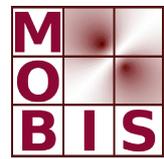
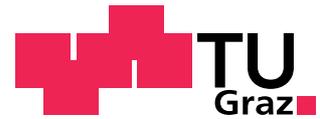




SpezialForschungsBereich F 32



Karl-Franzens Universität Graz
Technische Universität Graz
Medizinische Universität Graz



**Recovering piecewise smooth
multichannel images by
minimization of convex
functionals with total generalized
variation penalty**

Kristian Bredies

SFB-Report No. 2012-006

April 2012

A-8010 GRAZ, HEINRICHSTRASSE 36, AUSTRIA

Supported by the
Austrian Science Fund (FWF)

FWF Der Wissenschaftsfonds.

SFB sponsors:

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



Recovering piecewise smooth multichannel images by minimization of convex functionals with total generalized variation penalty

Kristian Bredies

Institute of Mathematics and Scientific Computing, University of Graz,
Heinrichstraße 36, A-8010 Graz, Austria,

`kristian.bredies@uni-graz.at`,

WWW home page: <http://www.uni-graz.at/~bredies>

Abstract. We study and extend the recently introduced total generalized variation (TGV) functional for multichannel images. This functional has already been established to constitute a well-suited convex model for piecewise smooth scalar images. It comprises exactly the functions of bounded variation but is, unlike purely total-variation based functionals, also aware of higher-order smoothness. For the multichannel version which is developed in this paper, basic properties and existence of minimizers for associated variational problems regularized with second-order TGV is shown. Furthermore, we address the design of numerical solution methods for the minimization of functionals with TGV^2 penalty and present, in particular, a class of primal-dual algorithms. Finally, the concrete realization for various image processing problems, such as image denoising, deblurring, zooming, dequantization and compressive imaging, are discussed and numerical experiments are presented.

Keywords: Total generalized variation, multichannel images, primal-dual algorithms, image denoising, image deblurring, zooming, dequantization, compressive imaging.

1 Introduction

Many imaging problems are nowadays solved by variational methods, i.e., by finding a minimizer of a functional which models the problem in terms of encouraging potential solutions of the problems by low values and penalizing unsuitable images by high values. Typically, the variational problems are cast in the form

$$\min_u F(u) + \Psi(u)$$

where data term F models the fitness of the image u with respect to some given data and the regularization functional Ψ represents an underlying image model incorporating the essential features of the sought class of images. The latter functional is responsible for the qualitative properties of the solutions, it is therefore important to choose it appropriately. As images often possess multiple

channels such as red, green, blue (RGB) or cyan, magenta, yellow, black (CMYK), such a model should also account for multichannel data. Moreover, regarding the efficient computation of numerical solutions, convexity of the objective functional is of great significance. In terms of algorithms, it is therefore favorable to consider convex models Ψ . This work is concerned with a multichannel version of the *total generalized variation* (TGV), which has been introduced, in its scalar form, in [4] and provides a well-suited convex model for piecewise smooth images. In particular, we study methods for the efficient global minimization of associated variational problems.

Let us discuss, along TGV, some existing regularization functionals for scalar images which are well-known and used in mathematical image processing. The most widely used is probably the *total variation* seminorm [28] which reads as

$$\text{TV}(u) = \int_{\Omega} d|\nabla u| = \sup \left\{ \int_{\Omega} u \operatorname{div} v \, dx \mid v \in \mathcal{C}_c^1(\Omega, \mathbf{R}^d), \|v\|_{\infty} \leq 1 \right\}.$$

where $|Du|$ denotes the variation-measure of the distributional derivative Du which is a vector-valued Radon measure. Its main feature is the incorporation of discontinuities along hypersurfaces making it a suitable model for images with edges. Indeed, solutions of variational problems with total-variation regularization admit many desirable properties, most notably the appearance of sharp edges. Unfortunately, one can also observe typical artifacts which are associated with the regularization with TV. The most prominent of these artifacts is the so-called *staircasing effect*, i.e., the undesired appearance of edges [21, 36]. This is a side-effect of the model assumption that an image consists is piecewise constant up to a discontinuity set. Natural images are, however, often piecewise smooth due to shading, for instance. Several modified models have been suggested to overcome this limitation. The most famous is the Mumford-Shah model [20] which reads as

$$\Psi_{\text{MS}}(u) = \int_{\Omega \setminus \Gamma} |\nabla u|^2 \, dx + \beta \mathcal{H}^{d-1}(\Gamma)$$

and measures piecewise smoothness on Ω up to the discontinuity set Γ whose length (or surface measure) is also penalized. This functional can be well-defined on the set of special functions of bounded variation $\text{SBV}(\Omega)$. However, Ψ_{MS} is non-convex which implies considerable analytical and practical effort when solving associated variational problems [1, 24]. As we are interested in algorithms which can efficiently and globally solve variational imaging problems, we focus, in the following, on convex problems as their stationary points are always global minimizers. An illustration of the effect for most of these models applied to the image denoising problem with L^2 -discrepancy term can be found in Figure 1.

A first approach to reduce staircasing artifacts in a convex way is to smooth out the singularity at 0 for the penalization of the gradient [35]:

$$\text{TV}_{\varepsilon}(u) = \int_{\Omega} d\varphi_{\varepsilon}(\nabla u), \quad \varphi_{\varepsilon}(t) = \sqrt{|t|^2 + \varepsilon^2} - \varepsilon.$$

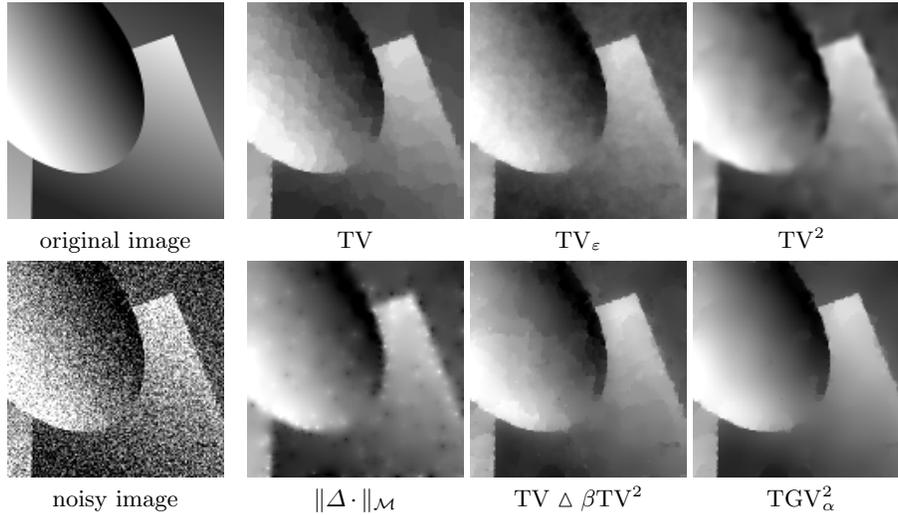


Fig. 1. Comparison of different first- and second-order image models for variational image denoising with L^2 -discrepancy. Left column: The original image (top) and noisy input image (bottom). Columns 2-4: Results for variational denoising with different regularization terms. The parameters were optimized for best PSNR.

Here, $\varphi_\varepsilon(d\nabla u)$ has to be interpreted in the sense of a function of a Radon measure. It can also be expressed as a dual functional:

$$\text{TV}_\varepsilon(u) = \sup \left\{ \int_{\Omega} u \operatorname{div} v - \varphi_\varepsilon^*(v) \, dx \mid v \in \mathcal{C}_c^1(\Omega, \mathbf{R}^d) \right\},$$

with

$$\varphi_\varepsilon^*(t) = \begin{cases} \varepsilon(1 - \sqrt{1 - |t|^2}) & \text{for } |t| < 1, \\ \infty & \text{else.} \end{cases}$$

This reduces the tendency towards piecewise constant solutions. However, as φ_ε still grows as fast as $|\cdot|$, discontinuities and consequently, the staircasing effect still appears. Such an observation can generally be made for first-order functionals penalizing the measure-valued gradient with linear growth at ∞ .

One approach to overcome these defects is to incorporate higher-order derivatives into the image model. An obvious choice is taking the total variation of second order [19, 14] which can also be expressed in a dual formulation using symmetric matrix fields $v : \Omega \rightarrow S^{d \times d}$:

$$\text{TV}^2(u) = \int_{\Omega} d|\nabla^2 u| = \sup \left\{ \int_{\Omega} u \operatorname{div}^2 v \, dx \mid u \in \mathcal{C}_c^2(\Omega, S^{d \times d}), \|v\|_\infty \leq 1 \right\}.$$

Here, the derivative ∇u is modeled to be piecewise constant which has the consequence that it is itself a regular function and can thus not have any discontinuities. Therefore, solutions of variational problems with TV^2 penalty cannot have jump discontinuities and object boundaries become inevitably blurry.

These effects cannot be overcome by changing the underlying second-order differentiation operator. For instance, taking the Laplacian

$$\|\Delta u\|_{\mathcal{M}} = \int_{\Omega} d|\Delta u| = \sup \left\{ \int_{\Omega} u \Delta v \, dx \mid v \in \mathcal{C}_c^2(\Omega), \|v\|_{\infty} \leq 1 \right\}$$

leads to a significantly weaker smoothness measure whose kernel is the set of harmonic functions on Ω which are arbitrarily smooth. Moreover, regularity theory for elliptic equations [30] tells us that each function $u \in L^1(\Omega)$ with $\Delta u \in \mathcal{M}(\Omega)$ also belongs to the Sobolev space $W_{\text{loc}}^{1,q}(\Omega)$ for each $1 \leq q < d/(d-1)$ and hence, u can also not contain jump discontinuities.

One approach to incorporate smoothness information on different scales is to combine first- and second-order derivatives. This can, for instance, be realized by considering the weighted sum of TV and TV^2 [23]. Another possibility is to interpret an image $u = u^1 + u^2$ as the sum of a piecewise constant function u^1 and piecewise smooth function u^2 . This results in *infimal convolution* models, for instance, with TV and TV^2 :

$$(\text{TV} \triangle \beta \text{TV}^2)(u) = \inf_{u=u^1+u^2} \int_{\Omega} d|\nabla u^1| + \beta \int_{\Omega} d|\nabla^2 u^2|.$$

Indeed, this model yields piecewise smooth functions [8]. However, plugging this functional into variational problems given solutions which still admit staircasing artifacts in situations where they also appear for TV-model. It seems that the tendency of TV^2 to incorporate the smoothness information of u is not strong enough such that the TV-term is still responsible for the overall impression of the solution. Again, changing the second-order term in the infimal convolution to $\|\Delta u\|_{\mathcal{M}}$, for instance [10], leads to results which are comparable to TV-TV² infimal convolution.

The *total generalized variation* model [4] can now be motivated by the dual formulation of $(\text{TV} \triangle \beta \text{TV}^2)$ which reads as

$$\begin{aligned} (\text{TV} \triangle \beta \text{TV}^2)(u) = \sup \left\{ \int_{\Omega} uw \, dx \mid v^1 \in \mathcal{C}_c^1(\Omega, \mathbf{R}^d), \|v^1\|_{\infty} \leq 1, \right. \\ \left. v^2 \in \mathcal{C}_c^2(\Omega, S^{d \times d}), \|v^2\|_{\infty} \leq \beta, \right. \\ \left. w = \text{div} v^1 = \text{div}^2 v^2 \right\}. \end{aligned}$$

The total generalized variation of second order now arises from the introduction of the additional constraint $v^1 = \text{div} v^2$:

$$\text{TGV}_{(\beta,1)}^2(u) = \sup \left\{ \int_{\Omega} u \text{div}^2 v \, dx \mid v \in \mathcal{C}_c^2(\Omega, S^{d \times d}), \|v\|_{\infty} \leq \beta, \|\text{div} v\|_{\infty} \leq 1 \right\}.$$

It is a special case of the total generalized variation of order k and positive weights $\alpha = (\alpha_0, \dots, \alpha_{k-1})$ which is defined as follows:

$$\begin{aligned} \text{TGV}_{\alpha}^k(u) = \sup \left\{ \int_{\Omega} u \text{div}^k v \, dx \mid v \in \mathcal{C}_c^k(\Omega, \text{Sym}^k(\mathbf{R}^d)), \right. \\ \left. \|\text{div}^{\kappa} v\|_{\infty} \leq \alpha_{\kappa}, \kappa = 0, \dots, k-1 \right\}. \quad (1) \end{aligned}$$

For a detailed description of the notions utilized here, we ask for the reader's patience until Section 2. This functional can be interpreted to measure smooth regions as well as jump discontinuities in a convex manner. In particular, it leads to solutions which desirable properties when used as a regularization term for variational imaging problems.

The models presented above also extend to the multichannel case. Among the several choices which are possible, the most obvious is just summing up the respective regularization functionals over each channel. In this situation, the qualitative properties of the solutions of variational problems are comparable to the scalar versions. However, there are other choices which realize coupling between the channels, e.g. for TV, see [7, 37]. Nevertheless, these models again lead to results which are similar to the scalar case in particular, the typical staircasing artifacts are also present. We can therefore expect that the qualitative properties of scalar models are generally reflected in respective multichannel versions. This motivates to define the total generalized variation functional also for multichannel data.

The aim of the present paper is, on the one hand, to shortly review as well as to extend the notion of total generalized variation to multichannel images and thus to provide a framework for color images. This is done in Section 2. Moreover, we present and discuss, in Section 3 a class of numerical methods which are easy to implement and suitable to solve general convex variational imaging problems with TGV_α^2 -penalty. These are applied in Section 4 to a variety of imaging problems: denoising, deblurring, zooming, dequantization and compressive imaging. This includes in particular a specific numerical algorithm for each of these problems. Finally, conclusions are drawn in Section 5.

