MATHEMATICS OF COMPUTATION
Volume 86, Number 306, July 2017, Pages 1769–1791
http://dx.doi.org/10.1090/mcom/3138
Article electronically published on October 20, 2016

A POSTERIORI ERROR CONTROL FOR THE BINARY MUMFORD-SHAH MODEL

BENJAMIN BERKELS, ALEXANDER EFFLAND, AND MARTIN RUMPF

Abstract. The binary Mumford-Shah model is a widespread tool for image segmentation and can be considered as a basic model in shape optimization with a broad range of applications in computer vision, ranging from basic segmentation and labeling to object reconstruction. This paper presents robust a posteriori error estimates for a natural error quantity, namely the area of the non-properly segmented region. To this end, a suitable uniformly convex and non-constrained relaxation of the originally non-convex functional is investigated and Repin's functional approach for a posteriori error estimation is used to control the numerical error for the relaxed problem in the L^2-norm. In combination with a suitable cut out argument, fully practical estimates for the area mismatch are derived. This estimate is incorporated in an adaptive mesh refinement strategy. Two different adaptive primal-dual finite element schemes, a dual gradient descent scheme, and the most frequently used finite difference discretization are investigated and compared. Numerical experiments show qualitative and quantitative properties of the estimates and demonstrate their usefulness in practical applications.

1. Introduction

Since the introduction of the image denoising and edge segmentation model by Mumford and Shah in the late 1980's [36], there has been much effort to find effective and efficient numerical algorithms to compute minimizers of different variants of this variational problem. The original model is based on the functional $E_{\mathrm{MS}}[u,K] = \int_{\Omega\setminus K} |\nabla u|^2 + \alpha(u-u_0)^2 \,\mathrm{d}x + \beta\mathcal{H}^{n-1}(K)$ with $\alpha,\beta>0$, where $u_0 \in L^\infty(\Omega,[0,1])$ is a scalar image intensity on the bounded image domain $\Omega\subset\mathbb{R}^n$, u the reconstructed image intensity and K the associated set of edges, on which the image intensity u jumps. Here, $\mathcal{H}^{n-1}$ denotes the $(n-1)$-dimensional Hausdorff measure. The space of functions of bounded variation $\mathrm{BV}(\Omega)$ turned out to be the proper space to formulate the problem in a mathematically rigorous way. Indeed, existence in the context of the space of special functions of bounded variation $\mathrm{SBV}(\Omega)$ was proved by Ambrosio (see [2, Theorem 4.2]). For details on these spaces we refer to [4]. Restricting u to be piecewise constant instead of piecewise smooth, one is led to a basic and widespread image segmentation model. This model is discussed from a geometric perspective in the book by Morel and Solimini [35]. In the case of just two intensity values $c_1,c_2\in[0,1]$, the associated energy can be rewritten in terms of a characteristic function $\chi\in\mathrm{BV}(\Omega,\{0,1\})$ as

$$E[\chi] = \int_\Omega \theta_1\chi + \theta_2(1-\chi)\,\mathrm{d}x + |\mathrm{D}\chi|(\Omega)\,. \tag{1.1}$$

Received by the editor May 19, 2015 and, in revised form, December 8, 2015 and December 17, 2015.

2010 *Mathematics Subject Classification.* Primary 49M25, 53C22, 65D18, 65L20.

Here, $\theta_i = \frac{1}{\nu}(c_i - u_0)^2$ for $i = 1, 2$, the new weight $\nu = \beta/\alpha$ and the resulting binary model is given by $u = c_1\chi + c_2(1-\chi)$. For fixed χ, one immediately obtains the optimal constants

(1.2)
$$c_1[\chi] = \left(\int_\Omega \chi \,\mathrm{d}x\right)^{-1} \int_\Omega \chi u_0 \,\mathrm{d}x \quad \text{and} \quad c_2[\chi] = \left(\int_\Omega 1-\chi \,\mathrm{d}x\right)^{-1} \int_\Omega (1-\chi) u_0 \,\mathrm{d}x\,.$$

For fixed c_1 and c_2 one aims to minimize the energy over the non-convex set of characteristic functions $\chi \in \mathrm{BV}(\Omega, \{0,1\})$. In the general case, one is interested in a triple (χ, c_1, c_2) as a minimizer of E w.r.t. the set $\mathrm{BV}(\Omega) \times [0,1]^2$. Henceforth, if not otherwise stated we assume the intensity values to be fixed.

Nikolova, Esedoglu and Chan [21] showed that the non-convex minimization problem for χ can be solved via relaxation and thresholding—a breakthrough for both reliable and fast algorithms in computer vision [17, 41]. Here, at first one asks for a minimizer of E over all $u \in \mathrm{BV}(\Omega, [0,1])$ and then thresholds u for any threshold value $s \in [0,1)$ to obtain the solution $\chi = \chi_{[u>s]}$ of the original minimization problem. The relaxed problem coincides with a constrained version of the classical image denoising model by Rudin, Osher and Fatemi (ROF) [45]. Numerical schemes for an effective and efficient minimization of this model have been extensively studied. Making use of a dual formulation, Chambolle [14] introduced an iterative finite difference scheme and proved its convergence. Hintermüller and Kunisch [31] proposed a predual formulation for a generalized ROF model and applied a semismooth Newton method for a regularized variant. Chambolle and Pock [20] deduced a primal-dual algorithm with guaranteed first order convergence and applied their approach to different variational models in BV such as image denoising, deblurring and interpolation. The scheme is based on an alternating discrete gradient scheme for the discrete primal and the discrete dual problems. Bartels [6] used the embedding $\mathrm{BV}(\Omega) \cap L^\infty(\Omega) \hookrightarrow H^{\frac{1}{2}}(\Omega)$ for finite element functions to improve the step-size restriction for BV functionals. Wang and Lucier [47] employed a finite difference approximation of the ROF model and derived an a priori error estimate for the discrete solution based on suitable projection operators. Following Dobson and Vogel [24], the total variation regularization can be approximated smoothly via $\sqrt{|\nabla u|^2 + \epsilon}$. In [27], the convergence of the L^2-gradient flow of this smooth approximation to the TV flow in L^2 is shown under strong regularity assumptions on the solution.

Furthermore, approximations of the original Mumford-Shah model have been studied extensively. An early overview of different approximation and discretization strategies was given by Chambolle in [13]. Ambrosio and Tortorelli [3] proposed a phase field approximation of this functional and proved its Γ-convergence. Chambolle and Dal Maso [18] proposed a discrete finite element approximation and established its Γ-convergence. Bourdin and Chambolle [9] picked up this approach and studied the generation of adaptive meshes iteratively adapted in accordance to an anisotropic metric depending on the current approximate solution. In [46], Shen introduced a Γ-converging approximation of the piecewise constant Mumford-Shah segmentation, where the length term in the Mumford-Shah model is approximated via an approach originating from the phase field model by Modica and Mortola [34]. A simple and widespread level set approach was proposed by Chan and Vese [22].

The goal of this paper is to derive a posteriori error estimates for the characteristic function χ. To this end, we proceed as follows: We take into account a suitable *uniformly* convex relaxation of the binary Mumford-Shah functional already studied in [8], which is related to more general relaxation approaches suggested by Chambolle [15] and does not require any constraint in the minimization. For this relaxation, we consider its predual and set up a corresponding primal-dual algorithm [5, 20, 30]. Then, following Bartels [5], we use Repin's primal-dual approach [42, 43] to derive functional a posteriori error estimates for the relaxed solution based on upper bounds of the duality gap (*cf.* also the book by Han [29] with respect to mechanical applications) (see Section 2). These estimates can be used together with a suitable cut out argument to derive an a posteriori estimate for the characteristic function χ minimizing the original functional (1.1). In addition, a sensitivity analysis of χ depending on c_1, c_2 and of c_1, c_2 depending on χ is studied (see Section 3). Moreover, two adaptive finite element discretization schemes and one conventional, non-adaptive finite difference scheme are investigated (see Section 4). Based on these discretization schemes, a primal-dual algorithm and a dual gradient descent are introduced in Section 5. Finally, we apply the resulting estimate to these schemes incorporating an appropriate post-smoothing and present the numerical results (see Section 6).

2. Uniformly convex relaxation and functional error estimates

Henceforth, we use the notation χ_A to denote the indicator function of a measurable set $A \subset \Omega$ and define $[u > c] := \{x \in \Omega : u(x) > c\}$. We use generic constants c and C throughout this paper. Furthermore, if not stated otherwise, we assume the intensity values c_1 and c_2 to be fixed with $c_1, c_2 \in [0, 1]$ and $c_1 \neq c_2$. Rewriting the binary Mumford-Shah functional (1.1) as

$$E[\chi] = \int_\Omega (\theta_1 - \theta_2)\chi \,\mathrm{d}x + |\mathrm{D}\chi|(\Omega) + \int_\Omega \theta_2 \,\mathrm{d}x\,, \tag{2.1}$$

one observes that adding a constant to θ_1 and θ_2 leaves the minimizers χ unchanged. Thus, we may assume that $\theta_1, \theta_2 \geq c > 0$. Let us introduce the relaxed functional

$$E^{\mathrm{rel}}[u] = \int_\Omega u^2\theta_1 + (1-u)^2\theta_2 \,\mathrm{d}x + |\mathrm{D}u|(\Omega)\,, \tag{2.2}$$

which is supposed to be minimized over all $u \in \mathrm{BV}(\Omega, \mathbb{R})$. Indeed, $E^{\mathrm{rel}}[\chi] = E[\chi]$ for characteristic functions χ, and one retrieves the original binary Mumford-Shah model. Proving existence of minimizers of (2.2) via the direct method in the calculus of variations is straightforward; for details we refer to [4, 26]. Furthermore, (2.2) is uniformly convex by our above assumptions on θ_1 and θ_2. Loosely speaking, a preference for the values 0 and 1 for u is encoded in the quadratically growing data term. The minimizers of both functionals (1.1) and (2.2) are related in the following sense (*cf.* [8]).

Proposition 2.1 (Convex relaxation and thresholding)**.** *Under the above assumptions, a minimizer $u \in \mathrm{BV}(\Omega)$ of the functional E^{rel} exists, $u(x) \in [0, 1]$ for a.e. $x \in \Omega$ and*

$$\chi_{[u>0.5]} \in \operatorname*{argmin}_{\chi \in \mathrm{BV}(\Omega,\{0,1\})} E[\chi]\,.$$

Proposition 2.1 is an instance of a more general result, which can be found in [15, 19, 32]. In fact, let $\Psi : \Omega \times \mathbb{R} \to \mathbb{R}$ be measurable, $\Psi(\cdot, t) \in L^1(\Omega)$ for a.e. $t \in \mathbb{R}$,

$\Psi(x,\cdot) \in C^1(\mathbb{R})$ be strictly convex for a.e. $x \in \Omega$, $\Psi(x,t) \geq c|t| - C$ and $E_\Psi[u] = \int_\Omega \Psi(x,u)\,\mathrm{d}x + |\mathrm{D}u|(\Omega)$. Then, there exists a minimizer $u \in \operatorname{argmin}_{\tilde{u}\in \mathrm{BV}(\Omega)} E_\Psi[\tilde{u}]$ and for $s \in \mathbb{R}$,

$$\chi^s := \chi_{[u>s]} \in \operatorname*{argmin}_{\chi\in \mathrm{BV}(\Omega,\{0,1\})} \int_\Omega \partial_t \Psi(x,s)\chi \,\mathrm{d}x + |\mathrm{D}\chi|(\Omega)\,.$$

For $\Psi(x,t) = t^2\theta_1(x) + (1-t)^2\theta_2(x)$ the property $u(x) \in [0,1]$ follows directly when comparing with the energy of the function $\min(1,\max(0,u))$. Furthermore, choosing $s = \frac{1}{2}$ allows us to verify the main claim of Proposition 2.1. Indeed, for $t \in \mathbb{R}$ and $\chi \in \mathrm{BV}(\Omega,\{0,1\})$, let $E_t^{\mathrm{rel}}[\chi] := \int_\Omega \partial_t\Psi(x,t)\chi\,\mathrm{d}x + |\mathrm{D}\chi|(\Omega)$. For our specific choice of Ψ,

$$E_t^{\mathrm{rel}}[\chi] = \int_\Omega \left(2t(\theta_1(x)+\theta_2(x)) - 2\theta_2(x)\right)\chi\,\mathrm{d}x + |\mathrm{D}\chi|(\Omega)$$

implies that minimizing the functional $E_{\frac{1}{2}}^{\mathrm{rel}}$ is equivalent to minimizing the functional E because $E_{\frac{1}{2}}^{\mathrm{rel}}[\chi] = E[\chi] - \int_\Omega \theta_2\,\mathrm{d}x$.

