SIAM J. IMAGING SCIENCES
Vol. 3, No. 3, pp. 492–526

Total Generalized Variation*

Kristian Bredies†, Karl Kunisch†, and Thomas Pock‡

Abstract. The novel concept of total generalized variation of a function u is introduced, and some of its essential properties are proved. Differently from the bounded variation seminorm, the new concept involves higher-order derivatives of u. Numerical examples illustrate the high quality of this functional as a regularization term for mathematical imaging problems. In particular this functional selectively regularizes on different regularity levels and, as a side effect, does not lead to a staircasing effect.

Key words. bounded variation, total generalized variation, tensor fields, regularization, image denoising

AMS subject classifications. 49J52, 49N45, 68U10

DOI. 10.1137/090769521

1. Introduction.

Most mathematical formulations of inverse problems and, in particular, of mathematical imaging problems are cast in the form

$$\min_u \ \mathcal{F}(u) + \mathcal{R}(u), \tag{1.1}$$

where $\mathcal{F}$ represents the data fidelity and $\mathcal{R}$ the regularization term. If G denotes the forward modeling operator, then the most common fidelity term is of the form

$$\mathcal{F}(u) = \frac{1}{2}\|G(u) - z\|^2, \tag{1.2}$$

where z stands for the possibly error-prone data and $\|\cdot\|$ denotes an appropriately chosen Hilbertian norm. Similarly, the most frequently chosen regularization term is given by

$$\mathcal{R}(u) = \frac{\alpha}{2}|u|^2, \tag{1.3}$$

where α is the regularization parameter and $|\cdot|$ again denotes a Hilbertian norm or seminorm. It is now becoming well accepted that the mathematical and computational simplicity of the norm-of-squares terms must be put into perspective noting some serious shortcomings. If the errors in the data contain outliers or if the error is of impulsive type, the fidelity terms suggested by methods from robust statistics should be preferred over (1.2). Similarly (1.3)

*Received by the editors September 1, 2009; accepted for publication (in revised form) May 12, 2010; published electronically September 9, 2010.
http://www.siam.org/journals/siims/3-3/76952.html
†Institute of Mathematics and Scientific Computing, University of Graz, Heinrichstraße 36, A-8010 Graz, Austria (kristian.bredies@uni-graz.at, karl.kunisch@uni-graz.at). The first two authors' work was supported by the Austrian Science Fund (FWF) under grant SFB F32 (SFB Research Center "Mathematical Optimization and Applications in Biomedical Sciences").
‡Institute for Computer Graphics and Vision, Graz University of Technology, Inffeldgasse 16, A-8010 Graz, Austria (pock@icg.tugraz.at). This author's work was supported by the Austrian Science Fund (FWF) under the doctoral program "Confluence of Vision and Graphics."

is frequently not an appropriate choice. In fact, the regularization term penalizes a certain property of u, which is quantified in the choice of the norm $|\cdot|$, and the natural proportionality by which this should enter into $\mathcal{R}$ would be 1-homogeneous rather than quadratic.

In the present paper the focus will be on the choice of $\mathcal{R}$. One of the early proposals for a refined choice was given in [ROF]. It uses the bounded variation seminorm

$$\mathcal{R}(u) = \alpha \int_\Omega |Du|, \tag{1.4}$$

where u is defined on the bounded domain $\Omega \subset \mathbb{R}^d$. This choice is highly effective when compared to, e.g., $\mathcal{R}(u) = \frac{\alpha}{2}\int_\Omega |\nabla u|^2 \, \mathrm{d}x$ if the data z to be reconstructed are piecewise constant, since it is more apt to preserve corners and edges. The bounded variation seminorm, however, also has some shortcomings, most notably the staircasing phenomenon. To briefly explain this effect, we assume that $G = I$ so that (1.1) describes the imaging denoising problem. If the true image contains not only flat but also slanted regions, then the image reconstructed on the basis of the bounded variation seminorm tends to be piecewise constant (staircasing). This staircasing effect was rigorously established in [Ni, CNN, R], for example. For various other aspects on the topic of constructing appropriate regularization or filter functionals in image reconstruction, we refer the reader to [SGGHL] and the references given there.

In this paper we propose and analyze the regularization term of the form

$$\mathrm{TGV}^k_\alpha(u) = \sup\left\{ \int_\Omega u \operatorname{div}^k v \, \mathrm{d}x \;\middle|\; v \in \mathcal{C}^k_{\mathrm{c}}(\Omega, \mathrm{Sym}^k(\mathbb{R}^d)), \; \|\operatorname{div}^l v\|_\infty \le \alpha_l, \; l = 0, \dots, k-1 \right\}, \tag{1.5}$$

where $\mathrm{Sym}^k(\mathbb{R}^d)$ denotes the space of symmetric tensors of order k with arguments in $\mathbb{R}^d$, and α_l are fixed positive parameters. For the definition of the remaining quantities, we ask for the reader's patience until section 2. Suffice it to say at this moment that for $k = 1$, $\alpha_0 = 1$ the seminorm TGV^k_α coincides with the bounded variation seminorm. We refer to TGV^k_α as total generalized bounded variation of order k with weight $\alpha \in \mathbb{R}^k$. From the definition of TGV^k_α it is immediately clear that it involves (generalized) derivatives of u of order $i = 1, \dots, k$, and that the kernel of TGV^k_α is the set of polynomials of degree less than k. Intuitively the total generalized bounded variation further automatically balances the first to the kth derivatives of u among themselves. It will be shown that TGV^k_α shares some properties of TV: It is also rotationally invariant and for $k = 2$, the total generalized variation of the indicator function of a smooth set $\Omega' \subset\subset \Omega$ equals $\alpha_1 \operatorname{Per}_{\Omega'} = \alpha_1 \operatorname{TV}(\chi_{\Omega'})$, where $\operatorname{Per}_{\Omega'}$ denotes the perimeter of Ω'. It differs, however, for functions which are not piecewise constant.

As a further preview we point out that in dimension 1, with $\Omega =]0, L[$, $k = 2$, $\alpha_0, \alpha_1 > 0$, we have for

$$u(x) = \sum_{i=1}^{2} p_i(x)\chi_{\Omega_i}, \qquad \text{with } p_i(x) = a_i x + b_i,$$

for each $a_1, a_2, b_1, b_2 \in \mathbb{R}$ and $\Omega_1 =]0, c[$, $\Omega_2 =]c, 1[$ that

$$\mathrm{TGV}^2_\alpha(u) = \alpha_1 |p_2(c) - p_1(c)| + \alpha_0 |p_1'(c) - p_2'(c)|,$$

provided that the jump point c is not near the boundary, i.e., $c \in [\alpha_0/\alpha_1, L - \alpha_0/\alpha_1]$. In particular, $\mathrm{TGV}^2_\alpha(u)$ does not penalize the derivative of order $l = 0, 1$ unless it jumps at c.

As already mentioned, our motivation for studying $\mathrm{TGV}^2_\alpha(u)$ is based on the fact that it involves and balances higher-order derivatives of u. As a consequence, it reduces the staircasing effect of the bounded variation functional. This will be demonstrated in our numerical experiments. The use of higher-order derivatives with the aim of reducing staircasing is not new. In [CL] the infimal-convolution functional

$$\min_{u_1+u_2=u} \int_\Omega |\nabla u_1| + \alpha |\nabla(\nabla u_2)| \, \mathrm{d}x$$

was proposed and proved to be practically efficient, eliminating the staircasing effect, for denoising problems with images which contain various grey levels as well as edges and corners. This idea was followed in a modified form in [CEP], where the regularization term is of the form

$$\min_{u_1+u_2=u} \int_\Omega |\nabla u_1| + \alpha |\Delta u_2| \, \mathrm{d}x;$$

i.e., the second derivative is replaced by the Laplacian, and a dual method for its numerical realization is derived.

A different functional was proposed and tested in [CMM]. It is given by

$$\int_\Omega |\nabla u| + \alpha \Phi\big(|\nabla u|\big)\big(\mathcal{L}(u)\big)^2 \, \mathrm{d}x,$$

where Φ is a real-valued function that reflects the presence of edges in the sense that its value approaches 0 when the gradient $|\nabla(u)|$ is large, and $\mathcal{L}(u)$ is an elliptic operator. For this choice of regularization functional the absence of the staircasing effect was verified in [DFLM]. In [PS] regularization terms of the form

$$\mathcal{R}(u) = \int_\Omega \left| D\nabla^{l-1}(u) \right|,$$

where $\int_\Omega |D\nabla^{l-1}(u)|$ denotes the total variation of the $(l-1)$th derivative of $u \in W^{l-1,1}$, were considered, and special structures of minimizers of the resulting problems (1.1) were investigated. Higher-order regularization functionals in the discrete setting which are related to our computations for the second-order case were further proposed and tested in [SS].

As we shall see in section 3, even for the case $k = 2$ the proposed functional TGV^k_α does not agree with those regularization functionals which were considered earlier in the literature. In particular, although convex, it is structurally different from any infimal-convolution functional, especially the approaches of [CL, CEP].

Let us give a brief outline of the following sections. Section 2 contains a compact treatise of tensor and, in particular, symmetric tensor analysis in a manner that is useful for the variational analysis context. The definition of total generalized variation norms and some of their basic properties are given in section 3. Based on Fenchel duality, an equivalent description of $\mathrm{TGV}^k_\alpha(u)$ is derived for u sufficiently regular. Moreover, the relationship to an appropriately defined k-fold infimal convolution is obtained. A subsection is devoted to the

special case $k = 2$. The description of the numerical procedure that was employed, as well as carefully selected numerical denoising experiments, is contained in section 4. Appendix A contains some basic results involving symmetric k-tensor fields.

2. Spaces of symmetric tensor fields. This section is mainly devoted to the introduction of the notions we are going to utilize in the main parts of this article. For many of the considerations which are going to follow, the concept of symmetric tensor fields plays a central role. Therefore, we give a rather extensive introduction to make this paper more self-contained and also for the convenience for those readers who are familiar with tensor analysis. Most of the results, although scattered in several chapters, can also be found in the books [BG, S].

We restrict ourselves mainly to a general introduction of symmetric tensor fields. Some more specific results can be found in Appendix A.

Throughout the paper, $d \geq 1$ denotes the dimension, which is typically 2 or 3, in applications. Let

$$\mathcal{T}^k(\mathbb{R}^d) = \big\{\xi : \underbrace{\mathbb{R}^d \times \cdots \times \mathbb{R}^d}_{k \text{ times}} \to \mathbb{R} \;\big|\; \xi \text{ } k\text{-linear}\big\},$$
$$\mathrm{Sym}^k(\mathbb{R}^d) = \big\{\xi : \underbrace{\mathbb{R}^d \times \cdots \times \mathbb{R}^d}_{k \text{ times}} \to \mathbb{R} \;\big|\; \xi \text{ } k\text{-linear and symmetric}\big\}$$

be the vector space of *k-tensors* and *symmetric k-tensors*, respectively (actually, these are spaces of $(0,k)$-tensors, but since we deal only with covariant vectors, we omit the 0). Here $\xi \in \mathcal{T}^k(\mathbb{R}^d)$ is called symmetric if $\xi(a_1, \ldots, a_k) = \xi(a_{\pi(1)}, \ldots, a_{\pi(k)})$ for all $\pi \in S_k$, where S_k denotes the permutation group of $\{1, \ldots, k\}$.

The case $k = 0$ corresponds to scalar values; for $k = 1$, $\mathrm{Sym}^1(\mathbb{R}^d) = \mathbb{R}^d$; and for $k = 2$, $\mathrm{Sym}^2(\mathbb{R}^d) = S^{d\times d}$; i.e., the space corresponds to symmetric matrices.

Note three basic operations for k-tensors. For $\xi \in \mathcal{T}^k(\mathbb{R}^d)$ and $\eta \in \mathcal{T}^l(\mathbb{R}^d)$ the *tensor product*

$$(\xi \otimes \eta)(a_1, \ldots, a_{k+l}) = \xi(a_1, \ldots, a_k)\eta(a_{k+1}, \ldots, a_{k+l})$$

yields an element of $\mathcal{T}^{k+l}(\mathbb{R}^d)$. We define the *trace* $\mathrm{tr}(\xi) \in \mathcal{T}^{k-2}(\mathbb{R}^d)$ of $\xi \in \mathcal{T}^k(\mathbb{R}^d)$ with $k \geq 2$ by

$$\mathrm{tr}(\xi)(a_1, \ldots, a_{k-2}) = \sum_{i=1}^{d} \xi(e_i, a_1, \ldots, a_{k-2}, e_i),$$

where e_i denotes the ith standard basis vector. This operation can be iterated; for example, $\mathrm{tr}^l(\xi \otimes \eta)$ for $\xi \in \mathrm{Sym}^{k+l}(\mathbb{R}^d)$ and $\eta \in \mathrm{Sym}^l(\mathbb{R}^d)$ yields a symmetric k-tensor. Every k-tensor $\xi \in \mathcal{T}^k(\mathbb{R}^d)$ can be symmetrized by

$$(|||\xi)(a_1, \ldots, a_k) = \frac{1}{k!} \sum_{\pi \in S_k} \xi(a_{\pi(1)}, \ldots, a_{\pi(k)}).$$

The symmetrization is a projection, i.e., $|||^2\xi = |||\xi$.

The spaces $\mathcal{T}^k(\mathbb{R}^d)$ and, consequently, $\mathrm{Sym}^k(\mathbb{R}^d)$ will be equipped with the scalar product

$$\xi \cdot \eta = \mathrm{tr}^k(\bar{\xi} \otimes \eta) = \sum_{p \in \{1,\dots,d\}^k} \xi(e_{p_1}, \dots, e_{p_k}) \eta(e_{p_1}, \dots, e_{p_k}),$$
$$\bar{\xi}(a_1, \dots, a_k) = \xi(a_k, \dots, a_1),$$

leading canonically to the norm $|\xi| = \sqrt{\xi \cdot \xi}$. Again, this is the absolute value for $k = 0$, for $k = 1$, this corresponds to the Euclidean norm in $\mathbb{R}^d$, and in case $k = 2$, we can identify $\xi \in \mathrm{Sym}^2(\mathbb{R}^d)$ with

$$\xi = \begin{pmatrix} \xi_{11} & \cdots & \xi_{1d} \\ \vdots & \ddots & \vdots \\ \xi_{1d} & \cdots & \xi_{dd} \end{pmatrix}, \qquad |\xi| = \left(\sum_{i=1}^d \xi_{ii}^2 + 2 \sum_{i<j} \xi_{ij}^2 \right)^{1/2}.$$

The scalar product, moreover, possesses the property that the symmetrization of a k-tensor becomes the orthogonal projection onto $\mathrm{Sym}^k(\mathbb{R}^d)$. Indeed, for $\xi \in \mathcal{T}^k(\mathbb{R}^d)$ and $\eta \in \mathrm{Sym}^k(\mathbb{R}^d)$ we have

$$\begin{aligned} \|\!| \xi \cdot \eta &= \frac{1}{k!} \sum_{\pi \in S_k} \sum_{p \in \{1,\dots,d\}^k} \xi(e_{p_{\pi(1)}}, \dots, e_{p_{\pi(k)}}) \eta(e_{p_1}, \dots, e_{p_k}) \\ &= \frac{1}{k!} \sum_{p \in \{1,\dots,d\}^k} \xi(e_{p_1}, \dots, e_{p_k}) \sum_{\pi \in S_k} \eta(e_{p_{\pi(1)}}, \dots, e_{p_{\pi(k)}}) = \xi \cdot \eta. \end{aligned}$$

In order to describe the structure of $\mathrm{Sym}^k(\mathbb{R}^d)$ as a vector space, it is useful to consider the following relation σ from $p \in \{1, \dots, d\}^k$ to multi-indices $\beta \in \mathbb{N}^d \cup \{0\}$ with $|\beta| = \sum_{i=1}^d \beta_i = k$, which denumerates the number of linearly independent symmetric tensors in $\mathrm{Sym}^k(\mathbb{R}^d)$ compared to all tensors of order k. It is given by

$$\sigma : \{1, \dots, d\}^k \to \mathbb{N}^d, \qquad \sigma(p)_i = \#\{j \mid p_j = i\}.$$

For each $\beta \in \mathbb{N}^d$ with $|\beta| = k$, one can associate a $p \in \{1, \dots, d\}^k$ by

$$\sigma^{-1}(\beta)_j = \min\left\{ m \;\middle|\; \sum_{n=1}^m \beta_n \geq j \right\},$$

which is only a right inverse of σ. In fact, there are several p for which $\sigma(p) = \beta$ with $|\beta| = k$. The cardinality of the set containing those p is known to be

$$\#\{p \mid \sigma(p) = \beta\} = \frac{k!}{\beta!} = \frac{k!}{\beta_1! \cdots \beta_d!}.$$

The multi-index notation reflects the fact that the order of elements does not matter for the symmetry we are considering. It is known that a basis of $\mathrm{Sym}^k(\mathbb{R}^d)$ can be obtained by

$$\beta \in \mathbb{N}^d,\ |\beta| = k : e_\beta(a_1, \dots, a_k) = \sum_{\substack{p \in \{1,\dots,d\}^k \\ \sigma(p) = \beta}} \prod_{j=1}^k a_{j,p_j}.$$

In this basis the representation of a symmetric tensor is given by

$$\xi = \sum_{\substack{\beta\in\mathbb{N}^d\\ |\beta|=k}} \xi_\beta e_\beta, \qquad \text{where } \xi_\beta = \xi(e_{p_1},\dots,e_{p_k}) \text{ and } p=\sigma^{-1}(\beta).$$

The tensor product of some $\xi\in \mathrm{Sym}^k(\mathbb{R}^d)$ and $\eta\in\mathrm{Sym}^l(\mathbb{R}^d)$ can, moreover, be expressed by

$$(\xi\otimes\eta) = \sum_{\substack{\beta\in\mathbb{N}^d\\ |\beta|=k}}\sum_{\substack{\gamma\in\mathbb{N}^d\\ |\gamma|=l}} (\xi\otimes\eta)_{\beta,\gamma}(e_\beta\otimes e_\gamma), \qquad (\xi\otimes\eta)_{\beta,\gamma}=\xi_\beta\eta_\gamma.$$

Hence, counting multiplicities, the l-trace of $\xi\in\mathrm{Sym}^{k+l}(\mathbb{R}^d)$ and $\eta\in\mathrm{Sym}^l(\mathbb{R}^d)$ obeys

$$\mathrm{tr}^l(\xi\otimes\eta)_\beta = \sum_{p\in\{1,\dots,d\}^l} \xi_{\beta+\sigma(p)}\eta_{\sigma(p)} = \sum_{|\gamma|=l}\frac{l!}{\gamma!}\xi_{\beta+\gamma}\eta_\gamma$$

for the basis coefficient associated with a $\beta\in\mathbb{N}^d$ with $|\beta|=k$. In particular, the scalar product on $\mathrm{Sym}^k(\mathbb{R}^d)$ is $\xi\cdot\eta=\sum_{|\beta|=k}\frac{k!}{\beta!}\xi_\beta\eta_\beta$.

