# Convex Half-Quadratic Criteria and Interacting Auxiliary Variables for Image Restoration

Jérôme Idier



***Abstract*—This paper deals with convex half-quadratic criteria and associated minimization algorithms for the purpose of image restoration. It brings a number of original elements within a unified mathematical presentation based on convex duality. Firstly, Geman and Yang's [1] and Geman and Reynolds's [2] constructions are revisited, with a view to establish convexity properties of the resulting half-quadratic augmented criteria, when the original nonquadratic criterion is already convex. Secondly, a family of convex Gibbsian energies that incorporate interacting auxiliary variables is revealed as a potentially fruitful extension of Geman and Reynolds's construction.**

***Index Terms*—Convex duality, coordinate descent algorithms, edge-preserving restoration, Gibbs–Markov models, line processes.**

## I. INTRODUCTION

HANDLING *half-quadratic* (HQ) criteria has recently stood out as a powerful numerical device in the field of edge-preserving image restoration, either formulated in the stochastic framework of Gibbs–Markov random field estimation [1]–[5], or in the deterministic, energetic counterpart [6]–[8]. In this paper, the terminology rather refers to the former formulation.

A function $K$ is said to be HQ if it depends on two sets of variables, say, $\boldsymbol{x}$ and $\boldsymbol{b}$, so that $K$ is a quadratic function of $\boldsymbol{x}$. We shall assume that $\boldsymbol{b}$ and $\boldsymbol{x}$ are column vectors, of respective size $M$ and $N$. For instance, consider a probability density function $f_{\boldsymbol{X},\boldsymbol{B}}$ whose conditional density $f_{\boldsymbol{X}|\boldsymbol{B}}$ is Gaussian. Then, the negative logdensity $-\log f_{\boldsymbol{X},\boldsymbol{B}}(\boldsymbol{x},\boldsymbol{b}) = -\log f_{\boldsymbol{X}|\boldsymbol{B}}(\boldsymbol{x}|\boldsymbol{b}) - \log f_{\boldsymbol{B}}(\boldsymbol{b})$ is clearly HQ.

HQ criteria can be traced back to piecewise Gaussian Gibbs–Markov models that incorporate binary line processes, either interacting [9]–[11] or decoupled [12]. In the latter case, Blake & Zisserman showed that a HQ criterion $K$, expressing the idea of a weak continuity constraint, could be considered as an *augmented* version of another criterion $J$ (involving the truncated quadratic), in the sense that

$$\inf_{\boldsymbol{b}\in\{0,1\}^M} K(\cdot,\boldsymbol{b}) = J.$$

As a consequence, $J$ and $K$ share the same minima, that can be sought using any suitable numerical device working on either $J$ or $K$. In practice, Blake and Zisserman defined the HQ function $K$ as the objective function, deduced $J$ from $K$, and proposed a *graduated nonconvexity* approach to minimize the resulting nonquadratic function $J$.

Later, Geman *et al.*'s contributions [1], [2] generalized Blake and Zisserman's construction to a larger class of decoupled auxiliary processes. In fact, they also reversed the construction process: they showed that there exist augmented HQ counterparts $K$ for a wide range of nonquadratic edge-preserving Gibbsian criteria $J$. Each of the two references proposes its own way to construct an augmented HQ function $K$, say $K_{\mathrm{GR}}$ and $K_{\mathrm{GY}}$, from some nonquadratic criteria $J$, so that

$$\min_{\boldsymbol{b}\in B} K(\cdot,\boldsymbol{b}) = J \tag{1}$$

for an appropriate set $B$, say $B_{\mathrm{GR}}$ and $B_{\mathrm{GY}}$, respectively (the resulting auxiliary process $\boldsymbol{b}$ may not be binary).

Moreover, [1] and [2] supported the idea that minimizing $K$ rather than $J$ has some structural advantages. More precisely, one can benefit from half-quadraticity by alternating updates of $\boldsymbol{x}$ given $\boldsymbol{b}$, and of $\boldsymbol{b}$ given $\boldsymbol{x}$, provided that the latter be a simple enough operation. In [1] and [2], simulated annealing based on alternate Gibbs sampling was explored. Other contributors rather developed deterministic counterparts, whether on $K_{\mathrm{GR}}$ [3]–[5], [8] or on $K_{\mathrm{GY}}$ [3], [6], [7]. The resulting deterministic algorithms fall into the well-known class of coordinate descent algorithms (i.e., *relaxation* algorithms) [13]–[15]. For instance, Charbonnier *et al.* [3] introduce block coordinate descent algorithms called ARTUR and LEGEND to minimize $K_{\mathrm{GR}}$ and $K_{\mathrm{GY}}$, respectively. If the original criterion $J$ presents local minima, it is not difficult to check that deterministic HQ algorithms can be stuck on corresponding local minimizers (a thorough convergence analysis is conducted in [16]).

This paper rather focuses on the case where the energy function $J$ is convex, and studies whether the convexity of $J$ is structurally transferred to $K_{\mathrm{GR}}$ and to $K_{\mathrm{GY}}$. Be the answer positive, then the convergence analysis of deterministic HQ algorithms such as ARTUR and LEGEND becomes a straightforward application of existing and well-documented results about the relaxation of convex criteria (for instance, see [13] and [14] for pioneering contributions, and [15] for a recent overview).

In [5], it is already shown that ARTUR provides a sequence of images converging toward $\widehat{\boldsymbol{x}}$ under some technical conditions, including the strict convexity of $J$. Yet, it can easily be checked that $K_{\mathrm{GR}}$ is not necessarily a convex function, even when $J$ is convex (see Remark 1). One of the main contributions of this paper is actually to provide a change of auxiliary variables that

Manuscript received March 31, 1999; revised March 28, 2001. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Timothy J. Schulz.

J. Idier is with the Laboratoire des Signaux et Systèmes (CNRS-SUPÉLEC-UPS), 91192 Gif-sur-Yvette Cedex, France (e-mail: idier@lss.supelec.fr).

Publisher Item Identifier S 1057-7149(01)05437-9.

makes $K_{\mathrm{GR}}$ convex (see Section IV). The case of $K_{\mathrm{GY}}$ is simpler, since convexity holds under appropriate technical conditions (see Section III). As a consequence, guarantees of global convergence (toward $\hat{\boldsymbol{x}}$) are provided for many versions of deterministic algorithms that operate on either of the criteria $K_{\mathrm{GR}}$ or $K_{\mathrm{GY}}$.

This paper is organized as follows. In Section II, the mathematical structure of criteria $J$ falling into the scope of our study is introduced. The corresponding form of HQ augmented criteria $K_{\mathrm{GR}}$ and $K_{\mathrm{GY}}$ is given, and a connection is made between HQ minimization and *reweighted least squares*. In Sections III and IV, Geman and Yang's and Geman and Reynolds's constructions are, respectively, revisited. Although more recent, Geman and Yang's form of augmented criterion is studied first, because its convexity properties turn out to be simpler to analyze. Most of the new results brought by Sections III and IV pertain to numerical analysis. On the contrary, Section V focuses on the design of a new class of image models: on the ground of Section IV, a family of convex Gibbsian energies that incorporate interacting auxiliary variables is introduced.

## II. Framework

### A. General Setting

Let us consider the case of energy functions that read

$$J(\boldsymbol{x}) = J_0(\boldsymbol{x}) + \sum_{c\in\mathcal{C}} \phi\left(\boldsymbol{v}_c^{\mathrm{t}}\boldsymbol{x} - w_c\right), \quad \boldsymbol{x} \in \mathbb{R}^N \tag{2}$$

where $\mathcal{C}$ is a set of cliques on a finite grid $\mathcal{S}$ ($|\mathcal{S}| = N$ and $|\mathcal{C}| = M$), the quantities $w_c$ are real-valued, and each of the vectors $\boldsymbol{v}_c \in \mathbb{R}^N$ has a support restricted to clique $c$. We also assume that $J_0$ is a strictly convex quadratic function, which reads

$$J_0(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{t}}\boldsymbol{M}_0\boldsymbol{x} - 2\boldsymbol{m}_0^{\mathrm{t}}\boldsymbol{x} + \mu_0 \tag{3}$$

with $\boldsymbol{M}_0 \geq 0$, without loss of generality, and that the function $\phi: \mathbb{R} \to \mathbb{R}$ is not quadratic. For the sake of notational simplicity, the same function $\phi$ is assigned to every clique, but the whole derivation admits an immediate inhomogeneous extension. We assume that the problem to solve is

$$\text{minimize } J(\boldsymbol{x}) \quad \text{subject to } \boldsymbol{x} \in X \tag{4}$$

where $X = X_1 \times X_2 \times \cdots \times X_N$ is a Cartesian product of closed convex sets of $\mathbb{R}$. Typical cases are $X = [x_{\min}, x_{\max}]^N$ (pixels have a known finite range), $X = \mathbb{R}_+^N$ (pixels have positive values) and $X = \mathbb{R}^N$ (unconstrained case).

