



**HAL**  
open science

# On the primal and dual formulations of the discrete Mumford-Shah functional

Nelly Pustelnik

► **To cite this version:**

Nelly Pustelnik. On the primal and dual formulations of the discrete Mumford-Shah functional. IEEE ICASSP, Jun 2023, Rhodes (Grèce), Greece. hal-03912074v2

**HAL Id: hal-03912074**

**<https://hal.science/hal-03912074v2>**

Submitted on 16 Nov 2023

**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.

# ON THE PRIMAL AND DUAL FORMULATIONS OF THE DISCRETE MUMFORD-SHAH FUNCTIONAL

*Nelly Pustelnik*

ENSL, CNRS, Laboratoire de Physique, F-69342 Lyon, France.

## ABSTRACT

This work focuses on the discrete Mumford-Shah (D-MS) functional which aims to perform jointly image reconstruction and contour detection but at the price of minimizing a non-convex objective function. This functional was of main interest during the 90's but was then forsaken in order to focus on the unique restoration task relying on non-smooth convex minimization. Recent advances about D-MS were dedicated to alternative objective functions for which efficient numerical solution based on proximal iterations can be designed. In the 90's literature about D-MS, equivalences between primal and dual formulations were derived. However, in the framework obtained by more recent developments dedicated to D-MS such an equivalence was not yet derived and it is the goal of this work. By providing both a primal and dual formulation, a large panel of algorithms can be employed including recent proximal-based algorithms benefiting of good convergence behavior, especially due to KL properties and also most standard methods such as BFGS.

**Index Terms**— Image restoration, contour detection, Discrete Mumford-Shah, proximal algorithms, dual formulation.

## 1. INTRODUCTION

The resolution of an inverse problem in image processing consists in finding an image  $\hat{\mathbf{u}} \in \mathbb{R}^{|\Omega|}$  as close as possible from an (unknown true) image  $\bar{\mathbf{u}} \in \mathbb{R}^{|\Omega|}$  using information contained in the observation  $\mathbf{z} = \mathbf{A}\bar{\mathbf{u}} + \epsilon$  (where  $\mathbf{A} \in \mathbb{R}^{M \times |\Omega|}$  models a linear degradation and  $\epsilon$  denotes a random, possibly Gaussian, perturbation). This inversion task is generally a preliminary step before other processing such as segmentation or interface detection [1, 2]. In this work, we focus on the second type of post-processing when it is combined with the inversion step, that is of main interest for physics application, for instance to extract contact surfaces when studying multiphasic flows [3].

During the 90's, an important research effort was dedicated to joint image denoising and interface detection, starting with the contribution of Mumford and Shah (MS) [4] and

similar ideas formulated by Geman and Geman in a probabilistic and discrete setting [5]. The main advantages of these formulations is to favor piece-wise smooth images.

In [6], the authors proposed an extension of Geman and Geman energy expressed as

$$\min_{\substack{\mathbf{u} \in \mathbb{R}^{|\Omega|} \\ \mathbf{e} \in [0,1]^{|\mathbb{E}|}}} \frac{1}{2} \|\mathbf{A}\mathbf{u} - \mathbf{z}\|_2^2 + \frac{\lambda^2}{2} \sum_{\substack{s=(i,j) \\ s' \in \mathcal{N}_s}} \phi(u_s - u_{s'}) \varphi(e_{s,s'}) + \sum_{s,s'} \psi(e_{s,s'}), \quad (1)$$

where  $\mathbf{u} = (u_s)_{s \in \Omega}$  denotes the variable associated with the image to recover,  $\mathbf{e} = (e_{s,s'})_{(s,s') \in \mathbb{E}}$  is the variable modeling the edges located on the lattice  $\mathbb{E}$ ,  $\mathcal{N}_s$  denotes the neighboring pixels of the pixel located at  $s$ , and  $\phi$ ,  $\varphi$ , and  $\psi$  are functions from  $\mathbb{R}$  to  $] - \infty, +\infty]$ . The first term is the standard data-fidelity term, the second one aims to insure smoothness everywhere except at the location where  $\varphi(e_{s,s'}) = 0$ , and the third term penalizes the edges in order to control the length of the contours. Interestingly, when  $\phi$  models a quadratic form, and for specific choices of  $\psi$ , equivalences have been established between the so-called primal formulation (1) and the dual formulation (2) involving a penalization  $\tilde{\phi}$ :