Remark. The particular advantage of our model compared to the relaxation approach by Nikolova, Esedoglu and Chan [21] is that the relaxed problem does not need to be constrained to functions u with values in $[0,1]$. One could also consider an ROF type functional choosing $\Psi(x,t) = \frac{1}{2}(t - (\theta_2(x) - \theta_1(x)))^2$ (*cf.* [7]) and obtain the functional $E_0^{\mathrm{rel}}[\chi]$ for the threshold value $s = 0$, but in this case the L^∞ bound of the relaxed solution depends on the L^∞ bounds of $\frac{1}{\nu}(c_i - u_0)^2$ (*cf.* Proposition 2.1) and requires a more involved cutoff scheme (see Section 3).

In what follows, we make use of convex analysis to derive a duality formulation for the minimization problem of the relaxed functional (2.2) and derive functional a posteriori estimates for this problem. Primal and dual formulations will later be used in the a posteriori estimates. The dual of $\mathrm{BV}(\Omega)$ is very difficult to characterize and not suitable for computational purposes. Thus, for a generalized ROF model Hintermüller and Kunisch [30] proposed considering the corresponding BV functional as the dual of another functional, which we refer to as the *predual functional*. Bartels [5] made use of this approach in the context of a posteriori estimates for the ROF model. Here, we follow this procedure and investigate the predual of (2.2).

Recall that the *Fenchel conjugate* J^* of a functional $J : X \to \bar{\mathbb{R}}$ on a Banach space X with $\bar{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$ is a functional on the dual space X' with values in $\bar{\mathbb{R}}$, defined as $J^*[x'] = \sup_{x\in X}\{\langle x', x\rangle - J[x]\}$, where $\langle\cdot,\cdot\rangle$ denotes the duality pairing. Furthermore, we denote by $\Lambda^* \in \mathcal{L}(Y', X')$ the *adjoint operator* of $\Lambda \in \mathcal{L}(X,Y)$ and by ∂J the *subgradient* of J (*cf.* [25]).

Now, we investigate an energy functional

$$D^{\mathrm{rel}}[q] = F[q] + G[\Lambda q], \qquad q \in \mathcal{Q}\,, \tag{2.3}$$

with $F : \mathcal{Q} \to \bar{\mathbb{R}}$ and $G : \mathcal{V} \to \bar{\mathbb{R}}$ being proper, convex and lower semicontinuous functionals, $\mathcal{V}$ and $\mathcal{Q}$ being reflexive Banach spaces and $\Lambda \in \mathcal{L}(\mathcal{Q},\mathcal{V})$. In our case, the predual of the convex relaxed binary Mumford-Shah model is given by

$$F[q] = I_{\overline{B}_1}[q] = \begin{cases} 0 & \text{if } |q| \leq 1 \text{ a.e.,} \\ +\infty & \text{else,} \end{cases} \qquad G[v] = \int_\Omega \frac{\frac{1}{4}v^2 + v\theta_2 - \theta_1\theta_2}{\theta_1+\theta_2}\,\mathrm{d}x\,,$$

with $\Lambda = \operatorname{div}$, $\mathcal{Q} = H_N(\operatorname{div},\Omega)$ and $\mathcal{V} = L^2(\Omega)$. Recall the definition of the spaces $H(\operatorname{div},\Omega) = \{q \in L^2(\Omega,\mathbb{R}^n) : \operatorname{div} q \in L^2(\Omega)\}$, endowed with the norm $\|q\|^2_{H(\operatorname{div},\Omega)} = \|q\|^2_{L^2(\Omega)} + \|\operatorname{div} q\|^2_{L^2(\Omega)}$, and $H_N(\operatorname{div},\Omega) = H(\operatorname{div},\Omega) \cap \{q \cdot \nu = 0 \text{ on } \partial\Omega\}$, where ν is the outer normal on $\partial\Omega$ and the operator div is understood in the weak sense. Moreover, $\Lambda^* = -\nabla$ holds in the sense

$$(\Lambda^* v, q)_{L^2(\Omega)} = (v, \operatorname{div} q)_{L^2(\Omega)} \quad \forall v \in \mathcal{V},\ q \in \mathcal{Q}\,.$$

Based on this duality and for the particular choice of D^{rel}, we easily verify that $(D^{\mathrm{rel}})^* = E^{\mathrm{rel}}$. Indeed, from the general theory in [25, pp. 58ff.], we can deduce $(D^{\mathrm{rel}})^*[v] = F^*[-\Lambda^* v] + G^*[v]$. As a result of the denseness of $C^1_c(\Omega)$ in $H_N(\operatorname{div},\Omega)$ with respect to the norm $\|\cdot\|_{H(\operatorname{div},\Omega)}$, we can infer for any $v \in \mathrm{BV}(\Omega)$

$$|\mathrm{D}v|(\Omega) = \sup_{q\in\mathcal{Q},\|q\|_\infty\le 1} \int_\Omega v \operatorname{div} q \,\mathrm{d}x = \sup_{q\in\mathcal{Q}} \left(- \int_\Omega v \operatorname{div} q \,\mathrm{d}x - I_{\overline{B}_1}[q] \right),$$

which leads to

$$F^*[-\Lambda^* v] = \sup_{q\in\mathcal{Q}} \left(- \int_\Omega v \operatorname{div} q \,\mathrm{d}x - I_{\overline{B}_1}[q] \right) = |\mathrm{D}v|(\Omega)\,.$$

On the other hand, the Fenchel conjugate of G can be computed as follows:

$$G^*[v] = \sup_{w\in L^2(\Omega)} \big((v,w)_{L^2(\Omega)} - G[w] \big) = \int_\Omega v^2\theta_1 + (1-v)^2\theta_2 \,\mathrm{d}x\,,$$

where the supremum is attained for $w = 2v(\theta_1+\theta_2) - 2\theta_2$. This verifies the assertion.

The central insight is that

$$D^{\mathrm{rel}}[p] = -(D^{\mathrm{rel}})^*[u] \tag{2.4}$$

for a minimizer p of D^{rel} and a minimizer u of $(D^{\mathrm{rel}})^*$. A rigorous verification can be found in [25, Chapter III.4] (see also [5, 42, 44]). Furthermore, one obtains that $\bar q \in \mathcal{Q}$ and $\bar v \in \mathcal{V}$ are optimal if and only if $-\Lambda^*\bar v \in \partial F[\bar q]$ and $\bar v \in \partial G[\Lambda \bar q]$, which can be deduced from the equivalence $J[x] + J^*[x'] = \langle x', x\rangle \iff x' \in \partial J[x]$ (see [25, Proposition I.5.1]).

In what follows, we investigate a posteriori error estimates associated with the energy $D^{\mathrm{rel}}[q] = F[q] + G[\Lambda q]$ and its dual $E^{\mathrm{rel}}[v] = F^*[-\Lambda^* v] + G^*[v]$ (for fixed intensity values c_1 and c_2). A crucial prerequisite is the uniform convexity of G^*, which is linked to the specific choice of the relaxed model E^{rel}.

Recall that a functional $J : X \to \mathbb{R}$ is *uniformly convex* if there exists a continuous functional $\Phi_J : X \to [0,\infty)$ such that $J\big[\frac{x_1+x_2}{2}\big] + \Phi_J(x_2-x_1) \le \frac{1}{2}(J[x_1]+J[x_2])$ for all $x_1, x_2 \in X$ and $\Phi_J(x) = 0$ if and only if $x = 0$. Furthermore, we denote by Ψ_J a non-negative functional such that $\langle x', x_2 - x_1\rangle + \Psi_J(x_2-x_1) \le J[x_2] - J[x_1]$ for all $x' \in \partial J[x_1]$. Hence, Ψ_J allows a quantification of the strict monotonicity of J. If $J \in C^2$ and λ_{min} denotes the smallest eigenvalue of D^2J, then Φ_J and Ψ_J admit the representation $\Phi_J(x) = \frac{1}{8}\lambda_{min}(D^2J)\|x\|^2$ and $\Psi_J(x) = \frac{1}{2}\lambda_{min}(D^2J)\|x\|^2$, which follows readily via a Taylor expansion.

Remark. The optimal Ψ_J^* coincides with the *Bregman distance* (*cf.* [11]), *i.e.*

$$\Psi_J(x_2-x_1) \le \Psi_J^*(x_2-x_1) := J[x_2] - J[x_1] - \sup_{x'\in\partial J[x_1]} \langle x', x_2 - x_1\rangle$$

for any other Ψ_J and $x_1, x_2 \in X$.

Now, the a posteriori error estimate is based on the following direct application of a general result by Repin [42]: Let $u \in \operatorname{argmin}_{\tilde{v}\in\mathcal{V}} E^{\mathrm{rel}}[\tilde{v}]$ and $q \in \mathcal{Q}$, $v \in \mathcal{V}' = \mathcal{V} = L^2(\Omega)$. Then,

$$\Phi_{G^*}(v-u) + \Phi_{F^*}(-\Lambda^*(v-u)) + \Psi_{E^{\mathrm{rel}}}\left(\frac{v-u}{2}\right) \leq \frac{1}{2}(E^{\mathrm{rel}}[v] + D^{\mathrm{rel}}[q])\,. \tag{2.5}$$

The proof of (2.5) relies on the above strict convexity estimates and the fundamental relation $E^{\mathrm{rel}}[u] \geq -D^{\mathrm{rel}}[q]$ known as the *weak complementarity principle* [25], and can be found in [5, 42].

In the case of the binary Mumford-Shah model, we easily verify that

$$\Phi_{F^*} \equiv 0,\ \Phi_{G^*}(v) = \frac{1}{4}\int_\Omega v^2(\theta_1+\theta_2)\,\mathrm{d}x,\ \Psi_{E^{\mathrm{rel}}}(v) = \int_\Omega v^2(\theta_1+\theta_2)\,\mathrm{d}x\,,$$

and the estimate (2.5) implies for any $v \in \mathcal{V}$ and $q \in \mathcal{Q}$

$$\int_\Omega (u-v)^2(\theta_1+\theta_2)\,\mathrm{d}x \leq E^{\mathrm{rel}}[v] + D^{\mathrm{rel}}[q]\,.$$

Finally, $\frac{1}{2}(a-b)^2 \leq a^2+b^2$ with $a = c_1 - u_0$ and $b = c_2 - u_0$ yields $\frac{1}{2\nu}(c_1-c_2)^2 \leq \theta_1+\theta_2$. Thus, we obtain the following theorem.

Theorem 2.2. *Let $u \in \mathcal{V}$ be the minimizer of E^{rel}. Then, for any $v \in \mathcal{V}$ and $q \in \mathcal{Q}$ it holds that*

$$\|u-v\|^2_{L^2(\Omega)} \leq \operatorname{err}^2_u[v,q,c_1,c_2] := \frac{2\nu}{(c_1-c_2)^2}\left(E^{\mathrm{rel}}[v] + D^{\mathrm{rel}}[q]\right)\,. \tag{2.6}$$

In the application, one asks for the (post-processed) discrete primal v and dual solution q which ensure a small right hand side. Additionally, the estimator err_u is *consistent*, *i.e.* $\operatorname{err}^2_u[v,q,c_1,c_2] \to 0$ provided v and q converge to the extrema of the corresponding energy functionals w.r.t. the topology of the associated Banach spaces.

