Noname manuscript No.
(will be inserted by the editor)

Convergence Analysis of Block Majorize-Minimize Subspace Approach

Emilie Chouzenoux and Jean-Baptiste Fest

the date of receipt and acceptance should be inserted later

Abstract We consider the minimization of a differentiable Lipschitz gradient but non necessarily convex, function F defined on $\mathbb{R}^N$. We propose an accelerated gradient descent approach which combines three strategies, namely (i) a variable metric derived from the majorization-minimization principle ; (ii) a subspace strategy incorporating information from the past iterates ; (iii) a block alternating update. Under the assumption that F satisfies the Kurdyka-Łojasiewicz property, we give conditions under which the sequence generated by the resulting block majorize-minimize subspace algorithm converges to a critical point of the objective function, and we exhibit convergence rates for its iterates.

Keywords Block alternating method, majorization-minimization, memory gradient, quasi-Newton, non-convex optimization, Kurdyka-Lojasiewicz.

1 Introduction

Our work focuses on the resolution of

$$\text{minimize}_{\boldsymbol{x}\in\mathbb{R}^N} F(\boldsymbol{x}), \tag{1}$$

with $F : \mathbb{R}^N \to \mathbb{R}$ is a differentiable Lipschitz gradient function which is not assumed to be convex. Instead, we address the case when F satisfies the Kurdyka-Łojasiewicz (KL) inequality [1,2].

In the case of large scale optimization problems, one major concern is to find an optimization algorithm able to deliver reliable numerical solutions in a reasonable time. Numerous works have been devoted to accelerate the first order gradient descent technique. These methods aim to increase the convergence rate while preserving theoretical guarantees and limited computational cost/memory burden per iteration. Three main families of acceleration strategies can be distinguished in the literature. The first approach, adopted for example in the well-known L-BFGS [3] and non-linear conjugate gradient [4] methods, relies on subspace acceleration [5,6]. The convergence rate is improved by using information from past iterates for the construction of new estimates. Another efficient way to accelerate the convergence of a minimization algorithm is based on a variable metric (i.e., preconditioning) strategy [7,8]. The underlying metric is modified at each iteration thanks to a preconditioning matrix, which may incorporate structural second-order information about the function to minimize. The third technique to limit the dependence of an optimization algorithm on the dimension of the problem, is to adopt a block alternating scheme where, at each iteration, only a subset of the variables are updated [9].

Among various choices for preconditioning first-order methods, an important class of techniques rely on the principle of Majorization-Minimization (MM) [10,11]. At each iteration, a quadratic convex surrogate function majorizing F is constructed. The inverse of its curvature (i.e., Hessian) matrix then serves to define a weighted Euclidean metric used for updating the next iterate. This idea is at the core of the half-quadratic algorithm [12,13] for image restoration. It has also been exploited in [14] to build an accelerated proximal gradient method for non smooth optimization, with guaranteed convergence of the iterates to a stationary point, in the non-convex case. The

Laboratoire CVN, CentraleSupélec, Inria, Université Paris-Saclay, 9 rue Joliot Curie, Gif-sur-Yvette, France
Corresponding author: Jean-Baptiste Fest, `jean-baptiste.fest@inria.fr`

latter result has been then extended in [15], where block alternating updates are introduced. Block alternating MM approaches have also been explored in [16,17,18,19], although without established convergence guarantees on their iterates in the non-convex setting. MM metrics are also well suited to the construction of efficient subspace optimization methods [20,21,22,10]. In [20], subspace acceleration is employed to reduce the complexity of an MM algorithm in large scale image processing problems, and in [21], convergence guarantees are obtained on the iterates under the KL assumption. This algorithm has recently been extended in [23] to the resolution of convex constrained optimization problems.

In this letter, we propose to bridge the gap between the theoretical analysis from [15] and [20]. We introduce the Block MM Subspace (B2MS) algorithm to solve (1), that incorporates the three aforementioned catalyzing effects, namely (i) MM-based preconditioning, (ii) subspace acceleration, (iii) block alternating update. We show the convergence of its iterates to a critical point of F. We furthermore perform its convergence rate analysis, relying on the KL exponent properties from [24].

The paper is organized as follows. Section 2 introduces the notation, the proposed B2MS algorithm, and the considered assumptions. Section 3 provides technical descent lemmas essential to our analysis. Section 4 presents our main contribution, namely the proof of convergence for B2MS, and a study of its convergence rate.

2 Block MM subspace algorithm

2.1 Notation

We consider the Euclidean space $\mathbb{R}^N$, endowed with the scalar product $\langle\cdot|\cdot\rangle$ and norm $\|\cdot\|$. $\boldsymbol{I}_N$ states for the identity matrix of $\mathbb{R}^N$. For any $\boldsymbol{A} \in \mathbb{R}^{N\times N}$ symmetric definite positive (SDP), we also introduce the weighted norm $\|\cdot\|_{\boldsymbol{A}} = \sqrt{\langle\cdot|\boldsymbol{A}\cdot\rangle}$. Let $\mathcal{S} \subset \{1,\dots,N\} \triangleq [\![1,N]\!]$ with cardinal $|\mathcal{S}|$ and complementary set $\overline{\mathcal{S}} \triangleq [\![1,N]\!]/\mathcal{S}$. For all $\boldsymbol{x} = (x^n)_{n\in[\![1,N]\!]} \in \mathbb{R}^N$, we denote $\boldsymbol{x}^{(\mathcal{S})} \triangleq \left(\boldsymbol{x}^i\right)_{i\in\mathcal{S}} \in \mathbb{R}^{|\mathcal{S}|}$. Similarly, the restriction to block $\mathcal{S}$ of the gradient of F, at some $\boldsymbol{x} \in \mathbb{R}^N$ reads $\nabla F^{(\mathcal{S})}(\boldsymbol{x}) \in \mathbb{R}^{|\mathcal{S}|}$. For any $\boldsymbol{x} \in \mathbb{R}^N$, we finally introduce function $F^{(\mathcal{S})}(.\,,\boldsymbol{x}) : \boldsymbol{v} \in \mathbb{R}^{|\mathcal{S}|} \mapsto F(\boldsymbol{u})$ where $\boldsymbol{u}^{(\mathcal{S})} = \boldsymbol{v}$ and $\boldsymbol{u}^{(\overline{\mathcal{S}})} = \boldsymbol{x}^{(\overline{\mathcal{S}})}$.

