



Adaptive Image Processing: First Order PDE Constraint Regularizers and a Bilevel Training Scheme

Elisa Davoli¹ · Irene Fonseca² · Pan Liu³

Received: 13 November 2020 / Accepted: 14 February 2023 / Published online: 3 March 2023
© The Author(s) 2023

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 under a conditional uniform bound on the trace constant of the operators and a finite-null-space condition. Some first examples and numerical results are given.

Keywords Image processing · Optimal training scheme · First order differential operators · Γ -convergence

Mathematics Subject Classification 26B30 · 94A08 · 47J20

Contents

1 Introduction	2
2 Notations and Preliminary Results	7

Communicated by Paul Newton.

✉ Elisa Davoli
elisa.davoli@tuwien.ac.at
Irene Fonseca
fonseca@andrew.cmu.edu
Pan Liu
dragonrider.liupan@gmail.com

- ¹ Institute of Analysis and Scientific Computing, TU Wien, Wiedner Hauptstrasse 8-10, 1040 Vienna, Austria
- ² Department of Mathematics, Center of Nonlinear Analysis, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA
- ³ Department of Radiology, First Medical Center of Chinese PLA General Hospital, Beijing 100853, China

3 The Space of Functions with Bounded *PGV*-Seminorm 8
 3.1 The Space $BV_{\mathcal{B}}$ and the Class of Admissible Operators 8
 3.2 The *PGV*-Total Generalized Variation 9
 4 Γ -Convergence of Functionals Defined by *PGV*-Total Generalized Variation Seminorms 12
 5 The Bilevel Training Scheme with *PGV*-Regularizers 17
 6 Training Set $\Sigma[\mathcal{A}]$ Based on $(\mathcal{A}, \mathcal{B})$ Training Operators Pairs 20
 6.1 A Subcollection of Π Characterized by $(\mathcal{A}, \mathcal{B})$ Training Operators Pairs 20
 6.2 Training Scheme with Fixed and Multiple Operators \mathcal{A} 27
 7 Explicit Examples and Numerical Observations 29
 7.1 The Existence of Fundamental Solutions of Operators \mathcal{A} 29
 7.2 The Unified Approach to TGV^2 and $NsTGV^2$: An Example of $\Sigma[\mathcal{A}]$ 31
 7.2.1 Comparison with Other Works 32
 7.3 Numerical Simulations and Observations 34
 8 Conclusions 36
 References 37

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

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

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}_α 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 one to reconstruct a “clean” image based on the functional properties of the regularizer \mathcal{R}_α .

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

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

An example is the ROF model (Rudin et al. 1992), 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. Ambrosio et al. (2000, Chapter 4)], $\alpha \in \mathbb{R}^+$ is the tuning parameter, and we have

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

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 might 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 De Los Reyes et al. (2016, 2017) the authors proposed a bilevel training scheme, which was originally introduced in Machine Learning and later adopted by the imaging processing community (see Chen et al. 2013, 2014; Domke 2012; Tappen et al. 2007). 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

$$\text{Level 1. } \quad \tilde{\alpha} \in \arg \min \left\{ \|u_{\alpha,TV} - u_c\|_{L^2(Q)}^2 : \alpha \in \mathbb{R}^+ \right\}, \tag{T-L1}$$

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

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}_α 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_α^2 , defined as

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

has been characterized in Bredies et al. (2010), where Du denotes the distributional gradient of u , $(\text{sym}\nabla)v := (\nabla v + \nabla^T v)/2$, $\mathcal{M}_b(Q; \mathbb{R}^{N \times N})$ 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_α^2 -seminorm defined above, namely the $NsTGV_\alpha^2$ functional (see e.g., Valkonen et al. 2013; Valkonen 2017). The different regularizers have been shown to have several perks and drawbacks for image reconstruction. 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 (T) by inserting different regularizers inside the training level 2 in (T-L1). For example, one could set

$$\begin{aligned}
 \text{Level 1.} \quad & (\tilde{\mathcal{R}}_\alpha) := \arg \min \left\{ \|u_{\alpha, \mathcal{R}} - u_c\|_{L^2(Q)}^2 : \mathcal{R}_\alpha \in \{\alpha TV, TGV_\alpha^2, NsTGV_\alpha^2\} \right\}, \\
 \text{Level 2.} \quad & u_{\alpha, \mathcal{R}} := \arg \min \left\{ \|u - u_\eta\|_{L^2(Q)}^2 + \mathcal{R}_\alpha(u) : u \in L^1(Q) \right\}. \tag{1.3}
 \end{aligned}$$

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 (T) finitely many times, at each time with a different regularizer \mathcal{R}_α . In turn, some possible texture effects for which an “intermediate” (or interpolated) reconstruction between the one provided by, say, TGV_α^2 and $NsTGV_\alpha^2$, 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 $PGV_{\alpha, \mathcal{B}}^2$, defined as

$$\begin{aligned}
 PGV_{\alpha, \mathcal{B}}^2(u) := \inf \left\{ \alpha_0 |Du - v|_{\mathcal{M}_b(Q; \mathbb{R}^N)} + \alpha_1 |\mathcal{B}v|_{\mathcal{M}_b(Q; \mathbb{R}^{N \times N})} : \right. \\
 \left. v \in L^1(Q; \mathbb{R}^N), \mathcal{B}v \in \mathcal{M}_b(Q; \mathbb{R}^{N \times N}) \right\}, \tag{1.4}
 \end{aligned}$$

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

$$BPGV_{\alpha, \mathcal{B}}^2(Q) := \left\{ u \in L^1(Q) : PGV_{\alpha, \mathcal{B}}^2(u) < +\infty \right\}.$$

Note that if $\mathcal{B} := \text{sym}\nabla$, then the operator $PGV_{\alpha, \mathcal{B}}^2$ defined in (1.4) coincides with the operator TGV_α^2 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 Brinkmann et al. (2019) (see Sect. 7.2). Moreover, the collection of functionals described in (1.4) naturally incorporates the recent PDE-based approach to image denoising formulated in Barbu and Marinoschi (2017) via nonconvex optimal control problem, thus offering a very general and abstract framework to simultaneously describe a variety of different image-processing techniques. Adding to the model higher-order regularizations which can be different from the symmetric gradient additionally allows to enhance image reconstruction in one direction more than in the others, thus paving the way for furthering the study of anisotropic noise-reduction.

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 $PGV_{\alpha, \mathcal{B}}^2(u)$ in the class described in (1.4) and an optimal tuning parameter α , so that the corresponding reconstructed image $u_{\alpha, \mathcal{B}}$, obtained in Level 2 of the (\mathcal{T}_θ^2) -scheme (see $(\mathcal{T}_\theta^2$ -L2) below), minimizes the L^2 -distance from the original clean image u_c . To be precise, in Sects. 3, 4, and 5 we study the improved training scheme \mathcal{T}_θ^2 for $\theta \in (0, 1)$, defined as follows

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

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

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

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

is continuous in the L^1 topology, in the sense of Γ -convergence, with respect to the parameters α and the operators \mathcal{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}_θ^2) admits at least one solution $(\tilde{\alpha}, \tilde{\mathcal{B}}) \in [\theta, \frac{1}{\theta}]^2 \times \Sigma$, and provides an associated optimally reconstructed image $u_{\tilde{\alpha}, \tilde{\mathcal{B}}} \in BV(Q)$.*

The collection Σ of operators \mathcal{B} used in $(\mathcal{T}_\theta^2\text{-L1})$ has to satisfy several natural regularity and ellipticity assumptions, which are fulfilled by $\mathcal{B} := \nabla$ and $\mathcal{B} := \text{sym} \nabla$ (see Sect. 7.2.1). The general requirements on \mathcal{B} that allow scheme (\mathcal{T}_θ^2) to have a solution are listed on Assumptions 3.2 and 3.3. Later in Sect. 6, as the third main contribution of this article, we provide in Definition 6.1 a collection of operators \mathcal{B} satisfying Assumptions 3.2 and 3.3 under some uniform bounds on the behavior of their traces and under finiteness of their null spaces. A simplified statement of our result is the following (see Theorem 6.5 for the detailed formulation).

