MINIMAL PARTITIONS AND IMAGE CLASSIFICATION USING A GRADIENT-FREE PERIMETER APPROXIMATION

S. AMSTUTZ, A. A. NOVOTNY, AND N. VAN GOETHEM

Abstract. In this paper a new mathematically-founded method for the optimal partitioning of domains, with applications to the classification of greyscale and color images, is proposed. Since optimal partition problems are in general ill-posed, some regularization strategy is required. Here we regularize by a non-standard approximation of the total interface length, which does not involve the gradient of approximate characteristic functions, in contrast to the classical Modica-Mortola approximation. Instead, it involves a system of uncoupled linear partial differential equations and nevertheless shows Γ-convergence properties in appropriate function spaces. This approach leads to an alternating algorithm that ensures a decrease of the objective function at each iteration, and which always provides a partition, even during the iterations. The efficiency of this algorithm is illustrated by various numerical examples. Among them we consider binary and multilabel minimal partition problems including supervised or automatic image classification, inpainting, texture pattern identification and deblurring.

1. Introduction

Image processing is a huge field of research ranging from code-based algorithms to advanced mathematical tools, and where mainly two classes of problems are addressed. The first one is image *restoration* whose aim is to remove all effects responsible for an image degradation: noise, blur, missing parts, etc. Another family of problems can be referred to as image *segmentation*, where the constituents of a given image (damaged or not) are identified: it can be different colors, intensities or texture regions. Image *classification* is a particular form of image segmentation, where it is emphasized that the image characteristics are sought within a prescribed number of components or phases, identified by labels.

The standard greyscale image processing problem can be stated as restore and/or segment $f = A\bar{u} + \nu$ where $f : \Omega \mapsto [0, 1]$ is the observed image, $\bar{u}$ its idealized version (the undamaged image), A is a known or unknown mask operator (blur kernel or a projection operator away from the missing parts of $\bar{u}$), and ν is the noise. According to the application, one wishes to find a u which is either a continuous restoration of $\bar{u}$ or a segmented version of $\bar{u}$.

Image restoration and image segmentation are hard problems per se, mainly due to the fact that the minimization problem $\min_{u\in H(\Omega)} J(u) := \|Au - f\|_{H(\Omega)}$ is in general ill-posed (w.r.t. f if A is known and also w.r.t. A if the mask is unknown) in the sense that small perturbations in the data may produce unbounded variations in the solution, but also because it is not necessarily a convex problem, especially when the set $H(\Omega)$ contains discrete levels. Of course, simultaneous segmentation and restoration of a blurred and noisy image is even a harder task, in particular in the case of blind deconvolution. Let us emphasize that an objective assessment of segmentation algorithms is hardly found, essentially because there is no unique ground-truth classification of an image against which the output of an algorithm may be compared.

Several mathematical models coexist in the literature to provide as output a restoration and/or a segmentation of a given image. Let us just mention the Mumford-Shah [4, 32] and the TV-L^2 (the so-called Rudin-Osher-Fatemi model) and TV-L^1 functionals [16, 35]. It is today widely recognized [14, 15, 38] that in order to obtain a solution which is smooth enough while preserving the edges, one should consider the problem $\min_{u\in H(\Omega)} J(u) + \alpha|Du|(\Omega)$ where $|Du|(\Omega)$ is the total variation of u in Ω, with Du, its distributional derivative, i.e., a measure with a diffuse part identified with ∇u outside the edges and a concentrated part on the edges, and with α a weight on the total variation of the image. For specific purposes, some authors have proposed to replace the total variation by anisotropic variants [22, 23, 25]. Segmentation problems, mainly in the bi-label case, can also be

2010 *Mathematics Subject Classification.* 49Q10,49Q20,49M25,35J05,35J25,65K10.

Key words and phrases. Image classification, deblurring, optimal partitions, perimeter approximation.

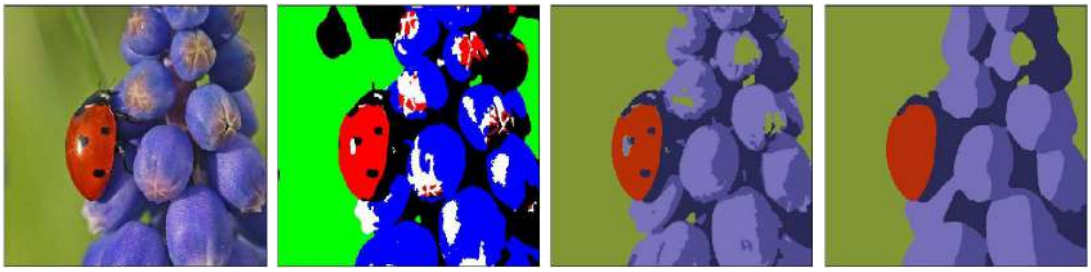


FIGURE 1. Color image classification, from left to right. First: original image (611×471 pixels). Second: supervised solution for $\alpha = 5.10^{-4}\sqrt{n}$ (n is the number of pixels). Third and fourth: unsupervised solution for $\alpha = 5.10^{-4}\sqrt{n}$ and $\alpha = 10^{-2}\sqrt{n}$, respectively. TV-model with $H = L^1$, $A = I$ and 5 labels.

FIGURE 2. Greylevel image classification, from left to right. First: original picture (390×390 pixels). Second: damaged image with salt and pepper noise. Third: unsupervised solution for $\alpha = 4.10^{-4}\sqrt{n}$. TV-model with $H = L^1$, $A = I$ and 4 levels of grey.

addressed by methods of moving interfaces, like level-sets [19, 26, 33] or snakes [3]. A totally different (purely discrete) approach consists in viewing the image as a graph and segmentation as a minimal cut problem [20, 36, 39].

In this paper we propose a novel mathematically-founded method for image classification purposes. More precisely, our method assigns to each pixel of a given image a label, with a prescribed upper bound on the number of labels. Typically, each label corresponds to a grey or color level. The classification can be supervised, i.e., the intensity of grey or the color for each label is fixed, or unsupervised, meaning that these values are automatically chosen by the algorithm (as done, e.g., in [17, 26]). In Fig. 1 the resulting classification of a color image is shown with and without level updates, and where the effect of the parameter α is shown: a smaller value will provide a solution with a greater perimeter, that is with more details. Let us emphasize that the value of α should be specifically adapted for each kind of image to reconstruct. In particular a "large enough" value of α should be used for noisy images, in order to remove the spurious perimeter created by the noise. As an example, an unsupervised greylevel simultaneous classification and denoising of a picture with our method (and the TV-L^1 model) is shown in Fig. 2. Later, we will also show examples of black and white image deblurring with an assumed known blur kernel (as is the case in many applications, the blur being due to diffraction, motion, zoom, of a measure apparatus etc.). We will also apply our method to perform combined deblurring and denoising and to classify an image in terms of prescribed textures.

Our method consists in a multiphase (piecewise constant) joint classification and restoration of an image, as based on

- a gradient-free approximation of the total variation in the functional setting,
- an optimal partitioning algorithm.

Here, we mean by *functional setting* the fact that we approximate the continuous total variation, instead of its discretized counterpart on a chosen mesh. It is also meant that the approximating functionals converge, as a small parameter ε tends to zero, in some *variational sense* (the Γ-convergence) to the total variation. By *gradient-free* we mean that, as opposed to Modica-Mortola/Ambrosio-Tortorelli-type regularizations [5, 28, 29, 30, 31, 34], the gradient of the image does not appear in the approximating functionals. Hence, from a numerical point of view, our approach is better suited to the reconstruction of discontinuous images. It has the other feature to force discrete prescribed values. This is of course a major advantage as to the applications to *optimal partition* problems (see [9, 17]), which are in some sense primary to classification problems and whose basic concepts are recalled hereafter. Consider a bounded domain Ω of $\mathbb{R}^2$, a number $N \in \mathbb{N}$, functions $g_1, ..., g_N \in L^1(\Omega)$, and a parameter $\alpha > 0$. A prototype problem of minimal partition reads:

$$\min_{\Omega_1,...,\Omega_N} \sum_{i=1}^{N} \left[\int_{\Omega_i} g_i(x)dx + \frac{\alpha}{2} \mathrm{Per}(\Omega_i) \right], \tag{1.1}$$

where the minimum is searched among all partitions $(\Omega_1, ..., \Omega_N)$ of Ω by subsets of finite perimeter. Here, $\mathrm{Per}(\Omega_i)$ is the relative perimeter of Ω_i in Ω, i.e., under regularity conditions, $\mathrm{Per}(\Omega_i) = |\partial\Omega_i \cap \Omega|$. Hence we have

$$\frac{\alpha}{2} \sum_{i=1}^{N} \mathrm{Per}(\Omega_i) = \alpha \sum_{i<j} |\partial\Omega_i \cap \partial\Omega_j|.$$

Optimal partition problems in imaging are known to be challenging. Other methods as found in the literature are based on convexification, whose major difficulty is to construct an auxiliary cost function which does not to depart too much from the convex envelope of the original one. In particular these methods solve the optimal partition problem exactly only for two levels (see [17, 40]). Other mathematical tools, such as topological asymptotic analysis [11], can be applied to address image classification.

In essence, our method is not designed for the restoration of continuous images, but it applies whenever a segmented restoration is sought, two examples of which are shown in Figs. 1 and 2. It is *a priori* also not well-suited to process images with texture parts that must be finely identified, since the optimal partitioning will rather identify homogeneous regions. Nevertheless it can be observed that the segmented image of Fig. 2 features some fine characteristics, as e.g., the shade on Lena's hat and some residual texture on its feather tuft. We are aware that more efficient methods exist to address these restoration problems, in particular if the image to restore shows sharp edges of connected subregions [10, 12, 18].

Nonetheless, our method is adequate for all problems whose underlying mathematical structure is related to optimal partition. In particular it is adapted to classify an image without sharp edges (a pioneering reference on this subject is [19]), a simple example of which is shown in Fig. 10(b). Let us remark that addressing denoising for this kind of images is in general cumbersome, since meaningful portions of the image are completely disconnected without being considered as noise. Our method shows particular good behavior in this case. Another example where our method is well suited is the following. Assume that a detail of an image is occluded, i.e., is missing for some reason, and that this detail shows a structure which cannot be recovered by, e.g., harmonic expansion [12]. Then, an optimal partition formulation of the problem allows us to reconstruct the missing information, as shown in Fig. 5 for the triple and quadruple-point examples.

The paper is organized as follows. The leading ideas are introduced in Section 2 in the context of binary minimal partition. The method is then extended to the multilabel partitioning problem in Section 3. The image classification problem is addressed in Sections 4 and 5 for greyscale and color images, respectively. Section 6 deals with the classification of images based on the analysis of anisotropic textures. The deblurring problem is discussed in Section 7.