$$\min_{\mathbf{u} \in \mathbb{R}^{|\Omega|}} \frac{1}{2} \|\mathbf{A}\mathbf{u} - \mathbf{z}\|_2^2 + \frac{\lambda^2}{2} \sum_{\substack{s=(i,j) \\ s' \in \mathcal{N}_s}} \tilde{\phi}(u_s - u_{s'}) \quad (2)$$

whose first term is still the data-term insuring that the estimate is close to the data  $\mathbf{z}$  and the second term is a penalization term over the difference between neighboring pixels. Table 1 lists several functions encountered in the literature [7, 8, 6, 9] allowing to establish this link between primal and dual formulations.

**Algorithmic procedure** – The benefit of this twofold primal or dual formulations is to offer a large panel of algorithmic strategies to perform joint restoration and contour detection. Geman and Reynold [6] performed minimization by stochastic relaxation with annealing (Gibbs sampling). In [9], a Gauss-Seidel alternating minimization strategy is designed. An acceleration based on conjugate gradient is provided in [10]. Recently, Storath and Weimann [11] provided several interesting scheme to deal with Blake-Zisserman or Potts model (i.e.  $\phi_\alpha = \|\cdot\|_0$ ) based on dynamic programming allowing them to provide a fast and convergent algorithmic solution to the global minimum in 1D. An extension to dimension 2 by having recourse to ADMM iterations is also provided, at the price of weaker convergence guarantees.

This work is supported by the ANR (Agence Nationale de la Recherche) from France ANR-19-CE48-0009 Multisc'In.

**Contributions and outline** – In this work, we focus on a more recent alternative formulation of (1), improving over the references listed before, inspired from the discrete Ambrosio-Tortorelli formulation provided in [12, 13, 14] which is described in Problem 1 (Section 2) and for which we establish the dual formulation. By providing both a primal and dual formulation, a large panel of algorithms can be employed, thus Section 3 will discuss the pros and cons of both formulations through the optimization point of view and lists algorithms including recent proximal-based algorithms benefiting of good convergence behavior, especially due to KL properties but also most standard methods. Section 4 illustrates the numerical behavior of the algorithms for minimizing either the primal or the dual formulation. A BFGS strategy is also implemented and compared in order to illustrate the benefit to go beyond first order schemes. Conclusions are provided in Section 6. Proof are detailed in Section 7.

## 2. RELATION BETWEEN PRIMAL AND DUAL FORMULATIONS.

The primal minimization formulation considered in this work is provided in Problem 1.

**Problem 1** Let  $\mathbf{A} \in \mathbb{R}^{M \times |\Omega|}$  and  $\mathbf{z} = \mathcal{D}(\mathbf{A}\mathbf{x}) \in \mathbb{R}^M$ , where  $\mathcal{D}$  denotes a stochastic perturbation<sup>1</sup>,  $\mathbf{D} \in \mathbb{R}^{|\mathbb{E}| \times |\Omega|}$  and such as the  $\ell$ -row of  $\mathbf{D}$  is denoted  $\mathbf{D}_\ell$ ,  $\lambda > 0$ , and  $\alpha > 0$ . We assume that  $f(\cdot, \mathbf{z}): \mathbb{R}^M \rightarrow ]-\infty, +\infty]$  is convex, continuous, and proper;  $\phi: \mathbb{R} \rightarrow ]-\infty, +\infty]$ , and  $\psi: \mathbb{R} \rightarrow ]-\infty, +\infty]$  are proper; l.s.c such that  $\inf \psi = 0$ ,  $\psi(0) = 0$  and  $\inf \phi = 0$ . We aim to solve

$$(\hat{\mathbf{u}}, \hat{\mathbf{e}}) \in \underset{\substack{\mathbf{u} \in \mathbb{R}^{|\Omega|}, \\ \mathbf{e} \in \mathbb{R}^{|\mathbb{E}|}}}{\text{Argmin}} f(\mathbf{A}\mathbf{u}, \mathbf{z}) + \frac{\lambda^2}{2} \sum_{\ell=1}^{|\mathbb{E}|} \phi(\mathbf{D}_\ell \mathbf{u}) (1 - e_\ell)^2 + \alpha \sum_{\ell} \psi(e_\ell). \quad (3)$$

$\lambda > 0$  and  $\alpha > 0$  are the regularization parameters. The main advantage of this alternative formulation of (1) relies on the implicit assumption that  $e_\ell \leq 1$  imposed by the term  $(1 - \cdot)^2$  and  $e_\ell \geq 0$  imposed by the definition of  $\psi$  allowing us to relax the dynamic range constraint on  $e_\ell$ .

