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

PAN LIU

Centre of Mathematical Imaging and Healthcare, Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK

March 19, 2019

ABSTRACT. A class of mixed-order PDE-constraint regularizer for image processing problem is proposed, generalizing the standard first order total variation (TV). A semi-supervised (bilevel) training scheme, which provides a simultaneous optimization with respect to parameters and new class of regularizers, is studied. Also, a finite approximation method, which used to solve the global optimization solutions of such training scheme, is introduced and analyzed.

CONTENTS

1. Introduction 2
2. The space of functions with bounded $PV$-seminorm 5
2.1. The PDE-constraint total variation defined by operator $\mathcal{B}$ 5
3. Analytic properties of PDE-constraint variations 9
3.1. $\Gamma$-convergence of functionals defined by $PV$ seminorms 9
3.2. The point-wise characterization of sub-differential of $PV_{\mathcal{B}}$ 12
4. Learning the optimal operator $\mathcal{B}$ in imaging processing problems 16
4.1. The bilevel training scheme with the $PV_{\mathcal{B}}$ regularizer 16
4.2. Numerical realization and finite approximation of scheme $T$ 18
4.3. Examples of Training ground 27
5. Experimental insights, further extensions, and upcoming works 29
5.1. Numerical simulations 29
5.2. Further generalization of $PV_{\mathcal{B}}$ regularizer 30
5.3. Upcoming works 32
References 33

E-mail address: panliu.0923@maths.cam.ac.uk.

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

Key words and phrases. image processing, optimal training scheme, higher order differential operators, $\Gamma$-convergence.
1. Introduction

The use of variational technics with non-smooth regularizers in image processing has become popular in the last decades. One of the most successful approaches is introduced in the celebrated work [20] which relies on the so called ROF total-variational functional

\[ I(u) := \|u - u_\eta\|^2_{L^2(Q)} + \alpha TV(u), \]  

(1.1)

where \(u_\eta \in L^2(Q)\) is a given corrupted image, \(Q := (0,1)^2\) represents the unit square, \(\alpha \in \mathbb{R}^+\) is an intensity parameter, and \(TV(u)\) stands for the total variation of \(u\) in \(Q\) (see [14]). In the simple case that \(u \in W^{1,1}(Q)\), we have

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

(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 has been introduced and studied:

1. using higher-order extensions ([3, 9]);
2. changing the underlying Euclidean norm ([22]);
3. introducing fractional order derivatives [18, 10].

These methods introduces collections of regularizers which generalizes TV seminorm. For example, in [22], the underlying Euclidean norm of TV seminorm is generalized from \(p = 2\), used in (1.2), to \(p \in [1, +\infty]\) by

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

In [18], the order of derivative is generalized from \( r = 1 \), used in (1.2), to \( r \in \mathbb{R}^+ \), by

\[ TV^r(u) = \int_Q |\nabla^r u| \, dx = \int_Q \left( |\partial^r_1 u(x)|^2 + |\partial^r_2 u(x)|^2 \right)^{1/2} \, dx, \]

in which the fractional order derivative is realized by using the Riemann-Liouville fractional order derivative (see [21] for definition). In both [18, 10], it has been shown that for given corrupted image \(u_\eta\), a carefully selected regularizer parameter \(p \in [1, +\infty]\) (resp. \(r \in \mathbb{R}^+\)) allows TV\(_p\) (resp. TV\(_r\)) to provide improved imaging processing result, 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 imaging processing results would certainly be improved if we could further expand the collections of regularizers. To this purpose, in this paper we introduce a family of novel TV-like PDE-constraint regularizer (semi-norm), say \(PV_B\), by

\[ PV_B(u) := |B u|_{\mathcal{M}_0(Q; \mathbb{R}^K)}. \]  

(1.3)

Here \(|\cdot|_{\mathcal{M}_0}\) denotes the Radon norm of a measure, and \(B : L^1(Q) \to D'(Q; \mathbb{R}^K)\) is a linear differential operator (see Notation 2.1). In the simple case \(B = \nabla\), we recover the total variation TV seminorm. 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 aim of this paper is threefold. First, we provide a rigorous and detailed analysis of the properties of the $PV_{B}$ seminorm, such as the approximation by smooth functions, lower semi-continuity with respect to both function $u$ and operator $B$, and a point-wise characterization of the sub-gradient of $PV_{B}$.

The second result is the study of the aforementioned selection mechanism, realized by a semi-supervised (bilevel) training scheme defined in machine learning (see [7, 8, 12, 23, 11, 17]). 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 $u_{\eta}$ 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), and $\eta$ encodes noise, and we call $(u_{\eta}, u_{c})$ as training set. Then, a bilevel training scheme, say Scheme $B$, for determining the optimal intensity parameter $\alpha$ can be formulated as follows:

Level 1. $\alpha_{T} \in A[T] := \arg\min \left\{ \|u_{\alpha} - u_{c}\|^{2}_{L^{2}} : \alpha \in T \right\}$, \hspace{1cm} (B-L1)

Level 2. $u_{\alpha} := \arg\min \left\{ \|u - u_{\eta}\|^{2}_{L^{2}(Q)} + \alpha TV_{2}(u) : u \in BV(Q) \right\}$, \hspace{1cm} (B-L2)

where $T := \text{cl}(\mathbb{R}^{+})$, used in (B-L1), is called the training ground. Roughly speaking, Level 1 problem in (B-L1) looks for an $\alpha$ that solves the minimum $L^{2}$-distance to the clean image $u_{c}$, subject to the minimizing problem (B-L2). That is, scheme $B$ is able to optimally adapt itself to the given “perfect data” $u_{c}$.

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

Level 1. $(\alpha_{T}, B_{T}) \in A[T] := \arg\min \left\{ \|u_{c} - u_{\alpha, B}\|^{2}_{L^{2}(Q)} : (\alpha, B) \in T \right\}$, \hspace{1cm} (T-L1)

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

In (T-L1), we expand the training ground to $T := \text{cl}(\mathbb{R}^{+}) \times \Sigma$ to incorporate the new parameter $B \in \Sigma$, where $\Sigma$ denotes a closed collection of operators $B$ (see Notation 2.1, Notation 4.1, and (4.5) for details). We remark that the expanded training ground $T$ allows scheme $T$ to optimize the regularizer $PV_{B}(u)$ and intensity parameter $\alpha$ simultaneously. We summarize the main result in the following theorem.

**Theorem 1.1** (see Theorem 4.4). The training scheme $T$ admits at least one solution $(\alpha_{T}, B_{T}) \in T$, and provides an associated optimally reconstructed image $u_{\alpha_{T}, B_{T}} \in BV_{B_{T}}(Q)$.

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

$$A(\alpha, B) := \|u_{\alpha, B} - u_{c}\|^{2}_{L^{2}(Q)},$$

where $u_{\alpha, B}$ is obtained from (T-L2). However, as shown in [22] that even in the simplest case with $B = \nabla$ (i.e. $PV_{B} = TV$), the assessment function $A(\alpha, \nabla)$ is not quasi-convex (in the sense of [16], or simply convex), and hence the 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 $(\alpha', B')$ is an acceptable optimal solution of
scheme $\mathcal{T}$ with the given error $\varepsilon > 0$ if
\[
|A(\alpha_T, B_T) - A(\alpha', B')| < \varepsilon,
\]
(1.5)
where $(\alpha_T, B_T)$ is a global minimum obtained from $(\mathcal{T}$-L1).

To compute such acceptable optimal solution we use a finite approximation method, originally introduced and studied in [22], and generalized 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 $\mathcal{T}$.

- The intensity parameter $\alpha$ is contained in a closet interval $[0, P]$, where the box-constraint constant $P > 0$ can be chosen by the user;
- the collection $\Sigma$ of operator $B$ satisfies an additional continuity assumptions, such as, for any $B_1, B_2 \in \Sigma$,
\[
|PV_{B_1}(u) - PV_{B_2}(u)| \leq O \left(|B_1 - B_2| \min \{PV_{B_1}(u), PV_{B_2}(u)\}\right),
\]
where $O(\cdot)$ denotes the big-O notation.

Then, the finite approximation method is constructed based on a sequence of (finite) training sets $\mathcal{T}_l$, indexed by $l \in \mathbb{N}$, in which (where $\mathcal{H}^0(\cdot)$ denotes the counting measure)
\[
\mathcal{H}^0(\mathcal{T}_l) < +\infty \text{ and } \mathcal{T} \subset \text{cl} \left( \bigcup_{l \in \mathbb{N}} \mathcal{T}_l \right).
\]

For the precise definition of $\mathcal{T}_l$ we refer readers to Definition 4.7. We remark that, since $\mathcal{H}^0(\mathcal{T}_l) < +\infty$ for each $l \in \mathbb{N}$ fixed, we could evaluate $A(\alpha, B)$ at each element of $\mathcal{T}_l$ and determine the optimal solution(s)
\[
(\alpha_T, B_T) \in A[\mathcal{T}_l] := \arg \min \{A(\alpha, B) : (\alpha, B) \in \mathcal{T}_l\}
\]
precisely. The following theorem is established in order to achieve (1.5).

**Theorem 1.2** (see Theorem 4.9). *Let a training ground $\mathcal{T}$ satisfies above box-constraint. Then the following assertions hold:*

1. *we have*
\[
\lim_{l \to \infty} \text{dist}(A[\mathcal{T}], A[\mathcal{T}_l]) = 0.
\]
2. *Let $\varepsilon > 0$ be given. Then for each $l \in \mathbb{N}$ we have*
\[
|A(\alpha_T, B_T) - A(\alpha_T, B_T)| \leq 4KP \left[O \left(\frac{P}{l}\right) + \frac{1}{l}\right]^{1/2} \|u\|_{W^{1/2}}^{1/2} / \varepsilon^d + \varepsilon/2,
\]
*where the value of right hand side can be computed explicitly.*

That is, for any given $\varepsilon > 0$, we could compute $l \in \mathbb{N}$ that is large enough so that the corresponding optimal solution $(\alpha_T, B_T)$ 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 $\mathcal{T}$ is still sufficiently large to encompass interesting operator. We finally remark that, although this work focuses mainly on the theoretical analysis of the operators $PV_{B}$ and the training scheme $\mathcal{T}$, in Section 5.1 a primal-dual algorithm for solving $(\mathcal{T}$-L2) is discussed, and some preliminary numerical demonstration of scheme $\mathcal{T}$ are provided.

Our article is organized as follows. In Section 2 we analyze the functional properties of the $PV_{B}$-seminorms. The $\Gamma$-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 demonstrate several numerical implementations, and in Section 5.2 some possible extensions of $PV_B$.

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 unit open cube in $\mathbb{R}^N$. $M^{N^N}$ is the space of matrices with dimension $N \times N \times \cdots \times N$ ($n$ times) with elements in $\mathbb{R}$. For the convenience of the presentation of this article, we identify the matrix space $M^{N^N}$ by vector space $\mathbb{R}^{N^n}$, where $N^n = N \cdot N \cdots N$ ($n$ times). Moreover, $\mathcal{D}'(Q, \mathbb{R}^n)$ represents the space of distributions with values in $\mathbb{R}^n$.

Notation 2.1. We collect some notation which will be adopted in connection with linear differential operators.

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

2. For each $h = 1, \ldots, d$, we let $B^h$ be matrix mapping from $\mathbb{R}^{N^h}$ to $\mathbb{R}^{N^h}$ and

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

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

$$\mathcal{B}u := \sum_{h \in \mathbb{N}, h \leq d} B^h(H^h u); \quad (2.1)$$

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

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

4. We denote the bilinear operator $\circ \mathcal{B}$, induced by $\mathcal{B}$, such that

$$\mathcal{B}(uw) = w\mathcal{B}u + u \circ \mathcal{B} w \quad (2.2)$$

5. Given a sequence of operators $\{\mathcal{B}_n\}_{n=1}^\infty$ and an operator $\mathcal{B}$, with coefficients $\{B_n\}_{n=1}^\infty$ and $B$, respectively, we say that $\mathcal{B}_n \to \mathcal{B}$ in $\ell^\infty$ if

$$|\mathcal{B}_n - \mathcal{B}| := \sum_{h \leq d} |B^h_n - B^h|_{\ell^\infty} \to 0,$$

where $|\cdot|_{\ell^\infty}$ stands for the $\ell^\infty$ matrix norm.

Definition 2.2. Let $d \in \mathbb{N}$ be fixed. We denote by $\Pi^d$ the collection of operator $\mathcal{B}$ defined in notation 2.1, with order at most $d$.

2.1. The PDE-constraint total variation defined by operator $\mathcal{B}$. We generalize the standard total variation seminorm by using the $d$-th order differential operators $\mathcal{B} \in \Pi^d$ defined in Definition 2.2.

Definition 2.3. Let $u \in L^1(Q)$ and operator $\mathcal{B} \in \Pi^d$ be given.

1. We define the PDE-constraint seminorm, say $PV_{\mathcal{B}}$, by

$$PV_{\mathcal{B}}(u) := \sup \left\{ \int_Q u \mathcal{B}^* \varphi \, dx : \varphi \in C^\infty_c(Q; \mathbb{R}^K), |\varphi| \leq 1 \right\}; \quad (2.3)$$
2. We define the space 

\[ BV_B(Q) := \left\{ u \in L^1(Q) : PV_B(u) < +\infty \right\}, \]

and we equip it with the norm 

\[ \|u\|_{BV_B(Q)} := \|u\|_{L^1(Q)} + PV_B(u). \]  

(2.4)

In next proposition we collect several preliminary results regarding functions in space \( BV_B(Q) \).

**Proposition 2.4.** Let operator \( B \in \Pi^d \) and \( u \in BV_B(Q) \) be given.

1. For any sequence \( \{u_n\}_{n=1}^\infty \subset L^1(Q) \) and function \( u \in L^1(Q) \) that satisfying one of the following conditions:
   - i. \( \{u_n\}_{n=1}^\infty \) is locally uniformly integrable and \( u_n \to u \) a.e..
   - ii. \( u_n \rightharpoonup u \) in \( M_b(Q) \).

