as well as IRLS) can be formulated as a (constant step-size) quasi-Newton descent algorithm.

[Fig. 3 about here.]

[Fig. 4 about here.]

Let us remark on Fig. C.3 that the computation of the MS estimate is more time demanding (70 seconds) than that of the SS estimate (20 seconds), since there are more unknown spectral amplitudes and also more auxiliary variables to update for MS restoration. Furthermore, each normal equation requires more multiplications and additions, as shown in Section 4.3.

Finally, the better SS and MS estimates, depicted in Fig. C.4(b) and Fig. C.5(b) have been obtained for $\mu = 10\mu_{\text{sup}}$ and the other parameters unchanged. For such a value of μ, convexity of $\mathcal{J}_{\text{S}}$ and $\mathcal{J}_{\text{M}}$ is not ensured. Then, the computed spectra does not necessarily correspond to a global minimizer. Nonetheless, in terms of numerical cost, the same conclusion as before can be drawn *i.e.,* RSD is an appealing alternative to the well-known PCG algorithm, even if non convexity implies slower numerical convergence.

[Fig. 5 about here.]

6 Conclusion

In the context of LS recovering, we showed that IRLS is in turn a BCD method minimizing a HQ criteria, derived from Geman and Reynolds contruction. Then, we proved that IRLS is the method of choice, *i.e.,* it converges faster than gradient based methods. As a BCD method, simpler convergence results of IRLS than existing ones (Byrd and Payne, 1979) have been stated. Unfortunately, we outlined that IRLS cannot be implemented in SS and MS cases.

Since IRLS failed in such situations, we developed another algorithm to fill this gap. The proposed numerical tool is actually an original RSD method (Yarlagadda et al., 1985), even if it seems to be closer to LEGEND (Charbonnier et al., 1994), since it is a BCD method minimizing a HQ criteria derived from Geman and Yang's construction. Whatever the form of the penalty function, provided that it is convex, convergence of RSD was proved. Then, the performances of RSD were compared to IRLS and PCG. In case of separable regularization, two different conclusions were drawn regarding differentiability of the penalization function. If the latter was *smooth* enough, RSD behaves as IRLS, whereas in the opposite case, RSD behaves as PCG. For SS and MS estimation, we demonstrated that RSD is competitive with PCG. We also highlighted that the computational burden is heavier for MS restoration since there are more variables than in case of SS estimation.

The last concern of our study was devoted to propose over-relaxed schemes of BCD methods, since over-relaxation is normally able to accelerate numerical convergence. Such a procedure was successfully implemented on IRLS and RSD. From our practical experience, it gave the expected effect but IRLS was not sensible to over-relaxation of auxiliary variables, contrary to RSD.

A Proof of Proposition 4

Let $\boldsymbol{u} \in \mathbb{C}^M$ with $u_i = |u_i|\, e^{j\theta_i}$ and $(|u_i|, \theta_i) \in \mathbb{R}_+ \times [0, 2\pi)$, for $i \in \mathbb{N}_M$. Let us also define the vector of phases $\boldsymbol{\theta} = [\theta_0, \theta_1, \ldots, \theta_{M-1}]^{\mathrm{t}} \in [0, 2\pi)^M$. Given that

ϕ is circular, we have $\forall\, \boldsymbol{v} \in \mathbb{C}^M$:

$$\begin{aligned}
\zeta_\alpha(\boldsymbol{v}) &= \sup_{(|\boldsymbol{u}|,\boldsymbol{\theta})\in\mathbb{R}_+^M\times[0,2\pi)^M} \left(\phi_\alpha(|\boldsymbol{u}|) - \sum_{i=0}^{M-1} \frac{\left||u_i|\ e^{j\theta_i} - v_i\right|^2}{2}\right) \\
&= \sup_{|\boldsymbol{u}|\in\mathbb{R}_+^M} \left(\phi_\alpha(|\boldsymbol{u}|) - \sum_{i=0}^{M-1} \inf_{\theta_i\in[0,2\pi)} \frac{\left||u_i|\ e^{j\theta_i} - v_i\right|^2}{2}\right) \qquad \text{(A.1)} \\
&= \sup_{|\boldsymbol{u}|\in\mathbb{R}_+^M} \left(\phi_\alpha(|\boldsymbol{u}|) - \sum_{i=0}^{M-1} \frac{(|u_i| - |v_i|)^2}{2}\right) = \zeta_\alpha(|\boldsymbol{v}|),
\end{aligned}$$

where the infima in (A.1) are reached for $\theta_i = arg(v_i)$.

