

STABILITY AND EXPERIMENTAL COMPARISON OF PROTOTYPICAL ITERATIVE SCHEMES FOR TOTAL VARIATION REGULARIZED PROBLEMS

SÖREN BARTELS AND MARIJO MILICEVIC

ABSTRACT. Various iterative methods are available for the approximate solution of nonsmooth minimization problems. For a popular nonsmooth minimization problem arising in image processing the suitable application of three prototypical methods and their stability is discussed. The methods are compared experimentally with a focus on choice of stopping criteria, influence of rough initial data, step sizes as well as mesh sizes and an overview of existing algorithms is given.

1. INTRODUCTION

In this paper we deal with the minimization of the TV - L^2 functional

$$I(u) = |Du|(\Omega) + \frac{\alpha}{2} \|u - g\|_{L^2(\Omega)}^2$$

which is a prototypical model problem for total variation regularized minimization problems. The minimization of functionals that involve the BV -seminorm $|Du|(\Omega)$ are particularly interesting in applications where functions with discontinuities are desired, such as in image denoising where sharp edges of images should be preserved, or in the modelling of perfect elastoplasticity, damage and fracture in continuum mechanics where spatial jumps should be allowed for the symmetric gradient of the displacement field or the damage variable, respectively, see, e.g., [6, 21, 40].

The minimization of I , the so called *unconstrained ROF problem* (or simply ROF problem), has been first proposed in [39] in the context of an image denoising problem. Despite the seemingly simple structure of the functional I , its minimization by numerical methods poses a challenging problem due to the non-differentiability of the BV -seminorm. Since its introduction in the aforementioned paper many algorithms for the minimization of I have been developed many of which motivated by image processing problems.

I.A. Regularization. The authors in [39] considered a regularization of the BV -seminorm and employed an explicit time discretization of the parabolic

Date: January 7, 2016.

1991 Mathematics Subject Classification. 65K15 (49M29).

PDE with homogeneous Neumann boundary condition

$$\partial_t u = -\operatorname{div}\left(\frac{\nabla u}{\sqrt{|\nabla u|^2 + \varepsilon^2}}\right) + \alpha(u - g)$$

with a finite difference discretization of the involved differential operators. The drawbacks of this approach are that we lose the possibility of sharp jumps of minimizers across lower-dimensional subsets on the one hand and that the condition on the time-step size is very restrictive, namely $\tau \leq ch^2$ with h being the mesh size of the underlying grid, on the other hand. A similar technique has been proposed in [42, 23] where the authors applied a fixed-point iteration to solve the perturbed Euler-Lagrange equation

$$-\operatorname{div}\left(\frac{\nabla u}{|\nabla u|_\varepsilon}\right) + \alpha(u - g) = 0$$

with $|\nabla u|_\varepsilon = \sqrt{|\nabla u|^2 + \varepsilon^2}$. The authors in [17] considered a Newton method for the primal-dual system

$$\begin{aligned} |\nabla u|_\varepsilon p - \nabla u &= 0, \\ -\operatorname{div} p + \alpha(u - g) &= 0, \end{aligned}$$

and observed that the Newton iteration for this system is better behaved than the Newton iteration for the Euler-Lagrange equation. However, it remains unclear how the convergence depends on the regularization parameter ε .

In general, the performance of algorithms based on regularization approaches depends critically on the regularization parameter ε . By introducing an auxiliary variable s_h we devise a fully practical, unconditionally stable iterative method to solve the regularized Euler-Lagrange equations.

I.B. Splitting methods. Another class of methods is based on a splitting ansatz where a new variable $\sigma = \nabla u$ is introduced which transforms the minimization of the TV - L^2 functional with respect to u into the constrained minimization problem

$$\inf_{\sigma, u} \int_{\Omega} |\sigma| \, dx + \frac{\alpha}{2} \|u - g\|_{L^2(\Omega)}^2 \quad \text{s.t. } \sigma = \nabla u.$$

In [44] the authors proposed an alternating minimization scheme with respect to the variables u and σ for a functional that results from the L^1 - L^2 functional after adding the penalization term $\frac{\delta^{-1}}{2} \|\sigma - \nabla u\|_{L^2(\Omega)}$ that enforces the constraint when decreasing the penalization parameter δ . In [32] the authors introduce the so-called *split Brègman method* which employs a Brègman iterative scheme [11] in order to approximate a minimizer of the constrained L^1 - L^2 minimization problem and to enforce the equality constraint strictly via the Brègman iteration. A closely related method is the augmented Lagrangian method [29] that has been applied in [45] to the constrained L^1 - L^2 minimization problem, where, in addition to the penalization term, the term $(\lambda, \sigma - \nabla u)_{L^2}$ with λ being a Lagrange multiplier is added in

order to strictly enforce the constraint $\sigma = \nabla u$. In [45], it has also been noted that the iteration scheme in the augmented Lagrangian method is equivalent to that in the split Brègman method. Another alternating minimization scheme, which is closely related to the augmented Lagrangian method, has been proposed in [41] where the variable u_h is updated via minimization of the Lagrangian functional instead of the augmented Lagrangian functional, i.e., omitting the penalization term in the update rule for u_h , while σ_h is updated by minimizing the augmented Lagrangian functional. Yet, this algorithm requires a step size $\tau = \mathcal{O}(h^2)$ for convergence. In [31] the authors discuss acceleration techniques for the augmented Lagrangian method and the alternating minimization algorithm proposed in [41]. In order to enforce stability, a restart condition has to be included in the accelerated augmented Lagrangian method which may negatively affect the performance.

A critical aspect of most proposed splitting methods is that the consistency error $\sigma - \nabla u$ is measured in the L^2 -norm. However, since the minimizer u of I lies only in $BV(\Omega)$, we cannot in general expect sequences $(\nabla u_h)_{h>0}$ of approximations to be bounded in $L^2(\Omega)$. We therefore consider the augmented Lagrangian method with a weighted L^2 -norm for which the sequence $(\nabla u_h)_{h>0}$ is guaranteed to be bounded due to an inverse estimate and its boundedness in $L^1(\Omega)$.

I.C. Saddle-point approach. Another approach is based on the definition of the BV -seminorm as the operator norm of the distributional derivative Du and converts the TV - L^2 minimization problem into the problem of finding a saddle-point of the functional

$$S(u, p) = - \int_{\Omega} u \operatorname{div} p \, dx + \frac{\alpha}{2} \|u - g\|_{L^2(\Omega)}^2 - I_{K_1(0)}(p).$$

In [14], starting from a finite difference discretization of the functional I , the authors defined a primal-dual algorithm, which is a proximal-point algorithm for finding a saddle-point of S . Stability is guaranteed by choosing the involved step size as $\tau = \mathcal{O}(h)$. Various acceleration techniques are also discussed in [14] where variable step sizes and variable extrapolation parameters are considered. In [4], the author adapted the ideas from [14] to define a primal-dual algorithm for the approximation of a saddle-point of S using finite elements which takes the form of a semi-implicit time discretization of an L^2 -(sub-)gradient flow based on the optimality conditions for a saddle-point of S . The restrictive step size $\tau = \mathcal{O}(h)$ is required for stability as well. Recently, the author in [3] noted that if $g \in L^\infty(\Omega)$ then the discrete minimizers of I are uniformly bounded in $C(\bar{\Omega}) \cap BV(\Omega)$. This motivates to use a discrete variant of the inner product in $H^{1/2}(\Omega)$ as a preconditioner for the linear system of equations associated to the optimality condition for u_h and yields a weaker restriction on the step size, namely $\tau = \mathcal{O}(h^{1/2})$.

I.D. Other approaches. Further methods are based on the dual functional of I , the so called *dual methods*. In [12], the author considers the dual functional of a finite difference discretization of I and proposes a semi-implicit gradient-descent algorithm for solving the associated constrained optimization problem which requires a step size of order $\tau = \mathcal{O}(h^2)$. In [15], the authors start from a finite difference discretization of I and introduce two new variables u_1 and u_2 in such a way that a forward difference quotient in horizontal direction is applied to u_1 and a forward difference quotient in vertical direction is applied to u_2 . The resulting constraints are enforced with Lagrange multipliers. The obtained saddle-point problem is then converted into a maximization problem in the two Lagrange multipliers.

In [35] the authors reformulate the ROF problem as a bilaterally constrained optimization problem by considering the predual of the $TV-L^2$ functional with anisotropic BV -seminorm, that is, the ℓ^1 -norm of ∇u . They propose semi-smooth Newton methods (cf. [34]) for regularized versions of the predual problem.

In [30] the constrained *ROF* problem (cf. [39]) is formulated as a second-order cone program which is in turn solved by an interior-point algorithm. This method is related to the approach in [17], as observed in [46].

Finally, starting from a discretization of I , the authors in [22] decompose the variable u into its level sets and transform the minimization problem into independent binary Markov Random Fields associated to each level set. The minimization is then realized with a graph cut algorithm.

I.E. Objectives. This paper aims at answering the question which PDE-based methods may be the most appropriate, accurate and effective ones for the minimization of total variation regularized minimization problems in order to have a clear and unified statement for total variation minimization problems arising both in image processing and in other applications such as, e.g., continuum mechanics. Based on the given overview it seems sufficient to take iterative schemes for regularized variants of I , primal-dual schemes for the saddle-point problem defined by S and splitting methods into consideration since these are prototypical methods in the context of PDEs and for which a rigorous numerical analysis can be carried out in the sense that a stability estimate is available which is robust in the discretization parameters.

I.F. Outline. The paper is structured as follows: in Section 2 we state the model problem, review important properties of the minimizer of the functional I , state some duality relations and recall the finite element discretization of the ROF problem. We also introduce some basic notation regarding iterative solution schemes and review identities that are important in the solution of subdifferential inclusions. In Section 3 we introduce the *Heron method*, which is an iterative scheme for solving the discrete minimization problem associated with a regularization of I and discuss its stability. The properly weighted augmented Lagrangian method is introduced in Section 4

and we establish an adequately weighted L^2 -norm in order to guarantee stability of the method before we review the primal-dual method proposed in [3, 4] in Section 5. In Section 6 we show experimental results for two examples with different data functions g and different values of α . In Section 7 we give for each approach stated in the introduction an explanation why we decided to not include them into our comparison. After concluding the paper in Section 8, we discuss the computation of roots of quartic algebraic equations which is needed in the Heron method in Appendix A.

2. PRELIMINARIES

II.A. ROF model problem. For a bounded Lipschitz domain $\Omega \subset \mathbb{R}^d$, a given function $g \in L^\infty(\Omega)$, and a parameter $\alpha > 0$ we consider the minimization of

$$I(u) = \int_{\Omega} |Du| + \frac{\alpha}{2} \|u - g\|^2.$$

Here, $\|\cdot\|$ denotes the L^2 -norm on Ω and the first term on the right-hand side is the total variation of $u \in L^1(\Omega)$, given by

$$\int_{\Omega} |Du| = \sup \left\{ - \int_{\Omega} u \operatorname{div} q \, dx : q \in C_c^1(\Omega; \mathbb{R}^d), |q| \leq 1 \text{ a.e.} \right\},$$

where $|\cdot|$ denotes the Euclidean norm. Due to the convexity of the total variation and strong convexity of the squared L^2 -norm there exists a unique minimizer $u \in BV(\Omega) \cap L^2(\Omega)$ for I such that

$$\frac{\alpha}{2} \|u - v\|^2 \leq I(v) - I(u)$$

for every $v \in BV(\Omega) \cap L^2(\Omega)$. The space $BV(\Omega) \subset L^1(\Omega)$ consists of all $v \in L^1(\Omega)$ with finite total variation. For more details concerning the space $BV(\Omega)$ see, e.g., [1, 2]. A cut-off argument and the chain rule in $BV(\Omega)$ imply that we have

$$\|u\|_{L^\infty(\Omega)} \leq \|g\|_{L^\infty(\Omega)}.$$

For $u \in BV(\Omega) \cap L^2(\Omega)$ the supremum in the characterization of the total variation of u can be taken in the set of functions $q \in H_N(\operatorname{div}; \Omega)$, where

$$H_N(\operatorname{div}; \Omega) = \{q \in L^2(\Omega; \mathbb{R}^d) : \operatorname{div} q \in L^2(\Omega), q \cdot n = 0 \text{ on } \partial\Omega\}.$$

The minimization of I is thus equivalent to a saddle-point problem defined by the functional

$$\widehat{S}(u, p) = - \int_{\Omega} u \operatorname{div} p \, dx + \frac{\alpha}{2} \|u - g\|^2 - I_{K_1(0)}(p)$$

with the indicator functional $I_{K_1(0)}$ of the set $K_1(0)$ consisting of all vector fields in $L^2(\Omega; \mathbb{R}^d)$ with length uniformly bounded by one. The optimality conditions for a saddle point $(u, p) \in BV(\Omega) \cap L^2(\Omega) \times H_N(\operatorname{div}; \Omega)$ read

$$- \operatorname{div} p + \alpha(u - g) = 0, \quad p \in \partial|\nabla u|.$$

The subdifferential inclusion for p is equivalent to

$$p(x) \in \begin{cases} \{\nabla u(x)/|\nabla u(x)|\} & \text{if } \nabla u(x) \neq 0, \\ K_1(0) & \text{if } \nabla u(x) = 0, \end{cases}$$

for almost every $x \in \Omega$. Eliminating $u = (\operatorname{div} p + \alpha g)/\alpha$ from the saddle-point functional we see that p is maximal for

$$\begin{aligned} D(p) &= -\frac{1}{\alpha} \int_{\Omega} (\operatorname{div} p + \alpha g) \operatorname{div} p \, dx + \frac{1}{2\alpha} \|\operatorname{div} p\|^2 - I_{K_1(0)}(p) \\ &= -\frac{1}{2\alpha} \|\operatorname{div} p + \alpha g\|^2 + \frac{\alpha}{2} \|g\|^2 - I_{K_1(0)}(p). \end{aligned}$$

This dual problem admits solutions which are nonunique in general. It can be shown that strong duality, i.e., $\inf_u I(u) = \sup_p D(p)$, holds, cf. [25].

