SIAM J. SCI. COMPUT.
Vol. 28, No. 1, pp. 1–23

AN INFEASIBLE PRIMAL-DUAL ALGORITHM FOR TOTAL BOUNDED VARIATION–BASED INF-CONVOLUTION-TYPE IMAGE RESTORATION*

M. HINTERMÜLLER† AND G. STADLER‡

Abstract. In this paper, a primal-dual algorithm for total bounded variation (TV)–type image restoration is analyzed and tested. Analytically it turns out that employing a global $\boldsymbol{L}^s$-regularization, with $1 < s \leq 2$, in the dual problem results in a local smoothing of the TV-regularization term in the primal problem. The local smoothing can alternatively be obtained as the infimal convolution of the ℓ_r-norm, with $r^{-1} + s^{-1} = 1$, and a smooth function. In the case $r = s = 2$, this results in Gauss-TV–type image restoration. The globalized primal-dual algorithm introduced in this paper works with generalized derivatives, converges locally at a superlinear rate, and is stable with respect to noise in the data. In addition, it utilizes a projection technique which reduces the size of the linear system that has to be solved per iteration. A comprehensive numerical study ends the paper.

Key words. Fenchel duality, generalized Newton-type methods, image restoration, total bounded variation

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

DOI. 10.1137/040613263

1. Introduction.

Image denoising based on total bounded variation regularization (TV regularization) was introduced in [26]. Since then it has received a considerable amount of attention and is widely accepted as a reliable tool in edge preserving image restoration; see the selected papers [1, 4, 5, 6, 9, 10, 14, 23, 29] and the recent monograph [32] for further references.

Compared to classical H^1-type regularization, such as $\int_\Omega |\nabla u|^2_{\ell^2}\, dx$ with u representing the reconstructed image, or other techniques based on smooth (at least C^1) functionals, the TV term $\int_\Omega |Du|$ is nondifferentiable in the Fréchet sense. This nondifferentiability is responsible for preserving edges in images, but at the same time, it poses significant numerical challenges. Early computational techniques employed global smoothing of the bounded variation (BV) seminorm to overcome the nondifferentiability [9, 33]; see also the more recent work [2]. In many cases a C^∞-function like $\int_\Omega (|\nabla u|^2_{\ell^2} + \epsilon)^{1/2}\, dx$, with $\epsilon > 0$ fixed, is used as a smooth approximation of the nondifferentiable TV term. For small smoothing parameter ϵ, this technique introduces severe ill-conditioning of second derivatives and adversely affects, e.g., Newton's method for solving the first order conditions associated with the smoothed TV-type problem. To remedy this ill-conditioning, in [6] a (dual) flux variable w is introduced together with the requirement that the iterates w^k stay in the interior of the unit ball (strict feasibility) in the pointwise almost everywhere sense. Another type of global regularization is introduced in the primal affine scaling technique in [21], where

*Received by the editors August 11, 2004; accepted for publication (in revised form) February 22, 2005; published electronically March 3, 2006. This work was supported in part by the "Fonds zur Förderung der wissenschaftlichen Forschung" under "SRC 03, Optimization and Control."
http://www.siam.org/journals/sisc/28-1/61326.html

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

‡Institut für Mathematik und Wissenschaftliches Rechnen, Karl-Franzens-Universität Graz, Heinrichstraße 36, A-8010 Graz, Austria (ge.stadler@uni-graz.at).

the interior point treatment acts like a global regularization within a continuation framework.

As the above-mentioned smoothing of the BV seminorm introduces *artificial* global nonlinearity to the problem, several authors studied local smoothing [5, 10, 28]; see [31] for further references. The idea is to keep the original problem formulation to the largest possible extent. While the above remarks on ill-conditioning persist, in the local case the BV seminorm is changed only close to kinks. The utilized functionals are typically only once continuously differentiable as opposed to the C^∞-type global smoothing.

Yet another approach that avoids global smoothing is based on active set techniques; see [15, 17, 18] for closely related algorithms. Utilizing augmented Lagrangian formulations, these methods predict the set of points where the BV seminorm is nondifferentiable, fix the gradient of the primal variable on this set, and compute it on the complement set. It turns out that the original problem is approximated by a sequence of problems which are almost as difficult to solve as the original one. In [15] this drawback is overcome by using a penalization technique, which, to some extent, acts like an additional regularization. In [18] a heuristic lumping technique is applied for solving the auxiliary problems. In numerical practice these active set-type methods are significantly faster than the widely used gradient descent schemes [26] or fixed point–type iterations [9, 33].

In this paper, we adopt the Hilbert space perspective of [15, 17, 18]; i.e., we consider the problem

$$(\mathcal{P}) \qquad \min_{u\in H_0^1(\Omega)} \frac{\mu}{2}\int_\Omega |\nabla u|_2^2\,dx + \frac{1}{2}\int_\Omega |Ku-z|^2\,dx + \alpha\int_\Omega |\nabla u|_r\,dx,$$

where $\Omega\subset\mathbb{R}^2$ is an open, bounded (image) domain with Lipschitz-continuous boundary $\partial\Omega$, $K : L^2(\Omega)\to L^2(\Omega)$ is a continuous linear operator with K^* its adjoint, $z\in L^2(\Omega)$ represents given noisy data, $\mu,\alpha>0$, and $|\nabla u|_r = (|u_x|^r+|u_y|^r)^{1/r}$, with $1\le r<+\infty$. Typically $0<\mu\ll\alpha$ such that $(\mathcal{P})$ is a close approximation to the TV-regularized problem

$$(\mathcal{P}_{BV}) \qquad \min_{u\in \mathrm{BV}(\Omega)} \frac{1}{2}\int_\Omega |Ku-z|^2\,dx + \alpha\int_\Omega |Du|_r,$$

where $\mathrm{BV}(\Omega)$ denotes the space of functions of bounded variation; i.e., a function u is in $\mathrm{BV}(\Omega)$ if the BV seminorm

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

with $r^{-1}+s^{-1}=1$, is finite. We recall that $\mathrm{BV}(\Omega)\subset L^2(\Omega)$ for $\Omega\subset\mathbb{R}^2$ [12]. For the discretized version of $(\mathcal{P})$ and as long as K^*K is invertible, we may even choose $\mu=0$ and obtain stable results; see section 5. A particular instance with K^*K invertible is given by image denoising where $K=\mathrm{id}$.

The aim of the present paper is to devise an implementable algorithm which is based on a primal-dual approach, overcomes the difficulties of the methods in [15, 17, 18], is locally rapidly convergent, and, in contrast to, e.g., [6], takes advantage of local regularization only. The primal-dual framework is obtained by considering the Fenchel dual to $(\mathcal{P})$, which turns out to be a bound-constrained concave quadratic maximization problem. The objective function of the dual problem involves the divergence operator, which is responsible for the solution set not to be a singleton.

To avoid this nonuniqueness property, we propose a simple $\boldsymbol{L}^s$-regularization, with $1 < s \leq 2$ and $\boldsymbol{L}^s(\Omega) = (L^s(\Omega))^2$, such that the resulting regularized dual problem admits a unique solution. Let $\gamma > 0$ denote the corresponding regularization parameter. Then, in the context of the primal problem, for $w \in \mathbb{R}^n$ this regularization is given by the inf-convolution

$$\phi_{r,\gamma}(w) = \inf\{|\hat{w}|_r + (r\gamma)^{-1}|w - \hat{w}|_r^r : \hat{w} \in \mathbb{R}^n\}.$$

As a consequence, the regularized version of $(\mathcal{P})$ involves the term $\int_\Omega \Phi_{r,\gamma}(\nabla u)\,dx$, with $\Phi_{r,\gamma}(\nabla u)(x) = \phi_{r,\gamma}(\nabla u(x))$, rather than $\int_\Omega |\nabla u|_r\,dx$. In the special case $s = 2$ we obtain Gauss-TV regularization [28], which was recently shown to be an excellent choice for medical imaging [20]. The latter statement is mainly based on the fact that the Gauss-TV regularization employs a local smoothing of $|\cdot|_2$ and allows avoidance of the adverse staircase or bulkiness effect well known in pure TV regularization; see, e.g., [3]. Similar effects are obtained for $1 < s < 2$. In our model, the choice of γ influences the amount of local smoothing. Indeed, for small γ one finds that $\Phi_{r,\gamma}$ is close to the BV seminorm, whereas for large γ, strong smoothing of $|\cdot|_r$ takes place. For further techniques to reduce the staircase effect we refer the reader to [19].

The rest of the paper is organized as follows. In section 2 we derive the Fenchel dual problem of $(\mathcal{P})$, introduce the dual regularization, and discuss some of its aspects. The primal-dual algorithm is defined in section 3, where we also provide a discussion of global as well as fast local convergence. In section 4 we consider aspects of the implementation, and in section 5 we report on our numerical experience with the new algorithm, validate the theoretical results, and we study the qualitative behavior of our dual regularization technique.

Throughout the paper we shall frequently use the following notation. For a vector $v \in \mathbb{R}^n$, and $r \in \mathbb{R}$ we define $|v|^r := (|v_1|^r, \ldots, |v_n|^r)^\top$ and $\sigma(v) \in \mathbb{R}^n$ with $\sigma(v) = (\operatorname{sign}(v_1), \ldots, \operatorname{sign}(v_n))^\top$. We use $|\cdot|_s$ for the ℓ_s-vector norm in $\mathbb{R}^n$. Further we denote by $\star$ the Hadamard product of vectors, i.e., $v \star w = (v_1 w_1, \ldots, v_n w_n)^\top$. For vector-valued functions we write $\vec{p}$ and $p(x) \in \mathbb{R}^n$ for its value at x. For a symmetric matrix $M \in \mathbb{R}^{m,m}$, $\lambda_{\min}(M)$ denotes its smallest eigenvalue. In the case when M is nonsymmetric, $\lambda_{\min}(M)$ is the smallest eigenvalue of $\frac{1}{2}(M + M^\top)$, the symmetrization of M.

2. The Fenchel dual of $(\mathcal{P})$ and its regularization. For the reader's convenience we now recall the Fenchel duality theorem in infinite-dimensional spaces; see, e.g., [11]. For this purpose let V and Y be Banach spaces with topological duals V^* and Y^*, respectively. Further, let $\Lambda \in \mathcal{L}(V, Y)$ be the space of continuous linear operators from V to Y, and let $\mathcal{F} : V \to \mathbb{R} \cup \{\infty\}$, $\mathcal{G} : Y \to \mathbb{R} \cup \{\infty\}$ be convex, proper, and lower semicontinuous such that there exists $v_0 \in V$ with $\mathcal{F}(v_0) < \infty$, $\mathcal{G}(\Lambda v_0) < \infty$, and $\mathcal{G}$ continuous at Λv_0. Then

$$\inf_{v \in V} \{\mathcal{F}(v) + \mathcal{G}(\Lambda v)\} = \sup_{q \in Y^*} \{-\mathcal{F}^*(\Lambda^* q) - \mathcal{G}^*(-q)\}, \tag{2.1}$$

where $\Lambda^* \in \mathcal{L}(Y^*, V^*)$ is the adjoint of Λ. The convex conjugates $\mathcal{F}^* : V^* \to \mathbb{R} \cup \{\infty\}$, $\mathcal{G}^* : Y^* \to \mathbb{R} \cup \{\infty\}$ of $\mathcal{F}$ and $\mathcal{G}$, respectively, are defined by

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

and analogously for $\mathcal{G}^*$. The conditions imposed on $\mathcal{F}$ and $\mathcal{G}$ guarantee that the dual problem (which is the problem on the right-hand side of (2.1)) admits a solution.

Furthermore, the solutions $\bar v \in V$ and $\bar q \in Y^*$ are characterized by the optimality conditions

$$\begin{aligned} \Lambda^* \bar q &\in \partial \mathcal{F}(\bar v), \\ -\bar q &\in \partial \mathcal{G}(\Lambda \bar v), \end{aligned} \tag{2.2}$$

where ∂ denotes the subdifferential from convex analysis.