2 Total generalized variation

2.1 General theory for scalar functions

Let us first review the concept of total generalized variation (1) for the scalar case as introduced [4], starting with a more detailed explanation of the notions involved in its definition.

Throughout this section, we assume that $d \in \mathbf{N}$, $d \geq 1$ is a fixed space dimension, usually, for images, we have $d = 2$. Moreover, let $\Omega \subset \mathbf{R}^d$ be a domain, i.e., a non-empty, open and connected set. We need the space of *symmetric tensors* on \mathbf{R}^d , denoted by $\text{Sym}^k(\mathbf{R}^d)$. The latter is defined, for each $k \in \mathbf{N}$, as

$$\text{Sym}^k(\mathbf{R}^d) = \left\{ \xi : \underbrace{\mathbf{R}^d \times \cdots \times \mathbf{R}^d}_{k \text{ times}} \rightarrow \mathbf{R} \mid \xi \text{ multilinear and symmetric} \right\}.$$

It is, however, convenient to identify elements $\xi \in \text{Sym}^k(\mathbf{R}^d)$ with its coefficients $\{\xi_\beta\}_{\beta \in M_k}$ where

$$M_k = \left\{ \beta \in \mathbf{N}^d \mid |\beta| = \sum_{i=1}^d \beta_i = k \right\},$$

is the set of multiindices of order k . This allows to define the spaces of compactly supported symmetric tensor fields $\mathcal{C}_c^m(\Omega, \text{Sym}^k(\mathbf{R}^d))$ for $m, k \in \mathbf{N}$. For symmetric k -tensor fields which are smooth enough, iterated divergence operators are defined componentwise by

$$(\text{div}^\kappa v)_\beta = \sum_{\gamma \in M_\kappa} \frac{\kappa!}{\gamma!} \frac{\partial^\kappa v_{\beta+\gamma}}{\partial x^\gamma} \quad \text{for each component } \beta \in M_{k-\kappa}.$$

Moreover, we define the supremum norm of a compactly supported continuous symmetric tensor field $v \in \mathcal{C}_c(\Omega, \text{Sym}^k(\mathbf{R}^d))$ as

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

which corresponds to the ∞ -norm with respect to the pointwise Frobenius norm for tensors.

With these prerequisites, TGV_α^k according to (1) makes sense for any scalar function $u \in L_{\text{loc}}^1(\Omega)$, any order $k \in \mathbf{N}$, $k \geq 1$ and any set of weights $\alpha = (\alpha_0, \dots, \alpha_{k-1})$ satisfying $\alpha_\kappa > 0$ for $\kappa = 0, \dots, k-1$.

As we will later focus on total generalized variation of second order, i.e., $k = 2$, let us elaborate on the above notions in this specific case. It turns out that TGV_α^2 can equally be written as

$$\text{TGV}_\alpha^2(u) = \sup \left\{ \int_\Omega u \text{div}^2 v \, dx \mid v \in \mathcal{C}_c^2(\Omega, S^{d \times d}), \|v\|_\infty \leq \alpha_0, \|\text{div} v\|_\infty \leq \alpha_1 \right\}$$

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

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

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

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

$$\|w\|_\infty = \sup_{x \in \Omega} \left\{ \left(\sum_{i=1}^d w_i(x)^2 \right)^{1/2} \right\}.$$

Let us now summarize some basic properties of the total generalized variation and discuss solvability of inverse problems with TGV-regularization for the second-order case.

First, observe that for $k = 1$, it follows from the definition (1) that the total generalized variation coincides, up to a factor, with the total variation, i.e., we

have $\text{TGV}_\alpha^k = \alpha_0 \text{TV}$. Hence, one can indeed speak of a generalization of the total variation.

Having defined TGV_α^k according to (1), one can show that it constitutes a proper, convex and lower semi-continuous functional on each $L^p(\Omega)$ space ($1 \leq p < \infty$) which is moreover translation and rotation invariant. The space

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

is a Banach space which is independent of the weights $\alpha_0, \dots, \alpha_{k-1}$ chosen in the definition of TGV_α^k . On this space, TGV_α^k is a semi-norm which vanishes exactly on $\mathcal{P}_{k-1}(\Omega)$, the space of polynomials of degree less than or equal to $k-1$. It can be interpreted as a model for piecewise smooth functions as follows. Let u be piecewise polynomials of maximal degree $k-1$, i.e.,

$$u = \sum_{i=1}^n \chi_{\Omega_i} q_i$$

where $\Omega_1, \dots, \Omega_n \subset \Omega$ are disjoint Lipschitz subdomains such that $\bar{\Omega} = \bigcup_{i=1}^n \bar{\Omega}_i$ and $q_i \in \mathcal{P}_{k-1}(\Omega)$ for $i = 1, \dots, n$. Then, we have that TGV_α^k is finite and measures the jump of the derivatives, from the zeroth to the $(k-1)$ -st order, of these polynomials only at the interfaces $\Gamma_{i,j} = \partial\Omega_i \cap \partial\Omega_j \cap \Omega$:

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

where $\left\| \cdot \right\|$ denotes the symmetrization of a tensor and ν_i the outer normal of Ω_i . In some cases, the estimate can be proven to be sharp. Again, see [4] for the proofs and more details.

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

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

Here, $\text{BD}(\Omega)$ denotes the space of vector fields of bounded deformation [33], i.e., the set of vector fields whose weak symmetrized derivative $\mathcal{E}(p) = \frac{1}{2}(\nabla p + \nabla p^T)$ is a matrix-valued Radon measure. Moreover, $\|\cdot\|_{\mathcal{M}}$ denotes the Radon norm for vector-valued and matrix-valued Radon measures, respectively.

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

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

We therefore have also the usual embeddings $\text{BGV}^2(\Omega) \hookrightarrow L^p(\Omega)$ for $1 \leq p \leq d/(d-1)$ which are compact for $p < d/(d-1)$. Finally, there is also a variant of the Poincaré-Friedrichs inequality available which states that for a linear

projection $P : L^{d/(d-1)}(\Omega) \rightarrow \mathcal{P}_1(\Omega)$ onto $\mathcal{P}_1(\Omega)$, we can find a constant $C > 0$, only depending on Ω , P and α such that

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

This can be used to solve the linear inverse problem

$$Ku = f$$

with TGV_α^2 -regularization. Indeed, existence of solutions for the Tikhonov functional

$$\min_{u \in L^p(\Omega)} \frac{\|Ku - f\|_H^2}{2} + \text{TGV}_\alpha^2(u)$$

where $p \in (1, \infty)$, $p \leq d/(d-1)$, $K : L^p(\Omega) \rightarrow H$ is linear, continuous and H is a Hilbert space can be shown under the assumption that K is injective on $\ker(\text{TGV}_\alpha^2) = \mathcal{P}_1(\Omega)$. Proofs and more details regarding these results can be found in [5].

2.2 Extension to multichannel images

In order to deal with, e.g., color images, it is convenient to have a notion of total generalized variation also for vector-valued images $u : \Omega \rightarrow \mathbf{R}^L$ for some $L \geq 1$. Here, we assume that color information can be encoded by a linear space, for instance \mathbf{R}^3 for the RGB or YUV color space and \mathbf{R}^4 for the CYMK color space. The space \mathbf{R}^L is then equipped with a norm $|\cdot|_\circ$ which provides a way to compare colors allowing to incorporate the characteristics of different color channels.

In order to define TGV_α^k for vector-valued images, we need to know the dual norm of $|\cdot|_\circ$ which is given, for $y \in \mathbf{R}^L$, by

$$|y|_* = \sup_{x \in \mathbf{R}^L, |x|_\circ \leq 1} x \cdot y.$$

This dual norm can now be extended to L -tuples of symmetric tensors $\xi \in \text{Sym}^k(\mathbf{R}^d)^L$ by setting

$$|\xi|_{*,k} = \|\xi\|_*, \quad \text{where} \quad |\xi| = \left(\sum_{\beta \in M_k} \frac{k!}{\beta!} \xi_\beta^2 \right)^{1/2}. \quad (3)$$

Consequently, the ∞ -norms for compactly supported symmetric tensor fields of order k read as

$$v \in \mathcal{C}_c(\Omega, \text{Sym}^k(\mathbf{R}^d)^L) : \quad \|v\|_{\infty,*,k} = \sup_{x \in \Omega} |v(x)|_{*,k}.$$

A vector-valued version of TGV_α^k for a $u \in L_{\text{loc}}^1(\Omega, \mathbf{R}^L)$ can then be defined as (abusing notation, we do not utilize a different notation)

$$\text{TGV}_\alpha^k(u) = \sup \left\{ \int_\Omega \sum_{l=1}^L u_l(\text{div}^k v_l) \, dx \mid v \in \mathcal{C}_c^k(\Omega, \text{Sym}^k(\mathbf{R}^d)^L), \right. \\ \left. \|\text{div}^\kappa v\|_{\infty,*,\kappa} \leq \alpha_\kappa, \kappa = 0, \dots, k-1 \right\}. \quad (4)$$

Example 1. An obvious choice for $\|\cdot\|_\circ$ is, of course, the Euclidean norm $\|x\|_\circ = (\sum_{l=1}^L x_l^2)^{1/2}$. It is dual to itself, so

$$\xi \in \text{Sym}^k(\mathbf{R}^d)^L : \quad |\xi|_{*,k} = \left(\sum_{l=1}^L \sum_{\beta \in M_k} \frac{k!}{\beta!} \xi_{l,\beta}^2 \right)^{1/2}$$

which is a generalization of the Frobenius norm to $\text{Sym}^k(\mathbf{R}^d)^L$ and therefore making this space a Hilbert space. The associated TGV_α^k involves, consequently, only Hilbert norms which can be exploited in numerical computations.

With more or less effort we can see that this functional possesses basically the same properties as TGV_α^k for the scalar case. As we will use them in the sequel, we highlight some of these properties. We start with basic observations.

Proposition 1. *The functional TGV_α^k is non-negative on $L_{\text{loc}}^1(\Omega, \mathbf{R}^L)$. For each $1 \leq p \leq \infty$, TGV_α^k restricted to $L^p(\Omega, \mathbf{R}^L)$ is proper, convex and lower semi-continuous.*

Proof. Note that for each $v \in \mathcal{C}_c^k(\Omega, \text{Sym}^k(\mathbf{R}^d)^L)$ which satisfies the constraints $\|\text{div}^k v\|_{\infty, \kappa} \leq \alpha_\kappa$, we also have that $-v$ satisfies the same constraints. Plugging in both v and $-v$, we see that we can replace $\int_\Omega \sum_{l=1}^L u_l \text{div}^k v_l \, dx$ by its absolute value in (4) without changing the supremum. Hence, $\text{TGV}_\alpha^k(u) \geq 0$.

To see that TGV_α^k is proper, observe that $\text{TGV}_\alpha^k(0) = 0$. Finally, for each $v \in \mathcal{C}_c^k(\Omega, \text{Sym}^k(\mathbf{R}^d)^L)$ which satisfies the constraints in (4), we have $\text{div}^k v \in \mathcal{C}_c(\Omega, \mathbf{R}^L)$ and hence, the mapping

$$u \mapsto \int_\Omega \sum_{l=1}^L u_l \text{div}^k v_l \, dx$$

is in the dual space of $L^p(\Omega, \mathbf{R}^L)$. Consequently, TGV_α^k is a pointwise supremum of convex and continuous functionals on $L^p(\Omega, \mathbf{R}^L)$ which implies the convexity and lower semi-continuity. \square

The next is the observation that each of the vector-valued TGV_α^k are equivalent in the following sense:

Proposition 2. *There are constants $0 < c < C < \infty$ such that for each $u \in L_{\text{loc}}^1(\Omega, \mathbf{R}^L)$, we have*

$$c \sum_{l=1}^L \text{TGV}_\alpha^k(u_l) \leq \text{TGV}_\alpha^k(u) \leq C \sum_{l=1}^L \text{TGV}_\alpha^k(u_l). \quad (5)$$

Proof. Denote by $|x|_1 = \sum_{l=1}^L |x_l|$ the 1-norm on \mathbf{R}^L whose dual is the ∞ -norm $|y|_\infty = \max_{l=1, \dots, L} |y_l|$. Note that by (3),

$$\|\xi\|_{\infty, k} = \max_{l=1, \dots, L} |\xi|_l \quad \text{for all } \xi \in \text{Sym}^k(\mathbf{R}^d)^L. \quad (6)$$

For a fixed k and an arbitrary norm $|\cdot|_o$ in \mathbf{R}^L , the corresponding tensor norms $|\cdot|_{*,\kappa}$ are equivalent to $|\cdot|_{\infty,\kappa}$, i.e., there exist $0 < c < C < \infty$ such that

$$\kappa = 0, \dots, k-1 \quad \text{and} \quad \xi \in \text{Sym}^\kappa(\mathbf{R}^d)^L : \quad C^{-1}|\xi|_{\infty,\kappa} \leq |\xi|_{*,\kappa} \leq c^{-1}|\xi|_{\infty,\kappa}.$$

This implies for $v \in \mathcal{C}_c^k(\Omega, \text{Sym}^k(\mathbf{R}^d)^L)$ that

$$\begin{aligned} \|\text{div}^\kappa v\|_{\infty,*,\kappa} \leq \alpha_\kappa &\quad \Rightarrow \quad \|\text{div}^\kappa v\|_{\infty,\infty,\kappa} \leq C\alpha_\kappa, \\ \|\text{div}^\kappa v\|_{\infty,\infty,\kappa} \leq c\alpha_\kappa &\quad \Rightarrow \quad \|\text{div}^\kappa v\|_{\infty,*,\kappa} \leq \alpha_\kappa. \end{aligned}$$

