HAL open science

Variable metric forward-backward algorithm for minimizing the sum of a differentiable function and a convex function

Emilie Chouzenoux, Jean-Christophe Pesquet, Audrey Repetti

▶ To cite this version:

Emilie Chouzenoux, Jean-Christophe Pesquet, Audrey Repetti. Variable metric forward-backward algorithm for minimizing the sum of a differentiable function and a convex function. Journal of Optimization Theory and Applications, 2014, 162 (1), pp.107-132. 10.1007/s10957-013-0465-7 . hal-00789970

HAL Id: hal-00789970
https://hal.science/hal-00789970

Submitted on 19 Feb 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Variable Metric Forward-Backward algorithm for minimizing the sum of a differentiable function and a convex function

Emilie Chouzenoux, Jean-Christophe Pesquet and Audrey Repetti *

Abstract

We consider the minimization of a function G defined on $\mathbb{R}^N$, which is the sum of a (non necessarily convex) differentiable function and a (non necessarily differentiable) convex function. Moreover, we assume that G satisfies the Kurdyka-Łojasiewicz property. Such a problem can be solved with the Forward-Backward algorithm. However, the latter algorithm may suffer from slow convergence. We propose an acceleration strategy based on the use of variable metrics and of the Majorize-Minimize principle. We give conditions under which the sequence generated by the resulting Variable Metric Forward-Backward algorithm converges to a critical point of G. Numerical results illustrate the performance of the proposed algorithm in an image reconstruction application.

1 Introduction

We consider the following problem:

$$\text{Find} \quad \hat{x} \in \text{Argmin } G, \tag{1}$$

where $G : \mathbb{R}^N \to (-\infty, +\infty]$ is a coercive (i.e. $\lim_{\|x\|\to+\infty} G(x) = +\infty$) function. In addition, we assume that G can be split as

$$G = F + R, \tag{2}$$

where F is a differentiable function and R is a proper lower semicontinuous convex function. A standard approach in this context consists of using the proximal Forward-Backward (FB) algorithm [1, 2], which generates a sequence $(x_k)_{k\in\mathbb{N}}$ by the following iterations:

$$\begin{array}{l} x_0 \in \mathbb{R}^N \\ \text{For } k = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} y_k = \text{prox}_{\gamma_k R}(x_k - \gamma_k \nabla F(x_k)), \\ x_{k+1} = x_k + \lambda_k (y_k - x_k), \end{array} \right. \end{array} \tag{3}$$

where, for every $k \in \mathbb{N}$, $(\gamma_k, \lambda_k) \in (0, +\infty)^2$, $\nabla F(x_k)$ is the gradient of F at x_k, and $\text{prox}_{\gamma_k R}$ denotes the so-called *proximity operator* of $\gamma_k R$. Let us introduce the weighted norm:

$$(\forall x \in \mathbb{R}^N) \qquad \|x\|_U = \left(x^\top U x\right)^{1/2}, \tag{4}$$

where $U \in \mathbb{R}^{N\times N}$ is some symmetric positive definite matrix. Then, the proximity operator ([3, Sec. XV.4], [4]) is defined as follows:

*E. Chouzenoux, J.-C. Pesquet and A. Repetti are with the Université Paris-Est, LIGM, CNRS-UMR 8049, 77454 Marne-la-Vallée Cedex 2, France. Phone: +33 1 60 95 72 88, E-mail: {emilie.chouzenoux,jean-christophe.pesquet,audrey.repetti}@univ-paris-est.fr.

Definition 1.1. *Let $\psi\colon \mathbb{R}^N \to (-\infty,+\infty]$ be a proper, lower semicontinuous, convex function, let $U \in \mathbb{R}^{N\times N}$ be a symmetric positive definite matrix, and let $x \in \mathbb{R}^N$. The proximity operator of ψ at x relative to the metric induced by U is the unique minimizer of $\psi + \frac{1}{2}\|\cdot - x\|_U^2$, and it is denoted by $\operatorname{prox}_{U,\psi}(x)$. If U is equal to I_N, the identity matrix of $\mathbb{R}^{N\times N}$, then $\operatorname{prox}_\psi \equiv \operatorname{prox}_{\mathrm{I}_N,\psi}$ is the proximity operator originally defined in [5].*

When F is a convex function having an L-Lipschitzian gradient with $L > 0$, the sequence $(x_k)_{k\in\mathbb{N}}$ given by (3) converges to a solution to Problem (1), under the following assumptions on the step lengths $(\gamma_k)_{k\in\mathbb{N}}$ and the relaxation parameters $(\lambda_k)_{k\in\mathbb{N}}$ [6, 7]:

- $0 < \inf_{l\in\mathbb{N}} \gamma_l \leq \sup_{l\in\mathbb{N}} \gamma_l < 2L^{-1}$,
- $(\forall k \in \mathbb{N})\ 0 < \inf_{l\in\mathbb{N}} \lambda_l \leq \lambda_k \leq 1$.

The convergence properties of the FB algorithm have been recently extended to the case of non-convex functions F and R in [8,9] when $\lambda_k \equiv 1$. The convergence results in [8,9] mainly rely on the assumption that the objective function G satisfies the Kurdyka-Łojasiewicz (KL) inequality [10]. The interesting point is that this inequality holds for a wide class of functions. In particular, it is satisfied by real analytic functions, semi-algebraic functions and many others [10–13].

In the case of large scale optimization problems such as those encountered in image restoration, one major concern is to find an optimization algorithm able to deliver reliable numerical solutions in a reasonable time. The FB algorithm is characterized by a low computational cost per iteration. However, as many first-order minimization methods, it may suffer from slow convergence [1]. Two families of acceleration strategies can be distinguished in the literature. The first approach, adopted for example in the FISTA method, relies on subspace acceleration [14–18]. In such methods, the convergence rate is improved by using informations from previous iterates for the construction of the new estimate. Another efficient way to accelerate the convergence of the FB algorithm is based on a variable metric strategy [1, 19–24]. The underlying metric of FB is modified at each iteration, giving rise to the so-called Variable Metric Forward-Backward (VMFB) algorithm:

$$
\begin{array}{l}
x_0 \in \mathbb{R}^N \\
\text{For } k = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
y_k = \operatorname{prox}_{\gamma_k^{-1} A_k, R}(x_k - \gamma_k A_k^{-1} \nabla F(x_k)), \\
x_{k+1} = x_k + \lambda_k (y_k - x_k),
\end{array}
\right.
\end{array}
\tag{5}
$$

where, for every $k \in \mathbb{N}$, $A_k \in \mathbb{R}^{N\times N}$ is a symmetric positive definite matrix. On the one hand, when A_k is the identity matrix, the FB algorithm (3) is recovered. On the other hand, when $R \equiv 0$, Algorithm (5) corresponds to a preconditioned gradient algorithm. If F is a twice differentiable convex function, the preconditioning matrix A_k is then usually chosen as an approximation of the Hessian of F at x_k. This amounts to performing a change of variables leading to a function whose Hessian has more clustered eigenvalues ([25, Sec.1.3.], [26, 27]). A convergence analysis of Algorithm (5) is provided in [23], under the assumptions that F and R are convex functions and that there exists a positive bounded sequence $(\eta_k)_{k\in\mathbb{N}}$ such that, for every $k \in \mathbb{N}$,

$$
(\forall x \in \mathbb{R}^N) \quad (1 + \eta_k)(x^\top A_{k+1} x) \geq x^\top A_k x. \tag{6}
$$

More specific convergence results are available in the literature in the particular case when R is the indicator function of a convex set [28–30]. However, in the aforementioned works, the convergence study is limited to the case of a *convex* smooth function F. As pointed out in [24], for an arbitrary matrix A_k, the proximal step in (5) is not explicit in general, and sub-iterations are thus needed. Our contribution in this paper is to derive an inexact version of the VMFB algorithm, based on majorize-minimize arguments. The convergence of this algorithm is established for a non necessarily convex smooth function F.

The rest of the paper is organized as follows: Section 2 introduces the assumptions made in the paper and presents the proposed inexact VMFB strategy. In Section 3, we investigate the convergence properties of the proposed algorithm. Finally, Section 4 provides some numerical results and a discussion of the algorithm performance by means of experiments concerning image recovery problems.

2 Proposed optimization method

2.1 Background and assumptions

Let us first recall some definitions and notations that will be used throughout the paper.

Definition 2.1. *Let ψ be a function from $\mathbb{R}^N$ to $(-\infty,+\infty]$. The* domain *of ψ is* $\operatorname{dom}\psi = \left\{x \in \mathbb{R}^N \mid \psi(x) < +\infty\right\}$. *Function ψ is* proper *if $\operatorname{dom}\psi$ is nonempty. The* level set *of ψ at height $\delta \in \mathbb{R}$ is* $\operatorname{lev}_{\leq\delta}\psi = \{x \in \mathbb{R}^N \mid \psi(x) \leq \delta\}$.

Definition 2.2. [31, Def. 8.3], [32, Sec.1.3] *Let $\psi\colon \mathbb{R}^N \to (-\infty,+\infty]$ be a proper function and let $x \in \operatorname{dom}\psi$. The* Fréchet sub-differential *of ψ at x is the following set:*

$$\hat{\partial}\psi(x) = \left\{ t \in \mathbb{R}^N \mid \liminf_{\substack{y\to x\\ y\neq x}} \frac{1}{\|x-y\|}\left(\psi(y) - \psi(x) - (y-x)^\top t\right) \geq 0 \right\}.$$

If $x \notin \operatorname{dom}\psi$, then $\hat{\partial}\psi(x) = \varnothing$.
The sub-differential *of ψ at x is defined as*

$$\partial\psi(x) = \left\{\hat{t} \in \mathbb{R}^N \mid \exists y_k \to x,\ \psi(y_k) \to \psi(x),\ t_k \in \hat{\partial}\psi(y_k) \to \hat{t}\right\}.$$

Recall that a necessary condition for $x \in \mathbb{R}^N$ to be a minimizer of ψ is that x is a *critical point* of ψ, i.e. $0 \in \partial\psi(x)$. Moreover, if ψ is convex, this condition is sufficient.

