LEARNING OPTIMAL ORDERS OF THE UNDERLYING EUCLIDEAN NORM IN TOTAL VARIATION IMAGE DENOISING

PAN LIU
Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK

CAROLA-BIBIANE SCHÖNLEB
Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK

March 28, 2019

Abstract. A novel class of semi-norms, generalising the notion of the isotropic total variation \(TV_2\) and the an-isotropic total variation \(TV_1\) is introduced. A supervised learning method via bilevel optimisation is proposed for the computation of optimal parameters for this class of regularizers. Existence of solutions to the bilevel optimisation approach is proven. Moreover, a finite-dimensional approximation scheme for the bilevel optimisation approach is introduced that can numerically compute a global optimizer to any given accuracy.

Contents

1. Introduction 2
2. The \(\ell^p\)-anisotropic total variation, \(\Gamma\)-convergence, and an optimal training scheme 4
   2.1. The \(\ell^p\)-(an)-isotropic total variation 5
   2.2. \(\Gamma\)-convergence of functionals defined by \(TV_p\) seminorms 5
   2.3. Bilevel training scheme \(T\) and existence of solutions 7
   2.4. Extension of \(\ell^p\)-anisotropic total variation via Finsler metrics 11
3. Learning of acceptable optimal solutions 12
   3.1. Non-convexity of the assessment function and counterexamples 12
   3.2. A finite approximation of scheme \(T\) 14
4. Numerical simulations and conclusions 23
   4.1. Simulations and insights 23
   4.2. Conclusions and future works 25
References 27

E-mail addresses: panliu.0923@maths.cam.ac.uk, cbs31@cam.ac.uk.
2010 Mathematics Subject Classification. 26B30, 94A08, 47J20.
Key words and phrases. total variation, optimization and control, computer vision and pattern recognition.
1. Introduction

Total variation denoising is given by the minimization problem

$$u_\alpha := \arg \min \left\{ \| u - u_n \|_{L^2(Q)}^2 + \alpha \text{TV}(u) : u \in BV(Q) \right\}, \quad (1.1)$$

where $u_n \in L^2(Q)$ denotes a given noisy image on the domain $Q := (0, 1)^N$, $N \in \mathbb{N}$, $\alpha \in \mathbb{R}^+$ denotes the regularization parameter, and $\text{TV}(u)$ is the (isotropic) total variation defined by

$$\text{TV}(u) := \sup \left\{ \int_Q u \, \text{div} \varphi \, dx : \varphi \in C^\infty_c(Q; \mathbb{R}^N), \ |\varphi| \leq 1 \right\}$$

which is $< +\infty$ for $u \in BV(Q)$, the space of functions of bounded variation. Here $|\cdot|$ refers to the $\ell^2$-Euclidean norm, that is, for $x = (x_1, x_2) \in \mathbb{R}^2$ we have $|x| = (x_1^2 + x_2^2)^{1/2}$. The TV denoising model (1.1) is also called ROF model, named after the pioneering paper [27] of Rudin, Osher, and Fatemi. The TV denoising model is known for its ability to both denoise an image and at the same time preserve discontinuities. Due to this edge-preserving property the ROF model has established itself in the image processing literature.

Next to these desirable denoising properties the ROF model, however, also comes with disadvantages. One of those is the tendency of the ROF model to generate unnecessary edges. These turn originally smoothly changing image intensities into piecewise constant intensity areas which create blocky-like artefacts also known as stair-casing. Another disadvantage of the ROF model is that it leads to a contrast loss near edges that mainly depends on their curvature.