2. Binary minimal partition

2.1. **Motivation: binary image classification.** We shall introduce our method on a simple example of image processing which consists of binary image classification. Let us begin with some definitions and notation. Let Ω be an open rectangle of $\mathbb{R}^2$. We define the set

$$\mathcal{E} = L^\infty(\Omega, \{0, 1\})$$

of characteristic functions in Ω and the functional $F : \mathcal{E} \to \mathbb{R} \cup \{+\infty\}$ such that

$$F(u) = \begin{cases} \frac{1}{2}|Du|(\Omega) & \text{if } u \in BV(\Omega, \{0,1\}), \\ +\infty & \text{otherwise.} \end{cases} \tag{2.1}$$

We recall that the total variation of $u \in L^1(\Omega)$ is defined as

$$|Du|(\Omega) = \sup\{\langle u, \operatorname{div} \xi\rangle : \ \xi \in \mathcal{C}^1_c(\Omega), |\xi(x)| \leq 1 \ \forall x \in \Omega\}, \tag{2.2}$$

and u is said of *bounded variation*, denoted $u \in BV(\Omega)$, when $|Du|(\Omega) < \infty$. Throughout we use the notation

$$\langle u, v\rangle := \int_\Omega u\, v\, dx$$

for every pair of functions u, v having suitable regularity. If u is the characteristic function of $E \subset \Omega$, $|Du|(\Omega)$ is called the relative perimeter of E in Ω, and E is said of *finite perimeter* when $u \in BV(\Omega)$. In this case it is shown that $|Du|(\Omega) = \mathcal{H}^1(\partial_M E \cap \Omega)$, where $\partial_M E$ is the measure theoretical (or essential) boundary of E and $\mathcal{H}^1$ is the one-dimensional Hausdorff measure on $\mathbb{R}^2$. We recall (see, e.g., [4, 8]) that $\partial_M E$ is a subset of the topological boundary ∂E of E, defined by

$$\partial_M E = \left\{ x \in \mathbb{R}^2 : \limsup_{\rho \to 0} \frac{\mathcal{L}^2(B(x,\rho) \cap E)}{\mathcal{L}^2(B(x,\rho))} > 0 \text{ and } \limsup_{\rho \to 0} \frac{\mathcal{L}^2(B(x,\rho) \setminus E)}{\mathcal{L}^2(B(x,\rho))} > 0 \right\},$$

with $\mathcal{L}^2$ the Lebesgue measure on $\mathbb{R}^2$. Moreover, E admits a generalized unit inner normal n_E at $\mathcal{H}^1$-almost every point $x \in \partial_M E$, see again [4, 8].

The following calculus illustrates our method on a simple case. Let $f \in L^\infty(\Omega, [0,1])$ be a given image and $c_1, c_2 \in [0,1]$ be two classes representing different intensities of grey. We want to approximate f by the piecewise constant image $u\, c_1 + (1-u)\, c_2$, with $u \in \mathcal{E}$. Given $p \in [1, +\infty[$, the binary image classification problem is formulated as: find a minimizer for

$$I := \min_{u \in \mathcal{E}} \left\{ \|uc_1 + (1-u)c_2 - f\|^p_{L^p(\Omega)} + \alpha F(u) \right\} . \tag{2.3}$$

Since u is a characteristic function, we have

$$\begin{aligned} \int_\Omega |uc_1 + (1-u)c_2 - f|^p dx &= \int_\Omega |u(c_1 - f) + (1-u)(c_2 - f)|^p dx \\ &= \int_\Omega u|c_1 - f|^p dx + \int_\Omega (1-u)|c_2 - f|^p dx. \end{aligned}$$

We can then redefine I, up to an additive constant, as

$$I = \min_{u \in \mathcal{E}} \{\mathcal{I}(u) := \langle u, g\rangle + \alpha F(u)\} \ , \tag{2.4}$$

with

$$g = |c_1 - f|^p - |c_2 - f|^p. \tag{2.5}$$

More generally, we shall address the numerical solution of (2.4) for an arbitrary function $g \in L^1(\Omega)$. We refer to this problem as a binary minimal partition problem.

As already said, a number of specific methods provide satisfactory solutions to this problem, or some of its variants. We have quoted level-sets [19, 26, 33], snakes [3] and graph cutting [20, 36, 39], and we must also mention the convex reformulation of (2.4) given in Prop. 2.6 of [16], which can be solved for instance by Chambolle's algorithm [14]. The main advantage of our approach, as we will see later, is its straightforward generalization to the multiphase case as well as to the presence of a mask operator.

2.2. **Approximation of the perimeter term.** Let us define

$$\tilde{\mathcal{E}} = L^\infty(\Omega, [0,1]),$$

the convex hull of $\mathcal{E}$, and the function $\tilde{F} : \tilde{\mathcal{E}} \to \mathbb{R} \cup \{+\infty\}$ such that

$$\tilde{F}(u) = \begin{cases} F(u) & \text{if } u \in \mathcal{E}, \\ +\infty & \text{otherwise.} \end{cases}$$

It is shown in [7] that a suitable approximation of $\tilde{F}$ is provided by the functional $\tilde{F}_\varepsilon$ defined as

$$\tilde{F}_\varepsilon(u) = \inf_{v\in H^1(\Omega)} \left\{ \varepsilon \|\nabla v\|^2_{L^2(\Omega)} + \frac{1}{\varepsilon}\left(\|v\|^2_{L^2(\Omega)} + \langle u, 1-2v\rangle \right) \right\}. \tag{2.6}$$

The above minimization problem is easily solved, and we have the alternative expression

$$\tilde{F}_\varepsilon(u) = \frac{1}{\varepsilon}\langle 1 - L_\varepsilon u, u\rangle, \tag{2.7}$$

where $L_\varepsilon u$ is the (weak) solution of the boundary value problem with unknown $v \in H^1(\Omega)$:

$$\left\{ \begin{array}{rcll} -\varepsilon^2 \Delta v + v &=& u & \text{in} \quad \Omega, \\ \partial_n v &=& 0 & \text{on} \quad \partial\Omega. \end{array} \right. \tag{2.8}$$

Note that, on choosing classically $\min(0, v)$ as test function in the variational formulation of (2.8), then considering the function $1 - v$ as unknown, we obtain that $L_\varepsilon u \in \tilde{\mathcal{E}}$ for all $u \in \tilde{\mathcal{E}}$. Substituting $\tilde{F}_\varepsilon$ for $\tilde{F}$ leads to the approximate minimization problem:

$$I_\varepsilon = \min_{u\in\tilde{\mathcal{E}}} \left\{ \mathcal{I}_\varepsilon(u) := \langle u, g\rangle + \alpha \tilde{F}_\varepsilon(u) \right\} . \tag{2.9}$$

2.3. **Mathematical properties.** We first recall some key properties of the functional $\tilde{F}_\varepsilon$ (Proposition 2.1 and Theorems 2.3, 2.4 and 2.5) proved in [7]. Nevertheless, the proof of Theorem 2.3 given in [7] relies on a result from [37], and does not lend itself to the generalization we will need later. Therefore we provide an original and well-adapted proof of Theorem 2.3. However, we point out that this proof is valid only in dimension 2 for a rectangular domain Ω.

Let us start by defining the functional

$$F_\varepsilon(u) = \inf_{v\in H^1(\Omega)} \left\{ \varepsilon \|\nabla v\|^2_{L^2(\Omega)} + \frac{1}{\varepsilon}\|v - u\|^2_{L^2(\Omega)} \right\}. \tag{2.10}$$

It is straightforward to check that, for every $u \in L^2(\Omega)$,

$$\tilde{F}_\varepsilon(u) = F_\varepsilon(u) + \frac{1}{\varepsilon}\langle u, 1-u\rangle \, . \tag{2.11}$$

Proposition 2.1. *The function $\tilde{F}_\varepsilon : \tilde{\mathcal{E}} \to \mathbb{R}$ is continuous on $\tilde{\mathcal{E}}$ for the weak-$*$ topology of $L^\infty(\Omega)$, and it is the relaxation (i.e. the lower semi-continuous envelope) of the function*

$$u \in \tilde{\mathcal{E}} \mapsto \left\{ \begin{array}{ll} F_\varepsilon(u) & \text{if } u \in \mathcal{E}, \\ +\infty & \text{if } u \notin \mathcal{E}. \end{array} \right. \tag{2.12}$$

The notion of Γ-convergence (see, e.g., [8, 13, 21]) is a powerful tool of calculus of variations in function spaces. Given a metrizable space (X, d) (in our case $X = \tilde{\mathcal{E}}$ endowed with the distance induced by the L^1 norm) one would like the maps

$$f \mapsto \inf_X f \quad \text{and} \quad f \mapsto \operatorname*{argmin}_X f$$

to be sequentially continuous on the space of extended real-valued functions $f : X \to \mathbb{R} \cup \{+\infty\}$. Throughout we will simply index sequences by ε, to actually represent any positive sequence of indices (ε_k) such that $\lim_{k\to+\infty} \varepsilon_k = 0$.

Definition 2.2 (Γ-convergence)**.** Let (f_ε) be a sequence of functions $f_\varepsilon : X \to \mathbb{R} \cup \{+\infty\}$ and $f : X \to \mathbb{R} \cup \{+\infty\}$. We say that f_ε Γ-converges to f if and only if, for all $u \in X$, the following two conditions hold:

(1) for all sequences $(u^\varepsilon) \in X$ such that $d(u^\varepsilon, u) \to 0$ it holds $f(u) \le \liminf_{\varepsilon\to 0} f_\varepsilon(u^\varepsilon)$,
(2) there exists a sequence $(\bar{u}^\varepsilon) \in X$ such that $d(\bar{u}^\varepsilon, u) \to 0$ and $f(u) \ge \limsup_{\varepsilon\to 0} f_\varepsilon(\bar{u}^\varepsilon)$.

Theorem 2.3 (Γ-convergence of the approximating functionals)**.** *As $\varepsilon \to 0$, the functionals $\tilde{F}_\varepsilon$ $\Gamma-$converge to $\tilde{F}$ in $\tilde{\mathcal{E}}$, strongly in $L^1(\Omega)$.*

Theorem 2.4 (Equicoercivity)**.** *Let (u^ε) be a sequence of functions of $\tilde{\mathcal{E}}$ such that $\sup_\varepsilon \tilde{F}_\varepsilon(u^\varepsilon) < +\infty$. There exists $u \in \mathcal{E}$ such that $u^\varepsilon \to u$ strongly in $L^1(\Omega)$ for a subsequence.*

Theorem 2.5. *Let u^ε be an approximate minimizer of (2.9), i.e.*

$$\mathcal{I}_\varepsilon(u^\varepsilon) = \langle u^\varepsilon, g\rangle + \alpha \tilde{F}_\varepsilon(u^\varepsilon) \leq I_\varepsilon + \lambda_\varepsilon,$$

with $u^\varepsilon \in \tilde{\mathcal{E}}$ and $\lim_{\varepsilon\to 0} \lambda_\varepsilon = 0$. Then we have

$$\mathcal{I}_\varepsilon(u^\varepsilon) \to I.$$

Moreover, the sequence (u^ε) admits cluster points, and each of these cluster points is a minimizer of (2.4).

