

## Analysis of Half-Quadratic Minimization Methods for Signal and Image Recovery

Nikolova, Mila; Ng, Michael K.

*Published in:*  
SIAM Journal on Scientific Computing

*DOI:*  
[10.1137/030600862](https://doi.org/10.1137/030600862)

Published: 01/05/2005

*Document Version:*  
Publisher's PDF, also known as Version of record

[Link to publication](#)

*Citation for published version (APA):*  
Nikolova, M., & Ng, M. K. (2005). Analysis of Half-Quadratic Minimization Methods for Signal and Image Recovery. *SIAM Journal on Scientific Computing*, 27(3), 937-966. <https://doi.org/10.1137/030600862>

### General rights

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent publication URLs

## ANALYSIS OF HALF-QUADRATIC MINIMIZATION METHODS FOR SIGNAL AND IMAGE RECOVERY\*

MILA NIKOLOVA<sup>†</sup> AND MICHAEL K. NG<sup>‡</sup>

**Abstract.** We address the minimization of regularized convex cost functions which are customarily used for edge-preserving restoration and reconstruction of signals and images. In order to accelerate computation, the *multiplicative* and the *additive* half-quadratic reformulation of the original cost-function have been pioneered in Geman and Reynolds [*IEEE Trans. Pattern Anal. Machine Intelligence*, 14 (1992), pp. 367–383] and Geman and Yang [*IEEE Trans. Image Process.*, 4 (1995), pp. 932–946]. The alternate minimization of the resultant (augmented) cost-functions has a simple explicit form. The goal of this paper is to provide a systematic analysis of the convergence rate achieved by these methods. For the multiplicative and additive half-quadratic regularizations, we determine their upper bounds for their root-convergence factors. The bound for the multiplicative form is seen to be always smaller than the bound for the additive form. Experiments show that the number of iterations required for convergence for the multiplicative form is always less than that for the additive form. However, the computational cost of each iteration is much higher for the multiplicative form than for the additive form. The global assessment is that minimization using the additive form of half-quadratic regularization is faster than using the multiplicative form. When the additive form is applicable, it is hence recommended. Extensive experiments demonstrate that in our MATLAB implementation, both methods are substantially faster (in terms of computational times) than the standard MATLAB OPTIMIZATION TOOLBOX routines used in our comparison study.

**Key words.** convergence analysis, half-quadratic regularization, inverse problems, maximum a posteriori estimation, optimization, preconditioning, signal and image restoration, variational methods

**AMS subject classifications.** 35A15, 49N45, 65K10, 65F22, 65Y20, 94A08, 94A12

**DOI.** 10.1137/030600862

**1. Introduction.** This work addresses a wide class of problems where a sought vector  $\hat{x} \in \mathbb{R}^p$  (e.g., an image or a signal) is estimated based on degraded data  $y \in \mathbb{R}^q$  by minimizing a cost function  $J : \mathbb{R}^p \rightarrow \mathbb{R}$  which combines a quadratic data-fidelity term and a regularization term  $\Phi$ , weighted by a parameter  $\beta > 0$ :

$$(1.1) \quad \hat{x} = \min_{x \in \mathbb{R}^p} J(x),$$

$$(1.2) \quad J(x) = \|Ax - y\|^2 + \beta\Phi(x).$$

In a statistical framework, this data-fidelity term assumes that data  $y$  are obtained from the original  $x$  by a linear transform, modeled by  $A \in \mathbb{R}^{q \times p}$ , and that they are contaminated by white Gaussian noise. Such data-fidelity terms are popular in denoising, in deblurring, and in numerous inverse problems such as seismic imaging,

\*Received by the editors September 5, 2003; accepted for publication (in revised form) March 29, 2005; published electronically December 8, 2005. The work was supported by a grant from the PROCORE - France/Hong Kong Joint Research Scheme sponsored by the Research Grants Council of Hong Kong; the Consulate General of France in Hong Kong F-HK 30/04T; RGC grants HKU 7130/02P, 7046/03P, 7035/04P; and HKU CRCG grants 10205775 and 10204437.

<http://www.siam.org/journals/sisc/27-3/60086.html>

<sup>†</sup>Centre de Mathématiques et de Leurs Applications (CNRS-UMR 8536), ENS de Cachan, 61 avenue du Président Wilson, 94235 Cachan Cedex, France (nikolova@cmla.ens-cachan.fr).

<sup>‡</sup>Department of Mathematics, The University of Hong Kong, Pokfulam Road, Hong Kong and Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong (mng@maths.hku.hk).

nondestructive evaluation, and X-ray tomography [1, 12, 28, 29, 31]. We focus on regularization terms  $\Phi$  of the form

$$(1.3) \quad \Phi(x) = \sum_{i=1}^r \phi(g_i^T x),$$

where  $g_i^T : \mathbb{R}^p \rightarrow \mathbb{R}$ , for  $i = 1, \dots, r$ , are linear operators. Typically,  $\{g_i^T x\}$  are first- or second-order differences between neighboring samples in  $x$ . For example, if  $x$  is a one-dimensional signal, usually  $g_i^T x = x_i - x_{i+1}$ , for  $i = 1, \dots, p-1$ . Let  $G$  denote the  $r \times p$  matrix whose  $i$ th row is  $g_i^T$ , for  $i = 1, \dots, r$ . Some basic requirements are

$$(1.4) \quad \phi \not\equiv 0 \quad \text{and} \quad \ker(A^T A) \cap \ker(G^T G) = \{0\}.$$

We focus on convex, *edge-preserving* potential functions  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  in (1.3), because they give rise to image and signal estimates of high quality, involving edges and homogeneous regions. Examples of such functions are [4, 5, 8, 17, 22]

$$(1.5) \quad \phi(t) = |t|^\alpha, \quad 1 < \alpha \leq 2,$$

$$(1.6) \quad \phi(t) = \sqrt{\alpha + t^2},$$

$$(1.7) \quad \phi(t) = \log(\cosh(\alpha t)),$$

$$(1.8) \quad \phi(t) = |t|/\alpha - \log(1 + |t|/\alpha),$$

$$(1.9) \quad \phi(t) = \begin{cases} t^2/2 & \text{if } |t| \leq \alpha, \\ \alpha|t| - \alpha^2/2 & \text{if } |t| > \alpha, \end{cases}$$

where  $\alpha > 0$  is a parameter. We will systematically consider that  $\phi$  is convex, even, and  $\mathcal{C}^2$  and that

$$(1.10) \quad A^T A \text{ is invertible} \quad \text{and/or} \quad \phi''(t) > 0 \quad \forall t \in \mathbb{R}.$$

REMARK 1. *It is easy to see that the assumptions in (1.10) and (1.4) guarantee that for every  $y \in \mathbb{R}^p$ , the function  $J$  has a unique minimum and that the latter is strict.*

However, the minimizers  $\hat{x}$  of cost-functions  $J$  involving edge-preserving regularization are nonlinear with respect to  $y$  and their computation is costly, especially when  $A$  has many nonzero entries and/or  $A$  is ill-conditioned. In order to cope with the computation, *half-quadratic* reformulation of  $J$  was pioneered, in two different ways, in [13, 14]. The idea is to construct an *augmented cost function*  $\mathcal{J} : \mathbb{R}^p \times \mathbb{R}^r \rightarrow \mathbb{R}$  which involves an auxiliary variable  $b \in \mathbb{R}^r$ ,

$$(1.11) \quad \mathcal{J}(x, b) = \|Ax - y\|^2 + \beta \sum_{i=1}^r (Q(g_i^T x, b_i) + \psi(b_i)),$$

where  $Q(\cdot, s) : \mathbb{R} \rightarrow \mathbb{R}$  is quadratic for any  $s \in \mathbb{R}$  and  $\psi : \mathbb{R} \rightarrow \mathbb{R}$  satisfies

$$(1.12) \quad \phi(t) = \min_{s \in \mathbb{R}} \{Q(t, s) + \psi(s)\} \quad \forall t \in \mathbb{R}.$$

Such a *dual* potential function  $\psi$  is determined using the theory of convex conjugacy [18, 27]. Condition (1.12) ensures that

$$(1.13) \quad J(x) = \min_{b \in \mathbb{R}^r} \mathcal{J}(x, b) \quad \forall x \in \mathbb{R}^p.$$

The regularization term involved in  $\mathcal{J}$  is half-quadratic, and hence the name of the method. The minimizer  $(\hat{x}, \hat{b})$  of  $\mathcal{J}$  is calculated using *alternate minimization*. Let the solution at iteration  $(k-1)$  read  $(x^{(k-1)}, b^{(k-1)})$ . At iteration  $k$  we calculate

$$(1.14) \quad b^{(k)} \text{ such that } \mathcal{J}(x^{(k-1)}, b^{(k)}) \leq \mathcal{J}(x^{(k-1)}, b) \quad \forall b \in \mathbb{R}^r,$$

$$(1.15) \quad x^{(k)} \text{ such that } \mathcal{J}(x^{(k)}, b^{(k)}) \leq \mathcal{J}(x, b^{(k)}) \quad \forall x \in \mathbb{R}^p.$$

This amounts to finding  $b^{(k)}$  and  $x^{(k)}$  according to

$$(1.16) \quad \begin{aligned} b^{(k)} &= \left[ \sigma(g_i^T x^{(k-1)}) \right]_{i=1}^r, \\ x^{(k)} &= \chi(b^{(k)}), \end{aligned}$$

where  $\sigma : \mathbb{R} \rightarrow \mathbb{R}$  and  $\chi : \mathbb{R}^r \rightarrow \mathbb{R}^p$  are minimizer functions which are defined in sections 2 and 3. For every  $t \in \mathbb{R}$ , the scalar  $\sigma(t)$  is such that

$$(1.17) \quad Q(t, \sigma(t)) + \psi(\sigma(t)) \leq Q(t, s) + \psi(s) \quad \forall s \in \mathbb{R}.$$

For every  $b \in \mathbb{R}^r$ , the vector  $\chi(b) \in \mathbb{R}^p$  is such that

$$(1.18) \quad \|A\chi(b) - y\|^2 + \beta \sum_{i=1}^r Q(g_i^T \chi(b), b_i) \leq \|Ax - y\|^2 + \beta \sum_{i=1}^r Q(g_i^T x, b_i) \quad \forall x \in \mathbb{R}^p.$$

The key points are that each component  $b_i^{(k)}$  is calculated using a scalar function  $\sigma$  and that  $x^{(k)}$  is calculated using an affine function of  $y$ , namely  $\chi$ . Moreover, both functions  $\sigma$  and  $\chi$  have an explicit form. Thus  $b^{(k)}$  and  $x^{(k)}$  are easy to compute and allow  $J$  to be decreased.

REMARK 2. From (1.14)–(1.15), for all  $k \in \mathbb{N}$ , we have

$$(1.19) \quad \mathcal{J}(x^{(k+1)}, b^{(k+1)}) \leq \mathcal{J}(x^{(k)}, b^{(k+1)}) \leq \mathcal{J}(x^{(k)}, b^{(k)}).$$

Since  $\mathcal{J}$  is bounded below, the sequence

$$\{\dots, \mathcal{J}(x^{(k)}, b^{(k)}), \mathcal{J}(x^{(k)}, b^{(k+1)}), \mathcal{J}(x^{(k+1)}, b^{(k+1)}), \dots\}$$

converges as  $k \rightarrow \infty$ . In particular,  $J(x^{(k+1)}) \leq J(x^{(k)})$ , for all  $k$ , and the sequence  $\{J(x^{(k)})\}$  is convergent.

By extending the notion of line-variable [15], Geman and Reynolds [13] first considered quadratic terms of the *multiplicative form*,

$$(1.20) \quad Q(t, s) = \frac{1}{2}t^2s \quad \text{for } t \in \mathbb{R}, \quad s \in \mathbb{R}_+.$$

Later, Geman and Yang proposed in [14] the *additive form* for  $Q$ ,

$$(1.21) \quad Q(t, s) = (t - s)^2 \quad \text{for } t \in \mathbb{R}, \quad s \in \mathbb{R}.$$

In [13, 14], half-quadratic regularization was used to simplify simulated annealing minimization in cases when  $\phi$  is nonconvex and  $A$  has many nonzero entries. These fruitful ideas have been pursued and deepened by many authors. The multiplicative form (1.20) was considered in [7, 8, 11, 19, 21, 30]. The additive form (1.21) seems to have suggested a narrower interest. It was considered in [3, 7, 10, 19]. In [10],

the auxiliary variable is introduced in a nonconvex data-fidelity term. Extensions of the multiplicative form to interacting line variables  $b$  were proposed in [30, 19]. The convergence of the alternate minimization (1.14)–(1.15) of  $\mathcal{J}$  towards the sought minimizer  $\hat{x}$  of the original  $J$  for the multiplicative form is considered in [8, 11, 19] and for the additive form in [3] in a continuous setting.

In [3, 7, 10, 13, 14], the numerical results have shown that minimization using half-quadratic (HQ) regularization can speed up computation compared with the steepest descent method. However, the convergence rate of HQ minimization methods has never been analyzed in a systematic way. The effectiveness of the two HQ formulations (1.20) and (1.21) has never been compared theoretically. The main objective of this paper is to characterize the speed of convergence and estimate the bound on the convergence rate relevant to each form of HQ regularization and to compare it with standard minimization methods. We show that minimization using HQ regularization amounts to finding the fixed point  $\hat{x}$  of a mapping  $F : \mathbb{R}^p \rightarrow \mathbb{R}^p$  of the form

$$F(x) = x - (\mathcal{H}(x))^{-1}DJ(x),$$

where  $\mathcal{H}(x) \in \mathbb{R}^{p \times p}$  performs a pertinent correction of the steepest descent direction  $-DJ(x)$ . We analyze the contraction properties of the mapping  $F$  corresponding to the multiplicative and the additive forms of HQ regularization. In order to compare the iterations corresponding to these two forms, we focus on the root-convergence factor [9, 26], namely

$$(1.22) \quad C(F, \hat{x}) = \sup \left\{ \limsup_{k \rightarrow \infty} \|x^{(k)} - \hat{x}\|^{\frac{1}{k}} : x^{(0)} \in \mathbb{R}^p \right\}.$$

Let us emphasize that  $C(F, \hat{x})$  characterizes the convergence speed for any norm used to measure the residual  $\|x^{(k)} - \hat{x}\|$ ; see, e.g., [26]. Using the linear convergence theorem [26, p. 301], there are conditions ensuring that

$$C(F, \hat{x}) = \text{Spectral Radius}(DF(\hat{x})).$$

For each form we determine an upper bound of the spectral radius of  $DF$  at  $\hat{x}$  which is shown to be strictly smaller than 1. The bound for the multiplicative form is revealed to be smaller than the bound for the additive form. This result suggests that the multiplicative form (1.20) needs fewer iterations to find the minimum than the additive form (1.21). However, the cost per iteration is much smaller for the additive form than for the multiplicative form. By combining these two factors, we demonstrate that minimization using the additive form is faster than using the multiplicative form. When the additive form is applicable, it is more attractive. We also propose a fast method to effectively recover signals and images using regularization with the Huber function  $\phi$ —cf. (1.9)—and a nonsmooth augmented cost-function  $\mathcal{J}$ . Furthermore, we discuss different ways to improve the computational cost of HQ minimization. Extensive experiments based on the MATLAB computing environment demonstrate that the additive form achieves minimization faster than the multiplicative form, and that both forms in our MATLAB implementation substantially outperform in terms of computational times the routines in the MATLAB OPTIMIZATION TOOLBOX. (These results must be qualified by the observation that MATLAB run times for the same operations can vary by orders of magnitude depending on the particular coding used.)

