

A preconditioned Forward-Backward approach with application to large-scale nonconvex spectral unmixing problems

Audrey Repetti, Emilie Chouzenoux, Jean-Christophe Pesquet

▶ **To cite this version:**

Audrey Repetti, Emilie Chouzenoux, Jean-Christophe Pesquet. A preconditioned Forward-Backward approach with application to large-scale nonconvex spectral unmixing problems. 39th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2014)., May 2014, Florence, Italy. pp.1498 - 1502, 10.1109/ICASSP.2014.6853847 . hal-01077329

HAL Id: hal-01077329
https://hal.science/hal-01077329

Submitted on 24 Oct 2014

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

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

A PRECONDITIONED FORWARD-BACKWARD APPROACH WITH APPLICATION TO LARGE-SCALE NONCONVEX SPECTRAL UNMIXING PROBLEMS

Audrey Repetti, Emilie Chouzenoux, and Jean-Christophe Pesquet

Université Paris-Est, Lab. d'Informatique Gaspard Monge, UMR CNRS 8049,
Champs-sur-Marne, 77454 Marne-la-Vallée, France
{first.last}@univ-paris-est.fr

ABSTRACT

Many inverse problems require to minimize a criterion being the sum of a non necessarily smooth function and a Lipschitz differentiable function. Such an optimization problem can be solved with the Forward-Backward algorithm which can be accelerated thanks to the use of variable metrics derived from the Majorize-Minimize principle. The convergence of this approach is guaranteed provided that the criterion satisfies some additional technical conditions. Combining this method with an alternating minimization strategy will be shown to allow us to address a broad class of optimization problems involving large-size signals. An application example to a nonconvex spectral unmixing problem will be presented.

***Index Terms*—** Block coordinate algorithm, Forward-Backward algorithm, Nonconvex optimization, Nonsmooth optimization, Large-scale problems.

1. INTRODUCTION

In the context of inverse problems, an estimation of the object of interest can be efficiently obtained by solving the following optimization problem :

$$\underset{x\in\mathbb{R}^N}{\text{minimize}} \ (G(x) = F(x) + R(x)), \tag{1}$$

where F is a data-fidelity term, and R is a regularization function. In this paper, we focus on the case when $F\colon \mathbb{R}^N \to \mathbb{R}$ is a differentiable function and $R\colon \mathbb{R}^N \to (-\infty, +\infty]$ is a proper lower semicontinuous function, which may be nonsmooth. For example, the latter function may model a hard constraint on the target signal or it may be a sparsity promoting measure. A standard approach in this context consists of using the *Forward-Backward* (FB) algorithm [1–4], which alternates a gradient step on F and a proximal step on R.

In the case of large scale inverse problems such as those encountered in image restoration, one major concern is to find an optimization algorithm able to deliver reliable numerical solutions in a reasonable time. As many first-order minimization methods, the FB algorithm may suffer from slow convergence [1]. A possible way to accelerate its convergence is to modify the underlying metric at each iteration thanks to a preconditioning matrix, giving rise to the so-called *Variable Metric Forward-Backward* (VMFB) algorithm [1, 5–12].

When the regularization function has a block separable structure, another efficient strategy for solving (1) is to alternate between blocks. The *Block Coordinate Descent* (BCD) algorithm is obtained when, at each iteration, G is minimized exactly within the current block [13]. As pointed out in [13], the BCD algorithm is not guaranteed to converge under the sole assumption that the criterion is convex with respect to each block. In the proximal version of the BCD algorithm [14], this limitation is overcome. The convergence of its iterates has been established in [15] for a non necessarily convex criterion, and further generalized in [4] to the case of a variable metric. For more flexibility, at each iteration, it is often easier to replace the proximal step by a FB step which leads to the so-called *Block Coordinate Forward-Backward* (BC-FB) algorithm [16–20]. Recently, the convergence to a critical point of (1) has been established in [20] when F and R are not necessarily convex and when the blocks are updated following a *cyclic rule*.