Denoting

$$\begin{aligned} K_\alpha^k &= \{v \in \mathcal{C}_c^k(\Omega, \text{Sym}^k(\mathbf{R}^d)) \mid \|\text{div}^\kappa v\|_\infty \leq \alpha_\kappa \text{ for } \kappa = 0, \dots, k-1\}, \\ K_{\alpha,*}^k &= \{v \in \mathcal{C}_c^k(\Omega, \text{Sym}^k(\mathbf{R}^d)^L) \mid \|\text{div}^\kappa v\|_{\infty,*,\kappa} \leq \alpha_\kappa \text{ for } \kappa = 0, \dots, k-1\} \end{aligned}$$

and $K_{\alpha,\infty}^k$ analogously to $K_{\alpha,*}^k$, the latter leads to $cK_{\alpha,\infty}^k \subset K_{\alpha,*}^k \subset CK_{\alpha,\infty}^k$ and, consequently,

$$c \sup_{v \in K_{\alpha,\infty}^k} \int_\Omega \sum_{l=1}^L u_l \text{div}^k v_l \, dx \leq \text{TGV}_\alpha^k(u) \leq C \sup_{v \in K_{\alpha,*}^k} \int_\Omega \sum_{l=1}^L u_l \text{div}^k v_l \, dx. \quad (7)$$

From (6) now follows that $K_{\alpha,\infty}^k = (K_\alpha^k)^L$, hence

$$\sup_{v \in K_{\alpha,\infty}^k} \int_\Omega \sum_{l=1}^L \int_\Omega u_l (\text{div}^k v_l) \, dx = \sum_{l=1}^L \sup_{v_l \in K_\alpha^k} \int_\Omega u_l \text{div}^k v_l \, dx = \sum_{l=1}^L \text{TGV}_\alpha^k(u_l).$$

Together with (7), this gives (5). \square

Corollary 1. *The kernel of TGV_α^k for multichannel data reads as*

$$\begin{aligned} \ker(\text{TGV}_\alpha^k) &= \{u \in L_{\text{loc}}^1(\Omega)^L \mid \text{TGV}_\alpha^k(u) = 0\} \\ &= \left\{ u(x) = \sum_{|\alpha| \leq k-1} a_\alpha x^\alpha \text{ a.e. in } \Omega \mid a_\alpha \in \mathbf{R}^L \text{ for } \alpha \in \mathbf{N}^d, |\alpha| \leq k-1 \right\} \end{aligned}$$

Proof. Observe that the norm equivalence (5) gives that $\text{TGV}_\alpha^k(u) = 0$ if and only if $\text{TGV}_\alpha^k(u_l) = 0$ for $l = 1, \dots, L$. This is in turn equivalent to $u_l(x) = \sum_{|\alpha| \leq k-1} a_{\alpha,l} x^\alpha$ a.e. in Ω for $a_{\alpha,l} \in \mathbf{R}$ and each $\alpha \in \mathbf{N}^d$ with $|\alpha| \leq k-1$ and each $l = 1, \dots, L$. Arranging each $\{a_{\alpha,l}\}$ to a vector $a_\alpha \in \mathbf{R}^L$ yields the result. \square

Remark 1. For the case $k = 2$, the statement of Corollary 1 reads as

$$\ker(\text{TGV}_\alpha^2) = \{u(x) = Ax + b \text{ a.e. in } \Omega \mid A \in \mathbf{R}^{L \times d}, b \in \mathbf{R}^L\}. \quad (8)$$

Remark 2. An even more general definition for TGV for multichannel data would arise from choosing, for each $k = 0, 1, \dots$, a norm $|\cdot|_{o,k}$ on $\text{Sym}^k(\mathbf{R}^d)^L$, setting $|\cdot|_{*,k}$ its dual norm and defining again TGV_α^k as in (4), utilizing the supremum

norms associated to $|\cdot|_{*,k}$. However, this requires, besides a model how to measure the distance of colors in the respective representation in \mathbf{R}^L , also a model for each $\text{Sym}^k(\mathbf{R}^d)^L$. In this view, setting the dual norm as in (3) seems quite natural.

Nevertheless, such a model is, for example, given by choosing $|\cdot|_{\circ}$ as above and defining the dual norm $|\cdot|_{*,k}$ for $\text{Sym}^k(\mathbf{R}^d)^L$ as

$$|\xi|_{*,k} = \sup_{|x|_{\circ} \leq 1, |y|_* \leq 1} x \cdot \underbrace{\xi(y, \dots, y)}_{k \text{ times}}$$

where each $\xi_l \in \text{Sym}^k(\mathbf{R}^d)$ is interpreted as a multilinear mapping. This can be interpreted as a generalization of the spectral norm for tensors of arbitrary order (which is also known as the least cross norm). It is, however, not clear how to treat these norms in concrete numerical implementations.

2.3 Solution of variational problems

Let us now consider variational problems for recovering multichannel images supported on the bounded Lipschitz domain $\Omega \subset \mathbf{R}^d$ which are regularized with a TGV_{α}^2 penalty. The general problem reads as

$$\min_{u \in L^p(\Omega, \mathbf{R}^L)} F(u) + \text{TGV}_{\alpha}^2(u) \quad (9)$$

where $p \in (1, \infty)$, $p \leq d/(d-1)$ and $F : L^p(\Omega, \mathbf{R}^L) \rightarrow (-\infty, \infty]$ is a proper, convex and lower semi-continuous functional which is also bounded from below.

For the following, we choose projection operators which project on the kernel of TGV_{α}^2 which is given as follows (recall (8)):

$$\{u : \Omega \rightarrow \mathbf{R}^L \mid u(x) = Ax + b \text{ a.e. for some } A \in \mathbf{R}^{L \times d}, b \in \mathbf{R}^L\}.$$

A projection operator onto the kernel is then a mapping P which satisfies

$$P : L^p(\Omega, \mathbf{R}^L) \rightarrow \ker(\text{TGV}_{\alpha}^2) \text{ linear, } P|_{\ker(\text{TGV}_{\alpha}^2)} = \text{id, } P^2 = P.$$

To obtain existence of solutions for (9), the following coercivity assumption on F is made: For any sequence $\{u^n\}$ in $L^p(\Omega, \mathbf{R}^L)$ it follows that

$$\|Pu^n\|_p \rightarrow \infty \quad \text{and} \quad \{\|(\text{id} - P)u^n\|_p\} \text{ bounded} \quad \Rightarrow \quad F(u^n) \rightarrow \infty. \quad (10)$$

Under these prerequisites, there exists at least one minimizer.

Theorem 1. *If (10) is satisfied, then (9) possesses a minimizer.*

Proof. The case where $F + \text{TGV}_{\alpha}^2$ is constant ∞ is trivial, so let $F + \text{TGV}_{\alpha}^2$ be proper. Let $\{u^n\}$ be a minimizing sequence in $L^p(\Omega)^L$, i.e., $\lim_{n \rightarrow \infty} (F + \text{TGV}_{\alpha}^2)(u^n) = \inf (F + \text{TGV}_{\alpha}^2)$. Then, $\{\text{TGV}_{\alpha}^2(u^n)\}$ has to be bounded as $\{F(u^n)\}$ is bounded from below.

Now, for each $l = 1, \dots, L$, we have that $P_l : u \mapsto (Pu)_l$ is a projection onto $\mathcal{P}_1(\Omega)$, so there exists a constant $C_1 > 0$ such that $\|u_l - P_l u\|_{d/(d-1)} \leq$

$C_1 \text{TGV}_\alpha^k(u_l)$, for each $u \in L^{d/(d-1)}(\Omega, \mathbf{R}^L)$ see (2), hence, in view of Proposition 2 it follows that

$$\begin{aligned} \|u^n - Pu^n\|_{d/(d-1)} &\leq C_2 \sum_{l=1}^L \|u_l^n - Pu_l^n\|_{d/(d-1)} \\ &\leq C_1 C_2 \sum_{l=1}^L \text{TGV}_\alpha^2(u_l^n) \leq c^{-1} C_1 C_2 \text{TGV}_\alpha^2(u^n). \end{aligned}$$

This implies, by continuous embedding $L^{d/(d-1)}(\Omega, \mathbf{R}^L) \hookrightarrow L^p(\Omega, \mathbf{R}^L)$, that $\{ \|(\text{id} - P)u^n\|_p \}$ is bounded. In addition, we can exclude that $\|Pu^n\|_p$ is unbounded: If there is an unbounded subsequence, then by restricting to that subsequence without relabeling, we can achieve that $\{u^n\}$ is still a minimizing sequence with $\|Pu^n\|_p \rightarrow \infty$ and $\|(\text{id} - P)u^n\|_p$ is bounded. By assumption, $F(u^n) \rightarrow \infty$ and $(F + \text{TGV}_\alpha^2)(u^n) \rightarrow \infty$ which is a contradiction to $\{u^n\}$ being a minimizing sequence. Hence, $\|Pu^n\|_p$ is bounded and consequently, $\{u^n\}$ as $u^n = Pu^n + (\text{id} - P)u^n$ for each n . By reflexivity of $L^p(\Omega, \mathbf{R}^L)$, a subsequence of $\{u^n\}$ converges weakly to some $u^* \in L^p(\Omega, \mathbf{R}^L)$. The sum $F + \text{TGV}_\alpha^2$ is convex and lower semi-continuous as its summands are, according to the assumptions as well as Proposition 1, which implies weak lower semi-continuity and finally

$$(F + \text{TGV}_\alpha^2)(u^*) \leq \liminf_{n \rightarrow \infty} (F + \text{TGV}_\alpha^2)(u^n) = \inf_{u \in L^p(\Omega, \mathbf{R}^L)} (F + \text{TGV}_\alpha^2)(u).$$

Therefore, u^* is the sought minimizer. \square

3 Numerical approximation and minimization

3.1 Discretization as a convex-concave saddle-point problem

For the numerical solution, we first discretize the problem (9). For simplicity, we assume that $\Omega = (0, N_1) \times (0, N_2) \subset \mathbf{R}^2$ for some positive $N_1, N_2 \in \mathbf{N}$ as it is easy to generalize to arbitrary domains and dimensions.

Following essentially the presentation in [4], we first replace Ω by the discretized grid

$$\Omega_h = \{(i, j) \mid i, j \in \mathbf{N}, 1 \leq i \leq N_1, 1 \leq j \leq N_2\}.$$

The TGV_α^2 functional will be discretized by finite differences where we also choose step-size 1, again for the sake of simplicity. For this purpose, we introduce the respective forward and backward operators which are, up to the factor -1 , adjoint to each other:

$$\begin{aligned} (\partial_x^+ u)_{i,j} &= \begin{cases} u_{i+1,j} - u_{i,j} & \text{for } 1 \leq i < N_1, \\ 0 & \text{for } i = N_1, \end{cases} \\ (\partial_y^+ u)_{i,j} &= \begin{cases} u_{i,j+1} - u_{i,j} & \text{for } 1 \leq j < N_2, \\ 0 & \text{for } j = N_2, \end{cases} \end{aligned}$$

as well as

$$\begin{aligned} (\partial_x^- u)_{i,j} &= \begin{cases} u_{1,j} & \text{if } i = 1 \\ u_{i,j} - u_{i-1,j} & \text{for } 1 < i < N_1, \\ -u_{N_1-1,j} & \text{for } i = N_1, \end{cases} \\ (\partial_y^- u)_{i,j} &= \begin{cases} u_{i,1} & \text{for } j = 1, \\ u_{i,j} - u_{i,j-1} & \text{for } 1 < j < N_2, \\ -u_{i,N_2-1} & \text{for } j = N_2. \end{cases} \end{aligned}$$

Let us further introduce the appropriate vector spaces of functions, vector and tensor fields. For $L \geq 1$, define

$$U = \{u : \Omega_h \rightarrow \mathbf{R}\}^L, \quad V = \{u : \Omega_h \rightarrow \mathbf{R}^2\}^L, \quad W = \{u : \Omega_h \rightarrow \text{Sym}^2(\mathbf{R}^2)\}^L.$$

We will denote $v = (v^1, \dots, v^L) \in V$ and its components $(v^l)^1$ and $(v^l)^2$. Likewise the components of $w = (w^1, \dots, w^L) \in W$ are $(w^l)^{11}, (w^l)^{12}$ and $(w^l)^{22}$. For convenience, we introduce

$$a, b : \Omega_h \rightarrow \mathbf{R} : \quad \langle a, b \rangle = \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} a_{i,j} b_{i,j}$$

The spaces U, V and W will be interpreted as Hilbert spaces with the scalar products

$$u, r \in U : \langle u, r \rangle_U = \sum_{l=1}^L \langle u^l, r^l \rangle,$$

$$v, p \in V : \langle v, p \rangle_V = \sum_{l=1}^L \langle (v^l)^1, (p^l)^1 \rangle + \langle (v^l)^2, (p^l)^2 \rangle,$$

$$w, q \in W : \langle w, q \rangle_W = \sum_{l=1}^L \langle (w^l)^{11}, (q^l)^{11} \rangle + \langle (w^l)^{22}, (q^l)^{22} \rangle + 2 \langle (w^l)^{12}, (q^l)^{12} \rangle.$$

