On Global and Local Convergence of Half-Quadratic Algorithms

Marc Allain, Jérôme Idier, and Yves Goussard

***Abstract*—This paper provides original results on the global and local convergence properties of half-quadratic (HQ) algorithms resulting from the Geman and Yang (GY) and Geman and Reynolds (GR) primal-dual constructions. First, we show that the convergence domain of the GY algorithm can be extended with the benefit of an improved convergence rate. Second, we provide a precise comparison of the convergence rates for both algorithms. This analysis shows that the GR form does not benefit from a better convergence rate in general. Moreover, the GY iterates often take advantage of a low cost implementation. In this case, the GY form is usually faster than the GR form from the CPU time viewpoint.**

***Index Terms*—Algorithms, asymptotic rate, convergence analysis, half-quadratic (HQ) iterations, image reconstruction, image restoration, robust statistics.**

I. Introduction

THE SOLUTION to many image restoration and reconstruction problems is defined as follows:

$$\boldsymbol{x}^\star \in \arg\min_{\boldsymbol{x}\in\mathbb{R}^N} J(\boldsymbol{x}) \tag{1}$$

where $J : \mathbb{R}^N \to \mathbb{R}$ is an estimation criterion (or objective function) that must be minimized. The solution $\boldsymbol{x}^\star$ can rarely be expressed in closed form, and its approximate evaluation is generally carried out with an iterative numerical procedure. In the past 12 years, several such procedures based on the notion of Fenchel duality [1] have been proposed [2], [3]. The approach consists of reformulating problem (1) as the minimization of an *augmented criterion* which presents a half-quadratic (HQ) structure. Minimization of HQ objective functions can be implemented easily using an iterative block-relaxation procedure. Each iteration is naturally expressed in closed-form, thereby avoiding the determination of a stepsize parameter using a nested, line-search procedure. Such a simplification gives an appealing, ready-made feature to the resulting algorithms. This is a major reason to the ever-growing use of algorithms based on HQ constructions. Several signal and image processing applications are concerned, such as synthetic aperture radar [4], magnetic resonance imaging [5], and spectrometry [6], to mention some recent examples.

Currently, two types of derivations, respectively referred to as the Geman and Reynolds (GR) and Geman and Yang (GY) constructions, yield two distinct forms of HQ criteria. Since their introduction in the early 1990s, these constructions and the corresponding algorithms have been mainly used in the area of penalized image reconstruction and restoration. However, in the field of robust statistics, essentially the same algorithms have been proposed without introducing HQ criteria, but by linearizing the optimality condition of problem (1) [7]. Global convergence of these algorithms, referred to as *Iterative Reweighted Least-Square* (IRLS) and *Residual Steepest Descent* (RSD), relies on the construction of a local quadratic model that bounds criterion J from above. Therefore, these procedures appear as special cases of *generalized Weiszfeld algorithms* introduced as early as 1937 [8].

In this paper, we show that convergence for the same algorithms can also be studied in the general framework of unconstrained optimization. This approach provides new results on the impact of some parameters associated with either the GR or the GY construction. More precisely, two original results are presented: 1) a relaxation of the global convergence conditions of GY algorithms which yields faster algorithmic forms, and 2) a precise direct comparison of the asymptotic convergence speeds of GR and GY algorithms which provides guidelines for selection of one form or the other according to the characteristics of the problem at hand.

In this paper, the objective function J in (1) is assumed to take the following form:

$$J(\boldsymbol{x}) = Q(\boldsymbol{x}) + \Phi(\boldsymbol{x}) \tag{2}$$

where Q is a *convex quadratic form* that can be expressed as

$$Q(\boldsymbol{x}) = \boldsymbol{x}^t\boldsymbol{Q}\boldsymbol{x} - 2\boldsymbol{q}^t\boldsymbol{x} + \mu. \tag{3}$$

In the previous expression, $\boldsymbol{Q} \in \mathbb{R}^{N\times N}$ is a *symmetric, positive semidefinite* matrix, and $\cdot^t$ denotes the transpose operator. Function Φ is defined as

$$\Phi(\boldsymbol{x}) = \sum_{i=1}^{I} \phi(\delta_i)$$
$$\delta_i \triangleq [\boldsymbol{V}^t\boldsymbol{x} - \boldsymbol{w}]_i$$

where $\phi : \mathbb{R} \to \mathbb{R}$ is a scalar function, and where $\boldsymbol{V}$ and $\boldsymbol{w}$, respectively, denote an $(N \times I)$ matrix and a $(I \times 1)$ vector with

Manuscript received August 11, 2004; revised April 25, 2005. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Nicolas Rougon.

M. Allain and J. Idier are with the Institut de Recherche en Communications et en Cybernétique de Nantes (IRCCyN), BP 92 101-44321 Nantes Cedex 03, France (e-mail: marc.allain@irccyn.ec-nantes.fr; Jerome.Idier@irccyn.ec-nantes.fr).

Y. Goussard is with the Department of Electrical Engineering, École Polytechnique de Montréal, Succ. Centre-ville Montréal, QC H3C 3A7, Canada (e-mail: yves.goussard@polymtl.ca).

Digital Object Identifier 10.1109/TIP.2005.864173

real components. Notation $[\boldsymbol{u}]_i$ is used to denote the ith component of any vector $\boldsymbol{u}$. The previous framework is general enough to encompass a wide class of imaging and data processing problems for which a *penalized least-squares* solution is sought: typically, the solution is defined as the minimum of a compound criterion that can be expressed as follows:

$$J(\boldsymbol{x}) = ||\boldsymbol{G}\boldsymbol{x} - \boldsymbol{y}||^2 + \alpha\mathcal{P}(\boldsymbol{x}), \quad \alpha \geq 0 \tag{4}$$

where $\boldsymbol{y} \in \mathbb{R}^M$ and $\boldsymbol{G}$, respectively, denote the observed data vector and a linear operator that models the data formation process, e.g., a convolution matrix in image restoration [2], [3], [9], [10], or a projection operator in tomographic reconstruction [11]–[13]. The role of the penalization term $\mathcal{P}$, whose weight is set through regularization parameter α, is to guarantee the uniqueness of the solution, its robustness with respect to (w.r.t.) observation noise and its fidelity to some priors [14]. In image restoration, one may choose to obtain *locally smooth* solutions by penalizing the intensity differences between adjacent pairs of pixels, i.e.,

$$\mathcal{P}(\boldsymbol{x}) = \sum_{c=1}^{C} \psi\left([\boldsymbol{D}^t\boldsymbol{x}]_c\right) \tag{5}$$

where $\boldsymbol{D}^t \in \mathbb{R}^{C\times N}$ is a difference operator between pairs $c \in \{1,\ldots,C\}$ of neighboring pixels and where $\psi : \mathbb{R} \to \mathbb{R}$ typically represents an *edge-preserving* function, e.g., l_p norm with $1 \leq p < 2$, Huber, hyperbolic, or Geman and McClure functions [7], [15]–[17]. Note that our convergence analysis can be extended to the *inhomogeneous* case $\Phi = \sum_i \phi_i$ in a straightforward manner. This allows one to address signal processing applications such as autoregressive modeling [18] or digital filter synthesis [19].

The paper is organized as follows. In Section II, the two algorithms studied in the paper are defined precisely; the two classical convergence analysis frameworks are also briefly presented in order to put our work in perspective. Section III is dedicated to the study of the global convergence of the GR and GY algorithms; original convergence results are derived using unconstrained optimization tools. In Section IV, the asymptotic convergence speed of the algorithms is investigated as well as some issues related to the cost of implementation. Numerical examples are used to illustrate some of our results. Finally, a discussion of this work and conclusions are presented in Section V.

II. Studied Algorithms: Different Mathematical Standpoints

A. Iterative Scheme of Interest

We first introduce the two matrix operators $\boldsymbol{B}^a_{\mathrm{GY}}$ and $\boldsymbol{B}_{\mathrm{GR}}$ that will help us define the algorithms studied in this paper. We have

$$\boldsymbol{B}^a_{\mathrm{GY}} \triangleq 2\boldsymbol{Q} + \frac{1}{a}\boldsymbol{V}\boldsymbol{V}^t \tag{6}$$

where $a > 0$ is a free parameter, and

$$\boldsymbol{B}_{\mathrm{GR}}(\boldsymbol{x}) \triangleq 2\boldsymbol{Q} + \boldsymbol{V}\boldsymbol{L}(\boldsymbol{x})\boldsymbol{V}^t \tag{7}$$

with $\boldsymbol{L}(\boldsymbol{x}) \triangleq \mathrm{diag}(\phi'(\delta_i)/\delta_i)$. For the sake of simplicity, we borrowed notations from the HQ framework: the "GY" and "GR" subscripts refer to the HQ constructions presented in Section II-B. Provided that operators (6) and (7) are invertible and bounded, two iterative forms of interest are now defined.

Given an initial vector $\boldsymbol{x}^{(0)}$, the following algorithm will be called *GY iterative scheme*:

$$\begin{aligned} \forall k \geq 0, \quad \boldsymbol{\xi}^{(k)}_{\mathrm{GY}} &= -\left(\boldsymbol{B}^a_{\mathrm{GY}}\right)^{-1}\boldsymbol{\nabla}J^{(k)} \\ \boldsymbol{x}^{(k+1)} &= \boldsymbol{x}^{(k)} + \theta\boldsymbol{\xi}^{(k)}_{\mathrm{GY}} \end{aligned} \tag{8}$$

where $\boldsymbol{\nabla}J^{(k)} \equiv \boldsymbol{\nabla}J(\boldsymbol{x}^{(k)})$ is the gradient of J for the current update, and $\theta > 0$ is a *constant* stepsize. Let us also define the *GR iterative scheme* as follows:

$$\begin{aligned} \forall k \geq 0, \quad \boldsymbol{\xi}^{(k)}_{\mathrm{GR}} &= -\left(\boldsymbol{B}^{(k)}_{\mathrm{GR}}\right)^{-1}\boldsymbol{\nabla}J^{(k)} \\ \boldsymbol{x}^{(k+1)} &= \boldsymbol{x}^{(k)} + \theta\boldsymbol{\xi}^{(k)}_{\mathrm{GR}} \end{aligned} \tag{9}$$

with $\boldsymbol{B}^{(k)}_{\mathrm{GR}} \equiv \boldsymbol{B}_{\mathrm{GR}}(\boldsymbol{x}^{(k)})$. Notice that both schemes are *Newtonian iterations with a constant stepsize* [20, Chap. 3].

B. Half-Quadratic Formalism

We now provide brief background information on the HQ formalism. The reader is referred to [2] and [3] for pioneering contributions, or to [21] for a synthetic overview. Within the HQ framework, an *augmented* criterion $J^* : \mathbb{R}^N \times L \to \mathbb{R}$ is first introduced. It is related to J by the relation

$$\inf_{\boldsymbol{\ell}\in L} J^*(\boldsymbol{x},\boldsymbol{\ell}) = J(\boldsymbol{x})$$

where $\boldsymbol{\ell} = (\ell_1,\ldots,\ell_I)^t$ is a vector of *dual* variables. By construction, J^* is quadratic w.r.t. $\boldsymbol{x}$ when $\boldsymbol{\ell}$ remains constant, hence the terminology "half-quadratic." Moreover, the minimizer of J^* w.r.t. $\boldsymbol{\ell}$ when $\boldsymbol{x}$ is constant can be expressed in closed-form in commonly found HQ constructions. Such properties naturally lead to *relaxation* schemes for the optimization of J^*, i.e., the HQ criterion is minimized along the primal and dual variables $(\boldsymbol{x},\boldsymbol{\ell})$ in an alternate fashion. More generally, *over-* or *under-relaxed* updates can be introduced: $\forall k \in \mathbb{N}$

$$\boldsymbol{\ell}^{(k+1)} = (1-\beta)\boldsymbol{\ell}^{(k)} + \beta\widehat{\boldsymbol{\ell}}^{(k+1)} \tag{10a}$$

$$\boldsymbol{x}^{(k+1)} = (1-\theta)\boldsymbol{x}^{(k)} + \theta\widehat{\boldsymbol{x}}^{(k+1)} \tag{10b}$$

where

$$\begin{aligned} \widehat{\boldsymbol{\ell}}^{(k+1)} &= \arg\min_{\boldsymbol{\ell}\in L} J^*\left(\boldsymbol{x}^{(k)},\boldsymbol{\ell}\right) \\ \widehat{\boldsymbol{x}}^{(k+1)} &= \arg\min_{\boldsymbol{x}\in\mathbb{R}^N} J^*\left(\boldsymbol{x},\boldsymbol{\ell}^{(k+1)}\right) \end{aligned} \tag{11}$$

and the parameters β, θ are relaxation coefficients. The expression of $\widehat{\boldsymbol{x}}^{(k+1)}$ and $\widehat{\boldsymbol{\ell}}^{(k+1)}$ depends on which of the GY [3] or GR [2] construction is chosen. Both versions were initially introduced for optimizing nonconvex criteria with simulated annealing. Subsequently, they have been mainly used in deterministic relaxation framework for convex or nonconvex optimization.

