# BILEVEL TRAINING SCHEMES IN IMAGING FOR TOTAL-VARIATION-TYPE FUNCTIONALS WITH CONVEX INTEGRANDS

VALERIO PAGLIARI, KOSTAS PAPAFITSOROS, BOGDAN RAIȚĂ, AND ANDREAS VIKELIS

Abstract. In the context of image processing, we study a class of integral regularizers defined in terms of spatially inhomogeneous integrands that depend on general linear differential operators. Particularly, the spatial dependence is assumed to be only measurable. The setting is made rigorous by means of the theory of Radon measures and of suitable function spaces modeled on BV. We prove the lower semicontinuity of the functionals at stake and existence of minimizers for the corresponding variational problems. Then, we embed the latter into a bilevel scheme in order to automatically compute the regularization parameters. These parameters are considered to be spatially varying, thus allowing for good flexibility and preservation of details in the reconstructed image. After identifying a series of spatially inhomogeneous regularization functionals commonly used in image processing that are included in our framework, we substantiate its feasibility by performing numerical denoising examples in which the spatial dependence of the integrand is measurable. Specifically, we use Huber versions of the first and second order total variation (and their sum) with both the Huber and the regularization parameter being spatially varying. Notably, the spatially varying version of second order total variation produces high quality reconstructions when compared to regularizations of similar type, and the introduction of the low regularity spatially dependent Huber parameter leads to a further enhancement of the image details. We expect that our theoretical investigations and our numerical feasibility study will support future work on setting up schemes where general differential operators with spatially dependent coefficients will also be part of the optimization scheme.

## 1. Introduction

In this contribution we study a general bilevel training scheme for the automatic selection of spatially varying regularization weights in the framework of variational image reconstruction. Specifically, given a suitably defined class Adm of admissible weights $\alpha$, we look for solutions to the problem

$$\alpha^* \in \operatorname{argmin} \{F(u_\alpha) : \alpha \in \mathrm{Adm}\}, \tag{1.1}$$

where $F$ is an assigned cost functional and $u_\alpha$ is an image reconstructed by minimizing

$$I[u;\alpha] \coloneqq \Phi_g(u) + \mathcal{R}(u;\alpha). \tag{1.2}$$

Here, $\Phi_g$ is a fidelity term that penalizes deviations of $u$ from the datum $g$, whereas $\mathcal{R}(u;\alpha)$ is a regularization functional whose strength can be tuned by an appropriate selection of the regularization parameter $\alpha$ belonging to the admissible set Adm. The datum $g$ is typically a corrupted version of some ground truth image $u_{\mathrm{gt}}$. Often, one has

$$g = Tu_{\mathrm{gt}} + \eta,$$

with $\eta$ denoting a random noise component and $T$ being a bounded linear operator that corresponds to a certain image reconstruction problem. For instance, $T$ is a blurring operator in the case of deblurring, a sub-sampled Fourier transform in magnetic resonance imaging (MRI), the Radon transform in tomography, or simply the identity in denoising tasks, on which we will be focusing here. The aim of solving a problem of the type (1.2) for suitable $\Phi_g$, $\mathcal{R}$ and $\alpha$ is to obtain an output $u$ which represents as well as possible the initial

ground truth image $u_{\mathrm{gt}}$. We concisely point out here that the main contributions of the present paper consist in establishing existence of solutions to the scheme for inhomogeneous regularizers of the type

$$\mathcal{R}(u;\alpha) = \int_\Omega \alpha(x) f(x, \mathrm{d}\mathscr{B}u), \tag{1.3}$$

drawing examples from the literature that fit this framework, and presenting numerical results for some of these examples when the $x$-dependence in $f$ has low regularity. Here, $\Omega \subset \mathbb{R}^n$ is the image domain, $\alpha \in \mathrm{Adm} \subset L^\infty(\Omega, [0,+\infty))$ belongs to a class of admissible weights, $f$ is a Carathéodory integrand that is convex in the second entry, and $\mathscr{B}$ is a linear, $k$-th order, homogeneous differential operator with constant coefficients. Before we report further on our contribution, and for the sake of readability, we proceed with a brief review of standard regularization functionals in image reconstruction.

Among classical regularization functionals we find the total variation (TV) [46, 16], as well as higher order extensions of it, like the second order total variation ($\mathrm{TV}^2$) [44, 4] and the total generalized variation (TGV) [8]. For a function $u \in L^1(\Omega)$, these functionals are defined by duality as follows:

$$\mathrm{TV}(u) = \sup\left\{ \int_\Omega u \,\mathrm{div}\phi \,\mathrm{d}x : \ \phi \in \mathrm{C}_c^\infty(\Omega, \mathbb{R}^n),\ \|\phi\|_\infty \leqslant 1 \right\}, \tag{1.4}$$

$$\mathrm{TV}^2(u) = \sup\left\{ \int_\Omega u \,\mathrm{div}^2\phi \,\mathrm{d}x : \ \phi \in \mathrm{C}_c^\infty(\Omega, \mathbb{R}^{n\times n}),\ \|\phi\|_\infty \leqslant 1 \right\}, \tag{1.5}$$

$$\mathrm{TGV}(u) = \sup\left\{ \int_\Omega u \,\mathrm{div}^2\phi \,\mathrm{d}x : \ \phi \in \mathrm{C}_c^\infty(\Omega, \mathcal{S}^{n\times n}),\ \|\phi\|_\infty \leqslant \alpha_0,\ \|\mathrm{div}^2\phi\|_\infty \leqslant \alpha_1 \right\}. \tag{1.6}$$

Here $\mathcal{S}^{n\times n}$ denotes the space of $n \times n$ symmetric matrices. Note that the scalar regularization parameters $\alpha_0, \alpha_1 > 0$ are inserted within the definition of TGV, while the other functionals admit a single weighting parameter $\alpha$ acting in a multiplicative way, i.e., $\alpha\mathrm{TV}$ and $\alpha\mathrm{TV}^2$. If the supremum in (1.4) is finite, then we say that $u \in \mathrm{BV}(\Omega)$, the space of functions of bounded variation [2], and $\mathrm{TV}(u) = |\mathrm{D}u|(\Omega)$, where $|\mathrm{D}u|$ is the total variation measure associated with the distributional derivative $\mathrm{D}u \in \mathcal{M}(\Omega, \mathbb{R}^n)$. Similarly, if the right-hand side in (1.5) is finite, then $u \in \mathrm{BV}^2(\Omega)$, the space of functions of bounded second variation [44, 4], and $\mathrm{TV}^2(u) = |\mathrm{D}^2u|(\Omega)$, with $\mathrm{D}^2u \in \mathcal{M}(\Omega, \mathcal{S}^{n\times n})$. Finally, it turns out that if the supremum in (1.6) is finite, then $u \in \mathrm{BV}(\Omega)$ as well and

$$\mathrm{TGV}(u) = \min_{w \in \mathrm{BD}(\Omega)} \left\{ \alpha_1 \int_\Omega \mathrm{d}|\mathrm{D}u - w| + \alpha_0 \int_\Omega \mathrm{d}|\mathscr{E}w| \right\},$$

see [9, 6]. In the previous formula, $\mathrm{BD}(\Omega)$ is the space of functions of bounded deformation and $\mathscr{E}w$ denotes the symmetrized gradient of $w$. The advantage of higher order regularizers lies in their capability to reduce an undesirable artifact typical of TV, the so-called staircasing effect, that is, the creation of cartoon-like piecewise constant structures in the reconstruction [43]. Additive versions of these regularizers have also been considered in the literature, e.g. the direct sum of TV and $\mathrm{TV}^2$ [44], the more general $\mathrm{TV}^{k_1}$-$\mathrm{TV}^{k_2}$ case [7], or anisotropic versions of those, tuned to regularize along specific directions [35, 38, 45].

Grounding on the concept of convex functions of Radon measures [24], variants of the above regularizers involving convex integrands have also been considered in the literature [44, 49, 34]. A widely used example is the one of Huber total variation $\mathrm{TV}_\gamma$, which is defined for $u \in \mathrm{BV}(\Omega)$ as

$$\mathrm{TV}_\gamma(u) = \int_\Omega f_\gamma(\mathrm{dD}u) = \int_\Omega f_\gamma(\nabla u)\mathrm{d}x + \int_\Omega \mathrm{d}|\mathrm{D}^s u|, \tag{1.7}$$

with $\nabla u$ and $\mathrm{D}^s u$ denoting respectively the absolutely continuous and the singular part of $\mathrm{D}u$ with respect to the Lebesgue measure. The function $f_\gamma\colon \mathbb{R}^n \to \mathbb{R}$ is given for $\gamma \geq 0$ by

$$f_\gamma(z) = \begin{cases} |z| - \dfrac{1}{2}\gamma, & \text{if } |z| \geq \gamma, \\ \dfrac{1}{2\gamma}|z|^2, & \text{if } |z| < \gamma. \end{cases} \tag{1.8}$$

This modification of TV, which corresponds to a smoothing of the $\ell_1$ norm in the discrete setting, is typically considered in order to employ classical smooth numerical solvers for the solution of the minimization problem (1.2). In this specific case, however, it also leads to a reduction of the staircasing effect by penalizing small gradients with the Tikhonov term $(2\gamma)^{-1}\int_\Omega |\nabla u|^2 \mathrm{d}x$, which promotes smooth reconstructions [34, 12]. A less trivial example of regularizer that can be regarded as convex function of the total variation measure is the structural total variation arising in multimodal imaging problems [25, 28], see Section 4 for more details.

In all these models, the choice of the weights in the regularization term is crucial to establish an adequate balance between data fitting and denoising. On one hand, the reconstructed $u$ may remain too noisy or have many artifacts if the regularization is too weak. On the other hand, a very strong regularization may result in an unnatural smoothing effect. In the last years, bilevel minimization methods have been employed to select these weights automatically. A subfamily of these methods assumes the existence of one or several training pairs $(u_{\mathrm{gt}}, g)$ consisting of the ground truth and its corrupted counterpart [14, 22, 21, 20, 41]. In these works the energy in the upper level problem (1.1) is usually given by

$$F_{\mathrm{PSNR}}(u) \coloneqq \|u - u_{\mathrm{gt}}\|^2_{L^2(\Omega)}, \tag{1.9}$$

and its minimization essentially corresponds to finding reconstructions that are closest to the ground truth in the least square sense. Since typical images generally feature both homogeneous regions and fine details, it is reasonable to assume that the optimal regularization intensity is not uniform throughout the domain. This matter of fact has prompted researchers to consider bilevel schemes that output space-dependent weights, i.e., functions $\alpha\colon \Omega \to [0, +\infty)$ [18, 23]. A recent series of papers [32, 33, 29, 31] deals with schemes for TV and TGV that yield such weights without resorting to the ground truth. If the corrupted datum $g$ is obtained by some additive Gaussian noise $\eta$ with variance $\sigma^2$, this is achieved by the introduction of the statistics-based upper level objective

$$F_{\mathrm{stat}}(u) \coloneqq \frac{1}{2}\int_\Omega \max(Ru - \overline{\sigma}^2, 0)^2 \mathrm{d}x + \frac{1}{2}\int_\Omega \min(Ru - \underline{\sigma}^2, 0)^2 \mathrm{d}x, \tag{1.10}$$

where $\underline{\sigma}^2 \coloneqq \sigma^2 - \epsilon$, $\overline{\sigma}^2 \coloneqq \sigma^2 + \epsilon$, and

$$Ru(x) \coloneqq \int_\Omega w(x,y)(u-g)^2(y)\mathrm{d}y \quad \text{for } w \in L^\infty(\Omega\times\Omega), \ \int_\Omega\int_\Omega w(x,y)\mathrm{d}x\mathrm{d}y = 1.$$

The idea is that if the reconstructed image $u$ is close to $u_{\mathrm{gt}}$, then it is expected that on average the value of $Ru(x)$ will be close to $\sigma^2$. This justifies the use of $F_{\mathrm{stat}}$, since its minimization forces the localized residuals $Ru$ to fall within the tight corridor $[\underline{\sigma}^2, \overline{\sigma}^2]$.

**Our contribution.** The contribution of this paper is connected to the aforementioned literature on several levels. Starting from an arbitrary $k$-th order, homogeneous, linear differential operator $\mathscr{B}$ between two finite dimensional Euclidean spaces $\mathbb{U}$ and $\mathbb{V}$, we introduce the general regularizer

$$\mathcal{R}(u;\alpha) = \sum_{i=1}^{k-1}\int_\Omega \alpha_i(x) f_i(x, \mathrm{D}^i u) + \int_\Omega \alpha_k(x) f_k(x, \mathrm{d}\mathscr{B}u), \tag{1.11}$$

$\alpha_i \colon \Omega \to [0, +\infty)$ being for $i = 1, \dots, k$ the spatially dependent weights. The functions $f_i$ are of linear growth and convex in the second variable. We assume them to be Carathéodory integrands, so that, in particular, their dependence on the spatial variable $x$ is *only measurable*. More details about the setting are to be found in Section 2.

As a first contribution, we prove lower semicontinuity of the functional in (1.11) with respect to a suitable weak-$*$ convergence, a necessary step towards the existence of solutions of the corresponding variational image reconstruction problem (1.2). Secondly, we introduce and prove existence of solutions to the bilevel scheme, which provides an optimal spatially dependent weight $\alpha := (\alpha_1, \dots, \alpha_k)$ and an associated reconstructed image $u$.

From the theoretical point of view, one of the main advantages of our approach is the fact that we can allow for *spatially dependent weights* and for *general convex integrands* in the regularizers. Since we require very low regularity of $f_i(\bullet, z)$, there is room for a lot of flexibility in choosing the convex functions $f_i(x, \bullet)$. We are thus in a position to adjust the regularization effects according to the characteristics of the specific problem at stake, as in the case of structural total variation or Huber versions of TV with spatially varying Huber parameter, see Section 4. Our hypotheses on the convex integrands are optimal, due to the use of Young measures for oscillation and concentration (see Appendix A). Indeed, in spite of the generality of the integrands, the use of these powerful tools enables us to prove our main results by short and elegant arguments. In the same vein, we can moreover consider TV-type integrals that depend on arbitrary differential operators (not just gradients and Hessians), without any extra assumptions on the regularizers. We expect that this will play a role in future work, as we explain below. It is also the case that our regularity assumptions on the weights are minimal, as can be seen from the discussion in Section 2.3, as well as from the numerical justification for this regularity, see Section 5.

Another technical novelty is that we allow for abstract differential operators modeled on total variation schemes. The employment of such a general theoretical framework brings several advantages. Firstly, by considering general differential operators we can aim to obtain a unifying umbrella for the models previously discussed [44, 7, 38, 45, 28]. These can also include nonstandard choices for the operator $\mathscr{B}$ like the symmetrized gradient and the Laplacian, which have already been discussed in the literature both in the context of additive and infimal convolution-type regularization [7, 11]. Secondly, the wide generality in which our proofs hold lays the foundations for future work. We aim at developing our theory to include optimization problems over linear PDE operators $\mathscr{B}$ with spatially dependent coefficients. While this task seems gargantuan at the moment, we expect that our lower semicontinuity result, Theorem 3.2, and the existence result for the bilevel schemes, Theorem 3.7, will serve as preliminary work in this direction. In fact, there has been work done concerning functionals depending on general differential operators. One example comes from the recent paper [19], where a bilevel scheme for first order differential operators $\mathscr{B}$ is developed. Interestingly, the authors identify classes of $\mathbb{C}$-elliptic operators $\mathscr{B}$ such that the scheme outputs an optimal reconstructed image and an optimal $\mathscr{B}$ for the upper level problem. In their analysis one always obtains minimizers that lie in the space of maps of bounded variation. In contrast, in our method the operator $\mathscr{B}$ is fixed, but it is allowed to be *arbitrary* (see Theorems 3.2 and 3.7). Of course, the difference comes from the fact that $\mathrm{BV} \subsetneq \mathrm{BV}^{\mathscr{B}}$ in general, see [42, 37].

Before performing numerical experiments, we proceed by revising some examples from the literature that are encompassed by our theoretical approach, that is, we identify regularizers that can be considered as integrands of spatially inhomogeneous functionals. Specifically, we discuss the cases of structural and directional total variation, as well as its (higher order) Huber versions with the Huber parameter being spatially dependent. Motivated by the facts that in the latter case the $x$-dependence of the integrand $f$ is nontrivial (it is only measurable) and that such a case has not been considered in the imaging literature,

we conclude the paper with a series of corresponding numerical examples. In particular, the examples deal with versions of the Huber TV and $\mathrm{TV}^2$ in which *both* the Huber and the regularization parameters are spatially dependent and we also provide examples with the sum of those (Huber TV-$\mathrm{TV}^2$). We devise a strategy to prefix the Huber parameter in a way that is adapted to the image, while the regularization parameter is computed automatically by the bilevel scheme. We also provide numerical justifications for the regularity (continuity) of these regularization weights. While the purpose of these examples is to support the applicability and versatility of our framework, they also lead to two new observations that are interesting from the reconstruction point of view. The first one is that the bilevel $\mathrm{TV}^2$ and TV-$\mathrm{TV}^2$ with spatially varying weights, in combination with the statistics-based upper level objective $F_{\mathrm{stat}}$, are able to produce high quality reconstructions, even outperforming TGV (both in its scalar and spatially varying versions). The second one is that the introduction of the spatially varying Huber parameter can further enhance the detailed areas in the reconstructed images.

**Structure of the paper.** In Section 2 we introduce the spaces of functions of bounded $\mathscr{B}$-variation, which provide the functional setting for our analysis. We then state the assumptions under which the general bilevel scheme is studied, and we justify our choice of admissible weights. In Section 3 we first prove lower semicontinuity and existence theorems concerning the lower level of the bilevel scheme, then we turn to the main existence result for optimal weights and reconstructed images. The connections between our work and the existing literature are presented in Section 4, where we show that our theory encompasses several regularization functionals already used in the literature. Eventually, numerical experiments for the Huber TV, $\mathrm{TV}^2$ and TV-$\mathrm{TV}^2$ regularizers with spatially varying weights are performed in Section 5, where we briefly describe the algorithmic set-up and present a series of numerical examples.

**Acknowledgements.** The authors are very grateful to Elisa Davoli for preliminary discussions at the early stage of the project. VP acknowledges the supports of the Austrian Science Fund (FWF) through project I4052 and of the BMBWF through the OeAD-WTZ num. CZ04/2019.

## 2. Mathematical setup

We collect in this section all the assumptions and the notations to be used in the sequel. We also include some heuristic motivation for the definition of the class of admissible weights.

### 2.1. Functional setting: $\mathrm{BV}^{\mathscr{B}}_p$ spaces.

We work in the $n$-dimensional Euclidean space $\mathbb{R}^n$, $n \geqslant 2$, that we endow with the Lebesgue measure $\mathscr{L}^n$. We let $\Omega \subset \mathbb{R}^n$ be a fixed open and bounded set with Lipschitz boundary, which stands as the image domain. In typical applications $n = 2$ and $\Omega$ is a rectangle. We suppose that the image functions take values in a finite dimensional inner product space $\mathbb{U}$, which, for instance, is $\mathbb{R}$ for grayscale images, $\mathbb{R}^3$ for RGB images, or it can be even more structured like e.g. $\mathcal{S}^{n\times n}$ for diffusion tensor imaging [48]. In order to describe further the functional setting in which our analysis is carried out, we need to introduce the class of differential operators that we consider.