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

The requirements collected in Definition 6.1 and the analysis in Sect. 6 move from the observation that a fundamental property that the admissible operators \mathcal{B} must satisfy is to ensure that the set of maps $v \in L^1(Q; \mathbb{R}^N)$ such that $\mathcal{B}v$ is a bounded

Radon measure (henceforth denoted by $BV_{\mathcal{B}}(Q; \mathbb{R}^N)$) must embed compactly in $L^1(Q; \mathbb{R}^N)$. In the case in which \mathcal{B} coincides with ∇ or $\text{sym}\nabla$, a crucial ingredient is Kolmogorov–Riesz compactness theorem (see Brezis 2011, Theorem 4.26 and Proposition 6.6). In particular, for $\mathcal{B} = \text{sym}\nabla$ the key point of the proof is to guarantee that bounded sets $\mathcal{F} \subset BD(Q)$ satisfy

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

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

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

where ϕ is a fundamental solution for curlcurl , and where δ and δ_h denote the Dirac deltas centered in the origin and in h , respectively. In the case in which $\mathcal{B} = \text{sym}\nabla$ 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 $\text{curlcurl}((\delta_h * \phi - \phi) * f)$ by means of the total variation of $(\text{sym}\nabla)f$ and owing to the regularity of the fundamental solution of curlcurl . The operator \mathcal{A} in Theorem 1.2 plays the role of curlcurl in the case in which $\text{sym}\nabla$ is replaced by a generic operator \mathcal{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 $(\mathcal{A}, \mathcal{B})$.

Finally, in Sect. 7.2 we give some explicit examples to show that our class of regularizers $PGV_{\alpha, \mathcal{B}}^2$ includes the seminorms TGV_{α}^2 and $NsTGV_{\alpha}^2$, 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 Davoli and Liu (2018) 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 the very first numerical simulations of the proposed bilevel training scheme. Although this work focuses mainly on the theoretical analysis of the operators $PGV_{\alpha, \mathcal{B}}^2$ and on showing the existence of optimal results for the training scheme (\mathcal{T}_{θ}^2) , in Sect. 7.3 a primal-dual algorithm for solving $(\mathcal{T}_{\theta}^2\text{-L2})$ 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 $(\tilde{\alpha}, \tilde{\mathcal{B}})$ for the training scheme (\mathcal{T}_{θ}^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 Fonseca and Liu 2017).

This paper is organized as follows: in Sect. 2 we collect some notations and preliminary results. In Sect. 3 we analyze the main properties of the $P GV_{\alpha, \mathcal{B}}^2$ -seminorms. The Γ -convergence result and the bilevel training scheme are the subjects of Sects. 4 and 5, respectively. We point out that the results in Sects. 3 and 4 are direct generalizations of the works in Bredies and Valkonen (2011), Bredies and Holler (1993). 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 Sects. 3–5, and proceed directly with the content of Sect. 6. Section 6 is devoted to the analysis of the space $BV_{\mathcal{B}}$ for suitable differential operators \mathcal{B} . The numerical implementation of some explicit examples is performed in Sect. 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 \mathcal{B} will be used for first order differential operators $\mathcal{B} : \mathcal{D}'(U; \mathbb{R}^N) \rightarrow \mathcal{D}'(U; \mathbb{R}^{N \times N})$ defined as

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

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 $i = 1, \dots, N$.

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

$$\|\mathcal{B}_n - \mathcal{B}\|_{\ell^\infty} := \sum_{i=1}^N \|B_n^i - B^i\| \rightarrow 0, \quad (2.2)$$

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 $BV_{\mathcal{B}}$ and the Class of Admissible Operators

We generalize the standard total variation seminorm by using first order differential operators $\mathcal{B}: \mathcal{D}'(Q; \mathbb{R}^N) \rightarrow \mathcal{D}'(Q; \mathbb{R}^{N \times N})$ in the form (2.1).

Definition 3.1 We define the space of tensor-valued functions $BV_{\mathcal{B}}(Q; \mathbb{R}^N)$ as

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

and we equip it with the norm

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

We refer to Raiřă (2019), Raiřă and Skorobogatova (2020) for some recent results on $BV_{\mathcal{B}}$ -spaces for elliptic and cancelling operators, as well as to Kristensen and Raiřă (2022) 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 $BV_{\mathcal{B}}$ are connected to the theory of $W^{1,p}$ -spaces, cf. (Gmeineder and Raiřă 2019; Gmeineder et al. 2019; Raiřă 2018, 2019). See also Guerra and Raiřă (2020) for a related compensated-compactness study.

In order to introduce the class of admissible operators, we first list some assumptions on the operator \mathcal{B} .

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

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

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

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

We point out that all requirements above are satisfied for $\mathcal{B} := \nabla$.

Assumption 3.3 The following compactness property applies to a collection of operators $\{\mathcal{B}_n\}_{n=1}^\infty$. For $\{v_n, \mathcal{B}_n\}_{n=1}^\infty$ such that \mathcal{B}_n satisfies Assumption 3.2 for every $n \in \mathbb{N}$, and

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

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

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

and

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

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

In Sect. 6 we will exhibit a subclass of operators $\mathcal{B} \in \Pi$ 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 $\mathcal{B}: \mathcal{D}'(Q; \mathbb{R}^N) \rightarrow \mathcal{D}'(Q; \mathbb{R}^{N \times N})$, $\mathcal{B} \in \Pi$, we consider the seminorm

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

where the space $BV_{\mathcal{B}}$ is introduced in Definition 3.1.

We note that for all $\alpha \in \mathbb{R}_+^2$, the seminorms $PGV_{\alpha, \mathcal{B}}^2$ are topologically equivalent. With a slight abuse of notation, in what follows we will write $PGV_{\mathcal{B}}^2$ instead of $PGV_{\alpha, \mathcal{B}}^2$ 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

$$BPGV_{\mathcal{B}}^2(Q) := \left\{ u \in L^1(Q) : PGV_{1, \mathcal{B}}^2(u) < +\infty \right\},$$

and we write

$$\|u\|_{BPGV_{\mathcal{B}}^2(Q)} := \|u\|_{L^1(Q)} + PGV_{1, \mathcal{B}}^2(u).$$

We next show that the $PGV_{\mathcal{B}}^2$ -seminorm is finite if and only if the TV -seminorm is. The next three propositions show some basic properties of the PGV^2 regularizers. The expert Reader might skip their proof and proceed directly to Sect. 4.

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

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

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

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

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

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

It suffices to observe that

$$\begin{aligned} |Du|_{\mathcal{M}_b(Q; \mathbb{R}^N)} &\leq |Du - \bar{v}|_{\mathcal{M}_b(Q; \mathbb{R}^N)} + \|\bar{v}\|_{L^1(Q; \mathbb{R}^N)} \\ &\leq PGV_{\mathcal{B}}^2(u) + 1 + \|\bar{v}\|_{L^1(Q; \mathbb{R}^N)} < +\infty. \end{aligned}$$

□

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

Proposition 3.8 *Let $u \in BV(Q)$ and let \mathcal{B} satisfy Assumption 3.2. Then, for $\alpha \in \mathbb{R}_+^2$ there exists a function $v \in BV_{\mathcal{B}}(Q; \mathbb{R}^N)$ 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 $PGV_{\mathcal{B}}^2(u) < +\infty$.

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

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

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

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

and

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

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 $\mathcal{B}v \in \mathcal{M}_b(Q; \mathbb{R}^{N \times N})$ such that, up to the extraction of a subsequence (not relabeled), there holds

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

and

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

The minimality of v follows by lower-semicontinuity. □

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

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

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

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

$$\lim_{n \rightarrow \infty} PGV_{\alpha_n, \mathcal{B}_n}^2(u) = PGV_{\alpha, \mathcal{B}}^2(u).$$

Proof We first claim that

$$\liminf_{n \rightarrow \infty} PGV_{\alpha_n, \mathcal{B}_n}^2(u) \geq PGV_{\alpha, \mathcal{B}}^2(u). \tag{3.8}$$

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

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

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

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

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

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

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

$$\begin{aligned} \liminf_{n \rightarrow \infty} PGV_{\alpha_n, \mathcal{B}_n}^2(u) &= \liminf_{n \rightarrow \infty} \left[\alpha_n^0 |Du - v_n|_{\mathcal{M}_b(Q; \mathbb{R}^N)} + \alpha_n^1 |\mathcal{B}_n v_n|_{\mathcal{M}_b(Q; \mathbb{R}^{N \times N})} \right] \\ &\geq (1 - \varepsilon)\alpha_0 \liminf_{n \rightarrow \infty} |Du - v_n|_{\mathcal{M}_b(Q; \mathbb{R}^N)} + (1 - \varepsilon)\alpha_1 \\ &\quad \liminf_{n \rightarrow \infty} |\mathcal{B}_n v_n|_{\mathcal{M}_b(Q; \mathbb{R}^{N \times N})} \\ &\geq (1 - \varepsilon)\alpha_0 |Du - v|_{\mathcal{M}_b(Q; \mathbb{R}^N)} + (1 - \varepsilon)\alpha_1 |\mathcal{B}v|_{\mathcal{M}_b(Q; \mathbb{R}^{N \times N})} \\ &\geq (1 - \varepsilon)PGV_{\alpha, \mathcal{B}}^2(u), \end{aligned}$$

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

We now claim that

$$\limsup_{n \rightarrow \infty} PGV_{\alpha_n, \mathcal{B}_n}^2(u) \leq PGV_{\alpha, \mathcal{B}}^2(u). \tag{3.10}$$

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

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

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,

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

Since

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

we obtain

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

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 \mathcal{B} . For $r > 0$ we denote [see (2.2)]

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

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

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

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

The following theorem is the main result of this section.

