COMMUN. MATH. SCI.
Vol. 20, No. 4, pp. 1117–1149

ADAPTIVE IMAGE PROCESSING: A BILEVEL STRUCTURE LEARNING APPROACH FOR MIXED-ORDER TOTAL VARIATION REGULARIZERS*

PAN LIU†, XIN Y. LU‡, AND WAN L. ZHU§

Abstract. We propose a class of mixed-order *PDE*-constraint regularizer for image processing problem, generalizing the standard first-order total variation (TV). Then, we study the corresponding semi-supervised (bilevel) training scheme, which provides a simultaneous optimization with respect to parameters and new class of regularizers. Finally, by relying on the finite approximation method, we solve the global optimization problem on such training scheme, and analyze the resulting numerical results.

Keywords. Image processing; optimal training scheme; higher order differential operators; Γ-convergence.

AMS subject classifications. 26B30; 94A08; 47J20.

1. Introduction

The use of variational techniques with non-smooth regularizers in image processing has become popular in the recent decades. One of the most successful approaches is introduced in the celebrated work [21], which relies on the so-called *ROF* total-variational functional

$$\mathcal{I}(u) := \|u - u_\eta\|^2_{L^2(Q)} + \alpha TV(u), \tag{1.1}$$

named after the authors Rudin, Osher, Fatemi. Here $u_\eta \in L^2(Q)$ is a given corrupted image, the unit square $Q := (0,1)^2$ is our domain, $\alpha \in \mathbb{R}^+$ is an *intensity parameter*, and $TV(u)$ stands for the total variation of u in Q (see [14]). For brevity, in this paper we will always use the notation $L^2(Q)$ to denote square integrable functions on Q, independent of their co-domain. In the simple case that $u \in W^{1,1}(Q)$, we have

$$TV(u) = \int_Q |\nabla u|\, dx = \int_Q \left(\left|\partial_1^1 u(x)\right|^2 + \left|\partial_2^1 u(x)\right|^2 \right)^{1/2} dx. \tag{1.2}$$

One advantage of using the TV regularization is it promotes piecewise constant reconstructions, thus preserving edges. However, this also leads to blocky-like artifacts in the reconstructed image, an effect known as stair-casing. To mitigate this effect, and also to explore possible improvements, the following methods have been introduced and studied:

(1) using higher-order extensions [3, 9];
(2) changing the underlying Euclidean norm [23];
(3) introducing fractional order derivatives [10, 20].

*Received: April 27, 2021; Accepted (in revised form): October 30, 2021. Communicated by Wotao Yin.
†China National Clinical Research Center for Neurological Diseases, Beijing Tiantan Hospital, Capital Medical University, Beijing, 100070, P.R. China (dragonrider.liupan@gmail.com).
‡Department of Mathematical Sciences, Lakehead University, 955 Oliver Road, Thunder Bay, ON, Canada (xlu8@lakeheadu.ca).
§China National Clinical Research Center for Neurological Diseases, Beijing Tiantan Hospital, Capital Medical University, Beijing, 100070, P.R. China (wanlin.zhu@163.com).

These methods introduce collections of regularizers which generalize the TV seminorm. For example, in [23], the underlying Euclidean norm is generalized from $p=2$, used in (1.2), to $p\in[1,+\infty]$ by defining

$$TV_p(u)=\int_Q |\nabla u|_p\,dx=\int_Q \left(\left|\partial_1^1 u(x)\right|^p+\left|\partial_2^1 u(x)\right|^p\right)^{1/p} dx.$$

In [20], the order of the derivative is generalized from $r=1$, used in (1.2), to $r\in\mathbb{R}^+$, by defining

$$TV^r(u)=\int_Q |\nabla^r u|\,dx=\int_Q \left(|\partial_1^r u(x)|^2+|\partial_2^r u(x)|^2\right)^{1/2} dx,$$

in which the fractional order derivative is realized by using the *Riemann-Liouville* fractional order derivative (see [22] for definition). In both works [10, 20], it has been shown that for given corrupted image u_η, a carefully selected *regularizer* parameter $p\in[1,+\infty]$ (resp. $r\in\mathbb{R}^+$) allows TV_p (resp. TV^r) to provide improved results, and such selection can be done automatically by using a bi-level training scheme, which will be detailed below.

In general, with a reliable selection mechanism, the image processing results would certainly be improved if we could further expand the collection of regularizers. To this purpose, in this paper we introduce a family of novel TV-like PDE-constraint regularizer (seminorm), say $PV_{\mathscr{B}}$, given by

$$PV_{\mathscr{B}}(u):=|\mathscr{B}u|_{\mathcal{M}_b(Q;\mathbb{R}^K)}. \tag{1.3}$$

Here $\mathcal{M}_b$ denotes the space of finite Radon measures, $|\cdot|_{\mathcal{M}_b}$ denotes the *Radon* norm of a measure, and $\mathscr{B}$: $L^1(Q)\to\mathcal{D}'(Q,\mathbb{R}^K)$ is a linear differential operator (see Notation 2.1). In the simple case $\mathscr{B}=\nabla$, we recover the total variation TV seminorm (see also [16, 17]). We remark that the abstract framework studied in (1.3) naturally incorporates the recent PDE-based approach to image denoising problems formulated in [1], and also allows us to simultaneously describe a variety of different image-processing techniques. The greater generality of the $PV_{\mathscr{B}}$ seminorm, compared to the classic TV seminorm, is beneficial in allowing for more a general functional, and hence better processing results, but, on the other hand, such increased generality is also the main issue in our analysis.

The aim of this paper is threefold. First, we provide a rigorous and detailed analysis of the properties of the $PV_{\mathscr{B}}$ seminorm, such as approximation by smooth functions, lower semi-continuity with respect to both the function u and the operator $\mathscr{B}$, and a point-wise characterization of the sub-gradient of $PV_{\mathscr{B}}$.

The second result is the analysis of the aforementioned selection mechanism, realized by a semi-supervised (bilevel) training scheme from machine learning (see [7, 8, 11, 12, 19, 24]). For example, we could apply the bilevel training scheme to determine the optimal value of $\alpha\in\mathbb{R}^+$ from (1.1), which controls the strength of the regularizer. More precisely, we assume that the corrupted image u_η can be decomposed as $u_\eta=u_c+\eta$ where $u_c\in L^2(Q)$ represents a noise-free clean image (the perfect data), η encodes the noise, and we refer to (u_η,u_c) as *training set*. Then, a bilevel training scheme, say Scheme $\mathcal{B}$, for determining the optimal intensity parameter α, can be formulated as follows:

$$\text{Level 1.}\qquad \alpha_{\mathbb{T}}\in\mathbb{A}[\mathbb{T}]:=\operatorname{arg\,min}\left\{\|u_\alpha-u_c\|^2_{L^2}:\alpha\in\mathbb{T}\right\}, \tag{$\mathcal{B}$-L1}$$

Level 2. $$u_\alpha := \arg\min\left\{\|u-u_\eta\|^2_{L^2(Q)} + \alpha TV_2(u) : u \in BV(Q)\right\}, \qquad (\mathcal{B}\text{-L2})$$

where $\mathbb{T} := [0,+\infty]$, used in ($\mathcal{B}$-L1), is called the *training ground*. Roughly speaking, Level 1 problem in ($\mathcal{B}$-L1) looks for an α that minimizes the L^2-distance to the clean image u_c, subject to the minimizing problem ($\mathcal{B}$-L2). That is, scheme $\mathcal{B}$ is able to optimally adapt itself to the given perfect data u_c.

In the same spirit, in order to identify the optimal operator $\mathscr{B}$ in $PV_{\mathscr{B}}$ for a given training set (u_η, u_c), we introduce the scheme $\mathcal{T}$ (($\mathcal{T}$-L1)-($\mathcal{T}$-L2)) defined as

Level 1. $$(\alpha_{\mathbb{T}}, \mathscr{B}_{\mathbb{T}}) \in \mathbb{A}[\mathbb{T}] := \arg\min\left\{\|u_c - u_{\alpha,\mathscr{B}}\|^2_{L^2(Q)} : (\alpha,\mathscr{B}) \in \mathbb{T}\right\}, \qquad (\mathcal{T}\text{-L1})$$

Level 2. $$u_{\alpha,\mathscr{B}} := \arg\min\left\{\|u-u_\eta\|^2_{L^2(Q)} + \alpha PV_{\mathscr{B}}(u) : u \in L^1(Q)\right\}. \qquad (\mathcal{T}\text{-L2})$$

In ($\mathcal{T}$-L1), we expand the training ground to $\mathbb{T} := \mathrm{cl}(\mathbb{R}^+) \times \Sigma$ to incorporate the new parameter $\mathscr{B} \in \Sigma$, where Σ denotes a closed collection of operators $\mathscr{B}$ (see Notation 2.1, Notation 4.1, and (4.1) for details). We remark that the expanded training ground $\mathbb{T}$ allows the scheme $\mathcal{T}$ to optimize both the regularizer $PV_{\mathscr{B}}(u)$ and intensity parameter α simultaneously. The main result is:

THEOREM 1.1 (see Theorem 4.1). *The training scheme $\mathcal{T}$ admits at least one solution $(\alpha_{\mathbb{T}}, \mathscr{B}_{\mathbb{T}}) \in \mathbb{T}$, and provides an associated optimally reconstructed image $u_{\alpha_{\mathbb{T}},\mathscr{B}_{\mathbb{T}}} \in BV_{\mathscr{B}_{\mathbb{T}}}(Q)$.*

In the third part of this article we focus on how to numerically determine the optimal solution of scheme $\mathcal{T}$, or equivalently, compute global minimizers of the *assessment function* $\mathcal{A}(\alpha,\mathscr{B})$: $\mathbb{T} \to \mathbb{R}^+$ defined as

$$\mathcal{A}(\alpha,\mathscr{B}) := \|u_{\alpha,\mathscr{B}} - u_c\|^2_{L^2(Q)}, \qquad (1.4)$$

where $u_{\alpha,\mathscr{B}}$ is obtained from ($\mathcal{T}$-L2). However, as shown in [23], even in the simplest case with $\mathscr{B} = \nabla$ (i.e. $PV_{\mathscr{B}} = TV$), the assessment function $\mathcal{A}(\alpha,\nabla)$ is not quasi-convex (in the sense of [18]), and hence methods such as *Newton's descent* or *Line search* might get trapped in a local minimum. To overcome this difficulty, we introduce the concept of the *acceptable optimal solution*. To be precise, we say the solution $(\alpha',\mathscr{B}')$ is an acceptable optimal solution of scheme $\mathcal{T}$ with the given error $\varepsilon > 0$ if

$$|\mathcal{A}(\alpha_{\mathbb{T}},\mathscr{B}_{\mathbb{T}}) - \mathcal{A}(\alpha',\mathscr{B}')| < \varepsilon, \qquad (1.5)$$

where $(\alpha_{\mathbb{T}},\mathscr{B}_{\mathbb{T}})$ is a global minimum obtained from ($\mathcal{T}$-L1).

To compute such an acceptable optimal solution, we use a finite approximation method, originally introduced and studied in [23], and generalize it in Section 4.2 to fit our new regularizer PV. To this aim, and also for the numerical realization of scheme $\mathcal{T}$, we add the following *box-constraint* on the training ground $\mathbb{T}$.

- The intensity parameter α is contained in a closed interval $[0,P]$, where the box-constraint constant $P > 0$ can be chosen by the user.
- The collection Σ of operators $\mathscr{B}$ satisfies an additional continuity assumption: For any $\mathscr{B}_1$, $\mathscr{B}_2 \in \Sigma$,

  $$|PV_{\mathscr{B}_1}(u) - PV_{\mathscr{B}_2}(u)| \le O(|\mathscr{B}_1 - \mathscr{B}_2|)\min\{PV_{\mathscr{B}_1}(u), PV_{\mathscr{B}_2}(u)\},$$

  where $O(\cdot)$ denotes the big-O notation.

Then, the finite approximation method is constructed based on a sequence of (finite) training sets $\mathbb{T}_l$, indexed by $l \in \mathbb{N}$, in which (where $\mathcal{H}^0(\cdot)$ denotes the counting measure)

$$\mathcal{H}^0(\mathbb{T}_l) < +\infty \text{ and } \mathbb{T} \subset \mathrm{cl}\left(\bigcup_{l\in\mathbb{N}} \mathbb{T}_l\right).$$

Here cl$(\cdot)$ denotes the topological closure. The precise definition of $\mathbb{T}_l$ will be presented in Definition 4.2 below. We remark that, since $\mathcal{H}^0(\mathbb{T}_l) < +\infty$ for each $l \in \mathbb{N}$ fixed, we could evaluate $\mathcal{A}(\alpha,\mathscr{B})$ at each element of $\mathbb{T}_l$ and determine the optimal solution(s)

$$(\alpha_{\mathbb{T}_l}, \mathscr{B}_{\mathbb{T}_l}) \in \mathbb{A}[\mathbb{T}_l] := \arg\min\{\mathcal{A}(\alpha,\mathscr{B}) : (\alpha,\mathscr{B}) \in \mathbb{T}_l\}$$

precisely. The following result, i.e. Theorem 1.2 below, is crucial for (1.5).

THEOREM 1.2 (see Theorem 4.2). *Let $\mathbb{T}$ be a training ground satisfying the above box-constraint. Then the following assertions hold:*

(1) *We have*

$$\lim_{l\to\infty} \mathrm{dist}(\mathbb{A}[\mathbb{T}], \mathbb{A}[\mathbb{T}_l]) = 0,$$

where dist *denotes the Hausdorff distance: Given two sets A, B, and a metric m, the Hausdorff distance between A, B is defined as*

$$\max\left\{\sup_{x\in A}\inf_{y\in B} m(x,y), \sup_{x\in B}\inf_{y\in A} m(x,y)\right\}.$$

(2) *Let $\varepsilon > 0$ be given. Then for each $l \in \mathbb{N}$ we have*

$$|\mathcal{A}(\alpha_{\mathbb{T}_l}, \mathscr{B}_{\mathbb{T}_l}) - \mathcal{A}(\alpha_{\mathbb{T}}, \mathscr{B}_{\mathbb{T}})| \le 4KP[O(P/l) + 1/l]^{1/2} \|u_\eta\|^{1/2}_{W^{d,1}(Q)}/\varepsilon^d + \varepsilon/2,$$

where the value of the right-hand side can be computed explicitly. Here d denotes the order of the considered differential operator $\mathscr{B}$, and the constant K is defined in the proof of Theorem 4.2.

This result states that, for any given $\varepsilon > 0$, we could compute a sufficiently large $l \in \mathbb{N}$ such that the corresponding optimal solution $(\alpha_{\mathbb{T}_l}, \mathscr{B}_{\mathbb{T}_l})$ is an acceptable optimal solution of scheme $\mathcal{T}$. Also, in Section 5.1 we show that, even with the box-constraint, the training ground $\mathbb{T}$ is still sufficiently large to encompass many interesting operators. We finally remark that, although this work focuses mainly on the theoretical analysis of the operators $PV_{\mathscr{B}}$ and the training scheme $\mathcal{T}$, in Section 5.1 a primal-dual algorithm for solving ($\mathcal{T}$-L2) is discussed, and some numerical realizations of scheme $\mathcal{T}$ are provided.

This article is structured as follows. In Section 2 we study the theoretical properties of $PV_{\mathscr{B}}$-seminorms. The Γ-convergence result, the bilevel training scheme, and the finite approximation are the subjects of Sections 3 and 4, respectively. Finally, in Section 5.1 we show several numerical implementations.

2. The space of functions with bounded PV-seminorm

Let d, $N \in \mathbb{N}$ be given, and let $Q := (0,1)^N$ be the open unit cube in $\mathbb{R}^N$. Denote by $\mathbb{M}^{N_n}$ the space of matrices/tensors with dimension $N \times \cdots \times N$ (n times), and real entries. For convenience, we identify the matrix space $\mathbb{M}^{N_n}$ with the vector space $\mathbb{R}^{N^n}$. Moreover, we denote by $\mathcal{D}'(Q,\mathbb{R}^n)$ the space of real-valued distributions on $\mathbb{R}^n$.

NOTATION 2.1. *We first collect some notations, which will be used in connection with linear differential operators.*

(1) *For $h \in \mathbb{N}$, we denote by $H^h\colon \mathcal{D}'(Q) \to \mathcal{D}'(Q;\mathbb{R}^{N^h})$ the h-th* Hessian *differential operator. For example, when $h=1$, we have $H^1 u = \nabla u$.*

(2) *For $h=1,\dots,d$, we let B^h be a matrix mapping from $\mathbb{R}^{N^h}$ to $\mathbb{R}^{N^h}$ and*

$$K := \sum_{h\in\mathbb{N}, h\le d} N^h.$$

We denote by $\mathscr{B}:\mathcal{D}'(Q)\to\mathcal{D}'(Q;\mathbb{R}^K)$ the d-th order differential operator

$$\mathscr{B}u := \sum_{h\in\mathbb{N}, h\le d} B^h(H^h u). \tag{2.1}$$

(3) *For $h=1,\dots,d$, we denote by $(B^h)^*$ the formal adjoint of the matrix B^h, and we define the differential operator $\mathscr{B}^*:\mathcal{D}'(Q;\mathbb{R}^K)\to\mathcal{D}'(Q)$ by*

$$\langle \mathscr{B}^* v, u\rangle_{\mathbb{R}} := \langle v, \mathscr{B}u\rangle_{\mathbb{R}^K}.$$

(4) *We denote by $\circ_{\mathscr{B}}$ the bilinear operator induced by $\mathscr{B}$, i.e.*

$$\mathscr{B}(uw) = w\mathscr{B}u + u\circ_{\mathscr{B}} w. \tag{2.2}$$

(5) *Given a sequence of operators $\{\mathscr{B}_n\}_{n=1}^\infty$ and an operator $\mathscr{B}$, with coefficients $\{B_n\}_{n=1}^\infty$ and B, respectively, we say that $\mathscr{B}_n\to\mathscr{B}$ in ℓ^∞ if*

$$|\mathscr{B}_n - \mathscr{B}| := \sum_{h\le d} \left|B_n^h - B^h\right|_{\ell^\infty} \to 0,$$

where $|\cdot|_{\ell^\infty}$ denotes the ℓ^∞ matrix norm.

