**SpezialForschungsBereich F 32**

Karl–Franzens Universität Graz
Technische Universität Graz
Medizinische Universität Graz

# Total Generalized Variation

K. Bredies K. Kunisch T. Pock

SFB-Report No. 2009–038 August 2009

A–8010 GRAZ, HEINRICHSTRASSE 36, AUSTRIA

Supported by the
Austrian Science Fund (FWF)

SFB sponsors:

- **Austrian Science Fund** (FWF)
- **University of Graz**
- **Graz University of Technology**
- **Medical University of Graz**
- **Government of Styria**
- **City of Graz**

# Total Generalized Variation

Kristian Bredies Karl Kunisch Thomas Pock

August 31, 2009

**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 semi-norm, 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 does not lead to a staircasing effect.

**Keywords:** Bounded variation, total generalized variation, tensor fields, regularization, image denoising.

**AMS Subject Classification:** 49J52, 49N45, 68U10.

## 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 $G$ denoted 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 $\alpha$ is the regularization parameter and $|\cdot|$ again denotes a Hilbertian norm or semi-norm. It is now becoming well-accepted that the mathematical and computational simplicity of the norm-of-squares terms must be put into

perspective with 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) 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 proposal for a refined choice was given in [ROF]. It uses the bounded variation semi-norm

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

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 semi-norm, 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 bounded variation semi-norm tends to be piecewise constant (staircasing). For one-dimensional images this staircasing effect was rigorously established in [DFLM], for example. For diverse other aspects on the topic of constructing appropriate regularization or filter functionals in image reconstruction we refer to [SGGHL] and the references given there.

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

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

where $\mathrm{Sym}^k(\mathbb{R}^d)$ denotes the space of symmetric tensors of order $k$ with arguments in $\mathbb{R}^d$, and $\alpha_l$ are fixed positive parameters. For the definition of the remaining quantities, we ask for the readers' patience until Section 2. Suffice it to say at this moment that for $k = 1$, $\alpha_0 = 1$ the semi-norm $\mathrm{TGV}^k_\alpha$ coincides with the bounded variation semi-norm. We refer to $\mathrm{TGV}^k_\alpha$ as total generalized bounded variation of order $k$ with weight $\alpha \in \mathbb{R}^k$. From the definition of $\mathrm{TGV}^k_\alpha$ it is immediately clear that it involves (generalized) derivatives of $u$ of order $i = 1, \dots, k$, and that the kernel of $\mathrm{TGV}^k_\alpha$ is the set of polynomials of order less than $k$. Intuitively the total generalized bounded variation further automatically balances the first to the $k$-th derivatives of $u$ among themselves. It will be shown that $\mathrm{TGV}^k_\alpha$ 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 $\Omega'$. If differs, however, for functions which are not piecewise constant.

As a further preview we point out that in dimension 1, with $\Omega = (0,1)$, $k = 2$, $\alpha_0, \alpha_1 > 0$ such that $\alpha_0/\alpha_1 < 1/2$, we have for

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

$a_1, a_2, b_1, b_2 \in \mathbb{R}$ and $\Omega_1 = ]0, c[$, $\Omega_2 = ]c, 1[$ for some $\alpha_0/\alpha_1 < c < 1 - \alpha_0/\alpha_1$, that

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

In particular, $\operatorname{TGV}^2_\alpha(u)$ does not penalize the derivative of order $l = 0, 1$ unless it jumps at $c$.

As already mentioned our motivation for studying $\operatorname{TGV}^2_\alpha(u)$ 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 inf-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 upon 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 $\Phi$ 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

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

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

As we shall see in Section 3, even for the case $k = 2$ the proposed functional $\mathrm{TGV}^k_\alpha$ does not agree with those regularization functionals which were considered earlier in the literature.

Let us give a briefly 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 its 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 inf-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 are contained in Section 4. The Appendix contains some basic results involving symmetric $k$-tensor fields.

# 2 Preliminaries

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.

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

## 2.1 Spaces of symmetric tensor fields

In order to introduce our definition of total generalized variation, let us first fix some notation. Throughout the paper, $d \geq 1$ denotes the dimension which is typically 2 or 3, in applications. Let

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

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

be the vector space of *$k$-tensors* and *symmetric $k$-tensors*, respectively (actually, these are spaces of $(0,k)$-tensors, but since we only deal with covariant vectors, we omit the 0). Here $\xi \in \mathcal{T}^k(\mathbb{R}^d)$ is called symmetric, if

$\xi(a_1,\dots,a_k) = \xi((\pi(a_1),\dots,\pi(a_k))$ for all $\pi \in S_k$, where $S_k$ denotes the permutation group of $\{1,\dots,k\}$.

The case $k = 0$ corresponds to scalar values, for $k = 1$, $\mathrm{Sym}(\mathbb{R}^d, \mathbb{R}) = \mathbb{R}^d$ while 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*

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

yields an element of $\mathcal{T}^{k+l}(\mathbb{R}^d) \in \mathcal{T}^k(\mathbb{R}^d)$. 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

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

where $e_i$ denotes the $i$-th 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

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

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

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

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 $\sigma$ from $p \in \{1,\dots,d\}^k$ to multiindices $\beta \in \mathbb{N}^d \cup \{0\}$ 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

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

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

$$\sigma^{-1}(\beta)_j = \min\{m \mid \sum_{n=1}^m \beta_n \geq j\},$$

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

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

The multiindex 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

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

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

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

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

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

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:

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

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

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

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

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

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

As it is also done in the mathematical theory of elasticity, for example, we are in particular 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 $\xi : \Omega \to \mathcal{T}^k(\mathbb{R}^d)$:

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

$$= \frac{1}{k+1} \sum_{j=1}^{k+1} \sum_{i=1}^{d} a_{j,i} \frac{1}{k!} \sum_{\substack{\pi \in S_{k+1} \\ \pi(1)=j}} \frac{\partial(|||\xi)}{\partial x_i}(a_{\pi(2)}, \ldots, a_{\pi(k+1)})$$

$$= \frac{1}{(k+1)!} \sum_{\pi \in S_{k+1}} \sum_{i=1}^{d} a_{\pi(1),i} \frac{\partial(|||\xi)}{\partial x_i}(a_{\pi(2)}, \ldots, a_{\pi(k+1)})$$

$$= |||(\nabla \otimes |||\xi)(a_1, \ldots, a_{k+1}).$$

Spaces of continuously differentiable functions in this sense are defined as:

$$\mathcal{C}^l(\overline{\Omega}, \mathrm{Sym}^k(\mathbb{R}^d)) = \{\xi : \overline{\Omega} \to \mathrm{Sym}^k(\mathbb{R}^d) \\ \big| \, \nabla^m \otimes \xi \text{ continuous on } \overline{\Omega}, m = 0, \ldots, l\}$$

$$\|\xi\|_{l,\infty} = \max_{m=0,\ldots,l} \|\mathcal{E}^m(\xi)\|_\infty.$$

We also use symmetric $k$-tensors fields with compact support in $\Omega$:

$$\mathcal{C}_{\mathrm{c}}^l(\Omega, \mathrm{Sym}^k(\mathbb{R}^d)) = \{\xi \in \mathcal{C}^k(\overline{\Omega}, \mathrm{Sym}^k(\mathbb{R}^d)) \;\big|\; \mathrm{supp}\, \xi \subset\subset \Omega\},$$
$$\mathcal{C}_{\mathrm{c}}^\infty(\Omega, \mathrm{Sym}^k(\mathbb{R}^d)) = \bigcap_{l \geq 1} \mathcal{C}_{\mathrm{c}}^l(\Omega, \mathrm{Sym}^k(\mathbb{R}^d)).$$