For multilabel problems, we will need special recovery sequences in the lim sup inequality of the Γ-convergence, which can be put together in order to form a partition of Ω. In fact, we will show a pointwise convergence result (Theorem 2.8). To do so, we will establish two preliminary results. The first one provides the $\Gamma - \liminf$ inequality of the Γ-convergence, and has already been proved in [7] using [37]. For completeness we propose here a new, self-contained, and much shorter proof.

Proposition 2.6. *Let $u \in \tilde{\mathcal{E}}$ and (u^ε) be a sequence of $\tilde{\mathcal{E}}$ such that $u^\varepsilon \to u$ strongly in $L^1(\Omega)$. Then we have*

$$\liminf_{\varepsilon\to 0} \tilde{F}_\varepsilon(u^\varepsilon) \geq \tilde{F}(u).$$

Proof. First, from (2.11), we have that

$$\tilde{F}_\varepsilon(u^\varepsilon) \geq \frac{1}{\varepsilon}\langle u^\varepsilon, 1 - u^\varepsilon\rangle.$$

Since $0 \leq u^\varepsilon \leq 1$, we also have $u^\varepsilon \to u$ in $L^2(\Omega)$, whereby $\liminf_{\varepsilon\to 0} \tilde{F}_\varepsilon(u^\varepsilon) = +\infty$ whenever $u \notin \mathcal{E}$. We now assume that $u \in \mathcal{E}$ and that $\liminf_{\varepsilon\to 0} \tilde{F}_\varepsilon(u^\varepsilon) = \ell < +\infty$. Up to extracting a (non-relabeled) subsequence, we further assume that $\ell = \lim_{\varepsilon\to 0} \tilde{F}_\varepsilon(u^\varepsilon)$. Like in [7] we introduce the potentials

$$W(s,t) = t^2 + s(1-2t), \qquad \mathcal{W}(t) = t^2 + \min(0, 1-2t),$$

which clearly satisfy $W(s,t) \geq \mathcal{W}(t)$ for all $(s,t) \in [0,1]\times[0,1]$ and $\mathcal{W}(1/2-\tau) = \mathcal{W}(1/2+\tau) \geq 0$ for all $\tau \in \mathbb{R}$. Set $v^\varepsilon = L_\varepsilon u^\varepsilon$. Then we have

$$\tilde{F}_\varepsilon(u^\varepsilon) = \int_\Omega \left[\varepsilon|\nabla v^\varepsilon|^2 + \frac{1}{\varepsilon}W(u^\varepsilon, v^\varepsilon)\right] dx \geq \int_\Omega \left[\varepsilon|\nabla v^\varepsilon|^2 + \frac{1}{\varepsilon}\mathcal{W}(v^\varepsilon)\right] dx.$$

The elementary Young inequality entails

$$\tilde{F}_\varepsilon(u^\varepsilon) \geq 2\int_\Omega |\nabla v_\varepsilon|\sqrt{\mathcal{W}(v^\varepsilon)}dx.$$

Let $\psi : \mathbb{R} \to \mathbb{R}$ be a function such that $\psi'(t) = \sqrt{\mathcal{W}(t)}$ for all $t \in \mathbb{R}$, and set $w^\varepsilon = \psi \circ v^\varepsilon$. We obtain by the chain rule

$$\tilde{F}_\varepsilon(u^\varepsilon) \geq 2\int_\Omega |\nabla w^\varepsilon| dx. \tag{2.13}$$

Next, from

$$\tilde{F}_\varepsilon(u^\varepsilon) \geq F_\varepsilon(u^\varepsilon) \geq \frac{1}{\varepsilon}\|u^\varepsilon - v^\varepsilon\|^2_{L^2(\Omega)},$$

we infer that $\lim_{\varepsilon\to 0} \|u^\varepsilon - v^\varepsilon\|_{L^2(\Omega)} = 0$. Obviously, the same limit holds for the L^1 norm. It follows that

$$\|v^\varepsilon - u\|_{L^1(\Omega)} \leq \|v^\varepsilon - u^\varepsilon\|_{L^1(\Omega)} + \|u^\varepsilon - u\|_{L^1(\Omega)} \to 0.$$

Using that $0 \leq \mathcal{W}(t) \leq 1/4$ for all $t \in [0,1]$ we obtain that ψ is $1/2$-Lipschitz on $[0,1]$, which implies that $w^\varepsilon \to w := \psi \circ u$ in $L^1(\Omega)$. In view of (2.13), the lower semicontinuity of the total variation yields that $w \in BV(\Omega)$ with

$$\liminf_{\varepsilon\to 0} \tilde{F}_\varepsilon(u^\varepsilon) \geq 2|Dw|(\Omega).$$

Yet, $u \in \mathcal{E}$ and $w = \psi \circ u$ entail $w \in BV(\Omega, \{\alpha, \beta\})$ with $\alpha = \psi(0)$ and $\beta = \psi(1)$. In addition, as u takes only $0-1$ values a.e. in Ω, we can write $w = \alpha + (\beta - \alpha)u$ and $u = (w-\alpha)/(\beta-\alpha)$, whereby $u \in BV(\Omega, \{0,1\})$. It also follows that $Dw = (\beta-\alpha)Du$, hence $|Dw|(\Omega) = (\beta-\alpha)|Du|(\Omega)$. We finally compute

$$\beta - \alpha = \int_0^1 \psi'(t)dt = \int_0^1 \sqrt{\mathcal{W}(t)}dt = 2\int_0^{1/2} t dt = \frac{1}{4},$$

which leads to

$$\liminf_{\varepsilon\to 0} \tilde{F}_\varepsilon(u^\varepsilon) \geq \frac{1}{2}|Du|(\Omega) = \tilde{F}(u).$$

The proof is thus achieved. □

The following proposition gives a uniform upper bound for the functionals F_ε.

Proposition 2.7. *For all $u \in BV(\Omega, [0,1])$ and all $\varepsilon > 0$ we have*

$$F_\varepsilon(u) \leq \frac{1}{2}|Du|(\Omega).$$

Proof. We proceed in two steps.
Step 1. Let us first assume that $u \in \tilde{\mathcal{E}} \cap H^1(\Omega)$. From (2.7) and (2.11) we obtain that

$$F_\varepsilon(u) = \frac{1}{\varepsilon}\langle u - v^\varepsilon, u\rangle, \tag{2.14}$$

with $v^\varepsilon = L_\varepsilon u$. The weak formulation of (2.8) gives the alternative expression

$$F_\varepsilon(u) = \varepsilon \int_\Omega \nabla v^\varepsilon . \nabla u dx. \tag{2.15}$$

Let $\tilde{u}$ be the extension of u to the whole $\mathbb{R}^2$ obtained by performing successive reflections with respect to the main axes. By uniqueness we have $v^\varepsilon = \Phi_\varepsilon * \tilde{u}$, where Φ_ε is the fundamental solution of the operator $-\varepsilon^2\Delta + I$, i.e., the solution in the sense of distributions of

$$-\varepsilon^2 \Delta \Phi_\varepsilon + \Phi_\varepsilon = \delta. \tag{2.16}$$

A change of variable in (2.16) indicates that $\Phi_\varepsilon(x) = \varepsilon^{-2}\Phi_1(\varepsilon^{-1}x)$, whereby

$$\nabla v^\varepsilon(x) = (\nabla \Phi_\varepsilon * \tilde{u})(x) = \frac{1}{\varepsilon^3}\int_{\mathbb{R}^2} \nabla\Phi_1(\frac{y}{\varepsilon})\tilde{u}(x-y)dy.$$

Then a change of variable in the above integral entails

$$\varepsilon \nabla v^\varepsilon(x) = \int_{\mathbb{R}^2} \nabla\Phi_1(z)\tilde{u}(x-\varepsilon z)dz.$$

Let ν be an arbitrary unit vector of $\mathbb{R}^2$. Using that $0 \leq \tilde{u} \leq 1$ we derive that

$$\varepsilon \nabla v^\varepsilon(x).\nu = \int_{\mathbb{R}^2} (\nabla\Phi_1(z).\nu)\,\tilde{u}(x-\varepsilon z)dz \leq \int_{\mathbb{R}^2} \max(0, \nabla\Phi_1(z).\nu)dz =: M_\nu. \tag{2.17}$$

We shall now compute this latter bound. The fundamental solution of the normalized screened Poisson equation is known to be given by

$$\Phi_1(z) = \frac{1}{2\pi}K_0(|z|),$$

where K_0 is the modified Bessel function (see [1]) defined by

$$K_0(r) = \int_0^{+\infty} \exp(-r\cosh t)dt.$$

Using polar coordinates with ν as first basis vector, we arrive at

$$M_\nu = \frac{1}{2\pi}\int_0^{+\infty}\left[\int_0^{2\pi} \max(0, K_0'(r)\cos\theta)rd\theta\right]dr.$$

Yet, we have

$$K_0'(r) = -\int_0^{+\infty} \exp(-r\cosh t)\cosh t\ dt \leq 0,$$

hence

$$M_\nu = \frac{1}{2\pi}\int_0^{+\infty}\left[\int_{\frac{\pi}{2}}^{\frac{3\pi}{2}} \cos\theta d\theta\right]K_0'(r)rdr = -\frac{1}{\pi}\int_0^{+\infty} K_0'(r)rdr.$$

Using that K_0 has an exponential decay at $+\infty$ and a logarithmic growth at 0, an integration by parts yields

$$M_\nu = \frac{1}{\pi}\int_0^{+\infty} K_0(r)dr = \frac{1}{\pi}\int_0^{+\infty}\left[\int_0^{+\infty}\exp(-r\cosh t)dt\right]dr.$$

By Fubini's theorem, the order of integration can be changed, which leads to

$$M_\nu = \frac{1}{\pi}\int_0^{+\infty} \frac{1}{\cosh t} dt = \frac{1}{2},$$

this latter integral being easily computed by the change of variable $s = e^t$. Coming back to (2.17), which holds true for any unit vector ν and a.e. $x \in \Omega$, we infer

$$\|\varepsilon \nabla v^\varepsilon\|_{L^\infty(\Omega)} \leq \frac{1}{2}.$$

Finally, (2.15) yields

$$F_\varepsilon(u) \leq \frac{1}{2}\|\nabla u\|_{L^1(\Omega)} = \frac{1}{2}|Du|(\Omega).$$