**Organization of the paper.** Section 2 is dedicated to the speed of convergence of the multiplicative form. In section 3, we study the speed of convergence relevant

to the additive form. The convergence properties of the two forms, multiplicative and additive, are compared in section 4. Section 5 summarizes the experimental results. Concluding remarks are presented in section 6.

**Notations.** When necessary, a vector  $x \in \mathbb{R}^p$  is also written as  $[x_i]_{i=1}^p$ . Given  $x \in \mathbb{R}^p$ ,  $\text{diag}(x) \equiv \text{diag}([x_i]_{i=1}^p)$  is the diagonal  $p \times p$  matrix whose diagonal elements are the entries of  $x$ . For a function  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f'(t^-)$  and  $f'(t^+)$  are its left-side and right-side derivatives at  $t$ , respectively. By  $\lim_{t \searrow 0} f(t)$  we mean the limit of  $f$  when  $t$  converges to 0 by positive values. The symbol  $\mathbb{R}_+$  addresses the nonnegative reals and  $\mathbb{R}_+^*$  the positive reals. If  $f : \mathbb{R}^p \times \mathbb{R}^q \rightarrow \mathbb{R}$  depends on two variables,  $D_j^k f$  is its  $k$ th differential with respect to the  $j$ th variable. The Euclidean norm of a vector  $u \in \mathbb{R}^p$  is denoted  $\|u\|$ . If  $H$  is a  $p \times p$  real matrix, its largest-in-magnitude eigenvalue is denoted by  $\rho(H)$ . Furthermore,  $\|H\| = \sqrt{\rho(H^T H)}$ , which is in fact the largest singular value of  $H$ . We write  $H \succ 0$  if  $H$  is symmetric and positive definite. The symbol  $I$  stands for identity matrix and  $\mathbb{1}$  for a vector whose entries are equal to 1. Given  $x \in \mathbb{R}^p$  and  $\delta > 0$ , we denote  $B(x, \delta) = \{u \in \mathbb{R}^p : \|x - u\| \leq \delta\}$ . Furthermore,  $S = \{x \in \mathbb{R}^p : \|x\| = 1\}$ . We also systematically denote

$$\nu^2 = \min \{ \mu \in \mathbb{R}_+ : \mu \text{ is an eigenvalue of } A^T A \},$$

where  $A$  is the matrix involved in (1.2).

**2. Multiplicative form of half-quadratic regularization.**

**2.1. Preliminary facts.** Based on previous papers, we suppose that

- (a)  $t \rightarrow \phi(t)$  is convex on  $\mathbb{R}$ ,
- (b)  $t \rightarrow \phi(\sqrt{t})$  is concave on  $\mathbb{R}_+$ ,
- (c)  $\phi(t) = \phi(-t) \forall t \in \mathbb{R}$ ,
- (d)  $\phi$  is  $\mathcal{C}^1$  on  $\mathbb{R}$ ,
- (e)  $\phi''(0^+) > 0$ ,
- (f)  $\lim_{t \rightarrow \infty} \phi(t)/t^2 = 0$ .

The dual function  $\psi$  is convex and reads

$$\psi(s) = \sup_{t \in \mathbb{R}} \left\{ -\frac{1}{2}st^2 + \phi(t) \right\}.$$

As required in (1.12), we have reciprocally (cf. section 7.1 in the appendix)

$$\phi(t) = \inf_{s \in \mathbb{R}} \left\{ \frac{1}{2}st^2 + \psi(s) \right\}.$$

The relation between  $\phi$  and  $\psi$  under different assumptions on  $\phi$  were analyzed in [8, 13, 19]. Using (1.11) and (1.20), the augmented cost-function  $\mathcal{J}$  reads

$$\begin{aligned} \mathcal{J}(x, b) &= \|Ax - y\|^2 + \beta \sum_{i=1}^r \left( \frac{b_i}{2} (g_i^T x)^2 + \psi(b_i) \right) \\ &= \|Ax - y\|^2 + \frac{\beta}{2} (Gx)^T \text{diag}(b) Gx + \beta \sum_{i=1}^r \psi(b_i). \end{aligned}$$

The minimizer function  $\sigma$ , defined by (1.17), admits an explicit form which was first determined in [8]:

$$(2.5) \quad \sigma(t) = \begin{cases} \phi''(0^+) & \text{if } t = 0, \\ \frac{\phi'(t)}{t} & \text{if } t \neq 0. \end{cases}$$

For its proof under (2.1), see section 7.1 in the appendix. The expression of  $\sigma$  for several functions  $\phi$  is given in Table 4.1. It is useful to notice its main properties.

REMARK 3. *Suppose that  $\phi$  satisfies all conditions given in (2.1).*

1. *Then the function  $t \rightarrow \sigma(t)$  is decreasing on  $\mathbb{R}_+$ . Indeed, (b) shows that  $t \rightarrow \frac{d}{dt}\phi(\sqrt{t}) = \phi'(\sqrt{t})/(2\sqrt{t}) = \sigma(\sqrt{t})/2$  is decreasing on  $\mathbb{R}_+$ .*
2. *Using (a), (c), and (e), we see that if  $t > 0$ , then  $\phi'(t) > 0$ , and hence  $\sigma(t) > 0$  for all  $t \in \mathbb{R}$ . Then*

$$(2.6) \quad \sigma(t) \in (0, \phi''(0^+)] \quad \forall t \in \mathbb{R}.$$

Hence, we can consider  $b$  in (2.4) only on  $(0, \phi''(0^+)]^r$ . For any  $b$  fixed, the minimizer function  $\chi$ , introduced in (1.18), is determined by solving with respect to  $x$  the equation  $D_1\mathcal{J}(x, b) = 0$ , that is,  $H(b)x = 2A^T y$ , where

$$(2.7) \quad H(b) = 2A^T A + \beta G^T \text{diag}(b)G.$$

LEMMA 2.1. *Consider the matrix-valued function  $H : \mathbb{R}^r \rightarrow \mathbb{R}^{p \times p}$  given in (2.7). If assumption (1.4) holds, then for every  $\zeta > 0$  there is a constant  $\lambda_\zeta > 0$  such that for any  $b \in [\zeta, \infty)^r$ , all eigenvalues of  $H(b)$  are greater than or equal to  $\lambda_\zeta$ .*

*Proof.* Given  $\zeta > 0$ , let  $\lambda_\zeta$  be the smallest eigenvalue of  $H(\zeta\mathbb{1})$  and  $u_0 \in S$  be the corresponding eigenvector. Clearly,  $\lambda_\zeta \geq 0$ . We have

$$u_0^T H(\zeta\mathbb{1})u_0 = \lambda_\zeta = 2\|Au_0\|^2 + \beta\zeta\|Gu_0\|^2.$$

If  $u_0 \in \ker(A^T A)$ , then (1.4) guarantees that  $\|Gu_0\|^2 > 0$ . By (1.4) again, if  $u_0 \in \ker(G^T G)$ , we have  $\|Au_0\|^2 > 0$ . In all cases,  $\lambda_\zeta > 0$ . Furthermore, for any  $b \in [\zeta, \infty)^r$  and  $u \in S$ ,

$$u^T H(b)u = 2\|Au\|^2 + \beta \sum_{i=1}^r b_i (g_i^T u)^2 \geq 2\|Au\|^2 + \beta\zeta\|Gu\|^2 = u^T H(\zeta\mathbb{1})u \geq \lambda_\zeta.$$

Hence we have the conclusion.  $\square$

If (1.4) and (1.10) hold, for every  $b \in (0, \infty)^r$  fixed,  $H(b)$  is invertible and  $\chi$  reads

$$(2.8) \quad \chi(b) = (H(b))^{-1} 2A^T y.$$

Although  $\mathcal{J}(x, \cdot)$  and  $\mathcal{J}(\cdot, b)$  are convex functions,  $(x, b) \rightarrow \mathcal{J}(x, b)$  is generally nonconvex.<sup>1</sup> The convergence of  $\{x^{(k)}\}$  in the alternate minimization sequence

<sup>1</sup>For  $x \in \mathbb{R}$ ,  $y \in \mathbb{R}$ ,  $A = 1$ , and  $g_1 = 1$ , we have  $J(x) = (x - y)^2 + \beta\phi(x)$ , and hence  $\mathcal{J}(x, b) = (x - y)^2 + \frac{\beta}{2}bx^2 + \beta\psi(b)$ . Consider that  $\psi$  is twice differentiable on  $\mathbb{R}_+^*$ . The second differential of  $\mathcal{J}$  at an arbitrary  $(x, b)$  reads

$$D^2\mathcal{J}(x, b) = \begin{bmatrix} 2 + \beta b & \beta x \\ \beta x & \beta\psi''(b) \end{bmatrix}.$$

$D^2\mathcal{J}(x, b)$  is not positive definite for all  $(x, b) \in \mathbb{R} \times \mathbb{R}_+^*$  since for any  $b > 0$  and  $|x| > (2/\beta + b)\psi''(b)$  we have  $\det D^2\mathcal{J}(x, b) < 0$ .

$\{(x^{(k)}, b^{(k)})\}$  towards the minimizer  $\hat{x}$  of  $J$  was initially proven in [8] under strong assumptions on  $\phi$  and later in [19] under conditions similar to (2.1). Convergence towards local minimizers of a nonconvex  $J$  was addressed in [11]. It is easy to see that under our assumptions, the sequence  $\{x^{(k)}\}$  is convergent.

LEMMA 2.2. *Consider  $J$  of the form (1.2)–(1.3), where (1.4) and (2.1) hold. Let  $\mathcal{J}$  read as in (2.2)–(2.4). Consider the alternate minimization sequence  $\{(x^{(k)}, b^{(k)})\}$  defined by (1.14)–(1.15). We have*

$$\|x^{(k)} - x^{(k+1)}\| \rightarrow 0 \quad \text{as } k \rightarrow \infty.$$

*Proof.* Since the function  $\mathcal{J}(\cdot, b)$  is quadratic, for any  $k \in \mathbb{N}$  we can write that

$$\begin{aligned} \mathcal{J}(x^{(k-1)}, b^{(k)}) - \mathcal{J}(x^{(k)}, b^{(k)}) &= D_1 \mathcal{J}(x^{(k)}, b^{(k)}) \left( x^{(k-1)} - x^{(k)} \right) \\ &\quad + \frac{1}{2} \left( x^{(k-1)} - x^{(k)} \right)^T H(b^{(k)}) \left( x^{(k-1)} - x^{(k)} \right), \end{aligned}$$

where the matrix  $H(b^{(k)})$  reads as in (2.7). Since  $x^{(k)}$  minimizes  $\mathcal{J}(\cdot, b^{(k)})$ , we have  $D_1 \mathcal{J}(x^{(k)}, b^{(k)}) = 0$ . Then

$$(2.9) \quad \mathcal{J}(x^{(k-1)}, b^{(k)}) - \mathcal{J}(x^{(k)}, b^{(k)}) = \frac{1}{2} \left( x^{(k-1)} - x^{(k)} \right)^T H(b^{(k)}) \left( x^{(k-1)} - x^{(k)} \right).$$

The rest of the proof uses the following simple observation.

REMARK 4. *If (1.4) holds, and if  $\phi$  is convex and even, then  $J$  is 0-coercive (i.e.,  $\lim_{\|x\| \rightarrow \infty} J(x) = \infty$ ). Since by Remark 2 the sequence  $\{J(x^{(k)})\}$  is decreasing, there is  $\delta \in (0, \infty)$  such that*

$$(2.10) \quad x^{(k)} \in B(0, \delta) \quad \forall k \in \mathbb{N}.$$

For this  $\delta > 0$ , put

$$(2.11) \quad \zeta = \sigma(\delta \|G\|);$$

then  $\zeta > 0$  by Remark 3 condition 2. Then we have

$$(2.12) \quad \begin{aligned} x \in B(0, \delta) &\Rightarrow |g_i^T x| \leq \delta \|G\| \quad \forall i \in \{1, \dots, r\}, \\ &\Rightarrow \sigma(g_i^T x) \geq \zeta > 0 \quad \forall i \in \{1, \dots, r\}, \end{aligned}$$

where the second implication comes from Remark 3 condition 1. Since  $b_i^{(k)} = \sigma(g_i^T x^{(k-1)})$  for all  $i$  and  $k$ , combining (2.10) and (2.12) yields

$$(2.13) \quad b_i^{(k)} \geq \zeta > 0 \quad \forall i \in \{1, \dots, r\}, \quad \forall k \in \mathbb{N}.$$

Let  $\lambda_\zeta > 0$  be the constant exhibited in Lemma 2.1 relevant to  $\zeta$ . Then (2.9) yields

$$\mathcal{J}(x^{(k-1)}, b^{(k)}) - \mathcal{J}(x^{(k)}, b^{(k)}) \geq \frac{\lambda_\zeta}{2} \|x^{(k-1)} - x^{(k)}\|^2 \quad \forall k \in \mathbb{N}.$$

By Remark 2,  $\mathcal{J}(x^{(k-1)}, b^{(k)}) - \mathcal{J}(x^{(k)}, b^{(k)}) \rightarrow 0$  as  $k \rightarrow \infty$ , and hence we have the conclusion.  $\square$



**2.2. Convergence analysis.** We will assume that in addition to (2.1),

$$(2.14) \quad \phi \text{ is } \mathcal{C}^2 \text{ on } \mathbb{R} \text{ and } \mathcal{C}^3 \text{ near zero.}$$

REMARK 5. *If  $\phi$  satisfies (2.1) and (2.14), the function  $\sigma$  in (2.5) is  $\mathcal{C}^1$  on  $\mathbb{R}$  and  $\sigma'(0) = 0$ . Indeed, for any  $t \in \mathbb{R} \setminus \{0\}$ , the function*

$$(2.15) \quad \sigma'(t) = \frac{\phi''(t)t - \phi'(t)}{t^2}$$

*is continuous. Its limit for  $t \rightarrow 0$  is calculated using l'Hôpital's rule,*

$$\lim_{t \rightarrow 0} \sigma'(t) = \lim_{t \rightarrow 0} \frac{\frac{d}{dt}(\phi''(t)t - \phi'(t))}{\frac{d}{dt}(t^2)} = \lim_{t \rightarrow 0} \frac{\phi'''(t)}{2} = 0.$$

Notice that  $\phi'''(0) = 0$  because  $\phi$  is even and  $\mathcal{C}^3$  near zero.

