Abstract. It is demonstrated that the pre-dual for problems with total bounded variation regularization terms can be expressed as bilaterally constrained optimization problem. Existence of a Lagrange multiplier and an optimality system are established. This allows 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 semi-smooth Newton method is analyzed and superlinear convergence is proved. As a by-product it is obtained 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, pre-dual, semi-smooth Newton methods, box constraints, image reconstruction.

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

1. Introduction and notation. This work is concerned with the study of the problem

\[
\begin{aligned}
\min_{u \in BV(\Omega)} \quad & \frac{1}{2} \int_{\Omega} |K u - f|^2 dx + \frac{\alpha}{2} \int_{\Omega} |u|^2 dx + \beta \int_{\Omega} |Du| \\
\end{aligned}
\]  

(1.1)

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

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

is finite. It is well-known that \( BV(\Omega) \subset L^2(\Omega) \) for \( \Omega \subset \mathbb{R}^2 \) [16], and that \( u \mapsto |u|_{L^2} + \int_{\Omega} |Du| \) defines a norm on \( BV(\Omega) \). If \( K = \text{identity} \) 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 that (1.1) admits a unique solution \( u^* \in BV(\Omega) \) [9]. 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. 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 of 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 non-reflexive 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 non-linear 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 non-reflexive Banach space problem (1.1) has a complicated measure theoretic structure, its pre-dual can be characterized in a well-known Hilbert space setting. Specifically the pre-dual 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 semi-smooth 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 post-processing on the multiplier such as thresholding or sharpening techniques.

Let us briefly mention a few alternatives that have been investigated to treat (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 \( \int_{\Omega} |Du| \) is replaced by

\[
\int_{\Omega} \sqrt{\delta + |\nabla u|^2} dx,
\]

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 we summarize the function space notation that will be used. In Section 2 we characterize the pre-dual of (1.1) in the sense of Fenchel. We shall point out the close connection, for 1D problems, between our algorithm and the taut-string algorithm well-known in non-parametric regression analysis [11, 21]. Section 3 is devoted to the description and convergence proof for a class of regularized problems. Semi-smooth Newton methods for the pre-dual problems are developed in Section 4. Superlinear convergence for the regularized infinite dimensional problems, and for the discretized pre-dual 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 \( \mathcal{F} : V \to \mathbb{R} \cup \{\infty\} \) and \( \mathcal{G} : Y \to \mathbb{R} \cup \{\infty\} \) be convex, lower semi-continuous functionals not identically equal to \( \infty \), 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 \( \Lambda v_0 \). Then we have

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

$$\Lambda^* \bar{p} \in \partial \mathcal{F}(\bar{u}),$$
$$- \bar{p} \in \partial \mathcal{G}(\Lambda \bar{u}),$$

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

To compute, formally, the Fenchel dual to (1.1) we set $\Lambda = \nabla$, $\mathcal{F}(u) = \frac{1}{2} |K u - f|^2 + \frac{\alpha}{2} |u|^2$ and $\mathcal{G}(\bar{p}) = \beta \int_{\Omega} |\bar{p}|_{1}, dx$,

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

$$\mathcal{F}^*(v) = \frac{1}{2} (v + K^* f, B^{-1} (v + K^* f)) - \frac{1}{2} |f|^2$$
and

$$\mathcal{G}^*(\bar{p}) = I_{[-\beta \vec{1}, \beta \vec{1}]}(\bar{p}),$$

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

$$I_{[-\beta \vec{1}, \beta \vec{1}]}(\bar{p}) = \begin{cases} 0 & \text{if } - \beta \vec{1} \leq \bar{p}(x) \leq \beta \vec{1} \text{ for a.e. } x \in \Omega, \\ \infty & \text{otherwise.} \end{cases}$$

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

$$\inf \left\{ \frac{1}{2} |\text{div} \bar{p} + K^* f|_B^2 \right\} \quad \text{s.t. } - \beta \vec{1} \leq \bar{p}(x) \leq \beta \vec{1} \text{ for a.e. } x \in \Omega,$$

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

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

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

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

\[
\mathbb{L}^2(\Omega) = \text{grad} \ H^1(\Omega) \oplus H_0(\text{div} 0),
\]

c.f. [10, p.216], for example. Moreover,

\[
H_0(\text{div}) = H_0(\text{div} 0)^\perp \oplus H_0(\text{div} 0),
\]

with

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

and \( \text{div} : H_0(\text{div} 0)^\perp \subset H_0(\text{div}) \rightarrow L^2_0(\Omega) \) is a homeomorphism. In fact, it is injective by construction and for every \( f \in L^2_0(\Omega) \) there exists, by the Lax-Milgram lemma, \( \varphi \in H^1(\Omega) \) such that

\[
\text{div} \nabla \varphi = f \text{ in } \Omega, \quad \nabla \varphi \cdot n = 0 \text{ on } \partial \Omega,
\]

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

\[
\text{div} \in \mathcal{L}(H_0(\text{div} 0)^\perp, L^2_0(\Omega)).
\]