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

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

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 $\mathrm{div}^k\, \mathrm{div}^l\, \xi = \mathrm{div}^{k+l}\, \xi$ whenever the expression makes sense. Let us moreover point out the special case $k = 2$ with the interpretation as symmetric matrix:

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

For $k = 1$, the divergence for mappings $\Omega \to \mathrm{Sym}^1(\mathbb{R}^d)$ 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 $\Omega$ be a bounded Lipschitz domain and $\xi \in \mathcal{C}^1(\overline{\Omega}, \mathrm{Sym}^{k+1}(\mathbb{R}^d))$ be a smooth $(k+1)$-symmetric tensor field. Then, for each smooth symmetric $k$-tensor field $\eta : \Omega \to \mathrm{Sym}^k(\mathbb{R}^d)$ we have:*

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

*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 $\sigma\big((p,i)\big) = \sigma(p) + e_i$. Therefore, with integration by parts and remembering that the symmetrization is the orthogonal projection onto the space of symmetric $k$-tensors,

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

yielding the desired identity. □

*Remark* 2.2. It is easy to see that if $\xi$ has compact support in $\Omega$, 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 $\xi \in L^1_{\mathrm{loc}}(\Omega, \mathrm{Sym}^k(\mathbb{R}^d))$, a symmetric $(k+l)$-tensor field $\eta \in L^1_{\mathrm{loc}}(\Omega, \mathrm{Sym}^{k+l}(\mathbb{R}^d))$ is called the *$l$-th weak symmetrized derivative* of $\xi$ if the identity

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

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

Note again that since we only test with symmetric $(k+l)$-tensor fields, we are only able to determine the symmetrized gradients. The Sobolev spaces

associated with this notion of derivative are then given by:

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

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

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

# 3 Total generalized variation semi-norms

## 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, $k \ge 1$ and $\alpha_0, \dots, \alpha_{k-1} > 0$. Then, the *total generalized variation* of order $k$ with weight $\alpha$ for $u \in L^1_{\mathrm{loc}}(\Omega)$ is defined as the value of the functional

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

with taking the value $\infty$ if the respective set is unbounded from above.

The space

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

is called the space of *functions of bounded generalized variation* of order $k$ with weight $\alpha$.

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

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

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. $\mathrm{TGV}_\alpha^k$ *is a semi-norm 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 order less than* $k$,

3. *for fixed $k$ and positive weights $\alpha, \tilde{\alpha} \in \mathbb{R}^k$, the semi-norms $\mathrm{TGV}^k_\alpha$ and $\mathrm{TGV}^k_{\tilde{\alpha}}$ are equivalent,*

4. *$\mathrm{TGV}^k_\alpha$ 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*

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

*Proof.* Let us begin proving the first statement. Note that TGV can be interpreted as the dual semi-norm in which the set

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

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

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

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

For the second statement, suppose $u$ is a polynomial of degree less that $k$ 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 $v \in \mathcal{C}^k_{\mathrm{c}}(\Omega, \mathrm{Sym}^k(\mathbb{R}^d))$

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

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 $\lambda v$ and $-\lambda v$ to get

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

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 $\Omega$ is connected.

The asserted equivalence of norms according to the third statement can be proven 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 $\alpha$ and $\tilde{\alpha}$ leads to the desired equivalence.

For proving the rotational invariance as stated in the fourth item, 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 the fifth statement 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 $v \in \mathcal{C}^l_{\mathrm{c}}(r^{-1}\Omega, \mathrm{Sym}^k(\mathbb{R}^d))$ 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 the third statement 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 only have to prove that $\mathrm{BGV}^k = \mathrm{BGV}^k_\alpha$ is complete for some weight $\alpha$.

To achieve this, we first show that $\mathrm{TGV}^k_\alpha$ always gives a lower semi-continuous 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 $v \in \mathcal{C}^k_{\mathrm{c}}(\Omega, \mathrm{Sym}^k(\Omega))$ with $\|\operatorname{div}^l v\|_\infty \le \alpha_l$ it follows that

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

Taking the supremum thus yields

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

meaning that $\mathrm{TGV}^k_\alpha$ is indeed lower semi-continuous 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 semi-continuity 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 only remains to show that $u$ is also the limit in the corresponding norm. But this follows again from the lower semi-continuity of $\mathrm{TGV}^k_\alpha$ on $L^1(\Omega)$: For each $\varepsilon > 0$ one chooses an $n$ such that for all $n' \geq n$ holds that $\mathrm{TGV}^k_\alpha(u^n - u^{n'}) \leq \varepsilon$, so letting $n' \to \infty$ gives

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

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 $u : \Omega \to \mathbb{R}$ 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 $q_i$ polynomials of maximal degree $k-1$. Then,*

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

*where $\Omega_0 = \Omega \setminus \bigcup_{i=1}^n \overline{\Omega_i}$, $q_0 = 0$ and $\Gamma_{ij} = \partial\Omega_i \cap \partial\Omega_j \cap \Omega$ and $\nu_i$ is the outer normal to $\Omega_i$.*

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

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

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

Since all $\Omega_i$ are disjoint, it is possible to write

$$\partial\Omega_i \cap \Omega = \partial\Omega_i \cap \partial(\Omega \setminus \Omega_i) \cap \Omega = \partial\Omega_i \cap \partial\Big(\bigcup_{j\neq i} \overline{\Omega_j}\Big) \cap \Omega = \bigcup_{j=1,\, j\neq i}^{n} \partial\Omega_i \cap \partial\Omega_j \cap \Omega,$$