Next, let $\Omega\subset\mathbb{R}^d$ be a fixed domain. We define *symmetric k-tensor fields* as mappings $\xi:\Omega\to\mathrm{Sym}^k(\mathbb{R}^d)$ and associate Lebesgue spaces with them:

$$L^p(\Omega,\mathrm{Sym}^k(\mathbb{R}^d)) = \big\{\xi:\Omega\to\mathrm{Sym}^k(\mathbb{R}^d) \text{ measurable, identified a.e. } \big| \|\xi\|_p<\infty\big\},$$
$$\|\xi\|_p = \left(\int_\Omega |\xi(x)|^p\,\mathrm{d}x\right)^{1/p} \quad \text{for } 1\le p<\infty, \qquad \|\xi\|_\infty = \operatorname*{ess\,sup}_{x\in\Omega}\ |\xi(x)|.$$

Let the spaces $L^p_{\mathrm{loc}}(\Omega,\mathrm{Sym}^k(\mathbb{R}^d))$ be defined through the usual modification:

$$L^p_{\mathrm{loc}}(\Omega,\mathrm{Sym}^k(\mathbb{R}^d)) = \big\{\xi:\Omega\to\mathrm{Sym}^k(\mathbb{R}^d) \text{ measurable, identified a.e. } \big| \ \|\xi|_{\Omega'}\|_p<\infty \text{ for each } \Omega'\subset\subset\Omega\big\}$$

with $\xi|_{\Omega'}$ denoting the restriction of ξ to Ω'. Note that, since the vector norm in $\mathrm{Sym}^k(\mathbb{R}^d)$ is induced by a scalar product, the usual duality holds: $L^p(\Omega,\mathrm{Sym}^k(\mathbb{R}^d))^* = L^{p^*}(\Omega,\mathrm{Sym}^k(\mathbb{R}^d))$ for $1\le p<\infty$ and $1/p+1/p^*=1$.

Moreover, denote by $\mathcal{C}(\overline{\Omega},\mathrm{Sym}^k(\mathbb{R}^d))$ the usual space of continuous functions equipped with $\|\cdot\|_\infty$ as well as

$$\mathcal{C}_{\mathrm{c}}(\Omega,\mathrm{Sym}^k(\mathbb{R}^d)) = \big\{\xi\in\mathcal{C}(\overline{\Omega},\mathrm{Sym}^k(\mathbb{R}^d)) \ \big|\ \operatorname{supp}\xi\subset\subset\Omega\big\},$$
$$\mathcal{C}_0(\Omega,\mathrm{Sym}^k(\mathbb{R}^d)) = \overline{\mathcal{C}_{\mathrm{c}}(\Omega,\mathrm{Sym}^k(\mathbb{R}^d))},$$

where the closure is taken in $\mathcal{C}(\overline{\Omega},\mathrm{Sym}^k(\mathbb{R}^d))$.

For spaces incorporating the (covariant) derivatives of a symmetric k-tensor field, the description is somewhat more involved, since the lth derivative, provided that it exists, is, in general, not symmetric with respect to all arguments. Nevertheless, it is a tensor field, for which we will use the notation

$$(\nabla^l\otimes\xi)(x)(a_1,\dots,a_{k+l}) = \big(\mathrm{D}^l\xi(x)(a_1,\dots,a_l)\big)(a_{l+1},\dots,a_{k+l}),$$

where $\mathrm{D}^l\xi : \Omega \to \mathcal{L}^l\big(\mathbb{R}^d, \mathrm{Sym}^k(\mathbb{R}^d)\big)$ denotes the lth Fréchet derivative of ξ and $\mathcal{L}^l\big(X, Y\big)$ is the space of l-linear and continuous mappings $X^l \to Y$.

As is also done in the mathematical theory of elasticity, for example, we are particularly interested in symmetrization of the derivative, i.e.,

$$\mathcal{E}^l(\xi) = |||(\nabla^l \otimes \xi) = \big(|||(\nabla\otimes)\big)^l \xi.$$

The last identity follows from the following observation for differentiable mappings $\xi : \Omega \to \mathcal{T}^k(\mathbb{R}^d)$:

$$\begin{aligned}
|||(\nabla \otimes \xi)(a_1, \dots, a_{k+1}) &= \frac{1}{(k+1)!} \sum_{\pi \in S_{k+1}} \sum_{i=1}^{d} a_{\pi(1),i} \frac{\partial \xi}{\partial x_i}(a_{\pi(2)}, \dots, a_{\pi(k+1)}) \\
&= \frac{1}{k+1} \sum_{j=1}^{k+1} \frac{1}{k!} \sum_{\substack{\pi \in S_{k+1} \\ \pi(1)=j}} \sum_{i=1}^{d} a_{j,i} \frac{\partial \xi}{\partial x_i}(a_{\pi(2)}, \dots, a_{\pi(k+1)}) \\
&= \frac{1}{k+1} \sum_{j=1}^{k+1} \sum_{i=1}^{d} a_{j,i} \frac{1}{k!} \sum_{\substack{\pi \in S_{k+1} \\ \pi(1)=j}} \frac{\partial(|||\xi)}{\partial x_i}(a_{\pi(2)}, \dots, a_{\pi(k+1)}) \\
&= \frac{1}{(k+1)!} \sum_{\pi \in S_{k+1}} \sum_{i=1}^{d} a_{\pi(1),i} \frac{\partial(|||\xi)}{\partial x_i}(a_{\pi(2)}, \dots, a_{\pi(k+1)}) \\
&= |||(\nabla \otimes |||\xi)(a_1, \dots, a_{k+1}).
\end{aligned}$$

Spaces of continuously differentiable functions in this sense are defined as

$$\begin{aligned}
\mathcal{C}^l(\overline{\Omega}, \mathrm{Sym}^k(\mathbb{R}^d)) = \big\{ \xi : \overline{\Omega} \to \mathrm{Sym}^k(\mathbb{R}^d) \\
\big| \ \nabla^m \otimes \xi \text{ continuous on } \overline{\Omega}, m = 0, \dots, l \big\}, \\
\|\xi\|_{l,\infty} = \max_{m=0,\dots,l} \|\mathcal{E}^m(\xi)\|_\infty.
\end{aligned}$$

We also use continuously differentiable symmetric k-tensor fields with compact support in Ω:

$$\begin{aligned}
\mathcal{C}^l_{\mathrm{c}}(\Omega, \mathrm{Sym}^k(\mathbb{R}^d)) &= \big\{ \xi \in \mathcal{C}^k(\overline{\Omega}, \mathrm{Sym}^k(\mathbb{R}^d)) \ \big| \ \mathrm{supp}\, \xi \subset\subset \Omega \big\}, \\
\mathcal{C}^\infty_{\mathrm{c}}(\Omega, \mathrm{Sym}^k(\mathbb{R}^d)) &= \bigcap_{l \geq 1} \mathcal{C}^l_{\mathrm{c}}(\Omega, \mathrm{Sym}^k(\mathbb{R}^d)).
\end{aligned}$$

One has, moreover, the notion of l-divergence of a symmetric $(k+l)$-tensor field ξ:

$$(\mathrm{div}^l\, \xi) = \mathrm{tr}^l(\nabla^l \otimes \xi), \quad \text{with} \quad (\mathrm{div}^l\, \xi)_\beta = \sum_{\gamma \in \mathbb{N}^d,\, |\gamma| = l} \frac{l!}{\gamma!} \frac{\partial^l \xi_{\beta+\gamma}}{\partial x^\gamma}, \tag{2.1}$$

where $\beta \in \mathbb{N}^d$, $|\beta| = k$. Note that this divergence operator corresponds to changing, via the standard metric, some index to a contravariant vector, taking the covariant derivative and contracting this index with the covariant index arising from differentiation. For the l-divergence,

this procedure is simply iterated; hence $\operatorname{div}^k \operatorname{div}^l \xi = \operatorname{div}^{k+l} \xi$ whenever the expression makes sense. Let us point out, moreover, the special case $k = 2$ with the interpretation as symmetric matrix:

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

For $k = 0$, this is the identity, while for $k = 1$, the divergence for mappings $\Omega \to \operatorname{Sym}^1(\mathbb{R}^d)$ coincides with the usual divergence.

With the definition of the divergence according to (2.1), the validity of a respective divergence theorem can be verified.

Proposition 2.1. *Let Ω be a bounded Lipschitz domain and let $\xi \in \mathcal{C}^1(\overline{\Omega}, \operatorname{Sym}^{k+1}(\mathbb{R}^d))$ be a smooth $(k+1)$-symmetric tensor field. Then, for each smooth symmetric k-tensor field $\eta : \Omega \to \operatorname{Sym}^k(\mathbb{R}^d)$ we have*

$$\int_\Omega \operatorname{div} \xi \cdot \eta \, \mathrm{d}x = \int_{\partial\Omega} \xi \cdot |||(\eta \otimes \nu) \, \mathrm{d}\mathcal{H}^{d-1}(x) - \int_\Omega \xi \cdot \mathcal{E}(\eta) \, \mathrm{d}x \tag{2.2}$$

with ν denoting the outer normal on $\partial\Omega$.

Proof. We just need to show that this statement corresponds to the usual integration by parts. We use again that $p \in \{1, \dots, d\}^k$ and $i = 1, \dots, d$ yields each $(p, i) \in \{1, \dots, d\}^{k+1}$, so we can express $\sigma((p, i)) = \sigma(p) + e_i$. Therefore, with integration by parts and remembering that the symmetrization is the orthogonal projection onto the space of symmetric k-tensors,

$$\begin{aligned}
\int_\Omega \operatorname{div} \xi \cdot \eta \, \mathrm{d}x &= \int_\Omega \operatorname{tr}^k\big(\operatorname{tr}(\nabla \otimes \xi) \otimes \eta\big) \, \mathrm{d}x \\
&= \int_\Omega \sum_{(p,i) \in \{1,\dots,d\}^{k+1}} \frac{\partial \xi_{\sigma(p)+e_i}}{\partial x_i} \eta_{\sigma(p)} \, \mathrm{d}x \\
&= \int_{\partial\Omega} \sum_{(p,i) \in \{1,\dots,d\}^{k+1}} \xi_{\sigma(p)+e_i} \nu_i \eta_{\sigma(p)} \, \mathrm{d}\mathcal{H}^{d-1}(x) \\
&\quad - \int_\Omega \sum_{(p,i) \in \{1,\dots,d\}^{k+1}} \xi_{\sigma((p,i))} \frac{\partial \eta_{\sigma(p)}}{\partial x_i} \, \mathrm{d}x \\
&= \int_{\partial\Omega} \xi \cdot (\eta \otimes \nu) \, \mathrm{d}\mathcal{H}^{d-1}(x) - \int_\Omega \xi \cdot (\nabla \otimes \eta) \, \mathrm{d}x \\
&= \int_{\partial\Omega} \xi \cdot |||(\eta \otimes \nu) \, \mathrm{d}\mathcal{H}^{d-1}(x) - \int_\Omega \xi \cdot \mathcal{E}(\eta) \, \mathrm{d}x,
\end{aligned}$$

yielding the desired identity. ∎

Remark 2.2. It is easy to see that if ξ has compact support in Ω, then (2.2) holds with the boundary term being zero and for arbitrary domains.

Having the device of integration by parts (2.2), we can define weak derivatives of symmetric k-tensor fields.

Definition 2.3. *For a given symmetric k-tensor field $\xi \in L^1_{\mathrm{loc}}(\Omega, \operatorname{Sym}^k(\mathbb{R}^d))$, a symmetric $(k+l)$-tensor field $\eta \in L^1_{\mathrm{loc}}(\Omega, \operatorname{Sym}^{k+l}(\mathbb{R}^d))$ is called the* lth weak symmetrized derivative *of*

ξ if the identity

$$\int_\Omega \eta \cdot \zeta \, \mathrm{d}x = (-1)^l \int_\Omega \xi \cdot \mathrm{div}^l \, \zeta \, \mathrm{d}x$$

is satisfied for all $\zeta \in \mathcal{C}_c^l(\Omega, \mathrm{Sym}^{k+l}(\mathbb{R}^d))$. In this case, we denote $\mathcal{E}^l(\xi) = \eta$.

Note again that since we test only with symmetric $(k+l)$-tensor fields, we are able to determine only the symmetrized gradients. The Sobolev spaces associated with this notion of derivative are then given by

$$H^{l,p}(\Omega, \mathrm{Sym}^k(\mathbb{R}^d)) = \{\xi \in L^p(\Omega, \mathrm{Sym}^k(\mathbb{R}^d)) \\ \big| \, \mathcal{E}^m(\xi) \in L^p(\Omega, \mathrm{Sym}^{k+m}(\mathbb{R}^d)), m = 0, \dots, l\},$$

$$\|\xi\|_{l,p} = \left(\sum_{m=0}^{l} \|\mathcal{E}^m(\xi)\|_p^p \right)^{1/p} \quad \text{for } 1 \le p < \infty, \quad \|\xi\|_{l,\infty} = \max_{m=0,\dots,l} \|\mathcal{E}^m(\xi)\|_\infty,$$

$$H_0^{l,p}(\Omega, \mathrm{Sym}^k(\mathbb{R}^d)) = \overline{\mathcal{C}_c^l(\Omega, \mathrm{Sym}^k(\mathbb{R}^d))} \subset H^{l,p}(\Omega, \mathrm{Sym}^k(\mathbb{R}^d)).$$

3. Total generalized variation seminorms.

3.1. Basic properties.

We are now able to formulate the definition of the total generalized variation.

Definition 3.1. *Let $\Omega \subset \mathbb{R}^d$ be a domain, let $k \ge 1$, and let $\alpha_0, \dots, \alpha_{k-1} > 0$. Then, the* total generalized variation *of order k with weight α for $u \in L^1_{\mathrm{loc}}(\Omega)$ is defined as the value of the functional*

$$\mathrm{TGV}_\alpha^k(u) = \sup \left\{ \int_\Omega u \, \mathrm{div}^k v \, \mathrm{d}x \;\middle|\; v \in \mathcal{C}_c^k(\Omega, \mathrm{Sym}^k(\mathbb{R}^d)), \right. \\ \left. \|\mathrm{div}^l v\|_\infty \le \alpha_l, \; l = 0, \dots, k-1 \right\}, \tag{3.1}$$

where the supremum admits the value ∞ where the respective set is unbounded from above.