3. A posteriori error estimates for the binary Mumford-Shah model

In the sequel, we expand the a posteriori theory to the binary Mumford-Shah model. The key observation is that for many images approximate solutions $u \in L^2(\Omega)$ of the relaxed model are characterized by steep profiles, where the actual solution of the original binary Mumford-Shah model jumps. We proceed as follows. We define

$$\mathrm{a}[v,\eta] = \left\| \chi_{[\frac{1}{2}-\eta \leq v \leq \frac{1}{2}+\eta]} \right\|_{L^1(\Omega)}$$

for $\eta \in \left(0, \frac{1}{2}\right)$, which measures the area of the preimage of the interval of size 2η centered at the threshold value $s = \frac{1}{2}$ (*cf.* Section 2). Based on the above observation, the set $\mathcal{S}_\eta = \left[\frac{1}{2} - \eta \leq v \leq \frac{1}{2} + \eta\right]$ can be regarded as the set of non-properly identified regions. Combining this definition with the thresholding argument presented in Proposition 2.1, we obtain the subsequent theorem:

Theorem 3.1 (A posteriori error estimator for χ)**.** *For fixed c_1 and c_2, let $u \in \mathrm{BV}(\Omega)$ and $\chi = \chi_{[u>\frac{1}{2}]} \in \mathrm{BV}(\Omega,\{0,1\})$ be a minimizer of the relaxed functional*

E^{rel} (see (2.2)) and the binary Mumford-Shah functional E (see (1.1)), respectively. Then for all $v \in \mathcal{V} = L^2(\Omega)$ and $q \in \mathcal{Q} = H_N(\mathrm{div}, \Omega)$ we have that

$$\left\| \chi - \chi_{[v>\frac{1}{2}]} \right\|_{L^1(\Omega)} \leq \mathrm{err}_\chi[v,q] := \inf_{\eta \in (0,\frac{1}{2})} \left(\mathrm{a}[v,\eta] + \frac{1}{\eta^2} \mathrm{err}_u^2[v,q,c_1,c_2] \right) . \tag{3.1}$$

Recall that $\chi_{[v>\frac{1}{2}]}$ is the indicator function of the set $[v > \frac{1}{2}]$. Let us remark that $\chi_{[v>\frac{1}{2}]}$ is the result of the same thresholding, which relates χ to the solution u of the relaxed problem (2.2), *i.e.* $\chi = \chi_{[u>\frac{1}{2}]}$, this time applied to v.

Proof. Any minimizer u of E^{rel} fulfills $0 \leq u \leq 1$ and $\chi_{[u>\frac{1}{2}]}$ minimizes E (see Proposition 2.1). For all $\eta \in (0, \frac{1}{2})$, we obtain the following set relation for the symmetric difference of the sets $[u > \frac{1}{2}]$ and $[v > \frac{1}{2}]$ (Δ denoting the symmetric difference of two sets):

$$\left[u > \tfrac{1}{2}\right] \Delta \left[v > \tfrac{1}{2}\right] \subseteq \left\{ x \in \Omega \,\middle|\, \tfrac{1}{2} - \eta \leq v(x) \leq \tfrac{1}{2} + \eta \right\} \cup \left\{ x \in \Omega \,\middle|\, |u(x) - v(x)| > \eta \right\} . \tag{3.2}$$

Now, using Theorem 2.2 the Lebesgue measure of the rightmost set can be estimated as

$$\mathcal{L}^n(|u-v| > \eta) \leq \int_{\{|u-v|>\eta\}} \frac{|u-v|^2}{\eta^2} \,\mathrm{d}x \leq \frac{1}{\eta^2} \mathrm{err}_u^2[v,q,c_1,c_2] \,,$$

where $\eta \in (0, \frac{1}{2})$. Finally, taking the infimum for all $\eta \in (0, \frac{1}{2})$ concludes the proof. □

In the application, the computational cost to find the optimal η is of the order of the degrees of freedom for the discrete solution and thus affordable. Let us emphasize that the error estimator err_χ is not tailored to a specific finite element approach. Indeed, we can project any primal and dual solution onto the spaces $\mathcal{V} = L^2(\Omega)$ and $\mathcal{Q} = H_N(\mathrm{div}, \Omega)$, respectively.

Remark. (i) We can obtain an a posteriori error estimate for the segmentation also for intensity values $\tilde{c}_i$, which are only known to be in intervals $[c_i - \epsilon, c_i + \epsilon]$ around some value c_i for $i = 1, 2$ and $\epsilon > 0$. Indeed, applying straightforward monotonicity arguments, we obtain the estimate $\sup_{\tilde{c}_i \in B_\epsilon(c_i),\, i=1,2} \mathrm{err}_u^2[v,q,\tilde{c}_1,\tilde{c}_2] \leq \mathrm{err}_u^{2,\epsilon}[v,q,c_1,c_2]$ for

$$\begin{aligned} \mathrm{err}_u^{2,\epsilon}[v,q,c_1,c_2] \quad := \quad & \int_\Omega \tfrac{2\nu}{(|c_1-c_2|-2\epsilon)^2} \Big(v^2 \theta_1^{max} + (1-v)^2 \theta_2^{max} + |\nabla v| + \tfrac{\frac{1}{4}(\mathrm{div}\, q)^2}{\theta_1^{min}+\theta_2^{min}} \\ & + \max\left\{ \tfrac{(\mathrm{div}\, q)\theta_2^{max}}{\theta_1^{min}+\theta_2^{min}}, \tfrac{(\mathrm{div}\, q)\theta_2^{min}}{\theta_1^{max}+\theta_2^{max}} \right\} - \tfrac{\theta_1^{min}\theta_2^{min}}{\theta_1^{max}+\theta_2^{max}} \Big) \,\mathrm{d}x \end{aligned}$$

provided that $2\epsilon < |c_1 - c_2|$. Here, $\theta_i^{max}(x) = \frac{1}{\nu} \max\{(c_i + \epsilon - u_0)^2, (c_i - \epsilon - u_0)^2\}$ and $\theta_i^{min}(x) = \frac{1}{\nu} \min\{(c_i + \epsilon - u_0)^2, (c_i - \epsilon - u_0)^2\}$ for $i = 1, 2$. Thus, for the minimizer $\tilde{\chi} \in \mathrm{BV}(\Omega, \{0,1\})$ of $E[\,\cdot\,, \tilde{c}_1, \tilde{c}_2]$ the a posteriori estimate

$$\left\| \tilde{\chi} - \chi_{[v>\frac{1}{2}]} \right\|_{L^1(\Omega)} \leq \mathrm{err}_\chi^\epsilon[v,q,c_1,c_2] := \inf_{\eta \in (0,\frac{1}{2})} \left(\mathrm{a}[v,\eta] + \frac{1}{\eta^2} \mathrm{err}_u^{2,\epsilon}[v,q,c_1,c_2] \right) \tag{3.3}$$

holds for all $v \in \mathcal{V} = L^2(\Omega)$ and $q \in \mathcal{Q} = H_N(\mathrm{div}, \Omega)$.

(ii) The optimal intensity values for a given characteristic function χ are given in (1.2). For the sensitivity of these values on χ, we straightforwardly obtain the

estimates

(3.4)
$$|c_1[\chi] - c_1[\tilde{\chi}]| \leq \frac{2\,\|\tilde{\chi}-\chi\|_{L^1(\Omega)}}{\|\chi\|_{L^1(\Omega)}-\|\tilde{\chi}-\chi\|_{L^1(\Omega)}}, \quad |c_2[\chi] - c_2[\tilde{\chi}]| \leq \frac{2\,\|(\tilde{\chi}-\chi)\|_{L^1(\Omega)}}{\|1-\chi\|_{L^1(\Omega)}-\|(\tilde{\chi}-\chi)\|_{L^1(\Omega)}}$$

assuming $\|(\tilde{\chi}-\chi)\|_{L^1(\Omega)} < \min\{\|\chi\|_{L^1(\Omega)}, \|1-\chi\|_{L^1(\Omega)}\}$.

(iii) Given the sensitivity results from (i) and (ii) one might ask for an a posteriori error estimate both for χ and the intensity values c_1, c_2. In fact, if (χ, c_1, c_2) is a minimizer of the in general non-convex energy

$$E = \int_\Omega \frac{1}{\nu}(c_1 - u_0)^2\chi + \frac{1}{\nu}(c_2 - u_0)^2(1-\chi)\,\mathrm{d}x + |\mathrm{D}\chi|(\Omega)$$

and one assumes a priori that each of the initially chosen intensity values is already in some ϵ neighborhood of the corresponding c_i–value, then the estimates (3.3) and (3.4) can be combined to obtain an a posteriori error estimate for the numerical approximation of (χ, c_1, c_2). Unfortunately, the estimate (3.3) is not sufficiently sharp to ensure that the resulting estimated error in the intensities does actually improve compared to the a priori assumption, and a bootstrapping argument could not be applied to further improve the resulting estimates. We refer to Section 6 for an explicit evaluation of the sensitivity of the relaxed solution.

4. Finite element and finite difference discretization

In this section, we investigate different numerical approximation schemes for the primal and the dual solution of the relaxed problem (2.2) on adaptive meshes and the refinement of the meshes based on the a posteriori error estimates derived in Section 3. In the context of image processing applications with input images usually given on a regular rectangular mesh, an adaptive quadtree for $n = 2$ (or octree for $n = 3$) turned out to be an effective choice for an adaptive mesh data structure. In what follows, we pick up the finite element approach for a variational problem on BV proposed by Bartels [6] and a simplified version of the latter. Furthermore, we consider the widespread finite difference scheme proposed by Chambolle [14]. In all numerical experiments in this paper, we choose $\Omega = [0,1]^2$.

(FE) Finite element scheme on an induced adaptive triangular grid. On the domain Ω, we consider an adaptive mesh $\mathcal{M}_h$ described by a quadtree with cells $\mathscr{C} \in \mathcal{M}_h$ being squares, which are recursively refined into four squares via an edge bisection. We suppose that the level of refinement between cells at edges differs at most by one. Thus, on a single edge at most one hanging node appears. Let h indicate the spatially varying mesh size function on Ω, where h on a grid cell $\mathscr{C}$ ranges from an initial mesh size $2^{-L_{\text{init}}}$ to a finest mesh size 2^{-L_0} (usually determined by the image resolution). For all discretization approaches investigated here, the degrees of freedom are associated with the non-hanging nodes. Let us denote by N_v the number of these nodes, which will coincide with the number of degrees of freedom of discrete primal functions. The finite element discretization is based on a triangular mesh $\mathcal{S}_h$ spread over the adaptive quadtree mesh via a splitting of each quadratic leaf cell into simplices $\mathscr{T}$ ("cross subdivision"). We ask for discrete primal functions u_h in the space of piecewise affine and globally continuous functions on $\mathcal{S}_h$ denoted by $\mathcal{V}_h$. Thus, for functions $v_h \in \mathcal{V}_h$ the values at hanging nodes are interpolated based on the values at adjacent non-hanging nodes, which are associated with the actual degrees of freedom. By $\mathcal{Q}_h = \{q_h \in \mathcal{V}_h^n : \; q_h \cdot \nu = 0 \text{ on } \partial\Omega\}$ we

denote the discrete counterpart of $\mathcal{Q}$. To accommodate this boundary condition, the boundary nodes are modified after each update of the dual solution in a post-processing step. On $\mathcal{V}_h$ and $\mathcal{Q}_h$, we define discrete counterparts of the continuous functionals F and G as follows:

$$G_h[v_h] := \int_\Omega \frac{\frac{1}{4}v_h^2 + v_h\theta_{2,h} - \theta_{1,h}\theta_{2,h}}{\theta_{1,h} + \theta_{2,h}}\,\mathrm{d}x\,, \quad F_h[q_h] := I_{\bar{B}_1}[q_h]\,,$$

where $\theta_{i,h} = \mathcal{I}_h(\theta_i) = \mathcal{I}_h(\frac{1}{\nu}(c_i - u_0)^2)$ for $i = 1, 2$ with $\mathcal{I}_h$ denoting the Lagrange interpolation. In the application on images, we suppose that $u_0 \in \mathcal{V}_0$, where $\mathcal{V}_0$ is the simplicial finite element space corresponding to the full resolution image on the finest grid level L_0 representing the full image resolution. Furthermore, we consider two different scalar products. On $\mathcal{V}_h$, we take into account the L^2-product and on $\mathcal{Q}_h$ the lumped mass product $(q_h, p_h) \mapsto \int_\Omega \mathcal{I}_h(q_h \cdot p_h)\,\mathrm{d}x$ and identify $\mathcal{V}_h$ and $\mathcal{Q}_h$ with their dual spaces with respect to the L^2- and the lumped mass product, respectively. Then, the associated dual operators are

$$G_h^*[v_h] = \int_\Omega v_h^2\theta_{1,h} + (1 - v_h)^2\theta_{2,h}\,\mathrm{d}x\,, \quad F_h^*[q_h] = \int_\Omega \mathcal{I}_h(|q_h|)\,\mathrm{d}x\,.$$

Finally, we define the discrete divergence $\Lambda_h : \mathcal{Q}_h \to \mathcal{V}_h$, $q_h \mapsto \mathcal{P}_h \operatorname{div} q_h$, where $\mathcal{P}_h$ denotes the L^2-projection $\mathcal{P}_h : L^2(\Omega) \to \mathcal{V}_h$. Following Bartels [5] and taking into account the above scalar products on $\mathcal{V}_h$ and on $\mathcal{Q}_h$, we obtain for the discrete gradient $-\Lambda_h^* : \mathcal{V}_h \to \mathcal{Q}_h$, $v_h \mapsto -\Lambda_h^* v_h$, the defining duality

$$\int_\Omega \mathcal{I}_h(\Lambda_h^* v_h \cdot q_h)\,\mathrm{d}x = \int_\Omega v_h \mathcal{P}_h \operatorname{div} q_h\,\mathrm{d}x \tag{4.1}$$

for all $q_h \in \mathcal{Q}_h$ and $v_h \in \mathcal{V}_h$.

