Automated Regularization Parameter Selection in Multi-Scale Total Variation Models for Image Restoration

**Yiqiu Dong · Michael Hintermüller ·
M. Monserrat Rincon-Camacho**

Abstract Multi-scale total variation models for image restoration are introduced. The models utilize a spatially dependent regularization parameter in order to enhance image regions containing details while still sufficiently smoothing homogeneous features. The fully automated adjustment strategy of the regularization parameter is based on local variance estimators. For robustness reasons, the decision on the acceptance or rejection of a local parameter value relies on a confidence interval technique based on the expected maximal local variance estimate. In order to improve the performance of the initial algorithm a generalized hierarchical decomposition of the restored image is used. The corresponding subproblems are solved by a superlinearly convergent algorithm based on Fenchel-duality and inexact semismooth Newton techniques. The paper ends by a report on numerical tests, a qualitative study of the proposed adjustment scheme and a comparison with popular total variation based restoration methods.

Y. Dong (✉)
Institute of Biomathematics and Biometry, Helmholtz Zentrum
München, Ingolstädter Landstraße 1, 85764 Neuherberg,
Germany
e-mail: yiqiu.dong@helmholtz-muenchen.de

M. Hintermüller
Department of Mathematics, Humboldt-University of Berlin,
Unter den Linden 6, 10099 Berlin, Germany
e-mail: michael.hintermueller@uni-graz.at

M. Hintermüller · M.M. Rincon-Camacho
START-Project "Interfaces and Free Boundaries" and SFB
"Mathematical Optimization and Applications in Biomedical
Science", Institute of Mathematics and Scientific Computing,
University of Graz, Heinrichsstrasse 36, 8010 Graz, Austria

M.M. Rincon-Camacho
e-mail: maria.rincon-camacho@uni-graz.at

1 Introduction

During acquisition and transmission images are often blurred and corrupted by Gaussian noise. In many applications, the deblurring and denoising of such images are fundamental for subsequent image processing operations, such as edge detection, segmentation, object recognition, and many more.

Suppose an image $\hat{u}$ is a real function defined on a bounded and piecewise smooth open subset Ω of $\mathbb{R}^2$ which, in applications, is typically only available in a degraded form z with

$$z = K\hat{u} + \eta. \tag{1.1}$$

Here, K is a linear and continuous blurring operator from $L^2(\Omega)$ to $L^2(\Omega)$, i.e., $K \in \mathcal{L}(L^2(\Omega))$, which we assume to be known. The quantity η represents white Gaussian noise with zero mean and standard deviation σ. The problem of restoring $\hat{u}$ from z with unknown η is known to be typically ill-posed [34]. Hence, stable reconstruction processes usually rely on regularization techniques which are based on prior information on $\hat{u}$.

In this direction and with the aim of preserving significant edges in images, in their seminal work [26] Rudin, Osher and Fatemi proposed total variation regularization for image restoration. In this approach (which we call the ROF-model in what follows), the recovery of the image $\hat{u}$ is based

on solving the constrained minimization problem

minimize $J(u) := \int_\Omega |Du|$ over $u \in BV(\Omega)$

subject to $\int_\Omega K u \, dx = \int_\Omega z \, dx,$ (1.2)

$\int_\Omega |Ku - z|^2 \, dx = \sigma^2 |\Omega|,$

where $BV(\Omega)$ denotes the space of functions of bounded variation, i.e. $u \in BV(\Omega)$ iff $u \in L^1(\Omega)$ and the BV-seminorm

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

is finite. Here, $(C_0^\infty(\Omega))^2$ is the space of vector-valued functions with compact support in Ω. The space $BV(\Omega)$ endowed with the norm $\|u\|_{BV(\Omega)} = \|u\|_{L^1(\Omega)} + \int_\Omega |Du|$ is a Banach space; see, e.g., [18]. Further, $|\Omega|$ denotes the volume or (two-dimensional) measure of the set Ω.

Usually, the ROF-model (1.2) is solved via the following optimization problem:

minimize $\int_\Omega |Du| + \frac{\lambda}{2} \int_\Omega |Ku - z|^2 \, dx$

over $u \in BV(\Omega)$ (1.3)

for a given $\lambda > 0$. Observe in (1.3) that the second constraint of (1.2) occurs in a penalized form. Moreover, assuming $K \cdot 1 = 1$ and $\int_\Omega |z|^2 \ge \sigma^2$ it is shown in [11] that the constraint $\int_\Omega K u \, dx = \int_\Omega z \, dx$ is automatically satisfied and that (1.2) and (1.3) are equivalent provided $\lambda \ge 0$ is chosen appropriately. In this case, λ represents the Lagrange multiplier associated with the corresponding constraint in (1.2). We also note that (1.3) can be equivalently expressed as

minimize $\frac{1}{2} \int_\Omega |Ku - z|^2 \, dx + \alpha \int_\Omega |Du|$

over $u \in BV(\Omega)$ (1.4)

where $\alpha = 1/\lambda > 0$ is a regularization parameter.

The properties of the TV-term $\int_\Omega |Du|$ are responsible for preserving edges during the reconstruction. This edge preservation ability is one of the reasons why the ROF-model is widely accepted as a reliable tool in image restoration. Over the years, various research efforts have been devoted to studying, solving and extending the ROF-model; see, e.g., [10–13, 20, 21, 23, 25, 30, 31] as well as the monograph [34] and the many references therein.

In the optimization problems (1.3) and (1.4), both the Lagrange multiplier λ and the regularization parameter α control the trade-off between a good fit of z and a smoothness requirement due to the total variation regularization. In general, images are comprised of multiple objects at different scales. This suggests that different values of λ and α localized at image features of different scales are desirable

to obtain better restoration results. Roughly speaking, for small features, large λ, or equivalently small α, leads to little smoothing and usually good detail preservation. On the other hand, for large features, small λ, or large α, leads to smoothing so that noise is removed considerably. For this reason and based on (1.3) or (1.4), in [3, 7, 29, 30] multi-scale total variation (MTV) models with a spatially varying choice of parameters were considered. The corresponding multi-scale versions of (1.3) and (1.4) read

minimize $\int_\Omega |Du| + \frac{1}{2} \int_\Omega \lambda(x) |Ku - z|^2(x) \, dx$

over $u \in BV(\Omega),$ (1.5)

and, for an appropriate function α,

minimize $\frac{1}{2} \int_\Omega |Ku - z|^2 dx + \int_\Omega \alpha(x) |Du|$

over $u \in BV(\Omega),$ (1.6)

respectively. In fact, in [30] the notion of a scale of an image feature (that is the ratio of the volume and the perimeter) is studied and a regularized gradient descent scheme for (1.6) is used. While [30] merely studies the influence of the scale on the choice of α, the subsequent work [29] proposes an update scheme for α. We note that the overall algorithm has to determine several reference parameters for α such as a scale recognition probe and a reference or threshold value and is primarily driven by geometric properties of image features. Moreover, the α-update rule neither depends on the noise statistics nor on local estimators for a robust adjustment scheme. The latter aspect is also true for the method proposed in [7], which uses a pre-segmented image and considers λ (in the framework of (1.5)) to be a piecewise constant function with the pieces defined by the segmentation output. In [7], for the solution of (1.5) with a piecewise constant λ a regularized gradient descent method is used. Then the λ-update rule follows an augmented Lagrangian scheme. Finally, in [3] the automated choice of λ is based on local constraints through local variance estimates. In a discrete setting, the method uses upper bounds of the expectancy of the maximal local squared residual. Thus, it relies on probabilistic arguments taking into account the noise statistics. The solution algorithm is finite dimensional, of proximal point type and converges at a linear rate with the latter depending on the proximal point regularization.

In this paper, we study (1.5) on the continuous, i.e. function space, level, and we propose a local variance estimator in order to decide, in a robust way, on the scales of the features contained in z. The decision on the acceptance or rejection of a local λ-value uses a confidence interval technique based on the expected maximal local variance estimate. The

latter is rigorously justified by the theory in [19]. Our results improve the ones, for instance, given in [3], where the upper bound on the expected maximal local image residual depends on $\ln(m) + \omega^2$, where $m \times m$ is the discrete image size and ω^2 is the number of pixels in the local window for generating the local squared residuals. The bound derived in [3] is typically too loose to yield accurate reconstructions. Instead, for the numerical results in [3] the heuristic bound $(1 + \delta)\sigma^2$ with $\delta \in [0, 1]$ is used and the choice of δ is empirical. This leads to infeasibility considerations for the associated minimization problem through the question of how many pixels satisfy the local constraints. This latter aspect is addressed in [15, Sect. 7.1]. Further, our λ-adjustment is fully automated and, thus, requires no user interaction. In order to accelerate the performance of the λ-update scheme we generalize the hierarchical decomposition approach proposed by [32, 33] to spatially dependent λ. The corresponding subproblems are solved by a superlinearly convergent algorithm based on Fenchel-duality and inexact semismooth Newton techniques. The latter extends earlier work in [21]. We note that, based on the relationship between (1.5) and (1.6), our subproblem solver can also be adapted easily to handle (1.6), as well. Further, we mention that besides our analysis of the localized constraints in the original function space context of (1.5), our solver strategy differs significantly from previous work such as, e.g., [3], where Uzawa's method is combined with the iterative scheme of [10] yielding a linearly convergent scheme only. In particular, the convergence of Uzawa's method is extremely slow with a rate of linear convergence rather close to 1. As noted above, our TV-solver converges locally superlinearly and when combined with the hierarchical decomposition scheme of [32] it converges extremely fast in practice.

The outline of the rest of the paper is as follows. In Sect. 2 we study the existence of a solution of a version of (1.2) with localized constraints and relate this problem to (1.5). Moreover, a first order optimality characterization is derived. In Sect. 3 we describe our new spatially adapted parameter selection in detail and discuss the statistics of the expected value of the maximal local residual estimates. The following Sect. 4 extends the algorithm in [33] to the case where λ is spatially dependent. Utilizing our new spatially adapted parameter selection rule, in Sect. 5 we introduce a primal-dual algorithm for solving the MTV-problem. Section 6 gives numerical results to demonstrate the performance of the new method. Moreover, we compare our method with several other popular methods. The numerical results indicate that our method has the potential to outperform the other approaches in both noise removal and detail preservation. Finally, conclusions are drawn in Sect. 7.

2 Spatially Adapted Regularization

Similar to [3, 17] we consider smoothed image residuals, which should ideally only contain noise after restoration, for extracting information on the scale that we then use for automatically adjusting λ in (1.5). Assume that w is a normalized filter, i.e. $w \in L^\infty(\Omega \times \Omega)$, $w \geq 0$ on $\Omega \times \Omega$ with

$$\int_\Omega \int_\Omega w(x, y) \, dy \, dx = 1 \quad \text{and}$$

$$\int_\Omega \int_\Omega w(x, y)\phi^2(y) \, dy \, dx \geq \epsilon \|\phi\|_{L^2(\Omega)}^2 \quad \forall \phi \in L^2(\Omega) \tag{2.1}$$

for some $\epsilon > 0$ (independent of ϕ). The second condition in (2.1) is required in the proof of the radial unboundedness result of Proposition 1. By $S(u)$ we denote the w-smoothed version of the residual which is

$$S(u)(x) := \int_\Omega w(x, y)(Ku - z)^2(y) \, dy. \tag{2.2}$$

Note that $S(u)(x)$ may be interpreted as a local variance. Observe that since $(Ku - z)^2 \in L^1(\Omega)$ and $w \in L^\infty(\Omega \times \Omega)$ we have $S(u) \in L^\infty(\Omega)$. Moreover, it can readily be shown that $S(\cdot)$ is continuous as a mapping from $L^2(\Omega)$ to $L^\infty(\Omega)$. The smoothed residual is now used to formulate a version of the ROF-model (1.2) with local constraints (instead of the original global constraint):

$$\begin{aligned} \text{minimize} \quad & J(u) \quad \text{over } u \in BV(\Omega) \\ \text{subject to} \quad & S(u) - \sigma^2 \leq 0 \quad \text{a.e. in } \Omega. \end{aligned} \tag{2.3}$$

Here and below 'a.e.' stands for 'almost everywhere'. For later use we define the feasible set

$$\mathcal{U} = \{u \in BV(\Omega) : S(u) \leq \sigma^2 \text{ a.e. in } \Omega\}. \tag{2.4}$$

It is straightforward to show that $\mathcal{U}$ is closed and convex.

2.1 Existence of a Solution

For the existence of a solution to (2.3) we start by adapting a result due to [2]. We provide a proof for the sake of keeping the paper self-contained.

Proposition 1 *Assume that K does not annihilate constant functions, i.e. $K\chi_\Omega \neq 0$, where $\chi_\Omega(x) = 1$ for $x \in \Omega$. Then $\|u\|_{BV} \to +\infty$ implies $\mathcal{J}(u) \to +\infty$ with*

$$\mathcal{J}(u) = J(u) + \int_\Omega \int_\Omega w(x, y)(Ku - z)^2(y) \, dy \, dx.$$

Proof Any $u \in BV(\Omega)$ can be decomposed according to

$$u = t + v \quad \text{with } t = \left(\frac{\int_\Omega u \, dx}{|\Omega|}\right)\chi_\Omega \text{ and } \int_\Omega v \, dx = 0. \tag{2.5}$$

Hence, we obtain

$$\|u\|_{BV} \le \|t\|_{BV} + \|v\|_{BV}$$

$$= \int_\Omega |t|\,dx + J(t) + \int_\Omega |v|\,dx + J(v)$$

$$\le \|t\|_{L^1(\Omega)} + \|v\|_{L^1(\Omega)} + J(v)$$

$$\le \|t\|_{L^1(\Omega)} + C_2 J(v)$$

for some $C_2 > 0$. Recalling that Ω is bounded with a piecewise smooth boundary, note here that we used the Sobolev inequality [18, p. 24] $\|v\|_{L^2(\Omega)} \le C_1 J(v)$, with $C_1 > 0$, to obtain the last inequality above. Since K does not annihilate constants, there exists $C_3 > 0$ independent of t such that $\|Kt\|_{L^2(\Omega)} \ge C_3\|t\|_{L^1(\Omega)}$. Then, by (2.1) we get

$$\mathcal{J}(u) \ge J(v) + \epsilon\|Kt + Kv - z\|_{L^2(\Omega)}^2$$

$$\ge J(v) + \epsilon\|Kt\|_{L^2(\Omega)}(\|Kt\|_{L^2(\Omega)}$$

$$- 2\|Kv - z\|_{L^2(\Omega)}).$$

Since $\|Kv - z\|_{L^2(\Omega)} \le \|K\|\|v\|_{L^2(\Omega)} + \|z\|_{L^2(\Omega)} \le C_1\|K\|J(v) + \|z\|_{L^2(\Omega)}$ we have

$$\mathcal{J}(u) \ge J(v) + \epsilon\|Kt\|_{L^2(\Omega)}\big(C_3\|t\|_{L^1(\Omega)}$$

$$- 2(\|K\|C_1 J(v) + \|z\|_{L^2(\Omega)})\big). \tag{2.6}$$

If $C_3\|t\|_{L^1(\Omega)} - 2(\|K\|C_1 J(v) + \|z\|_{L^2(\Omega)}) \ge 1$, then $\mathcal{J}(u) \ge J(v) + \epsilon\|Kt\|_{L^2(\Omega)}$ and

$$\|t\|_{L^1(\Omega)} \le \frac{1}{C_4}\mathcal{J}(u) \tag{2.7}$$

for $C_4 = \epsilon C_3 > 0$, and further

$$J(v) \le \mathcal{J}(u). \tag{2.8}$$

Then, (2.7) and (2.8) yield

$$\|u\|_{BV} \le \left(\frac{1}{C_4} + C_2\right)\mathcal{J}(u). \tag{2.9}$$

On the other hand, if $C_3\|t\|_{L^1(\Omega)} - 2(\|K\|C_1 J(v) + \|z\|_{L^2(\Omega)}) < 1$, then

$$\|t\|_{L^1(\Omega)} < \frac{1 + 2(\|K\|C_1 J(v) + \|z\|_{L^2(\Omega)})}{C_3}$$

and hence

$$\|u\|_{BV} - \frac{1 + 2\|z\|_{L^2(\Omega)}}{C_3} \le \left(\frac{2\|K\|C_1}{C_3} + C_2\right)\mathcal{J}(u). \tag{2.10}$$

Thus, (2.9) and (2.10) yield the assertion. □

Based on Proposition 1 it is immediate to argue existence of a solution to (2.3).

