SIAM J. NUMER. ANAL.
Vol. 50, No. 3, pp. 1162–1180

TOTAL VARIATION MINIMIZATION WITH FINITE ELEMENTS: CONVERGENCE AND ITERATIVE SOLUTION*

SÖREN BARTELS†

Abstract. The numerical solution of a convex minimization problem involving the nonsmooth total variation norm is analyzed. Consistent finite element discretizations that avoid regularizations lead to simple convergence proofs in the case of piecewise affine, globally continuous finite elements. For the approximation with piecewise constant finite elements it is proved that convergence to the exact solution cannot be expected in general. The iterative solution is based on a regularized L^2 flow of the energy functional, and convergence of the iteration to a stationary point is proved under a moderate constraint on the time-step size. The extension of the techniques to an energy functional that involves a negative order term is discussed. Numerical experiments that illustrate the theoretical results are presented.

Key words. total variation, finite elements, convergence, partial differential equations, image processing

AMS subject classifications. 65N12, 65N30, 35Q68

DOI. 10.1137/11083277X

1. Introduction.

As a model problem for total variation minimization, we consider the energy functional $E : BV(\Omega) \cap L^2(\Omega) \to \mathbb{R}$ defined by

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

for a bounded Lipschitz domain $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, a function $g \in L^2(\Omega)$, and a parameter $\alpha > 0$. Here, $BV(\Omega) \subset L^1(\Omega)$ consists of all functions $v \in L^1(\Omega)$ whose distributional gradient is a Radon measure with bounded total variation, i.e., with $|\cdot|$ denoting the Euclidean norm in $\mathbb{R}^d$,

$$\|Dv\| = \sup\left\{\int_\Omega v \operatorname{div} p \,\mathrm{d}x : p \in C_0^\infty(\Omega;\mathbb{R}^d),\ |p| \le 1\right\} < \infty.$$

Since this quantity defines a weakly* lower semicontinuous functional on $BV(\Omega)$, it is easy to verify that the functional E has a minimizer in $BV(\Omega)$ which is unique owing to the strict convexity of the quadratic part of E.

Piecewise constant as well as globally continuous, piecewise affine finite elements are included in $BV(\Omega)$ so that the restriction of the functional E to these spaces admits unique minimizers. The questions we address are whether discrete solutions converge to the exact solution, how the total variation norm can be evaluated practically, and whether the discrete problems can be solved efficiently without the use of regularizations of the functional E.

It is a straightforward task to check that piecewise constant and piecewise affine, globally continuous finite element spaces are dense in $BV(\Omega)$ with respect to weak* convergence in $BV(\Omega)$. This means that for every function $v \in BV(\Omega)$ there exists a sequence $(v_h)_{h>0}$ in the respective finite element spaces that converges strongly in

*Received by the editors May 2, 2011; accepted for publication (in revised form) January 24, 2012; published electronically May 15, 2012.
http://www.siam.org/journals/sinum/50-3/83277.html
†Institute for Numerical Simulation, Rheinische Friedrich-Wilhelms-Universität Bonn, Wegelerstr. 6, 53115 Bonn, Germany (bartels@ins.uni-bonn.de).

$L^1(\Omega)$ to v and whose distributional gradients are measures with uniformly bounded total variation. Unfortunately, this is not sufficient to deduce that discrete minimizers accumulate at exact solutions. For this, the stronger notion of strict convergence is needed, which additionally requires that $\|Dv_h\| \to \|Dv\|$ as the mesh-size h tends to zero. In the case of $W^{1,1}$-conforming finite elements this is easy to establish, and if $\Omega = (0,1)^2$ and $g \in L^\infty(\Omega)$, then the error estimate $\|u - u_h\|_{L^2(\Omega)} \le ch^{1/6}$ can be proved following an argument from [WL09]. The employed finite element setting and the resulting consistent treatment of the energy functional leads to very simple proofs of these statements. In the language of Γ-convergence we show that the restriction of the total variation norm to the finite element space converges to the total variation norm in the topology of $L^1(\Omega)$. For the approximation with piecewise constant finite elements this may not be true and we demonstrate that in general convergence of approximations to the exact solution cannot be expected. This is surprising since the approximation of functions in $BV(\Omega)$ with piecewise constant finite elements appears natural at first glance. The proof shows that for a generic function $u \in BV(\Omega)$ and a generic sequence of triangulations the simultaneous convergence $\|u - u_h\|_{L^1(\Omega)} \to 0$ and $\|Du_h\| \to \|Du\|$ is impossible for any sequence of piecewise constant finite element functions. Nevertheless, the approximation with piecewise constants may be sufficient in some applications, e.g., in image denoising.

In order to compute discrete minimizers of E we consider its L^2 gradient flow, which is formally given by

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

with the subdifferential $\partial|\cdot|$ and supplemented by Neumann boundary conditions and initial data. The equivalence

$$p \in \partial|\nabla u| \iff \nabla u \in \partial I_B(p),$$

where $B = \{p \in L^1(\Omega;\mathbb{R}^d) : |p| \le 1\}$ and I_B denotes its indicator functional, provides motivation to consider for $\sigma > 0$ the system of evolution equations

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

A semi-implicit discretization decouples the equations and leads to a scheme that is stable and convergent to a stationary point if the time-step size τ and mesh-size h satisfy the condition $\tau \le c\sigma h$. The proof of this statement is inspired by recent results in [CP11]. Owing to the simple structure of the set B the variational inclusion can be solved pointwise; i.e., for the proposed discretization we have (for $\sigma = 1$)

$$p_h^{n+1} = \frac{p_h^n + \tau\nabla_h \widetilde{u}_h^{n+1}}{\max\{1, |p_h^n + \tau\nabla_h \widetilde{u}_h^{n+1}|\}}, \quad \frac{1}{\tau}(u_h^{n+1} - u_h^n) - \operatorname{div}_h p^{n+1} = -\alpha(u_h^{n+1} - g_h),$$

where $\widetilde{u}_h^{n+1}$ is the extrapolated function $\widetilde{u}_h^{n+1} = 2u_h^n - u_h^{n-1}$, ∇_h and div_h are discrete versions of the gradient and the divergence operator, and d_t is a backward difference operator. Notice that this scheme is fully explicit in the sense that it only requires the inversion of a (lumped) mass matrix in each step.

The correct interpretation of the abstract L^2 flow stated above requires the choice of a discrete space with inner product. For piecewise affine, globally continuous finite element functions subordinated to a triangulation $\mathcal{T}_h$ of Ω into triangles or tetrahedra this choice will be based on the identity

$$\|Dv_h\| = \|\nabla v_h\|_{L^1(\Omega)} = \sum_{T\in\mathcal{T}_h} |T| \sup_{\xi\in\mathbb{R}^d, |\xi|\le 1} \nabla v_h|_T \cdot \xi = \sup_{q_h \in \mathcal{L}^0(\mathcal{T}_h)^d, |q_h|\le 1} \int_\Omega \nabla v_h \cdot q_h \,\mathrm{d}x,$$

where $\mathcal{L}^0(\mathcal{T}_h)^d$ denotes the space of piecewise constant vector fields. In view of this identity it is natural to equip $\mathcal{L}^0(\mathcal{T}_h)^d$ with the L^2 scalar product. Although we make no explicit use of the fact that the pair $\mathcal{L}^0(\mathcal{T}_h)^d \times \mathcal{S}^1(\mathcal{T}_h)$ of finite element spaces is stable for the mixed finite element discretization of the Poisson problem, the inf-sup condition appears to be relevant for the discretization of the total variation norm since for the related Fortin interpolant $I_F : C^\infty(\overline{\Omega};\mathbb{R}^d) \to \mathcal{L}^0(\mathcal{T}_h)^d$ we have

$$\int_\Omega \mathrm{div} p\, v_h \,\mathrm{d}x = \int_\Omega \mathrm{div} I_F p\, v_h \,\mathrm{d}x$$

and this identity leads to consistent discretizations of the total variation norm.

The regularized L^2 flow may also be regarded as a simultaneous gradient flow for a saddle-point formulation of E. We have

$$\inf_{v_h} E(v_h) = \inf_{v_h}\sup_{q_h} \int_\Omega \nabla v_h \cdot q_h \,\mathrm{d}x + (\alpha/2)\|v_h - g\|^2_{L^2(\Omega)} - I_B(q_h) = \inf_{v_h}\sup_{q_h} L(v_h, q_h).$$

The saddle-point problem can be iteratively solved with first order primal-dual algorithms which are discretizations of the system

$$\partial_t u_h = -\partial_v L(u_h, p_h), \quad \sigma \partial_t p_h \in \partial_p L(u_h, p_h).$$

While the subdifferential of L with respect to the first variable is smooth, the second equation defines a variational inclusion. We refer the reader to [CP11] for details and a discussion of relations to proximal point methods introduced in [Roc76].

A general approach to the iterative minimization of a discretization of E includes the following steps: (i) restriction of E to a finite-dimensional subset X_h of $BV(\Omega)$, (ii) identification of a discrete space Y_h that allows the interpretation of the gradient as an operator $\nabla_h : X_h \to Y_h$, and (iii) reformulation of the primal minimization problem as a saddle-point formulation with techniques from convex analysis. A proof of convergence for the iterative solution in an abstract, simple situation that follows [CP11] is discussed in the appendix to illustrate this general concept. The formal derivation of the iterative scheme as a regularized L^2 flow has the advantage that it can cover more general problems involving nonconvex contributions to the energy functional.

The presented finite element framework is particularly useful when the energy functional involves terms that contain differential operators. To illustrate this aspect we also discuss the approximation of the functional

$$E'(v) = \|Dv\| + (\lambda/2)\|v - g\|^2_{-1},$$

where $\|v\|_{-1}$ is the operator norm of v when interpreted as a functional on $H^1(\Omega)$. We prove convergence to the minimizer of E' for an iterative scheme and show that numerical approximations converge to the minimizer of the continuous problem. The discretization of E with finite elements is similar to discretizations based on finite difference methods on uniform grids which are popular and efficient in image processing. For problems in mechanics the approach with finite elements seems preferable.

The energy functional E has been proposed in [ROF92] for the restoration or reconstruction of a true image u from an observation g that is a noisy or blurred version of u. The use of the total variation norm rather than a (smooth) $W^{1,p}$ seminorm for $p > 1$ is essential since it allows for discontinuities along curves so that edges and contours are kept in the minimizers. The modification E' that uses the weaker H^{-1} norm to define the fidelity term is appropriate to represent textures or oscillatory

patterns and has been proposed in [Mey01, OSV03]. We refer the reader to the papers [ROF92, OSV03] for interesting numerical studies of these models. A minor disadvantage of the functionals E and E' is that they may lead to the formation of so-called staircasing. To avoid this, one may regularize the total variation norm by replacing the modulus $|\cdot|$ by a continuously differentiable approximation $|\cdot|_\varepsilon$. Although this leads to a smooth minimization problem which can be treated with standard methods, the performance of the resulting algorithms is poor and numerical schemes should not make explicit use of this regularization. Using that, for example, $|\cdot|_\varepsilon^* = I_B + (\varepsilon/2)|\cdot|^2$, the methods discussed in this paper can be modified to include the regularization leading to schemes whose performance is independent of ε; cf. [CP11]. A different approach to avoid staircasing is the use of higher order derivatives; cf. [CL97, BKP10].

