SIAM J. APPL. MATH.
Vol. 64, No. 4, pp. 1311–1333

TOTAL BOUNDED VARIATION REGULARIZATION AS A BILATERALLY CONSTRAINED OPTIMIZATION PROBLEM*

M. HINTERMÜLLER† AND K. KUNISCH‡

Abstract. It is demonstrated that the predual for problems with total bounded variation regularization terms can be expressed as a bilaterally constrained optimization problem. Existence of a Lagrange multiplier and an optimality system are established. This allows us to utilize efficient optimization methods developed for problems with box constraints in the context of bounded variation formulations. Here, in particular, the primal-dual active set method, considered as a semismooth Newton method, is analyzed, and superlinear convergence is proved. As a by-product we obtain that the Lagrange multiplier associated with the box constraints acts as an edge detector. Numerical results for image denoising and zooming/resizing show the efficiency of the new approach.

Key words. total bounded variation, predual, semismooth Newton methods, box constraints, image reconstruction

AMS subject classifications. 94A08, 49M29, 65K05

DOI. 10.1137/S0036139903422784

1. Introduction and notation.

This work is concerned with the study of the problem

$$\left\{ \begin{array}{ll} \min & \frac{1}{2}\int_\Omega |Ku-f|^2 dx + \frac{\alpha}{2}\int_\Omega |u|^2 dx + \beta \int_\Omega |\mathrm{D}u| \\ \text{over} & u \in \mathrm{BV}(\Omega), \end{array} \right. \tag{1.1}$$

where Ω is a simply connected domain in $\mathbb{R}^2$ with Lipschitz continuous boundary $\partial\Omega$, $f \in \mathrm{L}^2(\Omega)$, $\beta > 0$, $\alpha \geq 0$ are given, and $\mathrm{K} \in \mathcal{L}(\mathrm{L}^2(\Omega))$. By K^* we denote the adjoint of K. We assume that K^*K is invertible or $\alpha > 0$. Further, $\mathrm{BV}(\Omega)$ denotes the space of functions of bounded variation. A function u is in $\mathrm{BV}(\Omega)$ if the BV seminorm defined by

$$\int_\Omega |Du| = \sup\left\{ \int_\Omega u \ \mathrm{div}\, \vec{v} : \vec{v} \in (C_0^\infty(\Omega))^2,\ |\vec{v}(x)|_{\ell^\infty} \leq 1 \right\}$$

is finite. It is well known [16] that $\mathrm{BV}(\Omega) \subset \mathrm{L}^2(\Omega)$ for $\Omega \subset \mathbb{R}^2$, and that $u \mapsto |u|_{L^2} + \int_\Omega |\mathrm{D}u|$ defines a norm on $\mathrm{BV}(\Omega)$. If $K = identity$, then (1.1) is the well-known image restoration problem with BV-regularization term. It consists of recovering the true image u from the noisy image f. It is well known [9] that (1.1) admits a unique solution $u^* \in \mathrm{BV}(\Omega)$. BV-regularization, differently from regularization by means of $\int_\Omega |\nabla u|^2\, dx$, for example, is known to be preferable due to its ability to preserve edges in the original image during the reconstruction process. Since the pioneering work in [23], the literature on (1.1) has grown tremendously. We give some selected references [1, 5, 7, 12, 14, 18] and refer to the recent monograph [25] for further references.

*Received by the editors February 14, 2003; accepted for publication (in revised form) December 18, 2003; published electronically May 5, 2004. This research was supported by the special research grant SFB "Optimierung und Kontrolle."

http://www.siam.org/journals/siap/64-4/42278.html

†Department of Computational and Applied Mathematics, Rice University, CAAM - MS 134, 6100 S. Main Street, Houston, TX 77005 (hint@caam.rice.edu).

‡Department of Mathematics, University of Graz, Heinrichstr. 36, A-8010 Graz, Austria (karl.kunisch@uni-graz.at).

The original formulation has been extended in various directions including concepts of reconstruction of images with multiple scales; see, e.g., [2, 4, 6, 19].

Despite its favorable properties for reconstruction of images, and especially images with blocky structure, problem (1.1) poses some severe difficulties. On the analytical level these are related to the fact that (1.1) is posed in a nonreflexive Banach space, the dual of which is difficult to characterize [16, 19], and on the numerical level the optimality system related to (1.1) consists of a nonlinear partial differential equation, which is not directly amenable to numerical implementations.

In the present work we show the remarkable result that while the dual of the nonreflexive Banach space problem (1.1) has a complicated measure theoretic structure, its predual can be characterized in a well-known Hilbert space setting. Specifically, the predual to (1.1) is a quadratic optimization problem with bilateral constraints. For such problems the literature provides a variety of possible algorithms. Here we describe and analyze two variants of semismooth Newton methods. We prove their superlinear convergence and provide numerical examples for some denoising and zooming problems. In practice these algorithms are globally convergent without the need for line searches. As a by-product we obtain that the Lagrange multiplier associated with the box constraints acts as an edge detector. We show numerically that the edge detecting property does not require any postprocessing on the multiplier such as thresholding or sharpening techniques.

Let us briefly mention a few alternatives that have been investigated for treating (1.1) numerically. In [23] a time marching scheme to solve the necessary optimality condition related to (1.1) is used. Time marching is also essential for the work in, e.g., [6]. In [19, 26] fixed point iteration schemes are applied to the optimality system using primal variables only. The optimality system based on the primal and dual variables is the basis for the schemes in [19] and [8]. In the former an augmented Lagrangian-based active set strategy is used; in the latter a Newton method is applied. Compared to the formulations used in earlier work, ours appears to have the advantage of being of significantly simpler structure since only a quadratic problem with affine box constraints must be solved. In earlier work, if analysis is carried out, then frequently $\int_\Omega |\mathrm{D}u|$ is replaced by

$$\int_\Omega \sqrt{\delta + |\nabla u|^2}\, dx, \tag{1.2}$$

for $\delta > 0$. In our approach the algorithms are well posed for $\delta = 0$, and for the discretized formulations we have superlinear convergence, still with $\delta = 0$.

The paper is organized as follows. In the remainder of this section we recall some facts from convex analysis and summarize the function space notation that will be used. In section 2 we characterize the predual of (1.1) in the sense of Fenchel. We shall point out the close connection, for 1D (one–dimensional) problems, between our algorithm and the taut-string algorithm well known in nonparametric regression analysis [11, 21]. Section 3 is devoted to the description and convergence proof for a class of regularized problems. Semismooth Newton methods for the predual problems are developed in section 4. Superlinear convergence for the regularized infinite-dimensional problems, and for the discretized predual problems without extra regularization, is proved. Section 5 is devoted to a numerical feasibility study of our results.

We recall the Fenchel duality theorem in infinite-dimensional spaces in a form that is convenient for our work; see, e.g., [3, 13] for details. Let V and Y be Banach spaces with topological duals denoted by V^* and Y^*, respectively. Further, let $\Lambda \in \mathcal{L}(V, Y)$

and let $\mathcal{F} : V \to \mathbb{R} \cup \{\infty\}$, $\mathcal{G} : Y \to \mathbb{R} \cup \{\infty\}$ be convex lower semicontinuous functionals not identically equal to ∞, and assume that there exists $v_0 \in V$ such that $\mathcal{F}(v_0) < \infty$, $\mathcal{G}(\Lambda v_0) < \infty$, and $\mathcal{G}$ is continuous at Λv_0. Then we have

$$\inf_{u \in V} \mathcal{F}(u) + \mathcal{G}(\Lambda u) = \sup_{p \in Y^*} -\mathcal{F}^*(\Lambda^* p) - \mathcal{G}^*(-p), \tag{1.3}$$

where $\mathcal{F}^* : V^* \to \mathbb{R} \cup \{\infty\}$ denotes the conjugate of $\mathcal{F}$ defined by

$$\mathcal{F}^*(v^*) = \sup_{v \in V} \langle v, v^* \rangle_{V,V^*} - \mathcal{F}(v).$$

Under the conditions imposed on $\mathcal{F}$ and $\mathcal{G}$, it is known that the problem on the right-hand side of (1.3) admits a solution. Moreover, $(\bar{u}, \bar{p})$ are solutions to the two optimization problems in (1.3) if and only if

$$\Lambda^* \bar{p} \in \partial \mathcal{F}(\bar{u}), \tag{1.4a}$$

$$-\bar{p} \in \partial \mathcal{G}(\Lambda \bar{u}), \tag{1.4b}$$

where $\partial \mathcal{F}$ denotes the subdifferential of the convex functional $\mathcal{F}$.

To compute, formally, the Fenchel dual to (1.1) we set $\Lambda = \nabla$,

$$\mathcal{F}(u) = \frac{1}{2}|\mathrm{K}u - f|^2 + \frac{\alpha}{2}|u|^2 \qquad \textit{and} \qquad \mathcal{G}(\vec{p}) = \beta \int_\Omega |\vec{p}|_{\ell^1} dx,$$

where u and $\vec{p}$ denote a scalar and a 2D vector-valued function, respectively. Further, $|\cdot|$ denotes the $L^2(\Omega)$-norm and $|\cdot|_{\ell^1}$ stands for the ℓ^1-norm on $\mathbb{R}^n$. For the convex conjugates we find

$$\mathcal{F}^*(v) = \frac{1}{2}(v + \mathrm{K}^* f, B^{-1}(v + \mathrm{K}^* f)) - \frac{1}{2}|f|^2 \quad \text{and} \quad \mathcal{G}^*(\vec{p}) = \mathrm{I}_{[-\beta\vec{1},\beta\vec{1}]}(\vec{p}),$$

where $\vec{1}$ is the 2D vector field with 1 in both coordinates, $B = \alpha\,\mathrm{I} + \mathrm{K}^*\mathrm{K}$, and

$$\mathrm{I}_{[-\beta\vec{1},\beta\vec{1}]}(\vec{p}) = \begin{cases} 0 & \text{if } -\beta\vec{1} \le \vec{p}(x) \le \beta\vec{1} \qquad \text{for almost every (a.e.) } x \in \Omega, \\ \infty & \text{otherwise.} \end{cases}$$

Thus, formally the dual to (1.1) is given by

$$\begin{cases} \inf \ \frac{1}{2}|\operatorname{div} \vec{p} + \mathrm{K}^* f|_B^2 \\ \text{s.t.} \ -\beta\vec{1} \le \vec{p}(x) \le \beta\vec{1} \qquad \text{for a.e. } x \in \Omega, \end{cases} \tag{1.5}$$

where $|v|_B^2 = (v, B^{-1}v)$, and the relationship (1.4) applied to the solutions of (1.1) and (1.5) implies that

$$\operatorname{div} \vec{p} = Bu - \mathrm{K}^* f, \quad \vec{p} = \beta \left(\frac{u_{x_i}}{|u_{x_i}|} \right)_{i=1}^n \qquad \text{on } \{x : u_{x_i}(x) \neq 0 \text{ for all } i\}. \tag{1.6}$$

The functional analytic statement corresponding to (1.6) is given in (2.3), (2.4) below.

We note that nondifferentiability due to the BV-term in (1.1) is replaced by the bilateral constraints in the formal dual (1.5).