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

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

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

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

and, consequently, to the desired estimate. □

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

*Remark* 3.8. In a formal sense one can, for smooth functions, interpret $\mathrm{TGV}_\alpha^k$ as an infimal convolution of functionals incorporating the total variation (in the measure-theoretic sense) of each derivative up to order $k$. Indeed, defining for $l = 0, \dots, k-1$ and $\bar{\alpha} > 0$ the sets

$$K_{\bar{\alpha}}^l = \{ \operatorname{div}^k v \;\big|\; v \in \mathcal{C}_\mathrm{c}^k(\Omega, \operatorname{Sym}^k(\mathbb{R}^d)) \;,\; \|\operatorname{div}^l v\|_\infty \leq \bar{\alpha} \},$$

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

$$\mathrm{TGV}_\alpha^k = \Big( \sum_{l=0}^{k-1} I_{K_{\alpha_l}^l} \Big)^* .$$

The range of $\operatorname{div}^l$ is dense in the orthogonal complement of the kernel of its adjoint $(-1)^l \mathcal{E}^l$. Hence, one can express, for smooth $u$, the Fenchel dual for $I_{K_{\bar{\alpha}}^l}$ as the infimal convolution

$$\begin{aligned} I_{K_{\bar{\alpha}}^l}^*(u) &= \sup \Big\{ \int_\Omega u \operatorname{div}^k v \, \mathrm{d}x \;\Big|\; v \in \mathcal{C}_\mathrm{c}^k(\Omega, \operatorname{Sym}^k(\mathbb{R}^d)) \;,\; \|\operatorname{div}^l v\|_\infty \leq \bar{\alpha} \Big\} \\ &= (I_{K_{\bar{\alpha}}^l} + I_{range(\operatorname{div}^l)})^* ((-1)^{k-l} \nabla^{k-l} u) \\ &= \inf \Big\{ \bar{\alpha} \int_\Omega |\nabla^{k-l} u + w| \, \mathrm{d}x \;\Big|\; w \in \mathcal{C}^l(\Omega, \operatorname{Sym}^{k-l}(\mathbb{R}^d)),\; \mathcal{E}^l(w) = 0 \Big\}, \end{aligned}$$

where $I$ denotes the indicator function of the set specified by the index. By $k$-fold inf-convolution this implies that

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

One can interpret this representation as follows. The total generalized variation looks, in a certain sense, for the best decomposition of $u$ into summands $u_l$ with $l = 1, \dots, k$ for which the weighted sum of $l$-th derivatives is minimal with respect to the Lebesgue 1-norm (w.r.t. $x$) and up to some functions which lie in the respective kernels. This way, it automatically adapts to the smoothness of $u$: If some derivative of $u$ is large or oscillatory, resulting in a large $\|\nabla^l u\|_1$, the major features of $u$ will be distributed among $u_0, \dots, u_{k-l-1}$. The other way around, if $\nabla^l u$ is almost flat, differentiation may result in a small value of $\|\nabla^l u\|_1$ and hence one can reduce the right-hand side in (3.5) by distributing a major part of $u$ to $u^{k-l-1}$. Roughly speaking, if there exist some minimizing arguments $u_0, \dots, u_{k-1}$ for (3.5), these may indicate the smoothness of $u$ in a way that $u \approx u_l$ where $u$ is $k-l$ smooth.

*Remark* 3.9. The "dualization" in the definition of the functional $\mathrm{TGV}_\alpha^k$ can also be formally interpreted in terms of the Fenchel duality formula.

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

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

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

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

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

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

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

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

$$= \inf_{u_{k-\kappa+1}\in\mathcal{C}^{\kappa-1}(\overline{\Omega},\mathrm{Sym}^{k-\kappa+1}(\mathbb{R}^d))} \alpha_{\kappa-1}\|\mathcal{E}(u_{k-\kappa}) - u_{k-\kappa+1}\|_1 + \Big(\sum_{l=0}^{\kappa-2} I_{K_l}\Big)^*\big((-1)^{\kappa-1}\mathcal{E}^{\kappa-1}(u_{k-\kappa+1})\big)$$

$$= \inf_{u_{k-\kappa+1}\in\mathcal{C}^{\kappa-1}(\overline{\Omega},\mathrm{Sym}^{k-\kappa+1}(\mathbb{R}^d))} \alpha_{\kappa-1}\|\mathcal{E}(u_{k-\kappa}) - u_{k-\kappa+1}\|_1 + \mathrm{GV}_\alpha^{\kappa-1}\big(\mathcal{E}(u_{k-\kappa+1})\big).$$

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

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

and consequently

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

This representation can be interpreted analogously to (3.5); for simplicity we restrict our argument 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 $\Omega'$ 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 $\mathrm{TGV}_\alpha^k$ 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 $\mathcal{C}_\mathrm{c}^k(\Omega, \mathrm{Sym}^k(\mathbb{R}^d))$, influences the functional. If we would have taken $\mathcal{C}_\mathrm{c}^k(\Omega, \mathcal{T}^k(\mathbb{R}^d))$ in Definition 3.1 instead, i.e.

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

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 to modify (3.6) is to restrict the functions $u_l$ to $l$-th 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

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

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.

## 3.2 Second-order total generalized variation

In order to get some more intuition on how the total generalized variation measures functions, we make some observations for the case $k = 2$. Specifically, $\mathrm{TGV}^2_\alpha$ 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 $]0, 1[$.

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

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

*Proof.* First observe that (3.4) immediately gives that

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

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

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

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 $\sigma^\varepsilon$ is continuously differentiable in a neigborhood of $\partial\Omega'$ since $\Omega'$ 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, $\mathrm{supp}\,\sigma^\varepsilon \subset\subset \Omega$ as well as $|\nabla\sigma^\varepsilon(x)| \leq 1$ almost everywhere in $\Omega$. Choosing a standard mollifier $G \in \mathcal{C}^\infty_0(B_1(0))$ and denoting by $G_\varepsilon$ its dilated versions, it is immediate that $v^\varepsilon_0 = \alpha_1 \sigma^\varepsilon * G_{\varepsilon/4}$ satisfies $v^\varepsilon_0 \in \mathcal{C}^\infty_0(\Omega)$.