The iterative solution of functionals involving the total variation norm has been discussed in a number of papers. The paper [Cha04] proves convergence of a semi-implicit discretization of the L^2 flow of the dual problem for E. This leads to simple problems in each time step but requires that the time-step size satisfy the restrictive condition $\tau \le ch^2$. The implicit discretization of gradient flows of regularized versions of E and E' has been investigated in [FP03, FvOP05, ES09]. Other methods such as nonsmooth Newton methods have been studied in [DV97, CGM99, HK04] and seem to require the use of certain regularizations. The recent paper [CP11] proves convergence of a primal-dual method for the iterative minimization of a finite difference discretization of E and leads to an algorithm similar to the one presented in this paper. For acceleration methods and the application to modifications of the functional E we refer the reader to [CP11] and the references therein. The first error bound for a discretization of E has been given in [WL09] using finite differences.

The outline of this article is as follows. In section 2 we recall some elementary facts about finite element spaces, discrete time derivatives, convex analysis, and functions of bounded variation. Piecewise affine, globally continuous approximations are discussed in section 3 and piecewise constant, discontinuous methods in section 4. Section 5 is devoted to the numerical analysis of a model involving a negative norm. Numerical experiments for all methods and energy functionals studied in the paper are reported in section 6. Appendix A reviews the convergence proof of the primal-dual algorithm from [CP11] in a simplified setting.

2. Preliminaries. We recall some elementary identities and results that are relevant in the subsequent sections. Throughout this paper we denote by c a constant that is independent of discretization parameters and by $(\cdot,\cdot)$ the L^2 scalar product.

2.1. Discrete time derivatives. Given a time-step size $\tau > 0$ and a sequence of functions or real numbers $(v^n)_{n\in\mathbb{N}}$ we define

$$d_t v^{n+1} = (v^{n+1} - v^n)/\tau$$

and notice that we have

$$d_t v^{n+1} \cdot (v^{n+1} - v) = \frac{d_t}{2}\|v - v^{n+1}\|^2 + \frac{\tau}{2}\|d_t v^{n+1}\|^2.$$

A proof follows from considering $\widetilde{v}^{n+1} = v^{n+1} - v$ and noting that $2\tau(d_t\widetilde{v}^{n+1})\cdot\widetilde{v}^{n+1} = 2(\widetilde{v}^{n+1} - \widetilde{v}^n)\cdot\widetilde{v}^{n+1} = (\widetilde{v}^{n+1} - \widetilde{v}^n)\cdot(\widetilde{v}^{n+1} - \widetilde{v}^n) + (\widetilde{v}^{n+1} - \widetilde{v}^n)\cdot(\widetilde{v}^{n+1} + \widetilde{v}^n)$. We also note that for sequences $(a^n)_{n\in\mathbb{N}}$ and $(b^n)_{n\in\mathbb{N}}$ we have $d_t(a^{n+1}\cdot b^{n+1}) = (d_t a^{n+1})\cdot b^{n+1} +$

$a^n \cdot (d_t b^{n+1})$ and hence

$$\tau \sum_{n=0}^{N} \big\{ (d_t a^{n+1}) \cdot b^{n+1} + a^n \cdot (d_t b^{n+1}) \big\} = a^{N+1} \cdot b^{N+1} - a^0 \cdot b^0.$$

2.2. Convex analysis. Given a proper, convex, and lower semicontinuous functional $F : X \to \mathbb{R} \cup \{+\infty\}$ on a normed linear space X its conjugate $F^* : X' \to \mathbb{R} \cup \{+\infty\}$ is defined by $F^*(x') = \sup_{x \in X} \langle x', x \rangle - F(x)$. We have $F^{**} = F$ when F^{**} is restricted to the image of the canonical embedding of X into X''. The subdifferential $\partial F(x)$ of F at $x \in X$ is given by

$$\partial F(x) = \{x' \in X' : \langle x', z - x \rangle \le F(z) - F(x) \text{ for all } z \in X\}.$$

We have that $x' \in \partial F(x)$ if and only if $x \in \partial F^*(x')$.

2.3. Finite element spaces. For a sequence of regular triangulations $(\mathcal{T}_h)_{h>0}$ of Ω into triangles or tetrahedra with maximal diameters $h = \max_{T \in \mathcal{T}_h} \operatorname{diam}(T)$ we define

$$\mathcal{L}^0(\mathcal{T}_h) = \{q_h \in L^1(\Omega) : q_h|_T \text{ is constant for each } T \in \mathcal{T}_h\},$$
$$\mathcal{S}^1(\mathcal{T}_h) = \{v_h \in C(\overline{\Omega}) : v_h|_T \text{ is affine for each } T \in \mathcal{T}_h\}.$$

Let $h_{\min} = \min_{T \in \mathcal{T}_h} \operatorname{diam}(T)$. An elementwise inverse estimate shows that there exists $c > 0$ such that $\|\nabla v_h\|_{L^2(\Omega)} \le c h_{\min}^{-1} \|v_h\|_{L^2(\Omega)}$ for all $v_h \in \mathcal{S}^1(\mathcal{T}_h)$. The nodal interpolant $\mathcal{I}_h v \in \mathcal{S}^1(\mathcal{T}_h)$ of a function $v \in W^{2,p}(\Omega)$, with $d/2 < p \le \infty$ or $d = 2$ if $p = 1$, satisfies (cf., e.g., [BS08])

$$\|v - \mathcal{I}_h v\|_{L^p(\Omega)} + h \|\nabla(v - \mathcal{I}_h v)\|_{L^p(\Omega)} \le c h^2 \|D^2 v\|_{L^p(\Omega)}.$$

2.4. Approximation by smooth functions. The space $BV(\Omega)$ is continuously embedded in $L^p(\Omega)$ for $p \le d/(d-1)$, i.e., we have $\|v\|_{L^p(\Omega)} \le c(\|v\|_{L^1(\Omega)} + \|Dv\|)$ for all $v \in BV(\Omega)$, and this embedding is compact if $p < d/(d-1)$; cf. [AFP00] for details. Smooth functions are dense in $BV(\Omega) \cap L^p(\Omega)$, $1 \le p < \infty$, with respect to strict convergence in the sense that for $v \in BV(\Omega) \cap L^p(\Omega)$ and $\delta > 0$ there exists $\varepsilon_0 > 0$ and functions $(v_\varepsilon)_{\varepsilon > 0} \subset C^\infty(\Omega) \cap BV(\Omega) \cap L^p(\Omega)$ such that for all $\varepsilon \le \varepsilon_0$ we have

$$\begin{aligned} \|\nabla v_\varepsilon\|_{L^1(\Omega)} &\le \|Dv\| + c_0 \delta,\ \|v - v_\varepsilon\|_{L^p(\Omega)} \le c_1 \delta,\ \|D^2 v_\varepsilon\|_{L^p(\Omega)} \\ &\le c\varepsilon^{-2} \|v\|_{L^p(\Omega)},\ \|D^2 v_\varepsilon\|_{L^1(\Omega)} \le c\varepsilon^{-1} \|Dv\|. \end{aligned}$$

If $\Omega = (0,1)^d$, we identify a function $v \in L^1(\Omega)$ with its extension to $\mathbb{R}^d$ obtained by reflection and periodification and write $v \in \operatorname{Lip}(\beta; L^p(\Omega))$ if for $0 \le \beta \le 1$ we have $\sup_{t>0} t^{-\beta} \sup_{|y| \le t} \big(\int_\Omega |v(x+y) - v(x)|^p \, \mathrm{d}x \big)^{1/p} < \infty$. In this case we can choose $c_0 = 0$ and have $\|v - v_\varepsilon\|_{L^p(\Omega)} \le c_1 \varepsilon^\beta$; cf. [WL09].

3. $W^{1,1}$ conforming approximation. Let $u \in BV(\Omega) \cap L^2(\Omega)$ and $u_h \in \mathcal{S}^1(\mathcal{T}_h)$ be the unique minimizers of the functional

$$E(v) = \|Dv\| + (\alpha/2) \|v - g\|_{L^2(\Omega)}^2$$

in $BV(\Omega) \cap L^2(\Omega)$ and $\mathcal{S}^1(\mathcal{T}_h)$, respectively.

THEOREM 3.1 (approximation). *We have $u_h \to u$ in $L^2(\Omega)$ as $h \to 0$, and if $d = 2$, $\Omega = (0,1)^2$, and $u \in \operatorname{Lip}(\beta; L^2(\Omega))$ for some $0 < \beta \le 1$, then $\|u - u_h\|_{L^2(\Omega)}^2 \le c h^{\beta/(1+\beta)}$.*

Proof. Owing to the strict convexity of E and $0 \in \partial E(u)$ we have

$$(\alpha/2)\|u-u_h\|^2_{L^2(\Omega)} \le E(u_h) - E(u).$$

For $\delta > 0$ let $u_\varepsilon \in C^\infty(\Omega) \cap L^2(\Omega)$ be as in section 2.4 and $\mathcal{I}_h u_\varepsilon$ its nodal interpolant. Since u_h is minimal we have the following, using nodal interpolation estimates and the bounds $\|u\|_{L^2(\Omega)}, \|\mathcal{I}_h u_\varepsilon\|_{L^2(\Omega)} \le c$, which hold provided that $h \le c\varepsilon$:

$$\begin{aligned}
&(\alpha/2)\|u-u_h\|^2_{L^2(\Omega)}\\
&\quad\le E(\mathcal{I}_h u_\varepsilon) - E(u)\\
&\quad= \|\nabla \mathcal{I}_h u_\varepsilon\|_{L^1(\Omega)} + (\alpha/2)\|\mathcal{I}_h u_\varepsilon - g\|^2_{L^2(\Omega)} - \|Du\| - (\alpha/2)\|u-g\|^2_{L^2(\Omega)}\\
&\quad\le \|\nabla \mathcal{I}_h u_\varepsilon\|_{L^1(\Omega)} - \|\nabla u_\varepsilon\|_{L^1(\Omega)} + c_0\delta + (\alpha/2)\int_\Omega (\mathcal{I}_h u_\varepsilon - u)(\mathcal{I}_h u_\varepsilon + u + 2g)\,\mathrm{d}x\\
&\quad\le \|\nabla(\mathcal{I}_h u_\varepsilon - u_\varepsilon)\|_{L^1(\Omega)} + c_0\delta + c(\alpha/2)(\|\mathcal{I}_h u_\varepsilon - u_\varepsilon\|_{L^2(\Omega)} + \|u_\varepsilon - u\|_{L^2(\Omega)})\\
&\quad\le ch\|D^2 u_\varepsilon\|_{L^2(\Omega)} + c_0\delta + c(h^2\|D^2 u_\varepsilon\|_{L^2(\Omega)} + \|u_\varepsilon - u\|_{L^2(\Omega)})\\
&\quad\le c\big((h/\varepsilon^2) + c_0\delta + (h/\varepsilon)^2 + c_1\delta\big).
\end{aligned}$$