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

\[\textbf{2. The Fenchel pre-dual.}\] The section is devoted to the study of the problems

\[
\left\{ \begin{array}{l}
\min \frac{1}{2} |\text{div} \vec{p} + K^* f|_{L^2_B}^2 \quad \text{over } \vec{p} \in H_0(\text{div}) \\
\text{s.t. } -\beta \vec{1} \leq \vec{p} \leq \beta \vec{1},
\end{array} \right. \quad (2.1)
\]

and

\[
\left\{ \begin{array}{l}
\min \frac{1}{2} |\text{div} \vec{p} + K^* f|_{L^2_B}^2 + \frac{\gamma}{2} |P_{\text{div}} \vec{p}|^2 \quad \text{over } \vec{p} \in H_0(\text{div}) \\
\text{s.t. } -\beta \vec{1} \leq \vec{p} \leq \beta \vec{1},
\end{array} \right. \quad (2.2)
\]

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

**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 \( J(\vec{p}) = \frac{1}{2} |\text{div} \vec{p} + K^* f|_{L^2_B}^2 + \frac{\gamma}{2} |P_{\text{div}} \vec{p}|^2 \). To ascertain strict convexity of \( J \) we use the fact that the second derivative satisfies

\[
J''(\vec{p}, \vec{p}) = |\text{div} \vec{p}|_{L^2_B}^2 + \gamma |P_{\text{div}} \vec{p}|^2 \geq \kappa |\vec{p}|_{H_0(\text{div})}^2
\]

for a constant \( \kappa > 0 \) independent of \( \vec{p} \in H_0(\text{div}) \). Here we 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

\[
Bu^* = \text{div} \vec{p}^* + K^* f
\]

\[
\langle (\text{div} u^*)^*, \vec{p} - \vec{p}^* \rangle_{H_0(\text{div})^*, H_0(\text{div})} \leq 0 \quad \text{for all } \vec{p} \in H_0(\text{div}), \quad (2.3)
\]

\[
\langle (\text{div} u^*)^*, \vec{p} - \vec{p}^* \rangle_{H_0(\text{div})^*, H_0(\text{div})} \leq 0 \quad \text{for all } \vec{p} \in H_0(\text{div}), \quad (2.4)
\]
with \(-\beta \tilde{I} \leq \tilde{p} \leq \beta \tilde{I}\). Alternatively (2.1) can be considered as the pre-dual of the original problem (1.1). If (2.1) is a zero-residue problem, i.e., \(\tilde{p}^*\) satisfies \(\text{div} \tilde{p}^* = -K^* f\), then the additional penalty term in (2.2) chooses among all solutions, the one which minimizes \(|P_{\text{div}} \tilde{p}^*|\).

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

\[
G^*(v) = \frac{1}{2}|Kv + f|^2 + \frac{\alpha}{2}|v|^2 - \frac{1}{2}|f|^2.
\]

Further the conjugate \(F^* : H_0(\text{div})^* \rightarrow \mathbb{R}\) of \(F\) is given by

\[
F^*(\tilde{q}) = \sup_{\tilde{p} \in S_1} \langle \tilde{q}, \tilde{p} \rangle |_{H_0(\text{div})^*, H_0(\text{div})}, \quad \text{for } \tilde{q} \in H_0(\text{div})^*, \tag{2.5}
\]

where \(S_1 = \{\tilde{p} \in H_0(\text{div}) : -\beta \tilde{I} \leq \tilde{p} \leq \beta \tilde{I}\}\). Let us set

\[
S_2 = \{\tilde{p} \in C_0^1(\Omega) \times C_0^1(\Omega) : -\beta \tilde{I} \leq \tilde{p} \leq \beta \tilde{I}\}.
\]

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

\[
|\tilde{p}_n - P\tilde{p}_n|_{H_0(\text{div})} \leq |\tilde{p}_n - \tilde{p}|_{H_0(\text{div})} + |\tilde{p} - P\tilde{p}_n|_{H_0(\text{div})} \leq 2|\tilde{p} - \tilde{p}_n|_{H_0(\text{div})} \rightarrow 0 \text{ for } n \rightarrow \infty.
\]

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

\[
F^*((- \text{div})^* v) = \sup_{\tilde{p} \in S_2} \langle v, - \text{div} \tilde{p} \rangle,
\]

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

\[
F^*((- \text{div})^* v) = \beta \int_{\Omega} |Dv| < \infty \quad \text{for } v \in BV(\Omega).
\]

The dual problem to (2.1) is found to be

\[
\min \frac{1}{2}|Ku - f|^2 + \frac{\alpha}{2}|u|^2 + \beta \int_{\Omega} |Du| \quad \text{over } u \in BV(\Omega).
\]

From (1.4) moreover we find

\[
\langle (- \text{div})^* u^*, \tilde{p} - \tilde{p}^* \rangle |_{H_0(\text{div})^*, H_0(\text{div})} \leq 0 \quad \text{for all } \tilde{p} \in S_1
\]

and

\[
Bu^* = \text{div} \tilde{p}^* + K^* f.
\]
We obtain the following optimality system.