Let $\mathbb{V}$ be another finite dimensional inner product space and let $\mathrm{Lin}(\mathbb{U},\mathbb{V})$ be the space of linear maps from $\mathbb{U}$ to $\mathbb{V}$. Hereafter, for $k \in \mathbb{N}\setminus\{0\}$, $\mathscr{B}$ denotes a $k$-th order, homogeneous and linear differential operator with constant coefficients. Explicitly, given $B_i \in \mathrm{Lin}(\mathbb{U},\mathbb{V})$ for any $n$-dimensional multi-index $i = (i_1,\dots,i_n) \in \mathbb{N}^n$, we define for a smooth function $u\colon \mathbb{R}^n \to \mathbb{U}$

$$\mathscr{B}u \coloneqq \sum_{|i|=k} B_i \partial^i u.$$

(recall that $|i| \coloneqq \sum_j i_j$). When $u$ is less regular, we interpret $\mathscr{B}u$ in the distributional sense. In particular, we are interested in the case in which $\mathscr{B}u$ is a finite Radon measure.

Given a generic open set $O \subset \mathbb{R}^n$, we recall that a finite ($\mathbb{U}$-valued) Radon measure on $O$ is a measure on the $\sigma$-algebra of the Borel sets of $O$. We denote the space of such measures by $\mathcal{M}(O, \mathbb{U})$, and, by means of the classical Riesz's representation theorem, we can identify it as the dual of the space

$$\mathrm{C}_0(O, \mathbb{U}) \coloneqq \big\{u \colon O \to \mathbb{U} : \{|u| > \delta\} \text{ is relatively compact for all } \delta > 0\big\},$$

equipped with the uniform norm. The dual norm induced on $\mathcal{M}(O, \mathbb{U})$ turns out to be the one associated with the total variation, which we denote by $|\bullet|$. We refer to [2, Chapter 1] for further reading on measure theory.

In the case we consider, given $\mu \in \mathcal{M}(\Omega, \mathbb{U})$, we have that $\mathscr{B}\mu \in \mathcal{M}(\Omega, \mathbb{V})$ if and only if there exists $\nu \in \mathcal{M}(\Omega, \mathbb{V})$ such that

$$\langle \nu, \phi \rangle = \int_\Omega \mathscr{B}^* \phi \, \mathrm{d}\mu \qquad \text{for all } \phi \in \mathrm{C}_c^\infty(\Omega, \mathbb{V}),$$

where $\langle \bullet, \bullet \rangle$ denotes the duality pairing and $\mathscr{B}^*$ is the formal adjoint of $\mathscr{B}$, i.e.

$$\mathscr{B}^* \phi \coloneqq - \sum_{|i|=k} B_i^* \partial^i \phi \qquad \text{for all } \phi \in \mathrm{C}_c^\infty(\mathbb{R}^n, \mathbb{V}),$$

$B_i^*$ being the transpose of $B_i$.

It is convenient to have at our disposal a specific notation for the spaces that we are going to work with. For $\Omega$, $\mathbb{U}$ and $\mathbb{V}$ as above, and for $p \in (1, +\infty)$, we set

$$\mathrm{BV}_p^{\mathscr{B}}(\Omega) \coloneqq \{u \in L^p(\Omega, \mathbb{U}) \colon \mathscr{B}u \in \mathcal{M}(\Omega, \mathbb{V})\},$$

and we abbreviate $\mathrm{BV}^{\mathscr{B}}(\Omega) \coloneqq \mathrm{BV}_p^{\mathscr{B}}(\Omega)$ when $p = 1$. The spaces above are naturally endowed with weak-$*$ notions of convergence, namely

$$u_j \overset{*}{\rightharpoonup} u \text{ in } \mathrm{BV}_p^{\mathscr{B}}(\Omega) \quad \text{if and only if} \quad u_j \rightharpoonup u \text{ in } L^p(\Omega) \text{ and } \mathscr{B}u_j \overset{*}{\rightharpoonup} \mathscr{B}u \text{ in } \mathcal{M}(\Omega, \mathbb{V}).$$

Stronger convergences may be retrieved if the class of differential operators is restricted. We give a brief account on this point in the following lines.

Due to the interaction between Fourier transform and linear PDE, often analytic properties of $\mathrm{BV}^{\mathscr{B}}$ spaces (and of the equation $\mathscr{B}u = v$ in general) can be expressed in terms of algebraic properties of the characteristic polynomial. We recall that the characteristic polynomial, or symbol, of $\mathscr{B}$ is

$$\mathscr{B}(\xi) \coloneqq \sum_{|i|=k} \xi^i B_i \in \mathrm{Lin}(\mathbb{U}, \mathbb{V}), \qquad \xi \in \mathbb{C}^n,$$

where $\xi^i \coloneqq \xi_1^{i_1} \cdots \xi_n^{i_n}$. The following property, ensuring higher integrability for the minimizers of the lower level problem can be singled out, cf. Theorem 3.5:

**Definition 2.1** ([47, 10, 27])**.** *An operator $\mathscr{B}$ is said to be $\mathbb{C}$-elliptic if*

$$\ker_{\mathbb{C}} \mathscr{B}(\xi) = \{0\} \quad \text{for all } \xi \in \mathbb{C}^n \setminus \{0\}.$$

It was shown in [47] that $\mathbb{C}$-ellipticity is equivalent with full Sobolev regularity for the equation $\mathscr{B}u = v$ on domains, provided that $v \in L^p(\Omega, \mathbb{V})$, $p \in (1, +\infty)$. For $p = 1$ we have the counterpart:

**Theorem 2.2** ([27])**.** *Let $\Omega \subset \mathbb{R}^n$ be a Lipschitz domain. An operator $\mathscr{B}$ is $\mathbb{C}$-elliptic if and only if*

$$\|u\|_{\mathrm{W}^{k-1,n/(n-1)}(\Omega)} \leqslant c \left(|\mathscr{B}u|(\Omega) + \|u\|_{L^1(\Omega)}\right) \quad \text{for } u \in \mathrm{BV}^{\mathscr{B}}(\Omega),$$

*where $|\mathscr{B}u|$ is the total variation measure associated with $\mathscr{B}u$.*

2.2. **The bilevel scheme.** We are now in a position to formulate our problem rigorously. Let us fix $p \in (1, +\infty)$. As we have touched upon in the introduction, our goal is to provide an existence result for solutions to the following training scheme: given $g \in L^p(\Omega, \mathbb{U})$,

$$\text{find } \alpha^* \in \operatorname{argmin} \{F(u_\alpha) : \alpha \in \mathrm{Adm}\} \tag{L1}$$

$$\text{such that } u_\alpha \in \operatorname{argmin} \left\{ I[u;\alpha] \colon u \in \mathrm{BV}^{\mathscr{B}}_p(\Omega) \right\}, \tag{L2}$$

where

$$I[u;\alpha] \coloneqq \Phi_g(u) + \int_\Omega \alpha(x) f(x, \mathrm{d}\mathscr{B}u). \tag{2.2}$$

All the due definitions and assumptions are collected below.

– *Cost functional:* As for the upper level problem (L1), $F \colon L^p(\Omega, \mathbb{U}) \to \mathbb{R}$ is a proper, convex and weakly lower semicontinuous functional. Typical choices for this functional are the PSNR maximizing $F_{\mathrm{PSNR}}$ in (1.9), which makes use of the ground truth $u_{\mathrm{gt}}$, and the statistics-based, ground truth-free $F_{\mathrm{stat}}$ in (1.10), in the spirit of supervised and unsupervised learning respectively.

– *Fidelity term:* The assumptions on the functional $\Phi_g \colon L^p(\Omega, \mathbb{U}) \to \mathbb{R}$ in (2.2) are similar to the ones on $F$, namely $\Phi_g$ is a proper, convex and weakly lower semicontinuous functional that is also coercive. This means that

$$\lim_{j \to +\infty} \|u_j - g\|_{L^p(\Omega,\mathbb{U})} = +\infty \quad \text{implies} \quad \lim_{j \to +\infty} \Phi_g(u_j) = +\infty.$$

In particular,

$$\Phi_g(u) = \|u_j - g\|^p_{L^p(\Omega,\mathbb{U})}$$

is a simple instance of fidelity term.

– *Weights:* Given $\underline{\alpha}, \overline{\alpha} \geqslant 0$ with $\underline{\alpha} < \overline{\alpha}$, the scalar fields $\alpha \in \mathrm{C}(\bar{\Omega}, [\underline{\alpha}, \overline{\alpha}])$ are supposed to share the same uniform modulus of continuity $\omega$, that is, an increasing function $\omega \colon [0, +\infty) \to [0, +\infty)$ such that $\omega(0) = 0$. As a consequence, the class of admissible weights

$$\mathrm{Adm} \coloneqq \left\{ \alpha \in \mathrm{C}(\bar{\Omega}, [\underline{\alpha}, \overline{\alpha}]) : |\alpha(x) - \alpha(y)| \leqslant \omega(|x - y|) \text{ for every } x, y \in \bar{\Omega} \right\} \tag{2.3}$$

is compact with respect to the uniform norm by Arzelà–Ascoli theorem. We will motivate the definition of the set Adm below, see Subsection 2.3.

– *Integrand:* The function $f \colon \Omega \times \mathbb{V} \to [0, +\infty)$ is a *Carathéodory* integrand such that $z \mapsto f(x, z)$ is convex for $\mathscr{L}^n$-a.e. $x \in \Omega$. Here, Carathéodory integrand means jointly Borel measurable and continuous in the second variable. We also suppose that the integrand satisfies the linear coercivity and growth bounds

$$c(|\bullet| - 1) \leqslant f(x, \bullet) \leqslant C(1 + |\bullet|) \quad \text{for } \mathscr{L}^n\text{-a.e. } x \in \Omega, \tag{2.4}$$

for some $c, C \geqslant 0$.

In order to make sense of (2.2), we are still left to define the applications of convex functions to measures, as in the term $f(x, \mathrm{d}\mathscr{B}u)$. For an integrand $f \colon \Omega \times \mathbb{V} \to [0, +\infty)$ satisfying (2.4), we define the *recession function*

$$f^\infty(x, z) \coloneqq \lim_{(x', z', t) \to (x, z, +\infty)} \frac{f(x', tz')}{t} \quad \text{for } (x, z) \in \bar{\Omega} \times \mathbb{V}, \tag{2.5}$$

which we assume to exist. We then set for $\mu \in \mathcal{M}(\Omega, \mathbb{V})$

$$\int_\Omega f(x, \mathrm{d}\mu) \coloneqq \int_\Omega f\left(x, \frac{\mathrm{d}\mu}{\mathrm{d}\mathscr{L}^n}(x)\right) \mathrm{d}x + \int_\Omega f^\infty\left(x, \frac{\mathrm{d}\mu^s}{\mathrm{d}|\mu|}(x)\right) \mathrm{d}|\mu|(x), \tag{2.6}$$

where $\mu^s$ denotes the singular part of $\mu$ with respect to Lebesgue measure and $\mathrm{d}\mu/\mathrm{d}\nu$ is the Radon-Nikodým derivative of $\mu$ with respect to the measure $\nu$.

2.3. **Rationale for the definition of the set of admissible weights.** In order to highlight the main technical obstacles that are encountered in the analysis of bilevel training schemes with space-dependent weights, we start with an example involving the total variation with spatially varying weight, which, in spite of its simplicity, exhibits the typical features of such class of problems. The model we address has been already studied in [32] (with a different approach from the one we outline).

Let $\Omega \subset \mathbb{R}^n$ be a bounded open set with Lipschitz boundary. For $p \in [1, \frac{n}{n-1})$, we suppose that a training pair $(u_{\mathrm{gt}}, g) \in L^2(\Omega, \mathbb{R}) \times L^p(\Omega, \mathbb{R})$ is assigned, where $u_{\mathrm{gt}}$ and $g$ encode respectively the ground truth and the corrupted datum. We also fix two positive parameters $\underline{\alpha}$ and $\overline{\alpha}$, and we provisionally allow the regularizing weights to vary in $\mathrm{LSC}(\Omega, [\underline{\alpha}, \overline{\alpha}])$, the space of lower semicontinuous functions on $\Omega$ with range in $[\underline{\alpha}, \overline{\alpha}]$.

For $u \in \mathrm{BV}(\Omega)$ and $\alpha \in \mathrm{LSC}(\Omega, [\underline{\alpha}, \overline{\alpha}])$, we introduce the first order functional

$$J[u; \alpha] \coloneqq \int_\Omega |u - g|^p \mathrm{d}x + \int_\Omega \alpha(x) \mathrm{d}|\mathrm{D}u|(x), \tag{2.7}$$

and the ensuing corresponding training scheme:

$$\text{find } \alpha^* \in \operatorname{argmin} \{F_{\mathrm{PSNR}}(u_\alpha) : \alpha \in \mathrm{LSC}(\Omega, [\underline{\alpha}, \overline{\alpha}])\} \tag{2.8}$$

$$\text{such that } u_\alpha \in \operatorname{argmin} \{J[u; \alpha] \colon u \in \mathrm{BV}(\Omega)\}. \tag{2.9}$$

The functional in (2.7) is reminiscent of the one considered in [3], where, motivated by vortex density models, the authors studied the property of minimizers, i.e., of solutions to (2.9).

Before discussing the existence of solutions to the scheme (2.8)–(2.9) as a whole, let us justify the choice of the class of weights in (2.8). Note that the definition of $J$ itself calls for some degree of regularity for $\alpha$. Indeed, if in (2.7) $\alpha \colon \Omega \to [\underline{\alpha}, \overline{\alpha}]$ is a given function and $u$ is allowed to vary in $\mathrm{BV}(\Omega)$ (as it is the case of (2.9)), there might be choices of $u$ for which the coupling

$$\int_\Omega \alpha(x) \mathrm{d}|\mathrm{D}u|(x)$$

is not well-defined. Prescribing lower semicontinuity for the admissible weights $\alpha$ allows to circumvent the issue, because lower semicontinuous functions are Borel measurable and $\mathrm{D}u \in \mathcal{M}(\Omega, \mathbb{R}^n)$ is a Borel measure. Besides, for any $\alpha \in \mathrm{LSC}(\Omega, [\underline{\alpha}, \overline{\alpha}])$ the existence of a solution $u_\alpha$ to (2.9) follows by the direct method of the calculus of variations. Indeed, we firstly observe that the coercivity of $J$ in $L^1$ is deduced by the following standard result (see e.g. [2, Theorem 3.23]):

**Theorem 2.3** (Compactness in BV)**.** *Let $\Omega \subset \mathbb{R}^n$ be a bounded Lipschitz domain and let $(u_j)_j$ be a bounded sequence in $\mathrm{BV}(\Omega)$. Then, there exist $u \in \mathrm{BV}(\Omega)$ and a subsequence $(j_k)_k$ such that $(u_{j_k})$ weakly-$*$ converges to $u$, that is, $u_{j_k} \to u$ in $L^1(\Omega)$ and*

$$\lim_{k \to +\infty} \int_\Omega \phi \, \mathrm{d}\mathrm{D}u_{j_k} = \int_\Omega \phi \, \mathrm{d}\mathrm{D}u \qquad \text{for all } \phi \in \mathrm{C}_0(\Omega).$$

Secondly, we notice that $J[\bullet; \alpha]$ is lower semicontinuous with respect to the $L^1$-convergence, because, when $\alpha \in \mathrm{LSC}(\Omega, [\underline{\alpha}, \overline{\alpha}])$, general lower semicontinuity results in BV may be invoked (see e.g. [26]; and also [1] for lower semicontinuity and relaxation results with BV integrands).

Once we know that for lower semicontinuous weights, solutions to (2.9) exist, we can try to tackle the complete scheme. So, let $(\alpha_j)_j \subset \mathrm{LSC}(\Omega, [\underline{\alpha}, \overline{\alpha}])$ be a minimizing sequence

for (2.8). Then, by definition, the integrals

$$F_{\mathrm{PSNR}}(u_j) = \|u_j - u_{\mathrm{gt}}\|^2_{L^2(\Omega)} \qquad \text{with } u_j \coloneqq u_{\alpha_j}$$

converge, and we deduce that $(u_j)_j$ is a bounded sequence in $L^2(\Omega)$. Denote by $u \in L^2(\Omega)$ the weak $L^2$-limit of (a subsequence of) $(u_j)_j$. By lower semicontinuity of the $L^2$-norm, we obtain

$$F_{\mathrm{PSNR}}(u) \leqslant \inf\left\{F_{\mathrm{PSNR}}(u_j) : \alpha \in \mathrm{LSC}(\Omega, [\underline{\alpha}, \overline{\alpha}])\right\}.$$

If we manage to show that $u = u_{\alpha^*}$ for some admissible $\alpha^*$, then the latter is a solution to (2.8). The natural choice for $\alpha^*$ would be the weak-$*$ limit of $(\alpha_j)_j$ in $L^\infty(\Omega)$, which can fail in general to have any lower semicontinuous representative. On the positive side, $J[u;\bullet]$ is continuous with respect to a suitable weak-$*$ convergence. Indeed, if $(\alpha_j)_j \subset \mathrm{LSC}(\Omega, [\underline{\alpha}, \overline{\alpha}])$ is bounded and $u \in \mathrm{BV}(\Omega)$, then there exist a subsequence, which we do not relabel, and $\alpha^* \in L^\infty(\Omega, [\underline{\alpha}, \overline{\alpha}]; |\mathrm{D}u|)$ such that

$$\lim_{j\to+\infty} \int_\Omega \alpha_j(x)\phi(x)\mathrm{d}|\mathrm{D}u|(x) = \int_\Omega \alpha^*(x)\phi(x)\mathrm{d}|\mathrm{D}u|(x) \qquad \text{for all } \phi \in L^1(\Omega; |\mathrm{D}u|).$$

In particular,

$$\lim_{j\to+\infty} J[u;\alpha_j] = J[u;\alpha^*]. \tag{2.10}$$

The previous lines suggest that what is missing to solve the scheme (2.8)–(2.9) is a compactness property for the class of admissible weights. This leads us to reduce ourselves to the problem

$$\begin{aligned} &\text{find } \alpha^* \in \operatorname{argmin}\left\{F_{\mathrm{PSNR}}(u_\alpha) : \alpha \in \mathrm{Adm}\right\} \\ &\text{with } u_\alpha \in \operatorname{argmin}\left\{J[u;\alpha] \colon u \in \mathrm{BV}(\Omega)\right\}, \end{aligned} \tag{2.11}$$

where we assume *a priori* that $\mathrm{Adm} \subset \mathrm{C}(\bar{\Omega}, [\underline{\alpha}, \overline{\alpha}])$ is compact with respect to the uniform convergence. Under the compactness assumptions on the class of admissible weights, if $(\alpha_j)_j \subset \mathrm{Adm}$ is a minimizing sequence for (2.11), and if $(u_j)_j$ and $u$ are constructed as above, we are actually able to prove that $u = u_{\alpha^*}$, where $\alpha^* \in \mathrm{Adm}$ is the uniform limit of $(\alpha_j)$. In other words, the couple $(\alpha^*, u)$ is a solution to the scheme consisting of (2.11)–(2.9).

To prove the claim, we need to show that

$$J[u;\alpha^*] \leqslant J[v;\alpha^*] \qquad \text{for any } v \in \mathrm{BV}(\Omega). \tag{2.12}$$

We start from observing that the definition of $u_j$ grants

$$J[u_j;\alpha_j] \leqslant J[v;\alpha_j] \qquad \text{for any } v \in \mathrm{BV}(\Omega),$$

and hence, for any $v \in \mathrm{BV}(\Omega)$,

$$\liminf_{j\to+\infty} J[u_j;\alpha_j] \leqslant \liminf_{j\to+\infty} J[v;\alpha_j] = J[v;\alpha^*], \tag{2.13}$$

where the equality follows by (2.10). In particular,

$$\liminf_{j\to+\infty} J[u_j;\alpha_j] < +\infty.$$

Then, the uniform lower bound $\alpha_j \geqslant \underline{\alpha}$ and Theorem 2.3 yield that $(u_j)$ converges weakly-$*$ in $\mathrm{BV}(\Omega)$ (again upon extraction of subsequences) to a limit function which is necessarily $u$. We thereby infer

$$u \in \mathrm{BV}(\Omega) \cap L^2(\Omega).$$

From the uniform convergence of $(\alpha_j)$ and the weak-$*$ convergence of $(u_j)$ we obtain

$$J[u;\alpha^*] \leqslant \liminf_{j\to+\infty} J[u_j;\alpha_j]. \tag{2.14}$$

On the whole, owing to (2.13), we deduce (2.12).