For h sufficiently small so that $h/\varepsilon^2 \le \delta$ we deduce that $u_h \to u$ in $L^2(\Omega)$. If $d = 2$, $\Omega = (0,1)^d$, and $u \in \mathrm{Lip}(\beta; L^2(\Omega))$, then $c_0 = 0$, $c_1\delta$ can be replaced by $c_1\varepsilon^\beta$, and we may estimate $\|\nabla(\mathcal{I}_h u_\varepsilon - u_\varepsilon)\|_{L^1(\Omega)} \le c(h/\varepsilon)\|Du\|$. With $\varepsilon = h^{1/(\beta+1)}$ we verify the asserted convergence rate. ∎

Remark 3.1. (i) The assumption $u \in \mathrm{Lip}(\beta; L^2(\Omega))$ can be derived in a periodic setting if $g \in \mathrm{Lip}(\beta; L^2(\Omega))$ using that for minimizers $u, u' \in BV(\Omega) \cap L^2(\Omega)$ of E subject to the data $g, g' \in L^2(\Omega)$ we have $\|u-u'\|_{L^2(\Omega)} \le \|g-g'\|_{L^2(\Omega)}$ and that for $g' = g(\cdot - y)$ we have $u' = u(\cdot - y)$. This follows from the identities $-\operatorname{div} p_j + \alpha(u_j - g_j) = 0$, $p_j \in \partial|\nabla u_j|$, and $\langle p_1 - p_2, \nabla(u_1 - u_2)\rangle \ge 0$.

(ii) If $g \in L^\infty(\Omega)$, then by convexity we see that $u \in L^\infty(\Omega) \cap BV(\Omega)$, which implies for $d = 2$ that $u \in \mathrm{Lip}(1/2; L^2(\Omega))$ (cf. [DJL92]); i.e., $\|u - u_h\| \le ch^{1/6}$. If $d = 3$, we may choose $\varepsilon = h^{1/(\beta+2)}$.

We next state optimality conditions for the minimization of E in $\mathcal{S}^1(\mathcal{T}_h)$.

LEMMA 3.1 (optimality). *The function $u_h \in \mathcal{S}^1(\mathcal{T}_h)$ minimizes E in $\mathcal{S}^1(\mathcal{T}_h)$ if and only if there exists $p_h \in \mathcal{L}^0(\mathcal{T}_h)^d$ with $|p_h| \le 1$ in Ω such that*

$$(p_h, \nabla v_h) = -\alpha(u_h - g, v_h), \quad (\nabla u_h, q_h - p_h) \le 0$$

for all $(v_h, q_h) \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{L}^0(\mathcal{T}_h)^d$ with $|q_h| \le 1$ in Ω.

Proof. We have

$$\inf_{v_h \in \mathcal{S}^1(\mathcal{T}_h)} E(v_h) = \inf_{v_h \in \mathcal{S}^1(\mathcal{T}_h)} \sup_{q_h \in \mathcal{L}^0(\mathcal{T}_h)^d} \int_\Omega \nabla v_h \cdot q_h\,\mathrm{d}x + (\alpha/2)\|v_h - g\|^2_{L^2(\Omega)} - I_B(q_h),$$

where I_B is the indicator functional of $B = \{q \in L^1(\Omega)^d : |q| \le 1\}$. The existence of a saddle point $(u_h, p_h) \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{L}^0(\mathcal{T}_h)^d$ follows from the fact that the Lagrangian is a closed, proper, convex-concave function and these are the solutions of the Kuhn–Tucker conditions stated in the lemma; cf. [Roc97] for details. ∎

The optimality conditions depend on the choice of a scalar product on the space $\mathcal{L}^0(\mathcal{T}_h)^d$. For the choice of the L^2 scalar product we can formulate the discrete, regularized L^2 flow of the functional E as follows.

ALGORITHM $(P1)$. Let $\sigma > 0$, $(u_h^0, p_h^0) \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{L}^0(\mathcal{T}_h)^d$, set $d_t u_h^0 = 0$, and solve for $n = 0, 1, \dots$ with $\widetilde{u}_h^{n+1} = u_h^n + \tau d_t u_h^n$ the equations

$$(-\sigma d_t p_h^{n+1} + \nabla \widetilde{u}_h^{n+1}, q_h - p_h^{n+1}) \le 0, \quad (d_t u_h^{n+1}, v_h) + (p_h^{n+1}, \nabla v_h^{n+1}) = -\alpha(u_h^{n+1} - g, v_h)$$

subject to $|p_h^{n+1}| \le 1$ in Ω for all $(v_h, q_h) \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{L}^0(\mathcal{T}_h)^d$ with $|q_h| \le 1$ in Ω.

Remark 3.2. Notice that p_h^{n+1} is the unique minimizer of

$$q_h \mapsto \sigma \|q_h - p_h\|^2/(2\tau) - (q_h, \nabla \widetilde{u}_h^{n+1}) + I_B(q_h)$$

and is given by $p_h^{n+1} = (p_h^n + (\tau/\sigma)\nabla \widetilde{u}_h^{n+1})/\max\{1, |p_h^n + (\tau/\sigma) d_t \widetilde{u}_h^{n+1}|\}$.

The iterates of Algorithm $(P1)$ converge to a stationary point, e.g., if $\sigma = 1$ and $\tau \le c h_{\min}$. We denote $\|\nabla\| = \sup_{0 \ne v_h \in \mathcal{S}^1(\mathcal{T}_h)} \|\nabla v_h\|_{L^2(\Omega)}/\|v_h\|_{L^2(\Omega)} \le c h_{\min}^{-1}$.

PROPOSITION 3.1 (convergence). *Let $u_h \in \mathcal{S}^1(\mathcal{T}_h)$ be minimal for E in $\mathcal{S}^1(\mathcal{T}_h)$. If $\theta = \tau^2 \|\nabla\|^2/\sigma \le 1$, then the iterates of Algorithm $(P1)$ satisfy for every $N \ge 1$*

$$\tau \sum_{n=0}^{N} \Big((1-\theta)\frac{\tau}{2} \|d_t u_h^{n+1}\|_{L^2(\Omega)}^2 + \alpha \|u_h - u_h^{n+1}\|_{L^2(\Omega)}^2 \Big) \le C.$$

Proof. Let $p_h \in \mathcal{L}^0(\mathcal{T}_h)^d$ be as in Lemma 3.1. Upon choosing $v_h = u_h - u_h^{n+1}$ and $q_h = p_h$ in Algorithm $(P1)$ and $q_h = p_h^{n+1}$ in Lemma 3.1 and using

$$(u_h^{n+1} - g, u_h - u_h^{n+1}) + \|u_h - u_h^{n+1}\|_{L^2(\Omega)}^2 = (u_h - g, u_h - u_h^{n+1})$$

we find that

$$\begin{aligned}
&\frac{d_t}{2}\big(\|u_h - u_h^{n+1}\|_{L^2(\Omega)}^2 + \sigma \|p_h - p_h^{n+1}\|_{L^2(\Omega)}^2\big) \\
&\quad + \frac{\tau}{2}\big(\|d_t u_h^{n+1}\|_{L^2(\Omega)}^2 + \sigma \|d_t p_h^{n+1}\|_{L^2(\Omega)}^2\big) + \alpha \|u_h - u_h^{n+1}\|_{L^2(\Omega)}^2 \\
&= -(d_t u_h^{n+1}, u_h - u_h^{n+1}) - \sigma (d_t p_h^{n+1}, p_h - p_h^{n+1}) + \alpha \|u_h - u_h^{n+1}\|_{L^2(\Omega)}^2 \\
&\le (p_h^{n+1}, \nabla(u_h - u_h^{n+1})) + \alpha (u_h^{n+1} - g, u_h - u_h^{n+1}) - (p_h - p_h^{n+1}, \nabla \widetilde{u}_h^{n+1}) \\
&\quad + \alpha \|u_h - u_h^{n+1}\|_{L^2(\Omega)}^2 \\
&= (p_h^{n+1}, \nabla(u_h - u_h^{n+1})) - (p_h - p_h^{n+1}, \nabla \widetilde{u}_h^{n+1}) + \alpha (u_h - g, u_h - u_h^{n+1}) \\
&= (p_h^{n+1}, \nabla(u_h - u_h^{n+1})) - (p_h - p_h^{n+1}, \nabla \widetilde{u}_h^{n+1}) - (p_h, \nabla(u_h - u_h^{n+1})) \\
&= (p_h - p_h^{n+1}, \nabla(u_h^{n+1} - \widetilde{u}_h^{n+1})) + (p_h^{n+1} - p_h, \nabla u_h) \\
&\le (p_h - p_h^{n+1}, \nabla(u_h^{n+1} - \widetilde{u}_h^{n+1})) = \tau^2 (p_h - p_h^{n+1}, \nabla d_t^2 u_h^{n+1}),
\end{aligned}$$

where we used $u_h^{n+1} - \widetilde{u}_h^{n+1} = \tau^2 d_t^2 u_h^{n+1}$. Multiplication by τ, summation over $n = 0, \dots, N$, discrete integration by parts, Young's inequality, and $d_t u_h^0 = 0$ show that for the right-hand side we have

$$\begin{aligned}
\tau^3 \sum_{n=0}^{N} (p_h - p_h^{n+1}, \nabla d_t^2 u_h^{n+1}) &= \tau^3 \sum_{n=0}^{N} (d_t p_h^{n+1}, \nabla d_t u_h^n) + \tau^2 (p_h - p_h^n, \nabla d_t u_h^n)\big|_{n=0}^{N+1} \\
&\le \frac{\tau^2}{2} \left(\sum_{n=0}^{N} \frac{\tau^2}{\sigma} \|\nabla d_t u_h^n\|_{L^2(\Omega)}^2 + \sigma \|d_t p_h^{n+1}\|_{L^2(\Omega)}^2 \right) \\
&\quad + \frac{\sigma}{2} \|p_h - p_h^{N+1}\|_{L^2(\Omega)}^2 + \frac{\tau^4}{2\sigma} \|\nabla d_t u_h^{N+1}\|_{L^2(\Omega)}^2.
\end{aligned}$$

A combination of the estimates proves the theorem. ∎

Remark 3.3. (i) Notice that we cannot expect convergence $p_h^n \to p_h$ since p_h is not unique in general.