(FE') Finite element scheme based on a simple gradient operator. Instead of the above defined discrete gradient operator $-\Lambda_h^*$, we alternatively consider the piecewise constant gradient ∇v_h on the simplices $\mathscr{T}$ of the simplicial mesh for functions $v_h \in \mathcal{V}_h$. To this end, we choose $\mathcal{Q}_h$ as the space of piecewise constant functions on the simplicial mesh and take into account the standard L^2-product on both spaces. The above definitions of the functionals G_h and F_h are still valid. Moreover, G_h^* remains the same, only F_h^* changes to $F_h^*[q_h] = \int_\Omega |q_h|\,\mathrm{d}x$. The discrete divergence $\Lambda_h : \mathcal{Q}_h \to \mathcal{V}_h$ is defined via duality starting from the preset discrete gradient as

$$\int_\Omega \Lambda_h q_h v_h\,\mathrm{d}x = -\int_\Omega q_h \cdot \nabla v_h\,\mathrm{d}x\,,$$

which indeed ensures that $-\Lambda_h^* v_h = \nabla v_h$. This simplified ansatz leads to a non-conforming iterative solution scheme (see Section 5), since the space of piecewise constant finite elements is not contained in $H_N(\operatorname{div}, \Omega)$ (*cf.* [6]). After each modification of the (piecewise constant) dual solution the values on the corresponding boundary cells are set to 0 to satisfy the boundary condition. To apply the derived a posteriori error estimates a projection onto the space $H_N(\operatorname{div}, \Omega)$ is required. To this end, we replace the solution $p_h \in \mathcal{Q}_h$ by its L^2-projection onto the space $\mathcal{V}_h^n$ after each execution of the algorithm.

(FD) Finite difference scheme on a regular mesh. The finite difference scheme for the numerical solution of functionals on BV proposed by Chambolle [14] is extensively used in many computer vision applications and applies to image data defined on a structured non-adaptive mesh. We compare the a posteriori error estimator for this scheme on non-adaptive meshes with the above finite element schemes on adaptive meshes. To this end, we denote by $\mathbf{V}_h \in \mathbb{R}^{N_v}$ and $\mathbf{Q}_h \in \mathbb{R}^{nN_v}$ nodal vectors on the regular lattice for primal and dual solutions, respectively. Here, $N_v = (h^{-1}+1)^n$, where h denotes the fixed grid size of the finite difference lattice. Integration is replaced by summation, and we obtain the following discrete analogues $\mathbf{G}_h$ and $\mathbf{F}_h$ of the continuous functionals F and G as functions on $\mathbb{R}^{N_v}$ and $\mathbb{R}^{nN_v}$, respectively:

$$\mathbf{G}_h[\mathbf{V}_h] := \sum_{i=1}^{N_v} \left(\frac{\frac{1}{4}(\mathbf{V}_h^i)^2 + \mathbf{V}_h^i \Theta_{2,h}^i - \Theta_{1,h}^i \Theta_{2,h}^i}{\Theta_{1,h}^i + \Theta_{2,h}^i} \right), \quad \mathbf{F}_h[\mathbf{Q}_h] := \max_{i=1,\dots,N_v} \mathbf{I}_{\bar{B}_1}[\mathbf{Q}_h^i]$$

with $\Theta_{1,h}^i$, $\Theta_{2,h}^i$ denoting the pointwise evaluation of θ_1 and θ_2, respectively, and $\mathbf{I}_{\bar{B}_1}[\mathbf{Q}_h^i] = 0$ for $|\mathbf{Q}_h^i| \leq 1$ and $+\infty$ otherwise. The associated dual operators for the standard Euclidean product as the duality pairing are

$$\mathbf{G}_h^*[\mathbf{V}_h] = \sum_{i=1}^{N_v} (\mathbf{V}_h^i)^2 \Theta_{1,h}^i + (1-\mathbf{V}_h^i)^2 \Theta_{2,h}^i, \quad \mathbf{F}_h^*[\mathbf{Q}_h] = \sum_{i=1}^{N_v} \left|\mathbf{Q}_h^i\right|.$$

Finally, we take into account periodic boundary conditions (by identifying degrees of freedom on opposite boundary segments) and use forward difference quotients to define the discrete gradient operator $-\mathbf{\Lambda}_h^* : \mathbb{R}^{N_v} \to \mathbb{R}^{nN_v}$, *i.e.*

$$((-\mathbf{\Lambda}_h^*)\mathbf{V}_h)^i = \left(\frac{\mathbf{V}_h^{\mathscr{N}(i,j)} - \mathbf{V}_h^i}{h} \right)_{j=1,\dots,n},$$

where $\mathscr{N}(i,j)$ is the index of the neighboring node in the direction of the jth coordinate vector. As a consequence, the matrix representing the discrete divergence operator $\mathbf{\Lambda}_h : \mathbb{R}^{nN_v} \to \mathbb{R}^{N_v}$ is just the negative transpose of the matrix representing the discrete gradient and thus corresponds to a discrete divergence based on backward difference quotients.

To use the a posteriori error estimate in the finite difference context, we consider as a simplest choice the piecewise bilinear functions u_h and p_h uniquely defined by the solution vectors $\mathbf{U}_h$ and $\mathbf{P}_h$, respectively. The boundary condition for p_h is taken care of in exactly the same way as in the case (FE).

5. Duality-based algorithms

For the numerical solution of the different discrete variational problems, we primarily use the primal-dual algorithm proposed by Chambolle and Pock [20, Algorithm 1], which computes both a discrete primal and a discrete dual solution to be used in the a posteriori error estimates. Note that we use [20, Algorithm 1] instead of [20, Algorithm 2] even though $\mathbf{G}_h^*$ is uniformly convex. As we will see later, evaluating $(\mathrm{Id} + \tau \partial \mathbf{G}_h^*)^{-1}$ requires the inversion of a matrix depending on τ. In Algorithm 1, τ is fixed and the inverse can be computed once using a Cholesky decomposition for the sparse, symmetric and positive-definite matrix (for details see [23]), whereas in Algorithm 2 the decomposition of the linear system has to be performed in each iteration. Moreover, using diagonal preconditioning one can

improve the convergence speed of Algorithm 1 without any further step size control (see [39]). There is a variety of alternative algorithms to solve this convex minimization problem, *e.g.* the split Bregman method [28], the semi-implicit dual gradient descent [14], the alternating descent method for the Lagrangian [5, Algorithm A'] or the alternating direction method of multipliers (ADMM) (see [10] and the references therein). For distinct images we compare below the aforementioned algorithm by Chambolle and Pock with a dual gradient descent in terms of the quantity of the error estimator.

Before we discuss the algorithm due to Chambolle and Pock in the more conventional matrix-vector notation, let us rewrite the finite element approaches correspondingly. Let $N_v = \dim \mathcal{V}_h$ (the number of non-hanging nodes) and $N_q = \dim \mathcal{Q}_h$ (for (FE) $N_q = nN_v$ and for (FE') N_q is n times the number of simplices). In what follows, we will use uppercase letters to denote a vector of nodal values, *e.g.* $\mathbf{V}_h^i = v_h(X^i)$ if X^i is the ith non-hanging node. The two scalar products are encoded via mass matrices. Here, $\mathbf{M}_h \in \mathbb{R}^{N_v,N_v}$ represents the standard L^2-product on $\mathcal{V}_h$ and is given by $\mathbf{M}_h\mathbf{V}_h \cdot \mathbf{U}_h = \int_\Omega v_h u_h \,\mathrm{d}x$ for all v_h, $u_h \in \mathcal{V}_h$. Furthermore, $\tilde{\mathbf{M}}_h \in \mathbb{R}^{N_q,N_q}$ is the mass matrix associated with the space $\mathcal{Q}_h$.

For the approach (FE) this is given as the lumped mass matrix with $\tilde{\mathbf{M}}_h\mathbf{P}_h \cdot \mathbf{Q}_h = \int_\Omega \mathcal{I}_h(p_h \cdot q_h)\,\mathrm{d}x$ for all p_h, $q_h \in \mathcal{Q}_h$, whereas for the discretization (FE') $\tilde{\mathbf{M}}_h\mathbf{P}_h \cdot \mathbf{Q}_h = \int_\Omega p_h \cdot q_h \,\mathrm{d}x$ for all p_h, $q_h \in \mathcal{Q}_h$ defines a classical (diagonal) mass matrix. For the matrix representations $\mathbf{\Lambda}_h$ and $-\mathbf{\Lambda}_h^*$ of the discrete divergence and the discrete gradient, respectively, we obtain the relation (*cf.* [5])

$$\mathbf{\Lambda}_h^* = \tilde{\mathbf{M}}_h^{-1}\mathbf{\Lambda}_h^T\mathbf{M}_h\,. \tag{5.1}$$

Moreover, for the discretization (FD) we have $\mathbf{\Lambda}_h^* = \mathbf{\Lambda}_h^T$. Altogether, the discrete predual energy $\mathbf{D}_h^{\mathrm{rel}} : \mathbb{R}^{N_q} \to \mathbb{R}$ and the discrete energy $\mathbf{E}_h^{\mathrm{rel}} : \mathbb{R}^{N_v} \to \mathbb{R}$ are defined as follows:

$$\mathbf{D}_h^{\mathrm{rel}}[\mathbf{Q}_h] = \mathbf{F}_h[\mathbf{Q}_h] + \mathbf{G}_h[\mathbf{\Lambda}_h\mathbf{Q}_h]\,,$$
$$\mathbf{E}_h^{\mathrm{rel}}[\mathbf{V}_h] = \mathbf{F}_h^*[-\mathbf{\Lambda}_h^*\mathbf{V}_h] + \mathbf{G}_h^*[\mathbf{V}_h]\,.$$

In the case of all finite element discretizations, the functionals $\mathbf{G}_h, \mathbf{F}_h, \mathbf{G}_h^*$ and $\mathbf{F}_h^*$ are defined using the corresponding functions on the finite element spaces, *e.g.* $\mathbf{G}_h^*[\mathbf{V}_h] := G_h^*[v_h]$.