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

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

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

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

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

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

(Liminf inequality) If

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

then

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

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

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

and

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

We subdivide the proof of Theorem 4.2 into two propositions.

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

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

Note that this projection operator is well defined owing to the assumption that $\mathcal{N}(\mathcal{B})$ is finite dimensional [see Brezis (2011, p. 38, Definition and Example 2) and Breit et al. (2017, Subsection 3.1)].

Next we have an enhanced version of Korn’s inequality.

Proposition 4.3 *Let $\mathcal{B} \in \Pi$ and let $r > 0$ be small enough so that elements of $(\mathcal{B})_r$ have finite dimensional kernel. Then, under Assumption 3.3 there exists a constant $C = C(\mathcal{B}, Q, r)$, depending only on \mathcal{B} , on the domain Q , and on r , such that*

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

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

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

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

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

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

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

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

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

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

Thus, by (4.3) we have

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

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

By the joint continuity of the projection operator, by (4.4) we have

$$\|\mathbb{P}_{\tilde{\mathcal{B}}}(\tilde{v})\|_{L^1(Q)} = \|\mathbb{P}_{\mathcal{B}_n}(\tilde{v}_n) - \mathbb{P}_{\tilde{\mathcal{B}}}(\tilde{v})\|_{L^1(Q)} \leq C\|\mathcal{B}_n - \tilde{\mathcal{B}}\|_{\ell^\infty} + \|\tilde{v} - \tilde{v}_n\|_{L^1(Q)} \rightarrow 0.$$

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

The following proposition is instrumental for establishing the liminf inequality.

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

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

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

$$\liminf_{n \rightarrow \infty} PPGV_{\alpha_n, \mathcal{B}_n}^2(u_n) \geq PPGV_{\alpha, \mathcal{B}}^2(u).$$

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

$$\|u\|_{BPGV_{\mathcal{B}}^2(Q)} \leq \|u\|_{BV(Q)} \leq C_r \|u\|_{BPGV_{\mathcal{B}}^2(Q)}, \tag{4.8}$$

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

Indeed, by Definitions 3.5 and 3.6 we always have

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

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

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

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

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

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

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

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

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 $u_n \rightarrow 0$ strongly in $L^1(Q)$ and

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

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

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

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

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

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.,

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

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

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

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

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

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

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

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

and we conclude (4.13).

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

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

In particular, in view of (4.6) we have

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

Since $\mathcal{B}_n \rightarrow \mathcal{B}$ in ℓ^∞ , choosing $r > 0$ small enough there exists $N > 0$ such that $\mathcal{B}_n \subset (\mathcal{B})_1$ for all $n \geq N$. Thus, by (4.8) and (4.16), we infer that

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

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 $v_n \in BV_{\mathcal{B}_n}(Q; \mathbb{R}^N)$ such that,

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

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

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

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. \square

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

$$\limsup_{n \rightarrow \infty} PGV_{\alpha_n, \mathcal{B}_n}^2(u_n) \leq PGV_{\alpha, \mathcal{B}}^2(u).$$

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

We close Sect. 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. \square

5 The Bilevel Training Scheme with PGV-Regularizers

In this section, we introduce a bilevel training scheme associated to our class of regularizers and show its well-posedness. 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 Sect. 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}_θ^2) are

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

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

We first show that the Level 2 problem in $(T_\theta^2\text{-L2})$ 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 $\mathcal{B} \in \Sigma$, and let $\alpha \in \mathbb{R}_+^2$. Then there exists $u_{\alpha, \mathcal{B}} \in BV(Q)$ such that*

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

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

$$\|u_n - u_\eta\|_{L^2(Q)}^2 + PGV_{\mathcal{B}}^2(u_n) \leq \inf \left\{ \|u - u_\eta\|_{L^2(Q)}^2 + PGV_{\mathcal{B}}^2(u) : u \in BV(Q) \right\} + 1/n, \tag{5.1}$$

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

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

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

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

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

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

Setting

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

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

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

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

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

such that

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

and

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

Additionally, (5.5) and (5.7) yield

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

and

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

Since by (5.8) $D\tilde{u}_n \rightarrow 0$ 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

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

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

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

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

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

□

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

Proof By the boundedness and closedness of Σ in ℓ^∞ , up to a subsequence (not relabeled), there exists $(\tilde{\alpha}, \tilde{\mathcal{B}}) \in [\theta, 1/\theta]^2 \times \Sigma$ such that $\alpha_n \rightarrow \tilde{\alpha}$ in \mathbb{R}^2 and $\mathcal{B}_n \rightarrow \tilde{\mathcal{B}}$ in ℓ^∞ . Therefore, in view of Theorem 4.2 and the Fundamental Theorem of Γ -convergence (see, e.g. Dal Maso 1993), we have

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

where $u_{\alpha_n, \mathcal{B}_n}$ and $u_{\tilde{\alpha}, \tilde{\mathcal{B}}}$ are defined in (T_θ²-L2).

By (5.10), we have

$$\|u_{\tilde{\alpha}, \tilde{\mathcal{B}}} - u_c\|_{L^2(Q)} \leq \liminf_{n \rightarrow \infty} \|u_{\alpha_n, \mathcal{B}_n} - u_c\|_{L^2(Q)},$$

which completes the proof. □

6 Training Set Σ[\mathcal{A}] Based on (\mathcal{A} , \mathcal{B}) Training Operators Pairs

This section is devoted to providing a class of operators \mathcal{B} belonging to Π (see Definition 3.4), satisfying Assumption 3.3, and being closed with respect to the convergence in (2.2). Recall that $Q = (-\frac{1}{2}, \frac{1}{2})^N$.

6.1 A Subcollection of Π Characterized by (\mathcal{A} , \mathcal{B}) Training Operators Pairs

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

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

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

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

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 \mathcal{B} be a first order differential operator, $\mathcal{B} : \mathcal{D}'(U; \mathbb{R}^N) \rightarrow \mathcal{D}'(U; \mathbb{R}^{N \times N})$, given by

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

where $B^i \in \mathbb{M}^{N^3}$ for each $i = 1, \dots, N$, 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 (\mathcal{A} , \mathcal{B}) satisfying the ellipticity assumptions below.

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

1. For every $\lambda \in \{-1, 1\}^N$, the operator \mathcal{A} has a fundamental solution $P_\lambda \in L^1(\mathbb{R}^N; \mathbb{R}^N)$ such that:
 - a. $\mathcal{A}P_\lambda = \lambda\delta$, 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 $\frac{\partial^a}{\partial x^a} P_\lambda \in L^1_{loc}(\mathbb{R}^N; \mathbb{R}^N)$ for every multi-index $a \in \mathbb{N}^N$ with $|a| \leq d - 1$ (where d is the order of the operator \mathcal{A}).
- 2. For every open set $U \subset \mathbb{R}^N$ such that $Q \subset U$, and for every $u \in W^{d-1,1}(U; \mathbb{R}^N)$ and $v \in C_c^\infty(U; \mathbb{R}^N)$

$$\|(\mathcal{A}u)_i * v_i\|_{L^1(U)} \leq C_{\mathcal{A}} \left[\sum_{|a| \leq d-1} \left\| \frac{\partial^a}{\partial x^a} u \right\|_{L^1(U; \mathbb{R}^N)} \right] |\mathcal{B}v|_{\mathcal{M}_b(U; \mathbb{R}^N \times \mathbb{R}^N)}, \tag{6.1}$$

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

Note that, in view of 1b. one has, directly, the following property: 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

$$\sum_{|a|=d-1} \left\| \tau_h \left(\frac{\partial^a}{\partial x^a} P_\lambda \right) - \frac{\partial^a}{\partial x^a} P_\lambda \right\|_{L^1(U; \mathbb{R}^N)} =: M_{\mathcal{A}}(U; h) \rightarrow 0 \text{ as } |h| \rightarrow 0, \tag{6.2}$$

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

$$\tau_h w(x) := w(x + h) \text{ for every } w \in L^1(\mathbb{R}^N; \mathbb{R}^N) \text{ and for a.e. } x \in \mathbb{R}^N. \tag{6.3}$$

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

$$\mathcal{A}u := \left(\frac{\partial^2 u_1}{\partial x_1^2} \quad \frac{\partial^2 u_2}{\partial x_2^2} \right)^T \text{ for every } u = (u_1, u_2)^T \in D'(U; \mathbb{R}^2).$$

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

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

for every $i = 1, 2$. In other words, the pair $(\mathcal{A}, \mathcal{B})$ satisfies (6.1) with $C_{\mathcal{A}} = 1$.