Then, we have

\[ \lim inf_{n \to \infty} PV_B(u_n) \geq PV_B(u). \]  

(2.5)

2. There exists a Radon measure \( \mu \) and a \( \mu \)-measurable function \( \sigma : Q \to \mathbb{R}^K \) such that
   - i. \( |\sigma(x)| = 1 \) \( \mu \)-a.e.;
   - ii. for all \( \varphi \in C_c^\infty(Q; \mathbb{R}^K) \), there holds

\[ \int_Q u B^* \varphi \, dx = -\int_Q \varphi \cdot \sigma \, d\mu. \]

**Proof.** We prove Assertion 1 first. If

\[ \lim inf_{n \to \infty} PV_B(u_n) = +\infty, \]

there is nothing to prove. Assume not, then we have, for arbitrary \( \varphi \in C_c^\infty(Q; \mathbb{R}^K) \), that

\[ \lim inf_{n \to \infty} PV_B(u_n) \geq \lim inf_{n \to \infty} \int_Q u_n B^* \varphi \, dx = \int_Q u B^* \varphi \, dx, \]

where the last equality can be deduced either from condition 1(i) or 1(ii), independently. Hence, we conclude (2.5) in view of the arbitrariness of \( \varphi \).

We next prove Assertion 2. We define the linear functional \( L : C_c^\infty(Q; \mathbb{R}^K) \to \mathbb{R} \) such that

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

Then, since \( u \in BV_B(Q) \), we have that

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

which implies that

\[ |L(\varphi)| \leq PV_B(u) \|\varphi\|_{L^\infty(Q)}. \]  

(2.6)

Now, for arbitrary \( \varphi \in C_c(Q; \mathbb{R}^K) \), we define the mollifications \( \varphi_\varepsilon := \varphi * \eta_\varepsilon \), for some mollifier \( \eta_\varepsilon \) with \( \varepsilon < \text{dist} (\text{spt} (\varphi), \partial Q) \). Then we have, by [14, Theorem 1, item (ii), Section 4.2], that \( \varphi_\varepsilon \to \varphi \) uniformly on \( Q \). Therefore, by defining

\[ \tilde{L}(\varphi) := \lim_{\varepsilon \to 0} L(\varphi_\varepsilon) \text{ for } \varphi \in C_c(Q; \mathbb{R}^K), \]
and together with (2.6), we conclude that
\[ \sup \{ \bar{L}(\varphi) : \varphi \in C_c(Q; \mathbb{R}^K) \text{ and } |\varphi| \leq 1 \} < +\infty. \]
Thus, in view of the Riesz representation theorem (see [14, Section 1.8]), the proof is complete. □

**Remark 2.5.** We henceforth write \(|\mathcal{B}u|\) by \(\mu\) and have
\[ \int_Q u \mathcal{B}^* \varphi \, dx = - \int_Q \varphi \cdot \sigma \, d|\mathcal{B}u| \]
for arbitrary \(\varphi \in C^\infty_c(Q; \mathbb{R}^K)\).

**Theorem 2.6** (local approximation by smooth functions). Let \(p \geq 1\) and \(u \in \text{BV}_\mathcal{B}(Q) \cap L^p(Q)\) be given. There exists a sequence \(\{u_n\}_{n=1}^\infty \subset C^\infty_c(Q; \mathbb{R}^K)\) such that the following assertions hold.

1. \(u_n \to u\) strongly in \(L^p(Q)\);
2. \(\text{PV}_\mathcal{B}(u_n) \to \text{PV}_\mathcal{B}(u)\).
3. \(u_n \in C^\infty(\bar{Q})\) for each \(n \in \mathbb{N}\).

**Remark.** Assertion 3 only asserts that for each fixed \(n \in \mathbb{N}\) that \(u_n \in C^\infty(\bar{Q})\) but it is possible that \(\|u_n\|_{L^1(\partial Q)} \to \infty\) as \(n \to \infty\). In another word, we make no conclusions with respect to the trace value of \(u\) from Theorem 2.6.

**Proof.** The construction of approximation sequence \(\{u_n\}_{n=1}^\infty\) is almost same to the approximation sequence used in the standard \(BV\) case as presented in [14, Theorem 2, Page 172]. We shall only concentrated on showing that Assertion 3 holds, but for reader’s convenience, we shall outline the construction of approximation sequence and key steps.

Let \(u \in \text{BV}_\mathcal{B}(Q)\) be given, and let \(Q_k\) be the cube centered at point \(q = (1/2, 1/2)^N\) with side length \(1 - 1/(k+M)\). Let arbitrary \(\varepsilon > 0\) be given, we choose \(M > 0\) large enough such that
\[ |\mathcal{B}u|(Q \setminus Q_{0+M}) < \varepsilon/2. \]

Define \(Q_0 = Q_{0+M}\) and
\[ V_k := Q_{k+1} \setminus \bar{Q}_k \text{ for } k \in \mathbb{N}. \]

Let \(\{\zeta_k\}_{k=1}^\infty \subset C^\infty_c(Q)\) be the partition of unity such that
\[ \zeta_k \in C^\infty_c(V_k), \text{ such that } 0 \leq \zeta_k \leq 1 \]
\[ \sum_{k \geq 1} \zeta_k(x) = 1 \text{ for each } x \in Q. \]

Let \(\eta_\varepsilon\) be the standard mollifier, and for each \(k\), we choose \(\varepsilon_k\) small enough such that
\[ \text{spt}(\eta_{\varepsilon_k} * (u \zeta_k)) \subset V_k \]  \hspace{1cm} (2.7)
\[ \|\eta_{\varepsilon_k} * (u \zeta_k) - u \zeta_k\|_{L^p(Q)} < \varepsilon/2^{k+1} \]  \hspace{1cm} (2.8)
\[ \|\eta_{\varepsilon_k} * (u \mathcal{B}\zeta_k) - u \mathcal{B}\zeta_k\|_{L^1(Q)} < \varepsilon/2^{k+1}, \]  \hspace{1cm} (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 $u_\varepsilon \to u$ strongly in $L^p(Q)$.

This, and together with Assertion 1, Proposition 2.4, we conclude that

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

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

$$\langle \eta_{\varepsilon k} * (u \cdot \zeta_k), \mathcal{B}^* \varphi \rangle = \langle u \cdot \zeta_k, \mathcal{B}^* (\eta_{\varepsilon k} * \varphi) \rangle = \langle u, \mathcal{B}^* (\zeta_k (\eta_{\varepsilon k} * \varphi)) \rangle - \langle u, (\eta_{\varepsilon k} * \varphi) \circ \mathcal{B}^* \nabla \zeta_k \rangle,$$

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

$$\langle u_\varepsilon, \mathcal{B}^* \varphi \rangle = \sum_{k \geq 1} \langle \eta_{\varepsilon k} * (u \cdot \zeta_k), \mathcal{B}^* \varphi \rangle = \sum_{k \geq 1} \langle u, \mathcal{B}^* (\zeta_k (\eta_{\varepsilon k} * \varphi)) \rangle - \sum_{k \geq 1} \langle u, (\eta_{\varepsilon k} * \varphi) \circ \mathcal{B}^* \nabla \zeta_k \rangle.$$

Following the same computation used in [14, Theorem 2, Page 172] and use (2.9), we deduce that

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

Hence, in view of the arbitrariness of $\varphi$, we obtain that

$$\limsup_{\varepsilon \to 0} PV_\mathcal{B}(u_\varepsilon) \leq PV_\mathcal{B}(u),$$

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

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

Consequentially, we have $u_{\varepsilon, \delta} \to u_\varepsilon$ in $L^p$ strong and $PV_\mathcal{B}(u_{\varepsilon, \delta}) \to PV_\mathcal{B}(u_\varepsilon)$, as $\delta \to 0$. Hence, by using a diagonal argument, we could extract a subsequence $\{u_{\delta_n}\}_{n \geq 1}$ such that $u_{\delta_n} \to u$ strongly in $L^p$ and $PV_\mathcal{B}(u_{\delta_n}) \to PV_\mathcal{B}(u)$.

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

**Remark.** The construction of $u_{\varepsilon, \delta}$ in (2.10) is possible because of the simple geometry of domain $Q$. However, for domain with arbitrary geometry, even with Lipschitz boundary, such construction is not available. We refer readers to [4, 15] for alternative constructions with, however, operator $\mathcal{B}$ with several additional restriction.

**Corollary 2.7.** Let a finite set of $\mathcal{B}_i$, $i = 1, \ldots, M$, be given and

$$u \in \bigcap_{i=1}^MBV_{\mathcal{B}_i}(Q).$$

Then, there exists a sequence

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

such that the following assertions hold.

1. $u_n \to u$ strongly in $L^1(Q)$;
2. $PV_{\mathcal{B}_i}(u_n) \to PV_{\mathcal{B}_i}(u)$, for each $i = 1, \ldots, M$ uniformly;
3. $u_n \in C^\infty(Q)$ for each $n \in \mathbb{N}$. 
Proof. We only need to change (2.9) to
\[ \sum_{i=1}^{M} \| \eta_{\varepsilon_k} * (\eta B_i \zeta_k) - \eta B_i \zeta_k \|_{L^1(Q)} < \varepsilon / 2^{k+1}, \]
and the rest follows with the same argument used in the proof of Theorem 2.6. \qed

We close this section by stating the l.s.c. result of PV semi-norm.

**Proposition 2.8.** Let \( u \in L^1(Q) \) and sequence \( \{ B_n \}_{n=1}^{\infty} \) such that \( B_n \to B \) in \( \ell^\infty \) be given. Then, we have that
\[
\lim_{n \to \infty} \text{PV}_{B_n}(u) \geq \text{PV}_B(u).
\]

**Proof.** First of all, if \( \lim \inf_{n \to \infty} \text{PV}_{B_n}(u) = +\infty \), then there is nothing to prove. Suppose \( \sup \{ \text{PV}_{B_n}(u) : n \in \mathbb{N} \} := M < +\infty \), then, for arbitrary \( \varphi \in C_c^\infty(Q; \mathbb{R}^K) \), we have that
\[
+\infty > \lim_{n \to \infty} \text{inf} \text{PV}_{B_n}(u) \geq \lim_{n \to \infty} \int_Q u B_n \varphi \, dx = \int_Q u B^* \varphi \, dx.
\]
Hence, by taking supremum with respect to \( \varphi \) on the right hand side of above inequality, we conclude that
\[
\lim_{n \to \infty} \text{inf} \text{PV}_{B_n}(u) \geq \text{PV}_B(u),
\]
as desired. \qed

### 3. Analytic properties of PDE-constraint variations

#### 3.1. \( \Gamma \)-convergence of functionals defined by PV seminorms.

In this section we prove a \( \Gamma \)-convergence result with respect to the intensity parameter \( \alpha \) and operator \( \mathcal{B} \).

**Definition 3.1.** We define the functional \( \mathcal{I}_{\alpha, \mathcal{B}} : L^1(Q) \to [0, +\infty] \) as
\[
\mathcal{I}_{\alpha, \mathcal{B}}(u) := \begin{cases} 
\| u - u_\eta \|_{L^2(Q)}^2 + \alpha \text{PV}_B(u) & \text{if } u \in BV_B(Q), \\
+\infty & \text{otherwise}.
\end{cases}
\]

The following theorem is the main result of this section.

**Theorem 3.2.** Let sequences \( \{ B_n \}_{n=1}^{\infty} \) and \( \{ \alpha_n \}_{n=1}^{\infty} \) be given such that \( B_n \to B_0 \) in \( \ell^\infty \) and \( \alpha_n \to \alpha_0 \in \mathbb{R}^+ \). Then, the functional \( \mathcal{I}_{\alpha_n, B_n} \) \( \Gamma \)-converges to \( \mathcal{I}_{\alpha, \mathcal{B}} \) in the weak \( L^2 \) topology. To be precise, for every \( u \in L^1(Q) \) the following two conditions hold:

(Lower semi-continuity) If \( u_n \rightharpoonup u \) weakly in \( L^2(Q) \)
then
\[
\mathcal{I}_{\alpha, \mathcal{B}}(u) \leq \liminf_{n \to +\infty} \mathcal{I}_{\alpha_n, B_n}(u_n).
\]

(Recovery sequence) For each \( u \in BV(Q) \), there exists \( \{ u_n \}_{n=1}^{\infty} \subset L^1(Q) \) such that \( u_n \rightharpoonup u \) weakly in \( L^2(Q) \)
and
\[
\limsup_{n \to +\infty} I_{\alpha_n, \mathscr{B}_n}(u_n) \leq I_{\alpha, \mathscr{B}}(u).
\]

We subdivide the proof of Theorem 3.2 into two propositions.

The following proposition is instrumental for establishing the liminf inequality.

**Proposition 3.3.** 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 \( \ell^\infty \) and \( \alpha_n \to \alpha_0 \in \mathbb{R}^+ \). Let \( \{u_n\}_{n=1}^\infty \subset L^p(Q) \) be given such that there exists \( p \in (1, +\infty] \) and
\[
\sup \left\{ \|u_n\|_{L^p(Q)} + PV_{\mathscr{B}_n}(u_n) : n \in \mathbb{N} \right\} < +\infty. \tag{3.1}
\]
Then there exists \( u_0 \in BV_{\mathscr{B}_0}(Q) \) such that, up to the extraction of a subsequence (not relabeled),
\[
u_n \rightharpoonup u_0 \text{ weakly in } L^p(Q), \tag{3.2}
\]
and
\[
\liminf_{n \to \infty} PV_{\mathscr{B}_n}(u_n) \geq PV_{\mathscr{B}_0}(u_0). \tag{3.3}
\]

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

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

Next, for arbitrary \( \varphi \in C_0^\infty(Q; \mathbb{R}^K) \), we observe that
\[
\limsup_{n \to \infty} \left| \int_Q u_n \mathcal{B}_n^* \varphi dx - \int_Q u_n \mathcal{B}_0^* \varphi dx \right| \leq \limsup_{n \to \infty} \int_Q |u_n| |\mathcal{B}_n^* \varphi - \mathcal{B}_0^* \varphi| dx \leq \left( \sup_{n \geq 0} \|u_n\|_{L^p} \right) \left( \limsup_{n \to \infty} \|\mathcal{B}_n^* \varphi - \mathcal{B}_0^* \varphi\|_{L^{p'}} \right) = 0, \tag{3.4}
\]
where at the last we used the fact that \( \varphi \in C_0^\infty(Q; \mathbb{R}^K) \) and the Lebesgue dominated convergence theorem.

