



HAL
open science

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. IEEE Signal Processing Letters, 2019, 26 (9), pp.1275-1279. <10.1109/lsp.2019.2917518>. <hal-01948570v2>

HAL Id: hal-01948570

<https://hal.science/hal-01948570v2>

Submitted on 27 May 2019

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.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Discrete Mumford-Shah on graph for mixing matrix estimation

Yacouba Kaloga, Marion Foare, Nelly Pustelnik, Pablo Jensen ^{*†‡}

May 21, 2019

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.

Keywords – Mumford-Shah, graph, mixing matrix estimation, nonconvex optimisation

1 Introduction

This work focuses on mixing matrix estimation on a graph in the context where the sources are known, but that the mixing matrices differ for each node of the graph. This general setting involves a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where \mathcal{E} denotes the edges, \mathcal{V} the vertices, and $|\mathcal{V}|$ the total number of vertices. Our goal is to estimate, for each node $n \in \mathcal{V}$, a mixing matrix $\mathbf{M}_n \in \mathbb{R}^{P \times Q}$ from the observations $\mathbf{z} = (z_n)_{n \in \mathcal{V}} \in \mathbb{R}^{|\mathcal{V}|}$ and the sources $\mathbf{s} = (s_n)_{n \in \mathcal{V}} \in \mathbb{R}^{|\mathcal{V}|}$ such that

$$(\forall n \in \mathcal{V}) \quad z_n = \mathbf{M}_n \mathbf{s}_n. \quad (1)$$

We set $\mathbf{M} = (\mathbf{M}_1, \dots, \mathbf{M}_{|\mathcal{V}|}) \in \mathbb{R}^{P \times Q \times |\mathcal{V}|}$. A specific application of (1), when \mathcal{G} models a regular grid, is encountered in hyperspectral unmixing imagery [1], for which recent contributions in the convex setting, involving total variation, have been proposed (see [2] and references therein). The framework of problem (1) also fits the challenging question of vote transfer matrix estimation at each polling location, for which no state-of-the-art method is able to provide such a local estimate \mathbf{M}_n , this problem being especially ill-posed [3, 4].

The contributions of this paper are: (i) to encompass the mixing matrix estimation problem into the Mumford-Shah formalism to be able to estimate \mathbf{M}_n with limited assumptions: two matrices \mathbf{M}_n and $\mathbf{M}_{n'}$ should be similar if n and n' are neighbor nodes in the graph, except at some locations where sharp behaviors can be captured (Section 2); (ii) a new proximal alternating minimization scheme with convergence guarantees as well as a block version of it, to handle large datasets (Section 3); (iii) experimental validation of the proposed numerical schemes and robustness to the initialization (Section 4).

^{*}This work is supported by CNRS Défi InFinTi 2018 Projet OptMatGeo and ANR GRAPHSIP and by the ACADEMICS Grant of the IDEXLYON project of the Université de Lyon – PIA operated by ANR-16-IDEX-0005.

[†]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.

2 Discrete Mumford-Shah on graph

2.1 Generalities on Mumford-Shah

Proposed in the late 80's, the Mumford-Shah (MS) model [5] 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 $z: \Omega \rightarrow [0, 1]$ a corrupted input image, the Mumford-Shah model reads [5]:

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

In (2), the first term is the data-term, ensuring similarities between the noisy image and the estimate \hat{u}^1 , 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 $\beta, \lambda > 0$ allows us to adjust the contribution of each term in the estimation.

Among the large literature dedicated to the resolution of MS in image processing, we can identify three types of methods. The first class of methods is restricted to piecewise constant approximations such as the Chan-Vese formulation [8], the Potts model [9–12] or its convex relaxation known as the ROF model [13]. Efficient iterative strategies have been designed and a recent work by Cai and Steidl [14] establishes the link between a thresholded-ROF strategy and the MS model. The second class relies on the Blake-Zisserman model [15–17] which handles the piecewise smoothness by considering a different approximation of MS known as the discrete membrane energy. Convergence to local minimizers was proved for the 1D formulation in [16]. However, the major drawbacks of the extensions to image processing derived in [15] and [17] are the weak convergence guarantees, and the lack of flexibility in the data-term choice, whose proximity operator should have a closed form expression. Finally, for handling with the exact MS model, which estimates jointly the contours and the image, we can refer to Ambrosio-Tortorelli alternatives [18–20], at the price of a huge computational time, or the strategy we proposed in [21]. Note that in the literature, the extension of MS to graphs relies on a Chan-Vese formulation [22]. Thus, it is limited to foreground/background segmentation, with weak convergence guarantees due to the structure of the proposed objective function, and cannot handle more complex data-term.