Definition 2.1. *Let $d\in\mathbb{N}$ be fixed. We denote by Π^d the collection of operators $\mathscr{B}$ defined in Notation 2.1, with order at most d.*

2.1. The PDE-constraint total variation defined by an operator $\mathscr{B}$.

We generalize the standard total variation seminorm by using the d-th order differential operators $\mathscr{B}\in\Pi^d$ defined in Definition 2.1.

Definition 2.2. *Let $u\in L^1(Q)$ and $\mathscr{B}\in\Pi^d$ be given.*

(1) *We define the* PDE-constraint *seminorm, say $PV_{\mathscr{B}}$, by*

$$PV_{\mathscr{B}}(u) := \sup\left\{\int_Q u\mathscr{B}^*\varphi\, dx : \varphi\in C_c^\infty(Q;\mathbb{R}^K),\, |\varphi|\le 1\right\}; \tag{2.3}$$

(2) *We define the space*

$$BV_{\mathscr{B}}(Q) := \left\{u\in L^1(Q) : PV_{\mathscr{B}}(u) < +\infty\right\},$$

and equip it with the norm

$$\|u\|_{BV_{\mathscr{B}}(Q)} := \|u\|_{L^1(Q)} + PV_{\mathscr{B}}(u). \tag{2.4}$$

In the next proposition we collect several preliminary results on functions in $BV_{\mathscr{B}}(Q)$.

Proposition 2.1. *Let $\mathscr{B}\in\Pi^d$ and $u\in BV_{\mathscr{B}}(Q)$ be given.*

(1) *For any sequence $\{u_n\}_{n=1}^{\infty}\subset L^1(Q)$ and function $u\in L^1(Q)$, we have*

$$\liminf_{n\to\infty} PV_{\mathscr{B}}(u_n)\geq PV_{\mathscr{B}}(u), \tag{2.5}$$

provided that one of the following conditions is satisfied:

(i) *$\{u_n\}_{n=1}^{\infty}$ is locally uniformly integrable and $u_n\to u$ a.e.;*

(ii) *$u_n \overset{*}{\rightharpoonup} u$ in $\mathcal{M}_b(Q)$.*

(2) *There exists a* Radon *measure μ, and a μ-measurable function σ: $Q\to\mathbb{R}^K$, such that*

(i) *$|\sigma(x)|=1$ μ-a.e.;*

(ii) *for all $\varphi\in C_c^\infty(Q;\mathbb{R}^K)$, it holds*

$$\int_Q u\mathscr{B}^*\varphi\,dx=-\int_Q\varphi\cdot\sigma\,d\mu.$$

Proof. We prove Assertion 1 first. If

$$\liminf_{n\to\infty} PV_{\mathscr{B}}(u_n)=+\infty,$$

then there is nothing to prove. Assume the opposite, i.e. for an arbitrary $\varphi\in C_c^\infty(Q;\mathbb{R}^K)$,

$$\liminf_{n\to\infty} PV_{\mathscr{B}}(u_n)\geq\liminf_{n\to\infty}\int_Q u_n\mathscr{B}^*\varphi\,dx=\int_Q u\mathscr{B}^*\varphi\,dx,$$

where the last equality can be deduced from either condition 1(i) or 1(ii). Hence, we conclude (2.5) by the arbitrariness of φ.

We next prove Assertion 2. We define the linear functional

$$L: C_c^\infty(Q;\mathbb{R}^K)\to\mathbb{R},\qquad L(\varphi):=-\int_Q u\mathscr{B}^*\varphi\,dx,\text{ for }\varphi\in C_c^\infty(Q;\mathbb{R}^K).$$

Then, since $u\in BV_{\mathscr{B}}(Q)$, we have that

$$\sup\left\{\frac{1}{\|\varphi\|_{L^\infty(Q)}}\int_Q u\mathscr{B}^*\varphi\,dx:\text{ for }\varphi\in C_c^\infty(Q;\mathbb{R}^K)\right\}=PV_{\mathscr{B}}(u)<+\infty,$$

which implies that

$$|L(\varphi)|\leq PV_{\mathscr{B}}(u)\,\|\varphi\|_{L^\infty(Q)}. \tag{2.6}$$

Now, for an arbitrary $\varphi\in C_c(Q;\mathbb{R}^K)$, we define $\varphi_\varepsilon:=\varphi*\eta_\varepsilon$, for some suitable mollifier η_ε with $\varepsilon<\operatorname{dist}(\operatorname{spt}(\varphi),\partial Q)$. The particular form of η_ε is not really relevant, but we point out that, for instance, a possible choice for such η_ε can be a Gaussian distribution with mean zero and variance $1/\varepsilon^2$. Then we have, by [14, Theorem 1, item (ii), Section 4.2], that $\varphi_\varepsilon\to\varphi$ uniformly on Q. Then, define

$$\bar{L}(\varphi):=\lim_{\varepsilon\to 0}L(\varphi_\varepsilon)\text{ for }\varphi\in C_c(Q;\mathbb{R}^K),$$

and together with (2.6), we conclude that

$$\sup\left\{\bar{L}(\varphi):\ \text{for } \varphi\in C_c(Q;\mathbb{R}^K) \text{ and } |\varphi|\le 1\right\}<+\infty.$$

Thus, by the Riesz representation theorem (see [14, Section 1.8]), the proof is complete. ∎

REMARK 2.1. We henceforth have

$$\int_Q u\mathscr{B}^*\varphi\,dx=-\int_Q \varphi\cdot\sigma\,d|\mathscr{B}u|$$

for arbitrary $\varphi\in C_c^\infty(Q;\mathbb{R}^K)$.

THEOREM 2.1 (local approximation by smooth functions). *Let $p\ge 1$ and $u\in BV_{\mathscr{B}}(Q)\cap L^p(Q)$ be given. There exists a sequence $\{u_n\}_{n=1}^\infty\subset C^\infty(Q)\cap BV_{\mathscr{B}}(Q)$ such that the following assertions hold.*

(1) *$u_n\to u$ strongly in $L^p(Q)$;*

(2) *$PV_{\mathscr{B}}(u_n)\to PV_{\mathscr{B}}(u)$;*

(3) *$u_n\in C^\infty(\bar{Q})$ for each $n\in\mathbb{N}$.*

REMARK 2.2. Assertion 3 only states that, for each fixed $n\in\mathbb{N}$, $u_n\in C^\infty(\bar{Q})$, but it is possible that $\|u_n\|_{L^1(\partial Q)}\to\infty$ as $n\to\infty$. That is, we make no conclusions about the trace of u.

Proof. The construction of the approximation sequence $\{u_n\}_{n=1}^\infty$ is almost identical to that for the standard BV case, presented in [14, Theorem 2, Page 172]. We shall only concentrate on showing Assertion 3, but for readers' convenience, we quickly outline the construction of approximation sequence, and its key steps.

Let $u\in BV_{\mathscr{B}}(Q)$ be given, and let Q_k be the cube centered at point $q=(1/2,\cdots,1/2)$ with side length $1-1/(k+M)$. Given an arbitrary $\varepsilon>0$, we choose $M>0$ large enough such that

$$|\mathscr{B}u|(Q\setminus Q_{0+M})<\varepsilon/2.$$

Define $Q_0=Q_{0+M}$ and

$$V_k:=Q_{k+1}\setminus\bar{Q}_{k-1}\ \text{for } k\in\mathbb{N}.$$

Let $\{\zeta_k\}_{k=1}^\infty\subset C_c^\infty(Q)$ be the partition of unity such that

$$\begin{aligned}&\zeta_k\in C_c^\infty(V_k)\ \text{such that } 0\le\zeta_k\le 1,\\&\sum_{k\ge1}\zeta_k(x)=1\ \text{for each } x\in Q.\end{aligned}$$

Let η_ε be a suitable mollifier, and for each k, we choose ε_k small enough such that

$$\operatorname{spt}(\eta_{\varepsilon_k}*(u\zeta_k))\subset V_k, \tag{2.7}$$

$$\|\eta_{\varepsilon_k}*(u\zeta_k)-u\zeta_k\|_{L^p(Q)}<\varepsilon/2^{k+1}, \tag{2.8}$$

$$\|\eta_{\varepsilon_k}*(u\mathscr{B}\zeta_k)-u\mathscr{B}\zeta_k\|_{L^1(Q)}<\varepsilon/2^{k+1}, \tag{2.9}$$

and we define

$$u_\varepsilon:=\sum_{k=1}^\infty\eta_{\varepsilon_k}*(u\zeta_k).$$

We observe that (2.7) implies that $u_\varepsilon \in C^\infty(Q)$, and (2.8) implies that

$$u_\varepsilon \to u \text{ strongly in } L^p(Q).$$

This, combined with Assertion 1 of Proposition 2.1, gives

$$\liminf_{\varepsilon\to 0} PV_{\mathscr{B}}(u_\varepsilon) \geq PV_{\mathscr{B}}(u).$$

Next, for arbitrary $\varphi \in C_c^\infty(Q;\mathbb{R}^N)$, we observe that,

$$\langle \eta_{\varepsilon_k} * (u\zeta_k), \mathscr{B}^*\varphi\rangle = \langle u\zeta_k, \mathscr{B}^*(\eta_{\varepsilon_k} * \varphi)\rangle = \langle u, \mathscr{B}^*(\zeta_k(\eta_{\xi_k} * \varphi))\rangle - \langle u, (\eta_{\varepsilon_k} * \varphi) \circ_{\mathscr{B}^*} \nabla\zeta_k\rangle,$$

where, at the first equality we used the linearity of the convolution operator, and at the last equality we used (2.2). Thus, we have

$$\langle u_\varepsilon, \mathscr{B}^*\varphi\rangle = \sum_{k\geq 1}\langle \eta_{\varepsilon_k} * (u\zeta_k), \mathscr{B}^*\varphi\rangle = \sum_{k\geq 1}\langle u, \mathscr{B}^*(\zeta_k(\eta_{\xi_k} * \varphi))\rangle - \sum_{k\geq 1}\langle u, (\eta_{\varepsilon_k} * \varphi) \circ_{\mathscr{B}^*} \nabla\zeta_k\rangle.$$

Following the same computations from [14, Theorem 2, Page 172], and using (2.9), we deduce that

$$\langle u_\varepsilon, \mathscr{B}^*\varphi\rangle \leq PV_{\mathscr{B}}(u) + \varepsilon,$$

Hence, in view of the arbitrariness of φ, we obtain that

$$\limsup_{\varepsilon\to 0} PV_{\mathscr{B}}(u_\varepsilon) \leq PV_{\mathscr{B}}(u).$$

Finally, we further modify the sequence $\{u_\varepsilon\}_{\varepsilon>0}$, so that $u_\varepsilon \in C^\infty(\bar{Q})$ for each $\varepsilon > 0$. Let $\delta > 0$ be given, and define

$$u_{\varepsilon,\delta}(x) := u_\varepsilon((x-q)/(1+\delta)), \text{ for } x \in Q. \tag{2.10}$$

As a consequence, $u_{\varepsilon,\delta} \to u_\varepsilon$ strongly in $L^p(Q)$, and $PV_{\mathscr{B}}(u_{\varepsilon,\delta}) \to PV_{\mathscr{B}}(u_\varepsilon)$, as $\delta \to 0$. Hence, by a diagonal argument, we could extract a subsequence $\{u_{\delta_\varepsilon}\}_{\varepsilon>0}$ such that

$$u_{\delta_\varepsilon} \to u \text{ strongly in } L^p(Q) \text{ and } PV_{\mathscr{B}}(u_{\delta_\varepsilon}) \to PV_{\mathscr{B}}(u).$$

On the other hand, by the definition of u_{δ_ε}, we have $u_{\delta_\varepsilon} \in C^\infty(\bar{Q})$, which concludes Assertion 3, as desired. □

REMARK 2.3. The construction of $u_{\varepsilon,\delta}$ in (2.10) is only possible due to the simple geometry of Q. However, for domains with more complicated geometry, even with Lipschitz regular boundary, such construction is not available. We refer the interested reader to [4, 15] for alternative constructions which, however, require several additional conditions on the operator $\mathscr{B}$.

COROLLARY 2.1. *Let $\mathscr{B}_i$, $i = 1, \ldots M$ be given, and let*

$$u \in \bigcap_{i=1}^{M} BV_{\mathscr{B}_i}(Q).$$

Then, there exists a sequence

$$\{u_n\}_{n=1}^\infty \subset C^\infty(Q) \cap \bigcap_{i=1}^{M} BV_{\mathscr{B}_i}(Q)$$

such that the following assertions hold.

(1) $u_n \to u$ *strongly in* $L^1(Q)$;
(2) $PV_{\mathscr{B}_i}(u_n) \to PV_{\mathscr{B}_i}(u)$, *for each* $i = 1, \ldots, M$ *uniformly;*
(3) $u_n \in C^\infty(\bar{Q})$ *for each* $n \in \mathbb{N}$.

Proof. We only need to change (2.9) to

$$\sum_{i=1}^{M} \|\eta_{\varepsilon_k} * (u\mathscr{B}_i\zeta_k) - u\mathscr{B}_i\zeta_k\|_{L^1(Q)} < \varepsilon/2^{k+1},$$

and the rest follows with the same arguments used in the proof of Theorem 2.1. □

We close this section with a lower semi-continuity result for $PV_{\mathscr{B}}$.

PROPOSITION 2.2. *Given* $u \in L^1(Q)$, *and a sequence* $\{\mathscr{B}_n\}_{n=1}^\infty$ *such that* $\mathscr{B}_n \to \mathscr{B}$ *in* ℓ^∞, *it holds*

$$\liminf_{n\to\infty} PV_{\mathscr{B}_n}(u) \geq PV_{\mathscr{B}}(u).$$

Proof. First, if

$$\liminf_{n\to\infty} PV_{\mathscr{B}_n}(u) = +\infty,$$

then the thesis is trivial. Assume the opposite, i.e.

$$\sup\{PV_{\mathscr{B}_n}(u) : n \in \mathbb{N}\} := M < +\infty.$$

For arbitrary $\varphi \in C_c^\infty(Q;\mathbb{R}^K)$, we have

$$+\infty > \liminf_{n\to\infty} PV_{\mathscr{B}_n}(u) \geq \liminf_{n\to\infty} \int_Q u\mathscr{B}_n^*\varphi\, dx = \int_Q u\mathscr{B}^*\varphi\, dx.$$

Hence, by taking the supremum with respect to φ on the right-hand side, we conclude

$$\liminf_{n\to\infty} PV_{\mathscr{B}_n}(u) \geq PV_{\mathscr{B}}(u),$$

as desired. □

3. Analytic properties of PDE-constraint variations

3.1. Γ-convergence of functionals defined by PV seminorms.

In this section we prove a Γ-convergence result with respect to the intensity parameter α and operator $\mathscr{B}$.

DEFINITION 3.1. *We define the functional*

$$\mathcal{I}_{\alpha,\mathscr{B}} : L^1(Q) \to [0,+\infty], \qquad \mathcal{I}_{\alpha,\mathscr{B}}(u) := \begin{cases} \|u - u_\eta\|_{L^2(Q)}^2 + \alpha PV_{\mathscr{B}}(u) & \text{if } u \in BV_{\mathscr{B}}(Q), \\ +\infty & \text{otherwise.} \end{cases}$$

The following theorem is the main result of this section.

THEOREM 3.1. *Let sequences* $\{\mathscr{B}_n\}_{n=1}^\infty$ *and* $\{\alpha_n\}_{n=1}^\infty$ *be given such that* $\mathscr{B}_n \to \mathscr{B}_0$ *in* ℓ^∞ *and* $\alpha_n \to \alpha_0 \in \mathbb{R}^+$. *Then, the functional* $\mathcal{I}_{\alpha_n,\mathscr{B}_n}$ Γ*-converges to* $\mathcal{I}_{\alpha,\mathscr{B}}$ *in the weak* L^2 *topology. More precisely, for any* $u \in L^1(Q)$, *the following two assertions hold.*

(Lower semi-continuity) *For any sequence*

$$u_n \rightharpoonup u \text{ weakly in } L^2(Q),$$

we have

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

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

$$u_n\rightharpoonup u \text{ weakly in } L^2(Q)$$

and

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

We split the proof of Theorem 3.1 into two propositions.

The following result is instrumental in establishing the liminf inequality.