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

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

The following extension result in $BV_{\mathcal{B}}$ is a corollary of the properties of the trace operator defined in Breit et al. (2017, Section 4).

Lemma 6.3 *Let $\mathcal{B} \in \Pi_{\mathcal{A}}$, and let $BV_{\mathcal{B}}(Q; \mathbb{R}^N)$ be the space introduced in Definition 3.1. Then there exists a continuous extension operator $\mathbb{T} : BV_{\mathcal{B}}(Q; \mathbb{R}^N) \rightarrow BV_{\mathcal{B}}(\mathbb{R}^N; \mathbb{R}^N)$ such that $\mathbb{T}u = u$ almost everywhere in Q for every $u \in BV_{\mathcal{B}}(Q; \mathbb{R}^N)$.*

Proof Since $\mathcal{N}(\mathcal{B})$ is finite dimensional, in view of Breit et al. (2017, (4.9) and Theorem 1.1) there exists a continuous trace operator $\text{tr} : BV_{\mathcal{B}}(Q; \mathbb{R}^N) \rightarrow L^1(\partial Q; \mathbb{R}^N)$. By the classical results by E. Gagliardo (see Gagliardo (1957)) there exists a linear and continuous extension operator $E : L^1(\partial Q; \mathbb{R}^N) \rightarrow W^{1,1}(\mathbb{R}^N \setminus Q; \mathbb{R}^N)$. The statement follows by setting

$$\mathbb{T}u := u\chi_Q + E(\text{tr}(u))\chi_{\mathbb{R}^N \setminus Q}, \tag{6.4}$$

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 Breit et al. (2017, Corollary 4.21). \square

Remark 6.4 We point out that, as a direct consequence of Lemma 6.3, we obtain

$$|\mathcal{B}(\mathbb{T}u)|_{\mathcal{M}_b(\mathbb{R}^N; \mathbb{R}^{N \times N})} \leq C_{\mathcal{B}} \|u\|_{BV_{\mathcal{B}}(Q; \mathbb{R}^N)}. \tag{6.5}$$

In particular, from (6.4) and Theorem Breit et al. (2017, Corollary 4.21), the constant $C_{\mathcal{B}}$ in the inequality above is obtained by the following estimate:

$$\begin{aligned} |\mathcal{B}(\mathbb{T}(u))|_{\mathcal{M}_b(\mathbb{R}^N; \mathbb{R}^{N \times N})} &\leq |\mathcal{B}u|_{\mathcal{M}_b(Q; \mathbb{R}^{N \times N})} + |\mathcal{B}(E(\text{tr}(u)))|_{\mathcal{M}_b(\mathbb{R}^N \setminus Q; \mathbb{R}^{N \times N})} \\ &\leq |\mathcal{B}u|_{\mathcal{M}_b(Q; \mathbb{R}^{N \times N})} + \|\mathcal{B}\|_{\ell^\infty} C_G \|\text{tr}(u)\|_{L^1(\partial Q; \mathbb{R}^N)} \leq (1 + \|\mathcal{B}\|_{\ell^\infty} C_G C_{T_{\mathcal{B}}}) \|u\|_{BV_{\mathcal{B}}(Q)}, \end{aligned}$$

where C_G is the constant associated to the classical Gagliardo’s extension in $W^{1,1}$ (see Gagliardo 1957) and is thus independent of \mathcal{B} , whereas $C_{T_{\mathcal{B}}}$ is the constant associated to the trace operator in $BV_{\mathcal{B}}$. Hence, $C_{\mathcal{B}} = (1 + \|\mathcal{B}\|_{\ell^\infty} C_G C_{T_{\mathcal{B}}})$.

The main result of this section is the following.

Theorem 6.5 *Let \mathcal{A} be as in Definition 6.1. Let Π and $\Pi_{\mathcal{A}}$ be the collections of first order operators introduced in Definitions 3.4 and 6.2, respectively. Then every operator $\mathcal{B} \in \Pi_{\mathcal{A}}$ satisfies Assumption 3.2. Additionally, every subset of operators in $\Pi_{\mathcal{A}}$ for which the constants in (6.5) are uniformly bounded fulfills Assumption 3.3.*

We proceed by first recalling two preliminary results from the literature. The next proposition, that may be found in Brezis (2011, Theorem 4.26), will be instrumental in the proof of a regularity result for distributions with bounded \mathcal{B} -total-variation (see Proposition 6.9).

Proposition 6.6 *Let \mathcal{F} be a bounded set in $L^p(\mathbb{R}^N)$ with $1 \leq p < +\infty$. Assume that*

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

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

We also recall some basic properties of the space $BV_{\mathcal{B}}(Q; \mathbb{R}^N)$ for $\mathcal{B} \in \Pi_{\mathcal{A}}$ [see Definition 3.1 and Breit et al. (2017, Section 2)].

Proposition 6.7 *Let $\mathcal{B} \in \Pi_{\mathcal{A}}$. Let U be an open set in \mathbb{R}^N . Then*

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

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

Before we establish Theorem 6.5, we prove a technical lemma.

Lemma 6.8 *Let $k \in \mathbb{N}$. Then there exists a constant $C > 0$ such that, for every $h \in \mathbb{R}^N$ and $w \in W_{\text{loc}}^{k,1}(\mathbb{R}^N; \mathbb{R}^N)$, there holds*

$$\begin{aligned} & \limsup_{|h| \rightarrow 0} \sum_{|a| \leq k} \left\| \tau_h \left(\frac{\partial^a}{\partial x^a} w \right) - \frac{\partial^a}{\partial x^a} w \right\|_{L^1(Q; \mathbb{R}^N)} \\ & \leq \limsup_{|h| \rightarrow 0} C \sum_{|a|=k} \left\| \tau_h \left(\frac{\partial^a}{\partial x^a} w \right) - \frac{\partial^a}{\partial x^a} w \right\|_{L^1(Q; \mathbb{R}^N)}, \end{aligned}$$

where τ_h is the operator defined in (6.3).

Proof By the linearity of τ_h , we have

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

On the one hand, by the Sobolev embedding theorem (see, e.g., Leoni 2009), we have

$$\begin{aligned} & \sum_{|a| \leq k} \left\| \tau_h \left(\frac{\partial^a}{\partial x^a} w \right) - \frac{\partial^a}{\partial x^a} w \right\|_{L^1(Q; \mathbb{R}^N)} \\ & = \sum_{|a| \leq k} \left\| \frac{\partial^a}{\partial x^a} (\tau_h w - w) \right\|_{L^1(Q; \mathbb{R}^N)} \\ & \leq C \|\tau_h(w) - w\|_{L^1(Q; \mathbb{R}^N)} + C \sum_{|a|=k} \left\| \frac{\partial^a}{\partial x^a} (\tau_h w - w) \right\|_{L^1(Q; \mathbb{R}^N)}. \end{aligned} \tag{6.6}$$

On the other hand, by the continuity of the translation operator in L^1 (see, e.g., Brezis (2011, Lemma 4.3) for a proof in \mathbb{R}^N , the analogous argument holds on bounded open sets) we have

$$\limsup_{|h| \rightarrow 0} \|\tau_h(w) - w\|_{L^1(Q; \mathbb{R}^N)} = 0. \tag{6.7}$$

The result follows by combining (6.6) and (6.7). □

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

Proposition 6.9 *Let $\mathcal{B} \in \Pi_{\mathcal{A}}$, and let $BV_{\mathcal{B}}(Q; \mathbb{R}^N)$ be the space introduced in Definition 3.1. Then the injection of $BV_{\mathcal{B}}(Q; \mathbb{R}^N)$ into $L^1(Q; \mathbb{R}^N)$ is compact.*

Proof For every $u \in BV_{\mathcal{B}}(Q; \mathbb{R}^N)$ we still denote by u its extension to $BV_{\mathcal{B}}(2Q; \mathbb{R}^N)$ provided by Lemma 6.3. In view of Proposition 6.7, for every $u \in BV_{\mathcal{B}}(Q; \mathbb{R}^N)$ we then find a sequence of maps $\{v_u^n\}_{n=1}^\infty \subset C^\infty(2Q; \mathbb{R}^N)$ such that

$$\|v_u^n - u\|_{L^1(2Q; \mathbb{R}^N)} + \left| \|\mathcal{B}v_u^n\|_{L^1(2Q; \mathbb{R}^N \times \mathbb{R}^N)} - |\mathcal{B}u|_{\mathcal{M}_b(2Q; \mathbb{R}^N \times \mathbb{R}^N)} \right| \leq \frac{1}{n}. \tag{6.8}$$

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. Fefferman 2007), where d is the order of the operator \mathcal{A} . Without loss of generality, up to a multiplication by a cut-off function, we can assume that $v_u^n \in C_c^d(3Q; \mathbb{R}^N)$ for every $n \in \mathbb{N}$.