II.B. Finite element discretization. For a regular triangulation \mathcal{T}_h of Ω into triangles or tetrahedra of maximal diameter $h > 0$, we define the space of continuous, elementwise affine functions on \mathcal{T}_h via

$$\mathcal{S}^1(\mathcal{T}_h) = \{v_h \in C(\overline{\Omega}) : v_h|_T \text{ affine for all } T \in \mathcal{T}_h\}.$$

Elementwise constant functions are contained in the space

$$\mathcal{L}^0(\mathcal{T}_h) = \{q_h \in L^\infty(\Omega) : q_h|_T \text{ constant for all } T \in \mathcal{T}_h\}$$

and the set of elementwise constant vector fields is denoted by $\mathcal{L}^0(\mathcal{T}_h)^d$. Note that unless \mathcal{T}_h consists of one element only, we have $\mathcal{L}^0(\mathcal{T}_h)^d \not\subset H_N(\operatorname{div}; \Omega)$. The restriction of the minimization of I to $\mathcal{S}^1(\mathcal{T}_h) \subset W^{1,1}(\Omega)$ leads to minimizing

$$I(u_h) = \int_{\Omega} |\nabla u_h| \, dx + \frac{\alpha}{2} \|u_h - g\|^2$$

and admits a unique minimizer for which we have the error estimate

$$\frac{\alpha}{2} \|u - u_h\|^2 \leq I(u_h) - I(u) \leq \min_{v_h \in \mathcal{S}^1(\mathcal{T}_h)} I(v_h) - I(u) \leq ch^{1/2}, \quad (1)$$

cf. [43, 7]. The error estimate is suboptimal in the sense that for every $u \in BV(\Omega) \cap L^\infty(\Omega)$ we have

$$\min_{v_h \in \mathcal{S}^1(\mathcal{T}_h)} \|u - v_h\|^2 \leq ch,$$

cf. [5], but this rate cannot be expected in general. Numerical approximations of u appear to be closer to the true solution than this estimate guarantees but this seems to be related to the staircasing effect of the ROF model which implies that solutions tend to develop steps instead of smooth transitions. A discrete saddle-point formulation is defined by the functional

$$S(u_h, p_h) = \int_{\Omega} \nabla u_h \cdot p_h \, dx - I_{K_1(0)}(p_h) + \frac{\alpha}{2} \|u_h - g\|^2$$

for $(u_h, p_h) \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{L}^0(\mathcal{T}_h)^d$. The corresponding nonconforming discrete dual formulation seeks $p_h \in \mathcal{L}^0(\mathcal{T}_h)^d$ as a maximizer for

$$D_h(p_h) = -\frac{1}{2\alpha} \|\nabla'_h p_h + \alpha g\|^2 + \frac{\alpha}{2} \|g\|^2 - I_{K_1(0)}(p_h).$$

The operator $\nabla'_h : \mathcal{L}^0(\mathcal{T}_h)^d \rightarrow \mathcal{S}^1(\mathcal{T}_h)$ is for $q_h \in \mathcal{L}^0(\mathcal{T}_h)^d$ defined by $w_h = \nabla'_h q_h$ with $w_h \in \mathcal{S}^1(\mathcal{T}_h)$ such that

$$\int_{\Omega} v_h w_h \, dx = - \int_{\Omega} \nabla v_h \cdot q_h \, dx$$

for all $v_h \in \mathcal{S}^1(\mathcal{T}_h)$ and is an approximation of the divergence operator subject to homogeneous Neumann boundary conditions.

II.C. Iterative solution. The iterative schemes for the numerical solution of the discretized ROF problem discussed below may be regarded as discretizations of evolution equations. Accordingly, we use the backward difference quotient operator d_t , defined for a step size $\tau > 0$ and a sequence $(a^j)_{j \geq 0}$ via

$$d_t a^{j+1} = (a^{j+1} - a^j) / \tau$$

for $j = 0, 1, \dots$. We often abbreviate the L^2 -inner product by

$$(v, w) = \int_{\Omega} v \cdot w \, dx$$

for functions or vector fields $v, w \in L^2(\Omega; \mathbb{R}^\ell)$, $\ell \in \{1, d\}$. Implicitly discretized gradient or subdifferential flows lead to sequences of minimization problems which seek for given p^j the minimizer p^{j+1} of the mapping

$$p \mapsto \frac{1}{2\tau} \|p - p^j\|^2 + F(p).$$

The related optimality condition reads

$$-d_t p^{j+1} \in \partial F(p^{j+1}).$$

An important ingredient for the development of efficient numerical methods for the ROF model is that certain related nonlinearities in such minimization problems can be solved explicitly, e.g., the minimization of

$$p \mapsto \frac{1}{2} \|p - q\|^2 + I_{K_1(0)}(p)$$

is solved by the best approximation of q within $K_1(0)$ and given by the pointwise shrinkage operation

$$p = \frac{q}{\max\{1, |q|\}}.$$

Via convex duality this is related to the solution of minimization problems of the form

$$s \mapsto \frac{1}{2} \|s - r\|^2 + c_1 \int_{\Omega} |s| \, dx.$$

The strong convexity of the squared L^2 -norm implies the unique solvability with the pointwise optimality condition

$$r - s \in c_1 \partial|s|,$$

where $\partial|s|$ denotes the subdifferential of the modulus at s . For $s \neq 0$ it follows that

$$r - s = c_1 \frac{s}{|s|} \iff r = (c_1 + |s|) \frac{s}{|s|},$$

i.e., s is parallel to r with modulus given by

$$|r| = c_1 + |s| \iff |s| = |r| - c_1.$$

If $s = 0$ then it follows that

$$r \in c_1 \overline{B_1(0)} \iff |r| \leq c_1.$$

Hence, the minimizing s is given by the pointwise operation

$$s = (|r| - c_1)_+ \frac{r}{|r|},$$

where $(t)_+ = \max\{t, 0\}$.

3. REGULARIZATION

A canonical way to deal with the nondifferentiability of the minimization follows from a regularization of the functional I , i.e., considering for a given $\varepsilon > 0$ the minimization of the functional

$$I_\varepsilon(u_h) = \int_{\Omega} (|\nabla u_h|^2 + \varepsilon^2)^{1/2} dx + \frac{\alpha}{2} \|u_h - g\|^2$$

for $u_h \in \mathcal{S}^1(\mathcal{T}_h)$. Noting that

$$|a| \leq (|a|^2 + \varepsilon^2)^{1/2} \leq |a| + \varepsilon$$

it follows that for the numerical solution $u_{\varepsilon,h} \in \mathcal{S}^1(\mathcal{T}_h)$ of the regularized problem we have

$$\begin{aligned} \frac{\alpha}{2} \|u - u_{\varepsilon,h}\|^2 &\leq I(u_{\varepsilon,h}) - I(u) \\ &\leq \min_{v_h \in \mathcal{S}^1(\mathcal{T}_h)} I_\varepsilon(v_h) - I(u) \\ &\leq \min_{v_h \in \mathcal{S}^1(\mathcal{T}_h)} I(v_h) - I(u) + \varepsilon|\Omega| \\ &\leq ch^{1/2} + \varepsilon|\Omega|, \end{aligned}$$

cf. [5], which suggests to choose $\varepsilon = \mathcal{O}(h^{1/2})$ to retain the same qualitative approximation properties as for the unregularized problem. In order not to violate the best possible convergence rate $\mathcal{O}(h^{1/2})$ we choose $\varepsilon = \mathcal{O}(h)$. Similarly, we have for the discrete minimizer $u_h \in \mathcal{S}^1(\mathcal{T}_h)$ of I

$$\frac{\alpha}{2} \|u_h - u_{\varepsilon,h}\|^2 \leq \varepsilon|\Omega|. \quad (2)$$

The inequality (2) will be useful in our experimental comparison. The practical computation of $u_{\varepsilon,h}$ is difficult since classical iterative schemes such as Newton or descent methods may depend critically on ε . To construct a stable numerical method we note that for every $a \geq 0$ we have

$$a^{1/2} = \inf_{s \neq 0} \frac{a}{2s^2} + \frac{s^2}{2}.$$

The optimal s satisfies $s = a^{1/4}$. This motivates to consider the augmented functional

$$\widehat{I}_\varepsilon(u_h, s_h) = \int_\Omega \frac{|\nabla u_h|^2 + \varepsilon^2}{2s_h^2} + \frac{s_h^2}{2} dx + \frac{\alpha}{2} \|u_h - g\|^2$$

for $(u_h, s_h) \in \mathcal{S}^1(\mathcal{T}_h) \in \mathcal{L}^0(\mathcal{T}_h)$. Note that we have

$$\min_{s_h \in \mathcal{L}^0(\mathcal{T}_h)} \widehat{I}_\varepsilon(u_h, s_h) = I_\varepsilon(u_h)$$

for every $u_h \in \mathcal{S}^1(\mathcal{T}_h)$ with optimal s_h given by

$$s_h = (|\nabla u_h|^2 + \varepsilon^2)^{1/4}.$$

An important aspect is that for a sequence $(u_h)_{h>0}$ which is bounded in $W^{1,1}(\Omega)$ we have that the corresponding sequence $(s_h)_{h>0}$ is bounded in $L^2(\Omega)$. This is not the case for the frequently employed variant

$$\widehat{I}'_\varepsilon(u_h, s_h) = \int_\Omega \frac{s_h(|\nabla u_h|^2 + \varepsilon^2)}{2} + \frac{1}{2s_h} dx + \frac{\alpha}{2} \|u_h - g\|^2,$$

often referred to as a *half-quadratic functional*, where for given u_h the optimal s_h is given by $s_h = (|\nabla u_h|^2 + \varepsilon^2)^{-1/2}$, cf. [23, 18]. A rigorous functional analytical framework is important to define a robust iterative scheme for minimizing \widehat{I}_ε . Noting that \widehat{I}_ε is separately convex, we use a decoupled gradient descent method in both variables, defined by the following algorithm.

Algorithm III.1 (Heron method). *Choose $\tau > 0$ and $(u_h^0, s_h^0) \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{L}^0(\mathcal{T}_h)$ with $s_h^0 > 0$ in Ω , set $j = 0$.*

(1) *Compute the minimizer $u_h^{j+1} \in \mathcal{S}^1(\mathcal{T}_h)$ of the mapping $u_h \mapsto \widehat{I}_\varepsilon(u_h, s_h^j) + \frac{1}{2\tau} \|u_h - u_h^j\|^2$, i.e., u_h^{j+1} satisfies*

$$(d_t u_h^{j+1}, v_h) = - \int_\Omega \frac{\nabla u_h^{j+1} \cdot \nabla v_h}{(s_h^j)^2} dx - \alpha \int_\Omega (u_h^{j+1} - g) v_h dx$$

for all $v_h \in \mathcal{S}^1(\mathcal{T}_h)$.

(2) *Compute the minimizer $s_h^{j+1} \in \mathcal{L}^0(\mathcal{T}_h)$, $s_h^{j+1} > 0$, of the mapping $s_h \mapsto \widehat{I}_\varepsilon(u_h^{j+1}, s_h) + \frac{1}{2\tau} \|s_h - s_h^j\|^2$, i.e., s_h^{j+1} satisfies*

$$(d_t s_h^{j+1}, r_h) = \int_\Omega \left(\frac{|\nabla u_h^{j+1}|^2 + \varepsilon^2}{(s_h^{j+1})^3} - s_h^{j+1} \right) r_h dx$$

for all $r_h \in \mathcal{L}^0(\mathcal{T}_h)$, and $s_h^{j+1} > 0$.

(3) Stop if $\|d_t u_h^{j+1}\| + \|d_t s_h^{j+1}\| \leq \varepsilon_{stop}$; increase $j \rightarrow j+1$ and continue with (1) otherwise.

The iteration is well-defined. While u_h^{j+1} is a solution of a linear system of equations, the equation defining s_h^{j+1} is nonlinear but can be solved explicitly elementwise with a uniformly positive solution.

Lemma III.2. *If $s_h^0 > 0$, then the function $s_h^{j+1} \in \mathcal{L}^0(\mathcal{T}_h)$ satisfies $s_h^{j+1} > 0$ and*

$$s_h^{j+1} - s_h^j = \tau \frac{|\nabla u_h^{j+1}|^2 + \varepsilon^2}{(s_h^{j+1})^3} - \tau s_h^{j+1}.$$

If we initialize s_h^0 such that $s_h^0 \geq \sqrt{\varepsilon}$, we have $s_h^{j+1} \geq \sqrt{\varepsilon}$ for all $j \geq 0$.

Proof. Choosing the characteristic function $r_h = \chi_T$ for arbitrary $T \in \mathcal{T}_h$ in Algorithm III.1 we obtain the first assertion. Note that for every $T \in \mathcal{T}_h$, the value $s_h^{j+1}|_T$ is the unique positive zero of the continuous, monotonically increasing function $f_T^{j+1} : (0, \infty) \rightarrow \mathbb{R}$,

$$f_T^{j+1}(x) := \left(1 + \frac{1}{\tau}\right)x - \frac{|\nabla u_h^{j+1}|_T|^2 + \varepsilon^2}{x^3} - \frac{s_h^j|_T}{\tau}.$$

It follows from the fact that for $x = \left(\frac{|\nabla u_h^{j+1}|_T|^2 + \varepsilon^2}{1 + \frac{1}{\tau}}\right)^{1/4}$ we have

$$f_T^{j+1}(x) = -\frac{s_h^j|_T}{\tau} < 0$$

since $s_h^0 > 0$. By the strict monotonicity of f_T^{j+1} it follows that $s_h^{j+1} > \left(\frac{|\nabla u_h^{j+1}|_T|^2 + \varepsilon^2}{1 + \frac{1}{\tau}}\right)^{1/4}$. If we initialize s_h^0 such that $s_h^0 \geq \sqrt{\varepsilon}$, we get

$$f_T^{j+1}(\sqrt{\varepsilon}) = \frac{\sqrt{\varepsilon} - s_h^j|_T}{\tau} - \frac{|\nabla u_h^{j+1}|_T|^2}{\varepsilon^{3/2}} \leq \frac{\sqrt{\varepsilon} - s_h^j|_T}{\tau} \leq 0,$$

and thus $s_h^{j+1} \geq \sqrt{\varepsilon}$ for all $j \geq 0$. \square

Remark III.3. *Consider the fourth order polynomial*

$$\widehat{f}_T^{j+1}(x) := x^4 - \frac{s_h^j|_T}{1 + \tau}x^3 - \frac{\tau}{1 + \tau}(|\nabla u_h^{j+1}|_T|^2 + \varepsilon^2). \quad (3)$$

The fact that $x = 0$ is not a zero of \widehat{f}_T^{j+1} implies that f_T^{j+1} - extended to $\mathbb{R} \setminus \{0\}$ - and \widehat{f}_T^{j+1} share the same real zeros. The properties of f_T^{j+1} therefore yield that \widehat{f}_T^{j+1} has one positive, one negative and two complex conjugate zeros and $s_h^{j+1}|_T$ is the positive zero of $\widehat{f}_T^{j+1}(x)$. This information enables us to explicitly compute $s_h^{j+1}|_T$ on each element $T \in \mathcal{T}_h$ and in Appendix A we provide a corresponding formula.