2.2 B2MS scheme

The proposed B2MS algorithm solves (1) through a block alternating minimization approach. Let $\mathbb{S}$ a family of $C \geq 1$ nonempty subsets of $[\![1,N]\!]$ (not necessarily disjoints). Let $\boldsymbol{x}_0 \in \mathbb{R}^N$. At every iteration $k \in \mathbb{N}$, the entries of the current iterate $\boldsymbol{x}_k$ within a selected block $\mathcal{S}_k \in \mathbb{S}$ are updated using one iteration of the MM subspace algorithm [20] on the restriction of F to the k-th block $F^{(\mathcal{S}_k)}(\cdot,\boldsymbol{x}_k)$. The entries of $\boldsymbol{x}_k$ within the complementary set $\overline{S}_k$ remain constant.

To implement the MM subspace update, we first build the following *quadratic majorant approximation* [10] of $F^{(\mathcal{S}_k)}(\cdot,\boldsymbol{x}_k)$ at $\boldsymbol{x}_k$,

$$(\forall \boldsymbol{v} \in \mathbb{R}^{|\mathcal{S}_k|}) \quad Q^{(\mathcal{S}_k)}(\boldsymbol{v},\boldsymbol{x}_k) \triangleq F(\boldsymbol{x}_k) + \langle \nabla F^{(\mathcal{S}_k)}(\boldsymbol{x}_k), \boldsymbol{v} - \boldsymbol{x}_k^{(\mathcal{S}_k)}\rangle + \frac{1}{2}\|\boldsymbol{v} - \boldsymbol{x}_k^{(\mathcal{S}_k)}\|^2_{\boldsymbol{A}^{(\mathcal{S}_k)}(\boldsymbol{x}_k)}, \tag{2}$$

where $\boldsymbol{A}^{(\mathcal{S}_k)}(\boldsymbol{x}_k) \in \mathbb{R}^{|\mathcal{S}_k|\times|\mathcal{S}_k|}$ is an SDP matrix such that:

$$(\forall \boldsymbol{v} \in \mathbb{R}^{|\mathcal{S}_k|}) \quad F^{(\mathcal{S}_k)}(\boldsymbol{v},\boldsymbol{x}_k) \leq Q^{(\mathcal{S}_k)}(\boldsymbol{v},\boldsymbol{x}_k). \tag{3}$$

Second, we choose a *subspace acceleration* matrix $\boldsymbol{D}_k \in \mathbb{R}^{|\mathcal{S}_k|\times M_k}$ [5]. The block update $\boldsymbol{x}_{k+1}^{(\mathcal{S}_k)}$ is then defined as a minimizer of $Q^{(\mathcal{S}_k)}(\cdot,\boldsymbol{x}_k)$ within the vectorial subspace spanned by the columns of $\boldsymbol{D}_k$. Iterating the above procedure in a block alternating fashion yields Algorithm (4):

$$\begin{array}{l} \text{Initialize} \quad \boldsymbol{x}_0 \in \mathbb{R}^N. \\ \text{For} \quad k = 0,1,2,\dots \\ \textit{(B2MS)} \left\lfloor \begin{array}{l} \text{Choose } \mathcal{S}_k \in \mathbb{S} \text{ and } \boldsymbol{D}_k \in \mathbb{R}^{|\mathcal{S}_k|\times M_k} \\ \boldsymbol{u}_k \in \underset{\boldsymbol{u}\in\mathbb{R}^{M_k}}{\arg\min}\, Q^{(\mathcal{S}_k)}\left(\boldsymbol{x}_k^{(\mathcal{S}_k)} + \boldsymbol{D}_k\boldsymbol{u}, \boldsymbol{x}_k\right) \\ \boldsymbol{x}_{k+1}^{(\mathcal{S}_k)} = \boldsymbol{x}_k^{(\mathcal{S}_k)} + \boldsymbol{D}_k\boldsymbol{u}_k \\ \boldsymbol{x}_{k+1}^{(\overline{\mathcal{S}_k})} = \boldsymbol{x}_k^{(\overline{\mathcal{S}_k})} \end{array}\right. \end{array} \tag{4}$$

If $\mathbb{S} = \{[\![1, N]\!]\}$ (so $C = 1$), we retrieve the MM subspace algorithm from [20,21]. When $\boldsymbol{D}_k = \boldsymbol{I}_{|\mathcal{S}_k|}$, the above approach can be viewed as a particular case of the BSUM scheme [16,25] using quadratic surrogates, or of the approach from [15] using a null proximal term.

2.3 Assumptions

Assumption 1.
Family $\mathbb{S}$ verifies $\bigcup_{\mathcal{S}\in\mathbb{S}} \mathcal{S} = [\![1, N]\!]$. Moreover, there exists $K \in \mathbb{N}^*$ such that, for all $k \in \mathbb{N}$, every $\mathcal{S} \in \mathbb{S}$ belongs to $\{\mathcal{S}_k, \ldots, \mathcal{S}_{k+K-1}\}$.

Assumption 2. *F is $\mathcal{C}^1$ and coercive on $\mathbb{R}^N$. Moreover, it is B-Lipschitz gradient, that is there exists $B > 0$ such that*

$$(\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N) \quad \|\nabla F(\boldsymbol{x}) - \nabla F(\boldsymbol{y})\| \leq B\|\boldsymbol{x} - \boldsymbol{y}\|. \tag{5}$$

Assumption 3.
(i) For all $k \in \mathbb{N}$, $\boldsymbol{D}_k$ has full column rank.
(ii) There exists $(\gamma_0, \gamma_1) > 0$ such that, for all $k \in \mathbb{N}$,

$$(\boldsymbol{d}_k)^\top \nabla F^{(\mathcal{S}_k)}(\boldsymbol{x}_k) \leq -\gamma_0 \|\nabla F^{(\mathcal{S}_k)}(\boldsymbol{x}_k)\|^2, \tag{6}$$

$$\|\boldsymbol{d}_k\| \leq \gamma_1 \|\nabla F^{(\mathcal{S}_k)}(\boldsymbol{x}_k)\|, \tag{7}$$

with $\boldsymbol{d}_k \in \mathbb{R}^{|\mathcal{S}_k|}$ the first column of $\boldsymbol{D}_k$.