The following two subsections give a brief overview of GY and GR HQ constructions, respectively.

1) GY HQ Construction: Let g be the scalar function defined by $g(u;a) \triangleq u^2/2 - a\phi(u)$. The "scale parameter" $a > 0$ is a free parameter that modifies the convergence properties of GY algorithm. Provided that the following convexity condition holds:

$$g(\cdot;a) \text{ is convex} \tag{12}$$

it is possible to introduce an augmented HQ criterion $J^*_{\mathrm{GY}}(\boldsymbol{x},\boldsymbol{\ell})$ (for an exact definition of J^*_{GY}, see [3]). According to [21, Lem. 1], the set of values of a that fulfill (12) is an interval $(0;\widehat{a})$, where

$$\widehat{a} = \sup\{a > 0 : g(\cdot;a) \text{ is convex}\}. \tag{13}$$

Given the structural properties of J^*_{GY}, the updating scheme (11) admits the more explicit formulation [21, Sec. III-A]

$$\widehat{\ell}_i^{(k+1)} = \delta_i^{(k)} - a\phi'\left(\delta_i^{(k)}\right), \quad i = 1,\dots,I \tag{14a}$$

$$\widehat{\boldsymbol{x}}^{(k+1)} = (\boldsymbol{B}^a_{\mathrm{GY}})^{-1}\left(2\boldsymbol{q} + \frac{\boldsymbol{V}\left(\boldsymbol{\ell}^{(k+1)} + \boldsymbol{w}\right)}{a}\right). \tag{14b}$$

When $\beta = 1$, i.e., without dual over- or under-relaxation,[1] substituting (14a) into (14b) yields that (10b) coincides with (8); see [22]. Within the HQ framework, the algorithm (14) has been proposed in [23] under the name of LEGEND. Convergence results derived from structural properties of J^*_{GY} are presented in [21]; provided that (12) holds *strictly* and ϕ is strictly convex with the additional condition

$$\lim_{|u|\to\infty} \frac{\phi(u)}{u^2} < \frac{1}{2a} \tag{15}$$

J^*_{GY} is shown to exist and to be strictly convex and continuously differentiable (C^1). Then, the relaxation scheme (10) converges to $\boldsymbol{x}^\star$ according to [21, Sec. III-C] if $\beta \in (0;1)$ and (θ,a) belongs to

$$\Gamma_{\mathrm{s}} \triangleq (0;2) \times (0;\widehat{a}). \tag{16}$$

Finally, let us mention that the GY construction can be further generalized to a larger family of algorithms [24].

2) GR HQ Construction: Another type of HQ construction was introduced by Geman and Reynolds [2]: it yields an augmented HQ criterion J^*_{GR} provided that ϕ fulfills the following three conditions:

$$\begin{cases} \phi \text{ is even} & \text{(17a)} \\ \phi \text{ is } C^1 & \text{(17b)} \\ \phi(\sqrt{\cdot}) \text{ is concave on } \mathbb{R}^+. & \text{(17c)} \end{cases}$$

According to [21, Sec. IV-A], alternate minimization of J^*_{GR} reads as follows:

$$\widehat{\ell}_i^{(k+1)} = \frac{\phi'\left(\delta_i^{(k)}\right)}{\delta_i^{(k)}}, \quad i = 1,\dots,I \tag{18a}$$

$$\widehat{\boldsymbol{x}}^{(k+1)} = \left(\boldsymbol{B}^{(k)}_{\mathrm{GR}}\right)^{-1}\left(2\boldsymbol{q} + \boldsymbol{V}\boldsymbol{L}^{(k+1)}\boldsymbol{w}\right). \tag{18b}$$

[1]Choosing $\beta \neq 1$ is of marginal interest: on the one hand, convergence is not proved for $\beta > 1$. On the other hand, practical evidence indicates that $\beta < 1$ slows the convergence rate down.

For $\beta = 1$, substituting (18a) within (18b) yields that (10b) coincides with (9); e.g., see [4, p. 627]. Within the HQ framework, the resulting algorithm was proposed by Charbonnier *et al.* under the name of ARTUR [16], [23]. Provided that condition (17c) holds *strictly* and that ϕ is convex and fulfills the additional technical conditions

$$\begin{cases} \lim\limits_{|u|\to\infty} \frac{\phi'(u)}{u} = 0 & \text{(19a)} \\ \lim\limits_{|u|\to 0} \frac{\phi'(u)}{u} < \infty & \text{(19b)} \end{cases}$$

then (10) converges to $\boldsymbol{x}^\star$ for $\beta \in (0;1)$ and $\theta \in (0;2)$. In the nonconvex case, weaker results can still be obtained according to [25].

C. IRLS, RSD, and Majorizing Quadratic Approximations

Another way of introducing GY and GR iterative schemes relies on the notion of *local quadratic model*. More precisely, let us introduce the following second-order approximation of J in the neighborhood of $\boldsymbol{x}$:

$$\widehat{J}(\boldsymbol{x}^+;\boldsymbol{x}) \triangleq J(\boldsymbol{x}) + (\boldsymbol{x}^+ - \boldsymbol{x})^t \boldsymbol{\nabla} J(\boldsymbol{x}) + \frac{1}{2}(\boldsymbol{x}^+ - \boldsymbol{x})^t \boldsymbol{B}(\boldsymbol{x})(\boldsymbol{x}^+ - \boldsymbol{x}) \tag{20}$$

where $\boldsymbol{B} : \mathbb{R}^N \to \mathbb{R}^{N\times N}$ is *positive definite* (PD). Then, let an iterative scheme be defined as follows:

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \theta\left(\widehat{\boldsymbol{x}}^{(k+1)} - \boldsymbol{x}^{(k)}\right) \tag{21}$$

where

$$\begin{aligned} \widehat{\boldsymbol{x}}^{(k+1)} &= \arg\min_{\boldsymbol{x}\in\mathbb{R}^N} \widehat{J}\left(\boldsymbol{x};\boldsymbol{x}^{(k)}\right) \\ &= \boldsymbol{x}^{(k)} - \left(\boldsymbol{B}\left(\boldsymbol{x}^{(k)}\right)\right)^{-1} \boldsymbol{\nabla} J\left(\boldsymbol{x}^{(k)}\right). \end{aligned}$$

When $\boldsymbol{B} = \boldsymbol{B}^a_{\mathrm{GY}}$, the iterative scheme (21) is clearly of the GY type (8). Similarly, it is a GR scheme (9) when $\boldsymbol{B} = \boldsymbol{B}_{\mathrm{GR}}$. Such schemes were introduced in the late 1970s in the context of robust estimation, under the name of *residual steepest descent* and *iterative reweighted least-square* algorithms, respectively. In the literature on robust estimation, the *majorizing* character of $\widehat{J}$ w.r.t. J is the essential element used to prove the convergence of iterative schemes such as (8) and (9) (see [26, A-7] and [7, Sec. 7.8]). More precisely, we have the following result, which extends [27, Lem. 4.3] to cope with both GY and GR schemes.

Proposition 1: Let J be defined by (2) where ϕ is a convex, C^1 function. If $\boldsymbol{B} \equiv \boldsymbol{B}^a_{\mathrm{GY}}$ [resp. $\boldsymbol{B} \equiv \boldsymbol{B}_{\mathrm{GR}}$] and assumption (12) [resp. (17)] holds, then

$$\forall \boldsymbol{x}, \boldsymbol{x}^+ \in \mathbb{R}^N, \quad \widehat{J}(\boldsymbol{x}^+;\boldsymbol{x}) \geq J(\boldsymbol{x}^+). \tag{22}$$

Proof: See Appendix A. ∎

The majorization approach traces back to *generalized Weiszfeld* algorithms [8], [27], [28]. It is a powerful framework to analyze important algorithmic structures such as EM type algorithms [29], in which cases the majorizing approximation is not necessarily quadratic. Specifically, quadratic majorizing approximations have also been fruitful in image reconstruction or restoration [30]–[33].

The majorization framework provides slightly weaker convergence conditions than the HQ setting. In particular, convergence can be proved for GY and GR schemes without tech-

nical conditions (15) and (19a), respectively (see [24, Sec. X.4], whose study is based on [28]). Yet, convergence conditions can still be weakened within a more general framework, as shown in the next section.

III. New Convergence Results

In this section, we introduce general tools to study *constant stepsize*, *gradient related iterative schemes*, as found in a fundamental contribution such as [35]. We then check that GY and GR algorithms are special cases of such schemes. We finally deduce new global convergence results, which bring a substantial improvement in the GY case.

In this section, it is assumed that ϕ is a C^1 function, i.e., nondifferentiable penalization functions ϕ fall out of the scope of our study. More specifically, this excludes $\phi(u) = |u|$ as found in image restoration using *total variation* regularization, unless nondifferentiability is gradually introduced (e.g., see [36]).

A. Preliminary Results

Let us begin by giving some background results about Newtonian schemes [20, Chap. 3]. Given an initial vector $\boldsymbol{x}^{(0)}$, one generates a sequence $\{\boldsymbol{x}^{(k)}\}_{k\in\mathbb{N}}$ defined by

$$\begin{aligned} \forall k \geq 0, \quad \boldsymbol{\xi}^{(k)} &= -\left(\boldsymbol{B}^{(k)}\right)^{-1} \boldsymbol{\nabla} J^{(k)} \\ \boldsymbol{x}^{(k+1)} &= \boldsymbol{x}^{(k)} + \theta^{(k)} \boldsymbol{\xi}^{(k)} \end{aligned} \tag{23}$$

where $\boldsymbol{B}^{(k)} \equiv \boldsymbol{B}(\boldsymbol{x}^{(k)}) : \mathbb{R}^N \to \mathbb{R}^{N\times N}$ is a PD operator, and $\theta^{(k)} > 0$ denotes the stepsize for the current direction $\boldsymbol{\xi}^{(k)}$. For sake of brevity, let us introduce the following compact notations:

$$\chi \equiv \left\{\boldsymbol{x}^{(k)}\right\}_{k\in\mathbb{N}}, \quad \Xi \equiv \left\{\boldsymbol{\xi}^{(k)}\right\}_{k\in\mathbb{N}}, \; \Theta \equiv \left\{\theta^{(k)}\right\}_{k\in\mathbb{N}}.$$

In the sequel, the triplet (χ, Ξ, Θ) will be called an *iterative scheme*. The goal is to build series Ξ and Θ such that χ converges (preferably rapidly) to $\boldsymbol{x}^\star$. Here, we focus on *descent methods*, i.e., on iterative schemes such that $J(\boldsymbol{x}^{(k)})$ is nonincreasing with k. This property does not suffice to ensure that χ converges to $\boldsymbol{x}^\star$; however, the following additional properties on sequences Ξ and Θ will lead to convergence [35, Chap. 14]:

$$\begin{cases} \Xi \text{ is gradient related to } \chi & \text{(24a)} \\ \Theta \text{ is admissible (in a sense precised below).} & \text{(24b)} \end{cases}$$

Property (24a) means that all subsequences extracted from Ξ remain bounded, and that none of them asymptotically becomes orthogonal to $\boldsymbol{\nabla} J$. The following proposition gives useful sufficient conditions to ensure that Ξ is gradient related to χ [35, Sec. 14.4.1].