Now, we are in position to formulate the primal-dual algorithm. For a fixed mesh, fixed intensity values c_1, c_2 and initial data $(\mathbf{U}_h^0, \mathbf{P}_h^0) \in \mathbb{R}^{N_v} \times \mathbb{R}^{N_q}$, the Algorithm 1 proposed by Chambolle and Pock [20, Algorithm 1] computes a sequence $(\mathbf{U}_h^k, \mathbf{P}_h^k)$, which converges to the tuple $(\mathbf{U}_h, \mathbf{P}_h)$ of the discrete primal and dual solutions provided $\tau\sigma\|\mathbf{\Lambda}_h\|^2 < 1$. Indeed, by using inverse estimates for finite elements (see

$k = 0$;
while $\|\mathbf{U}_h^{k+1} - \mathbf{U}_h^k\|_\infty > \mathit{THRESHOLD}$ **do**
  $\mathbf{P}_h^{k+1} = (\mathrm{Id} + \sigma\partial\mathbf{F}_h)^{-1}[\mathbf{P}_h^k - \sigma\mathbf{\Lambda}_h^*\bar{\mathbf{U}}_h^k]$;
  $\mathbf{U}_h^{k+1} = (\mathrm{Id} + \tau\partial\mathbf{G}_h^*)^{-1}[\mathbf{U}_h^k + \tau\mathbf{\Lambda}_h\mathbf{P}_h^{k+1}]$;
  $\bar{\mathbf{U}}_h^{k+1} = 2\mathbf{U}_h^{k+1} - \mathbf{U}_h^k$;
  $k = k + 1$;
end

Algorithm 1: The primal-dual algorithm used to minimize $\mathbf{E}_h^{\mathrm{rel}}$.

[37] for a computation of the constants) the operator norm can be bounded in the case (FE') for $n = 2$ as follows: $\|\mathbf{\Lambda}_h\|^2 \leq 48(3+2\sqrt{2})h_{min}^{-2} \approx 279.8\, h_{min}^{-2}$, where h_{min} denotes the minimal mesh size occurring in $\mathcal{M}_h$. Moreover, to estimate the operator norm for the case (FE) we use (4.1) and obtain

$$\begin{aligned}\|\mathbf{\Lambda}_h\|^2 &= \left(\max_{v_h\in\mathcal{V}_h,\|v_h\|_{L^2}=1}\ \max_{q_h\in\mathcal{Q}_h,\|q_h\|_h=1}\int_\Omega \mathcal{I}_h(-\Lambda_h^* v_h\cdot q_h)\,\mathrm{d}x\right)^2\\ &= \left(\max_{v_h\in\mathcal{V}_h,\|v_h\|_{L^2}=1}\ \max_{q_h\in\mathcal{Q}_h,\|q_h\|_h=1}\int_\Omega v_h\mathcal{P}_h\operatorname{div} q_h\,\mathrm{d}x\right)^2\\ &\leq \max_{\substack{q_h\in\mathcal{Q}_h,\\ \|q_h\|_h=1}}\|\mathcal{P}_h\operatorname{div} q_h\|_{L^2(\Omega)}^2 \leq \max_{\substack{q_h\in\mathcal{Q}_h,\\ \|q_h\|_h=1}}\|\operatorname{div} q_h\|_{L^2(\Omega)}^2 \leq 96(3+2\sqrt{2})h_{min}^{-2}\,,\end{aligned}$$

where $\|q_h\|_h^2 = \int_\Omega \mathcal{I}_h(|q_h|^2)\,\mathrm{d}x$. Finally, following [14] we can estimate $\|\mathbf{\Lambda}_h\|^2 \leq 8h_{min}^{-2}$ for the discretization (FD).

Suitable stopping criteria are a threshold on the maximum norm of the difference of successive solutions $\mathbf{U}_h^{k+1} - \mathbf{U}_h^k$ (which we apply here) or on the primal-dual gap $\mathbf{E}_h^{\mathrm{rel}}[\mathbf{U}_h^k] + \mathbf{D}_h^{\mathrm{rel}}[\mathbf{P}_h^k]$. To compute the resolvents $(\mathrm{Id}+\partial\mathbf{F}_h)^{-1}[\mathbf{Q}_h]$ and $(\mathrm{Id}+\partial\mathbf{G}_h^*)^{-1}[\mathbf{V}_h]$ we use a variational ansatz (for details see [44]); *i.e.* for the resolvent of a subdifferentiable functional J with an underlying scalar product $(\cdot,\cdot)$ we have that

$$(\mathrm{Id}+\tau\partial J)^{-1}[x] = \operatorname*{argmin}_y (x-y,x-y) + 2\tau J(y)\,.$$

The resolvent of $\mathbf{F}_h$ for the approaches (FE) and (FD) is given by

$$(\mathrm{Id}+\sigma\partial\mathbf{F}_h)^{-1}[\mathbf{Q}_h] = \left(\frac{\mathbf{Q}_h^i}{\max\{|\mathbf{Q}_h^i|,1\}}\right)_{i=1,\dots N_v} \tag{5.2}$$

with $\mathbf{Q}_h^i = q_h(X^i)$. In the case (FE'), the above evaluation is performed on each cell. For the discretizations (FE) and (FE'), we denote by $\mathbf{M}_h[\mathbf{W}_h]\mathbf{U}_h\cdot\mathbf{V}_h = \int_\Omega w_h\,u_h v_h\,\mathrm{d}x$ the weighted mass matrix for functions $u_h, v_h \in \mathcal{V}_h$ and a weight $w_h \in \mathcal{V}_h$. Then, the resolvent of $\mathbf{G}_h$ reads as

$$(\mathrm{Id}+\tau\partial\mathbf{G}_h^*)^{-1}[\mathbf{V}_h] = \left(\mathbf{M}_h[1+2\tau(\Theta_{1,h}+\Theta_{2,h})]\right)^{-1}\mathbf{M}_h\left(\mathbf{V}_h+2\tau\Theta_{2,h}\right)\,.$$

In the case (FD), the resolvent is given by

$$(\mathrm{Id}+\tau\partial\mathbf{G}_h^*)^{-1}[\mathbf{V}_h^i] = \frac{\mathbf{V}_h^i+2\tau\Theta_{2,h}^i}{1+2\tau\left(\Theta_{1,h}^i+\Theta_{2,h}^i\right)} \qquad \text{for } 1\leq i\leq N_v\,.$$

In our numerical experiments, we have chosen THRESHOLD $= 10^{-8}$.

In the sequel, we consider a projected dual gradient descent for the minimization of E^{rel} using the discretization (FE) and picking up Chambolle's semi-implicit gradient descent w.r.t. the dual variable for the ROF model using a finite difference scheme [14]. We remark that the algorithm can analogously be derived for the discretization scheme (FE'). Starting from the first order condition $\kappa := -(2u\theta_1 + 2(u-1)\theta_2) \in \partial\mathrm{TV}[u]$ for E^{rel} with $\mathrm{TV}[u] = |\mathrm{D}u|(\Omega)$, we can infer $u \in \partial\mathrm{TV}^*[\kappa]$ due to [25, Chapter I, Corollary 5.2], which is the first order condition of the functional

$$\int_\Omega \frac{1}{4(\theta_1+\theta_2)}\left(\kappa-2\theta_2\right)^2\,\mathrm{d}x + \mathrm{TV}^*[\kappa]\,.$$

Thus, the unique minimizer is given by $\kappa = \mathcal{P}_S[2\theta_2]$, where $\mathcal{P}_S$ denotes the orthogonal projection onto the set $S = \left\{ \operatorname{div} p \,\middle|\, p \in H_N(\operatorname{div}, \Omega), \|p\|_{L^\infty(\Omega)} \leq 1 \right\}$ w.r.t. the weighted L^2-space with weight $w = (4(\theta_1 + \theta_2))^{-1}$, denoted by $L^2(\Omega, w)$. Note that $w \in L^\infty(\Omega, \mathbb{R}^{>0})$ due to the assumptions regarding θ_i. The primal solution is given by $u = \frac{2\theta_2 - \mathcal{P}_S[2\theta_2]}{2(\theta_1+\theta_2)}$. For the computation of $\mathcal{P}_S[2\theta_2]$ we have to solve $\operatorname{argmin}_{\iota \in S} \|\iota - 2\theta_2\|^2_{L^2(\Omega,w)}$, which is computed by alternatingly performing a gradient descent for the unconstrained problem and a projection onto S. To be precise, for a fixed step size $\tau > 0$ the usual gradient descent update scheme reads $p^{k+1} = p^k - \tau \nabla(w(\operatorname{div} p^k - 2\theta_2))$. After a multiplication by a test function $\vartheta \in \mathcal{Q}_h$, applying $\int_\Omega \mathcal{I}_h(\,\cdot\,)\,\mathrm{d}x$ on both sides, replacing ∇ and div by the differential operators $-\Lambda_h^*$ and Λ_h as introduced in the discretization (FE), respectively, and using (4.1) one obtains

$$\int_\Omega \mathcal{I}_h(p^{k+1} \cdot \vartheta)\,\mathrm{d}x = \int_\Omega \mathcal{I}_h(p^k \cdot \vartheta)\,\mathrm{d}x + \tau \int_\Omega (w(\Lambda_h p^k - 2\theta_2))\mathcal{P}_h \operatorname{div} \vartheta \,\mathrm{d}x\,. \tag{5.3}$$

Let $\tilde{\mathbf{M}}_h$ be the lumped mass matrix, $\mathbf{L}_j\mathbf{U}_h \cdot \mathbf{V}_h = \int_\Omega u_h \cdot \partial_j v_h \,\mathrm{d}x$, $j = 1, 2$, for all u_h, $v_h \in \mathcal{V}_h$, and $\mathbf{W}_h^i = w(X^i)$. Then (5.3) implies

$$[\mathbf{P}_h^{k+1}]_j = [\mathbf{P}_h^k]_j + \tau \tilde{\mathbf{M}}_h^{-1} \mathbf{L}_j(\mathbf{W}_h(\mathbf{\Lambda}_h \mathbf{P}_h^k - 2\Theta_{2,h})) \quad \text{for } j = 1, 2\,. \tag{5.4}$$

After each update of the dual variable (5.4) a projection onto S is performed (*cf.* (5.2)). The step size is taken as $\tau = \gamma \cdot \min_{\mathscr{C} \in \mathcal{M}_h} h(\mathscr{C})$. The stopping criterion for the resulting algorithm relies on the L^∞-distance of two successive dual solutions (in our case, THRESHOLD $= 10^{-8}$). For a convergence analysis of this approach we refer to [12]. In the sequel, we will denote results obtained with the discretization scheme (FE) and this algorithm by (FE_D).

The adaptive mesh refinement is implemented as follows. Given a mesh, fixed intensity values and initial data for the primal and dual solutions, we run one of the above algorithms and compute the relaxed discrete primal-dual solution pair (u_h, p_h). In the case of the finite difference approach (FD), we define them as the multilinear interpolation on the cells $\mathscr{C}$ of the regular mesh. The corresponding discrete solution of the original problem (2.1) is then given as $\chi_h = \chi_{[u_h > \frac{1}{2}]}$. Based on u_h and p_h, we evaluate the local error estimator for every cell $\mathscr{C}_0$ of the full resolution image grid as follows:

$$\begin{aligned} \operatorname{err}^2_{u,\mathscr{C}_0}[u_h, p_h] \;:=\; & \frac{2\nu}{(c_1 - c_2)^2} \Bigg(\int_{\mathscr{C}_0} u_h^2\theta_1 + (1 - u_h)^2\theta_2 + |\nabla u_h| \\ & + \frac{\frac{1}{4}(\operatorname{div} p_h)^2 + (\operatorname{div} p_h)\theta_2 - \theta_1\theta_2}{\theta_1 + \theta_2}\,\mathrm{d}x \Bigg)\,. \end{aligned}$$

To this end, a higher order Gaussian quadrature is used. In fact, for (FE), (FE_D) and (FE') we use a Gaussian quadrature of order 4 on the simplices $\mathscr{T}_0$ composing the cell $\mathscr{C}_0$ on the finest mesh with full image resolution, where the θ_i ($i = 1, 2$) are originally defined, and for (FD) we use a Gaussian quadrature of order 5 directly on the cells $\mathscr{C}_0$. The resulting local error estimator for a cell $\mathscr{C} \in \mathcal{M}_h$ and the global estimator are given as

$$\operatorname{err}^2_{u,\mathscr{C}}[u_h, p_h] = \sum_{\mathscr{C}_0 \subset \mathscr{C}} \operatorname{err}^2_{u,\mathscr{C}_0}[u_h, p_h] \quad \text{and} \quad \operatorname{err}^2_u[u_h, p_h] = \sum_{\mathscr{C} \in \mathcal{M}_h} \operatorname{err}^2_{u,\mathscr{C}}[u_h, p_h]\,,$$

respectively. We mark those cells $\mathscr{C}$ for refinement for which

$$\operatorname{err}^2_{u,\mathscr{C}}[u_h, p_h] \geq \alpha \max_{\mathscr{C}' \in \mathcal{M}_h} \operatorname{err}^2_{u,\mathscr{C}'}[u_h, p_h], \tag{5.5}$$

where α is a fixed threshold in $(0, 1)$. Since this method is prone to outliers, we additionally sort all local estimators $\operatorname{err}^2_{u,\mathscr{C}}$ according to their size (starting with the smallest) and mark the cells in the upper decile for refinement as well. For the input data from Figure 1 we refine up to the resolution of the initial image.

