

Discrete Mumford-Shah on graph for mixing matrix estimation

Yacouba Kaloga, Marion Foare, Nelly Pustelnik, Pablo Jensen

▶ To cite this version:

Yacouba Kaloga, Marion Foare, Nelly Pustelnik, Pablo Jensen. Discrete Mumford-Shah on graph for mixing matrix estimation. 2018. hal-01948570v1

HAL Id: hal-01948570 https://hal.science/hal-01948570v1

Preprint submitted on 7 Dec 2018 (v1), last revised 27 May 2019 (v2)

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.

Discrete Mumford-Shah on graph for mixing matrix estimation

Yacouba Kaloga, Marion Foare, Member, IEEE, Nelly Pustelnik, Member, IEEE, Pablo Jensen

Abstract—The discrete Mumford-Shah formalism has been introduced for the image denoising problem, allowing to capture both smooth behavior inside an object and sharp transitions on the boundary. In the present work, we propose first to extend this formalism to graphs and to the problem of mixing matrix estimation. New algorithmic schemes with convergence guarantees relying on proximal alternating minimization strategies are derived and their efficiency (good estimation and robustness to initialization) are evaluated on simulated data, in the context of vote transfer matrix estimation.

Index Terms—Mumford-Shah, graph, mixing matrice estimation, nonconvex optimisation

I. INTRODUCTION

Mixing matrix estimation is mainly encountered in the context of blind source separation where the problem can be described as, for every sample $n, z_n = Ms_n$, where $z_n =$ $(z_{n,1},\ldots,z_{n,P}) \in \mathbb{R}^P$ are the observed signals and $s_n = (s_{n,1},\ldots,s_{n,Q}) \in \mathbb{R}^Q$ are the unknown original source signals. M denotes the unknown mixture matrix with full row rank. The main objective of blind source separation is to estimate the mixing matrix M and the sources $(s_n)_n$. A vast literature is dedicated to this subject going from ICA methods used in the context of overdetermined case (i.e. $Q \leq P$) [1] to sparse component analysis for the underdetermined context [2]. In this work, we consider a slightly different problem, for which we assume the sources to be known, but that the mixing matrices differ for each n. To formulate our problem in a general setting, we propose to write it on a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where \mathcal{E} denotes the edges and \mathcal{V} the vertices. Our goal is to estimate, for each node $n \in \mathcal{V}$, a mixing matrix $M_n \in \mathbb{R}^{P imes Q}$ from the observations $\boldsymbol{z} = (z_n)_{n \in \mathcal{V}}$ and the sources $\boldsymbol{s} = (s_n)_{n \in \mathcal{V}}$ such that

$$(\forall n \in \mathcal{V}) \quad z_n = \boldsymbol{M}_n s_n.$$

A specific application of this problem, when \mathcal{G} models a regular grid, is encountered in hyperspectral unmixing imagery [3], for which recent contributions in the convex setting, involving total variation, have been proposed (see [4] and references therein).

Copyright (c) 2018 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

This work is supported by CNRS Defi InFIniTI 2018 Projet OptMatGeo and ANR GRAPHSIP.

M. Foare (Corresponding Author) is with Univ. Lyon, CPE Lyon, F-69616, Villeurbanne, France and with Univ. Lyon, Inria, CNRS, ENS de Lyon, UCB Lyon 1, LIP UMR 5668, F-69342, Lyon, France. E-mail: marion.foare@ens-lyon.fr

Y. Kaloga, N. Pustelnik and P. Jensen are with Univ Lyon, ENS de Lyon, Univ Claude Bernard, CNRS, Laboratoire de Physique, F-69342 Lyon, France. This framework also fits the challenging question of vote transfer matrix estimation at each polling location for which no stateof-the-art method is able to provide such a local estimate, this problem being especially ill-posed [5], [6]).

The contributions of this paper are: (i) to encompass the mixing matrix estimation problem into the Mumford-Shah formalism to be able to estimate M_n with limited assumptions: two matrices M_n and $M_{n'}$ should be similar if n and n' are neighbor nodes in the graph, except at some locations where sharp behaviors can be captured (Section II); (ii) a new proximal alternating minimization scheme with convergence guarantees as well as a block version of it, to handle large datasets (Section III); (iii) experimental validation of the proposed numerical schemes and robustness to the initialization (Section IV).