In this paper we present a *Block Coordinate Variable Metric Forward-Backward* (BC-VMFB) algorithm, which combines the VMFB algorithm [11] with the alternating minimization approach of BC algorithms. Up to the best of our knowledge, the convergence of BC-VMFB has only been investigated for convex functions F and R, and a *random rule* on the blocks [21, 22]. Our contribution is to generalize the BC-VMFB algorithm to the case when the criterion is non necessarily convex in the context of 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).

The rest of the paper is organized as follows : Section 2 introduces the problem formulation. Section 3 describes the proposed BC-VMFB algorithm and investigates its convergence properties. A discussion of the algorithm performance by means of experiments concerning a large-size hyperspectral unmixing problem is provided in Section 4. Finally, some conclusions are drawn in Section 5.

This work was supported by the MASTODONS project by CNRS (grant 2013MesureHD)

2. PROBLEM FORMULATION

2.1. Problem formulation

We consider Problem (1) where G is coercive (i.e. $\lim_{\|x\|\to+\infty} G(x) = +\infty$) and F is differentiable with an L-Lipschitzian gradient ($L > 0$) on the domain $\mathrm{dom}\, R$ of function R, i.e.

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

where $\|\cdot\|$ denotes the standard Euclidean norm of $\mathbb{R}^N$, and ∇F is the gradient of F. Moreover, we assume a block separable structure on R. More precisely, let us define an integer $J \geq 2$ and some positive integers $N_1, \ldots, N_J$ such that $\sum_{j=1}^{J} N_j = N$. Any element x of the product space $\mathbb{R}^{N_1} \times \ldots \times \mathbb{R}^{N_J}$ is denoted hereafter by $x = \big(x^{(j)}\big)_{1\leq j\leq J}$, where, for every $j \in \{1,\ldots,J\}$, $x^{(j)} \in \mathbb{R}^{N_j}$. Then, we assume that R reads

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

where, for every $j \in \{1,\ldots,J\}$, $R_j \colon \mathbb{R}^{N_j} \to (-\infty,+\infty]$ is proper, lower semicontinuous, bounded from below by an affine function, and its restriction to its domain is continuous. Moreover, for every $j \in \{1,\ldots,J\}$, we denote by $\overline{j}$ the complementary set of j on $\{1,\ldots,J\}$, i.e. $\overline{j} = \{1,\ldots,J\} \setminus \{j\}$.

2.2. Optimization tools

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

$$(\forall x \in \mathbb{R}^M) \quad \|x\|_U = \langle x, Ux\rangle^{1/2}, \tag{3}$$

where $U \in \mathbb{R}^{M\times M}$ is some symmetric positive definite (SPD) matrix, and $\langle\cdot,\cdot\rangle$ denotes the usual scalar product. The proximity operator ([23, Sec. XV.4], [24] and [4]) is defined as follows :

Definition 2.1. *Let $\psi \colon \mathbb{R}^M \to (-\infty,+\infty]$ be a proper, lower semicontinuous function, let $U \in \mathbb{R}^{M\times M}$ be a SPD matrix, and let $x \in \mathbb{R}^M$. The proximity operator of ψ at x relative to the metric induced by U is given by* $\mathrm{prox}_{U,\psi}(x) = \underset{y\in\mathbb{R}^M}{\mathrm{Argmin}}\ \psi(y) + \frac{1}{2}\|y-x\|_U^2$.

The following definition is useful to deal with a nonconvex cost function [3, 4, 11, 15, 20]. As emphasized in [3], it is satisfied for a wide class of functions, such as, in particular, real analytic and semi-algebraic functions.