The gradient, symmetrized gradient as well as the divergence operator for vector and tensor fields can then be expressed as

$$\begin{aligned} \nabla_h : U &\rightarrow V, & (\nabla_h u)^l &= \begin{pmatrix} \partial_x^+ u^l \\ \partial_y^+ u^l \end{pmatrix}, \\ \mathcal{E}_h : V &\rightarrow W, & (\mathcal{E}_h(v))^l &= \begin{pmatrix} \partial_x^-(v^l)^1 & \frac{1}{2}(\partial_y^-(v^l)^1 + \partial_x^-(v^l)^2) \\ \frac{1}{2}(\partial_y^-(v^l)^1 + \partial_x^-(v^l)^2) & \partial_y^-(v^l)^2 \end{pmatrix}, \end{aligned}$$

and

$$\begin{aligned} \text{div}_h : V &\rightarrow U, & (\text{div}_h v)^l &= \partial_x^-(v^l)^1 + \partial_y^-(v^l)^2, \\ \text{div}_h : W &\rightarrow V, & (\text{div}_h w)^l &= \begin{pmatrix} \partial_x^+(w^l)^{11} + \partial_y^+(w^l)^{12} \\ \partial_x^+(w^l)^{12} + \partial_y^+(w^l)^{22} \end{pmatrix}. \end{aligned}$$

Note that with the scalar products introduced above, it holds that $(\nabla_h)^* = -\text{div}_h$ as well as $(\mathcal{E}_h)^* = -\text{div}_h$. In order to define a discrete version of TGV_α^2 , we still need the norms

$$\begin{aligned} v \in V : \quad \|v\|_\infty &= \max_{(i,j) \in \Omega_h} \left(\sum_{l=1}^L ((v^l)_{i,j}^1)^2 + ((v^l)_{i,j}^2)^2 \right)^{1/2}, \\ w \in W : \quad \|w\|_\infty &= \max_{(i,j) \in \Omega_h} \left(\sum_{l=1}^L ((w^l)_{i,j}^{11})^2 + ((w^l)_{i,j}^{22})^2 + 2((w^l)_{i,j}^{12})^2 \right)^{1/2}. \end{aligned}$$

These norms correspond to discrete ∞ -norms with respect to the norms according to (3) where $|\cdot|_o$ is the Euclidean norm on \mathbf{R}^L , also see Example 1. With the constraint $\text{div}_h w = v$, we can now deduce a discrete version of TGV_α^2 :

$$\text{TGV}_\alpha^2(u) = \max \{ \langle u, \text{div}_h v \rangle_U \mid (v, w) \in V \times W, \text{div}_h w = v, \|w\|_\infty \leq \alpha_0, \|w\|_\infty \leq \alpha_1 \}.$$

Introducing indicator functionals, i.e.,

$$\mathcal{I}_K(x) = \begin{cases} 0 & \text{if } x \in K, \\ \infty & \text{else} \end{cases}$$

and observing that

$$-\mathcal{I}_{\{0\}}(\text{div}_h w - v) = \min_{p \in V} \langle p, \text{div}_h w - v \rangle_V,$$

the discrete functional can be rewritten to

$$\text{TGV}_\alpha^2(u) = \max_{(v,w) \in V \times W} \min_{p \in V} \langle u, \text{div}_h v \rangle_U + \langle p, \text{div}_h w - v \rangle_V - \mathcal{I}_{\{\|\cdot\|_\infty \leq \alpha_0\}}(w) - \mathcal{I}_{\{\|\cdot\|_\infty \leq \alpha_1\}}(v).$$

One can show that the maximum and minimum can be interchanged. Moreover, the constraints are symmetric around 0, so the above can be rewritten to

$$\text{TGV}_\alpha^2(u) = \min_{p \in V} \max_{(v,w) \in V \times W} \langle \nabla_h u - p, v \rangle_V + \langle \mathcal{E}_h(p), w \rangle_W - \mathcal{I}_{\{\|\cdot\|_\infty \leq \alpha_0\}}(w) - \mathcal{I}_{\{\|\cdot\|_\infty \leq \alpha_1\}}(v). \quad (11)$$

Next, assume that $F_h : \Omega_h \rightarrow (-\infty, \infty]$ is proper, convex and lower semi-continuous and corresponds to a discretized version of the data term F in (9). Then, a discretization of the variational problem (9) is given by the saddle-point problem

$$\min_{(u,p) \in U \times V} \max_{(v,w) \in V \times W} \langle \nabla_h u - p, v \rangle_V + \langle \mathcal{E}_h(p), w \rangle_W + F_h(u) - \mathcal{I}_{\{\|\cdot\|_\infty \leq \alpha_0\}}(w) - \mathcal{I}_{\{\|\cdot\|_\infty \leq \alpha_1\}}(v). \quad (12)$$

We also like to consider the situation where F_h is also given by the solution of a certain maximization problem:

$$F_h(u) = \max_{\lambda \in \Lambda} \langle Ku, \lambda \rangle_\Lambda + \tilde{F}_h(u) - G_h(\lambda) \quad (13)$$

where Λ is a finite-dimensional Hilbert space, $K : U \rightarrow \Lambda$ a linear mapping and $\tilde{F}_h : U \rightarrow (-\infty, \infty]$, $G_h : \Lambda \rightarrow (-\infty, \infty]$ are proper, convex and lower semi-continuous functionals. In this case, we like to solve the saddle-point problem

$$\min_{(u,p) \in U \times V} \max_{(v,w,\lambda) \in V \times W \times \Lambda} \langle \nabla_h u - p, v \rangle_V + \langle \mathcal{E}_h(p), w \rangle_W + \langle Ku, \lambda \rangle_\Lambda + \tilde{F}_h(u) - \mathcal{I}_{\{\|\cdot\|_\infty \leq \alpha_0\}}(w) - \mathcal{I}_{\{\|\cdot\|_\infty \leq \alpha_1\}}(v) - G_h(\lambda). \quad (14)$$

3.2 A numerical algorithm

For the solution of (12) and (14), any kind of numerical algorithm for the solution of convex-concave saddle point problems can be used. Here, we chose to employ the primal-dual ascent-descent method with primal extragradient according to [9]. The main reason is its applicability for a wide range of problems as we will see in Section 4. However, for the solution of specialized problems, other algorithms might be suited and efficient as well. Basically, every convergent method which finds a zero of a maximal monotone operator or the sum to two maximally monotone operators may work [27, 18, 13]. In its general form, the method finds a saddle point for the problem

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \langle \mathcal{K}x, y \rangle_{\mathcal{Y}} + \mathcal{F}(x) - \mathcal{G}(y) \quad (15)$$

where \mathcal{X}, \mathcal{Y} are Hilbert spaces, $\mathcal{K} : \mathcal{X} \rightarrow \mathcal{Y}$ is a linear and continuous mapping, and $\mathcal{F} : \mathcal{X} \rightarrow (-\infty, \infty]$, $\mathcal{G} : \mathcal{Y} \rightarrow (-\infty, \infty]$ are proper, convex and lower semi-continuous functionals.

In order to state the algorithm, we need the notion of resolvent operators $(\text{id} + \tau \partial \mathcal{F})^{-1}$ and $(\text{id} + \sigma \partial \mathcal{G})^{-1}$ for the subgradients of \mathcal{F} and \mathcal{G} , respectively. They can be characterized as the solutions of

$$\begin{aligned} x^* = (\text{id} + \tau \partial \mathcal{F})^{-1}(\bar{x}) &\Leftrightarrow x^* = \arg \min_{x \in \mathcal{X}} \frac{\|x - \bar{x}\|_{\mathcal{X}}^2}{2} + \tau \mathcal{F}(x), \\ y^* = (\text{id} + \sigma \partial \mathcal{G})^{-1}(\bar{y}) &\Leftrightarrow y^* = \arg \min_{y \in \mathcal{Y}} \frac{\|y - \bar{y}\|_{\mathcal{Y}}^2}{2} + \sigma \mathcal{G}(y) \end{aligned}$$

where $\sigma, \tau > 0$. These resolvent operators are assumed to be computationally accessible. We will discuss some examples in Section 4.

The iteration procedure for the solution of (15) can be written as follows. Choose $\sigma, \tau > 0$ such that $\sigma\tau\|\mathcal{K}\|^2 < 1$. For initial values $(x^0, y^0) \in \mathcal{X} \times \mathcal{Y}$ and $\bar{x}^0 = x^0$, the iteration reads as

$$\begin{cases} y^{n+1} = (\text{id} + \sigma \partial \mathcal{G})^{-1}(y^n + \sigma \mathcal{K} \bar{x}^n), \\ x^{n+1} = (\text{id} + \tau \partial \mathcal{F})^{-1}(x^n - \tau \mathcal{K}^* y^{n+1}), \\ \bar{x}^{n+1} = 2x^{n+1} - x^n. \end{cases} \quad (16)$$

If \mathcal{X} and \mathcal{Y} are finite dimensional, this algorithm is known [9] to converge to a saddle point (x^*, y^*) of (15) provided that a saddle point exists.

We like to apply this algorithm for the solution of (12) and (14). First, let us address the problem (12) which admits the structure (15) if one chooses

$$\mathcal{X} = U \times V, \quad \mathcal{Y} = V \times W, \quad \mathcal{K} = \begin{bmatrix} \nabla_h & -\text{id} \\ 0 & \mathcal{E}_h \end{bmatrix} \Rightarrow \mathcal{K}^* = \begin{bmatrix} -\text{div}_h & 0 \\ -\text{id} & -\text{div}_h \end{bmatrix}$$

as well as

$$\begin{aligned} \mathcal{F}(x) &= \mathcal{F}(u, p) = F_h(u), \\ \mathcal{G}(y) &= \mathcal{G}(v, w) = \mathcal{I}_{\{\|\cdot\|_\infty \leq \alpha_1\}}(v) + \mathcal{I}_{\{\|\cdot\|_\infty \leq \alpha_0\}}(w). \end{aligned}$$

As the functionals \mathcal{F} and \mathcal{G} are the sum of functionals which only depend on one component of x and y , respectively, the resolvent operators decouple meaning that they can be performed componentwise. For \mathcal{G} , they correspond to projection operators on the respective constraint sets. They can be seen to correspond to

$$\begin{aligned} v^* &= \arg \min_{\|v\|_\infty \leq \alpha_1} \frac{\|v - \bar{v}\|_V^2}{2} & \Leftrightarrow & \quad v^* = \mathcal{P}_{\alpha_1}(\bar{v}) = \frac{\bar{v}}{\max(1, \frac{|\bar{v}|}{\alpha_1})}, \\ w^* &= \arg \min_{\|w\|_\infty \leq \alpha_0} \frac{\|w - \bar{w}\|_V^2}{2} & \Leftrightarrow & \quad w^* = \mathcal{P}_{\alpha_0}(\bar{w}) = \frac{\bar{w}}{\max(1, \frac{|\bar{w}|}{\alpha_0})} \end{aligned}$$

where the operations on the right-hand side have to be interpreted in the pointwise sense with $|\bar{v}|$ and $|\bar{w}|$ according to

$$\begin{aligned} \bar{v} \in V : \quad |\bar{v}|_{i,j} &= \left(\sum_{l=1}^L ((\bar{v}^l)_{i,j}^1)^2 + ((\bar{v}^l)_{i,j}^2)^2 \right)^{1/2}, \\ \bar{w} \in W : \quad |\bar{w}|_{i,j} &= \left(\sum_{l=1}^L ((\bar{w}^l)_{i,j}^{11})^2 + ((\bar{w}^l)_{i,j}^{22})^2 + 2((\bar{w}^l)_{i,j}^{12})^2 \right)^{1/2}. \end{aligned}$$

For \mathcal{F} , the componentwise resolvents just correspond to $(\text{id} + \sigma \partial F_h)^{-1}$ and, as the functional is independent of p , to the identity on V , respectively.

Finally, in order to choose the step sizes σ and τ , we need an estimate for the norm of \mathcal{K} . One can see that $\|\nabla_h\|^2 < 8$ and $\|\mathcal{E}_h\|^2 < 8$ which leads, after some computations, to the estimate

$$\|\mathcal{K}\|^2 < \frac{17 + \sqrt{33}}{2} < 12.$$

Hence, the primal-dual method for the saddle-point problem (12) reads as follows.

Algorithm 1.

Objective: Solve $\min_{u \in U} F_h(u) + \text{TGV}_\alpha^2(u)$

1. Choose $\sigma > 0, \tau > 0$ such that $\sigma\tau\frac{1}{2}(17 + \sqrt{33}) \leq 1$.
2. Choose $(u^0, p^0) \in U \times V, (v^0, w^0) \in V \times W$ and set $\bar{u}^0 = u^0, \bar{p}^0 = p^0$.
3. For $n = 0, 1, 2, \dots$ iterate according to

$$\begin{cases} v^{n+1} = \mathcal{P}_{\alpha_1}(v^n + \sigma(\nabla_h \bar{u}^n - \bar{p}^n)), \\ w^{n+1} = \mathcal{P}_{\alpha_0}(w^n + \sigma \mathcal{E}_h(\bar{p}^n)), \\ u^{n+1} = (\text{id} + \tau \partial F_h)^{-1}(u^n + \tau \text{div}_h v^{n+1}), \\ p^{n+1} = p^n + \tau(v^{n+1} + \text{div}_h w^{n+1}), \\ \bar{u}^{n+1} = 2u^{n+1} - u^n, \\ \bar{p}^{n+1} = 2p^{n+1} - p^n. \end{cases}$$

4. Return u^N for some large N .

Note that the problem of choosing N , i.e., finding an appropriate stopping criterion, remains. However, as our major goal is to demonstrate the applicability and efficiency of algorithms suitable for the minimization of TGV_α^2 , we do not discuss this issue here. Let us nevertheless remark that it is possible to compute estimates for the primal-dual gap for the underlying saddle-point problem which allows to estimate the distance of the current iterate to the minimizer in terms of the functional values. These estimates could be used to implement a stopping criterion, see, for instance [3].