THEOREM 2.3. *For  $J$  of the form (1.2)–(1.3), suppose that (1.4) and (1.10) hold and that  $\phi$  satisfies (2.1) and (2.14). Then the sequence  $\{(x^{(k)}, b^{(k)})\}$  generated by (1.16), where  $\sigma$  and  $\chi$  are given in (2.5) and (2.7)–(2.8), respectively, satisfies the following:*

(i) *for all  $k \in \mathbb{N}$ , we have*

$$(2.16) \quad x^{(k)} = F(x^{(k-1)}),$$

*where  $F : \mathbb{R}^p \rightarrow \mathbb{R}^p$  is  $\mathcal{C}^1$  on  $B(0, \delta)$ , for any  $\delta \in (0, \infty)$ , and reads*

$$(2.17) \quad F(x) = x - (\mathcal{H}(x))^{-1}DJ(x),$$

$$(2.18) \quad \mathcal{H}(x) = 2A^T A + \beta G^T \text{diag} \left( [\sigma(g_i^T x)]_{i=1}^r \right) G \succ 0,$$

*and  $DJ$  is the differential of the original cost-function  $J$  given in (1.2)–(1.3);*

(ii) *as  $k \rightarrow \infty$ , the sequence  $\{x^{(k)}\}$  converges to the unique minimizer  $\hat{x}$  of  $J$ ;*  
 (iii) *the spectral radius of  $DF$  at  $\hat{x}$  satisfies*

$$(2.19) \quad \rho(DF(\hat{x})) \leq K \max_{1 \leq i \leq r} \mathcal{R}(|g_i^T \hat{x}|) \leq K \sup_{0 \leq t \leq \mu} \mathcal{R}(t) < 1,$$

*where  $\mu = \max_{1 \leq i \leq r} |g_i^T \hat{x}|$ ,  $K \in [0, 1]$ , and  $\mathcal{R} : \mathbb{R}_+ \rightarrow [0, 1]$  are continuous and read*

$$(2.20) \quad K = \frac{\beta \phi''(0) \|G\|^2}{2\nu^2 + \beta \phi''(0) \|G\|^2},$$

$$(2.21) \quad \mathcal{R}(t) = 1 - \frac{\phi''(t)}{\sigma(t)}.$$

The shape of  $\mathcal{R}$  for different functions  $\phi$  can be seen in Table 4.2 and Figures 4.1 and 4.2. Notice that  $\mathcal{R}$  is monotone increasing provided that  $\phi''$  is monotone decreasing on  $\mathbb{R}_+$  and that the latter is true for all potential functions cited in (1.6)–(1.9).

*Proof.* Let  $\delta > 0$  be the radius exhibited in Remark 4 and let  $\zeta$  be defined as in (2.11). The first step of iteration  $k$  is to calculate

$$(2.22) \quad b_i^{(k)} = \sigma(g_i^T x^{(k-1)}) \quad \forall i \in \{1, \dots, r\}.$$

Noticing that  $b^{(k)}$  satisfies (2.13), the matrix  $H(b^{(k)})$  defined by (2.7) is invertible (cf. Lemma 2.1). The second step of iteration  $k$  is to calculate  $x^{(k)}$  using (2.8):

$$(2.23) \quad x^{(k)} = \left( H(b^{(k)}) \right)^{-1} 2A^T y = \left( 2A^T A + \beta G^T \text{diag}(b^{(k)}) G \right)^{-1} 2A^T y.$$

Introducing (2.22) into (2.23) yields

$$x^{(k)} = F(x^{(k-1)}),$$

where  $F : \mathbb{R}^p \rightarrow \mathbb{R}^p$  reads

$$(2.24) \quad F(x) = (\mathcal{H}(x))^{-1} 2A^T y,$$

and  $\mathcal{H}$  is as given in (2.18). From (1.2)–(1.3),

$$(2.25) \quad DJ(x) = 2A^T Ax + \beta G^T [\phi'(g_i^T x)]_{i=1}^r - 2A^T y.$$

By (2.5), for every  $i \in \{1, \dots, r\}$  we can write

$$\phi'(g_i^T x) = \sigma(g_i^T x) g_i^T x.$$

Then (2.25) yields

$$\begin{aligned} 2A^T y &= 2A^T Ax + \beta G^T \text{diag} \left( [\sigma(g_i^T x)]_{i=1}^r \right) Gx - DJ(x) \\ &= \mathcal{H}(x) x - DJ(x). \end{aligned}$$

Inserting this expression into (2.24) yields (2.17).

Let us fix  $\delta > 0$  arbitrarily. Then define  $\zeta$  according to (2.11). Based on Lemma 2.1,  $b \rightarrow (H(b))^{-1}$  is  $\mathcal{C}^1$  on  $[\zeta, \phi''(0)]^r$  since  $b \rightarrow H(b)$  is  $\mathcal{C}^\infty$  and

$$D(H(b))^{-1} = -(H(b))^{-1} DH(b) (H(b))^{-1}$$

is continuous. Using (2.12) we have  $[\sigma(g_i^T x)]_{i=1}^r \in [\zeta, \phi''(0)]^r$  for all  $x \in B(0, \delta)$  and, since  $\sigma$  is  $\mathcal{C}^1$  on  $\mathbb{R}$  (see Remark 5), it follows that

$$x \rightarrow (\mathcal{H}(x))^{-1} = \left( H \left( [\sigma(g_i^T x)]_{i=1}^r \right) \right)^{-1}$$

is  $\mathcal{C}^1$  on  $B(0, \delta)$ . Combining the latter with the fact that  $J$  is  $\mathcal{C}^2$  shows that  $F$  is  $\mathcal{C}^1$  on  $B(0, \delta)$ . The proof of (i) is complete.

By Lemma 2.2, the sequence  $\{x^{(k)}\}$  converges; let  $\hat{x}$  denote its limit as  $k \rightarrow \infty$ . Since  $F$  is continuous, considering (2.16) for  $k \rightarrow \infty$  yields

$$(2.26) \quad \hat{x} = F(\hat{x}).$$

Introducing this into (2.17) shows that  $(\mathcal{H}(\hat{x}))^{-1} DJ(\hat{x}) = 0$  and consequently that  $DJ(\hat{x}) = 0$ . Since  $J$  has a unique minimizer (cf. Remark 1), it follows that  $J$  reaches its minimum at  $\hat{x}$ . Hence we prove (ii).

Next, we determine  $DF$  at an arbitrary  $x$ . Using (2.24),

$$\mathcal{H}(x)F(x) = 2A^T y \quad \forall x \in B(0, \delta).$$

Differentiating both sides of this equation with respect to  $x$  yields

$$(2.27) \quad D\mathcal{H}(x)F(x) + \mathcal{H}(x)DF(x) = 0 \quad \forall x \in B(0, \delta).$$

Then  $DF(x) = -(\mathcal{H}(x))^{-1}D\mathcal{H}(x)F(x)$ , where

$$\begin{aligned} D\mathcal{H}(x)F(x) &= \lim_{\epsilon \rightarrow 0} \frac{\mathcal{H}(x + \epsilon F(x)) - \mathcal{H}(x)}{\epsilon} = \beta G^T \text{diag} \left( [\sigma'(g_i^T x) g_i^T F(x)]_{i=1}^r \right) G \\ &= -\beta G^T \text{diag} \left( [|\sigma'(g_i^T x) g_i^T x|]_{i=1}^r \right) G \\ &\quad -\beta G^T \text{diag} \left( [\sigma'(g_i^T x) g_i^T (x - F(x))]_{i=1}^r \right) G, \end{aligned}$$

where we use the fact that  $\sigma'(t)t \leq 0$  for all  $t \in \mathbb{R}$ . It follows that

$$(2.28) \quad DF(x) = M(x) + E(x),$$

where

$$(2.29) \quad M(x) = (\mathcal{H}(x))^{-1} \beta G^T \text{diag} \left( [|\sigma'(g_i^T x) g_i^T x|]_{i=1}^r \right) G,$$

$$(2.30) \quad E(x) = (\mathcal{H}(x))^{-1} \beta G^T \text{diag} \left( [\sigma'(g_i^T x) g_i^T (x - F(x))]_{i=1}^r \right) G.$$

The eigenvalues of  $M(x)$ , for  $x \in \mathbb{R}^p$ , are considered next.

LEMMA 2.4. *Under the assumptions of Theorem 2.3, then we have the spectral radius of the matrix-valued function  $M : \mathbb{R}^p \rightarrow \mathbb{R}^{p \times p}$  defined by (2.29) satisfies*

$$(2.31) \quad \rho(M(x)) \leq K \max_{1 \leq i \leq r} \mathcal{R}(|g_i^T x|) \quad \forall x \in \mathbb{R}^p,$$

where  $K$  and  $\mathcal{R}$  are given in (2.20) and (2.21), respectively.

*Proof of Lemma 2.4.* Let us first focus on the function  $\mathcal{R}$  defined in (2.21). Clearly,  $\mathcal{R}(t) \leq 1$  for all  $t \in \mathbb{R}_+$  since  $\phi''(t) \geq 0$  and  $\sigma(t) > 0$  for all  $t \in \mathbb{R}$ . Using (2.5), (2.6), and (2.15), for any  $t \geq 0$  we can write

$$(2.32) \quad \mathcal{R}(t) = \frac{1}{\sigma(t)} \left( \frac{\phi'(t)}{t} - \phi''(t) \right) = \frac{-\sigma'(t)t}{\sigma(t)},$$

where in particular  $\mathcal{R}(0) = 0$ . For all  $t \in \mathbb{R}_+$ , Remark 3 condition 1 says that  $\sigma'(t) \leq 0$ , and hence  $\mathcal{R}(t) \geq 0$ . This allows us to write down that

$$(2.33) \quad \mathcal{R}(|t|) = \frac{|\sigma'(t)t|}{\sigma(t)} \quad \forall t \in \mathbb{R}.$$

Given  $x \in \mathbb{R}^p$ , let  $\lambda \in \mathbb{C}$  be the largest-in-magnitude eigenvector of  $M(x)$ . Let  $u \in \mathbb{C}^p$  be such that  $\|u\| = 1$  and  $M(x)u = \lambda u$ . Using the expression for  $M$ , we obtain

$$(2.34) \quad \beta(Gu)^T \text{diag} \left( [|\sigma'(g_i^T x) g_i^T x|]_{i=1}^r \right) Gu = \lambda u^T \mathcal{H}(x) u \geq 0.$$

The last inequality comes from the fact that  $G^T \text{diag} \left( [|\sigma'(g_i^T x) g_i^T x|]_{i=1}^r \right) G$  is semi-positive definite and that  $\mathcal{H}(x)$  is positive definite. It implies that  $\lambda$  is real with  $\lambda \geq 0$

and that  $u \in S$ . If  $\lambda = 0$ , then (2.31) is satisfied. In the following, consider that  $\lambda > 0$ . Inserting (2.18) into (2.34) yields

$$\beta(Gu)^T \text{diag} \left( [ |\sigma'(g_i^T x) g_i^T x| - \lambda \sigma(g_i^T x) ]_{i=1}^r \right) Gu = 2\lambda \|Au\|^2 \geq 2\lambda \nu^2 \geq 0.$$

Using (2.33), this is equivalent to

$$(2.35) \quad \beta \sum_{i=1}^r \sigma(g_i^T x) (\mathcal{R}(g_i^T x) - \lambda) (g_i^T u)^2 = 2\lambda \|Au\|^2 \geq 2\lambda \nu^2 \geq 0.$$

Let  $j \in \{1, \dots, r\}$  be such that

$$\sigma(g_j^T x) (\mathcal{R}(g_j^T x) - \lambda) \geq \sigma(g_i^T x) (\mathcal{R}(g_i^T x) - \lambda) \quad \forall i \in \{1, \dots, r\}.$$

From (2.35) we find that  $\sigma(g_j^T x) (\mathcal{R}(g_j^T x) - \lambda) \geq 0$ . Using (2.35) yet again,

$$\beta \sigma(g_j^T x) (\mathcal{R}(g_j^T x) - \lambda) \|G\|^2 \geq \beta \sigma(g_j^T x) (\mathcal{R}(g_j^T x) - \lambda) \|Gu\|^2 \geq 2\lambda \nu^2.$$

Since  $\sigma(t) \leq \phi''(0)$  for all  $t \in \mathbb{R}$  (cf. Remark 3 condition 2), we obtain

$$\mathcal{R}(|g_j^T x|) - \lambda \geq \frac{2\lambda \nu^2}{\beta \phi''(0) \|G\|^2}.$$

It follows that

$$\lambda \leq K \mathcal{R}(|g_j^T x|) \leq K \max_{1 \leq i \leq r} \mathcal{R}(|g_i^T x|),$$

where  $K \leq 1$  is the constant given in (2.20).  $\square$

We pursue the proof of the theorem. From (2.26) it follows that  $E(\hat{x}) = 0$ , and hence

$$(2.36) \quad DF(\hat{x}) = M(\hat{x}).$$

On the other hand, the function  $\mathcal{R}$  is continuous on  $\mathbb{R}_+$  with  $\mathcal{R}(0) = 0$ . The first two inequalities in (2.19) are an immediate consequence of (2.36) and Lemma 2.4. If  $\phi''(t) > 0$ , for all  $t \in \mathbb{R}$ , (2.6) and (2.21) show that  $\mathcal{R}(t) < 1$ , for all  $t \in \mathbb{R}_+$ , and hence

$$\sup_{0 \leq t \leq \mu} \mathcal{R}(t) < 1,$$

since  $[0, \mu]$  is compact. Thus, the assumption in (1.10) ensures that

$$K < 1 \quad \text{and/or} \quad \sup_{0 \leq t \leq \mu} \mathcal{R}(t) < 1,$$

and hence (2.19).  $\square$

REMARK 6. *Observe that  $F$  is not necessarily a contraction. By (2.28),  $DF$  is composed of two terms,  $M$  and  $E$ . Since  $\|E(x^{(k)})\|$  is controlled by the residual  $\|x^{(k)} - F(x^{(k)})\|$ , if  $x^{(k)}$  is far from  $\hat{x}$ , we can have  $\rho(DF(x^{(k)})) > 1$ . The method being convergent, as far as iterations progress,  $\|E(x^{(k)})\|$  diminishes and the convergence is controlled essentially by  $M$ , which is a contraction. Since  $DF$  is continuous at  $\hat{x}$ , (2.19) shows that there is  $\eta > 0$  such that*

$$\sup_{x \in B(\hat{x}, \eta)} \rho(DF(x)) < 1.$$

Furthermore, there is  $k_0 \in \mathbb{N}$  such that

$$x^{(k)} \in B(\hat{x}, \eta) \quad \forall k \geq k_0.$$

It follows that  $F$  is a contraction on  $B(\hat{x}, \eta)$ , i.e., in the vicinity of the minimizer  $\hat{x}$ .

