J Glob Optim (2016) 66:457–485
DOI 10.1007/s10898-016-0405-9

A block coordinate variable metric forward–backward algorithm

Emilie Chouzenoux[1] · **Jean-Christophe Pesquet**[1] · **Audrey Repetti**[2]

Received: 8 October 2014 / Accepted: 18 January 2016 / Published online: 10 February 2016

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 two terms: (i) a differentiable, but not necessarily convex, function and (ii) a function that is not necessarily convex, nor necessarily differentiable. 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.

Keywords Nonconvex optimization · Nonsmooth optimization · Proximity operator · Majorize–Minimize algorithm · Block coordinate descent · Alternating minimization · Phase retrieval · Inverse problems

Mathematics Subject Classification 90C25 · 90C26 · 65K10 · 65F08 · 49M27 · 68U10 · 94A08 · 90C05

This work was supported by the CNRS MASTODONS project under grant 2013MesureHD and by the CNRS Imag'in Project under Grant 2015OPTIMISME.

✉ Emilie Chouzenoux
emilie.chouzenoux@univ-mlv.fr

[1] Laboratoire d'Informatique Gaspard Monge and CNRS UMR 8049, Université Paris-Est Marne-la-Vallée, Champs-sur-Marne, 77454 Marne-la-Vallée, France

[2] Institute of Sensors, Signals and Systems, Heriot-Watt University, Riccarton, EH14 4AS Edinburgh, Scotland, UK

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 Argmin $G \neq \varnothing$ denotes the set of minimizers of 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 $\left(\boldsymbol{x}^{(j)}\right)_{1\le j\le J} \in \mathbb{R}^{N_1} \times \ldots \times \mathbb{R}^{N_J}$, where, for every $j \in \{1,\ldots,J\}$, $\boldsymbol{x}^{(j)} = \left(x^{(n)}\right)_{n\in\mathbb{J}_j} \in \mathbb{R}^{N_j}$. 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 $j \in \{1,\ldots,J\}$, $R_j : \mathbb{R}^{N_j} \to (-\infty, +\infty]$.

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

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

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

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

The BCD method (3) is described in various reference books [9,35,43,62] 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 [29], it is sometimes also referred to as a *nonlinear Gauss-Seidel method* ([9, Chap.2], [43, 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 [58] 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 minimizer with respect to each block. As pointed out in [58], 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 [45]). The proximal version of the BCD algorithm, introduced in [5], allows this limitation to be overcome. It is defined as follows:

$$
\begin{array}{l}
\text{Let } \boldsymbol{x}_0 \in \mathbb{R}^N, \\
\text{For } \ell = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\text{Let } j_\ell \in \{1, \ldots, J\}, \\
\boldsymbol{x}_{\ell+1}^{(j_\ell)} \in \operatorname{prox}^{\boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell)/\gamma_\ell}_{F_{j_\ell}(\cdot, \boldsymbol{x}_\ell^{(\overline{\jmath}_\ell)}) + R_{j_\ell}} \left(\boldsymbol{x}_\ell^{(j_\ell)} \right), \\
\boldsymbol{x}_{\ell+1}^{(\overline{\jmath}_\ell)} = \boldsymbol{x}_\ell^{(\overline{\jmath}_\ell)},
\end{array}
\right.
\end{array}
\tag{6}
$$

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 matrix. Hereabove, $\operatorname{prox}^{\boldsymbol{U}}_{\psi}$ denotes the so-called *proximity operator* of a proper lower semicontinuous function $\psi : \mathbb{R}^M \to \mathbb{R}$ relative to the metric induced by a symmetric positive definite matrix $\boldsymbol{U} \in \mathbb{R}^{M \times M}$ (see Sect. 2.1). Note that Algorithm (6) has been extended in [8] 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, when $F \equiv 0$ and, for every $j \in \{1, \ldots, J\}$, R_j is the indicator function of a convex set, Algorithm (6) allows us to recover the celebrated POCS (*Projection Onto Convex Sets*) algorithm [14].

The convergence of the sequence $(\boldsymbol{x}_\ell)_{\ell \in \mathbb{N}}$ generated by Algorithm (6) to a solution to (1) has been established in [5] for a convex Lipschitz differentiable function F and proper lower semicontinous 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 [3] when $(\boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell))_{\ell \in \mathbb{N}}$ are identity matrices, and then generalized in [4] for general symmetric positive definite matrices $(\boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell))_{\ell \in \mathbb{N}}$, again assuming a cyclic rule. The convergence studies in [3,4] mainly rely on the assumption that the objective function G satisfies the Kurdyka-Łojasiewicz (KL) inequality [34]. 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 [10,11,33,34]. Since the proximal step in (6) is not explicit in general, an inexact version of the proximal BCD method is also considered in [4], 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:

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

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 [16] for the minimization of the Burg entropy function under linear constraints, and then extended to the more general case of a smooth function F [36,37]. Recently, the convergence of this algorithm has been studied in the case of an arbitrary nonsmooth function R under the assumptions that G satisfies the KL inequality and F is Lipschitz differentiable [13,27,60].

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 [60] in the case when F and R are respectively convex and convex w.r.t. each block variable, and generalized in [13] 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 [13]. A variant of PALM algorithm with similar convergence guarantees has been recently proposed in [30], alternating between Forward–Backward and proximal steps. Another related work is [61], where the convergence properties of PALM in the case of an essentially cyclic rule are studied.

An exact (resp. inexact) version of Algorithm (7) with general symmetric positive definite matrices $(\boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell))_{\ell\in\mathbb{N}}$ is studied in [51] (resp. [57]), 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 $1 - \delta$ (see also [20] for almost sure convergence results). Finally, let us emphasize that, as already noticed in [47], 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 [25,53,56] in the context of image reconstruction. Therefore, some convergence properties of Algorithm (7) can be deduced from those derived in [32] in the case when R_j are indicator functions of closed convex subsets of $\mathbb{R}^{N_j}$, and in [47] 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 [32,47] 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 the KL inequality similarly to [4]). Note that this convergence study generalizes our previous work [18] (see also [42] for a related approach, and [22] for the case when the functions are convex) which was restricted to an inexact Variable Metric Forward–Backward algorithm without block alternation (i.e. $J = 1$ and $N_1 = N$).

In a recent work [27], other authors have independently and concurrently established the convergence of the iterates generated by a version of Algorithm (7) for a class of nonconvex problems that encompasses the one we consider here. The main difference with respect to our work is that their approach is restricted to the use of a cyclic updating rule for the sequence $(j_\ell)_{\ell\in\mathbb{N}}$. By contrast, our analysis allows more flexibility in the choice of the blocks, since the essentially cyclic rule assumption we adopt makes it possible to update some of the target variables more frequently than others. Such a strategy appears to be of major interest in terms of numerical performance in some applications (see, for instance, [48]). Due to this fact, our convergence study significantly differs from the one conducted in [27]. The application to phase reconstruction provided in Sect. 4, which deals with an important problem in signal processing, is also completely novel. Table 1 hereafter summarizes the differences/similarities between our work and existing works, by precising whether convergence results are available for the sequence of iterates, or only for the sequence of objective function values.

