ADAPTIVE IMAGE PROCESSING: FIRST ORDER PDE CONSTRAINT REGULARIZERS AND A BILEVEL TRAINING SCHEME

ELISA DAVOLI

Institute of Analysis and Scientific Computing, TU Wien
Wiedner Hauptstrasse 8-10, 1040 Vienna, Austria

IRENE FONSECA

Center of Nonlinear Analysis, Department of Mathematics,
Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA, 15213, USA

PAN LIU

Department of Radiology,
First Medical Center of Chinese PLA General Hospital, Beijing 100853, China

November 13, 2020

Abstract. A bilevel training scheme is used to introduce a novel class of regularizers, providing a unified approach to standard regularizers TGV^2 and $NsTGV^2$. Optimal parameters and regularizers are identified, and the existence of a solution for any given set of training imaging data is proved by Γ-convergence. Explicit examples and numerical results are given.

Contents

E-mail addresses: elisa.davoli@tuwien.ac.at, fonseca@andrew.cmu.edu, dragonrider.liupan@gmail.com.

2010 *Mathematics Subject Classification.* 26B30, 94A08, 47J20.

Key words and phrases. image processing, optimal training scheme, first order differential operators, Γ-convergence.

1. Introduction

Image processing aims at the reconstruction of an original "clean" image starting from a "distorted one", namely from a datum which has been deteriorated or corrupted by noise effects or damaged digital transmission. The key idea of variational formulations in image-processing consists in rephrasing this problem as the minimization of an underlying functional of the form

$$\mathcal{I}(u) := \|u - u_\eta\|^2_{L^2(Q)} + \mathcal{R}_\alpha(u),$$

where u_η is a given corrupted image, $Q := (-1/2, 1/2)^N$ is the N-dimensional unit square (in image processing we usually take $N = 2$, i.e., Q represents the domain of a square image) and $\mathcal{R}_\alpha$ is a regularizing functional, with α denoting the intensity parameter (which could be a positive scalar or a vector). Minimizing the functional $\mathcal{I}$ allows to reconstruct a "clean" image based on the functional properties of the regularizer $\mathcal{R}_\alpha$.

Within the context of image denoising, for a fixed regularizer $\mathcal{R}_\alpha$ we seek to identify

$$u_{\alpha,\mathcal{R}} := \arg\min\left\{\|u - u_\eta\|^2_{L^2(Q)} + \mathcal{R}_\alpha(u) : \ u \in L^2(Q)\right\}.$$

An example is the ROF model ([30]), in which the regularizer is taken to be $\mathcal{R}_\alpha(u) := \alpha TV(u)$, where $TV(u)$ is the total variation of u (see, e.g. [1, Chapter 4]), $\alpha \in \mathbb{R}^+$ is the tuning parameter, and we have

$$u_{\alpha,TV} := \arg\min\left\{\|u - u_\eta\|^2_{L^2(Q)} + \alpha TV(u) : \ u \in L^2(Q)\right\}. \tag{1.1}$$

In view of the coercivity of the minimized functional, the natural class of competitors in (1.1) is $BV(Q)$, the space of real-valued functions of bounded variation in Q. The trade-off between the denoising effects of the ROF-functional and its feature-preserving capabilities is encoded by the tuning parameter $\alpha \in \mathbb{R}^+$. Indeed, high values of α lead to a strong penalization of the total variation of u, which in turn determines an over-smoothing effect and a resulting loss of information on the internal edges of the reconstructed image, while small values of α cause an unsatisfactory noise removal.

In order to determine the optimal α, say $\tilde{\alpha}$, in [13, 14] the authors proposed a bilevel training scheme, which was originally introduced in Machine Learning and later adopted by the imaging processing community (see [10, 11, 15, 31]). The bilevel training scheme is a semi-supervised training scheme that optimally adapts itself to the given "clean data". To be precise, let (u_η, u_c) be a pair of given images, where u_η represents the corrupted version and u_c stands for the original version, or the "clean" image. This training scheme searches for the optimal α so that the recovered image $u_{\alpha,TV}$, obtained in (1.1), minimizes the L^2-distance from the clean image u_c. An implementation of such training scheme, denoted by $(\mathcal{T})$, equipped with total variation TV is

$$\text{Level 1.} \qquad \tilde{\alpha} \in \arg\min\left\{\|u_{\alpha,TV} - u_c\|^2_{L^2(Q)} :\ \alpha \in \mathbb{R}^+\right\},$$

$$\text{Level 2.} \qquad u_{\alpha,TV} := \arg\min\left\{\|u - u_\eta\|^2_{L^2(Q)} + \alpha TV(u) :\ u \in BV(Q)\right\}. \qquad (\mathcal{T}\text{-L2})$$

An important observation is that the geometric properties of the regularizer TV play an essential role in the identification of the reconstructed image $u_{\alpha,TV}$ and may lead to a loss of some fine texture in the image. The choice of a given regularizer $\mathcal{R}_\alpha$ is indeed a crucial step in the formulation of the denoising problem: on the one hand, the structure of the regularizer must be such that the removal of undesired noise effects is guaranteed, and on the other hand the disruption of essential details of the image must be prevented. For these reasons, various choices of regularizers have been proposed in the literature. For example, the second order total generalized variation, TGV^2_α, defined as

$$TGV^2_\alpha(u) := \inf\Big\{\alpha_0\,|Du - v|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + \alpha_1\,|(\operatorname{sym}\nabla)v|_{\mathcal{M}_b(Q;\,\mathbb{R}^{N\times N})} :\ v \in L^1(Q;\mathbb{R}^N),\ (\operatorname{sym}\nabla)v \in \mathcal{M}_b(Q;\mathbb{R}^{N\times N})\Big\}, \qquad (1.2)$$

has been characterized in [4], where Du denotes the distributional gradient of u, $(\operatorname{sym}\nabla)v := (\nabla v + \nabla^T v)/2$, $\mathcal{M}_b(Q;\mathbb{R}^{N\times N})$ is the space of bounded Radon measures in Q with values in $\mathbb{R}^{N\times N}$, α_0 and α_1 are positive tuning parameters, and $\alpha := (\alpha_0, \alpha_1)$. A further possible choice for the regularizer is the non-symmetric counterpart of the TGV^2_α- seminorm defined above, namely the $NsTGV^2_\alpha$ functional (see e.g., [34, 33]). It has been shown that a reconstructed image presents several perks and drawbacks according to the different regularizers. An important question is thus how to identify the regularizer that might provide the best possible image denoising for a given class of corrupted images.

To address this problem, it is natural to use a straightforward modification of scheme $(\mathcal{T})$ by inserting different regularizers inside the training level 2 in ($\mathcal{T}$-L2). For example, one could set

$$\text{Level 1.} \qquad (\tilde{\mathcal{R}}_\alpha) := \arg\min\left\{\|u_{\alpha,\mathcal{R}} - u_c\|^2_{L^2(Q)} :\ \mathcal{R}_\alpha \in \left\{\alpha TV, TGV^2_\alpha, NsTGV^2_\alpha\right\}\right\}, \qquad (1.3)$$

$$\text{Level 2.} \qquad u_{\alpha,\mathcal{R}} := \arg\min\left\{\|u - u_\eta\|^2_{L^2(Q)} + \mathcal{R}_\alpha(u) :\ u \in L^1(Q)\right\}.$$

However, the finite number of possible choices for the regularizer within this training scheme would imply that the optimal regularizer $\tilde{\mathcal{R}}_\alpha$ would simply be determined by performing scheme $(\mathcal{T})$ finitely many times, at each time with a different regularizer $\mathcal{R}_\alpha$. In turn, some possible texture effects for which an "intermediate" (or interpolated) reconstruction between the one provided by, say, TGV^2_α and $NsTGV^2_\alpha$, might be more accurate, would then be neglected in the optimization procedure. Therefore, one main challenge in the setup of such a training scheme is to give a meaningful interpolation between the regularizers used in (1.3), and also to guarantee that the collection of the corresponding functional spaces exhibits compactness and lower semicontinuity properties.

The aim of this paper is threefold. First, we propose a novel class of image-processing operators, the PDE-constrained total generalized variation operators, or $PGV^2_{\alpha,\mathscr{B}}$, defined as

$$PGV^2_{\alpha,\mathscr{B}}(u) := \inf\Big\{\alpha_0\,|Du - v|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + \alpha_1\,|\mathscr{B}v|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})}:\\ v\in L^1(Q;\mathbb{R}^N),\ \mathscr{B}v\in\mathcal{M}_b(Q;\mathbb{R}^{N\times N})\Big\}, \quad (1.4)$$

for each $u \in L^1(Q;\mathbb{R}^N)$, where $\mathscr{B}$ is a linear differential operator (see Section 2 and Definition 3.5) and $\alpha := (\alpha_0, \alpha_1)$, with α_0, $\alpha_1 \in (0, +\infty)$. We also define the space of functions with bounded second order $PGV^2_{\alpha,\mathscr{B}}$-seminorms

$$BPGV^2_{\alpha,\mathscr{B}}(Q) := \left\{u \in L^1(Q) :\ PGV^2_{\alpha,\mathscr{B}}(u) < +\infty\right\}.$$

Note that if $\mathscr{B} := \mathrm{sym}\nabla$, then the operator $PGV^2_{\alpha,\mathscr{B}}$ defined in (1.4) coincides with the operator TGV^2_α mentioned in (1.2). In fact, we will show that, under appropriate assumptions (see Definition 6.1), the class described in (1.4) provides a unified approach to some of the standard regularizers mentioned in (1.3), generalizing the results in [8] (see Section 7.2). Moreover, the collection of functionals described in (1.4) naturally incorporates the recent PDE-based approach to image denoising formulated in [2] via nonconvex optimal control problem, thus offering a very general and abstract framework to simultaneously describe a variety of different image-processing techniques.

The second main goal of this article is the study of a training scheme optimizing the trade-off between effective reconstruction and fine image-detail preservation. That is, we propose a new bilevel training scheme that simultaneously yields the optimal regularizer $PGV^2_{\alpha,\mathscr{B}}(u)$ in the class described in (1.4) and an optimal tuning parameter α, so that the corresponding reconstructed image $u_{\alpha,\mathscr{B}}$, obtained in Level 2 of the $(\mathcal{T}^2_\theta)$-scheme (see $(\mathcal{T}^2_\theta$-L2) below), minimizes the L^2-distance from the original clean image u_c. To be precise, in Sections 3, 4, and 5 we study the improved training scheme $\mathcal{T}^2_\theta$ for $\theta \in (0,1)$, defined as follows

Level 1. $$(\tilde\alpha, \tilde{\mathscr{B}}) := \arg\min\left\{\|u_c - u_{\alpha,\mathscr{B}}\|^2_{L^2(Q)} :\ \alpha \in \left[\theta, \frac{1}{\theta}\right]^2,\ \mathscr{B}\in\Sigma\right\}, \quad (\mathcal{T}^2_\theta\text{-L1})$$

Level 2. $$u_{\alpha,\mathscr{B}} := \arg\min\left\{\|u - u_\eta\|^2_{L^2(Q)} + PGV^2_{\alpha,\mathscr{B}}(u),\ u \in BPGV^2_{\alpha,\mathscr{B}}(Q)\right\}, \quad (\mathcal{T}^2_\theta\text{-L2})$$

where Σ is an infinite collection of first order linear differential operators $\mathscr{B}$ (see Definition 3.4 and Definition 5.1). We prove the existence of optimal solutions to $(\mathcal{T}^2_\theta$-L1) by showing that the functional

$$\mathcal{I}_{\alpha,\mathscr{B}}(u) := \|u - u_\eta\|^2_{L^2} + PGV^2_{\alpha,\mathscr{B}}(u) \quad (1.5)$$

is continuous in the L^1 topology, in the sense of Γ-convergence, with respect to the parameters α and the operators $\mathscr{B}$ (see Theorem 4.2). A simplified statement of our main result (see Theorem 5.4) is the following.

Theorem 1.1. *Let $\theta \in (0,1)$. Then, the training scheme $(\mathcal{T}^2_\theta)$ admits at least one solution $(\tilde\alpha, \tilde{\mathscr{B}}) \in \left[\theta, \frac{1}{\theta}\right]^2 \times \Sigma$, and provides an associated optimally reconstructed image $u_{\tilde\alpha,\tilde{\mathscr{B}}} \in BV(Q)$.*

The collection Σ of operators $\mathscr{B}$ used in $(\mathcal{T}^2_\theta$-L1) has to satisfy several natural regularity and ellipticity assumptions, which are fulfilled by $\mathscr{B} := \nabla$ and $\mathscr{B} := \mathrm{sym}\,\nabla$ (see Section 7.2.1). The general requirements on $\mathscr{B}$ that allow scheme $(\mathcal{T}^2_\theta)$ to have a solution are listed on Assumptions 3.2 and 3.3. Later in Section 6, as the third main contribution of this article, we provide in Definition 6.1 a collection of operators $\mathscr{B}$ satisfying Assumptions 3.2 and 3.3. A simplified statement of our result is the following (see Theorem 6.3 for the detailed formulation).

Theorem 1.2. *Let $\mathscr{B}$ be a first order differential operator such that there exists a differential operator $\mathscr{A}$ for which $(\mathscr{A}, \mathscr{B})$ is a training operator pair, namely $\mathscr{A}$ admits a fundamental solution having suitable regularity assumptions, and the pair $(\mathscr{A}, \mathscr{B})$ fulfills a suitable integration-by-parts formula (see Definition 6.1 for the precise conditions). Then $\mathscr{B}$ is such that the training scheme $(\mathcal{T}_\theta^2)$ admits a solution.*

The requirements collected in Definition 6.1 and the analysis in Section 6 move from the observation that a fundamental property that the admissible operators $\mathscr{B}$ must satisfy is to ensure that the set of maps $v \in L^1(Q;\mathbb{R}^N)$ such that $\mathscr{B}v$ is a bounded Radon measure (henceforth denoted by $BV_{\mathscr{B}}(Q;\mathbb{R}^N)$) must embed compactly in $L^1(Q;\mathbb{R}^N)$. In the case in which $\mathscr{B}$ coincides with ∇ or $\operatorname{sym}\nabla$, a crucial ingredient is Kolmogorov-Riesz compactness theorem (see [7, Theorem 4.26] and Proposition 6.4). In particular, for $\mathscr{B} = \operatorname{sym}\nabla$ the key point of the proof is to guarantee that bounded sets $\mathcal{F} \subset BD(Q)$ satisfy

$$\lim_{|h|\to 0} \|\tau_h f - f\|_{L^1(\mathbb{R}^N)} = 0 \text{ uniformly in } \mathcal{F},$$

where $\tau_h f(\cdot) := f(\cdot - h)$. This in turn relies on the formal computation

$$\begin{aligned} \tau_h f - f = \delta_h * f - f = \delta_h * (\delta * f) - (\delta * f) = \delta_h * ((\operatorname{curlcurl}\phi) * f) - (\operatorname{curlcurl}\phi) * f \\ = \operatorname{curlcurl}((\delta_h * \phi - \phi) * f), \end{aligned}$$

where ϕ is a fundamental solution for $\operatorname{curlcurl}$, and where δ and δ_h denote the Dirac deltas centered in the origin and in h, respectively. In the case in which $\mathscr{B} = \operatorname{sym}\nabla$ the conclusion then follows from the fact that one can perform an "integration by parts" in the right-hand side of the above formula, and estimate the quantity $\operatorname{curlcurl}((\delta_h * \phi - \phi) * f)$ by means of the total variation of $(\operatorname{sym}\nabla)f$ and owing to the regularity of the fundamental solution of $\operatorname{curlcurl}$. The operator $\mathscr{A}$ in Theorem 1.2 plays the role of $\operatorname{curlcurl}$ in the case in which $\operatorname{sym}\nabla$ is replaced by a generic operator $\mathscr{B}$. Definition 6.1 is given in such a way as to guarantee that the above formal argument is rigorously justified for a pair of operators $(\mathscr{A}, \mathscr{B})$.
Finally, in Section 7.2 we give some explicit examples to show that our class of regularizers $PGV^2_{\alpha,\mathscr{B}}$ includes the seminorms TGV^2_α and $NsTGV^2_\alpha$, as well as smooth interpolations between them.

We remark that the task of determining not only the optimal tuning parameter but also the optimal regularizer for given training image data (u_η, u_c), has been undertaken in [12] where we have introduced one dimensional real order TGV^r regularizers, $r \in [1, +\infty)$, as well as a bilevel training scheme that simultaneously provides the optimal intensity parameters and order of derivation for one-dimensional signals.

Our analysis is complemented by very first numerical simulations of the proposed bilevel training scheme. Although this work focuses mainly on the theoretical analysis of the operators $PGV^2_{\alpha,\mathscr{B}}$ and on showing the existence of optimal results for the training scheme $(\mathcal{T}_\theta^2)$, in Section 7.3 a primal-dual algorithm for solving ($\mathcal{T}_\theta^2$-L2) is discussed, and some preliminary numerical examples, such as image denoising, are provided.

