The Equivalence of Half-Quadratic Minimization and the Gradient Linearization Iteration

Mila Nikolova and Raymond H. Chan

***Abstract*—A popular way to restore images comprising edges is to minimize a cost function combining a quadratic data-fidelity term and an edge-preserving (possibly nonconvex) regularization term. Mainly because of the latter term, the calculation of the solution is slow and cumbersome. Half-quadratic (HQ) minimization (multiplicative form) was pioneered by Geman and Reynolds (1992) in order to alleviate the computational task in the context of image reconstruction with nonconvex regularization. By promoting the idea of locally homogeneous image models with a continuous-valued line process, they reformulated the optimization problem in terms of an augmented cost function which is quadratic with respect to the image and separable with respect to the line process, hence the name "half quadratic." Since then, a large amount of papers were dedicated to HQ minimization and important results—including edge-preservation along with convex regularization and convergence—have been obtained. In this paper, we show that HQ minimization (multiplicative form) is equivalent to the most simple and basic method where the gradient of the cost function is linearized at each iteration step. In fact, both methods give exactly the same iterations. Furthermore, connections of HQ minimization with other methods, such as the quasi-Newton method and the generalized Weiszfeld's method, are straightforward.**

***Index Terms*—Gradient linearization, half-quadratic (HQ) regularization, inverse problems, optimization, signal and image restoration, variational methods.**

I. INTRODUCTION

LET data $y \in \mathbb{R}^q$ be obtained from an original unknown image or signal x^* via $y = Ax^* + \text{noise}$ where $A \in \mathbb{R}^{q \times p}$ is a linear transform. This simple model addresses various applications, such as denoising, deblurring, image reconstruction in tomography, and other inverse problems [1]–[3]. Since [4]–[6], the sought-after $\hat{x} \in \mathbb{R}^p$ is defined as the minimizer of an objective function $J : \mathbb{R}^p \to \mathbb{R}$ of the form

$$J(x) = \|Ax - y\|^2 + \beta\Phi(x) \tag{1}$$

$$\Phi(x) = \sum_{i=1}^{r} \varphi(\|G_i x\|) \tag{2}$$

Manuscript received September 24, 2005; revised January 22, 2007. This work was supported by the HKRGC under Grants CUHK 400503 and CUHK DAG 2060257. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Jelena Kovačević.

M. Nikolova is with Centre de Mathématiques et de Leurs Applications (CNRS-UMR 8536), ENS de Cachan, 94235 Cachan Cedex, France (e-mail: nikolova@cmla.ens-cachan.fr).

R. H. Chan is with the Department of Mathematics, The Chinese University of Hong Kong, Shatin, Hong Kong (e-mail: rchan@math.cuhk.edu.hk).

Digital Object Identifier 10.1109/TIP.2007.896622

TABLE I
COMMONLY USED FUNCTIONS $\varphi : \mathbb{R}_+ \to \mathbb{R}$ WHERE $\alpha > 0$ IS A PARAMETER AND $\varphi'(0) = 0$

	Convex PFs		Nonconvex PFs
(f1)	$t^\alpha,\ 1<\alpha\le 2$	(f6)	$1 - \exp(-\alpha t^2)$
(f2)	$\sqrt{\alpha + t^2}$	(f7)	$\dfrac{\alpha t^2}{1+\alpha t^2}$
(f3)	$\log(\cosh(\alpha t))$	(f8)	$\log(\alpha t^2 + 1)$
(f4)	$t/\alpha - \log(1 + t/\alpha)$	(f9)	$\min\{\alpha t^2, 1\}$
(f5)	$\begin{cases} \frac{t^2}{2} \text{ if } 0 \le t \le \alpha \\ \alpha\lvert t\rvert - \frac{\alpha^2}{2} \text{ if } t > \alpha \end{cases}$	(f10)	$\begin{cases} \sin(\alpha t^2),\ 0 \le t \le \sqrt{\frac{\pi}{2\alpha}} \\ 1 \text{ if } t > \sqrt{\frac{\pi}{2\alpha}} \end{cases}$

