Semi-Linearized Proximal Alternating Minimization for a Discrete Mumford–Shah Model

Marion Foare, Nelly Pustelnik, Laurent Condat

Semi-Linearized Proximal Alternating Minimization for a Discrete Mumford–Shah Model

Marion Foare, Nelly Pustelnik, and Laurent Condat *

April 30, 2018

Abstract

The Mumford–Shah model is a standard model in image segmentation and many approximations have been proposed in order to approximate it. The major interest of this functional is to be able to perform jointly image restoration and contour detection. In this work, we propose a general formulation of the discrete counterpart of the Mumford–Shah functional, adapted to nonsmooth penalizations, fitting the assumptions required by the Proximal Alternating Linearized Minimization (PALM), with convergence guarantees. A second contribution aims to relax some assumptions on the involved functionals and derive a novel Semi-Linearized Proximal Alternated Minimization (SL-PAM) algorithm, with proved convergence. We compare the performances of the algorithm with several nonsmooth penalizations, for Gaussian and Poisson denoising, image restoration and RGB-color denoising. We compare the results with state-of-the-art convex relaxations of the Mumford–Shah functional, and a discrete version of the Ambrosio–Tortorelli functional. We show that the SL-PAM algorithm is faster than the original PALM algorithm, and leads to competitive denoising, restoration and segmentation results.

Keywords – Segmentation, restoration, inverse problems, nonsmooth optimization, nonconvex optimization, proximal algorithms, PALM, Mumford–Shah.

1 Introduction

The topic of inverse problems is of major interest for a large panel of applications going from microscopy (see e.g. [2, 3]) or tomography (see [4, 5, 6, 7] and the reference therein) to atmospheric science and oceanography [8]. The pioneering regularization approaches to solve inverse problems

*M. Foare (corresponding author) and N. Pustelnik are with Univ. Lyon, ENS de Lyon, Univ. Lyon 1, CNRS, Laboratoire de Physique, F-69342 Lyon, France. L. Condat is with Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France.

This work is supported by Defi Imag'In SIROCCO.

Part of this work has been presented in [1]. In this extended paper, the added contributions are 1) a general discrete Mumford-Shah model, 2) the convergence proof of the algorithm proposed in [1] and more experimental results.

can be traced back to the works by Tikhonov [9] and by Geman and Geman [10]. The major challenge of this topic consists in designing jointly a cost function and an algorithm (to estimate its minimum) in order to obtain a solution that is the closest to the original unknown one. The recent development of proximal algorithms [11, 12] led to significant advances, thanks to the possibility to efficiently deal with large-size data and nonsmooth objective functions (e.g., nonlocal total-variation constraints, analysis-synthesis formulation, Kullback-Leibler divergence) [13].

In this work, we focus on image restoration and we denote the multicomponent image to recover by $\overline{\mathbf{u}} = (\overline{\mathbf{u}}_m)_{1 \leq m \leq M} \in \mathbb{R}^{NM}$, where each column of $\overline{\mathbf{u}}$ is the vectorized representation of the m-th component. The degradation model we consider takes the form:

$$(\forall m \in \{1, \ldots, M\}) \qquad \mathbf{z}_m = \mathcal{D}_\alpha(A_m \overline{\mathbf{u}}_m), \tag{1}$$

where $A_m \in \mathbb{R}^{L \times N}$ models a linear degradation (e.g. a blur, a compressed sensing matrix, a wrapping matrix) and $\mathcal{D}_\alpha \colon \mathbb{R}^L \to \mathbb{R}^L$ denotes a random degradation that can be white Gaussian noise, leading to an additive model, or Poisson noise. The objective of this work is to estimate jointly the restored image $\widehat{\mathbf{u}}$ and its contours, denoted by $\widehat{\mathbf{e}}$ in the following, from the degraded data $\mathbf{z}$.

One of the standard (variational) approach to solving such an ill-posed inverse problem consists in dealing with a regularization of the problem, by minimizing a sum of functionals. The variational formulation of this problem, when white Gaussian noise is involved, reads:

$$\widehat{\mathbf{u}} = \operatorname*{argmin}_{\mathbf{u}} \frac{1}{2\sigma^2} \|A\mathbf{u} - \mathbf{z}\|_2^2 + \rho(\mathbf{u}), \tag{2}$$

where ρ is a "well-chosen" regularizing functional, which allows us to denoise, while preserving the discontinuities. Hence, it generally involves the gradient of the estimate. A classical choice is $\rho(\mathbf{u}) = \|D\mathbf{u}\|_0$, where D models the finite difference operator and $\|\cdot\|_0$ is the pseudo-norm ℓ_0, which is known as the L2-Potts model [14], or $\rho(\mathbf{u}) = \mathrm{TV}(\mathbf{u})$, the Total Variation model [15], which is convex. However, these models are restricted to piecewise constant estimates, and do not integrate contour detection in the variational formulation, which is performed as post-processing step. The main limitation of such a two-step procedure for contour detection is the difficulty of appropriately selecting the thresholding rule used for edge detection.

Mumford and Shah proposed to consider a more general regularizing term, depending on both the gradient of the estimate and the set of discontinuities [16]. The latter becomes an unknown in the problem. Since the Mumford–Shah (MS) formalism is generally formulated in a continuous setting, we denote by $\Omega \subset \mathbb{R}^2$ the image domain. The MS model aims at estimating both $\widehat{u} \in \mathrm{W}^{1,2}(\Omega)$[1], a piecewise smooth approximation of an image $z \in \mathrm{L}^\infty(\Omega)$, and the set of discontinuities $K \subset \Omega$, such that the pair $(\widehat{u}, K)$ is an optimal solution of:

$$\operatorname*{minimize}_{u,K} \frac{1}{2} \int_\Omega (u - z)^2 dxdy + \beta \int_{\Omega \setminus K} |\nabla u|^2 dxdy + \lambda |K|, \tag{3}$$

where the first term acts as a data fidelity term and forces the approximation u to be close to z, the second term penalizes strong variations except at the locations K of the strong edges, and $|K|$

[1] $\mathrm{W}^{1,2}(\Omega) = \left\{ u \in \mathrm{L}^2(\Omega) \ : \ \partial u \in \mathrm{L}^2(\Omega) \right\}$ where ∂ denotes the weak derivative operator.

denotes the total length of the arcs forming K, thus the minimization of this functional implies that $|K|$ is small at a solution. Finally, $\beta > 0$ and $\lambda > 0$ denote regularization parameters controlling the smoothness and the length of K respectively.

Following discretization ideas proposed in the original paper of Mumford and Shah [16], we assume that u and z are functions on a lattice instead of functions on a two-dimensional region, and we denote them by $\mathbf{u}$ and $\mathbf{z}$, respectively (referring to (1)). K models the path made up of lines between all pairs of adjacent lattice points where $\mathbf{u}$ has sharp transitions, as illustrated in Figure 1. In a discrete setting, K is thus replaced by the variable $\mathbf{e} \in \mathbb{R}^{|\mathbb{E}|}$, which denotes the edges between nodes (e.g. if the set of edges are limited to the horizontal and vertical edges between two pixels, then $|\mathbb{E}| = 2N - N_1 - N_2$, where $N = N_1 \times N_2$ is the size of the grid), and whose value is 1 when a contour change is detected, and 0 otherwise. A discrete counterpart of (3) can be written:

$$\underset{\mathbf{u}\in\mathbb{R}^{NM},\mathbf{e}\in\mathbb{R}^{|\mathbb{E}|}}{\text{minimize}} \ \frac{1}{2}\|\mathbf{u}-\mathbf{z}\|_2^2 + \beta\|(1-\mathbf{e})\odot D\mathbf{u}\|^2 + \lambda\mathcal{R}(\mathbf{e}), \tag{4}$$

where $D \in \mathbb{R}^{|\mathbb{E}|\times N}$ models a finite difference operator and $\mathcal{R}$ denotes a penalization term, that favors sparse solutions, which is a discrete translation of "short $|K|$". Note that there is no need to add additional constraints on $\mathbf{e}$, since both $(1-\mathbf{e})$ and $\mathcal{R}(\mathbf{e})$ should force it to stay between 0 and 1.

Related works. One of the most popular convex relaxation of the MS functional is the Total Variation (TV) functional [15, 17], which favors piecewise constant results, while preserving the discontinuities. Its ℓ_0 counterpart is studied in [18, 19], leading to the L_2-Potts formulation (ℓ_0-penalization on Du). In the same spirit, but for the original piecewise smooth case, Strekalovskiy et al. proposed in [20] to replace $\beta\mathcal{S} + \lambda\mathcal{R}$ by a single function depending on Du, defined by $\mathcal{R}_{MS}(Du) = \min\{\beta|Du|^2, \lambda\}$. Similar ideas have been derived in [21]. The authors proposed two convex relaxations of the MS functional designed for discrete domain with continuous labels [21]. Nonetheless, this relaxation is not able to detect the contours. As emphasized by the authors, proper convergence may be difficult to achieve for some parameterization, and these two methods are not able to detect the contours. For these reasons, we don't consider it in further comparisons. The Chan–Vese model can also be considered as a relaxation of the MS model, whose main limitation is due to a prior label number, and a piecewise constant estimate [22, 23].

Recently, Li et al. [24] proposed a nonlocal TV model, similar to the AT functional, where the gradient is computed in a weighted neighborhood. Convergence is proved, but the contours are obtained by post-processing the estimated image. The approach of Strekalovskiy et al. [20] relies on a truncated quadratic penalization of the gradient of the estimate. They derive a heuristic algorithm, based on a convex relaxation of the functional they propose, and extract the contours by thresholding. The first author and her collaborators [25, 26] proposed a new formulation of the AT functional in the framework of Discrete Calculus. They obtain true 1-dimensional contours. But since they still have to deal with the ε parameter, their algorithm is particularly slow.

Contributions and outline. In order to jointly identify the edges and to restore the image, our contributions are 1) to define a theoretical framework making the bridge between a discrete version of the Mumford–Shah model, called D-MS, and the objective function handled by the Proximal Alternating Linearized Minimization (PALM) algorithm [27], 2) to provide a new algo-

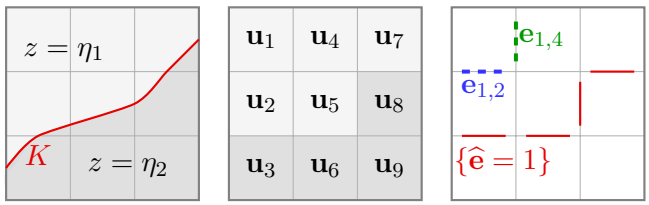


Figure 1: Continuous versus discrete formulations of the MS model (4). In the discrete setting, when $D = [D_h^\top, D_v^\top]^\top$ models the concatenation of the horizontal and vertical difference operators, the values of $D_h\mathbf{u}$ (resp. $D_v\mathbf{u}$) live on the horizontal (resp. vertical) midgrid, and so does $\mathbf{e}$. K and $\{\widehat{\mathbf{e}} = 1\}$ are delineated in red.

rithmic scheme, called Semi-Linearized Proximal Alternating Minimization (SL-PAM), aiming to combine one step of PAM [28] with one step of PALM [27], allowing to relax condition of a stepsize parameter, and having convergence guarantees. The convergence proof is derived. The efficiency of the proposed algorithmic scheme is illustrated on several restoration examples: Gaussian denoising, Poisson denoising, color denoising and image restoration. Comparisons to state-of-the-art approaches are performed on the color denoising example.