(ii) Algorithm $(P1)$ may not be as simple if rectangles or parallelepipeds are used instead of triangles or tetrahedra owing to the fact that gradients are no longer piecewise constant.

4. Piecewise constant approximation. We let $\mathcal{S}_h$ denote the set of interior sides of elements, i.e., all $(d-1)$-dimensional simplices S with $S = T_+ \cap T_-$ for $T_+, T_- \in \mathcal{T}_h$. The surface measure of S is denoted by $|S|$, and for each $S \in \mathcal{S}_h$ we choose a unit normal $\nu_S \in \mathbb{R}^d$ to S. For $v_h \in \mathcal{L}^0(\mathcal{T}_h)$ we define the jump $[v_h]_S \in \mathbb{R}$ across S by

$$[v_h]_S = v_h|_{T_+} - v_h|_{T_-}$$

if ν_S points from T_- into T_+.

LEMMA 4.1 (consistency). *For $v_h \in \mathcal{L}^0(\mathcal{T}_h)$ we have*

$$\|Dv_h\| = \sum_{S\in\mathcal{S}_h} |S|\big|[v_h]_S\big| = \sup_{(\alpha_S)_{S\in\mathcal{S}_h}, |\alpha_S|\le 1} \sum_{S\in\mathcal{S}_h} |S|\alpha_S[v_h]_S.$$

Proof. For every $p \in C_0^\infty(\Omega;\mathbb{R}^d)$ an elementwise integration by parts and a rearrangement of the sum over boundaries of elements shows

$$\int_\Omega v_h \operatorname{div} p \,\mathrm{d}x = \sum_{T\in\mathcal{T}_h} \int_{\partial T} v_h p\cdot\nu_T \,\mathrm{d}s = \sum_{S\in\mathcal{S}_h} [v_h]_S \int_S p\cdot\nu_S \,\mathrm{d}s.$$

A maximization of this identity over $p \in C_0^\infty(\Omega;\mathbb{R}^d)$ with $|p| \le 1$ in Ω proves the first identity. The second identity is straightforward to verify. ∎

We equip the space $\mathcal{L}^0(\mathcal{S}_h)$ of piecewise constant functions on $\mathcal{S}_h$ with the inner product

$$(p_h, q_h)_h = \sum_{S\in\mathcal{S}_h} |S|^{d/(d-1)} p_h|_S q_h|_S,$$

and define the discrete gradient $\nabla_h : \mathcal{L}^0(\mathcal{T}_h) \to \mathcal{L}^0(\mathcal{S}_h)$ by $\nabla_h v_h = [v_h]_S/|S|^{1/(d-1)}$. This leads to the following algorithm.

ALGORITHM $(P0)$. Let $\sigma > 0$, $(u_h^0, p_h^0) \in \mathcal{L}^0(\mathcal{T}_h) \times \mathcal{L}^0(\mathcal{S}_h)$, set $d_t u_h^0 = 0$, and solve for $n = 0, 1, \dots$ with $\widetilde{u}_h^{n+1} = u_h^n + \tau d_t u_h^n$ the equations

$$\begin{aligned} (-\sigma d_t p_h^{n+1} + \nabla_h \widetilde{u}_h^{n+1}, q_h - p_h^{n+1})_h &\le 0, \\ (d_t u_h^{n+1}, v_h) + (p_h^{n+1}, \nabla_h v_h^{n+1})_h &= -\alpha(u_h^{n+1} - g, v_h) \end{aligned}$$

for all $(v_h, q_h) \in \mathcal{L}^0(\mathcal{T}_h) \times \mathcal{L}^0(\mathcal{S}_h)^d$ with $\big||q_h|_S\big| \le 1$ for all $S \in \mathcal{S}_h$ and subject to $\big|p_h|_S\big| \le 1$ for all $S \in \mathcal{S}_h$.

The proof of the following proposition follows the lines of the proof of Proposition 3.1.

PROPOSITION 4.1 (convergence). *If $\tau \le c\sigma h$ then the iteration of Algorithm $(P0)$ converges to the minimizer $u_h \in \mathcal{L}^0(\mathcal{T}_h)$ of E restricted to $\mathcal{L}^0(\mathcal{T}_h)$.*

Remark 4.1. (i) The algorithm can also be derived employing the lowest order Raviart–Thomas finite element space $\mathcal{RT}_0(\mathcal{T}_h)$ equipped with the scalar product $(p_h, q_h)_h = \sum_{S\in\mathcal{S}_h} |S|^{d/(d-1)} (p_h(x_S)\cdot\nu_S)(q_h(x_S)\cdot\nu_S)$, where x_S is the midpoint of the

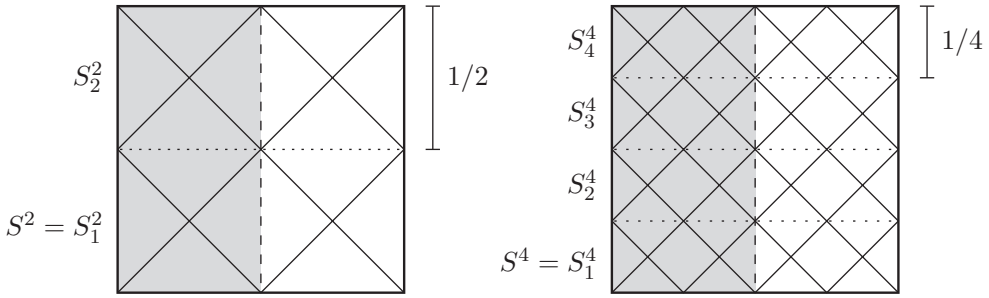

FIG. 1. *Construction of a sequence of triangulations* $(n = 2, 4)$ *on which piecewise constant finite element functions are not dense in* $BV(\Omega)$ *with respect to strict convergence. The employed function is the characteristic function of the gray-shaded region. Triangles are only used at the boundary of* $\Omega = (-1/2, 1/2) \times (0, 1)$.

side $S \in \mathcal{S}_h$. The space $\mathcal{RT}_0(\mathcal{T}_h)$ can be identified with $\mathcal{L}^0(\mathcal{S}_h)$ via the isomorphism $p_h \to \big(p_h(x_S) \cdot \nu_S\big)_{S \in \mathcal{S}_h}$. In view of the identity $\int_\Omega v_h \operatorname{div} p \, \mathrm{d}x = \int_\Omega v_h \operatorname{div} I_F p \, \mathrm{d}x$ for $v_h \in \mathcal{L}^0(\mathcal{T}_h)$ and $p \in C_0^\infty(\Omega; \mathbb{R}^d)$ and the Fortin operator $I_F : C_0^\infty(\Omega; \mathbb{R}^d) \to \mathcal{RT}_0(\mathcal{T}_h)$ this seems natural.

(ii) The choice of the inner product on $\mathcal{L}^0(\mathcal{S}_h)$ may be regarded as a lumped version of the L^2 inner product restricted to $\mathcal{RT}_0(\mathcal{T}_h)$ and leads to an efficient algorithm since (for $\sigma = 1$) we have $p_h^{n+1} = (p_h^n + \tau \nabla_h \widetilde{u}_h^{n+1}) / \max\{1, |p_h^n + \tau \nabla_h \widetilde{u}_h^{n+1}|\}$ on every side $S \in \mathcal{S}_h$. The discrete L^2 flow may not, however, have a meaningful limit as $h \to 0$.

The following example shows that in general we cannot expect that the approximation with piecewise constant finite elements converges to the right solution.

Example 4.1 (nonapproximation). Let $\Omega = (-1/2, 1/2) \times (0, 1)$ and let $u \in BV(\Omega)$ be defined by $u(x) = 1$ for $x \in (-1/2, 0) \times (0, 1)$ and $u(x) = 0$ for $x \in (0, 1/2) \times (0, 1)$. For each $n \geq 1$ let $\mathcal{T}_n$ be a triangulation of Ω into squares and triangles of diameter $h_n = 1/n$ (and edge lengths $h_n/\sqrt{2}$) as shown in Figure 1. Then there is no sequence $(u_n)_{n \in \mathbb{N}} \subset L^1(\Omega)$ with $u_n \in \mathcal{L}^0(\mathcal{T}_n)$ for all $n \in \mathbb{N}$ such that $u_n \to u$ in $L^1(\Omega)$ and $\|Du_n\| \to \|Du\| = 1$ as $n \to \infty$.

Proof. Let $(u_n)_{n \in \mathbb{N}}$ be a sequence with $u_n \in \mathcal{L}^0(\mathcal{T}_n)$ such that $\|u_n - u\|_{L^1(\Omega)} \to 0$ and $\|Du_n\| \leq c$ for all $n \in \mathbb{N}$. Let $n \in \mathbb{N}$. For $j = 1, \ldots, n$ we define

$$S_j^n = \{(x, y) \in \Omega : (j-1)/n < y < j/n\}$$

and set $S^n = S_1^n$. Let $\overline{u}_n \in L^1(S^n)$ be the average of u_n over all strips, i.e., for $(x, y) \in S^n$ set

$$\overline{u}_n(x, y) = \frac{1}{n} \sum_{j=1}^{n} u_n(x, y + j/n),$$

and reflect $\overline{u}_n$ across the x-axis, i.e., $\overline{u}_n(x, -y) = \overline{u}_n(x, y)$ for $(x, y) \in S^n$. We then define $\widetilde{u}_n \in L^1(\Omega)$ by periodically extending $\overline{u}_n$ with period $2/n$ in the y-direction. Then $\widetilde{u}_n \in L^1(\Omega)$ is continuous across the interfaces $\overline{S}_j^n \cap \overline{S}_{j+1}^n$ for $j = 1, \ldots, n-1$, and we have $\|\widetilde{u}_n - u\|_{L^1(S_j^n)} = \|\overline{u}_n - u\|_{L^1(S^n)}$ and $|D\widetilde{u}_n|(S_j^n) = |D\overline{u}_n|(S^n)$ for $j = 1, \ldots, n$, where $|D\overline{u}_n|(S^n)$ denotes the total variation of $D\overline{u}_n$ on S^n. Therefore,

$$\|D\widetilde{u}_n\| = n|D\overline{u}_n|(S^n) \leq \|Du_n\|, \quad \|\widetilde{u}_n - u\|_{L^1(\Omega)} = n\|\overline{u}_n - u\|_{L^1(S^n)} \leq \|u_n - u\|_{L^1(\Omega)}.$$

For every $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that $\|u_n - u\|_{L^1(\Omega)} < \varepsilon$ for all $n \geq N$, i.e.,

$$\|\overline{u}_n - u\|_{L^1(S^n)} < \varepsilon/n.$$

For each $n \geq N$ there exist squares $T_\pm \in \mathcal{T}_n$ with $T_\pm \subset S^n$ with $\overline{u}_n|_{T_+} \geq 1 - 4\varepsilon$ and $u_n|_{T_-} \leq 4\varepsilon$, since otherwise we have $\|\overline{u}_n - u\|_{L^1(S^n)} \geq \varepsilon/n$. The triangle inequality implies that $|D\overline{u}_n|(S^n) \geq (1 - 8\varepsilon)\sqrt{2}/n$ and $\|Du_n\| \geq \|D\widetilde{u}_n\| \geq (1 - 8\varepsilon)\sqrt{2}$ for all $n \geq N$, i.e., we have $\|Du_n\| \not\to 1$ as $n \to \infty$. ∎

Remark 4.2. The example is easily modified to prove a similar statement when only triangles are used, e.g., by dividing all elements in the above example along the direction $(1, 0)$.