The rest of the paper is organized as follows: Sect. 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

Table 1 List of existing convergence results for the BC-VMFB algorithm. Last line summarizes the paper's contribution

	Variable metric	Block update rule	Convergence
[13]	Scalar	Cyclic	Iterates
[61]	Scalar	Essentially cyclic	Iterates
[27]	Matrix	Cyclic	Iterates
[57]	Matrix	random	Objective function
Here	Matrix	Essentially cyclic	Iterates

is studied. Finally, Sect. 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:

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

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 symmetric positive definite matrix. Moreover, for every $\boldsymbol{U}_1 \in \mathbb{R}^{N\times N}$ and $\boldsymbol{U}_2 \in \mathbb{R}^{N\times N}$, we define the *Loewner partial order* on $\mathbb{R}^{N\times N}$ as

$$\boldsymbol{U}_1 \preceq \boldsymbol{U}_2 \quad \Leftrightarrow \quad (\forall \boldsymbol{x} \in \mathbb{R}^N) \quad \langle \boldsymbol{x}, \boldsymbol{U}_1\boldsymbol{x}\rangle \le \langle \boldsymbol{x}, \boldsymbol{U}_2\boldsymbol{x}\rangle .$$

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

Definition 2.2 [52, Def. 8.3],[39, Sec.1.3] Let $\psi : \mathbb{R}^N \to (-\infty,+\infty]$ 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:

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

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 [4], 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 ([31, Sec. XV.4], [21] and [4]) 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 symmetric positive definite 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

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

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 $\operatorname{prox}_{\psi}^{\boldsymbol{U}}$ is a nonempty set.

(ii) If $\boldsymbol{U}$ is equal to $\mathbf{I}_N$, the identity matrix of $\mathbb{R}^{N\times N}$, then $\operatorname{prox}_{\psi} \equiv \operatorname{prox}_{\psi}^{\mathbf{I}_N}$ is the proximity operator employed in [4]. 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 $\operatorname{prox}_{\psi} \equiv \operatorname{prox}_{\psi}^{\mathbf{I}_N}$ is the proximity operator originally defined in [40].

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 dom R where $L > 0$, i.e.,

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

(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 [4,23]. In particular, if dom R is compact and F is twice continuously differentiable, Assumption 2.1(ii) holds.
(ii) According to Assumption 2.1(ii), dom $R \subset$ dom $F = \mathbb{R}^N$. Thus, as a consequence of Assumption 2.1(i), dom G = dom R is nonempty.
(iii) Under Assumption 2.1, G is proper and lower semicontinuous, and its restriction to its domain is continuous. In particular, due to the coercivity of G, for every $\boldsymbol{x} \in$ dom R, $\operatorname{lev}_{\le G(\boldsymbol{x})} G$ 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 $0^0 = 0$).

Remark 2.4 Note that a more general local version of Assumption 2.2 can be found in the literature [11,12]. Nonetheless, as emphasized in [2], Assumption 2.2 is satisfied for a very wide class of functions, such as, in particular, real analytic and semi-algebraic functions.

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

Assumption 2.3

(i) The quadratic function defined as

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

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

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

(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 2.5

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

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

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

where $\boldsymbol{B}(\boldsymbol{x}') \in \mathbb{R}^{N \times N}$ is a symmetric positive definite matrix. Then, Assumption 2.3(i) is satisfied for $\boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell) = (B(\boldsymbol{x}_\ell)^{(n,n')})_{(n,n') \in \mathbb{J}_{j_\ell}^2}$, 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 $\boldsymbol{x}' \in \operatorname{dom} R$, $\underline{\nu} \mathbf{I}_N \preceq \boldsymbol{B}(\boldsymbol{x}') \preceq \overline{\nu} \mathbf{I}_N$, then Assumption 2.3(ii) is also satisfied.

(iii) If $\operatorname{dom} R$ is convex, the existence of the majorant function (11) is ensured when F satisfies Assumption 2.1(ii) (see [18, 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 [58]:

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 There exists $(\underline{\gamma}, \overline{\gamma}) \in (0, +\infty)^2$ such that, for every $\ell \in \mathbb{N}$, one of the following statements holds:

(i) $\underline{\gamma} \leq \gamma_\ell \leq 1 - \overline{\gamma}$,
(ii) R_{j_ℓ} is a convex function and $\underline{\gamma} \leq \gamma_\ell \leq 2(1 - \overline{\gamma})$.

Remark 2.6 Assumption 2.5 can be interpreted as the fact that, for every $j \in \{1, \ldots, J\}$, larger stepsizes can be used when R_j is convex. More precisely, if R_j is nonconvex, the stepsize is restricted to $(0, 1)$, whereas it can belong to $(0, 2)$ if R_j is convex.

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 $\ell = 0, 1, \ldots$

$$
\left\lfloor
\begin{array}{l}
\text{Let } j_\ell \in \{1, \ldots, J\}, \\
\text{Find } \boldsymbol{x}_{\ell+1}^{(j_\ell)} \in \mathbb{R}^{N_{j_\ell}} \text{ and } \boldsymbol{r}_{\ell+1}^{(j_\ell)} \in \partial R_{j_\ell}(\boldsymbol{x}_{\ell+1}^{(j_\ell)}) \text{ such that} \\
\quad R_{j_\ell}(\boldsymbol{x}_{\ell+1}^{(j_\ell)}) + \left\langle \boldsymbol{x}_{\ell+1}^{(j_\ell)} - \boldsymbol{x}_\ell^{(j_\ell)}, \nabla_{j_\ell} F(\boldsymbol{x}_\ell) \right\rangle + \alpha \left\| \boldsymbol{x}_{\ell+1}^{(j_\ell)} - \boldsymbol{x}_\ell^{(j_\ell)} \right\|^2_{\boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell)} \\
\qquad\qquad\qquad\qquad\qquad \leq R_{j_\ell}(\boldsymbol{x}_\ell^{(j_\ell)}), \qquad\qquad (12a) \\
\quad \left\| \nabla_{j_\ell} F(\boldsymbol{x}_\ell) + \boldsymbol{r}_{\ell+1}^{(j_\ell)} \right\| \leq \beta \left\| \boldsymbol{x}_{\ell+1}^{(j_\ell)} - \boldsymbol{x}_\ell^{(j_\ell)} \right\|_{\boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell)}, \qquad\qquad (12b) \\
\boldsymbol{x}_{\ell+1}^{(\overline{\jmath}_\ell)} = \boldsymbol{x}_\ell^{(\overline{\jmath}_\ell)}. \qquad\qquad (12c)
\end{array}
\right.
$$

Remark 2.7 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). Let $\ell \in \mathbb{N}$.

(i) Suppose that Assumption 2.5(i) holds. Due to the definition of the proximity operator, we have,

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

so that the sufficient-decrease condition (12a) holds with $\alpha = (1 - \overline{\gamma})^{-1}/2$ (as $\gamma_\ell^{-1} \geq (1 - \overline{\gamma})^{-1} > 1$).

(ii) Suppose now that Assumption 2.5(ii) holds. 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

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

which yields

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

so that the sufficient-decrease condition (12a) holds with the same value of α as in case (i) (since $\gamma_\ell^{-1} \geq (2 - 2\overline{\gamma})^{-1} > 1/2$).

Secondly, 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

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

Using Assumptions 2.3(ii) and 2.5, we obtain

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

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

3 Convergence analysis

3.1 Descent properties

In this section, we provide some technical results concerning the behavior of the sequence $\big(G(\boldsymbol{x}_\ell)\big)_{\ell \in \mathbb{N}}$ 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 $\ell \in \mathbb{N}$,*

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

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),

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

On the other hand, using (12c),

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

Then, using (12a), we obtain

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

Therefore, combining (14) and (15) yields

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

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). □

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

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

□

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 $\mathcal{J}_{\ell_0}$ be a subset of $\{1,\ldots,J\}$ containing j_{ℓ_0}. Then, under Assumptions 2.1 and 2.3, we have*

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

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

$$\sum_{j\in\mathcal{J}_{\ell_0}} \left\| \nabla_j F(\boldsymbol{x}_{\ell_0+1}) + \boldsymbol{r}_{\ell_0+1}^{(j)} \right\|^2 \le 2\sum_{j\in\mathcal{J}_{\ell_0}} \|\nabla_j F(\boldsymbol{x}_{\ell_0+1}) - \nabla_j F(\boldsymbol{x}_{\ell_0})\|^2 + 2\sum_{j\in\mathcal{J}_{\ell_0}} \left\| \nabla_j F(\boldsymbol{x}_{\ell_0}) + \boldsymbol{r}_{\ell_0+1}^{(j)} \right\|^2. \tag{19}$$