PROPOSITION 3.1. *Let sequences* $\{\mathscr{B}_n\}_{n=1}^{\infty}$ *and* $\{\alpha_n\}_{n=1}^{\infty}$ *be given, such that* $\mathscr{B}_n\to\mathscr{B}_0$ *in* ℓ^∞ *and* $\alpha_n\to\alpha_0\in\mathbb{R}^+$*. Let* $\{u_n\}_{n=1}^{\infty}\subset L^1(Q)$ *be given such that*

$$\sup\left\{\|u_n\|_{L^p(Q)}+PV_{\mathscr{B}_n}(u_n): n\in\mathbb{N}\right\}<+\infty \tag{3.1}$$

for some $p\in(1,+\infty]$*. Then there exists* $u_0\in BV_{\mathscr{B}_0}(Q)$ *such that, up to a subsequence (not relabeled),*

$$u_n\rightharpoonup u_0 \text{ weakly in } L^p(Q) \tag{3.2}$$

and

$$\liminf_{n\to\infty}PV_{\mathscr{B}_n}(u_n)\ge PV_{\mathscr{B}_0}(u_0). \tag{3.3}$$

Proof. Without loss of generality, we assume that $\alpha_n=1$ for every $n\in\mathbb{N}$, as the general case for α_n and $\alpha_0\in\mathbb{R}^+$ can be argued with straightforward adaptations.

From (3.1), and the fact that $p>1$, we have, up to a subsequence, the existence of $u_0\in L^p(Q)$ such that (3.2) holds.

Next, for arbitrary $\varphi\in C_c^\infty(Q;\mathbb{R}^K)$, we observe that

$$\begin{aligned}
&\limsup_{n\to\infty}\left|\int_Q u_n\mathscr{B}_n^*\varphi\,dx-\int_Q u_n\mathscr{B}_0^*\varphi\,dx\right|\\
&\le\limsup_{n\to\infty}\int_Q|u_n|\,|\mathscr{B}_n^*\varphi-\mathscr{B}_0^*\varphi|\,dx\\
&\le\left(\sup_{n\ge 0}\|u_n\|_{L^p}\right)\left(\limsup_{n\to\infty}\|\mathscr{B}_n^*\varphi-\mathscr{B}_0^*\varphi\|_{L^{p'}}\right)=0,
\end{aligned}\tag{3.4}$$

where we used the fact that $\varphi\in C_c^\infty(Q;\mathbb{R}^K)$, and the dominated convergence theorem.

Hence, we could obtain that

$$\begin{aligned}
\liminf_{n\to\infty}PV_{\mathscr{B}_n}(u_n)&\ge\liminf_{n\to\infty}\int_Q u_n\mathscr{B}_n^*\varphi\,dx\\
&\ge\liminf_{n\to\infty}\int_Q u_n\mathscr{B}_0^*\varphi\,dx+\liminf_{n\to\infty}\int_Q u_n(\mathscr{B}_n^*-\mathscr{B}_0^*)\varphi\,dx\ge\int_Q u_0\mathscr{B}_0^*\varphi\,dx,
\end{aligned}$$

where, at the last inequality we used (3.2) and (3.4). By the arbitrariness of $\varphi \in C_c^\infty(Q;\mathbb{R}^K)$, we conclude (3.3). ◻

Proposition 3.2 (Γ-limsup inequality). *Given sequences $\{\mathscr{B}_n\}_{n=1}^\infty$ and $\{\alpha_n\}_{n=1}^\infty$, such that $\mathscr{B}_n \to \mathscr{B}_0$ in ℓ^∞ and $\alpha_n \to \alpha_0 \in \mathbb{R}^+$, for any $u_0 \in BV_{\mathscr{B}_0}(Q)$ there exists a sequence $\{u_n\}_{n=1}^\infty \subset BV_{\mathscr{B}_n}(Q)$, such that (up to a subsequence) $u_n \rightharpoonup u_0$ in $L^p(Q)$, and*

$$\limsup_{n\to\infty} PV_{\mathscr{B}_n}(u_n) \le PV_{\mathscr{B}_0}(u_0).$$

Proof. If $PV_{\mathscr{B}_0}(u) = \infty$, then the thesis is trivial. Assume the opposite, and suppose for now that $u_0 \in C^\infty(\bar{Q})$, which gives that $u_0 \in BV_{\mathscr{B}_n}(Q)$ for each $n \in \mathbb{N}$. Fix $\delta > 0$, and choose $\varphi_{\delta,n} \in C_c^\infty(Q;\mathbb{R}^K)$ such that

$$PV_{\mathscr{B}_n}(u) \le \int_Q u \mathscr{B}_n^* \varphi_{\delta,n} dx + \delta. \tag{3.5}$$

We observe that

$$\left|\int_Q u_0 \mathscr{B}_n^* \varphi_{\delta,n} dx\right| = \left|\int_Q [\mathscr{B}_n u_0] \varphi_{\delta,n} dx\right| \le \|\varphi_{\delta,n}\|_{L^\infty(Q)} \int_Q |\mathscr{B}_n u_0|\, dx \le \int_Q |\mathscr{B}_n u_0|\, dx, \tag{3.6}$$

where, at the last inequality we used the fact that $\varphi_{\delta,n}$ satisfies (2.3). Next, since $u \in C^\infty(\bar{Q})$ and $\mathscr{B}_n \to \mathscr{B}$ in ℓ^∞, we have

$$|\mathscr{B}_n u(x)| \le \sup\{|\mathscr{B}_n|_{\ell^\infty} : n \in \mathbb{N}\} \cdot \sum_{h\le d} \left|H^h u_0(x)\right|_{\ell^\infty},$$

which implies that

$$\int_Q \sum_{h\le d} \left|H^h u(x)\right|_{\ell^\infty} dx \le \|u\|_{W^{d,+\infty}(Q)} < +\infty.$$

Thus, we could apply the dominated convergence theorem to conclude that

$$\limsup_{n\to\infty} \int_Q |\mathscr{B}_n u_0|\, dx \le \int_Q \limsup_{n\to\infty} |\mathscr{B}_n u_0|\, dx = \int_Q |\mathscr{B}_0 u_0|\, dx.$$

This, combined with (3.5) and (3.6), gives

$$\limsup_{n\to\infty} PV_{\mathscr{B}_n}(u_0) \le \limsup_{n\to\infty} \int_Q u_0 \mathscr{B}_n^* \varphi_{\delta,n} dx + \delta \le \int_Q |\mathscr{B}_0 u_0|\, dx + \delta = PV_{\mathscr{B}_0}(u_0) + \delta,$$

which implies, by taking $\delta \searrow 0$, that

$$\limsup_{n\to\infty} PV_{\mathscr{B}_n}(u_0) \le PV_{\mathscr{B}_0}(u_0). \tag{3.7}$$

Next, by Theorem 2.1, we could construct an approximation sequence $\{u_\varepsilon\}_{\varepsilon>0} \subset C^\infty(\bar{Q})$ such that $u_\varepsilon \to u$ in $L^p(Q)$ and

$$PV_{\mathscr{B}_0}(u_\varepsilon) \to PV_{\mathscr{B}_0}(u), \text{ or } PV_{\mathscr{B}_0}(u_\varepsilon) \le PV_{\mathscr{B}_0}(u) + O(\varepsilon).$$

Also, by (3.7), we have

$$\limsup_{n\to\infty} PV_{\mathscr{B}_n}(u_\varepsilon) \le PV_{\mathscr{B}_0}(u_\varepsilon) \le PV_{\mathscr{B}_0}(u) + O(\varepsilon).$$

Thus, by a diagonal argument, we can obtain a sequence $\{\mathscr{B}_{n_\varepsilon}\}_{\varepsilon>0}$ such that

$$PV_{\mathscr{B}_{n_\varepsilon}}(u_\varepsilon) \le PV_{\mathscr{B}_0}(u) + O(\varepsilon).$$

That is, we have

$$\limsup_{\varepsilon\to 0} PV_{\mathscr{B}_{n_\varepsilon}}(u_\varepsilon) \le PV_{\mathscr{B}_0}(u),$$

which concludes the proof. ∎

Proof. **(Proof of Theorem 3.1.)** Property *(Lower semi-continuity)* holds in view of Proposition 3.1, and Property *(Recovery sequence)* follows from Proposition 3.2. ∎

3.2. The point-wise characterization of sub-differential of $PV_{\mathscr{B}}$. We first recall a few notations, preliminary results, and definitions.

DEFINITION 3.2 ([13, Definition 4.1 & 5.1]). *Let F be a given function on a normed space V, taking values in $\bar{\mathbb{R}}$.*

(1) *We define the* polar *function of F, denoted by F^*, by*

$$F^*(u^*) = \sup\left\{\langle v, u^*\rangle_{V,V^*} - F(v) : v \in V\right\}.$$

(2) *We define the* bipolar *function, say F^{**}, of F by*

$$F^{**} = (F^*)^*.$$

(3) *We say F is* sub-differentiable *at point $u \in V$ if $F(u)$ is finite and there exists $u^* \in V^*$ such that*

$$\langle v - u, u^*\rangle_{V,V^*} + F(u) \le F(v)$$

for all $v \in V$. Then such $u^* \in V^*$ is called a sub-gradient *of F at u, and the set of sub-gradients at u is called the* sub-differential *at u, and is denoted by $\partial F(u)$.*

For brevity, when the sub-differential is a singleton, we might identify it with its only element, and write "$\partial F(u) = v$" instead of "$\partial F(u) = \{v\}$".

PROPOSITION 3.3 ([13, Propositions 4.1 and 5.1]). *Let $F : V \longrightarrow \bar{\mathbb{R}}$ be a given function, and let F^* be its polar. Then the following assertions hold.*

(1) *We have $u^* \in \partial F(u)$ if and only if*

$$F(u) + F^*(u^*) = \langle u, u^*\rangle.$$

(2) *The (possibly empty) set $\partial F(u)$ is convex and closed.*

(3) *If in addition F is convex, then $F^{**} = F$.*

DEFINITION 3.3. *Let $p \in [1, +\infty)$, $v \in L^p(Q;\mathbb{R}^K)$, and $\mathscr{B}$ be given.*

(1) *We say that $\mathscr{B}^* v$ in $L^p(Q)$ if there exists $w \in L^p(Q)$ such that for all $\varphi \in C_c^\infty(Q;\mathbb{R}^K)$*

$$\int_Q \mathscr{B}\varphi \cdot v\, dx = -\int_Q \varphi w\, dx.$$

(2) *We define the space*

$$W^p[\mathscr{B}](Q;\mathbb{R}^K) := \left\{v \in L^p(Q;\mathbb{R}^K) : \mathscr{B}^* v \in L^p(Q)\right\}$$

endowed with the norm

$$\|v\|^p_{W^p(\mathscr{B})} := \|v\|^p_{L^p(Q)} + \|\mathscr{B}^* v\|^p_{L^p(Q)}.$$

(3) *We define*

$$W_0^p[\mathscr{B}](Q;\mathbb{R}^K) := \mathrm{cl}(C_c^\infty(Q;\mathbb{R}^K))_{\|\cdot\|_{W^p(\mathscr{B})}},$$

i.e., the closure of $C_c^\infty(Q;\mathbb{R}^K)$ *with respect to the norm* $\|\cdot\|_{W^p(\mathscr{B})}$.

(4) *We define*

$$C_c^\infty[\mathscr{B}](Q) := \left\{\mathscr{B}^*\varphi : \varphi \in C_c^\infty(Q;\mathbb{R}^K),\ \|\varphi\|_{L^\infty(Q)} \le 1\right\}.$$

and

$$K^p[\mathscr{B}](Q) := \left\{\mathscr{B}^* v : v \in W_0^p[\mathscr{B}](Q;\mathbb{R}^K),\ \|v\|_{L^\infty(Q)} \le 1\right\}.$$

The main result of Section 3.2 is:

THEOREM 3.2. *Let* $p>1$, $q=p/(p-1)$, $u\in L^p(Q)$, *and* $\tilde{u}\in L^q(Q)$ *be given. Then* $\tilde{u}\in\partial PV_{\mathscr{B}}(u)$ *if and only if the following two conditions hold:*

(1) $u \in BV_{\mathscr{B}}(Q)$;

(2) *there exist* $v\in W_0^q[\mathscr{B}](Q;\mathbb{R}^K)$ *such that* $\|v\|_{L^\infty(Q)}\le 1$, $\tilde{u}=\mathscr{B}^* v$, *and*

$$PV_{\mathscr{B}}(u) = \int_Q u\mathscr{B}^* v\, dx.$$

We prove Theorem 3.2 over several propositions.

PROPOSITION 3.4. *Let* $p\in(1,+\infty)$ *be given. Then the closure of* $C_{\mathscr{B}}(Q)$ *under the* L^q *norm is equal to* $W_0^q[\mathscr{B}](Q)$, *i.e.,*

$$\mathrm{cl}(C_{\mathscr{B}}(Q))_{L^q(Q)} = W_0^q[\mathscr{B}](Q).$$

Proof. We claim

$$\mathrm{cl}(C_{\mathscr{B}}(Q))_{L^q(Q)} \subset W_0^q[\mathscr{B}](Q) \tag{3.8}$$

first, and we do it by showing the space $W_0^q[\mathscr{B}](Q)$ is closed with respect to the L^q norm. Let $g\in \mathrm{cl}(W^q[\mathscr{B}](Q;\mathbb{R}^K))$ be given, and consider a sequence $\{v_n\}_{n=1}^\infty \subset W_0^q[\mathscr{B}](Q;\mathbb{R}^K)$ such that

$$\|\mathscr{B}^* v_n - g\|_{L^q(Q)} \to 0. \tag{3.9}$$

Since $\{v_n\}_{n=1}^\infty \subset W^q[\mathscr{B}](Q;\mathbb{R}^K)$, we have $\|v_n\|_{L^\infty(Q)}\le 1$ and hence, up to a subsequence, there exists $v_0\in L^\infty(Q)$ such that

$$v_n \rightharpoonup v_0 \text{ weakly in } L^q(Q) \text{ and } \|v_0\|_{L^\infty(Q)} \le 1.$$

Next, let $\phi\in C_c^\infty(Q)$ be given, and we observe that

$$\int_Q \mathscr{B}^* v_n \phi\, dx = -\int_Q v_n \mathscr{B}\phi\, dx \to -\int_Q v_0\mathscr{B}\phi\, dx,$$

which, combined with (3.9), gives

$$\int_Q g\phi\, dx = -\int_Q v_0 \mathscr{B}\phi\, dx,$$

and hence $g = \mathscr{B}^* v_0$. Thus, we have $v_0 \in W_0^q[\mathscr{B}](Q;\mathbb{R}^K)$. Next, since the set

$$\left\{(v, \mathscr{B}^* v) : v \in W_0^q[\mathscr{B}](Q;\mathbb{R}^K)\right\} \subset L^q(Q;\mathbb{R}^K \times \mathbb{R})$$

is convex and closed, hence by [5, Theorem 3.7], it is weakly closed. Thus, we conclude that $v_0 \in W_0^q[\mathscr{B}](Q)$, which implies that $g \in W_0^q[\mathscr{B}](Q)$, i.e. $W_0^q[\mathscr{B}](Q)$ is closed with respect to the L^q norm, and also (3.8) too, as desired.

We next claim that

$$\mathrm{cl}(C_{\mathscr{B}}(Q))_{L^q} \supset W_0^q[\mathscr{B}](Q). \tag{3.10}$$

We prove (3.10) by following the arguments used in [14, Theorem 2, Page 125]. Let $g \in W_0^q[\mathscr{B}](Q)$ be given. That is, there exists $v \in W_0^q[\mathscr{B}](Q;\mathbb{R}^K)$, $\|v\|_{L^\infty(Q)} \leq 1$, and $g = \mathscr{B}^* v$. From the definition of $W^q[\mathscr{B}](Q)$, there exists $\{v_n\}_{n=1}^\infty \subset C_c^\infty(Q;\mathbb{R}^K)$ such that

$$\|v_n - v\|_{W^q[\mathscr{B}](Q)} \to 0. \tag{3.11}$$

Next, define the truncated function

$$\bar{v}_n(x) := \begin{cases} 1, & \text{if } v_n(x) \geq 1, \\ v_n(x), & \text{if } -1 \leq v_n(x) \leq 1, \\ -1, & \text{if } v_n(x) \leq -1, \end{cases} \tag{3.12}$$

and we note that

$$\bar{v}_n \to v \text{ a.e., and } \mathscr{B}^* \bar{v}_n \rightharpoonup \mathscr{B}^* v_0 \text{ weakly in } L^q(Q). \tag{3.13}$$

Using an argument similar to that from the proof of Proposition 2.1, combined with the fact that $\bar{v}_n \to v$ a.e., we obtain

$$\liminf_{n\to\infty} \|\mathscr{B}^* \bar{v}_n\|_{L^q(Q)} \geq \|\mathscr{B}^* v_0\|_{L^q(Q)}.$$

On the other hand, by (3.12), we have

$$\|\mathscr{B}^* \bar{v}_n\|_{L^q(Q)} \leq \|\mathscr{B}^* v_0\|_{L^q(Q)},$$

and hence

$$\lim_{n\to\infty} \|\mathscr{B}^* \bar{v}_n\|_{L^q(Q)} = \|\mathscr{B}^* v_0\|_{L^q(Q)}.$$

Combining with the second part in (3.13), and using [5, page 124, Exercise 4.19, 1], gives

$$\lim_{n\to\infty} \|\mathscr{B}^* \bar{v}_n - \mathscr{B}^* v_0\|_{L^q(Q)} = 0.$$

We next modify the sequence $\{\bar{v}_n\}_{n=1}^\infty$ so that $\{\mathscr{B}^* \bar{v}_n\}_{\varepsilon>0} \subset C_c^\infty[\mathscr{B}](Q)$. From the arguments used in Theorem 2.1, we obtain a sequence of sets V_k, $k \in \mathbb{N}$, and a partition of unity $\zeta_k \in C_c^\infty(Q)$. Next, for each k, we choose ε_k small enough such that

$$\mathrm{spt}(\eta_{\varepsilon_k} * (\bar{v}_n \zeta_k)) \subset V_k, \tag{3.14}$$

$$\|\eta_{\varepsilon_k} * (\bar{v}_n \zeta_k) - \bar{v}_n \zeta_k\|_{L^q(Q)} < \varepsilon / 2^{k+1}, \tag{3.15}$$

$$\|\eta_{\varepsilon_k} * (\mathscr{B}^*(\bar{v}_n\zeta_k)) - \mathscr{B}^*(\bar{v}_n\zeta_k)\|_{L^q(Q)} < \varepsilon/2^{k+1}, \tag{3.16}$$

and

$$\varepsilon_k \leq \operatorname{dist}(\partial Q, \operatorname{spt}(\bar{v}_n))/8. \tag{3.17}$$

Here $\operatorname{spt}(\cdot)$ denotes the support of the function. Then, we define

$$v_{\varepsilon,n} := \sum_{k=1}^{\infty} \eta_{\varepsilon_k} * (\bar{v}_n\zeta_k),$$

and (3.17) gives that $v_{\varepsilon,n} \in C_c^\infty(Q)$. Following the same arguments in [14, Theorem 2, Page 125], we have that

$$\lim_{\varepsilon\to 0} \|v_{\varepsilon,n} - \bar{v}_n\|_{W^q[\mathscr{B}](Q)} = 0,$$

which, combined with (3.11), allows us to construct, by a diagonal argument, a sequence $\{v_{\varepsilon_n}\}_{n=1}^\infty$ such that

$$\lim_{\varepsilon\to 0} \|v_{\varepsilon_n} - v\|_{W^q[\mathscr{B}](Q)} = 0.$$

Moreover, we observe that

$$|v_{\varepsilon_n}| \leq \left| \sum_{k=1}^{\infty} \eta_{\varepsilon_k} * (\bar{v}_n\zeta_k) \right| \leq |\bar{v}_n| \leq 1,$$

which shows that $\{\mathscr{B}^* v_{\varepsilon_n}\}_{n=1}^\infty \subset C_c^\infty[\mathscr{B}](Q)$, and hence (3.10), and the proof is complete. □

Now we ready to prove Theorem 3.2.

