BROKEN SOBOLEV SPACE ITERATION FOR TOTAL VARIATION REGULARIZED MINIMIZATION PROBLEMS

SÖREN BARTELS

ABSTRACT. We devise an improved iterative scheme for the numerical solution of total variation regularized minimization problems. The numerical method realizes a primal-dual iteration with discrete metrics that allow for large step sizes.

1. Introduction

Major progress has recently been made in the development of iterative schemes for the approximate solution of infinite-dimensional, nonsmooth, convex optimization problems, see [NN13, CP11]. By employing an equivalent formulation as a saddle-point problem, the convergence of certain primal-dual methods has been rigorously established. The schemes may be regarded as proximal point algorithms or sub-differential flows of a Lagrange functional and the employed metrics determine the speed of convergence to the stationary saddle-point. Since the numerical schemes are semi-implicit discretizations of continuous evolution problems, the conditions on discretization parameters that guarantee stability of the method crucially depend on the chosen metric. We refer the reader to [Gül10, BC11] for a general overview of related techniques in optimization and to [ROF92, CL97, CKP99, OBG$^+$05] for numerical methods and applications of total variation regularization.
We consider here a model problem that arises in image processing and qualitatively also in the modeling of damage. It consists in the minimization of the functional

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

in the set of functions $u \in BV(\Omega)\cap L^2(\Omega)$. Here, $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, is a bounded Lipschitz domain, $\alpha > 0$ is a given parameter, and $g \in L^2(\Omega)$ is a given noisy image. The first term on the right-hand side is the total variation of the distributional derivative of u which coincides with the $W^{1,1}(\Omega)$ seminorm if u belongs to this space, see [AFP00, ABM06]. The Fenchel conjugate of the $L^1(\Omega)$ norm with respect to the inner product in $L^2(\Omega;\mathbb{R}^d)$ is the indicator functional $I_{K_1(0)}$ of the closed unit ball in $L^\infty(\Omega;\mathbb{R}^d)$. The saddle-point problem associated to the minimization of I thus seeks a pair $(u,p) \in L^2(\Omega)\times H_N(\mathrm{div};\Omega)$ that is stationary for

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

where $(\cdot,\cdot)$ and $\|\cdot\|$ denote the inner product and the norm in $L^2(\Omega)$, respectively and $H_N(\mathrm{div};\Omega)$ is the space of vector fields in $L^2(\Omega;\mathbb{R}^d)$ with square integrable

Date: August 11, 2014.
1991 *Mathematics Subject Classification.* 65K15 (49M29).
Key words and phrases. Total variation, primal-dual method, finite elements, iterative solution.

distributional divergence and vanishing normal component on $\partial\Omega$. The structure of the functional L as the sum of an indicator functional and a linear-quadratic functional makes primal-dual methods attractive. With arbitrary inner products $\langle\cdot,\cdot\rangle$ and $\langle\!\langle\cdot,\cdot\rangle\!\rangle$ on subspaces of $L^2(\Omega)$ and $L^2(\Omega;\mathbb{R}^d)$, respectively, the considered numerical schemes are interpreted as discretizations of the formal evolution problem

$$\begin{aligned} \langle u', v\rangle = -\partial_u L(u,p) = &\ (v, \operatorname{div} p) - \alpha(u-g,v),\\ \langle\!\langle p', q\rangle\!\rangle \ \in \quad &\partial_p L(u,p) = -\partial I_{K_1(0)}(p)[q] - (u, \operatorname{div} q), \end{aligned}$$

where u' and p' denote generalized time-derivatives of u and p. The choice of the inner products needs to be done in such a way that (i) the problems resulting from a discretization in space and time can be solved efficiently, (ii) the discrete time-derivatives are bounded in sufficiently strong norms that allows error terms to be absorbed, and (iii) they do not limit relevant discontinuities of solutions.

