DYADIC PARTITION-BASED TRAINING SCHEMES FOR TV/TGV DENOISING

ELISA DAVOLI, RITA FERREIRA, IRENE FONSECA, AND JOSÉ A. IGLESIAS

ABSTRACT. Due to their ability to handle discontinuous images while having a well-understood behavior, regularizations with total variation (TV) and total generalized variation (TGV) are some of the best-known methods in image denoising. However, like other variational models including a fidelity term, they crucially depend on the choice of their tuning parameters. A remedy is to choose these automatically through multilevel approaches, for example by optimizing performance on noisy/clean image pairs. In this work, we consider such methods with space-dependent parameters which are piecewise constant on dyadic grids, with the grid itself being part of the minimization. We prove existence of minimizers for fixed discontinuous parameters, that box constraints for the values of the parameters lead to existence of finite optimal partitions, and converse results for well-prepared data. On the numerical side, we consider a simple subdivision scheme for optimal partitions built on top of any other bilevel optimization method for scalar parameters, and demonstrate its improved performance on some representative test images when compared with constant optimized parameters.

DATE: May 12, 2023.

CONTENTS

2020 *Mathematics Subject Classification.* 2020 Mathematics Subject Classification: 68U10, 26A45, 49J10, 94A08.

Key words and phrases. total variation, total generalized variation, discontinuous weights, spatially-dependent regularization parameters, box constraint, bilevel optimization.

1. Introduction

A fundamental problem in image processing is the restoration of a given "noisy" image. Images are often deteriorated due to several factors occurring, for instance, in the process of transmission or acquisition, such as blur caused by motion or a deficient lens adjustment.

A well-established and successful approach for image restoration is hinged on variational PDE methods, where minimizers of certain energy functionals provide the sought "clean" and "sharp" images. In the particular case where the degradation consists of additive noise, these energy functionals usually take the form

$$E(u) := \|u - u_\eta\|_X^p + R_\alpha(u) \quad \text{for } u \in \widetilde{X}, \tag{1.1}$$

where u_η represents the given noisy image and $\widetilde{X}$ is the class of possible reconstructions of u_η. The first term in (1.1), $\|u - u_\eta\|_X^p$, is the fidelity or data fitting term that, in a minimization process, controls the distance between u and u_η in some space X. The second term, $R_\alpha(u)$, is the so-called filter term, and is responsible for the regularization of the images. The parameter α is often called a tuning or regularization parameter, and accounts for a balance between the fidelity and filter terms.

A milestone approach in imaging denoising is due to Rudin, Osher, and Fatemi [51], who proposed an energy functional of the type (1.1) with $X := L^2(Q)$, $p := 2$, $\widetilde{X} := BV(Q)$, and $R_\alpha(u) := \alpha TV(u, Q)$ with $\alpha > 0$, where $Q \subset \mathbb{R}^2$ is the image domain and $TV(u, Q)$ is the total variation in Q of a function of bounded variation $u \in BV(Q)$. Precisely, given an observed noisy version u_η of a true image, Rudin, Osher, and Fatemi [51] suggested to find the best reconstruction of the original clean image as the solution of the minimization problem

$$\min_{\substack{u \in BV(Q)\\ u_\eta - u \in L^2(Q)}} \Big\{ \|u - u_\eta\|_{L^2(Q)}^2 + \alpha TV(u, Q) \Big\}, \tag{1.2}$$

which is known as the TV or ROF model. A striking feature of the TV model is that it removes noise while preserving images' edges. This model has been extended in several ways, including higher-order and vectorial settings to address color images, and gave rise to numerous related filter terms seeking to overcome some of its drawbacks, such as blurring and the staircasing effect (see, for instance, [2, 20, 8] for an overview).

In a nutshell, the TV model yields functions u that best fit the data, measured in terms of the L^2 norm, and whose gradient (total variation) is low so that noise is removed. The choice of the parameter α plays a decisive role in the success of this and similar variational approaches, as it balances the fitting and regularization features of such models. In fact, higher values of α in (1.2) lead to an oversmoothed reconstruction of u_η because the total variation has to be "small" to compensate for high values of α; conversely, lower values of α in (1.2) inhibit noise removal and, in particular, the reconstructed image provided by (1.2) converges to u_η as $\alpha \to 0$ (see [31]).

Until recently, the "optimal" parameter α was chosen individually for each noisy image. This raises an obvious practical question given the huge amount of images generated daily. To address this issue, a partial automatic selection of an "optimal" parameter α was proposed in [31, 32] (see also [21, 22, 34, 52]) in the flavour of Machine Learning optimization schemes. This automatic selection is based on a bilevel optimization scheme searching for the optimal α that minimizes the distance, in some space, between the reconstruction of a noisy image and the original clean image. In this setting, both the noisy image, u_η, and the original clean image, u_c, are known a priori and called the training data. The rationale is to use the same parameter α to reconstruct noisy images that are *qualitatively similar* to that of the training scheme and thus expected to require a similar balance between fitting and regularization effects.

In the context of the TV model in (1.2), one such bilevel optimization scheme reads as follows. Here, and in the sequel, $\mathbb{R}^+$ stands for the set of positive real numbers, $(0, \infty)$.

$(\mathcal{LS})_{TV}$ **TV learning scheme** (1.3)

Level 1. Find

$$\bar{\alpha} = \operatorname{arginf} \left\{ \int_Q |u_c - u_\alpha|^2 \, \mathrm{d}x : \alpha \in \mathbb{R}^+ \right\};$$

Level 2. Given $\alpha \in \mathbb{R}^+$, find

$$u_\alpha = \operatorname{argmin} \left\{ \int_Q |u_\eta - u|^2 \, \mathrm{d}x + \alpha TV(u, Q) : u \in BV(Q) \right\}.$$

This approach yields a unified way of identifying the best fitting parameters for every class of training data lying in the same L^2-neighborhood. However, the learning scheme (1.3) does not address a major drawback of the TV and similar models using scalar regularization parameters. In fact, it does not take into account possible inhomogeneous noise (occurring, e.g., in magnetic resonances for brain images) and other local features in a given deteriorated image that would benefit from an adapted treatment.

A solution to this issue consists in resorting to adaptive methods and varying fitting parameters instead. The mathematical literature in this direction is vast, from which we single out the following contributions: [40, 49] for results in the finite-dimensional case and for optimal image filters, [30] for bilevel learning in function spaces and development of numerical optimization, [37] for a study of adaptive total variation denoising, [27, 26, 43, 42, 44] for a study of optimal regularizers, [45] for a bilevel analysis of novel classes of semi-norms, [46] for an approach via Young measures, [12] and the references therein for an overview.

As originally pointed out in Pan Liu's PhD thesis [41], a relevant question in image reconstruction is the possibility of adapting the fitting parameters to the specific features of a given class of noisy images by performing, e.g., a stronger regularization in areas which have been highly deteriorated and by tuning down the filtering actions in portions that, instead, have been left unaffected.

Here, starting from the ideas in [41], we propose space-dependent learning schemes that locally search for the optimal level of refinement and the optimal regularization parameters. The optimal level of refinement translates into finding an optimal partition of the noisy image's domain that takes into account its local features. Precisely, as before, $Q = (0,1)^2$ represents the images' domain. We say that $\mathscr{L}$ is an admissible partition of Q if it consists of dyadic squares, each of which we often denote by L. Note that an admissible partition might be more or less refined in different parts of the domain. We denote by $\mathscr{P}$ the class of all such admissible partitions $\mathscr{L}$ of Q. Finally, let $(u_\eta, u_c) \in BV(Q) \times BV(Q)$ be a training pair of noisy and clean images. The first space-dependent learning scheme that we propose to restore u_η, based on the a priori knowledge of u_c, is as follows.

$(\mathscr{L}\mathscr{S})_{TV_\omega}$ **Weighted-TV learning scheme** (1.4)

Level 3. (optimal local training parameter) Fix $\mathscr{L} \in \mathscr{P}$; for each $L \in \mathscr{L}$, find

$$\alpha_L := \inf\left\{\operatorname{arginf}\left\{\int_L |u_c - u_\alpha|^2 \, \mathrm{d}x \colon \alpha \in \mathbb{R}^+\right\}\right\}, \tag{1.5}$$

where, for $\alpha \in \mathbb{R}^+$,

$$u_\alpha := \operatorname{argmin}\left\{\int_L |u_\eta - u|^2 \, \mathrm{d}x + \alpha TV(u, L) \colon u \in BV(L)\right\}. \tag{1.6}$$

Level 2. (space-dependent image denoising) For each $\mathscr{L} \in \mathscr{P}$, find

$$u_{\mathscr{L}} := \operatorname{argmin}\left\{\int_Q |u_\eta - u|^2 \, \mathrm{d}x + TV_{\omega_{\mathscr{L}}}(u, Q) \colon u \in BV_{\omega_{\mathscr{L}}}(Q)\right\}, \tag{1.7}$$

where we consider the piecewise constant weight $\omega_{\mathscr{L}}$ defined by

$$\omega_{\mathscr{L}}(x) := \sum_{L \in \mathscr{L}} \alpha_L \chi_L(x) \quad \text{with } \alpha_L \text{ given by Level 3,} \tag{1.8}$$

and $BV_{\omega_{\mathscr{L}}}$ is the space of $\omega_{\mathscr{L}}$-weighted BV-functions (see Section 3.2).

Level 1. (optimal partition and image restoration) Find

$$u^* \in \operatorname{argmin}\left\{\int_Q |u_c - u_{\mathscr{L}}|^2 \, \mathrm{d}x : \mathscr{L} \in \mathscr{P}\right\} \quad \text{with } u_{\mathscr{L}} \text{ given by Level 2.}$$

Remark 1.1. **(i)** We observe that by taking the infimum in (1.5), the corresponding parameter α_L is always well defined. On the other hand, if $TV(u_\eta, L) > TV(u_c, L)$ and $\|u_\eta - u_c\|^2_{L^2(L)} < \|[u_\eta]_L - u_c\|^2_{L^2(L)}$, with $[u_\eta]_L := \frac{1}{|L|}\int_L u_\eta \, \mathrm{d}x$, we prove in Theorem 3.8 that there exists $\tilde{\alpha}_L \in (0, \infty)$ satisfying

$$\tilde{\alpha}_L \in \operatorname{argmin}\left\{\int_L |u_c - u_\alpha|^2 \, \mathrm{d}x \colon \alpha \in \mathbb{R}^+\right\}$$

(see [31] for similar statements), in which case the infimum on such $\tilde{\alpha}_L$ as in (1.5) may be regarded as a choice criterium on the optimal parameter.

(ii) We refer to Section 3.2 for the definition and discussion of the space $BV_{\omega_{\mathscr{L}}}$ of $\omega_{\mathscr{L}}$-weighted BV-functions, as introduced in [3]. In particular, using the results in [3] (and also [14, 13]), we prove under appropriate conditions that $u_{\mathscr{L}} \in BV(Q)$ and

$$TV_{\omega_{\mathscr{L}}}(u_{\mathscr{L}}, Q) = \int_Q \omega_{\mathscr{L}}^{sc^-}(x)\,\mathrm{d}|Du_{\mathscr{L}}|(x), \tag{1.9}$$

where $\omega_{\mathscr{L}}^{sc^-}$ denotes the lower-semicontinuous envelope of $\omega_{\mathscr{L}}$.

The existence of solutions to the learning scheme $(\mathscr{LS})_{TV_\omega}$ in (1.4) is intimately related to the existence of a stopping criterion for the refinement of the admissible partitions or, in other words, a lower bound on the size of the dyadic squares $L \in \mathscr{L}$, with $\mathscr{L} \in \mathscr{P}$. This notion is made precise in the following definition.

Definition 1.2 (stopping criterion for the refinement of the admissible partitions). *We say that a condition* $(\mathcal{S})$ *on* $\mathscr{P}$ *is a stopping criterion for the refinement of the admissible partitions if there exist* $\kappa \in \mathbb{N}$ *and* $\mathscr{L}_1, ..., \mathscr{L}_\kappa \in \mathscr{P}$ *such that*

$$\operatorname{argmin}\left\{\int_Q |u_c - u_{\mathscr{L}}|^2\,\mathrm{d}x : \mathscr{L} \in \mathscr{P}\right\} = \operatorname{argmin}\left\{\int_Q |u_c - u_{\mathscr{L}_i}|^2\,\mathrm{d}x : i \in \{1, ..., \kappa\}\right\}$$

provided that $(\mathcal{S})$ *holds, where* $u_{\mathscr{L}}$ *and* $u_{\mathscr{L}_i}$ *are given by* (1.7). *In this case, we write* $\bar{\mathscr{P}} := \bigcup_{i=1}^{\kappa}\{\mathscr{L}_i\}$.

We refer to Section 3.4 for examples of stopping criteria as in Definition 1.2, from which we highlight the box-constraint that we discuss next.

Remark 1.3 (box constraint as a stopping criterion). To prove the existence of a solution to the learning scheme $(\mathscr{LS})_{TV_\omega}$ in (1.4), we adopt the usual box-constraint approach in which we replace $\alpha \in \mathbb{R}^+$ by

$$\alpha \in \left[c_0, \frac{1}{c_0}\right] \quad \text{for some } c_0 \in \mathbb{R}^+. \tag{1.10}$$

In this case, the analog of (1.5) becomes by

$$\bar{\alpha}_L = \inf\left\{\operatorname{argmin}\left\{\int_L |u_c - u_\alpha|^2\,\mathrm{d}x : \alpha \in \left[c_0, \tfrac{1}{c_0}\right]\right\}\right\}. \tag{1.11}$$

Under some assumptions on the training data, we prove in Subsection 3.4 (see Theorem 1.4 below) that this box constraint is equivalent to the existence of a stopping criterion for the refinement of the admissible partitions as in Definition 1.2.

Theorem 1.4 (Equivalence between box constraint and stopping criterion). *Consider the learning scheme* $(\mathscr{LS})_{TV_\omega}$ *in* (1.4). *The two following conditions hold:*

(a) If we replace (1.5) *by* (1.11), *then there exists a stopping criterion* $(\mathcal{S})$ *for the refinement of the admissible partitions as in Definition 1.2.*

(b) Assume that there exists a stopping criterion $(\mathcal{S})$ *for the refinement of the admissible partitions as in Definition 1.2 such that the training data satisfies for all* $L \in \bigcup_{\mathscr{L}\in\bar{\mathscr{P}}}\mathscr{L}$, *with* $\bar{\mathscr{P}}$ *as in Definition 1.2, the conditions*
 (i) $TV(u_c, L) < TV(u_\eta, L)$;
 (ii) $\|u_\eta - u_c\|^2_{L^2(L)} < \|[u_\eta]_L - u_c\|^2_{L^2(L)}$, *where* $[u_\eta]_L = \frac{1}{|L|}\int_L u_\eta\,\mathrm{d}x$.

Then, there exists $c_0 \in \mathbb{R}^+$ *such that the optimal solution* u^* *provided by* $(\mathscr{LS})_{TV_\omega}$ *with* $\mathscr{P}$ *replaced by* $\bar{\mathscr{P}}$ *coincides with the optimal solution* u^* *provided by* $(\mathscr{LS})_{TV_\omega}$ *with* (1.5) *replaced by* (1.11).

Next, we state our main theorem regarding existence of solutions for the learning scheme $(\mathscr{LS})_{TV_\omega}$ in (1.4). We state this result under the box-constraint condition. However, in view of Theorem 1.4, this result holds true under any stopping criterion for the refinement of the admissible partitions provided that the training data satisfies suitable conditions.

Theorem 1.5 (Existence of solutions to $(\mathscr{LS})_{TV_\omega}$). *There exists an optimal solution* u^* *to the learning scheme* $(\mathscr{LS})_{TV_\omega}$ *in* (1.4) *with* (1.5) *replaced by* (1.11).

The proofs of Theorems 1.4 and 1.5 are presented in Section 3, where we also explore alternative stopping criteria.

As shown in [39, Theorem 2.4.17], given a positive, bounded, and Lipschitz continuous function $\omega : Q \to (0,\infty)$ with $\nabla\omega \in BV(Q;\mathbb{R}^2)$, the solution of (1.7) with $\omega_{\mathscr{L}}$ replaced by ω may exhibit jumps inherited from the weight ω that are not present in the data u_η. Because $\omega_{\mathscr{L}}$ in Level 2 is constructed using the local optimal parameters given by Level 3, we heuristically expect that these extra jumps are, in some sense, negligible. However, this possible issue has led us to consider two alternative adaptive space-dependent learning schemes.

First, we consider a learning scheme based on $(\mathscr{LS})_{TV_\omega}$ in (1.4) with $\omega_{\mathscr{L}}$ replaced by a smooth regularization $(\omega_\varepsilon)_{\mathscr{L}}$ (see the regularized weighted TV learning scheme $(\mathscr{LS})_{TV_{\omega_\varepsilon}}$ in (1.12) below). Second, using the fact that the minimizer in (1.6) coincides with

$$\operatorname{argmin}\left\{\frac{1}{\alpha}\int_L |u_\eta - u|^2 \,\mathrm{d}x + TV(u,L)\colon\ u \in BV(L)\right\},$$

we consider the weighted-fidelity learning scheme $(\mathscr{LS})_{TV-Fid_\omega}$ in (1.16) below, where the weight appears in the fidelity term. We begin by describing the regularized scenario.

$(\mathscr{LS})_{TV_{\omega_\varepsilon}}$ **Regularized weighted-TV learning scheme** (1.12)

Level 3. (optimal local training parameter) Fix $\mathscr{L} \in \mathscr{P}$; for each $L \in \mathscr{L}$, find

$$\alpha_L = \inf\left\{\operatorname{arginf}\left\{\int_L |u_c - u_\alpha|^2 \,\mathrm{d}x\colon\ \alpha \in \mathbb{R}^+\right\}\right\}, \tag{1.13}$$

where, for $\alpha \in \mathbb{R}^+$,

$$u_\alpha := \operatorname{argmin}\left\{\int_L |u_\eta - u|^2 \,\mathrm{d}x + \alpha TV(u,L)\colon\ u \in BV(L)\right\}.$$

Level 2. (space-dependent image denoising) For each $\mathscr{L} \in \mathscr{P}$ and for $\varepsilon > 0$ fixed, find

$$u^\varepsilon_{\mathscr{L}} := \operatorname{argmin}\left\{\int_Q |u_\eta - u|^2 \,\mathrm{d}x + TV_{\omega^\varepsilon_{\mathscr{L}}}(u,Q)\colon\ u \in BV_{\omega^\varepsilon_{\mathscr{L}}}(Q)\right\},$$

where we consider a regularized weight $\omega^\varepsilon_{\mathscr{L}} : Q \to [0,\infty)$ of $\omega_{\mathscr{L}}$ in (1.8) such that

$$\omega^\varepsilon_{\mathscr{L}} \in C^1(Q) \quad \text{and} \quad \omega^\varepsilon_{\mathscr{L}} \nearrow \omega_{\mathscr{L}} \text{ as } \varepsilon \to 0^+ \text{ and a.e. in } Q. \tag{1.14}$$

Level 1. (optimal partition and image restoration) Find

$$u^*_\varepsilon \in \operatorname{argmin}\left\{\int_Q |u_c - u^\varepsilon_{\mathscr{L}}|^2 \,\mathrm{d}x : \mathscr{L} \in \mathscr{P}\right\} \quad \text{with } u^\varepsilon_{\mathscr{L}} \text{ given by Level 2.}$$

For each $\varepsilon > 0$ fixed, similar results to those regarding the learning scheme $(\mathscr{LS})_{TV_\omega}$ in (1.4) hold for the learning scheme $(\mathscr{LS})_{TV_{\omega_\varepsilon}}$ in (1.12). A natural question is whether the sequence of optimal solutions of the latter, $\{u^*_\varepsilon\}_\varepsilon$, converge in some sense to the optimal solution of the former, u^*, as $\varepsilon \to 0^+$. This turns out to be a challenging mathematical question (see Open Problem 4.3), which we partially address in the following proposition.

Proposition 1.6 (On the energies in $(\mathscr{LS})_{TV_{\omega_\varepsilon}}$ as $\varepsilon \to 0^+$). *Under the setup of the learning schemes $(\mathscr{LS})_{TV_\omega}$ and $(\mathscr{LS})_{TV_{\omega_\varepsilon}}$ above, fix $\mathscr{L} \in \mathscr{P}$ and let $E_{\mathscr{L}} : L^1(Q) \to [0,\infty]$ and $E^\varepsilon_{\mathscr{L}} : L^1(Q) \to [0,\infty]$ be the functionals defined for $u \in L^1(Q)$ by*

$$E_{\mathscr{L}}[u] := \begin{cases} \displaystyle\int_Q |u_\eta - u|^2 \,\mathrm{d}x + TV_{\omega_{\mathscr{L}}}(u,Q) & \text{if } u \in BV_{\omega_{\mathscr{L}}}(Q), \\ +\infty & \text{otherwise,} \end{cases}$$

$$E^\varepsilon_{\mathscr{L}}[u] := \begin{cases} \displaystyle\int_Q |u_\eta - u|^2 \,\mathrm{d}x + TV_{\omega^\varepsilon_{\mathscr{L}}}(u,Q) & \text{if } u \in BV_{\omega^\varepsilon_{\mathscr{L}}}(Q), \\ +\infty & \text{otherwise.} \end{cases}$$

If (1.14) holds, then

$$\Gamma(L^1(Q)) - \limsup_{\varepsilon\to 0^+} E^\varepsilon_{\mathscr{L}} \leqslant E_{\mathscr{L}}. \tag{1.15}$$

The proof of this proposition and an analytical discussion of the learning scheme $(\mathcal{LS})_{TV_{\omega_\varepsilon}}$ in (1.12) can be found in Section 4, while the corresponding numerical scheme is detailed in Section 6.

Next, we study the weighted-fidelity learning scheme $(\mathcal{LS})_{TV-Fid_\omega}$ motivated above.

$(\mathcal{LS})_{TV-Fid_\omega}$ **Weighted-fidelity learning scheme** (1.16)

Level 3. (optimal local training parameter) Fix $\mathcal{L} \in \mathcal{P}$; for each $L \in \mathcal{L}$, find

$$\alpha_L = \inf \left\{ \operatorname{arginf} \left\{ \int_L |u_c - u_\alpha|^2 \, dx \colon \alpha \in \mathbb{R}^+ \right\} \right\}, \tag{1.17}$$

where, for $\alpha \in \mathbb{R}^+$,

$$u_\alpha := \operatorname{argmin} \left\{ \int_L \frac{1}{\alpha} |u_\eta - u|^2 \, dx + TV(u, L) \colon u \in BV(L) \right\}. \tag{1.18}$$

Level 2. (space-dependent image denoising) For each $\mathcal{L} \in \mathcal{P}$, find

$$u_{\mathcal{L}} := \operatorname{argmin} \left\{ \int_Q \frac{1}{\omega_{\mathcal{L}}} |u_\eta - u|^2 \, dx + TV(u, Q) \colon u \in BV_{\omega_{\mathcal{L}}}(Q) \right\},$$

where, similarly to (1.8), $\omega_{\mathcal{L}}$ is defined by

$$\omega_{\mathcal{L}}(x) := \sum_{L \in \mathcal{L}} \alpha_L \chi_L(x) \quad \text{with } \alpha_L \text{ given by Level 3.}$$

Level 1. (optimal partition and image restoration) Find

$$u^* \in \operatorname{argmin} \left\{ \int_Q |u_c - u_{\mathcal{L}}|^2 \, dx : \mathcal{L} \in \mathcal{P} \right\} \quad \text{with } u_{\mathcal{L}} \text{ given by Level 2.}$$

Once more, similar results to those regarding the learning scheme $(\mathcal{LS})_{TV_\omega}$ in (1.4) hold for the learning scheme $(\mathcal{LS})_{TV-Fid_\omega}$ in (1.16). In particular, the box constraint here is essential to guarantee that **Level 2** of the scheme is well posed. This analysis is undertaken in Section 4, while the corresponding numerical study is addressed in Section 6.

The last theoretical result of this paper concerns replacing the TV term in our space-dependent bilevel learning schemes with a higher-order regularizer. A well-known drawback of the ROF model is the possible occurrence of staircasing effects whenever two neighboring areas of an image are both smoothed out and an abrupt spurious discontinuity is produced in the denoising process. A solution to counteract this effect consists in resorting to higher-order derivatives in the regularizer (see, e.g., [11, 19, 25, 47, 8]). We consider here the total generalized variation (TGV) model introduced in [9], which is considered to be one of the most effective image-reconstruction models among those involving mixed first- and higher-order terms, cf. [47, 10, 50, 38] for some theoretical results about its solutions.

For a function $u \in BV(Q)$ and $\alpha = (\alpha_0, \alpha_1) \in \mathbb{R}^+ \times \mathbb{R}^+$, the second-order TGV functional is given by

$$TGV_{\alpha_0,\alpha_1}(u) := \inf \left\{ \alpha_0 |Du - v|(Q) + \alpha_1 |\mathcal{E}v|(Q) : v \in BD(Q) \right\}, \tag{1.19}$$

where, as before, Du denotes the distributional gradient of u, $|\mu|(Q)$ is the total variation on Q of a Radon measure μ, $\mathcal{E}$ is the symmetric part of the distributional gradient, and BD indicates the space of vector-valued functions with bounded deformation, cf. [53]. In this setting, our learning scheme reads as follows.

$(\mathcal{LS})_{TGV_\omega}$ **Weighted-TGV learning scheme** (1.20)

Level 3. (optimal local regularization parameter) Fix $\mathcal{L} \in \mathcal{P}$; for each $L \in \mathcal{L}$, find

$$\alpha_L = \big((\alpha_L)_0, (\alpha_L)_1\big) := \inf \left\{ \operatorname{arginf} \left\{ \int_L |u_c - u_\alpha|^2 \, dx \colon \alpha = (\alpha_0, \alpha_1) \in \mathbb{R}^+ \times \mathbb{R}^+ \right\} \right\}, \tag{1.21}$$

where, for $\alpha = (\alpha_0, \alpha_1) \in \mathbb{R}^+ \times \mathbb{R}^+$,

$$u_\alpha := \operatorname{argmin} \left\{ \int_L |u_\eta - u|^2 \, dx + TGV_{\alpha_0,\alpha_1}(u, L) \colon u \in BV(L) \right\}, \tag{1.22}$$

and where the infimum in (1.21) is meant with respect to the lexicographic order in $\mathbb{R}^2$.

Level 2. (space-dependent TGV image denoising) For each $\mathscr{L} \in \mathscr{P}$, find

$$u_{\mathscr{L}} := \operatorname{argmin}\left\{ \int_Q |u_\eta - u|^2 \,\mathrm{d}x + TGV_{\omega^0_{\mathscr{L}}, \omega^1_{\mathscr{L}}}(u, Q) \colon\ u \in BV_{\omega^0_{\mathscr{L}}}(Q) \right\}, \tag{1.23}$$

where, for $i \in \{0, 1\}$, the weight $\omega^i_{\mathscr{L}}$ is defined by

$$\omega^i_{\mathscr{L}}(x) := \sum_{L \in \mathscr{L}} (\alpha_L)_i \, \chi_L(x) \quad \text{with } \alpha_L \text{ given by Level 3.}$$

In the expression above,

$$TGV_{\omega^0_{\mathscr{L}}, \omega^1_{\mathscr{L}}}(u, Q) := \inf_{v \in BD_{\omega^1_{\mathscr{L}}}(Q)} \left\{ \mathscr{V}_{\omega^0_{\mathscr{L}}}(Du - v, Q) + \mathscr{V}_{\omega^1_{\mathscr{L}}}(\mathcal{E}v, Q) \right\}, \tag{1.24}$$

where the quantities $\mathscr{V}_{\omega^0_{\mathscr{L}}}$ and $\mathscr{V}_{\omega^1_{\mathscr{L}}}$ are weighted counterparts to the classical total variation of Radon measures. We refer to Sections 2 and 5 for the precise definition and properties of these quantities. In particular, we will prove that

$$\mathscr{V}_{\omega^0_{\mathscr{L}}}(Du - v, Q) = \int_Q (\omega^0_{\mathscr{L}})^{\mathrm{sc}^-} \,\mathrm{d}|Du - v|, \tag{1.25}$$

and

$$\mathscr{V}_{\omega^1_{\mathscr{L}}}(\mathcal{E}v, Q) = \int_Q (\omega^1_{\mathscr{L}})^{\mathrm{sc}^-} \,\mathrm{d}|\mathcal{E}v|, \tag{1.26}$$

where $BV_{\omega^0_{\mathscr{L}}}$ is the space of $\omega^0_{\mathscr{L}}$-weighted BV-functions (see Subsection 3.2) and $BD_{\omega^1_{\mathscr{L}}}$ is the space of $\omega^1_{\mathscr{L}}$-weighted BD-functions (see Section 5).

Level 1. (optimal partition and image restoration) Find

$$u^* \in \operatorname{argmin}\left\{ \int_Q |u_c - u_{\mathscr{L}}|^2 \,\mathrm{d}x : \mathscr{L} \in \mathscr{P} \right\} \quad \text{with } u_{\mathscr{L}} \text{ given by Level 2.}$$

Analogously to $(\mathscr{L}\mathcal{S})_{TV-Fid_\omega}$, we can also consider a weighted-fidelity TGV scheme, which we use in our numerical results and describe next.