Theorem 2 *Assume that* $K \in \mathcal{L}(L^2(\Omega))$ *does not annihilate constant functions. Then problem* (2.3) *admits a solution.*

Proof We first note that J is bounded from below and choose an infimal sequence $\{u_n\} \subset \mathcal{U}$. Due to Proposition 1 $\{u_n\}$ is bounded in $BV(\Omega)$. Hence, there exists a subsequence $\{u_{n_k}\}$ which converges weakly in $L^2(\Omega)$ to some $\tilde{u} \in L^2(\Omega)$, and $\{Du_{n_k}\}$ converges weakly as a measure to $D\tilde{u}$ [6, p. 47]. By the weak lower semicontinuity of J we obtain that

$$J(\tilde{u}) \le \liminf_{k\to\infty} J(u_{n_k}) = \inf_{u\in\mathcal{U}} J(u). \tag{2.11}$$

Since K is a continuous linear operator, $\{Ku_{n_k}\}$ converges weakly to $K\tilde{u}$. Moreover, since $\mathcal{U}$ is closed and convex, we have $S(\tilde{u}) \le \sigma^2$ a.e. in Ω. □

Next we establish a uniqueness result. For this purpose we require the following property of the filter w.

Assumption 3 Let $u_1, u_2 \in BV(\Omega)$ denote two solutions of (2.3) with $u_1 \ne u_2$. If there exist $\delta > 0$ and $\Omega_\delta \subset \Omega$ with $|\Omega_\delta| > 0$ such that

$$\left(\frac{1}{2}K(u_1 + u_2) - z\right)^2$$

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

then there exists $\epsilon_\delta > 0$ such that

$$\int_\Omega w(x, y)\left(\frac{1}{2}K(u_1 + u_2) - z\right)^2(y)\,dy \le \sigma^2 - \epsilon_\delta$$

$$\text{for almost all } x \in \Omega. \tag{2.12}$$

We note that Assumption 3 is satisfied for the mean filter

$$w(x, y) = \begin{cases} \frac{1}{w_\epsilon^2} & \text{if } \|y - x\|_\infty \le \frac{\omega}{2}, \\ \epsilon_0 & \text{else,} \end{cases}$$

where $x \in \Omega$ is fixed, $\omega > 0$ sufficiently small is the essential width of the filter window, $0 < \epsilon_0 \ll \min(1, \frac{1}{w_\epsilon^2})$ and w_ϵ such that $\int_\Omega \int_\Omega w(x, y)\,dy\,dx = 1$. In this case we have $\epsilon_\delta = \epsilon_0 \delta |\Omega_\delta|$. It can also be shown that Assumption 3 holds true for the Gaussian filter.

The following uniqueness result generalizes a finite dimensional version due to [3].

Theorem 4 *Let the assumptions of Theorem 2 hold true and suppose* $K1 = 1$. *In addition we suppose that Assumption 3 is satisfied and that*

$$\inf_{c\in\mathbb{R}} \int_\Omega w(x, y)(c - z)^2(y)\,dy > \sigma^2 \quad \text{a.e. in } \Omega. \tag{2.13}$$

Then, for every solution $\tilde{u}$ *of* (2.3) $K\tilde{u}$ *has the same value.*

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

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

If the inequality holds as an equality a.e. in Ω, then $Ku_1 = Ku_2$ a.e. in Ω; otherwise there exist a $\delta > 0$ and a set $\Omega_\delta \subset \Omega$ of positive measure such that (2.12) holds true for a suitable $\epsilon_\delta > 0$. Define $u_s := s\bar{u}$ for $s \in [0, 1]$. Then, for s close to 1, we have $u_s \in \mathcal{U}$ and $J(u_s) = sJ(\bar{u}) < J(\bar{u})$ for all $s \in [0, 1)$, unless $J(\bar{u}) = 0$. This implies $\bar{u} \equiv \bar{c}$, and therefore $K\bar{u} = \bar{c}$ for some $\bar{c} \in \mathbb{R}$. This, however, is impossible since $\inf_{c \in \mathbb{R}} \int_\Omega w(x, y)(c - z)^2(y) dy > \sigma^2$ a.e. in Ω by assumption. Hence, $Ku_1 = Ku_2$ a.e. in Ω. □

Note that (2.13) is almost surely satisfied when z is the addition of some regular (non-constant) image and a white random field with variance σ^2.

2.2 First-order Optimality Characterization

We continue by characterizing a solution of (2.3) and relate the problem to (1.5). For this purpose we define the penalty problem

$$\text{minimize} \quad \mathcal{F}_\gamma(u) := J(u) + \frac{\gamma}{2}\int_\Omega \max(S(u) - \sigma^2, 0)^2 dx$$

$$\text{over } u \in BV(\Omega). \tag{2.14}$$

Here, $\gamma > 0$ denotes a penalty parameter.

Proposition 5 *Let the assumptions of Theorem 2 be satisfied. Then problem (2.14) admits a solution $u_\gamma \in BV(\Omega)$ for every $\gamma > 0$. Moreover, for $\gamma \to +\infty$ $\{u_\gamma\}$ converges along a subsequence weakly in $L^2(\Omega)$ to a solution of (2.3).*

Proof Note that due to the continuity and (pointwise) convexity of $S : L^2(\Omega) \to L^\infty(\Omega)$ as well as $\max(\cdot, 0) : L^2(\Omega) \to L^2(\Omega)$ and the weak lower semicontinuity of $J(u)$ according to [5], $\mathcal{F}_\gamma : BV(\Omega) \to \mathbb{R}$ is weakly lower semicontinuous. Let $\{u_n\} \subset BV(\Omega)$ denote an infimal sequence, and let $\tilde{u}$ be a solution of (2.3). Then, for all sufficiently large n we have $\mathcal{F}_\gamma(u_n) \leq \mathcal{F}_\gamma(\tilde{u}) + 1 = J(\tilde{u}) + 1$. Since $S(u) \geq 0$ a.e. in Ω for any $u \in BV(\Omega)$, there exists a constant C (independent of n and γ) such that $\|S(u_n)\|_{L^2(\Omega)} \leq C$. By Proposition 1 $\{u_n\}$ is bounded in $BV(\Omega)$. Now similar arguments as in the proof of Theorem 2 yield the existence of a solution $u_\gamma \in BV(\Omega)$.

Concerning the convergence result we first note that similarly to the first part of this proof one argues the boundedness of $\{u_\gamma\}$ in $BV(\Omega)$. Then by lower semicontinuity we have

$$J(\tilde{u}_\gamma) \leq \liminf_{\gamma \to +\infty} \mathcal{F}_\gamma(u_\gamma) \leq J(\tilde{u}) = \inf_{u \in \mathcal{U}} J(u),$$

where $\tilde{u}_\gamma$ is a weak limit of a subsequence of $\{u_\gamma\}$ in $L^2(\Omega)$ (which we still denote by $\{u_\gamma\}$). It remains to show that $\tilde{u}_\gamma \in \mathcal{U}$. For this we observe that for all $\gamma > 0$

$$\frac{\gamma}{2}\int_\Omega \max(S(u_\gamma) - \sigma^2, 0)^2 dx \leq J(\tilde{u}).$$

As a consequence, we obtain

$$\int_\Omega \max(S(u_\gamma) - \sigma^2, 0)^2 dx \to 0 \quad \text{as } \gamma \to \infty$$

and by the continuity of K, weak lower semicontinuity and Fatou's Lemma $S(\tilde{u}_\gamma) \leq \sigma^2$ a.e. in Ω. □

Observe that the arguments of the previous proof yield

$$\|\max(S(u_\gamma) - \sigma^2, 0)\|_{L^2(\Omega)} = \mathcal{O}(1/\sqrt{\gamma}), \tag{2.15}$$

where $\mathcal{O}(a_n)/a_n \to 0$ for a sequence $\{a_n\} \subset \mathbb{R}_+$ with $a_n \to 0$.

For our subsequent results, for arbitrarily fixed $\gamma > 0$ we define

$$\lambda_\gamma^\circ := \gamma \max(S(u_\gamma) - \sigma^2, 0), \tag{2.16}$$

$$\lambda_\gamma := \int_\Omega w(x, y)\lambda_\gamma^\circ(x) dx. \tag{2.17}$$

Note that λ_γ is related to the Fréchet-derivative of the penalty term in (2.14). This derivative at u_γ applied to some direction $v \in L^2(\Omega)$ is given by

$$\int_\Omega [2(Ku_\gamma - z)Kv](y)\int_\Omega \gamma \max(S(u_\gamma)$$
$$- \sigma^2, 0)(x)w(x, y) dx dy$$
$$= \int_\Omega [2(Ku_\gamma - z)Kv](y)\lambda_\gamma(y) dy.$$

Here we use the Fréchet-derivative $S'(\cdot)$ of $S(\cdot) : L^2(\Omega) \to L^2(\Omega)$ with its action on $v \in L^2(\Omega)$ given by

$$S'(u)v = 2\int_\Omega w(x, y)[(Ku - z)Kv](y) dy.$$

Now we are ready to state the first-order optimality characterization of a solution to (2.3).

Theorem 6 *Let the assumptions of Theorem 2 hold, and let $\tilde{u}$ denote a weak limit point of $\{u_{\gamma_n}\}$ as $\gamma_n \to +\infty$. Moreover, we assume that $\|Ku_{\gamma_n}\|_{L^2(\Omega)} \to \|K\tilde{u}\|_{L^2(\Omega)}$ as $\gamma_n \to +\infty$ and that there exists $C > 0$ such that $\gamma_n\|\max(S(u_{\gamma_n}) - \sigma^2, 0)\|_{L^1(\Omega)} \leq C$ for all $n \in \mathbb{N}$. Then there exist $\tilde{\lambda} \in L^\infty(\Omega)$, a bounded Borel measure $\tilde{\lambda}^\circ$ and a subsequence $\{\gamma_{n_k}\}$ such that the following properties hold true:*

(i) $\int_\Omega \lambda_{\gamma_{n_k}} f \, dx \to \int_\Omega \tilde\lambda f \, dx$ for all $f \in L^1(\Omega)$ and $\tilde\lambda \geq 0$ a.e. in Ω.

(ii) *There exists $j(\tilde u) \in \partial J(\tilde u)$ such that*

$$\langle j(\tilde u), v \rangle_{BV(\Omega)^*, BV(\Omega)} + 2 \int_\Omega (K^*\tilde\lambda(K\tilde u - z))v \, dx = 0$$

for all $v \in BV(\Omega)$.

(iii) $\int_\Omega \varphi \lambda_{\gamma_{n_k}}^\circ \, dx \to \int_\Omega \varphi d\tilde\lambda^\circ$ for all $\varphi \in C(\bar\Omega)$, $\tilde\lambda^\circ \geq 0$ and $\int_\Omega \lambda_{\gamma_n}^\circ (S(u_{\gamma_n}) - \sigma^2) \, dx \to 0$.

Proof We start by proving (i). First note that due to the properties of S, (2.17) and $w \in L^\infty(\Omega \times \Omega)$, we have $\lambda_{\gamma_n} \in L^\infty(\Omega)$. Thus, $\int_\Omega \lambda_{\gamma_n} f \, dx$ is well-defined for all $f \in L^1(\Omega)$. Under our assumptions there exists a constant $C' > 0$ independent of γ_n such that

$$\|\lambda_{\gamma_n}\|_{L^\infty(\Omega)} \leq \gamma_n \|w\|_{L^\infty(\Omega \times \Omega)} \| \max(S(u_{\gamma_n}) - \sigma^2, 0)\|_{L^1(\Omega)}$$
$$\leq C'.$$

Now the first part of (i) follows from the weak* sequential compactness of the closed unit ball in $L^\infty(\Omega)$ (according to the Banach-Alaoglu theorem; see [27, p. 66]). The non-negativity is an immediate consequence of the definition of λ_γ.

Concerning (ii) we recall that from the proof of Proposition 5 we get the boundedness of $\{u_{\gamma_n}\}$ in $BV(\Omega)$. From this and the continuity of $J(\cdot)$ at u_{γ_n}, we infer the uniform (w.r.t. γ_n) boundedness of $\partial J(u_{\gamma_n})$. Now note that the first order (necessary and sufficient) optimality condition for (2.14) is given by

$$0 \in \partial J(u_{\gamma_n}) + \gamma_n S'(u_{\gamma_n})^* \max(S(u_{\gamma_n}) - \sigma^2, 0), \qquad (2.18)$$

where $S'(\cdot)^*$ denotes the adjoint operator of $S'(\cdot)$. From the boundedness of $\{u_{\gamma_n}\}$ in $BV(\Omega)$ and of $\{\partial J(u_{\gamma_n})\}$ in $BV(\Omega)^*$ we infer

$$\gamma_n \|S'(u_{\gamma_n})^* \max(S(u_{\gamma_n}) - \sigma^2, 0)\|_{BV(\Omega)^*} \leq C'' \qquad (2.19)$$

for some constant $C'' > 0$ independent of γ_n. Moreover, for $v \in BV(\Omega)$ we have

$$\frac{\gamma_{n_k}}{2} \langle S'(u_{\gamma_{n_k}})^* \max(S(u_{\gamma_{n_k}}) - \sigma^2, 0), v \rangle_{BV(\Omega)^*, BV(\Omega)}$$

$$= \int_\Omega (K u_{\gamma_{n_k}} - z)(Kv)\lambda_{\gamma_{n_k}} \, dy$$

$$= \int_\Omega (K u_{\gamma_{n_k}} - K\tilde u)(Kv)\lambda_{\gamma_{n_k}} \, dy$$

$$\quad + \int_\Omega (K\tilde u - z)(Kv)(\lambda_{\gamma_{n_k}} - \tilde\lambda) \, dy$$

$$\quad + \int_\Omega (K\tilde u - z)(Kv)\tilde\lambda \, dy$$

$$\to \int_\Omega (K\tilde u - z)(Kv)\tilde\lambda \, dy \quad \text{as } k \to +\infty,$$

where, without loss of generality, $\tilde u \in BV(\Omega)$ denotes the weak limit of $\{u_{\gamma_{n_k}}\}$ in $L^2(\Omega)$. This proves (ii).