Assumption 4. *There exists $(\eta, \nu) > 0$ s.t*

$$(\forall k \in \mathbb{N}) \quad \eta \boldsymbol{I}_{|\mathcal{S}_k|} \preceq \boldsymbol{A}^{(\mathcal{S}_k)}(\boldsymbol{x}_k) \preceq \nu \boldsymbol{I}_{|\mathcal{S}_k|}. \tag{8}$$

Assumption 5. *For $\xi \in \mathbb{R}$ and any bounded $E \subset \mathbb{R}^N$, there exists $(\kappa, \zeta, \theta) \in \mathbb{R}_+^* \times \mathbb{R}_+^* \times]0, 1[$ such that, for all $\boldsymbol{x} \in E$ with $|F(\boldsymbol{x}) - \xi| \leq \zeta$,*

$$\|\nabla F(\boldsymbol{x})\| \geq \kappa \, |F(\boldsymbol{x}) - \xi|^\theta \,. \tag{9}$$

Assumption 1, also adopted in [15], implies that every set of $\mathbb{S}$ is updated at least once during any K-length cycle. It is also known as quasi-cyclic or acyclic rule [18], the cyclic rule being a special case of it. Assumption 2 ensures the existence of a minimizer for F, as well as the existence for quadratic majorant approximations for F (by descent lemma [26]). Assumption 3(ii) is equivalent to imposing a gradient-related condition [26] on the first column of $\boldsymbol{D}_k$. This is satisfied by a large number of typical subspace acceleration matrices [20][Tab. I]. In particular, setting $\boldsymbol{D}_k = [-\nabla F(\boldsymbol{x}_k) | \boldsymbol{x}_k - \boldsymbol{x}_{k-1}] \in \mathbb{R}^{N\times 2}$ for $k \in \mathbb{N}^*$ yields the memory gradient subspace [27, 28] which has strong connections with the non-linear conjugate gradient method [29]. Assumption 4 is rather mild, and inherent to the stability of quadratic MM schemes [20,14]. Finally, Assumption 5 is usually refereed to as the KL inequality [1,2], and arises from the literature of non-smooth analysis. It is satisfied by a large variety of functions, non necessarily convex, such as semi-algebraic or analytical functions, to name a few. Its use has become popular in the last decade, as it provides a key tool for establishing convergence of iterates for descent methods in the nonconvex setting [24, 15,8].

3 Technical lemmas

This section presents technical lemmas that turn out to be essential to our convergence analysis.

Lemma 1 *Under Assumptions 1-4, the B2MS sequence $(\boldsymbol{x}_k)_{k\in\mathbb{N}}$ satisfies:*

$$(\forall k \in \mathbb{N}) \quad F(\boldsymbol{x}_k) - F(\boldsymbol{x}_{k+1}) \geq \frac{\gamma_0^2}{2\gamma_1^2 \nu} \|\nabla F^{(\mathcal{S}_k)}(\boldsymbol{x}_k)\|^2, \tag{10}$$

$$\sum_{k=0}^{+\infty} \|\nabla F^{(\mathcal{S}_k)}(\boldsymbol{x}_k)\|^2 < +\infty, \tag{11}$$

$$(\forall k \in \mathbb{N}) \quad \|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|^2 \leq \frac{1}{\eta^2} \|\nabla F^{(\mathcal{S}_k)}(\boldsymbol{x}_k)\|^2, \tag{12}$$

$$(\forall k \in \mathbb{N}) \quad \|\nabla F^{(\mathcal{S}_k)}(\boldsymbol{x}_k)\|^2 \leq \frac{\gamma_1^2 \nu^2}{\gamma_0^2} \|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|^2. \tag{13}$$

Proof. Consider $k \in \mathbb{N}$. Since $\boldsymbol{x}_{k+1}$ is the concatenation of $\boldsymbol{x}_{k+1}^{(\mathcal{S}_k)}$, $\boldsymbol{x}_k^{(\overline{\mathcal{S}_k})}$, the use of (3) gives,

$$F(\boldsymbol{x}_{k+1}) = F^{(\mathcal{S}_k)}(\boldsymbol{x}_{k+1}^{(\mathcal{S}_k)}, \boldsymbol{x}_k) \leq Q^{(\mathcal{S}_k)}\left(\boldsymbol{x}_{k+1}^{(\mathcal{S}_k)}, \boldsymbol{x}_k\right) = Q^{(\mathcal{S}_k)}\left(\boldsymbol{x}_k^{(\mathcal{S}_k)} + \boldsymbol{D}_k \boldsymbol{u}_k, \boldsymbol{x}_k\right). \tag{14}$$

Let $\boldsymbol{e} = (1, 0, \dots, 0)^\top \in \mathbb{R}^{M_k}$. Since $\boldsymbol{u}_k$ minimizes $Q^{(\mathcal{S}_k)}\left(\boldsymbol{x}_k^{(\mathcal{S}_k)} + \boldsymbol{D}_k\ .\ , \boldsymbol{x}_k\right)$, any $t \in \mathbb{R}$ satisfies

$$Q^{(\mathcal{S}_k)}\left(\boldsymbol{x}_k^{(\mathcal{S}_k)} + \boldsymbol{D}_k \boldsymbol{u}_k, \boldsymbol{x}_k\right) \leq Q^{(\mathcal{S}_k)}\left(\boldsymbol{x}_k^{(\mathcal{S}_k)} + \boldsymbol{D}_k t\boldsymbol{e}, \boldsymbol{x}_k\right) = Q^{(\mathcal{S}_k)}\left(\boldsymbol{x}_k^{(\mathcal{S}_k)} + t\ \boldsymbol{d}_k, \boldsymbol{x}_k\right).$$

Moreover, $t \mapsto Q^{(\mathcal{S}_k)}\left(\boldsymbol{x}_k^{(\mathcal{S}_k)} + t\ \boldsymbol{d}_k, \boldsymbol{x}_k\right)$ is scalar quadratic with $F(\boldsymbol{x}_k) - \frac{\left\langle \nabla F^{(\mathcal{S}_k)}(\boldsymbol{x}^k), \boldsymbol{d}_k \right\rangle^2}{2\|\boldsymbol{d}_k\|^2_{\boldsymbol{A}^{(\mathcal{S}_k)}(\boldsymbol{x}^k)}}$ as minimal value. Hence, Assumptions 3-4 leads to

$$Q^{(\mathcal{S}_k)}\left(\boldsymbol{x}_k^{(\mathcal{S}_k)} + \boldsymbol{D}_k \boldsymbol{u}_k, \boldsymbol{x}_k\right) \leq F(\boldsymbol{x}_k) - \frac{\left\langle \nabla F^{(\mathcal{S}_k)}(\boldsymbol{x}^k), \boldsymbol{d}_k \right\rangle^2}{2\|\boldsymbol{d}_k\|^2_{\boldsymbol{A}^{(\mathcal{S}_k)}(\boldsymbol{x}^k)}} \leq F(\boldsymbol{x}_k) - \frac{\gamma_0^2}{2\gamma_1^2 \nu}\|\nabla F^{(\mathcal{S}_k)}(\boldsymbol{x}_k)\|^2. \tag{15}$$