Even though in our theoretical investigations the admissible set Adm is defined as a compact subset of $\mathrm{C}(\bar{\Omega}, [\underline{\alpha}, \overline{\alpha}])$, we already mention that in the numerical examples we will employ a slightly different class of weights: as we discuss in detail at the end of Section 5.1, Adm will be substituted by suitable $H^1$-projections of the regularization weights along the iterations of the numerical iterative scheme we will consider. Although the projections are performed within the numerical algorithm, the outcome is coherent with the analytical treatment, since a regularity result about the $H^1$-projection combined with a box constraint will guarantee that the weights belong to a Hölder space $\mathrm{C}^{0,\beta}(\bar{\Omega})$ for some $\beta \in (0,1)$.

*Remark* 2.4. In the absence of compactness for the set Adm under uniform convergence, the analysis becomes more delicate. We outline here some of the issues.

Keeping in force the notation above, let $u \in \mathrm{BV}(\Omega)$ be the weak-$*$ limit of $(u_j)$ and let $\alpha^* \in L^\infty(\Omega, [\underline{\alpha}, \overline{\alpha}]; |\mathrm{D}u|)$ be the weak-$*$ limit of $(\alpha_j)$. Proving the optimality of $u$, i.e. $u = u_{\alpha^*}$, means

$$J[u; \alpha^*] \leqslant J[v; \alpha^*] \qquad \text{for any } v \in \mathrm{BV}(\Omega).$$

However, the right-hand side might be not well-defined. Intuitively, the point is that an ideal class of weights should be *a priori* "sufficiently compact", and at the same it should give rise to "well-behaved" weighted BV functions.

Another passage that is needed in the proof of existence (cf. (3.8), (3.9)) is the following semicontinuity inequality:

$$J[u; \alpha^*] \leqslant \liminf_{j \to +\infty} J[u_j; \alpha^{(j)}].$$

Knowing that $(u_j)$ weakly-$*$ converges to $u$, its validity is undermined if only weak-$*$ convergence is available for the weights.

## 3. The bilevel training scheme in the space $\mathrm{BV}^{\mathscr{B}}_p$

We devote this section to the proof of our main theoretical result, that is, the existence of solutions to the bilevel scheme (L1)–(L2). The study of the lower level problem will be addressed by Theorem 3.2. A variant involving the regularization in (1.11) will also be presented, see Theorem 3.5 and Remark 3.8.

We begin with a general lower semicontinuity result for convex integrands with measurable $x$-dependence, a minimal assumption which is difficult to obtain with analytical techniques different from Young measures.

**Proposition 3.1.** *Let $f\colon \Omega \times \mathbb{V} \to [0, +\infty)$ be a Borel measurable integrand such that $f(x, \bullet)$ is convex for almost every $x \in \Omega$. Suppose that the recession function $f^\infty$ in (2.5) exists for all $(x, z) \in \bar{\Omega} \times \mathbb{V}$. Then*

$$\mu_j \overset{*}{\rightharpoonup} \mu \text{ in } \mathcal{M}(\Omega, \mathbb{V}) \implies \liminf_{j \to \infty} \int_\Omega f(x, \mathrm{d}\mu_j(x)) \geqslant \int_\Omega f(x, \mathrm{d}\mu(x)).$$

Note that the restriction on the existence of the recession function of $f$ implies both the joint continuity of $f^\infty$ and the linear growth of $f$ from above.

*Proof.* We consider a Young measure $\boldsymbol{\nu}$ generated by $(\mu_j)_j$. Bearing in mind that a convex function is continuous in the interior of its domain, by Proposition A.5 and Jensen's inequality, we have that

$$\begin{aligned}
\liminf_{j \to \infty} \int_\Omega f(x, \mathrm{d}\mu_j(x)) &\geqslant \int_\Omega \langle \nu_x, f(x, \bullet) \rangle \mathrm{d}x + \int_{\bar{\Omega}} \langle \nu_x^\infty, f^\infty(x, \bullet) \rangle \mathrm{d}\lambda \\
&\geqslant \int_\Omega f(x, \bar{\nu}_x) \mathrm{d}x + \int_{\bar{\Omega}} f^\infty(x, \bar{\nu}_x^\infty) \mathrm{d}\lambda \\
&= \int_\Omega \big[ f(x, \bar{\nu}_x) + \lambda^a(x) f^\infty(x, \bar{\nu}_x^\infty) \big] \mathrm{d}x + \int_{\bar{\Omega}} f^\infty(x, \bar{\nu}_x^\infty) \mathrm{d}\lambda^s
\end{aligned}$$

Using the inequality $g(z)+tg^\infty(w)\geqslant g(z+tw)$ which holds for any convex function $g$ and $t\geqslant 0$, we have

$$\begin{aligned}\liminf_{j\to\infty}\int_\Omega f(x,\mathrm{d}\mu_j(x)) &\geqslant \int_\Omega f(x,\bar\nu_x+\lambda^a(x)\bar\nu_x^\infty)\mathrm{d}x+\int_{\bar\Omega} f^\infty(x,\bar\nu_x^\infty)\mathrm{d}\lambda^s\\ &\geqslant \int_\Omega f(x,\mu^a(x))\mathrm{d}x+\int_\Omega f^\infty(x,\mathrm{d}\mu^s)\\ &= \int_\Omega f(x,\mathrm{d}\mu).\end{aligned}$$

where the inequality follows from the nonnegativity of $f$ and Lemma A.4, which implies that $\mu=((\bar\nu_x+\lambda^a(x)\bar\nu_x^\infty)\mathscr{L}^n+\bar\nu_x^\infty\lambda^s)\llcorner\Omega$. □

We next prove two general existence results for convex integrals defined on $\mathrm{BV}_p^{\mathscr{B}}$ spaces. The first one holds for *arbitrary* operators $\mathscr{B}$.

**Theorem 3.2.** *Let us fix $p\in(1,+\infty)$, $g\in L^p(\Omega,\mathbb{U})$, $\overline{\alpha}>\underline{\alpha}>0$ and $\alpha\in\mathrm{C}(\bar\Omega,[\underline{\alpha},\overline{\alpha}])$. Let $f\colon\Omega\times\mathbb{V}\to[0,+\infty)$ be an integrand satisfying the assumptions outlined in Subsection 2.2, and suppose further that the recession function $f^\infty$ in (2.5) exists for all $(x,z)\in\bar\Omega\times\mathbb{V}$. Then, the functional in (2.2) is weakly-$*$ lower semicontinuous and admits a minimizer $u\in\mathrm{BV}_p^{\mathscr{B}}(\Omega)$. If the fidelity term is strictly convex, then the minimizer is unique.*

*Proof.* There is no loss of generality in assuming that $\alpha\equiv 1$. We will employ the direct method in the calculus of variations. Since $I$ is bounded from below, there exists a minimizing sequence $(u_j)_j\subset\mathrm{BV}_p^{\mathscr{B}}(\Omega)$ such that the limit of $I[u_j;\alpha]$ as $j\to+\infty$ is finite. Since $\Phi_g$ is coercive on $L^p(\Omega,\mathbb{U}$ and $f$ satisfies the growth condition in (2.4), $(u_j)_j$ must be bounded in $\mathrm{BV}_p^{\mathscr{B}}(\Omega)$. Thus, on a subsequence that we do not relabel, we have $u_j\rightharpoonup u$ in $L^p(\Omega,\mathbb{U})$ and $\mathscr{B}u_j\overset{*}{\rightharpoonup}\mathscr{B}u$ in $\mathcal{M}(\Omega,\mathbb{V})$. By the weak lower semicontinuity of the fidelity term we have that

$$\Phi_g(u)\leqslant\liminf_{j\to+\infty}\Phi_g(u_j), \tag{3.1}$$

whereas by Proposition 3.1 we obtain

$$\int_\Omega f(x,\mathrm{d}\mathscr{B}u)\leqslant\liminf_{j\to\infty}\int_\Omega f(x,\mathrm{d}\mathscr{B}u_j). \tag{3.2}$$

On the whole, we deduce

$$I[u;\alpha]\leqslant\liminf_{j\to+\infty}I[u_j;\alpha],$$

and we conclude that $u\in\mathrm{BV}_p^{\mathscr{B}}(\Omega)$ is a minimizer of $I$.

Uniqueness follows easily when $\Phi_g$ is strictly convex. Indeed, let $u^1$, $u^2$ be distinct minimizers and $u^0:=(u^1+u^2)/2$. Then, $2\Phi_g(u^0)<\Phi_g(u^1)+\Phi_g(u^2)$, while the convexity of the second term gives

$$2\int_\Omega f(x,\mathrm{d}\mathscr{B}u^0)\leqslant\int_\Omega f(x,\mathrm{d}\mathscr{B}u^1)+\int_\Omega f(x,\mathrm{d}\mathscr{B}u^2).$$

By adding the last two inequalities we infer $I[u^0;\alpha]<\min I$, a contradiction. □

*Remark* 3.3. Notably, in the previous theorem uniqueness holds for $\Phi_g(u)=\|u-g\|^p_{L^p(\Omega)}$. Indeed, for instance by the uniform convexity of the $L^p$ spaces, we have for $u^0:=(u^1+u^2)/2$

$$2\int_\Omega|u^0-g|^p\mathrm{d}x<\int_\Omega|u^1-g|^p\mathrm{d}x+\int_\Omega|u^2-g|^p\mathrm{d}x.$$

*Remark* 3.4. There is no immediate counterpart of Theorem 3.2 when $p = 1$, because in general bounded sequences in $\mathrm{BV}^{\mathscr{B}}$ are not weakly-$*$ precompact. One possibility would be to embed $\mathrm{BV}^{\mathscr{B}}$ in the larger space of measures $\{\mu \in \mathcal{M}(\Omega, \mathbb{U}) : \mathscr{B}\mu \in \mathcal{M}(\Omega, \mathbb{V})\}$. A second option is to assume $\mathscr{B}$ to be $\mathbb{C}$-elliptic, as we do below.

The second existence result involves the smaller class of $\mathbb{C}$-elliptic operators, which was introduced in Definition 2.1. In this case, we are able to treat regularizers that also involve lower order terms, see (1.11), and we can obtain much more precise information on the minimizers. We make the unconventional convention that $\frac{n}{n-k} = +\infty$ if $k \geqslant n$, and we denote by $\mathrm{sym}^i(\mathbb{R}^n, \mathbb{U})$ the space of symmetric $\mathbb{U}$-valued $i$-linear maps on $\mathbb{R}^n$. Also in the following we use the notation $\alpha := (\alpha_1, \dots, \alpha_k)$.

**Theorem 3.5.** *Let us fix $p \in [1, \frac{n}{n-k})$, $g \in L^p(\Omega, \mathbb{U})$, $\overline{\alpha} > \underline{\alpha} > 0$, $\alpha_i \in \mathrm{C}(\bar{\Omega}, [0, \overline{\alpha}])$ for $i = 1, \dots, k-1$ and $\alpha_k \in \mathrm{C}(\bar{\Omega}, [\underline{\alpha}, \overline{\alpha}])$. Let $f_i \colon \Omega \times \mathrm{sym}^i(\mathbb{R}^n, \mathbb{U}) \to \mathbb{R}$, $i = 1, \dots, k-1$, and $f_k \colon \Omega \times \mathbb{V} \to [0, +\infty)$ be Carathéodory integrands such that for all $i$ $f_i(x, \bullet)$ is convex and $f_i(x, \bullet) \leqslant C(1 + |\bullet|)$ for almost every $x \in \Omega$. Suppose in addition that $f_k$ satisfies the coercivity bound $c(|\bullet| - 1) \leqslant f_k(x, \bullet)$ and that*

$$f_k^\infty(x, z) := \lim_{(x', z', t) \to (x, z, +\infty)} \frac{f_k(x', tz')}{t}$$

*exists for all $(x, z) \in \bar{\Omega} \times \mathbb{V}$. Then, if $\mathscr{B}$ is $\mathbb{C}$-elliptic, the functional*

$$\tilde{I}[u; \alpha] := \Phi_g(u) + \sum_{i=1}^{k-1} \int_\Omega \alpha_i(x) f_i(x, \nabla^i u(x)) \mathrm{d}x + \int_\Omega \alpha_k(x) f_k(x, \mathrm{d}\mathscr{B}u). \tag{3.3}$$

*is weakly-$*$ lower semicontinuous in $\mathrm{BV}^{\mathscr{B}}$ and admits a minimizer*

$$u \in \mathrm{BV}^{\mathscr{B}}(\Omega) \cap \mathrm{W}^{k-1, n/(n-1)}(\Omega, \mathbb{U}).$$

*If the fidelity term is strictly convex, then the minimizer is unique.*

*Proof.* If $k = 1$ the statement collapses to Theorem 3.2. The $\mathbb{C}$-ellipticity of $\mathscr{B}$ is still needed to make use of Theorem 2.2, which grants that $u \in L^{n/(n-1)}(\Omega, \mathbb{U})$. We now turn to the case $k \geqslant 2$.

If $(u_j)_j \subset \mathrm{BV}^{\mathscr{B}}(\Omega)$ is a minimizing sequence, then $(u_j)_j$ is bounded in $\mathrm{BV}^{\mathscr{B}}(\Omega)$ by the coercivity assumptions, and hence also in $\mathrm{W}^{k-1, n/(n-1)}(\Omega, \mathbb{U})$ thanks to Theorem 2.2. Let $u \in \mathrm{BV}^{\mathscr{B}}$ be a weak-$*$ limit point of $(u_j)_j$. By the same reasoning as in the proof of Theorem 3.2 we have that (3.1) and (3.2) with $f = f_k$ hold. We now fix $1 \leqslant i \leqslant k-1$ and look at a Young measure $\boldsymbol{\nu}$ generated by $(\nabla^i u_j)_j$, which is bounded in $L^{n/(n-1)}(\Omega)$. By the growth bound on $f_i$ and the de la Vallée Poussin criterion, we have that $(\alpha_i f_i(\bullet, \nabla^i u_j))_j$ is uniformly integrable. We can thus employ Proposition A.6 for $f = \alpha_i f_i$ to obtain

$$\begin{aligned}
\liminf_{j \to \infty} \int_\Omega \alpha_i(x) f_i(x, \nabla^i u_j(x)) \mathrm{d}x &= \int_\Omega \alpha_i(x) \langle \nu_x, f_i(x, \bullet) \rangle \mathrm{d}x \\
&\geqslant \int_\Omega \alpha_i(x) f_i(x, \bar{\nu}_x) \mathrm{d}x \\
&= \int_\Omega \alpha_i(x) f_i(x, \nabla^i u(x)) \mathrm{d}x,
\end{aligned}$$

where we used Jensen's inequality and Lemma A.4. We infer that

$$\liminf_{j \to +\infty} \tilde{I}[u_j; \alpha] \geqslant \tilde{I}[u; \alpha],$$

and we can conclude that $u \in \mathrm{BV}^{\mathscr{B}}(\Omega) \subset \mathrm{W}^{k-1, n/(n-1)}(\Omega, \mathbb{U})$ is a minimizer of $\tilde{I}[\bullet, \alpha]$.

The uniqueness follows exactly by the same argument as in the proof of Theorem 3.2, so the conclusion is achieved. □

*Remark* 3.6. If $\Phi_g(u) = \|u - g\|_{L^1(\Omega)}$, uniqueness might fail in Theorem 3.5.

Our main result reads as follows.

**Theorem 3.7.** *Let us fix $p \in (1, +\infty)$ and $g \in L^p(\Omega, \mathbb{U})$. Let $f\colon \Omega \times \mathbb{V} \to [0, +\infty)$ be an integrand satisfying the assumptions outlined in Subsection 2.2, and suppose further that the recession function $f^\infty$ in (2.5) exists for all $(x, z) \in \bar{\Omega} \times \mathbb{V}$. Then, the training scheme (L1)–(L2) in Subsection 2.2 admits a solution $\alpha^* \in \mathrm{Adm}$ and it provides an associated optimally reconstructed image $u_{\alpha^*} \in \mathrm{BV}^{\mathscr{B}}_p(\Omega)$.*

*Proof.* Let $(\alpha_j)_j \subset \mathrm{Adm}$ be a minimizing sequence for the upper level objective $F$. Under our assumptions on the admissible weights, we may suppose that $\alpha_j \to \alpha^* \in \mathrm{Adm}$ uniformly in $\bar{\Omega}$. To prove the result, it suffices to show that

$$F(u_{\alpha^*}) \leqslant \liminf_{j \to +\infty} F(u_j), \tag{3.4}$$

where we abbreviated $u_j \coloneqq u_{\alpha_j}$ for a minimizer of (L2) with respect to the weight $\alpha_j$, which exists in the light of Theorem 3.2.

We firstly show that $(u_j)$ is weakly-$*$ precompact in $\mathrm{BV}^{\mathscr{B}}_p(\Omega)$. To see this, we observe that by definition of $u_j$ we have

$$I[u_j; \alpha_j] \leqslant I[v; \alpha_j] \quad \text{for any } v \in \mathrm{BV}^{\mathscr{B}}_p(\Omega). \tag{3.5}$$

In particular, by selecting $v = 0$ and recalling that $\|\alpha_j\|_{L^\infty} \leqslant \overline{\alpha}$, we find that $I[u_j; \alpha_j] \leqslant C$ for some $C \geqslant 0$ independent of $j$. Then, owing to the coercivity of $I$, $(u_j)_j$ is bounded in $\mathrm{BV}^{\mathscr{B}}_p(\Omega)$ and there exists $u \in \mathrm{BV}^{\mathscr{B}}_p(\Omega)$ such that, upon extraction of subsequences, $u_j \overset{*}{\rightharpoonup} u$ in $\mathrm{BV}^{\mathscr{B}}_p(\Omega)$. If we prove that $u = u_{\alpha^*}$, the conclusion is then achieved, since (3.4) would then follow by the lower semicontinuity of $F$.

We thus need to show that

$$I[u; \alpha^*] \leqslant I[v; \alpha^*] \quad \text{for any } v \in \mathrm{BV}^{\mathscr{B}}_p(\Omega). \tag{3.6}$$

The uniform convergence of $(\alpha_j)_j$ along with (3.5) yields

$$\liminf_{j \to +\infty} I[u_j; \alpha_j] \leqslant \liminf_{j \to +\infty} I[v; \alpha_j] = I[v; \alpha^*] \quad \text{for any } v \in \mathrm{BV}^{\mathscr{B}}_p(\Omega). \tag{3.7}$$

Further, in view of the growth condition (2.4) and of the bound of $(u_j)_j$ in $\mathrm{BV}^{\mathscr{B}}_p(\Omega)$ we obtain the estimates

$$\begin{aligned} |I[u_j; \alpha_j] - I[u_j; \alpha^*]| &\leqslant \int_\Omega |\alpha_j - \alpha^*| f(x, \mathrm{d}\mathscr{B}u_j) \\ &\leqslant C(1 + |\mathscr{B}u_j|(\Omega))\|\alpha_j - \alpha^*\|_{L^\infty} \\ &\leqslant C\|\alpha_j - \alpha^*\|_{L^\infty}, \end{aligned}$$

whence

$$\lim_{j \to +\infty} \big|I[u_j; \alpha_j] - I[u_j; \alpha^*]\big| = 0. \tag{3.8}$$

Finally, by the lower semicontinuity result in Theorem 3.2,

$$I[u; \alpha^*] \leqslant \liminf_{j \to \infty} I[u_j; \alpha^*], \tag{3.9}$$

so that, collecting (3.7)–(3.9), we obtain (3.6). The proof is thus complete. □

If in the lower level problem (2.9) the functional $I$ is replaced by $\tilde{I}$ as in Theorem 3.5, a result in the same spirit of the one above holds. We only sketch it in the next remark, since it parallels closely Theorem 3.7.