Hence, we could obtain that
\[
\liminf_{n \to \infty} PV_{\mathcal{B}_n}(u_n) \geq \liminf_{n \to \infty} \int_Q u_n \mathcal{B}_n^* \varphi dx \\
\geq \liminf_{n \to \infty} \int_Q u_n \mathcal{B}_0^* \varphi dx + \liminf_{n \to \infty} \int_Q u_n \left( \mathcal{B}_n^* \varphi - \mathcal{B}_0^* \varphi \right) dx \geq \int_Q u_0 \mathcal{B}_0^* \varphi dx,
\]
where at the last inequality we used (3.2) and (3.4). Thus, by the arbitrariness of \( \varphi \in C_0^\infty(Q; \mathbb{R}^K) \), we conclude (3.3), and hence the thesis. \( \square \)

**Proposition 3.4 (\( \Gamma \)-lim sup inequality).** 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 \( \ell^\infty \) and \( \alpha_n \to \alpha_0 \in \mathbb{R}^+ \). Then, for every \( u_0 \in BV_{\mathscr{B}_0}(Q) \) there exist \( \{u_n\}_{n=1}^\infty \subset BV_{\mathscr{B}_n}(Q) \) and, up to a subsequence of \( \{\mathscr{B}_n\}_{n=1}^\infty \), such that \( u_n \rightharpoonup u_0 \) in \( L^p \) and
\[
\limsup_{n \to \infty} PV_{\mathscr{B}_n}(u_n) \leq PV_{\mathscr{B}_0}(u_0).
\]
Proof. If \( PV_{\mathcal{B}_0}(u) = \infty \), there is nothing to prove. Suppose not, and assume for a moment that 
\( u_0 \in C^\infty(\bar{Q}) \), which indicates that \( u_0 \in BV_{\mathcal{B}_n}(Q) \) for each \( n \in \mathbb{N} \). Fix \( \delta > 0 \), and chose \( \varphi_{\delta,n} \in C^\infty_c(Q;\mathbb{R}^K) \) such that \( PV_{\mathcal{B}_n}(u) \leq \int_Q u\mathcal{B}_n^*\varphi_{\delta,n}dx + \delta. \)

We observe that \( \int_Q u_0\mathcal{B}_n^*\varphi_{\delta,n}dx = \int_Q [\mathcal{B}_n u_0]\varphi_{\delta,n}dx \leq \|\varphi_{\delta,n}\|_{L^\infty(Q)} \int_Q |\mathcal{B}_n u_0|dx \leq \int_Q |\mathcal{B}_n u_0|dx, \)

where at the last inequality we used the fact that \( \varphi_{\delta,n} \) satisfies (2.4). Next, since \( u \in C^\infty(\bar{Q}) \) and \( \mathcal{B}_n \to \mathcal{B} \) in \( \ell^\infty \), we have

\[
|\mathcal{B}_n u(x)| \leq \sup \{|\mathcal{B}_n|_{\ell^\infty} : n \in \mathbb{N}\} \cdot \sum_{h \leq d} |H^h u_0(x)|_{\ell^\infty},
\]

which implies that

\[
\int_Q \sum_{h \leq d} |H^h u(x)|_{\ell^\infty} dx \leq \|u\|_{W^{d,\infty}(Q)} < +\infty.
\]

Thus, we could apply the Lebesgue dominate convergence theorem to conclude that

\[
\limsup_{n \to \infty} \int_Q |\mathcal{B}_n u_0| dx \leq \int_Q \limsup_{n \to \infty} |\mathcal{B}_n u_0| dx = \int_Q |\mathcal{B}_0 u_0| dx.
\]

This, and together with (3.5) and (3.6), we observe that

\[
\limsup_{n \to \infty} PV_{\mathcal{B}_n}(u_0) \leq \limsup_{n \to \infty} \int_Q u_0\mathcal{B}_n^*\varphi_{\delta,n}dx + \delta \leq \int_Q |\mathcal{B}_0 u_0| dx + \delta = PV_{\mathcal{B}_0}(u_0) + \delta,
\]

which implies, by sending \( \delta \searrow 0 \) second, that

\[
\limsup_{n \to \infty} PV_{\mathcal{B}_n}(u_0) \leq PV_{\mathcal{B}_0}(u_0). \tag{3.7}
\]

Next, by Theorem 2.6, 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

\[
PV_{\mathcal{B}_0}(u_\varepsilon) \to PV_{\mathcal{B}_0}(u), \text{ or } PV_{\mathcal{B}_0}(u_\varepsilon) \leq PV_{\mathcal{B}_0}(u) + O(\varepsilon).
\]

Also, by (3.7), we have

\[
\limsup_{n \to \infty} PV_{\mathcal{B}_n}(u_\varepsilon) \leq PV_{\mathcal{B}_0}(u_\varepsilon) \leq PV_{\mathcal{B}_0}(u) + O(\varepsilon).
\]

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

\[
PV_{\mathcal{B}_{n_\varepsilon}}(u_\varepsilon) \leq PV_{\mathcal{B}_0}(u) + O(\varepsilon).
\]

That is, we have

\[
\limsup_{\varepsilon \to 0} PV_{\mathcal{B}_{n_\varepsilon}}(u_\varepsilon) \leq PV_{\mathcal{B}_0}(u),
\]

which concludes our thesis. \( \Box \)

We close Section 3.1 by proving Theorem 3.2.

Proof of Theorem 3.2. Property (Lower semi-continuity) hold in view of Proposition 3.3, and Property (Recovery sequence) follows from Proposition 3.4. \( \Box \)
3.2. The point-wise characterization of sub-differential of $PV_{\mathcal{B}}$. We recall few notations and preliminary results and definitions first.

**Definition 3.5** ([13, Definition 4.1 & 5.1]). Let $F$ be a function of normed space $V$ into $\mathbb{R}$ be given.

1. We define the *polar* function of $F$, denoted by $F^*$, by
   \[ F^*(u^*) = \sup\{\langle v, u^* \rangle_{V^*,V} - F(v) : v \in V \} \]

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 $\langle v - u, u^* \rangle_{V^*,V} + F(u) \leq F(v)$ for all $v \in V$. Then we call 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 $\partial F(u)$.

**Proposition 3.6** ([13, Proposition 4.1 & 5.1]). Let $F$ be a function of $V$ into $\mathbb{R}$ and $F^*$ 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 set $\partial F(u)$ (possible empty) is convex and closed.
3. If in addition $F$ is convex, then $F^{**} = F$.

**Definition 3.7.** Let $p \in [1, +\infty)$, $v \in L^p(Q; \mathbb{R}^K)$, and operator $\mathcal{B}$ be given.

1. we say that $\mathcal{B}^*v$ in $L^p(Q)$ if there exists $w \in L^p(Q)$ such that for all $\varphi \in C_c^\infty(Q; \mathbb{R}^K)$
   \[ \int_Q \mathcal{B}\varphi \cdot v \, dx = -\int_Q \varphi w \, dx. \]
2. we define the space $W^p[\mathcal{B}](Q; \mathbb{R}^K) := \{v \in L^p(Q; \mathbb{R}^K) : \mathcal{B}^*v \in L^p(Q)\}$
   with the norm $\|v\|_{W^p[\mathcal{B}]} := \|v\|_{L^p(Q)} + \|\mathcal{B}^*v\|_{L^p(Q)}$.
3. we define $W_{0,\mathcal{B}}^p(Q; \mathbb{R}^K) := \text{cl}(C_c^\infty(Q; \mathbb{R}^K)\|_{W^p[\mathcal{B}]}$, i.e., the closure of function space $C_c^\infty(Q; \mathbb{R}^N)$ with respect to $\|\|_{W^p[\mathcal{B}]}$ norm.
4. we define $C_c^\infty[\mathcal{B}](Q) := \{\mathcal{B}^*\varphi : \varphi \in C_c^\infty(Q; \mathbb{R}^K), \|\varphi\|_{L^\infty(Q)} \leq 1\}$.
   and $K^p[\mathcal{B}](Q) := \{\mathcal{B}^*v : v \in W_{0,\mathcal{B}}^p(Q; \mathbb{R}^K), \|v\|_{L^\infty(Q)} \leq 1\}$.

The main result of Section 3.2 reads as follows.

**Theorem 3.8.** Let $p > 1$, $q = p/(p-1)$, and $u \in L^p(Q)$, $\bar{u} \in L^q(Q)$. Then $\bar{u} \in \partial PV_{\mathcal{B}}(u)$ if and only if

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

\[
PV_{\mathcal{B}}(u) = \int_Q u \mathcal{B}^*v \, dx.
\]

We prove Theorem 3.8 in several propositions.

**Proposition 3.9.** Let \( p \in (1, +\infty) \) be given. Then we have the closure of function space \( C_{\mathcal{B}}(Q) \) under \( L^q \) norm equals to the function space \( W_0^q[\mathcal{B}](Q) \), i.e.,

\[
\text{cl}(C_{\mathcal{B}}(Q))_{L^q} = W_0^q[\mathcal{B}](Q).
\]

**Proof.** We claim

\[
\text{cl}(C_{\mathcal{B}}(Q))_{L^q} \subset W_0^q[\mathcal{B}](Q)
\]

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

\[
\|\mathcal{B}^*v_n - g\|_{L^q(Q)} \to 0.
\]

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

\[
v_n \rightharpoonup v_0 \text{ weakly in } L^q \text{ and } \|v_0\|_{L^\infty} \leq 1.
\]

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

\[
\int_Q \mathcal{B}^*v_n \phi \, dx = -\int_Q v_n \mathcal{B} \phi \, dx \to -\int_Q v_0 \mathcal{B} \phi \, dx,
\]

and together with (3.9), we have

\[
\int_Q g \phi \, dx = -\int_Q v_0 \mathcal{B} \phi \, dx,
\]

which implies that \( g = \mathcal{B}^*v_0 \). Thus, we have \( v_0 \in W_0^q[\mathcal{B}](Q; \mathbb{R}^K) \). Next, since the set

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

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

We next claim that

\[
\text{cl}(C_{\mathcal{B}}(Q))_{L^q} \supset W_0^q[\mathcal{B}](Q).
\]

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

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

Next, define the truncation function

\[
\tilde{v}_n := -1 \vee v_n \wedge 1,
\]

and we note that

\[
\tilde{v}_n \rightharpoonup v \text{ a.e., and } \mathcal{B}^*\tilde{v}_n \rightharpoonup \mathcal{B}^*v_0 \text{ weakly in } L^q.
\]
Using a similar argument used in Proposition 2.4, and together with \( \tilde{v}_n \to v \) a.e., we obtain that
\[
\liminf_{n \to \infty} \| B^* \tilde{v}_n \|_{L^\infty(Q)} \geq \| B^* v_0 \|_{L^\infty(Q)}.
\]

On the other hand, by (3.12), we have
\[
\| B^* \tilde{v}_n \|_{L^\infty(Q)} \leq \| B^* v_0 \|_{L^\infty(Q)},
\]
and hence
\[
\lim_{n \to \infty} \| B^* \tilde{v}_n \|_{L^\infty(Q)} = \| B^* v_0 \|_{L^\infty(Q)}.
\]
This, and together with the second part in (3.13), and using [5, Exercise 4.19, p. 124], we conclude that
\[
\lim_{n \to \infty} \| B^* \tilde{v}_n - B^* v_0 \|_{L^\infty(Q)} = 0.
\]
We next modify the sequence \( \{ \tilde{v}_n \}_{n=1}^\infty \) so that \( \{ B^* \tilde{v}_n \}_{\varepsilon > 0} \subset C_c^\infty[\mathcal{B}](Q) \). We obtain sequence of sets \( V_k, k \in \mathbb{N} \), and partition of unity \( \zeta_k \in C_c^\infty(Q) \) from the argument used in Theorem 2.6. Next, for each \( k \), we choose \( \varepsilon_k \) small enough such that
\[
\text{spt}(\eta_{\varepsilon_k} * (\tilde{v}_n \zeta_k)) \subset V_k
\]
\[
\| \eta_{\varepsilon_k} * (\tilde{v}_n \zeta_k) - \tilde{v}_n \zeta_k \|_{L^\infty(Q; \mathbb{R}^N)} < \varepsilon/2^{k+1}
\]
\[
\| \eta_{\varepsilon_k} * ((B^* \tilde{v}_n \zeta_k)) - B^* (\tilde{v}_n \zeta_k) \|_{L^\infty(Q)} < \varepsilon/2^{k+1},
\]
and in addition, we choose \( \varepsilon_k \) small that
\[
\varepsilon_k \leq \text{dist}(\partial Q, \text{spt}(\tilde{v}_n))/8. \tag{3.14}
\]
Then, we define
\[
v_{\varepsilon,n} := \sum_{k=1}^\infty \eta_{\varepsilon_k} * (\tilde{v}_n \zeta_k)
\]
and (3.14) indicates that \( v_{\varepsilon,n} \in C_c^\infty(Q) \). Following the same calculation in [14, Theorem 2, p. 125], we have that
\[
\lim_{\varepsilon \to 0} \| v_{\varepsilon,n} - \tilde{v}_n \|_{W^s[\mathcal{B}](Q)} = 0,
\]
together with (3.11), and a diagonal argument, we could construct a sequence \( \{ v_{\varepsilon,n} \}_{n=1}^\infty \) such that
\[
\lim_{\varepsilon \to 0} \| v_{\varepsilon,n} - v \|_{W^s[\mathcal{B}](Q)} = 0,
\]
Moreover, we observe that
\[
|v_{\varepsilon,n}| \leq \sum_{k=1}^\infty \eta_{\varepsilon_k} * (\tilde{v}_n \zeta_k) \leq |\tilde{v}_n| \leq 1,
\]
which proves that \( \{ B^* v_{\varepsilon,n} \}_{n=1}^\infty \subset C_c^\infty[\mathcal{B}](Q) \), and hence (3.10), and our thesis. \( \square \)

Now we are ready to prove Theorem 3.8.

