A Block Coordinate Variable Metric Forward-Backward Algorithm

Emilie Chouzenoux, Jean-Christophe Pesquet and Audrey Repetti *

Abstract

A number of recent works have emphasized the prominent role played by the Kurdyka-Łojasiewicz inequality for proving the convergence of iterative algorithms solving possibly nonsmooth/nonconvex optimization problems. In this work, we consider the minimization of an objective function satisfying this property, which is a sum of a non necessarily convex differentiable function and a non necessarily differentiable or convex function. The latter function is expressed as a separable sum of functions of blocks of variables. Such an optimization problem can be addressed with the Forward-Backward algorithm which can be accelerated thanks to the use of variable metrics derived from the Majorize-Minimize principle. We propose to combine the latter acceleration technique with an alternating minimization strategy which relies upon a flexible update rule. We give conditions under which the sequence generated by the resulting Block Coordinate Variable Metric Forward-Backward algorithm converges to a critical point of the objective function. An application example to a nonconvex phase retrieval problem encountered in signal/image processing shows the efficiency of the proposed optimization method.

1 Introduction

In this work, we are interested in the following optimization problem:

$$\text{Find} \quad \widehat{\boldsymbol{x}} \in \text{Argmin} \, (G := F + R), \tag{1}$$

where $G : \mathbb{R}^N \to (-\infty, +\infty]$ is a coercive function (i.e. $\lim_{\|\boldsymbol{x}\| \to +\infty} G(\boldsymbol{x}) = +\infty$), F is a differentiable function, R is a proper lower semicontinuous function which is additively block separable, and $\text{Argmin} \, G \neq \varnothing$ denotes the set of minimizers G. More precisely, let $(\mathbb{J}_j)_{1 \le j \le J}$ be a partition of $\{1, \ldots, N\}$ into $J \ge 2$ subsets, and for every $j \in \{1, \ldots, J\}$, let $N_j \neq 0$ be the cardinality of $\mathbb{J}_j$. Any vector $\boldsymbol{x} \in \mathbb{R}^N$ with elements $(x^{(n)})_{1 \le n \le N}$ is block-decomposed into $(\boldsymbol{x}^{(j)})_{1 \le j \le J} \in \mathbb{R}^{N_1} \times \ldots \times \mathbb{R}^{N_J}$, where, for every $j \in \{1, \ldots, J\}$, $\boldsymbol{x}^{(j)} = (x^{(n)})_{n \in \mathbb{J}_j} \in \mathbb{R}^{N_j}$. With this notation, we assume that

$$(\forall \boldsymbol{x} \in \mathbb{R}^N) \quad R(\boldsymbol{x}) := \sum_{j=1}^{J} R_j(\boldsymbol{x}^{(j)}), \tag{2}$$

where, for every $j \in \{1, \ldots, J\}$, $R_j \colon \mathbb{R}^{N_j} \to (-\infty, +\infty]$.

A standard approach for solving (1) in this context consists of using a *Block Coordinate Descent* (BCD) algorithm, where, at each iteration $\ell \in \mathbb{N}$, G is minimized with respect to the

---
*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.

j_ℓ block coordinates with $j_\ell \in \{1, \ldots, J\}$, while the others remain fixed, leading to the following iterations:

$$
\begin{aligned}
&\text{Let } \boldsymbol{x}_0 \in \mathbb{R}^N, \\
&\text{For } \ell = 0, 1, \ldots \\
&\quad \left| \begin{aligned}
&\text{Let } j_\ell \in \{1, \ldots, J\}, \\
&\boldsymbol{x}_{\ell+1}^{(j_\ell)} \in \underset{\boldsymbol{y} \in \mathbb{R}^{N_{j_\ell}}}{\operatorname{Argmin}} \left(F_{j_\ell}(\boldsymbol{y}, \boldsymbol{x}_\ell^{(\overline{j}_\ell)}) + R_{j_\ell}(\boldsymbol{y}) \right), \\
&\boldsymbol{x}_{\ell+1}^{(\overline{j}_\ell)} = \boldsymbol{x}_\ell^{(\overline{j}_\ell)}.
\end{aligned} \right.
\end{aligned}
\tag{3}
$$

In the above algorithm, for every $j \in \{1, \ldots, J\}$, $\overline{j}$ denotes the complementary set of j on $\{1, \ldots, J\}$, i.e. $\overline{j} := \{1, \ldots, J\} \setminus \{j\}$, and for every $\boldsymbol{x} \in \mathbb{R}^N$, $\boldsymbol{x}^{(\overline{j})} := (\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(j-1)}, \boldsymbol{x}^{(j+1)}, \ldots, \boldsymbol{x}^{(J)})$. Moreover, for a given $\boldsymbol{x}^{(\overline{j})} \in \bigtimes_{i \in \overline{j}} \mathbb{R}^{N_i}$, function $F_j(\cdot, \boldsymbol{x}^{(\overline{j})}) \colon \mathbb{R}^{N_j} \to \mathbb{R}$ is the partial function defined as

$$
(\forall \boldsymbol{y} \in \mathbb{R}^{N_j}) \quad F_j(\boldsymbol{y}, \boldsymbol{x}^{(\overline{j})}) := F(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(j-1)}, \boldsymbol{y}, \boldsymbol{x}^{(j+1)}, \ldots, \boldsymbol{x}^{(J)}).
\tag{4}
$$

The BCD method (3) is described in various reference books [8, 29, 36, 48] assuming a *cyclic rule*, i.e.

$$
(\forall \ell \in \mathbb{N}) \quad j_\ell - 1 = \ell \mod (J).
\tag{5}
$$

In this case, since Algorithm (3) can be viewed as a generalization of the Gauss-Seidel strategy for solving linear systems [24], it is sometimes also referred to as a *nonlinear Gauss-Seidel method* ([8, Chap.2], [36, Chap.7]). Up to the best of our knowledge, one of the most general convergence results for the BCD algorithm (3) has been established in [45] under the assumptions that (i) G is quasi-convex and hemivariate regular in each block, (ii) $(j_\ell)_{\ell \in \mathbb{N}}$ follows an *essentially cyclic rule* (i.e. blocks can be updated in an arbitrary manner as far as each of them is updated at least once within a given number of iterations) and (iii) either G is pseudoconvex in every pair of blocks or has at most one minimum with respect to each block. As pointed out in [45], the last assumption is sharp in the sense that the algorithm may not converge if we only assume that G is convex w.r.t. each block (see an illustration in [37]). The proximal version of the BCD algorithm, introduced in [4], allows this limitation to be overcome. It is defined as follows:

$$
\begin{aligned}
&\text{Let } \boldsymbol{x}_0 \in \mathbb{R}^N, \\
&\text{For } \ell = 0, 1, \ldots \\
&\quad \left| \begin{aligned}
&\text{Let } j_\ell \in \{1, \ldots, J\}, \\
&\boldsymbol{x}_{\ell+1}^{(j_\ell)} \in \operatorname{prox}_{\gamma_\ell^{-1} \boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell), F_{j_\ell}(\cdot, \boldsymbol{x}_\ell^{(\overline{j}_\ell)}) + R_{j_\ell}} \left(\boldsymbol{x}_\ell^{(j_\ell)} \right), \\
&\boldsymbol{x}_{\ell+1}^{(\overline{j}_\ell)} = \boldsymbol{x}_\ell^{(\overline{j}_\ell)},
\end{aligned} \right.
\end{aligned}
\tag{6}
$$

where, for every $\ell \in \mathbb{N}$, $\gamma_\ell \in (0, +\infty)$ and $\boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell) \in \mathbb{R}^{N_{j_\ell} \times N_{j_\ell}}$ is a symmetric positive definite (SPD) matrix. Hereabove, $\operatorname{prox}_{U, \psi}$ denotes the so-called *proximity operator* of a proper lower semicontinuous function $\psi \colon \mathbb{R}^M \to \mathbb{R}$ relative to the metric induced by a SPD matrix $U \in \mathbb{R}^{M \times M}$ (see Section 2.1). Note that Algorithm (6) has been extended in [7] for Bregman projection operators, in the case when $J = 2$, F is a Bregman distance and R_1, R_2 are convex functions. Note also that the celebrated POCS (*Projection Onto Convex Sets*) algorithm [13] can be seen as a particular case of Algorithm (6) when $F \equiv 0$ and, for every $j \in \{1, \ldots, J\}$, R_j is the indicator function of a convex set.

The convergence of the sequence $(\boldsymbol{x}_\ell)_{\ell \in \mathbb{N}}$ generated by Algorithm (6) to a solution of (1) has been established in [4] for a convex Lipschitz differentiable function F and proper lower-semicontinuous convex functions $(R_j)_{1 \le j \le J}$, in the case when $(j_\ell)_{\ell \in \mathbb{N}}$ follows a cyclic rule, and $(\boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell))_{\ell \in \mathbb{N}}$ are identity matrices. Recently, the convergence of the proximal BCD iterates to a critical point of G in the case of nonconvex functions F and $(R_j)_{1 \le j \le J}$, has been proved in [2] when $(\boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell))_{\ell \in \mathbb{N}}$ are identity matrices, and then generalized in [3] for general SPD matrices $(\boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell))_{\ell \in \mathbb{N}}$, again assuming a cyclic rule. The convergence studies in [2, 3] mainly rely on the assumption that the objective function G satisfies the Kurdyka-Łojasiewicz (KL)

inequality [28]. The interesting point is that this inequality holds for a wide class of functions such as real analytic functions, semi-algebraic functions and many others [9,10,27,28]. Since the proximal step in (6) is not explicit in general, an inexact version of the proximal BCD method is also considered in [3], with similar convergence guarantees.

Another strategy to circumvent the difficulty of solving the block subproblems in (6) is to replace, at each iteration, the proximal step by a Forward-Backward step, thus leading to the so-called *Block Coordinate Variable Metric Forward-Backward* (BC-VMFB) algorithm:

$$
\begin{aligned}
&\text{Let } \boldsymbol{x}_0 \in \mathbb{R}^N, \\
&\text{For } \ell = 0, 1, \ldots \\
&\quad\left|\begin{aligned}
&\text{Let } j_\ell \in \{1, \ldots, J\}, \\
&\boldsymbol{x}_{\ell+1}^{(j_\ell)} \in \operatorname{prox}_{\gamma_\ell^{-1} \boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell), R_{j_\ell}} \left(\boldsymbol{x}_\ell^{(j_\ell)} - \gamma_\ell \left(\boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell)\right)^{-1} \nabla_{j_\ell} F(\boldsymbol{x}_\ell) \right), \\
&\boldsymbol{x}_{\ell+1}^{(\overline{j}_\ell)} = \boldsymbol{x}_\ell^{(\overline{j}_\ell)},
\end{aligned}\right.
\end{aligned}
\tag{7}
$$

where, for every $\boldsymbol{x} \in \mathbb{R}^N$ and $j \in \{1, \ldots, J\}$, $\nabla_j F(\boldsymbol{x}) \in \mathbb{R}^{N_j}$ is the partial gradient of F with respect to $\boldsymbol{x}^{(j)}$ computed at $\boldsymbol{x}$. Algorithm (7) was firstly introduced in [15] for the minimization of the Burg entropy function under linear constraints, and then extended to the more general case of a smooth function F [30,31]. Recently, the convergence of this algorithm has been studied in the case of an arbitrary nonsmooth function R under the assumptions that G satisfies KL inequality and F is Lipschitz differentiable [12,47]. The convergence of the sequence $(\boldsymbol{x}_\ell)_{\ell \in \mathbb{N}}$ generated by (7) to a critical point of (1) has been proved in [47] in the case when F and R are respectively convex and convex w.r.t. each block variable, and generalized in [12] when neither F nor R is necessarily convex. Note that the aforementioned works considered actually a simplified version of Algorithm (7) where $(\boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell))_{\ell \in \mathbb{N}}$ are identity matrices and the sequence $(j_\ell)_{\ell \in \mathbb{N}}$ follows a cyclic rule. The BC-VMFB algorithm is then referred to as the *Proximal Alternating Linearized Minimization* (PALM) algorithm [12]. An exact (resp. inexact) version of Algorithm (7) with general SPD matrices $(\boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell))_{\ell \in \mathbb{N}}$ is studied in [39] (resp. [44]), in the context of a *random rule*, i.e., for every $\ell \in \mathbb{N}$, j_ℓ is a realization of a uniform random variable. Assuming that F and R_j are convex, the authors establish the convergence of the sequence $(G(\boldsymbol{x}_\ell))_{\ell \in \mathbb{N}}$ in the sense that, for all $\delta \geq 0$ and $\epsilon \geq 0$, there exists $\ell_0 \in \mathbb{N}$ such that the probability of having $G(\boldsymbol{x}_{\ell_0}) - G(\widehat{\boldsymbol{x}}) \leq \epsilon$ is greater than $1 - \delta$. Finally, let us emphasize that, as already noticed in [38], for carefully chosen matrices $(\boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell))_{\ell \in \mathbb{N}}$, the BC-VMFB algorithm can be viewed as a particular form of the block alternating majorize-minimize (MM) approach proposed in [21, 41, 43] in the context of image reconstruction. Therefore, some convergence properties of Algorithm (7) can be deduced from those derived in [26] in the case when R_j are indicator functions of closed convex subsets of $\mathbb{R}^{N_j}$, and in [38] for arbitrary nonsmooth convex functions R_j. However, it should be noticed that the convergence of $(\boldsymbol{x}_\ell)_{\ell \in \mathbb{N}}$ to a solution to (1) is only proved in [26,38] under specific assumptions, in particular the uniqueness of solutions to each block subproblem and to the initial problem (1) is required.

In this paper, we consider an inexact version of (7) where the preconditioning matrices $(\boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell))_{\ell \in \mathbb{N}}$ are chosen according to MM arguments. The convergence of the proposed algorithm is established for blocks following an essentially cyclic rule, under weak assumptions on the involved functions (G is mainly assumed to satisfy KL inequality similarly to [3]). Note that this convergence study generalizes our previous work [16] (see also [35] for a related approach) which was restricted to an inexact Variable Metric Forward-Backward algorithm without block alternation (i.e. $J = 1$ and $N_1 = N$).

The rest of the paper is organized as follows: Section 2 introduces the assumptions made in the paper and presents the proposed inexact BC-VMFB strategy. Section 3 investigates the convergence properties. In particular, the convergence rate of the proposed algorithm is studied. Finally, Section 4 provides some numerical results and a discussion of the algorithm performance by means of experiments concerning a large-size image reconstruction problem.

2 Proposed Optimization Method

2.1 Analysis background

Let us first recall some definitions and the notation that will be used throughout the paper. We define the weighted norm:

$$\left(\forall \boldsymbol{x} \in \mathbb{R}^N\right) \qquad \|\boldsymbol{x}\|_{\boldsymbol{U}} := \langle \boldsymbol{x}, \boldsymbol{U}\boldsymbol{x} \rangle^{1/2}, \tag{8}$$

where $\langle \cdot, \cdot \rangle$ is the standard scalar product of $\mathbb{R}^N$ and $\boldsymbol{U} \in \mathbb{R}^{N \times N}$ is some SPD matrix.