Proposition 1 establishes the dual formulation of Problem 1 providing a new relation between  $\phi$ ,  $\psi$ , and  $\tilde{\phi}$ . The dual expression appears to be simply formulated by making use of proximity operator.

**Proposition 1** The minimization Problem 1 is equivalent to

$$\hat{\mathbf{u}} \in \underset{\mathbf{u} \in \mathbb{R}^{|\Omega|}}{\text{Argmin}} f(\mathbf{A}\mathbf{u}, \mathbf{z}) + \lambda^2 \sum_{\ell=1}^{|\mathbb{E}|} \tilde{\phi}_{\alpha/\lambda^2}(\phi(\mathbf{D}_\ell \mathbf{u})),$$

and  $\hat{\mathbf{e}} = (\hat{e}_\ell)_{\ell \in \mathbb{E}}$  with

$$\hat{e}_\ell = \begin{cases} \text{prox}_{\frac{\alpha}{\lambda^2 \phi(\mathbf{D}_\ell \hat{\mathbf{u}})}} \psi(1) & \text{if } \phi(\mathbf{D}_\ell \hat{\mathbf{u}}) > 0, \\ 0 & \text{otherwise,} \end{cases}$$

<sup>1</sup>not only restricted to additive Gaussian noise

	$\tilde{\phi}(t)$	$\psi(w)$	$\varphi(u)$
Geman&McClure	$\frac{t^2}{1+t^2}$	$w - 2\sqrt{w} + 1$	$u$
Hebert&Leavy	$\log(1+t^2)$	$w - \log w - 1$	$u$
Charbonnier et al.	$2\sqrt{1+t^2} - 2$	$w + \frac{1}{w} - 2$	$u$
Geman&Reynold	$\frac{-1}{1+ t }$	$\frac{w-3\sqrt{w}}{2}$	$\frac{u^{3/2}}{2(1-\sqrt{b})}$

**Table 1.** Choice of  $\tilde{\phi}$ ,  $\psi$  and  $\varphi$  encountered in the literature leading to the equivalence between (1) and (2) when  $\phi = (\cdot)^2$ .

where, for every  $\eta > 0$ ,

$$\tilde{\phi}_{\alpha/\lambda^2}(\eta) = \frac{\eta}{2} (1 - \text{prox}_{\frac{\alpha}{\lambda^2 \eta}} \psi(1))^2 + \frac{\alpha}{\lambda^2} \psi(\text{prox}_{\frac{\alpha}{\lambda^2 \eta}} \psi(1))$$

and 0 otherwise.

To facilitate the interpretation of the penalization  $\tilde{\phi}_{\alpha/\lambda^2}$ , we propose to focus on specific choices of  $\phi$  and  $\psi$ . Lemma 1 focuses on D-MS- $\ell_1$  proposed in [13] while Lemma 2 allows us to establish the link with Blake-Zisserman functional [15].

**Lemma 1** Let  $\psi = |\cdot|$  and  $\phi = |\cdot|^2$ , the minimization Problem 1 can be reformulated as

$$\begin{cases} \hat{\mathbf{u}} \in \underset{\mathbf{u} \in \mathbb{R}^{|\Omega|}}{\text{Argmin}} f(\mathbf{A}\mathbf{u}, \mathbf{z}) + \frac{\lambda^2}{2} \sum_{\ell=1}^{|\mathbb{E}|} \tilde{\phi}_{\alpha/\lambda^2}((\mathbf{D}_\ell \mathbf{u})^2), \\ (\forall \ell) \quad \hat{e}_\ell = \begin{cases} \text{prox}_{\frac{\alpha}{\lambda^2 (\mathbf{D}_\ell \hat{\mathbf{u}})^2} |\cdot|} (1) & \text{if } (\mathbf{D}_\ell \hat{\mathbf{u}})^2 > 0, \\ 0 & \text{otherwise,} \end{cases} \end{cases}$$