6. Numerical results

In what follows, we show numerical results for four different input images shown in Figure 1. Prior to executing the primal-dual scheme (Algorithm 1) or the dual gradient descent scheme, we choose suitable values for c_1 and c_2 by applying Lloyd's Algorithm (see [33]) for the computation of a 2-means clustering (with initial values 1 and 0). The resulting intensity values are given in Figure 1 together with the values for ν.

image	(a)	(b)	(c)	(d)
resolution	2049×2049	2049×2049	2049×2049	513×513
c_1	0.999772	0.893734	0.664404	0.602566
c_2	$1.99 \cdot 10^{-4}$	0.030416	0.167763	0.092273
ν	$5 \cdot 10^{-3}$	$5 \cdot 10^{-3}$	10^{-3}	$5 \cdot 10^{-3}$

Figure 1. Input images together with the corresponding image resolution and the model parameters c_1, c_2, and ν. (Flower image: "Leucanthemum vulgare" by Derek Ramsey/Chanticleer Garden (desaturated from original), used under CC BY-SA 3.0, https://commons.wikimedia.org/wiki/File:Leucanthemum_vulgare_%27Filigran%27_Flower_2200px.jpg. Cameraman image: copyright by Massachusetts Institute of Technology, used with permission.)

The pixels of the input images are interpreted as nodal values of the function u_0 on a uniform mesh with mesh size $h = 2^{-L_0}$ ($L_0 = 9$ for (d), $L_0 = 11$ else). The algorithm is then started on a uniform mesh of mesh size $h = 2^{-L_{\text{init}}}$ ($L_{\text{init}} = 3$ for (d), $L_{\text{init}} = 5$ else). In all computations we use $\tau = 10^{-5}$ and $\sigma = 5 \cdot 10^{-5}$ (Algorithm 1), $\gamma = 0.05$ for the dual gradient descent and $\alpha = 0.2$. We perform 10 cycles of the adaptive algorithm and refine cells until the depth L_0 of the input image is reached.

We observe slight local oscillations for the finite element approaches (FE), (FE_D) and (FE'), which deteriorate the result of the a posteriori estimator (*cf.* the numerical results in [5]). Thus, in a post-processing step, we compensate these oscillations prior to the evaluation of the estimator by an application of a smoothing filter. The filter is defined via an implicit time step of the discrete heat equation using affine finite elements on the underlying adaptive mesh; *i.e.* we apply the operator

$(\mathbf{M}_h + \iota \mathbf{S}_h)^{-1}\mathbf{M}_h$ to the solutions, where $\mathbf{S}_h$ denotes the stiffness matrix. For the discretizations (FE) and (FE$_D$), we choose $\iota = c \cdot h_{min}^2$, where h_{min} denotes the minimal mesh size of the current adaptive grid, with $c = 3$ and $c = 6$ for the primal and the dual solution, respectively. Moreover, in the case (FE') the smoothing is only applied to the dual solution with parameter $\iota = 0.75 \cdot h_a^{0.9}$, where h_a denotes the average cell size on the adaptive mesh. In our experiments, we observed that these smoothing methods and parameters outperformed other tested choices for the corresponding discretizations. We call the resulting post-processed functions $\bar{u}_h$ and $\bar{p}_h$, respectively, and replace the local error estimator by $\mathrm{err}^2_{u,\mathscr{C}}[\bar{u}_h, \bar{p}_h]$.

TABLE 1. Rescaled dual and primal energy evaluated on the discrete solution (u_h, p_h), error estimator for the relaxed solution, optimal threshold $\eta_{optimal}$ computed for u_h and the resulting a posteriori estimator err_χ for the L^1-error of the characteristic function χ (after 10 cycles of the adaptive algorithm).

		(a)	(b)	(c)	(d)
$\frac{2\nu}{(c_1-c_2)^2}E[u_h]$	(FE)	0.022422	0.078461	0.124137	0.203797
	(FE$_D$)	0.022421	0.078632	0.124659	0.204353
	(FE')	0.022249	0.077981	0.122572	0.202494
	(FD)	0.022493	0.078814	0.122777	0.205645
$\frac{2\nu}{(c_1-c_2)^2}D[p_h]$	(FE)	-0.021736	-0.075971	-0.117785	-0.183819
	(FE$_D$)	-0.021757	-0.075870	-0.117973	-0.181720
	(FE')	-0.020973	-0.071333	-0.111234	-0.166114
	(FD)	-0.021520	-0.075455	-0.119865	-0.188600
err^2_u	(FE)	6.86e-04	0.002490	0.006352	0.019978
	(FE$_D$)	6.63e-04	0.002761	0.006686	0.022633
	(FE')	0.001276	0.006647	0.011338	0.036380
	(FD)	9.73e-04	0.003359	0.002912	0.017045
$\eta_{optimal}$	(FE)	0.39	0.3225	0.2825	0.3075
	(FE$_D$)	0.385	0.325	0.2725	0.315
	(FE')	0.45	0.3675	0.31	0.3275
	(FD)	0.4375	0.345	0.24	0.3025
err_χ	(FE)	0.008847	0.038502	0.160912	0.373173
	(FE$_D$)	0.008704	0.041019	0.167892	0.394317
	(FE')	0.009816	0.068514	0.227060	0.545201
	(FD)	0.008223	0.0425225	0.109039	0.339256

Table 1 lists (scaled) primal and dual energies, err^2_u, $\eta_{optimal}$ (the η value corresponding to the optimal a posteriori error bound for given err^2_u), as well as err_χ for all input images after the 10th refinement step of the adaptive algorithm. The value of err_χ peaks for the application (d) due to the relatively low image resolution. Figure 2 plots the error estimator err^2_u after each refinement step for all input images and all finite element discretizations. In most of our numerical experiments, the scheme (FE) performs comparably to the discretization (FE$_D$), but slightly better than the scheme (FE'). For the flower image, the sequence of adaptive meshes and solutions resulting from the adaptive algorithm for the discretization (FE') is depicted in Figure 3.

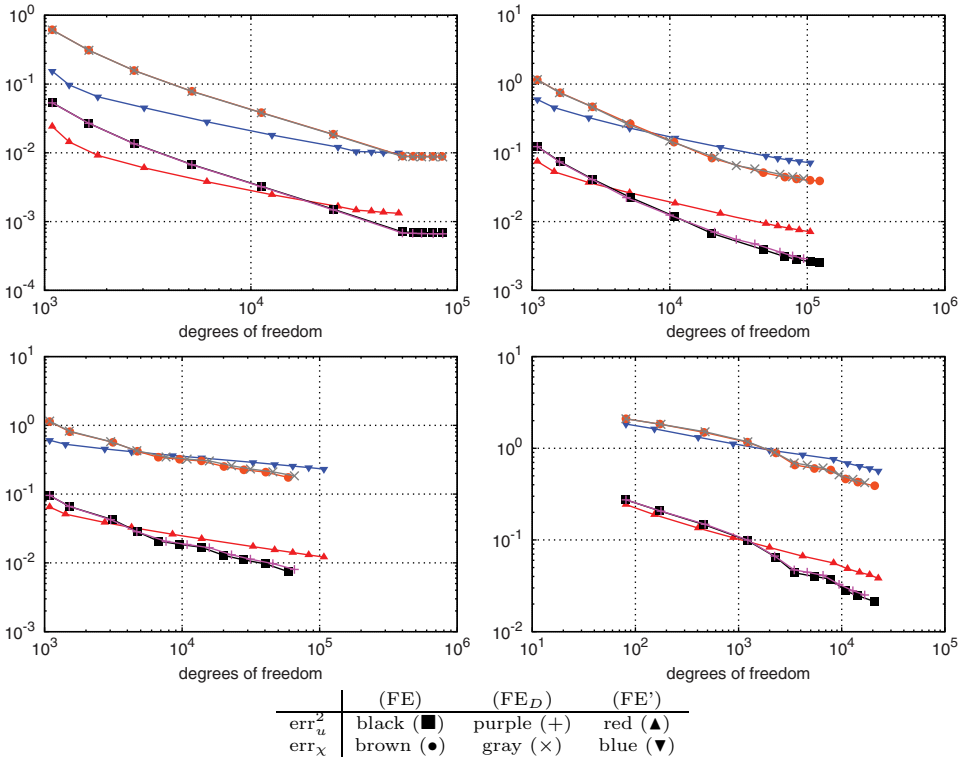

FIGURE 2. The values of err_u^2 and err_χ are displayed in relation to the number of degrees of freedom in a log-log plot for the applications (a) (upper left), (b) (upper right), (c) (lower left) and (d) (lower right).

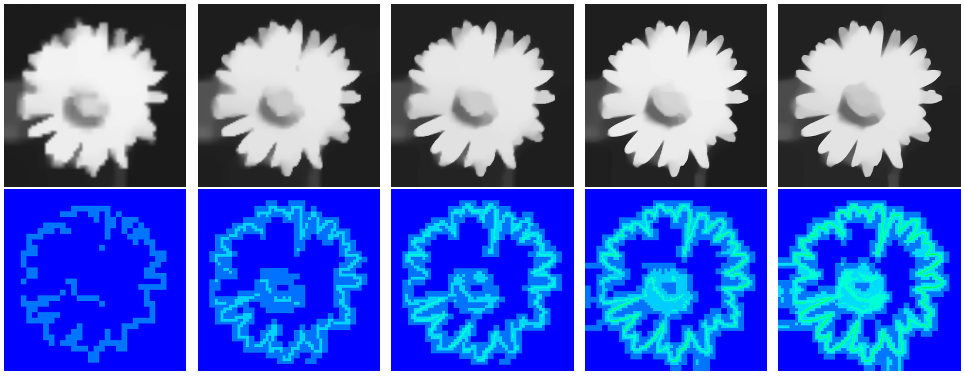

FIGURE 3. The sequence of solutions u_h and a color coding of the corresponding fineness of the adaptive meshes at the 1st, 2nd, 3rd, 4th and 5th iterations of the adaptive algorithm applied to the input image (c) and computed using the (FE') discretization.

