IMAGE DENOISING: LEARNING NOISE DISTRIBUTION VIA PDE-CONSTRAINED OPTIMIZATION

JUAN CARLOS DE LOS REYES[*] AND CAROLA-BIBIANE SCHÖNLIEB[†]

Abstract. We propose a PDE-constrained optimization approach for the determination of noise distribution in total variation (TV) image denoising. An optimization problem for the determination of the weights correspondent to different types of noise distributions is stated and existence of an optimal solution is proved. A tailored regularization approach for the approximation of the optimal parameter values is proposed thereafter and its consistency studied. Additionally, the differentiability of the solution operator is proved and an optimality system characterizing the optimal solutions of each regularized problem is derived. The optimal parameter values are numerically computed by using a quasi-Newton method, together with semismooth Newton type algorithms for the solution of the TV-subproblems.

1. Introduction. Let $f \in L^p(\Omega)$, $p = 1$ or 2 with $\Omega \subset \mathbb{R}^2$, be a given noisy image. Depending on the application at hand the type of noise, i.e., the noise distribution, changes [5]. Examples for noise distributions are Gaussian noise, which typically appears in, e.g. MRI (Magnetic Resonance Tomography), Poisson noise in, e.g. radar measurements or PET (Positron Emission Tomography), and impulse noise usually due to transmission errors or malfunctioning pixel elements in camera sensors. To remove the noise a total variation (TV) regularization is frequently considered [3, 11, 12, 13, 20, 37] that amounts to reconstruct a denoised version u of f as a minimiser of the generic functional

$$\mathcal{J}(u) = |Du|(\Omega) + \lambda \phi(u, f), \tag{1.1}$$

with

$$|Du|(\Omega) = \sup_{\mathbf{g} \in C_0^\infty(\Omega; \mathbb{R}^2), \|g\|_\infty \leq 1} \int_\Omega u \, \nabla \cdot \mathbf{g} \, dx \tag{1.2}$$

the total variation of u in Ω, λ a positive parameter and ϕ a suitable distance function called the data fidelity term. The latter depends on the statistics of the data f, which can be either estimated or approximated by a noise model known from the physics behind the acquisition of f. For normally distributed f, i.e. the interferences in f are Gaussian noise, this distance function is the squared L^2 norm of $u - f$. If a Poisson noise distribution is present, $\phi(u, f) = \int_\Omega \lambda \, (u - f \log u) \, dx$, which corresponds to the Kullback-Leibler distance between u and f [29, 33]. In the presence of impulse noise,

the correct data fidelity term turns out to be the L^1 norm of $u-f$ [32, 21]. Other noise models have been considered as well, cf. e.g. [2]. The size of the parameter λ depends on the strength of the noise, i.e. it models the trade-off between regularisation and fidelity to the measured datum f.

A key issue in total variation denoising is an adequate choice of the correct noise model, i.e. the choice of ϕ, and of the size of the parameter λ. Depending on this choice, different results are obtained. The term ϕ is usually modelled from the physics behind the acquisition process. Several strategies, both heuristic and statistically grounded, have been considered for choosing the weight λ, cf. e.g. [11, 22, 23, 24, 25, 35]. In this paper we propose an optimal control strategy for choosing both ϕ and λ. To do so we extend model (1.1) to a more general model, that allows for mixed noise distributions in the data. Namely, instead of (1.1) we consider

$$\min_u \left(|Du|(\Omega) + \sum_{i=1}^{d} \int_\Omega \lambda_i \; \phi_i(u,f) \; dx \right). \tag{1.3}$$

where ϕ_i, $i = 1, \ldots, d$, are convex differentiable functions in u, and λ_i are positive parameters. The functions ϕ_i model different choices of data fidelities. In the case of mixed Gaussian and impulse noise $d = 2$, $\phi_1(u,f) = \|u - f\|_{L^2(\Omega)}^2$ and $\phi_2(u,f) = \|u - f\|_{L^1(\Omega)}$. The parameters λ_i weight the different noise models ϕ_i and the regularising term against each other. As such, the choice of these parameters depends on the amount and strength of noise of different distributions in f. Typically, the λ_i are chosen to be real parameters. However, in some applications, it may be more favourable to choose them to be spatially dependent functions $\lambda_i : \Omega \to \mathbb{R}^+$, cf. e.g. [1, 4, 23, 25, 35].

We propose a PDE-constrained optimization approach to determine the weights λ_i of the noise distribution and, in that manner, learn the noise distribution present in the measured datum f for both $d = 1$ and mixed noise models $d > 1$. To do so, we treat (1.3) as a constraint and state an optimization problem governed by (1.3) for the optimal determination of weights. When possible, we replace the optimization problem by a necessary and sufficient optimality condition (in form of a variational inequality (VI)) as a constraint.

Schematically, we proceed in the following way:

1. We consider a training set of pairs (f_k, u_k), $k = 1, 2, \ldots, N$. Here, f_k's are noisy images, which have been measured with a fixed device with fixed settings, and the images u_k represent the ground truth or images that approximate the ground truth within a desirable tolerance.
2. We determine the optimal choice of functions λ_i by solving the following problem for $k = 1, 2, \ldots, N$

$$\min_{\lambda_i \geq 0, \; i=1,\ldots,d} \|\tilde{u} - u_k\|_{L^2(\Omega)}^2 + \beta \sum_{i=1}^{d} \|\lambda_i\|_X^2, \tag{1.4}$$

where $\tilde{u}$ solves the minimization problem (1.3) for a given f_k, X corresponds to $\mathbb{R}$ in the case of scalar parameters or to, e.g., $L^2(\Omega)$ in the case of distributed functions, and $0 < \beta \ll 1$ is a given weight.

The reasonability of assuming to have a such a training set is motivated by certain applications, where the accuracy and as such the noise level in the measurements can be tuned to a certain extent. In MRI or PET, for example, the accuracy of the measurements depends on the setup of the experiment, e.g., the acquisition time. Hence, such a training set can be provided by a series of measurements using phantoms. Then, the u_k's are measured with the maximal accuracy practically possible and the f_k's are measured within a usual clinical setup. For instance, dictionary based image reconstruction methods are already used in the medical imaging community. There, good quality measurements or template shapes are used as priors for reconstructing $\tilde{u}$, cf. e.g. [36], or for image segmentation, cf. e.g. [34] and references therein.

Up to our knowledge this paper is the first one to approach the estimation of the noise distribution as an optimal control problem. By incorporating more than one ϕ_i into the model (1.3) our approach automatically chooses the correct one(s) through an optimal choice of the weights λ_i in terms of (1.4).

Organisation of the paper:. We continue with the analysis of the optimization problem (2.4)–(2.4b) in Section 2. After proving existence of an optimal solution and convergence of the Huber-regularized minimisers to a minimiser of the original total variation problem, the optimization problem is transferred to a Hilbert space setting where the rest of our analysis takes place in Section 3. This further smoothing of the regularizer turns out to be necessary in order to prove continuity of the solution map in a strong enough topology and to verify convergence of our procedure. Moreover, differentiability of the regularized solution operator is thereafter proved, which leads to a first order optimality system characterization of the regularized minimisers. The paper ends with three detailed numerical experiments where the suitability of our approach is computationally verified.

2. Optimization problem in $BV(\Omega)$. We are interested in the solution of the following bilevel optimization problem

$$\min_{\lambda_i \geq 0, \ i=1,\ldots,d} g(\tilde{u}) + \beta \sum_{i=1}^{d} \|\lambda_i\|_X^2 \qquad (2.1a)$$

subject to

$$\tilde{u} \quad = \quad \operatorname{argmin}_{u \in BV \cap \mathcal{A}} \left\{ \mathcal{J}(u) = |Du|(\Omega) + \sum_{i=1}^{d} \int_{\Omega} \lambda_i \phi_i(u, f) \, dx \right\}, \quad (2.1b)$$

where the space X corresponds to $\mathbb{R}$ in the case of scalar parameters or to a function space such that $X \hookrightarrow L^2(\Omega)$ (where $\hookrightarrow$ stands for continuous injection) in the case of distributed functions, $g : L^2(\Omega) \mapsto \mathbb{R}$ is a C^1 functional to be minimised and $\beta > 0$. The admissible set of functions $\mathcal{A}$ is chosen according to the data fidelities ϕ_i. In particular, $BV(\Omega) \cap \mathcal{A}$ restricts the set of BV functions on Ω to those for which the ϕ_i's are well defined, cf. examples below. Moreover, we assume that the functions ϕ_i are differentiable and convex in u, are bounded from below, and fulfil the following coercivity assumption

$$\int_{\Omega} \phi_i(u, f) \, dx \geq C_1 \|u\|_{L^p}^p - C_2, \quad \forall u \in L^p(\Omega) \cap \mathcal{A} \qquad (2.2)$$

for nonnegative constants C_1, C_2 and at least one $p = 1$ or $p = 2$. Examples of ϕ_i's that fulfill these assumptions and that are considered in the paper are

- The Gaussian noise model, where $\int_\Omega \phi(u, f) \, dx = \|u - f\|^2_{L^2(\Omega)}$ fulfills the coercivity constraint for $p = 2$ and the admissible set $\mathcal{A} = L^2(\Omega)$.
- The Poisson noise model, where $\phi(u, f) = u - f \log u$ and $\mathcal{A} = \{u \in L^1(\Omega) \mid u \geq 0\}$. This ϕ is convex and differentiable and fulfils the coercivity condition for $p = 1$. More precisely, we have for $u \geq 0$

$$\int_\Omega (u - f \log u) \, dx \geq \|u\|_{L^1(\Omega)} - \|f\|_{L^\infty(\Omega)} \cdot \log \|u\|_{L^1(\Omega)},$$

  where we have used Jensen's inequality, i.e., for $u \geq 0$

$$\log \left(\int_\Omega u \, dx \right) \geq \int_\Omega \log u \, dx.$$

- The impulse noise model, where $\int_\Omega \phi(u, f) \, dx = \|u - f\|_{L^1(\Omega)}$ fulfills the coercivity constraint for $p = 1$.