In the choice of the metrics we exploit the observation that if the data function satisfies $g \in L^\infty(\Omega)$ then owing to monotonicity properties of the total variation and a truncation argument, see, e.g., [ABM06], the minimizer $u \in BV(\Omega)\cap L^2(\Omega)$ of the functional I satisfies $u \in L^\infty(\Omega)$. In an algebraic sense, the minimizer thus belongs to the interpolation space $H^{1/2}(\Omega)$, but this inclusion does not hold analytically, as functions in $BV(\Omega)\cap L^\infty(\Omega)$ may be discontinuous. For $P1$ finite element functions we prove a discrete variant of this inclusion. We then employ a corresponding discrete version of the inner product in $H^{1/2}(\Omega)$ which enables us to improve the step size restriction for discretizations to $\tau = O(h^{1/2})$ as opposed to the more restrictive condition $\tau = O(h)$ when the weaker inner product of $L^2(\Omega)$ is employed, see [CP11, Bar12]. The choice of the inner product of $L^2(\Omega;\mathbb{R}^d)$ for the evolution of the dual variable p allows for a pointwise explicit solution of the variational inequality defined by the second equation. A drawback of the modified iteration is that a nontrivial but sparse linear system of equations has to be solved in every step. For iterations based on a lumped version of the inner product of $L^2(\Omega)$ the steps are fully explicit provided that the term $(\alpha/2)\|u-g\|^2$ is discretized using numerical quadrature.

The outline of this article is as follows. In Section 2 we introduce the required notation and prove an auxiliary discrete interpolation estimate. The modified scheme and its analysis are presented in Section 3. Numerical experiments that study the influence of different inner products and choices of initial data on the performance of the numerical scheme are presented in Section 4.

2. Preliminaries

2.1. **Finite element discretization.** Given a regular triangulation $\mathcal{T}_h$ of the polyhedral Lipschitz domain $\Omega \subset \mathbb{R}^d$ into triangles or tetrahedra with vertices $\mathcal{N}_h$ we define the spaces of continuous, elementwise affine and discontinuous, elementwise constant finite element functions by

$$\begin{aligned} \mathcal{S}^1(\mathcal{T}_h) &= \{v_h \in C(\overline{\Omega}) : v_h|_T \text{ affine for all } T \in \mathcal{T}_h\},\\ \mathcal{L}^0(\mathcal{T}_h)^d &= \{r_h \in L^1(\Omega;\mathbb{R}^d) : r_h|_T \text{ constant for all } T \in \mathcal{T}_h\}, \end{aligned}$$

respectively. We let $\mathcal{I}_h : C(\overline{\Omega}) \to \mathcal{S}^1(\mathcal{T}_h)$ denote the nodal interpolation operator on $\mathcal{T}_h$ which is also applied to bounded, piecewise continuous functions in the numerical experiments. We refer the reader to [BS08] for details on finite element spaces.

With these spaces a nonconforming discretization of the saddle-point formulation is defined by the following lemma.

Lemma 2.1 (Existence of approximations)**.** *There exists a pair* $(u_h, p_h) \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{L}^0(\mathcal{T}_h)^d$ *that is stationary for*

$$L_h(u_h, p_h) = -I_{K_1(0)}(p_h) + (\nabla u_h, p_h) + \frac{\alpha}{2}\|u_h - g\|^2,$$

where $I_{K_1(0)}(p_h) = 0$ *if* $|p_h| \le 1$ *almost everywhere in* Ω *and* $I_{K_1(0)}(p_h) = +\infty$ *otherwise. The pair* $(u_h, p_h) \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{L}^0(\mathcal{T}_h)^d$ *satisfies* $|p_h| \le 1$ *almost everywhere in* Ω *and*

$$(\nabla u_h, q_h - p_h) \le 0, \quad (p_h, \nabla v_h) + \alpha(u_h - g, v_h) = 0 \tag{2.1}$$

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

Proof. The result is an immediate consequence of standard assertions for saddle-point problems, see, e.g., [ET99, Chap. VI] or [BC11, Chap. 15]. □

Remark 2.1. *Convergence* $u_h \to u$ *in* $L^2(\Omega)$ *as* $h \to 0$ *is a consequence of the strong convexity of* I*, see* [Bar12].

2.2. **Discrete maximum principle.** Although we make no explicit use of it, we prove a discrete maximum principle under a structural assumption on the triangulation $\mathcal{T}_h$ and for the case that reduced integration (or mass lumping) is used for the fidelity term, i.e., the norm $\|\cdot\|_\ell$ induced by the scalar product

$$(\phi, \psi)_\ell = \sum_{T \in \mathcal{T}_h} \frac{|T|}{d+1} \sum_{z \in \mathcal{N}_h \cap T} \phi(z)\psi(z)$$

for elementwise continuous functions $\phi, \psi \in L^\infty(\Omega)$.