In the next section we shall put (1.5) into a proper functional analytical framework. For this purpose we require some notation which we summarize next. Let $\mathbb{L}^2(\Omega) = \mathrm{L}^2(\Omega) \times \mathrm{L}^2(\Omega)$ endowed with the Hilbert space inner product structure and

norm. If the context suggests to do so, then we shall distinguish between vector fields $\vec{v} \in \mathbb{L}^2(\Omega)$ and scalar functions $v \in \mathrm{L}^2(\Omega)$ by using an arrow on top of the letter. Analogously we set $\mathbb{H}_0^1(\Omega) = H_0^1(\Omega) \times H_0^1(\Omega)$. We denote $L_0^2(\Omega) = \{v \in \mathrm{L}^2(\Omega) : \int_\Omega v dx = 0\}$, $H_0(\mathrm{div}) = \{\vec{v} \in \mathbb{L}^2(\Omega) : \mathrm{div}\,\vec{v} \in \mathrm{L}^2(\Omega), \vec{v} \cdot n = 0 \text{ on } \partial\Omega\}$, where n is the outer normal to $\partial\Omega$. The space $H_0(\mathrm{div})$ is endowed with $|\vec{v}|^2_{H_0(\mathrm{div})} = |\vec{v}|^2_{\mathbb{L}^2(\Omega)} + |\,\mathrm{div}\,\vec{v}|^2_{L^2}$ as norm. Further, we put $H_0(\mathrm{div}\,0) = \{\vec{v} \in H_0(\mathrm{div}) : \mathrm{div}\,\vec{v} = 0 \text{ almost everywhere in } \Omega\}$. It is well known that

$$\mathbb{L}^2(\Omega) = \mathrm{grad}\; H^1(\Omega) \oplus H_0(\mathrm{div}\,0); \tag{1.7}$$

cf. [10, p. 216], for example. Moreover,

$$H_0(\mathrm{div}) = H_0(\mathrm{div}\,0)^\perp \oplus H_0(\mathrm{div}\,0), \tag{1.8}$$

with

$$H_0(\mathrm{div}\,0)^\perp = \{\vec{v} \in \mathrm{grad}\, H^1(\Omega) : \mathrm{div}\,\vec{v} \in L^2(\Omega), \vec{v} \cdot n = 0 \text{ on } \partial\Omega\},$$

and $\mathrm{div} : H_0(\mathrm{div}\,0)^\perp \subset H_0(\mathrm{div}) \to L_0^2(\Omega)$ is a homeomorphism. In fact, it is injective by construction, and for every $f \in L_0^2(\Omega)$ there exists, by the Lax–Milgram lemma, $\varphi \in H^1(\Omega)$ such that

$$\mathrm{div}\,\nabla\varphi = f \text{ in } \Omega, \qquad \nabla\varphi \cdot n = 0 \text{ on } \partial\Omega,$$

with $\nabla\varphi \in H_0(\mathrm{div}\,0)^\perp$. Hence, by the closed mapping theorem we have

$$\mathrm{div} \in \mathcal{L}(H_0(\mathrm{div})^\perp, L_0^2(\Omega)).$$

Finally, let $\mathrm{P}_{\mathrm{div}}$ and $\mathrm{P}_{\mathrm{div}^\perp}$ denote the orthogonal projections in $\mathbb{L}^2(\Omega)$ onto $H_0(\mathrm{div}\,0)$ and $\mathrm{grad}\, H^1(\Omega)$, respectively. Note that the restrictions of $\mathrm{P}_{\mathrm{div}}$ and $\mathrm{P}_{\mathrm{div}^\perp}$ to $H_0(\mathrm{div}\,0)$ coincide with the orthogonal projections in $H_0(\mathrm{div})$ onto $H_0(\mathrm{div}\,0)$ and $H_0(\mathrm{div}\,0)^\perp$.

2. The Fenchel predual. The section is devoted to the study of the problems

$$\begin{cases} \min & \dfrac{1}{2}|\,\mathrm{div}\,\vec{p} + \mathrm{K}^* f|_B^2 \quad \text{over } \vec{p} \in H_0(\mathrm{div}) \\ \text{s.t.} & -\beta\vec{1} \le \vec{p} \le \beta\vec{1}, \end{cases} \tag{2.1}$$

and

$$\begin{cases} \min & \dfrac{1}{2}|\,\mathrm{div}\,\vec{p} + \mathrm{K}^* f|_B^2 + \frac{\gamma}{2}|\mathrm{P}_{\mathrm{div}}\vec{p}|^2 \quad \text{over } \vec{p} \in H_0(\mathrm{div}) \\ \text{s.t.} & -\beta\vec{1} \le \vec{p} \le \beta\vec{1}, \end{cases} \tag{2.2}$$

where $\gamma > 0$ is given, and we recall that for $v \in \mathrm{L}^2(\Omega)$ we set $|v|_B^2 = (v, B^{-1}v)_{L^2}$.

PROPOSITION 2.1. *Both* (2.1) *and* (2.2)*, admit a solution. The solution to* (2.2) *is unique.*

Proof. Existence of a solution to (2.1) as well as (2.2) can be proved by standard arguments. To verify uniqueness of the solution to (2.2) we note that the set of feasible $\vec{p}$ is convex. Hence it suffices to verify strict convexity of $J(\vec{p}) = \frac{1}{2}|\,\mathrm{div}\,\vec{p} + \mathrm{K}^* f|_B^2 + \frac{\gamma}{2}|\mathrm{P}_{\mathrm{div}}\vec{p}|^2$. To ascertain strict convexity of J we use the fact that the second derivative satisfies

$$J''(\vec{p}, \vec{p}) = |\,\mathrm{div}\,\vec{p}\,|_B^2 + \gamma|\mathrm{P}_{\mathrm{div}}\vec{p}\,|^2 \ge \kappa|\vec{p}\,|^2_{H_0(\mathrm{div})}$$

for a constant $\kappa > 0$ independent of $\vec{p} \in H_0(\mathrm{div})$. Here we have used (1.6) and the subsequent comments. Hence J is even uniformly convex, and uniqueness follows. ∎

THEOREM 2.2. *The Fenchel dual to* (2.1) *is given by* (1.1)*, and the solutions* u^* *of* (1.1) *and* $\vec{p}^*$ *of* (2.1) *are related by*

$$Bu^* = \mathrm{div}\,\vec{p}^* + \mathrm{K}^* f, \tag{2.3}$$

$$\langle (-\mathrm{div})^* u^*, \vec{p} - \vec{p}^* \rangle_{H_0(\mathrm{div})^*, H_0(\mathrm{div})} \le 0 \quad \textit{for all } \vec{p} \in H_0(\mathrm{div}), \tag{2.4}$$

with $-\beta\vec{1} \le \vec{p} \le \beta\vec{1}$.

Alternatively, (2.1) can be considered as the predual of the original problem (1.1). If (2.1) is a zero-residue problem, i.e., $\vec{p}^*$ satisfies $\mathrm{div}\,\vec{p}^* = -\mathrm{K}^* f$, then the additional penalty term in (2.2) chooses from among all solutions the one which minimizes $|\mathrm{P}_{\mathrm{div}}\vec{p}^*|$.

Proof of Theorem 2.2. We apply Fenchel duality as recalled in section 1 with $V = H_0(\mathrm{div})$, $Y = Y^* = \mathrm{L}^2(\Omega)$, $\Lambda = -\mathrm{div}$, $\mathcal{G} : Y \to \mathbb{R}$ given by $\mathcal{G}(v) = \frac{1}{2}|v - \mathrm{K}^* f|_B^2$, and $\mathcal{F} : V \to \mathbb{R}$ defined by $\mathcal{F}(\vec{p}\,) = \mathrm{I}_{[-\beta\vec{1},\beta\vec{1}]}(\vec{p}\,)$. The convex conjugate $\mathcal{G}^* : L^2(\Omega) \to \mathbb{R}$ of $\mathcal{G}$ is given by

$$\mathcal{G}^*(v) = \frac{1}{2}|\mathrm{K}v + f|^2 + \frac{\alpha}{2}|v|^2 - \frac{1}{2}|f|^2.$$

Further, the conjugate $\mathcal{F}^* : H_0(\mathrm{div})^* \to \mathbb{R}$ of $\mathcal{F}$ is given by

$$\mathcal{F}^*(\vec{q}) = \sup_{\vec{p} \in S_1} \langle \vec{q}, \vec{p}\,\rangle_{H_0(\mathrm{div})^*, H_0(\mathrm{div})} \quad \text{for } \vec{q} \in H_0(\mathrm{div})^*, \tag{2.5}$$

where $S_1 = \{\vec{p} \in H_0(\mathrm{div}) : -\beta\vec{1} \le \vec{p} \le \beta\vec{1}\}$. Let us set

$$S_2 = \{\vec{p} \in C_0^1(\Omega) \times C_0^1(\Omega) : -\beta\vec{1} \le \vec{p} \le \beta\vec{1}\}.$$

The set S_2 is dense in the topology of $H_0(\mathrm{div})$ in S_1. In fact, let $\vec{p}$ be an arbitrary element of S_1. Since $(\mathcal{D}(\Omega))^2$ is dense in $H_0(\mathrm{div})$ (see, e.g., [15, p. 26]), there exists a sequence $\vec{p}_n \in (\mathcal{D}(\Omega))^2$ converging in $H_0(\mathrm{div})$ to $\vec{p}$. Let $\mathcal{P}$ denote the canonical projection in $H_0(\mathrm{div})$ onto the closed convex subset S_1 and note that, since $\vec{p} \in S_1$,

$$\begin{aligned} |\vec{p} - \mathcal{P}\vec{p}_n|_{H_0(\mathrm{div})} &\le |\vec{p} - \vec{p}_n|_{H_0(\mathrm{div})} + |\vec{p}_n - \mathcal{P}\vec{p}_n|_{H_0(\mathrm{div})} \\ &\le 2|\vec{p} - \vec{p}_n|_{H_0(\mathrm{div})} \to 0 \qquad \text{for } n \to \infty. \end{aligned}$$

Hence $\lim_{n\to\infty} |\vec{p} - \mathcal{P}\vec{p}_n|_{H_0(\mathrm{div})} = 0$ and S_2 is dense in S_1. Returning to (2.5), we have for $v \in \mathrm{L}^2(\Omega)$ and $(-\mathrm{div})^* \in \mathcal{L}(\mathrm{L}^2(\Omega), V^*)$,

$$\mathcal{F}^*((-\mathrm{div})^* v) = \sup_{\vec{p} \in S_2} (v, -\mathrm{div}\,\vec{p}\,),$$

which can be $+\infty$. By the definition of the functions of bounded variation it is finite if and only if $v \in \mathrm{BV}(\Omega)$ (see [16, p. 3]) and

$$\mathcal{F}^*((-\mathrm{div})^* v) = \beta \int_\Omega |\mathrm{D}v| < \infty \quad \text{for } v \in \mathrm{BV}(\Omega).$$

The dual problem to (2.1) is found to be

$$\min\ \frac{1}{2}|\mathrm{K}u - f|^2 + \frac{\alpha}{2}|u|^2 + \beta \int_\Omega |\mathrm{D}u| \quad \text{over } u \in \mathrm{BV}(\Omega).$$

From (1.4), moreover, we find

$$\langle(-\operatorname{div})^* u^*, \vec{p} - \vec{p}^*\rangle_{H_0(\operatorname{div})^*, H_0(\operatorname{div})} \leq 0 \quad \text{for all } p \in S_1$$

and

$$Bu^* = \operatorname{div} \vec{p}^* + \mathrm{K}^* f. \qquad \square$$

We obtain the following optimality system.

COROLLARY 2.3. *Let $\vec{p}^* \in H_0(\operatorname{div})$ be a solution to (2.1). Then there exists $\vec{\lambda}^* \in H_0(\operatorname{div})^*$ such that*

$$\operatorname{div}^* B^{-1} \operatorname{div} \vec{p}^* + \operatorname{div}^* B^{-1} \mathrm{K}^* f + \vec{\lambda}^* = 0, \tag{2.6}$$

$$\langle \vec{\lambda}^*, \vec{p} - \vec{p}^*\rangle_{H_0(\operatorname{div})^*, H_0(\operatorname{div})} \leq 0 \quad \textit{for all } \vec{p} \in H_0(\operatorname{div}), \tag{2.7}$$

with $-\beta\vec{1} \leq \vec{p} \leq \beta\vec{1}$.

For convenience we also specify the variational form of (2.6) which holds in $H_0(\operatorname{div})^*$:

$$(B^{-1} \operatorname{div} \vec{p}^*, \operatorname{div} \vec{v}\,)_{L^2} + (B^{-1}\mathrm{K}^* f, \operatorname{div} \vec{v}\,)_{L^2} + \langle \vec{\lambda}^*, \vec{v}\,\rangle_{H_0(\operatorname{div})^*, H_0(\operatorname{div})} = 0$$

for all $\vec{v} \in H_0(\operatorname{div})$.