Our general D-MS model is defined in Section 2. PALM formulated to solve D-MS is defined in Section 3.1 as well as additional assumptions under the D-MS objective function allowing to ensure convergence. The proposed SL-PAM is derived in Section 3.2. Experiments and comparisons are provided in Section 4.

2 Generalized Discrete Mumford–Shah model

The Discrete-MS (D-MS) model proposed in this work is expressed as follows.

Problem 1. *Let $\mathbf{z} \in \mathbb{R}^{LM}$ and $A \in \mathbb{R}^{LM \times NM}$. Let $\mathcal{L}(A\cdot, \mathbf{z})\colon \mathbb{R}^{NM} \to (-\infty, +\infty]$, be a fidelity term to the data $\mathbf{z}$, and $\mathcal{R}\colon \mathbb{R}^{|\mathbb{E}|} \to (-\infty, +\infty]$, be a regularizer term which enforces sparsity and acts as a length term, both being proper and lower semicontinuous functions. Let $\mathcal{S}\colon \mathbb{R}^{|\mathbb{E}|} \times \mathbb{R}^{NM} \to \mathbb{R}$, be the coupling term, which penalizes strong variations, except at edges, be a C^1 function and such that $\nabla\mathcal{S}$ is Lipschitz continuous on bounded subsets of $\mathbb{R}^{|\mathbb{E}|} \times \mathbb{R}^{NM}$. The general D-MS-like problem we aim to solve reads:*

$$\underset{\mathbf{u}\in\mathbb{R}^{NM}, \mathbf{e}\in\mathbb{R}^{|\mathbb{E}|}}{\textit{minimize}} \; \Psi(\mathbf{u}, \mathbf{e}) := \mathcal{L}(A\mathbf{u}, \mathbf{z}) + \beta\mathcal{S}\big(\mathbf{e}, \mathbf{u}\big) + \lambda\mathcal{R}(\mathbf{e}). \tag{5}$$

The specificities of this problem compared to the state-of-the-art formulation are :

- the generalization of the data-term, allowing to deal with linear degradation and not restricted to the Euclidean norm. For instance, $\mathcal{L}(A\mathbf{u}, \mathbf{z}) = \sum_m \|A_m\mathbf{u}_m - \mathbf{z}_m\|^2$ suited to data corrupted by both a linear degradation and white Gaussian noise [29, 30, 31]. A choice $\mathcal{L}(\mathbf{u}, \mathbf{z}) = \sum_m \|\mathbf{u}_m - \mathbf{z}_m\|_1$ fits data degraded with impulse noise [31], while the choice of the Kullback-Leibler divergence $\mathcal{L}(\mathbf{u}, \mathbf{z}) = \sum_m \mathrm{DKL}(\mathbf{u}_m, \mathbf{z}_m)$ is employed for data corrupted by

Poisson noise [32, 33].

- the possibility to deal with a large panel of regularization terms $\mathcal{R}$. One of the most popular choice of $\mathcal{R}$ encountered in the literature is $\mathcal{R}_{AT}(\mathbf{e}) = \varepsilon\|D\mathbf{e}\|_2^2 + \frac{1}{4\varepsilon}\|\mathbf{e}\|_2^2$, with $\varepsilon > 0$ and D being a difference operator, proposed by Ambrosio and Tortorelli [34, 35]. Such a contour penalization makes (4) Γ-converge to the MS functional as ε tends to 0. As a matter of fact, large values of ε lead to thick contours but help to detect the set of discontinuities. Then, as ε tends to 0, the penalization of $\|\mathbf{e}\|_2^2$ increases and enforces $\mathbf{e}$ to become sparser and sparser, and thus contours becoming thinner and thinner. Numerically, however, it is not possible for ε to be arbitrarily small since it controls the thickness of the contours.

- the flexibility in the coupling term $\mathcal{S}$. Since the MS functional is originally designed with a L_2-penalization of the gradient of $\mathbf{u}$, a common choice for the coupling term is $\mathcal{S}\big(\mathbf{e}, \mathbf{u}\big) = \sum_m \|(1-\mathbf{e}) \odot D\mathbf{u}_m\|_2^2$ [35, 36], where D is defined as in (4). However, in [37], Shah proposed to replace the L_2-norm with a coupling term involving the L_1-norm such as $\mathcal{S}\big(\mathbf{e}, \mathbf{u}\big) = \sum_m \|(1 - \mathbf{e}) \odot (1 - \mathbf{e}) \odot D\mathbf{u}_m\|_1$, combined with the AT regularizer. Alicandro et al. [38] proved the Γ-convergence of this particular functional to a variant of the MS functional, involving the Cantor part of $D\mathbf{u}$. Experiments show that this TV-like coupling term is more robust to image gradients, but eliminates high-frequency content. More recently, Li et al. [24] suggested to set $\mathbf{e}_p = \{\mathbf{e}_p^{(q)}\}_{q\in\mathcal{B}}$, where $\mathcal{B}$ is a box centered at the pixel p, as weights of the dissimilarity $D_p^{(q)}\mathbf{u}_m = \mathbf{u}_{m,p} - \mathbf{u}_{m,p+q}$. The regularization functional is thus a nonlocal TV of the form $\mathcal{S}(\mathbf{e}, \mathbf{u}) = \sum_m \sum_{p\in\mathbb{E}} \sqrt{\Sigma_{q\in\mathcal{B}} \mathbf{e}_q^p (D_p^{(q)}\mathbf{u}_m)^2}$. In this approach, the contours are not obtained from $\mathbf{e}$ but by thresholding $\mathbf{u}$, leading to less accurate estimation.

We can remark that, when dealing with multivariate images, contours can be defined either as similar edges through all the components, or as distinct edges, leading to a path K that may be different for all the components. In order to facilitate the understanding and the reading, we formulate Problem 1 in the context of similar edges. But Problem 1, as well as the following results, can be similarly derived for distinct edges considering $\mathbf{e} \in \mathbb{R}^{|E|\times M}$. When $M = 1$, both formalisms are equivalent.

3 Algorithms

In order to solve Problem 1, we propose two algorithmic strategies. The first one relies on the PALM algorithm [27], and requires additional assumptions on the function involved in order to ensure convergence guarantees. The second is an alternative to PALM, that we called SL-PAM, allowing to relax some of the assumptions made on the coupling term.

3.1 PALM for D-MS

The following Algorithm 1, which is an instance of the generic algorithm PALM [27], is tailored to solving Problem 1:

Algorithm 1 (PALM) for solving D-MS (5)

Set $\mathbf{u}^{[0]} \in \mathbb{R}^{NM}$ and $\mathbf{e}^{[0]} \in \mathbb{R}^{|\mathbb{E}|}$.
For $k \in \mathbb{N}$
- Set $\gamma > 1$ and $c_k = \gamma\nu(\mathbf{e}^{[k]})$.
- $\mathbf{u}^{[k+1]} \in \operatorname{prox}_{\frac{1}{c_k}\mathcal{L}(A\cdot,\mathbf{z})}\left(\mathbf{u}^{[k]} - \frac{1}{c_k}\nabla_{\mathbf{u}}\mathcal{S}\big(\mathbf{e}^{[k]}, \mathbf{u}^{[k]}\big)\right)$
- Set $\delta > 1$ and $d_k = \delta\varepsilon(\mathbf{u}^{[k+1]})$.
- $\mathbf{e}^{[k+1]} \in \operatorname{prox}_{\frac{1}{d_k}\lambda\mathcal{R}}\left(\mathbf{e}^{[k]} - \frac{1}{d_k}\nabla_{\mathbf{e}}\mathcal{S}\big(\mathbf{e}^{[k]}, \mathbf{u}^{[k+1]}\big)\right)$

It consists in updating alternately the image $\mathbf{u}^{[k]}$ and the edges $\mathbf{e}^{[k]}$ by means of proximity operator steps, defined as,

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad \operatorname{prox}_f(\mathbf{x}) = \operatorname*{argmin}_{\mathbf{y}\in\mathbb{R}^N} \frac{1}{2}\|\mathbf{y}-\mathbf{x}\|_2^2 + f(\mathbf{y}), \tag{6}$$

where $f\colon \mathbb{R}^N \to (-\infty, +\infty]$ denotes a proper and lower semi-continuous function. Algorithm 1 converges under some assumptions listed in the following proposition:

Proposition 1. *The sequence $(\mathbf{u}^{[k]}, \mathbf{e}^{[k]})_{k\in\mathbb{N}}$ generated by Algorithm 1 converges to a critical point of Problem 1 if*

i) the updating steps of $\mathbf{u}^{[k+1]}$ and $\mathbf{e}^{[k+1]}$ have closed form expressions;

ii) the sequence $(\mathbf{u}^{[k]}, \mathbf{e}^{[k]})_{k\in\mathbb{N}}$ generated by Algorithm 1 is bounded;

iii) $\mathcal{L}(A\cdot, \mathbf{z})$, $\mathcal{R}$ and $\Psi(\cdot,\cdot)$ are bounded below;

iv) Ψ is a Kurdyka-Łojasiewicz function [27, Definition 2.3];

v) $\nabla_{\mathbf{u}}\mathcal{S}$ and $\nabla_{\mathbf{e}}\mathcal{S}$ are globally Lipschitz continuous with moduli $\nu\big(\mathbf{e}\big)$ and $\varepsilon\big(\mathbf{u}\big)$ respectively, and for all $k \in \mathbb{N}$, $\nu\big(\mathbf{e}^{[k]}\big)$ and $\varepsilon\big(\mathbf{u}^{[k]}\big)$ are bounded by strictly positive constants.