The assumption in (2.14) that  $\phi$  is  $\mathcal{C}^3$  near zero is only technical. The proof is easy to extend to functions  $\phi$  which are  $\mathcal{C}^2$  and have  $\phi'''(0^+)$  finite. Even if  $\sigma$  is nondifferentiable at zero, its side derivatives there are then finite. Noticing that the function  $t \rightarrow \sigma(t)t$  is continuous, the analysis of  $M$  in Lemma 2.4 remains unchanged. An example of a function  $\phi$  with  $\sigma$  nonsmooth at zero is  $\phi(t) = |t|/\alpha - \log(1 + |t|/\alpha)$  since we have  $\sigma'(0^+) = \alpha = -\sigma'(0^-)$ . Extending the proof to functions  $\phi$  as discussed above would considerably complicate the presentation.

**COROLLARY 2.5.** *Under the assumptions and the notations of Theorem 2.3, then we have the root-convergence factor  $C(F, \hat{x})$ , given in (1.22), satisfies*

$$C(F, \hat{x}) \leq K \max_{1 \leq i \leq r} \mathcal{R}(|g_i^T \hat{x}|) \leq K \sup_{0 \leq t \leq \mu} \mathcal{R}(t) < 1,$$

whatever norm  $\|\cdot\|$  is used in (1.22).

*Proof.* It is easy to check that  $F$  satisfies all conditions required in the linear convergence theorem [26, p. 301]. By the latter theorem,  $C(F, \hat{x}) = \rho(DF(\hat{x}))$ . The result follows from Theorem 2.3.  $\square$

### 3. Additive form of half-quadratic regularization.

**3.1. Preliminary facts.** This form is considered under the following conditions:

- (a)  $\phi$  is convex,
- (b)  $c > 0$  is such that  $t \rightarrow ct^2/2 - \phi(t)$  is convex,
- (c)  $\phi(t) = \phi(-t) \quad \forall t \in \mathbb{R}$ ,
- (d)  $\phi$  is continuous on  $\mathbb{R}$ ,
- (e)  $\lim_{|t| \rightarrow \infty} \phi(t)/t^2 < c/2$ .

By (a)–(b),  $\phi$  is differentiable.<sup>2</sup> The dual function  $\psi$  is given by

$$\psi(s) = \sup_{t \in \mathbb{R}} \left\{ -\frac{1}{2} \left( t\sqrt{c} - \frac{s}{\sqrt{c}} \right)^2 + \phi(t) \right\},$$

where  $c > 0$  is as specified in (3.1). Under (3.1),  $\psi$  is convex and the requirement in (1.12) is satisfied since

$$\phi(t) = \inf_{s \in \mathbb{R}} \left\{ \frac{1}{2} \left( t\sqrt{c} - \frac{s}{\sqrt{c}} \right)^2 + \psi(s) \right\}.$$

The connections between  $\phi$  and  $\psi$  under different assumptions on  $\phi$  were considered in [3, 10, 14, 19]. For completeness, these are sketched in section 7.2 in the appendix.

<sup>2</sup>By (a),  $\phi'(t^-) \leq \phi'(t^+)$  for all  $t$ . If for some  $t$  the latter inequality is strict, (b) cannot be satisfied for any  $c$ . It follows that  $\phi'(t^-) = \phi'(t^+) = \phi'(t)$ . For example, if we apply (3.2) to  $\phi(t) = |t|$  with  $c = 1$ , we get  $\psi(s) = |s| + 1/2$ . Then (3.3) yields  $\min_s (\psi(s) + (s-t)^2/2) \neq \phi(t)$ .

According to (1.11) and (1.21),  $\mathcal{J}$  reads

$$\begin{aligned} \mathcal{J}(x, b) &= \|Ax - y\|^2 + \beta \sum_{i=1}^r \left( \frac{1}{2} \left( \sqrt{c} g_i^T x - \frac{1}{\sqrt{c}} b_i \right)^2 + \psi(b_i) \right) \\ (3.4) \qquad &= \|Ax - y\|^2 + \frac{\beta}{2} \left\| \sqrt{c} Gx - \frac{1}{\sqrt{c}} b \right\|^2 + \beta \sum_{i=1}^r \psi(b_i). \end{aligned}$$

The expression of  $\sigma$  for several functions  $\phi$  is presented in Table 4.1. The minimizer function  $\sigma$  defined by (1.17) admits an explicit form [7, 10, 3] and its derivation is recalled in section 7.2 in the appendix:

$$(3.5) \qquad \qquad \qquad \sigma(t) = ct - \phi'(t).$$

If (1.4) holds, the minimizer function  $\chi$  defined by (1.18) reads

$$(3.6) \qquad \qquad \qquad \chi(b) = \mathcal{H}^{-1} (2A^T y + \beta G^T b),$$

where

$$(3.7) \qquad \qquad \qquad \mathcal{H} = 2A^T A + \beta c G^T G \succ 0$$

is invertible thanks to (1.4).

As exhibited in [19],  $\mathcal{J}$  is convex on  $\mathbb{R}^p \times \mathbb{R}^r$ , and hence the convergence of alternate minimization towards the minimizer  $\hat{x}$  of  $J$ . Convergence when  $x$  and  $y$  are defined on a bounded domain of  $\mathbb{R}^2$ —in the context of image restoration—was established in [3] under appropriate assumptions on  $\phi$ .

**3.2. Convergence analysis.** The analysis here is developed in a similar fashion as for the multiplicative form.

**THEOREM 3.1.** *For  $J$  of the form (1.2)–(1.3), suppose that (1.4) and (1.10) hold and that  $\phi$  is  $\mathcal{C}^2$  on  $\mathbb{R}$  and satisfies (3.1). The sequence  $\{(x^{(k)}, b^{(k)})\}$  generated by (1.16), where  $\sigma$  and  $\chi$  are given in (3.5) and (3.6)–(3.7), respectively, satisfies the following:*

(i) *for all  $k \in \mathbb{N}$ , we have*

$$(3.8) \qquad \qquad \qquad x^{(k)} = F(x^{(k-1)}),$$

where  $F : \mathbb{R}^p \rightarrow \mathbb{R}^p$  is  $\mathcal{C}^1$  and reads

$$(3.9) \qquad \qquad \qquad F(x) = x - \mathcal{H}^{-1} DJ(x),$$

$$(3.10) \qquad \qquad \qquad \mathcal{H} = 2A^T A + \beta c G^T G \succ 0;$$

(ii) *for every  $x \in \mathbb{R}^p$ ,*

$$(3.11) \qquad \rho(DF(x)) \leq K \max_{1 \leq i \leq r} \mathcal{R}(|g_i^T x|) \leq K \sup_{0 \leq t \leq \mu} \mathcal{R}(t) < 1,$$

where  $\mu = \max_{1 \leq i \leq r} |g_i^T x|$ ,  $K \in [0, 1]$ , and  $\mathcal{R} : \mathbb{R}_+ \rightarrow [0, 1]$  is continuous and read

$$(3.12) \qquad \qquad \qquad K = \frac{\beta c \|G\|^2}{2\nu^2 + \beta c \|G\|^2},$$

$$(3.13) \qquad \qquad \qquad \mathcal{R}(t) = 1 - \frac{\phi''(t)}{c};$$

(iii) as  $k \rightarrow \infty$ ,  $\{x^{(k)}\}$  converges to the unique minimizer  $\hat{x}$  of  $J$ .

The function  $\mathcal{R}$  for several potential functions  $\phi$  is presented in Table 4.2 and Figures 4.1 and 4.2. For all potential functions cited in (1.6)–(1.9),  $\phi''$  is monotone decreasing on  $\mathbb{R}_+$ , and hence  $\mathcal{R}$  is monotone increasing on  $\mathbb{R}_+$ .

*Proof.* Given  $x^{(k-1)}$ , iterate  $b^{(k)}$  is calculated using (3.5):

$$(3.14) \quad b^{(k)} = \left[ \sigma(g_i^T x^{(k-1)}) \right]_{i=1}^r = cGx^{(k-1)} - \left[ \phi'(g_i^T x^{(k-1)}) \right]_{i=1}^r.$$

Next,  $x^{(k)}$  is calculated using (3.6)–(3.7):

$$x^{(k)} = \mathcal{H}^{-1}(2A^T y + \beta G^T b^{(k)}).$$

Inserting (3.14) into the expression given above yields

$$(3.15) \quad x^{(k)} = F(x^{(k-1)}),$$

where  $F : \mathbb{R}^p \rightarrow \mathbb{R}^p$  reads

$$(3.16) \quad F(x) = \mathcal{H}^{-1} \left( 2A^T y + \beta cG^T Gx - \beta G^T \left[ \phi'(g_i^T x) \right]_{i=1}^r \right).$$

The expression for  $DJ$  given in (2.25) yields

$$\begin{aligned} 2A^T y &= 2A^T Ax + \beta G^T \left[ \phi'(g_i^T x) \right]_{i=1}^r - DJ(x) \\ &= \mathcal{H}x - \beta cG^T Gx + \beta G^T \left[ \phi'(g_i^T x) \right]_{i=1}^r - DJ(x). \end{aligned}$$

Introducing this into (3.16) leads to  $F(x) = \mathcal{H}^{-1}(\mathcal{H}x - DJ(x))$ , and hence to (3.9). Since  $J$  is  $\mathcal{C}^2$ , it is straightforward that  $F$  is  $\mathcal{C}^1$  on  $\mathbb{R}^p$ . The proof of (i) is complete.

The differential of  $F$  at an arbitrary  $x \in \mathbb{R}^p$  reads

$$(3.17) \quad \begin{aligned} DF(x) &= I - \mathcal{H}^{-1} \left( 2A^T A + \beta G^T \text{diag} \left( \left[ \phi''(g_i^T x) \right]_{i=1}^r \right) G \right) \\ &= \beta c \mathcal{H}^{-1} G^T \text{diag} \left( \left[ \mathcal{R}(g_i^T x) \right]_{i=1}^r \right) G, \end{aligned}$$

where  $\mathcal{R}$  was introduced in (3.13). The eigenvalues of  $DF(x)$ , for any  $x \in \mathbb{R}^p$ , are analyzed below.

LEMMA 3.2. *Under the assumption of Theorem 3.1, we have*

$$(3.18) \quad \rho(DF(x)) \leq K \max_{1 \leq i \leq r} \mathcal{R}(|g_i^T x|) \quad \forall x \in \mathbb{R}^p,$$

where  $K$  and  $\mathcal{R}$  are given in (3.12) and (3.13), respectively.

*Proof of Lemma 3.2.* Consider first the function  $\mathcal{R}$  introduced in (3.13). Since  $\phi$  is  $\mathcal{C}^2$ , the constant  $c$  in (3.1)(b) satisfies

$$(3.19) \quad \phi''(t) \leq c \quad \forall t \in \mathbb{R},$$

and hence  $0 \leq \mathcal{R}(t) \leq 1$  for all  $t \in \mathbb{R}$ .

For  $x \in \mathbb{R}^p$ , let  $\lambda \in \mathbb{C}$  be the largest-in-magnitude eigenvector of  $DF(x)$ . Let  $u \in \mathbb{C}^p$  be such that  $\|u\| = 1$  and  $DF(x)u = \lambda u$ . Then using (3.17) we get

$$(3.20) \quad \beta c (Gu)^T \text{diag} \left( \left[ \mathcal{R}(g_i^T x) \right]_{i=1}^r \right) Gu = \lambda u^T \mathcal{H} u \geq 0.$$

The last inequality above comes from the facts that  $\mathcal{H}$  is positive definite and that  $G^T \text{diag}([\mathcal{R}(g_i^T x)]_{i=1}^r) G$  is semipositive definite. It shows that  $\lambda \geq 0$  and that  $u \in S$ . If  $\lambda = 0$ , then (3.18) is satisfied. In the following we consider that  $\lambda > 0$ . Introducing (3.10) into (3.20) yields

$$(3.21) \quad \beta c \sum_{i=1}^r (\mathcal{R}(g_i^T x) - \lambda) (g_i^T u)^2 = 2\lambda \|Au\|^2 \geq 2\lambda \nu^2 \geq 0.$$

Let  $j \in \{1, \dots, r\}$  be such that

$$\mathcal{R}(g_j^T x) - \lambda \geq \mathcal{R}(g_i^T x) - \lambda \quad \forall i \in \{1, \dots, r\}.$$

Then (3.21) shows that  $\mathcal{R}(g_j^T x) - \lambda \geq 0$  and we can write down that

$$\beta c (\mathcal{R}(g_j^T x) - \lambda) \|G\|^2 \geq \beta c (\mathcal{R}(g_j^T x) - \lambda) \|Gu\|^2 \geq 2\lambda \|Au\|^2 \geq 2\lambda \nu^2.$$

It follows that

$$\lambda \leq \frac{\beta c \|G\|^2}{2\nu^2 + \beta c \|G\|^2} \mathcal{R}(|g_j^T x|) \leq K \max_{1 \leq i \leq r} \mathcal{R}(|g_i^T x|).$$

The lemma is proved.  $\square$

We pursue the proof of the theorem. If  $\phi''(t) > 0$  for all  $t \in \mathbb{R}$ , noticing that  $\mathcal{R}$  is continuous on  $\mathbb{R}$  and that  $[0, \mu]$  is compact, (3.13) and (3.19) show that  $\sup_{0 \leq t \leq \mu} \mathcal{R}(t) < 1$ . By the assumption given in (1.10) we have

$$K < 1 \quad \text{and/or} \quad \sup_{0 \leq t \leq \mu} \mathcal{R}(t) < 1.$$

Combining this with Lemma 3.2 proves (ii).