$(\mathscr{L}\mathcal{S})_{TGV-Fid_\omega}$ **TGV weighted-fidelity learning scheme** (1.27)

With $\alpha_0, \alpha_1 \in \mathbb{R}^+$ fixed throughout:

Level 3. (optimal local training parameter) Fix $\mathscr{L} \in \mathscr{P}$; for each $L \in \mathscr{L}$, find

$$\lambda_L = \inf\left\{ \operatorname{arginf}\left\{ \int_L |u_c - u_\lambda|^2 \,\mathrm{d}x \colon\ \lambda \in \mathbb{R}^+ \right\} \right\}, \tag{1.28}$$

where, for $\lambda \in \mathbb{R}^+$,

$$u_\lambda := \operatorname{argmin}\left\{ \lambda \int_L |u_\eta - u|^2 \,\mathrm{d}x + TGV_{\alpha_0, \alpha_1}(u, L) \colon\ u \in BV(L) \right\}.$$

Level 2. (space-dependent image denoising) For each $\mathscr{L} \in \mathscr{P}$, find

$$u_{\mathscr{L}} := \operatorname{argmin}\left\{ \int_Q \omega_{\mathscr{L}} |u_\eta - u|^2 \,\mathrm{d}x + TGV_{\alpha_0, \alpha_1}(u, Q) \colon\ u \in BV_{\omega_{\mathscr{L}}}(Q) \right\},$$

where $\omega_{\mathscr{L}}$ is defined by

$$\omega_{\mathscr{L}}(x) := \sum_{L \in \mathscr{L}} \lambda_L \chi_L(x) \quad \text{with } \lambda_L \text{ given by Level 3.}$$

Level 1. (optimal partition and image restoration) Find

$$u^* \in \operatorname{argmin}\left\{ \int_Q |u_c - u_{\mathscr{L}}|^2 \,\mathrm{d}x : \mathscr{L} \in \mathscr{P} \right\} \quad \text{with } u_{\mathscr{L}} \text{ given by Level 2.}$$

As in the case of our learning schemes for the weighted total variation, the analysis of $(\mathscr{L}\mathcal{S})_{TGV_\omega}$ and $(\mathscr{L}\mathcal{S})_{TGV-\mathrm{Fid}_\omega}$ is performed under a box constraint assumption, which for the first case reads as

$$\alpha = (\alpha_0, \alpha_1) \in \left[c_0, \frac{1}{c_0}\right] \times \left[c_1, \frac{1}{c_1}\right]. \tag{1.29}$$

Our main result for the weighted-TGV scheme is the following.

Theorem 1.7. *There exists an optimal solution u^* to the learning scheme $(\mathcal{LS})_{TGV_\omega}$ in (1.20) with the minimization in (1.21) restricted by (1.29).*

Analogously, we infer the ensuing theorem for the TGV with weighted fidelity.

Theorem 1.8 (Existence of solutions to $(\mathcal{LS})_{TGV-Fid_\omega}$). *There exists an optimal solution u^* to the learning scheme $(\mathcal{LS})_{TGV-Fid_\omega}$ in (1.27) with the minimization in (1.28) restricted by the box constraint $\lambda \in \left[c, \frac{1}{c}\right]$, for some $c > 0$.*

Also in the case of weighted-TGV learning schemes, we provide a connection between stopping criteria and existence of a box constraint. To be precise, we show that if (1.29) is imposed, then a stopping criterion can be naturally imposed on the schemes. Concerning the converse implication, we show that if a suitable stopping criterion is enforced, then $(\alpha_L)_0$ and $(\alpha_L)_1$ are both always bounded from below by a positive constant, and that they cannot simultaneously blow up to infinity. The weaker nature of this latter implication is due to one main reason: the upper bound established on the optimal parameters for the weighted TV scheme is hinged upon a suitable Poincaré inequality for the total variation functional, cf. Proposition 3.5; in the TGV case, the analogous argument only provides a bound from above for the minimum between $(\alpha_L)_0$ and $(\alpha_L)_1$, and thus does not allow to conclude the existence of a uniform upper bound on either component, cf. Proposition 5.11. We refer to Subsection 5.3 for a discussion of this issue and for the details of this argument. For completeness, we mention that a result related to Proposition 5.11 has been proven in [48, Proposition 6]. In Proposition 5.11, we make this study quantitative and keep track of the dependence on the cell size through the Poincaré constant.

The paper is organized as follows: in Section 2, we collect some notation which will be employed throughout the paper. The focus of Sections 3 and 4 is on our weighted-TV scheme, as well as on the two variants thereof, including a regularization of the weight and a weighted fidelity, respectively. Section 5 is devoted to the study of our weighted-TGV learning scheme and of the corresponding TGV scheme with weighted fidelity. Section 6 contains some numerical results for the various learning schemes presented in the paper and a comparison of their performances.

2. Glossary

Here we collect some notation that will be used throughout the paper, and introduce some energy functionals that will be studied.

In what follows, $A \subset \mathbb{R}^n$ is an open and bounded set and $\mathbb{X}$ stands for either $\mathbb{R}$, $\mathbb{R}^n$, or $\mathbb{R}^{n\times n}_{sym}$, where the latter is the space of all $n \times n$ symmetric matrices and $n \in \mathbb{N}$. We denote by $\mathcal{M}(A;\mathbb{X})$ the space of all finite Radon measures in A with values on $\mathbb{X}$, and by $|\mu| \in \mathcal{M}(A;\mathbb{R}^+_0)$ the total variation of $\mu \in \mathcal{M}(A;\mathbb{X})$, which is defined for each measurable set $B \subset A$ by

$$|\mu|(B) := \sup\left\{ \sum_{i=1}^{\infty} |\mu(B_i)| : \ \{B_i\}_{i\in\mathbb{N}} \text{ is a partition of } B \right\}.$$

Using the Riesz representation theorem, $\mathcal{M}(A;\mathbb{X})$ can be identified with the dual of $C_0(A;\mathbb{X}')$, the closure with respect to the supremum norm of the set of all continuous functions on A with compact support. In particular, the total variation of a Radon measure $\mu \in \mathcal{M}(A;\mathbb{X})$ is alternatively given by

$$|\mu|(B) = \sup\left\{ \int_B \varphi(x)\cdot \mathrm{d}\mu(x) : \ \varphi \in C_0(B;\mathbb{X}'),\ \|\varphi\|_{L^\infty(B;\mathbb{X}')} \leqslant 1 \right\},\quad B \subset A \text{ measurable}, \tag{2.1}$$

where $\cdot$ represent the duality product between an element of $\mathbb{X}'$ and an element of $\mathbb{X}$. With the trivial identification of column vectors with row vectors, we will often write $\mathbb{X}$ in place of $\mathbb{X}'$.

In the case in which $\mu = Du \in \mathcal{M}(A;\mathbb{R}^n)$ for some $u \in BV(A)$, a density argument show that (2.1) is equivalent to

$$|Du|(B) = \sup\left\{ \int_B u(x) \operatorname{div}\varphi(x)\,\mathrm{d}x : \ \varphi \in \mathrm{Lip}_c(B;\mathbb{R}^n),\ \|\varphi\|_{L^\infty(B;\mathbb{R}^n)} \leqslant 1 \right\}, \tag{2.2}$$

and we often write $TV(u,B)$ in place of $|Du|(B)$. In the preceding expression, and throughout this manuscript, $\mathrm{Lip}_c(B;\mathbb{X})$ represents the space of all $\mathbb{X}$-valued Lipschitz functions with compact support in B.

Similarly, in the case in which $\mu = \mathcal{E}v \in \mathcal{M}(A;\mathbb{R}^{n\times n}_{sym})$ for some $v \in BD(A)$ and $\mathcal{E}$ the symmetrical part of the distributional derivative, then (2.1) is equivalent to

$$|\mathcal{E}v|(B) = \sup\left\{ \int_B v(x)\cdot \operatorname{div}\varphi(x)\,\mathrm{d}x :\ \varphi \in \operatorname{Lip}_c(B;\mathbb{R}^{n\times n}_{sym}),\ \|\varphi\|_{L^\infty(B;\mathbb{R}^{n\times n}_{sym})} \leqslant 1 \right\}, \tag{2.3}$$

where $(\operatorname{div}\varphi)_j = \sum_{k=1}^n \frac{\partial \varphi_{jk}}{\partial x_k}$ for each $j \in \{1,\dots,n\}$.

At the core of the present manuscript are weighted versions of the spaces of bounded variation and of bounded deformation. These weighted versions rely on a generalization of (2.2) and (2.3) that cannot be derived directly from the Riesz representation theorem, and thus need a careful analysis to prove the variational identities stated in (1.9) and (1.25)–(1.26), addressed in Sections 3 and 5, respectively.

Given a Radon measure $\mu \in \mathcal{M}(A;\mathbb{X})$ and a locally integrable function $\omega : A \to [0,\infty)$, we define the ω-weighted variation of μ on A, written $\mathscr{V}_\omega(\mu, A)$, by

$$\mathscr{V}_\omega(\mu, A) := \sup\left\{ \int_A \varphi(x)\cdot \mathrm{d}\mu(x) :\ \varphi \in \operatorname{Lip}_c(A;\mathbb{X}'),\ |\varphi| \leqslant \omega \right\}. \tag{2.4}$$

As before, if $\mu = Du \in \mathcal{M}(A;\mathbb{R}^n)$ for some $u \in BV(A)$, then (2.4) is equivalent to

$$\mathscr{V}_\omega(Du, A) = \sup\left\{ \int_A u(x)\ \operatorname{div}\varphi(x)\,\mathrm{d}x :\ \varphi \in \operatorname{Lip}_c(A;\mathbb{R}^n),\ |\varphi| \leqslant \omega \right\},$$

which we often represent by $TV_\omega(u, A)$. Also, if $\mu = Du - v := Du - v\mathcal{L}^n\lfloor A \in \mathcal{M}(A;\mathbb{R}^n)$ for some $u \in BV(A)$ and $v \in L^1(A;\mathbb{R}^n)$, then (2.4) is equivalent to

$$\mathscr{V}_\omega(Du - v, A) = \sup\left\{ \int_A \big(u(x)\ \operatorname{div}\varphi(x) + v(x)\cdot\varphi(x)\big)\,\mathrm{d}x :\ \varphi \in \operatorname{Lip}_c(A;\mathbb{R}^n),\ |\varphi| \leqslant \omega \right\}.$$

Moreover, if $\mu = \mathcal{E}v \in \mathcal{M}(A;\mathbb{R}^{n\times n}_{sym})$ for some $v \in BD(A)$, then (2.4) is equivalent to

$$\mathscr{V}_\omega(\mathcal{E}v, A) = \sup\left\{ \int_A v(x)\cdot \operatorname{div}\varphi(x)\,\mathrm{d}x :\ \varphi \in \operatorname{Lip}_c(A;\mathbb{R}^{n\times n}_{sym}),\ |\varphi| \leqslant \omega \right\}.$$

The energy functional associated with the analogue to the ROF's model, where we use a weighted-TV regularizer on $\Omega \subset \mathbb{R}^2$ instead of the total variation (TV), is denoted by (see Theorem 3.2)

$$E[u] := \int_\Omega |u_\eta - u|^2\,\mathrm{d}x + TV_\omega(u, \Omega).$$

To highlight the dependence on a partition $\mathscr{L}$ of Q made of dyadic cubes, the extension of the preceding functional (for a weight $\omega_{\mathscr{L}}$ and $\Omega = Q$) to $L^1(Q)$ is represented by

$$E_{\mathscr{L}}[u] := \begin{cases} \displaystyle\int_Q |u_\eta - u|^2\,\mathrm{d}x + TV_{\omega_{\mathscr{L}}}(u, Q) & \text{if } u \in BV_{\omega_{\mathscr{L}}}(Q), \\ +\infty & \text{otherwise.} \end{cases}$$

Moreover, for the ε-dependent regularized weight $\omega^\varepsilon_{\mathscr{L}}$, introduced in (1.14), the energy above is written as

$$E^\varepsilon_{\mathscr{L}}[u] := \begin{cases} \displaystyle\int_Q |u_\eta - u|^2\,\mathrm{d}x + TV_{\omega^\varepsilon_{\mathscr{L}}}(u, Q) & \text{if } u \in BV_{\omega^\varepsilon_{\mathscr{L}}}(Q), \\ +\infty & \text{otherwise.} \end{cases}$$

The two preceding functionals are introduced in Proposition 1.6, where we address the relationship between the weighted-TV and the regularized weighted-TV learning schemes in (1.4) and (1.12), respectively.

For a fixed image domain Ω, the optimal tuning parameter α in Level 3 of any of the TV learning schemes addressed here is found by minimizing the cost function $I : (0,\infty) \to \mathbb{R}$ defined by

$$I(\alpha) := \int_\Omega |u_c - u_\alpha|^2\,\mathrm{d}x \text{ for } \alpha \in (0, +\infty), \tag{2.5}$$

where u_c is the clean image and u_α is the reconstructed image obtained as the minimizer of the denoising model in aforementioned Level 3. The lower-semicontinuous envelope of I on $[0,\infty]$, $I^{sc^-} : [0,+\infty] \to [0,+\infty]$, is defined for $\bar\alpha \in [0,+\infty]$ by

$$I^{sc^-}(\bar\alpha) := \inf\left\{ \liminf_{j\to\infty} I(\alpha_j) \colon\ (\alpha_j)_{j\in\mathbb{N}} \subset (0,+\infty),\ \alpha_j \to \bar\alpha \text{ in } [0,+\infty] \right\}.$$

The study of existence of minimizers for I and the characterization of I^{sc^-} for the weighted-TV learning scheme in (1.4) is addressed in Theorem 3.8 and Lemma 3.11, respectively. This study relies on the Γ-convergence of the family, parametrized by $\alpha \in (0,\infty)$, of energy functionals associated with ROF's model,

$$F_\alpha[u] := \begin{cases} \displaystyle\int_\Omega |u_\eta - u|^2 \,\mathrm{d}x + \alpha TV(u,\Omega) & \text{if } u \in BV(\Omega),\\ +\infty & \text{otherwise.}\end{cases}$$

In turn, this Γ-convergence analysis naturally involves the extreme points $\bar\alpha = 0$ and $\bar\alpha = +\infty$, which are associated with the energies

$$F_0[u] := \begin{cases} \displaystyle\int_\Omega |u_\eta - u|^2 \,\mathrm{d}x & \text{if } u \in L^2(\Omega),\\ +\infty & \text{otherwise,}\end{cases} \quad \text{and} \quad F_\infty[u] := \begin{cases} \displaystyle\int_\Omega |u_\eta - c|^2 \,\mathrm{d}x & \text{if } u \equiv c \in \mathbb{R},\\ +\infty & \text{otherwise,}\end{cases}$$

respectively (see Lemma 3.9 for the weighted-TV learning scheme in (1.4)).

Regarding the TGV case, to obtain the existence of optimal parameters for Level 3 of the schemes (1.20) and (1.27), stated in Theorem 5.13, we are led to study Γ-convergence of the family of functionals, parametrized by $\alpha = (\alpha_0, \alpha_1) \in (0,+\infty)^2$, defined as

$$G_\alpha[u] := \begin{cases} \int_\Omega |u_\eta - u|^2 \,\mathrm{d}x + TGV_{\alpha_0,\alpha_1}(u,\Omega) & \text{if } u \in BV(\Omega),\\ +\infty & \text{otherwise.}\end{cases}$$

In this case, the Γ-convergence result is more involved because it includes different combinations of $\bar\alpha_i = 0$, $\bar\alpha_i \in \mathbb{R}^+$, or $\bar\alpha_i = +\infty$ for $i = 0$ and $i = 1$. The expressions for the ensuing limits can be found in the statement of Lemma 5.15.

The characterization of the lower semicontinuous envelope of the TGV analog of (2.5), denoted by $J(\alpha)$ for $\alpha = (\alpha_0, \alpha_1)$, is contained in Lemma 5.18.

In the sequel, we use both the average of a function $u : \Omega \to \mathbb{R}$ on a subdomain $L \subset \Omega$,

$$[u]_L := \frac{1}{|L|}\int_L u(x)\,\mathrm{d}x,$$

and its projection onto affine functions $\langle u\rangle_L$, which is the unique solution to the minimum problem

$$\min\left\{\int_L |u-v|^2 \,\mathrm{d}x : \, v \text{ is affine in } L\right\},$$

where in both cases the subscript may be omitted when $L = \Omega$.

3. Analysis of the Weighted-TV learning scheme $(\mathscr{LS})_{TV_\omega}$

Here, we prove existence of solutions to the weighted-TV learning scheme, $(\mathscr{LS})_{TV_\omega}$, introduced in (1.4). We analyze each level in the three subsequent subsections. In particular, we prove Theorem 1.5 in Subsection 3.3. Then, in Subsection 3.4, we prove Theorem 1.4 and we provide different examples of stopping criteria for the refinement of the admissible partitions introduced in Definition 1.2.

3.1. On Level 3.

In this section, we discuss the main features of Level 3, and variants thereof, of the learning scheme $(\mathscr{LS})_{TV_\omega}$ in (1.4).

3.1.1. As we mentioned in Remark 1.1, the parameter α_L in (1.5) is uniquely determined by definition, with $\alpha_L \in [0,+\infty]$.

3.1.2. In view of Theorem 3.8 (see Subsection 3.4), if $L \in \mathscr{L}$ is such that

$$TV(u_c, L) < TV(u_\eta, L) \quad \text{and} \quad \|u_\eta - u_c\|^2_{L^2(L)} < \|[u_\eta]_L - u_c\|^2_{L^2(L)}, \tag{3.1}$$

then

$$\operatorname{arginf}\left\{\int_L |u_c - u_\alpha|^2 \,\mathrm{d}x \colon \alpha \in \mathbb{R}^+\right\} = \operatorname{argmin}\left\{\int_L |u_c - u_\alpha|^2 \,\mathrm{d}x \colon \alpha \in \big[c_L, C_Q\|u_\eta\|_{L^2(L)}\big]\right\},$$

where c_L and C_Q are positive constants, with c_Q depending only on Q. In particular, we have that $\alpha_L \in \big[c_L, C_Q\|u_\eta\|_{L^2(L)}\big]$. Furthermore, because each partition $\mathscr{L} \in \mathscr{P}$ is finite, it follows that if (3.1) holds for all $L \in \mathscr{L}$, then

$$\alpha_L \in K_{\mathscr{L}} := \Big[\min_{L\in\mathscr{L}} c_L, C_Q \max_{L\in\mathscr{L}} \|u_\eta\|_{L^2(L)}\Big] \subset (0,+\infty)$$

for every $L \in \mathscr{L}$, which yields a natural box constraint for a fixed partition. Note, however, that the box constraint given by the compact set $K_{\mathscr{L}}$ may vary according to the choice of the partition $\mathscr{L}$.

3.1.3. If we consider Level 3 with (1.5) replaced by (1.11), then the minimum

$$\min_{\alpha \in [c_0, \frac{1}{c_0}]} \int_L |u_c - u_\alpha|^2 \, dx$$

exists as the minimum of a lower semicontinuous function (see Lemma 3.11 in Subsection 3.4) on a compact set. In particular, $\bar{\alpha}_L$ is uniquely determined, with

$$\bar{\alpha}_L \in \left[c_0, \frac{1}{c_0}\right] \text{ for all } L \in \mathscr{L} \text{ and } \mathscr{L} \in \mathscr{P}.$$

3.2. **On Level 2.** Here, we discuss existence and uniqueness of solutions to the minimization problem in (1.7). A key step in this discussion is the study of the space $BV_\omega(\Omega)$ of ω-weighted BV-functions in an open set $\Omega \subset \mathbb{R}^n$, where the weight $\omega : \Omega \to [0, \infty)$ is assumed to be a locally integrable function. We adopt the approach introduced in [3], and further analyzed in [14, 13].

Given a ω-weighted locally integrable function in Ω, $u \in L^1_{\omega,\mathrm{loc}}(\Omega)$, where

$$L^1_{\omega,\mathrm{loc}}(\Omega) := \left\{ v : \Omega \to \mathbb{R} \text{ measurable} : \int_K |v(x)|\, \omega(x)\, dx < \infty \text{ for every compact set } K \subset \Omega \right\}, \tag{3.2}$$

we define its ω-weighted total variation in Ω, $TV_\omega(u, \Omega)$, by

$$TV_\omega(u, \Omega) := \sup \left\{ \int_\Omega u \operatorname{div} \varphi \, dx \colon \varphi \in \mathrm{Lip}_c(\Omega; \mathbb{R}^2),\ |\varphi| \leqslant \omega \right\} \tag{3.3}$$

(see also Section 2). Accordingly, we define the space $BV_\omega(\Omega)$ of ω-weighted BV-functions in Ω by

$$BV_\omega(\Omega) := \left\{ u \in L^1_\omega(\Omega) \colon TV_\omega(u, \Omega) < \infty \right\},$$

endowed with the norm

$$\|u\|_{BV_\omega(\Omega)} := \|u\|_{L^1_\omega(\Omega)} + TV_\omega(u, \Omega), \quad \text{where } \|u\|_{L^1_\omega(\Omega)} := \int_\Omega |u(x)|\, \omega(x)\, dx.$$

Clearly, if $\omega \equiv 1$, then we recover the usual space BV of functions of bounded variation. Next, we collect some properties of $BV_\omega(\Omega)$, proved in [3, 14, 13], that will be used in our analysis.

Theorem 3.1. *Let $\Omega \subset \mathbb{R}^n$ be an open set and let $\omega : \Omega \to [0, \infty)$ be a locally integrable function. Then, the following hold:*

(i) The map $u \mapsto TV_\omega(u, \Omega)$ is lower-semicontinuous with respect to the (strong) convergence in $L^1_{\omega,\mathrm{loc}}(\Omega)$.

(ii) Given $u \in L^1_{\omega,\mathrm{loc}}(\Omega)$, we have that $TV_\omega(u, \Omega) = TV_{\omega^{sc^-}}(u, \Omega)$, where ω^{sc^-} denotes the lower-semicontinuous envelope of ω.

(iii) Assume that ω is lower-semicontinuous and strictly positive. Then, we have that $u \in L^1_{\mathrm{loc}}(\Omega)$ and $TV_\omega(u, \Omega) < \infty$ if and only if $u \in BV_{\mathrm{loc}}(\Omega)$ and $\omega \in L^1(\Omega; |Du|)$. If any of these two equivalent conditions hold, then we have

$$TV_\omega(u, B) = \int_B \omega(x)\, d|Du|(x)$$

for every Borel set $B \subset \Omega$.

Proof. The proof of (i)–(iii) may be found in [3] under the additional assumption that ω satisfies a Muckenhoupt A_1 condition (see [3] for the details). Without assuming this extra assumption on ω, the proof of (i) may be found in [13, Proposition 1.3.1 and Remark 1.3.2]; the proof of (ii) follows from [13, Proposition 2.1.1 and Theorem 2.1.2]; finally, (iii) is shown in [13, Theorem 2.1.5]. □

The existence and uniqueness of solutions of Level 2 of the learning scheme $(\mathscr{LS})_{TV_\omega}$ in (1.4) with (1.5) replaced by (1.11) are hinged on the following theorem.

Theorem 3.2. *Let $v \in L^2(\Omega)$ and let $\omega : \Omega \to (0,\infty)$ be such that $0 < \inf_\Omega \omega \leqslant \sup_\Omega \omega < \infty$. Then, there exists a unique $\bar u \in BV_\omega(\Omega)$ satisfying*

$$\int_\Omega |v-\bar u|^2 \,\mathrm{d}x + TV_\omega(\bar u,\Omega) = \min_{u\in BV_\omega(\Omega)} \left\{ \int_\Omega |v-u|^2\,\mathrm{d}x + TV_\omega(u,\Omega)\right\}.$$

Moreover, denoting by ω^{sc^-} the lower-semicontinuous envelope of ω, we have $\bar u \in BV_\omega(\Omega)\cap BV(\Omega)\cap BV_{\omega^{sc^-}}(\Omega)$ and

$$TV_\omega(\bar u,\Omega) = \int_\Omega \omega^{sc^-}(x)\,\mathrm{d}|D\bar u|(x).$$

Proof. For $u \in BV_\omega(\Omega)$, set

$$E[u] := \int_\Omega |v-u|^2\,\mathrm{d}x + TV_\omega(u,\Omega),$$

and let

$$\mathscr{m} := \inf_{u\in BV_\omega(\Omega)} E[u].$$

Note that $0 \leqslant \mathscr{m} \leqslant E[0] = \|v\|^2_{L^2(\Omega)}$, and consider $(u_n)_{n\in\mathbb{N}} \subset BV_\omega(\Omega)$ such that

$$\mathscr{m} = \lim_{n\to\infty} E[u_n]. \tag{3.4}$$

By hypothesis, there exist c_1, $c_2 \in \mathbb{R}^+$ such that for a.e. $x\in\Omega$, we have

$$c_1 \leqslant \omega(x) \leqslant c_2. \tag{3.5}$$

Consequently, for all $x \in \Omega$,

$$c_1 \leqslant \omega^{sc^-}(x) \leqslant c_2. \tag{3.6}$$

Then, in view of (3.4) and Theorem 3.1 *(ii)*–*(iii)*, for all $n\in\mathbb{N}$ sufficiently large, we have

$$\begin{aligned}\mathscr{m}+1 &\geqslant \int_\Omega |v-u_n|^2\,\mathrm{d}x + TV_\omega(u_n,\Omega) = \int_\Omega |v-u_n|^2\,\mathrm{d}x + TV_{\omega^{sc^-}}(u_n,\Omega)\\ &= \int_\Omega |v-u_n|^2\,\mathrm{d}x + \int_\Omega \omega^{sc^-}(x)\,\mathrm{d}|Du_n|(x) \geqslant \int_\Omega |v-u_n|^2\,\mathrm{d}x + c_1|Du_n|(\Omega).\end{aligned}$$

Thus, extracting a subsequence if necessary (not relabeled), there exists $\bar u \in BV(\Omega)$ such that

$$u_n \overset{*}{\rightharpoonup} \bar u \text{ in } BV(\Omega), \quad u_n \rightharpoonup \bar u \text{ in } L^2(\Omega), \quad u_n \to \bar u \text{ in } L^1(\Omega).$$

Moreover, by (3.5)–(3.6) and Theorem 3.1, we have also $\bar u \in BV(\Omega)\cap BV_{\omega^{sc^-}}(\Omega)$, with

$$TV_\omega(\bar u,\Omega) = \int_\Omega \omega^{sc^-}(x)\,\mathrm{d}|D\bar u|(x),$$

and

$$\begin{aligned}\mathscr{m} \leqslant E[\bar u] &= \int_\Omega |v-\bar u|^2\,\mathrm{d}x + TV_\omega(\bar u,\Omega)\\ &\leqslant \liminf_{n\to\infty}\left(\int_\Omega |v-u_n|^2\,\mathrm{d}x + TV_{\omega^{sc^-}}(u_n,\Omega)\right) = \lim_{n\to\infty} E[u_n] = \mathscr{m}.\end{aligned}$$

Because $|\cdot|^2$ is strictly convex, $\bar u$ is the unique minimizer of $E[\cdot]$ over $BV_\omega(\Omega)$. ☐

Corollary 3.3. *There exists a unique solution $u_{\mathscr{L}} \in BV_{\omega_{\mathscr{L}}}(\Omega)\cap BV(\Omega)\cap BV_{\omega_{\mathscr{L}}^{sc^-}}(\Omega)$ to Level 2 of the learning scheme $(\mathcal{LS})_{TV_\omega}$ in (1.4) with (1.5) replaced by (1.11), where $\omega_{\mathscr{L}}^{sc^-}$ denotes the lower-semicontinuous envelope of $\omega_{\mathscr{L}}$. Moreover,*

$$\min\left\{\int_Q |u_\eta - u|^2\,\mathrm{d}x + TV_{\omega_{\mathscr{L}}}(u,Q)\colon\, u\in BV_{\omega_{\mathscr{L}}}(Q)\right\} = \int_Q |u_\eta - u_{\mathscr{L}}|^2\,\mathrm{d}x + \int_\Omega \omega_{\mathscr{L}}^{sc^-}(x)\,\mathrm{d}|Du_{\mathscr{L}}|(x).$$

Proof. Using the analysis in Subsection 3.1.3, the function $\omega_{\mathscr{L}}$ in (1.8) satisfies the bounds $c_0 \leqslant \omega_{\mathscr{L}} \leqslant \frac{1}{c_0}$ in Q, which, together with Theorem 3.2, concludes the proof. ☐

Remark 3.4. Recalling the analysis in Subsection 3.1.2, the previous corollary still holds if we assume that (3.1) holds for all $L \in \mathscr{L}$ instead of replacing (1.5) by (1.11).

3.3. **On Level 1.** Here, we prove that Level 1 of the learning scheme $(\mathcal{LS})_{TV_\omega}$ admits a solution provided we consider a stopping criterion as in Definition 1.2. We start by checking that the box constraint (1.10) yields such a stopping criterion, after which we establish the converse statement. We then explore alternative stopping criteria.

To prove that the box constraint (1.10) yields a stopping criterion for the refinement of the admissible partitions, we first recall the existence of a smallness condition on the tuning parameter under which the restored image given by the TV model is constant.