**Corollary 2.3**. Let \( \bar{p}^* \in H_0(\text{div}) \) be a solution to (2.1). Then there exists \( \bar{\lambda}^* \in H_0(\text{div})^* \) such that

\[
\begin{align*}
\text{div}^* B^{-1} \text{div} \bar{p}^* + \text{div}^* B^{-1} K^* f + \bar{\lambda}^* &= 0 \quad (2.6) \\
\langle \bar{\lambda}^*, \bar{p} - \bar{p}^* \rangle_{H_0(\text{div}), H_0(\text{div})} &\leq 0 \quad \text{for all } \bar{p} \in H_0(\text{div}), \quad (2.7)
\end{align*}
\]

with \(-\beta \bar{\lambda} \leq \bar{p} \leq \beta \bar{\lambda}\). For convenience we also specify the variational form of equation (2.6) which holds in \( H_0(\text{div})^* \):

\[
\langle B^{-1} \text{div} \bar{p}^*, \text{div} \bar{v} \rangle_{L^2} + \langle B^{-1} K^* f, \text{div} \bar{v} \rangle_{L^2} + \langle \bar{\lambda}^*, \bar{v} \rangle_{H_0(\text{div}), H_0(\text{div})} = 0
\]

for all \( \bar{v} \in H_0(\text{div}) \).

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

The optimality system for (2.2) is given next.

**Corollary 2.4**. Let \( \bar{p}^* \in H_0(\text{div})^* \) denote the solution to (2.2). Then there exists \( \bar{\lambda}^* \in H_0(\text{div})^* \) such that

\[
\begin{align*}
\text{div}^* B^{-1} \text{div} \bar{p}^* + \text{div}^* B^{-1} K^* f + \gamma P_{\text{div}} \bar{p}^* + \bar{\lambda}^* &= 0, \quad (2.8) \\
\langle \bar{\lambda}^*, \bar{p} - \bar{p}^* \rangle_{H_0(\text{div}), H_0(\text{div})} &\leq 0 \quad \text{for all } \bar{p} \in H_0(\text{div}), \quad (2.9)
\end{align*}
\]

with \(-\beta \bar{\lambda} \leq \bar{p} \leq \beta \bar{\lambda}\).

**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 \( \bar{v} \in H_0(\text{div}) \) can be decomposed according to \( \bar{v} = \bar{v}_1 + \bar{v}_2 \in H_0(\text{div}0)^{\perp} \oplus H_0(\text{div}0) \). The functional in (2.2) is then separable and (2.2) can be expressed as

\[
\min_{\bar{p} \in H_0(\text{div})} \mathcal{F}(\bar{p}) + \mathcal{G}_1(\Lambda_1 \bar{p}_1) + \mathcal{G}_2(\Lambda_2 \bar{p}_2),
\]

where \( \mathcal{F} \) is defined in the proof of Theorem 2.2, \( \mathcal{G}_1 \) and \( \Lambda_1 \) coincide with \( \mathcal{G} \) and \( \Lambda \) from the proof of Theorem 2.2, and we set

\[
\mathcal{G}_2 : \mathbb{L}^2(\Omega) \rightarrow \mathbb{R}, \quad \mathcal{G}_2(\bar{p}) = \frac{\gamma}{2} |\bar{p}|_{\mathbb{L}^2(\Omega)}^2,
\]

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

\[
\begin{align*}
B \bar{u}_1^* &= \text{div} \bar{p}_1^* + K^* f = \text{div} \bar{p}^* + K^* f, \\
-\bar{u}_2^* &= \gamma \bar{p}_2^* = \gamma P_{\text{div}} \bar{p}^*,
\end{align*}
\]

and

\[
\langle -\text{div}^* \bar{u}_1^* + \bar{u}_2^* \rangle_{H_0(\text{div}), H_0(\text{div})} \leq 0 \quad \text{for all } \bar{p} \in S_1.
\]

The claim follows with \( \bar{\lambda}^* = -\text{div}^* \bar{u}_1^* + \bar{u}_2^* \). \( \square \)

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 \(-\beta \bar{\lambda} \leq \bar{p} \leq \beta \bar{\lambda}\) 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 \( K = I, \alpha = 0 \). A continuous version of the taut string algorithm can be expressed as
\[
\begin{align*}
\min \int_0^1 \sqrt{1 + w_x^2} \, dx,
\text{s.t. } F - \beta &\leq w \leq F + \beta,
\end{align*}
\] (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
\[
\begin{align*}
\min \int_0^1 \sqrt{1 + p_x^2 + f^2} \, dx,
\text{s.t. } -\beta &\leq p \leq \beta.
\end{align*}
\] (2.11)
and \( u = p_x + f \). Thus, except for the square root in (2.11), we obtain (2.1).