Proof. **(Proof of Theorem 3.2.)** We first claim that the convex conjugate of $PV_{\mathscr{B}}$, say $PV_{\mathscr{B}}^*$, has the form

$$PV_{\mathscr{B}}^*(v) = I_{W_0^q[\mathscr{B}](Q)}(v) =: \begin{cases} 0 & \text{if } v \in W_0^q[\mathscr{B}](Q), \\ +\infty & \text{if } v \notin W_0^q[\mathscr{B}](Q). \end{cases}$$

By Definition 3.2 and Proposition 3.4, we have that

$$I^*_{W_0^q[\mathscr{B}](Q)}(u) = PV_{\mathscr{B}}(u).$$

Next, since the seminorm $PV_{\mathscr{B}}$ and the indicator function $I_{\mathrm{cl}(C_{\mathscr{B}}(Q))_{L^q(Q)}}$ are both convex and lower semi-continuous, we have

$$PV_{\mathscr{B}}^*(v) = (I^*_{W_0^q[\mathscr{B}](Q)})^* = I_{W_0^q[\mathscr{B}](Q)}.$$

Finally, in view of Proposition 3.3,

$$u^* \in \partial PV_{\mathscr{B}}(u)$$

holds if and only if

$$PV_{\mathscr{B}}(u) + PV_{\mathscr{B}}^*(u^*) = \langle u, u^* \rangle,$$

concluding the proof. ∎

REMARK 3.1. In view of Proposition 3.4, we have actually showed that, for any $v \in W_0^p[\mathscr{B}](Q)$, with $\|v\|_{L^\infty(Q)} \le 1$, it holds

$$\int_Q u \mathscr{B}^* v \, dx \le PV_{\mathscr{B}}(u).$$

THEOREM 3.3 (The point-wise characterization of $\partial PV_{\mathscr{B}}$). *Let $u \in L^p(Q) \cap BV_{\mathscr{B}}(Q)$, $p > 1$, be given. Let $v \in W_0^p[\mathscr{B}](Q)$ be such that $\mathscr{B}^* v \in \partial PV_{\mathscr{B}}(u)$. Then we have*

$$v = \sigma_u \text{ a.e. } x \in Q,$$

where σ_u is the density of $\mathscr{B}u$ with respect to $|\mathscr{B}u|$ (see Remark 2.1).

Proof. Let $u \in L^p(Q) \cap BV_{\mathscr{B}}(Q)$ be given, and let $v \in W_0^p[\mathscr{B}](Q)$ be the function obtained from Theorem 3.2. Then, by the definition of $W_0^p[\mathscr{B}](Q)$, we could obtain a sequence $\{v_n\}_{n=1}^\infty \subset C_c^\infty[\mathscr{B}^*](Q)$ such that $\mathscr{B}^* v_n \to \mathscr{B}^* v$ strongly in $L^p(Q)$.

We claim that

$$\|\sigma_u - v_n\|_{L^p(Q,|\mathscr{B}u|)} \to 0. \tag{3.18}$$

From the definition of $PV_{\mathscr{B}}$ and Theorem 2.1, we have that

$$\int_Q u \mathscr{B}^* v_n \, dx = \int_Q v_n \cdot \sigma_u d|\mathscr{B}u|. \tag{3.19}$$

On the other hand, since $\{v_n\}_{n=1}^\infty \subset C_c^\infty[\mathscr{B}^*](Q)$, we have $\|v_n\|_{L^\infty(Q)} \le 1$ and hence, together with the fact that $|\sigma_u| = 1$ $|\mathscr{B}u|$ a.e., we observe that

$$\begin{aligned} 1 - (\sigma_u \cdot v_n) &= \frac{1}{2}|\sigma_u|^2 - (\sigma_u \cdot v_n) + \frac{1}{2}|v_n|^2 + \frac{1}{2}|\sigma_u|^2 - \frac{1}{2}|v_n|^2 \\ &= \frac{1}{2}|\sigma_u - v_n|^2 + \frac{1}{2}|\sigma_u|^2 - \frac{1}{2}|v_n|^2 \ge \frac{1}{2}|\sigma_u - v_n|^2 \ge 0. \end{aligned}$$

Therefore, we could compute that

$$\begin{aligned} \int_Q |v_n - \sigma_u| d|\mathscr{B}u| &= \int_Q 1 \cdot |v_n - \sigma_u| d|\mathscr{B}u| \\ &\le \left(\int_Q 1 \cdot d|\mathscr{B}u|\right)^{1/2} \cdot \left(\int_Q |v_n - \sigma_u|^2 d|\mathscr{B}u|\right)^{1/2} \\ &\le [PV_{\mathscr{B}}(u)]^{1/2} \cdot \left(\int_Q 1 - (\sigma_u \cdot v_n) d|\mathscr{B}u|\right)^{1/2}. \end{aligned} \tag{3.20}$$

Next, from (3.19), we have that

$$\lim_{n\to\infty} \int_Q v_n \cdot \sigma_u d|\mathscr{B}u| = \int_Q u \mathscr{B}^* v \, dx = PV_{\mathscr{B}}(u) = \int_Q 1 \cdot d|\mathscr{B}u|.$$

This, combined with (3.20), gives (3.18), as desired. ∎

PROPOSITION 3.5. *Let $u \in BV_{\mathscr{B}}(Q)$ and $V \subset\subset Q$ be given. Then, for $u^* \in \partial PV_{\mathscr{B}}(u)$ and $u_V^* \in \partial PV_{\mathscr{B}}(u)\lfloor_V$, we have*

$$u^*(x) = u_V^*(x) \text{ for } |\mathscr{B}u|\text{-a.e. } x \in V.$$

Proof. Let v and $v_V \in W_0^p[\mathscr{B}](Q;\mathbb{R}^K)$ be such that Assertions 1 and 2 hold for $PV_{\mathscr{B}}(u)$ and $PV_{\mathscr{B}}(u)\lfloor_V$, respectively. Then, by Theorem 3.3, both $v(x)$ and $v_V(x)$ can be represented by the density of $\mathscr{B}u$ with respect to $|\mathscr{B}u|$, concluding the proof. ∎

4. Learning the optimal operator $\mathscr{B}$ in image processing problems

In this section we use the bilevel training scheme introduced in Section 1 to determine the optimal setting of $PV_{\mathscr{B}}$ for a given *training pair* (u_c,u_η), where $u_\eta\in L^2(Q)$ and $u_c\in BV(Q)$ represent the corrupted and clean images, respectively.

4.1. The bilevel training scheme with the $PV_{\mathscr{B}}$ regularizer.

We collect few notations first.

NOTATION 4.1. *Recall the definition of $\mathscr{B}$ from Notation 2.1.*

(1) *We denote by Σ the collection of operators $\mathscr{B}$ such that*

$$\Sigma:=\{\mathscr{B}: |\mathscr{B}|_{\ell^\infty}\le 1\}.$$

(2) *We denote the* Training Ground $\mathbb{T}$ *by*

$$\mathbb{T}:=\mathrm{cl}(\mathbb{R}^+)\times\Sigma.$$

We state below the definition of training scheme $\mathcal{T}$ and associated notations.

DEFINITION 4.1. *We define the training scheme $\mathcal{T}$ with underlying training ground $\mathbb{T}$ by*

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

$$\textit{Level 2.}\quad u_{\alpha,\mathscr{B}}:=\arg\min\left\{\|u-u_\eta\|^2_{L^2(Q)}+\alpha PV_{\mathscr{B}}(u),\ u\in L^1(Q)\right\}.\qquad(\mathcal{T}\text{-L2})$$

In particular, for the case that $\alpha=+\infty$, we define

$$u_{+\infty}:=\arg\min\left\{\|u-u_c\|^2_{L^2(Q)}:u\in\mathcal{N}\right\}\ \textit{where}\ \mathcal{N}:=\mathrm{conv}\left(\bigcup_{\mathscr{B}\in\Sigma}\mathcal{N}(\mathscr{B})\right).\qquad(4.1)$$

Here conv *denotes the convex envelope. In ($\mathcal{T}$-L1), we denoted by $\mathbb{A}[\mathbb{T}]$ the collection of optimal solution(s) of $\mathcal{T}$ with underlying training ground $\mathbb{T}$, and $(\alpha_{\mathbb{T}},\mathscr{B}_{\mathbb{T}})\in\mathbb{A}[\mathbb{T}]$ is an optimal solution obtained from the training ground $\mathbb{T}$.*

We first show that the Level 2 problem ($\mathcal{T}$-L2) admits a unique solution.

PROPOSITION 4.1. *Let $\alpha\in\mathbb{R}^+$ and $\mathscr{B}\in\Sigma$ be given. Then, there exists a unique $u_{\alpha,\mathscr{B}}\in BV_{\mathscr{B}}(Q)$ such that*

$$u_{\alpha,\mathscr{B}}=\arg\min\left\{\|u-u_\eta\|^2_{L^2(Q)}+\alpha PV_{\mathscr{B}}(u):u\in L^1(Q)\right\}.$$

Proof. The proof can be obtained by Proposition 2.2 and the fact that $PV_{\mathscr{B}}$ is convex. □

THEOREM 4.1. *Let the training ground $\mathbb{T}$ be given. Then the training scheme $\mathcal{T}$ admits at least one solution $(\alpha_{\mathbb{T}},\mathscr{B}_{\mathbb{T}})\in\mathbb{T}$, and provides an associated optimally reconstructed image $u_{\alpha_{\mathbb{T}},\mathscr{B}_{\mathbb{T}}}\in BV_{\mathscr{B}_{\mathbb{T}}}(Q)$.*

Proof. Let $\{\alpha_n,\mathscr{B}_n\}_{n=1}^\infty\subset\mathbb{T}$ be a minimizing sequence obtained from ($\mathcal{T}$-L1). Then, by the boundedness and closedness of Σ in ℓ^∞, up to a (not relabeled) subsequence, there exists $(\alpha_{\mathbb{T}},\mathscr{B}_{\mathbb{T}})\in\mathrm{cl}(\mathbb{R}^+)\times\Sigma$ such that $\alpha_n\to\alpha_{\mathbb{T}}$ in $\mathbb{R}$, $\mathscr{B}_n\to\mathscr{B}_{\mathbb{T}}$ in ℓ^∞, and

$$\lim_{n\to\infty}\|u_c-u_{\alpha_n,\mathscr{B}_n}\|^2_{L^2(Q)}\to m:=\inf\left\{\|u_c-u_{\alpha,\mathscr{B}}\|^2_{L^2(Q)}:(\alpha,\mathscr{B})\in\mathbb{T}\right\}.\qquad(4.2)$$

We divide our arguments into three cases.

Case 1: $\alpha_{\mathbb{T}}>0$. By Theorem 3.1 and the properties of Γ-convergence, we have

$$u_{\alpha_n,\mathscr{B}_n} \rightharpoonup u_{\alpha_{\mathbb{T}},\mathscr{B}_{\mathbb{T}}} \text{ weakly in } L^2(Q), \tag{4.3}$$

where $u_{\alpha_n,\mathscr{B}_n}$ and $u_{\alpha_{\mathbb{T}},\mathscr{B}_{\mathbb{T}}}$ are obtained from ($\mathcal{T}$-L2). Thus, we deduce that

$$\|u_{\alpha_{\mathbb{T}},\mathscr{B}_{\mathbb{T}}}-u_c\|_{L^2(Q)} \le \liminf_{n\to\infty} \|u_{\alpha_n,\mathscr{B}_n}-u_c\|_{L^2(Q)}=m,$$

as desired.

Case 2: $\alpha_{\mathbb{T}}=0$. Then by (4.2), up to a subsequence, there exists $\bar{u}\in L^2(Q)$ such that $u_{\alpha_n,\mathscr{B}_n}\rightharpoonup \bar{u}$ weakly in $L^2(Q)$. We claim that $u_{\alpha_n,\mathscr{B}_n}\to u_\eta$ strongly in $L^2(Q)$. Extend u_η by zero outside Q, and define

$$u_\eta^\varepsilon := u_\eta * \eta_\varepsilon$$

where η_ε is some suitable mollifier, whose particular expression is not relevant. For instance, a possible choice for such η_ε can be a Gaussian distribution with mean zero and variance $1/\varepsilon^2$. Then we have $u_\eta^\varepsilon\in C_c^\infty(\mathbb{R}^N)$ and $u_\eta^\varepsilon\to u_\eta$ strongly in $L^2(\mathbb{R}^N)$. By the optimality condition of ($\mathcal{T}$-L2), we have

$$\begin{aligned}\|u_{\alpha_n,\mathscr{B}_n}-u_\eta\|^2_{L^2(Q)}&+\alpha_n PV_{\mathscr{B}_n}(u_{\alpha_n,\mathscr{B}_n})\\ &\le \left\|u_\eta^\varepsilon-u_\eta\right\|^2_{L^2(Q)}+\alpha_n PV_{\mathscr{B}_n}(u_\eta^\varepsilon)\\ &\le \left\|u_\eta^\varepsilon-u_\eta\right\|^2_{L^2(Q)}+\alpha_n N^d\left\|u_\eta^\varepsilon\right\|_{W^{d,1}(\mathbb{R}^N)}.\end{aligned}$$

That is, we have

$$\|u_{\alpha_n,\mathscr{B}_n}-u_\eta\|^2_{L^2(Q)} \le \left\|u_\eta^\varepsilon-u_\eta\right\|^2_{L^2(Q)}+\alpha_n N^d\left\|u_\eta^\varepsilon\right\|_{W^{d,1}(\mathbb{R}^N)},$$

and we conclude by letting $\alpha_n\to 0$ and $\varepsilon\to 0$.

Case 3: $\alpha_{\mathbb{T}}=+\infty$. By arguing as in Case 2, we have again the existence of $\bar{u}\in L^2(Q)$ such that

$$u_{\alpha_n,\mathscr{B}_n}\rightharpoonup \bar{u} \text{ and } PV_{\mathscr{B}_{\mathbb{T}}}(\bar{u})=0.$$

Then, by (4.1), we have

$$m=\liminf_{n\to\infty}\|u_{\alpha_n,\mathscr{B}_n}-u_c\|_{L^2(Q)} \ge \|\bar{u}-u_c\|_{L^\infty(Q)} \ge \|u_{+\infty}-u_c\|_{L^2(Q)},$$

as desired. ∎

4.2. Numerical realization and finite approximation of scheme $\mathcal{T}$. For the numerical realization of training scheme $\mathcal{T}$, we impose the extra requirement that the training ground $\mathbb{T}$ satisfies also the following assumption.

ASSUMPTION 4.1. *Let the order $d\in\mathbb{N}$ be given.*

(1) *We assume the intensity parameter α satisfies the* box-constraint *(see e.g. [2, 11]). That is, there exists a constant $P\in\mathbb{R}^+$, chosen by the user, such that $\alpha\in[0,P]$.*

(2) *We assume the collection Σ of operator $\mathscr{B}$ satisfies the following two conditions.*

(a) *Each operator $\mathscr{B}\in\Sigma$ has at most order d (i.e., a box-constraint on the order of $\mathscr{B}$);*

(b) *For any $\mathscr{B}_1$, $\mathscr{B}_2\in\Sigma$, the continuity assumptions*

$$|PV_{\mathscr{B}_1}(u)-PV_{\mathscr{B}_2}(u)|\leq O(|\mathscr{B}_1-\mathscr{B}_2|_{\ell^\infty})PV_{\mathscr{B}_1}(u) \tag{4.4}$$

and

$$|PV_{\mathscr{B}_1}(u)-PV_{\mathscr{B}_2}(u)|\leq O(|\mathscr{B}_1-\mathscr{B}_2|_{\ell^\infty})PV_{\mathscr{B}_2}(u) \tag{4.5}$$

hold.