It can be seen in Algorithm 1 that the resolvent $(\text{id} + \tau \partial F_h)^{-1}$ is needed in order to perform the computational procedure. In some cases, this resolvent is not computationally accessible or expensive to compute. It might, however, be possible to write F_h in terms of (13) where the resolvents $(\text{id} + \tau \partial \tilde{F}_h)^{-1}$ and $(\text{id} + \sigma \partial G_h)^{-1}$ are easy to compute. In such a case, the algorithm can be modified in order to accommodate for this situation: Indeed, the associated saddle-point problem (14) can be represented by (15) if one chooses

$$\mathcal{X} = U \times V, \quad \mathcal{Y} = V \times W \times \Lambda, \quad \mathcal{K} = \begin{bmatrix} \nabla_h & -\text{id} \\ 0 & \mathcal{E}_h \\ K & 0 \end{bmatrix} \Rightarrow \mathcal{K}^* = \begin{bmatrix} -\text{div}_h & 0 & K^* \\ -\text{id} & -\text{div}_h & 0 \end{bmatrix}$$

as well as

$$\begin{aligned} \mathcal{F}(x) &= \mathcal{F}(u, p) = \tilde{F}_h(u), \\ \mathcal{G}(y) &= \mathcal{G}(v, w, \lambda) = \mathcal{I}_{\{\|\cdot\|_\infty \leq \alpha_1\}}(v) + \mathcal{I}_{\{\|\cdot\|_\infty \leq \alpha_0\}}(w) + G_h(\lambda). \end{aligned}$$

As in the previous case, all the resolvents decouple and each component of x and y can be updated individually. However, one has to be more careful when choosing σ and τ as the norm of \mathcal{K} can only be guaranteed to obey

$$\|\mathcal{K}\|^2 < \frac{\sqrt{(\|K\|^2 - 1)^2 + 32} + \|K\|^2 + 17}{2} < \|K\|^2 + 12.$$

The primal-dual method then corresponds to the following.

Algorithm 2.

$$\text{Objective:} \quad \text{Solve} \quad \begin{cases} \min_{u \in U} F_h(u) + \text{TGV}_\alpha^2(u), \\ F_h(u) = \max_{\lambda \in \Lambda} \langle Ku, \lambda \rangle_\Lambda + \tilde{F}_h(u) - G_h(\lambda) \end{cases}$$

1. Choose $\sigma > 0, \tau > 0$ such that $\sigma\tau\frac{1}{2}(\sqrt{(\|K\|^2 - 1)^2 + 32} + \|K\|^2 + 17) \leq 1$.
2. Choose $(u^0, p^0) \in U \times V, (v^0, w^0, \lambda^0) \in V \times W \times \Lambda$.
Set $\bar{u}^0 = u^0, \bar{p}^0 = p^0$.
3. For $n = 0, 1, 2, \dots$ iterate according to

$$\begin{cases} v^{n+1} = \mathcal{P}_{\alpha_1}(v^n + \sigma(\nabla_h \bar{u}^n - \bar{p}^n)), \\ w^{n+1} = \mathcal{P}_{\alpha_0}(w^n + \sigma \mathcal{E}_h(\bar{p}^n)), \\ \lambda^{n+1} = (\text{id} + \sigma \partial G_h)^{-1}(\lambda^n + \sigma K \bar{u}^n) \\ u^{n+1} = (\text{id} + \tau \partial \tilde{F}_h)^{-1}(u^n + \tau(\text{div}_h v^{n+1} - K^* \lambda^{n+1})), \\ p^{n+1} = p^n + \tau(v^{n+1} + \text{div}_h w^{n+1}), \\ \bar{u}^{n+1} = 2u^{n+1} - u^n, \\ \bar{p}^{n+1} = 2p^{n+1} - p^n. \end{cases}$$

4. Return u^N for some large N .

Again, the procedure converges to a saddle-point, so u^N for N large enough is close to a solution of the original problem.

4 Application to mathematical imaging problems

Now, we aim at applying the total generalized variation model to some well-known variational problems. On the one hand, we show how existence in the continuous setting can be ensured using the results of Section 2. On the other hand, it is also discussed how the algorithms in Section 3 can be realized and how they perform in numerical experiments.

4.1 Denoising

We first look at the TGV_α^2 -regularized multichannel denoising problem for a noisy image $f \in L^q(\Omega, \mathbf{R}^L)$ where $q \in [1, \infty)$. Let the norm in $L^q(\Omega, \mathbf{R}^L)$ be based on the vector norm $|\cdot|_\circ$. Then, the variational denoising problem with L^q -data term, is to solve

$$\min_{u \in L^q(\Omega, \mathbf{R}^L)} F(u) + \text{TGV}_\alpha^2(u), \quad F(u) = \frac{1}{q} \int_\Omega |u - f|_\circ^q \, dx = \frac{\|u - f\|_q^q}{q}. \quad (17)$$

We like to verify existence of a minimizer in $L^p(\Omega, \mathbf{R}^L)$ for some $p \in (1, \infty)$ with $p \leq d/(d-1)$. For this purpose, observe that it is easy to see that F is non-negative, proper, convex and lower semi-continuous on $L^p(\Omega, \mathbf{R}^L)$, the latter

with the help of Fatou's lemma. To establish property (10), set $r = \min(p, q)$ and choose $P : L^r(\Omega, \mathbf{R}^L) \rightarrow \ker(\text{TGV}_\alpha^2)$ as a linear and continuous projection. If, for a sequence $\{u^n\}$ in $L^p(\Omega, \mathbf{R}^L)$ it holds that $\|Pu^n\|_p \rightarrow \infty$, then also $\|Pu^n\|_q \rightarrow \infty$ since all norms are equivalent on the finite-dimensional space $\ker(\text{TGV}_\alpha^2)$. Consequently, as P is continuous on $L^q(\Omega, \mathbf{R}^L)$,

$$\|u^n - f\|_q \geq \|u^n\|_q - \|f\|_q \geq c\|Pu^n\|_q - M$$

for some constants $c > 0$ and $M > 0$. Hence, $F(u^n) \rightarrow \infty$ and (10) is satisfied. By Theorem 1, there exists a minimizer.

Let us now discretize (17) according to Section 3. For this purpose, we choose F_h for some data $f \in U$ according to

$$F_h(u) = \frac{1}{q} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} |u_{i,j} - f_{i,j}|_o^q$$

which is in accordance with a discretization step-size of 1. We like to use Algorithm 1 which needs the resolvent $(\text{id} + \tau \partial F_h)^{-1}$. For $|\cdot|_o = |\cdot|$ the Euclidean norm on \mathbf{R}^L and $q \in \{1, 2\}$, this operator can be computed:

$$u^* = (\text{id} + \tau \partial F_h)^{-1}(\bar{u}) \quad \Leftrightarrow \quad u^* = \begin{cases} \bar{u} + \tau f & \text{if } q = 2, \\ \frac{\bar{u} + \tau f}{1 + \tau} & \text{if } q = 1. \end{cases}$$

where \mathcal{S}_τ is the pointwise shrinkage operator

$$\mathcal{S}_\tau(u) = \frac{u}{|u|} \max(0, |u| - \tau).$$

This leads to the following iteration for $\{u^n\}$ in Algorithm 1:

$$u^{n+1} = \begin{cases} \frac{u^n + \tau(\text{div}_h v^{n+1} + f)}{1 + \tau} & \text{if } q = 2, \\ f + \mathcal{S}_\tau(u^n + \tau \text{div}_h v^n - f) & \text{if } q = 1. \end{cases}$$

The resulting algorithm was implemented in PYTHON [25] using Scientific Tools for Python (SCIPY) [34] and graphics-processing unit (GPU) acceleration based on NVIDIA's CUDATM Toolkit [22] via the PYTHON interface PYCUDA [16]. Computations were performed on a AMD PhenomTM 9950 Quad-Core Processor with a NVIDIA GeForce GTX 280 GPU with 1 Gigabyte of memory. The outcome of the TGV²-based denoising procedure, for the L^2 -norm and L^1 -norm as well as a comparison to the standard TV-based counterparts, are depicted in the Figures 2 and 3, respectively. In order to compare the models, the parameters were chosen to give the best peak signal-to-noise ratio (PSNR). The actual values as well as the image size, noise level, iteration count and computation time can be found in the respective captions. One can observe that the multichannel TGV² image model is able to recover smooth regions as well as discontinuities at object boundaries. In particular, artifacts which are typical for TV-based

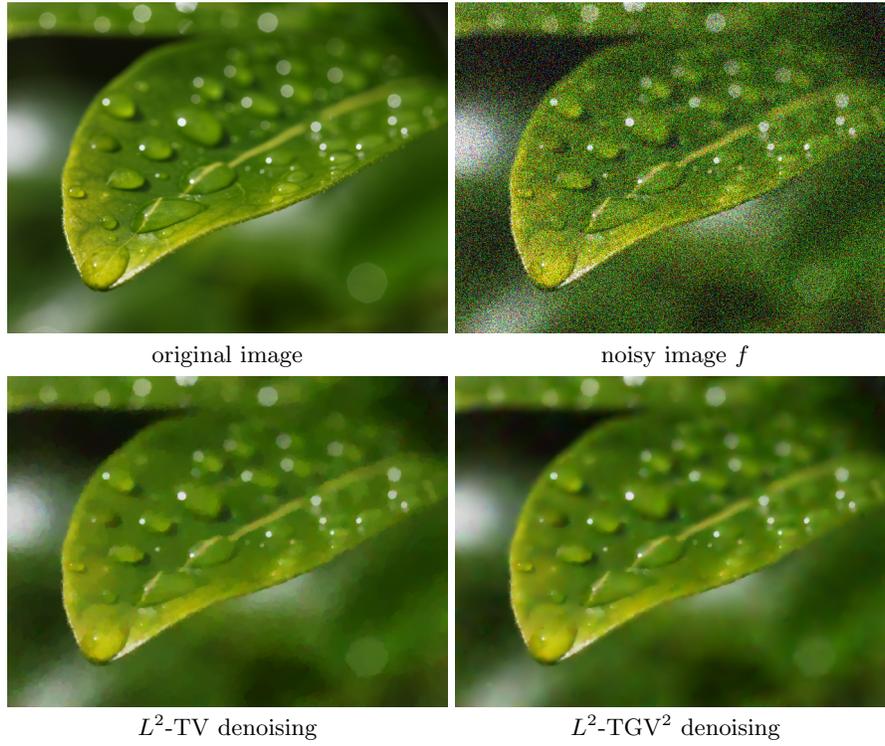


Fig. 2. Example for variational denoising according to (17) with L^2 data fitting term and TV/TGV² image model. Top: The original image [31] (left, 640×480 pixels, RGB) and a noise-perturbed version (right, additive Gaussian noise, standard deviation $\sigma = 0.2$). Bottom: Result of TV-based denoising (left, PSNR=31.91 dB, 500 iterations, computation time: 0.48 seconds), and TGV²-based denoising (right, PSNR=32.29 dB, 500 iterations, computation time: 1.29 seconds). Images licenced under CreativeCommons-by-sa-2.0 (<http://creativecommons.org/licenses/by-sa/2.0/>).

denoising, do not appear when TGV² is used. The improved image quality is also reflected by a slightly higher PSNR value. However, since this distance measure is essentially based on pointwise comparison and is not incorporating neighborhood information, more accurate recovery of smooth regions does not lead to an significantly higher PSNR, although the differences can noticeably be visually perceived. Of course, usage of the TGV² image model come with higher computational effort. In the case of denoising, TV and TGV² roughly need the same number of iterations such that TGV² needs about 2 to 3 times more computation time. The absolute computation time is, however, still quite low thanks to the parallelization provided by the GPU.

Remark 3. In order to preserve more details, one approach is to introduce a spatially dependent parameter which serves as a weight for the L^2 -discrepancy,

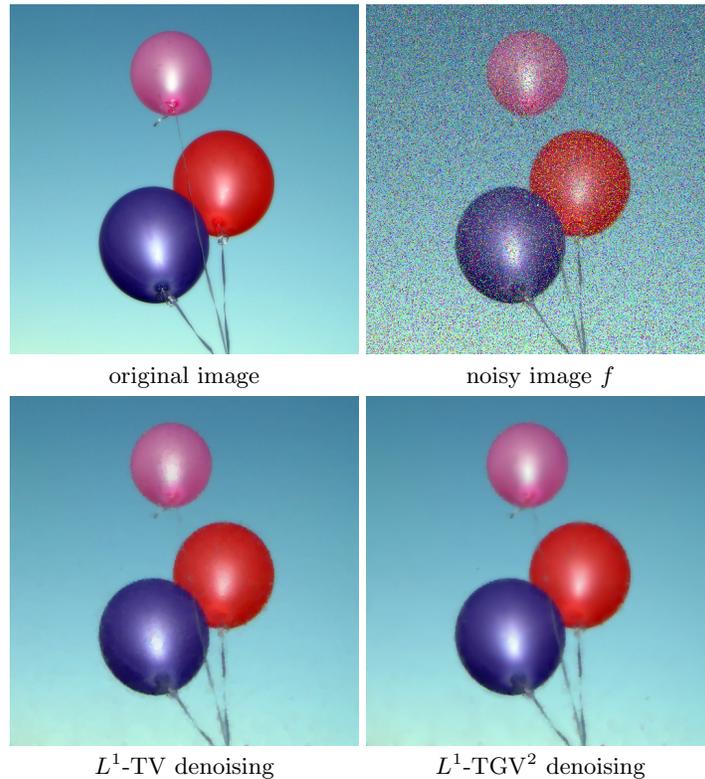


Fig. 3. Example for variational denoising with L^1 data fitting term and TV/TGV² image model. Top: The original image [29] (left, 512×512 pixels, RGB) and a noise-perturbed version (right, 33.3% of the pixels replaced by random values). Bottom: Result of TV-based denoising (left, PSNR=33.23 dB, 1000 iterations, computation time: 0.98 seconds), and TGV²-based denoising (right, PSNR=33.77 dB, 1000 iterations, computation time: 2.46 seconds).

i.e.,

$$F(u) = \frac{1}{2} \int_{\Omega} \lambda |u - f|^2 dx, \quad \lambda \in L^{\infty}(\Omega).$$

An appropriate choice of λ then leads, in conjunction with TV-regularization, to denoised images which indeed admit more details [11]. Recently, this framework has been extended to TGV _{α} ²-regularization which yields further improvements, see [2] for details.