Finally, from the boundedness assumption of this theorem, (2.16) and [5, Cor. 2.4.3] we obtain the first result in (iii). The non-negativity of $\tilde\lambda^\circ$ is an immediate consequence of the definition of λ_γ°. Then, based on

$$\left| \int_\Omega \lambda_{\gamma_n}^\circ \big(S(u_{\gamma_n}) - \sigma^2\big) \, dx \right| = \gamma \, \| \max(S(u_{\gamma_n}) - \sigma^2, 0)\|_{L^2(\Omega)}^2,$$

the third assertion in (iii), $\int_\Omega \lambda_{\gamma_n}^\circ (S(u_{\gamma_n}) - \sigma^2) \, dx \to 0$, follows from (2.15). $\qquad \square$

We note that if (2.15) holds true with $\mathcal{O}(1/\sqrt\gamma)$ replaced by $\mathcal{O}(1/\gamma)$, then $\{\lambda_{\gamma_n}^\circ\}$, with $\{\gamma_n\}$ as in Theorem 6, is uniformly bounded in $L^2(\Omega)$. As a consequence, $\tilde\lambda^\circ \in L^2(\Omega)$ is the weak limit of a subsequence $\{\lambda_{\gamma_{n_k}}^\circ\}$. In this case the system of Theorem 6(iii) becomes

$$\tilde\lambda^\circ \geq 0 \quad \text{a.e. in } \Omega, \qquad S(\tilde u) \leq \sigma^2 \quad \text{a.e. in } \Omega,$$

$$\lim_{n \to \infty} \int_\Omega \lambda_{\gamma_n}^\circ (S(u_{\gamma_n}) - \sigma^2) \, dx = 0.$$

If the last relation above holds as $\int_\Omega \tilde\lambda^\circ (S(\tilde u) - \sigma^2) \, dx = 0$, then we may equivalently write

$$\tilde\lambda^\circ \geq 0 \quad \text{a.e. in } \Omega, \qquad \tilde\lambda^\circ = \tilde\lambda^\circ + \rho \max(S(\tilde u) - \sigma^2, 0)$$
$$(2.20)$$

for arbitrary and fixed $\rho > 0$.

Setting $\lambda = 2\tilde\lambda$ in (1.5) we find that the first order optimality condition for (1.5) coincides with Theorem 6(ii). This relates the constrained problem (2.3) and the unconstrained problem (1.5) formally. Note that for the existence proof for (1.5) we need $\tilde\lambda \geq \tilde\epsilon > 0$ a.e. in Ω. A rigorous investigation when (2.3) admits the existence of such a multiplier $\tilde\lambda$, however, goes beyond the scope of the present paper.

3 Spatial Adaptation by Local Variance Estimators

We suppose that the variance σ^2 of the Gaussian noise is at our disposal. In practice the variance can be estimated, e.g., from homogeneous parts; see [4, 16] for various estimation techniques. While the solution of the ROF-model satisfies the global constraint

$$\int_\Omega |Ku - z|^2 \, dx = \sigma^2 |\Omega|, \qquad (3.1)$$

Fig. 1 Local variance estimator S^ω with different window sizes: (**a**) Original image, (**b**) Noisy image ($\sigma = 0.2$, K = identity matrix), (**c**) Restored image obtained from solving (1.3) with $\lambda = 2.4$, (**d**) Residual, (**e**) S^5, (**f**) S^7, (**g**) S^9
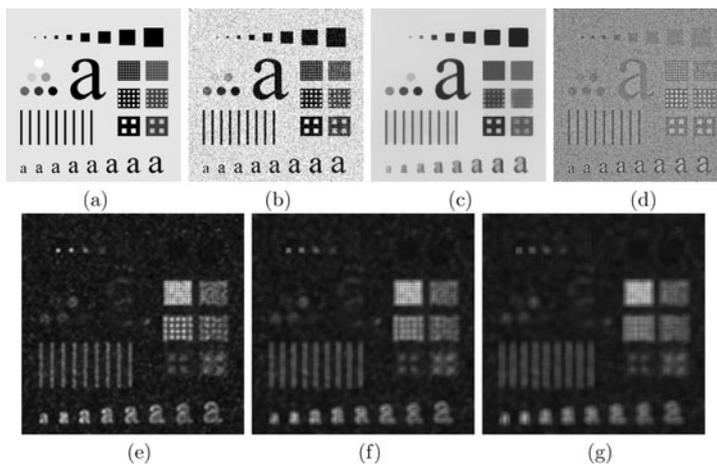
(1.5) represents a localized version by allowing $\lambda = \lambda(x)$. In order to enhance image details while preserving homogeneous regions, the choice of λ must be based on local image features. Hence, we search for a reconstruction where the variance of the residual is close to the noise variance in both the detail regions and the homogeneous parts. In order to achieve this goal we introduce local variance estimators for an automated adaptive choice of λ. Our adjustment rule makes use of the constraint in (2.3).

3.1 Local Variance Estimator

Our subsequent considerations are exemplarily based on the mean filter introduced earlier. We mention, however, that the Wiener filter employed in [17] or the Gaussian filter of the non-local means approach [9] may be used as well. Moreover, from now on we proceed in discrete terms, but, for the sake of simplicity, we keep the notations from the continuous context. We assume that the discrete image domain Ω contains $m \times m$ pixels. Let $r = z - Ku$ denote the discrete residual image with r, z, $u \in \mathbb{R}^{m^2}$ and $K \in \mathbb{R}^{m^2 \times m^2}$. For convenience, for the remainder of this section we reshape r, z, and Ku as $m \times m$-matrices. We also note that in the discrete mean filter we may choose $\epsilon_0 = 0$ and ω is an odd integer as the discrete version of Assumption 3 still holds true in this case. Further, let $\Omega_{i,j}^{\omega}$ denote the set of pixel-coordinates in a ω-by-ω window centered at (i, j) (with a symmetric extension at the boundary), i.e.

$$\Omega_{i,j}^{\omega} = \left\{ (s+i, t+j) : -\frac{\omega-1}{2} \leq s, t \leq \frac{\omega-1}{2} \right\}.$$

Then we apply the mean filter to the residual and obtain

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

$$= \frac{1}{\omega^2} \sum_{(s,t) \in \Omega_{i,j}^{\omega}} \left(r_{s,t} \right)^2. \tag{3.2}$$

Based on the current estimate λ and the pertinent reconstruction u, S^ω is a local variance estimator, which allows us to decide on the amount of details contained in the window around (i, j). For illustration purposes, Fig. 1 depicts S^ω for $\omega = 5, 7, 9$, respectively; see the plots (e)–(g). The corresponding reconstruction u comes from solving the discrete version of (1.3) with $\lambda = 2.4$ by the primal-dual algorithm introduced in [21]. With this small λ, the restored image u is over-smoothed, and the residual contains noise and details. Observe that S^ω is typically large (indicated in light gray) in image regions which contain fine scale details. Moreover we find that for fixed contrast, S^ω is the larger the finer (smaller) the scale is. In order to distinguish a region containing just noise from a region containing details we propose to employ the confidence interval technique well-known from statistics [22, 24].

3.2 Upper Bound for the Local Variance

In the discrete setting, η (see (1.1)) can be regarded as an array of independent normally distributed random variables with zero mean and variance σ^2. Then 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 $u = \hat{u}$ satisfies $\eta = z - K\hat{u}$, then

$$S_{i,j}^{\omega} = \frac{1}{\omega^2} \sum_{(s,t) \in \Omega_{i,j}^{\omega}} (z_{s,t} - (K\hat{u})_{s,t})^2$$

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

If u is an over-smoothed restored image, then the residual $z - Ku$ contains details, and we expect

$$
\begin{aligned}
S_{i,j}^{\omega} &= \frac{1}{\omega^2} \sum_{(s,t)\in\Omega_{i,j}^{\omega}} (z_{s,t} - (Ku)_{s,t})^2 \\
&> \frac{1}{\omega^2} \sum_{(s,t)\in\Omega_{i,j}^{\omega}} (\eta_{s,t})^2 = \frac{\sigma^2}{\omega^2} T_{i,j}^{\omega}.
\end{aligned}
$$

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. For ease of notation, below we write $T_k^{\omega} := T_{i,j}^{\omega}$ with $k = i + (m-1)j$ for $i, j = 1, \dots, m$.

Given $m \times m$, the total number of pixels in the image, we propose to consider the expected maximum of the m^2 random variables $\frac{\sigma^2}{\omega^2} T_k^{\omega}$, $k = 1, \dots, m^2$, where—as before—each T_k^{ω} has the χ^2-distribution with ω^2 degrees of freedom. The bound B depends on the size of the image ($m \times m$) and on the size of the window ($\omega \times \omega$). Thus, we write $B^{\omega,m} = B(\omega, m)$ with

$$
B^{\omega,m} := \frac{\sigma^2}{\omega^2} \mathfrak{E}\left(\max_{k=1,\dots,m^2} T_k^{\omega} \right), \tag{3.3}
$$

where $\mathfrak{E}$ represents the expected value of a random variable.

3.2.1 Distribution of the Maximum of N Random Variables

In order to compute the expected value of the maximum of the $N = m^2$ random variables T_k^{ω}, $k = 1, \dots, N$, we use [19], disregarding certain dependencies. For the moment we drop the superscript ω. Let $\mathfrak{f}$ be the χ^2-distribution with ω^2 degrees of freedom, and let $\mathfrak{F}$ denote its cumulative distribution function, i.e.

$$
\mathfrak{F}(T) = \int_{-\infty}^{T} \mathfrak{f}(z) \, dz. \tag{3.4}
$$

The maximum value of N observations distributed along $\mathfrak{f}$ is denoted by $T_{\max}$. Our goal is to describe the distribution $\mathfrak{f}_{\max}$ of this maximum value. According to [19] this distribution is given by

$$
\mathfrak{f}_{\max}[y(T_{\max})] = N\mathfrak{f}(T_{\mathrm{dom}})e^{-y(T_{\max})-e^{-y(T_{\max})}}, \tag{3.5}
$$

where $y(T) = N\mathfrak{f}(T_{\mathrm{dom}})(T - T_{\mathrm{dom}})$ is the standardization of the variable T. Here, T_{dom} is the so-called dominant value, which is such that

$$
\mathfrak{F}(T_{\mathrm{dom}}) = 1 - \frac{1}{N}. \tag{3.6}
$$

From (3.5), the cumulative distribution function $\mathfrak{F}_{\max}$ of $T_{\max}$ reads

$$
\mathfrak{F}_{\max}(T) = P(T_{\max} \leq T) = e^{-e^{-y(T)}}. \tag{3.7}
$$

The expected value and the standard deviation of the standardized variable $y(T_{\max})$ are

$$
\mathfrak{E}(y(T_{\max})) = \kappa \quad \text{and} \quad \mathfrak{d}(y(T_{\max})) = \frac{\pi}{\sqrt{6}}, \tag{3.8}
$$

respectively, where $\kappa = 0.577215$ is the Euler-Mascheroni constant; for further details see [19]. According to the transformation of $T_{\max}$ we have that its expected value and its standard deviation are

$$
\mathfrak{E}(T_{\max}) = T_{\mathrm{dom}} + \frac{\kappa}{\beta_{\max}} \quad \text{and}
$$
$$
\mathfrak{d}(T_{\max}) = \frac{\pi}{\beta_{\max}\sqrt{6}} \quad \text{with } \beta_{\max} = N\mathfrak{f}_{\max}(T_{\mathrm{dom}}). \tag{3.9}
$$

3.2.2 Confidence Interval

We observe that the size of the image influences the expected maximum value given by (3.9). In fact, if $N_1 \leq N_2$, then

$$
\mathfrak{E}(T_{\max}^1) \leq \mathfrak{E}(T_{\max}^2),
$$

where $T_{\max}^i$ corresponds to the maximum of N_i observations, $i = 1, 2$. In view of our above findings, the following two choices for the bound $B^{\omega,m}$ are natural: Either we take the expected value of the random variables T_k^{ω}, $k = 1, \dots, N$, which corresponds to (3.3) and our earlier discussion, or we re-define $B^{\omega,m}$ by adding the corresponding standard deviation $\mathfrak{d}(T_{\max})$ to the first choice. The latter option is taken in (3.12) below.

The confidence level of these bounds is given by the cumulative distribution (3.7). The probability that the maximum value is below or equal to $\mathfrak{E}(T_{\max})$ is

$$
P(T_{\max} \leq \mathfrak{E}(T_{\max})) = e^{-e^{-y(\mathfrak{E}(T_{\max}))}} = e^{-e^{-\kappa}} = 0.57037 \tag{3.10}
$$

and that the maximum value is not larger than $\mathfrak{E}(T_{\max}) + \mathfrak{d}(T_{\max})$ is

$$
\begin{aligned}
P(T_{\max} \leq \mathfrak{E}(T_{\max}) + \mathfrak{d}(T_{\max})) &= e^{-e^{-y(\mathfrak{E}(T_{\max})+\mathfrak{d}(T_{\max}))}} \\
&= e^{-e^{-\kappa-\frac{\pi}{\sqrt{6}}}} = 0.85580,
\end{aligned} \tag{3.11}
$$

since

$$
\begin{aligned}
y(\mathfrak{E}(T_{\max}) + \mathfrak{d}(T_{\max})) &= N\,\mathfrak{f}(T_{\mathrm{dom}})\left(T_{\mathrm{dom}} + \frac{\kappa}{N\,\mathfrak{f}(T_{\mathrm{dom}})} \right.\\
&\quad \left. + \frac{\pi}{N\,\mathfrak{f}(T_{\mathrm{dom}})\sqrt{6}} - T_{\mathrm{dom}} \right)\\
&= \kappa + \frac{\pi}{\sqrt{6}}.
\end{aligned}
$$
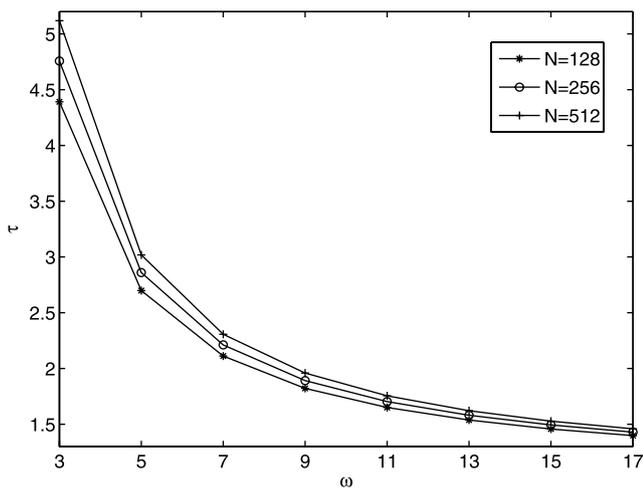
Fig. 2 Different values of τ for different window sizes ω and image sizes N

Observe that (3.10) specifies the probability that all of the smoothed image residuals T_k^ω, $k = 1, \ldots, m^2$, satisfy the constraints. A similar reasoning holds true for (3.11), which yields a higher probability as the upper bound is relaxed. Then, even if there is only noise left in the residual, using the first upper bound the constraints are satisfied by all T_k^ω only with a probability of 0.57. In the reconstruction process, this leads to difficulties in distinguishing whether the violation of constraints is due to noise or image details still contained in the local residuals. Therefore, subsequently we choose the second bound described in (3.11) above yielding

$$B^{\omega,m} := \frac{\sigma^2}{\omega^2} \left(\mathfrak{E}(T_{\max}) + \mathfrak{d}(T_{\max}) \right). \tag{3.12}$$

3.2.3 Window Size

The previous results are valid for any window size ω. We write

$$B^{\omega,m} = \tau \sigma^2, \tag{3.13}$$

then, from (3.12), we obtain

$$\tau = \frac{1}{\omega^2} \left(\mathfrak{E}(T_{\max}) + \mathfrak{d}(T_{\max}) \right). \tag{3.14}$$

In Fig. 2 we show the dependence of τ on the window size ω and the image size N. We note that τ is always larger than 1, i.e. $B^{\omega,m}$ is always larger than σ^2. Our choice of ω should avoid the following unfavorable cases: (i) $B^{\omega,m} \gg \sigma^2$. In this case there is a relatively large chance that regions with details are not recognized properly. (ii) $B^{\omega,m} \approx \sigma^2$. In this case there is a likeliness that a region is identified which seemingly contains details although this is not true. In addition, upon inspection of the graphs we find that the larger the window size is, the tighter the bound on the local variance

estimator becomes. Moreover, in view of (2.17), increasing the window size reduces the "sharpness" of λ. Conversely, a small window size yields a rather large bound and yields a λ, which better reflects the location of image details. Further, from Fig. 2 we observe that the upper bound depends on the window size relatively to the image size. In our numerical experiments in Sect. 6 we also study the effect of the window size on the reconstruction quality; see, e.g., Fig. 11.