Remark 2.1. *Definition 2.2 implies that $\partial\psi$ is closed* [9]. *More precisely, we have the following property:*
Let $(y_k,\hat{t}_k)_{k\in\mathbb{N}}$ be a sequence of $\operatorname{Graph}\partial\psi = \left\{(x,\hat{t}) \in \mathbb{R}^N \times \mathbb{R}^N \mid \hat{t} \in \partial\psi(x)\right\}$. *If $\left(y_k,\hat{t}_k\right)$ converges to $\left(x,\hat{t}\right)$ and $\psi(y_k)$ converges to $\psi(x)$, then $(x,\hat{t}) \in \operatorname{Graph}\partial\psi$.*

Let us introduce our notation for linear operators. $\mathcal{S}_N$ denotes the space of symmetric matrices of $\mathbb{R}^{N\times N}$. The Loewner partial ordering on $\mathbb{R}^{N\times N}$ is defined as

$$(\forall U_1 \in \mathbb{R}^{N\times N})(\forall U_2 \in \mathbb{R}^{N\times N}) \quad U_1 \succcurlyeq U_2 \quad \Leftrightarrow \quad (\forall x \in \mathbb{R}^N) \quad x^\top U_1 x \geq x^\top U_2 x.$$

In the remainder of this work, we will focus on functions F and R satisfying the following assumptions:

Assumption 2.1.

(i) $R\colon \mathbb{R}^N \to (-\infty,+\infty]$ *is proper, lower semicontinuous and convex, and its restriction to its domain is continuous.*

(ii) $F\colon \mathbb{R}^N \to \mathbb{R}$ *is differentiable. Moreover,* F *has an* L*-Lipschitzian gradient on* $\operatorname{dom} R$ *where* $L > 0$*, i.e.*

$$\big(\forall (x,y) \in (\operatorname{dom} R)^2\big) \quad \|\nabla F(x) - \nabla F(y)\| \leq L\|x-y\|.$$

(iii) G *as defined by* (2) *is coercive.*

Some comments on these assumptions which will be useful in the rest of the paper are made below.

Remark 2.2.

(i) *Assumption 2.1(ii) is weaker than the assumption of Lipschitz differentiability of* F *usually adopted to prove the convergence of the FB algorithm [6, 9]. In particular, if* $\operatorname{dom} R$ *is compact and* F *is twice continuously differentiable, Assumption 2.1(ii) holds.*

(ii) *According to Assumption 2.1(ii),* $\operatorname{dom} R \subset \operatorname{dom} F$*. Then, as a consequence of Assumption 2.1(i),* $\operatorname{dom} G = \operatorname{dom} R$ *is a nonempty convex set.*

(iii) *Under Assumption 2.1,* G *is proper and lower semicontinuous, and its restriction to its domain is continuous. Hence, due to the coercivity of* G*, for every* $x \in \operatorname{dom} R$*,* $\operatorname{lev}_{\leq G(x)} G$ *is a compact set. Moreover, the set of minimizers of* G *is nonempty and compact.*

Assumption 2.2.
Function G *satisfies the Kurdyka-Łojasiewicz inequality i.e., for every* $\xi \in \mathbb{R}$*, and, for every bounded subset* E *of* $\mathbb{R}^N$*, there exist three constants* $\kappa > 0$*,* $\zeta > 0$ *and* $\theta \in [0,1)$ *such that*

$$\big(\forall t(x) \in \partial G(x)\big) \qquad \|t(x)\| \geq \kappa |G(x) - \xi|^{\theta},$$

for every $x \in E$ *such that* $|G(x) - \xi| \leq \zeta$ *(with the convention* $0^0 = 0$*).*

Note that other forms of the KL inequality can be found in the literature [12, 33].

2.2 Majorize-Minimize metric

Some matrices serving to define some appropriate variable metric will play a central role in the algorithm proposed in this work. More specifically, let $(x_k)_{k\in\mathbb{N}}$ be some given sequence of $\operatorname{dom} R$ and let $(A_k)_{k\in\mathbb{N}}$ be a sequence of matrices of $\mathcal{S}_N$ that fulfill the following so-called *majorization conditions*:

Assumption 2.3.

(i) *For every* $k \in \mathbb{N}$*, the quadratic function defined as*

$$(\forall x \in \mathbb{R}^N) \quad Q(x, x_k) = F(x_k) + (x - x_k)^\top \nabla F(x_k) + \frac{1}{2}(x - x_k)^\top A_k (x - x_k),$$

is a majorant function *of* F *at* x_k *on* $\operatorname{dom} R$*, i.e.,*

$$(\forall x \in \operatorname{dom} R) \quad F(x) \leq Q(x, x_k).$$

(ii) *There exists* $(\underline{\nu}, \overline{\nu}) \in (0, +\infty)^2$ *such that*

$$(\forall k \in \mathbb{N}) \quad \underline{\nu}\, \mathrm{I}_N \preceq A_k \preceq \overline{\nu}\, \mathrm{I}_N\,.$$

The following lemma points out the existence of such a sequence of matrices:

Lemma 2.1. *Suppose that Assumption 2.1 holds and, for every $k \in \mathbb{N}$, let A_k be equal to $L\,\mathrm{I}_N$, where $L > 0$ is the Lipschitz constant of ∇F. Then, $(A_k)_{k\in\mathbb{N}}$ satisfies Assumption 2.3 with $\underline{\nu} = \overline{\nu} = L$.*

Proof. Under Assumption 2.1, $\operatorname{dom} R$ is a convex set and, since F is Lipschitz-differentiable on $\operatorname{dom} R$, the *Descent Lemma* [25, Prop.A.24] applies, yielding:

$$\big(\forall (x,y) \in (\operatorname{dom} R)^2\big) \quad F(x) \le F(y) + (x-y)^\top \nabla F(y) + \frac{L}{2}\|x-y\|^2.$$

Consequently, when $(\forall k \in \mathbb{N})$ $A_k = L\,\mathrm{I}_N$, Assumption 2.3(i) is satisfied while Assumption 2.3(ii) obviously holds.. $\square$

Although the above lemma provides a simple choice for sequence $(A_k)_{k\in\mathbb{N}}$, it is worth noticing that other choices have been investigated in the literature [34, 35] for some subclasses of functions F.

2.3 Inexact Variable Metric Forward-Backward algorithm

In general, the proximity operator relative to an arbitrary metric does not have a closed form expression. To circumvent this difficulty, we propose to solve Problem 1 by introducing the following inexact version of the Variable Metric FB method:

$$
\begin{array}{l}
\tau \in (0,+\infty), x_0 \in \operatorname{dom} R\\
\text{For } k = 0,1,\ldots\\
\left\lfloor
\begin{array}{ll}
\text{Find } y_k \in \mathbb{R}^N \text{ and } r(y_k) \in \partial R(y_k) \text{ such that} & \\
\quad R(y_k) + (y_k - x_k)^\top \nabla F(x_k) + \gamma_k^{-1}\|y_k - x_k\|^2_{A_k} \le R(x_k), & \text{(7a)}\\
\quad \|\nabla F(x_k) + r(y_k)\| \le \tau \|y_k - x_k\|_{A_k}, & \text{(7b)}\\
x_{k+1} = (1-\lambda_k)x_k + \lambda_k y_k, & \text{(7c)}
\end{array}
\right.
\end{array}
$$

where $(A_k)_{k\in\mathbb{N}}$ is a sequence of $\mathcal{S}_N$ associated with $(x_k)_{k\in\mathbb{N}}$ for which Assumption 2.3 holds. In addition, $(\gamma_k)_{k\in\mathbb{N}}$ and $(\lambda_k)_{k\in\mathbb{N}}$ are sequences of nonnegative reals satisfying the following two assumptions:

Assumption 2.4.

(i) *There exists $(\underline{\eta}, \overline{\eta}) \in (0,+\infty)^2$ such that, for every $k \in \mathbb{N}$, $\underline{\eta} \le \gamma_k\lambda_k \le 2 - \overline{\eta}$.*

(ii) *There exists $\underline{\lambda} \in (0,+\infty)$ such that, for every $k \in \mathbb{N}$, $\underline{\lambda} \le \lambda_k \le 1$.*

Assumption 2.5.
There exists $\underline{\alpha} \in (0,1]$ such that, for every $k \in \mathbb{N}$,

$$G(x_{k+1}) \le (1-\underline{\alpha})G(x_k) + \underline{\alpha}G(y_k).$$

Remark 2.3.

(i) *If, for every $k \in \mathbb{N}$, one chooses x_{k+1} such that $G(x_{k+1}) \le G(y_k)$, then Assumption 2.5 holds for $\underline{\alpha} = 1$.*

(ii) *$(\lambda_k)_{k\in\mathbb{N}}$ can always be chosen such that Assumption 2.5 is satisfied (by taking for every $k \in \mathbb{N}$, $\lambda_k = \underline{\alpha} = 1$).*

(iii) *Under Assumption 2.1(i),* $\operatorname{dom} R$ *is convex. Hence, for every* $k \in \mathbb{N}$, *both* x_k *and* y_k *belong to* $\operatorname{dom} R$.

(iv) *As already mentioned, Algorithm* (7) *can be viewed as an inexact version of Algorithm* (5)*. Indeed, let* $(x_k)_{k\in\mathbb{N}}$ *and* $(y_k)_{k\in\mathbb{N}}$ *be sequences generated by Algorithm* (5)*. On the one hand, due to the variational characterization of the proximity operator and the convexity of* R*, for every* $k \in \mathbb{N}$*, there exists* $r(y_k) \in \partial R(y_k)$ *such that*

$$\begin{cases} r(y_k) = -\nabla F(x_k) + \gamma_k^{-1} A_k(x_k - y_k) \\ (y_k - x_k)^\top r(y_k) \geq R(y_k) - R(x_k), \end{cases} \tag{8}$$

which yields

$$R(y_k) + (y_k - x_k)^\top \nabla F(x_k) + \gamma_k^{-1} \|y_k - x_k\|^2_{A_k} \leq R(x_k).$$

So the sufficient-decrease condition (7a) *holds. On the other hand, let us assume that Assumption 2.4 holds. According to* (8)*, we have*

$$\|\nabla F(x_k) + r(y_k)\| = \gamma_k^{-1} \|A_k(y_k - x_k)\| \leq \underline{\eta}^{-1} \sqrt{\overline{\nu}} \|y_k - x_k\|_{A_k},$$

which is the inexact optimality condition (7b) *with* $\tau = \underline{\eta}^{-1}\sqrt{\overline{\nu}}$.