where $G_i : \mathbb{R}^p \to \mathbb{R}^s$, $s \ge 1$ for $i = 1, \ldots, r$, are linear operators, $\|.\|$ is the ℓ_2 norm, $\varphi : \mathbb{R}_+ = \{t | t \ge 0\} \to \mathbb{R}$ is called a *potential function* (PF) and $\beta > 0$ is a parameter. Typically,[1] either $s = 1$ and operators G_i give rise to finite differences between neighboring samples of x, or $s = 2$ and every G_i yields a discrete approximation of the gradient of x at i. One can observe that $s = 1$ is used along with Markov random field models for Bayesian inference, e.g., [1] and [6]–[10], while $s = 2$ along with variational formulations, e.g., [3] and [11]–[14]. The rationale of the image and signal restoration approach (1), (2) has widely been discussed in the literature; let us evoke [1]–[4], [6], and [9], among many others. Let G denote the $rs \times p$ matrix obtained by vertical concatenation of the matrices G_i, that is $G = [G_1^T, \ldots, G_r^T]^T$. A basic condition in order to have regularization is that

$$\text{Ker}A^T A \cap \text{Ker}G^T G = \{0\}. \tag{3}$$

Many different functions φ can be found in the literature [3], [6], [9], [13], [16], [17], some of the most popular ones are given in Table I. Observe that all these PFs satisfy the general condition as follows.

H1 φ is continuous and increasing on $\mathbb{R}_+$ with $\varphi \not\equiv 0$ and $\varphi(0) = 0$.

The effect of the choice of φ on the solution has been discussed, e.g., in [13] and [16]–[19]. One important requirement is that φ allows large differences ("edges") to be recovered in the solution $\hat{x}$. From the references cited, it is well known that edge-preserving functions are nearly affine beyond a neighborhood of the origin. Then φ' and, hence, ∇J involve almost flat regions where standard minimization methods progress very slowly. The computation of a minimizer $\hat{x}$ of J presents a real

[1] Let x be an $m \times n$ image. If $s = 1$, for every $(i, j) \in \{1, \ldots, m\} \times \{1, \ldots, n\}$, there are associated two, or possibly four operators G_i that yield the differences $x_{i,j} - x_{i-1,j}$ and $x_{i,j} - x_{i,j-1}$, and possibly also $x_{i,j} - x_{i-1,j+1}$ and $x_{i,j} - x_{i+1,j-1}$. When $s = 2$, for every (i, j), the operator $G_{i,j}$ is usually defined by $G_{i,j}x = [x_{i,j} - x_{i-1,j}, x_{i,j} - x_{i,j-1}]^T$. The operators relevant to the boundaries are defined according to the boundary conditions.

challenge because in addition to this, the dimension p of x is high and $A^T A$ is usually ill conditioned.

This paper focuses precisely on the computation of a minimizer $\hat{x}$ of J when PF φ is a smooth edge-preserving function. In their inaugural paper [8], Geman and Reynolds have shown that $\hat{x}$ is also given by

$$(\hat{x}, \hat{b}) = \arg\min_{(x,b)} \mathcal{J}(x, b)$$

where $b \in \mathbb{R}^r_+$ is an auxiliary variable (called also a line process) and $\mathcal{J}$ is of the form

$$\mathcal{J}(x, b) = \|Ax - y\|^2 + \beta \sum_{i=1}^{r} \left(\frac{b_i}{2} \|G_i x\|^2 + \psi(b_i) \right) \quad (4)$$

for $\psi : \mathbb{R}_+ \to \mathbb{R}$ a function outlined in Section II. One can notice that $x \to \mathcal{J}(x, b)$ is quadratic and that $b \to \mathcal{J}(x, b)$ can be minimized separately for each b_i. Then a two-step alternate minimization scheme is used: if the $(k-1)$th iterate is $(b^{(k-1)}, x^{(k-1)})$, the next one is defined by