Proof. The form of Problem 1 and the assumptions in Proposition 1 fit the requirements for convergence of the PALM algorithm described in [27, Assumptions A-B, Theorem 3.1]. □

From the practical point of view, the major challenge regarding the assumptions in Proposition 1 is to ensure that $\mathcal{L}$ (resp. $\mathcal{R}$) has a closed form expression for the associated proximity operator. A large number of functions having a closed form expression of their proximal maps is listed in [12, 39, 40] going from ℓ_p-norm to gamma divergences. The main difficulty is due to the linear operator A. Indeed, the proximity operator of a function composed with a linear operator has a closed form expression if

- $\mathcal{L}(\cdot, \mathbf{z}) = \|\cdot - \mathbf{z}\|_2^2$ and A^*A is invertible [39], leading to $(\forall \gamma > 0)(\forall \mathbf{u} \in \mathbb{R}^{NM})$,

$$\gamma \mathcal{L}(A\cdot, \mathbf{z})\mathbf{u} = (\mathbb{I} + \gamma A^*A)^{-1}(\mathbf{u} + \gamma A^* z); \tag{7}$$

- A models a frame (or a semi-orthogonal) linear operator [13], i.e. $A^*A = \mu\mathbb{I}$ with $\mu > 0$, taking the form $(\forall \gamma > 0)(\forall \mathbf{u} \in \mathbb{R}^{NM})$,

$$\text{prox}_{\gamma \mathcal{L}(A\cdot, \mathbf{z})}(\mathbf{u}) = \mathbf{u} + \mu^{-1} A^* (\text{prox}_{\gamma\mu\mathcal{L}}(A\mathbf{u}) - A\mathbf{u}). \tag{8}$$

Moreover, assumption *ii)* in Proposition 1 holds in several scenarios, such as when the functions $\mathcal{L}(A., \mathbf{z})$ and $\mathcal{R}$ have bounded level sets. The reader could refer to [28, Remark 5] and [27, Remark 3.4] for more details about this boundedness assumption.

3.2 Proposed SL-PAM

We propose an alternative to PALM, where the update $\mathbf{u}^{[k+1]}$ exploits the linearization and where the update $\mathbf{e}^{[k+1]}$ relies on the proximity operator of the function $\beta\mathcal{S}(\cdot, \mathbf{u}^{[k+1]}) + \lambda\mathcal{R}$. The resulting Semi-Linearized PAM (SL-PAM) is described in Algorithm 2, that does not require $\varepsilon(\mathbf{u}^{[k]})$ to be bounded and allows us to choose larger d_k.

Algorithm 2 (SL-PAM) algorithm for solving D-MS (5)

Set $\mathbf{u}^{[0]} \in \mathbb{R}^{NM}$ and $\mathbf{e}^{[0]} \in \mathbb{R}^{|\mathbb{E}|}$.
For $k \in \mathbb{N}$
- Set $\gamma > 1$ and $c_k = \gamma\nu(\mathbf{e}^{[k]})$.
- $\mathbf{u}^{[k+1]} \in \text{prox}_{\frac{1}{c_k}\mathcal{L}(A\cdot, \mathbf{z})}\left(\mathbf{u}^{[k]} - \frac{1}{c_k}\nabla_{\mathbf{u}}\mathcal{S}(\mathbf{e}^{[k]}, \mathbf{u}^{[k]})\right)$
- Set $d_k > 0$.
- $\mathbf{e}^{[k+1]} \in \text{prox}_{\frac{1}{d_k}\left(\lambda\mathcal{R} + \beta\mathcal{S}(\cdot, \mathbf{u}^{[k+1]})\right)}\left(\mathbf{e}^{[k]}\right)$

The convergence of Algorithm 2 is ensured under Assumption 1.

Assumption 1. *i) The updating steps of $\mathbf{u}^{[k+1]}$ and $\mathbf{e}^{[k+1]}$ have closed form expressions;*

ii) Ψ is a Kurdyka-Łojasiewicz function;

iii) $\mathcal{L}(A\cdot, \mathbf{z})$, $\mathcal{R}$ and Ψ are bounded below;

iv) $\nabla_{\mathbf{u}}\mathcal{S}$ is globally Lipschitz continuous with moduli $\nu(\mathbf{e}^{[k]})$ $k \in \mathbb{N}$ and there exists $\nu^-, \nu^+ > 0$ such that $\nu^- \leq \nu(\mathbf{e}^{[k]}) \leq \nu^+$;

v) $(d_k)_{k\in\mathbb{N}}$ is a positive sequence such that the stepsizes d_k belong to (d^-, d^+), for some positive $d^- \leq d^+$.

Proposition 2. *Under Assumption 1, and let assume that the sequence $\{\mathbf{x}^{[k]}\}_{k\in\mathbb{N}} = \{(\mathbf{u}^{[k]}, \mathbf{e}^{[k]})\}_{k\in\mathbb{N}}$ generated by Algorithm 2 is bounded. Then*

i) $\Sigma_{k=1}^{\infty}\|\mathbf{x}^{[k+1]} - \mathbf{x}^{[k]}\| < \infty$;

ii) $\{\mathbf{x}^{[k]}\}_{k\in\mathbb{N}}$ *converges to a critical point* $(\mathbf{u}^*, \mathbf{e}^*)$ *of* Ψ.

The proof relies on the general proof recipe given in [27], divided into three main steps: (i) sufficient decrease property, (ii) subgradient lower bound for the iterate gap, and (iii) Kurdyka-Łojasiewicz property. These three steps are detailed thereafter, where we set $\mathbf{x}^{[k]} = (\mathbf{u}^{[k]}, \mathbf{e}^{[k]})$.

3.2.1 Sufficient decrease property

The objective is to find $\rho_1 > 0$ such that

$$(\forall k \in \mathbb{N}) \qquad \frac{\rho_1}{2}\|\mathbf{x}^{[k+1]} - \mathbf{x}^{[k]}\|^2 \leq \Psi(\mathbf{x}^{[k]}) - \Psi(\mathbf{x}^{[k+1]}). \tag{9}$$

This results relies on the following Lemma.

Lemma 1. *Let* $\{\mathbf{x}^{[k]}\}_{k\in\mathbb{N}}$ *be a sequence generated by Algorithm 2. Then*

i) the sequence $\{\Psi(\mathbf{x}^{[k]})\}_{k\in\mathbb{N}}$ *is nonincreasing, and in particular*

$$(\forall k \in \mathbb{N}) \quad \frac{\rho_1}{2}\|\mathbf{x}^{[k+1]} - \mathbf{x}^{[k]}\|^2 \leq \Psi(\mathbf{x}^{[k]}) - \Psi(\mathbf{x}^{[k+1]}),$$

where $\rho_1 = \min\{(\gamma - 1)\nu^-, d^-\}$*;*

ii) $\Sigma_{k=0}^{\infty}\|\mathbf{x}^{[k+1]} - \mathbf{x}^{[k]}\|^2 < \infty$ *and* $\lim_{k\to\infty}\|\mathbf{x}^{[k+1]} - \mathbf{x}^{[k]}\| = 0$.

The proof is given in Appendix 6.1.

3.2.2 A subgradient lower bound for the iterates gap

This step relies on Lemma 2.

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

$$A_{\mathbf{u}}^k := c_{k-1}(\mathbf{u}^{[k-1]} - \mathbf{u}^{[k]}) + \nabla_{\mathbf{u}}\mathcal{S}\big(\mathbf{e}^{[k]}, \mathbf{u}^{[k]}\big) - \nabla_{\mathbf{u}}\mathcal{S}\big(\mathbf{e}^{[k-1]}, \mathbf{u}^{[k-1]}\big), \tag{10}$$

$$A_{\mathbf{e}}^k := d_{k-1}(\mathbf{e}^{[k-1]} - \mathbf{e}^{[k]}). \tag{11}$$

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

$$\|(A_{\mathbf{u}}^k, A_{\mathbf{e}}^k)\| \leq \|A_{\mathbf{u}}^k\| + \|A_{\mathbf{e}}^k\| \leq 2(M + \rho_2)\|\mathbf{x}^{[k-1]} - \mathbf{x}^{[k]}\| \tag{12}$$

where $\rho_2 = \gamma_k\nu^+ + d^+$.

The proof is given in Appendix 6.2.

3.2.3 Kurdyka-Łojasiewicz property

This step relies on the assumption that Ψ is a Kurdyka-Łojasiewicz (KL) function, and proves that the minimizing sequence $\{\mathbf{x}^{[k]}\}_{k\in\mathbb{N}}$ is a Cauchy sequence. According to [27, Theorem 5.1], if $\Psi\colon \mathbb{R}^{NM} \to \mathbb{R}$ is a proper, lower semi-continuous (l.s.c.), and semi-algebraic function, then it satisfies the KL property at any point of $\mathrm{dom}\Psi$. The proof of this step is the same as for [27, Lemma 3.6].

3.3 Additional comments on Assumption 1-i)

The conditions to obtain a closed form expression for the update of $\mathbf{u}^{[k+1]}$ are similar to the ones detailed in Section 3.1. The tedious part concerns the update of $\mathbf{e}^{[k+1]}$ for which a closed form expression is provided in Proposition 3 for specific choices of $\mathcal{S}$ and $\mathcal{R}$.

Proposition 3. *Let $D\colon \mathbb{R}^{|\mathbb{E}|\times N}$. For every $(\mathbf{u},\mathbf{e}) \in \mathbb{R}^{NM}\times\mathbb{R}^{|\mathbb{E}|}$, we assume that*

$$S(\mathbf{e},\mathbf{u}) = \|(1-\mathbf{e})\odot D\mathbf{u}\|_2^2, \tag{13}$$

and that $\mathcal{R}$ is a separable function such that

$$(\forall \mathbf{e}{=}(\mathbf{e}_i)_{1\le i\le|\mathbb{E}|}) \qquad \mathcal{R}(\mathbf{e}) = \sum_{i=1}^{|\mathbb{E}|}\sigma_i(\mathbf{e}_i), \tag{14}$$

where $\sigma_i : \mathbb{R}^{|\mathbb{E}|} \to (-\infty;+\infty]$, and whose proximity operator has a closed form expression. At the iteration $k\in\mathbb{N}$, with $d_k > 0$, $\beta > 0$ and $\lambda > 0$, the updating step on $\mathbf{e}^{[k+1]}$ in Algorithm 2 is equivalent to, for all $i \in \{1,...,|\mathbb{E}|\}$,

$$\mathbf{e}_i^{[k+1]} \in \mathrm{prox}_{\frac{\lambda\sigma_i}{2\beta(D\mathbf{u}^{[k+1]})_i^2+d_k}}\left(\frac{\beta\big(D\mathbf{u}^{[k+1]}\big)_i^2 + \frac{d_k\mathbf{e}_i^{[k]}}{2}}{\beta\big(D\mathbf{u}^{[k+1]}\big)_i^2 + \frac{d_k}{2}}\right). \tag{15}$$

The proof is given in Appendix 6.4.

4 Experiments

Based on the results derived in the previous section, it is now possible to provide efficient algorithmic schemes in order to deal with D-MS, with the possibility of having $\mathcal{R}$ modeling a nonsmooth penalization. To the best of our knowledge, this has never been proposed before.

4.1 Specific choice of D-MS

In our experiments we suggest to choose

$$\mathcal{S}(\mathbf{e},\mathbf{u}) = \|(1-\mathbf{e})\odot D\mathbf{u}\|_2^2, \tag{16}$$

which is C^1 and has Lipschitz continuous gradients. The regularization term is chosen as

$$\mathcal{R}(\mathbf{e}) = \sum_{i=1}^{|\mathbb{E}|} \max\Big\{|\mathbf{e}_i|^p, \frac{|\mathbf{e}_i|^q}{4\epsilon}\Big\}, \tag{17}$$

where $\epsilon > 0$, whose particular cases are:

- the ℓ_0-pseudo norm when $p = 0$ and $\epsilon \to \infty$;
- the ℓ_1-norm when $p = 1$ and $\epsilon \to \infty$;
- the quadratic ℓ_1 penalization, $p = 1$, $q = 2$ and $0 < \epsilon < 1$, derived in [1], which aims to model the quadratic behavior of $\frac{1}{4\epsilon}\|.\|_2^2$ for small ϵ and enforce sparsity.