Proposition 2: Let J be C^1 on a compact set $D_0 \subset \mathbb{R}^N$, and $\boldsymbol{B} : D_0 \to \mathbb{R}^{N\times N}$ be a PD operator for which $\exists \gamma_2 \geq \gamma_1 > 0$ such that

$$\forall \boldsymbol{u} \in D_0, \, \boldsymbol{v} \in \mathbb{R}^N, \quad \gamma_2 \|\boldsymbol{v}\|^2 \geq \boldsymbol{v}^t \boldsymbol{B}(\boldsymbol{u}) \boldsymbol{v} \geq \gamma_1 \|\boldsymbol{v}\|^2.$$

If Ξ is generated by (23), then Ξ is gradient related to χ.

Property (24b) is relative to a *stepsize rule* that ensures, for all iterations k, that $\theta^{(k)}$ is neither too large (otherwise χ could diverge) nor too small (otherwise χ could converge to a nonstationary point[2]). Checking *Armijo inequality* is a simple solution to prevent $\theta^{(k)}$ from being too large. It corresponds to

$$J\left(\boldsymbol{x}^{(k+1)}\right) - J\left(\boldsymbol{x}^{(k)}\right) \leq \omega \theta^{(k)} \boldsymbol{\nabla}^t J^{(k)} \boldsymbol{\xi}^{(k)} \tag{25}$$

where $\omega \in (0;1)$ is independent of k (see Fig. 1).

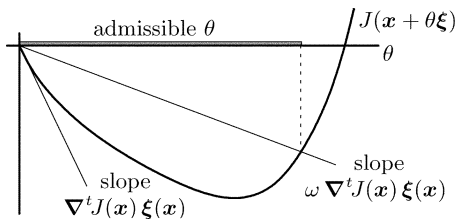


Fig. 1. Illustration of Armijo inequality: for any strictly descending direction $\boldsymbol{\xi}$, (25) selects those values of $\theta > 0$ for which the decrease of J is at least a fraction ω of the value resulting from the local linear approximation at the current point.

Additionally, the *backtracking technique* [37, p. 29] prevents the values of $\theta^{(k)}$ to become too small: the principle is to test (25) using candidates of the form $\theta^{(k)} = s\tau^m$, $m = 0, 1, \ldots$, where $s > 0$ and $\tau \in (0;1)$ are constant parameters (they must not vary with k). The retained value for $\theta^{(k)}$ is the largest (i.e., the first found) for which (25) holds. Here, we adopt Bertsekas' definition of *Armijo rule*, which incorporates the backtracking technique [37, p. 29].

Definition 1 [Armijo Rule]: Let J be a C^1 function and (χ, Ξ, Θ) an iterative scheme. The sequence Θ is admissible for (χ, Ξ) in the sense of Armijo if there exists $\omega \in (0;1)$, $s > 0$ and $\tau \in (0;1)$ such that, for all $k \in \mathbb{N}$, $\theta^{(k)}$ is obtained by the backtracking technique on inequality (25).

If Ξ is gradient related to χ and Θ is admissible for (χ, Ξ) in the sense of Armijo, then every accumulation point is a stationary point [37, Prop. 1.2.1]. However, such a result does not provide a global convergence guarantee, and the following proposition is of practical interest.

Proposition 3: Let J be a C^1 function bounded from below and coercive.[3] Let Ξ be gradient related to χ and Θ be admissible in the sense of Armijo. If the local minimizers of J are isolated, then χ converges to a local minimizer for any $\boldsymbol{x}^{(0)}$. Moreover, if J is strictly convex, then χ converges to the global minimizer $\boldsymbol{x}^\star$ for any $\boldsymbol{x}^{(0)}$.

Proof: See Appendix B. ∎

The boundedness assumption is a minimal prerequisite for a correctly formulated optimization problem. Coerciveness is also fairly natural, since it ensures that J is minimized for finite solutions. Given the structure (2) of the criterion, this property holds either if $\boldsymbol{Q}$ is invertible, or if $\boldsymbol{Q} + \boldsymbol{V}\boldsymbol{V}^t$ is invertible and ϕ is coercive.

If J is strictly convex and coercive, global convergence occurs; in our context, these conditions hold if

$$\begin{cases} \boldsymbol{Q} + \boldsymbol{V}\boldsymbol{V}^t \text{ is invertible} & \text{(26a)} \\ \phi \text{ is strictly convex, coercive.} & \text{(26b)} \end{cases}$$

[2]Stationary points are those that cancel $\boldsymbol{\nabla} J$.

[3]Coerciveness means that $J(\boldsymbol{x})$ tends to infinity for any $\boldsymbol{x}$ that tends to infinity: $\lim_{n\to\infty} J(\boldsymbol{x}_n) = \infty$ for all series $\{\boldsymbol{x}_n\}_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty} \|\boldsymbol{x}_n\| = \infty$.

Finally, the only pathological case of practical interest left out by Proposition 3 is when some local minimizers form a continuum. In this case, global convergence to a single minimizer is not granted, see [37, p. 73] for further considerations on such singular problems.

Clearly, admissibility of Θ is a cornerstone for convergence of a gradient related sequence Ξ (see Proposition 3). In the following, the Armijo rule will be used to establish the convergence of the HQ iterations.

B. Armijo Rule for Constant Stepsize Iterative Schemes

According to (8) and (9), it is clear that GY and GR algorithms are iterative schemes (23) with constant stepsizes. In the sequel, iterative schemes with constant stepsizes are denoted (χ, Ξ, θ), where $\theta > 0$ is the stepsize. GY and GR schemes are more specifically denoted $(\chi, \Xi, \theta)_{\mathrm{GY}}$ and $(\chi, \Xi, \theta)_{\mathrm{GR}}$, respectively. Let us now investigate conditions on (χ, Ξ, θ) such that the Armijo rule holds for θ.

Proposition 4: Let (χ, Ξ, θ) be a constant stepsize iterative scheme. If $\exists \omega \in (0;1)$ such that

$$\forall \boldsymbol{x} \in \mathbb{R}^N,\ J(\boldsymbol{x} + \theta \boldsymbol{\xi}(\boldsymbol{x})) - J(\boldsymbol{x}) \leq \omega \theta \boldsymbol{\nabla}^t J(\boldsymbol{x}) \boldsymbol{\xi}(\boldsymbol{x}) \quad (27)$$

then the constant sequence $\Theta = \theta$ is admissible in the sense of Armijo.

Proof: See Appendix C. ∎

In the sequel, a constant stepsize θ is said to be admissible in the sense of Armijo if (27) holds for some $\omega \in (0;1)$.

The following proposition shows that (22) implies (27), i.e., admissibility in the sense of Armijo is a potentially weaker condition than the majorization tool introduced in Section II-C.

Proposition 5: Let $\widehat{J}$ be defined by (20), and for all $\boldsymbol{x}$, let

$$\boldsymbol{\xi}(\boldsymbol{x}) = -\left(\boldsymbol{B}(\boldsymbol{x})\right)^{-1} \boldsymbol{\nabla} J(\boldsymbol{x}). \quad (28)$$

If inequality (22) holds, then Armijo inequality (27) is true for $\omega = 1 - (\theta/2)$.

Proof: See Appendix D. ∎

C. Convergence Analysis for GY Algorithm

For $(\chi, \Xi, \theta)_{\mathrm{GY}}$, we establish now that Ξ is gradient related to χ, and that θ is admissible in the sense of Armijo, under suitable conditions.

Proposition 6: If (26a) holds, then the GY scheme generates a series Ξ that is gradient related to χ.

Proof: See Appendix F. ∎

Proposition 7: Let ϕ be C^1 with a $(1/\widehat{a})$–Lipschitz[4] derivative. Then θ is admissible in the sense of Armijo for the GY scheme if (θ, a) belongs to

$$\Gamma_{\mathrm{e}} \triangleq \left\{ (\theta, a) : \theta \in (0;2), a \in \left(0; \frac{2\widehat{a}}{\theta}\right) \right\}. \quad (29)$$

Proof: See Appendix G. ∎

Based on the previous two propositions, Proposition 4 readily applies, and convergence for the GY iterations is ensured, according to Proposition 3.

[4]function $f : \mathbb{R}^n \to \mathbb{R}^m$ is said to be *L-Lipschitz* if $\exists L > 0 : \forall \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$, $\|f(\boldsymbol{u}) - f(\boldsymbol{v})\| \leq L \|\boldsymbol{u} - \boldsymbol{v}\|$.

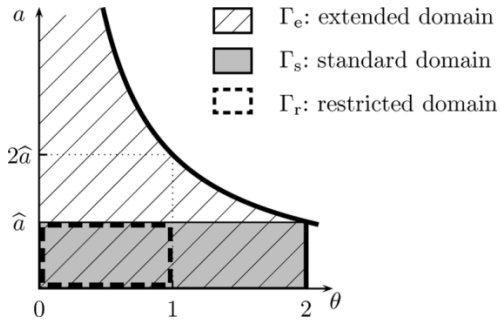


Fig. 2. Comparison between the new *extended* convergence domain Γ_{e} and the *standard* convergence domain $\Gamma_{\mathrm{s}} \subset \Gamma_{\mathrm{e}}$ obtained using either the HQ framework or the majorization approach. The depicted *restricted* domain $\Gamma_{\mathrm{r}} \subset \Gamma_{\mathrm{s}}$ is relative to comparison between convergence rates (see Section IV).

For a nonconvex criterion, these convergence results are difficult to compare with previous works. Nonetheless, a direct comparison in the *convex case* is possible, provided that a useful equivalence between condition (12) and the Lipschitz character of ϕ' is established.

Lemma 1: Let ϕ be a convex, C^1 function. The following two assertions are equivalent:

$$g(\cdot; \widehat{a}) = \frac{(\cdot)^2}{2} - \widehat{a}\phi \text{ is convex} \quad (30a)$$

$$\phi' \text{ is } L - \text{Lipschitz with } L = \frac{1}{\widehat{a}}. \quad (30b)$$

Moreover, if (30) holds, then $\forall a \in (0; \widehat{a})$, $g(\cdot; a)$ is strictly convex and $|\phi'(u) - \phi'(v)| < L|u - v|$ for all $u \neq v$.

Proof: See Appendix E. ∎

Lemma 1 shows that (30a) and (30b) are equivalent means to ensure that GY iterations can be made convergent under suitable conditions. However, it should be noted that the HQ framework (and the majorization approach) provides more stringent convergence conditions in comparison with our analysis. Whereas the former requires that (θ, a) belongs to Γ_{s} [cf. (16)], the latter establishes convergence in a significantly extended domain Γ_{e} [cf. (29)]; see Fig. 2. Moreover, extending the convergence domain has an important practical interest: it will be checked in Section IV that the best convergence rate is usually achieved in the extended part of Γ_{e} (i.e., in $\Gamma_{\mathrm{e}} \setminus \Gamma_{\mathrm{s}}$). Finally, remark that technical condition (15) is not required in our analysis (it was a prerequisite in the HQ framework only).

D. Convergence Analysis for GR Algorithm

The framework of Section III-B applies to GR scheme as well. However, it is not as fruitful as in the GY case, since it does not significantly weaken existing convergence conditions as obtained in the convex case by [27, Lem. 4.3], or in the nonconvex case by [25]. To save space, convergence of the GR iterations is now established for the convex case only.

Proposition 8: If (26a) holds, and ϕ is a convex, even, and C^1 function such that $\forall u \in \mathbb{R}$, $\phi'(u)/u < \infty$, then the GR scheme generates a series Ξ that is gradient related to χ.

Proof: See [34, p. 181]. ∎

Proposition 9: Let ϕ be an even and C^1 function such that $\phi(\sqrt{\cdot})$ is concave on $\mathbb{R}^+$. Then θ is admissible in the sense of Armijo for the GR scheme if $\theta \in (0;2)$.

Proof: See [34, p. 181]. ∎

Based on the previous two propositions, Proposition 4 readily applies, and convergence for the GR iterations is ensured in the convex case, according to Proposition 3.

IV. Convergence Speed

Until now, the speed of convergence of GY and GR schemes has been scarcely addressed. In [27], it is established that the convergence order of the GR algorithm is linear. More recently, a similar result has been proved for the GY algorithm by Nikolova *et al.* [38], [39]. On the other hand, very few comparison results exist between the convergence rate of both schemes, except [39, Sec. 4.2], which compares bounds on *quotient*-convergence[5] rates. Akin to previous works devoted to the convergence speed of EM type algorithms [30], [31], the present study rather relies on *root*-convergence[6] factors, for which closed-form expressions are available.