For the numerical solution of (1.3) we want to use derivative-based iterative methods. To do so, the gradient of the total variation denoising model has to be uniquely defined. That is, a minimiser of (1.3) is uniquely characterised by the solution of the corresponding Euler-Lagrange equation. Since the total variation regulariser is not differentiable but its "derivative" can be only characterised by a set of subgradients (the subdifferential), we (from now on) shall use a regularised version of the total variation. More precisely, we consider for $\gamma \gg 1$ the Huber-type regularisation of the total variation with

$$|\nabla u|_\gamma = \begin{cases} |\nabla u| - \frac{1}{2\gamma} & \text{if } |\nabla u| \geq \frac{1}{\gamma} \\ |\nabla u|^2 \frac{\gamma}{2} & \text{if } |\nabla u| < \frac{1}{\gamma} \end{cases} \tag{2.3}$$

and the following regularised version of (2.1)-(2.1b)

$$\min_{\lambda_i \geq 0, \; i=1,\ldots,d} g(\tilde{u}) + \beta \sum_{i=1}^d \|\lambda_i\|_X^2 \tag{2.4a}$$

subject to

$$\tilde{u} = \text{argmin}_{u \in W^{1,1} \cap \mathcal{A}} \left\{ \mathcal{J}^\gamma(u) = \int_\Omega |\nabla u|_\gamma \, dx + \sum_{i=1}^d \int_\Omega \lambda_i \phi_i(u, f) \, dx \right\}, \tag{2.4b}$$

where the space X, g, ϕ_i's and $\beta > 0$ are defined as before. The admissible set of functions $\mathcal{A}$ is assumed to be convex and closed subset of $W^{1,1}(\Omega)$ and is chosen according to the data fidelities ϕ_i, cf. examples above. The existence of an optimal solution for (2.4b) is proven by the method of relaxation. To do so we extend the definition of $\mathcal{J}^\gamma$ to $BV(\Omega)$ as

$$\mathcal{J}^\gamma_{ext}(u) = \begin{cases} \mathcal{J}^\gamma(u) & u \in W^{1,1}(\Omega) \cap \mathcal{A} \\ +\infty & u \in BV(\Omega) \setminus (W^{1,1} \cap \mathcal{A}) \end{cases}$$

and prove the existence of a minimiser for the lower-semicontinuous envelope of $\mathcal{J}^\gamma_{ext}$ as follows. We have the following existence result.

THEOREM 2.1. *Let $f \in L^2(\Omega)$ and $\lambda_i \geq 0$ fixed. Then there exists a unique solution $u \in BV(\Omega) \cap \mathcal{A}$ of the minimisation problem*

$$\min_{u \in BV(\Omega) \cap \mathcal{A}} \mathcal{J}_{relax}^{\gamma}(u),$$

where

$$\mathcal{J}_{relax}^{\gamma}(u) = \int_{\Omega} |\nabla u|_{\gamma} \; dx + C \int_{\Omega} |D_s u| + \sum_{i=1}^{d} \int_{\Omega} \lambda_i \phi_i(u, f) \; dx. \qquad (2.5)$$

is the relaxed functional of $\mathcal{J}_{ext}^{\gamma}$ on $BV - w*$.

REMARK 2.1. *Note that*

$$\mathcal{J}_{relax}^{\gamma}(u) \leq \mathcal{J}_{ext}^{\gamma}(u), \quad u \in BV(\Omega)$$

and $\mathcal{J}_{relax}^{\gamma}(u) = \mathcal{J}_{ext}^{\gamma}(u)$ for $u \in W^{1,1}(\Omega) \cap \mathcal{A}$. Moreover, the relaxation result from Theorem 2.1 means that

$$\mathcal{J}_{relax}^{\gamma}(u) = \inf \left\{ \liminf_{n} \mathcal{J}_{ext}^{\gamma}(u_n) : u_n \in BV(\Omega), \, u_n \to u \text{ in } BV - w* \right\},$$

i.e. $\mathcal{J}_{relax}^{\gamma}$ is the greatest $BV - w*$ lower semicontinuous functional less than or equal to $\mathcal{J}_{ext}^{\gamma}$.

Proof. Let u_n be a minimising sequence for $\mathcal{J}_{relax}^{\gamma}$. We start by stating the fact that $|\cdot|_{\gamma}$ is coercive and at most linear. That is

$$\text{For } |x| \geq \frac{1}{\gamma} : \; A|x| - B \leq |x|_{\gamma} = |x| - \frac{1}{2\gamma} \leq |x| + 1$$

$$\text{For } |x| < \frac{1}{\gamma} < 1 : \; A|x| - B \leq = |x|_{\gamma} = |x|^2 \frac{\gamma}{2} < |x| \frac{\gamma}{2}.$$

Hence,

$$|Du_n|(\Omega) = \int_{\Omega} |\nabla u_n| \; dx + \int_{\Omega} |D_s u| \leq \int_{\Omega} |\nabla u|_{\gamma} \; dx + C \int_{\Omega} |D_s u| \leq M, \quad \forall n \geq 1.$$

Moreover, u_n is uniformly bounded in $L^p(\Omega)$ for $p = 1$ or $p = 2$ because of the coercivity assumption (2.2) on ϕ_i and therefore u_n is uniformly bounded in $BV(\Omega)$. Because $BV(\Omega)$ can be compactly embedded in $L^1(\Omega)$ this gives that u_n converges weak $*$ to a function u in $BV(\Omega)$ and (by passing to another subsequence) strongly converges in $L^1(\Omega)$. From the convergence in $L^1(\Omega)$, Ω bounded, we get that u_n (up to a subsequence) converges pointwise a.e. in Ω. Moreover, since ϕ_i is continuous, also $\phi_i(u_n, f)$ converges pointwise to $\phi_i(u, f)$. Then, lower-semicontinuity of $R(|Du|) = \int_{\Omega} |\nabla u|_{\gamma} \; dx + C \int_{\Omega} |D_s u|$ w.r.t. strong convergence in L^1 [19] and Fatou's lemma together with pointwise convergence applied to $\int_{\Omega} \phi_i(u_n, f) \; dx$ gives that

$$\mathcal{J}_{relax}^{\gamma}(u) = \int_{\Omega} |\nabla u|_{\gamma} \; dx + C \int_{\Omega} |D_s u| + \sum_{i=1}^{d} \int_{\Omega} \lambda_i \phi_i(u, f) \; dx$$

$$\leq \liminf_{n} \mathcal{J}_{relax}^{\gamma}(u_n) = \liminf_{n} \left(\int_{\Omega} |\nabla u_n|_{\gamma} \; dx + C \int_{\Omega} |D_s u_n| + \sum_{i=1}^{d} \int_{\Omega} \lambda_i \phi_i(u_n, f) \; dx \right).$$

To see that the minimiser lies in the admissible set $\mathcal{A}$ it is enough to observe that the set $\mathcal{A}$ is a convex and closed subset of $BV(\Omega)$ and hence it is weakly closed by Mazur's Theorem. This gives that $u \in \mathcal{A}$. To see that in fact $\mathcal{J}_{relax}^{\gamma}$ is the greatest lower-semicontinuous envelope of $\mathcal{J}_{ext}^{\gamma}$ see [19, 7, 8, 9, 6]. $\square$

THEOREM 2.2. *There exists an optimal solution to*

$$\min_{\lambda_i \geq 0,\ i=1,\dots,d} g(\tilde{u}) + \beta \sum_{i=1}^{d} \|\lambda_i\|_X^2 \qquad (2.6a)$$

subject to

$$\tilde{u} = \operatorname{argmin}_{u \in BV(\Omega) \cap \mathcal{A}} \mathcal{J}_{relax}^{\gamma}(u). \qquad (2.6b)$$

Proof. Since the cost functional is bounded from below, there exists a minimizing sequence $\{\lambda_n\} = \{\lambda_n(u_n)\} \subset X^d$. Due to the Tikhonov term in the cost functional, we get that $\{\lambda_n\}$ is bounded in X^d. Let u_n be a minimiser of $\mathcal{J}_{relax}^{\gamma}$ for a corresponding λ_n. Such a minimiser exists because of Theorem 2.1. Hence,

$$\mathcal{J}_{relax}^{\gamma}(u_n) \leq \mathcal{J}_{relax}^{\gamma}(0)$$

$$\int_\Omega |\nabla u_n|_\gamma \, dx + C \int_\Omega |D_s u_n| + \sum_{i=1}^{d} \int_\Omega (\lambda_i)_n \, \phi_i(u_n, f) \, dx \leq \sum_{i=1}^{d} \int_\Omega (\lambda_i)_n \, \phi_i(0, f) \, dx$$

As before, from the coercivity condition on f and the uniform bound on λ_n, we deduce that

$$C|Du_n|(\Omega) \leq C|Du_n|(\Omega) + \sum_{i=1}^{d} \int_\Omega (\lambda_i)_n \, \phi_i(u_n, f) \, dx \leq \sum_{i=1}^{d} \int_\Omega (\lambda_i)_n \, \phi_i(0, f) \, dx$$

$$\leq \frac{1}{2} \left(\sum_{i=1}^{d} \|(\lambda_i)_n\|_X + \|\phi_i(0, f)\|_{L^2}^2 \right)$$

$$\leq C.$$

Moreover, from the coercivity of ϕ_i in u_n we get with a similar calculation that u_n is uniformly bounded in L^p for $p = 1$ or 2, and hence in particular in L^1. In sum, u_n is uniformly bounded in $BV(\Omega)$ and hence, converges weakly $*$ in $BV(\Omega)$ and strongly in $L^1(\Omega)$. The latter also gives pointwise convergence of u_n and consequently $\phi_i(u_n, f)$ a.e. and hence we have

$$\mathcal{J}_{relax}^{\gamma}(\hat{u}, \hat{\lambda})$$

$$:= \int_\Omega |\nabla \hat{u}|_\gamma \, dx + C \int_\Omega |D_s \hat{u}| + \sum_{i=1}^{d} \int_\Omega \hat{\lambda}_i \phi_i(\hat{u}, f) \, dx$$

$$\leq \liminf_n \left(\int_\Omega |\nabla u_n|_\gamma \, dx + C \int_\Omega |D_s u_n| + \sum_{i=1}^{d} \int_\Omega (\lambda_i)_n \, \phi_i(u_n, f) \, dx \right)$$

$$= \liminf_n \mathcal{J}_{relax}^{\gamma}(u_n, \lambda_n).$$

Since the cost functional is w.l.s.c., it follows, together with the fact that $\{\lambda : \lambda \geq 0\}$ is weakly closed, that $(\hat{\lambda}, \hat{u}) \in (X \cap \{\lambda_i \geq 0\} \times (BV(\Omega) \cap \mathcal{A})$ is optimal for (2.4). $\square$