Step 2. We now consider an arbitrary function $u \in BV(\Omega, [0,1])$. By density of $\mathcal{C}^\infty(\overline{\Omega})$ in $BV(\Omega)$ for the intermediate convergence (see, e.g., [8] Theorem 10.1.2), there exists a sequence of functions $u_k \in \mathcal{C}^\infty(\overline{\Omega})$ such that $u_k \to u$ in $L^1(\Omega)$ and $|Du_k|(\Omega) \to |Du|(\Omega)$. Set $\bar{u}_k = P_{[0,1]}(u_k)$, with $P_{[0,1]}(t) = \max(0, \min(1,t))$ the projection onto $[0,1]$. As $P_{[0,1]}$ is 1-Lipschitz we have $\|\bar{u}_k - u\|_{L^1(\Omega)} = \|P_{[0,1]}(u_k) - P_{[0,1]}(u)\|_{L^1(\Omega)} \leq \|u_k - u\|_{L^1(\Omega)}$, and since $u_k \in H^1(\Omega)$ we also have

$$|D\bar{u}_k|(\Omega) = \int_\Omega |\nabla \bar{u}_k| dx = \int_\Omega |\nabla u_k| \chi_{\{0 \leq u_k \leq 1\}} dx \leq \int_\Omega |\nabla u_k| dx = |Du_k|(\Omega).$$

As $\bar{u}_k \in \tilde{\mathcal{E}} \cap H^1(\Omega)$, we get from step 1

$$F_\varepsilon(\bar{u}_k) \leq \frac{1}{2}|D\bar{u}_k|(\Omega) \leq \frac{1}{2}|Du_k|(\Omega). \tag{2.18}$$

From the expression (2.14) and the continuity of $L_\varepsilon : L^2(\Omega) \to L^2(\Omega)$, it stems that F_ε is continuous in $L^2(\Omega)$, thus also in $L^1(\Omega, [0,1])$. Passing to the limit in (2.18) entails

$$F_\varepsilon(u) \leq \frac{1}{2}|Du|(\Omega),$$

and the proof is complete. □

We now arrive at the desired pointwise convergence result.

Theorem 2.8. *For all $u \in \tilde{\mathcal{E}}$ it holds*

$$\lim_{\varepsilon \to 0} \tilde{F}_\varepsilon(u) = \tilde{F}(u).$$

Proof. In view of Proposition 2.6 applied to the constant sequence $u^\varepsilon = u$, it suffices to prove that

$$\limsup_{\varepsilon \to 0} \tilde{F}_\varepsilon(u) \leq \tilde{F}(u). \tag{2.19}$$

If $u \notin BV(\Omega, \{0,1\})$, then (2.19) is obviously satisfied since $\tilde{F}(u) = +\infty$. If now $u \in BV(\Omega, \{0,1\})$, then $\tilde{F}_\varepsilon(u) = F_\varepsilon(u)$ for each $\varepsilon > 0$, thus (2.19) is a direct consequence of Proposition 2.7. □

Note that, as announced in the beginning of this section, Theorem 2.3 is immediately retrieved by combining Proposition 2.6 and Theorem 2.8.

Finally, by the direct method of the calculus of variations, we straightforwardly prove the following existence result (see [7] for details).

Proposition 2.9. *The infima of $\mathcal{I}$ and $\mathcal{I}_\varepsilon$ in* (2.4) *and* (2.9) *are finite and attained in $BV(\Omega, \{0,1\})$ and $\tilde{\mathcal{E}}$, respectively.*

2.4. **Algorithm.** Our algorithm is based on a continuation method, namely we construct a decreasing sequence (ε_m) of positive numbers tending to zero, and, for each $\varepsilon = \varepsilon_m$, we find an approximate minimizer of (2.9) using as initialization the solution obtained at iteration $m-1$. In the sequel, the subscript m will be dropped for simplicity.

2.4.1. *Description of the algorithm in the function space setting.* Plugging (2.6) into (2.9), we obtain that the subproblem at ε fixed consists is solving the following two-level minimization problem

$$I_\varepsilon = \min_{u\in\tilde{\mathcal{E}}} \inf_{v\in H^1(\Omega)} \left\{ \langle u, g\rangle + \alpha \left[\varepsilon \|\nabla v\|^2_{L^2(\Omega)} + \frac{1}{\varepsilon}\left(\|v\|^2_{L^2(\Omega)} + \langle u, 1-2v\rangle \right) \right] \right\}. \tag{2.20}$$

The simple structure of this problem with respect to each variable u and v leads us to use an alternating minimization algorithm. We refer to [2, 7] for efficient applications of such algorithms to problems of similar form in topology optimization. The superscript k is used to designate variables computed at iteration k. The iteration k, $k \geq 1$, consists in the two steps described below.

(1) The minimization with respect to v is straightforward. It consists in solving the boundary value problem

$$\left\{ \begin{array}{rcll} -\varepsilon^2 \Delta v^k + v^k &=& u^{k-1} & \text{in} \quad \Omega, \\ \partial_n v^k &=& 0 & \text{on} \quad \partial\Omega. \end{array} \right. \tag{2.21}$$

(2) The minimization with respect to u is a linear programming problem in a convex set. Therefore a minimizer can always be found among the extreme points of $\tilde{\mathcal{E}}$. More precisely here, we have to minimize at each point $x \in \Omega$ the linear function

$$\phi(s) = s\, g(x) + \frac{\alpha}{\varepsilon} s(1 - 2v^k(x))$$

over $s \in [0, 1]$. Setting

$$\zeta^k(x) = g(x) + \frac{\alpha}{\varepsilon}(1 - 2v^k(x)) \ ,$$

a solution is immediately found as

$$u^k(x) = \left\{ \begin{array}{ll} 0 & \text{if } \zeta^k(x) \geq 0, \\ 1 & \text{otherwise.} \end{array} \right.$$

In other terms, u^k is the characteristic function of the level-set $\{\zeta^k < 0\}$, denoted by

$$u^k = \chi_{\{\zeta^k<0\}}.$$

This algorithm ensures a decrease of the objective function at each iteration. Moreover, each cluster point (in the weak-$*$ topology for u and the H^1 norm topology for v) is a stationary point. Of course, as the coupled problem in (u, v) is not convex, local minimizers are theoretically not excluded. An outstanding feature of this algorithm is that u^k is always a characteristic function during the iterations.

2.4.2. *Discrete version.* For solving the boundary value problem (2.8) we use finite elements on a Cartesian mesh with $Q1$ shape functions. The mesh nodes coincide with the image pixels, and without any loss of generality, the mesh size is fixed to 1. The discrete counterparts of the variables u and v are therefore vectors of $\mathbb{R}^n$ where n is the number of pixels. Denoting by K and M the stiffness and mass matrices for the (negative) Laplacian, the discrete problem reads

$$I = \min_{u\in\mathbb{R}^n} \min_{v\in\mathbb{R}^n} \left\{ Mu \cdot g + \alpha \left[\varepsilon Kv \cdot v + \frac{1}{\varepsilon}\left(Mv \cdot v + Mu \cdot (\mathbb{1} - 2v) \right) \right] \right\}.$$

In the above expression, the dot stands for the standard dot product of $\mathbb{R}^n$, and $\mathbb{1} = (1, ..., 1)^T$.

In this framework the two steps of the algorithm consist of solving the linear system

$$(\varepsilon^2 K + M)v^k = u^{k-1}, \tag{2.22}$$

and setting

$$\zeta^k = M\left(g + \frac{\alpha}{\varepsilon}(\mathbb{1} - 2v^k) \right) \ , \tag{2.23}$$

$$u^k = \chi_{\{\zeta^k<0\}}. \tag{2.24}$$

This algorithm enjoys the same monotonicity and convergence properties as its infinite-dimensional counterpart. The stopping criterion we choose corresponds to a relative variation of the vector u (in squared ℓ^2 norm) between two successive iterations smaller than some threshold, fixed to 10^{-5}.

The linear system (2.22) is solved in an efficient way with the help of the fast Fourier transform (FFT), according to the following procedure. First, symmetries of the image u are performed in both axial directions, in such a way that we consider a domain $\hat{\Omega}$ with double width and height. Then periodicity conditions are assumed on the boundary of $\hat{\Omega}$, which is a convenient way of implementing Neumann boundary conditions. In this framework, the matrix products Kv and Mv represent

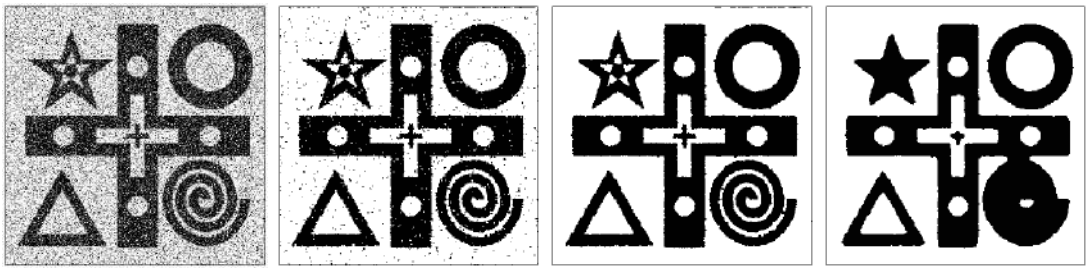


FIGURE 3. Left: original black and white noisy image (320×320 pixels). From left to right: obtained solutions for $\alpha = 2.10^{-4}\sqrt{n}$ (incomplete denoising), $\alpha = 2.10^{-3}\sqrt{n}$ (optimal perimeter weight), $\alpha = 9.10^{-3}\sqrt{n}$ (excessive perimeter weight), respectively.

FIGURE 4. From left to right: original image (320×480 pixels), initialization ($\varepsilon = \infty$), and obtained solutions for $\varepsilon = \varepsilon_{max}/16$ and $\varepsilon = \varepsilon_{min}$, respectively (with $\alpha = 9.10^{-3}\sqrt{n}$).

bidimensional discrete convolutions, which are easily transferred to the Fourier domain. The Fourier transform of v is thus obtained, and v itself is retrieved by inverse FFT.

2.4.3. *Update of ε.* The parameter ε has the dimension of a length. In fact, in view of (2.21), it is a characteristic width of the diffuse interface represented by the *slow* variable v. Thus we start with a characteristic size of Ω, namely $\varepsilon_0 = \varepsilon_{max} = \sqrt{n}$. Then we divide ε by two at each iteration of an outer loop, that is, we choose $\varepsilon_m = \varepsilon_{max}/2^m$. In order to approximate (2.21) properly, ε must not be taken significantly smaller than the grid size. Thus we stop the algorithm as soon as $\varepsilon_m \leq \varepsilon_{min} = 0.1$. In fact, numerical tests show that almost no more evolution occurs when ε goes below this value.

2.4.4. *Initialization.* The initialization of u is performed by the expressions (2.23)-(2.24) with $\varepsilon \to +\infty$, that is, we set

$$\zeta^0 = Mg, \qquad u^0 = \chi_{\{\zeta^0<0\}}.$$

2.5. **Numerical examples.** In Figure 3, we present an example of binary image classification using the above procedure. The function g is defined by (2.5) with the two levels c_1 and c_2 corresponding to black and white. Since this particular example aims at denoising, $p = 1$ is chosen in (2.3)-(2.5).