In this section, iterations are granted to converge to the global minimizer or to a local one, according to Proposition 3; in the sequel, $\boldsymbol{x}^\infty$ is introduced as a generic notation for the minimizer reached by the iterations.

A. Asymptotic Convergence Factors

In the whole section, it is assumed that ϕ is twice continuously differentiable. In order to study the convergence speed of constant stepsize iterative schemes (χ, Ξ, θ), let us define the application $\mathcal{M}_\theta$ according to

$$\mathcal{M}_\theta(\boldsymbol{x}) \triangleq \boldsymbol{x} - \theta \boldsymbol{B}^{-1}(\boldsymbol{x}) \nabla J(\boldsymbol{x}) \tag{31}$$

which allow to recursively define χ from a *one-step stationary iteration*

$$\boldsymbol{x}^{(k+1)} = \mathcal{M}_\theta\left(\boldsymbol{x}^{(k)}\right) \tag{32}$$

in the terminology of [35, Chap. 10]. Let us also introduce the following spectral radius:

$$\sigma(\theta) \triangleq \rho\left(\boldsymbol{I} - \theta \boldsymbol{B}^{-1}(\boldsymbol{x}^\infty)\boldsymbol{H}(\boldsymbol{x}^\infty)\right) \tag{33}$$

where $\boldsymbol{I}$ is the identity matrix and $\boldsymbol{H}$ is the Hessian matrix of J

$$\boldsymbol{H}(\boldsymbol{x}) = 2\boldsymbol{Q} + \boldsymbol{V}\mathrm{diag}\left(\phi''(\delta_i)\right)\boldsymbol{V}^t. \tag{34}$$

Then, the *linear convergence theorem* [35, Th. 10.1.4] ensures that root-convergence of (32) is at least linear, and that $\sigma(\theta)$ is the corresponding rate, provided that $\sigma(\theta) < 1$ (recall that the convergence speed decreases as the convergence rate increases).

Following [37, A.13], let us remark that (33) also reads

$$\sigma(\theta) = \max\{1 - \theta m, \theta M - 1\} \tag{35}$$

[5]We say that the convergence in quotient is at least linear with a convergence rate γ if there is a constant $\gamma \in (0,1)$ such that $\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^\star\| \leq \gamma \|\boldsymbol{x}^{(k)} - \boldsymbol{x}^\star\|$ for all k sufficiently large.

[6]We say that the convergence in root is at least linear with a convergence rate σ if there is a constant $\sigma \in (0,1)$ such that $\limsup_{k\to\infty} \|\boldsymbol{x}^{(k)} - \boldsymbol{x}^\star\|^{1/k} \leq \sigma$; see [35, Sec. 9] for an extensive discussion of both convergence rates.

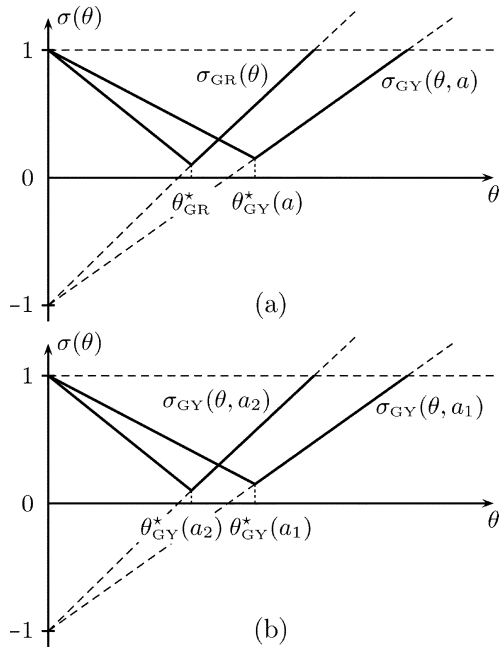


Fig. 3. Consequences of inequalities (37) and (38) on the root-convergence rates (35) for (a) GR scheme versus GY scheme for $a \in (0; \widehat{a}]$, and (b) for GY scheme with two different values a_1 and a_2; $a_1 \leq a_2$ of the scale parameter.

with

$$\begin{aligned} m &= \lambda_{\min}\left(\boldsymbol{B}^{-1}(\boldsymbol{x}^\infty)\boldsymbol{H}(\boldsymbol{x}^\infty)\right) \\ M &= \lambda_{\max}\left(\boldsymbol{B}^{-1}(\boldsymbol{x}^\infty)\boldsymbol{H}(\boldsymbol{x}^\infty)\right) \end{aligned} \tag{36}$$

where $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ are respectively the smallest and largest eigenvalue of a given square matrix. From (35), it appears that $\sigma(\theta)$ is a piecewise linear function of θ (see Fig. 3).

The following proposition ensures that $\sigma(\theta) < 1$ for both GY and GR schemes.

Proposition 10: Let $\sigma_{\mathrm{GY}}(\theta, a)$ [resp. $\sigma_{\mathrm{GR}}(\theta)$] be defined by (33) for $\boldsymbol{B} \equiv \boldsymbol{B}^a_{\mathrm{GY}}$ [resp. $\boldsymbol{B} \equiv \boldsymbol{B}_{\mathrm{GR}}(\boldsymbol{x}^\infty)$]. Inequality $\sigma_{\mathrm{GY}}(\theta, a) < 1$ [resp. $\sigma_{\mathrm{GR}}(\theta) < 1$] holds under the conditions of Propositions 6 and 7 [resp. Propositions 8 and 9].

Proof: See Appendix H. ∎

Regarding linear convergence of the GY scheme, previous results (based on quotient-convergence factors) were limited to Γ_{s} [38], [39], whereas ours is available over Γ_{e}.

In the purely quadratic case $[\phi(u) = u^2]$, both GY and GR schemes can reach a *superlinear* convergence since they identify with Newton iterates for suitable values of tuning parameters: on the one hand, $\sigma_{\mathrm{GR}} = 0$ for $\theta = 1$, and on the other hand, $\sigma_{\mathrm{GY}} = 0$ for $\theta = 1$ and $a = \widehat{a} = 0.5$.

In realistic nonquadratic cases, the convergence order is only linear and the two algorithms have distinct convergence rates. Convergence rates are significant elements for measuring the overall efficiency of algorithmic structures. The convergence rates of GY and GR schemes are more thoroughly analyzed in the following two subsections. Nonetheless, it must be kept in mind that convergence rates are not the only elements to consider. First, they are meaningless about the algorithmic efficiency during earlier iterations. Second, they do not take the computational cost per iteration into account. The latter question is discussed in Subsection IV-C.

The following proposition provides basic inequalities to compare the convergence rates of GY and GR schemes.

Proposition 11: Let m^a_{GY}, M^a_{GY} [resp. m_{GR}, M_{GR}] the smallest and largest eigenvalues of $(\boldsymbol{B}^a_{\mathrm{GY}})^{-1}\boldsymbol{H}(\boldsymbol{x}^\infty)$ [resp. $\boldsymbol{B}^{-1}_{\mathrm{GR}}(\boldsymbol{x}^\infty)\boldsymbol{H}(\boldsymbol{x}^\infty)$]. We have

$$\forall a \leq \widehat{a},\ \begin{cases} m^a_{\mathrm{GY}} \leq m_{\mathrm{GR}} & \text{(37a)} \\ M^a_{\mathrm{GY}} \leq M_{\mathrm{GR}} & \text{(37b)} \end{cases}$$

and

$$\forall a_1, a_2,\ a_1 \leq a_2 \Longrightarrow \begin{cases} m^{a_1}_{\mathrm{GY}} \leq m^{a_2}_{\mathrm{GY}} & \text{(38a)} \\ M^{a_1}_{\mathrm{GY}} \leq M^{a_2}_{\mathrm{GY}}. & \text{(38b)} \end{cases}$$

Proof: See Appendix I. ∎

Given (35) and (37), Fig. 3(a) depicts the evolution of σ_{GR} and σ_{GY} as functions of θ within Γ_{s}. Fig. 3(a) also displays the minimizer

$$\theta^\star_{\mathrm{GY}}(a) = \frac{2}{m^a_{\mathrm{GY}} + M^a_{\mathrm{GY}}}, \quad \theta^\star_{\mathrm{GR}} = \frac{2}{m_{\mathrm{GR}} + M_{\mathrm{GR}}} \qquad (39)$$

of σ_{GR} and $\sigma_{\mathrm{GY}}(\cdot, a)$, respectively. Note that the corresponding optimal rate takes the following expression:

$$\sigma^\star_{\mathrm{GY}}(a) = \min_\theta \sigma_{\mathrm{GY}}(\theta, a) = \frac{M^a_{\mathrm{GY}} - m^a_{\mathrm{GY}}}{m^a_{\mathrm{GY}} + M^a_{\mathrm{GY}}}$$
$$\sigma^\star_{\mathrm{GR}} = \min_\theta \sigma_{\mathrm{GR}}(\theta) = \frac{M_{\mathrm{GR}} - m_{\mathrm{GR}}}{m_{\mathrm{GR}} + M_{\mathrm{GR}}}$$

Similarly, Fig. 3(b) depicts the evolution of σ_{GY} for two distinct values a_1 and a_2 of the scale parameter. Remark that Fig. 3(b) is not only valid within Γ_{s} but in the whole set Γ_{e}.

Inequalities $\sigma_{\mathrm{GR}}(\theta) \leq \sigma_{\mathrm{GY}}(\theta, a)$ or $\min_\theta \sigma_{\mathrm{GR}}(\theta) \leq \min_{\theta,a} \sigma_{\mathrm{GY}}(\theta, a)$ are not true in general. Nonetheless, some restricted inequalities can be established on the basis of Proposition 11, according to the following corollary.

Corollary 1: We have the following inequalities:

$$\forall (\theta, a) \in \Gamma_{\mathrm{s}}, \quad \sigma_{\mathrm{GR}}(\theta) \leq \sigma_{\mathrm{GY}}(\theta, a), \quad \text{if } \mathrm{Ker}(\boldsymbol{V}^t) \neq \{\boldsymbol{0}\} \qquad (40a)$$

$$\forall (\theta, a) \in \Gamma_{\mathrm{r}}, \quad \sigma_{\mathrm{GR}}(1) \leq \sigma_{\mathrm{GR}}(\theta) \leq \sigma_{\mathrm{GY}}(\theta, a) \qquad (40b)$$

$$\forall (\theta, a_2) \in \Gamma_{\mathrm{s}}, \forall a_1 \leq a_2, \quad \sigma_{\mathrm{GY}}(\theta, a_2) \leq \sigma_{\mathrm{GY}}(\theta, a_1) \quad \text{if } \mathrm{Ker}\,(\boldsymbol{V}^t) \neq \{\boldsymbol{0}\} \qquad (40c)$$

$$\forall (\theta, a_2) \in \Gamma_{\mathrm{r}}, \forall a_1 \leq a_2, \quad \sigma_{\mathrm{GY}}(1, a_2) \leq \sigma_{\mathrm{GY}}(\theta, a_2) \leq \sigma_{\mathrm{GY}}(\theta, a_1) \qquad (40d)$$

where $\Gamma_{\mathrm{r}} \triangleq \{(\theta, a) \in (0;1) \times (0; \widehat{a})\}$ (see Fig. 2).

Proof: See Appendix J. ∎

Corollary 1 provides some useful information. In particular, it appears that

- $\sigma_{\mathrm{GR}}(\theta) \leq \sigma_{\mathrm{GY}}(\theta, a)$ is only true under restrictive conditions (numerical examples, e.g., Fig. 7, confirms that this inequality is not always true);
- $\theta < 1$ is never optimal for the GR scheme, which corroborates the empirical observation that over-relaxation yields faster convergence;
- similarly, $(\theta, a) \in \Gamma_{\mathrm{r}}$ is never optimal for the GY scheme;
- according to (40c), $a < \widehat{a}$ is never optimal for the GY scheme if $\mathrm{Ker}(\boldsymbol{V}^t) \neq \{\boldsymbol{0}\}$, which covers many useful cases where $\boldsymbol{V}^t$ corresponds to a finite difference matrix (since $\boldsymbol{V}^t\boldsymbol{1} = \boldsymbol{0}$ usually holds then).