B Proof of Proposition 12

First, apply a basic theorem regarding the composition of convex functions (Ciuciu et al., 2001, Th. 1) in order to state strict convexity of $f_\alpha(\boldsymbol{m}_{\boldsymbol{u}_p})$: since $\boldsymbol{\varphi}$ is defined from φ_ε, each of its components is strictly convex. On the other hand, following Corollary 7, t_α is a strictly convex and coordinatewise increasing function, if conditions (43) are fulfilled. It follows that $t_\alpha \circ \boldsymbol{\varphi}$ and then $f_\alpha(|\boldsymbol{u}_p|)$ are strictly convex on $\mathbb{R}_+^2$.

Second, since φ_ε is increasing on $\mathbb{R}_+$, so is $f_\alpha(|\boldsymbol{u}_p|)$. Finally, from Proposition 6 f_α is strictly convex on $\mathbb{C}^2$ when conditions (43) hold.

C Optimization algorithm under its Matlab Code form

The following MATLAB-code summarizes both algorithms for computing line and smooth spectra. The respective hyperparameters, (λ, τ_0) and $(\lambda, \tau_1, \mu, \varepsilon)$, are supposed to be set to the values given in Section 5.3 and Section 5.4. The latter parameter α has been chosen close to the value given by (27) (see also Example 10) and by (43) (see also Note 1) according to the adressed case.

0. Define stopping rules

```
NBITER=5e2;alpha1=1e-7;alpha2=1e-6;% Thresholds on J and x
```

1. Initialization:e.g., zero padded periodogram:

```
Ypad=[y;zeros(P-N,1)];x0=fft(Ypad)/N;
omb1=1;omb2=1.95;omx=1.9;
```

2. Choose what kind of spectrum you wish to estimate

```
opt='ls'; % Line Spectra
%opt='ss';% Smooth Spectra
```

3. Save in memory:

```
coef1=2*N/P;ybis=coef1*y;x=x0;
if strcmp(opt,'ls')
    Hpar=[lambda,tau0];alpha=.99*tau0;
    lb=lambda/alpha;coef2=lb;
elseif strcmp(opt,'ss')
    Hpar=[lambda,tau1,mu,epsilon];
    alpha=.99*min(epsilon/(1+2*mu),tau1/(4*mu));
    lb=lambda/alpha;coef2=2*lb;
else
    error('Unrecognized string opt!');
end
coef3=sqrt(N);i=0;DF=1;Dx=1;
```

4. Compute $\mathcal{J}$ with subroutine `fun`.

```
J0= feval('fun',y,x0,Hpar,opt);
```

5. Core algorithm

```
while (i<NBITER) & ((DF>alpha1)|(Dx>alpha2))
```

a. Parallelized update of $\boldsymbol{b}^{(i)}$ or of $(\boldsymbol{b}^{(i)+} + \boldsymbol{b}^{(i)-})$:

```
    g=feval('grad',x,Hpar,opt);
    b1=coef2/lb*x-alpha*g;
    %over-relaxation of b
    omb=omb1+omb2*(1-log(2)/log(i+1));
    b=omb*b1+(1-omb)*b; %1<=omb<=2
```

b.Compute $\xi(\boldsymbol{b}^{(i)})$:

```
    xi=lb*b;Fxi=ifft(xi,P)*coef3;
    Fxi(1:N)=Fxi(1:N)+ybis;
    DMFxi=[coef1*Fxi(1:N),coef2*Fxi(N+1:P)];
```

c. Global update of $\boldsymbol{x}^{(i)}$:

```
    x1=fft(DMFxi,P)/coef3;
    %Over relaxation of x^(i):
    x=omx*x1+(1-omx)*x; %1<=omx<=2
```

d. Compute stopping criteria:

```
    J1=feval('fun',y,x,Hpar,opt);
    DF=(J0-J1)/J0;
    Dx=sum(abs(x-x0))/sum(abs(x));
```