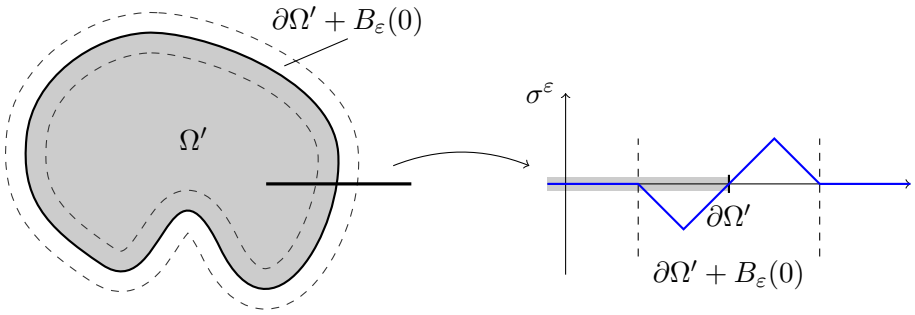



Figure 1: Illustration of the construction of the signed distance function $\sigma^\varepsilon$. On the left-hand side, the smooth set $\Omega'$, its boundary and the corresponding $\varepsilon$-tube $\partial\Omega' + B_\varepsilon(0)$ is depicted. On the right-hand side you can see, qualitatively,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

One can then compute that $v_0^\varepsilon \to 0$ uniformly in $\Omega$ as well as $\nabla v_0^\varepsilon(x) \to \alpha_1 \nu(x)$ for $x \in \partial\Omega'$ as $\varepsilon \to 0$. Consequently, choosing $\varepsilon$ small enough yields $\|v_0^\varepsilon\|_\infty \leq \alpha_0/\sqrt{d}$ and, since the Lipschitz constant of $\sigma^\varepsilon$ 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 $v^\varepsilon$ 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 follows indeed that $\mathrm{TGV}_\alpha^2(u) \geq \alpha_1 \operatorname{TV}(u)$ what was to show. □

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

**Example 3.12.** Fix $k = 2$ and $\alpha_0, \alpha_1 > 0$ such that $\alpha_0/\alpha_1 < 1/2$. Let $u : ]0,1[ \to \mathbb{R}$ be

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

with $a_1, a_2, b_1, b_2 \in \mathbb{R}$ and $\Omega_1 = ]0, c[$, $\Omega_2 = ]c, 1[$ for some $\alpha_0/\alpha_1 < c < 1 - \alpha_0/\alpha_1$.

We compute $\mathrm{TGV}_\alpha^2$ for some $\alpha_0, \alpha_1 > 0$. For this purpose, choose a $v \in \mathcal{C}_c^2(]0,1[)$ and apply integration by parts twice to get

$$\int_0^1 u v'' \, \mathrm{d}x = \big(p_2(c) - p_1(c)\big) v'(c) + \big(p_1'(c) - p_2'(c)\big) v(c). \tag{3.8}$$

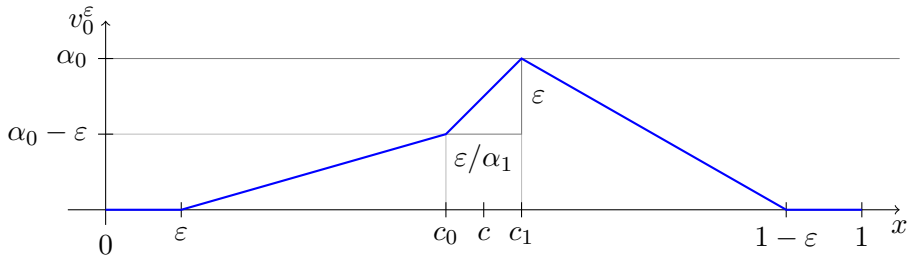



Figure 2: Schematic illustration of the functions $v_0^\varepsilon$.

By Proposition 3.6 we already have

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

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

$$v_0^\varepsilon(x) = \begin{cases} 0 & \text{for } 0 \leq x \leq \varepsilon \\ (\alpha_0 - \varepsilon)(x - \varepsilon)/(c_0 - \varepsilon) & \text{for } \varepsilon \leq x \leq c_0 \\ \alpha_0 - \varepsilon + \alpha_1 (x - c_0) & \text{for } c_0 \leq x \leq c_1 \\ \alpha_0 (x + \varepsilon - 1)/(c_1 + \varepsilon - 1) & \text{for } c_1 \leq x \leq 1 - \varepsilon \\ 0 & \text{for } 1 - \varepsilon \leq x \leq 1. \end{cases} \tag{3.9}$$

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 $\|v_0^\varepsilon\|_\infty \leq \alpha_0$ as well as $\|(v_0^\varepsilon)'\|_\infty \leq \alpha_1$, the latter taking the choice of $c$ and $\alpha_0/\alpha_1 < 1/2$ 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 $G_\varepsilon$ its dilated versions, the functions $v^\varepsilon = v_0^\varepsilon * G_{\varepsilon/2}$ satisfy $v^\varepsilon \in \mathcal{C}_\mathrm{c}^2(]0,1[)$ with

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

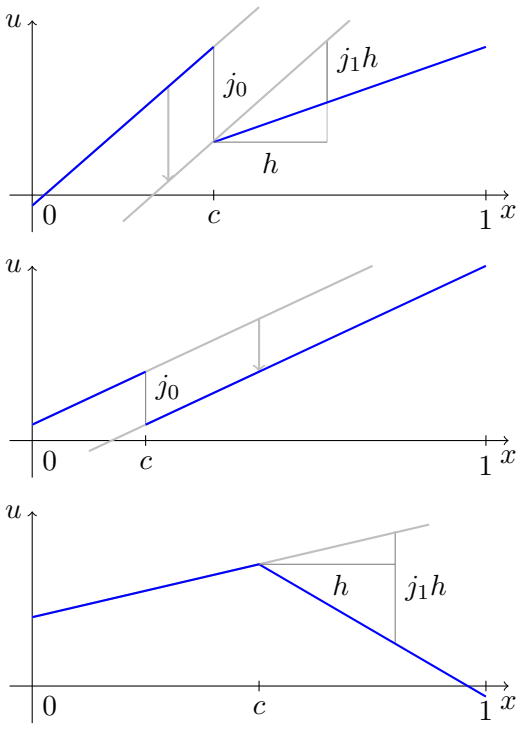
So, plugging the sequence into (3.8) gives that the estimate is sharp, i.e.

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

With analog constructions for the cases where $p_2(c) - p_1(c) \leq 0$ or $p_1'(c) - p_2'(c) \leq 0$, this follows for all $u$ of the form (3.7).

Note that if $p_1'(c) = p_2'(c)$, then one need not take care of $v_0^\varepsilon(c) \to \alpha_0$, hence constructing $v_0^\varepsilon$ with $\alpha_0 = 0$ in (3.9) gives that $\|v_0^\varepsilon\|_\infty \to 0$ as $\varepsilon \to 0$ and, even without the restrictions for $\alpha_0$, $\alpha_1$ and $c$, $\|(v_0^\varepsilon)'\|_\infty \leq \alpha_1$ for small $\varepsilon > 0$. Consequently, the above equation for $\mathrm{TGV}_\alpha^2(u)$ is also true in a more general case. As one can easily see, however, the conditions cannot be relaxed in the case $p_1(c) = p_2(c)$.

Finally, Figure 3 depicts some cases of $u$ and how the values of $\mathrm{TGV}_\alpha^2(u)$ can be expressed.