This function is bounded below, proper, l.s.c., separable, and semi-algebraic (see [27, Example 5.3]). The associated proximity operator of the quadratic ℓ_1-penalization is:

Proposition 4. *For every $\eta \in \mathbb{R}$,*

$$\operatorname{prox}_{\tau \max\left\{|.|, \frac{|.|^2}{4\epsilon}\right\}}(\eta) = \operatorname{sign}(\eta)\ \max\Big\{0, \min\Big[|\eta| - \tau, \max\Big(4\epsilon, \frac{|\eta|}{\frac{\tau}{2\epsilon}+1}\Big)\Big]\Big\}. \tag{18}$$

The choice of $\mathcal{L}(\cdot, \mathbf{z})$ will be dependent on the restoration problem considered and it will be given in each subsection.

4.2 Gray-scale white Gaussian noise denoising

Experimental setting – For this first set of experiments, we assume that $\mathcal{D}_\alpha$ models white Gaussian noise with standard deviation denoted $\alpha > 0$. In the context of gray-scale denoising, $M = 1$, $L = N$ and $A_1 \equiv \mathbb{I}_N$. As commonly used in image restoration when Gaussian noise is involved, the data-term is a squared Euclidean norm, i.e. $\mathcal{L}(\mathbf{u}, \mathbf{z}) = \frac{1}{2}\|\mathbf{u} - \mathbf{z}\|_2^2$, which is bounded below, proper, l.s.c and semi-algebraic [27]. By definition of Ψ and since the finite sum of semi-algebraic functions is semi-algebraic, we deduce that Ψ satisfies Assumptions 1-*i)*, *ii)*, *iii)*. In addition, this particular choice of A implies that $\mathcal{L}(A., \mathbf{z})$ satisfies the boundedness assumption in Proposition 2.

Let us consider the ground truth image in Figure 2 (left), where the contours are obtained by binarization and computation of the gradients. In this section, we evaluate the denoising and contour detection performances obtained with the proposed D-MS performed with Algorithm 2, when the input corresponds to Figure 2 with an additive white Gaussian noise of standard deviation $\alpha \in \{0.2^2, 0.4^2\}$.

Regarding the algorithms step-size, we set c_k and d_k constant. We first compute $\nu\big(\mathbf{e}^{[k]}\big)$, assuming that $\mathbf{e}$ is not equal to 1 everywhere. This assumption is not restrictive in general, since its means that we do not have contours everywhere. We have $\nu\big(\mathbf{e}^{[k]}\big) = \beta(1 - \mathbf{e}^{[k]})^2\|D\|^2 \leq \beta\|D\|^2$, where the upper bound is attained when $\mathbf{e}^{[k]} \equiv 0$. Hence, we choose, for both PALM and SL-PAM,

$c_k \equiv 1.01 * \beta\|D\|^2$. In the other hand, $\varepsilon(\mathbf{u}^{[k+1]}) = \beta(1-\mathbf{e})\|D\mathbf{u}^{[k+1]}\|^2 \leq \beta(1-\mathbf{e})\|D\|^2\|\mathbf{u}^{[k+1]}\|^2$. If we normalize $\mathbf{z}$, then $\forall k \in \mathbb{N}$, $\|\mathbf{u}^{[k]}\|^2 \leq 1$. Thus we set, for PALM, $d_k \equiv 1.01 * \beta\|D\|^2$. Finally, for SL-PAM, we set $d_k \equiv 1.01 * \beta\|D\|^2 * 10^{-3}$. This choice will be discussed below (see Figure 5).

We compare the results obtained with various regularization terms: the ℓ_0 pseudo-norm, the ℓ_1 norm, and the quadratic-ℓ_1 penalization.

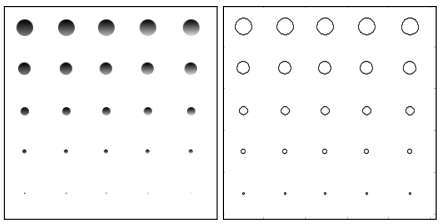


Figure 2: Ground truth image of size 256×256 with real contours.

Performances evaluation – The performances are evaluated in terms of signal-to-noise-ratio (SNR), Structural Similarity Index (SSIM) [41], and Jaccard index [42], for β varying in $[1, 50]$ and λ varying in $[0.0001, 0.9]$. The resulting scores are summarized in a map as the one displayed in Figure 3.

We performe the experiments on a 3.2GHz Intel Core i5 CPU, and stop when $|(\Psi(\mathbf{u}^{[k+1]}, \mathbf{e}^{[k+1]}) - \Psi(\mathbf{u}^{[k]}, \mathbf{e}^{[k]})| < 10^{-4}$.

The best performance with the quadratic-ℓ_1 penalization, according to each measure (SNR, SSIM and Jaccard index), is summarized in Table 1. It is displayed in Figure 6 for $\alpha = 0.2^2$ and in Figure 7 for $\alpha = 0.4^2$. From Figures 6 and 7, we first observe that the SNR and the SSIM lead to similar denoising and segmentation results. For small α, they do not allow us to extract a sparse 1-dimensional contour, while the result obtained from Jaccard index provides the best denoising and segmentation result. However, for strong noise, the SNR and the SSIM both outperform the Jaccard index for denoising purpose, with satisfying denoising and contour detection results.

Choice of $\mathcal{R}$ – From Tables 1 and 2, we notice that the best performances are obtained using either the ℓ_1 norm or the quadratic-ℓ_1 penalization. Since the latest provides the best segmentation results, we propose in the sequel to use the quadratic-ℓ_1 penalization together with the SSIM or Jaccard index, depending on the noise level.

Table 1: SL-PAM performances according to the SNR, the SSIM, and the Jaccard index.

α		ℓ_0	ℓ_1	$\ell_{1,Q}$
0.2^2	SNR	39.67	**41.79**	41.57
	SSIM	0.991	**0.995**	0.994
	Jaccard index	0.557	0.596	**0.605**
0.4^2	SNR	30.19	**30.6**	30.58
	SSIM	0.929	0.933	**0.935**
	Jaccard index	0.396	0.475	**0.494**

Sensitivity to the initialization – We propose to evaluate the robustness of the proposed

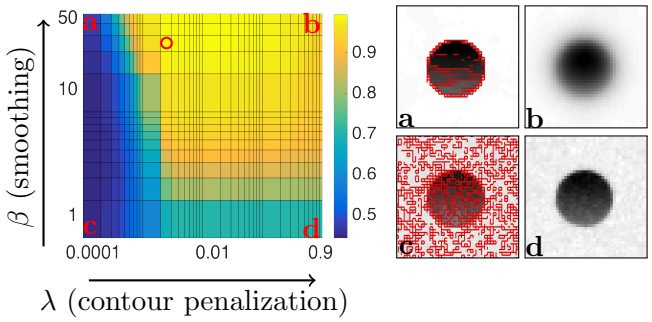


Figure 3: Example of score map with corresponding results on the right. Contours are delineated in red. The red circle on the map represents the best score. We can observe that larger β leads to a smoother estimate and that a larger value of λ implies less contours.

Table 2: SL-PAM computational times.

α		ℓ_0	ℓ_1	$\ell_{1,Q}$
0.2^2	SNR	**0.46**s	1.98s	3.24s
	SSIM	**0.93**s	1.74s	3.24s
	Jaccard index	**0.06**s	0.09s	0.49s
0.4^2	SNR	**0.91**s	1.67s	4.50s
	SSIM	**0.83**s	1.54s	2.86s
	Jaccard index	**0.15**s	0.25s	0.59s

algorithmic scheme with respect to the initialization. We compare different choices for $\mathbf{u}^{[0]}$: $\mathbf{u}^{[0]} = \mathbf{z}$, $\mathbf{u}^{[0]} \sim \mathcal{N}(0, \mathbb{I}_N)$ and $\mathbf{u}^{[0]} \equiv \zeta_{\mathbf{u}} \in [\min(\mathbf{z}), \max(\mathbf{z})]$. Similarly, we propose to deal with either $\mathbf{e}^{[0]} \equiv \{0,1\}^{|\mathbb{E}|}$, $\mathbf{e}^{[0]} \sim \mathcal{B}(0.5)$ or $\mathbf{e}^{[0]} \equiv \zeta_{\mathbf{e}} \in (0,1)$. We show the mean convergence results for 10 realizations in Figure 4, and we observe that the best initializing pair for Gaussian denoising is $(\mathbf{u}^{[0]}, \mathbf{e}^{[0]}) = (\mathbf{z}, 1^{|\mathbb{E}|})$. Notice that, whatever the initialization, all the run converge to the same value, which leads to a robust estimation, despite the resolution of a nonconvex problem.

SL-PAM versus PALM – We now compare in Figure 5 the performances of the PALM algorithm 1, to those of our SL-PAM algorithm 2, with decreasing $d_k \in 1.01 * \beta \|D\|^2 * \{1, 10^{-1}, 10^{-2}, 10^{-3}\}$, and a quadratic-$\ell_1$ regularization. We first notice that PALM and SL-PAM converge to the same minimum. In particular, they converge the same way when the descent parameters c_k and d_k are identically chosen for both of them. Nonetheless, SL-PAM outperforms the PALM algorithm for d_k set such that $\delta < 1$.

4.3 Color denoising and comparisons with state-of-the-art methods

In this section, we propose to perform RGB color image denoising involving white Gaussian noise. In this case, we consider $\mathbf{u} = (\mathbf{u}_R, \mathbf{u}_G, \mathbf{u}_B)$, $M = 3$ and $\mathbf{e} \in \mathbb{R}^{|\mathbb{E}|}$ common to the three components of $\mathbf{u}$. We compare the proposed method with state-of-the-art approaches, including TV minimization, the MS relaxation proposed in [20] (disabling the GPU implementation for computational time comparison) and the Discrete AT formulation [25]. Since the TV minimization does not allow us

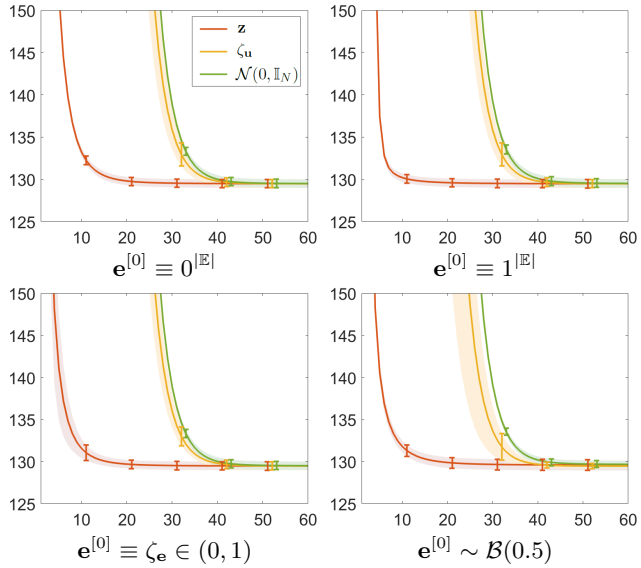


Figure 4: Performances of SL-PAM with different initial values of $\mathbf{u}^{[0]}$ and $\mathbf{e}^{[0]}$, when the input is the image in Fig. 2, degraded by additive white Gaussian noise with standard deviation $\alpha = 0.2^2$. SL-PAM is not sensitive to the initalization.