The space

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

is called the space of functions of bounded generalized variation *of order k with weight α.*

Remark 3.2. For $k = 1$ and $\alpha > 0$, we see that

$$\begin{aligned} \mathrm{TGV}_\alpha^1(u) &= \alpha \sup \left\{ \int_\Omega u \, \mathrm{div} \, v \, \mathrm{d}x \;\middle|\; v \in \mathcal{C}_c^1(\Omega, \mathrm{Sym}^1(\mathbb{R}^d)), \; \|v\|_\infty \le 1 \right\} \\ &= \alpha \, \mathrm{TV}(u). \end{aligned}$$

Thus one can indeed speak of a generalization of the total variation.

In the following, we will derive some basic properties of the total generalized variation.

Proposition 3.3. *The following statements hold:*

1. *TGV_α^k is a seminorm on the normed space $\mathrm{BGV}_\alpha^k(\Omega)$.*
2. *For $u \in L^1_{\mathrm{loc}}(\Omega)$, $\mathrm{TGV}_\alpha^k(u) = 0$ if and only if u is a polynomial of degree less than k.*

3. *For fixed k and positive weights $\alpha, \tilde{\alpha} \in \mathbb{R}^k$, the seminorms TGV^k_α and $\mathrm{TGV}^k_{\tilde{\alpha}}$ are equivalent.*

4. TGV^k_α *is rotationally invariant; i.e., for each orthonormal matrix $O \in \mathbb{R}^{d\times d}$ and $u \in \mathrm{BGV}^k_\alpha(\Omega)$, we have that $\tilde{u} \in \mathrm{BGV}^k_\alpha(O^{\mathrm{T}}\Omega)$ with $\mathrm{TGV}^k_\alpha(\tilde{u}) = \mathrm{TGV}^k_\alpha(u)$, where $\tilde{u}(x) = u(Ox)$.*

5. *For $r > 0$ and $u \in \mathrm{BGV}^k_\alpha(\Omega)$, we have, defining $\tilde{u}(x) = u(rx)$, that $\tilde{u} \in \mathrm{BGV}^k_\alpha(r^{-1}\Omega)$ with*

$$\mathrm{TGV}^k_\alpha(\tilde{u}) = r^{-d}\,\mathrm{TGV}^k_{\tilde{\alpha}}(u), \quad \tilde{\alpha}_l = \alpha_l r^{k-l}.$$

Proof. Let us begin by proving statement 1. Note that TGV can be interpreted as the dual seminorm in which the set

$$K^k_\alpha(\Omega) = \left\{ \operatorname{div}^k v \,\middle|\, v \in \mathcal{C}^k_{\mathrm{c}}(\Omega, \mathrm{Sym}^k(\mathbb{R}^d)),\ \|\operatorname{div}^l v\|_\infty \le \alpha_l,\ l = 0, \dots, k-1 \right\} \tag{3.2}$$

is taking the role of the "predual unit ball":

$$\mathrm{TGV}(u) = \sup_{w \in K^k_\alpha} \int_\Omega uw \,\mathrm{d}x. \tag{3.3}$$

It is easy to see that $K^k_\alpha(\Omega)$ is balanced and convex. The former implies that TGV^k_α is positively 1-homogeneous, while the latter yields its convexity and, consequently, the triangle inequality. This proves the seminorm property as well as the assertion that $\mathrm{BGV}^k_\alpha(\Omega)$ is a normed linear space.

For statement 2, suppose u is a polynomial of degree less than k, which means that $\nabla^k u = \mathcal{E}^k(u) = 0$. Using the defining integral (3.1) and the divergence formula (2.2) therefore yields, for $v \in \mathcal{C}^k_{\mathrm{c}}(\Omega, \mathrm{Sym}^k(\mathbb{R}^d))$,

$$\int_\Omega u \operatorname{div}^k v \,\mathrm{d}x = (-1)^k \int_\Omega \nabla^k u \cdot v \,\mathrm{d}x = 0 \quad \text{implies that} \quad \mathrm{TGV}(u) = 0.$$

Now suppose that $\mathrm{TGV}(u) = 0$ for some $u \in L^1_{\mathrm{loc}}(\Omega)$. For each $v \in \mathcal{C}^k_{\mathrm{c}}(\Omega, \mathrm{Sym}^k(\mathbb{R}^d))$, one can find a $\lambda > 0$ such that $\lambda v \in K^k_\alpha(\Omega)$, and test with λv and $-\lambda v$ to get

$$\int_\Omega u \operatorname{div}^k v \,\mathrm{d}x = 0 \qquad \text{for all } v \in \mathcal{C}^k_{\mathrm{c}}(\Omega, \mathrm{Sym}^k(\mathbb{R}^d)).$$

Hence, $\nabla^k u = 0$ in the weak sense, which immediately implies, via induction, that u is a polynomial of degree less than k since Ω is connected.

The asserted equivalence of norms according to statement 3 can be proved by the following observation:

$$c = \frac{\min_k \tilde{\alpha}_k}{\max_k \alpha_k} \quad \Rightarrow \quad cK^k_\alpha(\Omega) \subset K^k_{\tilde{\alpha}}(\Omega) \quad \Rightarrow \quad c\,\mathrm{TGV}^k_\alpha(u) \le \mathrm{TGV}^k_{\tilde{\alpha}}(u)$$

for each $u \in L^1_{\mathrm{loc}}(\Omega)$. Interchanging the roles of α and $\tilde{\alpha}$ leads to the desired equivalence.

For proving the rotational invariance as stated in statement 4, we use (A.4) to see that for orthonormal $O \in \mathbb{R}^{d\times d}$

$$v \in \mathcal{C}^k_{\mathrm{c}}(O^{\mathrm{T}}\Omega, \mathrm{Sym}^k(\mathbb{R}^d)) \quad \Leftrightarrow \quad \tilde{v} = (v \circ O^{\mathrm{T}})O^{\mathrm{T}} \in \mathcal{C}^k_{\mathrm{c}}(\Omega, \mathrm{Sym}^k(\mathbb{R}^d))$$

with $\operatorname{div}^l \tilde{v} = \big((\operatorname{div}^l v) \circ O^{\mathrm{T}}\big) O^{\mathrm{T}}$. Hence,

$$\|\operatorname{div}^l \tilde{v}\|_\infty = \big\|\big((\operatorname{div}^l v) \circ O^{\mathrm{T}}\big) O^{\mathrm{T}}\big\|_\infty = \|(\operatorname{div}^l v) \circ O^{\mathrm{T}}\|_\infty = \|\operatorname{div}^l v\|_\infty,$$

implying that $v \in K^k_\alpha(O^{\mathrm{T}}\Omega)$ if and only if $\tilde{v} \in K^k_\alpha(\Omega)$. Eventually, for each $v \in K^k_\alpha(O^{\mathrm{T}}\Omega)$ we have

$$\int_{O^{\mathrm{T}}\Omega} u(Ox) \operatorname{div}^k v(x)\, \mathrm{d}x = \int_\Omega u(x) \operatorname{div}^k v(O^{\mathrm{T}}x)\, \mathrm{d}x = \int_\Omega u(x) \operatorname{div}^k \tilde{v}(x)\, \mathrm{d}x$$

and, consequently, $\mathrm{TGV}^k_\alpha(u \circ O) = \mathrm{TGV}^k_\alpha(u)$.

The scaling behavior asserted in statement 5 can be seen as follows. Observe the identity

$$\operatorname{div}^l(v \circ r^{-1} I) = r^{-l} (\operatorname{div}^l v) \circ r^{-1} I$$

such that for $v \in \mathcal{C}^l_{\mathrm{c}}(r^{-1}\Omega, \operatorname{Sym}^k(\mathbb{R}^d))$ and $\hat{v} = r^k v \circ r^{-1} I$ we have $\|\operatorname{div}^l \hat{v}\|_\infty = r^{k-l} \|\operatorname{div}^l v\|_\infty$. Consequently, $v \in K^k_\alpha(r^{-1}\Omega)$ if and only if $\hat{v} \in K^k_{\tilde{\alpha}}(\Omega)$ as well as

$$\begin{aligned}\int_{r^{-1}\Omega} u(rx) \operatorname{div}^k v(x)\, \mathrm{d}x &= r^{-d} \int_\Omega u(x) \operatorname{div}^k v(r^{-1}x)\, \mathrm{d}x \\ &= r^{-d} \int_\Omega u(x) \operatorname{div}^k \hat{v}(x)\, \mathrm{d}x;\end{aligned}$$

hence $\mathrm{TGV}^k_\alpha(u \circ rI) = r^{-d}\, \mathrm{TGV}^k_{\tilde{\alpha}}(u)$. ∎

Remark 3.4. Because of statement 3 of Proposition 3.3, we write $\mathrm{BGV}^k(\Omega)$ instead of $\mathrm{BGV}^k_\alpha(\Omega)$, since the spaces are equivalent.

Proposition 3.5. *Each* $\mathrm{BGV}^k(\Omega)$ *becomes a Banach space when equipped with* $\|\cdot\|_{\mathrm{BGV}^k_\alpha}$ *for some weight* $\alpha > 0$.

Proof. Due to Remark 3.4 we have to prove only that $\mathrm{BGV}^k = \mathrm{BGV}^k_\alpha$ is complete for some weight α.

To achieve this, we first show that TGV^k_α always gives a lower semicontinuous functional with respect to $L^1(\Omega)$. For that purpose, let the sequence $\{u^n\}$ be in $\mathrm{BGV}^k(\Omega)$ such that $u^n \to u$ in $L^1(\Omega)$. Then, for each $v \in \mathcal{C}^k_{\mathrm{c}}(\Omega, \operatorname{Sym}^k(\mathbb{R}^d))$ with $\|\operatorname{div}^l v\|_\infty \leq \alpha_l$, it follows that

$$\int_\Omega u \cdot \operatorname{div}^k v\, \mathrm{d}x = \lim_{n\to\infty} \int_\Omega u^n \cdot \operatorname{div}^k v\, \mathrm{d}x \leq \liminf_{n\to\infty} \mathrm{TGV}^k_\alpha(u^n).$$

Taking the supremum thus yields

$$\mathrm{TGV}^k_\alpha(u) \leq \liminf_{n\to\infty} \mathrm{TGV}^k_\alpha(u^n),$$

meaning that TGV^k_α is indeed lower semicontinuous as stated.

Now, let $\{u^n\}$ be a Cauchy sequence in $\mathrm{BGV}^k(\Omega)$. It follows immediately that $\{u^n\}$ is a Cauchy sequence in $L^1(\Omega)$; hence a limit $u \in L^1(\Omega)$ exists for which the lower semicontinuity yields $\mathrm{TGV}^k_\alpha(u) \leq \liminf_{n\to\infty} \mathrm{TGV}^k_\alpha(u^n)$. Consequently, $u \in \mathrm{BGV}^k(\Omega)$, and it remains only to show that u is also the limit in the corresponding norm. But this follows again from the

lower semicontinuity of TGV_α^k on $L^1(\Omega)$: For each $\varepsilon > 0$ one chooses an n such that for all $n' \geq n$ it holds that $\mathrm{TGV}_\alpha^k(u^n - u^{n'}) \leq \varepsilon$, so letting $n' \to \infty$ gives

$$\mathrm{TGV}_\alpha^k(u^n - u) \leq \liminf_{n' \to \infty} \mathrm{TGV}_\alpha^k(u^n - u^{n'}) \leq \varepsilon,$$

implying that $u^n \to u$ in $\mathrm{BGV}^k(\Omega)$. ∎

Proposition 3.6. *Let $k \geq 1$, $\alpha_0, \ldots, \alpha_{k-1} > 0$, and let $u : \Omega \to \mathbb{R}$ be such that*

$$u(x) = \sum_{i=1}^{n} \chi_{\Omega_i} q_i(x),$$

where $\Omega_i \subset \Omega$ are disjoint Lipschitz domains and q_i are polynomials of degree less than k. Then,

$$\mathrm{TGV}_\alpha^k(u) \leq \frac{1}{2} \sum_{i,j=0}^{n} \int_{\Gamma_{ij}} \sum_{l=0}^{k-1} \alpha_l \left| \left| \left| (\nabla^{k-1-l}(q_i - q_j) \otimes \nu_i) \right| \right| \right| \, \mathrm{d}\mathcal{H}^{d-1}(x), \tag{3.4}$$

where $\Omega_0 = \Omega \backslash \bigcup_{i=1}^n \overline{\Omega_i}$; $q_0 = 0$; $\Gamma_{ij} = \partial\Omega_i \cap \partial\Omega_j \cap \Omega$; and ν_i is the outer normal to Ω_i.

In Proposition 3.11 a special case where equality holds in (3.4) is given.

Proof. Fix a $v \in \mathcal{C}_c^k(\Omega, \mathrm{Sym}^k(\mathbb{R}^d))$ for which $\|\mathrm{div}^l v\|_\infty \leq \alpha_l$ and integrate over Ω_i for $i = 0, \ldots, n$. Using the divergence theorem (2.2) k times, we deduce that

$$\int_{\Omega_i} u \, \mathrm{div}^k v \, \mathrm{d}x = \sum_{l=0}^{k-1} (-1)^l \int_{\partial\Omega_i \cap \Omega} \left| \left| \left| (\nabla^{k-1-l} q_i \otimes \nu_i) \cdot \mathrm{div}^l v \right| \right| \right| \, \mathrm{d}\mathcal{H}^{d-1}(x).$$

Since all Ω_i are disjoint, it is possible to write

$$\partial\Omega_i \cap \Omega = \partial\Omega_i \cap \partial(\Omega \backslash \Omega_i) \cap \Omega = \partial\Omega_i \cap \partial \left(\bigcup_{j \neq i} \overline{\Omega_j} \right) \cap \Omega = \bigcup_{j=1, j \neq i}^{n} \partial\Omega_i \cap \partial\Omega_j \cap \Omega;$$

hence, summation over the corresponding integral for $i, j \in \{0, \ldots, n\}$ gives, after rearranging and filling in zeros for $i = j$,

$$2 \int_\Omega u \, \mathrm{div}^k v \, \mathrm{d}x = \sum_{i,j=0}^{n} \sum_{l=0}^{k-1} (-1)^l \int_{\Gamma_{ij}} \left| \left| \left| (\nabla^{k-1-l}(q_i - q_j) \otimes \nu_i) \cdot \mathrm{div}^l v \right| \right| \right| \, \mathrm{d}\mathcal{H}^{d-1}(x)$$

since each boundary part is counted twice. Now, on the respective boundaries, $\left|\mathrm{div}^l v(x)\right| \leq \alpha_l$, leading to

$$\int_\Omega u \, \mathrm{div}^k v \, \mathrm{d}x \leq \frac{1}{2} \sum_{i,j=0}^{n} \int_{\Gamma_{ij}} \sum_{l=0}^{k-1} \alpha_l \left| \left| \left| (\nabla^{k-1-l}(q_i - q_j) \otimes \nu_i) \right| \right| \right| \, \mathrm{d}\mathcal{H}^{d-1}(x)$$

and, consequently, to the desired estimate. ∎

Remark 3.7. Estimate (3.4) tells us that the total generalized variation measures piecewise polynomial functions in terms of the jumps of the derivatives at the respective boundaries of Ω_i. In particular, TGV_α^k will not penalize if some $\nabla^l u$, $l = 0, \dots, k-1$, does not jump on some part of the boundary of Ω_i.