$\mathrm{TGV}^2_\alpha(u) = \alpha_1 j_0 + \alpha_0 j_1$

$\mathrm{TGV}^2_\alpha(u) = \alpha_1 j_0$

$\mathrm{TGV}^2_\alpha(u) = \alpha_0 j_1$

Figure 3: Illustration of some piecewise affine functions and the corresponding values of $\mathrm{TGV}^2_\alpha$.

# 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 $\mathrm{TGV}^2_\alpha$ regularization functional with a quadratic $L^2$ data fidelity term. We have two reasons for that. First, the $\mathrm{TGV}^2_\alpha$-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 an order greater than two.

The $\mathrm{TGV}^2_\alpha - L^2$ 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 non-smooth minimization problem can be obtained by solving the Fenchel predual problem, an approach which recently became 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

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

Solutions $u^*$ and $v^*$ of (4.1) and (4.2), respectively, satisfy $u^* = f - \operatorname{div}^2 v^*$, 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 simple projected-gradient based algorithm FISTA recently proposed in [BT].

The subsequent section then shows numerical experiments we carried out using this algorithm. As expected the proposed $\mathrm{TGV}^2_\alpha - L^2$ image denoising model is able to restore piecewise affine functions. and in contrast to the usual total variation denoising model [ROF] ($\mathrm{TGV}^1_\alpha - L^2$ in our terms), it does not exhibit the staircasing effect. Furthermore we will show that replacing $\mathrm{TGV}^2_\alpha$ in (4.1) by a $\mathrm{TGV}^3_\alpha$ 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 only consider the case $d = 2$, i.e. the case of two dimensional images. 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

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

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 only need to 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 analog of the (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 $\mathrm{TGV}^2_\alpha$ $-L^2$-denoised version of $f^h$. It can be equivalently formulated as

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

The discrete $\infty$-norms on $V^h$ and $W^h$ can be expressed by

$$\begin{aligned}
v^h \in V^h: \qquad & \|v^h\|_\infty = \max_{i,j} \left( (v^h_1)^2_{i,j} + (v^h_2)^2_{i,j} + 2(v^h_3)^2_{i,j} \right)^{1/2}, \\
w^h \in W^h: \qquad & \|w^h\|_\infty = \max_{i,j} \left( (w^h_1)^2_{i,j} + (w^h_2)^2_{i,j} \right)^{1/2}.
\end{aligned}$$

Moreover, for the discretization of the divergence operator we use a recursive application of forward and backward differences in such a way that the

outmost 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

$$\mathrm{div}^h : V^h \to W^h \quad , \quad (\mathrm{div}^h\, v^h)_{i,j} = \begin{pmatrix} (\delta^h_{x+} v^h_1)_{i,j} + (\delta^h_{y+} v^h_3)_{i,j} \\ (\delta^h_{x+} v^h_3)_{i,j} + (\delta^h_{y+} v^h_2)_{i,j} \end{pmatrix},$$

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

$$\begin{aligned} \big((\mathrm{div}^h)^2 v^h\big)_{i,j} &= \big(\mathrm{div}^h(\mathrm{div}^h\, v^h)\big)_{i,j} \\ &= (\delta^h_{x-}\delta^h_{x+} v^h_1)_{i,j} + (\delta^h_{y-}\delta^h_{y+} v^h_2)_{i,j} + \big((\delta^h_{y-}\delta^h_{x+} + \delta^h_{x-}\delta^h_{y+}) v^h_3\big)_{i,j}, \end{aligned}$$

where the forward and backward differences are defined as

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

and

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

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 $(\mathcal{E}^h)^2 = \big((\mathrm{div}^h)^2\big)^*$. By computing the adjoint of $(\mathrm{div}^h)^2$ and taking into account the symmetry of $v^h$, we arrive at

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

Note that by our choice 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. 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 Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) recently proposed in [BT], which shares a convergence rate of $O(1/k^2)$. This means that one needs $O(1/\sqrt{\varepsilon})$ iterations to compute an $\varepsilon$-accurate solution in terms of the functional values. Furthermore, FISTA is easy to implement and can be efficiently parallelized. The outline of FISTA applied to (4.3) is as follows: We choose $v_0^h = 0$, $\bar{v}_0^h = 0$ and $t_0 = 1$. Then, for $k \geq 0$ we let

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

where $\tau > 0$ is some prescibed step-size and $\Pi_{K^h}$ 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 $\{v_k^h\}$ generated by algorithm (4.5) is such that for any $k \geq 0$*

$$0 \leq E(v_k^h) - E\big((v^h)^*\big) \leq \frac{2L^2\|v_0^h - (v^h)^*\|^2}{(k+1)^2}, \tag{4.6}$$

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

$$\begin{aligned} (v_1^h)_{-1,j} &= (v_1^h)_{0,j}, & (v_1^h)_{M,j} &= (v_1^h)_{M-1,j}, \\ (v_2^h)_{i,-1} &= (v_2^h)_{i,0}, & (v_2^h)_{i,N} &= (v_2^h)_{i,N-1}, \\ (v_3^h)_{i,-1} &= (v_3^h)_{-1,j} = 0, & (v_3^h)_{i,N} &= (v_3^h)_{M,j} = 0. \end{aligned}$$

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 the following. 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 $(a+b)^2 \leq 2a^2 + 2b^2$,

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

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

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

Together, we find

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

Then, by substitution into (4.7) we get $L^2 \leq \frac{64}{h^4}$. 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 moreover know that the corresponding sequence $\{u_{k_l}^h\}$ converges to a $(u^h)^* = f^h - (\mathrm{div}^h)^2(v^h)^*$. 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 the $v^h$ 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

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

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.

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

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

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

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

We adopt a variant of FISTA to compute the minimizer of (4.9) where we exploit 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 $k \geq 0$ 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 $\sigma = \frac{h^2}{8} \leq \|\operatorname{div}^h\|^{-2}$ 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 feasibilty error of $\Pi_{K^h}(\hat{v}^h)$ is below a threshold $\varepsilon_p$.

# 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 ($\mathrm{TGV}^2_\alpha - L^2$) consistently outperforms both the standard TV model of [ROF] (which is equivalent to the $\mathrm{TGV}^1_\alpha - L^2$ model) and the Inf-Conv model of [CL]. Finally we show that a third order model ($\mathrm{TGV}^3_\alpha - L^2$) can further improve the results.

## 5.1 Efficiency of the Numerical Algorithm

In our first experiment, we compare the theoretical efficiency estimate of FISTA with the practical implementation using a simple synthetic input image. As stated above, FISTA allows to compute a bound on the accuracy $\varepsilon$ of the function 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 $\varepsilon$, 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 $\mathrm{TGV}^2_\alpha - L^2$ 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 ran 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 our first order minimization algorithm (FISTA). In order to compute the true minimal function value $E((v^h)^*)$ we ran FISTA 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.