Definition 2.2. *Let $\psi \colon \mathbb{R}^M \to (-\infty,+\infty]$ be a proper, lower semicontinuous function. ψ satisfies the* Kurdyka-Łojasiewicz *(KL) inequality iff, for every $\xi \in \mathbb{R}$, and, for every bounded subset E of $\mathbb{R}^M$, there exist three constants $\kappa \in (0,+\infty)$, $\zeta \in (0,+\infty)$ and $\theta \in [0,1)$ such that $\|t\| \geq \kappa|\psi(x)-\xi|^\theta$, for every $t \in \partial\psi(x)$, and for every $x \in E$ such that $|\psi(x)-\xi| \leq \zeta$ (with the convention $0^0 = 0$).*

3. PROPOSED METHOD

3.1. BC-VMFB algorithm

An efficient approach for solving the minimization problem (1)-(2) consists of using the BC-VMFB algorithm :

Algorithm 1 BC-VMFB algorithm.

For every $\ell \in \mathbb{N}$, let $\gamma_\ell \in (0,+\infty)$.
Let $x_0 \in \mathrm{dom}\, R$.
Iterations :
For $\ell = 0, 1, \ldots$
- Let $j_\ell \in \{1,\ldots,J\}$.
- Let $A_{j_\ell}(x_\ell) \in \mathbb{R}^{N_{j_\ell}\times N_{j_\ell}}$ be a SPD matrix.
- $\tilde{x}_\ell^{(j_\ell)} = x_\ell^{(j_\ell)} - \gamma_\ell A_{j_\ell}(x_\ell)^{-1}\nabla_{j_\ell} F(x_\ell)$,
- $x_{\ell+1}^{(j_\ell)} \in \mathrm{prox}_{\gamma_\ell^{-1}A_{j_\ell}(x_\ell),R_{j_\ell}}\left(\tilde{x}_\ell^{(j_\ell)}\right)$,
- $x_{\ell+1}^{(\overline{j}_\ell)} = x_\ell^{(\overline{j}_\ell)}$.

In the above algorithm, for every $\ell \in \mathbb{N}$, $\nabla_{j_\ell} F(x_\ell) \in \mathbb{R}^{N_{j_\ell}}$ is the partial gradient of F with respect to $x^{(j_\ell)}$ computed at x_ℓ. In other words, at each iteration ℓ, block j_ℓ is activated and a VMFB step is performed on the associated variables, while all the other blocks with indices in $\overline{j}_\ell$ are kept unchanged.

In the particular case when a single block is used ($J = 1$), Algorithm 1 is equivalent to the VMFB algorithm [11]. In addition, if $A_1(x_\ell) \equiv \mathrm{I}_N$, where I_N denotes the identity matrix of $\mathbb{R}^N$, then the usual FB algorithm is recovered.

In general, the proximity operator relative to an arbitrary metric does not have a closed form expression. Although it is not detailed in this paper, it is worth mentioning that an inexact version of Algorithm 1 can be derived.

3.2. Assumptions

Matrices $(A_{j_\ell}(x_\ell))_{\ell\in\mathbb{N}}$ serving to define some appropriate variable metric will play a central role in the convergence analysis of Algorithm 1. More specifically, these matrices fulfill the following so-called *majorization condition* :

Assumption 3.1.
Let $\ell \in \mathbb{N}$ and let $j_\ell \in \{1,\ldots,J\}$. The quadratic function defined as : For every $x^{(j_\ell)} \in \mathbb{R}^{N_{j_\ell}}$,

$$Q_{j_\ell}(x^{(j_\ell)} \mid x_\ell) = F(x_\ell) + \left\langle x^{(j_\ell)} - x_\ell^{(j_\ell)}, \nabla_{j_\ell} F(x_\ell)\right\rangle + \frac{1}{2}\|x^{(j_\ell)} - x_\ell^{(j_\ell)}\|^2_{A_{j_\ell}(x_\ell)},$$

is a majorant function *of the restriction of F to its j_ℓ-th block on $\mathrm{dom}\, R_{j_\ell}$, i.e., for every $x^{(j_\ell)} \in \mathrm{dom}\, R_{j_\ell}$,*

$$F(x_\ell^{(1)},\ldots,x_\ell^{(j_\ell-1)},x^{(j_\ell)},x_\ell^{(j_\ell+1)},\ldots,x_\ell^{(J)}) \leq Q_{j_\ell}(x^{(j_\ell)} \mid x_\ell).$$

Moreover, the eigenvalues of $A_{j_\ell}(x_\ell)$ are lower and upper bounded by positive values.