Proof of Corollary 2.3. Set $\vec{\lambda}^* = -\operatorname{div}^* u^* \in H_0(\operatorname{div})^*$ and apply $\operatorname{div}^* B^{-1}$ to obtain (2.6). For this choice of $\vec{\lambda}^*$, equation (2.7) follows from (2.4). $\square$

The optimality system for (2.2) is given next.

COROLLARY 2.4. *Let $\vec{p}^* \in H_0(\operatorname{div})^*$ denote the solution to (2.2). Then there exists $\vec{\lambda}^* \in H_0(\operatorname{div})^*$ such that*

$$\operatorname{div}^* B^{-1} \operatorname{div} \vec{p}^* + \operatorname{div}^* B^{-1} \mathrm{K}^* f + \gamma P_{\operatorname{div}} \vec{p}^* + \vec{\lambda}^* = 0, \tag{2.8}$$

$$\langle \vec{\lambda}^*, \vec{p} - \vec{p}^*\rangle_{H_0(\operatorname{div})^*, H_0(\operatorname{div})} \leq 0 \quad \textit{for all } \vec{p} \in H_0(\operatorname{div}), \tag{2.9}$$

with $-\beta\vec{1} \leq \vec{p} \leq \beta\vec{1}$.

Proof. We only sketch the proof here since the assertion will also follow from the proof of Theorem 3.1 below. By (1.6), every $\vec{v} \in H_0(\operatorname{div})$ can be decomposed according to $\vec{v} = \vec{v}_1 + \vec{v}_2 \in H_0(\operatorname{div} 0)^\perp \oplus H_0(\operatorname{div} 0)$. The functional in (2.2) is then separable, and (2.2) can be expressed as

$$\min_{\vec{p} \in H_0(\operatorname{div})} \mathcal{F}(\vec{p}) + \mathcal{G}_1(\Lambda_1 \vec{p}_1) + \mathcal{G}_2(\Lambda_2 \vec{p}_2),$$

where $\mathcal{F}$ is defined in the proof of Theorem 2.2, $\mathcal{G}_1$ and Λ_1 coincide with $\mathcal{G}$ and Λ from the proof of Theorem 2.2, and we set

$$\mathcal{G}_2 : \mathbb{L}^2(\Omega) \to \mathbb{R}, \quad \mathcal{G}_2(\vec{p}) = \frac{\gamma}{2} |\vec{p}\,|^2_{\mathbb{L}^2(\Omega)},$$

$\Lambda_2 \in \mathcal{L}(H_0(\operatorname{div} 0), \mathbb{L}^2(\Omega))$ with Λ_2 the canonical injection. From general results in convex analysis (e.g., [13, p. 61]), there exist $\vec{u}_1^* \in \mathbb{L}^2(\Omega)$ and $\vec{u}_2^* \in \mathbb{L}^2(\Omega)$ such that

$$\begin{aligned} B\vec{u}_1^* &= \operatorname{div} \vec{p}_1^* + \mathrm{K}^* f = \operatorname{div} \vec{p}^* + \mathrm{K}^* f, \\ -\vec{u}_2^* &= \gamma \vec{p}_2^* = \gamma P_{\operatorname{div}} \vec{p}^* \end{aligned}$$

and

$$\langle -\operatorname{div}^* \vec{u}_1^* + \vec{u}_2^*, \vec{p} - \vec{p}^* \rangle_{H_0(\mathrm{div})^*, H_0(\mathrm{div})} \le 0 \qquad \text{for all } \vec{p} \in S_1.$$

The claim follows with $\vec{\lambda}^* = -\operatorname{div}^* \vec{u}_1^* + \vec{u}_2^*$. □

We end this section with the following remarks.

Remark 1.

- In our numerical tests, in many cases we can set $\gamma = 0$. This suggests the conjecture that the constraints $-\beta\vec{1} \le \vec{p} \le \beta\vec{1}$ imply some type of uniqueness.
- We point out the close connection between (2.1) and the taut-string algorithm well known in regression analysis [11, 21]. Here we have $\mathrm{K} = \mathrm{I}$, $\alpha = 0$. A continuous version of the taut-string algorithm can be expressed as

$$\begin{cases} \min \displaystyle\int_0^1 \sqrt{1 + |w_x|^2}\, dx, \\ \text{s.t. } F - \beta \le w \le F + \beta, \end{cases} \tag{2.10}$$

where $F(x) = \int_0^x f(s)ds$. The denoised image u is obtained from $u = w_x$. Observe that the change of variables $p = w - F$ transforms (2.10) into

$$\begin{cases} \min \displaystyle\int_0^1 \sqrt{1 + |p_x + f|^2}\, dx, \\ \text{s.t. } -\beta \le p \le \beta \end{cases} \tag{2.11}$$

and $u = p_x + f$. Thus, except for the square root in (2.11), we obtain (2.1).

3. A family of regularized problems. To treat (2.1) and (2.2) numerically one can discretize these box constrained problems and implement one's algorithm of choice for the resulting finite-dimensional quadratic optimization problems with affine constraints. With such an approach the infinite-dimensional structure tends to get covered up. One of the features that can be pointed out by considering (2.6) and (2.8) of the optimality systems is that the leading differential operator is not smoothing (see (1.7)) as it is for obstacle-type problems, nor is it a compact perturbation of the identity operator as, for instance, for control constrained optimal control problems [17]. This complicates the convergence analysis for semismooth Newton algorithms; see [17, 24]. Therefore we describe in this section a family of approximating problems which have more amenable properties for Newton-type algorithms in an infinite-dimensional setting. A second difficulty with (2.1), (2.2) is related to the fact that β will typically be chosen as a small constant so that the resulting problems are close to bottleneck problems. We shall see in section 5 that the algorithms we propose are able to deal efficiently with such constraints.

As announced above, we focus in this section on a family of approximating problems given by

$$\begin{cases} \min \ \dfrac{1}{2c}|\nabla\vec{p}|^2 + \dfrac{1}{2}|\operatorname{div}\vec{p} + \mathrm{K}^* f|_B^2 + \dfrac{\gamma}{2}|\mathrm{P}_{\mathrm{div}}\vec{p}|^2 \\ \qquad + \dfrac{1}{2c}|\max(0, c(\vec{p} - \beta\vec{1}))|^2 + \dfrac{1}{2c}|\min(0, c(\vec{p} + \beta\vec{1}))|^2 \text{ over } \vec{p} \in \mathbb{H}_0^1(\Omega), \end{cases} \tag{3.1}$$

where $c > 0$. Let $\vec{p}_c$ denote the unique solution to (3.1). It satisfies the optimality

condition

$$-\frac{1}{c}\Delta\vec{p}_c - \nabla B^{-1}\operatorname{div}\vec{p}_c - \nabla B^{-1}\mathrm{K}^* f + \gamma \mathrm{P}_{\mathrm{div}}\vec{p}_c + \vec{\lambda}_c = 0, \tag{3.2a}$$

$$\vec{\lambda}_c = \max(0, c(\vec{p}_c - \beta\vec{1})) + \min(0, c(\vec{p}_c + \beta\vec{1})). \tag{3.2b}$$

Next we address convergence as $c \to \infty$.

THEOREM 3.1. *The family* $\{(\vec{p}_c, \vec{\lambda}_c)\}_{c>0}$ *converges weakly in* $H_0(\mathrm{div}) \times \mathbb{H}_0^1(\Omega)^*$ *to the unique solution* $(\vec{p}^*, \vec{\lambda}^*)$ *of* (2.8), (2.9). *Moreover, the convergence of* $\vec{p}_c$ *to* $\vec{p}^*$ *is strong in* $H_0(\mathrm{div})$.

Proof. Recall the variational form of (2.8) given by

$$(\operatorname{div}\vec{p}^*, \operatorname{div}\vec{v}\,)_B + (\mathrm{K}^* f, \operatorname{div}\vec{v}\,)_B + \gamma(\mathrm{P}_{\mathrm{div}}\vec{p}^*, \mathrm{P}_{\mathrm{div}}\vec{v}\,) + \langle \vec{\lambda}^*, \vec{v}\,\rangle_{H_0(\mathrm{div})^*, H_0(\mathrm{div})} = 0 \tag{3.3}$$

for all $\vec{v} \in H_0(\mathrm{div})$. To verify uniqueness, let us suppose that $(\vec{p}_i, \vec{\lambda}_i) \in H_0(\mathrm{div}) \times H_0(\mathrm{div})^*$, $i = 1, 2$, are two solution pairs to (2.8), (2.9). For $\delta\vec{p} = \vec{p}_2 - \vec{p}_1$, $\delta\vec{\lambda} = \vec{\lambda}_2 - \vec{\lambda}_1$ we have

$$(B^{-1}\operatorname{div}\delta\vec{p}, \operatorname{div}\vec{v}\,) + \gamma(\mathrm{P}_{\mathrm{div}}\delta\vec{p}, \mathrm{P}_{\mathrm{div}}\vec{v}\,) + \langle \delta\vec{\lambda}, \vec{v}\,\rangle_{H_0(\mathrm{div})^*, H_0(\mathrm{div})} = 0 \tag{3.4}$$

for all $\vec{v} \in H_0(\mathrm{div})$, and

$$\langle \delta\vec{\lambda}, \delta\vec{p}\,\rangle_{H_0(\mathrm{div})^*, H_0(\mathrm{div})} \geq 0.$$

With $\vec{v} = \delta\vec{p}$ in (3.4) we obtain

$$|B^{-1}\operatorname{div}\delta\vec{p}\,|^2 + \gamma|\mathrm{P}_{\mathrm{div}}\delta\vec{p}\,|^2 \leq 0,$$

and hence $\vec{p}_1 = \vec{p}_2$. From (3.3) we deduce that $\vec{\lambda}_1 = \vec{\lambda}_2$. Thus uniqueness is established, and we can henceforth rely on subsequential arguments.

In the following computation we consider the coordinates $\vec{\lambda}_c^i$, $i = 1, 2$, of $\vec{\lambda}_c$. We have for the pointwise a.e. evaluation at $x \in \Omega$

$$\begin{aligned}\vec{\lambda}_c^i\,\vec{p}_c^i &= \big(\max(0, c(\vec{p}_c^i - \beta)) + \min(0, c(\vec{p}_c^i + \beta))\big)\,\vec{p}_c^i \\ &= \begin{cases} c(\vec{p}_c^i - \beta)\vec{p}_c^i & \text{if } \vec{p}_c^i \geq \beta, \\ 0 & \text{if } |\vec{p}_c^i| = \beta, \\ c(\vec{p}_c^i + \beta)\vec{p}_c^i & \text{if } \vec{p}_c^i \leq \beta. \end{cases}\end{aligned}$$

It follows that

$$(\vec{\lambda}_c^i, \vec{p}_c^i)_{\mathrm{L}^2(\Omega)} \geq \frac{1}{c}|\vec{\lambda}_c^i|^2_{\mathrm{L}^2(\Omega)} \quad \text{for } i = 1, 2,$$

and consequently

$$(\vec{\lambda}_c, \vec{p}_c)_{\mathbb{L}^2(\Omega)} \geq \frac{1}{c}|\vec{\lambda}_c|^2_{\mathbb{L}^2(\Omega)} \quad \text{for every } c > 0. \tag{3.5}$$

From (3.2) and (3.5) we deduce that

$$\frac{1}{c}|\nabla\vec{p}_c|^2 + |\operatorname{div}\vec{p}_c|_B^2 + \gamma|\mathrm{P}_{\mathrm{div}}\vec{p}_c|^2 \leq |\operatorname{div}\vec{p}_c|_B|\mathrm{K}^* f|_B$$

and hence

$$\frac{1}{c}|\nabla \vec{p}_c|^2 + \frac{1}{2}|\operatorname{div} \vec{p}_c|_B^2 + \gamma |\mathrm{P}_{\mathrm{div}} \vec{p}_c|^2 \le \frac{1}{2}|\mathrm{K}^* f|_B. \tag{3.6}$$