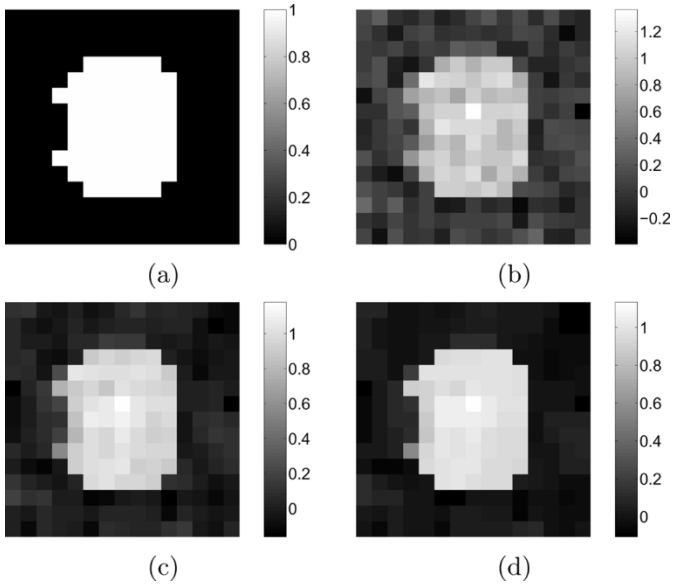


Fig. 4. Transmission tomography example. (a) Binary phantom $\boldsymbol{x}^{\mathrm{true}}$, and (b)–(d) three reconstruction results $\boldsymbol{x}^\infty_1$, $\boldsymbol{x}^\infty_2$, $\boldsymbol{x}^\infty_3$ obtained by minimization of criterion (4), with $\alpha = 5$, and $s = 1$, $s = 0.1$, and $s = 0.01$, respectively.

According to the latter point, exploring the extended part of the convergence domain (i.e., $a > \widehat{a}$) could yield faster versions of GY scheme in practically important cases. However, we have been unable to establish mathematical inequalities between convergence rates when $a > \widehat{a}$. Hence, we proceed with our study using numerical examples.

B. Numerical Simulations

In this subsection, the behavior of the asymptotic rates $\sigma_{GY}(\theta, a)$ and $\sigma_{\mathrm{GR}}(\theta)$ is illustrated through a transmission tomography problem.

1) Problem Formulation: Tomographic reconstruction from a small number of projections is an ill-posed inverse problem. Hence, a regularized solution is computed through the minimization of the compound criterion (4), where $\boldsymbol{G}$ is a sparse matrix produced by the *Radon transform* of a pixel-based object; see [40].

Data $\boldsymbol{y}$ consists of 36 noisy projections generated from the synthetic binary object[7] presented in Fig. 4(a); the additive noise is Gaussian and white, with a zero mean and unity standard deviation. This corresponds to an empirical signal-to-noise ratio close to 20 dB.

The penalization term $\mathcal{P}$ is given by (5), where $\boldsymbol{D}^t$ corresponds to a first-order difference operator. The chosen convex "edge preserving" function is given by $\psi(u; s) = \sqrt{u^2 + s^2}$ (see Fig. 5), in which case, $\widehat{a} = s$.

The solution $\boldsymbol{x}^\infty$ of the reconstruction problem is computed by an iterative scheme (either GY or GR). The adopted stopping rule is $J(\boldsymbol{x}^{(k)}) - J(\boldsymbol{x}^{(k+1)}) < \varepsilon$, where we have chosen $\varepsilon = 10^{-12}$. Such a value was empirically determined as a limit of

[7]A small-sized imaging problem was chosen in order to allow the computation of the spectral radius arising from the root convergence rate expression.

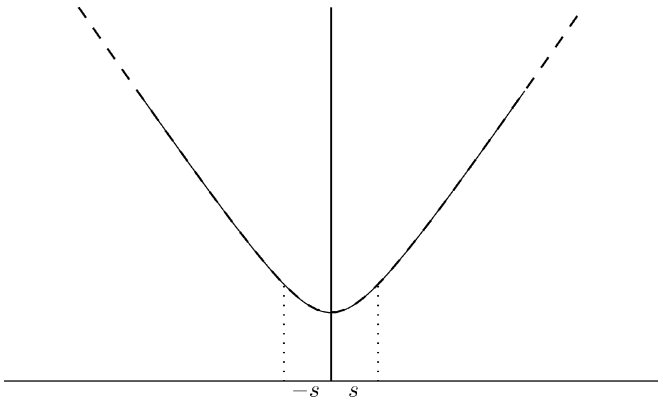


Fig. 5. Edge preserving function $\psi(u;s) = \sqrt{u^2+s^2}$. For $u \in [-s, s]$, ψ roughly behaves like the quadratic function $s + u^2/2s$. For larger values of u, it has rather a linear behavior.

numerical reliability (lower values of ε may produce unstable iterates).

2) Results: The results represent the typical behavior of the algorithms as observed on a larger set of experiments. The reconstructions $\boldsymbol{x}_1^\infty$, $\boldsymbol{x}_2^\infty$, and $\boldsymbol{x}_3^\infty$ shown, respectively, in Fig. 4 were produced with $\alpha = 5$ and three distinct values of parameter $s \in \{1, 0.1, 0.01\}$. In each case, the level sets of $\sigma_{\mathrm{GY}}(\theta, a)$ computed from (33) are shown in Fig. 6, where the symbols ⓔ, Ⓢ, and ⓡ locate the best rate for Γ_{e}, Γ_{s}, and Γ_{r}, respectively. These rates and the rate for the GR iterations are gathered in Table I. Finally, Tables II and III, group the number of iterations and the CPU time (the `cputime` command within MATLAB) needed for convergence, respectively.

For reconstruction $\boldsymbol{x}_1^\infty$, the penalization ψ behaves almost like a *quadratic* function since the dynamic in the image is much lower than the "threshold" $s = 1$. Hence, as it was noticed in Section IV-A, the best rate is achieved for $a \simeq \widehat{a}$, with σ nearly zero (i.e., superlinear) if θ is equal to one; cf. Table I.

As the parameter s decreases (solutions $\boldsymbol{x}_2^\infty$, and $\boldsymbol{x}_3^\infty$), ψ takes more values in the low curvature areas. The point ⓔ moves away from $\widehat{a}$, reaches the hyperbolic boundary of the convergence domain [Fig. 6(b)], and follows it from left to right [Fig. 6(c)].

The "intermediate" situation $\boldsymbol{x}_2^\infty$ is of practical interest since it represents a satisfactory tradeoff between the convergence speed of the algorithm and the edge-preserving behavior of the objective function. It should be underlined that a significant reduction in the asymptotic rate of the GY algorithm when $a > \widehat{a}$ can be expected precisely in this situation. Note also that even a limited decrease in the rate can significantly reduce the number of iterations needed for convergence. For instance, the computation of $\boldsymbol{x}_2^\infty$ needs 30% more iterations when the rate changes from 0.782 (Ⓢ) to 0.771 (ⓔ).

For reconstruction $\boldsymbol{x}_3^\infty$, the optimum rate in Γ_{e} is very close to the best rate in Γ_{s}. In this case, choosing a outside Γ_{s} can only produce marginal benefit.

The convergence rate greatly deteriorates when the maximum curvature (the Lipschitz constant) of the problem increases. Hence, small s leads to slower convergence for both algorithmic forms. However, the GR rate seems to be more robust suggesting that the curvature information captured by matrix $\boldsymbol{B}_{\mathrm{GR}}$ allows to better deal with "quasi-nondifferentiable" ψ; see also [39, Sec. 4.2].

TABLE I
VALUES OF CONVERGENCE RATE σ_{GY} CORRESPONDING TO REMARKABLE POINTS IN FIG. 6(a)–(c), RESPECTIVELY

		$s = 1$	$s = 0.1$	$s = 0.01$
ⓔ	$\sigma_{\mathrm{GY}} \simeq$	**0.146**	**0.771**	**0.977**
	$(\theta, a) \simeq$	$(1.11, 1.04)$	$(1.15, 0.17)$	$(1.974, 0.0101)$
Ⓢ	$\sigma_{\mathrm{GY}} \simeq$	**0.151**	**0.782**	**0.9773**
	$\theta \simeq$	1.15	1.78	1.977
ⓡ	$\sigma_{\mathrm{GY}} \simeq$	**0.2624**	**0.8774**	**0.989**
GR	$\sigma^\star_{\mathrm{GR}} \simeq$	**0.09**	**0.46**	**0.80**
	$\theta \simeq$	1.09	1.46	1.80

TABLE II
NUMBER OF ITERATIONS FOR CONVERGENCE CORRESPONDING TO TABLE I

	$s = 1$	$s = 0.1$	$s = 0.01$
ⓔ	8 it.	45 it.	563 it.
Ⓢ	9 it.	64 it.	632 it.
ⓡ	9 it.	85 it.	761 it.
GR	7 it.	21 it.	72 it.

TABLE III
CPU TIME IN SECONDS FOR CONVERGENCE CORRESPONDING TO TABLE I

	$s = 1$	$s = 0.1$	$s = 0.01$
ⓔ	0.09 s.	0.53 s.	6.97 s.
Ⓢ	0.11 s.	0.84 s.	7.35 s.
ⓡ	0.11 s.	1.02 s.	9.41 s.
GR	0.78 s.	2.36 s.	7.92 s.

In practice, the rate also greatly depends on the tuning of the regularization parameter. For ill-conditioned problems, for instance, a better rate should arise as the parameter grows provided that the penalization improves the condition number. Our analysis does not convey insight on this important problem, and further investigations are required.

In these numerical examples, the best rate is always achieved by the GR scheme (cf. Table I). This is the most common situation in practice; however, some counter-examples show that inequalities (40a) and (40b) cannot be generalized to Γ_{e}. For instance, minimizing criterion (4), with $s = 1$ and $\alpha = 0.005$, reaches an under-regularized tomographic reconstruction (not shown here) with the rates $\sigma_{\mathrm{GR}}(\theta)$ and $\min_a \sigma_{\mathrm{GY}}(\theta, a)$ depicted in Fig. 7.

C. Asymptotic Rate and Computation Burden

From Table III, GY iterates benefit from a faster convergence *w.r.t. the CPU time* since their computation cost is much lower [the inverse of $\boldsymbol{B}^a_{\mathrm{GY}}$ was computed once for all in the update (14b)].

For more realistic problem sizes, computing the inverse of $\boldsymbol{B}^a_{\mathrm{GY}}$ is untractable. However, $\boldsymbol{B}^a_{\mathrm{GY}}$ often enjoys structural properties that lead to low-cost updates. This is the case for some denoising [24], [41] and deconvolution problems [39], [42]. Moreover, *approximately* solving each linear system (14b)

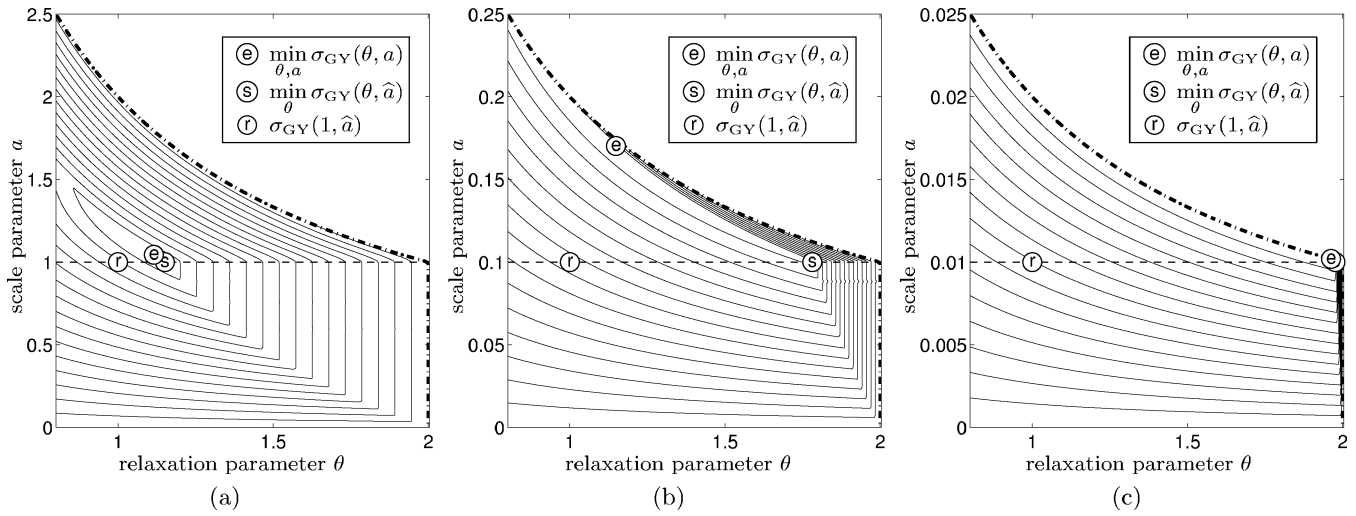


Fig. 6. Level sets of $\sigma_{\mathrm{GY}}(\theta, a)$ computed from (33) on a 200×200 grid covering all the domain Γ_{e}. These level sets are associated with the reconstructions $\boldsymbol{x}_1^{\infty}$, $\boldsymbol{x}_2^{\infty}$, and $\boldsymbol{x}_3^{\infty}$ shown in Fig. 4. The dashed line corresponds to $a = \widehat{a}$; according to Corollary 1, the best rate achieved in Γ_{e}, Γ_{s}, and Γ_{r} corresponds to the symbol ⓔ, ⓢ, and ⓡ, respectively.

