
Higher-Order TV Methods - Enhancement via Bregman Iteration

Martin Benning · Christoph Brune ·
Martin Burger · Jahn Müller

In honor of Stanley Osher for his 70th birthday

June 7, 2012

Abstract In this work we analyze and compare two recent variational models for image denoising and improve their reconstructions by applying a Bregman iteration strategy. One of the standard techniques in image denoising, the ROF-model (cf. [ROF92]), is well known for recovering sharp edges of a signal or image, but also for producing staircase-like artifacts. In order to overcome these model-dependent deficiencies, total variation modifications that incorporate higher-order derivatives have been proposed (cf. [CL97, BKP10]). These models reduce staircasing for reasonable parameter choices. However, the combination of derivatives of different order leads to other undesired side effects, which we shall also highlight in several examples.

The goal of this paper is to analyze capabilities and limitations of the different models and to improve their reconstructions in quality by introducing Bregman iterations. Besides general modeling and analysis we discuss efficient numerical realizations of Bregman iterations and modified versions thereof.

Keywords Total Variation Regularization, Higher Order Methods, Staircasing, Exact Solutions, Bregman Iteration

1 Introduction

Total variation regularization, dating back to the fundamental work of Rudin, Osher, and Fatemi (cf. [ROF92]) has become a standard technique in image processing. As a precursor of the nowadays omnipresent ℓ^1 -minimization techniques, the approach is seeking sparsity of the gradient and is hence particularly successful in producing blocky images with sharp edges (cartoons).

Compared to other approaches like filtering in some frame system such as wavelets, curvelets, or shearlets, total variation methods can realize significantly sharper edges and overall more visually pleasing images, but on the other hand tend to create piecewise-constant images even in regions with smooth transitions of grey or color values in the original image. The latter artifact, usually called *staircasing*, is undesirable for some applications such

Martin Benning · Martin Burger · Jahn Müller
Westfälische Wilhelms-Universität Münster, Institut für Numerische und Angewandte Mathematik, Einsteinstr. 62, 48149 Münster, Germany
E-mail: {martin.benning, martin.burger, jahn.mueller}@wwu.de

Christoph Brune
Department of Mathematics, University of California Los Angeles, 405 Hilgard Avenue, Los Angeles, CA 90095-1555, USA
E-mail: brune@math.ucla.edu

as denoising, zooming, inpainting, or deblurring high resolution images or images with a natural gradient such as those in MRI. Although several improved methods tailored specially for denoising exist (cf. e.g. [BCM05, CS05, GO08]), total variation approaches still remain of highest importance for image processing and reconstruction, since unlike others the variational approach can be generalized directly to a variety of imaging tasks. It is hence important to study properties of total variation and modifications theoretically and numerically, where usually image denoising with additive Gaussian noise remains the model problem to gain understanding. We shall take the same route in this paper and consider modified total variation approaches for image denoising, although we have applications to other imaging problems in mind, which however would shadow the main arguments about the regularizations by the necessity to introduce and analyze more complicated data fidelities. Hence, we shall investigate variational problems of the form

$$u = \arg \min_{u \in \text{dom}(J)} \left\{ \frac{1}{2} \|u - f\|_{L^2(\Omega)}^2 + \alpha J(u) \right\}, \quad (1.1)$$

for different convex regularization functionals J . A special case of this general denoising scheme is the ROF-model (cf. [ROF92])

$$u_R = \arg \min_{u \in \text{BV}(\Omega)} \left\{ \frac{1}{2} \|u - f\|_{L^2(\Omega)}^2 + \alpha \text{TV}(u) \right\}, \quad (1.2)$$

with $\text{TV}(u)$ being the total variation of u defined as

$$\text{TV}(u) := \sup_{\substack{p \in C_0^\infty(\Omega; \mathbb{R}^n) \\ \|p\|_{L^\infty(\Omega; \mathbb{R}^n)} \leq 1}} \int_{\Omega} u \operatorname{div} p \, dx, \quad (1.3)$$

and with $\text{BV}(\Omega)$ denoting the space of functions u for which $\text{TV}(u)$ is finite, i.e.

$$\text{BV}(\Omega) = \{u \in L^2(\Omega) \mid \text{TV}(u) < \infty\}.$$

We shall investigate functionals J extending the original idea of total variation regularization, introduced in particular to cure staircasing effects. Such approaches based on combinations with higher order total variation functionals have been proposed and investigated recently, in particular the infimal convolution model (cf. [CL97]), denoted by ICTV, and a generalized total variation model (cf. [SS08, BKP10, SST11]), denoted by GTV, in the following. These techniques seem to yield visual improvement in some cases, on the other hand there are still some deficiencies visible in several examples and a detailed analysis of the models seems to be missing. In this paper we shall perform a detailed structural analysis of the ICTV and the GTV model, which highlights some important issues in both approaches and in particular explains why choosing weighting parameters between TV and higher order TV is delicate. Our analysis is based on comparing formulations and general properties of the functionals and associated optimality conditions for variational image denoising on the one hand, and on exact solutions of the denoising problems, related to eigenfunctions of the regularization functionals, on the other hand. The study of eigenfunctions allows to highlight what kind of images are favoured in variational methods with different regularization functionals and provides hints on remaining systematic errors of each approach. We also mention that we shall always work with a fixed parameter α , which we assume to be appropriately chosen. Since our main focus is to compare the impact of the regularization methods, we can compare minimizers of (1.1) at fixed α and avoid the discussion of choosing the parameter optimally in dependence of the noise and of asymptotics as α tends to zero.

As it was well-understood first for total variation regularization and later for all kinds of singular regularization methods, the systematic errors can be reduced or in some cases even completely eliminated by introducing Bregman iterations or inverse scale space methods (cf. [OBG⁺05, BGOX06]). In the context of the ROF-model, the

Bregman iteration allows to overcome the loss of contrast while still suppressing noise efficiently. Thus, it seems natural to try to overcome the deficiencies arising from the incorporation of higher-order differential operators by Bregmanizing the variational approach. We are going to demonstrate that in combination with higher-order total variation regularization the Bregman iteration will allow similar improvements and furthermore reduce undesired slope changes in homogeneous regions.

The paper is organized as follows: First of all we introduce the different models and regularization functionals, characterize their subdifferentials and provide several estimates between the functionals and the solutions of the variational problems. Afterwards we focus on exact solutions of the different models in order to highlight advantages and disadvantages with respect to each other. Subsequently we explain the computational realization of the different methods and compare computational and analytical results. In the final part of this paper we discuss Bregman iterations for infimal convolution and generalized total variation regularization. We give a brief motivation for its use and advantages towards standard regularization. We explain different ways of computational realization and compare the Bregman-iterated reconstructions with standard reconstructions.

2 Higher-Order TV Functionals and their Roles in $BV(\Omega)$

In the following we discuss some basic properties of the variational denoising scheme (1.1) for three different regularization functionals corresponding to the different models described in the introduction, namely the higher-order total variation, the infimal convolution regularization and the generalized total variation regularization. We verify that - up to a very low-dimensional nullspace - the latter two introduce the same topology as BV and give estimates between the functional values. Moreover, we discuss the subdifferentials of each functional and the optimality conditions for the variational denoising problems. Finally we provide estimates between the minimizers of the different functionals, yielding also insight into sources of difference between them.

2.1 Higher-Order Total Variation (TV^l)

A key property of modified total variation regularizations is that higher-order differential operators are incorporated into the regularizer. Thus, we want to extend the concept of total variation to higher-order derivatives first. The higher-order total variation regularization TV^l (cf. e.g. [Sch98, CMM00]) is defined as

$$TV^l(u) := \sup_{\substack{p \in C_0^\infty(\Omega; \text{Sym}^l(\mathbb{R}^n)) \\ \|p\|_\infty \leq 1}} \int_{\Omega} u \operatorname{div}^l p \, dx, \quad (2.1)$$

with $\|\cdot\|_\infty$ denoting the $L^\infty(\Omega; \text{Sym}^l(\mathbb{R}^n))$ -norm, and with $\text{Sym}^l(\mathbb{R}^n)$ denoting the space of symmetric l -tensors. For a definition of symmetric tensor fields we would like to refer to [BG80, Chapter2]; a detailed introduction of tensor spaces in the context of higher-order total variation methods can be found in [BKP10, Section 2]. Note that the $L^\infty(\Omega; \text{Sym}^l(\mathbb{R}^n))$ -norm and thus the characterization of dual variables heavily depends on the choice of the inner tensor-norm. Throughout this work we will focus on the ℓ^1 -tensor-norm only, leading to anisotropic total variation methods, which simplifies the analysis in some points. However we note that isotropic versions, i.e. ℓ^2 -type norms, are at least of equal interest in practice.

The subdifferential ∂TV^l can be characterized as

$$\partial TV^l(u) = \left\{ \operatorname{div}^l p \mid p \in L_0^\infty(\Omega; \text{Sym}^l(\mathbb{R}^n)), \|p\|_\infty \leq 1, \langle u, \operatorname{div}^l p \rangle = TV^l(u) \right\},$$

which can be seen by analogous reasoning as in case of the subdifferential of the standard total variation (i.e. $l = 1$) described in [Mey01, BO12].

However, in the following we will focus on the cases $l = 1$ and $l = 2$, for which the subdifferentials are

$$\partial \text{TV}(u) = \{ \text{div} p \mid p \in L^\infty(\Omega; \mathbb{R}^n), \|p\|_\infty \leq 1, p \cdot n = 0 \text{ on } \partial\Omega, \langle u, \text{div} p \rangle = \text{TV}(u) \} \quad (2.2)$$

and

$$\begin{aligned} \partial \text{TV}^2(u) = \{ & \text{div}^2 p \mid p \in L^\infty(\Omega; \text{Sym}^2(\mathbb{R}^n)), \|p\|_\infty \leq 1, p \cdot n = 0, \text{ and } (\text{div} p) \cdot n = 0 \text{ on } \partial\Omega, \\ & \langle u, \text{div}^2 p \rangle = \text{TV}^2(u) \}, \end{aligned} \quad (2.3)$$

with n denoting the outer unit normal to the boundary $\partial\Omega$.

Usually, promoting sparsity with respect to one particular higher order derivative is not desirable, e.g. TV^2 promotes piecewise linear solutions. However, combinations of various derivatives have been subject of recent interest and the higher order total variation as well as the original first-order one are the building blocks of those approaches. Two particular models for combining the TV and higher order TV^l will be presented in the two upcoming subsections.

2.2 Infimal Convolution Regularization (ICTV)

A regularization approach that has been introduced in the context of total variation denoising by Chambolle and Lions ([CL97]) is the use of infimal convolutions as regularization functionals. The infimal convolution of two functionals Φ and Ψ is defined as

$$(\Phi \square \Psi)(u) := \inf_{u=v+w} \Phi(v) + \Psi(w).$$

Although infimal convolution regularization can be used for very general functionals Φ and Ψ (cf. [BC11]), the particular interest in this work will lie on the infimal convolution of TV and TV^2 and its extensions.

Hence, we define the infimal convolution model (ICTV) concerning TV and TV^2 , by using of (2.1), as

$$\begin{aligned} \text{ICTV}_\beta(u) &:= (\text{TV} \square \text{TV}^2)(u) = \inf_{u=v+w} (\text{TV}(v) + \beta \text{TV}^2(w)) \\ &= \inf_{w \in \text{BV}^2(\Omega)} \sup_{\substack{p \in C_0^\infty(\Omega; \mathbb{R}^n) \\ \|p\|_\infty \leq 1}} \int_{\Omega} (u-w) \text{div}(p) \, dx + \beta \sup_{\substack{q \in C_0^\infty(\Omega; \text{Sym}^2(\mathbb{R}^n)) \\ \|q\|_\infty \leq 1}} \int_{\Omega} w \text{div}^2(q) \, dx, \end{aligned} \quad (2.4)$$

where we have abused the notation $\|\cdot\|_\infty$ for both the $L^\infty(\Omega; \text{Sym}^2(\mathbb{R}^n))$ - as well as the $L^\infty(\Omega; \text{Sym}^1(\mathbb{R}^n))$ -norm, for the sake of brevity. Equation (2.4) can be rewritten to

$$\text{ICTV}_\beta(u) = \sup_{\substack{p \in C_0^\infty(\Omega; \mathbb{R}^n) \\ \|p\|_\infty \leq 1}} \sup_{\substack{q \in C_0^\infty(\Omega; \text{Sym}^2(\mathbb{R}^n)) \\ \|q\|_\infty \leq 1}} \inf_w \int_{\Omega} (u-w) \text{div}(p) + \beta w \text{div}^2(q) \, dx$$

with the infimum for w being attained for $\text{div}(p) = \beta \text{div}^2(q)$. By inserting this relation and by substituting q with βq we end up with

$$\text{ICTV}_\beta(u) = \sup_{\substack{p \in C_0^\infty(\Omega; \mathbb{R}^n) \\ q \in C_0^\infty(\Omega; \text{Sym}^2(\mathbb{R}^n)) \\ \|p\|_\infty \leq 1, \|q\|_\infty \leq 1 \\ \beta \text{div}^2(q) = \text{div}(p)}} \int_{\Omega} u \text{div}(p) \, dx = \sup_{\substack{p \in C_0^\infty(\Omega; \mathbb{R}^n) \\ q \in C_0^\infty(\Omega; \text{Sym}^2(\mathbb{R}^n)) \\ \|p\|_\infty \leq 1, \|q\|_\infty \leq \beta \\ \text{div}^2(q) = \text{div}(p)}} \int_{\Omega} u \text{div}^2(q) \, dx. \quad (2.5)$$

Using the dual definition it is quite straight-forward to characterize the subdifferential of (2.5) as

$$\partial \text{ICTV}_\beta(u) = \left\{ \text{div}^2 q \mid q \in L^\infty(\Omega; \text{Sym}^2(\mathbb{R}^n)), p \in L^\infty(\Omega; \mathbb{R}^n), \|q\|_\infty \leq \beta, \|p\|_\infty \leq 1, \right. \\ \left. q \cdot n = 0, (\text{div} q) \cdot n, p \cdot n = 0 \text{ on } \partial\Omega, \text{div}^2 q = \text{div} p, \langle u, \text{div}^2 q \rangle = \text{ICTV}_\beta(u) \right\}. \quad (2.6)$$

In higher dimensions it has been observed that image reconstruction models with infimal convolution regularization do not separate a function into its TV and TV^2 structures properly. Consequently, a modified convolution model has been proposed that we are going to recall in the following subsection.

2.3 Generalized Total Variation (GTV)

Since the infimal convolution of TV and TV^2 did not yield the desired optimal separations of piecewise constant and piecewise linear regions for higher dimensions, modifications of (2.5) have been proposed in [BKP10] and [SST11]. We follow the definition of [BKP10] based on the dual formulation, while [SST11] discussed a modification in the discrete setting for the primal version of (2.5). Under appropriate discretizations of the functional in [BKP10], called *total generalized variation* there, both modifications coincide however at least in the case of TV and TV^2 .

The functional of [BKP10], which we rather call generalized total variation (GTV) for unification with TV and ICTV, is defined in the second order case as

$$\text{GTV}_\beta(u) := \sup_{\substack{q \in C_0^\infty(\Omega; \text{Sym}^2(\mathbb{R}^n)) \\ \|q\|_\infty \leq \beta, \|\text{div}(q)\|_\infty \leq 1}} \int_\Omega u \text{div}^2(q) \, dx. \quad (2.7)$$

Hence, comparing equation (2.7) with (2.5), the regularization functionals $\text{GTV}_\beta(u)$ and $\text{ICTV}_\beta(u)$ coincide if the constraint $p = \text{div}(q)$ for the dual functions holds.

The subdifferential of (2.7) is very similar to (2.6), with the difference that we have only a single dual variable and hence, no further equality constraint. The subdifferential reads as follows

$$\partial \text{GTV}_\beta(u) = \left\{ \text{div}^2 q \mid q \in L^\infty(\Omega; \text{Sym}^2(\mathbb{R}^n)), \|q\|_\infty \leq \beta, \|\text{div} q\|_\infty \leq 1, \right. \\ \left. q \cdot n = 0, (\text{div} q) \cdot n = 0 \text{ on } \partial\Omega, \langle u, \text{div}^2 q \rangle = \text{GTV}_\beta(u) \right\}. \quad (2.8)$$

2.4 Functional Estimates

In this section we compare basic properties of the three versions of total variation TV, ICTV_β , and GTV_β . A first result obtained from the way the parameter settings are used in the definition of $\text{ICTV}_\beta(u)$ and $\text{GTV}_\beta(u)$ is an ordering of the functional values:

Lemma 1 (Functional estimates) *For functions u in $BV(\Omega)$ and for a positive weighting parameter $\beta > 0$ the following estimates hold:*

$$\text{GTV}_\beta(u) \leq \text{ICTV}_\beta(u) \leq \text{TV}(u). \quad (2.9)$$

Proof The feasible set for the supremum in the definition of GTV_β in (2.7) is smaller than the feasible set in the definition of $ICTV_\beta$ in (2.5). In the special case of $p = \text{div}(q)$, the definitions coincide. Hence, we obtain

$$GTV_\beta(u) \leq ICTV_\beta(u).$$

Moreover, the feasible set for the supremum in the definition of $TV(u)$ includes the feasible set in $ICTV_\beta(u)$. This implies

$$ICTV_\beta(u) \leq TV(u).$$

It seems intuitive that one could expect estimates in the opposite direction using appropriate constants and appropriate subspaces of $BV(\Omega)$. For this sake we first characterize the nullspaces of the functionals:

Lemma 2 (Nullspace) *Let u be a function of bounded total variation. Then, the following equivalences hold*

- (i) $TV(u) = 0 \iff u$ is constant
- (ii) $ICTV_\beta(u) = 0 \iff u$ is affinely linear
- (iii) $GTV_\beta(u) = 0 \iff u$ is affinely linear.

Proof (i) is a well-known result (cf. e.g. [BO12] for full details).

(ii) Let $u \in BV(\Omega)$ be a function annihilating the infimal-convolution functional $ICTV_\beta(u)$. By considering the definition in (2.4), we obtain for $w \in BV^2(\Omega)$

$$TV^2(w) = 0 \quad \text{and} \quad TV(u - w) = 0,$$

because those functionals are positive in general. This implies that w is affinely linear and from (i) we obtain $u = w + c$ where $c \in \mathbb{R}$ denotes a positive constant. Hence we can conclude that u is affinely linear as well.

(iii) Let $GTV_\beta(u) = 0$, then from the definition of the generalized total variation we see that

$$\int_{\Omega} u \text{div}^2(q) \, dx = 0$$

for all $q \in C_0^\infty(\Omega)$ with

$$\|\text{div}(q)\|_{L^\infty(\Omega)} \leq 1 \quad \|q\|_{L^\infty(\Omega)} \leq \beta.$$

For arbitrary $p \in C_0^\infty(\Omega)$ however we can always find $c > 0$ such that $q = cp$ satisfies the above constraints and thus

$$\int_{\Omega} u \text{div}^2(p) \, dx = \frac{1}{c} \int_{\Omega} u \text{div}^2(q) \, dx = 0,$$

which implies that $\nabla^2 u = 0$ in the distributional sense, and hence u is affinely linear.