In Figure 4, we present another example, also with black and white levels and L^1 norm. Intermediate iterations are displayed.

3. Multilabel minimal partition

In this section we extend the previous approach to the multilabel problem. We define the set

$$\mathcal{E}_N = \left\{ (u_1, ..., u_N) \in \mathcal{E}^N, \sum_{i=1}^{N} u_i = 1 \right\}.$$

This is the set of N-tuples of characteristic functions $(u_1, ..., u_N) = (\chi_{\Omega_1}, ..., \chi_{\Omega_N})$, where $(\Omega_1, ..., \Omega_N)$ form a partition of Ω. Given functions $g_1, ..., g_N \in L^1(\Omega)$, the minimal partition problem (1.1) reads

$$I := \min_{(u_1,...,u_N)\in\mathcal{E}_N} \left\{ \mathcal{I}(u) := \sum_{i=1}^{N} \langle u_i, g_i \rangle + \alpha F(u_i) \right\}. \tag{3.1}$$

In the same way as in Section 2, this problem is approximated by

$$I_\varepsilon := \min_{(u_1,\dots,u_N)\in\tilde{\mathcal{E}}_N} \left\{ \mathcal{I}_\varepsilon(u) := \sum_{i=1}^N \langle u_i, g_i\rangle + \alpha \tilde{F}_\varepsilon(u_i) \right\} \tag{3.2}$$

with

$$\tilde{\mathcal{E}}_N = \left\{ (u_1,...,u_N) \in \tilde{\mathcal{E}}^N, \sum_{i=1}^N u_i = 1 \right\}.$$

The following result is an extension of its binary counterpart presented in Section 2, namely Theorem 2.5.

Theorem 3.1. *Let $u^\varepsilon = (u_i^\varepsilon)_{1\le i\le N}$ be an approximate minimizer of (3.2), i.e.,*

$$\mathcal{I}_\varepsilon(u^\varepsilon) = \sum_{i=1}^N \langle u_i^\varepsilon, g_i\rangle + \alpha\tilde{F}_\varepsilon(u_i^\varepsilon) \le I_\varepsilon + \lambda_\varepsilon,$$

with $u_i^\varepsilon \in \tilde{\mathcal{E}}_N$ and $\lim_{\varepsilon\to 0}\lambda_\varepsilon = 0$. Then we have

$$\mathcal{I}_\varepsilon(u^\varepsilon) \to I.$$

Moreover, the sequence (u^ε) admits cluster points, and each of these cluster points is a minimizer of (3.1).

Proof. As usual for this kind of results, the proof relies on the Γ-convergence of the functionals $u \in \tilde{\mathcal{E}}_N \mapsto \sum_{i=1}^N \tilde{F}_\varepsilon(u_i)$ along with the compactness of the sequences of approximate minimizers (see, e.g., Theorem 12.1.1 of [8]). The lim inf inequality of the Γ-convergence immediately passes to the sum. As to the lim sup inequality, one has to construct a recovery sequence which belongs to $\tilde{\mathcal{E}}_N$, which is not automatically achieved by gathering independent recovery sequences for each variable u_i. However, the pointwise convergence result of Theorem 2.8 shows that the constant sequence is a trivial recovery sequence.

From this latter property, there exists $\bar{u}^\varepsilon \in \tilde{\mathcal{E}}_N$ such that $\bar{u}^\varepsilon \to (0,...,0,1)$ in $L^1(\Omega)^N$ and $\sum_{i=1}^N \tilde{F}_\varepsilon(\bar{u}_i^\varepsilon) \to 0$. Let (u^ε) be a sequence of approximate minimizers of (3.2). Using

$$\sum_{i=1}^N \left[\langle u_i^\varepsilon, g_i\rangle + \alpha\tilde{F}_\varepsilon(u_i^\varepsilon)\right] \le \sum_{i=1}^N \left[\langle \bar{u}_i^\varepsilon, g_i\rangle + \alpha\tilde{F}_\varepsilon(\bar{u}_i^\varepsilon)\right] + \lambda_\varepsilon$$

we derive that $\tilde{F}_\varepsilon(u_i^\varepsilon)$ is uniformly bounded for every i. The equicoercivity property of Theorem 2.4 implies that, for a subsequence, $u_i^\varepsilon \to u_i \in \mathcal{E}$ in $L^1(\Omega)$ for each i. As $u^\varepsilon \in \tilde{\mathcal{E}}_N$ we infer that $u = (u_1,...,u_N) \in \mathcal{E}_N$. □

3.1. **Algorithm.** For ε fixed we have to solve the approximate problem

$$I_\varepsilon = \min_{(u_1,\dots,u_N)\in\tilde{\mathcal{E}}_N} \inf_{(v_1,\dots,v_N)\in H^1(\Omega)^N} \sum_{i=1}^N \left\{ \langle u_i, g_i\rangle + \alpha\left[\varepsilon\|\nabla v_i\|^2_{L^2(\Omega)} + \frac{1}{\varepsilon}\left(\|v_i\|^2_{L^2(\Omega)} + \langle u_i, 1-2v_i\rangle\right)\right]\right\}.$$

We use again an alternating minimization algorithm with respect to the two N-tuples of variables $(u_1,...,u_N)$ and $(v_1,...,v_N)$. The superscript k is again used to designate these vectors at iteration k.

(1) The minimization with respect to $(v_1,...,v_N)$ consists in solving the N boundary value problems

$$\left\{ \begin{array}{rcll} -\varepsilon^2\Delta v_i^k + v_i^k &=& u_i^{k-1} & \text{in } \Omega, \\ \partial_n v_i^k &=& 0 & \text{on } \partial\Omega. \end{array}\right. \tag{3.3}$$

(2) The minimization with respect to $(u_1,...,u_N)$ is a linear programming problem. Minimizers can be found by exploring the vertices of the polyhedron $\tilde{\mathcal{E}}_N$, that is, $\mathcal{E}_N$. The practical procedure is the following. Set

$$\zeta_i^k = g_i + \frac{\alpha}{\varepsilon}(1-2v_i^k).$$

At each point $x\in\Omega$ we find an index $i(x)$ such that

$$\zeta_{i(x)}^k(x) = \min\{\zeta_1^k(x),...,\zeta_N^k(x)\}. \tag{3.4}$$

We then set

$$u_i^k(x) = \begin{cases} 1 & \text{if } i = i(x) \\ 0 & \text{otherwise.} \end{cases}$$

The discrete counterpart is easily obtained in the same fashion as in the binary case. Again, an outstanding feature of this algorithm is that the functions $u_1^k, ..., u_N^k$ are always characteristic functions of a partition of Ω.

3.2. **Numerical validation.** The three examples of Fig. 5 are taken from and compared to [17]. Let $E_0, E_1, ..., E_N$ be a given partition of Ω. We define g_i, $i = 1, ..., N$, by

$$g_i = \sum_{\substack{0 \le j \le N \\ j \ne i}} \chi_{E_j} = 1 - \chi_{E_i}.$$

This means that, in the set E_i, $i \ge 1$, the label i is favored, whereas in the set E_0 there is no preference, or, said otherwise, no information on which label to choose. From this point of view the problem is related to the wide field of image inpainting, see e.g. [22] and the references therein. In the subsequent examples the domain Ω is the unit square with a 400×400 discretization and we choose $\alpha = 0.1\sqrt{n}$.

3.2.1. *Example with 3 labels.* The partition is as shown in Fig. 5(a) (left). The set E_0 is the black disc, while each E_i, $i = 1, 2, 3$, is assigned to a specific color, namely red, green and blue, respectively. We retrieve the triple junction, which is known to be the theoretical solution [17].

3.2.2. *Examples with 4 labels.* We consider now the partitions of Fig. 5(b) and 5(c), with E_0 the black disc and 4 other subsets. In these cases, the solution is not unique, and the algorithm chooses a particular one. This choice stems from the selection of a particular minimizer in (3.4). This is in contrast with the results obtained in [17] which, due to the convexification method employed, are mixtures of minimizers.

4. Multilabel image classification

4.1. **Formulation as a minimal partition problem.** We come back to the classification problem presented in Section 2.1, but this time with N grey levels $c_1, ..., c_N \in [0, 1]$. We are given an image $f \in L^\infty(\Omega, [0, 1])$, and consider the piecewise constant image

$$w = \sum_{i=1}^{N} u_i c_i,$$

where each u_i is the characteristic function of a subset Ω_i of Ω such that $(\Omega_1, ..., \Omega_N)$ forms a partition of Ω. We have for any L^p norm on Ω:

$$\|w - f\|_{L^p(\Omega)}^p = \left\| \sum_{i=1}^{N} u_i c_i - f \right\|_{L^p(\Omega)}^p = \left\| \sum_{i=1}^{N} u_i (c_i - f) \right\|_{L^p(\Omega)}^p = \sum_{i=1}^{N} \int_\Omega u_i |c_i - f|^p dx.$$

The difference between the piecewise constant and original images is measured by:

$$\|w - f\|_{L^p(\Omega)}^p = \sum_{i=1}^{N} \langle u_i, g_i \rangle, \qquad g_i = |c_i - f|^p.$$

When the levels c_i are fixed, we can directly apply the algorithm of Section 3.1.

4.2. **Update of levels.** However, it is often desirable to determine the grey levels within the classes automatically. Thus, we include a third step in the alternating minimization algorithm, consisting of solving

$$\min_{(c_1, ..., c_N) \in [0,1]^N} \sum_{i=1}^{N} \langle u_i, g_i \rangle = \sum_{i=1}^{N} \int_\Omega u_i |c_i - f|^p dx.$$

This problem is separable in its variables $c_1, ..., c_N$, and each c_i must satisfy

$$c_i \in \operatorname*{argmin}_{c \in [0,1]} \int_\Omega u_i |c - f|^p dx.$$

Note that, since $0 \le f \le 1$, the constraint $0 \le c \le 1$ can be removed. We distinguish between the two cases of practical interest, namely $p = 2$ and $p = 1$.