## 5.2 Synthetic Images

In our second experiment we evaluate the performance of the $\mathrm{TGV}^2_\alpha - L^2$ model using a piecewise affine test image. We compare our model to the standard ROF model and the Inf-Conv model. The parameters of each model were optimized to achieve the best reconstruction with respect to the root mean squared error. Figure 5 shows the input images and the

| | Test image | |
|---|---|---|
| Model | Piecewise affine | Piecewise smooth |
| ROF | 0.0130 | 0.0176 |
| Inf-Conv | 0.0110 | 0.0125 |
| $\mathrm{TGV}^2_\alpha - L^2$ | **0.0081** | 0.0083 |
| $\mathrm{TGV}^3_\alpha - L^2$ | - | **0.0081** |

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.

reconstructed images. For better visualization we additionally provide 3D 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-Conf model performs better but also has some staircasing near discontinuities. The proposed $\mathrm{TGV}^2_\alpha - L^2$ model performs best. It leads to a natural piecewise affine approximation of the noisy data with sharp discontinuities inbetween. The quantitative evaluation of the reconstruction emphasizes the visual impression of the results. See Table 1 for the exact RMSE values.

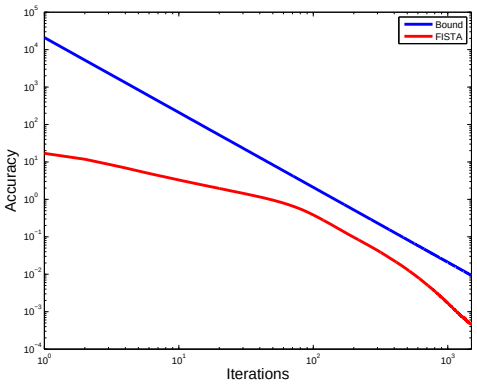
In our third experiment, we apply total generalized variation regularization up to order three for the reconstruction of a piecewise smooth test image. Again, we compare our models to the ROF model and the Inf-Conv model and, as in the previous experiment, all parameters were optimized in order to meet 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 $\mathrm{TGV}^2_\alpha - L^2$ and $\mathrm{TGV}^3_\alpha - L^2$ 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

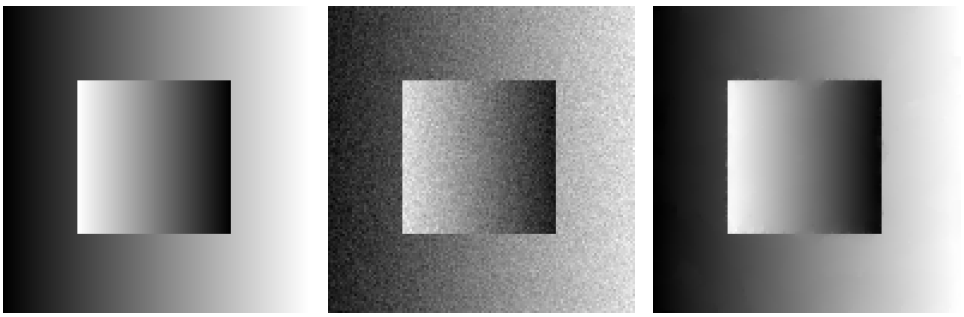
Finally, we apply our total generalized variation models to denoising of natural images. Figure 7 shows a noisy image of a penguin in front of a blurry background. While the textured parts (e.g. the rock) of the image are equally

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 closeups make clear the characteristics of the three models. ROF denoising leads to a piecewise constant, $\mathrm{TGV}^2_\alpha - L^2$ denoising leads to a piecewise affine and $\mathrm{TGV}^3_\alpha - L^2$ denoising leads to a piecewise quadratic approximation of the image function.

Figure 8 shows the denoising capabilities of the proposed models in case of severe noise. While ROF denoising leads to very blocky results, the proposed total generalized variation models leads to significantly better results. The closeups show that in regions of high curvature, the third order model leads to further improvements over the second order model.

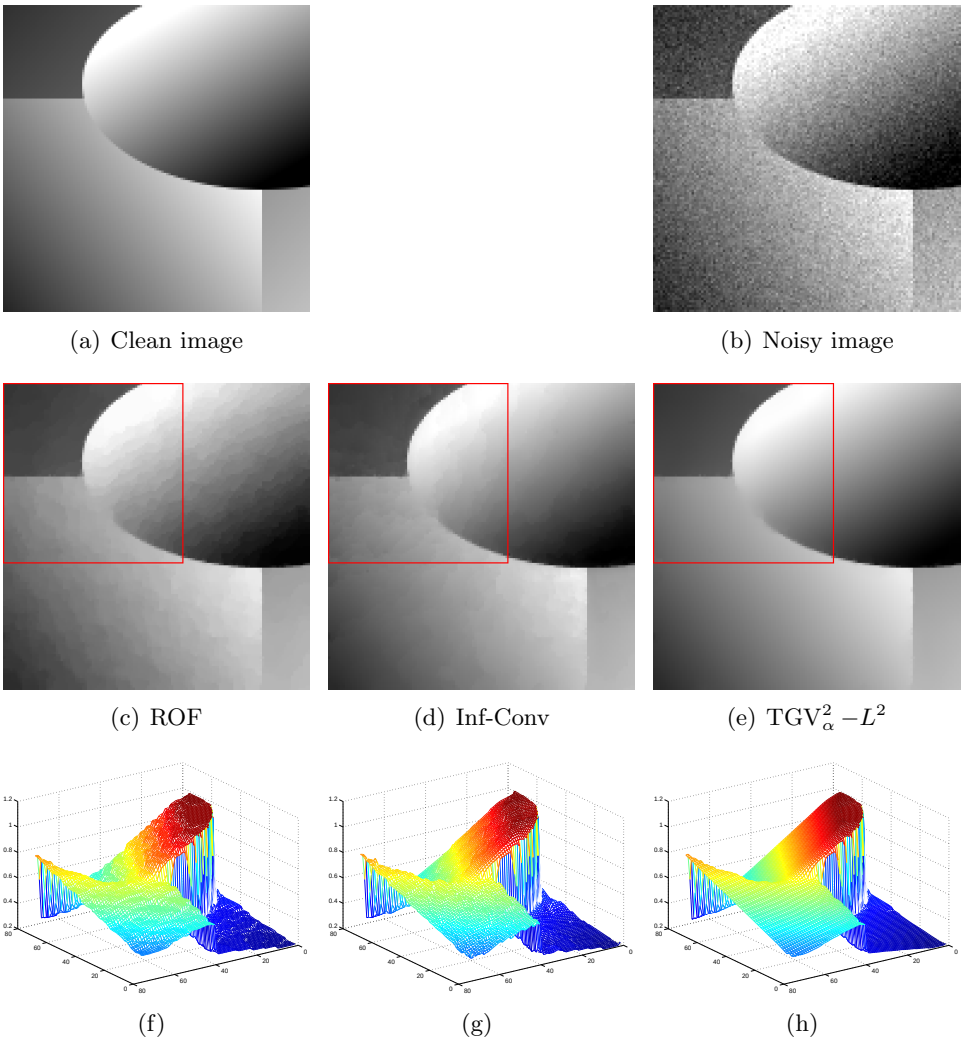



(a)



(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 FISTA versus the empirical observation. The theoretical bound provides an useful estimate sharing the same asymptotic behavior, but is still outperformed in practice. (b) Input image of size $128 \times 128$. (c) Degraded image containing 5% zero mean Gaussian noise. (d) Reconstructed image using the proposed $\mathrm{TGV}^2_\alpha - L^2$ model. Note that the proposed method almost perfectly reconstructs the input image.