We first show that, setting

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

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

$$\limsup_{|h| \rightarrow 0} \sup_{u \in \mathcal{F}} \left\{ \|\tau_h v_u^n - v_u^n\|_{L^1(2Q; \mathbb{R}^N)} \right\} = 0, \tag{6.9}$$

where we recall τ_h from Theorem 6.6, and where for fixed $u \in \mathcal{F}$, v_u^n is as above and satisfying (6.8).

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

$$\begin{aligned} \tau_h(\lambda_i v_{u,i}^n) &= \delta_h * \lambda_i v_{u,i}^n = \delta_h * (\lambda_i \delta * v_{u,i}^n) = \delta_h * ((\mathcal{A} P_\lambda)_i * v_{u,i}^n) \\ &= (\delta_h * (\mathcal{A} P_\lambda))_i * v_{u,i}^n = (\mathcal{A}(\delta_h * (P_\lambda)))_i * v_{u,i}^n, \end{aligned}$$

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

$$\begin{aligned} &\|\tau_h(\lambda_i v_{u,i}^n) - \lambda_i v_{u,i}^n\|_{L^1(2Q; \mathbb{R}^N)} \\ &= \|(\mathcal{A}(\delta_h * (P_\lambda)))_i * v_{u,i}^n - (\mathcal{A} P_\lambda)_i * v_{u,i}^n\|_{L^1(2Q; \mathbb{R}^N)} \\ &= \|(\mathcal{A}(\delta_h * (P_\lambda) - P_\lambda))_i * v_{u,i}^n\|_{L^1(2Q; \mathbb{R}^N)} \end{aligned}$$

$$\leq C_{\mathcal{A}} \left[\sum_{|a| \leq d-1} \left\| \tau_h \left(\frac{\partial^a}{\partial x^a} P_\lambda \right) - \frac{\partial^a}{\partial x^a} P_\lambda \right\|_{L^1(2Q; \mathbb{R}^N)} \right] |\mathcal{B}v_u^n|_{\mathcal{M}_b(2Q; \mathbb{R}^{N \times N})} \quad (6.10)$$

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, the identity $\tau_h \left(\frac{\partial^a}{\partial x^a} P_\lambda \right) = \frac{\partial^a}{\partial x^a} (\tau_h P_\lambda)$, as well as Definition 6.1, Assertion 2.

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

$$\begin{aligned} & \sup_{u \in \mathcal{F}} \left\{ \|\tau_h(v_u^n) - v_u^n\|_{L^1(2Q; \mathbb{R}^N)} \right\} \\ & \leq C_{\mathcal{A}} \left(1 + \frac{1}{n} \right) \sum_{|a| \leq d-1} \left\| \tau_h \left(\frac{\partial^a}{\partial x^a} P_{\bar{\lambda}} \right) - \frac{\partial^a}{\partial x^a} P_{\bar{\lambda}} \right\|_{L^1(2Q; \mathbb{R}^N)}, \end{aligned}$$

and, in view of (6.2) and Lemma 6.8, we conclude that

$$\begin{aligned} & \lim_{|h| \rightarrow 0} \sup_{u \in \mathcal{F}} \left\{ \|\tau_h(v_u^n) - v_u^n\|_{L^1(2Q; \mathbb{R}^N)} \right\} \\ & \leq C_{\mathcal{A}} \left(1 + \frac{1}{n} \right) \lim_{|h| \rightarrow 0} \sum_{|a| = d-1} \left\| \tau_h \left(\frac{\partial^a}{\partial x^a} P_{\bar{\lambda}} \right) - \frac{\partial^a}{\partial x^a} P_{\bar{\lambda}} \right\|_{L^1(2Q; \mathbb{R}^N)} = 0 \end{aligned}$$

for every $n \in \mathbb{N}$, which yields (6.9).

By (6.8), 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

$$\begin{aligned} \|\tau_h u - u\|_{L^1(\frac{3}{2}Q; \mathbb{R}^N)} & \leq \|\tau_h u - \tau_h v_u^n\|_{L^1(\frac{3}{2}Q; \mathbb{R}^N)} \\ & \quad + \|\tau_h v_u^n - v_u^n\|_{L^1(\frac{3}{2}Q; \mathbb{R}^N)} + \|v_u^n - u\|_{L^1(\frac{3}{2}Q; \mathbb{R}^N)} \\ & \leq 2 \|v_u^n - u\|_{L^1(2Q; \mathbb{R}^N)} + \|\tau_h v_u^n - v_u^n\|_{L^1(2Q; \mathbb{R}^N)} \end{aligned}$$

The thesis follows then by (6.8), (6.9), and Proposition 6.6. □

We close this subsection by proving a compactness and lower-semicontinuity result for functions with uniformly bounded $BV_{\mathcal{B}_n}$ norms. We recall that the definition of $M_{\mathcal{A}}$ is found in (6.2).

Proposition 6.10 *Let $\{\mathcal{B}_n\}_{n=1}^\infty \subset \Pi_{\mathcal{A}}$ be such that $\mathcal{B}_n \rightarrow \mathcal{B}$ in ℓ^∞ and the constants $C_{\mathcal{B}_n}$ in (6.5) are uniformly bounded. For every $n \in \mathbb{N}$ let $v_n \in BV_{\mathcal{B}_n}(Q; \mathbb{R}^N)$ be such that*

$$\sup \left\{ \|v_n\|_{BV_{\mathcal{B}_n}(Q; \mathbb{R}^N)} : n \in \mathbb{N} \right\} < +\infty. \quad (6.11)$$

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

$$v_n \rightarrow v \text{ strongly in } L^1(Q; \mathbb{R}^N), \quad (6.12)$$

and

$$\mathcal{B}_n v_n \overset{*}{\rightharpoonup} \mathcal{B}v \text{ weakly* in } \mathcal{M}_b(Q; \mathbb{R}^{N \times N}). \tag{6.13}$$

Proof Let v_n satisfy (6.11). With a slight abuse of notation we still indicate by v_n the $BV_{\mathcal{B}_n}$ continuous extension of the above maps to \mathbb{R}^N (see Lemma 6.3). 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 $\text{supp } \tilde{v}_n \subset\subset 2Q$. Additionally, by Lemma 6.3 there holds

$$\begin{aligned} \|\tilde{v}_n\|_{BV_{\mathcal{B}_n}(2Q; \mathbb{R}^N)} &\leq \|v_n\|_{L^1(2Q; \mathbb{R}^N)} + |\mathcal{B}_n v_n|_{\mathcal{M}_b(2Q; \mathbb{R}^{N \times N})} \\ &\quad + \left\| \sum_{i=1}^N B_n^i \frac{\partial \phi}{\partial x_i} \right\|_{L^\infty(2Q; \mathbb{M}^{N^3})} \|v_n\|_{L^1(2Q; \mathbb{R}^N)} \\ &\leq C_1 \|v_n\|_{BV_{\mathcal{B}_n}(2Q; \mathbb{R}^N)} \leq C_2 \|v_n\|_{BV_{\mathcal{B}_n}(Q; \mathbb{R}^N)}, \end{aligned} \tag{6.14}$$

where in the last inequality we used Lemma 6.3, and where the constants C_1 and C_2 depend only on the cut-off function ϕ . To prove (6.12) we first show that

$$\limsup_{|h| \rightarrow 0} \sup_{n \in \mathbb{N}} \{ \|\tau_h \tilde{v}_n - \tilde{v}_n\|_{L^1(\mathbb{R}^N; \mathbb{R}^N)} \} = 0, \tag{6.15}$$

where we recall τ_h from Theorem 6.6. Arguing as in the proof of (6.10), by (6.14) we deduce that for $|h|$ small enough, since $\text{supp } \mathcal{E} \subset\subset 2Q$,

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

for every $n \in \mathbb{N}$. Property (6.15) follows by (6.2). Owing to Proposition 6.6, we deduce (6.12).