Unconditional global convergence, termination and energy decay of the Heron method follows from choosing $v_h = d_t u_h^{j+1}$ and $r_h = d_t s_h^{j+1}$ in the optimality conditions for the iterates so that due to the separate convexity of \widehat{I}_ε we have

$$\begin{aligned} & \|d_t u_h^{j+1}\|^2 + \|d_t s_h^{j+1}\|^2 \\ &= -\delta_u \widehat{I}_\varepsilon(u_h^{j+1}, s_h^j)[d_t u_h^{j+1}] - \delta_s \widehat{I}_\varepsilon(u_h^{j+1}, s_h^{j+1})[d_t s_h^{j+1}] \\ &\leq \frac{1}{\tau} (\widehat{I}_\varepsilon(u_h^j, s_h^j) - \widehat{I}_\varepsilon(u_h^{j+1}, s_h^j)) + \frac{1}{\tau} (\widehat{I}_\varepsilon(u_h^{j+1}, s_h^j) - \widehat{I}_\varepsilon(u_h^{j+1}, s_h^{j+1})). \end{aligned}$$

A summation over $j = 0, 1, \dots, J$ and multiplication by τ imply that

$$\widehat{I}_\varepsilon(u_h^{J+1}, s_h^{J+1}) + \tau \sum_{j=0}^J (\|d_t u_h^{j+1}\|^2 + \|d_t s_h^{j+1}\|^2) \leq \widehat{I}_\varepsilon(u_h^0, s_h^0). \quad (4)$$

Hence, $d_t u_h^{j+1} \rightarrow 0$ and $d_t s_h^{j+1} \rightarrow 0$ so that the algorithm terminates. Moreover, these quantities are the residuals in the equations so that (u_h^j, s_h^j) converge to a discrete minimizer for \widehat{I}_ε . Note that the right-hand side of (4) remains bounded as $\varepsilon, h \rightarrow 0$ if the sequence $(\nabla u_h^0)_{h>0}$ is uniformly bounded in $L^1(\Omega)$ and if, e.g., $s_h^0 = (|\nabla u_h^0|^2 + \varepsilon^2)^{1/4}$.

Remark III.4. *One may also use other regularizations of the BV-seminorm, e.g., the Huber regularization*

$$|\nabla u_h|_\varepsilon = \begin{cases} |\nabla u_h|^2, & \text{if } |\nabla u_h| \leq \varepsilon, \\ 2\varepsilon, & \\ |\nabla u_h| - \frac{\varepsilon}{2}, & \text{if } |\nabla u_h| > \varepsilon. \end{cases}$$

However, the quality of computed approximations should not depend on the employed regularization but only on the degree of regularization determined by ε .

4. AUGMENTATION

By introducing the variable $\sigma_h = \nabla u_h$ and enforcing this identity via a Lagrange multiplier and a stabilizing term, the determination of the minimizer u_h of I is equivalent to computing a saddle point $(u_h, \sigma_h; \lambda_h) \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{L}^0(\mathcal{T}_h)^d \times \mathcal{L}^0(\mathcal{T}_h)^d$ for

$$L_\tau(u_h, \sigma_h; \lambda_h) = \int_\Omega |\sigma_h| \, dx + \frac{\alpha}{2} \|u_h - g\|^2 + (\lambda_h, \sigma_h - \nabla u_h)_h + \frac{\tau}{2} \|\sigma_h - \nabla u_h\|_h^2$$

with a scalar product $(\cdot, \cdot)_h$ on $\mathcal{L}^0(\mathcal{T}_h)^d$ and corresponding norm $\|\cdot\|_h$. We then have

$$\min_{u_h} I(u_h) = \min_{(u_h, \sigma_h)} \max_{\lambda_h} L_\tau(u_h, \sigma_h; \lambda_h).$$

The following algorithm for computing a saddle point for L_τ is unconditionally convergent and performs ascent steps in the variable λ_h .

Algorithm IV.1 (Splitting method). Choose $\tau > 0$ and $(\sigma_h^0, \lambda_h^0)$, and set $j = 0$. Let $\|\cdot\|_h = c_w^{1/2} \|\cdot\|$.

(1) Compute the minimizer u_h^{j+1} of the mapping $u_h \mapsto L_\tau(u_h, \sigma_h^j; \lambda_h^j)$, i.e., compute $u_h^{j+1} \in \mathcal{S}^1(\mathcal{T}_h)$ such that

$$\alpha \int_{\Omega} (u_h^{j+1} - g) v_h \, dx - (\lambda_h^j, \nabla v_h)_h + \tau (\nabla u_h^{j+1} - \sigma_h^j, v_h)_h = 0$$

for all $v_h \in \mathcal{S}^1(\mathcal{T}_h)$.

(2) Compute the minimizer σ_h^{j+1} of the mapping $\sigma_h \mapsto L_\tau(u_h^{j+1}, \sigma_h; \lambda_h^j)$, i.e., $\sigma_h^{j+1} \in \mathcal{L}^0(\mathcal{T}_h)^d$ with

$$-\lambda_h^j + \tau (\nabla u_h^{j+1} - \sigma_h^{j+1}) \in \frac{1}{c_w} \partial |\sigma_h^{j+1}|$$

with solution given by

$$\sigma_h^{j+1} = \frac{1}{\tau} (|\tau \nabla u_h^{j+1} - \lambda_h^j| - c_w^{-1})_+ \frac{\tau \nabla u_h^{j+1} - \lambda_h^j}{|\tau \nabla u_h^{j+1} - \lambda_h^j|}.$$

(3) Compute the maximizer of the mapping $\lambda_h \mapsto L_\tau(u_h^{j+1}, \sigma_h^{j+1}; \lambda_h) - \frac{1}{2\tau} \|\lambda_h - \lambda_h^j\|^2$, i.e., set

$$\lambda_h^{j+1} = \lambda_h^j + \tau (\sigma_h^{j+1} - \nabla u_h^{j+1}),$$

and stop if $\|\lambda_h^{j+1} - \lambda_h^j\|_h + \tau \|\sigma_h^{j+1} - \sigma_h^j\|_h \leq \varepsilon_{stop}$; otherwise increase $j \rightarrow j+1$, and continue with (1).

The scalar product $(\cdot, \cdot)_h$ on $\mathcal{L}^0(\mathcal{T}_h)^d$ has to be carefully chosen in order to have a stable numerical method. The frequently employed choice of the L^2 -scalar product cannot be expected to lead to a uniformly stable method since $(\nabla u_h)_{h>0}$ is not uniformly bounded in $L^2(\Omega)$. Instead the sequence is bounded in $L^1(\Omega)$ and an inverse estimate shows that for the weighted L^2 -norm

$$\|q_h\|_h = h^{d/2} \|q_h\|$$

the sequence $(\nabla u_h)_{h>0}$ remains bounded as $h \rightarrow 0$. This is related to a softer treatment of the constraint $\nabla u_h = \sigma_h$. Omitting the weighting factor $c_w = h^d$ overpenalizes the constraint and results in a locking of the method. This effect is visible in the stability estimate for the splitting method, which guarantees that for every saddle point $(u_h, \sigma_h; \lambda_h)$ with $\sigma_h = \nabla u_h$ we have

$$\begin{aligned} \tau \sum_{j=0}^J \left(\alpha \|u_h - u_h^{j+1}\|^2 + \frac{\tau}{2} \|\nabla u_h^{j+1} - \sigma_h^{j+1}\|_h^2 + \tau^3 \|d_t \sigma_h^{j+1}\|_h^2 \right) \\ \leq \frac{1}{2} \left(\|\lambda_h - \lambda_h^0\|_h^2 + \tau^2 \|\nabla u_h - \sigma_h^0\|_h^2 \right), \end{aligned} \quad (5)$$

cf. [28, 29]. In particular, $\|\lambda_h^{j+1} - \lambda_h^j\|_h + \tau \|\sigma_h^{j+1} - \sigma_h^j\|_h \rightarrow 0$ as $j \rightarrow \infty$ and the splitting method terminates. Moreover, due to the choice of the weighted norm, the right-hand side of (5) remains bounded as $h \rightarrow 0$.

Remark IV.2. *If we have at the j -th iteration $\lambda_h^{j+1} - \lambda_h^j = 0$ and $\tau(\sigma_h^{j+1} - \sigma_h^j) = 0$, then, noting that $\tau(\sigma_h^{j+1} - \nabla u_h^{j+1}) = \lambda_h^{j+1} - \lambda_h^j = 0$, the variational (in-)equalities defining the iterates u_h^{j+1} , σ_h^{j+1} and λ_h^{j+1} in Algorithm IV.1 read*

$$\begin{aligned} 0 &= \alpha(u_h^{j+1} - g, v_h) - (\lambda_h^j, \nabla v_h)_h + \tau(\nabla u_h^{j+1} - \sigma_h^j, v_h)_h \\ &= \alpha(u_h^{j+1} - g, v_h) - (\lambda_h^{j+1}, \nabla v_h)_h + (\tau(\nabla u_h^{j+1} - \sigma_h^{j+1}), v_h)_h \\ &= \alpha(u_h^{j+1} - g, v_h) - (\lambda_h^{j+1}, \nabla v_h)_h \end{aligned}$$

for all $v_h \in \mathcal{S}^1(\mathcal{T}_h)$, as well as

$$\lambda_h^{j+1} = \lambda_h^j + \tau(\sigma_h^{j+1} - \nabla u_h^{j+1}) \in \frac{1}{c_w} \partial |\sigma_h^{j+1}|$$

and

$$0 = \lambda_h^{j+1} - \lambda_h^j = \tau(\sigma_h^{j+1} - \nabla u_h^{j+1}) \iff \nabla u_h^{j+1} = \sigma_h^{j+1}.$$

Since these are exactly the optimality conditions for a saddle-point of L_τ we have that $(u_h^{j+1}, \sigma_h^{j+1}; \lambda_h^{j+1})$ is a saddle-point of L_τ . This means that the pair $(\lambda_h^{j+1} - \lambda_h^j, \tau(\sigma_h^{j+1} - \sigma_h^j))$ can be regarded as the residual of the system that is being solved and that $\|\lambda_h^{j+1} - \lambda_h^j\|_h + \tau\|\sigma_h^{j+1} - \sigma_h^j\|_h$ is an accurate measure of optimality.

Remark IV.3. *Algorithm IV.1 belongs to the class of alternating direction method of multipliers (ADMM) and is a special case of the Douglas-Rachford splitting method, cf., e.g., [9, 24, 37].*

In the sequel we refer to Algorithm IV.1 as the splitting method or the augmented Lagrangian method.

5. SADDLE-POINT FORMULATION

Due to the strong convexity, the minimizer u_h for I can be determined by computing a saddle point $(u_h, p_h) \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{L}^0(\mathcal{T}_h)^d$ for the functional

$$S(u_h, p_h) = \int_{\Omega} \nabla u_h \cdot p_h \, dx - I_{K_1(0)}(p_h) + \frac{\alpha}{2} \|u_h - g\|^2.$$

The following primal-dual method chooses inner products on $\mathcal{S}^1(\mathcal{T}_h)$ and $\mathcal{L}^0(\mathcal{T}_h)$ and alternately performs descent and ascent steps for the variables u_h and p_h .

Algorithm V.1 (Primal-dual method). *Choose $\tau > 0$ and $(u_h^0, p_h^0) \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{L}^0(\mathcal{T}_h)^d$, and set $j = 0$ and $d_t u_h^0 = 0$.*

- (1) *Set $\tilde{u}_h^{j+1} = u_h^j + \tau d_t u_h^j$.*
- (2) *Compute $p_h^{j+1} \in \mathcal{L}^0(\mathcal{T}_h)^d$ maximal for $p_h \mapsto S(\tilde{u}_h^{j+1}, p_h) - \frac{1}{2\tau} \|p_h - p_h^j\|^2$, i.e., set*

$$p_h^{j+1} = \frac{p_h^j + \tau \nabla \tilde{u}_h^{j+1}}{\max\{1, |p_h^j + \tau \nabla \tilde{u}_h^{j+1}|\}}.$$

(3) Compute $u_h^{j+1} \in \mathcal{S}^1(\mathcal{T}_h)$ minimal for $u_h \mapsto S(u_h, p_h^{j+1}) + \frac{1}{2\tau} \|u_h - u_h^j\|_h^2$, i.e.,

$$(d_t u_h^{j+1}, v_h)_h = - \int_{\Omega} \nabla v_h \cdot p_h^{j+1} \, dx - \alpha \int_{\Omega} (u_h^{j+1} - g) v_h \, dx$$

for all $v_h \in \mathcal{S}^1(\mathcal{T}_h)$.

(4) Stop if $\|d_t u_h^{j+1}\|_h + \|d_t p_h^{j+1}\| \leq \varepsilon_{stop}$; increase $j \rightarrow j + 1$ and continue with (1) otherwise.

The inner product $(\cdot, \cdot)_h$ should be chosen such that the iteration is stable under moderate conditions on the step size. On the other hand, it has to be simple enough to allow for an efficient solution of the equations and the discrete saddle points should be uniformly bounded in these norms as $h \rightarrow 0$. Choosing the L^2 -norm, the iteration is well-defined and convergent under the condition $\tau \leq c'h$. Observing that a discrete interpolation estimate controls a discrete version of the $H^{1/2}$ -norm by the $W^{1,1}$ -seminorm and the L^∞ -norm, one may use the inner product

$$(v_h, w_h)_h = (v_h, w_h) + h(\nabla v_h, \nabla w_h) \quad (6)$$

for which the primal-dual method is convergent under the less restrictive condition $\tau \leq ch^{1/2}$. In particular, we have the stability estimate

$$\begin{aligned} C \frac{\tau^2}{2} \left(\sum_{j=1}^J \|d_t u_h^j\|_h^2 + \|d_t p_h^j\|^2 \right) + \alpha \tau \sum_{j=1}^J \|u_h - u_h^j\|^2 \\ \leq \frac{1}{2} \|u_h - u_h^0\|_h^2 + \frac{1}{2} \|p_h - p_h^0\|^2, \end{aligned} \quad (7)$$

where (u_h, p_h) is a saddle point for S , cf. [3, 5]. The discrete interpolation estimate shows that the right-hand side of (7) remains bounded as $h \rightarrow 0$, cf. [3].