The following corollary is a direct consequence of Theorem 4.1.

Corollary 4.1. *The training scheme $\mathcal{T}$, with an underlying training ground $\mathbb{T}$, satisfies Assumption (4.1). It admits at least one solution $(\alpha_{\mathbb{T}},\mathscr{B}_{\mathbb{T}})\in\mathbb{T}$, and provides an associated optimally reconstructed image $u_{\alpha_{\mathbb{T}},\mathscr{B}_{\mathbb{T}}}\in BV_{\mathscr{B}_{\mathbb{T}}}(Q)$.*

Proof. The proof is identical to that used for Cases 1 and 2 in Theorem 4.1. □

Recall the definition of the *assessment operator* from (1.4) that

$$\mathcal{A}(\alpha,\mathscr{B}):=\|u_c-u_{\alpha,\mathscr{B}}\|^2_{L^2(Q)},\ \text{for } (\alpha,\mathscr{B})\in\mathbb{T}.$$

As discussed in Section 1, the Level 1 problem ($\mathcal{T}$-L1) for scheme $\mathcal{T}$ is equivalent to finding global minimizers of $\mathcal{A}(\alpha,\mathscr{B})$ in the training ground $\mathbb{T}$. However, in view of the counter-example provided in [23], the assessment function $\mathcal{A}(\cdot)$ is not convex, and hence the traditional methods like *Newton's descent* or *Line search* could be trapped into local minima.

We overcome this issue by using a finite approximation method originally introduced in [23]. Recall the constant $P>0$ given in the box-constraint, as stated in Assumption 4.1.

Definition 4.2 (The *Finite Training Ground* and *Finite Grid*). *Let $l\in\mathbb{N}$ be given.*

(1) *We define the* step size *δ_l by*

$$\delta_l:=P/l.$$

(2) *We define the finite set*

$$\mathbb{T}_l[P]\subset[0,P],\qquad \mathbb{T}_l[P]:=\{0,\,\delta_l,\,2\delta_l,\,\dots,\,i\delta_l,\dots,P\}.$$

(3) *We define the finite set*

$$\mathbb{T}_l[\Sigma]\subset\Sigma,\qquad \mathbb{T}_l[\Sigma]:=\bigcup_{k\geq1}T_k[\Sigma]$$

where each $T_k[\Sigma]$ is a singleton containing one operator $\mathscr{B}\in\Sigma$ and defined recursively in the following steps.

Step 1. *Define*

$$\mathscr{B}_0\in\arg\min\{\|\mathscr{B}\|:\mathscr{B}\in\Sigma\},\,T_0[\Sigma]=\{\mathscr{B}_0\},\ \textit{and}\ \Sigma_0:=\Sigma.$$

We also denote by $Q_l[\mathscr{B}_0]\subset\Sigma_0$ the cube centered at $\mathscr{B}_0$ with side length Δ_l.

Step 2. *Define*

$$\Sigma_1 := \Sigma_0 \setminus Q_l[\mathscr{B}_0],\ \mathscr{B}_1 \in \arg\min\{|\mathscr{B}| : \mathscr{B} \in \Sigma_1\},$$

and

$$T_1[\Sigma] := \{\mathscr{B}_1\};$$

$\vdots$

Step j. Define

$$\Sigma_k := \Sigma_{k-1} \setminus Q_l[\mathscr{B}_{k-1}],\ \mathscr{B}_j \in \arg\min\{\|\mathscr{B}\| : \mathscr{B} \in \Sigma_j\},$$

and

$$T_k[\Sigma] := \{\mathscr{B}_j\}.$$

Repeat until $\Sigma_k = \emptyset$.

(4) *Define the* Finite Training Ground $\mathbb{T}_l$ *at step $l \in \mathbb{N}$ by*

$$\mathbb{T}_l := \mathbb{T}_l[\alpha] \times \mathbb{T}_l[\Sigma].$$

(5) *For $i,\ j \in \mathbb{N}$, we define the (i,j)-th* Finite Grid *at step l by*

$$\mathbb{G}_l(i,j) := [i\Delta_l, (i+1)\Delta_l] \times Q_l[\mathscr{B}_j]. \tag{4.6}$$

Remark 4.1. From the definition of Σ_k and Q_l, we have, for fixed $l \in \mathbb{N}$, the existence of an upper bound $M \in \mathbb{N}$, depending on l, such that $\Sigma_M = \emptyset$. That is, we have $\mathcal{H}^0(\mathbb{T}_l[\mathscr{B}]) < +\infty$, and hence

$$\mathcal{H}^0(\mathbb{T}_l) < +\infty, \text{ for each } l \in \mathbb{N} \text{ fixed.}$$

Then, the optimal parameters of scheme $\mathcal{T}$ (global minimizers of $\mathcal{A}(\cdot,\cdot)$) over finite training ground $\mathbb{T}_l$

$$(\alpha_{\mathbb{T}_l}, \mathscr{B}_{\mathbb{T}_l}) \in \mathbb{A}[\mathbb{T}_l] := \arg\min\left\{\|u_{\alpha,\mathscr{B}} - u_c\|^2_{L^2(Q)} : (\alpha,\mathscr{B}) \in \mathbb{T}_l\right\},$$

can be determined exactly by evaluating $\mathcal{A}(\cdot)$ over each element of $\mathbb{T}_l$.

The main result of Section 4.2 reads as follows.

Theorem 4.2 (Finite approximation and error estimation). *Let a training ground $\mathbb{T}$ that satisfies Assumption 4.1 be given, and $\mathbb{T}_l \subset \mathbb{T}$ be constructed as in Definition 4.2. Then the following assertions hold.*

(1) *First,*

$$\lim_{l\to\infty} \operatorname{dist}(\mathbb{A}[\mathbb{T}], \mathbb{A}[\mathbb{T}_l]) = 0. \tag{4.7}$$

(2) *Second, let $\delta > 0$ be given. Then, for each $l \in \mathbb{N}$,*

$$\mathcal{A}(\alpha_{\mathbb{T}_l}, \mathscr{B}_{\mathbb{T}_l}) - \mathcal{A}(\alpha_{\mathbb{T}}, \mathscr{B}_{\mathbb{T}}) \le 4KP[O(P/l) + 1/l]^{1/2} \|u_\eta\|^{1/2}_{W^{d,1}(Q)}/\delta^d + \delta/2, \tag{4.8}$$

for any $(\alpha_{\mathbb{T}}, \mathscr{B}_{\mathbb{T}}) \in \mathbb{A}[\mathbb{T}]$ and $(\alpha_{\mathbb{T}_l}, \mathscr{B}_{\mathbb{T}_l}) \in \mathbb{A}[\mathbb{T}_l]$.

We split our argument into Sections 4.2.1 and 4.2.2, where we will discuss the properties of the reconstructed image $u_{\alpha,\mathscr{B}}$ with fixed $\mathscr{B} \in \Sigma$ and $\alpha \in \mathbb{R}^+$, respectively.

4.2.1. Properties of reconstructed image $u_{\alpha,\mathscr{B}}$ with respect to $\alpha \in \mathbb{R}^+$. Since $\mathscr{B} \in \Sigma$ is fixed, we abbreviate $u_{\alpha,\mathscr{B}}$ and $PV_{\mathscr{B}}$ by u_α and PV, respectively, in Section 4.2.1.

PROPOSITION 4.2. *We collect two auxiliary results in this proposition.*

(1) *The function $g(\alpha) := PV(u_\alpha)$ is continuous and decreasing.*

(2) *Assume in addition that*

$$PV(u_\eta) > PV(u_c). \tag{4.9}$$

Then, there exists $\alpha > 0$ such that

$$\|u_\alpha - u_c\|_{L^2(Q)} < \|u_\eta - u_c\|_{L^2(Q)}. \tag{4.10}$$

Proof. We show Assertion 1 first. The continuity of $g(\alpha)$ can be deduced from Theorem 3.1. Next, let $0 \le \alpha_1 < \alpha_2 < +\infty$ be given, we observe, from the optimality condition of ($\mathcal{T}$-L2), that

$$\|u_{\alpha_1} - u_\eta\|^2_{L^2(Q)} + \alpha_1 PV(u_{\alpha_1}) \le \|u_{\alpha_2} - u_\eta\|^2_{L^2(Q)} + \alpha_1 PV(u_{\alpha_2})$$

and

$$\|u_{\alpha_2} - u_\eta\|^2_{L^2(Q)} + \alpha_2 PV(u_{\alpha_2}) \le \|u_{\alpha_1} - u_\eta\|^2_{L^2(Q)} + \alpha_2 PV(u_{\alpha_1}).$$

Adding up the previous two inequalities yields

$$\alpha_1 PV(u_{\alpha_1}) + \alpha_2 PV(u_{\alpha_2}) \le \alpha_1 PV(u_{\alpha_2}) + \alpha_2 PV(u_{\alpha_1}),$$

which implies that $PV(u_{\alpha_2}) \le PV(u_{\alpha_1})$, as desired.

Now we claim Assertion 2. From Theorem 3.2, we have that $\partial PV(u_\alpha)$, the sub-differential of PV at u_α, is well defined. We observe that, for any $\alpha > 0$,

$$\begin{aligned}
&\|u_\eta - u_c\|^2_{L^2(Q)} - \|u_\alpha - u_c\|^2_{L^2(Q)} \\
&\quad = 2\langle u_\eta - u_\alpha, u_\alpha - u_c\rangle + \|u_\eta - u_\alpha\|^2_{L^2(Q)} \\
&\quad = 2\alpha\langle \partial PV(u_\alpha), u_\alpha - u_c\rangle + \|u_\eta - u_\alpha\|^2_{L^2(Q)} \\
&\quad = 2\alpha\langle \partial PV(u_\alpha), u_\alpha\rangle - 2\alpha\langle \partial PV(u_\alpha), u_c\rangle + \|u_\eta - u_\alpha\|^2_{L^2(Q)} \\
&\quad \ge 2\alpha[PV(u_\alpha) - PV(u_c)] + \|u_\eta - u_\alpha\|^2_{L^2(Q)},
\end{aligned}$$

where, at the last inequality we used the definition of sub-differential. Thus

$$\|u_\eta - u_c\|^2_{L^2(Q)} - \|u_\alpha - u_c\|^2_{L^2(Q)} \ge 2\alpha[PV(u_\alpha) - PV(u_c)] + \|u_\eta - u_\alpha\|^2_{L^2(Q)}. \tag{4.11}$$

Next, in view of Assertion 1, we have that $PV(u_\alpha)$ is continuous and decreasing. Hence, combined with (4.9), we infer the existence of $\bar{\alpha} > 0$ such that

$$PV(u_{\bar{\alpha}}) - PV(u_c) \ge \frac{1}{4}[PV(u_\eta) - PV(u_c)] > 0. \tag{4.12}$$

Hence, we conclude (4.10) by combining (4.11) and (4.12). $\square$

PROPOSITION 4.3. *Let α_1 and $\alpha_2 \in \mathbb{R}^+$ be given. Then we have*

$$\|u_{\alpha_1} - u_{\alpha_2}\|^2_{L^2(Q)} \leq |\alpha_1 - \alpha_2| (PV(u_{\alpha_1}) + PV(u_{\alpha_2})).$$

Proof. Without loss of generality, we assume that $\alpha_1 < \alpha_2$. In view of Theorem 3.2, and the optimality condition of ($\mathcal{T}$-L2), we have

$$u_{\alpha_1} - u_\eta = -\alpha_1 \partial PV(u_{\alpha_1}) \text{ and } u_{\alpha_2} - u_\eta = -\alpha_2 \partial PV(u_{\alpha_2}).$$

Subtracting one from another, then multiplying by $u_{\alpha_1} - u_{\alpha_2}$, and finally integrating over Q, gives

$$\begin{aligned}\|u_{\alpha_1} - u_{\alpha_2}\|^2_{L^2(Q)} = \alpha_1 \langle \partial PV(u_{\alpha_2}) - \partial PV(u_{\alpha_1}), u_{\alpha_1} - u_{\alpha_2} \rangle \\ + (\alpha_2 - \alpha_1) \langle \partial PV(u_{\alpha_2}), u_{\alpha_1} - u_{\alpha_2} \rangle. \end{aligned} \tag{4.13}$$

Since the seminorm PV is proper, lower semi-continuous, and convex, ∂PV is a monotone maximal operator, thus

$$\langle \partial PV(u_{\alpha_2}) - \partial PV(u_{\alpha_1}), u_{\alpha_2} - u_{\alpha_1} \rangle \geq 0.$$

Combining with (4.13) and Assertion 1 from Proposition 4.2 gives

$$\begin{aligned}\|u_{\alpha_1} - u_{\alpha_2}\|^2_{L^2(Q)} &\leq (\alpha_2 - \alpha_1) \langle \partial PV(u_{\alpha_2}), u_{\alpha_1} - u_{\alpha_2} \rangle \\ &\leq (\alpha_2 - \alpha_1) PV(u_{\alpha_1} - u_{\alpha_2}) \leq (\alpha_2 - \alpha_1) [PV(u_{\alpha_1}) + PV(u_{\alpha_2})],\end{aligned}$$

where, at the second last inequality we used Remark 3.1, concluding the proof. ∎

4.2.2. Properties of the reconstructed image $u_{\alpha,\mathscr{B}}$ with respect to $\mathscr{B} \in \Sigma$. Similarly to Section 4.2.1, we still abbreviate $u_{\alpha,\mathscr{B}}$ by $u_{\mathscr{B}}$, for $\alpha \in \mathbb{R}^+$ fixed. Recall the structure of $\mathscr{B}$ from Notation 2.1.

Moreover, we further restrict the corrupted image $u_\eta \in L^2(Q)$, by requiring it to satisfy the following additional assumption: There must exist $0 < M_1 < M_2 < +\infty$ such that

$$0 < M_1 \leq u_\eta(x) \leq M_2 < +\infty, \text{ for a.e. } x \in Q. \tag{4.14}$$

In this way, we have that the reconstructed image

$$u_{\mathscr{B}} = \arg\min \left\{ \|u - u_\eta\|^2_{L^2(Q)} + PV_{\mathscr{B}}(u) \colon u \in L^1(Q) \right\}$$

also satisfies that

$$M_1 \leq u_{\mathscr{B}}(x) \leq M_2, \text{ for a.e. } x \in Q. \tag{4.15}$$

We recall the following result regarding Lebesgue points.

THEOREM 4.3 (Lebesgue-Besicovitch differentiation theorem). *Let μ be a Radon measure on $\mathbb{R}^N$ and $f \in L^1_{\text{loc}}(\mathbb{R}^N, \mu)$. Then*

$$\lim_{r \to 0} \fint_{B(x,r)} f \, d\mu = f(x)$$

for μ a.e. $x \in \mathbb{R}^N$.

PROPOSITION 4.4. *Given $u_\eta \in L^2(Q)$ satisfying (4.14), and operators $\mathscr{B}_1$ and $\mathscr{B}_2 \in \Sigma$, it holds*

$$\|u_{\mathscr{B}_1} - u_{\mathscr{B}_2}\|^2_{L^2(Q)} \le O(|\mathscr{B}_1 - \mathscr{B}_2|_{\ell^\infty})[PV_{\mathscr{B}_1}(u_{\mathscr{B}_1}) + PV_{\mathscr{B}_1}(u_{\mathscr{B}_2})],$$

where $u_{\mathscr{B}}$ is defined in ($\mathcal{T}$-L2).