II. DISCRETE MUMFORD-SHAH ON GRAPH

A. Generalities on Mumford-Shah

Proposed in the late 80's, the Mumford-Shah (MS) model [7] is one of the most studied models in image processing, since it allows both to recover a piecewise smooth approximation of an input image corrupted with white Gaussian noise, and to extract the contours at the same time. Formally, let $\Omega \subset \mathbb{R}^2$ be a bounded, open set and $z: \Omega \to [0, 1]$ a corrupted input image, the Mumford-Shah model reads [8]:

$$\underset{u,\Gamma}{\text{minimize}} \ \frac{1}{2} \int_{\Omega} (u-z)^2 dx + \beta \int_{\Omega \setminus \Gamma} |\nabla u|^2 dx + \lambda \mathcal{H}^1(\Gamma \cap \Omega)$$
(1)

The first term is the data-term, ensuring similarities between the noisy image and the estimate \hat{u} , the second term favors solutions having a smooth behavior everywhere except on contours Γ and the third term controls the length of Γ by means of the 1D Hausdorff measure. The parameters $\beta > 0$ and $\lambda > 0$ allows us to adjust the contribution of each term in the estimation.

The nonconvexity of the MS functional leads to numerical difficulties, and most of the state-of-the-art methods work with piecewise constant approximations: e.g. the famous ROF model [9], which involves the Total Variation (TV) of the image or the Chan-Vese formulation [10]. However, none of these methods perform simultaneously the contour detection and the estimation of the denoised \hat{u} . More recently, we can refer to [8], [11], [12] for alternatives relying on the estimation of both \hat{u} and $\hat{\Gamma}$. Cai and Steidl [11] update iteratively the threshold from the ROF solution, Strekalovskiy et al. [12] provide an heuristic algorithm to solve a nonconvex formulation while in [8], a new discrete MS formulation allows us to fit proximal alternating minimization having convergence guarantees to a critical point. For this reason, we focus on this last strategy that estimate both \hat{u} and $\hat{\Gamma}$ simultaneously and that has theoretical guarantees.

B. Extension to graphs and mixing matrix estimation

In this work, we extend the discrete MS model proposed in [8] for image denoising to mixing matrix estimation on graph. We propose to formulate the data term as

$$ig(orall oldsymbol{M} \in \mathbb{R}^{P imes Q imes |\mathcal{V}|}ig) \quad \mathcal{L}(oldsymbol{M}) = \sum_{n \in \mathcal{V}} \|z_n - oldsymbol{M}_n s_n\|_2^2.$$

The second term aims to promote small variations between two adjacent matrices M_n and $M_{n'}$:

$$\mathcal{S}(\boldsymbol{M}, \boldsymbol{e}) = \beta \sum_{n \in \mathcal{V}} \sum_{n' \in \mathcal{N}(n)} (1 - e_{n,n'})^2 \|\boldsymbol{M}_n - \boldsymbol{M}_{n'}\|_2^2$$

where $\mathcal{N}(n)$ models the neighborhood of n and where $e_{n,n'} \in \mathbb{R}$ denotes the value of the edge between the nodes n and n'. This penalization tends to favor $e_{n,n'} \equiv 1$. In order to control the length of the contour, another penalization term favoring sparsity is required, e.g. a ℓ_1 -norm or a quadratic- ℓ_1 penalization [13]. This penalization over e, controlled with a parameter $\lambda > 0$, will be denoted by \mathcal{R} , and ensures $e_{n,n'} \in [0, 1]$. Another difference between image denoising and mixing matrix estimation on graph is the necessity to integrate additional hard constraints over M, e.g. a simplex constraint on the columns. The resulting D-MS model (1) reads

$$\min_{\boldsymbol{M}\in\mathbb{R}^{P\times Q\times|\mathcal{V}|}, e\in\mathbb{R}^{|\mathcal{E}|}} \Psi(\boldsymbol{M}, e) := \mathcal{L}(\boldsymbol{M}) + \mathcal{S}(\boldsymbol{M}, e) + \mathcal{R}(e) + \iota_C(\boldsymbol{M})$$
(2)

where $C \subset \mathbb{R}^{P \times Q \times |\mathcal{V}|}$ and ι_C denotes an indicator function ¹ separable in n, i.e. $\iota_C(\mathbf{M}) = \sum_{n=1}^{|\mathcal{V}|} \iota_{C_n}(\mathbf{M}_n)$, where, $\forall n \in \mathcal{V}, C_n \subset \mathbb{R}^{P \times Q}$, such that $C = C_1 \times \cdots \times C_n$. This minimization problem is nonconvex with respect to (\mathbf{M}, e) . In the next section we propose an algorithmic scheme relying on proximal alternating minimization strategy allowing to build a sequence that converges to a critical point of (2).