Lemma 2.2 (Discrete maximum principle)**.** *Assume that* $g \in L^\infty(\Omega)$ *is elementwise continuous,* $\mathcal{T}_h$ *is acute, i.e., that* $k_{zy} = (\nabla\varphi_z, \nabla\varphi_y) \le 0$ *for all distinct nodes* $z, y \in \mathcal{N}_h$ *with associated nodal basis functions* $\varphi_z, \varphi_y \in \mathcal{S}^1(\mathcal{T}_h)$*. If* $u_h \in \mathcal{S}^1(\mathcal{T}_h)$ *is a minimizer for*

$$I_h(u_h) = \int_\Omega |\nabla u_h| \,\mathrm{d}x + \frac{\alpha}{2}\|u_h - g\|_\ell^2,$$

then we have that $\|u_h\|_{L^\infty(\Omega)} \le \|g\|_{L^\infty(\Omega)}$*.*

Proof. We let $u_h \in \mathcal{S}^1(\mathcal{T}_h)$ be the minimizer of I_h and define $\widetilde{u}_h \in \mathcal{S}^1(\mathcal{T}_h)$ via the truncated nodal values

$$\widetilde{u}_h(z) = \begin{cases} u_h(z) & \text{if } |u_h(z)| \le \|g\|_{L^\infty(\Omega)}, \\ \operatorname{sign}\big(u_h(z)\big)\|g\|_{L^\infty(\Omega)} & \text{if } |u_h(z)| > \|g\|_{L^\infty(\Omega)}, \end{cases}$$

for all $z \in \mathcal{N}_h$. Note that we have $\widetilde{u}_h(z) = G\big(u_h(z)\big)$ for all $z \in \mathcal{N}_h$ with a Lipschitz continuous operator $G : \mathbb{R} \to \mathbb{R}$ satisfying $\|G'\|_{L^\infty(\mathbb{R})} \le 1$. We have that $\|\widetilde{u}_h\|_{L^\infty(\Omega)} \le \|g\|_{L^\infty(\Omega)}$ and $\big|\widetilde{u}_h(z) - g|_T(z)\big| \le \big|u_h(z) - g|_T(z)\big|$ for all $T \in \mathcal{T}_h$ and $z \in \mathcal{N}_h \cap T$. For every $v_h \in \mathcal{S}^1(\mathcal{T}_h)$ and $T \in \mathcal{T}_h$ we note, using $k_{zy} = k_{yz}$ and $\sum_{y \in \mathcal{N}_h \cap T} k_{zy} = 0$, that

$$|T|\big|\nabla v_h|_T\big|^2 = -\frac{1}{2} \sum_{z,y \in \mathcal{N}_h \cap T} k_{zy}|v_h(z) - v_h(y)|^2.$$

With the Lipschitz continuity of G this implies that $\big|\nabla \widetilde{u}_h|_T\big| \le \big|\nabla u_h|_T\big|$ for every $T \in \mathcal{T}_h$. We thus have that $I_h(\widetilde{u}_h) \le I_h(u_h)$. Since u_h is assumed to be minimial we deduce that $u_h = \widetilde{u}_h$, and hence $\|u_h\|_{L^\infty(\Omega)} \le \|g\|_{L^\infty(\Omega)}$. □

Remarks 2.1. *(i) The condition $k_{zy} \le 0$ is satisfied if all angles between adjacent edges or faces of triangles or tetrahedra in $\mathcal{T}_h$ are bounded by $\pi/2$.*
(ii) In general, a discrete maximum principle $\|u_h\|_{L^\infty(\Omega)} \le c\|g\|_{L^\infty(\Omega)}$ cannot be expected to hold with $c = 1$.

2.3. **Discrete interpolation.** The following lemma is a discrete version of the fact that functions in $BV(\Omega) \cap L^\infty(\Omega)$ belong to the Besov space $B^{1/2,2}_\infty(\Omega)$ which nearly coincides with the broken Sobolev space $H^{1/2}(\Omega)$, see, e.g., [Tar07, Chapter 38] for details. It implies that if $(u_h)_{h>0}$ is a bounded sequence of finite element functions associated to a shape regular family of regular triangulations in $W^{1,1}(\Omega) \cap L^\infty(\Omega)$, then the weighted norm

$$\|h_{\mathcal{T}}^{1/2} \nabla u_h\|$$

remains bounded as $h \to 0$, i.e., that $\|\nabla u_h\| \le c h_{\min}^{-1/2}$. Here, $h_{\mathcal{T}} \in L^\infty(\Omega)$ is the elementwise constant mesh-size function defined by $h_{\mathcal{T}}|_T = h_T = \operatorname{diam}(T)$ for all $T \in \mathcal{T}_h$ and we denote $h_{\min} = \min h_{\mathcal{T}}$ and $h = h_{\max} = \max h_{\mathcal{T}}$.