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

Definition 2.2. *[40, Def. 8.3], [33, Sec.1.3] Let $\psi \colon \mathbb{R}^N \to (-\infty, +\infty]$ be a proper function and let $\boldsymbol{x} \in \operatorname{dom} \psi$. The* Fréchet sub-differential *of ψ at $\boldsymbol{x}$ is the following set:*

$$\widehat{\partial}\psi(\boldsymbol{x}) := \left\{ \widehat{\boldsymbol{t}} \in \mathbb{R}^N \; : \; \liminf_{\substack{\boldsymbol{y} \to \boldsymbol{x} \\ \boldsymbol{y} \neq \boldsymbol{x}}} \frac{1}{\|\boldsymbol{x} - \boldsymbol{y}\|} \left(\psi(\boldsymbol{y}) - \psi(\boldsymbol{x}) - \left\langle \boldsymbol{y} - \boldsymbol{x}, \widehat{\boldsymbol{t}} \right\rangle \right) \geq 0 \right\}.$$

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

$$\partial\psi(\boldsymbol{x}) := \left\{ \boldsymbol{t} \in \mathbb{R}^N \; : \; \exists \boldsymbol{y}_k \to \boldsymbol{x}, \; \psi(\boldsymbol{y}_k) \to \psi(\boldsymbol{x}), \; \widehat{\boldsymbol{t}}_k \in \widehat{\partial}\psi(\boldsymbol{y}_k) \to \boldsymbol{t} \right\}.$$

Remark 2.1.

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

(ii) *Definition 2.2 implies that $\partial\psi$ is closed [3], that is:*
Let $(\boldsymbol{y}_k, \boldsymbol{t}_k)_{k \in \mathbb{N}}$ be a sequence of $\operatorname{Graph} \partial\psi := \left\{ (\boldsymbol{x}, \boldsymbol{t}) \in \mathbb{R}^N \times \mathbb{R}^N : \boldsymbol{t} \in \partial\psi(\boldsymbol{x}) \right\}$. If $(\boldsymbol{y}_k, \boldsymbol{t}_k)$ converges to $(\boldsymbol{x}, \boldsymbol{t})$ and $\psi(\boldsymbol{y}_k)$ converges to $\psi(\boldsymbol{x})$, then $(\boldsymbol{x}, \boldsymbol{t}) \in \operatorname{Graph} \partial\psi$.

The proximity operator ([25, Sec. XV.4], [18] and [3]) is defined as follows:

Definition 2.3. *Let $\psi \colon \mathbb{R}^N \to (-\infty, +\infty]$ be a proper, lower semicontinuous function, let $\boldsymbol{U} \in \mathbb{R}^{N \times N}$ be a SPD matrix, and let $\boldsymbol{x} \in \mathbb{R}^N$. The* proximity operator *of ψ at $\boldsymbol{x}$ relative to the metric induced by $\boldsymbol{U}$ is defined as*

$$\operatorname{prox}_{\boldsymbol{U},\psi}(\boldsymbol{x}) := \operatorname*{Argmin}_{\boldsymbol{y} \in \mathbb{R}^N} \psi(\boldsymbol{y}) + \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{x}\|_{\boldsymbol{U}}^2. \tag{9}$$

Remark 2.2.

(i) *In the above definition, since $\| \cdot \|_{\boldsymbol{U}}^2$ is coercice and ψ is proper and lower semicontinuous, if ψ is bounded from below by an affine function, then $\operatorname{prox}_{\boldsymbol{U},\psi}$ is a nonempty set.*

(ii) *If $\boldsymbol{U}$ is equal to $\mathbf{I}_N$, the identity matrix of $\mathbb{R}^{N \times N}$, then $\operatorname{prox}_\psi \equiv \operatorname{prox}_{\mathbf{I}_N,\psi}$ is the proximity operator employed in [3]. In addition, if ψ is a convex function, then the minimizer of $\psi + \frac{1}{2}\| \cdot - \boldsymbol{x}\|_{\boldsymbol{U}}^2$ is unique and $\operatorname{prox}_\psi \equiv \operatorname{prox}_{\mathbf{I}_N,\psi}$ is the proximity operator originally defined in [34].*

2.2 Assumptions

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

Assumption 2.1.

(i) *For every $j \in \{1, \dots, J\}$, $R_j \colon \mathbb{R}^{N_j} \to (-\infty, +\infty]$ is proper, lower semicontinuous, bounded from below by an affine function 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 $\mathrm{dom}\, R$ where $L > 0$, i.e.,*
$$\big(\forall (\boldsymbol{x}, \boldsymbol{y}) \in (\mathrm{dom}\, R)^2\big) \quad \|\nabla F(\boldsymbol{x}) - \nabla F(\boldsymbol{y})\| \leq L \|\boldsymbol{x} - \boldsymbol{y}\|.$$

(iii) *G is coercive.*

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

Remark 2.3.

(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 [3, 19]. In particular, if $\mathrm{dom}\, R$ is compact and F is twice continuously differentiable, Assumption 2.1(ii) holds.*

(ii) *According to Assumption 2.1(ii), $\mathrm{dom}\, R \subset \mathrm{dom}\, F = \mathbb{R}^N$. Thus, as a consequence of Assumption 2.1(i), $\mathrm{dom}\, G = \mathrm{dom}\, R$ is nonempty.*

(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 $\boldsymbol{x} \in \mathrm{dom}\, R$, $\mathrm{lev}_{\leq G(\boldsymbol{x})}\, G$ is a compact set. Moreover, the set of minimizers of G is nonempty and compact.*

(iv) *If, for every $j \in \{1, \dots, J\}$, R_j is proper, lower semicontinuous and convex, then R_j is bounded from below by an affine function.*

Assumption 2.2.
Function G satisfies the Kurdyka-Łojasiewicz (KL) inequality i.e., for every $\xi \in \mathbb{R}$, and, for every bounded subset E of $\mathbb{R}^N$, there exist three constants $\kappa \in (0, +\infty)$, $\zeta \in (0, +\infty)$ and $\theta \in [0, 1)$ such that
$$\big(\forall \boldsymbol{t} \in \partial G(\boldsymbol{x})\big) \qquad \|\boldsymbol{t}\| \geq \kappa |G(\boldsymbol{x}) - \xi|^\theta, \tag{10}$$
for every $\boldsymbol{x} \in E$ such that $|G(\boldsymbol{x}) - \xi| \leq \zeta$ (with the convention $0^0 = 0$).

As emphasized in [1], where this form of KL inequality was used to establish convergence results, Assumption 2.2 is satisfied for a wide class of functions, such as, in particular, real analytic and semi-algebraic functions. Note however that a more general local version of Assumption 2.2 can be found in the literature [10, 11].

Some matrices serving to define some appropriate variable metric will play a central role in the algorithm proposed in this work. More specifically, let $j_\ell \in \{1, \dots, J\}$ be the index of the block selected at iteration $\ell \in \mathbb{N}$ of Algorithm (7), let $\boldsymbol{x}_\ell \in \mathrm{dom}\, R$ be the associated iterate and let $\boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell) \in \mathbb{R}^{N_{j_\ell} \times N_{j_\ell}}$ be a SPD matrix that fulfills the following so-called *majorization condition*:

Assumption 2.3.

(i) *The quadratic function defined as*

$$(\forall \boldsymbol{y} \in \mathbb{R}^{N_{j_\ell}}) \quad Q_{j_\ell}(\boldsymbol{y}, \boldsymbol{x}_\ell) := F(\boldsymbol{x}_\ell) + \left\langle \boldsymbol{y} - \boldsymbol{x}_\ell^{(j_\ell)}, \nabla_{j_\ell} F(\boldsymbol{x}_\ell) \right\rangle$$
$$+ \frac{1}{2} \left\langle \boldsymbol{y} - \boldsymbol{x}_\ell^{(j_\ell)}, \boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell)(\boldsymbol{y} - \boldsymbol{x}_\ell^{(j_\ell)}) \right\rangle,$$

is a majorant function of $F_{j_\ell}(\cdot, \boldsymbol{x}_\ell^{(\bar{\jmath}_\ell)})$ at $\boldsymbol{x}_\ell^{(j_\ell)}$ on $\mathrm{dom}\, R_{j_\ell}$, i.e.,

$$(\forall \boldsymbol{y} \in \mathrm{dom}\, R_{j_\ell}) \quad F_{j_\ell}(\boldsymbol{y}, \boldsymbol{x}_\ell^{(\bar{\jmath}_\ell)}) \leq Q_{j_\ell}(\boldsymbol{y}, \boldsymbol{x}_\ell).$$

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

$$(\forall \ell \in \mathbb{N}) \quad \underline{\nu} \mathbf{I}_{N_{j_\ell}} \preceq \boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell) \preceq \overline{\nu} \mathbf{I}_{N_{j_\ell}}.$$

Remark 2.4.

(i) *Note that it is not necessary to build a quadratic majorant of $F_j(\cdot, \boldsymbol{x}^{(\bar{\jmath})})$ on $\mathrm{dom}\, R_j$, for every $j \in \{1, \ldots, J\}$ and for every $\boldsymbol{x}^{(\bar{\jmath})} \in \times_{i \in \bar{\jmath}} \mathrm{dom}\, R_i$.*

(ii) *Suppose that, for every $\boldsymbol{x}' \in \mathrm{dom}\, R$, a quadratic majorant function of F on $\mathrm{dom}\, R$ is given by*

$$(\forall \boldsymbol{x} \in \mathbb{R}^N) \quad Q(\boldsymbol{x}, \boldsymbol{x}') := F(\boldsymbol{x}') + \langle \boldsymbol{x} - \boldsymbol{x}', \nabla F(\boldsymbol{x}') \rangle + \frac{1}{2} \langle \boldsymbol{x} - \boldsymbol{x}', \boldsymbol{B}(\boldsymbol{x}')(\boldsymbol{x} - \boldsymbol{x}') \rangle, \quad (11)$$

where $\boldsymbol{B}(\boldsymbol{x}') \in \mathbb{R}^{N \times N}$ is a SPD matrix. Then, Assumption 2.3(i) is satisfied for $\boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell) = \left(B(\boldsymbol{x}_\ell)^{(n,n')} \right)_{(n,n') \in \mathbb{J}_{j_\ell}^2}$, where, for every $(n, n') \in \{1, \ldots, N\}^2$, $B(\boldsymbol{x}_\ell)^{(n,n')}$ denotes the (n, n') element of matrix $\boldsymbol{B}(\boldsymbol{x}_\ell)$. Moreover, if there exists $(\underline{\nu}, \overline{\nu}) \in (0, +\infty)^2$ such that, for every $\boldsymbol{x}' \in \mathrm{dom}\, R$, $\underline{\nu} \mathbf{I}_N \preceq \boldsymbol{B}(\boldsymbol{x}') \preceq \overline{\nu} \mathbf{I}_N$, then Assumption 2.3(ii) is also satisfied.

(iii) *If $\mathrm{dom}\, R$ is convex, the existence of the majorant function (11) is ensured when F satisfies Assumption 2.1(ii) (see [16, Lem. 3.1]).*

Moreover, in order to ensure that each block is updated an infinite number of times, we make the following assumption, which is equivalent to the essentially cyclic rule from [45]:

Assumption 2.4.
Let $(j_\ell)_{\ell \in \mathbb{N}}$ be the sequence of updated block indices. There exists a constant $K \geq J$ such that, for every $\ell \in \mathbb{N}$, $\{1, \ldots, J\} \subset \{j_\ell, \ldots, j_{\ell+K-1}\}$.

Note that the blocks do not need to be updated in any specific order.

Finally, we suppose that, for every $\ell \in \mathbb{N}$, the stepsize γ_ℓ involved in Algorithm 7 satisfies the following assumption:

Assumption 2.5.
One of the following statements holds:

(i) *There exists $(\underline{\gamma}, \overline{\gamma}) \in (0, +\infty)^2$ such that, for every $\ell \in \mathbb{N}$, $\underline{\gamma} \leq \gamma_\ell \leq 1 - \overline{\gamma}$.*

(ii) *For every $j \in \{1, \ldots, J\}$, R_j is a convex function and there exists $(\underline{\gamma}, \overline{\gamma}) \in (0, +\infty)^2$ such that, for every $\ell \in \mathbb{N}$, $\underline{\gamma} \leq \gamma_\ell \leq 2 - \overline{\gamma}$.*

2.3 Inexact BC-VMFB 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 Algorithm (7):

Let $\alpha \in (1/2, +\infty), \beta \in (0, +\infty),$ and $x_0 \in \operatorname{dom} R,$

For $\ell = 0, 1, \ldots$

$\quad$ Let $j_\ell \in \{1, \ldots, J\},$

$\quad$ Find $\boldsymbol{x}_{\ell+1}^{(j_\ell)} \in \mathbb{R}^{N_{j_\ell}}$ and $\boldsymbol{r}_{\ell+1}^{(j_\ell)} \in \partial R_{j_\ell}(\boldsymbol{x}_{\ell+1}^{(j_\ell)})$ such that

$$R_{j_\ell}(\boldsymbol{x}_{\ell+1}^{(j_\ell)}) + \left\langle \boldsymbol{x}_{\ell+1}^{(j_\ell)} - \boldsymbol{x}_\ell^{(j_\ell)}, \nabla_{j_\ell} F(\boldsymbol{x}_\ell) \right\rangle + \alpha \|\boldsymbol{x}_{\ell+1}^{(j_\ell)} - \boldsymbol{x}_\ell^{(j_\ell)}\|_{\boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell)}^2 \le R_{j_\ell}(\boldsymbol{x}_\ell^{(j_\ell)}), \qquad (12\text{a})$$

$$\|\nabla_{j_\ell} F(\boldsymbol{x}_\ell) + \boldsymbol{r}_{\ell+1}^{(j_\ell)}\| \le \beta \|\boldsymbol{x}_{\ell+1}^{(j_\ell)} - \boldsymbol{x}_\ell^{(j_\ell)}\|_{\boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell)}, \qquad (12\text{b})$$

$$\boldsymbol{x}_{\ell+1}^{(\overline{j}_\ell)} = \boldsymbol{x}_\ell^{(\overline{j}_\ell)}. \qquad (12\text{c})$$

Remark 2.5.

As already mentioned, under our working assumptions, Algorithm (12) can be viewed as an inexact version of Algorithm (7). To see this, let us consider sequences $(\boldsymbol{x}_\ell)_{\ell \in \mathbb{N}}$ and $(j_\ell)_{\ell \in \mathbb{N}}$ generated by Algorithm (7).

(i) *First, suppose that Assumption 2.5(i) holds. On the one hand, due to the definition of the proximity operator, we have, for every $\ell \in \mathbb{N}$,*

$$R_{j_\ell}(\boldsymbol{x}_{\ell+1}^{(j_\ell)}) + \left\langle \boldsymbol{x}_{\ell+1}^{(j_\ell)} - \boldsymbol{x}_\ell^{(j_\ell)}, \nabla_{j_\ell} F(\boldsymbol{x}_\ell) \right\rangle + \frac{\gamma_\ell^{-1}}{2} \|\boldsymbol{x}_{\ell+1}^{(j_\ell)} - \boldsymbol{x}_\ell^{(j_\ell)}\|_{\boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell)}^2 \le R_{j_\ell}(\boldsymbol{x}_\ell^{(j_\ell)}),$$

so that the sufficient-decrease condition (12a) holds with $\alpha = (1 - \overline{\gamma})^{-1}/2$ (as $\gamma_\ell^{-1} \ge (1 - \overline{\gamma})^{-1} > 1$). On the other hand, according to the variational characterization of the proximity operator, there exists $\boldsymbol{r}_{\ell+1}^{(j_\ell)} \in \partial R_{j_\ell}(\boldsymbol{x}_{\ell+1}^{(j_\ell)})$ such that

$$\boldsymbol{r}_{\ell+1}^{(j_\ell)} = -\nabla_{j_\ell} F(\boldsymbol{x}_\ell) + \gamma_\ell^{-1} \boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell)(\boldsymbol{x}_\ell^{(j_\ell)} - \boldsymbol{x}_{\ell+1}^{(j_\ell)}).$$

Using Assumptions 2.3(ii) and 2.5(i), we obtain

$$\|\boldsymbol{r}_{\ell+1}^{(j_\ell)} + \nabla_{j_\ell} F(\boldsymbol{x}_\ell)\| = \gamma_\ell^{-1} \|\boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell)(\boldsymbol{x}_\ell^{(j_\ell)} - \boldsymbol{x}_{\ell+1}^{(j_\ell)})\| \le \underline{\gamma}^{-1} \sqrt{\overline{\nu}} \|\boldsymbol{x}_\ell^{(j_\ell)} - \boldsymbol{x}_{\ell+1}^{(j_\ell)}\|_{\boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell)},$$

which is the inexact optimality condition (12b) with $\beta = \underline{\gamma}^{-1} \sqrt{\overline{\nu}}$.