III. PROPOSED ALGORITHM

A. Semi-Linearized Proximal Alternating Method (SL-PAM)

The most encountered strategy to find an estimate (\dot{M}, \hat{e}) of (2) relies on the Gauss-Seidel scheme (i.e., coordinate descent method), whose convergence guarantees require the minimum to be uniquely attained at each update of the algorithm, e.g. by assuming the strict convexity with respect to one argument when the other one is fixed. Such a condition being difficult to satisfy in practice, it has been proposed to perform proximal regularization of the Gauss-Seidel scheme. This yields to

$$\begin{split} \boldsymbol{M}^{[k+1]} &\in \arg\min_{\boldsymbol{M}} \Psi(\boldsymbol{M}, e^{[k]}) + \frac{\mu_k}{2} \|\boldsymbol{M} - \boldsymbol{M}^{[k]}\|_2^2 \\ e^{[k+1]} &\in \arg\min_{\boldsymbol{e}} \Psi(\boldsymbol{M}^{[k+1]}, \boldsymbol{e}) + \frac{\eta_k}{2} \|\boldsymbol{e} - e^{[k]}\|_2^2, \end{split}$$

where μ_k and η_k are positive real numbers [14]. The proof of convergence to a critical point of such a scheme is due to a recent work by Attouch et al. [15]. The proof is provided in the nonsmooth and nonconvex setting. The main practical issue is to have a closed form expression of both proximity operators². This has been relaxed by Bolte et al. [16], who derived a proximal alternating linearized scheme (PALM). A hybrid version, named SL-PAM for semi-linearized proximal alternating direction method, has been proposed in [8], especially adapted to the resolution of the Discrete Mumford-Shah problem. The iterations of SL-PAM to solve (2) would read

$$\boldsymbol{M}^{[k+1]} \in \operatorname{prox}_{\frac{\mathcal{L}}{\mu_{k}}+\iota_{C}} \left(\boldsymbol{M}^{[k]} - \frac{1}{\mu_{k}} \nabla_{1} \mathcal{S} \left(\boldsymbol{M}^{[k]}, e^{[k]} \right) \right),$$
$$e^{[k+1]} \in \operatorname{prox}_{\frac{1}{\eta_{k}} \left(\mathcal{R} + \mathcal{S} \left(\boldsymbol{M}^{[k+1]}, \cdot \right) \right)} \left(e^{[k]} \right),$$

where ∇_1 denotes the gradient with respect to the first variable. In [8], a closed form expression has been derived for the update of *e* while, for the basic image denoising MS, the first step does not involve the proximity operator of a sum of functions but only the proximity operator of \mathcal{L} , which is much easier to handle with. The proximity operator of a sum of functions is known to have closed form expression for a very limited number of functions (see e.g. [17]–[19] and references therein). Thus, in order to design an algorithmic scheme with convergence guarantees to solve (2), we need to propose a new proximal alternating scheme.

B. New algorithmic scheme – SL2-PAM

In the context of mixing matrix estimation on graph considered in this study, the data-term is differentiable with a Lipschitz gradient, we can thus derive the iterations summarized in Algorithm 1, where P_C denotes the projection onto the set C, whose convergence guarantees are provided in Proposition III.2 and the proof is given in Appendix VI-A.

Algorithm 1 (SL2-PAM) for solving D-MS for mixing matrix estimation (2)

Set $\overline{M^{[0]} \in C}$ and $e^{[0]} \in \mathbb{R}^{|\mathcal{E}|}$. For k = 0, 1, ...Parameter setting: Let $\nu(e^{[k]})$ be the Lipschitz constant of $\nabla_1 \mathcal{S}(\cdot, e^{[k]}) + \nabla \mathcal{L}$. Set $\gamma > 1$, $\mu_k = \gamma \nu(e^{[k]})$ and $\eta_k > 0$. Variable updates: $M^{[k+1]} \in P_C \left(M^{[k]} - \frac{1}{\mu_k} \nabla_1 \mathcal{S}(M^{[k]}, e^{[k]}) - \frac{1}{\mu_k} \nabla \mathcal{L}(M^{[k]}) \right)$ $e^{[k+1]} \in \operatorname{prox}_{\frac{1}{\eta_k}} (\mathcal{R} + \mathcal{S}(M^{[k+1]}, \cdot)) (e^{[k]})$