Proof. By Theorem 3.2, the sub-differentials $\partial PV_{\mathscr{B}_1}$ and $\partial PV_{\mathscr{B}_2}$ are well defined. Then, by the optimality condition of ($\mathcal{T}$-L2) we have that

$$u_{\mathscr{B}_1} - u_\eta = -\partial PV_{\mathscr{B}_1}(u_{\mathscr{B}_1}) \text{ and } u_{\mathscr{B}_2} - u_\eta = -\partial PV_{\mathscr{B}_2}(u_{\mathscr{B}_2}),$$

Subtracting one from another gives

$$\begin{aligned} u_{\mathscr{B}_1} - u_{\mathscr{B}_2} &= \partial PV_{\mathscr{B}_2}(u_{\mathscr{B}_2}) - \partial PV_{\mathscr{B}_1}(u_{\mathscr{B}_1}) \\ &= \partial PV_{\mathscr{B}_2}(u_{\mathscr{B}_2}) - \partial PV_{\mathscr{B}_2}(u_{\mathscr{B}_1}) + \partial PV_{\mathscr{B}_2}(u_{\mathscr{B}_1}) - \partial PV_{\mathscr{B}_1}(u_{\mathscr{B}_1}). \end{aligned}$$

Multiplying both sides by $u_{\mathscr{B}_1} - u_{\mathscr{B}_2}$ and integrating over Q gives

$$\begin{aligned} \|u_{\mathscr{B}_2} - u_{\mathscr{B}_1}\|^2_{L^2(Q)} = &-\langle \partial PV_{\mathscr{B}_2}(u_{\mathscr{B}_2}) - \partial PV_{\mathscr{B}_2}(u_{\mathscr{B}_1}), u_{\mathscr{B}_2} - u_{\mathscr{B}_1}\rangle \\ &+ \langle \partial PV_{\mathscr{B}_2}(u_{\mathscr{B}_1}) - \partial PV_{\mathscr{B}_1}(u_{\mathscr{B}_1}), u_{\mathscr{B}_1} - u_{\mathscr{B}_2}\rangle. \end{aligned} \tag{4.16}$$

Since $PV_{\mathscr{B}}$ is convex, $\partial PV_{\mathscr{B}}$ is a monotone maximal operator. Therefore,

$$\langle \partial PV_{\mathscr{B}_2}(u_{\mathscr{B}_2}) - \partial PV_{\mathscr{B}_2}(u_{\mathscr{B}_1}), u_{\mathscr{B}_2} - u_{\mathscr{B}_1}\rangle \ge 0. \tag{4.17}$$

We next estimate the second part of (4.16). First, by the definition of sub-gradient, we have

$$\langle \partial PV_{\mathscr{B}_2}(u_{\mathscr{B}_1}) - \partial PV_{\mathscr{B}_1}(u_{\mathscr{B}_1}), u_{\mathscr{B}_1}\rangle = PV_{\mathscr{B}_2}(u_{\mathscr{B}_1}) - PV_{\mathscr{B}_1}(u_{\mathscr{B}_1}) \le cPV_{\mathscr{B}_1}(u_{\mathscr{B}_1}). \tag{4.18}$$

where

$$c = O(|\mathscr{B}_1 - \mathscr{B}_2|_{\ell^\infty})$$

is the constant used in (4.4). Moreover, from (4.4) we also deduce that

$$-c|\mathscr{B}_2 u_{\mathscr{B}_1}| \le |\mathscr{B}_2 u_{\mathscr{B}_1}|(V) - |\mathscr{B}_1 u_{\mathscr{B}_1}|(V) \le c|\mathscr{B}_2 u_{\mathscr{B}_1}|. \tag{4.19}$$

Next, let $v_{\mathscr{B}_1}$ and $v_{\mathscr{B}_2}$ be obtained from Proposition 3.2 as sub-gradients of $PV_{\mathscr{B}_1}(u_{\mathscr{B}_1})$ and $PV_{\mathscr{B}_2}(u_{\mathscr{B}_1})$, respectively. Then, by Proposition 3.5, for any open set $V \subset Q$ we have that

$$|\mathscr{B}_2 u_{\mathscr{B}_1}|(V) = \int_V u_{\mathscr{B}_1}[\mathscr{B}_2^* v_{\mathscr{B}_2}]\,dx \text{ and } |\mathscr{B}_1 u_{\mathscr{B}_1}|(V) = \int_V u_{\mathscr{B}_1}[\mathscr{B}_1^* v_{\mathscr{B}_1}]\,dx.$$

Combining with (4.19) gives

$$-c\int_V u_{\mathscr{B}_1}[\mathscr{B}_2^* v_{\mathscr{B}_2}]\,dx \le \int_V u_{\mathscr{B}_1}[\mathscr{B}_2^* v_{\mathscr{B}_2}]\,dx - \int_V u_{\mathscr{B}_1}[\mathscr{B}_1^* v_{\mathscr{B}_1}]\,dx \le c\int_V u_{\mathscr{B}_1}[\mathscr{B}_1^* v_{\mathscr{B}_1}]\,dx.$$

Thus, we could further write, by taking $Q(x,\delta) := [x-\delta, x+\delta]^N$, a cube centered at x with side length 2δ, that

$$-c\fint_{Q(x,\delta)} u_{\mathscr{B}_1}[\mathscr{B}_2^* v_{\mathscr{B}_2}]\,dx \le \fint_{Q(x,\delta)} (u_{\mathscr{B}_1}[\mathscr{B}_2^* v_{\mathscr{B}_2}] - u_{\mathscr{B}_1}[\mathscr{B}_1^* v_{\mathscr{B}_1}])\,dx$$

$$= \fint_{Q(x,\delta)} u_{\mathscr{B}_1}[\mathscr{B}_2^* v_{\mathscr{B}_2}]\,dx - \fint_{Q(x,\delta)} u_{\mathscr{B}_1}[\mathscr{B}_1^* v_{\mathscr{B}_1}]\,dx$$
$$\le c \fint_{Q(x,\delta)} u_{\mathscr{B}_1}[\mathscr{B}_1^* v_{\mathscr{B}_1}]\,dx.$$

By Assertion 2 of Theorem 3.2, we have $\mathscr{B}_1^* v_{\mathscr{B}_1} \in L^1(Q)$. Since $u_{\mathscr{B}_1} \in L^\infty(Q)$, we have $u_{\mathscr{B}_1}\mathscr{B}_1^* v_{\mathscr{B}_1} \in L^1(Q)$. Thus, we could apply Theorem 4.3 and take $\delta \to 0$ to conclude that

$$-cu_{\mathscr{B}_1}[\mathscr{B}_2^* v_{\mathscr{B}_2}] \le u_{\mathscr{B}_1}[\mathscr{B}_2^* v_{\mathscr{B}_2}] - u_{\mathscr{B}_1}[\mathscr{B}_1^* v_{\mathscr{B}_1}] \le cu_{\mathscr{B}_1}[\mathscr{B}_1^* v_{\mathscr{B}_1}],$$

for a.e. $x \in Q$. That is, we have

$$-cu_{\mathscr{B}_1}[\mathscr{B}_2^* v_{\mathscr{B}_2}] \le u_{\mathscr{B}_1}[\mathscr{B}_2^* v_{\mathscr{B}_2} - \mathscr{B}_1^* v_{\mathscr{B}_1}] \le cu_{\mathscr{B}_1}[\mathscr{B}_1^* v_{\mathscr{B}_1}],$$

and, combined with the fact that $u_{\mathscr{B}_1} \ge 1$ (see (4.15)), we deduce that

$$-c[\mathscr{B}_2^* v_{\mathscr{B}_2}] \le [\mathscr{B}_2^* v_{\mathscr{B}_2} - \mathscr{B}_1^* v_{\mathscr{B}_1}] \le c[\mathscr{B}_1^* v_{\mathscr{B}_1}], \tag{4.20}$$

for a.e. $x \in Q$.

On the other hand, again by (4.15), we have $-u_{\mathscr{B}_1} + 2M_2 > 1$, and hence

$$\begin{aligned}
&\langle \partial PV_{\mathscr{B}_2}(u_{\mathscr{B}_1}) - \partial PV_{\mathscr{B}_1}(u_{\mathscr{B}_1}), -u_{\mathscr{B}_2}\rangle \\
&= \langle \partial PV_{\mathscr{B}_2}(u_{\mathscr{B}_1}) - \partial PV_{\mathscr{B}_1}(u_{\mathscr{B}_1}), -u_{\mathscr{B}_2} + 2M_2 - 2M_2\rangle \\
&= \langle \partial PV_{\mathscr{B}_2}(u_{\mathscr{B}_1}) - \partial PV_{\mathscr{B}_1}(u_{\mathscr{B}_1}), -u_{\mathscr{B}_2} + 2M_2\rangle + \langle \partial PV_{\mathscr{B}_2}(u_{\mathscr{B}_1}) - \partial PV_{\mathscr{B}_1}(u_{\mathscr{B}_1}), -2M_2\rangle \\
&\le c\langle [\mathscr{B}_1^* v_{\mathscr{B}_1}], -u_{\mathscr{B}_2} + 2M_2\rangle + \langle \partial PV_{\mathscr{B}_2}(u_{\mathscr{B}_1}) - \partial PV_{\mathscr{B}_1}(u_{\mathscr{B}_1}), -2M_2\rangle.
\end{aligned} \tag{4.21}$$

Note that, as $v_{\mathscr{B}_1} \in W_0^p[\mathscr{B}_1](Q)$, by Remark 3.1 we observe that

$$\langle \mathscr{B}_1^* v_{\mathscr{B}_1}, -u_{\mathscr{B}_2} + 2M_2\rangle \le PV_{\mathscr{B}_1}(-u_{\mathscr{B}_2} + 2M_2) = PV_{\mathscr{B}_1}(u_{\mathscr{B}_2}), \tag{4.22}$$

and, since constants belong to the kernel of $PV_{\mathscr{B}_2}$,

$$\langle \partial PV_{\mathscr{B}_2}(u_{\mathscr{B}_1}) - \partial PV_{\mathscr{B}_1}(u_{\mathscr{B}_1}), -2M_2\rangle = 0. \tag{4.23}$$

Therefore, by combining (4.21), (4.22), and (4.23), we obtain that

$$\langle \partial PV_{\mathscr{B}_2}(u_{\mathscr{B}_1}) - \partial PV_{\mathscr{B}_1}(u_{\mathscr{B}_1}), -u_{\mathscr{B}_2}\rangle \le cPV_{\mathscr{B}_1}(u_{\mathscr{B}_2}).$$

Combining with (4.16), (4.17), and (4.18), concludes the proof. ∎

4.2.3. L^2-distance estimate of the reconstructed image $u_{\alpha,\mathscr{B}}$. We start with a relaxation result regarding the corrupted image u_η.

PROPOSITION 4.5. *Let $u_\eta \in L^2(Q)$ be given. Let $\{u_\eta^\varepsilon\}_{\varepsilon>0} \subset L^2(Q)$ such that $u_\eta^\varepsilon \to u_\eta$ strongly in L^2. For arbitrary $(\alpha, \mathscr{B}) \in \mathbb{T}$, define*

$$u_{\alpha,\mathscr{B}}^\varepsilon := \operatorname{arg\,min}\left\{ \left\| u - u_\eta^\varepsilon \right\|_{L^2}^2 + \alpha PV_{\mathscr{B}}(u) : u \in L^1(Q) \right\}. \tag{4.24}$$

Then we have

$$\left\| u_{\alpha,\mathscr{B}} - u_{\alpha,\mathscr{B}}^\varepsilon \right\|_{L^2(Q)} \le \left\| u_\eta^\varepsilon - u_\eta \right\|_{L^2(Q)} \tag{4.25}$$

and

$$\lim_{\varepsilon\to 0} PV_{\mathscr{B}}(u^{\varepsilon}_{\alpha,\mathscr{B}}) = PV_{\mathscr{B}}(u_{\alpha,\mathscr{B}}), \tag{4.26}$$

where $u_{\alpha,\mathscr{B}}$ *is defined in* ($\mathcal{T}$-L2).

Proof. From the optimality condition of (4.24) and ($\mathcal{T}$-L2), we have

$$u_{\alpha,\mathscr{B}} - u^{\varepsilon}_{\alpha,\mathscr{B}} + u^{\varepsilon}_{\eta} - u_{\varepsilon} = \alpha\partial PV_{\mathscr{B}}(u^{\varepsilon}_{\alpha,\mathscr{B}}) - \alpha\partial PV_{\mathscr{B}}(u_{\alpha,\mathscr{B}}).$$

Multiplying by $u_{\alpha,\mathscr{B}} - u^{\varepsilon}_{\alpha,\mathscr{B}}$ on both the sides gives

$$\begin{aligned}\left\|u_{\alpha,\mathscr{B}} - u^{\varepsilon}_{\alpha,\mathscr{B}}\right\|^2_{L^2(Q)} &+ \left\langle u^{\varepsilon}_{\eta} - u_{\varepsilon}, u_{\alpha,\mathscr{B}} - u^{\varepsilon}_{\alpha,\mathscr{B}}\right\rangle \\ &= \alpha\left\langle \partial PV_{\mathscr{B}}(u^{\varepsilon}_{\alpha,\mathscr{B}}) - \partial PV_{\mathscr{B}}(u_{\alpha,\mathscr{B}}), u_{\alpha,\mathscr{B}} - u^{\varepsilon}_{\alpha,\mathscr{B}}\right\rangle \le 0,\end{aligned}$$

where, at the last inequality we used the fact that $\partial PV_{\mathscr{B}}$ is a maximal monotone operator, and we conclude (4.25), as desired.

We next show (4.26). We assume that $\alpha\in\mathbb{R}^+$, otherwise there is nothing to prove. By (4.25),

$$u_{\alpha,\mathscr{B}} \to u^{\varepsilon}_{\alpha,\mathscr{B}} \text{ strongly in } L^2(Q). \tag{4.27}$$

Together with Proposition 2.1, we deduce that

$$\liminf_{\varepsilon\to 0} PV_{\mathscr{B}}(u^{\varepsilon}_{\alpha,\mathscr{B}}) \ge PV_{\mathscr{B}}(u_{\alpha,\mathscr{B}}). \tag{4.28}$$

On the other hand, in view of the optimality condition of (4.24) again, we have

$$\left\|u^{\varepsilon}_{\alpha,\mathscr{B}} - u^{\varepsilon}_{\eta}\right\|^2_{L^2(Q)} + \alpha PV_{\mathscr{B}}(u^{\varepsilon}_{\alpha,\mathscr{B}}) \le \left\|u_{\alpha,\mathscr{B}} - u^{\varepsilon}_{\eta}\right\|^2_{L^2(Q)} + \alpha PV_{\mathscr{B}}(u_{\alpha,\mathscr{B}}),$$

i.e.,

$$\alpha PV_{\mathscr{B}}(u^{\varepsilon}_{\alpha,\mathscr{B}}) \le \left\|u_{\alpha,\mathscr{B}} - u^{\varepsilon}_{\eta}\right\|^2_{L^2(Q)} - \left\|u^{\varepsilon}_{\alpha,\mathscr{B}} - u^{\varepsilon}_{\eta}\right\|^2_{L^2(Q)} + \alpha PV_{\mathscr{B}}(u_{\alpha,\mathscr{B}}).$$

Hence, by (4.27), we have that

$$\begin{aligned}\limsup_{\varepsilon\to 0}\ &\alpha PV_{\mathscr{B}}(u^{\varepsilon}_{\alpha,\mathscr{B}}) \\ &\le \limsup_{\varepsilon\to 0}\left[\left\|u_{\alpha,\mathscr{B}} - u^{\varepsilon}_{\eta}\right\|^2_{L^2(Q)} - \left\|u^{\varepsilon}_{\alpha,\mathscr{B}} - u^{\varepsilon}_{\eta}\right\|^2_{L^2(Q)}\right] + \alpha PV_{\mathscr{B}}(u_{\alpha,\mathscr{B}}) \\ &= \alpha PV_{\mathscr{B}}(u_{\alpha,\mathscr{B}}).\end{aligned}$$

Combining with (4.28) allows us to infer (4.26), as desired. ∎

We next show an improved version of Proposition 4.4, in which we remove the requirement that u_η needs to satisfy the boundedness assumption (4.14).

COROLLARY 4.2. *Let* $u_\eta\in L^2(Q)$, $\alpha\in\mathbb{R}^+$, *and* $\mathscr{B}_1$, $\mathscr{B}_2\in\Sigma$ *be given. Then the following estimation holds.*

$$\|u_{\alpha,\mathscr{B}_1} - u_{\alpha,\mathscr{B}_2}\|^2_{L^2(Q)} \le \alpha\cdot O(|\mathscr{B}_1 - \mathscr{B}_2|_{\ell^\infty})\left[PV_{\mathscr{B}_1}(u_{\alpha,\mathscr{B}_1}) + PV_{\mathscr{B}_1}(u_{\alpha,\mathscr{B}_2})\right],$$

where $u_{\alpha,\mathscr{B}}$ *is defined in* ($\mathcal{T}$-L2).

Proof. Let $M \in \mathbb{N}$ be given, and define

$$u_\eta^M(x) := \begin{cases} M, & \text{if } u_\eta(x) \geq M, \\ u_\eta(x), & \text{if } -M \leq u_\eta(x) \leq M, \\ -M, & \text{if } u_\eta(x) \leq -M. \end{cases}$$

Also, define

$$u_{\alpha,\mathscr{B}}^M := \operatorname{arg\,min}\left\{ \left\| u - u_\eta^M \right\|_{L^2(Q)}^2 + \alpha PV_{\mathscr{B}}(u) : u \in L^1(Q) \right\}, \tag{4.29}$$

and

$$\bar{u}_{\alpha,\mathscr{B}}^M := \operatorname{arg\,min}\left\{ \left\| u - (u_\eta^M + 2M) \right\|_{L^2(Q)}^2 + \alpha PV_{\mathscr{B}}(u) : u \in L^1(Q) \right\}. \tag{4.30}$$

We claim that

$$\bar{u}_{\alpha,\mathscr{B}}^M = u_{\alpha,\mathscr{B}}^M + 2M. \tag{4.31}$$

We observe that

$$\begin{aligned} & \left\| \bar{u}_{\alpha,\mathscr{B}}^M - \left(u_\eta^M + 2M\right) \right\|_{L^2(Q)}^2 + \alpha PV_{\mathscr{B}}\left(\bar{u}_{\alpha,\mathscr{B}}^M\right) \\ \leq & \left\| u_{\alpha,\mathscr{B}}^M + 2M - \left(u_\eta^M + 2M\right) \right\|_{L^2(Q)}^2 + \alpha PV_{\mathscr{B}}\left(u_{\alpha,\mathscr{B}}^M + 2M\right) \\ = & \left\| u_{\alpha,\mathscr{B}}^M - u_\eta^M \right\|_{L^2(Q)}^2 + \alpha PV_{\mathscr{B}}\left(u_{\alpha,\mathscr{B}}^M\right) \\ \leq & \left\| \bar{u}_{\alpha,\mathscr{B}}^M - 2M - u_\eta^M \right\|_{L^2(Q)}^2 + \alpha PV_{\mathscr{B}}\left(\bar{u}_{\alpha,\mathscr{B}}^M - 2M\right) \\ = & \left\| \bar{u}_{\alpha,\mathscr{B}}^M - \left(u_\eta^M + 2M\right) \right\|_{L^2(Q)}^2 + \alpha PV_{\mathscr{B}}\left(\bar{u}_{\alpha,\mathscr{B}}^M\right), \end{aligned}$$

where, at the first inequality we used the optimality condition in (4.30), and at the last inequality we used the optimality condition in (4.29). Thus,

$$\begin{aligned} & \left\| \bar{u}_{\alpha,\mathscr{B}}^M - (u_\eta^M + 2M) \right\|_{L^2(Q)}^2 + \alpha PV_{\mathscr{B}}(\bar{u}_{\alpha,\mathscr{B}}^M) \\ = & \left\| u_{\alpha,\mathscr{B}}^M + 2M - (u_\eta^M + 2M) \right\|_{L^2(Q)}^2 + \alpha PV_{\mathscr{B}}(u_{\alpha,\mathscr{B}}^M + 2M), \end{aligned}$$

and we conclude (4.31) due of the uniqueness of the minimizer. Thus,

$$\left\| u_{\alpha,\mathscr{B}_1}^M - u_{\alpha,\mathscr{B}_2}^M \right\|_{L^2(Q)} = \left\| \bar{u}_{\alpha,\mathscr{B}_1}^M - \bar{u}_{\alpha,\mathscr{B}_2}^M \right\|_{L^2(Q)}.$$

Therefore, we could assume, without loss of generality, $u_\eta^M \geq M > 0$. That is, u_η^M satisfies (4.14).