(ii) *Now, suppose that Assumption 2.5(ii) holds. Let $\ell \in \mathbb{N}$. Due to the variational characterization of the proximity operator and the convexity of R_{j_ℓ}, there exists $\boldsymbol{r}_{\ell+1}^{(j_\ell)} \in \partial R_{j_\ell}(\boldsymbol{x}_{\ell+1}^{(j_\ell)})$ such that*

$$\begin{cases} \boldsymbol{r}_{\ell+1}^{(j_\ell)} = -\nabla_{j_\ell} F(\boldsymbol{x}_\ell) + \gamma_\ell^{-1} \boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell)(\boldsymbol{x}_\ell^{(j_\ell)} - \boldsymbol{x}_{\ell+1}^{(j_\ell)}) \\ \left\langle \boldsymbol{x}_{\ell+1}^{(j_\ell)} - \boldsymbol{x}_\ell^{(j_\ell)}, \boldsymbol{r}_{\ell+1}^{(j_\ell)} \right\rangle \ge R_{j_\ell}(\boldsymbol{x}_{\ell+1}^{(j_\ell)}) - R_{j_\ell}(\boldsymbol{x}_\ell^{(j_\ell)}), \end{cases}$$

which yields

$$R_{j_\ell}(\boldsymbol{x}_{\ell+1}^{(j_\ell)}) + \left\langle \boldsymbol{x}_{\ell+1}^{(j_\ell)} - \boldsymbol{x}_\ell^{(j_\ell)}, \nabla_{j_\ell} F(\boldsymbol{x}_\ell) \right\rangle + \gamma_\ell^{-1} \|\boldsymbol{x}_{\ell+1}^{(j_\ell)} - \boldsymbol{x}_\ell^{(j_\ell)}\|_{\boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell)}^2 \le R_{j_\ell}(\boldsymbol{x}_\ell^{(j_\ell)}),$$

so that the sufficient-decrease condition (12a) holds with $\alpha = (2 - \overline{\gamma})^{-1}$ (since $\gamma_\ell^{-1} \ge (2 - \overline{\gamma})^{-1} > 1/2$). The inexact optimality condition (12b) is obtained using similar arguments as in (i).

3 Convergence Analysis

3.1 Descent properties

In this section, we provide some technical results concerning the behavior of the sequence $(G(\boldsymbol{x}_\ell))_{\ell \in \mathbb{N}}$ generated by Algorithm (12), which will be useful in proving the convergence of the proposed algorithm.

Lemma 3.1. *Let $(\boldsymbol{x}_\ell)_{\ell \in \mathbb{N}}$ be a sequence generated by Algorithm* (12)*. Under Assumptions 2.1 and 2.3, there exists $\mu \in (0, +\infty)$ such that, for every $\ell \in \mathbb{N}$,*

$$G(\boldsymbol{x}_{\ell+1}) \le G(\boldsymbol{x}_\ell) - \frac{\mu}{2}\|\boldsymbol{x}_{\ell+1}^{(j_\ell)} - \boldsymbol{x}_\ell^{(j_\ell)}\|^2 = G(\boldsymbol{x}_\ell) - \frac{\mu}{2}\|\boldsymbol{x}_{\ell+1} - \boldsymbol{x}_\ell\|^2. \tag{13}$$

Proof. Let $\ell \in \mathbb{N}$. We have

$$G(\boldsymbol{x}_{\ell+1}) = F(\boldsymbol{x}_{\ell+1}) + R(\boldsymbol{x}_{\ell+1}).$$

On the one hand, according to Assumption 2.3(i),

$$F(\boldsymbol{x}_{\ell+1}) \le F(\boldsymbol{x}_\ell) + \left\langle \boldsymbol{x}_{\ell+1}^{(j_\ell)} - \boldsymbol{x}_\ell^{(j_\ell)}, \nabla_{j_\ell} F(\boldsymbol{x}_\ell) \right\rangle + \frac{1}{2}\|\boldsymbol{x}_{\ell+1}^{(j_\ell)} - \boldsymbol{x}_\ell^{(j_\ell)}\|_{\boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell)}^2. \tag{14}$$

On the other hand, using (12c),

$$\begin{aligned}
R(\boldsymbol{x}_{\ell+1}) &= R_{j_\ell}(\boldsymbol{x}_{\ell+1}^{(j_\ell)}) + \sum_{j \in \overline{\jmath}_\ell} R_j(\boldsymbol{x}_{\ell+1}^{(j)}) \\
&= R_{j_\ell}(\boldsymbol{x}_{\ell+1}^{(j_\ell)}) + \sum_{j \in \overline{\jmath}_\ell} R_j(\boldsymbol{x}_\ell^{(j)}) \\
&= R(\boldsymbol{x}_\ell) + \left(R_{j_\ell}(\boldsymbol{x}_{\ell+1}^{(j_\ell)}) - R_{j_\ell}(\boldsymbol{x}_\ell^{(j_\ell)}) \right).
\end{aligned}$$

Then, using (12a), we obtain

$$R(\boldsymbol{x}_{\ell+1}) \le R(\boldsymbol{x}_\ell) - \left\langle \boldsymbol{x}_{\ell+1}^{(j_\ell)} - \boldsymbol{x}_\ell^{(j_\ell)}, \nabla_{j_\ell} F(\boldsymbol{x}_\ell) \right\rangle - \alpha\|\boldsymbol{x}_{\ell+1}^{(j_\ell)} - \boldsymbol{x}_\ell^{(j_\ell)}\|_{\boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell)}^2. \tag{15}$$

Therefore, combining (14) and (15) yields

$$G(\boldsymbol{x}_{\ell+1}) \le G(\boldsymbol{x}_\ell) - \left(\alpha - \frac{1}{2} \right) \|\boldsymbol{x}_{\ell+1}^{(j_\ell)} - \boldsymbol{x}_\ell^{(j_\ell)}\|_{\boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell)}^2. \tag{16}$$

Finally, (13) is deduced from Assumption 2.3(ii) and the fact that $\alpha \in (1/2, +\infty)$, by setting $\mu = \underline{\nu}(2\alpha - 1)$, and using (12c). $\qquad\square$

Let the sequence $(\boldsymbol{\chi}_\ell)_{\ell \in \mathbb{N}}$ be defined as

$$(\forall \ell \in \mathbb{N}) \quad \boldsymbol{\chi}_\ell = (\boldsymbol{x}_{\ell+k+1} - \boldsymbol{x}_{\ell+k})_{0 \le k \le K-1} \in (\mathbb{R}^N)^K, \tag{17}$$

where $(\boldsymbol{x}_\ell)_{\ell \in \mathbb{N}}$ is a sequence generated by Algorithm (12) and K is the integer constant from Assumption 2.4. Then,

$$\|\boldsymbol{\chi}_\ell\|^2 = \sum_{k=0}^{K-1} \|\boldsymbol{x}_{\ell+k+1} - \boldsymbol{x}_{\ell+k}\|^2,$$

and the following property holds.

Lemma 3.2. *Let $(\boldsymbol{x}_\ell)_{\ell \in \mathbb{N}}$ be a sequence generated by Algorithm* (12)*. Under Assumptions 2.1, 2.3 and 2.4, for every $\ell \in \mathbb{N}$,*

$$G(\boldsymbol{x}_{\ell+K}) \le G(\boldsymbol{x}_\ell) - \frac{\mu}{2}\|\boldsymbol{\chi}_\ell\|^2,$$

where $\mu \in (0, +\infty)$ is the same constant as in Lemma 3.1.

Proof. Let $\ell \in \mathbb{N}$. According to Lemma 3.1, we have

$$
\begin{aligned}
G(\boldsymbol{x}_{\ell+K}) &\leq G(\boldsymbol{x}_{\ell+K-1}) - \frac{\mu}{2}\|\boldsymbol{x}_{\ell+K} - \boldsymbol{x}_{\ell+K-1}\|^2 \\
&\leq G(\boldsymbol{x}_{\ell+K-2}) - \frac{\mu}{2}\left(\|\boldsymbol{x}_{\ell+K-1} - \boldsymbol{x}_{\ell+K-2}\|^2 + \|\boldsymbol{x}_{\ell+K} - \boldsymbol{x}_{\ell+K-1}\|^2\right) \\
&\vdots \\
&\leq G(\boldsymbol{x}_\ell) - \frac{\mu}{2}\sum_{k=0}^{K-1}\|\boldsymbol{x}_{\ell+k+1} - \boldsymbol{x}_{\ell+k}\|^2.
\end{aligned}
$$

$\square$

3.2 Convergence theorem

We first state the following two lemmas which will be useful to handle the essentially cyclic rule:

Lemma 3.3. *Let $(\boldsymbol{x}_\ell)_{\ell \in \mathbb{N}}$ be a sequence of iterates generated by Algorithm (12). Let $\ell_0 \in \mathbb{N}$ and let $\mathscr{J}_{\ell_0}$ be a subset of $\{1, \dots, J\}$ containing j_{ℓ_0}. Then, under Assumptions 2.1 and 2.3, we have*

$$
\sum_{j \in \mathscr{J}_{\ell_0}} \|\nabla_j F(\boldsymbol{x}_{\ell_0+1}) + \boldsymbol{r}_{\ell_0+1}^{(j)}\|^2 \leq 2\left(L^2 + \beta^2\overline{\nu}\right)\|\boldsymbol{x}_{\ell_0+1} - \boldsymbol{x}_{\ell_0}\|^2 + 2\sum_{j \in \mathscr{J}_{\ell_0}\setminus\{j_{\ell_0}\}} \|\nabla_j F(\boldsymbol{x}_{\ell_0}) + \boldsymbol{r}_{\ell_0}^{(j)}\|^2, \quad (18)
$$

where $\boldsymbol{r}_{\ell_0+1}^{(j_{\ell_0})}$ is defined by Algorithm (12) and, for every $j \in \mathscr{J}_{\ell_0} \setminus \{j_{\ell_0}\}$, $\boldsymbol{r}_{\ell_0+1}^{(j)} \in \partial R_j(\boldsymbol{x}_{\ell_0+1}^{(j)})$ and $\boldsymbol{r}_{\ell_0}^{(j)} \in \partial R_j(\boldsymbol{x}_{\ell_0}^{(j)})$.

Proof. Let $\ell_0 \in \mathbb{N}$. According to Jensen's inequality,

$$
\begin{aligned}
\sum_{j \in \mathscr{J}_{\ell_0}} \|\nabla_j F(\boldsymbol{x}_{\ell_0+1}) + \boldsymbol{r}_{\ell_0+1}^{(j)}\|^2 &\leq 2\sum_{j \in \mathscr{J}_{\ell_0}} \|\nabla_j F(\boldsymbol{x}_{\ell_0+1}) - \nabla_j F(\boldsymbol{x}_{\ell_0})\|^2 \\
&\quad + 2\sum_{j \in \mathscr{J}_{\ell_0}} \|\nabla_j F(\boldsymbol{x}_{\ell_0}) + \boldsymbol{r}_{\ell_0+1}^{(j)}\|^2. \quad (19)
\end{aligned}
$$

On the one hand, since $\sum_{j=1}^{J} \|\nabla_j F(\boldsymbol{x}_{\ell_0+1}) - \nabla_j F(\boldsymbol{x}_{\ell_0})\|^2 = \|\nabla F(\boldsymbol{x}_{\ell_0+1}) - \nabla F(\boldsymbol{x}_{\ell_0})\|^2$, Assumption 2.1(ii) leads to

$$
\sum_{j \in \mathscr{J}_{\ell_0}} \|\nabla_j F(\boldsymbol{x}_{\ell_0+1}) - \nabla_j F(\boldsymbol{x}_{\ell_0})\|^2 \leq L^2\|\boldsymbol{x}_{\ell_0+1} - \boldsymbol{x}_{\ell_0}\|^2. \quad (20)
$$

On the other hand, since $j_{\ell_0} \in \mathscr{J}_{\ell_0}$

$$
\sum_{j \in \mathscr{J}_{\ell_0}} \|\nabla_j F(\boldsymbol{x}_{\ell_0}) + \boldsymbol{r}_{\ell_0+1}^{(j)}\|^2 = \|\nabla_{j_{\ell_0}} F(\boldsymbol{x}_{\ell_0}) + \boldsymbol{r}_{\ell_0+1}^{(j_{\ell_0})}\|^2 + \sum_{j \in \mathscr{J}_{\ell_0}\setminus\{j_{\ell_0}\}} \|\nabla_j F(\boldsymbol{x}_{\ell_0}) + \boldsymbol{r}_{\ell_0+1}^{(j)}\|^2.
$$

Moreover, using (12b) and Assumption 2.3(ii), and since, for every $j \in \mathscr{J}_{\ell_0} \setminus \{j_{\ell_0}\}$, $\boldsymbol{x}_{\ell_0+1}^{(j)} = \boldsymbol{x}_{\ell_0}^{(j)}$,

$$
\sum_{j \in \mathscr{J}_{\ell_0}} \|\nabla_j F(\boldsymbol{x}_{\ell_0}) + \boldsymbol{r}_{\ell_0+1}^{(j)}\|^2 \leq \beta^2\overline{\nu}\|\boldsymbol{x}_{\ell_0+1} - \boldsymbol{x}_{\ell_0}\|^2 + \sum_{j \in \mathscr{J}_{\ell_0}\setminus\{j_{\ell_0}\}} \|\nabla_j F(\boldsymbol{x}_{\ell_0}) + \boldsymbol{r}_{\ell_0}^{(j)}\|^2. \quad (21)
$$

Finally, (18) results from (19), (20) and (21). $\square$

Lemma 3.4. *Let $(\boldsymbol{x}_\ell)_{\ell\in\mathbb{N}}$ be a sequence of iterates generated by Algorithm (12). Let $(\ell_0,\ell_0')\in\mathbb{N}^2$ be such that $\ell_0\le\ell_0'$ and let $\mathcal{J}_{\ell_0,\ell_0'}\subset\{1,\ldots,J\}$ be such that, for every $\ell\in\{\ell_0,\ldots,\ell_0'\}$, $j_\ell\in\mathcal{J}_{\ell_0,\ell_0'}$. Then, under Assumptions 2.1 and 2.3, we have*

$$\sum_{j\in\mathcal{J}_{\ell_0,\ell_0'}}\|\nabla_j F(\boldsymbol{x}_{\ell_0'+1})+\boldsymbol{r}_{\ell_0'+1}^{(j)}\|^2$$

$$\le \left(L^2+\beta^2\overline{\nu}\right)\sum_{\ell=\ell_0}^{\ell_0'}2^{\ell_0'+1-\ell}\|\boldsymbol{x}_{\ell+1}-\boldsymbol{x}_\ell\|^2+2^{\ell_0'+1-\ell_0}\sum_{j\in\mathcal{J}_{\ell_0,\ell_0'}\setminus\{j_{\ell_0}\}}\|\nabla_j F(\boldsymbol{x}_{\ell_0})+\boldsymbol{r}_{\ell_0}^{(j)}\|^2,$$

where $\boldsymbol{r}_{\ell_0'+1}^{(j_{\ell_0'})}$ is defined by Algorithm (12), for every $j\in\mathcal{J}_{\ell_0,\ell_0'}\setminus\{j_{\ell_0'}\}$, $\boldsymbol{r}_{\ell_0'+1}^{(j)}\in\partial R_j(\boldsymbol{x}_{\ell_0'+1}^{(j)})$ and, for every $j\in\mathcal{J}_{\ell_0,\ell_0'}\setminus\{j_{\ell_0}\}$, $\boldsymbol{r}_{\ell_0}^{(j)}\in\partial R_j(\boldsymbol{x}_{\ell_0}^{(j)})$.

