SPATIALLY DEPENDENT REGULARIZATION PARAMETER SELECTION IN TOTAL GENERALIZED VARIATION MODELS FOR IMAGE RESTORATION

KRISTIAN BREDIES, YIQIU DONG, AND MICHAEL HINTERMÜLLER

ABSTRACT. The automated spatially dependent regularization parameter selection framework of [9] for multi-scale image restoration is applied to total generalized variation (TGV) of order two. Well-posedness of the underlying continuous models is discussed and an algorithm for the numerical solution is developed. Experiments confirm that due to the spatially adapted regularization parameter the method allows for a faithful and simultaneous recovery of fine structures and smooth regions in images. Moreover, because of the TGV regularization term, the adverse staircasing effect, which is a well-known drawback of the total variation regularization, is avoided.

1. INTRODUCTION

Image acquisition or transmission typically introduce degradation effects, respectively, which need to be undone or reduced for performing further image processing tasks. For a mathematical description of such degradations suppose that an image $\hat{u}$ is a real function defined on a bounded and piecewise smooth open subset (image domain) $\Omega \subset \mathbb{R}^2$, i.e., $\hat{u} : \Omega \to \mathbb{R}$. The degraded image z is given by

$$z = K\hat{u} + \eta,$$

where $K \in \mathcal{L}(L^2(\Omega))$ is a known linear and continuous blurring operator, and η represents white Gaussian noise with zero mean and standard deviation σ. Thus, given the data z, we are interested in recovering $\hat{u}$, which is well-known to be an ill-posed problem, in general; see, e.g., [18].

In order to restore the image stably, regularization techniques are usually applied, and the choice of the associated regularization term is based on *prior* information on $\hat{u}$. Given the statistical properties of η, a rather general paradigm in image restoration requires to solve

1

the constrained minimization problem

$$
\begin{aligned}
&\text{minimize } R(u) \text{ over } u \in L^2(\Omega) \\
&\text{subject to } \int_\Omega |Ku - z|^2 \ \mathrm{d}x \leq \sigma^2 |\Omega|,
\end{aligned}
\tag{1.1}
$$

where the structure of the constraint results from the Gaussian noise model, and $R(u)$ is based on prior information on $\hat{u}$. Further, $|\Omega|$ denotes the (Lebesgue) measure of Ω. In order to avoid the realm of constrained optimization (and to avoid knowledge of σ), usually (1.1) is solved via the following unconstrained optimization problem

$$
\text{minimize } \frac{\lambda}{2} \int_\Omega |Ku - z|^2(x) \ \mathrm{d}x + R(u) \text{ over } u \in L^2(\Omega),
\tag{1.2}
$$

where $\lambda \geq 0$ is the Lagrange multiplier associated with the constraint in (1.1). It is well-known that λ is also strongly related to the regularization parameter in regularization techniques.

Concerning the choice of the regularization term $R(u)$, one of the most straightforward choices is Tikhonov regularization [17], i.e., $R(u) := \int_\Omega |\nabla u|_2^2$. This choice leads to rather smooth images. Adversely, in particular edges in images appear rather diffuse than sharp. With the goal of preserving significant edges in images, total variation (TV) regularization $R(u) := \int_\Omega |Du|$ was proposed in [13]. Although TV regularization is very effective when restoring piecewise constant images, it has several shortcomings. Probably, the most prominent one is the appearance of staircasing artifacts in slanted regions [2, 12]. To overcome this problem, higher-order derivatives have been used; see [6, 8]. In [4], recently total generalized variation of order k, $\mathrm{TGV}_\alpha^k(u)$, was proposed as a concept which preserves edges while still imposing some higher order smoothness in regions away from edges. In fact, in homogeneous regions it imposes smoothness from the first up to the k-th derivative. When k equals 1, it yields TV regularization. Substituting it into the model (1.2) and considering the case of $k = 2$, we obtain the total generalized variation model

$$
\text{minimize } \frac{\lambda}{2} \int_\Omega |Ku - z|^2(x) \ \mathrm{d}x + \mathrm{TGV}_\alpha^2(u) \text{ over } u \in \mathrm{BGV}^2(\Omega),
\tag{1.3}
$$

where $\mathrm{BGV}^2(\Omega)$ is the space of functions of bounded generalized variation of order 2, which is, endowed with the norm $\|u\|_{\mathrm{BGV}^2(\Omega)} = \|u\|_1 + \mathrm{TGV}_\alpha^2(u)$, a Banach space; see [4] for details.

In the optimization problem (1.3), the Lagrange multiplier λ controls the trade-off between a good fit of z and a smoothness requirement due to the total generalized variation regularization. The fact that in general images are comprised of multiple objects at different scales suggests that different values of λ localized at image features of different scales are desirable to obtain better restoration results. For this reason, the use of a spatially dependent regularization parameter $\lambda \in L^\infty(\Omega)$ was proposed in [1, 9, 14]. Combining this concept with the TGV restoration model (1.3), we obtain a spatially adapted total generalized variation model (SA-TGV model) given by

$$
\text{minimize } \frac{1}{2} \int_\Omega \lambda(x) |Ku - z|^2(x) \ \mathrm{d}x + \mathrm{TGV}_\alpha^2(u) \text{ over } u \in \mathrm{BGV}^2(\Omega)
\tag{1.4}
$$

with $\lambda \geq 0$ almost everywhere (a.e.) in Ω.

In this paper, we study the existence and uniqueness of a solution to (1.4). Based on the properties of the second order total generalized variation, we are able to extend the theoretical results of [9] to the case of TGV regularization. In combination with the spatially dependent parameter choice rule proposed in [9], we introduce an algorithmic framework

for solving (1.4) and updating $\lambda \in L^\infty(\Omega)$ based on statistical methods (SA-TGV method). Furthermore, in order to reduce the number of iterations and to obtain robustness with respect to the initial choice of λ, a hierarchical decomposition scheme similar to the one considered in [15] is introduced. The numerical results in this paper show that the SA-TGV method overcomes the drawback of the staircasing effect when compared to the SA-TV method of [9]. Furthermore, with the spatially dependent parameter the SA-TGV method outperforms the usual TGV method, which restores images by solving the model (1.3), with respect to both noise removal and detail preservation.

The rest of the paper is organized as follows. Section 2 gives a summary of the total generalized variation regularization and provides its main properties. In Section 3, we study the existence and uniqueness of a solution of the model (1.4). Moreover, we extend the spatially dependent parameter choice rule proposed in [9] to the model with TGV regularization. Based on an extension of the hierarchical decomposition concept proposed in [15], we describe the SA-TGV method in Section 4. The numerical results shown in Section 5 demonstrate the effectivity of the new method. Finally, conclusions are drawn in Section 6.