2.1. Fenchel dual of ($\mathcal{P}$). Now we apply the Fenchel calculus to

$$\begin{aligned} \mathcal{F}(u) &:= \frac{1}{2}\int_\Omega |Ku - z|^2\, dx + \frac{\mu}{2}\int_\Omega |\nabla u|_2^2 dx, \\ \mathcal{G}(\vec q) &:= \alpha \int_\Omega |\vec q|_r\, dx, \;\; \Lambda u := (u_x, u_y)^\top, \end{aligned}$$

with $V := H_0^1(\Omega)$, $Y := \boldsymbol{L}^2(\Omega)$, $1 \le r < +\infty$. As a result we obtain

$$\begin{aligned} \mathcal{F}^*(u^*) &:= \frac{1}{2}|||u^* + K^* z|||_{H^{-1}}^2 - \frac{1}{2}\|z\|_{L^2}^2, \\ \mathcal{G}^*(\vec q^*) &:= I_{\{\vec w \in \boldsymbol{L}^2(\Omega) : |\vec w|_s \le \alpha \text{ a.e. in } \Omega\}}(\vec q^*), \end{aligned}$$

where $r^{-1} + s^{-1} = 1$, I_S is the indicator function of the set $S \subset \boldsymbol{L}^2(\Omega)$, and $|||v|||_{H^{-1}} = \langle H_{\mu,K} v, v\rangle_{H_0^1, H^{-1}}$, $v \in H^{-1}(\Omega)$, with $H_{\mu,K} = (K^*K - \mu\Delta)^{-1}$. Here $\langle \cdot, \cdot \rangle_{H_0^1, H^{-1}}$ denotes the duality pairing between $H_0^1(\Omega)$ and its dual $H^{-1}(\Omega)$, and $\Delta : H_0^1(\Omega) \to H^{-1}(\Omega)$.

Setting $\vec p = -\vec q$, then, according to (2.1), the dual problem of ($\mathcal{P}$) is

$$(\mathcal{P}^\star) \qquad \sup_{\substack{\vec p \in \boldsymbol{L}^2(\Omega), \\ |\vec p|_s \le \alpha \text{ a.e. in } \Omega}} \; -\frac{1}{2}|||\operatorname{div} \vec p + K^* z|||_{H^{-1}}^2 + \frac{1}{2}\|z\|_{L^2}^2.$$

Note that the Fenchel dual of ($\mathcal{P}$) is a bound constrained maximization problem with a concave quadratic objective function. Further observe that the solution to ($\mathcal{P}^\star$) is not unique since the divergence operator has a nontrivial kernel.

From (2.2) we obtain the following characterization of solutions $\bar u$ and $\bar{\vec p}$ of ($\mathcal{P}$) and ($\mathcal{P}^\star$), respectively:

$$-\mu\Delta \bar u + K^*K\bar u - \operatorname{div} \bar{\vec p} = K^* z \;\text{ in } H^{-1}(\Omega), \tag{2.3a}$$

$$\left.\begin{aligned} -\alpha(\sigma(\nabla \bar u) \star |\nabla \bar u|^{r-1}) + |\nabla \bar u|_r^{r-1}\bar{\vec p} \;&= 0 \quad \text{if} \quad |\bar{\vec p}|_s = \alpha, \\ \nabla \bar u \;&= 0 \quad \text{if} \quad |\bar{\vec p}|_s < \alpha \end{aligned}\right\} \;\text{ in } \boldsymbol{L}^2(\Omega). \tag{2.3b}$$

In (2.3a)–(2.3b) the expression $|\nabla \bar u|_r^{r-1}$ can be interpreted as a Lagrange multiplier for the inequality constraint in the dual problem ($\mathcal{P}^\star$). We summarize the above results in the following theorem.

THEOREM 2.1. *The problems* ($\mathcal{P}$) *and* ($\mathcal{P}^\star$) *are mutually dual, i.e.,* (2.1) *holds and the solutions* $\bar u$ *of* ($\mathcal{P}$) *and* $\bar{\vec p}$ *of* ($\mathcal{P}^\star$) *are characterized by* (2.3a)–(2.3b). *The solution of* ($\mathcal{P}$) *is unique, while for two solutions* $\bar{\vec p}^{\,1}$, $\bar{\vec p}^{\,2}$ *of* ($\mathcal{P}^\star$) *there holds* $\operatorname{div}(\bar{\vec p}^{\,1} - \bar{\vec p}^{\,2}) = 0$ *and* $\bar{\vec p}^{\,1} = \bar{\vec p}^{\,2}$ *on sets where* $\nabla \bar u \neq 0$.

With respect to the special case $\mu = 0$ we point out that

- in [15, Remark 2.1], for $K = \mathrm{id}$, it is shown that the solution $\bar u = \bar u(\mu)$ satisfies $\bar u(\mu) \rightharpoonup \bar u_{\mathrm{BV}}$ weakly in $L^2(\Omega)$ as $\mu \to 0$, where $\bar u_{BV}$ denotes the unique solution to ($\mathcal{P}_{BV}$);

- for $\mu = 0$, in [14] the Fenchel predual problem of $(\mathcal{P})$ in an ℓ_1-vector-norm setting is derived. Surprisingly it turns out that the predual of the TV problem $(\mathcal{P}_{BV})$ is posed in a Hilbert space. Its structure is similar to $(\mathcal{P}^\star)$. The underlying function space is $H_0(\mathrm{div}) = \{\vec{v} \in \boldsymbol{L}^2(\Omega) : \mathrm{div}\,\vec{v} \in L^2(\Omega),\ \vec{v}\cdot\vec{n} = 0 \text{ on } \partial\Omega\}$, where $\vec{n}$ denotes the outward unit normal to $\partial\Omega$. For a convergence analysis of the proposed semismooth Newton method, however, an H^1-type regularization of the dual problem is necessary; for details see [14, section 3].

2.2. Regularization. Nonuniqueness of the solution to the dual problem may cause instabilities in a numerical scheme utilizing primal and dual variables. As a remedy we propose the following simple Tikhonov-type regularization of $(\mathcal{P}^\star)$:

$$(\mathcal{P}^\star_\gamma) \qquad \sup_{\substack{\vec{p}\in\boldsymbol{L}^2(\Omega),\\ |\vec{p}|_s \le \alpha \text{ a.e. in } \Omega}} -\frac{1}{2}|||\mathrm{div}\,\vec{p} + K^*z|||^2_{H^{-1}} + \frac{1}{2}\|z\|^2_{L^2} - \frac{\gamma^{s-1}}{s\alpha^{s-1}}\|\vec{p}\|^s_{\boldsymbol{L}^s},$$

where $\gamma > 0$ is the regularization parameter and $1 < s \le 2$. Subtracting the term $s^{-1}\alpha^{1-s}\gamma^{s-1}\|\vec{p}\|^s_{\boldsymbol{L}^s}$ from the objective functional of $(\mathcal{P}^\star)$ results in an $\boldsymbol{L}^s$-uniformly concave objective functional in $(\mathcal{P}^\star_\gamma)$. Thus, $(\mathcal{P}^\star_\gamma)$ admits a unique solution $\bar{\vec{p}}_\gamma \in \boldsymbol{L}^2(\Omega)$ for every fixed $\gamma > 0$.

Due to the reflexivity of the involved function spaces, we compute the Fenchel dual of $(\mathcal{P}^\star_\gamma)$ and study the effect of the dual regularization in the primal problem. For the Fenchel calculus, we set $\vec{p} = -\vec{q}$ and

$$\begin{aligned}\mathcal{F}(\vec{q}) &:= I_{\{\vec{w}\in\boldsymbol{L}^2(\Omega):|\vec{w}|_s\le\alpha\}}(\vec{q}) + \frac{\gamma^{s-1}}{s\alpha^{s-1}}\|\vec{q}\|^s_{\boldsymbol{L}^s},\\ \mathcal{G}(\Lambda\vec{q}) &:= \frac{1}{2}|||K^*z - \mathrm{div}\,\vec{q}|||^2_{H^{-1}} - \frac{1}{2}\|z\|^2_{L^2},\ \ \Lambda\vec{q} := \mathrm{div}\,\vec{q},\end{aligned}$$

with $Y := H^{-1}(\Omega)$, $V := \boldsymbol{L}^2(\Omega)$. The Fenchel dual problem of $(\mathcal{P}^\star_\gamma)$ is given by

$$(\mathcal{P}_\gamma) \qquad \min_{u\in H^1_0(\Omega)} \frac{\mu}{2}\int_\Omega |\nabla u|^2\,dx + \frac{1}{2}\int_\Omega |Ku - z|^2\,dx + \alpha\int_\Omega \Phi_{r,\gamma}(\nabla u)\,dx,$$

where $r^{-1} + s^{-1} = 1$ and, for $\vec{w} \in \boldsymbol{L}^2(\Omega)$,

$$\Phi_{r,\gamma}(\vec{w})(x) := \begin{cases} |w(x)|_r - \frac{r-1}{r}\gamma^{1/(r-1)} & \text{if } |w(x)|_r \ge \gamma^{1/(r-1)},\\ \frac{1}{r\gamma}|w(x)|^r_r & \text{if } |w(x)|_r < \gamma^{1/(r-1)}.\end{cases}$$

The function $\Phi_{r,\gamma}$ is continuously differentiable. Hence, the objective functional in $(\mathcal{P}_\gamma)$ is continuously differentiable and strictly convex. Thus, $(\mathcal{P}_\gamma)$ admits a unique solution $\bar{u}_\gamma \in H^1_0(\Omega)$. Further, $\Phi_{r,\gamma}(\nabla u)$ represents a local smoothing of $\int_\Omega |\nabla u|_r\,dx$ and, as our analysis shows, it is connected to a particular regularization of the Fenchel dual of $(\mathcal{P}^\star)$. In the case $r = 2$ we have

$$\Phi_{2,\gamma}(\vec{w})(x) := \begin{cases} |w(x)|_2 - \frac{\gamma}{2} & \text{if } |w(x)|_2 \ge \gamma,\\ \frac{1}{2\gamma}|w(x)|^2_2 & \text{if } |w(x)|_2 < \gamma.\end{cases}$$

This choice is related to results in [5, p. 186] and [28], and it is known as Gauss-TV regularization. In medical imaging it is in many cases superior to the Perona–Malik filter [24], edge-flat-gray filter [16], and pure TV filtering. Let us also mention that $\Phi_{2,\gamma}$ is sometimes referred to as Huber function; see [10, 32].

By (2.2), the solutions $\bar{u}_\gamma$ and $\bar{\vec{p}}_\gamma$ of the regularized problems $(\mathcal{P}_\gamma)$ and $(\mathcal{P}_\gamma^\star)$, respectively, satisfy

$$-\mu\Delta\bar{u}_\gamma + K^*K\bar{u}_\gamma - \operatorname{div}\bar{\vec{p}}_\gamma = K^*z \ \text{ in } H^{-1}(\Omega), \tag{2.4a}$$

$$\left.\begin{array}{rll} \gamma\bar{\vec{p}}_\gamma - \alpha(|\nabla\bar{u}_\gamma|^{r-1}\star\sigma(\nabla\bar{u}_\gamma)) &= 0 \quad \text{if} \quad |\bar{\vec{p}}_\gamma|_s < \alpha, \\ |\nabla\bar{u}_\gamma|_r^{r-1}\bar{\vec{p}}_\gamma - \alpha(|\nabla\bar{u}_\gamma|^{r-1}\star\sigma(\nabla\bar{u}_\gamma)) &= 0 \quad \text{if} \quad |\bar{\vec{p}}_\gamma|_s = \alpha \end{array}\right\} \ \text{in } \boldsymbol{L}^2(\Omega). \tag{2.4b}$$

One can easily verify that (2.4b) can be equivalently expressed as

$$\max(\gamma, |\nabla\bar{u}_\gamma|_r^{r-1})\bar{\vec{p}}_\gamma - \alpha(|\nabla\bar{u}_\gamma|^{r-1}\star\sigma(\nabla\bar{u}_\gamma)) = 0. \tag{2.5}$$

This reformulation is fundamental for the algorithm which we propose in the next section.