3 Convergence analysis

3.1 Descent properties

The present section gathers some technical results concerning the behaviour of the sequences $\big(G(y_k)\big)_{k\in\mathbb{N}}$ and $\big(G(x_k)\big)_{k\in\mathbb{N}}$ generated by Algorithm (7), which will be used to prove the convergence of the proposed algorithm.

Lemma 3.1. *Under Assumptions 2.1, 2.3 and 2.4, there exists* $\mu_1 \in (0, +\infty)$ *such that for every* $k \in \mathbb{N}$,

$$G(x_{k+1}) \leq G(x_k) - \frac{\mu_1}{2} \|x_{k+1} - x_k\|^2 \tag{9}$$

$$\leq G(x_k) - \underline{\lambda}^2 \frac{\mu_1}{2} \|y_k - x_k\|^2. \tag{10}$$

Proof. For every $k \in \mathbb{N}$, the update equation (7c) yields

$$G(x_{k+1}) = F(x_{k+1}) + R\left((1 - \lambda_k)x_k + \lambda_k y_k\right).$$

The convexity of R and Assumption 2.3(i) allow us to deduce that

$$\begin{aligned} G(x_{k+1}) &\leq F(x_{k+1}) + (1 - \lambda_k)R(x_k) + \lambda_k R(y_k) \\ &\leq F(x_k) + (x_{k+1} - x_k)^\top \nabla F(x_k) + \frac{1}{2}\|x_{k+1} - x_k\|^2_{A_k} \\ &\quad + (1 - \lambda_k)R(x_k) + \lambda_k R(y_k). \end{aligned} \tag{11}$$

In addition, according to (7c),

$$x_{k+1} - x_k = \lambda_k (y_k - x_k). \tag{12}$$

Using (7a) and (12) leads to the following inequality:

$$(x_{k+1} - x_k)^\top \nabla F(x_k) \leq -\gamma_k^{-1} \lambda_k^{-1} \|x_{k+1} - x_k\|^2_{A_k} + \lambda_k \big(R(x_k) - R(y_k)\big). \tag{13}$$

Therefore, by combining (11) and (13), we obtain

$$\begin{aligned} G(x_{k+1}) &\leq G(x_k) - (\gamma_k^{-1}\lambda_k^{-1} - \frac{1}{2})\|x_{k+1} - x_k\|^2_{A_k} \\ &\leq G(x_k) - \frac{1}{2}\frac{\overline{\eta}}{2-\overline{\eta}}\|x_{k+1} - x_k\|^2_{A_k}, \end{aligned}$$

where the last inequality follows from Assumption 2.4(i). Then, the lower bound in Assumption 2.3(ii) allows us to derive (9) by setting $\mu_1 = \dfrac{\underline{\nu}\overline{\eta}}{2-\overline{\eta}} > 0$. Inequality (10) results from (12) and Assumption 2.4(ii). □

As a consequence of the above lemma, Assumption 2.5 can be reexpressed in a different form:

Corollary 3.1. *Let $\underline{\alpha} \in (0, 1]$ and let $k \in \mathbb{N}$. Under Assumptions 2.1, 2.3 and 2.4, Assumption 2.5 is satisfied if and only if there exists $\alpha_k \in [\underline{\alpha}, 1]$ such that*

$$G(x_{k+1}) \leq (1 - \alpha_k)G(x_k) + \alpha_k G(y_k). \tag{14}$$

Proof. Under Assumption 2.5, (14) holds if we take, for every $k \in \mathbb{N}$, $\alpha_k = \underline{\alpha}$. Conversely, (14) is equivalent to

$$\alpha_k(G(x_k) - G(y_k)) \leq G(x_k) - G(x_{k+1}). \tag{15}$$

If the above inequality holds with $\alpha_k \in [\underline{\alpha}, 1]$, then two cases may arise:

(i) *Case when $G(x_k) \leq G(y_k)$.* From Lemma 3.1 we have $G(x_{k+1}) \leq G(x_k)$. Thus,

$$\underline{\alpha}(G(x_k) - G(y_k)) \leq 0 \leq G(x_k) - G(x_{k+1}).$$

(ii) *Case when $G(x_k) \geq G(y_k)$.* Then, (15) yields

$$\underline{\alpha}(G(x_k) - G(y_k)) \leq \alpha_k(G(x_k) - G(y_k)) \leq G(x_k) - G(x_{k+1}).$$

This shows that, if (14) holds, then Assumption 2.5 is satisfied. □

When G satisfies some convexity property, we recover standard assumptions on the relaxation parameter as shown below.

Corollary 3.2. *Under Assumptions 2.1, 2.3 and 2.4, if G is convex on $[x_k, y_k]$ for every $k \in \mathbb{N}$, then Assumption 2.5 holds.*

Proof. According to (7c), we have

$$(\forall k \in \mathbb{N}) \qquad G(x_{k+1}) = G\big((1 - \lambda_k)x_k + \lambda_k y_k\big),$$

where $\lambda_k \in (0, 1]$. If G is convex on $[x_k, y_k]$ for every $k \in \mathbb{N}$, then

$$G(x_{k+1}) \leq (1 - \lambda_k)G(x_k) + \lambda_k G(y_k).$$

Using Corollary 3.1 and the fact that, for every $k \in \mathbb{N}$, λ_k is lower-bounded by $\underline{\lambda} > 0$, we conclude that Assumption 2.5 is satisfied. □

The next result will allow us to evaluate the variations of G when going from x_k to y_k at each iteration k of Algorithm (7).

Lemma 3.2. *Under Assumptions 2.1, 2.3 and 2.4, there exists $\mu_2 \in \mathbb{R}$ such that*

$$(\forall k \in \mathbb{N}) \quad G(y_k) \leq G(x_k) - \mu_2 \|y_k - x_k\|^2.$$

Proof. According to Assumption 2.3(i) and (7a), we have

$$F(y_k) \leq Q(y_k, x_k) \leq F(x_k) + R(x_k) - R(y_k) - (\gamma_k^{-1} - \frac{1}{2})\|y_k - x_k\|^2_{A_k},$$

which, by using Assumption 2.4, yields,

$$G(y_k) \leq G(x_k) - (\frac{\underline{\lambda}}{2 - \overline{\eta}} - \frac{1}{2})\|y_k - x_k\|^2_{A_k}.$$

The result then follows from Assumption 2.3(ii) by setting

$$\mu_2 = \begin{cases} \underline{\nu}(\frac{\underline{\lambda}}{2 - \overline{\eta}} - \frac{1}{2}) & \text{if } 2\underline{\lambda} + \overline{\eta} \geq 2 \\ \overline{\nu}(\frac{\underline{\lambda}}{2 - \overline{\eta}} - \frac{1}{2}) & \text{otherwise.} \end{cases}$$

□

3.2 Convergence result

Our convergence proof hinges upon the following preliminary result:

Lemma 3.3. *Let $(u_k)_{k\in\mathbb{N}}$, $(g_k)_{k\in\mathbb{N}}$, $(g'_k)_{k\in\mathbb{N}}$ and $(\Delta_k)_{k\in\mathbb{N}}$ be sequences of nonnegative reals and let $\theta \in (0, 1)$. Assume that*

(i) *For every $k \in \mathbb{N}$, $u_k^2 \leq g_k^\theta \Delta_k$.*

(ii) *$(\Delta_k)_{k\in\mathbb{N}}$ is summable.*

(iii) *For every $k \in \mathbb{N}$, $g_{k+1} \leq (1 - \underline{\alpha})g_k + g'_k$ where $\underline{\alpha} \in (0, 1]$.*

(iv) *For every $k \geq k^*$, $(g'_k)^\theta \leq \beta u_k$ where $\beta > 0$ and $k^* \in \mathbb{N}$.*

Then, $(u_k)_{k\in\mathbb{N}}$ is a summable sequence.

Proof. According to (iii), for every $k \in \mathbb{N}$,

$$g_{k+1}^\theta \leq (1 - \underline{\alpha})^\theta g_k^\theta + (g'_k)^\theta.$$

Assumption (iv) then yields

$$(\forall k \geq k^*) \qquad g_{k+1}^\theta \leq (1 - \underline{\alpha})^\theta g_k^\theta + \beta u_k,$$

which implies that, for every $K > k^*$,

$$\begin{aligned} &\sum_{k=k^*+1}^{K} g_k^\theta \leq (1 - \underline{\alpha})^\theta \sum_{k=k^*}^{K-1} g_k^\theta + \beta \sum_{k=k^*}^{K-1} u_k \\ \Leftrightarrow \quad &\big(1 - (1 - \underline{\alpha})^\theta\big) \sum_{k=k^*}^{K-1} g_k^\theta \leq g_{k^*}^\theta - g_K^\theta + \beta \sum_{k=k^*}^{K-1} u_k. \end{aligned} \tag{16}$$

On the other hand, (i) can be rewritten as

$$(\forall k \in \mathbb{N}) \qquad u_k^2 \leq \Big(\beta^{-1}\big(1-(1-\underline{\alpha})^\theta\big)g_k^\theta\Big)\Big(\beta\big(1-(1-\underline{\alpha})^\theta\big)^{-1}\Delta_k\Big).$$

By using now the inequality $\big(\forall (v,v') \in [0,+\infty)^2\big)$ $\sqrt{vv'} \leq (v+v')/2$ and since, for every $k \in \mathbb{N}$, $u_k \geq 0$, we get

$$(\forall k \in \mathbb{N}) \qquad u_k \leq \frac{1}{2}\Big(\beta^{-1}\big(1-(1-\underline{\alpha})^\theta\big)g_k^\theta + \beta\big(1-(1-\underline{\alpha})^\theta\big)^{-1}\Delta_k\Big). \tag{17}$$