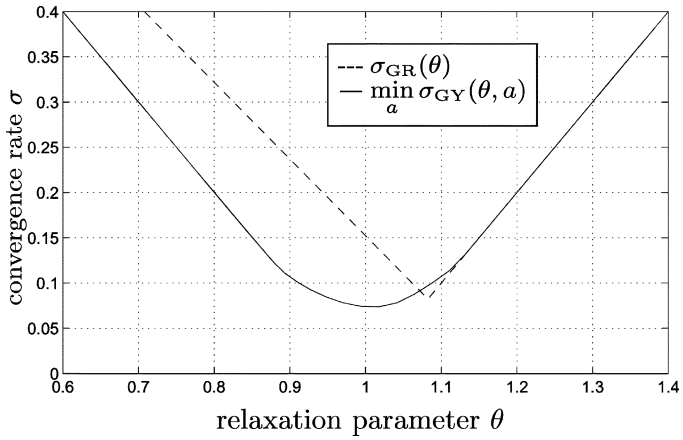


Fig. 7. Counter-example to the conjecture: "the best optimum rate is always achieved by the GR iterates."

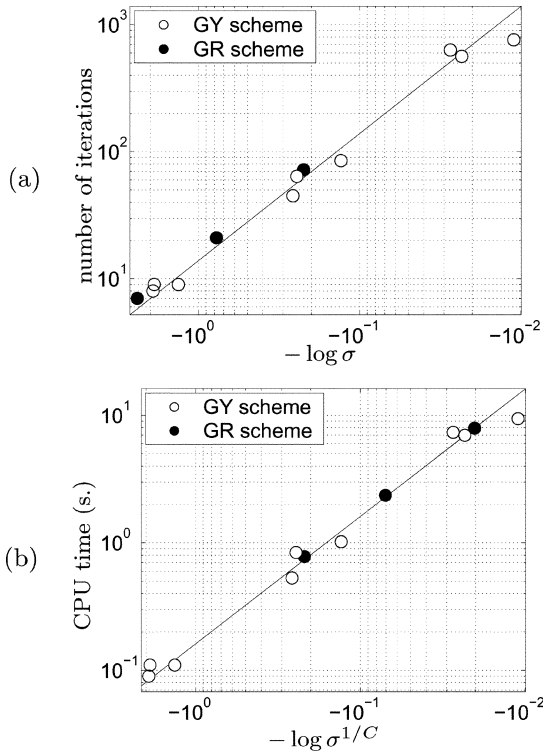


Fig. 8. (a) Number of iterations and (b) CPU time needed for convergence in monoton relationship with (a) the convergence rate and (b) the rate per operation.

allows one to adapt the GY form to more general problems, for instance in the fields of deconvolution [43] and tomography [16], [25].

These elements reveal that the convergence rate only partly measures the "merit" of an algorithm. A direct extension of the *quotient*-convergence rate is now introduced in order to take into account both the convergence rate *and* the computation burden per iteration. Let us recall that an algorithm with linear quotient-convergence rate γ satisfies the following inequality [35, Sec. 9]:

$$\left\| \boldsymbol{x}^{(k)} - \boldsymbol{x}^{\star} \right\| \leq \gamma \left\| \boldsymbol{x}^{(k-1)} - \boldsymbol{x}^{\star} \right\|$$

w.r.t. a norm $\| \cdot \|$ on $\mathbb{R}^N$. Therefore, if each iteration costs C operations, the *convergence rate per operation* $\gamma^{1/C}$ is a natural means for taking the computer burden into consideration.

Note, however, that it is not clear that a similar measure can be introduced so simply from the *root*-convergence rate. Nevertheless, it has been verified in practice that the quantity $\sigma^{1/C}$ is a good indicator of the CPU time needed for an algorithm with σ root-convergence rate. The tomographic problem presented in Section IV-B can be used as an example. For both algorithmic forms (nine instances for the GY scheme and three for the GR scheme, cf. Table II), Fig. 8(a) depicts the number of iterations K needed to achieve convergence as a function of the convergence rate σ (Table I). These points follow the same monotonic curve. This was expected since the number of iterations depends only on the convergence rate and is independent of the considered algorithm. The relationship between K and σ is easily deduced from Fig. 8(a) and reads $K(\sigma) \propto$

$1/\log\sigma$, which is the theoretical relationship expected for the quotient-convergence rate. The CPU time $T_{\mathrm{CPU}} = C \times K \propto C/\log\sigma$ (Table III) depends on the rate and on the computation cost C of the algorithm. In contrast, the rate per operation $\sigma_{\mathrm{op}} = \sigma^{1/C}$ is in a monotonic relationship with T_{CPU} and does not depend on the algorithm since $T_{\mathrm{CPU}} \propto 1/\log\sigma^{1/C}$, as shown in Fig. 8(b). Hence, two iterative forms with distinct rates and computation costs can be formally compared by mean of the rate per operation.

V. Conclusion

The GY and GR iterative forms are remarkable since they benefit from global convergence without line-search and from a decent asymptotic rate. In this paper, our main objective was to sharpen our knowledge of the convergence properties of these algorithms. Standard results from unconstrained optimization were used in order to perform a new analysis. Most of the new results concern the GY form. First, the sufficient conditions for global convergence of this algorithm can be noticeably weakened. Second, our study shows that faster forms exist with the GY algorithm. Moreover, the GY is a very attractive choice when matrix $\boldsymbol{B}^a_{\mathrm{GY}}$ can be inverted at a low cost. For many imaging problems, however, the fast inversion of $\boldsymbol{B}^a_{\mathrm{GY}}$ is not possible, and we are presently developing distinct HQ forms in order to preserve both the numerical efficiency and the global convergence without line-search; see [24] for an example.

Appendix

Hereafter, the notation $\boldsymbol{A} \succ 0$ (resp. $\boldsymbol{A} \succeq 0$) indicates that $\boldsymbol{A}$ is a positive definite (resp. positive semidefinite) matrix.

A. Proof of Proposition 1

Since $J = Q + \Phi$ and Q is quadratic, it is easy to deduce the following expression of $\widehat{J}$ from (20): $\forall \boldsymbol{x}, \boldsymbol{x}^+ \in \mathbb{R}^N$

$$\widehat{J}(\boldsymbol{x}^+;\boldsymbol{x}) - J(\boldsymbol{x}^+) = \Phi(\boldsymbol{x}) - \Phi(\boldsymbol{x}^+) + (\boldsymbol{x}^+ - \boldsymbol{x})^t \boldsymbol{\nabla}\Phi(\boldsymbol{x}) + \frac{1}{2}(\boldsymbol{x}^+ - \boldsymbol{x})^t \boldsymbol{P}(\boldsymbol{x})(\boldsymbol{x}^+ - \boldsymbol{x}) \quad (41)$$

with $\boldsymbol{P}(\boldsymbol{x}) = \boldsymbol{V}\boldsymbol{V}^t/a$ if $\widehat{J} \equiv \widehat{J}_{\mathrm{GY}}$ and $\boldsymbol{P}(\boldsymbol{x}) = \boldsymbol{V}\mathrm{diag}(\phi'(\delta_i)/\delta_i)\boldsymbol{V}^t$ if $\widehat{J} \equiv \widehat{J}_{\mathrm{GR}}$, respectively. It remains to show that (41) is nonnegative under assumption (12) [resp., (17)]. In the case $\boldsymbol{B} \equiv \boldsymbol{B}_{\mathrm{GR}}$, the proof is almost identical to that of [27, Lem. 4.3], and it will not be reproduced here.

In the other case $\boldsymbol{B} \equiv \boldsymbol{B}_{\mathrm{GY}}$, let us first notice that (41) also reads

$$\widehat{J}(\boldsymbol{x}^+;\boldsymbol{x}) - J(\boldsymbol{x}^+) = \sum_i \left(\phi(\delta_i) - \phi\left(\delta_i^+\right) + \Delta_i \phi'(\delta_i) + \frac{\Delta_i^2}{2a} \right)$$

where $\Delta_i = \delta_i^+ - \delta_i$ and $\delta_i^+ = [\boldsymbol{V}^t\boldsymbol{x}^+ - \boldsymbol{w}]_i$. On the other hand, let us remark that ϕ' is L–Lipschitz with $L = 1/\widehat{a}$ according to (13) and to Lemma 1. Then, the *descent lemma* [37, Prop. A.24] yields

$$\phi(\delta_i) - \phi\left(\delta_i^+\right) + \Delta_i \phi'(\delta_i) + \frac{\Delta_i^2}{2\widehat{a}} \geq 0$$

which shows $\widehat{J}(\boldsymbol{x}^+;\boldsymbol{x}) \geq J(\boldsymbol{x}^+)$ since $a < \widehat{a}$.

B. Proof of Proposition 3

Since J is coercive and bounded from below, all the level sets are compacts, and convergence toward at least one accumulation point is granted, according to Cauchy Theorem. The gradient related character of Ξ being assumed, it can be shown that every accumulation point is stationary [37, Prop. 1.2.1]. If J has only isolated mimina, convergence to multiple stationnary point would contradict the "capture theorem" [37, 1.2.5], and χ converges to a single stationnary point $\boldsymbol{x}^\infty$. Finally, $\boldsymbol{x}^\infty$ is the global minimizer $\boldsymbol{x}^\star$ in the strictly convex case, since this property ensures that every stationary point is the global minimum.

C. Proof of Proposition 4

Condition (27) ensures that (25) holds with $\theta^{(n)} = \theta$ for $n = 0$ and for all subsequent iterations. Moreover, it is obvious that $s = \theta$ can be considered as obtained from the backtracking technique. Hence, the constant series $\Theta = \theta$ is admissible in the sense of Armijo.

D. Proof of Proposition 5

For $\boldsymbol{x}^+ = \boldsymbol{x} + \theta\boldsymbol{\xi}$, the majorizing character (22) of $\widehat{J}$ reads: $\forall \boldsymbol{x}, \theta, \boldsymbol{\xi}$

$$J(\boldsymbol{x}) + \theta\boldsymbol{\xi}^t\boldsymbol{\nabla}J(\boldsymbol{x}) + \frac{\theta^2\boldsymbol{\xi}^t\boldsymbol{B}(\boldsymbol{x})\boldsymbol{\xi}}{2} \geq J(\boldsymbol{x}^+).$$

In particular, let us consider $\boldsymbol{\xi} = \boldsymbol{\xi}(\boldsymbol{x})$, as a solution of the normal equation (28): $\forall \boldsymbol{x}, \theta$

$$J(\boldsymbol{x}) - J(\boldsymbol{x}^+) + \theta\left(1 - \frac{\theta}{2}\right)\boldsymbol{\xi}^t(\boldsymbol{x})\boldsymbol{\nabla}J(\boldsymbol{x}) \geq 0.$$

The latter identifies with the Armijo inequality (25) for $\omega = 1 - \theta/2$.