Proof. Let $(\ell_0,\ell_0')\in\mathbb{N}^2$ be such that $\ell_0\le\ell_0'$. Under the considered assumptions, by applying successively Lemma 3.3 for $\ell_0',\ell_0'-1,\ldots,\ell_0$, we have

$$\sum_{j\in\mathcal{J}_{\ell_0,\ell_0'}}\|\nabla_j F(\boldsymbol{x}_{\ell_0'+1})+\boldsymbol{r}_{\ell_0'+1}^{(j)}\|^2$$

$$\le \left(L^2+\beta^2\overline{\nu}\right)2\|\boldsymbol{x}_{\ell_0'+1}-\boldsymbol{x}_{\ell_0'}\|^2+2\sum_{j\in\mathcal{J}_{\ell_0,\ell_0'}\setminus\{j_{\ell_0'}\}}\|\nabla_j F(\boldsymbol{x}_{\ell_0'})+\boldsymbol{r}_{\ell_0'}^{(j)}\|^2$$

$$\le \left(L^2+\beta^2\overline{\nu}\right)2\|\boldsymbol{x}_{\ell_0'+1}-\boldsymbol{x}_{\ell_0'}\|^2+2\sum_{j\in\mathcal{J}_{\ell_0,\ell_0'}}\|\nabla_j F(\boldsymbol{x}_{\ell_0'})+\boldsymbol{r}_{\ell_0'}^{(j)}\|^2$$

$$\le \left(L^2+\beta^2\overline{\nu}\right)\left(2\|\boldsymbol{x}_{\ell_0'+1}-\boldsymbol{x}_{\ell_0'}\|^2+2^2\|\boldsymbol{x}_{\ell_0'}-\boldsymbol{x}_{\ell_0'-1}\|^2\right)$$
$$+2^2\sum_{j\in\mathcal{J}_{\ell_0,\ell_0'}\setminus\{j_{\ell_0'-1}\}}\|\nabla_j F(\boldsymbol{x}_{\ell_0'-1})+\boldsymbol{r}_{\ell_0'-1}^{(j)}\|^2$$

$$\le \left(L^2+\beta^2\overline{\nu}\right)\left(2\|\boldsymbol{x}_{\ell_0'+1}-\boldsymbol{x}_{\ell_0'}\|^2+2^2\|\boldsymbol{x}_{\ell_0'}-\boldsymbol{x}_{\ell_0'-1}\|^2+2^3\|\boldsymbol{x}_{\ell_0'-1}-\boldsymbol{x}_{\ell_0'-2}\|^2\right)$$
$$+2^3\sum_{j\in\mathcal{J}_{\ell_0,\ell_0'}\setminus\{j_{\ell_0'-2}\}}\|\nabla_j F(\boldsymbol{x}_{\ell_0'-2})+\boldsymbol{r}_{\ell_0'-2}^{(j)}\|^2$$

$$\vdots$$

$$\le \left(L^2+\beta^2\overline{\nu}\right)\sum_{\ell=\ell_0}^{\ell_0'}2^{\ell_0'+1-\ell}\|\boldsymbol{x}_{\ell+1}-\boldsymbol{x}_\ell\|^2+2^{\ell_0'+1-\ell_0}\sum_{j\in\mathcal{J}_{\ell_0,\ell_0'}\setminus\{j_{\ell_0}\}}\|\nabla_j F(\boldsymbol{x}_{\ell_0})+\boldsymbol{r}_{\ell_0}^{(j)}\|^2.$$

$\square$

Some notation will be needed in the remainder. Let $j\in\{1,\ldots,J\}$, let $\ell\in\mathbb{N}$, and let $K>0$ be defined by Assumption 2.4. We denote by

$$k_{\ell,j}=\min\left\{k\in\{0,\ldots,K-1\}\mid j_{\ell+k}=j\right\},\tag{22}$$

the first time the j-th block is updated after the ℓ-th iteration of Algorithm (12). Moreover, we define the permutation $\sigma_\ell\colon\{1,\ldots,J\}\to\{1,\ldots,J\}$ ensuring that $(k_{\ell,\sigma_\ell(i)})_{1\le i\le J}$ is increasing.

Our main result concerning the asymptotic behavior of Algorithm (12) is given below:

Theorem 3.1. *Let $(\boldsymbol{x}_\ell)_{\ell\in\mathbb{N}}$ be defined by (12). Under Assumptions 2.1-2.4, the following hold.*

(i) *The sequence $(\boldsymbol{x}_\ell)_{\ell\in\mathbb{N}}$ converges to a critical point $\widehat{\boldsymbol{x}}$ of G.*

(ii) *This sequence has a finite length in the sense that*

$$\sum_{\ell=0}^{+\infty} \|\boldsymbol{x}_{\ell+1} - \boldsymbol{x}_\ell\| < +\infty.$$

(iii) $\big(G(\boldsymbol{x}_\ell)\big)_{\ell \in \mathbb{N}}$ *is a nonincreasing sequence converging to* $G(\widehat{\boldsymbol{x}})$.

Proof. According to Lemma 3.1, we have

$$(\forall \ell \in \mathbb{N}) \quad G(\boldsymbol{x}_{\ell+1}) \leq G(\boldsymbol{x}_\ell),$$

thus, $(G(\boldsymbol{x}_\ell))_{\ell \in \mathbb{N}}$ is a nonincreasing sequence. In addition, since $\boldsymbol{x}_0 \in \operatorname{dom} R$, by Remark 2.3(iii), the sequence $(\boldsymbol{x}_\ell)_{\ell \in \mathbb{N}}$ belongs to the compact subset $E = \operatorname{lev}_{\leq G(\boldsymbol{x}_0)} G \subset \operatorname{dom} R$ and G is lower bounded. Thus, $\big(G(\boldsymbol{x}_\ell)\big)_{\ell \in \mathbb{N}}$ converges to a real ξ, and $\big(G(\boldsymbol{x}_\ell) - \xi\big)_{\ell \in \mathbb{N}}$ is a nonnegative sequence converging to 0.
Moreover, by invoking Lemma 3.2, we have

$$(\forall \ell \in \mathbb{N}) \quad \frac{\mu}{2}\|\boldsymbol{\chi}_\ell\|^2 \leq (G(\boldsymbol{x}_\ell) - \xi) - (G(\boldsymbol{x}_{\ell+K}) - \xi). \tag{23}$$

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

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

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

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

Using the latter inequality with $u = G(\boldsymbol{x}_\ell) - \xi$ and $v = G(\boldsymbol{x}_{\ell+K}) - \xi$ leads to

$$(\forall \ell \in \mathbb{N}) \quad \big(G(\boldsymbol{x}_\ell) - \xi\big) - \big(G(\boldsymbol{x}_{\ell+K}) - \xi\big) \leq (1 - \theta)^{-1} \big(G(\boldsymbol{x}_\ell) - \xi\big)^\theta \Delta_\ell,$$

where

$$(\forall \ell \in \mathbb{N}) \quad \Delta_\ell = \big(G(\boldsymbol{x}_\ell) - \xi\big)^{1-\theta} - \big(G(\boldsymbol{x}_{\ell+K}) - \xi\big)^{1-\theta}.$$

Thus, combining the above inequality with (23) yields

$$(\forall \ell \in \mathbb{N}) \quad \|\boldsymbol{\chi}_\ell\|^2 \leq 2\mu^{-1}(1-\theta)^{-1}\big(G(\boldsymbol{x}_\ell) - \xi\big)^\theta \Delta_\ell. \tag{24}$$

Let us define

$$(\forall \ell \in \mathbb{N}) \qquad \boldsymbol{t}_\ell = \Big(\nabla_j F(\boldsymbol{x}_\ell) + \boldsymbol{r}_\ell^{(j)}\Big)_{1 \leq j \leq J} \in \mathbb{R}^{N_1} \times \ldots \times \mathbb{R}^{N_J},$$

where, for every $j \in \{1, \ldots, J\}$, $\boldsymbol{r}_\ell^{(j)} \in \partial R_j(\boldsymbol{x}_\ell^{(j)})$. Using the differentiation rule for separable functions, we have $\boldsymbol{r}_\ell = \big(\boldsymbol{r}_\ell^{(j)}\big)_{1 \leq j \leq J} \in \partial R(\boldsymbol{x}_\ell)$. Thus, for every $\ell \in \mathbb{N}$,

$$\boldsymbol{t}_\ell \in \partial G(\boldsymbol{x}_\ell). \tag{25}$$

Since E is bounded and Assumption 2.2 holds, there exist constants $\kappa > 0$, $\zeta > 0$ and $\theta \in [0,1)$ such that (10) holds for every $\boldsymbol{x} \in E$ for which the inequality $|G(\boldsymbol{x}) - \xi| \leq \zeta$ is satisfied. Since $\big(G(\boldsymbol{x}_\ell)\big)_{\ell \in \mathbb{N}}$ converges to ξ, there exists $\ell^* \in \mathbb{N}$, such that, for every $\ell \geq \ell^*$, $|G(\boldsymbol{x}_\ell) - \xi| < \zeta$. Hence, we have

$$(\forall \ell \geq \ell^*) \quad \kappa |G(\boldsymbol{x}_\ell) - \xi|^\theta \leq \|\boldsymbol{t}_\ell\|. \tag{26}$$

Let K be defined by Assumption 2.4. For every $\ell \in \mathbb{N}$,

$$\|\boldsymbol{t}_{\ell+K}\|^2 = \left\| \left(\nabla_j F(\boldsymbol{x}_{\ell+K}) + \boldsymbol{r}_{\ell+K}^{(j)} \right)_{1 \le j \le J} \right\|^2 = \sum_{j=1}^{J} \left\| \nabla_j F(\boldsymbol{x}_{\ell+K}) + \boldsymbol{r}_{\ell+K}^{(j)} \right\|^2.$$

For every $k \in \{\ell + k_{\ell,\sigma_\ell(J)}, \ldots, \ell + K - 1\}$, let $\boldsymbol{r}_{k+1}^{(j_k)} \in \partial R_{j_k}(\boldsymbol{x}_{k+1}^{(j_k)})$ be defined as in Algorithm (12). Thus, Lemma 3.4 with $\ell_0 = \ell + k_{\ell,\sigma_\ell(J)}$, $\ell_0' = \ell + K - 1$ and $\mathcal{J}_{\ell_0,\ell_0'} = \{1, \ldots, J\}$ leads to

$$\|\boldsymbol{t}_{\ell+K}\|^2 \le (L^2 + \beta^2 \overline{\nu}) \sum_{k=\ell+k_{\ell,\sigma_\ell(J)}}^{\ell+K-1} 2^{\ell+K-k} \|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|^2$$

$$+ 2^{K-k_{\ell,\sigma_\ell(J)}} \sum_{\substack{j=1 \\ j \ne \sigma_\ell(J)}}^{J} \|\nabla_j F(\boldsymbol{x}_{\ell+k_{\ell,\sigma_\ell(J)}}) + \boldsymbol{r}_{\ell+k_{\ell,\sigma_\ell(J)}}^{(j)}\|^2.$$

Using again Lemma 3.4 on $\sum_{\substack{j=1 \\ j \ne \sigma_\ell(J)}}^{J} \|\nabla_j F(\boldsymbol{x}_{\ell+k_{\ell,\sigma_\ell(J)}}) + \boldsymbol{r}_{\ell+k_{\ell,\sigma_\ell(J)}}^{(j)}\|^2$ with $\ell_0 = \ell + k_{\ell,\sigma_\ell(J-1)}$, $\ell_0' = \ell + k_{\ell,\sigma_\ell(J)} - 1$ and $\mathcal{J}_{\ell_0,\ell_0'} = \{1, \ldots, J\} \setminus \{\sigma_\ell(J)\}$, we obtain

$$\|\boldsymbol{t}_{\ell+K}\|^2 \le (L^2 + \beta^2 \overline{\nu}) \sum_{k=\ell+k_{\ell,\sigma_\ell(J)}}^{\ell+K-1} 2^{\ell+K-k} \|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|^2$$

$$+ (L^2 + \beta^2 \overline{\nu}) \sum_{k=\ell+k_{\ell,\sigma_\ell(J-1)}}^{\ell+k_{\ell,\sigma_\ell(J)}-1} 2^{\ell+K-k} \|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|^2$$

$$+ 2^{K-k_{\ell,\sigma_\ell(J-1)}} \sum_{\substack{j=1 \\ j \ne \sigma_\ell(i), i \in \{J-1, J\}}}^{J} \|\nabla_j F(\boldsymbol{x}_{\ell+k_{\ell,\sigma_\ell(J-1)}}) + \boldsymbol{r}_{\ell+k_{\ell,\sigma_\ell(J-1)}}^{(j)}\|^2.$$

Proceeding similarly for $i \in \{1, \ldots, J-2\}$, we get

$$\|\boldsymbol{t}_{\ell+K}\|^2 \le (L^2 + \beta^2 \overline{\nu}) \sum_{k=\ell+k_{\ell,\sigma_\ell(J)}}^{\ell+K-1} 2^{\ell+K-k} \|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|^2$$

$$+ (L^2 + \beta^2 \overline{\nu}) \sum_{i=1}^{J-1} \sum_{k=\ell+k_{\ell,\sigma_\ell(j)}}^{\ell+k_{\ell,\sigma_\ell(j+1)}-1} 2^{\ell+K-k} \|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|^2, \tag{27}$$

where we have used the fact that $\{1, \ldots, J\} \setminus \{\sigma_\ell(1), \ldots, \sigma_\ell(J)\} = \varnothing$, thus

$$\sum_{\substack{j=1 \\ j \ne \sigma_\ell(i), i \in \{1, \ldots, J\}}}^{J} \|\nabla_j F(\boldsymbol{x}_\ell) + \boldsymbol{r}_\ell^{(j)}\|^2 = 0.$$

Since $k_{\ell,\sigma_\ell(1)} = 0$ and, for every $k \in \{\ell, \ldots, \ell+K-1\}$, $2^{\ell+K-k} \le 2^K$, it follows from (17) and (27) that

$$\|\boldsymbol{t}_{\ell+K}\|^2 \le 2^K (L^2 + \beta^2 \overline{\nu}) \sum_{k=\ell}^{\ell+K-1} \|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|^2 = 2^K (L^2 + \beta^2 \overline{\nu}) \|\boldsymbol{\chi}_\ell\|^2. \tag{28}$$

Combining (24), (26) and (28) yields

$$(\forall \ell \geq \max\{\ell^*, K\}) \quad \|\boldsymbol{\chi}_\ell\|^2 \leq 2\mu^{-1}(1-\theta)^{-1}\kappa^{-1}2^{K/2}(L^2 + \beta^2\overline{\nu})^{1/2}\|\boldsymbol{\chi}_{\ell-K}\|\Delta_\ell.$$

By using the fact that

$$(\forall (u,v) \in [0,+\infty)^2) \quad (uv)^{1/2} \leq \frac{1}{2}(u+v),$$

and by setting $u = \|\boldsymbol{\chi}_{\ell-K}\|$ and $v = 2\mu^{-1}(1-\theta)^{-1}\kappa^{-1}2^{K/2}(L^2 + \beta^2\overline{\nu})^{1/2}\Delta_\ell$, we obtain

$$(\forall \ell \geq \max\{\ell^*, K\}) \quad \|\boldsymbol{\chi}_\ell\| \leq \frac{1}{2}\|\boldsymbol{\chi}_{\ell-K}\| + \mu^{-1}(1-\theta)^{-1}\kappa^{-1}2^{K/2}(L^2 + \beta^2\overline{\nu})^{1/2}\Delta_\ell. \tag{29}$$

Furthermore, it can be noticed that

$$\sum_{\ell=\ell^*}^{+\infty} \Delta_\ell = \sum_{\ell=\ell^*}^{+\infty} \left(G(\boldsymbol{x}_\ell) - \xi\right)^{1-\theta} - \left(G(\boldsymbol{x}_{\ell+K}) - \xi\right)^{1-\theta}$$
$$= \sum_{\ell=\ell^*}^{\ell^*+K-1} \left(G(\boldsymbol{x}_\ell) - \xi\right)^{1-\theta},$$

which shows that $(\Delta_\ell)_{\ell\in\mathbb{N}}$ is a summable sequence. As $(\|\boldsymbol{\chi}_\ell\|)_{\ell\geq\max\{\ell^*,K\}}$ satisfies inequality (29), $(\|\boldsymbol{\chi}_\ell\|)_{\ell\in\mathbb{N}}$ is also a summable sequence. According to (17),

$$(\forall \ell \in \mathbb{N}) \quad \|\boldsymbol{x}_{\ell+1} - \boldsymbol{x}_\ell\| \leq \|\boldsymbol{\chi}_\ell\|,$$

and $(\|\boldsymbol{x}_{\ell+1} - \boldsymbol{x}_\ell\|)_{\ell\in\mathbb{N}}$ is a summable sequence.