e. Updates:

```
    J0=J1;x0=x;i=i+1;
end
```

Subroutine fun:computing the global criterion $\mathcal{J}$

```
    function J=fun(y,x,Hpar,opt)
    if nargin~=4
        error('Bad number of arguments !');
    end
    N=length(y);
```

```
%Jfd:Fidelity to data term
xt=ifft(x)*sqrt(N); % xt:time series
xt0=xt(1:N); % W_{NP}x
Jfd=sum(abs(y-xt0).^2);
%Jreg:Regularization term
z=abs(x);
if strcmp(opt,'ls')
    Jreg= Hpar(1)*sum(sqrt(z.^2+Hpar(2)^2)-Hpar(2));
elseif strcmp(opt,'ss')
    q=sqrt(z.^2+Hpar(4)^2);
    %symdif.m:[x1;...;xn]-->[x1-x2;...;xn-1-xn;xn-x1]
    Jreg=sqrt(symdif(q).^2+Hpar(2)^2)-Hpar(2));
    Jreg=Hpar(1)*sum(q+Hpar(3)*Jreg);
end
J=Jfd+Jreg;
```

Subroutine grad:computing the gradient of the penalization function $\mathcal{R}$

```
function g=grad(x,Hpar,opt)
if nargin~=3
    error('Bad number of arguments !');
end
z=abs(x);
if strcmp(opt,'ls')
g=x./sqrt(z.^2+Hpar(2)^2);
elseif strcmp(opt,'ss')
    q=sqrt(z.^2+Hpar(4)^2);
    grs=x./q; % gradient of the separable term in (38)
    syz=symdif(z);
    sign1=sign(syz);
    grm1=sign1.*abs(syz);
    grm1 = grm1./sqrt(abs(syz).^2+Hpar(2)^2);
    %symdif2.m:[x1;...;xn]-->[xn-x1;x1-x2;...;xn-1-xn]
    syz2=symdif2(z);
    sign2=sign(syz2);
    grm2=sign2.*abs(syz2);
    grm2 = grm2./sqrt(abs(syz2).^2+Hpar(2)^2);
    g=grs.*(1+Hpar(3).*(grm1-grm2));
end
```

References

Aubert, G., Vese, L., October 1997. A variational method in image recovery. SIAM J. Num. Anal. 34 (5), 1948–1979.

Bertsekas, D., 1975. Nondifferentiable optimization approximation. In: Mathematical Programming Studies. Vol. 3. Balinski, M.L. and Wolfe, P., Amsterdam, The Netherlands, pp. 1–25.

Bertsekas, D. P., 1995. Nonlinear programming. Athena Scientific, Belmont, MA.

Bouman, C. A., Sauer, K. D., July 1993. A generalized Gaussian image model for edge-preserving MAP estimation. IEEE Trans. Image Processing 2 (3), 296–310.

Byrd, R. H., Payne, D. A., June 1979. Convergence of the iteratively reweighted least squares algorithm for robust regression. Tech. Rep. Tech. Rep. 313, The Johns Hopkins Univ., Baltimore, MD.

Cabrera, S. D., Parks, T. W., April 1991. Extrapolation and spectral estimation with iterative weighted norm modification. IEEE Trans. Signal Processing 39 (4), 842–851.

Charbonnier, P., Blanc-Féraud, L., Aubert, G., Barlaud, M., November 1994. Two deterministic half-quadratic regularization algorithms for computed imaging. In: Proc. IEEE ICIP. Vol. 2. Austin, TX, pp. 168–172.

Charbonnier, P., Blanc-Féraud, L., Aubert, G., Barlaud, M., February 1997. Deterministic edge-preserving regularization in computed imaging. IEEE Trans. Image Processing 6 (2), 298–311.

Ciuciu, P., Idier, J., Giovannelli, J.-F., October 2001. Regularized estimation of mixed spectra using a circular Gibbs-Markov model. IEEE Trans. Signal Processing 49 (10), 2202–2213.

Delaney, A. H., Bresler, Y., February 1998. Globally convergent edge-preserving regularized reconstruction: an application to limited-angle tomography. IEEE Trans. Image Processing 7 (2), 204–221.