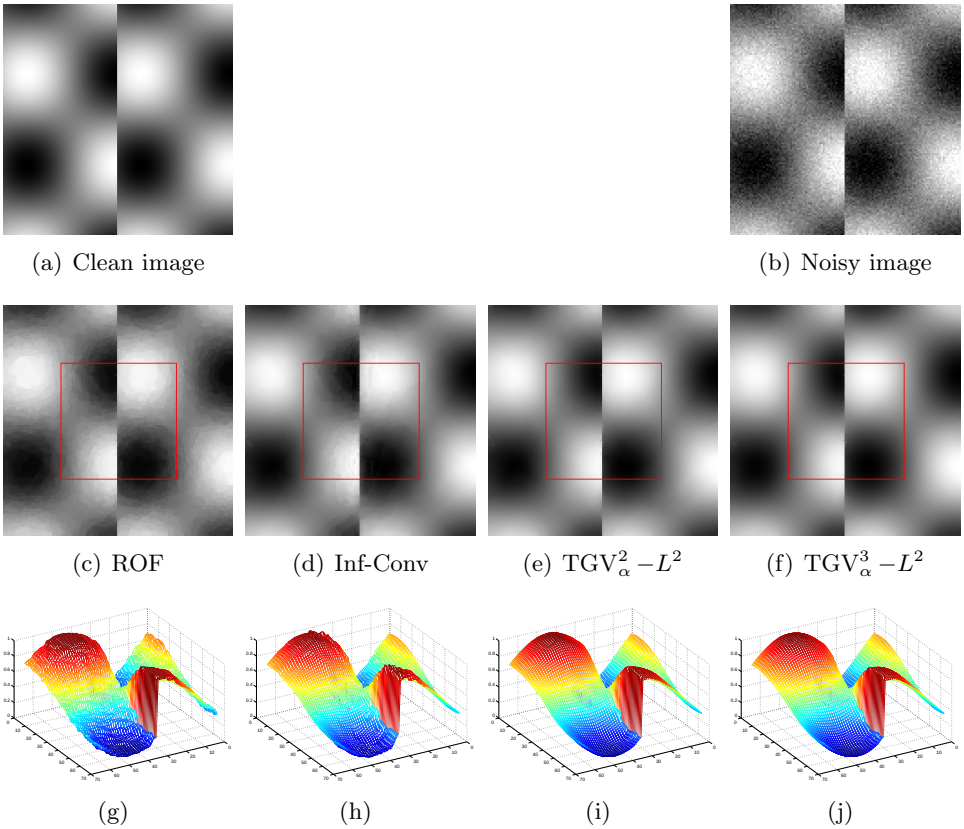


(a) Clean image (b) Noisy image

(c) ROF (d) Inf-Conv (e) $\mathrm{TGV}^2_\alpha - L^2$

(f) (g) (h)

Figure 5: Reconstruction of a piecewise affine image using different image regularization models. (a) and (b) show the $128 \times 128$ input image and the noisy version containing 5% zero mean Gaussian noise. (c)-(e) show the results of ROF, Inf-Conv and $\mathrm{TGV}^2_\alpha - L^2$ image regularization and (f)-(h) are the respective 3D closeups of the upper left region. Note that the $\mathrm{TGV}^2_\alpha - L^2$ is the only model which is able to reconstruct the piecewise affine structure of the image.

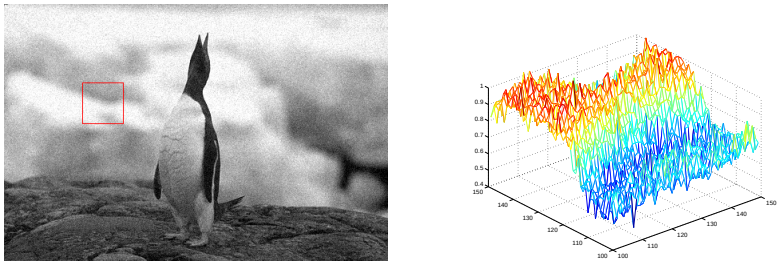


(a) Clean image (b) Noisy image

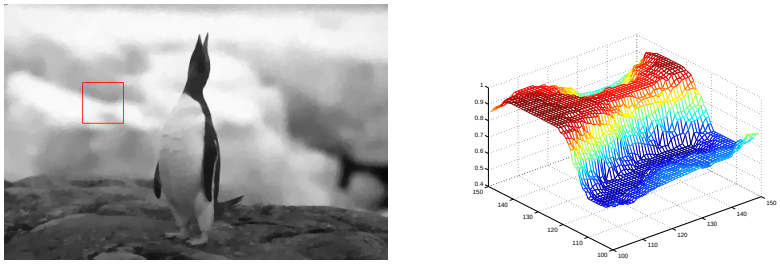
(c) ROF (d) Inf-Conv (e) $\mathrm{TGV}^2_\alpha - L^2$ (f) $\mathrm{TGV}^3_\alpha - L^2$

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

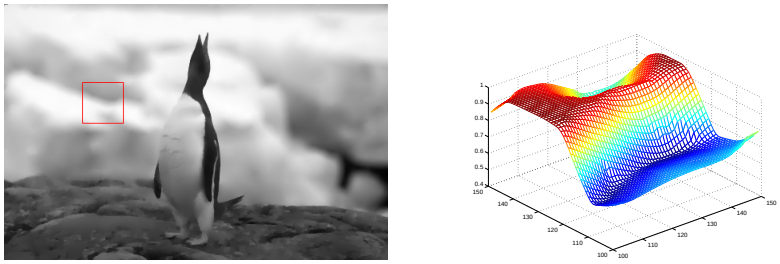
Figure 6: Reconstruction of a piecewise smooth image using different image regularization models. (a) and (b) show the $128 \times 128$ input image and the noisy version containing 5% zero mean Gaussian noise. (c)-(f) show the results of ROF, Inf-Conv, $\mathrm{TGV}^2_\alpha - L^2$ and $\mathrm{TGV}^3_\alpha - L^2$ image regularization and (g)-(j) are the respective 3D closeups 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 $\mathrm{TGV}^2_\alpha - L^2$ and $\mathrm{TGV}^3_\alpha - L^2$ models are both able to faithfully restore the image.