With this article we initiate our study of the combination of PDE-constraints and bilevel training schemes in image processing. Future goals will be:

- the construction of a finite grid approximation in which the optimal result $(\tilde{\alpha}, \tilde{\mathscr{B}})$ for the training scheme $(\mathcal{T}_\theta^2)$ can be efficiently determined, with an estimation of the approximation accuracy;
- spatially dependent differential operators and multi-layer training schemes. This will allow to specialize the regularization according to the position in the image, providing a more accurate analysis of complex textures and of images alternating areas with finer details with parts having sharpest contours (see also [23]).

This paper is organized as follows: in Section 2 we collect some notations and preliminary results. In Section 3 we analyze the main properties of the $PGV^2_{\alpha,\mathscr{B}}$-seminorms. The Γ-convergence result and the bilevel training scheme are the subjects of Sections 4 and 5, respectively. We point out that the results in Sections 3 and 4 are direct generalizations of the works in [5, 3]. The novelty of our approach consists in providing a slightly stronger analysis of the behavior of the functionals in (1.5) by showing not only convergence of minimizers under convergence of parameters and regularizers, but exhibiting also a complete Γ-convergence result.
The expert Reader might skip Sections 3–5, and proceed directly with the content of Section 6. Section 6 is devoted to the analysis of the space $BV_{\mathscr{B}}$ for suitable differential operators $\mathscr{B}$. The numerical implementation of some explicit examples is performed in Section 7.3.

2. Notations and preliminary results

We collect below some notation that will be adopted in connection with differential operators. Let $N \in \mathbb{N}$ be given, and let $Q := (-1/2, 1/2)^N$ be the unit open cube in $\mathbb{R}^N$ centered in the origin and with sides parallel to the coordinate axes. $\mathbb{M}^{N^3}$ is the space of real tensors of order $N \times N \times N$. Also, $\mathcal{D}'(Q, \mathbb{R}^N)$ and $\mathcal{D}'(Q, \mathbb{R}^{N\times N})$ stand for the spaces of distributions with values in $\mathbb{R}^N$ and $\mathbb{R}^{N\times N}$, respectively, and $\mathbb{R}^N_+$ denotes the set of vectors in $\mathbb{R}^N$ having positive entries.

For every open set $U \subset \mathbb{R}^N$, the notation $\mathscr{B}$ will be used for first order differential operators $\mathscr{B} : \mathcal{D}'(U; \mathbb{R}^N) \to \mathcal{D}'(U; \mathbb{R}^{N\times N})$ defined as

$$(\mathscr{B}v)_{lj} := \sum_{i,k=1}^{N} B^i_{ljk} \frac{\partial}{\partial x_i} v_k \quad \text{for every } v \in \mathcal{D}'(U; \mathbb{R}^N), \quad l, j = 1, \dots, N, \tag{2.1}$$

where $\frac{\partial}{\partial x_i}$ denotes the distributional derivative with respect to the i-th variable, and where $B^i \in \mathbb{M}^{N^3}$ for each $i = 1, \dots, N$. We additionally write the symbol of $\mathscr{B}$ as

$$\mathbb{B}[\xi] := \sum_{i=1}^{N} \xi_i B^i \quad \text{for every } \xi = (\xi_1, \dots, \xi_N) \in \mathbb{S}^{N-1}. \tag{2.2}$$

Given a sequence $\{\mathscr{B}_n\}_{n=1}^\infty$ of first order differential operators and a first order differential operator $\mathscr{B}$, with coefficients $\{B^i_n\}_{n=1}^\infty$ and B^i, $i = 1, \dots, N$, respectively, we say that $\mathscr{B}_n \to \mathscr{B}$ in ℓ^∞ if

$$\|\mathscr{B}_n - \mathscr{B}\|_{\ell^\infty} := \sum_{i=1}^{N} \left\| B^i_n - B^i \right\| \to 0, \tag{2.3}$$

where for $B \in \mathbb{M}^{N^3}$, $\|B\|$ stands for its Euclidean norm.

3. The space of functions with bounded PGV - seminorm

3.1. The space $BV_{\mathscr{B}}$ and the class of admissible operators. We generalize the standard total variation seminorm by using first order differential operators $\mathscr{B}\colon \mathcal{D}'(Q;\mathbb{R}^N) \to \mathcal{D}'(Q;\mathbb{R}^{N\times N})$ in the form (2.1).

Definition 3.1. For every $l \in \mathbb{N}$, we define the space of tensor-valued functions $BV_{\mathscr{B}}(Q;\mathbb{R}^N)$ as

$$BV_{\mathscr{B}}(Q;\mathbb{R}^N) := \left\{u \in L^1(Q;\mathbb{R}^N) :\ \mathscr{B}u \in \mathcal{M}_b(Q,\mathbb{R}^{N\times N})\right\}, \tag{3.1}$$

and we equip it with the norm

$$\|u\|_{BV_{\mathscr{B}}(Q;\mathbb{R}^N)} := \|u\|_{L^1(Q;\mathbb{R}^N)} + |\mathscr{B}u|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})}. \tag{3.2}$$

We refer to [28, 29] for some recents results on $BV_{\mathscr{B}}$-spaces for elliptic and cancelling operators, as well as to [24] for a study of associated Young measures. We point out that in the same way in which BV spaces relate to $W^{1,p}$-spaces, the spaces $BV_{\mathscr{B}}$ are connected to the theory of $W^{1,p}_{\mathscr{B}}$-spaces, cf. [18, 19, 26, 27]. See also [21] for a related compensated-compactness study.
In order to introduce the class of admissible operators, we first list some assumptions on the operator $\mathscr{B}$.

Assumption 3.2. 1. The space $BV_{\mathscr{B}}(Q;\mathbb{R}^N)$ is a Banach space with respect to the norm defined in (3.1).

2. The space $C^\infty(\overline{Q},\mathbb{R}^N)$ is dense in $BV_{\mathscr{B}}(Q;\mathbb{R}^N)$ in the strict topology. In other words, for every $u \in BV_{\mathscr{B}}(Q;\mathbb{R}^N)$ there exists $\{u_n\}_{n=1}^\infty \subset C^\infty(\bar{Q};\mathbb{R}^N)$ such that

$$u_n \to u \text{ strongly in } L^1(Q;\mathbb{R}^N) \text{ and } |\mathscr{B}u_n|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})} \to |\mathscr{B}u|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})}.$$

3. (Compactness) The injection of $BV_{\mathscr{B}}(Q;\mathbb{R}^N)$ into $L^1(Q;\mathbb{R}^N)$ is compact.

We point out that, for $l = 1$, Requirement 3 above is satisfied for $\mathscr{B} := \nabla$.

The following compactness property applies to a collection of operators $\{\mathscr{B}_n\}_{n=1}^\infty$.

Assumption 3.3. Let $\{v_n, \mathscr{B}_n\}_{n=1}^\infty$ be such that $\mathscr{B}_n$ satisfies Assumption 3.2 for every $n \in \mathbb{N}$, and

$$\sup\left\{\|\mathscr{B}_n\|_{\ell^\infty} + \|v_n\|_{BV_{\mathscr{B}_n}(Q;\mathbb{R}^N)} :\ n \in \mathbb{N}\right\} < +\infty.$$

Then there exist $\mathscr{B}$ and $v \in BV_{\mathscr{B}}(Q;\mathbb{R}^N)$ such that, up to a subsequence (not relabeled),

$$v_n \to v \text{ strongly in } L^1(Q;\mathbb{R}^N),$$

and

$$\mathscr{B}_n v_n \overset{*}{\rightharpoonup} \mathscr{B}v \text{ weakly}^* \text{ in } \mathcal{M}_b(Q;\mathbb{R}^{N\times N}).$$

Definition 3.4. We denote by Π the collection of operators $\mathscr{B}$ defined in (2.1), with finite dimensional null-space $\mathcal{N}(\mathscr{B})$, and satisfying Assumption 3.2.

In Section 6 we will exhibit a subclass of operators $\mathscr{B} \in \Pi$ additionally fulfilling the compactness and closure Assumption 3.3.

3.2. **The PGV- total generalized variation.** We introduce below the definition of the PDE-constrained total generalized variation seminorms.

Definition 3.5. Let $u \in L^1(Q)$ be given. For every $\alpha = (\alpha_0, \alpha_1) \in \mathbb{R}^2_+$ and $\mathscr{B}\colon \mathcal{D}'(Q;\mathbb{R}^N) \to \mathcal{D}'(Q;\mathbb{R}^{N\times N})$, $\mathscr{B} \in \Pi$, we consider the seminorm

$$PGV^2_{\alpha,\mathscr{B}}(u) := \inf \left\{ \alpha_0 \, |Du - v|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + \alpha_1 \, |\mathscr{B}v|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})} : \ v \in BV_{\mathscr{B}}(Q;\mathbb{R}^N) \right\}, \tag{3.3}$$

where the space $BV_{\mathscr{B}}$ is introduced in Definition (3.1).

We note that for all $\alpha \in \mathbb{R}^2_+$, the seminorms $PGV^2_{\alpha,\mathscr{B}}$ are topologically equivalent. With a slight abuse of notation, in what follows we will write $PGV^2_{\mathscr{B}}$ instead of $PGV^2_{\alpha,\mathscr{B}}$ whenever the dependence of the seminorm on a specific multi-index $\alpha \in \mathbb{R}^2_+$ will not be relevant for the presentation of the results.

We introduce below the set of functions with bounded PDE-generalized variation-seminorms.

Definition 3.6. We define

$$BPGV^2_{\mathscr{B}}(Q) := \left\{ u \in L^1(Q) : \ PGV^2_{1,\mathscr{B}}(u) < +\infty \right\},$$

and we write

$$\|u\|_{BPGV^2_{\mathscr{B}}(Q)} := \|u\|_{L^1(Q)} + PGV^2_{1,\mathscr{B}}(u).$$

We next show that the $PGV^2_{\mathscr{B}}$-seminorm is finite if and only if the TV-seminorm is.

Proposition 3.7. Let $u \in L^1(Q)$ and recall $PGV^2_{\mathscr{B}}(u)$ from Definition 3.5. Then, $PGV^2_{\mathscr{B}}(u) < +\infty$ if and only if $u \in BV(Q)$.

Proof. We notice that by setting $v = 0$ in (3.3), we have

$$PGV^2_{\mathscr{B}}(u) \le |Du|_{\mathcal{M}_b(Q;\mathbb{R}^N)} \tag{3.4}$$

for every $u \in L^1(Q)$. Thus, if $u \in BV(Q)$ then $PGV^2_{\mathscr{B}}(u) < +\infty$.

Conversely, assume that $PGV^2_{\mathscr{B}}(u) < +\infty$. Then, there exists $\bar{v} \in BV_{\mathscr{B}}(Q)$ such that

$$PGV^2_{\mathscr{B}}(u) \ge |Du - \bar{v}|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + |\mathscr{B}\bar{v}|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})} - 1.$$

It suffices to observe that

$$|Du|_{\mathcal{M}_b(Q;\mathbb{R}^N)} \le |Du - \bar{v}|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + \|\bar{v}\|_{L^1(Q;\mathbb{R}^N)} \le PGV^2_{\mathscr{B}}(u) + 1 + \|\bar{v}\|_{L^1(Q;\mathbb{R}^N)} < +\infty.$$

□

We prove that the infimum problem in the right-hand side of (3.3) has a solution.

Proposition 3.8. Let $u \in BV(Q)$. Then, for $\alpha \in \mathbb{R}^2_+$ there exists a function $v \in BV_{\mathscr{B}}(Q;\mathbb{R}^N)$ attaining the infimum in (3.3).

Proof. Let $u \in BV(Q)$ and, without loss of generality, assume that $\alpha = (1,1)$. In view of Proposition 3.7 we have $PGV^2_{\mathscr{B}}(u) < +\infty$.

The existence of a minimizer $v \in L^1(Q;\mathbb{R}^N)$ with $\mathscr{B}v \in \mathcal{M}_b(Q;\mathbb{R}^{N\times N})$ follows from the Direct Method of the calculus of variations. Indeed, let $\{v_n\}_{n=1}^\infty \subset BV_{\mathscr{B}}(Q;\mathbb{R}^N)$ be such that

$$|Du - v_n|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + |\mathscr{B}v_n|_{\mathcal{M}_b(\Omega;\mathbb{R}^{N\times N})} \leq PGV^2_{\mathscr{B}}(u) + 1/n$$

for every $n \in \mathbb{N}$. Then,

$$\|v_n\|_{L^1(Q;\mathbb{R}^N)} \leq |Du - v_n|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + |Du|_{\mathcal{M}_b(Q;\mathbb{R}^N)} \leq PGV^2_{\mathscr{B}}(u) + |Du|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + 1/n, \tag{3.5}$$

and

$$|\mathscr{B}v_n|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})} \leq PGV^2_{\mathscr{B}}(u) + 1/n, \tag{3.6}$$

for every $n \in \mathbb{N}$. In view of Assumption 3.2, and together with (3.5) and (3.6), we obtain a function $v \in L^1(Q;\mathbb{R}^N)$ with $\mathscr{B}v \in \mathcal{M}_b(Q;\mathbb{R}^{N\times N})$ such that, up to the extraction of a subsequence (not relabeled), there holds

$$v_n \to v \quad \text{strongly in } L^1(Q;\mathbb{R}^N),$$

and

$$\liminf_{n\to\infty} |\mathscr{B}v_n|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})} \geq |\mathscr{B}v|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})}.$$

The minimality of v follows by lower-semicontinuity.

□

We close this section by studying the asymptotic behavior of the $PGV^2_{\mathscr{B}}$ seminorms in terms of the operator $\mathscr{B}$ for subclasses of Π satisfying Assumption 3.3.

Proposition 3.9. Let $u \in BV(Q)$. Let $\{\mathscr{B}_n\}_{n=1}^\infty \subset \Pi$ and $\{\alpha_n\}_{n=1}^\infty \subset \mathbb{R}^2_+$ be such that $\mathscr{B}_n \to \mathscr{B}$ in ℓ^∞ and

$$\alpha_n \to \alpha \in \mathbb{R}^2_+. \tag{3.7}$$

Assume that $\{\mathscr{B}_n\}_{n=1}^\infty$ satisfies Assumption 3.3. Then

$$\lim_{n\to\infty} PGV^2_{\alpha_n,\mathscr{B}_n}(u) = PGV^2_{\alpha,\mathscr{B}}(u).$$

Proof. We first claim that

$$\liminf_{n\to\infty} PGV^2_{\alpha_n,\mathscr{B}_n}(u) \geq PGV^2_{\alpha,\mathscr{B}}(u). \tag{3.8}$$

Indeed, by Proposition 3.8 for each $n \in \mathbb{N}$ there exists $v_n \in BV_{\mathscr{B}_n}(Q;\mathbb{R}^N)$ such that, setting $\alpha_n = (\alpha_n^0, \alpha_n^1)$,

$$PGV^2_{\alpha_n,\mathscr{B}_n}(u) = \alpha_n^0 \,|Du - v_n|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + \alpha_n^1 \,|\mathscr{B}_n v_n|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})}.$$

From (3.4) and (3.7),we see that

$$\alpha_n^0 \,|Du - v_n|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + \alpha_n^1 \,|\mathscr{B}_n v_n|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})} \leq \alpha_n^0 \,|Du| < +\infty,$$

which from (3.7) implies that $\sup\{\|v_n\|_{BV_{\mathscr{B}_n}(Q;\mathbb{R}^N)} + \|\mathscr{B}_n\|_{\ell^\infty}\}$ is finite. Therefore, by Assumption 3.3 there exist $\mathscr{B} \in \Pi$ and $v \in BV_{\mathscr{B}}(Q)$ such that $v_n \to v$ strongly in $L^1(Q;\mathbb{R}^N)$ and

$$\liminf_{n\to\infty} |\mathscr{B}_n v_n|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})} \geq |\mathscr{B}v|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})}. \tag{3.9}$$

Fix $\varepsilon > 0$. By (3.7), for n big enough there holds $\alpha_n^0 \geq (1-\varepsilon)\alpha_0$, and $\alpha_n^1 \geq (1-\varepsilon)\alpha_1$. Thus, by (3.9) we have

$$\liminf_{n\to\infty} PGV^2_{\alpha_n,\mathscr{B}_n}(u) = \liminf_{n\to\infty} \left[\alpha_n^0 \,|Du - v_n|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + \alpha_n^1 \,|\mathscr{B}_n v_n|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})}\right]$$

$$\geq (1-\varepsilon)\alpha_0 \liminf_{n\to\infty} |Du - v_n|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + (1-\varepsilon)\alpha_1 \liminf_{n\to\infty} |\mathscr{B}_n v_n|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})}$$
$$\geq (1-\varepsilon)\alpha_0 |Du - v|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + (1-\varepsilon)\alpha_1 |\mathscr{B}v|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})}$$
$$\geq (1-\varepsilon) PGV^2_{\alpha,\mathscr{B}}(u),$$

where in the last inequality we used (3.3). The arbitrariness of ε concludes the proof of (3.8).