Comparing (2.4b) and (2.3b) one finds that our regularization changes the optimality conditions only on subsets of Ω where $|\nabla\bar{u}_\gamma|_r < \gamma^{1/(r-1)}$, i.e., on sets where the pointwise constraint $|\bar{\vec{p}}_\gamma|_s \le \alpha$ is inactive. At sharp edges in the image, $|\nabla\bar{u}_\gamma|_r$ is large and (2.4b) becomes

$$|\nabla\bar{u}_\gamma|_r^{r-1}\bar{\vec{p}}_\gamma = \alpha(|\nabla\bar{u}_\gamma|^{r-1}\star\sigma(\nabla\bar{u})),$$

as it would be the case for the original problem $(\mathcal{P})$. As a consequence, our regularization does not smear sharp edges in the image. In addition, our local smoothing prevents the adverse *staircase* or *bulkiness* effect which typically occurs in TV regularization [16, 20]; see the discussion in section 5.2.

The following convergence result for $(\bar{u}_\gamma, \bar{\vec{p}}_\gamma)$ as $\gamma \to 0$ justifies our dual regularization analytically. The proof is given in Appendix A.

THEOREM 2.2. *The solutions $\bar{u}_\gamma$ of $(\mathcal{P}_\gamma)$ converge to the solution $\bar{u}$ of $(\mathcal{P})$ strongly in $H_0^1(\Omega)$ as $\gamma \to 0$. Moreover, as $\gamma \to 0$, the solutions $\bar{\vec{p}}_\gamma$ of $(\mathcal{P}_\gamma^\star)$ converge to a solution $\bar{\vec{p}}$ of $(\mathcal{P}^\star)$ weakly in $\boldsymbol{L}^2(\Omega)$.*

We conclude this section by pointing out another interesting relation between the dual regularization, $\Phi_{r,\gamma}$, and inf-convolution. In fact, $\Phi_{r,\gamma}$ can be obtained in a purely primal fashion. For this purpose consider for $w \in \mathbb{R}^n$ the inf-convolution

$$\phi_{r,\gamma}(w) := \inf\{|\hat{w}|_r + (r\gamma)^{-1}|w - \hat{w}|_r^r : \hat{w} \in \mathbb{R}^n\}.$$

One finds

$$\phi_{r,\gamma}(w) = \begin{cases} |w|_r - \frac{r-1}{r}\gamma^{1/(r-1)} & \text{if } |w|_r > \gamma^{1/(r-1)}, \\ \frac{1}{r\gamma}|w|_r^r & \text{if } |w|_r \le \gamma^{1/(r-1)}. \end{cases}$$

Thus, $\Phi_{r,\gamma}(\vec{q})(x) = \phi_{r,\gamma}(q(x))$ for $x \in \Omega$.

REMARK 2.3. *In $(\mathcal{P}_\gamma^\star)$ we could have chosen a different form of the regularization parameter. If we replace $\gamma^{s-1}/(s\alpha^{s-1})$ by $\tilde{\gamma}/s$, $\tilde{\gamma} > 0$, then (2.5) becomes*

$$\max(\alpha\tilde{\gamma}^{r-1}, |\nabla\bar{u}_{\tilde{\gamma}}|_r^{r-1})\bar{\vec{p}}_{\tilde{\gamma}} - \alpha(|\nabla\bar{u}_{\tilde{\gamma}}|^{r-1}\star\sigma(\nabla\bar{u}_{\tilde{\gamma}})) = 0. \tag{2.6}$$

Choosing $\tilde{\gamma} = \gamma^{s-1}/\alpha^{s-1}$, (2.6) is identical to (2.5).

3. Primal-dual algorithm and its convergence. As it turned out in recent work, primal-dual schemes are typically superior to purely primal methods; see [6, 14]. Thus, we take (2.4a) and (2.5) as the starting point for devising an algorithm for solving the discrete analogue of $(\mathcal{P}_\gamma)$. In this way we naturally involve the primal variable

u and the dual variable $\vec{p}$. Consequently, we compute the solutions to the discrete versions of $(\mathcal{P}_\gamma)$ and $(\mathcal{P}_\gamma^\star)$ simultaneously. The algorithm proposed here is related to semismooth Newton methods [13, 25, 30]. In what follows we denote discretized quantities by superscript h. For a vector $w \in \mathbb{R}^l$ we denote by $D(w) := \operatorname{diag}(w)$ the $l \times l$ diagonal matrix with diagonal entries w_i.

The discrete analogues of (2.4a) and (2.5) are

$$B_\mu^h \bar{u}_\gamma^h - \operatorname{div}^h \bar{p}_\gamma^h = (K^h)^\top z^h, \tag{3.1a}$$

$$\max(\gamma e^h, \eta_r(\nabla^h \bar{u}_\gamma^h)) \star \bar{p}_\gamma^h - \alpha\big(|\nabla^h \bar{u}_\gamma^h|^{r-1} \star \sigma(\nabla^h \bar{u}^h)\big) = 0, \tag{3.1b}$$

where $\bar{u}_\gamma^h \in \mathbb{R}^m$, $\bar{p}_\gamma^h \in \mathbb{R}^{2m}$, for some $m \in \mathbb{N}$ which depends on the image size, solve the discrete analogues of $(\mathcal{P}_\gamma)$ and $(\mathcal{P}_\gamma^\star)$. Further, $z^h \in \mathbb{R}^m$ and $\nabla^h \in \mathbb{R}^{2m\times m}$, and $e^h \in \mathbb{R}^{2m}$ is the vector of all ones. We use $B_\mu^h = -\mu\Delta^h + (K^h)^\top K^h$, with $\Delta^h, (K^h)^\top K^h \in \mathbb{R}^{m\times m}$. Throughout we assume that B_μ^h is symmetric and positive definite. Note that this allows $\mu = 0$ if $(K^h)^\top K^h$ is invertible. The function $\eta_r : \mathbb{R}^{2m} \to \mathbb{R}^{2m}$ is defined by

$$(\eta_r(v))_i = (\eta_r(v))_{i+m} = |(v_i, v_{i+m})^\top|_r^{r-1} \quad \text{for } v \in \mathbb{R}^{2m},\ 1 \le i \le m.$$

The discrete gradient operator is composed as

$$\nabla^h = \begin{pmatrix} \nabla_x^h \\ \nabla_y^h \end{pmatrix} \in \mathbb{R}^{2m\times m},$$

where $\nabla_x^h \in \mathbb{R}^{m\times m}$ corresponds to the discrete derivative in the x-direction and $\nabla_y^h \in \mathbb{R}^{m\times m}$ corresponds to the derivative in the y-direction. In our implementation we use

$$\operatorname{div}^h = -(\nabla^h)^\top$$

and the implied ordering for the components of $\bar{p}_\gamma^h$. Subsequently we will frequently use the mapping $\xi_t : \mathbb{R}^{2m} \to \mathbb{R}^{2m}$ defined by

$$\xi_t(v)_i = \xi_t(v)_{i+m} := |(v_i, v_{i+m})^\top|_t$$

for $v \in \mathbb{R}^{2m}$, $i = 1, \dots, m$, and $1 < t < +\infty$. Note that $|\xi_t(v)|^{t-1} = \eta_t(v)$.

3.1. Approximate generalized Newton step. In the following we present our algorithm for the solution of (3.1a)–(3.1b). We restrict our discussion to the case $2 \le r < +\infty$ (and thus $1 < s \le 2$). It is well known that the max-operator and the ℓ_r-norm involved in (3.1b) are generalized differentiable and semismooth [13, 22, 25]. This is also true for the composition of the semismooth functions occurring in (3.1b). A particular element of the generalized Jacobian (see [8, 13, 25]) of $\max : \mathbb{R}^l \to \mathbb{R}^l$ is the diagonal matrix $G_{\max} \in \mathbb{R}^{l\times l}$ with

$$(G_{\max}(w))_{ii} := \begin{cases} 1 & \text{if } w_i \ge 0, \\ 0 & \text{if } w_i < 0 \end{cases} \quad \text{for } 1 \le i \le l.$$

Given approximations u_k^h, p_k^h of $\bar{u}_\gamma^h$, $\bar{p}_\gamma^h$, the results on generalized differentiability and semismoothness alluded to above and our assumption that $r \ge 2$ justify the

application of a Newton step to (3.1a)–(3.1b) at (u_k^h, p_k^h):

$$\begin{pmatrix} B_\mu^h & (\nabla^h)^\top \\ \big(-\alpha(r-1)D(|\nabla^h u_k^h|^{r-2}) + \chi_{\mathcal{A}_{k+1}} D(p_k^h) N_r^h(\nabla^h u_k^h)\big)\nabla^h & D(m_k^h) \end{pmatrix} \begin{pmatrix} \delta_u \\ \delta_p \end{pmatrix} = \begin{pmatrix} -B_\mu^h u_k^h - (\nabla^h)^\top p_k^h + (K^h)^\top z^h \\ \alpha\big(|\nabla^h u_k^h|^{r-1} \star \sigma(\nabla^h u_k^h)\big) - D(m_k^h) p_k^h \end{pmatrix}, \tag{3.2}$$

where we use $m_k^h = \max(\gamma e^h, \eta_r(\nabla^h u_k^h)) \in \mathbb{R}^{2m}$ for convenience, and $\chi_{\mathcal{A}_{k+1}} = D(t_k^h) \in \mathbb{R}^{2m\times 2m}$ with

$$(t_k^h)_i := \begin{cases} 1 & \text{if } \eta_r(\nabla^h u_k^h)_i \geq \gamma, \\ 0 & \text{else.} \end{cases} \tag{3.3}$$

Further, N_r^h denotes the Jacobian of η_r; i.e., for $v = (v_x, v_y)^\top \in \mathbb{R}^{2m}$, with $v_x, v_y \in \mathbb{R}^m$, we have

$$N_r^h(v) = (r-1)\big(D(\xi_r(v))\big)^{-1} \begin{pmatrix} D(|v_x|^{r-1} \star \sigma(v_x)) & D(|v_y|^{r-1} \star \sigma(v_y)) \\ D(|v_x|^{r-1} \star \sigma(v_x)) & D(|v_y|^{r-1} \star \sigma(v_y)) \end{pmatrix}.$$

Since $D(m_k^h)$ is invertible, we can eliminate δ_p from the Newton system (3.2). The remaining equation for δ_u is written as

$$H_k \delta_u = f_k, \tag{3.4}$$

where the matrix H_k and the right-hand side f_k are defined by

$$H_k := B_\mu^h + (\nabla^h)^\top D(m_k^h)^{-1}\big(\alpha(r-1)D(|\nabla^h u_k^h|^{r-2}) - \chi_{\mathcal{A}_{k+1}} D(p_k^h) N_r^h(\nabla^h u_k^h)\big)\nabla^h,$$

$$f_k := -B_\mu^h u_k^h + (K^h)^\top z^h - \alpha \nabla^h D(m_k^h)^{-1}\big(|\nabla^h u_k^h|^{r-1} \star \sigma(\nabla^h u_k^h)\big).$$

Note that H_k is the Schur complement of $D(m_k^h)$ in the system matrix in (3.2). Further properties of H_k are addressed next.