Next, by the optimality condition in (4.29), we have that

$$\frac{1}{\alpha}(u_{\alpha,\mathscr{B}_1}^M - u_\eta) = -\partial PV_{\mathscr{B}_1}(u_{\alpha,\mathscr{B}_1}^M) \text{ and } \frac{1}{\alpha}(u_{\alpha,\mathscr{B}_2}^M - u_\eta) = -\partial PV_{\mathscr{B}_2}(u_{\alpha,\mathscr{B}_2}^M).$$

Following the exact same arguments from Proposition 4.4 (in (4.21) we use $2M$ instead of M_2), we obtain

$$\frac{1}{\alpha}\left\| u_{\alpha,\mathscr{B}_1}^M - u_{\alpha,\mathscr{B}_2}^M \right\|_{L^2(Q)}^2 \leq O\left(|\mathscr{B}_1 - \mathscr{B}_2|_{\ell^\infty}\right)\left[PV_{\mathscr{B}_1}(u_{\alpha,\mathscr{B}_1}^M) + PV_{\mathscr{B}_2}(u_{\alpha,\mathscr{B}_1}^M)\right].$$

Finally, we note that

$$\begin{aligned}&\frac{1}{\alpha}\|u_{\alpha,\mathscr{B}_1}-u_{\alpha,\mathscr{B}_2}\|^2_{L^2(Q)}\\ &\le\frac{1}{\alpha}\|u^M_{\alpha,\mathscr{B}_1}-u^M_{\alpha,\mathscr{B}_2}\|^2_{L^2(Q)}+\frac{1}{\alpha}\|u_{\alpha,\mathscr{B}_1}-u^M_{\alpha,\mathscr{B}_1}\|^2_{L^2(Q)}+\|u_{\alpha,\mathscr{B}_2}-u^M_{\alpha,\mathscr{B}_2}\|^2_{L^2(Q)}\\ &\le O(|\mathscr{B}_1-\mathscr{B}_2|_{\ell^\infty})\left[PV_{\mathscr{B}_1}(u^M_{\alpha,\mathscr{B}_1})+PV_{\mathscr{B}_2}(u^M_{\alpha,\mathscr{B}_1})\right]\\ &\quad+\frac{1}{\alpha}\|u_{\alpha,\mathscr{B}_1}-u^M_{\alpha,\mathscr{B}_1}\|^2_{L^2(Q)}+\|u_{\alpha,\mathscr{B}_2}-u^M_{\alpha,\mathscr{B}_2}\|^2_{L^2(Q)}.\end{aligned}$$

Then, using Proposition 4.5, with u^ε_η replaced by u^M_η, we conclude the proof by sending $M\nearrow+\infty$ in the above inequality. □

PROPOSITION 4.6. *Let $(\alpha_1,\mathscr{B}_1)$ and $(\alpha_2,\mathscr{B}_2)\in\mathbb{T}$ be given. Then we have*

$$\begin{aligned}&\|u_{\alpha_1,\mathscr{B}_1}-u_{\alpha_2,\mathscr{B}_2}\|^2_{L^2(Q)}\\ &\le 4[\alpha_1O(|\mathscr{B}_1-\mathscr{B}_2|_{\ell^\infty})+|\alpha_1-\alpha_2|][PV_{\mathscr{B}_1}(u_{\alpha_1,\mathscr{B}_1})+PV_{\mathscr{B}_2}(u_{\alpha_1,\mathscr{B}_2})].\end{aligned}$$

Proof. Direct computations give

$$\begin{aligned}&\|u_{\alpha_1,\mathscr{B}_1}-u_{\alpha_2,\mathscr{B}_2}\|^2_{L^2(Q)}\le 2\|u_{\alpha_1,\mathscr{B}_1}-u_{\alpha_1,\mathscr{B}_2}\|^2_{L^2(Q)}+2\|u_{\alpha_1,\mathscr{B}_2}-u_{\alpha_2,\mathscr{B}_2}\|^2_{L^2(Q)}\\ &\le 2\alpha_1O(|\mathscr{B}_1-\mathscr{B}_2|_{\ell^\infty})[PV_{\mathscr{B}_1}(u_{\alpha_1,\mathscr{B}_1})+PV_{\mathscr{B}_1}(u_{\alpha_1,\mathscr{B}_2})]+4|\alpha_1-\alpha_2|PV_{\mathscr{B}_2}(u_{\alpha_1,\mathscr{B}_2})\\ &\le[2\alpha_1O(|\mathscr{B}_1-\mathscr{B}_2|_{\ell^\infty})+4|\alpha_1-\alpha_2|][PV_{\mathscr{B}_1}(u_{\alpha_1,\mathscr{B}_1})+PV_{\mathscr{B}_1}(u_{\alpha_1,\mathscr{B}_2})].\end{aligned}\tag{4.32}$$

Moreover, from (4.5), we have

$$|PV_{\mathscr{B}_1}(u_{\alpha_1,\mathscr{B}_2})-PV_{\mathscr{B}_2}(u_{\alpha_1,\mathscr{B}_2})|\le O(|\mathscr{B}_1-\mathscr{B}_2|_{\ell^\infty})PV_{\mathscr{B}_2}(u_{\alpha_1,\mathscr{B}_2}).$$

Combining with (4.32) gives

$$\begin{aligned}&\|u_{\alpha_1,\mathscr{B}_1}-u_{\alpha_2,\mathscr{B}_2}\|^2_{L^2(Q)}\\ &\le 4[2\alpha_1O(|\mathscr{B}_1-\mathscr{B}_2|_{\ell^\infty})+|\alpha_1-\alpha_2|][PV_{\mathscr{B}_1}(u_{\alpha_1,\mathscr{B}_1})+PV_{\mathscr{B}_2}(u_{\alpha_1,\mathscr{B}_2})],\end{aligned}$$

as desired. □

We close this section by proving Theorem 4.2.

Proof. **(Proof of Theorem 4.2.)** Assertion 1 is a direct consequence of Theorem 3.1.

We next claim (4.8). We first assume that $u\in C^\infty(\bar{Q})$. Indeed, for any $(\alpha,\mathscr{B})\in\mathbb{T}$, we could extract a sequence $\{(\alpha_l,\mathscr{B}_l)\}_{l=1}^\infty\subset\mathbb{T}$, where for each $l\in\mathbb{N}$, $(\alpha_l,\mathscr{B}_l)\in\mathbb{T}_l$, such that $(\alpha_l,\mathscr{B}_l)\to(\alpha,\mathscr{B})$. We observe that, by Proposition 4.6,

$$\begin{aligned}&|\mathcal{A}(\alpha_l,\mathscr{B}_l)-\mathcal{A}(\alpha,\mathscr{B})|\\ &=\left|\|u_{\alpha_l,\mathscr{B}_l}-u_c\|_{L^2(Q)}-\|u_{\alpha,\mathscr{B}}-u_c\|_{L^2(Q)}\right|\le\|u_{\alpha_l,\mathscr{B}_l}-u_{\alpha,\mathscr{B}}\|_{L^2(Q)}\\ &\le 2[\alpha_l2O(|\mathscr{B}_1-\mathscr{B}_2|_{\ell^\infty})+|\alpha_l-\alpha|]^{1/2}[PV_{\mathscr{B}_l}(u_{\alpha_l,\mathscr{B}_l})+PV_{\mathscr{B}}(u_{\alpha_l,\mathscr{B}})]^{1/2}\\ &\le 4K[\alpha_l2O(|\mathscr{B}_l-\mathscr{B}|_{\ell^\infty})+|\alpha_l-\alpha|]^{1/2}\|u_\eta\|^{1/2}_{W^{d,1}(Q)}.\end{aligned}\tag{4.33}$$

Here K is the optimal constant, which might depend on the L^2 norm of u_η, such that

$$[PV_{\mathscr{B}_l}(u_{\alpha_l,\mathscr{B}_l})]^{1/2}\le K\|u_\eta\|^{1/2}_{W^{d,1}(Q)}.$$

Next, take any optimal solution $(\alpha_{\mathbb{T}},\mathscr{B}_{\mathbb{T}})$ from (4.7), and by Assertion 1 we could obtain a sequence $\{(\alpha_{\mathbb{T}_l},\mathscr{B}_{\mathbb{T}_l})\}_{l=1}^{\infty}$, where, at each step $l\in\mathbb{N}$, $(\alpha_{\mathbb{T}_l},\mathscr{B}_{\mathbb{T}_l})\in\mathbb{T}_l$ is determined in (4.7), such that

$$(\alpha_{\mathbb{T}_l},\mathscr{B}_{\mathbb{T}_l})\to(\alpha_{\mathbb{T}},\mathscr{B}_{\mathbb{T}}).$$

Also, at each step $l\in\mathbb{N}$, we let the grid $\mathbb{G}_l(i_l,j_l)$ be such that

$$(\alpha_{\mathbb{T}},\mathscr{B}_{\mathbb{T}})\in\mathbb{G}_l(i_l,j_l), \tag{4.34}$$

where $\mathbb{G}_l(i_l,j_l)$ is defined in (4.6). Then, in view of (4.33), we have that

$$\begin{aligned}&\max\{\mathcal{A}(\alpha,\mathscr{B})\colon(\alpha,\mathscr{B})\in\mathbb{G}_l(i_l,j_l)\}-\min\{\mathcal{A}(\alpha,\mathscr{B})\colon(\alpha,\mathscr{B})\in\mathbb{G}_l(i_l,j_l)\}\\ &\le 4KP[O(P/l)+1/l]^{1/2}\|u_\eta\|^{1/2}_{W^{d,1}(Q)}.\end{aligned}\tag{4.35}$$

We have the following two cases.

Case 1: Assume that, at step l, it holds $(\alpha_{\mathbb{T}_l},\mathscr{B}_{\mathbb{T}_l})\in\mathbb{G}_l(i_l,j_l)$. In this case we could directly deduce

$$\begin{aligned}&\mathcal{A}(\alpha_{\mathbb{T}_l},\mathscr{B}_{\mathbb{T}_l})-\mathcal{A}(\alpha_{\mathbb{T}},\mathscr{B}_{\mathbb{T}})\le\max\{\mathcal{A}(\alpha,\mathscr{B})\colon(\alpha,\mathscr{B})\in\mathbb{G}_l(i_l,j_l)\}-\mathcal{A}(\alpha_{\mathbb{T}_l},\mathscr{B}_{\mathbb{T}_l})\\ &\le\max\{\mathcal{A}(\alpha,\mathscr{B})\colon(\alpha,\mathscr{B})\in\mathbb{G}_l(i_l,j_l)\}-\min\{\mathcal{A}(\alpha,\mathscr{B})\colon(\alpha,\mathscr{B})\in\mathbb{G}_l(i_l,j_l)\};\end{aligned}$$

Case 2: Assume that, at step l, it holds $(\alpha_{\mathbb{T}_l},\mathscr{B}_{\mathbb{T}_l})\notin\mathbb{G}_l(i_l,j_l)$. In this case, in view of the definition of $(\alpha_{\mathbb{T}_l},\mathscr{B}_{\mathbb{T}_l})$, we must have

$$\max\{\mathcal{A}(\alpha,\mathscr{B})\colon(\alpha,\mathscr{B})\in\mathbb{G}_l(i_l,j_l)\cap\mathbb{T}_l\}\ge\mathcal{A}(\alpha_{\mathbb{T}_l},\mathscr{B}_{\mathbb{T}_l}),\tag{4.36}$$

as the opposite would imply that $(\alpha_{\mathbb{T}_l},\mathscr{B}_{\mathbb{T}_l})$ is not a global minimizer over $\mathbb{T}_l$, which is a contradiction. Therefore, by (4.36), we have again

$$\begin{aligned}&\mathcal{A}(\alpha_{\mathbb{T}_l},\mathscr{B}_{\mathbb{T}_l})-\mathcal{A}(\alpha_{\mathbb{T}},\mathscr{B}_{\mathbb{T}})\le\max\{\mathcal{A}(\alpha,\mathscr{B})\colon(\alpha,\mathscr{B})\in\mathbb{G}_l(i_l,j_l)\cap\mathbb{T}_l\}-\mathcal{A}(\alpha_{\mathbb{T}},\mathscr{B}_{\mathbb{T}})\\ &\le\max\{\mathcal{A}(\alpha,\mathscr{B})\colon(\alpha,\mathscr{B})\in\mathbb{G}_l(i_l,j_l)\}-\min\{\mathcal{A}(\alpha,\mathscr{B})\colon(\alpha,\mathscr{B})\in\mathbb{G}_l(i_l,j_l)\},\end{aligned}$$

where, at the last inequality we used (4.34). Combining the two cases discussed above with (4.35) gives

$$\begin{aligned}&\mathcal{A}(\alpha_{\mathbb{T}_l},\mathscr{B}_{\mathbb{T}_l})-\mathcal{A}(\alpha_{\mathbb{T}},\mathscr{B}_{\mathbb{T}})\\ &\le\max\{\mathcal{A}(\alpha,\mathscr{B})\colon(\alpha,\mathscr{B})\in\partial\mathbb{G}_l(i_l,j_l)\}-\min\{\mathcal{A}(\alpha,\mathscr{B})\colon(\alpha,\mathscr{B})\in\mathbb{G}_l(i_l,j_l)\}\\ &\le 4KP[O(P/l)+1/l]^{1/2}\|u_\eta\|^{1/2}_{W^{d,1}(Q)},\end{aligned}\tag{4.37}$$

and hence the thesis.

Now we remove the assumption $u_\eta\in C^\infty(\bar{Q})$. Let $u_\eta^\varepsilon\in C^\infty(\bar{Q})$ be defined as in Case 2 of the proof of Theorem 4.1. Define

$$u^\varepsilon_{\alpha,\mathscr{B}}:=\arg\min\left\{\left\|u-u^\varepsilon_\eta\right\|^2_{L^2(Q)}+\alpha PV_{\mathscr{B}}(u)\colon u\in L^1(Q)\right\}.$$

Then by Proposition 4.5 we have that

$$\left\|u_{\alpha,\mathscr{B}}-u^\varepsilon_{\alpha,\mathscr{B}}\right\|_{L^2(Q)}\le\left\|u^\varepsilon_\eta-u_\varepsilon\right\|_{L^2(Q)},$$

for any $(\alpha,\mathscr{B})\in\mathbb{T}$. Then, for any fixed $\delta>0$, we could choose $\varepsilon>0$ small enough such that

$$\left\|u_\eta^\varepsilon - u_\varepsilon\right\|_{L^2(Q)} < \delta/4 \text{ and } \left\|u_\eta^\varepsilon\right\|_{W^{d,1}(Q)} \le \|u_\eta\|_{L^1(Q)}/\delta^d.$$

This, together with (4.37), gives

$$\begin{aligned}\mathcal{A}(\alpha_{\mathbb{T}_l},\mathscr{B}_{\mathbb{T}_l}) - \mathcal{A}(\alpha_{\mathbb{T}},\mathscr{B}_{\mathbb{T}}) &\le 4KP[O(P/l)+1/l]^{1/2}\|u_\eta\|_{W^{d,1}(Q)}^{1/2}+\delta/2\\ &\le 4KP[O(P/l)+1/l]^{1/2}\|u_\eta\|_{W^{d,1}(Q)}^{1/2}/\delta^d+\delta/2,\end{aligned}$$

as desired. □

4.3. Examples of Training ground. In this section we give some examples of collection Σ that satisfy Assumption 4.1. Recall the structure of operator $\mathscr{B}$ from Notation 2.1.