Let  $\delta \in (0, \infty)$  be the radius exhibited in Remark 4. Define  $\eta = \delta \|G\|$ ; then

$$|g_i^T x| \leq \eta \quad \forall i \in \{1, \dots, r\}, \quad \forall x \in B(0, \delta).$$

Using Lemma 3.2 yet again,

$$(3.22) \quad \sup_{x \in B(0, \delta)} \rho(DF(x)) \leq K \sup_{x \in B(0, \delta)} \max_{1 \leq i \leq r} \mathcal{R}(|g_i^T x|) \leq K \sup_{0 \leq t \leq \eta} \mathcal{R}(t) < 1.$$

It follows that the sequence  $\{x^{(k)}\}$  converges; let  $\lim_{k \rightarrow \infty} x^{(k)} = \hat{x}$ . Since  $F$  is continuous, letting  $k \rightarrow \infty$  in (3.8) shows that  $\hat{x} = F(\hat{x})$ . Introducing the latter result into (3.9) shows that  $DJ(\hat{x}) = 0$ . Since  $J$  has a unique minimizer (Remark 1), it follows that  $J$  reaches its minimum at  $\hat{x}$ . Hence we prove (iii).  $\square$

**COROLLARY 3.3.** *Under the assumptions and the notations of Theorem 3.1, then we have the root-convergence factor  $C(F, \hat{x})$ , given in (1.22), satisfies*

$$C(F, \hat{x}) \leq K \max_{1 \leq i \leq r} \mathcal{R}(|g_i^T \hat{x}|) \leq K \sup_{0 \leq t \leq \mu} \mathcal{R}(t) < 1,$$

whatever norm  $\|\cdot\|$  is used in (1.22).

*Proof.* It is easy to see that the linear convergence theorem [26] can be applied, which shows that  $C(F, \hat{x}) = \rho(DF(\hat{x}))$ . The result follows from Theorem 3.1.  $\square$



REMARK 7 (the optimal value of  $c$ ). Based on (3.19), any  $c \geq \sup_{t \in \mathbb{R}} \phi''(t)$  can be used to construct  $\psi$  in (3.2). However, (3.12) and (3.13) show that both  $K$  and  $\mathcal{R}(t)$ , for all  $t \in \mathbb{R}$ , increase with  $c$ . It follows that when  $c$  is small, the bound on the root-convergence factor is improved. This shows that the best choice for  $c$  is

$$(3.23) \quad c = \sup_{t \in \mathbb{R}} \phi''(t).$$

Noticing that edge-preserving functions  $\phi$  satisfy  $\phi''(0) = \sup_{t \in \mathbb{R}} \phi''(t)$ , (3.23) becomes

$$(3.24) \quad c = \phi''(0).$$

This recommendation is corroborated by the experiments in section 5.

**3.3. Fast additive form based on nonsmooth dual function.** We now concentrate on the Huber potential function  $\phi$ :

$$(3.25) \quad \phi(t) = \begin{cases} \frac{1}{2}t^2 & \text{if } |t| \leq \alpha, \\ \alpha|t| - \frac{1}{2}\alpha^2 & \text{if } |t| > \alpha. \end{cases}$$

This potential function is edge preserving, since it is affine beyond  $\alpha$ . Unlike total variation regularization  $\phi(t) = |t|$  which yields a stair-casing effect, the quadratic part of  $\phi$  in (3.25) near 0 allows smoothly varying regions in  $\hat{x}$  to be restored [25]. Homogeneous regions in  $\hat{x}$  correspond with  $|g_i^T \hat{x}| \leq \alpha$  while edges correspond with  $|g_i^T \hat{x}| > \alpha$ .

REMARK 8. If  $A^T A$  is invertible, for every  $y \in \mathbb{R}^q$ , the relevant  $J$  has a unique minimizer  $\hat{x}$ . Analyzing the Hessian matrix  $D^2 J$  shows that the minimum  $\hat{x}$  of  $J$  is usually strict, even if  $A^T A$  is singular. Notice that  $\phi''$  is discontinuous at  $-\alpha$  and  $\alpha$ . If  $A^T A$  is invertible, it is easy to see that all  $y \in \mathbb{R}^q$ , yielding a minimizer  $\hat{x}$  such that  $|g_i^T \hat{x}| = \alpha$  for a certain number of indexes  $i$ , belong to a closed, negligible subset of  $\mathbb{R}^q$ . The chance of acquiring data placed in such subsets is null. For almost every  $y$ ,  $J$  is  $\mathcal{C}^\infty$  on a neighborhood of  $\hat{x}$ . This remains generally true even if  $A^T A$  is singular.

From (3.25) we have  $\max_{t \in \mathbb{R}} \{\phi''(t^-), \phi''(t^+)\} = 1$ . For  $c > 1$ , the dual  $\psi$  reads

$$\psi(s) = \begin{cases} \frac{s^2}{(2c(c-1))} & \text{if } |s| \leq (c-1)\alpha, \\ \frac{\alpha|s|}{c - \alpha^2(c-1)/(2c)} & \text{if } |s| > (c-1)\alpha. \end{cases}$$

Previous papers [7, 19] recommend taking  $c > 1$  in order to ensure that  $\psi$  is  $\mathcal{C}^1$ . Using (3.5), the minimizer function  $\sigma$  is

$$(3.26) \quad \sigma(t) = \begin{cases} (c-1)t & \text{if } |t| \leq \alpha, \\ ct - \alpha \operatorname{sign}(t) & \text{if } |t| > \alpha. \end{cases}$$

Instead, for  $c = 1$ , we get

$$(3.27) \quad \psi(s) = \alpha|s|.$$

It is easy to check<sup>3</sup> that applying (3.3) leads to  $\phi$  as given in (3.25). The function  $\sigma$  now takes a particularly simple form:

$$(3.28) \quad \sigma(t) = \begin{cases} 0 & \text{if } |t| \leq \alpha, \\ t - \alpha \operatorname{sign}(t) & \text{if } |t| > \alpha. \end{cases}$$

The computation of each  $b_i^{(k)} = \sigma(g_i^T x^{(k)})$  using (3.28) is much easier than using (3.26). Typically we have  $|g_i^T x^{(k)}| \leq \alpha$  for most of the indexes  $i \in \{1, \dots, r\}$  since most of the pixels are in homogeneous regions. The relevant  $b_i^{(k)}$  are set to zero after verifying that  $|g_i^T x^{(k)}| \leq \alpha$ : thus, they need no calculation.

The augmented cost-function  $\mathcal{J}$  now reads

$$(3.29) \quad \mathcal{J}(x, s) = \|Ax - y\|^2 + \beta \sum_{i=1}^r \left( \frac{1}{2} (g_i^T x - s_i)^2 + |s_i| \right).$$

The convergence of alternate minimization is easily deduced from Theorem 1.4, p. 73 in [16]. Furthermore, Theorem 3.1 can be extended to potential functions  $\phi$  which satisfy (3.1) and which at several points  $\alpha_i$  have  $\phi''(\alpha_i^-)$  and  $\phi''(\alpha_i^+)$  different but finite. Instead of  $DF$ , the subdifferential of  $F$  must be considered in (3.17). The inequalities presented in the proof of Theorem 3.1 remain basically the same.

**4. Theoretical comparisons of two forms.**

The expressions for  $\sigma$  and  $\mathcal{R}$  for both forms are displayed in Tables 4.1 and 4.2.

**4.1. Extension to alternative regularization terms.** In some image processing applications, the regularization term  $\Phi$  in (1.2) reads [2]

$$(4.1) \quad \Phi(x) = \sum_{i=1}^p \phi(\|G_i x\|),$$

where every  $G_i x$  is a vector approximating the gradient of  $x$  at pixel  $i$ . We wish to know whether the resultant  $J$  can be minimized using HQ regularization methods. According to (1.11)–(1.12), an auxiliary variable  $b_i$  is associated with each  $\|G_i x\|$ .

- In the multiplicative form of HQ regularization, the augmented cost-function  $\mathcal{J}$  is similar to (2.4):

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

Notice that  $\mathcal{J}(x, \cdot)$  is separable and the minimization over  $b$  is given by  $p$  scalar minimizer functions  $\sigma$  of the form (2.5). Since the function  $\mathcal{J}(\cdot, b)$  is quadratic, the minimizer function  $\chi$ , defined in (1.18), is linear in  $y$  and has an explicit form similar to (2.7)–(2.8). Minimization using the multiplicative form of HQ regularization is straightforward.

<sup>3</sup>Given  $t \in \mathbb{R}$ , define  $h_t : \mathbb{R} \rightarrow \mathbb{R}$  by  $h_t(s) = (s - t)^2/2 + \alpha|s|$ . The function  $h_t$  reaches its minimum at  $\hat{s} \in \mathbb{R}$  if and only if [18]

$$\begin{aligned} \hat{s} = 0 & \Rightarrow h'_t(0^-) \leq 0 \leq h'_t(0^+), & \text{i.e., } -t - \alpha \leq 0 \leq -t + \alpha; \\ \hat{s} \neq 0 & \Rightarrow h'_t(\hat{s}) = 0, & \text{i.e., } \hat{s} - t + \alpha \operatorname{sign}(\hat{s}) = 0. \end{aligned}$$

It is easily deduced that

$$\begin{aligned} |t| \leq \alpha & \Rightarrow \hat{s} = 0 & \Rightarrow \inf_s h_t(\hat{s}) = t^2/2, \\ |t| > \alpha & \Rightarrow \hat{s} = t - \alpha \operatorname{sign}(t) & \Rightarrow \inf_s h_t(\hat{s}) = \alpha|t| - \alpha^2/2. \end{aligned}$$

In other words,  $\inf_{s \in \mathbb{R}} ((s - t)^2/2 + \alpha|s|) = \phi(t)$  where  $\phi$  is given in (3.25).

TABLE 4.1

Minimizer function  $\sigma$  relevant to multiplicative and additive form of HQ regularization (cf. (2.5) and (3.5)) for different potential functions  $\phi$ .

Potential function	Multiplicative form	Additive form
$\phi(t) =  t ^\alpha, \quad \alpha \in (1, 2]$	$\sigma(t) = \alpha t ^{\alpha-2}$	not applicable
$\phi(t) = \sqrt{\alpha + t^2}$	$\sigma(t) = \frac{1}{\sqrt{\alpha + t^2}}$	$\sigma(t) = ct - \frac{t}{\sqrt{\alpha + t^2}}$
$\phi(t) = \frac{ t }{\alpha} - \log\left(1 + \frac{ t }{\alpha}\right)$	$\sigma(t) = \frac{1}{\alpha(\alpha +  t )}$	$\sigma(t) = ct - \frac{t}{\alpha(\alpha +  t )}$
$\phi(t) = \begin{cases} \frac{t^2}{2}, &  t  \leq \alpha \\ \alpha t  - \frac{\alpha^2}{2}, &  t  > \alpha \end{cases}$	$\sigma(t) = \begin{cases} \frac{1}{\alpha}, &  t  \leq \alpha \\ \frac{1}{ t }, &  t  > \alpha \end{cases}$	$\sigma(t) = \begin{cases} (c-1)t, &  t  \leq \alpha \\ ct - \alpha \text{sign}(t), &  t  > \alpha \end{cases}$
$\phi(t) = \log(\cosh(\alpha t))$	$\sigma(t) = \alpha \frac{\tanh(\alpha t)}{t}$	$\sigma(t) = ct - \alpha \tanh(\alpha t)$

- In the additive form of HQ regularization,  $\mathcal{J}$  reads

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

where for simplicity we take  $c = 1$ . Now the function  $\mathcal{J}(\cdot, b)$  is nonquadratic since it is nonsmooth. The minimizer function  $\chi$  defined by (1.18) is nonlinear in  $y$  and is difficult to calculate numerically. There is no advantage to consider minimization using the additive form of HQ regularization when  $\Phi$  is of the form (4.1).

**4.2. The bound on convergence rate.** The minimization of a  $J$  of the form (1.2)–(1.3) using either the multiplicative or the additive form of HQ regularization amounts to constructing a sequence of iterates  $\{x^{(k)}\}$  as

$$(4.2) \quad x^{(k)} = F(x^{(k-1)}),$$

where  $F : \mathbb{R}^p \rightarrow \mathbb{R}^p$  is of the form

$$F(x) = x - (\mathcal{H}(x))^{-1} DJ(x).$$

The matrix  $\mathcal{H}(x)$  here provides a correction of the steepest descent direction  $-DJ(x)$ . Recall that the classical Newton method corresponds to  $\mathcal{H}(x) = D^2J(x)$ , where

$$(4.3) \quad D^2J(x) = 2A^T A + \beta G^T \text{diag} \left( [\phi''(g_i^T x)]_{i=1}^p \right) G.$$

In the multiplicative form, the matrix  $\mathcal{H}(x)$  given in (2.18) differs from  $D^2J(x)$  only in that the entries of the diagonal matrix above are  $\sigma(g_i^T x) = \phi'(g_i^T x)/g_i^T x$ . Since  $\sigma(t) \approx \phi''(t)$  if  $t \approx 0$ , the entries relevant to small differences  $|g_i^T x|$  are almost the

same for  $\mathcal{H}$  and for  $D^2J$ . For  $|g_i^T x|$  large,  $\phi''(g_i^T x) \approx 0$  because  $\phi$  is edge preserving; such entries heavily deteriorate the conditioning of  $D^2J$ . On the other hand, (2.21) shows that  $\sigma(t) > \phi''(t)$  if  $t \neq 0$ . In fact,  $\sigma(g_i^T x)$  remains larger than  $\phi''(g_i^T x)$  for large differences  $|g_i^T x|$ , thus preventing  $\mathcal{H}(x)$  from being as ill-conditioned as  $D^2J(x)$ .

In the additive form,  $\mathcal{H}$  is constant for all  $x$ ; cf. (3.10). If we choose  $c = \phi''(0)$ , as recommended in (3.24), then  $\mathcal{H} \approx D^2J(\mathbf{1})$ . Here again, the entries of the diagonal matrix in (4.3) relevant to small differences  $|g_i^T x| \approx 0$  are the same for  $D^2J$  and  $\mathcal{H}$ .

The calculation of  $x^{(k)}$ , as given by (4.2), can also be put into the form

$$x^{(k)} = F^k(x^{(0)}), \quad \text{where } F^k = \underbrace{F \circ \dots \circ F}_{k \text{ times}}.$$

In the multiplicative form of HQ regularization,  $F$  is not necessarily a contraction except in the vicinity of the minimizer  $\hat{x}$ —see Remark 6. We can expect that in early iterations the decrease of  $J$  is relatively slow. In the additive form of HQ regularization, Theorem 3.1 and Lemma 3.2 show that  $F$  is a contraction.