TV$p$, $1 \leq p \leq \infty$. In this paper we consider a generalised notion of the total variation in which we replace the underlying $\ell^2$-Euclidean norm by the $\ell^p$-Euclidean norm for $1 \leq p < \infty$. We therefore write $|\cdot|_p$ for $|\cdot|$ and $\text{TV}_p(\cdot)$ for $\text{TV}(\cdot)$. In [17], for instance, another variant of the total variation has been proposed, by switching the underlying Euclidean norm from $\ell^2$ to $\ell^1$ ([30]), i.e.,

$$\text{TV}_1(u) := \sup \left\{ \int_Q u \, \text{div} \varphi \, dx : \varphi \in C^\infty_c(Q; \mathbb{R}^N), \ |\varphi|_1 \leq 1 \right\} < +\infty,$$

where $|\cdot|_1$ denotes the Euclidean 1-norm in the sense that for $x = (x_1, x_2) \in \mathbb{R}^2$, $|x|_1 = |x_1| + |x_2|$, and $|\cdot|_1^*$ denotes the dual norm associated with $|\cdot|_1$. Successful applications of $TV_1$, also called the anisotropic total variation, can be found in [17, 25, 22, 15, 28]. In particular, in [17], it has been observed that $TV_1$ has the ability to suppress the stair-casing effect which is a typical artifact induced by isotropic total variation. Total variations defined by Euclidean $\ell^p$-norms, for $p \notin \{1, 2\}$, have rarely been analyzed, and hence their performance is largely unknown.

Bilevel optimization. The quality of a reconstructed image $u_\alpha$ obtained from (1.1) highly depends on the choice of the regularization parameter $\alpha$. If $\alpha$ is too large then $TV(u)$ is penalized too much and the image is over-smoothed, resulting in a loss of information in the reconstructed image. On the other hand, if $\alpha$ is too small then the reconstructed image is under-regularized and noise is left in the reconstruction. Classical approaches to choose an appropriate regularisation parameter are Morozov’s discrepancy principle [26], generalised cross-validation [16] or L-curve [19] just to name a few [13]. A recent approach to determine the optimal $\alpha$ is bilevel optimization (see e.g.,[18, 29, 11, 9, 21, 6, 7]). Here, an optimality criterion for the denoised image is used, given in terms of a loss function for the minimiser
that given a training set of noisy images \((u^i_n)_{i=1}^I\) and corresponding noise-free images \((u^i_c)_{i=1}^I\), bilevel optimization for the regularization parameter \(\alpha\) in (1.1) reads

\[
\begin{align*}
\text{Level 1. } & \quad \tilde{\alpha} \in \arg \min \left\{ \sum_{i=1}^I \|u^i_n - u^i_c\|_{L^2(Q)}^2 : \alpha \in \mathbb{R}^+ \right\}, \quad \text{s.t.} \quad (B-L1) \\
\text{Level 2. } & \quad u^i_\alpha := \arg \min \left\{ \|u - u^i_c\|_{L^2(Q)}^2 + \alpha TV_2(u) : u \in BV(Q) \right\}, \quad i = 1, \ldots, (B-L2)
\end{align*}
\]

The Level 1 problem in \((B-L1)\) looks for an \(\alpha\) that minimizes the average \(L^2\)-distance between minimizers \(u^\alpha\) of the Level 2 problem \((B-L2)\) and clean images \(u^c\). It has been proven in [10] that \((B-L1)\) admits at least one positive solution \(\tilde{\alpha} \in \mathbb{R}^+\) provided that \(TV_2(u^\alpha_n) > TV_2(u^\alpha_c)\), which is a reasonable assumption for image denoising.

For simplicity, in what follows we will omit the index \(i\) from the training set and perform our analysis for a single pair \((u^\alpha, u^c)\) of noisy and noise-free image, respectively. Everything we will discuss, however, will still hold for the case of multiple images in the training set.

**Bilevel optimisation for \(TV_p\).** For the purpose of studying \(TV_p\) for \(p \in [1, +\infty]\) we extend the bilevel training scheme \(B\) to scheme \(T\) as

\[
\begin{align*}
\text{Level 1. } & \quad (\alpha_T, p_T) \in \mathcal{A}[T] := \arg \min \left\{ \|u_{\alpha,p} - u^c\|_{L^2(Q)}^2 : (\alpha, p) \in \mathbb{T} \right\}, \quad (T-L1) \\
\text{Level 2. } & \quad u_{\alpha,p} := \arg \min \left\{ \|u - u^c\|_{L^2(Q)}^2 + \alpha TV_p(u) : u \in BV(Q) \right\}, \quad (T-L2)
\end{align*}
\]

where, and in what follows, we call the set

\[
\mathbb{T} := \mathbb{R}^+ \times [1, +\infty]
\]

the *training ground*, in which we optimize parameters \(\alpha\) and \(p\), and the set \(\mathcal{A}[\mathbb{T}]\) the optimal set associated with \(\mathbb{T}\), which contains the optimal parameters produced by the scheme \(T\).