Remark V.2 (Choice of c). In [3] the convergence of Algorithm V.1 is proven. Using the inner product (6), we observe that for any $v_h \in \mathcal{S}^1(\mathcal{T}_h)$ we have

$$\|\nabla v_h\|^2 \leq \frac{1}{h} (\|v_h\|^2 + h \|\nabla v_h\|^2) = \frac{1}{h} \|v_h\|_h^2,$$

so this inverse estimate holds with constant 1 and factor $h^{-1/2}$ when defining the inner product with h being the maximal mesh size of the underlying triangulation. However, the discrete interpolation estimate stated in [3] reads

$$h_{min} \|\nabla v_h\|^2 \leq \tilde{c} \|\nabla v_h\|_{L^1(\Omega)} \|v_h\|_{L^\infty(\Omega)}$$

with \tilde{c} depending on the minimal angle occurring in the triangulation and h_{min} the minimal diameter appearing in the triangulation. Therefore, the right-hand side in (7) remains bounded as $h \rightarrow 0$ if the ratio h/h_{min} is uniformly bounded from above, i.e., if \mathcal{T}_h is quasi-uniform. When making use of Young's inequality

$$\tau \gamma^{1/2} \|\nabla d_t u_h^{k-1}\| \gamma^{-1/2} \|d_t p_h^k\| \leq \tau^2 \frac{\gamma}{2} \|\nabla d_t u_h^{k-1}\|^2 + \frac{1}{2\gamma} \|d_t p_h^k\|^2$$

for arbitrary $1 < \gamma < 2$ and of the inverse estimate mentioned above in the convergence proof of Algorithm V.1 we obtain convergence of the algorithm for any $\tau < (h/\gamma)^{1/2}$, and hence for any $\tau < h^{1/2}$. In our experiments presented below the choice $\tau = h^{1/2}$ also guaranteed convergence.

Remarks V.3. (1) The primal-dual method belongs to the class of proximal-point algorithms, cf., e.g., [14, 33, 38].

(2) In [14] the convergence rate $\mathcal{O}(1/J)$ for the primal-dual algorithm has been shown. The same convergence rate holds for the $H^{1/2}$ -primal-dual method, however, the h -dependent constant entering the convergence rate is smaller for the $H^{1/2}$ -primal-dual method due to the less restrictive condition on the step size.

In what follows we refer to the primal-dual method with scalar product $(v_h, w_h)_h = (v_h, w_h) + h(\nabla v_h, \nabla w_h)$ as the $H^{1/2}$ -primal-dual method whereas the primal-dual method with scalar product $(v_h, w_h)_h = (v_h, w_h)$ will be simply referred to as the *primal-dual method*.

6. COMPARISON

VI.A. Setup. We tested the performance of Algorithms III.1, IV.1 and V.1 ($H^{1/2}$ -primal-dual method) for two different choices of data g and α . We assume that all occurring linear systems of equations can be solved with linear complexity. In our realization all systems are solved using MATLAB's backslash operator. We therefore restrict to comparing iteration numbers. In each experiment, we investigated the influence of the following aspects:

- **Meshes:** The computations were done using uniform triangulations \mathcal{T}_h consisting of halved squares with mesh sizes $h = \sqrt{2}2^{-\ell}$ and refinement levels $\ell \in \{5, \dots, 9\}$. We denote by \mathcal{T}_ℓ the triangulation \mathcal{T}_h with mesh size $h = \sqrt{2}2^{-\ell}$.
- **Initialization:** We started the algorithms with two different initializations of u_h^0 , namely $u_h^0 = 0$ and $u_h^0 = \mathcal{I}_h g$. In both experiments, we used $p_h^0 = 0$ for the $H^{1/2}$ -primal-dual method, $s_h^0 = \varepsilon^{1/2}$ for the Heron method and $\sigma_h^0 = 0$ and $\lambda_h^0 = 0$ for the splitting method. The regularization parameter is chosen to be $\varepsilon = h$ in order to allow for the optimal convergence rate for the primal variable in the L^2 -norm.
- **Step sizes:** We ran the methods using different step sizes τ for the splitting algorithm and the Heron method. For the $H^{1/2}$ -primal-dual method the step size has been chosen as $\tau = \sqrt{h}$ and as $\tau = \sqrt{3h/8}$.
- **Common stopping criterion:** In each experiment we used two different stopping criteria to stop the iteration. To determine how many iteration steps in the algorithms are needed to achieve a given quality, we compared the difference of the iterates to a precomputed reference solution \tilde{u}_h obtained with the primal-dual method after

10^4 iterations. Recalling the inequality (2) we may choose

$$\|\tilde{u}_h - u_h^n\| \leq \left(\frac{2\varepsilon|\Omega|}{\alpha} \right)^{1/2} \quad (8)$$

as a stopping criterion. Here, u_h^n denotes the n -th iterate of any of the algorithms. This stopping criterion enables us to do a fair comparison between all three algorithms since, for fixed h , the primal-dual method and the splitting algorithm approximate $u_h \approx \tilde{u}_h$ whereas the Heron method approximates $u_{\varepsilon,h}$, i.e., the L^2 -distance between $u_h \approx \tilde{u}_h$ and the iterates generated by the Heron method is not guaranteed to strictly fall below the error tolerance in (8), cf. (2). Note that the approximations corresponding to the stopping criterion (8) with $\varepsilon = h$ which are generated by each of the algorithms will qualitatively be of the same accuracy as u_h , cf. (1) and (2).

- **General stopping criteria:** The following stopping criteria are derived by the optimality conditions for the iterates in each algorithm. Our motivation was to control meaningful residual quantities that are guaranteed to converge in the iterations.

$H^{1/2}$ -primal-dual method: We note that if $(u_h^{j+1}, p_h^{j+1}) \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{L}^0(\mathcal{T}_h)^d$ is such that $d_t u_h^{j+1} = 0$ and $d_t p_h^{j+1} = 0$, we have that (u_h^{j+1}, p_h^{j+1}) is a saddle-point of S . This gives rise to the stopping criterion

$$\left(\frac{\|d_t u_h^{j+1}\|_h^2 + \|d_t p_h^{j+1}\|_h^2}{I(\hat{u})} \right)^{1/2} \leq \frac{h^{1/2}}{50}, \quad (9)$$

where $I(\hat{u}) = \frac{\alpha}{2}\|g\|^2$ with $\hat{u} \equiv 0$ is an upper bound for the energy $I(u_h)$ and serves as a normalization factor.

Heron method: We use the stopping criterion

$$\left(\frac{\|d_t u_h^{j+1}\|_h^2 + \|d_t s_h^{j+1}\|_h^2}{I(\hat{u})} \right)^{1/2} \leq \min\left(1, \frac{1}{\tau}\right) \frac{h^{1/2}}{50}. \quad (10)$$

Here, we use the factor $\min(1, \frac{1}{\tau})$ on the right-hand side because otherwise the stopping criterion would be artificially fulfilled after very few steps if τ was being increased significantly while the output would still not be close to the minimizer (u_h, s_h) .

Splitting algorithm: The observations made in Remark IV.2 justify to choose the stopping criterion as

$$\begin{aligned} & \left(\frac{\|\lambda_h^{j+1} - \lambda_h^j\|_h^2 + \tau^2 \|\sigma_h^{j+1} - \sigma_h^j\|_h^2}{I(\hat{u})} \right)^{1/2} \\ & \leq c_w^{-1/2} \min(\tau c_w, 1) \frac{h^{1/2}}{50}. \quad (11) \end{aligned}$$

Here, the additional factors $c_w^{-1/2}$ and $\min(\tau c_w, 1)$ on the right-hand side are necessary in order to obtain meaningful outputs in the sense that the gradient of u_h is sufficiently well approximated and in order to generate similar outputs independently of the choice of the step size τ .

VI.B. Noisy image. Our first comparison uses a characteristic function of a disk that is randomly perturbed at the pixel scale of a given triangulation. We compare different stopping criteria and the influence of rough initial data and different choices of step sizes.

Example VI.1. We let $d = 2$, $\Omega = (0, 1)^2$ and $\alpha = 20$. Let \mathcal{T}_0 be the triangulation of Ω into two triangles with diagonal parallel to $(-1, 1)^T$. Consider the triangulation \mathcal{T}_5 of Ω generated from five red refinements of \mathcal{T}_0 . Denoting by $x_\Omega = (1/2, 1/2)^T$ the center of Ω , we define the function g as the continuous, piecewise linear function $g = \mathcal{I}_h(\chi_{B_{1/5}(x_\Omega)} + \xi_h)$ with $\xi_h \in \mathcal{S}^1(\mathcal{T}_5)$ whose coefficient vector is a sample of an in the interval $[-1/10, 1/10]$ uniformly distributed random variable.

Table 1 and Table 2 display the number of iterations needed until the stopping criteria (9), (10), (11) and the stopping criterion (8), respectively, were satisfied for Example VI.1. A dash (–) means that we did not run the algorithm for the corresponding parameters because on the previous refinement level the number of iterations has already exceeded 10^3 .

We observe that the Heron method reaches the prescribed accuracy for both choices of initial data u_h^0 and all refinement levels after a smaller number of iterations than the $H^{1/2}$ -primal-dual method and the augmented Lagrangian (splitting) method, yet, one has to be aware of the fact that, for fixed h , the Heron method approximates the discrete solution of the regularized functional I_ε while the $H^{1/2}$ -primal-dual method and the augmented Lagrangian method approximate the minimizer $u_h \in \mathcal{S}^1(\mathcal{T}_\ell)$ of the functional I . However, the minimization of I over $\mathcal{S}^1(\mathcal{T}_\ell)$ is already qualitatively a regularization since it is $W^{1,1}$ -conforming due to the continuity property of functions in $\mathcal{S}^1(\mathcal{T}_\ell)$.

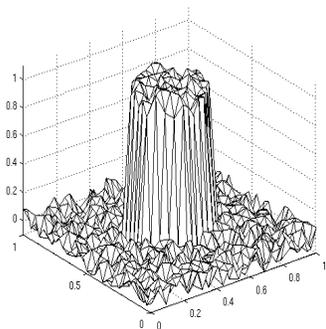
The choice $\tau = h^{1/2}$ for the $H^{1/2}$ -primal-dual method, although not justified by the stability estimate (7) but for every $\tau < h^{1/2}$, and $\tau = h^{-3/2}$ for the splitting method yield the smallest iteration numbers and seem to have a similar or even the same behavior. While the iteration of the splitting method is independent of u_h^0 , cf. Algorithm IV.1 and (5), the initialization of u_h^0 enters the right-hand side of the stability estimate (7) of the $H^{1/2}$ -primal-dual method. In [3], the author elaborated that the right-hand side of (7) is bounded h -independently if $u_h - u_h^0$ is uniformly bounded in $L^\infty(\Omega) \cap W^{1,1}(\Omega)$. This does not need to be the case for the initialization $u_h^0 = \mathcal{I}_h g$ since we only require $g \in L^\infty(\Omega)$ and the numerical experiments in [3] show that the iteration numbers may be larger for $u_h^0 = \mathcal{I}_h g$ than for $u_h^0 = 0$. The

		$H^{1/2}$ -primal-dual				Heron						Augmented Lagrangian ($c_w = h^d$)				
τ	$h^{1/2}$	$(\frac{3}{8}h)^{1/2}$	h	1	h^{-1}	h^{-2}	1	$h^{-1/2}$	h^{-1}	$h^{-3/2}$	h^{-2}	$h^{-5/2}$				
u_h^0	0	$T_{h,g}$	0	$T_{h,g}$	0	$T_{h,g}$	0	0	0	0	0	0				
h	65	42	107	68	345	343	35	35	32	32	32	32				
$\sqrt{2}/25$	105	84	173	137	959	957	47	47	44	44	44	44				
$\sqrt{2}/26$	170	168	276	274	3074	3071	69	69	64	64	64	64				
$\sqrt{2}/27$	272	273	442	444	—	—	97	97	90	90	90	90				
$\sqrt{2}/28$	354	349	576	569	—	—	135	135	126	126	126	126				
$\sqrt{2}/29$																

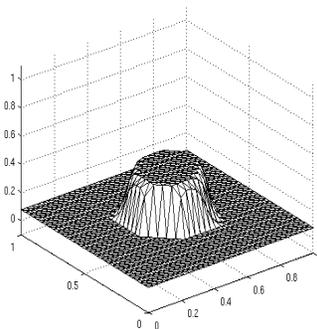
TABLE 1. Stopping by residual control, cf. (9), (10), (11), for Example VI.1.

		$H^{1/2}$ -primal-dual		Heron						Augmented Lagrangian ($c_w = h^d$)				
τ	$h^{1/2}$	$(\frac{3}{8}h)^{1/2}$	h	1	h^{-1}	h^{-2}	1	$h^{-1/2}$	h^{-1}	$h^{-3/2}$	h^{-2}	$h^{-5/2}$		
u_h^0	0	$T_{h,g}$	0	$T_{h,g}$	0	$T_{h,g}$	0	0	0	0	0	0		
h	2	7	1	11	44	42	11	11	10	10	10	10		
$\sqrt{2}/25$	4	7	7	11	110	105	16	16	13	13	13	13		
$\sqrt{2}/26$	7	8	12	13	255	242	19	19	16	16	16	16		
$\sqrt{2}/27$	11	12	17	19	585	554	23	23	19	19	19	19		
$\sqrt{2}/28$	16	16	25	25	1352	1275	27	27	22	22	22	22		
$\sqrt{2}/29$														

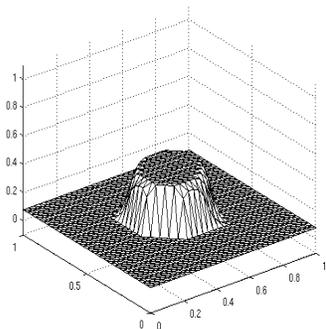
TABLE 2. Stopping by comparison with reference solution, cf. (8), for Example VI.1.



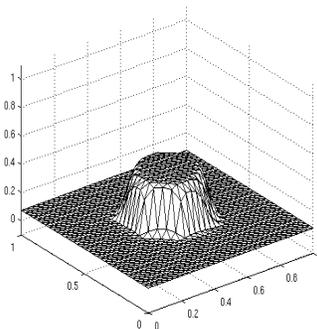
(A) Noisy image $\mathcal{I}_h g$



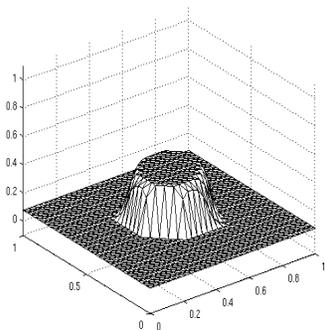
(B) Reference solution \tilde{u}_h



(C) $H^{1/2} - PD$



(D) Heron



(E) Splitting

FIGURE 1. Noisy image $\mathcal{I}_h g$, reference solution \tilde{u}_h and outputs of the three algorithms with residual control, cf. (9), (10), (11), $h = \sqrt{22}^{-5}$ and $u_h^0 = 0$ for Example VI.1.