A useful result is the following rewritten version of the generalized total variation, which allows an augmentation by skew-symmetric tensors:

Lemma 3 *Let $n = 2$ and $u \in BV(\Omega)$, then*

$$GTV_\beta(u) \geq \sup_{\substack{q \in C^2(\Omega; \text{Sym}^2(\mathbb{R}^2)), r \in C^2(\Omega) \\ \|q\|_\infty \leq \beta, \|\text{div}(q)\|_\infty \leq 1 \\ (q+S(r)) \cdot n|_{\partial\Omega} = 0, \nabla \cdot (q+S(r)) \cdot n|_{\partial\Omega} = 0}} \int_{\Omega} u \text{div}^2(q) \, dx \quad (2.10)$$

with the matrix

$$S(r) = \begin{pmatrix} 0 & -r \\ r & 0 \end{pmatrix}.$$

Proof Let $u \in C^2(\Omega)$, then using integration by parts and a density argument it is straight-forward to see that

$$\text{GTV}_\beta(u) = \sup_{\substack{q \in C^2(\Omega; \text{Sym}^2(\mathbb{R}^2)) \\ \|q\|_\infty \leq \beta, \|\text{div}(q)\|_\infty \leq 1}} \int_{\Omega} \nabla^2 u : q \, dx,$$

where the colon indicates tensor contraction (generalized trace for tensors) involving two indices. Now we observe that for the skew-symmetric matrix $S(r)$ and the symmetric matrix $\nabla^2 u$ we have $\nabla^2 u : S(r) = 0$. Thus,

$$\text{GTV}_\beta(u) = \sup_{\substack{q \in C^2(\Omega; \text{Sym}^2(\mathbb{R}^2)), r \in C^2(\Omega) \\ \|q\|_\infty \leq \beta, \|\text{div}(q)\|_\infty \leq 1}} \int_{\Omega} \nabla^2 u : (q + S(r)) \, dx.$$

Now, restricting to the class of functions such that the boundary values $(q + S(r)) \cdot n$ and $\nabla \cdot (q + S(r)) \cdot n$ vanish the supremum does not increase and integrating by parts twice we obtain

$$\text{GTV}_\beta(u) \geq \sup_{\substack{q \in C^2(\Omega; \text{Sym}^2(\mathbb{R}^2)), r \in C^2(\Omega) \\ \|q\|_\infty \leq \beta, \|\text{div}(q)\|_\infty \leq 1 \\ (q+S(r)) \cdot n|_{\partial\Omega} = 0, \nabla \cdot (q+S(r)) \cdot n|_{\partial\Omega} = 0}} \int_{\Omega} u \, \text{div}^2(q + S(r)) \, dx = \sup_{\substack{q \in C^2(\Omega; \text{Sym}^2(\mathbb{R}^2)), r \in C^2(\Omega) \\ \|q\|_\infty \leq \beta, \|\text{div}(q)\|_\infty \leq 1 \\ (q+S(r)) \cdot n|_{\partial\Omega} = 0, \nabla \cdot (q+S(r)) \cdot n|_{\partial\Omega} = 0}} \int_{\Omega} u \, \text{div}^2(q) \, dx.$$

Lemma 4 (Functional Estimate, GTV and TV) *Let $\Omega = [a_1, b_1] \times [a_2, b_2] \subset \mathbb{R}^2$ and $\beta > 0$. Then there exists a constant $C = C(\beta) > 0$ such that for all $u \in BV$ with vanishing zeroth and first moment, i.e.*

$$\int_{\Omega} u(x) \, dx = 0 \quad \text{and} \quad \int_{\Omega} x u(x) \, dx = 0,$$

the estimate

$$\text{GTV}_\beta(u) \geq C \cdot \text{TV}(u) \tag{2.11}$$

holds.

Proof First of all we observe that the norms induced by $\text{GTV}_\beta(u)$ on the subspace of functions with vanishing zeroth and first moments for different β are equivalent, thus we may assume β to be sufficiently large. Furthermore, standard density arguments as for the analysis of total variation yield

$$\sup_{\substack{p \in C_0^\infty(\Omega; \mathbb{R}^2) \\ \|p\|_\infty \leq 1 \\ \int p \, dx = 0}} \int_{\Omega} u \, \text{div}(p) \, dx = \sup_{\substack{p \in C^1(\Omega; \mathbb{R}^2) \\ \|p\|_\infty \leq 1 \\ p \cdot n|_{\partial\Omega} = 0, \int p \, dx = 0}} \int_{\Omega} u \, \text{div}(p) \, dx$$

and

$$\text{TV}(u) = \sup_{\substack{p \in C_0^\infty(\Omega; \mathbb{R}^2) \\ \|p\|_\infty \leq 1}} \int_{\Omega} u \, \text{div}(p) \, dx = \sup_{\substack{p \in C^1(\Omega; \mathbb{R}^2) \\ \|p\|_\infty \leq 1 \\ p \cdot n|_{\partial\Omega} = 0}} \int_{\Omega} u \, \text{div}(p) \, dx.$$

Now let $x = (x_1, x_2)$ and $P_i(x_1, x_2)$ for $i = 1, 2$ be the second-order polynomial with

$$\begin{aligned} P_1(a_1, x_2) &= 0, & P_1(b_1, x_2) &= 0 & \forall x_2 \\ P_2(x_1, a_2) &= 0, & P_2(x_1, b_2) &= 0 & \forall x_1 \end{aligned}$$

and satisfying $\int_{\Omega} P_i(x) \, dx = 1$. For arbitrary $p \in C^1(\Omega; \mathbb{R}^2)$ such that $p \cdot n = 0$ on $\partial\Omega$ define

$$\tilde{p}_i(x) := \frac{1}{\gamma} \left(p_i(x) - P_i(x) \int_{\Omega} p_i(y) \, dy \right) \tag{2.12}$$

with $\gamma \in \mathbb{R}$. Then $\tilde{p} \in C^1(\Omega; \mathbb{R}^2)$ and by construction we see

$$\tilde{p} \cdot n|_{\partial\Omega} = 0, \quad \int_{\Omega} \tilde{p} \, dx = 0.$$

Additionally $\|\tilde{p}\|_{\infty} \leq 1$ holds for

$$\gamma \geq 1 + |\Omega| \max_{x \in \Omega} |P_i(x)|, \quad i = 1, 2.$$

Moreover, since $\operatorname{div}(P_1, P_2)$ is linear we find

$$\int_{\Omega} u \operatorname{div}(p) \, dx = \int_{\Omega} u \operatorname{div}(p) \, dx - \int_{\Omega} p \, dy \underbrace{\int_{\Omega} u \operatorname{div}(P_1, P_2) \, dx}_{=0} \stackrel{(2.12)}{=} \gamma \int_{\Omega} u \operatorname{div}(\tilde{p}) \, dx.$$

Hence,

$$\frac{1}{\gamma} \operatorname{TV}(u) = \sup_{\substack{p \in C^1(\Omega; \mathbb{R}^2) \\ \|p\|_{\infty} \leq 1 \\ p \cdot n|_{\partial\Omega} = 0}} \int_{\Omega} u \operatorname{div}\left(\frac{1}{\gamma} p\right) \, dx \leq \sup_{\substack{p \in C^1(\Omega; \mathbb{R}^2) \\ \|p\|_{\infty} \leq 1 \\ p \cdot n|_{\partial\Omega} = 0, \int_{\Omega} p \, dx = 0}} \int_{\Omega} u \operatorname{div}(p) \, dx. \quad (2.13)$$

Now we are going to apply Lemma 3. For this sake let $p \in C^1(\Omega; \mathbb{R}^2)$ with $\int_{\Omega} p \, dx = 0$ and $p \cdot n = 0$ on $\partial\Omega$. Define $w_i \in C^3(\Omega; \mathbb{R}^2)$ with mean zero as the solution of $\lambda \Delta w_i = p_i$ in Ω with homogeneous Neumann boundary conditions, which exists since the right-hand sides p_i fulfill the solvability condition for the Neumann problem. Now we define

$$q_{11} := \partial_{x_1} w_1, \quad q_{22} := \partial_{x_2} w_2, \quad q_{12} = q_{21} := \frac{1}{2}(\partial_{x_2} w_1 + \partial_{x_1} w_2), \quad r := \frac{1}{2}(-\partial_{x_2} w_1 + \partial_{x_1} w_2).$$

Then, by the regularity properties of the Poisson equation it is straightforward to show that $\|q\|_{\infty} \leq \beta$ for β sufficiently large and $\|\operatorname{div}(q)\|_{\infty} \leq 1$ for appropriate choice of λ independent of p . Moreover, one easily checks that $\lambda \operatorname{div}^2(q) = \operatorname{div}(p)$. Thus,

$$\sup_{\substack{p \in C^1(\Omega; \mathbb{R}^2) \\ \|p\|_{\infty} \leq 1 \\ p \cdot n|_{\partial\Omega} = 0, \int_{\Omega} p \, dx = 0}} \int_{\Omega} u \operatorname{div}(p) \, dx = \sup_{\substack{q \in C^2(\Omega; \operatorname{Sym}^2(\mathbb{R}^2)), r \in C^2(\Omega) \\ \|q\|_{\infty} \leq \beta, \|\operatorname{div}(q)\|_{\infty} \leq 1 \\ (q+S(r)) \cdot n|_{\partial\Omega} = 0, \nabla \cdot (q+S(r)) \cdot n|_{\partial\Omega} = 0}} \lambda \int_{\Omega} u \operatorname{div}^2(q) \, dx \stackrel{\text{Lemma 3}}{\leq} \lambda \operatorname{GTV}_{\beta}(u),$$

which together with (2.13) finishes the proof for $C = 1/(\lambda\gamma)$.

We mention that (2.11) can be shown easily also in spatial dimension one and with few technical additions also for rectangles in three dimensions, we chose the case of dimension two here in order to illustrate the proof technique.

The conclusion from the above results is that restricted to BV functions with vanishing zeroth and first moment the (semi-)norms of TV, ICTV and GTV are equivalent norms. The minimizers of the associated variational problems (1.1) can differ however, which we shall investigate below. Before going on, we verify that the functionals $\operatorname{GTV}_{\beta}$ and $\operatorname{ICTV}_{\beta}$ only differ in multiple dimensions:

Theorem 1 *Let $\Omega \subset \mathbb{R}$. Then for all $u \in BV(\Omega)$ the equality*

$$\operatorname{GTV}_{\beta}(u) = \operatorname{ICTV}_{\beta}(u) \quad (2.14)$$

holds.

Proof In the one-dimensional case we have

$$\text{ICTV}_\beta(u) = \sup_{\substack{p \in C_0^\infty(\Omega) \\ q \in C_0^\infty(\Omega) \\ \|p\|_\infty \leq 1, \|q\|_\infty \leq \beta \\ q'' = p'}} \int_{\Omega} u \operatorname{div}^2(q) \, dx. \quad (2.15)$$

Thus we obtain that q' and p in the feasible set for the supremum differ at most by a constant, which on the other hand must be zero since both have compact support. Thus, $p = q'$ and the definition coincides with the one for $\text{GTV}_\beta(u)$.

2.5 Variational Problems

In this section we return to the denoising scheme (1.1) with the functionals GTV_β and ICTV_β defined above. To fix notation we write

$$u_I = \arg \min_{u \in \text{BV}(\Omega)} \left\{ \frac{1}{2} \|u - f\|_{L^2(\Omega)}^2 + \alpha \text{ICTV}_\beta(u) \right\}, \quad (2.16)$$

and

$$u_G = \arg \min_{u \in \text{BV}(\Omega)} \left\{ \frac{1}{2} \|u - f\|_{L^2(\Omega)}^2 + \alpha \text{GTV}_\beta(u) \right\}. \quad (2.17)$$

The equivalence of the seminorms allows to adapt the proofs for the ROF-functional in a straight-forward way and conclude the existence and uniqueness of a minimizer. Moreover, using the characterization of nullspaces we immediately conclude the following property (again completely analogous to the proof in the ROF-case):

Proposition 1 *Let u_I, u_G be defined as above, then*

$$\int_{\Omega} f(x) \, dx = \int_{\Omega} u_I(x) \, dx = \int_{\Omega} u_G(x) \, dx \quad (2.18)$$

and

$$\int_{\Omega} x f(x) \, dx = \int_{\Omega} x u_I(x) \, dx = \int_{\Omega} x u_G(x) \, dx \quad (2.19)$$

Using the characterization of subdifferentials above it is straight-forward to state the optimality conditions for the denoising scheme in a more detailed way. Note that the optimality for (1.1) is given by

$$u - f + \alpha w = 0 \quad w \in \partial J(u). \quad (2.20)$$

Thus, we obtain an optimality condition for u_I (with associated vector field p_I and tensor q_I) as

$$\begin{aligned} u_I - f + \alpha \operatorname{div}^2 q_I &= 0 \\ \operatorname{div}^2 q_I - \operatorname{div} p_I &= 0 \\ \|p_I\|_\infty &\leq 1, \quad \|q_I\|_\infty \leq \beta \\ q_I \cdot n|_{\partial\Omega} &= 0, \quad (\operatorname{div} q_I) \cdot n|_{\partial\Omega} = p_I \cdot n|_{\partial\Omega} = 0 \\ \langle u_I, \operatorname{div}^2 q_I \rangle - \text{ICTV}_\beta(u_I) &= 0. \end{aligned}$$

In the same way we conclude the optimality for u_G (with associated tensor q_G) as

$$\begin{aligned} u_G - f + \alpha \operatorname{div}^2 q_G &= 0 \\ \|\operatorname{div} q_G\|_\infty &\leq 1, \quad \|q_G\|_\infty \leq \beta \\ q_G \cdot n|_{\partial\Omega} &= 0, \quad (\operatorname{div} q_G) \cdot n|_{\partial\Omega} = 0 \\ \langle u_G, \operatorname{div}^2 q_G \rangle - \operatorname{GTV}_\beta(u_G) &= 0. \end{aligned}$$

We mention that clearly u_I respectively u_G can be eliminated from the first line of the optimality condition to obtain a purely dual problem, similar to the case of the ROF functional (cf. e.g. [Cha04]), which is however not our main interest here.

The optimality condition also allows to gain some insight into structural properties of minimizers, as the next result shows:

Theorem 2 *Let $D \subset \Omega$ be an open simply connected set. If $\|q_G\| \leq \beta - \varepsilon$ and $\|\operatorname{div} q_G\| \leq 1 - \varepsilon$ in D for some $\varepsilon > 0$, then u_G is affinely linear in D . If $\|q_I\| \leq \beta - \varepsilon$ and $\|p_I\| \leq 1 - \varepsilon$ in D for some $\varepsilon > 0$, then u_I is affinely linear in D .*

Proof We only prove the result for u_G , the proof for u_I is analogous. Let $r \in C_0^\infty(D; \operatorname{Sym}^2(\mathbb{R}^n))$, and denote also its zero extension outside D by r . Then for $\delta > 0$ sufficiently small, we have

$$\|q_G \pm \delta r\| \leq \beta, \quad \|\operatorname{div}(q_G \pm \delta r)\| \leq 1.$$

Thus,

$$\int_\Omega u_G \operatorname{div}^2 q_G \, dx \geq \int_\Omega u_G \operatorname{div}^2 (q_G \pm \delta r) \, dx,$$

which implies

$$\int_D u_G \operatorname{div}^2 r \, dx = 0, \quad \forall r \in C_0^\infty(D; \operatorname{Sym}^2(\mathbb{R}^n)).$$

Hence $\nabla^2 u_G = 0$ in D , which implies that u_G is affinely linear in D .

Theorem 2 indicates that piecewise affinely linear is a typical structure for solutions u_G or u_I , in particular if the constraints on the tensors are not active on a set of positive measure. In spatial dimension one an analogous result as for the ROF model can be shown: For the ROF model it is known (cf. [CCN07]) that locally $u = f$ or u is constant. Since we have

$$q_G'' = \frac{1}{\alpha}(f - u_G) \in L^2(\Omega)$$

We conclude $q_G \in H^2(\Omega)$ and by embedding $q_G \in C^1(\Omega)$. Thus, by continuity $|q_G| = \beta$ on an open simply connected subset D is only possible if $q_G = \beta$ or $q_G = -\beta$ everywhere, i.e. q_G is constant in D . In analogous way we conclude that q_G' is constant in D if $|q_G'| = 1$ everywhere in D . In either case we obtain $q_G'' = 0$ and thus, from the optimality condition $u = f$.

We proceed to a comparison result between the minimizers of (1.1). For this sake we need the notion of the generalized Bregman distance (cf. [Kiw97]) related to a convex functional J . For a subgradient $w \in \partial J(u)$ we have

$$D_J^w(v, u) := J(v) - J(u) - \langle w, v - u \rangle. \quad (2.21)$$

In particular in singular cases like total variation regularization, Bregman distances have evolved as standard measures for error estimates after [BO04]. We can derive an error estimate from the optimality condition of (1.1) with functionals J_1 and J_2 such that $J_1(u) \leq J_2(u)$ for all u . The optimality condition is

$$u_i - f + \alpha w_i = 0 \quad w_i \in \partial J_i(u_i), \quad (2.22)$$

and subtracting it for $i = 1, 2$ we have

$$u_1 - u_2 + \alpha(w_1 - w_2) = 0 \quad w_i \in \partial J_i(u_i).$$

Now a duality product with $u_1 - u_2$ as well as adding or subtracting terms of the form $\alpha J_i(u_j)$ yields

$$\|u_1 - u_2\|^2 + \alpha D_{J_1}^{w_1}(u_2, u_1) + \alpha D_{J_2}^{w_2}(u_1, u_2) \leq \alpha(J_2(u_1) - J_1(u_1) + J_1(u_2) - J_2(u_2)).$$

Using that $J_1 \leq J_2$ we conclude the following result:

Theorem 3 *Let J_1 and J_2 be convex functionals such that $J_1 \leq J_2$ and let u_i be a minimizer of (1.1) for $J = J_i$. Then the estimate*

$$\|u_1 - u_2\|^2 + \alpha D_{J_1}^{w_1}(u_2, u_1) + \alpha D_{J_2}^{w_2}(u_1, u_2) \leq \alpha(J_2(u_1) - J_1(u_1))$$

holds. In particular, if $J_2(u_1) = J_1(u_1)$, then $u_1 = u_2$.

Theorem 3 provides an option to compare minimizers of different regularizations and since we have the ordering (2.9) we can apply the result with $J_1 = \text{GTV}_\beta$ and $J_2 = \text{TV}$ or $J_2 = \text{ICTV}_\beta$, as well as $J_1 = \text{ICTV}_\beta$ and $J_2 = \text{TV}$. In particular if $\text{GTV}_\beta(u_G) = \text{TV}(u_G)$, then $u_G = u_l = u_R$.