Under slightly different technical conditions, [1] and [2] respectively propose the following HQ functions that satisfy (1)

$$K_{\mathrm{GR}}(\boldsymbol{x}, \boldsymbol{b}) = J_0(\boldsymbol{x}) + \sum_{c\in\mathcal{C}} \left(b_c\left(\boldsymbol{v}_c^{\mathrm{t}}\boldsymbol{x} - w_c\right)^2 + \psi(b_c)\right) \tag{5}$$

$$K_{\mathrm{GY}}(\boldsymbol{x}, \boldsymbol{b}) = J_0(\boldsymbol{x}) + \sum_{c\in\mathcal{C}} \left(\frac{1}{2}\left(\boldsymbol{v}_c^{\mathrm{t}}\boldsymbol{x} - w_c - b_c\right)^2 + \zeta(b_c)\right) \tag{6}$$

where $\boldsymbol{b} = (b_c)_{c\in\mathcal{C}}$, $B_{\mathrm{GR}} = \mathbb{R}_+^M$, $B_{\mathrm{GY}} = \mathbb{R}^M$, and both functions $\psi$ and $\zeta$ can be defined from $\phi$ through convex duality relations, according to Sections IV and III, respectively.

### B. Image Restoration

The energy function (2) encompasses the posterior negative loglikelihood of many existing Gibbs–Markov models in image restoration and computed imaging. More specifically, a frequent case is obtained when $\boldsymbol{x}$ is an unknown two-dimensional (2-D) or three-dimensional (3-D) image, $J_0(\boldsymbol{x})$ is a quadratic fidelity term with respect to (w.r.t.) some data set $\boldsymbol{z}$, $w_c = 0$, and the vectors $\boldsymbol{v}_c$ define finite difference operators. This is precisely the starting point of existing works devoted to HQ regularization for image restoration.

One basic, widespread model involves pairwise interactions and has the following posterior energy of a 2-D image of size $I \times J$ (up to an additional constant):

$$\begin{aligned} J(\boldsymbol{x}) = &\|\boldsymbol{z} - \boldsymbol{H}\boldsymbol{x}\|^2 + \sum_{i=1}^{I}\sum_{j=2}^{J} \phi(x_{i,j} - x_{i,j-1}) \\ &+ \sum_{i=2}^{I}\sum_{j=1}^{J} \phi(x_{i,j} - x_{i-1,j}) \end{aligned} \tag{7}$$

where matrix $\boldsymbol{H}$ stands for a linear observation operator. For instance, the weak membrane corresponds to (7) when $\phi$ is the truncated quadratic function. In the probabilistic interpretation of (7), the first term corresponds to the neg-loglikelihood of the data $\boldsymbol{z}$ when it is assumed that the descrepancy between $\boldsymbol{z}$ and $\boldsymbol{H}\boldsymbol{x}$ is due to additive zero-mean white Gaussian noise.

### C. Signal Processing

The *reweighted least squares* method for one-dimensional (1-D) linear prediction [17] is a pioneering example of a HQ approach, developped in the context of geophysical $l_p$ deconvolution. Basically, the problem is to find the optimal predictor $\boldsymbol{a} = [a_1, \ldots, a_N]^{\mathrm{t}}$ in the $l_p$ norm sense, i.e., to minimize

$$J(\boldsymbol{a}) = \sum_{m=1}^{M}\left|z_m - \sum_{n=1}^{N} a_n z_{m-n}\right|^p \tag{8}$$

where $[z_1, \ldots, z_M]$ is a data vector, possibly windowed. Clearly, (8) identifies with (2) for

$$\begin{aligned} \boldsymbol{x} = &\boldsymbol{a}, \quad \phi = |.|^p, \quad \mathcal{C} = \{1, \ldots, M\}, \quad w_m = z_m, \\ \boldsymbol{v}_m = &[z_{m-1}, \ldots, z_{m-N}]^{\mathrm{t}}. \end{aligned}$$

Actually, the *iteratively reweighted least squares* (IRLS) algorithm proposed in [17] is formally equivalent to a numerical scheme that performs block coordinate descent on the HQ criterion $K_{\mathrm{GR}}$ built from (8). It is even more surprising to find that the *residual steepest descent* (RSD) method proposed after IRLS in the same paper [17] implicitly performs block coordinate descent on $K_{\mathrm{GY}}$. However, no reference is made to the concept of augmented criterion in [17].

A more recent application of IRLS can be found in [18]. Apparently, the recent contributions to HQ image regularization were made independently of the reweighted least squares approach developed earlier for signal estimation. Here, our primary concern stays in the field of 2-D and 3-D image restoration, but the connection between HQ and reweighted least squares approaches makes clear that the unified methodology applies to a very wide range of signal and image processing issues.

## III. Geman and Yang's Form of Augmented Criterion [2]

### A. Introduction and General Properties

Following Geman and Yang [2], consider functions $\phi$ such that $g(x) = x^2/2 - \phi(x)$ defines a closed proper convex function. Let $g^*$ be the conjugate of $g$ [19, Sect. 12], i.e.,

$$g^*(b) = \sup_{x \in \mathbb{R}} (bx - g(x))$$

and $\zeta(b) = g^*(b) - b^2/2$. Then, we have

$$\zeta(b) = \sup_{x \in \mathbb{R}} \left( \phi(x) - \frac{1}{2}(x-b)^2 \right). \tag{9}$$

Reciprocally, we have also (see [2])

$$g(x) = \sup_{b \in \mathbb{R}} (bx - g^*(b))$$

so that

$$\phi(x) = \inf_{b \in \mathbb{R}} \left( \frac{1}{2}(x-b)^2 + \zeta(b) \right) \tag{10}$$

and (1) holds for $K = K_{\text{GY}}$. According to (10), $\phi$ is the *infimal convolution* of $\zeta$ with the quadratic function $(\cdot)^2/2$, in the terminology of convex analysis [19].

In [3], an algorithm called LEGEND is proposed for computed imaging. It performs block coordinate descent on $K_{\text{GY}}$ in the unconstrained case ($X = \mathbb{R}^N$), based on the convexity of $K_{\text{GY}}(\boldsymbol{x}, \boldsymbol{b})$ in $\boldsymbol{x}$ when $\boldsymbol{b}$ is fixed and in $\boldsymbol{b}$ when $\boldsymbol{x}$ is fixed. An essentially similar scheme is proposed in [6], in the field of computer vision. Such algorithms proceed in two steps.

- As a function of $\boldsymbol{x}$, $K_{\text{GY}}(\boldsymbol{x}, \boldsymbol{b})$ is quadratic, and the associated normal matrix does not depend on $\boldsymbol{b}$: given (3) and (6), the gradient of $K_{\text{GY}}$ w.r.t. $\boldsymbol{x}$ vanishes for

$$\left(2\boldsymbol{M}_0 + \boldsymbol{V}\boldsymbol{V}^{\text{t}}\right)\boldsymbol{x} = 2\boldsymbol{m}_0 + \boldsymbol{V}(\boldsymbol{w} + \boldsymbol{b}) \tag{11}$$

  where $\boldsymbol{V} \triangleq [\boldsymbol{v}_1 | \cdots | \boldsymbol{v}_M]$ and $\boldsymbol{w} \triangleq [w_1, \ldots, w_M]^{\text{t}}$. This property is interesting from the numerical viewpoint since, for instance, the normal matrix can be inverted or factored once for all.
- Since the variables $b_c$ do not interact within $K_{\text{GY}}$, the second step can be solved in a parallel form. From basic results of duality theory [19], the updating equation for each $b_c$ is explicit if $\phi$ is differentiable, and the infimum of (10) is uniquely reached at

$$\widetilde{b}_x = g'(x) = x - \phi'(x). \tag{12}$$

  Hence, no explicit expression of $\zeta$ is needed to compute $\widetilde{b}_x$. This is fortunate since closed-form expressions of $\zeta$ are difficult or even impossible to establish for most well-known edge-preserving functions $\phi$, as depicted in Table I.