We now claim that

$$\limsup_{n\to\infty} PGV^2_{\alpha_n,\mathscr{B}_n}(u) \leq PGV^2_{\alpha,\mathscr{B}}(u). \tag{3.10}$$

By Proposition 3.8 there exists $v \in BV_{\mathscr{B}}(Q;\mathbb{R}^N)$ such that

$$PGV^2_{\alpha,\mathscr{B}}(u) = \alpha_0 |Du - v|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + \alpha_1 |\mathscr{B}v|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})}.$$

In view of the density result in Assumption 3.2, Statement 2, we may assume that $v \in C^\infty(Q;\mathbb{R}^N)$ and, for $\varepsilon > 0$ small,

$$PGV^2_{\alpha,\mathscr{B}}(u) \geq \alpha_0 |Du - v|_{\mathcal{M}_b(\Omega;\mathbb{R}^N)} + \alpha_1 |\mathscr{B}v|_{\mathcal{M}_b(\Omega;\mathbb{R}^{N\times N})} - \varepsilon. \tag{3.11}$$

Since

$$PGV^2_{\alpha_n,\mathscr{B}_n}(u) \leq \alpha_n^0 |Du - v|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + \alpha_n^1 |\mathscr{B}_n v|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})},$$

we obtain

$$\limsup_{n\to\infty} PGV^2_{\alpha_n,\mathscr{B}_n}(u) \leq \alpha_0 |Du - v|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + \limsup_{n\to\infty} \alpha_n^1 |\mathscr{B}_n v|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})}$$
$$\leq \alpha_0 |Du - v|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + \alpha_1 |\mathscr{B}v|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})}$$
$$\leq PGV^2_{\alpha,\mathscr{B}}(u) + \varepsilon,$$

where in the last inequality we used (3.11). Claim (3.10) is now asserted by the arbitrariness of $\varepsilon > 0$.

□

4. Γ-convergence of functionals defined by PGV- total generalized variation seminorms

In this section we prove a Γ-convergence result with respect to the operator $\mathscr{B}$. For $r > 0$ we denote (see (2.3))

$$(\mathscr{B})_r := \{\mathscr{B}' \in \Pi : \ \|\mathscr{B}' - \mathscr{B}\|_{\ell^\infty} \leq r\}. \tag{4.1}$$

Throughout this section, let $u_\eta \in L^2(Q)$ be a given datum representing a corrupted image.

Definition 4.1. Let $\mathscr{B} \in \Pi$, $\alpha \in \mathbb{R}^2_+$. We define the functional $\mathcal{I}_{\alpha,\mathscr{B}} : L^1(Q) \to [0,+\infty]$ as

$$\mathcal{I}_{\alpha,\mathscr{B}}(u) := \begin{cases} \|u - u_\eta\|^2_{L^2(Q)} + PGV^2_{\alpha,\mathscr{B}}(u) & \text{if } u \in BV(Q), \\ +\infty & \text{otherwise.} \end{cases}$$

The following theorem is the main result of this section.

Theorem 4.2. *Let $\{\mathscr{B}_n\}_{n=1}^{\infty} \subset \Pi$ satisfy Assumption 3.3, and let $\{\alpha_n\}_{n=1}^{\infty} \subset \mathbb{R}^2_+$ be such that $\mathscr{B}_n \to \mathscr{B}$ in ℓ^∞ and $\alpha_n \to \alpha \in \mathbb{R}^2_+$. Then the functionals $\mathcal{I}_{\alpha_n,\mathscr{B}_n}$ satisfy the following compactness properties:*

(Compactness) *Let $u_n \in BV(Q)$, $n \in \mathbb{N}$, be such that*

$$\sup\{\mathcal{I}_{\alpha_n,\mathscr{B}_n}(u_n) : \ n \in \mathbb{N}\} < +\infty.$$

Then there exists $u \in BV(Q)$ such that, up to the extraction of a subsequence (not relabeled),

$$u_n \overset{*}{\rightharpoonup} u \text{ weakly}^* \text{ in } BV(Q).$$

Additionally, $\mathcal{I}_{\alpha_n,\mathscr{B}_n}$ Γ-converges to $\mathcal{I}_{\alpha,\mathscr{B}}$ in the L^1 topology. To be precise, for every $u \in BV(Q)$ the following two conditions hold:

(Liminf inequality) *If*

$$u_n \to u \text{ in } L^1(Q)$$

then

$$\mathcal{I}_{\alpha,\mathscr{B}}(u) \le \liminf_{n\to+\infty} \mathcal{I}_{\alpha_n,\mathscr{B}_n}(u_n).$$

(Recovery sequence) *For each $u \in BV(Q)$, there exists $\{u_n\}_{n=1}^{\infty} \subset BV(Q)$ such that*

$$u_n \to u \text{ in } L^1(Q)$$

and

$$\limsup_{n\to+\infty} \mathcal{I}_{\alpha_n,\mathscr{B}_n}(u_n) \le \mathcal{I}_{\alpha,\mathscr{B}}(u).$$

We subdivide the proof of Theorem 4.2 into two propositions.

For $\mathscr{B} \in \Pi$, we consider the projection operator

$$\mathbb{P}_{\mathscr{B}} : L^1(Q;\mathbb{R}^N) \to \mathcal{N}(\mathscr{B}).$$

Note that this projection operator is well defined owing to the assumption that $\mathcal{N}(\mathscr{B})$ is finite dimensional (see [7, page 38, Definition and Example 2] and [6, Subsection 3.1]).
Next we have an enhanced version of Korn's inequality.

Proposition 4.3. Let $\mathscr{B} \in \Pi$ and let $r > 0$. Then there exists a constant $C = C(\mathscr{B}, Q)$, depending only on $\mathscr{B}$ and on the domain Q, such that

$$\|v - \mathbb{P}_{\mathscr{B}'}(v)\|_{L^1(Q;\mathbb{R}^N)} \le C\,|\mathscr{B}'v|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})}, \tag{4.2}$$

for all $v \in L^1(Q)$ and $\mathscr{B}' \in (\mathscr{B})_r$.

Proof. Suppose that (4.2) fails. Then there exist sequences $\{\mathscr{B}_n\}_{n=1}^{\infty} \subset (\mathscr{B})_r$ and $\{v_n\}_{n=1}^{\infty} \subset L^1(Q)$ such that

$$\|v_n - \mathbb{P}_{\mathscr{B}_n}(v_n)\|_{L^1(Q;\mathbb{R}^N)} \ge n\,|\mathscr{B}_n v_n|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})}$$

for every $n \in \mathbb{N}$. Up to a normalization, we can assume that

$$\|v_n - \mathbb{P}_{\mathscr{B}_n}(v_n)\|_{L^1(Q;\mathbb{R}^N)} = 1 \text{ and } |\mathscr{B}_n v_n|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})} \le 1/n \tag{4.3}$$

for every $n \in \mathbb{N}$. Since $\{\mathscr{B}_n\}_{n=1}^{\infty} \subset (\mathscr{B})_r$, up to a subsequence (not relabeled), we have $\mathscr{B}_n \to \tilde{\mathscr{B}}$ in ℓ^∞, for some $\tilde{\mathscr{B}} \in (\mathscr{B})_r$. Next, let

$$\tilde{v}_n := v_n - \mathbb{P}_{\mathscr{B}_n}(v_n).$$

Note that for each $n \in \mathbb{N}$

$$\mathbb{P}_{\mathscr{B}_n}(\tilde{v}_n) = 0. \tag{4.4}$$

Thus, by (4.3) we have

$$\|\tilde{v}_n\|_{L^1(Q;\mathbb{R}^N)} = 1 \text{ and } |\mathscr{B}_n \tilde{v}_n|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})} \leq 1/n. \tag{4.5}$$

In view of Assumption 3.3, up to a further subsequence (not relabeled), there exists $\tilde{v} \in BV_{\tilde{\mathscr{B}}}(Q;\mathbb{R}^N)$ such that $\tilde{v}_n \to \tilde{v}$ strongly in $L^1(Q)$ and $|\tilde{\mathscr{B}}\tilde{v}|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})} = 0$. Moreover, in view of (4.5), we also have $\|\tilde{v}\|_{L^1(Q;\mathbb{R}^N)} = 1$.

Since the projection operator is Lipschitz, by (4.4) we have

$$\left\|\mathbb{P}_{\tilde{\mathscr{B}}}(\tilde{v})\right\|_{L^1(Q)} = \left\|\mathbb{P}_{\mathscr{B}_n}(\tilde{v}_n) - \mathbb{P}_{\tilde{\mathscr{B}}}(\tilde{v})\right\|_{L^1(Q)} \leq C\,\|\tilde{v} - \tilde{v}_n\|_{L^1(Q)} \to 0.$$

Thus, $\mathbb{P}_{\tilde{\mathscr{B}}}(\tilde{v}) = 0$. However, $|\tilde{\mathscr{B}}\tilde{v}|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})} = 0$ implies that $\tilde{v} \in \mathcal{N}[\tilde{\mathscr{B}}]$ with $\tilde{v} = P_{\tilde{\mathscr{B}}}(\tilde{v})$, and hence we must have $\tilde{v} = 0$, contradicting the fact that $\|\tilde{v}\|_{L^1(Q;\mathbb{R}^N)} = 1$. □

The following proposition is instrumental for establishing the liminf inequality.

Proposition 4.4. Let $\{\mathscr{B}_n\}_{n=1}^{\infty} \subset \Pi$ satisfy Assumption 3.3, and let $\{\alpha_n\}_{n=1}^{\infty} \subset \mathbb{R}^2_+$ be such that $\mathscr{B}_n \to \mathscr{B}$ in ℓ^∞ and $\alpha_n \to \alpha \in \mathbb{R}^2_+$. For every $n \in \mathbb{N}$ let $u_n \in BV(Q)$ be such that

$$\sup\{\mathcal{I}_{\alpha_n,\mathscr{B}_n}(u_n) : \ n \in \mathbb{N}\} < +\infty. \tag{4.6}$$

Then there exists $u \in BV(Q)$ such that, up to the extraction of a subsequence (not relabeled),

$$u_n \overset{*}{\rightharpoonup} u \text{ weakly}^* \text{ in } BV(Q) \tag{4.7}$$

and

$$\liminf_{n\to\infty} PGV^2_{\alpha_n,\mathscr{B}_n}(u_n) \geq PGV^2_{\alpha,\mathscr{B}}(u),$$

with

$$\liminf_{n\to\infty} \mathcal{I}_{\alpha_n,\mathscr{B}_n}(u_n) \geq \mathcal{I}_{\alpha,\mathscr{B}}(u).$$

Proof. Fix $r > 0$ and recall the definition of $(\mathscr{B})_r$ from (4.1). We claim that if r is small enough then there exists $C_r > 0$ such that

$$\|u\|_{BPGV^2_{\mathscr{B}'}(Q)} \leq \|u\|_{BV(Q)} \leq C_r\,\|u\|_{BPGV^2_{\mathscr{B}'}(Q)}, \tag{4.8}$$

for all $u \in BV(Q)$ and $\mathscr{B}' \in (\mathscr{B})_r$.

Indeed, by Definitions 3.5 and 3.6 we always have

$$\|u\|_{BPGV^2_{\mathscr{B}'}(Q)} \leq \|u\|_{BV(Q)},$$

for all $\mathscr{B}' \in \Pi$ and $u \in BV(Q)$.

The crucial step is to prove that the second inequality in (4.8) holds. Set

$$\mathcal{N}_r(\mathscr{B}) := \{\omega \in L^1(Q;\mathbb{R}^N) : \ \text{there exists } \mathscr{B}' \in (\mathscr{B})_r \text{ for which } \omega \in \mathcal{N}(\mathscr{B}')\}.$$

We claim that there exists $C>0$, depending on r, such that for each $u\in BV(Q)$ and $\omega\in\mathcal{N}_r(\mathscr{B})$ we have

$$|Du|_{\mathcal{M}_b(Q;\mathbb{R}^N)} \leq C\left(|Du-\omega|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + \|u\|_{L^1(Q)}\right). \tag{4.9}$$

Suppose that (4.9) fails. Then we find sequences $\{u_n\}_{n=1}^\infty\subset BV(Q)$ and $\{\omega_n\}_{n=1}^\infty\subset\mathcal{N}_r(\mathscr{B})$ such that

$$|Du_n|_{\mathcal{M}_b(Q;\mathbb{R}^N)} \geq n\left(|Du_n-\omega_n|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + \|u_n\|_{L^1(Q)}\right)$$

for every $n\in\mathbb{N}$. Thus, up to a normalization, we can assume that

$$|Du_n|_{\mathcal{M}_b(Q;\mathbb{R}^N)} = 1 \tag{4.10}$$

and

$$|Du_n-\omega_n|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + \|u_n\|_{L^1(Q)} \leq 1/n, \tag{4.11}$$

which implies that $u_n\to 0$ strongly in $L^1(Q)$ and

$$|Du_n-\omega_n|_{\mathcal{M}_b(Q;\mathbb{R}^N)} \to 0. \tag{4.12}$$

By (4.10) and (4.11), it follows that $|\omega_n|_{\mathcal{M}_b(Q;\mathbb{R}^N)}$ is uniformly bounded, and hence, up to a subsequence (not relabeled), there exists $\omega\in\mathcal{M}_b(Q;\mathbb{R}^N)$ such that $\omega_n\overset{*}{\rightharpoonup}\omega$ in $\mathcal{M}_b(Q;\mathbb{R}^N)$. For every $n\in\mathbb{N}$ let $\mathscr{B}'_n\in(\mathscr{B})_r$ be such that $\omega_n\in\mathcal{N}(\mathscr{B}'_n)$. Then $\mathscr{B}'_n\omega_n=0$ for all $n\in\mathbb{N}$. Since $\|\mathscr{B}'_n-\mathscr{B}\|_{\ell^\infty}<r$, in particular the sequence $\{\omega_n,\mathscr{B}'_n\}_{n=1}^\infty\subset L^1(\mathcal{M}_b(Q;\mathbb{R}^N))\times\Pi$ fulfills Assumption 3.3, and hence, upon extracting a further subsequence (not relabeled), there holds

$$\omega_n\to\omega_0 \quad \text{strongly in } L^1(Q;\mathbb{R}^N).$$

Additionally, since $u_n\to 0$ strongly in $L^1(Q)$, we infer that $Du_n\to 0$ in the sense of distributions. Therefore, by (4.12) we deduce that $\omega_0=0$. Using again (4.11), we conclude that

$$|Du_n|_{\mathcal{M}_b(Q;\mathbb{R}^N)}\to 0,$$

which contradicts (4.10). This completes the proof of (4.9).

We are now ready to prove the second inequality in (4.8), i.e.,

$$\|u\|_{BV(Q)} \leq C_r\,\|u\|_{BPGV^2_{\mathscr{B}'}(Q)} \tag{4.13}$$

for some constant $C_r>0$, and for all $\mathscr{B}'\in(\mathscr{B})_r$.

Fix $\mathscr{B}'\in(\mathscr{B})_r$, and by Proposition 3.8 let $v_{\mathscr{B}'}$ satisfy

$$PGV^2_{\mathscr{B}'}(u) = |Du-v_{\mathscr{B}'}|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + |\mathscr{B}'v_{\mathscr{B}'}|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})}. \tag{4.14}$$

Since $\mathbb{P}_{\mathscr{B}'}[v_{\mathscr{B}'}]\in\mathcal{N}_r(\mathscr{B})$, we have

$$\begin{aligned}
|Du|_{\mathcal{M}_b(Q;\mathbb{R}^N)} &\leq C(|Du-\mathbb{P}_{\mathscr{B}'}[v_{\mathscr{B}'}]|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + \|u\|_{L^1(Q)})\\
&\leq C(|Du-v_{\mathscr{B}'}|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + |v_{\mathscr{B}'}-\mathbb{P}_{\mathscr{B}'}[v_{\mathscr{B}'}]|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + \|u\|_{L^1(Q)})\\
&\leq C(|Du-v_{\mathscr{B}'}|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + C'\,|\mathscr{B}'v_{\mathscr{B}'}|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})} + \|u\|_{L^1(Q)})\\
&\leq (C+C')\left[|Du-v_{\mathscr{B}'}|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + |\mathscr{B}'v_{\mathscr{B}'}|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})} + \|u\|_{L^1(Q)}\right]\\
&= (C+C')\left[PGV^2_{\mathscr{B}'}(u) + \|u\|_{L^1(Q)}\right],
\end{aligned}$$

where in the first inequality we used (4.9), the third inequality follows by (4.2), and in the last equality we invoked (4.14). Defining $C_r := C + C' + 1$, we obtain

$$\|u\|_{BV(Q)} = \|u\|_{L^1(Q)} + |Du|_{\mathcal{M}_b(Q;\mathbb{R}^N)} \le C_r(PGV^2_{\mathscr{B}'}(u) + \|u\|_{L^1(Q)}) = C_r \|u\|_{BPGV^2_{\mathscr{B}'}(Q)}$$

and we conclude (4.13).