4.2 Deblurring

Next, let us discuss the deblurring of a multichannel image. We model this problem as the general linear inverse problem of finding a solution to

$$Ku = f$$

where, for some $p \in (1, \infty)$, $p \leq d/(d-1)$ the operator $K : L^p(\Omega, \mathbf{R}^L) \rightarrow H$ is a linear and continuous mapping into a Hilbert space H in which the data $f \in H$ is also given. We like to regularize this generally ill-posed problem with TGV_α^2 and solve the associated Tikhonov minimization problem, i.e.,

$$\min_{u \in L^p(\Omega, \mathbf{R}^L)} F(u) + \text{TGV}_\alpha^2(u), \quad F(u) = \frac{\|Ku - f\|_H^2}{2}. \quad (18)$$

This problem turns out to have a solution as soon as K is injective on $\ker(\text{TGV}_\alpha^2)$: Let us assume that

$$Ku = 0 \quad \text{for some } u \in \ker(\text{TGV}_\alpha^2) \quad \Rightarrow \quad u = 0,$$

which is, as $\ker(\text{TGV}_\alpha^2)$ is finite-dimensional, equivalent to the existence of a $c > 0$ such that $\|Ku\|_H \geq c\|u\|_p$ for each $u \in \ker(\text{TGV}_\alpha^2)$. For an arbitrary projection operator $P : L^p(\Omega, \mathbf{R}^L) \rightarrow \ker(\text{TGV}_\alpha^2)$ and a sequence $\{u^n\}$ in $L^p(\Omega, \mathbf{R}^L)$ such that $\|Pu^n\|_p \rightarrow \infty$ and $\{\|(id - P)u^n\|_p\}$ is bounded, we then have

$$\|Ku - f\|_H \geq \|KPu^n\|_H - \|K(id - P)u^n\|_H - \|f\|_H \geq c\|Pu^n\|_H - M$$

for some $M > 0$ and the above $c > 0$. Hence $F(u^n) \rightarrow \infty$ as $n \rightarrow \infty$. Together with the observation that F is non-negative, proper, convex, continuous (and hence, lower semi-continuous), this implies by Theorem 1 that there is always a solution to (18).

In the concrete case of deblurring, we model the forward operator K as the convolution with a kernel $k \in L^\infty(\mathbf{R}^d)$ with compact support Ω_0 . The data space is $H = L^2(\Omega', \mathbf{R}^L)$ where Ω' is a non-empty open set which models the region on which the blurred image is measured. We assume that only data on Ω is convolved which is satisfied if

$$\Omega' - \Omega_0 = \{x - y \mid x \in \Omega', y \in \Omega_0\} \subset \Omega.$$

Furthermore, let $k \geq 0$ almost everywhere and such that $\int_{\Omega_0} k \, dx = 1$. The operator K is then given by

$$Ku = u * k, \quad (u * k)(x) = \int_{\Omega_0} u(x - y)k(y) \, dy \quad \text{for } x \in \Omega'. \quad (19)$$

Note that $\|u * k\|_\infty \leq \|u\|_1 \|k\|_\infty$, so K is in particular continuous between $L^p(\Omega, \mathbf{R}^L)$ and H . It remains to verify that K is injective on $\ker(\text{TGV}_\alpha^2)$. For this purpose, let $u(x) = Ax + b$ for $A \in \mathbf{R}^{L \times d}$, $b \in \mathbf{R}^L$ such that $Ku = 0$. Then, for $m \in \mathbf{R}^d$, $m_i = \int_{\Omega_0} y_i \, dy$, we have

$$\int_{\Omega_0} (A(x - y) + b)k(y) \, dy = Ax + (b - Am) \quad \text{for all } x \in \Omega'$$

meaning that Ku is an affine linear function on Ω' . As Ω' contains a non-empty open set, $Ku = 0$ is only possible if $A = 0$ and $b - Am = b = 0$, implying that

$u = 0$. This shows the injectivity, hence (18) can always be solved for the blurring operator according to (19).

Let us now discuss the numerical realization of the solution of (18) in the framework of Section 3. Regarding the general problem, we assume that K can be discretized to a $K_h : U \rightarrow A$ where the Hilbert space A corresponds to the discretized data space H . The discrepancy functional for discrete data $f \in A$ then reads as

$$F_h(u) = \frac{\|K_h u - f\|_A^2}{2}.$$

To describe A and K_h for the blurring operator (19), let $k \in \mathbf{R}^{(2M+1) \times (2M+1)}$ a discrete convolution kernel which is indexed through $-M, \dots, M$. The data can then be measured on the set

$$\Omega'_h = \{(i, j) \mid i, j \in \mathbf{N}, M+1 \leq i \leq N_1 - M, M+1 \leq j \leq N_2 - M\}.$$

Consequently, we let $A = \{\Omega_h \rightarrow \mathbf{R}\}^L$ such that the discrete convolution operator becomes

$$K_h u = u * k, \quad (u * k)_{i,j}^l = \sum_{i'=-M}^M \sum_{j'=-M}^M u_{i-i', j-j'}^l k_{i',j'} \quad \text{for } (i, j) \in \Omega'_h.$$

One can easily see that if k is normalized, i.e., $k \geq 0$ componentwise and $\sum_{i=-M}^M \sum_{j=-M}^M k_{i,j} = 1$, then $\|K_h\| \leq 1$.

There is now the choice whether to take Algorithm 1 or 2 for the numerical solution. Let us shortly discuss Algorithm 1. Here, one has again to evaluate the resolvent operator which corresponds to

$$u^* = (\text{id} + \tau \partial F_h)^{-1}(\bar{u}) \quad \Leftrightarrow \quad u^* = (\text{id} + \tau K_h^* K_h)^{-1}(\bar{u} + \tau K_h^* f).$$

Hence, the iteration step for u^{n+1} reads as

$$u^{n+1} = (\text{id} + \tau K_h^* K_h)^{-1}(u^n + \tau(\text{div}_h v^{n+1} + K_h^* f))$$

which involves the solution of a linear equation. As this might be quite costly, in particular if it has to be done iteratively and the evaluation of K_h or K_h^* is expensive, we also discuss Algorithm 2 which, as it turns out, does not involve such an inversion step.

It bases on the observation that F_h can be written as

$$F_h(u) = \max_{\lambda \in A} \langle K_h u, \lambda \rangle_A - \left(\frac{\|\lambda\|_A^2}{2} + \langle f, \lambda \rangle_A \right)$$

which is of the form (13) with $\tilde{F}_h(u) = 0$ and $G_h(\lambda) = \frac{1}{2}\|\lambda\|_A^2 + \langle f, \lambda \rangle_A$. The resolvent associated with the subgradient of \tilde{F}_h again turns out to be the identity while

$$\lambda^* = (\text{id} + \sigma \partial G_h)^{-1}(\bar{\lambda}) \quad \Leftrightarrow \quad \lambda^* = \frac{\bar{\lambda} - \sigma f}{1 + \sigma}.$$

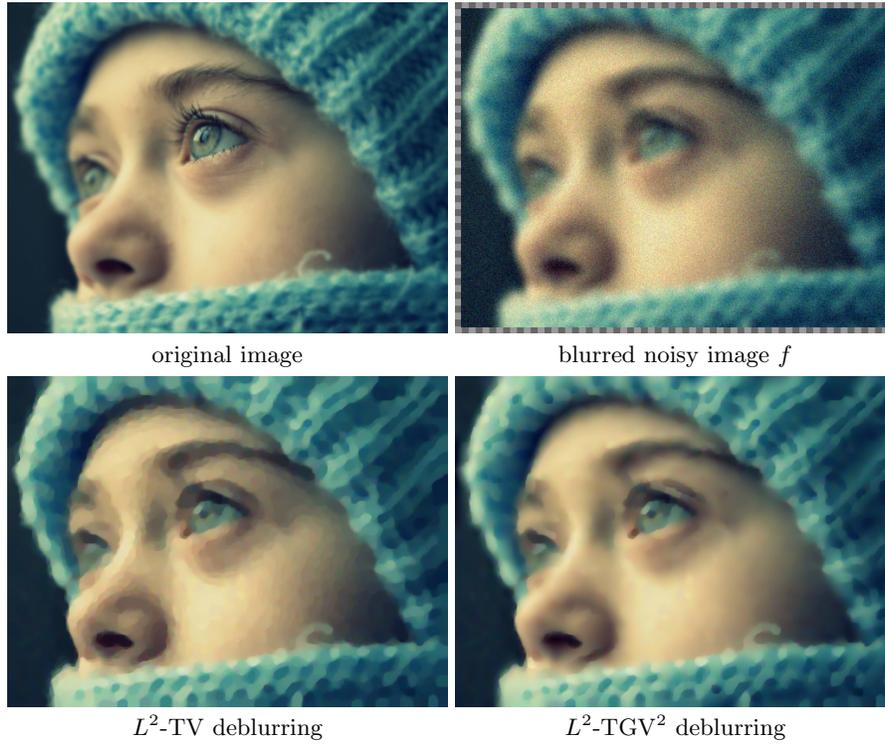


Fig. 4. Example for variational deblurring according to (18) with L^2 data fitting term and TV/TGV² image model. Top: The original image [15] (left, 512×384 pixels, RGB) and a blurred, noise-perturbed version (right, out-of-focus kernel with 15 pixels diameter, additive Gaussian noise, standard deviation $\sigma = 0.05$). Bottom: Result of TV-based deblurring (left, PSNR=31.45 dB, 1000 iterations, computation time: 45.63 seconds), and TGV²-based deblurring (right, PSNR=32.05 dB, 1000 iterations, computation time: 46.72 seconds).

Hence, the iteration steps for λ and u in Algorithm 2 read as

$$\begin{cases} \lambda^{n+1} = \frac{\lambda^n + \sigma(K_h \bar{u}^n - f)}{1 + \sigma}, \\ u^{n+1} = u^n + \tau(\operatorname{div}_h v^{n+1} - K_h^* \lambda^{n+1}). \end{cases}$$

This variant provides an alternative in which only one evaluation of K_h and K_h^* is necessary in each iteration step. However, one needs to have an estimate for $\|K_h\|$ in order to choose step-sizes σ and τ such that convergence can be ensured. In the case of the discrete convolution operator introduced above, one obtains again the estimate $\|K_h\| \leq 1$ for normalized kernels.

As in Subsection 4.1, this method has again been implemented in PYTHON using PYCUDA. Computations have been performed to deconvolve a blurred image which has additionally been contaminated by noise. The same configuration

as for the denoising experiments has been used, the outcome as well as the details are shown in Figure 4. Again, one can observe the ability of the TGV²-model to nicely resolve smooth regions as well as sharp edges. One can also observe that the computation time is only slightly higher compared to TV-based deblurring. This is due to the fact that most of the time is spent in evaluating the forward and adjoint operator K_h and K_h^* , respectively.

4.3 Zooming

The following deals with the problem of recovering an image from a projected version which can be interpreted as a zooming problem. In order to describe the setting, let $p \in (1, \infty)$ with $p \leq d/(d-1)$ and $Z \subset L^p(\Omega, \mathbf{R}^L)$ be a closed subspace which is modelling images at a low-resolution in which a low-resolution image $f \in Z$ is given. Furthermore, let the operator $P_Z : L^p(\Omega, \mathbf{R}^L) \rightarrow Z$ be a continuous projection onto Z which is modelling the way the resolution is reduced. The corresponding zooming problem then reads as

$$\min_{u \in L^p(\Omega, \mathbf{R}^L)} F(u) + \text{TGV}_\alpha^2(u), \quad F(u) = \mathcal{I}_{\{0\}}(P_Z u - f). \quad (20)$$

Let us discuss existence of solutions. As P_Z is continuous and $\text{rg}(P_Z) = Z$, it is obvious that the indicator functional F is non-negative, proper, convex and lower semi-continuous. To establish the desired coercivity property, we need the assumption

$$\ker(P_Z) \cap \ker(\text{TGV}_\alpha^2) = \{0\}.$$

Note that this implies that P_Z is injective on $\ker(\text{TGV}_\alpha^2)$ as $P_Z u = 0$ and $u \in \ker(\text{TGV}_\alpha^2)$ implies $u = 0$. Now, if $P : L^p(\Omega, \mathbf{R}^L) \rightarrow \ker(\text{TGV}_\alpha^2)$ is a linear, continuous projection, then there is a constant $c > 0$ such that $\|P_Z P u\|_p \geq c \|P u\|_p$ for each $u \in L^p(\Omega, \mathbf{R}^L)$. Thus, for each sequence $\{u^n\}$ in $L^p(\Omega, \mathbf{R}^L)$ such that $\|P u^n\|_p \rightarrow \infty$ and $\{\|(\text{id} - P)u^n\|_p\}$ is bounded, it is impossible that $P_Z u^n - f = 0$ for each n : If this is the case, then

$$P_Z P u^n + P_Z(\text{id} - P)u^n = P_Z u^n = f$$

and consequently,

$$c \|P u^n\|_p \leq \|P_Z P u^n\|_p = \|P_Z(P - \text{id})u^n + f\|_p \leq C + \|f\|_p$$

which implies that $\|P u^n\|_p$ is bounded, a contradiction. Hence, $F(u^n) \rightarrow \infty$ as this argumentation also applies to each subsequence. This establishes (10) and, by Theorem 1, existence of a minimizer.