*Remark* 3.8. Within the general framework of Subsection 2.2, we introduce a variant of the scheme (L1)–(L2). For $k \in \mathbb{N}$, $k \geqslant 2$, we define the sets

$$\mathrm{Adm}_{\mathrm{low}} \coloneqq \left\{\alpha \in \mathrm{C}(\bar{\Omega}, [0,\overline{\alpha}]^{k-1}) : |\alpha_i(x) - \alpha_i(y)| \leqslant \omega(|x-y|) \text{ for every } i \text{ and } x, y \in \bar{\Omega}\right\},$$

$$\widetilde{\mathrm{Adm}} \coloneqq \mathrm{Adm}_{\mathrm{low}} \times \mathrm{Adm},$$

where $\omega$ is the same modulus of uniform continuity as in (2.3). We consider the bilevel problem

$$\text{find } \alpha^* \in \operatorname{argmin}\left\{F(u_\alpha) : \alpha \in \widetilde{\mathrm{Adm}}\right\} \tag{3.10}$$

$$\text{such that } u_\alpha \in \operatorname{argmin}\left\{\tilde{I}[u;\alpha] \colon u \in \mathrm{BV}^{\mathscr{B}}(\Omega)\right\}, \tag{3.11}$$

where $\tilde{I}$ is as in (3.3). Under the assumptions of Theorem 3.5, notably $\mathbb{C}$-ellipticity for $\mathscr{B}$, we are able to prove the existence of a solution, that is, an optimal regularizer $\alpha^* \in \widetilde{\mathrm{Adm}}$ for (3.10). Let us outline the argument.

If $(\alpha^j)_j \subset \widetilde{\mathrm{Adm}}$ is a minimizing sequence, we may assume that $\alpha^j \to \alpha^* \in \widetilde{\mathrm{Adm}}$ uniformly. By Theorem 3.5 we can pick a sequence $(u_j)_j \subset \mathrm{BV}^{\mathscr{B}}(\Omega)$ made of minimizers for (3.11) associated with $(\alpha^j)_j$. As a consequence of the coercivity of $\tilde{I}$, $(u_j)_j$ is bounded in $\mathrm{BV}^{\mathscr{B}}_p(\Omega)$, and thus, owing to Theorem 2.2, also in $\mathrm{W}^{k-1,n/(n-1)}(\Omega)$. Denoting by $u \in \mathrm{BV}^{\mathscr{B}}(\Omega)$ the weak-$*$ limit (up to subsequences) of $(u_j)_j$, the remainder of the proof follows the one of Theorem 3.8, the most significant difference being the use of Theorem 3.5 instead of Theorem 3.2 to obtain the analogue of (3.9).

## 4. Examples of spatially inhomogeneous regularization functionals

In order to clarify how our theoretical setup relates to existing contributions, in this section we focus on some previously introduced regularizers which can be expressed via a convex, spatially inhomogeneous function of some differential operator and we show that they fall within our general framework. For the last family of regularizers we discuss here, that is, Huber versions of TV, $\mathrm{TV}^2$ with spatially varying Huber parameter, as well as for the sum $\mathrm{TV} + \mathrm{TV}^2$, we also perform extended numerical examples in Section 5, since no such examples are available in the literature.

**Structural total variation.** In certain image reconstruction problems it may be assumed that some information on the structure of the ground truth is available. It is then reasonable to take into account regularization terms that encode the prior knowledge, and to this purpose a suitable functional framework for *structural total variation* was proposed in [28]. In the case of combined MR-PET reconstructions, for instance, the following regularizer was put forth for $u \in \mathrm{W}^{1,1}(\Omega)$:

$$J[u] \coloneqq \int_\Omega \sqrt{1 - \eta^2 \left(w(x) \cdot \frac{\nabla u(x)}{|\nabla u(x)|}\right)^2}\, |\nabla u(x)| \mathrm{d}x. \tag{4.1}$$

Here, $w \in \mathrm{L}^1_{\mathrm{loc}}(\Omega, \mathbb{R}^2)$ is a vector field that records information on hand (in this case a reconstructed MR image) and has modulus strictly smaller than 1, whereas $\eta \in (0,1)$ is a parameter that ensures coercivity:

$$J[u] \geqslant \sqrt{1-\eta^2} \int_\Omega |\nabla u(x)| \mathrm{d}x.$$

For a general $w$, such regularizer is outside the scopes of our theory. However, upon mollification, we can suppose that $w \in \mathrm{C}(\bar{\Omega}, \mathbb{R}^2)$, and it is easy to see that under this

assumptions the integrand

$$f_w(x,z) \coloneqq \Big[|z|^2 - \eta^2\big(w(x)\cdot z\big)^2\Big]^{\frac{1}{2}} = \Big[1-\eta^2\left(w(x)\cdot \hat{z}\right)^2\Big]^{\frac{1}{2}}|z|, \qquad \text{with } \hat{z} \coloneqq \frac{z}{|z|},$$

admits a recession function as in (2.5), which coincides with $f_w$ itself. Therefore, when $w$ is continuous, by relaxation results in BV (see e.g. [2, Theorem 5.54] and references therein), the $L^1$-relaxation of the functional in (4.1) is given for $u \in \mathrm{BV}(\Omega)$ by

$$\bar{J}[u] = \int_\Omega f_w\big(x, \nabla u(x)\big)\mathrm{d}x + \int_\Omega f_w\left(x, \frac{\mathrm{D}^s u}{\mathrm{d}|\mathrm{D}u|}(x)\right)\mathrm{d}|\mathrm{D}u|(x).$$

Different interesting aspects emerge from our approach to this specific example. Firstly, by taking into account the explicit formula for the relaxation, which was instead left implicit in [28], we can integrate the functional in (4.1) into our general theory. Secondly, Theorem 3.2 provides an alternative proof of the existence of solutions to the primal problem in [28, Theorem 5.4]. Lastly, by Theorem 3.7, we are able to embed such problem in a bilevel scheme and solve it.

**Directional total variation.** Still in the spirit of employing prior information on data, a notion of *directional total variation* was introduced in [38] in relation to fibered materials. Pictorial representations of such materials feature textures with a specific alignment and call for regularizations that feature directionality. Given the parameters $a \in (0,1]$ and $\theta \in [0, 2\pi)$, for $u \in \mathrm{BV}(\Omega)$, $\Omega \subset \mathbb{R}^2$, the authors of [38] addressed this issue by introducing the functional

$$\mathrm{DTV}(u) \coloneqq \sup\left\{\int_\Omega u\, \mathscr{B}^*_{a,\theta}\phi\,\mathrm{d}x : \phi \in \mathrm{C}^\infty_c(\Omega,\mathbb{R}^2), \|\phi\|_\infty \leqslant 1\right\},$$

with

$$\begin{aligned}\mathscr{B}^*_{a,\theta}\phi &\coloneqq -\begin{pmatrix}\cos(\theta)\\ -a\sin(\theta)\end{pmatrix}\cdot\partial_1\phi - \begin{pmatrix}\sin(\theta)\\ a\cos(\theta)\end{pmatrix}\cdot\partial_2\phi \\ &= -\operatorname{trace}\left[\begin{pmatrix}\cos(\theta) & -a\sin(\theta)\\ \sin(\theta) & a\cos(\theta)\end{pmatrix}\nabla\phi\right].\end{aligned} \tag{4.2}$$

In image reconstruction problems the parameters $a$ and $\theta$ are encode the prior information about the geometry of the datum $g$ and are suitably fixed before running the restoration algorithm [38]. We note that, in wider generality, DTV may be defined analogously for $u \in L^p(\Omega)$. Then, it easy to see that $\mathrm{DTV}(u)$ is finite if and only if $u \in \mathrm{BV}^{\mathscr{B}_{a,\theta}}_p(\Omega)$, where

$$\mathscr{B}_{a,\theta}u \coloneqq \begin{pmatrix}\cos(\theta)\\ -a\sin(\theta)\end{pmatrix}\partial_1 u + \begin{pmatrix}\sin(\theta)\\ a\cos(\theta)\end{pmatrix}\partial_2 u. \quad \text{for every } u \in \mathrm{C}^1_{\mathrm{c}}(\mathbb{R}^2,\mathbb{R}).$$

Therefore, our theory encompasses image reconstruction problems in which the regularization term is DTV.

Note that we could replace the matrix that weights $\nabla\phi$ in (4.2) with any another matrix $M$, thereby obtaining a weighted divergence operator in the spirit of [45]. The fundamental difference with this contribution is that our treatment does not cover the case of space-dependent fields $M\colon \Omega \to \mathbb{R}^{2\times 2}$, because that would give rise to an operator $\mathscr{B}$ with nonconstant coefficients. Investigations in this direction are surely an attractive topic for future research. On the positive side, we point out that we can however deal with some sort of directional total variation. Indeed, for $u \in \mathrm{BV}(\Omega)$, regularizers of the form

$$\int_\Omega f(x, \mathrm{D}u) \quad \text{with } f(x,z) \coloneqq |M(x)z|, M\colon \Omega \to \mathbb{R}^{2\times 2} \text{ measurable},$$

are included in our theory, provided a recession function exists.

**Spatially varying Huber versions of TV and TV$^2$.** Huber versions of total variation have already been used, either to reduce the staircasing effect of TV or to facilitate the use of second order numerical solvers for the corresponding minimization problem [12, 13, 34]. However, in all the previous approaches the Huber parameter $\gamma$ is always considered to be a fixed scalar quantity. Here we describe how our framework can accommodate definitions of Huber versions of the TV and TV$^2$ functionals where the Huber parameter is spatially distributed.

Let $\gamma \in L^\infty(\Omega)$, with $\gamma \geq 0$, be fixed. We define the spatially varying Huber function $f_\gamma \colon \Omega \times \mathbb{R}^n \to [0, +\infty)$ as follows:

$$f_\gamma(x,z) = \begin{cases} |z| - \dfrac{1}{2}\gamma(x), & \text{if } |z| \geq \gamma(x), \\ \dfrac{1}{2\gamma(x)}|z|^2, & \text{if } |z| < \gamma(x). \end{cases} \tag{4.3}$$

Obviously, for all $z \in \mathbb{R}^n$ and for almost all $x \in \Omega$, $f_\gamma$ satisfies the coercivity and growth conditions in (2.4), namely

$$|z| - \|\gamma\|_{L^\infty(\Omega)} \leq f_\gamma(x,z) \leq |z|. \tag{4.4}$$

Then, if $u \in \mathrm{BV}(\Omega)$, we define the ensuing convex function of the measure $\mathrm{D}u$ with the alternative notations

$$\mathrm{TV}_\gamma(u) \coloneqq |f_\gamma(\mathrm{D}u)|(\Omega) \coloneqq \int_\Omega f_\gamma(x, d\mathrm{D}u).$$

An easy check shows that the recession function of $f_\gamma$ (cf. (2.5)) is $f_\gamma^\infty(x,z) = |z|$. Thus, all the assumptions of Theorem 3.7 are trivially satisfied. Consequently, $\mathrm{TV}_\gamma$ is indeed well-defined as

$$\mathrm{TV}_\gamma(u) = \int_\Omega f_\gamma(x, \nabla u)\mathrm{d}x + \int_\Omega d|\mathrm{D}^s u|,$$

and for $\alpha \in C(\bar{\Omega})$ with $\alpha(x) \geq \underline{\alpha} > 0$ we can define its weighted version

$$\mathrm{TV}_{\alpha,\gamma}(u) = \int_\Omega \alpha f_\gamma(x, \nabla u)\mathrm{d}x + \int_\Omega \alpha\, d|\mathrm{D}^s u|. \tag{4.5}$$

Note that $\mathrm{TV}_{\alpha,\gamma}(u)$ can be equivalently defined via duality [24]:

$$\mathrm{TV}_{\alpha,\gamma}(u) = \sup\left\{ \int_\Omega u \,\mathrm{div}\phi \,\mathrm{d}x - \mathcal{I}_{\{|\bullet(x)|\leq\alpha(x)|\}}(\phi) - \frac{1}{2}\int_\Omega \frac{\gamma}{\alpha}|\phi|^2 \,\mathrm{d}x : \ \phi \in \mathrm{C}_c^\infty(\Omega, \mathbb{R}^n) \right\}. \tag{4.6}$$

This is an immediate extension of the standard Huber TV functional mentioned in the introduction. Similarly, for a function $u \in \mathrm{BV}^2(\Omega) \coloneqq \{u \in W^{1,1}(\Omega) : \mathrm{D}^2 u \in \mathcal{M}(\Omega, \mathcal{S}^{n\times n})\}$ and $\alpha \in \mathrm{C}(\bar{\Omega})$ with $\alpha(x) \geqslant \underline{\alpha} > 0$, we define the spatially varying Huber TV$^2$ functional as

$$\mathrm{TV}^2_{\alpha,\gamma}(u) = \int_\Omega \alpha f_\gamma(x, \nabla^2 u)\,\mathrm{d}x + \int_\Omega \alpha\,\mathrm{d}|(\mathrm{D}^2)^s u|, \tag{4.7}$$

where $f_\gamma$ now is a function defined on $\Omega \times \mathbb{R}^{n\times n}$ defined by the natural analogue of (4.3). Again via duality, we find the equivalent expression

$$\mathrm{TV}^2_{\alpha,\gamma}(u) = \sup\left\{ \int_\Omega u \,\mathrm{div}^2\phi \,\mathrm{d}x - \mathcal{I}_{\{|\bullet(x)|\leq\alpha(x)\}}(\phi) - \frac{1}{2}\int_\Omega \frac{\gamma}{\alpha}|\phi|^2 \,\mathrm{d}x : \ \phi \in \mathrm{C}_c^\infty(\Omega, \mathbb{R}^n) \right\}. \tag{4.8}$$

## 5. Numerical examples for spatially varying Huber versions of TV and $TV^2$

Grounding on the analysis above, in this section we provide some numerical results for image reconstruction. Specifically, since this case has not been considered in the literature, we focus on the last class of examples of the previous section, that is, Huber versions of TV and $TV^2$ with spatially varying Huber parameter $\gamma$, as well as their sum $TV + TV^2$. We stress that, while the purpose of our numerical examples is to show the applicability and versatility of our bilevel framework involving spatially inhomogeneous regularization functionals, it also turns out that our regularizers are able to yield results that are comparable, and in certain cases even better, than the ones obtained by using some standard high quality regularizers, such as the Total Generalized Variation (TGV) [8] and its version with spatially varying weights [31]. Since our main target here is to evaluate the performance of these regularizers, we restrict ourselves to three particular cases of image denoising. Firstly, in the class of first-order functionals, we consider the Huber-type TV regularization, with both the regularization parameter $\alpha_1$ and the Huber parameter $\gamma$ being spatially dependent. As we saw in the previous section, this can be thought as a functional that incorporates a local choice between TV and Tikhonov regularization. The second example is a Huber $TV^2$ regularization again with both the regularization parameter $\alpha_2$ and the Huber parameter $\gamma$ being spatially dependent. $TV^2$ was introduced in the literature as a second-order functional with the capability to eliminate the undesirable staircasing effect of TV [4, 44]. Even though in theory the $TV^2$ regularization is not able to preserve sharp edges, we will see that its spatially varying version produces high quality results and can even outperform both the scalar and the spatially varying versions of TGV. Finally, to be consistent with the analytical results above, which also deal with sums of regularization functionals, we provide numerical examples for the sum $TV + TV^2$ with spatially varying regularization parameters $\alpha_1$ and $\alpha_2$, whose scalar version was introduced in [44]. In summary, our examples concern the following lower level image denoising problems:

$$I^1_\gamma[u;\alpha_1] \coloneqq \int_\Omega (u-g)^2 \mathrm{d}x + \mathrm{TV}_{\alpha_1,\gamma}(u), \tag{5.1}$$

$$I^2_\gamma[u;\alpha_1] \coloneqq \int_\Omega (u-g)^2 \mathrm{d}x + \mathrm{TV}^2_{\alpha_2,\gamma}(u). \tag{5.2}$$

$$I^{1,2}_\gamma[u;\alpha_1,\alpha_2] \coloneqq \int_\Omega (u-g)^2 \mathrm{d}x + \mathrm{TV}_{\alpha_1,\gamma}(u) + \mathrm{TV}^2_{\alpha_2,\gamma}(u). \tag{5.3}$$

5.1. **The bilevel problems.** The family of bilevel problems for the automatic computation of the spatial regularization parameter $\alpha_1$, $\alpha_2$ associated with the functionals above are

$$\text{find } \alpha^* \in \operatorname{argmin}\{F(u_\alpha) :\ \alpha \in \mathrm{Adm}\} \tag{5.4}$$

$$\text{such that } u_\alpha = \operatorname{argmin}\left\{I_\gamma[u;\alpha] \colon u \in \mathrm{BV}^i(\Omega)\right\}. \tag{5.5}$$

where $I_\gamma$ is either of the functionals $I^1_\gamma$, $I^2_\gamma$, $I^{1,2}_\gamma$, $i = 1$ for the $\mathrm{TV}_{\alpha_1,\gamma}$ and $i = 2$ for either of $\mathrm{TV}^2_{\alpha_2,\gamma}$ and $\mathrm{TV}_{\alpha_1,\gamma} + \mathrm{TV}^2_{\alpha_2,\gamma}$ cases respectively. Correspondingly, $\alpha$ denotes either set of regularization weights $\{\alpha_1\}$, $\{\alpha_2\}$ or $\{\alpha_1,\alpha_2\}$.

In view of (4.4), the lower level problems (5.5) are well-defined, while from Theorem 3.7 we know that the overall schemes (5.4)–(5.5) admit a solution for the two alternative upper level objectives $F$ considered next. As we discussed in the introduction and repeat here for the sake of reading flow, we take into account two alternatives for the upper level objective functional $F$:

$$F_{\mathrm{PSNR}}(u) = \int_\Omega |u - u_{\mathrm{gt}}|^2 \mathrm{d}x, \tag{5.6}$$

$$F_{\text{stat}}(u) = \frac{1}{2}\int_\Omega \max(Ru - \overline{\sigma}^2, 0)^2 \mathrm{d}x + \frac{1}{2}\int_\Omega \min(Ru - \underline{\sigma}^2, 0)^2 \mathrm{d}x, \tag{5.7}$$

$$\text{where } Ru(x) \coloneqq \int_\Omega w(x,y)(u-g)^2(y)\mathrm{d}y \quad \text{for } w \in L^\infty(\Omega\times\Omega), \quad \int_\Omega\int_\Omega w(x,y)\mathrm{d}x\mathrm{d}y = 1.$$

The first cost functional corresponds to a maximization of the PSNR of the reconstruction and requires the knowledge of the ground truth $u_{\text{gt}}$ [14, 22, 21, 20, 41], while the second enforces the localized residuals $Ru$ to belong in a certain tight corridor $[\underline{\sigma}^2, \overline{\sigma}^2] \coloneqq [\sigma^2 - \epsilon, \sigma^2 + \epsilon]$, $\sigma^2$ being the variance of the noise $\eta$, which is assumed here to be Gaussian, see also [32, 33, 29, 31]. The latter option has the advantage of being ground truth free, but knowledge or a good estimate for the noise variance $\sigma^2$ is needed. For the discrete version of the averaging filter $w$ in the definition of the localized residuals (5.7) we use a filter of size $n_w \times n_w$, with entries of equal value that sum to one.