We further estimate

$$\begin{aligned} |\vec{\lambda}_c|_{\mathbb{H}_0^1(\Omega)^*} &= \sup_{|\vec{v}\,|_{\mathbb{H}_0^1(\Omega)}=1} \langle \vec{\lambda}_c, \vec{v}\,\rangle_{\mathbb{H}_0^1(\Omega)^*, \mathbb{H}_0^1(\Omega)} \\ &\le \sup_{|\vec{v}|_{\mathbb{H}_0^1(\Omega)}=1} \Big\{ \frac{1}{c}|\nabla \vec{p}_c||\nabla \vec{v}\,| + |\operatorname{div} \vec{p}_c|_B |\operatorname{div} \vec{v}\,|_B + |\mathrm{K}^* f|_B |\operatorname{div} \vec{v}\,|_B \\ &\qquad + \gamma |\mathrm{P}_{\mathrm{div}} \vec{p}_c|\,|\mathrm{P}_{\mathrm{div}} \vec{v}\,| \Big\}. \end{aligned}$$

From (3.6) we deduce the existence of a constant K independent of $c \ge 1$ such that

$$|\vec{\lambda}_c|_{\mathbb{H}_0^1(\Omega)^*} \le K. \tag{3.7}$$

Combining (3.6) and (3.7), we can assert the existence of $(\vec{p}^*, \vec{\lambda}^*) \in H_0(\mathrm{div}) \times \mathbb{H}_0^1(\Omega)^*$ such that for a subsequence denoted by the same symbol

$$(\vec{p}_c, \vec{\lambda}_c) \rightharpoonup (\vec{p}^*, \vec{\lambda}^*) \qquad \text{weakly in } H_0(\mathrm{div}) \times \mathbb{H}_0^1(\Omega)^*. \tag{3.8}$$

We recall the variational form of (3.2), i.e.,

$$\begin{aligned} \frac{1}{c}(\nabla \vec{p}_c, \nabla \vec{v}\,) + (\operatorname{div} \vec{p}_c, \operatorname{div} \vec{v}\,)_B + (\mathrm{K}^* f, \operatorname{div} \vec{v}\,)_B + \gamma (\mathrm{P}_{\mathrm{div}} \vec{p}_c, \mathrm{P}_{\mathrm{div}} \vec{v}\,) \\ + (\vec{\lambda}_c, \vec{v}\,) = 0 \quad \text{for all } \vec{v} \in \mathbb{H}_0^1(\Omega). \end{aligned}$$

Passing to the limit $c \to \infty$, using (3.6) and (3.8) we have

$$\begin{aligned} (\operatorname{div} \vec{p}^*, \operatorname{div} \vec{v}\,)_B + (\mathrm{K}^* f, \operatorname{div} \vec{v}\,)_B + \gamma (\mathrm{P}_{\mathrm{div}} \vec{p}^*, \mathrm{P}_{\mathrm{div}} \vec{v}\,) \\ + \langle \vec{\lambda}^*, \vec{v}\,\rangle_{\mathbb{H}_0^1(\Omega)^*, \mathbb{H}_0^1(\Omega)} = 0 \quad \text{for all } \vec{v} \in \mathbb{H}_0^1(\Omega). \end{aligned} \tag{3.9}$$

Since $\mathbb{H}_0^1(\Omega)$ is dense in $H_0(\mathrm{div})$ and $\vec{p}^* \in H_0(\mathrm{div})$, we have that (3.9) holds for all $\vec{v} \in H_0(\mathrm{div})$. Consequently $\vec{\lambda}^*$ can be identified with an element in $H_0(\mathrm{div})^*$, and $\langle \cdot, \cdot \rangle_{\mathbb{H}_0^1(\Omega)^*, \mathbb{H}_0^1(\Omega)}$ in (3.9) can be replaced by $\langle \cdot, \cdot \rangle_{H_0(\mathrm{div})^*, H_0(\mathrm{div})}$. We next verify that $\vec{p}^*$ is feasible. For this purpose note that

$$(\vec{\lambda}_c, \vec{p} - \vec{p}_c) = \big(\max(0, c(\vec{p}_c - \beta \vec{1})) + \min(0, c(\vec{p}_c + \beta \vec{1}), \vec{p} - \vec{p}_c \big) \le 0 \tag{3.10}$$

for all $-\beta \vec{1} \le \vec{p} \le \beta \vec{1}$. From (3.1) we have

$$\frac{1}{c}|\nabla \vec{p}_c|^2 + |\operatorname{div} \vec{p}_c + \mathrm{K}^* f|_B^2 + \gamma |\mathrm{P}_{\mathrm{div}} \vec{p}_c|^2 + \frac{1}{c}|\vec{\lambda}_c|^2 \le |\mathrm{K}^* f|_B^2. \tag{3.11}$$

Consequently, $\frac{1}{c}|\vec{\lambda}_c|^2 \le |\mathrm{K}^* f|_B^2$ for all $c > 0$. Note that

$$\frac{1}{c}|\vec{\lambda}_c|_{\mathbb{L}^2(\Omega)}^2 = c|\max(0, \vec{p}_c - \beta \vec{1})|_{\mathbb{L}^2(\Omega)}^2 + c|\min(0, \vec{p}_c + \beta \vec{1})|_{\mathbb{L}^2(\Omega)}^2$$

and thus

$$\begin{aligned} |\max(0, (\vec{p}_c - \beta \vec{1}))|_{\mathbb{L}^2(\Omega)}^2 &\overset{c \to \infty}{\to} 0, \\ |\min(0, (\vec{p}_c + \beta \vec{1}))|_{\mathbb{L}^2(\Omega)}^2 &\overset{c \to \infty}{\to} 0. \end{aligned} \tag{3.12}$$

Recall that $\vec{p}_c \rightharpoonup \vec{p}^*$ weakly in $\mathbb{L}^2(\Omega)$. Weak lower semicontinuity of the convex functional $\vec{p} \mapsto |\max(0, \vec{p} - \beta\vec{1})|_{\mathbb{L}^2(\Omega)}$ and (3.12) imply that

$$\int_\Omega |\max(0, \vec{p}^* - \beta\vec{1})|^2 dx \le \liminf_{c\to\infty} \int_\Omega |\max(0, \vec{p}_c - \beta\vec{1})|^2 dx = 0.$$

Consequently, $\vec{p}^* \le \beta\vec{1}$, and analogously one verifies that $-\beta\vec{1} \le \vec{p}^*$. In particular, $\vec{p}^*$ is feasible, and from (3.10) we conclude that

$$\langle \vec{\lambda}_c, \vec{p}^* - \vec{p}_c \rangle_{H_0(\mathrm{div})^*, H_0(\mathrm{div})} \le 0 \quad \text{for all } c > 0. \tag{3.13}$$

By optimality of $\vec{p}_c$ for (3.1) we have

$$\limsup_{c\to\infty} \left(\frac{1}{2} |\operatorname{div} \vec{p}_c + \mathrm{K}^* f|_B^2 + \frac{\gamma}{2} |\mathrm{P}_{\mathrm{div}} \vec{p}_c|^2 \right) \le \frac{1}{2} |\operatorname{div} \vec{p} + \mathrm{K}^* f|_B^2 + \frac{\gamma}{2} |\mathrm{P}_{\mathrm{div}} \vec{p}\,|^2 \tag{3.14}$$

for all $\vec{p} \in S_2 = \{\vec{p} \in (C_0^1(\Omega))^2 : -\beta\vec{1} \le \vec{p} \le \beta\vec{1}\}$. Density of S_2 in $S_1 = \{\vec{p} \in H_0(\mathrm{div}) : -\beta\vec{1} \le \vec{p} \le \beta\vec{1}\}$ in the norm of $H_0(\mathrm{div})$ implies that (3.14) holds for all $\vec{p} \in S_1$ and consequently

$$\begin{aligned} \limsup_{c\to\infty} \left(\frac{1}{2} |\operatorname{div} \vec{p}_c + \mathrm{K}^* f|_B^2 + \frac{\gamma}{2} |\mathrm{P}_{\mathrm{div}} \vec{p}_c|^2 \right) &\le \frac{1}{2} |\operatorname{div} \vec{p}^* + \mathrm{K}^* f|_B^2 + \frac{\gamma}{2} |\mathrm{P}_{\mathrm{div}} \vec{p}^*|^2 \\ &\le \liminf_{c\to\infty} \left(\frac{1}{2} |\operatorname{div} \vec{p}_c + \mathrm{K}^* f|_B^2 + \frac{\gamma}{2} |\mathrm{P}_{\mathrm{div}} \vec{p}_c|^2 \right), \end{aligned}$$

where for the last inequality weak lower semicontinuity of norms is used. The above inequalities together with weak convergence of $\vec{p}_c$ to $\vec{p}^*$ in $H_0(\mathrm{div})$ imply strong convergence of $\vec{p}_c$ to $\vec{p}^*$ in $H_0(\mathrm{div})$. Finally we aim at passing to the limit in (3.13). This is impeded by the fact that we only established $\vec{\lambda}_c \rightharpoonup \vec{\lambda}^*$ in $\mathbb{H}_0^1(\Omega)^*$. Note from (3.2) that $\{-\frac{1}{c}\Delta\vec{p}_c + \vec{\lambda}_c\}_{c\ge 1}$ is bounded in $H_0(\mathrm{div})$. Hence there exists $\vec{\mu}^* \in H_0(\mathrm{div})^*$ such that

$$-\frac{1}{c}\Delta\vec{p}_c + \vec{\lambda}_c \rightharpoonup \vec{\mu}^* \quad \text{weakly in } H_0(\mathrm{div})^*,$$

and consequently also in $\mathbb{H}_0^1(\Omega)^*$. Moreover, $\{\frac{1}{\sqrt{c}} |\nabla \vec{p}_c|\}_{c\ge 1}$ is bounded and hence

$$-\frac{1}{c}\Delta\vec{p}_c \rightharpoonup 0 \quad \text{weakly in } \mathbb{H}_0^1(\Omega)^*$$

as $c \to \infty$. Since $\vec{\lambda}_c \rightharpoonup \vec{\lambda}^*$ weakly in $\mathbb{H}_0^1(\Omega)^*$, it follows that

$$\langle \vec{\lambda}^* - \vec{\mu}^*, \vec{v} \rangle_{\mathbb{H}_0^1(\Omega)^*, \mathbb{H}_0^1(\Omega)} = 0 \quad \text{for all } \vec{v} \in \mathbb{H}_0^1(\Omega).$$

Since both $\vec{\lambda}^*$ and $\vec{\mu}^*$ are elements of $H_0(\mathrm{div})^*$ and since $\mathbb{H}_0^1(\Omega)$ is dense in $H_0(\mathrm{div})$, it follows that $\vec{\lambda}^* = \vec{\mu}^*$ in $H_0(\mathrm{div})^*$. For $\vec{p} \in S_2$ we have

$$\begin{aligned} \langle \vec{\lambda}^*, \vec{p} - \vec{p}^* \rangle_{H_0(\mathrm{div})^*, H_0(\mathrm{div})} &= \langle \mu^*, \vec{p} - \vec{p}^* \rangle_{H_0(\mathrm{div})^*, H_0(\mathrm{div})} \\ &= \lim_{c\to\infty} \left\langle -\frac{1}{c}\Delta\vec{p}_c + \vec{\lambda}_c, \vec{p} - \vec{p}_c \right\rangle_{H_0(\mathrm{div})^*, H_0(\mathrm{div})} \\ &= \lim_{c\to\infty} \left(\frac{1}{c} (\nabla\vec{p}_c, \nabla(\vec{p} - \vec{p}_c)) + (\vec{\lambda}_c, \vec{p} - \vec{p}_c) \right) \\ &\le \lim_{c\to\infty} \left(\frac{1}{c} (\nabla\vec{p}_c, \nabla\vec{p}) + (\vec{\lambda}_c, \vec{p} - \vec{p}_c) \right) \le 0 \end{aligned}$$

by (3.10) and (3.11). Since S_2 is dense in S_1, we find

$$\langle \vec{\lambda}^*, \vec{p} - \vec{p}^* \rangle_{H_0(\mathrm{div})^*, H_0(\mathrm{div})} \leq 0 \quad \text{for all } \vec{p} \in S_1. \qquad \square$$