Example 2. Let $\Omega = (0, M_1) \times (0, M_2)$ with $M_1, M_2 \geq 2$. Denote by $Q_{i,j} = (i-1, i) \times (j-1, j)$ and set

$$Z = \left\{ \sum_{i=1}^{M_1} \sum_{j=1}^{M_2} c_{i,j} \chi_{Q_{i,j}} \mid c_{i,j} \in \mathbf{R}^L \right\}$$

which is modelling $N \times M$ pixel images. A projection onto Z is then given by

$$P_Z u = \sum_{i=1}^{M_1} \sum_{j=1}^{M_2} \left(\int_{Q_{i,j}} u \, dx \right) \chi_{Q_{i,j}}.$$

A $u \in \ker(\text{TGV}_\alpha^2)$ can be expressed by $u(x) = a_1 x_1 + a_2 x_2 + b$ where $a_1, a_2, b \in \mathbf{R}^L$. We then see that

$$c_{i,j} = \int_{Q_{i,j}} u \, dx = \frac{2i-1}{2} a_1 + \frac{2j-1}{2} a_2 + b$$

so $a_1 = c_{2,1} - c_{1,1}$, $a_2 = c_{1,2} - c_{1,1}$ and $b = 2c_{1,1} - \frac{1}{2}(c_{1,2} + c_{2,1})$. Hence, if $P_Z u = 0$, then $c_{1,1} = c_{1,2} = c_{2,2} = 0$ and consequently, $u = 0$. This shows $\ker(P_Z) \cap \ker(\text{TGV}_\alpha^2) = \{0\}$, thus a solution for the zooming problem (20) with averaging over squares exists.

Example 3. Let $\Omega = (0, \pi)^2$ and $M_1, M_2 \geq 2$ and denote by

$$z_{i,j}(x) = \zeta_i(x_1) \zeta_j(x_2), \quad \zeta_i(x) = \begin{cases} \sqrt{\frac{1}{\pi}} & \text{if } i = 0, \\ \sqrt{\frac{2}{\pi}} \cos(ix) & \text{if } i > 0, \end{cases}$$

which is corresponding to the cosine orthonormal basis of $L^2(\Omega)$. The space Z and a projection P_Z is then given by

$$Z = \left\{ \sum_{i=0}^{M_1-1} \sum_{j=0}^{M_2-1} c_{i,j} z_{i,j} \mid c_{i,j} \in \mathbf{R}^L \right\}, \quad P_Z u = \sum_{i=0}^{M_1-1} \sum_{j=0}^{M_2-1} \left(\int_{\Omega} z_{i,j} u \, dx \right) z_{i,j}.$$

For a $u \in \ker(\text{TGV}_\alpha^2)$, i.e., $u(x) = a_1 x_1 + a_2 x_2 + b$, $a_1, a_2, b \in \mathbf{R}^L$ we see that

$$c_{0,0} = \frac{\pi^2}{2} (a_1 + a_2) + \pi b, \quad c_{1,0} = -4a_1, \quad c_{0,1} = -4a_2$$

which implies that if $P_Z u = 0$, then also $u = 0$. Again, we thus have $\ker(P_Z) \cap \ker(\text{TGV}_\alpha^2) = \{0\}$ and consequently, a minimizer for the zooming problem (20) with cosine low-pass filter exists.

For a numerical realization, we have to discretize the space Z . In case of Example 3, a good choice is the corresponding two-dimensional discrete cosine basis. For a discrete low-pass image $f \in (\mathbf{R}^{M_1 \times M_2})^L$, $2 \leq M_1 \leq N_1$, $2 \leq M_2 \leq N_2$, and with DCT denoting the associated parallel discrete cosine transform operator, the discrete functional F_h reads as

$$F_h(u) = \begin{cases} 0 & \text{if DCT}(u)_{i,j} = f_{i,j} \text{ for } 0 \leq i \leq M_1 - 1, 0 \leq j \leq M_2 - 1, \\ \infty & \text{else.} \end{cases}$$

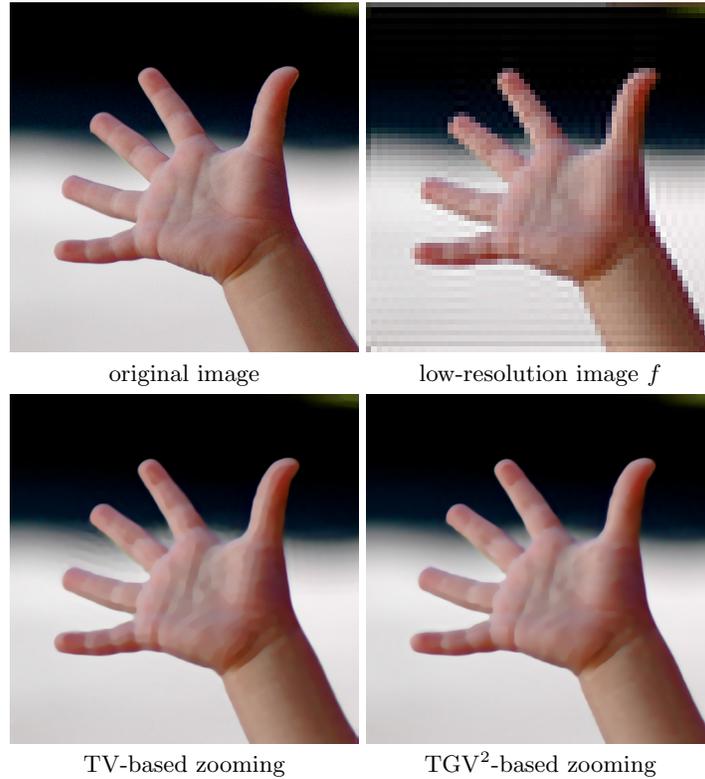


Fig. 5. Example for variational zooming with the TV/TGV² image model according to 4.3. Top: The original image [17] (left, 512×512 pixels, RGB) and a low-resolution representation (right, 64×64 pixels, DCT low-pass coefficients). Bottom: Result of TV-based zooming (left, PSNR=36.60 dB, 2000 iterations, computation time: 4.44 seconds), and TGV²-based zooming (right, PSNR=37.30 dB, 2000 iterations, computation time: 7.46 seconds). Images licenced under CreativeCommons-by-sa-2.0 (<http://creativecommons.org/licenses/by-sa/2.0/>).

Consequently, since DCT is an orthonormal mapping, the resolvent reads as

$$(\text{id} + \tau \partial F_h)^{-1}(u) = \text{DCT}^{-1}(\tilde{c}),$$

$$\tilde{c}_{i,j} = \begin{cases} f_{i,j} & \text{if } 0 \leq i \leq M_1 - 1, 0 \leq j \leq M_2 - 1, \\ \text{DCT}(u)_{i,j} & \text{else.} \end{cases}$$

The iteration step which computes u^{n+1} in Algorithm 1 then reads as

$$\begin{cases} c^{n+1} = \text{DCT}(u^n + \tau \text{div}_h v^{n+1}), \\ \tilde{c}_{i,j}^{n+1} = \begin{cases} f_{i,j} & \text{for } 0 \leq i \leq M_1 - 1, 0 \leq j \leq M_2 - 1, \\ c_{i,j}^{n+1} & \text{else,} \end{cases} \\ u^{n+1} = \text{DCT}^{-1}(\tilde{c}^{n+1}). \end{cases}$$

This procedure was again implemented in PYTHON/PYCUDA and tested on the same machine as for the experiments in the previous subsections. The outcome of a zooming experiment with the factor 8 can be seen in Figure 5. Compared to TV-based zooming, it is interesting to observe that neither models lead to pronounced staircasing artifacts. However, one sees that the TGV^2 model nevertheless leads to a solution which appears less blocky.

4.4 Dequantization

Although images are often modelled as functions which admit continuous values, their digital representation is often quantized to a finite number of bins. The most common case is the 256-level representation which corresponds to 8 bits per pixel and color channel. In the case where significantly less bins are available, the image representing the pixelwise centers of the respective bins appears blocky and unnatural. Hence, one is interested in restoring a more natural image from a quantized image. We assume that each bin has the form

$$[a, b] = \{c \in \mathbf{R}^L \mid a_l \leq c_l \leq b_l, l = 1, \dots, L\} \quad \text{for some } a, b \in \mathbf{R}^L.$$

The given data can then be represented by the respective lower and upper bound functions $f_{\text{lower}}, f_{\text{upper}} \in L^p(\Omega, \mathbf{R}^L)$ for some $p \in (1, \infty)$, $p \leq d/(d-1)$ which have to satisfy $f_{\text{lower}} \leq f_{\text{upper}}$ (componentwise) almost everywhere in Ω . The feasible images are then given by $u \in L^p(\Omega, \mathbf{R}^L)$ such that $f_{\text{lower}} \leq u \leq f_{\text{upper}}$ almost everywhere in Ω . As the standard dequantization $\tilde{f} = \frac{1}{2}(f_{\text{lower}} + f_{\text{upper}})$ is the image which is most probable if for almost every pixel $x \in \Omega$, the values are uniformly distributed in $[f_{\text{lower}}(x), f_{\text{upper}}(x)]$, we like to minimize TGV_α^2 under the above constraint penalizing also the distance to the standard dequantization. This results in the minimization problem

$$\begin{cases} \min_{u \in L^p(\Omega, \mathbf{R}^L)} F(u) + \text{TGV}_\alpha^2(u), \\ F(u) = \mathcal{I}_{\{f_{\text{lower}} \leq u \leq f_{\text{upper}} \text{ a.e.}\}}(u) + \frac{1}{p} \int_\Omega |u - \tilde{f}|_o^p dx. \end{cases} \quad (21)$$

The functional $F : L^p(\Omega, \mathbf{R}^L) \rightarrow (-\infty, \infty]$ can easily be seen to be non-negative, proper, convex and lower semi-continuous, see also Subsection 4.4. Moreover, as we have $F(u) \geq \frac{1}{p} \|u - \tilde{f}\|_p^p$ for each u , the proof of (10) in Subsection 4.1 directly leads to (10) for the above F and, consequently, to the existence of minimizers by virtue of Theorem 1.

The functional F can be discretized in a straightforward way:

$$F_h(u) = \begin{cases} \frac{1}{p} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} |u_{i,j} - \tilde{f}_{i,j}|_o^p & \text{if } (f_{\text{lower}}) \leq u \leq (f_{\text{upper}}), \\ \infty & \text{else.} \end{cases}$$

The associated resolvent operator is explicitly computable in case $|\cdot|_o$ is the

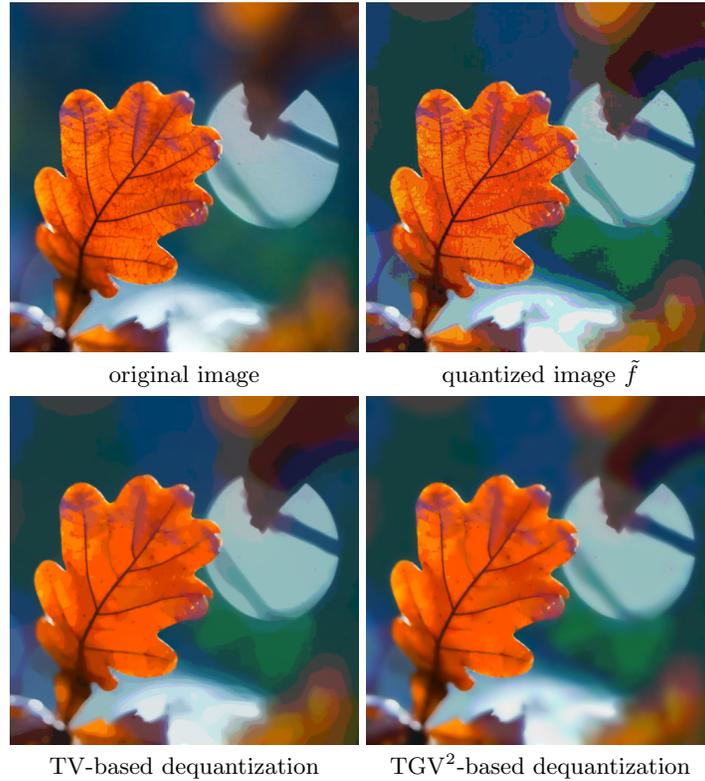


Fig. 6. Example for variational dequantization with the TV/TGV² image model according to (21). Top: The original image [32] (left, 512 × 512 pixels, RGB) and a quantized representation (right, 6 bins per color channel). Bottom: Result of TV-based dequantization (left, PSNR=29.59 dB, 2000 iterations, computation time: 1.88 seconds), and TGV²-based dequantization (right, PSNR=29.87 dB, 2000 iterations, computation time: 4.76 seconds). Images licenced under CreativeCommons-by-sa-2.0 (<http://creativecommons.org/licenses/by-sa/2.0/>).