**Proof of Theorem 3.8.** We first claim that the convex conjugate of \( PV_{\mathcal{B}} \), say \( PV^*_{\mathcal{B}} \), has the form that
\[
PV^*_{\mathcal{B}}(v) = I_{W^s[\mathcal{B}](Q)}(v) = \begin{cases} 0 & \text{if } v \in W^s_0[\mathcal{B}](Q) \\ +\infty & \text{if } v \notin W^s_0[\mathcal{B}](Q). \end{cases}
\]
By Definition 3.5 and Proposition 3.9, we have that
\[
I_{W^s[\mathcal{B}](Q)}^*(u) = PV_{\mathcal{B}}(u).
\]
Next, since the seminorm $PV_{\mathcal{B}}$ and indicator function $I_{c((C_c(Q)))_{L^p}}$ are convex and lower semi-continuity, we have
\[ PV_{\mathcal{B}}^*(v) = (I_{W_0^1[\mathcal{B}]}^*)^p = I_{W_0^1[\mathcal{B}]}^p. \]
Finally, in view of Proposition 3.6, we have that
\[ u^* \in \partial PV_{\mathcal{B}}(u) \]
if and only if
\[ PV_{\mathcal{B}}(u) + PV_{\mathcal{B}}^*(u^*) = \langle u, u^* \rangle, \]
and we are done. \hfill \Box

Remark 3.10. We have actually showed, in view of Proposition 3.9, that for any $v \in \dot{W}^p_0(\mathcal{B}(Q))$, $\|v\|_{L^\infty} \leq 1$, the inequality
\[ \int_Q u \mathcal{B}^* v \, dx \leq PV_{\mathcal{B}}(u) \]
holds.

Theorem 3.11 (The point-wise characterization of $\partial PV_{\mathcal{B}}$). Let $u \in L^p(Q) \cap BV_{\mathcal{B}}(Q)$, $p > 1$, be given. Let $v \in \dot{W}^p_0(\mathcal{B})(Q)$ such that $\mathcal{B}^* v \in \partial PV_{\mathcal{B}}(u)$. Then we have
\[ v = \sigma_u \text{ a.e. } x \in Q, \]
where $\sigma_u$ is the density of $\mathcal{B}u$ with respect to $|\mathcal{B}u|$ (see Remark 2.5).

Proof. Let $u \in L^p(Q) \cap BV_{\mathcal{B}}(Q)$ be given and $v \in \dot{W}^p_0(\mathcal{B})(Q)$ be obtained from Theorem 3.8. Then, by the definition of $\dot{W}^p_0(\mathcal{B})(Q)$, we could obtain a sequence $\{v_n\}_{n=1}^\infty \subset C_c^\infty(\mathcal{B}^*, Q)$ such that $\mathcal{B}^* v_n \to \mathcal{B}^* v$ strongly in $L^p$.

We claim that
\[ \|\sigma_u - v_n\|_{L^p(Q, |\mathcal{B}u|)} \to 0. \tag{3.15} \]
From the definition of $PV_{\mathcal{B}}$ and Theorem 2.5, we have that
\[ \int_Q u \mathcal{B}^* v_n \, dx = \int_Q v_n \cdot \sigma_u \, d|\mathcal{B}u|. \tag{3.16} \]
On the order hand, since $\{v_n\}_{n=1}^\infty \subset C_c^\infty(\mathcal{B}^*, Q)$, we have $\|v_n\|_{L^\infty} \leq 1$ and hence, together with the fact that $|\sigma_u| = 1$ a.e., we observe that
\[ 1 - (\sigma_u \cdot v_n) = \frac{1}{2} |\sigma_u|^2 - (\sigma_u \cdot v_n) + \frac{1}{2} |v_n|^2 + \frac{1}{2} |\sigma_u|^2 - \frac{1}{2} |v_n|^2 \]
\[ = \frac{1}{2} |\sigma_u - v_n|^2 + \frac{1}{2} |\sigma_u|^2 - \frac{1}{2} |v_n|^2 \geq \frac{1}{2} |\sigma_u - v_n|^2 \geq 0. \]
Therefore, we could compute that
\[ \int_Q |v_n - \sigma_u| \, d|\mathcal{B}u| = \int_Q 1 \cdot |v_n - \sigma_u| \, d|\mathcal{B}u| \]
\[ \leq \left( \int_Q 1 \, d|\mathcal{B}u| \right)^{1/2} \cdot \left( \int_Q |v_n - \sigma_u|^2 \, d|\mathcal{B}u| \right)^{1/2} \]
\[ \leq [PV_{\mathcal{B}}(u)]^{1/2} \cdot \left( \int_Q 1 - (\sigma_u \cdot v_n) \, d|\mathcal{B}u| \right)^{1/2}. \tag{3.17} \]
Next, from (3.16), we have that
\[
\lim_{n \to \infty} \int_Q v_n \cdot \sigma_u d|B u| = \int_Q u B^* v dx = PV_B(u) = \int_Q 1d|B u|.
\]
This, and together with (3.17), we conclude (3.15) as desired. \hfill \Box

**Proposition 3.12.** Let \( u \in BV_B(Q) \) and \( V \subseteq Q \) be given. Let \( u^* \in \partial PV_B(u) \) and \( u_V^* \in \partial PV_B(u)|_V \), then we have
\[
u^*(x) = u_V^*(x) \quad \text{for } |B u|-\text{a.e. } x \in V.
\]

**Proof.** We obtain \( v \) and \( v_V \in W^p_0[\Omega](Q; \mathbb{R}^K) \) such that Assertions 1 and 2 hold for \( PV_B(u) \) and \( PV_B(u)|_V \), respectively. Then, by Theorem 3.11 we have both \( v(x) \) and \( v_V(x) \) can be represented point-wisely by the density of \( B u \) with respect to \( |B u| \), and we are done. \hfill \Box

4. **Learning the Optimal Operator \( B \) in Imaging Processing Problems**

In this section we use the bilevel training scheme introduced in Section 1 to determine the optimal setting of \( PV_B \) for a given training pairs \( (u_c, u_q) \), where \( u_q \in L^2(Q) \) and \( u_c \in BV(Q) \) represents the corrupted and clean image, respectively.

4.1. **The bilevel training scheme with the \( PV_B \) regularizer.** We collect few notations first.

**Notation 4.1.** Recall the definition of \( B \) from Notation 2.1.

1. We denote by \( \Sigma \) the collection of operators \( B \) such that
\[
\Sigma := \{ B : |B|_{L^2} \leq 1 \}
\]

2. We denote the TrainingGround \( T \) by
\[
T := \text{cl}(\mathbb{R}^+) \times \Sigma.
\]

We state below the definition of training scheme \( T \) and associated notations.

**Definition 4.2.** We define the training scheme \( T \) with underlying training ground \( T \) by Level 1. \( (\alpha_T, B_T) \in A[T] := \arg \min \left\{ \|u_c - u_{\alpha, B}\|_{L^2(Q)} : (\alpha, B) \in T \right\}, \) (T-L1)

Level 2. \( u_{\alpha, B} := \arg \min \left\{ \|u - u_q\|_{L^2} + \alpha PV_B(u), \ u \in L^1(Q) \right\}. \) (T-L2)

In particular, for the case that \( \alpha = +\infty \), we define
\[
u_{+\infty} := \arg \min \left\{ \|u - u_c\|_{L^2(Q)} : u \in \mathcal{N} \right\} \quad \text{where } \mathcal{N} := \text{conv} \left( \bigcup_{B \in \Sigma} \mathcal{N}(B) \right). \tag{4.1}
\]

In (T-L1), we denote by notation \( A[T] \) the collection of optimal solution(s) of scheme \( T \) with underlying training ground \( T \), and \( (\alpha_T, B_T) \in A[T] \) is an optimal solution obtained from training ground \( T \).

We first show that the Level 2 problem (T-L2) admits a unique solution.

**Proposition 4.3.** Let \( \alpha \in \mathbb{R}^+ \) and \( B \in \Sigma \) be given. Then, there exists a unique \( u_{\alpha, B} \in BV_B(Q) \) such that
\[
u_{\alpha, B} = \arg \min \left\{ \|u - u_q\|_{L^2(Q)} + \alpha PV_B(u), \ u \in L^1(Q) \right\}.
\]

**Proof.** The proof can be obtained by Proposition 2.8 and the fact that \( PV_B \) is convex. \hfill \Box
Theorem 4.4. Let the training ground $T$ be given. Then the training scheme $T$ admits at least one solution $(\alpha_T, B_T) \in T$, and provides an associated optimally reconstructed image $u_{\alpha_T, B_T} \in BV_B(T)$.

Proof. Let $\{\alpha_n, B_n\}_{n=1}^\infty \subset T$ be a minimizing sequence obtained from (T-L1). Then, by the boundedness and closedness of $\Sigma$ in $\ell^\infty$, up to a subsequence (not relabeled), there exists $(\alpha_T, B_T) \in cl(\mathbb{R}^+) \times \Sigma$ such that $\alpha_n \to \alpha_T$ in $\mathbb{R}$, $B_n \to B_T$ in $\ell^\infty$, and

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

(4.2)

We divide our arguments into three cases.

Case 1: Assume $\alpha_T > 0$. Then, in view of Theorem 3.2 and the properties of $\Gamma$-convergence, we have

$$
u_{\alpha_n, B_n} \to u_{\alpha_T, B_T} \text{ weakly in } L^2(Q),$$

where $u_{\alpha_n, B_n}$ and $u_{\alpha_T, B_T}$ are obtained from (T-L2). Thus, we deduce that

$$
\|u_{\alpha_T, B_T} - u_c\|_{L^2(Q)} \leq \liminf_{n \to \infty} \|u_{\alpha_n, B_n} - u_c\|_{L^2(Q)} = m,
$$

which completes the thesis.

Case 2: Assume $\alpha_T = 0$. Then by (4.2), up to a subsequence, there exists $\bar{u} \in L^2(Q)$ such that $u_{\alpha_n, B_n} \to \bar{u}$ weakly in $L^2$. We claim that $u_{\alpha_n, B_n} \to u_\eta$ in $L^2$ strong. Extend $u_\eta$ by zero outside $Q$ and we define

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

where $\eta_\varepsilon$ is the standard mollifier. 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 (T-L2), we have

$$
\|u_{\alpha_n, B_n} - u_\eta\|_{L^2(Q)}^2 + \alpha_n \text{PV}_{B_n}(u_{\alpha_n, B_n}) \\
\leq \|u_\eta^\varepsilon - u_\eta\|_{L^2(Q)}^2 + \alpha_n \text{PV}_{B_n}(u_\eta^\varepsilon) \leq \|u_\eta^\varepsilon - u_\eta\|_{L^2(Q)}^2 + \alpha_n N d \|u_\eta^\varepsilon\|_{W^{1,1}(\mathbb{R}^N)}.
$$

That is, we have

$$
\|u_{\alpha_n, B_n} - u_\eta\|_{L^2(Q)}^2 \leq \|u_\eta^\varepsilon - u_\eta\|_{L^2(Q)}^2 + \alpha_n N d \|u_\eta^\varepsilon\|_{W^{1,1}(\mathbb{R}^N)},
$$

and we are done by letting $\alpha_n \to 0$ first and $\varepsilon \to 0$ second.

Case 3: Assume $\alpha_T = +\infty$. Reasoning as in Case 2, we have again that there exists $\bar{u} \in L^2(Q)$ such that $u_{\alpha_n, B_n} \to \bar{u}$ and $\text{PV}_{B_n}(\bar{u}) = 0$.

Then, in this case we have, by (4.1), that

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

as desired. $\square$
4.2. Numerical realization and finite approximation of scheme $\mathcal{T}$. For the numerical realization of training scheme $\mathcal{T}$, we in addition require that the training ground $\mathcal{T}$ satisfies the following assumption.

**Assumption 4.5.** Let the order $d \in \mathbb{N}$ be given.

1. We assume the intensity parameter $\alpha$ satisfies the box-constraint (see, e.g. [2, 11]). That is, there exists a constant $P \in \mathbb{R}^+$, chosen by user, such that $\alpha \in [0, P]$.
2. We assume the collection $\Sigma$ of operator $\mathcal{B}$ satisfies the following two conditions.
   a. Each operator $\mathcal{B} \in \Sigma$ has at most order $d$ (the box-constraint on order of $\mathcal{B}$);
   b. For any $\mathcal{B}_1, \mathcal{B}_2 \in \Sigma$, the continuity assumption
      \begin{align*}
      |PV_{\mathcal{B}_1}(u) - PV_{\mathcal{B}_2}(u)| \leq O(\|\mathcal{B}_1 - \mathcal{B}_2\|_{\infty}) PV_{\mathcal{B}_1}(u)
      \tag{4.3}
      
      \text{and}
      
      |PV_{\mathcal{B}_1}(u) - PV_{\mathcal{B}_2}(u)| \leq O(\|\mathcal{B}_1 - \mathcal{B}_2\|_{\infty}) PV_{\mathcal{B}_2}(u)
      \tag{4.4}
      \end{align*}

   holds.

The following corollary is a direct consequence of Theorem 4.4

**Corollary 4.6.** The training scheme $\mathcal{T}$, with a underlying training ground $\mathcal{T}$ satisfies Assumption (4.5), admits at least one solution $(\alpha_T, \mathcal{B}_T) \in \mathcal{T}$, and provides an associated optimally reconstructed image $u_{\alpha_T, \mathcal{B}_T} \in BV_{\mathcal{B}_T}(Q)$.

*Proof.* The argument is identical to the argument used in Theorem 4.4, Case 1 & Case 2. □

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

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

As discussed in Section 1, the Level 1 problem ($\mathcal{T}$-L1) for scheme $\mathcal{T}$ is equivalent to find global minimizers of $A(\alpha, \mathcal{B})$ among the training ground $\mathcal{T}$. However, in view of the counter-example provided in [22], the assessment function $A(\cdot)$ is not convex, and hence the traditional methods like Newton’s descent or Line search could trapped into local minimums, but not convergence to global minimums.

We overcome this problem by using a finite approximation method original introduced in [22]. Recall the constant $P > 0$ given in box-constraint stated in Assumption 4.5.

**Definition 4.7** (The Finite TrainingGround and Finite Grid). Let $l \in \mathbb{N}$ be given.

1. We define the step size $\delta_l$ by
   $$\delta_l := P/l;$$
2. we define the finite set $T_l[P] \subset [0, P]$ via
   $$T_l[P] := \{0, \delta_l, 2\delta_l, \ldots, i\delta_l, \ldots, P\}.$$ 
3. we define the finite set $T_l[\Sigma] \subset \Sigma$ via
   $$T_l[\Sigma] := \bigcup_{k \geq 1} T_k[\Sigma]$$

where each $T_k[\Sigma]$ is a singleton contains one operator $\mathcal{B} \in \Sigma$ and defined recursively in the following steps.
Step 1. Define
$$B_0 \in \text{arg min}\{\|B\| : B \in \Sigma\}, \ T_0[\Sigma] = \{B_0\}, \text{ and } \Sigma_0 := \Sigma.$$
We also denote $Q_l[B_0] \subset \Sigma_0$ be the cube centered at $B_0$ with side length $\Delta_l$.