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

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 semi-smooth 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 \( \beta \) will typically be chosen as a small constant so that the resulting problems are close to bottle neck 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
\[
\begin{align*}
\min \frac{1}{c} |\nabla \tilde{p}_c|^2 + \frac{1}{2} |\text{div} \tilde{p}_c + K^* f|^2 + \frac{1}{2} |\text{P_{div} \tilde{p}_c}|^2 +
\frac{1}{c} |\max(0, c(\tilde{p}_c - \beta \tilde{1}))|^2 + \frac{1}{c} |\min(0, c(\tilde{p}_c + \beta \tilde{1}))|^2 \text{ over } \tilde{p}_c \in H^1_0(\Omega),
\end{align*}
\] (3.1)
where \( c > 0 \). Let \( \tilde{p}_c \) denote the unique solution to (3.1). It satisfies the optimality condition
\[
\begin{align*}
-\frac{1}{c} \Delta \tilde{p}_c - \nabla B^{-1} \text{div} \tilde{p}_c - \nabla B^{-1} K^* f + \gamma \text{P_{div} \tilde{p}_c} + \tilde{\lambda}_c &= 0, \\
\tilde{\lambda}_c &= \max(0, c(\tilde{p}_c - \beta \tilde{1})) + \min(0, c(\tilde{p}_c + \beta \tilde{1})).
\end{align*}
\] (3.2a)
(3.2b)
Next we address convergence as \( c \to \infty \).

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

**Proof.** Recall the variational form of (2.8) given by
\[
(\text{div} \tilde{p}^*, \text{div} \tilde{v})_B + (K^* f, \text{div} \tilde{v})_B + \gamma (\text{P_{div} \tilde{p}^*}, \text{P_{div} \tilde{v}}) + \langle \tilde{\lambda}^*, \tilde{v} \rangle_{H_0(\text{div})^* \times H_0(\text{div})} = 0,
\] (3.3)
for all \( \vec{v} \in H_0(\text{div}) \). To verify uniqueness, let us suppose that \( (\vec{p}_i, \vec{\lambda}_i) \in H_0(\text{div}) \times H_0(\text{div})^*, i = 1, 2 \), 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

\[
(B^{-1} \text{div} \delta \vec{p}, \text{div} \vec{v}) + \gamma (P_{\text{div}} \delta \vec{p}, P_{\text{div}} \vec{v}) + \langle \delta \vec{\lambda}, \vec{v} \rangle_{H_0(\text{div})^*, H_0(\text{div})} = 0,
\]

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

\[
\langle \delta \vec{\lambda}, \delta \vec{p} \rangle_{H_0(\text{div})^*, H_0(\text{div})} \geq 0.
\]

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

\[
|B^{-1} \text{div} \delta \vec{p}|^2 + \gamma |P_{\text{div}} \delta \vec{p}|^2 \leq 0
\]

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 \( \vec{\lambda}_i, i = 1, 2 \), of \( \vec{\lambda}_c \). We have for the pointwise a.e. evaluation at \( x \in \Omega \)

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

It follows that

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

and consequently

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

From (3.2) and (3.5) we deduce that

\[
\frac{1}{c} |\nabla \vec{p}_c|^2 + |\text{div} \vec{p}_c|^2 + \gamma |P_{\text{div}} \vec{p}_c|^2 \leq |\text{div} \vec{p}_c| B |K^* f| B
\]

and hence

\[
\frac{1}{c} |\nabla \vec{p}_c|^2 + \frac{1}{2} |\text{div} \vec{p}_c|^2 + \gamma |P_{\text{div}} \vec{p}_c|^2 \leq \frac{1}{2} |K^* f| B. \tag{3.6}
\]

We further estimate

\[
|\vec{\lambda}_c|_{H^1_0(\Omega)^*} \leq \sup_{|\vec{v}|_{H^1_0(\Omega)} = 1} \langle \vec{\lambda}_c, \vec{v} \rangle_{H^1_0(\Omega)^*, H^1_0(\Omega)} \leq \sup_{|\vec{v}|_{H^1_0(\Omega)} = 1} \left\{ \frac{1}{c} |\nabla \vec{p}_c| |\nabla \vec{v}| + |\text{div} \vec{p}_c| B |\text{div} \vec{v}| B + |K^* f| B |\text{div} \vec{v}| B + \gamma |P_{\text{div}} \vec{p}_c| |P_{\text{div}} \vec{v}| \right\}.
\]

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

\[
|\vec{\lambda}_c|_{H^1_0(\Omega)^*} \leq K. \tag{3.7}
\]
Combining (3.6) and (3.7) we can assert the existence of \((\bar{p}^*, \bar{\lambda}^*) \in H_0(\text{div}) \times \mathbb{H}^1_0(\Omega)^*\) such that for a subsequence denoted by the same symbol

\[
(\bar{p}_c^*, \bar{\lambda}_c) \rightharpoonup (\bar{p}^*, \bar{\lambda}^*) \text{ weakly in } H_0(\text{div}) \times \mathbb{H}^1_0(\Omega)^*.
\]

(3.8)