3.3 Selection of λ

We use the confidence interval technique for S^ω to reduce the effect of noise on the local variance estimators for detecting image details.

Recall that $S_{i,j}^\omega$ represents the mean value of the squared residual in a given window $\Omega_{i,j}^\omega$. Ideally, the residual should contain noise only. In this case

$$S_{i,j}^\omega \in \left[0, B^{\omega,m}\right), \tag{3.15}$$

where $B^{\omega,m}$ is given by (3.12). On the other hand, if (3.15) is not satisfied, we suppose that this is due to image details contained in the residual image in $\Omega_{i,j}^\omega$. This motivates the introduction of the following modified local variance estimator $\tilde{S}^\omega$ defined by

$$\tilde{S}_{i,j}^\omega := \begin{cases} S_{i,j}^\omega & \text{if } S_{i,j}^\omega \geq B^{\omega,m}, \\ \sigma^2 & \text{otherwise.} \end{cases} \tag{3.16}$$

In the second row of Fig. 3, the quantity $\tilde{S}^\omega$ is depicted, which is based on the noisy image and the restored image shown in Fig. 1 (b) and (c), respectively. In order to understand the performance of $B^{\omega,m}$ better according to (3.13) and (3.14) we compare this bound with the upper bound proposed in the recent work [15]. In that paper, the authors consider local constraints similar to ours in (2.3). We should mention that in [15] a bilateral bound is determined based on the constraint which is used for defining a stopping condition rather than for updating the regularization parameter as in our case. Here, we only choose the upper bound of [15] for a comparison with our bound in order to show that for distinguishing the detail regions from other features in the residual the bound $B^{\omega,m}$ as defined in (3.12) is more suitable. Referring to [15], the value of $(1 + \alpha)0.64\sigma^2$, with $\alpha > 0$ such that $(1 + \alpha)0.64 < 1$, is used as an upper bound for localized variance estimates (instead of σ^2). It is motivated by the observation that in order to avoid loss of texture in images one should accept a higher total variation. The latter can be achieved by employing a lower bound to the true variance, thus yielding $(1 + \alpha)0.64 < 1$. Based on [15], there exists the relation $P(S^\omega \leq (1 + \alpha)0.64\sigma^2) = P(\chi^2(\omega^2) \leq (1 + \alpha)\omega^2) = \gamma(\frac{\omega^2}{2}, \frac{(1+\alpha)\omega^2}{2})/\Gamma(\frac{\omega^2}{2})$, where $\gamma(a, x) = \int_0^x t^{a-1}e^{-t}\,dt$ is the incomplete gamma function

(a)   (b)   (c)
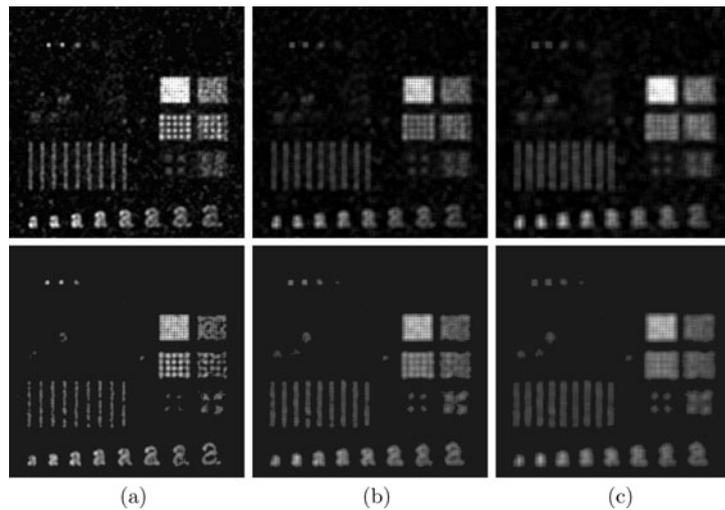
and $\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}\,dt$. It provides the expected fraction of the total number of image pixels satisfying the local constraints. In order to compare with our upper bound, we set $P(S^\omega \leq (1+\alpha)0.64\sigma^2) = 0.8$, and calculate the value α with $\omega = 5, 7, 9$, respectively. Observe that a probability of 0.8 for the window sizes under investigation keeps α in the range where $(1+\alpha)0.64 < 1$. We note that an upper bound which is too tight—such as the one in [15] which is even less than σ^2—produces a modified local variance estimator which appears strongly influenced by noise; see the first row in Fig. 3 and compare it with the second row in Fig. 3. In view of our update strategy of λ which we discuss below, the bound of [15] would yield too large local λ-values and, hence, too little regularization in homogeneous features. In such image regions, the noise removal would be adversely affected. Our upper bound is adjusted automatically based on the window size, the image size and the statistical properties of the maximum value of the variance estimators. Thus, compared to S^ω (see Fig. 1), the influence of noise is significantly reduced and $\tilde{S}^\omega_{i,j}$ is more adaptive, i.e., it is larger than the noise level primarily in image regions containing details; compare the second row of Fig. 3.

For adapting λ algorithmically we proceed as follows. Initially we assign a small positive value to λ, in order to obtain an over-smoothed restored image and to keep most details in the residual. Then we restore the image iteratively by increasing λ according to the following rule: Let $\tilde{\lambda}_k$ denote a given discrete approximation of $\tilde{\lambda}^\circ$ in (2.20). Then we set

$$(\tilde{\lambda}_{k+1})_{i,j} := (\tilde{\lambda}_k)_{i,j} + \rho\max((\tilde{S}^\omega_k)_{i,j} - \sigma^2, 0)$$
$$= (\tilde{\lambda}_k)_{i,j} + \rho((\tilde{S}^\omega_k)_{i,j} - \sigma^2), \qquad (3.17a)$$

$$(\lambda_{k+1})_{i,j} = \frac{1}{\omega^2}\sum_{(s,t)\in\Omega^\omega_{i,j}}(\tilde{\lambda}_{k+1})_{s,t}, \qquad (3.17b)$$

where $\rho > 0$. Observe that (3.17a) is motivated by a discrete version of (2.20) and (3.17b) by (2.17). Here and below, $\tilde{S}^\omega_k$ is the modified local variance estimator obtained from u_k. Based on the definition in (3.16), $\tilde{S}^\omega$ is always larger than σ^2, which leads to the rightmost consequence in (3.17a). We set $\rho = \rho_k = ||\tilde{\lambda}_k||_\infty/\sigma$ in order to keep the new $\tilde{\lambda}_{k+1}$ at the same scale as $\tilde{\lambda}_k$.

Based on the iterative update of λ in (3.17), we have the following basic multi-scale total variation algorithm.

Basic MTV-Algorithm.
1: Initialize $\lambda_0 := \tilde{\lambda}_0 \in \mathbb{R}^{m\times m}_+$ and set $k := 0$.
2: Let u_k denote the solution of the discrete version of the minimization problem (1.5) with discrete $\lambda = 2\lambda_k$.
3: Update λ_{k+1} based on u_k and (3.17).
4: Stop, or set $k := k+1$ and return to step 2.

Our numerical experience indicates that this basic MTV-algorithm exhibits a rather slow adjustment of λ, which, in particular for rather small initial λ_0, leads to an unacceptably large number of iterations; compare Table 1. In order to remedy this effect, in the next section we extend the notion of hierarchical decompositions in image processing (see [32, 33]) to spatially dependent regularization and combine it with our basic MTV-framework. This results in a tremendous acceleration of the basic MTV-algorithm and an enhancement in the reconstruction of image details. As a consequence, the associated algorithm (SA-TV, below) favorably competes with several other image restoration techniques; see Sect. 6 for the results and a comparison.

4 A Hierarchical Decomposition with Spatially Dependent λ

In [32, 33] Tadmor, Nezzar and Vese (TNV) introduced a method for hierarchically decomposing an image into

scales. They utilize concepts from interpolation theory to represent a noisy and blurry image as the sum of "atoms" u_k, where every u_k extracts features at a scale finer than for the previous u_{k-1}. This method acts like an iterative regularization scheme, i.e. up to some iteration index $\bar{k}$ the method yields improving reconstruction results with a deterioration (due to noise influence and ill-conditioning) beyond $\bar{k}$; see, e.g., the iteration sequence displayed in column (a) of Fig. 8. In our context we use the TNV-scheme to improve the basic MTV-algorithm. This results in a reduction of the number of iterations until successful termination and in a robust method with respect to the choice of the initial λ_0. We also mention that while the approach in [33] relies on a scalar regularization parameter λ, we extend the concept to a spatially varying one.

Considering dyadic scales, the resulting algorithm is as follows:

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

$$u_0 := \arg \min_{u \in BV(\Omega)} \int_\Omega |Du| + \frac{1}{2} \int_\Omega \lambda_0 (Ku - z)^2 \, dx. \tag{4.1}$$

(ii) For $j = 0, 1, 2 \ldots$ set $\lambda_j = 2^j \lambda_0$ and $v_j = z - Ku_j$. Then compute

$$\hat{u}_j := \arg \min_{u \in BV(\Omega)} \int_\Omega |Du| + \frac{1}{2} \int_\Omega \lambda_{j+1} (Ku - v_j)^2 \, dx,$$

$$u_{j+1} := u_j + \hat{u}_j. \tag{4.2}$$

Note that we assume here for simplicity that u_0 and $\hat{u}_j$, $j = 0, 1, \ldots$, are unique. The following results extending those in [33] can easily be proved. For the sake of completeness we provide the proofs in Appendix A. For the formulation of these results we introduce $\| \cdot \|_*$, the dual of the seminorm $\int_\Omega |Du|$, i.e.,

$$\|u\|_* := \sup_{\int_\Omega |D\varphi| \neq 0} \frac{\int_\Omega u \, \varphi \, dx}{\int_\Omega |D\varphi|}. \tag{4.3}$$

Thus, we have

$$\int_\Omega u \, \varphi \, dx \leq \|u\|_* \int_\Omega |D\varphi|, \quad \forall \varphi \in BV(\Omega). \tag{4.4}$$

The pair (u, φ) is called extremal if equality holds in (4.4). The following theorem characterizes a solution of (1.5) in terms of $\| \cdot \|_*$.

Theorem 7

(i) u is a minimizer of (1.5) if and only if

$$\int_\Omega u \, K^* \lambda(z - Ku) \, dx = \|K^* \lambda(z - Ku)\|_* \int_\Omega |Du|$$

$$= \int_\Omega |Du|. \tag{4.5}$$

(ii) $\|K^* \lambda z\|_* \leq 1$ *if and only if* $u = 0$ *is a minimizer of* (1.5).

(iii) *Assume that* $1 < \|K^* \lambda z\|_* < +\infty$. *Then* u *is a minimizer of* (1.5) *if and only if* $(K^* \lambda(z - Ku), u)$ *is an extremal pair and* $\|K^* \lambda(z - Ku)\|_* = 1$.

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

$$\left\| K^* \lambda_0 \left(z - \sum_{j=0}^k K \hat{u}_j \right) \right\|_* = \frac{1}{2^k}. \tag{4.6}$$

In order to extract image features at a finer scale in the spirit of the above localized version of the TNV technique, we correspondingly modify the iterative adaption of λ in (3.17) by setting

$$(\tilde{\lambda}_{k+1})_{i,j} := \zeta \min \left((\tilde{\lambda}_k)_{i,j} + \rho \left(\sqrt{(\tilde{S}_k^\omega)_{i,j}} - \sigma \right), L \right), \tag{4.7a}$$

$$(\lambda_{k+1})_{i,j} = \frac{1}{\omega^2} \sum_{(s,t) \in \Omega_{i,j}^\omega} (\tilde{\lambda}_{k+1})_{s,t}, \tag{4.7b}$$

where $\zeta \geq 1$ and L is a large positive value to ensure uniform boundedness of $\{\tilde{\lambda}_k\}$; otherwise if $\tilde{\lambda}_k$ would become unbounded (possibly only on a non-empty subset of the discrete Ω), then the local regularization effect would vanish and significant noise would remain. In our numerics we choose $\zeta = 2$, which comes from the notion of dyadic scales in the TNV-algorithm presented above, to accelerate the adjustment of λ. In addition, since all images in our numerical tests have a dynamic range of $[0, 1]$, and, thus, $0 < \sigma < 1$, for scaling purposes we replace $(\tilde{S}_k^\omega)_{i,j} - \sigma^2$ by $\sqrt{(\tilde{S}_k^\omega)_{i,j}} - \sigma$.

In the next section we propose an algorithm which uses (4.7) to accelerate the parameter adjustment and, hence, the image restoration.

5 Spatially Adapted TV-Algorithm

Based on the hierarchical decomposition of Sect. 4 and the local variance estimators and confidence interval technique of Sect. 3 we propose the following algorithm.

SA-TV-Algorithm.
1: Initialize $u_0 = 0 \in \mathbb{R}^{m \times m}$, $\lambda_0 = \tilde{\lambda}_0 \in \mathbb{R}_+^{m \times m}$ and set $k := 0$.
2: If $k = 0$, solve the discrete version of the minimization problem in (4.1), else compute $v_k = z - Ku_k$ and solve

the discrete version of

$$\min_{u \in BV(\Omega)} \int_\Omega |Du| + \frac{1}{2} \int_\Omega \lambda (Ku - v_k)^2 \, dx$$

with the discretization of λ equal to λ_k. Let $\hat{u}_k$ denote the corresponding solution.

3: Update $u_{k+1} = u_k + \hat{u}_k$.

4: Based on u_{k+1} and (4.7), update λ_{k+1}.

5: Stop, or set $k := k + 1$ and return to step 2.

A few remarks on the algorithm are in order: (i) We initialize λ by a relatively small constant, i.e. $\lambda_0 = \bar{\lambda}^0 \mathbf{1}$ with $\bar{\lambda}^0 > 0$ small and $\mathbf{1} \in \mathbb{R}^{m \times m}$ the matrix with all entries equal to 1. Extensive numerical results suggest that with a small $\bar{\lambda}^0$ our method is robust with respect to the choice of λ_0; see Sect. 6. (ii) The solution of the TV-problems in step 2 is obtained by a superlinear convergent semismooth Newton method; see [21] for scalar λ and the following subsection as well as Appendix B for an extension to the spatially dependent case. The parameters in this method are chosen as in [21]. (iii) In our numerical practice a 11-by-11 window turned out to yield reliable results. In order to support our choice, in Sect. 6 we study the influence of the window size on the restoration quality. (iv) Similar to the Bregman iteration proposed in [23], we stop the iterative procedure as soon as the residual $\|z - Ku_k\|_2$ drops below $\xi\sigma$, where $\xi > 1$ relates to the image size. For $m \to \infty$ we have $\xi \to 1$.

5.1 Primal-dual Approach to Spatially Adapted Total Variation

In [21] an infeasible primal-dual algorithm of generalized Newton-type was proposed for solving (1.5) with a scalar λ. In the sequel we extend its key features to the case where $\lambda = \lambda(x)$. Thus, the method serves as a solver for the problems in step 2 of our SA-TV-algorithm.