Step 2. Define
$$\Sigma_1 := \Sigma_0 \setminus Q_l[B_0], \ B_1 \in \text{arg min}\{\|B\| : B \in \Sigma_1\},$$
and
$$T_1[\Sigma] := \{B_1\};$$

Step $j$. Define
$$\Sigma_k := \Sigma_{k-1} \setminus Q_l[B_{k-1}], \ B_j \in \text{arg min}\{\|B\| : B \in \Sigma_j\},$$
and
$$T_k[\Sigma] := \{B_j\}.$$
Repeat until $\Sigma_k = \emptyset$.

4. we define the Finite TrainingGround $T_l$ at step $l \in \mathbb{N}$ by
$$T_l := T_l[\alpha] \times T_l[\Sigma].$$

5. for $i, j \in \mathbb{N}$, we define the $(i,j)$-th FiniteGrid at step $l$ by
$$G_l(i,j) := [i\Delta_l, (i+1)\Delta_l] \times Q_l[B_j].$$

Remark 4.8. From the definition of $\Sigma_k$ and $Q_l$, we have, for any $l \in \mathbb{N}$ fixed, there exists an upper bound $M \in \mathbb{N}$, depends on $l$, such that $\Sigma_M = \emptyset$. In another word, we have $\mathcal{H}^0(T_l[B]) < +\infty$ and hence
$$\mathcal{H}^0(T_l) < +\infty,$$ for each $l \in \mathbb{N}$ fixed.

Then, the optimal parameters of scheme $\mathcal{T}$ (global minimizers of $\mathcal{A}(\cdot, \cdot)$) over finite training ground $T_l$
$$(\alpha_{T_l}, B_{T_l}) \in \mathcal{A}[T_l] := \text{arg min} \left\{\|u_{\alpha,B} - u_c\|_{L^2(Q)}^2 : (\alpha, B) \in T_l\right\},$$
can be determined exactly by evaluating $\mathcal{A}(\cdot)$ over each elements of $T_l$.

The main result of Section 4.2 reads as follows.

Theorem 4.9 (finite approximation and error estimation). Let a training ground $T$ satisfy Assumption 4.5 be given, and $T_l \subset T$ be constructed as in Definition 4.7. Then the following assertions hold:

1. as $l \to \infty$, we have
$$\text{dist}(\mathcal{A}[T], \mathcal{A}[T_l]) \to 0.$$  

2. let $\delta > 0$ be given. Then for each $l \in \mathbb{N}$ we have the following estimation hold
$$\mathcal{A}(\alpha_{T_l}, B_{T_l}) - \mathcal{A}(\alpha_T, B_T) \leq 4KP[O(P/l) + 1/l]^{1/2} \|u_\eta\|_{W^{2,1}(Q)}^{1/2} / \delta^d + \delta/2,$$

for any $(\alpha_T, B_T) \in \mathcal{A}[T]$ and $(\alpha_{T_l}, B_{T_l}) \in \mathcal{A}[T_l]$.

We sub-divide our argument into Section 4.2.1 and Section 4.2.2, in which we discuss the properties of reconstructed image $u_{\alpha,B}$ with $B \in \Sigma$ fixed and $\alpha \in \mathbb{R}^+$ fixed, respectively.
4.2.1. Properties of reconstructed image $u_{\alpha, \mathcal{B}}$ with respect to $\alpha \in \mathbb{R}^+$. Since $\mathcal{B} \in \Sigma$ is fixed, we abbreviate $u_{\alpha, \mathcal{B}}$ and $PV_{\mathcal{B}}$ by $u_\alpha$ and $PV$, respectively, in Section 4.2.1.

**Proposition 4.10.** We collect two auxiliary results in this proposition.
1. The function $g(\alpha) := PV(u_\alpha)$ is continuous decreasing;
2. Assume in addition that $PV(u_\alpha) > PV(u_c)$.

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

$$\|u_\alpha - u_c\|_{L^2(Q)} < \|u_\eta - u_c\|.$$ (4.8)

**Proof.** We show Assertion 1 first. The continuity of $g(\alpha)$ can be deduced from Theorem 3.2. Next, let $0 \leq \alpha_1 < \alpha_2 < +\infty$ be given, we observe, from the optimality condition of (T-L2), that

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

and

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

Adding up the previous two inequalities yields

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

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

Now we claim Assertion 2. From Theorem 3.8, we have $\partial PV(u_\alpha)$, the sub-differential of $PV$ at $u_\alpha$, is well defined. We observe, for any $\alpha > 0$, that

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

$$= 2 \langle u_\eta - u_\alpha, u_\alpha - u_c \rangle + \|u_\eta - u_\alpha\|_{L^2(Q)}^2$$

$$= 2\alpha \langle \partial PV(u_\alpha), u_\alpha - u_c \rangle + \|u_\eta - u_\alpha\|_{L^2(Q)}^2$$

$$\geq 2\alpha [PV(u_\alpha) - PV(u_c)] + \|u_\eta - u_\alpha\|_{L^2(Q)}^2,$$

where at the last inequality we use the property of sub-differential operator, and we obtain that

$$\|u_\eta - u_c\|_{L^2(Q)}^2 - \|u_\alpha - u_c\|_{L^2(Q)}^2 \geq 2\alpha [PV(u_\alpha) - PV(u_c)] + \|u_\eta - u_\alpha\|_{L^2(Q)}^2.$$ (4.10)

Next, in view of Assertion 1, we have that $PV(u_\alpha)$ is continuous decreasing and hence, together with (4.8), there exists $\bar{\alpha} > 0$ such that

$$PV(u_{\bar{\alpha}}) - PV(u_c) \geq \frac{1}{4}[PV(u_\eta) - PV(u_c)] > 0.$$ (4.11)

Hence, we conclude (4.9) by combining (4.10) and (4.11).

**Proposition 4.11.** Let $\alpha_1$ and $\alpha_2 \in \mathbb{R}^+$ be given. Then we have that

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

as desired.
Proof. Without lose of generality we assume that $\alpha_1 < \alpha_2$. In view of Theorem 3.8, and from the optimality condition of (T-L2) we have

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

Subtracting one from another and multiplying with $u_{\alpha_1} - u_{\alpha_2}$ and integration over $Q$, we obtain that

$$\|u_{\alpha_1} - u_{\alpha_2}\|_{L^2(Q)}^2 = \alpha_1 \langle \partial PV(u_{\alpha_2}) - \partial PV(u_{\alpha_1}), u_{\alpha_1} - u_{\alpha_2} \rangle + (\alpha_2 - \alpha_1) \langle \partial PV(u_{\alpha_2}), u_{\alpha_1} - u_{\alpha_2} \rangle.$$  \hfill (4.12)

Since the seminorm $PV$ is proper, l.s.c., and convex, we have $\partial PV$ is a monotone maximal operator and hence

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

This, together with (4.12) and Assertion 1 from Proposition 4.10, we obtain that

$$\|u_{\alpha_1} - u_{\alpha_2}\|_{L^2(Q)}^2 \leq (\alpha_2 - \alpha_1) \langle \partial PV(u_{\alpha_2}), u_{\alpha_1} - u_{\alpha_2} \rangle \leq (\alpha_2 - \alpha_1) \|PV(u_{\alpha_1}) + PV(u_{\alpha_2})\|_{L^1(Q)},$$

where at the second last inequality we used Remark 3.10, and hence the thesis. \hfill \Box

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

Moreover, in Section 4.2.2, we further restrict the corrupted image $u_\eta \in L^2(Q)$ satisfies that there exists $0 < M_1 < M_2 < +\infty$ such that

$$0 < M_1 \leq u_\eta(x) \leq M_2 < +\infty, \text{ for a.e. } x \in Q. \hfill (4.13)$$

In this way, we have that the reconstructed image

$$u_B = \arg\min \left\{ \|u - u_\eta\|_{L^2(Q)}^2 + PV_B(u) : u \in L^1(Q) \right\}$$

also satisfies that

$$M_1 \leq u_B(x) \leq M_2, \text{ for a.e. } x \in Q. \hfill (4.14)$$

Before we move to next proposition, we call the following result regarding the Lebesgue point.

**Theorem 4.12** (Lebesgue-besicovitch differentiation theorem). Let $\mu$ be a Radon measure on $\mathbb{R}^N$ and $f \in L^1_{\text{loc}}(\mathbb{R}^N, \mu)$. Then

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

for $\mu$ a.e. $x \in \mathbb{R}^N$.

**Proposition 4.13.** Let $u_\eta \in L^2(Q)$ satisfies (4.13) be given. Let $B_1$ and $B_2 \in \Sigma$ be given. Then we have that

$$\|u_{B_1} - u_{B_2}\|_{L^2(Q)}^2 \leq O(|B_1 - B_2|_{L^\infty}) \left[ PV_{B_1}(u_{\alpha_1}) + PV_{B_1}(u_{\alpha_2}) \right],$$

where $u_{B}$ is defined in (T-L2).
Proof. By Theorem 3.8, we have the sub-differential $\partial PV_{\mathcal{B}_1}$ and $\partial PV_{\mathcal{B}_2}$ are well defined. Then, by the optimality condition of (T-L2) we have that

$$u_{\mathcal{B}_1} - u_\eta = -\partial PV_{\mathcal{B}_1}(u_{\mathcal{B}_1})$$

and

$$u_{\mathcal{B}_2} - u_\eta = -\partial PV_{\mathcal{B}_2}(u_{\mathcal{B}_2}).$$

Subtracting one from another, we have that

$$u_{\mathcal{B}_1} - u_{\mathcal{B}_2} = \partial PV_{\mathcal{B}_2}(u_{\mathcal{B}_2}) - \partial PV_{\mathcal{B}_1}(u_{\mathcal{B}_1})$$

$$= \partial PV_{\mathcal{B}_2}(u_{\mathcal{B}_2}) - \partial PV_{\mathcal{B}_2}(u_{\mathcal{B}_1}) + \partial PV_{\mathcal{B}_2}(u_{\mathcal{B}_1}) - \partial PV_{\mathcal{B}_1}(u_{\mathcal{B}_1}).$$

Multiplying both side by $u_{\mathcal{B}_1} - u_{\mathcal{B}_2}$ and integrate over $Q$, we obtain that

$$\|u_{\mathcal{B}_2} - u_{\mathcal{B}_1}\|_{L^2(Q)}^2 = -\langle \partial PV_{\mathcal{B}_2}(u_{\mathcal{B}_2}) - \partial PV_{\mathcal{B}_2}(u_{\mathcal{B}_1}), u_{\mathcal{B}_2} - u_{\mathcal{B}_1} \rangle$$

$$+ \langle \partial PV_{\mathcal{B}_2}(u_{\mathcal{B}_1}) - \partial PV_{\mathcal{B}_1}(u_{\mathcal{B}_1}), u_{\mathcal{B}_1} - u_{\mathcal{B}_2} \rangle. \quad (4.15)$$

Since $PV_{\mathcal{B}_2}$ is convex, $\partial PV_{\mathcal{B}_2}$ is a monotone maximal operator. Therefore, we have

$$\langle \partial PV_{\mathcal{B}_2}(u_{\mathcal{B}_2}) - \partial PV_{\mathcal{B}_2}(u_{\mathcal{B}_1}), u_{\mathcal{B}_2} - u_{\mathcal{B}_1} \rangle \geq 0. \quad (4.16)$$

We next estimate the second part of (4.15). Firstly, from the definition of sub-gradient, we have

$$\langle \partial PV_{\mathcal{B}_2}(u_{\mathcal{B}_1}) - \partial PV_{\mathcal{B}_1}(u_{\mathcal{B}_1}), u_{\mathcal{B}_2} - u_{\mathcal{B}_1} \rangle = PV_{\mathcal{B}_2}(u_{\mathcal{B}_2}) - PV_{\mathcal{B}_1}(u_{\mathcal{B}_1}) \leq cPV_{\mathcal{B}_1}(u_{\mathcal{B}_1}). \quad (4.17)$$

where

$$c := O \left( |\mathcal{B}_1 - \mathcal{B}_2|_{L^\infty} \right)$$

is the constant used in (4.3). Moreover, from (4.3) we also deduce that

$$-c|\mathcal{B}_2 u_{\mathcal{B}_1}| \leq |\mathcal{B}_2 u_{\mathcal{B}_1}|(V) - |\mathcal{B}_1 u_{\mathcal{B}_1}|(V) \leq c|\mathcal{B}_2 u_{\mathcal{B}_1}|. \quad (4.18)$$

Next, Let $v_{\mathcal{B}_1}$ and $v_{\mathcal{B}_2}$ be obtained from Proposition 3.8 as the sub-differential of $PV_{\mathcal{B}_1}(u_{\mathcal{B}_1})$ and $PV_{\mathcal{B}_2}(u_{\mathcal{B}_1})$, respectively. Then, by Proposition 3.12, for any open set $V \subset Q$ we have that

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

This, and together with (4.18), conclude

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

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

$$-c\int_{Q(x, \delta)} u_{\mathcal{B}_1} [\mathcal{B}_2^* v_{\mathcal{B}_2}] \, dx \leq \int_{Q(x, \delta)} (u_{\mathcal{B}_1} [\mathcal{B}_2^* v_{\mathcal{B}_2}] - u_{\mathcal{B}_1} [\mathcal{B}_1^* v_{\mathcal{B}_1}]) \, dx$$

$$= \int_{Q(x, \delta)} u_{\mathcal{B}_1} [\mathcal{B}_2^* v_{\mathcal{B}_2}] \, dx - \int_{Q(x, \delta)} u_{\mathcal{B}_1} [\mathcal{B}_1^* v_{\mathcal{B}_1}] \, dx$$

$$\leq c\int_{Q(x, \delta)} u_{\mathcal{B}_1} [\mathcal{B}_1^* v_{\mathcal{B}_1}] \, dx.$$

By Assertion 2, Theorem 3.8, we have $\mathcal{B}_1^* v_{\mathcal{B}_1} \in L^1(Q)$. Since $u_{\mathcal{B}_1} \in L^\infty(Q)$, we have $u_{\mathcal{B}_1}, \mathcal{B}_1^* v_{\mathcal{B}_1} \in L^1(Q)$. Thus, we could apply the Lebesgue point in Theorem 4.12 and take $\delta \to 0$ to conclude that

$$-c u_{\mathcal{B}_1} [\mathcal{B}_2^* v_{\mathcal{B}_2}] \leq u_{\mathcal{B}_1} [\mathcal{B}_2^* v_{\mathcal{B}_2}] - u_{\mathcal{B}_1} [\mathcal{B}_1^* v_{\mathcal{B}_1}] \leq c u_{\mathcal{B}_1} [\mathcal{B}_1^* v_{\mathcal{B}_1}],$$

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

$$-c u_{\mathcal{B}_1} [\mathcal{B}_2^* v_{\mathcal{B}_2}] \leq u_{\mathcal{B}_1} [\mathcal{B}_2^* v_{\mathcal{B}_2} - \mathcal{B}_1^* v_{\mathcal{B}_1}] \leq c u_{\mathcal{B}_1} [\mathcal{B}_1^* v_{\mathcal{B}_1}],$$
and together with the fact that \( u_{\mathcal{B}} \geq 1 \) (see (4.14)), we deduce that

\[
-c \left\| \mathcal{B}^* u_{\mathcal{B}} - \mathcal{B}^* v_{\mathcal{B}} \right\| \leq \left\| \mathcal{B}^* v_{\mathcal{B}} - \mathcal{B}^* v_{\mathcal{B}} \right\| \leq c \left\| \mathcal{B}^* v_{\mathcal{B}} \right\|
\]

for a.e. \( x \in Q \).