Remark 3.8. In an informal sense one can, for smooth functions, interpret TGV_α^k as a k-fold infimal convolution of inf-L^1-type functionals evaluated at $\nabla^k u$. Indeed, defining for $l = 0, \dots, k-1$ and $\bar{\alpha} > 0$ the sets

$$K_{\bar{\alpha}}^l = \left\{ v \;\middle|\; v \in \mathcal{C}_c^k(\Omega, \mathrm{Sym}^k(\mathbb{R}^d)),\ \|\mathrm{div}^l v\|_\infty \le \bar{\alpha} \right\},$$

one can see, for $\alpha = (\alpha_0, \dots, \alpha_{k-1})$, that in terms of indicator functionals and Fenchel duality,

$$\begin{aligned} \mathrm{TGV}_\alpha^k(u) &= \sup \left\{ \int_\Omega v(-\nabla)^k u \, \mathrm{d}x \;\middle|\; v \in K_{\alpha_l}^l,\ k = 0, \dots, k-1 \right\} \\ &= \left(\sum_{l=0}^{k-1} I_{K_{\alpha_l}^l} \right)^* \left((-\nabla)^k u \right). \end{aligned}$$

Hence, informally employing the Fenchel–Rockafellar duality formula, we have

$$\begin{aligned} I_{K_{\bar{\alpha}}^l}^*(w) &= \sup_{v \in \mathcal{C}_c^k(\Omega, \mathrm{Sym}^k(\mathbb{R}^d))} - \int_\Omega (-w) \cdot v \, \mathrm{d}x - I_{\{\|\bar{v}\|_\infty \le \bar{\alpha}\}}(\mathrm{div}^l v) \\ &= \inf_{u_l \in \mathcal{C}^l(\overline{\Omega}, \mathrm{Sym}^{k-l}(\mathbb{R}^d))} I_{\{w\}}\left((-\mathcal{E})^l(u_l) \right) + \bar{\alpha} \|u_l\|_1 \\ &= \inf_{\substack{u_l \in \mathcal{C}^l(\overline{\Omega}, \mathrm{Sym}^{k-l}(\mathbb{R}^d)) \\ \mathcal{E}^l(u_l) = w}} \bar{\alpha} \|u_l\|_1, \end{aligned}$$

where I denotes the indicator function of the set specified by the subscript in the respective space. By k-fold infimal convolution this implies that

$$\mathrm{TGV}_\alpha^k(u) = \inf_{\substack{\nabla^k u = u_0 + \mathcal{E}(u_1) + \\ \cdots + \mathcal{E}^{k-1}(u_{k-1})}} \sum_{l=0}^{k-1} \alpha_l \|u_l\|_1. \tag{3.5}$$

Note that the infimum in each $I_{K_{\alpha_l}^l}^*$ is absorbed by the infimal convolution. The resulting functional can therefore be called an infimal convolution of nontrivial infimal L^1-type functionals which is structurally different from each straightforward infimal convolution of L^1-type norms.

Unfortunately, the representation is hard to interpret since it operates on the decomposition of $\nabla^k u$ into the ranges of $\mathcal{E}^l$ for all orders from 0 to $k-1$, i.e., on the highest level of differentiation. We therefore derive, in the following, a different representation which is based on successive differentiation performed k times.

Remark 3.9. The "dualization" in the definition of the functional TGV_α^k can also be informally interpreted in terms of iterated Fenchel duality.

To see this connection, let $\kappa = 1, \dots, k$ and introduce the functionals

$$\mathrm{GV}_\alpha^\kappa(w_{k-\kappa+1}) = \sup \left\{ -\int_\Omega w_{k-\kappa+1} \cdot \operatorname{div}^{\kappa-1} v \, \mathrm{d}x \;\middle|\; v \in \mathcal{C}_\mathrm{c}^k(\Omega, \operatorname{Sym}^k(\mathbb{R}^d)), \right. \\ \left. \|\operatorname{div}^l v\|_\infty \le \alpha_l, \; l = 0, \dots, \kappa-1 \right\},$$

where $w_{k-\kappa+1} : \Omega \to \operatorname{Sym}^{k-\kappa+1}(\mathbb{R}^d)$ is locally integrable. Note that $\mathrm{TGV}_\alpha^k(u) = \mathrm{GV}_\alpha^k(\nabla u)$ for sufficiently smooth u. With the sets

$$K_l = \left\{ v \in \mathcal{C}_\mathrm{c}^k(\Omega, \operatorname{Sym}^k(\mathbb{R}^d)) \;\middle|\; \|\operatorname{div}^l v\|_\infty \le \alpha_l \right\},$$

and $\kappa - 1$ times integration by parts, the functional becomes

$$\mathrm{GV}_\alpha^\kappa(w_{k-\kappa+1}) = \left(\sum_{l=0}^{\kappa-1} I_{K_l} \right)^* \left((-1)^\kappa \mathcal{E}^{\kappa-1}(w_{k-\kappa+1}) \right).$$

Thus, for smooth $u_{k-\kappa} : \Omega \to \operatorname{Sym}^{k-\kappa}(\mathbb{R}^d)$, we deduce with the help of the Fenchel duality formula (for the operator $\operatorname{div}^{\kappa-1}$) that

$$\begin{aligned}
&\mathrm{GV}_\alpha^\kappa\big(\mathcal{E}(u_{k-\kappa})\big) \\
&\quad = \sup_{v \in \mathcal{C}_\mathrm{c}^k(\Omega, \operatorname{Sym}^k(\mathbb{R}^d))} -\int_\Omega \mathcal{E}(u_{k-\kappa}) \cdot \operatorname{div}^{\kappa-1} v \, \mathrm{d}x - I_{K_{\kappa-1}}(v) - \left(\sum_{l=0}^{\kappa-2} I_{K_l} \right)(v) \\
&\quad = \sup_{v \in \mathcal{C}_\mathrm{c}^k(\Omega, \operatorname{Sym}^k(\mathbb{R}^d))} -\big(I_{\{\|\cdot\|_\infty \le \alpha_{\kappa-1}\}} + \langle \mathcal{E}(u_{k-\kappa}), \, \cdot \, \rangle \big)(\operatorname{div}^{\kappa-1} v) - \left(\sum_{l=0}^{\kappa-2} I_{K_l} \right)(v) \\
&\quad = \inf_{u_{k-\kappa+1} \in \mathcal{C}^{\kappa-1}(\overline{\Omega}, \operatorname{Sym}^{k-\kappa+1}(\mathbb{R}^d))} \alpha_{\kappa-1} \|\mathcal{E}(u_{k-\kappa}) - u_{k-\kappa+1}\|_1 \\
&\qquad\qquad + \left(\sum_{l=0}^{\kappa-2} I_{K_l} \right)^* \left((-1)^{\kappa-1} \mathcal{E}^{\kappa-1}(u_{k-\kappa+1}) \right) \\
&\quad = \inf_{u_{k-\kappa+1} \in \mathcal{C}^{\kappa-1}(\overline{\Omega}, \operatorname{Sym}^{k-\kappa+1}(\mathbb{R}^d))} \alpha_{\kappa-1} \|\mathcal{E}(u_{k-\kappa}) - u_{k-\kappa+1}\|_1 + \mathrm{GV}_\alpha^{\kappa-1}\big(\mathcal{E}(u_{k-\kappa+1})\big).
\end{aligned}$$

Iterating this procedure through $\kappa = k, \dots, 2$ and observing the identity $\mathrm{GV}_\alpha^1\big(\mathcal{E}(u_{k-1})\big) = \alpha_0 \|\mathcal{E}(u_{k-1})\|_1$ leads to

$$\mathrm{GV}_\alpha^k\big(\mathcal{E}(u_0)\big) = \inf_{\substack{u_l \in \mathcal{C}^{k-l}(\overline{\Omega}, \operatorname{Sym}^l(\mathbb{R}^d)) \\ l=1,\dots,k-1}} \left(\sum_{l=1}^{k-1} \alpha_{k-l} \|\mathcal{E}(u_{l-1}) - u_l\|_1 \right) + \alpha_0 \|\mathcal{E}(u_{k-1})\|_1$$

and, consequently,

$$\mathrm{TGV}_\alpha^k(u) = \inf_{\substack{u_l \in \mathcal{C}^{k-l}(\overline{\Omega}, \operatorname{Sym}^l(\mathbb{R}^d)) \\ l=1,\dots,k-1, \; u_0 = u, \; u_k = 0}} \sum_{l=1}^{k} \alpha_{k-l} \|\mathcal{E}(u_{l-1}) - u_l\|_1. \tag{3.6}$$

Note that the tensor fields u_l are in different spaces for varying l. Moreover, the operators $\mathcal{E}$ (for different tensor orders) are involved, in contrast to a straightforward k-fold infimal convolution where the corresponding operator is the identity. Again, this implies that regularization with TGV_α^k differs significantly from all straightforward infimal-convolution based L^1-type regularization approaches.

The representation (3.6) can be seen as a balancing between the derivatives of order $1, \dots, k$. For simplicity we restrict our argumentation to $k = 2$: The gradient $\mathcal{E}(u_0) = \nabla u$ is decomposed into $\mathcal{E}(u_0) - u_1$, and $\mathrm{TGV}_\alpha^2(u)$ involves the 1-norm of $\mathcal{E}(u_0) - u_1$ and $\mathcal{E}(u_1)$ with appropriate weights. So, with $\mathcal{E}^2 u = \mathcal{E}(\mathcal{E}u_0 - u_1) + \mathcal{E}u_1$ in mind, if locally, i.e., on some subdomain Ω' with $\overline{\Omega'} \subset\subset \Omega$, it holds that $\|\nabla^2 u\|_1 \gg \|\nabla u\|_1$, then choosing $u_1 \approx 0$ locally might already minimize (3.6), and hence, the functional locally resembles the total variation. If, on the other hand, $\mathcal{E}(u_0)$ is locally flat, then it is favorable to choose $u_1 \approx \mathcal{E}(u_0)$ since $\|\mathcal{E}(u_1)\|_1 \approx \|\mathcal{E}^2(u_0)\|_1 = \|\nabla^2 u\|_1$ will be locally much lower than $\|\mathcal{E}(u_0) - u_1\|_1$. In this case, the functional behaves more like the 1-norm of the second derivative. Arguing recursively, one can again say that TGV_α^k adapts to the smoothness of u (up to the order k) in a certain sense.

Remark 3.10. From (3.6) it also becomes clear how the symmetry of the test functions, i.e., the space $\mathcal{C}_c^k(\Omega, \mathrm{Sym}^k(\mathbb{R}^d))$, influences the functional. If we had taken $\mathcal{C}_c^k(\Omega, \mathcal{T}^k(\mathbb{R}^d))$ in Definition 3.1 instead, i.e.,

$$\neg\,\mathrm{symTGV}_\alpha^k(u) = \sup\left\{ \int_\Omega u \,\mathrm{div}^k v \,\mathrm{d}x \;\middle|\; v \in \mathcal{C}_c^k(\Omega, \mathcal{T}^k(\mathbb{R}^d)), \right.$$
$$\left. \|\mathrm{div}^l v\|_\infty \le \alpha_l,\ l = 0, \dots, k-1 \right\},$$

we would have ended in

$$\neg\,\mathrm{symTGV}_\alpha^k(u) = \inf_{\substack{u_l \in \mathcal{C}^{k-l}(\Omega, \mathcal{T}^l(\mathbb{R}^d)) \\ l=1,\dots,k-1,\ u_0 = u,\ u_k = 0}} \sum_{l=1}^k \alpha_{k-l} \|\nabla u_{l-1} - u_l\|_1,$$

where the norm of the full derivative instead of the symmetrized derivative is taken.

Another possibility for modifying (3.6) is to restrict the functions u_l to lth gradient fields of $\mathcal{C}^k(\Omega)$ functions which are clearly in $\mathcal{C}^{k-l}(\Omega, \mathrm{Sym}^l(\mathbb{R}^d))$. Such an approach leads to

$$\mathrm{gradTGV}_\alpha^k(u) = \inf_{\substack{u_l \in \mathcal{C}^k(\Omega),\ l=1,\dots,k-1 \\ u_0 = u,\ u_k = 0}} \sum_{l=1}^k \alpha_{k-l} \|\nabla^l (u_{l-1} - u_l)\|_1.$$

Since u_l are now functions in the same space for all l and the operator ∇^l can be integrated into the functional, this is exactly the k-fold infimal convolution of $\alpha_{k-l}\|\nabla^l \cdot\|_1$ for $l = 1, \dots, k$.

In particular, for $k = 2$, this corresponds to the infimal-convolution of $\|\nabla u\|_1$ and $\|\nabla^2 u\|_1$ as proposed in [CL] as a regularization term for image denoising, which also reads as the dual

functional

$$\sup\left\{\int_\Omega uv_2\,\mathrm{d}x \;\middle|\; v_l\in\mathcal{C}_c^l(\Omega,\mathrm{Sym}^l(\mathbb{R}^d)),\ l=0,1,2,\ v_2=\operatorname{div}v_1=\operatorname{div}^2 v_0,\ \|v_0\|_\infty\le\alpha_0,\ \|v_1\|_\infty\le\alpha_1\right\}.$$

The functional considered in [CEP], in its dual formulation corresponding to

$$\sup\left\{\int_\Omega uv_2\,\mathrm{d}x \;\middle|\; v_2\in\mathcal{C}_c(\Omega),\ v_1\in\mathcal{C}_c^1(\Omega,\mathbb{R}^d),\ v_0\in\mathcal{C}_c^2(\Omega),\ v_2=\operatorname{div}v_1=\Delta v_0,\ \|v_0\|_\infty\le\alpha_0,\ \|v_1\|_\infty\le\alpha_1\right\},$$

also fits into the latter framework, with different differential operators involved.

3.2. Second-order total generalized variation. In order to gain more intuition on how the total generalized variation measures functions, we make some observations for the case $k=2$. Specifically, TGV_α^2 for characteristic functions on some compactly embedded smooth set in arbitrary dimensions is computed. We also examine the one-dimensional case, i.e., some classes of functions on the interval $]0,L[$.

Proposition 3.11. *Let $\emptyset\neq\Omega'\subset\subset\Omega$ have $\mathcal{C}^{1,1}$ boundary, with $\alpha_0,\alpha_1>0$. Then $u=\chi_{\Omega'}$ is of bounded total generalized variation (of order 2) and*

$$\mathrm{TGV}_\alpha^2(u)=\alpha_1\,\mathrm{TV}(u)=\alpha_1\,\mathrm{Per}_{\Omega'}.$$

Proof. First observe that (3.4) immediately gives that

$$\mathrm{TGV}_\alpha^2(u)\le\alpha_1\int_{\partial\Omega'}1\,\mathrm{d}\mathcal{H}^{d-1}=\alpha_1\,\mathrm{TV}(u)=\alpha_1\,\mathrm{Per}_{\Omega'},$$

so we only have to construct a sequence of feasible v^ε such that the right-hand side is attained. Choose ε such that $0<\varepsilon<\operatorname{dist}(\Omega',\partial\Omega)/2$ and denote by $\sigma:\Omega\to\mathbb{R}$ a compactly supported signed distance function associated with Ω', i.e.,

$$\sigma^\varepsilon(x)=(1-2\chi_{\Omega'})\operatorname{dist}\big(x,\partial\Omega'\cup\complement(\partial\Omega'+B_\varepsilon(0))\big),$$

where $B_\varepsilon(0))=\{x:|x|<\varepsilon\}$.

See Figure 1 for an illustration of this construction. It is known [DZ, Theorem 5.4.3] that each σ^ε is continuously differentiable in a neighborhood of $\partial\Omega'$ since Ω' has a $\mathcal{C}^{1,1}$ boundary. Also, each gradient coincides with the outer normal on $\partial\Omega'$, i.e., $\nabla\sigma^\varepsilon(x)=\nu(x)$ for $x\in\partial\Omega'$. Eventually, $\operatorname{supp}\sigma^\varepsilon\subset\subset\Omega$ and $|\nabla\sigma^\varepsilon(x)|\le1$ a.e. in Ω. Choosing a standard mollifier $G\in\mathcal{C}_0^\infty(B_1(0))$ and denoting by G_ε its dilated versions, it is immediate that $v_0^\varepsilon=\alpha_1\sigma^\varepsilon*G_{\varepsilon/4}$ satisfies $v_0^\varepsilon\in\mathcal{C}_0^\infty(\Omega)$.