$$b^{(k)} = \arg\min_{b} \mathcal{J}\left(x^{(k-1)}, b\right) \quad (5)$$

$$x^{(k)} = \arg\min_{x} \mathcal{J}\left(x, b^{(k)}\right). \quad (6)$$

The resulting minimization method is called *half quadratic* (HQ). It can also be qualified as *multiplicative*[2] because the line variables b_i in (4) multiply the differences $G_i x$. HQ minimization (4)–(6) has been considered in a large number of papers, e.g., [13], [14], and [20]–[23]. Connections of HQ method with other well-known methods have also been explored, most notably, with generalized Weiszfeld's method [25], with statistical EM algorithms [26], with Lagrangian unconstrained optimization in recursive robust fitting [27], with quasi-Newton minimization [15], [28] and as a residual steepest descent method [28].

The contribution of this paper is to show that the HQ minimization defined by (4) and (5) is equivalent to the very classical *gradient linearization* approach, known also as the *relaxed fixed point iteration*: in order to solve the equation $\nabla J(\hat{x}) = 0$, we write

$$\nabla J(x) = L(x)x - z \quad (7)$$

where z is independent of x. Then at each iteration k, one finds $x^{(k)}$ by solving the linear problem

$$L\left(x^{(k-1)}\right) x^{(k)} = z. \quad (8)$$

A connection between both approaches has been mentioned by Vogel in [29] for the particular case when $\varphi(t) = \sqrt{\alpha + t^2}$. As we show in the following, equivalence holds in general. In turn, convergence results on HQ regularization can now be applied directly to the basically heuristic gradient linearization method in (7) and (8).

[2]Let us mention that another form of HQ minimization, where b is involved additively via terms of the form $(b_i + G_i x)^2$, was initiated in [10] and studied, e.g., in [15], [20], [23], and [24]; it is out of the scope of this paper.

The outline of the paper is as follows. A concise review of the multiplicative form of HQ minimization is given in Section II. Then, in Section III, we formally show that HQ minimization and the simple gradient linearization approach define exactly the same iterates. In Section IV, we give some implications of this equivalence. Conclusions are given in Section V.

II. Half-Quadratic Regularization (Multiplicative Form)

To make the paper self-contained, below we present the derivation of the augmented objective $\mathcal{J} : \mathbb{R}^p \times \mathbb{R}^r$ in (4) by synthesizing the results obtained in many previous papers [8], [13]–[15], [20]–[23]. A fundamental assumption for what follows is that:

H2 $t \to \varphi(\sqrt{t})$ is concave.

Put $\theta(t) = -\varphi(\sqrt{t})$, then θ is convex by H2 and continuous on $\mathbb{R}_+$ by H1. Its convex conjugate [30], [31] is $\theta^*(b) = \sup_{t \geq 0}\{bt - \theta(t)\}$ where $b \in \mathbb{R}$. Define $\psi(b) = \theta^*(-(1/2)b)$ which means that

$$\psi(b) = \sup_{t \geq 0} \left\{ -\frac{1}{2}bt - \theta(t) \right\} = \sup_{t \geq 0} \left\{ -\frac{1}{2}bt^2 + \varphi(t) \right\}. \quad (9)$$

By the Fenchel–Moreau theorem [30], [31], the convex conjugate of θ^* satisfies $(\theta^*)^* = \theta$. Calculating $(\theta^*)^*$ at t^2 yields

$$-\varphi(t) = \theta(t^2) = \sup_{b \leq 0} \left\{ bt^2 - \theta^*(b) \right\} = \sup_{b \geq 0} \left\{ -\frac{1}{2}bt^2 - \psi(b) \right\}.$$

Since $\theta(t) \leq 0$ on $\mathbb{R}_+$, we have $\theta^*(b) = +\infty$ if $b > 0$ and then the supremum of $b \to bt^2 - \theta^*(b)$ in the middle expression above necessarily corresponds to $b \leq 0$. Finally

$$\varphi(t) = \inf_{b \geq 0} \left\{ \frac{1}{2}bt^2 + \psi(b) \right\}. \quad (10)$$

The conclusion is that *under H1 and H2*, (9) *holds if and only if* (10) *holds*. This equivalence was first exhibited in [8] for nonconvex and bounded functions φ. It was established under different conditions on φ in [13]–[15], [23], and [32].

