

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

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

ABSTRACT. In the context of image processing, given a k -th order, homogeneous and linear differential operator with constant coefficients, we study a class of variational problems whose regularizing terms depend on the operator. Precisely, the regularizers are integrals of spatially inhomogeneous integrands with convex dependence on the differential operator applied to the image function. 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 space-dependent regularization parameters, thus allowing for good flexibility and preservation of details in the reconstructed image. We establish existence of optima for the scheme and we finally substantiate its feasibility by numerical examples in image denoising. The cases that we treat are Huber versions of the first and second order total variation with both the Huber and the regularization parameter being spatially dependent. Notably the spatially dependent version of second order total variation produces high quality reconstructions when compared to regularizations of similar type, and the introduction of the spatially dependent Huber parameter leads to a further enhancement of the image details.

1. INTRODUCTION

In this contribution we study a 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 α , we look for solutions to the problem

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

where F is an assigned cost functional and u_α is an image reconstructed by minimizing

$$(1.2) \quad I[u; \alpha] := \Phi_g(u) + \mathcal{R}(u; \alpha).$$

Here, Φ_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 α belonging to the admissible set Adm . The datum g is typically a corrupted version of some ground truth image u_{gt} . Often, one has

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

with η 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 Φ_g , \mathcal{R} and α is to obtain an output u which represents as well as possible the initial ground truth image u_{gt} . We concisely point out here that the main novelty of the present paper consists in establishing existence of solutions to the scheme for inhomogeneous regularizers

of the type

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

and present numerical results that fit this framework. Here, $\Omega \subset \mathbb{R}^n$ is the image domain, $\alpha \in \text{Adm} \subset L^\infty(\Omega, [0, +\infty))$ belongs to a class of admissible weights, f is a Carathéodory integrand that is convex in the second entry, and \mathcal{B} is a linear, k -th order, homogeneous differential operator with constant coefficients. Before we report further on our contribution, we proceed with a brief review of regularization functionals in image reconstruction.

Among classical regularization functionals we find the total variation (TV) [34, 13], as well as higher order or anisotropic extensions of it. Particularly relevant for this work are the second order total variation (TV²) [33, 4] and the total generalized variation (TGV) [7]. For a function $u \in L^1(\Omega)$, these functionals are defined by duality as follows:

$$(1.4) \quad \text{TV}(u) = \sup \left\{ \int_{\Omega} u \operatorname{div} \phi \, dx : \phi \in C_c^\infty(\Omega, \mathbb{R}^n), \|\phi\|_\infty \leq 1 \right\},$$

$$(1.5) \quad \text{TV}^2(u) = \sup \left\{ \int_{\Omega} u \operatorname{div}^2 \phi \, dx : \phi \in C_c^\infty(\Omega, \mathbb{R}^{n \times n}), \|\phi\|_\infty \leq 1 \right\},$$

$$(1.6) \quad \text{TGV}(u) = \sup \left\{ \int_{\Omega} u \operatorname{div}^2 \phi \, dx : \phi \in C_c^\infty(\Omega, \mathcal{S}^{n \times n}), \|\phi\|_\infty \leq \alpha_0, \|\operatorname{div}^2 \phi\|_\infty \leq \alpha_1 \right\}.$$

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 α acting in a multiplicative way, i.e., αTV and αTV^2 . If the supremum in (1.4) is finite, then we say that $u \in \text{BV}(\Omega)$, the space of functions of bounded variation [2], and $\text{TV}(u) = |Du|(\Omega)$, where $|Du|$ is the total variation measure associated with the distributional derivative $Du \in \mathcal{M}(\Omega, \mathbb{R}^n)$. Similarly, if the right-hand side in (1.5) is finite, then $u \in \text{BV}^2(\Omega)$, the space of functions of bounded second variation [33, 4], and $\text{TV}^2(u) = |D^2u|(\Omega)$, with $D^2u \in \mathcal{M}(\Omega, \mathcal{S}^{n \times n})$. Finally, it turns out that if the supremum in (1.6) is finite, then $u \in \text{BV}(\Omega)$ as well and

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

see [8, 6]. In the previous formula, $\text{BD}(\Omega)$ is the space of functions of bounded deformation and $\mathcal{E}w$ 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 [32].

Grounding on the concept of convex functions of Radon measures [21], variants of the above regularizers involving convex integrands have also been considered in the literature [33, 37, 28]. A widely used example is the one of Huber total variation TV_γ , which is defined for $u \in \text{BV}(\Omega)$ as

$$(1.7) \quad \text{TV}_\gamma(u) = \int_{\Omega} f_\gamma(dDu) = \int_{\Omega} f_\gamma(\nabla u) dx + \int_{\Omega} d|D^s u|,$$

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

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

Note that $\text{TV}_\gamma(u)$ can be equivalently defined via duality as

$$(1.9) \quad \text{TV}_\gamma(u) = \sup \left\{ \int_\Omega u \operatorname{div} \phi \, dx - \frac{1}{2} \gamma \int_\Omega |\phi|^2 dx : \phi \in C_c^\infty(\Omega, \mathbb{R}^n), \|\phi\|_\infty \leq 1 \right\},$$

see [21]. This modification of TV, which corresponds to a smoothing of the ℓ_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 $(2\gamma)^{-1} \int_\Omega |\nabla u|^2 dx$, which promotes smooth reconstructions [28, 10].

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_{gt}, g) consisting of the ground truth and its corrupted counterpart [11, 19, 18, 17, 31]. In these works the energy in the upper level problem (1.1) is usually given by

$$(1.10) \quad F_{\text{PSNR}}(u) := \|u - u_{\text{gt}}\|_{L^2(\Omega)}^2,$$

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 $\alpha: \Omega \rightarrow [0, +\infty)$ [15, 20]. A recent series of papers [26, 27, 24, 25] 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 η with variance σ^2 , this is achieved by the introduction of the statistics-based upper level objective

$$(1.11) \quad F_{\text{stat}}(u) := \frac{1}{2} \int_\Omega \max(Ru - \bar{\sigma}^2, 0)^2 dx + \frac{1}{2} \int_\Omega \min(Ru - \underline{\sigma}^2, 0)^2 dx,$$

where $\underline{\sigma}^2 := \sigma^2 - \epsilon$, $\bar{\sigma}^2 := \sigma^2 + \epsilon$, and

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

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

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 \mathcal{B} between two finite dimensional Euclidean spaces \mathbb{U} and \mathbb{V} , we introduce the general regularizer

$$(1.12) \quad \mathcal{R}(u; \alpha) = \sum_{i=1}^{k-1} \int_\Omega \alpha_i(x) f_i(x, dD^i u) + \int_\Omega \alpha_k(x) f_k(x, d\mathcal{B}u),$$

$\alpha_i: \Omega \rightarrow [0, +\infty)$ 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, or in other words, they explicitly depend on the spatial variable x in a measurable way. More details about the setting are to be found in Section 2. As a first contribution, we prove lower semicontinuity of the functional in (1.12) 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 and an associated reconstructed image.

Not much work has been done for functionals depending on general differential operators. One example comes from the recent preprint [16], where a bilevel scheme for first order differential operators \mathcal{B} is developed. Interestingly, the authors identify classes of operators \mathcal{B} such that the scheme outputs an optimal reconstructed image and an optimal \mathcal{B} for the upper level problem. However, in their analysis one always obtains BV minimizers. In contrast, in our method the operator \mathcal{B} is fixed, but it is allowed to be *arbitrary* (see Theorems 3.3 and 4.1).

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. Our hypotheses on the convex integrands are optimal, due to the use of Young measures for oscillation and concentration, see Appendix A. From an analytical point of view, our regularity assumptions on the weights are minimal, as can be seen from Section 2.3. In the future, we aim to develop our theory to include optimization problems over linear PDE operators \mathcal{B} that satisfy as few assumptions as possible. We expect that our lower semicontinuity result, Theorem 3.3, and the existence result for the bilevel schemes, Theorem 4.1 will serve as preliminary work in this direction.

We conclude the paper with a series of numerical examples that deal with versions of the Huber TV and TV^2 in which both the Huber and the regularization parameter are spatially dependent. We devise a strategy to prefix the former in a sensible way, while the latter is computed automatically by the bilevel scheme. Even though the main purpose of these numerical examples is to support the applicability and versatility of the framework, we are able to draw two interesting conclusions. The first one is that the bilevel TV^2 with spatially varying weight, in combination with the statistics-based upper level objective F_{stat} is 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 \mathcal{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 prove lower semicontinuity and existence theorems concerning the lower level of the bilevel scheme, while Section 4 is devoted to the main existence result for optimal weights and reconstructed images. Eventually, numerical experiments for the Huber TV and 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 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: $BV_p^{\mathcal{B}}$ spaces. We work in the n -dimensional Euclidean space \mathbb{R}^n , $n \geq 2$, that we endow with the Lebesgue measure \mathcal{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 Ω 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 [36]. 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 $\text{Lin}(\mathbb{U}, \mathbb{V})$ be the space of linear maps from \mathbb{U} to \mathbb{V} . Hereafter, for $k \in \mathbb{N} \setminus \{0\}$, \mathcal{B} denotes a k -th order, homogeneous and linear differential operator with constant coefficients. Explicitly, given $B_i \in \text{Lin}(\mathbb{U}, \mathbb{V})$ for any n -dimensional multi-index $i = (i_1, \dots, i_n) \in \mathbb{N}^n$, we define for a smooth function $u: \mathbb{R}^n \rightarrow \mathbb{U}$

$$\mathcal{B}u := \sum_{|i|=k} B_i \partial^i u.$$

(recall that $|i| := \sum_j i_j$). When u is less regular, we interpret $\mathcal{B}u$ in the distributional sense. In particular, we are interested in the case in which $\mathcal{B}u$ 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 σ -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

$$C_0(O, \mathbb{U}) := \{u: O \rightarrow \mathbb{U} : \{|u| > \delta\} \text{ is relatively compact for all } \delta > 0\},$$