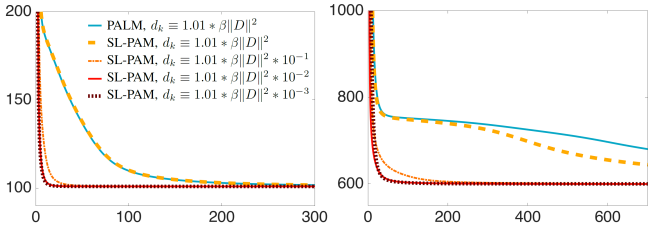
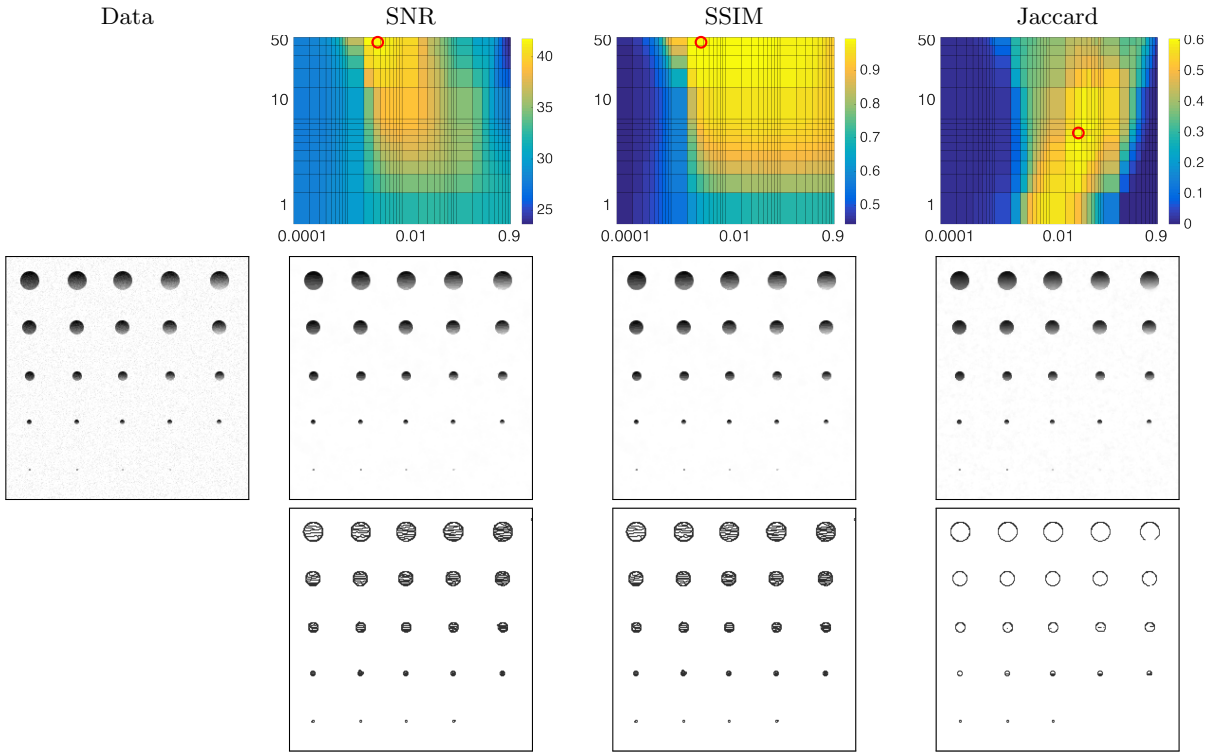


Figure 5: Comparison of PALM and SL-PAM convergence rates, with fixed c_k identically choosen for both of them, and decreasing d_k for SL-PAM, when the input data is the image in Fig. 2, degraded by additive white Gaussian noise with standard deviantion $\alpha = 0.2^2$ (left), and $\alpha = 0.4^2$ (right).

to directly extract the contours, we compute them by thresholding the gradient of the estimate, and we do not include this method in the scores' comparison. The best results according to SNR and SSIM are presented in Figures 8. We do not provide the Jaccard index since the real contour in this case is hard to define.

Despite good SNR and SSIM results, we observe the typical staircasing artifacts using the TV minimization, which are visually disturbing and unpleasant, while the other ones provide piecewise smooth results, and poor contour detection. The D-MS allows us to extract 1-dimensional contours similar to those obtained with the MS relaxation and the Discrete AT, but with some additional contours detected, especially in the necklace or around the lips. Regarding computational time, D-MS outperforms the Discrete AT method. The MS relaxation relies on an very efficient algorithm (even with the CPU implementation), but without convergence guarantees.



Data	SNR	SSIM	Jaccard
SNR = 30.61 dB SSIM = 0.681	**SNR = 41.57 dB** SSIM = 0.994 Jaccard = 0.323 Time = 3.24s	SNR = 41.57 dB **SSIM = 0.994** Jaccard = 0.323 Time = 3.24s	SNR = 37.48 dB SSIM = 0.944 **Jaccard = 0.605** Time = 0.49s

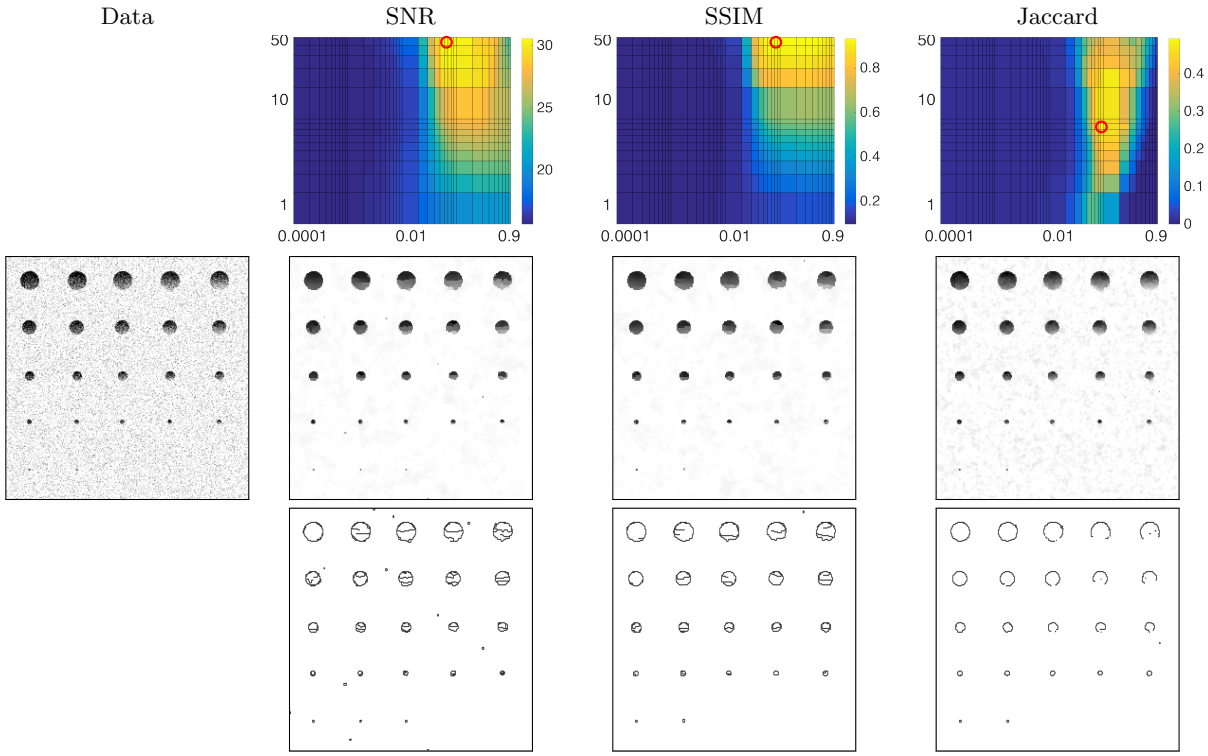
Figure 6: Denoising with SL-PAM and the quadratic-ℓ_1 penalization on the image in Fig. 2, degraded by additive white Gaussian noise with standard deviation $\alpha = 0.2^2$. The best results for each score are presented.

4.4 Poisson denoising and image restoration

Since Problem 1 allows us to deal with more complex data fidelity terms, we propose here to illustrate the results obtained when (i) data are corrupted by Poisson noise and (ii) data are degraded by both a blur and Gaussian noise. Since the experiments in Section 4 showed that the proposed approach outperforms the TV minimization, we do not present TV results in the following.

Poisson denoising – The choice of the Kullback-Leibler divergence $\mathcal{L}(\mathbf{u}, \mathbf{z}) = \sum_m \mathrm{DKL}(\mathbf{u}_m, \mathbf{z}_m)$ fits data corrupted by Poisson noise [43, 32, 33]. This data-term is bounded below and l.s.c. Thus Ψ satisfies Assumption 1-*i)*, *ii)*, *iii)*.

We first consider the image in Figure 2 corrupted by a Poisson noise with parameter $\sigma = 100$. The best results according to (SNR,SSIM,Jaccard index) using the quadratic-ℓ_1 regularization are shown in Figure 9. In Figure 10, we present the Poisson denoising results of a real image with the quadratic-ℓ_1 regularization. The performances are comparable with Gaussian denoising, with higher computational time, due to the use of the Kullback-Leibler divergence.



SNR = 18.68 dB
SSIM = 0.173

SNR = 30.58 dB
SSIM = 0.928
Jaccard = 0.361
Time = 4.5s

SNR = 30.43 dB
SSIM = 0.935
Jaccard = 0.380
Time = 2.86s

SNR = 27.45 dB
SSIM = 0.568
Jaccard = 0.494
Time = 0.59s

Figure 7: Denoising with SL-PAM and the quadratic-ℓ_1 penalization on the image in Fig. 2, degraded by additive white Gaussian noise with standard deviation $\alpha = 0.4^2$. The best results for each score are presented.