For ψ defined by (9), the function $\mathcal{J}$ in (4) clearly satisfies $J(x) = \inf_{b \in \mathbb{R}^r_+} \mathcal{J}(x, b)$ for every $x \in \mathbb{R}^p$, because of (10). In what follows, it is supposed that:

H3 φ is $\mathcal{C}^1$ on $[0, +\infty)$ with $\varphi'(0) = 0$.

Then the regularization term Φ in (2) is smooth. Furthermore, we adopt the classical assumption for edge preservation which states that φ grows less fast than a quadratic function [3], [13], [33]:

H4 $\lim_{t \to \infty} \varphi(t)/t^2 = 0$.

Next, we focus on the possibility to achieve the supremum in (9) jointly with the infimum in (10). For any $\hat{b} > 0$, define $f_{\hat{b}} : \mathbb{R}_+ \to \mathbb{R}$ by $f_{\hat{b}}(t) = (1/2)\hat{b}t + \theta(t)$, then (9) yields $\psi(\hat{b}) = -\inf_{t \geq 0} f_{\hat{b}}(t)$. Observe that $f_{\hat{b}}$ is convex by H2 with $f_{\hat{b}}(0) = 0$ by H1 and $\lim_{t \to \infty} f_{\hat{b}}(t) = +\infty$ by H4; hence, $f_{\hat{b}}$ has a unique minimum reached at a $\hat{t} \geq 0$. According to (9), $\psi(\hat{b}) = -(1/2)\hat{b}\hat{t}^2 + \varphi(\hat{t})$, then equivalently the infimum in

(10) is reached for $\hat{b}$ since $\varphi(\hat{t}) = (1/2)\hat{b}\hat{t}^2 + \psi(\hat{b})$. Then one finds[3] that $\hat{b} = \sigma(\hat{t})$ for

$$\sigma(t) = \begin{cases} \frac{\varphi'(t)}{t}, & \text{if } t > 0 \\ \varphi''(0^+), & \text{if } t = 0 \end{cases} \tag{11}$$

where the expression for $t = 0$ corresponds to $\lim_{t\searrow 0} \sigma(t)$ and uses the fact that $\varphi'(0) = 0$. Notice that by H1, $\sigma(t) \geq 0$ for all $t \geq 0$ and that by H2, σ is decreasing on $\mathbb{R}_+$. Hence, $\sigma(t) \in [0, \varphi''(0^+)]$. Let us now consider the possibility that $\hat{b} = 0$. Put $\varphi_\infty = \lim_{t\to\infty} \varphi(t)$, then $\varphi_\infty = \sup_{t\geq 0} \varphi(t) \in (0, \infty]$ by H1. If $\varphi(t) < \varphi_\infty$ for all $t \geq 0$ [e.g., φ is strictly increasing as (f1)–(f8) in Table I], for $\hat{b} = 0$ the term between the braces in (10) equals φ_∞, so (10) cannot be realized for any finite $\hat{t} \geq 0$; then the infimum in (10) necessarily corresponds to $\hat{b} > 0$. Otherwise, by H1, there is $T > 0$ such that $\varphi(t) = \varphi_\infty \in (0, \infty)$ for all $t \geq T$, and then $\hat{b} = 0$ in (9) corresponds to $\hat{t} \geq T$ in (10) and $\sigma(\hat{t}) = 0$ in (11) since $\varphi'(t) = 0$ if $t \geq T$ [e.g., (f9) in Table I].