Proposition 3.5. *There exists a positive constant, C_Q, depending only on Q, such that for any dyadic cube $L \subset Q$ and for all $\alpha \geqslant C_Q \|u_\eta\|_{L^2(L)}$, the solution u_α of (1.6) is constant, with $u_\alpha \equiv [u_\eta]_L$.*

Proof. The proof is a simple consequence of [39, Proposition 2.5.7] combined with the scaling invariance of the constant in the 2-dimensional Poincaré–Wirtinger inequality in BV (see [1, Remark 3.50]). □

Theorem 3.6. *Consider the learning scheme $(\mathcal{LS})_{TV_\omega}$ in (1.4) with (1.5) replaced by (1.11). Then, there exist $\kappa \in \mathbb{N}$ and $\mathcal{L}_1, ..., \mathcal{L}_\kappa \in \mathcal{P}$ such that*

$$\operatorname{argmin}\left\{\int_Q |u_c - u_{\mathcal{L}}|^2 \,\mathrm{d}x : \mathcal{L} \in \mathcal{P}\right\} = \operatorname{argmin}\left\{\int_Q |u_c - u_{\mathcal{L}_i}|^2 \,\mathrm{d}x : i \in \{1, ..., \kappa\}\right\}. \tag{3.7}$$

Proof. We use Proposition 3.5 to prove that if a partition contains dyadic squares of side length smaller than a certain threshold, then it can be replaced by a partition of dyadic squares of side length greater than that threshold without changing the minimizer at Level 2.

Let $\bar\varepsilon \in (0,1)$ be such that for every measurable set $E \subset Q$ with $|E| \leqslant \bar\varepsilon$, we have

$$\|u_\eta\|_{L^2(E)} \leqslant \frac{c_0}{C_Q}, \tag{3.8}$$

where c_0 is the constant in (1.11) and C_Q is the constant given by Proposition 3.5. Set

$$\bar k := \min\left\{k \in \mathbb{N}\colon \frac{1}{4^k} \leqslant \bar\varepsilon\right\} \quad \text{and} \quad \bar{\mathcal{P}} := \left\{\mathcal{L} \in \mathcal{P}\colon |L| \geqslant \frac{1}{4^{\bar k}} \text{ for all } L \in \mathcal{L}\right\}.$$

Note that $\bar{\mathcal{P}}$ has finite cardinality. Finally, define

$$\mathcal{P}^* := \mathcal{P} \setminus \bar{\mathcal{P}}.$$

Fix $\mathcal{L}^* \in \mathcal{P}^*$, and let

$$\mathcal{L}^*_- := \{L^* \in \mathcal{L}^*\colon |\tilde L^*| \geqslant |L^*| \text{ for all } \tilde L^* \in \mathcal{L}^*\}$$

be the collection of all dyadic squares with the smallest side length in $\mathcal{L}^*$. Then, there exists $k^* \in \mathbb{N}$, with $k^* > \bar k$, such that $|L^*| = \frac{1}{4^{k^*}}$ for all $L^* \in \mathcal{L}^*_-$. Moreover, by construction of our admissible partitions, we can write

$$\mathcal{L}^*_- = \cup_{j=1}^{\ell} \{L^*_{j,i}\}_{i=1}^4 \quad \text{for some } \ell \in \mathbb{N},$$

where, for each $j \in \{1, ..., \ell\}$,

$$\cup_{i=1}^4 L^*_{j,i} =: \bar L^*_j \quad \text{is a dyadic square with } |\bar L^*_j| = \frac{1}{4^{k^*-1}}.$$

Note that $k^* - 1 \geqslant \bar k$. Then, for any $\alpha \in [c_0, 1/c_0]$, Proposition 3.5 and (3.8) yield

$$\int_{L^*_{j,i}} |u_c - u_\alpha| \,\mathrm{d}x = \int_{L^*_{j,i}} |u_c - [u_\eta]_{L^*_{j,i}}| \,\mathrm{d}x \quad \text{and} \quad \int_{\bar L^*_j} |u_c - u_\alpha| \,\mathrm{d}x = \int_{\bar L^*_j} |u_c - [u_\eta]_{\bar L^*_j}| \,\mathrm{d}x$$

for all $j \in \{1, ..., \ell\}$ and $i \in \{1, ..., 4\}$. Thus, by (1.11),

$$\alpha_{L^*_{j,i}} = \alpha_{\bar L^*_j} = c_0$$

for all $j \in \{1, ..., \ell\}$ and $i \in \{1, ..., 4\}$. Consequently (see Figure 1), defining

$$\bar{\mathcal{L}}^* := (\mathcal{L}^* \setminus \mathcal{L}^*_-) \cup \cup_{j=1}^{\ell} \bar L^*_j,$$

we have $\bar{\mathcal{L}}^* \in \mathcal{P}$ and, recalling Level 2,

$$\omega_{\bar{\mathcal{L}}^*} \equiv \omega_{\mathcal{L}^*} \quad \text{and} \quad u_{\omega_{\bar{\mathcal{L}}^*}} \equiv u_{\omega_{\mathcal{L}^*}}.$$

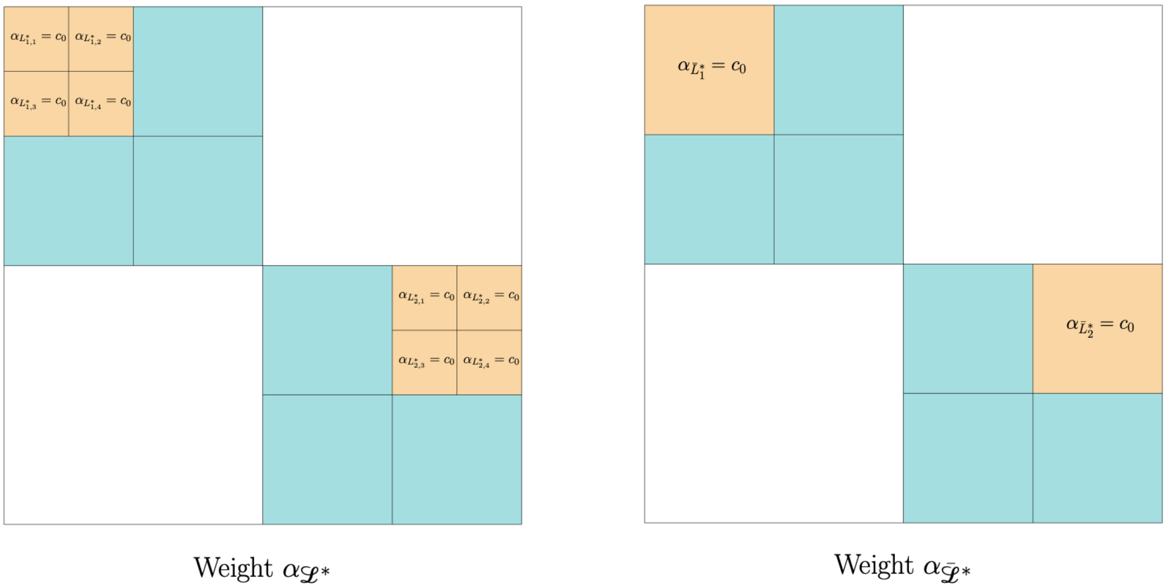

FIGURE 1. Example of two partitions, $\mathscr{L}^*$ and $\bar{\mathscr{L}}^*$, that yield the same solution at Level 2.

Note also that $|\bar{L}^*| \geqslant \frac{1}{4^{k^*-1}}$ for all $\bar{L}^* \in \bar{\mathscr{L}}^*$. If $k^* - 1 = \bar{k}$, we conclude that $\bar{\mathscr{L}}^* \in \bar{\mathscr{P}}$. Otherwise, if $k^* - 1 > \bar{k}$, we repeat the construction above $k^* - 1 - \bar{k}$ times to obtain a partition $\hat{\mathscr{L}}^* \in \bar{\mathscr{P}}$ for which

$$u_{\omega_{\hat{\mathscr{L}}^*}} \equiv u_{\omega_{\mathscr{L}^*}}.$$

Repeating this argument for each $\mathscr{L}^* \in \mathscr{P}^*$, and recalling that $\bar{\mathscr{P}}$ has finite cardinality, we deduce (3.7). $\square$

Remark 3.7. We have shown in the previous proof that the box constraint condition yields a threshold on the minimum side length of the dyadic squares of the possible optimal partitions $\mathscr{L}$ of Q. In other words, the box constraint condition yields the following stopping criterion for the refinement of the admissible partitions:

($\mathcal{S}$) There exists $\kappa \in \mathbb{N}$ such that $|L| \geqslant \frac{1}{4^\kappa}$ for all $L \in \mathscr{L}$.

In the next subsection, we establish the converse of this implication (see the proof of Theorem 1.4).

We conclude this section by proving Theorem 1.5 that shows the existence of an optimal solution to the learning scheme $(\mathcal{LS})_{TV_\omega}$.

Proof of Theorem 1.5 . This result is an immediate consequence of Subsection 3.1.3, Corollary 3.3, and Theorem 3.6. $\square$

3.4. **Stopping Criteria and Box Constraint.** In this subsection, we provide different examples of stopping criteria for the refinement of the admissible partitions, which notion was introduced in Definition 1.2, and we prove Theorem 1.4. The latter is based on the following theorem that yields a natural box constraint for the optimal parameter α associated with the TV model, provided the training data satisfy some mild conditions. The proof of (3.9) in Theorem 3.8 below uses arguments from [28] that are alternative to those in [31].

Theorem 3.8. *Let $\Omega \subset \mathbb{R}^2$ be a bounded, Lipschitz domain and, for each $\alpha \in (0, +\infty)$, let $u_\alpha \in BV(\Omega)$ be given by (1.6) with L replaced by Ω. Assume that the two following conditions on the training data hold:*

i) $TV(u_c, \Omega) < TV(u_\eta, \Omega)$;

ii) $\|u_\eta - u_c\|^2_{L^2(\Omega)} < \|[u_\eta]_\Omega - u_c\|^2_{L^2(\Omega)}$.

Then, there exists $\alpha^*_\Omega \in (0, +\infty)$ such that

$$I(\alpha^*_\Omega) = \min_{\alpha \in (0,+\infty)} I(\alpha) \quad \textit{where} \quad I(\alpha) := \int_\Omega |u_c - u_\alpha|^2 \, dx \ \textit{for } \alpha \in (0, +\infty). \tag{3.9}$$

Moreover, there exist positive constants c_Ω and C_Ω, such that any minimizer, α^*_Ω, of I over $(0,+\infty)$ satisfies $c_\Omega \leqslant \alpha^*_\Omega < C_\Omega \|u_\eta\|_{L^2(\Omega)}$. Furthermore, if $\Omega = L$ with $L \subset Q$ a dyadic square, then there exists a positive constant c_L such that any minimizer, α^*_L of I over $(0,+\infty)$ satisfies $c_L \leqslant \alpha^*_L < C_Q \|u_\eta\|_{L^2(L)}$, where C_Q is the constant given by Proposition 3.5. In particular, $\alpha^*_L \to 0$ as $|L| \to 0$.

The proof of Theorem 3.8 is hinged on the next few lemmata, the first of which is a Γ-convergence result involving the TV-model.

Lemma 3.9. *Let $\Omega \subset \mathbb{R}^2$ be a bounded, Lipschitz domain and, for each $\alpha \in (0,+\infty)$, let $u_\alpha \in BV(\Omega)$ be given by (1.6) with L replaced by Ω. Consider the family of functionals $(F_{\bar\alpha})_{\bar\alpha \in [0,+\infty]}$, where $F_{\bar\alpha} : L^1(\Omega) \to [0,+\infty]$ is defined by*

$$F_\alpha[u] := \begin{cases} \int_\Omega |u_\eta - u|^2 \,\mathrm{d}x + \alpha TV(u,\Omega) & \text{if } u \in BV(\Omega), \\ +\infty & \text{otherwise,} \end{cases} \qquad \text{for } \bar\alpha = \alpha \in (0,+\infty),$$

$$F_0[u] := \begin{cases} \int_\Omega |u_\eta - u|^2 \,\mathrm{d}x & \text{if } u \in L^2(\Omega), \\ +\infty & \text{otherwise,} \end{cases} \qquad \text{for } \bar\alpha = 0,$$

$$F_\infty[u] := \begin{cases} \int_\Omega |u_\eta - c|^2 \,\mathrm{d}x & \text{if } u \equiv c \in \mathbb{R}, \\ +\infty & \text{otherwise,} \end{cases} \qquad \text{for } \bar\alpha = +\infty.$$

Let $(\alpha_j)_{j\in\mathbb{N}} \subset (0,+\infty)$ and $\bar\alpha \in [0,\infty]$ be such that $\alpha_j \to \bar\alpha$ in $[0,+\infty]$. Then, $(F_{\alpha_j})_{j\in\mathbb{N}}$ Γ-converges to $F_{\bar\alpha}$ in $L^1(\Omega)$.

Proof. We start by showing that if $(u_j)_{j\in\mathbb{N}} \subset L^1(\Omega)$ and $u \in L^1(\Omega)$ are such that $u_j \to u$ in $L^1(\Omega)$, then

$$F_{\bar\alpha}[u] \leqslant \liminf_{j\to\infty} F_{\alpha_j}[u_j]. \tag{3.10}$$

To prove (3.10), we may assume without loss of generality that

$$\liminf_{j\to\infty} F_{\alpha_j}[u_j] = \lim_{j\to\infty} F_{\alpha_j}[u_j] < +\infty \quad \text{and} \quad \sup_{j\in\mathbb{N}} F_{\alpha_j}[u_j] < +\infty.$$

Then, $u_j \in BV(\Omega)$ for all $j \in \mathbb{N}$, $\sup_{j\in\mathbb{N}} \int_\Omega |u_\eta - u_j|^2 \,\mathrm{d}x < +\infty$ and $\sup_{j\in\mathbb{N}} \alpha_j TV(u_j,\Omega) < +\infty$. Hence, $u \in L^2(\Omega)$ and $u_j \rightharpoonup u$ weakly in $L^2(\Omega)$. Moreover,

(i) If $\bar\alpha = 0$, then (3.10) holds by the lower-semicontinuity of the L^2-norm with respect to the weak convergence in $L^2(\Omega)$.

(ii) If $\bar\alpha = \alpha \in (0,+\infty)$, then $(u_j)_{j\in\mathbb{N}}$ is bounded in $BV(\Omega)$. Thus, $u \in BV(\Omega)$ and $u_j \rightharpoonup u$ weakly-$\star$ in $BV(\Omega)$, from which (3.10) follows.

(iii) If $\bar\alpha = +\infty$, then $(u_j)_{j\in\mathbb{N}}$ is bounded in $BV(\Omega)$ with $\lim_{j\to\infty} TV(u_j,\Omega) = 0$. Thus, $u \in BV(\Omega)$, $u_j \rightharpoonup u$ weakly-$\star$ in $BV(\Omega)$, and $TV(u,\Omega) = 0$. Hence, u is constant and (3.10) holds by the lower-semicontinuity of the L^2-norm with respect to the weak convergence in $L^2(\Omega)$.

Next, we show that for any $u \in L^1(\Omega)$, there exists $(u_j)_{j\in\mathbb{N}} \subset L^1(\Omega)$ such that $u_j \to u$ in $L^1(\Omega)$ and

$$F_{\bar\alpha}[u] \geqslant \limsup_{j\to\infty} F_{\alpha_j}[u_j]. \tag{3.11}$$

To prove (3.11), we may assume, without loss of generality, that

(a) $u \in L^2(\Omega)$ if $\bar\alpha = 0$. Then, because Ω is a bounded, Lipschitz domain, we can find a sequence $u_j \in C^\infty(\overline{\Omega}) \subset BV(\Omega)$ such that $u_j \to u$ in $L^2(\Omega)$ and $\lim_{j\to\infty} \alpha_j TV(u_j,\Omega) = 0$. Consequently, (3.11) holds.

(b) $u \in BV(\Omega)$ if $\bar\alpha = \alpha \in (0,+\infty)$. Then, it suffices to take $u_j = u$ for all $j \in \mathbb{N}$ to obtain (3.11).

(c) $u \equiv c \in \mathbb{R}$ if $\bar\alpha = +\infty$. Then, as in (b), it suffices to take $u_j = \mathfrak{c}$ for all $j \in \mathbb{N}$ to obtain (3.11).

From (3.10) and (3.11), we conclude that $(F_{\alpha_j})_{j\in\mathbb{N}}$ Γ-converges to $F_{\bar\alpha}$ in $L^1(\Omega)$. $\square$

Lemma 3.10. *Let $\Omega \subset \mathbb{R}^2$ be a bounded, Lipschitz domain and let $(F_{\bar\alpha})_{\bar\alpha\in[0,+\infty]}$ the family of functionals introduced in Lemma 3.9. Given $\bar\alpha \in [0,\infty]$, set $u_{\bar\alpha} := \operatorname{argmin}_{u\in L^1(\Omega)} F_{\bar\alpha}[u]$. Then, there exists a sequence of positive numbers, $(\alpha_j)_{j\in\mathbb{N}} \subset (0,+\infty)$, such that $\alpha_j \to \bar\alpha$ in $[0,+\infty]$ as $j \to \infty$ and*

$$\lim_{j\to\infty} \int_\Omega |u_{\alpha_j} - u_{\bar\alpha}|^2 \,\mathrm{d}x = 0, \tag{3.12}$$

where $u_{\alpha_j} := \operatorname{argmin}_{u\in L^1(\Omega)} F_{\alpha_j}[u]$ *for all* $j \in \mathbb{N}$.

Proof. We first note that

$$u_{\bar\alpha} = \begin{cases} u_\alpha & \text{if } \bar\alpha = \alpha, \\ u_\eta & \text{if } \bar\alpha = 0, \\ [u_\eta]_\Omega & \text{if } \bar\alpha = +\infty. \end{cases} \tag{3.13}$$

Then,

(1) If $\bar\alpha = 0$, we choose any sequence $(\alpha_j)_{j\in\mathbb{N}} \subset (0,+\infty)$ such that $\alpha_j \to 0$, and denote by $(v_j)_{j\in\mathbb{N}}$ the sequence constructed in (a) in the proof of Lemma 3.9 with $u := u_{\bar\alpha} = u_0 = u_\eta$. Then, using the minimality of u_{α_j}, we conclude that

$$\begin{aligned}\limsup_{j\to\infty}\left(\int_\Omega |u_{\alpha_j} - u_\eta|^2 \,\mathrm{d}x + \alpha_j TV(u_{\alpha_j},\Omega)\right) &= \limsup_{j\to\infty} F_{\alpha_j}[u_{\alpha_j}] \\ &\leqslant \limsup_{j\to\infty} F_{\alpha_j}[v_j] \leqslant F_0[u_\eta] = 0.\end{aligned}$$

Thus, (3.12) holds.

(2) If $\bar\alpha = \alpha$, it suffices to take $\alpha_j = \alpha$ and $u_{\alpha_j} = u_\alpha$ for all $j \in \mathbb{N}$.

(3) If $\bar\alpha = +\infty$, we choose any sequence $(\alpha_j)_{j\in\mathbb{N}} \subset (0,+\infty)$ such that $\alpha_j \to +\infty$, and observe that the uniform bounds in $BV(\Omega)$ proved in Lemma 3.9 assert that $(F_{\alpha_j})_{j\in\mathbb{N}}$ is an equi-coercive sequence in $L^1(\Omega)$. Thus, by well-known properties of Γ-convergence on the convergence of minimizing sequences and energies (see [24, Corollary 7.20 and Theorem 7.8]), together with the uniqueness of minimizers of F_{α_j} and F_∞, we have that $u_{\alpha_j} \rightharpoonup u_{+\infty} = [u_\eta]_\Omega$ weakly-$\star$ in $BV(\Omega)$ and $\lim_{j\to\infty} F_{\alpha_j}[u_{\alpha_j}] = F_\infty[[u_\eta]_\Omega]$. In particular, $u_{\alpha_j} \rightharpoonup [u_\eta]_\Omega$ weakly in $L^2(\Omega)$ and, using the L^2-weak lower semicontinuity of the L^2 norm and the definition of F_{α_j} and F_∞,

$$\begin{aligned}\int_\Omega |[u_\eta]_\Omega - u_\eta|^2 \,\mathrm{d}x &\leqslant \liminf_{j\to\infty}\int_\Omega |u_{\alpha_j} - u_\eta|^2 \,\mathrm{d}x \leqslant \limsup_{j\to\infty}\int_\Omega |u_{\alpha_j} - u_\eta|^2 \,\mathrm{d}x \\ &\leqslant \lim_{j\to\infty} F_{\alpha_j}[u_{\alpha_j}] = F_\infty[[u_\eta]_\Omega] = \int_\Omega |[u_\eta]_\Omega - u_\eta|^2 \,\mathrm{d}x.\end{aligned}$$

Thus, $\lim_{j\to\infty}\int_\Omega |u_{\alpha_j}|^2 \,\mathrm{d}x = \int_\Omega |[u_\eta]_\Omega|^2 \,\mathrm{d}x$, which together with $u_{\alpha_j} \rightharpoonup [u_\eta]_\Omega$ weakly in $L^2(\Omega)$, proves (3.12). □

Lemma 3.11. *Let* $\Omega \subset \mathbb{R}^2$ *be a bounded, Lipschitz domain, and let* $I : (0,+\infty) \to [0,+\infty)$ *be the function defined in* (3.9)*. Then, the lower-semicontinuous envelope of* I *on* $[0,\infty]$*,* $I^{sc^-} : [0,+\infty] \to [0,+\infty]$*, defined for* $\bar\alpha \in [0,+\infty]$ *by*

$$I^{sc^-}(\bar\alpha) := \inf\Big\{\liminf_{j\to\infty} I(\alpha_j) \colon\, (\alpha_j)_{j\in\mathbb{N}} \subset (0,+\infty),\ \alpha_j \to \bar\alpha \text{ in } [0,+\infty]\Big\}, \tag{3.14}$$

satisfies

$$I^{sc^-}(\bar\alpha) = \begin{cases} I(\alpha) = \|u_\alpha - u_c\|^2_{L^2(\Omega)} & \text{if } \bar\alpha = \alpha \in (0,+\infty), \\ \|u_\eta - u_c\|^2_{L^2(\Omega)} & \text{if } \bar\alpha = 0, \\ \|[u_\eta]_L - u_c\|^2_{L^2(\Omega)} & \text{if } \bar\alpha = +\infty. \end{cases} \tag{3.15}$$

Proof. We first note that the function I^{sc^-} in (3.14) is lower-semicontinuous on $[0,\infty]$ and $I^{sc^-} \leqslant I$ in $(0,+\infty)$. Next, we denote by $\widetilde{I}$ the function on $[0,+\infty]$ defined by the right-hand side of (3.15), and observe that

$$\widetilde{I}(\bar\alpha) = \|u_{\bar\alpha} - u_c\|^2_{L^2(\Omega)},$$

where $u_{\bar\alpha} := \operatorname{argmin}_{u\in L^1(\Omega)} F_{\bar\alpha}(u)$ is given by (3.13). We want to show that $I^{sc^-} \equiv \widetilde{I}$.

We distinguish two cases:

– *Case* $\bar\alpha \in \{0,+\infty\}$*:* Let $(\alpha_j)_{j\in\mathbb{N}} \subset (0,+\infty)$ be such that $\alpha_j \to \bar\alpha$ in $[0,+\infty]$. As proved in (1) and (3) in Lemma 3.10, we have

$$\widetilde{I}(\bar\alpha) = \|u_{\bar\alpha} - u_c\|^2_{L^2(\Omega)} = \lim_{j\to\infty}\|u_{\alpha_j} - u_c\|^2_{L^2(\Omega)} = \lim_{j\to\infty} I(\alpha_j). \tag{3.16}$$

Thus, taking the infimum over all such sequences $(\alpha_j)_{j\in\mathbb{N}} \subset (0,+\infty)$, we conclude that $I^{sc^-}(\bar\alpha) = \widetilde{I}(\bar\alpha)$ for $\bar\alpha \in \{0,+\infty\}$.

– *Case* $\bar\alpha = \alpha \in (0,+\infty)$*:* By Lemma 3.10, there exists a sequence $(\alpha_j)_{j\in\mathbb{N}} \subset (0,+\infty)$ such that $\alpha_j \to \alpha$ and for which (3.16) holds. Thus, $\widetilde{I}(\alpha) \geqslant I^{sc^-}(\alpha)$. Conversely, let $(\alpha_j)_{j\in\mathbb{N}} \subset (0,+\infty)$ be such that $\alpha_j \to \alpha$. As argued before, we observe that the uniform bounds in $BV(\Omega)$ proved in Lemma 3.9 assert that $(F_{\alpha_j})_{j\in\mathbb{N}}$ is an equi-coercive sequence in $L^1(\Omega)$. Thus, as before, by well-known properties of Γ-convergence on the convergence of minimizing sequences and energies (see [24, Corollary 7.20 and Theorem 7.8]), together with the uniqueness of minimizers of F_{α_j} and F_α, we have that $u_{\alpha_j} \rightharpoonup u_\alpha$ weakly-$\star$ in $BV(\Omega)$ and $\lim_{j\to\infty} F_{\alpha_j}[u_{\alpha_j}] = F_\alpha[u_\alpha]$. In particular, $u_{\alpha_j} \rightharpoonup u_\alpha$ weakly in $L^2(\Omega)$. Thus,

$$\widetilde{I}(\alpha) = \|u_\alpha - u_c\|^2_{L^2(\Omega)} \leqslant \liminf_{j\to\infty} \|u_{\alpha_j} - u_c\|^2_{L^2(\Omega)} = \liminf_{j\to\infty} I(\alpha_j).$$

Taking the infimum of all such sequences $(\alpha_j)_{j\in\mathbb{N}} \subset (0,+\infty)$, we conclude that $\widetilde{I}(\alpha) \leqslant I^{sc^-}(\alpha)$. □

Proof of Theorem 3.8. We will proceed in three steps.

<u>*Step 1.*</u> *We prove that if condition i) in the statement holds (i.e.,* $TV(u_\eta,\Omega) - TV(u_c,\Omega) > 0$*), then there exists* $\alpha \in (0,+\infty)$ *such that*

$$\|u_\alpha - u_c\|^2_{L^2(\Omega)} < \|u_\eta - u_c\|^2_{L^2(\Omega)}. \tag{3.17}$$

To show (3.17), we first recall (see [16]) that for any $\alpha \in (0,+\infty)$, there exists a unique $u_\alpha \in BV(\Omega) \subset L^2(\Omega)$ such that

$$u_\alpha = \operatorname{argmin}_{u\in L^1(\Omega)} F_\alpha[u] = \operatorname{argmin}_{u\in L^2(\Omega)} F_\alpha[u], \tag{3.18}$$

which allow us to regard F_α as a sum of two convex functionals on $L^2(\Omega)$ with values in $[0,+\infty]$. Precisely,

$$F_\alpha[u] = F^1_\alpha[u] + F^2_\alpha[u],$$

where, for $u \in L^2(\Omega)$,

$$F^1_\alpha[u] := \|u - u_\eta\|^2_{L^2(\Omega)} \quad \text{and} \quad F^2_\alpha[u] := \begin{cases} \alpha TV(u,\Omega) & \text{if } u \in BV(\Omega), \\ +\infty & \text{otherwise.} \end{cases}$$

Denoting by $\partial F(v) \in (L^2(\Omega))' \cong L^2(\Omega)$ the subdifferential of a convex functional $F : L^2(\Omega) \to [0,+\infty]$ at $v \in L^2(\Omega)$, we conclude from (3.18) that

$$0 \in \partial F_\alpha(u_\alpha) \quad \text{or, equivalently,} \quad 2(u_\eta - u_\alpha) \in \partial F^2_\alpha(u_\alpha).$$

Consequently,

$$\begin{aligned} 0 &\geqslant F^2_\alpha[u_\alpha] - F^2_\alpha[u_c] + \int_\Omega 2(u_\eta - u_\alpha)(u_c - u_\alpha)\,\mathrm{d}x \\ &\geqslant F^2_\alpha[u_\alpha] - F^2_\alpha[u_c] + \int_\Omega 2(u_\eta - u_\alpha)(u_c - u_\alpha)\,\mathrm{d}x - \|u_\alpha - u_\eta\|^2_{L^2(\Omega)} \\ &= \alpha\big(TV(u_\alpha,\Omega) - TV(u_c,\Omega)\big) + \|u_\alpha - u_c\|^2_{L^2(\Omega)} - \|u_\eta - u_c\|^2_{L^2(\Omega)}. \end{aligned}$$

Hence,

$$\|u_\eta - u_c\|^2_{L^2(\Omega)} - \|u_\alpha - u_c\|^2_{L^2(\Omega)} \geqslant \alpha\big(TV(u_\alpha,\Omega) - TV(u_c,\Omega)\big). \tag{3.19}$$

We claim that

$$TV(u_\alpha,\Omega) \nearrow TV(u_\eta,\Omega) \quad \text{as} \quad \alpha \searrow 0. \tag{3.20}$$

Assuming that the preceding claim holds, the condition $TV(u_\eta,\Omega) - TV(u_c,\Omega) > 0$ allows us to find $\tilde\alpha \in (0,+\infty)$ for which the left-hand side of (3.19) with $\alpha = \tilde\alpha$ is strictly positive. Thus, $\|u_\eta - u_c\|^2_{L^2(\Omega)} > \|u_{\tilde\alpha} - u_c\|^2_{L^2(\Omega)}$, which proves (3.17).