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

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

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

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

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

$$\sum_{j\in\mathcal{J}_{\ell_0}} \left\| \nabla_j F(\boldsymbol{x}_{\ell_0}) + \boldsymbol{r}_{\ell_0+1}^{(j)} \right\|^2 \le \beta^2\overline{\nu}\|\boldsymbol{x}_{\ell_0+1} - \boldsymbol{x}_{\ell_0}\|^2 + \sum_{j\in\mathcal{J}_{\ell_0}\setminus\{j_{\ell_0}\}} \left\| \nabla_j F(\boldsymbol{x}_{\ell_0}) + \boldsymbol{r}_{\ell_0}^{(j)} \right\|^2. \tag{21}$$

Finally, (18) results from (19), (20) and (21). □

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*

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

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

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

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

$$\begin{aligned}
&\sum_{j\in\mathcal{J}_{\ell_0,\ell_0'}} \left\|\nabla_j F(\boldsymbol{x}_{\ell_0'+1}) + \boldsymbol{r}^{(j)}_{\ell_0'+1}\right\|^2 \\
&\le \left(L^2+\beta^2\overline{\nu}\right)2\|\boldsymbol{x}_{\ell_0'+1}-\boldsymbol{x}_{\ell_0'}\|^2 + 2\sum_{j\in\mathcal{J}_{\ell_0,\ell_0'}\setminus\{j_{\ell_0'}\}}\left\|\nabla_j F(\boldsymbol{x}_{\ell_0'}) + \boldsymbol{r}^{(j)}_{\ell_0'}\right\|^2 \\
&\le \left(L^2+\beta^2\overline{\nu}\right)2\|\boldsymbol{x}_{\ell_0'+1}-\boldsymbol{x}_{\ell_0'}\|^2 + 2\sum_{j\in\mathcal{J}_{\ell_0,\ell_0'}}\left\|\nabla_j F(\boldsymbol{x}_{\ell_0'}) + \boldsymbol{r}^{(j)}_{\ell_0'}\right\|^2 \\
&\le \left(L^2+\beta^2\overline{\nu}\right)\left(2\|\boldsymbol{x}_{\ell_0'+1}-\boldsymbol{x}_{\ell_0'}\|^2 + 2^2\|\boldsymbol{x}_{\ell_0'}-\boldsymbol{x}_{\ell_0'-1}\|^2\right) \\
&\quad + 2^2\sum_{j\in\mathcal{J}_{\ell_0,\ell_0'}\setminus\{j_{\ell_0'-1}\}}\left\|\nabla_j F(\boldsymbol{x}_{\ell_0'-1}) + \boldsymbol{r}^{(j)}_{\ell_0'-1}\right\|^2 \\
&\le \left(L^2+\beta^2\overline{\nu}\right)\left(2\|\boldsymbol{x}_{\ell_0'+1}-\boldsymbol{x}_{\ell_0'}\|^2 + 2^2\|\boldsymbol{x}_{\ell_0'}-\boldsymbol{x}_{\ell_0'-1}\|^2 + 2^3\|\boldsymbol{x}_{\ell_0'-1}-\boldsymbol{x}_{\ell_0'-2}\|^2\right) \\
&\quad + 2^3\sum_{j\in\mathcal{J}_{\ell_0,\ell_0'}\setminus\{j_{\ell_0'-2}\}}\left\|\nabla_j F(\boldsymbol{x}_{\ell_0'-2}) + \boldsymbol{r}^{(j)}_{\ell_0'-2}\right\|^2 \\
&\;\;\vdots \\
&\le \left(L^2+\beta^2\overline{\nu}\right)\sum_{\ell=\ell_0}^{\ell_0'} 2^{\ell_0'+1-\ell}\|\boldsymbol{x}_{\ell+1}-\boldsymbol{x}_\ell\|^2 \\
&\quad + 2^{\ell_0'+1-\ell_0}\sum_{j\in\mathcal{J}_{\ell_0,\ell_0'}\setminus\{j_{\ell_0}\}}\left\|\nabla_j F(\boldsymbol{x}_{\ell_0}) + \boldsymbol{r}^{(j)}_{\ell_0}\right\|^2.
\end{aligned}$$

□

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

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

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

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

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 $\big(\boldsymbol{x}_\ell\big)_{\ell\in\mathbb{N}}$ belongs to the compact subset $E = \operatorname{lev}_{\le G(\boldsymbol{x}_0)} G \subset \operatorname{dom} R$ 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

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

where $K > 0$ is defined in Assumption 2.4. 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 $u = G(\boldsymbol{x}_\ell) - \xi$ and $v = G(\boldsymbol{x}_{\ell+K}) - \xi$ leads to

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

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

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

Let us define

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

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

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

For every $k \in \{\ell + k_{\ell,\sigma_\ell(J)}, \ldots, \ell + K - 1\}$, let $\boldsymbol{r}^{(j_k)}_{k+1} \in \partial R_{j_k}(\boldsymbol{x}^{(j_k)}_{k+1})$ 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

$$\begin{aligned}\|\boldsymbol{t}_{\ell+K}\|^2 \leq{}& (L^2 + \beta^2\overline{\nu}) \sum_{k=\ell+k_{\ell,\sigma_\ell(J)}}^{\ell+K-1} 2^{\ell+K-k}\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|^2 \\ &+ 2^{K-k_{\ell,\sigma_\ell(J)}} \sum_{\substack{j=1\\ j\neq\sigma_\ell(J)}}^{J} \Big\|\nabla_j F(\boldsymbol{x}_{\ell+k_{\ell,\sigma_\ell(J)}}) + \boldsymbol{r}^{(j)}_{\ell+k_{\ell,\sigma_\ell(J)}}\Big\|^2.\end{aligned}$$

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

$$\begin{aligned}\|\boldsymbol{t}_{\ell+K}\|^2 \leq{}& (L^2 + \beta^2\overline{\nu}) \sum_{k=\ell+k_{\ell,\sigma_\ell(J)}}^{\ell+K-1} 2^{\ell+K-k}\big\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\big\|^2 \\ &+ (L^2 + \beta^2\overline{\nu}) \sum_{k=\ell+k_{\ell,\sigma_\ell(J-1)}}^{\ell+k_{\ell,\sigma_\ell(J)}-1} 2^{\ell+K-k}\big\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\big\|^2 \\ &+ 2^{K-k_{\ell,\sigma_\ell(J-1)}} \sum_{\substack{j=1\\ j\neq\sigma_\ell(i), i\in\{J-1,J\}}}^{J} \big\|\nabla_j F(\boldsymbol{x}_{\ell+k_{\ell,\sigma_\ell(J-1)}}) + \boldsymbol{r}^{(j)}_{\ell+k_{\ell,\sigma_\ell(J-1)}}\big\|^2.\end{aligned}$$

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

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

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

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

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

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

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

$$\begin{aligned}\sum_{\ell=\ell^*}^{+\infty}\Delta_\ell&=\sum_{\ell=\ell^*}^{+\infty}\big(G(\boldsymbol{x}_\ell)-\xi\big)^{1-\theta}-\big(G(\boldsymbol{x}_{\ell+K})-\xi\big)^{1-\theta}\\&=\sum_{\ell=\ell^*}^{\ell^*+K-1}\big(G(\boldsymbol{x}_\ell)-\xi\big)^{1-\theta},\end{aligned}$$

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 $\ell\in\mathbb{N}$,

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

In addition, since the sequence $\big(\|\boldsymbol{\chi}_\ell\|\big)_{\ell\in\mathbb{N}}$ 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 $\ell\in\mathbb{N}$, $\boldsymbol{x}_\ell\in\operatorname{dom}G$, 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), $(\widehat{\boldsymbol{x}}, \mathbf{0}) \in \operatorname{Graph} \partial G$ i.e., $\widehat{\boldsymbol{x}}$ is a critical point of G. □

Remark 3.1 In the case when the blocks are updated according to a cyclic rule and the proximity operator is computed exactly, one can obtain similar convergence results without assuming the continuity of functions $(R_j)_{1\leq j\leq J}$, by using similar arguments to those in the proof of [13, Lem. 5 (i)].