First, note that H_k is in general not symmetric, since

$$C_k^h := \alpha(r-1)D(|\nabla^h u_k^h|^{r-2}) - \chi_{\mathcal{A}_{k+1}} D(p_k^h) N_r^h(\nabla^h u_k^h)$$

is not. However, in the solution, i.e., for $(u_k^h, p_k^h) = (\bar{u}_\gamma^h, \bar{p}_\gamma^h)$ we infer from (3.1b) that $(p_k^h)_i = \alpha(\eta_r \nabla^h u_k^h))_i^{-1} |\nabla^h u_k^h|_i^{r-2} \sigma(\nabla^h u_k^h)_i$ holds for all $i \in \{1, \dots, 2m\}$ with $\eta_r(\nabla^h u_k^h)_i \geq \gamma$. Thus,

$$\begin{aligned} &\chi_{\mathcal{A}^{k+1}} D(p_k^h) N_r^h(\nabla^h u_k^h) \\ &\quad = (r-1)\chi_{\mathcal{A}^{k+1}}\big(D(\xi_r(\nabla^h u_k^h))\big)^{-1} D(p_k^h) \\ &\qquad\qquad \cdot \begin{pmatrix} D(|\nabla_x^h u_k^h|^{r-1} \star \sigma(\nabla_x^h u_k^h)) & D(|\nabla_y^h u_k^h|^{r-1} \star \sigma(\nabla_y^h u_k^h)) \\ D(|\nabla_x^h u_k^h|^{r-1} \star \sigma(\nabla_x^h u_k^h)) & D(|\nabla_y^h u_k^h|^{r-1} \star \sigma(\nabla_y^h u_k^h)) \end{pmatrix} \\ &\quad = \alpha(r-1)\chi_{\mathcal{A}^{k+1}}\big(D(\xi_r(\nabla^h u_k^h))\big)^{-r} \begin{pmatrix} D_1 & D_2 \\ D_2 & D_3 \end{pmatrix}, \end{aligned}$$

where

$$\begin{aligned} D_1 &= D\big(|\nabla_x^h u_k^h|^{2r-2}\big), \\ D_2 &= D\big(|\nabla_x^h u_k^h|^{r-1} \star |\nabla_y^h u_k^h|^{r-1} \star \sigma(\nabla_x^h u_k^h) \star \sigma(\nabla_y^h u_k^h)\big), \\ D_3 &= D\big(|\nabla_y^h u_k^h|^{2r-2}\big). \end{aligned}$$

Hence, in the solution the matrix C_k^h—and thus the entire system matrix H_k—is symmetric.

In what follows we investigate definiteness properties of the matrix H_k. Whenever H_k is *not* positive definite, these observations lead us to simple modifications to obtain a uniformly positive definite matrix. This guarantees that in each (approximate) Newton step a descent direction is computed.

We start by considering properties of C_k^h. Note that one can reorder the indices such that $D(p_k^h)N_r^h(\nabla^h u_k^h)$ becomes a block-diagonal matrix where every 2×2 diagonal block has the structure

$$(n_k^h)_i := \frac{r-1}{(\xi_r(\nabla^h u_k^h))_i}\begin{pmatrix} (p_k^h)_i|\nabla_x^h u_k^h|_i^{r-1}\sigma(\nabla_x^h u_k^h)_i & (p_k^h)_i|\nabla_y^h u_k^h|_i^{r-1}\sigma(\nabla_y^h u_k^h)_i \\ (p_k^h)_{i+m}|\nabla_x^h u_k^h|_i^{r-1}\sigma(\nabla_x^h u_k^h)_i & (p_k^h)_{i+m}|\nabla_y^h u_k^h|_i^{r-1}\sigma(\nabla_y^h u_k^h)_i \end{pmatrix}$$

for $1 \le i \le m$. By this reordering, C_k^h is transformed into a block-diagonal matrix with diagonal blocks

$$(c_k^h)_i := \alpha(r-1)\begin{pmatrix} |\nabla_x^h u_k^h|_i^{r-2} & 0 \\ 0 & |\nabla_y^h u_k^h|_i^{r-2} \end{pmatrix} - (t_k^h)_i(n_k^h)_i, \tag{3.5}$$

where $(t_k^h)_i \in \{0,1\}$ is defined in (3.3). Obviously, $(c_k^h)_i$ is positive semidefinite for all inactive indices, i.e., for all i with $(t_k^h)_i = 0$. It remains to find conditions which imply that $(c_k^h)_i$ is positive semidefinite whenever $(t_k^h)_i = 1$. It is well known that positive semidefiniteness of $(c_k^h)_i$ is characterized by nonnegativity of all eigenvalues of its symmetrization. Therefore, we compute the eigenvalues $\lambda((\hat{c}_h^k)_i)^+$ and $\lambda((\hat{c}_h^k)_i)^-$ of $(\hat{c}_k^h)_i := \big((c_k^h)_i + (c_k^h)_i^\top\big)/2$, where

$$(c_k^h)_i = (r-1)\begin{pmatrix} a_{k,i} & b_{k,i} \\ c_{k,i} & d_{k,i} \end{pmatrix}$$

with

$$\begin{aligned}
a_{k,i} &= \alpha|\nabla_x^h u_k^h|_i^{r-2} - (\xi_r(\nabla^h u_k^h))_i^{-1}(p_k^h)_i|\nabla_x^h u_k^h|_i^{r-1}\sigma(\nabla_x^h u_k^h)_i, \\
b_{k,i} &= -(\xi_r(\nabla^h u_k^h))_i^{-1}(p_k^h)_i|\nabla_y^h u_k^h|_i^{r-1}\sigma(\nabla_y^h u_k^h)_i, \\
c_{k,i} &= -(\xi_r(\nabla^h u_k^h))_i^{-1}(p_k^h)_{i+m}|\nabla_x^h u_k^h|_i^{r-1}\sigma(\nabla_x^h u_k^h)_i, \\
d_{k,i} &= \alpha|\nabla_y^h u_k^h|_i^{r-2} - (\xi_r(\nabla^h u_k^h))_i^{-1}(p_k^h)_{i+m}|\nabla_y^h u_k^h|_i^{r-1}\sigma(\nabla_y^h u_k^h)_i.
\end{aligned}$$

Since we are considering the case $(t_k^h)_i = 1$, from (3.3) it follows that $(\xi_r(\nabla^h u_k^h))_i > 0$ and thus the above expressions are well defined. For the eigenvalues of $(\hat{c}_k^h)_i$ we obtain

$$\lambda((\hat{c}_k^h)_i)^\pm = \frac{a_{k,i}+d_{k,i}}{2} \pm \sqrt{\left(\frac{a_{k,i}+d_{k,i}}{2}\right)^2 + \left(\frac{b_{k,i}+c_{k,i}}{2}\right)^2 - a_{k,i}d_{k,i}}. \tag{3.6}$$

Let us now assume that the following two conditions hold:

$$\text{(C)} \qquad \xi_s(p_k^h)_i \le \alpha \quad \text{and} \quad (b_{k,i}+c_{k,i})^2 \le 4a_{k,i}d_{k,i}.$$

The first condition implies that $a_{k,i} \ge 0$ and $d_{k,i} \ge 0$. The second condition yields that $(b_{k,i}+c_{k,i})^2/4 - a_{k,i}d_{k,i}$ is nonpositive. Comparing with (3.6) we obtain that both eigenvalues $\lambda((\hat{c}_k^h)_i)^\pm$ are nonnegative real numbers, and hence $(c_k^h)_i$ is positive semidefinite.

Let us briefly interpret the conditions in (C). The condition $\xi_s(p_k^h)_i \le \alpha$ corresponds to feasibility of the dual variable. Assuming that this condition holds, the latter condition in (C) is more likely to be satisfied if the matrix $(c_k^h)_i$ is close to a symmetric matrix, which can be seen as follows: assuming $b_{k,i} = c_{k,i}$ we obtain

(3.7)
$$\begin{aligned}\frac{(b_{k,i}+c_{k,i})^2}{4} - a_{k,i}d_{k,i} &= b_{k,i}c_{k,i} - a_{k,i}d_{k,i}\\ &= -\alpha^2|\nabla_x^h u_k^h|_i^{r-2}|\nabla_y^h u_k^h|_i^{r-2} + \alpha(\xi_r(\nabla^h u_k^h))_i^{-1}\\ &\qquad \cdot |\nabla_x^h u_k^h|_i^{r-2}|\nabla_y^h u_k^h|_i^{r-2}\left((p_k^h)_{i+m}(\nabla_y^h u_k^h)_i + (p_k^h)_i(\nabla_x^h u_k^h)_i\right)\\ &\le \alpha|\nabla_x^h u_k^h|_i^{r-2}|\nabla_y^h u_k^h|_i^{r-2}(-\alpha + (\xi_s(p_k^h))_i)\\ &\le 0.\end{aligned}$$

Above, the Hölder inequality and the assumption that $\xi_s(p_k^h)_i \le \alpha$ are used. We summarize our results in the following lemma.

LEMMA 3.1. *Assume that condition* (C) *holds for all* $i \in \{1,\dots,m\}$. *Then the matrix* C_k^h *is positive semidefinite.*

An immediate corollary of Lemma 3.1 is stated next.

COROLLARY 3.2. *Let the assumption* (C) *hold for all* $i \in \{1,\dots,m\}$ *and* $k \in \mathbb{N}$. *Then, for all* $k \in \mathbb{N}$, *the matrix* H_k *is positive definite, and* $\lambda_{\min}(H_k) \ge \lambda_{\min}(B_\mu^h) > 0$. *Moreover, the sequence* $\{H_k^{-1}\}_{k\in\mathbb{N}}$ *is uniformly bounded.*

Proof. The assertions follow from the fact that $B_\mu^h = -\mu\Delta^h + (K^h)^\top K^h$ is symmetric and positive definite by assumption and from Lemma 3.1. ∎

In the case $r = s = 2$ the second assumption in (C) is automatically satisfied, as can be seen from the next lemma.

LEMMA 3.3. *In the case* $r = s = 2$, *the first condition in* (C) *implies the latter; i.e., feasibility of the dual variable implies positive semidefiniteness of the matrix* C_k^h.

Proof. Suppose that $r = s = 2$ and that the first condition in (C) is satisfied. Then, for $k \in \mathbb{N}$ and $i \in \{1,\dots,m\}$,

$$\begin{aligned}&(b_{k,i}+c_{k,i})^2/4 - a_{k,i}d_{k,i}\\ &\quad= (2\xi_2(\nabla^h u_k^h)_i)^{-2}\Big((p_k^h)_i^2(\nabla_y u_k^h)_i^2 + (p_k^h)_{i+m}^2(\nabla_x u_k^h)_i^2\\ &\qquad\qquad - 2(p_k^h)_i(p_k^h)_{i+m}(\nabla_x u_k^h)_i(\nabla_y u_k^h)_i\Big) - \alpha^2\\ &\qquad + \alpha\xi_2(\nabla^h u_k^h)_i^{-1}\left((p_k^h)_i(\nabla_x u_k^h)_i + (p_k^h)_{i+m}(\nabla_y u_k^h)_i\right)\\ &\quad= (2\xi_2(\nabla^h u_k^h)_i)^{-2}\Big((p_k^h)_i^2 + (p_k^h)_{i+m}^2\big)\big((\nabla_x u_k^h)_i^2 + (\nabla_y u_k^h)_i^2\big)\\ &\qquad\qquad - \big((p_k^h)_i(\nabla_x u_k^h)_i + (p_k^h)_{i+m}(\nabla_y u_k^h)_i\big)^2\Big) - \alpha^2\\ &\qquad + \alpha\xi_2(\nabla^h u_k^h)_i^{-1}\left((p_k^h)_i(\nabla_x u_k^h)_i + (p_k^h)_{i+m}(\nabla_y u_k^h)_i\right)\\ &\quad= -\Big(\alpha - (2\xi_2(\nabla^h u_k^h)_i)^{-1}\big((p_k^h)_i(\nabla_x u_k^h)_i + (p_k^h)_{i+m}(\nabla_y u_k^h)_i\big)\Big)^2\\ &\qquad + \big((p_k^h)_i^2 + (p_k^h)_{i+m}^2\big)/4\\ &\quad\le -\alpha^2/4 + \big((p_k^h)_i^2 + (p_k^h)_{i+m}^2\big)/4\\ &\quad\le 0.\end{aligned}$$

Thus, for $r = s = 2$, the second condition in (C) follows from $\xi_s(p_k^h)_i \le \alpha$ as asserted in the lemma. ∎

Whenever (C) is satisfied, Lemma 3.1 and Corollary 3.2 guarantee that the solution of (3.4) exists for all k and that it is a descent direction for the objective function in $(\mathcal{P}_\gamma)$. In general, however, the assumptions in (C) are unlikely to hold for all $i \in \{1, \dots, m\}$ and $k \in \mathbb{N}$. To guarantee the positive definiteness of the system matrix in our Newton-type step, in the subsequent algorithm we modify the term involving $D(p_k^h)N_r^h(\nabla^h u_k^h)$ for indices i which violate (C).