2. Total Generalized Variation – TGV

The total generalized variation recently introduced in [4] is a convex variational image model which is able to incorporate smoothness up to a certain differential order while still accounting for edges. It can be seen as a generalization of the well-known total variation model [13]. In the following, we give a short introduction and summarize the results on TGV which are relevant for the present paper.

In this section, we consider a general space dimension $d \in \mathbb{N}$ and denote by $\Omega \subset \mathbb{R}^d$ a bounded domain. The total generalized variation is then defined as

(2.1)
$$\mathrm{TGV}_\alpha^k(u) = \sup\left\{ \int_\Omega u \operatorname{div}^k v \, \mathrm{d}x : v \in C_c^k(\Omega, \mathrm{Sym}^k(\mathbb{R}^d)), \ \|\operatorname{div}^l v\|_\infty \le \alpha_l \text{ for } l = 0, \ldots, k-1 \right\}.$$

It depends on an order $k \in \mathbb{N}$ and weights $\alpha = (\alpha_0, \ldots, \alpha_{k-1})$ which are assumed to be positive. Here, $\mathrm{Sym}^k(\mathbb{R}^d)$ denotes the space of symmetric k-tensors. The l-divergence of a symmetric k-tensor field is a symmetric $(k-l)$-tensor field which is given by

$$(\operatorname{div}^l v)_\beta = \sum_{\gamma \in M_l} \frac{l!}{\gamma!} \frac{\partial^l v_{\beta+\gamma}}{\partial x^\gamma} \qquad \text{for each component } \beta \in M_{k-l},$$

where M_k are the multiindices of order k, i.e., $M_k = \{\beta \in \mathbb{N}^d \mid \sum_{i=1}^d \beta_i = k\}$. The ∞-norm for symmetric k-vector fields is given by

$$\|v\|_\infty = \sup_{x \in \Omega} \left\{ \left(\sum_{\beta \in M_k} \frac{k!}{\beta!} v_\beta(x)^2 \right)^{1/2} \right\}$$

which corresponds to the ∞-norm with respect to the pointwise Frobenius norm for tensors. It is obvious from the definition that in the case where $k = 1$, the total generalized variation coincides, up to a factor, with the total variation, i.e., we have $\mathrm{TGV}_\alpha^k = \alpha_0 \mathrm{TV}$. Having defined TGV_α^k according to (2.1), one can show that it constitutes a proper, convex and (weak) lower semi-continuous functional on each $L^p(\Omega)$ space ($1 \le p < \infty$) which is moreover translation and rotation invariant. The space

$$\mathrm{BGV}^k(\Omega) = \{u \in L^1(\Omega) \mid \mathrm{TGV}_\alpha^k(u) < \infty\}, \qquad \|u\|_{\mathrm{BGV}^k} = \|u\|_1 + \mathrm{TGV}_\alpha^k(u)$$

is a Banach space which is independent of the weights $\alpha_0, \ldots, \alpha_{k-1}$ chosen in the definition of TGV_α^k. On this space TGV_α^k is a semi-norm which vanishes exactly on $\mathcal{P}_{k-1}(\Omega)$, the space of polynomials of degree less than or equal to $k-1$. See [4] for proofs and more details.

In this paper, we will focus on total generalized variation of second order, i.e., $k = 2$. In this case, (2.1) can equally be written as

$$\mathrm{TGV}_\alpha^2(u) = \sup \left\{ \int_\Omega u \, \mathrm{div}^2 \, v \, \mathrm{d}x \,\Big|\, v \in C_\mathrm{c}^2(\Omega, S^{d \times d}), \ \|v\|_\infty \leq \alpha_0, \ \|\mathrm{div} \, v\|_\infty \leq \alpha_1 \right\}$$

with $S^{d \times d}$ denoting the space of symmetric $d \times d$ matrices. The first and second divergences of a symmetric matrix field are then vector and scalar fields, respectively, given by

$$(\mathrm{div} \, v)_i = \sum_{j=1}^d \frac{\partial v_{ij}}{\partial x_j}, \qquad (\mathrm{div}^2 \, v) = \sum_{i=1}^d \frac{\partial v_{ii}}{\partial x_i^2} + 2 \sum_{i=1}^2 \sum_{j<i} \frac{\partial v_{ij}}{\partial x_i \partial x_j}.$$

Likewise, the ∞-norms of matrix and vector fields v and w, respectively, used here are given by

$$\|v\|_\infty = \sup_{x \in \Omega} \left\{ \left(\sum_{i=1}^d v_{ii}(x)^2 + 2 \sum_{i=1}^d \sum_{j<i} v_{ij}(x)^2 \right)^{1/2} \right\}, \quad \|w\|_\infty = \sup_{x \in \Omega} \left\{ \left(\sum_{i=1}^d w_i(x)^2 \right)^{1/2} \right\}.$$

For the second-order total generalized variation it was shown in [5] that for $u \in L^1(\Omega)$ one has

$$\mathrm{TGV}_\alpha^2(u) = \min_{w \in \mathrm{BD}(\Omega)} \alpha_1 \|\nabla u - w\|_\mathcal{M} + \alpha_0 \|\mathcal{E}w\|_\mathcal{M}.$$

Here, $\mathrm{BD}(\Omega)$ denotes the space of vector fields of bounded deformation [16], i.e., the set of vector fields whose weak symmetrized derivative $\mathcal{E}w = \frac{1}{2}(\nabla w + \nabla w^\top)$ is a matrix-valued Radon measure. Moreover, $\|\cdot\|_\mathcal{M}$ denotes the Radon norm for vector-valued and matrix-valued Radon measures, respectively.

Furthermore, for bounded Lipschitz domains $\Omega \subset \mathbb{R}^d$, $\mathrm{BGV}^2(\Omega)$ coincides with $\mathrm{BV}(\Omega)$, the space of functions of bounded variation, in the topological sense, i.e., there exist $0 < c < C < \infty$ such that for each $u \in \mathrm{BV}(\Omega)$,

$$c\big(\|u\|_1 + \mathrm{TGV}_\alpha^2(u)\big) \leq \|u\|_1 + \mathrm{TV}(u) \leq C\big(\|u\|_1 + \mathrm{TGV}_\alpha^2(u)\big).$$

We therefore have also the usual embeddings $\mathrm{BGV}^2(\Omega) \hookrightarrow L^p(\Omega)$ for $1 \leq p \leq d/(d-1)$ which are compact for $p < d/(d-1)$. Finally, there is also a variant of the Poincaré-Friedrichs inequality available which states that for a linear projection $P : L^{d/(d-1)}(\Omega) \to \mathcal{P}_1(\Omega)$ onto $\mathcal{P}_1(\Omega)$, the space of polynomials of degree less or equal than 1, we can find a constant $C > 0$, only depending on Ω, P and α such that

$$(2.2) \qquad \|u - Pu\|_{d/(d-1)} \leq C \, \mathrm{TGV}_\alpha^2(u) \qquad \text{for all } u \in \mathrm{BV}(\Omega).$$

Proofs and more details regarding these results can be found in [5].