2.2 Extension to graphs and mixing matrix estimation

In this work, we extend the Discrete MS-type (D-MS) model proposed in [21] for image denoising to mixing matrix estimation on graph. We propose to formulate the data-term as

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

The second term aims to promote small variations between two adjacent matrices \mathbf{M}_n and $\mathbf{M}_{n'}$:

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

where $\|\cdot\|_F$ stands for the Frobenius norm, $\mathcal{N}(n)$ 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 single penalization ensures $e_{n,n'} \equiv 1$, which cancels this term during the minimization. Thus, another penalization term on e favoring sparsity is required, e.g. a ℓ_1 -norm or a quadratic- ℓ_1 penalization [23]. This penalization over e , controlled by a parameter $\lambda > 0$, will be denoted by \mathcal{R} . Intuitively, \mathcal{R} tends to favor $e_{n,n'} = 0$, except at the locations of strong variations (that is, the contour), where \mathcal{S} must vanish, meaning

¹Note that Problem (2) depends on a pair (u, Γ) of a priori independent objects, and requires to work with a relaxed version of (2) in the class of special functions with bounded variations, depending only on u to ensure the existence of \hat{u} . More details in [6, 7].

$e_{n,n'} \approx 1$. Note that the balance between \mathcal{S} and \mathcal{R} ensures $e_{n,n'} \in [0, 1]$, since $e_{n,n'} < 0$ increases \mathcal{S} while $e_{n,n'} > 1$ increases \mathcal{R} . Unlike image denoising tasks, mixing matrix estimation on graph requires additional hard constraints C over \mathbf{M} , e.g. a simplex constraint on the columns of each \mathbf{M}_n . The resulting D-MS model in the context of mixing matrix estimation on graph reads

$$\begin{aligned} & \text{minimize } \Psi(\mathbf{M}, e) := \mathcal{L}(\mathbf{M}) + \mathcal{S}(\mathbf{M}, e) + \mathcal{R}(e) + \iota_C(\mathbf{M}) \\ & \text{where } \mathbf{M} \in \mathbb{R}^{P \times Q \times |\mathcal{V}|} \text{ and } \iota_C \text{ denotes an indicator function}^2 \text{ separable in } n, \text{ i.e. } \iota_C(\mathbf{M}) = \sum_{n=1}^{|\mathcal{V}|} \iota_{C_n}(\mathbf{M}_n), \\ & \text{where, } \forall n \in \mathcal{V}, C_n \subset \mathbb{R}^{P \times Q}, \text{ such that } C = C_1 \times \dots \times C_{|\mathcal{V}|}. \end{aligned} \quad (3)$$

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.

3 Proposed algorithm

3.1 Semi-Linearized Proximal Alternating Method (SL-PAM)

The most encountered strategy to find a solution $(\widehat{\mathbf{M}}, \widehat{e})$ of (3) relies on the Gauss-Seidel scheme (i.e., coordinate descent method), whose convergence guarantees require the minimum of (3) 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

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

where μ_k and η_k are positive real numbers [24]. The proof of convergence to a critical point of such a scheme is due to a recent work by Attouch et al. [25]. The proof is provided in the nonsmooth and nonconvex setting. The main practical issue is to have a closed form expression of both proximity operators³. This has been relaxed by Bolte et al. [26], 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 [21], especially adapted to the resolution of the D-MS problem. The iterations of SL-PAM to solve (3) would read

$$\begin{aligned} \mathbf{M}^{[k+1]} & \in \text{prox}_{\frac{\epsilon}{\mu_k} + \iota_C} \left(\mathbf{M}^{[k]} - \frac{1}{\mu_k} \nabla_1 \mathcal{S}(\mathbf{M}^{[k]}, e^{[k]}) \right), \\ e^{[k+1]} & \in \text{prox}_{\frac{1}{\eta_k}} (\mathcal{R} + \mathcal{S}(\mathbf{M}^{[k+1]}, \cdot)) (e^{[k]}), \end{aligned}$$

where ∇_1 denotes the gradient with respect to the first variable. In [21], a closed form expression has been derived for the update of e while, for the basic image denoising MS, the update of \mathbf{M} only involves the proximity operator of \mathcal{L} . When dealing with mixing matrix estimation, it is thus of almost importance of either having a closed form expression of the proximity operator involved in the update of $\mathbf{M}^{[k+1]}$ or providing a new iterative scheme without such an updating step. However, 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. [27–29] and references therein), which does not fit to the update of $\mathbf{M}^{[k+1]}$. We thus propose a new proximal alternating scheme.