Now we prove the compactness property. Fix $\varepsilon > 0$. We first observe that, since $\alpha_n \to \alpha \in \mathbb{R}^2_+$, for $\alpha_n = (\alpha_n^0, \alpha_n^1)$, and for n small enough there holds

$$\alpha_n^0 \ge (1-\varepsilon)\alpha_0 \quad \text{and} \quad \alpha_n^1 \ge (1-\varepsilon)\alpha_1. \tag{4.15}$$

In particular, in view of (4.6) we have

$$(1-\varepsilon)\min\{\alpha_0,\alpha_1\}\sup\left\{\|u_n\|_{BPGV^2_{\mathscr{B}_n}(Q)} :\ n\in\mathbb{N}\right\} < +\infty. \tag{4.16}$$

Since $\mathscr{B}_n \to \mathscr{B}$ in ℓ^∞, choosing $r = 1$ there exists $N > 0$ such that $\mathscr{B}_n \subset (\mathscr{B})_1$ for all $n \ge N$. Thus, by (4.8) and (4.16), we infer that

$$\sup\left\{\|u_n\|_{BV(Q)} :\ n\in\mathbb{N}\right\} \le C_1 \sup\left\{\|u_n\|_{BPGV^2_{\mathscr{B}_n}(Q)} :\ n\in\mathbb{N}\right\} < +\infty,$$

and thus we may find $u \in BV(Q)$ such that, up to a subsequence (not relabeled), $u_n \overset{*}{\rightharpoonup} u$ in $BV(Q)$.

Additionally, again from Proposition 3.8, for every $n \in \mathbb{N}$ there exists $v_n \in BV_{\mathscr{B}_n}(Q;\mathbb{R}^N)$ such that,

$$PGV^2_{\alpha_n,\mathscr{B}_n}(u_n) = \alpha_n^0\,|Du_n - v_n|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + \alpha_n^1\,|\mathscr{B}_n v_n|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})}.$$

By (4.6) and (4.7), and in view of Assumption 3.3, we find $v \in BV_{\mathscr{B}}(Q;\mathbb{R}^N)$ such that, up to a subsequence (not relabeled), $v_n \to v$ strongly in L^1. Therefore, we have

$$\begin{aligned}\liminf_{n\to\infty} PGV^2_{\alpha_n,\mathscr{B}_n}(u_n) &\ge \liminf_{n\to\infty}\alpha_n^0\,|Du_n - v_n|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + \liminf_{n\to\infty}\alpha_n^1\,|\mathscr{B}_n v_n|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})}\\ &\ge (1-\varepsilon)\alpha_0\,|Du - v|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + (1-\varepsilon)\alpha_1\,|\mathscr{B}v|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})}\\ &\ge (1-\varepsilon)PGV^2_{\alpha,\mathscr{B}}(u),\end{aligned}$$

where in the second to last inequality we used Assumption 3.3 and (4.15). The arbitrariness of ε concludes the proof of the proposition. □

Proposition 4.5. Let $\{\mathscr{B}_n\}_{n=1}^\infty \subset \Pi$ satisfy Assumption 3.3, and let $\{\alpha_n\}_{n=1}^\infty \subset \mathbb{R}^2_+$ be such that $\mathscr{B}_n \to \mathscr{B}$ in ℓ^∞ and $\alpha_n \to \alpha \in \mathbb{R}^2_+$. Then for every $u \in BV(Q)$ there exists $\{u_n\}_{n=1}^\infty \subset BV(Q)$ such that $u_n \to u$ in $L^1(Q)$ and

$$\limsup_{n\to\infty} PGV^2_{\alpha_n,\mathscr{B}_n}(u_n) \le PGV^2_{\alpha,\mathscr{B}}(u).$$

Proof. This is a direct consequence of Proposition 3.9 by choosing $u_n := u$. □

We close Section 4 by proving Theorem 4.2.

Proof of Theorem 4.2. Properties (Compactness) and (Liminf inequality) hold in view of Proposition 4.4, and Property (Recovery sequence) follows from Proposition 4.5. □

5. The bilevel training scheme with PGV- regularizers

Let $u_\eta \in L^2(Q)$ and $u_c \in BV(Q)$ be the corrupted and clean images, respectively. In what follows we will refer to pairs (u_c, u_η) as training pairs. We recall that Π was introduced in Definition 3.4.

Definition 5.1. We say that $\Sigma \subset \Pi$ is a training set if the operators in Σ satisfy Assumption 3.3, and if Σ is closed and bounded in ℓ^∞.

Examples of training sets are provided in Section 7. We introduce the following bilevel training scheme.

Definition 5.2. Let $\theta \in (0,1)$ and let Σ be a training set. The two levels of the scheme $(\mathcal{T}_\theta^2)$ are

Level 1. $$(\tilde\alpha, \tilde{\mathscr{B}}) := \arg\min \left\{ \|u_c - u_{\alpha,\mathscr{B}}\|^2_{L^2(Q)} :\ \alpha \in \left[\theta, \frac{1}{\theta}\right]^2,\ \mathscr{B} \in \Sigma \right\},$$

Level 2. $$u_{\alpha,\mathscr{B}} := \arg\min \left\{ \|u - u_\eta\|^2_{L^2(Q)} + PGV^2_{\alpha,\mathscr{B}}(u),\ u \in BV(Q) \right\}. \qquad (\mathcal{T}_\theta^2\text{-L2})$$

We first show that the Level 2 problem in ($\mathcal{T}_\theta^2$-L2) admits a solution for every given $u_\eta \in L^2(Q)$, and for every $\alpha \in \mathbb{R}^2_+$.

Proposition 5.3. Let $u_\eta \in L^2(Q)$. Let $\mathscr{B} \in \Sigma$, and let $\alpha \in \mathbb{R}^2_+$. Then there exists $u_{\alpha,\mathscr{B}} \in BV(Q)$ such that

$$\|u_{\alpha,\mathscr{B}} - u_\eta\|^2_{L^2(Q)} + PGV^2_{\alpha,\mathscr{B}}(u_{\alpha,\mathscr{B}}) = \min \left\{ \|u - u_\eta\|^2_{L^2(Q)} + PGV^2_{\alpha,\mathscr{B}}(u) :\ u \in BV(Q) \right\}.$$

Proof. Without loss of generality, we assume that $\alpha := (1,1)$. Let $\{u_n\}_{n=1}^\infty \subset BV(Q)$ be such that

$$\|u_n - u_\eta\|^2_{L^2(Q)} + PGV^2_{\mathscr{B}}(u_n) \le \inf \left\{ \|u - u_\eta\|^2_{L^2(Q)} + PGV^2_{\mathscr{B}}(u) :\ u \in BV(Q) \right\} + 1/n, \qquad (5.1)$$

for every $n \in \mathbb{N}$, and let $\{v_n\} \subset BV_{\mathscr{B}}(Q)$ be the associated sequence of maps provided by Proposition 3.8. In view of (5.1), there exists a constant C such that

$$\|u_n - u_\eta\|^2_{L^2(Q)} + |Du_n - v_n|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + |\mathscr{B} v_n|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})} \le C \qquad (5.2)$$

for every $n \in \mathbb{N}$. We claim that

$$\sup \left\{ \|v_n\|_{L^1(Q;\mathbb{R}^N)} :\ n \in \mathbb{N} \right\} < +\infty. \qquad (5.3)$$

Indeed, if (5.3) does not hold, then, up to the extraction of a subsequence (not relabeled), we have

$$\lim_{n\to+\infty} \|v_n\|_{L^1(Q;\mathbb{R}^N)} = +\infty.$$

Setting

$$\tilde u_n := \frac{u_n}{\|v_n\|_{L^1(Q;\mathbb{R}^N)}} \quad \text{and} \quad \tilde v_n := \frac{v_n}{\|v_n\|_{L^1(Q;\mathbb{R}^N)}} \quad \text{for every } n \in \mathbb{N}, \qquad (5.4)$$

and dividing both sides of (5.2) by $\|v_n\|_{L^1(Q)}$, we deduce that

$$\lim_{n\to+\infty} \left[\left\| \tilde u_n - \frac{u_\eta}{\|v_n\|_{L^1(Q;\mathbb{R}^N)}} \right\|^2_{L^2(Q)} + |D\tilde u_n - \tilde v_n|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + |\mathscr{B}\tilde v_n|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})} \right] = 0. \qquad (5.5)$$

In view of (5.4) and (5.5), and by Assumption 3.3, there exists $\tilde v \in BV_{\mathscr{B}}(Q;\mathbb{R}^N)$, with

$$\|\tilde v\|_{L^1(Q;\mathbb{R}^N)} = 1, \qquad (5.6)$$

such that

$$\tilde{v}_n \to \tilde{v} \quad \text{strongly in } L^1(Q;\mathbb{R}^N), \tag{5.7}$$

and

$$\mathscr{B}\tilde{v}_n \overset{*}{\rightharpoonup} \mathscr{B}\tilde{v} \quad \text{weakly}^* \text{ in } \mathcal{M}_b(Q;\mathbb{R}^{N\times N}).$$

Additionally, (5.5) and (5.7) yield

$$\tilde{u}_n \to 0 \quad \text{strongly in } L^2(Q), \tag{5.8}$$

and

$$\limsup_{n\to+\infty} |D\tilde{u}_n - \tilde{v}|_{\mathcal{M}_b(Q;\mathbb{R}^N)} \le \lim_{n\to+\infty} |D\tilde{u}_n - \tilde{v}_n|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + \lim_{n\to+\infty} \|\tilde{v}_n - \tilde{v}\|_{L^1(Q;\mathbb{R}^N)} = 0. \tag{5.9}$$

Since by (5.8) $D\tilde{u}_n \to 0$ in the sense of distribution, we deduce from (5.9) that $\tilde{v} = 0$. This contradicts (5.6), and implies claim (5.3).

By combining (5.2) and (5.3), we obtain the uniform bound

$$|Du_n|_{\mathcal{M}_b(Q;\mathbb{R}^N)} \le |Du_n - v_n|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + \|v_n\|_{L^1(Q;\mathbb{R}^N)} \le C$$

for every $n \in \mathbb{N}$ and some $C > 0$. Thus, by (5.2) and Assumption 3.2 there exist $u_{\mathscr{B}} \in BV(Q)$ and $v \in BV_{\mathscr{B}}(Q)$ such that, up to the extraction of a subsequence (not relabeled),

$$\begin{aligned}
&u_n \rightharpoonup u_{\mathscr{B}} \quad \text{weakly in } L^2(Q),\\
&u_n \overset{*}{\rightharpoonup} u_{\mathscr{B}} \quad \text{weakly}^* \text{ in } BV(Q),\\
&v_n \to v \quad \text{strongly in } L^1(Q;\mathbb{R}^N),\\
&\mathscr{B}v_n \overset{*}{\rightharpoonup} \mathscr{B}v \text{ weakly}^* \text{ in } \mathcal{M}_b(Q;\mathbb{R}^{N\times N}).
\end{aligned}$$

In view of (5.1), and by lower-semicontinuity, we obtain the inequality

$$\begin{aligned}
&\|u_{\mathscr{B}} - u_0\|^2_{L^2(Q)} + |Du_{\mathscr{B}} - v|_{\mathcal{M}_b(Q;\mathbb{R}^N)} + |\mathscr{B}v|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})}\\
&\quad \le \inf\left\{\|u - u_\eta\|^2_{L^2(Q)} + PGV^2_{\mathscr{B}}(u) :\, u \in BV(Q)\right\}.
\end{aligned}$$

□

Theorem 5.4. *The training scheme* $(\mathcal{T}^2_\theta)$ *admits at least one solution* $(\tilde{\alpha}, \tilde{\mathscr{B}}) \in \left[\theta, 1/\theta\right]^2 \times \Sigma$, *and provides an associated optimally reconstructed image* $u_{\tilde{\alpha},\tilde{\mathscr{B}}} \in BV(Q)$.

Proof. By the boundedness and closedness of Σ in ℓ^∞, up to a subsequence (not relabeled), there exists $(\tilde{\alpha}, \tilde{\mathscr{B}}) \in [\theta, 1/\theta]^2 \times \Sigma$ such that $\alpha_n \to \tilde{\alpha}$ in $\mathbb{R}^2$ and $\mathscr{B}_n \to \tilde{\mathscr{B}}$ in ℓ^∞. Therefore, in view of Theorem 4.2 and the properties of Γ-convergence, we have

$$u_{\alpha_n,\mathscr{B}_n} \overset{*}{\rightharpoonup} u_{\tilde{\alpha},\tilde{\mathscr{B}}} \text{ weakly}^* \text{ in } BV(Q) \text{ and strongly in } L^1(Q), \tag{5.10}$$

where $u_{\alpha_n,\mathscr{B}_n}$ and $u_{\tilde{\alpha},\tilde{\mathscr{B}}}$ are defined in ($\mathcal{T}^2_\theta$-L2).

By (5.10), we have

$$\left\|u_{\tilde{\alpha},\tilde{\mathscr{B}}} - u_c\right\|_{L^2(Q)} \le \liminf_{n\to\infty} \|u_{\alpha_n,\mathscr{B}_n} - u_c\|_{L^2(Q)},$$

which completes the proof. □

6. Training set $\Sigma[\mathscr{A}]$ based on $(\mathscr{A}, \mathscr{B})$ training operators pairs

This section is devoted to providing a class of operators $\mathscr{B}$ belonging to Π (see Definition 3.4), satisfying Assumption 3.3, and being closed with respect to the convergence in (2.3). Recall that $Q = \left(-\frac{1}{2}, \frac{1}{2}\right)^N$.

6.1. A subcollection of Π characterized by $(\mathscr{A}, \mathscr{B})$ training operators pairs.

Let U be an open set in $\mathbb{R}^N$, and let $\mathscr{A} : \mathcal{D}'(U; \mathbb{R}^N) \to \mathcal{D}'(U; \mathbb{R}^N)$ be a d-th order differential operator, defined as

$$\mathscr{A} u := \sum_{|a| \leq d} A_a \frac{\partial^a}{\partial x^a} u \quad \text{for every } u \in \mathcal{D}'(U; \mathbb{R}^N),$$

where, for every multi-index $a = (a^1, a^2, \ldots, a^N) \in \mathbb{N}^N$,

$$\frac{\partial^a}{\partial x^a} := \frac{\partial^{a^1}}{\partial x_1^{a^1}} \frac{\partial^{a^2}}{\partial x_2^{a^2}} \cdots \frac{\partial^{a^N}}{\partial x_N^{a^N}}$$

is meant in the sense of distributional derivatives, and A_a is a linear operator mapping from $\mathbb{R}^N$ to $\mathbb{R}^N$. Let $\mathscr{B}$ be a first order differential operator, $\mathscr{B} : \mathcal{D}'(U; \mathbb{R}^N) \to \mathcal{D}'(U; \mathbb{R}^{N \times N})$, given by

$$\mathscr{B} v := \sum_{i=1}^{N} B^i \frac{\partial}{\partial x_i} v \quad \text{for every } v \in \mathcal{D}'(U; \mathbb{R}^N),$$

where $B^i \in \mathbb{M}^{N^3}$ for each $i = 1, \ldots, N$, and where $\frac{\partial}{\partial x_i}$ denotes the distributional derivative with respect to the i-th variable. We will restrict our analysis to elliptic pairs $(\mathscr{A}, \mathscr{B})$ satisfying the ellipticity assumptions below.