It thus remains to get a deeper understanding of why and when the difference between the values GTV_β and ICTV_β can typically occur, which we want to understand better in the case of the anisotropic total variation. The canonical structures for the modified total variation functionals appear to be piecewise linear functions with potential discontinuities in between. It can easily be generalized from the arguments for characteristic functions in [BKP10] that for $u(x) = a + b\chi_D(x)$ with χ_D being the characteristic function of a subdomain $D \subset \Omega$ with sufficiently regular boundary one has

$$\text{GTV}_\beta(u) = \text{ICTV}_\beta(u) = \text{TV}(u). \quad (2.23)$$

A more interesting example is of the form

$$u(x) = (a + c \cdot x) + (b + d \cdot x)\chi_D(x). \quad (2.24)$$

It is straightforward to see that

$$\int_{\Omega} u \text{div}^2 q \, dx = \int_{\partial D \cap \Omega} ([u] \text{div} q - q[\nabla u]) \cdot n \, d\sigma,$$

where $[u]$ and $[\nabla u]$ denote the jump of the function value and gradient, respectively, across ∂D . In the setup of the generalized total variation one can also show that

$$\int_{\Omega} u \text{div}^2 q \, dx = \int_{\partial D \cap \Omega} ([u]p - q[\nabla u]) \cdot n \, d\sigma.$$

Hence, $\text{GTV}_\beta(u)$ respectively $\text{ICTV}_\beta(u)$ can be computed by maximizing a functional concentrated on ∂D with the above constraints.

For the anisotropic total variation with ℓ^1 -tensor and -vector norms, canonical subdomains D are rectangles parallel to the coordinate axes. Moreover, canonical vectors c and d are multiples of the unit vectors. An example of this form with $d = (1, 0)$ in spatial dimension two is a standard test example for several image analysis tasks and has also been used to demonstrate advantages of modified TV functionals in [SST11]. Thus, let $\Omega = (-L, L)^2$ with $L > 1$, $D = (-1, 1)^2$, and $d = (1, 0)$. Then we have

$$\partial D = \Gamma_1 \cup \Gamma_{-1} \cup \Gamma_2 \cup \Gamma_{-2},$$

with

$$\Gamma_{\pm 1} = \{x \in \partial D \mid x_1 = \pm 1\}, \quad \Gamma_{\pm 2} = \{x \in \partial D \mid x_2 = \pm 1\}.$$

On $\Gamma_{\pm 1}$ we have $n = (\pm 1, 0)$ and hence

$$\int_{\Gamma_{\pm 1}} ([u] \operatorname{div} q - q[\nabla u]) \cdot n \, d\sigma = \pm \int_{\Gamma_{\pm 1}} ((b \pm d)(\partial_{x_1} q_{11} + \partial_{x_2} q_{12}) - dq_{11}) \, dx_2.$$

On $\Gamma_{\pm 2}$ we have $n = (0, \pm 1)$ and hence

$$\int_{\Gamma_{\pm 2}} ([u] \operatorname{div} q - q[\nabla u]) \cdot n \, d\sigma = \pm \int_{\Gamma_{\pm 2}} ((b + dx_1)(\partial_{x_1} q_{12} + \partial_{x_2} q_{22}) - dq_{12}) \, dx_1.$$

In spatial dimension two it is straightforward to characterize the nullspace of the divergence as rotations of a compactly supported scalar v , i.e. $(-\partial_{x_2} v, \partial_{x_1} v)$. I.e., in the case of ICTV, where $\operatorname{div}^2 q = \operatorname{div} p$ holds, we have

$$\partial_{x_1} q_{11} + \partial_{x_2} q_{12} = p_1 - \partial_{x_2} v \quad \partial_{x_1} q_{12} + \partial_{x_2} q_{22} = p_2 + \partial_{x_1} v.$$

Thus on, $\Gamma_{\pm 1}$ we find

$$\begin{aligned} \int_{\Gamma_{\pm 1}} (b \pm d)(\partial_{x_1} q_{11} + \partial_{x_2} q_{12}) \, dx_2 &= \int_{\Gamma_{\pm 1}} (b \pm d)p_1 \, dx_2 - (b \pm d) \int_{\Gamma_{\pm 1}} \partial_{x_2} v \, dx_2 \\ &= \pm \int_{\Gamma_{\pm 1}} [u] p \cdot n \, d\sigma - (b \pm d)(v(\pm 1, 1) - v(\pm 1, -1)). \end{aligned}$$

On $\Gamma_{\pm 2}$ we have

$$\begin{aligned} \int_{\Gamma_{\pm 2}} (b + dx_1)(\partial_{x_1} q_{12} + \partial_{x_2} q_{22}) \, dx_1 &= \int_{\Gamma_{\pm 2}} (b + dx_1)(p_2) \, dx_1 + \int_{\Gamma_{\pm 2}} (b + dx_1) \partial_{x_1} v \, dx_1 \\ &= \pm \int_{\Gamma_{\pm 2}} [u] p \cdot n \, d\sigma \pm (b \pm d)v(\pm 1, \pm 1) \mp (b \mp d)v(\mp 1, \pm 1) \mp d \int_{\Gamma_{\pm 2}} v \, dx_1. \end{aligned}$$

Thus, we find for admissible tensors and vector fields in the ICTV case

$$\int_{\Omega} u \operatorname{div}^2 q \, dx = \int_{\partial D \cap \Omega} ([u] p - q[\nabla u]) \cdot n \, d\sigma - d \int_{\Gamma_{+2}} v \, dx_1 + d \int_{\Gamma_{-2}} v \, dx_1.$$

This computation indicates that the main differences between $\operatorname{GTV}_{\beta}(u)$ and $\operatorname{ICTV}_{\beta}(u)$ will be caused by the second and third term, i.e. due to the difference of vector fields around $\Gamma_{\pm 2}$, i.e. the part where the jump of the function u is not constant across the boundary. As we shall see in the numerical examples, indeed staircasing can remain in the result of (2.16) close to this part of the discontinuity set, while eliminated in the solution of (2.17).

3 Exact Solutions

A famous example was given by Meyer [Mey01] in the case of the two-dimensional, isotropic ROF-model on an infinite dimensional set. There the characteristic function of a circle can be recovered almost exactly, only suffering a loss of contrast. Further examples of exact solutions are given by Strong (cf. [SC03]) and systematically in the context of eigenfunctions in [ACC05, Ben11]. It is also well known that the application of inverse scale space methods or Bregman iteration can compensate this loss of contrast in finite time or a finite number of iterations, respectively (cf. [OBG⁺05, BGOX06, Ben11]). In this section we therefore want to compare exact solutions for the different models in a one-dimensional setting, in order to give an intuition of what kind of solutions we can expect to be recovered almost exactly by variational denoising schemes, respectively exactly with a Bregman iteration strategy as we shall see in Section 6, even in the presence of noise.

3.1 Total Variation Regularization ($l = 1$)

As in Meyer's example we want to define the function $u_{\text{block}} : [-L, L] \subset \mathbb{R} \rightarrow \{-1/\sqrt{2L}, 1/\sqrt{2L}\}$ with

$$u_{\text{block}}(x) := \begin{cases} \frac{1}{\sqrt{2L}} & x \in [-\frac{L}{2}, \frac{L}{2}] \\ -\frac{1}{\sqrt{2L}} & \text{else} \end{cases}, \quad (3.1)$$

and we are going to show that in case of $f = u_{\text{block}}$ the unique solution of the ROF-model satisfies $\hat{u} = cu_{\text{block}}$ with $c = 1 - 2\alpha\sqrt{2/L}$, for $\alpha < \sqrt{L/2}/2$.

Computing the optimality condition of ROF and inserting $f = u_{\text{block}}$ and $\hat{u} = cu_{\text{block}}$ yields

$$\hat{w} = \frac{1}{\alpha} (f - \hat{u}) = 2u_{\text{block}}\sqrt{2/L}, \quad (3.2)$$

for $\hat{w} \in \partial\text{TV}(\hat{u}) = \partial\text{TV}(u_{\text{block}})$. Thus, it remains to be shown that $2u_{\text{block}}\sqrt{2/L}$ indeed is a subgradient. The subdifferential of TV can be characterized via (2.2). Hence, in one dimension we simply have to find a function p such that $p' = 2u_{\text{block}}\sqrt{2/L}$ holds in a weak sense, and with p satisfying $\|p\|_{\infty} = 1$, $p(L) = p(-L) = 0$ and $\langle p', \hat{u} \rangle = \text{TV}(\hat{u})$. If we consider $p : [-L, L] \rightarrow [-1, 1]$ with

$$p(x) := \frac{2}{L} \begin{cases} x & x \in [-\frac{L}{2}, \frac{L}{2}] \\ -x + L & x \in [\frac{L}{2}, L] \\ -x - L & x \in [-L, -\frac{L}{2}] \end{cases}, \quad (3.3)$$

we easily see that p satisfies $p(-L) = p(L) = 0$, $\|p\|_{\infty} = p(L/2) = 1$, $p' = (2u_{\text{block}})\sqrt{2/L}$ (in a weak sense) and $\langle p', \hat{u} \rangle = c\langle p', u_{\text{block}} \rangle = 2c\sqrt{2/L} \int_{-L}^L u_{\text{block}}^2 dx = 2c\sqrt{2/L} = c\text{TV}(u_{\text{block}}) = \text{TV}(\hat{u})$. Thus, $2u_{\text{block}}\sqrt{2/L}$ is a subgradient, and therefore (3.2) is fulfilled.

3.2 Second-Order Total Variation Regularization ($l = 2$)

After having considered a simple example that remains - despite the loss of contrast - almost invariant with respect to ROF minimization we want to find a similar example in case of second-order total variation TV^2 , with its subdifferential being characterized by (2.3). Again, we restrict ourselves to one dimension and consider the function $u_{\text{hat}} : [-L, L] \rightarrow [-\sqrt{3/(2L)}, \sqrt{3/(2L)}]$ with

$$u_{\text{hat}}(x) := \sqrt{\frac{6}{L^3}} \left(\frac{L}{2} - |x| \right). \quad (3.4)$$

We are going to show that the unique solution \hat{u} of (1.1) with $J(u) = \text{TV}^2(u)$ is given via $\hat{u} = cu_{\text{hat}}$, with $c = 1 - 2\alpha\sqrt{6/(L^3)}$, for $\alpha < \sqrt{L^3}/24$. Similar to the previous section we simply have to verify that the optimality condition

$$\hat{w} = \frac{1}{\alpha} (f - \hat{u}) = 2u_{\text{hat}}\sqrt{6/(L^3)},$$

for $\hat{w} \in \partial\text{TV}^2(\hat{u}) = \partial\text{TV}^2(u_{\text{hat}})$, is satisfied. In case of $\partial\text{TV}^2(u_{\text{hat}})$ we therefore have to find a function q satisfying $q'' = 2u_{\text{hat}}\sqrt{6/(L^3)}$ (in a weak sense), $q(L) = q(-L) = 0$, $\|q\|_{\infty} \leq 1$ and $\langle q'', u_{\text{hat}} \rangle = \text{TV}^2(u_{\text{hat}})$. By considering the function $q : [-L, L] \rightarrow [0, 1]$ with

$$q(x) := \frac{3}{L^2}x^2 - \frac{2}{L^3}|x|^3 - 1$$

we easily see that this specific q satisfies all these conditions, since we have $\text{TV}^2(u_{\text{hat}}) = 2\sqrt{\frac{6}{L^3}}$.

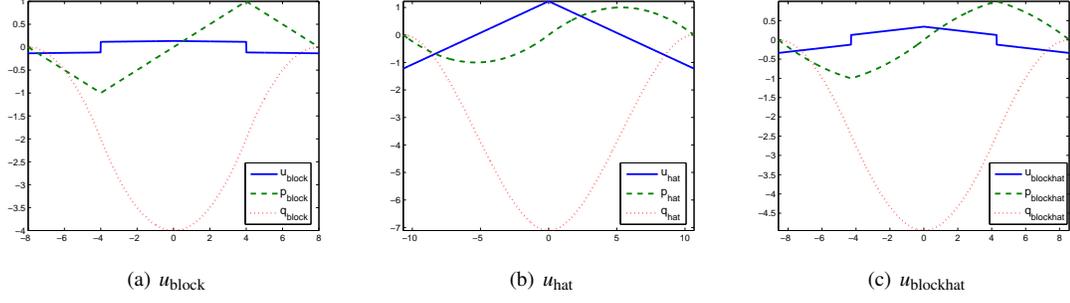


Fig. 1 The functions u_{block} , u_{hat} and u_{blockhat} , and their corresponding dual variables, for the interval lengths $L = 8$, $L = 32/3$ and $L = 32/(2 + \sqrt{3})$, respectively.

3.3 Infimal Convolution and Generalized Total Variation Regularization

In the previous sections we have seen two functions which can be recovered almost exactly by the ROF model (1.2) and model (1.1) with $J(u) = \text{TV}^2(u)$ as a regularizer respectively, suffering only a loss of contrast. However, neither is the ROF model able to recover u_{hat} as defined in (3.4), nor does model (1.1) with $J(u) = \text{TV}^2(u)$ allow to recover u_{block} as defined in (3.1) without introducing a slope into its constant parts and eliminating the discontinuity. However, it is easy to compute that both (3.1) and (3.4) can be recovered by (2.16) or (2.17), respectively (cf. [Ben11]). The remaining dual variables in case of infimal convolution, respectively the only dual variables in case of generalized total variation can be characterized via

$$q(x) := \frac{2}{L} \begin{cases} \frac{1}{2}x^2 - \frac{1}{4}L^2 & x \in [-\frac{L}{2}, \frac{L}{2}] \\ Lx - \frac{1}{2}x^2 - \frac{1}{2}L^2 & x \in]\frac{L}{2}, L] \\ -Lx - \frac{1}{2}x^2 - \frac{1}{2}L^2 & x \in [-L, -\frac{L}{2}[\end{cases},$$

for the choice of u_{block} , with $\beta \geq L/2$, and

$$q(x) := \frac{2}{L}x^2 - \frac{4}{3L^2}|x|^3 - \frac{2}{3}L, \quad (3.5)$$

for the choice of u_{hat} , with $\beta \geq (2L)/3$. In the latter case the loss of contrast modifies to $c = 1 - \alpha\sqrt{32/(3L)}$, due to the requirement $\|q'\|_\infty \leq 1$.

Moreover, we can combine these functions additively, to obtain a function that can neither be recovered by ROF nor by model (1.1) with $J(u) = \text{TV}^2(u)$, but by (1.1) with either $J(u) = \text{ICTV}_\beta(u)$ or $J(u) = \text{GTV}_\beta(u)$. We therefore want to validate that for input data given in terms of the function

$$u_{\text{blockhat}} : [-L, L] \rightarrow \left[-\sqrt{(1 + \sqrt{3})/(2L(2 + \sqrt{3}))}, \sqrt{(1 + \sqrt{3})/(2L(2 + \sqrt{3}))} \right]$$

defined as

$$u_{\text{blockhat}}(x) := \frac{1}{\sqrt{2 + \sqrt{3}}} \begin{cases} \frac{1}{\sqrt{2L}} + \sqrt{\frac{6}{L^3}} \left(\frac{L}{2} - |x| \right) & x \in [-\frac{L}{2}, \frac{L}{2}] \\ -\frac{1}{\sqrt{2L}} + \sqrt{\frac{6}{L^3}} \left(\frac{L}{2} - |x| \right) & \text{else} \end{cases}$$

the solution of (1.1) with $J(u) = \text{ICTV}_\beta(u)$ satisfies $\hat{u} = cu_{\text{blockhat}}$, for $c = 1 - 4\alpha\sqrt{2/(L(2+\sqrt{3}))}$ and $\beta \geq L/\sqrt{3}$.

The subdifferential of (2.5) can be characterized via (2.6). Thus, as in the previous examples we need to ensure the existence of two functions q and p that satisfy the subdifferential properties and for which we obtain $q'' = p' = (\sqrt{L}(2+\sqrt{3})u_{\text{blockhat}})/(4(1+\sqrt{3}))$. If we consider the functions

$$p(x) = \frac{4\sqrt{2}}{(2+\sqrt{3})\sqrt{L}} \begin{cases} \frac{1}{\sqrt{2L}}x + \sqrt{\frac{6}{L^3}}\left(\frac{L}{2}x - \frac{1}{2}\text{sign}(x)|x|^2\right) & x \in [-\frac{L}{2}, \frac{L}{2}] \\ -\frac{1}{\sqrt{2L}}x + \sqrt{\frac{6}{L^3}}\left(\frac{L}{2}x - \frac{1}{2}x^2\right) + \sqrt{\frac{L}{2}} & x \in]\frac{L}{2}, L] \\ -\frac{1}{\sqrt{2L}}x + \sqrt{\frac{6}{L^3}}\left(\frac{L}{2}x + \frac{1}{2}x^2\right) - \sqrt{\frac{L}{2}} & x \in [-L, -\frac{L}{2}[\end{cases}$$

and

$$q(x) = \frac{4}{(2+\sqrt{3})L} \begin{cases} \frac{1+\sqrt{3}}{2}x^2 - \frac{1}{\sqrt{3L}}|x|^3 - L^2\left(\frac{2+\sqrt{3}}{4\sqrt{3}}\right) & x \in [-\frac{L}{2}, \frac{L}{2}] \\ -\frac{1-\sqrt{3}}{2}x^2 - \frac{1}{\sqrt{3L}}x^3 + Lx - \frac{L^2}{2}\left(\frac{1+\sqrt{3}}{\sqrt{3}}\right) & x \in]\frac{L}{2}, L] \\ -\frac{1-\sqrt{3}}{2}x^2 + \frac{1}{\sqrt{3L}}x^3 - Lx - \frac{L^2}{2}\left(\frac{1+\sqrt{3}}{\sqrt{3}}\right) & x \in [-L, -\frac{L}{2}[\end{cases},$$

we see that $q(L) = q(-L) = p(L) = p(-L) = 0$ holds, and that we obtain $\|p\|_\infty = 1$ and $\|q\|_\infty = L/\sqrt{3} = \beta$. Furthermore, we can verify either $q'' = p' = (4u_{\text{blockhat}})/\sqrt{2/(L(2+\sqrt{3}))}$ and $\langle u_{\text{blockhat}}, q'' \rangle = \text{ICTV}_{\frac{L}{\sqrt{3}}}(u_{\text{blockhat}})$, since

$$\text{ICTV}_{\frac{L}{\sqrt{3}}}(u_{\text{blockhat}}) = \text{TV}(u_{\text{block}}) + (L \text{TV}^2(u_{\text{hat}}))/\sqrt{3} = \left(2\sqrt{2/L} + L/\sqrt{3}\sqrt{6/(L^3)}\right) / \sqrt{2+\sqrt{3}} = 4\sqrt{2/(L(2+\sqrt{3}))}$$

is satisfied.