Lemma 2.3 (Interpolation inequality)**.** *There exists $c > 0$ such that for all $v_h \in \mathcal{S}^1(\mathcal{T}_h)$ we have*

$$\|h_{\mathcal{T}}^{1/2} \nabla v_h\|^2 \le c \|\nabla v_h\|_{L^1(\Omega)} \|v_h\|_{L^\infty(\Omega)}.$$

Proof. For $T \in \mathcal{T}_h$ an integration by parts on T shows

$$h_T \int_T |\nabla v_h|^2 \,\mathrm{d}x = -h_T \int_{\partial T} v_h (\nabla v_h \cdot \nu_T) \,\mathrm{d}s \le \|v_h\|_{L^\infty(\Omega)} h_T \|\nabla v_h\|_{L^1(\partial T)}.$$

Noting $h_T|\partial T| \le c|T|$ and summing over all $T \in \mathcal{T}_h$ implies the assertion. □

For $s \in [0,1]$ and $u_h, v_h \in \mathcal{S}^1(\mathcal{T}_h)$ we define the inner product

$$(u_h, v_h)_{h,s} = \int_\Omega u_h v_h \,\mathrm{d}x + \int_\Omega h_{\mathcal{T}}^{(1-s)/s} \nabla u_h \cdot \nabla v_h \,\mathrm{d}x$$

and let $\|u_h\|_{h,s}$ denote the corresponding norm on $\mathcal{S}^1(\mathcal{T}_h)$. We use the convention that $h_{\mathcal{T}}^{(1-s)/s} = 0$ for $s = 0$. It then follows that $(\cdot,\cdot)_{h,s}$ coincides with the inner product in $L^2(\Omega)$ if $s = 0$ and with the inner product in $H^1(\Omega)$ if $s = 1$.

Lemma 2.4 (Inverse estimate)**.** *There exists $c > 0$ such that for all $v_h \in \mathcal{S}^1(\mathcal{T}_h)$ we have*

$$\|\nabla v_h\| \le c h_{\min}^{-\min\{1,(1-s)/(2s)\}} \|v_h\|_{h,s}.$$

Proof. For $s > 0$ we have by definition of $\|\cdot\|_{h,s}$ that

$$\|\nabla v_h\| \le h_{\min}^{-(1-s)/(2s)} \|h_{\mathcal{T}}^{(1-s)/(2s)} \nabla v_h\| \le h_{\min}^{-(1-s)/(2s)} \|v_h\|_{h,s}$$

for all $v_h \in \mathcal{S}^1(\mathcal{T}_h)$ while for $s \ge 0$ the inverse estimate $\|\nabla v_h\| \le c h_{\min}^{-1} \|v_h\|$, see [BS08], implies

$$\|\nabla v_h\| \le c h_{\min}^{-1} \|v_h\|_{h,s}.$$

The combination of the estimates proves the asserted inequality. □

2.4. **Difference quotients.** For a step-size $\tau > 0$ we let d_t denote the backward difference quotient defined by $d_t a^n = (a^n - a^{n-1})/\tau$ for a sequence $(a^n)_{n=0,\dots,N}$. We note that we have

$$(a^n, d_t a^n) = \frac{d_t}{2}\|a^n\|^2 + \frac{\tau}{2}\|d_t a^n\|^2 \tag{2.2}$$

for $n = 1, 2, \dots, N$. For sequences $(a^n)_{n=0,\dots,N}$ and $(b_n)_{n=0,\dots,N}$ we have Leibniz' summation by parts formula

$$\tau \sum_{n=1}^{N} (d_t a^n) b^n = \tau \sum_{n=1}^{N} a^{n-1}(d_t b^n) + a^N b^N - a^0 b^0 \tag{2.3}$$

which follows from a summation of the discrete product rule $d_t(a^n b^n) = (d_t a^n) b^n + a^{n-1}(d_t b^n)$.

3. Modified primal-dual method

The following algorithm is a modified version of the algorithms used in [CP11, Bar12]. It has the interpretation of a discretization of the formal evolution problem

$$\begin{aligned}(u', v)_s &= (\operatorname{div} p, v) - \alpha(u - g, v),\\ (-p', q - p) &\le (u, \operatorname{div}(q - p)) + I_{K_1(0)}(q) - I_{K_1(0)}(p)\end{aligned}$$

with initial conditions $u(0) = u_0$ and $p(0) = p_0$. For $u' = 0$ and $p' = 0$, the equations coincide with the optimality conditions (2.1).