Hence, the sequence $(\boldsymbol{x}_\ell)_{\ell\in\mathbb{N}}$ satisfies the finite length property. In addition, since this latter condition implies that $(\boldsymbol{x}_\ell)_{\ell\in\mathbb{N}}$ is a Cauchy sequence, it converges towards a point $\widehat{\boldsymbol{x}}$.

It remains us to show that the limit $\widehat{\boldsymbol{x}}$ is a critical point of G. According to (25), we have, for every $\ell \in \mathbb{N}$,

$$(\boldsymbol{x}_\ell, \boldsymbol{t}_\ell) \in \text{Graph}\,\partial G.$$

In addition, since the sequence $(\|\boldsymbol{\chi}_\ell\|)_{\ell\in\mathbb{N}}$ is summable, it converges to 0. Moreover, according to (28), we have

$$\|\boldsymbol{t}_\ell\| \leq 2^{K/2}(L^2 + \beta^2\overline{\nu})^{1/2}\|\boldsymbol{\chi}_{\ell-K}\|,$$

hence $(\boldsymbol{x}_\ell, \boldsymbol{t}_\ell)_{\ell\in\mathbb{N}}$ converges to $(\widehat{\boldsymbol{x}}, \boldsymbol{0})$. Furthermore, according to Remark 2.3(iii), the restriction of G to its domain is continuous. Thus, as, for every $\ell \in \mathbb{N}$, $\boldsymbol{x}_\ell \in \text{dom}\,G$, the sequence $(G(\boldsymbol{x}_\ell))_{\ell\in\mathbb{N}}$ converges to $G(\widehat{\boldsymbol{x}})$. Finally, according to the closedness property of ∂G (see Remark 2.1), $(\widehat{\boldsymbol{x}}, \boldsymbol{0}) \in \text{Graph}\,\partial G$ i.e., $\widehat{\boldsymbol{x}}$ is a critical point of G. $\qquad\square$

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

Corollary 3.1. *Suppose that $(\boldsymbol{x}_\ell)_{\ell\in\mathbb{N}}$ is a sequence generated by Algorithm (12), and suppose that Assumptions 2.1-2.4 hold. There exists $\upsilon \in (0, +\infty)$ such that, if*

$$G(\boldsymbol{x}_0) \leq \inf_{\boldsymbol{x}\in\mathbb{R}^N} G(\boldsymbol{x}) + \upsilon,$$

then $(\boldsymbol{x}_\ell)_{\ell\in\mathbb{N}}$ converges to a solution to Problem (1).

Proof. Same proof as in [16, Cor. 3.2]. $\qquad\square$

3.3 Convergence rate

According to Theorem 3.1, the limit $\widehat{\boldsymbol{x}}$ of a sequence $(\boldsymbol{x}_\ell)_{\ell \in \mathbb{N}}$ generated by Algorithm (12) is a critical point of G, under Assumptions 2.1-2.4. Thus, proceeding similarly to the derivation of (26), there exists $\zeta \in (0, +\infty)$ such that for every $\boldsymbol{x} \in \mathbb{R}^N$ with $G(\boldsymbol{x}) \leq G(\widehat{\boldsymbol{x}}) + \zeta$, (10) is satisfied for some $\kappa \in (0, +\infty)$ and $\theta \in [0, 1)$. The number θ is then called a *Łojasiewicz exponent of G at $\widehat{\boldsymbol{x}}$*. Similarly to other algorithms based on Kurdyka-Łojasiewicz inequality [1, 2], the local convergence rate of the BC-VMFB algorithm depends on this exponent.

The following lemma is instrumental to establish the convergence rate:

Lemma 3.5. [1] *Let $(\Lambda_m)_{m \in \mathbb{N}}$ be a nonnegative sequence of reals decreasing to 0. Assume that there exist $m^* \in \mathbb{N}^*$ and $C \in (0, +\infty)$ such that, for every $m \geq m^*$,*

$$\Lambda_m \leq (\Lambda_{m-1} - \Lambda_m) + C(\Lambda_{m-1} - \Lambda_m)^{\frac{1-\theta}{\theta}}, \tag{30}$$

where $\theta \in (0, 1)$.
If $\theta \in (\frac{1}{2}, 1)$, then there exists $\lambda \in (0, +\infty)$ such that

$$(\forall m \geq 1) \qquad \Lambda_m \leq \lambda m^{-\frac{1-\theta}{2\theta - 1}}.$$

If $\theta \in (0, \frac{1}{2}]$, then there exist $\lambda \in (0, +\infty)$ and $\tau \in [0, 1)$ such that

$$(\forall m \in \mathbb{N}) \qquad \Lambda_m \leq \lambda \tau^m.$$

Theorem 3.2. *Let $(\boldsymbol{x}_\ell)_{\ell \in \mathbb{N}}$ be a sequence generated by Algorithm (12) and suppose that Assumptions 2.1-2.4 hold. Let θ be a Łojasiewicz exponent of G at the limit point $\widehat{\boldsymbol{x}}$ of $(\boldsymbol{x}_\ell)_{\ell \in \mathbb{N}}$. The following properties hold:*

(i) *If $\theta \in (\frac{1}{2}, 1)$, then there exists $(\lambda', \lambda'') \in (0, +\infty)^2$ such that*

$$(\forall \ell > K) \qquad \|\boldsymbol{x}_\ell - \widehat{\boldsymbol{x}}\| \leq \lambda'\Big(\frac{\ell}{K} - 1\Big)^{-\frac{1-\theta}{2\theta - 1}}, \tag{31}$$

$$(\forall \ell > 2K) \qquad G(\boldsymbol{x}_\ell) - G(\widehat{\boldsymbol{x}}) \leq \lambda''\Big(\frac{\ell}{K} - 2\Big)^{-\frac{1-\theta}{\theta(2\theta - 1)}}. \tag{32}$$

(ii) *If $\theta \in (0, \frac{1}{2}]$, then there exist $(\lambda', \lambda'') \in (0, +\infty)^2$ and $\tau' \in [0, 1)$ such that*

$$(\forall \ell \in \mathbb{N}) \qquad \|\boldsymbol{x}_\ell - \widehat{\boldsymbol{x}}\| \leq \lambda'(\tau')^\ell, \tag{33}$$

$$G(\boldsymbol{x}_\ell) - G(\widehat{\boldsymbol{x}}) \leq \lambda''(\tau')^{\frac{\ell}{\theta}}. \tag{34}$$

(iii) *If $\theta = 0$, then the sequence $(\boldsymbol{x}_\ell)_{\ell \in \mathbb{N}}$ converges in a finite number of steps.*

Proof. We use the same notation as in the proof of Theorem 3.1. Let K be given by Assumption 2.4. For every $\ell \in \mathbb{N}$, there exist $m \in \mathbb{N}$ and $k \in \{0, \ldots, K-1\}$ such that $\ell = mK + k$. Then, according to the triangle inequality,

$$\|\boldsymbol{x}_\ell - \widehat{\boldsymbol{x}}\| \leq \|\boldsymbol{x}_{mK} - \widehat{\boldsymbol{x}}\| + \|\boldsymbol{x}_\ell - \boldsymbol{x}_{mK}\|. \tag{35}$$

Moreover, using again the triangle inequality, we have

$$\|\boldsymbol{x}_{mK} - \widehat{\boldsymbol{x}}\| = \Big\| \sum_{p=m}^{+\infty} \big(\boldsymbol{x}_{(p+1)K} - \boldsymbol{x}_{pK}\big) \Big\|$$

$$= \Big\| \sum_{p=m}^{+\infty} \sum_{k'=0}^{K-1} \big(\boldsymbol{x}_{pK+k'+1} - \boldsymbol{x}_{pK+k'}\big) \Big\|$$

$$\leq \sum_{p=m}^{+\infty} \Big\| \sum_{k'=0}^{K-1} \big(\boldsymbol{x}_{pK+k'+1} - \boldsymbol{x}_{pK+k'}\big) \Big\|, \tag{36}$$

and according to Jensen's inequality and (17),

$$(\forall p \geq m) \quad \big\| \sum_{k'=0}^{K-1} \big(\boldsymbol{x}_{pK+k'+1} - \boldsymbol{x}_{pK+k'}\big) \big\|^2 \leq K\|\boldsymbol{\chi}_{pK}\|^2. \tag{37}$$

For every $m' \in \mathbb{N}$, let $\Lambda_{m'} = \sum_{p=m'}^{+\infty} \|\boldsymbol{\chi}_{pK}\|$ which is finite by Theorem 3.1. Hence, the last two inequalities yield

$$\|\boldsymbol{x}_{mK} - \widehat{\boldsymbol{x}}\| \leq \sqrt{K}\Lambda_m. \tag{38}$$

Involving again Jensen's inequality, we have

$$\|\boldsymbol{x}_{mK} - \boldsymbol{x}_\ell\|^2 = \big\| \sum_{k'=0}^{k-1} \big(\boldsymbol{x}_{mK+k'+1} - \boldsymbol{x}_{mK+k'}\big) \big\|^2$$
$$\leq k \sum_{k'=0}^{k-1} \big\|\boldsymbol{x}_{mK+k'+1} - \boldsymbol{x}_{mK+k'}\big\|^2 \leq (K-1)\|\boldsymbol{\chi}_{mK}\|^2. \tag{39}$$

Altogether, (35), (38), and (39) lead to

$$(\forall \ell \in \mathbb{N}) \quad \|\boldsymbol{x}_\ell - \widehat{\boldsymbol{x}}\| \leq \sqrt{K}\Lambda_m + \sqrt{K-1}\|\boldsymbol{\chi}_{mK}\| \leq 2\sqrt{K}\Lambda_m. \tag{40}$$

Using (29), we have, for every $m \geq \max\{\ell^*/K, 1\}$,

$$\|\boldsymbol{\chi}_{mK}\| \leq \frac{1}{2}\|\boldsymbol{\chi}_{(m-1)K}\| + \mu^{-1}(1-\theta)^{-1}\kappa^{-1}2^{K/2}(L^2 + \beta^2\overline{\nu})^{1/2}\Delta_{mK},$$

where $\Delta_{mK} = \big(G(\boldsymbol{x}_{mK}) - G(\widehat{\boldsymbol{x}})\big)^{1-\theta} - \big(G(\boldsymbol{x}_{(m+1)K}) - G(\widehat{\boldsymbol{x}})\big)^{1-\theta}$. Thus, since $\big(G(\boldsymbol{x}_\ell) - G(\widehat{\boldsymbol{x}})\big)_{\ell \in \mathbb{N}}$ is a nonnegative sequence converging to 0, we obtain

$$\Lambda_m \leq (\Lambda_{m-1} - \Lambda_m) + 2\mu^{-1}(1-\theta)^{-1}\kappa^{-1}2^{K/2}(L^2 + \beta^2\overline{\nu})^{1/2}\big(G(\boldsymbol{x}_{mK}) - G(\widehat{\boldsymbol{x}})\big)^{1-\theta}.$$

Let us now assume that $\theta \neq 0$. According to (26) and (28), we have

$$\kappa\,(G(\boldsymbol{x}_{mK}) - G(\widehat{\boldsymbol{x}}))^\theta \leq \big(2^K(L^2 + \beta^2\overline{\nu})\big)^{1/2}\|\boldsymbol{\chi}_{(m-1)K}\|,$$

so that

$$\big(G(\boldsymbol{x}_{mK}) - G(\widehat{\boldsymbol{x}})\big)^{1-\theta} \leq \kappa^{-\frac{1-\theta}{\theta}}\big(2^K(L^2 + \beta^2\overline{\nu})\big)^{\frac{1-\theta}{2\theta}}\|\boldsymbol{\chi}_{(m-1)K}\|^{\frac{1-\theta}{\theta}}. \tag{41}$$

Thus, by defining

$$C = 2\mu^{-1}(1-\theta)^{-1}\kappa^{-\frac{1}{\theta}}\big(2^K(L^2 + \beta^2\overline{\nu})\big)^{\frac{1}{2\theta}}, \tag{42}$$

we get, for every $m \geq \max\{\ell^*/K, 1\}$,

$$\Lambda_m \leq (\Lambda_{m-1} - \Lambda_m) + C\|\boldsymbol{\chi}_{(m-1)K}\|^{\frac{1-\theta}{\theta}},$$

and (30) is satisfied.

Thus, according to Lemma 3.5 and (40), if $\theta \in (\frac{1}{2}, 1)$, there exists $\lambda \in (0, +\infty)$ such that

$$(\forall \ell > K) \quad \|\boldsymbol{x}_\ell - \widehat{\boldsymbol{x}}\| \leq 2\sqrt{K}\lambda m^{-\frac{1-\theta}{2\theta-1}} \leq 2\sqrt{K}\lambda\big(\frac{\ell}{K} - 1\big)^{-\frac{1-\theta}{2\theta-1}},$$

where m is the lower integer part of ℓ/K. Inequality (31) is thus obtained by setting $\lambda' = 2\sqrt{K}\lambda$. Similarly, if $\theta \in (0, \frac{1}{2}]$, then there exist $\lambda \in (0, +\infty)$ and $\tau \in [0, 1)$ such that

$$(\forall \ell > K) \quad \|\boldsymbol{x}_\ell - \widehat{\boldsymbol{x}}\| \leq 2\sqrt{K}\lambda\tau^m \leq 2\sqrt{K}\lambda\tau^{\ell/K-1}.$$

Hence, if $\tau \neq 0$, (33) is satisfied by setting $\lambda' = 2\sqrt{K}\lambda/\tau$ and $\tau' = \tau^{1/K}$, while (33) also holds trivially when $\tau = 0$.