The problem formulation (3.1) contains two limiting processes: the $\mathbb{H}_0^1(\Omega)$ smoothing, which will be used to guarantee superlinear convergence of semismooth Newton methods applied to the first order optimality conditions (3.2) of (3.1) in function spaces (see section 4), and a penalization of the constraints $-\beta\vec{1} \leq \vec{p} \leq \beta\vec{1}$ resulting in the max- and min-terms. There is no need to utilize the same parameter c for both limiting processes. Rather, if $\frac{1}{2c}|\nabla\vec{p}|^2$ is replaced by $\frac{1}{2\bar{c}}|\nabla\vec{p}|^2$, then $(\vec{p}_{\bar{c},c}, \vec{\lambda}_{\bar{c},c})$ converges to $(\vec{p}^*, \vec{\lambda}^*)$ weakly in $H_0(\mathrm{div}) \times \mathbb{H}_0^1(\Omega)^*$, where $(\vec{p}_{\bar{c},c}, \vec{\lambda}_{\bar{c},c})$ denotes the solution of (3.2) with $\frac{1}{c}\Delta\vec{p}_c$ replaced by $\frac{1}{\bar{c}}\Delta\vec{p}_{\bar{c},c}$, as $c \to \infty$ and $\bar{c} \to \infty$.

4. Semismooth Newton methods. Here we shall describe two algorithms, one for a discretized form of (2.2) and another one for (3.1). Both algorithms are locally superlinearly convergent.

First we consider the unregularized problem (2.2). After discretization it is of the form

$$\begin{cases} \min \quad \dfrac{1}{2}|A_1 p + \tilde{f}|^2 + \dfrac{\gamma}{2}|A_2 p|^2 \\ \text{s.t.} \ -\beta\mathbf{1} \leq p \leq \beta\mathbf{1}, \end{cases} \tag{4.1}$$

where $p \in \mathbb{R}^m$, for some $m \in \mathbb{N}$ with coordinates p_i. Further, A_1, A_2 are $m \times m$-matrices, $\tilde{f} \in \mathbb{R}^m$, and $\mathbf{1} \in \mathbb{R}^m$ denotes the vector with all entries equal to 1. We assume that $\ker A_1 \cap \ker A_2 = 0$. The optimality condition for (4.1) is given by

$$\begin{aligned} &A_1^T A_1 p + \gamma A_2^T A_2 p + A_1^T \tilde{f} + \lambda = 0, \\ &\lambda = \max(0, \lambda + c(p - \beta\mathbf{1})) + \min(0, \lambda + c(p + \beta\mathbf{1})), \end{aligned} \tag{4.2}$$

where $c > 0$ is arbitrary and fixed. The primal-dual active set strategy, or equivalently the semismooth Newton algorithm applied to (4.2), is specified next.

ALGORITHM A.

(1) *Choose* $p_0, \lambda_0 \in \mathbb{R}^m$ *and set* $k = 0$.

(2) *Define*

$$\begin{aligned} \mathcal{A}_{k+1}^+ &= \{i : (\lambda_k + c(p_k - \beta\mathbf{1}))_i > 0\}, \\ \mathcal{A}_{k+1}^- &= \{i : (\lambda_k + c(p_k + \beta\mathbf{1}))_i < 0\}, \\ \mathcal{I}_{k+1}^i &= \{i : i \notin \mathcal{A}_{k+1}^\pm\}. \end{aligned}$$

(3) *Solve for* p_{k+1}, λ_{k+1}

$$\begin{aligned} &A_1^T A_1 p_{k+1} + \gamma A_2^T A_2 p_{k+1} + A_1^T \tilde{f} + \lambda_{k+1} = 0, \\ &(\lambda_{k+1})_i = 0 \text{ for } i \in \mathcal{I}_{k+1}, \\ &(p_{k+1})_i = \beta \text{ for } i \in \mathcal{A}_{k+1}^+, \quad (p_{k+1})_i = -\beta \text{ for } i \in \mathcal{A}_{k+1}^-. \end{aligned}$$

(4) *Stop, or set* $k = k + 1$ *and go to* (2).

This algorithm can be obtained by applying a formal Newton step to (4.2), choosing as generalized derivative for the function $s \mapsto \max(0, s)$ the value 1 if $s \geq 0$ and 0 if $s < 0$, and making an analogous choice for $s \mapsto \min(0, s)$.

For the following result we suppose that, given p_0, the first equation in (4.2) is used to compute λ_0. Then we have the following result.

THEOREM 4.1. *If $|p_0 - p^*|_{\mathbb{R}^m}$ is sufficiently small, then the iterates $\{(p_k, \lambda_k)\}_{k=1}^{\infty}$ of Algorithm A converge superlinearly to the solution (p^*, λ^*) of* (4.2).

The result can be verified by standard techniques from semismooth Newton methods; see, e.g., [17]. We do not enter into the details here but rather for Algorithm B below, where they are more involved.

We turn to the algorithmic treatment of the infinite-dimensional problem (3.1), for which we propose the following algorithm.

ALGORITHM B.

(1) *Choose $\vec{p}_0 \in \mathbb{H}_0^1(\Omega)$ and set $k = 0$.*

(2) *Set, for $i = 1, 2$,*

$$\begin{aligned} \mathcal{A}_{k+1}^{+,i} &= \{x : (\vec{p}_k^{\,i} - \beta\vec{1})(x) > 0\}, \\ \mathcal{A}_{k+1}^{-,i} &= \{x : (\vec{p}_k^{\,i} + \beta\vec{1})(x) < 0\}, \\ \mathcal{I}_{k+1}^{i} &= \Omega \setminus (\mathcal{A}_{k+1}^{+,i} \cup \mathcal{A}_{k+1}^{-,i}). \end{aligned}$$

(3) *Solve for $\vec{p} \in \mathbb{H}_0^1(\Omega)$ and set $\vec{p}_{k+1} = \vec{p}$, where*

$$\begin{aligned} \frac{1}{c}(\nabla\vec{p}, \nabla\vec{v}\,) + (\operatorname{div}\vec{p}, \operatorname{div}\vec{v}\,)_B + (\mathrm{K}^* f, \operatorname{div}\vec{v}\,)_B + \gamma(\mathrm{P}_{\mathrm{div}}\vec{p}, \mathrm{P}_{\mathrm{div}}\vec{v}\,) \\ + (c(\vec{p} - \beta\vec{1}\,)\chi_{\mathcal{A}_{k+1}^{+}}, \vec{v}\,) + (c(\vec{p} + \beta\vec{1}\,)\chi_{\mathcal{A}_{k+1}^{-}}, \vec{v}\,) = 0 \end{aligned} \tag{4.3}$$

for all $\vec{v} \in \mathbb{H}_0^1(\Omega)$.

(4) *Set*

$$\vec{\lambda}_{k+1}^{i} = \begin{cases} 0 & \text{on } \mathcal{I}_{k+1}^{i}, \\ c(\vec{p}_{k+1}^{\,i} - \beta\vec{1}) & \text{on } \mathcal{A}_{k+1}^{+,i}, \\ c(\vec{p}_{k+1}^{\,i} + \beta\vec{1}) & \text{on } \mathcal{A}_{k+1}^{-,i}, \end{cases}$$

for $i = 1, 2$.

(5) *Stop, or set $k = k + 1$ and go to* (2).

In the above, $\chi_{\mathcal{A}_{k+1}^{+}}$ stands for

$$\chi_{\mathcal{A}_{k+1}^{+}}^{i} = \begin{cases} 1 & \text{if } x \in \mathcal{A}_{k+1}^{+,i}, \\ 0 & \text{if } x \notin \mathcal{A}_{k+1}^{+,i}, \end{cases}$$

and analogously for $\mathcal{A}_{k+1}^{-}$. The superscript i, $i = 1, 2$, refers to the respective component. We note that (4.3) admits a solution $\vec{p}_{k+1} \in \mathbb{H}_0^1(\Omega)$. Step (4) is included for the sake of the analysis of the algorithm. Let $C : \mathbb{H}_0^1(\Omega) \to H^{-1}(\Omega) \times H^{-1}(\Omega)$ stand for the operator

$$C = -\frac{1}{c}\Delta - \nabla B^{-1} \operatorname{div} + \gamma \mathrm{P}_{\mathrm{div}}.$$

It is a homeomorphism for every $c > 0$ and allows us to express (3.2) as

$$C\vec{p} - \nabla B^{-1}\mathrm{K}^* f + c\max(0, \vec{p} - \beta\vec{1}) + c\min(0, \vec{p} + \beta\vec{1}) = 0, \tag{4.4}$$

where we drop the index in the notation for $\vec{p}_c$. For $\varphi \in \mathrm{L}^2(\Omega)$ we define

$$\mathrm{Dmax}(0,\varphi)(x) = \begin{cases} 1 & \text{if } \varphi(x) > 0, \\ 0 & \text{if } \varphi(x) \le 0, \end{cases} \tag{4.5}$$

and

$$\mathrm{Dmin}(0,\varphi)(x) = \begin{cases} 1 & \text{if } \varphi(x) < 0, \\ 0 & \text{if } \varphi(x) \ge 0. \end{cases} \tag{4.6}$$

These operators define semismooth derivatives; see, e.g., [22] for the finite-dimensional case and [17, 24] for the infinite dimensions. Using (4.5), (4.6) as generalized derivatives for the max and the min operations in (4.4), the semismooth Newton step can be expressed as

$$C\vec{p}_{k+1} + c(\vec{p}_{k+1} - \beta\vec{1})\chi_{\mathcal{A}^+_{k+1}} + c(\vec{p}_{k+1} + \beta\vec{1})\chi_{\mathcal{A}^-_{k+1}} - \nabla B^{-1}\mathrm{K}^* f = 0, \tag{4.7}$$

and $\vec{\lambda}_{k+1}$ from step (4) of Algorithm B is given by

$$\vec{\lambda}_{k+1} = c(\vec{p}_{k+1} - \beta\vec{1})\chi_{\mathcal{A}^+_{k+1}} + c(\vec{p}_{k+1} + \beta\vec{1})\chi_{\mathcal{A}^-_{k+1}}. \tag{4.8}$$

The iteration of Algorithm B can also be expressed with respect to the variable $\vec{\lambda}$ rather than $\vec{p}$. For this purpose we define

$$F(\vec{\lambda}) = \vec{\lambda} - c\max(0, C^{-1}(\nabla\hat{f} - \vec{\lambda}) - \beta\vec{1}) - c\min(0, C^{-1}(\nabla\hat{f} - \vec{\lambda}) + \beta\vec{1}), \tag{4.9}$$

where we put $\hat{f} = B^{-1}\mathrm{K}^* f$. Setting $\vec{p}_k = C^{-1}(\nabla\hat{f} - \vec{\lambda}_k)$, the semismooth Newton step applied to $F(\vec{\lambda}) = 0$ at $\vec{\lambda} = \vec{\lambda}_k$ results in

$$\vec{\lambda}_{k+1} = c(C^{-1}(\nabla\hat{f} - \vec{\lambda}_{k+1}) - \beta\vec{1})\chi_{\mathcal{A}^+_{k+1}} + c(C^{-1}(\nabla\hat{f} - \vec{\lambda}_{k+1}) + \beta\vec{1})\chi_{\mathcal{A}^-_{k+1}},$$

which coincides with (4.8). Therefore the semismooth Newton iterations according to Algorithm B and for $F(\vec{\lambda}) = 0$ coincide, provided that the initializations are related by $C\vec{p}_0 - \nabla\hat{f} + \vec{\lambda}_0 = 0$. The mapping F is slantly differentiable; i.e., for every $\vec{\lambda} \in \mathbb{L}^2(\Omega)$

$$|F(\vec{\lambda} + \vec{h}) - F(\vec{\lambda}) - DF(\vec{\lambda} + \vec{h})h|_{\mathbb{L}^2(\Omega)} = o(|\vec{h}|_{\mathbb{L}^2(\Omega)}) \tag{4.10}$$

for $|\vec{h}|_{\mathbb{L}^2(\Omega)} \to 0$ (see [17]). Here D denotes the derivative of F defined by means of (4.5) and (4.6). For (4.10) to hold the smoothing property of C^{-1} in the sense of an embedding from $\mathbb{L}^2(\Omega)$ into $\mathbb{L}^p(\Omega)$ for some $p > 2$ is essential. The following result now follows from standard arguments.