5. Minimization with H^{-1} fidelity term. Let $H^1(\Omega)'$ denote the dual of $H^1(\Omega)$ and $\langle \eta, v\rangle$ the duality pairing for $\eta \in H^1(\Omega)'$ and $v \in H^1(\Omega)$ such that $\langle \eta, v\rangle = \int_\Omega \eta v \,\mathrm{d}x$ if $\eta \in L^{6/5}(\Omega)$. For $\eta \in H^1(\Omega)'$ with $\langle \eta, 1\rangle = 0$ the inverse of the negative Laplace operator subject to homogeneous Neumann boundary conditions applied to η is the function $\mathcal{G}\eta \in H^1(\Omega)$ with $\int_\Omega \mathcal{G}\eta \,\mathrm{d}x = 0$ such that

$$\big(\nabla \mathcal{G}\eta, \nabla\varphi\big) = \langle \eta, \varphi\rangle$$

for all $\varphi \in H^1(\Omega)$ with $(\varphi, 1) = 0$, and we set $\|\eta\|_{-1} = \|\nabla \mathcal{G}\eta\|_{L^2(\Omega)}$. Given $\lambda > 0$ and $g \in H^1(\Omega)'$ with $\langle g, 1\rangle = 0$ we consider the functional

$$E'(v) = \|Dv\| + (\lambda/2)\|v - g\|_{-1}^2$$

defined for $v \in BV(\Omega)$ with $\int_\Omega v \,\mathrm{d}x = 0$. Notice that owing to the continuous embeddings $BV(\Omega) \to L^{3/2}(\Omega)$ and $L^{6/5}(\Omega) \to H^1(\Omega)'$ the functional is well defined for $d \leq 3$. The discrete inverse Laplacian $\mathcal{G}_h\eta \in \mathring{\mathcal{S}}^1(\mathcal{T}_h) = \big\{v_h \in \mathcal{S}^1(\mathcal{T}_h) : \int_\Omega v_h \,\mathrm{d}x = 0\big\}$ is for $\eta \in H^1(\Omega)'$ with $\langle \eta, 1\rangle = 0$ defined by

$$\big(\nabla \mathcal{G}_h\eta, \nabla\varphi_h\big) = \langle \eta, \varphi_h\rangle$$

for all $\varphi_h \in \mathring{\mathcal{S}}^1(\mathcal{T}_h)$, and we set $\|\eta\|_{-1,h} = \|\nabla \mathcal{G}_h\eta\|_{L^2(\Omega)}$. We let $P_h : H^1(\Omega)' \to \mathring{\mathcal{S}}^1(\mathcal{T}_h)$ denote the L^2 projection onto $\mathring{\mathcal{S}}^1(\mathcal{T}_h)$ and define $g_h = P_h g$. We have that $\|P_h\eta\|_{-1,h} = \|\eta\|_{-1,h} \leq \|\eta\|_{-1}$ for all $\eta \in H^1(\Omega)'$ with $\langle \eta, 1\rangle = 0$. The discrete problem consists in the minimization of

$$E_h'(v_h) = \|Dv_h\| + (\lambda/2)\|v_h - g_h\|_{-1,h}^2$$

among $v_h \in \mathring{\mathcal{S}}^1(\mathcal{T}_h)$.

THEOREM 5.1 (approximation). *Assume $d = 2$, $g \in L^2(\Omega)$, and that there exists $\gamma > 0$ such that $\|[\mathcal{G} - \mathcal{G}_h]\eta\|_{L^2(\Omega)} \leq ch^\gamma \|\eta\|_{-1}$ for all $\eta \in H^1(\Omega)'$ with $\langle \eta, 1\rangle = 0$. Then $u_h \to u$ in $L^p(\Omega)$ for every $1 \leq p < 2$ as $h \to 0$.*

Proof. Owing to the strict convexity of E' and $0 \in \partial E'(u)$, we have

$$(\lambda/2)\|u - u_h\|_{-1}^2 \leq E'(u_h) - E'(u) = E_h'(u_h) - E'(u) + E'(u_h) - E_h'(u_h).$$

For $\delta > 0$ let $u_\varepsilon \in C^\infty(\Omega) \cap BV(\Omega)$ be as in section 2.4 and $\mathring{\mathcal{I}}_h u_\varepsilon = \mathcal{I}_h u_\varepsilon - \int_\Omega \mathcal{I}_h u_\varepsilon \,\mathrm{d}x / \int_\Omega 1 \,\mathrm{d}x$. Since u_h is minimal for E_h' in $\mathring{\mathcal{S}}^1(\mathcal{T}_h)$ and since $E_h'(v_h) \leq E'(v_h)$ for all $v_h \in \mathring{\mathcal{S}}^1(\mathcal{T}_h)$ we have that

$$\begin{aligned}
&(\lambda/2)\|u - u_h\|_{-1}^2 \\
&\quad \leq E'(\mathring{\mathcal{I}}_h u_\varepsilon) - E'(u) + E'(u_h) - E_h'(u_h) \\
&\quad = \|\nabla \mathcal{I}_h u_\varepsilon\|_{L^1(\Omega)} - \|\nabla u_\varepsilon\|_{L^1(\Omega)} + c_0\delta \\
&\qquad + (\lambda/2)\big(\|\mathring{\mathcal{I}}_h u_\varepsilon - g\|_{-1}^2 - \|u - g\|_{-1}^2 + \|u_h - g\|_{-1}^2 - \|u_h - g_h\|_{-1,h}^2\big).
\end{aligned}$$

The difference of the first two terms on the right-hand side is bounded by ch/ε, and it remains to control the terms inside the brackets. Using Galerkin orthogonality and the definition of $\mathcal{G}$ we find

$$\begin{aligned}\|u_h-g\|_{-1}^2-\|u_h-g_h\|_{-1,h}^2&=\big(\nabla[\mathcal{G}-\mathcal{G}_h](u_h-g),\nabla[\mathcal{G}+\mathcal{G}_h](u_h-g)\big)\\&=\big(\nabla[\mathcal{G}-\mathcal{G}_h](u_h-g),\nabla\mathcal{G}(u_h-g)\big)\\&=\big([\mathcal{G}-\mathcal{G}_h](u_h-g),u_h-g\big)\\&\le\|[\mathcal{G}-\mathcal{G}_h](u_h-g)\|_{L^2(\Omega)}\|u_h-g\|_{L^2(\Omega)}\\&\le ch^\gamma\|u_h-g\|_{L^2(\Omega)}^2,\end{aligned}$$

and the right-hand side vanishes as $h\to 0$ owing to continuity of the embedding $BV(\Omega)\to L^2(\Omega)$. Finally, we note that

$$\|\mathring{\mathcal{I}}_h u_\varepsilon-g\|_{-1}^2-\|u-g\|_{-1}^2=\|\mathring{\mathcal{I}}_h u_\varepsilon-u\|_{-1}\|\mathring{\mathcal{I}}_h u_\varepsilon+u+2g\|_{-1},$$

and it is straightforward to show that this expression converges to 0 as $h\to 0$. The boundedness of $(u_h)_{h>0}$ in $BV(\Omega)$ and the compactness of the embedding $BV(\Omega)\to L^p(\Omega)$ prove the assertion. ∎

Remark 5.1. (i) If Ω is convex, then the assumption of the theorem follows from the estimate $\|[\mathcal{G}-\mathcal{G}_h]\eta\|_{L^2(\Omega)}\le ch^2\|\eta\|_{L^2(\Omega)}$, i.e., $\gamma=2$.

(ii) If $d=3$, then it is not clear how to control the term $\|\nabla(\mathcal{I}_h u_\varepsilon-u_\varepsilon)\|_{L^1(\Omega)}$.

We rewrite the minimization of E_h on $\mathring{\mathcal{S}}^1(\mathcal{T}_h)$ as the saddle-point formulation

$$\inf_{v_h\in\mathring{\mathcal{S}}^1(\mathcal{T}_h)}\sup_{q_h\in\mathcal{L}^0(\mathcal{T}_h)^d}\int_\Omega\nabla v_h\cdot q_h\,\mathrm{d}x+(\lambda/2)\|u_h-g_h\|_{-1,h}^2-I_B(q_h).$$

To define the iterative solution procedure we choose the scalar product $(\nabla\mathcal{G}_h\cdot,\nabla\mathcal{G}_h\cdot)$ on $\mathring{\mathcal{S}}^1(\mathcal{T}_h)$ and the L^2 inner product $(\cdot,\cdot)$ on $\mathcal{L}^0(\mathcal{T}_h)^d$. A discrete, regularized H^{-1} flow for E_h' then reads with $\widetilde{u}_h^{n+1}=u_h^n+\tau d_t u_h^n$ and $\sigma>0$,

$$\begin{aligned}-\sigma(d_t p_h^{n+1},q_h-p_h^{n+1})+(\nabla\widetilde{u}_h^{n+1},q_h-p_h^{n+1})&\le I_B(q_h)-I_B(p_h^{n+1}),\\(\nabla\mathcal{G}_h d_t u_h^{n+1},\nabla\mathcal{G}_h v_h)+(p_h^{n+1},\nabla v_h)&=-\lambda(\nabla\mathcal{G}_h(u_h^{n+1}-g_h),\nabla\mathcal{G}_h v_h)\end{aligned}$$

for all $q_h\in\mathcal{L}^0(\mathcal{T}_h)^d$ and $v_h\in\mathring{\mathcal{S}}^1(\mathcal{T}_h)$. Upon introducing $\varphi_h=\mathcal{G}_h(d_t u_h^{n+1}+\lambda u_h^{n+1}-\lambda g_h)$ we obtain the following algorithm.