Before continuing with two dimensional examples we want to point out that a piecewise-linear function orthogonal to any affine-linear function cannot automatically be recovered by (2.16) or (2.17) in the above sense. Let us, for instance, consider the function

$$u(x) := \begin{cases} -\frac{7}{4}x & x \in [-L/2, L/2] \\ \frac{1}{4}x & \text{else} \end{cases},$$

for $x \in [-L, L]$. We make the following attempt for the corresponding dual variable. We define

$$q(x) := \frac{1}{L^2} \begin{cases} -\frac{7}{3}x^3 + L^2x & x \in [-L/2, L/2] \\ \frac{1}{3}x^3 - L^2x + \frac{2}{3}L^3 & x \in]L/2, L] \\ \frac{1}{3}x^3 - L^2x - \frac{2}{3}L^3 & x \in [-L, -L/2[\end{cases}.$$

We can easily compute that q satisfies the properties $q'' = \lambda u$ for $\lambda = 8/(L^2)$ (in a weak sense), $q(L) = q'(-L) = q(-L) = 0$, $\|q\|_\infty = (2L\sqrt{7})/21 := \beta$ and $\|q'\|_\infty = 1$. However, we cannot guarantee

$$\langle u, q'' \rangle = \text{GTV}_{(2L\sqrt{7})/21}(u) = \text{ICTV}_{(2L\sqrt{7})/21}(u).$$

Computing the dual product actually yields $\langle u, q'' \rangle = (7L)/3$. In addition, due to (2.4), we would actually assume $\text{ICTV}_{(2L\sqrt{7})/21}(u) = \text{TV}(v) + \beta \text{TV}^2(w)$ to hold, for

$$v(x) = \begin{cases} 0 & x \in [-L/2, L/2] \\ L & x \in]L/2, L] \\ -L & x \in [-L, -L/2[\end{cases}$$

and

$$w(x) = \begin{cases} -\frac{7}{4}x & x \in [-L/2, L/2] \\ \frac{1}{4}x - L & x \in]L/2, L] \\ \frac{1}{4}x + L & x \in [-L, -L/2[\end{cases}.$$

However, for these two functions we obtain $\text{TV}(v) = 2L$ and $\text{TV}^2(w) = L$, and thus,

$$\text{GTV}_{(2L\sqrt{7})/21}(u) = \text{ICTV}_{(2L\sqrt{7})/21}(u) = 2L + (2L^2\sqrt{7})/21 \neq (7L)/3.$$

As a consequence, we have failed to construct a function q satisfying all subdifferential properties and hence, λu is not a subgradient. Therefore we cannot expect to recover a multiple of u by applying (1.1) with either $J = \text{GTV}_\beta$ or $J = \text{ICTV}_\beta$.

Due to Theorem 1 the solutions of ICTV_β and GTV_β do not differ in one spatial dimension. Thus, in order to observe differences between ICTV_β and GTV_β at least the two-dimensional case needs to be considered. First of all, we can simply extend the one-dimensional examples from above to two dimensions by a simple constant extension of the corresponding dual variables in each dimension. If we pick e.g. the corresponding dual variable p for the function u_{hat} defined as the derivative of (3.5), i.e.

$$p(x) = \frac{4}{L} \left(x - \frac{1}{L^2} \text{sign}(x)|x|^2 \right),$$

we can define a vector field based on p simply via $p_x(x, y) = p(x)$ and $p_y(x, y) = p(y)$. The corresponding function u_{hat}^{2D} obtained via $c \text{div } p = u_{\text{hat}}^{2D}$, with p denoting (p_x, p_y) in this case and c being a constant guaranteeing $\|u_{\text{hat}}^{2D}\|_{L^2(\Omega)} = 1$, thus reads as

$$u_{\text{hat}}^{2D}(x, y) = \frac{\sqrt{3/2}}{L^2} (L - |x| - |y|),$$

for $(x, y) \in [-L, L]^2$. In a similar fashion we can extend u_{block} to

$$u_{\text{block}}^{2D}(x, y) = \frac{1}{L\sqrt{2}} \begin{cases} 1 & (x, y) \in [-\frac{L}{2}, \frac{L}{2}]^2 \\ 0 & ((|x| > \frac{L}{2}) \wedge (|y| \leq \frac{L}{2})) \vee ((|y| > \frac{L}{2}) \wedge (|x| \leq \frac{L}{2})) \\ -1 & \text{else} \end{cases},$$

and u_{blockhat} to

$$u_{\text{blockhat}}^{2D}(x, y) = \sqrt{\frac{2}{4 - \sqrt{3}}} (u_{\text{block}}^{2D}(x, y) + u_{\text{hat}}^{2D}(x, y)).$$

3.4 Staircasing

In [CL97, BKP10, SST11] it has already been pointed out that the infimal convolution- as well as the generalized total variation-model significantly reduce staircasing artifacts. We want to illustrate this fact with a small example based on the function u_{blockhat} of the previous section.

Let us therefore assume that we either want to solve (2.16) or (2.17) with input data f given in terms of $f(x) = u_{\text{blockhat}}(x) + n(x)$, with an additional noise term defined as

$$n(x) := A \left(2 \cos \left(\frac{2\pi k}{L} x \right)^2 - 1 \right), \quad (3.6)$$

for $A \in \mathbb{R}$ being the amplitude and $k \in \mathbb{Z} \setminus \{0\}$ representing the frequency of the noise. It is completely clear that for this kind of data the ROF model will produce staircasing artifacts in the reconstruction (for any choice of α), which we are also going to demonstrate in Section 5. For now we want to investigate under which conditions on A and k it is however possible to recover u_{blockhat} only with a loss of contrast via (2.16) or (2.17). If we again consider the optimality condition

$$\hat{w} = \frac{1}{\alpha} (f - \hat{u}) = \frac{1}{\alpha} (u_{\text{blockhat}} + n - \hat{u}),$$

for $\hat{w} \in \partial \text{GTV}_\beta(\hat{u})$ (or $\hat{w} \in \partial \text{ICTV}_\beta(\hat{u})$ respectively), and assume $\hat{u} = c u_{\text{blockhat}}$, then we obtain

$$\hat{w} = \frac{1-c}{\alpha} u_{\text{blockhat}} + \frac{1}{\alpha} n.$$

We know from the previous section that $\lambda u_{\text{blockhat}} \in \partial \text{GTV}_{L/\sqrt{3}}(u_{\text{blockhat}}) = \partial \text{GTV}_{L/\sqrt{3}}(\hat{u})$ is a subgradient, for $\lambda = 4\sqrt{2/(L(2+\sqrt{3}))}$. Thus, for $\alpha = (1-c)/\lambda$ we have to ensure that n/α does not affect the subdifferential properties in order to guarantee that $\lambda u_{\text{blockhat}} + n/\alpha$ is a subgradient as well. It is easy to see that for N defined as

$$N(x) := \frac{AL^2}{8\pi^2 k^2} \sin \left(\frac{2\pi k}{L} x \right)^2$$

we have

$$N'(x) = \frac{AL}{2\pi k} \sin \left(\frac{2\pi k}{L} x \right) \cos \left(\frac{2\pi k}{L} x \right),$$

and $N''(x) = n(x)$. Moreover, we immediately discover $N(L) = N(-L) = 0$, $N'(L) = N'(-L) = 0$, and $\int_{-L}^L N'(x) dx = \int_{-L}^L n(x) dx = 0$. Consequently, adding up N and N' to the dual variables p and q (defined as in the previous section) does neither affect the boundary conditions, nor does it violate the condition $\langle q'' + n, u_{\text{blockhat}} \rangle = \text{GTV}_{L/\sqrt{3}}(u_{\text{blockhat}})$. The only conditions that might get violated are $\|p\|_\infty = 1$ and $\|q\|_\infty = L/\sqrt{3}$. Note however that we have $N(0) = 0$ and $N'(-L/2) = N'(L/2) = 0$ and thus, the noise does not directly affect the critical points of the dual variables. Hence, for properly chosen parameters c (or α , respectively), A and k it is possible to recover u_{blockhat} only with a loss of contrast and without staircasing artifacts. We are going to support this little theoretic result by computational results in Section 5.

Prior to that - in order to obtain more interesting and complicated examples - we shall consider numerical solutions motivated by the above analysis. For this sake we shall discuss some methods for the numerical solution of the variational problems in the next section and then proceed to numerical results.

4 Numerical Methods

In the following we discuss two approaches to compute minimizers of (2.16) and (2.17). We describe only one of these models in details, i.e. GTV, and briefly mention the modifications for the other one. For simplicity we give formal statements instead of all detailed notions of functionsspaces.

4.1 Newton Method for TV, ICTV and GTV

We first discuss a Newton-type method for the optimality conditions using additional approximation of constraints to obtain at least semismoothness. The constraints on q and $\operatorname{div} q$ (or p) can be realized via the so-called penalty method. The regularization term (2.7) can be approximated by

$$\sup_q \int_{\Omega} u \left[\operatorname{div}^2 q - \frac{1}{\varepsilon} P(\|q\| - \beta) - \frac{1}{\varepsilon} P(\|\operatorname{div} q\| - 1) \right] dx$$

with $\varepsilon > 0$ tending to zero and a term P penalizing if the constraint is violated. A typical example for such a P is

$$P(s) = \frac{1}{2} \max\{s, 0\}^2.$$

We will use the ℓ^1 tensor-norm and hence set the penalty term for an n dimensional entry

$$P_c(s) := \frac{1}{2} \sum_{i=1}^n \max\{|s_i| - c, 0\}^2.$$

Setting $p = \operatorname{div} q$ and realizing this new equality constraint via the method of Lagrange multipliers we obtain

$$\inf_{\mu} \sup_{p, q} \int_{\Omega} u \left[\operatorname{div}^2 q - \frac{1}{\varepsilon} P_{\beta}(q) - \frac{1}{\varepsilon} P_1(p) \right] dx + \int_{\Omega} \mu (p - \operatorname{div} q). \quad (4.1)$$

ICTV can be approximated in a similar way, the only difference is that the equality constraint here is given by $\operatorname{div}^2 q = \operatorname{div} p$. For first-order total variation the approximation is even simpler since there is only one constraint on p and no additional equality constraint.

Regarding now the denoising problem 2.17 with $\alpha = 1/\lambda$ and using a penalty approximation as before, yields the following saddle point problem:

$$\inf_{u, \mu} \sup_{p, q} \frac{\lambda}{2} \int_{\Omega} (u - f)^2 dx + \int_{\Omega} \left[u \operatorname{div}^2 q - \frac{1}{\varepsilon} P_{\beta}(q) - \frac{1}{\varepsilon} P_1(p) \right] dx + \int_{\Omega} \mu (p - \operatorname{div} q) dx.$$

The first order optimality conditions then yield

$$\begin{aligned} \lambda(u - f) + \operatorname{div}^2 q &= 0 \\ p - \operatorname{div} q &= 0 \\ -\frac{1}{\varepsilon} P'_1(p) + \mu &= 0 \\ -\frac{1}{\varepsilon} P'_{\beta}(q) + \nabla^2 u + \nabla \mu &= 0. \end{aligned}$$

We now linearize the nonlinear terms $P'_c(s)$ via a first-order Taylor-approximation, i.e.

$$P'_c(s^{k+1}) \approx P'_c(s^k) + P''_c(s^k)(s^{k+1} - s^k),$$

where we use P'' in the sense of a semismooth Newton method (cf. [HPUU09]). Adding damping terms for p, q and μ we have to solve the following linear system in each step:

$$\begin{aligned} \lambda(u^{k+1} - f) + \operatorname{div}^2 q^{k+1} &= 0 \\ p^{k+1} - \operatorname{div} q^{k+1} &= 0 \\ -\frac{1}{\varepsilon} \left(P'_1(p^k)(p^{k+1} - p^k) - P'_1(p^k) \right) - \tau_p^k(p^{k+1} - p^k) + \mu^{k+1} - \tau_\mu^k(\mu^{k+1} - \mu^k) &= 0 \\ -\frac{1}{\varepsilon} \left(P'_\beta(q^k)(q^{k+1} - q^k) - P'_\beta(q^k) \right) - \tau_q^k(q^{k+1} - q^k) + \nabla^2 u^{k+1} + \nabla \mu^{k+1} &= 0. \end{aligned}$$

We mention that in the course of the algorithm it is not necessary to restrict to constant equal values ε , but we can rather choose two different sequences ε_p^k and ε_q^k tending to zero.

As usual for Newton-type methods it remains to efficiently solve the linear system in each iteration step. After discretization the arising matrix equation $Ax = b$ is given by

$$A = \begin{pmatrix} C & D \\ D^t & B \end{pmatrix} \quad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

with

$$\begin{aligned} C &= \begin{pmatrix} \lambda I_N & 0 & 0 \\ 0 & -\frac{1}{\varepsilon} P'_1(p^k) I_{2N} & I_{2N} \\ 0 & I_{2N} & \tau_\mu I_{2N} \end{pmatrix} \quad D = \begin{pmatrix} \operatorname{div}^2 \\ 0 \\ -\operatorname{div} \end{pmatrix} \quad B = \begin{pmatrix} -\frac{1}{\varepsilon} P'_\beta(q) I_{2N} \end{pmatrix} \\ x_1 &= \begin{pmatrix} u^{k+1} \\ p^{k+1} \\ \mu^{k+1} \end{pmatrix} \quad x_2 = (q^{k+1}) \quad b_1 = \begin{pmatrix} \lambda f \\ (-\frac{1}{\varepsilon} P'_1(p^k) - \tau_p^k) p^k + P'_1(p^k) \\ \tau_\mu \mu^k \end{pmatrix} \quad b_2 = \left(\left(-\frac{1}{\varepsilon} P'_\beta(q^k) - \tau_q^k \right) q^k + P'_\beta(q^k) \right). \end{aligned}$$

Here I_N denotes the identity matrix of size $N = \prod_{i=1}^n n_i$ where n_i is the number of pixels in dimension i . For simplicity we will discuss only the two dimensional case (i.e. $n = 2$ and $N = n_1 * n_2$). As the basis for the discretization of the divergence matrices we will use the forward difference matrix with Neumann boundary conditions:

$$D_i := \begin{pmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & 0 & \dots & 0 \\ \vdots & 0 & -1 & 1 & \ddots & \vdots \\ 0 & \dots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & -1 & 1 \\ 0 & \dots & \dots & 0 & 0 & 0 \end{pmatrix} \in \mathbb{R}^{n_i \times n_i} \quad \text{for } i = 1, 2.$$

Then the derivatives in two dimensions can be written as:

$$\begin{aligned} D_x &:= I_{n_2} \otimes D_1, & D_{xx} &:= I_{n_2} \otimes -D_1^t D_1 \\ D_y &:= D_2 \otimes I_{n_1}, & D_{yy} &:= -D^t D \otimes I_{n_1} \end{aligned}$$

and hence under the aspect that the negative adjoint of this discrete gradient corresponds to the discrete divergence and neglecting the mixed derivatives D_{xy} and D_{yx} we obtain

$$-\operatorname{div} = \begin{pmatrix} D_x^t & 0 \\ 0 & D_y^t \end{pmatrix}, \quad \operatorname{div}^2 = (D_{xx}^t \ D_{yy}^t). \quad (4.2)$$

Since the matrix C is quite easy to invert one may compute its Schur complement matrix $S := B - D^t C^{-1} D$ and use this to solve the linear system $Sx_2 = b_2 - D^t C^{-1} b_1$. We then obtain x_1 from $x_1 = C^{-1}(b_1 - Dx_2)$.

4.2 Split Inexact Uzawa Method for GTV

An alternative to Newton-type methods are first-order approaches based on the introduction of appropriate novel variables and splitting. In the following we will see that an analogous primal view on GTV allows for easy decoupling of complicated terms in the variational setting. To motivate the "primal version" of GTV denoising we recall that

$$\text{TV}(u) = \int_{\Omega} |\nabla u| dx,$$

for functions $u \in W^{1,1}(\Omega)$.

In this sense the following (primal) GTV regularization generalizes TV by an additional higher order term in a decomposition approach and an additional minimization problem.

$$\begin{aligned} \text{GTV}_{\beta}(u) &= \inf_v \int_{\Omega} |\nabla u - v| dx + \beta \int_{\Omega} |\varepsilon(v)| dx \\ &= \inf_{\substack{v,w \\ \nabla u = v+w}} \int_{\Omega} |w| dx + \beta \int_{\Omega} |\varepsilon(v)| dx, \end{aligned} \quad (4.3)$$

where $\varepsilon(v) := \frac{1}{2}(\nabla v + \nabla v^t)$ denotes the symmetric derivative.

Hence in a similar fashion to the infimal-convolution, compare Section 2.2, we can observe a decomposition of the gradient of u in the regularization via GTV. For further details we refer for instance to [SST11, BKP10, KBPS11]. In Section 6 we will concentrate on this decomposition from a numerical point of view. With the primal interpretation (4.3) the variational GTV denoising (2.17) defined in section 2 reads as follows

$$u_G = \arg \min_{u \in \text{BV}(\Omega)} \left\{ \frac{1}{2} \|u - f\|_{L^2(\Omega)}^2 + \alpha \left\{ \inf_{\nabla u = v+w} \int_{\Omega} |w| dx + \beta \int_{\Omega} |\varepsilon(v)| dx \right\} \right\}. \quad (4.4)$$

A key idea to obtain efficient splitting methods is to decouple complicated nested variational terms. Usually this is realized by introducing additional constraints (often linear), which reveal simple sub-steps in an Augmented Lagrangian setting. By introducing an additional constraint substituting $\varepsilon(v)$ in (4.4) we obtain a decoupled minimization problem with simple linear constraints

$$\min_{(u,v), (w,y)} \frac{1}{2} \|u - f\|_{L^2(\Omega)}^2 + \alpha \left(\|w\|_{L^1(\Omega)} + \beta \|y\|_{L^1(\Omega)} \right) \quad (4.5)$$

$$\text{s.t. } w = \nabla u - v \quad (4.6)$$

$$y = \varepsilon(v). \quad (4.7)$$

Introducing Lagrange multiplier functions $p = (p_1, p_2)$ for the constraints yields a saddle point problem (related to the KKT optimality system) of the following form

$$\max_p \min_{\bar{u}, z} (L(\bar{u}, z; p) = H(\bar{u}) + J(z) + \langle p, A\bar{u} - z \rangle)$$

with the Lagrangian L , the data fidelity and regularization part in (4.5) denoted by $H(\bar{u})$, and $J(z)$, as well as $z = A\bar{u}$ to represent the constraints in (4.6), (4.7) in a compact form. The primal functions $\bar{u} := (u, v)$ and $z := (w, y)$ are used to simplify notations.