E. Proof of Lemma 1

Function $g(\cdot;a) = (\cdot)^2/2 - a\phi$ is C^1 since ϕ is C^1. For all a, u, v, $u \leq v$

$$\begin{aligned} g'(u,a) - g'(v,a) &= a\left(\phi'(v) - \phi'(u)\right) - (v-u) \\ &= a\left|\phi'(v) - \phi'(u)\right| - |v-u| \end{aligned} \quad (42)$$

where the latter equality holds because ϕ' is nondecreasing. From (42), it becomes obvious that (30a) and (30b) are equivalent, i.e., $g'(\cdot,\widehat{a})$ is nondecreasing if and only if ϕ' is $(1/\widehat{a})$–Lipschitz. Finally, strict convexity of $g(\cdot;a)$ holds for $0 < a < \widehat{a}$ [21, Sec. III-B], and it is clear from (42) that the Lipschitz inequality becomes strict.

F. Proof of Proposition 6

Proposition 6 is a consequence of Proposition 2. Since matrix $\boldsymbol{B}^a_{\mathrm{GY}} = 2\boldsymbol{Q} + \boldsymbol{V}\boldsymbol{V}^t/a$ remains constant during the iterations, its eigenvalues take constant finite values. Moreover, since $\boldsymbol{Q} \succeq 0$ and $\boldsymbol{V}\boldsymbol{V}^t \succeq 0$, then $\boldsymbol{B}^a_{\mathrm{GY}} \succ 0$ if and only if $\boldsymbol{Q} + \boldsymbol{V}\boldsymbol{V}^t \succ 0$.

G. Proof of Proposition 7

Let $\boldsymbol{x} \in \mathbb{R}^N$, $\boldsymbol{\xi} = -(\boldsymbol{B}^a_{\mathrm{GY}})^{-1}\boldsymbol{\nabla}J(\boldsymbol{x})$ and let $A(\boldsymbol{x};\omega) = J(\boldsymbol{x}) - J(\boldsymbol{x} + \theta\boldsymbol{\xi}) + \omega\theta\boldsymbol{\nabla}^t J(\boldsymbol{x})\boldsymbol{\xi}$. In order to show (25), our goal is to establish the stronger condition (27), i.e., $A(\cdot;\omega) \geq 0$

for some $\omega \in (0;1)$. It is easy to obtain $A(\cdot;\omega) = A_0(\cdot;\omega) + \sum_i A_i(\cdot;\omega)$ with

$$A_0(\boldsymbol{x};\omega) = \theta(2-2\omega-\theta)\boldsymbol{\xi}^t\boldsymbol{Q}\boldsymbol{\xi}$$
$$\forall i, \quad A_i(\boldsymbol{x};\omega) = \frac{1-\omega}{a\theta}\Delta_i^2 + \phi(\delta_i) - \phi\left(\delta_i^+\right) - \Delta_i\phi'(\delta_i)$$

where $\delta_i^+ = [\boldsymbol{V}^t(\boldsymbol{x}+\theta\boldsymbol{\xi}) - \boldsymbol{w}]_i$ and $\Delta_i = \delta_i^+ - \delta_i$.

On the one hand, $A_0(\cdot;\omega) \geq 0$ if $\omega \leq \omega_0 = 1 - (\theta/2)$ since $\boldsymbol{Q} \succeq 0$. On the other hand, since ϕ' is $(1/\widehat{a})$-Lipschitz, the *descent lemma* [37, Prop. A.24] yields

$$\forall i, \boldsymbol{x}, \quad A_i(\boldsymbol{x};\omega) \geq \left(\frac{1-\omega}{\theta a} - \frac{1}{2\widehat{a}}\right)\Delta_i^2.$$

The latter bound is nonnegative provided that $\omega \leq \omega_1 = 1 - (\theta a/2\widehat{a})$.

Finally, the condition $(\theta, a) \in \Gamma_{\mathrm{e}}$ ensures that $\omega_0 \in (0;1)$ and $\omega_1 \in (0;1)$, so that $A(\cdot;\omega) \geq 0\ \forall \omega \in (0; \min\{\omega_0, \omega_1\}) \subset (0;1)$.

H. Proof of Proposition 10

1) GY Algorithm: We would like to show that $\sigma_{\mathrm{GY}}(\theta, a) < 1$ for all $(\theta, a) \in \Gamma_{\mathrm{e}}$ (i.e., such that $0 < \theta < 2$ and $\theta a < 2\widehat{a}$). Let m_{GY}^a and M_{GY}^a, respectively, denote the smallest and largest eigenvalue of $(\boldsymbol{B}_{\mathrm{GY}}^a)^{-1}\boldsymbol{H}(\boldsymbol{x}^\infty)$. According to (35)

$$\sigma_{\mathrm{GY}}(\theta, a) = \begin{cases} 1 - \theta m_{\mathrm{GY}}^a, & \text{if } (m_{\mathrm{GY}}^a + M_{\mathrm{GY}}^a)\,\theta \leq 2 \\ \theta M_{\mathrm{GY}}^a - 1, & \text{otherwise.} \end{cases}$$

In the first case, the conclusion is obvious since $m_{\mathrm{GY}}^a > 0$. In the other case, it remains to show that $\theta M_{\mathrm{GY}}^a < 2$. Let us define

$$\boldsymbol{C}_{\mathrm{GY}}^a(\boldsymbol{x}) \triangleq \boldsymbol{B}_{\mathrm{GY}}^a - \boldsymbol{H}(\boldsymbol{x}) = \boldsymbol{V}\mathrm{diag}\left(\frac{1}{a} - \phi''(\delta_i)\right)\boldsymbol{V}^t. \quad (43)$$

Two situations are to be distinguished, whether $a \leq \widehat{a}$ (i.e., $(\theta, a) \in \Gamma_{\mathrm{s}}$) or not.

1) If $a \leq \widehat{a}$, then $\boldsymbol{C}_{\mathrm{GY}}^a \succeq 0$ since $\widehat{a} = \inf_u 1/\phi''(u)$ according to (13). Then

$$M_{\mathrm{GY}}^a = \rho\left((\boldsymbol{B}_{\mathrm{GY}}^a)^{-1}\boldsymbol{H}(\boldsymbol{x}^\infty)\right) = 1 - \rho\left((\boldsymbol{B}_{\mathrm{GY}}^a)^{-1}\boldsymbol{C}_{\mathrm{GY}}^a(\boldsymbol{x}^\infty)\right) \leq 1$$

and the conclusion holds since $\theta < 2$.

2) If $a > \widehat{a}$, we have $M_{\mathrm{GY}}^a \leq \rho((\boldsymbol{B}_{\mathrm{GY}}^a)^{-1}\boldsymbol{B}_{\mathrm{GY}}^{\widehat{a}})$ since $\boldsymbol{C}_{\mathrm{GY}}^{\widehat{a}} = \boldsymbol{B}_{\mathrm{GY}}^{\widehat{a}} - \boldsymbol{H} \succeq 0$. According to (6), we also have

$$\left(\frac{a}{\widehat{a}}\right)\boldsymbol{B}_{\mathrm{GY}}^a - \boldsymbol{B}_{\mathrm{GY}}^{\hat{a}} = 2\left(\frac{a}{\widehat{a}} - 1\right)\boldsymbol{Q} \succeq 0$$

which yields that $M_{\mathrm{GY}}^a \leq a/\widehat{a}$. In conjunction with $\theta a < 2\widehat{a}$, we can conclude that $\theta M_{\mathrm{GY}}^a < 2$.

2) GR Algorithm: A paraphrase of the first part of the proof valid for the GY case suffices to conclude in the GR case. In particular

$$\boldsymbol{C}_{\mathrm{GR}}(\boldsymbol{x}) \triangleq \boldsymbol{B}_{\mathrm{GR}}(\boldsymbol{x}) - \boldsymbol{H}(\boldsymbol{x}) = \boldsymbol{V}\mathrm{diag}\left(\frac{\phi'(\delta_i)}{\delta_i} - \phi''(\delta_i)\right)\boldsymbol{V}^t \quad (44)$$

is positive semidefinite since $\forall u, \phi'(u)/u \geq \phi''(u)$ [44, Sec. II]. Hence

$$M_{\mathrm{GR}} = \rho\left(\bar{\boldsymbol{B}}_{\mathrm{GR}}^{-1}(\boldsymbol{x}^\infty)\bar{\boldsymbol{H}}(\boldsymbol{x}^\infty)\right) = 1 - \rho\left(\boldsymbol{B}_{\mathrm{GR}}^{-1}(\boldsymbol{x}^\infty)\boldsymbol{C}_{\mathrm{GR}}(\boldsymbol{x}^\infty)\right) \leq 1.$$

I. Proof of Proposition 11

In order to establish (37), the keypoint is the positive semidefiniteness of $\boldsymbol{B}_{\mathrm{GY}}^a - \boldsymbol{B}_{\mathrm{GR}}$

$$\boldsymbol{B}_{\mathrm{GY}}^a - \boldsymbol{B}_{\mathrm{GR}} = \boldsymbol{V}\mathrm{diag}\left(\frac{1}{a} - \frac{\phi'(\delta_i)}{\delta_i}\right)\boldsymbol{V}^t \succeq 0 \quad (45)$$

since $\forall u, \phi'(u)/u \leq \phi''(0)$ [44, Sec. II] and $a \leq \widehat{a} = 1/\phi''(0)$. Then $\boldsymbol{B}_{\mathrm{GR}}^{-1} - (\boldsymbol{B}_{\mathrm{GY}}^a)^{-1}$ is also positive semidefinite (which is yielded by left and right multiplication of (45) by $\boldsymbol{B}_{\mathrm{GR}}^{-1}$ and $(\boldsymbol{B}_{\mathrm{GY}}^a)^{-1}$, respectively) as well as $\bar{\boldsymbol{B}}_{\mathrm{GR}}^{-1}\bar{\boldsymbol{H}} - (\bar{\boldsymbol{B}}_{\mathrm{GY}}^a)^{-1}\bar{\boldsymbol{H}}$. In turn, the latter result implies (37b), since

$$M_{\mathrm{GY}}^a = \rho\left((\boldsymbol{B}_{\mathrm{GY}}^a)^{-1}\boldsymbol{H}\right) = \boldsymbol{v}_{\mathrm{GY}}^{\mathrm{t}}\left((\boldsymbol{B}_{\mathrm{GY}}^a)^{-1}\boldsymbol{H}\right)\boldsymbol{v}_{\mathrm{GY}} \leq \boldsymbol{v}_{\mathrm{GY}}^{\mathrm{t}}\left(\boldsymbol{B}_{\mathrm{GR}}^{-1}\boldsymbol{H}\right)\boldsymbol{v}_{\mathrm{GY}} \leq \rho\left(\boldsymbol{B}_{\mathrm{GR}}^{-1}\boldsymbol{H}\right) = M_{\mathrm{GR}}$$

where $\boldsymbol{v}_{\mathrm{GY}}$ is a normalized eigenvector of $(\boldsymbol{B}_{\mathrm{GY}}^a)^{-1}\boldsymbol{H}$ associated to M_{GY}^a. Similarly, $\boldsymbol{H}^{-1}\boldsymbol{B}_{\mathrm{GY}}^a - \boldsymbol{H}^{-1}\boldsymbol{B}_{\mathrm{GR}} \succeq 0$ as a consequence of (45). Then, (37a) holds since

$$m_{\mathrm{GR}}^{-1} = \rho(\boldsymbol{H}^{-1}\boldsymbol{B}_{\mathrm{GR}}) = \boldsymbol{v}_{\mathrm{GR}}^{\mathrm{t}}(\boldsymbol{H}^{-1}\boldsymbol{B}_{\mathrm{GR}})\boldsymbol{v}_{\mathrm{GR}} \leq \boldsymbol{v}_{\mathrm{GR}}^{\mathrm{t}}\left(\boldsymbol{H}^{-1}\boldsymbol{B}_{\mathrm{GY}}^a\right)\boldsymbol{v}_{\mathrm{GR}} \leq \rho\left(\boldsymbol{H}^{-1}\boldsymbol{B}_{\mathrm{GY}}^a\right) = (m_{\mathrm{GY}}^a)^{-1}$$

where $\boldsymbol{v}_{\mathrm{GR}}$ is a normalized eigenvector of $\boldsymbol{H}^{-1}\boldsymbol{B}_{\mathrm{GR}}$ associated to m_{GR}^{-1}.