We deduce from (16) and (17) that, for every $K > k^*$,

$$\sum_{k=k^*}^{K-1} u_k \leq \frac{1}{2}\Big(\sum_{k=k^*}^{K-1} u_k + \beta^{-1}(g_{k^*}^\theta - g_K^\theta) + \beta\big(1-(1-\underline{\alpha})^\theta\big)^{-1}\sum_{k=k^*}^{K-1}\Delta_k\Big)$$
$$\Rightarrow \sum_{k=k^*}^{K-1} u_k \leq \beta^{-1}g_{k^*}^\theta + \beta\big(1-(1-\underline{\alpha})^\theta\big)^{-1}\sum_{k=k^*}^{K-1}\Delta_k.$$

The summability of $(u_k)_{k\in\mathbb{N}}$ then follows from (ii). □

Our main result concerning the asymptotic behaviour of Algorithm (7) can now be stated:

Theorem 3.1. *Under Assumptions 2.1-2.5, the following hold.*

(i) *The sequences $(x_k)_{k\in\mathbb{N}}$ and $(y_k)_{k\in\mathbb{N}}$ defined by (7) both converge to a critical point $\hat{x}$ of G.*

(ii) *These sequences have a finite length in the sense that*

$$\sum_{k=0}^{+\infty}\|x_{k+1}-x_k\| < +\infty \quad \textit{and} \quad \sum_{k=0}^{+\infty}\|y_{k+1}-y_k\| < +\infty.$$

(iii) *$\big(G(x_k)\big)_{k\in\mathbb{N}}$ and $\big(G(y_k)\big)_{k\in\mathbb{N}}$ are sequences converging to $G(\hat{x})$. Moreover, $\big(G(x_k)\big)_{k\in\mathbb{N}}$ is a nonincreasing sequence.*

Proof. According to Lemma 3.1, we have

$$(\forall k \in \mathbb{N}) \quad G(x_{k+1}) \leq G(x_k),$$

thus, $(G(x_k))_{k\in\mathbb{N}}$ is a nonincreasing sequence. In addition, by Remark 2.2(iii) and Remark 2.3(iii), the sequence $\big(x_k\big)_{k\in\mathbb{N}}$ belongs to a compact subset E of $\operatorname{lev}_{\leq G(x_0)} G \subset \operatorname{dom} R$ and G is lower bounded. Thus, $\big(G(x_k)\big)_{k\in\mathbb{N}}$ converges to a real ξ, and $\big(G(x_k)-\xi\big)_{k\in\mathbb{N}}$ is a nonnegative sequence converging to 0.
Moreover, by invoking again Lemma 3.1, we have

$$(\forall k \in \mathbb{N}) \quad \underline{\lambda}^2\frac{\mu_1}{2}\|y_k - x_k\|^2 \leq (G(x_k)-\xi) - (G(x_{k+1})-\xi). \tag{18}$$

Hence, the sequence $(y_k - x_k)_{k\in\mathbb{N}}$ converges to 0.
On the other hand, Assumption 2.5 implies that, for every $k \in \mathbb{N}$,

$$G(x_{k+1}) - \xi \leq (1-\underline{\alpha})(G(x_k)-\xi) + \underline{\alpha}(G(y_k)-\xi).$$

Then, combining the last inequality with Lemma 3.2, we obtain that, for every $k \in \mathbb{N}$,

$$\underline{\alpha}^{-1}\big(G(x_{k+1}) - \xi - (1-\underline{\alpha})(G(x_k) - \xi)\big) \le G(y_k) - \xi \le G(x_k) - \xi - \mu_2 \|y_k - x_k\|^2.$$

Thus, since $(y_k - x_k)_{k\in\mathbb{N}}$ and $(G(x_k) - \xi)_{k\in\mathbb{N}}$ both converge to 0, the sequence $(G(y_k))_{k\in\mathbb{N}}$ converges to ξ.

Let us come back to (18) and let us apply to the convex function $\psi\colon [0,+\infty) \to [0,+\infty)\colon u \mapsto u^{\frac{1}{1-\theta}}$, with $\theta \in [0,1)$, the gradient inequality

$$(\forall (u,v) \in [0,+\infty)^2) \quad \psi(u) - \psi(v) \le \dot{\psi}(u)(u-v),$$

which, after a change of variables, can be rewritten as

$$(\forall (u,v) \in [0,+\infty)^2) \quad u - v \le (1-\theta)^{-1} u^\theta (u^{1-\theta} - v^{1-\theta}).$$

Using the latter inequality with $u = G(x_k) - \xi$ and $v = G(x_{k+1}) - \xi$ leads to

$$(\forall k \in \mathbb{N}) \quad \big(G(x_k) - \xi\big) - \big(G(x_{k+1}) - \xi\big) \le (1-\theta)^{-1}\big(G(x_k) - \xi\big)^\theta \Delta'_k,$$

where

$$(\forall k \in \mathbb{N}) \quad \Delta'_k = \big(G(x_k) - \xi\big)^{1-\theta} - \big(G(x_{k+1}) - \xi\big)^{1-\theta}.$$

Thus, combining the above inequality with (18) yields

$$(\forall k \in \mathbb{N}) \quad \|y_k - x_k\|^2 \le 2\underline{\lambda}^{-2}\mu_1^{-1}(1-\theta)^{-1}\big(G(x_k) - \xi\big)^\theta \Delta'_k. \tag{19}$$

On the other hand, since E is bounded and Assumption 2.2 holds, there exist constants $\kappa > 0$, $\zeta > 0$ and $\theta \in [0,1)$ such that

$$\big(\forall r(x) \in \partial R(x)\big) \qquad \kappa |G(x) - \xi|^\theta \le \|\nabla F(x) + r(x)\|, \tag{20}$$

for every $x \in E$ such that $|G(x) - \xi| \le \zeta$. Since $\big(G(y_k)\big)_{k\in\mathbb{N}}$ converges to ξ, there exists $k^* \in \mathbb{N}$, such that, for every $k \ge k^*$, $|G(y_k) - \xi| < \zeta$. Hence, we have, for every $r(y_k) \in \partial R(y_k)$,

$$(\forall k \ge k^*) \quad \kappa |G(y_k) - \xi|^\theta \le \|\nabla F(y_k) + r(y_k)\|.$$

Let $r(y_k)$ be defined as in Algorithm (7). Then, we have

$$\begin{aligned}
\kappa |G(y_k) - \xi|^\theta &\le \|\nabla F(y_k) - \nabla F(x_k) + \nabla F(x_k) + r(y_k)\| \\
&\le \|\nabla F(y_k) - \nabla F(x_k)\| + \|\nabla F(x_k) + r(y_k)\| \\
&\le \|\nabla F(y_k) - \nabla F(x_k)\| + \tau \|x_k - y_k\|_{A_k}.
\end{aligned} \tag{21}$$

Thus, by using Assumptions 2.1(ii) and 2.3(ii), we get

$$|G(y_k) - \xi|^\theta \le \kappa^{-1}(L + \tau\sqrt{\overline{\nu}})\|x_k - y_k\|. \tag{22}$$

In addition, according to Assumption 2.5,

$$G(x_{k+1}) - \xi \le (1-\underline{\alpha})(G(x_k) - \xi) + |G(y_k) - \xi|. \tag{23}$$

Besides, it can be noticed that

$$\begin{aligned}
\sum_{k=k^*}^{+\infty} \Delta'_k &= \sum_{k=k^*}^{+\infty} \big(G(x_k) - \xi\big)^{1-\theta} - \big(G(x_{k+1}) - \xi\big)^{1-\theta} \\
&= \big(G(x_{k^*}) - \xi\big)^{1-\theta},
\end{aligned}$$

which shows that $(\Delta'_k)_{k\in\mathbb{N}}$ is a summable sequence. From (19), the summability of $(\Delta'_k)_{k\in\mathbb{N}}$, (23), and (22), and by setting

$$(\forall k\in\mathbb{N})\qquad \begin{cases} u_k = \|y_k - x_k\| \\ g_k = G(x_k) - \xi \geq 0 \\ g'_k = |G(y_k) - \xi| \geq 0 \\ \beta = \kappa^{-1}(L+\tau\sqrt{\overline{\nu}}) > 0 \\ \Delta_k = 2\underline{\lambda}^{-2}\mu_1^{-1}(1-\theta)^{-1}\Delta'_k, \end{cases}$$

Lemma 3.3 allows us to conclude that $(\|y_k - x_k\|)_{k\in\mathbb{N}}$ is summable when $\theta \neq 0$. When $\theta = 0$, as $x_k - y_k \to 0$, there exists $k^{**} \geq k^*$ such that

$$(\forall k \geq k^{**})\quad \kappa^{-1}(L+\tau\sqrt{\overline{\nu}})\|x_k - y_k\| < 1.$$

Hence, according to (22) (recall that $0^0 = 0$), one necessarily has, for every $k \geq k^{**}$, $G(y_k) = \xi$. Then, according to (19), for every $k \geq k^{**}$, $x_k = y_k$, which trivially shows that $(\|y_k - x_k\|)_{k\in\mathbb{N}}$ is summable.