Several splitting methods, e.g. the Split Bregman algorithm (ADMM) by Goldstein and Osher [GO09], are based on alternating minimization of the Augmented Lagrangian with respect to the different primal and dual unknowns. The Split Inexact Uzawa methods, which are related to BOS, cf. [ZBO11], extend this idea by taking into account certain primal or dual preconditioning techniques to simplify and/or accelerate the splitting algorithm.

The Split Inexact Uzawa method (with primal preconditioning $P_{\delta,\tau} := \frac{1}{\delta} - \tau A^*A$) reads as follows:

$$\bar{u}^{k+1} = \arg \min_{\bar{u}} L(\bar{u}, z^k; p^k) + \frac{1}{2\delta} \|A\bar{u} - z^k\|_2^2 + \frac{1}{2} \|\bar{u} - \bar{u}^k\|_{P_{\delta,\tau}}^2 \quad (4.8)$$

$$= \arg \min_{\bar{u}} H(\bar{u}) + \langle A^* p^k, \bar{u} \rangle + \frac{1}{2\delta} \|\bar{u} - \bar{u}^k + \tau \delta A^*(A\bar{u}^k - z^k)\|_2^2 \quad (4.9)$$

$$z^{k+1} = \arg \min_z L(\bar{u}^{k+1}, z; p^k) + \frac{\tau}{2} \|A\bar{u}^{k+1} - z\|_2^2 = \arg \min_z J(z) + \frac{\tau}{2} \left\| \frac{1}{\tau} p^k + A\bar{u}^{k+1} - z \right\|_2^2 \quad (4.10)$$

$$p^{k+1} = p^k + \tau(A\bar{u}^{k+1} - z^{k+1}). \quad (4.11)$$

By replacing $\tau(A\bar{u}^k - z^k)$ with $p^k - p^{k-1}$ in the \bar{u} update (following the dual update) and combining p^{k+1} with w^{k+1} , the Split Inexact Uzawa approach can be rewritten as

$$\bar{u}^{k+1} = \arg \min_{\bar{u}} H(\bar{u}) + \langle A^* p^k, \bar{u} \rangle + \frac{1}{2\delta} \|\bar{u} - \bar{u}^k + \delta A^*(p^k - p^{k-1})\|_2^2 \quad (4.12)$$

$$p^{k+1} = p^k + \tau A\bar{u}^{k+1} - \tau \arg \min_z J(z) + \frac{\tau}{2} \|z - \tau^{-1}(p^k + \tau A\bar{u}^{k+1})\|_2^2. \quad (4.13)$$

To relate the Split Inexact Uzawa approach to another class of primal-dual splitting methods we found it useful to apply Moreau's decomposition for $g := p^k + \tau A\bar{u}^{k+1}$, i.e.

$$g = \arg \min_p J^*(p) + \frac{1}{2\tau} \|p - g\|_2^2 + \tau \arg \min_z J(z) + \frac{\tau}{2} \|z - \tau^{-1}g\|_2^2, \quad (4.14)$$

to the p update in (4.13). Moreover, adding terms independent of \bar{u} and p in (4.12) and (4.13) respectively implies

$$\bar{u}^{k+1} = \arg \min_{\bar{u}} H(\bar{u}) + \langle A^*(2p^k - p^{k-1}), \bar{u} \rangle + \frac{1}{2\delta} \|\bar{u} - \bar{u}^k\|_2^2 \quad (4.15)$$

$$p^{k+1} = \arg \min_p J^*(p) - \langle p, A\bar{u}^{k+1} \rangle + \frac{1}{2\tau} \|p - p^k\|_2^2. \quad (4.16)$$

Please note that, in the case of denoising with GTV, (4.8)-(4.16) can be realized very easily, since H in (4.9) is a simple L^2 data fidelity, and because the minimization in (4.10) can be solved explicitly via soft-shrinkage (thresholding).

If we replace the relaxation $2p^k - p^{k-1}$ in the u update in (4.15) simply by p^k we obtain the Primal Dual Hybrid Gradient (PDHG) splitting method, cf. [CP11, SST11, EZC10]. In other words the Split Inexact Uzawa method for our special choice of $P_{\delta,\tau}$ coincides with a modified version of PDHG.

By introducing an additional separate relaxation step, compare with the recent work [CP11],

$$\bar{p}^{k+1} = \bar{p}^k + \theta(p^{k+1} - p^k)$$

we obtain Split Inexact Uzawa above for $\theta = 1$ and PDHG for GTV, [SST11], for $\theta = 0$. In Algorithm 1 we summarized all the ideas. As possible extensions one can think of different preconditioning methods by adapting $P_{\delta,\tau}$ and of accelerating the iteration process by variable step sizes θ , τ and γ in analogy to [CP11].

Algorithm 1 Split Inexact Uzawa (SIU) for GTV Denoising (4.4),(2.17)**Parameters:** noisy data f , reg. param. $\alpha \geq 0$, weight. param. $\beta \geq 0$ **Initialization:** $\bar{u}^0 = f$, $v^0 = \frac{1}{2}\varepsilon(\nabla f)$, $w^0 = \frac{1}{2}\nabla f$, $p_1^0 = p_2^0 = \bar{p}_1^0 = \bar{p}_2^0 = 0$
 $\theta = 1$, $\tau = 0.017$, $\gamma = 9$ For $k = 0, \dots$ repeat until a primal/dual stopping criterion is fulfilled:

$$\begin{array}{l}
u^{k+1} = (1 + \delta)^{-1} \left(\delta f + u^k + \delta \tau \operatorname{div}(\bar{p}_1^k) \right) \\
v^{k+1} = v^k + \delta \tau \left(\bar{p}_1^k - \varepsilon^*(\bar{p}_2^k) \right) \\
w^{k+1} = S_{\frac{\alpha_1}{\tau}} \left(\bar{p}_1^k + \nabla u^{k+1} - v^{k+1} \right) \\
y^{k+1} = S_{\frac{\alpha_2}{\tau}} \left(\bar{p}_2^k + \varepsilon(v^{k+1}) \right) \\
p_1^{k+1} = p_1^k + \nabla u^{k+1} - v^{k+1} - w^{k+1} \\
p_2^{k+1} = p_2^k + \varepsilon(v^{k+1}) - y^{k+1} \\
\bar{p}_1^{k+1} = \bar{p}_1^k + \theta(p_1^{k+1} - p_1^k) \\
\bar{p}_2^{k+1} = \bar{p}_2^k + \theta(p_2^{k+1} - p_2^k)
\end{array}
\left. \begin{array}{l} \\ \\ \\ \\ \\ \\ \\ \\ \end{array} \right\} \begin{array}{l} \text{primal updates, i.e. updates for} \\ \bar{u} = (u, v) \text{ and } z = (w, y) \\ \\ \\ \text{dual updates, i.e. update for } p = (p_1, p_2) \\ \\ \text{relaxation, compare e.g. [CP11]} \end{array}$$

End

return $\bar{u}^{k+1}, w^{k+1}, v^{k+1}$.**5 Numerical Results for Higher-Order TV Methods**

In this section we present one- and two-dimensional results in particular, to verify the theoretical examples of Section 3. We restrict ourselves to the one-dimensional case here to highlight the main improvements with respect to pure TV regularization and the main difficulties appearing in the variational methods. Further two-dimensional examples will be given in the next section, when we compare with the results from Bregman iterations.

In the case of one-dimensional data we want to focus on the functions u_{block} and u_{blockhat} as defined in Section 3, and compare computational solutions of (1.2) and (1.1) with $J = \text{GTV}_\beta$, for varying β . Note that due to Theorem 1 it is sufficient to choose only (1.1) with either $J = \text{GTV}_\beta$ or $J = \text{ICTV}_\beta$.

Starting with u_{block} as an example, we define f as u_{block} corrupted by Gaussian distributed noise, with mean zero. We compute ROF reconstructions for $\alpha = 1/8$, $\alpha = 1/4$ and $\alpha = 1/2$, and reconstructions of (1.1) with $J = \text{GTV}_\beta$, with the same values for α , and $\beta = 4$ as well as $\beta = 2$. The results can be seen in Figure 2. Note that the interval length has been chosen as $L = 8$ in order to guarantee that the loss of contrast c is approximately $c = 1 - \alpha$ for $\beta \geq 4$ (approximately, since the analytical computations of Section 3 have been made without noise assumption). It is clearly visible that for $\beta = 4$ the TV- and GTV_β -reconstruction coincide and recover cu_{block} as predicted. In case of $\beta = 2$ the bound on the dual variable however is too restrictive and thus, no multiple of u_{block} , but an piecewise-constant-piecewise-affine-linear approximation, is recovered.

In the following we want to concentrate on the function u_{blockhat} , and proceed in the same fashion as for u_{block} . We define f as a noisy version of u_{block} and compare the ROF and (1.1) reconstructions for $\alpha = 1/30$, $\alpha = 1/8$, $\alpha = 1/4$ and $\alpha = 1/2$. The parameter β is set to $\beta = L/\sqrt{3}$ and $\beta = L/\sqrt{3} - 1$ respectively, for $L = 32/(2 + \sqrt{3})$. The results can be seen in Figure 3. It is easy to see that the ROF model fails in recovering some multiple of u_{blockhat} , but instead introduced the well-known staircasing artifacts. In case of $\beta = L/\sqrt{3}$, the GTV reconstruction does a perfect job in exact recovery, recovering u_{blockhat} with the predicted loss of contrast. For $\beta = L/\sqrt{3} - 1$ the model fails in recovering multiples but rather tips the linear parts or even eliminates the edges for large α .

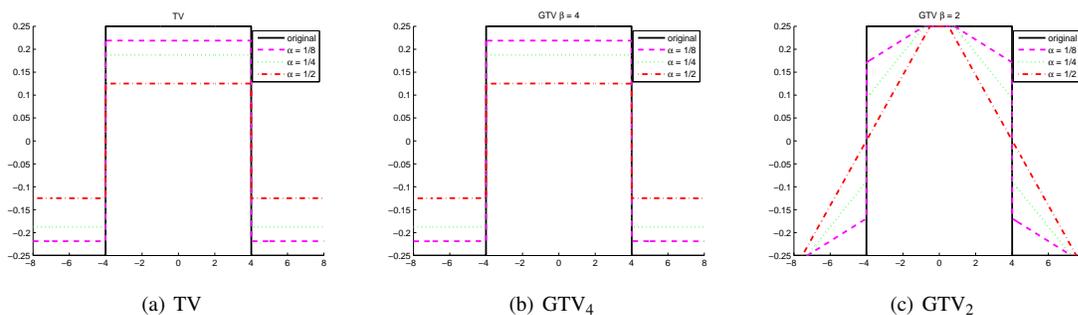


Fig. 2 Computational solutions of (1.2) and (1.1) with $J = \text{GTV}_\beta$ on the interval $[-8, 8]$, for $f = u_{\text{block}}$ corrupted by Gaussian distributed noise. It is easy to see that for $\beta = 4$ the GTV_4 -regularized reconstructions coincide with the solutions of the ROF model. For $\beta = 2$ however it is clearly visible that GTV_2 favors piecewise-constant-piecewise-affine-linear approximations.

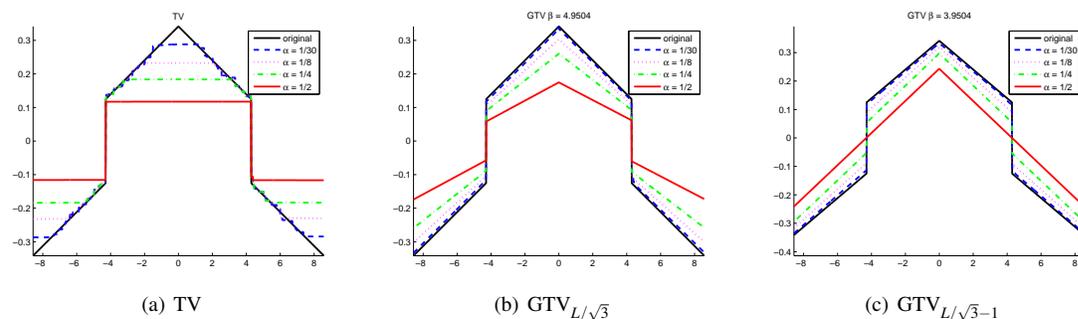


Fig. 3 Computational solutions of (1.2) and (1.1) with $J = \text{GTV}_\beta$ on the interval $[-32/(2 + \sqrt{3}), 32/(2 + \sqrt{3})]$, for $f = u_{\text{blockhat}}$ corrupted by Gaussian distributed noise. The ROF reconstruction produces staircasing artifacts and fails in recovering the linear parts of u_{blockhat} . For $\beta = L/\sqrt{3}$ the $\text{GTV}_{L/\sqrt{3}}$ -regularized reconstructions coincide with the analytical solutions predicted in Section 3. In case of $\beta = L/\sqrt{3} - 1$ however it is clearly visible that $\text{GTV}_{L/\sqrt{3}-1}$ does not recover multiples of u_{blockhat} anymore.

As a final 1D example we want to corrupt u_{blockhat} by the noise defined in Section 3.4, i.e. $f = u_{\text{blockhat}} + n$ with n being defined as in (3.6). The noise parameters are set to $A = 0.05$ and $k = 35$, and again we define the interval length to be $L = 32/(2 + \sqrt{3})$. Similar to the previous examples we compute reconstructions for $\alpha = 1/30$, $\alpha = 1/8$, $\alpha = 1/4$ and $\alpha = 1/2$, and set the parameter β to $\beta = L/\sqrt{3}$. The results can be seen in Figure 4. We observe that in contrast to the total variation reconstructions the generalized total variation reconstructions do not produce staircasing artifacts, though - if the regularization parameter is too small - a systematic bias occurs.

6 Higher Order Inverse Scale Space Methods

In this section we want to discuss the use of inverse scale space methods in the context of the generalized total variation (2.7) as a regularizer. In case of total variation and ℓ^1 regularization, inverse scale space methods have been extensively studied and discovered to correct the loss of contrast ([OBG⁺05, BGOX06]). The inverse scale

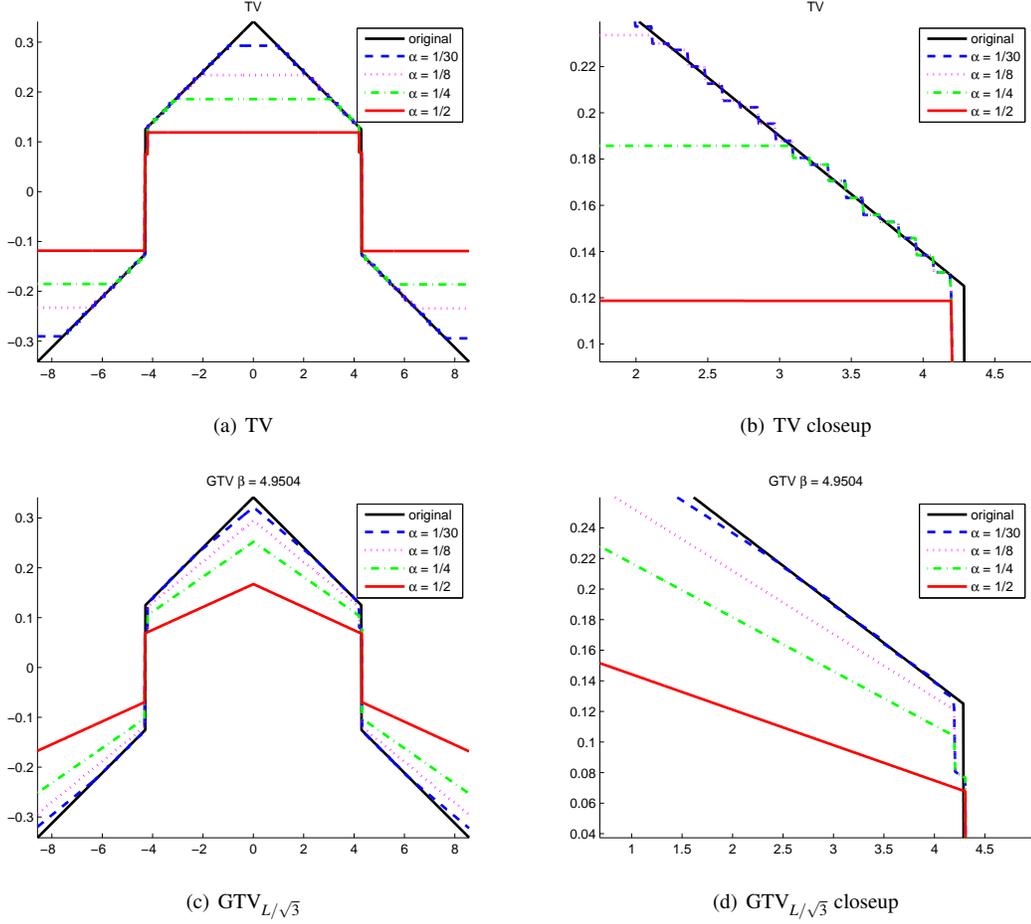


Fig. 4 Computational solutions of (1.2) and (1.1) with $J = GTV_\beta$ on the interval $[-32/(2 + \sqrt{3}), 32/(2 + \sqrt{3})]$, for $f = u_{\text{blockhat}}$ corrupted by (3.6) with $A = 0.05$ and $k = 35$. The ROF reconstruction produces staircasing artifacts and fails in recovering the linear parts of u_{blockhat} , which is clearly visible in the closeup of the reconstruction. In contrast, for $\beta = L/\sqrt{3}$ the $GTV_{L/\sqrt{3}}$ -regularized reconstructions again coincide with the analytical solutions predicted in Section 3 if α is large enough, which supports the considerations of Section 3.4. If α is small enough we face some bias of the reconstruction in contrast to the analytical solutions, but still no staircasing artifacts are produced as in the total variation case.

space for a general, convex regularization functional J is defined as

$$\frac{\partial}{\partial t} p(t) = f - u(t), \quad (6.1)$$

for $p(t) \in \partial J(u(t))$, assuming $u(t=0) = p(t=0) = 0$. The discrete analogue to (6.1) is the so-called Bregman iteration, which can be seen as a simple backward Euler discretization of (6.1), i.e.

$$\frac{p_k - p_{k-1}}{h} = f - u_k,$$

with the stepsize h defined as $h := 1/\alpha$. Then, this discretized inverse scale space flow is equivalent to solving the iterative scheme

$$u_k = \arg \min_{u \in \text{dom}(J)} \left\{ \frac{1}{2} \|u - f\|_{L^2(\Omega)}^2 + \alpha D_f^{p_{k-1}}(u, u_{k-1}) \right\}, \quad (6.2)$$

for $u_0 = p_0 = 0$, and with $D_f^p(u, v)$ denoting the Bregman distance. Equation (6.2) can be rewritten to (cf. [OBG⁺05])

$$\begin{aligned} u_k &= \arg \min_{u \in \text{dom}(J)} \left\{ \frac{1}{2} \|u - (f + v_{k-1})\|_{L^2(\Omega)}^2 + \alpha J(u) \right\}, \\ v_k &= v_{k-1} - (u_k - f) \end{aligned} \quad (6.3)$$

for $u_0 = v_0 = 0$, which will be useful below.