3.2 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,

²Let \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.

³Let $\varphi: \mathcal{H} \rightarrow]-\infty, +\infty]$ be a proper and lower semicontinuous function. Given $u \in \mathcal{H}$ and $\gamma > 0$, the proximal map associated to φ is defined by $\text{prox}_{\gamma\varphi}(u) = \arg \min_{v \in \mathcal{H}} \varphi(v) + \frac{\gamma}{2} \|u - v\|^2$.

where P_C denotes the projection onto the set C , whose convergence guarantees are provided in Proposition 3.2 and the proof is given in Appendix 6.1.

Algorithm 1 (SL2-PAM) for solving D-MS (3)

Set $\mathbf{M}^{[0]} \in C$ and $e^{[0]} \in \mathbb{R}^{|\mathcal{E}|}$.

For $k = 0, 1, \dots$

– 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 $\eta_k > 0$.
 – Variables updates –
 $\mathbf{M}^{[k+1]} \in P_C \left(\mathbf{M}^{[k]} - \frac{1}{\mu_k} \nabla_1 \mathcal{S}(\mathbf{M}^{[k]}, e^{[k]}) - \frac{1}{\mu_k} \nabla \mathcal{L}(\mathbf{M}^{[k]}) \right)$
 $e^{[k+1]} \in \text{prox}_{\frac{1}{\eta_k}}(\mathcal{R} + \mathcal{S}(\mathbf{M}^{[k+1]}, \cdot))(e^{[k]})$

Assumption 3.1 *i) Ψ is a Kurdyka-Łojasiewicz function [26, Def. 2.3],*

ii) Ψ and \mathcal{R} are bounded below;

iii) the updating steps of $\mathbf{M}^{[k]}, e^{[k]}$ 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) $\forall n \in \mathcal{V}, \forall k \in \mathbb{N}, \nabla \mathcal{L} + \nabla_1 \mathcal{S}(\cdot, e^{[k]})$ is globally Lipschitz continuous with moduli $\nu(e^{[k]})$, and there exists $\nu^-, \nu^+ > 0$ such that $\nu^- \leq \nu(e^{[k]}) \leq \nu^+$.

Proposition 3.2 *Assume that the sequence $(\mathbf{M}^{[k]}, e^{[k]})_{k \in \mathbb{N}}$ generated by Algorithm 1 is bounded, and Assumption 3.1. Then this sequence converges to a critical point of Problem (3).*

3.3 Block updates

In real applications, $\nu(\cdot)$ 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 $\Psi(\cdot, e)$ on each node $n \in \mathcal{V}$. Problem (3) can be equivalently written

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

with $\mathcal{S}_n(\mathbf{M}_n, e) = \beta \sum_{n' \in \mathcal{N}(n)} (1 - e_{n,n'})^2 \|\mathbf{M}_n - \mathbf{M}_{n'}\|_F^2$ and $\mathcal{L}_n(\mathbf{M}_n) = \|z_n - \mathbf{M}_n s_n\|_2^2$. The iterations are provided in Algorithm 2, the convergence results is given in Proposition 3.3 and the proof is given in Appendix 6.1.

Proposition 3.3 *Assume that the sequence $(\mathbf{M}^{[k]}, e^{[k]})_{k \in \mathbb{N}}$ generated by Algorithm 2 is bounded, and Assumption 3.1. Then this sequence converges to a critical point of Problem (3).*

4 Numerical experiments

4.1 Context: estimation of the voting transfer matrices

We illustrate the performance of the proposed D-MS model on graph to the estimation of the vote transfer matrices between two elections [30]. 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

Algorithm 2 (Block-SL2-PAM) for solving D-MS (3)