Moreover, according to (7c) and Assumption 2.4(ii), we have

$$(\forall k\in\mathbb{N})\quad \|x_{k+1} - x_k\| = \lambda_k\|y_k - x_k\| \leq \|y_k - x_k\|.$$

Hence, the sequence $(x_k)_{k\in\mathbb{N}}$ satisfies the finite length property. In addition, since this latter condition implies that $(x_k)_{k\in\mathbb{N}}$ is a Cauchy sequence, it converges towards a point $\hat{x}$. As $(x_k - y_k)_{k\in\mathbb{N}}$ converges to 0, $(y_k)_{k\in\mathbb{N}}$ also converges to the same limit $\hat{x}$, and $(y_k)_{k\in\mathbb{N}}$ satisfies the finite length property since

$$\begin{aligned}(\forall k\in\mathbb{N})\quad \|y_{k+1} - y_k\| &\leq \|y_{k+1} - x_k\| + \|x_k - y_k\| \\ &\leq \|y_{k+1} - x_{k+1}\| + \|x_{k+1} - x_k\| + \|x_k - y_k\|.\end{aligned}$$

It remains us to show that the limit $\hat{x}$ is a critical point of G. To this end, let us define

$$(\forall k\in\mathbb{N})\qquad t(y_k) = \nabla F(y_k) + r(y_k),$$

where $r(y_k)$ is given by (7), so that $(y_k, t(y_k)) \in \operatorname{Graph}\partial G$. In addition, by proceeding like in (21), we obtain

$$(\forall k\in\mathbb{N})\qquad \|t(y_k)\| \leq (L+\tau\sqrt{\overline{\nu}})\|x_k - y_k\|.$$

Since the sequences $(x_k)_{k\in\mathbb{N}}$ and $(y_k)_{k\in\mathbb{N}}$ both converge to $\hat{x}$, $(y_k, t(y_k))_{k\in\mathbb{N}}$ converges to $(\hat{x}, 0)$. Furthermore, according to Remark 2.2(iii), the restriction of G to its domain is continuous. Thus, as $(\forall k\in\mathbb{N})$ $y_k \in \operatorname{dom} G$, the sequence $(G(y_k))_{k\in\mathbb{N}}$ converges to $G(\hat{x})$. Finally, according to the closedness property of ∂G (see Remark 2.1), $(\hat{x}, 0) \in \operatorname{Graph}\partial G$ i.e., $\hat{x}$ is a critical point of G. □

As an offspring of the previous theorem, the proposed algorithm can be shown to locally converge to a global minimizer of G:

Corollary 3.3. *Suppose that Assumptions 2.1-2.5 hold, and suppose that $(x_k)_{k\in\mathbb{N}}$ and $(y_k)_{k\in\mathbb{N}}$ are sequences generated by Algorithm (7). There exists $\upsilon > 0$ such that, if*

$$G(x_0) \leq \inf_{x\in\mathbb{R}^N} G(x) + \upsilon, \tag{24}$$

then $(x_k)_{k\in\mathbb{N}}$ and $(y_k)_{k\in\mathbb{N}}$ both converge to a solution to Problem (1).

Proof. According to Remark 2.2(iii),

$$\xi = \inf_{x \in \mathbb{R}^N} G(x) < +\infty.$$

Let $E = \operatorname{lev}_{\xi+\delta} G$, where $\delta > 0$. As a consequence of Assumption 2.1(iii), E is bounded. In view of Assumption 2.2, there exists constants $\kappa > 0$, $\zeta > 0$ and $\theta \in [0, 1)$ such that (20) holds for every $x \in E$ such that $|G(x) - \xi| \leq \zeta$, that is, $G(x) \leq \xi + \zeta$ since, by definition of ξ, we always have $G(x) \geq \xi$. Let us now set $\upsilon = \min\{\delta, \zeta\} > 0$ and choose x_0 satisfying (24). It follows from Theorem 3.1(iii) that, for every $k \in \mathbb{N}$,

$$G(x_k) \leq \xi + \upsilon.$$

By continuity of the restriction of G to its domain and Theorem 3.1(i), $(x_k)_{k\in\mathbb{N}}$ and $(y_k)_{k\in\mathbb{N}}$ converge to $\hat{x}$ which is such that $G(\hat{x}) \leq \xi + \upsilon$. In other words, the Kurdyka-Łojasiewicz inequality is satisfied at $\hat{x}$:

$$(\forall t(\hat{x}) \in \partial G(\hat{x})) \quad \|t(\hat{x})\| \geq \kappa |G(\hat{x}) - \xi|^{\theta}.$$

As $\hat{x}$ is a critical point of G, 0 belongs to $\partial G(\hat{x})$, and then we have

$$|G(\hat{x}) - \xi|^{\theta} \leq 0.$$

This shows that $G(\hat{x}) = \inf_{x\in\mathbb{R}^N} G(x)$. ☐

We now comment on the differences between the results in some related works and our results.

Remark 3.1.

(i) *In* [23]*, the convergence is established under the assumption that G is convex, while our study relies on the fulfillment of Assumption 2.2. Moreover, it can be noticed that Assumption 2.3 on the matrices $(A_k)_{k\in\mathbb{N}}$ are less restrictive than Condition* (6) *considered in* [23]*.*

(ii) *Note that the convergence of* (7) *was established in* [9] *in the non-preconditioned case, i.e. $A_k \equiv L\,\mathrm{I}_N$, but for a non necessarily convex function R. Constant values of the relaxation parameters $(\lambda_k)_{k\in\mathbb{N}}$ and of the step sizes $(\gamma_k)_{k\in\mathbb{N}}$ were considered.*

4 Application to image reconstruction

4.1 Optimization problem

In this section, we consider an inverse problem where a degraded image $z = (z^{(m)})_{1\leq m\leq M} \in \mathbb{R}^M$ related to an original image $\overline{x} \in [0, +\infty)^N$ is observed through the model:

$$(\forall m \in \{1, \ldots, M\}) \qquad z^{(m)} = [H\overline{x}]^{(m)} + \sigma^{(m)}([H\overline{x}]^{(m)})\, w^{(m)},$$

where $H \in \mathbb{R}^{M\times N}$ is a matrix with non-negative elements and, for every $m \in \{1, \ldots, M\}$, $[H\overline{x}]^{(m)}$ denotes the m-th component of $H\overline{x}$. Moreover, $(w^{(m)})_{1\leq m\leq M}$ is a realization of a Gaussian random vector with zero-mean and covariance matrix I_M, and

$$(\forall m \in \{1, \ldots, M\}) \quad \begin{aligned} \sigma^{(m)} \colon [0, +\infty) &\to (0, +\infty) \\ u &\mapsto \sqrt{a^{(m)} u + b^{(m)}} \end{aligned} \tag{25}$$

with $(a^{(m)})_{1\le m\le M} \in [0,+\infty)^M$, $(b^{(m)})_{1\le m\le M} \in (0,+\infty)^M$. Such a noise model arises in a number of digital imaging devices [36–38] where the acquired image is contaminated by signal-dependent Photon shot noise and by independent electrical or thermal noise. Signal-dependent Gaussian noise can also be viewed as a second-order approximation of Poisson-Gauss noise which is frequently encountered in astronomy, medicine and biology [39, 40]. Our objective is to produce an estimate $\hat{x} \in [0,+\infty)^N$ of the target image $\overline{x}$ from the observed data z.

The original image can be estimated by solving (1) where F is a so-called data fidelity term and R is a penalty function serving to incorporate *a priori* information. In the Bayesian framework, this is equivalent to compute a maximum a posteriori (MAP) estimate [41] of the original image. In this context, a usual choice for the data fidelity term is the neg-log-likelihood of the data which is expressed as

$$(\forall x \in \mathbb{R}^N) \quad F(x) = \begin{cases} F_1(x) + F_2(x) & \text{if } x \in [0,+\infty)^N \\ +\infty & \text{otherwise,} \end{cases}$$

where

$$(\forall x \in [0,+\infty)^N) \quad F_1(x) = \frac{1}{2}\sum_{m=1}^{M} \rho_1^{(m)}([Hx]^{(m)}), \tag{26}$$

$$F_2(x) = \frac{1}{2}\sum_{m=1}^{M} \rho_2^{(m)}([Hx]^{(m)}), \tag{27}$$

and

$$(\forall m \in \{1,\dots,M\})(\forall u \in [0,+\infty)) \quad \rho_1^{(m)}(u) = \frac{1}{2}\frac{\left(u - z^{(m)}\right)^2}{a^{(m)}u + b^{(m)}} \tag{28}$$

$$\rho_2^{(m)}(u) = \frac{1}{2}\widehat{\log}(a^{(m)}u + b^{(m)}). \tag{29}$$

In the equation (29), $\widehat{\log}$ is a semi-algebraic approximation of the logarithm defined on $(0,+\infty)$, which, like the original function, is concave and Lipschitz differentiable on any interval $[\underline{b},+\infty)$ with $\underline{b} \in (0,+\infty)$. Such approximations are commonly used in numerical implementations of the logarithmic function [42, Chap.4].

Furthermore, a hybrid penalty function, made up of two terms $R = R_1 + R_2$ is considered. First, in order to take into account the dynamic range of the target image, we define $R_1 = \iota_C$, where $C = [x_{\min}, x_{\max}]^N$, $x_{\min} \in [0,+\infty)$ and $x_{\max} \in (x_{\min},+\infty)$ are the minimal and the maximal values of the components of $\overline{x}$, respectively, and ι_C is the indicator function of C defined as

$$\iota_C(x) = \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{otherwise.} \end{cases}$$

Secondly, a sparsity prior in an analysis frame [43–45] is introduced by taking

$$(\forall x \in \mathbb{R}^N) \quad R_2(x) = \sum_{j=1}^{J} \vartheta^{(j)} |[Wx]^{(j)}|,$$

where $(\vartheta^{(j)})_{1\le j\le J} \in [0,+\infty)^J$ and $W \in \mathbb{R}^{J\times N}$ is a tight frame operator, i.e. there exists $\mu_W \in (0,+\infty)$ such that $W^\top W = \mu_W\, \mathrm{I}_N$.

It is clear that Assumption 2.1(i) holds. For every $m \in \{1, \dots, M\}$, the first and second derivatives[1] of $\rho_1^{(m)}$ at $u \in [0, +\infty)$ are

$$\dot{\rho}_1^{(m)}(u) = \frac{(u - z^{(m)})(a^{(m)}u + a^{(m)}z^{(m)} + 2b^{(m)})}{2(a^{(m)}u + b^{(m)})^2}, \tag{30}$$

$$\ddot{\rho}_1^{(m)}(u) = \frac{(a^{(m)}z^{(m)} + b^{(m)})^2}{(a^{(m)}u + b^{(m)})^3}. \tag{31}$$

Hence, $\ddot{\rho}_1^{(m)}$ is bounded and, as $\dot{\rho}_2^{(m)}$ is Lipschitzian, Assumption 2.1(ii) is satisfied. Assumption 2.1(iii) follows from the fact that $\operatorname{dom} R = C$ is bounded.

Finally, since F_1, F_2 and R_2 are semi-algebraic functions and C is a semi-algebraic set, G is a semi-algebraic function and Assumption 2.2 holds.