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

\[
\frac{1}{c}(\nabla \bar{p}_c, \nabla \bar{v}) + (\text{div} \bar{p}_c, \text{div} \bar{v})_B + (K^* f, \text{div} \bar{v})_B + \gamma(P_{\text{div}} \bar{p}_c, P_{\text{div}} \bar{v}) +
\langle \bar{\lambda}_c, \bar{v} \rangle = 0 \text{ for all } \bar{v} \in \mathbb{H}^1_0(\Omega).
\]

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

\[
(\text{div} \bar{p}^*, \text{div} \bar{v})_B + (K^* f, \text{div} \bar{v})_B + \gamma(P_{\text{div}} \bar{p}^*, P_{\text{div}} \bar{v}) +
\langle \bar{\lambda}^*, \bar{v} \rangle_{\mathbb{H}^1_0(\Omega)^*, \mathbb{H}^1_0(\Omega)} = 0 \text{ for all } \bar{v} \in \mathbb{H}^1_0(\Omega).
\]

(3.9)

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

\[
(\bar{\lambda}_c, \bar{p} - \bar{p}_c) = (\max(0, c(\bar{p}_c - \beta \bar{\lambda}^*))) + \min(0, c(\bar{p}_c + \beta \bar{\lambda}^*)) \leq 0
\]

(3.10)

for all \(-\beta \bar{\lambda}^* \leq \bar{p} \leq \beta \bar{\lambda}^*\). From (3.1) we have

\[
\frac{1}{c} |\nabla \bar{p}_c|^2 + |\text{div} \bar{p}_c + K^* f|_B^2 + \gamma |P_{\text{div}} \bar{p}_c|^2 + \frac{1}{c} |\bar{\lambda}_c|^2 \leq |K^* f|_B^2.
\]

(3.11)

Consequently, \(\frac{1}{c} |\bar{\lambda}_c|^2 \leq |K^* f|_B^2\) for all \(c > 0\). Note that

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

and thus

\[
|\max(0, (\bar{p}_c - \beta \bar{\lambda}^*))|^2_{\mathbb{L}^2(\Omega)} \xrightarrow[c \to \infty]{} 0,
\]

\[
|\min(0, (\bar{p}_c + \beta \bar{\lambda}^*))|^2_{\mathbb{L}^2(\Omega)} \xrightarrow[c \to \infty]{} 0.
\]

(3.12)

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

\[
\int_\Omega |\max(0, \bar{p}^* - \beta \bar{\lambda}^*)|^2 dx \leq \liminf_{c \to \infty} \int_\Omega |\max(0, \bar{p}_c - \beta \bar{\lambda}^*)|^2 dx = 0.
\]

Consequently, \(\bar{p}^* \leq \beta \bar{\lambda}^*\) and analogously one verifies that \(-\beta \bar{\lambda}^* \leq \bar{p}^*\). In particular \(\bar{p}^*\) is feasible and from (3.10) we conclude that

\[
\langle \bar{\lambda}_c, \bar{p}^* - \bar{p}_c \rangle_{H_0(\text{div})^*, H_0(\text{div})} \leq 0 \text{ for all } c > 0.
\]

(3.13)

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

\[
\limsup_{c \to \infty} \left( \frac{1}{2} |\text{div} \bar{p}_c + K^* f|_B^2 + \frac{\gamma}{2} |P_{\text{div}} \bar{p}_c|^2 \right) \leq \frac{1}{2} |\text{div} \bar{p} + K^* f|_B^2 + \frac{\gamma}{2} |P_{\text{div}} \bar{p}|^2
\]

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

$$\limsup_{c \to \infty} \left( \frac{1}{2} \text{div} \vec{p}_c + K^* |f|^2_B + \frac{\gamma}{2} |\text{div} \vec{p}_c|^2 \right) \leq \frac{1}{2} \text{div} \vec{p}^* + K^* |f|^2_B + \frac{\gamma}{2} |\text{div} \vec{p}^*|^2$$

$$\leq \liminf_{c \to \infty} \left( \frac{1}{2} \text{div} \vec{p}_c + K^* |f|^2_B + \frac{\gamma}{2} |\text{div} \vec{p}_c|^2 \right),$$

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

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

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

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

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

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

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

$$\langle \vec{\lambda}^*, \vec{p} - \vec{p}^* \rangle_{H_0(\text{div})^*, H_0(\text{div})} = \langle \vec{\mu}^*, \vec{p} - \vec{p}^* \rangle_{H_0(\text{div})^*, H_0(\text{div})}$$

$$= \lim_{c \to \infty} \langle -\frac{1}{c} \Delta \vec{p}_c + \vec{\lambda}_c, \vec{p} - \vec{p}_c \rangle_{H_0(\text{div})^*, H_0(\text{div})}$$

$$= \lim_{c \to \infty} \left( \frac{1}{c} \langle \nabla \vec{p}_c, \nabla (\vec{p} - \vec{p}_c) \rangle + \langle \vec{\lambda}_c, \vec{p} - \vec{p}_c \rangle \right)$$

$$\leq \lim_{c \to \infty} \left( \frac{1}{c} \langle \nabla \vec{p}_c, \nabla \vec{p} \rangle + \langle \vec{\lambda}_c, \vec{p} - \vec{p}_c \rangle \right) \leq 0$$