where

$$(\forall \eta \geq 0) \quad \tilde{\phi}_{\alpha/\lambda^2}(\eta) = \begin{cases} \frac{\alpha}{\lambda^2} (2 - \frac{\alpha}{\lambda^2 \eta}) & \text{if } \eta > \frac{\alpha}{\lambda^2}, \\ \eta & \text{if } 0 < \eta \leq \frac{\alpha}{\lambda^2}, \\ 0 & \text{if } \eta = 0. \end{cases}$$

**Lemma 2** When  $\psi(\eta) = \{1 \text{ if } \eta \neq 0 \text{ and } 0 \text{ if } \eta = 0\}$  (i.e.  $\sum_{\ell} \psi$  denotes the  $\ell_0$  penalization), the minimization Problem 1 can be reformulated as

$$\begin{cases} \hat{\mathbf{u}} \in \underset{\mathbf{u} \in \mathbb{R}^N}{\text{Argmin}} f(\mathbf{A}\mathbf{u}, \mathbf{z}) + \frac{\lambda^2}{2} \sum_{\ell} \min(\phi(\mathbf{D}_\ell \mathbf{u}), \frac{2\alpha}{\lambda^2}) \\ (\forall \ell \in \{1, \dots, |\mathbb{E}|\}) \quad \hat{e}_\ell = \begin{cases} 0 & \text{if } \phi(\mathbf{D}_\ell \hat{\mathbf{u}}) < \frac{2\alpha}{\lambda^2}, \\ 1 & \text{if } \phi(\mathbf{D}_\ell \hat{\mathbf{u}}) > \frac{2\alpha}{\lambda^2}, \\ [0, 1] & \text{if } \phi(\mathbf{D}_\ell \hat{\mathbf{u}}) = \frac{2\alpha}{\lambda^2}. \end{cases} \end{cases}$$

When  $\phi = |\cdot|^2$ , the resulting non-convex but smooth penalization is known as truncated- $\ell_2$  penalization, and can also refer to Blake-Zisserman functional [15].

## 3. MINIMIZATION STRATEGIES

In this section, we propose minimization algorithms dedicated to both primal and dual formulations in the context of

Lemma 1 when  $\phi = |\cdot|^2$  and  $\psi = |\cdot|$ . We assume that  $f \circ \mathbf{A}$  can be either activated considering its gradient (assuming  $L_f$ -smooth functions  $f$ ) or its proximity operator. Algorithmic strategies relying either on alternating minimization (for solving the primal problem) or gradient-based iterations (for the dual) are derived. All algorithmic procedures benefit from global convergence to a critical point considering Kurdyka-Łojasiewicz properties [16, 17, 18]. Proximal methods could be useful in order to either handle with non-smooth functions but also in order to make the algorithmic procedure faster than gradient based approaches when smooth functions are involved [?].

Both minimization problems either the primal or the dual ones involved non-convex objective functions with different levels of complexity. On the other hand, the primal formulation is non-convex, however when  $\phi = |\cdot|^2$  and  $\psi = |\cdot|$ , this problem turns out to be biconvex and is generally solved by an alternate minimization procedure. On the other hand, in the context of Lemma 1 and when  $f$  is a quadratic function, typically encountered in restoration problem with Gaussian noise, the dual problem turns out to be smooth and make the resolution even simpler according to the exhaustive literature dedicated to this subject [19].

**Alternating Minimization for the primal** – The semi-linearized PAM procedure [16, 17] can be formulated either by a gradient activation of the data-fidelity term or by a proximal activation. In both cases, the sequence  $\{\mathbf{u}^{[k]}, \mathbf{e}^{[k]}\}_{k \in \mathbb{N}}$  converges to a critical point of the primal problem.