4.2 Construction of the majorant

Let us now present a family of diagonal matrices $(A_k)_{k\in\mathbb{N}}$ that fulfill Assumption 2.3. First, note that, on its domain, F is the sum of the convex function F_1 and the concave function F_2, respectively defined by (26) and (27). For every $k \in \mathbb{N}$, let x_k be generated by the k-th iteration of Algorithm (7). A majorant function of F_2 on $[0, +\infty)^N$ at x_k is

$$(\forall x \in \mathbb{R}^N) \quad Q_2(x, x_k) = F_2(x_k) + (x - x_k)^\top \nabla F_2(x_k), \tag{32}$$

where $\nabla F_2(x_k)$ is the gradient of F_2 at x_k. The next lemma allows us to construct a majorant function of F_1 at x_k. Before stating this lemma, we introduce the function $\omega\colon [0, +\infty) \to \mathbb{R}^M \colon u \mapsto \big(\omega^{(m)}(u)\big)_{1\le m\le M}$ where, for every $m \in \{1, \dots, M\}$,

$$(\forall u \in [0, +\infty)) \quad \omega^{(m)}(u) = \begin{cases} \ddot{\rho}_1^{(m)}(0) & \text{if} \quad u = 0, \\ 2\,\frac{\rho_1^{(m)}(0) - \rho_1^{(m)}(u) + u\dot{\rho}_1^{(m)}(u)}{u^2} & \text{if} \quad u > 0, \end{cases} \tag{33}$$

and ρ_1 is defined by (28).

Lemma 4.1. *Let F_1 be defined by (26) where $H = \big(H^{(m,n)}\big)_{1\le m\le M, 1\le n\le N} \in [0, +\infty)^{M\times N}$. For every $k \in \mathbb{N}$, let*

$$A_k = \operatorname{Diag}\big(P^\top \omega(Hx_k)\big) + \varepsilon\, \mathrm{I}_N, \tag{34}$$

where $\varepsilon \in [0, +\infty)$ and $P = \big(P^{(m,n)}\big)_{1\le m\le M, 1\le n\le N}$ is the matrix whose elements are given by

$$(\forall m \in \{1, \dots, M\})(\forall n \in \{1, \dots, N\}) \quad P^{(m,n)} = H^{(m,n)} \sum_{p=1}^{N} H^{(m,p)}. \tag{35}$$

Then, Q_1 defined as

$$(\forall x \in \mathbb{R}^N) \quad Q_1(x, x_k) = F_1(x_k) + (x - x_k)^\top \nabla F_1(x_k) + \frac{1}{2}(x - x_k)^\top A_k (x - x_k),$$

is a majorant function of F_1 on $[0, +\infty)^N$ at x_k.

Proof. For every $m \in \{1, \dots, M\}$, $\rho_1^{(m)}$ is convex and infinitely derivable on $[0, +\infty)$. Let us define

$$(\forall (u, u') \in [0, +\infty)^2)$$
$$q^{(m)}(u, u') = \rho_1^{(m)}(u') + (u - u')\dot{\rho}_1^{(m)}(u') + \frac{1}{2}\omega^{(m)}(u')\,(u - u')^2, \tag{36}$$

[1] We consider right derivatives when $u = 0$.

where functions $\dot{\rho}_1^{(m)}$ and $\omega^{(m)}$ are respectively given by (28) and (33). If $a^{(m)} = 0$, then $\rho_1^{(m)}$ is a quadratic function and we have

$$(\forall (u, u') \in [0, +\infty)^2) \quad \rho_1^{(m)}(u) = q^{(m)}(u, u').$$

Let us now assume that $a^{(m)} \in (0, +\infty)$. The third derivative of $\rho_1^{(m)}$ is given by $\dddot{\rho}_1^{(m)} : u \in [0, +\infty) \mapsto -3a^{(m)} \dfrac{(a^{(m)} z^{(m)} + b^{(m)})^2}{(a^{(m)} u + b^{(m)})^4}$, which is negative. Then, $\dot{\rho}_1^{(m)}$ is a strictly concave function on $[0, +\infty)$. By using a simplified version of [35, App.B.], we will prove the positivity of function $q^{(m)}(\cdot, u') - \rho_1^{(m)}$ on $[0, +\infty)$, for every $u' \in (0, +\infty)$. The second derivative of the latter function is given by

$$(\forall u \in [0, +\infty)) \quad \ddot{q}^{(m)}(u, u') - \ddot{\rho}_1^{(m)}(u) = \omega^{(m)}(u') - \ddot{\rho}_1^{(m)}(u). \tag{37}$$

Moreover, according to second-order Taylor's formula, there exists $\tilde{u} \in (0, u')$ such that

$$\rho_1^{(m)}(0) = \rho_1^{(m)}(u') - u' \dot{\rho}_1^{(m)}(u') + \frac{1}{2} u'^2 \ddot{\rho}_1^{(m)}(\tilde{u}),$$

hence, using (33) and (37),

$$(\forall u \in [0, +\infty)) \quad \ddot{q}^{(m)}(u, u') - \ddot{\rho}_1^{(m)}(u) = \ddot{\rho}_1^{(m)}(\tilde{u}) - \ddot{\rho}_1^{(m)}(u). \tag{38}$$

Since $\dddot{\rho}_1^{(m)}$ is negative, $\ddot{\rho}_1^{(m)}$ is strictly decreasing, and (38) implies that $\dot{q}^{(m)}(\cdot, u') - \dot{\rho}_1^{(m)}$ is first strictly decreasing on $(0, \tilde{u})$, then strictly increasing on $(\tilde{u}, +\infty)$. On the one hand, (33) and (36) yield

$$\begin{cases} q^{(m)}(u', u') - \rho_1^{(m)}(u') & = 0, \\ q^{(m)}(0, u') - \rho_1^{(m)}(0) & = 0. \end{cases} \tag{39}$$

Thus, according to the mean value theorem, there exists $u^* \in (0, u')$ such that $\dot{q}^{(m)}(u^*, u') - \dot{\rho}_1^{(m)}(u^*) = 0$. On the other hand, according to (36),

$$\dot{q}^{(m)}(u', u') - \dot{\rho}_1^{(m)}(u') = 0.$$

Therefore, from the monotonicity properties of $\dot{q}^{(m)}(\cdot, u') - \dot{\rho}_1^{(m)}$, we deduce that u^* is the unique zero of this function on $(0, u')$, and

$$\begin{cases} \dot{q}^{(m)}(u, u') - \dot{\rho}_1^{(m)}(u') & > 0, \qquad \forall u \in [0, u^*), \\ \dot{q}^{(m)}(u, u') - \dot{\rho}_1^{(m)}(u') & < 0, \qquad \forall u \in (u^*, u'), \\ \dot{q}^{(m)}(u, u') - \dot{\rho}_1^{(m)}(u') & > 0, \qquad \forall u \in (u', +\infty). \end{cases} \tag{40}$$

Equation (40) implies that $q^{(m)}(\cdot, u') - \rho_1^{(m)}(\cdot)$ is strictly increasing on $[0, u^*)$, strictly decreasing on (u^*, u') and strictly increasing on $(u', +\infty)$. Thus, given (39),

$$(\forall (u, u') \in [0, +\infty) \times (0, +\infty)) \quad \rho_1^{(m)}(u) \leq q^{(m)}(u, u'). \tag{41}$$

Moreover, from the expression of $\omega^{(m)}(0)$ in (33) and from (31), it is easy to show that

$$(\forall u \in [0, +\infty)) \quad \rho_1^{(m)}(u) \leq q^{(m)}(u, 0). \tag{42}$$

Therefore, by gathering (41) and (42), we obtain

$$(\forall (u, u') \in [0, +\infty)^2) \quad \rho_1^{(m)}(u) \leq q^{(m)}(u, u'), \tag{43}$$

which, as pointed out before, is still valid when $a^{(m)} = 0$.

Majoration (43) implies that, for every $k \in \mathbb{N}$,

$$(\forall x \in [0, +\infty)^N) \quad F_1(x) \le F_1(x_k) + (x - x_k)^\top \nabla F_1(x_k) \\ + \frac{1}{2}(Hx - Hx_k)^\top \operatorname{Diag}\big(\omega(Hx_k)\big)(Hx - Hx_k).$$

Let $(\Lambda^{(m,n)})_{1 \le m \le M, 1 \le n \le N} \in [0, +\infty)^{M \times N}$ be such that

$$(\forall (m,n) \in \{1, \ldots, M\} \times \{1, \ldots, N\}) \quad \Lambda^{(m,n)} = \begin{cases} 0 & \text{if } H^{(m,n)} = 0 \\ \dfrac{H^{(m,n)}}{\sum_{p=1}^N H^{(m,p)}} & \text{if } H^{(m,n)} > 0. \end{cases}$$

According to Jensen's inequality and (35), for every $m \in \{1, \ldots, M\}$,

$$\begin{aligned} &\big([\Lambda^{(m,1)}, \ldots, \Lambda^{(m,N)}](x - x_k)\big)^2 \le \sum_{n=1}^N \Lambda^{(m,n)} \big(x^{(n)} - x_k^{(n)}\big)^2 \\ \Leftrightarrow \quad &\big([H^{(m,1)}, \ldots, H^{(m,N)}](x - x_k)\big)^2 \le \sum_{n=1}^N P^{(m,n)} \big(x^{(n)} - x_k^{(n)}\big)^2. \end{aligned}$$

Since the convexity of $\rho_1^{(m)}$ for $m \in \{1, \ldots, M\}$ implies the positivity of $\omega^{(m)}$ on $[0, +\infty)$, we deduce that

$$(Hx - Hx_k)^\top \operatorname{Diag}\big(\omega(Hx_k)\big)(Hx - Hx_k) \le (x - x_k)^\top \operatorname{Diag}\big(P^\top \omega(Hx_k)\big)(x - x_k),$$

which yields the result. □

It can be deduced from (32) and Lemma 4.1 that Assumption 2.3(i) is satisfied for $Q = Q_1 + Q_2$.