3. Spatial Dependent Parameter Selection

In the following, the space dimension is fixed to $d = 2$ as well as the order of the total generalized variation functional, i.e., $k = 2$. We recall that $\Omega \subset \mathbb{R}^2$ is bounded with a piecewise smooth boundary. For $z \in L^2(\Omega)$ we assume that

$$(3.1) \qquad \text{there exists a } u \in \mathrm{BGV}^2(\Omega) \text{ such that } |Ku - z|(x) \leq \sigma \text{ for almost every } x \in \Omega.$$

In order to enhance image details while preserving homogeneous regions, instead of utilizing the global constraint in (1.1) we consider that the constraint is satisfied locally. For this purpose, assume that w is a normalized filter, i.e. $w \in L^\infty(\Omega \times \Omega)$, $w \geq 0$ on $\Omega \times \Omega$ with

$$(3.2) \qquad \int_\Omega w(x,y) \, \mathrm{d}y \equiv 1 \quad \text{and} \quad \int_\Omega \int_\Omega w(x,y)\phi^2(y) \, \mathrm{d}y \, \mathrm{d}x \geq \epsilon \|\phi\|_{L^2(\Omega)}^2 \quad \forall \phi \in L^2(\Omega)$$

for some $\epsilon > 0$ (independent of ϕ). The local variance of the residual can be interpreted as

$$(3.3) \qquad S(u)(x) := \int_\Omega w(x,y)(Ku-z)^2(y) \, \mathrm{d}y.$$

Referring to (1.1), our image restoration model with total generalized variation and local constraints is defined as

$$(3.4) \qquad \begin{aligned} &\text{minimize } \mathrm{TGV}_\alpha^2(u) \quad \text{over } u \in \mathrm{BGV}^2(\Omega) \\ &\text{subject to } S(u) - \sigma^2 \leq 0 \quad \text{a.e. in } \Omega. \end{aligned}$$

Obviously, the feasible set $\mathcal{U} := \{u \in \mathrm{BGV}^2(\Omega) \mid S(u) \leq \sigma^2 \text{ a.e. in } \Omega\}$ is closed and convex.

3.1. **Existence and uniqueness of a solution.** Based on the properties of total generalized variation and the space of bounded generalized variation functions, we prove the existence of a solution of (3.4).

Theorem 1. *Assume that $K \in \mathcal{L}(L^2(\Omega))$ does not annihilate affine functions, i.e., $Ku \neq 0$ whenever $u \in \mathcal{P}_1(\Omega)$ is non-zero. Then, the problem (3.4) admits a solution.*

Proof. With the assumption (3.1) on z and (3.2) we are able to find at least one image $u \in \mathcal{U}$, which has finite $\mathrm{TGV}_\alpha^2(u)$. Then, there exists a minimizing sequence $\{u_n : n = 1, 2, \cdots\} \subset \mathcal{U}$ for TGV_α^2. So $\{\mathrm{TGV}_\alpha^2(u_n)\}$ with $n = 1, 2, \cdots$ is bounded. Moreover, $\mathrm{TGV}_\alpha^2(u_n - Pu_n) = \mathrm{TGV}_\alpha^2(u_n)$, where $P : L^2(\Omega) \to \mathcal{P}_1(\Omega)$ is a linear projector onto the space of affine functions. As Ω is a Lipschitz domain, we can employ inequality (2.2) to obtain that $g_n = u_n - Pu_n$ is bounded in $L^2(\Omega)$. From the properties of the filter w and the fact that $u_n \in \mathcal{U}$ for each n we get

$$\int_\Omega \sigma^2 \, \mathrm{d}x \geq \int_\Omega S(u)(x) \, \mathrm{d}x = \int_\Omega \int_\Omega w(x,y)(Ku_n - z)^2(y) \, \mathrm{d}y \, \mathrm{d}x \geq \epsilon \|Ku_n - z\|_2^2.$$

Recalling that Ω is bounded it follows that $\|Ku_n - z\|_2^2$ is bounded from above by $\sigma^2 |\Omega|/\epsilon$. With $z \in L^2(\Omega)$, $\{Ku_n\}$ is bounded in $L^2(\Omega)$. Since $K \in \mathcal{L}(L^2(\Omega))$ is continuous, $Kg_n = Ku_n - KPu_n$ must give a bounded sequence in $L^2(\Omega)$ and, consequently, $\{KPu_n\}$ is bounded. As K is injective on the finite-dimensional space $\mathcal{P}_1(\Omega)$ by assumption, K can be inverted continuously on $\mathcal{P}_1(\Omega)$ and hence, $\{Pu_n\}$ must be bounded in $L^2(\Omega)$. Together with the boundedness of $\{g_n\}$, this leads to the boundedness of $\{u_n\}$ in $L^2(\Omega)$. Thus, there exists a weakly convergent subsequence $\{u_{n_l}\}$ with limit $u_* \in L^2(\Omega)$. Due to the weak lower-semicontinuity of TGV_α^2, we obtain

$$\mathrm{TGV}^2(u_*) \leq \liminf_{l \to \infty} \mathrm{TGV}_\alpha^2(u_{n_l}),$$

implying in particular that $u_* \in \mathrm{BGV}^2(\Omega)$. Moreover, since $\mathcal{U}$ is closed and convex, we have $S(u_*) \leq \sigma^2$ a.e. in Ω. As $\{u_n\}$ was a minimizing sequence, u_* is a solution to (3.4). $\qquad \square$

With an extra assumption on the filter w we can establish a uniqueness result.

Assumption 2. For $u_1, u_2 \in \mathrm{BGV}^2(\Omega)$, two solutions of (3.4) with $u_1 \neq u_2$, the existence of a $\delta > 0$ and $\Omega_\delta \subset \Omega$ with $|\Omega_\delta| > 0$ such that

$$\left(\frac{1}{2}K(u_1 + u_2) - z\right)^2 \leq \frac{1}{2}\left((Ku_1 - z)^2 + (Ku_2 - z)^2\right) - \delta \quad \text{a.e. in } \Omega_\delta$$

implies that there exists a $\epsilon_\delta > 0$ such that

$$(3.5) \qquad \int_\Omega w(x, y)\left(\frac{1}{2}K(u_1 + u_2) - z\right)^2(y)\,\mathrm{d}y \leq \sigma^2 - \epsilon_\delta \quad \text{for almost all } x \in \Omega.$$

Assumption 2 is for instance satisfied if there exists a $w_0 > 0$ such that $w \geq w_0$ almost everywhere in $\Omega \times \Omega$: Then, it can be seen that (3.5) holds with $\epsilon_\delta = |\Omega_\delta| w_0 \delta$. Uniqueness of solutions up to the kernel of K is now shown in the following theorem.