THEOREM 4.2. *If $|\vec{\lambda}_c - \vec{\lambda}_0|_{\mathbb{L}^2(\Omega)}$ is sufficiently small, then the iterates $\{(\vec{p}_k, \vec{\lambda}_k)\}_{k=1}^{\infty}$ of Algorithm B converge superlinearly in $\mathbb{H}^1_0(\Omega) \times \mathbb{L}^2(\Omega)$ to the solution $(\vec{p}_c, \vec{\lambda}_c)$ of* (3.1).

5. Discretization and numerical examples. We report now on numerical results attained by Algorithms A and B. In the examples below we choose $K = \mathrm{I}$ and $\alpha = 0$ for image denoising. We also include results for image zooming; see, e.g., [20] for a general description. In this case, we use given data f which correspond to a coarse (low-pixel–based) approximation of a given image. Then the aim is to

reconstruct the original image at the original pixel-scale. As a consequence, $K \neq \mathrm{I}$ with $\ker(K)$ typically nontrivial, which requires us to choose $\alpha > 0$.

In our tests, for the div-operator we use backward differences with quadratic extrapolation on the left boundary for Algorithm B. For Algorithm A we use symmetric differences with quadratic extrapolation on the boundary, where A_1 denotes the discretized divergence operator. The discrete grad $-$ div-operator is taken as $A_1^T A_1$. For Algorithm B we need the discrete Laplacian with homogeneous Dirichlet boundary conditions. We use the standard five-point stencil for its discretization. The projection $P_{\mathrm{div}}\vec{p}$ is obtained by solving a Neumann problem as stated at the end of section 1. Again we use the five-point stencil for discretizing the Laplace operator with symmetric differences for the discretization of the Neumann boundary condition.

To investigate possible ill-conditioning due to the parameter c appearing in Algorithm B we also tested a first order augmented Lagrangian variant of Algorithm B. To specify the algorithm we define

$$L(\vec{p},\vec{\lambda}) = \frac{1}{2\bar{c}}|\nabla\vec{p}\,|^2 + \frac{1}{2}|\operatorname{div}\vec{p} + \mathrm{K}^* f|_B^2 + \tfrac{\gamma}{2}|\mathrm{P}_{\mathrm{div}}\vec{p}\,|^2 + \phi_c(\vec{p},\vec{\lambda}),$$

where ϕ_c is the generalized Moreau–Yosida regularization of the indicator function ϕ of the set $\{\vec{p} \in \mathbb{L}^2(\Omega) : -\beta\vec{1} \le \vec{p} \le \beta\vec{1}\}$. We have

$$\phi_c(\vec{p},\vec{\lambda}) = \inf_{\vec{q}\in\mathbb{L}^2(\Omega)} \phi(\vec{p}-\vec{q}\,) + (\vec{\lambda},\vec{q}\,)_{\mathbb{L}^2(\Omega)} + \frac{c}{2}|\vec{q}\,|^2_{\mathbb{L}^2(\Omega)}$$

for $c > 0$ and $\vec{\lambda} \in \mathbb{L}^2(\Omega)$. Some simple manipulations result in

$$\begin{aligned}\phi_c(\vec{p},\vec{\lambda}) = \frac{1}{2c}\big|\max\big(0,\vec{\lambda}+c(\vec{p}-\beta\vec{1})\big)\big|^2_{\mathbb{L}^2(\Omega)} \\ + \frac{1}{2c}\big|\min\big(0,\vec{\lambda}+c(\vec{p}+\beta\vec{1})\big)\big|^2_{\mathbb{L}^2(\Omega)} - \frac{1}{2c}|\vec{\lambda}|^2_{\mathbb{L}^2(\Omega)}.\end{aligned}$$

Augmented Lagrangian Method (ALM).

(1) *Choose* $\vec{\lambda}_0 \in \mathbb{L}^2(\Omega)$, $c > 0$, *and* $n = 0$.

(2) *Given* $\vec{\lambda}_n \in \mathbb{L}^2(\Omega)$, *determine*

$$\vec{p}_n = \operatorname{argmin}\{L(\vec{p},\vec{\lambda}_n) : \vec{p} \in \mathbb{L}^2(\Omega)\}.$$

(3) *Update* $\vec{\lambda}_n$ *by* $\vec{\lambda}_{n+1} = \phi_c'(\vec{p}_n,\vec{\lambda}_n)$.

(4) *If convergence is not achieved, set* $n = n+1$ *and go to step* (2).

In step (3) we have

$$\phi_c'(\vec{p},\vec{\lambda}) = \max\big(0,\vec{\lambda}+c(\vec{p}-\beta\vec{1})\big) + \min\big(0,\vec{\lambda}+c(\vec{p}+\beta\vec{1})\big).$$

Note that the auxiliary problems in step (2) of ALM coincide with (3.1) except for the shift by $\vec{\lambda}_n$ in the max/min operations. In our numerical tests below we typically choose $\bar{c} = c$.

The algorithms in sections 3 and 4 are stated in terms of exact system solutions. Our numerical implementation utilizes inexact Newton techniques to underscore the feasibility of the proposed methods for large scale problems. In order to describe our approach let r_k denote the residual of the respective system, i.e., (4.2) for Algorithm A and (4.3) for Algorithm B. We resolve the respective system with the preconditioned conjugate gradient method (CG-method). The preconditioner involves the (vector)

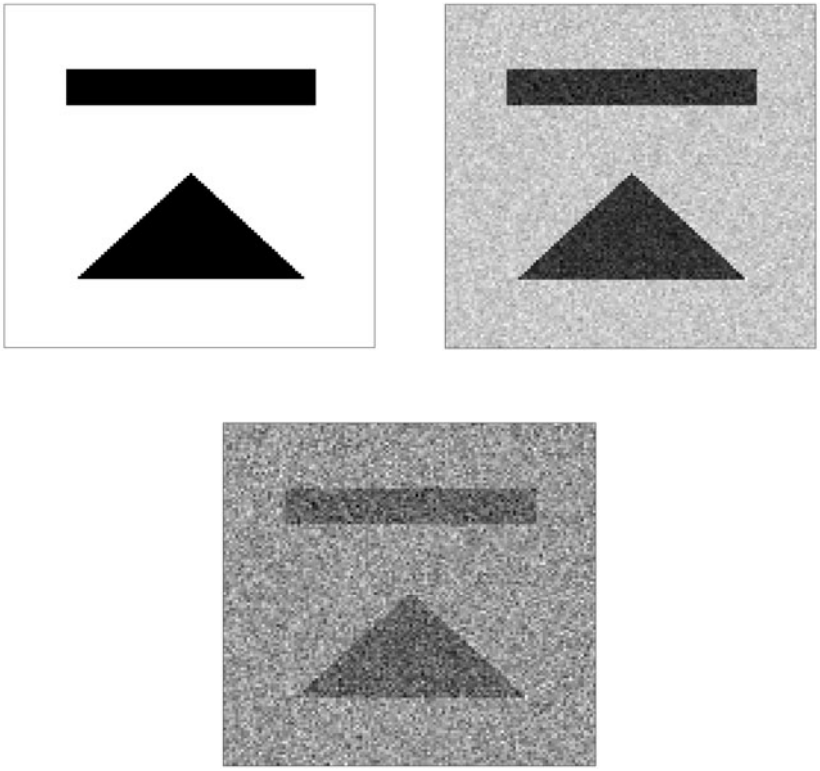

FIG. 1. *Upper left: original image (*128×128 *pixel). Upper right: noisy image with* 10% *noise. Lower: noisy image with* 50% *noise.*

Laplacian and, for Algorithm B, the terms involving the indicator functions of $\mathcal{A}^{\pm}_{k+1}$. The stopping tolerance for the CG-method in iteration $k+1$ is given by

$$\text{tol}_{k+1} = 0.1 \min(r_k^{1.25}, r_k).$$

This choice is motivated by the locally superlinear convergence rate of our algorithms.

Example 1. The test images for our first image denoising example are displayed in Figure 1. The upper left image is the original image, which is similar to the one in [8]. It has a dynamic range of $[0, 255]$. The other two images contain Gaussian white noise. The upper right one has 10% noise, and the remaining image contains 50% noise; i.e., we add Gaussian noise with standard deviation of 25.5 and 127.5, respectively. In the subsequent tables we denote by #as the total number of active set iterations, by #cg the total number of CG-iterations, and by #alm the total number of iterations updating $\vec{\lambda}_n$ for ALM. We stopped each algorithm as soon as the discrete L_2-norm of the residual dropped below `tol`$= \sqrt{\epsilon_M}$, with ϵ_M the machine precision, or when the difference between two successive residuals was smaller than `tol`, i.e., no further progress was observed.

Let us first report on the results obtained for denoising the image with 10% noise. For all algorithms we choose $c = 1E4$. However, let us note that Algorithm A does not require large c since c is not linked to a regularization term. Rather it is a parameter associated with the reformulation of the complementarity system induced by the box

TABLE 5.1
Results for 10% *noise.*

Algorithm	#as	#cg	#alm
Algorithm A	16	64	-
Algorithm B	9	23	-
ALM	14	27	3

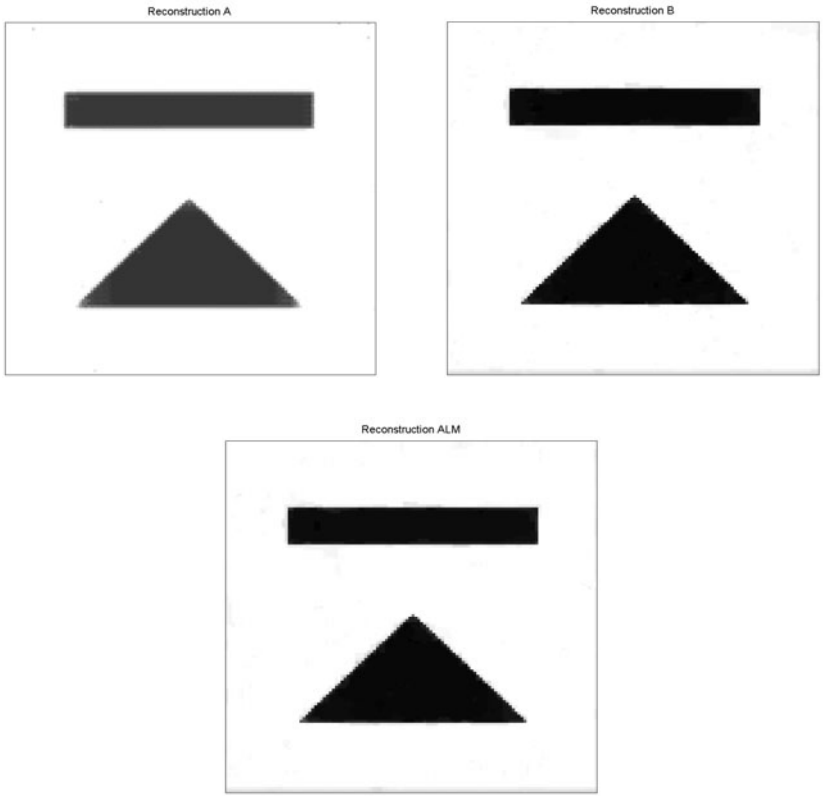

FIG. 2. *Upper left: result of Algorithm A. Upper right: result of Algorithm B. Lower: result of ALM.*

constraints. Further, for all three algorithms we chose $\beta = 0.2$, $\gamma = 0$ for ALM and Algorithm B, and $\gamma = 1E\text{-}3$ for Algorithm A. In general, for ALM and Algorithm B, γ had no noticeable effect on the results attained. However, Algorithm A is more sensitive to γ. This can be attributed to the fact that the system matrix in Algorithm A is singular for $\gamma = 0$. In Table 5.1 we report on the iteration numbers for the respective algorithms.