- *Primal-SLPAM*

$$\begin{aligned} \mathbf{u}^{[k+1]} &= \text{prox}_{\frac{1}{c_k} f(\mathbf{A}\cdot, \mathbf{z})} \left( \mathbf{u}^{[k]} - \frac{2\lambda^2}{c_k} \mathbf{D}^* \Lambda_{\mathbf{e}^{[k]}} \mathbf{D} \mathbf{u}^{[k]} \right) \\ \mathbf{e}^{[k+1]} &= \left( \text{prox}_{\frac{\alpha\psi}{\lambda^2(\mathbf{D}_\ell \mathbf{u}^{[k+1]})^2 + d_k}} \left( \frac{\lambda^2(\mathbf{D}_\ell \mathbf{u}^{[k+1]})^2 + d_k \mathbf{e}_\ell^{[k]}}{\lambda^2(\mathbf{D}_\ell \mathbf{u}^{[k+1]})^2 + d_k} \right) \right)_\ell \end{aligned}$$

where  $\Lambda_{\mathbf{e}^{[k]}}$  is a diagonal matrix with diagonal values set to  $(1 - \mathbf{e}_\ell^{[k]})^2$ ,  $d_k > 0$ , and  $c_k > 0$  and upper bounded by the local Lipschitz constant of the coupling term.

- Gradient update for  $\mathbf{u}^{[k]}$  (Primal-SLPAM-GD)

$$\begin{aligned} \mathbf{u}^{[k+1]} &= \mathbf{u}^{[k]} - \frac{1}{c_k} \mathbf{A}^* \nabla f(\mathbf{A} \mathbf{u}^{[k]}, \mathbf{z}) - \frac{2\lambda^2}{c_k} \mathbf{D}^* \Lambda_{\mathbf{e}^{[k]}} \mathbf{D} \mathbf{u}^{[k]} \\ \mathbf{e}^{[k+1]} &= \left( \text{prox}_{\frac{\alpha\psi}{\lambda^2(\mathbf{D}_\ell \mathbf{u}^{[k+1]})^2 + d_k}} \left( \frac{\lambda^2((\mathbf{D} \mathbf{u}^{[k+1]})_\ell)^2 + d_k \mathbf{e}_\ell^{[k]}}{\lambda^2(\mathbf{D}_\ell \mathbf{u}^{[k+1]})^2 + d_k} \right) \right)_\ell \end{aligned}$$

where  $d_k > 0$ , and  $c_k > 0$  and upper bounded by the local Lipschitz constant of the sum of the data-fidelity term and the coupling term.

**Gradient based algorithm for the dual** – Under specific assumptions of Lemma 1, the dual formulation reads:

$$\hat{\mathbf{u}} = \arg \min_{\mathbf{u}} f(\mathbf{A} \mathbf{u}, \mathbf{z}) + \frac{\lambda^2}{2} \sum_{\ell=1}^{|\mathbb{E}|} q(\mathbf{D}_\ell \mathbf{u}) \quad (4)$$

where, for every  $\eta \in \mathbb{R}$ ,

$$q(\eta) = \begin{cases} \frac{\alpha}{\lambda^2} (2 - \frac{\alpha}{\lambda^2 \eta^2}) & \text{if } \eta^2 > \frac{\alpha}{\lambda^2}, \\ \eta^2 & \text{if } 0 \leq \eta^2 \leq \frac{\alpha}{\lambda^2}, \end{cases}$$

which turns to be a non-convex but smooth minimization problem. We can derive a gradient based and a proximal-gradient based iterations, such that the sequences  $\{\mathbf{u}^{[k]}, \mathbf{e}^{[k]}\}_{k \in \mathbb{N}}$  converge to a critical point of the dual problem [18].

- Dual gradient descent (*Dual-GD*):

$$\mathbf{u}^{[k+1]} = \mathbf{u}^{[k]} - \tau \mathbf{A}^* \nabla f(\mathbf{A} \mathbf{u}^{[k]}, \mathbf{z}) - \tau \mathbf{D}^* (q'(\mathbf{D}_\ell \mathbf{u}^{[k]}))_{1 \leq \ell \leq |\mathbb{E}|}$$

with  $\tau < (L_f \|\mathbf{A}\|^2 + L_q \|\mathbf{D}\|^2)^{-1}$ .