We point out that the new training scheme \(T\) simultaneously optimizes both the parameter \(\alpha\) and the order \(p\).

**Our contribution.** This paper contains two main results. The first result is contained in Theorem 2.7 and proves that the scheme \(T\) \((T-L1)-(T-L2)\) admits at least one optimal solution \((\alpha_T, p_T) \in \mathbb{T}\). This existence result is based on Theorem 2.3 where we show that the functionals

\[
\mathcal{I}_{\alpha, p}(u) := \|u - u^c\|_{L^2(Q)}^2 + \alpha TV_p(u)
\]

are continuous, in the sense of \(\Gamma\)-convergence in the weak* topology of \(BV(Q)\) (see [3, 8]), with respect to the parameters \((\alpha, p)\). We prove this by showing that the collection of new spaces, induced from \(TV_p\) semi-norms, itself exhibits certain compactness and lower semi-continuity properties.

Our second contribution is a proposal for how to numerically determine the optimal solution of scheme \(T\), or equivalently compute global minimizers of the *assessment function* \(\mathcal{A}(\alpha, p) : \mathbb{T} \to \mathbb{R}^+\) defined as

\[
\mathcal{A}(\alpha, p) := \| u_{\alpha,p} - u^c \|_{L^2(Q)}^2
\]

where \(u_{\alpha,p}\) is obtained from \((T-L2)\). We note that computing such global minimizers would be straightforward if \(\mathcal{A}(\alpha, p)\) is quasi-convex in the sense of [20], or simply convex. In
this case Newton’s descent method or line search could be applied to compute a the global minimizer. However, as we shall later show in Figure 1 even for a fixed \( p_0 \in [1, +\infty) \) the assessment function \( A(\alpha, p_0) \) is not quasi-convex, and hence those methods mentioned above might get trapped in a local minimum. To overcome this difficulty, we introduce the concept of an acceptable optimal solution that is a solution to \( T \) with a prescribed error. To be precise, we say the solution \((\hat{\alpha}, \hat{\beta})\) is acceptable with error \( \varepsilon > 0 \) if

\[
|A(\hat{\alpha}, \hat{\beta}) - A(\alpha^T, p_T)| < \varepsilon,
\]

where \((\alpha^T, p_T) \in A(T)\) is a globally optimal solution obtained from the scheme \( T \).

For computing such an acceptable optimal solution, we propose in Section 3 a finite approximation method. We construct a sequence of finite sets \( T_l \) indexed by \( l \in \mathbb{N} \), such that \# \( \{T_l\} < +\infty \). For the precise definition of \( T_l \) we refer to Definition 3.1. We point out here that, since \# \( \{T_l\} < +\infty \), the optimal solution(s)

\[
(\alpha^T_l, p^T_l) \in A(T_l) := \arg \min \{A(\alpha, p) : (\alpha, p) \in T_l\}
\]

can be determined precisely by evaluating \( A(\alpha, p) \) at each \((\alpha, p) \in T_l\). From there, it is not hard to prove that \((\alpha^T_l, p^T_l) \to (\alpha^T, p_T)\) and \( A(\alpha^T_l, p^T_l) \to A(\alpha^T, p_T)\), as \( l \to \infty \), by using standard \( \Gamma \)-convergence techniques. This is, however, still not enough to allow the computation of an acceptable solution as in (1.4). To achieve such a result, we prove in Theorem 3.3 an estimate for a fixed index \( l \in \mathbb{N} \), which gives an estimate of the form

\[
|A(\alpha^T_l, p^T_l) - A(\alpha^T, p_T)| \leq \sqrt{\alpha_U} \left[ (1/l)^{1/2} + 2 \left( 1 - N^{-1/\sqrt{l}} \right)^{1/2} \right] (TV_1(u_\eta))^1/2,
\]

in which \( \alpha_U \) can be determined numerically (see Proposition 2.10). Therefore, by using estimate (1.5), we can acquire the desired index \( l \in \mathbb{N} \) so that the associated optimal solution \((\alpha^T_l, p^T_l) \in T_l\) is an acceptable optimal solution for the error \( \varepsilon > 0 \).