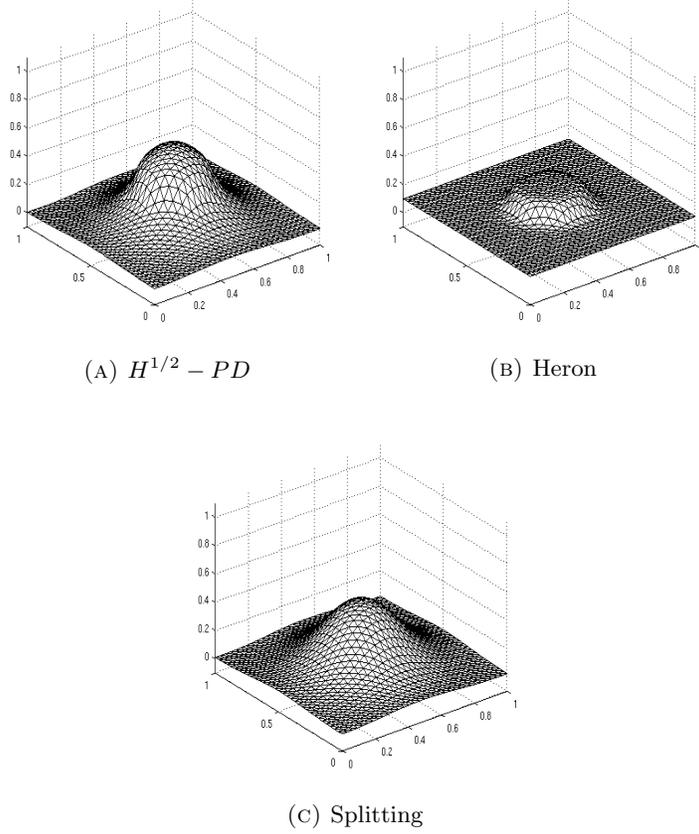


FIGURE 2. Outputs of the three algorithms with comparison to reference solution, cf. (8), $h = \sqrt{22}^{-5}$ and $u_h^0 = 0$ for Example VI.1.

experiments in [3] also indicate that this effect is even more pronounced when the metrics on which the gradient flow is based are too strong, e.g., for the choice $(v_h, w_h)_h = (v_h, w_h) + (\nabla v_h, \nabla w_h)$. This seems to contradict the results in Tables 1 and 2 at the first glance since the iteration numbers for $u_h^0 = \mathcal{I}_h g$ are at most as large as the iteration numbers for $u_h^0 = 0$. However, this can be explained by the larger step size for the $H^{1/2}$ -primal-dual method used in our experiments which are 7 to 10 times larger than the step size $\tau = h^{1/2}/10$ considered for the $H^{1/2}$ -primal-dual method in [3]. Since the step size influences the distance between two consecutive iterates u_h^{n+1} and u_h^n with respect to $\|\cdot\|_h$, a larger step size enables the method to faster diverge from u_h^0 . Note that $\sqrt{3h}/8$ for $h = \sqrt{22}^{-9}$ is still larger than $\sqrt{h}/10$ for $h = \sqrt{22}^{-4}$. The closeness of u_h and g in the L^2 -sense is in turn determined by the parameter α .

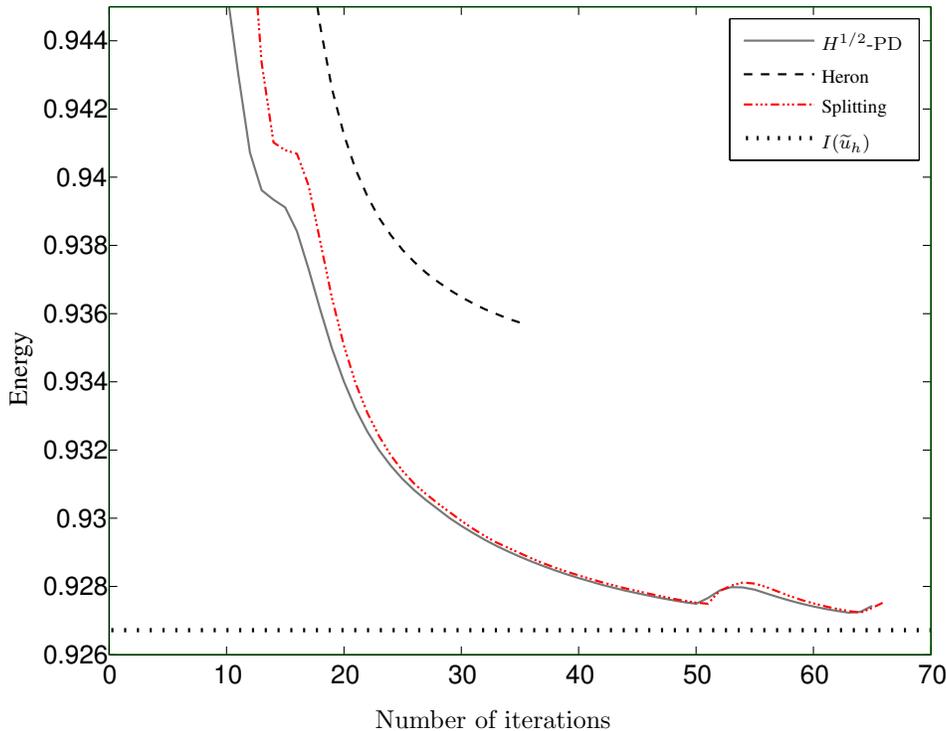


FIGURE 3. Energy I of iterates u_h^n for the three algorithms with residual control, cf. (9), (10), (11), $h = \sqrt{2}2^{-5}$, $u_h^0 = 0$, and energy of the reference solution \tilde{u}_h for Example VI.1.

Let us finally observe that the Heron method appears to be robust regarding the choice of the initialization of u_h^0 . Yet, it is not ensured that the right-hand side of the stability estimate (4) for the Heron method is bounded h -independently when $u_h^0 = \mathcal{I}_h g$ is chosen, so that the choice $u_h^0 = 0$ seems to be reasonable both theoretically and practically.

Figure 1 shows the plot of the noisy image $\mathcal{I}_h g$, the reference solution \tilde{u}_h , the outputs of the $H^{1/2}$ -primal-dual method with $\tau = \sqrt{h}$, the Heron method with $\tau = 1$ and the splitting method with $\tau = h^{-3/2}$ for Example VI.1 and for the residual-based stopping criteria (9), (10) and (11) on \mathcal{T}_5 . We observe that the outputs of all three algorithms are nearly indistinguishable from the reference solution. This is also underlined by the energy plot in Figure 3 which particularly indicates that the residual-based stopping criteria (9), (10) and (11) define an accurate measure of optimality. Figure 2 shows the outputs of the three iterative methods using the L^2 -stopping criterion (8) on \mathcal{T}_5 . Note that the outputs visually considerably differ from the reference solution which is also recognizable in the corresponding energy plot in Figure 4. This is, however, due to the fact that the approximations were computed on a coarse grid. For smaller $\varepsilon = h$ the outputs corresponding to

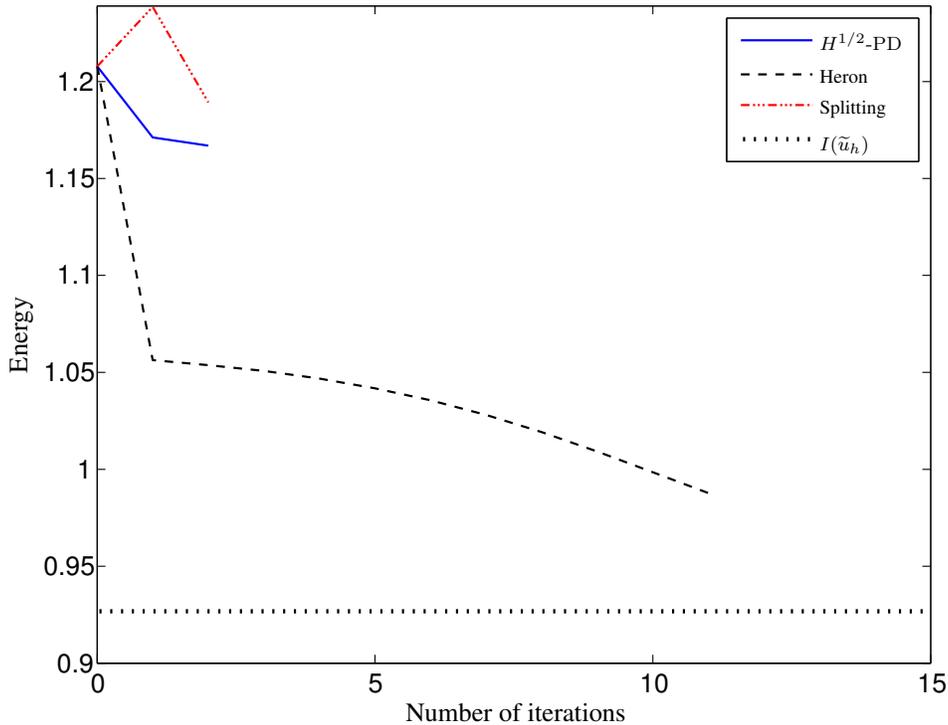


FIGURE 4. Energy of iterates u_h^n for the three algorithms with L^2 -based stopping criterion (8), $h = \sqrt{2}2^{-5}$, $u_h^0 = 0$, and energy of the reference solution \tilde{u}_h for Example VI.1.

the L^2 -stopping criterion (8) do not significantly differ from the reference solution. It is noteworthy that the energy of the final iterate of the Heron method using the L^2 -stopping criterion is much closer to the energy of the reference solution than the final iterates of the other two algorithms.

VI.C. Real image. We finally report numerical results for the benchmark image "cameraman".

Example VI.2. We let $d = 2$, $\Omega = (0, 1)^2$ and $\alpha = 500$. Let \mathcal{T}_0 be as before and \mathcal{T}_6 generated from six red refinements of \mathcal{T}_0 . Given a low resolution of the image "cameraman" consisting of 4225 pixels we define the function $\tilde{g}_h \in \mathcal{S}^1(\mathcal{T}_6)$ by using the (scaled) gray-values ranging from 0 to 1 as coefficients of the nodal basis functions φ_z of the corresponding nodes z of the triangulation \mathcal{T}_6 . Then g is defined as $g = \tilde{g}_h + \xi_h$ with $\xi_h \in \mathcal{S}^1(\mathcal{T}_6)$ as is Example VI.1.

The experiments for Example VI.2 show similar results as before in Example VI.1. Again, the Heron method outperforms the other two methods in terms of required iterations for both sets of stopping criteria, which can be seen in Tables 3 and 4. Moreover, the Heron method is robust with respect to the choice of u_h^0 . It is worth mentioning that the iteration numbers

τ	$H^{1/2}$ -primal-dual		Heron				Augmented Lagrangian ($c_w = h^d$)								
	$h^{1/2}$	$(\frac{3}{8}h)^{1/2}$	h	1	h^{-1}	h^{-2}	1	$h^{-1/2}$	h^{-1}	$h^{-3/2}$	h^{-2}	$h^{-5/2}$			
u_h^0	0	\mathcal{I}_{hg}	0	\mathcal{I}_{hg}	0	\mathcal{I}_{hg}	0	0	0	0	0	0			
h															
$\sqrt{2}/2^6$	47	49	77	79	208	208	14	13	13	13	2055	307	48	49	316
$\sqrt{2}/2^7$	78	80	127	128	601	601	18	17	17	17	—	731	79	71	659
$\sqrt{2}/2^8$	132	132	215	216	1791	1791	23	22	22	22	—	1750	132	100	1324
$\sqrt{2}/2^9$	213	213	348	348	—	—	30	30	28	28	—	—	213	137	—

TABLE 3. Stopping by residual control, cf. (9), (10), (11), for Example VI.2.

τ	$H^{1/2}$ -primal-dual		Heron				Augmented Lagrangian ($c_w = h^d$)								
	$h^{1/2}$	$(\frac{3}{8}h)^{1/2}$	h	1	h^{-1}	h^{-2}	1	$h^{-1/2}$	h^{-1}	$h^{-3/2}$	h^{-2}	$h^{-5/2}$			
u_h^0	0	\mathcal{I}_{hg}	0	\mathcal{I}_{hg}	0	\mathcal{I}_{hg}	0	0	0	0	0	0			
h															
$\sqrt{2}/2^6$	6	9	10	15	54	54	9	8	8	8	1120	167	25	23	149
$\sqrt{2}/2^7$	12	12	18	20	138	138	12	10	10	10	—	863	91	38	349
$\sqrt{2}/2^8$	24	24	38	38	358	358	16	13	13	13	—	3936	293	68	903
$\sqrt{2}/2^9$	47	47	76	76	960	960	20	16	16	16	—	—	870	47	104

TABLE 4. Stopping by comparison with reference solution, cf. (8), for Example VI.2.

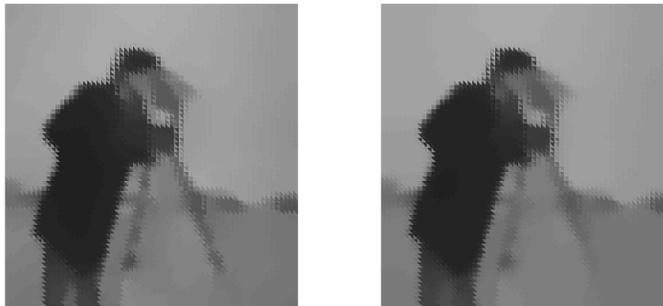
(A) Noisy image $\mathcal{I}_h g$ (B) Reference solution \tilde{u}_h (C) $H^{1/2} - PD$ 

(D) Heron



(E) Splitting

FIGURE 5. Noisy image $\mathcal{I}_h g$, reference solution \tilde{u}_h and outputs of the three algorithms with residual control, cf. (9), (10), (11), $h = \sqrt{2}2^{-6}$ and $u_h^0 = 0$ for Example VI.2.



(A) $H^{1/2} - PD$

(B) Heron



(c) Splitting

FIGURE 6. Outputs of the three algorithms with comparison to reference solution, cf. (8), $h = \sqrt{2}2^{-6}$ and $u_h^0 = 0$ for Example VI.2.

for the $H^{1/2}$ -primal-dual method seem not to depend on the initialization of u_h^0 . Our explanation for this observation is that the large value of α , namely $\alpha = 500$, forces the first iterate u_h^1 to be immediately very close to g and it therefore does not make a difference if we start the algorithm with $u_h^0 = 0$ or $u_h^0 = \mathcal{I}_h g$. In general, of course, the interplay between the step size τ and the parameter α and the L^2 -distance between u_h and u_h^0 play a crucial role regarding the performance of the ($H^{1/2}$ -) primal-dual method. Let us also note that the difference between the iteration numbers in Table 3 and Table 4 is not as pronounced as in Example VI.1 which is on the one hand due to the stricter error tolerance in (8) because of the larger α and on the other hand because the minimization of I is dominated by the fidelity term $\alpha \|u - g\|^2$ due to the large α .