Combination of (15) with (14) directly gives (10) in Lemma 1. $(F(\boldsymbol{x}_k))_{k\in\mathbb{N}}$ is then a decreasing sequence. Coercivity of F thus guarantees its convergence to a limit $F^\infty \in \mathbb{R}$. Moreover,

$$\sum_{k=0}^{+\infty} \|\nabla F^{(\mathcal{S}_k)}(\boldsymbol{x}_k)\|^2 \leq \frac{2\gamma_1^2 \nu}{\gamma_0^2}\left(F^\infty - F(\boldsymbol{x}_0)\right), \tag{16}$$

so that (11) in Lemma 1 is obtained. Since function $Q^{(\mathcal{S}_k)}\left(\boldsymbol{x}_k^{(\mathcal{S}_k)} + \boldsymbol{D}_k\ .\ , \boldsymbol{x}_k\right)$ is quadratic, we also deduce that its minimizer $\boldsymbol{u}_k$ satisfies:

$$\left(\nabla F^{(\mathcal{S}_k)}(\boldsymbol{x}_k)\right)^\top \boldsymbol{D}_k^{(\mathcal{S}_k)} \boldsymbol{u}_k = \left(\boldsymbol{u}^k\right)^\top \|\boldsymbol{D}_k\|^2_{A^{(\mathcal{S}_k)}(\boldsymbol{x}_k)}. \tag{17}$$

Equality $\boldsymbol{D}_k \boldsymbol{u}_k = \boldsymbol{x}_{k+1} - \boldsymbol{x}_k$, and Assumption 4, lead to

$$Q^{(\mathcal{S}_k)}\left(\boldsymbol{x}_k^{(\mathcal{S}_k)} + \boldsymbol{D}_k \boldsymbol{u}_k, \boldsymbol{x}_k\right) = F(\boldsymbol{x}_k) - \frac{1}{2}\|\boldsymbol{D}_k \boldsymbol{u}^k\|^2_{\boldsymbol{A}^{(\mathcal{S}_k)}(\boldsymbol{x}_k)} \geq F(\boldsymbol{x}_k) - \frac{\nu}{2}\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|^2. \tag{18}$$

Equation (13) of Lemma 1 then comes by plugging (18) into (15). Using again the expression of $\boldsymbol{u}_k$ as a minimizer of a quadratic form, we can rewrite one iteration of B2MS scheme as

$$\boldsymbol{x}_{k+1} - \boldsymbol{x}_k = -\boldsymbol{B}_k\ \nabla F^{(\mathcal{S}_k)}(\boldsymbol{x}_k), \tag{19}$$

with $\boldsymbol{B}_k \triangleq \boldsymbol{D}_k \left(\boldsymbol{D}_k^\top \boldsymbol{A}^{(\mathcal{S}_k)}(\boldsymbol{x}_k) \boldsymbol{D}_k\right)^{-1} \boldsymbol{D}_k^\top$. Remark that Assumption 3(i) ensures that $\boldsymbol{B}_k$ is a well-defined symmetric definite positive matrix. Then, by Assumption 4,

$$\boldsymbol{B}_k \preceq \frac{1}{\eta}\boldsymbol{D}_k \left(\boldsymbol{D}_k^\top \boldsymbol{D}_k\right)^{-1} \boldsymbol{D}_k^\top \preceq \frac{1}{\eta}\boldsymbol{I}_{|\mathcal{S}_k|}. \tag{20}$$

Plugging (20) into (19) and taking the squared norm of the quantities gives (12) of Lemma 1. □

Lemma 2 *Let* $(u_k)_{k\in\mathbb{N}}, (v_k)_{k\in\mathbb{N}}$ *two sequences of positive real. If there exists* $k^* \geq K$ *such that*

$$(\forall k \geq k^*) \quad u_k \leq \rho \sum_{i=k-K}^{k-1} u_i + v_{k-1}, \tag{21}$$

if $\rho < \frac{1}{K}$ *and* $\sum_{k=0}^{+\infty} v_k < +\infty$ *, then* $\sum_{k=0}^{+\infty} u_k < +\infty$.

Proof. Summing (21) from k^* to $n \geq k^*$ leads to

$$\sum_{k=k^*}^{n} u_k \leq \rho \sum_{k=k^*}^{n} \sum_{i=k-K}^{k-1} u_i + \sum_{k=k^*}^{n} v_{k-1}, \tag{22}$$

with

$$\sum_{k=k^*}^{n} \sum_{i=k-K}^{k-1} u_i = \sum_{k=k^*}^{n} \sum_{i=1}^{K} u_{k-i} = \sum_{i=1}^{K} \sum_{k=k^*-i}^{n-i} u_k \leq \sum_{i=1}^{K} \sum_{k=0}^{n} u_k. \tag{23}$$

Plugging (23) into (22), yields

$$\sum_{k=k^*}^{n} u_k \leq \rho K \sum_{k=k*}^{n} u_k + \left(\rho K \sum_{k=0}^{k^*-1} u_k + \sum_{k=k^*}^{n} v_{k-1}\right) \leq \rho K \sum_{k=k*}^{n} u_k + \left(\rho K \sum_{k=0}^{k^*-1} u_k + \sum_{k=0}^{+\infty} v_k\right), \tag{24}$$

that is $(1-\rho K)\sum_{k=k^*}^{n} u_k \leq \rho K \sum_{k=0}^{k^*-1} u_k + \sum_{k=0}^{+\infty} v_k$. With $0 < 1-\rho K < 1$, we deduce the summability of $(u_k)_{k\in\mathbb{N}}$. □

Lemma 1 gathers the different inequalities and descent properties which result from B2MS scheme (4). Equations (10)-(11) can be interpreted as a generalized block version of [20, Theorem 1]. Lemma 2 is an alternative of [30, Lemma 3], [31, Lemma 5.1].

4 Asymptotical behaviour

For the sake of clarity, our presentation for the convergence analysis of scheme (4) is divided into three parts. First, we establish the convergence of the gradient of the B2MS iterates to zero, under Assumptions 1-4. Second, under the additional Assumption 5 (i.e., KL inequality), we show the convergence of the iterates of B2MS to a stationary point. Third, we establish a convergence rate result involving the KL exponent θ of function F.

4.1 Global convergence

Theorem 1 *Let $(\boldsymbol{x}_k)_{k\in\mathbb{N}}$ the B2MS sequence. Under Assumptions 1-4, sequence $(\|\nabla F(\boldsymbol{x}_k)\|)_{k\in\mathbb{N}}$ converges to 0. Moreover, there exists $\boldsymbol{x}^*$, a stationary point of F, such that $F(\boldsymbol{x}_k) \underset{k\to+\infty}{\longrightarrow} F(\boldsymbol{x}^*)$.*

Proof. Let $k \geq K$. For all $\mathcal{S} \in \mathbb{S}$, Assumption 1 ensures that $\{t \in [\![k-K, k-1]\!] \ / \ \mathcal{S}_t = \mathcal{S}\}$ is a non-empty set. We can thus rewrite every set of $\mathbb{S}$ as $\mathcal{S} = \mathcal{S}_{T_k^{\mathcal{S}}}$ with

$$T_k^{\mathcal{S}} \triangleq \max\{t \in [\![k-K, k-1]\!] \ / \ \mathcal{S}_t = \mathcal{S}\}.$$