To conclude Step 4, we are left to prove (3.20). Using (3.18), for all α, $\beta \in (0,+\infty)$ with $\alpha < \beta$, we have that

$$\beta TV(u_\beta,\Omega) \leqslant F_\beta[u_\beta] \leqslant F_\beta[u_\eta] = \beta TV(u_\eta,\Omega)$$

and

$$\begin{aligned} \|u_\alpha - u_\eta\|^2_{L^2(\Omega} + \alpha TV(u_\alpha,\Omega) &\leqslant \|u_\beta - u_\eta\|^2_{L^2(\Omega} + \alpha TV(u_\beta,\Omega) \\ &= \|u_\beta - u_\eta\|^2_{L^2(\Omega} + \beta TV(u_\beta,\Omega) + (\alpha-\beta)TV(u_\beta,\Omega) \\ &\leqslant \|u_\alpha - u_\eta\|^2_{L^2(\Omega} + \beta TV(u_\alpha,\Omega) + (\alpha-\beta)TV(u_\beta,\Omega). \end{aligned}$$

Hence, $TV(u_\beta,\Omega) \leqslant TV(u_\eta,\Omega)$ and $TV(u_\alpha,\Omega) \geqslant TV(u_\beta,\Omega)$. Finally, using the first of these estimates and (1) in the proof of Lemma 3.10 with an arbitrary decreasing sequence $(\beta_j)_{j\in\mathbb{N}}$ converging to 0, the lower-semicontinuity of the total variation with respect to the strong convergence in L^1 yields

$$TV(u_\eta,\Omega) \geqslant \limsup_{j\to\infty} TV(u_{\beta_j},\Omega) \geqslant \liminf_{j\to\infty} TV(u_{\beta_j},\Omega) \geqslant TV(u_\eta,\Omega).$$

This concludes the proof of (3.20).

<u>Step 2.</u> *We prove that if condition ii) in the statement holds, (i.e., $\|u_\eta - u_c\|^2_{L^2(\Omega)} < \|[u_\eta]_\Omega - u_c\|^2_{L^2(\Omega)}$), then there exits $\alpha \in (0,+\infty)$ such that*

$$\|u_\alpha - u_c\|^2_{L^2(\Omega)} < \|[u_\eta]_\Omega - u_c\|^2_{L^2(\Omega)}. \tag{3.21}$$

Using Lemma 3.10 with $\bar\alpha = 0$ together with *ii)*, we obtain

$$\begin{aligned}\limsup_{j\to\infty} \|u_{\alpha_j} - u_c\|^2_{L^2(\Omega)} &\leqslant \limsup_{j\to\infty} \Big(\|u_{\alpha_j} - u_\eta\|^2_{L^2(\Omega)} + \|u_\eta - u_c\|^2_{L^2(\Omega)}\Big)\\ &= \|u_\eta - u_c\|^2_{L^2(\Omega)} < \|[u_\eta]_\Omega - u_c\|^2_{L^2(\Omega)},\end{aligned}$$

from which (3.21) follows.

<u>Step 3.</u> *We conclude the proof of Theorem 3.8.*

We first show (3.9). Because I^{sc^-} is a lower-semicontinuos function on the compact set $[0,+\infty]$, I^{sc^-} attains a minimum on $[0,+\infty]$. By (3.15), (3.17), and (3.21), we conclude that I^{sc^-} attains its minimum at some $\alpha^* \in (0,+\infty)$. Thus, using (3.15) once more,

$$I(\alpha^*) = I^{sc^-}(\alpha^*) = \min_{\bar\alpha\in[0,+\infty]} I^{sc^-}(\bar\alpha) = \min_{\alpha\in(0,+\infty)} I^{sc^-}(\alpha) = \min_{\alpha\in(0,+\infty)} I(\alpha), \tag{3.22}$$

which yields (3.9).

Next, to prove the existence of c_Ω as stated, assume that there exist $(\alpha^*_j)_{j\in\mathbb{N}} \subset (0,+\infty)$ such that $\alpha^*_j \to 0$ and (3.22) holds with $\alpha^* = \alpha^*_j$. Then, using the lower semi-continuity of I^{sc^-} on $[0,+\infty]$,

$$\min_{\bar\alpha\in[0,+\infty]} I^{sc^-}(\bar\alpha) \leqslant I^{sc^-}(0) \leqslant \liminf_{j\to\infty} I^{sc^-}(\alpha^*_j) = \min_{\bar\alpha\in[0,+\infty]} I^{sc^-}(\bar\alpha),$$

which is false by (3.17). This establishes the existence of the constant c_Ω.

On the other hand, as mentioned in the proof of Proposition 3.5, [39, Proposition 2.5.7] yields a positive constant, C_Ω, such that $u_\alpha \equiv [u_\eta]_\Omega$ for all $\alpha \geqslant C_\Omega\|u_\eta\|_{L^2(\Omega)}$. This fact, (3.21), and (3.22) show that we must have $\alpha^*_\Omega < C_\Omega\|u_\eta\|_{L^2(\Omega)}$. Finally, the $\Omega = L$ case follows from Proposition 3.5. □

Next, we prove Theorem 1.4.

Proof of Theorem 1.4. In view of Theorem 3.6 (also see Remark 3.7), the statement in (a) follows. Conversely, the statement in (b) can be proved arguing as in Subsection 3.1.2 and defining

$$c_0 := \min\left\{\min_{L\in\mathscr{L}\in\bar{\mathcal{P}}} c_L, \left(c_Q\|u_\eta\|^2_{L^2(Q)}\right)^{-1}\right\},$$

where c_L and C_Q are the constants given by Theorem 3.8. □

We conclude this section with some examples of stopping criteria for the refinement of the admissible partitions as defined in Definition 1.2.

Example 3.12. Here, we give an example of a stopping criteria that, heuristically, means that we only refine a given dyadic square L, if the distance of the restored image in L to the clean image is greater than or equal to the sum of the distances of the restored images in each of the subdivisions of L to the clean image, modulo a threshold that is determined by the user.

To make this idea precise, we introduce some notation. Given a dyadic square $L^{(1)} \subset Q$ of side length $\frac{1}{2^{k+1}}$, we can find three other dyadic squares, which we denote by $L^{(2)}$, $L^{(3)}$, and $L^{(4)}$, of side length $\frac{1}{2^{k+1}}$ and such that $L := \cup_{i=1}^4 L^{(i)}$ is a dyadic square of side length $\frac{1}{2^k}$. We observe further that $L^{(2)}$, $L^{(3)}$, and $L^{(4)}$ are uniquely determined by the requirement that L is a dyadic square. Using this notation, we fix $\delta > 0$ and set up an admissible criteria as follows:

($\mathcal{S}$) (*i*) Q is admissible;

(*ii*) If $L \subset Q$ is an admissible dyadic square, then each dyadic square $L^{(i)} \subset L$, with $i \in \{1, ..., 4\}$

and $\cup_{i=1}^4 L^{(i)} = L$, is admissible if

$$\|u_c - u_L\|^2_{L^2(L)} \geqslant \sum_{i=1}^{4} \|u_c - u_{L^{(i)}}\|^2_{L^2(L^{(i)})} + \delta. \tag{3.23}$$

As we prove next,

$$\bar{\mathcal{P}} := \big\{\mathcal{L} \in \mathcal{P} \colon\ L \text{ satisfies } (\mathcal{S}) \text{ for all } L \in \mathcal{L}\big\}$$

has finite cardinality, which shows that $(\mathcal{S})$ as above provides a stopping criteria for the refinement of the admissible partition.

To show that $\bar{\mathcal{P}}$ has finite cardinality, we first observe that if L satisfies $(\mathcal{S})$, then we can find k dyadic squares, $L_1, ..., L_k$, where $k \in \mathbb{N}$ is such that $|L| = \frac{1}{4^k}$, satisfying

$$Q = L_1 \supset .... \supset L_k \supset L, \qquad |L_k| = \frac{1}{4^{k-1}}, \qquad L_k \text{ satisfies } (\mathcal{S}).$$

Then, using (3.23), we conclude that

$$\|u_c - u_Q\|^2_{L^2(Q)} = \|u_c - u_{L_1}\|^2_{L^2(L_1)} \geqslant c_k + k\delta$$

for some positive constant c_k, which can only hold true if k is small enough. In other words, there exists $k_\delta \in \mathbb{N}$ such that if L satisfies $(\mathcal{S})$, then $|L| \geqslant \frac{1}{4^{k_\delta}}$. Hence, $\bar{\mathcal{P}}$ has finite cardinality.

4. Analysis of the Regularized Weighted-TV and Weighted-Fidelity learning schemes $(\mathcal{LS})_{TV_{\omega_\varepsilon}}$ and $(\mathcal{LS})_{TV-Fid_\omega}$

The results proved in the preceding section for the weighted-TV learning scheme can be easily adapted to the case of the regularized weighted-TV and the weighted-fidelity learning schemes, $(\mathcal{LS})_{TV_{\omega_\varepsilon}}$ and $(\mathcal{LS})_{TV-Fid_\omega}$. For the former, we prove here only Proposition 1.6 and provide an example of a sequence of regularized weights satisfying the conditions assumed in this result. Moreover, we highlight an open problem that is intimately related to the convergence of the solutions to $(\mathcal{LS})_{TV_{\omega_\varepsilon}}$ as $\varepsilon \to 0^+$ (see Subsection 4.1 below). Regarding $(\mathcal{LS})_{TV-Fid_\omega}$, and for completeness, we state the analogue existence and equivalence results for the weighted-fidelity learning scheme (see Subsection 4.2 below).

4.1. The $(\mathcal{LS})_{TV_{\omega_\varepsilon}}$ learning scheme.

Next, we prove Proposition 1.6 and provide an example of a sequence $(\omega^\varepsilon_{\mathcal{L}})_\varepsilon$ as in (1.14).

Proof of Proposition 1.6. We show that

$$E^\varepsilon_{\mathcal{L}}[u] \leqslant E_{\mathcal{L}}[u] \tag{4.1}$$

for all $u \in L^1(Q)$, from which (1.15) follows.

Let $u \in L^1(Q)$ be such that $E_{\mathcal{L}}[u] < \infty$. Then, $u \in BV_{\omega_{\mathcal{L}}}(Q)$ and recalling the definition and properties of the space of weighted BV-function discussed in Section 3.2, we have that $u \in BV_{\omega^\varepsilon_{\mathcal{L}}}(Q)$ with $TV_{\omega^\varepsilon_{\mathcal{L}}}(u, Q) \leqslant TV_{\omega_{\mathcal{L}}}(u, Q)$, using the estimate $\omega^\varepsilon_{\mathcal{L}} \leqslant \omega_{\mathcal{L}}$ a.e. in Q in (1.14). Thus, (4.1) holds. ☐

Example 4.1. An example of a sequence $(\omega^\varepsilon_{\mathcal{L}})_\varepsilon$ as in (1.14) can be constructed combining a diagonalization argument with a mollification of a Moreau–Yosida type approximation of $\omega^{sc^-}_{\mathcal{L}}$. Precisely, for each $k \in \mathbb{N}$, let $\omega_k : Q \to (0, \infty)$ be given by

$$\omega_k(x) := \inf\big\{\omega^{sc^-}_{\mathcal{L}}(y) + k|x - y| \colon\ y \in Q\big\} \quad \text{for } x \in Q. \tag{4.2}$$

We recall that each ω_k is a k-Lipschitz function, and we have (see [13, Theorem 2.1.2] for instance)

$$\omega_k \nearrow \omega^{sc^-}_{\mathcal{L}} \quad \text{pointwise everywhere in } Q. \tag{4.3}$$

Moreover, as we show next,

$$\lim_{k\to\infty} \|\omega_k - \omega^{sc^-}_{\mathcal{L}}\|_{L^\infty(K)} = 0 \tag{4.4}$$

for any compact set K such that $K \subset \mathrm{int}(L)$, where $L \in \mathcal{L}$ is arbitrary.

In fact, let $L \in \mathcal{L}$ and let K be a compact set such that $K \subset \mathrm{int}(L)$. Fix $\tau > 0$ and set $\delta := \frac{\mathrm{dist}(K, \partial L)}{2}$. Note that $\delta > 0$ and

$$\omega^{sc^-}_{\mathcal{L}}(x) = \alpha_L \quad \text{for all } x \in \mathrm{int}(L) \tag{4.5}$$

because $\omega_{\mathcal{L}}(x) = \alpha_L$ for all $x \in L$. Moreover, using (4.2), given $\bar{x} \in K$ we can find $y_k \in Q$ such that

$$\omega_k(\bar{x}) + \tau \geqslant \omega^{sc^-}_{\mathcal{L}}(y_k) + \ k|\bar{x} - y_k|. \tag{4.6}$$

Hence, using (4.3) and nonnegativity of ω^{sc^-}, we obtain

$$|\bar{x} - y_k| \leqslant \frac{\omega_k(\bar{x}) + \tau - \omega_{\mathscr{L}}^{sc^-}(y_k)}{k} \leqslant \frac{\|\omega_{\mathscr{L}}^{sc^-}\|_{L^\infty(Q)} + \tau}{k} < \delta$$

for all $k \geqslant k_0$ and for some $k_0 \in \mathbb{N}$ that is independent of $\bar{x}$. Then, $y_k \in \operatorname{int}(L)$ for all $k \geqslant k_0$. Consequently, (4.5)–(4.6) then yield

$$\omega_k(\bar{x}) + \tau \geqslant \omega_{\mathscr{L}}^{sc^-}(y_k) = \alpha_L = \omega_{\mathscr{L}}^{sc^-}(\bar{x})$$

for all $k \geqslant k_0$. Hence,

$$0 \leqslant \omega_{\mathscr{L}}^{sc^-}(\bar{x}) - \omega_k(\bar{x}) \leqslant \tau$$

for all $k \geqslant k_0$. Taking the supremum on $\bar{x} \in K$ in the preceding estimate yields (4.4).

On the other hand, for each $k \in \mathbb{N}$, a standard mollification argument yields a sequence $(\omega_\varepsilon^{(k)})_\varepsilon \subset C^\infty(\overline{Q})$ such that

$$\lim_{\varepsilon \to 0^+} \|\omega_\varepsilon^{(k)} - \omega_k\|_{L^\infty(Q)} = 0. \tag{4.7}$$

Finally, denoting by $Q(x, \delta)$ the open square centered at $x \in \mathbb{R}^2$ and side-length δ, we can write $\mathscr{L} = \cup_{i=1}^{\ell} L_i$ with $\operatorname{int}(L_i) = Q(x_i, \delta_i)$, for some $\ell \in \mathbb{N}$, $x_i \in L_i$, and $\delta_i > 0$. Then, exploiting the countability of the family

$$\mathcal{K} := \cup_{i=1}^{\ell} \{K_i := \overline{Q(x_i, r_i)} \colon\ r_i \in \mathbb{Q} \cap (0, \delta_i)\} \tag{4.8}$$

and a diagonalization argument together with (4.4) and (4.7), we can find a sequence $(\omega_{\mathscr{L}}^\varepsilon)_\varepsilon$ such that

$$\lim_{\varepsilon \to 0^+} \|\omega_{\mathscr{L}}^\varepsilon - \omega_{\mathscr{L}}^{sc^-}\|_{L^\infty(K)} = 0 \tag{4.9}$$

for all compact set $K \in \mathcal{K}$. From the definition of $\mathcal{K}$ in (4.8), we get that (4.9) also holds for all compact set $K \subset \operatorname{int}(L)$ and for any $L \in \mathscr{L}$. Furthermore, using the fact that mollification preserves monotonicity, we deduce from (4.3) and (4.7) that $\omega_{\mathscr{L}}^\varepsilon \nearrow \omega_{\mathscr{L}}^{sc^-}$ everywhere in Q.

To conclude that (1.14) also holds, it suffices to observe that $\omega_{\mathscr{L}}^{sc^-} \leqslant \omega_{\mathscr{L}}$ in Q, $\omega_{\mathscr{L}}^{sc^-} \equiv \omega_{\mathscr{L}}$ in $\cup_{i=1}^{\ell} \operatorname{int}(L_i)$, $|Q \setminus \cup_{i=1}^{\ell} \operatorname{int}(L_i)| = 0$.

Remark 4.2 (Existence of solutions to the learning scheme $(\mathcal{LS})_{TV_{\omega_\varepsilon}}$). For fixed ε, we can apply the results proved in Section 3. In particular, there exists an optimal solution u_ε^* to the learning scheme $(\mathcal{LS})_{TV_{\omega_\varepsilon}}$ in (1.12) with (1.13) replaced by (1.11) (cf. Theorem 1.5).

Open Problem 4.3. *An interesting open problem is whether condition* (1.14) *yields the convergence*

$$\lim_{\varepsilon \to 0^+} TV_{\omega_{\mathscr{L}}^\varepsilon}(u, Q) = TV_{\omega_{\mathscr{L}}}(u, Q) \tag{4.10}$$

for all $u \in BV_{\omega_{\mathscr{L}}}(Q)$. Because sets of zero Lebesgue measure may not have zero $|Du|$ measure, we do not expect (4.10) *to hold unless the almost everywhere pointwise convergence in* (1.14) *is replaced by everywhere pointwise convergence.*

To the best of our knowledge, the closest result in this direction is [13, Lemma 2.1.4]*, which shows the following. If $\tilde{\omega} \geqslant 0$ is lower semi-continuous in Q and $u : Q \to \mathbb{R}$ is measurable, then we can find a sequence of Lipschitz weights, $(\tilde{\omega}_k^{(u)})_{k \in \mathbb{N}}$, depending on u, such that $\tilde{\omega}_k^{(u)} \nearrow \tilde{\omega}$ pointwise everywhere in Q and* (4.10) *holds (with $\omega_{\mathscr{L}}^\varepsilon$ and $\omega_{\mathscr{L}}$ replaced by $\tilde{\omega}_k^{(u)}$ and $\tilde{\omega}$, respectively).*

4.2. **The $(\mathcal{LS})_{TV-Fid_\omega}$ learning scheme.** Given a dyadic square $L \subset Q$ and $\alpha \in (0, \infty)$, we have

$$\begin{aligned} &\operatorname{argmin}\left\{\frac{1}{\alpha}\int_L |u_\eta - u|^2 \, dx + TV(u, L) \colon\ u \in BV(L)\right\} \\ &\qquad = \operatorname{argmin}\left\{\int_L |u_\eta - u|^2 \, dx + \alpha TV(u, L) \colon\ u \in BV(L)\right\}. \end{aligned}$$

Consequently, Proposition 3.5 and Theorem 3.8 remain unchanged if we replace (1.6) by (1.18). These two results are the main tools to prove Theorems 1.4 and 1.5. Using this observation, the arguments used in Section 3 can be reproduced here for the weighted-fidelity learning scheme to conclude the two following theorems.

Theorem 4.4 (Existence of solutions to $(\mathcal{LS})_{TV-Fid_\omega}$). *There exists an optimal solution u^* to the learning scheme $(\mathcal{LS})_{TV-Fid_\omega}$ in (1.16) with (1.17) replaced by (1.11).*

As before, the previous existence theorem holds true under any stopping criterion for the refinement of the admissible partitions provided that the training data satisfies suitable conditions, as stated in the next result.

Theorem 4.5 (Equivalence between box constraint and stopping criterion). *Consider the learning scheme $(\mathcal{LS})_{TV-Fid_\omega}$ in (1.16). The two following conditions hold:*

(a) If we replace (1.17) by (1.11), then there exists a stopping criterion $(\mathcal{S})$ for the refinement of the admissible partitions as in Definition 1.2.

(b) Assume that there exists a stopping criterion $(\mathcal{S})$ for the refinement of the admissible partitions as in Definition 1.2 such that the training data satisfies for all $L \in \bigcup_{\mathcal{L}\in\bar{\mathcal{P}}}\mathcal{L}$, with $\bar{\mathcal{P}}$ as in Definition 1.2, the conditions

(i) $TV(u_c, L) < TV(u_\eta, L)$;

(ii) $\|u_\eta - u_c\|^2_{L^2(L)} < \|[u_\eta]_L - u_c\|^2_{L^2(L)}$.

Then, there exists $c_0 \in \mathbb{R}^+$ such that the optimal solution u^* provided by $(\mathcal{LS})_{TV-Fid_\omega}$ with $\mathcal{P}$ replaced by $\bar{\mathcal{P}}$ coincides with the optimal solution u^* provided by $(\mathcal{LS})_{TV-Fid_\omega}$ with (1.17) replaced by (1.11).

5. Analysis of the Weighted-TGV learning scheme $(\mathcal{LS})_{TGV_\omega}$

This section is devoted to proving the existence of solutions to the training scheme $(\mathcal{LS})_{TGV_\omega}$ described in (1.20). We begin by providing the precise definition of the quantities $\mathcal{V}_{\omega^0_{\mathcal{L}}}$ and $\mathcal{V}_{\omega^1_{\mathcal{L}}}$ in (1.24), which are particular instances of the general definition of the weighted variation of a Radon measure introduced in Section 2 (see (2.4)).

Definition 5.1. *Let Ω be an open set in $\mathbb{R}^n$ and $\omega : \Omega \to [0, +\infty)$ a locally integrable function. Given $u \in L^1_{\omega,\mathrm{loc}}(\Omega)$ and $v \in L^1_{\omega,\mathrm{loc}}(\Omega;\mathbb{R}^n)$ (see (3.2)), we set*

$$\mathcal{V}_\omega(Du - v, \Omega) := \sup\left\{\int_\Omega (u \operatorname{div}\varphi + v\cdot\varphi)\,\mathrm{d}x : \varphi \in \mathrm{Lip}_c(\Omega;\mathbb{R}^n),\ |\varphi| \leqslant \omega\right\} \tag{5.1}$$

and

$$\mathcal{V}_\omega(\mathcal{E}v, \Omega) := \sup\left\{\int_\Omega (v\cdot\operatorname{div}\xi)\,\mathrm{d}x : \xi \in \mathrm{Lip}_c(\Omega;\mathbb{R}^{n\times n}_{sym}),\ |\xi| \leqslant \omega\right\}, \tag{5.2}$$

where $(\operatorname{div}\xi)_j = \sum_{k=1}^n \frac{\partial \xi_{jk}}{\partial x_k}$ for each $j \in \{1, \dots, n\}$.

Remark 5.2. Recalling (2.4), we are using an abuse of notation in the preceding definition as we are not requiring Du nor $\mathcal{E}v$ to be Radon measures. However, if $u \in BV(\Omega)$, then (5.1) is the ω-weighted variation of the Radon measure $Du - v := Du - v\mathcal{L}^n\lfloor\Omega \in \mathcal{M}(\Omega;\mathbb{R}^n)$ in the sense of (2.4). Similarly, if $v \in BD(\Omega)$, then (5.2) is the ω-weighted variation of the Radon measure $\mathcal{E}v \in \mathcal{M}(\Omega;\mathbb{R}^{n\times n}_{sym})$ in the sense of (2.4).

Analogously to the $(\mathcal{LS})_{TV_\omega}$ case, we analyze each level of $(\mathcal{LS})_{TGV_\omega}$ in a dedicated subsection.

To prove existence of a solution to the learning scheme $(\mathcal{LS})_{TGV_\omega}$ in (1.20), we argue by a box-constraint approach in which we replace the requirement $\alpha = (\alpha_0, \alpha_1) \in \mathbb{R}^+ \times \mathbb{R}^+$ by the stricter condition (1.29). In this case, we replace (1.21) by

$$\bar{\alpha}_L = \inf\left\{\operatorname{argmin}\left\{\int_L |u_c - u_\alpha|^2\,\mathrm{d}x : \alpha \in \left[c_0, \tfrac{1}{c_0}\right] \times \left[c_1, \tfrac{1}{c_1}\right]\right\}\right\}. \tag{5.3}$$

Throughout this section, for $u \in L^2(\Omega)$, we denote by $\langle u\rangle_\Omega$ the affine projection of u given by the unique solution to the minimization problem

$$\min\left\{\int_\Omega |u - v|^2\,\mathrm{d}x : v \text{ is affine in } \Omega\right\}, \tag{5.4}$$

which will play an analogous role to the average $[u]_\Omega$ in the TV case treated in Section 3. Note that we have the orthogonality property

$$\int_\Omega (u - \langle u\rangle_\Omega)\langle u\rangle_\Omega \,\mathrm{d}x = 0 \tag{5.5}$$

for every $u \in L^2(\Omega)$, since $\langle u\rangle_\Omega$ is the Hilbert projection of u onto a finite dimensional subspace of $L^2(\Omega)$.

5.1. **On Level 3.** We provide here an analysis of Level 3, and minor variants thereof, of the learning scheme $(\mathcal{LS})_{TGV_\omega}$ in (1.20).

5.1.1. As in the weighted TV scheme case, the parameter α_L in Level 3 of $(\mathcal{LS})_{TGV_\omega}$ (see (1.21)) is uniquely determined by definition, and it satisfies $\alpha_L \in [0,+\infty]^2$.

5.1.2. In view of Theorem 5.13 (see Subsection 5.4), if $L \in \mathcal{L}$ is such that

$$TGV_{\hat\alpha_0,\hat\alpha_1}(u_c, L) < TGV_{\hat\alpha_0,\hat\alpha_1}(u_\eta, L) \quad \text{and} \quad \|u_\eta - u_c\|^2_{L^2(L)} < \|\langle u_\eta\rangle_L - u_c\|^2_{L^2(L)} \tag{5.6}$$

for some $\hat\alpha = (\hat\alpha_0, \hat\alpha_1)$, then

$$\begin{aligned}\operatorname{arginf}&\left\{\int_L |u_c - u_\alpha|^2 \,\mathrm{d}x\colon\ \alpha \in \mathbb{R}^+ \times \mathbb{R}^+\right\} = \\ &\operatorname{argmin}\left\{\int_L |u_c - u_\alpha|^2 \,\mathrm{d}x\colon\ \alpha \in \mathbb{R}^+ \times \mathbb{R}^+ \text{ is s.t. } c_L \leqslant \min\{\alpha_0,\alpha_1\} < C_Q\|u_\eta\|_{L^2(L)}\right\},\end{aligned}$$

where c_L and C_Q are positive constants, with c_Q depending only on Q. Furthermore, because each partition $\mathcal{L} \in \mathcal{P}$ is finite, it follows that if (5.6) holds for all $L \in \mathcal{L}$, then

$$\alpha_L \in \alpha_L \in \left[\min_{L\in\mathcal{L}} c_L, +\infty\right]^2 \setminus \{(+\infty,+\infty)\}. \tag{5.7}$$

5.1.3. If we consider Level 3 with (1.21) replaced by (5.3), then the minimum

$$\min_{\alpha\in[c_0,\frac{1}{c_0}]\times[c_1,\frac{1}{c_1}]} \int_L |u_c - u_\alpha|^2 \,\mathrm{d}x$$

exists as the minimum of a lower semicontinuous function (see Lemma 5.18 in Subsection 5.4) on a compact set. In particular, $\bar\alpha_L$ in (5.3) is uniquely determined and belongs to the set in (1.29).

5.2. **On Level 2.** In this subsection, we discuss the existence of solutions to (1.23). In what follows, let $\Omega \subset \mathbb{R}^n$ be an open set and $\omega : \Omega \to [0,\infty)$ a locally integrable function. Recalling the definition of $L^1_{\omega,\mathrm{loc}}(\Omega)$ and $\|\cdot\|_{L^1_\omega(\Omega)}$ in Subsection 3.2, as well as (5.2), we define the space $BD_\omega(\Omega)$ of ω-weighted BD functions in Ω by

$$BD_\omega(\Omega) := \big\{v \in L^1_\omega(\Omega;\mathbb{R}^n) : \mathcal{V}_\omega(\mathcal{E}v,\Omega) < \infty\big\},$$

and we endow it with the norm

$$\|v\|_{BD_\omega(\Omega)} := \|v\|_{L^1_\omega(\Omega;\mathbb{R}^n)} + \mathcal{V}_\omega(\mathcal{E}v,\Omega).$$

Note that BD_ω with $\omega \equiv 1$ coincides with the classical space of functions with bounded deformation, c.f. [53] for instance. The instrumental properties of BD_ω for our analysis are collected in the ensuing result.

Theorem 5.3. *Let $\Omega \subset \mathbb{R}^n$ be an open set and $\omega : \Omega \to [0,\infty)$ a locally integrable function. Then, the following statements hold:*

- *(i) If $\inf_\Omega \omega > 0$, then the map $v \mapsto \mathcal{V}_\omega(\mathcal{E}v,\Omega)$ is lower-semicontinuous with respect to the (strong) convergence in $L^1_{\omega,\mathrm{loc}}(\Omega;\mathbb{R}^n)$.*
- *(ii) Given $v \in L^1_{\omega,\mathrm{loc}}(\Omega;\mathbb{R}^n)$, we have $\mathcal{V}_\omega(\mathcal{E}v,\Omega) = \mathcal{V}_{\omega^{sc^-}}(\mathcal{E}v,\Omega)$, where ω^{sc^-} denotes the lower-semicontinuous envelope of ω.*
- *(iii) Assume ω is lower-semicontinuous and strictly positive. Then, $v \in L^1_{\mathrm{loc}}(\Omega;\mathbb{R}^n)$ and $\mathcal{V}_\omega(\mathcal{E}v,\Omega) < \infty$ if and only if $v \in BD_{\mathrm{loc}}(\Omega)$ and $\omega \in L^1(\Omega;|\mathcal{E}v|)$. If any of these two equivalent conditions hold, we have*

$$\mathcal{V}_\omega(\mathcal{E}v,B) = \int_B \omega(x)\,\mathrm{d}|\mathcal{E}v|(x)$$

for every Borel set $B \subset \Omega$.

(iv) *If $\omega \in L^\infty_{\rm loc}(\Omega)$ is lower-semicontinuous and strictly positive, then bounded sequences in $BD_\omega(\Omega)$ are precompact in the strong $L^1_{\omega,\rm loc}$-topology.*

Proof. Accounting for the fact that test functions here take values in $\mathbb{R}^{n\times n}_{\rm sym}$, the proof of (i), (ii), and (iii) may be obtained by mimicking that of [13, Proposition 1.3.1], [13, Proposition 2.1.1], and [13, Theorem 2.1.5], respectively.