### B. Effect of a Change of Scale

Let us study the effect of altering the scale of $\phi$ in the definition of $\zeta$, by substituting $\phi_\alpha \triangleq \alpha\phi$ for $\phi$ in the previous subsection. Since expression (2) of criterion $J$ can be rewritten

$$J(\boldsymbol{x}) = J_0(\boldsymbol{x}) + \frac{1}{\alpha} \sum_{c \in \mathcal{C}} \phi_\alpha \left( \boldsymbol{v}_c^{\text{t}} \boldsymbol{x} - w_c \right)$$

for any $\alpha \neq 0$, the introduction of $\alpha$ corresponds to a true degree of freedom in the definition of the augmented criterion. Indeed, it is shown below that the resulting effect is not trivial and that accounting for it is not a secondary matter. Provided that $g_\alpha(x) = x^2/2 - \phi_\alpha(x)$ still define a convex function, the change of scale gives rise to a family of augmented criteria

$$K_{\text{GY}}(\boldsymbol{x}, \boldsymbol{b}; \alpha) = J_0(\boldsymbol{x}) + \frac{1}{\alpha} \sum_{c \in \mathcal{C}} \left( \frac{1}{2} \left( \boldsymbol{v}_c^{\text{t}} \boldsymbol{x} - w_c - b_c \right)^2 + \zeta_\alpha(b_c) \right)$$

which satisfy the original requirement $\inf_{\boldsymbol{b} \in \mathbb{R}^M} K_{\text{GY}}(\boldsymbol{x}, \boldsymbol{b}; \alpha) = J(\boldsymbol{x})$ for every $\alpha > 0$, if $\zeta_\alpha$ is defined by $\zeta_\alpha(b) = \sup_{x \in \mathbb{R}} (\alpha\phi(x) - (x-b)^2/2)$.

The fulfillment of the convexity condition on $g_\alpha$ clearly depends on $\alpha$. It is not difficult to prove the following lemma.

*Lemma 1:* For a given function $\phi$ that takes finite values on $\mathbb{R}$, the set of nonnegative values of $\alpha$ that render $g_\alpha$ convex is a non empty interval of the form $[0, \alpha_{\max}]$ ($\alpha_{\max} = \infty$ if $\phi$ is concave). For all $\alpha < \alpha_{\max}$, $g_\alpha$ is even strictly convex.

Such a result meets a qualitative comment of Cohen [6, Remark 6]. It ensures that Geman and Yang's construction is available for a wide range of functions $\phi$, provided that the scale factor $\alpha$ be suitably tuned.

Although no simple operation allows one to deduce $\zeta_\alpha$ from $\zeta$, the simplicity of the updating (11) and (12) is maintained, according to

$$\left(2\boldsymbol{M}_0 + \frac{1}{\alpha}\boldsymbol{V}\boldsymbol{V}^{\text{t}}\right)\boldsymbol{x} = 2\boldsymbol{m}_0 + \frac{1}{\alpha}\boldsymbol{V}(\boldsymbol{w} + \boldsymbol{b})$$

and

$$\widetilde{b} = x - \alpha\phi'(x)$$

respectively. It would be interesting to study the speed of convergence of a given descent algorithm on $K_{\text{GY}}(\boldsymbol{x}, \boldsymbol{b}; \alpha)$ as a function of $\alpha$. Intuitively, it seems preferable to choose $\alpha$ close to $\alpha_{\max}$. However, if $g_{\alpha_{\max}}$ is not strictly convex, then $\zeta_{\alpha_{\max}}$ is not a $C^1$ function.

Finally, let us remark that a similar change of scale on Geman and Reynolds's alternative would only yield the trivial modification $\psi_\alpha = \alpha\psi(\cdot/\alpha)$.

### C. Additional Properties When $\phi$ is Convex

If $\phi$ is convex, the following theorem provides the appropriate technical basis to show the convergence of coordinate descent algorithms toward $\widehat{\boldsymbol{x}}$.

TABLE I
FOUR EXAMPLES OF CONVEX EDGE-PRESERVING FUNCTIONS $\phi$, AND CORRESPONDING EXPRESSIONS OF $\zeta_\alpha(b) = \sup_{x\in\mathbb{R}} (\alpha\phi(x) - (x-b)^2/2)$

| | | Geman & Yang's construction | |
|---|---|---|---|
| | $\phi(x)$ | $\zeta_\alpha(b)$, $\alpha \in [0, \alpha_{\max}]$ | $\alpha_{\max}$ |
| (a) | $\sqrt{s^2+x^2}$ | $\alpha\phi(\tilde{x}_b) - \frac{1}{2}(\tilde{x}_b - b)^2$ ; $\tilde{x}_b$ : zero of $(s^2+x^2)(x-b)^2 - \alpha^2x^2$ | $s$ |
| (b) | $1 + \frac{\|x\|}{s} - \log\left(1+\frac{\|x\|}{s}\right)$ | $\alpha\phi(\tilde{x}_b) - \frac{1}{2}(\tilde{x}_b - \|b\|)^2$ ; $\tilde{x}_b \geq 0$ : zero of $x^2 - (\|b\| + \alpha/s - s)x - s\,\|b\|$ | 1 |
| (c) | $\begin{cases} x^2 & \text{if } \|x\| < s \\ 2s\,\|x\| - s^2 & \text{otherwise} \end{cases}$ | $\begin{cases} \alpha b^2/(1-2\alpha) & \text{if } \|b\| < (1-2\alpha)s \\ 2\alpha s\,\|b\| - (1-2\alpha)\alpha s^2 & \text{otherwise} \end{cases}$ | $\frac{1}{2}$ |
| (d) | $\|x\|^p$ , $1 \leq p < 2$ | not applicable | 0 |

*Theorem 1:* Let the function $\phi$ be continuous, convex (resp. strictly convex) and satisfy

$$\lim_{|x|\to\infty} \frac{\phi(x)}{x^2} < \frac{1}{2}$$

and suppose that $g$ is convex. Then the function $\zeta$ defined by (9) is convex (resp. strictly convex).

*Proof:* See Appendix A. □

*Corollary 1:* Let us assume $\phi$ meets the conditions of Theorem 1. Then the criterion $K_{\text{GY}}$ defined by (6) is convex in $(\boldsymbol{x}, \boldsymbol{b})$. In addition, if $J_0$ is strictly convex, or if $J$ and $\phi$ are strictly convex, then $K_{\text{GY}}$ is strictly convex. On the other hand, if $g$ is strictly convex, then $K_{\text{GY}}$ is $C^1$.

*Proof:* See Appendix B. □

Corollary 1 provides a sufficient condition to ensure that $K_{\text{GY}}$ is a strictly convex $C^1$ criterion. As a consequence, methods of coordinate descent on $K_{\text{GY}}$ are globally convergent toward the unique global minimizer in $X \times B_{\text{GY}}$, according to classical studies (for instance, [15, Prop. 2.7.1]).

Among the examples of Table I, the three first scaled functions $\phi_\alpha = \alpha\phi$ satisfy the conditions of Theorem 1 for $\alpha \leq \alpha_{\max}$. Moreover, save the limit case $\alpha = \alpha_{\max} = 1/2$ for Huber's function (c) (in which case $\zeta_{1/2} = s|\cdot|$), $g_\alpha = (\cdot)^2/2 - \phi_\alpha$ is strictly convex, so the resulting HQ criterion $K_{\text{GY}}$ is $C^1$.