For every $n \in \mathcal{V}$, set $\mathbf{M}_n^{[0]} \in C_n$ and $e^{[0]} \in \mathbb{R}^{|\mathcal{E}|}$.
For $k = 0, 1, \dots$
– 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 $\eta_k > 0$.
– Variables updates –
For $n = 1, \dots, |\mathcal{V}|$
 $[\mathbf{M}_n^{[k+1]} \in P_{C_n}(\mathbf{M}_n^{[k]} - \frac{1}{\mu_{k,n}} (\nabla_1 \mathcal{S}(\mathbf{M}_n^{[k]}, e^{[k]}) + \nabla \mathcal{L}(\mathbf{M}_n^{[k]}))$
 $e^{[k+1]} \in \text{prox}_{\frac{1}{\eta_k}}(\mathcal{R} + \mathcal{S}(\mathbf{M}^{[k+1]}, \cdot))(e^{[k]})$

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 \mathbf{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 \mathbf{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 [31]. Moreover, in our experiments, $\mathcal{R} = \lambda \|\cdot\|_1$ and the closed form expression of $\text{prox}_{\frac{1}{\eta_k}}(\mathcal{R} + \mathcal{S}(\mathbf{M}_n^{[k]}, \cdot))$ is given in [21], so that Assumption 3.1 *iii*) is fulfilled by the proposed functional. This specific choice implies \mathcal{R} to be a bounded below, proper, lower semi-continuous, and semi-algebraic function, and so does Ψ . Then it satisfies Assumption 3.1 *i*)-*ii*) (see [26, Th. 5.1]), while Assumption 3.1 *iv*) is ensured by the definition of \mathcal{S} provided in Section III.B.

4.2 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 $\overline{\mathbf{M}}_{top}^{(i)}$, $\overline{\mathbf{M}}_{bottom}^{(i)}$, and we set the ground truth mixing matrix of each polling place $n \in \mathcal{V}^{(i)}$ as $\overline{\mathbf{M}}_n^{(i)} = P_C(\omega_n \overline{\mathbf{M}}_{bottom}^{(i)} + (1 - \omega_n) \overline{\mathbf{M}}_{top}^{(i)} + \varepsilon_n)$. 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 $(L_k)_{k=1, \dots, 20}$, 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.

4.3 Results

State-of-the-art for voting transfer matrices – To the best of our knowledge, it is not possible to provide fair comparisons to the state-of-the-art, because there are no comparable methods capable of dealing with a large number of nodes and recovering matrix gradients. Indeed, one can either encounter estimation of a unique transfer matrix or, at the price of MCMC simulations, provide clusters of vertices sharing *the same* matrix [4], therefore, it cannot recover matrix gradients as consider here.

$$\begin{aligned} \overline{M}_{top}^{(1)} &= \begin{pmatrix} 0.9 & 0.5 & 3.0 & 3.4 \\ 2.7 & 4.3 & 0 & 2.8 \\ 6.4 & 5.2 & 0 & 3.8 \end{pmatrix} \cdot 10^{-1} & \overline{M}_{top}^{(2)} &= \begin{pmatrix} 9.3 & 2.5 & 8.4 & 2 \\ 0.6 & 1.2 & 0 & 6.7 \\ 0.1 & 6.3 & 1.2 & 1.4 \end{pmatrix} \cdot 10^{-1} \\ \overline{M}_{bottom}^{(1)} &= \begin{pmatrix} 0.9 & 0.5 & 3.0 & 3.4 \\ 2.7 & 4.3 & 9.5 & 2.8 \\ 6.4 & 5.2 & 0 & 3.8 \end{pmatrix} \cdot 10^{-1} & \overline{M}_{bottom}^{(2)} &= \begin{pmatrix} 9.3 & 2.5 & 8.4 & 2 \\ 0.6 & 1.2 & 9.5 & 6.7 \\ 0.1 & 6.3 & 1.2 & 1.4 \end{pmatrix} \cdot 10^{-1} \end{aligned}$$

Figure 1: Black (resp. white) nodes model the polling places in region $\mathcal{V}^{(1)}$ (resp. $\mathcal{V}^{(2)}$). $\overline{M}_{top}^{(\cdot)}, \overline{M}_{bottom}^{(\cdot)}$ 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 is both non trivial and irregular.