Geman, D., Reynolds, G., March 1992. Constrained restoration and the recovery of discontinuities. IEEE Trans. Pattern Anal. Mach. Intell. 14 (3), 367–383.

Geman, D., Yang, C., July 1995. Nonlinear image recovery with half-quadratic regularization. IEEE Trans. Image Processing 4 (7), 932–946.

Giovannelli, J.-F., Idier, J., July 2001. Bayesian interpretation of periodograms. IEEE Trans. Signal Processing 49 (7), 1988–1996.

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

Golub, G. H., Van Loan, C. F., 1989. Matrix computations (2nd Edition). The John Hopkins University Press, Baltimore, Maryland.

Idier, J., July 2001. Convex half-quadratic criteria and interacting auxiliary variables for image restoration. IEEE Trans. Image Processing 10 (7), 1001–1009.

Kay, S. M., Marple, S. L., November 1981. Spectrum analysis – a modern perpective. Proc. IEEE 69 (11), 1380–1419.

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

Rockafellar, R. T., 1970. Convex Analysis. Princeton Univ. Press.

Ruzinsky, S. A., Olsen, E. T., February 1989. L_1 and L_∞ minimization via a variant of Karmarkar's algorithm. IEEE Trans. Signal Processing 37 (2), 245–253.

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

Tikhonov, A., Arsenin, V., 1977. Solutions of Ill-Posed Problems. Winston, Washington, DC.

Vogel, R. V., Oman, M. E., January 1996. Iterative methods for total variation denoising. SIAM J. Sci. Comput. 17 (1), 227–238.

Vogel, R. V., Oman, M. E., June 1998. Fast, robust total variation-based reconstruction of noisy, blurred images. IEEE Trans. Image Processing 7 (6), 813–823.

Yarlagadda, R., Bednar, J. B., Watt, T. L., February 1985. Fast algorithms for l_p deconvolution. IEEE Trans. Acoust. Speech, Signal Processing ASSP-33 (1), 174–182.

List of Figures

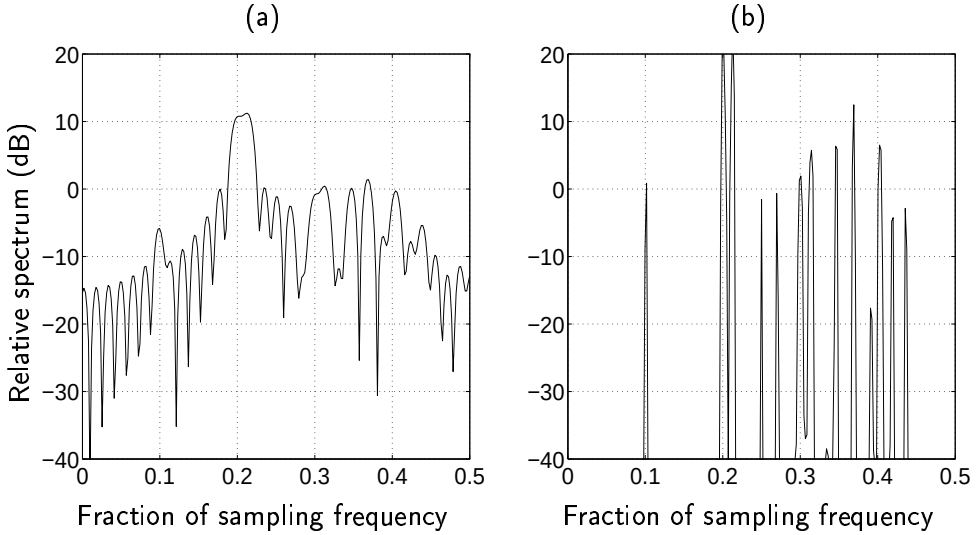

Fig. C.1. Spectra reconstructed with separable regularization. (a): zero-padded periodogram, (b) line spectra reconstructed with the hyperbolic potential, $(\lambda, \tau_0) = (0.06, 0.002)$.