Theorem 3. *Assume that $K \in \mathcal{L}(L^2(\Omega))$ does not annihilate affine functions and that Assumption 2 is satisfied. In addition we suppose that*

$$(3.6) \qquad \inf_{u \in \mathcal{P}_1(\Omega)} S(u)(x) > \sigma^2 \ \text{a.e. in } \Omega.$$

Then, for every solution u_* of (3.4) Ku_* has the same value a.e. in Ω.

Proof. Let $u_1, u_2 \in \mathrm{BGV}^2(\Omega)$ denote two solutions with $u_1 \neq u_2$. Define $\bar{u} = \frac{1}{2}(u_1 + u_2)$. By convexity, we have

$$(K\bar{u} - z)^2 \leq \frac{1}{2}\left((Ku_1 - z)^2 + (Ku_2 - z)^2\right).$$

The inequality becomes an equality a.e. in Ω if and only if $Ku_1 = Ku_2$ a.e. in Ω. Thus, if $Ku_1 \neq Ku_2$, then strict inequality holds on a non-null set and it is possible to find $\delta > 0$ and $\Omega_\delta \subset \Omega$, $|\Omega_\delta| > 0$, such that

$$(K\bar{u} - z)^2 \leq \frac{1}{2}\left((Ku_1 - z)^2 + (Ku_2 - z)^2\right) - \delta \qquad \text{a.e. on } \Omega_\delta.$$

By Assumption 2 it follows that $S(\bar{u}) \leq \sigma - \epsilon_\delta$ almost everywhere in Ω for $\epsilon_\delta > 0$. In particular, as $S : L^2(\Omega) \to L^\infty(\Omega)$ is continuous, it is possible to find an s such that $0 < s < 1$ and $s\bar{u} \in \mathcal{U}$. Based on the definition of TGV_α^2, we get the contradiction

$$\mathrm{TGV}_\alpha^2(s\bar{u}) = s\,\mathrm{TGV}_\alpha^2(\bar{u}) < \mathrm{TGV}_\alpha^2(\bar{u}) = \min_{u \in \mathcal{U}}\,\mathrm{TGV}_\alpha^2(u),$$

unless $\mathrm{TGV}_\alpha^2(\bar{u}) = 0$, i.e., $\bar{u} \in \mathcal{P}_1(\Omega)$. But, as $S(\bar{u}) > \sigma^2$ a.e. in Ω by (3.6), this also results in a contradiction. Hence, $Ku_1 = Ku_2$ a.e. in Ω. $\qquad\square$

3.2. Spatial adaptation by local variance estimators.
Since the restoration capability of the model (1.4) hinges on the selection of the parameter λ, in this section we review the choice rule proposed in [9] and extend it to the model (1.4).

In order to preserve image details as well as to remove noise, the choice of λ must be based on local image features. To this end, relying on the local constraints in (3.4), we search for a reconstruction where the variance of the residual is close to the noise variance in both the detail regions and the homogeneous regions.

From now on we confine ourselves to the discrete version of the problem. Hence, we assume that the discrete image domain Ω contains $m_1 \times m_2$ pixels. For the sake of simplicity,

however, we keep the notations from the continuous context in what follows. We define the local variance estimator at the pixel (i, j) and depending on a ω-by-ω window by

$$(3.7) \qquad S_{i,j}^{\omega} := \frac{1}{\omega^2} \sum_{(s,t) \in \Omega_{i,j}^{\omega}} (z_{s,t} - (Ku)_{s,t})^2,$$

where $\Omega_{i,j}^{\omega} = \left\{ (s+i, t+j) : -\frac{\omega-1}{2} \leq s, t \leq \frac{\omega-1}{2} \right\}$, and a Neumann boundary condition is applied at boundary pixels. In view of (3.3), the local variance estimators approximate the values of $S(u)$ in the constraints by employing a mean filter (w in (3.2)) in a $\omega \times \omega$ filter window. Then, with the local constraints in (3.4) we are able to decide on the amount of details contained in the window around (i, j).

Since the noise η can be regarded as an array of independent normally distributed random variables with zero mean and variance σ^2, the random variable

$$T_{i,j}^{\omega} = \frac{1}{\sigma^2} \sum_{(s,t) \in \Omega_{i,j}^{\omega}} (\eta_{s,t})^2$$

has the χ^2-distribution with ω^2 degrees of freedom, i.e. $T_{i,j}^{\omega} \sim \chi_{\omega^2}^2$. If the restored image is $\hat{u}$, the real image, then we should have

$$S_{i,j}^{\omega} = \frac{1}{\omega^2} \sum_{(s,t) \in \Omega_{i,j}^{\omega}} (\eta_{s,t})^2 = \frac{\sigma^2}{\omega^2} T_{i,j}^{\omega}.$$

But if u is an over-smoothed restored image, then the residual $z - Ku$ contains details, and in the detail regions we expect

$$S_{i,j}^{\omega} > \frac{1}{\omega^2} \sum_{(s,t) \in \Omega_{i,j}^{\omega}} (\eta_{s,t})^2 = \frac{\sigma^2}{\omega^2} T_{i,j}^{\omega}.$$

Therefore, we search for a bound B such that $S_{i,j}^{\omega} > B$ for some pixel (i, j) implies that in the residual there are some details left in the neighborhood of this pixel.

In order to reduce the effect due to the noise and to obtain a robust way (with respect to the selection of the window size ω) to distinguish detail regions from homogenous regions, similarly as in [9], we choose the bound as

$$(3.8) \qquad B^{\omega} := \frac{\sigma^2}{\omega^2} \left(\mathfrak{E}(T_{\max}) + \mathfrak{d}(T_{\max}) \right),$$

where $T_{\max} = \max_{(i,j) \in \Omega} T_{i,j}^{\omega}$, $\mathfrak{E}$ and $\mathfrak{d}$ represent the expected value and the standard deviation of a random variable, respectively.

We note that the calculations leading to the bound (3.8) depend on an independence assumption with respect to $T_{i,j}^{\omega}$. Due to the overlap of neighboring filter windows these random variables violate the independence assumption. While theoretically this would need to be accounted for and provides a future research direction, in numerical practice we find that our parameter choice rule yields results clearly superior to those obtained from scalar regularization parameters.