THEOREM 2.3. *The sequence of functionals $\mathcal{J}_{relax}^{\gamma}$ in (2.5) converges in the Γ- sense to the functional*

$$\mathcal{J}_{relax}(u) = \begin{cases} \int_\Omega |\nabla u| + \sum_{i=1}^d \int_\Omega \lambda_i \phi_i(u) \ dx & u \in W^{1,1}(\Omega) \cap \mathcal{A} \\ +\infty & u \in BV(\Omega) \setminus (W^{1,1}(\Omega) \cap \mathcal{A}) \end{cases}$$

as $\gamma \to \infty$. Therefore, the unique minimiser of $\mathcal{J}_{relax}^{\gamma}$ converges to the unique minimiser of $\mathcal{J}_{relax}$ as γ goes to infinity.

Proof. The proof is a standard result that follows from the fact that a decreasing point wise converging sequence of functionals Γ- converges to the lower semicontinuous envelope of the point wise limit [16, Proposition 5.7]. In fact, $\int_\Omega |\nabla u|_\gamma + \frac{1}{2\gamma}$ decreases in γ and converges pointwise to $\int_\Omega |\nabla u|$. Then, for $u \in BV(\Omega)$ the functional $\mathcal{J}_{relax}^{\gamma}$ (being the lower-semicontinuous envelope of $\mathcal{J}_{ext}^{\gamma}$) Γ- converges to the functional $\mathcal{J}_{relax}$ in Theorem 2.3. The latter is the lower-semicontinous envelope of the functional in (1.3). $\square$

Although Theorem 2.3 provides a convergence result for the regularized TV subproblems, it is not sufficient to conclude convergence of the optimal regularized weights. For this we need the continuity of the solution map $\lambda \to u(\lambda)$. Up to our knowledge, no sufficient continuity results for the control-to-state map in the case of a total variation minimiser as the state are known. There are various contributions in this directions [14, 31, 38, 39] which are – as they stand – not strong enough to prove the desired result in our case. Indeed, this is a matter of future research.

3. Optimization problem in $H_0^1(\Omega)$. In order to obtain continuity of the solution map and, hence, convergence of the regularized optimal parameters, we proceed in an alternative way and move, from now on, to a Hilbert space setting. Specifically, we replace the minimisation problem (1.3) by the following elliptic-regularized version of it:

$$\min_u \left(\frac{\varepsilon}{2} \|Du\|_{L^2}^2 + |Du|(\Omega) + \sum_{i=1}^d \int_\Omega \lambda_i \ \phi_i(u) \ dx \right). \tag{3.1}$$

where $0 < \varepsilon \ll 1$ is an artificial diffusion parameter.

A necessary and sufficient optimality condition for (3.1) is given by the following elliptic variational inequality:

$$\varepsilon(Du, D(v-u))_{L^2} + \sum_{i=1}^d \int_\Omega \lambda_i \phi_i'(u)(v-u) \ dx$$
$$+ \int_\Omega |Dv| \ dx - \int_\Omega |Du| \ dx \geq 0 \ \text{ for all } v \in H_0^1(\Omega). \tag{3.2}$$

Note that by adding the coercive term, we implicitly impose the solution space $H_0^1(\Omega)$ (see [26]).

7

Our aim is to determine the optimal choice of parameters $\lambda_i,\ i = 1, ..., d$, by solving the following optimization problem:

$$\min_{\lambda_i \geq 0,\ i=1,...,d} g(u) + \beta \sum_{i=1}^{d} \|\lambda_i\|_X^2 \tag{3.3a}$$

subject to

$$\varepsilon(Du, D(v - u))_{L^2} + \sum_{i=1}^{d} \int_{\Omega} \lambda_i\ \phi_i'(u)(v - u)\ dx$$

$$+ \int_{\Omega} |Dv|\ dx - \int_{\Omega} |Du|\ dx \geq 0 \ \text{ for all } v \in H_0^1(\Omega), \tag{3.3b}$$

where the space X corresponds to $\mathbb{R}$ in the case of scalar parameters or to a Hilbert function space in the case of distributed functions. Problem 3.3 corresponds to an optimization problem governed by a variational inequality of the second kind (see [17] and the references therein).

Next, we perform the analysis of the optimization problem (3.3). After proving existence of an optimal solution, a regularization approach will be also proposed in this context. We will prove the continuity of the control-to-state map and, based on it, convergence of the regularized images and the optimal regularized parameters. In the case of a smoother regularization of the TV term, also differentiablity of the solution operator will be verified, which will lead us afterwards to a first order optimality system characterizing the optimal solution to (3.3).

We start with the following existence theorem.

THEOREM 3.1. *There exists an optimal solution for problem* (3.3).

Proof. Let $\{\lambda_n\} \subset X^d$ be a minimizing sequence. Due to the Tikhonov term in the cost functional, we get that $\{\lambda_n\}$ is bounded in X^d. From (3.3b) we additionally get that the sequence of images $\{u_n\}$ satisfy

$$\varepsilon\|u_n\|_{H_0^1}^2 + \sum_{i=1}^{d} \int_{\Omega} \lambda_{ni}\ [\phi_i'(u_n) - \phi_i'(0)]\, u_n\ dx$$

$$+ \int_{\Omega} |Du_n|\ dx \leq - \sum_{i=1}^{d} \int_{\Omega} \lambda_{ni}\ \phi_i'(0)u_n\ dx, \tag{3.4}$$

which, due to the monotonicity of the operators on the left hand side, implies that

$$\varepsilon\|u_n\|_{H_0^1}^2 \leq \sum_{i=1}^{d} \|\lambda_{ni}\|_X \|\phi_i'(0)\|_{L^r} \|u_n\|_{L^p}, \tag{3.5}$$

for $2 < p < +\infty$ and $r = \frac{2p}{p-2}$. Thanks to the embedding $H_0^1(\Omega) \hookrightarrow L^p(\Omega)$, for all $1 \leq p < +\infty$, we get that

$$\|u_n\|_{H_0^1} \leq C \sum_{i=1}^{d} \|\lambda_{ni}\|_X \|\phi_i'(0)\|_{L^r}, \tag{3.6}$$

8

for some constant $C > 0$. Consequently,

$$\{u_n\} \text{ is uniformly bounded in } H_0^1(\Omega). \tag{3.7}$$

Therefore, there exists a subsequence $\{(u_n, \lambda_n)\}$ which converges weakly in $H_0^1(\Omega) \times X^d$ to a limit point $(\hat{u}, \hat{\lambda})$. Moreover, $u_n \to \hat{u}$ strongly in $L^p(\Omega)$ and, thanks to the continuity of ϕ', also

$$\phi'(u_n)u_n \rightharpoonup \phi'(\hat{u})\hat{u} \qquad \text{strongly in } L^{\frac{p}{2}}(\Omega) \tag{3.8}$$

Consequently, thanks to the continuity of ϕ' and the properties of $a(\cdot, \cdot)$, we get

$$a(\hat{u}, \hat{u}) + \sum_{i=1}^{d} \int_{\Omega} \hat{\lambda}_i \phi_i'(\hat{u})\hat{u} + \int_{\Omega} |D\hat{u}|$$

$$\leq \liminf a(u_n, u_n) + \sum_{i=1}^{d} \int_{\Omega} \lambda_{ni} \phi_i'(u_n)u_n + \int_{\Omega} |Du_n|$$

$$\leq \liminf a(u_n, v) + \sum_{i=1}^{d} \int_{\Omega} \lambda_{ni} \phi_i'(u_n)v + \int_{\Omega} |Dv|$$

$$= a(\hat{u}, v) + \sum_{i=1}^{d} \int_{\Omega} \hat{\lambda}_i \phi_i'(\hat{u})v + \int_{\Omega} |Dv|.$$

Since the cost functional is w.l.s.c., it follows, together with the fact that $\{\lambda : \lambda \geq 0\}$ is weakly closed, that $(\hat{\lambda}, \hat{u})$ is optimal for (3.3). $\square$

Next, we consider the following family of regularized problems:

$$\epsilon(Du_\gamma, Dv) + (h_\gamma(Du_\gamma), Dv) + \sum_{i=1}^{d} \int_{\Omega} \lambda_i \ \phi_i'(u_\gamma)v = 0, \forall v \in H_0^1(\Omega), \tag{3.9}$$

where $h_\gamma(Du_\gamma)$ corresponds to an active-inactive-set approximation of the subdifferential of $|Du_\gamma|$, i.e., h_γ coincides with an element of the subdifferential up to a small neighborhood of 0. The most natural choice is the function

$$h_\gamma(z) = \frac{\gamma z}{\max(1, \gamma|z|)}, \tag{3.10}$$

which corresponds to the derivative of the Huber function defined in (2.3). An alternative regularization is given by the C^1 function

$$h_\gamma(z) = \begin{cases} g\frac{z}{|z|} & \text{if } \gamma|z| \geq g + \frac{1}{2\gamma} \\ \frac{z}{|z|}(g - \frac{\gamma}{2}(g - \gamma|z| + \frac{1}{2\gamma})^2) & \text{if } g - \frac{1}{2\gamma} \leq \gamma|z| \leq g + \frac{1}{2\gamma} \\ \gamma z & \text{if } \gamma|z| \leq g - \frac{1}{2\gamma}, \end{cases} \tag{3.11}$$

where g is a positive parameter. The latter smoothing for the total variation is going to be used in Proposition 3.4 where differentiability of the regulariser is needed.