As above, we work with the reordered system matrix and consider for $i \in \{1, \dots, m\}$ the 2×2 matrix $(c_k^h)_i$ as defined in (3.5). To enforce the first condition in (C) we replace $\big((p_k^h)_i, (p_k^h)_{i+m}\big)$ by $\alpha \max\big(\alpha, \xi_s(p_k^h)_i\big)^{-1}\big((p_k^h)_i, (p_k^h)_{i+m}\big)$ when assembling the system matrix; i.e., we project p_k^h onto the feasible set. The resulting 2×2 matrix is denoted by

$$(\tilde{c}_k^h)_i = \begin{pmatrix} \tilde{a}_{k,i} & \tilde{b}_{k,i} \\ \tilde{c}_{k,i} & \tilde{d}_{k,i} \end{pmatrix}.$$

Due to the projection step we have $\tilde{a}_{k,i} \ge 0$ and $\tilde{d}_{k,i} \ge 0$. Then we check whether $(\tilde{b}_{k,i} + \tilde{c}_{k,i})^2/4 - \tilde{a}_{k,i}\tilde{d}_{k,i} \le 0$ is satisfied. If this is not the case we dampen the off-diagonal elements by $\nu_{k,i} = 2\sqrt{\tilde{a}_{k,i}\tilde{d}_{k,i}}/|\tilde{b}_{k,i} + \tilde{c}_{k,i}|$. From this we obtain

$$(\tilde{\tilde{c}}_k^h)_i = \begin{pmatrix} \tilde{a}_{k,i} & \nu_{k,i}\tilde{b}_{k,i} \\ \nu_{k,i}\tilde{c}_{k,i} & \tilde{d}_{k,i} \end{pmatrix}.$$

The resulting 2×2 matrix is positive semidefinite; i.e., its symmetrization has nonnegative eigenvalues. Revoking the reordering of the indices results in a modified system matrix which we denote by H_k^+. It replaces H_k in (3.4). We have the following result (compare with Corollary 3.2, also for a proof).

LEMMA 3.4. *The matrix H_k^+ is positive definite and $\lambda_{\min}(H_k^+) \ge \lambda_{\min}(B_\mu^h) > 0$, respectively, for every $k \in \mathbb{N}$. Moreover, the sequence $\{(H_k^+)^{-1}\}_{k\in\mathbb{N}}$ is uniformly bounded.*

3.2. Algorithm. Now we are prepared for defining our approximate generalized Newton method for solving (3.1a)–(3.1b). Whenever (C) is not satisfied, it operates with H_k^+; otherwise the true system matrix H_k is used. In the next section we show that the new algorithm converges at a locally superlinear rate. This convergence result parallels the one for generalized Newton methods; see, e.g., [13, 30].

We propose the following algorithm.

ALGORITHM.

1. Initialize $(u_0^h, p_0^h) \in \mathbb{R}^m \times \mathbb{R}^{2m}$ and set $k = 0$.
2. Estimate the active sets, i.e., determine $\chi_{\mathcal{A}_{k+1}} \in \mathbb{R}^{2m\times2m}$.
3. Compute H_k^+ if (C) is not satisfied for all $i = \{1, \dots, m\}$; otherwise set $H_k^+ = H_k$. Solve
$$H_k^+ \delta_u = f_k$$
for δ_u.
4. Compute δ_p according to (3.2).
5. Update $u_{k+1}^h := u_k^h + \delta_u$, $p_{k+1} := p_k^h + \delta_p$.
6. Stop, or set $k := k + 1$ and go to step 2.

Note that during the iteration of our algorithm the dual variable p is allowed to become infeasible with respect to the discrete version of the Fenchel dual problem $(\mathcal{P}_\gamma^\star)$. This differs significantly from the method in [6], where the dual iterates are forced to stay in the interior of the feasible region. Such a restriction may lead to prohibitively small step sizes in Newton's method.

3.3. Convergence analysis. In this section we investigate convergence properties of the above algorithm.

LEMMA 3.5. *Let $\bar{u}_\gamma^h$, $\bar{p}_\gamma^h$ be the solutions to the discrete analogues of $(\mathcal{P}_\gamma)$ and $(\mathcal{P}_\gamma^\star)$, respectively, and assume that $p_k^h \to \bar{p}_\gamma^h$ and $u_k^h \to \bar{u}_\gamma^h$. Then the modified system matrices H_k^+ converge to H_k as $k \to \infty$.*

Proof. Our argumentation again utilizes the reordered system. As above we argue for each 2×2 block separately. For inactive indices, i.e., for i with $(t_k^h)_i = 0$, the original and the modified 2×2 blocks coincide.

Let us now turn to active indices, i.e., to indices where $(t_k^h)_i = 1$. First note that the assumption $p_k^h \to \bar{p}_\gamma^h$ implies that $(\xi_s(p_k^h))_i \to (\xi_s(\bar{p}_\gamma^h))_i \leq \alpha$. Hence, for the projected dual variables it holds that

$$\alpha \max\big(\alpha, \xi_s(p_k^h)\big)^{-1}\big((p_k^h)_i, (p_k^h)_{i+m}\big) \to \big((\bar{p}_\gamma^h)_i, (\bar{p}_\gamma^h)_{i+m}\big).$$

Moreover, the matrix $(c_k^h)_i$ converges to a symmetric matrix as $(u_k^h, p_k^h) \to (\bar{u}_\gamma^h, \bar{p}_\gamma^h)$ (recall the discussion on page 8). Hence, the estimate (3.7) implies that $e_{k,i} := (b_{k,i} + c_{k,i})^2/4 - a_{k,i}d_{k,i} \to e_{\gamma,i} \leq 0$ for all i. If $e_{\gamma,i} < 0$, then $e_{k,i} < 0$ for all $k \geq k_0$ with some $k_0 \in \mathbb{N}$. Thus, for all $k \geq k_0$ no dampening of the off-diagonal elements occurs. If $e_{\gamma,i} = 0$, it is easy to see that $\nu_{k,i} \to 1$ as $p_k^h \to \bar{p}_\gamma^h$ and $u_k^h \to \bar{u}_\gamma^h$. Hence, all modified 2×2 diagonal blocks converge to the original blocks as $k \to \infty$. Revoking the reordering of the indices ends the proof. ∎

The local convergence of the above algorithm is the subject of the next theorem.

THEOREM 3.6. *The iterates (u_k^h, p_k^h) of the algorithm converge superlinearly to $(\bar{u}_\gamma^h, \bar{p}_\gamma^h)$ provided that (u_0^h, p_0^h) is sufficiently close to $(\bar{u}_\gamma^h, \bar{p}_\gamma^h)$.*

Proof. Recall that (3.1a)–(3.1b) are semismooth. However, since we possibly modify our system matrix, we end up with an approximate semismooth Newton method, and fast local convergence does not follow from standard arguments.

Utilizing a technique similar to that in the proof of Lemma 3.5 we can show that for each $\Delta > 0$ there exists a radius $\rho > 0$ such that if (u_k^h, p_k^h) is in a ρ-ball around $(\bar{u}_\gamma^h, \bar{p}_\gamma^h)$, then there exists $C_2 > 0$ with

$$\|H_k - H_k^+\| \leq \Delta \ \text{ and } \ \|H_k^+\| \leq C_2.$$

Thus, the assumptions of Theorem 4.1. in [30] are satisfied. Using this theorem we obtain that the iterates (u_k^h, p_k^h) converge to $(\bar{u}_\gamma^h, \bar{p}_\gamma^h)$ at a linear rate, provided that (u_0^h, p_0^h) is sufficiently close to $(\bar{u}_\gamma^h, \bar{p}_\gamma^h)$. Then, Lemma 3.5 implies that $H_k^+ \to H_k$. As a consequence, the assumptions of Theorem 4.2. in [30] are satisfied, which yields that the iterates converge locally superlinearly. ∎

As a globalization strategy for the above algorithm we use an Armijo linesearch technique with backtracking for the functional of the primal regularized problem $(\mathcal{P}_\gamma)$. Let us point out, however, that in our numerical experiments the step size is almost always equal to one; i.e., a full step of our generalized Newton algorithm leads to a sufficient decrease in the functional.

4. Aspects of the implementation. To illustrate the feasibility of our approach for large scale systems, our implementation utilizes an inexact solve of the linear system arising in step 3 of the algorithm.

In the case $r = 2$ (Gauss-TV regularization) the system matrix arising in our algorithm simplifies and no dampening of the off-diagonal elements is necessary to obtain positive definiteness of the system matrix (see Lemma 3.3). However, an immediate application of the preconditioned conjugate gradient (PCG) method for the solution of the linear system in each Newton step is not feasible since H_k^+ (like H_k) is in general not symmetric. Therefore, in our implementation we replace $N_r^h(\nabla^h u_k^h)$ by its symmetrization $\tilde{N}_k^h(\nabla^h u_k^h) := 1/2\big(N_r^h(\nabla^h u_k^h) + N_r^h(\nabla^h u_k^h)^\top\big)$. As a result, the system matrix $\tilde{H}_k^+$ is symmetric and the PCG method can be applied.

The symmetrization of H_k^+ is justified by the fact that in the solution the system matrix H_k (respectively, H_k^+) is symmetric and therefore coincides with its symmetrization $\tilde{H}_k$ (see page 8). Thus, using the symmetrization of H_k^+ in step 3 of our algorithm still results in a locally superlinearly convergent algorithm. Obviously, a similar observation holds true for the case $r > 2$. However, our numerical experience shows that proceeding with the nonsymmetric matrix gives slightly better results with respect to the number of primal-dual iterations (usually 0–3 fewer iterations are needed). Therefore, we utilize the biconjugate gradient stabilized (BICGSTAB) algorithm (see, e.g., [27]) to solve the nonsymmetric positive definite linear system arising in each Newton step. We point out that alternatively one may use specialized techniques such as the recently developed semiconjugate direction method; see [34].

Next we address the accuracy required when solving the Newton-type system in every iteration. For this purpose let res_k denote the residual of the nonlinear system. Motivated by the superlinear convergence rate we stop the PCG or the BICGSTAB iteration as soon as the norm of the residual of the linearized system drops below

$$\mathrm{tol}_{k+1} := 0.1 \min\left\{ \left(\frac{\mathrm{res}_k}{\mathrm{res}_0}\right)^{3/2}, \frac{\mathrm{res}_k}{\mathrm{res}_0} \right\}.$$

As a preconditioner for the system matrix we use an incomplete Choleski factor of the symmetrization of H_k^+. The algorithm is initialized with a Gaussian smoothing of the noisy image z and with $p_0^h = 0$. We stop the Newton iteration as soon as the initial residual is reduced by a factor of 10^{-6}.

5. Numerical results. In this section we present numerical results for the Gauss-TV case $r = 2$, as well as for the case $r > 2$. We discretize the Laplace operator with the standard five-point stencil with homogenous Dirichlet boundary conditions and use forward differences for the gradient and the corresponding discrete adjoint scheme for the divergence operator. Unless otherwise specified we concentrate on image denoising, i.e., $K = \mathrm{id}$. The choice of the parameters α and γ is based on the quality of the restored images, i.e., the parameters are adjusted manually.

5.1. Behavior of the algorithm. Here, we study the behavior of our proposed method with respect to its reconstruction ability, convergence properties such as speed of convergence, stability with respect to the noise level, the choice of r, and possible violations of the dual bound constraint during the iteration. In the following section we provide a qualitative study of the influence of the regularization parameter γ on, e.g., the staircase effect.