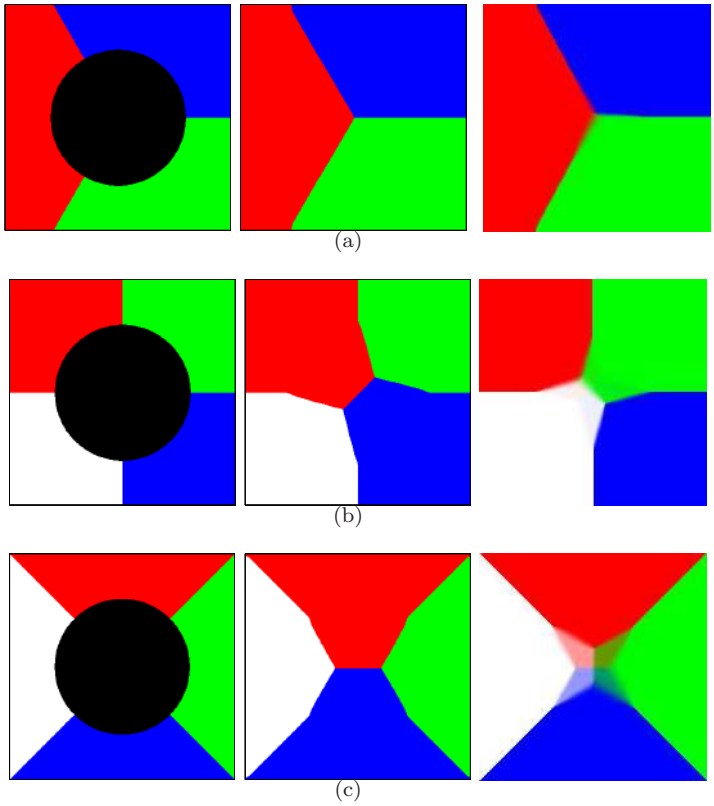


FIGURE 5. 5(a) Triple point: given partition (left), obtained solution with our algorithm (middle), and solution from [17] (right); 5(b) and 5(c) Two quadruple points: given partition (left), obtained solution with our algorithm (middle), and solution from [17] (right).

- $p = 2$**.** This is a standard problem which results in computing the arithmetic mean

$$c_i = \frac{\int_\Omega u_i f dx}{\int_\Omega u_i dx}.$$

- $p = 1$**.** Minimizing a sum of L^1 distances classically amounts to computing a median, but here, due to the weights u_i this is a little more involved. Details of the procedure in the discrete setting are given in Appendix A. Note that there may be several solutions. In this case we take the half-sum of the extreme points of the minimizing set.

For the initialization, the levels are equidistributed in $[0, 1]$, i.e., we choose

$$c_i^0 = \frac{i-1}{N-1}.$$

4.3. **Examples.** We begin with two synthetic examples taken from [28], for comparison. In [28] the authors consider a multiphase "sine-sinc" extension of the classical Modica-Mortola perimeter approximation applied to the same $TV - L^2$ model as ours. Therefore we choose $p = 2$. The example depicted in Fig. 6 is a 3 label classification of a noisy image, and Fig. 7 shows a 5 label classification. In each case we have represented the "reconstructed" image $\sum_{i=1}^N u_i c_i$ but, of course, our algorithm provides the N characteristic functions $u_1, ..., u_N$. However the fact that the $u_i's$ are indeed characteristic functions is clearly visible in the images, in contrast to [28].

Examples with the L^1 norm are shown in Figs. 8-10. In Fig. 8, an example of denoising with 3 labels is shown. In Fig. 9, the classification of the swan picture is shown with 2 and 3 labels. This result is to be compared with Fig. 4. We have chosen a smaller value of α to take into account the

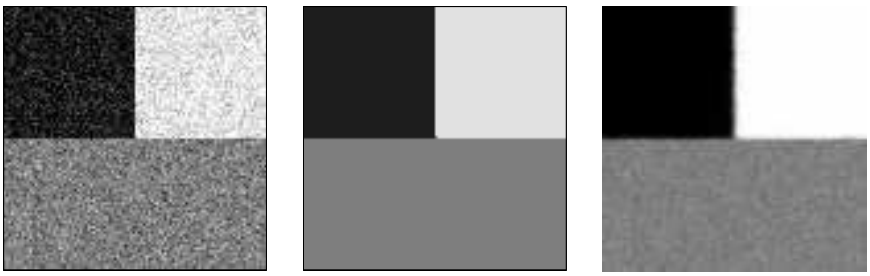


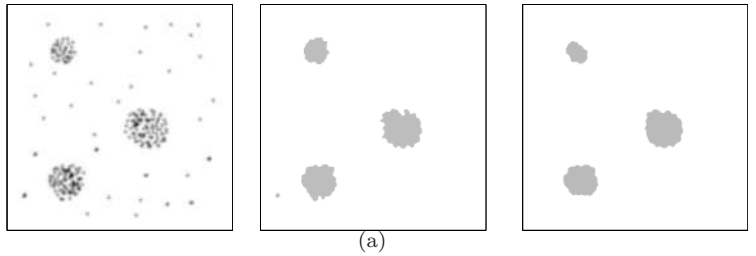
FIGURE 6. Greylevel image denoising with L^2 norm: original image (left, 100×100 pixels), solution with 3 labels and $\alpha = 10^{-3}\sqrt{n}$ (middle), solution taken from [28] (right).

FIGURE 7. Greylevel image denoising with L^2 norm: original image (left, 240×240 pixels), solution with 5 labels and $\alpha = 10^{-3}\sqrt{n}$ (middle), solution taken from [28] (right).

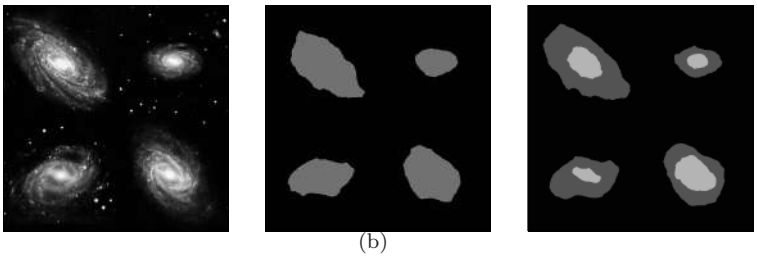
FIGURE 8. Greylevel image denoising with L^1 norm: original image (left), noisy image (middle), obtained solution with 3 labels and $\alpha = 5.10^{-3}\sqrt{n}$ (right).

FIGURE 9. From left to right: original image (320×480 pixels) and obtained solutions with 2 and 3 labels, respectively (L^1 norm, $\alpha = 5.10^{-3}\sqrt{n}$).

decrease of the fidelity term due to the update of levels. In Fig. 10 two examples with disconnected geometric elements are featured.



(a)



(b)

FIGURE 10. 10(a) Image classification with 2 labels and L^1 norm: original image (left, 512×512 pixels), obtained images for $\alpha = 2.10^{-3}\sqrt{n}$ (middle) and $\alpha = 5.10^{-3}\sqrt{n}$ (right); 10(b) Image classification with L^1 norm, $\alpha = 5.10^{-3}\sqrt{n}$: original image (left, 600×600 pixels), obtained images with 2 labels (middle) and with 3 labels (right).

5. Multilabel classification of color images

The original image f is represented by the three channels $(f_1, f_2, f_3) \in L^\infty(\Omega, [0,1])^3$ representing the intensity of red, green and blue, respectively. Each phase Ω_i is associated to a color $(c_{i1}, c_{i2}, c_{i3}) \in [0,1]^3$ in the same RGB system. The reconstructed image $w = (w_1, w_2, w_3)$ is given by

$$w_j = \sum_{i=1}^{N} u_i c_{ij},$$

where u_i is the characteristic function of Ω_i. We have for each channel

$$\|w_j - f_j\|_{L^p}^p = \left\| \sum_{i=1}^{N} u_i c_{ij} - f_j \right\|_{L^p}^p = \left\| \sum_{i=1}^{N} u_i (c_{ij} - f_j) \right\|_{L^p}^p = \sum_{i=1}^{N} \int_\Omega u_i |c_{ij} - f_j|^p.$$

The difference between the segmented and original images is measured by

$$\sum_{j=1}^{3} \|w_j - f_j\|_{L^p}^p = \sum_{i=1}^{N} \langle u_i, g_i \rangle, \qquad g_i = \sum_{j=1}^{3} |c_{ij} - f_j|^p.$$

We then apply the same algorithm as in Section 4. Note that the geometrical variable $u = (u_1, ..., u_N)$ as well as the auxiliary variable $v = (v_1, ..., v_N)$ remain N-dimensional vectors, and the update of levels is separable in the channels.

The first example (see Figure 11) is a two-label problem, initialized with a pure black phase and a pure white phase.

In the second example (see Figure 12), we first choose 2 labels, then 5 labels. In this latter case, the phases are initialized by the pure colors black, red, green, blue and white. We observe that only 4 labels remain at convergence, which is an effect of the perimeter penalization.

In the third example (see Figure 13), we choose again 5 labels. Results obtained with different values of α are depicted.

6. Anisotropy-based image classification

We come back to a greyscale image $f \in L^\infty(\Omega, [0,1])$. We are given N vectors $\xi_1, ..., \xi_N \in \mathcal{S}_2$, where $\mathcal{S}_2$ is the unit sphere of $\mathbb{R}^2$. In order to detect fluctuations oriented along the direction ξ_i, we

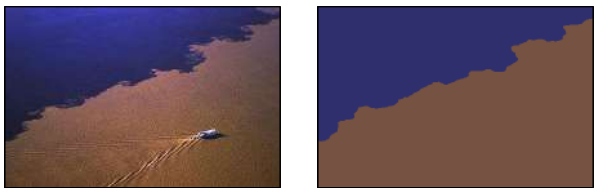


FIGURE 11. Unsupervised color image classification with 2 labels and L^1 norm, $\alpha = 5.10^{-3}\sqrt{n}$: original image (left, 327×500 pixels), obtained image (right).

FIGURE 12. Unsupervised color image classification with L^1 norm, $\alpha = 5.10^{-3}\sqrt{n}$: original image (left, 231×221 pixels), obtained image with 2 labels (middle), obtained image with 5 labels (right).

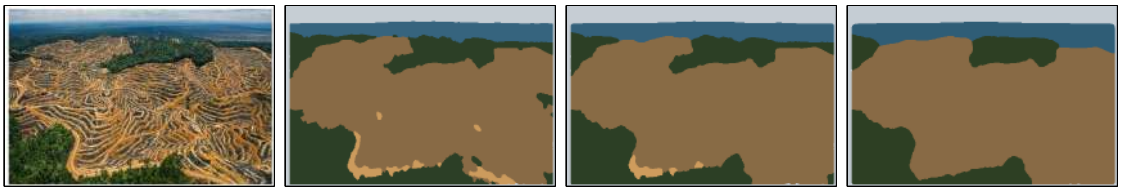


FIGURE 13. Unsupervised color image classification with L^1 norm and 5 labels, from left to right: original image (320×584 pixels) and obtained images for $\alpha = 5.10^{-3}\sqrt{n}$, $\alpha = 10^{-2}\sqrt{n}$, $\alpha = 2.10^{-2}\sqrt{n}$.

first regularize f by solving

$$\left\{ \begin{array}{rcccc} -\Delta\phi + \phi & = & f & \text{in} & \Omega, \\ \partial_n \phi & = & 0 & \text{on} & \partial\Omega, \end{array} \right. \tag{6.1}$$

then we set

$$g_i = -(\nabla\phi.\xi_i)^2. \tag{6.2}$$

In practice, (6.1) is solved by finite elements, in the same way as described in Section 2.4.2. Then (6.2) is computed at each node using a discrete gradient based on finite differences. We point out that the function g_i can be interpreted as the topological sensitivity of the energy functional

$$\frac{1}{2}\|\nabla\phi\|^2_{L^2(\Omega)} + \frac{1}{2}\|\phi - f\|^2_{L^2(\Omega)}$$

with respect to the creation of an insulating crack normal to ξ_i, see [6] and [12] for various applications of this concept in image processing.