- **Assumption III.1** *i*) Ψ *is a Kurdyka-Łojasiewcz function* [16, Def. 2.3],
- *ii)* Ψ and \mathcal{R} are bounded below;
- iii) the updating steps of $M^{[k]}$ and $e^{[k]}$ have closed form expressions;
- iv) $(\eta_k)_{k \in \mathbb{N}}$ is a positive sequence such that the stepsizes η_k belong to (η^-, η^+) for some positive $\eta^- \leq \eta^+$;
- v) the sequence $(\mathbf{M}^{[k]}, e^{[k]})_{k \in \mathbb{N}}$ generated by Algorithm 1 is bounded.

Proposition III.2 Under Assumption III.1 and assuming that, $\forall n \in \mathcal{V}, \forall k \in \mathbb{N}, \nabla \mathcal{L} + \nabla_1 \mathcal{S}(\cdot, e^{[k]})$ is globally Lipschitz continuous with moduli $\nu(e^{[k]})$, and there exists $\nu^-, \nu^+ > 0$ such that $\nu^- \leq \nu(e^{[k]}) \leq \nu^+$. Then the sequence $(\mathbf{M}^{[k]}, e^{[k]})_{k \in \mathbb{N}}$ generated by Algorithm 1 converges to a critical point of Problem (2).

¹Let \mathcal{H} be a finite dimensional Hilbert space. The indicator function of a nonempty closed subset $C \subset \mathcal{H}$ is denoted ι_C and is such that, for every $u \in \mathcal{H}, \iota_C(u) = 0$ if $u \in C$ or $+\infty$ otherwise.

²Let $\varphi: \mathcal{H} \to]-\infty, +\infty]$ be a proper and lower semicontinuous function. Given $u \in \mathcal{H}$ and $\gamma > 0$, the proximal map associated to φ is defined by $\operatorname{prox}_{\gamma\varphi}(u) = \arg\min_{v \in \mathcal{H}} \varphi(v) + \frac{\gamma}{2} ||u - v||^2$.

C. Block updates

However, $\nu(\cdot)$ may be very large, essentially due to the large value of the Lipschitz constant of the data-term \mathcal{L} , which implies a very small descent stepsize in the first step of Algorithm 1. In order to accelerate the convergence, we suggest to derive a block-coordinate SL2-PAM relying on the separability of $\Psi(M, e)$ on each node $n \in \mathcal{V}$. Problem (2) can be equivalently written

$$\Psi(\boldsymbol{M}, e) := \sum_{n=1}^{|\nu|} \left\{ \mathcal{L}_n(\boldsymbol{M}_n) + \mathcal{S}_n(\boldsymbol{M}_n, e) + \iota_{C_n}(\boldsymbol{M}_n) \right\} + \mathcal{R}(e)$$

with $\mathcal{S}_n(\boldsymbol{M}_n, e) = \beta \sum_{n=1}^{|\nu|} (1 - e_{n,n'})^2 \|\boldsymbol{M}_n - \boldsymbol{M}_{n'}\|_2^2$ and

 $\mathcal{L}_n(M_n) = ||z_n - M_n s_n||_2^2$. The iterations are provided in Algorithm 2, the convergence results is given in Proposition III.3 and the proof is given in Appendix VI-A.

Algorithm 2 (Block-SL2-PAM) for solving D-MS for mixing matrix estimation (2)

For every
$$n \in \mathcal{V}$$
, set $M_n^{[0]} \in C_n$ and $e^{[0]} \in \mathbb{R}^{|\mathcal{E}|}$.
For $k = 0, 1, ...$
– Parameter setting –
Let $\nu_n(e^{[k]})$ the Lipschitz constant of $\nabla_1 \mathcal{S}_n(\cdot, e^{[k]}) + \nabla \mathcal{L}_n$.
Set $\gamma > 1$, $\mu_{k,n} = \gamma \nu_n(e^{[k]})$ and $\eta_k > 0$.
–Variable updates –
For $n = 1, ..., |\mathcal{V}|$
 $[M_n^{[k+1]} \in P_{C_n}(M_n^{[k]} - \frac{1}{\mu_{k,n}}(\nabla_1 \mathcal{S}(M_n^{[k]}, e^{[k]}) + \nabla \mathcal{L}(M_n^{[k]})))$
 $e^{[k+1]} \in \operatorname{prox}_{\frac{1}{\eta_k}}(\mathcal{R} + \mathcal{S}(M^{[k+1]}, \cdot))$ $(e^{[k]})$