To prove (iv), we observe that for each compact set $K \subset \Omega$, there exists a positive constant c_K such that $0< \frac{1}{c_K} \leqslant \omega \leqslant c_K$ in K because $\omega \in L^\infty_{\rm loc}(\Omega)$ and strictly positive lower-semicontinuous functions are locally bounded away from zero. Then, using (iii), we have for every $v \in BD_\omega(\Omega)$ that

$$\mathscr{V}_\omega(\mathcal{E}v, K) = \int_K \omega(x)\, \mathrm{d}|\mathcal{E}v|(x) \begin{cases} \geqslant \frac{1}{c_K}|\mathcal{E}v|(K), \\ \leqslant c_K|\mathcal{E}v|(K), \end{cases}$$

$$\|v\|_{L^1_\omega(K;\mathbb{R}^n)} = \int_\Omega |v(x)|\,\omega(x)\,\mathrm{d}x \begin{cases} \geqslant \frac{1}{c_K}\|v\|_{L^1(K;\mathbb{R}^n)}, \\ \leqslant c_K\|v\|_{L^1(K;\mathbb{R}^n)}. \end{cases}$$

The preceding estimates and the compact embedding of $BD(K)$ into $L^1(K;\mathbb{R}^n)$ (cf. [53]) yield (iv). □

Remark 5.4. If $\omega : \Omega \to (0,\infty)$ is a lower-semicontinuous function satisfying $0 < c \leqslant \inf_\Omega \omega \leqslant \sup_\Omega \omega \leqslant c^{-1}$ for some positive constant c, then the arguments in the preceding proof show that Theorem 5.3 (iv) holds globally in Ω. In other words, bounded sequences in $BD_\omega(\Omega)$ are precompact in the strong $L^1_\omega(\Omega;\mathbb{R}^n)$-topology.

Remark 5.5. Differently from the weighted-TV case (cf. Theorem 3.1), we need the weights ω in Theorem 5.3 to be bounded from below away from zero for item (i) to hold. This is because one cannot resort to arguments based on coarea formulas in the symmetrized gradient case, which prevents us to adapt the arguments in [13, Remark 1.3.2 and Theorem 3.1.13] to this framework.

The next result collects some basic properties of the quantity $\mathscr{V}_\omega(Du - v, \Omega)$ given by (5.1).

Theorem 5.6. *Let $\Omega \subset \mathbb{R}^n$ be an open set and $\omega : \Omega \to [0,\infty)$ a locally integrable function. Let $u \in BV_\omega(\Omega)$. Then, the following statements hold:*

(i) *The map $v \to \mathscr{V}_\omega(Du - v, \Omega)$ is lower semicontinuous with respect to the strong convergence in $L^1_{\omega,\rm loc}(\Omega;\mathbb{R}^n)$.*

(ii) *Given $v \in L^1_{\omega,\rm loc}(\Omega;\mathbb{R}^n)$, we have $\mathscr{V}_\omega(Du - v, \Omega) = \mathscr{V}_{\omega^{sc^-}}(Du - v, \Omega)$, where ω^{sc^-} denotes the lower-semicontinuous envelope of ω.*

(iii) *If $v \in L^1_{\omega,\rm loc}(\Omega;\mathbb{R}^n)$ and $\omega \in L^1(\Omega; |Du - v|)$ is lower-semicontinuous and strictly positive, then*

$$\mathscr{V}_\omega(Du - v, B) = \int_B \omega(x)\, \mathrm{d}|Du - v|(x) \tag{5.8}$$

for every Borel set $B \subset \Omega$.

Proof. To prove (i), let $(v_k)_{k\in\mathbb{N}} \subset L^1_{\omega,\rm loc}(\Omega;\mathbb{R}^n)$ be such that $v_k \to v$ strongly in $L^1_{\omega,\rm loc}(\Omega;\mathbb{R}^n)$. Then, by Definition 5.1,

$$\mathscr{V}_\omega(Du - v_k, \Omega) \geqslant \int_\Omega \big(u \operatorname{div}\varphi + v_k\cdot\varphi\big)\,\mathrm{d}x$$

for every $\varphi \in \mathrm{Lip}_c(\Omega;\mathbb{R}^n)$ with $|\varphi| \leqslant \omega$ in Ω. Moreover, for all such φ,

$$\int_\Omega |v_k - v||\varphi|\,\mathrm{d}x \leqslant \int_{\operatorname{supp}\varphi} |v_k - v|\,\omega\,\mathrm{d}x \to 0 \text{ as } k \to +\infty.$$

Hence,

$$\liminf_{k\to+\infty} \mathscr{V}_\omega(Du - v_k, \Omega) \geqslant \int_\Omega \big(u \operatorname{div}\varphi + v\cdot\varphi\big)\,\mathrm{d}x,$$

from which the conclusion follows by taking the supremum over all test functions $\varphi \in \mathrm{Lip}_c(\Omega;\mathbb{R}^n)$ with $|\varphi| \leqslant \omega$ in Ω.

The proof of (ii) follows by Definition 5.1, observing that every map $\varphi \in \mathrm{Lip}_c(\Omega;\mathbb{R}^n)$ with $|\varphi| \leqslant \omega$ in Ω also satisfies $|\varphi| \leqslant (\omega^{\rm sc})^-$ in Ω.

As we discuss next, the proof of *(iii)* is an adaptation of [13, Theorem 2.1.5]. In fact, because $u \in BV_\omega(\Omega)$ and strictly positive lower-semicontinuous functions are locally bounded away from zero, we have $u \in BV_{\mathrm{loc}}(\Omega)$. Then, for every $\varphi \in \mathrm{Lip}_c(\Omega;\mathbb{R}^n)$ with $|\varphi| \leqslant \omega$ in Ω, we have that

$$\int_\Omega \big(u \operatorname{div} \varphi + v\cdot\varphi\big)\,\mathrm{d}x \leqslant \int_\Omega \omega\,\mathrm{d}|Du - v|;$$

hence, $\mathscr{V}_\omega(Du - v, \Omega) \leqslant \int_\Omega \omega\,\mathrm{d}|Du-v|$. Conversely, since $\omega \in L^1(\Omega;|Du-v|)$, we infer that

$$\int_\Omega \omega\,\mathrm{d}|Du-v| = |\omega(Du-v)|(\Omega) = \sup\left\{\int_\Omega \omega\,\psi\cdot\,\mathrm{d}(Du-v) : \psi \in \mathrm{Lip}_c(\Omega;\mathbb{R}^n),\ |\psi|\leqslant 1\right\}. \tag{5.9}$$

Let $(\omega_k)_{k\in\mathbb{N}}$ be an increasing sequence of k-Lipschitz functions converging to ω in Ω as in Example 4.1 (see also [13, Theorem 2.1.2]). Then, for every $\psi \in \mathrm{Lip}_c(\Omega;\mathbb{R}^n)$ with $|\psi|\leqslant 1$ in Ω, we have $\omega_k\,\psi \in \mathrm{Lip}_c(\Omega;\mathbb{R}^n)$ with $|\omega_k\,\psi| \leqslant \omega_k \leqslant \omega$ in Ω; thus, using the Lebesgue dominated convergence theorem and recalling (5.1), we find that

$$\int_\Omega \omega\,\psi\cdot\,\mathrm{d}(Du-v) = \lim_{k\to\infty}\int_\Omega \omega_k\,\psi\cdot\,\mathrm{d}(Du-v) = -\lim_{k\to\infty}\int_\Omega \big(u\operatorname{div}(\omega_k\,\psi) + \omega_k\,\psi\cdot v\big)\,\mathrm{d}x \leqslant \mathscr{V}_\omega(Du-v,\Omega).$$

From this estimate and (5.9), we deduce that $\int_\Omega \omega\,\mathrm{d}|Du-v| \leqslant \mathscr{V}_\omega(Du-v,\Omega)$, which concludes the proof of (5.8) when $B = \Omega$. The proof that this identity holds for every Borel set $B \subset \Omega$ can be done exactly as in [13, Theorem 2.1.5]. □

We proceed by showing that the infimum in

$$TGV_{\omega_0,\omega_1}(u,Q) := \inf_{v\in BD_{\omega_1}(Q)} \big\{\mathscr{V}_{\omega_0}(Du-v,Q) + \mathscr{V}_{\omega_1}(\mathcal{E}v,Q)\big\}, \tag{5.10}$$

where $\omega_0,\,\omega_1 : Q \to (0,+\infty)$ are bounded functions and $u \in L^1_{\omega_0}(Q)$, is actually a minimum, and that the contributions due to $\mathscr{V}_{\omega_0}$ and $\mathscr{V}_{\omega_1}$ can be expressed in a simplified way in terms of the lower semicontinuous envelopes of the weights ω_0 and ω_1. We begin with a technical lemma.

Lemma 5.7. *Let $c_0 > 0$ be a positive constant. For $i \in \{0,1\}$, let $\omega_i : Q \to (0,+\infty)$ be such that $c_0 < \inf_Q \omega_i < \sup_Q \omega_i < \frac{1}{c_0}$, and let $u \in L^1_{\omega_0,\mathrm{loc}}(Q)$. Then, for every $v \in L^1_{\omega_1}(Q;\mathbb{R}^2)$, we have*

$$\|v\|_{L^1_{\omega_1}(Q;\mathbb{R}^n)} \leqslant \frac{1}{c_0^2}\big(\mathscr{V}_{\omega_0}(Du-v,Q) + TV_{\omega_0}(u,Q)\big). \tag{5.11}$$

Proof. Fix $v \in L^1_{\omega_1}(Q;\mathbb{R}^2)$. Note that the uniform bounds on ω_1 yield

$$c_0\int_Q |v(x)|\,\mathrm{d}x \leqslant \int_Q \omega_1(x)|v(x)|\,\mathrm{d}x = \|v\|_{L^1_{\omega_1}(Q;\mathbb{R}^n)} \leqslant \frac{1}{c_0}\int_Q |v(x)|\,\mathrm{d}x. \tag{5.12}$$

In particular, $v \in L^1(Q;\mathbb{R}^2)$; thus,

$$\begin{aligned} c_0\int_Q |v(x)|\,\mathrm{d}x &= c_0 \sup\left\{\int_Q \psi(x)\cdot v(x)\,\mathrm{d}x : \psi \in \mathrm{Lip}_c(Q;\mathbb{R}^2),\ \|\psi\|_{L^\infty(Q;\mathbb{R}^2)} \leqslant 1\right\} \\ &= \sup\left\{\int_Q \tilde\psi(x)\cdot v(x)\,\mathrm{d}x : \tilde\psi \in \mathrm{Lip}_c(Q;\mathbb{R}^2),\ \|\tilde\psi\|_{L^\infty(Q;\mathbb{R}^2)} \leqslant c_0\right\} \\ &\leqslant \sup\left\{\int_Q \varphi(x)\cdot v(x)\,\mathrm{d}x : \varphi \in \mathrm{Lip}_c(Q;\mathbb{R}^2),\ |\varphi| \leqslant \omega_0\right\} \\ &\leqslant \mathscr{V}_{\omega_0}(Du-v,Q) + TV_{\omega_0}(u,Q), \end{aligned} \tag{5.13}$$

where we used Definition 5.1 together with the subadditivity of the supremum in the last estimate, and the bound $c_0 \leqslant \inf_Q \omega_0$ in the preceding one. We then obtain (5.11) by combining (5.12) and (5.13). □

Under the same assumptions of Lemma 5.7, the infimum problem in (1.24) is actually a minimum.

Lemma 5.8. *Let $c_0 > 0$ be a positive constant. For $i \in \{0,1\}$, let $\omega_i : Q \to (0,+\infty)$ be such that $c_0 < \inf_Q \omega_i < \sup_Q \omega_i < \frac{1}{c_0}$, and let $u \in L^1_{\omega_0}(Q)$. Then, there exists $u^* \in BD_{\omega_1}(Q)$ such that*

$$TGV_{\omega_0,\omega_1}(u,Q) = \mathscr{V}_{\omega_0}(Du-u^*,Q) + \mathscr{V}_{\omega_1}(\mathcal{E}u^*,Q). \tag{5.14}$$

Proof. We claim that $TGV_{\omega_0,\omega_1}(u,Q)$ is finite if and only if $u \in BV_{\omega_0}(Q)$. In fact, choosing $v = 0$ as a competitor in (5.10), we infer that $TGV_{\omega_0,\omega_1}(u,Q) \leqslant TV_{\omega_0}(u,Q)$. On the other hand, recalling (3.3), we have for any $v \in BD_{\omega_1}(Q)$ that

$$\begin{aligned} TV_{\omega_0}(u,Q) &= \sup\left\{\int_Q (u \operatorname{div}\varphi + v\cdot\varphi - v\cdot\varphi)\,\mathrm{d}x :\ \varphi \in \mathrm{Lip}_c(\Omega;\mathbb{R}^2),\ |\varphi| \leqslant \omega_0\right\} \\ &\leqslant \mathscr{V}_{\omega_0}(Du - v, Q) + \|v\|_{L^1_{\omega_0}(Q;\mathbb{R}^2)} \\ &\leqslant \mathscr{V}_{\omega_0}(Du - v, Q) + \frac{1}{c_0^2}\|v\|_{BD_{\omega_1}(Q)}, \end{aligned}$$

where we used the subadditivity of the supremum combined with Definition 5.1 in the first inequality, and the bounds on the two weights in the second inequality. Thus, $TV_{\omega_0}(u,Q) \leqslant \max\{1, c_0^{-2}\}\, TGV_{\omega_0,\omega_1}(u,Q)$, which concludes the proof of the claim.

To show (5.14), we may assume without loss of generality that $TGV_{\omega_0,\omega_1}(u,Q) < \infty$, in which case $u \in BV_{\omega_0}(Q)$. Moreover, we may find a sequence $(v_n) \subset BD_{\omega_1}(Q)$ such that

$$TGV_{\omega_0,\omega_1}(u,Q) = \lim_{n\to+\infty}\left\{\mathscr{V}_{\omega_0}(Du - v_n, Q) + \mathscr{V}_{\omega_1}(\mathcal{E}v_n, Q)\right\} \leqslant C \tag{5.15}$$

for some positive constant C. From Lemma 5.7 and (5.15) we infer that $\sup_{n\in\mathbb{N}} \|v_n\|_{BD_{\omega_1}(Q)} < +\infty$. Using the uniform bounds on ω_1, which are inherited by its lower semicontinuous envelope $(\omega_1)^{\mathrm{sc}^-}$, and Theorem 5.3 (ii), also

$$\sup_{n\in\mathbb{N}} \|v_n\|_{BD_{(\omega_1)^{\mathrm{sc}^-}}(Q)} < +\infty.$$

Moreover, by Theorem 5.3 (i), (ii), and (iv) (also see Remark 5.4), there exists $u^* \in BD_{\omega_1}(Q) \cap BD_{(\omega_1)^{\mathrm{sc}^-}}(Q)$ such that

$$\begin{aligned} &v_n \to u^* \quad \text{strongly in } L^1_{(\omega_1)^{\mathrm{sc}^-}}(Q;\mathbb{R}^2), \\ &\mathscr{V}_{\omega_1}(\mathcal{E}u^*, \Omega) = \mathscr{V}_{(\omega_1)^{\mathrm{sc}^-}}(\mathcal{E}u^*, \Omega) \leqslant \liminf_{n\to+\infty} \mathscr{V}_{(\omega_1)^{\mathrm{sc}^-}}(\mathcal{E}v_n, \Omega) = \liminf_{n\to+\infty} \mathscr{V}_{\omega_1}(\mathcal{E}v_n, \Omega). \end{aligned} \tag{5.16}$$

Using the uniform bounds on both weights once more, we also have $v_n \to u^*$ strongly in $L^1_{\omega_0}(Q;\mathbb{R}^2)$. The minimality of u^* is then a direct consequence of Theorem 5.6 (i), (5.16), and (5.15). ☐

The next result provides a characterization of the infimum problem in Level 2 of our learning scheme.

Proposition 5.9. *Let $\phi \in L^2(Q)$, and let $c_0 > 0$ be a positive constant. For $i \in \{0,1\}$, let $\omega_i : Q \to [0,+\infty)$ be such that $c_0 < \inf_Q \omega_i < \sup_Q \omega_i < \frac{1}{c_0}$. Then, there exists a unique $\bar u \in BV_{\omega_0}(Q)$ such that*

$$\int_Q |\phi - \bar u|\ \mathrm{d}x + TGV_{\omega_0,\omega_1}(\bar u, Q) = \min_{u\in BV_{\omega_0}(Q)}\left\{\int_Q |\phi - u|^2\ \mathrm{d}x + TGV_{\omega_0,\omega_1}(u,Q)\right\}.$$

Moreover, denoting by $(\omega_i)^{\mathrm{sc}^-}$ the lower semicontinuous envelope of ω_i, $i \in \{0,1\}$, we have $\bar u \in BV(Q) \cap BV_{(\omega_0)^{\mathrm{sc}^-}}(Q)$, and

$$TGV_{\omega_0,\omega_1}(\bar u) = \int_Q (\omega_0)^{\mathrm{sc}^-}\ \mathrm{d}|D\bar u - u^*| + \int_Q (\omega_1)^{\mathrm{sc}^-}\ \mathrm{d}|\mathcal{E}u^*|,$$

where $u^* \in BD_{\omega_1}(Q) \cap BD_{(\omega_1)^{\mathrm{sc}^-}}(Q)$ is a minimizer of (5.10) associated to $\bar u$.

Proof. For $u \in BV_{\omega_0}(Q)$, we define

$$H[u] := \int_Q |\phi - u|^2\ \mathrm{d}x + TGV_{\omega_0,\omega_1}(u,Q),$$

and we set

$$\mu := \inf_{u\in BV_{\omega_0}(Q)} H[u].$$

We have $0 \leqslant \mu \leqslant F[0] = \|\phi\|^2_{L^2(Q)}$, and we may take a sequence $(u_n)_{n\in\mathbb{N}} \subset BV_{\omega_0}(Q)$ such that

$$\mu = \lim_{n\to+\infty} H[u_n].$$

Moreover, the boundedness assumptions on the weights ω_i, $i \in \{0,1\}$, yield for all $x \in Q$ that

$$c_0 \leqslant (\omega_i)^{\mathrm{sc}^-}(x) \leqslant \frac{1}{c_0}.$$

Thus, by Lemma 5.8 and Theorems 5.3 and 5.6, we find for all $n \in \mathbb{N}$ large enough that

$$\begin{aligned}\mu + 1 \geqslant H[u_n] &= \int_Q |\phi - u_n|^2 \,\mathrm{d}x + \mathscr{V}_{\omega_0}(Du_n - u_n^*, Q) + \mathscr{V}_{\omega_1}(\mathcal{E}u_n^*, Q)\\ &= \int_Q |\phi - u_n|^2 \,\mathrm{d}x + \mathscr{V}_{(\omega_0)^{\mathrm{sc}^-}}(Du_n - u_n^*, Q) + \mathscr{V}_{(\omega_1)^{\mathrm{sc}^-}}(\mathcal{E}u_n^*, Q)\\ &= \int_Q |\phi - u_n|^2 \,\mathrm{d}x + \int_Q (\omega_0)^{\mathrm{sc}^-}\,\mathrm{d}|Du_n - u_n^*| + \int_Q (\omega_1)^{\mathrm{sc}^-}\,\mathrm{d}|\mathcal{E}u_n^*|\\ &\geqslant \int_Q |\phi - u_n|^2 \,\mathrm{d}x + c_0|Du_n - u_n^*|(Q) + c_0|\mathcal{E}u_n^*|(Q).\end{aligned}$$

An argument by contradiction as in the classical TGV case and variants thereof (see, e.g., [26, Proposition 5.3]) yields that the sequences $(u_n^*)_{n\in\mathbb{N}}$ and $(u_n)_{n\in\mathbb{N}}$ are uniformly bounded in $BD(Q)$ and $BV(Q)$, respectively. Thus, there exist $\bar{u}^* \in BD(Q)$ and $u \in BV(Q)$ such that, up to extracting a not relabelled subsequence,

$$\begin{aligned}u_n &\overset{*}{\rightharpoonup} \bar{u} \quad \text{weakly* in } BV(Q),\\ u_n^* &\overset{*}{\rightharpoonup} \bar{u}^* \quad \text{weakly* in } BD(Q).\end{aligned}$$

By the bounds on the weights, and their lower-semicontinuous envelopes, and Theorems 5.3 and 5.6, we deduce that $\bar{u} \in BV_{(\omega_0)^{\mathrm{sc}^-}}(Q) \cap BV_{\omega_0}(Q) \cap BV(Q)$ and $\bar{u}^* \in BD_{(\omega_1)^{\mathrm{sc}^-}}(Q) \cap BD_{\omega_1}(Q) \cap BD(Q)$, with

$$\begin{aligned}\mu \leqslant H[\bar{u}] &\leqslant \int_Q |\phi - \bar{u}|^2 \,\mathrm{d}x + \mathscr{V}_{(\omega_0)^{\mathrm{sc}^-}}(D\bar{u} - \bar{u}^*, Q) + \mathscr{V}_{(\omega_1)^{\mathrm{sc}^-}}(\mathcal{E}\bar{u}^*, Q)\\ &\leqslant \lim_{n\to+\infty} H[u_n] = \mu.\end{aligned} \tag{5.17}$$

Because of the strict convexity of the L^2-norm, we infer the uniqueness of $\bar{u}$. Finally, by (5.17),

$$TGV_{\omega_0,\omega_1}(\bar{u}, Q) = \mathscr{V}_{(\omega_0)^{\mathrm{sc}^-}}(D\bar{u} - \bar{u}^*, Q) + \mathscr{V}_{(\omega_1)^{\mathrm{sc}^-}}(\mathcal{E}\bar{u}^*, Q).$$

The last part of the statement is then a consequence of Theorems 5.3 and 5.6. ☐

5.3. **On Level** 1. As we address next, and similarly to the $(\mathcal{LS})_{TV_\omega}$ case, the box constraint provides a stopping criterion for the TGV-learning scheme.

To proceed as in Theorem 3.6, we need an analog to Proposition 3.5, which we now prove. Recalling that L represents a cell in a dyadic partition of Q, we will use the Sobolev inequality in $BV(L)$ yielding for every $u \in BV(L)$ that

$$\|u - c_u\|_{L^2(L)} \leqslant C_Q^{BV}|Du|(L), \tag{5.18}$$

where $c_u \in \mathbb{R}$ is the average of u in L, and the constant C_Q^{BV} depends only on the shape of Q because of scale invariance of the embedding BV in L^2 in dimension $d = 2$. Moreover, we also have for any $w \in BD(L)$ that

$$\|w - R_{M_w} - v_w\|_{L^2(L)} \leqslant C_Q^{BD}|\mathcal{E}w|(L), \tag{5.19}$$

where $v_w \in \mathbb{R}^2$ and R_{M_w} denotes the function defined for $M_w \in \mathbb{R}^{2\times 2}$ by $R_{M_w}(x) = M_w x$.

Lemma 5.10. *Let $L \subset Q$ be a dyadic square. Then, there is a constant $C_Q^{rot} > 0$ such that for every $u \in BV(L)$ and for every antisymmetric matrix $M \in \mathbb{R}^{2\times 2}$ with $M + M^\top = 0$, we have*

$$C_Q^{rot}|Du|(L) \leqslant |Du - R_M|(L). \tag{5.20}$$

Proof. Suppose that (5.20) does not hold; then, we may find functions $u_n \in BV(L)$ with $|Du_n|(L) = 1$ and antisymmetric matrices $M_n \in \mathbb{R}^{2\times 2}$ for which

$$\frac{1}{n} = \frac{1}{n}|Du_n|(L) > |Du_n - R_{M_n}|(L). \tag{5.21}$$

Then, in particular, $\|R_{M_n}\|_{L^1(L)} \leqslant 2$; consequently, since $\{R_M \,|\, M + M^\top = 0\}$ is a finite-dimensional set, we can assume that $R_{M_n} \to R_{M_\infty}$ for some antisymmetric matrix M_∞, up to taking a not relabelled subsequence.

On the other hand, recalling (5.18), there are constants $c_n \in \mathbb{R}$ satisfying

$$\|u_n - c_n\|_{L^2(L)} \leqslant C_Q^{BV}|Du_n|(L);$$

thus, up to taking a not relabelled further subsequence, we have that $u_n - c_n \overset{*}{\rightharpoonup} u_\infty \in BV(L)$ for some $u_\infty \in BV(L)$. Using (5.21) once more, we must have $Du_\infty = R_{M_\infty}$. At this point, we can distinguish two cases, $M_\infty = 0$ or $M_\infty \neq 0$.

If $M_\infty = 0$, then

$$\frac{1}{n} = \frac{1}{n}|Du_n|(L) > |Du_n - R_{M_n}|(L) \to 1,$$

which cannot be.

If $M_\infty \neq 0$, then, using the antisymmetry of $DR_{M_\infty} = M_\infty$, we again arrive at a contradiction, since

$$\operatorname{curl} Du_\infty = 0 \quad \text{but} \quad |\operatorname{curl} R_{M_\infty}| = \sqrt{2}|M_\infty| > 0.$$

To see that the last equality holds, just notice that in the two dimensional case under consideration we must have

$$M_\infty = \begin{pmatrix} 0 & a \\ -a & 0 \end{pmatrix} \text{ for some } a \neq 0, \text{ which implies } \operatorname{curl} R_{M_\infty} = -2a.$$

Thus, we have proved that there is a constant C_L, possibly depending on L, such that

$$C_L|Du|(L) \leqslant |Du - R_M|(L) \text{ for all } M \text{ with } M + M^\top = 0.$$

To see that C_L is independent of the size of L, we just notice that this inequality holds for all M and that upon rescaling $x \mapsto rx$ it is enough to replace M by M/r to maintain the inequality. ☐

The next proposition guarantees that if a dyadic square $L \subset Q$ is small enough, then a solution u_{α_0,α_1} of Level 3 of our TGV learning scheme in (1.20) is affine for every $(\alpha_0, \alpha_1) \in \left[c_0, \frac{1}{c_0}\right] \times \left[c_1, \frac{1}{c_1}\right]$. Let us remark that a related result is contained in [48, Proposition 6], which we make quantitative and with a scaling that enables us to draw conclusions on the cell size.

Proposition 5.11. *Fix c_0, $c_1 > 0$ and $L \subset Q$ a dyadic square. Let $\bar{\alpha}_L$ be the optimal parameter given by (5.3), where u_α is defined by (1.22) and (1.19) (with Q replaced by L), and let C_Q^{BV}, C_Q^{BD}, and C_Q^{rot} be the constants in (5.18), (5.19), and (5.20), respectively. If*

$$\|u_\eta\|_{L^2(L)} < \min\left(c_0, \frac{c_1}{C_Q^{BD}|L|^{1/2}}\right)\frac{C_Q^{rot}}{C_Q^{BV}}, \tag{5.22}$$

then $\bar{\alpha}_L := (\overline{\alpha}_0, \overline{\alpha}_1) = (c_0, c_1)$ and $u_{\bar{\alpha}_L} := u_{\overline{\alpha}_0,\overline{\alpha}_1}$ is affine on L, with $u_{\overline{\alpha}_0,\overline{\alpha}_1} = \langle u_\eta \rangle_L$.

Proof. To simplify the notation in the proof, we omit the dependence of TGV_{α_0,α_1} on L by writing $TGV_{\alpha_0,\alpha_1}(\cdot)$ in place of $TGV_{\alpha_0,\alpha_1}(\cdot, L)$.