- Dual proximal gradient descent (*Dual-PG*):

$$\mathbf{u}^{[k+1]} = \text{prox}_{\frac{1}{c_k} f(\mathbf{A}\cdot, \mathbf{z})} \left( \mathbf{u}^{[k]} - \tau \mathbf{D}^* (q'(\mathbf{D}_\ell \mathbf{u}^{[k]}))_{1 \leq \ell \leq |\mathbb{E}|} \right)$$

with  $\tau < (L_q \|\mathbf{D}\|^2)^{-1}$ .

#### 4. NUMERICAL EXPERIMENTS

We focus on image denoising or restoration problems where the degradation model is  $\mathbf{z} = \mathbf{A}\bar{\mathbf{x}} + \epsilon$  where  $\epsilon \sim \mathcal{N}(0, \sigma^2 \mathbb{I})$ . We illustrate the performance of the different algorithmic strategies in this context.

Additionally to the algorithms described in the previous section having convergence guarantees using KL properties, we also include a minimisation of the dual function relying on BFGS to highlight the benefice of considering higher order optimization schemes.

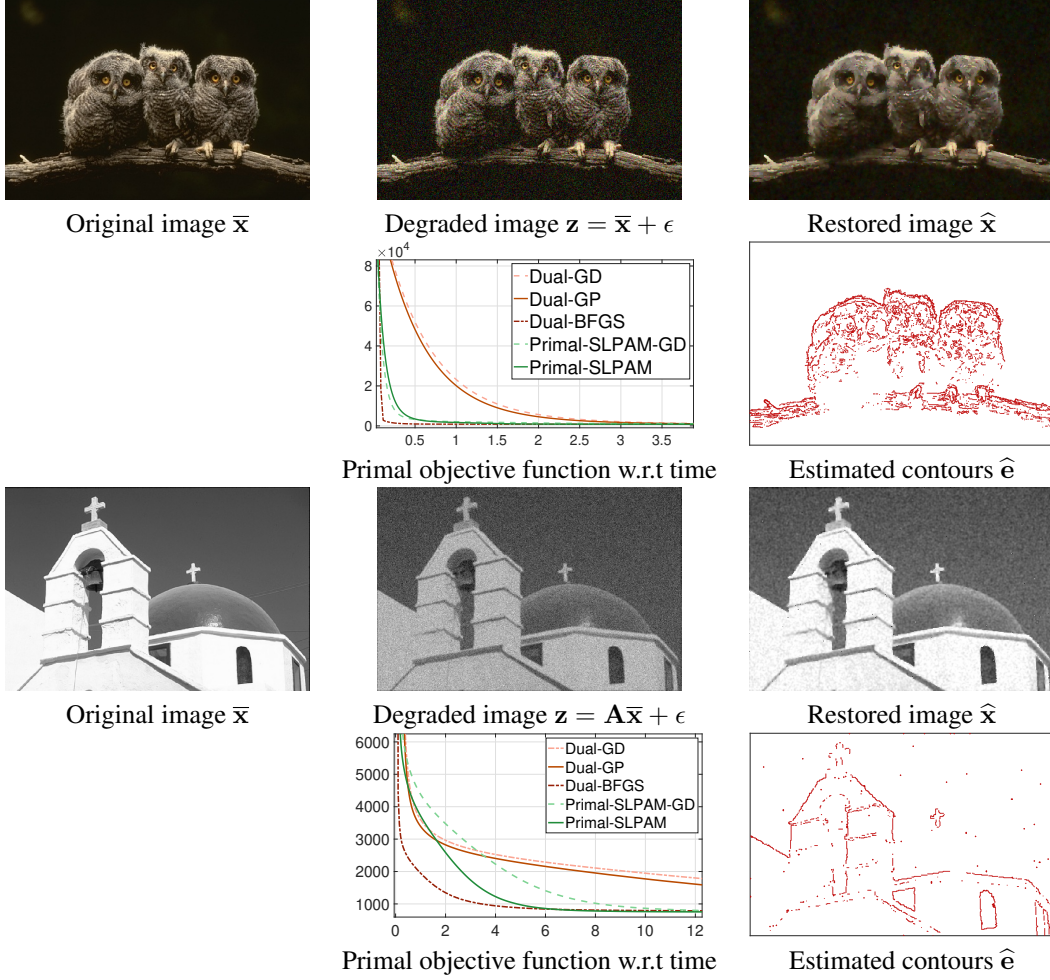
We display in Fig. 1 the convergence behavior of the primal objective function with respect to time in a context of (top) denoising with  $\sigma = 1$  and where hyperparameters are set to extract meaningful contours such that  $\lambda = \sqrt{1.4}$  and  $\alpha = 0.0015$  (bottom) restoration with Gaussian blur of standard deviation of 3 and Gaussian noise with standard deviation set to 0.1 where  $\lambda = \sqrt{12}$  and  $\alpha = 0.009$ .