Proposition III.3 Under Assumption III.1 and assuming that, $\forall n \in \mathcal{V}, \forall k \in \mathbb{N}, \nabla \mathcal{L}_n + \nabla_1 \mathcal{S}_n(\cdot, e^{[k]})$ is globally Lipschitz continuous with moduli $\nu_n(e^{[k]})$, and there exists $\nu_n^-, \nu_n^+ > 0$ such that $\nu_n^- \leq \nu_n(e^{[k]}) \leq \nu_n^+$. Then the sequence $(\mathbf{M}^{[k]}, e^{[k]})_{k \in \mathbb{N}}$ generated by Algorithm 2 converges to a critical point of Problem (2).

IV. NUMERICAL EXPERIMENTS

A. Context: estimation of the voting transfer matrices

We propose to illustrate the performance of the proposed Mumford-Shah model on graph to the estimation of the vote transfer matrices between two elections [20]. While in most of the state-of-the-art studies this estimation is performed globally, the challenge here is to provide an estimate of the voting transfer matrices at each location n. In this case, the set of vertices \mathcal{V} of the graph \mathcal{G} models the polling locations, and \mathcal{E} is the set of edges between to nodes associated with the 8 nearest neighbors. The resulting graph is a directed acyclic graph with $|\mathcal{E}| = 8|\mathcal{V}|$. The involved matrices M_n are of size $P^{(2)} \times P^{(1)}$ where $P^{(1)}$ denotes the number of candidates (including null votes and abstention) competing in the first election $E^{(1)}$, and $P^{(2)}$ the number in the second one, denoted by $E^{(2)}$. Hence, s_n (resp. z_n) denotes a vector containing the number of votes cast for each candidate at the *n*-th polling location in election $E^{(1)}$ (resp. $E^{(2)}$). The coefficients in the matrices M_n denote transfer percentage. Thus, for every $n \in \mathcal{V}$, the constraint C_n is selected to be a simplex constraint over the columns to impose value between 0 and 1 with a sum equals to 1. The projection P_{C_n} is computed by means of the efficient implementation provided in [21]. Moreover, in our experiments, \mathcal{R} is set as a ℓ_1 -norm, i.e. $\mathcal{R} = \lambda \| \cdot \|_1$ and the closed form expression of $\operatorname{prox}_{\frac{1}{n_k}} (\mathcal{R} + \mathcal{S}(\mathcal{M}_n^{[k]}, e^{[k]}))$ is given in [8].

B. Synthetic data

The graph is built on the polling locations of Lyon (France) leading to n = 283. To simplify the interpretation and accurately measure the performance of the proposed method, we consider synthetic data, with $P^{(1)} = 4$ and $P^{(2)} = 3$, and two piecewise smooth regions $\mathcal{V}^{(1)}$ and $\mathcal{V}^{(2)}$ such that $\mathcal{V} = \mathcal{V}^{(1)} \cup \mathcal{V}^{(2)}$, with a frontier in between (see Figure 1). We set the sources $(s_n)_{n \in \mathcal{V}}$ at their real value, and generate $(z_n)_{n \in \mathcal{V}}$ signals as follows. For region $i \in \{1, 2\}$, we define two matrices $\overline{M}_{top}^{(i)}$, $\overline{M}_{bottom}^{(i)}$, and we set the ground truth mixing matrix of each polling place $n \in \mathcal{V}^{(i)}$ as $\overline{M}_n^{(i)} = P_C(\omega_n \overline{M}_{bottom}^{(i)} + (1 - \omega_n) \overline{M}_{top}^{(i)} + \varepsilon_n)$. ε_n denotes a white Gaussian noise with standard deviation σ , to model uncertainties as in real data, and, $\forall n \in \mathcal{V}, \omega_n \in [0, 1]$ depends on the latitude l_n of the *n*-th polling place. That is, we divide the interval of all the latitudes into 20 uniformly distributed subsets $(L_k)_{k=1,...,20}$, and we set $\omega_n = \frac{k}{20}$ if $l_n \in L_k$. Notice that, in this case, the gradient models the spatial correlation of voting behavior.