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

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

\[
\begin{align*}
\min & \quad \frac{1}{2} |A_1 p + \tilde{f}|^2 + \frac{\gamma}{2} |A_2 p|^2 \\
\text{s.t.} & \quad -\beta \mathbf{1} \leq p \leq \beta \mathbf{1},
\end{align*}
\]

(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

\[
A_1^T A_1 p + \gamma A_2^T A_2 p + A_1^T \tilde{f} + \lambda = 0,
\]

\[
\lambda = \max(0, \lambda + c(p - \beta \mathbf{1})) + \min(0, \lambda + c(p + \beta \mathbf{1})),
\]

(4.2)

where \( c > 0 \) is arbitrary and fixed. The primal-dual active set strategy or equivalently the semi-smooth 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

\[
A_{k+1}^+ = \{ i : (\lambda_k + c(p_k - \beta \mathbf{1}))_i > 0 \},
\]

\[
A_{k+1}^- = \{ i : (\lambda_k + c(p_k + \beta \mathbf{1}))_i < 0 \},
\]

\[
I_{k+1} = \{ i : i \notin A_{k+1}^+ \cap A_{k+1}^- \}.
\]

(3) Solve for \( p_{k+1}, \lambda_{k+1} \)

\[
A_1^T A_1 p_{k+1} + \gamma A_2^T A_2 p_{k+1} + A_1^T \tilde{f} + \lambda_{k+1} = 0,
\]

\[
(\lambda_{k+1})_i = 0 \quad \text{for} \quad i \in I_{k+1},
\]

\[
(p_{k+1})_i = \beta \quad \text{for} \quad i \in A_{k+1}^+ \setminus I_{k+1},
\]

\[
(p_{k+1})_i = -\beta \quad \text{for} \quad i \in A_{k+1}^- \setminus I_{k+1}.
\]

(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 \( \lambda_0 \). Then we have:

**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 \( (p^*, \lambda^*) \) of (4.2).

The result can be verified by standard techniques from semi-smooth 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 \( \bar{p}_0 \in H_0^1(\Omega) \) and set \( k = 0 \).
(2) Set, for $i = 1, 2$,

$$A_{k+1}^{+,i} = \{ x : (\bar{p}_{k}^{i} - \beta \bar{1})(x) > 0 \},$$

$$A_{k+1}^{-,i} = \{ x : (\bar{p}_{k}^{i} + \beta \bar{1})(x) < 0 \},$$

$$I_{k+1}^{i} = \Omega \setminus (A_{k+1}^{+,i} \cup A_{k+1}^{-,i}).$$

(3) Solve for $\bar{p} \in H_{0}^{1}(\Omega)$ and set $\bar{p}_{k+1} = \bar{p}$ where

$$\frac{1}{c}((\nabla \bar{p}, \nabla \bar{v}) + (\text{div} \bar{p}, \text{div} \bar{v})_{\Omega} + (K^{*}f, \text{div} \bar{v})_{\Omega} + \gamma(\text{div} \bar{p}, \text{div} \bar{v}) + (c(\bar{p} - \beta \bar{1})\chi_{A_{k+1}^{+,i}}, \bar{v}) + (c(\bar{p} + \beta \bar{1})\chi_{A_{k+1}^{-,i}}, \bar{v}) = 0$$

for all $\bar{v} \in H_{0}^{1}(\Omega)$.

(4) Set

$$\bar{x}_{k+1}^{i} = \begin{cases} 0 & \text{on } I_{k+1}^{i}, \\ c(\bar{p}_{k+1}^{i} - \beta \bar{1}) & \text{on } A_{k+1}^{+,i}, \\ c(\bar{p}_{k+1}^{i} + \beta \bar{1}) & \text{on } A_{k+1}^{-,i}. \end{cases}$$

for $i = 1, 2$.

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

Above $\chi_{A_{k+1}^{+,i}}$ stands for

$$\chi_{A_{k+1}^{+,i}} = \begin{cases} 1 & \text{if } x \in A_{k+1}^{+,i}, \\ 0 & \text{if } x \not\in A_{k+1}^{+,i}, \end{cases}$$

and analogously for $A_{k+1}^{-,i}$. The superscript $i, i = 1, 2$, refers to the respective component. We note that (4.3) admits a solution $\bar{p}_{k+1} \in H_{0}^{1}(\Omega)$. Step (4) is included for the sake of the analysis of the algorithm. Let $C : H_{0}^{1}(\Omega) \to H^{-1}(\Omega) \times H^{-1}(\Omega)$ stand for the operator

$$C = -\frac{1}{c} \Delta - \nabla B^{-1} \text{div} + \gamma \text{div}.$$

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

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

where we drop the index in the notation for $\bar{p}_{c}$. For $\varphi \in L^{2}(\Omega)$ we define

$$D_{\max}(0, \varphi)(x) = \begin{cases} 1 & \text{if } \varphi(x) > 0, \\ 0 & \text{if } \varphi(x) \leq 0, \end{cases}$$

and

$$D_{\min}(0, \varphi)(x) = \begin{cases} 1 & \text{if } \varphi(x) < 0, \\ 0 & \text{if } \varphi(x) \geq 0. \end{cases}$$