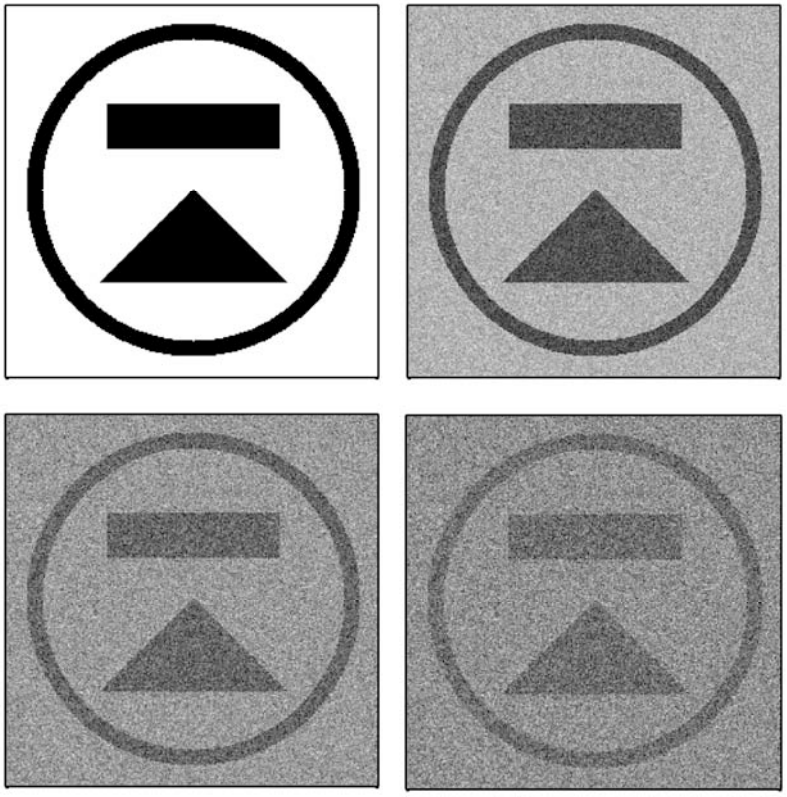

FIG. 5.1. *Example* 1: *Original image (*256×256 *pixels) and image with* 20% *Gaussian white noise (upper row), images containing* 50% *and* 80% *noise (lower row).*

TABLE 5.1
Example 1: *Number of primal-dual iterations (*#PD*) and overall number of CG iterations (*#CG*) for different noise levels.*

Noise	#PD	#CG
20%	11	48
50%	12	58
80%	13	61

5.1.1. Example 1. Our first test examples are displayed in Figure 5.1. On the upper left-hand side the original 256×256 image, which is similar to the ones used in [6, 14], is shown. A degraded image containing 20% of Gaussian white noise is given by the upper right plot. In the lower row images with a noise level of 50% respectively 80% are depicted. First we consider Gauss-TV regularization, i.e., $r = 2$ in $(\mathcal{P}_\gamma)$. We choose $\mu = 0$, $\gamma = 10^{-3}$, and $\alpha = 0.35$ for denoising the image with 20% noise. For the images containing 50% and 80% Gaussian white noise $\alpha = 0.90$ and $\alpha = 1.35$ are chosen, respectively. The numbers of primal-dual iterations of the algorithm for the three different noise levels are shown in Table 5.1, where we also provide the overall number of CG iterations. Note that our method behaves very stably with respect to the noise level. In our tests we also observe that the algorithm's performance is

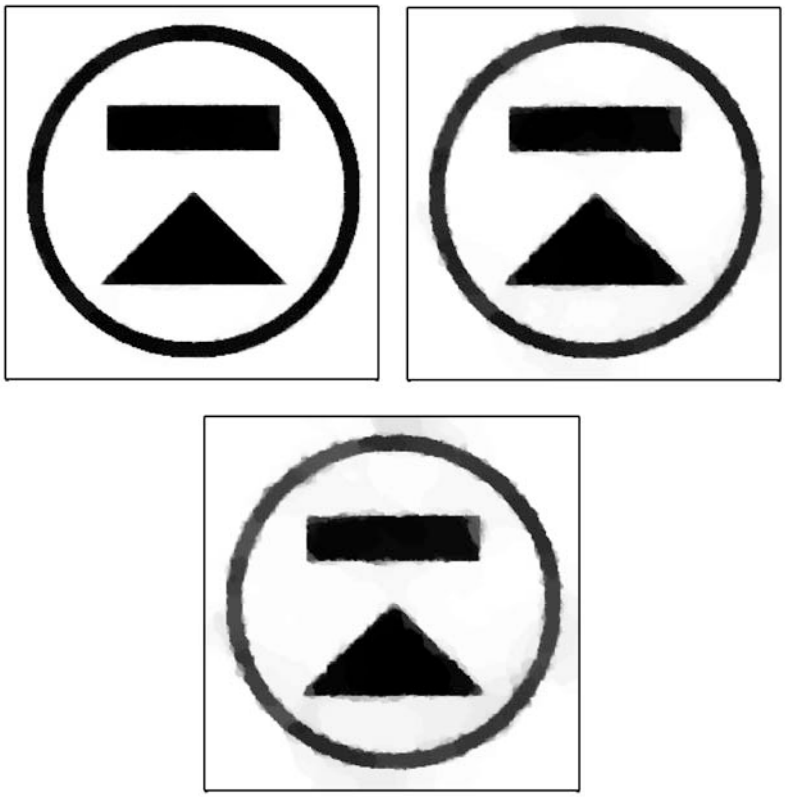

FIG. 5.2. *Example* 1: *Results of our algorithm with* $r = 2$ *for the image with* 20% *noise (upper row, left), with* 50% *noise (upper row, right), and with* 80% *noise (lower row).*

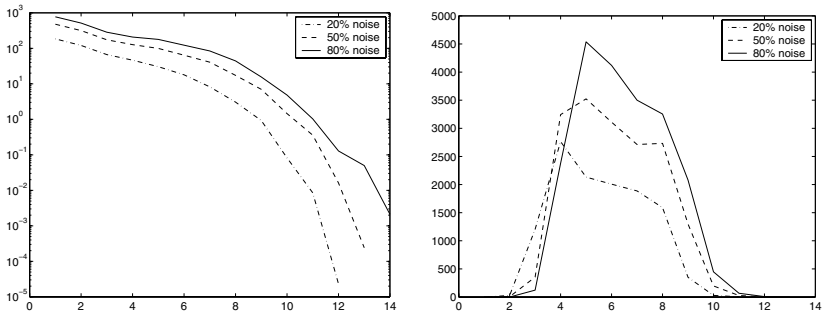

FIG. 5.3. *Example* 1: *Norm of the residual (left) and number of pixels where the dual variable is infeasible (right) versus iterations.*

rather independent of its initialization. Even for seemingly poor initializations u_0^h the algorithm converges and requires only two to three additional iterations when compared to a run initialized by the smoothed noisy image. The denoised images produced by our algorithm are shown in Figure 5.2. To document the convergence behavior we show the residual of the nonlinear system for the three noise levels in a logarithmic plot (Figure 5.3, left). We observe a fast decrease towards the end of

the iterations, which indicates a rapid local convergence of the method. To emphasize that our method, in contrast to the algorithm in, e.g., [6], allows violations of the dual constraints, in the right plot of Figure 5.3 we depict the number of points, where the iterates p_k^h are infeasible, i.e., where $|p_k^h|_s > (1+\varepsilon)\alpha$. Here $\varepsilon = 10^{-5}$ is introduced to compensate roundoff errors. Since the dual variable is initialized with $p_0^h = 0$, the number of infeasible points is small for the starting iterations, then it increases before it tends to zero as the algorithm approaches the solution.

Dependence on μ and γ. In our numerical tests we observe that the performance of the algorithm and the reconstructions are comparable for μ with $0 \le \mu \ll \alpha$. In particular, our choice $\mu = 0$ appears unproblematic. For large values of μ (i.e., $\mu \sim \alpha$ and larger), the quality of the reconstruction degrades (sharp edges are smoothed) and the number of overall PCG steps decreases. Concerning the influence of the parameter γ on the convergence behavior we observe that the method is not very sensitive with respect to γ. However, let us point out that large γ tends to decrease the iteration numbers.

Dependence on the image resolution. We also tested our algorithm for images having different resolutions. For this purpose, the image containing 50% noise in Figure 5.1 was chosen, and reconstructions for analog images having 128×128 and 512×512 pixels were computed. Our algorithm stopped after 11, 12, and 11 iterations and overall 51, 58, and 57 PCG iterations for the resolutions 128×128, 256×256, and 512×512, respectively. This shows a remarkable stability with respect to the image size.

Case $r > 2$. Next we test our algorithm with $r = 2.5$. For our test we use the image from Figure 5.1 containing 50% of Gaussian white noise and a rotated version of the same image. The parameters are set to $\mu = 0$, $\alpha = 0.9$, and $\gamma = 2.5 \cdot 10^{-5}$. We point out that for $r > 2$ the parameter γ must be chosen smaller as for $r = 2$ in order to obtain a reconstruction of comparable quality. This can be seen from the dual problem formulation ($\mathcal{P}_\gamma^\star$), where s also influences the weight of the regularizing term; see the discussion in Remark 2.3. For both, the original and the rotated image, our algorithm terminates after 20 primal-dual iterations. Overall, 61 and 65 BICGSTAB iterations are needed, respectively. We observe local fast convergence of the iterates as expected from Theorem 3.6. Furthermore, during the iterations many components of the dual iterates violate the pointwise constraints. Since we are in the case $r > 2$, off-diagonal entries may be dampened in order to guarantee positive definiteness of the system matrix (see the discussion on page 11). In our computations, for the rotated image only in the first 10 iterations a few (at most 71) off-diagonal elements are scaled by dampening factors between 0.7 and 1. After iteration 10 no dampening is necessary. We contribute this behavior to the fact that the system matrix is already close to a symmetric matrix for which no dampening of off-diagonal elements is required (see the discussions on page 8 and (3.7)).

Results of our computations are shown in Figure 5.4. The reconstruction of the rotated image differs from the rotated reconstruction of the original image. This is partly due to the fact that the unit ball of the ℓ_r-norm is not invariant with respect to rotations for $r \neq 2$. Obviously, there are other reasons for the different reconstructions as well: rotating introduces errors to the image. Furthermore, one-sided differences for the gradients influence the flow of information in the reconstruction process.

Simultaneous image deblurring and denoising. Finally, we illustrate the restoration ability of our algorithm for blurred noisy images, i.e., for $K \neq \mathrm{id}$. We consider the image from Figure 5.1 without noise and choose a block Toeplitz matrix (see, e.g., [32]) for the discrete blurring operator K^h. Block Toeplitz matrices occur as

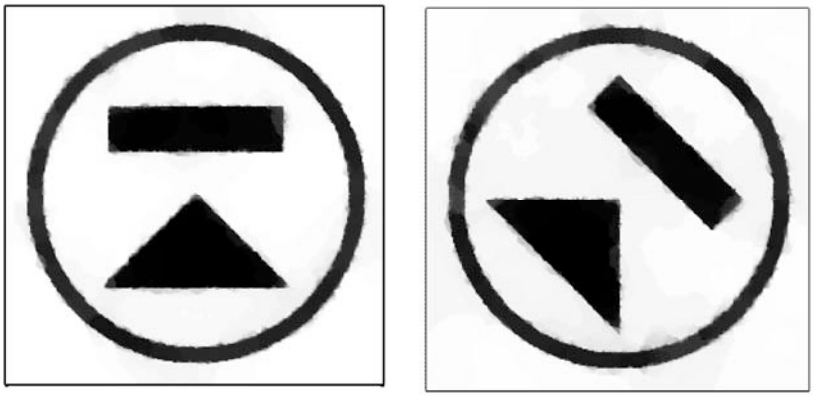

FIG. 5.4. *Example* 1: *Results of our calculations with* $r = 2.5$ *for the image from Figure* 5.1 *containing* 50% *of noise (left) and a rotated version of this image (right).*

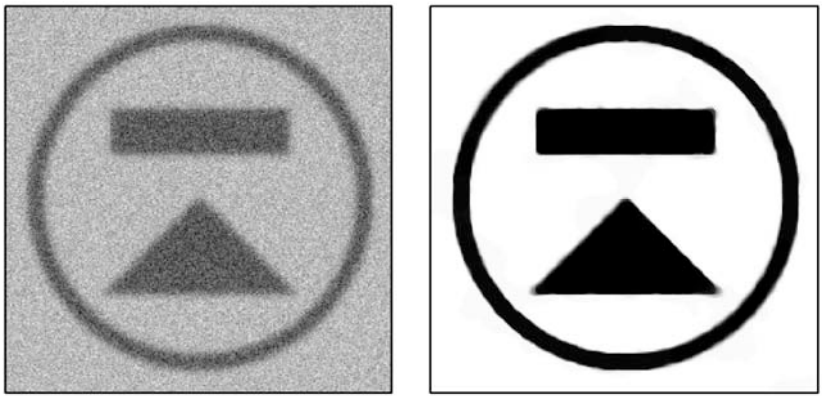

FIG. 5.5. *Example* 1: *Blurred noisy image (left) and reconstruction with* $r = 2$, $\alpha = 0.3$, *and* $\gamma = 10^{-3}$ *(right).*

matrix representations of discrete convolution operators, which represent a model for blurring of images. After applying K^h to the image, 20% of Gaussian white noise is added. The resulting degraded image can be seen on the left of Figure 5.5. For the reconstruction process we choose $r = 2$, $\alpha = 0.3$, $\gamma = 10^{-3}$, and $\mu = 0$. The result obtained after 16 primal-dual iterations can be seen on the right side of Figure 5.5.