Fig. 1. The black nodes represent the polling places in region $\mathcal{V}^{(1)}$, and the white ones represent those in region $\mathcal{V}^{(2)}$. The top and bottom matrices differ only on coefficient (2,3). However, the additive noise and the projection onto C_n will change all the coefficients in the generated $\overline{M}_n^{(i)}$. Hence, the gradient from north to south is both non trivial and irregular.

C. Results

Good estimation – In Figure 2, we display the results obtained with the Algorithm 2. We performed a grid search on (λ, β) for the Jaccard index³ and the MSE⁴. We notice that for a good choice of regularization parameters (λ, β) the proposed approach is able to provide a good estimate of both the edges (white lines) and the matrix coefficients, here $(\overline{M}_{n,2,3})_n$.

Block versus global – Figures 3(a,b) compare the algorithmic behavior of Algorithms 1 and 2 in terms of iterations (we observe similar behavior when the plots are displayed w.r.t time) for the optimal (β, λ) . As expected, the block strategy needs twice less iterations than the original SL2-PAM to achieve the same accuracy, that is, $\frac{|\Psi(M^{[k+1]}e^{[k+1]})-\Psi(M^{[k]}e^{[k]})|}{|\Psi(M^{[k]}e^{[k]})|} < 10^{-10}$ (cf. Figure 3(b)). The more *n* increases, the more important the gain is.

³If A and B are nonempty sets,
$$\operatorname{Jacc}(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

⁴ $MSE = \frac{1}{|\mathcal{V}|P^{(1)}P^{(2)}} \sum_{n} \sum_{i,j} \|\overline{M_n}(i,j) - M_n(i,j)\|_2^2$



Fig. 2. Numerical results on synthetic data with $\sigma = 0.1$. (a) Ground truth $(\overline{M}_{n,2,3})_n$ and \overline{e} , (b) Result obtained with the proposed method: estimated $(\widehat{M}_{n,2,3})_n$ values displayed by means of the red-blue color on the graph while estimated \widehat{e} is displayed by a white line between two nodes when the estimated value is 1. The colorbar represents the value of coefficient (2,3) in the local matrices. (c) Jaccard index map. (d) MSE map. The red box denotes the optimal choice for (β, λ) associated to the result displayed in (b).



Fig. 3. (a) Comparison of $\Psi(\boldsymbol{M}^{[k]}, e^{[k]})$ w.r.t the iterations for SL2-PAM and Block-SL2-PAM. (b) Comparison of $\frac{|\Psi(\boldsymbol{M}^{[k+1]}, e^{[k+1]}) - \Psi(\boldsymbol{M}^{[k]}, e^{[k]})|}{|\Psi(\boldsymbol{M}^{[k]}, e^{[k]})|}$ w.r.t $|\Psi(M^{[k]}, e^{[k]})|$ the iterations, (c) Comparison of the evolution of $\Psi(M^{[k]}, e^{[k]})$ in a log-log scale w.r.t the iterations for different initializations.

Robustness – We display in Figure 3(c) the evolution of the objective function for Algorithm 2 with different initialization choices for the optimal (β, λ) . Note that the one used to generate Figure 3(a,b) corresponds to the solid red plot. Although the proposed method is nonconvex, we observe that, whatever the initialization, it converges to the same value of the objective function.

V. CONCLUSION AND PERSPECTIVES

In this work, we propose 1) an extension of the famous MS model to graph signal processing, with an application to mixing matrix estimation, 2) two new block-coordinate proximal algorithms with convergence guarantees, 3) the comparison of these two schemes and their robustness in estimating the vote transfer matrices at each polling location, for which we clearly see the efficiency to estimate properly both the matrices M and the transitions e. The Algorithm 2 outperforms the Algorithm 1 in terms of computational cost, and allows us to deal with real data, involving larger datasets.