3.3. **Selection of λ in** (1.4). Based on the local constraints in (3.4), in the previous section we have defined the local variance estimator and have introduced a way to distinguish detail regions from the homogeneous regions. Now, we need to relate (3.4) to (1.4) in order to choose the spatially dependent parameter λ.

For this purpose we define the penalty problem

$$(3.9) \qquad \text{minimize} \quad \mathrm{TGV}_\alpha^2(u) + \frac{\gamma}{4} \int_\Omega \max(S(u) - \sigma^2, 0)^2 \, \mathrm{d}x \quad \text{over} \quad u \in \mathrm{BGV}^2(\Omega),$$

where $\gamma > 0$ is a penalty parameter. Based on the weak lower semicontinuity of the objective in $\mathrm{BGV}^k(\Omega)$, we extend Proposition 5 in [9] to the present case.

Theorem 4. *Assume that $K \in \mathcal{L}(L^2(\Omega))$ does not annihilate affine functions. Then, problem (3.9) admits a solution $u_\gamma \in \mathrm{BGV}^2(\Omega)$ for every $\gamma > 0$. Moreover, as $\gamma \to +\infty$ there exists a subsequence of $\{u_\gamma\}$ converging weakly in $L^2(\Omega)$ to a solution of (3.4).*

We calculate the Fréchet-derivative of the penalty term in (3.9) at u_γ in direction $v \in L^2(\Omega)$, and get

$$\int_\Omega \int_\Omega w(x,y)\gamma \max(S(u_\gamma) - \sigma^2, 0)(x) \, \mathrm{d}x [(Ku_\gamma - z)Kv](y) \, \mathrm{d}y$$

$$(3.10) \qquad\qquad\qquad = \int_\Omega \lambda_\gamma(y)[(Ku_\gamma - z)Kv](y) \, \mathrm{d}y$$

with

$$(3.11) \qquad \hat{\lambda}_\gamma := \gamma \max(S(u_\gamma) - \sigma^2, 0), \quad \lambda_\gamma := \int_\Omega w(x,y)\hat{\lambda}_\gamma \, \mathrm{d}x.$$

Since $w \geq 0$ and $\gamma > 0$, we have $\lambda_\gamma \geq 0$. We compare the right side of (3.10) with the Fréchet-derivative of the data-fitting term in (1.4) and notice that they are clearly related as $\lambda_\gamma = \lambda$. With the same TGV_α^2 term, we find that the first order optimality condition of (1.4) coincides with that of (3.9). Considering that (3.9) is the unconstrained format of the constrained minimization problem (3.4), we have established a relation between the total generalized variation model (1.4) and (3.4). As a consequence, we are able to choose the spatially dependent parameter λ in (1.4) in connection with the satisfaction of the local constraints in (3.4) and with the relations (3.11).

Our adaptation rule for λ in (1.4) operates as follows: Initially we assign a small positive value to λ_0, in order to obtain an over-smoothed restored image, u_0, and to keep most details in the image residual $z - Ku_0$. Then we restore the image iteratively by increasing λ according to the following rule:

$$(3.12\mathrm{a}) \qquad (\hat{\lambda}_{k+1})_{i,j} := \min\left((\hat{\lambda}_k)_{i,j} + \rho_k \left(\sqrt{(\tilde{S}_k^\omega)_{i,j}} - \sigma \right), L \right),$$

$$(3.12\mathrm{b}) \qquad (\lambda_{k+1})_{i,j} = \frac{1}{\omega^2} \sum_{(s,t) \in \Omega_{i,j}^\omega} (\hat{\lambda}_{k+1})_{s,t},$$

where $\rho_k = \|\hat{\lambda}_k\|_\infty / \sigma$ in order to keep $\hat{\lambda}_{k+1}$ at the same scale as $\hat{\lambda}_k$, and L is a large positive value to ensure uniform boundedness of $\{\hat{\lambda}_k\}$. Furthermore,

$$(\tilde{S}_k^\omega)_{i,j} := \begin{cases} (S_k^\omega)_{i,j} & \text{if } (S_k^\omega)_{i,j} \geq B^\omega, \\ \sigma^2 & \text{otherwise}, \end{cases}$$

and we notice that $\tilde{S}^\omega$ is always larger than or equals to σ^2, which helps to remove the maximum operator in (3.11). Since all images in our numerical tests have a dynamic range of $[0,1]$, for scaling purposes we replace $\tilde{S}_k^\omega - \sigma^2$ by $\sqrt{\tilde{S}_k^\omega} - \sigma$.

4. The SA-TGV Method

In this section we first introduce a speed-up of our restoration procedure and then state the spatially adapted total generalized variation (SA-TGV) method.

4.1. Hierarchical decomposition for TGV model with spatially dependent parameter.

In order to reduce the number of iterations and obtain an algorithm which is robust with respect to the initial value of λ, referring to [9] we intertwine a hierarchical decomposition method proposed in [15] (TNV-method) with the model (1.4). The resulting algorithm is as follows:

(i) Choose $\lambda_0 > 0$, $\lambda_0 \in L^\infty(\Omega)$ and compute

$$(4.1) \qquad u_0 := \operatorname*{arg\,min}_{u \in \mathrm{BGV}^2(\Omega)} \mathrm{TGV}_\alpha^2(u) + \frac{1}{2} \int_\Omega \lambda_0 (Ku - z)^2 \ \mathrm{d}x.$$

Set $\hat{u}_0 = u_0$.

(ii) For $j = 1, 2 \ldots$ set $\lambda_j = 2^j \lambda_0$ and $v_{j-1} = z - K\hat{u}_{j-1}$. Then compute

$$(4.2) \qquad u_j := \operatorname*{arg\,min}_{u \in \mathrm{BGV}^2(\Omega)} \mathrm{TGV}_\alpha^2(u) + \frac{1}{2} \int_\Omega \lambda_j (Ku - v_{j-1})^2 \ \mathrm{d}x, \quad \hat{u}_j := u_j + \hat{u}_{j-1}.$$

In addition, we are able to extend the following results in [15] to the total generalized variation regularization case. For the fomulation of these results, we introduce $\|u\|_*$, the dual of the seminorm $\mathrm{TGV}_\alpha^k(u)$, i.e.,

$$(4.3) \qquad \|u\|_* := \sup_{\mathrm{TGV}_\alpha^2(\varphi) \neq 0} \frac{\int_\Omega u\,\varphi \ \mathrm{d}x}{\mathrm{TGV}_\alpha^2(\varphi)}.$$

Then, the following theorem characterizes a solution of (1.4) in terms of $\|\cdot\|_*$ with a given spatially dependent parameter $\lambda \in L^\infty(\Omega)$ and $\lambda \geq 0$ a.e. in Ω.