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 $|\cdot|$. 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 $\mathcal{B}\mu \in \mathcal{M}(\Omega, \mathbb{V})$ if and only if there exists $\nu \in \mathcal{M}(\Omega, \mathbb{V})$ such that

$$\langle \nu, \phi \rangle = \int_{\Omega} \mathcal{B}^* \phi \, d\mu \quad \text{for all } \phi \in C_c^\infty(\Omega, \mathbb{V}),$$

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

$$\mathcal{B}^* \phi := - \sum_{|i|=k} B_i^* \partial^i \phi \quad \text{for all } \phi \in C_c^\infty(\mathbb{R}^n, \mathbb{V}),$$

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 Ω , \mathbb{U} and \mathbb{V} as above, and for $p \in (1, +\infty)$, we set

$$BV_p^{\mathcal{B}}(\Omega) := \{u \in L^p(\Omega, \mathbb{U}) : \mathcal{B}u \in \mathcal{M}(\Omega, \mathbb{V})\},$$

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

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

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 $BV^{\mathcal{B}}$ spaces (and of the equation $\mathcal{B}u = v$ in general) can be expressed in terms of algebraic properties of the characteristic polynomial. We recall that the characteristic

polynomial, or symbol, of \mathcal{B} is

$$\mathcal{B}(\xi) := \sum_{|i|=k} \xi^i B_i \in \text{Lin}(\mathbb{U}, \mathbb{V}), \quad \xi \in \mathbb{C}^n,$$

where $\xi^i := \xi_1^{i_1} \cdots \xi_n^{i_n}$. In our study, the following property will be particularly relevant:

Definition 2.1 ([35, 9, 23]). *An operator \mathcal{B} is said to be \mathbb{C} -elliptic if*

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

It was shown in [35] that \mathbb{C} -ellipticity is equivalent with full Sobolev regularity for the equation $\mathcal{B}u = v$ 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 ([23]). *Let $\Omega \subset \mathbb{R}^n$ be a Lipschitz domain. An operator \mathcal{B} is \mathbb{C} -elliptic if and only if*

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

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

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 $g \in L^p(\Omega, \mathbb{U})$,

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

$$(L2) \quad \text{such that } u_\alpha \in \text{argmin} \left\{ I[u; \alpha] : u \in \text{BV}_p^{\mathcal{B}}(\Omega) \right\},$$

where

$$(2.2) \quad I[u; \alpha] := \Phi_g(u) + \int_{\Omega} \alpha(x) f(x, d\mathcal{B}u).$$

All the due definitions and assumptions are collected below.

– *Cost functional:* As for the upper level problem (L1), $F: L^p(\Omega, \mathbb{U}) \rightarrow \mathbb{R}$ is a proper, convex and weakly lower semicontinuous functional. Typical choices for this functional are the PSNR maximizing F_{PSNR} in (1.10), which makes use of the ground truth u_{gt} , and the statistics-based, ground truth-free F_{stat} in (1.11), in the spirit of supervised and unsupervised learning respectively.

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

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

In particular,

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

is a simple instance of fidelity term.

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

$$(2.3) \quad \text{Adm} := \{ \alpha \in C(\bar{\Omega}, [\underline{\alpha}, \bar{\alpha}]) : |\alpha(x) - \alpha(y)| \leq \omega(|x - y|) \text{ for every } x, y \in \bar{\Omega} \}$$

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 $f: \Omega \times \mathbb{V} \rightarrow \mathbb{R}$ is a *Carathéodory* integrand such that $z \mapsto f(x, z)$ is convex for \mathcal{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

$$(2.4) \quad c(|\cdot| - 1) \leq f(x, \cdot) \leq C(1 + |\cdot|) \quad \text{for } \mathcal{L}^n\text{-a.e. } x \in \Omega,$$

for some $c, C \geq 0$.

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 $f(x, d\mathcal{B}u)$. For an integrand $f: \Omega \times \mathbb{V} \rightarrow \mathbb{R}$ satisfying (2.4), we define the *recession function*

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

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

$$(2.6) \quad \int_{\Omega} f(x, d\mu) := \int_{\Omega} f\left(x, \frac{d\mu}{d\mathcal{L}^n}(x)\right) dx + \int_{\Omega} f^\infty\left(x, \frac{d\mu^s}{d|\mu|}(x)\right) d|\mu|(x),$$

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

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 [26] (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 $(u_{\text{gt}}, g) \in L^2(\Omega, \mathbb{R}) \times L^p(\Omega, \mathbb{R})$ is assigned, where u_{gt} and g encode respectively the ground truth and the corrupted datum. We also fix two positive parameters $\underline{\alpha}$ and $\bar{\alpha}$, and we provisionally allow the regularizing weights to vary in $\text{LSC}(\Omega, [\underline{\alpha}, \bar{\alpha}])$, the space of lower semicontinuous functions on Ω with range in $[\underline{\alpha}, \bar{\alpha}]$.

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

$$(2.7) \quad J[u; \alpha] := \int_{\Omega} |u - g|^p dx + \int_{\Omega} \alpha(x) d|Du|(x),$$

and the ensuing corresponding training scheme:

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

$$(2.9) \quad \text{such that } u_\alpha \in \text{argmin} \{J[u; \alpha] : u \in \text{BV}(\Omega)\}.$$

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 α . Indeed, if in (2.7) $\alpha: \Omega \rightarrow [\underline{\alpha}, \bar{\alpha}]$ is a given function and u is allowed to vary in $\text{BV}(\Omega)$ (as it is the case of (2.9)), there might be choices of u for which the coupling

$$\int_{\Omega} \alpha(x) d|Du|(x)$$

is not well-defined. Prescribing lower semicontinuity for the admissible weights α allows to circumvent the issue, because lower semicontinuous functions are Borel measurable and $Du \in \mathcal{M}(\Omega, \mathbb{R}^n)$ is a Borel measure. Besides, for any $\alpha \in \text{LSC}(\Omega, [\underline{\alpha}, \bar{\alpha}])$ the existence of a

solution u_α 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 $\text{BV}(\Omega)$. Then, there exist $u \in \text{BV}(\Omega)$ and a subsequence $(j_k)_k$ such that (u_{j_k}) weakly-* converges to u , that is, $u_{j_k} \rightarrow u$ in $L^1(\Omega)$ and*

$$\lim_{k \rightarrow +\infty} \int_{\Omega} \phi \, d\text{D}u_{j_k} = \int_{\Omega} \phi \, d\text{D}u \quad \text{for all } \phi \in C_0(\Omega).$$

Secondly, we notice that $J[\cdot; \alpha]$ is lower semicontinuous with respect to the L^1 -convergence, because, when $\alpha \in \text{LSC}(\Omega, [\underline{\alpha}, \bar{\alpha}])$, general lower semicontinuity results in BV may be invoked (see e.g. [22]; 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 $(\alpha_j)_j \subset \text{LSC}(\Omega, [\underline{\alpha}, \bar{\alpha}])$ be a minimizing sequence for (2.8). Then, by definition, the integrals

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

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

$$F_{\text{PSNR}}(u) \leq \inf \{F_{\text{PSNR}}(u_j) : \alpha \in \text{LSC}(\Omega, [\underline{\alpha}, \bar{\alpha}])\}.$$

If we manage to show that $u = u_{\alpha^*}$ for some admissible α^* , then the latter is a solution to (2.8). The natural choice for α^* 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, $J[u; \cdot]$ is continuous with respect to a suitable weak-* convergence. Indeed, if $(\alpha_j)_j \subset \text{LSC}(\Omega, [\underline{\alpha}, \bar{\alpha}])$ is bounded and $u \in \text{BV}(\Omega)$, then there exist a subsequence, which we do not relabel, and $\alpha^* \in L^\infty(\Omega, [\underline{\alpha}, \bar{\alpha}]; |Du|)$ such that

$$\lim_{j \rightarrow +\infty} \int_{\Omega} \alpha_j(x) \phi(x) \, d|Du|(x) = \int_{\Omega} \alpha^*(x) \phi(x) \, d|Du|(x) \quad \text{for all } \phi \in L^1(\Omega; |Du|).$$

In particular,

$$(2.10) \quad \lim_{j \rightarrow +\infty} J[u; \alpha_j] = J[u; \alpha^*].$$

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

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

where we assume *a priori* that $\text{Adm} \subset C(\bar{\Omega}, [\underline{\alpha}, \bar{\alpha}])$ is compact with respect to the uniform convergence. Under the compactness assumptions on the class of admissible weights, if $(\alpha_j)_j \subset \text{Adm}$ is a minimizing sequence for (2.3), and if $(u_j)_j$ and u are constructed as above, we are actually able to prove that $u = u_{\alpha^*}$, where $\alpha^* \in \text{Adm}$ is the uniform limit of (α_j) . In other words, the couple (α^*, u) is a solution to the scheme consisting of (2.3)–(2.9).

To prove the claim, we need to show that

$$(2.11) \quad J[u; \alpha^*] \leq J[v; \alpha^*] \quad \text{for any } v \in \text{BV}(\Omega).$$

We start from observing that the definition of u_j grants

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

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

$$(2.12) \quad \liminf_{j \rightarrow +\infty} J[u_j; \alpha_j] \leq \liminf_{j \rightarrow +\infty} J[v; \alpha_j] = J[v; \alpha^*],$$

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

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

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

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

From the uniform convergence of (α_j) and the weak-* convergence of (u_j) we obtain

$$(2.13) \quad J[u; \alpha^*] \leq \liminf_{j \rightarrow +\infty} J[u_j; \alpha^{(j)}].$$

On the whole, owing to (2.12), we deduce (2.11).