4.3.1. Operator $\mathscr{B}$ with invertible matrix. Let $P\in\mathbb{R}^+$ from Assumption 4.1 be given. We define the collection Σ_P by

$$\Sigma_P := \left\{\mathscr{B} : \left|(B^h)^{-1}\right| \le P, \text{ for each } 1\le h\le d\right\}. \tag{4.38}$$

We define the h-order total variation, say TV^h, of u by

$$TV^h(u) = \left|H^h u\right|_{\mathcal{M}_b(Q;\mathbb{M}^{N^h})}.$$

where H^h is the h-order Hessian operator defined in Notation 2.1. We also define the space $BV^d(Q)$ by

$$BV^d(Q) := \left\{u\in L^1(Q) : TV^d(u) < +\infty\right\},$$

endowed with norm

$$\|u\|_{BV^d(Q)} := \|u\|_{L^1(Q)} + TV^d(u).$$

PROPOSITION 4.7. *Let $\mathscr{B}\in\Sigma_P$ be given. Then the space $BV_{\mathscr{B}}(Q)$ is equivalent to the space $BV^d(Q)$.*

Proof. Without loss of generality, we assume that $u\in BV^d(Q)\cap C^\infty(Q)$, and in view of the structure of operator $\mathscr{B}$, we have

$$\begin{aligned}PV_{\mathscr{B}}(u) &= \sum_{h=1}^d \left|B^h H^h u\right| dx \le \sum_{h=1}^d \left|B^h\right|\left|H^h u\right| dx\\ &\le \sum_{h=1}^d TV^h(u) \le C\left(\|u\|_{L^1(Q)} + TV^d(u)\right),\end{aligned} \tag{4.39}$$

where the estimate is due to Sobolev inequality.

On the other hand,

$$TV^d(u) = \int_Q \left|H^d u\right| dx = \int_Q \left|(B^d)^{-1}B^d H^d u\right| dx \le \left|(B^d)^{-1}\right|_{\ell^\infty}\int_Q \left|B^d H^d u\right| dx \le PV_{\mathscr{B}}(u).$$

This, together with (4.39), allows us to conclude the proof. □

In the following proposition we show that Σ_P satisfies Assertion 2 of Assumption 4.1.

PROPOSITION 4.8. *Let $\mathscr{B}_1$ and $\mathscr{B}_2 \in \Sigma_P$, and $u \in BV^d(Q)$ be given. Then we have*

$$|PV_{\mathscr{B}_1}(u) - PV_{\mathscr{B}_2}(u)| \leq \left[\sqrt{K} |\mathscr{B}_1 - \mathscr{B}_2|_{\ell^\infty} \sum_{h \leq d} \left|(B_1^h)^{-1}\right|_{\ell^\infty} \right] PV_{\mathscr{B}_1}(u), \tag{4.40}$$

and

$$|PV_{\mathscr{B}_1}(u) - PV_{\mathscr{B}_2}(u)| \leq \left[\sqrt{K} |\mathscr{B}_1 - \mathscr{B}_2|_{\ell^\infty} \sum_{h \leq d} \left|(B_2^h)^{-1}\right|_{\ell^\infty} \right] PV_{\mathscr{B}_2}(u). \tag{4.41}$$

Proof. Assume first that $u \in C^\infty(\bar{Q})$. Direct computations give

$$|PV_{\mathscr{B}_1}(u) - PV_{\mathscr{B}_2}(u)| = \left| \int_Q |\mathscr{B}_1 u| \, dx - \int_Q |\mathscr{B}_2 u| \, dx \right| \leq \int_Q |\mathscr{B}_1 u - \mathscr{B}_2 u| \, dx.$$

Next, we observe that, for $x \in Q$,

$$(\mathscr{B}_1 - \mathscr{B}_2) u(x) = \sum_{h \leq d} (B_1^h - B_2^h) H^h u(x) = \sum_{h \leq d} (B_1^h - B_2^h)(B_1^h)^{-1} B_1^h H^h u(x),$$

and

$$\begin{aligned} \int_Q |(\mathscr{B}_1 - \mathscr{B}_2) u| \, dx &\leq \left[\sum_{h \leq d} \left|(B_1^h - B_2^h)\right|_{\ell^\infty} \left|(B_1^h)^{-1}\right|_{\ell^\infty} \right] \sum_{h \leq d} \left|B_1^h H^h u\right|_{\mathcal{M}_b(Q;\mathbb{R}^{N^h})} \\ &\leq \sqrt{K} \left[\sum_{h \leq d} \left|(B_1^h - B_2^h)\right|_{\ell^\infty} \left|(B_1^h)^{-1}\right|_{\ell^\infty} \right] \int_Q |\mathscr{B}_1 u| \, dx. \end{aligned}$$

Thus,

$$\begin{aligned} |PV_{\mathscr{B}_1}(u) - PV_{\mathscr{B}_2}(u)| &\leq \int_Q |\mathscr{B}_1 u - \mathscr{B}_2 u| \, dx \\ &\leq \left[\sqrt{K} |\mathscr{B}_1 - \mathscr{B}_2|_{\ell^\infty} \sum_{h \leq d} \left|(B_1^h)^{-1}\right|_{\ell^\infty} \right] PV_{\mathscr{B}_1}(u). \end{aligned} \tag{4.42}$$

To conclude, we use an approximation sequence $\{u_\varepsilon\}_{\varepsilon > 0} \subset C^\infty(\bar{Q})$ from Corollary 2.1, such that $u_\varepsilon \to u$ in $L^1(Q)$ and

$$PV_{\mathscr{B}_1}(u_\varepsilon) \to PV_{\mathscr{B}_1}(u) \text{ and } PV_{\mathscr{B}_2}(u_\varepsilon) \to PV_{\mathscr{B}_2}(u).$$

Combining with (4.42) gives (4.40), as desired. Finally, we remark that (4.41) is obtained in the same way, and the proof is complete. ∎

REMARK 4.2. By (4.38), (4.40), and (4.41), we conclude that

$$|PV_{\mathscr{B}_1}(u) - PV_{\mathscr{B}_2}(u)| \leq \left[d\sqrt{K} P |\mathscr{B}_1 - \mathscr{B}_2|_{\ell^\infty} \right] PV_{\mathscr{B}_1}(u),$$

and

$$|PV_{\mathscr{B}_1}(u) - PV_{\mathscr{B}_2}(u)| \le \left[d\sqrt{K} P |\mathscr{B}_1 - \mathscr{B}_2|_{\ell^\infty} \right] PV_{\mathscr{B}_2}(u).$$

Thus, by setting

$$O(|\mathscr{B}_1 - \mathscr{B}_2|_{\ell^\infty}) = d\sqrt{K} P |\mathscr{B}_1 - \mathscr{B}_2|_{\ell^\infty},$$

we conclude that Σ_P satisfies Assumption 4.1.

5. Experimental insights, further extensions, and upcoming works

5.1. Numerical simulations. We remark that the reconstructed image $u_{\alpha,\mathscr{B}}$ defined in ($\mathcal{T}$-L2), for a given $(\alpha,\mathscr{B}) \in \mathbb{T}$, can be computed using the primal-dual algorithm presented in [6]. Indeed, we could recast the minimizing problem ($\mathcal{T}$-L2) as the min-max problem

$$\min \left\{ \max \left\{ \|u - u_\eta\|^2_{L^2(Q)} + \alpha \langle u, \mathscr{B}^* \varphi \rangle : \varphi \in C_c^\infty(Q;\mathbb{R}^K) \right\} : u \in L^1(Q) \right\},$$

and then the primal-dual method presented in [6] can be applied.

Next, we present a practical application of Theorem 4.2.

Let $u_c \in L^2(Q), u_\eta \in L^2(Q)$ be given. Let an acceptable error threshold $\varepsilon > 0$ be given.

- Initialization: Choose $\varepsilon > 0$, and the box-constraint constant $P > 0$.
- Step 1: Let $\delta = \varepsilon/2$, and increase step $l \in \mathbb{N}$ until the training error, given in Assertion 2 of Theorem 4.2, does not exceed $\varepsilon/2$.
- Step 2: Determine one global minimizer $(\alpha_{\mathbb{T}_l}, \mathscr{B}_{\mathbb{T}_l})$ of assessment function $\mathcal{A}(\alpha,\mathscr{B})$ over the finite training ground $\mathbb{T}_l$. Then, by Theorem 4.2,

$$|\mathcal{A}(\alpha_{\mathbb{T}_l}, \mathscr{B}_{\mathbb{T}_l}) - \mathcal{A}(\alpha_{\mathbb{T}}, \mathscr{B}_{\mathbb{T}})| \le \varepsilon,$$

- Step 3: The reconstructed image $u_{\alpha_{\mathbb{T}_l}, \mathscr{B}_{\mathbb{T}_l}}$ is then a desired optimal reconstructed result within the acceptable error range.

To make an appropriate comparison, we apply our proposed training scheme $\mathcal{T}$ (($\mathcal{T}$-L1)-($\mathcal{T}$-L2)) on the image in Figure 5.1, with the following training grounds

$$\mathbb{T}^0 := [0,1] \times \{\mathscr{B}_0\}, \text{ where } \mathscr{B}_0 := [1,0;0,1], \tag{5.1}$$

$$\mathbb{T}^1 := [0,1] \times \{\mathscr{B}_s : -0.5 \le s \le 0.5\}, \text{ where } \mathscr{B}_s := [1,s;0,1], \tag{5.2}$$

$$\mathbb{T}^2 := [0,1] \times \{\mathscr{B}_{s,t} : -0.5 \le s,t \le 0.5\}, \text{ where } \mathscr{B}_{s,t} := [1,s;t,1], \tag{5.3}$$

where we use the super-script to avoid confusion with the finite training ground $\mathbb{T}_l$. Note that the training ground $\mathbb{T}^0$ gives the original training scheme $\mathcal{B}$ (($\mathcal{B}$-L1)-($\mathcal{B}$-L2)) with TV regularizer only. We perform numerical simulations on the images shown in Figure 5.1: The first image represents a clean image u_c, whereas the second one is a noisy version u_η. We summarize our simulation results in Table 5.1 below. We observe that, from Table 5.1, as the training ground expanded, the minimum value of assessment function $\mathcal{A}(\alpha,\mathscr{B})$ decreased. That is, our new regularizer $PV_{\mathscr{B}}$ really provides an improved reconstructed result compared to TV. However, we remark that while the extension of training ground results in a considerably larger amount of CPU

Training ground	optimal solution	minimum assessment value
$\mathbb{T}^0$	$\alpha_{\mathbb{T}^0} = 0.048$	14.8575
$\mathbb{T}^1$	$\alpha_{\mathbb{T}^1} = 0.052,\ s_{\mathbb{T}^1} = 0.4$	12.8382
$\mathbb{T}^2$	$\alpha_{\mathbb{T}^2} = 0.052,\ s_{\mathbb{T}^2} = -0.2,\ t_{\mathbb{T}^2} = 0.5$	12.2369

TABLE 5.1. *Minimum assessment value for scheme $\mathcal{T}$ over the training grounds defined in* (5.1), (5.2), *and* (5.3).

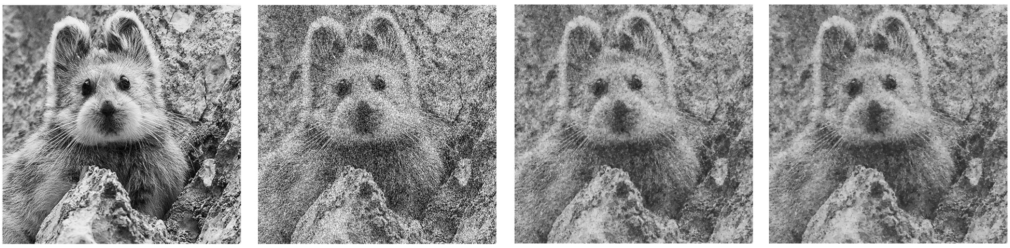

FIG. 5.1. *From left to right: clean image u_c; corrupted image u_η (with heavy artificial Gaussian noise); the optimally reconstructed image at $\alpha_{\mathbb{T}^0}$; the optimally reconstructed image at $(\alpha_{\mathbb{T}^2}, \mathscr{B}_{\mathbb{T}^2})$.*

time, this is a minor issue in real applications, since it is only computed once for a given data set, and the structure of finite training ground $\mathbb{T}_l$ allows us to use parallel computing very efficiently, hence reducing CPU usage.

We remark that the introduction of $PV_{\mathscr{B}}$ regularizers into the training scheme is only meant to expand the training choices, rather than to provide a superior seminorm with respect to the standard TV. Whether the optimal regularizer is TV, or an intermediate one, is completely dependent on the training image $u_\eta = u_c + \eta$. Moreover, we remark that the results in this article are not restricted to image processing problems. They can be generally applied to parameter estimation problems of variational inequalities, as long as a suitable assessment function can be found.

Acknowledgments. PL acknowledges support from the EPSRC Centre Nr. EP/N014588/1 and the Leverhulme Trust project on Breaking the non-convexity barrier. XYL acknowledges the support of his NSERC Discovery Grant "Regularity of minimizers and pattern formation in geometric minimization problems". Part of this research was performed when PL was affiliated with the Centre of Mathematical Imaging and Healthcare, University of Cambridge.

REFERENCES

[1] T. Barbu and G. Marinoschi, *Image denoising by a nonlinear control technique*, Int. J. Control, 90:1005–1017, 2017. 1

[2] M. Bergounioux, *Optimal control of problems governed by abstract elliptic variational inequalities with state constraints*, SIAM J. Control Optim., 36(1):273–289 (electronic), 1998. 1

[3] K. Bredies, K. Kunisch, and T. Pock, *Total generalized variation*, SIAM J. Imaging Sci., 3(3):492–526, 2010. 1

[4] D. Breit, L. Diening, and F. Gmeineder, *Traces of functions of bounded A-variation and variational problems with linear growth*, arXiv preprint, arXiv:1707.06804, 2017. 2.3

[5] H. Brezis, *Functional Analysis, Sobolev Spaces and Partial Differential Equations*, Universitext, Springer, New York, 2011. 3.2, 3.2

[6] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vis., 40(1):120–145, 2011. 5.1

[7] Y. Chen, T. Pock, R. Ranftl, and H. Bischof, *Revisiting loss-specific training of filter-based MRFs for image restoration*, in J. Weickert, M. Hein, and B. Schiele (eds.), Pattern Recognition, GCPR 2013, Lecture Notes in Computer Science, vol. 8142. Springer, Berlin, Heidelberg, 271–281, 2013. 1

[8] Y. Chen, R. Ranftl, and T. Pock, *Insights into analysis operator learning: From patch-based sparse models to higher order MRFs*, IEEE Trans. Image Process., 23(3):1060–1072, 2014. 1

[9] E. Davoli, I. Fonseca, and P. Liu, *Adaptive image processing: first order* PDE *constraint regularizers and a bilevel training scheme*, arXiv preprint, arXiv:1902.01122, 2019. 1

[10] E. Davoli and P. Liu, *One dimensional fractional order TGV: gamma-convergence and bilevel training scheme*, Commun. Math. Sci., 16(1):213–237, 2018. 3, 1

[11] J.C. De los Reyes and C.-B. Schönlieb, *Image denoising: learning the noise model via nonsmooth PDE-constrained optimization*, Inverse Probl. Imaging, 7(4):1183–1214, 2013. 1, 1

[12] J. Domke, *Generic methods for optimization-based modeling*, Proc. Mach. Learn. Res., 22:318–326, 2012. 1

[13] I. Ekeland and R. Temam, *Convex Analysis and Variational Problems*, SIAM, 1976. 3.2, 3.3

[14] L.C. Evans and R.F. Gariepy, *Measure Theory and Fine Properties of Functions, Studies in Advanced Mathematics*, CRC Press, Boca Raton, FL, 1992. 1, 2.1, 2.1, 2.1, 3.2, 3.2

[15] I. Fonseca and S. Müller, *Relaxation of quasiconvex functionals in* $\mathrm{BV}(\Omega,\mathbb{R}^p)$ *for integrands* $f(x,u,\nabla u)$, Arch. Ration. Mech. Anal., 123(1):1–49, 1993. 2.3

[16] Y. Gao, *Global strong solution with BV derivatives to singular solid-on-solid model with exponential nonlinearity*, J. Differ. Equ., 267(7):4429–4447, 2019. 1

[17] Y. Gao, G. Jin, and J.-G Liu, *Inbetweening auto-animation via Fokker-Planck dynamics and thresholding*, Inverse Probl. Imaging, 15(5):843–864, 2021. 1

[18] K.C. Kiwiel, *Convergence and efficiency of subgradient methods for quasiconvex minimization*, Math. Program., 90(1):1–25, 2001. 1

[19] K. Kunisch and T. Pock, *A bilevel optimization approach for parameter learning in variational models*, SIAM J. Imaging Sci., 6(2):938–983, 2013. 1

[20] P. Liu and X.Y. Lu, *Real order (an)-isotropic total variation with application to the learning of optimal structures*, arXiv preprint, arXiv:1903.08513, 2019. 3, 1

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

[22] S.G. Samko, A.A. Kilbas, and O.I. Marichev, *Fractional Integrals and Derivatives: Theory and Applications*, Gordon and Breach Science Publishers, Yverdon, 1993. 1

[23] C.-B. Schönlieb and P. Liu, *Learning optimal orders of the underlying Euclidean norm in total variation image processing*, arXiv preprint, arXiv:1903.11953v1, 2019. 2, 1, 1, 1, 4.2

[24] M.F. Tappen, C. Liu, E.H. Adelson, and W.T. Freeman, *Learning Gaussian conditional random fields for low-level vision*, IEEE Conf. Comput. Vis. Pattern Recognit., 1–8, 2007. 1