The application of $k - T_k^{\mathcal{S}}$ Jensen's inequalities on $\|\nabla F^{\mathcal{S}}(\boldsymbol{x}_k)\|^2$ leads to

$$\|\nabla F^{(\mathcal{S})}(\boldsymbol{x}_k)\|^2 \leq \sum_{i=T_k^{\mathcal{S}}}^{k-1} 2^{k-i}\|\nabla F^{(\mathcal{S})}(\boldsymbol{x}_{i+1}) - \nabla F^{(\mathcal{S})}F(\boldsymbol{x}_i)\|^2 + 2^{k-T_k^{\mathcal{S}}}\|\nabla F^{(\mathcal{S})}F(\boldsymbol{x}_{T_k^{\mathcal{S}}})\|^2. \tag{25}$$

We now majorize both parts of the right term of (25). For the former, we apply successively (5) and (12). The latter part is handled by using $\mathcal{S} = \mathcal{S}_{T_k^{\mathcal{S}}}$ and noticing that $k-i \in]0, K[$ for $i \in [\![T_k^{\mathcal{S}}, k-1]\!]$. It yields

$$\begin{aligned}
\|\nabla F^{(\mathcal{S})}(\boldsymbol{x}_k)\|^2 &\leq \frac{2^K B^2}{\eta^2}\sum_{i=T_k^{\mathcal{S}}}^{k-1}\|\nabla F^{(\mathcal{S}_i)}(\boldsymbol{x}_i)\|^2 + 2^K\|\nabla F^{(\mathcal{S}_{T_k^{\mathcal{S}}})}(\boldsymbol{x}_{T_k^{\mathcal{S}}})\|^2,\\
&\leq 2^K\left(\frac{B^2}{\eta^2}+1\right)\sum_{i=T_k^{\mathcal{S}}}^{k-1}\|\nabla F^{(\mathcal{S}_i)}(\boldsymbol{x}_i)\|^2,\\
&\leq 2^K\left(\frac{B^2}{\eta^2}+1\right)\sum_{i=k-K}^{k-1}\|\nabla F^{(\mathcal{S}_i)}(\boldsymbol{x}_i)\|^2.
\end{aligned} \tag{26}$$

Then from (11) we have $\|\nabla F^{(\mathcal{S})}(\boldsymbol{x}_k)\|^2 \underset{k\to+\infty}{\longrightarrow} 0$. Since $[\![1, N]\!] = \bigcup_{\mathcal{S}\in\mathbb{S}} \mathcal{S}$ (by Ass. 1) and $\mathbb{S}$ finite,

$$\|\nabla F(\boldsymbol{x}_k)\|^2 \leq \sum_{\mathcal{S}\in\mathbb{S}}\|\nabla F^{(\mathcal{S})}(\boldsymbol{x}_k)\|^2 \underset{k\to+\infty}{\longrightarrow} 0. \tag{27}$$

Through Lemma 1 proof, we have already shown that $(F(\boldsymbol{x}_k))_{k\in\mathbb{N}}$ converges to a finite limit F^∞. The coercivity of F (Ass. 2) also ensures the boundedness of $(\boldsymbol{x}_k)_{k\in\mathbb{N}}$. Thus, there exists an accumulation point of $(\boldsymbol{x}_k)_{k\in\mathbb{N}}$, that we denote $\boldsymbol{x}^*$. As $\|\nabla F(\boldsymbol{x}_k)\| \underset{k\to+\infty}{\longrightarrow} 0$ and ∇F is continuous (by Ass. 2), $\nabla F(\boldsymbol{x}^*) = \boldsymbol{0}$ directly follows, so that $\boldsymbol{x}^*$ is a stationary point of F. Moreover, $F^\infty = F(\boldsymbol{x}^*)$ since F is continuous, which concludes our proof. □

Theorem 1 shows a classical behaviour for a descent method applied to non-convex Lipschitz gradient objective function [32].

4.2 Sequence convergence

This part is dedicated to refine the result of Theorem 1, when we additionally introduce Assumption 5. We first state a technical inequality giving a direct relation between the gradient at the current iterate and the differences of past iterates over a cycle (i.e a K-length period)

Proposition 1. *Let $(\boldsymbol{x}_k)_{k\in\mathbb{N}}$ the B2MS sequence. Under Assumptions 1-4, for all $k \geq K$,*

$$\|\nabla F(\boldsymbol{x}_k)\|^2 \leq 2^K C \left(B^2 + \frac{\gamma_1^2}{\gamma_0^2}\nu^2\right) \sum_{i=k-K}^{k-1} \|\boldsymbol{x}_{i+1} - \boldsymbol{x}_i\|^2. \tag{28}$$

Proof. Let $k \geq K$. The beginning of the proof is identical to that of Theorem 1 until Eq.(25). We then derive a new majoration for the quantity involved in the right term of (25). We first majorize using Ass. 2 combined with $\mathcal{S} = \mathcal{S}_{T_k^{\mathcal{S}}}$. We then use (13) of Lemma 1. This yields:

$$\begin{aligned}
\|\nabla F^{(\mathcal{S})}(\boldsymbol{x}_k)\|^2 &\leq 2^K B^2 \sum_{i=T_k^{\mathcal{S}}}^{k-1} \|\boldsymbol{x}_{i+1} - \boldsymbol{x}_i\|^2 + 2^K \|\nabla F^{(\mathcal{S}_{T_k^{\mathcal{S}}})}(\boldsymbol{x}_{T_k^{\mathcal{S}}})\|^2, \\
&\leq 2^K B^2 \sum_{i=T_k^{\mathcal{S}}}^{k-1} \|\boldsymbol{x}_{i+1} - \boldsymbol{x}_i\|^2 + 2^K \frac{\nu^2\gamma_1^2}{\gamma_0^2} \|\boldsymbol{x}_{T_k^{\mathcal{S}}+1} - \boldsymbol{x}_{T_k^{\mathcal{S}}}\|^2, \\
&\leq 2^K \left(B^2 + \frac{\gamma_1^2}{\gamma_0^2}\nu^2\right) \sum_{i=T_k^{\mathcal{S}}}^{k-1} \|\boldsymbol{x}_{i+1} - \boldsymbol{x}_i\|^2, \\
&\leq 2^K \left(B^2 + \frac{\gamma_1^2}{\gamma_0^2}\nu^2\right) \sum_{i=k-K}^{k-1} \|\boldsymbol{x}_{i+1} - \boldsymbol{x}_i\|^2. \qquad (29)
\end{aligned}$$

We then sum (29) over $\mathbb{S}$ (of cardinal C). The right term being independent from $\mathcal{S}$, it is simply multiplied by C. Then, noting that $\|\nabla F(\boldsymbol{x}_k)\|^2 \leq \sum_{\mathcal{S}\in\mathbb{S}} \|\nabla F^{(\mathcal{S})}(\boldsymbol{x}_k)\|^2$ concludes our proof. □

Remark that an alternative proof of Theorem 1 could be obtained from Prop. 1. However, inequality (26) is more direct to demonstrate than (28).