Theorem 5.

(i) *u is a minimizer of* (1.4) *if and only if*

$$(4.4) \qquad \int_\Omega u\, K^*\lambda(z - Ku) \ \mathrm{d}x = \|K^*\lambda(z - Ku)\|_* \, \mathrm{TGV}_\alpha^2(u) = \mathrm{TGV}_\alpha^2(u).$$

(ii) $\|K^*\lambda z\|_* \leq 1$ *if and only if an affine function u which satisfies $Ku = 0$ is a minimizer of* (1.4).

(iii) *Assume that $1 < \|K^*\lambda z\|_* < +\infty$. Then u is a minimizer of* (1.4) *if and only if* $\int_\Omega u\, K^*\lambda(z - Ku) \ \mathrm{d}x = \|K^*\lambda(z - Ku)\|_* \, \mathrm{TGV}_\alpha^2(u)$ *and* $\|K^*\lambda(z - Ku)\|_* = 1$.

(iv) *If $\|K^*\lambda z\|_* > 1$, then the decomposition* (4.1)–(4.2) *yields* $\lambda_0 z = \sum_{j=0}^\infty \lambda_0 Ku_j$ *and further*

$$(4.5) \qquad \left\| K^*\lambda_0 \left(z - \sum_{j=0}^{k+1} Ku_j \right) \right\|_* = \frac{1}{2^{k+1}}.$$

Proof. (i) Since TGV_α^2 is a seminorm, similar as in [9] the result can be easily proved.

(ii) Let $\|K^*\lambda z\|_* \leq 1$. Based on (4.4), a minimizer u satisfies

$$\mathrm{TGV}_\alpha^2(u) = \int_\Omega uK^*\lambda(z - Ku) \ \mathrm{d}x$$

$$\leq \|K^*\lambda z\|_* \mathrm{TGV}_\alpha^2(u) - \int_\Omega (\sqrt{\lambda}Ku)^2 \ \mathrm{d}x$$

$$\leq \mathrm{TGV}^2(u) - \int_\Omega (\sqrt{\lambda}Ku)^2 \ \mathrm{d}x.$$

Thus, $Ku = 0$ and

$$\mathrm{TGV}_\alpha^2(u) + \frac{1}{2}\int_\Omega \lambda(Ku - z)^2 \ \mathrm{d}x = \mathrm{TGV}_\alpha^2(u) + \frac{1}{2}\int_\Omega \lambda z^2 \ \mathrm{d}x$$

$$\leq \mathrm{TGV}_\alpha^2(0) + \frac{1}{2}\int_\Omega \lambda(K0 - z)^2 \ \mathrm{d}x = \frac{1}{2}\int_\Omega \lambda z^2 \ \mathrm{d}x$$

as u is a minimizer. It follows that $\mathrm{TGV}_\alpha^2(u) = 0$ and hence u is an affine function which satisfies, in addition, $Ku = 0$.

Now, assume u is an affine function and satisfies $Ku = 0$, then we also have $\mathrm{TGV}_\alpha^2(u) = 0$. If u is a minimizer of (1.4), then following the same way utilized in [9] we get $\|K^*\lambda z\|_* \leq 1$.

(iii) Based on (i), we have to show that $\mathrm{TGV}_\alpha^2(u)$ does not vanish.

Let us suppose it does vanish. Then u is an affine function. For any affine function $\tilde{u}$ and any $\varphi \in \mathrm{BGV}^2(\Omega)$ with $\mathrm{TGV}_\alpha^2(\varphi) > 0$ we have

$$\|K^*\lambda z\|_* \geq \frac{\int_\Omega K^*\lambda z(\varphi + \tilde{u}) \ \mathrm{d}x}{\mathrm{TGV}_\alpha^2(\varphi + \tilde{u})} = \frac{\int_\Omega K^*\lambda z\varphi \ \mathrm{d}x}{\mathrm{TGV}_\alpha^2(\varphi)} + \frac{\int_\Omega K^*\lambda z\tilde{u} \ \mathrm{d}x}{\mathrm{TGV}_\alpha^2(\varphi)}$$

$$\geq \frac{\int_\Omega K^*\lambda z\tilde{u} \ \mathrm{d}x}{\mathrm{TGV}_\alpha^2(\varphi)} - \|K^*\lambda z\|_*.$$

Since $\|K^*\lambda z\|_* < +\infty$, we obtain that $\int_\Omega K^*\lambda z\tilde{u} \ \mathrm{d}x$ is bounded with a bound independent from any affine function $\tilde{u}$ and, consequently, has to vanish. Therefore, we have

$$\frac{1}{2}\int_\Omega \lambda(Ku - z)^2 \ \mathrm{d}x = \frac{1}{2}\int_\Omega \lambda z^2 \ \mathrm{d}x + \frac{1}{2}\int_\Omega (\sqrt{\lambda}Ku)^2 \ \mathrm{d}x,$$

which reaches its minimum when $Ku = 0$. From (ii) we conclude $\|K^*\lambda z\|_* \leq 1$, which yields a contradiction to our assumption.

(iv) Based on (iii), from (4.1) we have $\|K^*\lambda_0(z - Ku_0)\|_* = 1$, and from (4.2) we have $\|K^*\lambda_{k+1}(v_k - Ku_{k+1})\|_* = 1$. Since $v_k = z - K\hat{u}_k$, $\lambda_{k+1} = 2\lambda_k$, $\hat{u}_j = u_j + \hat{u}_{j-1}$ for $j = 1, \ldots, k$, and $\hat{u}_0 = u_0$, we obtain

$$1 = \|K^*\lambda_{k+1}(z - K\hat{u}_k - Ku_{k+1})\|_*$$

$$= 2\|K^*\lambda_k(z - K\hat{u}_{k-1} - Ku_k - Ku_{k+1})\|_*$$

$$= 2^{k+1}\left\|K^*\lambda_0\left(z - \sum_{j=0}^{k+1} Ku_j\right)\right\|_*$$

for $k = 0, 1, \cdots$.

$\square$

4.2. **SA-TGV Algorithm.** Utilizing the hierarchical decomposition method for solving problem (1.4), and combining it with the spatially dependent regularization parameter selection results in the following algorithm:

SA-TGV-Algorithm.

1: Initialize $\lambda_0 = \hat{\lambda}_0 \in \mathbb{R}_+^{m_1 \times m_2}$ and set $j := 0$.