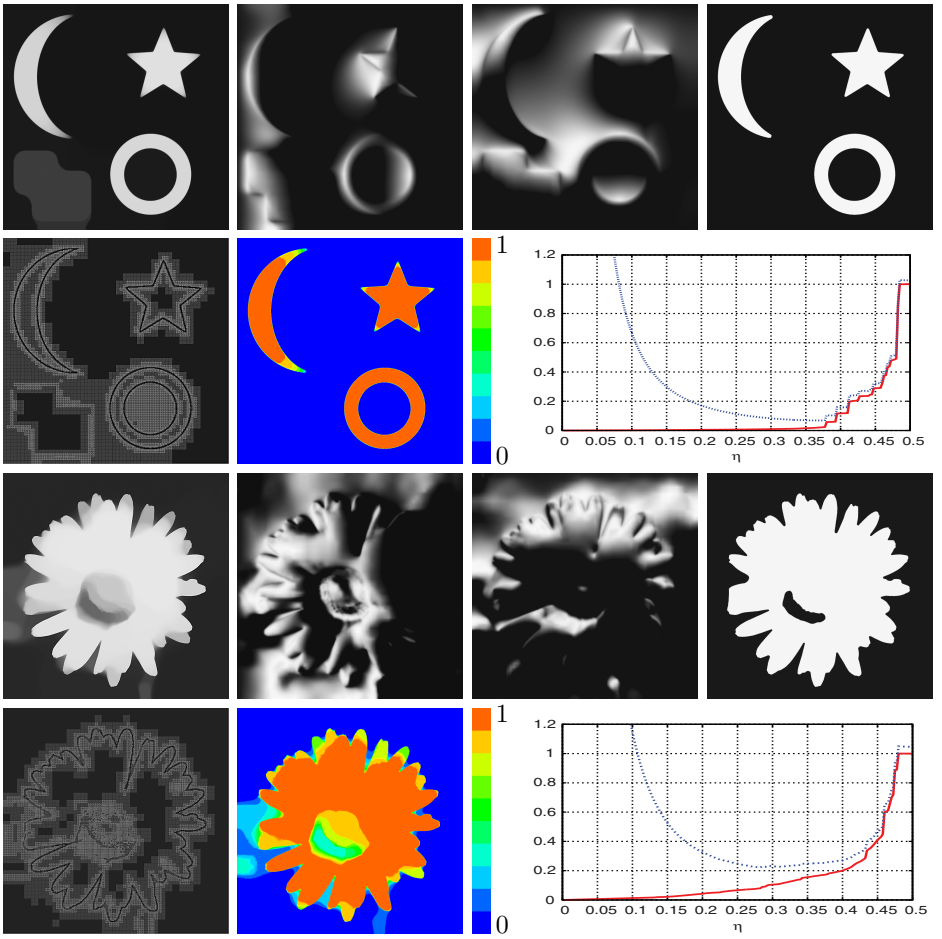

FIGURE 4. For images (b) (first and second row) and (c) (last two rows) and the discretization (FE'), the components of the relaxed solution u_h, $(p_h)_1$, $(p_h)_2$ and the resulting solution χ_h are shown after the 10th iteration of the adaptive scheme. In the second and fourth rows, the adaptive grid (after the 6th refinement step), deciles of the discrete solution u_h encoded with different colors, and the functions $\eta \mapsto \mathrm{a}[\mathbf{U}_h, \eta]$ (red solid line) and $\eta \mapsto \mathrm{err}_\chi$ (blue dashed line) are rendered.

Figure 4 displays solutions for the discretization (FE') and the corresponding adaptive meshes together with color coded deciles of u_h, and the graphs of $\eta \mapsto \mathrm{a}[\mathbf{U}_h, \eta]$ and $\eta \mapsto \mathrm{err}_\chi$. Note that the displayed deciles explicitly indicate the sets S_η for $\eta = 0.1, 0.2, 0.3, 0.4$.

Figures 5 and 6 show the relaxed solution u_h and the thresholded solution χ_h for the input images (b) to (d) using the discretization schemes (FE) and (FD). The corresponding results for (FE_D) are not depicted since they are almost indistinguishable from the results for (FE).

It is known that solutions of the Mumford-Shah problem are in general not unique. We pick up a classical example for this non-uniqueness in Figure 7 with alternating intensity values 0 and 1 on quadrants. We demonstrate the sensitivity

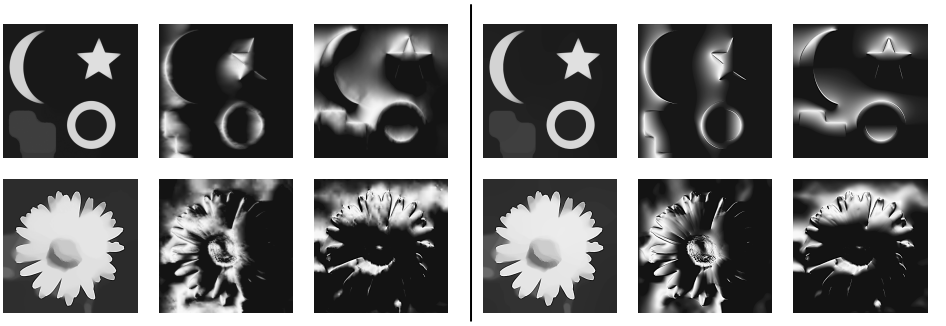

FIGURE 5. Relaxed solution u_h, $(p_h)_1$, $(p_h)_2$ for the input image (b) (top row) and (c) (bottom row) using the discretization (FE) (left, after 10 iterations of the adaptive algorithm) and (FD) (right).

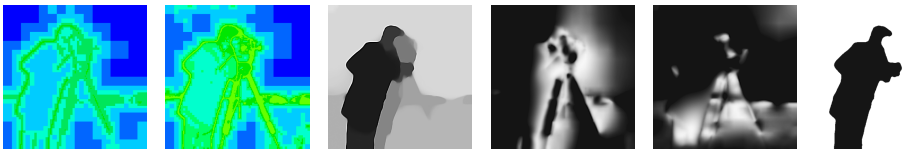

FIGURE 6. The mesh in the 5th and 10th iterations, the relaxed solution u_h, $(p_h)_1$, $(p_h)_2$ and χ_h for the input image (d) using the discretization (FE') after 10 iterations of the algorithm.

of our adaptive scheme with respect to the topology of the segmentation. In fact, we compute a segmentation using the discretization (FE) (with initial intensity values $c_1 = 1$, $c_2 = 0$ and $\nu = 0.01$) for slightly perturbed versions of the original image. That is, solely the four pixels in the center of the image of resolution 2049×2049 are either set black or white, respectively. The resulting segmentations are shown in Figure 7 along with the adaptive mesh and the decile plots. The adaptive algorithm is capable of detecting properly the decision for one of the two segmentation solutions. The issue of non-uniqueness is closely related to the flatness of the relaxed solution in the center, in particular leading to an increase of the a posteriori error contribution $\mathrm{a}[v, \eta]$.

Moreover, we applied the above methods to an analytic function consisting of a weighted sum of two Gaussian kernels. To this end, in each step the functionals and the error estimator are evaluated on the current adaptive grid and not on a prefixed full resolution grid. The results are shown in Figure 8 (with parameters $c_1 = 0.495349$, $c_2 = 0.0568447$ and $\nu = 5 \cdot 10^{-3}$).

In addition, we evaluated the sensitivity of the error estimates with respect to variations in the intensity values. Figure 9 shows the function plot of $\epsilon \mapsto \mathrm{err}_u^{2,\epsilon}[u_h, p_h, c_1, c_2]$ with fixed primal u_h and dual solution p_h obtained with the discretization (FE) after the 10th iteration for the images (b) and (c). The error estimator for image (b) is less sensitive to small fluctuations in the intensity values compared to image (c) due to the stronger variation of intensity values along the boundary of the segmentation in image (b).

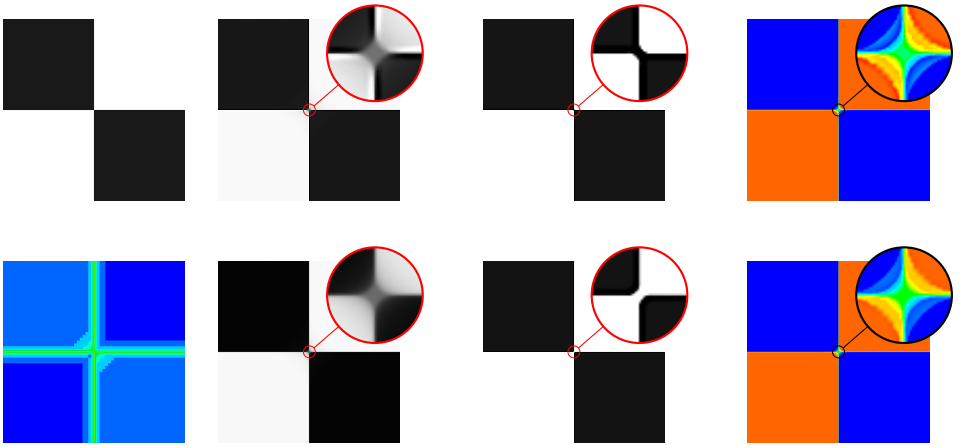

FIGURE 7. The original input image with resolution 2049×2049 and the adaptive mesh after the 6th iteration with white pixels in the center of the input image (first column). Second to fourth column: the relaxed solution u_h, the thresholded solution χ_h and the decile plots after the 10th iteration for black pixels in the center (first row) and white pixels in the center (second row) along with the corresponding zoom of the center (with zoom factor 8) using the discretization (FE).

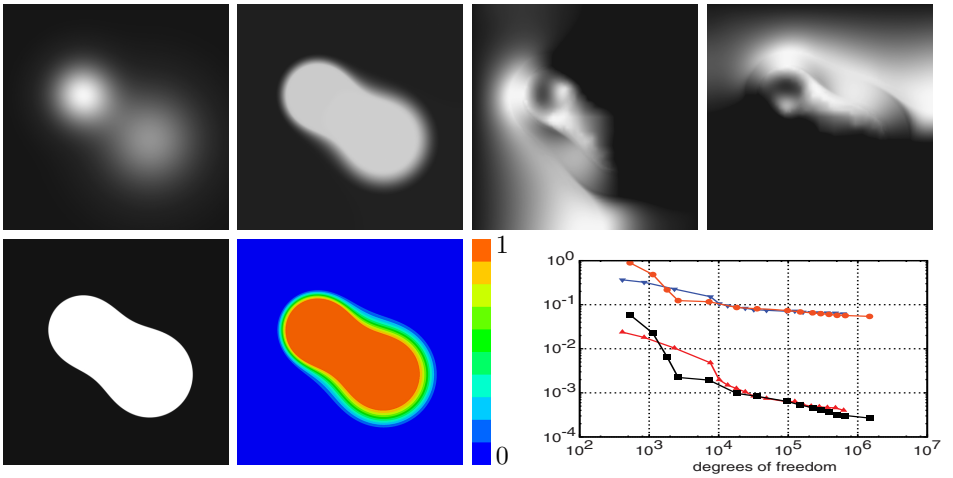

FIGURE 8. First row: Input image u_0 composed by the superposition of two Gaussian kernels, numerical solutions u_h, $(p_h)_1$ and $(p_h)_2$ computed via the adaptive algorithm using discretization (FE'). Second row: χ_h, deciles and the error estimators in a log-log plot (err_u^2 in red ▲, err_χ in blue ▼ for the discretization (FE'), and err_u^2 in black ■, err_χ in brown • for the discretization (FE)).

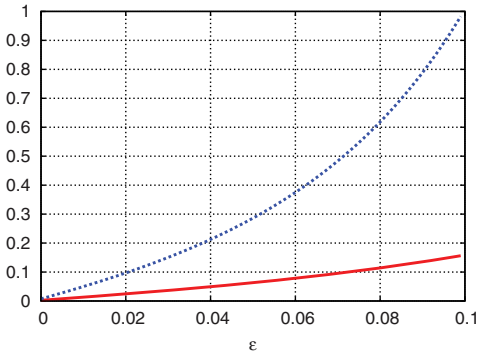

FIGURE 9. The function plot $\epsilon \mapsto \mathrm{err}_u^{2,\epsilon}$ for the images (b) (red solid line) and (c) (blue dashed line) using the discretization (FE) after the 10th iteration.

Finally, the proposed discretization schemes were compared in terms of the relative CPU time for the images (b) and (d) in the last iteration. To enforce comparable conditions the stopping criterion was set to $\|\mathbf{P}_h^{k+1} - \mathbf{P}_h^k\|_\infty < 10^{-6}$, and the primal and dual solutions were initialized with constant values. In comparison with the discretization scheme (FE), the scheme (FE') required comparable CPU time (image (b): -6.7%, image (d): $+6.5\%$), whereas (FE_D) performed slower for larger images (image (b): $+30.4\%$, image (d): -0.1%).

7. Conclusions

We have investigated the a posteriori error estimation for the binary Mumford-Shah model and applied this estimate to three different adaptive finite element discretizations in comparison to a non-adaptive finite difference scheme on a regular grid. The proposed finite element discretizations in combination with the adaptive mesh refinement strategy lead to a substantial reduction of the required degrees of freedom with error values err_u^2 and err_χ of about the same magnitude as for a standard finite difference scheme on a non-adaptive mesh with mesh size equal to the finest mesh size of the adaptive meshes. To improve the resulting estimate of the duality gap $E^{\mathrm{rel}}[v] + D^{\mathrm{rel}}[q]$, the finite element schemes require some oscillation damping smoothing in a post-processing step.