We now prove (6.13). Let $\varphi \in C_c^\infty(Q; \mathbb{R}^{N \times N})$ be such that $|\varphi| \leq 1$. Then

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

$$= - \sum_{i,j,k,l=1}^N \int_Q v_l (B)_{ijl}^k \frac{\partial \varphi_{ij}}{\partial x_k} dx$$

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

This completes the proof of (6.13) and of the proposition. □

Proof of Theorem 6.5 Let $\mathcal{B} \in \Pi_{\mathcal{A}}$ be given. The fact that \mathcal{B} satisfies Assumption 3.2 follows by Propositions 6.7 and 6.9. The fulfillment of Assumption 3.3 is a direct consequence of Proposition 6.10. □

6.2 Training Scheme with Fixed and Multiple Operators \mathcal{A}

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

Definition 6.11 Let \mathcal{A} be a differential operator of order $d \in \mathbb{N}$. We denote by $\hat{\Sigma}[\mathcal{A}]$ the set

$$\hat{\Sigma}[\mathcal{A}] := \{ \mathcal{B} \in \Pi_{\mathcal{A}} : \|\mathcal{B}\|_{\ell^\infty} \leq 1 \}.$$

The first result of this subsection is the following.

Theorem 6.12 *Let \mathcal{A} be a differential operator of order $d \in \mathbb{N}$, and assume that $\Sigma[\mathcal{A}]$ is a non-empty subset of $\hat{\Sigma}[\mathcal{A}]$ which is closed in the ℓ^∞ convergence with respect of the property of having finite-dimensional null space. Then the collection $\Sigma[\mathcal{A}]$ is a training set (see Definition 5.1).*

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

The fact that $\mathcal{N}(\mathcal{B})$ is finite-dimensional follows by definition. To conclude the proof of the theorem we still need to show that $(\mathcal{A}, \mathcal{B})$ 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 $v \in BV_{\mathcal{B}}(U; \mathbb{R}^N)$. By Proposition 6.7 there exists $\{v_k\}_{k=1}^\infty \subset C^\infty(U; \mathbb{R}^N)$ such that

$$v_k \rightarrow v \text{ strongly in } L^1(U; \mathbb{R}^N) \text{ and } |\mathcal{B}v_k|_{\mathcal{M}_b(U; \mathbb{R}^{N \times N})} \rightarrow |\mathcal{B}v|_{\mathcal{M}_b(U; \mathbb{R}^{N \times N})}. \tag{6.16}$$

Integrating by parts we obtain

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

for every $i = 1, \dots, N$. Taking the limit as $n \rightarrow \infty$ first, and then as $k \rightarrow \infty$, since $\mathcal{B}_n \rightarrow \mathcal{B}$ in ℓ^∞ and in view of (6.16), we conclude that

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

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

Remark 6.13 We note that the result of Theorem 6.12 still holds if we replace the upper bound 1 in Definition 6.11 with an arbitrary positive constant.

We additionally point out that requiring that the finite-dimensional-kernel property is preserved in the limit passage automatically ensures the existence of a lower bound on the ℓ^∞ -norms of the operators. In other words, the null operator is not included in our analysis.

As a final remark, we stress that, if $\hat{\Sigma}[\mathcal{A}]$ contains an operator $\bar{\mathcal{B}}$ with finite-dimensional null space, then a training set $\Sigma[\mathcal{A}] \subset \hat{\Sigma}[\mathcal{A}]$ being closed in the ℓ^∞ -norm with respect to the property of having finite-dimensional null space can be constructed by taking the intersection of $\hat{\Sigma}[\mathcal{A}]$ with a small enough neighborhood of $\bar{\mathcal{B}}$ in the ℓ^∞ -topology.

In fact, denoting by $B^i \in \mathbb{M}^{N^3}$, $i = 1, \dots, N$, the coefficients of $\bar{\mathcal{B}}$, the symbol of $\bar{\mathcal{B}}$ is defined as

$$\mathbb{B}[\xi] := \sum_{i=1}^N \xi_i B^i \quad \text{for every } \xi = (\xi_1, \dots, \xi_N) \in \mathbb{C}^N \setminus \{0\}.$$

The condition of having finite-dimensional null space is equivalent to the so-called \mathbb{C} -ellipticity condition, which consists in the injectivity of the map $\mathbb{B}[\xi]$ as a linear map on $\mathbb{C}^N \setminus \{0\}$ for every $\xi \in \mathbb{C}^N \setminus \{0\}$ [see Breit et al. (2017, Section 2.3)]. By linearity, this, in turn, can be reduced to the condition of the map $\mathbb{B}[\xi]$ being injective on $\mathbb{C}^N \setminus \{0\}$ for every ξ in $B_{\mathbb{C}}(0, 1) \setminus \{0\}$, where $B_{\mathbb{C}}(0, 1)$ is the unit ball centered in the origin in the complex plane. In particular, it is a stable condition with respect to small ℓ^∞ -perturbations of the coefficients.

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

Definition 6.14 We say that collection \mathcal{A} of differential operators \mathcal{A} is a training set builder if

$$\sup \{C_{\mathcal{A}} : \mathcal{A} \in \mathcal{A}\} < +\infty \text{ and } \lim_{|h| \rightarrow 0} \sup \{M_{\mathcal{A}}(h) : \mathcal{A} \in \mathcal{A}\} = 0, \quad (6.17)$$

where $C_{\mathcal{A}}$ and $M_{\mathcal{A}}(h)$ are defined in (6.1) and (7.5), respectively.

We then define the class $\Sigma[\mathcal{A}]$ via

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

where for every $\mathcal{A} \in \mathcal{A}$, $\Sigma[\mathcal{A}]$ is the class defined in Definition 6.11.

We close this section by proving the following theorem.

Theorem 6.15 *Let \mathcal{A} be a training set builder. Then $\Sigma[\mathcal{A}]$ is a training set.*

Proof The proof of this theorem follows the argument in the proof of Theorem 6.12 using the fact that the two critical constants $M_{\mathcal{A}}(h)$ and $C_{\mathcal{A}}$, in (6.2) and (6.1), respectively, are uniformly bounded due to (6.17). □

7 Explicit Examples and Numerical Observations

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

7.1 The Existence of Fundamental Solutions of Operators \mathcal{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 \mathcal{A} . A result in this direction can be found in Hsiao and Wendland (2008, p. 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 $N = 2$, \mathcal{A} has order 2 and satisfies the assumptions in Hsiao and Wendland (2008, p. 351, Section 6.3), the fundamental solution P_λ can be written as

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

where L denotes the fundamental solution of Laplace’s equation, $R_{\mathcal{A}}$ denotes a constant depending on \mathcal{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

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

the fundamental solution P_α , with $\mathcal{A}P_\alpha = \alpha\delta$ 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 $R_{\mathcal{A}}$ in (7.1) is a constant, ∇P_λ must have the same homogeneity as ∇P_α , which is $1 - N$.

Proposition 7.2 *Let \mathcal{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 property (6.2) 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 Temam (1983, Lemma 1.4) we deduce the estimate

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

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 $\mathcal{A} P_\lambda = \lambda \delta$, 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

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

and (6.2) is established. □

Remark 7.3 As a corollary of Proposition 7.2 and Remark 7.1, we deduce that all operators \mathcal{A} satisfying the assumptions in Hsiao and Wendland (2008, p. 351, Section 6.3) comply with Definition 6.1, Assertion 1. In particular, differential operators \mathcal{A} which can be written in the form $\mathcal{A} = \mathcal{B}^* \circ \mathcal{C}$, where \mathcal{B}^* is the first order differential operator associated to \mathcal{B} and having as coefficients the transpose of the matrices B^i , $i = 1, \dots, N$, and where \mathcal{C} is a differential operator of order $d - 1$ having constant coefficients, are such that $(\mathcal{A}, \mathcal{B})$ complies with Definition 6.1.

7.2 The Unified Approach to TGV^2 and $NsTGV^2$: An Example of $\Sigma[\mathcal{A}]$

In this section we give an explicit construction of an operator \mathcal{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 $\mathcal{B} \in \Sigma[\mathcal{A}]$.