Let us point out that Assumption 3.1 is not restrictive since it is satisfied when, for every $\ell \in \mathbb{N}$, $A_{j_\ell}(x_\ell) = L\,\mathrm{I}_{N_{j_\ell}}$. However, in order to obtain good numerical performance, the matrices $(A_{j_\ell}(x_\ell))_{\ell\in\mathbb{N}}$ should be simple to compute and built in such a way that, for every $\ell \in \mathbb{N}$, the quadratic function $Q_{j_\ell}(\cdot \mid x_\ell)$ is as close as possible to the restriction of F to its j_ℓ-th block. Some useful techniques for constructing such efficient metrics are presented in [25] for some subclasses of functions F.

In order to ensure that each block is updated an infinite number of times, we also make the following assumption :

Assumption 3.2.
Blocks $(j_\ell)_{\ell\in\mathbb{N}}$ are updated following an essentially cyclic rule, i.e., there exists $K \geq J$ such that, for every $\ell \in \mathbb{N}$, $\{1,\ldots,J\} \subset \{j_\ell,\ldots,j_{\ell+K-1}\}$.

Note that the usual cyclic rule defined, for every $\ell \in \mathbb{N}$, by $j_\ell - 1 = \ell \mod (J)$, satisfies Assumption 3.2 with $K = J$. Finally, we suppose that, for every $\ell \in \mathbb{N}$, the stepsize γ_ℓ is a positive real satisfying the following assumption :

Assumption 3.3.
One of the following statements holds :

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

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

3.3. Convergence results

The proposed metric construction leads to the following descent property, describing the behaviour of $(G(x_\ell))_{\ell\in\mathbb{N}}$, where $(x_\ell)_{\ell\in\mathbb{N}}$ is a sequence of iterates generated by Algorithm 1 :

Proposition 3.1. [26] *Under Assumptions 3.1-3.3, there exists $\mu \in (0,+\infty)$ such that, for every $\ell \in \mathbb{N}$,*

$$G(x_{\ell+K}) \leq G(x_\ell) - \frac{\mu}{2}\|\chi_\ell\|^2, \tag{4}$$

where $\|\chi_\ell\|^2 = \sum_{k=0}^{K-1} \|x_{\ell+k+1} - x_{\ell+k}\|^2$, and K is the constant given in Assumption 3.2.

Our main result concerning the asymptotic behaviour of Algorithm 1 is provided by the following theorem :

Theorem 3.1. [26] *Assume that Assumptions 3.1-3.3 hold and that G satisfies the KL inequality. Then the sequence $(x_\ell)_{\ell\in\mathbb{N}}$ converges to a critical point $\widehat{x}$ of G. Moreover, $\big(G(x_\ell)\big)_{\ell\in\mathbb{N}}$ is a nonincreasing sequence converging to $G(\widehat{x})$.*