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 [18, Cor. 3.2]. □

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 [2,3], the local convergence rate of the BC-VMFB algorithm depends on this exponent.

The following lemma, which can be deduced from [2, Thm. 2], is instrumental to establish the convergence rate:

Lemma 3.5 *Let $(\Lambda_m)_{m\in\mathbb{N}}$ be a nonnegative sequence of reals decreasing to 0. Assume that there exist $m^* \in \mathbb{N}\setminus\{0\}$ 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 $\theta \in \left(\frac{1}{2}, 1\right)$, 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 $\theta \in \left(0, \frac{1}{2}\right]$, 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*

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

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

(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,

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

Moreover, using again the triangle inequality, we have

$$\begin{aligned}
\|\boldsymbol{x}_{mK} - \widehat{\boldsymbol{x}}\| &= \left\| \sum_{p=m}^{+\infty} \left(\boldsymbol{x}_{(p+1)K} - \boldsymbol{x}_{pK}\right) \right\| \\
&= \left\| \sum_{p=m}^{+\infty} \sum_{k'=0}^{K-1} \left(\boldsymbol{x}_{pK+k'+1} - \boldsymbol{x}_{pK+k'}\right) \right\| \\
&\le \sum_{p=m}^{+\infty} \left\| \sum_{k'=0}^{K-1} \left(\boldsymbol{x}_{pK+k'+1} - \boldsymbol{x}_{pK+k'}\right) \right\|,
\end{aligned} \tag{36}$$

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

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

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

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

Involving again Jensen's inequality, we have

$$\begin{aligned}
\|\boldsymbol{x}_{mK} - \boldsymbol{x}_\ell\|^2 &= \left\| \sum_{k'=0}^{k-1} \left(\boldsymbol{x}_{mK+k'+1} - \boldsymbol{x}_{mK+k'}\right) \right\|^2 \\
&\le k \sum_{k'=0}^{k-1} \left\|\boldsymbol{x}_{mK+k'+1} - \boldsymbol{x}_{mK+k'}\right\|^2 \le (K-1)\|\boldsymbol{\chi}_{mK}\|^2.
\end{aligned} \tag{39}$$

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

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

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

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

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

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

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

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

so that

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

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 $m \ge \max\{\ell^*/K, 1\}$,

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

and (30) is satisfied.

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

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

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

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

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 $\ell \in \mathbb{N}$,

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

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

$$\begin{aligned}
G(\boldsymbol{x}_\ell) - G(\widehat{\boldsymbol{x}}) &\le \kappa^{-1/\theta}\Big(2^K(L^2+\beta^2\overline{\nu})\Big)^{\frac{1}{2\theta}}\|\boldsymbol{\chi}_{(m-1)K}\|^{1/\theta}\\
&\le \kappa^{-1/\theta}\Big(2^K(L^2+\beta^2\overline{\nu})\Big)^{\frac{1}{2\theta}}\Lambda_{m-1}^{1/\theta}.
\end{aligned}$$

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

$$\begin{aligned}
G(\boldsymbol{x}_\ell) - G(\widehat{\boldsymbol{x}}) &\le \kappa^{-1/\theta}\Big(2^K(L^2+\beta^2\overline{\nu})\Big)^{\frac{1}{2\theta}}\lambda(m-1)^{-\frac{1-\theta}{\theta(2\theta-1)}}\\
&\le \kappa^{-1/\theta}\Big(2^K(L^2+\beta^2\overline{\nu})\Big)^{\frac{1}{2\theta}}\lambda\Big(\frac{\ell}{K}-2\Big)^{-\frac{1-\theta}{\theta(2\theta-1)}}.
\end{aligned}$$

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

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

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

Remark 3.2

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

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

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

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 $\boldsymbol{z} = \left(z^{(s)}\right)_{1 \leq s \leq S} \in [0, +\infty)^S$ be a degraded signal related to an original unknown signal $\overline{\boldsymbol{v}} \in \mathbb{R}^M$ through the model

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

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

Such a problem is of paramount importance in numerous areas of applied physics and engineering [7,15,24,54,59]. Note that unlike many existing works [6,15,26,28], it is not assumed that $\boldsymbol{H}$ is a Fourier transform matrix.

Set $\widehat{\boldsymbol{v}} = \boldsymbol{W}\widehat{\boldsymbol{x}}$ where $\boldsymbol{W} \in \mathbb{R}^{M\times N}$, $N \geq M$, is a given frame synthesis operator (e.g. a possibly redundant wavelet synthesis operator) [38]. Then, following a synthesis approach, the frame coefficient vector $\widehat{\boldsymbol{x}}$ can be estimated by solving Problem (1) where F is the so-called data fidelity term of the form:

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

Hereabove, for every $s \in \{1,\dots,S\}$, $\varphi^{(s)}\colon [0,+\infty) \to \mathbb{R}$, and $[\boldsymbol{H}\boldsymbol{W}\boldsymbol{x}]^{(s)}$ is the s-th component of $\boldsymbol{H}\boldsymbol{W}\boldsymbol{x} \in \mathbb{C}^S$. 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 $s \in \{1,\dots,S\}$, $\varphi^{(s)} := \varphi_1^{(s)} + \varphi_2^{(s)}$, where

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

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

with $\delta > 0$ and $z^{(s)}$, the s-th component of $\boldsymbol{z}$. Thus, the data fidelity term (43) is split as $F = F_1 + F_2$ where

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

For every $s \in \{1,\dots,S\}$, the first and second order derivatives of $\varphi_1^{(s)}$ and $\varphi_2^{(s)}$ with respect to ω are, respectively, [1]

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

$$\dot{\varphi}_2^{(s)}(\omega) = -z^{(s)}\omega\left(\omega^2+\delta^2\right)^{-1/2}, \tag{48}$$

and

$$(\forall \omega \in [0,+\infty)) \quad \ddot{\varphi}_1^{(s)}(\omega) = 1, \tag{49}$$

$$\ddot{\varphi}_2^{(s)}(\omega) = -z^{(s)}\delta^2(\omega^2+\delta^2)^{-3/2}. \tag{50}$$

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 [26] is recovered (i.e. $F = \frac{1}{2}\||\boldsymbol{H}\boldsymbol{W}\cdot| - \boldsymbol{z}\|^2$), which shows that the proposed function can be viewed as a smoothed version of it.

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

In addition, following [17,46], the following penalization term is employed on the wavelet coefficients:

$$(\forall \boldsymbol{x} = (x^{(n)})_{1 \le n \le N} \in \mathbb{R}^N) \quad R(\boldsymbol{x}) := \sum_{n=1}^{N} \rho^{(n)}(x^{(n)}), \tag{51}$$

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

$$(\forall \omega \in \mathbb{R}) \quad \rho^{(n)}(\omega) := \begin{cases} \vartheta_n |\omega - \overline{\omega}_n|^{\pi_n} & \text{if } \underline{\eta}_n \le \omega \le \overline{\eta}_n, \\ +\infty & \text{otherwise,} \end{cases} \tag{52}$$

and, for every $n \in \{1, \dots, N\}$, $\vartheta_n \in (0, +\infty)$, $\pi_n \in \mathbb{N} \setminus \{0\}$, $\underline{\eta}_n \in [-\infty, +\infty)$, $\overline{\eta}_n \in [\underline{\eta}_n, +\infty]$, 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 $\boldsymbol{T} := \boldsymbol{H}\boldsymbol{W} = (T^{(s,n)})_{1 \le s \le S, 1 \le n \le N} \in \mathbb{C}^{S \times N}$.