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

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

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

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

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

$$\begin{aligned} (\mathcal{A}w) * v &= \sum_{j=1}^N [\Delta w_j + \partial_j \text{div}(w)] * v_j = \sum_{i,j=1}^N (\partial_i w_j + \partial_j w_i) * \partial_i v_j \\ &= \frac{1}{2} \sum_{i,j=1}^N (\partial_i w_j + \partial_j w_i) * (\partial_i v_j + \partial_j v_i) = \frac{1}{2} (\mathcal{B}_{\text{sym}}w) * (\mathcal{B}_{\text{sym}}v), \end{aligned} \tag{7.7}$$

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

$$\begin{aligned} |(\mathcal{A}w) * v|_{\mathcal{M}_b(U; \mathbb{R}^2)} &\leq \frac{1}{2} |(\mathcal{B}_{\text{sym}}w) * (\mathcal{B}_{\text{sym}}v)|_{\mathcal{M}_b(U; \mathbb{M}^{2 \times 2})} \\ &\leq \frac{1}{2} \|\nabla w\|_{L^1(U; \mathbb{M}^{2 \times 2})} |\mathcal{B}_{\text{sym}}(v)|_{\mathcal{M}_b(U; \mathbb{M}^{2 \times 2})}. \end{aligned}$$

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

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

$$B_t := \begin{bmatrix} 1 & 0 \\ 0 & (1-t) \end{bmatrix} \begin{matrix} t & 0 \\ 0 & 0 \end{matrix} \quad \text{and} \quad B_s := \begin{bmatrix} 1 & 0 \\ 0 & 1-s \end{bmatrix} \begin{matrix} s & 0 \\ 0 & 0 \end{matrix},$$

and we set

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

By a straightforward computation, we obtain that $\mathcal{N}(\mathcal{B}_{s,t})$ is finite dimensional for every $0 \leq s, t \leq 1$. 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

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

which implies that

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

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

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

is a training set according to Definition 6.11. We remark that $\Sigma[\mathcal{A}]$ 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 (T_θ^2) with training set $\Sigma[\mathcal{A}]$ 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 Brinkmann et al. (2019) the authors analyze a range of first order linear operators generated by diagonal matrixes. To be precise, letting $B = \text{diag}(\beta_1, \beta_2, \beta_3, \beta_4)$,

Brinkmann et al. (2019) treats first order operators \mathcal{B} defined as

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

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 Brinkmann et al. (2019) ∇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

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

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

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

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

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

That is, we would have

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

which are not diagonal matrices. Hence, this example shows that our model indeed covers more operators than those discussed in Brinkmann et al. (2019).

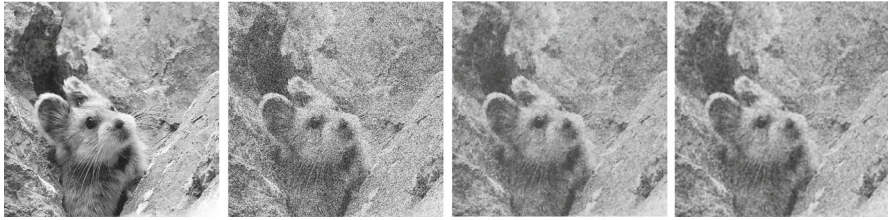


Fig. 1 From left to right: the test image of a Pika; a noised version (with heavy artificial Gaussian noise); the optimally reconstructed image with *TGV* regularizer; the optimally reconstructed image with *PGV* regularizer

7.3 Numerical Simulations and Observations

Let \mathcal{A} be the operator defined in Sect. 7.2, and let

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

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

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

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

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

We perform numerical simulations of the images shown in Fig. 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 $u_{\alpha, \mathcal{B}}$ in Level 2 of our training scheme is computed by using the primal-dual algorithm presented in Chambolle and Pock (2011).

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 $u_{\tilde{\alpha}, \mathcal{B}_{\tilde{s}, \tilde{t}}}$ is the last image in Fig. 1, whereas the optimal result with $\mathcal{B}_{s,t} \equiv \mathcal{E}$, i.e., $u_{\tilde{\alpha}, TGV}$, is the third image in Fig. 1. Although the optimal reconstructed image $u_{\tilde{\alpha}, \mathcal{B}_{\tilde{s}, \tilde{t}}}$ and $u_{\tilde{\alpha}, \mathcal{E}}$ do not present too many differences to the naked eye, we do have that

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

Table 1 Minimum cost value with different regularizers

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

The minimum value of the cost function for the PGV^2 -regularizer is approximately 5% below that of the TGV^2 -regularizer

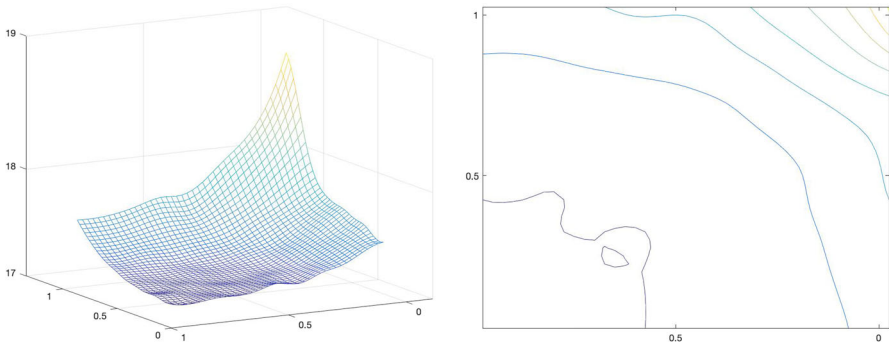


Fig. 2 From the left to the right: mesh and contour plot of the cost function $C(\bar{\alpha}, s, t)$ in which $\bar{\alpha} = (\bar{\alpha}_0, \bar{\alpha}_1)$ is fixed, $(s, t) \in [0, 1]^2$

(see also Table 1 below). That is, the reconstructed image $u_{\tilde{\alpha}, \mathcal{B}_{\tilde{s}, \tilde{t}}}$ 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 $\tilde{\alpha}_0 = 0.072$ and $\tilde{\alpha}_1 = 0.575$ and plot in Fig. 2 the mesh and contour plot of $C(\bar{\alpha}, s, t)$.

We again remark that the introduction of $PGV_{\alpha, \mathcal{B}[k]}$ regularizers into the training scheme is only meant to expand the training choices, but not to provide a superior seminorm with respect to the 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 $C(\alpha, s, t)$. We consider also in this second example the discrete box-constraint in (7.10), and we analyze the images shown in Fig. 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 $u_{\alpha, \mathcal{B}}$ in Level 2 of our training scheme is again computed by using the primal-dual algorithm presented in Chambolle and Pock (2011).

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 $u_{\tilde{\alpha}, \mathcal{B}_{\tilde{s}, \tilde{t}}}$ is the last image in Fig. 3, whereas the optimal result with $\mathcal{B}_{s, t} \equiv \mathcal{E}$, i.e., $u_{\tilde{\alpha}, TGV}$, is the third image in Fig. 3. Although the optimal reconstructed image $u_{\tilde{\alpha}, \mathcal{B}_{\tilde{s}, \tilde{t}}}$ and $u_{\tilde{\alpha}, TGV}$ do not present too many differences with respect to our eyesight, we do have, also in

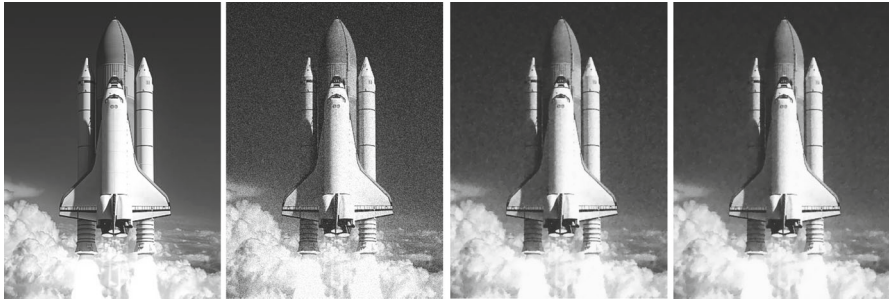


Fig. 3 From left to right: the test image of a space shuttle; a noised version (with artificial Gaussian noise); the optimally reconstructed image $u_{\tilde{\alpha}, \mathcal{E}}$, where $\tilde{\alpha}_0 = 5.2$, $\tilde{\alpha}_1 = 1.9$; the optimally reconstructed image $u_{\tilde{\alpha}, \mathcal{B}_{\tilde{s},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