## IV. GEMAN AND REYNOLDS'S FORM OF AUGMENTED CRITERION [1]

### A. Introduction and General Properties

Here, we introduce the HQ criterion $K_{\text{GR}}$ initially proposed by Geman and Reynolds [1], and we study some of its properties. In order to stress the common features in the construction of $K_{\text{GR}}$ and $K_{\text{GY}}$, our construction of $K_{\text{GR}}$ is based on convex duality. Besides, this allows us to benefit directly from the known properties of convex conjugate functions.

Let us consider the functions $\phi$ that satisfy the following hypotheses:

$$\phi \text{ is even} \tag{13}$$

$$\phi(\sqrt{\cdot}) \text{ is concave on } \mathbb{R}_+ \tag{14}$$

$$\phi \text{ is continuous near zero and } \mathcal{C}^1 \text{ on } \mathbb{R}^* = \mathbb{R}\backslash\{0\}. \tag{15}$$

Most regularizing functions known in the litterature as *edge-preserving* satisfy (13)–(15), including $l_p$ functions $(0 < p < 2)$ and Huber's function.

The *extended-valued* function $f$ defined by

$$\begin{cases} \forall x < 0, & f(x) = +\infty \\ \forall x \geq 0, & f(x) = -\phi(\sqrt{x}) \end{cases} \tag{16}$$

is convex on $\mathbb{R}$ and $C^1$ on $\mathbb{R}_+^*$. Let $f^*$ denote its conjugate. Because $f$ is a closed proper convex function, it is also the conjugate of $f^*$ [19, Cor. 12.2.1]. As a consequence, we have

$$\phi(x) = \inf_{b\in\mathbb{R}} \left(bx^2 + \psi(b)\right) \tag{17}$$

where

$$\psi(b) = f^*(-b) = \sup_{x\in\mathbb{R}} \left(\phi(x) - bx^2\right). \tag{18}$$

From (17), it is easy to deduce (1) for $K = K_{\text{GR}}$ as defined by (5).

The auxiliary process $\boldsymbol{b}$ within criterion $K_{\text{GR}}$ is clearly reminiscent of a line process. Basically, each component $\psi(b_c)$ introduces a price to pay for cancelling the quadratic potentiel attached to clique $c$.

Lemma 2 sums up the main properties of the function $\psi$ defined by (18), when it is additionally assumed that $\phi$ grows slower than $x^2$ at infinity.

*Lemma 2:* Let $\phi$ fulfill (13)–(15) and

$$\lim_{x\to\infty} \frac{\phi(x)}{x^2} = 0 \tag{19}$$

and let $b_\infty = -f'_+(0) = \lim_{x\to 0} \phi'(x)/2x$ ($b_\infty$ may be infinite). Then

1)
$$\begin{cases} \forall b \leq 0, \psi(b) = +\infty \\ \psi \text{ is strictly decreasing and strictly convex on } (0, b_\infty) \\ \psi \text{ is constant on } [b_\infty, +\infty). \end{cases}$$

2) $\psi$ is convex. Furthermore it is $C^1$ on $\mathbb{R}_+^*$ if (14) is replaced by the *strict* counterpart

$$\phi(\sqrt{\cdot}) \text{ is strictly concave on } \mathbb{R}_+. \tag{20}$$

3) Equation (17) can be replaced by

$$\phi(x) = \inf_{b\in(0,b_\infty]} \left(bx^2 + \psi(b)\right). \tag{21}$$

*Proof:* The proof is straightforward and most of its elements can be found elsewhere. Besides, several properties stated by Lemma 2 are counterparts of [1, Th. 1] or of [5, Th. 1], while some are obtained for weakened conditions. In particular, (15) is slightly weaker than continuous differentiability (assumed in [5] and also in [16]). Note also that $\lim_{x\to\infty} \phi(x)/x^2 = 0$ is equivalent to $\lim_{x\to\infty} \phi'(x)/x = 0$, i.e., [5, Theorem 1, Cond. 12(f)]. Finally, let us stress that point 2) is a direct consequence of properties attached to convex duality: $f^*$ is convex as the conjugate of $f$, and it is $C^1$ if $f$ is strictly convex [19, Th. 26.3]. □

The augmented criterion $K_{\text{GR}}$ is a quadratic function of $\boldsymbol{x}$ when $\boldsymbol{b}$ is fixed, and it is a convex function of $\boldsymbol{b}$ when $\boldsymbol{x}$ is fixed, as an immediate consequence of Lemma 2. In [3] and [5], an algorithm called ARTUR is proposed for computed imaging, which performs block coordinate descent on $K_{\mathrm{GR}}$ in the unconstrained case ($X = \mathbb{R}^N$).

- In a first step, $K_{\mathrm{GR}}$ is minimized as a quadratic function of $\boldsymbol{x}$, while $\boldsymbol{b}$ is held constant. The resulting normal equation reads

$$\left(\boldsymbol{M}_0 + \boldsymbol{V}\boldsymbol{B}\boldsymbol{V}^{\mathrm{t}}\right)\boldsymbol{x} = \boldsymbol{m}_0 + \boldsymbol{V}\boldsymbol{B}\boldsymbol{w} \tag{22}$$

where $\boldsymbol{b} \triangleq \operatorname{diag}\{b_1, \ldots, b_M\}$, and $\boldsymbol{V}$ and $\boldsymbol{w}$ have been defined in Section III.

- In a second step, $K_{\mathrm{GR}}$ is minimized as a function of $\boldsymbol{b}$. Since the variables $b_c$ do not interact within $K_{\mathrm{GR}}$, this step can be achieved in a parallel form. Moreover, the updating equation for each $b_c$ is explicit: the infimum of (21) is uniquely reached at

$$\hat{b}_x = \begin{cases} b_\infty, & \text{if } x = 0 \\ \frac{\phi'(x)}{2x}, & \text{otherwise} \end{cases} \tag{23}$$

and the expression of $\psi$ is not required to compute $\widehat{b}_x$.

It is not obvious to adapt such a block version to constrained cases ($X \subsetneq \mathbb{R}^N$), whereas the *single-site update* version proposed in [4] naturally accounts for each constraint $x_n \in X_n$ as the corresponding variable $x_n$ is updated.

Although $K_{\mathrm{GR}}(\boldsymbol{x}, \boldsymbol{b})$ is a convex function of $\boldsymbol{x}$ when $\boldsymbol{b}$ is fixed, and also a convex function of $\boldsymbol{b}$ when $\boldsymbol{x}$ is fixed, it is not necessarily a convex function of $(\boldsymbol{x}, \boldsymbol{b})$, so that global convergence toward the global minimizer is not guaranteed for such deterministic descent algorithms. In [5], a convergence study is conducted when $\phi$ is convex (additional technical conditions are required). Without the convexity of $\phi$, weaker results can still be obtained according to [16]. In fact, neither of the two studies is actually based on acknowledged properties of coordinate descent. In contrast, the next subsection establishes properties of the augmented criterion $K_{\mathrm{GR}}(\boldsymbol{x}, \boldsymbol{b})$ when $\phi$ is convex, so that global convergence properties of algorithms such as ARTUR can be obtained in the usual framework of coordinate descent.

### B. Additional Properties When $\phi$ is Convex

This subsection contains the main result of Section IV, that is Theorem 2. More specifically, the latter provides sufficient conditions to establish that $K_{\mathrm{GR}}$ is convex (see Corollary 2), up to the change of variable

$$(0, b_\infty] \longrightarrow \boldsymbol{I}_\phi = [\phi(0), +\infty), \quad b \longmapsto d = \psi(b).$$

Since the latter is one-to-one according to Lemma 2(1), (21) also reads

$$\phi(x) = \inf_{d \in I_\phi} \left(\psi^{-1}(d)x^2 + d\right).$$

Table II contains several examples of convex edge-preserving functions $\phi$ that yield closed-form expressions for $\psi$ and $\psi^{-1}$.

*Theorem 2:* If $\phi$ is a convex function on $\mathbb{R}$ that fulfills (13) and (14), then the function $\psi^{-1}(d)x^2$ is convex in $(x, d)$ on $\mathbb{R} \times I_\phi$.