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 $F_{j_\ell}(\cdot, \boldsymbol{x}_\ell^{(\overline{\jmath}_\ell)})$ 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 $F_{1,j_\ell}(\cdot, \boldsymbol{x}_\ell^{\overline{\jmath}_\ell})$ and $F_{2,j_\ell}(\cdot, \boldsymbol{x}_\ell^{\overline{\jmath}_\ell})$ 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 $\ell \in \mathbb{N}$, $F_{2,j_\ell}(\cdot, \boldsymbol{x}_\ell^{\overline{\jmath}_\ell})$ is majorized by

$$(\forall \boldsymbol{y} \in \mathbb{R}^{N_{j_\ell}}) \quad Q_{2,j_\ell}(\boldsymbol{y} \mid \boldsymbol{x}_\ell) := F_2(\boldsymbol{x}_\ell) + \left\langle \boldsymbol{y} - \boldsymbol{x}_\ell^{(j_\ell)}, \nabla_{j_\ell} F_2(\boldsymbol{x}_\ell) \right\rangle. \tag{53}$$

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

$$\begin{aligned}(\forall \boldsymbol{y} \in \mathbb{R}^{N_{j_\ell}}) \quad Q_{1,j_\ell}(\boldsymbol{y} \mid \boldsymbol{x}_\ell) := {} & F_1(\boldsymbol{x}_\ell) + \left\langle \boldsymbol{y} - \boldsymbol{x}_\ell^{(j_\ell)}, \nabla_{j_\ell} F_1(\boldsymbol{x}_\ell) \right\rangle \\ & + \frac{1}{2} \left\langle \boldsymbol{y} - \boldsymbol{x}_\ell^{(j_\ell)}, \boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell)(\boldsymbol{y} - \boldsymbol{x}_\ell^{(j_\ell)}) \right\rangle, \end{aligned} \tag{54}$$

is a majorant function of $F_{1,j_\ell}(\cdot, \boldsymbol{x}_\ell^{\overline{\jmath}_\ell})$. The following proposition allows us to propose a symmetric positive definite 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, Re$\{\cdot\}$ (resp. Im$\{\cdot\}$) designates the real (resp. imaginary) part of its argument.

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

$$(\forall \boldsymbol{x} \in \mathbb{R}^N) \quad Q_1(\boldsymbol{x} \mid \boldsymbol{u}) := F_1(\boldsymbol{u}) + \langle \boldsymbol{x} - \boldsymbol{u}, \nabla F_1(\boldsymbol{u}) \rangle + \frac{1}{2} \langle \boldsymbol{x} - \boldsymbol{u}, \boldsymbol{B}(\boldsymbol{x} - \boldsymbol{u}) \rangle, \tag{55}$$

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

$$(\forall s \in \{1, \ldots, S\})(\forall n \in \{1, \ldots, N\})$$
$$\Omega^{(s,n)} := |\mathrm{Re}\{T^{(s,n)}\}| \sum_{n'=1}^{N} |\mathrm{Re}\{T^{(s,n')}\}| + |\mathrm{Im}\{T^{(s,n)}\}| \sum_{n'=1}^{N} |\mathrm{Im}\{T^{(s,n')}\}|. \tag{56}$$

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

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

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

$$(\forall \boldsymbol{x} \in \mathbb{R}^N) \quad F_1(\boldsymbol{x}) = F_1(\boldsymbol{u}) + \langle \boldsymbol{x} - \boldsymbol{u}, \nabla F_1(\boldsymbol{u}) \rangle + \frac{1}{2} |||\boldsymbol{T}(\boldsymbol{x} - \boldsymbol{u})|||^2, \tag{57}$$

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

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

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

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

$$\begin{aligned}
\left| \sum_{n=1}^{N} T^{(s,n)} (x^{(n)} - u^{(n)}) \right|^2 &= \left(\sum_{n=1}^{N} \mathrm{Re}\{T^{(s,n)}\} (x^{(n)} - u^{(n)}) \right)^2 \\
&\quad + \left(\sum_{n=1}^{N} \mathrm{Im}\{T^{(s,n)}\} (x^{(n)} - u^{(n)}) \right)^2 \\
&= \left(\sum_{n \in \mathcal{S}_{\mathcal{R}}^{(s)}} V_{\mathcal{R}}^{(s,n)} \left(\frac{\mathrm{Re}\{T^{(s,n)}\}}{V_{\mathcal{R}}^{(s,n)}} (x^{(n)} - u^{(n)}) \right) \right)^2 \\
&\quad + \left(\sum_{n \in \mathcal{S}_{\mathcal{I}}^{(s)}} V_{\mathcal{I}}^{(s,n)} \left(\frac{\mathrm{Im}\{T^{(s,n)}\}}{V_{\mathcal{I}}^{(s,n)}} (x^{(n)} - u^{(n)}) \right) \right)^2 \\
&\le \sum_{n \in \mathcal{S}_{\mathcal{R}}^{(s)}} \frac{(\mathrm{Re}\{T^{(s,n)}\})^2}{V_{\mathcal{R}}^{(s,n)}} (x^{(n)} - u^{(n)})^2 \\
&\quad + \sum_{n \in \mathcal{S}_{\mathcal{I}}^{(s)}} \frac{(\mathrm{Im}\{T^{(s,n)}\})^2}{V_{\mathcal{I}}^{(s,n)}} (x^{(n)} - u^{(n)})^2.
\end{aligned} \tag{58}$$

Let us now choose

$$
\begin{aligned}
&(\forall (s,n) \in \{1,\ldots,S\} \times \{1,\ldots,N\}) \\
&V_{\mathcal{R}}^{(s,n)} = \begin{cases} 0, & \text{if } \mathrm{Re}\{T^{(s,n)}\} = 0, \\ \frac{|\mathrm{Re}\{T^{(s,n)}\}|}{\sum_{n'=1}^{N} |\mathrm{Re}\{T^{(s,n')}\}|}, & \text{otherwise,} \end{cases} \\
&V_{\mathcal{I}}^{(s,n)} = \begin{cases} 0, & \text{if } \mathrm{Im}\{T^{(s,n)}\} = 0, \\ \frac{|\mathrm{Im}\{T^{(s,n)}\}|}{\sum_{n'=1}^{N} |\mathrm{Im}\{T^{(s,n')}\}|}, & \text{otherwise.} \end{cases}
\end{aligned}
$$

It follows from (58) that, for every $s \in \{1,\ldots,S\}$,

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

It can be deduced that

$$
|||\boldsymbol{T}(\boldsymbol{x}-\boldsymbol{u})|||^2 \le \left\langle \boldsymbol{x}-\boldsymbol{u}, \mathrm{Diag}\left(\boldsymbol{\Omega}^\top \mathbf{1}_S\right)(\boldsymbol{x}-\boldsymbol{u})\right\rangle, \tag{59}
$$

where $\boldsymbol{\Omega}$ is defined by (56). Altogether, (57) and (59) lead to the desired majorization. □

Combining the above lemma with Remark 2.5(ii) leads to the construction, for every $\ell \in \mathbb{N}$, of a quadratic majorant of $F_{1,j_\ell}(\cdot, \boldsymbol{x}_\ell^{\overline{j}_\ell})$ at $\boldsymbol{x}_\ell$ of the form (54) with

$$
(\forall \ell \in \mathbb{N}) \quad \boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell) := \mathrm{Diag}\left(\boldsymbol{\Omega}_{j_\ell}^\top \mathbf{1}_S\right) + \varepsilon \mathbf{I}_{N_{j_\ell}}, \tag{60}
$$

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 (56). Note that Assumption 2.3(ii) is satisfied for matrices (60) with

$$
\begin{cases} \underline{\nu} = \varepsilon + \min\limits_{n \in \mathbb{J}_{j_\ell}} \sum_{s=1}^{S} \Omega^{(s,n)}, \\ \overline{\nu} = \varepsilon + \max\limits_{n \in \mathbb{J}_{j_\ell}} \sum_{s=1}^{S} \Omega^{(s,n)}. \end{cases} \tag{61}
$$