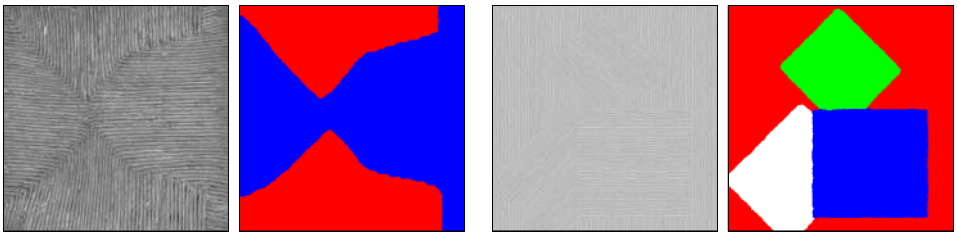
We then apply the multilabel minimal partition algorithm of Section 3 with the obtained functions $(g_1, ..., g_N)$.

In Figure 14(a) we show an example with 2 labels. We want to detect vertical and horizontal fluctuations, therefore we choose the vectors (ξ_1, ξ_2) as

$$\xi_1 = (1, 0)^T,\ \xi_2 = (0, 1)^T.$$

In Figure 14(b) we show an example with 4 labels, where the vectors ξ_i are given by

$$\xi_1 = (1, 0)^T,\ \xi_2 = \frac{\sqrt{2}}{2}(1, 1)^T, \xi_3 = (0, 1)^T,\ \xi_4 = \frac{\sqrt{2}}{2}(-1, 1)^T.$$



(a) (b)

FIGURE 14. (a) Anisotropy-based image classification with 2 labels: original image (left, 951×1000 pixels), obtained partition for $\alpha = 10^{-5}\sqrt{n}$ (right); (b) Anisotropy-based image classification with 4 labels: original image (left, 512×512 pixels), obtained partition for $\alpha = 10^{-6}\sqrt{n}$ (right).

7. Image deblurring

In this section we apply a variant of our minimal partition model for binary image deconvolution. The blurring kernel is represented by a linear and continuous operator $A : L^2(\Omega) \to L^2(\Omega)$ such that $A1 = 1$. The given blurred (greyscale) image is $f \in L^\infty(\Omega, [0,1])$, and the reconstructed image is $w = c_1 u + c_2(1-u)$, with $u \in \mathcal{E}$ and $0 \leq c_1 < c_2 \leq 1$. We have

$$Aw = c_1 Au + c_2 A(1-u) = (c_1 - c_2)Au + c_2,$$

and the deblurring problem reads

$$I := \min_{u \in \mathcal{E}} \{\mathcal{I}(u) = \Phi(u) + \alpha F(u)\} \tag{7.1}$$

with $\Phi : L^2(\Omega) \to L^2(\Omega)$ defined by

$$\Phi(u) = \|(c_1 - c_2)Au + c_2 - f\|^2_{L^2(\Omega)}.$$

In our model the blurring kernel A is assumed to be known, which occurs in some practical applications, like when the blur is generated by an optical device. Algorithms for blind deconvolution can be found in [22, 24], using the Modica-Mortola functional or wavelet-based extensions.

In our case we approximate (7.1) by

$$I_\varepsilon = \min_{u \in \tilde{\mathcal{E}}} \left\{ \mathcal{I}_\varepsilon(u) = \Phi(u) + \alpha \tilde{F}_\varepsilon(u) \right\}, \tag{7.2}$$

and use the same continuation procedure with respect to ε as before. Using that the function Φ is at the same time continuous on $\tilde{\mathcal{E}}$ endowed with the L^1 norm topology and, by convexity, lower semi-continuous on $\tilde{\mathcal{E}}$ endowed with the weak-$*$ topology of $L^\infty(\Omega)$, analogues to Theorem 2.5 and Proposition 2.9 can be straightforwardly derived, see [7] for full justifications. If A is compact, which happens whenever A is the convolution by a smoothing kernel, then the extension of the function $\Phi : \mathcal{E} \to \mathbb{R}$ to $\tilde{\mathcal{E}}$ can be argued by the following Proposition.

Proposition 7.1. *If $A : L^2(\Omega) \to L^2(\Omega)$ is compact, then Φ is continuous on $\tilde{\mathcal{E}}$ endowed with the weak-$*$ topology of $L^\infty(\Omega)$, and it is the relaxation of the function*

$$u \in \tilde{\mathcal{E}} \mapsto \begin{cases} \Phi(u) \text{ if } u \in \mathcal{E}, \\ +\infty \text{ if } u \notin \mathcal{E}. \end{cases}$$

Proof. The weak continuity is a direct consequence of the compactness of A. To obtain the relaxation result, it suffices (see e.g. Proposition 11.11 of [8]) to prove that for any $u \in \tilde{\mathcal{E}}$ there exists a sequence $(u_n) \in \mathcal{E}$ such that $u_n \rightharpoonup u$ and $\Phi(u) \geq \limsup_{n \to +\infty} \Phi(u_n)$. The existence of a sequence of characteristic functions such that $u_n \rightharpoonup u$ is obtained by a standard construction, see e.g. Proposition 7.2.14 of [27], while the second assertion stems from the weak continuity of Φ. □

For solving (7.2) at ε fixed, applying directly an alternating algorithm based on the formulation (2.6) would not be well-suited since the minimization with respect to u would be in itself a difficult problem. We could use instead a projected gradient algorithm, however, this would require a line search at every iteration. In fact, an efficient alternating algorithm can still be used thanks to the following lemma.

Lemma 7.2. *Let X be a Hilbert space and $\Phi : X \to \mathbb{R}$ be a differentiable function such that $\nabla\Phi$ is λ-Lipschitz. We have for all $u \in X$:*

$$\Phi(u) = \inf_{\hat{u}\in X} \Phi(\hat{u}) + \langle \nabla\Phi(\hat{u}), u - \hat{u}\rangle + \frac{\lambda}{2}\|u - \hat{u}\|^2.$$

Proof. On choosing $\hat{u} = u$, we immediately get that

$$\inf_{\hat{u}\in X} \Phi(\hat{u}) + \langle \nabla\Phi(\hat{u}), u - \hat{u}\rangle + \frac{\lambda}{2}\|u - \hat{u}\|^2 \leq \Phi(u).$$

The reverse inequality stems from

$$\begin{aligned}
\Phi(u) - \Phi(\hat{u}) &= \int_0^1 \langle \nabla\Phi(\hat{u} + t(u - \hat{u})), u - \hat{u}\rangle dt \\
&= \langle \nabla\Phi(\hat{u}), u - \hat{u}\rangle + \int_0^1 \langle \nabla\Phi(\hat{u} + t(u - \hat{u})) - \nabla\Phi(\hat{u}), u - \hat{u}\rangle dt \\
&\leq \langle \nabla\Phi(\hat{u}), u - \hat{u}\rangle + \frac{\lambda}{2}\|u - \hat{u}\|^2,
\end{aligned}$$

and the proof is complete. □

From Lemma 7.2 and (2.6), (7.2) becomes:

$$\begin{aligned}
I_\varepsilon = \min_{u\in\tilde{\mathcal{E}}} \inf_{\hat{u}\in L^2(\Omega)} \inf_{v\in H^1(\Omega)} \Phi(\hat{u}) + \langle \nabla\Phi(\hat{u}), u - \hat{u}\rangle + \frac{\lambda}{2}\|u - \hat{u}\|^2_{L^2(\Omega)} \\
+ \alpha\left\{\varepsilon\|\nabla v\|^2_{L^2(\Omega)} + \frac{1}{\varepsilon}\left(\|v\|^2_{L^2(\Omega)} + \langle u, 1 - 2v\rangle\right)\right\}. \quad (7.3)
\end{aligned}$$

We use an alternating minimization algorithm with respect to the three variables $u, \hat{u}, v$. The minimization with respect to v amounts to solving (2.21). The minimization with respect to $\hat{u}$ is simply achieved by setting $\hat{u}^k = u^{k-1}$. The minimization with respect to u is spatially uncoupled. A short calculation gives

$$u^k(x) = P_{[0,1]}\left(\hat{u}^k(x) - \frac{1}{\lambda}\left(\nabla\Phi(\hat{u}^k)(x) + \frac{\alpha}{\varepsilon}(1 - 2v^k)\right)\right),$$

with

$$P_{[0,1]}(\varphi) = \max(0, \min(1, \varphi)).$$

Here we have

$$\nabla\Phi(u) = 2(c_1 - c_2)A^*[(c_1 - c_2)Au + c_2 - f],$$

with A^* the adjoint operator of A, hence we can choose any $\lambda \geq 2(c_1 - c_2)^2\|A^*A\|$.

In our experiments we take the operator $A = A_0^q$, where A_0 is the discrete convolution operator by the kernel

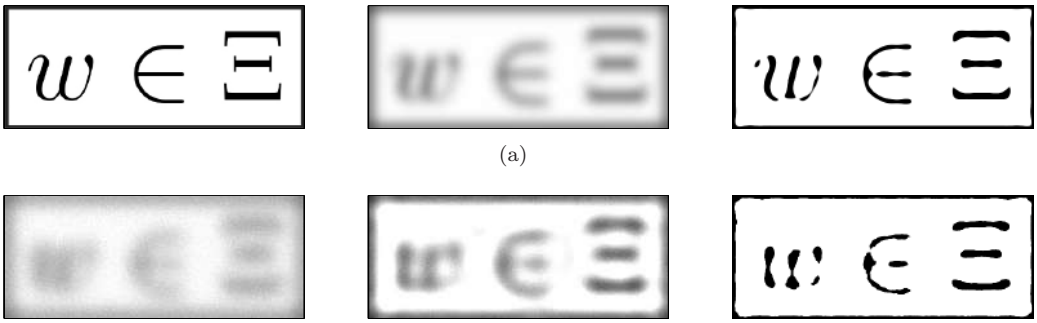
$$\kappa = \begin{pmatrix} 0 & a & 0 \\ a & 1 - 4a & a \\ 0 & a & 0 \end{pmatrix}, \qquad a = 0.15.$$

We recall that the unknown u is symmetrized and periodized, hence the convolution is performed without boundary effect. In addition, the computation of the product $Au = A_0^q u$ is efficiently performed through the FFT, without actually computing and storing the matrix of the operator A. We have that $\|A\| = 1$, therefore we choose $\lambda = 2(c_1 - c_2)^2$. Assuming that $c_1 \leq f \leq c_2$, the initialization is done by $u = (f - c_1)/(c_2 - c_1)$. The stopping criterion for the minimization of (7.3) is a threshold on the squared relative variation of u fixed to 10^{-6}.