Image restoration – We propose to discuss the potential of the SL-PAM algorithm for image restoration tasks. In presence of blur, the data fidelity term depends on the blur matrix A, and reads: $\mathcal{L}(A\mathbf{u}, \mathbf{z}) = \frac{1}{2}\|A\mathbf{u} - \mathbf{z}\|_2^2$. In our experiments, we consider a Gaussian blur of size $Q \times Q$ and standard deviation σ, and additive white Gaussian noise, with standard deviation α. This type of degradation allows us to ensure the boundedness assumption in Proposition 2. Figure 11 displays the restoration results on the image in Figure 2, when $\alpha = 0.2$ and $Q = 7$. Restoration results on a real image are presented in Figure 12, with $\alpha = 0.2$ and $Q = 7$. Except for the best results according to the SNR, we observe that the method is able to detect sharp contours and to recover thin structures.

5 Conclusion

In this work, we propose 1) a new discrete formulation of the MS functional, and 2) a new proximal algorithm, with proved convergence, to solve it. Numerical experiments show that the proposed method is able to detect sharp contours and to reconstruct piecewise smooth approximations with low computational cost, and that it is competitive with state-of-the-art approaches. The influence

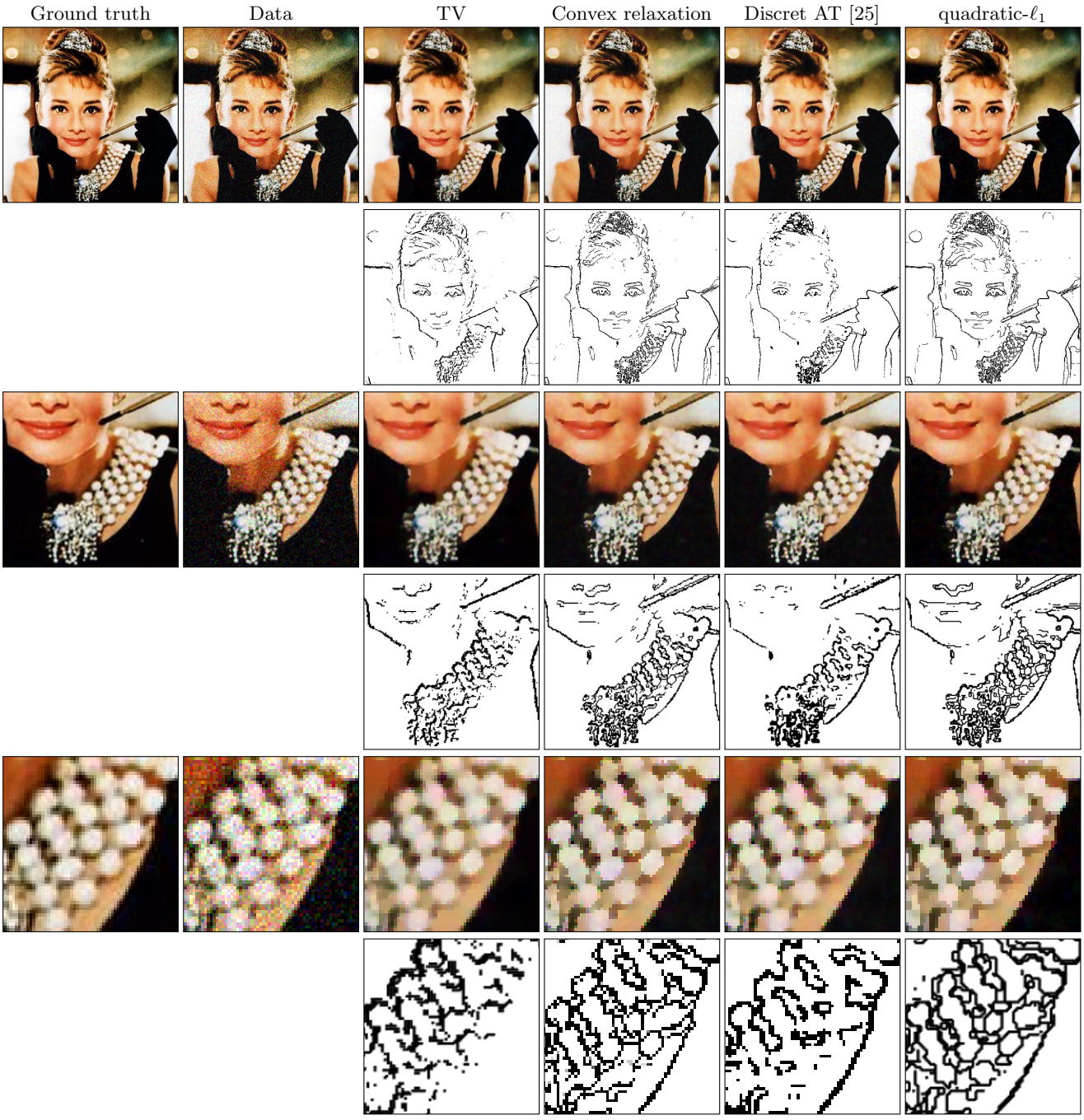
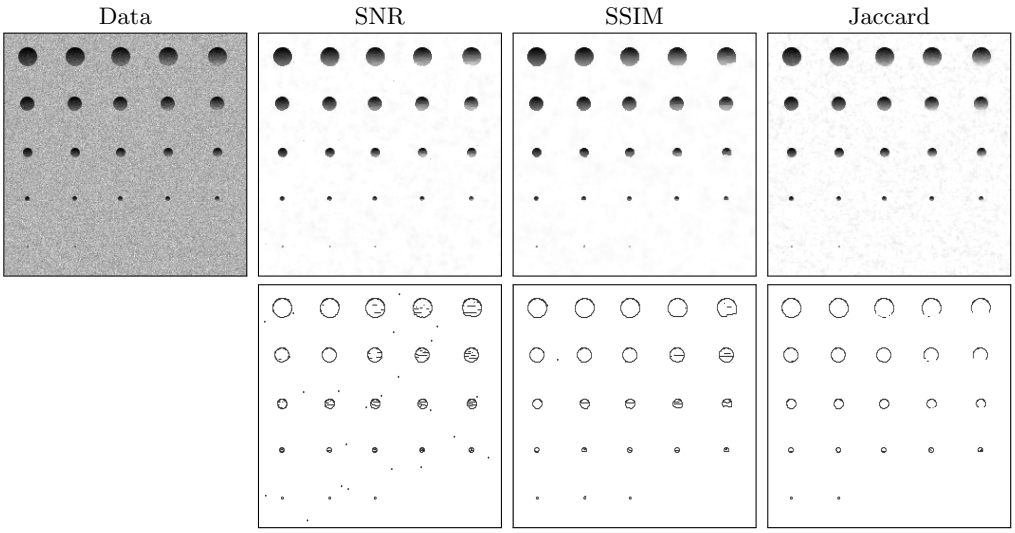


Figure 8: Comparison according to the SNR of color denoising performances, involving white Gaussian noise with standard deviation $\alpha = 0.1$, with state-of-the-art methods, from left to right: TV, the MS relaxation [20], the Discrete AT [25], and the proposed method. Regarding the scores comparison, we do not include the TV minimization since it does not perform joint restoration and segmentation.



SNR = 10.20 dB	**SNR = 34.32 dB**	SNR = 33.84 dB	SNR = 31.46 dB
SSIM = 0.222	SSIM = 0.942	**SSIM = 0.972**	SSIM = 0.741
	Jaccard = 0.488	Jaccard = 0.528	**Jaccard = 0.588**
	Time = 22.47s	Time = 41.89s	Time = 5.04s

Figure 9: Denoising with quadratic-ℓ_1 penalization with SL-PAM on the image in Fig. 2 degraded by Poisson noise with parameter $\alpha = 100$. The best results for each score are presented.

of the choice of the regularization parameters with respect to different performance measures is also provided. The proposed SL-PAM algorithm could be useful for other tasks, e.g. nonnegative matrix factorization.

6 Appendix

6.1 Proof of Lemma 1

(i) Let $k \geq 0$. Applying [27, Lemma 3.2] with $h = \mathcal{S}\big(\mathbf{e}^{[k]}, \cdot\big)$, $\sigma = \mathcal{L}$ and $t = c_k$ we obtain:

$$\begin{aligned}
&\mathcal{S}\big(\mathbf{e}^{[k]}, \mathbf{u}^{[k+1]}\big) + \mathcal{L}(\mathbf{u}^{[k+1]}) \\
\leq &\mathcal{S}\big(\mathbf{e}^{[k]}, \mathbf{u}^{[k]}\big) + \mathcal{L}(\mathbf{u}^{[k]}) - \frac{1}{2}(c_k - \nu(\mathbf{e}^{[k]}))\|\mathbf{u}^{[k+1]} - \mathbf{u}^{[k]}\|^2 \qquad (19)\\
\leq &\mathcal{S}\big(\mathbf{e}^{[k]}, \mathbf{u}^{[k]}\big) + \mathcal{L}(\mathbf{u}^{[k]}) - \frac{1}{2}(\gamma_k - 1)\nu(\mathbf{e}^{[k]})\|\mathbf{u}^{[k+1]} - \mathbf{u}^{[k]}\|^2 \qquad (20)
\end{aligned}$$

with $c_k = \gamma_k \nu(\mathbf{e}^{[k]})$. On the other hand, the update of $\mathbf{e}^{[k+1]}$ can be written

$$\mathbf{e}^{[k+1]} \in \operatorname*{argmin}_{\mathbf{e}} \frac{d_k}{2}\|\mathbf{e} - \mathbf{e}^{[k]}\|_2^2 + \lambda\mathcal{R}(\mathbf{e}) + \beta\mathcal{S}(\mathbf{e}, \mathbf{u}^{[k+1]}), \qquad (21)$$

leading to

$$\lambda\mathcal{R}(\mathbf{e}^{[k+1]}) + \beta\mathcal{S}\big(\mathbf{e}^{[k+1]}, \mathbf{u}^{[k+1]}\big) + \frac{d_k}{2}\|\mathbf{e}^{[k+1]} - \mathbf{e}^{[k]}\|^2$$

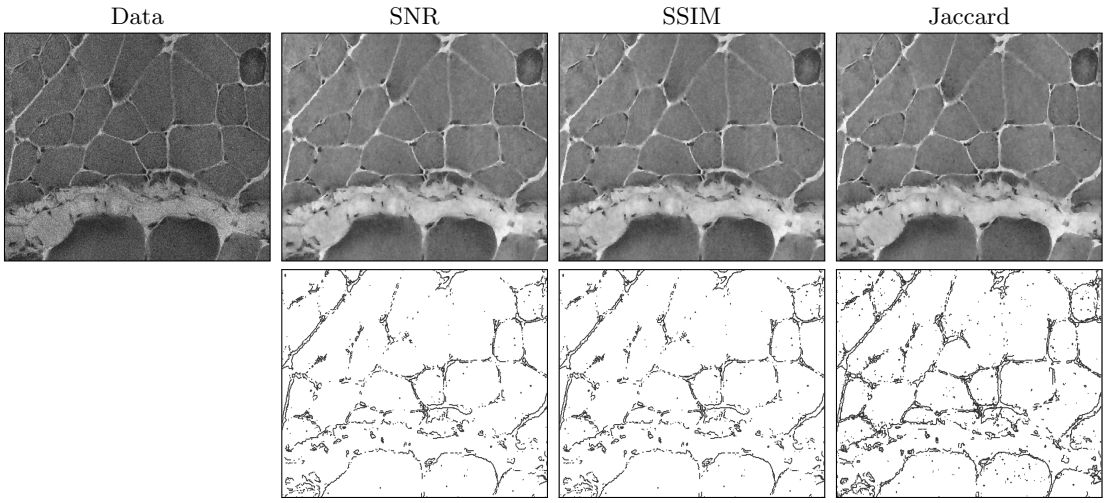


Figure 10: Denoising with quadratic-ℓ_1 penalization with SL-PAM on a muscle image degraded by Poisson noise with parameter $\alpha = 100$. The best results for each score are presented.