If each column of $\boldsymbol{T}$ is nonzero, then one can choose $\varepsilon = 0$ in (61). 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

$$
\left(\forall \boldsymbol{y} = (y^{(n)})_{n \in \mathbb{J}_{j_\ell}} \in \mathbb{R}^{N_{j_\ell}}\right) \quad \mathrm{prox}_{R_{j_\ell}}^{\boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell)/\gamma_\ell}(\boldsymbol{y}) = \left(\mathrm{prox}_{\gamma_\ell \rho^{(n)}/a_{j_\ell}^{(n)}}(y^{(n)})\right)_{n \in \mathbb{J}_{j_\ell}}. \tag{62}
$$

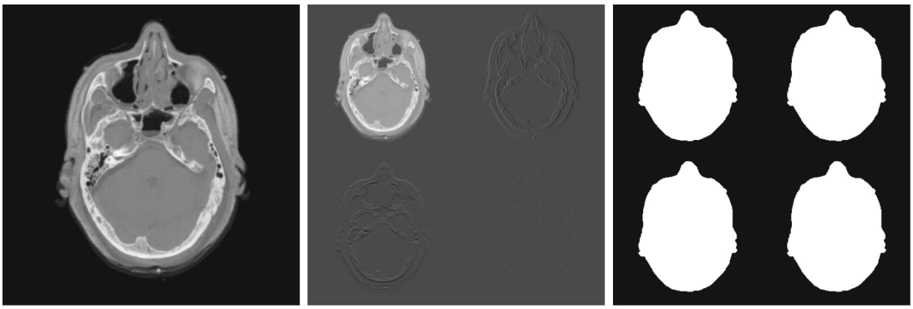
For every $n \in \mathbb{J}_{j_\ell}$, let $\varsigma_{j_\ell}^{(n)} := \gamma_\ell \vartheta_n \left(a_{j_\ell}^{(n)}\right)^{-1} > 0$. According to (52), we have then

$$
\begin{aligned}
(\forall \upsilon \in \mathbb{R}) \quad \operatorname{prox}_{\gamma_\ell \rho^{(n)}/a_{j_\ell}^{(n)}}(\upsilon) &= \underset{\underline{\eta}_n \le \omega \le \overline{\eta}_n}{\operatorname{argmin}} \left\{ \varsigma_{j_\ell}^{(n)} |\omega - \overline{\omega}_n|^{\pi_n} + \frac{1}{2}(\omega - \upsilon)^2 \right\} \\
&= \min \left\{ \overline{\eta}_n, \max \left\{ \underline{\eta}_n, \operatorname{prox}_{\varsigma_{j_\ell}^{(n)} |\cdot - \overline{\omega}_n|^{\pi_n}}(\upsilon) \right\} \right\} \\
&= \min \left\{ \overline{\eta}_n, \max \left\{ \underline{\eta}_n, \overline{\omega}_n + \operatorname{prox}_{\varsigma_{j_\ell}^{(n)} |\cdot|^{\pi_n}}(\upsilon - \overline{\omega}_n) \right\} \right\}. \qquad (63)
\end{aligned}
$$

Hence, provided that the proximity operator $\operatorname{prox}_{\varsigma_{j_\ell}^{(n)} |\cdot|^{\pi_n}}$ 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 $\boldsymbol{x} = (x^{(n)})_{1 \le n \le N} \in \mathbb{R}^N$, $(x^{(n)})_{1 \le n \le M}$ correspond to the approximation frame coefficients, whereas $(x^{(n)})_{pM+1 \le n \le (p+1)M}$ 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 $(\pi_n)_{1 \le n \le N}$ and $(\vartheta_n)_{1 \le n \le N}$, the proximity operator (63) has an explicit form [19]. The original image $\overline{\boldsymbol{v}}$, with size $M = 256 \times 256$, is shown in Fig. 1a. Although the Haar coefficient vector $\overline{\boldsymbol{x}}$ is not uniquely defined, an example is displayed in Fig. 1b. 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 Fig. 1c, and for coefficient indices $n \in \{1, \ldots, N\} \backslash \mathbb{E}$, we do not introduce specific range assumption by setting $\underline{\eta}_n = -\infty$ and $\overline{\eta}_n = +\infty$. Moreover, we



(a) **(b)** **(c)**

Fig. 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**)

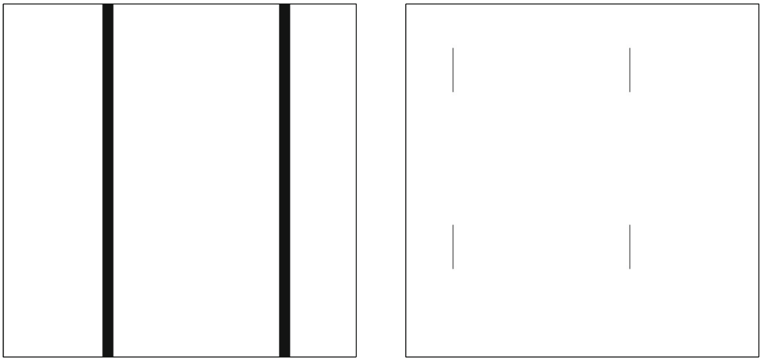


Fig. 2 An example of index set $\mathbb{J}_{j'}$ (*black*), for $P = 4096$ (*left*) and $P = 64$ (*right*), the frame coefficients being structured as depicted in Fig. 1b

take $\overline{\omega}_n = 0.8$, for every $n \in \{1, \ldots, M\}\backslash\mathbb{E}$, $\overline{\omega}_n = 0$ 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

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

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

$$(\forall j' \in \{1, \ldots, J\}) \quad \mathbb{J}_{j'} = \bigcup_{p=0}^{3} \{Mp + (j' - 1)P + 1, \ldots, Mp + j'P\}, \tag{64}$$

with $(J, P) \in (\mathbb{N}\backslash\{0\})^2$ such that $M = JP$. Thus, at each iteration $\ell \in \mathbb{N}$, the updated j_ℓ block is of constant size $N_{j_\ell} = 4P$. Figure 2 illustrates two examples of a resulting block index set $\mathbb{J}_{j'}$ for two different values of P.

Figure 3 (left) shows the reconstructed image with Algorithm (7), using the majorant curvature (60) where $\varepsilon = 0$, $P = 64$ and $\gamma_\ell \equiv 1.9$. We also present in Fig. 3 (right) the variations of the reconstruction time with respect to the block-size parameter P, 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 condition:

$$\|\boldsymbol{x}_\ell - \widehat{\boldsymbol{x}}\| \leq 10^{-3} \|\widehat{\boldsymbol{x}}\|, \tag{65}$$

where $\widehat{\boldsymbol{x}}$ is precomputed by running the algorithm, for each block size, until full stabilization of the iterates (up to the machine precision). The image $\widehat{\boldsymbol{x}}$ is a critical point of the criterion, since the convergence of the iterates of BC-VMFB to such a point is guaranteed, so that (65) aims at evaluating the computation time necessary to allow an iterate to be close enough to this limit point. Note that (65) is not led to be a practical stopping criterion for the method, since it requires two runs of the algorithm. A practical termination test could consist of controling the relative difference in norms between two consecutive iterates. One can observe on Fig. 3 (right) that the best compromise in terms of convergence speed is obtained for an