We note that the estimate (1.5) requires that \( u_\eta \in BV(Q) \), which usually does not hold for a noisy image \( u_\eta \). To overcome this, in Section 3.2.1 we show that, for any given \( \varepsilon > 0 \), even if \( u_\eta \in L^2(Q) \setminus BV(Q) \), we are still able to find \( l \in \mathbb{N} \) such that

\[
|A(\alpha^T_l, p^T_l) - A(\alpha^T, p_T)| \leq \varepsilon,
\]

i.e., the associated optimal solution \((\alpha^T_l, p^T_l) \in T_l\) is an acceptable optimal solution for the error \( \varepsilon > 0 \). We do so by introducing a piece-wise constant approximation of the corrupted image \( u_\eta \in L^2(Q) \setminus BV(Q) \), and we refer readers to Theorem 3.9 and Corollary 3.12 for details.

**Organisation of the paper.** The paper is organized as follows. In Section 2.1 we collect some notations and preliminary results. The \( \Gamma \)-convergence and the bilevel training scheme are the subjects of Sections 2.2 and 2.3, respectively. Section 3 is devoted to the analysis of the finite approximation training scheme and the proof of Theorem 3.3. Finally, in Section 4 some numerical simulations and insights.

2. THE \( \ell^p \)-ANISOtROPIC TOTAL VARIATION, \( \Gamma \)-CONVERGENCE, AND AN OPTIMAL TRAINING SCHEME

We recall that, throughout this article, \( u_\eta \in L^2(Q) \) denotes a given datum representing a noisy image, \( u_c \in BV(Q) \) represents the corresponding noise-free image, and \( u_{\alpha, p} \in BV(Q) \) is the reconstructed image obtained from (7-L2) for a given set \((\alpha, p) \in T\).
2.1. \textbf{The $\ell^p$-(an)-isotropic total variation.} We recall from \cite{14} that a function $u \in L^1(Q)$ has \textit{bounded variation} in $Q$ if

$$TV_2(u) := \sup \left\{ \int_Q u \, \text{div} \, \varphi \, dx : \varphi \in C_c^\infty(Q; \mathbb{R}^N), \ |\varphi|_2 \leq 1 \right\} < +\infty,$$

and write $BV(Q)$ to denote the space of functions of bounded variation. We also define the norm

$$\|u\|_{BV(Q)} := \|u\|_{L^1(Q)} + TV_2(u).$$

We next define the Euclidean $\ell^p$-norm for $p \in [1, +\infty]$ and for $x = (x_1, x_2, \ldots, x_N) \in \mathbb{R}^N$ as

$$|x|_p := (|x_1|^p + |x_2|^p + \cdots + |x_N|^p)^{1/p}. \quad (2.1)$$

We recall that $|\cdot|_p$ for $p \in [1, +\infty]$ are equivalent norms on $\mathbb{R}^N$. To be precise, for any $1 \leq p_1 < p_2 \leq +\infty$ and $x \in \mathbb{R}^N$, we have that

$$|x|_{p_2} \leq |x|_{p_1} \leq N^{1/p_2 - 1/p_1} |x|_{p_2}. \quad (2.2)$$

\textbf{Definition 2.1} (The $\ell^p$-an-isotropic total variation). Let $u \in L^1(Q)$ be given, we define, for $1 \leq p \leq +\infty$, the $\ell^p$ an-isotropic total variation $TV_p$ by

$$TV_p(u) := \sup \left\{ \int_Q u \, \text{div} \, \varphi \, dx : \varphi \in C_c^\infty(Q; \mathbb{R}^N), \ |\varphi|_p^* \leq 1 \right\},$$

where $|\cdot|_p^*$ denotes the dual norm associated with $|\cdot|_p$.