4. APPLICATION TO HYPERSPECTRAL UNMIXING

4.1. Problem formulation

In this section, we apply the proposed algorithm to the problem of hyperspectral data unmixing. Let us consider a hyperspectral data set $Y \in \mathbb{R}^{S\times M}$, modeling a set of images in $\mathbb{R}^M$ (columnwise reshaped), acquired in S different spectral bands. We assume the following linear mixing model :

$$Y = \overline{U}\,\overline{V} + E, \tag{5}$$

where the columns of $\overline{U} \in \mathbb{R}^{S\times P}$ represent spectra of P distinct components (endmembers) available in the image, $\overline{V} \in \mathbb{R}^{P\times M}$ their respective proportions (abundances) at each pixel, and $E \in \mathbb{R}^{S\times M}$ is the measurement noise. The problem of unmixing consists of estimating $\overline{U}$ and $\overline{V}$ from the observation Y. Although standard unmixing approaches are supervised, in the sense that the endmember spectra composing $\overline{U}$ are assumed to be part of an available spectral library or provided by an endmember extraction algorithm [27–30], there is an increasing interest in joint estimation methods based on nonnegative matrix factorization (NMF) [30]. In the standard NMF approach, $\overline{U}$ and $\overline{V}$ are estimated thanks to the minimization of a least squares objective function, under positivity constraints on the elements of both matrices [31]. As pointed out in [30, 32], a significant improvement of the quality of the results is obtained by incorporating some *a priori* knowledge on the sought matrices. Here, we focus on the case when each endmember spectra is a weighted combination of few components of a given large dictionary of size $Q > P$, here modeled by a matrix $\Omega \in \mathbb{R}^{S\times Q}$. Thus, the observation model is re-expressed as

$$Y = \Omega\,\overline{T}\,\overline{V} + E, \tag{6}$$

where $\overline{T} \in \mathbb{R}^{Q\times P}$ is the matrix of components of the endmember spectra assumed to be sparse.

4.2. Proposed algorithm

Estimates $\widehat{T}$ and $\widehat{V}$ of $\overline{T}$ and $\overline{V}$ result from

$$\underset{\substack{T\in\mathbb{R}^{Q\times P}\\ V\in\mathbb{R}^{P\times M}}}{\text{minimize}}\ (G(T,V) = F(T,V) + R_1(T) + R_2(V)), \tag{7}$$

where $F(T,V) = \frac{1}{2}\|Y - \Omega TV\|_F^2$, $\|\cdot\|_F$ denotes the Frobenius norm, and R_1 and R_2 are proper, lower semicontinuous regularization functions on T and V. Problem (7) takes the form of our general problem (1). In order to apply Algorithm 1, we need to define, for a given $(T',V') \in \operatorname{dom} R_1 \times \operatorname{dom} R_2$, quadratic majorants of $F(\cdot,V')$ (resp. $F(T',\cdot)$). Under the assumption that matrices T' and V' have positive elements, we can derive the following majorant functions, which

are reminiscent of the auxiliary functions proposed in [33] in the context of NMF :

$$Q_1(T \mid T', V') = F(T', V') + \mathrm{tr}\left((T - T')\nabla_1 F(T', V')^\top\right) + \frac{1}{2}\mathrm{tr}\left(((T - T') \odot A_1(T', V'))(T - T')^\top\right),$$

$$Q_2(V \mid T', V') = F(T', V') + \mathrm{tr}\left((V - V')\nabla_2 F(T', V')^\top\right) + \frac{1}{2}\mathrm{tr}\left(((V - V') \odot A_2(T', V'))(V - V')^\top\right),$$

with

$$\begin{aligned} A_1(T', V') &= ((\Omega^\top \Omega)T'(V'V'^\top)) \oslash T', \\ A_2(T', V') &= ((\Omega T')^\top \Omega T' V') \oslash V', \end{aligned} \tag{8}$$

where $\mathrm{tr}(\cdot)$ denotes the trace operator, and $\odot$ (resp. $\oslash$) the Hadamard product (resp. division) between matrices of the same size. Problem (7) is then solved by applying Algorithm 1 with $J = 2$ and $N = P(M + Q)$.