VI. APPENDIX

A. Common proof for Propositions III.2 and III.3

To prove Proposition III.2 (resp. Proposition III.3), we set $\mathbb{I} =$ $\{\{1, \dots, |\mathcal{V}|\}\}$ (resp. $\mathbb{I} = \{\{1\}, \dots, \{|\mathcal{V}|\}\}$). The proof relies on the proof of [8, Prop.2] (see also [16] for more details). Let $x^{[k]} = (\boldsymbol{M}^{[k]}, e^{[k]})$, the proof relies on:

- 1) A sufficient decrease property: find $\rho_1 > 0$ such that $(\forall k \in$ $\mathbb{N}), \ \frac{\rho_1}{2} \|x^{[k+1]} - x^{[k]}\|^2 \leq \Psi(x^{[k]}) - \Psi(x^{[k+1]});$
- 2) A subgradient lower bound for the iterates gap: assume that $(x^{[k]})_{k \in \mathbb{N}}$ is bounded and find $\rho_2 > 0$ such that $||w^{[k]}|| \le \rho_2 ||x^{[k]} - x^{[k-1]}||$, where $w^{[k]} \in \partial \Psi(x^{[k]})$;

3) Kurdyka-Łojasiewicz (KL) property: assume that Ψ is a KL function and prove that $(x^{[k]})_{k \in \mathbb{N}}$ is a Cauchy sequence.

The proof of steps 1 and 3 follows similar ideas than in [8, Prop.2], with $\rho_1 = \min \left\{ \sum_{n \in \mathbb{I}} (\gamma - 1) \nu_n^-, \eta^- \right\}$. Regarding step 2, we prove the following result:

Lemma VI.1 Assume that the sequence $\{x^{[k]}\}_{k\in\mathbb{N}}$ generated by Algorithm 2 is bounded. Let

$$\begin{aligned} A_{M}^{k} &:= \left(\mu_{k-1,n}(\boldsymbol{M}_{n}^{[k-1]} - \boldsymbol{M}_{n}^{[k]})\right)_{n \in \mathbb{I}} + \nabla \mathcal{L}(\boldsymbol{M}^{[k]}) \\ &+ \nabla_{1} \mathcal{S}(\boldsymbol{M}^{[k]}, e^{[k]}) - \nabla \mathcal{L}(\boldsymbol{M}^{[k-1]}) - \nabla_{2} \mathcal{S}(\boldsymbol{M}^{[k-1]}, e^{[k-1]}) \\ and \\ &A_{e}^{k} := \eta_{k-1}(e^{[k-1]} - e^{[k]}). \end{aligned}$$

Then
$$(A_M^k, A_e^k) \in \partial \Psi(\mathbf{M}^{[k]}, e^{[k]})$$
 and there exists $\chi > 0$ such that $\|(A_M^k, A_e^k)\| \le \|A_M^k\| + \|A_e^k\| \le 2(C + \rho_2)\|x^{[k-1]} - x^{[k]}\|$, where $\rho_2 = \max_{n \in \mathbb{I}} \gamma \nu_n^+ + \eta^+$.

Proof. The optimality condition for the updating step on M_n in Algoritm 2 is given by

$$(\forall n \in \mathbb{I}), \ \nabla \mathcal{L}_n(\boldsymbol{M}_n^{[k-1]}) + \nabla_1 \mathcal{S}_n(\boldsymbol{M}_n^{[k-1]}, e^{[k-1]}) + \mu_{k-1,n}(\boldsymbol{M}_n^{[k]} - \boldsymbol{M}_n^{[k-1]}) + u_n^{[k]} = 0,$$
(3)

where $u_n^{[k]} \in \partial \iota_{C_n}(\boldsymbol{M}_n^{[k]})$. Concatenating (3) on $n \in \mathcal{V}$ yields:

$$\nabla \mathcal{L}(\boldsymbol{M}^{[k-1]}) + \nabla_1 \mathcal{S}(\boldsymbol{M}^{[k-1]}, e^{[k-1]}) \\ + \left(\mu_{k-1,n}(\boldsymbol{M}_n^{[k]} - \boldsymbol{M}_n^{[k-1]})\right)_{n \in \mathbb{I}} + u^{[k]} = 0, \quad (4)$$

where $u^{[k]} \in \partial \iota_C(\mathbf{M}^{[k]})$. Hence, using the subdifferential property [16, Prop. 2.1] we obtain that $\nabla \mathcal{L}(M^{[k]})$ + $\nabla_1 \mathcal{S}(\boldsymbol{M}^{[k]}, e^{[k]}) + u^{[k]} \in \partial_1 \Psi(\boldsymbol{M}^{[k]}, e^{[k]}).$ Similarly, we prove that $\nabla_e \mathcal{S}(\boldsymbol{M}^{[k]}, e^{[k]}) + \xi^{[k]} \in \partial_e \Psi(\boldsymbol{M}^{[k]}, e^{[k]}),$ where $\xi^{[k]} \in \partial \mathcal{R}(e^{[k]})$. Finally, $(A_M^k, A_e^k) \in \partial \Psi(\mathbf{M}^{[k]}, e^{[k]})$.