REMARK 3.1. *For a fixed λ, it can be verified that (3.9) has a unique solution. Moreover, the sequence of regularized solutions $\{u_\gamma\}$ converges strongly in $H_0^1(\Omega)$ to the solution of (3.2) (cf. [17]).*

9

Based on the regularized problems (3.9), we now focus on the following optimization problem:

$$\min_{\lambda_i \geq 0, \ i=1,\ldots,d} g(u) + \beta \sum_{i=1}^{d} \|\lambda_i\|_X^2 \qquad (3.12a)$$

subject to

$$\epsilon(Du_\gamma, Dv) + (h_\gamma(Du_\gamma), Dv) + \sum_{i=1}^{d} \int_\Omega \lambda_i \ \phi_i'(u_\gamma)v = 0, \forall v \in H_0^1(\Omega), \qquad (3.12b)$$

where $0 < \epsilon \ll 1$. In this setting we can prove the following continuity and convergence results.

PROPOSITION 3.2. *Let $\{\lambda_n\}$ be a sequence in X^d such that $\lambda_n \rightharpoonup \hat{\lambda}$ weakly in X^d as $n \to \infty$. Further, let $u_n := u_\gamma(\lambda_n)$ denote the solution to (3.12b) associated with λ_n and $\hat{u} := u_\gamma(\hat{\lambda})$. Then*

$$u_n \to \hat{u} \ strongly \ in \ H_0^1(\Omega).$$

Proof. Since $\lambda_n \rightharpoonup \hat{\lambda}$ weakly in X, it follows, by the principle of uniform boundedness, that $\{\lambda_n\}$ is bounded in X. From (3.12b) we additionally get that

$$\varepsilon\|u_n\|_{H_0^1}^2 + \sum_{i=1}^{d} \int_\Omega \lambda_{ni} \ [\phi_i'(u_n) - \phi_i'(0)] \, u_n \, dx$$

$$+ (h_\gamma(Du_n), Du_n) \leq -\sum_{i=1}^{d} \int_\Omega \lambda_{ni} \ \phi_i'(0)u_n \, dx, \quad (3.13)$$

which, proceeding as in the proof of Theorem 3.1, implies that

$$\|u_n\|_{H_0^1} \leq C \sum_{i=1}^{d} \|\lambda_{ni}\|_X \|\phi_i'(0)\|_{L^r}, \qquad (3.14)$$

for some constant $C > 0$. Hence, $\{u_n\}$ is uniformly bounded in $H_0^1(\Omega)$.

Consequently, there exists a subsequence (denoted the same) and a limit $\hat{u}$ such that

$$u_n \rightharpoonup \hat{u} \ \text{in } H_0^1(\Omega) \qquad \text{and} \qquad u_n \to \hat{u} \ \text{in } L^p(\Omega), \ 1 \leq p < +\infty.$$

Thanks to the structure of the regularized VI (3.12b) it also follows (as in the proof of Theorem 3.1) that $\hat{u}$ is solution of the regularized VI associated with $\hat{\lambda}$. Since the solution to (3.12b) is unique, it additionally follows that the whole sequence $\{u_n\}$ converges weakly towards $\hat{u}$.

To verify strong convergence, we take the difference of the variational equations satisfied by u_n and $\hat{u}$ and obtain that

$$\varepsilon(Du_n - D\hat{u}, Dv) + (h_\gamma(Du_n) - h_\gamma(D\hat{u}), Dv)$$

$$= \int_\Omega \left[\hat{\lambda} \ \phi'(\hat{u}) - \lambda_n \ \phi'(u_n) \right] v \, dx, \forall v \in H_0^1(\Omega).$$

10

Adding the term $-\lambda_n \ \phi'(\hat{u})$ on both sides of the latter yields

$$\varepsilon(Du_n - D\hat{u}, Dv) + (h_\gamma(Du_n) - h_\gamma(D\hat{u}), Dv)$$
$$\int_\Omega [\lambda_n \ \phi'(u_n) - \lambda_n \ \phi'(\hat{u})] \ v \ dx = \int_\Omega \left[\hat{\lambda} \ \phi'(\hat{u}) - \lambda_n \ \phi'(\hat{u})\right] v \ dx, \forall v \in H_0^1(\Omega).$$

Choosing $v = u_n - \hat{u}$ and thanks to the monotonicity of the operator on the left hand side, we then obtain that

$$\varepsilon\|Du_n - D\hat{u}\|^2 \leq \left|\int_\Omega \left[\hat{\lambda} \ \phi'(\hat{u}) - \lambda_n \ \phi'(\hat{u})\right] (u_n - \hat{u}) \ dx\right|,$$

which thanks to the strong convergence $u_n \to \hat{u}$ in $L^p(\Omega)$, $1 \leq p < +\infty$, and the regularity $\phi'(\hat{u}) \in L^{\frac{p}{2}}(\Omega)$, implies the result. □

THEOREM 3.3. *There exists an optimal solution for each regularized problem (3.12). Moreover, the sequence $\{\lambda_\gamma\}$ of regularized optimal parameters is bounded in X^d and every weakly convergent subsequence converges towards an optimal solution of (3.3).*

Proof. Let $\{\lambda_n\}$ be a minimizing sequence. From the structure of the cost functional and the properties of (3.12b) it follows that the sequence is bounded. Consequently, there exists a subsequence (denoted the same) and a limit λ^* such that $\lambda_n \rightharpoonup \lambda^*$ weakly in X^d. From Proposition 3.2 and the weakly lower semicontinuity of the cost functional, optimality of λ^* follows, and, therefore, existence of an optimal solution.

Let now $\{\lambda_\gamma\}_{\gamma>0}$ be a sequence of optimal solutions to (3.12). Since $(0,0) \in H_0^1(\Omega) \times X^d$ is feasible for each $\gamma > 0$, it follows that

$$J(u_\gamma(\lambda_\gamma), \lambda_\gamma) \leq J(u_\gamma(0), 0) = J(0, 0).$$

Thanks to the Tikhonov term in the cost functional it then follows that $\{\lambda_\gamma\}_{\gamma>0}$ is bounded.

Let $\hat{\lambda}$ be the limit point of a weakly convergent subsequence (also denoted by $\{\lambda_\gamma\}$). From Remark 3.1 and Proposition 3.2 it follows, by using the triangle inequality, that

$$u_\gamma(\lambda_\gamma) \to \hat{u} \quad \text{strongly in } H_0^1(\Omega) \quad \text{as } \gamma \to \infty,$$

where $\hat{u}$ denotes the solution to (3.2) associated with $\hat{\lambda}$.

From the weakly lower semicontinuity of the cost functional, we finally get that

$$J(\hat{u}, \hat{\lambda}) \leq \liminf_{\gamma \to \infty} J(u_\gamma(\lambda_\gamma), \lambda_\gamma) \leq \liminf_{\gamma \to \infty} J(u_\gamma(\bar{\lambda}), \bar{\lambda}) = J(\bar{u}, \bar{\lambda}),$$

where $\bar{\lambda}$ is an optimal solution to (3.12). □

The next proposition is concerned with the differentiability of the solution operator. This result will lead us thereafter (see Theorem 3.5) to get an expression for the gradient of the cost functional and also to obtain an optimality system for the characterization of the optimal solutions to (3.3).

PROPOSITION 3.4. *Let $G_\gamma : X^d \mapsto H_0^1(\Omega)$ be the solution operator, which assigns to each parameter λ the corresponding solution to the regularized VI (3.9), with the function h_γ given by (3.11). Then the operator G_γ is Gâteaux differentiable and its*

11

derivative at $\bar\lambda$, in direction ξ, is given by the unique solution $z \in H_0^1(\Omega)$ of the following linearized equation:

$$\epsilon(Dz, Dv) + (h'_\gamma(D\bar u)Dz, Dv) + \sum_{i=1}^{d} \int_\Omega \lambda_i \ \phi''_i(\bar u) \ z \ v \ dx$$

$$+ \sum_{i=1}^{d} \int_\Omega \xi_i \ \phi'_i(\bar u) \ v \ dx = 0, \quad \text{for all } v \in H_0^1(\Omega). \quad (3.15)$$

Proof. Existence and uniqueness of a solution to (3.15) follows from Lax-Milgram theorem by making use of the monotonicity properties of h_γ and ϕ'_i.

Let $\xi \in X^d$, and let y_t and y be the unique solutions to (3.9) correspondent to $\lambda + t\xi$ and λ, respectively. By taking the difference between both equations, it follows that

$$\epsilon(D(u_t - u), D(u_t - u)) + (h_\gamma(Du_t) - h_\gamma(Du), D(u_t - u))$$
$$+ \lambda_t(\phi'(u_t) - \phi'(u), u_t - u) = -t(\xi\phi'(u), u_t - u), \quad (3.16)$$

which, by the monotonicity of h_γ and ϕ' yields that

$$\kappa\|u_t - u\|_{H_0^1}^2 \leq t \ \|\xi\|_{X^d}\|\phi'(u)\|_{L^r}\|u_t - u\|_{L^p}. \quad (3.17)$$

Therefore, the sequence $\{z_t\}_{t>0}$, with $z_t := \frac{y_t - y}{t}$, is bounded and there exists a subsequence (denoted the same) such that $z_t \rightharpoonup z$ weakly in $H_0^1(\Omega)$.

Using the mean value theorem in integral form we get that

$$a(z_t, w) + \frac{1}{t}\left(h_\gamma(Du_t) - h_\gamma(Du), Dw\right) + \frac{1}{t}\left(\lambda_t(\phi'(u_t) - \phi'(u)), w\right) \quad (3.18)$$

$$= a(z_t, w) + \int_\Omega \langle h'_\gamma(\vartheta_t)Dz_t, Dw\rangle \ dx + \int_\Omega \lambda_t\phi''(\zeta_t)w \ dx \quad (3.19)$$

$$= -(\xi\phi'(u), w), \quad \text{for all } w \in V, \quad (3.20)$$

where $\vartheta_t(x) = Du(x) + \rho_t(x)(Du_t(x) - Du(x))$, with $0 \leq \rho_t(x) \leq 1$, and $\zeta_t(x) = u(x) + \varrho_t(x)(u_t(x) - u(x))$, with $0 \leq \varrho_t(x) \leq 1$.