Algorithm (H^{-1}-$P1$). Let $\sigma>0$, $(u_h^0,p_h^0)\in\mathring{\mathcal{S}}^1(\mathcal{T}_h)\times\mathcal{L}^0(\mathcal{T}_h)^d$, set $d_t u_h^0=0$, and compute for $n=0,1,\dots$ the pair $(\varphi_h,u_h^{n+1})\in\mathring{\mathcal{S}}^1(\mathcal{T}_h)^2$ satisfying

$$\begin{aligned}(\nabla\varphi_h,\nabla\xi_h)-(d_t u_h^{n+1}+\lambda u_h^{n+1},\xi_h)&=-\lambda(g_h,\xi_h),\\-(\varphi_h,v_h)&=(p_h^{n+1},\nabla v_h)\end{aligned}$$

for all $(\xi_h,v_h)\in\mathring{\mathcal{S}}^1(\mathcal{T}_h)^2$, where for $\widetilde{u}_h^{n+1}=u_h^n+\tau d_t u_h^n$ we have

$$p_h^{n+1}=\frac{p_h^n+(\tau/\sigma)\nabla\widetilde{u}_h^{n+1}}{\max\{1,|p_h^n+(\tau/\sigma)\nabla\widetilde{u}_h^{n+1}|\}}.$$

Proposition 5.1 (convergence). *If $\tau\le c\sigma h_{\min}^2$, then the iteration of Algorithm (H^{-1}-$P1$) converges to the unique minimizer $u_h\in\mathring{\mathcal{S}}^1(\mathcal{T}_h)$ of E_h'.*

Proof. The proof follows along the lines of the proof of Proposition 3.1 using different inner products. The estimate $\|\nabla v_h\|\le ch_{\min}^{-2}\|v_h\|_{-1,h}$ for all $v_h\in\mathring{\mathcal{S}}^1(\mathcal{T}_h)$ follows from the spectral properties of the discrete Laplace operator and leads to the condition on the time-step size. ∎

Remark 5.2. The linear system of equations in Algorithm (H^{-1}-$P1$) can be split into two linear systems of equations which can be solved subsequently and only require the inversion of an augmented (lumped) mass matrix (incorporating the zero-mean constraint).

6. Numerical experiments. We tested and compared the practical performance of Algorithms ($P1$), ($P0$), and (H^{-1}-$P1$) with the following two-dimensional example. We use the parameter $p = 2$ unless stated otherwise.

Example 6.1. Let $d = 2$, $\Omega = (-0.5, 0.5)^2$, $\alpha = 1.0 \times 10^2$, $\lambda = 1.0 \times 10^5$, and $g = g_0 + \xi_h$ for $g_0 = \chi_{B_r^p(0)}$ for $B_r^p(0) = \{x \in \mathbb{R}^2 : |x|_{\ell^p} < r\}$ with $r = 0.2$, $p \in [1, \infty]$, and a mesh-dependent perturbation function ξ_h.

6.1. Choice of data. Given a triangulation $\mathcal{T}_h$ of Ω we define the function $\xi_h \in \mathcal{L}^0(\mathcal{T}_h)$ by associating to each element $T \in \mathcal{T}_h$ a random value ξ_T sampled from a Gaussian distribution with mean 0 and standard deviation 1 and then subtracted the integral mean of the resulting function. The employed function $g_h \in \mathcal{L}^0(\mathcal{T}_h)$ was defined by setting $g_h|_T = g_0(x_T) + \xi_h|_T$ for all $T \in \mathcal{T}_h$, where x_T denotes the midpoint of the triangle T. For Algorithm (H^{-1}-$P1$) this function was after subtraction of the mean of g_0 projected onto $\mathring{\mathcal{S}}^1(\mathcal{T}_h)$. The projection of the function g_h onto the respective finite element spaces was used to define the initial u_h^0 unless otherwise stated. Throughout this section we employ $\sigma = 1$.

6.2. Stopping criteria. We stopped the iteration of Algorithms ($P1$), ($P0$) and Algorithm (H^{-1}-$P1$) when

$$\frac{\|d_t u_h^n\|^2_{L^2(\Omega)}}{E(0)} \le \varepsilon_{stop} = 1.0 \times 10^{-6}, \qquad \frac{\|d_t u_h^n\|^2_{-1,h}}{E_h'(0)} \le \varepsilon_{stop} = 1.0 \times 10^{-6},$$

respectively. Other useful stopping criteria can be formulated in terms of the duality gap.

6.3. Comparison. Figure 2 displays the outputs of Algorithms ($P0$), ($P0$), and (H^{-1}-$P1$) obtained on a uniform triangulation consisting of 2048 triangles which are halved squares with edge length $\widehat{h} = 2^{-5}$. The time-step size was chosen as $\tau = \widehat{h}/10$ for Algorithms ($P1$) and ($P0$) and $\tau = (\widehat{h}/10)^2$ in the case of Algorithm (H^{-1}-$P1$). The iterates u_h^n for $n = 0, 10, 20, 40$ and n_{stop} are shown in Figure 3 from a different perspective. The number n_{stop} is the smallest integer for which the stopping criterion was satisfied; we have $n_{stop} = 149$, 112, 129 for Algorithms ($P0$), ($P1$), and (H^{-1}-$P1$), respectively. For a better comparison we added the mean $\pi/25$ of the function g_0 to the iterates obtained with Algorithm (H^{-1}-$P1$) in the plots. We observe sharp edges in the numerical solution obtained with piecewise constant functions (left column) and a transition region of width comparable to the mesh-size h in the case of piecewise affine, globally continuous functions for the functional involving the L^2 fidelity term (middle column). Notice that the circular edge of the original image is not accurately approximated in the case of piecewise constant finite elements. For the image denoising method with the H^{-1} fidelity we see in the output of the numerical experiments that certain oscillations are kept (right column). The scale of these oscillations is determined by the parameter λ.

6.4. Convergence of the energies. In the left plot of Figure 4 we displayed the energies $E(u_h^n)$ for the iterates u_h^n, $n = 0, 1, \ldots, n_{stop}$ of Algorithms ($P0$) and ($P1$). We see that the energy is monotonically decreasing in both cases. The curves approach a similar value. To illustrate that as $h \to 0$ the values may differ, we showed in the right

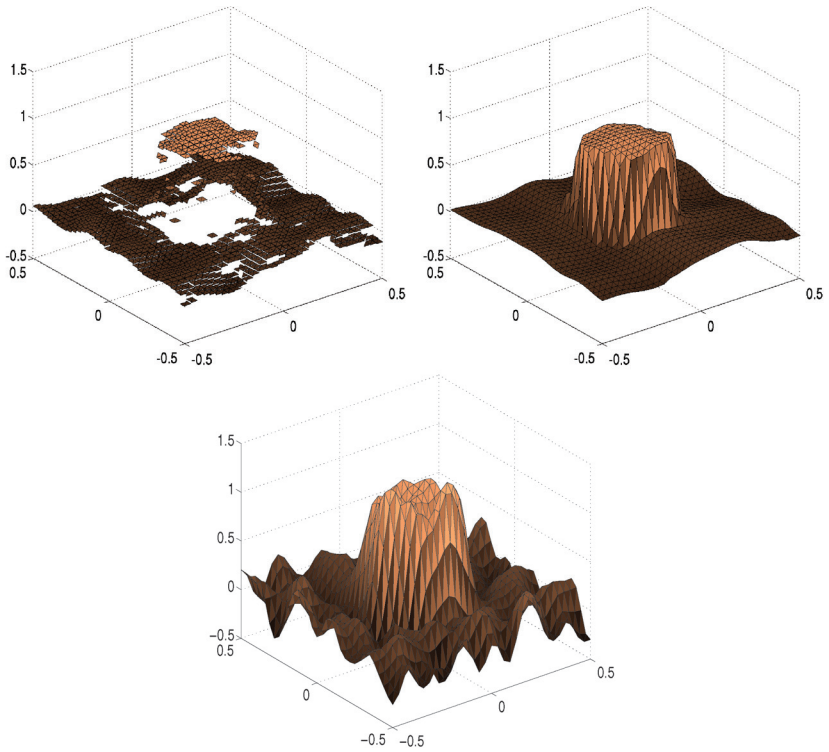

FIG. 2. *Output* $u_h^{n_{stop}}$ *of Algorithms* $(P0)$ *(top left),* $(P1)$ *(top right), and* $(H^{-1}\text{-}P1)$ *(bottom).*

plot of Figure 4 the final energies $E(u_h^{n_{stop}})$ for the outputs $u_h^{n_{stop}}$ of Algorithms $(P1)$ and $(P0)$ for a sequence of triangulations with mesh-size $\widehat{h} = 2^{-\ell}$ for $\ell = 5, 6, 7, 8, 9$. We observe that the different methods do not tend to the same value, and this is in agreement with Example 4.1, which showed that the approximation with piecewise constant functions cannot simultaneously converge to the exact solution and the exact energy.

6.5. Energy decay and iteration numbers. The behavior of the energies of the functional $E_h'(u_h^n)$ for the choice of initial data $u_h^0 = \mathring{g}_h$ and $u_h^0 = 0$ is illustrated in the plot of Figure 5. While for $u_h^0 = \mathring{g}_h$ the energy is monotonically decreasing, the energy initially grows for the choice $u_h^0 = 0$. This reflects the fact that the algorithm is not a gradient flow but a regularized one, and monotonicity of the energy depends on the choice of the discretization parameters.

In Table 1 we displayed the iteration numbers for Algorithms $(P0)$, $(P1)$, and $(H^{-1}\text{-}P1)$ when used on meshes with mesh-size $\widehat{h} = 2^{-\ell}$ for $\ell = 5, 6, 7, 8, 9$. We observe that the numbers of iterations required to satisfy the stopping criterion grow like $1/\tau$, i.e., proportional to the inverse of the mesh-size and the square of the mesh-size for the iterations involving the L^2 and H^{-1} fidelity terms, respectively. The direct solution of the saddle-point problems related to the constraint on the mean of the functions in the iteration of Algorithm $(H^{-1}\text{-}P1)$ did not allow us to test the algorithm on the mesh defined through $\widehat{h} = 2^{-9}$ owing to memory restrictions. It is expected that an Uzawa iteration for the iterative solution of the saddle-point problems avoids this effect. The CPU times for the iterations of Algorithms $(P0)$,