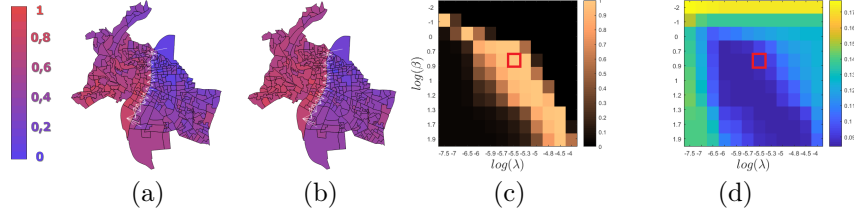


Figure 2: Numerical results on synthetic data with $\sigma = 0.1$. (a) Ground truth $(\overline{M}_{n,2,3})_n$ and \bar{e} , (b) Result obtained with the proposed method: estimated $(\widehat{M}_{n,2,3})_n$ values displayed by means of the red-blue color on the graph while estimated \widehat{e} is displayed 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).

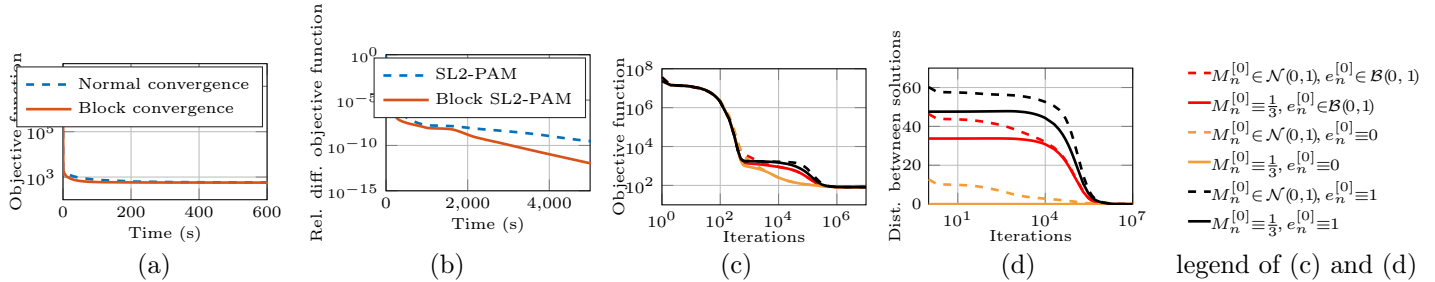


Figure 3: (a) Comparison of $\Psi(\mathbf{M}^{[k]}, e^{[k]})$ w.r.t the time for SL2-PAM and Block-SL2-PAM. (b) Comparison of $\frac{|\Psi(\mathbf{M}^{[k+1]}, e^{[k+1]}) - \Psi(\mathbf{M}^{[k]}, e^{[k]})|}{|\Psi(\mathbf{M}^{[k]}, e^{[k]})|}$ w.r.t the time. (c) Comparison of the evolution of $\Psi(\mathbf{M}^{[k]}, e^{[k]})$ in a log-log scale w.r.t the iterations for different initializations. (d) Distance between 10 solutions to the solution obtained in (c) w.r.t the iterations for different initializations.

Good estimation – In Figure 2, we display the results obtained with the Algorithm 2. We performed a grid search on (β, λ) for the Intersection Over Union (IOU) and the MSE⁴. We notice that for a good choice of regularization parameters (β, λ) 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 time (we observe similar behavior when the plots are displayed w.r.t iterations) for the optimal (β, λ) . The block strategy needs twice less time than the original SL2-PAM to achieve the same accuracy, that is, $\frac{|\Psi(\mathbf{M}^{[k+1]}, e^{[k+1]}) - \Psi(\mathbf{M}^{[k]}, e^{[k]})|}{|\Psi(\mathbf{M}^{[k]}, e^{[k]})|} < 10^{-10}$ (cf. Figure 3(b)). The more n increases, the more important the gain is.

Robustness – We display in Figure 3(c) the evolution of the objective function for Algorithm 2 with

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

different initialization choices for the optimal (β, λ) . We repeat the experiments 10 times, and we display in Figure 3(d) the average distance of the minimizers to the one we obtained in Figure 3(c). Note that the one used to generate Figure 3(a,b) corresponds to the solid orange plot. Although the proposed method is nonconvex, we observe that, whatever the initialization, it converges to the same value of the objective function, and the minimizers are very similar. Although this analysis does not guarantee the uniqueness of the minimizer, the proposed functional does not seem to get stuck in a non global local minimum.