4.3. Numerical results

In order to simulate realistic hyperspectral data, we define Ω as the pruned version available at `http://www.lx.it.pt/~bioucas`, of the U.S. Geological Survey library [34] composed with $Q = 62$ spectra of $S = 224$ spectral bands from 383 nm to 2508 nm. Matrix $\overline{U}$ is composed with $P = 5$ distinct spectra resulting from weighted combinations of few (typically 3) pure spectra randomly selected in the columns of Ω, the weights being stored in matrix $\overline{T}$. Each line of the abundance matrix $\overline{V}$ is then simulated as the superposition of 2D Gaussian patterns with size $M = 128^2$ pixels, with random location and variance, normalized to ensure the sum-to-one constraint. Finally, the resulting mixture is corrupted with a zero-mean white Gaussian noise, whose variance is set in order to get a signal-to-noise ratio (SNR) of 20 dB.

In order to promote the sparsity of T, we choose :

$$R_1(T) = \sum_{q=1}^{Q}\sum_{p=1}^{P}\left(\iota_{[T_{\min}, T_{\max}]}(T^{(q,p)}) + \eta\,\varphi_\beta(T^{(q,p)})\right),$$

where $(\eta, T_{\min}, T_{\max}) \in (0, +\infty)^3$, $\iota_{[T_{\min}, T_{\max}]}$ stands for the indicator function of interval $[T_{\min}, T_{\max}]$ and φ_β denotes the regularization function defined in [35], with parameter $\beta \in (0, 1]$. We recall that φ_β is convex if and only if $\beta = 1$, and that φ_1 reduces to the absolute value function. Moreover, we define R_2 as the indicator function of the set $\mathcal{V} \subset \mathbb{R}^{P \times M}$ given by

$$\mathcal{V} = \Big\{V \in \mathbb{R}^{P\times M} : (\forall m \in \{1, ..., M\})\ \textstyle\sum_{p=1}^{P} V^{(p,m)} = 1,$$
$$(\forall p \in \{1, ..., P\})(\forall m \in \{1, ..., M\}) \quad V^{(p,m)} \geq V_{\min}\Big\},$$

with $V_{\min} > 0$.

In practice, $\beta = 0.1$ has been observed to yield the best reconstruction performance. Fig. 1 shows the exact and reconstructed spectra $\overline{U}$ and $\widehat{U}$. In this example, the average

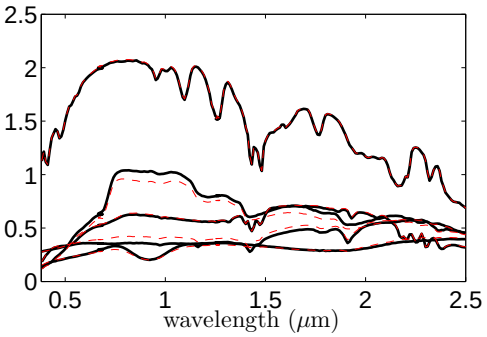


Fig. 1. Exact (continuous) and estimated endmembers (dashed).

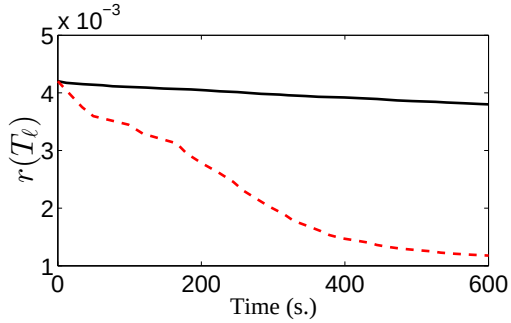


Fig. 2. Comparison of BC-FBVM algorithm (dashed) and PALM algorithm (continuous).

residual norm $r(\widehat{T})$ between $\widehat{T}$ and $\overline{T}$ (resp. $r(\widehat{V})$ between $\widehat{V}$ and $\overline{V}$) equals 1.1×10^{-3} (resp. 7.5×10^{-5}). For comparison, the NMF approach from [33] applied to Model (5) leads to $r(\widehat{V}) = 5 \times 10^{-4}$ (after renormalization of the columns of $\widehat{V}$).