The formula in (11) was initially obtained in [13] under special conditions on φ and was considered in later papers, e.g., [14], [15], and [23]. The derivation presented above is more general than these references. Let us resume its meaning: *if H1–H4 hold, for every* $t \geq 0$, *the infimum in* (10) *is reached for* $\hat{b} = \sigma(t) \in [0, \varphi''(0^+)]$, *where* σ *is defined in* (11), *and we have* $\hat{b} > 0$ *unless there is a* $T > 0$ *such that* $\varphi(t) = const$ *for all* $t \geq T$.

Let us come back to the augmented objective function $\mathcal{J}$ given in (4). The result given in (11) shows that $J(x) = \mathcal{J}(x, \hat{b})$ if $\hat{b}_i = \sigma(\|G_i\hat{x}\|)$ for all $i \in \{1, \ldots, r\}$. For convenience, let us now write $\mathcal{J}$ in (4) as

$$\mathcal{J}(x, b) = \|Ax - y\|^2 + \frac{\beta}{2}(Gx)^T\mathcal{D}(b)Gx + \beta\sum_{i=1}^{r}\psi(b_i)$$

where, for every $b \in \mathbb{R}_+^r$

$$\mathcal{D}(b) = \operatorname{diag}(b_1\mathbb{1}_s, \ldots, b_r\mathbb{1}_s) \quad \text{for} \quad \mathbb{1}_s = [1, \ldots, 1]^T \in \mathbb{R}^s. \tag{12}$$

Define $\mathcal{L} : \mathbb{R}_+^r \to \mathbb{R}^{p\times p}$ by

$$\mathcal{L}(b) = 2A^TA + \beta G^T\mathcal{D}(b)G. \tag{13}$$

Sufficient conditions for the invertibility of $\mathcal{L}(b)$ are that $b \in (0, \infty)^r$ and (3) holds, or that $\operatorname{rank} A = p$. Henceforth, we assume that $\mathcal{L}$ is invertible on the domain of b, namely $\{\sigma(t) : t \in \mathbb{R}_+\}^r$. Combining (11) with the necessary and sufficient condition for a minimum of $x \to \mathcal{J}(x, b)$ shows that the minimum of $\mathcal{J}$ is characterized by

$$\hat{b}_i = \sigma(\|G_i\hat{x}\|), \quad 1 \leq i \leq r \tag{14}$$

$$\mathcal{L}(\hat{b})\hat{x} = 2A^Ty. \tag{15}$$

[3]Notice that $\theta'(t) = -(\varphi'(\sqrt{t})/2\sqrt{t})$ and that $f_{\hat{b}}'(t) = (1/2)\hat{b} + \theta'(t)$ is increasing on $\mathbb{R}_+$ by H2. If $f_{\hat{b}}'(0^+) \geq 0$, i.e., if $\hat{b} \geq \varphi''(0^+)$, $f_{\hat{b}}$ reaches its minimum at $\hat{t} = 0$. Otherwise, its minimum is reached for a $\hat{t} > 0$ such that $f_{\hat{b}}'(\hat{t}) = 0$, i.e., $\hat{b} = -2\theta'(\hat{t})$. In this case, $t \to -(1/2)\hat{b}t^2 + \varphi(t)$ in the last expression of (9) reaches its supremum for a $\hat{t}$ that satisfies $\hat{b} = -2\theta'(\hat{t}^2)$, hence, (11).

The alternate minimization scheme mentioned in (5) and (6) is aimed at solving (14) and (15). Given $x^{(k-1)}$, the first step of iteration k, as defined in (5), has an explicit form

$$b_i^{(k)} = \sigma\left(\left\|G_ix^{(k-1)}\right\|\right), \quad i = 1, \ldots, r. \tag{16}$$

Its second step, defined by (6), amounts to

$$x^{(k)} = \left(\mathcal{L}\left(b^{(k)}\right)\right)^{-1} 2A^Ty. \tag{17}$$

The convergence of the resulting iterative scheme (16), (17) was considered under different assumptions on φ in [13], [21], and [23], while its speed was analyzed in [19] and [28].