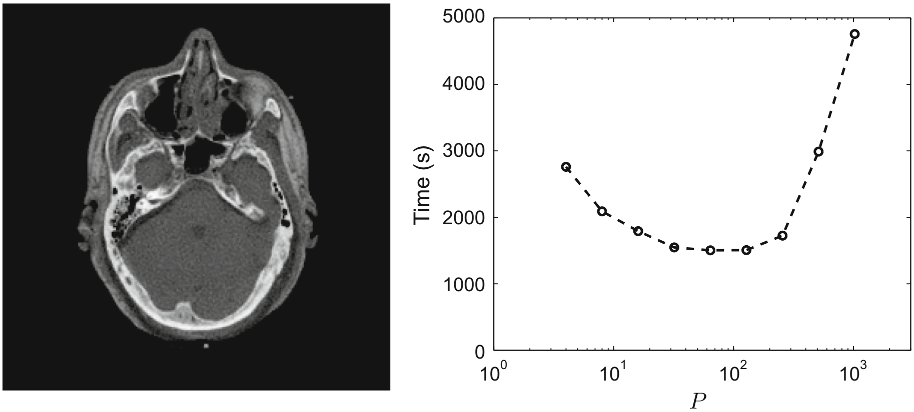


Fig. 3 Reconstructed image $\widehat{\boldsymbol{v}} = \boldsymbol{W}\widehat{\boldsymbol{x}}$ with SNR = 27.64 dB (*left*) and reconstruction time for different block-sizes (*right*)

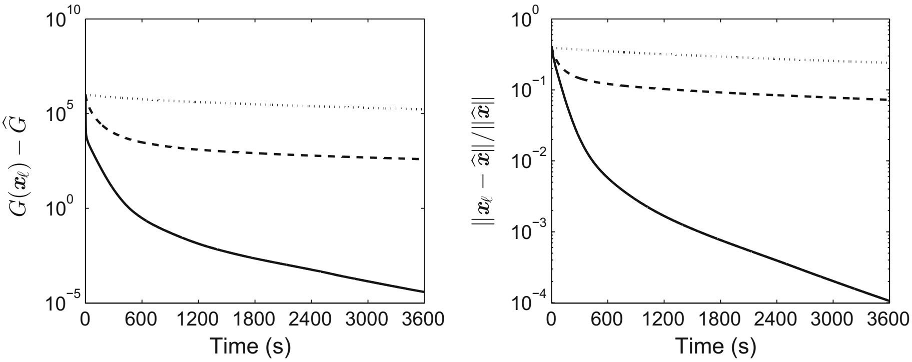


Fig. 4 Convergence profile of BC-VMFB algorithm (*solid line*), PALM algorithm (*dashed line*) and BC-FB algorithm (*dotted line*)

intermediate block-size, namely $P = 64$. Moreover, even if different values of P may result in different limit points $\widehat{\boldsymbol{x}}$ for the algorithm, we did not observe any significant variation in terms of reconstruction quality between these vectors. Figure 4 illustrates the variations of $\big(G(\boldsymbol{x}_\ell) - \widehat{G}\big)_\ell$ and $\big(\|\boldsymbol{x}_\ell - \widehat{\boldsymbol{x}}\|/\|\widehat{\boldsymbol{x}}\|\big)_\ell$ 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 (resp. PALM) algorithm can be viewed as a special instance of Algorithm (7) where the cyclic rule (5) is adopted and the preconditioning matrix is proportional to identity matrices, i.e.

$$(\forall \ell \in \mathbb{N}) \quad \boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell) = L\mathbf{I}_{N_{j_\ell}} \tag{66}$$

$$\text{(resp. } (\forall \ell \in \mathbb{N}) \quad \boldsymbol{A}_{j_\ell}(\boldsymbol{x}_\ell) = L_{j_\ell}\mathbf{I}_{N_{j_\ell}}), \tag{67}$$

where L is a Lipschitz modulus of ∇F (resp., for every $j \in \{1, \ldots, J\}$, L_j a Lipschitz modulus of $\nabla_j F(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(j-1)}, \cdot, \boldsymbol{x}^{(j+1)}, \ldots, \boldsymbol{x}^{(J)})$ [13]). All the algorithms lead asymptotically to solutions of similar quality in terms of SNR. Furthermore, one can observe on Fig. 4 that BC-VMFB algorithm requires less time than BC-FB and PALM algorithms to reach small values of $\left(G(\boldsymbol{x}_\ell) - \widehat{G}\right)_\ell$, and $\left(\|\boldsymbol{x}_\ell - \widehat{\boldsymbol{x}}\|/\|\widehat{\boldsymbol{x}}\|\right)_\ell$. This illustrates the fact that the metric strategy given by (60) leads to a significant acceleration in terms of decay of both the objective function and the error on the iterates. Note that the benefits of BC-VMFB over its non preconditioned versions have also been observed in the context of blind video deconvolution [1], spectral unmixing [49] and gene regulatory network inference [44].

Although the phase retrieval reconstruction problem has led to a large amount of works in the litterature [6,7,15,28,41,55,59], comparisons with the competing techniques were difficult to perform. Actually, the aforementioned methods tend to be sensitive to noise and/or to be less effective in the under-determined case and/or to be difficult to apply in a large scale non-Fourier context. On the one hand, when applied to our problem, the alternating projection algorithm from [28] and the regularized version [41] were extremely demanding in computational time and available memory. Moreover, they led to unsatisfactory results in terms of image quality. On the other hand, due to the large size of the data, and the complicated structure of $\boldsymbol{T}$, it appeared impossible to run the semidefinite programming phase retrieval technique from [59] or the greedy sparse technique from [55]. Similar conclusions were drawn when applying our method to a phase retrieval problem involving complex-valued images [50]. Finally, we would like to emphasize that, while this paper was under revision, we have been made aware of [15] where a nonconvex variational approach for phase reconstruction was developed in an independent manner. The advantage of our approach is to easily deal with a constraint or a regularization term so as to model prior knowledge on the sought solution, which is of major importance when the inverse problem is under-determined, as it is the case here.

References

1. Abboud, F., Chouzenoux, E., Pesquet, J.-C., Chenot, J.H., Laborelli, L.: A hybrid alternating proximal method for blind video restoration. In: Proceedings of European Signal Processing Conference (EUSIPCO 2014), pp. 1811–1815. Lisboa, Portugal (2014)
2. Attouch, H., Bolte, J.: On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. Math. Program. **116**, 5–16 (2009)
3. 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)
4. 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)
5. Auslender, A.: Asymptotic properties of the Fenchel dual functional and applications to decomposition problems. J. Optim. Theory Appl. **73**(3), 427–449 (1992)
6. 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. Am. A **19**(7), 1334–1345 (2002)
7. Bauschke, H.H., Combettes, P.L., Luke, D.R.: A new generation of iterative transform algorithms for phase contrast tomography. In: Proceedings of IEEE International Conference Acoust., Speech Signal Process. (ICASSP 2005), vol. 4, pp. 89–92. Philadelphia, PA (2005)
8. Bauschke, H.H., Combettes, P.L., Noll, D.: Joint minimization with alternating Bregman proximity operators. Pac. J. Optim. **2**(3), 401–424 (2006)
9. Bertsekas, D.P.: Nonlinear Programming, 2nd edn. Athena Scientific, Belmont, MA (1999)
10. 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)
11. Bolte, J., Daniilidis, A., Lewis, A., Shiota, M.: Clarke subgradients of stratifiable functions. SIAM J. Optim. **18**(2), 556–572 (2007)
12. Bolte, J., Daniilidis, A., Ley, O., Mazet, L.: Characterizations of Łojasiewicz inequalities: subgradient flows, talweg, convexity. Trans. Am. Math. Soc. **362**(6), 3319–3363 (2010)