From the continuity of the bilinear form it follows that $a(z_t, w) \to a(z, w)$, for all $w \in V$. Additionally, from the consistency of the regularization (see Remark 3.1), $u_t \to u$ strongly in $H_0^1(\Omega)$ and, therefore, $\vartheta_t \to Du$ strongly in $\mathbb{L}^2(\Omega)$ and $\zeta_t \to u$ strongly in $H_0^1(\Omega)$.

Introducing the function

$$\chi_\gamma(x) := \begin{cases} g & \text{if } \gamma|x| \geq g + \frac{1}{2\gamma}, \\ g - \frac{\gamma}{2}(g - \gamma|x| + \frac{1}{2\gamma})^2 & \text{if } |\gamma|x| - g| \leq \frac{1}{2\gamma}, \\ \gamma|x| & \text{if } \gamma|x| \leq g - \frac{1}{2\gamma}, \end{cases}$$

the regularizing function may be written as $h_\gamma(x) = \frac{x}{|x|} \ \chi_\gamma(x)$ and the second term

12

in (3.19) can be expressed as

$$\int_\Omega \langle h'_\gamma(\vartheta_t) Dz_t , Dw \rangle \ dx = \int_\Omega \chi_\gamma(\vartheta_t) \left\langle \frac{Dw}{|\vartheta_t|} - \frac{\langle \vartheta_t , Dw \rangle}{|\vartheta_t|^2} \frac{\vartheta_t}{|\vartheta_t|} , Dz_t \right\rangle \ dx$$

$$+ \int_\Omega \chi'_\gamma(\vartheta_t)[Dw] \left\langle \frac{\vartheta_t}{|\vartheta_t|} , Dz_t \right\rangle \ dx.$$

Let $\Phi : \mathbb{R}^2 \mapsto \mathbb{R}^2$ be the operator defined by

$$\Phi(\xi) := \chi_\gamma(\xi) \frac{w}{|\xi|} - \frac{\langle \xi , w \rangle}{|\xi|^2} \frac{\xi}{|\xi|} + \chi'_\gamma(\xi)[w] \frac{\xi}{|\xi|}, \tag{3.21}$$

with $w \in \mathbb{R}^2$. When considered from $\mathbb{L}^q(\Omega)$ to $\mathbb{L}^q(\Omega)$, Φ is a continuous superposition operator. Therefore, since $\vartheta_t \to Du$ strongly in $\mathbb{L}^2(\Omega)$ and thanks to the weak convergence of z_t and continuity of ϕ'', we may pass to the limit in (3.19)-(3.20) and obtain that

$$a(z, w) + \int_\Omega \langle h'_\gamma(Du) Dz , Dw \rangle \ dx$$

$$= - \int_\Omega \lambda \phi''(\zeta) w \ dx - \int_\Omega \xi \phi'(u) \ w \ dx, \text{ for all } w \in H_0^1(\Omega). \tag{3.22}$$

Consequently, $z \in H_0^1(\Omega)$ corresponds to the unique solution of the linearized equation.

Using again the function χ_γ, the operator on the left hand side of equation (3.22) may be written as

$$\int_\Omega Dw^T M(x) Dz \ dx,$$

where

$$M(x) := 2\mu I + \frac{\chi_\gamma(Dy)}{|Dy|} I - \frac{\chi_\gamma(Dy)}{|Dy|^3} \left[Dy(x) Dy(x)^T \right]$$

$$+ \chi_{\mathcal{S}^\gamma} \frac{\gamma^2}{|Dy|^2} (g - \gamma|Dy| + \frac{1}{2\gamma}) \left[Dy(x) Dy(x)^T \right] + \chi_{\mathcal{I}^\gamma} \frac{\gamma}{|Dy|^2} \left[Dy(x) Dy(x)^T \right], \tag{3.23}$$

where $\chi_{\mathcal{S}^\gamma}$ and $\chi_{\mathcal{I}^\gamma}$ correspond to the indicator functions of the sets $\mathcal{S}^\gamma := \{x \in \Omega : |\gamma|Dy| - g| < \frac{1}{2\gamma}\}$ and $\mathcal{I}^\gamma := \{x \in \Omega : \gamma|Dy| \le g - \frac{1}{2\gamma}\}$, respectively.

Similarly, by replacing Dy with ϑ_t in (3.23) a matrix denoted by M_t is obtained. Both matrices M and M_t are symmetric and positive definite. Using Cholesky decomposition we obtain lower triangular matrices L_t and L such that

$$M_t(x) = L_t(x) \ L_t^T(x) \quad \text{and} \quad M(x) = L(x) \ L^T(x).$$

Proceeding as in [10, pp. 30-31] (see also [17, Thm. 6.1]), strong convergence of $z_t \to z$ in $H_0^1(\Omega)$ is obtained, and, thus, also Gâteaux differentiability of G_γ. □

13

THEOREM 3.5 (Optimality system). *Let $(\bar{\lambda}, \bar{u})$ be an optimal solution to problem (3.12) with $X^d = \mathbb{R}^d$. There exist Lagrange multipliers $(p, \mu) \in H_0^1(\Omega) \times \mathbb{R}^d$ such that the following optimality system holds:*

$$\epsilon(D\bar{u}, Dv) + (h_\gamma(D\bar{u}), Dv) + \sum_{i=1}^{d} \int_\Omega \bar{\lambda}_i \; \phi_i'(\bar{u})v \; dx = 0, \forall v \in H_0^1(\Omega), \qquad (3.24a)$$

$$\epsilon(Dp, Dv) + (h_\gamma'(D\bar{u})^* Dp, Dv)$$
$$+ \sum_{i=1}^{d} \int_\Omega \lambda_i \; \phi_i''(\bar{u}) \; p \; v \; dx = -(g'(\bar{u}), v), \forall v \in H_0^1(\Omega), \quad (3.24b)$$

$$\mu_i = 2\beta\bar{\lambda}_i + \int_\Omega p\phi_i'(\bar{u}) \; dx, \qquad i = 1, ..., d, \qquad (3.24c)$$

$$\mu_i \geq 0, \; \lambda_i \geq 0, \; \mu_i\lambda_i = 0, \qquad i = 1, ..., d. \qquad (3.24d)$$

Proof. Consider the reduced cost functional

$$f(\lambda) = g(G_\gamma(\lambda)) + \beta\|\lambda\|_{\mathbb{R}^d}^2. \qquad (3.25)$$

Thanks to the optimality of $\bar{\lambda}$ and the differentiability of both G_γ and g it follows that

$$\nabla f(\bar{\lambda})^T(\xi - \bar{\lambda}) \geq 0, \; \text{for all } \xi \geq 0. \qquad (3.26)$$

Let $p \in H_0^1(\Omega)$ be the unique solution to the adjoint equation:

$$\epsilon(Dp, Dv) + (h_\gamma'(D\bar{u})^* Dp, Dv)$$
$$+ \sum_{i=1}^{d} \int_\Omega \lambda_i \; \phi_i''(\bar{u}) \; p \; v \; dx = -(g'(\bar{u}), v), \forall v \in H_0^1(\Omega). \quad (3.27)$$

Indeed, existence and uniqueness of a solution to (3.27) follows from the Lax-Milgram theorem, similarly as for the linearized equation.

Using the adjoint equation it follows that

$$\nabla f(\bar{\lambda})^T\xi = (g'(\bar{u}), G_\gamma'(\bar{u})\xi) + 2\beta\bar{\lambda}^T\xi$$
$$= 2\beta\bar{\lambda}^T\xi - \epsilon(Dp, Dv) - (h_\gamma'(D\bar{u})^* Dp, Dv) - \sum_{i=1}^{d} \int_\Omega \lambda_i \; \phi_i''(\bar{u}) \; p \; v \; dx,$$

which, utilizing the linearized equation (3.15), yields that

$$\nabla f(\bar{\lambda})^T\xi = 2\beta\bar{\lambda}^T\xi + \sum_{i=1}^{d} \int_\Omega \phi_i'(\bar{u})p \; \xi_i. \qquad (3.28)$$

14

Let $\mu_i := 2\beta\bar{\lambda}_i + \int_\Omega \phi_i'(\bar{u})p, \; i = 1, ..., d$. From (3.26) and (3.28) it then follows that

$$\mu^T(\xi - \bar{\lambda}) \geq 0, \; \text{for all } \xi \geq 0,$$

which is equivalent to the complementarity system (3.24d). $\square$

REMARK 3.2. *In the case of a general parameter Hilbert space X^d, an optimality system constituted by equations* (3.24a), (3.24b) *and the variational inequality*

$$2\beta(\bar{\lambda}, \xi - \bar{\lambda})_{X^d} + (\phi'(\bar{u})p, \xi - \bar{\lambda})_{\mathbb{L}^2} \geq 0, \; \text{for all } \xi \in X^d : \xi \geq 0 \; a.e.$$

is obtained.

4. Numerical solution of the optimization problem. In this section we focus on the numerical solution of the optimization problem (3.3). For the determination of the optimal parameter values we consider a projected BFGS (Broyden-Fletcher-Goldfarb-Shanno) method. In each computational experiment, the state equation is solved by means of a Newton type algorithm (specified in each case), while a forward finite differences quotient is used for the evaluation of the gradient of the cost functional. Due to the high computational cost of solving the state equation, a fixed line search parameter value $\alpha = 0.5$ was considered in combination with the BFGS method. The linear systems in each Newton iteration are solved exactly by using a LU decomposition for band matrices.