Definition 6.1. We say that $(\mathscr{A}, \mathscr{B})$ is a training operator pair if $\mathscr{B}$ has finite dimensional null-space $\mathcal{N}(\mathscr{B})$, and $(\mathscr{A}, \mathscr{B})$ satisfies the following assumptions:

1. For every $\lambda \in \{-1, 1\}^N$, the operator $\mathscr{A}$ has a fundamental solution $P_\lambda \in L^1(\mathbb{R}^N; \mathbb{R}^N)$ such that:
   a. $\mathscr{A} P_\lambda = \lambda \delta$, where δ denotes the Dirac measure centered at the origin;
   b. $P_\lambda \in C^\infty(\mathbb{R}^N \setminus \{0\}; \mathbb{R}^N)$ and $\frac{\partial^a}{\partial x^a} P_\lambda \in L^1(\mathbb{R}^N; \mathbb{R}^N)$ for every multi-index $a \in \mathbb{N}^N$ with $|a| \leq d - 1$ (where d is the order of the operator $\mathscr{A}$);
   c. for every $a \in \mathbb{N}^N$ with $|a| \leq d - 1$, and for every open set $U \subset \mathbb{R}^N$ such that $Q \subset U$, we have

   $$\sum_{|a| = d-1} \left\| \tau_h \left(\frac{\partial^a}{\partial x^a} P_\lambda \right) - \frac{\partial^a}{\partial x^a} P_\lambda \right\|_{L^1(U; \mathbb{R}^N)} =: M_{\mathscr{A}}(U; h) \to 0 \quad \text{as } |h| \to 0, \tag{6.1}$$

   where for $h \in \mathbb{R}^N$, the translation operator $\tau_h : L^1(\mathbb{R}^N; \mathbb{R}^N) \to L^1(\mathbb{R}^N; \mathbb{R}^N)$ is defined by

   $$\tau_h w(x) := w(x + h) \quad \text{for every } w \in L^1(\mathbb{R}^N; \mathbb{R}^N) \text{ and for a.e. } x \in \mathbb{R}^N. \tag{6.2}$$

2. For every open set $U \subset \mathbb{R}^N$ such that $Q \subset U$, and for every $u \in W^{d,1}(U; \mathbb{R}^N)$ and $v \in C_c^\infty(U; \mathbb{R}^N)$

   $$\|(\mathscr{A} u)_i * v_i\|_{L^1(U)} \leq C_{\mathscr{A}} \left[\sum_{|a| \leq d-1} \left\| \frac{\partial^a}{\partial x^a} u \right\|_{L^1(U; \mathbb{R}^N)} \right] |\mathscr{B} v|_{\mathcal{M}_b(U; \mathbb{R}^{N \times N})}, \tag{6.3}$$

   for every $i = 1, \ldots, N$, where the constant $C_{\mathscr{A}}$ depends only on the operator $\mathscr{A}$. The same property holds for $u \in C_c^\infty(U; \mathbb{R}^N)$ and $v \in BV_{\mathscr{B}}(U; \mathbb{R}^N)$ (see (3.1)).

Explicit examples of operators $\mathscr{A}$ and $\mathscr{B}$ satisfying Definition 6.1 are provided in Section 7. Condition 2. in Definition 6.1 can be interpreted as an "integration by parts-requirement", as highlighted by the example below. Let $N = 2$, $d = 2$, $\mathscr{B} = \nabla$, and let $U \subset \mathbb{R}^2$ be an open set such that $Q \subset U$. Consider the following second order differential operator

$$\mathscr{A}u := \begin{pmatrix} \frac{\partial^2 u_1}{\partial x_1^2} & \frac{\partial^2 u_2}{\partial x_2^2} \end{pmatrix}^\intercal \quad \text{for every} \quad u = (u_1, u_2)^\intercal \in D'(U; \mathbb{R}^2).$$

Then, for every $u \in W^{2,1}(U; \mathbb{R}^2)$ and $v \in C_c^\infty(U; \mathbb{R}^2)$ there holds

$$\begin{aligned} \|(\mathscr{A}u)_i * v_i\|_{L^1(U)} = \left\| \frac{\partial^2 u_i}{\partial x_i^2} * v_i \right\|_{L^1(U)} &= \left\| \frac{\partial u_i}{\partial x_i} * \frac{\partial v_i}{\partial x_i} \right\|_{L^1(U)} \leq \|\nabla u\|_{L^1(U;\mathbb{R}^{2\times 2})} \|\nabla v\|_{L^1(U;\mathbb{R}^{2\times 2})} \\ &= \|\nabla u\|_{L^1(U;\mathbb{R}^{2\times 2})} \|\mathscr{B}v\|_{L^1(U;\mathbb{R}^{2\times 2})}, \end{aligned}$$

for every $i = 1, 2$. In other words, the pair $(\mathscr{A}, \mathscr{B})$ satisfies (6.3) with $C_{\mathscr{A}} = 1$.

Definition 6.2. For every $\mathscr{A}$ as in Definition 6.1 we denote by $\Pi_{\mathscr{A}}$ the following collection of first order differential operators $\mathscr{B}$,

$$\Pi_{\mathscr{A}} := \{\mathscr{B} : (\mathscr{A}, \mathscr{B}) \text{ is a training operator pair}\}.$$

The main result of this section is the following.

Theorem 6.3. *Let $\mathscr{A}$ be as in Definition 6.1. Let Π and $\Pi_{\mathscr{A}}$ be the collections of first order operators introduced in Definition 3.4 and Definition 6.2, respectively. Then*

$$\Pi_{\mathscr{A}} \subset \Pi,$$

thus every operator $\mathscr{B} \in \Pi_{\mathscr{A}}$ satisfies Assumption 3.2. Additionally, the operators in $\Pi_{\mathscr{A}}$ fulfill Assumption 3.3.

We proceed by first recalling two preliminary results from the literature. The next proposition, that may be found in [7, Theorem 4.26], will be instrumental in the proof of a regularity result for distributions with bounded $\mathscr{B}$-total-variation (see Proposition 6.7).

Proposition 6.4. Let $\mathcal{F}$ be a bounded set in $L^p(\mathbb{R}^N)$ with $1 \leq p < +\infty$. Assume that

$$\lim_{|h| \to 0} \|\tau_h f - f\|_{L^p(\mathbb{R}^N)} = 0 \text{ uniformly in } \mathcal{F}.$$

Then, denoting by $\mathcal{F}\lfloor_Q$ the collection of the restrictions to Q of the functions in $\mathcal{F}$, the closure of $\mathcal{F}\lfloor_Q$ in $L^p(Q)$ is compact.

We also recall some basic properties of the space $BV_{\mathscr{B}}(Q; \mathbb{R}^N)$ for $\mathscr{B} \in \Pi_{\mathscr{A}}$ (see Definition 3.1 and [6, Section 2]) .

Proposition 6.5. Let $\mathscr{B} \in \Pi_{\mathscr{A}}$. Let U be an open set in $\mathbb{R}^N$. Then

1. $BV_{\mathscr{B}}(U; \mathbb{R}^N)$ is a Banach space with respect to the norm defined in (3.2);
2. $C^\infty(U, \mathbb{R}^N)$ is dense in $BV_{\mathscr{B}}(U; \mathbb{R}^N)$ in the strict topology, i.e., for every $u \in BV_{\mathscr{B}}(U; \mathbb{R}^N)$ there exists $\{u_n\}_{n=1}^\infty \subset C^\infty(U, \mathbb{R}^N)$ such that

$$u_n \to u \text{ strongly in } L^1(U; \mathbb{R}^N) \text{ and } |\mathscr{B}u_n|_{\mathcal{M}_b(U;\mathbb{R}^{N\times N})} \to |\mathscr{B}u|_{\mathcal{M}_b(U;\mathbb{R}^{N\times N})}.$$

Before we establish Theorem 6.3, we prove a technical lemma.

Lemma 6.6. *Let $k \in \mathbb{N}$. Then there exists a constant $C > 0$ such that, for every $h \in \mathbb{R}^N$ and $w \in W^{k,1}_{\mathrm{loc}}(\mathbb{R}^N;\mathbb{R}^N)$, there holds*

$$\limsup_{|h|\to 0} \sum_{|a|\le k} \left\| \tau_h\Big(\frac{\partial^a}{\partial x^a} w\Big) - \frac{\partial^a}{\partial x^a} w \right\|_{L^1(Q;\mathbb{R}^N)} \le \limsup_{|h|\to 0} C \sum_{|a|=k} \left\| \tau_h\Big(\frac{\partial^a}{\partial x^a} w\Big) - \frac{\partial^a}{\partial x^a} w \right\|_{L^1(Q;\mathbb{R}^N)},$$

where τ_h is the operator defined in (6.2).

Proof. By the linearity of τ_h, we have

$$\tau_h\Big(\frac{\partial^a}{\partial x^a} w\Big) - \frac{\partial^a}{\partial x^a} w = \frac{\partial^a}{\partial x^a}(\tau_h w - w).$$

On the one hand, by the Sobolev embedding theorem (see, e.g., [25]), we have

$$\begin{aligned} &\sum_{|a|\le k} \left\| \tau_h\Big(\frac{\partial^a}{\partial x^a} w\Big) - \frac{\partial^a}{\partial x^a} w \right\|_{L^1(Q;\mathbb{R}^N)} \\ &= \sum_{|a|\le k} \left\| \frac{\partial^a}{\partial x^a}(\tau_h w - w) \right\|_{L^1(Q;\mathbb{R}^N)} \\ &\le C \left\|\tau_h(w) - w\right\|_{L^1(Q;\mathbb{R}^N)} + C \sum_{|a|=k} \left\| \frac{\partial^a}{\partial x^a}(\tau_h w - w) \right\|_{L^1(Q;\mathbb{R}^N)}. \end{aligned} \tag{6.4}$$

On the other hand, by the continuity of the translation operator in L^1 (see, e.g., [7, Lemma 4.3] for a proof in $\mathbb{R}^N$, the analogous argument holds on bounded open sets) we have

$$\limsup_{|h|\to 0} \left\|\tau_h(w) - w\right\|_{L^1(Q;\mathbb{R}^N)} = 0. \tag{6.5}$$

The result follows by combining (6.4) and (6.5). □

The next proposition shows that operators in $\Pi_{\mathscr{A}}$ satisfy Assumption 3.2.

Proposition 6.7. Let $\mathscr{B} \in \Pi_{\mathscr{A}}$, and let $BV_{\mathscr{B}}(Q;\mathbb{R}^N)$ be the space introduced in Definition 3.1. Then the injection of $BV_{\mathscr{B}}(Q;\mathbb{R}^N)$ into $L^1(Q;\mathbb{R}^N)$ is compact.

Proof. In view of Proposition 6.5, for every $u \in BV_{\mathscr{B}}(Q;\mathbb{R}^N)$ there exists a sequence of maps $\{v_u^n\}_{n=1}^{\infty} \subset C^\infty(Q;\mathbb{R}^N)$ such that

$$\|v_u^n - u\|_{L^1(Q;\mathbb{R}^N)} + \left| \|\mathscr{B} v_u^n\|_{L^1(Q;\mathbb{R}^{N\times N})} - |\mathscr{B} u|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})} \right| \le \frac{1}{n}. \tag{6.6}$$

With a slight abuse of notation, we still denote by v_u^n the C^d-extension of the above maps to the whole $\mathbb{R}^N$ (see e.g. [16]), where d is the order of the operator $\mathscr{A}$. Without loss of generality, up to a multiplication by a cut-off function, we can assume that $v_u^n \in C_c^d(2Q;\mathbb{R}^N)$ for every $n \in \mathbb{N}$.

We first show that, setting

$$\mathcal{F} := \left\{ u \in L^1(Q;\mathbb{R}^N) : \ \|u\|_{BV_{\mathscr{B}}(Q;\mathbb{R}^N)} \le 1 \right\},$$

for every $n \in \mathbb{N}$ there holds

$$\lim_{|h|\to 0} \sup_{u\in\mathcal{F}} \left\{ \|\tau_h v_u^n - v_u^n\|_{L^1(Q;\mathbb{R}^N)} \right\} = 0, \tag{6.7}$$

where we recall τ_h from Theorem 6.4, and where for fixed $u \in \mathcal{F}$, v_u^n is as above and satisfying (6.6).

Let $h \in \mathbb{R}^N$ and let δ_h be the Dirac distribution centered at $h \in \mathbb{R}^N$. By the properties of the fundamental solution P_λ we deduce

$$\begin{aligned}\tau_h(\lambda_i v^n_{u,i}) &= \delta_h * \lambda_i v^n_{u,i} = \delta_h * (\lambda_i \delta * v^n_{u,i}) = \delta_h * \left((\mathscr{A} P_\lambda)_i * v^n_{u,i}\right)\\ &= (\delta_h * (\mathscr{A} P_\lambda)_i) * v^n_{u,i} = (\mathscr{A}\,(\delta_h * (P_\lambda)))_i * v^n_{u,i},\end{aligned}$$

for every $i = 1, \dots, N$, and every $\lambda \in \{-1,1\}^N$. Therefore, we obtain that

$$\begin{aligned}&\left\|\tau_h(\lambda_i v^n_{u,i}) - \lambda_i v^n_{u,i}\right\|_{L^1(Q;\mathbb{R}^N)} \qquad (6.8)\\ &= \left\|(\mathscr{A}\,(\delta_h * (P_\lambda)))_i * v^n_{u,i} - (\mathscr{A} P_\lambda)_i * v^n_{u,i}\right\|_{L^1(Q;\mathbb{R}^N)} = \left\|(\mathscr{A}\,(\delta_h * (P_\lambda) - P_\lambda))_i * v^n_{u,i}\right\|_{L^1(Q;\mathbb{R}^N)}\\ &\leq C_{\mathscr{A}} \left[\sum_{|a| \leq d-1} \left\|\tau_h\left(\frac{\partial^a}{\partial x^a} P_\lambda\right) - \frac{\partial^a}{\partial x^a} P_\lambda\right\|_{L^1(Q;\mathbb{R}^N)}\right] |\mathscr{B} v^n_u|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})}\end{aligned}$$

for every $\lambda \in \{-1,1\}^N$, where in the last inequality we used the fact that $\tau_h P_\lambda - P_\lambda \in W^{d-1,d}(\mathbb{R}^N;\mathbb{R}^N)$ owing to Definition 6.1, Assertion 1c, the identity $\tau_h\left(\frac{\partial^a}{\partial x^a} P_\lambda\right) = \frac{\partial^a}{\partial x^a}\left(\tau_h P_\lambda\right)$, as well as Definition 6.1, Assertion 2.

In particular, choosing $\bar{\lambda} := (1, \dots, 1)$ we have

$$\begin{aligned}&\sup_{u \in \mathcal{F}} \left\{\|\tau_h(v^n_u) - v^n_u\|_{L^1(Q;\mathbb{R}^N)}\right\}\\ &\quad\leq C_{\mathscr{A}}\left(1 + \frac{1}{n}\right) \sum_{|a| \leq d-1} \left\|\tau_h\left(\frac{\partial^a}{\partial x^a} P_{\bar{\lambda}}\right) - \frac{\partial^a}{\partial x^a} P_{\bar{\lambda}}\right\|_{L^1(Q;\mathbb{R}^N)},\end{aligned}$$

and, in view of (6.1) and Lemma 6.6, we conclude that

$$\begin{aligned}&\lim_{|h|\to 0} \sup_{u \in \mathcal{F}} \left\{\|\tau_h(v^n_u) - v^n_u\|_{L^1(Q;\mathbb{R}^N)}\right\}\\ &\leq C_{\mathscr{A}}\left(1 + \frac{1}{n}\right) \lim_{|h|\to 0} \sum_{|a| = d-1} \left\|\tau_h\left(\frac{\partial^a}{\partial x^a} P_{\bar{\lambda}}\right) - \frac{\partial^a}{\partial x^a} P_{\bar{\lambda}}\right\|_{L^1(Q;\mathbb{R}^N)} = 0\end{aligned}$$

for every $n \in \mathbb{N}$, which yields (6.7).

By (6.6), for $n \in \mathbb{N}$ fixed, for every $h \in \mathbb{R}^N$ with $|h| < 1$, and for every $u \in \mathcal{F}$ there holds

$$\begin{aligned}&\|\tau_h u - u\|_{L^1(Q;\mathbb{R}^N)} \leq \|\tau_h u - \tau_h v^n_u\|_{L^1(Q;\mathbb{R}^N)} + \|\tau_h v^n_u - v^n_u\|_{L^1(Q;\mathbb{R}^N)} + \|v^n_u - u\|_{L^1(Q;\mathbb{R}^N)}\\ &\quad\leq \frac{1}{n} + \|u\|_{L^1(Q_{|h|};\mathbb{R}^N)} + \|v^n_u\|_{L^1(Q_{|h|};\mathbb{R}^N)} + \|\tau_h v^n_u - v^n_u\|_{L^1(Q;\mathbb{R}^N)} \leq \frac{2}{n} + \sup_{w \in \mathcal{F}} \left\{\|\tau_h(v^n_w) - v^n_w\|_{L^1(Q;\mathbb{R}^N)}\right\},\end{aligned}$$

where we have still denoted by u the extension of the above map to zero on $\mathbb{R}^N \setminus Q$, and where $Q_{|h|} := \left(-\frac{1}{2} - |h|, \frac{1}{2} + |h|\right)^N \setminus \left(-\frac{1}{2} + |h|, \frac{1}{2} - |h|\right)^N$. By (6.7), and since $L^1(Q_{|h|}) \to 0$ as $|h| \to 0$, we deduce

$$\lim_{|h|\to 0} \sup_{u \in \mathcal{F}} \left\{\|\tau_h(u) - u\|_{L^1(Q;\mathbb{R}^N)}\right\} \leq \frac{2}{n},$$

and letting $n \to +\infty$ we get

$$\lim_{|h|\to 0} \sup_{u \in \mathcal{F}} \left\{\|\tau_h(u) - u\|_{L^1(Q;\mathbb{R}^N)}\right\} = 0.$$

Thus, recalling that $u=0$ on $\mathbb{R}^N\setminus Q$, we deduce the estimate

$$\lim_{|h|\to 0}\sup_{u\in\mathcal{F}}\left\{\|\tau_h(u)-u\|_{L^1(\mathbb{R}^N;\mathbb{R}^N)}\right\}$$
$$\leq \lim_{|h|\to 0}\sup_{u\in\mathcal{F}}\left\{\|\tau_h(u)-u\|_{L^1(Q;\mathbb{R}^N)}+C\|u\|_{L^1(Q_{|h|};\mathbb{R}^N)}\right\}=0.$$

The statement now follows from Proposition 6.4. □

The following extension result in $BV_{\mathscr{B}}$ is a corollary of the properties of the trace operator defined in [6, Section 4].