Algorithm 1 (Primal-dual iteration)**.** *Given $\tau > 0$ and $(u_h^0, p_h^0) \in \mathcal{S}^1(\mathcal{T}_h) \times \mathcal{L}^0(\mathcal{T}_h)^d$ set $d_t u_h^0 = 0 \in \mathcal{S}^1(\mathcal{T}_h)$ and $n = 1$ and iterate the following steps:*
(1) Set $\widetilde{u}_h^n = u_h^{n-1} + \tau d_t u_h^n$.
(2) Compute $p_h^n \in \mathcal{L}^0(\mathcal{T}_h)^d$ such that

$$(-d_t p_h^n + \nabla \widetilde{u}_h^n, q_h - p_h^n) + I_{K_1(0)}(p_h^n) \le I_{K_1(0)}(q_h)$$

for all $q_h \in \mathcal{L}^0(\mathcal{T}_h)^d$.
(3) Compute $u_h^n \in \mathcal{S}^1(\mathcal{T}_h)$ such that

$$(d_t u_h^n, v_h)_{h,s} + (p_h^n, \nabla v_h) = -\alpha(u_h^n - g, v_h)$$

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

$$\sup_{v_h \in \mathcal{S}^1(\mathcal{T}_h)} \frac{(d_t u_h^n, v_h)_{h,s}}{\|v_h\|} + \|d_t p_h^n\| \le \varepsilon_{\mathrm{stop}}.$$

Remarks 3.1. *(i) The equation in Step (2) is equivalent to seeking $p_h^n \in \mathcal{L}^0(\mathcal{T}_h)^d$ with $|p_h^n| \le 1$ almost everywhere in Ω and*

$$(-d_t p_h^n + \nabla \widetilde{u}_h^n, q_h - p_h^n) \le 0$$

for all $q_h \in \mathcal{L}^0(\mathcal{T}_h)^d$ with $|q_h| \le 1$ almost everywhere in Ω. This inequality characterizes the unique minimizer $p_h^n \in \mathcal{L}^0(\mathcal{T}_h)^d$ of the mapping

$$p_h \mapsto \frac{1}{2\tau}\|p_h - p_h^{n-1}\|^2 + I_{K_1(0)}(p_h) - (p_h, \nabla \widetilde{u}_h^n).$$

Owing to the choice of the inner product of $L^2(\Omega; \mathbb{R}^d)$ for the evolution of the dual variable we have

$$p_h^n = \frac{p_h^{n-1} + \tau \nabla \widetilde{u}_h^n}{\max\{1, |p_h^{n-1} + \tau \nabla \widetilde{u}_h^n|\}}$$

which can be evaluated elementwise.
(ii) The stopping criterion controls the residual of the equation of the primal variable u with respect to the norm in $L^2(\Omega)$. A stopping criterion of the form $\|d_t u_h^n\| \le \varepsilon_{\rm stop}$ corresponds to measuring the residual of the optimality conditions in a norm which is dual to $\|\cdot\|_{h,s}$ and the output u_h^* of the algorithm would critically depend on the parameter s.

Proposition 3.1 (Convergence)**.** *Let $(u_h,p_h) \in \mathcal{S}^1(\mathcal{T}_h)\times\mathcal{L}^0(\mathcal{T}_h)^d$ be a solution of the discrete saddle-point problem defined through L_h in Lemma 2.1. Provided that $\tau\Theta_h \le 1/2$ for*

$$\Theta_h = \sup_{v_h\in\mathcal{S}^1(\mathcal{T}_h)} \frac{\|\nabla v_h\|}{\|v_h\|_{h,s}} \le c h_{\min}^{-\min\{1,(1-s)/(2s)\}}$$

we have for the iterates $(u_h^n)_{n=0,\dots,N}$ and $(p_h^n)_{n=0,\dots,N}$ of Algorithm 1 that

$$\tau\sum_{n=1}^N \big(\alpha\|u_h^n-u_h\|^2 + \frac{\tau}{4}\|d_t u_h^n\|_{h,s}^2 + \frac{\tau}{4}\|d_t p_h^n\|^2\big) \le \frac12\big(\|u_h-u_h^0\|_{h,s}^2 + \|p_h-p_h^0\|^2\big).$$