In order to motivate the use of inverse scale space methods, we want to recall a result from [Ben11, Chapter 7, Section 2]. Assume a function u_λ to satisfy an Eigenfunction property with respect to a regularization functional J , i.e. $u_\lambda \neq 0$ satisfies

$$\lambda u_\lambda \in \partial J(u_\lambda) \quad (6.4)$$

for an Eigenvalue $\lambda \in \mathbb{R}_{>0}$. Then, for data given in terms of an Eigenfunction, we can state the following proposition.

Theorem 4 *Let $J : \text{dom}(J) \subseteq L^2(\Omega) \rightarrow \mathbb{R} \cup \{+\infty\}$ be a convex and one-homogeneous functional, and let $u_\lambda \neq 0$ be a function satisfying (6.4) with Eigenvalue λ . Then there exist times $t_* < t_{**} < \infty$ such that for given data $f = u_\lambda + n$ the solution of (6.1) at time t with $t_* \leq t < t_{**}$ is given via $u(t) = cu_\lambda$, for*

$$c = 1 + \frac{\lambda - \mu}{\eta},$$

if there exist constants μ and $\eta > 0$ such that

$$\mu u_\lambda + \eta n \in \partial J(u_\lambda)$$

is guaranteed.

Note that in case of noise-free data (i.e. $n = 0$) the constant c simply equals one, since η can be chosen arbitrarily and thus, μ can be set to equal λ . Moreover, for $n = 0$ the solution $u(t)$ remains u_λ for every $t \geq t_* = \lambda$ (i.e. $t_{**} = \infty$). Hence, Eigenfunctions u_λ can be recovered perfectly after finite time by the inverse scale space method (6.1). If we recall the examples of Section 3 we see that these examples were designed to be Eigenfunctions. Thus, with an inverse scale space strategy, with respect to the specific regularization functional, we would be able to recover them perfectly.

Due to the multivaluedness of subdifferentials in case of L^1 -type regularization functionals (as in case of higher-order total variation methods) we can even expect to find $\mu \approx \lambda$ in case of noisy data, yielding an almost perfect reconstruction for times $t_* \leq t < t_{**}$.

Hence, one reason for considering inverse scale space methods in the context of generalized total variation is the ability to improve the reconstruction of piecewise constant and piecewise linear functions, and additive combinations of it, by correcting the loss of contrast.

Another reason is that in many practical cases generalized total variation tends to tip homogeneous regions, which we can see by the following example. Let us consider the function $u_{\text{edge}} : [-L, L] \rightarrow \{-1/\sqrt{2L}, 1/\sqrt{2L}\}$ with

$$u_{\text{edge}}(x) = \begin{cases} \frac{1}{\sqrt{2L}} & x \in [0, L] \\ -\frac{1}{\sqrt{2L}} & x \in [-L, 0[\end{cases}.$$

We would expect generalized total variation to be able to recover u_{edge} only with a loss of contrast, since u_{edge} , in analogy to the function u_{block} defined in (3.1), is an Eigenfunction of the total variation of order one. However, the difference between both functions is that u_{edge} is not orthogonal to the affin-linear functions, which are the trivial ground states of GTV_β due to Lemma 2, since $\langle u_{\text{edge}}, x \rangle = \sqrt{L^3/2} \neq 0$ holds. And indeed, the solution of (1.1) with GTV_β as a regularizer can be characterized via

$$\hat{u}(x) = \begin{cases} \frac{6\alpha}{L^2} (x - \frac{2L}{3}) + \frac{1}{\sqrt{2L}} & x \in [0, L] \\ \frac{6\alpha}{L^2} (x + \frac{2L}{3}) - \frac{1}{\sqrt{2L}} & x \in [-L, 0[\end{cases},$$

for $\alpha < L/4$ and $\beta \geq (4L)/27$. Thus, the edges of u_{edge} get tipped. Obviously, we could argue that we do not need generalized total variation in order to recover u_{edge} . However, if we consider piecewise-constant-piecewise-linear functions, we would expect to discover similar effects, which are not desirable.

For that reason we want to consider what happens by applying a Bregman iteration strategy for the reconstruction. Let us assume that the solution of (6.3) at iteration 1 is given by $u_1 = \hat{u}$, which can be guaranteed by setting $\alpha < L/4$ and β fixed to a value larger or equal $(4L)/27$. The update for v then reads as

$$\begin{aligned} v_1 &= v_0 - (u_1 - f) \\ &= -\frac{6\alpha}{L^2} \begin{cases} x - \frac{2}{3}L & x \in [0, L] \\ x + \frac{2}{3}L & x \in [-L, 0[\end{cases} \end{aligned}$$

and thus, the data for the next update for u modifies to

$$(f + v_1)(x) = \begin{cases} \frac{1}{\sqrt{2L}} - \frac{6\alpha}{L^2} (x - \frac{2}{3}L) & x \in [0, L] \\ -\frac{1}{\sqrt{2L}} - \frac{6\alpha}{L^2} (x + \frac{2}{3}L) & x \in [-L, 0[\end{cases}.$$

As a consequence, $u_2 = u_{\text{edge}}$ is the solution for the next Bregman iteration, since u_{edge} satisfies the optimality condition

$$q''(x) = \frac{1}{\alpha} (f + v_1 - u_{\text{edge}})(x) = -\frac{6}{L^2} \begin{cases} x - \frac{2}{3}L & x \in [0, L] \\ x + \frac{2}{3}L & x \in [-L, 0[\end{cases}$$

(again, the derivation has to be considered in a weak sense), for the function q defined as

$$q(x) := -\begin{cases} \frac{1}{L^2}x^3 - \frac{2}{L}x^2 + x & x \in [0, L] \\ \frac{1}{L^2}x^3 + \frac{2}{L}x^2 + x & x \in [-L, 0[\end{cases}.$$

It is easy to see that q satisfies $\|q\|_\infty = (4L)/27 = \beta$, $\|q'\|_\infty = 1$, $q(-L) = q(L) = q'(-L) = q'(L) = 0$ and $\langle q'', u_{\text{edge}} \rangle = \sqrt{2/L} = \text{GTV}_{\frac{4L}{27}}(u_{\text{edge}}) = \text{TV}(u_{\text{edge}}) + (4L \text{TV}^2(0))/27 = \text{TV}(u_{\text{edge}})$, and thus, is a subgradient of the subdifferential $\partial \text{GTV}_{\frac{4L}{27}}(u_{\text{edge}})$.

Hence, the systematic tipping of homogeneous regions in case of u_{edge} as defined above can entirely be compensated by Bregman iteration. Note that this is true due to the fact that u_{edge} can simply be composed of two Eigenfunctions.

In the following we also want to consider a more complex two dimensional example for which the proposed Bregman-iterated GTV_β method will lead to results that are even superior to a Bregman-iterated ICTV_β method, in the absence and presence of noise. Note that due to their derivation, the corresponding vector fields of the two-dimensional Eigenfunction examples of Section 3, namely u_{block}^{2D} , u_{hat}^{2D} and u_{blockhat}^{2D} , are irrotational, i.e. $\text{curl}(p) = 0$ for p denoting a corresponding vector field of these examples. Consequently, due to the considerations of Section 2.4, we assume that both models, (1.1) with $J = \text{GTV}_\beta$ as well as with $J = \text{ICTV}_\beta$, are able to recover the functions u_{block}^{2D} , u_{hat}^{2D} and u_{blockhat}^{2D} almost perfectly, suffering only from a loss of contrast. Section 2.4 also suggests that the limitations of model (1.1) with $J = \text{ICTV}_\beta$ arise in the need for irrotational vector fields. Hence, we want to discover an image, for which the corresponding vector field is not irrotational, in order to have an analytical example which cannot be recovered well with ICTV_β -regularization.

For that reason we want to consider a vector field (p_x, p_y) based on the corresponding dual variable of u_{block} , which we modify to

$$p_x(x, y) := \frac{2}{L} \begin{cases} xy & x \in [-\frac{L}{2}, \frac{L}{2}] \\ y(L-x) & x \in]\frac{L}{2}, L] \\ -y(L+x) & x \in [-L, -\frac{L}{2}[\end{cases}$$

and

$$p_y(x, y) := \frac{2}{L} \begin{cases} xy & y \in [-\frac{L}{2}, \frac{L}{2}] \\ x(L-y) & y \in]\frac{L}{2}, L] \\ -x(L+y) & y \in [-L, -\frac{L}{2}[\end{cases}.$$

Consequently, p_x is no longer constant but linear with respect y , as is p_y with respect to x . Moreover, for $\text{curl}(p)$ we obtain

$$\text{curl}(p) = \frac{\partial p_y}{\partial x} - \frac{\partial p_x}{\partial y} \neq 0.$$

Computing $u = c \text{div } p$, with c guaranteeing $\|u\|_{L^2(\Omega)} = 1$, thus yields

$$u(x, y) = \frac{\sqrt{3}}{L^2\sqrt{8}} \begin{cases} x+y & (x, y) \in [-\frac{L}{2}, \frac{L}{2}]^2 \\ x-y & (x \notin [-\frac{L}{2}, \frac{L}{2}]) \wedge (y \in [-\frac{L}{2}, \frac{L}{2}]) \\ y-x & (x \in [-\frac{L}{2}, \frac{L}{2}]) \wedge (y \notin [-\frac{L}{2}, \frac{L}{2}]) \\ -(x+y) & \text{else} \end{cases}. \quad (6.5)$$

In the following, we want to investigate two approaches for the computational realization of Bregmanized GTV_β . The first approach aims in solving (6.3), by computing the primal update via a primal-dual Newton method; the second approach is based on a method closely related to the alternating direction of multipliers methods (ADMM). We are going to test these algorithms on the previously described examples, on the examples of Section 4 and on examples from [SST11].

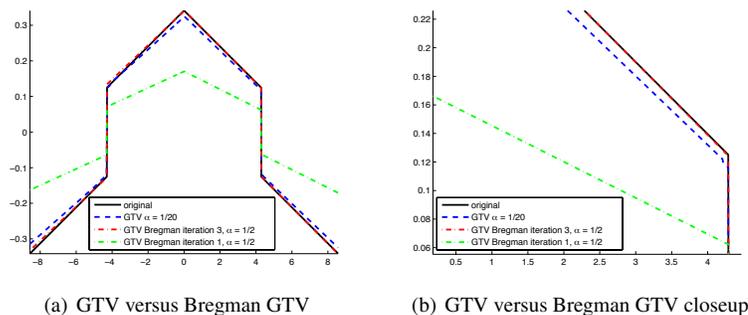


Fig. 5 Comparison of a standard GTV- and a Bregman-GTV-reconstruction in case of $f = u_{\text{blockhat}} + n$, with n denoting Gaussian distributed noise with mean zero. The function u_{blockhat} is defined on the interval $[-32/(2+\sqrt{3}), 32/(2+\sqrt{3})]$. The GTV-reconstruction was performed with the parameters $\alpha = 1/20$ and $\beta = L/\sqrt{3}$, while for the Bregman iterated solution 3 Bregman iterations have been performed, for the same value of β but with $\alpha = 1/2$.

6.1 Numerical Methods for Bregman-GTV

Bregman-GTV with Newton or Inexact Uzawa Method:

By rewriting the Bregman iteration scheme (6.2) to (6.3) we can simply use the Newton method or Split Inexact Uzawa method presented in Section 4.1 and Section 4.2 with modified input data (given in terms of $f + v_{k-1}$) in order to solve the Bregman iteration scheme. Thus, we basically have to solve a sequence of standard GTV problems, which have to be solved by the iterative Newton scheme or the iterative splitting method each.

6.2 Synthetic results for Bregman-GTV

6.2.1 The One-Dimensional Setup

Again, as in Section 5 we want to consider one-dimensional examples first, and therefore focus on the function u_{blockhat} of Section 3 again. In analogy to Section 5, we consider u_{blockhat} on the interval $[-L, L]$ with $L = 32/(2+\sqrt{3})$, and define f as a noisy version of u_{blockhat} . Figure 5 on the one hand shows a standard GTV reconstruction for small α ($\alpha = 1/20$) and $\beta = L/\sqrt{3}$. On the other hand we see a Bregman iterated reconstruction for comparison. The Bregman iterated version has been computed with the same value for β , but with $\alpha = 1/2$ and 3 Bregman iterations. In terms of u_{blockhat} , which is an Eigenfunction of the $\text{GTV}_{L/\sqrt{3}}$ -functional, the improvement is not stunning, but at least visible. The standard GTV reconstruction has a loss of contrast around the edges, and linear regions have slightly decreased slope in contrast to the Bregman iterated solution.

6.2.2 The Two-Dimensional Setup

First we investigate the two dimensional function u_{blockhat} defined in Section 3. As we can see in Fig. 6 and Fig. 6.2.2 the Bregman iterated GTV (Fig. 6(c)) as well as the Bregman iterated ICTV reconstruction (Fig. 6(d)) lead to the same result. The function u_{blockhat} is nearly exactly recovered despite of the noise in the image. However if we consider the function defined in 6.5 there are obvious differences in the Bregman reconstructions. While the Bregman iterated GTV reconstruction (Fig. 8(c)) again recovers the function nearly exactly, there are TV like

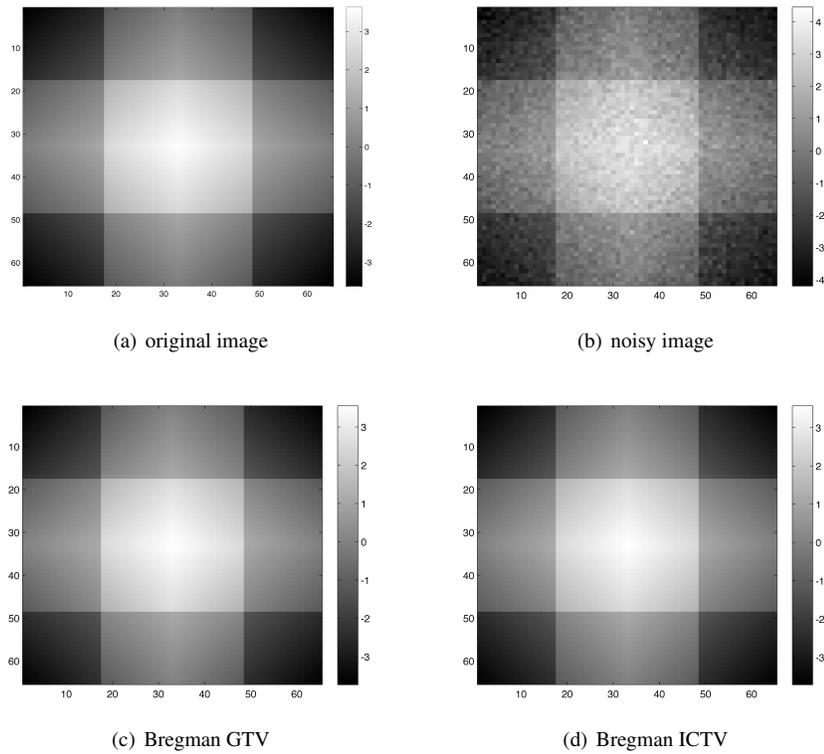


Fig. 6 a) Original image u (64x64 px), and b) given noisy image f corrupted by additive Gaussian noise. c) Bregman GTV (7 iterations) $\alpha = \frac{1}{2}$ and $\beta = \frac{L}{\sqrt{3}}$. d) Bregman ICTV (7 iterations) $\alpha = \frac{1}{2}$ and $\beta = \frac{L}{\sqrt{3}}$

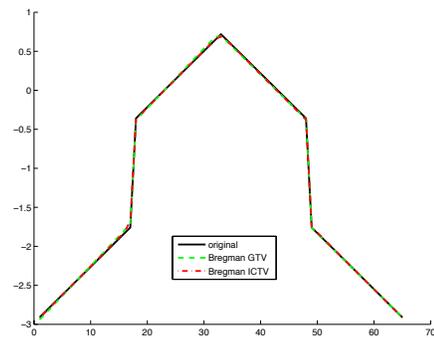


Fig. 7 Line profile through line 55 in Fig. 6

artifacts well visible in the Bregman iterated ICTV reconstruction (Fig. 8(d)). This is also apparent in the line profile in Fig. 9.

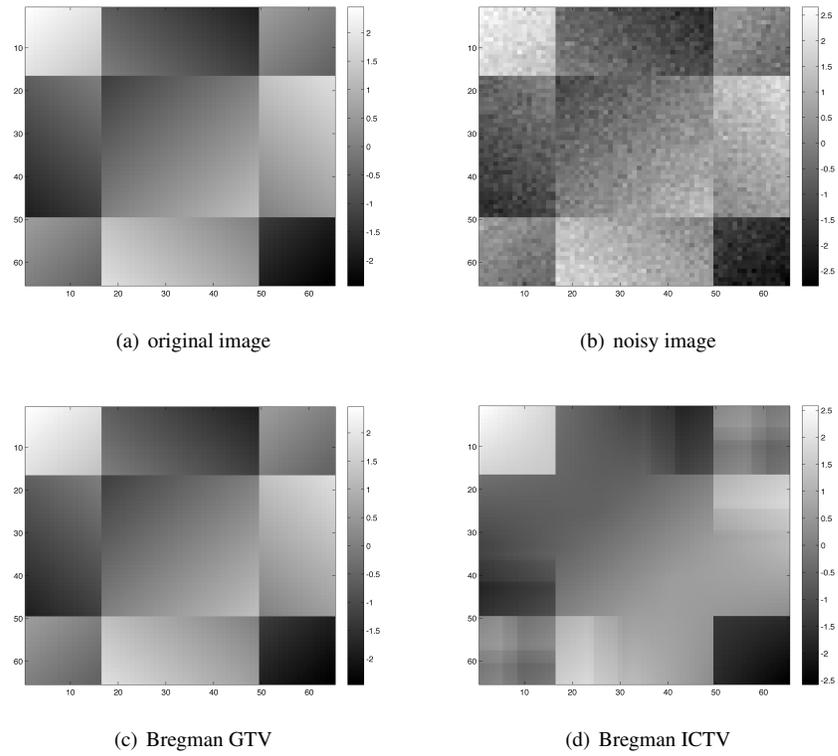


Fig. 8 a) Original image u (64x64 px), and b) given noisy image f corrupted by additive Gaussian noise. c) Bregman GTV (10 iterations) $\alpha = \frac{1}{2}$ and $\beta = \frac{L_u}{\sqrt{3}}$. d) Bregman (10 iterations) $\alpha = \frac{1}{2}$ and $\beta = \frac{L_u}{\sqrt{3}}$

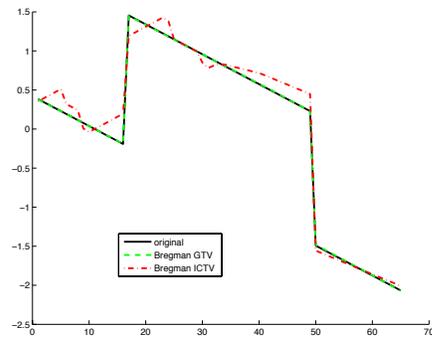


Fig. 9 Line profile through line 55 in Fig. 8

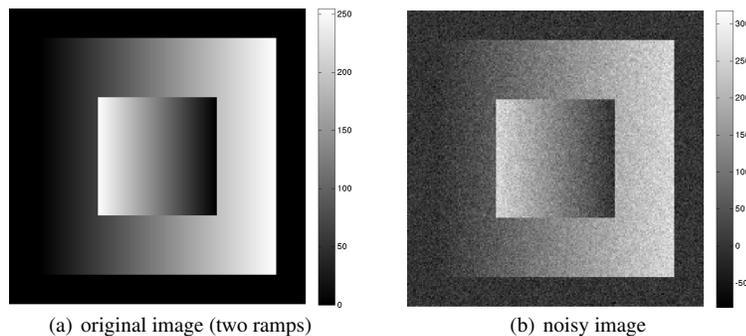


Fig. 10 (a): Original image u (200x200 px), and (b): given noisy image f corrupted by additive Gaussian noise of standard deviation 20. Compare with [SST11].