We finally state our main theoretical result, namely the convergence of the iterates of B2MS scheme.

Theorem 2 *Let $(\boldsymbol{x}_k)_{k\in\mathbb{N}}$ the B2MS sequence. Under Assumptions 1-5,*

$$\sum_{k=0}^{+\infty} \|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\| < +\infty. \tag{30}$$

Moreover, $(\boldsymbol{x}_k)_{k\in\mathbb{N}}$ converges to a stationary point of F.

Proof. Following the same notations as those of Theorem 1, let us apply (9) to set $E = \{\boldsymbol{x}_k \ / \ k \in \mathbb{N}\}$, which is a bounded set by our proof of Theorem 1, and $\xi = F^\infty$. Then, by KL inequality (Ass. 5), there exists $(\kappa, \zeta, \theta) \in \mathbb{R}_+^* \times \mathbb{R}_+^* \times]0, 1[$ such that for every $k \in \mathbb{N}$ verifying $|F(\boldsymbol{x}_k) - F^\infty| \leq \zeta$,

$$\|\nabla F(\boldsymbol{x}_k)\| \geq \kappa \left|F(\boldsymbol{x}_k) - F^\infty\right|^\theta. \tag{31}$$

Using (13) and (10) gives, for every $k \in \mathbb{N}$,

$$\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|^2 \leq \frac{2\gamma_1^2\nu}{\gamma_0^2\eta^2} \left(F(\boldsymbol{x}_k) - F(\boldsymbol{x}_{k+1})\right) = \frac{2\gamma_1^2\nu}{\gamma_0^2\eta^2} \left[\left(F(\boldsymbol{x}_k) - F^\infty\right) - \left(F(\boldsymbol{x}_{k+1}) - F^\infty\right)\right]. \tag{32}$$

Let us now invoke the convexity of $v \in \mathbb{R}_+ \mapsto v^{\frac{1}{1-\theta}}$. It follows that for all $v, w \in \mathbb{R}_+$ with $v \leq w$

$$w - v \leq (1-\theta)^{-1} w^\theta \left(w^{1-\theta} - v^{1-\theta}\right). \tag{33}$$

Plugging (33) in (32) with $w = F(\boldsymbol{x}_k) - F^\infty$ and $v = F(\boldsymbol{x}_{k+1}) - F^\infty$ yields, for every $k \in \mathbb{N}$,

$$\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|^2 \leq \frac{2\gamma_1^2\nu}{\gamma_0^2\eta^2(1-\theta)}[F(\boldsymbol{x}_k) - F^\infty]^\theta \Delta^k, \tag{34}$$

with $\Delta_k \triangleq \left(F(\boldsymbol{x}^k) - F^\infty\right)^{1-\theta} - \left(F(\boldsymbol{x}_{k+1}) - F^\infty\right)^{1-\theta}$. Since $F(\boldsymbol{x}_k) \underset{k\to+\infty}{\longrightarrow} F^\infty$, there exists $k_0 \geq K$ such that

$$(\forall k \geq k_0) \quad |F(\boldsymbol{x}_k) - F^\infty| \leq \zeta. \tag{35}$$

Thus, using (31),

$$(\forall k \geq k_0) \quad \|\nabla F(\boldsymbol{x}_k)\| \geq \kappa\, |F(\boldsymbol{x}_k) - F^\infty|^\theta. \tag{36}$$

Combining (36) and (34) leads to

$$(\forall k \geq k_0) \quad \|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|^2 \leq \frac{2\gamma_1^2\nu}{\gamma_0^2\eta^2\kappa(1-\theta)}\|\nabla F(\boldsymbol{x}_k)\|\Delta_k. \tag{37}$$

We now rely on the majoration of Lemma 1, to obtain

$$(\forall k \geq k_0) \quad \|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|^2 \leq \Lambda\sqrt{\sum_{i=k-K}^{k-1}\|\boldsymbol{x}_{i+1} - \boldsymbol{x}_i\|^2}\ \ \Delta_k \leq \Lambda \sum_{i=k-K}^{k-1}\|\boldsymbol{x}_{i+1} - \boldsymbol{x}_i\|\ \Delta_k, \tag{38}$$

with $\Lambda \triangleq \frac{2^{\frac{K}{2}+1}}{\kappa(1-\theta)}\frac{\gamma_1^2\nu}{\gamma_0^2\eta^2}\sqrt{C}\sqrt{B^2 + \frac{\gamma_1^2}{\gamma_0^2}\nu^2}$. We extract the square root of (38) and invoke the inequality $\sqrt{ab} \leq \frac{a}{c} + \frac{bc}{4}$ with $a = \sum\limits_{i=k-K}^{k-1}\|\boldsymbol{x}_{i+1} - \boldsymbol{x}_i\|$, $b = \Delta_k$ and some $c > 0$. Then,

$$(\forall k \geq k_0) \quad \|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\| \leq \frac{\sqrt{\Lambda}}{c}\sum_{i=k-K}^{k-1}\|\boldsymbol{x}_{i+1} - \boldsymbol{x}_i\| + \frac{c}{4}\sqrt{\Lambda}\Delta_k. \tag{39}$$

$(\Delta_k)_{k\in\mathbb{N}}$ is summable and $\sqrt{\Lambda}/c \in]0, 1/K[$ for $c > \sqrt{\Lambda}K$, we apply Lemma 2 with $k^* = k_0$. Finally, $(\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|)_{k\in\mathbb{N}}$ is summable and $(\boldsymbol{x}_k)_{k\in\mathbb{N}}$ is a Cauchy sequence possessing $\boldsymbol{x}^*$ as an accumulation point. Sequence $(\boldsymbol{x}_k)_{k\in\mathbb{N}}$ thus converges to $\boldsymbol{x}^*$. □

4.3 Convergence rate

As highlighted above, Theorem 2 guarantees the convergence of sequence $(\boldsymbol{x}_k)_{k\in\mathbb{N}}$ to $\boldsymbol{x}^*$, a stationary point of function F. Our last contribution lies in characterizing the convergence rate for B2MS algorithm. Hereagain, KL inequality (Ass. 5) is an anchor point of our analysis.