Among first order methods (Dual-GD, Dual-GP, Primal-SLPAM-GD, and Primal-SLPAM), the procedures relying on proximal activation of the data-fidelity term are always faster. Additionally, we observe that primal-SLPAM is always the faster one.

Now, when focusing on the first order method Primal-SLPAM and Dual-BFGS, we observe that both have very close performance as the fastest one depends on the configuration (denoising or restoration).

#### 5. CONCLUSIONS AND FUTURE WORKS

This work derives the dual formulation of D-MS formulated as in [12, 13]. In the context where  $\phi = |\cdot|^2$  and  $\psi = |\cdot|$ , the dual formulation leads to a smooth non-convex objective function adapted to several optimization schemes. We numerically compared several optimization schemes either solving the primal or the dual formulation. Primal-SLPAM and Dual-BFGS appear to be comparably efficient. A future work would be to establish, at the instar of [20], convergence rates in the context of strongly convex functions, to identify regimes where a method could be preferable to another one.



**Fig. 1.** Estimation and evolution of the primal objective function with several algorithmic schemes either applied on the dual functional or the primal one. (top) denoising with  $\sigma = 1$ . (bottom) restoration with Gaussian blur of standard deviation of 3 and Gaussian noise with standard deviation set to 0.1.

Additionally, deep learning is more and more focused on model-based architecture involving a minimization procedure. Consequently, for D-MS formalism, having a flexibility relying on either primal and dual formulation offer a larger range of possible architectures.

## 6. ACKNOWLEDGMENT

The author would like to thank Prof.Y. Nesterov for the insightful discussions that initiated this work.

## 7. ANNEX

### 7.1. Proof of Proposition 1

The minimization Problem 1 can be equivalently written as

$$\min_{\mathbf{u}} f(\mathbf{A}\mathbf{u}, \mathbf{z}) + \frac{\lambda^2}{2} \sum_{\ell} \underbrace{\inf_{e_{\ell}} \left\{ \phi(\mathbf{D}_{\ell}\mathbf{u})(1 - e_{\ell})^2 + \frac{2\alpha}{\lambda^2} \psi(e_{\ell}) \right\}}_{\tilde{\psi}(e_{\ell}, \phi(\mathbf{D}_{\ell}\mathbf{u}))}.$$

From proximity operator definition, we can write:

$$\arg \min_{e_{\ell}} \tilde{\psi}(e_{\ell}, \phi(\mathbf{D}_{\ell}\mathbf{u})) = \begin{cases} \text{prox}_{\frac{\alpha\psi(\cdot)}{\lambda^2\phi(\mathbf{D}_{\ell}\mathbf{u})}}(1) & \text{if } \phi(\mathbf{D}_{\ell}\mathbf{u}) > 0, \\ 0 & \text{otherwise,} \end{cases}$$

and thus, if  $\phi(\mathbf{D}_{\ell}\mathbf{u}) > 0$ ,

$$\begin{aligned} \inf_{e_{\ell}} \tilde{\psi}(e_{\ell}, \phi(\mathbf{D}_{\ell}\mathbf{u})) &= \phi(\mathbf{D}_{\ell}\mathbf{u}) \left( 1 - \text{prox}_{\frac{\alpha\psi(\cdot)}{\lambda^2\phi(\mathbf{D}_{\ell}\mathbf{u})}}(1) \right)^2 \\ &\quad + \frac{2\alpha}{\lambda^2} \psi \left( \text{prox}_{\frac{\alpha\psi(\cdot)}{\lambda^2\phi(\mathbf{D}_{\ell}\mathbf{u})}}(1) \right) \end{aligned} \quad (5)$$

and 0 if  $\phi(\mathbf{D}_{\ell}\mathbf{u}) = 0$ . This yields, for  $\eta > \frac{\alpha}{\lambda^2}$ , to

$$\text{prox}_{\frac{\alpha}{\lambda^2\eta}|\cdot|}(1) = 1 - \frac{\alpha}{\lambda^2\eta}$$

and to 0 if  $0 < \eta \leq \frac{\alpha}{\lambda^2}$  leading to the established result.