On the other hand, again by (4.14), we have that

\[
-\mathcal{B}^* u_{\mathcal{B}} - 2M_2 > 1, \quad \text{and hence}
\]

\[
\left\langle \partial PV_{\mathcal{B}}(u_{\mathcal{B}}), \mathcal{B}^* v_{\mathcal{B}} - u_{\mathcal{B}} \right\rangle = \left\langle \partial PV_{\mathcal{B}}(u_{\mathcal{B}}), \mathcal{B}^* v_{\mathcal{B}} - u_{\mathcal{B}} \right\rangle
\]

\[
= \left\langle \partial PV_{\mathcal{B}}(u_{\mathcal{B}}), \mathcal{B}^* u_{\mathcal{B}} + 2M_2 - 2M_2 \right\rangle
\]

\[
= \left\langle \partial PV_{\mathcal{B}}(u_{\mathcal{B}}), \mathcal{B}^* u_{\mathcal{B}} - 2M_2 \right\rangle
\]

\[
\leq c \left( \left\| \mathcal{B}^* v_{\mathcal{B}} \right\| - u_{\mathcal{B}} + 2M_2 \right) + \left( \partial PV_{\mathcal{B}}(u_{\mathcal{B}}) - \partial PV_{\mathcal{B}}(u_{\mathcal{B}}), -2M_2 \right).
\]

Note that, as \( v_{\mathcal{B}} \in W^0_{2,0} \), by Remark 3.10 we observe that

\[
\left\langle \mathcal{B}^* v_{\mathcal{B}} - u_{\mathcal{B}} + 2M_2 \right\rangle \leq PV_{\mathcal{B}}(u_{\mathcal{B}}) - u_{\mathcal{B}} + 2M_2 \}
\]

and, since the constants belongs to the kernel of \( PV_{\mathcal{B}} \),

\[
\left\langle \partial PV_{\mathcal{B}}(u_{\mathcal{B}}) - \partial PV_{\mathcal{B}}(u_{\mathcal{B}}), -2M_2 \right\rangle = 0.
\]

Therefore, by combing (4.19), (4.20), and (4.21), we obtain that

\[
\left\langle \partial PV_{\mathcal{B}}(u_{\mathcal{B}}) - \partial PV_{\mathcal{B}}(u_{\mathcal{B}}), -u_{\mathcal{B}} \right\rangle \leq c PV_{\mathcal{B}}(u_{\mathcal{B}}).
\]

This, together with (4.15), (4.16), and (4.17), we conclude our thesis. \( \square \)

4.2.3. \( L^2 \)-distance estimation of reconstructed image \( u_{\alpha, \mathcal{B}} \). We start with a relaxation result regarding to the corrupted image \( u_\eta \).

**Proposition 4.14.** 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 \( (\alpha, \mathcal{B}) \in \mathbb{T} \), define

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

Then we have

\[
\| u_{\alpha, \mathcal{B}} - u_{\alpha, \mathcal{B}} \|_{L^2(Q)} \leq \| u_\eta - u_\eta \|_{L^2(Q)}
\]

and

\[
\lim_{\varepsilon \to 0} PV_{\mathcal{B}}(u_{\alpha, \mathcal{B}}) = PV_{\mathcal{B}}(u_{\alpha, \mathcal{B}}),
\]

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

**Proof.** From the optimality condition of (4.22) and (T-L2), we have

\[
u_{\alpha, \mathcal{B}} - u_{\alpha, \mathcal{B}} + u_\eta - u_\eta = \alpha \partial PV_{\mathcal{B}}(u_{\alpha, \mathcal{B}}) - \alpha \partial PV_{\mathcal{B}}(u_{\alpha, \mathcal{B}}).
\]

Multiplying \( u_{\alpha, \mathcal{B}} - u_{\alpha, \mathcal{B}} \) on the both hand side, we have

\[
\left\| u_{\alpha, \mathcal{B}} - u_{\alpha, \mathcal{B}} \right\|_{L^2(Q)} = \left\langle u_{\alpha, \mathcal{B}} - u_{\alpha, \mathcal{B}} - \alpha \partial PV_{\mathcal{B}}(u_{\alpha, \mathcal{B}}), u_{\alpha, \mathcal{B}} - u_{\alpha, \mathcal{B}} \right\rangle \leq 0,
\]

where at the last inequality we used the fact that \( \partial PV_{\mathcal{B}} \) is a maximal monotone operator, and we conclude (4.23) as desired.
We next claim (4.24). We assume that \( \alpha \in \mathbb{R}^+ \), otherwise there is nothing to prove. By (4.23), we have that

\[
u_{\alpha, \mathcal{B}} \to u^*_{\alpha, \mathcal{B}} \text{ strongly in } L^2.
\]

(4.25)

This, and together with Proposition 2.4, we deduce that

\[
\liminf_{\varepsilon \to 0} PV_\mathcal{B}(\nu_{\alpha, \mathcal{B}}) \geq PV_\mathcal{B}(u_{\alpha, \mathcal{B}}).
\]

(4.26)

On the other hand, in view of the optimality condition of (4.22) again, we have

\[
\|u_{\alpha, \mathcal{B}} - u^*_\varepsilon\|_{L^2} + \alpha PV_\mathcal{B}(u^*_\varepsilon) \leq \|u_{\alpha, \mathcal{B}} - u^*_\eta\|_{L^2} + \alpha PV_\mathcal{B}(u_{\alpha, \mathcal{B}}),
\]

or

\[
\alpha PV_\mathcal{B}(u^*_\varepsilon) \leq \|u_{\alpha, \mathcal{B}} - u^*_\eta\|_{L^2} - \|u_{\alpha, \mathcal{B}} - u^*_\varepsilon\|_{L^2} + \alpha PV_\mathcal{B}(u_{\alpha, \mathcal{B}}).
\]

Hence, by (4.25), we have that

\[
\limsup_{\varepsilon \to 0} \alpha PV_\mathcal{B}(u^*_\varepsilon)
\leq \limsup_{\varepsilon \to 0} \left[ \|u_{\alpha, \mathcal{B}} - u^*_\eta\|_{L^2} - \|u_{\alpha, \mathcal{B}} - u^*_\varepsilon\|_{L^2} + \alpha PV_\mathcal{B}(u_{\alpha, \mathcal{B}}) \right]
= \alpha PV_\mathcal{B}(u_{\alpha, \mathcal{B}}).
\]

This, and (4.26), allows us to conclude (4.24) as desired. \( \square \)

We next present an improved version of Proposition 4.13, in which we remove the assumption that \( u_\eta \) need to satisfy the boundness assumption (4.13).

**Corollary 4.15.** Let \( u_\eta \in L^2(Q) \), \( \alpha \in \mathbb{R}^+ \), and \( \mathcal{B}_1, \mathcal{B}_2 \in \Sigma \) be given. Then the following estimation holds.

\[
\|u_{\alpha, \mathcal{B}_1} - u_{\alpha, \mathcal{B}_2}\|_{L^2(Q)}^2 \leq \alpha \cdot O(\|B_1 - B_2\|_{L^\infty}) [PV_{\mathcal{B}_1}(u_{\alpha, \mathcal{B}_1}) + PV_{\mathcal{B}_2}(u_{\alpha, \mathcal{B}_2})],
\]

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

**Proof.** Let \( M \in \mathbb{N} \) be given, and define

\[
u^M_\eta := \min \{ \|u - \eta\|_{L^2} + \alpha PV_\mathcal{B}(\eta) : u \in L^1(Q) \}.
\]

Also, we define that

\[
u^M_{\alpha, \mathcal{B}} := \arg \min \left\{ \|u - \nu^M_\eta\|_{L^2} + \alpha PV_\mathcal{B}(\nu^M_\eta) : u \in L^1(Q) \right\}
\]

(4.27)

and

\[
\tilde{u}^M_{\alpha, \mathcal{B}} := \arg \min \left\{ \|u - (\nu^M_\eta + 2M)\|_{L^2} + \alpha PV_\mathcal{B}(\nu^M_\eta) : u \in L^1(Q) \right\}.
\]

(4.28)

We claim that

\[
\tilde{u}^M_{\alpha, \mathcal{B}} = u^M_{\alpha, \mathcal{B}} + 2M.
\]

(4.29)

We observe that

\[
\|u^M_{\alpha, \mathcal{B}} - u^*_\eta\|_{L^2} + \alpha PV_\mathcal{B}(u^M_{\alpha, \mathcal{B}})
\leq \|u^M_{\alpha, \mathcal{B}} + 2M - (u^M_\eta + 2M)\|_{L^2} + \alpha PV_\mathcal{B}(u^M_{\alpha, \mathcal{B}}) + 2M
= \|u^M_{\alpha, \mathcal{B}} - u^*_\eta\|_{L^2} + \alpha PV_\mathcal{B}(u^M_{\alpha, \mathcal{B}})
\leq \|u^M_{\alpha, \mathcal{B}} - u^*_\eta\|_{L^2} + \alpha PV_\mathcal{B}(u^M_{\alpha, \mathcal{B}})
= \|u^M_{\alpha, \mathcal{B}} - (u^M_\eta + 2M)\|_{L^2} + \alpha PV_\mathcal{B}(u^M_{\alpha, \mathcal{B}}),
\]

(4.28)
where at the first inequality we used the optimality condition on (4.28), and at the last inequality we used the optimality condition on (4.27). Thus, we have
\[
\|\bar{u}_{\alpha, B}^M - (u_{\eta}^M + 2M)\|_{L^2}^2 + \alpha PV_B(\bar{u}_{\alpha, B}^M) = \|u_{\alpha, B}^M + 2M - (u_{\eta}^M + 2M)\|_{L^2}^2 + \alpha PV_B(u_{\alpha, B}^M + 2M),
\]
and we conclude (4.29) in view of the uniqueness of the minimizer. Thus, we have
\[
\|u_{\alpha, B}^M - u_{\eta, B}^M\|_{L^2(Q)} = \|\bar{u}_{\alpha, B}^M - \bar{u}_{\eta, B}^M\|_{L^2(Q)},
\]
Therefore, we could assume that, without lose of generality, \(u_{\eta}^M \geq M > 0\). In another word, we have \(u_{\eta}^M\) satisfies (4.13).

Next, by the optimality condition of (4.27), we have that
\[
\frac{1}{\alpha}(u_{\alpha, B}^M - u_{\eta}) = -\partial PV_B(u_{\alpha, B}^M) \text{ and } \frac{1}{\alpha}(u_{\alpha, B}^M - u_{\eta}) = -\partial PV_B(u_{\alpha, B}^M).
\]
Following exactly the same argument used in Proposition 4.13 (in (4.19) use 2M instead of \(M_2\)), we obtain that
\[
\frac{1}{\alpha} \|u_{\alpha, B}^M - u_{\alpha, B}^M\|_{L^2(Q)}^2 \leq O(\|B - B\|_{L^\infty}) \left[ PV_B(u_{\alpha, B}^M) + PV_B(u_{\alpha, B}^M) \right].
\]
In the end, we compute that
\[
\frac{1}{\alpha} \|u_{\alpha, B}^M - u_{\alpha, B}^M\|_{L^2(Q)}^2 \\
\leq \frac{1}{\alpha} \|u_{\alpha, B}^M - u_{\alpha, B}^M\|_{L^2(Q)}^2 + \frac{1}{\alpha} \|u_{\alpha, B}^M - u_{\alpha, B}^M\|_{L^2(Q)}^2 \\
\leq O(\|B - B\|_{L^\infty}) \left[ PV_B(u_{\alpha, B}^M) + PV_B(u_{\alpha, B}^M) \right] \\
+ \frac{1}{\alpha} \|u_{\alpha, B}^M - u_{\alpha, B}^M\|_{L^2(Q)}^2 + \|u_{\alpha, B}^M - u_{\alpha, B}^M\|_{L^2(Q)}^2.
\]
Then, by Proposition 4.14, in which \(u_{\eta}^M\) is replaced by \(u_{\eta}^M\), we conclude our thesis by sending \(M \to \infty\) on the right hand side on the above inequality. \(\square\)

**Proposition 4.16.** Let \((\alpha_1, B_1)\) and \((\alpha_2, B_2) \in \mathbb{T}\) be given. Then we have
\[
\|u_{\alpha_1, B_1} - u_{\alpha_2, B_2}\|_{L^2(Q)}^2 \\
\leq 4 \alpha_1 O(\|B_2 - B_2\|_{L^\infty}) + |\alpha_1 - \alpha_2| \left[ PV_B(u_{\alpha_1, B_1}) + PV_B(u_{\alpha_1, B_2}) \right].
\]

**Proof.** We compute that
\[
\|u_{\alpha_1, B_1} - u_{\alpha_2, B_2}\|_{L^2(Q)}^2 \\
\leq 2\alpha_1 O(\|B_1 - B_2\|_{L^\infty}) \left[ PV_B(u_{\alpha_1, B_1}) + PV_B(u_{\alpha_1, B_2}) \right] + 4 |\alpha_1 - \alpha_2| \left[ PV_B(u_{\alpha_1, B_1}) + PV_B(u_{\alpha_1, B_2}) \right] \tag{4.30}
\]
Moreover, from (4.4), we have
\[
|PV_B(u_{\alpha_1, B_1}) - PV_B(u_{\alpha_1, B_2})| \leq O(\|B_1 - B_2\|_{L^\infty}) \left[ PV_B(u_{\alpha_1, B_1}) + PV_B(u_{\alpha_1, B_2}) \right].
\]
Together with (4.30), we deduce that
\[
\|u_{\alpha_1, B_1} - u_{\alpha_2, B_2}\|_{L^2(Q)}^2
\]
\[ \leq 4[2\alpha_1 O(\|\mathcal{B}_1 - \mathcal{B}_2\|_{\infty}) + |\alpha_1 - \alpha_2|] [PV_{\mathcal{B}_1}(u_{\alpha_1, \mathcal{B}_1}) + PV_{\mathcal{B}_2}(u_{\alpha_1, \mathcal{B}_2})] \]

as desired. \hfill \Box

We close this section by proving Theorem 4.9.