It is worth it to recall the interpretation of the auxiliary array b pioneered in [8]. Since σ in (11) is decreasing on $\mathbb{R}_+$, (14) shows that a large $\hat{b}_i$ is attached to a small $\|G_i\hat{x}\|$, and vice versa. In this way, each $\hat{b}_i$ expresses the weight of the quadratic smoothness constraint imposed to the relevant $G_i\hat{x}$. Indeed, the auxiliary variable b has been interpreted by Geman and Reynolds [8] as a continuous-valued, noninteracting line process. This interpretation had a significant impact on the research on edge-preserving regularization during the last 15 years. From the PDE point of view, the auxiliary variables b_i, $1 \leq i \leq r$, can also be seen as the conduction coefficients in an anisotropic diffusion equation; see [12] and [34] for details.

III. Linearization of the Gradient

A natural alternative to simplify the search for a solution to $\nabla J(x) = 0$ is to linearize this gradient at each step, as sketched in (7) and (8). Below we develop these expressions for an objective function J of the form (1) and (2). The gradient of J reads

$$\nabla J(x) = 2A^TAx + \beta\sum_i G_i^T \frac{\varphi'(\|G_ix\|)}{\|G_ix\|}G_ix - 2A^Ty. \tag{18}$$

Since $x \to G_ix/\|G_ix\|$ is uniformly bounded on $\mathbb{R}^p$ and that $\varphi'(0) = 0$, we can write

$$\frac{\varphi'(\|G_ix\|)}{\|G_ix\|}G_ix = \begin{cases} 0, & \text{if } G_ix = 0 \\ \sigma(\|G_ix\|)G_ix, & \text{else} \end{cases} \tag{19}$$

where σ is the function in (11). Notice that the two cases in (19) are necessary because if $\sigma(0) = +\infty$, which occurs for instance for $\varphi(t) = t^\alpha$, $1 < \alpha < 2$, we can replace this infinite value by any positive real number without changing the expression in (18). Using (19), ∇J in (18) reads

$$\begin{aligned} \nabla J(x) &= 2A^TAx + \beta G^T\mathcal{D}\left([\sigma(\|G_ix\|)]_{i=1}^r\right)Gx - 2A^Ty \\ &= \left(2A^TA + \beta G^T\mathcal{D}\left([\sigma(\|G_ix\|)]_{i=1}^r\right)G\right)x - 2A^Ty \end{aligned}$$

where $b \to \mathcal{D}(b)$ is the application defined by (12). Then $\nabla J(x)$ is easily put into the form (7), namely $\nabla J(x) = L(x)x - z$, for

$$\begin{aligned} L(x) &= \mathcal{L}\left([\sigma(\|G_ix\|)]_{i=1}^r\right) \\ z &= 2A^Ty \end{aligned} \tag{20}$$

where $\mathcal{L}$ is the matrix-valued function defined in (13). The matrix $L(x)$ is invertible if $\mathcal{L}(b)$ is invertible on the domain of b, a question which was addressed already in Section II. Applying (8) actually yields

$$x^{(k)} = \left(\mathcal{L}\left(\left[\sigma\left(\left\|G_i x^{(k-1)}\right\|\right)\right]_{i=1}^{r}\right)\right)^{-1} 2A^T y. \tag{21}$$

This amounts to inserting the expression for the $b_i^{(k)}$ in (16) into the expression for $x^{(k)}$ in (17) in the HQ minimization methods. It follows that *the HQ minimization and the gradient linearization approach construct exactly the same sequence of iterates* $x^{(k)}$.

The gradient linearization method was used by Vogel and Oman in [29], [35] to minimize an approximate total variation regularization corresponding to $\varphi(t) = \sqrt{\alpha + t^2} \approx |t|$ and it was called the "lagged diffusivity fixed point iteration." In [29], the authors mention that it amounts to apply the multiplicative form of HQ minimization to this φ. As we have shown above, the equivalence holds for general functions φ applied either to the ℓ_1 or to the ℓ_2 norm of $G_i x$.