2: If $j = 0$, solve the discrete version of the minimization problem in (4.1) and set $\hat{u}_0 = u_0$, else solve the discrete version of (4.2) with $v_{j-1} = z - K\hat{u}_{j-1}$.

3: Update $\hat{u}_j = u_j + \hat{u}_{j-1}$.

4: Based on $\hat{u}_j$, update

$$(\hat{\lambda}_j)_{i,j} := \zeta \min\left((\hat{\lambda}_{j-1})_{i,j} + \rho_j \left(\sqrt{(\tilde{S}_j^\omega)_{i,j}} - \sigma\right), L\right),$$

$$(\lambda_j)_{i,j} = \frac{1}{\omega^2} \sum_{(s,t)\in\Omega_{i,j}^\omega} (\hat{\lambda}_j)_{s,t},$$

where $\zeta = 2$, which comes from the TNV algorithm.

5: Stop and output the restored image $\hat{u}_j = \sum_{i=0}^j u_i$, or set $j := j+1$ and return to step 2.

In the SA-TGV method, we initially set λ as a relatively small positive constant, in order to keep noise and most details in the residual. Similar to the Bregman iteration proposed in [11], we stop the iterative procedure as soon as the residual $\|z - K\hat{u}_j\|_2$ drops below $\xi\sigma$, where $\xi > 1$ relates to the image size. For $m \to \infty$ we have $\xi \to 1$. To solve the minimization problems (4.1) and (4.2), one can extend any method for solving the model (1.3) to the spatially dependent parameter case. Possible solvers can be found in [4] and [10].

5. Numerical Results

In this section, we provide numerical results to compare the total generalized variation model with the total variation model and our spatially dependent regularization parameter with scalar parameters. Referring to [10], we use the primal-dual algorithm proposed in [7] to solve the minimization problems (4.1) and (4.2) in the SA-TGV method. Moreover, all parameters and the stopping rule in the primal-dual algorithm are chosen as suggested in [10] except that the ratio $\frac{\alpha_0}{\alpha_1}$ in the definition of the total generalized variation is set to $\frac{1}{2}$. In addition, similar to what is mentioned in [9] we use the window size $\omega = 11$ and a small initial $\lambda_0 = \mathbf{1}$. The quality of the restoration results are shown quantitatively by using the peak signal-to-noise ratio (PSNR) [3]. A higher PSNR normally indicates that the restoration is of higher quality.

Example 1. First, we study the behavior of the SA-TGV method for removing Gaussian noise, and compare it with two current methods: the TGV method [4], which uses the scalar parameter, and the SA-TV method [9], which combines the total variation regularization with the spatially dependent parameter. In Figure 1 we show the original 8-bit gray-level images "Barbara" (262-by-262) and "Mushroom" (481-by-321), the noisy images which are corrupted by 10% white Gaussian noise, and the restoration results. For fair comparisons, we use the same scheme of the primal-dual algorithm proposed in [7] to solve the minimization problems in the TGV method and the SA-TV method. Furthermore, we use the same initial value λ_0 and the window size ω in the SA-TV method.

FIGURE 1. *Example 1:* Comparison of 10% Gaussian noise removal. (a) The original images, (b) the noisy images, (c) results of the TGV method with $\lambda = 10$ (row 1: PSNR=25.69 dB; row 2: PSNR=27.39 dB), (d) results of the SA-TV method (row 1: PSNR=26.39 dB; row 2: PSNR=27.12 dB), (e) final values of λ by the SA-TV method, (f) results of the SA-TGV method (row 1: PSNR=26.83 dB; row 2: PSNR=27.53 dB), (g) final values of λ by the SA-TGV method.

Concerning the results obtained by the different methods, we observe that the SA-TGV method produces the best restorations, visually and quantitatively. Comparing with the images restored by the TGV method, with spatially dependent parameter the SA-TGV method
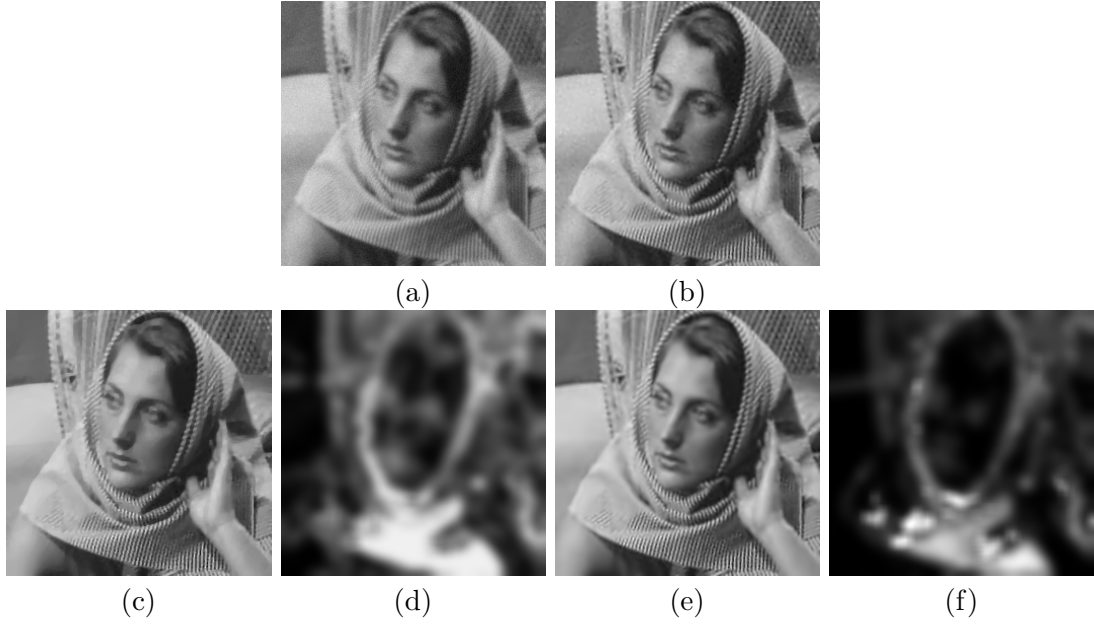
FIGURE 2. *Example 2:* Comparison of the blurry and noisy image restoration. (a) The degraded images, (b) result of the TGV method with $\lambda = 160$ (PSNR=27.69 dB), (c) result of the SA-TV method (PSNR=27.95 dB), (d) final values of λ by the SA-TV method, (e) result of the SA-TGV method (PSNR=28.20 dB), (f) final values of λ by the SA-TGV method.

suppresses noise successfully while preserving more details. For example, the features on the scarf of Barbara and the grass beside the mushroom are much better preserved, and the background in "Barbara" is much smoother without artifacts from the noise. Although the SA-TV method also utilizes the spatially dependent parameter, its denoising model uses the total variation regularization, which leads to the well-known staircasing effect; see, e.g., the face of Barbara and the mushroom. With the same initial value, both the SA-TV method and the SA-TGV method need 3 iterations to update λ. From the final values of λ shown in Figure 1, we note that λ is sufficiently large in detail regions in order to preserve the information inside that region accurately, and it is small in the homogeneous regions to remove noise sufficiently. Furthermore, comparing with the final values of λ obtained from the SA-TV method, the SA-TGV method distinguishes the detail regions more accurately.