Lemma 6.8. *Let $\mathscr{B}\in\Pi_{\mathscr{A}}$, and let $BV_{\mathscr{B}}(Q;\mathbb{R}^N)$ be the space introduced in Definition 3.1. Then there exists a continuous extension operator $\mathbb{T}: BV_{\mathscr{B}}(Q;\mathbb{R}^N)\to BV_{\mathscr{B}}(\mathbb{R}^N;\mathbb{R}^N)$ such that $\mathbb{T}u=u$ almost everywhere in Q for every $u\in BV_{\mathscr{B}}(Q;\mathbb{R}^N)$.*

Proof. Since $\mathcal{N}(\mathscr{B})$ is finite dimensional, in view of [6, (4.9) and Theorem 1.1] there exists a continuous trace operator $\mathrm{tr}: BV_{\mathscr{B}}(Q;\mathbb{R}^N)\to L^1(\partial Q;\mathbb{R}^N)$. By the classical results by E. Gagliardo (see [17]) there exists a linear and continuous extension operator $\mathrm{E}: L^1(\partial Q;\mathbb{R}^N)\to W^{1,1}(\mathbb{R}^N\setminus Q;\mathbb{R}^N)$. The statement follows by setting

$$\mathbb{T}u:=u\chi_Q+\mathrm{E}(\mathrm{tr}(u))\chi_{\mathbb{R}^N\setminus Q},$$

where χ_Q and $\chi_{\mathbb{R}^N\setminus Q}$ denote the characteristic functions of the sets Q and $\mathbb{R}^N\setminus Q$, respectively, and by Theorem [6, Corollary 4.21]. □

We point out that, as a direct consequence of Lemma 6.8, we obtain

$$|\mathscr{B}(\mathbb{T}u)|_{\mathcal{M}_b(\mathbb{R}^N;\mathbb{R}^{N\times N})}\leq C\,|\mathscr{B}u|_{\mathcal{M}_b(Q;\mathbb{R}^{N\times N})},$$

where the constant C depends only on Q and $|\mathscr{B}|_{\ell^\infty}$.
We close this subsection by proving a compactness and lower-semicontinuity result for functions with uniformly bounded $BV_{\mathscr{B}_n}$ norms. We recall that the definition of $M_{\mathscr{A}}$ is found in (6.1).

Proposition 6.9. Let $\{\mathscr{B}_n\}_{n=1}^\infty\subset\Pi_{\mathscr{A}}$ be such that $\mathscr{B}_n\to\mathscr{B}$ in ℓ^∞. For every $n\in\mathbb{N}$ let $v_n\in BV_{\mathscr{B}_n}(Q;\mathbb{R}^N)$ be such that

$$\sup\left\{\|v_n\|_{BV_{\mathscr{B}_n}(Q;\mathbb{R}^N)}:\ n\in\mathbb{N}\right\}<+\infty. \tag{6.9}$$

Then there exists $v\in BV_{\mathscr{B}}(Q;\mathbb{R}^N)$ such that, up to a subsequence (not relabeled),

$$v_n\to v \text{ strongly in } L^1(Q;\mathbb{R}^N), \tag{6.10}$$

and

$$\mathscr{B}_n v_n \overset{*}{\rightharpoonup}\mathscr{B}v \text{ weakly* in } \mathcal{M}_b(Q;\mathbb{R}^{N\times N}). \tag{6.11}$$

Proof. Let v_n satisfy (6.9). With a slight abuse of notation we still indicate by v_n the $BV_{\mathscr{B}}$ continuous extension of the above maps to $\mathbb{R}^N$ (see Lemma 6.8). Let $\phi\in C_c^\infty(2Q;\mathbb{R}^N)$ be a cut-off function such that $\phi\equiv 1$ on Q, and for every $n\in\mathbb{N}$ let $\tilde{v}_n$ be the map $\tilde{v}_n:=\phi v_n$. Note that $\mathrm{supp}\,\tilde{v}_n\subset\subset 2Q$. Additionally, by Lemma 6.8 there holds

$$\|\tilde{v}_n\|_{BV_{\mathscr{B}}(2Q;\mathbb{R}^N)}\leq\|v_n\|_{L^1(2Q;\mathbb{R}^N)}+|\mathscr{B}v_n|_{M_b(2Q;\mathbb{R}^{N\times N})}+\left\|\sum_{i=1}^N B^i\frac{\partial\phi}{\partial x_i}\right\|_{L^\infty(2Q;\mathbb{M}^{N^3})}\|v^n\|_{L^1(2Q;\mathbb{R}^N)} \tag{6.12}$$

$$\leq C_1 \|v_n\|_{BV_{\mathscr{B}}(2Q;\mathbb{R}^N)} \leq C_2 \|v_n\|_{BV_{\mathscr{B}}(Q;\mathbb{R}^N)},$$

where in the last inequality we used Lemma 6.8, and where the constants C_1 and C_2 depend only on the cut-off function ϕ. To prove (6.10) we first show that

$$\lim_{|h|\to 0} \sup_{n\in\mathbb{N}} \left\{ \|\tau_h \tilde{v}_n - \tilde{v}_n\|_{L^1(\mathbb{R}^N;\mathbb{R}^N)} \right\} = 0, \tag{6.13}$$

where we recall τ_h from Theorem 6.4. Arguing as in the proof of (6.8), by (6.12) we deduce that for $|h|$ small enough, since $\operatorname{supp}\phi \subset\subset 2\mathrm{Q}$,

$$\begin{aligned}
&\|\tau_h \tilde{v}_n - \tilde{v}_n\|_{L^1(\mathbb{R}^N;\mathbb{R}^N)} = \|\tau_h \tilde{v}_n - \tilde{v}_n\|_{L^1(2Q;\mathbb{R}^N)} \\
&\leq C \left[\sum_{|a|\leq d-1} \left\| \tau_h \left(\frac{\partial^a}{\partial x^a} P_\lambda \right) - \frac{\partial^a}{\partial x^a} P_\lambda \right\|_{L^1(2Q;\mathbb{R}^N)} \right] |\mathscr{B}\tilde{v}_n|_{\mathcal{M}_b(2Q;\mathbb{R}^{N\times N})} \\
&\leq C \left[\sum_{|a|\leq d-1} \left\| \tau_h \left(\frac{\partial^a}{\partial x^a} P_\lambda \right) - \frac{\partial^a}{\partial x^a} P_\lambda \right\|_{L^1(2Q;\mathbb{R}^N)} \right] \|v_n\|_{BV_{\mathscr{B}}(Q;\mathbb{R}^{N\times N})}
\end{aligned}$$

for every $n \in \mathbb{N}$. Property (6.13) follows by (6.1). Owing to Proposition 6.4, we deduce (6.10).

We now prove (6.11). Let $\varphi \in C_c^\infty(Q;\mathbb{R}^{N\times N})$ be such that $|\varphi| \leq 1$. Then

$$\begin{aligned}
\lim_{n\to\infty} \int_Q \varphi \cdot d(\mathscr{B}_n v_n) &= \lim_{n\to\infty} \sum_{i,j=1}^N \int_Q \varphi_{ij} d \left(\sum_{k,l=1}^N (B_n)_{ijl}^k \frac{\partial (v_n)_l}{\partial x_k} \right) \\
&= \lim_{n\to\infty} \sum_{i,j,k,l=1}^N \int_Q \varphi_{ij} d \left((B_n)_{ijl}^k \frac{\partial (v_n)_l}{\partial x_k} \right) \\
&= - \lim_{n\to\infty} \sum_{i,j,k,l=1}^N \int_Q (v_n)_l (B_n)_{ijl}^k \frac{\partial \varphi_{ij}}{\partial x_k} dx \\
&= - \sum_{i,j,k,l=1}^N \int_Q v_l (B)_{ijl}^k \frac{\partial \varphi_{ij}}{\partial x_k} dx
\end{aligned}$$

where in the last step we used the fact that $v_n \to v$ strongly in $L^1(Q)$ and $\mathscr{B}_n \to \mathscr{B}$ in ℓ^∞.

This completes the proof of (6.11) and of the proposition. □

Proof of Theorem 6.3. Let $\mathscr{B} \in \Pi_{\mathscr{A}}$ be given. The fact that $\mathscr{B}$ satisfies Assumption 3.2 follows by Propositions 6.5 and 6.7. The fulfillment of Assumption 3.3 is a direct consequence of Proposition 6.9. □

6.2. **Training scheme with fixed and multiple operators $\mathscr{A}$.** In this subsection we provide a construction of training sets associated to a given differential operator $\mathscr{A}$, namely collection of differential operators $\mathscr{B}$ for which our training scheme is well-posed (see Definitions 5.1 and 5.2). We first introduce a collection $\Sigma[\mathscr{A}]$ for a given operator $\mathscr{A}$ of order $d \in \mathbb{N}$.

Definition 6.10. Let $\mathscr{A}$ be a differential operator of order $d \in \mathbb{N}$. For every $\varepsilon > 0$ we denote by $\Sigma_\varepsilon[\mathscr{A}]$ the collection

$$\Sigma_\varepsilon[\mathscr{A}] := \{\mathscr{B} \in \Pi_{\mathscr{A}} : \varepsilon \leq \|\mathscr{B}\|_{\ell^\infty} \leq 1\}.$$

The first result of this subsection is the following.

Theorem 6.11. *Fix $\varepsilon > 0$. Let $\mathscr{A}$ be a differential operator of order $d \in \mathbb{N}$ such that $\Sigma_\varepsilon[\mathscr{A}]$ is non-empty. Then the collection $\Sigma_\varepsilon[\mathscr{A}]$ is a training set (see Definition 5.1).*

Proof. By the definition of $\Sigma_\varepsilon[\mathscr{A}]$ we just need to show that $\Sigma_\varepsilon[\mathscr{A}]$ is closed in ℓ^∞. Let $u \in C^\infty(Q;\mathbb{R}^N)$ and $\{\mathscr{B}_n\}_{n=1}^\infty \subset \Sigma_\varepsilon[\mathscr{A}]$ be given. Then, up to a subsequence (not relabeled), we may assume that $\mathscr{B}_n \to \mathscr{B}$ in ℓ^∞. We claim that $\mathscr{B} \in \Pi_{\mathscr{A}}$.

To prove that $\mathcal{N}(\mathscr{B})$ is finite-dimensional, we recall that this condition is equivalent to the injectivity of the symbol $\mathbb{B}(\xi)$ (see (2.2)) for all $\xi \in \mathbb{C}^N \setminus \{0\}$ (see [6, Remark 2.1]). Since for all $\xi \in \mathbb{C}^N \setminus \{0\}$ we have that $\mathbb{B}(\xi)$ is the uniform limit of the sequence of injective linear maps, namely the symbols $\{\mathbb{B}_n(\xi)\}_{n=1}^\infty$, either $\mathbb{B}(\xi)$ is constant or it is injective. On the other hand, the linearity of $\mathbb{B}(\xi)$ implies that it is constant only if it is identically zero. The fact that $\varepsilon \leq \|\mathscr{B}\|_{\ell^\infty} \leq 1$ for all $n \in \mathbb{N}$ guarantees that this cannot occur, and yields the injectivity of $\mathbb{B}(\xi)$ and hence the fact that the dimension of $\mathcal{N}(\mathscr{B})$ is finite.

To conclude the proof of the theorem we still need to show that $(\mathscr{A},\mathscr{B})$ satisfies Definition 6.1, Assertion 2. Let U be an open set in $\mathbb{R}^N$ such that $Q \subset U$. Let $u \in C_c^\infty(U;\mathbb{R}^{N\times N})$ and let $v \in BV_{\mathscr{B}}(U;\mathbb{R}^N)$. By Proposition 6.5 there exists $\{v_k\}_{k=1}^\infty \subset C^\infty(U;\mathbb{R}^N)$ such that

$$v_k \to v \text{ strongly in } L^1(U;\mathbb{R}^N) \text{ and } |\mathscr{B}v_k|_{\mathcal{M}_b(U;\mathbb{R}^{N\times N})} \to |\mathscr{B}v|_{\mathcal{M}_b(U;\mathbb{R}^{N\times N})}. \tag{6.14}$$

Integrating by parts we obtain

$$\|(\mathscr{A}u)_i * (v_k)_i\|_{L^1(U;\mathbb{R}^N)} \leq C_{\mathscr{A}} \left[\sum_{|a|\leq d-1} \left\|\frac{\partial^a}{\partial x^a} u\right\|_{L^1(U;\mathbb{R}^N)}\right] |\mathscr{B}_n v_k|_{\mathcal{M}_b(U;\mathbb{R}^{N\times N})},$$

for every $i = 1,\ldots,N$. Taking the limit as $n \to \infty$ first, and then as $k \to \infty$, since $\mathscr{B}_n \to \mathscr{B}$ in ℓ^∞ and in view of (6.14), we conclude that

$$\|(\mathscr{A}u)_i * (v_k)_i\|_{L^1(U;\mathbb{R}^N)} \leq C_{\mathscr{A}} \left[\sum_{|a|\leq d-1} \left\|\frac{\partial^a}{\partial x^a} u\right\|_{L^1(U;\mathbb{R}^N)}\right] |\mathscr{B}v|_{\mathcal{M}_b(U;\mathbb{R}^{N\times N})}.$$

The proof of the second part of Assertion 2 is analogous. This shows that $(\mathscr{A},\mathscr{B})$ satisfies Definition 6.1 and concludes the proof of the theorem. □

Remark 6.12. We note that the result of Theorem 6.11 still holds if we replace the upper bound 1 in Definition 6.10 with an arbitrary positive constant.

We now consider the case of multiple operators $\mathscr{A}$.

Definition 6.13. We say that collection $\mathcal{A}$ of differential operators $\mathscr{A}$ is a training set builder if

$$\sup\{C_{\mathscr{A}} :\ \mathscr{A} \in \mathcal{A}\} < +\infty \text{ and } \lim_{|h|\to 0} \sup\{M_{\mathscr{A}}(h) :\ \mathscr{A} \in \mathcal{A}\} = 0, \tag{6.15}$$

where $C_{\mathscr{A}}$ and $M_{\mathscr{A}}(h)$ are defined in (6.3) and (7.5), respectively.

For every $\varepsilon > 0$ we then define the class $\Sigma_\varepsilon[\mathcal{A}]$ via

$$\Sigma_\varepsilon[\mathcal{A}] := \text{convex hull}\left(\bigcup_{\mathscr{A}\in\mathcal{A}} \Sigma_\varepsilon[\mathscr{A}]\right),$$

where for every $\mathscr{A} \in \mathcal{A}$, $\Sigma_\varepsilon[\mathscr{A}]$ is the class defined in Definition 6.10.

We close this section by proving the following theorem.

Theorem 6.14. *Let $\mathcal{A}$ be a training set builder. Then $\Sigma_\varepsilon[\mathcal{A}]$ is a training set.*

Proof. The proof of this theorem follows the argument in the proof of Theorem 6.11 using the fact that the two critical constants $M_{\mathscr{A}}(h)$ and $C_{\mathscr{A}}$, in (6.1) and (6.3), respectively, are uniformly bounded due to (6.15). □

7. Explicit examples and numerical observations

In this section we exhibit several explicit examples of operators $\mathscr{A}$ and training sets $\Sigma_\varepsilon[\mathscr{A}]$, we provide numerical simulations and some observations derived from them.

7.1. The existence of fundamental solutions of operators $\mathscr{A}$.

One important requirement in Definition 6.1 is the existence of the fundamental solution $P_\lambda \in L^1(\mathbb{R}^N, \mathbb{R}^N)$ of a given operator $\mathscr{A}$. A result in this direction can be found in [22, Page 351, Section 6.3], where an explicit form of the fundamental solution for Agmon-Douglis-Nirenberg elliptic systems with constant coefficients is provided.