In addition, since $\big(G(\boldsymbol{x}_\ell) - G(\widehat{\boldsymbol{x}})\big)_{\ell \in \mathbb{N}}$ is a decreasing sequence, for every $\ell \in \mathbb{N}$,

$$G(\boldsymbol{x}_\ell) - G(\widehat{\boldsymbol{x}}) \leq G(\boldsymbol{x}_{mK}) - G(\widehat{\boldsymbol{x}}),$$

where m still denotes the lower integer part of ℓ/K. Using (41), if $m \geq \max\{\ell^*/K, 1\}$, then

$$G(\boldsymbol{x}_\ell) - G(\widehat{\boldsymbol{x}}) \leq \kappa^{-1/\theta} \left(2^K(L^2 + \beta^2\overline{\nu})\right)^{\frac{1}{2\theta}} \|\boldsymbol{\chi}_{(m-1)K}\|^{1/\theta}$$
$$\leq \kappa^{-1/\theta} \left(2^K(L^2 + \beta^2\overline{\nu})\right)^{\frac{1}{2\theta}} \Lambda_{m-1}^{1/\theta}.$$

So, if $\theta \in (\frac{1}{2}, 1)$, using again Lemma 3.5, there exists $\lambda \in (0, +\infty)$ such that, when $m > 2$,

$$G(\boldsymbol{x}_\ell) - G(\widehat{\boldsymbol{x}}) \leq \kappa^{-1/\theta} \left(2^K(L^2 + \beta^2\overline{\nu})\right)^{\frac{1}{2\theta}} \lambda(m-1)^{-\frac{1-\theta}{\theta(2\theta-1)}}$$
$$\leq \kappa^{-1/\theta} \left(2^K(L^2 + \beta^2\overline{\nu})\right)^{\frac{1}{2\theta}} \lambda\Big(\frac{\ell}{K} - 2\Big)^{-\frac{1-\theta}{\theta(2\theta-1)}}.$$

Hence, one can find $\lambda'' \in (0, +\infty)$ such that (32) holds for every $\ell > 2K$.

If $\theta \in (0, \frac{1}{2}]$, there exist $\lambda \in (0, +\infty)$ and $\tau \in [0, 1)$ such that

$$G(\boldsymbol{x}_\ell) - G(\widehat{\boldsymbol{x}}) \leq \kappa^{-1/\theta} \left(2^K(L^2 + \beta^2\overline{\nu})\right)^{\frac{1}{2\theta}} \lambda\tau^{\frac{m-1}{\theta}}$$
$$\leq \kappa^{-1/\theta} \left(2^K(L^2 + \beta^2\overline{\nu})\right)^{\frac{1}{2\theta}} \lambda\tau^{\frac{\ell/K-2}{\theta}}.$$

Therefore, one can find $\lambda'' \in (0, +\infty)$ such that (34) holds for every $\ell \in \mathbb{N}$.

Let us now prove Property (iii) by assuming that $\theta = 0$. Set $\mathcal{L} = \{\ell \in \mathbb{N} \,|\, \boldsymbol{x}_\ell \neq \widehat{\boldsymbol{x}}\}$, and let $\ell \geq \max\{\ell^*, K\}$ be in $\mathcal{L}$. According to Lemmas 3.1 and 3.2,

$$G(\boldsymbol{x}_{\ell+1}) \leq G(\boldsymbol{x}_\ell) - \frac{\mu}{2}\|\boldsymbol{x}_{\ell+1} - \boldsymbol{x}_\ell\|^2 \leq G(\boldsymbol{x}_{\ell-K}) - \frac{\mu}{2}\|\boldsymbol{\chi}_{\ell-K}\|^2.$$

Using (28), we obtain

$$G(\boldsymbol{x}_\ell) - G(\widehat{\boldsymbol{x}}) - \frac{\mu}{2}\|\boldsymbol{x}_{\ell+1} - \boldsymbol{x}_\ell\|^2 \leq G(\boldsymbol{x}_{\ell-K}) - G(\widehat{\boldsymbol{x}}) - \frac{\mu'}{2}\|\boldsymbol{t}_\ell\|^2,$$

where $\mu' \in (0, +\infty)$. Combined with (26), and since $\theta = 0$, this yields

$$G(\boldsymbol{x}_\ell) - G(\widehat{\boldsymbol{x}}) - \frac{\mu}{2}\|\boldsymbol{x}_{\ell+1} - \boldsymbol{x}_\ell\|^2 \leq G(\boldsymbol{x}_{\ell-K}) - G(\widehat{\boldsymbol{x}}) - \frac{\mu'}{2}\kappa^2|G(\boldsymbol{x}_\ell) - G(\widehat{\boldsymbol{x}})|^0,$$

that is,

$$G(\boldsymbol{x}_\ell) - G(\widehat{\boldsymbol{x}}) - \frac{\mu}{2}\|\boldsymbol{x}_{\ell+1} - \boldsymbol{x}_\ell\|^2 \leq G(\boldsymbol{x}_{\ell-K}) - G(\widehat{\boldsymbol{x}}) - \frac{\mu'}{2}\kappa^2.$$

Since $\lim\limits_{\ell \to +\infty} G(\boldsymbol{x}_\ell) = G(\widehat{\boldsymbol{x}})$, the above inequality implies that $\mathcal{L}$ is finite, and (iii) follows. $\qquad\square$

Remark 3.1.

(i) *Note that, when G is strongly convex, the Łojasiewicz exponent θ of G is equal to $1/2$. In this case, $\widehat{x}$ is a global minimizer of G and sequences $(\|x_\ell - \widehat{x}\|)_{\ell \in \mathbb{N}}$ and $(G(x_\ell) - G(\widehat{x}))_{\ell \in \mathbb{N}}$ converge linearly.*

(ii) *Note that, if $\theta \in (0, 1/2]$, then, for m large enough, (30) yields*

$$\Lambda_m \le (1 + C)(\Lambda_{m-1} - \Lambda_m),$$

so that the constant τ' in (33)-(34) can be chosen equal to $((1 + C)/(2 + C))^{1/K}$ where C is given by (42).

4 Application

4.1 Optimization problem

In this section, we consider a phase retrieval inverse problem which consists of estimating the phase of a complex-valued signal from measurements of its modulus and additional a priori information.

Let $z \in [0, +\infty)^S$ be a degraded signal related to an original unknown signal $\overline{v} \in \mathbb{R}^M$ through the model

$$z = |H\overline{v}| + w,$$

where $H \in \mathbb{C}^{S \times M}$ is an observation matrix with complex elements, $|\cdot|$ denotes the componentwise modulus operator, and $w \in [0, +\infty)^S$ is a realization of an additive noise. The objective is then to find an estimate $\widehat{v} \in \mathbb{R}^M$ of the target image $\overline{v}$ from the observed data z and the observation operator H.

Such a problem is of paramount importance in numerous areas of applied physics and engineering [6, 14, 20, 42, 46]. Note that unlike many existing works [5, 14, 22, 23], it is not assumed that H is a Fourier transform matrix.

Set $\widehat{v} = W\widehat{x}$ where $W \in \mathbb{R}^{M \times N}$, $N \ge M$, is a given frame synthesis operator (e.g. a possibly redundant wavelet synthesis operator) [32]. Then, following a synthesis approach, the frame coefficient vector $\widehat{x}$ can be estimated by solving Problem (1) where F is the so-called data fidelity term of the form:

$$(\forall x \in \mathbb{R}^N) \quad F(x) := \sum_{s=1}^{S} \varphi^{(s)}(|[HWx]^{(s)}|). \tag{43}$$

Hereabove, for every $s \in \{1, \ldots, S\}$, $\varphi^{(s)} \colon [0, +\infty) \to \mathbb{R}$, and $[HWx]^{(s)}$ is the s-th component of $HWx \in \mathbb{C}^S$. Moreover, in (1), a penalty function R is employed serving to incorporate a priori information on the frame coefficients.

We propose to choose, for every $s \in \{1, \ldots, S\}$, $\varphi^{(s)} := \varphi_1^{(s)} + \varphi_2^{(s)}$, where

$$\left(\forall \omega \in [0, +\infty)\right) \quad \varphi_1^{(s)}(\omega) := \frac{1}{2}\left(\omega^2 + (z^{(s)})^2\right), \tag{44}$$

$$\varphi_2^{(s)}(\omega) := -z^{(s)}\left(\omega^2 + \delta^2\right)^{1/2}, \tag{45}$$

with $\delta > 0$. Thus, the data fidelity term (43) is split as $F = F_1 + F_2$ where

$$(\forall x \in \mathbb{R}^N) \quad \begin{aligned} F_1(x) &:= \sum_{s=1}^{S} \varphi_1^{(s)}(|[HWx]^{(s)}|), \\ F_2(x) &:= \sum_{s=1}^{S} \varphi_2^{(s)}(|[HWx]^{(s)}|). \end{aligned} \tag{46}$$

For every $s \in \{1, \dots, S\}$, the second-order derivatives of $\varphi_1^{(s)}$ and $\varphi_2^{(s)}$ are[1]

$$(\forall \omega \in [0, +\infty)) \quad \ddot{\varphi}_1^{(s)}(\omega) = 1, \tag{47}$$

$$\ddot{\varphi}_2^{(s)}(\omega) = -z^{(s)} \delta^2 (\omega^2 + \delta^2)^{-3/2}. \tag{48}$$

Thus, $\varphi_2^{(s)}$ is concave on $[0, +\infty)$, while $\varphi^{(s)}$ is nonconvex. Moreover, $\varphi^{(s)}$ is Lipschitz differentiable, and Assumption 2.1(ii) is satisfied. Note that, in the limit case when $\delta = 0$, the usual nonconvex nonsmooth least squares data fidelity term [22] is recovered (i.e. $F = \frac{1}{2}\||\boldsymbol{HW} \cdot| - \boldsymbol{z}\|^2$), which shows that the proposed function can be viewed as a smoothed version of it.

In addition, the following penalization term is employed:

$$(\forall \boldsymbol{x} = (x^{(n)})_{1 \le n \le N} \in \mathbb{R}^N) \quad R(\boldsymbol{x}) := \sum_{n=1}^{N} \rho^{(n)}(x^{(n)}), \tag{49}$$

where, for every $n \in \{1, \dots, N\}$,

$$(\forall \omega \in \mathbb{R}) \quad \rho^{(n)}(\omega) := \begin{cases} \vartheta_n |\omega - \overline{\omega}_n|^{\pi_n} & \text{if } \underline{\eta}_n \le \omega \le \overline{\eta}_n, \\ +\infty & \text{otherwise,} \end{cases} \tag{50}$$

and, for every $n \in \{1, \dots, N\}$, $\vartheta_n \in (0, +\infty)$, $\pi_n \in \mathbb{N}^*$, $\underline{\eta}_n \in [-\infty, +\infty)$, $\overline{\eta}_n \in [\underline{\eta}_n, +\infty]$, and $\overline{\omega}_n \in \mathbb{R}$. Assumption 2.1 is thus satisfied. Moreover, since for every $n \in \{1, \dots, N\}$, $\rho^{(n)}$ is a semi-algebraic function, F is also a semi-algebraic function, and Assumption 2.2 holds.

In the following, in order to simplify the notation, we introduce the linear operator $\boldsymbol{T} := \boldsymbol{HW} = (T^{(s,n)})_{1 \le s \le S, 1 \le n \le N} \in \mathbb{C}^{S \times N}$.

4.2 Construction of the preconditioning matrices

The numerical efficiency of the proposed method relies on the use of quadratic majorants providing good approximations of $F_{j_\ell}(\cdot, \boldsymbol{x}_\ell^{(\overline{j}_\ell)})$ at iteration $\ell \in \mathbb{N}$, and whose curvature matrices $(\boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell))_{\ell \in \mathbb{N}}$ are simple to compute.

Similarly to (4), let us define, for every $\ell \in \mathbb{N}$, functions $F_{1,j_\ell}(\cdot, \boldsymbol{x}_\ell^{\overline{j}_\ell})$ and $F_{2,j_\ell}(\cdot, \boldsymbol{x}_\ell^{\overline{j}_\ell})$ associated with F_1 and F_2, respectively. It has already been noticed that, for every $s \in \{1, \dots, S\}$, $\varphi_2^{(s)}$ is concave. Hence, for every $\ell \in \mathbb{N}$, $F_{2,j_\ell}(\cdot, \boldsymbol{x}_\ell^{\overline{j}_\ell})$ is majorized by

$$(\forall \boldsymbol{y} \in \mathbb{R}^{N_{j_\ell}}) \quad Q_{2,j_\ell}(\boldsymbol{y}, \boldsymbol{x}_\ell) := F_2(\boldsymbol{x}_\ell) + \left\langle \boldsymbol{y} - \boldsymbol{x}_\ell^{(j_\ell)}, \nabla_{j_\ell} F_2(\boldsymbol{x}_\ell) \right\rangle. \tag{51}$$

Thus, there remains to find a family of SPD matrices $(\boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell))_{\ell \in \mathbb{N}}$ such that, for every $\ell \in \mathbb{N}$,

$$(\forall \boldsymbol{y} \in \mathbb{R}^{N_{j_\ell}}) \quad Q_{1,j_\ell}(\boldsymbol{y}, \boldsymbol{x}_\ell) := F_1(\boldsymbol{x}_\ell) + \left\langle \boldsymbol{y} - \boldsymbol{x}_\ell^{(j_\ell)}, \nabla_{j_\ell} F_1(\boldsymbol{x}_\ell) \right\rangle$$
$$+ \frac{1}{2} \left\langle \boldsymbol{y} - \boldsymbol{x}_\ell^{(j_\ell)}, \boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell)(\boldsymbol{y} - \boldsymbol{x}_\ell^{(j_\ell)}) \right\rangle, \tag{52}$$

is a majorant function of $F_{1,j_\ell}(\cdot, \boldsymbol{x}_\ell^{\overline{j}_\ell})$. The following proposition allows us to propose a SDP matrix $\boldsymbol{B} \in \mathbb{R}^{N \times N}$ for building majorizing approximations of F_1 at $\boldsymbol{x}_\ell$ for every $\ell \in \mathbb{N}$. Hereafter, $\text{Re}\{\cdot\}$ (resp. $\text{Im}\{\cdot\}$) designates the real (resp. imaginary) part of its argument.

---

[1] We consider right derivatives at $\omega = 0$.

Proposition 4.1. *Let $\boldsymbol{u} \in \mathbb{R}^N$. A quadratic majorant of F_1 at $\boldsymbol{u}$ is*

$$(\forall \boldsymbol{x} \in \mathbb{R}^N) \quad Q_1(\boldsymbol{x}, \boldsymbol{u}) := F_1(\boldsymbol{u}) + \langle \boldsymbol{x} - \boldsymbol{u}, \nabla F_1(\boldsymbol{u}) \rangle + \frac{1}{2} \langle \boldsymbol{x} - \boldsymbol{u}, \boldsymbol{B}(\boldsymbol{x} - \boldsymbol{u}) \rangle, \tag{53}$$

where $\boldsymbol{B} := \operatorname{Diag}\left(\boldsymbol{\Omega}^\top \mathbf{1}_S\right) + \varepsilon \mathbf{I}_N$, where $\mathbf{1}_S$ is the unit vector on $\mathbb{R}^S$, $\varepsilon \geq 0$, and $\boldsymbol{\Omega} = \left(\Omega^{(s,n)}\right)_{1 \leq s \leq S, 1 \leq n \leq N} \in \mathbb{R}^{S \times N}$ is given by

$$(\forall s \in \{1, \ldots, S\})(\forall n \in \{1, \ldots, N\})$$

$$\Omega^{(s,n)} := |\operatorname{Re}\{T^{(s,n)}\}| \sum_{n'=1}^{N} |\operatorname{Re}\{T^{(s,n')}\}| + |\operatorname{Im}\{T^{(s,n)}\}| \sum_{n'=1}^{N} |\operatorname{Im}\{T^{(s,n')}\}|. \tag{54}$$

Proof. Let $\boldsymbol{u} \in \mathbb{R}^N$. For every $s \in \{1, \ldots, S\}$, we have, for every $\boldsymbol{x} \in \mathbb{R}^N$,

$$\varphi_1^{(s)}\left(|\boldsymbol{T}^{(s)}\boldsymbol{x}|\right) = \varphi_1^{(s)}\left(|\boldsymbol{T}^{(s)}\boldsymbol{u}|\right) + \left\langle \boldsymbol{x} - \boldsymbol{u}, \operatorname{Re}\{(\boldsymbol{T}^{(s)})^*\boldsymbol{T}^{(s)}\}\boldsymbol{u} \right\rangle + \frac{1}{2}|\boldsymbol{T}^{(s)}(\boldsymbol{x} - \boldsymbol{u})|^2,$$

where $\boldsymbol{T}^{(s)}$ denotes line s of matrix $\boldsymbol{T}$ and $(\cdot)^*$ is the matrix trans-conjugate operation. Then, summing over $s \in \{1, \ldots, S\}$, we obtain

$$(\forall \boldsymbol{x} \in \mathbb{R}^N) \quad F_1(\boldsymbol{x}) = F_1(\boldsymbol{u}) + \langle \boldsymbol{x} - \boldsymbol{u}, \nabla F_1(\boldsymbol{u}) \rangle + \frac{1}{2}|||\boldsymbol{T}(\boldsymbol{x} - \boldsymbol{u})|||^2, \tag{55}$$

where $||| \cdot |||$ is the Hermitian norm of $\mathbb{C}^S$.