These operators define semi-smooth 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 semi-smooth Newton step can be expressed as
\[
C\tilde{p}_{k+1} + c(\tilde{p}_{k+1} - \beta\tilde{\lambda})\chi_{A_{k+1}^+} + c(\tilde{p}_{k+1} + \beta\tilde{\lambda})\chi_{A_{k+1}^-} - \nabla B^{-1}K^*f = 0, 
\]
and \(\tilde{\lambda}_{k+1}\) from step (4) of Algorithm B is given by
\[
\tilde{\lambda}_{k+1} = c(\tilde{p}_{k+1} - \beta\tilde{\lambda})\chi_{A_{k+1}^+} + c(\tilde{p}_{k+1} + \beta\tilde{\lambda})\chi_{A_{k+1}^-}. 
\]
The iteration of Algorithm B can also be expressed with respect to the variable \(\tilde{\lambda}\) rather than \(\tilde{p}\). For this purpose we define
\[
F(\tilde{\lambda}) = \tilde{\lambda} - c \max(0, C^{-1}(\nabla \tilde{f} - \tilde{\lambda}) - \beta\tilde{\lambda}) - c \min(0, C^{-1}(\nabla \tilde{f} - \tilde{\lambda}) + \beta\tilde{\lambda}), 
\]
where we put \(\tilde{f} = B^{-1}K^*f\). Setting \(\tilde{p}_k = C^{-1}(\nabla \tilde{f} - \tilde{\lambda}_k)\), the semi-smooth Newton step applied to \(F(\tilde{\lambda}) = 0\) at \(\tilde{\lambda} = \tilde{\lambda}_k\) results in
\[
\tilde{\lambda}_{k+1} = c(C^{-1}(\nabla \tilde{f} - \tilde{\lambda}_{k+1}) - \beta\tilde{\lambda})\chi_{A_{k+1}^+} + c(C^{-1}(\nabla \tilde{f} - \tilde{\lambda}_{k+1}) + \beta\tilde{\lambda})\chi_{A_{k+1}^-} 
\]
which coincides with (4.8). Therefore the semi-smooth Newton iterations according to Algorithm B and that for \(F(\tilde{\lambda}) = 0\) coincide, provided that the initializations are related by \(C\tilde{p}_0 - \nabla \tilde{f} + \tilde{\lambda}_0 = 0\). The mapping \(F\) is slantly differentiable, i.e., for every \(\tilde{\lambda} \in \mathbb{L}^2(\Omega)\)
\[
|F(\tilde{\lambda} + \tilde{h}) - F(\tilde{\lambda}) - DF(\tilde{\lambda} + \tilde{h})\tilde{h}|_{\mathbb{L}^2(\Omega)} = o(|\tilde{h}|_{\mathbb{L}^2(\Omega)}) \quad (4.10)
\]
for \(|\tilde{h}|_{\mathbb{L}^2(\Omega)} \to 0\), 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 \(|\tilde{\lambda}_e - \tilde{\lambda}_0|_{\mathbb{L}^2(\Omega)}\) is sufficiently small, then the iterates \(\{(\tilde{p}_k, \tilde{\lambda}_k)\}_{k=1}^\infty\) of Algorithm B converge superlinearly in \(\mathbb{H}^1(\Omega) \times \mathbb{L}^2(\Omega)\) to the solution \((\tilde{p}_e, \tilde{\lambda}_e)\) 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 \(K = I\) 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 \(K \neq I\) with \(\ker(K)\) typically nontrivial which requires 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 \(P_{\text{div}}\tilde{p}\) 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

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

where $\phi_c$ is the generalized Moreau-Yosida regularization of the indicator function $\phi$ of the set $\{\vec{p} \in L^2(\Omega) : -\beta \vec{1} \leq \vec{p} \leq \beta \vec{1}\}$. We have

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

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

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

**Augmented Lagrangian Method (ALM).**

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

$$\vec{p}_n = \arg\min \{L(\vec{p}, \vec{\lambda}_n) : \vec{p} \in L^2(\Omega)\},$$

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

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

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) Laplacian and, for Algorithm B, the terms involving the indicator functions of $A_{k+1}$.