**Proof of Theorem 4.9.** The Assertion 1 is the direct result of Theorem 3.2.

We next claim (4.7). We assume that \( u \in C^\infty(\bar{Q}) \) for a moment. Indeed, for any \((\alpha, \mathcal{B}) \in \mathbb{T}\), we could extract a sequence \( \{\mathcal{(\alpha_l, \mathcal{B}_l)}\}_{l=1}^{\infty} \subset \mathbb{T} \), where for each \( l \in \mathbb{N} \), \((\alpha_l, \mathcal{B}_l) \in \mathbb{T}_l \), such that \((\alpha_l, \mathcal{B}_l) \to (\alpha, \mathcal{B}) \). We observe that, by Proposition 4.16,

\[ \max \{A(\alpha_l, \mathcal{B}_l) - A(\alpha, \mathcal{B})\} \leq \max \{|\alpha_{l-1} - \alpha_l|\} \]

Next, take any optimal solution \((\alpha_T, \mathcal{B}_T)\) from (4.6) and by Assertion 1 we could obtain a sequence \( \{(\alpha_{T_1}, \mathcal{B}_{T_1})\}_{l=1}^{\infty} \), where, at each step \( l \in \mathbb{N} \), \((\alpha_{T_1}, \mathcal{B}_{T_1}) \in \mathbb{T}_l \) is determined in (4.6), such that

\[ (\alpha_{T_1}, \mathcal{B}_{T_1}) \to (\alpha_T, \mathcal{B}_T) \]

Also, at each step \( l \in \mathbb{N} \), we find the grid \( \mathcal{G}_l(i, j_l) \) be such that

\[ (\alpha_T, \mathcal{B}_T) \in \mathcal{G}_l(i, j_l) \]

where the grid \( \mathcal{G}_l(i, j_l) \) is defined in (4.5). Then, in view of (4.31), we have that

\[ \max \{A(\alpha, \mathcal{B}) : (\alpha, \mathcal{B}) \in \mathcal{G}_l(i, j_l)\} - \min \{A(\alpha, \mathcal{B}) : (\alpha, \mathcal{B}) \in \mathcal{G}_l(i, j_l)\} \leq 4K|\alpha_{T_1} - \alpha_{T_1}|^{1/2} \|\eta\|_{W^{1,1}(Q)}^{1/2} \]

We divide into two cases.

**Case 1:** Assume at step \( l \) that \((\alpha_{T_1}, \mathcal{B}_{T_1}) \in \mathcal{G}_l(i, j_l) \). In this case we could directly deduce that

\[ A(\alpha_{T_1}, \mathcal{B}_{T_1}) - A(\alpha_{T_1}, \mathcal{B}_{T_1}) \leq \max \{A(\alpha, \mathcal{B}) : (\alpha, \mathcal{B}) \in \mathcal{G}_l(i, j_l)\} - A(\alpha_{T_1}, \mathcal{B}_{T_1}) \]

**Case 2:** Assume at step \( l \) that \((\alpha_{T_1}, \mathcal{B}_{T_1}) \notin \mathcal{G}_l(i, j_l) \). In this case, however, in view of the definition of \((\alpha_{T_1}, \mathcal{B}_{T_1}) \), we must have

\[ \max \{A(\alpha, \mathcal{B}) : (\alpha, \mathcal{B}) \in \mathcal{G}_l(i, j_l) \cap \mathbb{T}_l\} \geq A(\alpha_{T_1}, \mathcal{B}_{T_1}) \]

Since if not, \((\alpha_{T_1}, \mathcal{B}_{T_1}) \) would not be a global minimizer among \( \mathbb{T}_l \), which is a contradiction. Therefore, by (4.34) again we have

\[ A(\alpha_{T_1}, \mathcal{B}_{T_1}) - A(\alpha_{T_1}, \mathcal{B}_{T_1}) \leq \max \{A(\alpha, \mathcal{B}) : (\alpha, \mathcal{B}) \in \mathcal{G}_l(i, j_l) \cap \mathbb{T}_l\} - A(\alpha_{T_1}, \mathcal{B}_{T_1}) \]

where at the last inequality we used the assumption (4.32). In the end, in view of those two cases discussed above and estimation (4.33), we observe that

\[ A(\alpha_{T_1}, \mathcal{B}_{T_1}) - A(\alpha_{T_1}, \mathcal{B}_{T_1}) \leq \max \{A(\alpha, \mathcal{B}) : (\alpha, \mathcal{B}) \in \partial \mathcal{G}_l(i, j_l)\} - \min \{A(\alpha, \mathcal{B}) : (\alpha, \mathcal{B}) \in \mathcal{G}_l(i, j_l)\} \]
Without lose of generality we assume that \( u_\eta \in C^\infty(Q) \). Let \( u_\eta^\varepsilon \in C^\infty(Q) \) be defined as in Case 2 in the argument used to prove Theorem 4.4. Define
\[
    u^\varepsilon_{\alpha, \mathcal{B}} := \arg\min \left\{ \| u - u_\eta^\varepsilon \|_{L^2(Q)}^2 + \alpha PV_{\mathcal{B}}(u) : u \in L^1(Q) \right\}.
\]
Then by Proposition 4.14 we have that
\[
    \| u^\varepsilon_{\alpha, \mathcal{B}} - u_\eta^\varepsilon \|_{L^2(Q)} \leq \| u_\eta^\varepsilon - u_\varepsilon \|_{L^2(Q)},
\]
for arbitrary \((\alpha, \mathcal{B}) \in \mathbb{T}\). Then, for any \( \delta > 0 \) be fixed, we could choose \( \varepsilon > 0 \) small enough such that
\[
    \| u^\varepsilon_{\eta} - u_\varepsilon \|_{L^2(Q)} < \delta/4 \quad \text{and} \quad \| u^\varepsilon_{\eta} \|_{W^{2,1}(Q)} \leq \| u_\eta \|_{L^1(Q)}/\delta^d.
\]
This, and together with (4.35), we conclude that
\[
    \mathcal{A}(\alpha_{T_1}, \mathcal{B}_{T_1}) - \mathcal{A}(\alpha_T, \mathcal{B}_T) \leq 4KP \left[ O(P/l) + 1/l \right]^{1/2} \| u_\eta \|_{W^{2,1}(Q)}^{1/2} + \delta/2
\]
\[
\leq 4KP \left[ O(P/l) + 1/l \right]^{1/2} \| u_\eta \|_{W^{2,1}(Q)}/\delta^d + \delta/2
\]
as desired. \( \square \)

4.3. Examples of Training ground. In this section we give some examples of collection \( \Sigma \) that satisfies Assumption 4.5. Recall the structure of operator \( \mathcal{B} \) from Notation 2.1.

4.3.1. Operator \( \mathcal{B} \) with invertible matrix. Let \( P \in \mathbb{R}^+ \) used in Assumption 4.5 be given. We define the collection \( \Sigma_P \) by
\[
    \Sigma_P := \{ \mathcal{B} : \left| (B^h)^{-1} \right| \leq P, \text{ for each } 1 \leq h \leq d \}, \quad (4.36)
\]
We define the \( h \)-order total variation, say TV\(^h\), of \( u \) by
\[
PV^d(u) := \left| H^h u \right|_{M_b(Q; \mathbb{R}^N^h)},
\]
where \( H^h \) is the \( h \)-order Hessian operator defined in Notation 2.1. We also define the space \( BV^d(Q) \) by
\[
BV^d(Q) := \left\{ u \in L^1(Q) : TV^d(u) < +\infty \right\},
\]
with norm
\[
\| u \|_{BV^d(Q)} := \| u \|_{L^1(Q)} + TV^d(u).
\]

Proposition 4.17. Let \( \mathcal{B} \in \Sigma_P \) be given. Then the space \( BV_{\mathcal{B}}(Q) \) is equivalent to the space \( BV^d(Q) \).

Proof. Without lose of generality we assume that \( u \in BV^d(Q) \cap C^\infty(Q) \), and in view of the structure of operator \( \mathcal{B} \), we have
\[
PV_{\mathcal{B}}(u) = \sum_{h=1}^d |B^h H^h u| dx \leq \sum_{h=1}^d |B^h| \left| H^h u \right| dx \leq \sum_{h=1}^d TV^h(u) \leq C \left( \| u \|_{L^1(Q)} + TV^d(u) \right), \quad (4.37)
\]
Thus, we have that

\[
\text{To conclude, we use an approximation sequence } \{u_\varepsilon\}_{\varepsilon > 0} \subset C^\infty(Q) \text{ from Corollary 2.7 such that } u_\varepsilon \to u \text{ in } L^1 \text{ and } \\
PV_{\mathcal{B}_1}(u_\varepsilon) \to PV_{\mathcal{B}_1}(u) \text{ and } PV_{\mathcal{B}_2}(u_\varepsilon) \to PV_{\mathcal{B}_2}(u). \\
\]

This, and together with (4.37), we are done. □

In the following proposition we show that \( \Sigma_p \) satisfies Assumption 4.5, Assertion 2.

**Proposition 4.18.** Let \( \mathcal{B}_1 \) and \( \mathcal{B}_2 \in \Sigma_p \), and \( u \in BV^d(Q) \) be given. Then we have

\[
|PV_{\mathcal{B}_1}(u) - PV_{\mathcal{B}_2}(u)| \leq \sqrt{K} |\mathcal{B}_1 - \mathcal{B}_2|_{\ell^\infty} \sum_{h \leq d} |(B^h_{\mathcal{B}_1})^{-1}|_{\ell^\infty} PV_{\mathcal{B}_1}(u), \quad (4.38)
\]

and

\[
|PV_{\mathcal{B}_1}(u) - PV_{\mathcal{B}_2}(u)| \leq \sqrt{K} |\mathcal{B}_1 - \mathcal{B}_2|_{\ell^\infty} \sum_{h \leq d} |(B^h_{\mathcal{B}_2})^{-1}|_{\ell^\infty} PV_{\mathcal{B}_2}(u), \quad (4.39)
\]

where \( K \) is defined in (5.4).

**Proof.** Assume for a moment that \( u \in C^\infty(Q) \), we compute that

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

Next, we observe that, for \( x \in Q \),

\[
(\mathcal{B}_1 - \mathcal{B}_2)u(x) = \sum_{h \leq d} (B^h_{\mathcal{B}_1} - B^h_{\mathcal{B}_2}) u(x) = \sum_{h \leq d} (B^h_{\mathcal{B}_1} - B^h_{\mathcal{B}_2})(B^{h}_{\mathcal{B}_1})^{-1} B^{h}_{\mathcal{B}_1} H^h u(x).
\]

Thus, we have

\[
\int_Q |(\mathcal{B}_1 - \mathcal{B}_2)u| \, dx \leq \sum_{h \leq d} |(B^h_{\mathcal{B}_1} - B^h_{\mathcal{B}_2})|_{\ell^\infty} \sum_{h \leq d} |(B^h_{\mathcal{B}_1})^{-1}|_{\ell^\infty} \int_Q |B^h_{\mathcal{B}_1} H^h u| \, dx.
\]

Thus,

\[
|PV_{\mathcal{B}_1}(u) - PV_{\mathcal{B}_2}(u)| \leq \int_Q |\mathcal{B}_1 u - \mathcal{B}_2 u| \, dx \quad (4.40)
\]

\[
\leq \sqrt{K} |\mathcal{B}_1 - \mathcal{B}_2|_{\ell^\infty} \sum_{h \leq d} |(B^h_{\mathcal{B}_1})^{-1}|_{\ell^\infty} PV_{\mathcal{B}_1}(u).
\]

To conclude, we use an approximation sequence \( \{u_\varepsilon\}_{\varepsilon > 0} \subset C^\infty(Q) \) from Corollary 2.7 such that \( u_\varepsilon \to u \) in \( L^1 \) and

\[
PV_{\mathcal{B}_1}(u_\varepsilon) \to PV_{\mathcal{B}_1}(u) \text{ and } PV_{\mathcal{B}_2}(u_\varepsilon) \to PV_{\mathcal{B}_2}(u).
\]

This, and together with (4.40), we obtain (4.38) as desired. Lastly, we remark that we could conclude (4.39) in the same way and hence the thesis. □
Remark. By (4.36), (4.38), and (4.39), we conclude that
\[ |PV_{\mathcal{B}_1}(u) - PV_{\mathcal{B}_2}(u)| \leq \left[ d\sqrt{KP} |\mathcal{B}_1 - \mathcal{B}_2|_{L^\infty} \right] PV_{\mathcal{B}_1}(u), \]
and
\[ |PV_{\mathcal{B}_1}(u) - PV_{\mathcal{B}_2}(u)| \leq \left[ d\sqrt{KP} |\mathcal{B}_1 - \mathcal{B}_2|_{L^\infty} \right] PV_{\mathcal{B}_2}(u). \]
Thus, by setting
\[ O(|\mathcal{B}_1 - \mathcal{B}_2|_{L^\infty}) = d\sqrt{KP} |\mathcal{B}_1 - \mathcal{B}_2|_{L^\infty}, \]
we conclude that \( \Sigma_P \) satisfies Assumption 4.5.