5 Conclusion and perspectives

In this work, we propose an extension of the MS model to graph signal processing, with an application to mixing matrix estimation, two new block-coordinate proximal algorithms with convergence guarantees, the comparison of these two schemes (the block strategy outperforms the global one in terms of computational cost) 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 \mathbf{M} and the transitions e . We are now working on real data with political scientists, to describe the geographical variations of vote matrices and relate these to political characteristics. A Matlab toolbox is available on the Github repository [mfoare/DMS-Graph](https://github.com/mfoare/DMS-Graph).

6 Appendix

6.1 Common proof for Propositions 3.2 and 3.3

To prove Proposition 3.2 (resp. Proposition 3.3), we set $\mathbb{I} = \{\{1, \dots, |\mathcal{V}|\}\}$ (resp. $\mathbb{I} = \{\{1\}, \dots, \{|\mathcal{V}|\}\}$). Let $x^{[k]} = (\mathbf{M}^{[k]}, e^{[k]})$, the proof relies on the proof of [21, Prop.2] (see also [26] for more details):

- 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 $\|w^{[k]}\| \leq \rho_2 \|x^{[k]} - x^{[k-1]}\|$, 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 [21, Prop.2], with $\rho_1 = \min \{ \sum_{n \in \mathbb{I}} (\gamma - 1) \nu_n^-, \eta^- \}$. Regarding step 2, we prove the following result:

Lemma 6.1 *Assume that the sequence $\{x^{[k]}\}_{k \in \mathbb{N}}$ generated by Algorithm 2 is bounded. Let*

$$A_M^k := (\mu_{k-1, n} (\mathbf{M}_n^{[k-1]} - \mathbf{M}_n^{[k]}))_{n \in \mathbb{I}} + \nabla \mathcal{L}(\mathbf{M}^{[k]}) \\ + \nabla_1 \mathcal{S}(\mathbf{M}^{[k]}, e^{[k]}) - \nabla \mathcal{L}(\mathbf{M}^{[k-1]}) - \nabla_2 \mathcal{S}(\mathbf{M}^{[k-1]}, e^{[k-1]})$$

and

$$A_e^k := \eta_{k-1} (e^{[k-1]} - e^{[k]}).$$

Then $(A_M^k, A_e^k) \in \partial\Psi(\mathbf{M}^{[k]}, e^{[k]})$ and there exists $\chi > 0$ such that

$$\|(A_M^k, A_e^k)\| \leq \|A_M^k\| + \|A_e^k\| \leq 2(C + \rho_2) \|x^{[k-1]} - x^{[k]}\|,$$

where $\rho_2 = \max_{n \in \mathbb{I}} \gamma \nu_n^+ + \eta^+$.

Proof. The optimality condition for the updating step on \mathbf{M}_n in Algorithm 2 is given by

$$(\forall n \in \mathbb{I}), \nabla \mathcal{L}_n(\mathbf{M}_n^{[k-1]}) + \nabla_1 \mathcal{S}_n(\mathbf{M}_n^{[k-1]}, e^{[k-1]}) + \mu_{k-1,n}(\mathbf{M}_n^{[k]} - \mathbf{M}_n^{[k-1]}) + u_n^{[k]} = 0, \quad (4)$$

where $u_n^{[k]} \in \partial \iota_{C_n}(\mathbf{M}_n^{[k]})$. Concatenating (4) on $n \in \mathcal{V}$ yields:

$$\nabla \mathcal{L}(\mathbf{M}^{[k-1]}) + \nabla_1 \mathcal{S}(\mathbf{M}^{[k-1]}, e^{[k-1]}) + (\mu_{k-1,n}(\mathbf{M}_n^{[k]} - \mathbf{M}_n^{[k-1]}))_{n \in \mathbb{I}} + u^{[k]} = 0, \quad (5)$$

where $u^{[k]} \in \partial \iota_C(\mathbf{M}^{[k]})$. Hence, using the subdifferential property [26, Prop. 2.1] we obtain that $\nabla \mathcal{L}(\mathbf{M}^{[k]}) + \nabla_1 \mathcal{S}(\mathbf{M}^{[k]}, e^{[k]}) + u^{[k]} \in \partial_1 \Psi(\mathbf{M}^{[k]}, e^{[k]})$. Similarly, we prove that $\nabla_e \mathcal{S}(\mathbf{M}^{[k]}, e^{[k]}) + \xi^{[k]} \in \partial_e \Psi(\mathbf{M}^{[k]}, e^{[k]})$, 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 [21, Lemma 2].