*Proof:* See Appendix C. □

TABLE II
SAME FUNCTIONS $\phi$ AS IN TABLE I, AND CORRESPONDING EXPRESSIONS OF $\psi(b) = \sup_{x \in \mathbb{R}} (\phi(x) - bx^2)$, $0 < b \leq b_\infty$ AND OF $\psi^{-1}(d)$, $\phi(0) \leq d < \infty$

| | | Geman & Reynolds's construction | | |
|---|---|---|---|---|
| | $\phi(x)$ | $\psi(b)$, $b \in (0, b_\infty]$ | $b_\infty$ | $\psi^{-1}(d)$, $d \in [\phi(0), \infty)$ |
| (a) | $\sqrt{s^2 + x^2}$ | $s^2 b + \frac{1}{4b}$ | $\frac{1}{2s}$ | $\frac{d - \sqrt{d^2 - s^2}}{2s^2}$ |
| (b) | $1 + \frac{\lvert x\rvert}{s} - \log\left(1 + \frac{\lvert x\rvert}{s}\right)$ | $s + \frac{2-s}{4sb} - s^2 b + \log 2sb$ | $\frac{1}{2}$ | no closed-form |
| (c) | $\begin{cases} x^2 & \text{if } \lvert x\rvert \leq s \\ 2s\lvert x\rvert - s^2 & \text{otherwise} \end{cases}$ | $\frac{s^2}{b} - s^2$ | 1 | $\frac{s^2}{d + s^2}$ |
| (d) | $\lvert x\rvert^p, 1 \leq p < 2$ | $\left(1 - \frac{p}{2}\right)\left(\frac{2b}{p}\right)^{\frac{p}{p-2}}$ | $\infty$ | $\frac{p}{2}\left(\frac{2d}{2-p}\right)^{\frac{p-2}{p}}$ |

*Remark 1:* Whereas the function $\psi^{-1}(d)x^2 + d$ is convex in $(x, d)$, the original function $bx^2 + \psi(b)$ is not convex in $(x, b)$. For the sake of simplicity, let us assume that $\psi$ is $C^2$ (i.e., continuously twice differentiable). Then the Hessian of $bx^2 + \psi(b)$ reads $2b\psi''(b) - 4x^2$, which takes negative values when $|x|$ increases with $b$ fixed. Of course, this does not contradict the fact that $bx^2 + \psi(b)$ is both convex in $x$ when $b$ is fixed and in $b$ when $x$ is fixed.

*Remark 2:* Had we supposed *strict* convexity of $\phi$ in Theorem 2, neither $\psi^{-1}(d)x^2 + d$ nor $\psi^{-1}(d)x^2$ would have become strictly convex, since their restrictions to the line $x = 0$ are linear anyway. However, their restrictions to any other line are strictly convex.

*Remark 3:* It is possible to show that Theorem 2 admits the following converse property: let $\psi$ fulfill the statements 1) and 2) of Lemma 2, and define $\phi$ according to (21). If $\psi^{-1}(d)x^2$ defines a convex function in $(x, d)$ on $\mathbb{R} \times I_\phi$, then $\phi$ is also a convex function (on $\mathbb{R}$). Moreover, it can be checked that (13), (14), and (19) are true, and (20) is also true if $\psi$ is $C^1$ on $\mathbb{R}_+^*$.

*Corollary 2:* Let us assume that $\phi$ meets the conditions of Theorem 2. Then the criterion $\mathcal{K}(\boldsymbol{x}, \boldsymbol{d}) = K_{\mathrm{GR}}\left(\boldsymbol{x}, \boldsymbol{\Psi}^{-1}(\boldsymbol{d})\right)$ (where $\boldsymbol{d} \triangleq \boldsymbol{\Psi}(\boldsymbol{b}) \triangleq (\psi(b_c))_{c \in \mathcal{C}}$) is convex in $(\boldsymbol{x}, \boldsymbol{d}) \in \mathbb{R}^N \times I_\phi^M$.

*Proof:* We have

$$\mathcal{K}(\boldsymbol{x}, \boldsymbol{d}) = J_0(\boldsymbol{x}) + \sum_{c \in \mathcal{C}} \left(\psi^{-1}(d_c)\left(\boldsymbol{v}_c^{\mathrm{t}}\boldsymbol{x} - w_c\right)^2 + d_c\right) \tag{24}$$

which is convex, as a sum of convex functions. □

### C. Convergence of Coordinate Descent Algorithms on $K_{\mathrm{GR}}$

Here, we study the convergence properties of coordinate descent methods on $K_{\mathrm{GR}}$ (or actually on $\mathcal{K}$) when $\phi$ is convex, in the light of the technical results yielded by the previous subsection.

*Theorem 3: (Convergence of Coordinate Descent on $K_{\mathrm{GR}}$)* Suppose that (13), (19), (20) and $b_\infty < +\infty$ hold, as well as the following property:

$$\forall n \in \{1, \ldots, N\}, \quad \exists\, c \in \mathcal{C} \text{ such that } (\boldsymbol{v}_c)_n \neq 0. \tag{25}$$

Then, every limit point of a series of iterates obtained by coordinate descent on $K_{\mathrm{GR}}(\boldsymbol{x}, \boldsymbol{b})$ in $X \times (0, b_\infty]^M$ minimizes $K_{\mathrm{GR}}$ over $X \times (0, b_\infty]^M$, and it solves (4).

*Proof:* See Appendix D. □

Theorem 3 shows that a block coordinate descent method such as ARTUR converges to a global minimizer of $K_{\mathrm{GR}}$ when the prescribed conditions are fulfilled (the latter are slightly weaker than their counterpart in [5] since [5, Conditions 12(h–i)] are actually unnecessary).

On the other hand, a slight adaptation of Theorem 3 is required to prove that the algorithm proposed in [4] is also convergent, because the latter performs an over-relaxed form of coordinate descent. On the ground of Appendix D (and with the same notations), it can be shown that the conclusion of Theorem 3 still holds if each iteration (33) is relaxed according to

$$\begin{aligned}\boldsymbol{y}_l^{k+1} =& (1-\omega_l)\boldsymbol{y}_l^k \\ &+ \omega_l \arg\min_{\xi \in Y_l} F\left(\boldsymbol{y}_1^{k+1}, \ldots, \boldsymbol{y}_{l-1}^{k+1}, \xi, \boldsymbol{y}_{l+1}^k, \boldsymbol{y}_L^k\right)\end{aligned}$$

where $0 < \omega_l < 2$ if $l \leq N$ (i.e., the descent can be either under- or over-relaxed on pixels) and $0 < \omega_l \leq 1$ otherwise (i.e., the descent can be under-relaxed on auxiliary variables). In the latter case, let us stress that under-relaxation must be performed on the transformed variables $d_c$, not on $b_c$, because the transformation $\psi$ is not linear.

Among the examples of Table II, only the first two satisfy the conditions of Theorem 3. Huber function (c) violates the strict concavity condition (20) for $\phi(\sqrt{\cdot})$. However, some refined convergence results are still available for the relaxation of the resulting nondifferentiable criterion [14, p. 73]. On the other hand, $l_p$ norms (d) are excluded because $b_\infty < +\infty$. Then, since $b_\infty = \psi^{-1}(\phi(0))$, $\psi^{-1}(d)x^2$ is not continuous near $(x,d) = (0, \phi(0))$, and coordinate descent methods on $\mathcal{K}$ can be trapped by any couple $(\boldsymbol{x}, \boldsymbol{d})$ such that $d_c = 0$ and $\boldsymbol{v}_c^{\mathrm{t}}\boldsymbol{x} = w_c$ for at least one clique $c$. The situation $b_\infty = +\infty$ was similarly excluded from convergence results in preexisting contributions involving augmented criteria in image restoration [5], [16], and the same conclusion was earlier drawn in [17] concerning $l_p$ minimization using IRLS. A converging alternative algorithm can be found in [20].