Since a numerical projection to the admissible set Adm is not practical, here we also follow [32, 33, 29, 31] and add instead a small $H^1$ term of the weight function $\alpha$ in the upper level objective, together with a supplementary box constraint $\mathcal{C} \coloneqq \{\alpha \in H^1(\Omega) : \underline{\alpha} \le \alpha \le \overline{\alpha}\}$ for some $\underline{\alpha}, \overline{\alpha} \in \mathbb{R}$ with $0 < \underline{\alpha} < \overline{\alpha}$. On the whole, we will use the following upper level objectives:

$$\mathcal{F}_{\text{PSNR}}(\alpha, u) = \int_\Omega |u - u_{\text{gt}}|^2 \mathrm{d}x + \frac{\lambda}{2}\|\alpha\|^2_{H^1(\Omega)},$$

$$\mathcal{F}_{\text{stat}}(\alpha, u) = \frac{1}{2}\int_\Omega \max(Ru - \overline{\sigma}^2, 0)^2 \mathrm{d}x + \frac{1}{2}\int_\Omega \min(Ru - \underline{\sigma}^2, 0)^2 \mathrm{d}x + \frac{\lambda}{2}\|\alpha\|^2_{H^1(\Omega)},$$

for some small $\lambda > 0$. We will denote by $\hat{\mathcal{F}}$ the corresponding reduced objective functionals, that is $\hat{\mathcal{F}}_{\text{PSNR/stat}}(\alpha) \coloneqq \mathcal{F}_{\text{PSNR/stat}}(\alpha, u_\alpha)$. That leads us to the bilevel minimization problems that we tackle numerically:

$$\text{find } \alpha^* \in \operatorname*{argmin}_\alpha \mathcal{F}_{\text{PSNR/stat}}(\alpha, u_\alpha) \tag{5.8}$$

$$\text{such that } \begin{cases} u_\alpha = \operatorname*{argmin}_u I_\gamma[u;\alpha], \\ \alpha \in \mathcal{C}. \end{cases} \tag{5.9}$$

Note that in this setting it is not guaranteed that $\alpha \in C(\bar{\Omega})$, since $H^1(\Omega)$ does not embed in that space for dimensions higher than 1. However, one can take advantage of the following regularity result of the $H^1$-projection onto $\mathcal{C}$, denoted by $P_{\mathcal{C}}$ see [33, Corollary 2.3] for a proof.

**Proposition 5.1.** *Let $\Omega \subset \mathbb{R}^\ell$ with $\ell = 1, 2, 3$ be a bounded convex set and let $\mathcal{C} = \{\alpha \in H^1(\Omega) : \underline{\alpha} \le \alpha \le \overline{\alpha}\}$, where $\underline{\alpha}, \overline{\alpha} \in H^2(\Omega)$ are such that $\underline{\alpha} \le \overline{\alpha}$ and $\frac{\partial \underline{\alpha}}{\partial \boldsymbol{\nu}} = \frac{\partial \overline{\alpha}}{\partial \boldsymbol{\nu}} = 0$ in $H^{1/2}(\partial\Omega)$. If*

$$\omega \in H^2(\Omega) \quad \textit{and} \quad \frac{\partial \omega}{\partial \boldsymbol{\nu}} = 0$$

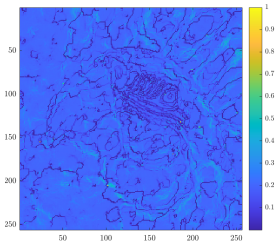
*and if*

$$\omega^* \coloneqq P_{\mathcal{C}}(\omega) \coloneqq \operatorname*{argmin}_{\alpha\in\mathcal{C}} \frac{1}{2}\|\alpha - \omega\|^2_{H^1(\Omega)},$$

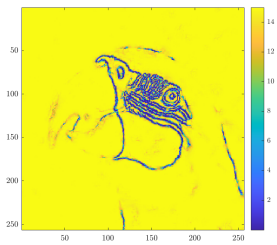
*then it holds*

$$\omega^* \in H^2(\Omega) \quad \textit{and} \quad \frac{\partial \omega^*}{\partial \boldsymbol{\nu}} = 0.$$

The projection $P_{\mathcal{C}}$ is applied to every iteration of the projected gradient algorithm, which is to be used for the numerical solution of (5.8)–(5.9) and is described next, see Algorithm 1. In that case, if $\underline{\alpha}, \overline{\alpha}$ as well as the initializations for the weights $\alpha^0$ are constant functions, it is ensured that the computed weight $\alpha^k$ at the $k$-th projected gradient iteration belongs to $H^2(\Omega)$, which for $n = 2$ embeds compactly into any Hölder space $\mathrm{C}^{0,\beta}(\bar{\Omega})$, $\beta \in (0, 1)$.

Huber TV denoising (left) with discontinuous weight $\alpha_1$ (right)
PSNR=26.84, SSIM=0.7547

Tikhonov denoising (left) with discontinuous weight $\tilde{\alpha}$ (right)
PSNR=27.87, SSIM=0.8361

FIGURE 1. TV and Tikhonov denoising with discontinuous regularization weights computed via the bilevel framework using $\mathcal{F}_{\text{stat}}$ as upper level objective. In the case of Huber TV denoising, the computed weight function is too irregular, having small values in arbitrary places, preserving part of the noise. On the other hand, in Tikhonov regularization, the weight remains large in parts of the edge set and as a result the edges are not preserved there.

As observed in [31, 33], where this type of upper level objective is also used, we also remark that in order to produce a meaningful spatially varying weight, its regularity should not be too strong, and hence a relatively small value for $\lambda$ should be chosen. As long as it is chosen relatively small, its precise value is not crucial and does not significantly affect the computed regularization weight, hence neither the image reconstruction.

We commented in Subsection 2.3 on the need of continuity for the regularization weights from the theoretical perspective. It is worth of note that such regularity is also relevant from the numerical point of view. From a practical imaging standpoint, we found that this regularity for the weights produces more stable reconstructions. For instance, in Figure 1 (left) we see the denoised image $u$ and the weight $\alpha$ that are computed by solving the bilevel problem with $\mathcal{F}_{\text{stat}}$ as an upper level objective upon substituting the $H^1$ regularization for $\alpha$ with a simple $L^2$ term. In that case, the corresponding projection is just a $L^2$ projection (pointwise thresholding), and the regularity result of Proposition 5.1 does not hold. We thus observe that the computed weight turns out to be too irregular, having small values in arbitrary points, and as a result part of the noise is preserved and the staircasing effect is pronounced. The relationship between the regularity of the weight in TV regularization and the creation of artifacts in the reconstruction was also studied in detail in [30, 36].

5.2. **Strategy for fixing $\gamma$.** Since in our set up the function $\gamma$ is not part of the minimizing variables, it has to be fixed from the start. Our rationale for fixing $\gamma$ is that we would like to regularize high detailed areas with a Tikhonov term with a spatially varying weight $\frac{1}{2}\int_\Omega \tilde{\alpha}|\nabla u|^2 \mathrm{d}x$, with $\tilde{\alpha}$ having as low regularity as possible, e.g. $L^\infty(\Omega)$, in order to increase flexibility in the regularization. In the other areas we would like to regularize using a TV or $\mathrm{TV}^2$ term with a spatially varying weight $\alpha$. This will happen if $\gamma$ is large in such detailed areas in order to allow for the second case in (4.3) and small otherwise. We thus adopt the following strategy: We first solve an auxiliary bilevel problem with a weighted Tikhonov regularizer using the upper level objective $\mathcal{F}_{\text{stat}}$. The output is a spatially varying $\tilde{\alpha}$ that essentially acts as an edge detector, see Figure 1 (right), since it is small on the edges and on the detailed areas of the image. We then invert this weight and set

$$\gamma = s\frac{1}{\tilde{\alpha}} \tag{5.10}$$

for some constant $s > 0$. By choosing the function $\gamma$ as in (5.10), we have that when $\tilde{\alpha}$ is small (fine scale details), $\gamma$ will be large and thus the second case in (4.3) will be selected with a weight $\frac{1}{2\gamma(x)} = \frac{s}{2}\tilde{\alpha}$ in front of the term $|\nabla u|^2$. On the other hand, when $\tilde{\alpha}$ is large,

then $\gamma$ will be small and thus a TV or $\mathrm{TV}^2$ term will be preferred, i.e., first case in (4.3). In the third images of the top rows of Figures 2 and 4, we see how the resulting $\gamma$ function looks like for the example images. Details for the computation of $\tilde{\alpha}$ via the auxiliary bilevel Tikhonov problem are given in the next section. We note however that, instead of solving a bilevel weighted Tikhonov problem to compute $\tilde{\alpha}$ and hence $\gamma$, one could alternatively employ some standard edge detector algorithms, like for instance the Canny method [15].

We finally remark that an attempt to combine the best characteristics of TV and Tikhonov regularization by considering their direct sum, that is, by considering the following lower lewel problem

$$\min_{u\in H^1(\Omega)} \int_\Omega (u-g)^2 \mathrm{d}x + \int_\Omega \alpha_1 \,\mathrm{d}|\mathrm{D}u| + \frac{1}{2}\int_\Omega \tilde{\alpha}|\nabla u|^2 \mathrm{d}x, \tag{5.11}$$

and employ a bilevel framework to automatically compute $\alpha_1$ and $\tilde{\alpha}$, would not work in practice. This can be seen already in Figure 1, where the lower level problem that corresponds to (5.11), by fixing $\alpha_1 = 0$, i.e., purely weighted Tikhonov, is considered. There, even though the computed (discontinuous) weight $\tilde{\alpha}$, takes small values in the detailed part of the image, it does not do so in the whole edge set. As a result, at some part of the edge set, $\tilde{\alpha}$ is significantly large and thus there the reconstruction is smooth. This can not be remedied by simply adding a TV term, since that term cannot reduce the regularity of the image. Nevertheless, as we described above, this computed $\tilde{\alpha}$ can still be used in order to define an appropriate spatially varying $\gamma$.

5.3. **Numerical algorithm for solving the bilevel problems.** In this section we describe the algorithm that we use for the numerical solution of the discrete versions of the bilevel problems (5.8)–(5.9). Similar algorithms were presented in [33, 31], so we limit ourselves to a brief description of the procedure, still providing all the necessary details to ensure reproducibility.

The lower level problems (5.1), (5.2) and (5.3) are substituted by their primal-dual optimality conditions, see [34, 31], and they respectively read:

$$u - g - \mathrm{div} p_1 = 0, \tag{5.12}$$

$$\max(|\nabla u|, \gamma) p_1 - \alpha_1 \nabla u = 0, \tag{5.13}$$

$$u - g + \mathrm{div}^2 p_2 = 0, \tag{5.14}$$

$$\max(|\nabla^2 u|, \gamma) p_2 - \alpha_2 \nabla^2 u = 0, \tag{5.15}$$

and

$$u - g - \mathrm{div} p_1 + \mathrm{div}^2 p_2 = 0, \tag{5.16}$$

$$\max(|\nabla u|, \gamma) p_1 - \alpha_1 \nabla u = 0, \tag{5.17}$$

$$\max(|\nabla^2 u|, \gamma) p_2 - \alpha_2 \nabla^2 u = 0, \tag{5.18}$$

with $p_1$, $p_2$ denoting the corresponding dual variables of each problem. For notation ease, we compactly write each of the above equations as $G(u, p) = 0$, where $p$ denotes either $p_1$, $p_2$ or both of these for the third problem. The application of the max function as well the multiplication in (5.13) and (5.15) are regarded component wise; note that here both $\alpha$ and $\gamma$ are spatially (i.e., pixel) dependent. We use standard forward and backward differences for the discretizations of $\nabla$ and div, see e.g. [31], and similarly for the discretizations of $\nabla^2$ and $\mathrm{div}^2$, see [44]. For the numerical solution of each of the problems (5.12)–(5.13), (5.14)–(5.15) and (5.16)–(5.18), we use a semismooth Newton algorithm as it is described in [31] for the TGV case. Note that we do not add an additional Laplacian term for $u$ as in [34, 31], and we do not smooth the max function. We terminate the semismooth Newton iterations when the Euclidean norm of both residuals is less than $10^{-4}$.

In order to solve the minimization problems in (5.8)–(5.9), where the lower level problems are substituted by either of (5.12)–(5.13), (5.14)–(5.15) and (5.16)–(5.18), we employ a discretized projected gradient approach with Armijo line search as it is described in [31], originated from [33]. The algorithm is summarized in Algorithm 1. We comment on the components that have not been clarified so far. The term $\Delta_N$ denotes the discrete Laplacian with zero Neumann boundary conditions. These are the desired boundary conditions for $\alpha$, as dictated by the regularity results for the $H^1$-projection $P_{\mathcal{C}}$. This projection is computed exactly as in [31] by using the same method and parameters mentioned there. In our numerical computations we set $\underline{a} = 10^{-8}$, $\overline{a} = 5$, $n_w = 7$, $\lambda = 10^{-11}$, $\tau^0 = 10^{-3}$, $c = 10^{-12}$, $\theta_- = 0.25$, $\theta_+ = 2$. As initializations for $\alpha_1$ and $\alpha_2$, we use the constant functions $\alpha_1^0 = 0.5$ and $\alpha_2^0 = 1$. Regarding $\underline{\sigma}, \overline{\sigma}$, we use the formulas $\underline{\sigma}^2 = \sigma^2(1 - \frac{\sqrt{2}}{n_w})$ and $\overline{\sigma}^2 = \sigma^2(1 + \frac{\sqrt{2}}{n_w})$, which are based on the statistics of the extremes, see [33, Section 4.2.1]. In all our noisy images, the Gaussian noise has zero mean and variance $\sigma^2 = 0.01$. We terminate Algorithm 1 after a fixed number of iterations maxit = 100, after which no noticeable reduction in the reduced objective function is observed, see also Figure 7.

In order to produce a spatially varying Huber parameter $\gamma$ as described before, we solve an auxiliary bilevel Tikhonov problem where the lower level problem corresponds to $G_T(u) \coloneqq u - \operatorname{div}(\tilde{\alpha}\nabla u) - g = 0$, i.e., the first order optimality condition of a variational denoising problem with $\frac{1}{2}\int_\Omega \tilde{\alpha}|\nabla u|^2 \mathrm{d}x$ as regularizer and $L^2$ fidelity term. In order to do so, we utilize again the projected gradient algorithm described in Algorithm 1, adjusted to this regularizer. We use no additional $H^1$ regularization for $\gamma$ and the $H^1$-projection $P_{\mathcal{C}}$ is substituted by a simple $L^2$ projection, that is, $\tilde{\alpha}^{k+1} = \max(\min(\tilde{\alpha}^k - \tau^k \nabla_{\tilde{\alpha}} \hat{\mathcal{F}}(\tilde{\alpha}^k), \overline{\alpha}), \underline{\alpha})$. We use 100 projected gradient iterations with $n_w$, $\underline{\sigma}, \overline{\sigma}$, $\tau^0$, $c$, $\theta_-$, $\theta_+$ as before, as well as $\underline{\alpha} = 10^{-8}$, $\overline{\alpha} = 15$, $\alpha_0 = 15$. The equation $G_T$ is solved exactly with a linear system solver. We use again 100 projected gradient iterations to get an output weight $\tilde{a}$. Then, as we mentioned in (5.10) we define $\gamma = s/\tilde{\alpha}$. In all our experiments, we set $s = 0.1$.

For comparison purposes, we also report TV and TV$^2$ denoising results with a scalar Huber parameter $\gamma$, which is always set $\gamma = 10^{-3}$. We do that for both scalar and spatially varying regularization parameters $\alpha_1$ and $\alpha_2$. In the first case, we manually select the parameter $\alpha$ that maximizes the PSNR of the denoised image, computed with a semismooth Newton method as previously mentioned. The second case is computed exactly as in Algorithm 1. We also report the TGV reconstructions, both scalar and spatially varying versions, which are computed with the Chambolle-Pock primal-dual method [17] as described in [5]. For the scalar case, again we manually select the TGV parameters $(\alpha_0, \alpha_1)$ that maximize the PSNR. For the spatially varying case, we use the weights $(\alpha_0, \alpha_1)$ as produced in [31] for the same image examples we are considering here. In that work, these weights were computed via the ground truth-free bilevel approach but using a regularized lower level problem, i.e. with additional $H_1$ regularizations in the primal variables of the TGV minimization problem. We remark that it turns out that when these weights are directly fed into the Chambolle-Pock algorithm for the non-regularized problem as we do here, they produce a result of higher quality, hence the discrepancy between the PSNR and SSIM values we report here and then ones reported in [31].

5.4. **Numerical results.** We focus initially on the Huber versions of TV and TV$^2$. In Figure 2 we report our numerical results on the *Parrot* image, see also Figure 3 for zoom-in details. Here the spatially varying regularization weights $\alpha_1$ and $\alpha_2$ are produced with the ground truth-free bilevel approach, i.e., using $\mathcal{F}_{\text{stat}}$ as an upper lever objective. Among the regularizers with scalar parameters, second row, first three images, the best reconstruction both in terms of PSNR and SSIM is achieved by the scalar TGV. The bilevel Huber TV reconstruction with spatially varying $\alpha_1$ and scalar $\gamma$ is able to better preserve the details around the eye of the parrot, third row first image. When we use the spatially varying

**Algorithm 1**
Projected gradient for the bilevel Huber TV / $\mathrm{TV}^2$ / $\mathrm{TV} + \mathrm{TV}^2$ problems

**Input**: $g$, $\underline{\alpha}$, $\overline{\alpha}$, $\underline{\sigma}$, $\overline{\sigma}$, $\lambda$, $\gamma$, $n_w$, $\tau^0$, $0 < c < 1$, $0 < \theta_- < 1 \leq \theta_+$
**Initialize**: $\alpha^0 \in \mathcal{C}$, and set $k = 0$.
**repeat**
Use Semismooth Newton to solve (5.12)–(5.13) or (5.14)–(5.15) or (5.16)–(5.17), i.e.

$$G(u^k, p^k) = 0$$

Solve for $(u^*, p^*)$ the adjoint equation

$$(D_{(u,p)}G(u^k, p^k))^\top (u^*, p^*) = -D_u \mathcal{F}(u^k, \alpha^k)$$

Compute the derivative of the reduced objective w.r.t. $\alpha$ as

$$\hat{\mathcal{F}}'(\alpha^k) = (D_\alpha G(\alpha^k))^\top (u^*, p^*) + D_\alpha \mathcal{F}(\alpha^k)$$

Compute the reduced gradient

$$\nabla_\alpha \hat{\mathcal{F}}(\alpha^k) = (I - \Delta_N)^{-1} \hat{\mathcal{F}}'(\alpha^k)$$

Compute the trial points

$$\alpha^{k+1} = P_{\mathcal{C}}\big(\alpha^k - \tau^k \nabla_\alpha \hat{\mathcal{F}}(\alpha^k)\big),$$

**while**

$$\hat{\mathcal{F}}(\alpha^{k+1}) > \hat{\mathcal{F}}(\alpha^k) + c\big(\hat{\mathcal{F}}'(\alpha^k)^\top (\alpha^{k+1} - \alpha^k)\big)$$

**do** (Armijo line search)
Set $\tau^k \coloneqq \theta_- \tau^k$, and re-compute

$$\alpha^{k+1} = P_{\mathcal{C}}\big(\alpha^k - \tau^k \nabla_\alpha \hat{\mathcal{F}}(\alpha^k)\big)$$

**end while**
Update $\tau^{k+1} = \theta_+ \tau^k$, and $k \coloneqq k + 1$
**until** some stopping condition is satisfied

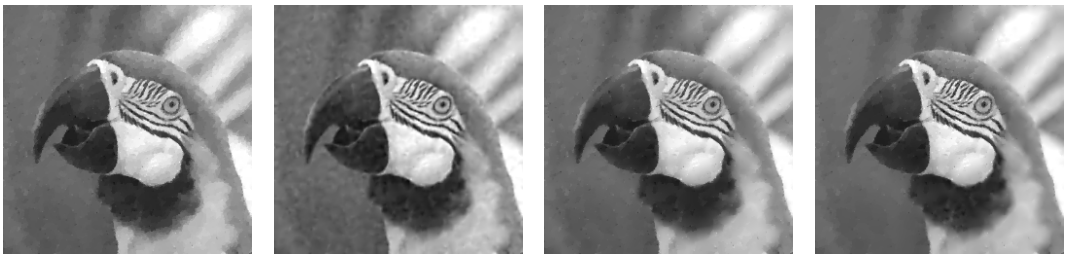
$\gamma$, the details in that area become even more pronounced, compare the first two images in the third row of Figure 3. This is also accompanied with a slight increase of the SSIM index but also with a decrease in PSNR. Observe that the weights $\alpha_1$ that are computed in these two cases are quite different, see first two images of the last row of Figure 2. The bilevel Huber $\mathrm{TV}^2$ approach with spatially varying $\alpha_2$, produces similar reconstructions for both the scalar (slighly higher SSIM) and the spatially varying $\gamma$ case (slightly higher PSNR). These reconstructions are of very good quality and even outperform the spatially varying TGV in terms of SSIM, having also the same PSNR. This is due to the fact that the combination of the statistics-based upper level objective and the second order TV is forcing the weight $\alpha_2$ to drop significantly in the detailed areas of the image, see the last two images of the last row of Figure 2. It is characteristic that while the PSNR of scalar $\mathrm{TV}^2$ is only 0.03 dB higher than the one of scalar TV, the PSNR of bilevel Huber $\mathrm{TV}^2$ with spatially varying $\alpha_2$ and scalar $\gamma$ is 0.61 dB higher compared to the corresponding Huber TV result.