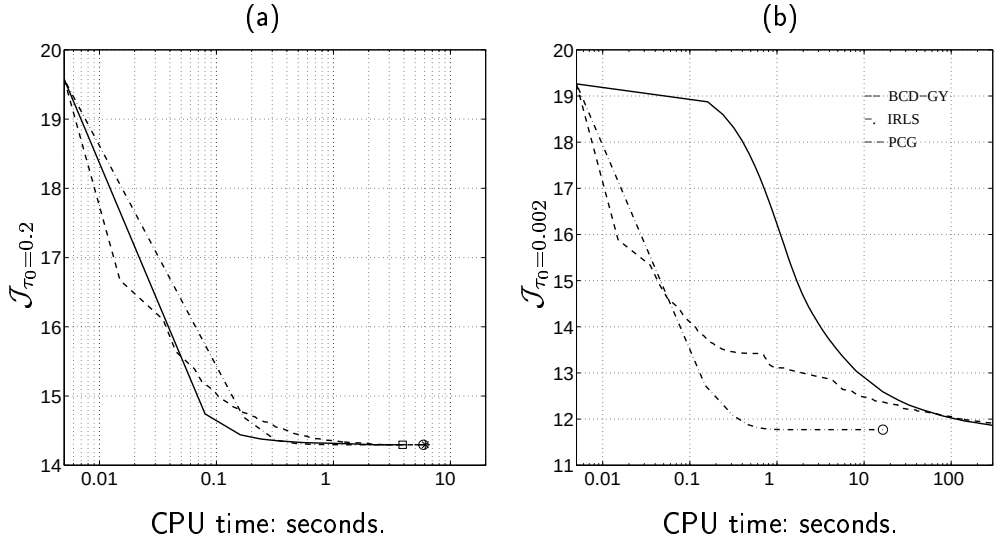

Fig. C.2. Performance of the IRLS, RSD and PCG algorithms for computing "separable" spectra. In (a), $\tau_0 = 0.2$ whereas in (b) $\tau_0 = 0.002$. Solid lines are for RSD, dash-dotted lines encode minimization with IRLS and dashed lines indicate that minimization is performed with PCG. Circles ($\circ$), squares ($\square$) and stars ($*$) depict the stopping points of IRLS, RSD and PCG.

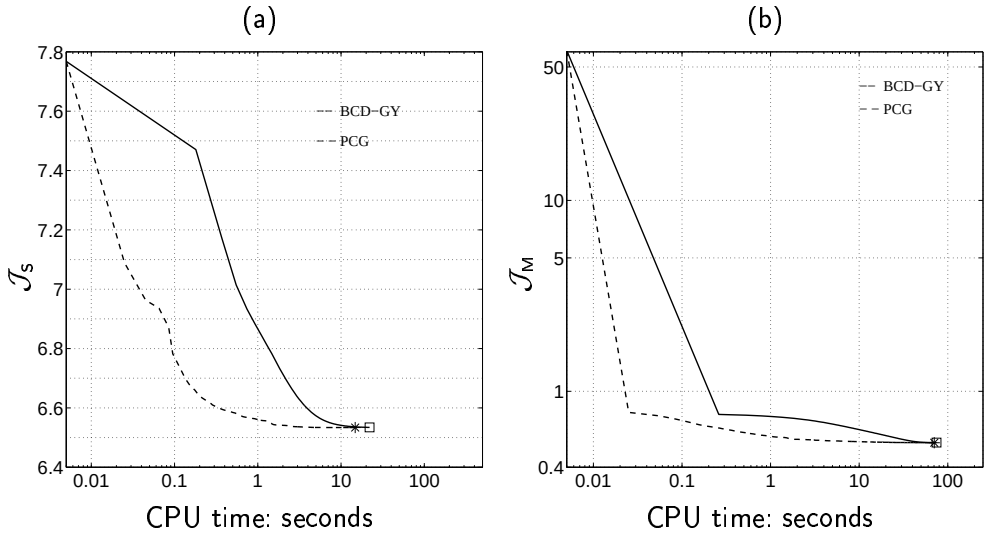

Fig. C.3. Performance of RSD and PCG algorithms for computing SS estimate on (a) and MS estimate on (b). The vertical axis represents the criterion values. Solid and dash-dotted lines are for minimization with RSD and PCG, respectively. The stopping points are depicted by a square ($\Box$) for RSD and by a star ($*$) for PCG.