The energy plots in Figures 7 and 8 once more provide an experimental justification of our choice of stopping criteria (9), (10) and (11).

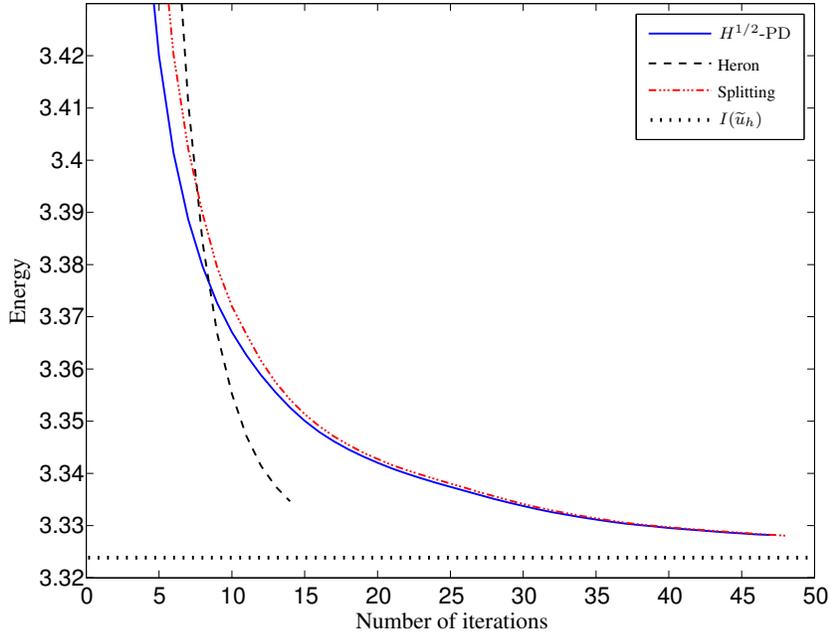


FIGURE 7. Energy I of iterates u_h^n for the three algorithms with residual control, cf. (9), (10), (11), $h = \sqrt{2}2^{-6}$, $u_h^0 = 0$, and energy of the reference solution \tilde{u}_h for Example VI.2.

7. DISCUSSION OF OTHER METHODS

VII.A. Semi-implicit L^2 -gradient flow. A frequently employed method to minimize the regularized ROF functional is the semi-implicit discretization of the L^2 -gradient flow, i.e.,

$$(d_t u_h^{j+1}, v_h) + \int_{\Omega} \frac{\nabla u_h^{j+1} \cdot \nabla v_h}{(|\nabla u_h^j|^2 + \varepsilon^2)^{1/2}} dx + \alpha(u_h^{j+1} - g, v_h) = 0,$$

cf., e.g., [27, 26]. Our experiments indicate a similar behavior of the semi-implicit L^2 -gradient flow compared to the Heron method and indicate stability and convergence independently of τ . Since the authors are unaware of a corresponding stability estimate we did not include this method in our comparison.

VII.B. Dual Method. Following [12], after introducing discrete gradient and discrete divergence operators ∇_h and div_h , respectively, the minimization of a finite difference discretization I_h of I can be transformed into a nonlinear projection problem in such a way that the discrete minimizer u_h of I_h can be written as $u_h = g - \frac{1}{\alpha} \text{div}_h p_h$ where p_h is a solution to the

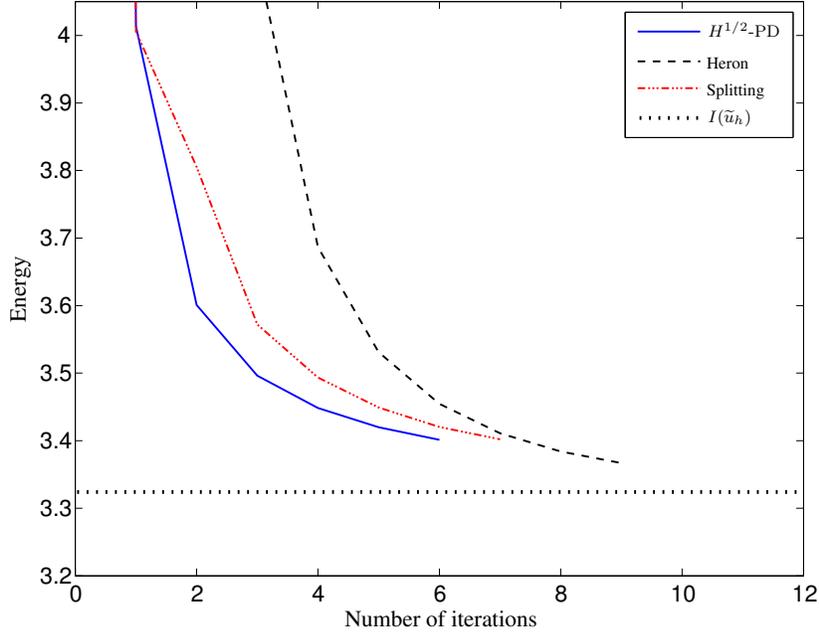


FIGURE 8. Energy of iterates u_h^n for the three algorithms with L^2 -based stopping criterion (8), $h = \sqrt{2}2^{-6}$, $u_h^0 = 0$, and energy of the reference solution \tilde{u}_h for Example VI.2.

discrete constrained minimization problem

$$\min_{p_h} \|\operatorname{div}_h p_h - \alpha g\|^2 \quad \text{s.t. } |p_h| \leq 1.$$

With a Lagrange multiplier λ_h the optimality condition reads

$$-\nabla_h(\operatorname{div}_h p_h - \alpha g) + \lambda_h p_h = 0.$$

The author observed that the relation $\lambda_h = |(\nabla_h(\operatorname{div}_h p_h - \alpha g))|$ holds point-wise and suggested the semi-implicit gradient descent algorithm

$$p_h^{n+1} = p_h^n + \tau((\nabla_h(\operatorname{div}_h p_h^n - \alpha g) - |(\nabla_h(\operatorname{div}_h p_h^n - \alpha g))|)p_h^{n+1})$$

with $\tau > 0$ and $p_h^0 = 0$. Convergence of the algorithm has been shown for $\tau \leq 1/8$. However, this is related to the definition of the differential operators via forward differences. A rescaling yields the condition $\tau = \mathcal{O}(h^2)$ on the step size with h being the mesh size. This restrictive condition on the step size causes a slow convergence behavior.

VII.C. Semi-smooth Newton method. In [35], the authors start from the TV - L^2 functional, where the anisotropic BV -seminorm is defined by

$$\int_{\Omega} |Du| = \sup \left\{ - \int_{\Omega} u \operatorname{div} q \, dx : q \in C_c^1(\Omega; \mathbb{R}^d), |q|_{\ell^\infty} \leq 1 \text{ a.e.} \right\},$$

and consider the predual problem given by

$$\inf_{p \in H_N(\operatorname{div}; \Omega)} \frac{1}{2} \|\operatorname{div} p + \alpha g\|^2 \text{ s.t. } -1 \leq p_i \leq 1 \text{ for all } 1 \leq i \leq d.$$

Since the predual problem may not have a unique solution a semi-smooth Newton method is suggested for a discretization of the regularized predual problem

$$\inf_{p \in H_N(\operatorname{div}; \Omega)} \frac{1}{2} \|\operatorname{div} p + \alpha g\|^2 + \frac{\gamma}{2} \|\Pi_{\operatorname{div}, 0} p\|^2 \text{ s.t. } -1 \leq p_i \leq 1, \quad (12)$$

with $\gamma > 0$ and $\Pi_{\operatorname{div}, 0} : L^2(\Omega) \rightarrow H_{N,0}(\operatorname{div}; \Omega)$ being the projection in $L^2(\Omega)$ onto the space of divergence-free vector fields in $H_N(\operatorname{div}; \Omega)$. Furthermore, in an infinite-dimensional setting, a semi-smooth Newton method is considered for the regularized problem

$$\begin{aligned} \inf_{p \in H_0^1(\Omega; \mathbb{R}^d)} \frac{1}{2c} \|\nabla p\|^2 + \frac{1}{2} \|\operatorname{div} p + \alpha g\|^2 + \frac{\gamma}{2} \|\Pi_{\operatorname{div}, 0} p\|^2 \\ + \frac{1}{2c} \|\max(0, c(p-1))\|^2 + \frac{1}{2c} \|\min(0, c(p+1))\|^2 \end{aligned} \quad (13)$$

with $c > 0$. In the formal limit $c \rightarrow \infty$, we arrive at (12), cf. [35]. It is proven that both semi-smooth Newton methods are locally superlinearly convergent. However, the dependence of the performance of the semi-smooth Newton method for (12) on the parameter γ and on the dimension of the problem is not clear since the convergence analysis for the semi-smooth Newton iteration for (13) is based on the additional regularizing terms.

VII.D. Accelerated primal-dual method. In [14] the authors discussed, apart from the primal-dual method, an acceleration technique for their primal-dual method by considering variable step sizes for both the primal and the dual variable and a variable extrapolation parameter. With given $\tau_0 > 0$, $\sigma_0 = ch^2/\tau$ we let

$$d_t u_h^{n+1} = \frac{u_h^{n+1} - u_h^n}{\tau^n}, \quad d_t p_h^{n+1} = \frac{p_h^{n+1} - p_h^n}{\sigma^n}$$

and $\tilde{u}_h^{n+1} = u_h^n + \theta_n(u_h^{n+1} - u_h^n)$. The step sizes and the extrapolation parameter θ_n are updated in each iteration according to $\theta_{n+1} = 1/\sqrt{1+2\gamma\tau_n}$, $\tau_{n+1} = \theta_{n+1}\tau_n$ and $\sigma_{n+1} = \sigma_n/\theta_{n+1}$, where γ can be chosen as, e.g., $\gamma = \alpha$, and stems from the strong convexity of the L^2 -data fidelity term. Yet, experiments show that the success of the acceleration procedure strongly depends on the initial step size τ_0 and on the given data g and α as well as on the choice of γ . This behavior is also theoretically predictable when taking the

error bound for $\|u_h - u_h^n\|^2$ in [14, Theorem 2] into account. While one can make the first summand in the error bound arbitrarily small by choosing τ_0 sufficiently large, one may influence the size of the second summand only by choosing p_h^0 close enough to p_h . In cases with a relatively small α , such as in Example VI.1, the acceleration procedure did not significantly decrease the number of iterations when compared to the method without adaptive step sizes, and, surprisingly, the choice of a small τ_0 led to less iterations than for a large τ_0 . Still, considering the stopping criterion (8), for examples where α was large, the acceleration scheme needed significantly fewer iterations (for $\tau_0 = 10^6$) than the primal-dual method without acceleration but required still more iterations than the $H^{1/2}$ -primal-dual method. A further disadvantage is that it is not clear which stopping criterion can be considered if we do not have a good guess of the solution u_h because it is not obvious at the first glance if $\|d_t u_h^n\|^2 + \|d_t p_h^n\|^2$ with $d_t u_h^n$ and $d_t p_h^n$ defined as above is a good measure of optimality.

VII.E. Graph Cut Algorithm. In [22], after discretizing I , the authors decompose the variable u into its level sets and reformulate the minimization problem into independent binary Markov random fields associated to each level set. In order for this approach to work the authors make the crucial assumption that the image u takes on values only in a discrete set L , where L is, for instance, having a grayscale image, the set $L = \{0, \dots, 255\}$. This, however, is an assumption that does not serve our purposes and we therefore did not include this approach in our comparison.

VII.F. Newton method for regularized saddle-point functional. In [17] the authors proposed a Newton iteration for the primal-dual system

$$\begin{aligned} |\nabla u|_\varepsilon p - \nabla u &= 0, \\ -\operatorname{div} p + \alpha(u - g) &= 0, \end{aligned}$$

below referred to as *CGM method*, where $|\nabla u|_\varepsilon = \sqrt{|\nabla u|^2 + \varepsilon^2}$. They suggested the following step size for the dual variable p in order to ensure solvability of the linear system of equations: if δp_h^{n+1} is the solution for the dual variable of the $(n+1)$ -th Newton iteration and if $|p_h^n| < 1$ holds for all $x \in \Omega$, then we set $p_h^{n+1} = p_h^n + \gamma \delta p_h^{n+1}$ with

$$\gamma = \rho \max\{s : |p_h^n + s \delta p_h^{n+1}| < 1 \forall x \in \Omega\}, \quad 0 < \rho < 1.$$

In our experiments the CGM method approached the solution very fast within the first few iterations but then developed oscillations, for instance, considering Example VI.1, it developed oscillations along the jump set and did not converge, cf. Figures 9 and 10. Due to missing statements concerning the relation of convergence of the method and the critical parameters ε and h , we decided not to compare the CGM method with the other algorithms.

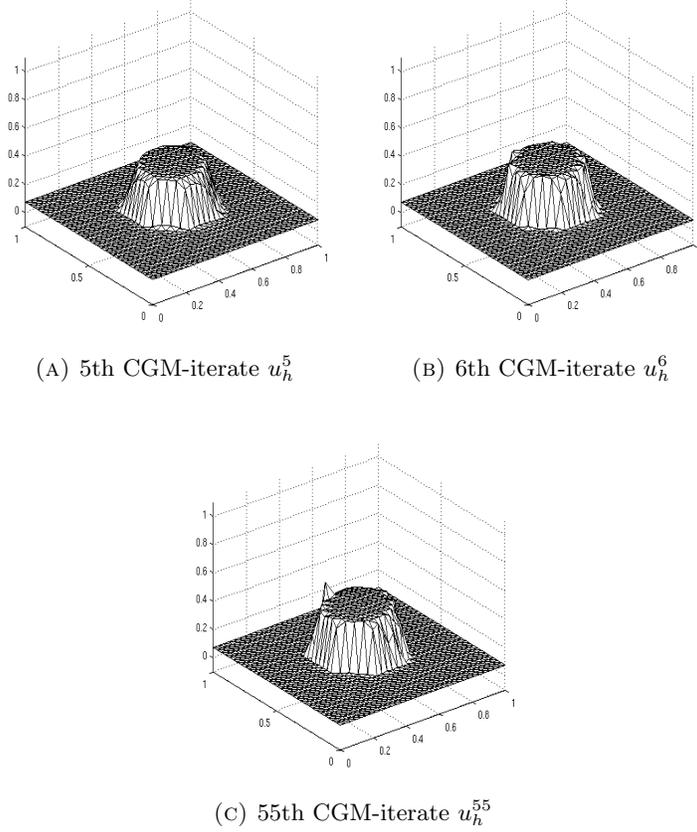


FIGURE 9. Outputs of the CGM method defined in VII.F after 5, 6 and 55 iterations, respectively, $h = \sqrt{2}2^{-5}$, $\varepsilon = h$, $\rho = 0.99$, $u_h^0 = 0$, $p_h^0 = 0$, for Example VI.1.