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 $u \in \text{BV}(\Omega)$ be the weak-* limit of (u_j) and let $\alpha^* \in L^\infty(\Omega, [\underline{\alpha}, \bar{\alpha}]; |Du|)$ be the weak-* limit of (α_j) . Proving the optimality of u , i.e. $u = u_{\alpha^*}$, means

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

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. (4.5), (4.6)) is the following semicontinuity inequality:

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

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

3. EXISTENCE THEOREMS FOR THE LOWER LEVEL PROBLEMS

We begin with a general lower semicontinuity result for convex integrands with rough x -dependence:

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

$$\mu_j \xrightarrow{*} \mu \text{ in } \mathcal{M}(\Omega, \mathbb{V}) \implies \liminf_{j \rightarrow \infty} \int_{\Omega} f(x, d\mu_j(x)) \geq \int_{\Omega} f(x, d\mu(x)).$$

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

Proof. We consider a Young measure ν generated by $(\mu_j)_j$. By Proposition A.5 and Jensen’s inequality, we have that

$$\begin{aligned} \liminf_{j \rightarrow \infty} \int_{\Omega} f(x, d\mu_j(x)) &\geq \int_{\Omega} \langle \nu_x, f(x, \cdot) \rangle dx + \int_{\bar{\Omega}} \langle \nu_x^\infty, f^\infty(x, \cdot) \rangle d\lambda \\ &\geq \int_{\Omega} f(x, \bar{\nu}_x) dx + \int_{\bar{\Omega}} f^\infty(x, \bar{\nu}_x^\infty) d\lambda \end{aligned}$$

$$= \int_{\Omega} [f(x, \bar{\nu}_x) + \lambda^a(x) f^{\infty}(x, \bar{\nu}_x^{\infty})] dx + \int_{\bar{\Omega}} f^{\infty}(x, \bar{\nu}_x^{\infty}) d\lambda^s$$

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

$$\begin{aligned} \liminf_{j \rightarrow \infty} \int_{\Omega} f(x, d\mu_j(x)) &\geq \int_{\Omega} f(x, \bar{\nu}_x + \lambda^a(x) \bar{\nu}_x^{\infty}) dx + \int_{\bar{\Omega}} f^{\infty}(x, \bar{\nu}_x^{\infty}) d\lambda^s \\ &\geq \int_{\Omega} f(x, \mu^a(x)) dx + \int_{\Omega} f^{\infty}(x, d\mu^s) \\ &= \int_{\Omega} f(x, d\mu). \end{aligned}$$

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

Remark 3.2. As it can be seen from the proof of Proposition 3.1, we can allow for signed integrands, as long as we consider $\mu_j \xrightarrow{*} \mu$ in $\mathcal{M}(\bar{\Omega}, \mathbb{V})$. In this case, the recession term in (2.6) should be an integral over $\bar{\Omega}$.

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

Theorem 3.3. *Let us fix $p \in (1, +\infty)$, $g \in L^p(\Omega, \mathbb{U})$, $\underline{\alpha} > 0$ and $\alpha \in C(\bar{\Omega}, [\underline{\alpha}, +\infty))$. Let $f: \Omega \times \mathbb{V} \rightarrow [0, +\infty)$ be an integrand satisfying the assumptions outlined in Subsection 2.2, and suppose further that the recession function f^{∞} 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 $u \in BV_p^{\mathcal{B}}(\Omega)$. 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 $(u_j)_j \subset BV_p^{\mathcal{B}}(\Omega)$ such that the limit of $I[u_j; \alpha]$ as $j \rightarrow +\infty$ is finite. Since Φ_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 $BV_p^{\mathcal{B}}(\Omega)$. Thus, on a subsequence that we do not relabel, we have $u_j \rightharpoonup u$ in $L^p(\Omega, \mathbb{U})$ and $\mathcal{B}u_j \xrightarrow{*} \mathcal{B}u$ in $\mathcal{M}(\Omega, \mathbb{V})$. By the weak lower semicontinuity of the fidelity term we have that

$$(3.1) \quad \Phi_g(u) \leq \liminf_{j \rightarrow +\infty} \Phi_g(u_j),$$

whereas by Proposition 3.1 we obtain

$$(3.2) \quad \int_{\Omega} f(x, d\mathcal{B}u) \leq \liminf_{j \rightarrow +\infty} \int_{\Omega} f(x, d\mathcal{B}u_j).$$

On the whole, we deduce

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

and we conclude that $u \in BV_p^{\mathcal{B}}(\Omega)$ is a minimizer of I .

Uniqueness follows easily when Φ_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

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

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

Remark 3.4. Notably, in the previous theorem uniqueness holds for $\Phi_g(u) = \|u - g\|_{L^p(\Omega)}^p$. Indeed, for instance by the uniform convexity of the L^p spaces, we have for $u^0 := (u^1 + u^2)/2$

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

Remark 3.5. There is no immediate counterpart of Theorem 3.3 when $p = 1$, because in general bounded sequences in $BV^{\mathcal{B}}$ are not weakly-* precompact. One possibility would be to embed $BV^{\mathcal{B}}$ in the larger space of measures $\{\mu \in \mathcal{M}(\Omega, \mathbb{U}) : \mathcal{B}\mu \in \mathcal{M}(\Omega, \mathbb{V})\}$. A second option is to assume \mathcal{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 (1.12), and we can obtain much more precise information on the minimizers. We make the unconventional convention that $\frac{n}{n-k} = +\infty$ if $k \geq n$, and we denote by $\text{sym}^i(\mathbb{R}^n, \mathbb{U})$ the space of symmetric \mathbb{U} -valued i -linear maps on \mathbb{R}^n .

Theorem 3.6. *Let us fix $p \in [1, \frac{n}{n-k})$, $g \in L^p(\Omega, \mathbb{U})$, $\underline{\alpha} > 0$, $\alpha_i \in C(\bar{\Omega}, [0, +\infty))$ for $i = 1, \dots, k-1$ and $\alpha_k \in C(\bar{\Omega}, [\underline{\alpha}, +\infty))$. Let $f_i: \Omega \times \text{sym}^i(\mathbb{R}^n, \mathbb{U}) \rightarrow \mathbb{R}$, $i = 1, \dots, k-1$, and $f_k: \Omega \times \mathbb{V} \rightarrow [0, +\infty)$ be Carathéodory integrands such that for all i $f_i(x, \cdot)$ is convex and $f_i(x, \cdot) \leq C(1 + |\cdot|)$ for almost every $x \in \Omega$. Suppose in addition that f_k satisfies the coercivity bound $c(|\cdot| - 1) \leq f_k(x, z)$ and that*

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

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

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

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

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

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

Proof. If $k = 1$ the statement collapses to Theorem 3.3. The \mathbb{C} -ellipticity of \mathcal{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 $k \geq 2$.

If $(u_j)_j \subset BV^{\mathcal{B}}(\Omega)$ is a minimizing sequence, then $(u_j)_j$ is bounded in $BV^{\mathcal{B}}(\Omega)$ by the coercivity assumptions, and hence also in $W^{k-1, n/(n-1)}(\Omega, \mathbb{U})$ thanks to Theorem 2.2. Let $u \in BV^{\mathcal{B}}$ be a weak-* limit point of $(u_j)_j$. By the same reasoning as in the proof of Theorem 3.3 we have that (3.1) and (3.2) with $f = f_k$ hold. We now fix $1 \leq i \leq k-1$ and look at a Young measure ν 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 $(\alpha_i f_i(\cdot, \nabla^i u_j))_j$ is uniformly integrable. We can thus employ Proposition A.6 for $f = \alpha_i f_i$ to obtain

$$\begin{aligned} \liminf_{j \rightarrow \infty} \int_{\Omega} \alpha_i(x) f_i(x, \nabla^i u_j(x)) dx &= \int_{\Omega} \alpha_i(x) \langle \nu_x, f_i(x, \cdot) \rangle dx \\ &\geq \int_{\Omega} \alpha_i(x) f_i(x, \bar{\nu}_x) dx \\ &= \int_{\Omega} \alpha_i(x) f_i(x, \nabla^i u(x)) dx, \end{aligned}$$

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

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

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

The uniqueness follows exactly by the same argument as in the proof of Theorem 3.3, so the conclusion is achieved. \square

Remark 3.7. If $\Phi_g(u) = \|u - g\|_{L^1(\Omega)}$, uniqueness might fail in Theorem 3.6.

4. THE BILEVEL TRAINING SCHEME IN THE SPACE $\text{BV}_p^{\mathcal{B}}$

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 3.3. A variant involving functionals as in Theorem 3.6 will also be presented, see Remark 4.2.

Theorem 4.1. *Let us fix $p \in (1, +\infty)$, $g \in L^p(\Omega, \mathbb{U})$, $\underline{\alpha} > 0$ and $\alpha \in C(\bar{\Omega}, [\underline{\alpha}, +\infty))$. Let $f: \Omega \times \mathbb{V} \rightarrow [0, +\infty)$ be an integrand satisfying the assumptions outlined in Subsection 2.2, and suppose further that the recession function f^∞ 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 $\alpha^* \in \text{Adm}$ and it provides an associated optimally reconstructed image $u_{\alpha^*} \in \text{BV}_p^{\mathcal{B}}(\Omega)$.*

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

$$(4.1) \quad F(u_{\alpha^*}) \leq \lim_{j \rightarrow +\infty} F(u_j),$$

where we abbreviated $u_j := u_{\alpha_j}$ for a minimizer of (L2) with respect to the weight α_j , which exists in the light of Theorem 3.3.

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

$$(4.2) \quad I[u_j; \alpha_j] \leq I[v; \alpha_j] \quad \text{for any } v \in \text{BV}_p^{\mathcal{B}}(\Omega).$$

In particular, by selecting $v = 0$ and recalling that $\|\alpha_j\|_{L^\infty} \leq \bar{\alpha}$, we find that $I[u_j; \alpha_j] \leq C$ for some $C \geq 0$ independent of j . Then, owing to the coercivity of I , $(u_j)_j$ is bounded in $\text{BV}_p^{\mathcal{B}}(\Omega)$ and there exists $u \in \text{BV}_p^{\mathcal{B}}(\Omega)$ such that, upon extraction of subsequences, $u_j \xrightarrow{*} u$ in $\text{BV}_p^{\mathcal{B}}(\Omega)$. If we prove that $u = u_{\alpha^*}$, the conclusion is then achieved, since (4.1) would then follow by the lower semicontinuity of F .