\textbf{Remark 2.2.} In view of (2.2), we have that the $TV_p$ semi-norms, for $1 \leq p \leq \infty$, are equivalent. That is, for $1 \leq p_1 < p_2 \leq +\infty$, we have that

$$TV_{p_2}(u) \leq TV_{p_1}(u) \leq N^{1/p_2 - 1/p_1} TV_{p_2}(u). \quad (2.3)$$

for all $u \in BV(Q)$. In particular, we have

$$TV(u) = TV_2(u) \leq N \cdot TV_p(u) \quad (2.4)$$

for any $p \in [1, +\infty]$ and $u \in BV(Q)$.

2.2. \textbf{$\Gamma$-convergence of functionals defined by $TV_p$ seminorms.} Let $p \in [1, +\infty]$ and $\alpha \in \mathbb{R}^+$. We define the functional $\mathcal{I}_{\alpha,p}: L^2(Q) \to [0, +\infty]$ as

$$\mathcal{I}_{\alpha,p}(u) := \begin{cases} \|u - u_0\|_{L^2(Q)}^2 + \alpha TV_p(u), & \text{if } u \in BV(Q) \\ +\infty, & \text{otherwise} \end{cases}. $$

The following theorem is the main result of Section 2.2.

\textbf{Theorem 2.3.} Let $\{p_n\}_{n=1}^\infty \subset [1, +\infty]$ and $\{\alpha_n\}_{n=1}^\infty \subset \mathbb{R}^+$ be given such that $p_n \to p_0$ and $\alpha_n \to \alpha_0 \in \mathbb{R}^+$. Then the functional $\mathcal{I}_{\alpha_n,p_n}$ \textit{$\Gamma$-converges to} $\mathcal{I}_{\alpha_0,p_0}$ in the weak* topology of $BV(Q)$. Namely, for every $u \in BV(Q)$ the following two assertions hold:

\begin{itemize}
  \item[(LI)] \textit{If}
  $$u_n \rightharpoonup u_0 \text{ weakly* in } BV(Q),$$
  \textit{then}
  $$\liminf_{n \to +\infty} \mathcal{I}_{\alpha_n,p_n}(u_n) \geq \mathcal{I}_{\alpha_0,p_0}(u_0).$$
  \item[(RS)] \textit{For each} $u_0 \in BV(Q)$, \textit{there exists} $\{u_n\}_{n=1}^\infty \subset BV(Q)$ \textit{such that}
  $$u_n \rightharpoonup u_0 \text{ weakly* in } BV(Q),$$
  \textit{and}
  $$\limsup_{n \to +\infty} \mathcal{I}_{\alpha_n,p_n}(u_n) \leq \mathcal{I}_{\alpha_0,p_0}(u_0).$$
\end{itemize}
We subdivide the proof of Theorem 2.3 into two propositions.

**Proposition 2.4** (Γ-limit inf inequality). Let \( \{p_n\}_{n=1}^{\infty} \subset [1, +\infty] \) and \( \{\alpha_n\}_{n=1}^{\infty} \subset \mathbb{R}^+ \) be such that \( p_n \to p \) and \( \alpha_n \to \alpha \in \mathbb{R}^+ \). Let \( \{u_n\}_{n=1}^{\infty} \subset BV(Q) \) be such that
\[
\sup \{ I_{\alpha_n,p_n} (u_n) : \ n \in \mathbb{N} \} < +\infty. \tag{2.5}
\]
Then, there exists \( u \in BV(Q) \) such that, up to the extraction of a (non-relabeled) subsequence, there holds
\[
u_n \rightharpoonup u \text{ in } BV(Q),
\]
with
\[
\liminf_{n \to +\infty} TV_{p_n}(u_n) \geq TV_p(u). \tag{2.6}
\]

**Proof.** We prove the statement for \( \alpha_n \equiv 1 \) only, as the general case for \( \alpha \in \mathbb{R}^+ \) can be argued with straightforward adaptations.