By (2.19) and (3.11), the convergence rate relevant to the multiplicative and the additive forms of HQ regularization is upper bounded by a term of the form

$$(4.4) \quad K \sup_{|t| \leq \mu} \mathcal{R}(t).$$

If in the additive form we choose  $c = \phi''(0)$ , as recommended in (3.24), then the constant  $K$  is the same for both forms; cf. (2.20) and (3.12). We notice that  $K$  decreases when  $\nu^2$  increases, which means that convergence is faster if  $A^T A$  is well-conditioned. For  $A^T A$  singular,  $K = 1$ . Furthermore, comparing (2.21) and (3.13) shows that the function  $\mathcal{R}$  has a similar expression for both the multiplicative and the additive forms of HQ regularization. Moreover, for all edge-preserving functions  $\phi$  used in practice  $\mathcal{R}$  is monotone increasing on  $\mathbb{R}_+$  since  $\phi''$  is monotone decreasing on  $\mathbb{R}_+$ . In the multiplicative form, using that  $\mathcal{R}(t) \geq 0$  for all  $t \in \mathbb{R}_+$  and Remark 3, we find that

$$\phi''(t) \leq \sigma(t) \leq \phi''(0) \quad \forall t \in \mathbb{R}.$$

These inequalities are strict in the usual situation when  $t \rightarrow \phi(\sqrt{t})$  is strictly concave and  $\phi''(t) > 0$  for all  $t \in \mathbb{R}$ . In such a case, (2.21) and (3.13) lead to

$$\begin{array}{ccc} \mathcal{R}(t) & < & \mathcal{R}(t) \\ \text{multipl. form} & & \text{additive form} \end{array} \quad \forall t \in \mathbb{R} \setminus \{0\},$$

with  $\mathcal{R}(0) = 0$  for both forms. Table 4.2 and Figures 4.1 and 4.2 illustrate this result. This suggests that the multiplicative form needs fewer iterations than the additive form in order to reach the minimizer  $\hat{x}$ . This is corroborated by the experiments presented in section 5. In Figure 4.2, we show  $\mathcal{R}$  corresponding to  $\phi(t) = \sqrt{\alpha + t^2}$ , for  $\alpha = 0.5$  and  $\alpha = 0.05$ . It is worth noticing that for both forms,  $\mathcal{R}$  decreases when  $\alpha$  increases, hence convergence is faster if  $\alpha$  is larger. The constant  $\mu \in (0, \infty)$  in (4.4) is determined essentially by the magnitude of the differences  $|g_i^T x|$ . Since  $\mathcal{R}$  is monotone increasing on  $\mathbb{R}_+$ , convergence is slower if  $\hat{x}$  involves large differences.

**4.3. Computational cost.** In both the multiplicative and the additive forms of HQ regularization, the calculation of  $b^{(k)}$  has basically the same complexity. In the multiplicative form, the new iterate  $x^{(k)}$  is the solution of a linear system

$$(4.5) \quad \mathcal{H}(b^{(k)})x = 2A^T y,$$

TABLE 4.2

The convergence factor  $\mathcal{R}$  for multiplicative and additive HQ regularization; cf. (2.21) and (3.13).

Potential function	$\mathcal{R}$ for the multiplicative form	$\mathcal{R}$ for the additive form
$\phi(t) = \sqrt{\alpha + t^2}$	$\frac{t^2}{\alpha + t^2}$	$1 - \frac{\alpha}{c(\alpha + t^2)\sqrt{\alpha + t^2}}$
$\phi(t) =  t /\alpha - \log(1 +  t /\alpha)$	$\frac{t}{\alpha + t}$	$1 - \frac{1}{c(\alpha + t)^2}$
$\phi(t) = \begin{cases} \frac{t^2}{2}, &  t  \leq \alpha \\ \alpha t  - \frac{\alpha^2}{2}, &  t  > \alpha \end{cases}$	$\begin{cases} 0, & t \leq \alpha \\ 1, & t > \alpha \end{cases}$	$\begin{cases} 1 - \frac{1}{c}, & t \leq \alpha \\ 1, & t > \alpha \end{cases}$
$\phi(t) = \log(\cosh(\alpha t))$	$1 - \frac{\alpha t}{\sinh(\alpha t) \cosh(\alpha t)}$	$1 - \frac{\alpha^2(1 - \tanh^2(\alpha t))}{c}$

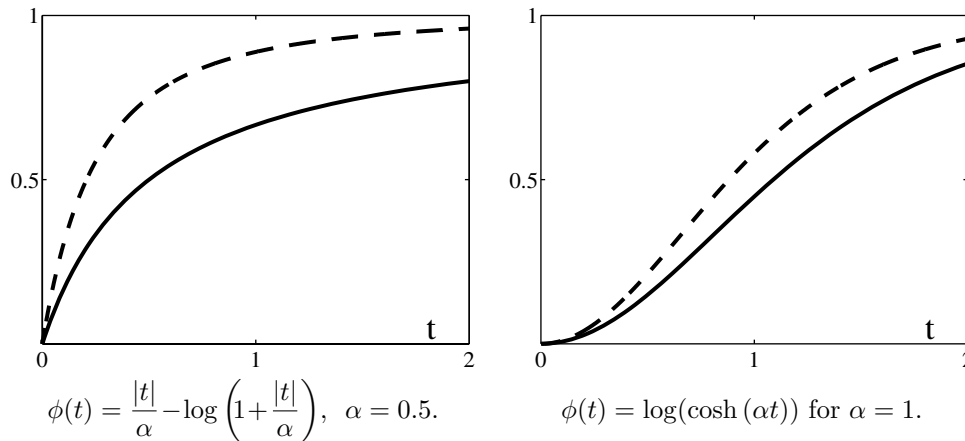


FIG. 4.1. The shape of  $\mathcal{R}$  for multiplicative (solid lines) and for additive (dotted lines) form of HQ regularization corresponding to different functions  $\phi$ . In the additive form,  $c = \phi''(0)$ .

where  $\mathcal{H}(b^{(k)})$  is given in (2.7). We note that the matrix  $\mathcal{H}(b^{(k)})$  is changing, but the right-hand side is fixed at each iteration. This system must be solved at each iteration. When  $p$  is small, (4.5) can be solved using Gaussian elimination. The complexity is then  $O(p^3)$ . In signal and image deblurring,  $A$  is a Toeplitz matrix. However,  $\mathcal{H}(b^{(k)})$  is not necessarily Toeplitz: no benefit can be made from the special structure of  $A$ . When  $p$  is large, the conjugate gradient method can be used to solve (4.5). Although invertible, the matrix  $\mathcal{H}(b^{(k)})$  can be ill-conditioned and/or there are a number of clusters in the small eigenvalues of the Hessian matrix. Solving (4.5) then needs a large number of iterations.

In the additive form,  $x^{(k)}$  is the solution of

$$(4.6) \quad \mathcal{H} x = 2A^T y + \beta G^T b^{(k)},$$

where  $\mathcal{H}$  is given in (3.7). Now  $\mathcal{H}$  is fixed, but the right-hand side is changing at each iteration. If  $p$  is small, we can compute  $\mathcal{H}^{-1}$  before starting iterations and

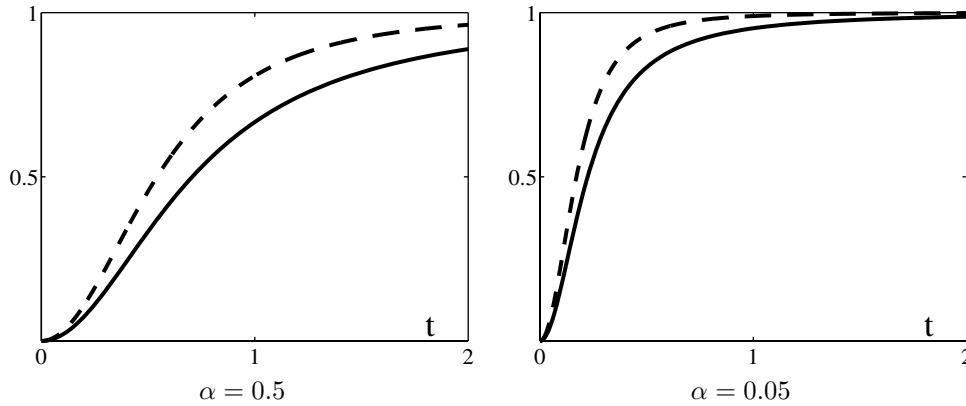


FIG. 4.2. The shape of  $\mathcal{R}$  as a function of  $\alpha$  for  $\phi(t) = \sqrt{\alpha + t^2}$ . Multiplicative form (solid lines) and additive form (dotted lines) for  $c = \phi''(0)$ .

then, at each iteration,  $x^{(k)} = \mathcal{H}^{-1}(2A^T y + \beta G^T b^{(k)})$ . A stable way to solve (4.6) is to first compute the Cholesky factorization of  $\mathcal{H}$  and then solve the corresponding triangular systems at each iteration. The factorization requires  $O(p^3)$  operations, but the complexity of solving a triangular system is only  $O(p^2)$ . Observe that if  $A$  is a Toeplitz matrix,  $\mathcal{H}$  is a Toeplitz-like matrix [20, 6] as well. We note that  $A^T A$  is a Toeplitz-like matrix and  $G$  is the discretization matrix of the first-order or the second-order differential operator. There are many direct or iterative Toeplitz solvers with costs ranging from  $O(p \log p)$  to  $O(p^2)$  [20, 6]. When  $p$  is large, the conjugate gradient method can be used to solve (4.6) at each iteration. Convergence can be improved using preconditioning techniques. When  $A$  is a Toeplitz-type or block-Toeplitz-Toeplitz-block-type matrix, transform-based preconditioning techniques have been proven to be very successful [24]. For example, if  $A$  is a blurring matrix generated by a symmetric point spread function,  $\mathcal{H}$  can be diagonalized by a fast transform matrix; then (4.6) is solved by using three fast transforms in  $O(p \log p)$  operations [24].

In conclusion, the computational cost of each iteration for the additive form of HQ regularization is smaller than for the multiplicative form. Moreover, it can substantially be improved using fast solvers and preconditioning.

## 5. Numerical results.

**5.1. One-dimensional signal.** Our goals are first to compare the convergence properties of the two forms of HQ regularization and second to compare them with standard minimization methods. All the computations are done in MATLAB 6.0. The MATLAB optimization routines are compared with our written MATLAB codes. Their computational times are reported in the tables in this section.

We give average results based on the restoration of 1000 generated original 128-length signals from blurred noisy data. If  $x_o$  is such an original signal, observed data read  $y = Ax_o + n$ , where

$$A_{i,j} = \gamma e^{-0.1(i-j)^2}, \quad i, j = 1, 2, \dots, p,$$

$\gamma$  is a normalization constant such that  $\sum_{i=1}^p A_{i,1} = 1$ , and  $n$  is white Gaussian noise yielding a signal-to-noise ratio (SNR) of 40dB. In Figure 5.1(a) we show an example of a randomly generated original signal  $x_o$  and in Figure 5.1(b) data  $y$  generated as explained above. For all minimization methods, the stopping criterion is

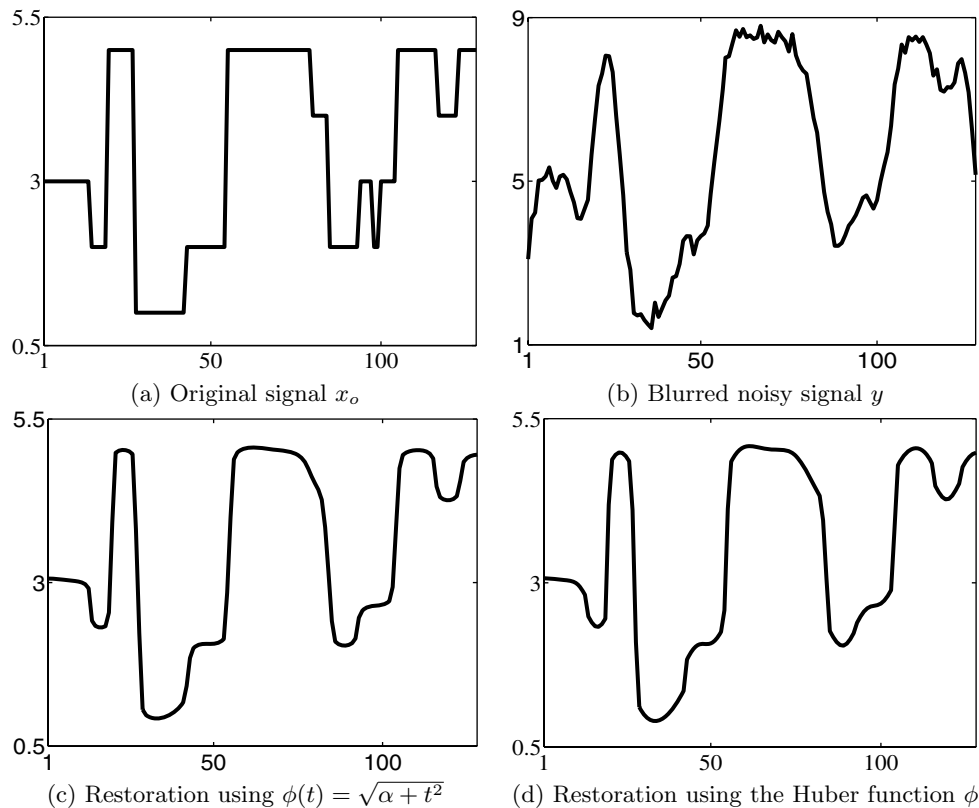


FIG. 5.1. Example of restoration of a randomly generated original signal (a) from blurred noisy data (b). Restored signals: (c)  $\phi(t) = \sqrt{\alpha + t^2}$ , and  $\alpha = 0.001$ ,  $\beta = 0.8$ ; (d)  $\phi(t) = t^2/2$  if  $|t| \leq \alpha$ ,  $\phi(t) = \alpha|t| - \alpha^2/2$  if  $|t| > \alpha$ , where  $\alpha = 0.1$ ,  $\beta = 10$ .

$\|x^{(k)} - x^{(k-1)}\|_2 < 1 \times 10^{-4} / \|x_o\|$ , while the maximum iteration number is 1000. For the restoration of each signal, we minimize cost-functions  $J$  defined using different potential functions  $\phi$  and different sets of parameters  $(\alpha, \beta)$ . Figures 5.1(c) and (d) show two restorations obtained from the data in Figure 5.1(b). Each cost-function  $J$  is minimized using the multiplicative form of HQ regularization, the additive form of HQ regularization where  $c = \phi''(0)$ , as well as using MATLAB optimization routines “fminunc.m.” In the latter optimization procedure, we use quasi-Newton updating schemes “bfgs” or “dfp,” the line search algorithm “cubicpoly,” as well as the steepest descent scheme “steepdesc.” Tables 5.1–5.5 summarize the average number of iterations (the first row of each set of parameters) and the average computational times in seconds (the second row of each set of parameters) for each minimization method. Below we briefly comment on the obtained results.