13. Bolte, J., Sabach, S., Teboulle, M.: Proximal alternating linearized minimization for nonconvex and nonsmooth problems. Math. Program. **146**(1), 459–494 (2014)
14. Brègman, L.M.: The method of successive projection for finding a common point of convex sets. Soviet Math. Dokl. **6**, 688–692 (1965)
15. Candès, E., Eldar, Y., Strohmer, T., Voroninski, V.: Phase retrieval via matrix completion. SIAM J. Imaging Sci. **6**(1), 199–225 (2013)
16. Censor, Y., Lent, A.: Optimization of $\log x$ entropy over linear equality constraints. SIAM J. Control Optim. **25**(4), 921–933 (1987)
17. Chaux, C., Combettes, P.L., Pesquet, J.-C., Wajs, V.R.: A variational formulation for frame based inverse problems. Inverse Probl. **23**(4), 1495–1518 (2007)
18. Chouzenoux, E., Pesquet, J.-C., Repetti, A.: Variable metric forward-backward algorithm for minimizing the sum of a differentiable function and a convex function. J. Optim. Theory Appl. **162**(1), 107–132 (2014)
19. Combettes, P.L., Pesquet, J.-C.: Proximal splitting methods in signal processing. In: Bauschke, H.H., Burachik, R., Combettes, P.L., Elser, V., Luke, D.R., Wolkowicz, H. (eds.) Fixed-Point Algorithms for Inverse Problems in Science and Engineering, pp. 185–212. Springer, New York (2010)
20. Combettes, P.L., Pesquet, J.-C.: Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping. SIAM J. Optim. **25**, 1221–1248 (2015)
21. Combettes, P.L., Vũ, B.C.: Variable metric quasi-Fejér monotonicity. Nonlinear Anal. **78**, 17–31 (2013)
22. Combettes, P.L., Vũ, B.C.: Variable metric forward-backward splitting with applications to monotone inclusions in duality. Optimization **63**(9), 1289–1318 (2014)
23. Combettes, P.L., Wajs, V.R.: Signal recovery by proximal forward-backward splitting. Multiscale Model. Simul. **4**(4), 1168–1200 (2005)
24. Dainty, J.C., Fienup, J.R.: Phase retrieval and image reconstruction for astronomy. In: Stark, H. (ed.) Image Recovery: Theory and Application, pp. 231–275. Academic Press, Orlando, FL (1987)
25. Fessler, J.A.: Grouped coordinate ascent algorithms for penalized-likelihood transmission image reconstruction. IEEE Trans. Med. Imag. **16**(2), 166–175 (1997)
26. Fienup, J.R.: Phase retrieval algorithms: a comparison. Appl. Opt. **21**(15), 2758–2769 (1982)
27. Frankel, P., Garrigos, G., Peypouquet, J.: Splitting methods with variable metric for Kurdyka-Łojasiewicz functions and general convergence rates. J. Optim. Theory Appl. **165**(3), 874–900 (2015)
28. 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)
29. Golub, G.H., Van Loan, C.F.: Matrix Computations, 3rd edn. Johns Hopkins University Press, Baltimore (1996)
30. Hesse, R., Luke, D.R., Sabach, S., Tam, M.K.: Proximal heterogeneous block input-output method and application to blind ptychographic diffraction imaging. Tech. rep. (2014). arXiv:1408.1887
31. Hiriart-Urruty, J.B., Lemaréchal, C.: Convex Analysis and Minimization Algorithms. Springer, New York (1993)
32. 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)
33. 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)
34. Łojasiewicz, S.: Une propriété topologique des sous-ensembles analytiques réels. Editions du centre National de la Recherche Scientifique, pp. 87–89 (1963)
35. Luenberger, D.G.: Linear and Nonlinear Programming. Addison-Wesley, Reading (1973)
36. 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)
37. 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)
38. Mallat, S.: A Wavelet Tour of Signal Processing, 3rd edn. Academic Press, Burlington (2009)
39. Mordukhovich, B.S.: Variational Analysis and Generalized Differentiation. Vol. I: Basic theory, Series of Comprehensive Studies in Mathematics, vol. 330. Springer, Berlin (2006)
40. Moreau, J.J.: Proximité et dualité dans un espace hilbertien. Bull. Soc. Math. France **93**, 273–299 (1965)
41. Mukherjee, S., Seelamantula, C.S.: An iterative algorithm for phase retrieval with sparsity constraints: application to frequency domain optical coherence tomography. In: Proceedings of IEEE Internationl Conference Acoust., Speech and Signal Process. (ICASSP 2012), pp. 553–556. Kyoto, Japan (2012)
42. Ochs, P., Chen, Y., Brox, T., Pock, T.: iPiano: inertial proximal algorithm for non-convex optimization. SIAM J. Imaging Sci. **7**(2), 1388–1419 (2014)
43. Ortega, J.M., Rheinboldt, W.C.: Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, New York (1970)

44. Pirayre, A., Couprie, C., Duval, L., Pesquet, J.-C.: Fast convex optimization for connectivity enforcement in gene regulatory network inference. In: Proceedings of IEEE International Conference Acoust., Speech Signal Process. (ICASSP 2015), pp. 1002–1006. Brisbane, Australia (2015)
45. Powell, M.J.D.: On search directions for minimization algorithms. Math. Program. **4**, 193–201 (1973)
46. Pustelnik, N., Benazza-Benhayia, A., Zheng, Y., Pesquet, J.-C.: Wavelet-based image deconvolution and reconstruction. To appear in Wiley Encyclopedia of Electrical and Electronics Engineering (2016). https://hal.archives-ouvertes.fr/hal-01164833v1
47. 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)
48. Repetti, A., Pham, M.Q., Duval, L., Chouzenoux, E., Pesquet, J.-C.: Euclid in a taxicab: Sparse blind deconvolution with smoothed ℓ_1/ℓ_2 regularization. IEEE Signal Process. Lett. **22**(5), 539–543 (2015)
49. Repetti, A., Chouzenoux, E., Pesquet, J.-C.: A preconditioned forward-backward approach with application to large-scale nonconvex spectral unmixing problems. In: Proceedings of IEEE International Conference Acoust., Speech Signal Process. (ICASSP 2014), pp. 1498–1502. Firenze, Italy (2014)
50. Repetti, A., Chouzenoux, E., Pesquet, J.-C.: A nonconvex regularized approach for phase retrieval. In: Proceedings of IEEE International Conference Image Process. (ICIP 2014), pp. 1753–1757. Paris, France (2014)
51. Richtárik, P., Talác, M.: Iteration complexity of randomized block-coordinate descent methods for minimizing a composite function. Math. Program. **144**(1), 1–38 (2014)
52. Rockafellar, R.T., Wets, R.J.B.: Variational Analysis, 1st edn. Springer, Berlin (1997)
53. Saquib, S., Zheng, J., Bouman, C.A., Sauer, K.D.: Parallel computation of sequential pixel updates in statistical tomographic reconstruction. In: Proceedings of IEEE International Conference Image Process. (ICIP 1995), vol. 2, 93–96. Washington, DC (1995)
54. Saxton, W.O.: Computer Techniques for Image Processing in Electron Microscopy. Academic Press, New York (1978)
55. Shechtman, Y., Beck, A., Eldar, Y.: GESPAR: efficient phase retrieval of sparse signals. IEEE Trans. Signal Process. **4**(62), 928–938 (2014)
56. Sotthivirat, S., Fessler, J.A.: Image recovery using partitioned-separable paraboloidal surrogate coordinate ascent algorithms. IEEE Trans. Signal Process. **11**(3), 306–317 (2002)
57. Tappenden, R., Richtárik, P., Gondzio, J.: Inexact coordinate descent: complexity and preconditioning. J. Optim. Theory Appl. (to appear). arXiv:1304.5530v2
58. Tseng, P.: Convergence of a block coordinate descent method for nondifferentiable minimization. J. Optim. Theory Appl. **109**(3), 475–494 (2001)
59. Waldspurger, I., d'Aspremont, A., Mallat, S.: Phase recovery, maxcut and complex semidefinite programming. Math. Program. **149**(1), 47–81 (2015)
60. 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)
61. Xu, Y., Yin, W.: A globally convergent algorithm for nonconvex optimization based on block coordinate update. Tech. rep. (2014). arXiv:1410.1386
62. Zangwill, W.I.: Nonlinear Programming. Prentice-Hall, Englewood Cliffs (1969)