Euclidean norm and $q = 2$:

$$(\text{id} + \tau \partial F_h)^{-1}(\bar{u}) = \min\left(f_{\text{upper}}, \max\left(f_{\text{lower}}, \frac{\bar{u} + \tau \tilde{f}}{1 + \tau}\right)\right),$$

hence, Algorithm 1 can be employed yielding an iteration step for u^{n+1} which reads as

$$u^{n+1} = \min\left(f_{\text{upper}}, \max\left(f_{\text{lower}}, \frac{\bar{u} + \tau(\text{div}_h v^{n+1} + \tilde{f})}{1 + \tau}\right)\right).$$

This algorithm has again been implemented. A numerical test on an image showing an oak leaf is depicted in Figure 6. One observes that the TV-based

dequantization coincides in essential parts with either f_{lower} or f_{upper} which cause the result to appear blocky. In contrast to that, the TGV_α^2 -based model yields smooth transitions where necessary. In both cases, however, some of the small details in the leaf are lost.

Remark 4. Let us remark that the need for appropriate dequantization also arises in artifact-free JPEG decompression. In this situation, the blockwise DCT-coefficients are only given in a quantized form. This problem can also be solved with TV and TGV_α^2 regularization, see [3, 6] for details.

4.5 Compressive imaging

The last problem we like to discuss is the compressive imaging or ‘single-pixel camera’ reconstruction of a one-channel image [12], a problem which is already set in finite dimensions. Here, an image is not observed directly but only the accumulated gray values over random pixel patterns are measured. One essential point is that the number of measurements is significantly smaller than the number of pixels. For a discrete image of size $N_1 \times N_2$, let $M \ll N_1 N_2$ and $f \in \mathbf{R}^M$ represent the data. For each $1 \leq m \leq M$, let $k^m \in \{0, 1\}^{N_1 \times N_2}$ be the random pixel pattern which is associated with the m -th measurement. The sought image $u \in U$ then obeys

$$Ku = f, \quad \text{where} \quad (Ku)_m = \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} k_{i,j}^m u_{i,k}, \quad 1 \leq m \leq M.$$

As this constraint is highly ambiguous, the compressive imaging approach assumes that the image u is sparse in a certain representation which is usually translated into the discrete total variation $\text{TV}(u)$ being small. A way to reconstruct u from f is then to solve

$$\min_{u \in U} \text{TV}(u) + F_h(u), \quad F_h(u) = \mathcal{I}_{\{0\}}(Ku - f).$$

We like to test the TGV_α^2 -model for this application and propose to solve, numerically, the problem

$$\min_{u \in U} \text{TGV}_\alpha^2(u) + F_h(u), \quad F_h(u) = \mathcal{I}_{\{0\}}(Ku - f). \quad (22)$$

For this purpose, we rewrite F_h according to (13):

$$F_h(u) = \sup_{\lambda \in \Lambda} \langle Ku, \lambda \rangle_\Lambda - G_h(\lambda), \quad G_h(u) = \langle f, \lambda \rangle_\Lambda$$

where $\Lambda = \mathbf{R}^M$. Hence, one can employ Algorithm 2 where the iteration steps for λ^{n+1} and u^{n+1} reads as

$$\begin{cases} \lambda^{n+1} = \lambda^n + \sigma(K\bar{u}^n - f), \\ u^{n+1} = u^n + \tau(\text{div}_h v^{n+1} - K^* \lambda^{n+1}). \end{cases}$$

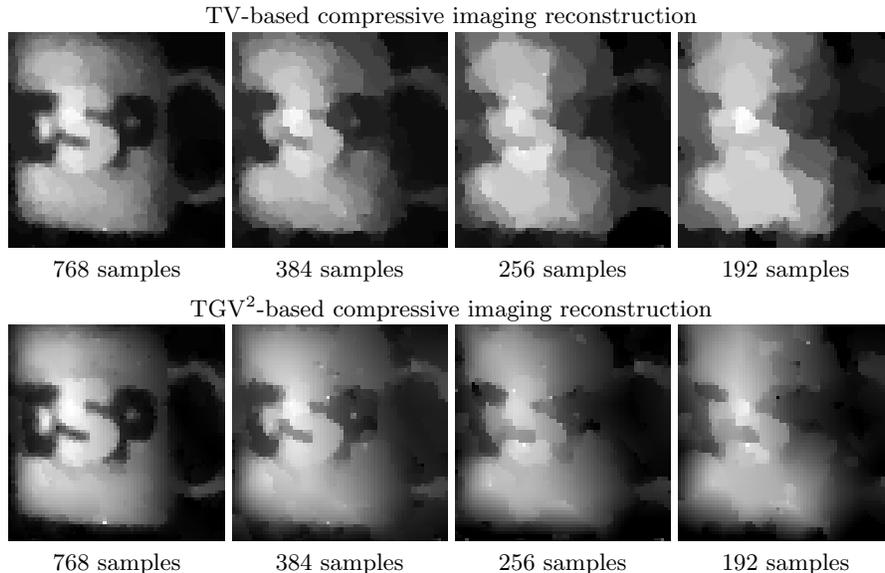


Fig. 7. Example for TV/TGV² compressive imaging reconstruction. Top: TV-based reconstruction of a 64×64 image from 18.75%, 9.375%, 6.25% and 4.6875% of the data (from left to right). Bottom: TGV²-based reconstruction obtained from the same data.

The norm of K can be estimated in terms of the Frobenius norm

$$\|K\| \leq \|K\|_F = \left(\sum_{m=1}^M |k^m|_2^2 \right)^{1/2}$$

which is easily computable from the given random pixel patterns k^m , $1 \leq m \leq M$.

This method has also been implemented and tested. The utilized test data was the compressed sensing camera ‘mug’ data set from Rice’s single-pixel camera project [26]. In Figure 7, the outcome of the computations is depicted. Comparing the results for TV and TGV², one can see a noticeable increase in visual quality as smooth regions are resolved better. Even a low number of samples, rough features of the object is still perceptible in the TGV _{α} ² reconstruction. One has to note, however, that the current method is only suitable to study the effect of the TGV²-model as it takes an extremely large number of iteration and, consequently, much computation time, in order to obtain the results. The reason seems to lie in the ill-posedness of the inversion of K , for which Algorithm 2 only performs a Landweber-type iteration (‘perturbed’ by the TGV _{α} ² regularization). In this case, it might be more efficient to utilize Algorithm 1.

5 Conclusions

The framework of total generalized variation, which already constitutes a convex model for piecewise smooth functions, can easily be extended to the multichannel

case. Compared to the scalar case, the same results hold with respect to basic properties and existence of solutions for associated convex variational problems. Furthermore, these problems can be discretized and solved numerically in a unified way by realizing a primal-dual method for associated convex-concave saddle point problems. This numerical framework is general enough to devise efficient algorithms for the solution of common low-level image processing tasks such as denoising, deblurring, zooming and dequantization. These algorithms can be implemented, without greater effort, in parallel architecture such as GPUs. Numerical experiments confirm that the multichannel version of TGV_α^2 is a suitable model for natural-looking color images and that incorporating it in variational problems leads to visually improved results in comparison to the total-variation counterpart. Moreover, the proof-of-concept application of the primal-dual algorithm to the single-pixel camera compressive imaging framework indicates that TGV_α^2 might also be a suitable model for compressed sensing. However, due to the ill-posedness in this situation, it seems that more effort has to be made in order to solve the reconstruction problem efficiently. This could be a topic of further studies.

Acknowledgements

Support by the Austrian Science Fund (FWF) under grant SFB F32 (SFB “Mathematical Optimization and Applications in Biomedical Sciences”) is gratefully acknowledged.

References

1. Alberti, G., Bouchitté, G., Maso, D.D.: The calibration method for the Mumford-Shah functional and free-discontinuity problems. *Calc. Var. Partial Dif.* 16(3), 299–333 (2003)
2. Bredies, K., Dong, Y., Hintermüller, M.: Spatially dependent regularization parameter selection in total generalized variation models for image restoration. SFB MOBIS Report 2012-002 (2012), University of Graz
3. Bredies, K., Holler, M.: A total variation-based JPEG decompression model. *SIAM J. Imaging Sci.* 5(1), 366–393 (2012)
4. Bredies, K., Kunisch, K., Pock, T.: Total generalized variation. *SIAM J. Imaging Sci.* 3, 492–526 (2011)
5. Bredies, K., Valkonen, T.: Inverse problems with second-order total generalized variation constraints. In: *Proceedings of SampTA 2011. 9th International Conference on Sampling Theory and Applications* (2011)
6. Bredies, K., Holler, M.: Artifact-free JPEG decompression with total generalized variation. In: *VISAPP 2012: Proceedings of the International Conference on Computer Vision Theory and Applications* (2012)
7. Bresson, X., Chan, T.F.C.: Fast dual minimization of the vectorial total variation norm and applications to color image processing. *Inverse Probl. Imag.* 2(4), 455–484 (2008)
8. Chambolle, A., Lions, P.L.: Image recovery via total variation minimization and related problems. *Numer. Math.* 76, 167–188 (1997)

9. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.* 40, 120–145 (2011)
10. Chan, T.F., Esedoglu, S., Park, F.E.: Image decomposition combining staircase reduction and texture extraction. *J. Visual Commun. Image R.* 18(6), 464 – 486 (2007)
11. Dong, Y., Hintermüller, M., Rincon-Camacho, M.: Automated regularization parameter selection in multi-scale total variation models for image restoration. *J. Math. Imaging Vis.* 40, 82–104 (2011)
12. Duarte, M.F., Davenport, M.A., Takhar, D., Laska, J., Sun, T., Kelly, K., Baraniuk, R.G.: Single-pixel imaging via compressive sampling. *IEEE Signal Proc. Mag.* 25(2), 83–91 (2008)
13. Eckstein, J., Bertsekas, D.P.: On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Program.* 55, 293–318 (1992)
14. Hinterberger, W., Scherzer, O.: Variational methods on the space of functions of bounded Hessian for convexification and denoising. *Computing* 76, 109–133 (2006)
15. ^@^ina@Flickr: Alina’s eye. Licenced under CreativeCommons-by-2.0 (<http://creativecommons.org/licenses/by/2.0/>) (2009), http://www.flickr.com/photos/angel_ina/3201337190
16. Klöckner, A.: PyCUDA 2011.2.2. <http://mathematician.de/software/pycuda>
17. Kubina, J.: Bubble popper. Licenced under CreativeCommons-by-sa-2.0 (<http://creativecommons.org/licenses/by-sa/2.0/>) (2005), <http://www.flickr.com/photos/kubina/42275122>
18. Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.* 16(6), 964–979 (1979)
19. Lysaker, M., Lundervold, A., Tai, X.C.: Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time. *IEEE T. Image Process.* 12, 1579–1590 (2003)
20. Mumford, D., Shah, J.: Optimal approximation by piecewise smooth functions and associated variational problems. *Commun. Pure Appl. Math.* 42, 577–685 (1989)
21. Nikolova, M.: Local strong homogeneity of a regularized estimator. *SIAM J. Appl. Math.* 61(2), 633–658 (2000)
22. NVIDIA Corporation: CUDA toolkit 4.0. <http://www.nvidia.com/getcuda>
23. Papafitsoros, K., Schönlieb, C.B.: A combined first and second order variational approach for image reconstruction. *ArXiv e-print 1202.6341* (2012)
24. Pock, T., Cremers, D., Bischof, H., Chambolle, A.: An algorithm for minimizing the Mumford-Shah functional. In: *IEEE 12th International Conference on Computer Vision (ICCV)*. pp. 1133–1140 (2009)
25. Python Software Foundation: Python programming language 2.6. <http://www.python.org/>
26. Rice Single-Pixel Camera Project: CS camera data. <http://dsp.rice.edu/cscamera>
27. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.* 14(5), 877–898 (1976)
28. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Physica D* 60, 259–268 (1992)
29. Schristia@Flickr: 1, 2, 3. Licenced under CreativeCommons-by-2.0 (<http://creativecommons.org/licenses/by/2.0/>) (2009), <http://www.flickr.com/photos/schristia/4057490235>
30. Stampacchia, G.: Le problème de Dirichlet pour les équations elliptiques du second ordre à coefficients discontinus. *Ann. I. Fourier (Grenoble)* 15(1), 189–258 (1965)

31. Steve-h@Flickr: Wet green smiles. Licenced under CreativeCommons-by-sa-2.0 (<http://creativecommons.org/licenses/by-sa/2.0/>) (2009), <http://www.flickr.com/photos/sbh/3859041020>
32. Steve-h@Flickr: Oak leaves and bokeh. Licenced under CreativeCommons-by-sa-2.0 (<http://creativecommons.org/licenses/by-sa/2.0/>) (2012), <http://www.flickr.com/photos/sbh/6802942537>
33. Temam, R.: *Mathematical Problems in Plasticity*. Bordas (1985)
34. The Scipy Community: Scientific tools for Python. <http://www.scipy.org/>
35. Vese, L.: A study in the BV space of a denoising-deblurring variational problem. *Appl. Math. Opt.* 44, 131–161 (2001)
36. W. Ring: Structural properties of solutions to Total Variation regularization problems. *ESAIM: Math. Model. Num.* 34(4), 799–810 (2000)
37. Yang, J., Yin, W., Zhang, Y., Wang, Y.: A fast algorithm for edge-preserving variational multichannel image restoration. *SIAM J. Imaging Sci.* 2(2), 569–592 (2009)