Here we also want to consider the two-dimensional synthetic **two-ramps example**, see Figure 10, discussed for ICTV and GTV in the work by Setzer et al. [SST11]. With this example we particularly aim for showing an improved image-decomposition performance of Higher-Order Inverse Scale Space methods, which results in higher reconstruction quality.

Figure 10 shows the original image u with two ramps intersecting and increasing in opposite directions. The surrounding frame has constant brightness to address the combination of piecewise affinely linear and piecewise constant parts. The resolution is 200x200 pixels and the image is normalized to $[0,255]$. In analogy to [SST11] we corrupted the original by additive Gaussian noise of standard deviation 20.

Setzer et al. [SST11] showed that GTV outperforms ICTV in this example. To get an idea of the natural remaining deficiencies of GTV we focus on applying image decomposition via Bregman-GTV to the noise-free original image first. Figure 11 shows the first step of the Bregman-GTV algorithm (with inner Split Inexact Uzawa), which is simply GTV. Compared to [SST11] we increased the regularization parameter by a factor of two, i.e. $\alpha = 120, \beta = 2.5$, to reveal remaining deficiencies of GTV even more. The first row shows the resulting decomposition $u = u_1 + u_2$ and the second and third row represent the decomposition of the gradient of u , i.e. $\nabla u = w + v$. Please note that the colorbar for the derivatives in row two and three is limited to a range of $[-10,+10]$ for better visualization.

A perfect reconstruction and decomposition of the image would show continuous edges in w_1 and w_2 , as well as nearly piecewise constant results for v_1 and v_2 . The lack of accuracy of the GTV result in Figure 11 is apparent for instance at the edges of the two ramps in u facing each other. This deficiency can also be verified by missing edge parts in w and remaining blocks in v .

In Figure 12 we present the result of Bregman-GTV after 3 iterations. Obviously, we now obtain a nearly perfect image decomposition and a strongly enhanced image reconstruction result.

In Figure 13 and Figure 14 we illustrate the same case, but now with noisy data. Interestingly, we could observe nearly the same great improvement of the image decomposition for Bregman-GTV, which again results in an improved reconstruction quality.

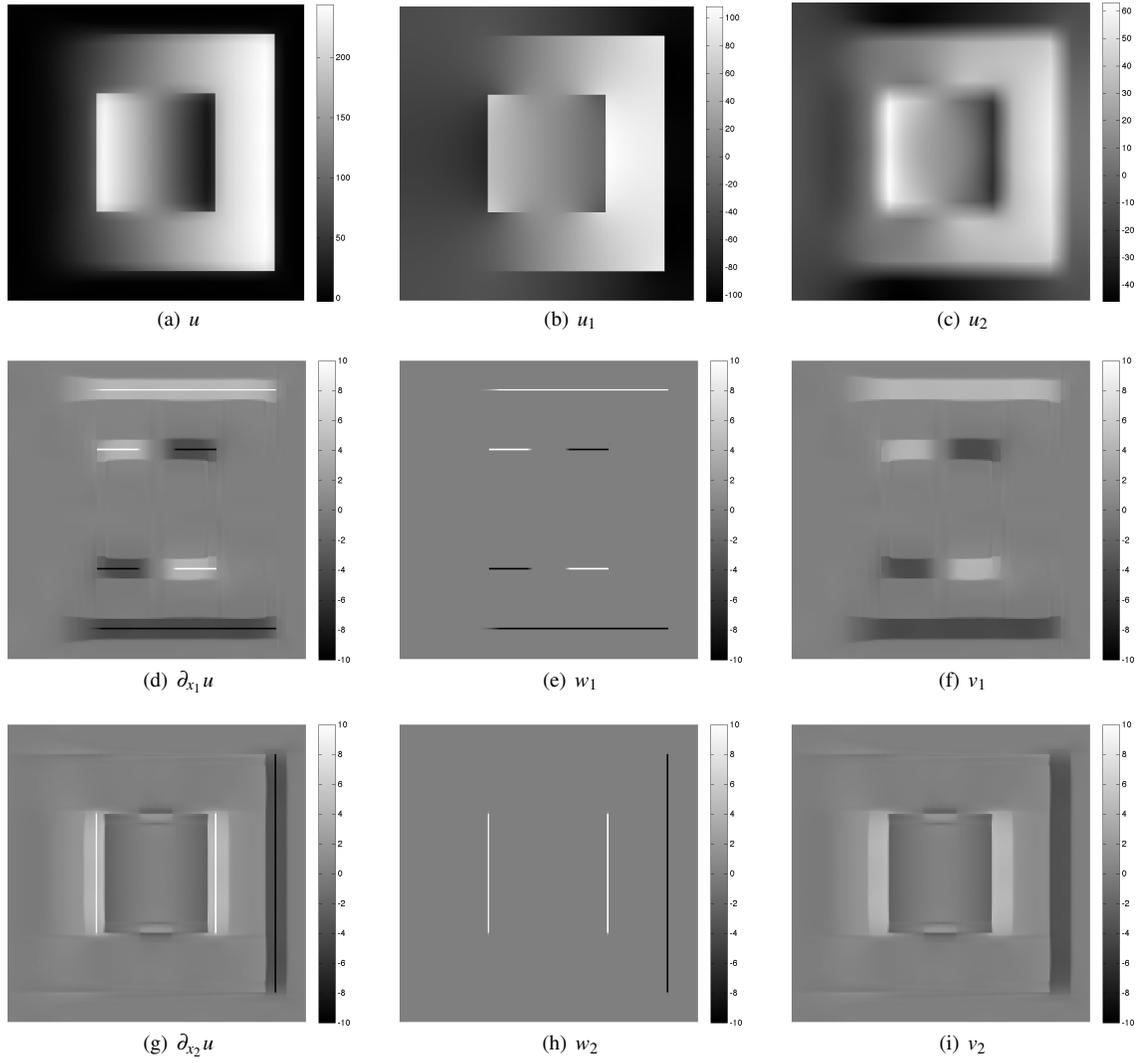


Fig. 11 Reconstructions for data without noise, 1st iteration of Bregman-GTV with $\alpha = 120, \beta = 2.5$ and 400 iterations.

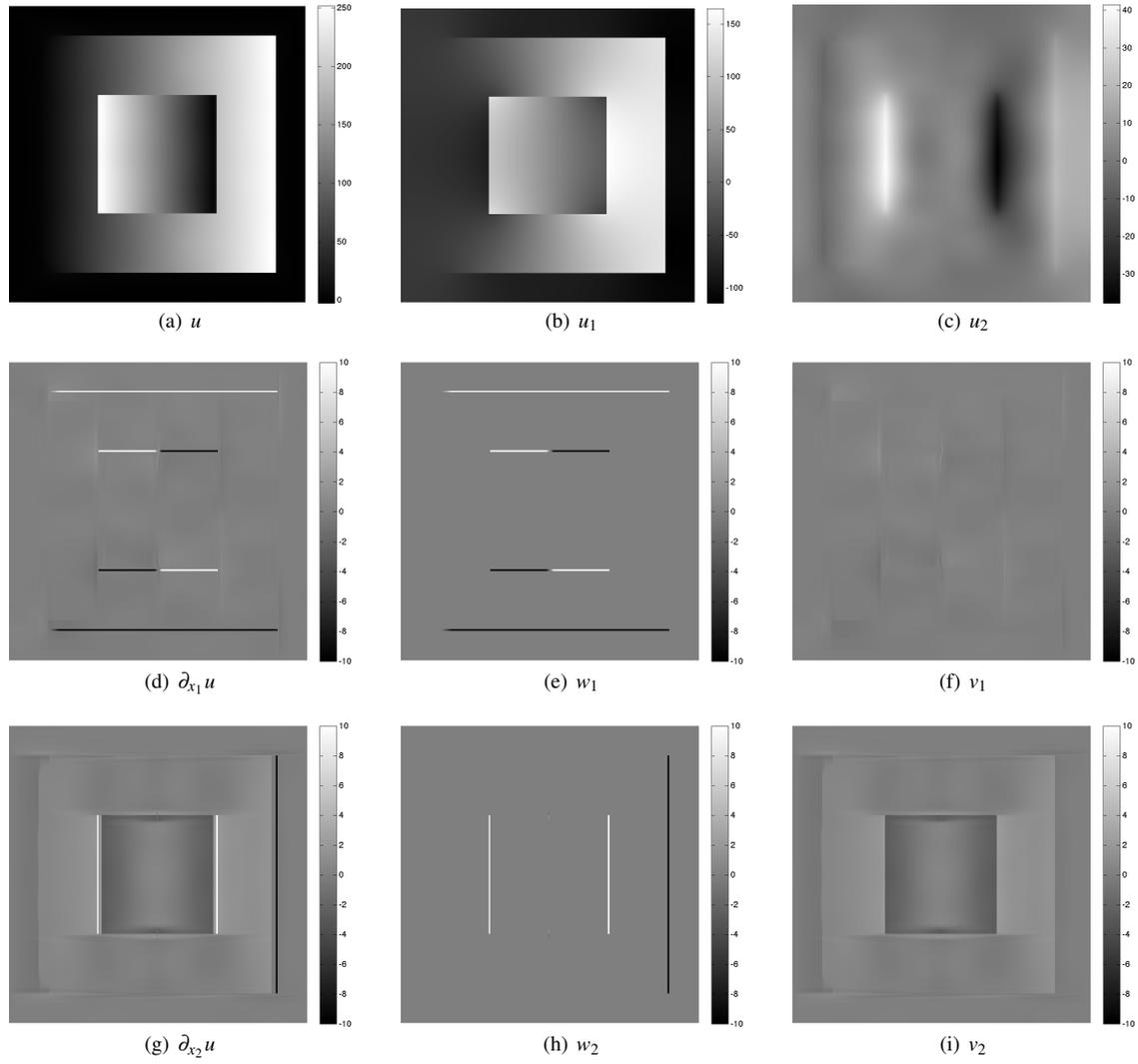


Fig. 12 Reconstructions for data without noise, 3rd iteration of Bregman-GTV with $\alpha = 120, \beta = 2.5$ and 400 iterations.

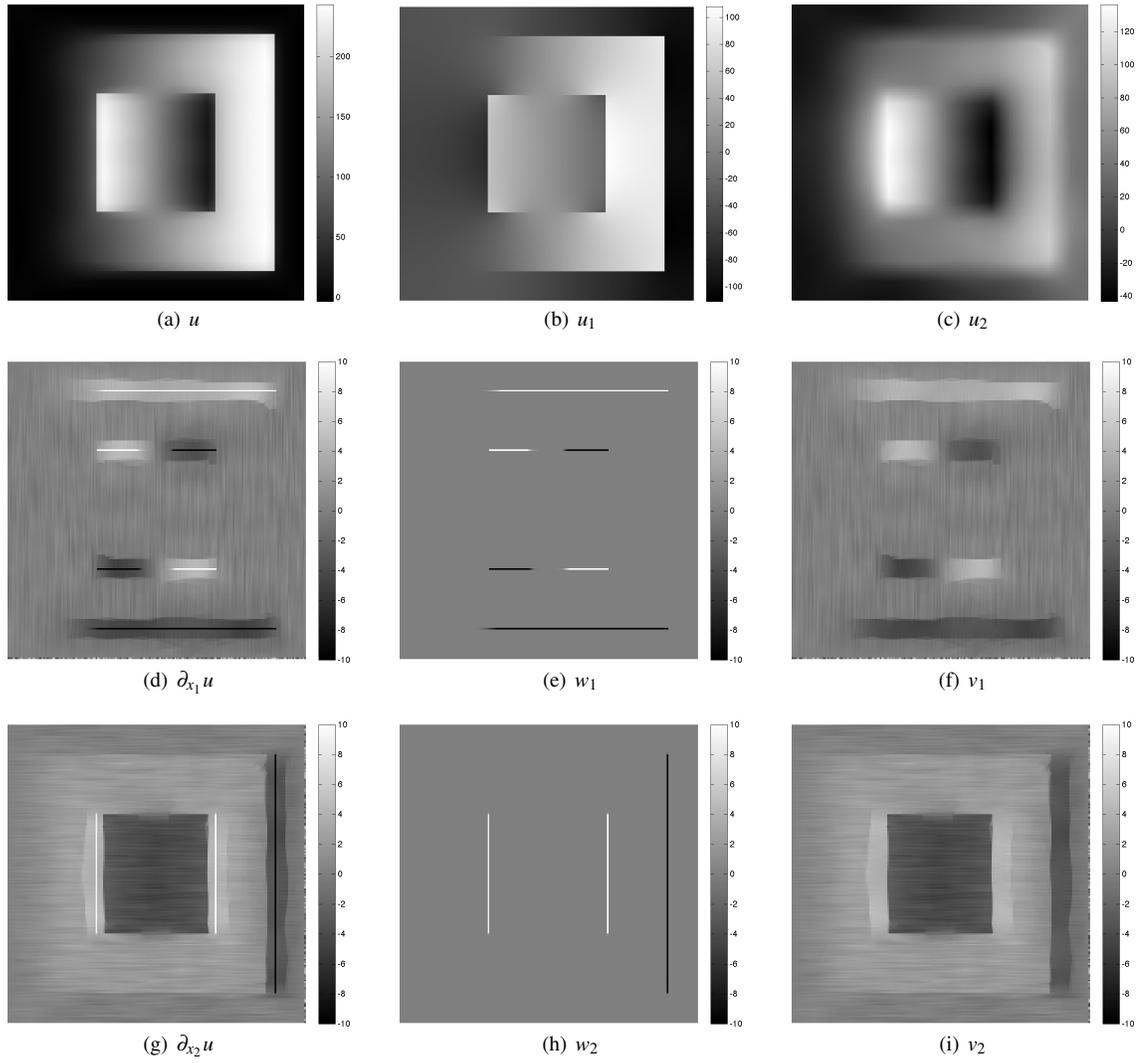


Fig. 13 Reconstructions for noisy data, 1st iteration of Bregman-GTV with $\alpha = 120$, $\beta = 2.5$ and 300 iterations.

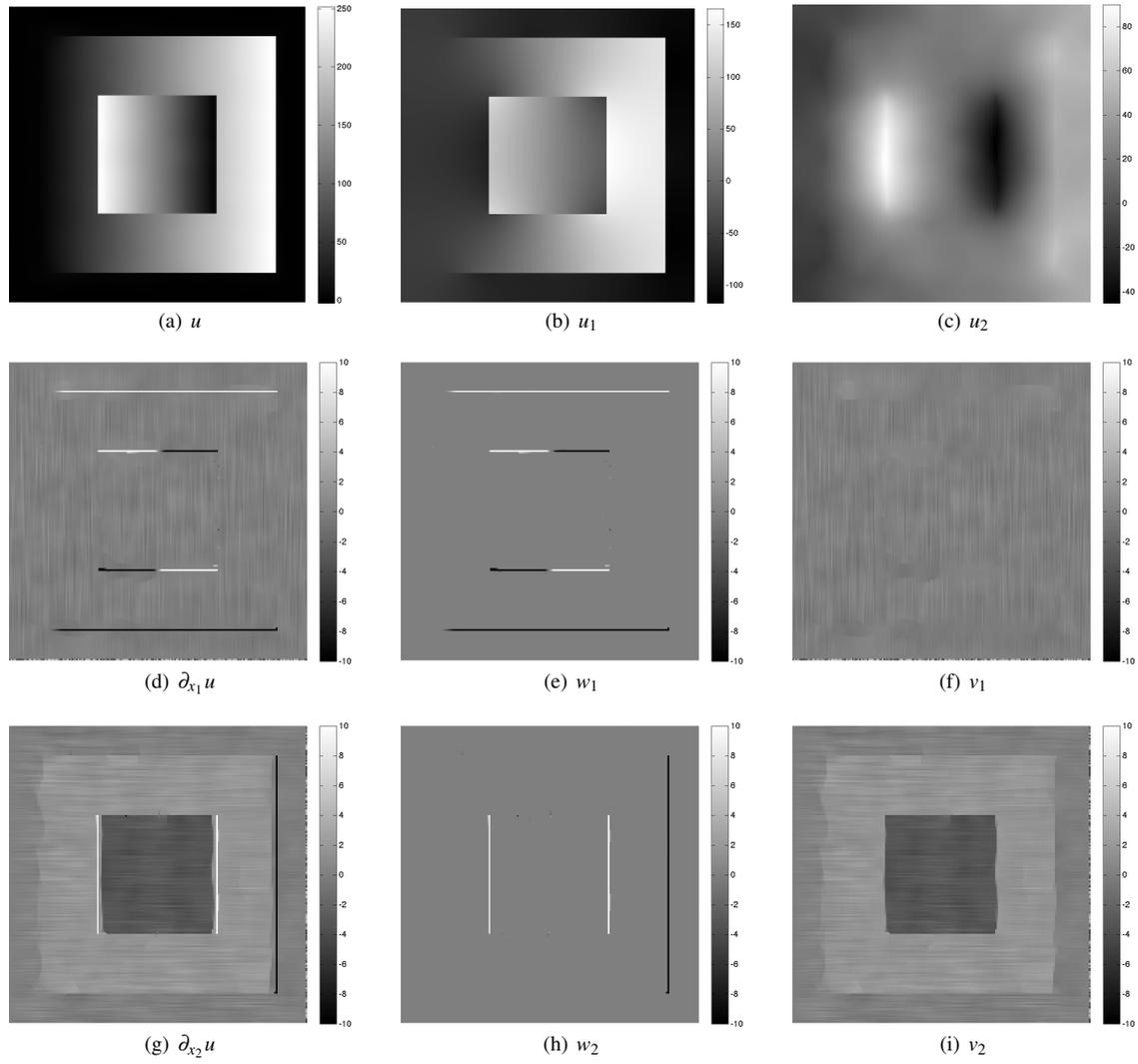


Fig. 14 Reconstructions for noisy data, 3rd iteration of Bregman-GTV with $\alpha = 120$, $\beta = 2.5$ and 300 iterations.