One can then compute that $v_0^\varepsilon\to0$ uniformly in Ω as well as $\nabla v_0^\varepsilon(x)\to\alpha_1\nu(x)$ for $x\in\partial\Omega'$ as $\varepsilon\to0$. Consequently, choosing ε small enough yields $\|v_0^\varepsilon\|_\infty\le\alpha_0/\sqrt{d}$ and, since

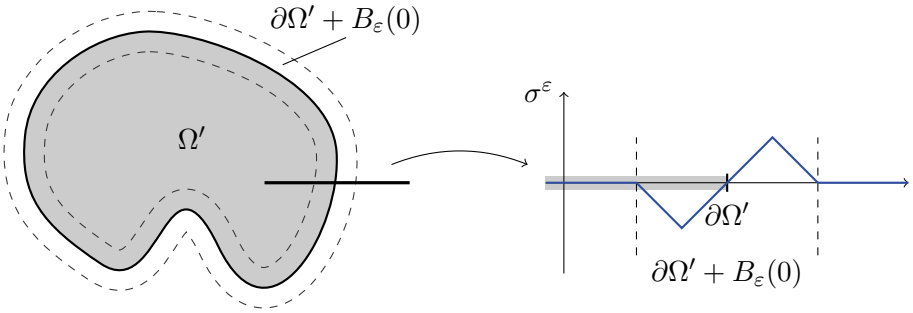


Figure 1. *Illustration of the construction of the signed distance function σ^ε. On the left-hand side, the smooth set Ω', its boundary, and the corresponding ε-tube $\partial\Omega' + B_\varepsilon(0)$ are depicted. On the right-hand side one can see, qualitatively, the values of σ^ε for the indicated section.*

the Lipschitz constant of σ^ε is 1, $\|\nabla v_0^\varepsilon\|_\infty \leq \alpha_1$. Defining the symmetric 2-tensor field (in matrix form) according to

$$v^\varepsilon(x) = v_0^\varepsilon(x) I$$

yields $\operatorname{div} v^\varepsilon(x) = \nabla v_0^\varepsilon$; hence v^ε are valid test functions for (3.1). Using the divergence theorem (2.2) then gives

$$\int_\Omega u \operatorname{div}^2 v^\varepsilon \, \mathrm{d}x = \int_{\partial\Omega'} \operatorname{div} v^\varepsilon \cdot \nu \, \mathrm{d}\mathcal{H}^{d-1}(x) = \int_{\partial\Omega'} \nabla v_0^\varepsilon \cdot \nu \, \mathrm{d}\mathcal{H}^{d-1}(x).$$

As $\varepsilon \to 0$, we have $\nabla v_0^\varepsilon \to \alpha_1 \nu$ on $\partial\Omega'$; it indeed follows that $\mathrm{TGV}_\alpha^2(u) \geq \alpha_1 \mathrm{TV}(u)$, which completes the proof. ∎

The following considerations are concerned with the one-dimensional case.

Example 3.12. Consider the interval $\Omega =]0, L[$ for some $L > 0$ and fix $k = 2$, $\alpha_0, \alpha_1 > 0$. To avoid boundary effects, we assume, moreover, that $\alpha_0/\alpha_1 < L/2$. Let $u :]0, L[\to \mathbb{R}$ be

$$u(x) = \sum_{i=1}^{2} p_i(x) \chi_{\Omega_i}, \qquad p_i(x) = a_i x + b_i \tag{3.7}$$

with $a_1, a_2, b_1, b_2 \in \mathbb{R}$ and $\Omega_1 =]0, c[$, $\Omega_2 =]c, L[$ for some $\alpha_0/\alpha_1 \leq c \leq L - \alpha_0/\alpha_1$.

We compute TGV_α^2 for some $\alpha_0, \alpha_1 > 0$. For this purpose, choose a $v \in \mathcal{C}_c^2(]0, L[)$ and apply integration by parts twice to get

$$\int_0^L uv'' \, \mathrm{d}x = \big(p_2(c) - p_1(c)\big)v'(c) + \big(p_1'(c) - p_2'(c)\big)v(c). \tag{3.8}$$

By Proposition 3.6 we already have

$$\mathrm{TGV}_\alpha^2(u) \leq \alpha_1 |p_2(c) - p_1(c)| + \alpha_0 \big|p_2'(c) - p_1'(c)\big|.$$

Assume that $p_2(c) - p_1(c) \geq 0$ as well as $p_1'(c) - p_2'(c) \geq 0$. Consider, for sufficiently small

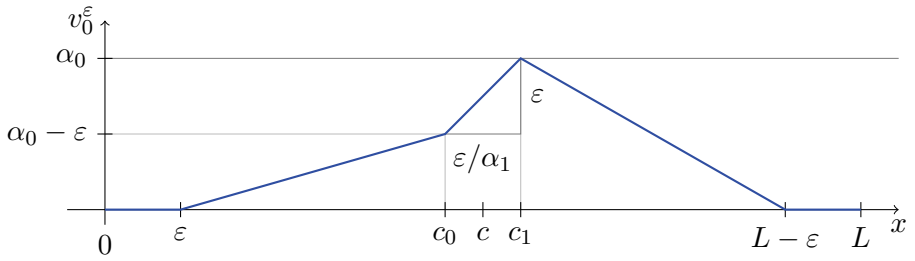


Figure 2. *Schematic illustration of the functions v_0^ε.*

$\varepsilon > 0$, the sequence of functions $\{v_0^\varepsilon\}$ according to

$$v_0^\varepsilon(x) = \begin{cases} 0 & \text{for } 0 \le x \le \varepsilon, \\ (\alpha_0 - \varepsilon)(x-\varepsilon)/(c_0 - \varepsilon) & \text{for } \varepsilon \le x \le c_0, \\ \alpha_0 - \varepsilon + \alpha_1(x - c_0) & \text{for } c_0 \le x \le c_1, \\ \alpha_0(x + \varepsilon - L)/(c_1 + \varepsilon - L) & \text{for } c_1 \le x \le L - \varepsilon, \\ 0 & \text{for } L - \varepsilon \le x \le L, \end{cases} \tag{3.9}$$

where $c_0 = c - \varepsilon/(2\alpha_1)$ and $c_1 = c + \varepsilon/(2\alpha_1)$; see Figure 2 for an illustration.

One can easily convince oneself that $\|v_0^\varepsilon\|_\infty \le \alpha_0$ as well as $\|(v_0^\varepsilon)'\|_\infty \le \alpha_1$, the latter taking the choice of c and the assumption on α_0/α_1 into account. Moreover, $v_0^\varepsilon(c) \to \alpha_0$ as $\varepsilon \to 0$ and $(v_0^\varepsilon)'(c) = \alpha_1$. Choosing a mollifier $G \in \mathcal{C}_0^\infty(]-1,1[)$ and denoting by G_ε its dilated versions, the functions $v^\varepsilon = v_0^\varepsilon * G_{\varepsilon/2}$ satisfy $v^\varepsilon \in \mathcal{C}_c^2(]0,L[)$ with

$$\|v^\varepsilon\|_\infty \le \alpha_0, \quad \|(v^\varepsilon)'\|_\infty \le \alpha_1, \quad v^\varepsilon(c) \to \alpha_0, \quad (v^\varepsilon)'(c) \to \alpha_1 \quad \text{as} \quad \varepsilon \to 0.$$

Thus, plugging the sequence into (3.8) gives that the estimate is sharp; i.e.,

$$\mathrm{TGV}_\alpha^2(u) = \alpha_1 |p_2(c) - p_1(c)| + \alpha_0 \big|p_1'(c) - p_2'(c)\big|.$$

With analogous constructions for the cases where $p_2(c) - p_1(c) \le 0$ or $p_1'(c) - p_2'(c) \le 0$, this follows for all u of the form (3.7). Figure 3 depicts some cases of u and how the values of $\mathrm{TGV}_\alpha^2(u)$ can be expressed.

Let us finally note the role of the restrictions $\alpha_0/\alpha_1 < L/2$ and $\alpha_0/\alpha_1 < c < L - \alpha_0/\alpha_1$. An arbitrary test function $v \in \mathcal{C}_c^2(]0,L[)$ with $\|v\|_\infty \le \alpha_0$ and $\|v'\|_\infty \le \alpha_1$ necessarily satisfies

$$|v(x)| \le v_{\max}(x) = \begin{cases} \alpha_1 x & \text{for } 0 \le x \le \alpha_0/\alpha_1, \\ \alpha_0 & \text{for } \alpha_0/\alpha_1 \le x \le L - \alpha_0/\alpha_1, \\ \alpha_1 \frac{L-x}{\alpha_0/\alpha_1 - L} & \text{for } L - \alpha_0/\alpha_1 \le x \le L. \end{cases}$$

In the general case for α_0, α_1, and c, we get a similar upper bound for $\mathrm{TGV}_\alpha^2(u)$ with $v_{\max}(c)$ instead of α_0 for which the sharpness can be obtained by a construction which is analogous to the above. This yields the more general identity

$$\mathrm{TGV}_\alpha^2(u) = \alpha_1 |p_2(c) - p_1(c)| + v_{\max}(c) \big|p_2'(c) - p_1'(c)\big|.$$

Consequently, the above assumptions therefore exactly reflect the case where $v_{\max}(c) = \alpha_0$.

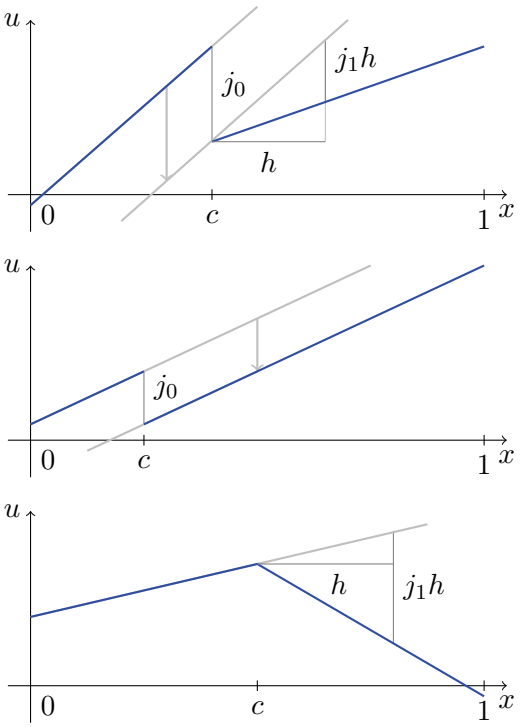


Figure 3. *Illustration of some piecewise affine functions and the corresponding values of* TGV^2_α.

4. Numerical methods. In this section we present numerical methods in order to solve total generalized variation based regularization models. In doing so, we will mainly concentrate on a TGV^2_α regularization functional with a quadratic L^2 data fidelity term. We have two reasons for this. First, the TGV^2_α-term is just simple enough to give a compact description of the numerics. On the other hand, it is general enough to enable the reader to apply the numerics to TGV models with a higher order, e.g., TGV^3_α.

The TGV^2_α-L^2 image denoising model is given by

$$\min_{u \in L^2(\Omega)} \mathrm{TGV}^2_\alpha(u) + \frac{\|u - f\|_2^2}{2} \tag{4.1}$$

for some positive $\alpha = (\alpha_0, \alpha_1)$. The solutions of this nonsmooth minimization problem can be obtained by solving the Fenchel predual problem, an approach which has recently become popular in the image processing literature [Ca, CGM, Ch, HK]. As usual, the predual problem can be rewritten as a projection problem, for instance, as

$$\min_{v \in H^2_{\mathrm{div},c}(\Omega, S^{d\times d})} \frac{\|f - \mathrm{div}^2 v\|_2^2}{2} + I_{\{\|v\|_\infty \le \alpha_0\}}(v) + I_{\{\|\mathrm{div}\, v\|_\infty \le \alpha_1\}}(v). \tag{4.2}$$

Solutions u^* and v^* of (4.1) and (4.2), respectively, satisfy $u^* = f - \mathrm{div}^2 v^*$, so one can solve the predual problem in order to obtain the unique solution of the denoising problem. For the minimization of (4.2) we will adapt the accelerated first-order method of Nesterov [Ne].

The subsequent section then shows numerical experiments we carried out using this algorithm. As expected, the proposed TGV^2_α-L^2 image denoising model is able to restore piecewise affine functions, and in contrast to the usual total variation denoising model [ROF] (TGV^1_α-L^2 in our terms), it does not exhibit the staircasing effect. Furthermore, we will show that replacing TGV^2_α in (4.1) by a TGV^3_α regularization restores piecewise smooth images and leads to further improvements.

4.1. Discrete setting. In order to implement (4.2) on a digital computer we need to introduce the discrete setting. For clarity of presentation we consider only the case $d = 2$, i.e., the case of two-dimensional images. Our approach will be based on finite-difference discretizations which are commonly used in image processing. We will utilize a two-dimensional regular Cartesian grid of size $M \times N$:

$$\Omega^h = \big\{(ih, jh) \;\big|\; (0,0) \le (i,j) < (M,N)\big\},$$

where h denotes the grid width and the pairs $(i,j) \in \mathbb{N}^2$ are the indices of the discrete locations (ih, jh) on the grid. In what follows we will denote the discrete quantities by the superscript h. Denote by U^h the Euclidean space $\mathbb{R}^{MN}$, by V^h the Euclidean space $\mathbb{R}^{3MN}$, and $W^h = \mathbb{R}^{2MN}$ equipped with the scalar products

$$\begin{aligned}
u^h, p^h \in U^h : \langle u^h,\, p^h\rangle &= \sum_{i,j} u^h_{i,j} p^h_{i,j},\\
v^h, q^h \in V^h : \langle v^h,\, w^h\rangle &= \sum_{i,j} (v^h_1)_{i,j}(q^h_1)_{i,j} + (v^h_2)_{i,j}(q^h_2)_{i,j}\\
&\quad + 2(v^h_3)_{i,j}(q^h_3)_{i,j},\\
w^h, \eta^h \in W^h : \langle w^h,\, \eta^h\rangle &= \sum_{i,j} (w^h_1)_{i,j}(\eta^h_1)_{i,j} + (w^h_2)_{i,j}(\eta^h_2)_{i,j},
\end{aligned}$$

respectively. Further, let $u^h \in U^h$ be the finite-dimensional approximation of the unknown function u in (4.1) and let $v^h = (v^h_1, v^h_2, v^h_3) \in V^h$ be the finite-dimensional approximation of the symmetric matrix field v in (4.2). Note that since we are dealing with symmetric matrices we need only store the entries of the upper triangle matrix. Here, v^h_1, v^h_2 stand for the diagonal entries, while v^h_3 models the off-diagonal entry (which is also reflected by the scalar product). The discrete analogue of (4.2) is given by

$$\min_{v^h \in K^h} \left\{ E(v^h) = \frac{\|f^h - (\mathrm{div}^h)^2 v^h\|^2}{2} \right\}, \tag{4.3}$$

where $\|\cdot\|$ denotes the standard L^2 vector norm, f^h is the discretized input image, and the convex set K^h is given by

$$K^h = \big\{v^h \in V^h \;\big|\; \|v^h\|_\infty \le \alpha_0,\ \|\mathrm{div}^h\, v^h\|_\infty \le \alpha_1\big\}.$$

The Fenchel dual problem is, of course, a discrete version of (4.1), whose solution we take as the discrete TGV^2_α-L^2-denoised version of f^h. It can be equivalently formulated as

$$\min_{u^h \in U^h} \left\{ \frac{\|f^h - u^h\|^2}{2} + \sup_{v^h \in K^h} \big\langle (\mathrm{div}^h)^2 v^h,\, u^h \big\rangle \right\}. \tag{4.4}$$

The discrete ∞-norms on V^h and W^h can be expressed by

$$v^h \in V^h : \|v^h\|_\infty = \max_{i,j} \left((v_1^h)_{i,j}^2 + (v_2^h)_{i,j}^2 + 2(v_3^h)_{i,j}^2 \right)^{1/2},$$
$$w^h \in W^h : \|w^h\|_\infty = \max_{i,j} \left((w_1^h)_{i,j}^2 + (w_2^h)_{i,j}^2 \right)^{1/2}.$$

Moreover, for the discretization of the divergence operator we use a recursive application of forward and backward differences in such a way that the outermost divergence operator is based on backward differences with homogeneous Dirichlet boundary conditions. Hence, the discrete versions of the first- and second-order divergence operators are given by

$$\operatorname{div}^h : V^h \to W^h, \qquad (\operatorname{div}^h v^h)_{i,j} = \begin{pmatrix} (\delta_{x+}^h v_1^h)_{i,j} + (\delta_{y+}^h v_3^h)_{i,j} \\ (\delta_{x+}^h v_3^h)_{i,j} + (\delta_{y+}^h v_2^h)_{i,j} \end{pmatrix},$$

and $(\operatorname{div}^h)^2 : V^h \to U^h$ with

$$\begin{aligned} \left((\operatorname{div}^h)^2 v^h\right)_{i,j} &= \left(\operatorname{div}^h(\operatorname{div}^h v^h)\right)_{i,j} \\ &= (\delta_{x-}^h \delta_{x+}^h v_1^h)_{i,j} + (\delta_{y-}^h \delta_{y+}^h v_2^h)_{i,j} + \left((\delta_{y-}^h \delta_{x+}^h + \delta_{x-}^h \delta_{y+}^h) v_3^h\right)_{i,j}, \end{aligned}$$

where the forward and backward differences are defined as

$$(\delta_{x+}^h v^h)_{i,j} = \begin{cases} (v_{i+1,j}^h - v_{i,j}^h)/h & \text{if} \quad 0 \le i < M-1, \\ 0 & \text{if} \quad i = M-1, \end{cases}$$
$$(\delta_{y+}^h v^h)_{i,j} = \begin{cases} (v_{i,j+1}^h - v_{i,j}^h)/h & \text{if} \quad 0 \le j < N-1, \\ 0 & \text{if} \quad j = N-1 \end{cases}$$

and

$$(\delta_{x-}^h v^h)_{i,j} = \begin{cases} -v_{i-1,j}^h/h & \text{if} \quad i = M-1, \\ (v_{i,j}^h - v_{i-1,j}^h)/h & \text{if} \quad 0 < i < M-1, \\ v_{i,j}^h/h & \text{if} \quad i = 0, \end{cases}$$
$$(\delta_{y-}^h v^h)_{i,j} = \begin{cases} -v_{i,j-1}^h/h & \text{if} \quad j = N-1, \\ (v_{i,j}^h - v_{i,j-1}^h)/h & \text{if} \quad 0 < j < N-1, \\ v_{i,j}^h/h & \text{if} \quad j = 0. \end{cases}$$

Furthermore, we need to introduce the discrete version of the symmetrized second-order derivative operator $(\mathcal{E}^h)^2$. We choose it in such a way that it is adjoint to the discrete divergence operator; that is, $(\mathcal{E}^h)^2 = ((\operatorname{div}^h)^2)^*$. By computing the adjoint of $(\operatorname{div}^h)^2$ and taking into account the symmetry of v^h, we arrive at

$$\left((\mathcal{E}^h)^2 u^h\right)_{i,j} = \begin{pmatrix} (\delta_{x-}^h \delta_{x+}^h u^h)_{i,j} & \dfrac{\left((\delta_{y-}^h \delta_{x+}^h + \delta_{x-}^h \delta_{y+}^h) u^h\right)_{i,j}}{2} \\ \dfrac{\left((\delta_{y-}^h \delta_{x+}^h + \delta_{x-}^h \delta_{y+}^h) u^h\right)_{i,j}}{2} & (\delta_{y-}^h \delta_{y+}^h u^h)_{i,j} \end{pmatrix}.$$

Note that, by our choosing that the outmost divergence is based on backward differences with homogeneous Dirichlet boundary conditions, the innermost derivative of the symmetrized derivative operator is now based on forward differences with Neumann boundary conditions. For example, for TGV^3_α, the $(\mathrm{div}^h)^3 v^h$ operator is computed as $\delta^h_{x-}\delta^h_{x+}\delta^h_{x-}v^h_1 + \cdots$. This corresponds to a natural replication of the image data, which is a common choice in image processing.