VII.G. Block coordinate descent algorithm. In the recent paper [15], the authors consider a finite difference discretization of the anisotropic BV -seminorm, i.e., defined via the ℓ^1 -norm of ∇u , and split it into the horizontal components TV_1 and the vertical components TV_2 . After introducing two new variables u_1 and u_2 , they obtain the saddle-point problem

$$\inf_{u_1, u_2, u} \sup_{\lambda_1, \lambda_2} TV_1(u_1) + (\lambda_1, u - u_1) + TV_2(u_2) + (\lambda_2, u - u_2) + \frac{\lambda}{2} \|u - g\|^2$$

with Lagrange multipliers λ_1 and λ_2 . Exchanging infimum and supremum and solving the inner minimization problem yields the dual problem

$$\sup_{\lambda_1, \lambda_2} \left(-TV_1^*(\lambda_1) - TV_2^*(\lambda_2) - \frac{1}{2\alpha} \|\lambda_1 + \lambda_2\|^2 + (\lambda_1 + \lambda_2, g) \right).$$

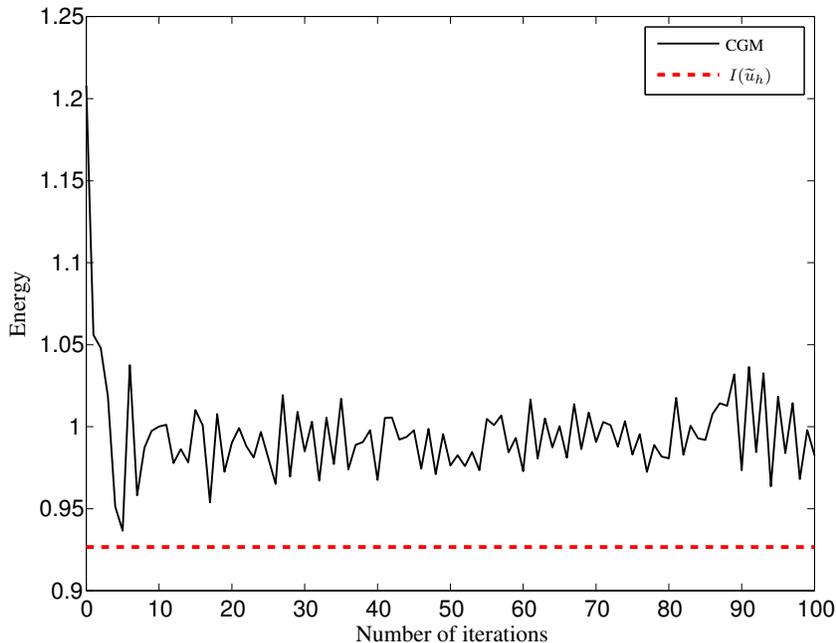


FIGURE 10. Energy of iterates u_h^n for the CGM method defined in VII.F, $0 \leq n \leq 100$, $h = \sqrt{2}2^{-5}$, $\varepsilon = h$, $\rho = 0.99$, $u_h^0 = 0$, $p_h^0 = 0$, and energy of the reference solution \tilde{u}_h for Example VI.1.

The authors suggest to solve this optimization problem via alternating minimization which, as they observe, simplifies to MN independent one-dimensional ROF problems, where $M \times N$ is the size of the image. Since this approach seems to be only applicable to the ROF problem with anisotropic BV -seminorm, we did not further consider this approach.

VII.H. Splitting and alternating minimization. In [44] a splitting approach is also the starting point for the definition of the algorithm suggested therein. As in the augmented Lagrangian method, the variable $\sigma = \nabla u$ is introduced and leads to the minimization of the functional

$$\int_{\Omega} |\sigma| dx + \frac{\delta^{-1}}{2} \|\sigma - \nabla u\|^2 + \frac{\alpha}{2} \|u - g\|^2$$

with $\delta > 0$. Here, the constraint $\sigma = \nabla u$ is included in the functional by penalization. The authors then propose an alternating minimization technique with respect to the variables σ and u . Since the constraint is not strictly enforced via a Lagrange multiplier the constraint is satisfied only in the limit $\delta \rightarrow 0$. This is, however, a major drawback since the involved matrices that arise in the linear systems of equations that have to be solved

in each step become ill-conditioned as δ tends to 0. Besides, we do not have such a relation as in (2) in order to determine an optimal scaling of δ to recover the optimal convergence rate. Particularly, we do not know how stopping criteria should look like to attain an accurate approximation.

The iteration scheme of the alternating minimization algorithm (AMA) proposed in [41] coincides with the iteration scheme of the augmented Lagrangian method in Algorithm IV.1 with $\|\cdot\|_h = \|\cdot\|$ except from the updating rule for the variable u_h , i.e., for $j \geq 0$, the iterate u_h^{j+1} is defined as the minimizer of the mapping

$$u_h \mapsto \int_{\Omega} |\sigma_h^j| dx + \frac{\alpha}{2} \|u_h - g\|^2 + (\lambda_h^j, \sigma_h^j - \nabla u_h, j).$$

However, this numerical method requires $\tau = \mathcal{O}(h^2)$, which is impractical. In [31] the authors discuss acceleration techniques for the augmented Lagrangian method and the AMA proposed in [41]. In order to enforce stability, a restart condition has to be included in the accelerated augmented Lagrangian method. In our experiments, the accelerated augmented Lagrangian method was faster than the standard augmented Lagrangian method in some cases. However, in some situations the number of restarts negatively affected the performance of the acceleration which resulted in higher iteration numbers than for the augmented Lagrangian method without acceleration. Since we cannot control the number of restarts we did not further consider the accelerated augmented Lagrangian method. The acceleration technique for the AMA was not able to compensate for the slow convergence due to the restrictive condition $\tau = \mathcal{O}(h^2)$.

VII.I. Second-order cone programming. Considering the constrained ROF problem, i.e., the minimization of $|Du|(\Omega)$ subject to the constraint $\|u - g\|^2 < \eta^2$ with η^2 being an estimate of the variance of the noise in the image g , a finite difference discretization of the constrained ROF problem has been reformulated as a second-order cone program (SOCP) in [30]. The reported numerical experiments indicate that the SOCP can compute accurate approximations within a few iterations. However, although not explicitly introduced, during the iterative solution of the SOCP by an interior-point method as proposed a regularization of the problem is implicitly introduced by adding a barrier functional to solve the second-order cone constraints, where the influence of the regularization parameter remains unclear. In our opinion, since the interior-point method uses Newton iterations the behavior of the SOCP with regards to mesh-dependent parameters is not clear. Moreover, since the reformulation of the ROF problem as a SOCP is based on a finite difference discretization of the ROF functional we decided to not further investigate the SOCP in this paper.

VII.J. Split Brègman Method. The split Brègman algorithm [32] proposed for the ROF functional is an iterative algorithm for the constrained

minimization problem

$$\min_{\sigma_h, u_h} \int_{\Omega} |\sigma_h| dx + \frac{\alpha}{2} \|u_h - g\|^2 \quad \text{s.t. } \sigma_h = \nabla u_h.$$

Applying Brègman iteration to the penalized functional

$$\int_{\Omega} |\sigma_h| dx + \frac{\alpha}{2} \|u_h - g\|^2 + \frac{\tau}{2} \|\sigma_h - \nabla u_h\|^2$$

defines the following iteration: given $\eta_h^0 \in \mathcal{S}^1(\mathcal{T}_h)$, $\mu_h^0 \in \mathcal{L}^0(\mathcal{T}_h)^d$, compute for $k \geq 0$

$$\begin{aligned} (u_h^{k+1}, \sigma_h^{k+1}) &= \underset{(u, \sigma)}{\operatorname{argmin}} D^{\eta_h^k, \mu_h^k}(u, \sigma; u_h^k, \sigma_h^k) + \frac{\tau}{2} \|\sigma - \nabla u\|^2 \\ &= \underset{(u, \sigma)}{\operatorname{argmin}} \|\sigma\|_{L^1(\Omega)} + \frac{\alpha}{2} \|u - g\|^2 - \|\sigma_h^k\|_{L^1(\Omega)} - \frac{\alpha}{2} \|u_h^k - g\|^2 \\ &\quad - (\eta_h^k, u - u_h^k) - (\mu_h^k, \sigma - \sigma_h^k) + \frac{\tau}{2} \|\sigma - \nabla u\|^2 \end{aligned}$$

and set

$$\eta_h^{k+1} = \eta_h^k + \tau \operatorname{div}_h(\sigma_h^{k+1} - \nabla u_h^{k+1}), \quad \mu_h^{k+1} = \mu_h^k + \tau(\nabla u_h^{k+1} - \sigma_h^{k+1})$$

where $\operatorname{div}_h : \mathcal{L}^0(\mathcal{T}_h)^d \rightarrow \mathcal{S}^1(\mathcal{T}_h)$ is the adjoint operator of $\nabla : \mathcal{S}^1(\mathcal{T}_h) \rightarrow \mathcal{L}^0(\mathcal{T}_h)^d$, i.e., $(\operatorname{div}_h q_h, v_h) = -(q_h, \nabla v_h)$ for all $q_h \in \mathcal{L}^0(\mathcal{T}_h)^d$, $v_h \in \mathcal{S}^1(\mathcal{T}_h)$. The function

$$\begin{aligned} D^{\eta_h^k, \mu_h^k}(u, \sigma; u_h^k, \sigma_h^k) &= \|\sigma\|_{L^1(\Omega)} + \frac{\alpha}{2} \|u - g\|^2 - \|\sigma_h^k\|_{L^1(\Omega)} - \frac{\alpha}{2} \|u_h^k - g\|^2 \\ &\quad - (\eta_h^k, u - u_h^k) - (\mu_h^k, \sigma - \sigma_h^k) \end{aligned}$$

is also known as the *Brègman distance* corresponding to the functional

$$\|\sigma\|_{L^1(\Omega)} + \frac{\alpha}{2} \|u - g\|^2$$

we are considering here, and (η_h^k, μ_h^k) is a subgradient of the aforementioned functional at the point (u_h^k, σ_h^k) . Due to the separate convexity of the functional that is being minimized in each iteration an alternating minimization can be employed to decouple the minimization problem. Thus, we get the following optimality conditions for the iterates:

$$\begin{aligned} \alpha(u_h^{k+1} - g, v_h) - (\eta_h^k, v_h) - \tau(\sigma_h^k - \nabla u_h^{k+1}, \nabla v_h) &= 0, \\ \tau(\nabla u_h^{k+1} - \sigma_h^{k+1} + \mu_h^k, q_h - \sigma_h^{k+1}) + \|\sigma_h^{k+1}\|_{L^1(\Omega)} &\leq \|q_h\|_{L^1(\Omega)} \end{aligned}$$

for all $(v_h, q_h) \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{L}^0(\mathcal{T}_h)^d$. Consider the k -th iterate λ_h^k of the Lagrange multiplier in the augmented Lagrangian method. If we set $\eta_h^k = \operatorname{div}_h \lambda_h^k$ and $\mu_h^k = -\lambda_h^k$ the split Brègman iteration coincides with the augmented Lagrangian method, as has been observed in [45]. Since here the constraint is also measured in the L^2 -norm and since sequences of minimizers $(\nabla u_h)_{h>0}$ may not be bounded in $L^2(\Omega)$, this method is not guaranteed to be robust or may require small step sizes.

8. CONCLUSIONS

We compared three approaches for the numerical solution of total variation regularized minimization problems by considering the prototypical problem of minimizing the ROF functional. The nonlinearity that arises in the formal Euler-Lagrange equation for the ROF problem and the nondifferentiability of the BV -seminorm cause considerable problems when trying to carry out a rigorous numerical analysis for approximation schemes. Although there exist many algorithms for approximately solving the ROF problem, which may also have a good performance with regards to the special application in image processing, the dependence of many of those algorithms on critical parameters is not well-understood. However, for the three schemes we considered in this paper, which are all of gradient descent and gradient ascent type with expected convergence behavior $\mathcal{O}(1/J)$, there is a rigorous analysis available. In particular, we have modified the Heron method and the augmented Lagrangian method to such an extent that, in view of the infinite-dimensional setting, robustness of the schemes is guaranteed. We have seen that the Heron method needs the least iteration numbers for a given accuracy, is robust regarding the initialization and its stability is independent of the step size. The $(H^{1/2}$ -)primal-dual method is slower due to the restriction on the step size, i.e., $\tau = \mathcal{O}(h)$ for the primal-dual method and $\tau = \mathcal{O}(h^{1/2})$ for the $H^{1/2}$ -primal-dual method. It is not based on an explicit regularization of the BV -seminorm in contrast to the Heron method. However, by seeking a discrete minimizer of the ROF functional in $\mathcal{S}^1(\mathcal{T}_h) \subset W^{1,1}(\Omega)$ we approximate the minimizer of the continuous problem by continuous functions which resembles a regularization of the problem which nevertheless allows for sharp approximations of discontinuities. More general approaches for the approximation of BV -regularized problems are splitting methods. We considered the augmented Lagrangian method with a weighted L^2 -norm in order to ensure unconditional stability and we experimentally observed that the augmented Lagrangian method performs as well as the $H^{1/2}$ -primal-dual method for the step size $\tau = h^{-3/2}$.