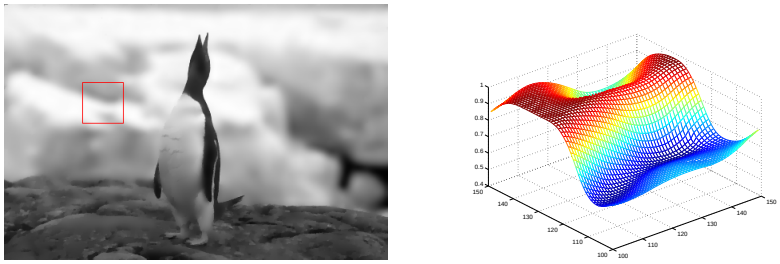
(a) Noisy Image



(b) ROF

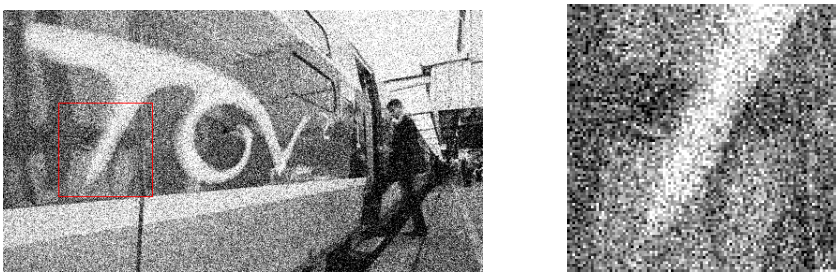


(c) $\mathrm{TGV}^2_\alpha - L^2$



(d) $\mathrm{TGV}^3_\alpha - L^2$

Figure 7: Denoising of a natural image. (a) shows the noisy input image containing 5% zero mean Gaussian noise. (b)-(d) show the results of ROF, $\mathrm{TGV}^2_\alpha - L^2$ and $\mathrm{TGV}^3_\alpha - L^2$ image denoising. The closeups of the blurry background regions show that ROF denoising leads to piecewise constant, $\mathrm{TGV}^2_\alpha - L^2$ denoising leads to piecewise affine and $\mathrm{TGV}^3_\alpha - L^2$ denoising leads to piecewise quadratic results.



(a) Noisy Image

(b) ROF

(c) $\mathrm{TGV}^2_\alpha - L^2$

(d) $\mathrm{TGV}^3_\alpha - L^2$

Figure 8: Denoising of a natural image containing severe noise. (a) shows the noisy input image containing 20% zero mean Gaussian noise. (b)-(d) show the results of ROF, $\mathrm{TGV}^2_\alpha - L^2$ and $\mathrm{TGV}^3_\alpha - L^2$ denoising. The closeups show that while ROF denoising leads to very blocky results, $\mathrm{TGV}^2_\alpha - L^2$ and $\mathrm{TGV}^3_\alpha - L^2$ models lead to smoother results.

# 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 semi-continuity. Due to the fact that $k$ is arbitrary, the framework allows 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 tensors fields. Such a restriction is natural in view of symmetry of derivatives provided that they exist. In case of non-smoothness, the proposed cost functional provides a weaker measure when compared to the non-symmetric analog.

The numerical examples show appealing properties of reconstructed test images. In particular, with the use of third order regularization, further improvements over second order is 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 $\mathrm{TGV}_\alpha^k$ and strategies for adapting the weights $\alpha$ should be the focus of further research.

# 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 lengthly computations, they are mainly presented here 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

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

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

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

Moreover, if $\xi \in \mathrm{Sym}^k(\mathbb{R}^d)$, then $\xi O \in \mathrm{Sym}^k(\mathbb{R}^d)$. Let us furthermore 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 $k \geq 2$:

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

Hence, right-multiplication respects orthonormality:

**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

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

□

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 $\xi : \Omega \to \mathcal{T}^k(\mathbb{R}^d)$ 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

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

□

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):

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

**Acknowledgements**

Kristian Bredies and Karl Kunisch acknowledge support by the SFB Research Center "Mathematical Optimization and Applications in Biomedical Sciences" at the University of Graz.

# References

[A] S.-F. Aujol, *Some First-Order Algorithms for Total Variation Based Image Restoration*, J. Math. Imaging Vision, Vol. 34, No. 3, 307–327, 2009.

[AFP] L. Ambrosio, Nicola Fusco, and Diego Pallara, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford, 2000.

[BL] J.M. Borwein and A.S. Lewis, *Convex Analysis and Nonlinear Optimization*, Canadian Mathematical Society Books in Mathematics, Springer, Berlin, 2006.

[BT] A. Beck, and M. Teboulle, *A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems*, SIAM J. Imaging Sci., Vol. 2, 183–202, 2009.

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

[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, 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., Vol. 20, No.6, 1964-1977, 1999.

[Ch] A. Chambolle, *An Algorithm for Total Variation Minimizations and Applications*, J. Math. Imaging Vision, Vol. 20, No. 1–2, 89–97, 2004.

[CL] A. Chambolle, and P.-L. Lions, *Image Recovery via Total Variation Minimization and Related Problems*, Numerische Mathematik, Vol. 76, pp. 167–188, 1997.

[CMM] T.F. Chan, A. Marquina, and P. Mulet, *Higher Order Total Variation-Based Image Restoration*, SIAM J. Sci. Comp., Vol. 22, 503–516, 2000.

[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., Vol. 40, 2351-2391, 2009.

[DZ] M. C. Delfour, and J.-P. Zolesio, *Shapes and Geometries*, SIAM, Philadelphia, 2001.

[HK] M. Hintermüller, and K. Kunisch, *Total Bounded Variation Regularization as Bilaterally Constrained Optimization Problems*, SIAM J. Appl. Mathematics, Vol. 64, 1311–1333, 2004.

[PS] C. Pöschl and O. Scherzer, *Characterization of Minimizers of Convex Regularization Functionals*, Contemporary Mathematics, Vol. 451, 219–248, 2008.

[ROF] L.I. Rudin, S. Osher, and E. Fatemi, *Nonlinear Total Variation Based Noise Removal Algorithms*, Physica D, Vol. 60, 259–268, 1992.

[SGGHL] O. Scherzer, Markus Grasmair, Harald Grossauer, Markus Haltmeier, Frank 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: M. Neamtu and L. L. Schumaker (Eds.), Approximation XII, Nashboro Press, Brentwood, 360–386, 2008.

[T] R. Temam, *Mathematical Problems in Plasticity*, Gauthier-Villars, Paris, 1985.