It can be further noticed that, for every $m \in \{1, \ldots, M\}$, the derivative of $\omega^{(m)}$ at $u' \in (0, +\infty)$ is

$$\begin{aligned} \dot{\omega}^{(m)}(u') &= 2\frac{u'^2 \ddot{\rho}_1^{(m)}(u') - 2\big(\rho_1^{(m)}(0) - \rho_1^{(m)}(u') + u' \dot{\rho}_1^{(m)}(u')\big)}{u'^3} \\ &= 2\frac{\ddot{\rho}_1^{(m)}(u') - \ddot{\rho}_1^{(m)}(\tilde{u})}{u'}, \end{aligned}$$

where $\tilde{u} \in (0, u')$. Since the third-order derivative of $\rho_1^{(m)}$ is negative, $\omega^{(m)}$ is a strictly decreasing positive function. According to the expression of R, $\operatorname{dom} R = [x_{\min}, x_{\max}]^N$, so that its image $\{Hx \mid x \in \operatorname{dom} R\}$ is a compact set. Thus, the function $x \mapsto \omega^{(m)}([Hx]^{(m)})$ admits a minimum value $\omega_{\min}^{(m)} > 0$. Then, Assumption 2.3(ii) is satisfied for the matrices $(A_k)_{k \in \mathbb{N}}$ defined by (34) with

$$\begin{cases} \underline{\nu} = \varepsilon + \min\limits_{1 \le n \le N} \sum_{m=1}^M P^{(m,n)} \, \omega_{\min}^{(m)}, \\ \overline{\nu} = \varepsilon + \max\limits_{1 \le n \le N} \sum_{m=1}^M P^{(m,n)} \, \ddot{\rho}_1^{(m)}(0). \end{cases} \tag{44}$$

Note that, if each column of H is non-zero, we can choose $\varepsilon = 0$ in (44). Otherwise, we must choose $\varepsilon > 0$.

4.3 Backward step

The implementation of the VMFB algorithm (in its exact form) requires to compute, at each iteration $k \in \mathbb{N}$, variable y_k corresponding to the proximity operator of R at $\tilde{x}_k = x_k - \gamma_k A_k^{-1} \nabla F(x_k)$, relative to the metric induced by $\gamma_k^{-1} A_k$, i.e.

$$y_k = \underset{x \in \mathbb{R}^N}{\operatorname{argmin}} \Big\{ \sum_{j=1}^{J} \vartheta^{(j)} |[Wx]^{(j)}| + \iota_C(x) + \frac{1}{2} \|x - \tilde{x}_k\|^2_{\gamma_k^{-1} A_k} \Big\}. \tag{45}$$

Due to the presence of matrix W, it is not possible to obtain an explicit expression for y_k. Sub-iterations are thus needed to compute it. Equation (45) is equivalent to

$$y_k = \gamma_k^{1/2} A_k^{-1/2} \underset{x \in \mathbb{R}^N}{\operatorname{argmin}} \Big\{ \frac{1}{2} \|x - \gamma_k^{-1/2} A_k^{1/2} \tilde{x}_k\|^2 + \sum_{j=1}^{J} \vartheta^{(j)} \gamma_k^{1/2} |[W A_k^{-1/2} x]^{(j)}| + \iota_C(\gamma_k^{1/2} A_k^{-1/2} x) \Big\}.$$

The above optimization problem can be solved by various algorithms. The numerical results provided in the next section have been obtained by using the Dual Forward-Backward algorithm [46].

4.4 Experimental results

We now demonstrate the practical performance of our algorithm on two image reconstruction scenarios. In the first scenario, the standard `Peppers` image of size 256×256 from `http://sipi.usc.edu/database/` is degraded by a blur operator H corresponding to a uniform convolution kernel of size 5×5, and further corrupted with the considered signal-dependent additive noise with standard-deviation given by (25) where, for every $m \in \{1, \dots, M\}$, $a^{(m)} = 0.5$ and $b^{(m)} = 1$. Here, $x_{\min} = 0$, $x_{\max} = 252$, and the Lipschitz constant of F is equal to $L = 1.8 \times 10^4$. In the second scenario, $\overline{x}$ corresponds to one slice of the standard `Zubal` phantom from [47] with dimensions 128×128 and H is the Radon matrix modeling $M = 16384$ parallel projections from 128 acquisition lines and 128 angles. The sinogram $H\overline{x}$ is corrupted with signal-dependent noise where, for every $m \in \{1, \dots, M\}$, $a^{(m)} = 0.01$ and $b^{(m)} = 0.1$. In this case, $x_{\min} = 0$, $x_{\max} = 1$, and the Lipschitz constant is $L = 6.0 \times 10^6$. For both experiments, the employed frame is a redundant (undecimated) wavelet transform using Daubechies' eight-taps filters over three resolution levels. The related tight frame constant is equal to $\mu_W = 64$. Parameters $\big(\vartheta^{(j)}\big)_{1 \le j \le J}$ are adjusted so as to maximize the signal-to-noise ratio (SNR) between the original image $\overline{x}$ and the reconstructed one $\hat{x}$, which is expressed as

$$\mathrm{SNR} = 20 \log_{10} \left(\frac{\|\overline{x}\|}{\|\hat{x} - \overline{x}\|} \right).$$

Figs. 1, 3 and. 4 show the degraded data, and the reconstructed images with VMFB, for the considered deblurring and reconstruction problems. We also present in Fig. 4 the reconstruction result obtained using the standard filtered back-projection approach [48]. The proposed VMFB algorithm is compared with FB [1] and FISTA [14] algorithms. The setting $\lambda_k \equiv 1$ has been chosen for FB and VMFB. Two values of the step-size are tested, namely $\gamma_k \equiv 1$ and $\gamma_k \equiv 1.9$. For these values of $(\lambda_k)_{k \in \mathbb{N}}$ and $(\gamma_k)_{k \in \mathbb{N}}$, Assumption 2.4 is satisfied. Moreover, according to Remark 2.3(ii), Assumption 2.5 also holds. Figs. 2

and 5 illustrate the variations of $\big(G(x_k) - G(\hat{x})\big)_k$ and $\big(\|x_k - \hat{x}\|\big)_k$ with respect to the computation time, using the proposed VMFB algorithm, FB and FISTA algorithms, when performing tests on an Intel(R) Xeon(R) CPU X5570 at 2.93GHz, in a single thread, using a Matlab 7 implementation. Note that the optimal solution $\hat{x}$ has been precomputed for each algorithm, using a large number of iterations. We can observe that, in the deblurring experiment, FISTA converges faster than FB algorithm, while, in the tomography example, the two methods behave similarly. Concerning the choice of the step-size, our results show that, for FB and VMFB algorithms, a faster convergence is obtained for $\gamma_k \equiv 1.9$ in both experiments.

In conclusion, the variable metric strategy leads to a significant acceleration in terms of decay of both the objective function and the error on the iterates in each experiment.

Figure 1: **Deblurring:** Degraded image, SNR=19.3 dB (left) and restored image (right) with the proposed algorithm, SNR=24.3 dB.

References

[1] Chen, G.H.G., Rockafellar, R.T.: Convergence rates in forward-backward splitting. SIAM J. Optim. **7**, 421–444 (1997)

[2] Tseng, P.: A modified forward-backward splitting method for maximal monotone mappings. SIAM J. Control Optim. **38**(2), 431–446 (1998)

[3] Hiriart-Urruty, J.B., Lemaréchal, C.: Convex Analysis and Minimization Algorithms. Springer-Verlag, New York (1993)

[4] Combettes, P.L., Vũ, B.C.: Variable metric quasi-Fejér monotonicity. Nonlinear Anal. **78**, 17–31 (2013)

[5] Moreau, J.J.: Proximité et dualité dans un espace hilbertien. Bull. Soc. Math. France **93**, 273–299 (1965)

[6] Combettes, P.L., Wajs, V.R.: Signal recovery by proximal forward-backward splitting. Multiscale Model. Simul. **4**(4), 1168–1200 (2005)

[7] Combettes, P.L., Pesquet, J.C.: Proximal thresholding algorithm for minimization over orthonormal bases. SIAM J. Optim. **18**(4), 1351–1376 (2007)

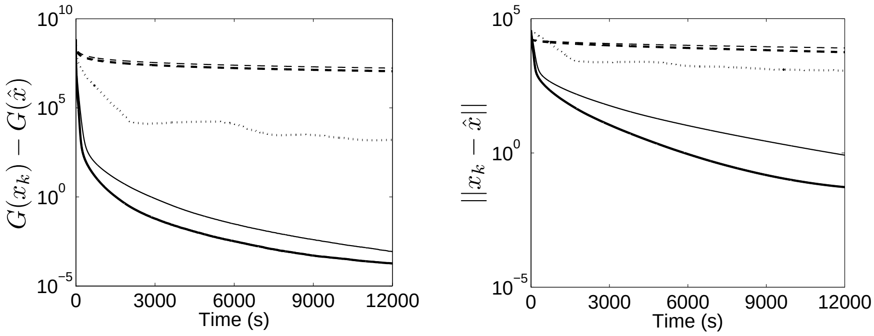

Figure 2: **Deblurring:** Comparison of VMFB algorithm with $\gamma_k \equiv 1$ (solid thin line) and $\gamma_k \equiv 1.9$ (solid thick line), FB algorithm with $\gamma_k \equiv 1$ (dashed thin line) and $\gamma_k \equiv 1.9$ (dashed thick line) and FISTA algorithm (dotted line).

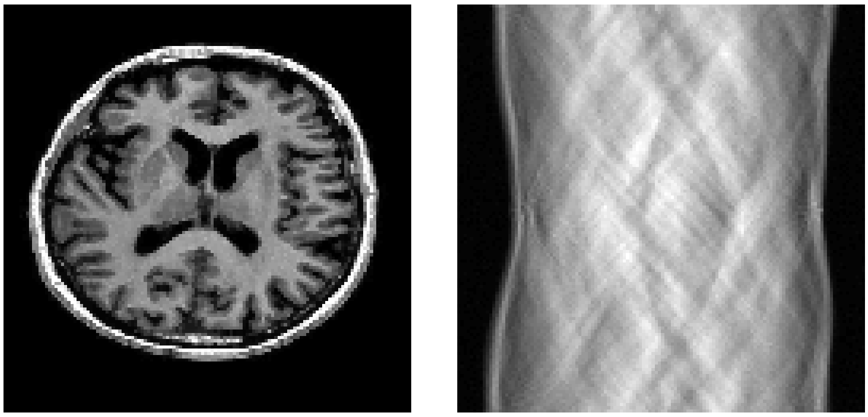

Figure 3: **Reconstruction:** Original image (left) and degraded sinogram (right).