Proof. We denote $\delta_u^n = u_h - u_h^n$ and $\delta_p^n = p_h - p_h^n$. The formula (2.2) and the identities $d_t\delta_u^n = -d_t u_h^n$ and $d_t\delta_p^n = -d_t p_h^n$ show

$$\begin{aligned}\Upsilon(n) &= \frac{d_t}{2}\big(\|\delta_u^n\|_{h,s}^2 + \|\delta_p^n\|^2\big) + \frac{\tau}{2}\big(\|d_t\delta_u^n\|_{h,s}^2 + \|d_t\delta_p^n\|^2\big) + \alpha\|\delta_u^n\|^2\\ &= (d_t\delta_u^n,\delta_u^n)_{h,s} + (d_t\delta_p^n,\delta_p^n) + \alpha\|\delta_u^n\|^2 = -(d_t u_h^n,\delta_u^n)_{h,s} - (d_t p_h^n,\delta_p^n) + \alpha\|\delta_u^n\|^2.\end{aligned}$$

Using the equations for $d_t u_h^n$ and $d_t p_h^n$ of Algorithm 1 leads to

$$\Upsilon(n) \le \alpha(u_h^n - g,\delta_u^n) + (p_h^n,\nabla\delta_u^n) - (\delta_p^n,\nabla\widetilde{u}_h^n) + \alpha\|\delta_u^n\|^2.$$

We note $\alpha\|\delta_u^n\|^2 = \alpha(u_h,\delta_u^n) - \alpha(u_h^n,\delta_u^n)$ and deduce that

$$\Upsilon(n) \le \alpha(u_h - g,\delta_u^n) + (p_h^n,\nabla\delta_u^n) - (\delta_p^n,\nabla\widetilde{u}_h^n).$$

The identity in (2.1) and a rearrangement yield

$$\begin{aligned}\Upsilon(n) &\le -(p_h,\nabla\delta_u^n) + (p_h^n,\nabla\delta_u^n) - (\delta_p^n,\nabla\widetilde{u}_h^n)\\ &= (\delta_p^n,\nabla(u_h^n-\widetilde{u}_h^n)) - (\delta_p^n,\nabla u_h).\end{aligned}$$

Since $|p_h^n| \le 1$ almost everywhere in Ω we deduce from the variational inequality in (2.1) that $(\nabla u_h,\delta_p^n)\ge 0$ and by incorporating the identity $u_h^n - \widetilde{u}_h^n = u_h^n - u_h^{n-1} - \tau d_t u_h^{n-1} = \tau^2 d_t^2 u_h^n$ we find

$$\Upsilon(n) \le (\delta_p^n,\nabla(u_h^n-\widetilde{u}_h^n)) \le \tau^2(\delta_p^n,\nabla d_t^2 u_h^n).$$

We sum this estimate over $n=1,2,...,N$ and multiply by τ to verify

$$\begin{aligned}\frac12\big(\|\delta_u^N\|_{h,s}^2 + \|\delta_p^N\|^2\big) + \tau\sum_{n=1}^N \frac{\tau}{2}\big(\|d_t\delta_u^n\|_{h,s}^2 + \|d_t\delta_p^n\|^2\big) + \alpha\tau\sum_{n=1}^N\|\delta_u^n\|^2&\\ \le \frac12\big(\|\delta_u^0\|_{h,s}^2 + \|\delta_p^0\|^2\big) + \tau^3\sum_{n=1}^N(\delta_p^n,\nabla d_t^2 u_h^n).&\end{aligned}$$

With Leibniz' formula (2.3), the identity $d_t u_h^{n-1} = -d_t\delta_u^{n-1}$, and $d_t u_h^0 = 0$ we find

$$\tau^3\sum_{n=1}^N(\delta_p^n,\nabla d_t^2 u_h^n) = \tau^3\sum_{n=1}^N(d_t\delta_p^n,\nabla d_t\delta_u^{n-1}) + \tau^2(\delta_p^N,\nabla d_t u_h^N).$$

Two applications of Young's inequality imply

$$\tau^3\sum_{n=1}^N(\delta_p^n,\nabla d_t^2 u_h^n)\le \tau^2\Big(\sum_{n=1}^N\tau^2\|\nabla d_t\delta_u^{n-1}\|^2+\frac14\|d_t\delta_p^n\|^2\Big)+\frac14\|\delta_p^N\|^2+\tau^4\|\nabla d_t\delta_u^N\|^2.$$