## V. Interacting Auxiliary Processes Within Convex Energies

For the purpose of image reconstruction, an interesting consequence of Theorem 2 is the possibility to design convex Gibbsian criteria incorporating interacting, continuous-valued processes. As pointed out in [1], criterion (5) can be interpreted as a compound Gibbsian energy function

$$K_{\mathrm{GR}}(\boldsymbol{x}, \boldsymbol{b}) = J_0(\boldsymbol{x}) + \sum_{c\in\mathcal{C}} b_c \left(\boldsymbol{v}_c^{\mathrm{t}}\boldsymbol{x} - w_c\right)^2 + \sum_{c\in\mathcal{C}} \psi(b_c)$$

where the separable term $\sum_{c\in\mathcal{C}} \psi(b_c)$ corresponds to a continuous-valued, decoupled, auxiliary process $\boldsymbol{b}$. Now let us introduce a non separable convex function $\Upsilon : I_\phi^M \to \mathbb{R}$, and define the HQ function

$$K_\Upsilon(\boldsymbol{x}, \boldsymbol{b}) = J_0(\boldsymbol{x}) + \sum_{c\in\mathcal{C}} b_c \left(\boldsymbol{v}_c^{\mathrm{t}}\boldsymbol{x} - w_c\right)^2 + \Upsilon(\Psi(\boldsymbol{b})). \quad (26)$$

Then, it is a straightforward generalization of Corollary 2 that

$$K_\Upsilon\left(\boldsymbol{x}, \boldsymbol{\Psi}^{-1}(\boldsymbol{d})\right) = J_0(x) + \sum_{c\in\mathcal{C}} \psi^{-1}(d_c)\left(\boldsymbol{v}_c^{\mathrm{t}}\boldsymbol{x} - w_c\right)^2 + \Upsilon(\boldsymbol{d})$$

is a convex function of $(\boldsymbol{x}, \boldsymbol{d})$, while $\boldsymbol{d}$ is a vector of interacting auxiliary variables.

For instance, let us consider the HQ criterion that corresponds to (7)

$$\begin{aligned}K_{\mathrm{GR}}\left(\boldsymbol{x}, \boldsymbol{b}^{\mathrm{v}}, \boldsymbol{b}^{\mathrm{h}}\right) =& \|\boldsymbol{z} - \boldsymbol{H}x\|^2 \\ &+ \sum_{i=1}^{I}\sum_{j=2}^{J}\Big(b_{i,j}^{\mathrm{v}}(x_{i,j} - x_{i,j-1})^2 \\ &\qquad + \psi\left(b_{i,j}^{\mathrm{v}}\right)\Big) \\ &+ \sum_{i=2}^{I}\sum_{j=1}^{J}\Big(b_{i,j}^{\mathrm{h}}(x_{i,j} - x_{i-1,j})^2 \\ &\qquad + \psi\left(b_{i,j}^{\mathrm{h}}\right)\Big). \end{aligned} \quad (27)$$

The continuous-valued variables $\left(b_{i,j}^{\mathrm{v}}\right) \cup \left(b_{i,j}^{\mathrm{h}}\right)$ constitute a *soft* version of a line process without interactions [1]. Let us design an interactive extension for such a process. Existing interacting line processes are binary $\left(b_{i,j}^{\mathrm{v}}, b_{i,j}^{\mathrm{h}} \in \{0,1\}\right)$, and one of the simplest is Geiger and Girosi's extended weak membrane model [11]: compared to (27), additional cliques $\left(\left\{b_{i-1,j}^{\mathrm{v}}, b_{i,j}^{\mathrm{v}}\right\}\right) \cup \left(\left\{b_{i,j-1}^{\mathrm{h}}, b_{i,j}^{\mathrm{h}}\right\}\right)$ are introduced to prevent discontinuous boundaries. If $\{b_1, b_2\}$ stands for a generic clique, the corresponding potential reads $\bar{b}_1 + \bar{b}_2 - \varepsilon\bar{b}_1\bar{b}_2$, with $\bar{b} = 1 - b$ and $\varepsilon > 0$, so that two neighboring lines ($\bar{b}_1 = \bar{b}_2 = 1$) cost less than double the cost of one isolated line. For a continuous-valued line process, a similar effect can be sought. One simple way is to define $\Upsilon$ according to

$$\begin{aligned}\Upsilon(\boldsymbol{d}^{\mathrm{v}}, \boldsymbol{d}^{\mathrm{h}}) =& \sum_{i=2}^{I}\sum_{j=2}^{J} \max\left\{d_{i-1,j}^{\mathrm{v}}, d_{i,j}^{\mathrm{v}}\right\} \\ &+ \sum_{i=2}^{I}\sum_{j=2}^{J} \max\left\{d_{i,j-1}^{\mathrm{h}}, d_{i,j}^{\mathrm{h}}\right\}.\end{aligned}$$

Since $\max\{d_1, d_2\}$ is a convex function of $(d_1, d_2)$, the resulting criterion $K_\Upsilon$ is a convex function of $(\boldsymbol{x}, \boldsymbol{d}^{\mathrm{v}}, \boldsymbol{d}^{\mathrm{h}})$.

The performance of the resulting interacting extension has been compared to the original, noninteracting model, on a synthetic deconvolution problem. Both visually and quantitatively, the results are quite disappointing. Although the introduction of line interactions does produced a favorable effect, it is almost negligible. More precisely, it has been observed on several data sets that the relative decrease of $l_1$ error norm is less than 5%, which is visually nearly imperceptible, and is obviously not worth the extra computational cost required to update interacting auxiliary variables.

It would probably be hazardous to draw definitive conclusions regarding continuous-valued, interacting processes, based on such a first attempt. In particular, a still interesting perspective for further research would be to design more sophisticated convex potentials, based on other existing interacting binary models, such as those proposed in [9] and [21].

## VI. Conclusion

This paper is devoted to convex *half-quadratic* criteria and associated minimization methods. Firstly, we have pointed out that half-quadratic algorithms such as ARTUR, LEGEND [3], iterative reweighted least squares and fast residual steepest descent [17], and others [4], [6], all resort to relaxation on augmented forms of the original criterion.

Under mild conditions, we have shown that the augmented version is convex when the original criterion is convex. This structural property is the keypoint of this study. Its most straightforward consequence is a proof that the various forms of HQ algorithms converge to the global minimizer under sufficient conditions linked to convexity.

Several aspects can be put forward to explain that HQ algorithms have rapidly spread as minimization tools in the field of edge-preserving image restoration, and more specifically in the case of convex penalization. An obvious reason is that HQ algorithms do provide competitive structures in term of convergence speed toward the unique minimizer, compared to well established schemes such as conjugate gradient methods. Until now, this has been ascertained by practice rather than by theoretical studies.

In particular, relaxation on $K_{\rm GY}$ can benefit from the fact that $\partial K_{\rm GY}/\partial \boldsymbol{x} = \boldsymbol{0}$ yields a constant normal matrix $\boldsymbol{M}_{\rm GY} = 2\boldsymbol{M}_0 + \boldsymbol{V}\boldsymbol{V}^{\rm t}$. For instance, consider the spectral estimation problem dealt with in [18] and [22]. Then, $\boldsymbol{M}_{\rm GY}$ is circulant and the normal equation can be inverted using fast Fourier transforms [22], whereas, in comparison, IRLS can only rely on Toeplitz inversion to solve $\partial K_{\rm GR}/\partial \boldsymbol{x} = \boldsymbol{0}$ [18]. Yet, relaxation on $K_{\rm GY}$ is not always faster than IRLS: it is extremely slow to minimize almost non differentiable criteria, whereas IRLS always compares very well with a conjugate gradient method [22].

On the other hand, HQ algorithms (more specifically, relaxation methods on $K_{\rm GR}$, hence the name, *iteratively reweighted least squares*) correspond to adaptive, iterated applications of quadratic regularization. As a consequence, if one is used to compute quadratic regularizers, HQ algorithms provide a natural, easy to implement transition toward the computation of convex, nonquadratic regularizers.