$$\leq \lambda \mathcal{R}(\mathbf{e}^{[k]}) + \beta \mathcal{S}\big(\mathbf{e}^{[k]}, \mathbf{u}^{[k]}\big) \tag{22}$$

Hence, combining (20) and (22), we get

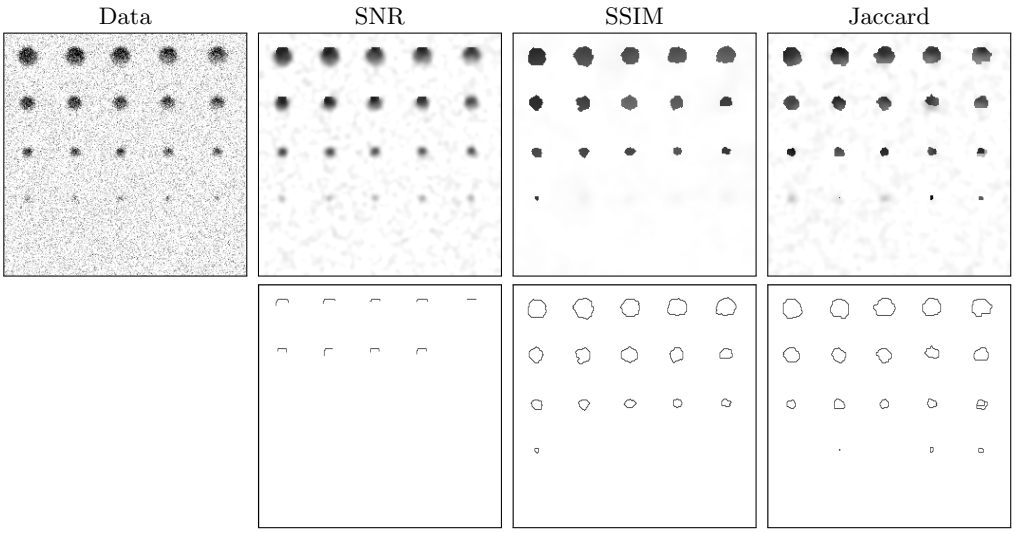
$$\begin{aligned}
&\Psi(\mathbf{x}^{[k]}) - \Psi(\mathbf{x}^{[k+1]}) && (23)\\
&= \mathcal{L}(\mathbf{u}^{[k]}) + \beta \mathcal{S}\big(\mathbf{e}^{[k]}, \mathbf{u}^{[k]}\big) + \lambda \mathcal{R}(\mathbf{e}^{[k]}) \\
&\quad - \mathcal{L}(\mathbf{u}^{[k+1]}) - \beta \mathcal{S}\big(\mathbf{e}^{[k+1]}, \mathbf{u}^{[k+1]}\big) - \lambda \mathcal{R}(\mathbf{e}^{[k+1]}) && (24)\\
&\geq \mathcal{L}(\mathbf{u}^{[k+1]}) + \beta \mathcal{S}\big(\mathbf{e}^{[k]}, \mathbf{u}^{[k+1]}\big) + \lambda \mathcal{R}(\mathbf{e}^{[k]}) \\
&\quad - \mathcal{L}(\mathbf{u}^{[k+1]}) - \beta \mathcal{S}\big(\mathbf{e}^{[k]}, \mathbf{u}^{[k+1]}\big) - \lambda \mathcal{R}(\mathbf{e}^{[k]}) \\
&\quad + \frac{1}{2}(\gamma_k - 1)\nu(\mathbf{e}^{[k]})\|\mathbf{u}^{[k+1]} - \mathbf{u}^{[k]}\|^2 + \frac{d_k}{2}\|\mathbf{e}^{[k+1]} - \mathbf{e}^{[k]}\|^2 && (25)\\
&\geq \frac{\rho 1}{2}\|\mathbf{x}^{[k+1]} - \mathbf{x}^{[k]}\|^2. && (26)
\end{aligned}$$

Combined with Assumptions 1-*iv)*, *v)*, it proves the result.

(ii) Since Ψ is bounded from below, Ψ converges to some $\underline{\Psi} \in \mathbb{R}$. Let now $N \in \mathbb{N}^*$. It follows from *(i)* that

$$\begin{aligned}
\sum_{k=0}^{N-1} \|\mathbf{x}^{[k+1]} - \mathbf{x}^{[k]}\|^2 &\leq \frac{2}{\rho_1}(\Psi(\mathbf{x}^{[0]}) - \Psi(\mathbf{x}^{[N]})) && (27)\\
&\leq \frac{2}{\rho_1}(\Psi(\mathbf{x}^{[0]}) - \underline{\Psi}) < \infty. && (28)
\end{aligned}$$

We conclude taking the limit as $N \to \infty$.



SNR = 23.90 dB	**SNR = 24.72 dB**	SNR = 24.04 dB	SNR = 24.12 dB
SSIM = 0.059	SSIM = 0.765	**SSIM = 0.941**	SSIM = 0.860
	Jaccard = 0.061	Jaccard = 0.146	**Jaccard = 0.188**
	Time = 3.5s	Time = 17.54s	Time = 8.67s

Figure 11: Image restoration with quadratic-ℓ_1 penalization on the image in Fig. 2 degraded by additive white Gaussian noise with standard deviation $\alpha = 0.2$, and a Gaussian blurring filter of size 7×7 and standard deviation $\sigma = 2$. The best results for each score are presented.

6.2 Proof of Lemma 2

Writing down the optimality conditions for the iterative steps of Algorithm 2, we get:

$$\beta\nabla_{\mathbf{u}}\mathcal{S}\big(\mathbf{e}^{[k-1]},\mathbf{u}^{[k-1]}\big)+c_{k-1}(\mathbf{u}^{[k]}-\mathbf{u}^{[k-1]})+v^{[k]}=0, \tag{29}$$

where $v^{[k]}\in\partial\mathcal{L}(\mathbf{u}^{[k]})$, and

$$\beta\nabla_{\mathbf{e}}\mathcal{S}\big(\mathbf{e}^{[k]},\mathbf{u}^{[k]}\big)+d_{k-1}(\mathbf{e}^{[k]}-\mathbf{e}^{[k-1]})+\xi^{[k]}=0, \tag{30}$$

where $\xi^{[k]}\in\partial(\lambda\mathcal{R}(\mathbf{e}^{[k]}))$.

Subdifferential property [27, Proposition 2.1] allows us to state that $\beta\nabla_{\mathbf{u}}\mathcal{S}\big(\mathbf{e}^{[k]},\mathbf{u}^{[k]}\big)+v^{[k]}\in\partial_{\mathbf{u}}\Psi(\mathbf{u}^{[k]},\mathbf{e}^{[k]})$ and $\beta\nabla_{\mathbf{e}}\mathcal{S}\big(\mathbf{e}^{[k]},\mathbf{u}^{[k]}\big)+\xi^{[k]}\in\partial_{\mathbf{e}}\Psi(\mathbf{u}^{[k]},\mathbf{e}^{[k]})$, and hence $(A^k_{\mathbf{u}},A^k_{\mathbf{e}})\in\partial\Psi(\mathbf{u}^{[k]},\mathbf{e}^{[k]})$.

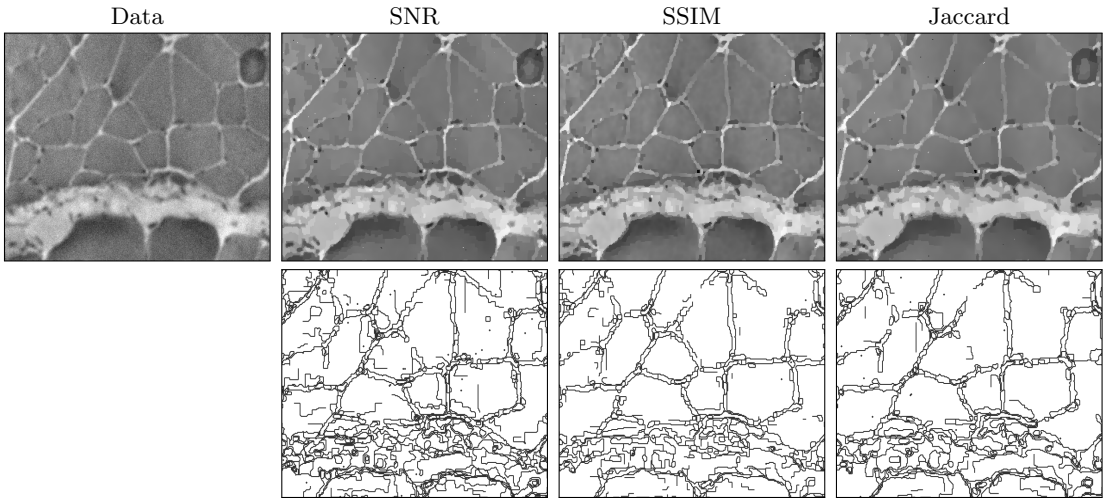
Combining Assumption 1-*iii)* with the assumption of Lipschitz continuity of $\nabla\mathcal{S}$, and following arguments in [27, Lemma 3.4], we can prove that there exists $M>0$ such that

$$\|A^k_{\mathbf{u}}\|\leq(2M+\gamma_k\nu^+)\|\mathbf{x}^{[k]}-\mathbf{x}^{[k-1]}\|. \tag{31}$$

On the other hand,

$$\|A^k_{\mathbf{e}}\|=d_{k-1}\|\mathbf{e}^{[k-1]}-\mathbf{e}^{[k]}\|\leq d^+\|\mathbf{x}^{[k]}-\mathbf{x}^{[k-1]}\|. \tag{32}$$

Summing up (31) and (32) we obtain the desired result with $\rho_2=\gamma_k\nu^++d^+$.



SNR = 18.10 dB SSIM = 0.394	**SNR = 19.47 dB** SSIM = 0.578 Jaccard = 0.135 Time = 150.2s	SNR = 19.34 dB **SSIM = 0.589** Jaccard = 0.135 Time = 42.1s	SNR = 19.34 dB SSIM = 0.573 **Jaccard = 0.141** Time = 121.9s

Figure 12: Image restoration with quadratic-ℓ_1 penalization on the image in Fig. 2 degraded by additive white Gaussian noise with standard deviation $\alpha = 0.2$ and a Gaussian blurring filter of size 7×7 and standard deviation $\sigma = 2$. The best results for each score are presented.