IV. Implications

The equivalence established in the last section shows in particular that the gradient linearization iteration is convergent for all objective functions for which convergence of HQ regularization has been proven [13], [21], [23]. We can also use the equivalence to connect the HQ minimization to other well-known methods. As an example, it has been pointed out in [15] and [28] that for convex PFs φ, the HQ minimization corresponds to a quasi-Newton minimization. The same holds now in our more general context since any gradient linearization method can always be viewed as a form of quasi-Newton method. Indeed, starting with (21), we derive

$$\begin{aligned} x^{(k)} &= \left(L\left(x^{(k-1)}\right)\right)^{-1} z \\ &= \left(L\left(x^{(k-1)}\right)\right)^{-1} \left(L\left(x^{(k-1)}\right) x^{(k-1)} \right. \\ &\qquad \left. - L\left(x^{(k-1)}\right) x^{(k-1)} + z\right) \\ &= x^{(k-1)} - \left(L\left(x^{(k-1)}\right)\right)^{-1} \nabla J\left(x^{(k-1)}\right) \end{aligned}$$

where, in the last equality, we use (7). This shows that HQ regularization (16), (17), or *equivalently* the gradient linearization iteration (21), performs a quasi-Newton iteration where the Hessian of J at $x^{(k-1)}$ is approximated by $L(x^{(k-1)})$ as defined in (20).

Following [25] and [28], we note further that if one defines

$$\mathcal{G}(y,x) = J(x) + (y-x)^T \nabla J(x) + \frac{1}{2}(y-x)^T L(x)(y-x)$$

then $\mathcal{G}(y,x)$ provides a majorizing quadratic approximation for $J(x)$. In the generalized Weiszfeld's algorithm, the iterates are given by $x^{(k)} = \arg\min_x \mathcal{G}(x, x^{(k-1)})$ which amounts to the gradient linearization method (21) and equivalently to the HQ minimization (16), (17) whenever L is given by (20).

V. Conclusion

HQ minimization has been studied and used by numerous authors. In this paper, we have shown that it is equivalent to the very basic approach where at each iteration a linear approximation of the gradient is used. We also demonstrate that it is a form of quasi-Newton method and is closely related to the gradient descent method and the generalized Weiszfeld's algorithm. This paper nicely shows a case where research follows a tortuous way to find a simple result.

References

[1] 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.

[2] A. Tarantola, *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*. Amsterdam, The Netherlands: Elsevier, 1987.

[3] G. Aubert and P. Kornprobst, *Mathematical Problems in Images Processing*. Berlin, Germany: Springer-Verlag, 2002.

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

[5] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 6, pp. 721–741, Nov. 1984.

[6] J. E. Besag, "Digital image processing: Towards Bayesian image analysis," *J. Appl. Statist.*, vol. 16, no. 3, pp. 395–407, 1989.

[7] D. Geman, "École d'Été de Probabilités de Saint-Flour XVIII—1988," in *Random Fields and Inverse Problems in Imaging*. New York: Springer-Verlag, 1990, vol. 1427, pp. 117–193, Lecture Notes in Mathematics edition.

[8] D. Geman and G. Reynolds, "Constrained restoration and recovery of discontinuities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 3, pp. 367–383, Mar. 1992.

[9] C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. Image Process.*, vol. 2, no. 3, pp. 296–310, Jul. 1993.

[10] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Image Process.*, vol. IP-4, no. 7, pp. 932–946, Jul. 1995.

[11] A. Blake and A. Zisserman, *Visual Reconstruction*. Cambridge, MA: MIT Press, 1987.

[12] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-12, pp. 629–639, Jul. 1990.

[13] 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.

[14] S. Teboul, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Variational approach for edge-preserving regularization using coupled PDE's," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 387–397, Mar. 1998.

[15] M. Nikolova and M. Ng, "Analysis of half-quadratic minimization methods for signal and image recovery," *SIAM J. Sci. Comput.*, vol. 27, no. 3, pp. 937–966, 2005.