The end of the proof is the same as for [8, Lemma 2].

REFERENCES

- [1] P. Common and C. Jutten, *Handbook of blind source separation*, Academic Press, New York, 2010.
- [2] Y. Li, S. Amari, A. Cichocki, D.W.C. Ho, and S. Xie, "Underdetermined blind source separation based on sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 423–437, 2006.
- [3] J.M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 5, no. 2, pp. 354–379, 2012.
- [4] A. Tiard, L. Condat, L. Drumetz, J. Chanussot, W. Yin, and X. Zhu, "Robust linear unmixing with enhanced sparsity," in *Proc. IEEE Int. Conf. Image Process.*, Beijing, China, Sept. 2017.
- [5] Adam N Glynn and Jon Walefield, "Ecological inference in the social sciences," *Statistical methodology*, vol. 7, no. 3, pp. 307–322, 2010.
- [6] X. Puig and J. Ginebra, "A cluster analysis of vote transitions," Computational Statistics & Data Analysis, vol. 70, pp. 328–344, 2014.
- [7] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Comm. Pure Applied Math.*, vol. 42, no. 5, pp. 577–685, 1989.
- [8] M. Foare, N. Pustelnik, and L. Condat, "Semi-Linearized Proximal Alternating Minimization for a Discrete Mumford–Shah Model," preprint, Apr. 2018.
- [9] L.I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys.D*, vol. 60, no. 1, pp. 259–268, 1992.
- [10] Tony F Chan and Luminita A Vese, "Active contours without edges," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, 2001.
- [11] X. Cai and G. Steidl, "Multiclass segmentation by iterated ROF thresholding," in Proc. Int. Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition, Sweden, Mar. 2013.
- [12] E. Strekalovskiy and D. Cremers, "Real-time minimization of the piecewise smooth Mumford–Shah functional," in *Proc. European Conference* on Computer Vision, pp. 127–141. Springer, Zurich, Germany, 2014.
- [13] M. Foare, N. Pustelnik, and L. Condat, "A new proximal method for joint image restoration and edge detection with the Mumford-Shah model," in *Proc. Int. Conf. Acoust., Speech Signal Process.*, Calgary, Canada, Apr. 2018.
- [14] A. Auslender, "Asymptotic properties of the Fenchel dual functional and applications to decomposition problems," *Journal of Optimization Theory and Applications*, vol. 73, pp. 427–449, 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," *Mathematics* of Operations Research, vol. 35, no. 2, pp. 438–457, 2010.
- [16] J. Bolte, S. Sabach, and M. Teboulle, "Proximal Alternating Linearized Minimization for nonconvex and nonsmooth problems," *Math. Pro*gram., vol. 146, no. 1-2, pp. 459–494, 2014.
- [17] C. Chaux, J.-C. Pesquet, and N. Pustelnik, "Nested iterative algorithms for convex constrained image recovery problems," *SIAM J. Imaging Sci.*, vol. 2, no. 2, pp. 730–762, Jun. 2009.
- [18] Y. Yu, "On decomposing the proximal map," Proc. Ann. Conf. Neur. Inform. Proc. Syst., pp. 91–99, Dec. 2013.
- [19] N. Pustelnik and L. Condat, "Proximity operator of a sum of functions; application to depth map estimation," *IEEE Signal Process. Lett.*, vol. 24, no. 12, pp. 1827–1831, 2017.
- [20] M. Foucault and J. Chiche, "Les reports de voix aux élections départementales de 2015," *Revue politique et parlementaire*, vol. 117, no. 1075, pp. 61–76, 2015.
- [21] L. Condat, "Fast projection onto the simplex and the 11 ball," Mathematical Programming Series A, vol. 158, no. 1, pp. 575–585, 2016.