Let $(V_\mathcal{R}^{(s,n)})_{1 \leq s \leq S, 1 \leq n \leq N} \in [0, +\infty)^{S \times N}$ and $(V_\mathcal{I}^{(s,n)})_{1 \leq s \leq S, 1 \leq n \leq N} \in [0, +\infty)^{S \times N}$ be such that, for every $s \in \{1, \ldots, S\}$, $\sum_{n \in \mathcal{S}_\mathcal{R}^{(s)}} V_\mathcal{R}^{(s,n)} \leq 1$, $\sum_{n \in \mathcal{S}_\mathcal{I}^{(s)}} V_\mathcal{I}^{(s,n)} \leq 1$ where

$$\mathcal{S}_\mathcal{R}^{(s)} := \{n \in \{1, \ldots, N\} : V_\mathcal{R}^{(s,n)} \neq 0\} = \{n \in \{1, \ldots, N\} : \operatorname{Re}\{T^{(s,n)}\} \neq 0\},$$
$$\mathcal{S}_\mathcal{I}^{(s)} := \{n \in \{1, \ldots, N\} : V_\mathcal{I}^{(s,n)} \neq 0\} = \{n \in \{1, \ldots, N\} : \operatorname{Im}\{T^{(s,n)}\} \neq 0\}.$$

Jensen's inequality yields, for every $s \in \{1, \ldots, S\}$,

$$\left| \sum_{n=1}^{N} T^{(s,n)}(x^{(n)} - u^{(n)}) \right|^2$$

$$= \left(\sum_{n=1}^{N} \operatorname{Re}\{T^{(s,n)}\}(x^{(n)} - u^{(n)}) \right)^2 + \left(\sum_{n=1}^{N} \operatorname{Im}\{T^{(s,n)}\}(x^{(n)} - u^{(n)}) \right)^2$$

$$= \left(\sum_{n \in \mathcal{S}_\mathcal{R}^{(s)}} V_\mathcal{R}^{(s,n)} \left(\frac{\operatorname{Re}\{T^{(s,n)}\}}{V_\mathcal{R}^{(s,n)}}(x^{(n)} - u^{(n)}) \right) \right)^2 + \left(\sum_{n \in \mathcal{S}_\mathcal{I}^{(s)}} V_\mathcal{I}^{(s,n)} \left(\frac{\operatorname{Im}\{T^{(s,n)}\}}{V_\mathcal{I}^{(s,n)}}(x^{(n)} - u^{(n)}) \right) \right)^2$$

$$\leq \sum_{n \in \mathcal{S}_\mathcal{R}^{(s)}} \frac{(\operatorname{Re}\{T^{(s,n)}\})^2}{V_\mathcal{R}^{(s,n)}}(x^{(n)} - u^{(n)})^2 + \sum_{n \in \mathcal{S}_\mathcal{I}^{(s)}} \frac{(\operatorname{Im}\{T^{(s,n)}\})^2}{V_\mathcal{I}^{(s,n)}}(x^{(n)} - u^{(n)})^2. \tag{56}$$

Let us now choose

$$(\forall (s,n) \in \{1, \ldots, S\} \times \{1, \ldots, N\})$$

$$V_\mathcal{R}^{(s,n)} = \begin{cases} 0, & \text{if } \operatorname{Re}\{T^{(s,n)}\} = 0, \\ \dfrac{|\operatorname{Re}\{T^{(s,n)}\}|}{\sum_{n'=1}^{N} |\operatorname{Re}\{T^{(s,n')}\}|}, & \text{otherwise,} \end{cases}$$

$$V_\mathcal{I}^{(s,n)} = \begin{cases} 0, & \text{if } \operatorname{Im}\{T^{(s,n)}\} = 0, \\ \dfrac{|\operatorname{Im}\{T^{(s,n)}\}|}{\sum_{n'=1}^{N} |\operatorname{Im}\{T^{(s,n')}\}|}, & \text{otherwise.} \end{cases}$$

It follows from (56) that, for every $s \in \{1, \ldots, S\}$,

$$\left| \sum_{n=1}^{N} T^{(s,n)}(x^{(n)} - u^{(n)}) \right|^2$$

$$\leq \sum_{n=1}^{N} \left(|\mathrm{Re}\{T^{(s,n)}\}| \sum_{n'=1}^{N} |\mathrm{Re}\{T^{(s,n')}\}| \right) (x^{(n)} - u^{(n)})^2$$

$$+ \sum_{n=1}^{N} \left(|\mathrm{Im}\{T^{(s,n)}\}| \sum_{n'=1}^{N} |\mathrm{Im}\{T^{(s,n')}\}| \right) (x^{(n)} - u^{(n)})^2.$$

It can be deduced that

$$|||\boldsymbol{T}(\boldsymbol{x} - \boldsymbol{u})|||^2 \leq \left\langle \boldsymbol{x} - \boldsymbol{u}, \mathrm{Diag}\left(\boldsymbol{\Omega}^\top \mathbf{1}_S\right)(\boldsymbol{x} - \boldsymbol{u}) \right\rangle, \tag{57}$$

where $\boldsymbol{\Omega}$ is defined by (54). Altogether, (55) and (57) lead to the desired majorization. $\square$

Combining the above lemma with Remark 2.4(ii) leads to the construction, for every $\ell \in \mathbb{N}$, of a quadratic majorant of $F_{1,j_\ell}(\cdot, \boldsymbol{x}_\ell^{\bar{\jmath}_\ell})$ at $\boldsymbol{x}_\ell$ of the form (52) with

$$(\forall \ell \in \mathbb{N}) \quad \boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell) := \mathrm{Diag}\left(\boldsymbol{\Omega}_{j_\ell}^\top \mathbf{1}_S\right) + \varepsilon \mathbf{I}_{N_{j_\ell}}, \tag{58}$$

where $\boldsymbol{\Omega}_{j_\ell} \in \mathbb{R}^{S \times N_{j_\ell}}$ is the matrix obtained by extracting the columns with indices in $\mathbb{J}_{j_\ell}$ from the matrix $\boldsymbol{\Omega}$ given by (54). Note that Assumption 2.3(ii) is satisfied for matrices (58) with

$$\begin{cases} \underline{\nu} = \varepsilon + \min\limits_{n \in \mathbb{J}_{j_\ell}} \sum_{s=1}^{S} \Omega^{(s,n)}, \\ \overline{\nu} = \varepsilon + \max\limits_{n \in \mathbb{J}_{j_\ell}} \sum_{s=1}^{S} \Omega^{(s,n)}. \end{cases} \tag{59}$$

If each column of $\boldsymbol{T}$ is nonzero, then one can choose $\varepsilon = 0$ in (59). Otherwise, we must choose $\varepsilon > 0$.

4.3 Implementation of the proximity operator of R

Let $\ell \in \mathbb{N}$, let $\boldsymbol{x}_\ell$ be the ℓ-th iterate in Algorithm (12) and let $j_\ell \in \{1, \ldots, J\}$ be the block selected at iteration ℓ. Since R_{j_ℓ} is an additive separable function, and $\boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell)$ reads $\mathrm{Diag}(a_{j_\ell}^{(1)}, \ldots, a_{j_\ell}^{(N_{j_\ell})})$, we have

$$(\forall \boldsymbol{y} = (y^{(n)})_{n \in \mathbb{J}_{j_\ell}} \in \mathbb{R}^{N_{j_\ell}}) \quad \mathrm{prox}_{\gamma_\ell^{-1} \boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell), R_{j_\ell}}(\boldsymbol{y}) = \left(\mathrm{prox}_{\gamma_\ell^{-1} a_{j_\ell}^{(n)}, \rho^{(n)}}(y^{(n)}) \right)_{n \in \mathbb{J}_{j_\ell}}. \tag{60}$$

For every $n \in \mathbb{J}_{j_\ell}$, let $\varsigma_{j_\ell}^{(n)} := \gamma_\ell \vartheta_n \left(a_{j_\ell}^{(n)}\right)^{-1} > 0$. According to (50), we have then

$$(\forall \upsilon \in \mathbb{R}) \quad \mathrm{prox}_{\gamma_\ell^{-1} a_{j_\ell}^{(n)}, \rho^{(n)}}(\upsilon) = \underset{\underline{\eta}_n \leq \omega \leq \overline{\eta}_n}{\mathrm{argmin}} \left\{ \varsigma_{j_\ell}^{(n)} |\omega - \overline{\omega}_n|^{\pi_n} + \frac{1}{2}(\omega - \upsilon)^2 \right\}$$

$$= \min \left\{ \overline{\eta}_n, \max \left\{ \underline{\eta}_n, \mathrm{prox}_{\varsigma_{j_\ell}^{(n)} |\cdot - \overline{\omega}_n|^{\pi_n}}(\upsilon) \right\} \right\}$$

$$= \min \left\{ \overline{\eta}_n, \max \left\{ \underline{\eta}_n, \overline{\omega}_n + \mathrm{prox}_{\varsigma_{j_\ell}^{(n)} |\cdot|^{\pi_n}}(\upsilon - \overline{\omega}_n) \right\} \right\}. \tag{61}$$

Hence, provided that the proximity operator $\mathrm{prox}_{\varsigma_{j_\ell}^{(n)} |\cdot|^{\pi_n}}$ has an explicit form, the exact version (7) of Algorithm (12) can be used.

4.4 Simulation results

We now demonstrate the practical performance of our algorithm on an image reconstruction problem. In our experiments, $\boldsymbol{W}$ is an overcomplete Haar synthesis operator performed on a single resolution level. Thus, $N = 4M$, and, for every $\boldsymbol{x} = (x^{(n)})_{1 \leq n \leq N} \in \mathbb{R}^N$, $(x^{(n)})_{1 \leq n \leq M}$ correspond to the approximation frame coefficients, whereas $(x^{(n)})_{pM+1 \leq n \leq (p+1)M}$ with $p \in \{1, 2, 3\}$ correspond to the horizontal, vertical and diagonal detail coefficients, respectively. We take, for every $n \in \{1, \ldots, M\}$, $(\pi_n, \vartheta_n) = (2, \vartheta^{\mathrm{a}})$ and, for every $n \in \{M+1, \ldots, N\}$, $(\pi_n, \vartheta_n) = (1, \vartheta^{\mathrm{d}})$, with $(\vartheta^{\mathrm{a}}, \vartheta^{\mathrm{d}}) \in (0, +\infty)^2$. Note that, for these choices of $(\pi_n)_{1 \leq n \leq N}$ and $(\vartheta_n)_{1 \leq n \leq N}$, the proximity operator (61) has an explicit form [17]. The original image $\overline{\boldsymbol{v}}$, with size $M = 256 \times 256$, is shown in Figure 1(a). Although the Haar coefficient vector $\overline{\boldsymbol{x}}$ is not uniquely defined, an example is displayed in Figure 1(b). The observation matrix is here $\boldsymbol{H} = \boldsymbol{H}_{\mathcal{R}} + \mathrm{i}\boldsymbol{H}_{\mathcal{I}}$ where $[\boldsymbol{H}_{\mathcal{R}}^{\top}, \boldsymbol{H}_{\mathcal{I}}^{\top}]^{\top} \in \mathbb{R}^{2S \times M}$ models $2S = 92160$ distinct projections from 256 parallel acquisition lines and 360 angles. The magnitude measurement vector $|\boldsymbol{H}\overline{\boldsymbol{v}}|$ is then corrupted with an additive real-valued white zero-mean Gaussian noise with variance equals to 0.1 which is truncated so as to guarantee the nonnegativity of the observed data. For every $n \in \{1, \ldots, N\}$, $(\underline{\eta}_n, \overline{\eta}_n, \overline{\omega}_n)$ are minimal, maximal and mean values, imposed on the sought frame coefficients. In order to set to zero the coefficients located in a subset $\mathbb{E} \subset \{1, \ldots, N\}$ corresponding to the object background, we choose, for every $n \in \mathbb{E}$, $\underline{\eta}_n = \overline{\eta}_n = 0$, as illustrated in Figure 1(c), and for coefficient indices $n \in \{1, \ldots, N\} \setminus \mathbb{E}$, we do not introduce specific range assumption by setting $\underline{\eta}_n = -\infty$ and $\overline{\eta}_n = +\infty$. Moreover, we take $\overline{\omega}_n = 0.8$, for every $n \in \{1, \ldots, M\} \setminus \mathbb{E}$, $\overline{\omega}_n = 0$ otherwise. Parameters ϑ^{a}, ϑ^{d} and δ are adjusted so as to maximize the signal-to-noise ratio (SNR) between the original image $\overline{\boldsymbol{v}}$ and the reconstructed one $\widehat{\boldsymbol{v}}$, expressed as

$$\mathrm{SNR} := 20 \log_{10} \left(\frac{\|\overline{\boldsymbol{v}}\|}{\|\widehat{\boldsymbol{v}} - \overline{\boldsymbol{v}}\|} \right).$$

| (a) | (b) | (c) |
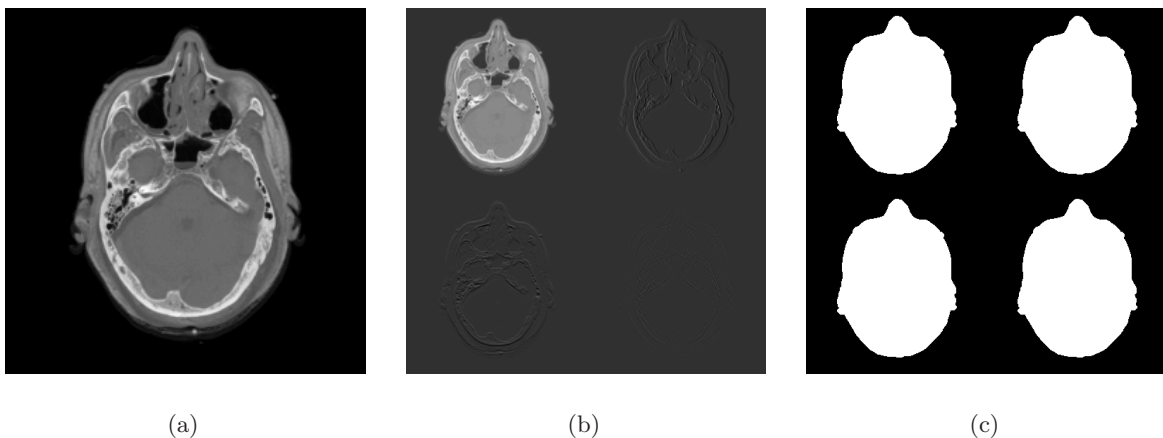
Figure 1: Original image $\overline{\boldsymbol{v}}$ (a), example of frame coefficient $\overline{\boldsymbol{x}}$ with approximation coefficients in top-left (b), and index set $\mathbb{E}$ in black (c).