The assumption on τ, i.e., $\tau\|\nabla v_h\|\le\tau\Theta_h\|v_h\|_{h,s}\le(1/2)\|v_h\|_{h,s}$ for all $v_h\in\mathcal{S}^1(\mathcal{T}_h)$ allows us to absorb all of the terms on the right-hand side and to deduce

$$\tau\sum_{n=1}^N\big(\alpha\|\delta_u^n\|^2+\frac{\tau}{2}(1-2\tau^2\Theta_h^2)\|d_t\delta_u^n\|_{h,s}^2+\frac{\tau}{4}\|d_t\delta_p^n\|^2\big)\le\frac12\big(\|\delta_u^0\|_{h,s}^2+\|\delta_p^0\|^2\big)$$

which, noting $1-2\tau^2\Theta_h^2\ge 1/2$, proves the asserted bound on the iterates. The estimate for Θ_h is a direct consequence of Lemma 2.4. □

Remark 3.1. *If the sequence $(u_h-u_h^0)_{h>0}$ is bounded uniformly in $L^\infty(\Omega)\cap W^{1,1}(\Omega)$ then according to Lemma 2.3 the right-hand side of the estimate of Proposition 3.1 is bounded h-independently if $0\le s\le 1/2$. The largest step-size τ is thus possible for $s=1/2$. No restriction on the step size is needed for $s=1$ but in this case the right-hand side of the estimate will typically grow like $h_{\min}^{-1/2}$.*

4. Numerical experiments

We tested Algorithm 1 with different choices of $s\in[0,1]$. For the practical implementation of the stopping criterion we employed the operator $\mathcal{A}_s:\mathcal{S}^1(\mathcal{T}_h)\to\mathcal{S}^1(\mathcal{T}_h)$ defined by

$$(\mathcal{A}_s v_h,w_h)=(v_h,w_h)_{h,s}$$

for all $w_h\in\mathcal{S}^1(\mathcal{T}_h)$. The stopping criterion is then equivalent to $\|\mathcal{A}_s d_t u_h^n\|\le\varepsilon_{\text{stop}}$. If M and S are the mass and stiffness matrix related to the nodal basis of the finite element space $\mathcal{S}^1(\mathcal{T}_h)$ then we have for the coefficient vector $d_t\widehat{U}_h^n$ of $d_t u_h^n$ that

$$\widehat{\mathcal{A}}_s d_t\widehat{U}_h^n=M^{-1}(M+h^{(1-s)/s}S)d_t\widehat{U}_h^n$$

in the case of a uniform triangulation. Mass lumping can be employed to avoid an additional inversion of the mass matrix in every iteration. In the numerical experiments we used the following specification of the model problem.

Example 4.1. *Let $d=2$, $\Omega=(-1,1)^2$, $\alpha=10$, and $g=\chi_{B_{1/2}(0)}+\xi_h$, where ξ_h is a mesh-dependent noise function.*

Table 1 displays the numbers of iterations needed to satisfy the stopping criterion with $\varepsilon_{\text{stop}}=10^{-2}$ for $s=0,1/2,1$, uniform triangulations consisting of halved squares with mesh-size $h=\sqrt{2}2^{-\ell}$ for $\ell=3,...,6$, and the following choices for u_h^0, i.e.,

(A) the smooth function $u_h^0=0$,
(B) the rough function $u_h^0=\mathcal{I}_h g$,
(C) the unique function $u_h^0=Q_h g\in\mathcal{S}^1(\mathcal{T}_h)$ satisfying

$$(\nabla u_h^0,\nabla v_h)+\alpha(u_h^0-\mathcal{I}_h g,v_h)=0$$

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

u_h^0	0			$\mathcal{I}_h g$			$Q_h g$		
s	0	1/2	1	0	1/2	1	0	1/2	1
$\ell = 3$	298	279	725	298	289	788	299	274	695
$\ell = 4$	603	645	2533	602	698	2763	603	601	2307
$\ell = 5$	1575	1065	5903	1572	1238	6811	1569	1074	5604
$\ell = 6$	4249	1394	9986	4225	1778	12524	4099	1559	9754

TABLE 1. Iteration numbers of Algorithm 1 for $\varepsilon_{\rm stop} = 10^{-2}$, different metrics $(\cdot,\cdot)_{h,s}$ specified by the parameter s, uniform triangulations with mesh-size $h = \sqrt{2}2^{-\ell}$, $\ell = 3, 4, ..., 6$, step-size $\tau = h^{1-s}/10$, and initialized with different functions u_h^0 specified by (A), (B), and (C).