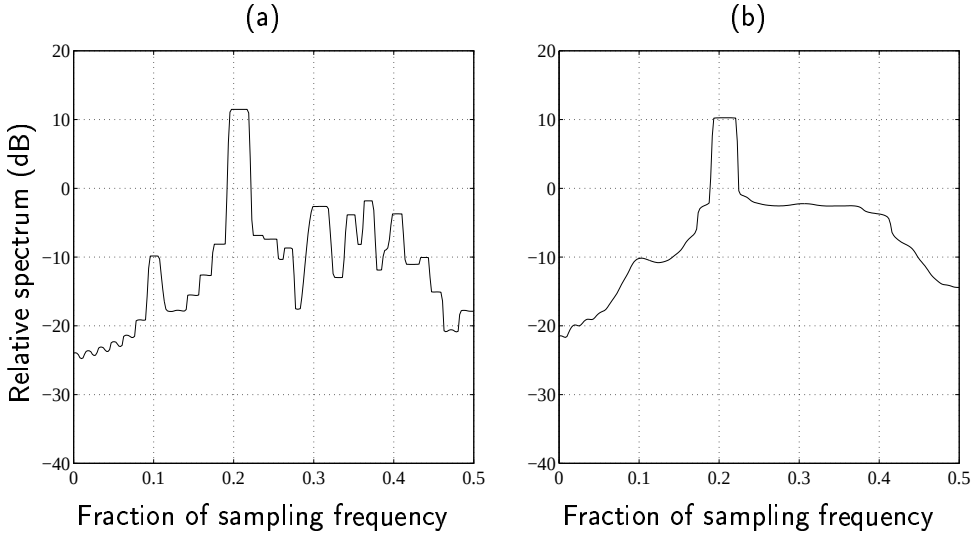

Fig. C.4. Smooth spectra reconstructed with a circular Gibbs-Markov penalty function, $(\lambda, \tau_1) = (0.05, 0.001)$; (a) convex case where $\mu = \mu_{\text{sup}} = 0.5,\ \varepsilon_K = 0.9,$ (b) nonconvex case where $\mu = 5,\ \varepsilon_K = 0.9.$

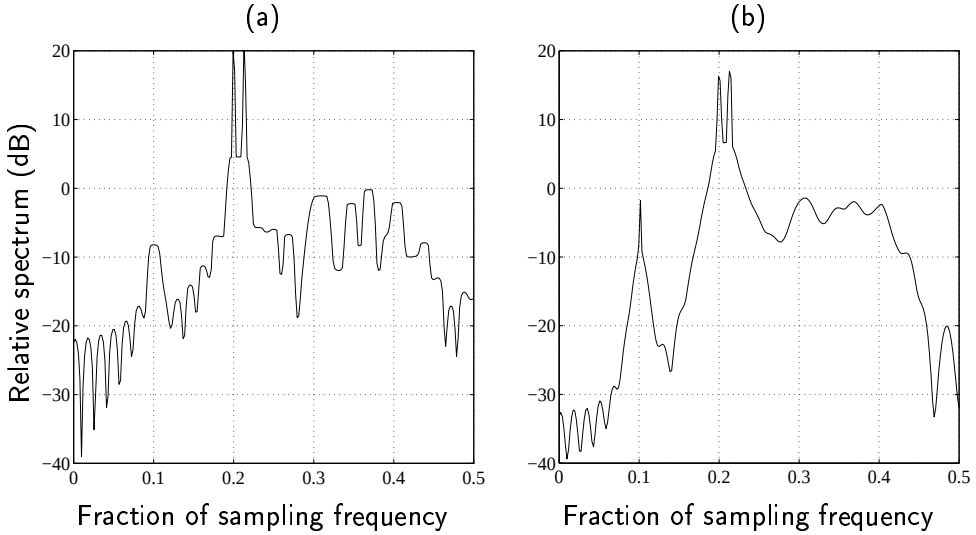

Fig. C.5. Mixed spectra. (a): convex case $\mu = 0.5$; (b) nonconvex extension $\mu = 5$; (c) and (d) correspond respectively to the line ($|\widehat{\boldsymbol{x}}_{\mathrm{L}}|^2$) and smooth ($|\widehat{\boldsymbol{x}}_{\mathrm{S}}|^2$) parts of $|\widehat{\boldsymbol{x}}|^2$ depicted in (b).