Fix $(\alpha_0, \alpha_1) \in \left[c_0, \frac{1}{c_0}\right] \times \left[c_1, \frac{1}{c_1}\right]$. The optimality condition for (1.22) reads as

$$u_\eta - u_{\alpha_0,\alpha_1} \in \partial TGV_{\alpha_0,\alpha_1}(u_{\alpha_0,\alpha_1}).$$

Since TGV_{α_0,α_1} is positively one-homogeneous, we have that

$$z \in \partial TGV_{\alpha_0,\alpha_1}(u_{\alpha_0,\alpha_1}) \text{ if and only if } z \in \partial TGV_{\alpha_0,\alpha_1}(0) \text{ and } \int_L z\, u_{\alpha_0,\alpha_1}\, \mathrm{d}x = TGV_{\alpha_0,\alpha_1}(u_{\alpha_0,\alpha_1}).$$

Furthermore, by the definition of subgradient,

$$z \in \partial TGV_{\alpha_0,\alpha_1}(0) \text{ if and only if } \int_L z\, \overline{u}\, \mathrm{d}x \leqslant TGV_{\alpha_0,\alpha_1}(\overline{u}) \text{ for all } \overline{u} \in L^2(L).$$

Now, given $v \in \mathbb{R}^2$ and $c \in \mathbb{R}$, we denote by $A_{v,c}$ the affine $A_{v,c}(x) = v \cdot x + c$. Because

$$TGV_{\alpha_0,\alpha_1}(A_{v,c}) = 0, \tag{5.23}$$

we deduce from the above with $z = u_\eta - u_{\alpha_0,\alpha_1}$ and $\overline{u} = \pm A_{v,c}$ that $\int_L (u_\eta - u_{\alpha_0,\alpha_1})A_{v,c}\, \mathrm{d}x = 0$ for any $v \in \mathbb{R}^2$ and $c \in \mathbb{R}$; moreover,

$$\begin{aligned} TGV_{\alpha_0,\alpha_1}(u_{\alpha_0,\alpha_1}) &= \int_L (u_\eta - u_{\alpha_0,\alpha_1})u_{\alpha_0,\alpha_1}\, \mathrm{d}x = \int_L (u_\eta - u_{\alpha_0,\alpha_1})(u_{\alpha_0,\alpha_1} - A_{v,c})\, \mathrm{d}x \\ &\leqslant \|u_\eta - u_{\alpha_0,\alpha_1}\|_{L^2(L)} \|u_{\alpha_0,\alpha_1} - A_{v,c}\|_{L^2(L)}. \end{aligned}$$

Thus, taking the infimum over $v \in \mathbb{R}^2$ and $c \in \mathbb{R}$ and recalling (5.4), we conclude that

$$TGV_{\alpha_0,\alpha_1}(u_{\alpha_0,\alpha_1}) \leqslant \|u_\eta - u_{\alpha_0,\alpha_1}\|_{L^2(L)} \|u_{\alpha_0,\alpha_1} - \langle u_{\alpha_0,\alpha_1}\rangle_L\|_{L^2(L)}. \tag{5.24}$$

On the other hand, since the infimum in the definition of TGV_{α_0,α_1} is attained, there is a $w_u \in BD(L)$ for which

$$\begin{aligned} TGV_{\alpha_0,\alpha_1}(u_{\alpha_0,\alpha_1}) &= \inf_{w\in BD(L)} \Big\{\alpha_0|Du_{\alpha_0,\alpha_1} - w|(L) + \alpha_1|\mathcal{E}w|(L)\Big\} \\ &= \alpha_0|Du_{\alpha_0,\alpha_1} - w_u|(L) + \alpha_1|\mathcal{E}w_u|(L) \\ &\geqslant \alpha_0|Du_{\alpha_0,\alpha_1} - w_u|(L) + \frac{\alpha_1}{C_Q^{BD}}\|w_u - R_{M_{w_u}} - v_{w_u}\|_{L^2(L)}, \end{aligned}$$

where we have used the inequality (5.19) for some antisymmetric matrix $M_{w_u} \in \mathbb{R}^{2\times 2}$ and vector $v_{w_u} \in \mathbb{R}^2$. Setting $R_u := R_{M_{w_u}}$ and $v_u := v_{w_u}$, we get that

$$\begin{aligned} TGV_{\alpha_0,\alpha_1}(u_{\alpha_0,\alpha_1}) &\geqslant \alpha_0|Du_{\alpha_0,\alpha_1} - w_u|(L) + \frac{\alpha_1}{C_Q^{BD}}\|w_u - R_u - v_u\|_{L^2(L)} \\ &\geqslant \alpha_0|Du_{\alpha_0,\alpha_1} - w_u|(L) + \frac{\alpha_1}{C_Q^{BD}|L|^{1/2}}\|w_u - R_u - v_u\|_{L^1(L)} \\ &\geqslant \min\left(c_0, \frac{c_1}{C_Q^{BD}|L|^{1/2}}\right)\Big[|Du_{\alpha_0,\alpha_1} - w_u|(L) + \|w_u - R_u - v_u\|_{L^1(L)}\Big] \\ &\geqslant \min\left(c_0, \frac{c_1}{C_Q^{BD}|L|^{1/2}}\right)|Du_{\alpha_0,\alpha_1} - R_u - v_u|(L) \\ &= \min\left(c_0, \frac{c_1}{C_Q^{BD}|L|^{1/2}}\right)|D(u_{\alpha_0,\alpha_1} - A_{v_u,0}) - R_u|(L). \end{aligned}$$

Now, we can apply Lemma 5.10 to $u_{\alpha_0,\alpha_1} - A_{v_u,0}$ and the Sobolev inequality (5.18) to obtain for some $c_u \in \mathbb{R}$ that

$$\begin{aligned} TGV_{\alpha_0,\alpha_1}(u_{\alpha_0,\alpha_1}) &\geqslant \min\left(c_0, \frac{c_1}{C_Q^{BD}|L|^{1/2}}\right) C_Q^{rot}|D(u_{\alpha_0,\alpha_1} - A_{v_u,0})|(L) \\ &\geqslant \min\left(c_0, \frac{c_1}{C_Q^{BD}|L|^{1/2}}\right) \frac{C_Q^{rot}}{C_Q^{BV}}\|u_{\alpha_0,\alpha_1} - A_{v_u,c_u}\|_{L^2(L)} \\ &\geqslant \min\left(c_0, \frac{c_1}{C_Q^{BD}|L|^{1/2}}\right) \frac{C_Q^{rot}}{C_Q^{BV}}\|u_{\alpha_0,\alpha_1} - \langle u_{\alpha_0,\alpha_1}\rangle_L\|_{L^2(L)}, \end{aligned} \tag{5.25}$$

where we used (5.4) once more. Then, if u_{α_0,α_1} was not affine, then $\|u_{\alpha_0,\alpha_1} - \langle u_{\alpha_0,\alpha_1}\rangle_L\|_{L^2(L)} > 0$, so we could combine (5.25) with the upper bound (5.24) and minimality of u_{α_0,α_1} in (1.22) to obtain

$$\min\left(c_0, \frac{c_1}{C_Q^{BD}|L|^{1/2}}\right) \frac{C_Q^{rot}}{C_Q^{BV}} \leqslant \|u_\eta - u_{\alpha_0,\alpha_1}\|_{L^2(L)} \leqslant \|u_\eta\|_{L^2(L)},$$

which contradicts (5.22). Thus, u_{α_0,α_1} must be affine.

Finally, using (5.23), (5.4), and $\langle u_\eta\rangle_L$ as a competitor in (1.22), we conclude that $u_{\alpha_0,\alpha_1} = \langle u_{\alpha_0,\alpha_1}\rangle_L = \langle u_\eta\rangle_L$. Hence, $\bar\alpha_L = (c_0, c_1)$, and this concludes the proof. □

Owing to Proposition 5.11, we are now in a position to reduce the minimum problem in Level 1 of our training scheme to a minimization over a finite set of admissible partitions.

Theorem 5.12. *Consider the learning scheme $(\mathcal{LS})_{TGV_\omega}$ in (1.20) with (1.21) restricted by (1.29) (see (5.3)). Then, there exist $\kappa \in \mathbb{N}$ and $\mathcal{L}_1, ..., \mathcal{L}_\kappa \in \mathcal{P}$ such that*

$$\operatorname{argmin}\left\{\int_Q |u_c - u_{\mathcal{L}}|^2\,dx : \mathcal{L} \in \mathcal{P}\right\} = \operatorname{argmin}\left\{\int_Q |u_c - u_{\mathcal{L}_i}|^2\,dx : i \in \{1, ..., \kappa\}\right\}.$$

Proof. The proof is analogous to that of Theorem 3.6, so we only provide a sketch of the argument. The only difference here is that instead of being a constant, the solution u_α of Level 1 is affine for any $\alpha := (\alpha_0, \alpha_1) \in \left[c_0, \frac{1}{c_0}\right] \times \left[c_1, \frac{1}{c_1}\right]$ on squares L on which (5.22) holds, due to Proposition 5.11. Moreover, $TGV_{\alpha_0,\alpha_1}(u_\alpha, L) = 0$ and, recalling (5.3), the optimal parameter given by (5.3) is $\bar\alpha_L = (c_0, c_1)$. As in the proof of Theorem 3.6, this observation allows us to replace any partition $\mathcal{L}^*$ containing such small dyadic squares with another partition $\overline{\mathcal{L}}^*$ whose dyadic squares have all side length above the threshold provided by (5.22) without affecting the minimizer of Level 2. We refer to Figure 1 for a graphical idea of the argument and to Theorem 3.6 for the details of the proof. □

We conclude this section by proving existence of an optimal solution to the learning scheme $(\mathcal{LS})_{TGV_\omega}$.

Proof of Theorem 1.7. The result follows directly by combining the analysis in Subsection 5.1.3, Proposition 5.9, and Theorem 5.12. □

5.4. **Stopping criteria and box constraint for TGV.** In this subsection, we prove a TGV-counterpart to Theorem 3.8. Our result reads as follows.

Theorem 5.13. *Let $\Omega \subset \mathbb{R}^2$ be a bounded, Lipschitz domain and, for each $\alpha \in (0, +\infty)^2$, let $u_\alpha \in BV(\Omega)$ be given by (1.22) with L replaced by Ω. Assume that the two following conditions on the training data hold:*

i) There exists $\hat{\alpha} \in (0, +\infty)^2$ such that $TGV_{\hat{\alpha}_0,\hat{\alpha}_1}(u_c, \Omega) < TGV_{\hat{\alpha}_0,\hat{\alpha}_1}(u_\eta, \Omega)$;
ii) $\|u_\eta - u_c\|^2_{L^2(\Omega)} < \|\langle u_\eta\rangle - u_c\|^2_{L^2(\Omega)}$.

Then, there exists

$$\alpha^*_\Omega \in (0, +\infty)^2 \cup \big(\{+\infty\} \times (0, +\infty)\big) \cup \big((0, +\infty) \times \{+\infty\}\big) \tag{5.26}$$

such that

$$J^{sc^-}(\alpha^*_\Omega) = \min_{\alpha \in [0,+\infty]^2} J^{sc^-}(\alpha), \tag{5.27}$$

where J^{sc^-} is the lower semicontinuous envelope on $[0, +\infty]^2$ (see (5.35) in Lemma 5.18 below) of the function $J : (0, +\infty)^2 \to [0, +\infty)$ defined by

$$J(\alpha) := \int_\Omega |u_c - u_\alpha|^2 \, \mathrm{d}x \quad \textit{for } \alpha \in (0, +\infty)^2. \tag{5.28}$$

Additionally, there exist positive constants, c_Ω and C_Ω, such that any minimizer, α^*_Ω, of J^{sc^-} over $[0, +\infty]^2$ satisfies $c_\Omega \leqslant \min\{(\alpha^*_\Omega)_0, (\alpha^*_\Omega)_1\} < C_\Omega \|u_\eta\|_{L^2(\Omega)}$.

In particular, if $\Omega = L$ with $L \subset Q$ is a dyadic square, then there exists a positive constant, c_L, such that any minimizer, α^*_L, of J^{sc^-} over $[0, +\infty]^2$ satisfies $c_L \leqslant \min\{(\alpha^*_L)_0, (\alpha^*_L)_1\} < C_Q \|u_\eta\|_{L^2(L)}$, where C_Q is a constant given by Proposition 5.11.

Owing to the orthogonality property (5.5), condition *ii)* in the statement of the theorem is equivalent to requiring that $\|u_c - \langle u_c\rangle - u_\eta + \langle u_\eta\rangle\|^2_{L^2(\Omega)} \leqslant \|u_c - \langle u_c\rangle\|^2_{L^2(\Omega)}$. In other words, *ii)* is satisfied provided that the perturbation which the noise causes on the non-affine portion of u_c is small in the L^2-sense compared to the original non-affine component of u_c. This is the case, for example, if $\eta = u_\eta - u_c$ and $\eta - \langle \eta \rangle$ has a small L^2-norm, regardless of the L^2-norm of $\langle \eta \rangle$.

We remark that the conclusion of the theorem in the general case is slightly weaker than the corresponding result for the TV-setting. Indeed, while we can show that both entries of optimal parameters must be uniformly far away from zero, we can only prove that their minimum is uniformly bounded from above but can't prevent that just one of the entries blows up to infinity. This is due to the fact that, without additional conditions, the maps u_α are not necessarily affine if just one of the entries of α becomes infinity, cf. also [48, Proposition 6] for comparison.

However, as a direct consequence of our result, we find a complete characterization for the case in which the analysis of TGV reduces to a one-dimensional problem.

Corollary 5.14. *Under the same assumption and with the same notation of Theorem 5.13, setting $u_\lambda := u_{\lambda(\hat{\alpha}_0,\hat{\alpha}_1)}$ for every $\lambda \in [0, +\infty]$, there exists $\lambda^*_\Omega \in (0, +\infty)$ such that*

$$J(\lambda^*_\Omega(\hat{\alpha}_0, \hat{\alpha}_1)) = \min_{\lambda \in (0,+\infty)} J(\lambda(\hat{\alpha}_0, \hat{\alpha}_1)).$$

Additionally, there exist positive constants, c_Ω and C_Ω, such that any minimizer λ^*_Ω satisfies $c_\Omega \leqslant \lambda^*_\Omega < C_\Omega \|u_\eta\|_{L^2(\Omega)}$.

In particular, if $\Omega = L$ with $L \subset Q$ is a dyadic square, then there exists a positive constant, c_L, such that any minimizer λ^*_L satisfies $c_L \leqslant \lambda^*_L < C_Q \|u_\eta\|_{L^2(L)}$, where C_Q is a constant given by Proposition 5.11.

As in the case of the total variation, we proceed by first studying the limiting behavior of the sum of fidelity and TGV-seminorm in the sense of Γ-convergence. To describe the situation in which the tuning coefficients approach $+\infty$, it is useful to recall that $\mathcal{M}_b(\Omega;\mathbb{R}^d)$ denotes the set of bounded Radon measures on Ω with values in $\mathbb{R}^d$ and $\operatorname{Ker}\mathcal{E}\,(\Omega;\mathbb{R}^d)$ is the set of all maps $\phi:\Omega\to\mathbb{R}^d$ such that $\mathcal{E}\phi=0$. In particular, $\phi\in\operatorname{Ker}\mathcal{E}\,(\Omega;\mathbb{R}^d)$ if and only if there exists $M\in\mathbb{R}^{d\times d}_{\text{skew}}$ and $m\in\mathbb{R}^d$ such that $\phi(x)=Mx+m$ for every $x\in\Omega$.

We also recall the function

$$m_{\mathcal{E}}:\mathcal{M}_b(\Omega;\mathbb{R}^d)\to\operatorname{Ker}\mathcal{E}\,(\Omega;\mathbb{R}^d),$$

introduced in [48, Proposition 3], and defined as the solution to the minimum problem

$$|\mu-m_{\mathcal{E}}(\mu)|(\Omega)=\min\{|\mu-\phi|(\Omega):\ \phi\in\operatorname{Ker}\mathcal{E}\,(\Omega;\mathbb{R}^d)\},\tag{5.29}$$

for every $\mu\in\mathcal{M}_b(\Omega;\mathbb{R}^d)$.

Lemma 5.15. *Let $\Omega\subset\mathbb{R}^2$ be a bounded, Lipschitz domain and, for each $\alpha\in(0,+\infty)^2$, let $u_\alpha\in BV(\Omega)$ be given by (1.22) and (1.19), with L and Q replaced by Ω. Consider the family of functionals $(G_{\bar\alpha})_{\bar\alpha\in[0,+\infty]^2}$, where $G_{\bar\alpha}:L^1(\Omega)\to[0,+\infty]$ is defined by*

$$G_\alpha[u]:=\begin{cases}\int_\Omega|u_\eta-u|^2\,\mathrm{d}x+TGV_{\alpha_0,\alpha_1}(u,\Omega) & \text{if } u\in BV(\Omega),\\ +\infty & \text{otherwise,}\end{cases}\qquad \text{for } \bar\alpha=:\alpha=(\alpha_0,\alpha_1)\in(0,+\infty)^2,$$

$$G_{0,\bar\alpha_1}[u]:=\begin{cases}\int_\Omega|u_\eta-u|^2\,\mathrm{d}x & \text{if } u\in L^2(\Omega),\\ +\infty & \text{otherwise,}\end{cases}\qquad \text{for } \bar\alpha_0=0 \text{ and } \bar\alpha_1\in[0,+\infty],$$

$$G_{\infty,\alpha_1}[u]:=\begin{cases}\int_\Omega|u_\eta-u|^2\,\mathrm{d}x+\alpha_1|D^2u|(\Omega) & \text{if } u\in BH(\Omega),\\ +\infty & \text{otherwise,}\end{cases}\qquad \text{for } \begin{array}{l}\bar\alpha_0=+\infty,\\ \bar\alpha_1=:\alpha_1\in(0,+\infty),\end{array}$$

$$G_{\bar\alpha_0,0}[u]:=\begin{cases}\int_\Omega|u_\eta-u|^2\,\mathrm{d}x & \text{if } u\in BV(\Omega),\\ +\infty & \text{otherwise,}\end{cases}\qquad \text{for } \bar\alpha_0\in(0,+\infty] \text{ and } \bar\alpha_1=0,$$

$$G_{\alpha_0,\infty}[u]:=\begin{cases}\int_\Omega|u_\eta-u|^2\,\mathrm{d}x+\alpha_0|Du-m_{\mathcal{E}}(Du)|(\Omega) & \text{if } u\in BV(\Omega),\\ +\infty & \text{otherwise,}\end{cases}\qquad \text{for } \begin{array}{l}\bar\alpha_0=:\alpha_0\in(0,\infty),\\ \bar\alpha_1=+\infty,\end{array}$$

$$G_{\infty,\infty}[u]:=\begin{cases}\int_\Omega|u_\eta-u|^2\,\mathrm{d}x & \text{if } Du\in\operatorname{Ker}\mathcal{E}(\Omega;\mathbb{R}^d),\\ +\infty & \text{otherwise,}\end{cases}\qquad \text{for } \bar\alpha_0=\bar\alpha_1=+\infty.$$

Let $(\alpha_j)_{j\in\mathbb{N}}\subset(0,+\infty)^2$ and $\bar\alpha\in[0,\infty]^2$ be such that $\alpha_j\to\bar\alpha$ in $[0,+\infty]^2$. Then, $(G_{\alpha_j})_{j\in\mathbb{N}}$ Γ-converges to $G_{\bar\alpha}$ in $L^1(\Omega)$.

Proof. We first prove that if $(u_j)_{j\in\mathbb{N}}\subset L^1(\Omega)$ and $u\in L^1(\Omega)$ are such that $u_j\to u$ in $L^1(\Omega)$, then

$$G_{\bar\alpha}[u]\leqslant\liminf_{j\to\infty}G_{\alpha_j}[u_j].\tag{5.30}$$

Without loss of generality, we work under the assumptions that

$$\liminf_{j\to\infty}G_{\alpha_j}[u_j]=\lim_{j\to\infty}G_{\alpha_j}[u_j]<+\infty\quad\text{and}\quad\sup_{j\in\mathbb{N}}G_{\alpha_j}[u_j]<+\infty.$$

Then, $u_j\in BV(\Omega)$ for all $j\in\mathbb{N}$, $\sup_{j\in\mathbb{N}}\int_\Omega|u_\eta-u_j|^2\,\mathrm{d}x<+\infty$ and $\sup_{j\in\mathbb{N}}TGV_{(\alpha_j)_0,(\alpha_j)_1}(u_j,\Omega)<+\infty$. Hence, $u\in L^2(\Omega)$ and $u_j\rightharpoonup u$ weakly in $L^2(\Omega)$. For each $j\in\mathbb{N}$, let $u_j^*\in BD(\Omega)$ be such that

$$TGV_{(\alpha_j)_0,(\alpha_j)_1}(u_j)=(\alpha_j)_0|Du_j-u_j^*|(\Omega)+(\alpha_j)_1|\mathcal{E}u_j^*|(\Omega).\tag{5.31}$$

We now consider each limiting behavior of the sequence $(\alpha_j)_{j\in\mathbb{N}}$ separately.

(i) If $\bar\alpha=\alpha\in(0,+\infty)^2$, then an argument by contradiction as the classical TGV case and variants thereof (see, e.g., [26, Proposition 5.3]) yields uniform bounds for sequences $(u_j)_{j\in\mathbb{N}}$ and $(u_j^*)_{j\in\mathbb{N}}$ in $BV(\Omega)$ and $BD(\Omega)$, respectively. Thus, $u\in BV(\Omega)$ and $u_j\rightharpoonup u$ weakly-$\star$ in $BV(\Omega)$. Additionally, there exists $u^*\in BD(\Omega)$ such that, up to extracting a further subsequence, $u_j^*\rightharpoonup u^*$ weakly-$\star$ in $BD(\Omega)$, from which (5.30) follows.

(ii) If $\bar\alpha_0=0$, then (5.30) holds by the lower-semicontinuity of the L^2-norm with respect to the weak convergence in $L^2(\Omega)$.

(iii) If $\bar{\alpha}_0 = +\infty$ and $\bar{\alpha}_1 \in (0, +\infty)$, then $(u_j^*)_{j\in\mathbb{N}}$ is bounded in $BD(\Omega)$. Thus, there exists $u^* \in BD(\Omega)$ such that, up to extracting a further subsequence, $u_j^* \rightharpoonup u^*$ weakly-$\star$ in $BD(\Omega)$. Additionally, $\lim_{j\to\infty} |Du_j - u_j^*|(\Omega) = 0$. Thus, $u_j \to u$ strongly in $BV(\Omega)$, $u \in BH(\Omega)$, and (5.30) holds by the lower-semicontinuity of the L^2-norm with respect to the strong convergence in $BV(\Omega)$.

(iv) If $\bar{\alpha}_0 \in (0, \infty]$ and $\bar{\alpha}_1 = 0$, then the situation is analogous to (ii).

(v) If $\bar{\alpha}_1 = +\infty$ and $\bar{\alpha}_0 \in (0, +\infty)$, then there exists u^* affine and such that $u_j^* \to u^*$ strongly in $BD(\Omega)$ and $(u_j)_{j\in\mathbb{N}}$ is uniformly bounded in $BV(\Omega)$, so that $u_j \overset{*}{\rightharpoonup} u$ weakly-$\star$ in $BV(\Omega)$. The statement follows from the lower semicontinuity of the total variation with respect to the weak-$\star$ convergence of measures, as well as from (5.29).

(vi) If $\bar{\alpha}_0 = \bar{\alpha}_1 = +\infty$, then there exists $u^* \in \operatorname{Ker} \mathcal{E}(\Omega; \mathbb{R}^d)$ such that $u_j^* \to u^*$ strongly in $BD(\Omega)$ and $Du_j \to u^*$ strongly in $\mathcal{M}_b(\Omega; \mathbb{R}^d)$. Thus, $Du = u^*$ and the statement follows.

Next, we show that for any $u \in L^1(\Omega)$, there exists $(u_j)_{j\in\mathbb{N}} \subset L^1(\Omega)$ such that $u_j \to u$ in $L^1(\Omega)$ and

$$G_{\bar{\alpha}}[u] \geqslant \limsup_{j\to\infty} G_{\alpha_j}[u_j]. \tag{5.32}$$

Again, we detail the argument in each case separately.

(i) If $\bar{\alpha} = \alpha \in (0, +\infty)^2$ then we can assume, without loss of generality, that $u \in BV(\Omega)$. The conclusion follows then by a classical argument relying on the continuity of TGV with respect to its tuning parameters (see, e.g., [26, Theorem 4.2]).

(ii) If $\bar{\alpha}_0 = 0$, then we consider for every $u \in L^2(\Omega)$ an approximating sequence $(u_k)_{k\in\mathbb{N}} \subset C_c^\infty(\Omega)$ such that $u_k \to u$ strongly in $L^2(\Omega)$. Choosing the null function as a competitor in the definition of TGV, we find that

$$TGV_{(\alpha_j)_0,(\alpha_j)_1}(u_k) \leqslant (\alpha_j)_0 |Du_k|(\Omega).$$

Thus,

$$\lim_{j\to+\infty} G_{\alpha_j}[u_k] = G_{0,\bar{\alpha}_1}[u_k]$$

for every $\bar{\alpha}_1 \in [0, +\infty]$ and every $k \in \mathbb{N}$. The thesis follows then by a classical diagonalization argument.

(iii) If $\bar{\alpha}_0 = +\infty$ and $\bar{\alpha}_1 = \alpha_1 \in (0, +\infty)$ then we can assume, without loss of generality, that $u \in BH(\Omega)$. In particular, $\nabla u \in BD(\Omega)$ which we can then use as a competitor in the definition of TGV to infer that

$$TGV_{(\alpha_j)_0,(\alpha_j)_1}(u) \leqslant (\alpha_j)_1 |D^2 u|(\Omega).$$

Thus,

$$\limsup_{j\to+\infty} G_{\alpha_j}[u] \leqslant \limsup_{j\to+\infty} \left(\int_\Omega |u_\eta - u|^2 \, dx + (\alpha_j)_1 |D^2 u|(\Omega) \right) = G_{\infty,\alpha_1}[u].$$

(iv) If $\bar{\alpha}_0 \in (0, +\infty]$ and $\bar{\alpha}_1 = 0$, arguing by approximation as in case (ii), we can assume without loss of generality that $u \in C_c^\infty(\Omega)$. Then, choosing ∇u as a competitor in the definition of TGV, we find that

$$TGV_{(\alpha_j)_0,(\alpha_j)_1}(u) \leqslant (\alpha_j)_1 |\nabla^2 u|(\Omega).$$

Hence, arguing as in case (ii) once more, yields (5.32).

(v) If $\bar{\alpha}_0 = \alpha_0 \in (0, +\infty)$ and $\bar{\alpha}_1 = +\infty$, then we can assume that $u \in BV(\Omega)$. Choosing $m_{\mathcal{E}}(Du)$ in the definition of TGV, yields

$$TGV_{(\alpha_j)_0,(\alpha_j)_1}(u) \leqslant (\alpha_j)_0 |Du - m_{\mathcal{E}}(Du)|(\Omega).$$

Hence, arguing as in case (ii), we infer (5.32).

(vi) If $\bar{\alpha}_0 = \bar{\alpha}_1 = +\infty$, then, without loss of generality, we can assume that $Du \in \operatorname{Ker} \mathcal{E}(\Omega; \mathbb{R}^d)$. Choosing Du as a competitor in the definition of TGV shows that $TGV_{(\alpha_j)_0,(\alpha_j)_1}(u) = 0$ for every $j \in \mathbb{N}$, from which (5.32) follows.

The Γ-convergence of $(G_{\alpha_j})_{j\in\mathbb{N}}$ to $G_{\bar{\alpha}}$ in $L^1(\Omega)$ is then a direct consequence of (5.30) and (5.32). $\square$

As a consequence of the previous result, we provide a characterization of the unique minimizer $u_{\bar{\alpha}}$ of $G_{\bar{\alpha}}$.

Corollary 5.16. *Under the same assumptions of Lemma 5.15, let* $u_{\bar\alpha} := \operatorname{argmin}_{u\in L^1(\Omega)} G_{\bar\alpha}[u]$ *for* $\bar\alpha \in [0,+\infty]^2$. *Then,*

$$u_{\bar\alpha} = \begin{cases} u_\alpha & \text{if } \bar\alpha = \alpha \in (0,+\infty)^2 \\ u_\eta & \text{if } \bar\alpha_0 = 0 \text{ or } \bar\alpha_1 = 0, \\ \langle u_\eta\rangle_\Omega & \text{if } \bar\alpha_0 = \bar\alpha_1 = +\infty. \end{cases} \tag{5.33}$$

Additionally, when just one among $\bar\alpha_0$ *and* $\bar\alpha_1$ *is infinite, then* $\langle u_{\bar\alpha}\rangle_\Omega = \langle u_\eta\rangle_\Omega$. *In these regimes, if additionally* $u_\eta = \langle u_\eta\rangle_\Omega$, *then* $u_{\bar\alpha} = \langle u_{\bar\alpha}\rangle_\Omega$.

Proof. The first claim follows directly from Lemma 5.15. We show the second statement only in the case in which $\bar\alpha_0 = \infty$ and $\bar\alpha_1$ is finite, being the case in which $\bar\alpha_1 = \infty$ analogous. The characterization of minimizers is then a consequence of the orthogonality property in (5.5) which, in turn, yields

$$G_{\infty,\bar\alpha_1}(u) = \int_\Omega |\langle u - u_\eta\rangle_\Omega|^2 \,dx + \int_\Omega \left[(u - \langle u\rangle_\Omega) - (u_\eta - \langle u_\eta\rangle_\Omega)\right]^2 \,dx + \bar\alpha_1 |D^2 (u - \langle u\rangle_\Omega)|$$

for every $u \in BH(\Omega)$. □

We proceed by proving the analogue to Lemma 3.10 in the TGV-setting.

Lemma 5.17. *Let* $\Omega \subset \mathbb{R}^2$ *be a bounded, Lipschitz domain and let* $(G_{\bar\alpha})_{\bar\alpha\in[0,+\infty]}$ *be the family of functionals introduced in Lemma 5.15. Given* $\bar\alpha \in [0,\infty]^2$, *set* $u_{\bar\alpha} := \operatorname{argmin}_{u\in L^1(\Omega)} G_{\bar\alpha}[u]$. *Then, there exists a sequence of pair of positive numbers,* $(\alpha_j)_{j\in\mathbb{N}} \subset (0,+\infty)^2$, *such that* $\alpha_j \to \bar\alpha$ *in* $[0,+\infty]^2$ *as* $j \to \infty$ *and*

$$\lim_{j\to\infty} \int_\Omega |u_{\alpha_j} - u_{\bar\alpha}|^2 \,dx = 0, \tag{5.34}$$

where $u_{\alpha_j} := \operatorname{argmin}_{u\in L^1(\Omega)} G_{\alpha_j}[u]$ *for all* $j \in \mathbb{N}$.

Proof. With the same notation as in the proof of Lemma 5.15, we detail the argument for each case separately.

(i) If $\bar\alpha = \alpha \in (0,+\infty)^2$, then the statement follows directly by choosing $\alpha_j = \alpha$ for every j.