We adopt the essentially cyclic rule described in Assumption 2.4 to update the $(K = J)$ blocks. Let $\ell \in \mathbb{N}$ be an iterate of the BC-VMFB algorithm, and $(m, j') \in \mathbb{N} \times \{1, \ldots, J\}$ be such that $\ell = mJ + j' - 1$. Then the block index j_ℓ is defined as $j_\ell = \sigma_m(j')$, where σ_m is a random permutation from $\{1, \ldots, J\}$ to $\{1, \ldots, J\}$, and

$$(\forall j' \in \{1, \ldots, J\}) \quad \mathbb{J}_{j'} = \bigcup_{p=0}^{3} \{Mp + (j'-1)Q + 1, \ldots, Mp + j'Q\}, \tag{62}$$

with $(J, Q) \in (\mathbb{N}^*)^2$ such that $M = JQ$. Thus, at each iteration $\ell \in \mathbb{N}$, the updated j_ℓ block is

of constant size $N_{j_\ell} = 4Q$. Figure 2 illustrates two examples of a resulting block index set $\mathbb{J}_{j'}$ for two different values of Q.
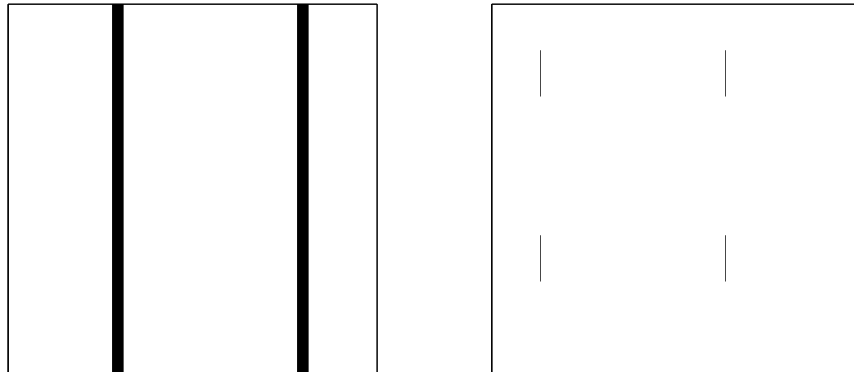

Figure 2: An example of index set $\mathbb{J}_{j'}$ (black), for $Q = 4096$ (left) and $Q = 64$ (right), the frame coefficients being structured as depicted in Figure 1(b).

Figure 3 (left) shows the reconstructed image with Algorithm (7), using the majorant curvature (58) where $\varepsilon = 0$, $Q = 64$ and $\gamma_\ell \equiv 1.9$. We also present in Figure 3 (right) the variations of the reconstruction time with respect to the block-size parameter Q, when performing tests on an Intel(R) Core(TM) i7-3520M @ 2.9GHz using a Matlab 7 implementation. The reconstruction time corresponds to the computation time necessary to fulfill the following stopping criterion:

$$\|\boldsymbol{x}_\ell - \widehat{\boldsymbol{x}}\| \leq 10^{-3}\|\widehat{\boldsymbol{x}}\|, \tag{63}$$

the optimal solution $\widehat{\boldsymbol{x}}$ being precomputed for each experiment, using a large number of iterations. One can observe that the best compromise in terms of convergence speed is obtained for an intermediate block-size, namely $Q = 64$. Figure 4 illustrates the variations of $\left(G(\boldsymbol{x}_\ell) - \widehat{G}\right)_\ell$ and $\left(\|\boldsymbol{x}_\ell - \widehat{\boldsymbol{x}}\|/\|\widehat{\boldsymbol{x}}\|\right)_\ell$ with respect to the computation time, using either the proposed BC-VMFB algorithm, BC-FB algorithm or PALM algorithm for the previous optimal block-size. Hereabove, $\widehat{G}$ denotes the minimum of the (possibly) different values $G(\widehat{\boldsymbol{x}})$ resulting from each simulation. Note that BC-FB and PALM algorithms can be viewed as special instances of Algorithm (7) where the cyclic rule (5) is adopted and the preconditioning matrices are equal to

$$(\forall \ell \in \mathbb{N}) \quad \boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell) = L\mathbf{I}_{N_{j_\ell}} \tag{64}$$

$$(\text{resp. } (\forall \ell \in \mathbb{N}) \quad \boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell) = L_{j_\ell}\mathbf{I}_{N_{j_\ell}}), \tag{65}$$

where L is a Lipschitz modulus of ∇F (resp., for every $j \in \{1, \dots, J\}$, L_j a Lipschitz modulus of $\nabla_j F(\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(j-1)}, \cdot, \boldsymbol{x}^{(j+1)}, \dots, \boldsymbol{x}^{(J)})$ [12]). These results illustrate the fact that the metric strategy given by (58) leads to a significant acceleration in terms of decay of both the objective function and the error on the iterates.

References

[1] Attouch, H., Bolte, J.: On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. Math. Program. **116**, 5–16 (2009)

[2] Attouch, H., Bolte, J., Redont, P., Soubeyran, A.: Proximal alternating minimization and projection methods for nonconvex problems. An approach based on the Kurdyka-Łojasiewicz inequality. Math. Oper. Res. **35**(2), 438–457 (2010)
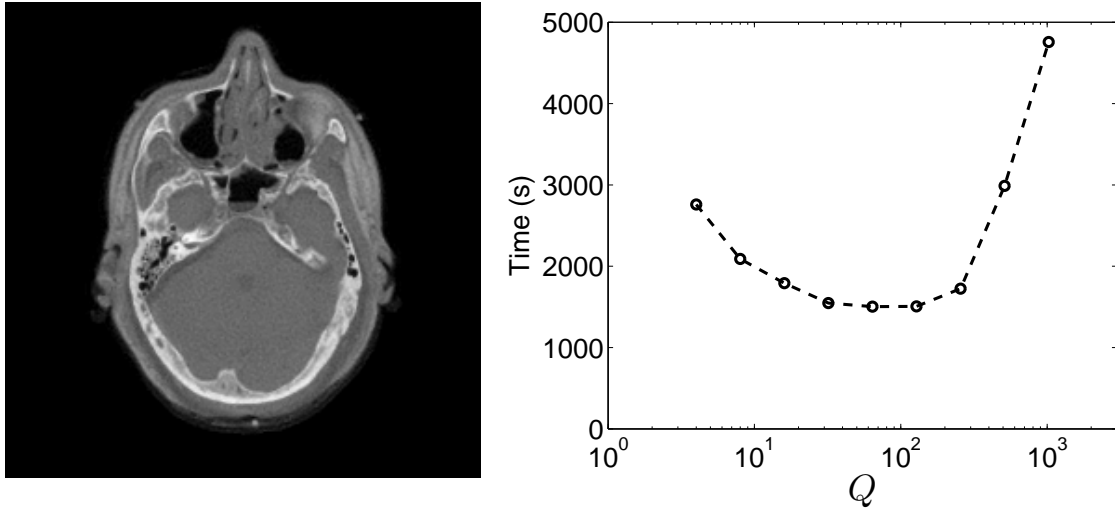
Figure 3: Reconstructed image $\widehat{\boldsymbol{v}} = \boldsymbol{W}\widehat{\boldsymbol{x}}$ with SNR = 27.64 dB (left) and reconstruction time for different block-sizes (right).

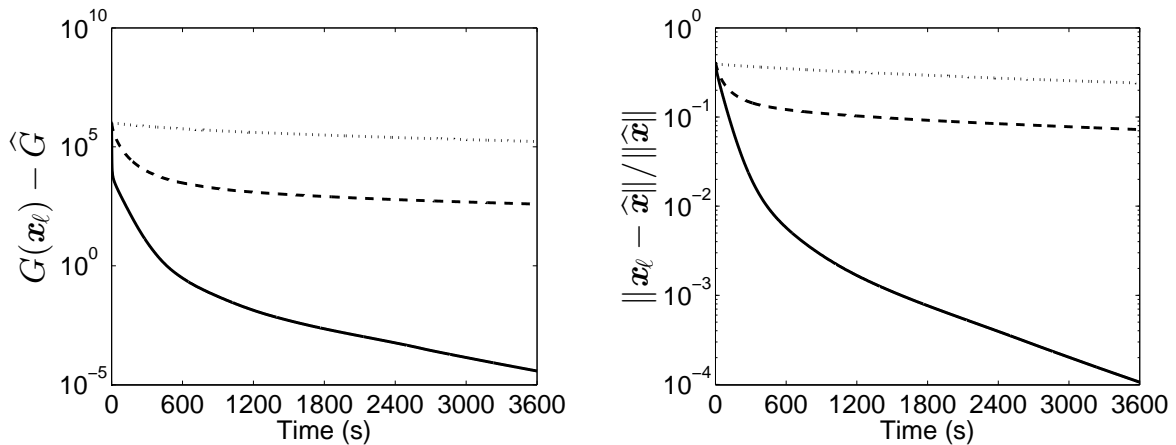

Figure 4: Convergence profile of BC-VMFB algorithm (solid line), PALM algorithm (dashed line) and BC-FB algorithm (dotted line).

[3] Attouch, H., Bolte, J., Svaiter, B.F.: Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods. Math. Program. **137**, 91–129 (2011)

[4] Auslender, A.: Asymptotic properties of the Fenchel dual functional and applications to decomposition problems. J. Optim. Theory Appl. **73**(3), 427–449 (1992)

[5] Bauschke, H.H., Combettes, P.L., Luke, D.R.: Phase retrieval, error reduction algorithm, and Fienup variants: a view from convex optimization. J. Opt. Soc. Amer. A **19**(7), 1334–1345 (2002)

[6] Bauschke, H.H., Combettes, P.L., Luke, D.R.: A new generation of iterative transform algorithms for phase contrast tomography. In: Proc. Int. Conf. Acoust., Speech Signal Process. (ICASSP 2005), vol. 4, pp. 89–92. Philadelphia, PA (2005)

[7] Bauschke, H.H., Combettes, P.L., Noll, D.: Joint minimization with alternating Bregman proximity operators. Pac. J. Optim. **2**(3), 401–424 (2006)

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

[9] 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)

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

[11] Bolte, J., Daniilidis, A., Ley, O., Mazet, L.: Characterizations of Łojasiewicz inequalities: subgradient flows, talweg, convexity. Trans. Amer. Math. Soc. **362**(6), 3319–3363 (2010)

[12] Bolte, J., Sabach, S., Teboulle, M.: Proximal alternating linearized minimization for nonconvex and nonsmooth problems. To appear in Math. Program. (2013). DOI 10.1007/s10107-013-0701-9

[13] Brègman, L.M.: The method of successive projection for finding a common point of convex sets. Soviet Math. Dokl. **6**, 688?692 (1965)

[14] Candès, E., Eldar, Y., Strohmer, T., Voroninski, V.: Phase retrieval via matrix completion. SIAM J. Imaging Sci. **6**(1), 199–225 (2013)

[15] Censor, Y., Lent, A.: Optimization of $\log x$ entropy over linear equality constraints. SIAM J. Control Optim. **25**(4), 921–933 (1987)

[16] Chouzenoux, E., Pesquet, J.C., Repetti, A.: Variable metric forward-backward algorithm for minimizing the sum of a differentiable function and a convex function. To appear in J. Optim. Theory Appl. (2013). DOI 10.1007/s10957-013-0465-7

[17] Combettes, P.L., Pesquet, J.C.: Proximal splitting methods in signal processing. In: H.H. Bauschke, R. Burachik, P.L. Combettes, V. Elser, D.R. Luke, H. Wolkowicz (eds.) Fixed-Point Algorithms for Inverse Problems in Science and Engineering, pp. 185–212. Springer-Verlag, New York (2010)

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

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

[20] Dainty, J.C., Fienup, J.R.: Phase retrieval and image reconstruction for astronomy. In: H. Stark (ed.) Image Recovery: Theory and Application, pp. 231–275. Academic Press, Orlando, FL (1987)

[21] Fessler, J.A.: Grouped coordinate ascent algorithms for penalized-likelihood transmission image reconstruction. IEEE Trans. Med. Imag. **16**, 166–175 (1997)

[22] Fienup, J.R.: Phase retrieval algorithms: A comparison. Appl. Opt. **21**, 2758–2769 (1982)

[23] Gerchberg, R.W., Saxton, W.O.: A practical algorithm for the determination of phase from image and diffraction plane pictures. Optik **35**, 237–246 (1972)

[24] Golub, G.H., Van Loan, C.F.: Matrix Computations, 3rd edn. Johns Hopkins University Press, Baltimore, Maryland (1996)

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

[26] Jacobson, M.W., Fessler, J.A.: An expanded theoretical treatment of iteration-dependent majorize-minimize algorithms. IEEE Trans. Image Process. **16**(10), 2411–2422 (2007)

[27] 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)

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

[29] Luenberger, D.G.: Linear and Nonlinear Programming. Addison-Wesley, Reading, Massachusetts (1973)

[30] Luo, Z.Q., Tseng, P.: On the convergence of the coordinate descent method for convex differentiable minimization. J. Optim. Theory Appl. **72**(1), 7–35 (1992)

[31] Luo, Z.Q., Tseng, P.: On the linear convergence of descent methods for convex essentially smooth minimization. SIAM J. Control Optim. **30**(2), 408–425 (1992)

[32] Mallat, S.: A Wavelet Tour of Signal Processing, 3rd edn. Academic Press, Burlington, MA (2009)

[33] 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)

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

[35] Ochs, P., Chen, Y., Brox, T., Pock, T.: iPiano: inertial proximal algorithm for non-convex optimization. Tech. rep. (2013). `http://lmb.informatik.uni-freiburg.de/ /Publications/2013/OB13`

[36] Ortega, J.M., Rheinboldt, W.C.: Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, New York, NY (1970)

[37] Powell, M.J.D.: On search directions for minimization algorithms. Math. Program. **4**, 193–201 (1973)

[38] Razaviyayn, M., Hong, M., Luo, Z.: A unified convergence analysis of block successive minimization methods for nonsmooth optimization. SIAM J. Optim. **23**(2), 1126–1153 (2013)

[39] Richtárik, P., Talác, M.: Iteration complexity of randomized block-coordinate descent methods for minimizing a composite function. Math. Program. **144**(1-2), 1–38 (2014)

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

[41] Saquib, S., Zheng, J., Bouman, C.A., Sauer, K.D.: Parallel computation of sequential pixel updates in statistical tomographic reconstruction. Proc. IEEE Int. Conf. Image Processing **2**, 93–96 (1995)

[42] Saxton, W.O.: Computer Techniques for Image Processing in Electron Microscopy. Academic Press, New York (1978)

[43] Sotthivirat, S., Fessler, J.A.: Image recovery using partitioned-separable paraboloidal surrogate coordinate ascent algorithms. IEEE Trans. Signal Process. **11**(3), 306–317 (2002)

[44] Tappenden, R., Richtárik, P., Gondzio, J.: Inexact coordinate descent: Complexity and preconditioning. Tech. rep. (2013). http://arxiv.org/abs/1304.5530

[45] Tseng, P.: Convergence of a block coordinate descent method for nondifferentiable minimization. J. Optim. Theory Appl. **109**(3), 475–494 (2001)

[46] Waldspurger, I., d'Aspremont, A., Mallat, S.: Phase recovery, maxcut and complex semidefinite programming. To appear in Math. Program. (2013). http://arxiv.org/abs/1206.0102

[47] Xu, Y., Yin, W.: A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. SIAM J. Imaging Sci. **6**(3), 1758–1789 (2013)

[48] Zangwill, W.I.: Nonlinear Programming. Prentice-Hall, Englewood Cliffs, New Jersey (1969)