4.2. A first-order minimization algorithm. Problem (4.3) poses a quadratic optimization problem with pointwise quadratic constraints. Hence, many algorithms can be used to compute the solution [A]. Here we employ the accelerated first-order method of Nesterov [Ne, BT], which shares a convergence rate of $O(1/k^2)$ in terms of the functional values. This means that one needs $O(1/\sqrt{\varepsilon})$ iterations to compute an approximate solution which minimizes the functional up to accuracy ε. Nesterov's method is easy to implement and can be efficiently parallelized, e.g., on recent graphics processing units. The outline of Nesterov's method applied to (4.3) is as follows: We choose $v^h_0 = 0$, $\bar{v}^h_0 = 0$, and $t_0 = 1$. Then, for $k \geq 0$ we let

$$\begin{cases} v^h_{k+1} = \Pi_{K^h}\left(\bar{v}^h_k + \tau\left((\mathcal{E}^h)^2\left(f^h - (\mathrm{div}^h)^2\bar{v}^h_k\right)\right)\right), \\ t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}, \\ \bar{v}^h_{k+1} = v^h_{k+1} + \left(\frac{t_k - 1}{t_{k+1}}\right)\left(v^h_{k+1} - v^h_k\right), \end{cases} \tag{4.5}$$

where $\tau > 0$ is some prescribed step-size and Π_{K^h} denotes the Euclidean projector onto the convex set K^h.

Proposition 4.1. *Let $\tau = \frac{1}{L^2}$, where $L^2 = \frac{64}{h^4} \geq \|(\mathrm{div}^h)^2\|^2$. Then, the sequence $\{v^h_k\}$ generated by algorithm (4.5) is such that for any $k \geq 0$*

$$0 \leq E(v^h_k) - E\big((v^h)^*\big) \leq \frac{2L^2\|v^h_0 - (v^h)^*\|^2}{(k+1)^2}, \tag{4.6}$$

with $(v^h)^* \in V^h$ being a solution of (4.3). Moreover, $u^h_k = f^h - \mathrm{div}^2 v^h_k \to (u^h)^*$, where $(u^h)^*$ is the solution of (4.4).

Proof. We begin with estimating the Lipschitz constant of $v^h \mapsto (\mathcal{E}^h)^2(f^h - (\mathrm{div}^h)^2 v^h)$ with respect to the associated norms in V^h and U^h. This amounts to estimating the operator norm defined as

$$L^2 = \|(\mathrm{div}^h)^2\|^2 = \sup_{v^h \in V^h, v^h \neq 0} \frac{\|(\mathrm{div}^h)^2 v^h\|^2}{\|v^h\|^2}. \tag{4.7}$$

In order to write $(\mathrm{div}^h)^2$ in terms of finite-difference schemes, we agree to set

$$\begin{aligned} (v^h_1)_{-1,j} &= (v^h_1)_{0,j}, & (v^h_1)_{M,j} &= (v^h_1)_{M-1,j}, \\ (v^h_2)_{i,-1} &= (v^h_2)_{i,0}, & (v^h_2)_{i,N} &= (v^h_2)_{i,N-1}, \\ (v^h_3)_{i,-1} &= (v^h_3)_{-1,j} = 0, & (v^h_3)_{i,N} &= (v^h_3)_{M,j} = 0. \end{aligned}$$

Moreover, one can see that $(\mathrm{div}^h)^2 v^h$ does not depend on the values $(v_3^h)_{M-1,i}$ and $(v_3^h)_{j,N-1}$, so we assume them to be zero in what follows. Then, $(\mathrm{div}^h)^2$ amounts to the application of the following finite difference scheme:

$$(\mathrm{div}^h)^2 v^h = \frac{1}{h^2}\underbrace{\begin{bmatrix}1 & -2 & 1\end{bmatrix}}_{D_1} v_1^h + \frac{1}{h^2}\underbrace{\begin{bmatrix}1\\ -2\\ 1\end{bmatrix}}_{D_2} v_2^h + \frac{1}{h^2}\underbrace{\begin{bmatrix}0 & 1 & -1\\ 1 & -2 & 1\\ -1 & 1 & 0\end{bmatrix}}_{D_3} v_3^h.$$

Let us estimate, by multiple use of $(a+b)^2 \le 2a^2 + 2b^2$,

$$\|D_1 v_1^h\|_2^2 \le \sum_{i,j} 2\big((2v_1^h)_{i,j}\big)^2 + 4\big((v_1^h)_{i-1,j}\big)^2 + 4\big((v_1^h)_{i+1,j}\big)^2 \le 16\|v_1^h\|_2^2$$

and, analogously, $\|D_2 v_2^h\|_2^2 \le 16\|v_2^h\|_2^2$. For the third term, consider

$$\begin{aligned}\|D_3 v_3^h\|_2^2 &\le 2\left\|\begin{bmatrix}0 & 1 & -1\\ 0 & -1 & 1\\ 0 & 0 & 0\end{bmatrix} v_3^h\right\|_2^2 + 2\left\|\begin{bmatrix}0 & 0 & 0\\ 1 & -1 & 0\\ -1 & 1 & 0\end{bmatrix} v_3^h\right\|_2^2\\ &\le 8\sum_{i,j} 2\big((v_3^h)_{i,j}\big)^2 + \big((v_3^h)_{i-1,j}\big)^2 + \big((v_3^h)_{i+1,j}\big)^2 + \big((v_3^h)_{i,j-1}\big)^2\\ &\quad + \big((v_3^h)_{i,j+1}\big)^2 + \big((v_3^h)_{i-1,j+1}\big)^2 + \big((v_3^h)_{i+1,j-1}\big)^2\\ &\le 64\|v_3^h\|_2^2.\end{aligned}$$

Together, we find

$$\begin{aligned}\|(\mathrm{div}^h)^2 v^h\|^2 &\le \frac{1}{h^4}\Big(4\|D_1 v_1^h\|_2^2 + 4\|D_2 v_2^h\|_2^2 + 2\|D_3 v_3^h\|_2^2\Big)\\ &\le \frac{64}{h^4}\Big(\|v_1^h\|_2^2 + \|v_2^h\|_2^2 + 2\|v_3^h\|_2^2\Big) = \frac{64}{h^4}\|v^h\|^2.\end{aligned}$$

Then, by substitution into (4.7) we get $L^2 \le \frac{64}{h^4}$. The proof of convergence for the functional values and the efficiency estimate of algorithm (4.5) are both presented in [BT, Theorem 4.1].

Finally, note that since K^h is bounded, each subsequence of $\{v_k^h\}$ has a convergent subsequence $\{v_{k_l}^h\}$ with some limit $(v^h)^*$. As the discrete divergence operator is continuous, we know, moreover, that the corresponding sequence $\{u_{k_l}^h\}$ converges to some $(u^h)^* = f^h - (\mathrm{div}^h)^2 (v^h)^*$. By the estimate (4.6), each subsequence is a minimizing sequence, and, hence, $(v^h)^*$ is a solution of (4.3). Consequently, as the solutions of (4.3) and (4.4) are in duality, $(u^h)^*$ is a solution of (4.4). Since the latter has to be unique by strict convexity, we deduce from the usual subsequence argument that the whole sequence satisfies $u_k^h \to (u^h)^*$. ∎

4.2.1. Computing the projection. The most costly part in (4.5) is the computation of the Euclidean projection of the dual variable onto the convex set K^h. Basically, the projection of a variable $\hat{v}^h$ is given by the minimizer of

$$\Pi_{K^h}(\hat{v}^h) = \operatorname*{argmin}_{v^h \in K^h} \frac{\|v^h - \hat{v}^h\|^2}{2}. \tag{4.8}$$

Since the convex set K^h involves inequality constraints on both v^h and $\operatorname{div}^h v^h$, we may also write

$$\Pi_{K^h}(\hat{v}^h) = \operatorname*{argmin}_{v^h \in V^h} \frac{\|v^h - \hat{v}^h\|^2}{2} + I_{\hat{K}^h}(\Lambda v^h)$$

with $\Lambda : V^h \to V^h \times W^h$ given by $\Lambda v^h = (v^h, \operatorname{div}^h v^h)$ and

$$\hat{K}^h = \big\{(v^h, w^h) \in V^h \times W^h \;\big|\; \|v^h\|_\infty \le \alpha_0 \;,\; \|w^h\|_\infty \le \alpha_1\big\}.$$

Applying Fenchel duality to this minimization problem yields the equivalent problem

$$\min_{(q^h,\eta^h)\in V^h\times W^h} \frac{\|q^h - \mathcal{E}^h(\eta^h) - \hat{v}^h\|^2}{2} + \alpha_0\|q^h\|_1 + \alpha_1\|\eta^h\|_1 \;, \tag{4.9}$$

where we used $\Lambda^*(q^h, \eta^h) = q^h - \mathcal{E}^h(\eta^h)$ and employed the following choices for the L^1-norms:

$$q^h \in V^h : \|q^h\|_1 = \sum_{i,j} \Big((q_1^h)_{i,j}^2 + (q_2^h)_{i,j}^2 + 2(q_3^h)_{i,j}^2 \Big)^{1/2},$$

$$\eta^h \in W^h : \|\eta^h\|_1 = \sum_{i,j} \Big((\eta_1^h)_{i,j}^2 + (\eta_2^h)_{i,j}^2 \Big)^{1/2}.$$

The projection (4.8) can hence be computed by obtaining a solution pair $((q^h)^*, (\eta^h)^*)$ of (4.9) and setting

$$\Pi_{K^h}(\hat{v}^h) = \hat{v}^h - \Lambda^*\big((q^h)^*, (\eta^h)^*\big) = \hat{v}^h - (q^h)^* + \mathcal{E}^h\big((\eta^h)^*\big).$$

For minimization of (4.9) we adopt the fast shrinkage thresholding algorithm (FISTA) recently proposed by Beck and Teboulle in [BT] as a variant of Nesterov's method [Ne]. We exploit the fact that minimization with respect to q^h is straightforward to compute using shrinkage operations. The outline of the algorithm is as follows: We choose $\eta_0^h, \bar{\eta}_0^h \in W^h$ and $t_0 = 1$. Then for each $k \ge 0$ we let

$$\begin{cases} q_{k+1}^h = \; \mathcal{S}_{\alpha_0}\left(\hat{v}^h + \mathcal{E}^h(\bar{\eta}_k^h)\right), \\ \eta_{k+1}^h = \; \mathcal{S}_{\sigma\alpha_1}\left(\bar{\eta}_k^h + \sigma \operatorname{div}^h\left(\hat{v}^h + \mathcal{E}^h(\bar{\eta}_k^h) - q_{k+1}^h\right)\right), \\ t_{k+1} = \; \frac{1+\sqrt{1+4t_k^2}}{2}, \\ \bar{\eta}_{k+1}^h = \; \eta_{k+1}^h + \left(\frac{t_k - 1}{t_{k+1}}\right)\left(\eta_{k+1}^h - \eta_k^h\right), \end{cases} \tag{4.10}$$

where $\sigma = \frac{h^2}{8} \le \|\operatorname{div}^h\|^{-2}$ denotes the step-width and $\mathcal{S}_\lambda(t)$ denotes the generalized shrinkage formula, which is given by $\mathcal{S}_\lambda(t)_{i,j} = (|t_{i,j}| - \lambda)^+ \frac{t_{i,j}}{|t_{i,j}|}$ with the respective absolute values. For each projection, we run the iterative projection algorithm until the maximal feasibility error of $\Pi_{K^h}(\hat{v}^h)$ is below a threshold ε_p.

5. Experimental results. In the following, we present numerical results of our total generalized variation models. We start by studying the efficiency of our first-order minimization algorithm. Then we present experimental results of synthetic and real images. It turns out that our second-order model (TGV^2_α-L^2) consistently outperforms both the standard TV model of [ROF] (which is equivalent to the TGV^1_α-L^2 model) and the Inf-Conv model of [CL]. Finally we show that a third-order model (TGV^3_α-L^2) can further improve the results in cases where the image intensity function cannot be well approximated by piecewise affine functions.

5.1. Efficiency of the first-order algorithm. In our first experiment, we compare the theoretical efficiency estimate of Nesterov's method for the functional values with the practical implementation using a simple synthetic input image. As already stated in section 4, Nesterov's first-order method allows us to compute a bound on the accuracy ε of the functional values for a given number of iterations. According to Proposition 4.1, the bound is given by

$$\varepsilon = E(v_k^h) - E((v^h)^*) \leq \frac{2L^2\|v_0^h - (v^h)^*\|^2}{(k+1)^2}, \tag{5.1}$$

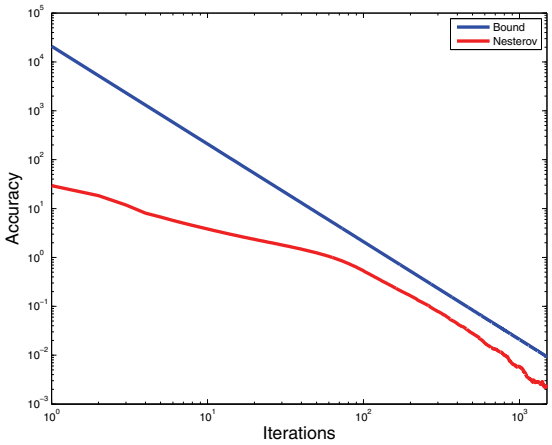
where k is the number of iterations. Hence, in order to compute a bound on ε, it remains to estimate the quantity $\|v_0^h - (v^h)^*\|^2$. Since we have that $v_0^h = 0$ and $\|v^h\|_\infty \leq \alpha_0$, we simply deduce that

$$\varepsilon \leq \frac{128\alpha_0^2 MN}{h^4(k+1)^2}. \tag{5.2}$$

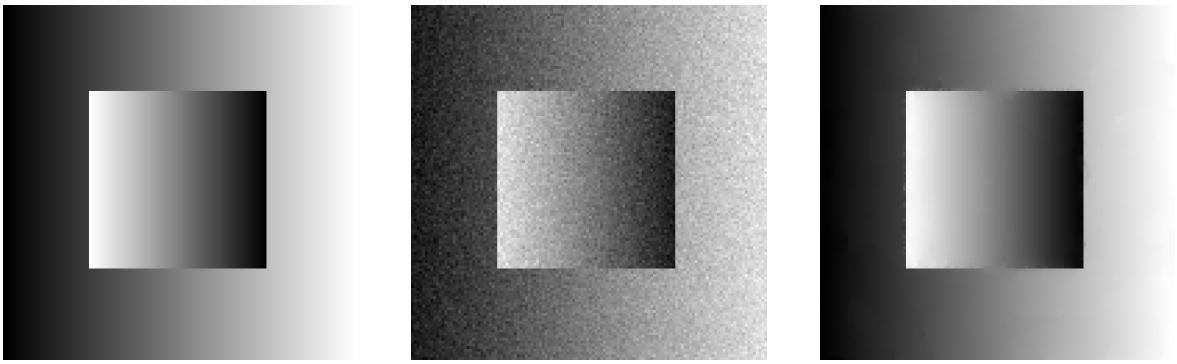
Figure 4(d) shows the denoising result of TGV_α^2-L^2 applied to the noisy input image shown in Figure 4(c). We set $(\alpha_0, \alpha_1) = (0.1, 0.05)$, $h = 1$, $\varepsilon_p = 10^{-4}$ and run algorithm (4.5) for $k = 1500$ iterations. This results in a theoretical accuracy of $\varepsilon \approx 10^{-2}$. Note that the proposed method almost perfectly reconstructs the piecewise affine input image. Figure 4(a) shows the accuracy $\varepsilon = E(v_k^h) - E((v^h)^*)$ of Nesterov's first-order minimization algorithm. The true minimal function value $E((v^h)^*)$ was determined by running the algorithm for a very large number of iterations. One can see that the theoretical bound on the accuracy is clearly outperformed in practice but shows a similar asymptotic behavior.