- The multiplicative form of HQ regularization allows us to minimize  $J$  using a smaller number of iterations than the additive form of HQ regularization.
- The overall computational times used by the additive form are less than those used by the multiplicative form.
- The performance of MATLAB optimization routines are poorer than those of both forms of HQ regularization in our MATLAB implementation. These results must be qualified by the observation that MATLAB run times for the

TABLE 5.1

Average number of iterations (upper row) and computational times (lower row) for  $\phi(t) = \sqrt{\alpha + t^2}$ .

$\beta$	$\alpha$	Multiplicative	Additive	BFGS	DFP	Steepest
0.8	0.00025	79.6	228.2	122.8	110.3	> 1000
		5.28	2.63	11.12	8.77	> 36.73
0.8	0.0005	74.6	207.4	110.5	101.6	> 1000
		4.94	2.39	9.96	8.08	> 36.71
0.8	0.001	68.8	182.9	100.1	89.4	990.2
		4.55	2.17	9.13	7.11	37.17
0.8	0.002	63.3	188.3	85.2	77.9	835.6
		4.15	2.19	7.79	6.23	30.81
0.8	0.004	56.4	181.4	76.0	65.4	697.5
		3.76	2.14	6.96	5.28	25.87
0.2	0.001	64.1	258.6	83.0	74.2	775.6
		4.24	2.90	7.62	6.03	28.49
0.4	0.001	64.7	226.3	92.9	81.7	886.5
		4.30	2.56	8.48	6.60	32.36
0.8	0.001	63.3	188.3	85.2	77.9	835.6
		4.15	2.19	7.79	6.23	30.81
1.6	0.001	72.6	174.0	111.8	100.4	> 1000
		4.78	2.00	10.06	7.98	> 36.74
3.2	0.001	84.7	168.0	126.4	112.4	> 1000
		5.58	1.93	11.30	8.85	36.89

TABLE 5.2

Average number of iterations and computational times for  $\phi(t) = t^2/2$ ,  $|t| \leq \alpha$  and  $\phi(t) = \alpha|t| - \alpha^2/2$ ,  $|t| > \alpha$ .

$\beta$	$\alpha$	Multiplicative	Additive	BFGS	DFP	Steepest
10	0.025	75.7	282.6	133.0	97.6	894.4
		5.03	3.23	11.67	7.68	32.65
10	0.05	75.8	267.4	109.8	86.6	721.8
		5.04	3.27	9.89	6.85	26.52
10	0.1	71.2	232.4	89.5	72.1	529.2
		4.74	2.81	7.96	5.61	19.49
10	0.2	54.3	147.2	63.7	47.3	329.1
		3.59	1.20	5.19	3.66	12.16
10	0.4	22.7	39.2	38.5	31.4	77.2
		1.54	0.62	3.49	2.43	2.92
2.5	0.1	62.3	195.6	56.5	59.5	323.0
		4.15	2.41	5.09	4.83	11.86
5	0.1	62.1	220.4	71.6	61.5	384.7
		4.13	2.68	6.40	4.87	14.08
10	0.1	71.2	232.4	89.5	72.1	529.2
		4.74	2.81	7.96	5.61	19.49
20	0.1	78.3	198.2	109.4	84.2	688.3
		5.19	2.46	9.75	6.56	25.47
40	0.1	96.0	154.6	132.4	105.6	961.7
		6.34	1.93	11.73	8.17	35.13

same operations can vary by orders of magnitude depending on the particular coding used. This result may show that HQ regularization is a powerful tool for the minimization of cost functions of the form (1.2)–(1.3).

- We have computed the condition numbers of the matrices  $\mathcal{H}(b^{(k)})$  and  $\mathcal{H}$  arising from the multiplicative form and the additive form of HQ regularization, respectively. As expected, the condition number of  $\mathcal{H}(b^{(k)})$  is always larger



TABLE 5.3

Average number of iterations and computational times for  $\phi(t) = |t|/\alpha - \log(1 + |t|/\alpha)$ .

$\beta$	$\alpha$	Multiplicative	Additive	BFGS	DFP	Steepest
0.005	0.025	45.4	182.9	68.1	52.1	608.3
		2.96	2.13	6.16	4.17	21.71
0.005	0.05	32.9	170.3	44.9	54.3	206.2
		2.19	2.00	4.19	4.46	7.55
0.005	0.1	25.1	111.4	42.6	54.7	224.1
		1.69	1.39	4.36	4.50	8.31
0.005	0.2	19.0	65.8	53.1	48.1	209.0
		1.27	0.90	4.89	3.85	7.49
0.005	0.4	13.8	36.6	59.4	58.3	252.8
		0.94	0.56	5.50	4.70	7.12
0.00125	0.1	24.8	117.4	69.6	61.6	290.5
		1.65	1.45	6.48	4.99	10.49
0.0025	0.1	24.2	114.7	55.6	54.6	260.7
		1.61	1.43	5.19	4.44	9.42
0.005	0.1	25.1	111.4	42.6	54.7	224.1
		1.69	1.39	4.36	4.50	8.31
0.01	0.1	26.2	111.9	38.4	44.9	161.3
		1.76	1.37	3.58	3.62	5.81
0.02	0.1	27.2	102.9	34.7	32.7	140.8
		1.82	1.28	3.24	2.69	5.14

TABLE 5.4

Average number of iterations and computational times for  $\phi(t) = \log(\cosh(\alpha t))$ .

$\beta$	$\alpha$	Multiplicative	Additive	BFGS	DFP	Steepest
0.5	0.25	3.0	3.0	32.2	324.1	132.6
		0.30	0.18	3.04	14.91	5.10
0.5	0.5	3.8	4.0	23.7	120.3	59.5
		0.34	0.20	2.17	5.65	2.36
0.5	1	5.5	6.1	213.5	211.7	30.8
		0.44	0.22	9.23	8.78	1.23
0.5	2	13.9	20.1	707.5	510.5	335.4
		0.98	0.40	30.70	22.74	12.49
0.5	4	30.5	63.7	> 1000	809.8	155.6
		2.07	0.90	42.08	36.91	6.02
0.125	1	6.9	8.2	220.2	120.6	66.5
		0.55	0.24	10.04	5.75	2.59
0.25	1	5.8	7.2	218.3	509.1	43.5
		0.47	0.24	9.72	21.45	1.72
0.5	1	5.5	6.1	213.5	211.7	30.8
		0.44	0.22	9.23	8.78	1.23
1	1	4.9	5.4	606.9	605.6	119.4
		0.41	0.22	25.75	25.32	4.51
2	1	4.5	4.9	902.7	803.3	804.9
		0.37	0.20	44.10	34.86	30.22

than that of  $\mathcal{H}$ . Some average results are reported in Table 5.6.

We have tested other examples where the blurred signal is corrupted with white Gaussian noise with SNR of 30dB or 50dB. We have also considered different point spread functions for  $A$ , such as

$$A_{i,j} = \begin{cases} \gamma e^{-0.1(i-j)^2}, & i, j = 1, 2, \dots, 10, \\ 0 & \text{otherwise,} \end{cases}$$

TABLE 5.5

Average number of iterations and computational times for  $\phi(t) = |t|^\alpha$ .

$\beta$	$\alpha$	Multiplicative	BFGS	DFP	Steepest
5	1.1	12.6	>1000	> 1000	> 1000
		1.36	38.39	37.89	37.93
5	1.15	18.9	>1000	> 1000	> 1000
		1.60	39.10	37.69	37.48
5	1.2	16.8	>1000	> 1000	> 1000
		1.26	38.76	38.48	38.12
5	1.25	14.4	>1000	> 1000	> 1000
		1.06	38.13	38.16	37.55
5	1.3	13.6	>1000	> 1000	> 1000
		0.98	37.91	37.56	37.12
1.25	1.2	18.2	>1000	> 1000	> 1000
		1.36	39.16	38.84	39.13
2.5	1.2	18.0	>1000	> 1000	> 1000
		1.28	39.01	38.45	38.56
5	1.2	16.8	>1000	> 1000	> 1000
		1.26	38.99	38.98	37.90
10	1.2	16.8	>1000	> 1000	> 1000
		1.17	38.59	38.39	38.67
20	1.2	16.3	>1000	> 1000	> 1000
		1.17	39.01	38.87	38.03

TABLE 5.6

Average condition numbers of  $\mathcal{H}(b^{(k)})$  in multiplicative form and  $\mathcal{H}$  in additive form; cf. (2.7) and (3.7). Left:  $\phi(t) = \sqrt{\alpha + t^2}$ . Right:  $\phi(t) = t^2/2$  if  $|t| \leq \alpha$ ,  $\phi(t) = \alpha|t| - \alpha^2/2$  if  $|t| > \alpha$ .

$\beta$	$\alpha$	Multiplicative	Additive	$\beta$	$\alpha$	Multiplicative	Additive
0.8	0.00025	58.68	25.34	10	0.025	51.37	13.83
0.8	0.0005	47.80	15.92	10	0.05	34.12	13.83
0.8	0.001	42.01	10.78	10	0.1	20.98	13.83
0.8	0.002	73.85	43.40	10	0.2	15.02	13.83
0.8	0.004	117.26	78.82	10	0.4	13.83	13.83
0.2	0.001	43.35	19.89	2.5	0.1	14.93	7.15
0.4	0.001	30.74	15.91	5	0.1	17.30	9.61
0.8	0.001	42.01	10.78	10	0.1	20.98	13.83
1.6	0.001	85.80	32.89	20	0.1	28.62	21.46
3.2	0.001	120.79	43.40	40	0.1	42.88	35.89

where  $\gamma$  is a normalization constant, and

$$A_{i,j} = \gamma_{i,j} e^{-0.1(i-j)^2}, \quad i, j = 1, 2, \dots, p,$$

where  $\gamma_{i,j} \in [0, 1]$  are random numbers. In the former case  $A$  is banded and in the latter  $A$  is not Toeplitz. All observations stated above have been corroborated.

We focused on the role of the constant  $c$  involved in the definition of  $\psi$  in the additive form of HQ regularization—see (3.1) and (3.2). The results presented in Figure 5.2 show that the convergence rate depends linearly on the value of  $c$ . This nicely corroborates our Remark 7 on the optimal choice for  $c$ .

**5.2. Two-dimensional images.** The original image and the data, degraded by blur and noise, are shown in Figure 5.3. We consider spatial-invariant blurring, in which case  $A$  is a Toeplitz-like matrix [23]. In the regularization term,  $G$  is the discretization matrix of the first-order differentiation operator. The matrix  $\mathcal{H}$  in the additive form can be diagonalized by the discrete cosine transform matrix. Thus the

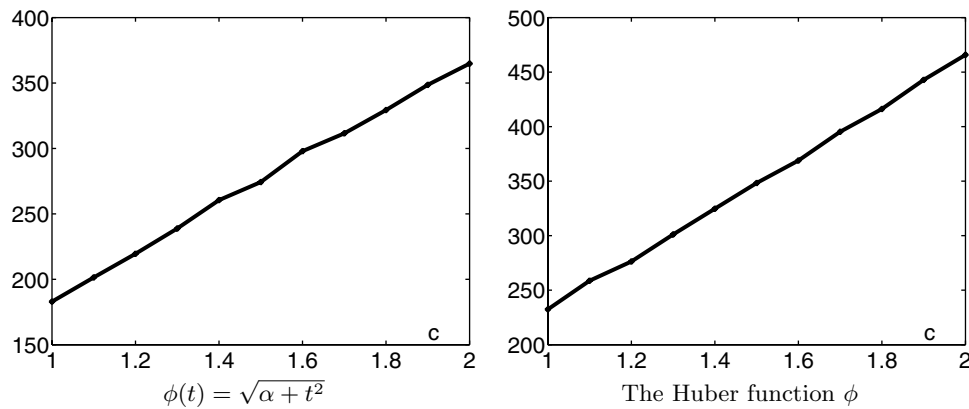


FIG. 5.2. Average number of iterations as a function of the values of  $c$  in the additive form: (left)  $\phi(t) = \sqrt{\alpha + t^2}$  ( $\alpha = 0.001$ ); (right)  $\phi(t) = t^2/2$  if  $|t| \leq \alpha$  and  $\phi(t) = \alpha|t| - \alpha^2/2$  if  $|t| > \alpha$  ( $\alpha = 0.1$ ).

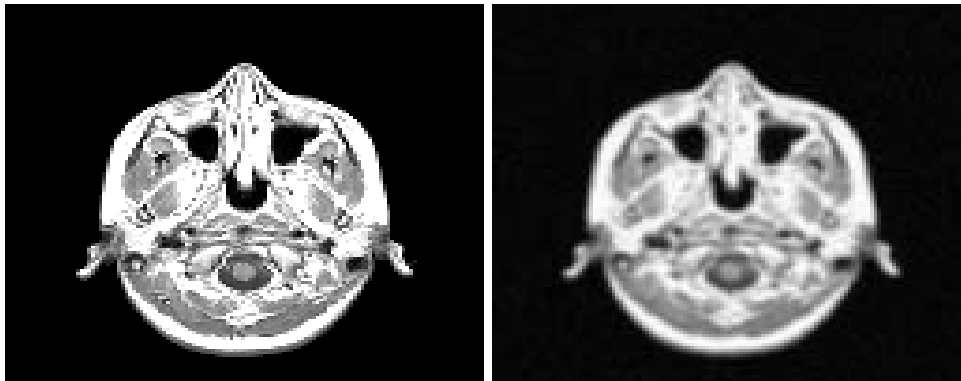


FIG. 5.3. Original image (left) and blurred and noisy image (right).

computational complexity to solve (4.6) at each iteration is  $O(p^2 \log p)$  operations for a  $p$ -by- $p$  image. In the multiplicative form, the matrix  $\mathcal{H}(b)$  cannot be diagonalized efficiently. At each iteration, we solve (4.5) using conjugate gradient methods. The computational complexity is  $O(p^2 \log p)$  per iteration.