Remark 7.1. In the case in which $N = 2$, $\mathscr{A}$ has order 2 and satisfies the assumptions in [22, Page 351, Section 6.3], the fundamental solution P_λ can be written as

$$P_\lambda(x,y) = \frac{1}{8\pi^2}(\Delta L_y)\int_{|\eta|=1,\eta\in\mathbb{R}^2} ((x-y)\cdot\eta)^2 \log|(x-y)\cdot\eta|\, R_{\mathscr{A}}\, d\omega_\eta, \tag{7.1}$$

where L denotes the fundamental solution of Laplace's equation, $R_{\mathscr{A}}$ denotes a constant depending on $\mathscr{A}$, and the integration is taken over the unit circle $|\eta| = 1$ with arc length element $d\omega_\eta$.

In the special case in which

$$\mathscr{A}w := \Delta w + \nabla(\text{div} w) \quad \text{for } w \in \mathcal{D}'(Q;\mathbb{R}^2), \tag{7.2}$$

the fundamental solution P_α, with $\mathscr{A}P_\alpha = \alpha\delta$ for $\alpha \in \mathbb{R}^2$, is given by

$$P_\alpha(x) := \frac{3\alpha}{8\pi}\log\frac{1}{|x|} + \frac{x}{8\pi}\frac{\alpha\cdot|x|}{|x|^2}.$$

We observe that ∇P_α is positively homogeneous of degree $-1(=1-N)$. Also, since $R_{\mathscr{A}}$ in (7.1) is a constant, ∇P_λ must have the same homogeneity as ∇P_α, which is $1-N$.

Proposition 7.2. Let $\mathscr{A}$ be a differential operator of order $d \in \mathbb{N}$, and assume that its fundamental solution P_λ is such that $\frac{\partial^a}{\partial x^a}P_\lambda$ is positively homogeneous of degree $1-N$ for all multi-indexes $a \in \mathbb{N}^N$ with $|a| = d-1$. Then Assertion 1c. in Definition 6.1 is satisfied.

Proof. Let $s \in (0,1)$ be fixed. Since $\frac{\partial^a}{\partial x^a} P_\lambda$ is positively homogeneous of degree $1-N$ for all multi-indexes $a \in \mathbb{N}^N$ with $|a| = d-1$, by [32, Lemma 1.4] we deduce the estimate

$$\begin{aligned}\sum_{|a|=d-1} \left| \tau_h \left(\frac{\partial^a}{\partial x^a} P_\lambda(x) \right) - \frac{\partial^a}{\partial x^a} P_\lambda(x) \right| \\ \leq C \left[\max \left\{ \sup \left\{ \left| \nabla^{d-1} P_\lambda(z) \right| : \ |z| = 1 \right\}, \ \sup \left\{ \left| \nabla^d P_\lambda(z) \right| : \ |z| = 1 \right\} \right\} \right] \cdot \\ \cdot |h|^s \left[\frac{1}{|x|^{N-1+s}} + \frac{1}{|x+h|^{N-1+s}} \right].\end{aligned} \tag{7.3}$$

for every $x \in \mathbb{R}^N$, $0 \leq s \leq 1$, and $|h| \leq 1/2$, where the constant C is independent of x and h.

Next, for every bounded open set $U \subset \mathbb{R}^N$ satisfying $Q \subset U$ we have

$$\begin{aligned}\int_U \frac{1}{|x|^{N-1+s}} dx \leq \int_{B(0,2)} \frac{1}{|x|^{N-1+s}} dx + \int_{U \setminus B(0,2)} \frac{1}{|x|^{N-1+s}} dx \\ \leq 2\pi \int_0^2 r^{-s} dr + \frac{1}{2^{N-1+s}} |U \setminus B(0,2)| < +\infty,\end{aligned} \tag{7.4}$$

The analogous computation holds for $\frac{1}{|x+h|^{N-1+s}}$. Since P_λ is a fundamental solution and $\mathscr{A} P_\lambda = \lambda \delta$, we have that $P_\lambda \in C^\infty(\mathbb{R}^N \setminus B(0,\varepsilon))$ for every $\varepsilon > 0$. In particular,

$$\max \left\{ \sup \left\{ \left| \nabla^{d-1} P(z) \right| : \ |z| = 1 \right\}, \ \sup \left\{ \left| \nabla^d P(z) \right| : \ |z| = 1 \right\} \right\} =: M < +\infty. \tag{7.5}$$

This, together with (7.3) and (7.4), yields

$$\left\| \sum_{|a|=d-1} \left| \tau_h \left(\frac{\partial^a}{\partial x^a} P_\lambda(x) \right) - \frac{\partial^a}{\partial x^a} P_\lambda(x) \right| \right\|_{L^1(U;\mathbb{R}^N)} \leq CM \, |h|^s,$$

for some $C > 0$, and thus

$$\lim_{h \to \infty} \left\| \sum_{|a|=d-1} \left| \tau_h \left(\frac{\partial^a}{\partial x^a} P_\lambda(x) \right) - \frac{\partial^a}{\partial x^a} P_\lambda(x) \right| \right\|_{L^1(U;\mathbb{R}^N)} = 0,$$

and (6.1) is established. □

Remark 7.3. As a corollary of Proposition 7.2 and Remark 7.1, we deduce that all operators $\mathscr{A}$ satisfying the assumptions in [22, Page 351, Section 6.3] comply with Definition 6.1, Assertion 1. In particular, differential operators $\mathscr{A}$ which can be written in the form $\mathscr{A} = \mathscr{B}^* \circ \mathscr{C}$, where $\mathscr{B}^*$ is the first order differential operator associated to $\mathscr{B}$ and having as coefficients the transpose of the matrices B^i, $i = 1, \ldots, N$, and where $\mathscr{C}$ is a differential operator of order $d-1$ having constant coefficients, are such that $(\mathscr{A}, \mathscr{B})$ complies with Definition 6.1.

7.2. **The unified approach to TGV^2 and $NsTGV^2$ - an example of $\Sigma[\mathscr{A}]$.** In this section we give an explicit construction of an operator $\mathscr{A}$ such that the seminorms $NsTGV^2$ and TGV^2, as well as a continuum of topologically equivalent seminorms connecting them, can be constructed as operators $\mathscr{B} \in \Sigma[\mathscr{A}]$.

We start by recalling the definition of the classical symmetrized gradient,

$$\mathcal{E}v = \frac{\nabla v + (\nabla v)^T}{2} = \begin{bmatrix} \partial_1 v_1 & \frac{(\partial_1 v_2 + \partial_2 v_1)}{2} \\ \frac{(\partial_1 v_2 + \partial_2 v_1)}{2} & \partial_2 v_2 \end{bmatrix}, \tag{7.6}$$

for $v = (v_1, v_2) \in C^\infty(Q; \mathbb{R}^2)$. Let

$$B^1_{\mathrm{sym}} = \left[\begin{array}{cc|cc} 1 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 0 \end{array}\right] \text{ and } B^2_{\mathrm{sym}} = \left[\begin{array}{cc|cc} 0 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1 \end{array}\right],$$

and let $\mathscr{B}_{\mathrm{sym}}(v)$ be defined as in (2.1) with B^1_{sym} and B^2_{sym} as above. Then $\mathscr{B}_{\mathrm{sym}}(v) = \mathcal{E}v$ for all $v \in C^\infty(Q; \mathbb{R}^2)$, and $\mathcal{N}(\mathscr{B}_{\mathrm{sym}})$ is finite dimensional. In particular,

$$\mathcal{N}(\mathscr{B}_{\mathrm{sym}}) = \left\{ v(x) = \alpha \begin{pmatrix} x_2 \\ -x_1 \end{pmatrix} + b : \alpha \in \mathbb{R} \quad \text{and} \quad b \in \mathbb{R}^2 \right\}.$$

The first part of Definition 6.1 follows from Remark 7.3. Next we verify that (6.3) holds. Indeed, choosing $\mathscr{A}$ as in (7.2), we first observe that

$$\begin{aligned} (\mathscr{A}w) * v &= \sum_{j=1}^N [\Delta w_j + \partial_j \mathrm{div}(w)] * v_j = \sum_{i,j=1}^N (\partial_i w_j + \partial_j w_i) * \partial_i v_j \\ &= \sum_{i,j=1}^N (\partial_i w_j + \partial_j w_i) * (\partial_i v_j + \partial_j v_i) = (\mathscr{B}_{\mathrm{sym}} w) * (\mathscr{B}_{\mathrm{sym}} v), \end{aligned} \tag{7.7}$$

for every $w \in W^{1,2}(Q; \mathbb{R}^2)$ and $v \in C^\infty_c(Q; \mathbb{R}^2)$. That is, for every open set $U \subset \mathbb{R}^N$ such that $Q \subset U$ we have

$$|(\mathscr{A}w) * v|_{\mathcal{M}_b(U;\mathbb{R}^2)} \le |(\mathscr{B}_{\mathrm{sym}} w) * (\mathscr{B}_{\mathrm{sym}} v)|_{\mathcal{M}_b(U;\mathbb{M}^{2\times 2})} \le \|\nabla w\|_{L^1(U;\mathbb{M}^{2\times 2})} |\mathscr{B}_{\mathrm{sym}}(v)|_{\mathcal{M}_b(U;\mathbb{M}^{2\times 2})}.$$

The same computation holds for $w \in C^\infty_c(Q; \mathbb{R}^2)$ and $v \in BV_{\mathscr{B}}(Q; \mathbb{R}^2)$. This proves that Assertion 2 in Definition 6.1 is also satisfied.

We finally construct an example of a training set $\Sigma[\mathscr{A}]$. For every $0 \le s, t \le 1$, we define

$$B_t := \left[\begin{array}{cc|cc} 1 & 0 & t & 0 \\ 0 & (1-t) & 0 & 0 \end{array}\right] \text{ and } B_s := \left[\begin{array}{cc|cc} 1 & 0 & s & 0 \\ 0 & 1-s & 0 & 0 \end{array}\right],$$

and we set

$$\mathscr{B}_{s,t}(v) := B_t \partial_1 v + B_s \partial_2 v = \begin{bmatrix} \partial_1 v_1 & (1-t)\partial_1 v_2 + (1-s)\partial_2 v_1 \\ t\partial_1 v_2 + s\partial_2 v_1 & \partial_2 v_2 \end{bmatrix}. \tag{7.8}$$

By a straightforward computation, we obtain that $\mathcal{N}(\mathscr{B}_{s,t})$ is finite dimensional for every $0 \le s, t \le 1$. Additionally, Assertion 1 in Definition 6.1 follows by adapting the arguments in Remark 7.3. Finally, arguing exactly as in (7.7), we obtain that

$$(\mathscr{A}w) * v = (\mathscr{B}_{t,s} w) * (\mathscr{B}_{s,t}(v)), \text{ for every } w, v \in C^\infty(\bar{Q}; \mathbb{R}^2),$$

which implies that

$$\begin{aligned} |(\mathscr{A}w) * v|_{\mathcal{M}_b(Q;\mathbb{R}^2)} &\le \|\mathscr{B}_{t,s} w\|_{L^1(Q;\mathbb{M}^{2\times 2})} |\mathscr{B}_{s,t}(v)|_{\mathcal{M}_b(Q;\mathbb{M}^{2\times 2})} \\ &\le 2 \|\nabla w\|_{L^1(Q;\mathbb{R}^{N\times N})} |\mathscr{B}_{s,t}(v)|_{\mathcal{M}_b(Q;\mathbb{M}^{2\times 2})}. \end{aligned}$$

Hence, we deduce again Statement 2 in Definition 6.1. Therefore, the collection $\Sigma[\mathscr{A}]$ given by

$$\Sigma[\mathscr{A}] := \{\mathscr{B}_{s,t} :\ 0 \le s, t \le 1\}$$

is a training set according to Definition 6.10. We remark that $\Sigma[\mathscr{A}]$ includes the operator TGV^2 (with $s = t = 1/2$) and the operator $NsTGV^2$ (with $t = 0$ and $s = 1$), as well as a collection of all "interpolating" regularizers. In other words, our training scheme $(\mathcal{T}^2_\theta)$ with training set $\Sigma[\mathscr{A}]$ is able to search for optimal results in a class of operators including the commonly used TGV^2 and $NsTGV^2$, as well as any interpolation regularizer.

7.2.1. *Comparison with other works.* In [8] the authors analyze a range of first order linear operators generated by diagonal matrixes. To be precise, letting $D = \mathrm{diag}(\beta_1, \beta_2, \beta_3, \beta_4)$, [8] treats first order operators $\mathscr{B}$ defined as

$$\mathscr{B}v := Q \cdot B \cdot Q \cdot (\nabla v)^T,$$

where

$$Q := \begin{bmatrix} 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & 1 \\ -1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \text{ and } \nabla v = [\partial_1 v_1, \partial_1 v_2, \partial_2 v_1, \partial_2 v_2].$$

That is, instead of viewing ∇v as a 2×2 matrix as we do, in [8] ∇v is represented as a vector in $\mathbb{R}^4$. In this way, the symmetric gradient $\mathcal{E}v$ in (7.6) can be written as

$$\begin{aligned}\mathcal{E}v = Q \cdot \mathrm{diag}(0, 1/2, 1/2, 1/2) \cdot Q \cdot (\nabla v)^T &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot [\partial_1 v_1, \partial_1 v_2, \partial_2 v_1, \partial_2 v_2]^T \\ &= [\partial_1 v_1, 0.5(\partial_1 v_2 + \partial_2 v_1), 0.5(\partial_1 v_2 + \partial_2 v_1), \partial_2 v_2].\end{aligned}$$

However, the representation above does not allow to consider skewed symmetric gradients $\mathscr{B}_{s,t}(v)$ with the structure introduced in (7.8). Indeed, let $s = t = 0.2$. We have

$$\mathscr{B}_{0.2,0.2}(v) = \begin{bmatrix} \partial_1 v_1 & 0.8\partial_1 v_2 + 0.8\partial_2 v_1 \\ 0.2\partial_1 v_2 + 0.2\partial_2 v_1 & \partial_2 v_2 \end{bmatrix}.$$

Rewriting the matrix above as a vector in $\mathbb{R}^4$, we obtain

$$\begin{aligned}\mathscr{B}_{0.2,0.2}(v) &= [\partial_1 v_1, 0.2(\partial_1 v_2 + \partial_2 v_1), 0.8(\partial_1 v_2 + \partial_2 v_1), \partial_2 v_2] \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.8 & 0.8 & 0 \\ 0 & 0.2 & 0.2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot [\partial_1 v_1, \partial_1 v_2, \partial_2 v_1, \partial_2 v_2]^T.\end{aligned}$$

That is, we would have

$$QD'Q = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.8 & 0.8 & 0 \\ 0 & 0.2 & 0.2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ or } D' = \begin{bmatrix} 0 & 0 & 0 & 0.3 \\ 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 \end{bmatrix},$$

which are not diagonal matrices. Hence, this example shows that our model indeed covers more operators that those discussed in [8].

7.3. **Numerical simulations and observations.** Let $\mathscr{A}$ be the operator defined in Subsection 7.2, and let

$$\Sigma[\mathscr{A}] := \{\mathscr{B}_{s,t} :\ s,t \in [0,1]\}$$

where, for $0 \leq s,t \leq 1$, $\mathscr{B}_{s,t}$ are the first order operators introduced in (7.8). As we remarked before, the seminorm $PGV^2_{\mathscr{B}_{s,t}}$ interpolates between the TGV^2 and $NsTGV^2$ regularizers. We define the cost function $\mathcal{C}(\alpha,s,t)$ to be

$$\mathcal{C}(\alpha,s,t) := \left\| u_{\alpha,\mathscr{B}_{s,t}} - u_c \right\|_{L^2(Q)}. \tag{7.9}$$

From Theorem 5.4 we have that $\mathcal{C}(\alpha,s,t)$ admits at least one minimizer $(\tilde\alpha,\tilde s,\tilde t) \in \mathbb{R}^+ \times [0,1] \times [0,1]$.

To explore the numerical landscapes of the cost function $\mathcal{C}(\alpha,s,t)$, we consider the discrete box-constraint

$$\begin{aligned}(\alpha_0,\alpha_1,s,t) &\in \{0.025,\, 0.05,\, 0.075,\ldots,1\}\\ &\times \{0.025,\, 0.05,\, 0.075,\ldots,1\} \times \{0,\, 0.025,\, 0.05,\,\ldots,\,1\} \times \{0,\, 0.025,\, 0.05,\,\ldots,\,1\}.\end{aligned} \tag{7.10}$$

We perform numerical simulations of the images shown in Figure 1: the first image represents a clean image u_c, whereas the second one is a noised version u_η, with heavy artificial Gaussian noise. The reconstructed image $u_{\alpha,\mathscr{B}}$ in Level 2 of our training scheme is computed by using the primal-dual algorithm presented in [9].