The superiority of the bilevel Huber $\mathrm{TV}^2$ with spatially varying weights, is even more evident in the second image example *Hatchling*, Figures 4 and 5. Here the reconstruction is more challenging due to the oscillatory nature of the ground truth image. The bilevel Huber $\mathrm{TV}^2$ with scalar $\gamma$ gives by far the best result with respect to both PSNR and SSIM. Again, the automatically computed regularization weights $\alpha_2$ have much lower values in bilevel $\mathrm{TV}^2$ than in bilevel TV, compare the first two versus the last two figures of the last row of Figure 4. In this example, the spatially varying $\gamma$ leads to a reduction of PSNR

Ground truth
PSNR=$\infty$, SSIM=1.000

Gaussian noise, $\sigma^2 = 0.01$
PSNR=20.04, SSIM=0.2773

Spatially varying $\gamma$



scalar Huber TV
PSNR=29.25, SSIM=0.8354

scalar Huber TV$^2$
PSNR=29.28, SSIM=0.8305

scalar TGV
PSNR=29.50, SSIM=0.8509

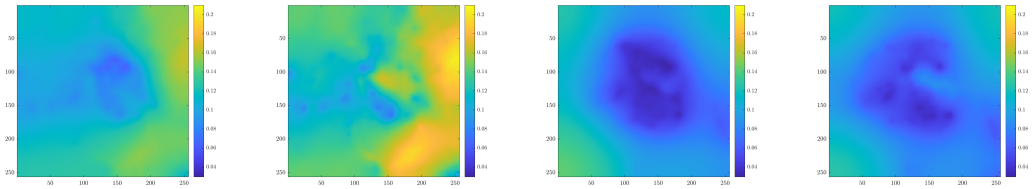
Bilevel TGV
with spatially varying $\alpha$ and $\beta$
PSNR=**29.84**, SSIM=0.8606

Bilevel Huber TV
with spatially varying $\alpha_1$
and scalar $\gamma$
PSNR=29.20, SSIM=0.8549

Bilevel Huber TV
with spatially varying $\alpha_1$
and spatially varying $\gamma$
PSNR=28.92, SSIM=0.8571

Bilevel Huber TV$^2$
with spatially varying $\alpha_2$
and scalar $\gamma$
PSNR=29.81, SSIM=**0.8705**

Bilevel Huber TV$^2$
with spatially varying $\alpha_2$
and spatially varying $\gamma$
PSNR=**29.84**, SSIM=0.8700



Weight $\alpha_1$ of Huber TV

Weight $\alpha_1$ of Huber TV

Weight $\alpha_2$ of Huber TV$^2$

Weight $\alpha_2$ of Huber TV$^2$

FIGURE 2. *Parrot* image: Huber TV and TV$^2$ denoising with spatially varying Huber parameter $\gamma$ and regularization parameters $\alpha_1$ and $\alpha_2$ respectively. The weights $\alpha_1$ and $\alpha_2$ are produced with the ground truth-free bilevel approach using $\mathcal{F}_{\text{stat}}$. The highest PSNR and SSIM values are highlighted in bold font.

and SSIM in all cases, but nevertheless also to more highlighted details in the eye area, see second and fourth images of the last row of Figure 5.

In order to verify further the regularization capabilities of these regularizers, we make another series of experiments with these two example images, using the ground truth-based upper level objective $\mathcal{F}_{\text{PSNR}}$, see Figure 6. In both images, the highest PSNR and SSIM is achieved by the bilevel TV$^2$ with spatially varying $\alpha_2$ and scalar Huber parameter $\gamma$, third

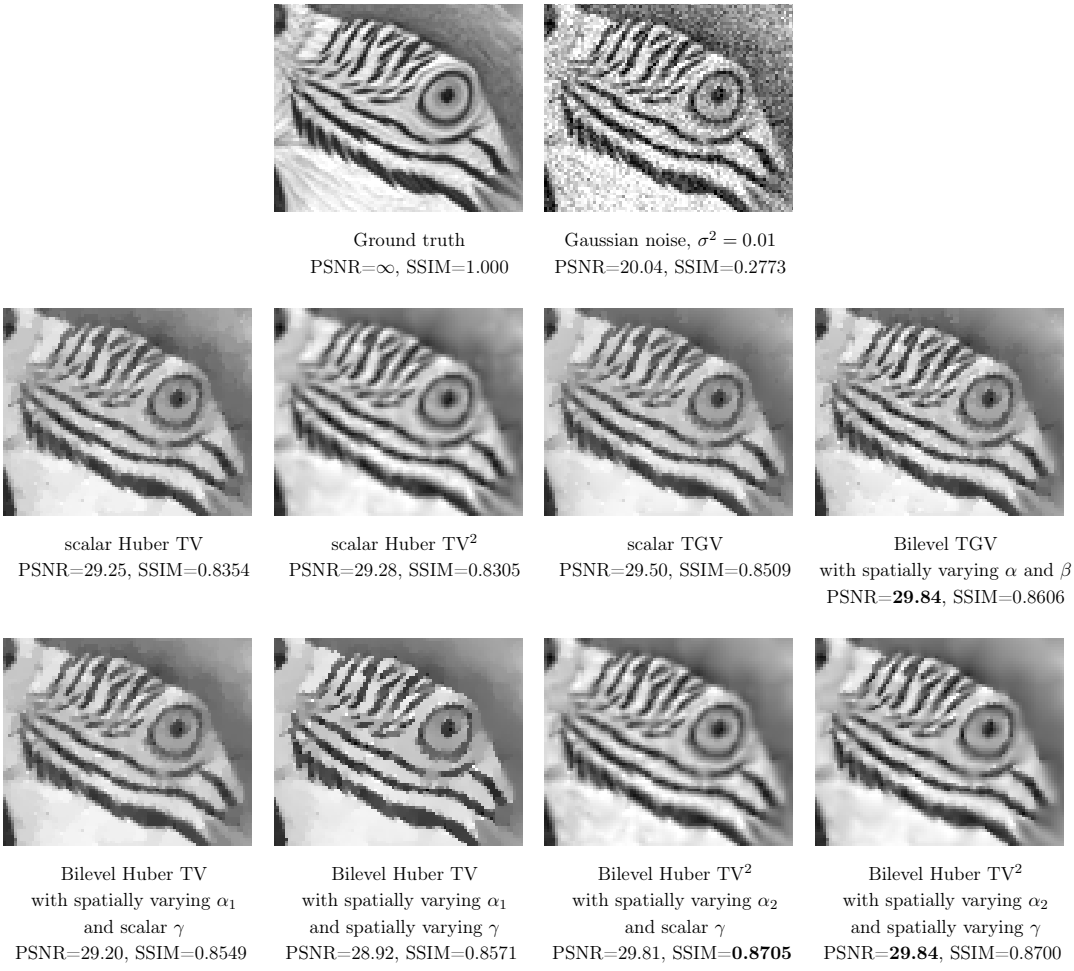
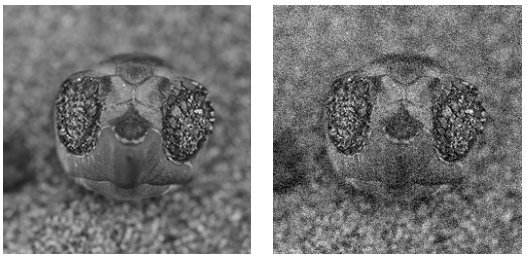



FIGURE 3. Details of images shown in Figure 2

images of first and third row, with the corresponding the regularization weight $\alpha_2$ having again smaller values compared to the TV one. Nevertheless, we observe that the spatially varying $\gamma$ results in higher SSIM in the Huber TV examples in both images, again with more pronounced features around the eye.

In Figure 7, we have plotted the values of the reduced objective $\hat{\mathcal{F}}(u^k)$ along the projected gradient iterations, for all bilevel Huber TV and TV$^2$ examples. The top row shows these plots for the reduced statistics-based upper level objective $\hat{\mathcal{F}}_{\text{stat}}$. We observe that in both images, the introduction of the spatially varying $\gamma$ in both Huber TV and Huber TV$^2$ functionals, helps towards a further reduction of this objective, compare red versus green and blue versus black plots. We observed already that in some cases this is accompanied with a larger SSIM index and more pronounced details in the images, but in most cases the PSNR in decreased. This is in accordance with the plots of the second row, where we see that the reduced PSNR-maximizing upper level objective $\hat{\mathcal{F}}_{\text{PSNR}}$ is not further decreased by the introduction of the spatially varying $\gamma$, compare again the red versus green and blue versus black plots.

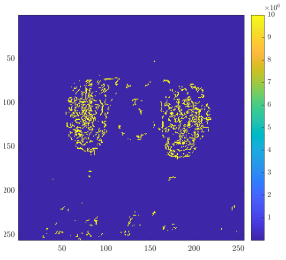
Finally, we provide numerical examples using (5.3) as a lower level problem, that is, the direct sum of TV and TV$^2$, each one weighted with a spatially varying parameter, respectively $\alpha_1$ and $\alpha_2$. We did so for a scalar choice of $\gamma = 10^{-3}$ and for both choices of the upper level objectives, $\mathcal{F}_{\text{stat}}$ and $\mathcal{F}_{\text{PSNR}}$. As we see in Figure 8, for the parrot image, this TV-TV$^2$ approach produces the best PSNR and the best SSIM among both the ground truth-free (using $\mathcal{F}_{\text{stat}}$) and the ground truth-based ($\mathcal{F}_{\text{PSNR}}$) reconstructions. This is not



Ground truth
PSNR=$\infty$, SSIM=1.000



Gaussian noise, $\sigma^2 = 0.01$
PSNR=20.00, SSIM=0.3349



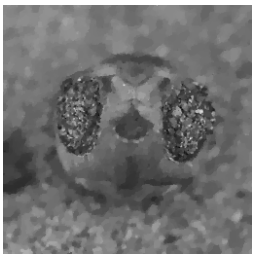
Spatially varying $\gamma$

scalar Huber TV
PSNR=27.75, SSIM=0.7701

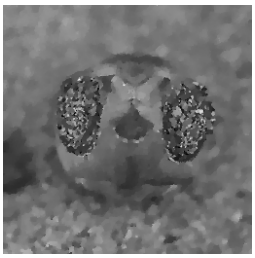
scalar Huber TV$^2$
PSNR=28.22, SSIM=0.8142

scalar TGV
PSNR=28.20, SSIM=0.8132

Bilevel TGV
with spatially varying $\alpha$ and $\beta$
PSNR=28.33, SSIM=0.8145



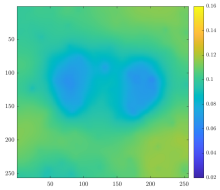
Bilevel Huber TV
with spatially varying $\alpha_1$
and scalar $\gamma$
PSNR=27.50, SSIM=0.7702



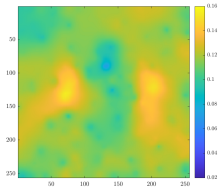
Bilevel Huber TV
with spatially varying $\alpha_1$
and spatially varying $\gamma$
PSNR=27.15, SSIM=0.7688

Bilevel Huber TV$^2$
with spatially varying $\alpha_2$
and scalar $\gamma$
PSNR=**28.66**, SSIM=**0.8367**

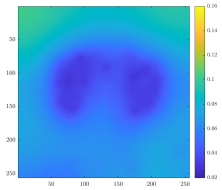
Bilevel Huber TV$^2$
with spatially varying $\alpha_2$
and spatially varying $\gamma$
PSNR=28.44, SSIM=0.8285



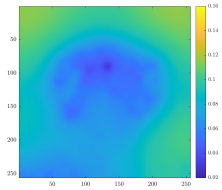
Weight $\alpha_1$ of Huber TV



Weight $\alpha_1$ of Huber TV



Weight $\alpha_2$ of Huber TV$^2$



Weight $\alpha_2$ of Huber TV$^2$

FIGURE 4. *Hatchling* image: Huber TV and TV$^2$ denoising with spatially varying Huber parameter $\gamma$ and regularization parameters $\alpha_1$ and $\alpha_2$ respectively. The weights $\alpha_1$ and $\alpha_2$ are produced with the ground truth-free bilevel approach using $\mathcal{F}_{\text{stat}}$. The highest PSNR and SSIM values are highlighted in bold font.

the case for the hatchling image, where the spatially varying TV$^2$ result is equally good to the TV-TV$^2$ one (both with PSNR=29.09) when the ground-truth based upper level objective is used. This is also reflected in the corresponding computed weights, see last images of the second and third row in Figure 8. The weight $\alpha_1$ of TV is significantly smaller than the one of TV$^2$, essentially rendering the TV regularization effect negligible and suggesting that the TV$^2$ regularization is preferred. On the other hand, when using

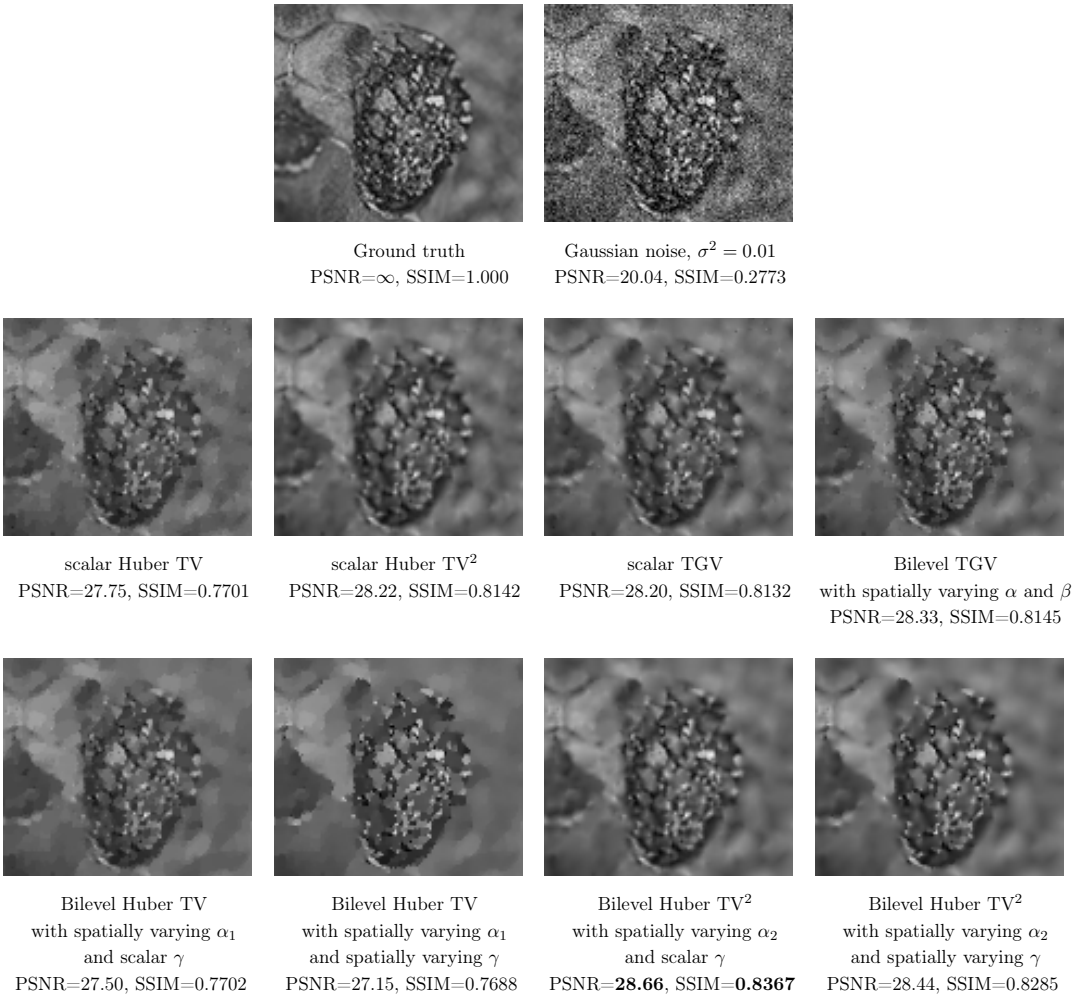



FIGURE 5. Details of images shown in Figure 4

the ground-truth free upper level objective $\mathcal{F}_{\text{stat}}$, the pure TV$^2$ reconstruction has a higher PSNR than the one of TV-TV$^2$ (28.66 versus 28.41). We stress again that minimization of $\mathcal{F}_{\text{stat}}$ does not necessarily correspond to maximization of the PSNR and the TV-TV$^2$ result could still have less $\mathcal{F}_{\text{stat}}$-energy than the the pure TV$^2$ one.

An overview of the PNSR and SSIM values for all the different reconstructions is provided in Table 1. In general, we conclude that the bilevel Huber TV$^2$ and sometimes the additive version of TV-TV$^2$ are able to produce remarkably good results. This is perhaps even surprising, as the use of their scalar versions is not that popular due to its inability to preserve sharp edges. We showed that the use of a spatially varying Huber parameter $\gamma$ can bring about improved results both quantitatively and qualitatively. We also stress that by no means our strategy for setting $\gamma$ is necessarily the optimal one. In fact, future work will involve setting up a bilevel framework where also this parameter is included in the upper level minimization variables along with the parameter $\alpha$, adding further flexibility to the regularization process. Finally we remark again that the purpose of this numerical section was to justify our rigorous analytical study on spatially inhomogeneous integrands acting on TV-type regularizers by choosing a particular example from the ones that were identified in Section 4. As we pointed there, several regularizers commonly used in image processing fit in our theoretical framework.