Algorithmic aspects are not the only motivation for studying HQ structures. Specifically, the purpose of Geman & Reynolds's construction is also image modeling: auxiliary variables constitute a continuous-valued extension of a line process without interactions [1] (no equivalent interpretation has been found yet for the auxiliary variables within Geman and Yang's construction). On this basis, a method has been proposed for the construction of convex Gibbsian energies that incorporate interacting auxiliary variables. At least conceptually, such Gibbs–Markov models gather several features that were previously not conciliable. On the one hand, because the resulting criteria are convex, the minimization step remains robust and fairly simple. On the other hand, interacting auxiliary variables provide a flexible mechanism in the field of image modeling. One simple example of construction has been proposed, but the brought improvement appears rather marginal compared to the noninteracting counterpart. A more thorough analysis is left for future research.

## Appendix I

## Proof of Theorem 1

According to the assumptions of Theorem 1, for any $b \in \mathbb{R}$, the function $(x-b)^2/2 - \phi(x)$ is continuous, convex in $x$, "infinite at infinity" as a function of $x$. Thus, there exists at least one finite value $\widetilde{x}_b$ of $x$ such that

$$\zeta(b) = -\frac{1}{2}(\tilde{x}_b - b)^2 + \phi(\tilde{x}_b). \tag{28}$$

In order to prove that $\zeta$ is convex, we have to check that $\forall\, \theta \in (0,1)$, $b_1$, $b_2$

$$\theta\zeta(b_1) + \bar{\theta}\zeta(b_2) - \zeta(b) \geq 0$$

where $\bar{\theta} = 1-\theta$ and $b = \theta b_1 + \bar{\theta} b_2$. According to (9) and (28), $\forall\, x_1,\, x_2 \in \mathbb{R}$

$$\begin{aligned}\theta\zeta(b_1) + \bar{\theta}\zeta(b_2) - \zeta(b) \geq{}& \theta\phi(x_1) + \bar{\theta}\phi(x_2) - \phi(\tilde{x}_b)\\ &- \frac{\theta}{2}(b_1 - x_1)^2 - \frac{\bar{\theta}}{2}(b_2 - x_2)^2 + \frac{1}{2}(b - \tilde{x}_b)^2.\end{aligned}$$

In particular, let us express the latter inequality for the pair $x_1 = \widetilde{x}_b + b_1 - b$, $x_2 = \widetilde{x}_b + b_2 - b$

$$\begin{aligned}\theta\zeta(b_1) + \bar{\theta}\zeta(b_2) - \zeta(b) \geq{}& \theta\phi(\tilde{x}_b + b_1 - b)\\ &+ \bar{\theta}\phi(\tilde{x}_b + b_2 - b) - \phi(\tilde{x}_b).\end{aligned}$$

Since we have $\widetilde{x}_b = \theta(\widetilde{x}_b + b_1 - b) + \bar{\theta}(\widetilde{x}_b + b_2 - b)$, the right-hand side is nonnegative according to the convexity of $\phi$, which proves that $\zeta$ is convex. If $\phi$ is strictly convex, $\zeta$ is obviously strictly convex.

## Appendix II

## Proof of Corollary 1

Given Theorem 1, the first assertion is obvious, since $K_{\rm GY}$ is a sum of convex terms. The proof of the second part of the corollary is not as straightforward. Let us derive the result by contraposition: assume that there exist two distinct pairs $(\boldsymbol{x}_1, \boldsymbol{b}_1)$, $(\boldsymbol{x}_2, \boldsymbol{b}_2)$ and $\theta \in (0,1)$ such that

$$K_{\rm GY}(\boldsymbol{x}, \boldsymbol{b}) = \theta K_{\rm GY}(\boldsymbol{x}_1, \boldsymbol{b}_1) + \bar{\theta} K_{\rm GY}(\boldsymbol{x}_2, \boldsymbol{b}_2)$$

where

$\bar{\theta} \quad = 1 - \theta,$

$\boldsymbol{x} \quad = \theta\boldsymbol{x}_1 + \bar{\theta}\boldsymbol{x}_2$, and

$\boldsymbol{b} \quad = \theta\boldsymbol{b}_1 + \bar{\theta}\boldsymbol{b}_2.$

If $J_0$ is strictly convex, then $\boldsymbol{x}_1 = \boldsymbol{x}_2$, which in turn implies $\boldsymbol{b}_1 = \boldsymbol{b}_2$ since $\sum_{c \in \mathcal{C}} (\boldsymbol{v}_c^{\rm t}\boldsymbol{x} - w_c - b_c)^2$ is a strictly convex function of $\boldsymbol{b}$. We are led to $(\boldsymbol{x}_1, \boldsymbol{b}_1) = (\boldsymbol{x}_2, \boldsymbol{b}_2)$, which contradicts the initial assumption.

Now assume instead that $J$ and $\phi$ are strictly convex. Since the sufficient conditions of Theorem 1 are met, including the strict convexity of $\phi$, $\zeta$ is strictly convex. Given (6), this implies $\boldsymbol{b}_1 = \boldsymbol{b}_2$ and also $\forall\, c \in \mathcal{C}$, $\boldsymbol{v}_c^{\rm t}\boldsymbol{x}_1 - (\boldsymbol{b}_1)_c = \boldsymbol{v}_c^{\rm t}\boldsymbol{x}_2 - (\boldsymbol{b}_2)_c$, and hence, $\forall\, c \in \mathcal{C}, \boldsymbol{v}_c^{\rm t}\boldsymbol{x}_1 = \boldsymbol{v}_c^{\rm t}\boldsymbol{x}_2$. In turn, the latter equality imposes that

$$\begin{aligned}\theta J(\boldsymbol{x}_1)+\bar{\theta}J(\boldsymbol{x}_2)-J(\boldsymbol{x}) &= \theta J_0(\boldsymbol{x}_1)+\bar{\theta}J_0(\boldsymbol{x}_2)-J_0(x)\\ &= \theta K_{\rm GY}(\boldsymbol{x}_1,\boldsymbol{b}_1)+\bar{\theta}K_{\rm GY}(\boldsymbol{x}_2,\boldsymbol{b}_2)\\ &\quad -K_{\rm GY}(\boldsymbol{x},\boldsymbol{b})\\ &= 0\end{aligned}$$

which contradicts the strict convexity of $J$.

Finally, $g^*$ and $\zeta$ are $C^1$ if $g$ is strictly convex [19, Th. 26.3]. Then, $K_{\rm GY}$ is also $C^1$.

## APPENDIX III

## PROOF OF THEOREM 2

Let us begin with a preliminary technical result.

*Lemma 3:* If $\phi$ is a convex function on $\mathbb{R}$ that fulfills (13) and (14), then it also fulfills (15).

*Proof:* First, let us show that $\phi$ necessarily takes finite values on $\mathbb{R}$. On one hand, the value $-\infty$ is excluded because $\phi$ is convex. On the other hand, the value $+\infty$ is also excluded because $\phi(\sqrt{\cdot})$ is concave on $\mathbb{R}_+$ and $\phi$ is even.

As a convex function, $\phi$ admits half-derivatives on $\mathbb{R}$, and $\phi'_-(x)\le\phi'_+(x)$ [19, Th. 24.1]. Simultaneously, the concavity of $\phi(\sqrt{\cdot})$ implies $\phi'_-(x)/x\ge\phi'_+(x)/x$ on $\mathbb{R}_+^*$. Hence, $\phi'_-=\phi'_+$, which shows that $\phi'$ exists as a continuous function on $\mathbb{R}_+^*$ (and on $\mathbb{R}^*$ because $\phi$ is even). □