6.3 Experimental results

In this subsection we present some real-world high-resolution color image denoising results with Bregman-GTV. The given noisy image in Figure 15 (a) shows Santa Monica beach near LA, California. The image has been taken by the authors with a smartphone camera. The data is a color image and has a resolution of $968 \times 1296 \times 3$ pixels. On the one hand, we can observe some smooth (higher-order) features in the given noisy image like the sky, clouds, as well as some sand regions with and without water. On the other hand, the image contains some sharp edges



(a) Santa Monica beach with pier and seagulls



(b) Best **GTV denoising** via Split Inexact Uzawa, $\alpha = 2, \beta = 2.5$



(c) **Bregman-GTV, 1st iter.**, with inner SIU, $\alpha = 64, \beta = 2.5$



(d) **Bregman-GTV, 4th iter.**, with inner SIU, $\alpha = 64, \beta = 2.5$



(e) **Bregman-GTV, 7th iter.**, with inner SIU, $\alpha = 64, \beta = 2.5$



(f) **Bregman-GTV, 17th iter.**, with inner SIU, $\alpha = 64, \beta = 2.5$

Fig. 15 Bregman-GTV denoising via Split Inexact Uzawa of a noisy color photo collected with a smartphone camera at the beach in Santa Monica, CA. Image resolution: $968 \times 1296 \times 3$ px.

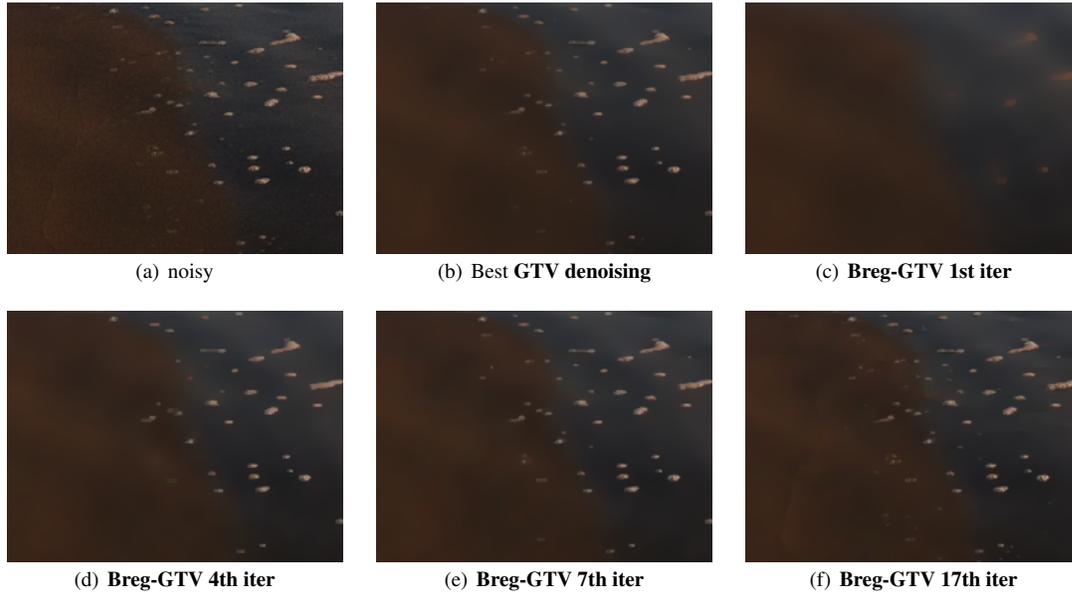


Fig. 16 Bottom left apertures corresponding to data and reconstructions in Fig. 15.

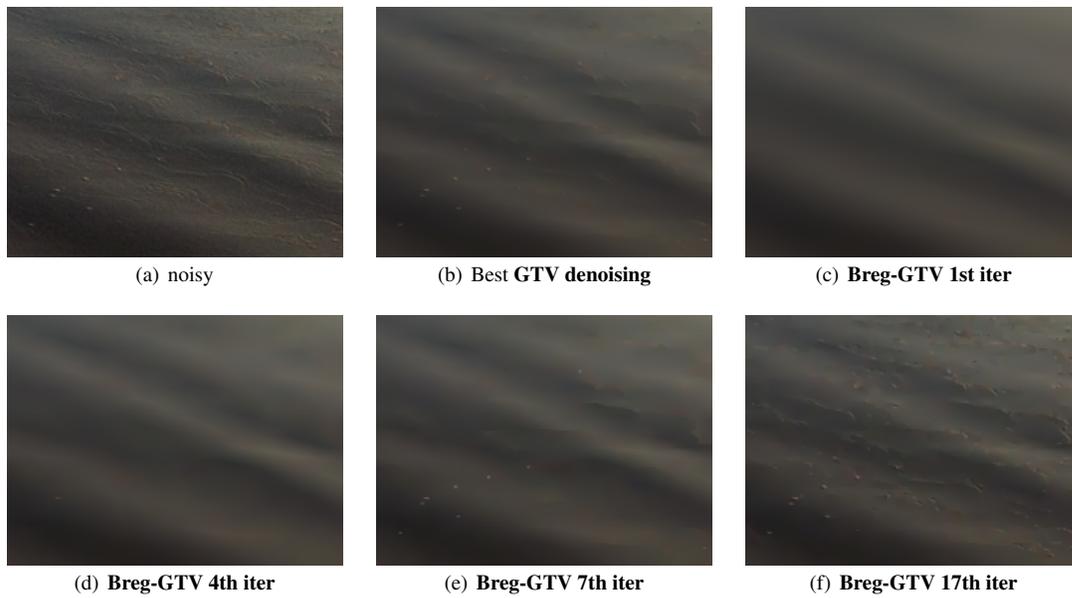


Fig. 17 Bottom right apertures corresponding to data and reconstructions in Fig. 15.

at interfaces and some interesting details and textures like the Santa Monica pier in the background, the waves, the seagulls or some fine structures in the still water close to the sand. Figure 15 illustrates the collected image, the best GTV denoising result we could obtain by tuning the regularization parameter, and the 1st, 4th, 7th and 17th iteration of Bregman-GTV with the Split Inexact Uzawa method for the inner GTV problem. Please note that for simplicity we applied our algorithms to all color channels separately in the same way. We can observe a nice removal of noise for the GTV denoising. However, with the higher-order inverse scale space method we can additionally recover important textures and sharp interfaces while taking care of the affinely linear parts in the image. This observation is underlined by the apertures in Figure 16 and Figure 17 taken from the data and reconstructions in Figure 15. Please note the revealed foam and line in the sand, as well as the thin lines in the water. In addition we plotted a line profile of the reconstructions at row 870 of 968 in Figure 18. This plot shows that Bregman-GTV nicely enhances fine structures like the foam peaks in the middle of Figure 18 (in between pixels 400 and 600), while regarding the smooth higher-order parts. The fact that we gain texture information in successive Bregman-GTV iterations can be explained by the underlying inverse scale space procedure reflecting a high-order gradient flow, compare (6.3). In each of the Bregman-GTV steps we update our reference function by a residual containing local texture information.

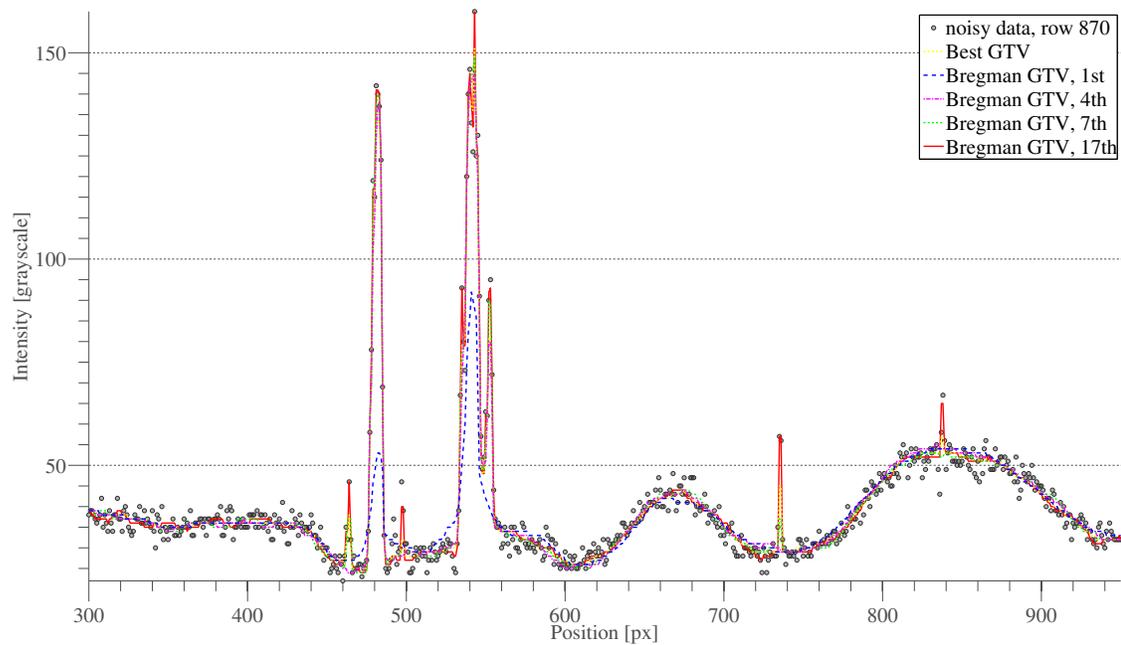


Fig. 18 This is a line profile at row 870 of 968 and a selected interval, pixels 300 – 950 in column direction, of the noisy data, best GTV reconstruction and Bregman-GTV reconstructions presented in Fig. 15 (after grayscale conversion for simplified visualization).

7 Conclusion and future directions

One of the standard regularization techniques in imaging science is the total variation. Although it has the desirable capability of recovering sharp edges in images and signals, TV can also produce staircase-like artifacts, which is less desirable, e.g. in some biomedical imaging applications. Modifications of total variation, incorporating derivatives of higher-order, have been proposed in literature, mainly to reduce this drawback. However, combining derivatives of different order still revealed other undesired side effects at the same time.

In this work we have focused on analyzing the capabilities and limitations of different higher-order TV models including TV^l , ICTV and GTV. We proved that their (semi-)norms are equivalent when restricting to BV functions with vanishing zeroth and first moment, whereas minimizers of associated variational problems can differ substantially. Thus, to compare capabilities of these models, we studied exact solutions with regard to eigenfunctions and presented several examples.

Besides general modeling and analysis of higher-order TV models, the second main contribution of this paper was to address the limitations of these models numerically. We improved higher-order TV reconstructions by introducing Bregman iterations. To realize Bregman-GTV efficiently, a primal-dual Newton method and a preconditioned Split Bregman method (Split Inexact Uzawa) have been proposed. With several synthetic and real-world denoising examples we underlined our analytical observations and illustrated the performance of the proposed Bregman-GTV algorithms.

Possible future directions are comparisons with other higher-order PDE methods in imaging science and the extension of Bregman-GTV to the case of new challenging inverse problems in biomedical imaging.

Acknowledgments

Financial support is acknowledged to the German Science Foundation (DFG) via grants SFB 656 (Subproject B2) and BU 2327/1. The third author thanks Stanley Osher for initiating his interest in variational methods for image processing.

References

- ACC05. F. Alter, V. Caselles, and A. Chambolle. Evolution of characteristic functions of convex sets in the plane by the minimizing total variation flow. *Interfaces Free Bound.*, 7(1):29–53, 2005. [12](#)
- BC11. Heinz H. Bauschke and Patrick L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, New York, 2011. With a foreword by Hedy Attouch. [4](#)
- BCM05. A. Buades, B. Coll, and J. M. Morel. A review of image denoising algorithms, with a new one. *Multiscale Model. Simul.*, 4(2):490–530, 2005. [2](#)
- Ben11. M. Benning. *Singular Regularization of Inverse Problems*. PhD thesis, University of Münster, Institute for Computational and Applied Mathematics, Einsteinstr. 62, 48149 Münster, May 2011. [12](#), [14](#), [25](#)
- BFOS07. M. Burger, K. Frick, S. Osher, and O. Scherzer. Inverse total variation flow. *Multiscale Model. Simul.*, 6(2):366–395, 2007.
- BG80. R. L. Bishop and S. I. Goldberg. *Tensor Analysis on Manifolds*. Dover Publications, December 1980. [3](#)
- BG04. Andrea L. Bertozzi and John B. Greer. Low-curvature image simplifiers: Global regularity of smooth solutions and Laplacian limiting schemes. *Communications on Pure and Applied Mathematics*, 57(6):764–790, June 2004.
- BGOV05. Andrea L. Bertozzi, John B. Greer, Stanley Osher, and Kevin Vixie. Nonlinear regularizations of TV based PDEs for image processing. *AMS Series of Contemporary Mathematics, Gui-Qiang Chen, George Gasper, and Joseph Jerome eds*, 371:29–40, 2005.
- BGOX06. M. Burger, G. Gilboa, S. Osher, and J. Xu. Nonlinear inverse scale space methods. *Comm. Math. Sci.*, 4:179–212, 2006. [2](#), [12](#), [23](#)
- BKP10. K. Bredies, K. Kunisch, and T. Pock. Total generalized variation. *SIAM Journal on Imaging Sciences*, 3(3):492–526, 2010. [1](#), [2](#), [3](#), [5](#), [11](#), [17](#), [20](#)
- BO04. M. Burger and S. Osher. Convergence rates of convex variational regularization. *Inverse Problems*, 20:1411–1421, 2004. [10](#)
- BO12. M. Burger and S. Osher. A guide to tv zoo 1: Models and analysis. In *Level Set and PDE-based Reconstruction Methods*. Springer, 2012. To appear. [3](#), [6](#)
- Bre67. L.M. Bregman. The relaxation method for finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Comp. Math. Math. Phys.*, 7:200–217, 1967.
- CCN07. Vicent Caselles, Antonin Chambolle, and Matteo Novaga. The discontinuity set of solutions of the TV denoising problem and some extensions. *Multiscale Model. Simul.*, 6(3):879–894, 2007. [10](#)
- CDVLL98. A. Chambolle, R. A. De Vore, N.-Y. Lee, and B. J. Lucier. Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage. *IEEE Trans. Image Process.*, 7(3):319–335, 1998.

- Cha04. Antonin Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imaging Vision*, 20(1-2):89–97, 2004. Special issue on mathematics and image analysis. [9](#)
- CL97. A. Chambolle and P.-L. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76:167–188, 1997. [1](#), [2](#), [4](#), [17](#)
- CMM00. Tony Chan, Antonio Marquina, and Pep Mulet. High-order total variation-based image restoration. *SIAM J. Sci. Comput.*, 22(2):503–516, 2000. [3](#)
- CP11. Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J Math Imaging Vis*, 40(1):120–145, 2011. [21](#), [22](#)
- CS05. T. F. Chan and J. Shen. *Image Processing and Analysis: Variational, PDE, Wavelet and Stochastic Methods*. SIAM, 2005. [2](#)
- EHN00. H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Mathematics and Its Applications. Kluwer Academic Publisher, 2000.
- ET76. I. Ekeland and R. Temam. *Convex Analysis and Variational Problems*, volume 1 of *Studies in Mathematics and Its Applications*. North-Holland Publishing Company, 1976.
- EZC10. Ernie Esser, Xiaoqun Zhang, and Tony F. Chan. A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM J. Imaging Sci.*, 3(4):1015–1046, 2010. [21](#)
- GO08. Guy Gilboa and Stanley Osher. Nonlocal operators with applications to image processing. *Multiscale Model. Simul.*, 7(3):1005–1028, 2008. [2](#)
- GO09. Tom Goldstein and Stanley Osher. The split bregman method for l1-regularized problems. *SIAM J. Imaging Sci.*, 2(2):323–343, 2009. [20](#)
- HPUU09. M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich. *Optimization with PDE constraints*, volume 23 of *Mathematical Modelling: Theory and Applications*. Springer, New York, 2009. [19](#)
- KBPS11. F. Knoll, K. Bredies, T. Pock, and R. Stollberger. Second order total generalized variation (TGV) for MRI. *Magnetic Resonance in Medicine*, 65(2):480–491, 2011. [20](#)
- Kiw97. Krzysztof C. Kiwiel. Proximal minimization methods with generalized Bregman functions. *SIAM J. Control Optim.*, 35(4):1142–1168, 1997. [10](#)
- Mey01. Y. Meyer. *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations: The Fifteenth Dean Jacqueline B. Lewis Memorial Lectures*, volume 22 of *University Lecture Series*. American Mathematical Society, Boston, MA, USA, 2001. [3](#), [12](#)
- OBG⁺05. S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total variation-based image restoration. *Multiscale Model. Simul.*, 4(2):460–489, 2005. [2](#), [12](#), [23](#), [25](#)
- ROF92. L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60:259–268, 1992. [1](#), [2](#)
- SC03. David Strong and Tony Chan. Edge-preserving and scale-dependent properties of total variation regularization. *Inverse Problems*, 19(6):S165–S187, 2003. Special section on imaging. [12](#)
- Sch98. O. Scherzer. Denoising with higher order derivatives of bounded variation and an application to parameter estimation. *Computing*, 60(1):1–27, 1998. [3](#)
- SRG10. W. Stefan, R. A. Renaut, and A. Gelb. Improved total variation-type regularization using higher order edge detectors. *SIAM Journal on Imaging Sciences*, 3(2):232–251, 2010.
- SS08. S. Setzer and G. Steidl. Variational methods with higher order derivatives in image processing. In M. Neamtu and L. L. Schumaker, editors, *Approximation XII*, Brentwood, 2008. Nashboro Press. [2](#)
- SST11. S. Setzer, G. Steidl, and T. Teuber. Infimal convolution regularizations with discrete l1-type functionals. *Communications in Mathematical Sciences*, 9(3):797–827, 2011. [2](#), [5](#), [11](#), [17](#), [20](#), [21](#), [27](#), [31](#)
- Vog02. C. R. Vogel. *Computational Methods for Inverse Problems*. Frontiers in Applied Mathematics. SIAM, 2002.
- ZBO11. Xiaoqun Zhang, Martin Burger, and Stanley Osher. A unified primal-dual algorithm framework based on bregman iteration. *Journal of Scientific Computing*, 46(1):20–46, 2011. [20](#)