5.1.2. Example 2 (validation of Theorem 3.6). In a second example we apply our method for enhancing the image shown in the left plot of Figure 5.6 that contains 10% of Gaussian white noise. Its resolution is 512×512 pixels. We apply Gauss-TV regularization, i.e., $r = 2$, with parameters $\mu = 0$, $\alpha = 0.08$, and $\gamma = 10^{-3}$. The result obtained after 11 iterations is shown in the right plot of Figure 5.6. To illustrate the convergence behavior of the method, in Table 5.2 we report on the norm of the nonlinear residual, the number of dually infeasible points, the step length, and the quotient

$$q_k = \frac{\|u_{k+1}^h - \bar{u}_\gamma^h\|}{\|u_k^h - \bar{u}_\gamma^h\|} \tag{5.1}$$

for each iteration. Above, $\bar{u}_\gamma^h$ denotes a beforehand calculated solution of the problem.

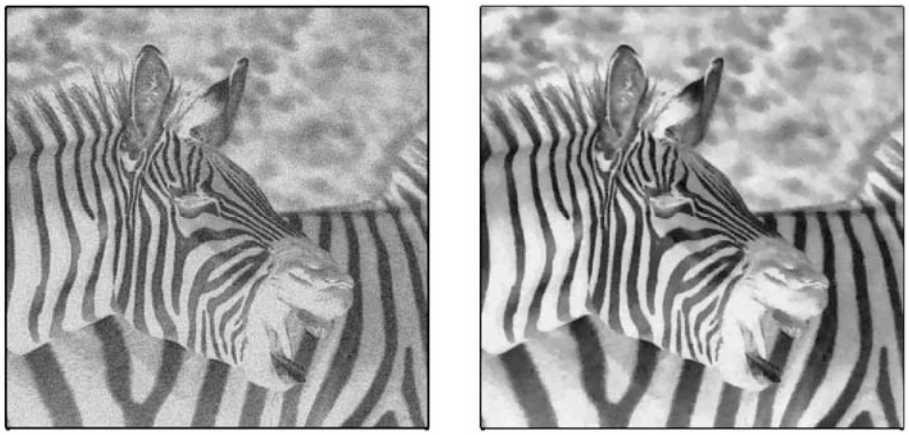

FIG. 5.6. *Example* 2: *Noisy* 512×512 *image (left) and results of our calculations with* $r = 2$, $\gamma = 10^{-3}$, *and* $\alpha = 0.08$ *(right).*

TABLE 5.2

Example 2: *For each iteration* k *we show the nonlinear residual* $\|res_k\|$, *the number* $\#CG_k$ *of PCG-iterations, the number* $\#|p_k^h|_2 > \alpha$ *of components of* p_k^h *that violate the dual constraint, the step size* β_k *and* q_k, *which is defined in* (5.1).

k	$\|res_k\|$	$\#\mathrm{CG}_k$	$\#\lvert p_k^h\rvert_2 > \alpha$	β_k	q_k
1	7.84e+1	2	0	1	0.412
2	4.12e+1	2	1425	1	0.391
3	2.01e+1	2	17284	1	0.361
4	1.31e+1	3	38990	1	0.315
5	9.48e 0	3	50246	1	0.265
6	7.10e 0	3	40267	1	0.249
7	3.68e 0	2	18641	1	0.272
8	1.37e 0	2	5346	1	0.239
9	3.81e–1	2	878	1	0.143
10	6.12e–2	3	73	1	0.086
11	3.44e–3	2	6	1	0.009
12	1.42e–5	2	0	1	0.000

We observe a decrease of q^k (except for iteration 7) which implies a locally superlinear convergence of the iterates (compare with Theorem 3.6). Again, the algorithm makes use of iterates that heavily violate the inequality constraints imposed on the dual variable.

Numerical experiments show that a robust and fast convergence of the method is essentially based on the replacement of the system matrix H_k by H_k^+. In fact, due to our approximation technique based on (C), a positive definite matrix is obtained which guarantees that δ_u is a descent direction reducing the objective function of $(\mathcal{P}_\gamma)$. Indeed, our line search always accepts step size $\beta_k = 1$ for achieving a sufficient decrease.

5.2. Qualitative study of the dual regularization. The image restoration model $(\mathcal{P}_{BV})$ tends to produce piecewise constant, i.e., "blocky" images. This may be useful in some applications. However, for images containing affine regions with noise, the result will be an adverse "staircase" effect. To prevent the method from

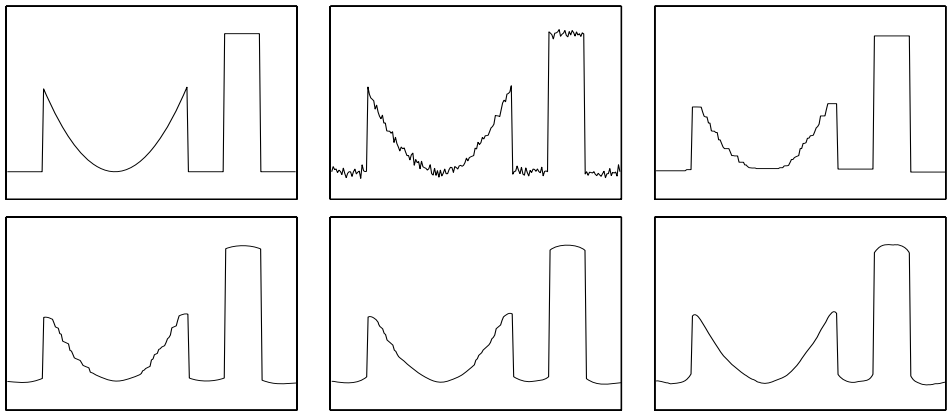

FIG. 5.7. *Upper row: Original one-dimensional image (left), noisy image (middle), and result of our calculations with* $\gamma = 10^{-10}$ *(right). Lower row: results obtained with* $\gamma = 0.1$ *(left),* $\gamma = 0.2$ *(middle), and* $\gamma = 0.5$ *(right).*

this oversharpening while preserving sharp edges, recently several strategies have been proposed and tested (see, e.g., [3, 7, 16]). In our approach, the parameter γ allows to control the BV model and the blocky structure in the image reconstruction. The regularization attached to γ induces a switch between a Gaussian-like regularization for small gradients and TV regularization for large gradients. To illustrate this influence of γ, a one-dimensional analogue to our algorithm was implemented. In Figure 5.7 the results for a one-dimensional image are shown, which were obtained by our algorithm with $\alpha = 10^{-2}$ and for various values of γ. The result for $\gamma = 0.5$ has exactly four active points at the solution. This implies that only at the four discontinuities of the original image TV regularization is applied, whereas all other regions are treated with a Gaussian regularization.

5.2.1. Example 3. Finally, we discuss the influence of γ on the two-dimensional noisy magnetic resonance (MR) image shown in Figure 5.8 (upper left). Its resolution is 256×256 pixels. A certain noise level in MR images is unavoidable due to background interference or other disturbances. Filters are used either to enhance the image quality without altering any image details or to prepare the image for a qualitative analysis, such as the segmentation of structures.

In [20] Gauss-TV denoising was shown to give good results for medical images. For this reason, we choose $r = 2$, the parameters $\mu = 0$ and $\alpha = 0.06$ and investigate the influence of the parameter γ on the reconstructed image. The result for $\gamma = 10^{-5}$ (see Figure 5.8, upper row, right) is obtained after 16 iterations. The reconstructions for $\gamma = 5 \cdot 10^{-3}$ and $\gamma = 5 \cdot 10^{-2}$ are obtained after 8 and 7 iterations of our algorithm, respectively. The corresponding results are shown in the lower row of Figure 5.8. Details of Figure 5.8 are shown in Figure 5.9. We observe that the reconstruction obtained with $\gamma = 10^{-5}$ is close to a piecewise constant image. This effect can be reduced by using larger γ values, e.g., $\gamma = 5 \cdot 10^{-3}$. However, depending on the noise level, if γ becomes too large, some of the noisy patterns may remain; see the result for $\gamma = 5 \cdot 10^{-2}$. In conclusion, this shows that the parameter γ can be useful in avoiding oversharpening of images while preserving sharp edges.

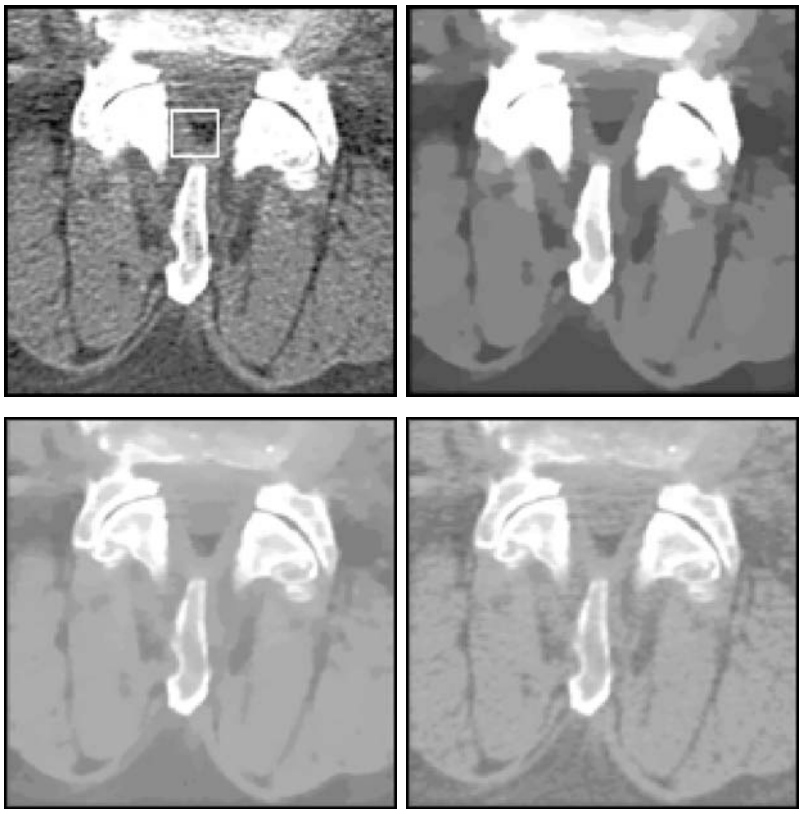

FIG. 5.8. *Example* 3: *Noisy magnetic resonance image (upper row, left, white box indicates the detail shown in Figure* 5.9*) and results of our calculations with* $r = 2$, $\mu = 0$, $\alpha = 0.06$ *and with* $\gamma = 10^{-5}$ *(upper row, right),* $\gamma = 5 \cdot 10^{-3}$ *(lower row, left), and* $\gamma = 5 \cdot 10^{-2}$ *(lower row, right).*

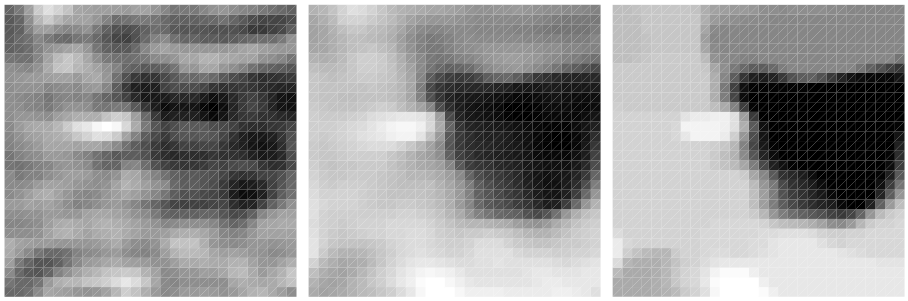

FIG. 5.9. *Example* 3: *Details of the images shown in Figure* 5.8. *Detail of the original image (right), reconstruction with* $\gamma = 5 \cdot 10^{-3}$ *(middle), and with* $\gamma = 10^{-5}$ *(right).*