The restored images using the Huber potential function are displayed in Figure 5.4 for the additive and the multiplicative forms with  $\alpha = \beta = 1$ . The stopping criterion of the HQ regularization iterations is  $\|x^{(k)} - x^{(k-1)}\|_2 < 1 \times 10^{-3} \|y\|_2$ . For the inner conjugate gradient iterations, we set the stopping criterion such that the  $\mathcal{H}(b)$ -norm of the solution error  $\|x^{(k)} - x_j^{(k)}\|_{\mathcal{H}(b)}$  is less than  $1 \times 10^{-4}$ , where  $x^{(k)}$  is the solution of the linear system (4.5) and  $x_j^{(k)}$  is the  $j$ th iterate of the conjugate gradient iteration [6]. Visually, two restored images using the additive and the multiplicative form are almost the same. From Table 5.7, we see that the multiplicative form is more effective than the additive form in terms of the number of iterations, the objective function values, and the relative errors. However, the differences are not significant. The computational times required by using the additive form are significantly less than those using the multiplicative form.

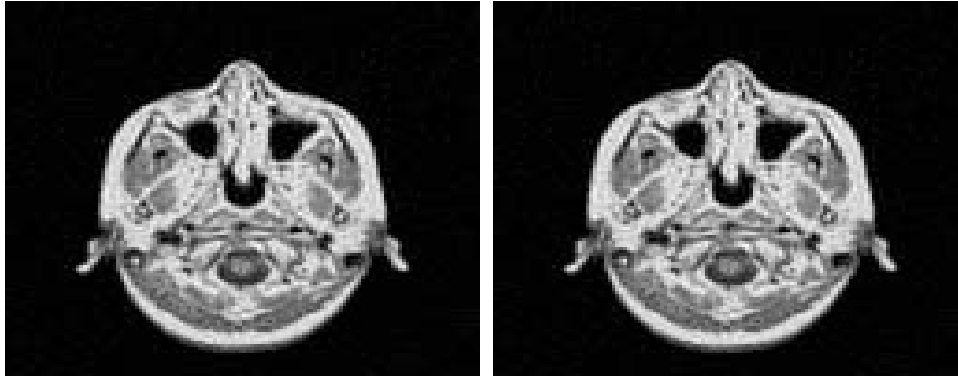


FIG. 5.4. Restored image using the multiplicative form (left) and using the additive form (right).

TABLE 5.7

The performance of the multiplicative form and the additive form for the Huber potential function.

$\beta = 1$ and $\alpha = 1$	Multiplicative	Additive
Number of iterations required	3	3
Number of inner conjugate gradient iterations required	79	–
Objective function value	457.1436	457.3545
Relative error of the restored image	0.1240	0.1241
CPU times required in seconds	16.23	0.54
$\beta = 1$ and $\alpha = 0.5$	Multiplicative	Additive
Number of iterations required	3	3
Number of inner conjugate gradient iterations required	78	–
Objective function value	456.4929	456.1902
Relative error of the restored image	0.1223	0.1237
CPU times required in seconds	16.02	0.54
$\beta = 1$ and $\alpha = 0.25$	Multiplicative	Additive
Number of iterations required	3	4
Number of inner conjugate gradient iterations required	79	–
Objective function value	430.4126	430.5825
Relative error of the restored image	0.1173	0.1229
CPU times required in seconds	16.20	0.71
$\beta = 0.5$ and $\alpha = 1$	Multiplicative	Additive
Number of iterations required	3	3
Number of inner conjugate gradient iterations required	80	–
Objective function value	325.2088	325.3879
Relative error of the restored image	0.1275	0.1339
CPU times required in seconds	16.01	0.54
$\beta = 2$ and $\alpha = 1$	Multiplicative	Additive
Number of iterations required	3	3
Number of inner conjugate gradient iterations required	78	–
Objective function value	682.3296	682.9721
Relative error of the restored image	0.1222	0.1224
CPU times required in seconds	16.06	0.54

**6. Concluding remarks.** We showed that the minimization of  $J$  using both the multiplicative and the additive form of HQ regularization amounts to finding the fixed point of a mapping involving a pertinent correction of the steepest descent direction. We presented a theoretical analysis of the convergence rate for each one of these minimization methods. More precisely, for both forms we derived an upper bound on the root-convergence factor. The bound for the multiplicative form was found to be smaller than the bound for the additive form. Our experimental results

show that minimization using the multiplicative form needs fewer iterations than the additive form, but that it requires a higher computation time. When applicable, the additive form is more attractive to use than the multiplicative form. Extensive experiments demonstrate that in our MATLAB implementation, minimization using HQ regularization is substantially faster (in terms of computational times) than standard routines, such as those in the MATLAB OPTIMIZATION TOOLBOX. These results must be qualified by the observation that MATLAB run times for the same operations can vary by orders of magnitude depending on the particular coding used.

**7. Appendix.** From the theory of convex conjugated functions [27, 18], given  $f : \mathbb{R} \rightarrow \mathbb{R}$ , its conjugated function  $g$  is defined by  $g(s) = \sup_{t \in \mathbb{R}} \{st - f(t)\}$  and is convex on  $\mathbb{R}$ . If  $f$  is continuous, convex and  $f \not\equiv \infty$  on  $\mathbb{R}$ , we have reciprocally  $\sup_{s \in \mathbb{R}} \{st - g(s)\} = f(t)$ .

**7.1. Functions  $\psi$  and  $\sigma$  for the multiplicative form.** We suppose that  $\phi$  satisfies all conditions given in (2.1). Define  $\theta : \mathbb{R} \rightarrow \mathbb{R}$  as

$$(7.1) \quad \theta(t) = \begin{cases} \phi(\sqrt{t}) & \text{if } t \geq 0, \\ -\infty & \text{if } t < 0. \end{cases}$$

By (2.1)(b), the function  $-\theta$  is convex on  $\mathbb{R}$ . Using convex conjugacy, if

$$(7.2) \quad \psi(s) = \sup_{t \in \mathbb{R}_+} \{ h(t, s) \}, \quad \text{where } h(t, s) = \left( -\frac{1}{2}s \right) t - (-\theta(t)),$$

we have reciprocally

$$(7.3) \quad \theta(t) = \inf_{s \in \mathbb{R}_+^*} \left\{ \frac{1}{2}st + \psi(s) \right\},$$

where the infimum is taken over  $\mathbb{R}_+^*$  because  $\psi(s) = +\infty$  if  $s \leq 0$ . Since  $t \in \mathbb{R}_+$  in (7.2) and (7.3), we can write  $t^2$  in place of  $t$ . Using that  $\theta(t^2) = \phi(t)$ , these expressions yield (2.2) and (2.3), respectively.

Since for any  $s > 0$  the function  $t \rightarrow h(t, s)$  is concave on  $\mathbb{R}_+$  by (2.1)(b) and satisfies  $\lim_{t \rightarrow \infty} h(t, s) = -\infty$  by (2.1)(f), it follows that  $h(\cdot, s)$  reaches its supremum for a finite  $t \in [0, \infty)$ . We determine  $\sigma$  by exhibiting all the pairs  $(\hat{t}, \hat{s})$  for which the supremum in (7.2) is reached, i.e.,  $\psi(\hat{s}) = h(\hat{t}, \hat{s})$ .

- For any  $\hat{s} \geq 2\theta'(0^+)$ , the function  $h(\cdot, \hat{s})$  reaches its maximum at  $\hat{t} = 0$  since  $\lim_{t \searrow 0} D_1 h(t, \hat{s}) \leq 0$  shows that  $h(\cdot, \hat{s})$  is decreasing on the right side of zero.
- Consider that  $\hat{s} \in (0, 2\theta'(0^+))$ . Since  $\lim_{t \searrow 0} D_1 h(t, \hat{s}) > 0$ , the function  $h(\cdot, \hat{s})$  increases on the right of zero. The supremum in (7.2) is then reached for a  $\hat{t} \in (0, \infty)$  such that  $D_1 h(\hat{t}, \hat{s}) = 0$ , hence  $\hat{s} = 2\theta'(\hat{t}) = \phi'(\sqrt{\hat{t}})/\sqrt{\hat{t}}$ .

Using  $\theta(t^2) = \phi(t)$  again shows that  $\sigma$  is as given in (2.5).

**7.2. Functions  $\psi$  and  $\sigma$  for the additive form.** We assume that  $\phi$  satisfies (3.1). Using (3.1)(b), convex conjugacy shows that if

$$(7.4) \quad \psi(s) + \frac{1}{2c}s^2 = \sup_{t \in \mathbb{R}} \{ h(t, s) \}, \quad \text{where } h(t, s) = ts - \left( \frac{c}{2}t^2 - \phi(t) \right),$$

we have reciprocally

$$\frac{c}{2}t^2 - \phi(t) = \sup_{s \in \mathbb{R}} \left( ts - \left( \psi(s) + \frac{1}{2c}s^2 \right) \right).$$

The equation in (7.4) is equivalent to (3.2) while the latter is equivalent to (3.3). Here again, the function  $\sigma$  is obtained by determining all the pairs  $(\hat{t}, \hat{s})$  which achieve the supremum in (7.4). Using (3.1)(b) and (3.1)(e), for any  $s \in \mathbb{R}$ , the function  $h(\cdot, s)$  is concave and satisfies  $\lim_{|t| \rightarrow \infty} h(t, s) = -\infty$ . It follows that for any  $\hat{s}$ , the supremum in (7.4) is reached for a finite  $\hat{t}$ . Since  $h$  is differentiable, any such  $\hat{t}$  satisfies  $D_1 h(\hat{t}, \hat{s}) = 0$ . Hence  $\hat{s} = \sigma(\hat{t})$  if  $\sigma$  reads as in (3.5).

## REFERENCES

- [1] R. ACAR AND C. VOGEL, *Analysis of bounded variation penalty methods for ill-posed problems*, IEEE Trans. Image Process., 10 (1994), pp. 1217–1229.
- [2] G. AUBERT AND P. KORNPBST, *Mathematical Problems in Image Processing*, Springer-Verlag, Berlin, 2002.
- [3] G. AUBERT AND L. VESE, *A variational method in image recovery*, SIAM J. Numer. Anal., 34 (1997), pp. 1948–1979.
- [4] M. BLACK AND A. RANGARAJAN, *On the unification of line processes, outlier rejection, and robust statistics with applications to early vision*, Internat. J. Computer Vision, 19 (1996), pp. 57–91.
- [5] C. BOUMAN AND K. SAUER, *A generalized Gaussian image model for edge-preserving MAP estimation*, IEEE Trans. Image Process., 2 (1993), pp. 296–310.
- [6] R. H. CHAN AND M. K. NG, *Conjugate gradient methods for Toeplitz systems*, SIAM Rev., 38 (1996), pp. 427–482.
- [7] P. CHARBONNIER, L. BLANC-FÉRAUD, G. AUBERT, AND M. BARLAUD, *Two deterministic half-quadratic regularization algorithms for computed imaging*, in Proceedings of IEEE ICIP, Vol. 2, 1994, pp. 168–172.
- [8] P. CHARBONNIER, L. BLANC-FÉRAUD, G. AUBERT, AND M. BARLAUD, *Deterministic edge-preserving regularization in computed imaging*, IEEE Trans. Image Process., 6 (1997), pp. 298–311.
- [9] P. G. CIARLET, *Introduction à l'analyse numérique matricielle et à l'optimisation*, Collection mathématiques appliquées pour la maîtrise, Masson, Paris, 1990.
- [10] L. D. COHEN, *Auxiliary variables and two-step iterative algorithms in computer vision problems*, J. Math. Imaging Vision, 6 (1996), pp. 376–386.
- [11] A. H. DELANEY AND Y. BRESLER, *Globally convergent edge-preserving regularized reconstruction: An application to limited-angle tomography*, IEEE Trans. Image Process., 7 (1998), pp. 204–221.
- [12] G. DEMOMENT, *Image reconstruction and restoration: Overview of common estimation structures and problems*, IEEE Trans. Acoustics Speech Signal Process., 37 (1989), pp. 2024–2036.
- [13] D. GEMAN AND G. REYNOLDS, *Constrained restoration and the recovery of discontinuities*, IEEE Trans. Pattern Anal. Machine Intelligence, 14 (1992), pp. 367–383.
- [14] D. GEMAN AND C. YANG, *Nonlinear image recovery with half-quadratic regularization*, IEEE Trans. Image Process., 4 (1995), pp. 932–946.
- [15] S. GEMAN AND D. GEMAN, *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*, IEEE Trans. Pattern Anal. Machine Intelligence, 6 (1984), pp. 721–741.
- [16] R. GLOWINSKI, J. LIONS, AND R. TRÉMOLIÈRES, *Analyse numérique des inéquations variationnelles*, Vol. 1, 1st ed., Dunod, Paris, 1976.
- [17] P. J. GREEN, *Bayesian reconstructions from emission tomography data using a modified EM algorithm*, IEEE Trans. Medical Imaging, 9 (1990), pp. 84–93.
- [18] J.-B. HIRIART-URRUTY AND C. LEMARÉCHAL, *Convex Analysis and Minimization Algorithms*, Vol. I, Springer-Verlag, Berlin, 1996.
- [19] J. IDIER, *Convex half-quadratic criteria and interacting auxiliary variables for image restoration*, IEEE Trans. Image Process., 10 (2001), pp. 1001–1009.
- [20] T. KAILATH AND A. H. SAYED, *Displacement structure: Theory and applications*, SIAM Rev., 37 (1995), pp. 297–386.
- [21] P. KORNPBST, R. DERICHE, AND G. AUBERT, *Image sequence analysis via partial differential equations*, J. Math. Imaging Vision, 11 (1999), pp. 5–26.
- [22] S. LI, *Markov Random Field Modeling in Computer Vision*, 1st ed., Springer-Verlag, New York, 1995.
- [23] M. NG, R. CHAN, T. CHAN, AND A. YIP, *Cosine transform preconditioners for high resolution image reconstruction*, Linear Algebra Appl., 316 (2000), pp. 89–104.

- [24] M. K. NG, R. H. CHAN, AND W.-C. TANG, *A fast algorithm for deblurring models with Neumann boundary conditions*, SIAM J. Sci. Comput., 21 (1999), pp. 851–866.
- [25] M. NIKOLOVA, *Local strong homogeneity of a regularized estimator*, SIAM J. Appl. Math., 61 (2000), pp. 633–658.
- [26] J. ORTEGA AND W. RHEINBOLDT, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.
- [27] R. T. ROCKAFELLAR, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
- [28] L. RUDIN, S. OSHER, AND E. FATEMI, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.
- [29] A. TARANTOLA, *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*, Elsevier–North Holland, Amsterdam, 1987.
- [30] S. TEBOUL, L. BLANC-FÉRAUD, G. AUBERT, AND M. BARLAUD, *Variational approach for edge-preserving regularization using coupled PDE's*, IEEE Trans. Image Process., 7 (1998), pp. 387–397.
- [31] A. TIKHONOV AND V. ARSEININ, *Solutions of Ill-Posed Problems*, Winston, Washington, DC, 1977.