Fig. 2 illustrates the variations of $(r(T_\ell))_\ell$ with respect to the computation time, when using either the proposed BC-VMFB algorithm with $\gamma_\ell \equiv 0.99$ or PALM algorithm [20], the latter being similar to Algorithm 1 where the preconditioning matrices (8) have been replaced by

$$(\forall \ell \in \mathbb{N})\ A_{j_\ell}(T_\ell, V_\ell) = \begin{cases} \|\Omega\|^2 \|V_\ell\|^2 \mathbf{1}, & \text{if } j_\ell = 1, \\ \|\Omega T_\ell\|^2 \mathbf{1}, & \text{if } j_\ell = 2, \end{cases} \tag{9}$$

and $\mathbf{1}$ is a constant matrix with entries equal to one. We can observe that the variable metric strategy leads to a significant acceleration in terms of decay of the residual.

5. CONCLUSION

The BC-VMFB introduced in this paper allows us to choose an iteration dependent metric based on Majorization-Minimization properties. Its convergence proof is an offspring of recent results in nonsmooth analysis. The application of this algorithm to hyperspectral unmixing demonstrates its ability to deal with sophisticated variational formulations, while exhibiting a good convergence rate.

6. REFERENCES

[1] G. H.-G. Chen and R. T. Rockafellar, "Convergence rates in forward-backward splitting," *SIAM J. Optim.*, vol. 7, pp. 421–444, 1997.

[2] P. Tseng, "A modified forward-backward splitting method for maximal monotone mappings," *SIAM J. Control Optim.*, vol. 38, no. 2, pp. 431–446, 1998.

[3] H. Attouch and J. Bolte, "On the convergence of the proximal algorithm for nonsmooth functions involving analytic features," *Math. Program.*, vol. 116, pp. 5–16, Jun. 2009.

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

[5] J. F. Bonnans, J. C. Gilbert, C. Lemaréchal, and C. A. Sagastizábal, "A family of variable metric proximal methods," *Math. Program.*, vol. 68, pp. 15–47, 1995.

[6] J. V. Burke and M. Qian, "A variable metric proximal point algorithm for monotone operators," *SIAM J. Control Optim.*, vol. 37, pp. 353–375, 1999.

[7] L. A. Parente, P. A. Lotito, and M. V. Solodov, "A class of inexact variable metric proximal point algorithms," *SIAM J. Optim.*, vol. 19, pp. 240–260, 2008.

[8] P. A. Lotito, L. A. Parente, and M. V. Solodov, "A class of variable metric decomposition methods for monotone variational inclusions," *J. Convex Anal.*, vol. 16, pp. 857–880, 2009.

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

[10] S. Becker and J. Fadili, "A quasi-Newton proximal splitting method," in *Advances in Neural Information Processing Systems (NIPS)*, P. Bartlett, F.C.N. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, Eds., 2012, vol. 25, pp. 2618–2626.

[11] E. Chouzenoux, J.-C. Pesquet, and A. Repetti, "Variable metric forward-backward algorithm for minimizing the sum of a differentiable function and a convex function," *J. Optim. Theory App.*, to appear, 2013, http ://www.optimization-online.org/DB_HTML/2013/01/3749.html.

[12] Q. Tran-Dinh, A. Kyrillidis, and V. Cevher, "Composite self-concordant minimization.," Tech. Rep., 2013, http ://arxiv.org/pdf/1308.2867v1.pdf.

[13] P. Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," *J. Optim. Theory Appl.*, vol. 109, no. 3, pp. 475–494, Jun. 2001.

[14] A. Auslender, "Asymptotic properties of the Fenchel dual functional and applications to decomposition problems," *J. Optim. Theory App.*, vol. 73, no. 3, pp. 427–449, Jun. 1992.