Finally, the proof of (38) is a paraphrase of the above derivation, after the substitution of $\boldsymbol{B}_{GY}^a$ and $\boldsymbol{B}_{\mathrm{GR}}$ by $\boldsymbol{B}_{\mathrm{GY}}^{a_1}$ for $\boldsymbol{B}_{\mathrm{GY}}^{a_2}$, respectively.

J. Proof of Corollary 1

In this appendix, we borrow some notations and results from Appendices H and I. Moreover, Fig. 3 is useful to visualize the following derivations.

Given (35) and (37a), (40a) stems from the fact that $\mathrm{Ker}(\boldsymbol{V}^t) \neq \{\boldsymbol{0}\}$ implies $M_{\mathrm{GY}}^a = M_{\mathrm{GR}} = 1$. The latter follows from

$$M_{\mathrm{GY}}^a = \left(1 + \lambda_{\min}\left(\mathbf{H}^{-1}\boldsymbol{C}_{\mathrm{GY}}^a\right)\right)^{-1}$$
$$M_{\mathrm{GR}} = \left(1 + \lambda_{\min}(\mathbf{H}^{-1}\boldsymbol{C}_{\mathrm{GR}})\right)^{-1}$$

where $\boldsymbol{C}_{\mathrm{GY}}^a$ and $\boldsymbol{C}_{\mathrm{GR}}$ are defined by (43) and (44), respectively, both matrices being rank-deficient positive semidefinite (hence, $\lambda_{\min} = 0$ in both cases).

In order to prove assertion (40b), let us first recall that $M_{\mathrm{GY}}^a \leq 1$ and $M_{\mathrm{GR}} \leq 1$. Moreover, we also have

$$m_{\mathrm{GY}}^a = \left(1 + \lambda_{\max}\left(\mathbf{H}^{-1}\boldsymbol{C}_{\mathrm{GY}}^a\right)\right)^{-1} \leq 1$$
$$m_{\mathrm{GR}} = \left(1 + \lambda_{\max}(\mathbf{H}^{-1}\boldsymbol{C}_{\mathrm{GR}})\right)^{-1} \leq 1.$$

Given (39), it is now clear that $\theta^{\star}_{\mathrm{GY}}(a) \geq 1$ and $\theta^{\star}_{\mathrm{GR}} \geq 1$. Hence, $\theta < 1$ implies $\sigma(\theta) = 1 - \theta m$ in both cases. Finally, (40b) is a direct consequence of (37a).

Straightforward adaptation of the latter proofs yields (40c) and (40d), respectively.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable comments.

REFERENCES

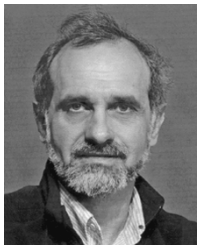
[1] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1970.
[2] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 3, pp. 367–383, Mar. 1992.
[3] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Image Process.*, vol. 4, no. 7, pp. 932–946, Jul. 1995.
[4] M. Çetin and W. Karl, "Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization," *IEEE Trans. Image Process.*, vol. 10, no. 4, pp. 623–631, Apr. 2001.
[5] E. Roullot, A. Herment, I. Bloch, A. de Cesare, M. Nikolova, and E. Mousseaux, "Modeling anisotropic undersampling of magnetic resonance angiographies and reconstruction of a high-resolution isotropic volume using half-quadratic regularization techniques," *Signal Process.*, vol. 84, pp. 743–762, Apr. 2004.
[6] V. Mazet, C. Carteret, D. Brie, J. Idier, and B. Humbert, "Background removal from spectra by designing and minimizing a nonquadratic cost function," *Chemometrics Intell. Lab. Syst.*, vol. 76, pp. 121–133, 2005.
[7] P. J. Huber, *Robust Statistics*. New York: Wiley, 1981.
[8] E. Weiszfeld, "Sur le point pour lequel la somme des distances de n points donnés est minimum," *Tôhoku Math. J.*, vol. 43, pp. 355–386, 1937.
[9] B. R. Hunt, "Bayesian methods in nonlinear digital image restoration," *IEEE Trans. Commun.*, vol. COM-26, no. 3, pp. 219–229, Mar. 1977.
[10] N. Villain, Y. Goussard, J. Idier, and M. Allain, "3D edge-preserving image enhancement for computed tomography," *IEEE Trans. Med. Imag.*, vol. 22, no. 10, pp. 1275–1287, Oct. 2003.
[11] G. T. Herman, *Image Reconstruction From Projections. The Fundamentals of Computerized Tomography*. New York: Academic, 1980.
[12] C. A. Bouman and K. D. Sauer, "A unified approach to statistical tomography using coordinate descent optimization," *IEEE Trans. Image Process.*, vol. 5, no. 3, pp. 480–492, Mar. 1996.
[13] M. Allain, J. Idier, and Y. Goussard, "Regularized approach in 3-D helical computed tomography," in *Proc. IEEE EMBS*, vol. 2, Houston, TX, Oct. 2002, pp. 843–844.
[14] G. Demoment, "Image reconstruction and restoration: overview of common estimation structure and problems," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. ASSP-37, no. 12, pp. 2024–2036, Dec. 1989.
[15] C. A. Bouman and K. D. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. Image Process.*, vol. 2, no. 3, pp. 296–310, Jul. 1993.
[16] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Trans. Image Process.*, vol. 6, no. 2, pp. 298–311, Feb. 1997.
[17] S. Geman and D. McClure, "Statistical methods for tomographic image reconstruction," in *Proc. 46th Session ICI, Bull. ICI*, vol. 52, 1987, pp. 5–21.
[18] R. Yarlagadda, J. B. Bednar, and T. L. Watt, "Fast algorithms for l_p deconvolution," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. ASSP-33, no. 1, pp. 174–182, Feb. 1985.
[19] C. Burrus, J. Barreto, and I. Selesnick, "Iterative reweighted least-squares design of FIR filters," *IEEE Trans. Signal Process.*, vol. 42, no. 11, pp. 2926–2936, Nov. 1994.
[20] J. Nocedal and S. J. Wright, *Numerical Optimization*, ser. Operations Research. New York: Springer Verlag, 1999.
[21] J. Idier, "Convex half-quadratic criteria and interacting auxiliary variables for image restoration," *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 1001–1009, Jul. 2001.
[22] M. Allain, J. Idier, and Y. Goussard, "On global and local convergence of half-quadratic algorithms," in *Proc. IEEE ICIP*, vol. 2, Rochester, MN, Sep. 2002.
[23] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Two deterministic half-quadratic regularization algorithms for computed imaging," in *Proc. IEEE ICIP*, vol. 2, Austin, TX, Nov. 1994, pp. 168–172.
[24] S. Husse, Y. Goussard, and J. Idier, "Extended forms of Geman and Yang algorithm: application to MRI reconstruction," in *Proc. IEEE ICASSP*, vol. III, Montreal, QC, Canada, May 2004, pp. 513–516.
[25] A. H. Delaney and Y. Bresler, "Globally convergent edge-preserving regularized reconstruction: an application to limited-angle tomography," *IEEE Trans. Image Process.*, vol. 7, no. 2, pp. 204–221, Feb. 1998.
[26] K. P. Bube and R. T. Langan, "Hybrid ℓ^1/ℓ^2 minimization with application to tomography," *Geophysics*, vol. 62, pp. 1183–1195, 1997.
[27] T. F. Chan and P. Mulet, "On the convergence of the lagged diffusivity fixed point method in total variation image restoration," *SIAM J. Numer. Anal.*, vol. 36, no. 2, pp. 354–367, 1999.
[28] H. Voss and H. Eckhardt, "Linear convergence of generalized Weiszfeld's method," *Computing*, vol. 25, pp. 243–251, 1980.
[29] K. Lange, D. R. Hunter, and I. Yang, "Optimization transfer using surrogate objective functions (with discussion)," *J. Comput. Graph. Statist.*, vol. 9, no. 1, pp. 1–20, Mar. 2000.
[30] J. Fessler, E. Ficaro, N. Clinthorne, and K. Lange, "Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction," *IEEE Trans. Med. Imag.*, vol. 16, no. 2, pp. 166–175, Apr. 1997.
[31] H. Erdogan and J. Fessler, "Monotonic algorithms for transmission tomography," *IEEE Trans. Med. Imag.*, vol. 18, no. 9, pp. 801–814, Sep. 1999.
[32] J. Zheng, S. S. Saquib, K. Sauer, and C. A. Bouman, "Parallelizable Bayesian tomography algorithms with rapid, guaranteed convergence," *IEEE Trans. Image Process.*, vol. 9, no. 10, pp. 1745–1759, Oct. 2000.
[33] S. Sotthivirat and J. Fessler, "Image recovery using partitioned-separable paraboloidal surrogate coordinate ascent algorithms," *IEEE Trans. Image Process.*, vol. 11, no. 3, pp. 306–317, Mar. 2002.
[34] M. Allain, "Approche pénalisée en tomographie hélicoïdale. Application à la conception d'une prothèse personnalisée du genou," Ph.D. dissertation (in French), Univ. de Paris-Sud, Orsay, France, Dec. 2002. [Online] http://tel.ccsd.cnrs.fr/documents/archives0/00/00/37/56.
[35] J. Ortega and W. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*. New York: Academic, 1970.
[36] T. F. Chan, H. M. Zhou, and R. H. Chan, "Continuation method for total variation denoising problems," in *Proc. SPIE, Conf. Adv. Signal Process. Algorithms*, vol. 2563, Jul. 1995, pp. 314–325.
[37] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA: Athena Scientific, 1999.
[38] M. Nikolova and M. Ng, "Comparison of the main forms of half quadratic regularization," in *Proc. IEEE ICIP*, Rochester, MN, Sep. 2002, pp. 349–352.
[39] ——, "Analysis of Half-Quadratic Minimization Methods for Signal and Image Recovery," *SIAM J. Sci. Comput.*, vol. 27, no. 3, pp. 937–966, 2005.
[40] K. D. Sauer and C. A. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 534–548, Feb. 1993.
[41] S. Husse and Y. Goussard, "Image reconstruction in MRI: regularized approach by Markov random fields," in *Proc. IEEE ICIP*, vol. 2, Barcelona, Spain, 2003, pp. 851–854.
[42] R. Deriche, P. Kornprobst, M. Nikolova, and M. Ng, "Half-quadratic regularization for MRI image restoration," in *Proc. IEEE ICIP*, vol. 6, 2003, pp. 585–588.
[43] M. Nikolova and M. Ng, "Fast image reconstruction algorithms combining half-quadratic regularization and preconditioning," in *Proc. IEEE ICIP*, Thessaloniki, Greece, Oct. 2001, pp. 277–280.
[44] J. Fessler, "Grouped-coordinate descent algorithms for robust edge-preserving image restoration," *Proc. SPIE*, vol. 3170, pp. 184–194, 1997.

Marc Allain received the Ph.D. degree in image processing and biomedical engineering from the University of Paris-Sud, Orsay, France, and École Polytechnique de Montréal, Montréal, QC, Canada, in 2002.

He is currently with the Institut de Recherche en Communications et Cybernétique, Nantes, France. His main subjects of interest are image reconstruction for medial, astronomical, and industrial applications.

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

Since 1991, he has been with the Centre National de la Recherche Scientifique. He is currently with the Institut de Recherche en Communications et Cybernétique, Nantes, France. His major scientific interest are in probabilistic approaches to inverse problems for signal and image processing.



Yves Goussard was born in Paris, France, in 1957. He graduated from the Eécole Nationale Supérieure de Techniques Avancées in 1980 and received the Docteur-Ingénieur and Ph.D. degrees from the Université de Paris-Sud, Orsay, France, in 1983 and 1989, respectively.

From 1983 to 1985, he was a Visiting Scholar at the Electrical Engineering and Computer Science Department, University of California, Berkeley. In 1985, he was appointed as a Chargé de Recherche at CNRS, Gif-sur-Yvette, France, and in 1992, he joined the Biomedical Engineering Institute and the Electrical Engineering Department of the École Polytechnique, Montréal, QC, Canada, where he is now a Full Professor. After some work on nonlinear system identification and modeling, his interests moved toward ill-posed problems in signal and image processing.