We note that Algorithm B requires the least number of AS-iterations. For ALM we point out that we initialized it with $\vec{\lambda}_0 \equiv 0$; then, typically, 8–10 AS-iterations were required in the first ALM-iteration. The subsequent ALM-iterations needed 2–3 AS-iterations.

In Figure 2 we display the reconstructions. The upper left and right correspond to Algorithms A and B. The lower image is the result obtained by Algorithm ALM. The quality of the reconstructions is equally good for all algorithms.

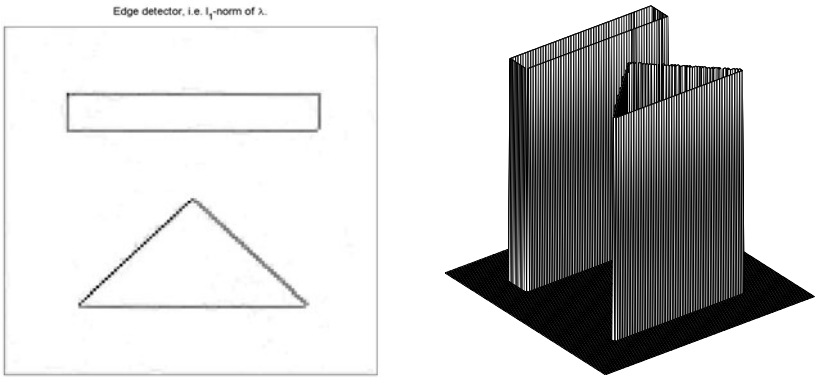

FIG. 3. *Left: Lagrange multiplier of Algorithm B. Right: corresponding edge detector.*

TABLE 5.2
Results for 50% *noise.*

Algorithm	#as	#cg	#alm
Algorithm A	15	59	-
Algorithm B	7	18	-
ALM	13	33	3

In the introduction we mentioned that the Lagrange multiplier associated with the box constraints serves as an edge detector. Figure 3 shows the ℓ_1-norm of the multiplier attained by Algorithm B and a resulting edge detector. The edge detector is obtained from a simple thresholding technique. In fact, as a threshold we took c and computed the edge detector λ_e as

$$\lambda_e(x_i) = \begin{cases} 1 & \text{if } |\vec{\lambda}^*(x_i)|_{\ell_1} \geq c, \\ 0 & \text{otherwise.} \end{cases}$$

In the above, x_i denotes the ith pixel of the image, and $\vec{\lambda}^*$ the multiplier upon termination of B. For the multipliers resulting from Algorithms A and ALM a similar observation holds true.

Now we turn to the results for the image containing 50% noise. The parameters had the values $c = 1E4$, $\beta = 0.9$, and $\gamma = 0$ for ALM and Algorithm B, and $c = 1E4$, $\beta = 0.75$, $\gamma = 1E$-3 for Algorithm A. Figure 4 shows the reconstructions obtained from our algorithms. As in the previous test case, the quality of the results for Algorithms B and ALM is comparable. Algorithm A appears to be slightly more sensitive to noise. This behavior could not be ruled out by tuning the parameters c, γ, and β. The iteration numbers are reported on in Table 5.2. As can be seen from these results, the number of iterations of the respective algorithm is rather stable with respect to the noise level.

In Figure 5 we display the ℓ_1-norm of the multiplier $\vec{\lambda}^*$ upon termination of Algorithm B. The related edge detector, which is obtained in the same way as explained previously, is given in the right image of Figure 5. We conclude that—without any thresholding—the Lagrange multiplier may act as an edge detector.

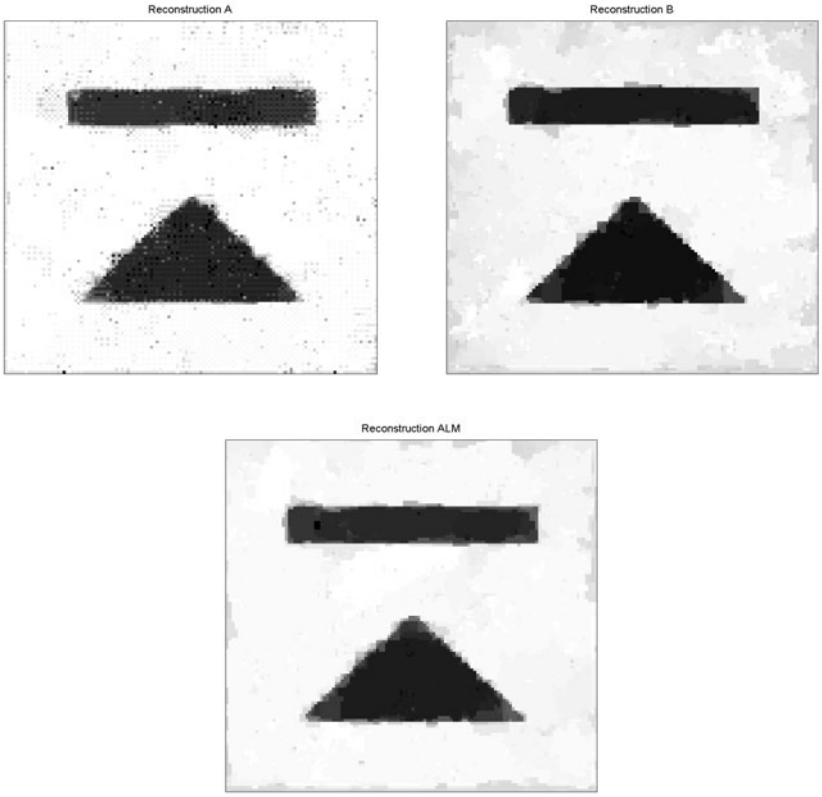

FIG. 4. *Upper left: result of Algorithm A. Upper right: result of Algorithm B. Lower: result of ALM.*

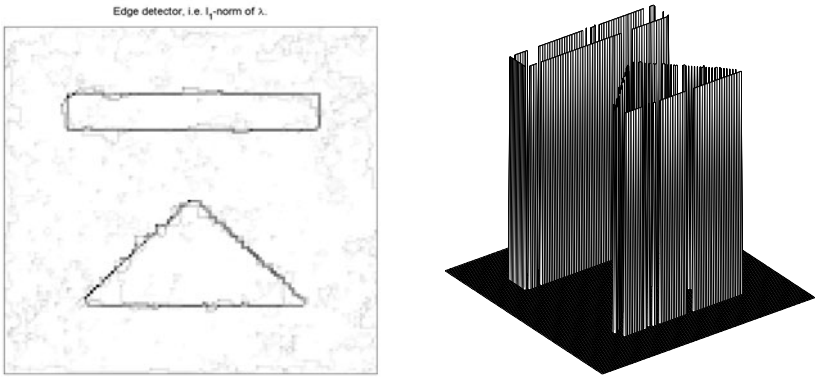

FIG. 5. *Left: Lagrange multiplier of Algorithm B. Right: corresponding edge detector.*

Let us briefly comment on the difference of stability with respect to β of Algorithms A and B compared to ALM. In general the choice of β influences the quality of the reconstruction. A large value for β decreases the number of active pixels, i.e., pixels at which $\vec{p}^*$ hits either the upper or the lower bound. As a consequence, details of the image are missed in the reconstruction. The right image in Figure 6 corresponds to the result attained by Algorithm B with $\beta = 1.5$ (compared to $\beta = 1.25$ in the

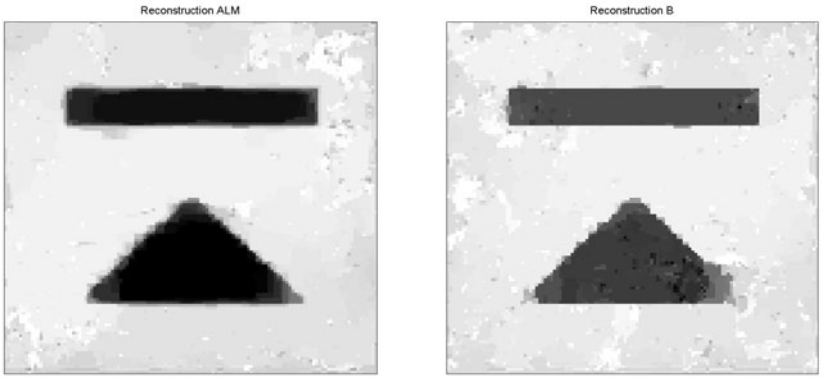

FIG. 6. *Reconstruction for $\beta = 1.5$. Left: ALM. Right: Algorithm B.*

TABLE 5.3
Convergence behavior of the residual.

#as	4.67E6	1.5E-1	9.6E-4	6.2E-7	1.8E-8	9.15E-9

previous run). Due to the larger β-value, the quality of the reconstruction degrades in the sense that details are missed, e.g., at the corners of the triangle. On the other hand, the left image in Figure 6 shows the result for ALM with $\beta = 1.5$.

Obviously the reconstruction is superior to the one obtained from Algorithm B. This reflects a general observation from our test runs, i.e., ALM is more stable with respect to the choice of β. The behavior of Algorithm A with respect to changes in β is comparable to that of Algorithm B.

Let us discuss the convergence behavior of our algorithms in terms of reductions of the residuals. From the results reported in Tables 5.1 and 5.2 we find that our algorithms require a rather small number of iterations which are even stable with respect to different noise levels. In Table 5.3 we show the behavior of the residual for Algorithm B for 50% noise, indicating a fast convergence. This fast convergence is also true for the numerical resolution of the auxiliary problem of ALM. A similar convergence behavior is obtained for Algorithm A. For smaller values of β the iterates converge superlinearly. Small values of β, however, imply a deterioration of the reconstruction. Here the ill-posedness in the problem becomes evident.

In [8] an inexact Newton method for solving a primal-dual formulation of the Euler–Lagrange equations associated with a regularized TV-based image reconstruction problem is proposed. The test problem in [8] involves the same geometry as in our test example. In Figure 7 (upper left plot) we show the noisy image containing Gaussian white noise with variance $\sigma^2 \approx 1200$, which gives a signal-to-noise ratio of approximately 1. This parallels the test setting in [8]. We also made an effort to adjust the stopping rule of Algorithm B for the comparison with the algorithm in [8]. Algorithm B requires nine iterations for obtaining the denoised image in the upper right plot of Figure 7. The algorithm in [8] with a line search and a continuation strategy with respect to δ in the regularization of the TV-seminorm of the type (1.2) is reported to need 12 iterations. The size of the systems which have to be solved per iteration in both algorithms is comparable. The edged detector, based on the ℓ_1-norm of $\vec{\lambda}$ upon termination of Algorithm B, is given in the last subplot of Figure 7.