In Figure 15, f is of the form $P_{[0,1]}\left(A(u^*c_1 + (1 - u^*)c_2 + \nu)\right)$, with u^* a characteristic function and ν a random noise. The grey levels are black and white.

The example featured in Fig. 16 also shows that though the underlying minimal partition algorithm is intended for the classification of plane regions, the proposed method also permits to restore a blurred and noisy text. In order to retrieve the fine details of the text, it is necessary to choose the penalization coefficient α quite small, hence it is difficult to accommodate with a high level of noise. In Fig. 16(b), a better result is obtained when the decrease of ε is stopped before reaching the value ε_{min}, although for this value the obtained image is not binary.

FIGURE 15. Deblurring and denoising: original image (left, 257×257 pixels), damaged image with blur and noise effects (middle), reconstructed image for $\alpha = 2.10^{-4}\sqrt{n}$ (right).



(a)

(b)

FIGURE 16. Deblurring of the text "$w \in \Xi$". 16(a) Original image (left, 245×600 pixels), blurred image (middle), reconstructed image for $\alpha = 2.10^{-5}\sqrt{n}$ (right). 16(b) blurred image with noise (left), its restoration with $\alpha = 2.10^{-5}\sqrt{n}$ for $\varepsilon = 1.99$ (middle) and for $\varepsilon = \varepsilon_{min}$ (right).

8. Concluding remarks

Whereas the computer vision community is very active in developing powerful algorithms, few of these approaches are theoretically justified. In this paper we propose a mathematically sound method to perform optimal partitions and apply it to image restoration and classification of greyscale and color images. Let us emphasize that the proposed continuation approach approximates the continuous total variation and not its discretized version. This is achieved through the solving of linear partial differential equations with constant coefficients, which is performed by finite elements and fast Fourier transforms. This task is from far the dominant part of the computational effort of our algorithms. Its implementation could certainly be further improved, but code optimization was not the aim of this work.

We consider the present paper as a first step of an ongoing work where an image partition algorithm and a gradient-free approximation of the perimeter is proposed to address piecewise constant image restoration and/or classification. In particular, we have applied our algorithm to denoising, deblurring and supervised texture identification problems, as well as for inpainting. Future extensions of the method could be multilabel deblurring and unsupervised texture identification.

Appendix A. Weighted median

Let $(x_1, ..., x_n) \in \mathbb{R}^n$, $(\alpha_1, ..., \alpha_n) \in \mathbb{R}^n$ be given. We assume that the x_i's are numbered in increasing order. We want to minimize

$$V(x) = \sum_{i=1}^{n} \alpha_i |x - x_i|.$$

This function is clearly convex and affine on each interval $[x_i, x_{i+1}]$. Therefore, the minimizing set is an interval of the form $[x_l, x_r]$, $1 \le l \le r \le n$. The subdifferential of V at the point x_j is

$$\partial V(x_j) = \sum_{i<j} \alpha_i - \sum_{i>j} \alpha_i + \alpha_j[-1, 1].$$

Then x_j is a minimizer of V if and only if $0 \in \partial V(x_j)$, i.e.

$$-\alpha_j \le \sum_{i<j} \alpha_i - \sum_{i>j} \alpha_i \le \alpha_j. \tag{A.1}$$

We obtain x_l and x_r by checking for each x_j if it satisfies (A.1).

Acknowledgements. The work of the third author has been supported in part by Fundação para a Ciência e a Tecnologia (Ciência 2007) and in part by the ERC Advanced Grant "Quasistatic and Dynamic Evolution Problems in Plasticity and Fracture" (grant agreement no. 290888).

References

[1] M. Abramowitz and I. A. Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55 of *National Bureau of Standards Applied Mathematics Series*. For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., 1964.

[2] G. Allaire. *Shape optimization by the homogenization method*, volume 146 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2002.

[3] L. Alvarez, L. Baumela, P. Márquez-Neila, and P. Henríquez. A real time morphological snakes algorithm. *IPOL*, 2012.

[4] L. Ambrosio, N. Fusco, and D. Palara. *Functions of bounded variation and free discontinuity problems*. Oxford Mathematical Monographs. Oxford, 2000.

[5] L. Ambrosio and V. M. Tortorelli. Approximation of functionals depending on jumps by elliptic functionals via Γ-convergence. *Comm. Pure Appl. Math.*, 43(8):999–1036, 1990.

[6] S. Amstutz, I. Horchani, and M. Masmoudi. Crack detection by the topological gradient method. *Control Cybernet.*, 34(1):81–101, 2005.

[7] S. Amstutz and N. Van Goethem. Topology optimization methods with gradient-free perimeter approximation. *Interfaces and Free Boundaries*, 14:401–430, 2012.

[8] H. Attouch, G. Buttazzo, and G. Michaille. *Variational analysis in Sobolev and BV spaces*, volume 6 of *MPS/SIAM Series on Optimization*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2006. Applications to PDEs and optimization.

[9] J.-F. Aujol and G. Aubert. Optimal partitions, regularized solutions, and application to image classification. *Applicable Analysis*, 84(1):15–35, 2005.

[10] J.-F. Aujol, G. Gilboa, T. Chan, and S. Osher. Structure-texture image decomposition-modeling, algorithms,and parameter selection. *Int. J. Comp. Vision*, 67(1):85–104, 2005.

[11] D. Auroux, L. Jaafar Belaid, and M. Masmoudi. A topological asymptotic analysis for the regularized grey-level image classification problem. *ESAIM, Math. Model. Numer. Anal.*, 41(3):607–625, 2007.

[12] D. Auroux and M. Masmoudi. Image processing by topological asymptotic expansion. *J. Math. Imaging Vision*, 33(2):122134, 2009.

[13] A. Braides. Γ-*convergence for beginners*, volume 22 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2002.

[14] A. Chambolle. An algorithm for total variation minimization and applications. Special issue on mathematics and image analysis. *J. Math. Imaging Vision*, 20(1-2):89–97, 2004.

[15] A. Chambolle, V. Caselles, D. Cremers, M. Novaga, and T. Pock. An introduction to total variation for image analysis. In *Theoretical foundations and numerical methods for sparse recovery*, volume 9 of *Radon Ser. Comput. Appl. Math.*, pages 263–340. Walter de Gruyter, Berlin, 2010.

[16] A. Chambolle, V. Caselles, and M. Novaga. Total variation in imaging. In O. Scherzer, editor, *Handbook of mathematical methods in imaging*. Springer Reference. Berlin: Springer, 2011.

[17] A. Chambolle, D. Cremers, and T. Pock. A convex approach to minimal partitions. *hal-00630947, version 1*, 11 Oct. 2011.

[18] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vision*, 40(1):120–145, 2011.

[19] T. Chan and L. Vese. Active contours without edges. *IEEE Trans. Image Processing*, 10(2):266–277, 2001.

[20] F. R. K. Chung. *Spectral graph theory*, volume 92 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC, 1997.

[21] G. Dal Maso. *An introduction to* Γ-*convergence*. Progress in Nonlinear Differential Equations and their Applications, 8. Birkhäuser Boston Inc., Boston, MA, 1993.

[22] J. A. Dobrosotskaya and A. L. Bertozzi. A wavelet-Laplace variational technique for image deconvolution and inpainting. *IEEE Trans. Image Process.*, 17(5):657–663, 2008.

[23] J. A. Dobrosotskaya and A. L. Bertozzi. Wavelet analogue of the Ginzburg-Landau energy and its Γ-convergence. *Interfaces Free Bound.*, 12(4):497–525, 2010.

[24] S. Esedoglu. Blind deconvolution of bar code signals. *Inverse Problems*, 20(1):121–135, 2004.
[25] S. Esedoḡlu and S. J. Osher. Decomposition of images by the anisotropic Rudin-Osher-Fatemi model. *Comm. Pure Appl. Math.*, 57(12):1609–1626, 2004.
[26] P. Getreuer. Chan-Vese segmentation. *IPOL*, 2012.
[27] A. Henrot and M. Pierre. *Variation et optimisation de formes*, volume 48 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*.
[28] Y. M. Jung, S. H. Kang, and J. Shen. Multiphase image segmentation via Modica-Mortola phase transition. *SIAM J. Appl. Math.*, 67(5):1213–1232, 2007.
[29] R. V. Kohn and P. Sternberg. Local minimisers and singular perturbations. *Proc. Roy. Soc. Edinburgh Sect. A*, 111(1-2):69–84, 1989.
[30] L. Modica. The gradient theory of phase transitions and the minimal interface criterion. *Arch. Rational Mech. Anal.*, 98(2):123–142, 1987.
[31] L. Modica and S. Mortola. Un esempio di Γ^{-}-convergenza. *Boll. Un. Mat. Ital. B (5)*, 14(1):285–299, 1977.
[32] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Commun. Pure Appl. Math.*, 42(5):577–685, 1989.
[33] S. Osher and J. A. Sethian. Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations. *J. Comput. Phys.*, 79(1):12–49, 1988.
[34] É. Oudet. Approximation of partitions of least perimeter by Γ-convergence: around Kelvin's conjecture. *Exp. Math.*, 20(3):260–270, 2011.
[35] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60(1-4):259–268, 1992.
[36] J. Shi and J. Malik. Normalize cuts and image segmentation. *IEEE Trans. Pat. Anal. Mach. Int.*, 22(8):888–905, 2000.
[37] M. Solci and E. Vitali. Variational models for phase separation. *Interfaces Free Bound.*, 5(1):27–46, 2003.
[38] L. Vese. A study in the BV space of a denoising-deblurring variational problem. *Appl. Math. Optimization*, 44(2):131–161, 2001.
[39] U. von Luxburg. A tutorial on spectral clustering. *Stat. Comput.*, 17(4):395–416, 2007.
[40] C. Zach, D. Gallup, J. Frahm, and M. Niethammer. Fast global labelling for real-time stereo using multiple plabe sweeps. In *Vision, Modeling, and Visualization*. IOS press, 2008.

(S. Amstutz) Laboratoire de Mathématiques d'Avignon, Université d'Avignon - Faculté des Sciences, 33 rue Louis Pasteur, 84000 Avignon, France
E-mail address: samuel.amstutz@univ-avignon.fr

(A. A. Novotny) Laboratório Nacional de Computação Científica LNCC/MCT, Coordenação de Matemática Aplicada e Computacional, Av. Getúlio Vargas 333, 25651-075 Petrópolis - RJ, Brasil.
E-mail address: novotny@lncc.br

(N. Van Goethem) SISSA, Via Bonomea 265, 34136 Trieste, Italia; and Universidade de Lisboa, Faculdade de Ciências, Departamento de Matemática, CMAF, Av. Prof. Gama Pinto 2, 1649-003 Lisboa, Portugal.
E-mail address: vangoeth@sissa.it