[16] M. Black and A. Rangarajan, "On the unification of line processes, outlier rejection, and robust statistics with applications to early vision," *Int. J. Comput. Vis.*, vol. 19, no. 1, pp. 57–91, 1996.

[17] S. Li, *Markov Random Field Modeling in Computer Vision*, 1st ed. New York: Springer-Verlag, 1995.

[18] M. Nikolova, "Weakly constrained minimization. Application to the estimation of images and signals involving constant regions," *J. Math. Imag. Vis.*, vol. 21, no. 2, pp. 155–175, Sep. 2004.

[19] M. Nikolova, "Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares," *SIAM J. Multiscale Model. Simul.*, vol. 4, no. 3, pp. 960–991, 2005.

[20] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Two deterministic half-quadratic regularization algorithms for computed imaging," in *Proc. IEEE Int. Conf. Image Processing*, 1994, vol. 2, pp. 168–172.

[21] 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.

[22] P. Kornprobst, R. Deriche, and G. Aubert, "Image sequence analysis via partial differential equations," *J. Math. Imag. Vis.*, vol. 11, no. 1, pp. 5–26, Oct. 1999.

[23] J. Idier, "Convex half-quadratic criteria and auxiliary interacting variables for image restoration," *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 1001–1009, Jul. 2001.
[24] G. Aubert and L. Vese, "A variational method in image recovery," *SIAM J. Numer. Anal.*, vol. 34, no. 5, pp. 1948–1979, 1997.
[25] T. 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.
[26] F. Champagnat and J. Idier, "A connection between half-quadratic criteria and em algorithms," *IEEE Signal Process. Lett.*, vol. 11, no. 9, pp. 709–712, Sep. 2004.
[27] J.-P. Tarel, S.-S. Ieng, and P. Charbonnier, "Using robust estimation algorithms for tracking explicit curves," in *Proc. ECCV*, 2002, vol. 2350, pp. 492–507, Lecture Notes Comput. Sci..
[28] M. Allain, J. Idier, and Y. Goussard, "On global and local convergence of half-quadratic algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 5, pp. 1130–1142, May 2006.
[29] C. R. Vogel and M. E. Oman, "Fast, robust total variation-based reconstruction of noisy, blurred images," *IEEE Trans. Image Process.*, vol. 7, no. 6, pp. 813–824, Jun. 1998.
[30] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1970.
[31] H. Brezis, "Collection mathématiques appliquées pour la maîtrise," in *Analyse fonctionnelle*. Paris, France: Masson, 1992.
[32] P. Ciuciu and J. Idier, "A half-quadratic block-coordinate descent method for spectral estimation," *Signal Process.*, no. 82, pp. 941–959, 2002.
[33] S. Z. Li, "On discontinuity-adaptive smoothness priors in computer vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-17, no. 6, pp. 576–586, Jun. 1995.
[34] L. Blanc-Féraud, P. Charbonnier, G. Aubert, and M. Barlaud, "Nonlinear image processing: Modelling and fast algorithm for regularisation with edge detection," in *Proc. IEEE Int. Conf. Image Processing*, 1995, vol. 2, pp. 474–477.
[35] C. R. Vogel and M. E. Oman, "Iterative method for total variation denoising," *SIAM J. Sci. Comput.*, vol. 17, no. 1, pp. 227–238, 1996.

Mila Nikolova received the Ph.D. degree in signal processing from the Université de Paris-Sud, Paris, France, in 1995.

Currently, she is a Senior Research Fellow with the National Center for Scientific Research (CNRS), France, and performs her research at the Centre de Mathématiques et de Leurs Applications (CMLA), ENS de Cachan, France. Her research interests are in Image and signal reconstruction, regularization and variational methods, and scientific computing.

Raymond H. Chan was born in 1958 in Hong Kong. He received the B.Sc. degree in mathematics from the Chinese University of Hong Kong and the M.Sc. and Ph.D. degrees in applied mathematics from New York University.

He is currently a Professor in the Department of Mathematics, The Chinese University of Hong Kong. His research interests include numerical linear algebra and image processing problems.