4.1. Gaussian noise. As a first example we consider the determination of a single regularization parameter, i.e., $\lambda \in \mathbb{R}$. The optimization problem takes the following form:

$$\min \; \frac{1}{2}\|u - u_o\|_{L^2}^2 + \beta\lambda^2 \tag{4.1a}$$

subject to:

$$\varepsilon(Du, D(v-u))_{L^2} + \int_\Omega \lambda(u - u_n)(v - u) \; dx$$
$$+ \int_\Omega |Dv| \; dx - \int_\Omega |Du| \; dx \geq 0, \forall v \in H_0^1(\Omega), \tag{4.1b}$$

where u_o and u_n denote the original and noisy images respectively. The problem consists therefore in the optimal choice of the TV regularization parameter, if the original image is known in advance. This is a toy example for proof of concept only. In practice this image would be replaced by a training set of images as motivated in the Introduction.

For the numerical solution of the regularized variational inequality we utilize the primal-dual algorithm developed in [28], which was proved to be globally as well as locally superlinear convergent.

The results for the parameter values $\beta = 1 \times 10^{-10}, \varepsilon = 1 \times 10^{-12}, \gamma = 100$ and $h = 1/177$ are shown in Figure 4.1 together with the noisy image distorted by Gaussian noise with zero mean and variance 0.002. The computed optimal parameter for this problem is $\lambda^* = 2980$ For a noise of mean 0 and variance 0.02, and the same regularization parameters as in the previous experiment, the optimal image is obtained with the weight $\lambda^* = 1770.9$. The noisy and denoised images are given in Figure 4.1.
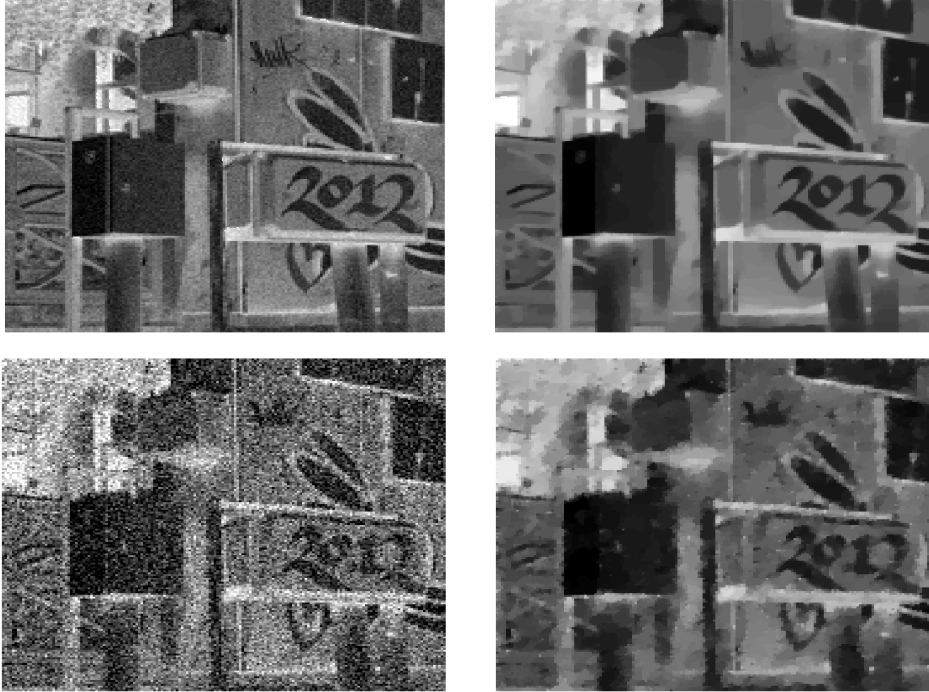
Fig. 4.1. *Noisy (left) and denoised (right) images. Noise variance: 0.002 (first row) and 0.02 (last row).*

#pixels	60	65	70	75	80	85
λ^*	674.6	742.9	788.9	855.0	885.8	933

TABLE 4.1
Optimal weight vs. mesh size; $\mu = 1e - 15$, $\gamma = 100$, $\beta = 1e - 10$.

By increasing the variance in the noise, the optimal values of the weight differ significantly. This is intuitively clear, since as the image becomes noisier there is less original information that can be directly obtained. When that happens, the TV regularization plays an increasingly important role.

The question of robustness of the optimal parameter value deserves also to be tested. In Table 4.1 we compute the optimal weight for different sources of the noisy image (different total number of pixels). The value of the optimal weight increases together with the size of the image from which the information is obtained. The variation remains however small, implying a robust behavior of the values.

4.2. Magnetic resonance imaging. Gaussian noise images typically arise within the framework of magnetic resonance imaging (MRI). The challenge in this case consists in training the machines so that a clearer image is obtained. The magnetic resonance images seem to be the natural choice for our methodology, since a training set of images is often at hand.

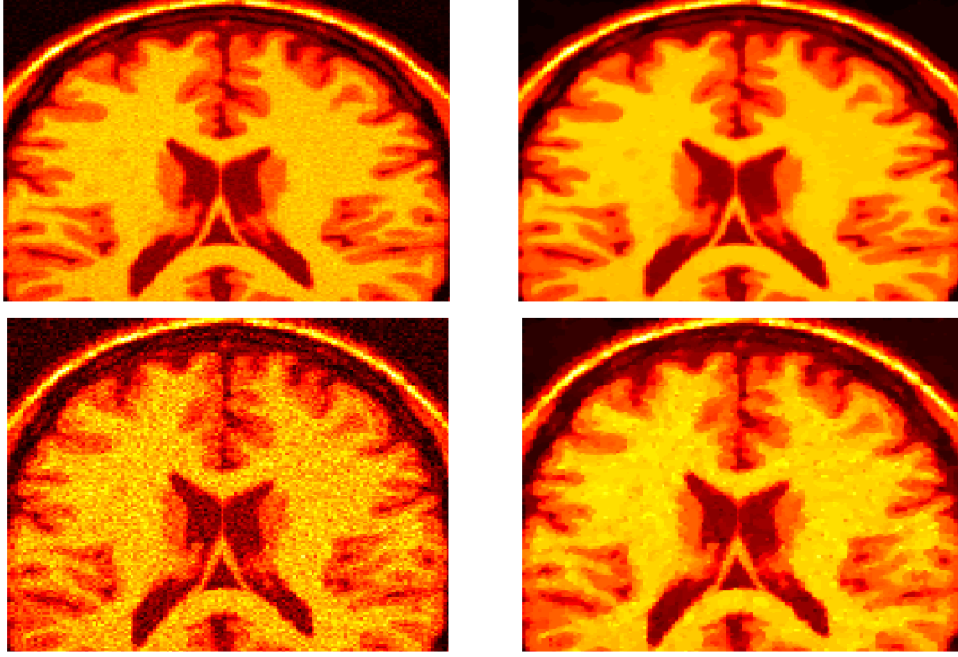For such a training set we consider the solution of problem (4.1). In Figure 4.2

FIG. 4.2. *Image with 3% noise (upper left) and its correspondent optimal denoised one (upper right); noisy image with 9% noise (lower left) and optimal denoised image (lower right)*

the noisy images together with the final optimized ones for a brain scan are shown. For this experiments a mesh step size of $h = 1/250$ was considered. The Tikhonov regularization parameter took the value $\beta = 10^{-10}$, while the Huber regularization parameter was chosen as $\gamma = 100$. With this values, the optimal parameter value for the MRI image with 3% of noise was $\lambda^* = 64.1448$. When the noise in the image was of 9%, the computed optimal weight was $\lambda^* = 26.7110$.

4.3. Training objective- multiple parameters (Gauss+Poisson). The importance of controlling the regularization in a denoising problem becomes clear once two different noise distributions, modeled by fidelity terms weighted by non-negative parameters λ_1 and λ_2, are present in an image. In particular, in the following experiment we shall solve the optimization problem

$$\min_{\lambda \geq 0} \ \frac{1}{2}\|u - u_o\|_{L^2}^2 + \beta \sum_{i=1}^{2} \|\lambda_i\|^2$$

subject to:

$$\min_{u \geq 0} \left\{ \frac{\varepsilon}{2}\|Du\|_{L^2}^2 + |Du|(\Omega) + \frac{\lambda_1}{2}\|u - u_n\|_{L^2}^2 + \lambda_2 \int_{\Omega} (u - u_n \log u) \ dx \right\}. \qquad (4.2)$$

17

For the characterization of a minimizer of (4.2) we (formally) get the following Euler-Lagrange equation

$$-\epsilon \Delta u - \mathrm{div}\left(\frac{\gamma \nabla u}{\max(\gamma|\nabla u|, 1)}\right) + \lambda_1(u - u_n) + \lambda_2(1 - \frac{u_n}{u}) - \alpha = 0$$

$$\alpha \cdot u = 0,$$

with non-negative Lagrange multiplier $\alpha \in L^2(\Omega)$, cf. [27]. As in [33] we multiply the first equation with u and get

$$u \cdot \left(-\epsilon \Delta u - \mathrm{div}\left(\frac{\gamma \nabla u}{\max(\gamma|\nabla u|, 1)}\right) + \lambda_1(u - u_n)\right) + \lambda_2(u - u_n) = 0,$$

where we have used the complementarity condition $\alpha \cdot u = 0$. Next, the solution u is computed iteratively by using a Newton type method.

For an appropriate initial guess u^0, an iteration of the semismooth Newton method consists in solving the system

$$\delta_u\left(-\epsilon \Delta u - \mathrm{div}q + \lambda_1(u - u_n)\right) + u\left(-\epsilon \Delta \delta_u - \mathrm{div}\delta_q + \lambda_1 \delta_u\right) \tag{4.3}$$
$$+\lambda_2 \delta_u = -u\left(-\epsilon \Delta u - \mathrm{div}q + \lambda_1(u - u_n)\right) - \lambda_2(u - u_n),$$

$$\delta_q - \frac{\gamma \nabla \delta_u}{\max(1, \gamma|\nabla u|)} + \chi_{\mathcal{A}_\gamma} \gamma^2 \frac{\nabla u^T \nabla \delta_u}{\max(1, \gamma|\nabla u|)^2} \frac{\nabla u}{|\nabla u|} = -q + \frac{\gamma \nabla u}{\max(1, \gamma|\nabla u|)}, \tag{4.4}$$

for the increments δ_u and δ_q. In equation (4.4), $\chi_{\mathcal{A}_\gamma}$ stands for the indicator function of the active set $\mathcal{A}_\gamma := \{x \in \Omega : \gamma|\nabla u(x)| \geq 1\}$.