Rather than operating on the original TV-model (1.5) the method is based on

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

$$+ \int_\Omega |\nabla u|_2 \, dx, \qquad (5.1)$$

where $0 < \epsilon \le \lambda(x) \le \bar{\lambda}$ for almost all $x \in \Omega$ and $0 < \mu \ll \bar{\lambda}^{-1}$. The μ-term serves the purpose of a function space regularization for a "convenient" dualization in a Hilbert space setting. Its effect on the restoration results is negligible since $\mu \ll \bar{\lambda}^{-1}$. We point out that for $\mu \to 0$ the solution of (5.1) converges weakly in $L^2(\Omega)$ to a solution of (1.5). Moreover, in our numerics we even use $\mu = 0$.

Applying the Fenchel-Legendre calculus [14] analogously as in [21], the Fenchel-dual of (5.1) reads

$$\sup_{\substack{\vec{p} \in \mathbf{L}^2(\Omega) \\ |\vec{p}(x)| \le 1 \text{ a.e. in } \Omega}} -\frac{1}{2} |||K^* \lambda z - \operatorname{div} \vec{p}|||_{H^{-1}}^2 + \frac{1}{2} \int_\Omega \lambda z^2 \, dx, \quad (P_0)$$

where $|||v|||_{H^{-1}}^2 = \langle H_{\mu, K} v, v \rangle_{H_0^1, H^{-1}}$, $v \in H^{-1}(\Omega)$ with $H_{\mu, K} = (K^* \lambda K - \mu \triangle)^{-1}$, $\triangle : H_0^1(\Omega) \to H^{-1}(\Omega)$, and $\langle \cdot, \cdot \rangle_{H_0^1, H^{-1}}$ denotes the duality pairing between $H_0^1(\Omega)$ and its dual $H^{-1}(\Omega)$. Moreover, $\mathbf{L}^2(\Omega) := (L^2(\Omega))^2$. In order to avoid the non-uniqueness of the solution of (P_0), following [21] we consider a dual regularization:

$$\sup_{\substack{\vec{p} \in \mathbf{L}^2(\Omega) \\ |\vec{p}(x)| \le 1 \text{ a.e. in } \Omega}} -\frac{1}{2} |||K^* \lambda z - \operatorname{div} \vec{p}|||_{H^{-1}}^2$$

$$+ \frac{1}{2} \int_\Omega \lambda z^2 \, dx - \frac{\beta}{2} \int_\Omega \|\vec{p}\|_{\mathbf{L}^2}^2, \qquad (P)$$

where $\beta > 0$ is the associated regularization parameter. In order to study the effect of the β-regularization of the Fenchel-dual we apply the Fenchel-Legendre calculus once more and find that the dual of (P) is given by

$$\min_{u \in H_0^1(\Omega)} \frac{\mu}{2} \int_\Omega |\nabla u|_2^2 \, dx$$

$$+ \frac{1}{2} \int_\Omega \lambda |Ku - z|^2 \, dx + \int_\Omega \Phi_\beta(\nabla u) \, dx, \qquad (P^*)$$

where for $\vec{w} \in \mathbf{L}^2(\Omega)$,

$$\Phi_\beta(\vec{w})(x) = \begin{cases} |w(x)|_2 - \frac{\beta}{2} & \text{if } |w(x)|_2 \ge \beta, \\ \frac{1}{2\beta} |w(x)|_2^2 & \text{if } |w(x)|_2 < \beta. \end{cases} \qquad (5.2)$$

Note that Φ_β represents a local smoothing of $\int_\Omega |\nabla u|_2 \, dx$ in (5.1) to obtain uniqueness of the dual solution $\vec{p}$. In our numerics, we choose $\beta = 10^{-3}$.

The first-order optimality conditions of (P^*) characterize the solution $\bar{u}$ and $\bar{\vec{p}}$ of (P^*) and (P), respectively, by

$$-\mu \triangle \bar{u} + K^* \lambda K \bar{u} - \operatorname{div} \bar{\vec{p}} = K^* \lambda z \quad \text{in } H^{-1}(\Omega), \qquad (5.3a)$$

$$\max(\beta, |\nabla \bar{u}|_2) \bar{\vec{p}} - \nabla \bar{u} = 0 \quad \text{in } \mathbf{L}^2(\Omega). \qquad (5.3b)$$

Note that the system (5.3) is non-smooth, i.e. not necessarily Fréchet-differentiable. The discrete version of this system can be solved efficiently by a semismooth Newton method; see Appendix B for the semismooth Newton algorithm and details on the involved numerical linear algebra. The generalized Newton solver converges globally, i.e. regardless of its initialization, and locally at a superlinear rate [21].

Fig. 4 *Example 1:* Original images. (**a**) "Cameraman", (**b**) Part 1 of "Barbara", (**c**) Part 2 of "Barbara"

(a)   (b)   (c)
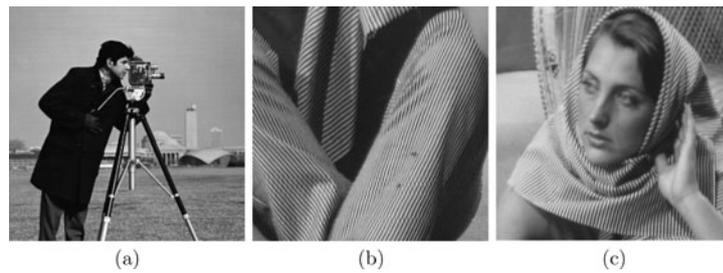
Fig. 5 *Example 1:* Results of different methods when restoring the noisy image "Cameraman". (**a**) Noisy image, (**b**) TV-method with [10] ($\alpha = 0.091$, PSNR = 27.04, MSSIM = 0.801), (**c**) TV-method ($\alpha = 0.07$, PSNR = 27.42, MSSIM = 0.783), (**d**) TV-method ($\alpha = 0.085$, PSNR = 27.19, MSSIM = 0.802), (**e**) Bregman iteration (PSNR = 27.34, MSSIM = 0.809), (**f**) TNV-method (PSNR = 26.96, MSSIM = 0.689), (**g**) Basic MTV-algorithm (PSNR = 26.94, MSSIM = 0.803), (**h**) Our method (PSNR = 27.90, MSSIM = 0.825)

(a)   (b)   (c)

(d)   (e)   (f)

(g)   (h)
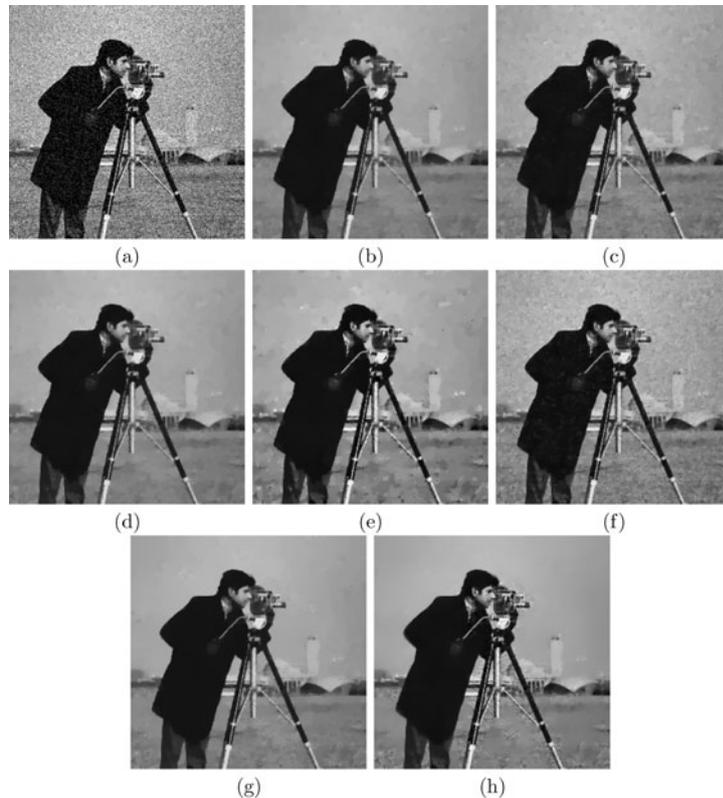
6 Numerical Results

In this section we provide numerical results to study the behavior of the SA-TV method with respect to its image restoration capabilities and its stability with respect to the choice of the initial λ and ω. Unless otherwise specified we concentrate on image denoising, i.e., K is the identity matrix, and use the window size $\omega = 11$, as mentioned earlier. Further, in all of our experiments reported on below the image intensity range is scaled to [0, 1].

6.1 Comparison with Other Restoration Techniques

We study the behavior of the SA-TV-method and compare it with the "classical" total variation (TV) method [26], the Bregman iteration [23], the TNV-method [32], and the basic MTV-method introduced in Sect. 3.3. While the first two methods operate with a scalar λ, the TNV-method yields a genuine hierarchical decomposition. The

performance of these methods is compared quantitatively by means of the peak signal-to-noise ratio (PSNR) [8], which is a widely used image quality assessment measure, and the recently proposed structural similarity measure (MSSIM) [35], which relates to perceived visual quality better than PSNR.

Example 1 Our first test examples are displayed in Fig. 4, where the original images "Cameraman" (256-by-256) and "Barbara" (512-by-512) are shown. For a study of our method in the case of texture-like structures we zoom into certain regions of the "Barbara"-image; see the middle and right plots. The original images can be found in [1]. In this example, we consider degraded images which are corrupted by Gaussian white noise with the noise level $\sigma = 0.1$. The noisy images are shown in Figs. 5, 6, and 7(a).

Fig. 6 *Example 1:* Results of different methods when restoring part of the noisy image "Barbara". (**a**) Noisy image, (**b**) TV-method with [10] ($\alpha = 0.064$, PSNR = 22.07, MSSIM = 0.659), (**c**) TV-method ($\alpha = 0.04$, PSNR = 22.81, MSSIM = 0.689), (**d**) TV-method ($\alpha = 0.045$, PSNR = 22.76, MSSIM = 0.691), (**e**) Bregman iteration (PSNR = 21.34, MSSIM = 0.688), (**f**) TNV-method (PSNR = 22.65, MSSIM = 0.691), (**g**) Basic MTV-algorithm (PSNR = 22.21, MSSIM = 0.675), (**h**) Our method (PSNR = 23.32, MSSIM = 0.767)
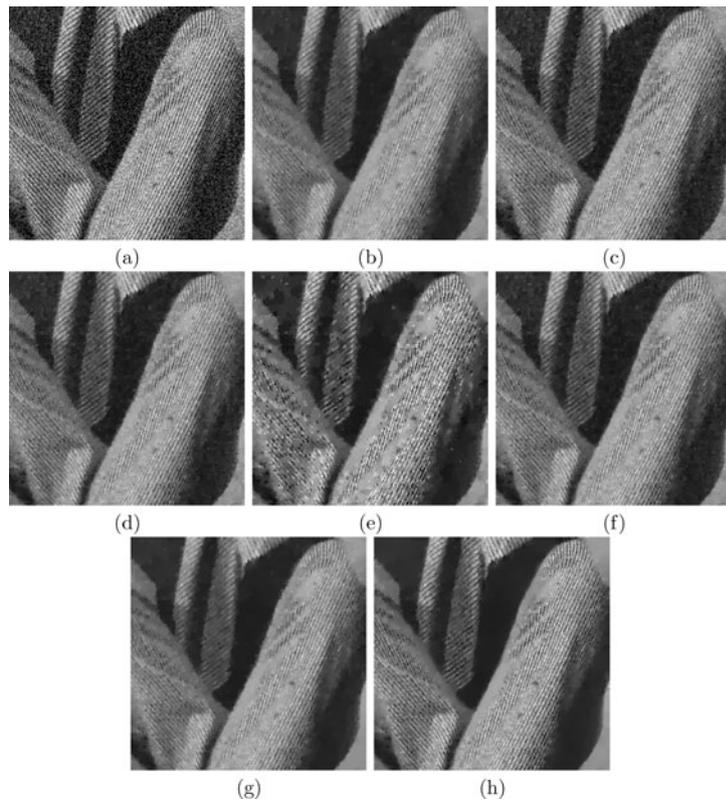
Fig. 7 *Example 1:* Results of different methods when restoring part of the noisy image "Barbara". (**a**) Noisy image, (**b**) TV-method with [10] ($\alpha = 0.088$, PSNR = 25.32, MSSIM = 0.723), (**c**) TV-method ($\alpha = 0.06$, PSNR = 25.97, MSSIM = 0.734), (**d**) TV-method ($\alpha = 0.07$, PSNR = 25.85, MSSIM = 0.742), (**e**) Bregman iteration (PSNR = 25.23, MSSIM = 0.747), (**f**) TNV-method (PSNR = 25.92, MSSIM = 0.716), (**g**) Basic MTV-algorithm (PSNR = 25.43, MSSIM = 0.732), (**h**) Our method (PSNR = 26.45, MSSIM = 0.785)

We start by comparing the SA-TV method with a few other methods listed in the beginning of this section. The results are shown in Figs. 5, 6, and 7. For the TV-method we show the restored image with α chosen such that the second constraint in (1.2) is satisfied; see [10]. In addition, after many experiments with different α-values in the model (1.4), the one with the best PSNR and the one with the best MSSIM are also presented here. Comparing these values we find that the results with the largest MSSIM values match the human visual system better than the one with the largest PSNR. The Bregman iteration is often used for contrast enhancement, but it is also an excellent method for noise removal. Therefore, we also list its results here for comparison. For a fair comparison, we use the same initial choices $\lambda^0 = 2.5$ or $\alpha^0 = 0.4$ and the same stopping rule for the Bregman iteration, the TNV-method, the basic MTV-method of Sect. 3.3 and the SA-TV-method of Sect. 5 with $\mu = 0$ and $\beta = 10^{-3}$, i.e, the respective algorithm is stopped as soon as the residual $\|z - u_k\|_2$ drops below the noise level σ.

From Figs. 5, 6, and 7 we find that our SA-TV-method performs best both visually and quantitatively. Note that in the images restored by the TV-method, we observe the usual result that small α preserves details, but at the same time some noticeable noise remains; otherwise, if α is large, the details are overregularized. Although the Bregman iteration removes most of the noise, it still gives more heterogeneous results; see in particular Fig. 5. Since the TNV-method performs a hierarchical image decomposition, image details are added back in every iteration, but at the same time some noise is also added back; see the background of Figs. 5, 6, and 7. For the images restored by the basic MTV-method we note that details and features at a larger scale are not as well resolved as in case of the TNV- or our new SA-TV method. This is due to a slower update of λ. In fact, when initializing by a rather small λ^0, our stopping rule involving the global variance of the current reconstruction terminates the iteration even before λ becomes sufficiently large in detail regions for recovering these regions sufficiently accurately; see, e.g., the camera in Fig. 5. Further, large scale features appear somewhat oversmoothed; see, e.g., the face of Barbara in Fig. 7. At the cost of additional iterations, a localized stopping rule improves the reconstructions. Note, however, that the basic MTV based recovery of details and the reconstruction of homogeneous regions still favorably compares with the results by the TV-method. The SA-TV-method, on the other hand, suppresses noise successfully while preserving significantly more details. For instance, the sky in Fig. 5 and the arms of Barbara in Fig. 7 are smooth and the details of the camera in Fig. 5 and the features on the scarf in Figs. 6 and 7 are preserved clearly without being degraded by noise. With respect to PSNR and MSSIM, we also find quantitatively that our method gives the best restoration results. On the other hand, a very close inspection shows, e.g., a slight halo around the camera and still some heterogene-

Table 1 CPU-time in seconds and the number of iterations by different methods

	"Cameraman"		"Barbara"	
	CPU-Time	k	CPU-Time	k
Bregman iteration	94.19	4	548.68	4
TNV-method	78.66	4	422.02	4
Basic MTV-method	213.38	22	1620.4	37
SA-TV-method	60.26	3	364.24	3

ity in the lawn area of Fig. 5. Related heterogeneous effects can be found in Figs. 6 and 7. One might hope that these effects may get removed by an anisotropic smoothing of λ or by adaptive local windows, which opens up interesting research perspectives.