Theorem 3 *Let $(\boldsymbol{x}_k)_{k\in\mathbb{N}}$ the B2MS sequence. Under Assumptions 1-5, the following holds.*

(i) if $\theta \in]1/2, 1[$, then $\|\boldsymbol{x}_k - \boldsymbol{x}^*\| \underset{k\to+\infty}{=} \mathcal{O}\left(k^{-\frac{1-\theta}{2\theta-1}}\right)$;
(ii) If $\theta \in]0, 1/2]$, then there exists $\varepsilon \in]0, 1[$ such that $\|\boldsymbol{x}_k - \boldsymbol{x}^*\| \underset{k\to+\infty}{=} \mathcal{O}\left(\varepsilon^k\right)$.

Proof. Keeping the same notations as previously, let $c > 0$ and $k \geq k_0$. We sum (39) from kK

$$\Gamma_k \leq \frac{\sqrt{\Lambda}}{c}\sum_{j=kK}^{+\infty}\sum_{i=j-K}^{j-1}\|\boldsymbol{x}_{i+1} - \boldsymbol{x}_i\| + \frac{c}{4}\sqrt{\Lambda}\sum_{j=kK}^{+\infty}\Delta_j, \tag{40}$$

with $\Gamma_k \triangleq \sum_{j=kK}^{+\infty}\|\boldsymbol{x}_{j+1} - \boldsymbol{x}_j\|$. On the one hand,

$$\begin{aligned}\sum_{j=kK}^{+\infty}\sum_{i=j-K}^{j-1}\|\boldsymbol{x}_{i+1} - \boldsymbol{x}_i\| &= \sum_{j=kK}^{+\infty}\sum_{i=0}^{K-1}\|\boldsymbol{x}_{i+j-K+1} - \boldsymbol{x}_{i+j-K}\|,\\ &= \sum_{i=0}^{K-1}\sum_{j=kK}^{+\infty}\|\boldsymbol{x}_{i+j-K+1} - \boldsymbol{x}_{i+j-K}\| \leq K\Gamma_{k-1}.\end{aligned} \tag{41}$$

On the other hand, using $F(\boldsymbol{x}) - F(\boldsymbol{x}^*) \underset{k\to+\infty}{\longrightarrow} 0$ (31) and Lemma 1 yields

$$\begin{aligned}
[F(\boldsymbol{x}_{kK}) - F(\boldsymbol{x}^*)]^{1-\theta} \leq \kappa^{\frac{\theta-1}{\theta}} \|\nabla F(\boldsymbol{x}_{kK})\|^{\frac{1-\theta}{\theta}} &\leq \Lambda^{'} \left(\sum_{j=(k-1)K}^{kK-1} \|\boldsymbol{x}_{j+1} - \boldsymbol{x}_j\|^2 \right)^{\frac{1-\theta}{2\theta}}, \\
&\leq \Lambda^{'} \left(\sum_{j=(k-1)K}^{kK-1} \|\boldsymbol{x}_{j+1} - \boldsymbol{x}_j\| \right)^{\frac{1-\theta}{\theta}}, \\
&= \Lambda^{'} \left(\Gamma_{k-1} - \Gamma_k\right)^{\frac{1-\theta}{\theta}}, \qquad (42)
\end{aligned}$$

with $\Lambda^{'} \triangleq \kappa^{\frac{\theta-1}{\theta}} \left(2^K C \left(B^2 + \frac{\gamma_1^2}{\gamma_0^2}\nu^2\right)\right)^{\frac{1-\theta}{2\theta}}$. Plugging (41), (42) in (40), with $c = 2\sqrt{\Lambda}K$ and $\sum_{j=kK}^{+\infty} \Delta_j = [F(\boldsymbol{x}_{kK}) - F(\boldsymbol{x}^*)]^{1-\theta}$ gives

$$\Gamma_k \leq \frac{1}{2}\Gamma_{k-1} + \frac{1}{2}K\Lambda\Lambda^{'} \left(\Gamma_{k-1} - \Gamma_k\right)^{\frac{1-\theta}{\theta}}. \qquad (43)$$

By multiplying (43) by 2 and removing Γ_k from each side,

$$\Gamma_k \leq (\Gamma_{k-1} - \Gamma_k) + K\Lambda\Lambda^{'} \left(\Gamma_{k-1} - \Gamma_k\right)^{\frac{1-\theta}{\theta}}. \qquad (44)$$

Since $(\Gamma_k)_{k\in\mathbb{N}}$ is a positive decreasing sequence to zero, we can apply [24, Theorem 2].

- If $\theta \in]1/2, 1[$, there exists $\lambda > 0$ such that

$$(\forall k \geq k_0) \quad \Gamma_k \leq \lambda k^{-\frac{1-\theta}{2\theta-1}}. \qquad (45)$$

Denoting $q(k), r(k)$ the quotient and remainder of the Euclidean division of k by K, leads to

$$\|\boldsymbol{x}_k - \boldsymbol{x}^*\| = \|\boldsymbol{x}_{q(k)K+r(k)} - \boldsymbol{x}^*\| \leq \Gamma_{q(k)}. \qquad (46)$$

Combining (46) with (45) gives, for k large enough,

$$\|\boldsymbol{x}_k - \boldsymbol{x}^*\| \leq \lambda q(k)^{-\frac{1-\theta}{2\theta-1}} = \lambda \left(\frac{k-r(k)}{K}\right)^{-\frac{1-\theta}{2\theta-1}} \leq \lambda \left(\frac{k}{K} - 1\right)^{-\frac{1-\theta}{2\theta-1}}, \qquad (47)$$

with $\lambda \left(\frac{k}{K} - 1\right)^{-\frac{1-\theta}{2\theta-1}} \underset{k\to+\infty}{=} \mathcal{O}\left(k^{-\frac{1-\theta}{2\theta-1}}\right)$.

- If $\theta \in]0, 1/2]$, there exist $\mu > 0$ and $\delta \in]0, 1[$ such that

$$(\forall k \geq k_0) \quad \Gamma_k \leq \mu\, \delta^k. \qquad (48)$$

Similarly with the previous case, for k large enough,

$$\|\boldsymbol{x}_k - \boldsymbol{x}^*\| \leq \mu\, \delta^{q(k)} = \mu\, \delta^{\frac{k-r(k)}{K}} \leq \mu\, \delta^{\frac{k}{K}}. \qquad (49)$$

Conclusion is obtained by taking $\varepsilon = \delta^{\frac{1}{K}} \in]0, 1[$.

□

One can notice the dependency of the convergence rate with the KL exponent θ. This result is similar to the one in [15], though with a more direct proof following naturally from our global convergence Theorem 2. We also notice that the gradient Lipschitz property (see Assumption 2) entails that the convergence rate obtained for $(\|\boldsymbol{x}_k - \boldsymbol{x}^*\|)_{k\in\mathbb{N}}$ also holds for the sequence $(\|\nabla F(\boldsymbol{x}_k)\|)_{k\in\mathbb{N}}$.