Finally, let us comment on the running times for TGV models of different order. Clearly, the running time increases with increasing order of the models. This is mainly due to the fact that computing the projection onto the set K^h is much more complicated for TGV^k models of order $k > 1$. In our nonoptimized MATLAB implementation, it turns out that the relative running times between TGV^1, TGV^2, and TGV^3 regularization behave roughly like $1 : 10 : 35$. However, as mentioned above, Nesterov's method can be easily parallelized. Therefore, we have also implemented a parallelized version of Nesterov's method for the TGV^2-L^2 model. On a Nvidia Tesla C1060 graphics processing unit (GPU), it takes only 1.4 seconds to perform 1500 iterations for the test image shown in Figure 4(a) (which ensures the accuracy $\varepsilon \approx 10^{-2}$ in terms of functional values). This includes all data transfers to the GPU and back.

5.2. Synthetic images. In our second experiment we evaluate the performance of the TGV_α^2-L^2 model using a piecewise affine test image. We compare our model to the standard Rudin–Osher–Fatemi (ROF) model and the Inf-Conv model (see Table 1). The parameters of each model were optimized to achieve the best reconstruction with respect to the root mean squared error (RMSE). Furthermore, for all regularizations, the number of iterations has been chosen according to (5.2) or an analogous estimate to ensure a comparable accuracy in terms of the functional values. Figure 5 shows the input images and the reconstructed images. For better visualization we additionally provide three-dimensional renderings of the upper left region. One can clearly see that the ROF model performs worst since it exhibits the well-known staircasing effect. The Inf-Conv model performs better but



(a)



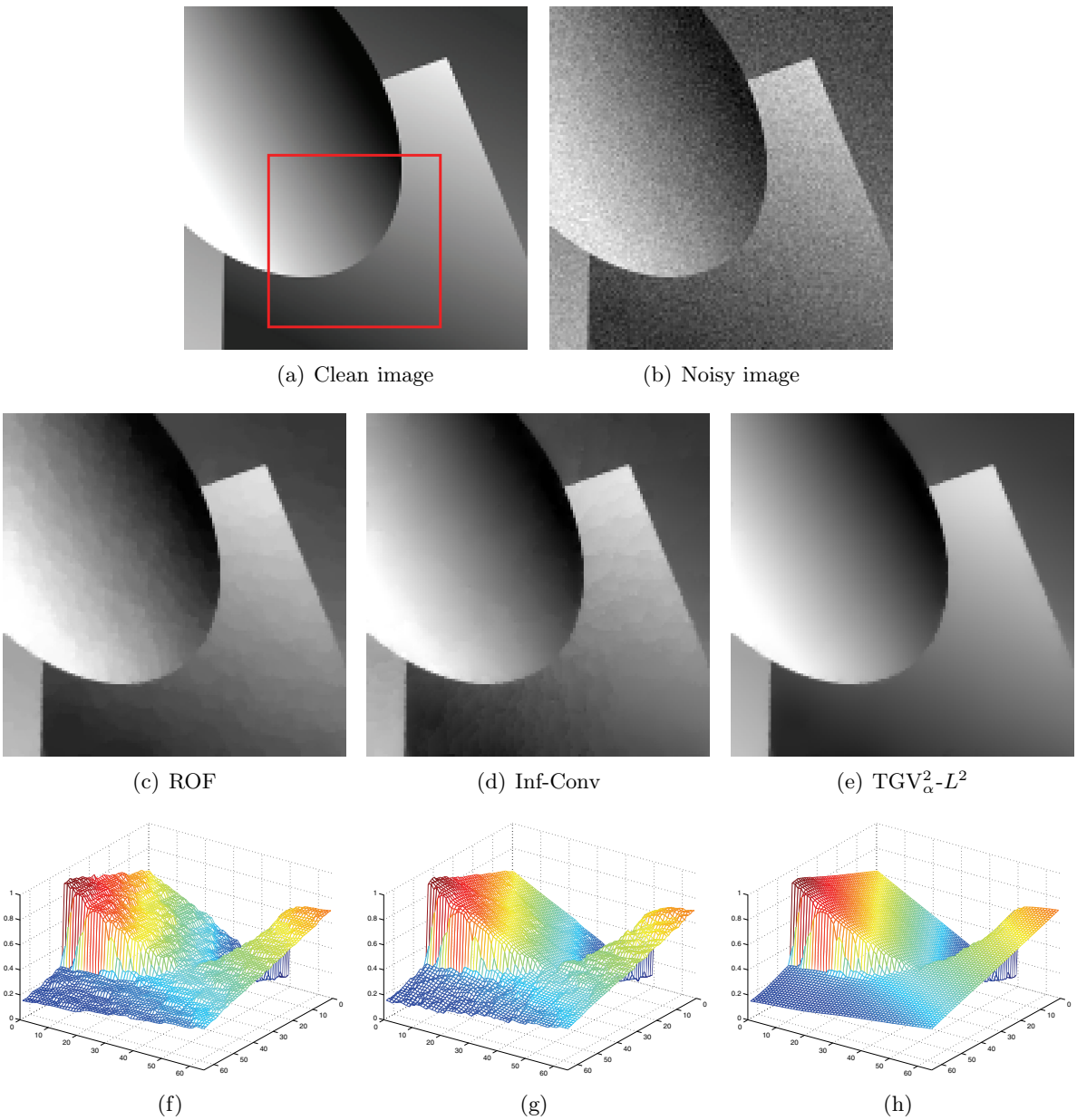
(b) (c) (d)

Figure 4. *Convergence of our first-order minimization algorithm applied to the restoration of a piecewise affine image.* (a) *Theoretical bound of Nesterov's method versus the empirical observation. The theoretical bound provides a useful estimate sharing the same asymptotic behavior, but is still outperformed in practice.* (b) *Input image of size* 128×128. (c) *Degraded image containing zero mean Gaussian noise with standard deviation* $\sigma = 0.05$. (d) *Reconstructed image using the proposed* TGV^2_α-L^2 *model. Note that the proposed method almost perfectly reconstructs the input image.*

Table 1

Quantitative evaluation of different image regularization models applied to piecewise affine and piecewise smooth image reconstruction (see Figures 5 and 6, respectively). The numbers represent the root mean squared error (RMSE) of the reconstructed images with respect to the clean input images.

	Test image	
Model	Piecewise affine	Piecewise smooth
ROF	0.0147	0.0176
Inf-Conv	0.0131	0.0125
TGV^2_α-L^2	**0.0107**	0.0080
TGV^3_α-L^2	—	**0.0074**



(a) Clean image (b) Noisy image

(c) ROF (d) Inf-Conv (e) TGV^2_α-L^2

(f) (g) (h)

Figure 5. *Reconstruction of a piecewise affine image using different image regularization models.* (a) *and* (b) *show the* 128×128 *input image with a marked close-up region and the noisy version containing zero mean Gaussian noise with standard deviation* $\sigma = 0.05$. (c)–(e) *show the results of ROF, Inf-Conv, and* TGV^2_α-L^2 *image regularization, and* (f)–(h) *are the respective three-dimensional close-ups. Note that* TGV^2_α-L^2 *is the only model which is able to reconstruct the piecewise affine structure of the image.*

also has some staircasing near discontinuities. The proposed TGV^2_α-L^2 model performs best. It leads to a natural piecewise affine approximation of the noisy data with sharp discontinuities in between. The quantitative evaluation of the reconstruction emphasizes the visual impression of the results. See Table 1 for the exact RMSE values. We point out that in [SS], a finite-difference approximation of the regularizer

$$\mathcal{R}(u) = \sup \left\{ \int_\Omega u \left(\frac{\partial^2 v_1}{\partial^2 x_1} + \frac{\partial^2 v_2}{\partial^2 x_2} \right) \mathrm{d}x \;\middle|\; v \in \mathcal{C}^2_{\mathrm{c}}(\Omega, \mathbb{R}^2),\ \|v\|_\infty \le \alpha_0, \left\| \left(\frac{\partial v_1}{\partial x_1}, \frac{\partial v_2}{\partial x_2} \right) \right\|_\infty \le \alpha_1 \right\}$$

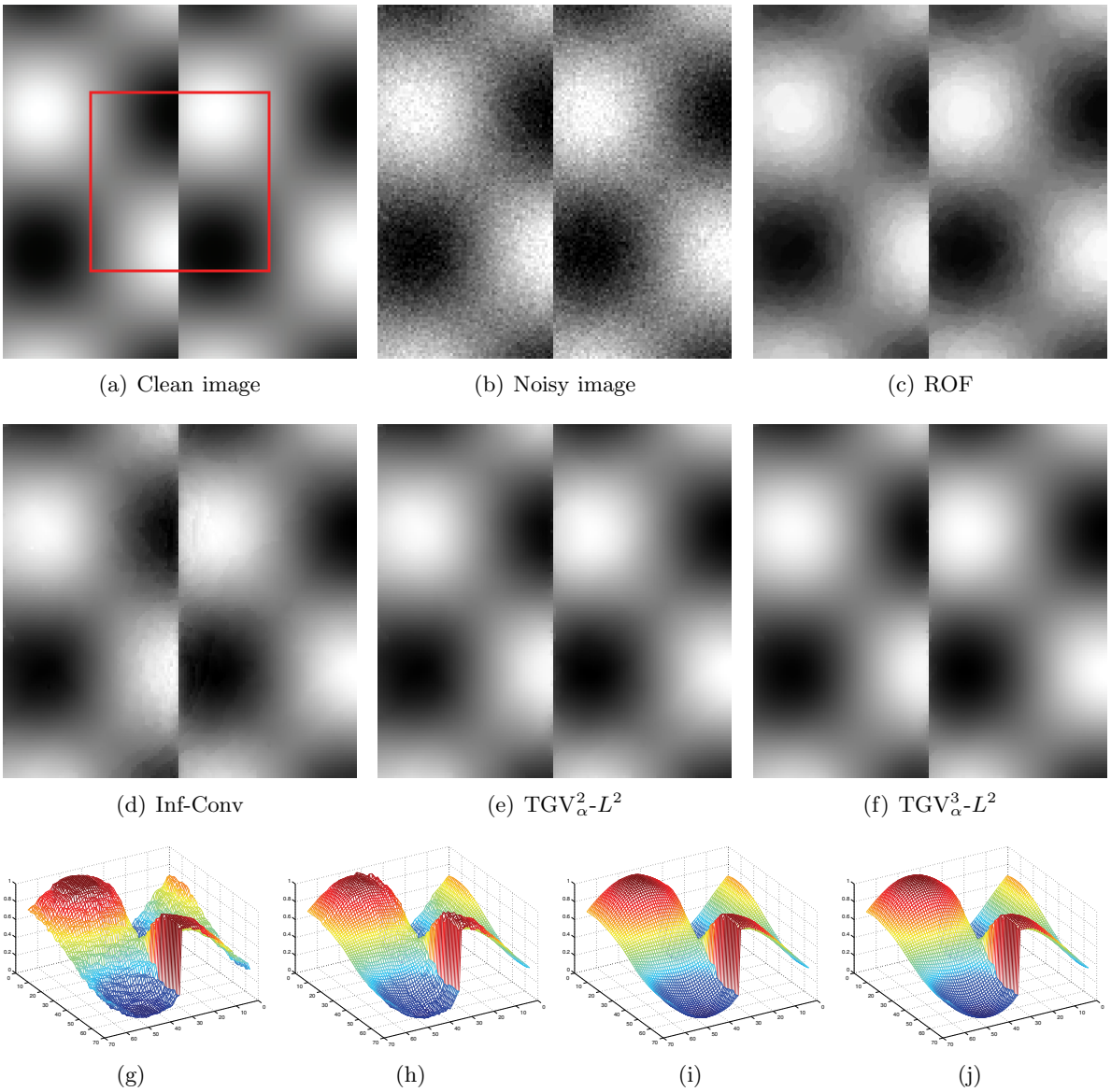
is utilized, leading to similar numerical results. In particular, a suppression of the staircasing effect is also observed.

In our third experiment, we apply total generalized variation regularization up to order 3 for the reconstruction of a piecewise smooth test image. Again, we compare our models to the ROF and Inf-Conv models and, as in the previous experiment, all parameters were optimized in order to meet the lowest RMSE values. Figure 6 shows the input images and the reconstructed images. One can see that the ROF model does not capture well the smooth parts of the image. The Inf-Conv model performs better in the smooth regions but exhibits some staircasing near discontinuities. The proposed TGV^2_α-L^2 and TGV^3_α-L^2 models perform significantly better. Qualitatively, both models perform equally well, but the quantitative evaluation shows that the third-order model has a slightly lower RMSE value (see Table 1). The reason is that the piecewise smooth image shown in Figure 6 contains curvilinear functions. While the second-order model tries to approximate the image based on affine functions, the third-order model additionally allows for quadratic functions, which is clearly better in this case.

5.3. Natural images. Finally, we apply our total generalized variation models to denoising of natural images. Figure 7 shows a noisy image of a penguin in front of a blurry background. While the textured parts (e.g., the rock) of the image are equally well reconstructed by all three models, the total generalized variation models perform significantly better in smooth regions (e.g., the penguin or the blurry background). Furthermore, the close-ups make clear the characteristics of the three models. ROF denoising leads to a piecewise constant, TGV^2_α-L^2 denoising leads to a piecewise affine, and TGV^3_α-L^2 denoising leads to a piecewise quadratic approximation of the image function (see also the close-ups in Figure 8).

Figure 9 shows the denoising capabilities of the proposed models in the case of severe noise. While ROF and Inf-Conv denoising leads to very blocky results, the proposed total generalized variation models lead to significantly better results. The close-ups show that in regions of high curvature, the third-order model leads to further changes compared to the second-order model. The shape of the edges is less blocky at the expense of possibly more regularized transitions (see also the close-ups in Figure 10).