We always used $p_h^0 = 0$. The discrete noise function $\xi_h \in \mathcal{S}^1(\mathcal{T}_h)$ was generated by defining its nodal values through normally distributed random numbers with vanishing mean and unit variance. The inner products (g, v_h) for $v_h \in \mathcal{S}^1(\mathcal{T}_h)$ were approximated by $(\mathcal{I}_h g, v_h)$. The step-size was chosen according to

$$\tau = \begin{cases} h/10 & \text{for } s = 0, \\ h^{1/2}/10 & \text{for } s = 1/2, \\ 1/10 & \text{for } s = 1, \end{cases}$$

i.e., $\tau = h^{1-s}/10$. For nearly all choices of initial data and refinement levels we obtain smaller iteration numbers for the choice $s = 1/2$ than for $s = 0$ and $s = 1$. This confirms the expected properties of the modified iteration. In the case $s = 1/2$ the rough initial function $u_h^0 = \mathcal{I}_h g$ of case (B) leads to larger iteration numbers than the functions defined in (A) and (C) which is in agreement with Remark 3.1. The iteration numbers grow superlinearly for $s = 0$, sublinearly for $s = 1/2$, and approximately linearly for $s = 1$ in the experiment. The overall conclusion from the experiments is that the choices $s = 1/2$ and $u_h^0 = 0$ lead to the smallest iteration numbers.

Remark 4.1. *If the fidelity term $(\alpha/2)\|u - g\|^2$ is discretized using mass lumping and if the corresponding inner product $(\cdot,\cdot)_\ell$ is used instead of $(\cdot,\cdot)_{h,s}$ in Step (3) of Algorithm 1, then only a linear system of equations with diagonal matrix has to be solved in every step. Compared to the use of the inner product $(\cdot,\cdot)_{h,s}$, the reduced numerical effort in the iterations may then compensate the larger number of required steps. On the other hand, only standard finite element matrices are involved in the linear system of equations which can be efficiently inverted.*

REFERENCES

[ABM06] Hedy Attouch, Giuseppe Buttazzo, and Gérard Michaille, *Variational analysis in Sobolev and BV spaces*, MPS/SIAM Series on Optimization, vol. 6, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Programming Society (MPS), Philadelphia, PA, 2006.

[AFP00] Luigi Ambrosio, Nicola Fusco, and Diego Pallara, *Functions of bounded variation and free discontinuity problems*, Oxford Mathematical Monographs, The Clarendon Press, Oxford University Press, New York, 2000.

[Bar12] Sören Bartels, *Total variation minimization with finite elements: convergence and iterative solution*, SIAM J. Numer. Anal. **50** (2012), no. 3, 1162–1180.

[BC11] Heinz H. Bauschke and Patrick L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, Springer, New York, 2011.

[BS08] Susanne C. Brenner and L. Ridgway Scott, *The mathematical theory of finite element methods*, third ed., Texts in Applied Mathematics, vol. 15, Springer, New York, 2008.

[CKP99] Eduardo Casas, Karl Kunisch, and Cecilia Pola, *Regularization by functions of bounded variation and applications to image enhancement*, Appl. Math. Optim. **40** (1999), no. 2, 229–257.

[CL97] Antonin Chambolle and Pierre-Louis Lions, *Image recovery via total variation minimization and related problems*, Numer. Math. **76** (1997), no. 2, 167–188.

[CP11] Antonin Chambolle and Thomas Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision **40** (2011), no. 1, 120–145.

[ET99] Ivar Ekeland and Roger Témam, *Convex analysis and variational problems*, english ed., Classics in Applied Mathematics, vol. 28, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1999.

[Gül10] Osman Güler, *Foundations of optimization*, Graduate Texts in Mathematics, vol. 258, Springer, New York, 2010.

[NN13] Yurii Nesterov and Arkadi Nemirovski, *On first-order algorithms for ℓ_1/nuclear norm minimization*, Acta Numer. **22** (2013), 509–575.

[OBG+05] Stanley Osher, Martin Burger, Donald Goldfarb, Jinjun Xu, and Wotao Yin, *An iterative regularization method for total variation-based image restoration*, Multiscale Model. Simul. **4** (2005), no. 2, 460–489.

[ROF92] Leonid I. Rudin, Stanley Osher, and Emad Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D **60** (1992), no. 1-4, 259–268.

[Tar07] Luc Tartar, *An introduction to Sobolev spaces and interpolation spaces*, Lecture Notes of the Unione Matematica Italiana, vol. 3, Springer, Berlin, 2007.

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