6.1.1 Computational Efficiency

In the SA-TV-method, we utilize the hierarchical decomposition concept of the TNV-method. As a result, our adaptive choice of λ not only improves the TNV-method with respect to its restoration capability (see Figs. 5, 6, and 7), but it also reduces the number of iterations and, hence, the CPU-time until successful termination. In Fig. 8, the restored image and the corresponding residual in each iteration of the TNV- and the SA-TV-method are shown, respectively. In the first iteration, since we utilize the same initial value of λ^0 for both algorithms, both restoration results and their pertinent residuals are identical; compare the first row of Fig. 8. We note that the residual contains most of the details. In the next iteration, for the TNV-method we have $\lambda_1 = 2\lambda_0$ in order to extract features at a finer scale. The SA-TV-method, however, is based on (4.7) and the associated λ_1 is larger than $2\lambda_0$ in image regions corresponding to details. Thus, these details are extracted better already in the second iteration; see the camera and the tripod in the second row of Fig. 8 (c) and (d). By similar reasons, in the third iteration more details are added back by the SA-TV-method than by the TNV-method; see, e.g., the buildings in the background. Then, the SA-TV-method satisfies the stopping condition first because it has extracted the various image features much faster. Furthermore, when the TNV-method satisfies the stopping condition, the result not only includes more details but also significant noise. For this aspect observe the background regions.

For comparing the computational time, in Table 1 we list the CPU-times consumed by the iterative methods in our comparison. All simulations are run in Matlab 7.5 (R2007b) on a PC equipped with P4 3.0 GHz CPU and 3G RAM memory. Since for all methods most of the computations within each iteration are spent for solving a total variation type problem, due to requiring the least number of iterations the SA-TV-method also spends least CPU-time. Moreover, we

Fig. 8 *Example 1:* Comparison of iterations of the TNV-method and the SA-TV-method when restoring the noisy image "Cameraman". (**a**) Result of TNV-method, (**b**) Residual of TNV-method, (**c**) Result of SA-TV-method, (**d**) Residual of SA-TV-method
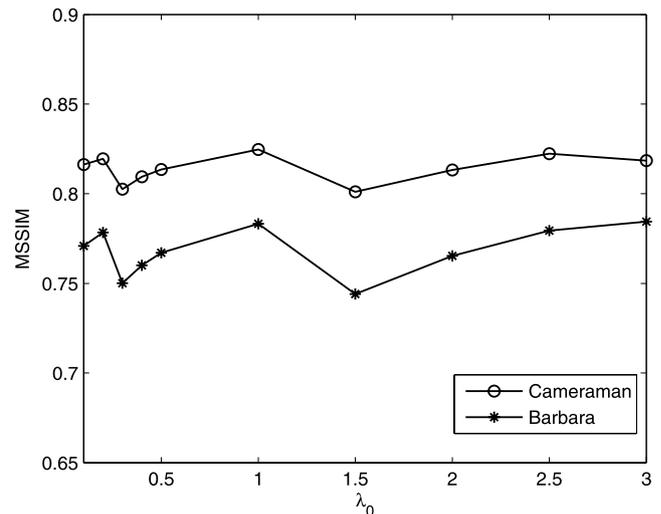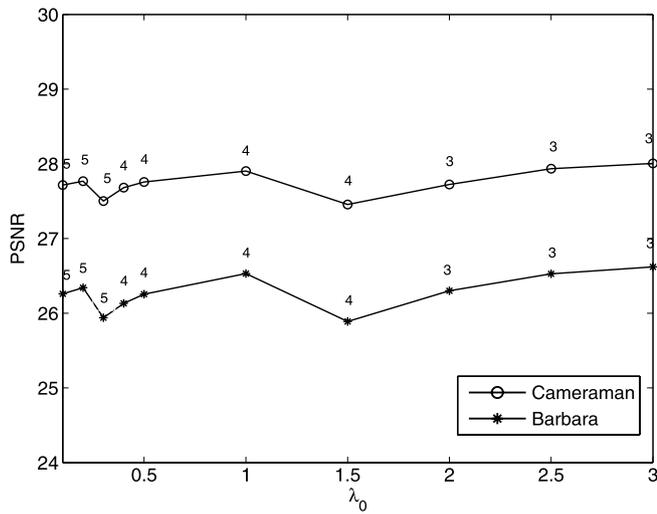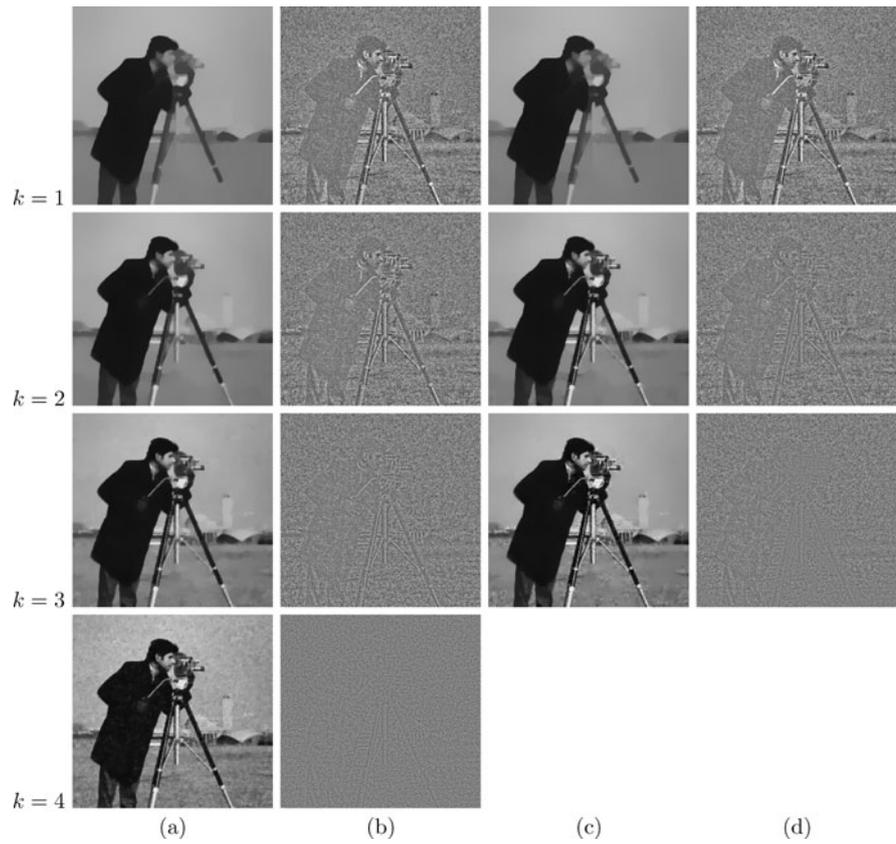
Fig. 9 *Example 1:* PSNR and MSSIM for images restored by our method for different initial $\bar{\lambda}^0$

find that with the original adaptive selection of λ in (3.17) the basic MTV-method needs significantly more iterations to meet the stopping condition.

6.1.2 Dependence on $\bar{\lambda}^0$

Concerning the influence of the initial parameter $\bar{\lambda}^0$ on the restoration behavior we observe that our method is rather

stable with respect to $\bar{\lambda}^0$. In Fig. 9, we plot the PSNR- and MSSIM-values for the images restored by our method with $\bar{\lambda}^0$ varying from 0.1 to 3. Since λ controls the trade-off between a good data fit and the regularization coming from the TV-term, it has large effect on the variance of the residual $\|z - u_k\|_2$. This can be seen from the left plot in Fig. 9 where we also specify the number of iterations for each PSNR-

Fig. 10 *Example 1:* Images restored by our method with different $\bar{\lambda}^0$. (**a**) $\bar{\lambda}^0 = 0.1$, (**b**) $\bar{\lambda}^0 = 0.5$, (**c**) $\bar{\lambda}^0 = 1$, (**d**) $\bar{\lambda}^0 = 2$

(a)　　　(b)　　　(c)　　　(d)
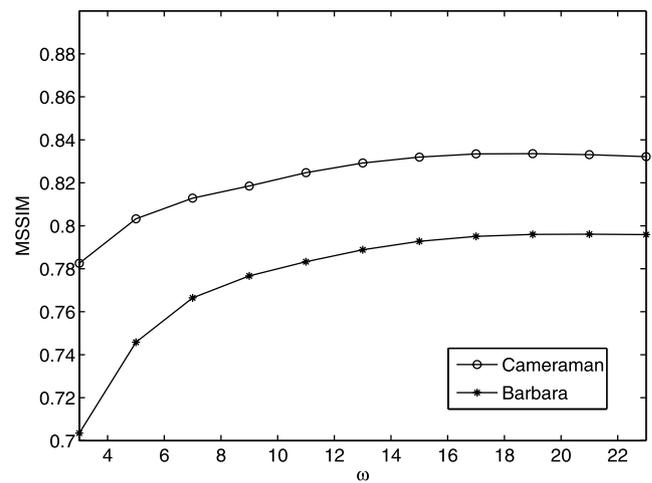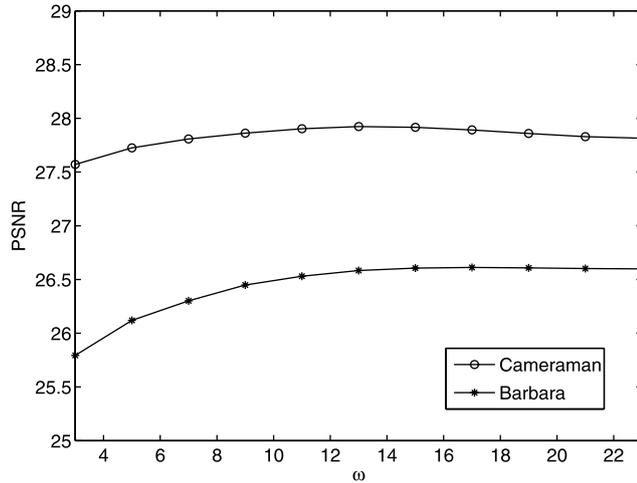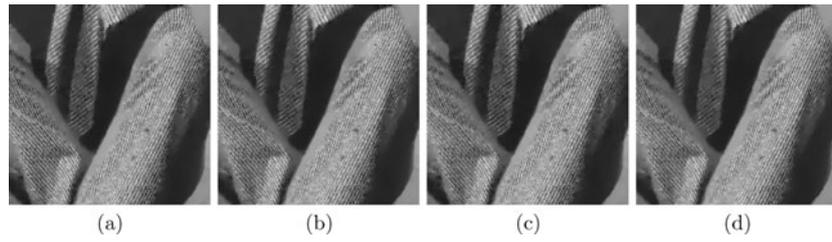
Fig. 11 *Example 1:* PSNR and MSSIM for results obtained by our method with different ω

Fig. 12 *Example 1:* Restored images by our method with different ω. (**a**) $\omega = 3$, (**b**) $\omega = 7$, (**c**) $\omega = 13$, (**d**) $\omega = 17$

(a)　　　(b)　　　(c)　　　(d)
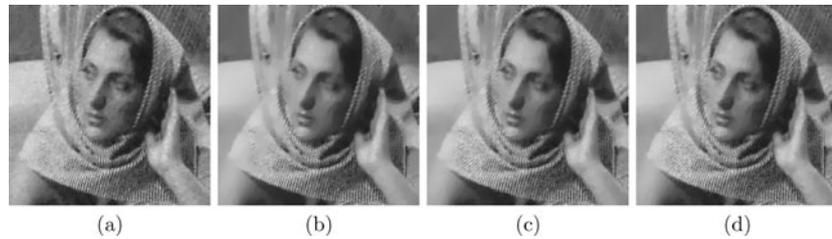
value. From the plots we can see that PSNR and MSSIM are rather stable. In order to illustrate the influence on the restored images, in Fig. 10 we show the results of part 1 of "Barbara" for several values of $\bar{\lambda}^0$.

6.1.3 Dependence on ω

We also test our method for different values of the window size ω varying from 3 to 23. Figure 11 shows the plots of the PSNR- and MSSIM-values of the denoising results for the same denoising problems as in Figs. 5 and 7 with $\bar{\lambda}^0 = 1$. Except for very small window size, we observe a remarkable stability with respect to ω. This can also be seen from the restored images in Fig. 12. Too small window sizes (here $\omega = 3$ and 5) yield comparatively small PSNR and MSSIM values, which is a consequence of the small sample sizes. Figure 12 shows that with $\omega = 3$ some noise persists, whereas sufficiently large ω reduces noise effects and

recovers details; compare the rather constant graphs for the PSNR- and MSSIM values for $\omega \geq 11$. If, however, ω becomes too large, then the regularization parameter choice becomes rather global than local which compromises image details.

Example 2 (Medical image restoration) In Fig. 13 we show the results obtained by the TV-method (c), the Bregman iteration (d), the TNV-method (e) and our algorithm (f), respectively, when denoising the magnetic resonance image of a rabbit heart at a resolution of 1024×1024 pixels (original (a) with (b) an enlarged part). For the TV-method we choose $\alpha = 1/\lambda$ based on the algorithm in [10]. In the other iteration methods, we estimate the noise level as the variance in a homogeneous region, and set the parameter $\bar{\lambda}^0 = 1$. For a better inspection of the result, we enlarge a part of the image in Fig. 13. We note that the classical TV-method is outperformed by the Bregman iteration, the TNV-method and our

Fig. 13 *Example 2:* Results when restoring the medical image of rabbit heart: (**a**) Original image, (**b**) Enlarged original part, (**c**) TV-method, (**d**) Bregman iteration ($k = 20$), (**e**) TNV-method ($k = 7$), (**f**) SA-TV-method ($k = 3$)

(a) (b) (c)

(d) (e) (f)

(a) (b) (c)

(d) (e) (f)
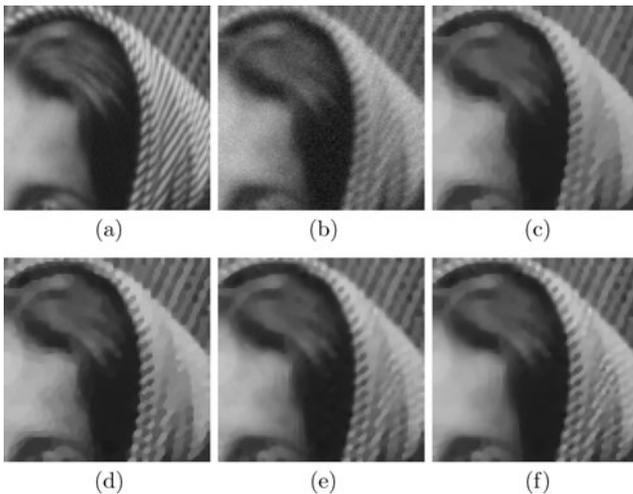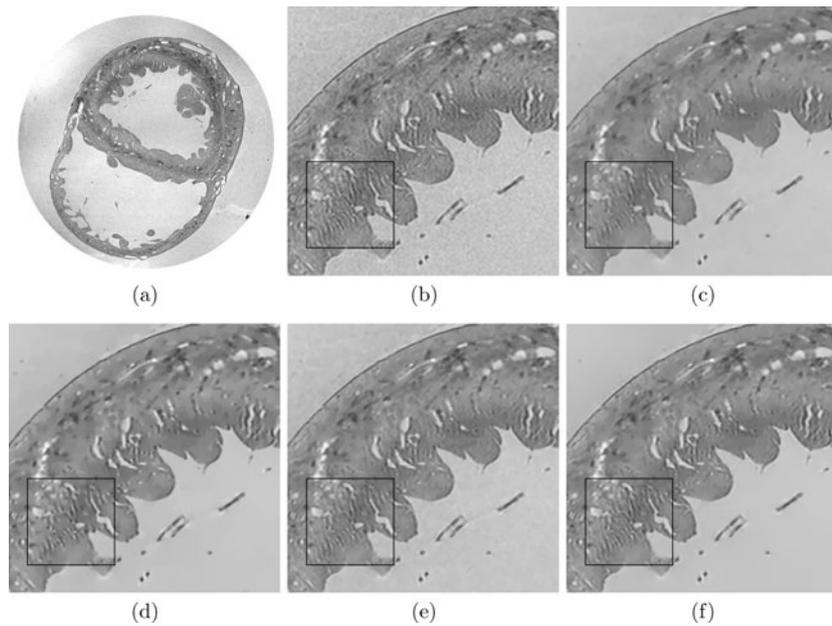
Fig. 14 *Example 3:* Results of different methods when restoring the blurred and noisy image "Barbara": (**a**) Original image, (**b**) Blurred noisy image, (**c**) TV-method (PSNR = 28.06, MSSIM = 0.823), (**d**) Bregman iteration (PSNR = 28.27, MSSIM = 0.839, $k = 76$), (**e**) TNV-method (PSNR = 29.15, MSSIM = 0.865, $k = 9$), (**f**) Our method (PSNR = 29.68, MSSIM = 0.883, $k = 4$)

method, with the Bregman iteration missing some of the details; see, e.g., the marked square in the lower left corner. In this highlighted region we find that in subplot (c) many details (fiber directions) are lost. In subplot (d) this effect is less pronounced, but nevertheless, when compared to (e) and (f), the quality of the restored details is significantly reduced. While the TNV-method and our method give visually similar results for this test image we observe that our method requires the smallest number of iterations (also when compared with the Bregman iteration).