6.3 Proof of Proposition 4

Let $\eta \in \mathbb{R}$. One has:

$$\operatorname{prox}_{\tau \max\left\{|.|, \frac{.^2}{4\epsilon}\right\}}(\eta) = \operatorname*{argmin}_x \frac{1}{2}\|x-\eta\|_2^2 + \tau \max\left\{|x|, \frac{x^2}{4\epsilon}\right\} \tag{33}$$

One must split cases:

- If $|x| \leq 4\epsilon$, then $\max\left\{|x|, \dfrac{x^2}{4\epsilon}\right\} = |x|$ and we have:

$$\begin{aligned}
\operatorname{prox}_{\tau \max\left\{|.|, \frac{.^2}{4\epsilon}\right\}}(\eta) &= \operatorname*{argmin}_x \frac{1}{2}\|x-\eta\|_2^2 + \tau|x| && (34)\\
&= \operatorname{prox}_{\tau|.|}(\eta) && (35)\\
&= \operatorname{sign}(\eta)\ \max(0, |\eta| - \tau) && (36)
\end{aligned}$$

  when $|\eta| \leq 4\epsilon + \tau$.

- If $|x| > 4\epsilon$, then $\max\left\{|x|, \dfrac{x^2}{4\epsilon}\right\} = \dfrac{x^2}{4\varepsilon}$:

$$\begin{aligned}
\operatorname{prox}_{\tau \max\left\{|.|, \frac{.^2}{4\epsilon}\right\}}(\eta) &= \operatorname*{argmin}_x \frac{1}{2}\|x-\eta\|_2^2 + \tau\frac{x^2}{4\epsilon} && (37)\\
&= \operatorname{prox}_{\frac{\tau}{2\epsilon}x^2}(\eta) && (38)
\end{aligned}$$

$$= \operatorname{sign}(\eta)\ \frac{|\eta|}{\frac{\tau}{2\epsilon}+1} \tag{39}$$

when $|\eta| > 4\epsilon + 2\tau$.

Finally, we obtain

$$\operatorname{prox}_{\tau \max\left\{|.|, \frac{.^2}{4\epsilon}\right\}}(\eta) = \begin{cases} \operatorname{sign}(\eta)\ \max(0, |\eta| - \tau) & \text{if } |\eta| < 4\epsilon + \tau, \\ 4\epsilon & \text{if } 4\epsilon + \tau \leq |\eta| \leq 4\epsilon + 2\tau, \\ \operatorname{sign}(\eta)\ \dfrac{|\eta|}{\frac{\tau}{2\epsilon}+1} & \text{if } |\eta| > 4\epsilon + 2\tau. \end{cases} \tag{40}$$

□

6.4 Proof of Proposition 3

For every $i \in \{1, ..., |\mathbb{E}|\}$,

$$= \operatorname*{argmin}_{\mathbf{e}} \frac{\lambda}{d_k}\sigma_i(\mathbf{e}) + \frac{\beta}{d_k}(1-\mathbf{e})^2\big(D\mathbf{u}^{[k+1]}\big)_i^2 + \frac{1}{2}\big(\mathbf{e} - \mathbf{e}_i^{[k]}\big)^2 \tag{41}$$

$$= \operatorname*{argmin}_{\mathbf{e}} \frac{\lambda}{d_k}\sigma_i(\mathbf{e}) + \frac{\beta}{d_k}(1 - 2\mathbf{e} + \mathbf{e}^2)\big(D\mathbf{u}^{[k+1]}\big)_i^2 \tag{42}$$

$$+ \frac{1}{2}\Big[\mathbf{e}^2 - 2\mathbf{e}\mathbf{e}_i^{[k]} + \big(\mathbf{e}_i^{[k]}\big)^2\Big] \tag{43}$$

$$= \operatorname*{argmin}_{\mathbf{e}} \frac{\lambda}{2\beta\big(D\mathbf{u}^{[k+1]}\big)_i^2 + d_k}\sigma_i(\mathbf{e}) \tag{44}$$

$$+ \frac{1}{2}\left(\mathbf{e} - \frac{\beta\big(D\mathbf{u}^{[k+1]}\big)_i^2 + \frac{d_k\mathbf{e}_i^{[k]}}{2}}{\beta g_i + \frac{d_k}{2}}\right)^2, \tag{45}$$

which concludes the proof. □

References

[1] 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.

[2] R. Gordon, R. Bender, and G. T. Herman, "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography," *Journal of Theoretical Biology*, vol. 29, no. 3, pp. 471–481, Dec. 1970.

[3] N. Dey, L. Blanc-Féraud, C. Zimmer, Z. Kam, P. Roux, J.C. Olivo-Marin, and J. Zerubia, "Richardson-Lucy algorithm with total variation regularization for 3D confocal microscope deconvolution," *Microscopy Research Technique*, vol. 69, pp. 260–266, 2006.

[4] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comput. Assist. Tomo.*, vol. 8, no. 2, pp. 306–316, Apr. 1984.

[5] J. Zheng, S. S. Saquib, K. Sauer, and C. A Bouman, "Parallelizable bayesian tomography algorithms with rapid, guaranteed convergence," *IEEE Trans. Image Process.*, vol. 9, no. 10, pp. 1745–1759, Oct. 2000.

[6] J. Verhaeghe, D. Van De Ville, I. Khalidov, Y. D'Asseler, I. Lemahieu, and M. Unser, "Dynamic PET reconstruction using wavelet regularization with adapted basis functions," *IEEE Trans. Med. Imag.*, vol. 27, no. 7, pp. 943–959, July 2008.

[7] N. Pustelnik, C. Chaux, J.-C. Pesquet, and C. Comtat, "Parallel algorithm and hybrid regularization for dynamic PET reconstruction," in *IEEE Med. Imag. Conf.*, Knoxville, Tennessee, 30, Oct - 6, Nov. 2010, 5p.

[8] S. Huang, J. Xiang, H. Du, and X. Cao, "Inverse problems in atmospheric science and their application," *Journal of Physics: Conference Series*, vol. 12, no. 1, pp. 4557, 2005.

[9] A. Tikhonov, "Tikhonov regularization of incorrectly posed problems," *Soviet Mathematics Doklady*, vol. 4, pp. 1624–1627, 1963.

[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] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds., pp. 185–212. Springer-Verlag, New York, 2011.

[12] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, New York, 2011.

[13] N. Pustelnik, A. Benazza-Benhayia, Y. Zheng, and J.-C. Pesquet, "Wavelet-based image deconvolution and reconstruction," *Wiley Encyclopedia of EEE*, 2016.

[14] M. Storath, A. Weinmann, and L. Demaret, "Jump-sparse and sparse recovery using Potts functionals," *IEEE Trans. Signal Process.*, vol. 62, no. 14, pp. 3654–3666, July 2014.

[15] 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.

[16] 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.

[17] L. Condat, "Discrete total variation: New definition and minimization," *SIAM J. Imaging Sci.*, vol. 10, no. 3, pp. 1258–1290, 2017.

[18] L. Xu, C. Lu, Y. Xu, and J. Jia, "Image smoothing via L0 gradient minimization," in *ACM Transactions on Graphics (TOG)*. ACM, 2011, vol. 30, p. 174.

[19] M. Storath, A. Weinmann, J. Frikel, and M. Unser, "Joint image reconstruction and segmentation using the Potts model," *Inverse Problems*, vol. 31, no. 2, pp. 025003, 2015.

[20] 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.

[21] C. Zach and C. Häne, "Discretized convex relaxations for the piecewise smooth Mumford-Shah model," in *Int. Conf. on Energy Minimization Methods in Computer Vision and Pattern Recognition*, 2017.

[22] T.F. Chan, S. Esedoglu, and M. Nikolova, "Algorithms for finding global minimizers of image segmentation and denoising models," *SIAM J. App. Math.*, vol. 66, no. 5, pp. 1632–1648, 2006.

[23] X. Bresson, S. Esedolu, P. Vandergheynst, J.-P. Thiran, and S. Osher, "Fast global minimization of the active contour/snake model," *J. Math. Imag. Vis.*, vol. 28, no. 2, pp. 151–167, 2007.

[24] Z. Li, F. Malgouyres, and T. Zeng, "Regularized non-local total variation and application in image restoration," *J. Math. Imag. Vis.*, vol. 59, no. 2, pp. 296–317, 2017.

[25] 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.

[26] M. Foare, J.-O. Lachaud, and H. Talbot, "Numerical implementation of the Ambrosio–Tortorelli functional using discrete calculus and application to image restoration and inpainting," in *Proc. 1st Workshop on Reproducible Research in Pattern Recognition (RRPR2016)*, Cancun, Mexico, 2016.

[27] 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.

[28] 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.

[29] L. Bar, N. Sochen, and N. Kiryati, "Semi-blind image restoration via Mumford–Shah regularization," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 483–493, 2006.

[30] L. Bar, B. Berkels, M. Rumpf, and G. Sapiro, "A variational framework for simultaneous motion estimation and restoration of motion-blurred video," in *Proc. IEEE Int. Conf. Comput. Vis.* IEEE, 2007, pp. 1–8.

[31] M. Jung, X. Bresson, T.F. Chan, and L.A. Vese, "Nonlocal Mumford–Shah regularizers for color image restoration," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1583–1598, 2011.

[32] N. Pustelnik, C. Chaux, and J.-C. Pesquet, "Hybrid regularization for data restoration in the presence of Poisson noise," in *Proc. Eur. Sig. Proc. Conference*. IEEE, 2009, pp. 1394–1398.

[33] G. Steidl and T. Teuber, "Removing multiplicative noise by Douglas-Rachford splitting methods," *J. Math. Imag. Vis.*, vol. 36, no. 2, pp. 168–184, 2010.

[34] 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.

[35] 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.

[36] L. Bar, T.F. Chan, G. Chung, M. Jung, N. Kiryati, R. Mohieddine, N. Sochen, and L.A. Vese, "Mumford and Shah model and its applications to image segmentation and image restoration," in *Handbook of mathematical methods in imaging*, pp. 1095–1157. Springer, 2011.

[37] J. Shah, "A common framework for curve evolution, segmentation and anisotropic diffusion," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 1996, pp. 136–142.

[38] R. Alicandro, A. Braides, and J. Shah, "Free-discontinuity problems via functionals involving the L1-norm of the gradient and their approximations," *Interfaces and free boundaries*, vol. 1, no. 1, pp. 17–37, 1999.

[39] P.L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-point algorithms for inverse problems in science and engineering*, pp. 185–212. Springer, 2011.

[40] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Optimization with Sparsity-Inducing Penalties," *Found. Trends Mach. Learn.*, vol. 4, no. 1, pp. 1–106, 2012.

[41] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.

[42] P. Jaccard, "Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines," *Bull. Soc. Vaudoise Sci. Nat.*, vol. 37, pp. 241–272, 1901.

[43] S. Kullback and R. A. Leibler, "On information and sufficiency," *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.