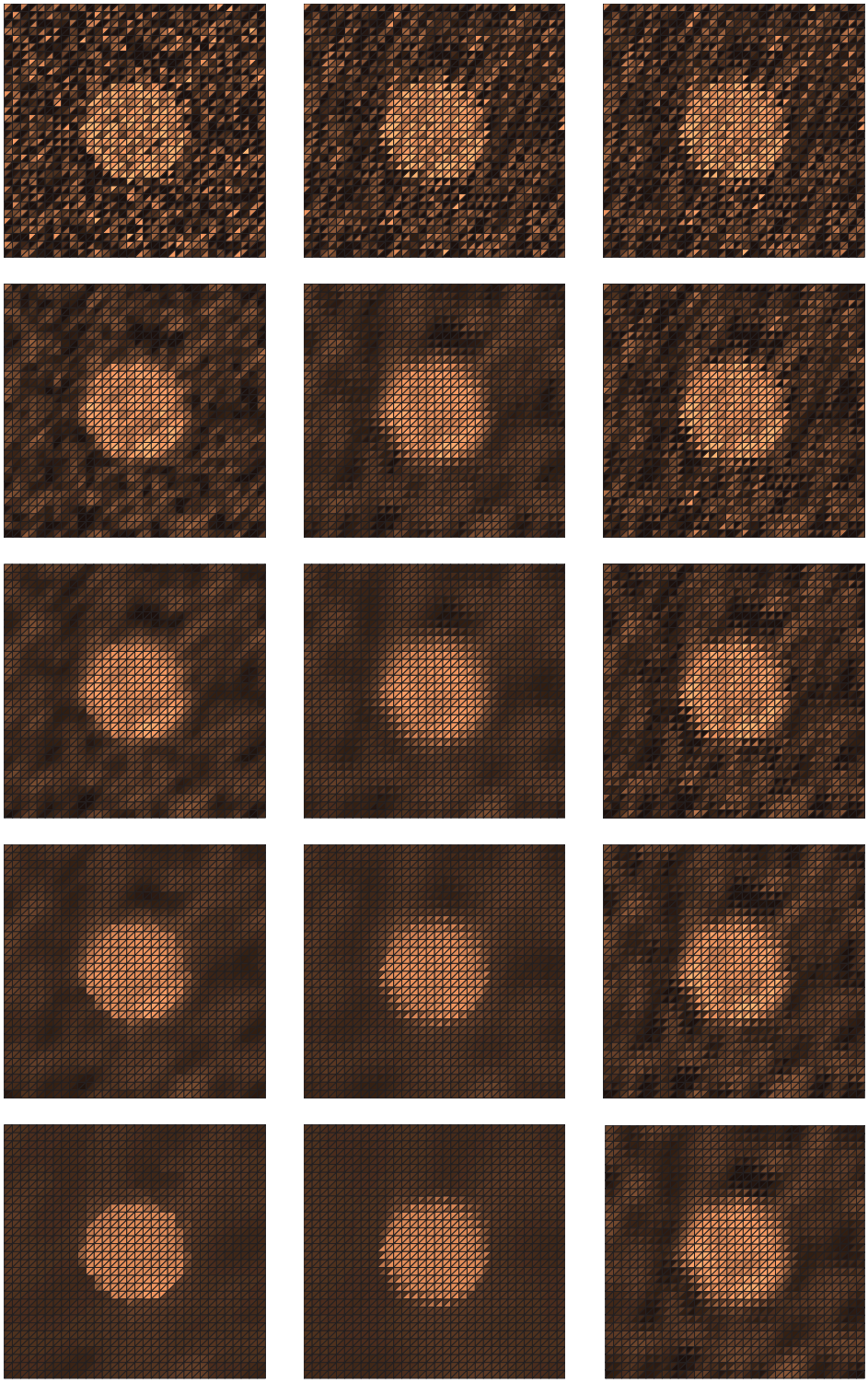

FIG. 3. *Iterates u_h^n obtained with Algorithms $(P0)$, $(P1)$, and $(H^{-1}\text{-}P1)$ (from left to right) for $n = 0, 10, 20, 40$, n_{stop} (from top to bottom), where n_{stop} is the number of iterations needed to satisfy the stopping criterion.*

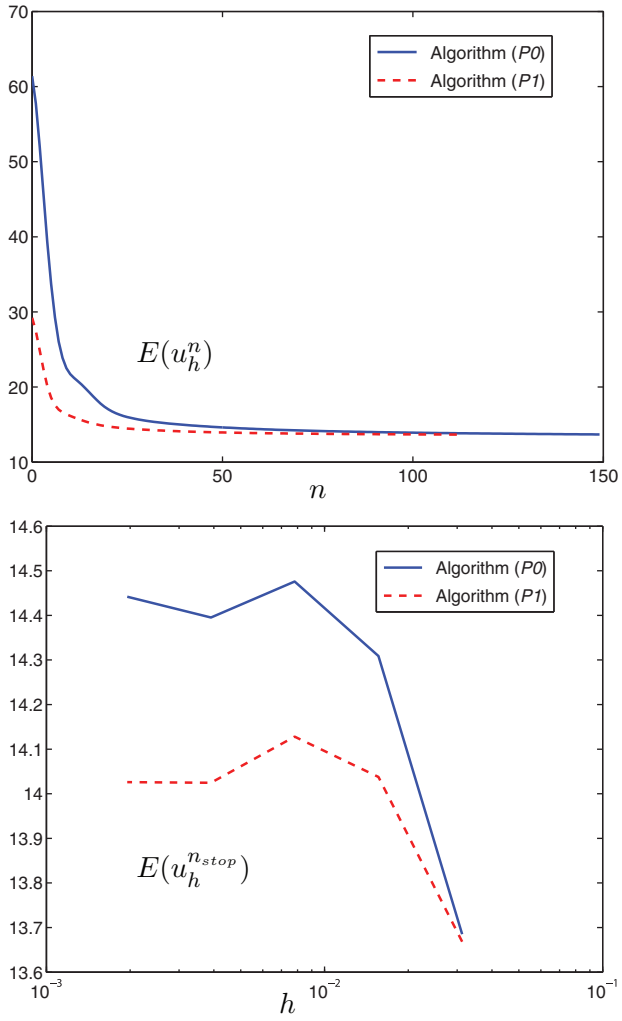

FIG. 4. *Energies* $E(u_h^n)$ *for the iterates* u_h^n, $n = 0, 1, \ldots, n_{stop}$ *of Algorithms* $(P0)$ *and* $(P1)$ *(top). Final energies* $E(u_h^{n_{stop}})$ *of the outputs of Algorithms* $(P0)$ *and* $(P1)$ *for* $\widehat{h} = 2^{-\ell}$, $\ell = 5, 6, 7, 8, 9$ *(bottom).*

$(P1)$, and $(H^{-1}$-$P1)$ were 0.8 s, 3.5 s, and 23.3 s for a mesh with 8192 triangles on a Dell Precision WorkStation T3500 (4 Intel Xeon CPU W3520, 2.67GHz, 6GB RAM) and an implementation in MATLAB.

6.6. Rectangular shapes. To experimentally study the approximation properties of the proposed finite element method for other shapes, we ran Algorithms $(P0)$, $(P1)$, and $(H^{-1}$-$P1)$ in Example 6.1 with $p = 1$ and $p = \infty$ on the uniform triangulation with edge length $\widehat{h} = 2^{-5}$. As above, the time-step size was chosen as $\tau = \widehat{h}/10$ for Algorithms $(P1)$ and $(P0)$ and $\tau = (\widehat{h}/10)^2$ in the case of Algorithm $(H^{-1}$-$P1)$. Figures 6 and 7 display the initial data and the output of the iterative scheme for $p = 1$ and $p = \infty$, respectively. We see that the results are similar to those obtained for the circular shape defined through $p = 2$.

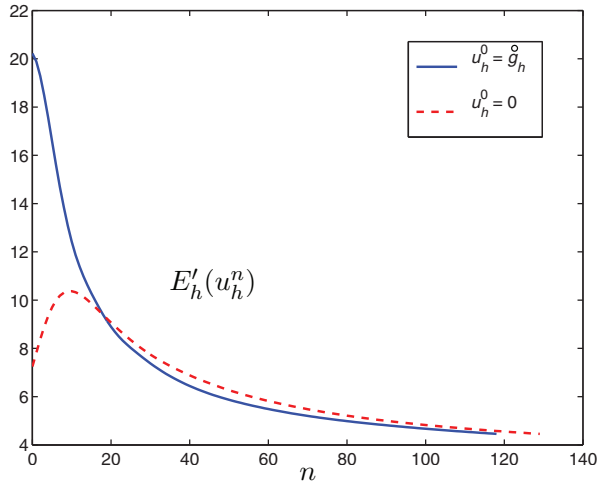

FIG. 5. *Energies $E_h'(u_h^n)$ for the iterates of Algorithms (H^{-1}-$P1$) for different choices of initial data u_h^0.*

TABLE 1
Iteration numbers needed to achieve the stopping criterion for Algorithms ($P0$), ($P1$), and (H^{-1}-$P1$). Notice that the perturbations are mesh-dependent.

$\widehat{h}$	($P0$)	($P1$)	(H^{-1}-$P1$)
2^{-5}	149	112	129
2^{-6}	196	158	249
2^{-7}	279	219	787
2^{-8}	457	350	2639
2^{-9}	750	557	–

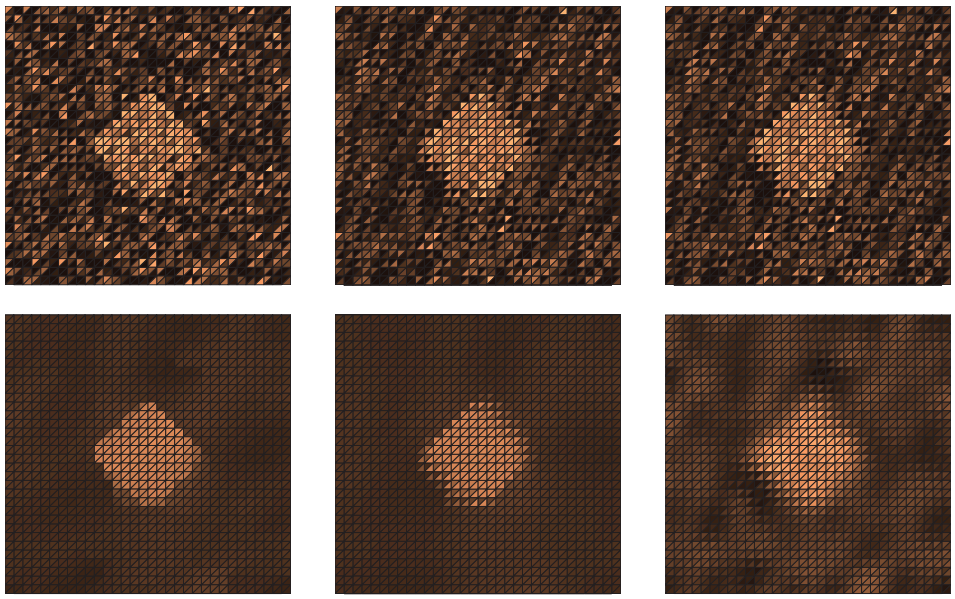

FIG. 6. *Initial function u_h^0 (top) and corresponding output $u_h^{n_{stop}}$ (bottom) obtained with Algorithms ($P0$), ($P1$), and (H^{-1}-$P1$) (from left to right) in Example 6.1 with $p = 1$, where $n_{stop} = 154, 123, 159$, respectively.*