[15] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, "Proximal alternating minimization and projection methods for nonconvex problems. An approach based on the Kurdyka-Łojasiewicz inequality," *Math. Oper. Res.*, vol. 35, no. 2, pp. 438–457, 2010.

[16] Z. Q. Luo and P. Tseng, "On the linear convergence of descent methods for convex essentially smooth minimization," *SIAM J. Control Optim.*, vol. 30, no. 2, pp. 408–425, 1992.

[17] Z. Q. Luo and P. Tseng, "On the convergence of the coordinate descent method for convex differentiable minimization," *J. Optim. Theory App.*, vol. 72, no. 1, pp. 7–35, 1992.

[18] Y. Xu and W. Yin, "A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion," *SIAM J. Imaging Sci.*, vol. 6, no. 3, pp. 1758–1789, 2013.

[19] J. Bolte, P. L. Combettes, and J.-C. Pesquet, "Alternating proximal algorithm for blind image recovery," in *17th IEEE Int. Conf. on Image Processing (ICIP 2010)*, Hong-Kong, China, 26-29 Sep. 2010, pp. 1673–1676.

[20] J. Bolte, S. Sabach, and M. Teboulle, "Proximal alternating linearized minimization fon nonconvex and nonsmooth problems," *Math. Program., Ser. A*, Jul. 2013.

[21] P. Richtárik and M. Talác, "Iteration complexity of randomized block-coordinate descent methods for minimizing a composite function," *Math. Program.*, vol. 144, no. 1-2, pp. 1–38, Apr. 2014.

[22] R. Tappenden, P. Richtárik, and J. Gondzio, "Inexact coordinate descent : Complexity and preconditioning," Tech. Rep., 2013, http ://arxiv.org/abs/1304.5530.

[23] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms*, Springer-Verlag, New York, 1993.

[24] P. L. Combettes and B. C. Vũ, "Variable metric quasi-Fejér monotonicity," *Nonlinear Anal.*, vol. 78, pp. 17–31, 2013.

[25] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *Amer. Stat.*, vol. 58, no. 1, pp. 30–37, Feb. 2004.

[26] E. Chouzenoux, J.-C. Pesquet, and A. Repetti, "A block coordinate variable metric forward-backward algorithm," Tech. Rep., 2013, http ://www.optimization-online.org/DB HTML/2013/12/4178.html.

[27] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, D. Qian, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview : Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 5, no. 2, pp. 354–379, April 2012.

[28] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, January 2002.

[29] E. Chouzenoux, M. Legendre, S. Moussaoui, and J. Idier, "Fast constrained least squares spectral unmixing using primal-dual interior-point optimization," *IEEE J. Sel. Topics Appl. Remote Sens.*, vol. 7, no. 1, pp. 59–69, 2014.

[30] W.-K. Ma, J. M. Bioucas-Dias, C. Tsung-Han, N. Gillis, P. Gader, A. Plaza, A. Ambikapathi, and C. Chong-Yung, "A signal processing perspective on hyperspectral unmixing : Insights from remote sensing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 67–81, Jan. 2014.

[31] P. Paatero and U. Tapper, "Positive matrix factorization : a nonnegative factor model with optimal utilization of error estimates of data values," *Environmetrics*, vol. 5, pp. 111–126, 1994.

[32] S. Jia and Y. Qian, "Constrained nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 161–173, 2009.

[33] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in Neural and Information Processing Systems*, 2001, vol. 13, pp. 556–562.

[34] R. N. Clark, G. A. Swayze, A. Gallagher, T. V. King, and W. M. Calvin, "The U.S. geological survey digital spectral library : version 1 : 0.2 to 3.0 μm," *U.S. Geological Survey, Denver, CO, Open File Rep. 93-592*, 1993.

[35] R. Chartrand, "Nonconvex splitting for regularized low-rank + sparse decomposition," *IEEE Trans. Signal Process.*, vol. 60, pp. 5810–5819, 2012.