4.3.2. Operators with Energy constraint. We briefly mention that in our previous work [9], a collection of first order operators \( \mathcal{B} \) is introduced, based on some natural quasi-convex constraint. The precise definition is pretty complicated so we decide not to report it again here but refer our readers to [9, Section 6] for future reference.

5. Experimental insights, further extensions, and upcoming works

5.1. Numerical simulations. We remark that the reconstructed image \( u_{\alpha, \mathcal{B}} \) defined in \((T-L2)\), for any given \((\alpha, \mathcal{B}) \in \mathcal{T}\), can be computed by using the primal-dual algorithm presented in [6]. Indeed, we could recast the minimizing problem \((T-L2)\) as the min-max problem
\[ \min \left\{ \max \left\{ \|u - u_\eta\|_{L^2}^2 + \alpha \langle u, \mathcal{B}^* \varphi \rangle : \varphi \in C^\infty_c(Q; \mathbb{R}^K) \right\} : u \in L^1(Q) \right\}, \]
and then the primal-dual method presented in [6] applied.

Next, we present how we put Theorem 4.9 into practical use.

Let \( u_\infty \in L^2(Q) \) and \( u_\eta \in L^2(Q) \) be given. Let an acceptable error \( \varepsilon > 0 \) be given.
- Initialization: Choose an acceptable error \( \varepsilon > 0 \). Choose 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, Theorem 4.9, less or equal to \( \varepsilon/2 \).
- Step 2: Determine one global minimizer \((\alpha_{\mathcal{T}_1}, \mathcal{B}_{\mathcal{T}_1})\) of assessment function \( A(\alpha, \mathcal{B}) \) over the finite training ground \( \mathcal{T}_1 \). Then, by Theorem 4.9 we have that
\[ |A(\alpha_{\mathcal{T}_1}, \mathcal{B}_{\mathcal{T}_1}) - A(\alpha_{\mathcal{T}}, \mathcal{B}_\mathcal{T})| \leq \varepsilon, \]
- Step 3: the reconstructed image \( u_{\alpha_{\mathcal{T}_1}, \mathcal{B}_{\mathcal{T}_1}} \) is then a desired optimal reconstructed result within the acceptable error range.

For the sake of appropriate comparison, we apply our proposed training scheme \( \mathcal{T} \) \(((T-L1)-(T-L2))\) on the image given in Figure 1 with the following training grounds
\[ \mathcal{T}^0 := [0, 1] \times \{ \mathcal{B}_0 \}, \text{ where } \mathcal{B}_0 := [1, 0; 0, 1], \]
(5.1)
\[ \mathcal{T}^1 := [0, 1] \times \{ \mathcal{B}_s : -0.5 \leq s \leq 0.5 \}, \text{ where } \mathcal{B}_s := [1, s; 0, 1], \]
(5.2)
\[ \mathcal{T}^2 := [0, 1] \times \{ \mathcal{B}_{s,t} : -0.5 \leq s, t \leq 0.5 \}, \text{ where } \mathcal{B}_{s,t} := [1, s, t; 1], \]
(5.3)
where we use super-script to avoid confusion with the finite training ground \( \mathcal{T}_1 \). Note that the training ground \( \mathcal{T}^0 \) gives the original training scheme \( \mathcal{B} \) \(((B-L1)-(B-L2))\) with \( TV \) regularizer only. We perform numerical simulations of the images shown in Figure 1: the first image represents a clean image \( u_\infty \), whereas the second one is a noised version \( u_\eta \). We summarize our simulation results...
in Table 1 below. We observe that, from Table 1, as the training ground expand, the minimum

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

Table 1. minimum assessment value for scheme $T$ over training ground defined in (5.1), (5.2), and (5.3)

table value of assessment function $A(\alpha, \mathcal{B})$ decreased. In another word, our new regularizer $PV_{\mathcal{B}}$ indeed provides an improved reconstructed result compare with $TV$ regularizer. However, we remark that the extension of training ground results in an increasing of considerable large amount of CPU time, this would not be a big problem for practical application since we only need to use it one time for a given data set, and the structure of finite training ground $T_l$ allows us to use parallel computing very efficiently and hence reduce the CPU usage.

![Figure 1](image1.png)

Figure 1. From left to right: clean image $u_c$; corrupted image $u_\eta$ (with heavy artificial Gaussian noise); the optimally reconstructed image at $\alpha_{T^0}$, the optimal reconstructed image at $(\alpha_{T^2}, \mathcal{B}_{T^2})$

To explore the numerical landscapes of the assessment function $A(\alpha, \mathcal{B})$ with respect to $\mathcal{B}$, we consider the following training ground with intensity parameter $\alpha$ fixed

$$T = \{0.025\} \times \{\mathcal{B}_{s,t} : -0.5 \leq s, t \leq 0.5\}.$$

and we plot in Figure 2 the mesh and contour images.

We remark that the introduction of $PV_{\mathcal{B}}$ regularizers into the training scheme is only meant to expand the training choices, but not to provide a superior semi-norm with respect to the standard $TV$ semi-norm. The fact whether the optimal regularizer is $TV$ or another intermediate regularizer, is completely dependent on the given training image $u_\eta = u_c + \eta$. Moreover, we remark that the results discovered in this article are not restricted to the imaging processing problems only. It can be generally applied to parameter estimation problems of variational inequalities, as long as a suitable assessment function can be defined.

5.2. Further generalization of $PV_{\mathcal{B}}$ regularizer.
Figure 2. From left to right: mesh and contour plot of the assessment function $A(0.052, s, t)$ in which $(s, t) \in [-0.5, 0.5]^2$. We see from above figures that indeed $A(\cdot)$ is not convex.

5.2.1. Extension with varying underlying Euclidean norm. For $x = (x_1, \ldots, x_K) \in \mathbb{R}^K$, we recall that, for $p \in [1, +\infty)$, the $p$-Euclidean norm of $x$ is defined as

$$|x|_p = \left(\sum_{i=1}^{K} |x_i|^p\right)^{1/p}$$

and $|x|_\infty = \max\{|x_i| : i = 1, \ldots, K\}$.

Note that for $p = 2$, we recover the standard Euclidean norm $|x| = |x|_2$, which is used in (2.4).

In the spirit of [22], we could generalize regularizer $PV_{\mathcal{B}}$ by varying the underlying Euclidean norm. To be precise, we define

$$PV_{p, \mathcal{B}} := \sup \left\{ \int_{Q} u \mathcal{B}^* \varphi \, dx : \varphi \in C_c^\infty(Q; \mathbb{R}^K), \; |\varphi(x)|_p^* \leq 1 \right\},$$

where $|\cdot|_p^*$ represents the dual norm of $|\cdot|_p$, as well as a new training scheme

Level 1. $(\alpha_T, p_T, \mathcal{B}_T) \in \arg\min \left\{ \|u_c - u_{\alpha, \mathcal{B}}\|_{L^2(Q)}^2 : (\alpha, p, \mathcal{B}) \in \mathbb{T} \right\},$

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

with the training ground

$$\mathbb{T} := \mathbb{R} \times [1, +\infty] \times \Sigma.$$  

We remark that, both Theorem 4.4 and Theorem 4.9 holds on this new training scheme with training ground $\mathbb{T}$ and $\mathbb{T} = [0, \kappa] \times [1, +\infty] \times \Sigma$, respectively. The proof is identical to the argument presented before and the argument used in [22] when deal with parameter $p$, so we decide to not report it here to avoid redundancy.
5.2.2. Extension with real order derivative. Let \( s = (s_1, s_2, \ldots, s_d) \in [0, 1)^d \) be given, we define the real \( s \)-order operator \( \mathcal{B}[s] : L^1(Q) \rightarrow D'(Q; \mathbb{R}^K) \) by
\[
\mathcal{B}[s]u := \sum_{h=1}^{d} B^h(H^{h-s_h}u) \quad \text{for every } u \in L^1(Q),
\]
(5.4)
where \( H^{h-s_h} \) represent the \( h-s_h \) order Hessian of \( u \). For example, for \( h = 1 \) and \( s_1 \in [0, 1) \), we have
\[
H^{1-s_1}u = [\partial_1^{1-s_1}u, \partial_2^{1-s_1}u],
\]
where \( \partial_1^{1-s_1}u \) represent the Riemann-Liouville fractional partial derivative with order \( 1-s_1 \notin \mathbb{N} \) (see [18, Definition 2.6]). Then we define
\[
PV_{s,p} := \sup \left\{ \int_{Q} u(\mathcal{B}^*[s])\varphi \, dx : \varphi \in C_0^\infty(Q; \mathbb{R}^K), |\varphi(x)|_p^s \leq 1 \right\},
\]
and the new scheme
Level 1. \((\alpha_T, p_T, s_T, \mathcal{B}_T) \in \arg \min \left\{ \|u_c - u_{\alpha,p,s,\mathcal{B}}\|_{L^2(Q)}^2 : (\alpha, p, s, \mathcal{B}) \in T \right\},\)
Level 2. \( u_{\alpha,p,s,\mathcal{B}} := \arg \min \left\{ \|u - u_\eta\|_{L^2}^2 + \alpha PV_{s,p}(u), u \in L^1(Q) \right\},\)
with training groud
\[
T := \text{cl} \left( \mathbb{R}^+ \right) \times [1, +\infty] \times [0, 1)^d \times \Sigma.
\]
We remark that Theorem 4.4 holds, with additional technics needed when deal with order parameter \( s \) which we report separately in [19]. However, dual to the complexity of fractional order derivative, we can not directly deduce an analogously version of Theorem 4.9.

5.3. Upcoming works. In [9], a PDE-constraint total generalized variation, say \( PGV^2_{\mathcal{B}} \), in multidimensions \( N \in \mathbb{N} \) is defined as follows
\[
PGV^2_{\mathcal{B}}(u) := \inf \left\{ |\nabla u - v_0|_{L^2(Q)} + |\mathcal{B}v_0|_{L^2(M^{N \times N})} : v_0 \in L^1(Q; M^{N \times N}) \right\},
\]
where \( \mathcal{B} : L^1(Q; \mathbb{R}^N) \rightarrow D'(Q; M^{N \times N}) \) is a first order differential operator with some natural PDE constraint (see Section 3 in [9]).

In our follow-up work, we propose to construct an unified approach to regularizers \( PGV^2_{\mathcal{B}} \) and \( PV_{s,\mathcal{B}} \) via
\[
PGV_{s,\mathcal{B},t}(u) := \inf \left\{ |\mathcal{B}[s]u - tv_0|_{L^2(M^{N \times N})} + t |\mathcal{B}[t]v_0|_{L^2(M^{N \times N})} : v_0 \in L^1(Q; M^{N \times N}) \right\}.
\]
We shall equip the training scheme \( T \) with this new regularizer and provide an analogously version of Theorem 4.4 and Theorem 4.9 in our follow-up work.

Acknowledgments. PL acknowledges support from the EPSRC Centre Nr. EP/N014588/1 and the Leverhulme Trust project on Breaking the non-convexity barrier. The author also gratefully acknowledge the support of NVIDIA Corporation with the donation of a Quadro P6000 GPU used for this research.
REFERENCES

[1] T. Barbu and G. Marineschi. Image denoising by a nonlinear control technique. *International Journal of Control*, 90:1005–1017, 2017.
[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.
[3] K. Bredies, K. Kunisch, and T. Pock. Total generalized variation. *SIAM J. Imaging Sci.*, 3(3):492–526, 2010.
[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.
[5] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Universitext. Springer, New York, 2011.
[6] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.
[7] Y. Chen, T. Pock, R. Ranftl, and H. Bischof. Revisiting loss-specific training of filter-based mrf s for image restoration. In *Pattern Recognition*, pages 271–281. Springer, 2013.
[8] Y. Chen, R. Ranftl, and T. Pock. Insights into analysis operator learning: From patch-based sparse models to higher order mrf s. *IEEE Transactions on Image Processing*, 23(3):1060–1072, March 2014.
[9] E. Davoli, I. Fonseca, and P. Liu. Adaptive image processing: first order PDE constraint regularizers and a bilevel training scheme. *arXiv*.
[10] E. Davoli and P. Liu. One dimensional fractional order tv: gamma-convergence and bilevel training scheme. *Commun. Math. Sci.*, 16(1):213–237, 2018.
[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.
[12] J. Domke. Generic methods for optimization-based modeling. In *AISTATS*, volume 22, pages 318–326, 2012.
[13] I. Ekeland and R. Temam. *Convex analysis and variational problems*. SIAM, 1976.
[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.
[15] I. Fonseca and S. Müller. Relaxation of quasiconvex functionals in BV(Ω, Rp) for integrands f(x, u, ∇u). *Arch. Rational Mech. Anal.*, 123(1):1–49, 1993.
[16] K. C. Kiwiel. Convergence and efficiency of subgradient methods for quasiconvex minimization. *Math. Program.*, 90(1, Ser. A):1–25, 2001.
[17] K. Kunisch and T. Pock. A bilevel optimization approach for parameter learning in variational models. *SIAM J. Imaging Sci.*, 6(2):938–983, 2013.
[18] P. Liu and X. Y. Lu. Real order (an)-isotropic total variation in image processing - Part I: analytical analysis and functional properties. *ArXiv e-prints*, May 2018.
[19] P. Liu and X. Y. Lu. Real order (an)-isotropic total variation in image processing - Part II: learning of optimal structures. *ArXiv e-prints*, 2019.
[20] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60(1-4):259–268, 1992. Experimental mathematics: computational issues in nonlinear science (Los Alamos, NM, 1991).
[21] S. G. Samko, A. A. Kilbas, and O. I. Marichev. *Fractional integrals and derivatives*. Gordon and Breach Science Publishers, Yverdon, 1993. Theory and applications, Edited and with a foreword by S. M. Nikol’skiĭ, Translated from the 1987 Russian original, Revised by the authors.
[22] C.-B. Schönlieb and P. Liu. Learning optimal orders of the underlying euclidean norm in total variation image processing. *arXiv*, 2018.
[23] M. F. Tappen, C. Liu, E. H. Adelson, and W. T. Freeman. Learning gaussian conditional random fields for low-level vision. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2007.