Example 3 (Simultaneous deblurring and denoising) Finally, we illustrate the restoration ability of our method for noisy blurred images. The blurring is due to a Gaussian convolution with a 9×9 window and a standard deviation of 1. Further we have Gaussian white noise with $\sigma = 0.02$. Figure 14 depicts a part of the noisy blurred "Barbara" image and the restoration results. For the TV-method we use $\alpha = 1/\lambda$ according to [10]. For the other methods we set $\bar{\lambda}^0 = 1$. Comparing the result obtained by our method with the others, we find that our method preserves details better; see, e.g., the features on the scarf. Based on PSNR and MSSIM, our method also outperforms the other methods. In addition, although we use the same value of $\bar{\lambda}^0$ for the Bregman iteration as well as the TNV-method and our algorithm, our method needs the smallest number of iterations. In this respect, we recall that our method intertwines the TNV-hierarchical decomposition concept [33], but with a spatially varying λ, with a confidence interval based λ-update. This fact is responsible for extracting details faster and, thus, resulting in a smaller iteration number than the pure TNV-method.

6.2 Stability of Our Regularization Parameter Choice Rule

Since the performance of multi-scale total variation is mainly influenced by the selection of the parameter λ, in this section we discuss the new spatially adaptive selection of λ in our method in case of various noise levels.

Example 4 We consider a 300-by-300 test image as shown in Fig. 15(a). The third row of Fig. 15 depicts the final values of the parameter λ upon termination of our method with $\bar{\lambda}^0 = 0.2$. In all cases λ is large in detail regions and small in
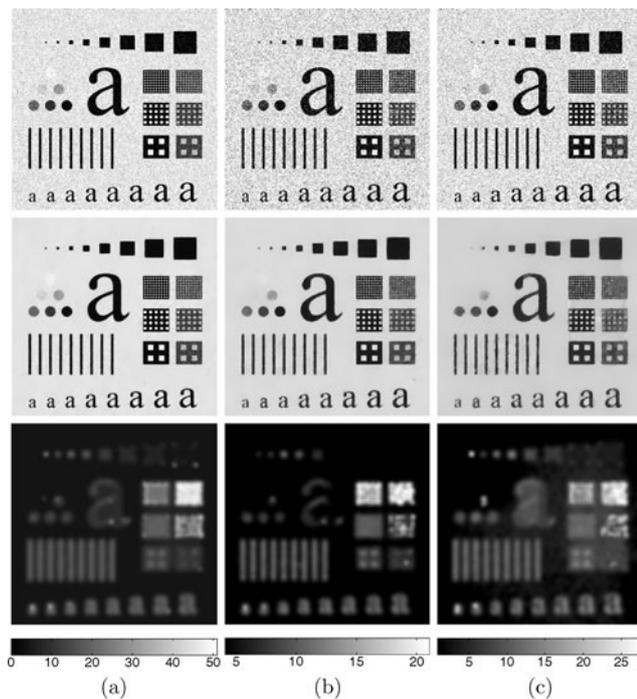
Fig. 15 *Example 4*: Results of the SA-TV-method for different noise levels σ (row 1: noisy images; row 2: restored images; row 3: final values of λ). (**a**) $\sigma = 0.1$, (**b**) $\sigma = 0.2$, (**c**) $\sigma = 0.3$

the homogeneous background. Moreover, small scale features lead to large λ. When comparing the restoration results of Fig. 15 (second row) for different noise levels, i.e., $\sigma = 0.1$, 0.2, and 0.3, respectively, we find that even for rather high noise level our method still is able to distinguish most of the detail regions and assigns automatically a large value to λ in these regions. Thus, our parameter choice rule appears to be robust with respect to noise.

In Fig. 3 we compared our upper bound for the local variance estimator with the bound in [15]. Here we continue the investigation of the stability of our upper bound by studying its behavior when over- or underestimating the variance σ^2. Figure 16 depicts the restoration results for underestimated (see the plots (a) and (b)) and for overestimated variance (see
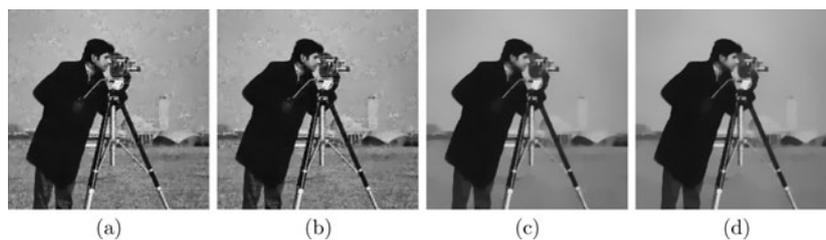
the plots (c) and (d)). We observe that an underestimated variance tightens the upper bound resulting in an overestimation of λ and, hence, too little regularization such that noise appears in the restoration; see plot (a) where the estimated σ is 0.08, whereas the true one is 0.1. On the other hand, if σ is overestimated, λ is underestimated and too much regularization takes place. The latter adversely affects the recovery of image details; compare plot (d) of Fig. 16. Obviously, the quality of the reconstruction depends on the quality of the estimate of σ, where overestimation appears less critical than underestimation.

In Fig. 17 we show the final values of λ obtained by our choice rule for the examples 1 to 3. Again in detail regions λ is large in order to preserve the details, and it is small in the homogeneous regions to remove noise. Furthermore, for the noisy blurred image our method is able to distinguish most of the detail regions properly; see Fig. 17(e).

7 Conclusions

A spatially adapted regularization parameter λ in the ROF-model is justified by considering an equivalent minimization problem subject to pointwise constraints. The introduction of a local variance estimator of the residual image turns out to be an accurate instrument for updating λ within an iterative procedure. Even though the update of the regularization parameter can be done gradually by using the local variance estimator only, combining the updating process with the hierarchical decomposition approach proposed by Tadmor, Nezar and Vese considerably reduces the number of iterations for adjusting λ according to the image scales and it yields even better results with respect to the recovery of image details when compared to the TNV-method as a stand-alone technique. Further, assuming that the noise variance σ^2 is known, the present algorithm is completely automatized, i.e., there is no necessity of tuning regularization parameters. The overall method combines the hierarchical decomposition based λ-adjustment scheme with an inexact semismooth Newton solver relying on Fenchel duality for

Fig. 16 *Example 4*: Results for restoring noisy image "Cameraman" with $\sigma = 0.1$ by SA-TV-method with inaccurate noise level estimation $\tilde{\sigma}$. (**a**) With $\tilde{\sigma} = 0.08$ (PSNR = 27.61, MSSIM = 0.793, $k = 3$),

(**b**) With $\tilde{\sigma} = 0.09$ (PSNR = 27.93, MSSIM = 0.824, $k = 3$), (**c**) With $\tilde{\sigma} = 0.11$ (PSNR = 25.90, MSSIM = 0.766, $k = 3$), (**d**) With $\tilde{\sigma} = 0.12$ (PSNR = 25.76, MSSIM = 0.763, $k = 3$)

Fig. 17 *Examples 1–3:* Final values of λ by our parameter choice rule for (**a**)–(**c**) Example 1, (**d**) Example 2, (**e**) Example 3
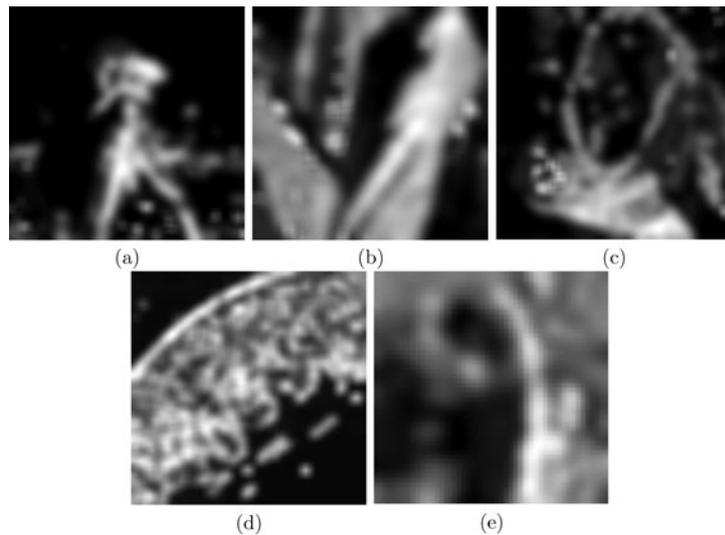
total variation regularized subproblems. The numerical results show that the new method outperforms several popular TV-based methods with respect to both noise removal and detail preservation.

Appendix A: Proofs of Sect. 4

We provide a proof of Theorem 7 which extends the results of [33].

Proof (i) Let $\epsilon > 0$ be sufficiently small. Then, if u is a minimizer of (1.5), we have for all $\varphi \in BV(\Omega)$

$$\int_{\Omega} |D(u + \epsilon\varphi)| + \frac{1}{2}\int_{\Omega} \lambda(K(u + \epsilon\varphi) - z)^2 \, dx$$
$$\geq \int_{\Omega} |Du| + \frac{1}{2}\int_{\Omega} \lambda(Ku - z)^2 \, dx, \tag{A.1}$$

$$\int_{\Omega} |D(u + \epsilon\varphi)| + \epsilon\int_{\Omega} \lambda(K\varphi)(Ku - z) \, dx$$
$$+ \frac{\epsilon^2}{2}\int_{\Omega} \lambda(K\varphi)^2 \, dx \geq \int_{\Omega} |Du|, \tag{A.2}$$

$$\int_{\Omega} |Du| + \epsilon\int_{\Omega} |D\varphi| + \epsilon\int_{\Omega} \lambda(K\varphi)(Ku - z) \, dx$$
$$+ \frac{\epsilon^2}{2}\int_{\Omega} \lambda(K\varphi)^2 \, dx \geq \int_{\Omega} |Du|. \tag{A.3}$$

Dividing (A.3) by $\epsilon > 0$ and letting $\epsilon \to 0_+$ we get

$$\int_{\Omega} |D\varphi| \geq \int_{\Omega} \lambda(K\varphi)(z - Ku) \, dx = \int_{\Omega} \varphi K^*\lambda(z - Ku) \, dx.$$

This yields

$$\sup_{\int_{\Omega} |D\varphi| \neq 0} \frac{\int_{\Omega} \varphi K^*\lambda(z - Ku) \, dx}{\int_{\Omega} |D\varphi|} \leq 1$$

and further $\|K^*\lambda(z - Ku)\|_* \leq 1$. For the reverse inequality set $\varphi = u$ and $-1 < \epsilon < 0$. Then starting from (A.2) we have

$$(1 + \epsilon)\int_{\Omega} |Du| + \frac{\epsilon^2}{2}\int_{\Omega} \lambda(Ku)^2 \, dx$$
$$\geq \int_{\Omega} |Du| + \epsilon\int_{\Omega} \lambda(Ku)(z - Ku) \, dx.$$

Dividing by $\epsilon < 0$ and letting $\epsilon \to 0_-$ we obtain

$$\int_{\Omega} |Du| \leq \int_{\Omega} uK^*\lambda(z - Ku) \, dx$$
$$\leq \|K^*\lambda(z - Ku)\|_* \int_{\Omega} |Du| \leq \int_{\Omega} |Du|.$$

For the sufficiency, we note that

$$\int_{\Omega} \lambda(z - K(u + \varphi))^2 \, dx$$
$$= \int_{\Omega} \lambda(z - Ku)^2 - 2\int_{\Omega} \lambda(z - Ku)(K(u + \varphi)) \, dx$$
$$+ 2\int_{\Omega} \lambda(z - Ku)(Ku) \, dx + \int_{\Omega} \lambda(K\varphi)^2 \, dx.$$

The second term of the right hand side above is estimated by

$$\int_{\Omega} K^*\lambda(z - Ku)(u + \varphi) \, dx$$

$$\leq \|K^*\lambda(z - Ku)\|_* \int_\Omega |D(u + \varphi)| \leq \int_\Omega |D(u + \varphi)|,$$

and because $(K^*\lambda(z - Ku), u)$ is an extremal pair, the third term implies

$$\int_\Omega K^*\lambda(z - Ku)u\,dx = \int_\Omega |Du|.$$

Hence, we have

$$\int_\Omega |D(u + \varphi)| + \frac{1}{2}\int_\Omega \lambda(z - K(u + \varphi))^2\,dx$$

$$\geq \int_\Omega |D(u + \varphi)| + \frac{1}{2}\int_\Omega \lambda(z - Ku)^2\,dx$$

$$- \int_\Omega |D(u + \varphi)| + \int_\Omega |Du| + \frac{1}{2}\int_\Omega \lambda(K\varphi)^2\,dx$$

$$\geq \frac{1}{2}\int_\Omega \lambda(z - Ku)^2 + \int_\Omega |Du|$$

which yields that u is a minimizer.

(ii) Let $\|K^*\lambda z\|_* \leq 1$. Given the characterization (4.5), we obtain

$$\int_\Omega |Du| = \int_\Omega uK^*\lambda z\,dx - \int_\Omega uK^*\lambda Ku$$

$$\leq \|K^*\lambda z\|_* \int_\Omega |Du| - \int_\Omega (\sqrt{\lambda}Ku)^2\,dx$$

$$\leq \int_\Omega |Du| - \int_\Omega (\sqrt{\lambda}Ku)^2\,dx.$$

Hence, $Ku = 0$ and

$$\int_\Omega |Du| + \frac{1}{2}\int_\Omega \lambda(Ku - z)^2\,dx = \int_\Omega |Du| + \frac{1}{2}\int_\Omega \lambda z^2\,dx.$$

Therefore, u must be a constant. Based on the properties of K, we have $u \equiv 0$.