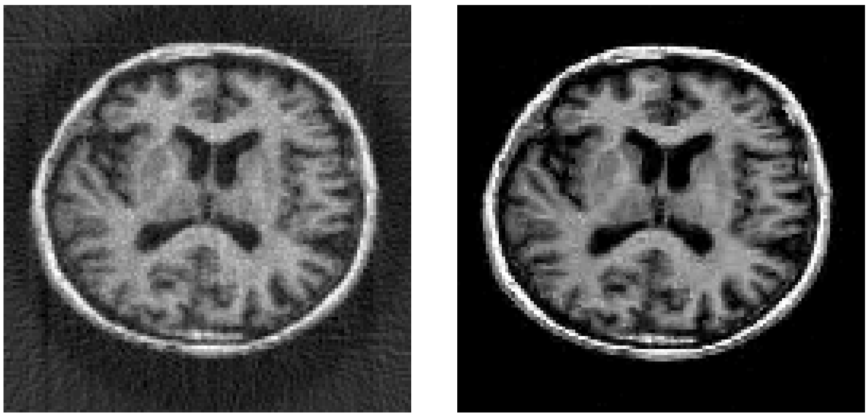

Figure 4: **Reconstruction:** Restored images with filtered back-projection, SNR=7 dB (left) and with the proposed algorithm, SNR=18.9 dB (right).

[8] Attouch, H., Bolte, J.: On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. Math. Program. **116**, 5–16 (2008)

[9] Attouch, H., Bolte, J., Svaiter, B.F.: Convergence of descent methods for semi-

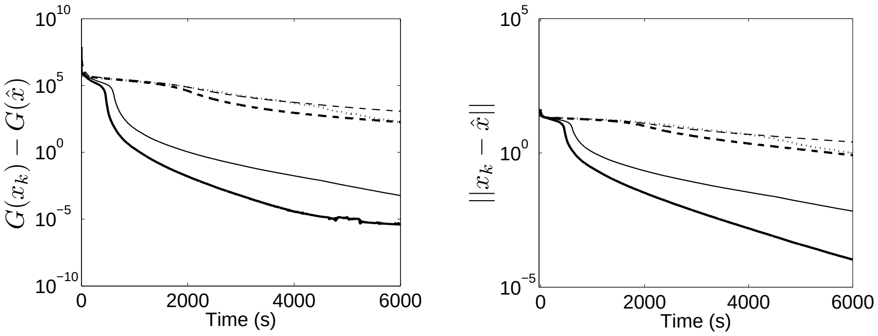

Figure 5: **Reconstruction:** Comparison of VMFB algorithm with $\gamma_k \equiv 1$ (solid thin line) and $\gamma_k \equiv 1.9$ (solid thick line), FB algorithm with $\gamma_k \equiv 1$ (dashed thin line) and $\gamma_k \equiv 1.9$ (dashed thick line) and FISTA algorithm (dotted line).

algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods. Math. Program. **137**, 91–129 (2011)

[10] Łojasiewicz, S.: Une propriété topologique des sous-ensembles analytiques réels, pp. 87–89. Editions du centre National de la Recherche Scientifique (1963)

[11] Bolte, J., Daniilidis, A., Lewis, A.: The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. SIAM J. Optim. **17**, 1205–1223 (2006)

[12] Bolte, J., Daniilidis, A., Lewis, A., Shiota, M.: Clarke subgradients of stratifiable functions. SIAM J. Optim. **18**(2), 556–572 (2007)

[13] Kurdyka, K., Parusinski, A.: w_f-stratification of subanalytic functions and the Łojasiewicz inequality. Comptes rendus de l'Académie des sciences. Série 1, Mathématique **318**(2), 129–133 (1994)

[14] Beck, A., Teboulle, M.: Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. IEEE Trans. Image Process. **18**(11), 2419 –2434 (2009)

[15] Bioucas-Dias, J.M., Figueiredo, M.A.: A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration. IEEE Trans. Image Process. **16**, 2992–3004 (2007)

[16] Zibulevsky, M., Elad, M.: ℓ_2-ℓ_1 optimization in signal and image processing. IEEE Signal Process. Mag. **27**(3), 76–88 (2010)

[17] Kowalski, M.: Proximal algorithm meets a conjugate descent. Tech. rep. (2010). `http://hal.archives-ouvertes.fr/docs/00/50/57/33/PDF/proxconj.pdf`

[18] Nesterov, Y.: Gradient methods for minimizing composite objective function (2007). `http://www.optimization-online.org/DB_HTML/2007/09/1784.html`

[19] Bonnans, J.F., Gilbert, J.C., Lemaréchal, C., Sagastizábal, C.A.: A family of variable metric proximal methods. Math. Program. **68**, 15–47 (1995)

[20] Burke, J.V., Qian, M.: A variable metric proximal point algorithm for monotone operators. SIAM J. Control Optim. **37**, 353–375 (1999)

[21] Parente, L.A., Lotito, P.A., Solodov, M.V.: A class of inexact variable metric proximal point algorithms. SIAM J. Optim. **19**, 240–260 (2008)

[22] Lotito, P.A., Parente, L.A., Solodov, M.V.: A class of variable metric decomposition methods for monotone variational inclusions. J. Convex Anal. **16**, 857–880 (2009)

[23] Combettes, P.L., Vũ, B.C.: Variable metric forward-backward splitting with applications to monotone inclusions in duality. to appear in Optimization (2013). `http://arxiv.org/abs/1206.6791`

[24] Becker, S., Fadili, J.: A quasi-Newton proximal splitting method. Tech. rep. (2012). `http://arxiv.org/abs/1206.1156`

[25] Bertsekas, D.P.: Nonlinear Programming, 2nd edn. Athena Scientific, Belmont, MA (1999)

[26] Chen, K.: Matrix Preconditioning Techniques and Applications, *Cambridge Monogr. Appl. Comput. Math.*, vol. 19. Cambridge University Press (2005)

[27] Nocedal, J., Wright, S.J.: Numerical Optimization. Springer-Verlag, New York, NY (1999)

[28] Bertsekas, D.P.: Projected Newton methods for optimization problems with simple constraints. SIAM J. Control Optim. **20**, 221–246 (1982)

[29] Birgin, E.G., Martínez, J.M., Raydan, M.: Nonmonotone spectral projected gradient methods on convex sets. SIAM J. Optim. **10**(4), 1196–1211 (2000)

[30] Bonettini, S., Zanella, R., Zanni, L.: A scaled gradient projection method for constrained image deblurring. Inverse Problems **25**(1), 015,002+ (2009)

[31] Rockafellar, R.T., Wets, R.J.B.: Variational Analysis, 1st edn. Springer-Verlag (1997)

[32] Mordukhovich, B.S.: Variational Analysis and Generalized Differentiation. Vol. I: Basic theory, *Series of Comprehensive Studies in Mathematics*, vol. 330. Springer-Verlag, Berlin-Heidelberg (2006)

[33] Bolte, J., Daniilidis, A., Ley, O., Mazet, L.: Characterizations of Łojasiewicz inequalities and applications. Trans. Amer. Math. Soc. **362**(6), 3319–3363 (2010)

[34] Hunter, D.R., Lange, K.: A tutorial on MM algorithms. Amer. Stat. **58**(1), 30–37 (2004)

[35] Erdogan, H., Fessler, J.A.: Monotonic algorithms for transmission tomography. IEEE Trans. Med. Imag. **18**(9), 801–814 (1999)

[36] Healey, G.E., Kondepudy, R.: Radiometric CCD camera calibration and noise estimation. IEEE Trans. Pattern Anal. Mach. Intell. **16**(3), 267–276 (1994)

[37] Tian, H., Fowler, B., Gamal, A.E.: Analysis of temporal noise in CMOS photodiode active pixel sensor. IEEE Solid State Circuits Mag. **36**(1), 92–101 (2001)

[38] Janesick, J.R.: Photon Transfer, vol. PM170. SPIE Press Book, Bellingham, WA (2007)

[39] Foi, A., Trimeche, M., Katkovnik, V., Egiazarian, K.: Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data. IEEE Trans. Image Process. **17**(10), 1737–1754 (2008)

[40] Li, J., Shen, Z., Jin, R., Zhang, X.: A reweighted ℓ_2 method for image restoration with Poisson and mixed Poisson-Gaussian noise. Tech. rep. (2012). UCLA Preprint, `ftp://ftp.math.ucla.edu/pub/camreport/cam12-84.pdf`

[41] Sorenson, H.W.: Parameter Estimation : Principles and Problems, *Control and systems theory*, vol. 9, 1st edn. M. Dekker, New York (1980)

[42] Dahlquist, G., Bjorck, A.: Numerical Methods. Dover Publication, Mineola, N.Y. (2003)

[43] Chaâri, L., Pustelnik, N., Chaux, C., Pesquet, J.C.: Solving inverse problems with overcomplete transforms and convex optimization techniques. In: SPIE, Wavelets XIII, vol. 7446. San Diego, CA, USA (2009). 14 p.

[44] Elad, M., Milanfar, P., Ron, R.: Analysis versus synthesis in signal priors. Inverse Problems **23**(3), 947–968 (2007)

[45] Pustelnik, N., Pesquet, J.C., Chaux, C.: Relaxing tight frame condition in parallel proximal methods for signal restoration. IEEE Trans. Signal Process. **60**(2), 968–973 (2012)

[46] Combettes, P.L., Dũng, D., Vũ, B.C.: Proximity for sums of composite functions. J. Math. Anal. Appl. **380**(2), 680–688 (2011)

[47] Zubal, I.G., Harrell, C.R., Smith, E.O., Rattner, Z., Gindi, G., Hoffer, P.B.: Computerized three-dimensional segmented human anatomy. Med. Phys. **21**, 299–302 (1994)

[48] Slaney, M., Kak, A.: Principles of computerized tomographic imaging. SIAM, Philadelphia (1988)