Similarly to [18], we consider a modification of the iteration based on the properties of the solution to (4.2). Specifically, noting that $q = \frac{\nabla u}{|\nabla u|}$ on the final active set and that $|q| \leq 1$, we replace the term $\frac{\nabla u}{|\nabla u|}$ by $\frac{q}{\max(1, \gamma|\nabla u|)}$ on the left hand side of the iteration system. The resulting iteration is then given by:

$$\delta_u\left(-\epsilon \Delta u - \mathrm{div}q + \lambda_1(u - u_n)\right) + u\left(-\epsilon \Delta \delta_u - \mathrm{div}\delta_q + \lambda_1 \delta_u\right) \tag{4.5}$$
$$+\lambda_2 \delta_u = -u\left(-\epsilon \Delta u - \mathrm{div}q + \lambda_1(u - u_n)\right) - \lambda_2(u - u_n),$$

$$\delta_q - \frac{\gamma \nabla \delta_u}{\max(1, \gamma|\nabla u|)} + \chi_{\mathcal{A}_\gamma} \gamma^2 \frac{\nabla u^T \nabla \delta_u}{\max(1, \gamma|\nabla u|)^2} \frac{q}{\max(1, |q|)} = -q + \frac{\gamma \nabla u}{\max(1, \gamma|\nabla u|)}, \tag{4.6}$$

The resulting algorithm exhibits global and local superlinear convergence properties. In Figure 4.3 the residuum of the algorithm in the last 4 iterations is depicted. The parameter values used are $\mu = 1e-4, \gamma = 50, \beta = 1e-10, h = 1/60, \lambda_1 = 769.2199, \lambda_2 = 30.6396$. From the behavior of the residdum, local superlinear convergence is inferred. In combination with the outer BFGS iteration, a competitive algorithm for the solution of the bilevel problem is obtained.

For the computational tests we consider the noisy zoomed image of a plane's wing (see Figure 4.4). Choosing the parameter values $\beta = 1e-10$, $\gamma = 100$ and $\epsilon = 1e-15$, the optimal weights $\lambda_1^* = 1847.75$ and $\lambda_2^* = 73.45$ were computed on a grid with mesh size step $h = 1/200$. From Figure 4.3 also a good match between the original and the denoised images can be observed. The noise appears to be succesfully removed with the computed optimal weights.
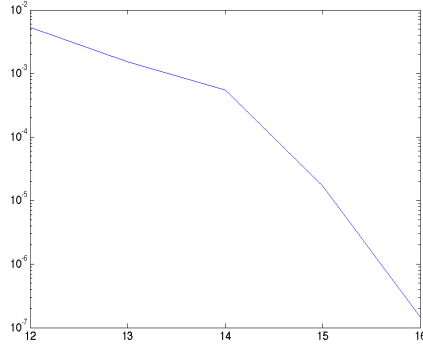
FIG. 4.3. *SSN residuum in the last 4 iterations.* $\mu = 1e-4, \gamma = 50, \beta = 1e-10, h = 1/60, u_1 = 769.2199, u_2 = 30.6396$
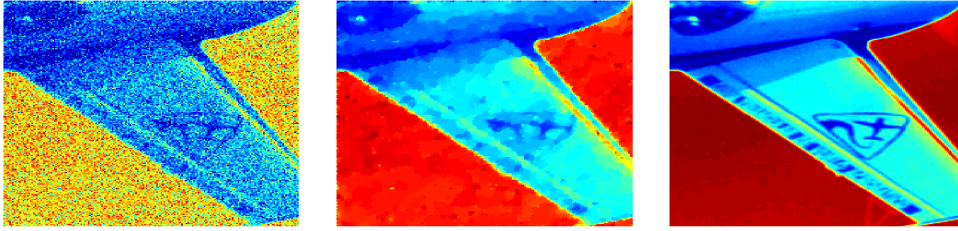

FIG. 4.4. *Noisy (left), denoised (center) and original (right) images*

Further, we tested the stability of the optimal parameter values with respect to changes in the source noisy image. This behavior is registered in Table 4.2. On the first data column the optimal weights corresponding to a zoomed image displaced by 10 grid points down and 10 grid points left is registered. The same kind of data is registered in the second column for a displacement of 5 grid points down and 5 left. In the fourth column the data corresponds to a displacement in the zoom by 5 grid points right and 5 up. Similarly, in the fifth column for 10 grid points. From Table 4.2 a robust behavior of the parameter values can be inferred. Indeed, by changing the source image by 10%, the optimal weights change less than 16%.

4.4. Impulse noise. For the last experiment we consider images with so-called impulse noise. Specifically, we aim to solve the following parameter estimation problem:

$$\min \; \frac{1}{2}\|u - u_o\|_{L^2}^2 + \beta\lambda^2 \tag{4.7}$$

subject to:

$$\varepsilon(Du, D(v-u))_{L^2} + \lambda \int_\Omega |v - u_n| \; dx - \lambda \int_\Omega |u - u_n| \; dx$$
$$+ \int_\Omega |Dv| \; dx - \int_\Omega |Du| \; dx \geq 0, \forall v \in H_0^1(\Omega). \tag{4.8}$$

19

%Displacement	-10	-5	0	5	10
% Sensitivity of λ^*	15.56	0.3	0	0.15	8.26
λ_1^*	1270.9	1103.4	1100.7	1099.8	1189.8
λ_2^*	26.39	48.24	46.37	44.92	28.13

TABLE 4.2

Optimal weight sensitivity by moving the sample image along the diagonal; $\mu = 1e-12$, $\gamma = 50$, $\beta = 1e-10$ $h = 1/100$.

Equation (4.8) corresponds to the necessary and sufficient optimality condition for the optimization problem:

$$\min_u \left\{ \frac{\varepsilon}{2} \|Du\|_{L^2}^2 + |Du|(\Omega) + \lambda \int_\Omega |u - u_n| \, dx \right\}. \qquad (4.9)$$

The L^1-norm is introduced to deal with the sparse impulse noise in the image. The presence of this norm adds, however, an additional nondifferentiability to the optimization problem.

For the numerical solution of the lower level problem we consider a Huber type regularization of both the TV term and the L^1-norm. Using a common regularization parameter γ, the resulting nonlinear PDE takes the following form:

$$-\epsilon \Delta u - \text{div}\left(\frac{\gamma \nabla u}{\max(\gamma |\nabla u|, 1)} \right) + \lambda \frac{\gamma(u - u_n)}{\max(1, \gamma |u - u_n|)} = 0, \qquad (4.10)$$

or, in primal-dual form,

$$-\epsilon \Delta u - \text{div } q + \lambda \, p = 0, \qquad (4.11)$$

$$q = \frac{\gamma \nabla u}{\max(\gamma |\nabla u|, 1)}, \qquad (4.12)$$

$$p = \frac{\gamma(u - u_n)}{\max(1, \gamma |u - u_n|)}. \qquad (4.13)$$

The nonlinearities in equation (4.10) are present both in the quasilinear and the semilinear terms. Both of them have to be be carefully treated in order to obtain a convergent numerical method for the solution.

Proceeding in a similar manner as in Section 5.2, a semismooth Newton iteration for the impulse noise lower level problem is given by

$$-\epsilon \Delta \delta_u - \text{div}\delta_q + \lambda \delta_p = -\left(-\epsilon \Delta u - \text{div}q + \lambda p \right), \qquad (4.14)$$

$$\delta_q - \frac{\gamma \nabla \delta_u}{\max(1, \gamma |\nabla u|)} + \chi_{\mathcal{A}_\gamma} \frac{\gamma^2 \nabla u^T \nabla \delta_u}{\max(1, \gamma |\nabla u|)^2} \frac{\nabla u}{|\nabla u|} = -q + \frac{\gamma \nabla u}{\max(1, \gamma |\nabla u|)}, \qquad (4.15)$$

$$\delta_p - \frac{\gamma \delta_u}{\max(1, \gamma |u - u_n|)} + \chi_{\mathcal{S}_\gamma} \frac{\gamma^2 (u - u_n) \delta_u}{\max(1, \gamma |u - u_n|)^2} \frac{(u - u_n)}{|u - u_n|}$$
$$= -p + \frac{\gamma(u - u_n)}{\max(1, \gamma |u - u_n|)}. \qquad (4.16)$$

Using a similar argumentation as for the Gauss+Poisson noise case, we consider the

Iteration	λ^*	Cost functional	Residuum	#SSN iterations
1	10.0015	0.0124	0.0015	15
2	16.7752	0.0124	0.0015	3
3	19.2223	0.0064	4.024e-4	14
4	10.0854	0.0048	5.496e-4	12
5	24.9562	0.0123	0.0014	17
6	26.2300	0.0018	1.124e-4	18
7	30.2286	0.0017	8.530e-5	10
8	45.8756	0.0013	6.794e-5	12
9	48.7340	7.83e-4	1.049e-5	13
10	73.2269	7.55e-4	9.397e-6	7
11	57.9833	8.12e-4	1.54e-5	12
12	58.2922	6.81e-4	3.20e-7	19

TABLE 4.3

Optimal weight sensitivity by moving the sample image along the diagonal; $\mu = 1e-12$, $\gamma = 50$, $\beta = 1e-10$ $h = 1/40$.

modified system

$$-\epsilon\Delta\delta_u - \mathrm{div}\delta_q + \lambda\delta_p = -\left(-\epsilon\Delta u - \mathrm{div}q + \lambda p\right), \qquad (4.17)$$

$$\delta_q - \frac{\gamma\nabla\delta_u}{\max(1,\gamma|\nabla u|)} + \chi_{\mathcal{A}_\gamma}\gamma^2\frac{\nabla u^T\nabla\delta_u}{\max(1,\gamma|\nabla u|)^2}\frac{q}{\max(1,|q|)}$$
$$= -q + \frac{\gamma\nabla u}{\max(1,\gamma|\nabla u|)}, \qquad (4.18)$$

$$\delta_p - \frac{\gamma\delta_u}{\max(1,\gamma|u-u_n|)} + \chi_{\mathcal{S}_\gamma}\frac{\gamma^2(u-u_n)\delta_u}{\max(1,\gamma|u-u_n|)^2}\frac{p}{\max(1,|p|)}$$
$$= -p + \frac{\gamma(u-u_n)}{\max(1,\gamma|u-u_n|)}. \qquad (4.19)$$

where we replaced the terms $\frac{\nabla u}{|\nabla u|}$ and $\frac{(u-u_n)}{|u-u_n|}$ on the left hand side by $\frac{q}{\max(1,|q|)}$ and $\frac{p}{\max(1,|p|)}$, respectively.