9. APPENDIX A

IX.A. Roots of Quartic Equations. Having (3) in mind we consider the problem of finding the roots of a quartic equation

$$x^4 + ax^3 + bx^2 + cx + d = 0. \quad (14)$$

with $a, b, c, d \in \mathbb{R}$. A comprehensive discussion of this topic can be found, e.g., in [10]. With the substitution $x = y - \frac{1}{4}a$ we equivalently obtain

$$y^4 + \tilde{b}y^2 + \tilde{c}y + \tilde{d} = 0 \quad (15)$$

with $\tilde{b} = -\frac{3}{8}a^2 + b$, $\tilde{c} = \frac{1}{8}a^3 - \frac{ab}{2} + c$ and $\tilde{d} = -\frac{3}{256}a^4 + \frac{a^2b}{16} - \frac{ac}{4} + d$. Adding $\frac{1}{4}\tilde{b}^2$ on both sides yields

$$\left(y^2 + \frac{1}{2}\tilde{b}\right)^2 = \frac{1}{4}\tilde{b}^2 - \tilde{c}y - \tilde{d}.$$

Our goal is to produce a perfect square on the right-hand side. The following procedure is due to *Lodovico Ferrari*. We introduce a new variable z , which is to be specified later, within the square on the left-hand side and obtain

$$\begin{aligned} \left(y^2 + \frac{1}{2}\tilde{b} + \frac{1}{2}z\right)^2 &= \frac{1}{4}\tilde{b}^2 - \tilde{c}y - \tilde{d} + \frac{1}{4}z^2 + z\left(y^2 + \frac{1}{2}\tilde{b}\right) \\ &= zy^2 - \tilde{c}y + \frac{1}{4}z^2 + \frac{1}{2}\tilde{b}z + \frac{1}{4}\tilde{b}^2 - \tilde{d}. \end{aligned} \quad (16)$$

Now the right-hand side is a perfect square with respect to y if and only if the associated discriminant vanishes, i.e., if and only if

$$0 = 4z\left(\frac{1}{4}z^2 + \frac{1}{2}\tilde{b}z + \frac{1}{4}\tilde{b}^2 - \tilde{d}\right) - \tilde{c}^2 \quad (17)$$

$$= z^3 + 2\tilde{b}z^2 + (\tilde{b}^2 - 4\tilde{d})z - \tilde{c}^2. \quad (18)$$

The cubic polynomial in (18) is called *cubic resolvent*. We see that with z being any root of the cubic resolvent the term in (16) simplifies to a perfect square. Indeed, we have using (17)

$$\left(y^2 + \frac{1}{2}\tilde{b} + \frac{1}{2}z\right)^2 = z\left(y^2 - \frac{\tilde{c}}{z}y + \frac{\tilde{c}^2}{4z^2}\right) = z\left(y - \frac{\tilde{c}}{2z}\right)^2.$$

We can take z to be, for instance, a real root of (18) and obtain

$$y^2 + \frac{1}{2}\tilde{b} + \frac{1}{2}z = \pm\sqrt{z}\left(y - \frac{\tilde{c}}{2z}\right),$$

so we have two quadratic algebraic equations in y and can compute the roots of (15). We obtain the roots $\{x_i\}_{1 \leq i \leq 4}$ of our original equation with the relation $x_i = y_i - \frac{1}{4}a$.

IX.B. Roots of Cubic Equations. In order to obtain the roots of (18) we briefly discuss how to compute the roots of a cubic algebraic equation. Given an equation of the form

$$x^3 + ax^2 + bx + c = 0 \quad (19)$$

we set $x = y - \frac{a}{3}$ and get

$$y^3 + \tilde{b}y + \tilde{c} = 0 \quad (20)$$

with $\tilde{b} = b - \frac{a^2}{3}$ and $\tilde{c} = c - \frac{ab}{3} + \frac{2a^3}{27}$. With a further substitution $y = u + v$, with - for the time being - arbitrary u and v , we get

$$\begin{aligned} y^3 &= (u + v)^3 = (u + v)(u^2 + 2uv + v^2) \\ &= u^3 + v^3 + 3uv(u + v) \\ &= u^3 + v^3 + 3uvy \\ &\Leftrightarrow y^3 - 3uvy - (u^3 + v^3) = 0. \end{aligned}$$

Comparing coefficients yields

$$\tilde{b} = -3uv, \quad \tilde{c} = -(u^3 + v^3).$$

According to Vieta's formulas, u^3 and v^3 are the roots of the quadratic algebraic equation

$$t^2 + \tilde{c}t - \frac{\tilde{b}^3}{27} = 0.$$

Hence we have

$$u = \sqrt[3]{-\frac{\tilde{c}}{2} + \sqrt{\frac{\tilde{c}^2}{4} + \frac{\tilde{b}^3}{27}}},$$

$$v = \sqrt[3]{-\frac{\tilde{c}}{2} - \sqrt{\frac{\tilde{c}^2}{4} + \frac{\tilde{b}^3}{27}}}.$$

The cube roots u and v have to be chosen such that $uv = -\frac{\tilde{b}}{3} \in \mathbb{R}$. Denoting by u_1, v_1 the cube roots defined by $\sqrt[3]{re^{i\varphi/3}}$ for a complex number $re^{i\varphi}$, the two other cube roots are given by $u_2 = u_1e^{2\pi i/3}$, $u_3 = u_1e^{4\pi i/3}$ and $v_2 = v_1e^{2\pi i/3}$, $v_3 = v_1e^{4\pi i/3}$, respectively. Since the product uv has to be real, we get the three feasible pairs (u_1, v_1) , (u_2, v_3) and (u_3, v_2) . Hence, the three roots of (20) are given by

$$y_1 = u_1 + v_1,$$

$$y_2 = u_2 + v_3,$$

$$y_3 = u_3 + v_2.$$

The roots $\{x_i\}_{1 \leq i \leq 3}$ of (19) are then given by $x_i = y_i - \frac{a}{3}$.

Acknowledgements. The authors acknowledge financial support for the project "Finite Element approximation of functions of bounded variation and application to model of damage, fracture and plasticity" (BA 2268/2-1) by the DFG via the priority program "Reliable Simulation Techniques in Solid Mechanics, Development of Non-standard Discretization Methods, Mechanical and Mathematical Analysis" (SPP 1748).

REFERENCES

- [1] L. Ambrosio, N. Fusco, D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford Mathematical Monographs, The Clarendon Press, Oxford University Press, New York, 2000.
- [2] H. Attouch, G. Buttazzo, G. Michaille, *Variational Analysis in Sobolev and BV Spaces*, MPS/SIAM Series on Optimization, Vol. 6, Philadelphia, 2006.
- [3] S. Bartels, *Broken Sobolev Space Iteration for Total Variation Regularized Minimization Problems*, IMA Journal of Numerical Analysis, doi:10.1093/imanum/drv023, 2015.
- [4] S. Bartels, *Total Variation Minimization with Finite Elements: Convergence and Iterative Solution*, SIAM Journal on Numerical Analysis, Vol. 50, 2012, pp. 1162-1180.

- [5] S. Bartels, *Numerical Methods for Nonlinear Partial Differential Equations*, Springer, Heidelberg, 2015.
- [6] S. Bartels, A. Mielke, T. Roubiřek, *Quasi-Static Small-Strain Plasticity in the Limit of Vanishing Hardening and its Numerical Approximation*, SIAM J. Numer. Anal., Vol. 50, No. 2, pp. 951-976.
- [7] S. Bartels, R. H. Nochetto, A. J. Salgado, *Discrete Total Variation Flows without Regularization*, SIAM J. Numer. Anal., Vol. 52, 2014, pp. 363-385.
- [8] A. Beck, M. Teboulle, *A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems*, SIAM J. Imaging Sci., Vol. 2, No. 1, 2009, pp. 183-202.
- [9] J.-D. Benamou, G. Carlier, *Augmented Lagrangian Methods for Transport Optimization, Mean Field Games and Degenerate Elliptic Equations*, Journal of Optimization Theory and Applications, Vol. 167, 2015, pp. 1-26.
- [10] J. Bewersdorff, *Algebra für Einsteiger*, 5. Auflage, Springer, Wiesbaden, 2013.
- [11] L. M. Bregman, *The Relaxation Method of Finding the Common Points of Convex Sets and its Application to the Solution of Problems in Convex Programming*, USSR Computational Mathematics and Mathematical Physics, Vol. 7, 1967, pp. 200-217.
- [12] A. Chambolle, *An Algorithm for Total Variation Minimization and Applications*, Journal of Mathematical Imaging and Vision, Vol. 20, 2004, pp. 89-97.
- [13] A. Chambolle, P.-L. Lions, *Image Recovery via Total Variation Minimization and Related Problems*, Numerische Mathematik, Vol. 76, 1997, pp. 167-188.
- [14] A. Chambolle, T. Pock, *A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging*, J Math Imaging Vis, Vol. 40, 2011, pp. 120-145.
- [15] A. Chambolle, T. Pock, *A Remark on Accelerated Block Coordinate Descent for Computing the Proximity Operators of a Sum of Convex Functions*, SMAI Journal of Computational Mathematics, Vol. 1, 2015, pp. 29-54.
- [16] R. H. Chan, H.-X. Liang, *A Fast and Efficient Half-Quadratic Algorithm for TV-L1 Restoration*, CHKU research report 370, 2010, available at <ftp://ftp.math.cuhk.edu.hk/report/2010-03.pdf>.
- [17] T. F. Chan, Gene H. Golub, P. Mulet, *A Nonlinear Primal-Dual Method for Total Variation-Based Image Restoration*, SIAM J. Sci. Comput., Vol. 20, 1999, pp. 1964-1977.
- [18] T. F. Chan, P. Mulet, *On the Convergence of the Lagged Diffusivity Fixed Point Method in Total Variation Image Restoration*, SIAM J. Numer. Anal., Vol. 36, 1999, pp. 354-367.
- [19] C. Clason, K. Kunisch, *A Duality-Based Approach to Elliptic Control Problems in Non-Reflexive Banach Spaces*, ESAIM Control, Optimisation and Calculus of Variations, Vol. 17, 2011, pp. 243-266.
- [20] S. Conti, J. Ginster, M. Rumpf, *A BV Functional and its Relaxation for Joint Motion Estimation and Image Sequence Recovery*, Preprint, to appear in: ESAIM Mathematical Modelling and Numerical Analysis, Vol. 49, 2015, pp. 1463-1487.
- [21] G. Dal Maso, A. DeSimone, M. G. Mora, *Quasistatic Evolution Problems for Linearly Elastic-Perfectly Plastic Materials*, Arch. Ration. Mech. Anal., Vol. 180, 2006, pp. 237-291.
- [22] J. Darbon, M. Sigelle, *A Fast and Exact Algorithm for Total Variation Minimization*, IbPRIA, LNCS 3522, 2005, pp. 351-359.
- [23] D. C. Dobson, C. R. Vogel, *Convergence of an Iterative Method for Total Variation Denoising*, SIAM J. Numer. Anal., Vol. 34, No. 5, 1997, pp. 1779-1791.
- [24] J. Eckstein, D. P. Bertsekas, *On the Douglas-Rachford Splitting Method and the Proximal Point Algorithm for Maximal Monotone Operators*, Math. Programming, Vol. 55, 1992, pp. 293-318.
- [25] I. Ekeland, R. Témam, *Convex Analysis and Variational Problems*, SIAM, Philadelphia, 1999.

- [26] C. M. Elliott, S. A. Smitheman, *Numerical Analysis of the TV Regularization and H^{-1} Fidelity Model for Decomposing an Image into Cartoon plus Texture*, IMA J. Numer. Anal., Vol. 29, 2009, pp. 651-689.
- [27] X. Feng, A. Prohl, *Analysis of Total Variation Flow and its Finite Element Approximations*, ESAIM Mathematical Modelling and Numerical Analysis, Vol. 37, 2003, pp. 533-556.
- [28] M. Fortin, R. Glowinski, *Augmented Lagrangian Methods*, Studies in Mathematics and its Applications, Vol. 15, North-Holland Publishing Co., Amsterdam, 1983.
- [29] R. Glowinski, P. Le Tallec, *Augmented Lagrangians and Operator-Splitting Methods in Nonlinear Mechanics*, SIAM, Philadelphia, 1989.
- [30] D. Goldfarb, W. Yin, *Second-Order Cone Programming Methods for Total Variation-Based Image Restoration*, SIAM J. Sci. Comput., Vol. 27, 2005, pp. 622-645.
- [31] T. Goldstein, B. O'Donoghue, S. Setzer, R. Baraniuk, *Fast Alternating Direction Optimization Methods*, SIAM J. Imaging Sci., Vol. 7, 2014, pp. 1588-1623.
- [32] T. Goldstein, S. Osher, *The Split Bregman Method for $L1$ Regularized Problems*, SIAM J. Imaging Sci., Vol. 2, 2009, pp. 323-343.
- [33] O. Güler, *Foundations of Optimization*, Graduate Texts in Mathematics, Vol. 258, Springer, New York, 2010.
- [34] M. Hintermüller, K. Ito, K. Kunisch, *The Primal-Dual Active Set Strategy as a Semismooth Newton Method*, SIAM J. Optim., Vol.13, 2003, pp. 865-888.
- [35] M. Hintermüller, K. Kunisch, *Total Bounded Variation Regularization as a Bilaterally Constrained Optimization Method*, SIAM J. Appl. Math., Vol. 64, 2004, pp. 1311-1333.
- [36] Y. Nesterov, *Smooth Minimization of Non-Smooth Functions*, Math. Program., Ser. A, Vol. 103, 2005, pp. 127-152.
- [37] N. Papadakis, G. Peyré, E. Oudet, *Optimal Transport with Proximal Splitting*, SIAM Journal on Imaging Sciences, Vol. 7, 2014, pp. 212-238.
- [38] R. T. Rockafellar, *Monotone Operators and the Proximal Point Algorithm*, SIAM J. Control and Optimization, Vol. 14, 1976, pp. 877-898.
- [39] L. I. Rudin, S. Osher, E. Fatemi, *Nonlinear Total Variation Based Noise Removal Algorithms*, Physica D, Vol. 60, 1992, pp. 259-268.
- [40] M. Thomas, *Quasistatic damage evolution with spatial BV-regularization*, Discr. Cont. Dyn. Syst., Series S, Vol. 6, 2013, pp. 235-255.
- [41] P. Tseng, *Applications of a Splitting Algorithm to Decomposition in Convex Programming and Variational Inequalities*, SIAM J. Control Optim., Vol. 29, 1991, pp. 119-138.
- [42] C. R. Vogel, M. E. Oman, *Iterative Methods for Total Variation Denoising*, SIAM J. Sci. Comput., Vol. 17, 1996, pp. 227-238.
- [43] J. Wang, B. J. Lucier, *Error Bounds for Finite-Difference Methods for Rudin-Osher-Fatemi Image Smoothing*, SIAM J. Numer. Anal., Vol. 49, 2011, pp. 845-868.
- [44] Y. Wang, J. Yang, W. Yin, Y. Zhang, *A New Alternating Minimization Algorithm for Total Variation Image Reconstruction*, SIAM Journal on Imaging Sciences, Vol. 1, 2008, pp. 248-272.
- [45] C. Wu, X.-C. Tai, *Augmented Lagrangian Method, Dual Methods, and Split Bregman Iteration for ROF, Vectorial TV, and Higher Order Models*, SIAM J. Imaging Sci., Vol. 3, 2010, pp. 300-339.
- [46] M. Zhu, *Fast Numerical Algorithms for Total Variation Based Image Restoration*, Ph.D. thesis, UCLA, 2008.

DEPARTMENT OF APPLIED MATHEMATICS, MATHEMATICAL INSTITUTE, UNIVERSITY OF
FREIBURG, HERMANN-HERDER-STR 9, 79104 FREIBURG I. BR., GERMANY
E-mail address: `bartels@mathematik.uni-freiburg.de`

DEPARTMENT OF APPLIED MATHEMATICS, MATHEMATICAL INSTITUTE, UNIVERSITY OF
FREIBURG, HERMANN-HERDER-STR 9, 79104 FREIBURG I. BR., GERMANY
E-mail address: `marijo.milicevic@mathematik.uni-freiburg.de`