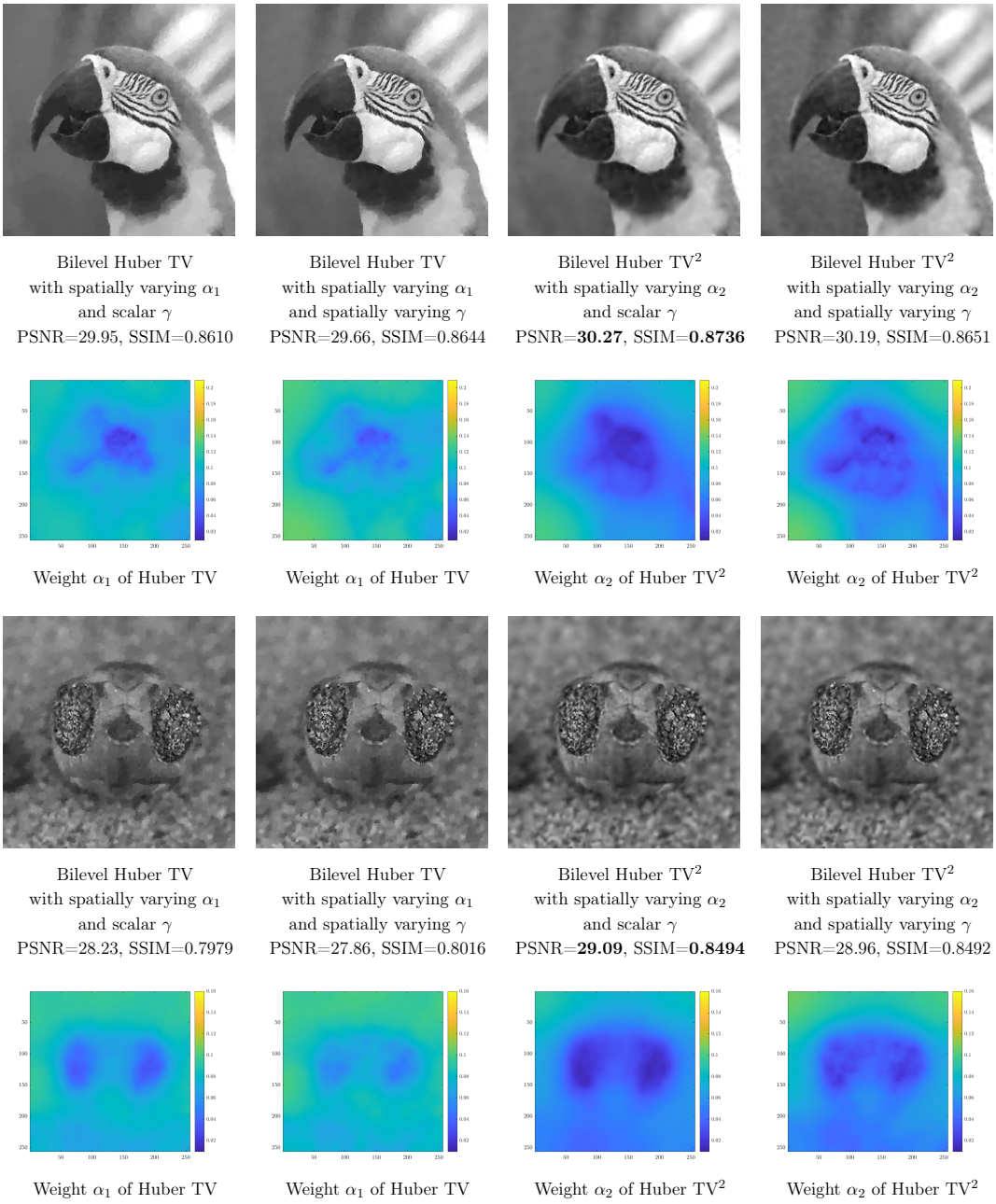
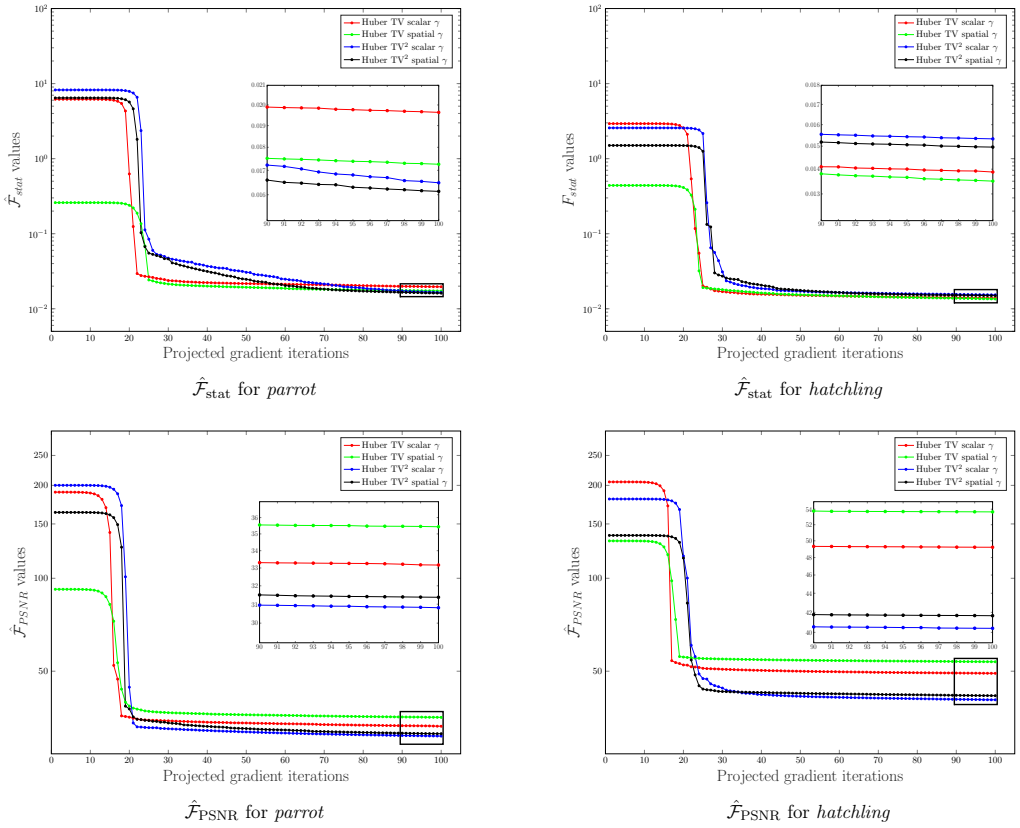



FIGURE 6. Huber TV and TV$^2$ denoising with spatially varying Huber parameter $\gamma$ and regularization parameters $\alpha_1$ and $\alpha_2$ respectively. The weights $\alpha_1$ and $\alpha_2$ are produced with the ground truth-based bilevel approach using $\mathcal{F}_{\text{PSNR}}$. The highest PSNR and SSIM values are highlighted in bold font.

## APPENDIX A. CONVEX INTEGRANDS AND GENERALIZED YOUNG MEASURES

This section loosely follows [39], where most proofs can be found. Let $\Omega \subset \mathbb{R}^n$ be a bounded open set with $\mathscr{L}^n(\partial\Omega) = 0$ and consider the space of integrands

$$\mathbb{E}(\Omega, \mathbb{V}) := \left\{ \Phi \in \mathrm{C}(\Omega \times \mathbb{V}) \colon \Phi^\infty(x, z) := \lim_{t \to \infty,\, x' \to x} \frac{\Phi(x', tz)}{t} \in \mathbb{R} \text{ uniformly in } \bar{\Omega} \times S_{\mathbb{V}} \right\},$$




$\hat{\mathcal{F}}_{\text{stat}}$ for *parrot* $\hat{\mathcal{F}}_{\text{stat}}$ for *hatchling*

$\hat{\mathcal{F}}_{\text{PSNR}}$ for *parrot* $\hat{\mathcal{F}}_{\text{PSNR}}$ for *hatchling*

FIGURE 7. Values of the reduced objective $\hat{\mathcal{F}}(u^k)$ along the projected gradient iterations. The inner boxes show zoom-in plots of the last 10 iterations.

which is naturally equipped with the norm

$$\|\Phi\|_{\mathbb{E}} \coloneqq \sup_{(x,z)\in\Omega\times\mathbb{V}} \frac{|\Phi(x,z)|}{1+|z|}.$$

It will thus be convenient to work with the coordinate transformations

$$S\colon \hat{z}\in B_{\mathbb{V}} \mapsto \frac{\hat{z}}{1-|\hat{z}|}\in\mathbb{V}, \quad S^{-1}\colon z\in\mathbb{V}\mapsto \frac{z}{1+|z|}\in B_{\mathbb{V}},$$

where we wrote $B_{\mathbb{V}}$ to denote the open unit ball in $\mathbb{V}$. With this notation, the space of integrands $\mathbb{E}(\Omega,\mathbb{V})$ can be identified with $\mathrm{C}(\overline{\Omega\times B_{\mathbb{V}}})$ via the linear isometric isomorphism

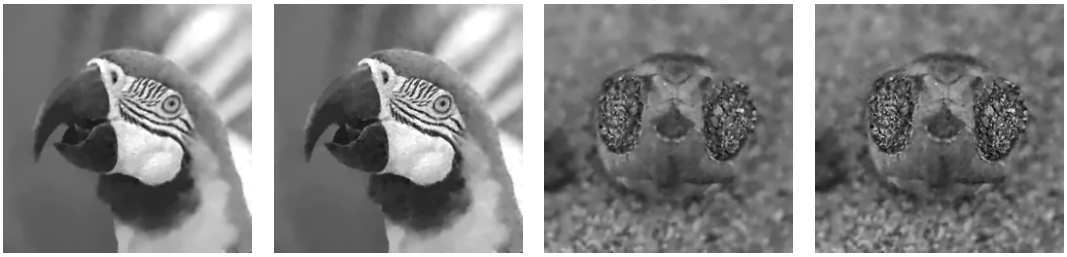
$$(T\Phi)(x,\hat{z}) \coloneqq (1-|\hat{z}|)\Phi\left(x,\frac{\hat{z}}{1-|\hat{z}|}\right), \quad \text{for } x\in\Omega,\ \hat{z}\in B_{\mathbb{V}}.$$

It follows that its adjoint, $T^*\colon \mathbb{E}(\Omega,\mathbb{V})^* \to \mathrm{C}(\overline{\Omega\times B_{\mathbb{V}}})^* \cong \mathcal{M}(\overline{\Omega\times B_{\mathbb{V}}})$ is also a linear isometric isomorphism. We embed $\mathcal{M}(\Omega,\mathbb{V})$ into $\mathbb{E}^*$ via

$$\boldsymbol{\varepsilon}_\mu(\Phi) \coloneqq \int_\Omega \Phi(x,\mathrm{d}\mu) = \int_\Omega \Phi(x,\mu^a(x))\mathrm{d}x + \int_\Omega \Phi^\infty\left(x,\frac{\mathrm{d}\mu^s}{\mathrm{d}\mu}(x)\right)\mathrm{d}|\mu|(x),$$

where $\mu=\mu^a\mathscr{L}^n\llcorner\Omega+\mu^s$ is the Radon–Nýkodim decomposition of $\mu\in\mathcal{M}(\Omega,\mathbb{V})$. By the sequential Banach–Alaoglu theorem, we can infer that bounded $L^p$ sequences are weakly-* compact in $\mathbb{E}^*$ under the above identification. In particular, if $(\mu_j)$ is bounded in $\mathcal{M}(\Omega,\mathbb{V})$, we know that along a subsequence we have $\boldsymbol{\varepsilon}_{\mu_j}\overset{*}{\rightharpoonup}\boldsymbol{\nu}$ in $\mathbb{E}(\Omega,\mathbb{V})^*$. We define $\sigma\coloneqq(T^{-1})^*\boldsymbol{\nu}\in\mathcal{M}(\overline{\Omega\times B_{\mathbb{V}}})$ and write for $\Phi\in\mathbb{E}(\Omega,\mathbb{V})$

$$\langle\!\langle\Phi,\boldsymbol{\nu}\rangle\!\rangle \coloneqq \langle\Phi,\boldsymbol{\nu}\rangle_{\mathbb{E},\mathbb{E}^*} = \langle T\Phi,\sigma\rangle$$

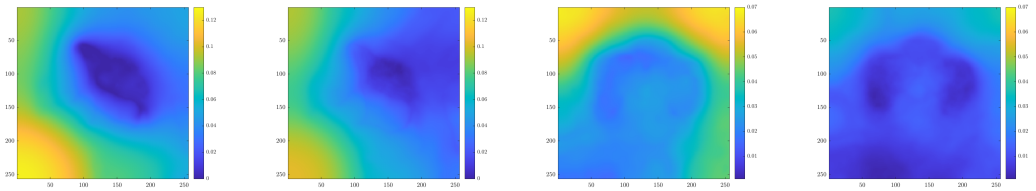


Bilevel Huber TV-TV$^2$ ($\mathcal{F}_{\text{stat}}$) with spatially varying $\alpha_1$ & $\alpha_2$ PSNR=29.93, SSIM=0.8750

Bilevel Huber TV-TV$^2$ ($\mathcal{F}_{\text{PSNR}}$) with spatially varying $\alpha_1$ & $\alpha_2$ PSNR=30.52, SSIM=0.8803

Bilevel Huber TV-TV$^2$ ($\mathcal{F}_{\text{stat}}$) with spatially varying $\alpha_1$ & $\alpha_2$ PSNR=28.41, SSIM=0.8205

Bilevel Huber TV-TV$^2$ ($\mathcal{F}_{\text{PSNR}}$) with spatially varying $\alpha_1$ & $\alpha_2$ PSNR=29.09, SSIM=0.8466

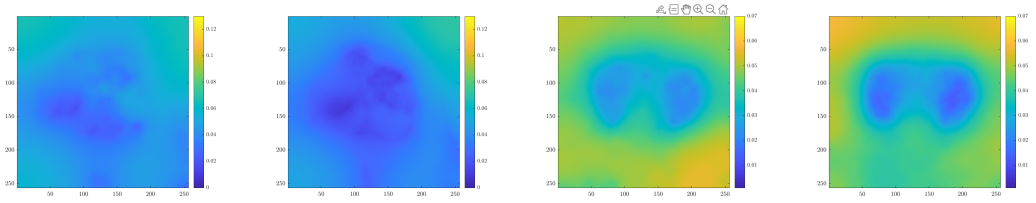


Weight $\alpha_1$ of Huber TV

Weight $\alpha_1$ of Huber TV

Weight $\alpha_1$ of Huber TV

Weight $\alpha_1$ of Huber TV



Weight $\alpha_2$ of Huber TV$^2$

Weight $\alpha_2$ of Huber TV$^2$

Weight $\alpha_2$ of Huber TV$^2$

Weight $\alpha_2$ of Huber TV$^2$

FIGURE 8. Huber TV-TV$^2$ denoising with spatially varying regularization parameters $\alpha_1$ and $\alpha_2$ and scalar $\gamma$. The weights $\alpha_1$ and $\alpha_2$ are produced both with the ground truth-free and the ground truth-based bilevel approach using $\mathcal{F}_{\text{stat}}$ and $\mathcal{F}_{\text{PSNR}}$ respectively.

$$= \int_{\bar{\Omega}\times B_{\mathbb{V}}} (1-|\hat{z}|)\Phi\left(x, \frac{\hat{z}}{1-|\hat{z}|}\right) \mathrm{d}\sigma(x.\hat{z}) + \int_{\bar{\Omega}\times S_{\mathbb{V}}} \Phi^{\infty}(x,\hat{z})\mathrm{d}\sigma(x,\hat{z}).$$

From this formula we derive two necessary conditions for the weakly-* limits of $\boldsymbol{\varepsilon}_{\mu_j}$, namely that $\sigma \geqslant 0$ in the sense of $\mathcal{M}(\overline{\Omega \times B_{\mathbb{V}}})$ and

$$\int_{\Omega} \varphi(x)\mathrm{d}x = \int_{\bar{\Omega}\times B_{\mathbb{V}}} \varphi(x)(1-|\hat{z}|)\mathrm{d}\sigma(x,\hat{z}) \text{ for all } \varphi \in \mathrm{C}(\bar{\Omega}). \tag{A.1}$$

Conversely, these conditions are sufficient to enable us to disintegrate $\sigma$ into appropriately parametrized (generalized Young) measures that detect both oscillation and concentration behavior of a weakly-* convergent sequence $(\mu_j)$. We define:

**Definition A.1.** *A parametrized measure* $\boldsymbol{\nu} = \big((\nu_x)_{x\in\Omega}, \lambda, (\nu_x^{\infty})_{x\in\bar{\Omega}}\big)$ *is said to be a* Young measure *(or* generalized Young measure*) whenever*

(a) $(\nu_x)_{x\in\Omega} \subset \mathcal{M}_1^+(\mathbb{V})$ *is weakly-* $\mathscr{L}^n$-measurable (the* oscillation measure*).*
(b) $\lambda \in \mathcal{M}^+(\bar{\Omega})$ *(the* concentration measure*).*
(c) $(\nu_x^{\infty})_{x\in\bar{\Omega}} \subset \mathcal{M}_1^+(\mathbb{V})$ *is weakly-* $\lambda$-measurable (the* concentration-angle measure*).*
(d) $\int_{\Omega}\int_{\mathbb{V}} |z|\mathrm{d}\nu_x(z)\mathrm{d}x < \infty$ *(the* moment condition *holds).*

*Then* $\boldsymbol{\nu}$ *acts linearly on* $\mathbb{E}(\Omega,\mathbb{V})$ *via*

$$\langle\!\langle \Phi, \boldsymbol{\nu} \rangle\!\rangle := \int_{\Omega}\int_{\mathbb{V}} \Phi(x,\bullet)\mathrm{d}\nu_x\mathrm{d}x + \int_{\bar{\Omega}}\int_{S_{\mathbb{V}}} \Phi^{\infty}(x,\bullet)\mathrm{d}\nu_x^{\infty}\mathrm{d}\lambda(x) \quad \textit{for } \Phi \in \mathbb{E}(\Omega,\mathbb{V}).$$

| | | Parrot | | Hatchling | |
|---|---|---|---|---|---|
| Upper level | Regularizer | PSNR | SSIM | PSNR | SSIM |
| - | TV sc. $\alpha_1$, sc. $\gamma$ | 29.25 | 0.8354 | 27.75 | 0.7701 |
| | $\mathrm{TV}^2$ sc. $\alpha_2$, sc. $\gamma$ | 29.28 | 0.8305 | 28.22 | 0.8142 |
| | TGV sc. $\alpha_0$, sc. $\alpha_1$ | 29.50 | 0.8509 | 28.20 | 0.8132 |
| $\mathcal{F}_{\text{stat}}$ | TGV s.v. $\alpha_0$, s.v. $\alpha_1$ | 29.84 | 0.8606 | 28.33 | 0.8145 |
| | TV s.v. $\alpha_1$, sc. $\gamma$ | 29.20 | 0.8549 | 27.50 | 0.7702 |
| | TV s.v. $\alpha_1$, s.v. $\gamma$ | 28.92 | 0.8571 | 27.15 | 0.7688 |
| | $\mathrm{TV}^2$ s.v. $\alpha_1$, sc. $\gamma$ | 29.81 | 0.8705 | **28.66** | **0.8367** |
| | $\mathrm{TV}^2$ s.v. $\alpha_1$, s.v. $\gamma$ | 29.84 | 0.8700 | 28.44 | 0.8285 |
| | TV-$\mathrm{TV}^2$ s.v. $\alpha_1$, s.v. $\alpha_2$, sc. $\gamma$ | **29.93** | **0.8750** | 28.41 | 0.8205 |
| $\mathcal{F}_{\text{PSNR}}$ | TV s.v. $\alpha_1$, sc. $\gamma$ | 29.95 | 0.8610 | 28.23 | 0.7979 |
| | TV s.v. $\alpha_1$, s.v. $\gamma$ | 29.66 | 0.8644 | 27.86 | 0.8016 |
| | $\mathrm{TV}^2$ s.v. $\alpha_1$, sc. $\gamma$ | 30.27 | 0.8736 | **29.09** | **0.8494** |
| | $\mathrm{TV}^2$ s.v. $\alpha_1$, s.v. $\gamma$ | 30.19 | 0.8651 | 28.96 | 0.8492 |
| | TV-$\mathrm{TV}^2$ s.v. $\alpha_1$, s.v. $\alpha_2$, sc. $\gamma$ | **30.52** | **0.8803** | **29.09** | 0.8466 |

TABLE 1. Summary of the PSNR and SSIM values for all the previous results. The abbreviations "sc." and "s.v." mean "scalar" and "spatially varying" respectively. The highest values for every upper level objective are highlighted with bold font.