We thus need to show that

$$(4.3) \quad I[u; \alpha^*] \leq I[v; \alpha^*] \quad \text{for any } v \in \text{BV}_p^{\mathcal{B}}(\Omega).$$

The uniform convergence of $(\alpha_j)_j$ along with (4.2) yields

$$(4.4) \quad \liminf_{j \rightarrow +\infty} I[u_j; \alpha_j] \leq \liminf_{j \rightarrow +\infty} I[v; \alpha_j] = I[v; \alpha^*] \quad \text{for any } v \in \text{BV}_p^{\mathcal{B}}(\Omega).$$

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

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

whence

$$(4.5) \quad \lim_{j \rightarrow +\infty} |I[u_j; \alpha_j] - I[u_j; \alpha^*]| = 0.$$

Finally, by the lower semicontinuity result in Theorem 3.3,

$$(4.6) \quad I[u; \alpha] \leq \liminf_{j \rightarrow \infty} I[u_j; \alpha^*],$$

so that, collecting (4.4)–(4.6), we obtain (4.3). The proof is thus complete. \square

If in the lower level problem (2.9) the functional I is replaced by \tilde{I} as in Theorem 3.6, a result in the same spirit of the one above holds. We only sketch it in the next remark, since it parallels closely Theorem 4.1.

Remark 4.2. Within the general framework of Subsection 2.2, we introduce a variant of the scheme (L1)–(L2). For $k \in \mathbb{N}$, $k \geq 2$, we define the sets

$$\begin{aligned} \text{Adm}_{\text{low}} &:= \left\{ \alpha \in C(\bar{\Omega}, [0, \bar{\alpha}]^{k-1}) : |\alpha_i(x) - \alpha_i(y)| \leq \omega(|x - y|) \text{ for every } i \text{ and } x, y \in \bar{\Omega} \right\}, \\ \widetilde{\text{Adm}} &:= \text{Adm}_{\text{low}} \times \text{Adm}, \end{aligned}$$

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

$$(4.7) \quad \text{find } \alpha^* \in \operatorname{argmin} \left\{ F(u_\alpha) : \alpha \in \widetilde{\text{Adm}} \right\}$$

$$(4.8) \quad \text{such that } u_\alpha \in \operatorname{argmin} \left\{ \tilde{I}[u; \alpha] : u \in \text{BV}^{\mathcal{B}}(\Omega) \right\},$$

where \tilde{I} is as in (3.3). Under the assumptions of Theorem 3.6, notably \mathbb{C} -ellipticity for \mathcal{B} , we are able to prove the existence of a solution, that is, an optimal regularizer $\alpha^* \in \widetilde{\text{Adm}}$ for (4.7). Let us outline the argument.

If $(\alpha^j)_j \subset \widetilde{\text{Adm}}$ is a minimizing sequence, we may assume that $\alpha^j \rightarrow \alpha^* \in \widetilde{\text{Adm}}$ uniformly. By Theorem 3.6 we can pick a sequence $(u_j)_j \subset \text{BV}^{\mathcal{B}}(\Omega)$ made of minimizers for (4.8) associated with $(\alpha^j)_j$. As a consequence of the coercivity of \tilde{I} , $(u_j)_j$ is bounded in $\text{BV}_p^{\mathcal{B}}(\Omega)$, and thus, owing to Theorem 2.2, also in $W^{k-1, n/(n-1)}(\Omega)$. Denoting by $u \in \text{BV}^{\mathcal{B}}(\Omega)$ the weak-* limit (up to subsequences) of $(u_j)_j$, the remainder of the proof follows the one of Theorem 4.2, the most significant difference being the use of Theorem 3.6 instead of Theorem 3.3 to obtain the analogue of (4.6).

5. NUMERICAL EXAMPLES

In this section we provide some numerical results for image reconstruction by focusing on some specific instances of the differential operators considered above. These numerical examples show the applicability and versatility of our approach, which, as we will see, is 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) [7] and its version with spatially varying weights [25]. Since our main target here is to evaluate the performance of the types of regularizers that we introduced, we restrict ourselves to two particular cases of image denoising. Firstly, in the class of first-order functionals, we consider a Huber-type TV regularization, with both the regularization parameter α and the Huber parameter γ being spatially dependent. This can be considered as a functional that incorporates a local choice between TV and Tikhonov regularization. The second example is a spatially varying TV^2 regularization, which is a second-order functional and has the capability to improve the reconstructions by eliminating the undesirable staircasing effect of TV [33]. 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.

5.1. Spatially varying Huber versions of TV and TV². Let $\gamma \in L^\infty(\Omega)$, with $\gamma \geq 0$, be fixed. We define the spatially varying Huber function $f_\gamma: \Omega \times \mathbb{R}^n \rightarrow \mathbb{R}$ as follows:

$$(5.1) \quad f_\gamma(x, z) = \begin{cases} |z| - \frac{1}{2}\gamma(x), & \text{if } |z| \geq \gamma(x), \\ \frac{1}{2\gamma(x)}|z|^2, & \text{if } |z| < \gamma(x). \end{cases}$$

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

$$(5.2) \quad |z| - \frac{1}{2}\|\gamma\|_{L^\infty(\Omega)} \leq f_\gamma(x, z) \leq |z|.$$

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

$$\text{TV}_\gamma(u) := |f_\gamma(Du)|(\Omega) := \int_\Omega f_\gamma(x, dDu).$$

An easy check shows that the recession function of f_γ (cf. (2.5)) is $f_\gamma^\infty(x, z) = |z|$. Thus, all the assumptions of Theorem 4.1 are trivially satisfied. Consequently, TV_γ is indeed well-defined as

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

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

$$(5.3) \quad \text{TV}_{\alpha, \gamma}(u) = \int_\Omega \alpha f_\gamma(x, \nabla u) dx + \int_\Omega \alpha d|D^s u|.$$

Note that $\text{TV}_{\alpha, \gamma}(u)$ can be equivalently defined via duality [21]:

$$(5.4) \quad \text{TV}_{\alpha, \gamma}(u) = \sup \left\{ \int_\Omega u \operatorname{div} \phi \, dx - \mathcal{I}_{\{|\cdot(x)| \leq \alpha(x)\}}(\phi) - \frac{1}{2} \int_\Omega \frac{\gamma}{\alpha} |\phi|^2 \, dx : \phi \in C_c^\infty(\Omega, \mathbb{R}^n) \right\}.$$

This is an immediate extension of the standard Huber TV functional (1.9) mentioned in the introduction. Similarly, for a function $u \in \text{BV}^2(\Omega) := \{u \in W^{1,1}(\Omega) : D^2 u \in \mathcal{M}(\Omega, \mathcal{S}^{n \times n})\}$ and $\alpha \in C(\overline{\Omega})$ with $\alpha(x) \geq \underline{\alpha} > 0$, we define the spatially varying Huber TV² functional as

$$(5.5) \quad \text{TV}_{\alpha, \gamma}^2(u) = \int_\Omega \alpha f_\gamma(x, \nabla^2 u) \, dx + \int_\Omega \alpha d|(D^2)^s u|,$$

where f_γ now is a function defined on $\Omega \times \mathbb{R}^{n \times n}$ defined by the natural analogue of (5.1). Again via duality, we find the equivalent expression

$$(5.6) \quad \text{TV}_{\alpha, \gamma}^2(u) = \sup \left\{ \int_\Omega u \operatorname{div}^2 \phi \, dx - \mathcal{I}_{\{|\cdot(x)| \leq \alpha(x)\}}(\phi) - \frac{1}{2} \int_\Omega \frac{\gamma}{\alpha} |\phi|^2 \, dx : \phi \in C_c^\infty(\Omega, \mathbb{R}^n) \right\}.$$

Our examples concern the following lower level image denoising problems:

$$(5.7) \quad I_\gamma^1[u; \alpha] := \int_\Omega (u - g)^2 \, dx + \text{TV}_{\alpha, \gamma}(u),$$

$$(5.8) \quad I_\gamma^2[u; \alpha] := \int_\Omega (u - g)^2 \, dx + \text{TV}_{\alpha, \gamma}^2(u).$$

5.2. The bilevel problems. The family of bilevel problems for the automatic computation of the spatial regularization parameter α associated with the functional I_γ^i for $i = 1, 2$ is:

$$(5.9) \quad \text{find } \alpha^* \in \operatorname{argmin} \{F(u_\alpha) : \alpha \in \operatorname{Adm}\}$$

$$(5.10) \quad \text{such that } u_\alpha = \operatorname{argmin} \{I_\gamma^i[u; \alpha] : u \in \operatorname{BV}^i(\Omega)\}.$$

In view of (5.2), the lower level problems (5.10) are well-defined, while from Theorem 4.1 we know that the overall schemes (5.9)–(5.10) 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 :

$$(5.11) \quad F_{\text{PSNR}}(u) = \int_{\Omega} |u - u_{\text{gt}}|^2 dx,$$

$$(5.12) \quad F_{\text{stat}}(u) = \frac{1}{2} \int_{\Omega} \max(Ru - \bar{\sigma}^2, 0)^2 dx + \frac{1}{2} \int_{\Omega} \min(Ru - \underline{\sigma}^2, 0)^2 dx,$$

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

The first cost functional corresponds to a maximization of the PSNR of the reconstruction and requires the knowledge of the ground truth u_{gt} [11, 19, 18, 17, 31], while the second enforces the localized residuals Ru to belong in a certain tight corridor $[\underline{\sigma}^2, \bar{\sigma}^2] := [\sigma^2 - \epsilon, \sigma^2 + \epsilon]$, σ^2 being the variance of the noise η , which is assumed here to be Gaussian, see also [26, 27, 24, 25]. The latter option has the advantage of being ground truth free, but knowledge or a good estimate for the noise variance σ^2 is needed. For the discrete version of the averaging filter w in the definition of the localized residuals (5.12) 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 [26, 27, 24, 25] and add instead a small H^1 term of the weight function α in the upper level objective, together with a supplementary box constraint $\mathcal{C} := \{\alpha \in H^1(\Omega) : \underline{\alpha} \leq \alpha \leq \bar{\alpha}\}$ for some $\underline{\alpha}, \bar{\alpha} \in \mathbb{R}$ with $0 < \underline{\alpha} < \bar{\alpha}$. On the whole, we will use the following upper level objectives:

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

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

for some small $\lambda > 0$. We will denote by $\hat{\mathcal{F}}$ the corresponding reduced objective functionals, that is $\hat{\mathcal{F}}_{\text{PSNR/stat}}(\alpha) := \mathcal{F}_{\text{PSNR/stat}}(\alpha, u_\alpha)$. That leads us to the bilevel minimization problems that we tackle numerically: for $i = 1, 2$

$$(5.13) \quad \text{find } \alpha^* \in \operatorname{argmin}_{\alpha} \mathcal{F}_{\text{PSNR/stat}}(\alpha, u_\alpha)$$

$$(5.14) \quad \text{such that } \begin{cases} u_\alpha = \operatorname{argmin}_u I_\gamma^i[u; \alpha], \\ \alpha \in \mathcal{C}. \end{cases}$$

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 a regularity result of the H^1 -projection onto \mathcal{C} , denoted by $P_{\mathcal{C}}$ see [27, Corollary 2.3]. This projection is applied to every iteration of the projected gradient algorithm, which is to be used for the numerical solution of (5.13)–(5.14) and is described next, see Algorithm 1. In that case, it is ensured that the computed weight α^k at the k -th projected gradient

iteration belongs to $H^2(\Omega)$, which for $n = 2$ embeds compactly into any Hölder space $C^\beta(\bar{\Omega})$, $\beta \in (0, 1)$.

5.3. Strategy for fixing γ . Since in our set up the function γ is not part of the minimizing variables, it has to be fixed from the start. Our rationale for fixing γ is that we would like to regularize high detailed areas with a Tikhonov term with a spatially varying weight $\frac{1}{2} \int_{\Omega} \tilde{\alpha} |\nabla u|^2 dx$, 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 TV² term with a spatially varying weight α . This will happen if γ is large in such detailed areas in order to allow for the second case in (5.1) 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, since it is small on the edges and on the detailed areas of the image. We then invert this weight and set

$$(5.15) \quad \gamma = s \frac{1}{\tilde{\alpha}}$$

for some constant $s > 0$. By choosing the function γ as in (5.15), we have that when $\tilde{\alpha}$ is small (fine scale details), γ will be large and thus the second case in (5.1) 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 γ will be small and thus a TV or TV² term will be preferred, i.e., first case in (5.1). In the third images of the top rows of Figures 1 and 3, we see how the resulting γ 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 γ , one could alternatively employ some standard edge detector algorithms, like for instance the Canny method [12].

5.4. 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 (5.13)–(5.14). Similar algorithms were presented in [27, 25], so we limit ourselves to a brief description of the procedure, still providing all the necessary details to ensure reproducibility.

The lower level problems (5.7) and (5.8) are substituted by their primal-dual optimality conditions, see [28, 25]:

$$(5.16) \quad u - g - \operatorname{div} p = 0,$$

$$(5.17) \quad \max(|\nabla u|, \gamma) p - \alpha \nabla u = 0,$$

and

$$(5.18) \quad u - g + \operatorname{div}^2 p = 0,$$

$$(5.19) \quad \max(|\nabla^2 u|, \gamma) p - \alpha \nabla^2 u = 0,$$

with p denoting the corresponding dual variable of each problem. For notation ease, we compactly write the above equations as $G_1(u, p) = 0$ and $G_2(u, p) = 0$. The application of the max function as well the multiplication in (5.17) and (5.19) are regarded component wise; note that here both α and γ are spatially (i.e. pixel) dependent. We use standard forward and backward differences for the discretizations of ∇ and div , see e.g. [25], and similarly for the discretizations of ∇^2 and div^2 , see [33]. For the numerical solution of (5.16)–(5.17), we use a semismooth Newton algorithm as it is described in [25] for the TGV case. Note that we do not add an additional Laplacian term for u as in [28, 25], 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 (5.13)–(5.14), where the lower level problems are substituted by (5.16)–(5.17) and (5.18)–(5.19), we employ a discretized projected gradient approach with Armijo line search as it is described in [25], originated from [27]. The algorithm is summarized in Algorithm 1. We comment on the components that have not been clarified so far. The term Δ_N denotes the discrete Laplacian with zero Neumann boundary conditions. These are the desired boundary conditions for α , as dictated by the regularity results for the H^1 -projection P_C . This projection is computed exactly as in [25] by using the same method and parameters mentioned there. In our numerical computations we set $\underline{\alpha} = 10^{-8}$, $\bar{\alpha} = 5$, $n_w = 7$, $\lambda = 10^{-11}$, $\tau^0 = 10^{-3}$, $c = 10^{-12}$, $\theta_- = 0.25$, $\theta_+ = 2$. As initializations for α , we use the constant functions $\alpha^0 = 0.5$ and $\alpha^0 = 1$ for the TV and TV² problems respectively. Regarding $\underline{\sigma}, \bar{\sigma}$, we use the formulas $\underline{\sigma}^2 = \sigma^2(1 - \frac{\sqrt{2}}{n_w})$ and $\bar{\sigma}^2 = \sigma^2(1 + \frac{\sqrt{2}}{n_w})$, which are based on the statistics of the extremes, see [27, 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 $\text{maxit} = 100$, after which no noticeable reduction in the reduced objective function is observed, see also Figure 6.

In order to produce a spatially varying Huber parameter γ as described before, we solve an auxiliary bilevel Tikhonov problem where the lower level problem corresponds to $G_T(u) := u - \text{div}(\tilde{\alpha}\nabla u) - g = 0$, i.e., the first order optimality condition of a variational denoising problem with $\frac{1}{2} \int_{\Omega} \tilde{\alpha} |\nabla u|^2 dx$ 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 γ and the H^1 -projection P_C is substituted by a simple L^2 projection, that is, $\alpha^{k+1} = \max(\min(\alpha^k - \tau^k \nabla_{\alpha} \hat{\mathcal{F}}(\alpha^k), \bar{\alpha}), \underline{\alpha})$. We use 100 projected gradient iterations with $n_w, \underline{\sigma}, \bar{\sigma}, \tau^0, c, \theta_-, \theta_+$ as before, as well as $\underline{\alpha} = 10^{-8}$, $\bar{\alpha} = 15$, $\alpha_0 = 15$. The equation G_T is solved exactly with a linear system solver. We use again 100 projected gradient iterations to get an output weight $\tilde{\alpha}$. Then, as we mentioned in (5.15) we define $\gamma = s/\tilde{\alpha}$. In all our experiments, we set $s = 0.1$.

For comparison purposes, we also report TV and TV² denoising results with a scalar Huber parameter γ , which is always set $\gamma = 10^{-3}$. We do that for both scalar and spatially varying regularization parameters α . In the first case, we manually select the parameter α 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 [14] as described in [5]. For the scalar case, again we manually select the TGV parameters α_0, α_1 that maximize the PSNR. For the spatially varying case, we use the weights α_0, α_1 as produced in [25] 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 [25].

5.5. Numerical results. In Figure 1 we report our numerical results on the *Parrot* image, see also Figure 2 for zoom-in details. Here the spatially varying regularization weights α 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 α and scalar γ is able to better preserve the details around the eye of the parrot, third row first image. When we use the spatially varying γ , the details in that area become even more pronounced, compare

Algorithm 1Projected gradient for the bilevel huber TV (resp. TV²) problems (5.13)–(5.14)**Input:** $g, \underline{\alpha}, \bar{\alpha}, \underline{\sigma}, \bar{\sigma}, \lambda, \gamma, n_w, \tau^0, 0 < c < 1, 0 < \theta_- < 1 \leq \theta_+$ **Initialize:** $\alpha^0 \in \mathcal{C}$, and set $k = 0$.**repeat**

Use Semismooth Newton to solve (5.16)–(5.17) (resp. (5.18)–(5.19)), i.e.

$$G_1(u^k, p^k) = 0, \quad (\text{resp. } G_2(u^k, p^k) = 0)$$

Solve for (u^*, p^*) the adjoint equation ($i = 1$, resp. $i = 2$)

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

Compute the derivative of the reduced objective w.r.t. α ($i = 1$, resp. $i = 2$) as

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

Compute the reduced gradient

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

Compute the trial points

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

while

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

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

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

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

the first two images in the third row of Figure 2. This is also accompanied with a slight increase of the SSIM index but also with a decrease in PSNR. Observe that the weights α that are computed in these two cases are quite different, see first two images of the last row of Figure 1. The bilevel Huber TV² approach with spatially varying α , produces similar reconstructions for both the scalar (slightly higher SSIM) and the spatially varying γ 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 α to drop significantly in the detailed areas of the image, see the last two images of the last row of Figure 1. It is characteristic that while the PSNR of scalar TV² is only 0.03 dB higher than the one of scalar TV, the PSNR of bilevel Huber TV² with spatially varying α and scalar γ is 0.61 dB higher compared to the corresponding Huber TV result.