Example 2. We consider the restoration of images that are simultaneously blurred and also corrupted by white Gaussian noise. In Figure 2, we show the degraded image and the restored results by the TGV method, the SA-TV method, and our SA-TGV method. Here, the blurring is due to a Gaussian convolution with a 9×9 window and a standard deviation of 1, and the noise level is 2%. In the TGV method, we set $\lambda = 160$, which provides the highest PSNR value of the restored image. Comparing with the results from the methods with spatially dependent parameter, we find that the restored image from the TGV method with large λ preserves most details, but also includes part of the noise; see the homogeneous regions in the background and the face of Barbara. Because of the usage of a spatially dependent parameter, the SA-TV method and the SA-TGV method reaching the stopping rule after 5 iterations provide much better results. Especially, the result obtained from the
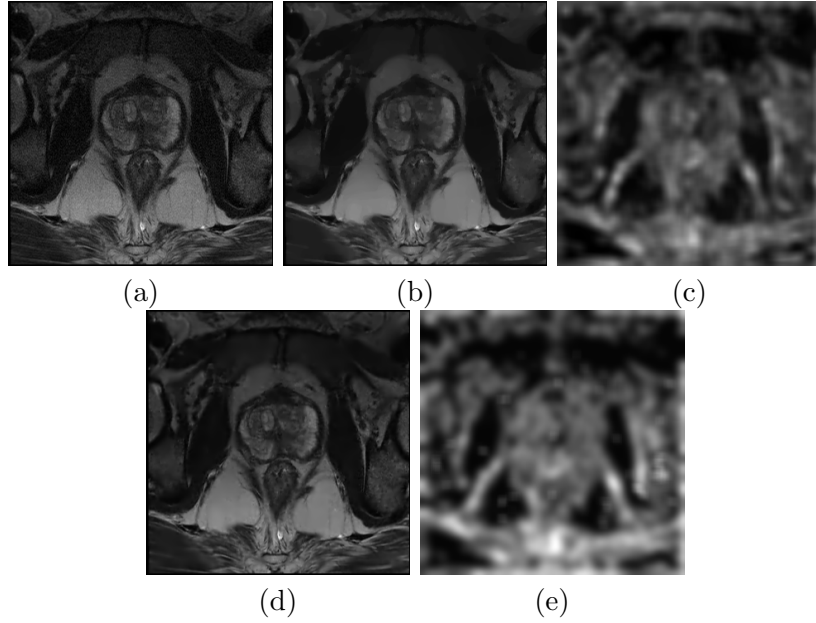
FIGURE 3. *Example 3:* Comparison of MR image restoration. (a) The noisy images, (b) the result of the SA-TV method, (c) the final λ by the SA-TV method, (d) the result of the SA-TGV method, (e) the final λ by the SA-TGV method.

SA-TGV method avoids the staircasing effect; see, e.g., the fingers and the nose of Barbara. Furthermore, concerning the final value of λ, our SA-TGV method differentiates well between detail regions and homogeneous regions. For example, in the regions of the arms, the hand and the face, the values of the final λ are much smaller. With respect to PSNR, we also find quantitatively that the SA-TGV method gives the best restoration result.

Example 3. In the last example, we compare the SA-TGV method with the SA-TV method for magnetic resonance (MR) image restoration. Figure 3(a) shows the noisy MR image, for which the noise level is estimated as $\sigma = 0.021$. Both methods remove most noise while preserving details efficiently. But in homogeneous regions the result by the SA-TV method is obviously corrupted by staircasing artifacts, because of the usage of total variation regularization. The SA-TGV method with the second order total generalized variation regularization provides better restoration results.

REFERENCES

[1] M. Bertalmio, V. Caselles, B. Rougé, and A. Solé. TV based image restoration with local constraints. *Journal of Scientific Computing*, 19:95–122, 2003.

[2] P. Blomgren, T. F. Chan, and P. Mulet. Extensions to total variation denoising. In *Proceedings of SPIE 1997*, San Diego, 1997.

[3] A. Bovik. *Handbook of Image and Video Processing*. Academic Press, 2000.

[4] K. Bredies, K. Kunisch, and T. Pock. Total generalized variation. *SIAM Journal on Imaging Sciences*, 3(3):492–526, 2010.

[5] K. Bredies and T. Valkonen. Inverse problems with second-order total generalized variation constraints. In *Proceedings of SampTA 2011 - 9th International Conference on Sampling Theory and Applications*, 2011.

[6] A. Chambolle and P-L. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76:167–188, 1997.

[7] A. Chambolle and T. Pock. A first order primal dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

[8] T. F. Chan, A. Marquina, and P. Mulet. Higher order total variation-based image restoration. *SIAM Journal on Scientific Computing*, 22:503–516, 2000.

[9] Y.Q. Dong, M. Hintermüller, and M. Rincon-Camacho. Automated regularization parameter selection in multi-scale total variation models for image restoration. *Journal of Mathematical Imaging and Vision*, 40:82–104, 2011.

[10] F. Knoll, K. Bredies, T. Pock, and R. Stollberger. Second order total generalized variation (TGV) for MRI. *Magnetic Resonance in Medicine*, 65(22):480–491, 2011.

[11] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total variation-based image restoration. *SIAM Multiscale Model. and Simu.*, 4:460–489, 2005.

[12] W. Ring. Structural properties of solutions to total variation regularization problems. *Mathematical Modelling and Numerical Analysis*, 34(4):799–810, 2000.

[13] L.I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[14] D. Strong and T. Chan. Spatially and scale adaptive total variation based regularization and anisotropic diffusion in image processing. Technical report, UCLA, 1996.

[15] E. Tadmor, S. Nezzar, and L. Vese. A multiscale image representation using hierarchical (BV, L^2) decompositions. *Multiscale Model. Simul.*, 2:554–579, 2004.

[16] Roger Temam. *Mathematical Problems in Plasticity*. Bordas, 1985.

[17] A. Tikhonov and V. Arsenin. *Solutions of Ill-Posed Problems*. Winston and Sons, Washington, D.C, 1977.

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