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 5.1. The upper left image is the original image which is similar to the one in [8]. It has a dynamic range $[0, 255]$. The other two images contain Gaussian white noise. The upper right image 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 the respective algorithm as soon as the discrete $L_2$-norm of the residual dropped below $\text{tol} = \sqrt{\epsilon_M}$, with $\epsilon_M$ the machine precision, or the difference between two successive residuals was smaller than $\text{tol}$, 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 constraints. Further, for all three algorithms we chose $\beta = 0.2$, $\gamma = 0$ for ALM and Algorithm B, and $\gamma = 1E-3$ for Algorithm A. In general, for ALM and Algorithm B $\gamma$ had no noticeable effect on the results attained. However, Algorithm A is more sensitive with respect to $\gamma$. 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 algorithm. We note that Algorithm B requires the least numbers 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 5.2 we display the reconstructions. The upper left and right corresponds
to Algorithm A and B. The lower image is the result obtained by Algorithm ALM. The quality of the reconstructions is equally good for all algorithms.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|}
\hline
Algorithm & #as & #cg & #alm \\
\hline
Algorithm A & 16 & 64 & - \\
Algorithm B & 9 & 23 & - \\
ALM & 14 & 27 & 3 \\
\hline
\end{tabular}
\caption{Results for 10% noise.}
\end{table}

In the introduction we mention that the Lagrange multiplier associated with the box constraints serves as an edge detector. Figure 5.3 shows the $\ell_1$-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 $\lambda_e$ as

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

Above $x_i$ denotes the $i$th pixel of the image and $\tilde{\lambda}^*$ the multiplier upon termination.
of B. For the multipliers resulting from Algorithms A and ALM a similar observation holds true.

Fig. 5.3. (Left) Lagrange multiplier of Algorithm B. (Right) Corresponding edge detector.

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 5.4 shows the reconstructions obtained from our algorithms. Like in the previous test case, the quality of the results for Algorithm B and ALM is comparable. Algorithm A appears to be slightly more sensitive with respect to noise. This behavior could not be ruled out by tuning the parameters $c$, $\gamma$ and $\beta$. 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 5.5 we display the $\ell_1$-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 5.5. We conclude that–without any thresholding–the Lagrange multiplier may act as an edge detector.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>#as</th>
<th>#cg</th>
<th>#alm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algorithm A</td>
<td>15</td>
<td>59</td>
<td>-</td>
</tr>
<tr>
<td>Algorithm B</td>
<td>7</td>
<td>18</td>
<td>-</td>
</tr>
<tr>
<td>ALM</td>
<td>13</td>
<td>33</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 5.2
Results for 50% noise.

Let us briefly comment on the difference of stability with respect to $\beta$ of Algorithms A and B compared to ALM. In general the choice of $\beta$ influences the quality of the reconstruction. A large value for $\beta$ 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 5.6 corresponds to the result attained by Algorithm B with $\beta = 1.5$ (compared to $\beta = 1.25$ in the previous run). Due to the larger $\beta$-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 5.6 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 in our test runs, i.e., ALM is more stable with respect to the
choice of $\beta$. The behavior of Algorithm A with respect to changes in $\beta$ is comparable to the one 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 numbers 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 $\beta$ the iterates converge superlinearly. Small values of $\beta$, 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 to 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 5.7 (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 9 iterations for obtaining the denoised image in the upper right plot of Figure 5.7. The algorithm in [8] with a line search and a continuation strategy w.r.t. $\delta$ 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 $\ell_1$-norm of $\lambda$ upon termination of Algorithm B, is given in the last subplot of Figure 5.7.

**Example 2.** Now we report on the behavior of Algorithm B for the benchmark problem in Figure 5.8. 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 9 AS-iterations (31 CG-iterations totally) with a residual of $6.2E-9$. The corresponding reconstruction is given in the lower left plot of Figure 5.8. The lower right plot displays the $\ell_1$-norm of the Lagrange multiplier associated with the box constraints. Like 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 $K \neq I$. The data $f$ correspond to a coarse version of the original image satisfying $f_{2i-1,2j-1} =$
Fig. 5.7. (Upper left) Noisy image (256×256 pixel). (Upper right) Result of Algorithm B. (Lower) $\ell_1$-norm of $\lambda$.

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 $\tilde{v}_{1,j} = u_{2i-1,2j-1}$ and $v_{2i-1,2j-1} = v_{2i,2j-1} = v_{2i-1,2j} = v_{2i,2j} = \tilde{v}_{1,j}$. 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 non-trivial kernel, we choose $\alpha = 1E-10$. Further we pick the parameter values $c = 1E5$, $\beta = 0.35$, and $\gamma = 0$. In Figure 5.9 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 2, and 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 non-reflexive Banach space BV and its pre-dual which is posed in the Hilbert space $H_0(\text{div})$. 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 pre-dual problem of the underlying BV-regularized minimization problem. By pre-dualization we obtain a box constrained minimization problem which—from the numer-
Fig. 5.8. (Upper left) Exact data (256×256 pixel). (Upper right) Noisy data. (Lower left) Reconstruction obtained by Algorithm B. (Lower right) $\ell_1$-norm of the Lagrange multiplier.

From a 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.

REFERENCES


Fig. 5.9. (Upper left) Exact data (256×256 pixel). (Upper right) Reconstruction obtained by Algorithm B. (Lower left) 128×128-pixel image expanded by a factor 2. (Lower right) Result using a nearest neighbor interpolation technique.