The superiority of the bilevel Huber TV² with spatially varying weights, is even more evident in the second image example *Hatchling*, Figures 3 and 4. Here the reconstruction is more challenging due to the oscillatory nature of the ground truth image. The bilevel Huber TV² with scalar γ gives by far the best result with respect to both PSNR and SSIM. Again, the automatically computed regularization weights α have much lower values in bilevel TV² than in bilevel TV, compare the first two versus the last two figures of the last row of Figure 3. In this example, the spatially varying γ leads to a reduction of PSNR

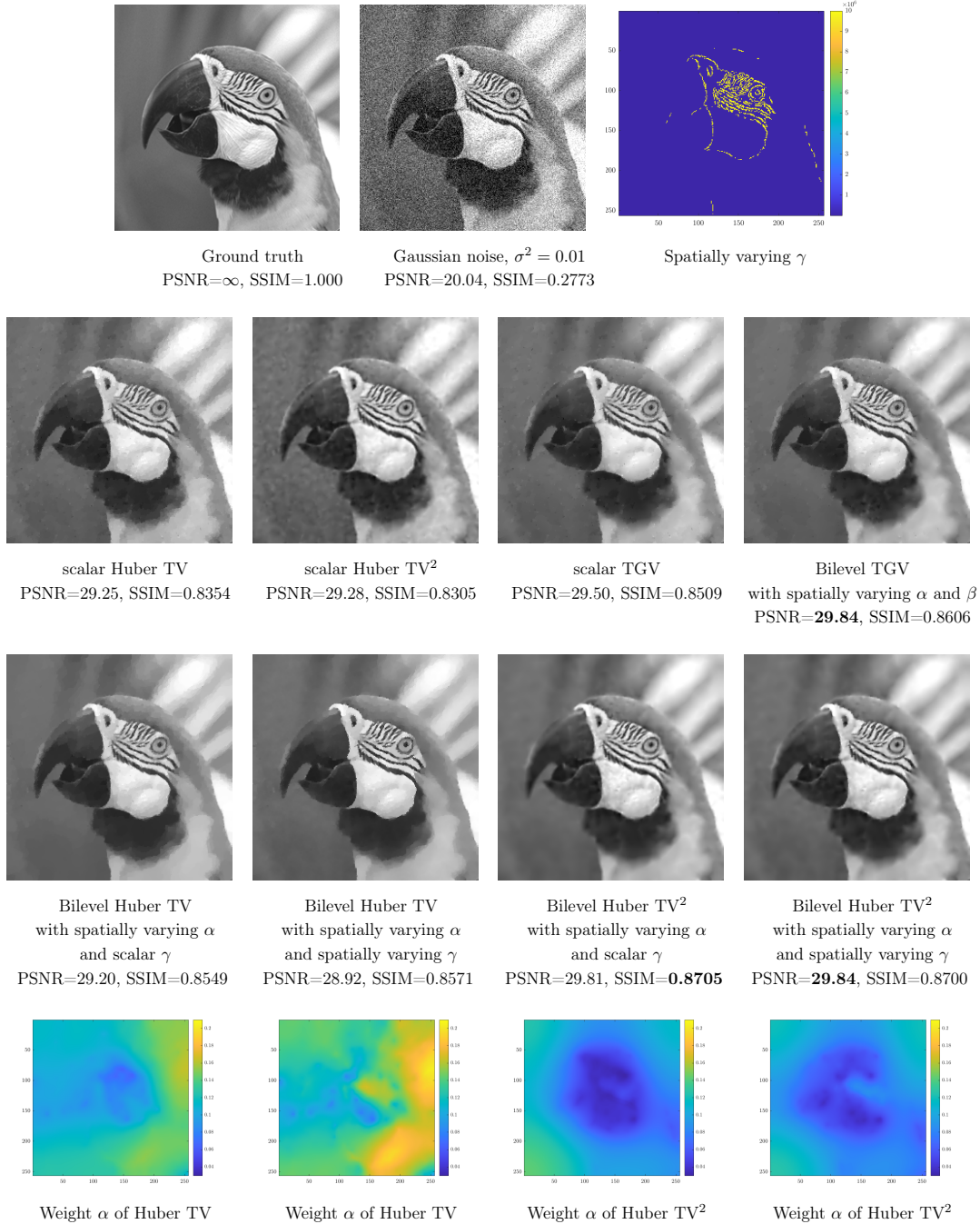


FIGURE 1. *Parrot* image: Huber TV and TV² denoising with spatially varying Huber parameter γ and regularization parameter α . The weights α are produced with the ground truth-free bilevel approach using $\mathcal{F}_{\text{stat}}$. The highest PSNR and SSIM values are highlighted in bold font.

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 4.

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 5. In both images, the highest PSNR and SSIM is achieved by the bilevel TV² with spatially varying α and scalar Huber parameter γ , third images of first and third row, with the corresponding the regularization weight α having

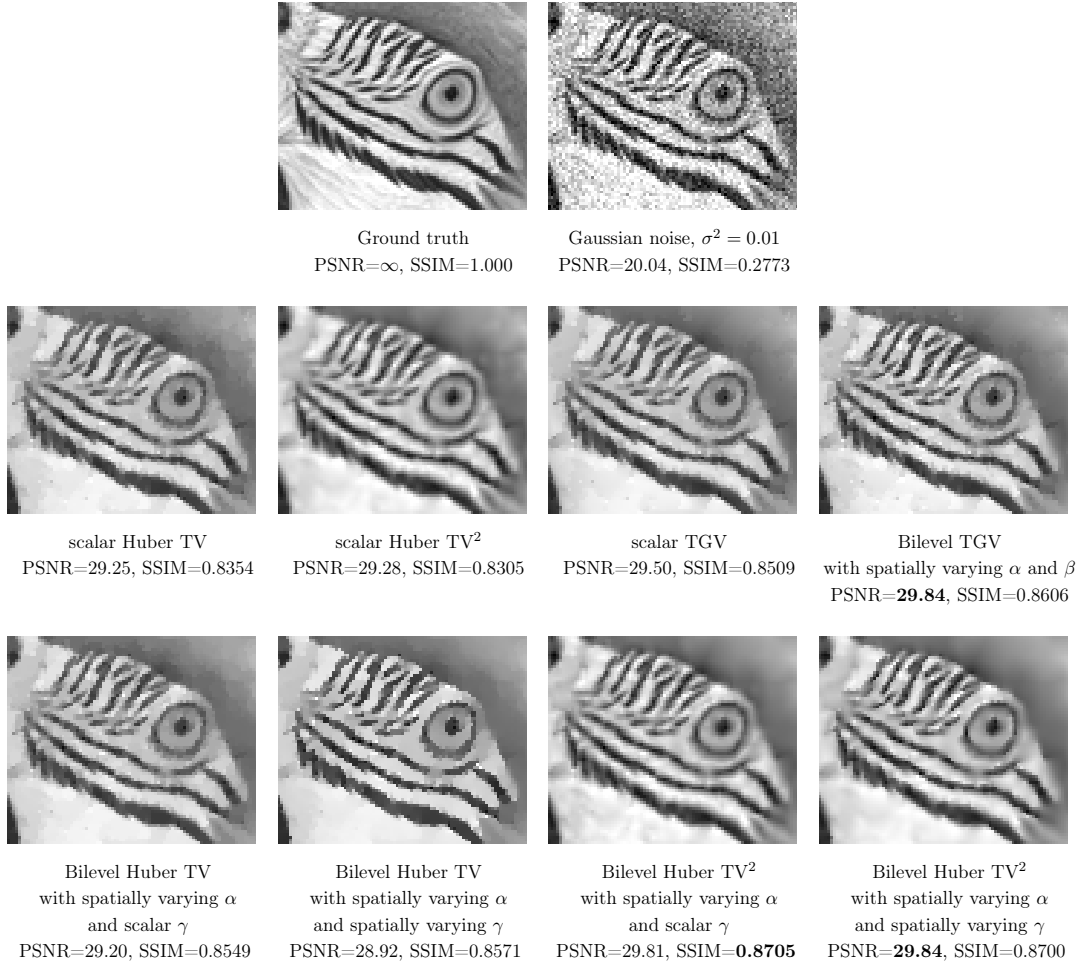


FIGURE 2. Details of images shown in Figure 1

again smaller values compared to the TV one. Nevertheless, we observe that the spatially varying γ results in higher SSIM in the Huber TV examples in both images, again with more pronounced features around the eye.

Finally in Figure 6, 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 TV² 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 γ in both Huber TV and Huber TV² 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 is 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 γ , compare again the red versus green and blue versus black plots.

We conclude that the bilevel Huber TV² is able to produce remarkably good results. This is perhaps even surprising as the use of its scalar version is not that popular due to its inability to preserve sharp edges. We showed that the use of a spatially varying Huber parameter γ can result in improved results both quantitatively and qualitatively, thus justifying our rigorous analytical study on spatially inhomogeneous integrands acting on TV-type regularizers. We also stress that by no means our strategy for setting γ is necessarily the optimal one. In fact, future work will involve setting up a bilevel framework