The proposed approach to a posteriori estimates for the binary Mumford-Shah model derived in this paper can be applied to more general problems in computer vision. In fact, the calibration method developed by Alberti, Bouchitté and Dal Maso [1] provides a convex relaxation of non-convex functionals of Mumford-Shah type via the lifting of a variational problem on a n-dimensional domain to a minimization problem over characteristic functions of subgraphs in $n+1$ dimensions. In the context of non-convex functionals in vision, this approach was studied by Pock *et al.* [40,41]. Applications of such functionals include the computation of minimal partitions [16, 38], the depth map identification from stereo images or the robust extraction of optimal flow fields [41]. Here, an adaptive mesh strategy is expected to have an even larger pay-off due to the increased dimension.

Acknowledgements

The second and third-named authors acknowledge support of the Collaborative Research Centre 1060 and the Hausdorff Center for Mathematics, both funded by the German Research Foundation. The first-named author was funded in part by the Excellence Initiative of the German federal and state governments. Furthermore, the authors would like to thank the anonymous referees for their valuable comments to improve this article.

References

[1] G. Alberti, G. Bouchitté, and G. Dal Maso, *The calibration method for the Mumford-Shah functional and free-discontinuity problems*, Calc. Var. Partial Differential Equations **16** (2003), no. 3, 299–333, DOI 10.1007/s005260100152. MR2001706 (2004k:49029)

[2] L. Ambrosio, *Variational problems in* SBV *and image segmentation*, Acta Appl. Math. **17** (1989), no. 1, 1–40, DOI 10.1007/BF00052492. MR1029833 (91d:49003)

[3] L. Ambrosio and V. M. Tortorelli, *On the approximation of free discontinuity problems* (English, with Italian summary), Boll. Un. Mat. Ital. B (7) **6** (1992), no. 1, 105–123. MR1164940 (93e:49010)

[4] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*, Oxford Mathematical Monographs, The Clarendon Press, Oxford University Press, New York, 2000. MR1857292 (2003a:49002)

[5] S. Bartels, *Error control and adaptivity for a variational model problem defined on functions of bounded variation*, Math. Comp. **84** (2015), no. 293, 1217–1240, DOI 10.1090/S0025-5718-2014-02893-7. MR3315506

[6] S. Bartels, *Broken Sobolev space iteration for total variation regularized minimization problems*, IMA J. Numer. Anal. **36** (2016), no. 2, 493–502, DOI 10.1093/imanum/drv023. MR3483094

[7] B. Berkels, *An unconstrained multiphase thresholding approach for image segmentation*, Proceedings of the Second International Conference on Scale Space Methods and Variational Methods in Computer Vision (SSVM 2009), Lecture Notes in Computer Science, vol. 5567, Springer, 2009, pp. 26–37.

[8] B. Berkels, *Joint methods in imaging based on diffuse image representations*, Dissertation, University of Bonn, 2010.

[9] B. Bourdin and A. Chambolle, *Implementation of an adaptive finite-element approximation of the Mumford-Shah functional*, Numer. Math. **85** (2000), no. 4, 609–646, DOI 10.1007/PL00005394. MR1771782 (2001g:65151)

[10] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Foundations and Trends in Machine Learning **3** (2010), 1–122.

[11] L. M. Brègman, *A relaxation method of finding a common point of convex sets and its application to the solution of problems in convex programming* (Russian), Ž. Vyčisl. Mat. i Mat. Fiz. **7** (1967), 620–631. MR0215617 (35 #6457)

[12] P. H. Calamai and J. J. Moré, *Projected gradient methods for linearly constrained problems*, Math. Programming **39** (1987), no. 1, 93–116, DOI 10.1007/BF02592073. MR909010 (89f:90132)

[13] A. Chambolle, *Image segmentation by variational methods: Mumford and Shah functional and the discrete approximations*, SIAM J. Appl. Math. **55** (1995), no. 3, 827–863, DOI 10.1137/S0036139993257132. MR1331589 (96h:68216)

[14] A. Chambolle, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vision **20** (2004), no. 1-2, 89–97, DOI 10.1023/B:JMIV.0000011320.81911.38. Special issue on mathematics and image analysis. MR2049783 (2005m:49058)

[15] A. Chambolle, *Total variation minimization and a class of binary MRF models*, Energy Minimization Methods in Computer Vision and Pattern Recognition (A. Rangarajan et al., eds.), Lecture Notes in Computer Science, vol. 3757, Springer, 2005, pp. 136–152.

[16] A. Chambolle, D. Cremers, and T. Pock, *A convex approach for computing minimal partitions*, Tech. Report 649, École Polytechnique, Centre de Mathématiques Appliquées, UMR CNRS 7641, November 2008.

[17] A. Chambolle, D. Cremers, and T. Pock, *A convex approach to minimal partitions*, SIAM J. Imaging Sci. **5** (2012), no. 4, 1113–1158, DOI 10.1137/110856733. MR3022190

[18] A. Chambolle and G. Dal Maso, *Discrete approximation of the Mumford-Shah functional in dimension two* (English, with English and French summaries), M2AN Math. Model. Numer. Anal. **33** (1999), no. 4, 651–672, DOI 10.1051/m2an:1999156. MR1726478 (2000j:49045)

[19] A. Chambolle and J. Darbon, *On total variation minimization and surface evolution using parametric maximum flows*, International Journal of Computer Vision **84** (2009), no. 3, 288–307.

[20] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision **40** (2011), no. 1, 120–145, DOI 10.1007/s10851-010-0251-1. MR2782122 (2012b:94004)

[21] T. F. Chan, S. Esedoḡlu, and M. Nikolova, *Algorithms for finding global minimizers of image segmentation and denoising models*, SIAM J. Appl. Math. **66** (2006), no. 5, 1632–1648, DOI 10.1137/040615286. MR2246072 (2007k:94012)

[22] T. F. Chan and L. A. Vese, *A level set algorithm for minimizing the Mumford-Shah functional in image processing*, IEEE/Computer Society Proceedings of the 1st IEEE Workshop on Variational and Level Set Methods in Computer Vision, 2001, pp. 161–168.

[23] Y. Chen, T. A. Davis, W. W. Hager, and S. Rajamanickam, *Algorithm 887: CHOLMOD, supernodal sparse Cholesky factorization and update/downdate*, ACM Trans. Math. Software **35** (2008), no. 3, Art. 22, 14, DOI 10.1145/1391989.1391995. MR2738209

[24] D. C. Dobson and C. R. Vogel, *Convergence of an iterative method for total variation denoising*, SIAM J. Numer. Anal. **34** (1997), no. 5, 1779–1791, DOI 10.1137/S003614299528701X. MR1472196 (99a:65081)

[25] I. Ekeland and R. Témam, *Convex Analysis and Variational Problems*, Corrected reprint of the 1976 English edition, Classics in Applied Mathematics, vol. 28, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1999. Translated from the French. MR1727362 (2000j:49001)

[26] L. C. Evans and R. F. Gariepy, *Measure Theory and Fine Properties of Functions*, Studies in Advanced Mathematics, CRC Press, Boca Raton, FL, 1992. MR1158660 (93f:28001)

[27] X. Feng and A. Prohl, *Analysis of total variation flow and its finite element approximations*, M2AN Math. Model. Numer. Anal. **37** (2003), no. 3, 533–556, DOI 10.1051/m2an:2003041. MR1994316 (2004f:65149)

[28] T. Goldstein and S. Osher, *The split Bregman method for L1-regularized problems*, SIAM J. Imaging Sci. **2** (2009), no. 2, 323–343, DOI 10.1137/080725891. MR2496060 (2010e:65087)

[29] W. Han, *A Posteriori Error Analysis via Duality Theory: With Applications in Modeling and Numerical Approximations*, Advances in Mechanics and Mathematics, vol. 8, Springer-Verlag, New York, 2005. MR2101057 (2005k:65003)

[30] M. Hintermüller and K. Kunisch, *Path-following methods for a class of constrained minimization problems in function space*, Tech. report, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, July 2004.

[31] M. Hintermüller and K. Kunisch, *Total bounded variation regularization as a bilaterally constrained optimization problem*, SIAM J. Appl. Math. **64** (2004), no. 4, 1311–1333, DOI 10.1137/S0036139903422784. MR2068672 (2005b:49062)

[32] K. Jalalzai, *Discontinuities of the minimizers of the weighted or anisotropic total variation for image reconstruction*, arXiv:1402.0026, 2014.

[33] S. P. Lloyd, *Least squares quantization in PCM*, IEEE Trans. Inform. Theory **28** (1982), no. 2, 129–137, DOI 10.1109/TIT.1982.1056489. MR651807 (84a:94012)

[34] L. Modica and S. Mortola, *Un esempio di Γ^--convergenza* (Italian, with English summary), Boll. Un. Mat. Ital. B (5) **14** (1977), no. 1, 285–299. MR0445362 (56 #3704)

[35] J.-M. Morel and S. Solimini, *Variational Methods in Image Segmentation: With Seven Image Processing Experiments*, Progress in Nonlinear Differential Equations and their Applications, 14, Birkhäuser Boston, Inc., Boston, MA, 1995. MR1321598

[36] D. Mumford and J. Shah, *Optimal approximations by piecewise smooth functions and associated variational problems*, Comm. Pure Appl. Math. **42** (1989), no. 5, 577–685, DOI 10.1002/cpa.3160420503. MR997568 (90g:49033)

[37] S. Özışık, Beatrice Riviere, and Tim Warburton, *On the constants in inverse inequalities in L^2*, Tech. Report 10-19, Department of Computational & Applied Mathematics, Rice University, 2010.

[38] T. Pock, T. Schoenemann, G. Graber, H. Bischof, and D. Cremers, *A convex formulation of continuous multi-label problems*, ECCV08, 2008, pp. III: 792–805.

[39] T. Pock, *Diagonal preconditioning for first order primal-dual algorithms in convex optimization*, IEEE International Conference on Computer Vision (ICCV), 2011, pp. 1762 – 1769.

[40] T. Pock, D. Cremers, H. Bischof, and A. Chambolle, *An algorithm for minimizing the Mumford-Shah functional*, IEEE 12th International Conference on Computer Vision, 2009.

[41] T. Pock, D. Cremers, H. Bischof, and A. Chambolle, *Global solutions of variational models with convex regularization*, SIAM J. Imaging Sci. **3** (2010), no. 4, 1122–1145, DOI 10.1137/090757617. MR2763711 (2012e:49063)

[42] S. I. Repin, *A posteriori error estimation for variational problems with uniformly convex functionals*, Math. Comp. **69** (2000), no. 230, 481–500, DOI 10.1090/S0025-5718-99-01190-4. MR1681096 (2000i:49046)

[43] S. I. Repin, *A Posteriori Estimates for Partial Differential Equations*, Radon Series on Computational and Applied Mathematics, vol. 4, Walter de Gruyter GmbH & Co. KG, Berlin, 2008. MR2458008

[44] R. T. Rockafellar, *Convex Analysis*, Princeton Landmarks in Mathematics, Princeton University Press, Princeton, NJ, 1997. Reprint of the 1970 original; Princeton Paperbacks. MR1451876 (97m:49001)

[45] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D **60** (1992), no. 1-4, 259–268. Experimental mathematics: computational issues in nonlinear science (Los Alamos, NM, 1991). MR3363401

[46] J. Shen, Γ*-convergence approximation to piecewise constant Mumford-Shah segmentation*, Proceedings of the 7th International Conference on Advanced Concepts for Intelligent Vision Systems (ACIVS 2005), Lecture Notes in Computer Science, vol. 3708, Springer, 2005, pp. 499–506.

[47] J. Wang and B. J. Lucier, *Error bounds for finite-difference methods for Rudin-Osher-Fatemi image smoothing*, SIAM J. Numer. Anal. **49** (2011), no. 2, 845–868, DOI 10.1137/090769594. MR2792398 (2012h:65244)

AICES Graduate School, RWTH Aachen University, 52062 Aachen, Germany
E-mail address: berkels@aices.rwth-aachen.de

Institute for Numerical Simulation, University of Bonn, 53115 Bonn, Germany
E-mail address: alexander.effland@ins.uni-bonn.de

Institute for Numerical Simulation, University of Bonn, 53115 Bonn, Germany
E-mail address: martin.rumpf@ins.uni-bonn.de