Appendix A. Proof of Theorem 2.2. Subtracting the weak form of (2.4a) from the weak form of (2.3a) results in

$$\int_\Omega (\bar{u} - \bar{u}_\gamma) v \, dx + \mu \int_\Omega \nabla(\bar{u} - \bar{u}_\gamma)^\top \nabla v \, dx = \int_\Omega (\bar{\vec{p}}_\gamma - \bar{\vec{p}})^\top \nabla v \, dx \tag{A.1}$$

for all $v \in H_0^1(\Omega)$. Choosing $v := \bar{u} - \bar{u}_\gamma$ yields

$$\int_\Omega (\bar{u} - \bar{u}_\gamma)^2 \, dx + \mu \int_\Omega |\nabla(\bar{u} - \bar{u}_\gamma)|_2^2 \, dx = \int_\Omega (\bar{\vec{p}}_\gamma - \bar{\vec{p}})^\top \nabla(\bar{u} - \bar{u}_\gamma) \, dx. \tag{A.2}$$

Pointwise bounds for the term $(\bar{\vec{p}}_\gamma - \bar{\vec{p}})^\top \nabla(\bar{u} - \bar{u}_\gamma)$ are established as follows: First we split Ω into disjoint sets according to the complementarity conditions (2.3b) and (2.4b); namely, $\Omega = \mathcal{A}_\gamma \cup \mathcal{I}_\gamma = \mathcal{A} \cup \mathcal{I}$ with

$$\begin{aligned} \mathcal{A}_\gamma &= \{x \in \Omega : |\nabla \bar{u}_\gamma|_r^{r-1} > \gamma\}, \quad & \mathcal{I}_\gamma &= \Omega \setminus \mathcal{A}_\gamma, \\ \mathcal{A} &= \{x \in \Omega : |\nabla \bar{u}|_r^{r-1} > 0\}, & \mathcal{I} &= \Omega \setminus \mathcal{A}. \end{aligned} \tag{A.3}$$

Then we bound $(\bar{\vec{p}}_\gamma - \bar{\vec{p}})^\top \nabla(\bar{u} - \bar{u}_\gamma)$ separately on the four sets $\mathcal{A}_\gamma \cap \mathcal{A}$, $\mathcal{A}_\gamma \cap \mathcal{I}$, $\mathcal{I}_\gamma \cap \mathcal{A}$, and $\mathcal{I}_\gamma \cap \mathcal{I}$. Subsequently, we frequently dismiss the argument (x) when referring to function values at x. We start with $\mathcal{A}_\gamma \cap \mathcal{A}$, where we use the facts that $|\bar{\vec{p}}|_s = |\bar{\vec{p}}_\gamma|_s = \alpha$ and that $\bar{\vec{p}}_\gamma = \alpha\big(|\nabla \bar{u}_\gamma|^{r-1} \star (|\nabla \bar{u}_\gamma|)\big)/|\nabla \bar{u}_\gamma|_r^{r-1}$ and $\bar{\vec{p}} = \alpha\big(|\nabla \bar{u}|^{r-1} \star (|\nabla \bar{u}|)\big)/|\nabla \bar{u}|_r^{r-1}$:

$$\begin{aligned} (\bar{\vec{p}}_\gamma - \bar{\vec{p}})^\top \nabla(\bar{u} - \bar{u}_\gamma) &\le |\bar{\vec{p}}_\gamma|_s |\nabla \bar{u}|_r - \frac{\alpha}{|\nabla \bar{u}_\gamma|_r^{r-1}} |\nabla \bar{u}_\gamma|_r^r - \frac{\alpha}{|\nabla \bar{u}|_r^{r-1}} |\nabla \bar{u}|_r^r + |\bar{\vec{p}}|_s |\nabla \bar{u}_\gamma|_r \\ &= \alpha |\nabla \bar{u}|_r - \alpha |\nabla \bar{u}_\gamma|_r - \alpha |\nabla \bar{u}|_r + \alpha |\nabla \bar{u}_\gamma|_r = 0. \end{aligned} \tag{A.4}$$

Next we turn to the set $\mathcal{A}_\gamma \cap \mathcal{I}$, where we exploit the facts that $\nabla \bar{u} = 0$ and $|\bar{\vec{p}}_\gamma|_s = \alpha$:

$$\begin{aligned} (\bar{\vec{p}}_\gamma - \bar{\vec{p}})^\top \nabla(\bar{u} - \bar{u}_\gamma) &= (\bar{\vec{p}} - \bar{\vec{p}}_\gamma)^\top \nabla \bar{u}_\gamma \\ &\le |\bar{\vec{p}}|_s |\nabla \bar{u}_\gamma|_r - \frac{\alpha}{|\nabla \bar{u}_\gamma|_r^{r-1}} |\nabla \bar{u}_\gamma|_r^r \\ &= |\bar{\vec{p}}|_s |\nabla \bar{u}_\gamma|_r - \alpha |\nabla \bar{u}_\gamma|_r \le 0. \end{aligned} \tag{A.5}$$

On the set $\mathcal{I}_\gamma \cap \mathcal{A}$ one uses $|\bar{\vec{p}}_\gamma|_s \le \alpha$ and obtains similar to the above that

$$(\bar{\vec{p}}_\gamma - \bar{\vec{p}})^\top \nabla(\bar{u} - \bar{u}_\gamma) \le 0. \tag{A.6}$$

Finally, we turn to the set $\mathcal{I}_\gamma \cap \mathcal{I}$ where we have $\nabla \bar{u} = 0$, $|\nabla \bar{u}_\gamma|_r^{r-1} \le \gamma$, $|\bar{\vec{p}}|_s \le \alpha$, and $|\bar{\vec{p}}_\gamma|_s \le \alpha$. We obtain

$$\begin{aligned} (\bar{\vec{p}}_\gamma - \bar{\vec{p}})^\top \nabla(\bar{u} - \bar{u}_\gamma) &= (\bar{\vec{p}} - \bar{\vec{p}}_\gamma)^\top \nabla \bar{u}_\gamma \\ &\le |\bar{\vec{p}} - \bar{\vec{p}}_\gamma|_s |\nabla \bar{u}_\gamma|_r \\ &\le 2\alpha \gamma^{1/(r-1)}. \end{aligned} \tag{A.7}$$

Since $(\mathcal{A}_\gamma \cap \mathcal{A})$, $(\mathcal{A}_\gamma \cap \mathcal{I})$, $(\mathcal{I}_\gamma \cap \mathcal{A})$, $(\mathcal{I}_\gamma \cap \mathcal{I})$ provide a disjoint partitioning of Ω, the estimates (A.4)–(A.7) yield, together with (A.2),

$$\int_\Omega (\bar{u} - \bar{u}_\gamma)^2 \, dx + \mu \int_\Omega |\nabla(\bar{u} - \bar{u}_\gamma)|_2^2 \, dx \le \int_\Omega 2\alpha \gamma^{1/(r-1)} \, dx. \tag{A.8}$$

Hence, we infer that $\bar{u}_\gamma \to \bar{u}$ strongly in $H_0^1(\Omega)$ as $\gamma \to 0$. From (A.1) we obtain the weak convergence of $\bar{\vec{p}}_\gamma \to \bar{\vec{p}}$ in grad $H_0^1(\Omega) \subset \boldsymbol{L}^2(\Omega)$. This ends the proof. ∎

REFERENCES

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

[2] M. Bertalmio, V. Caselles, B. Rouge, and A. Sole, *TV-based image restoration with local constraints*, J. Sci. Comput., 19 (2003), pp. 95–122.

[3] P. Blomgren, T. F. Chan, and P. Mulet, *Extensions to total variation denoising*, in Proceedings Society of Photo-optical Instrumentation Engineers (SPIE)'97, San Diego, 1997.

[4] 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.

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

[6] T. Chan, G. Golub, and P. Mulet, *A nonlinear primal-dual method for total variation-based image restoration*, SIAM J. Sci. Comput., 20 (1999), pp. 1964–1977.

[7] T. Chan, A. Marquina, and P. Mulet, *High-order total variation-based image restoration*, SIAM J. Sci. Comput., 22 (2000), pp. 503–516.

[8] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Canadian Mathematical Society Series of Monographs and Advanced Texts, John Wiley & Sons, New York, 1983.

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

[10] D. C. Dobson and F. Santosa, *An image-enhancement technique for electrical impedance tomography*, Inverse Problems, 10 (1994), pp. 317–334.

[11] I. Ekeland and R. Témam, *Convex Analysis and Variational Problems*, Classics Appl. Math. 28, SIAM, Philadelphia, 1999.

[12] E. Giusti, *Minimal Surfaces and Functions of Bounded Variation*, Monogr. Math. 80, Birkhäuser-Verlag, Basel, 1984.

[13] 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.

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

[15] 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.

[16] 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.

[17] T. Kärkkäinen and K. Majava, *Nonmonotone and monotone active-set methods for image restoration.* I. *Convergence analysis*, J. Optim. Theory Appl., 106 (2000), pp. 61–80.

[18] T. Kärkkäinen and K. Majava, *Nonmonotone and monotone active-set methods for image restoration.* II. *Numerical results*, J. Optim. Theory Appl., 106 (2000), pp. 81–105.

[19] T. Kärkkäinen, K. Majava, and M. M. Mäkelä, *Comparison of formulations and solution methods for image restoration problems*, Inverse Problems, 17 (2001), pp. 1977–1995.

[20] S. L. Keeling, *Total variation based convex filters for medical imaging*, Appl. Math. Comput., 139 (2003), pp. 101–119.

[21] Y. Li and F. Santosa, *An Affine Scaling Algorithm for Minimizing Total Variation in Image Enhancement*, Report CTC94TR201, Cornell Theory Center, Cornell University, Ithaca, NY, 1994.

[22] R. Mifflin, *Semismooth and semiconvex functions in constrained optimization*, SIAM J. Control Optim., 15 (1977), pp. 959–972.

[23] S. Osher, A. Solé, and L. Vese, *Image Decomposition and Restoration Using Total Variation Minimization and the H^{-1}-Norm*, CAM-Report 02-57, UCLA, Los Angeles, CA, 2002.

[24] P. Perona and J. Malik, *Scale-space and edge detection using anisotropic diffusion*, IEEE Trans. Pattern Analysis and Machine Intelligence, 12 (1990), pp. 629–639.

[25] L. Q. Qi and J. Sun, *A nonsmooth version of Newton's method*, Math. Program., 58 (1993), pp. 353–367.

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

[27] Y. Saad, *Iterative Methods for Sparse Linear Systems*, 2nd ed., SIAM, Philadelphia, 2003.

[28] C. Schnörr, *Unique reconstruction of piecewise-smooth images by minimizing strictly convex nonquadratic functionals*, J. Math. Imaging Vision, 4 (1994), pp. 189–198.

[29] D. Strong and T. Chan, *Edge-preserving and scale-dependent properties of total variation regularization*, Inverse Problems, 19 (2003), pp. S165–S187.

[30] D. Sun and J. Han, *Newton and quasi-Newton methods for a class of nonsmooth equations and related problems*, SIAM J. Optim., 7 (1997), pp. 463–480.

[31] C. R. Vogel, *Nonsmooth regularization*, in Inverse Problems in Geophysical Applications (Yosemite, CA, 1995), H. W. Engl, A. K. Louis, and W. Rundell, eds., SIAM, 1997, pp. 1–11.

[32] C. R. Vogel, *Computational Methods for Inverse Problems*, Frontiers Appl. Math. 23, SIAM, Philadelphia, 2002.

[33] C. R. Vogel and M. E. Oman, *Iterative methods for total variation denoising*, SIAM J. Sci. Comput., 17 (1996), pp. 227–238.

[34] J. Y. Yuan, G. H. Golub, R. J. Plemmons, and W. A. G. Cecílio, *Semi-conjugate direction methods for real positive definite systems*, BIT, 44 (2004), pp. 189–207.