Now, let us prove Theorem 2 itself. After the change of variable, (18) reads

$$d=\sup_{x\in\mathbb{R}}\left(\phi(x)-\psi^{-1}(d)x^2\right)$$

so, for any $x\in\mathbb{R}$ and any $d\in I_\phi$

$$\psi^{-1}(d)x^2\ge\phi(x)-d. \tag{29}$$

According to Lemma 3, (15) is fulfilled, which implies that at least one finite nonnegative value $\widehat{x}_d$ of $x$ exists such that

$$\psi^{-1}(d)\widehat{x}_d^2=\phi(\widehat{x}_d)-d. \tag{30}$$

We have to show that $\forall\,\theta\in(0,1)$, $x_1,x_2\in\mathbb{R}$, $d_1,d_2\in I_\phi$

$$\theta\psi^{-1}(d_1)x_1^2+\bar{\theta}\psi^{-1}(d_2)x_2^2\ge\psi^{-1}(d)x^2 \tag{31}$$

where

$\bar{\theta} \quad = 1-\theta;$

$x \quad = \theta x_1+\bar{\theta}x_2;$

$d \quad = \theta d_1+\bar{\theta}d_2.$

In the case $x=0$, (31) obviously holds. When $x\neq0$, let us introduce $k=\widehat{x}_d/x$, $x'_1=kx_1$ and $x'_2=kx_2$. Then we have

$$\begin{aligned}&\theta\psi^{-1}(d_1)x_1^2+\bar{\theta}\psi^{-1}(d_2)x_2^2\\ &\quad\ge\frac{\left(\theta\left(\phi(x'_1)-d_1\right)+\bar{\theta}\left(\phi(x'_2)-d_2\right)\right)}{k^2}\\ &\quad\ge\frac{\left(\phi(\widehat{x}_d)-d\right)}{k^2}\\ &\quad=x^2\psi^{-1}(d)\end{aligned}$$

where the last three lines, respectively, derive from (29), from the convexity of $\phi$ (remark that $\theta x'_1+\bar{\theta}x'_2=\widehat{x}_d$), and from (30).

## APPENDIX IV

## PROOF OF THEOREM 3

As a technical preliminary, coordinatewise properties of $\mathcal{K}$ are examined.

*Lemma 4:* Let us assume that the conditions of Theorem 2 apply, and that (25) holds. Then, for each $n$, $\mathcal{K}$ is a strictly convex function of $x_n$. □

*Proof:* The proof is obvious, since $(\boldsymbol{v}_c^{\rm t}\boldsymbol{x}-w_c)^2$ is a strictly convex function of $x_n$ if $(\boldsymbol{v}_c)_n\neq0$, and $\psi^{-1}>0$ in $I_\phi$.

*Lemma 5:* Suppose that the conditions of Theorem 2 apply. Then, for all $\boldsymbol{x}\in\mathbb{R}^N$, $c_0\in\mathcal{C}$, $(d_c)_{c\neq c_0}\in I_\phi^{M-1}$, $\mathcal{K}$ is hemivariate in $d_{c_0}$ on $I_\phi$, i.e., it is not constant along any segment $[\alpha,\beta]\subset I_\phi$ [13, Def. 14.1.1].

*Proof:* If $\boldsymbol{v}_{c_0}^{\rm t}\boldsymbol{x}-w_{c_0}\neq0$, $\mathcal{K}$ is strictly convex in $d_{c_0}$ on $I_\phi$ because $\psi^{-1}$ is strictly convex on $I_\phi$ according to Lemma 2; hence, it is hemivariate. Otherwise, it is an affine function of $d_{c_0}$ of slope one. □

Now, the main part of the proof relies on the following convergence result.

*Proposition 1:* Suppose that $F$ is $C^1$, convex over the Cartesian product

$$Y=Y_1\times Y_2\times\cdots\times Y_L \tag{32}$$

where each $Y_l$ is a closed convex set of $\mathbb{R}^{m_l}$. Furthermore, suppose that for each $l$, $F$ is an hemivariate function of $\boldsymbol{y}_l$. Consider the sequences generated by the block coordinate descent method

$$\boldsymbol{y}_l^{k+1}=\arg\min_{\xi\in Y_l}F\left(\boldsymbol{y}_1^{k+1},\ldots,\boldsymbol{y}_{l-1}^{k+1},\xi,\boldsymbol{y}_{l+1}^k,\boldsymbol{y}_L^k\right). \tag{33}$$

Every limit point of such sequences minimizes $F$ over $Y$.

Proposition 1 is almost a paraphrase of Bertsekas' Proposition 2.7.1 in [15], except that Bertsekas assumes that $F$ is block coordinatewise strictly convex, instead of block coordinatewise hemivariate only (remark that a strictly convex function is hemivariate). Incidentally, his demonstration of [15, Prop. 2.7.1] actually proves the slightly stronger Proposition 1, so the reader can refer to it as a valid proof of Proposition 1.

Such a result does not apply directly to $K_{\rm GR}$ (because $K_{\rm GR}$ is not necessarily convex), but it applies to $F=\mathcal{K}$ under the conditions of Theorem 3, for $\boldsymbol{y}=(\boldsymbol{x},\boldsymbol{d})$, $L=M+N, Y=X\times I_\phi^M$, since $\mathcal{K}$ is then

- $C^1$ because $\psi^{-1}$ is $C^1$ on $I_\phi$ (the latter holds if $b_\infty<+\infty$ only);
- convex according to Corollary 2;
- coordinate-wise hemivariate according to Lemmas 4 and 5.

Now the proof is completed with the following remark: if $(\boldsymbol{x}^k,\boldsymbol{b}^k)$ is a series of iterates obtained by coordinate descent on $K_{\rm GR}$, then $\left(\boldsymbol{x}^k,\boldsymbol{\psi}(\boldsymbol{b}^k)\right)$ would be obtained by coordinate descent on $\mathcal{K}$.

## REFERENCES

[1] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 367–383, 1992.

[2] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Image Processing*, vol. 4, pp. 932–946, 1995.
[3] P. Charbonnier, L. Blanck-Féraud, G. Aubert, and M. Barlaud, "Two deterministic half-quadratic regularization algorithms for computed imaging," in *Proc. IEEE Int. Conf. Image Processing*, vol. 2, Austin, TX, Nov. 1994, pp. 168–172.
[4] S. Brette and J. Idier, "Optimized single site update algorithms for image deblurring," in *Proc IEEE Int. Conf. Image Processing*, Lausanne, Switzerland, Sept. 1996, pp. 65–68.
[5] 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, 1997.
[6] L. D. Cohen, "Auxiliary variables and two-step iterative algorithms in computer vision problems," *J. Math. Imag. Vis.*, vol. 6, pp. 59–83, 1996.
[7] G. Aubert and L. Vese, "A variational method in image recovery," *SIAM J. Numer. Anal.*, vol. 34, pp. 1948–1979, 1997.
[8] R. V. Vogel and M. E. Oman, "Fast, robust total variation-based reconstruction of noisy, blurred images," *IEEE Trans. Image Processing*, vol. 7, pp. 813–823, June 1998.
[9] J. L. Marroquin, "Probabilistic solution of inverse problems," Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, 1985.
[10] F. C. Jeng and J. W. Woods, "Compound Gauss-Markov random fields for image estimation," *IEEE Trans. Signal Processing*, vol. 39, pp. 683–697, Mar. 1991.
[11] D. Geiger and F. Girosi, "Parallel and deterministic algorithms from MRF's: Surface reconstruction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, pp. 401–412, May 1991.
[12] A. Blake and A. Zisserman, *Visual Reconstruction*. Cambridge, MA: MIT Press, 1987.
[13] J. Ortega and W. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*. New York: Academic, 1970.
[14] R. Glowinski, J. L. Lions, and R. Trémolières, *Analyse numérique des inéquations variationnelles, tome 1 : Théorie générale, Méthodes mathématiques pour l'informatique*. Paris, France: Dunod, 1976.
[15] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena, 1995.
[16] A. H. Delaney and Y. Bresler, "Globally convergent edge-preserving regularized reconstruction: An application to limited-angle tomography," *IEEE Trans. Image Processing*, vol. 7, pp. 204–221, Jan. 1998.
[17] 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, 1985.
[18] 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, 1998.
[19] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1970.
[20] S. A. Ruzinsky and E. T. Olsen, "$L_1$ and $L_\infty$ minimization via a variant of Karmarkar's algorithm," *IEEE Trans. Signal Processing*, vol. 37, pp. 245–253, Feb. 1989.
[21] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 721–741, 1984.
[22] P. Ciuciu, "Méthodes markoviennes en estimation spectrale non paramétrique. Applications en imagerie radar Doppler," Ph.D. dissertation, Univ. Paris-Sud, Orsay, France, 2000.

**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, Gif-sur-Yvette. His major scientific interest is in probabilistic approaches to inverse problems for signal and image processing.