(ii) If $\bar\alpha_0 = 0$, then $u_{\bar\alpha} = u_\eta$ and $G_{\bar\alpha}[u_{\bar\alpha}] = G_{\bar\alpha}[u_\eta] = 0$. In view of Lemma 5.15, there exists a sequence $(u^j_\eta)_{j\in\mathbb{N}} \subset L^1(\Omega)$ such that

$$\limsup_{j\to+\infty} G_{\alpha_j}[u^j_\eta] \leqslant G_{\bar\alpha}[u_\eta].$$

Hence, for any sequence $(\alpha_j)_{j\in\mathbb{N}} \subset (0,+\infty)^2$ satisfying $\alpha_j \to \bar\alpha$, the minimality of u_{α_j} yields

$$\begin{aligned}\limsup_{j\to\infty} \left(\int_\Omega |u_{\alpha_j} - u_\eta|^2 \,dx + TGV_{(\alpha_j)_0,(\alpha_j)_1}(u_{\alpha_j},\Omega) \right) &= \limsup_{j\to\infty} G_{\alpha_j}[u_{\alpha_j}] \\ &\leqslant \limsup_{j\to\infty} G_{\alpha_j}[u^j_\eta] \leqslant G_{\bar\alpha}[u_\eta] = 0.\end{aligned}$$

Thus, we infer (5.34).

(iii) If $\bar\alpha_0 = +\infty$ and $\bar\alpha_1 = \alpha_1 \in (0,+\infty)$, then $u_{\bar\alpha} \in BH(\Omega)$. For every sequence $(\alpha^j_0)_{j\in\mathbb{N}}$ such that $\alpha^j_0 \to +\infty$ as $j \to +\infty$, setting $\alpha_j := (\alpha^j_0, \alpha_1)$, from the minimality of u_{α_j} and choosing $\nabla u_{\bar\alpha}$ as a competitor in the definition of TGV, we find

$$G_{\alpha_j}[u_{\alpha_j}] \leqslant G_{\alpha_j}[u_{\bar\alpha}] \leqslant \int_\Omega |u_{\bar\alpha} - u_\eta|^2 \,dx + \alpha_1 |D^2 u_{\bar\alpha}|(\Omega) = G_{\bar\alpha}[u_{\bar\alpha}].$$

By the fundamental theorem of Γ-convergence (see [24, Corollary 7.20 and Theorem 7.8]), the equi-coerciveness of the functionals G_{α_j} together with the uniqueness of minimizers yields that $u_{\alpha_j} \rightharpoonup u_{\bar\alpha}$ weakly in $L^2(\Omega)$. Property (5.34) follows then by arguing as in item *(iii)* in the first part of the proof of Lemma 5.15 and using the continuous embedding $BV(\Omega) \subset L^2(\Omega)$.

(iv) If $\bar\alpha_0 \in (0,+\infty]$ and $\bar\alpha_1 = 0$, then $u_{\bar\alpha} = u_\eta$. Let $(u^k_\eta)_{k\in\mathbb{N}} \subset C^\infty_c(\Omega)$ be such that $u^k_\eta \to u_\eta$ strongly in $L^2(\Omega)$. For every sequence $(\alpha_j)_{j\in\mathbb{N}} \subset (0,+\infty)^2$ satisfying $\alpha_j \to \bar\alpha$, we obtain from the minimality of u_{α_j} that

$$G_{\alpha_j}[u_{\alpha_j}] \leqslant G_{\alpha_j}[u^k_\eta] \leqslant \int_\Omega |u^k_\eta - u_\eta|^2 \,dx + (\alpha_1)_j \int_\Omega |\nabla^2 u^k_\eta| \,dx,$$

where the latter inequality follows by choosing ∇u_η^k as a competitor in the definition of TGV. Thus,

$$\limsup_{j\to+\infty} G_{\alpha_j}[u_{\alpha_j}] \leqslant \int_\Omega |u_\eta - u_\eta^k|^2 \, \mathrm{d}x$$

for every $k \in \mathbb{N}$. Passing to the limit as $k \to +\infty$, we infer that

$$\limsup_{j\to+\infty} G_{\alpha_j}[u_{\alpha_j}] = 0.$$

In turn, this implies (5.34).

(v) If $\bar\alpha_0 = \alpha_0 \in (0,+\infty)$ and $\bar\alpha_1 = +\infty$, then $u_{\bar\alpha} \in BV(\Omega)$. For $(\alpha_1^j)_{j\in\mathbb{N}} \subset (0,+\infty)$ such that $\alpha_1^j \to +\infty$, and setting $\alpha_j := (\alpha_0, \alpha_1^j)$, we deduce that

$$G_{\alpha_j}[u_{\alpha_j}] \leqslant G_{\alpha_j}[u_{\bar\alpha}] \leqslant \int_\Omega |u_{\bar\alpha} - u_\eta|^2 \, \mathrm{d}x + \alpha_0 |Du_{\bar\alpha} - m_{\mathcal{E}}(Du_{\bar\alpha})|(\Omega) = G_{\bar\alpha}[u_{\bar\alpha}].$$

By the fundamental theorem of Γ-convergence, we infer that $u_{\alpha_j} \to u_{\bar\alpha}$ strongly in $L^1(\Omega)$ and that $G_{\alpha_j}[u_{\alpha_j}] \to G_{\bar\alpha}[u_{\bar\alpha}]$. On the other hand, letting $u^*_{\alpha_j}$ be defined as in (5.31) with u_j replaced by u_{α_j}, the same argument as in item (v) in the first part of the proof of Lemma 5.15 yields $u_{\alpha_j} \overset{*}{\rightharpoonup} u$ weakly-$\star$ in $BV(\Omega)$, and $u^*_{\alpha_j} \to u^*$ strongly in $BD(\Omega)$ with u^* affine. By combining the above convergences, we deduce

$$\begin{aligned} G_{\bar\alpha}[u_{\bar\alpha}] &\leqslant \int_\Omega |u_{\bar\alpha} - u_\eta|^2 \, \mathrm{d}x + \alpha_0 |Du_{\bar\alpha} - u^*|(\Omega) \\ &\leqslant \liminf_{j\to+\infty} \int_\Omega |u_{\alpha_j} - u_\eta|^2 \, \mathrm{d}x + \alpha_0 |Du_{\alpha_j} - u^*_{\alpha_j}|(\Omega) \leqslant \lim_{j\to+\infty} G_{\alpha_j}[u_j] = G_{\bar\alpha}[u_{\bar\alpha}], \end{aligned}$$

where the first inequality follows by the definition of $m_{\mathcal{E}}$, cf. (5.29), whereas the second one is a consequence of the lower semicontinuity of the L^2-norm with respect to the weak L^2-convergence, as well as of the lower semicontinuity of the total variation with respect to the weak-$\star$ convergence of measures.

(vi) If $\bar\alpha_0 = \bar\alpha_1 = +\infty$, then $u_{\bar\alpha}$ is affine. Thus, for every sequence $(\alpha_j)_{j\in\mathbb{N}} \subset (0,+\infty)^2$ satisfying $\alpha_j \to \bar\alpha$,

$$G_{\alpha_j}[u_{\alpha_j}] \leqslant G_{\alpha_j}[u_{\bar\alpha}] = \int_\Omega |u_{\bar\alpha} - u_\eta|^2 \, \mathrm{d}x = G_{\bar\alpha}[u_{\bar\alpha}].$$

Property (5.34) is once again obtained arguing by the fundamental theorem of Γ-convergence, as in (iii). □

In view of the lemmas above, we obtain the following characterization of the lower semicontinuous envelope of J.

Lemma 5.18. *Let $\Omega \subset \mathbb{R}^2$ be a bounded, Lipschitz domain, and let $J : (0,+\infty)^2 \to [0,+\infty)$ be the function defined in (5.28). Then, the lower-semicontinuous envelope of J on $[0,\infty]^2$, $J^{sc^-} : [0,+\infty]^2 \to [0,+\infty]$, defined for $\bar\alpha \in [0,+\infty]^2$ by*

$$J^{sc^-}(\bar\alpha) := \inf\Big\{ \liminf_{j\to\infty} J(\alpha_j) \colon\ (\alpha_j)_{j\in\mathbb{N}} \subset (0,+\infty)^2,\ \alpha_j \to \bar\alpha \text{ in } [0,+\infty]^2 \Big\},$$

satisfies

$$J^{sc^-}(\bar\alpha) = \begin{cases} J(\alpha) = \|u_\alpha - u_c\|^2_{L^2(\Omega)} & \text{if } \bar\alpha = \alpha \in (0,+\infty)^2, \\ \|u_\eta - u_c\|^2_{L^2(\Omega)} & \text{if } \bar\alpha_0 = 0 \text{ or } \bar\alpha_1 = 0, \\ \|\langle u_\eta\rangle - u_c\|^2_{L^2(\Omega)} & \text{if } \bar\alpha_0 = \bar\alpha_1 = +\infty, \\ \|u_{\bar\alpha} - u_c\|^2_{L^2(\Omega)} \text{ with } \langle u_{\bar\alpha}\rangle = \langle u_\eta\rangle & \text{otherwise}, \end{cases} \tag{5.35}$$

where $u_{\bar\alpha}$ is the unique minimizer of $G_{\bar\alpha}$, cf. Corollary 5.16.

We omit the proof of the lemma above, for it follows by a direct adaptation of the proof of Lemma 3.11. We are now in a position to prove Theorem 5.13.

Proof of Theorem 5.13. The proof is subdivided into three steps.

Step 1. We prove that if condition *i)* in the statement holds, namely

$$TGV_{\hat\alpha_0,\hat\alpha_1}(u_\eta,\Omega)-TGV_{\hat\alpha_0,\hat\alpha_1}(u_c,\Omega)>0$$

for some $\hat\alpha\in(0,+\infty)^2$, then there exists $\bar\alpha\in(0,+\infty)^2$ such that

$$\|u_{\bar\alpha}-u_c\|^2_{L^2(\Omega)}<\|u_\eta-u_c\|^2_{L^2(\Omega)}. \tag{5.36}$$

From the convexity of the TGV-seminorm, arguing as in the proof of (3.17), we infer that

$$\|u_\eta-u_c\|^2_{L^2(\Omega)}-\|u_\alpha-u_c\|^2_{L^2(\Omega)}\leqslant TGV_{\alpha_0,\alpha_1}(u_\alpha,\Omega)-TGV_{\hat\alpha_0,\hat\alpha_1}(u_c,\Omega)$$

for every $\alpha\in(0,+\infty)^2$. Choosing $\alpha=\lambda\hat\alpha$, and denoting $u_{\lambda(\hat\alpha)}$ by u_λ, for simplicity, we find that

$$\|u_\eta-u_c\|^2_{L^2(\Omega)}-\|u_\lambda-u_c\|^2_{L^2(\Omega)}\leqslant \lambda\left(TGV_{\hat\alpha_0,\hat\alpha_1}(u_\lambda,\Omega)-TGV_{\hat\alpha_0,\hat\alpha_1}(u_c,\Omega)\right)$$

for every $\lambda\in(0,+\infty)$. By the proof of case (ii) of Lemma 5.17 and by Corollary 5.16, it follows that, up to (non-relabelled) subsequences, $u_\lambda\to u_\eta$ strongly in $L^2(\Omega)$ as $\lambda\to 0$. Fix $\varepsilon>0$; by the continuity of the TGV-seminorms with respect to the strong L^2-convergence, we conclude that

$$TGV_{\hat\alpha_0,\hat\alpha_1}(u_\lambda,\Omega)\geqslant TGV_{\hat\alpha_0,\hat\alpha_1}(u_\eta,\Omega)-\varepsilon(TGV_{\hat\alpha_0,\hat\alpha_1(u_\eta,\Omega)}-TGV_{\hat\alpha_0,\hat\alpha_1}(u_c,\Omega))$$

for λ small enough. Thus,

$$\|u_\eta-u_c\|^2_{L^2(\Omega)}-\|u_\lambda-u_c\|^2_{L^2(\Omega)}\geqslant \lambda(TGV_{\hat\alpha_0,\hat\alpha_1}(u_\eta,\Omega)-TGV_{\hat\alpha_0,\hat\alpha_1}(u_c,\Omega))(1-\varepsilon)$$

for λ small enough. This implies that there exists $\bar\lambda\in(0,+\infty)$ for which

$$\|u_\eta-u_c\|^2_{L^2(\Omega)}>\|u_\lambda-u_c\|^2_{L^2(\Omega)}.$$

The preceding estimate yields the thesis by choosing $\bar\alpha=\bar\lambda(\hat\alpha_0,\hat\alpha_1)$.

Step 2. We prove that if condition *ii)* in the statement holds, (i.e., $\|u_\eta-u_c\|^2_{L^2(\Omega)}<\|\langle u_\eta\rangle-u_c\|^2_{L^2(\Omega)}$), then there exits $\bar\alpha\in(0,+\infty)^2$ such that

$$\|u_{\bar\alpha}-u_c\|^2_{L^2(\Omega)}<\|\langle u_\eta\rangle-u_c\|^2_{L^2(\Omega)}. \tag{5.37}$$

In view of Step 1,

$$\lim_{\lambda\to 0}\|u_\lambda-u_c\|_{L^2(\Omega)}=\|u_\eta-u_c\|_{L^2(\Omega)}<\|\langle u_\eta\rangle-u_c\|_{L^2(\Omega)}.$$

By the proof of case (vi) of Lemma 5.17 and by Corollary 5.16, we obtain the existence of $\bar\lambda\in(0,+\infty)$ for which

$$\|u_{\bar\lambda}-u_c\|_{L^2(\Omega)}<\|\langle u_\eta\rangle-u_c\|_{L^2(\Omega)}.$$

The claim follows by choosing $\bar\alpha=\bar\lambda(\hat\alpha_0,\hat\alpha_1)$.

Step 3. We conclude the proof by establishing the bounds on the parameters stated in Theorem 5.13. From the lower semicontinuity of J^{sc^-}, we infer that there exists $\alpha^*\in[0,+\infty]^2$ where the minimum value is attained. By Corollary 5.16 and by the previous steps, α^* satisfies (5.26) and

$$J^{sc^-}(\alpha^*)=\min_{\bar\alpha\in[0,+\infty]^2}J^{sc^-}(\bar\alpha). \tag{5.38}$$

To prove the existence of the lower bound c_Ω, we argue by contradiction. We first assume that there exists a sequence $(\alpha^*_j)_{j\in\mathbb{N}}\subset(0,+\infty)^2$ such that $\alpha^*_j\to 0$ as $j\to+\infty$, and (5.38) holds for $\alpha^*=\alpha^*_j$ for all $j\in\mathbb{N}$. In view of the lower semi-continuity of J^{sc^-} on $[0,+\infty]^2$,

$$\min_{\bar\alpha\in[0,+\infty]^2}J^{sc^-}(\bar\alpha)\leqslant J^{sc^-}(0)\leqslant\liminf_{j\to\infty}J^{sc^-}(\alpha^*_j)=\min_{\bar\alpha\in[0,+\infty]^2}J^{sc^-}(\bar\alpha),$$

which is false by (5.36). This proves the existence of a constant $\hat c_\Omega$ such that $|\alpha^*|\geqslant\hat c_\Omega$ for every minimizer α^* of J^{sc^-}. The existence of the constant c_Ω as in the statement of the theorem follows by observing that the above argument can be repeated by considering sequences $(\alpha^*_j)_{j\in\mathbb{N}}$ for which just one of the entries converges to zero.

The bound from above on $\min\{\alpha^*_0,\alpha^*_1\}$ follows directly by Proposition 5.11. In fact, from (5.22), we infer the existence of a constant C_Ω such that u_{α^*} is affine if $C_\Omega\|u_\eta\|_{L^2(\Omega)}<\min\{\alpha^*_0,\alpha^*_1\}$. Now, assume by contradiction that there exists a sequence $(\alpha^*_j)_{j\in\mathbb{N}}\subset(0,+\infty)^2$ such that both entries of α^*_j blow up to infinity as $j \to +\infty$, and (5.38) holds for $\alpha^* = \alpha_j^*$ for all $j \in \mathbb{N}$. Using, once again, the lower semi-continuity of J^{sc^-} on $[0, +\infty]^2$, we find that

$$\min_{\bar{\alpha} \in [0,+\infty]^2} J^{sc^-}(\bar{\alpha}) \leqslant J^{sc^-}(+\infty, +\infty) \leqslant \liminf_{j \to \infty} J^{sc^-}(\alpha_j^*) = \min_{\bar{\alpha} \in [0,+\infty]^2} J^{sc^-}(\bar{\alpha}),$$

which is false by Corollary 5.16 and (5.37). ☐

5.5. **The $(\mathcal{LS})_{TGV-Fid_\omega}$ learning scheme.** Given a dyadic square $L \subset Q$ and $\lambda \in (0, \infty)$, we have

$$\begin{aligned} &\operatorname{argmin}\left\{\lambda \int_L |u_\eta - u|^2 \, dx + TGV_{1,1}(u, L) \colon u \in BV(L)\right\} \\ &= \operatorname{argmin}\left\{\int_L |u_\eta - u|^2 \, dx + TGV_{\frac{1}{\lambda}, \frac{1}{\lambda}}(u, L) \colon u \in BV(L)\right\}. \end{aligned}$$

The analysis in Subsections 5.1.1–5.3 applies also to the weighted-fidelity learning scheme and yields Theorem 1.8. As before, the previous existence theorem holds true under any stopping criterion for the refinement of the admissible partitions provided that the training data satisfies suitable conditions. We summarize the situation in the next result, which follows directly by the discussions in the previous subsection, in particular Corollary 5.14.

Theorem 5.19 (Equivalence between box constraint and stopping criterion). *Consider the learning scheme $(\mathcal{LS})_{TGV-Fid_\omega}$ in (1.27). The two following conditions hold:*

(a) If we replace (1.21) by (5.3), then there exists a stopping criterion $(\mathcal{S})$ for the refinement of the admissible partitions as in Definition 1.2.

(b) Assume that there exists a stopping criterion $(\mathcal{S})$ for the refinement of the admissible partitions as in Definition 1.2 such that the training data satisfies for all $L \in \bigcup_{\mathcal{L} \in \bar{\mathcal{P}}} \mathcal{L}$, with $\bar{\mathcal{P}}$ as in Definition 1.2, the conditions

(i) $TGV_{\alpha_0,\alpha_1}(u_c, L) < TGV_{\alpha_0,\alpha_1}(u_\eta, L)$,

(ii) $\|u_\eta - u_c\|^2_{L^2(L)} < \|\langle u_\eta \rangle_L - u_c\|^2_{L^2(L)}$.

Then, there exist $c_0, c_1 \in \mathbb{R}^+$ such that the optimal solution u^* provided by $(\mathcal{LS})_{TGV-Fid_\omega}$ with $\mathcal{P}$ replaced by $\bar{\mathcal{P}}$ coincides with the optimal solution u^* provided by $(\mathcal{LS})_{TGV-Fid_\omega}$ with (1.21) replaced by (5.3).

6. Numerical Treatment and Comparison of the learning schemes $(\mathcal{LS})_{TV_\omega}$, $(\mathcal{LS})_{TV_{\omega_\varepsilon}}$, $(\mathcal{LS})_{TV-Fid_\omega}$, and $(\mathcal{LS})_{TGV-Fid_\omega}$

6.1. **Common numerical framework for all schemes.** The focus of our article is on the use of space-dependent weights and, from the numerical point of view, our schemes require addressing weights that are piecewise constant on dyadic partitions. This stands in contrast to previous approaches for optimizing space-dependent parameters, which in most cases hinge on H^1-type penalizations of the weights, as done in [23, 36] for TV, [35] for TGV and [46] for some more general convex regularizers. The piecewise constant setting makes it possible to work in a modular fashion, building upon any numerical methods that are able to compute solutions to denoising with a weight (Level 2) and finding constant optimal regularization parameters (Level 3).

In our numerical examples, we have used a basic first-order finite difference discretization of the gradient and symmetrized gradient, on the regular grid arising from the discrete input images. For solving TV regularized denoising, either with constant or varying weights, we have opted for the standard primal-dual hybrid gradient (PDHG) descent scheme of [17]. The optimization for optimal constant parameters α of Level 3 is done with the 'piggyback' version of the same algorithm, which has been proposed in [18] to learn finite difference discretizations of TV with a high degree of isotropy, and further analyzed under smoothness assumptions on the energies in [4]. Essentially, it consists in evolving an adjoint state along with the main variables, to keep track of the sensitivity of the solution with respect to parameters. We remark that such sensitivity analysis in principle requires not just first but second derivatives of the cost functions involved, which in our case are only componentwise piecewise smooth. In any case, as already observed in [18, Appendix A], we do achieve an adequate performance in practice.

These methods are based on the saddle point formulation

$$\max_{y \in \mathcal{Y}} \min_{x \in \mathcal{X}} \langle y, Kx \rangle_{\ell^2} + \mathcal{G}(x) - \mathcal{F}^*(y),$$

with $\mathcal{G}$ representing the differentiable fidelity term and $\mathcal{F}^*$ being the projection onto a convex set, arising as the Fenchel conjugate of an ℓ^1-type norm. Denoting by $W = \mathbb{R}^{nm}$ the space of discrete scalar-valued functions, these read in the TV case as

$$\mathcal{X} = W,\ \mathcal{Y} = W^2,\ K = \nabla,\ \mathcal{G}(u) = \lambda \sum_{ij} \left(u^{ij} - u_\eta^{ij}\right)^2, \text{ and}$$
$$\mathcal{F}^*(p) = \mathcal{I}_{Q_{TV}} \text{ with } Q_{TV} = \left\{p \in \mathcal{Y} \,|\, (p_1^{ij})^2 + (p_2^{ij})^2 \leqslant \alpha \text{ for all } i,j\right\}.$$

For the TGV case, following the approach used in [5] and [7], we have used

$$\mathcal{X} = W \times W^2,\ \mathcal{Y} = W^2 \times W^3,\ K = \begin{pmatrix} \nabla u & -\mathrm{Id} \\ 0 & \mathcal{E} \end{pmatrix},\ \mathcal{G}(u,v) = \lambda \sum_{ij} \left(u^{ij} - u_\eta^{ij}\right)^2, \text{ and}$$

$$\mathcal{F}^*(p) = \mathcal{I}_{Q_{TGV}} \text{ for } Q_{TGV} = \left\{(p,q) \in \mathcal{Y} \,|\, (p_1^{ij})^2 + (p_2^{ij})^2 \leqslant \alpha_0, (q_{11}^{ij})^2 + 2(q_{12}^{ij})^2 + (q_{22}^{ij})^2 \leqslant \alpha_1 \text{ for all } i,j\right\}.$$

With this notation and denoting the subgradient by ∂, the PDHG algorithm [17, Algorithm 1] can be written as

$$\begin{cases} y^{k+1} = (\mathrm{Id} + \sigma \partial \mathcal{F}^*)^{-1}(y^k + \sigma K \bar{x}^k), \\ x^{k+1} = (\mathrm{Id} + \tau \partial \mathcal{G})^{-1}(x^k - \tau K^* y^{k+1}), \\ \bar{x}^{k+1} = x^{k+1} + \theta(x^{k+1} - x^k), \end{cases} \tag{6.1}$$

where the descent parameters satisfy $\sigma\tau\|K\| \leqslant 1$. In the TV case, this operator norm of ∇ can be bounded by $\sqrt{8}$ (cf. [15, Theorem 3.1]), while in the TGV case, we have $\|K\|^2 \leqslant (17 + \sqrt{33})/2$ (cf. [5, Section 3.2]). The piggyback algorithm of [18, 4] introduces one adjoint variable for each primal and dual variable above (denoted by $X \in \mathcal{X}$, $Y \in \mathcal{Y}$, $U \in W$, $P \in W^2$, $Q \in W^3$) and performs the same kind of updates also on these new variables to optimize the values of a loss function L, resulting in

$$\begin{cases} Y^{k+1} = D \operatorname{prox}_{\sigma \mathcal{F}^*}(y^k + \sigma K \bar{x}^k) \cdot \left[Y^k + \sigma K\left(\bar{X}^k + D_x L(x^k, y^k)\right)\right], \\ X^{k+1} = D \operatorname{prox}_{\tau \mathcal{G}}(x^k - \tau K^* y^{k+1}) \cdot \left[X^k - \tau K^*\left(Y^{k+1} + D_y L(x^k, y^k)\right)\right], \\ \bar{X}^{k+1} = X^{k+1} + \theta(X^{k+1} - X^k), \end{cases} \tag{6.2}$$

where $\operatorname{prox}_{\tau\mathcal{G}} = (\mathrm{Id} + \tau\partial\mathcal{G})^{-1}$ and $\operatorname{prox}_{\sigma\mathcal{F}^*} = (\mathrm{Id} + \tau\partial F^*)^{-1}$ as appearing in (6.1); the latter corresponds to a projection onto Q_{TV} or Q_{TGV} which, as already remarked, is not differentiable on the boundary of these sets.

In our case, we optimize the squared L^2 distance to u_c by varying the fidelity parameter $\lambda = 1/\alpha$, so that

$$L(u) = \frac{1}{2} \sum_{ij} (u^{ij} - u_c^{ij})^2 \ \text{ and } \ D_\lambda \mathcal{L}(\lambda) = \lambda \sum_{ij} \hat{U}^{ij} (\hat{u}^{ij} - u_\eta^{ij}) \ \text{ for } \ \mathcal{L}(\lambda) = L(\hat{u}(\lambda)),$$

where $\hat{u}$, $\hat{U}$ are the optimal image variable and corresponding adjoint obtained after convergence of (6.1) and (6.2). We have then used the derivative $D_\lambda \mathcal{L}$ to update λ with vanilla gradient descent steps.

6.2. Effect of parameter discontinuities in Level 2 of $(\mathcal{LS})_{TV_\omega}$, $(\mathcal{LS})_{TV_{\omega_\varepsilon}}$ and $(\mathcal{LS})_{TV-Fid_\omega}$.

In Figure 2, we present an example using large regularization parameters and a symmetric input image to demonstrate the effect of parameter discontinuities in Level 2 of the schemes $(\mathcal{LS})_{TV_\omega}$, $(\mathcal{LS})_{TV_{\omega_\varepsilon}}$ and $(\mathcal{LS})_{TV-Fid_\omega}$. In the weighted-$TV$ result, a jump in the weight results in a spurious discontinuity in the resulting image. Mollifying the weight smooths the transition slightly, and it shifts it to the side with lower weight. Using a weighted fidelity term does not introduce discontinuities besides those present in the input, but still creates visible artifacts near them.

6.3. Dyadic subdivision approach to Level 1.

In Algorithm 1, we summarize our approach to numerically treat Level 1. We remark that in comparison with the original formulations $(\mathcal{LS})_{TV_\omega}$ and $(\mathcal{LS})_{TGV_\omega}$ as formulated in the introduction, we do not search the entire space of partitions (which would be numerically intractable) and instead work by subdivision as in Example 3.12. This means that for any given cell L, we make a local decision whether to subdivide it or not, based on the training costs arising from it before and after subdividing it in four new cells. When performing this subdivision, the parameter from the original cell is used as initialization for the optimization on the newly created ones.

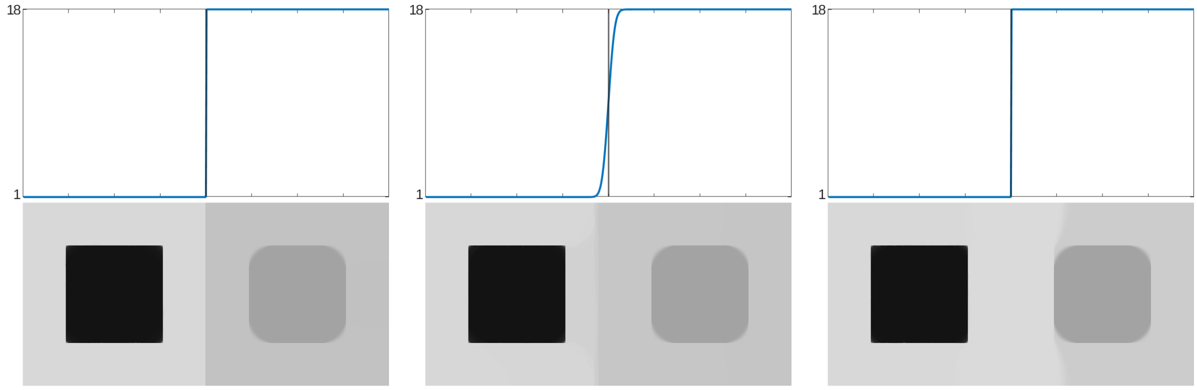

FIGURE 2. Oversmoothed denoising with a sharp change of weight and schemes corresponding to Level 2 of $(\mathscr{LS})_{TV_\omega}$, $(\mathscr{LS})_{TV_{\omega_\varepsilon}}$, and $(\mathscr{LS})_{TV-Fid_\omega}$, from left to right. Top row: weights $\omega(x) = \omega(x_1)$. Bottom row: results with each denoising scheme and the corresponding (not optimal) weight.