Now, if $u = 0$ is a minimizer of (1.5), then for any $\varphi \in BV(\Omega)$ we have

$$\frac{1}{2}\int_\Omega \lambda z^2\,dx \leq \int_\Omega |D\varphi| + \frac{1}{2}\int_\Omega \lambda(K\varphi - z)^2\,dx$$

and further

$$\int_\Omega \lambda z K\varphi\,dx \leq \int_\Omega |D\varphi| + \frac{1}{2}\int_\Omega (\sqrt{\lambda}K\varphi)^2\,dx.$$

From a rescaling by setting $\varphi \mapsto \epsilon\varphi$, we obtain

$$\epsilon \int_\Omega \lambda z K\varphi\,dx \leq \epsilon \int_\Omega |D\varphi| + \frac{\epsilon^2}{2}\int_\Omega (\sqrt{\lambda}K\varphi)^2\,dx.$$

Next we divide by ϵ and let $\epsilon \to 0_+$, which results in

$$\int_\Omega \lambda z K\varphi \leq \int_\Omega |D\varphi|$$

for all $\varphi \in BV(\Omega)$. We conclude that $\|K^*\lambda z\|_* \leq 1$.

(iii) Given the result (4.5) we have to show that $\int_\Omega |Du|$ does not vanish.

Let us suppose it does vanish. Then u is a constant. For any constant c and any $\varphi \in BV(\Omega)$ with $\int_\Omega |D\varphi|dx > 0$ we have

$$\frac{\int_\Omega K^*\lambda z(\varphi + c)\,dx}{\int_\Omega |D(\varphi + c)|} = \frac{\int_\Omega K^*\lambda z\varphi\,dx}{\int_\Omega |D\varphi|} + c\frac{\int_\Omega K^*\lambda z\,dx}{\int_\Omega |D\varphi|}.$$

Based on the assumption $\|K^*\lambda z\|_* < +\infty$ and (4.3), this implies $\int_\Omega K^*\lambda z c\,dx = 0$ for all constants c. Thus,

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

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

(iv) For u_0 we have $\|K^*\lambda_0(z - Ku_0)\|_* = 1$, and for $\hat{u}_k$ there holds

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

or, with $\lambda_{k+1} = 2\lambda_k$, we have

$$\|K^*\lambda_0(z - Ku_k - K\hat{u}_k)\|_*$$

$$= \|K^*\lambda_0(z - Ku_{k+1})\|_*$$

$$= \left\|K^*\lambda_0\left(z - \sum_{j=0}^k K\hat{u}_j\right)\right\|_* = \frac{1}{2^k}$$

for $k = 0, 1, 2, \ldots$. □

Appendix B: Semismooth Newton Method for Solving (5.3)

We explain our semismooth Newton solver by means of the discrete version of (5.3) and vector-valued variables. For this purpose let $u_\ell \in \mathbb{R}^M$, $p_\ell \in \mathbb{R}^{2M}$, $\lambda_\ell \in \mathbb{R}^M$, for some $M \in \mathbb{N}$ which depends on the image size $m \times m$, denote the discrete image intensity, dual variable and spatially dependent λ, respectively. The subscript ℓ refers to the ℓ-th element of a sequence generated by the semismooth Newton solver introduced below. Further, let $z \in \mathbb{R}^M$ denote the discrete data vector. We define the discrete gradient operator $\nabla \in \mathbb{R}^{2M \times M}$ and the discrete operator $B_{\mu,\lambda} = -\mu\triangle + K^\top D(\lambda)K$ with $\triangle, K^\top D(\lambda)K \in \mathbb{R}^{M \times M}$ and $K^\top$ the transpose of $K \in \mathbb{R}^{M \times M}$. Here $D(\lambda) = \text{diag}(\lambda_1, \ldots, \lambda_M)$. We use $m_\ell = \max(\beta e, \varrho(\nabla u_\ell)) \in \mathbb{R}^{2M}$, where $e = (1, 1, \ldots, 1)^\top \in \mathbb{R}^{2M}$ and $(\varrho(v))_i = (\varrho(v))_{i+M} = |((v_x)_i, (v_y)_i)^\top|_2 = \sqrt{|(v_x)_i|^2 + |(v_y)_i|^2}$ $(1 \leq i \leq M)$ for $v \in \mathbb{R}^{2M} = (v_x^\top, v_y^\top)^\top$

with $v_x, v_y \in \mathbb{R}^M$. Moreover, $\chi_{\mathcal{A}_{\ell+1}} = D(t_\ell) \in \mathbb{R}^{2M \times 2M}$ with

$$(t_\ell)_i = \begin{cases} 1, & \text{if } (\varrho(\nabla u_\ell))_i \geq \beta, \\ 0, & \text{else}, \end{cases}$$

and $\mathcal{N}$ denotes the Jacobian of the function ϱ, i.e.

$$\mathcal{N}(v) = (D(\varrho(v)))^{-1} \begin{pmatrix} D(v_x) \, D(v_y) \\ D(v_x) \, D(v_y) \end{pmatrix}.$$

The discrete version of (5.3) at (u_ℓ, p_ℓ) is given by

$$B_{\mu,\lambda} u + \nabla^\top p = K^\top D(\lambda) z, \tag{B.1a}$$

$$D(m_\ell) p - \nabla u = 0. \tag{B.1b}$$

Applying a generalized Newton step to (B.1) at (u_ℓ, p_ℓ) yields

$$\begin{pmatrix} B_{\mu,\lambda} & \nabla^\top \\ (-D(e) + \chi_{\mathcal{A}_{\ell+1}} D(p_\ell) \mathcal{N}(\nabla u_\ell)) \nabla & D(m_\ell) \end{pmatrix} \begin{pmatrix} \delta u \\ \delta p \end{pmatrix}$$
$$= \begin{pmatrix} -B_{\mu,\lambda} u_\ell - \nabla^\top p_\ell + K^\top D(\lambda) z \\ \nabla u_\ell - D(m_\ell) p_\ell \end{pmatrix}, \tag{B.2}$$

where $\delta u \in \mathbb{R}^M$ and $\delta p \in \mathbb{R}^{2M}$ denote the update directions.

Since $D(m_\ell)$ is invertible, we eliminate δp from (B.2) and obtain the reduced system

$$H_\ell \delta u = f_\ell, \tag{B.3}$$

where

$$H_\ell = B_{\mu,\lambda} + \nabla^\top D(m_\ell)^{-1} \big[D(e) - \chi_{\mathcal{A}_{\ell+1}} D(p_\ell) \mathcal{N}(\nabla u_\ell) \big] \nabla,$$

$$f_\ell = -B_{\mu,\lambda} u_\ell + K^\top D(\lambda_\ell) z - \nabla^\top D(m_\ell)^{-1} \nabla u_\ell.$$

We note that δu is a decent direction for the discrete objective in (5.1) if H_ℓ is positive definite. Similar to [21, Cor. 3.2] the positive definiteness is guaranteed under the following condition

$$\varrho(p_\ell)_i \leq 1 \quad \text{and} \quad (b_{\ell,i} + c_{\ell,i})^2 \leq 4 a_{\ell,i} d_{\ell,i}$$
$$\text{for all } i = 1, \ldots, M, \tag{C}$$

where

$$a_{\ell,i} = 1 - (\varrho(\nabla u_\ell))_i^{-1} (p_\ell)_i (\nabla_x u_\ell)_i,$$

$$b_{\ell,i} = - (\varrho(\nabla u_\ell))_i^{-1} (p_\ell)_i (\nabla_y u_\ell)_i,$$

$$c_{\ell,i} = - (\varrho(\nabla u_\ell))_i^{-1} (p_\ell)_{i+M} (\nabla_x u_\ell)_i,$$

$$d_{\ell,i} = 1 - (\varrho(\nabla u_\ell))_i^{-1} (p_\ell)_{i+M} (\nabla_y u_\ell)_i.$$

In the following theorem $\lambda_{\min}(\cdot)$ refers to the smallest eigenvalue of a matrix.

Theorem 8 *Let the condition* (C) *hold for all* $i \in \{1, \ldots, M\}$ *and* $\ell \in \mathbb{N}$. *Then, for all* $\ell \in \mathbb{N}$, *the matrix* H_ℓ *is positive definite, and* $\lambda_{\min}(H_\ell) \geq \lambda_{\min}(B_{\mu,\lambda}) > 0$. *Moreover, the sequence* $\{H_\ell^{-1}\}_{\ell \in \mathbb{N}}$ *is uniformly bounded.*

In order to satisfy (C) in our algorithm we proceed as follows: Replace $((p_\ell)_i, (p_\ell)_{i+M})$ by $\max(1, \varrho(p_\ell)_i)^{-1}((p_\ell)_i, (p_\ell)_{i+M}) =: ((\hat{p}_\ell)_i, (\hat{p}_\ell)_{i+M})$ and check whether $(b_{\ell,i} + c_{\ell,i})^2 \leq 4 a_{\ell,i} d_{\ell,i}$ is satisfied by $((\hat{p}_\ell)_i, (\hat{p}_\ell)_{i+M})$. If this is not the case, then $b_{\ell,i}$ and $c_{\ell,i}$ are replaced by $\nu_{\ell,i} b_{\ell,i}$ and $\nu_{\ell,i} c_{k,i}$ with $\nu_{\ell,i} = 2\sqrt{a_{\ell,i} d_{\ell,i}}/|b_{\ell,i} + c_{\ell,i}|$; otherwise $b_{\ell,i}$ and $c_{\ell,i}$ are kept. After these modification, we obtain a positive definite matrix H_ℓ^+ which replaces H_ℓ in (B.3). It can be shown that $H_\ell^+ \to H_\ell$ as u_ℓ converges to the solution of the discrete version of (5.1).

The above considerations result in the following semismooth Newton solver.

Semismooth Newton for step 2 of the SA-TV-algorithm.
1: Initialize $(u_0, p_0) \in \mathbb{R}^M \times \mathbb{R}^{2M}$ and set $\ell := 0$.
2: Estimate the active sets, i.e., determine $\chi_{\mathcal{A}_{\ell+1}} \in \mathbb{R}^{2M \times 2M}$.
3: If the condition (C) is not satisfied, then compute H_ℓ^+; otherwise set $H_\ell^+ := H_\ell$.
4: Solve $H_\ell^+ \delta u = f_\ell$ for δu and let δu_ℓ denote the solution.
5: Use δu_ℓ to compute δp_ℓ by means of the second equation of (B.2).
6: Update $u_{\ell+1} := u_\ell + \delta u_\ell$, $p_{\ell+1} := p_\ell + \delta p_\ell$.
7: Stop; or set $\ell := \ell + 1$ and return to step 2.

Similar as in [21] it can be shown that this algorithm converges at a superlinear rate provided that u_0 is sufficiently close to the solution of the discrete version of (5.1). Moreover, the algorithm may be equipped with a line search or damping procedure. In this case, in addition to the fast local convergence the method converges globally, i.e., regardless of the initial choice u_0. For details on the globalization we refer to [21].

Finally we note that the system in step 4 of the above algorithm is solved iteratively by the BICGSTAB-method [28]. We further point out that we use an inexact Newton solver, i.e., the stopping tolerances for the BICGSTAB-solver become more and more stringent as we approach the solution.

References

1. USC-SIPI image database. University of Southern California. http://sipi.usc.edu/services/database/Database.html
2. Acar, R., Vogel, C.R.: Analysis of bounded variation penalty methods for ill-posed problems. Inverse Probl. **10**, 1217–1229 (1994)

3. Almansa, A., Ballester, C., Caselles, V., Haro, G.: A TV based restoration model with local constraints. J. Sci. Comput. **34**(3), 209–236 (2008)

4. Andrews, H.C., Hunt, B.R.: Digital Image Restoration. Prentice Hall, New York (1977)

5. Attouch, H., Buttazzo, G., Michaille, G.: Variational Analysis in Sobolev and BV Spaces. SIAM, Philadelphia (2005)

6. Aubert, G., Kornprobst, P.: Mathematical Problems in Image Processing. Partial Differential Equations and the Calculus of Variations. Springer, New York (2002)

7. Bertalmio, M., Caselles, V., Rougé, B., Solé, A.: TV based image restoration with local constraints. J. Sci. Comput. **19**, 95–122 (2003)

8. Bovik, A.: Handbook of Image and Video Processing. Academic Press, San Diego (2000)

9. Buades, A., Coll, B., Morel, J.M.: A review of image denoising algorithms, with a new one. Multiscale Model. Simul. **4**(2), 490–530 (2005)

10. Chambolle, A.: An algorithm for total variation minimization and application. J. Math. Imaging Vis. **20**, 89–97 (2004)

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

12. Chang, Q., Chern, I.-L.: Acceleration methods for total variation-based image denoising. SIAM J. Appl. Math. **25**, 982–994 (2003)

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

14. Ekeland, I., Témam, R.: Convex Analysis and Variational Problems. Classics Appl. Math., vol. 28. SIAM, Philadelphia (1999)

15. Facciolo, G., Almansa, A., Aujol, J.-F., Caselles, V.: Irregular to regular sampling, denoising and deconvolution. Multiscale Model. Simul. **7**(4), 1574–1608 (2009)

16. Galatsanos, N.P., Ketsaggelos, A.K.: Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation. IEEE Trans. Image Process. **1**, 322–336 (1992)

17. Gilboa, G., Sochen, N., Zeevi, Y.Y.: Texture preserving variational denoising using an adaptive fidelity term. In: Proceeding of the IEEE Workshop on Variational, Geometric and Level Set Methods in Computer Vision, Nice, France, pp. 137–144 (2003)

18. Giusti, E.: Minimal Surfaces and Functions of Bounded Variation. Birkhäuser, Boston (1984)

19. Gumbel, E.J.: Les valeurs extrêmes des distributions statistiques. Ann. Inst. Henri Poincaré **5**(2), 115–158 (1935)

20. Hintermüller, M., Kunisch, K.: Total bounded variation regularization as bilaterally constrained optimization problem. SIAM J. Appl. Math. **64**, 1311–1333 (2004)

21. Hintermüller, M., Stadler, G.: An infeasible primal-dual algorithm for total bounded variation-based inf-convolution-type image restoration. SIAM J. Sci. Comput. **28**(1), 1–23 (2006)

22. Mood, A.: Introduction to the Theory of Statistics. McGraw-Hill, New York (1974)

23. Osher, S., Burger, M., Goldfarb, D., Xu, J., Yin, W.: An iterative regularization method for total variation-based image restoration. Multiscale Model. Simul. **4**, 460–489 (2005)

24. Papoulis, A.: Probability, Random Variables, Stochastic Processes. McGraw-Hill, New York (1991)

25. Rudin, L.: MTV-multiscale total variation principle for a PDE-based solution to nonsmooth ill-posed problem. Technical report, Cognitech, Inc. Talk presented at the Workshop on Mathematical Methods in Computer Vision, University of Minnesota, 1995

26. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60**, 259–268 (1992)

27. Rudin, W.: Functional Analysis. TATA McGraw-Hill Publishing Company LTD., Noida (1974)

28. Saad, Y.: Iterative Methods for Sparse Linear Systems, 2nd edn. SIAM, Philadelphia (2003)

29. Strong, D., Aujol, J.-F., Chan, T.: Scale recognition, regularization parameter selection, and Meyer's G norm in total variation regularization. Technical report, UCLA, 2005

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

31. Strong, D., Chan, T.: Edge-preserving and scale-dependent properties of total variation regularization. Inverse Probl. **19**, 165–187 (2003)

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

33. Tadmor, E., Nezzar, S., Vese, L.: Multiscale hierarchical decomposition of images with applications to deblurring, denoising and segmentation. Commun. Math. Sci. **6**, 1–26 (2008)

34. Vogel, C.R.: Computational Methods for Inverse Problems. Frontiers Appl. Math., vol. 23. SIAM, Philadelphia (2002)

35. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. IEEE Trans. Image Process. **13**, 600–612 (2004)