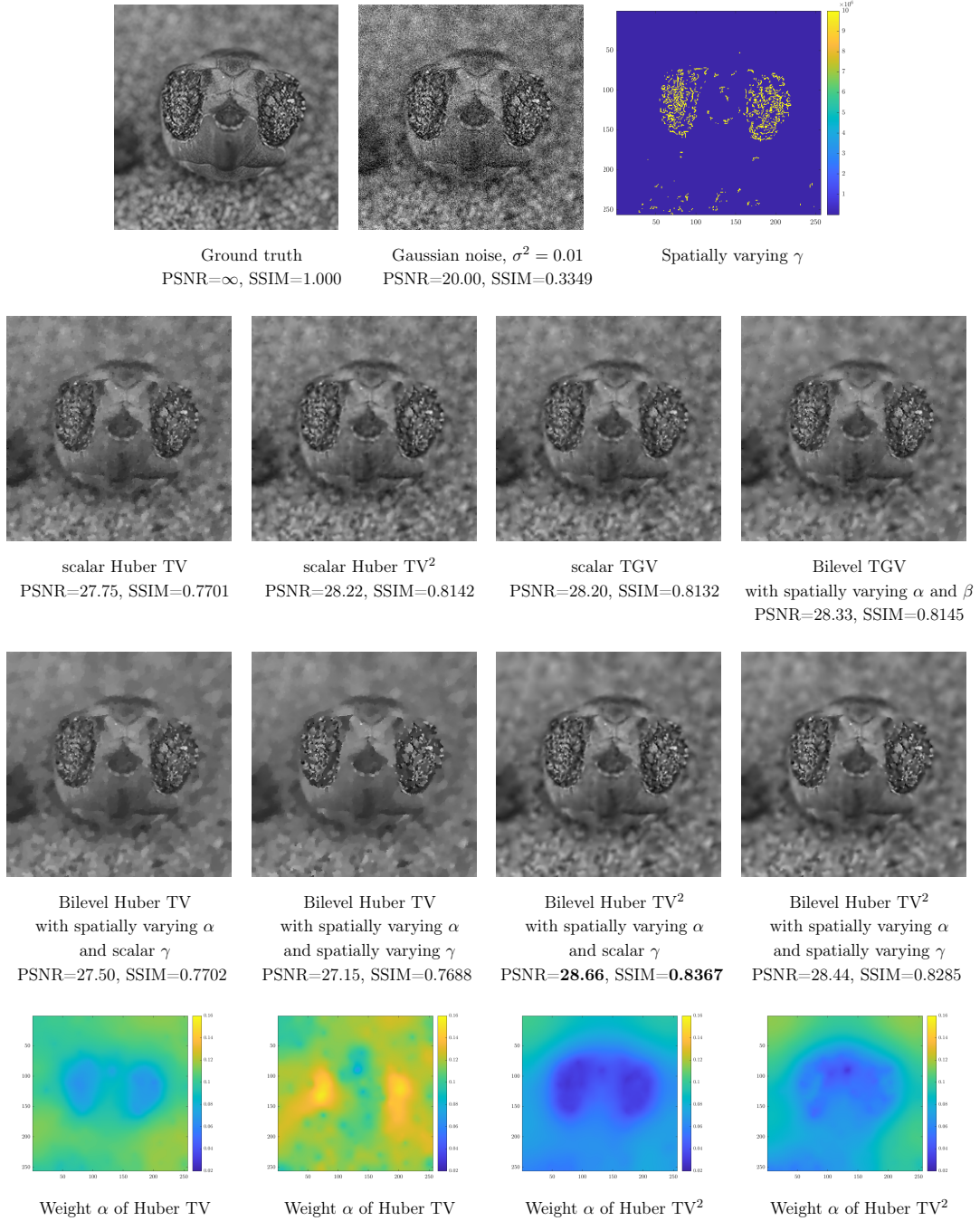


FIGURE 3. *Hatchling* image: Huber TV and TV² denoising with spatially varying Huber parameter γ and regularization parameter α . The weights α are produced with the ground truth-free bilevel approach using $\mathcal{F}_{\text{stat}}$. The highest PSNR and SSIM values are highlighted in bold font.

where also this parameter is included in the upper level minimization variables along with the parameter α , adding further flexibility to the regularization process.

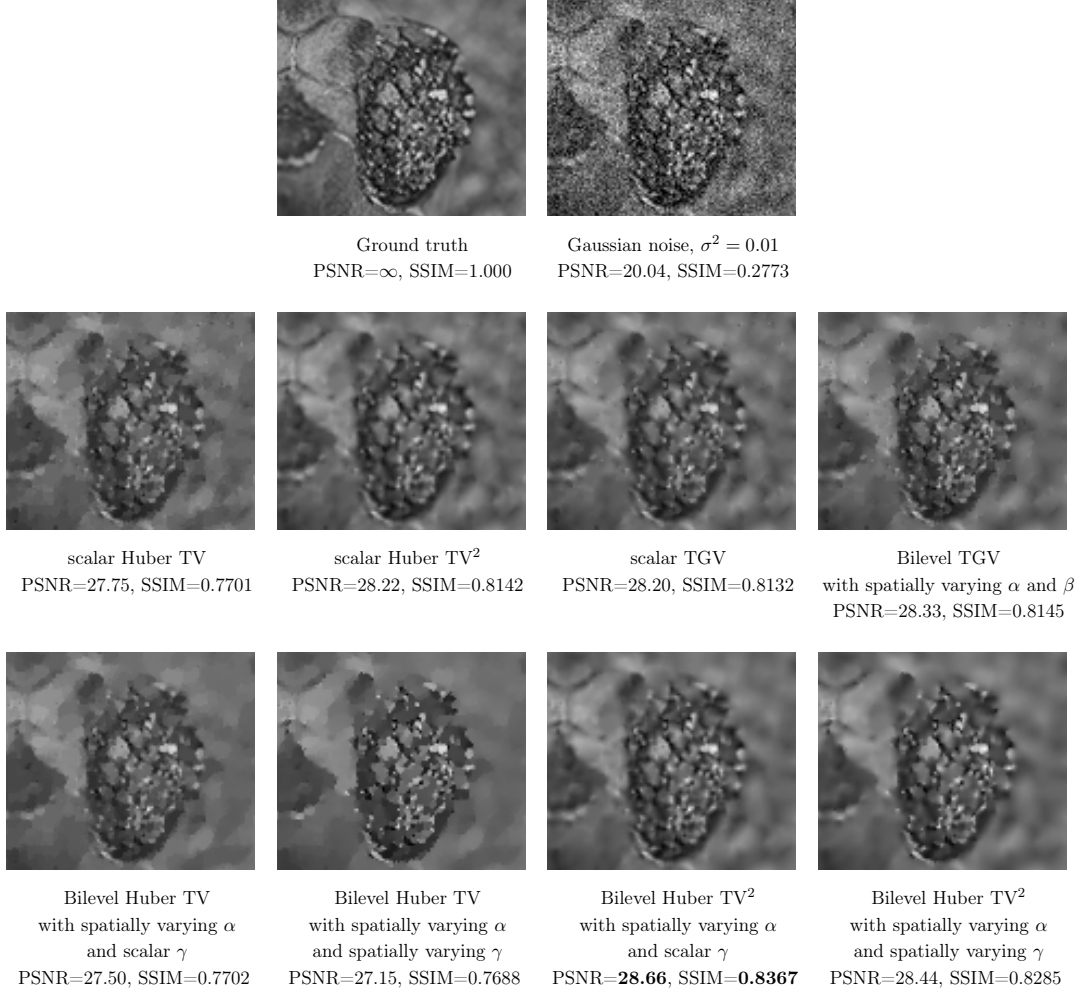


FIGURE 4. Details of images shown in Figure 3

APPENDIX A. CONVEX INTEGRANDS AND GENERALIZED YOUNG MEASURES

This section loosely follows [29], where most proofs can be found. Let $\Omega \subset \mathbb{R}^n$ be a bounded open set with $\mathcal{L}^n(\partial\Omega) = 0$ and consider the space of integrands

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

which is naturally equipped with the norm

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

It will thus be convenient to work with the coordinate transformations

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

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 $C(\bar{\Omega} \times B_{\mathbb{V}})$ via the linear isometric isomorphism

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

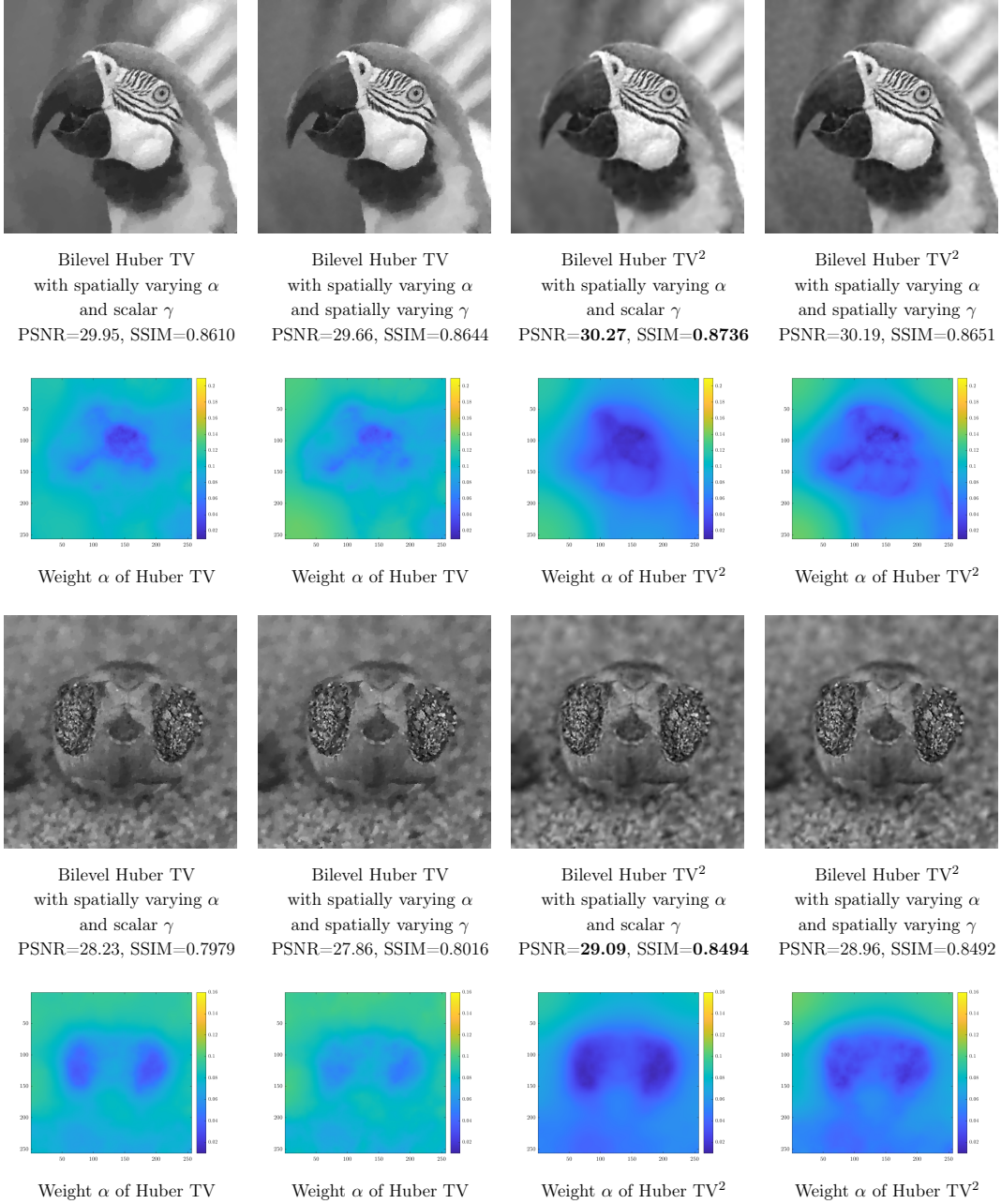


FIGURE 5. Huber TV and TV² denoising with spatially varying Huber parameter γ and regularization parameter α . The weights α are produced with the ground truth-based bilevel approach using $\mathcal{F}_{\text{PSNR}}$. The highest PSNR and SSIM values are highlighted in bold font.