Algorithm 1 Numerical approach to Level 1

Input: Noisy image u_η, clean (training) image u_c, subdivision tolerance ρ.
Output:

1. Set $\ell = 0$ and $\mathscr{L} = \{(0,1)^2\}$.
2. Compute the constant optimal $\omega_{(0,1)^2} \equiv \lambda_{(0,1)^2}$ or $\omega_{(0,1)^2} \equiv 1/\lambda_{(0,1)^2}$ using the numerical approach to Level 3 described in Section 6.1. Store the cost at the minimum as $c_{(0,1)^2}$.

while $\ell < \ell_{\max}$ **do**
  for all $L \in \mathscr{L}$ with $\mathrm{side}(L) = 2^{-\ell}$ **do**
    2. Denote by L_i for $i = 1, \ldots, 4$ the cells obtained by one dyadic subdivision of L.
    for $i = 1, \ldots, 4$ **do**
      3. Compute λ_{L_i} with the approach to Level 3 of Section 6.1, store the local optimal cost as $C_{L_i} := \|u_c - u_{L_i}\|^2_{L^2(L_i)}$.
    end for
    if $C_{L_1} + C_{L_2} + C_{L_3} + C_{L_4} < \rho C_L$ (cf. (3.23)) **then**
      4. Replace $\mathscr{L}$ by $\big(\mathscr{L} \setminus \{L\}\big) \bigcup_{i=1}^4 \{L_i\}$.
    end if
  end for
  5. Set $\omega_{\mathscr{L}}$ to be $\omega_{\mathscr{L}} = \lambda_L$ or $\omega_{\mathscr{L}} = 1/\lambda_L$ on each $L \in \mathscr{L}$.
  6. Set $\ell = \ell + 1$.
end while
7. Compute $u_{\mathscr{L}}$ with $\omega_{\mathscr{L}}$ and the numerical approach of Section 6.1 to Level 2 of $(\mathscr{LS})_{TGV-Fid_\omega}$ or $(\mathscr{LS})_{TV-Fid_\omega}$.

It is worth mentioning that methods to handle the bilevel optimization problems of Level 3 in a nonsmooth setting have been introduced in [29, 33, 6]. One could also use these in our subdivision scheme within Algorithm 1, and in fact the authors of the cited papers optimize for adaptive weights on regular dyadic grids refined uniformly. In contrast, our focus here is in the adaptive subdivision procedure.

6.4. **Numerical examples with the complete schemes** $(\mathscr{LS})_{TV-Fid_\omega}$ **and** $(\mathscr{LS})_{TGV-Fid_\omega}$**.**

In Figures 3, 4, 5, and 6, we present some illustrative examples resulting from the application of Algorithm 1 with $\ell_{\max} = 4$ to several images, for both TV and TGV regularization and optimizing for one adaptive parameter in the fidelity term, which is also shown along with the partitions overlaid on the noisy input images. In these, we generally see that the adapted fidelity parameter λ is higher in areas with finer details. Peak signal to noise ratios and SSIM values for each case are summarized in

Table 1. In all cases, *TGV* with adaptive fidelity produces the best results by these metrics, but there are several instances where the gains are very marginal or there are even ties with the corresponding adaptive *TV* results. Nevertheless, it may be argued that even in these cases the *TGV* results are more visually appealing due to reduced staircasing.

For the simple example of Figure 3, some more direct observations can be made. In it, we see that the spatially adaptive results manage to better preserve the fine structures inside the main object, while TGV greatly diminishes staircasing in regions where the original image is nearly linear. Observe that unlike the fine structures, the boundaries of the main object consisting of a sharp discontinuity along an interface with low curvature do not necessarily force further subdivision, as expected for *TV* or *TGV* regularization.

The synthetic image used in Figure 3 was created by the authors for this article. The lighthouse and parrot examples in Figures 4 and 6 have been cropped and converted to grayscale from images in the Kodak Lossless Image Suite. The cameraman image of Figure 5 is very widely used, but to our knowledge its origin is not quite clear.

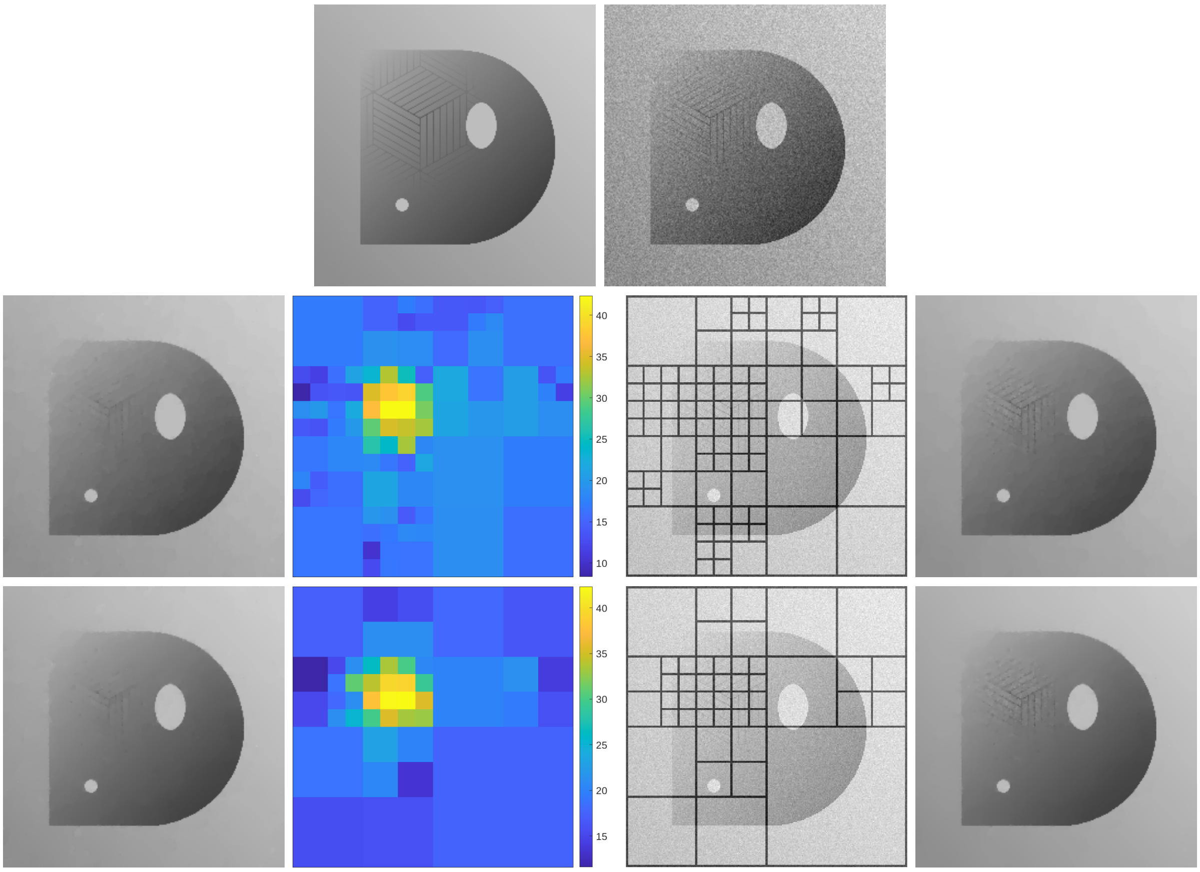

Figure 3. Synthetic example. Top row: Clean and noisy images u_c, u_η. Middle row, left to right: TV result with global parameter, partition and spatially-dependent λ arising from Algorithm 1, and corresponding result with weighted fidelity. Bottom row: TGV results, same order as in the middle row and with $\alpha_0 = 1, \alpha_1 = 10$.

Acknowledgements

The work of Elisa Davoli has been supported by the Austrian Science Fund (FWF) through grants F65, V 662, Y1292, and I 4052. Rita Ferreira was partially supported by King Abdullah University of Science and Technology (KAUST) baseline funds and KAUST OSR-CRG2021-4674. Irene Fonseca was partially supported under NSF-DMS1906238 and NSF-DMS2205627. A portion of this work was completed while José A. Iglesias was employed at the Johann Radon Institute for Computational and

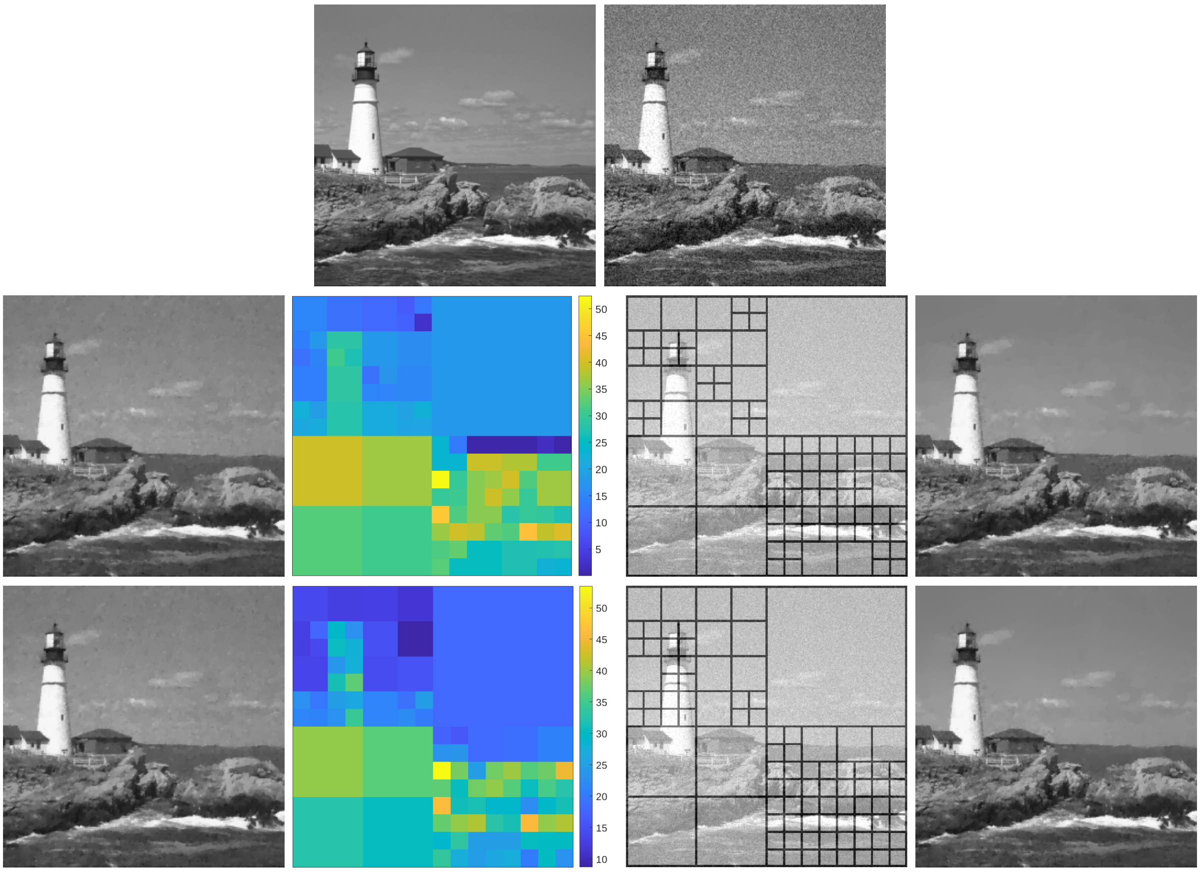

FIGURE 4. Lighthouse example. Top row: Clean and noisy images u_c, u_η. Middle row, left to right: TV result with global parameter, partition and spatially-dependent λ arising from Algorithm 1, and corresponding result with weighted fidelity. Bottom row: TGV results, same order as in the middle row and with $\alpha_0 = 1, \alpha_1 = 10$.

	Noisy	TV global	TV adaptive	TGV global	TGV adaptive
Synthetic	$26.05, 0.349$	$38.74, 0.946$	$39.41, 0.957$	$39.02, 0.949$	$39.80, 0.961$
Lighthouse	$24.64, 0.496$	$30.42, 0.853$	$30.82, 0.886$	$30.43, 0.854$	$30.88, 0.889$
Cameraman	$28.40, 0.642$	$32.86, 0.893$	$33.54, 0.925$	$32.86, 0.893$	$33.56, 0.925$
Parrot	$24.67, 0.447$	$31.86, 0.880$	$32.37, 0.898$	$31.87, 0.883$	$32.45, 0.899$

TABLE 1. PSNR and SSIM values for the examples of Figures 3, 4, 5, and 6.

Applied Mathematics (RICAM) of the Austrian Academy of Sciences (ÖAW), during which his work was partially supported by the State of Upper Austria.

All authors are thankful to Ilaria Perugia for some insightful comments on the topic of mesh refinements, as well as to Carolin Kreisbeck for some interesting discussions on the topic of box constraints.

REFERENCES

[1] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of bounded variation and free discontinuity problems.* Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York, 2000.

[2] G. Aubert and P. Kornprobst. *Mathematical problems in image processing. Partial differential equations and the calculus of variations. Foreword by Olivier Faugeras. 2nd ed.* New York, NY: Springer, 2006.

[3] A. Baldi. Weighted BV functions. *Houston J. Math.*, 27(3):683–705, 2001.

[4] L. Bogensperger, A. Chambolle, and T. Pock. Convergence of a piggyback-style method for the differentiation of solutions of standard saddle-point problems. *SIAM J. Math. Data Sci.*, 4(3):1003–1030, 2022.

[5] K. Bredies. Recovering piecewise smooth multichannel images by minimization of convex functionals with total generalized variation penalty. In A. Bruhn, T. Pock, and X.-C. Tai, editors, *Efficient Algorithms for Global Optimization Methods in Computer Vision*, pages 44–77. Springer Berlin Heidelberg, 2014.

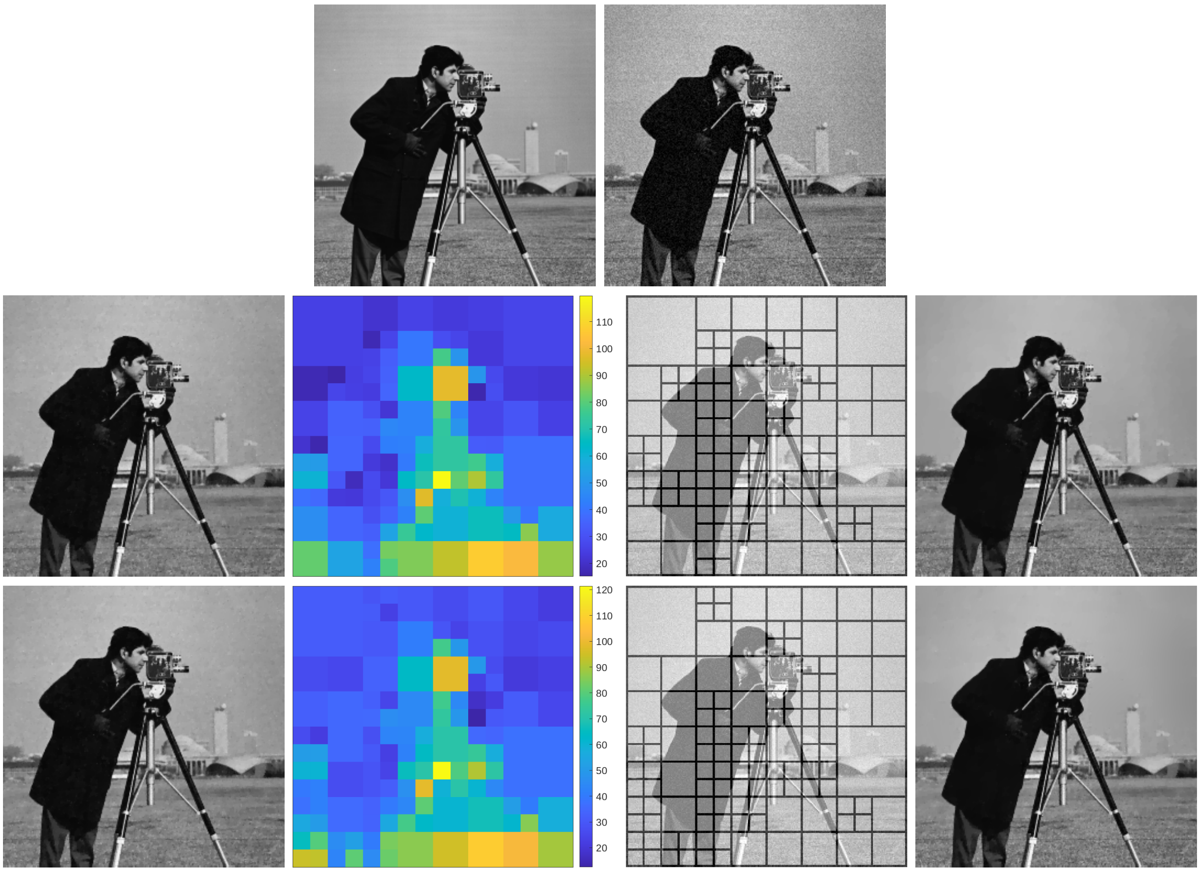

FIGURE 5. Cameraman example. Top row: Clean and noisy images u_c, u_η. Middle row, left to right: TV result with global parameter, partition and spatially-dependent λ arising from Algorithm 1, and corresponding result with weighted fidelity. Bottom row: TGV results, same order as in the middle row and with $\alpha_0 = 1, \alpha_1 = 10$.

[6] K. Bredies, E. Chenchene, and A. Hosseini. A hybrid proximal generalized conditional gradient method and application to total variation parameter learning. Preprint arXiv:2211.00997 [math.OC], 2022.

[7] K. Bredies and M. Holler. A TGV-based framework for variational image decompression, zooming, and reconstruction. Part II: Numerics. *SIAM J. Imaging Sci.*, 8(4):2851–2886, 2015.

[8] K. Bredies and M. Holler. Higher-order total variation approaches and generalisations. *Inverse Problems*, 36(12):123001, 128, 2020.

[9] K. Bredies, K. Kunisch, and T. Pock. Total generalized variation. *SIAM J. Imaging Sci.*, 3:492–526, 2010.

[10] K. Bredies, K. Kunisch, and T. Valkonen. Properties of L^1-TGV^2: the one-dimensional case. *J. Math. Anal. Appl.*, 398(1):438–454, 2013.

[11] 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. *J. Math. Imaging Vision*, 55:343–369, 2016.

[12] L. Calatroni, C. Cao, J. C. De los Reyes, C.-B. Schönlieb, and T. Valkonen. Bilevel approaches for learning of variational imaging models. In *Variational Methods: In Imaging and Geometric Control*, pages 252–290. De Gruyter, 2017.

[13] C. S. Camfield. *Comparison of BV norms in weighted Euclidean spaces and metric measure spaces*. PhD thesis, University of Cincinnati, 2008.

[14] C. S. Camfield. Comparison of BV norms in weighted Euclidean spaces. *J. Anal.*, 18:83–97, 2010.

[15] A. Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imaging Vision*, 20(1-2):89–97, 2004. Special issue on mathematics and image analysis.

[16] A. Chambolle and P.-L. Lions. Image recovery via total variation minimization and related problems. *Numer. Math.*, 76(2):167–188, 1997.

[17] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vision*, 40(1):120–145, 2011.

[18] A. Chambolle and T. Pock. Learning consistent discretizations of the total variation. *SIAM J. Imaging Sci.*, 14(2):778–813, 2021.

[19] T. Chan, A. Marquina, and P. Mulet. High-order total variation-based image restoration. *SIAM J. Sci. Comput.*, 22:503–516, 2000.

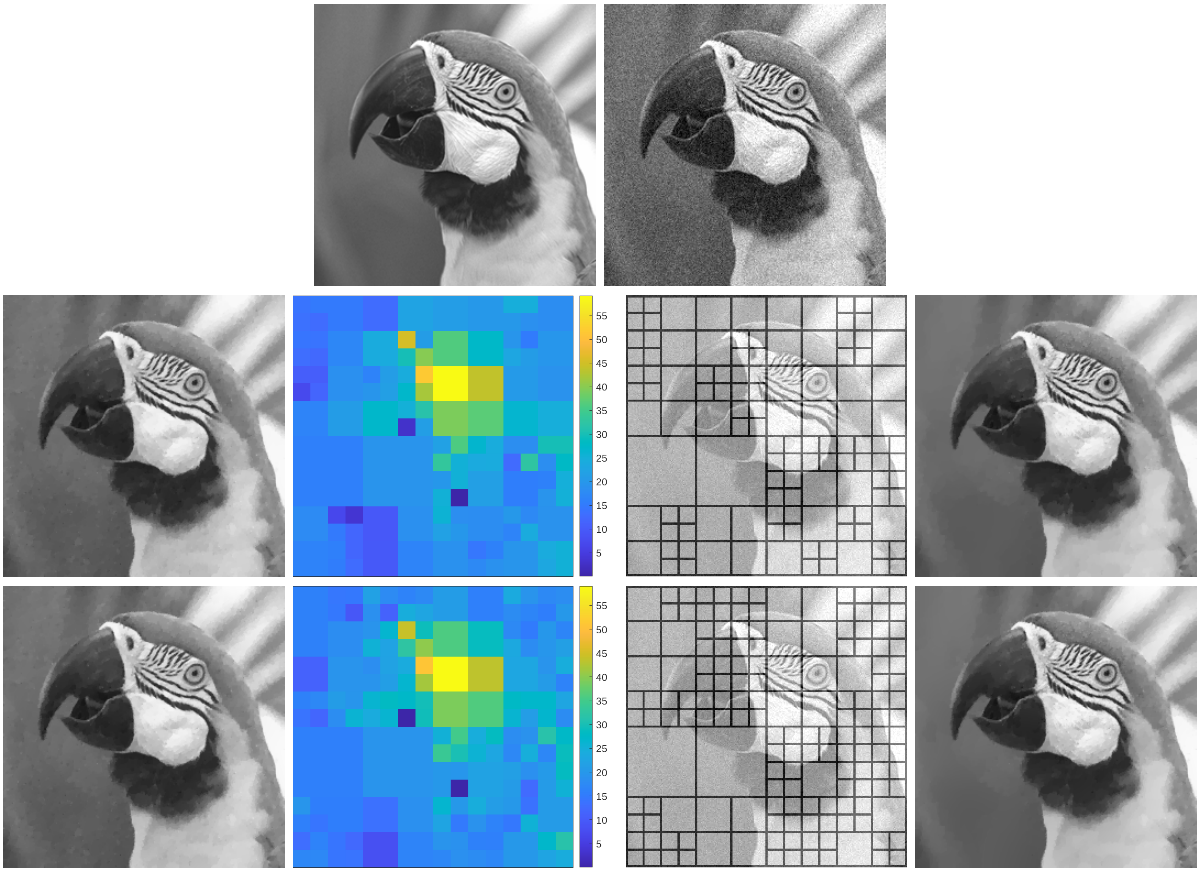

FIGURE 6. Parrot example. Top row: Clean and noisy images u_c, u_η. Middle row, left to right: TV result with global parameter, partition and spatially-dependent λ arising from Algorithm 1, and corresponding result with weighted fidelity. Bottom row: TGV results, same order as in the middle row and with $\alpha_0 = 1, \alpha_1 = 10$.

[20] T. F. Chan, S. H. Kang, and J. Shen. Total variation denoising and enhancement of color images based on the CB and HSV color models. *Journal of Visual Communication and Image Representation*, 12(4):422–435, 2001.
[21] Y. Chen, T. Pock, R. Ranftl, and H. Bischof. Revisiting loss-specific training of filter-based MRFs for image restoration. In J. Weickert, M. Hein, and B. Schiele, editors, *Pattern Recognition*, pages 271–281. Springer Berlin Heidelberg, 2013.
[22] Y. Chen, R. Ranftl, and T. Pock. Insights into analysis operator learning: from patch-based sparse models to higher order MRFs. *IEEE Trans. Image Process.*, 23(3):1060–1072, 2014.
[23] C. V. Chung, J. C. De los Reyes, and C. B. Schönlieb. Learning optimal spatially-dependent regularization parameters in total variation image denoising. Technical report, ModeMat, 2016.
[24] G. Dal Maso. *An introduction to* Γ*-convergence*. Progress in Nonlinear Differential Equations and their Applications, 8. Birkhäuser Boston Inc., Boston, MA, 1993.
[25] G. Dal Maso, I. Fonseca, G. Leoni, and M. Morini. A higher order model for image restoration: the one-dimensional case. *SIAM J. Math. Anal.*, 40:2351–2391, 2009.
[26] E. Davoli, I. Fonseca, and P. Liu. Adaptive image processing: first order PDE constraint regularizers and a bilevel training scheme. *J. Nonlinear Sci.*, 33(3):Paper No. 41, 38, 2023.
[27] E. Davoli and P. Liu. One dimensional fractional order TGV: Gamma-convergence and bilevel training scheme. *Commun. Math. Sci.*, 16(1):213–237, 2018.
[28] E. Davoli, Ferreira R., C. Kreisbeck, and H. Schönberger. Structural changes in nonlocal denoising models arising through bi-level parameter learning. *Applied Mathematics and Optimization*, 88:Art. 9, 2023.
[29] J. C. De los Reyes. Bilevel imaging learning problems as mathematical programs with complementarity constraints: Reformulation and theory. Preprint arXiv:2110.02273 [math.OC], 2021.
[30] J. C. De los Reyes and C.-B. Schönlieb. Image denoising: learning the noise model via nonsmooth PDE-constrained optimization. *Inverse Probl. Imaging*, 7(4):1183–1214, 2013.
[31] J. C. De Los Reyes, C.-B. Schönlieb, and T. Valkonen. The structure of optimal parameters for image restoration problems. *J. Math. Anal. Appl.*, 434(1):464–500, 2016.
[32] J. C. De los Reyes, C.-B. Schönlieb, and T. Valkonen. Bilevel parameter learning for higher-order total variation regularisation models. *J. Math. Imaging Vision*, 57(1):1–25, 2017.

[33] J. C. De los Reyes and D. Villacís. Optimality conditions for bilevel imaging learning problems with total variation regularization. *SIAM J. Imaging Sci.*, 15(4):1646–1689, 2022.

[34] J. Domke. Generic methods for optimization-based modeling. In N. D. Lawrence and M. Girolami, editors, *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*, volume 22 of *Proceedings of Machine Learning Research*, pages 318–326, La Palma, Canary Islands, 21–23 Apr 2012. PMLR.

[35] M. Hintermüller, K. Papafitsoros, C. N. Rautenberg, and H. Sun. Dualization and automatic distributed parameter selection of total generalized variation via bilevel optimization. *Numer. Funct. Anal. Optim.*, 43(8):887–932, 2022.

[36] 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. *J. Math. Imaging Vision*, 59(3):515–533, 2017.

[37] M. Hintermüller and T. Wu. Bilevel optimization for calibrating point spread functions in blind deconvolution. *Inverse Problems in Imaging*, 4(9):1139–1169, 2015.

[38] J. A. Iglesias and D. Walter. Extremal points of total generalized variation balls in 1D: characterization and applications. *J. Convex Anal.*, 29(4):1251–1290, 2022.

[39] K. Jalalzai. *Regularization of inverse problems in image processing*. PhD thesis, Ecole Polytechnique X, 2012.

[40] K. Kunisch and T. Pock. A bilevel optimization approach for parameter learning in variational models. *SIAM J. Imaging Sci.*, 2(6):938–983, 2012.

[41] P. Liu. *Variational and PDE Methods for Image Processing*. PhD thesis, Carnegie-Mellon University, 2017.

[42] P. Liu. Adaptive image processing: a bilevel structure learning approach for mixed-order total variation regularizers. Preprint arXiv:1902.01122 [math.OC], 2019.

[43] P. Liu and X. Y. Lu. Real order (an)-isotropic total variation in image processing - Part I: Analytical analysis and functional properties. Preprint arXiv:1805.06761 [math.AP], 2018.

[44] P. Liu and X. Y. Lu. Real order (an)-isotropic total variation in image processing - Part II: Learning of optimal structures. Preprint arXiv:1903.08513 [math.OC], 2019.

[45] P. Liu and C.-B. Schönlieb. Learning optimal orders of the underlying euclidean norm in total variation image denoising. Preprint arXiv:1903.11953 [math.AP], 2019.

[46] V. Pagliari, K. Papafitsoros, B. Raiţă, and A. Vikelis. Bilevel Training Schemes in Imaging for Total Variation–Type Functionals with Convex Integrands. *SIAM J. Imaging Sci.*, 15(4):1690–1728, 2022.

[47] K. Papafitsoros and K. Bredies. A study of the one dimensional total generalised variation regularisation problem. *Inverse Probl. Imaging*, 9:511–550, 2015.

[48] K. Papafitsoros and T. Valkonen. Asymptotic behaviour of total generalised variation. In *Scale space and variational methods in computer vision*, volume 9087 of *Lecture Notes in Comput. Sci.*, pages 720–714. Springer, Cham, 2015.

[49] T. Pock. On parameter learning in variational models. In *International Symposium on Mathematical Programming*, 2012.

[50] C. Pöschl and O. Scherzer. Exact solutions of one-dimensional total generalized variation. *Commun. Math. Sci.*, 13(1):171–202, 2015.

[51] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60(1-4):259–268, 1992.

[52] M. F. Tappen, C. Liu, E. Adelson, and W. T. Freeman. Learning gaussian conditional random fields for low-level vision. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007.

[53] R. Temam. *Problèmes mathématiques en plasticité*, volume 12 of *Méthodes Mathématiques de l'Informatique [Mathematical Methods of Information Science]*. Gauthier-Villars, Montrouge, 1983.

TU Wien, Institute of Analysis and Scientific Computing, Wiedner Hauptstrasse 8-10, 1040 Vienna, Austria
Email address: elisa.davoli@tuwien.ac.at

King Abdullah University of Science and Technology (KAUST), CEMSE Division, Thuwal 23955-6900, Saudi Arabia
Email address: rita.ferreira@kaust.edu.sa

Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA
Email address: fonseca@andrew.cmu.edu

Department of Applied Mathematics, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
Email address: jose.iglesias@utwente.nl