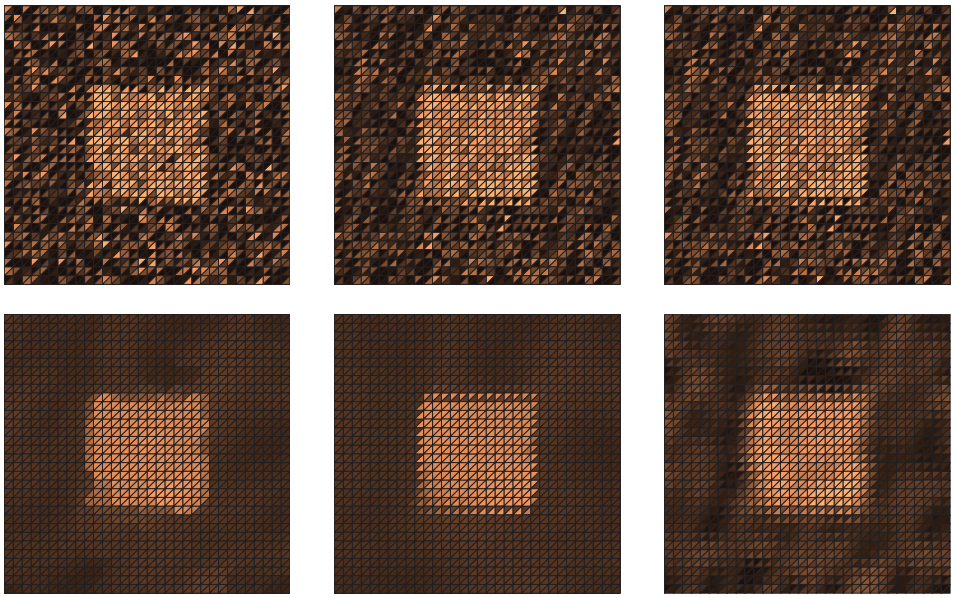

FIG. 7. *Initial function* u_h^0 *(top) and corresponding output* $u_h^{n_{stop}}$ *(bottom) obtained with Algorithms* $(P0)$*,* $(P1)$*, and* $(H^{-1}\text{-}P1)$ *(from left to right) in Example* 6.1 *with* $p = \infty$*, where* $n_{stop} = 144, 108, 103$*, respectively.*

Appendix A. Primal-dual algorithm. For Banach spaces X and Y, proper, convex, and lower semicontinuous functionals $G : X \to \mathbb{R}\cup\{+\infty\}$, $F : Y \to \mathbb{R}\cup\{+\infty\}$, and a bounded, linear operator $K : X \to Y$ we consider the saddle-point problem

$$\inf_{x\in X}\sup_{y'\in Y'}\langle Kx, y'\rangle - F^*(y') + G(x) = \inf_{x\in X}\sup_{y'\in Y'} L(x, y').$$

The related primal and dual problem consist in the minimization of the functionals

$$\mathcal{P}(x) = F(Kx) + G(x), \quad -\mathcal{D}(y') = G^*(-K'y') + F^*(y'),$$

respectively. We have $\mathcal{P}(x) - \mathcal{D}(y') \geq 0$ for all $(x, y') \in X \times Y'$ with equality if and only if (x, y') is a saddle point for L. We assume in the following that Y is a Hilbert space and identify Y and Y'. The equations $\partial_t x = -\partial_x L(x, y)$ and $\partial_t y = \partial_y L(x, y)$ motivate the following algorithm.

ALGORITHM (PD). Let $(x^0, y^0) \in X \times Y$, set $d_t x^0 = 0$, and solve for $n = 0, 1, \ldots$ the equations

$$\widetilde{x}^{n+1} = x^n + \tau d_t x^n, \quad -d_t y^{n+1} + K\widetilde{x}^{n+1} \in \partial F^*(y^{n+1}), \quad -d_t x^{n+1} - K'y^{n+1} \in \partial G(x^{n+1}).$$

Remark A.1. The equations in Algorithm (PD) are equivalent to the variational inequalities

$$\begin{aligned}\langle -d_t y^{n+1} + K\widetilde{x}^{n+1}, y - y^{n+1}\rangle &\leq F^*(y) - F^*(y^{n+1}),\\ \langle -d_t x^{n+1} - K'y^{n+1}, x - x^{n+1}\rangle &\leq G(x) - G(x^{n+1}) - (\alpha/2)\|x - x^{n+1}\|^2\end{aligned}$$

for all $(x, y) \in X \times Y$. Here, $\alpha > 0$ if G is uniformly convex.

THEOREM A.1 (convergence [CP11]). *Let $(\widehat{x},\widehat{y})$ be a saddle point for L. If $\tau\|K\|\le 1$, we have for $N\ge 0$*

$$\frac{1-\tau\|K\|}{2}\|\widehat{y}-y^{N+1}\|^2+\frac{1}{2}\|\widehat{x}-x^{N+1}\|^2+\tau\sum_{n=0}^{N}\frac{\alpha}{2}\|\widehat{x}-x^{n+1}\|^2\le\frac{1}{2}\|\widehat{y}-y^0\|^2+\frac{1}{2}\|\widehat{x}-x^0\|^2.$$

Proof. Using $\mathcal{P}(\widehat{x})-\mathcal{D}(\widehat{y})=0$ and $x^{n+1}-\widetilde{x}^{n+1}=\tau^2 d_t^2 x^{n+1}$ we have

$$\begin{aligned}
&\frac{d_t}{2}\big(\|\widehat{y}-y^{n+1}\|^2+\|\widehat{x}-x^{n+1}\|^2\big)+\frac{\tau}{2}\big(\|d_t x^{n+1}\|^2+\|d_t y^{n+1}\|^2\big)+(\alpha/2)\|\widehat{x}-x^{n+1}\|^2\\
&=-\langle d_t y^{n+1},\widehat{y}-y^{n+1}\rangle-\langle d_t x^{n+1},\widehat{x}-x^{n+1}\rangle+(\alpha/2)\|\widehat{x}-x^{n+1}\|^2\\
&\le F^*(\widehat{y})-F^*(y^{n+1})-\langle K\widetilde{x}^{n+1},\widehat{y}-y^{n+1}\rangle+G(\widehat{x})-G^*(x^{n+1})+\langle K' y^{n+1},\widehat{x}-x^{n+1}\rangle\\
&=\big[\langle K\widehat{x},y^{n+1}\rangle-F^*(y^{n+1})+G(\widehat{x})\big]-\big[\langle Kx^{n+1},\widehat{y}\rangle-F^*(\widehat{y})+G(x^{n+1})\big]\\
&\quad+\langle Kx^{n+1},\widehat{y}\rangle-\langle K\widetilde{x}^{n+1},\widehat{y}-y^{n+1}\rangle-\langle K' y^{n+1},x^{n+1}\rangle\\
&\le\mathcal{P}(\widehat{x})-\mathcal{D}(\widehat{y})+\langle K(x^{n+1}-\widetilde{x}^{n+1}),\widehat{y}-y^{n+1}\rangle=\tau^2\langle K d_t^2 x^{n+1},\widehat{y}-y^{n+1}\rangle.
\end{aligned}$$

Discrete integration by parts, $d_t x^0=0$, and Young's inequality show

$$\begin{aligned}
\tau^3\sum_{n=0}^{N}\langle K d_t^2 x^{n+1},\widehat{y}-y^{n+1}\rangle&=\tau^3\sum_{n=0}^{N}\langle K d_t x^n,d_t y^{n+1}\rangle+\tau^2\langle K d_t x^n,\widehat{y}-y^n\rangle|_{n=0}^{N+1}\\
&\le\frac{\tau^2}{2}\left(\sum_{n=0}^{N}\tau^2\|K d_t x^n\|^2+\|d_t y^{n+1}\|^2\right)\\
&\quad+\frac{\tau\|K\|}{2}\|\widehat{y}-y^{N+1}\|^2+\frac{\tau^3}{2\|K\|}\|K d_t x^{N+1}\|^2.
\end{aligned}$$

A combination of the estimates proves the theorem. ∎

Remark A.2. (i) Modifications of the algorithm allow us to prove unconditional stability but lead to unphysical evolutions or expensive problems in each step.

(ii) Acceleration methods are discussed in [CP11], and it is shown that uniform convexity of G is not necessary for convergence.

REFERENCES

[AFP00] 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.

[BKP10] K. BREDIES, K. KUNISCH, AND T. POCK, *Total generalized variation*, SIAM J. Imaging Sci., 3 (2010), pp. 492–526.

[BS08] S. C. BRENNER AND L. RIDGWAY SCOTT, *The Mathematical Theory of Finite Element Methods*, 3rd ed., Texts Appl. Math. 15, Springer, New York, 2008.

[CGM99] T. F. CHAN, G. H. GOLUB, AND P. MULET, *A nonlinear primal-dual method for total variation-based image restoration*, SIAM J. Sci. Comput., 20 (1999), pp. 1964–1977.

[Cha04] A. CHAMBOLLE, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vision, 20 (2004), pp. 89–97.

[CL97] A. CHAMBOLLE AND P.-L. LIONS, *Image recovery via total variation minimization and related problems*, Numer. Math., 76 (1997), pp. 167–188.

[CP11] A. CHAMBOLLE AND T. POCK, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision, 40 (2011), pp. 120–145.

[DJL92] R. A. DEVORE, B. JAWERTH, AND B. J. LUCIER, *Image compression through wavelet transform coding*, IEEE Trans. Inform. Theory, 38 (1992), pp. 719–746.

[DV97] D. C. Dobson and C. R. Vogel, *Convergence of an iterative method for total variation denoising*, SIAM J. Numer. Anal., 34 (1997), pp. 1779–1791.

[ES09] C. M. Elliott and 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., 29 (2009), pp. 651–689.

[FP03] X. Feng and A. Prohl, *Analysis of total variation flow and its finite element approximations*, M2AN Math. Model. Numer. Anal., 37 (2003), pp. 533–556.

[FvOP05] X. Feng, M. von Oehsen, and A. Prohl, *Rate of convergence of regularization procedures and finite element approximations for the total variation flow*, Numer. Math., 100 (2005), pp. 441–456.

[HK04] M. Hintermüller and K. Kunisch, *Total bounded variation regularization as a bilaterally constrained optimization problem*, SIAM J. Appl. Math., 64 (2004), pp. 1311–1333.

[Mey01] Y. Meyer, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*, Univ. Lecture Ser. 22, AMS, Providence, RI, 2001.

[OSV03] S. Osher, A. Solé, and L. Vese, *Image decomposition and restoration using total variation minimization and the H^{-1} norm*, Multiscale Model. Simul., 1 (2003), pp. 349–370.

[Roc76] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim., 14 (1976), pp. 877–898.

[Roc97] R. T. Rockafellar, *Convex Analysis*, Princeton Landmarks in Mathematics, Princeton University Press, Princeton, NJ, 1997; reprint of the 1970 original.

[ROF92] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.

[WL09] J. Wang and B. J. Lucier, *Error bounds for finite-difference methods for Rudin–Osher–Fatemi image smoothing*, SIAM J. Numer. Anal., 49 (2011), pp. 845–868.