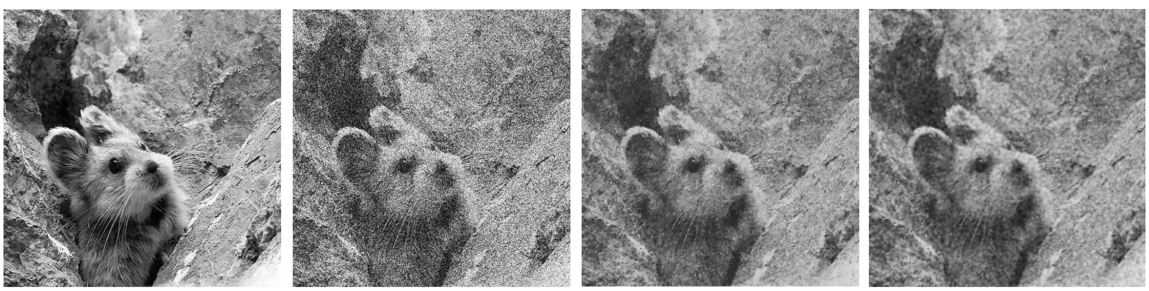

FIGURE 1. From left to right: the test image of a Pika; a noised version (with heavy artificial Gaussian noise); the optimally reconstructed image with TGV regularizer; the optimally reconstructed image with PGV regularizer.

It turns out that the minimum value of (7.9), taking values in (7.10), is achieved at $\tilde\alpha_0 = 0.072$, $\tilde\alpha_1 = 0.575$, $\tilde s = 0.95$, and $\tilde t = 0.05$. The optimal reconstruction $u_{\tilde\alpha,\mathscr{B}_{\tilde s,\tilde t}}$ is the last image in Figure 1, whereas the optimal result with $\mathscr{B}_{s,t} \equiv \mathcal{E}$, i.e., $u_{\tilde\alpha,TGV}$, is the third image in Figure 1. Although the optimal reconstructed image $u_{\tilde\alpha,\mathscr{B}_{\tilde s,\tilde t}}$ and $u_{\tilde\alpha,\mathcal{E}}$ do not present too many differences to the naked eye, we do have that

$$\mathcal{C}(\tilde\alpha,\tilde s,\tilde t) < \mathcal{C}(\tilde\alpha, 0.5, 0.5)$$

(see also Table 1 below). That is, the reconstructed image $u_{\tilde\alpha,\mathscr{B}_{\tilde s,\tilde t}}$ is indeed "better" in the sense of our training scheme (L^2-difference).
To visualize the change of cost function produced by different values of $(s,t) \in [0,1]^2$, we fix $\bar\alpha_0 = 0.072$ and $\bar\alpha_1 = 0.575$ and plot in Figure 2 the mesh and contour plot of $\mathcal{C}(\bar\alpha,s,t)$.
We again remark that the introduction of $PGV_{\alpha,\mathscr{B}[k]}$ regularizers into the training scheme is only meant to expand the training choices, but not to provide a superior seminorm with respect to the

Regularizer	optimal solution	minimum cost value
TGV^2	$\tilde{\alpha}_0 = 0.074$, $\tilde{\alpha}_1 = 0.625$	$\mathcal{C}(\tilde{\alpha}, 0.5, 0.5) = 18.653$
PGV^2	$\tilde{\alpha}_0 = 0.072$, $\tilde{\alpha}_1 = 0.575$, $\tilde{s} = 0.95$, $\tilde{t} = 0.05$	$\mathcal{C}(\tilde{\alpha}, \tilde{s}, \tilde{t}) = 17.6478$

TABLE 1. minimum cost value with different regularizers. The minimum value of the cost function for the PGV^2- regularizer is approximately 5% below that of the TGV^2- regularizer.

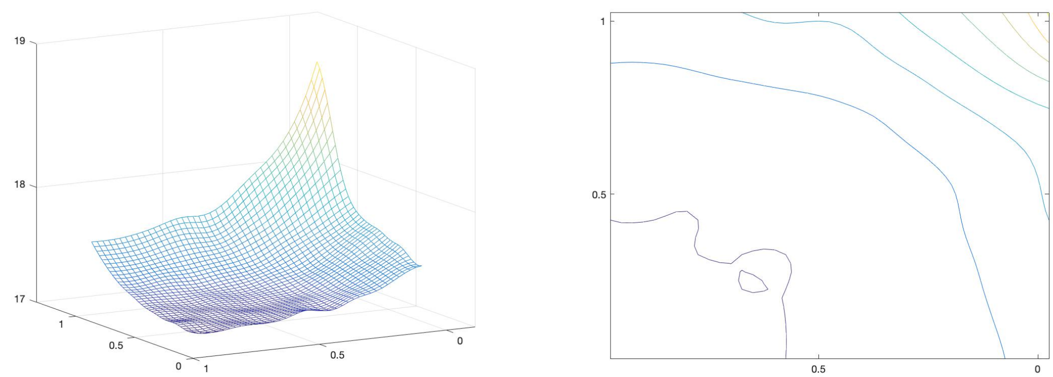

FIGURE 2. From the left to the right: mesh and contour plot of the cost function $\mathcal{C}(\bar{\alpha}, s, t)$ in which $\bar{\alpha} = (\bar{\alpha}_0, \bar{\alpha}_1)$ is fixed, $(s, t) \in [0,1]^2$.

popular choices TGV^2 or $NsTGV^2$. The fact whether the optimal regularizer is TGV^2, $NsTGV^2$ or an intermediate regularizer is completely dependent on the given training image $u_\eta = u_c + \eta$.

We conclude this section with a further study of the numerical landscapes associated to the cost function $\mathcal{C}(\alpha, s, t)$. We consider also in this second example the discrete box-constraint in (7.10), and we analyze the images shown in Figure 3: also in this second example the first image represents the clean image u_c, whereas the second one is a noised version u_η. The reconstructed image $u_{\alpha,\mathscr{B}}$ in Level 2 of our training scheme is again computed by using the primal-dual algorithm presented in [9].
We report that the minimum value of (7.9), taking values in (7.10), is achieved at $\tilde{\alpha}_0 = 5.6$, $\tilde{\alpha}_1 = 1.2$, $\tilde{s} = 0.8$, and $\tilde{t} = 0.2$. The optimal reconstruction $u_{\tilde{\alpha},\mathscr{B}_{\tilde{s},\tilde{t}}}$ is the last image in Figure 3, whereas the optimal result with $\mathscr{B}_{s,t} \equiv \mathcal{E}$, i.e., $u_{\tilde{\alpha},TGV}$, is the third image in Figure 3. Although the optimal reconstructed image $u_{\tilde{\alpha},\mathscr{B}_{\tilde{s},\tilde{t}}}$ and $u_{\tilde{\alpha},TGV}$ do not present too many differences with respect to our eyesight, we do have, also in this case, that

$$\left\| u_{\tilde{\alpha},\mathscr{B}_{\tilde{s},\tilde{t}}} - u_c \right\|_{L^2(Q)} < \left\| u_{\tilde{\alpha},TGV} - u_c \right\|_{L^2(Q)}.$$

Namely, the reconstructed image $u_{\tilde{\alpha},\mathscr{B}_{\tilde{s},\tilde{t}}}$ is indeed "better" in the sense of our training scheme (L^2-difference).

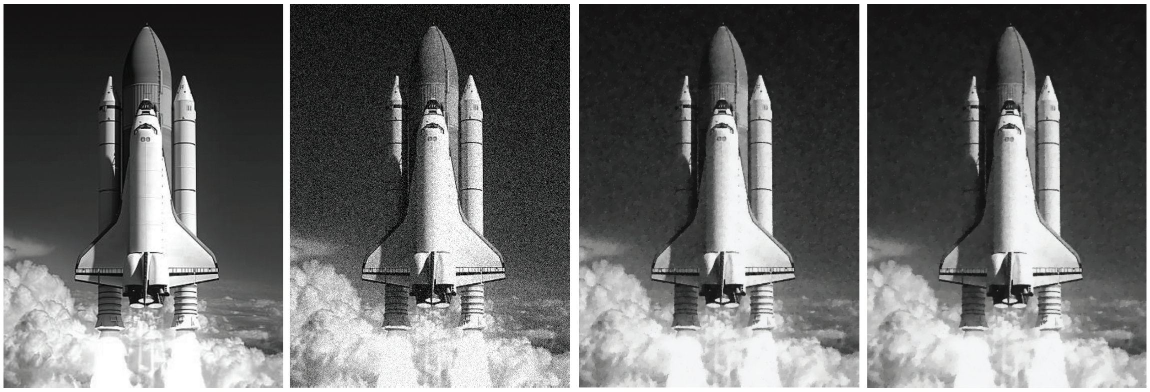

FIGURE 3. From left to right: the test image of a space shuttle; a noised version (with artificial Gaussian noise); the optimally reconstructed image $u_{\tilde{\alpha},\mathcal{E}}$, where $\tilde{\alpha}_0 = 5.2$, $\tilde{\alpha}_1 = 1.9$; the optimally reconstructed image $u_{\tilde{\alpha},\mathscr{B}_{\tilde{s},\tilde{t}}}$, where $\tilde{\alpha}_0 = 5.6$, $\tilde{\alpha}_1 = 1.2$, $\tilde{s} = 0.8$, and $\tilde{t} = 0.2$

To visualize the change of cost function produced by different values of $(s,t) \in [0,1]^2$, we fix $\bar{\alpha}_0 = 5.6$ and $\bar{\alpha}_1 = 1.9$ and plot in Figure 4 the mesh and contour plot of $\mathcal{C}(\bar{\alpha}, s, t)$ as follows.

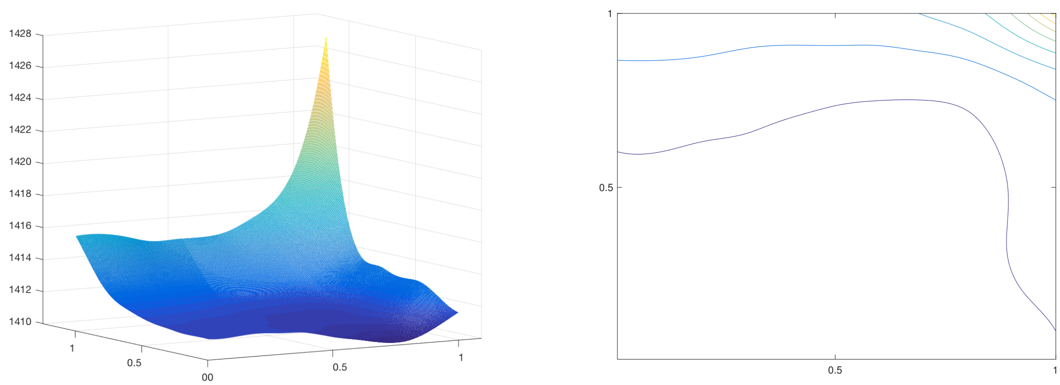

FIGURE 4. From left to right: mesh and contour plot of the cost function $\mathcal{C}(\bar{\alpha}, s, t)$ in which $\bar{\alpha} = (\bar{\alpha}_0, \bar{\alpha}_1)$ is fixed, $(s,t) \in [0,1]^2$. The function $\mathcal{C}(\bar{\alpha}, s, t)$ achieves the minimum at $\tilde{s} = 0.8$ and $\tilde{t} = 0.2$.

Acknowledgements

The work of Elisa Davoli has been funded by the Austrian Science Fund (FWF) projects F65, V 662, and I 4052, as well as by BMBWF through the OeAD-WTZ project CZ04/2019. Irene Fonseca thanks the Center for Nonlinear Analysis for its support during the preparation of the manuscript. She was supported by the National Science Foundation under Grants No. DMS-1411646 and DMS-1906238. The work of Pan Liu has been supported by the Centre of Mathematical Imaging and Healthcare and funded by the Grant "EPSRC Centre for Mathematical and Statistical Analysis of Multimodal Clinical Imaging" with No. EP/N014588/1. All authors are thankful to the Erwin Schrödinger Institute in Vienna, where part of this work has been developed during the workshop "New trends in the variational modeling of failure phenomena".

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] T. Barbu and G. Marinoschi. Image denoising by a nonlinear control technique. *International Journal of Control*, **90** (2017),1005–1017.

[3] K. Bredies and M. Holler. Regularization of linear inverse problems with total generalized variation. *Journal of inverse and ill-posed problems*, **22** (1993), 871–913.

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

[5] K. Bredies and T. Valkonen. Inverse problems with second-order total generalized variation constraints. *Proceedings of SampTA 2011 - 9th International Conference on Sampling Theory and Applications*, Singapore, (2011).

[6] D. Breit, L. Diening, and F. Gmeineder. Traces of functions of bounded A-variation and variational problems with linear growth. *Preprint arXiv:1707.06804*.

[7] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Universitext. Springer, New York, 2011.

[8] E.-M. Brinkmann, M. Burger, and J. S. Grah. Unified models for second-Order TV-type regularisation in imaging - a new perspective based on vector operators. *Journal of Mathematical Imaging and Vision*, **61** (2019), 571–601.

[9] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, **40** (2011), 120–145.

[10] Y. Chen, T. Pock, R. Ranftl, and H. Bischof. Revisiting loss-specific training of filter-based MRFs for image restoration. In *Pattern Recognition*, pages 271–281. Springer, 2013.

[11] Y. Chen, R. Ranftl, and T. Pock. Insights into analysis operator learning: From patch-based sparse models to higher order MRFs. *IEEE Transactions on Image Processing*, **23** (2014), 1060–1072.

[12] E. Davoli and P. Liu. One dimensional fractional order TGV : Gamma-convergence and bilevel training scheme. *Commun. Math. Sci.* **16** (2018), 213–237.

[13] 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** (2016), 464–500.

[14] J. C. De los Reyes, C.-B. Schönlieb, and T. Valkonen. Bilevel parameter learning for higher-order total variation regularisation models. *Journal of Mathematical Imaging and Vision*, **57** (2017), 1–25.

[15] J. Domke. Generic methods for optimization-based modeling. In *AISTATS*, volume 22, pages 318–326, 2012.

[16] C. Fefferman. C^m extension by linear operators. *Ann. of Math. (2)* **166** (2007), 779–835.

[17] E. Gagliardo. Caratterizzazioni delle tracce sulla frontiera relative ad alcune classi di funzioni in n variabili. *Rend. Sem. Mat. Univ. Padova*, **27** (1957), 284–305.

[18] F. Gmeineder and B. Raiţă. Embeddings for $\mathbb{A}$-weakly differentiable functions on domains. *Journal of Functional Analysis*, **277** (2019), 108278.

[19] F. Gmeineder, B. Raiţă, and J. Van Schaftingen. On limiting trace inequalities for vectorial differential operators. *Indiana Univ. Math. J.*, to appear.

[20] A. Guerra and B. Raiţă. On the necessity of the constant rank condition for L^p estimates. *Comptes Rendus Mathmatique*, to appear.

[21] A. Guerra, B. Raiţă, and M.R.I. Schrecker. Compensated compactness: continuity in optimal weak topologies. *Preprint arXiv: 2007.00564*.

[22] G. C. Hsiao and W. L. Wendland. *Boundary integral equations*. Springer, 2008.

[23] I. Fonseca and P. Liu. The weighted Ambrosio-Tortorelli approximation scheme. *SIAM J. Math. Anal.*, **49** (2017), 4491–4520.

[24] J. Kristensen and B. Raiţă. Oscillation and concentration in sequences of PDE constrained measures. *Preprint arXiv:1912.09190*

[25] G. Leoni. *A first course in Sobolev spaces*, volume 105 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2009.

[26] B. Raiţă. L^1-estimates for constant rank operators. *Preprint arXiv: 1811.10057*.

[27] B. Raiţă. Potentials for $\mathscr{A}$-quasiconvexity. *Calc. Var. PDEs* **58** (2019), article 105.

[28] B. Raiţă. Critical L^p-differentiability of $\mathrm{BV}^{\mathbb{A}}$-maps and canceling operators *Trans. Amer. Math. Soc.* **372** (2019), 7297–7326.

[29] B. Raiţă and A. Skorobogatova. Continuity and canceling operators of order n on $\mathbb{R}^n$. *Calc. Var. PDEs* **59** (2020), article 85.

[30] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, **60** (1992), 259–268.

[31] M. F. Tappen, C. Liu, E. H. 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, June 2007.

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

[33] T. Valkonen. The jump set under geometric regularisation. Part 2: Higher-order approaches. *Journal of Mathematical Analysis and Applications* **453** (2017), 1044–1085.

[34] T. Valkonen, K. Bredies and F. Knoll. Total generalized variation in diffusion tensor imaging. *SIAM J. Imaging Sci.* **6** (2013), 487–525.