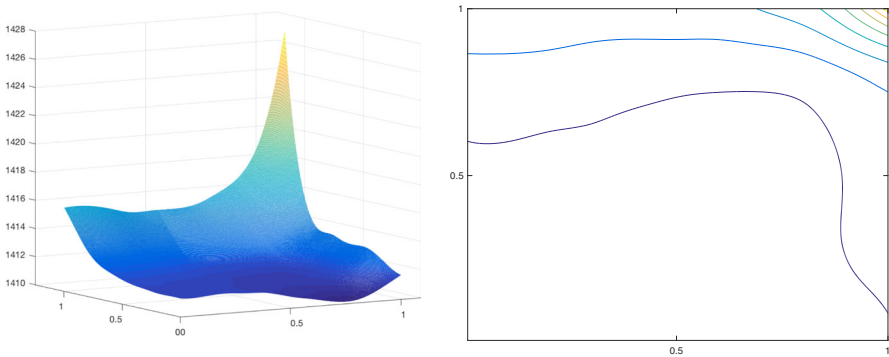


Fig. 4 From left to right: mesh and contour plot of the cost function $C(\bar{\alpha}, s, t)$ in which $\bar{\alpha} = (\bar{\alpha}_0, \bar{\alpha}_1)$ is fixed, $(s, t) \in [0, 1]^2$. The function $C(\bar{\alpha}, s, t)$ achieves the minimum at $\tilde{s} = 0.8$ and $\tilde{t} = 0.2$

this case, that

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

Namely, the reconstructed image $u_{\tilde{\alpha}, \mathcal{B}_{\tilde{s}, \tilde{t}}}$ 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 $\bar{\alpha}_0 = 5.6$ and $\bar{\alpha}_1 = 1.9$ and plot in Fig. 4 the mesh and contour plot of $C(\bar{\alpha}, s, t)$ as follows.

8 Conclusions

We have introduced a novel class of regularizers providing a generalization of TGV^2 to the case in which the higher-order operators can be different from the symmetric gradient. After establishing basic properties of this class of functionals, we have studied well-posedness of a bilevel learning scheme selecting the optimal regularizer in our class in terms of a quadratic cost function. Eventually, we have shown some very

first numerical simulations of our scheme. We point out that both examples in Figs. 1 and 3 do not present a clear distinction to the naked eye with respect to their TGV counterpart although performing much better in terms of the cost-function landscapes. We conjecture this behavior not to be the general case. Further numerical investigations are beyond the scope of this paper and will be the subject of forthcoming works.

Acknowledgements The work of Elisa Davoli has been funded by the Austrian Science Fund (FWF) projects F65, V 662, Y 1292, 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".

Funding Open access funding provided by Austrian Science Fund (FWF).

Data Availability Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Declarations

Conflict of interest The Authors confirm hereby that the only funding sources for this research are those summarized in the Acknowledgements section above and that they have no conflicts of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ambrosio, L., Fusco, N., Pallara, D.: Functions of Bounded Variation and Free Discontinuity Problems. Oxford Mathematical Monographs. The Clarendon Press/Oxford University Press, New York (2000)
- Barbu, T., Marinoschi, G.: Image denoising by a nonlinear control technique. *Int. J. Control* **90**, 1005–1017 (2017)
- Bredies, K., Holler, M.: Regularization of linear inverse problems with total generalized variation. *J. Inverse Ill Posed Probl.* **22**, 871–913 (1993)
- Bredies, K., Kunisch, K., Pock, T.: Total generalized variation. *SIAM J. Imaging Sci.* **3**, 492–526 (2010)
- Bredies, K., Valkonen, T.: Inverse problems with second-order total generalized variation constraints. In: Proceedings of SampTA 2011—9th International Conference on Sampling Theory and Applications, Singapore (2011)
- Breit, D., Diening, L., Gmeineder, F.: Traces of functions of bounded A-variation and variational problems with linear growth. Preprint [arXiv:1707.06804](https://arxiv.org/abs/1707.06804)
- Brezis, H.: Functional analysis, Sobolev spaces and partial differential equations. Universitext. Springer, New York (2011)
- Brinkmann, E.-M., Burger, M., Grah, J.S.: Unified models for second-Order TV-type regularisation in imaging—a new perspective based on vector operators. *J. Math. Imaging Vis.* **61**, 571–601 (2019)

- Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.* **40**, 120–145 (2011)
- Chen, Y., Pock, T., Ranftl, R., Bischof, H.: Revisiting loss-specific training of filter-based MRFs for image restoration. In: *Pattern Recognition*, pp. 271–281. Springer (2013)
- Chen, Y., Ranftl, R., Pock, T.: Insights into analysis operator learning: from patch-based sparse models to higher order MRFs. *IEEE Trans. Image Process.* **23**, 1060–1072 (2014)
- Dal Maso, G.: An introduction to Γ -convergence. Birkhäuser, Boston (1993)
- Davoli, E., Liu, P.: One dimensional fractional order TGV : Gamma-convergence and bilevel training scheme. *Commun. Math. Sci.* **16**, 213–237 (2018)
- De Los Reyes, J.C., Schönlieb, C.-B., Valkonen, T.: The structure of optimal parameters for image restoration problems. *J. Math. Anal. Appl.* **434**, 464–500 (2016)
- De los Reyes, J.C., Schönlieb, C.-B., Valkonen, T.: Bilevel parameter learning for higher-order total variation regularisation models. *J. Math. Imaging Vis.* **57**, 1–25 (2017)
- Domke, J.: Generic methods for optimization-based modeling. *AISTATS* **22**, 318–326 (2012)
- Fefferman, C.: C^m extension by linear operators. *Ann. Math. (2)* **166**, 779–835 (2007)
- Fonseca, I., Liu, P.: The weighted Ambrosio–Tortorelli approximation scheme. *SIAM J. Math. Anal.* **49**, 4491–4520 (2017)
- Gagliardo, E.: Caratterizzazioni delle tracce sulla frontiera relative ad alcune classi di funzioni in n variabili. *Rend. Sem. Mat. Univ. Padova* **27**, 284–305 (1957)
- Gmeineder, F., Raiță, B.: Embeddings for \mathbb{A} -weakly differentiable functions on domains. *J. Funct. Anal.* **277**, 108278 (2019)
- Gmeineder, F., Raiță, B., Van Schaftingen, J.: On limiting trace inequalities for vectorial differential operators. *Indiana Univ. Math. J.* **70**, 2133–2176 (2021)
- Guerra, A., Raiță, B.: On the necessity of the constant rank condition for L^p estimates. *Comptes Rendus. Mathématique*, **358**(9–10), 1091–1095 (2020)
- Guerra, A., Raiță, B., Schrecker, M.R.I.: Compensated compactness: continuity in optimal weak topologies. *J. Funct. Anal.* **283**, 109596 (2022). Preprint [arXiv:2007.00564](https://arxiv.org/abs/2007.00564)
- Hsiao, G.C., Wendland, W.L.: *Boundary Integral Equations*. Springer, New York (2008)
- Kristensen, J., Raiță, B.: Oscillation and concentration in sequences of PDE constrained measures. *Arch. Rational Mech. Anal.* **246**, 823–875 (2022). Preprint [arXiv:1912.09190](https://arxiv.org/abs/1912.09190)
- Leoni, G.: A first course in Sobolev spaces. In: *Graduate Studies in Mathematics*, vol. 105. American Mathematical Society, Providence (2009)
- Raiță, B.: L^1 -estimates for constant rank operators (2018). Preprint [arXiv:1811.10057](https://arxiv.org/abs/1811.10057)
- Raiță, B.: Potentials for \mathcal{A} -quasiconvexity. *Calc. Var. PDEs* **58**, (2019). article 105
- Raiță, B.: Critical L^p -differentiability of $BV^{\mathbb{A}}$ -maps and canceling operators *Trans. Am. Math. Soc.* **372**, 7297–7326 (2019)
- Raiță, B., Skorobogatova, A.: Continuity and canceling operators of order n on \mathbb{R}^n . *Calc. Var. PDEs* **59**, (2020). article 85
- Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Phys. D* **60**, 259–268 (1992)
- Tappe, M.F., Liu, C., Adelson, E.H., Freeman, W.T.: Learning gaussian conditional random fields for low-level vision. In: *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8 (June 2007)
- Temam, R.: *Problèmes mathématiques en plasticité. Méthodes Mathématiques de l'Informatique [Mathematical Methods of Information Science]*, vol. 12. Gauthier-Villars, Montrouge (1983)
- Valkonen, T.: The jump set under geometric regularisation. Part 2: higher-order approaches. *J. Math. Anal. Appl.* **453**, 1044–1085 (2017)
- Valkonen, T., Bredies, K., Knoll, F.: Total generalized variation in diffusion tensor imaging. *SIAM J. Imaging Sci.* **6**, 487–525 (2013)