References

- [1] 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.
- [2] 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.
- [3] Adam N Glynn and Jon Wakefield, “Ecological inference in the social sciences,” *Statistical methodology*, vol. 7, no. 3, pp. 307–322, 2010.
- [4] X. Puig and J. Ginebra, “A cluster analysis of vote transitions,” *Computational Statistics & Data Analysis*, vol. 70, pp. 328–344, 2014.
- [5] 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.
- [6] E. De Giorgi, M. Carriero, and A. Leaci, “Existence theorem for a minimum problem with free discontinuity set,” *Archive for Rational Mechanics and Analysis*, vol. 108, no. 4, pp. 195–218, 1989.
- [7] G. Dal Maso, J.-M. Morel, and S. Solimini, “A variational method in image segmentation: existence and approximation results,” *Acta Mathematica*, vol. 168, no. 1, pp. 89–151, 1992.
- [8] Tony F Chan and Luminita A Vese, “Active contours without edges,” *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, 2001.
- [9] R. B. Potts, “Some generalized order-disorder transformations,” in *Mathematical proceedings of the Cambridge philosophical society*. Cambridge University Press, 1952, vol. 48, pp. 106–109.
- [10] S. Geman and D. Geman, “Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images,” in *Readings in Computer Vision*, pp. 564–584. Elsevier, 1987.
- [11] M. Storath and A. Weinmann, “Fast partitioning of vector-valued images,” *SIAM J. Imaging Sci.*, vol. 7, no. 3, pp. 1826–1852, 2014.
- [12] L. Kiefer, M. Storath, and A. Weinmann, “Iterative Potts minimization for the recovery of signals with discontinuities from indirect measurements—the multivariate case,” *arXiv preprint arXiv:1812.00862*, 2018.

- [13] 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.
- [14] 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.
- [15] K. Hohm, M. Storath, and A. Weinmann, “An algorithmic framework for Mumford–Shah regularization of inverse problems in imaging,” *Inverse Problems*, vol. 31, no. 11, pp. 115011 (30pp), 2015.
- [16] A. Weinmann and M. Storath, “Iterative Potts and Blake–Zisserman minimization for the recovery of functions with discontinuities from indirect measurements,” *Proc. R. Soc. A*, vol. 471, no. 2176, pp. 20140638, 2015.
- [17] E. Strelakovsky 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.
- [18] L. Ambrosio and V.M. Tortorelli, “Approximation of functional depending on jumps by elliptic functional via t-convergence,” *Comm. Pure Applied Math.*, vol. 43, no. 8, pp. 999–1036, 1990.
- [19] L. Ambrosio and V.M. Tortorelli, “On the approximation of free discontinuity problems,” *Boll. Un. Mat. Ital.*, vol. 6, no. B, pp. 105–123, 1992.
- [20] M. Foare, J.-O. Lachaud, and H. Talbot, “Image restoration and segmentation using the Ambrosio–Tortorelli functional and discrete calculus,” in *Proc. Int. Conf. Patt. Rec.*, Cancun, Mexico, 2016.
- [21] M. Foare, N. Pustelnik, and L. Condat, “Semi-Linearized Proximal Alternating Minimization for a Discrete Mumford–Shah Model,” preprint hal-01782346, Apr. 2018.
- [22] L. Grady and C.V. Alvino, “The piecewise smooth Mumford–Shah functional on an arbitrary graph,” *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2547–2561, 2009.
- [23] 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.
- [24] 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.
- [25] 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.
- [26] J. Bolte, S. Sabach, and M. Teboulle, “Proximal Alternating Linearized Minimization for nonconvex and nonsmooth problems,” *Math. Program.*, vol. 146, no. 1-2, pp. 459–494, 2014.
- [27] C. Chau, 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.
- [28] Y. Yu, “On decomposing the proximal map,” *Proc. Ann. Conf. Neur. Inform. Proc. Syst.*, pp. 91–99, Dec. 2013.

- [29] 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.
- [30] 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.
- [31] L. Condat, “Fast projection onto the simplex and the l1 ball,” *Mathematical Programming Series A*, vol. 158, no. 1, pp. 575–585, 2016.