6. Conclusions. We have proposed a structurally transparent regularization functional which involves, in a weak sense, derivatives of order up to k of the desired object. It has the convenient properties of convexity and weak lower semicontinuity. Due to the fact that k is arbitrary, the framework allows us to adjust to a priori known regularity properties. A particularity of the notion of total generalized variation relates to the fact that the test functions are restricted to symmetric tensor fields. Such a restriction is natural in view of the



(a) Clean image (b) Noisy image (c) ROF

(d) Inf-Conv (e) TGV^2_α-L^2 (f) TGV^3_α-L^2

(g) (h) (i) (j)

Figure 6. *Reconstruction of a piecewise smooth image using different image regularization models.* (a) *and* (b) *show the* 128×128 *input image with a marked closeup region and the noisy version containing zero mean Gaussian noise with standard deviation* $\sigma = 0.05$. (c)–(f) *show the results of ROF, Inf-Conv,* TGV^2_α-L^2, *and* TGV^3_α-L^2 *image regularization, and* (g)–(j) *are the respective three-dimensional close-ups of the region containing the discontinuity. Note that the ROF model exhibits strong staircasing and the Inf-Conv model exhibits some staircasing near the discontinuities as well. The* TGV^2_α-L^2 *and* TGV^3_α-L^2 *models are both able to faithfully restore the image.*

(a) Clean Image

(b) Noisy Image

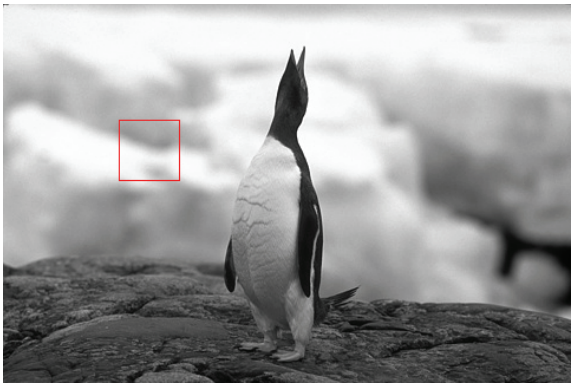
(c) ROF

(d) Inf-Conv

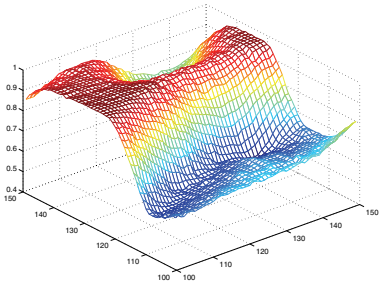
(e) TGV^2_α-L^2

(f) TGV^3_α-L^2

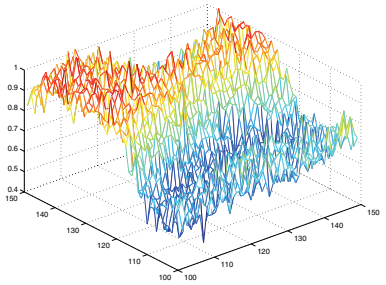
Figure 7. *Denoising of a natural image* (a). (b) *shows the* 481×321 *noisy input image containing zero mean Gaussian noise with standard deviation* $\sigma = 0.05$. (c)–(f) *show the results of ROF, Inf-Conv,* TGV^2_α-L^2, *and* TGV^3_α-L^2 *image denoising (see also Figure* 8*).*



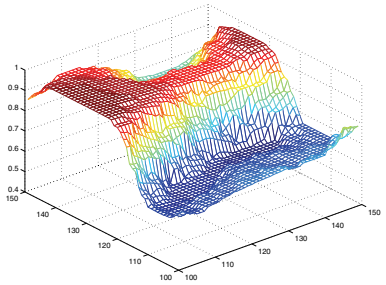
(a) Clean Image



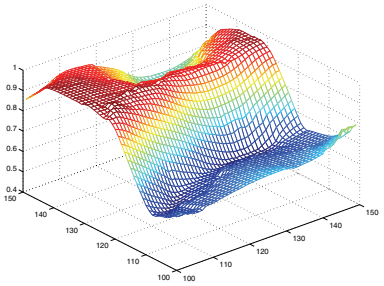
(b) Clean



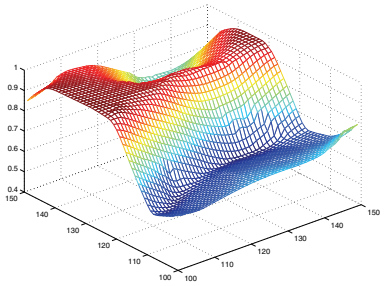
(c) Noisy



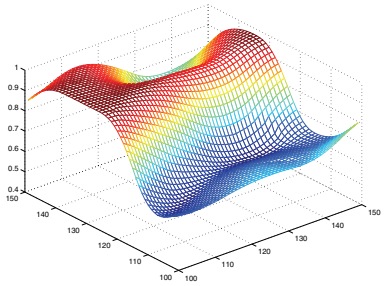
(d) ROF



(e) Inf-Conv



(f) TGV^2_α-L^2



(g) TGV^3_α-L^2

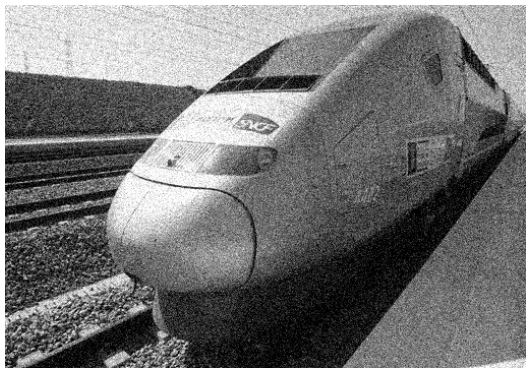
Figure 8. *The close-ups of the blurry background regions show that ROF denoising leads to piecewise constant,* TGV^2_α-L^2 *denoising leads to piecewise affine, and* TGV^3_α-L^2 *denoising leads to piecewise quadratic results.*

symmetry of derivatives, provided that they exist. In case of nonsmoothness, the proposed cost functional provides a weaker measure when compared to the nonsymmetric analogue.

The numerical examples show appealing properties of reconstructed test images. In particular, with the use of third-order regularization, further improvements over second-order are obtained. Besides the denoising application, we expect that the proposed approach would lead to a similar advance for a wider class of problems.

Further analysis including geometric properties of TGV^k_α and strategies for adapting the weights α should be the focus of future research. Another line of research will concentrate on

(a) Clean image



(b) Noisy image

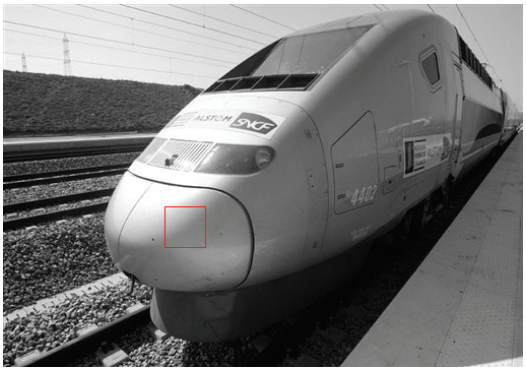
(c) ROF

(d) Inf-Conv

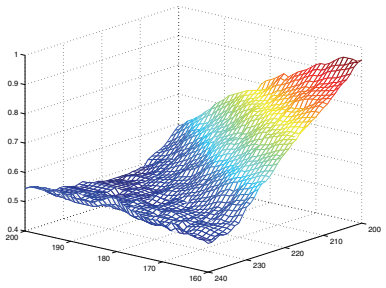
(e) TGV^2_α-L^2

(f) TGV^3_α-L^2

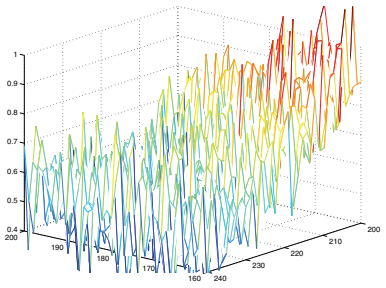
Figure 9. *Denoising of a natural image.* (a) *shows the* 512×357 *clean input image.* (b) *shows the noisy image containing zero mean Gaussian noise with standard deviation* $\sigma = 0.1$. (c)–(f) *show the results of ROF, Inf-Conv,* TGV^2_α-L^2, *and* TGV^3_α-L^2 *denoising. One can see that ROF and Inf-Conv lead to blocky results in smooth image regions.* TGV^2_α-L^2 *and* TGV^3_α-L^2 *models lead to piecewise smooth reconstructions (see also Figure 10).*



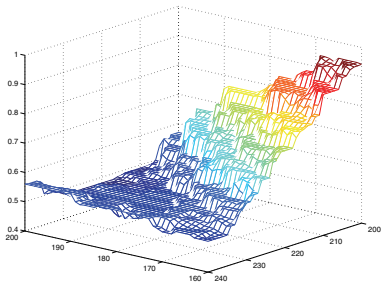
(a) Clean image



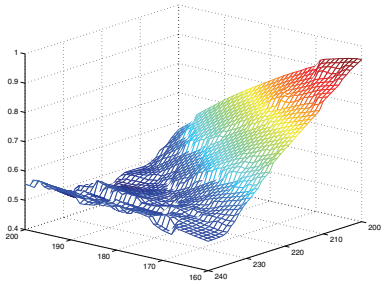
(b) Clean



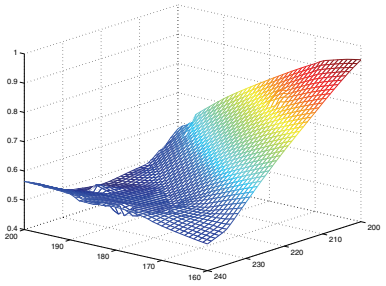
(c) Noisy



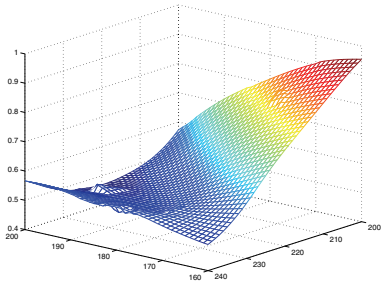
(d) ROF



(e) Inf-Conv



(f) TGV^2_α-L^2



(g) TGV^3_α-L^2

Figure 10. (a) *shows the clean image with a marked smoothly shaded region.* (b)–(g) *show three-dimensional representations of the marked region for the ROF, Inf-Conv,* TGV^2_α-L^2*, and* TGV^3_α-L^2 *models. In this representation one can clearly see that the* TGV *models lead to significantly better results.*

developing faster algorithms. The representation (3.6) depends only on first-order derivatives and hence serves as a good starting point for further investigations.

Appendix A. Tensor field calculus. In this section, we collect some basic results concerning the various operations one can perform on symmetric k-tensors and symmetric k-tensor fields, respectively, such as transformations or taking the l-gradient or l-divergence. As many of these results can be deduced by easy, but sometimes lengthy, computations, they are presented here mainly for the reader's convenience.

First, let us note that for a $\xi \in \mathcal{T}^k(\mathbb{R}^d)$ and a linear transformation $O \in \mathbb{R}^{d\times d}$, the

right-multiplication

$$(\xi O)(a_1, \dots, a_k) = \xi(Oa_1, \dots, Oa_k)$$

also gives a k-tensor for which one can see that with $p = \sigma^{-1}(\beta)$, the coefficients obey

$$(\xi O)_\beta = \sum_{q \in \{1,\dots,d\}^k} \prod_{j=1}^{k} o_{q_j p_j} \xi_{\sigma(q)}. \tag{A.1}$$

Moreover, if $\xi \in \operatorname{Sym}^k(\mathbb{R}^d)$, then $\xi O \in \operatorname{Sym}^k(\mathbb{R}^d)$. Furthermore, let us note that right-multiplication with an orthonormal $O \in \mathbb{R}^{d\times d}$ commutes with the trace of a $\xi \in \mathcal{T}^k(\mathbb{R}^d)$ with $k \geq 2$:

$$\begin{aligned} \operatorname{tr}(\xi O)(a_1, \dots, a_{k-2}) &= \sum_{i,j,j'=1}^{d} o_{ji} o_{j'i} \xi(e_j, Oa_1, \dots, Oa_{k-2}, e_{j'}) \\ &= \sum_{i=1}^{d} \xi(e_i, Oa_1, \dots, Oa_{k-2}, e_i) \\ &= \big(\operatorname{tr}(\xi) O\big)(a_1, \dots, a_{k-2}). \end{aligned} \tag{A.2}$$

Hence, right-multiplication respects orthonormality, as shown in the next lemma.

Lemma A.1. *If $O \in \mathbb{R}^{d\times d}$ is orthonormal, the operation $\xi \mapsto \xi O$ is an orthonormal transformation mapping* $\operatorname{Sym}^k(\mathbb{R}^d) \to \operatorname{Sym}^k(\mathbb{R}^d)$.

Proof. Applying (A.2) k times to $(\xi O \otimes \eta O) = (\xi \otimes \eta) O$ yields

$$\xi O \cdot \eta O = \operatorname{tr}^k\big((\xi \otimes \eta) O\big) = \operatorname{tr}^k(\xi \otimes \eta) O = \xi \cdot \eta. \qquad \blacksquare$$

It turns out that the right-multiplication is, moreover, an appropriate notion for describing the transformation behavior under linear coordinate changes.

Lemma A.2. *Let $O \in \mathbb{R}^{d\times d}$ and let $\xi : \Omega \to \mathcal{T}^k(\mathbb{R}^d)$ be an l times continuously differentiable mapping. Then*

$$\nabla^l \otimes \big((\xi \circ O) O\big) = \big((\nabla^l \otimes \xi) \circ O\big) O. \tag{A.3}$$

Proof. Set $\eta(x) = \xi(Ox)O$ and compute

$$\begin{aligned} (\nabla^l \otimes \eta)(x)(a_1, \dots, a_{k+l}) &= \big(\mathrm{D}^l(\xi \circ O)(x)(a_1, \dots, a_l)\big)(Oa_{l+1}, \dots, Oa_{k+l}) \\ &= \big(\mathrm{D}^l \xi(Ox)(Oa_1, \dots, Oa_l)\big)(Oa_{l+1}, \dots, Oa_{k+l}) \\ &= \big((\nabla^l \otimes \xi) O\big)(Ox)(a_1, \dots, a_{k+l}). \end{aligned} \qquad \blacksquare$$

In particular, for invertible O and $\eta(x) = \xi(Ox)O$, we have that $\eta \in \mathcal{C}^l(O^{-1}\overline{\Omega}, \operatorname{Sym}^k(\mathbb{R}^d))$ if and only if $\xi \in \mathcal{C}^l(\overline{\Omega}, \operatorname{Sym}^k(\mathbb{R}^d))$. The behavior of the l-divergence of a $\xi \in \mathcal{C}^l(\overline{\Omega}, \operatorname{Sym}^{k+l}(\mathbb{R}^d))$ under coordinate change with an orthonormal $O \in \mathbb{R}^{d\times d}$ follows, consequently, by combining (A.3) and (A.2):

$$\begin{aligned} \operatorname{div}^l\big((\xi \circ O) O\big) &= \operatorname{tr}^l\big(\nabla^l \otimes \big((\xi \circ O) O\big)\big) = \operatorname{tr}^l\big(\big((\nabla^l \otimes \xi) \circ O\big) O\big) \\ &= \operatorname{tr}^l\big((\nabla^l \otimes \xi) \circ O\big) O = \big((\operatorname{div}^l \xi) \circ O\big) O. \end{aligned} \tag{A.4}$$

REFERENCES

[A] S.-F. Aujol, *Some first-order algorithms for total variation based image restoration*, J. Math. Imaging Vision, 34 (2009), pp. 307–327.

[BT] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202.

[BG] R. L. Bishop and S. I. Goldberg, *Tensor Analysis on Manifolds*, Macmillan, New York, 1968.

[Ca] J. Carter, *Dual Methods for Total Variation-Based Image Restoration*, Ph.D. thesis, UCLA, Los Angeles, 2001.

[CNN] V. Caselles, A. Chambolle, and M. Novaga, *The discontinuity set of solutions of the TV denoising problem and some extensions*, Multiscale Model. Simul., 6 (2007), pp. 879–894.

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

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

[CEP] T. F. Chan, S. Esedoglu, and F. E. Park, *A Fourth Order Dual Method for Staircase Reduction in Texture Extraction and Image Restoration Problems*, UCLA CAM Report 05-28, UCLA, Los Angeles, 2005.

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

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

[DFLM] G. Dal Maso, I. Fonseca, G. Leoni, and M. Morini, *A higher order model for image restoration: The one-dimensional case*, SIAM J. Math. Anal., 40 (2009), pp. 2351–2391.

[DZ] M. C. Delfour and J.-P. Zolésio, *Shapes and Geometries: Analysis, Differential Calculus, and Optimization*, Adv. Des. Control 4, SIAM, Philadelphia, 2001.

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

[Ne] Yu. Nesterov, *A method for solving a convex programming problem with convergence rate* $O(1/k^2)$, Soviet Math. Dokl., 27 (1983), pp. 372–376.

[Ni] M. Nikolova, *Local strong homogeneity of a regularized estimator*, SIAM J. Appl. Math., 61 (2000), pp. 633–658.

[PS] C. Pöschl and O. Scherzer, *Characterization of minimizers of convex regularization functionals*, in Frames and Operator Theory in Analysis and Signal Processing, Contemp. Math. 451, AMS, Providence, RI, 2008, pp. 219–248.

[R] W. Ring, *Structural properties of solutions to total variation regularization problems*, M2AN Math. Model. Numer. Anal., 34 (2000), pp. 799–810.

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

[SGGHL] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen, *Variational Methods in Imaging*, Springer, New York, 2009.

[SS] S. Setzer and G. Steidl, *Variational methods with higher order derivatives in image processing*, in Approximation Theory XII: San Antonio 2007, M. Neamtu and L. L. Schumaker, eds., Nashboro Press, Brentwood, TN, 2008, pp. 360–385.

[S] V. A. Sharafutdinov, *Integral Geometry of Tensor Fields*, VSP, Utrecht, The Netherlands, 1994.