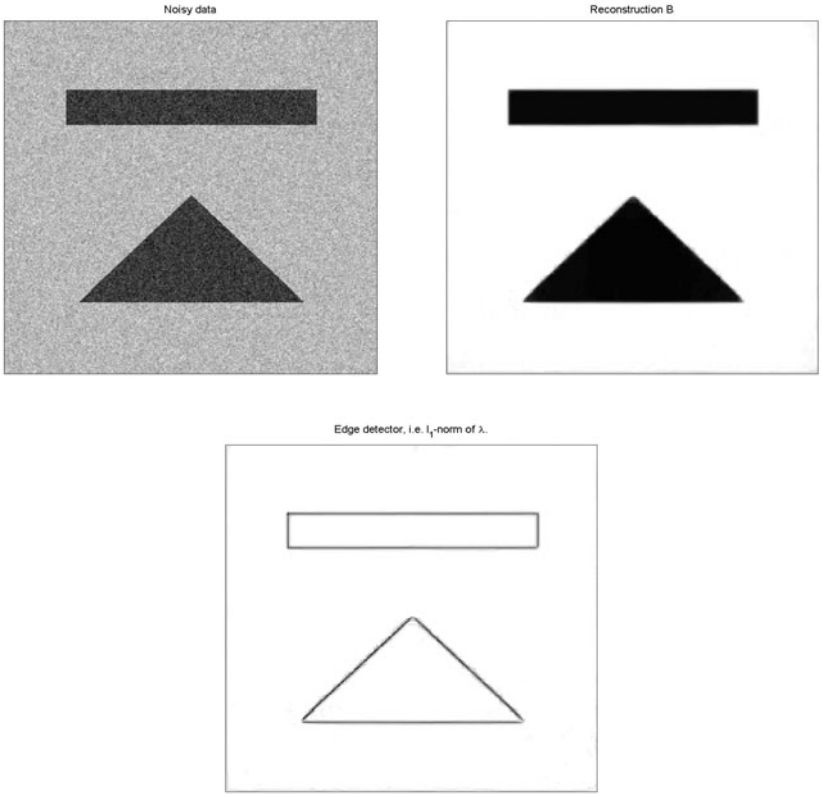

FIG. 7. *Upper left: noisy image (*256×256 *pixel). Upper right: result of Algorithm B. Lower:* ℓ_1*-norm of* $\vec{\lambda}$.

Example 2. Now we report on the behavior of Algorithm B for the benchmark problem in Figure 8. The upper left plot shows the original image. The upper right image contains 7.5% Gaussian white noise. The parameters had the values $c = 1E4$, $\gamma = 0$, and $\beta = 0.15$. The algorithm stopped after nine AS-iterations (31 CG-iterations total) with a residual of 6.2E-9. The corresponding reconstruction is given in the lower left plot of Figure 8. The lower right plot displays the ℓ_1-norm of the Lagrange multiplier associated with the box constraints. As in the previous examples, it behaves like an edge detector.

Example 3. We conclude our numerical section with the results obtained by Algorithm B for an image zooming/resizing problem. In this case, we have $K \neq \mathrm{I}$. The data f correspond to a *coarse* version of the original image satisfying $f_{2i-1,2j-1} = f_{2i,2j-1} = f_{2i-1,2j} = f_{2i,2j}$. For an arbitrary 256×256-pixel image u the application $v = Ku$ is related to a 128×128-pixel version $\tilde{v}$ of the image with $\tilde{v}_{i,j} = u_{2i-1,2j-1}$ and $v_{2i-1,2j-1} = v_{2i,2j-1} = v_{2i-1,2j} = v_{2i,2j} = \tilde{v}_{i,j}$. For more details on image zooming involving more advance operators K, we refer to [20]. Our aim is to use Algorithm B for reconstructing the fine image u from the given coarse image f. Since K has a nontrivial kernel, we choose $\alpha =$ 1E-10. Further, we pick the parameter values $c =$ 1E5, $\beta = 0.35$, and $\gamma = 0$. In Figure 9 we display the original image in the upper left plot. The result after 16 iterations of Algorithm B is shown in the upper right plot. The lower left plot shows the 128×128-pixel version expanded by a factor of 2, and

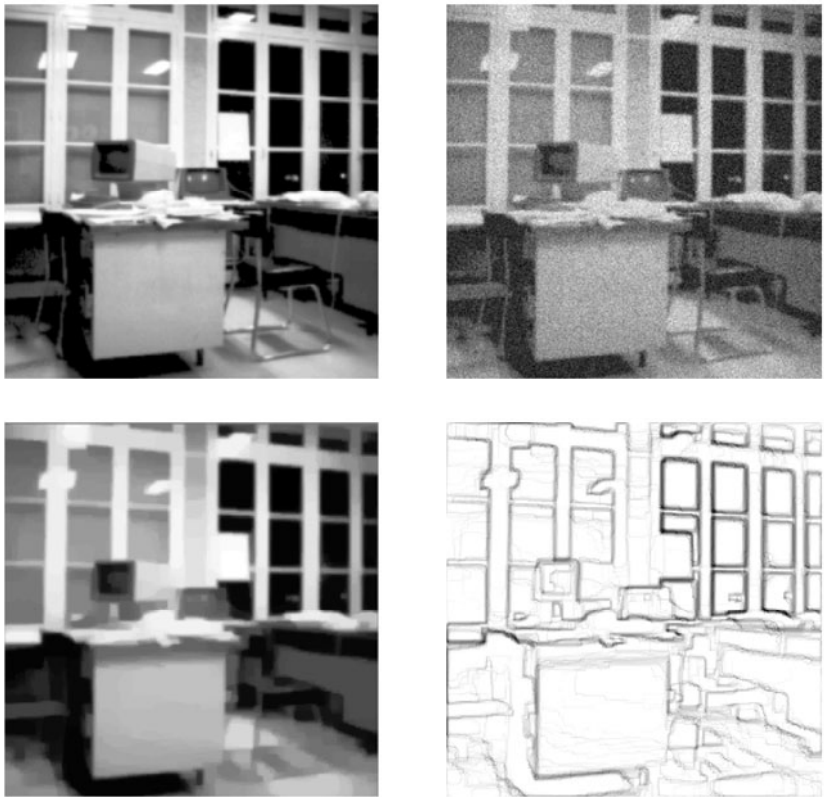

FIG. 8. *Upper left: exact data (*256×256 *pixel). Upper right: noisy data. Lower left: reconstruction obtained by Algorithm B. Lower right:* ℓ_1*-norm of the Lagrange multiplier.*

the lower right plot provides the result obtained by a nearest neighbor interpolation. Observe that the reconstructions differ quite noticeably along the boundaries of the person's left arm, for example.

6. Conclusions. The efficient numerical treatment of BV-regularization–based image restoration poses many challenges in theory as well as in the design of algorithms. In this paper we first establish the relationship between the primal problem in the nonreflexive Banach space BV and its predual which is posed in the Hilbert space $H_0(\mathrm{div})$. This analytical result appears to be of interest in its own right. We then introduce and study two semismooth Newton methods for solving the Fenchel predual problem of the underlying BV-regularized minimization problem. By predualization we obtain a box constrained minimization problem which—from the numerical optimization point of view—has the advantage that we can rely on sophisticated minimization algorithms. The convergence analysis of our semismooth Newton methods in function spaces relies on a smoothing procedure. The regularizing effect of our smoothing results in a two-norm property which is required for arguing locally superlinear convergence of our semismooth Newton method in an L^2-setting. Without smoothing we obtain a locally superlinearly convergent method on the discrete level.

Acknowledgment. We would like to thank Prof. O. Scherzer, University of Innsbruck, Austria, for making us aware of the taut string algorithm discussed in Remark 1.

FIG. 9. *Upper left: exact data (*256×256 *pixel). Upper right: reconstruction obtained by Algorithm B. Lower left:* 128×128*-pixel image expanded by a factor of* 2*. Lower right: result using a nearest neighbor interpolation technique.*

REFERENCES

[1] R. ACAR AND C. VOGEL, *Analysis of bounded variation penalty methods for ill-posed problems*, Inverse Problems, 10 (1994), pp. 1217–1229.

[2] G. AUBERT AND L. VESE, *A variational method in image recovery*, SIAM J. Numer. Anal., 34 (1997), pp. 1948–1979.

[3] V. BARBU AND T. PRECUPANU, *Convexity and Optimization in Banach Spaces*, 2nd rev. and extended ed., Mathematics and Its Applications (East European Series) 10, translated from the Romanian, D. Reidel, Dordrecht/Boston/Lancaster, Editura Academiei, Bucuresti, 1986.

[4] P. CARBONNIER, L. BLAUD-FÉRAND, G. AUBERT, AND M. BARLAUD, *Deterministic edge-preserving regularisation in computed imaging*, IEEE Trans. Image Process., 6 (1997), pp. 298–311.

[5] E. CASAS, K. KUNISCH, AND C. POLA, *Regularization by functions of bounded variation and applications to image enhancement*, Appl. Math. Optim., 40 (1999), pp. 229–257.

[6] F. CATTÉ, P.-L. LIONS, J.-M. MOREL, AND T. COLL, *Image selective smoothing and edge detection by nonlinear diffusion*, SIAM J. Numer. Anal., 29 (1992), pp. 182–193.

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

[8] 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.
[9] G. Chavent and K. Kunisch, *Regularization of linear least squares problems by total bounded variation*, ESAIM Control Optim. Calc. Var., 2 (1997), pp. 359–376.
[10] R. Dautray and J.-L. Lions, *Mathematical Analysis and Numerical Methods for Science and Technology. Vol.* 3*: Spectral Theory and Applications*, with the collaboration of Michel Artola and Michel Cessenat, translated from the French by John C. Amson, Springer-Verlag, Berlin, 2000.
[11] P. L. Davies and A. Kovac, *Local extremes, runs, strings and multiresolution*, Ann. Statist., 29 (2001), pp. 1–65.
[12] D. C. Dobson and F. Santosa, *Recovery of blocky images from noisy and blurred data*, SIAM J. Appl. Math., 56 (1996), pp. 1181–1198.
[13] I. Ekeland and R. Témam, *Convex Analysis and Variational Problems*, (corrected republication of the 1976 English original) Classics in Appl. Math. 28, SIAM, Philadelphia, 1999.
[14] D. Geman and C. Yang, *Nonlinear image recovery with half-quadratic regularization*, IEEE Trans. Image Process., 4 (1995), pp. 932–945.
[15] V. Girault and P.-A. Raviart, *Finite Element Methods for Navier–Stokes Equations. Theory and Algorithms*, Springer Ser. Comput. Math. 5, Springer-Verlag, Berlin, 1986.
[16] E. Giusti, *Minimal Surfaces and Functions of Bounded Variation*, Monogr. Math. 80, Birkhäuser, Boston, Basel, Stuttgart, 1984.
[17] M. Hintermüller, K. Ito, and K. Kunisch, *The primal-dual active set strategy as a semismooth Newton method*, SIAM J. Optim., 13 (2003), pp. 865–888.
[18] K. Ito and K. Kunisch, *An active set strategy based on the augmented Lagrangian formulation for image restoration*, M2AN Math. Model. Numer. Anal., 33 (1999), pp. 1–21.
[19] K. Ito and K. Kunisch, *BV-type regularization methods for convoluted objects with edge, flat and grey scales*, Inverse Problems, 16 (2000), pp. 909–928.
[20] F. Malgouyres and F. Guichard, *Edge direction preserving image zooming: A mathematical and numerical analysis*, SIAM J. Numer. Anal., 39 (2001), pp. 1–37.
[21] E. Mammen and S. van de Geer, *Locally adaptive regression splines*, Ann. Statist., 25 (1997), pp. 387–413.
[22] L. Q. Qi and J. Sun, *A nonsmooth version of Newton's method*, Math. Programming, 58 (1993), pp. 353–367.
[23] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.
[24] M. Ulbrich, *Semismooth Newton methods for operator equations in function spaces*, SIAM J. Optim., 13 (2003), pp. 805–841.
[25] C. R. Vogel, *Computational Methods for Inverse Problems*, Frontiers Appl. Math. 23, SIAM, Philadelphia, 2002.
[26] C. R. Vogel and M. E. Oman, *Fast, robust total variation-based reconstruction of noisy, blurred images*, IEEE Trans. Image Process., 7 (1998), pp. 813–824.