5 Conclusion

This work introduces a block alternating MM subspace algorithm and provides its convergence analysis in the non-convex case. When combined with a memory gradient subspace (see, for e.g., [28]), the proposed method can be viewed as a convergent block alternating non linear conjugate gradient algorithm for non-convex large scale optimization. Future work will be dedicated to building a distributed implementation for B2MS, for instance by adopting an asynchronous approach [17] where block updates are spread among core machines with possible update delay.

Declarations

Funding: This research work received funding support from the European Research Council Starting Grant MAJORIS ERC-2019-STG-850925.
Competing interests: The authors have no relevant financial or non-financial interests to disclose.

References

1. K. Kurdyka, On gradients of functions definable in o-minimal structures, in: Annales de l'institut Fourier, Vol. 48, 1998, pp. 769–783.
2. J. Bolte, A. Daniilidis, O. Ley, L. Mazet, Characterizations of łojasiewicz inequalities and applications, arXiv preprint arXiv:0802.0826 (2008).
3. D. C. Liu, J. Nocedal, On the limited memory BFGS method for large scale optimization, Mathematical programming 45 (1) (1989) 503–528.
4. R. Fletcher, C. M. Reeves, Function minimization by conjugate gradients, The Computer Journal 7 (2) (1964) 149–154.
5. Y.-X. Yuan, Subspace methods for large scale nonlinear equations and nonlinear least squares, Optimization and Engineering 10 (2) (2008) 207–218.
6. A. Wald, T. Schuster, Sequential subspace optimization for nonlinear inverse problems, Journal of Inverse and Ill-posed Problems 25 (1) (2017) 99–117.
7. S. Bonettini, F. Porta, M. Prato, S. Rebegoldi, V. Ruggiero, L. Zanni, Recent Advances in Variable Metric First-Order Methods, Springer International Publishing, Cham, 2019, pp. 1–31.
8. P. Frankel, G. Garrigos, J. Peypouquet, Splitting methods with variable metric for Kurdyka–Łojasiewicz functions and general convergence rates, Journal of Optimization Theory and Applications 165 (3) (2015) 874–900.
9. J. Bolte, S. Sabach, M. Teboulle, Proximal alternating linearized minimization for nonconvex and nonsmooth problems, Mathematical Programming 146 (1-2) (2014) 459–494.
10. Y. Sun, P. Babu, D. P. Palomar, Majorization-minimization algorithms in signal processing, communications, and machine learning, IEEE Transactions on Signal Processing 65 (3) (Aug. 2016).
11. Z. Zhang, J. T. Kwok, D.-Y. Yeung, Surrogate maximization/minimization algorithms and extensions, Machine Learning 69 (2007) 1–33.
12. M. C. Robini, Y. Zhu, Generic half-quadratic optimization for image reconstruction, SIAM Journal on Imaging Sciences 8 (3) (2015) 1752–1797.
13. M. Allain, J. Idier, Y. Goussard, On global and local convergence of half-quadratic algorithms, IEEE Transactions on Image Processing 15 (5) (2006) 1130–1142.
14. E. Chouzenoux, J.-C. Pesquet, A. Repetti, Variable metric forward–backward algorithm for minimizing the sum of a differentiable function and a convex function, Journal of Optimization Theory and Applications 162 (1) (2014) 107–132.
15. E. Chouzenoux, J.-C. Pesquet, A. Repetti, A block coordinate variable metric forward–backward algorithm, Journal of Global Optimization 66 (3) (2016) 457–485.
16. M. Hong, M. Razaviyayn, Z.-Q. Luo, J.-S. Pang, A unified algorithmic framework for block-structured optimization involving big data: With applications in machine learning and signal processing, IEEE Signal Processing Magazine 33 (1) (2015) 57–77.
17. G. Scutari, Y. Sun, Parallel and Distributed Successive Convex Approximation Methods for Big-Data Optimization, Springer Verlag Series, C.I.M.E Lecture Notes in Mathematics, 2018.
18. M. W. Jacobson, J. A. Fessler, An expanded theoretical treatment of iteration-dependent majorize-minimize algorithms, IEEE Transactions on Image Processing 16 (10) (2007) 2411–2422.
19. S. Sotthivirat, J. A. Fessler, Image recovery using partitioned-separable paraboloidal surrogate coordinate ascent algorithms, IEEE Transactions on Signal Processing 11 (3) (2002) 306–317.
20. E. Chouzenoux, J. Idier, S. Moussaoui, A majorize–minimize strategy for subspace optimization applied to image restoration, IEEE Transactions on Image Processing 20 (6) (2010) 1517–1528.
21. E. Chouzenoux, A. Jezierska, J.-C. Pesquet, H. Talbot, A majorize-minimize subspace approach for ℓ_2-ℓ_0 image regularization, SIAM Journal on Imaging Sciences 6 (1) (2013) 563–591.
22. E. Chouzenoux, J.-C. Pesquet, Convergence rate analysis of the majorize–minimize subspace algorithm, IEEE Signal Processing Letters 23 (9) (2016) 1284–1288.
23. E. Chouzenoux, S. Martin, J. Pesquet, A local MM subspace method for solving constrained variational problems in image recovery, Tech. rep., http://www.optimization-online.org/DB_HTML/2021/09/8595.html (2021).
24. H. Attouch, J. Bolte, On the convergence of the proximal algorithm for nonsmooth functions involving analytic features, Mathematical Programming 116 (1) (2009) 5–16.
25. M. Razaviyayn, M. Hong, Z.-Q. Luo, A unified convergence analysis of block successive minimization methods for nonsmooth optimization, SIAM Journal on Optimization 23 (2) (2013) 1126–1153.
26. D. P. Bertsekas, Nonlinear programming, Journal of the Operational Research Society 48 (3) (1997) 334–334.
27. A. Miele, J. Cantrell, Study on a memory gradient method for the minimization of functions, Journal of Optimization Theory and Applications 3 (6) (1969) 459–470.
28. A. Florescu, E. Chouzenoux, J.-C. Pesquet, P. Ciuciu, S. Ciochina, A majorize-minimize memory gradient method for complex-valued inverse problems, Signal Processing 103 (2014) 285–295.
29. Relation between the memory gradient method and the Fletcher-Reeves method, Journal of Optimization Theory and Applications 4 (1) (1969) 67–71.
30. R. I. Boţ, E. R. Csetnek, An inertial tseng's type proximal algorithm for nonsmooth and nonconvex optimization problems, Journal of Optimization Theory and Applications 171 (2) (2016) 600–616.
31. D. Davis, The asynchronous PALM algorithm for nonsmooth nonconvex problems, arXiv preprint arXiv:1604.00526 (2016).
32. J. Nocedal, S. Wright, Numerical optimization, Springer Science & Business Media, 2006.