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

$$\varepsilon_{\mu}(\Phi) := \int_{\Omega} \Phi(x, d\mu) = \int_{\Omega} \Phi(x, \mu^a(x)) dx + \int_{\Omega} \Phi^{\infty} \left(x, \frac{d\mu^s}{d\mu}(x) \right) d|\mu|(x),$$

where $\mu = \mu^a \mathcal{L}^n \llcorner \Omega + \mu^s$ is the Radon–Nykodim 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 (μ_j) is bounded in $\mathcal{M}(\Omega, \mathbb{V})$, we know that along a subsequence we have $\varepsilon_{\mu_j} \xrightarrow{*} \nu$ in $\mathbb{E}(\Omega, \mathbb{V})^*$. We define $\sigma := (T^{-1})^* \nu \in$

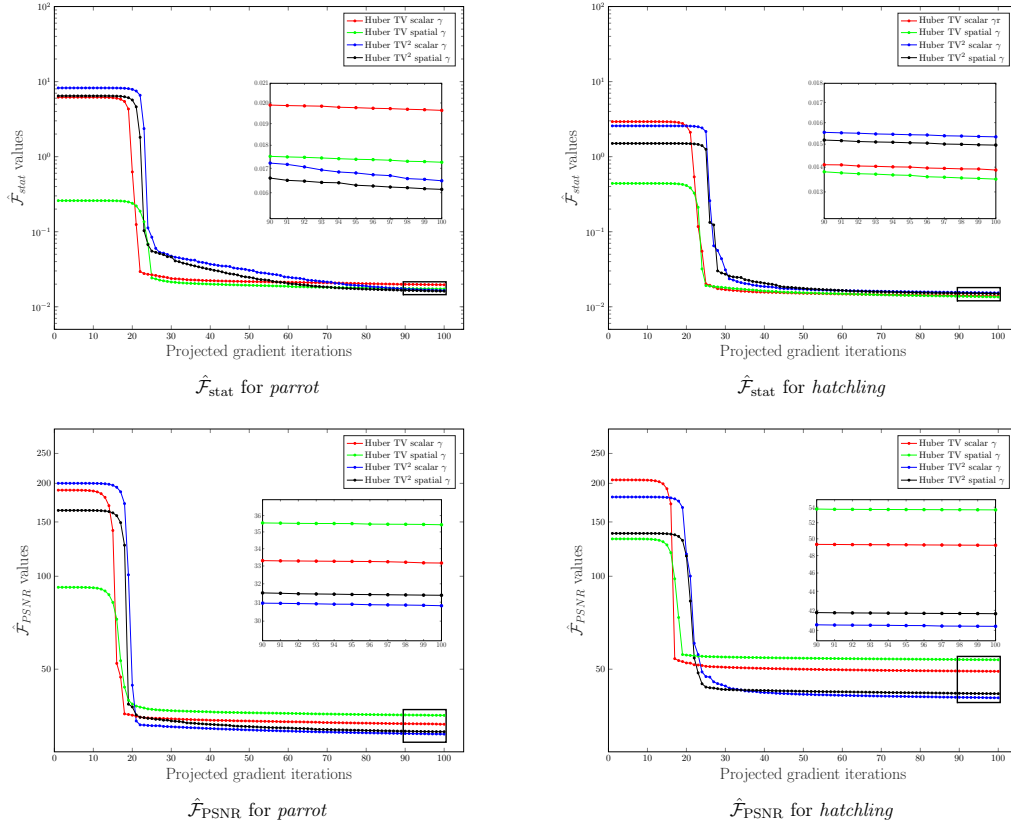


FIGURE 6. Values of the reduced objective $\hat{\mathcal{F}}(u^k)$ along the projected gradient iterations. The inner boxes show zoom-in plots of the last 10 iterations.

$\mathcal{M}(\bar{\Omega} \times B_{\mathbb{V}})$ and write for $\Phi \in \mathbb{E}(\Omega, \mathbb{V})$

$$\begin{aligned} \langle\langle \Phi, \nu \rangle\rangle &:= \langle \Phi, \nu \rangle_{\mathbb{E}, \mathbb{E}^*} = \langle T\Phi, \sigma \rangle \\ &= \int_{\bar{\Omega} \times B_{\mathbb{V}}} (1 - |\hat{z}|)^p \Phi \left(x, \frac{\hat{z}}{1 - |\hat{z}|} \right) d\sigma(x, \hat{z}) + \int_{\bar{\Omega} \times S_{\mathbb{V}}} \Phi^\infty(x, \hat{z}) d\sigma(x, \hat{z}). \end{aligned}$$

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

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

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

Definition A.1. A parametrized measure $\nu = ((\nu_x)_{x \in \Omega}, \lambda, (\nu_x^\infty)_{x \in \bar{\Omega}})$ 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-* \mathcal{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-* λ -measurable (the concentration-angle measure).
- (d) $\int_{\Omega} \int_{\mathbb{V}} |z| d\nu_x(z) dx < \infty$ (the moment condition holds).

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

$$\langle\langle \Phi, \nu \rangle\rangle := \int_{\Omega} \int_{\mathbb{V}} \Phi(x, \cdot) d\nu_x dx + \int_{\bar{\Omega}} \int_{S_{\mathbb{V}}} \Phi^\infty(x, \cdot) d\nu_x^\infty d\lambda(x) \quad \text{for } \Phi \in \mathbb{E}(\Omega, \mathbb{V}).$$

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

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

Theorem A.2. $Y(\Omega, \mathbb{V}) = T^*\{\sigma \in \mathcal{M}^+(\bar{\Omega} \times \bar{B}_{\mathbb{V}}) : \text{equation (A.1) holds}\}$.

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

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

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

for all $\Phi \in \mathbb{E}(\Omega, \mathbb{V})$. In this case, we say that (μ_j) generates ν .

In our analysis we repeatedly use the following:

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

$$\mu_j \xrightarrow{*} \bar{\nu}_x \mathcal{L}^n \llcorner \Omega + \bar{\nu}_x^{\infty} \lambda \text{ in } \mathcal{M}(\bar{\Omega}, \mathbb{V}).$$

The limit measure is referred to as the barycentre of ν .

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

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

for the expectations of the probability measures ν_x and ν_x^{∞} .

We also employ a general convergence result for Young measures:

Proposition A.5. [30, Prop. 2(i)] *Let $\Omega \subset \mathbb{R}^n$ be bounded and open, and $F: \Omega \times \mathbb{V} \rightarrow \mathbb{R}$ be a measurable integrand such that $f(x, \cdot)$ is continuous for almost every $x \in \Omega$ (a Carathéodory integrand). Suppose in addition that f has a regular recession function, i.e.,*

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

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

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

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 $F: \Omega \times \mathbb{V} \rightarrow \mathbb{R}$ be a measurable integrand such that $f(x, \cdot)$ is continuous for almost every $x \in \Omega$ (a Carathéodory integrand). Let $(v_j) \subset L^1(\Omega, \mathbb{V})$ generate $\nu \in Y(\Omega, \mathbb{V})$ be such that $(f(\cdot, v_j))_j$ is uniformly integrable. Then*

$$\lim_{j \rightarrow \infty} \int_{\Omega} f(x, v_j(x)) dx = \int_{\Omega} \langle \nu_x, f(x, \cdot) \rangle dx.$$

REFERENCES

1. M. Amar, V. De Cicco, and N. Fusco, *Lower semicontinuity and relaxation results in BV for integral functionals with BV integrands*, 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/AJN0.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. 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>.
8. 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.
9. D. Breit, L. Diening, and F. Gmeineder, *On the trace operator for functions of bounded Λ -variation*, Analysis & PDE **13** (2020), no. 2, 559 – 594, <https://doi.org/10.2140/apde.2020.13.559>.
10. M. Burger, K. Papafitsoros, E. Papoutsellis, and C.B. Schönlieb, *Infimal convolution regularisation functionals of BV and 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>.
11. 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>.
12. 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>.
13. 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>.
14. 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>.
15. 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>.
16. 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>.
17. 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>.
18. 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>.
19. 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>.
20. 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>.
21. F. Demengel and R. Temam, *Convex functions of a measure and applications*, Indiana University Mathematics Journal **33** (1984), 673–709.
22. 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>.
23. F. Gmeineder and B. Raiță, *Embeddings for α -weakly differentiable functions on domains*, Journal of Functional Analysis **277** (2019), no. 12, 108278, <https://www.sciencedirect.com/science/article/pii/S0022123619302411>.
24. 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.
25. M. Hintermüller, K. Papafitsoros, C.N. Rautenberg, and H. Sun, *Dualization and automatic distributed parameter selection of total generalized variation via bilevel optimization*, arXiv preprint arXiv:2002.05614 (2021), <https://arxiv.org/abs/2002.05614>.
 26. 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>.
 27. 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>.
 28. 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>.
 29. J. Kristensen and B. Raiță, *An introduction to generalized young measures*, Lecture Notes 45/2020, Max Planck Institute for Mathematics in the Sciences (2020).
 30. 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>.
 31. 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>.
 32. 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>.
 33. 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>.
 34. 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](http://dx.doi.org/10.1016/0167-2789(92)90242-F).
 35. K.T. Smith, *Formulas to represent functions by their derivatives*, Mathematische Annalen **188** (1970), no. 1, 53–77, <http://eudml.org/doc/162037>.
 36. 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>.
 37. L. Vese, *A study in the BV space of a denoising-deblurring variational problem*, Applied Mathematics and Optimization **44** (2001), no. 2, 131–161.

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.

E-mail: raita@mis.mpg.de

A. Vekelis.

University of Sussex,
Falmer, Brighton BN1 9RH, United Kingdom.

E-mail: A.Vekelis@sussex.ac.uk