## 8. REFERENCES

- [1] M. Storath, A. Weinmann, J. Friel, and M. Unser, “Joint image reconstruction and segmentation using the Potts model,” *Inverse Problems*, vol. 31, no. 2, pp. 025003, 2015.
- [2] C. Vacar and J.-F. Giovannelli, “Unsupervised joint deconvolution and segmentation method for textured images: a bayesian approach and an advanced sampling algorithm,” *EURASIP J. Adv. Signal Process.*, vol. 2019, no. 1, pp. 1–17, 2019.
- [3] B. Pascal, N. Pustelnik, P. Abry, M. Serres, and V. Vidal, “Joint estimation of local variance and local regularity for texture segmentation. application to multiphase flow characterization,” in *Proc. IEEE ICIP*, Athens, Greece, Oct 07-10, 2018, pp. 2092–2096.
- [4] D. Mumford and J. Shah, “Optimal approximations by piecewise smooth functions and associated variational problems,” vol. 42, no. 5, pp. 577–685, 1989.
- [5] S. Geman and D. Geman, “Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images,” *IEEE Trans. Pattern Anal. Match. Int.*, vol. 6, no. 6, pp. 721 – 741, Nov. 1984.
- [6] D. Geman and G. Reynolds, “Constrained image restoration and the recovery of discontinuities,” *IEEE Trans. PAMI*, vol. 14, no. 3, pp. 367–383, 1992.
- [7] S. Geman and D.E. McClure, “Bayesian image analysis: An application to single photon emission tomography,” *Proc. Statistical Computation Section, Amer. Statistical Assoc., Washington, DC*, p. 12–18, 1985.
- [8] T. Hebert and R. Leahy, “A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors,” *IEEE Trans. Med. Imag.*, p. 194–202, 1990.
- [9] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, “Deterministic edge-preserving regularization in computed imaging,” *IEEE Trans. Image Process.*, vol. 6, no. 2, pp. 298–311, 1997.
- [10] P. Lobel, C. Pichot, L. Blanc-Féraud, and M. Barlaud, “Conjugate gradient algorithm with edge-preserving regularization for image reconstruction from ipswitch data for mystery objects,” *IEEE Antennas Propagation Magazine*, vol. 39, no. 2, pp. 12–13, 1997.
- [11] 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.
- [12] M. Foare, J.-O. Lachaud, and H. Talbot, “Image restoration and segmentation using the Ambrosio–Tortorelli functional and discrete calculus,” Cancun, Mexico, 2016.
- [13] M. Foare, N. Pustelnik, and L. Condat, “Semi-linearized proximal alternating minimization for a discrete Mumford-Shah model,” *IEEE Trans. on Image Processing*, vol. 29, pp. 2176 – 2189, 2019.
- [14] H.T.V. Le, M. Foare, and N. Pustelnik, “Proximal based strategies for solving Discrete Mumford-Shah with Ambrosio-Tortorelli penalization on edges,” *IEEE Signal Process. Lett.*, vol. 29, pp. 952–956, 2022.
- [15] A. Blake and A. Zisserman, *Visual reconstruction*, MIT press, 1987.
- [16] H. Attouch and J. Bolte, “On the convergence of the proximal algorithm for nonsmooth functions involving analytic features,” *Math. Program.*, vol. 116, no. 1-2, pp. 5–16, 2009.
- [17] 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.
- [18] E. Chouzenoux, J.-C. Pesquet, and A. Repetti, “Variable metric forward-backward algorithm for minimizing the sum of a differentiable function and a convex function,” *Optim. Theory and Appl.*, pp. 107–132, 2014.
- [19] S. Wright and J. Nocedal, *Numerical optimization*, Springer, 1999.
- [20] Briceño-Arias and N. Pustelnik, “Theoretical and numerical comparison of first-order algorithms for co-coercive equations and smooth convex optimization,” *arXiv:2101.06152*, 2022.