*We write* $\mathrm{Y}(\Omega,\mathbb{V})$ *for the set of all such* $\boldsymbol{\nu}$.

It is then easy to check that a Young measure $\boldsymbol{\nu}$ actually lies in $\mathbb{E}^*$ and, moreover, that the inclusion $\mathrm{Y}(\Omega,\mathbb{V})\subset\mathbb{E}(\Omega,\mathbb{V})^*$ is strict. We have the disintegration theorem:

**Theorem A.2.** $\mathrm{Y}(\Omega,\mathbb{V})=T^*\{\sigma\in\mathcal{M}^+(\overline{\Omega\times B_{\mathbb{V}}})\colon$ *equation* (A.1) *holds*$\}$.

Consequently, $\mathrm{Y}(\Omega,\mathbb{V})$ is weakly-* closed in $\mathbb{E}(\Omega,\mathbb{V})^*$ and convex, therefore:

**Theorem A.3** (Fundamental Theorem of Young measures)**.** *Let* $(\mu_j)$ *be a bounded sequence in* $\mathcal{M}(\Omega,\mathbb{V})$. *Then there exists* $\boldsymbol{\nu}\in\mathrm{Y}(\Omega,\mathbb{V})$ *such that, along a subsequence,* $\boldsymbol{\varepsilon}_{\mu_j}\overset{*}{\rightharpoonup}\boldsymbol{\nu}$ *in* $\mathbb{E}(\Omega,\mathbb{V})^*$, *i.e.,*

$$\lim_{j\to\infty}\int_\Omega\Phi(x,\mathrm{d}\mu_j)=\int_\Omega\int_{\mathbb{V}}\Phi(x,z)\mathrm{d}\nu_x(z)\mathrm{d}x+\int_{\bar\Omega}\int_{S_{\mathbb{V}}}\Phi^\infty(x,z)\mathrm{d}\nu_x^\infty(z)\mathrm{d}\lambda(x)$$

*for all* $\Phi\in\mathbb{E}(\Omega,\mathbb{V})$. *In this case, we say that* $(\mu_j)$ generates $\boldsymbol{\nu}$.

In our analysis we repeatedly use the following:

**Lemma A.4.** *Let* $(\mu_j)\subset\mathcal{M}(\Omega,\mathbb{V})$ *generate a Young measure* $\boldsymbol{\nu}$. *Then*

$$\mu_j\overset{*}{\rightharpoonup}\bar\nu_x\mathscr{L}^n\llcorner\Omega+\bar\nu_x^\infty\lambda\ \text{ in }\mathcal{M}(\bar\Omega,\mathbb{V}).$$

*The limit measure is refered to as the* barycentre *of* $\boldsymbol{\nu}$.

This follows simply by taking $\Phi(x,z)=\varphi(x)z_i$ for $\varphi\in\mathrm{C}(\bar\Omega)$, where we wrote

$$\bar\nu_x=\int_{\mathbb{V}}z\mathrm{d}\nu_x(z)\quad\text{and}\quad\bar\nu_x^\infty=\int_{B_{\mathbb{V}}}z\mathrm{d}\nu_x^\infty(z)$$

for the expectations of the probability measures $\nu_x$ and $\nu_x^\infty$.

We also employ a general convergence result for Young measures:

**Proposition A.5.** [40, Prop. 2(i)] *Let $\Omega \subset \mathbb{R}^n$ be bounded and open, and $f\colon \Omega \times \mathbb{V} \to \mathbb{R}$ be a measurable integrand such that $f(x, \bullet)$ is continuous for almost every $x \in \Omega$ (a Carathéodory integrand). Suppose in addition that $f$ has a regular recession function, i.e.,*

$$f^\infty(x,z) \coloneqq \lim_{(x',z',t)\to(x,z,+\infty)} \frac{f(x',tz')}{t} \quad \text{for } (x,z) \in \bar{\Omega} \times \mathbb{V}$$

*exists. Let $(\mu_j) \subset \mathcal{M}(\Omega, \mathbb{V})$ generate $\boldsymbol{\nu} \in \mathrm{Y}(\Omega, \mathbb{V})$. Then*

$$\lim_{j\to\infty} \int_\Omega f(x, \mathrm{d}\mu_j(x)) = \int_\Omega \langle \nu_x, f(x, \bullet)\rangle \mathrm{d}x + \int_{\bar{\Omega}} \langle \nu_x^\infty, f^\infty(x, \bullet)\rangle \mathrm{d}\lambda(x).$$

Finally, we cite a variant of the fundamental theorem of Young measures:

**Proposition A.6.** *Let $\Omega \subset \mathbb{R}^n$ be bounded and open, and $F\colon \Omega \times \mathbb{V} \to \mathbb{R}$ be a measurable integrand such that $f(x, \bullet)$ is continuous for almost every $x \in \Omega$ (a Carathéodory integrand). Let $(v_j) \subset L^1(\Omega, \mathbb{V})$ generate $\boldsymbol{\nu} \in \mathrm{Y}(\Omega, \mathbb{V})$ be such that $(f(\bullet, v_j))_j$ is uniformly integrable. Then*

$$\lim_{j\to\infty} \int_\Omega f(x, v_j(x)) \mathrm{d}x = \int_\Omega \langle \nu_x, f(x, \bullet)\rangle \mathrm{d}x.$$

## References


1. M. Amar, V. De Cicco, and N. Fusco, *Lower semicontinuity and relaxation results in BV for integral functionals with BV integrants*, ESAIM: Control, Optimization and Calculus of Variations **14** (2008), no. 3, 456–477, https://doi.org/10.1051/cocv:2007061.
2. L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*, Oxford University Press, USA, 2000.
3. P. Athavale, R.L. Jerrard, M. Novaga, and G. Orlandi, *Weighted TV minimization and application to vortex density models*, Journal of Convex Analysis **24** (2017), no. 4, http://cvgmt.sns.it/media/doc/paper/2790/AJNO.pdf.
4. M. Bergounioux and L. Piffet, *A second-order model for image denoising*, Set-Valued and Variational Analysis **18** (2010), no. 3-4, 277–306, http://dx.doi.org/10.1007/s11228-010-0156-6.
5. K. Bredies, *Recovering piecewise smooth multichannel images by minimization of convex functionals with total generalized variation penalty*, Efficient Algorithms for Global Optimization Methods in Computer Vision, Lecture notes in Computer Science, Springer Berlin Heidelberg, 2014, http://dx.doi.org/10.1007/978-3-642-54774-4_3, pp. 44–77.
6. K. Bredies and M. Holler, *Regularization of linear inverse problems with total generalized variation*, Journal of Inverse and Ill-posed Problems **22** (2014), no. 6, 871–913, https://doi.org/10.1515/jip-2013-0068.
7. ———, *Higher-order total variation approaches and generalisations*, Inverse Problems **36** (2020), no. 12, 123001, https://doi.org/10.1088/1361-6420/ab8f80.
8. K. Bredies, K. Kunisch, and T. Pock, *Total generalized variation*, SIAM Journal on Imaging Sciences **3** (2010), no. 3, 492–526, http://dx.doi.org/10.1137/090769521.
9. K. Bredies and T. Valkonen, *Inverse problems with second-order total generalized variation constraints*, Proceedings of SampTA 2011 - 9th International Conference on Sampling Theory and Applications, Singapore, 2011.
10. D. Breit, L. Diening, and F. Gmeineder, *On the trace operator for functions of bounded $\mathbb{A}$-variation*, Analysis & PDE **13** (2020), no. 2, 559 – 594, https://doi.org/10.2140/apde.2020.13.559.
11. E.M. Brinkmann, M. Burger, and J.S. Grah, *Unified models for second-order TV-type regularisation in imaging: A new perspective based on vector operators.*, Journal of Mathematical Imaging and Vision **61** (2019), 571–601, https://doi.org/10.1007/s10851-018-0861-6.
12. M. Burger, K. Papafitsoros, E. Papoutsellis, and C.B. Schönlieb, *Infimal convolution regularisation functionals of* BV *and* $\mathrm{L}^p$ *spaces. Part I: The finite p case*, Journal of Mathematical Imaging and Vision **55** (2016), no. 3, 343–369, http://dx.doi.org/10.1007/s10851-015-0624-6.
13. M. Burger, K. Papafitsoros, E. Papoutsellis, and C.B. Schönlieb, *Infimal convolution regularisation functionals of* BV *and* $\mathrm{L}^p$ *spaces. the case $p = \infty$*, System Modeling and Optimization (Lorena Bociu, Jean-Antoine Désidéri, and Abderrahmane Habbal, eds.), Springer International Publishing, 2016, https://doi.org/10.1007/978-3-319-55795-3_15, pp. 169–179.
14. L. Calatroni, C. Chung, J.C. De Los Reyes, C.B. Schönlieb, and T. Valkonen, *Bilevel approaches for learning of variational imaging models*, RADON book Series on Computational and Applied Mathematics, vol. 18, Berlin, Boston: De Gruyter, 2017, https://www.degruyter.com/view/product/458544.

15. J. Canny, *A computational approach to edge detection*, IEEE Transactions on Pattern Analysis and Machine Intelligence **PAMI-8** (1986), no. 6, 679–698, http://dx.doi.org/10.1109/TPAMI.1986.4767851.
16. A. Chambolle and P.L. Lions, *Image recovery via total variation minimization and related problems*, Numerische Mathematik **76** (1997), 167–188, http://dx.doi.org/10.1007/s002110050258.
17. A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, Journal of Mathematical Imaging and Vision **40** (2011), no. 1, 120–145, http://dx.doi.org/10.1007/s10851-010-0251-1.
18. C.V. Chung, J.C. De los Reyes, and C.B. Schönlieb, *Learning optimal spatially-dependent regularization parameters in total variation image denoising*, Inverse Problems **33** (2017), no. 7, 074005, http://stacks.iop.org/0266-5611/33/i=7/a=074005.
19. E. Davoli, I. Fonseca, and P. Liu, *Adaptive image processing: first order PDE constraint regularizers and a bilevel training scheme*, 2019, arXiv preprint 1902.01122, https://arxiv.org/pdf/1902.01122.pdf.
20. J.C. De los Reyes and C.B. Schönlieb, *Image denoising: learning the noise model via nonsmooth PDE-constrained optimization*, Inverse Problems and Imaging **7** (2013), no. 4, 1183–1214, http://dx.doi.org/10.3934/ipi.2013.7.1183.
21. J.C. De Los Reyes, C.B. Schönlieb, and T. Valkonen, *The structure of optimal parameters for image restoration problems*, Journal of Mathematical Analysis and Applications **434** (2016), 464–500, https://doi.org/10.1016/j.jmaa.2015.09.023.
22. J.C. De Los Reyes, C.B. Schönlieb, and T. Valkonen, *Bilevel parameter learning for higher-order Total Variation regularisation models*, Journal of Mathematical Imaging and Vision **57** (2017), no. 1, 1–25, https://doi.org/10.1007/s10851-016-0662-8.
23. J.C. De los Reyes and D. Villacís, *Optimality conditions for bilevel imaging learning problems with total variation regularization*, arXiv preprint arXiv:2107.08100 (2021), https://arxiv.org/abs/2107.08100.
24. F. Demengel and R. Temam, *Convex functions of a measure and applications*, Indiana University Mathematics Journal **33** (1984), 673–709.
25. M.J. Ehrhardt and M.M. Betcke, *Multicontrast MRI reconstruction with structure-guided total variation*, SIAM Journal on Imaging Sciences **9** (2016), no. 3, 1084–1106, https://doi.org/10.1137/15M1047325.
26. I. Fonseca and G. Leoni, *On lower semicontinuity and relaxation*, Proceedings of the Royal Society of Edinburgh Section A Mathematics **131** (2001), no. 3, 519–565, https://doi.org/10.1017/S0308210500000998.
27. F. Gmeineder and B. Raiță, *Embeddings for a-weakly differentiable functions on domains*, Journal of Functional Analysis **277** (2019), no. 12, 108278, https://doi.org/10.1016/j.jfa.2019.108278.
28. M. Hintermüller, M. Holler, and K. Papafitsoros, *A function space framework for structural total variation regularization with applications in inverse problems*, Inverse Problems **34** (2018), no. 6, https://doi.org/10.1088/1361-6420/aab586.
29. M. Hintermüller and K. Papafitsoros, *Generating structured nonsmooth priors and associated primal-dual methods*, Processing, Analyzing and Learning of Images, Shapes, and Forms: Part 2 (Ron Kimmel and Xue-Cheng Tai, eds.), Handbook of Numerical Analysis, vol. 20, 2019, https://doi.org/10.1016/bs.hna.2019.08.001, pp. 437–502.
30. M. Hintermüller, K. Papafitsoros, and C.N. Rautenberg, *Analytical aspects of spatially adapted total variation regularisation*, Journal of Mathematical Analysis and Applications **454** (2017), no. 2, 891–935, https://doi.org/10.1016/j.jmaa.2017.05.025.
31. M. Hintermüller, K. Papafitsoros, C.N. Rautenberg, and H. Sun, *Dualization and automatic distributed parameter selection of total generalized variation via bilevel optimization*, Numerical Functional Analysis and Optimization (2022), 1–46, https://doi.org/10.1080/01630563.2022.2069812.
32. M. Hintermüller and C.N. Rautenberg, *Optimal selection of the regularization function in a weighted total variation model. Part I: Modelling and theory*, Journal of Mathematical Imaging and Vision **59** (2017), no. 3, 498–514, https://doi.org/10.1007/s10851-017-0744-2.
33. M. Hintermüller, C.N. Rautenberg, T. Wu, and A. Langer, *Optimal selection of the regularization function in a weighted total variation model. Part II: Algorithm, its analysis and numerical tests*, Journal of Mathematical Imaging and Vision **59** (2017), no. 3, 515–533, https://doi.org/10.1007/s10851-017-0736-2.
34. M. Hintermüller and G. Stadler, *An infeasible primal-dual algorithm for total bounded variation–based inf-convolution-type image restoration*, SIAM Journal on Scientific Computing **28** (2006), no. 1, 1–23, http://dx.doi.org/10.1137/040613263.
35. M. Holler and K. Kunisch, *On infimal convolution of TV-type functionals and applications to video and image reconstruction*, SIAM Journal on Imaging Sciences **7** (2014), no. 4, 2258–2300, https://doi.org/10.1137/130948793.

36. K. Jalalzai, *Discontinuities of the minimizers of the weighted or anisotropic total variation for image reconstruction*, arXiv preprint 1402.0026 (2014), http://arxiv.org/abs/1402.0026.
37. B. Kirchheim and J. Kristensen, *On rank one convex functions that are homogeneous of degree one*, Archive for rational mechanics and analysis **221** (2016), no. 1, 527–558, https://doi.org/10.1007/s00205-016-0967-1.
38. R.D. Kongskov, Y. Dong, and K. Knudsen, *Directional total generalized variation regularization*, BIT Numerical Mathematics **59** (2019), 903–928, https://doi.org/10.1007/s10543-019-00755-6.
39. J. Kristensen and B. Raiță, *An introduction to generalized young measures*, Lecture notes 45/2020, Max Planck Institute for Mathematics in the Sciences (2020).
40. J. Kristensen and F. Rindler, *Characterization of generalized gradient young measures generated by sequences in $W^{1,1}$ and $BV$*, Archive for rational mechanics and analysis **197** (2010), 539–598, https://doi.org/10.1007/s00205-009-0287-9.
41. K. Kunisch and T. Pock, *A bilevel optimization approach for parameter learning in variational models*, SIAM Journal on Imaging Sciences **6** (2013), no. 2, 938–983, http://dx.doi.org/10.1137/120882706.
42. D. Ornstein, *A non-inequality for differential operators in the* $L_1$ *norm*, Archive for Rational Mechanics and Analysis **11** (1962), no. 1, 40–49, https://doi.org/10.1007/BF00253928.
43. K. Papafitsoros, *Novel higher order regularisation methods for image reconstruction*, Ph.D. thesis, University of Cambridge, 2014, https://www.repository.cam.ac.uk/handle/1810/246692.
44. K. Papafitsoros and C.B. Schönlieb, *A combined first and second order variational approach for image reconstruction*, Journal of Mathematical Imaging and Vision **48** (2014), no. 2, 308–338, http://dx.doi.org/10.1007/s10851-013-0445-4.
45. S. Parisotto, S. Masnou, and C.B. Schönlieb, *Higher-order total directional variation: Analysis*, SIAM Journal on Imaging Sciences **13** (2020), no. 1, 474–496.
46. L.I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D: Nonlinear Phenomena **60** (1992), no. 1-4, 259–268, http://dx.doi.org/10.1016/0167-2789(92)90242-F.
47. K.T. Smith, *Formulas to represent functions by their derivatives*, Mathematische Annalen **188** (1970), no. 1, 53–77, http://eudml.org/doc/162037.
48. T. Valkonen, K. Bredies, and F. Knoll, *Total generalized variation in diffusion tensor imaging*, SIAM Journal on Imaging Sciences **6** (2013), no. 1, 487–525, http://dx.doi.org/10.1137/120867172.
49. L. Vese, *A study in the BV space of a denoising-deblurring variational problem*, Applied Mathematics and Optimization **44** (2001), no. 2, 131–161, https://doi.org/10.1007/s00245-001-0017-7.

V. Pagliari.
Institute of Analysis and Scientific Computing, TU Wien,
Wiedner Hauptstraße 8-10, 1040 Vienna, Austria.
E-mail: valerio.pagliari@tuwien.ac.at.

K. Papafitsoros.
Weierstrass Institute for Applied Analysis and Stochastics,
Mohrenstraße 39, 10117 Berlin, Germany.
E-mail: papafitsoros@wias-berlin.de.

B. Raiță.
Centro di Ricerca Matematica Ennio De Giorgi, Scuola Normale Superiore,
P.za dei Cavalieri, 3, 56126 Pisa PI, Italy;
Department of Mathematics, Alexandru-Ioan Cuza University of Iași,
Bd. Carol I, no. 11, 700506 Iași, Romania
E-mail: bogdanraita@gmail.com

A. Vikelis.
Faculty of Mathematics, University of Vienna
Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria.
&
University of Sussex,
Falmer, Brighton BN1 9RH, United Kingdom.
E-mail: andreas.panagiotis.vikelis@univie.ac.at