The behavior of the resulting BFGS-SSN algortihm is registered in Table 4.3. For the parameter values $\mu = 1e-12$, $\gamma = 50$, $\beta = 1e-10$ $h = 1/40$ the algorithm takes 12 iterations to converge. The number of iterations of the lower level algorithm, given through (4.17)- (4.19), is registered in the last column, from which the fast convergence of the method is experimentally verified.

REFERENCES

[1] A. Almansa, C. Ballester, V. Caselles, and G. Haro, *A TV based restoration model with local constraints*, J. Sci. Comput., 34(3), 209–236, 2008.
[2] G. Aubert, and J.-F. Aujol, *A Variational Approach to remove Multiplicative Noise*, SIAM Journal on Applied Mathematics, volume 68, number 4, 925–946, January 2008.
[3] G. AUBERT AND L. VESE, *A variational method in image recovery*, SIAM J. Numer. Anal. 34 (1997) pp. 1948–1979.
[4] M. Bertalmio, V. Caselles, B. Rougé, and A. Solé, *TV based image restoration with local constraints*, Journal of Scientific Computing, 19:95–122, 2003.
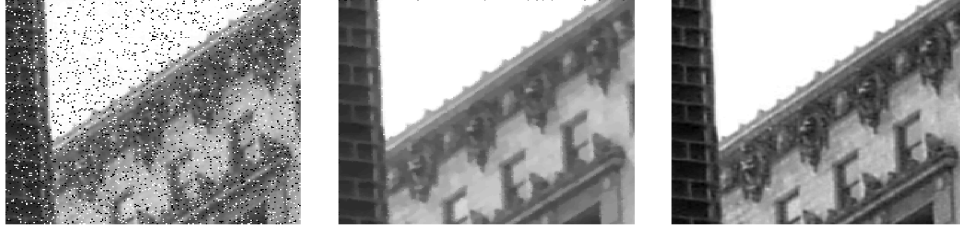
Fɪɢ. 4.5. *Noised (left), de-noised (center) and original (right) images, $\gamma = 50$, $\epsilon = 1e-12$, $h = 1/200$*

[5] A. Bovik, *Handbook of Image and Video Processing.* Academic Press, 2000.
Maass. *An optimal control problem in medical image processing.* Systems, Control, Modeling and Optimization Proceedings of the 22nd IFIP TC7 Conference held from July 18-22, 2005, in Turin, Italy, 2005.

[6] Bouchitté G, Braides A, Buttazzo G (1995) *Relaxation results for some free discontinuity problems.* J Reine Angew Math 458:118 5.

[7] Bouchitté G, Buttazzo G (1990) *New lower semicontinuity results for nonconvex functionals defined on measures.* Nonlinear Anal TMA 15(7):679692 6.

[8] Bouchitté G, Buttazzo G (1992) *Integral-representation of nonconvex functionals defined on measures.* Ann Inst H Poincaré 9(1):101117 7.

[9] Bouchitté G, Buttazzo G (1993) *Relaxation for a class of nonconvex functionals defined on measures.* Ann Inst H Poincaré 10(3):345361

[10] Casas E., Fernández L. *Distributed Control of Systems Governed by a General Class of Quasilinear Elliptic Equations.* J. Differential Equations, 104 (1993), pp. 20–47.

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

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

[13] A. Cʜᴀᴍʙᴏʟʟᴇ, V. Cᴀsᴇʟʟᴇs, D. Cʀᴇᴍᴇʀs, M. Nᴏᴠᴀɢᴀ, ᴀɴᴅ T. Pᴏᴄᴋ, *An Introduction to Total Variation for Image Analysis*, Theoretical Foundations and Numerical Methods for Sparse Recovery (M. Fornasier, ed.), Radon Series on Computational and Applied Mathematics, De Gruyter Verlag, 2010, pp. 263–340.

[14] T. F. Chan, and S. Esedoglu, *Aspects of total variation regularised L^1 function approximation*, Siam J. Appl. Math., Vol. 65, No. 5, pp. 18171837, 2005.

[15] T. F. Chan, and J. J. Shen, *Image Processing and Analysis - Variational, PDE, wavelet, and stochastic methods.* SIAM, (2005).

[16] G. Dal Maso, *An introduction to Gamma-convergence*, Birkhäuser, Boston, 1993.

[17] J.C. De Los Reyes. Optimal control of a class of variational inequalities of the second kind. *SIAM Journal on Control and Optimization*, Vol. 49, 1629-1658, 2011.

[18] J.C. De Los Reyes. Optimization of mixed variational inequalities arising in flow of viscoplastic materials. *Computational Optimization and Applications*, DOI: 10.1007/s10589-011-9435-x, 2011.

[19] Demengel F, Temam R (1984) *Convex functions of a measure and applications.* Indiana Univ Math J 33:673709.

[20] D. C. Dᴏʙsᴏɴ ᴀɴᴅ C. R. Vᴏɢᴇʟ, *Convergence of an iterative method for total variation denoising*, SIAM J. Numer. Anal. 34 (1997), pp. 1779–1791.

[21] V. Duval, J.-F. Aujol, and Y. Gousseau, *The TVL1 model: a geometric point of view*, SIAM Journal on Multiscale Modeling and Simulation, volume 8, number 1, 154–189, November 2009.

[22] K. Frick, P. Marnitz, A. Munk, *Statistical Multiresolution Dantzig Estimation in Imaging: Fundamental Concepts and Algorithmic Framework*, Electron. J. Stat., 6, 231–268, 2012.

[23] K. Frick, P. Marnitz, A. Munk, *Shape Constrained Regularisation by Statistical Multiresolution for Inverse Problems*, Inverse Problems, 28, 065006, 2012.

[24] K. Frick, P. Marnitz, A. Munk, *Statistical Multiresolution Estimation for Variational Imaging: With an Application in Poisson-Biophotonics*, J. Math. Imaging Vision. To appear.

[25] M. Hintermüller, Y. Dong, M.M. Rincon-Camacho, *Automated Regularization Parameter Selection in Multi-Scale Total Variation Models for Image Restoration*, Journal of Mathematical

Imaging and Vision 40 (1), pp. 82–104, 2011.

[26] M. Hintermüller and K. Kunisch *Total bounded variation regularization as a bilaterally constrained optimization problem.* *SIAM Journal on Applied Mathematics*, Vol. 64, 1311–1333, 2004.

[27] M. Hintermüller and K. Kunisch, *Stationary Optimal Control Problems with Pointwise State Constraints*, Lecture Notes in Computational Science and Engineering, 72, 2009.

[28] M. Hintermüller and G. Stadler *An Infeasible Primal-Dual Algorithm for Total Bounded Variation–Based Inf-Convolution-Type Image Restoration* *SIAM Journal on Scientific Computing*, Vol. 28, 1–23, 2006.

[29] T. Le, R. Chartrand, and T.J. Asaki, *A variational approach to reconstructing images corrupted by Poisson noise*, J. Math. Imaging Vision 27(3), 257–263, 2007.

[30] Risheng Liu, Zhouchen Lin, Wei Zhang and Zhixun Su, *Learning PDEs for Image Restoration via Optimal Control*, ECCV 2010.

[31] V.A. Morozov, *Regularization Methods for Ill–posed Problems*, CRC Press, Boca Raton, 1993.

[32] M. Nikolova, *A variational approach to remove outliers and impulse noise*, JMIV, vol. 20, 99–120, 2004.

[33] A. Sawatzky, C. Brune, J. Müller, M. Burger, *Total Variation Processing of Images with Poisson Statistics*, Proceedings of the 13th International Conference on Computer Analysis of Images and Patterns, Volume 5702, 533–540, July 2009.

[34] F. R. Schmidt, D. Cremers, *A Closed-Form Solution for Image Sequence Segmentation with Dynamical Shape Priors*, In Pattern Recognition (Proc. DAGM), 2009.

[35] D. Strong, J.-F. Aujol, and T. Chan, *Scale recognition, regularization parameter selection, and Meyers G norm in total variation regularization*, Technical report, UCLA, 2005.

[36] I. Tosic, I. Jovanovic, P. Frossard, M. Vetterli and N. Duric, *Ultrasound Tomography with Learned Dictionaries*, IEEE International Conference on Acoustics, Speech, and Signal Processing, Dallas, Texas, International Conference on Acoustics Speech and Signal Processing ICASSP, 2010.

[37] L. Vese, *A study in the BV space of a denoising-deblurring variational problem*, Appl Math Optim 44 (2001), pp. 131–161.

[38] C. R. Vogel and M. E. Oman, *Iterative methods for total variation denoising*, SIAM J. Sci. Comput. **17** (1996), no. 1, 227–238, Special issue on iterative methods in numerical linear algebra (Breckenridge, CO, 1994).

[39] A. M. Yip, and F. Park, *Solution Dynamics, Causality, and Critical Behavior of the Regularization Parameter in Total Variation Denoising Problems*, CAM reports 03-59, 2003.