By (2.4) we always have
\[
TV(u_n) \leq N \cdot TV_{p_n}(u_n) \leq N \cdot I_{1,p_n}(u_n).
\]
Thus, by (2.5) we have
\[
\sup \left\{ \|u_n\|_{BV(Q)} : \ n \in \mathbb{N} \right\} < +\infty,
\]
which implies that there exists \( u \in BV(Q) \) such that, up to extract a subsequence (not relabeled),
\[
u_n \rightharpoonup u \text{ in } BV(Q) \text{ and } u_n \to u \text{ in } L^1 \text{ and } a.e..
\]

Therefore, we conclude that
\[
\liminf_{n \to +\infty} TV_{p_n}(u_n) \geq \liminf_{n \to +\infty} N^{-1/p_n} TV_{p_n}(u_n) \geq TV_p(u),
\]
where in the first inequality we used (2.3). This concludes the proof of (2.6) and hence the proposition. \( \square \)

**Lemma 2.5.** Let \( p \in [1, +\infty] \) be fixed. Then the function
\[
f(\alpha) := TV_p(u_{\alpha,p}), \ \alpha \in \mathbb{R}^+
\]
is continuous and monotonically decreasing and the function
\[
g(\alpha) := \|u_{\alpha,p} - u_\eta\|_{L^2(Q)}, \ \alpha \in \mathbb{R}^+
\]
is continuous and non-decreasing.

**Proof.** We notice that by Proposition 2.4, we have \( f(\alpha) \) and \( g(\alpha) \) are continuous. Next, for any \( 0 \leq \alpha_1 < \alpha_2 < +\infty \), by minimality there holds
\[
\|u_{\alpha_1,p} - u_\eta\|_{L^2(Q)}^2 + \alpha_1 TV_p(u_{\alpha_1,p}) \leq \|u_{\alpha_2,p} - u_\eta\|_{L^2(Q)}^2 + \alpha_1 TV_p(u_{\alpha_2,p}) \tag{2.7}
\]
and
\[
\|u_{\alpha_2,p} - u_\eta\|_{L^2(Q)}^2 + \alpha_2 TV_p(u_{\alpha_2,p}) \leq \|u_{\alpha_1,p} - u_\eta\|_{L^2(Q)}^2 + \alpha_2 TV_p(u_{\alpha_1,p}).
\]

Adding up the previous two inequalities yields
\[
(\alpha_2 - \alpha_1) TV_p(u_{\alpha_2,p}) \leq (\alpha_2 - \alpha_1) TV_p(u_{\alpha_1,p}),
\]
which implies that
\[
f(\alpha_2) = TV_p(u_{\alpha_2,p}) \leq TV_p(u_{\alpha_1,p}) = f(\alpha_1).
\]
Moreover, in view of (2.7) and (2.8), we obtain that
\[ \|u_{\alpha_2,p} - u_0\|^2_{L^2(Q)} + \alpha_1 TV_p(u_{\alpha_2,p}) \]
\[ \leq \|u_{\alpha_2,p} - u_0\|^2_{L^2(Q)} + \alpha_1 TV_p(u_{\alpha_2,p}) \]
\[ \leq \|u_{\alpha_2,p} - u_0\|^2_{L^2(Q)} + \alpha_1 TV_p(u_{\alpha_1,p}), \]
which, in turn, yields \( \|u_{\alpha,\rho} - u_0\|^2_{L^2(Q)} \) is non-decreasing and we are done. \( \Box \)

**Proposition 2.6.** Let \( \{p_n\}_{n=1}^{\infty} \subset [1, +\infty] \) and \( \{\alpha_n\}_{n=1}^{\infty} \subset \mathbb{R}^+ \) be such that \( p_n \to p \) and \( \alpha_n \to \alpha \in \mathbb{R}^+ \). Then for every \( u \in BV(Q) \) there holds
\[ \limsup_{n \to \infty} \alpha_n TV_{p_n}(u) = \alpha TV_p(u). \]

**Proof.** For simplicity, we only analyze this proposition under assumption \( \alpha_n = 1 \) for all \( n \in \mathbb{N} \). All arguments also hold for a general sequence \( \{\alpha_n\}_{n=1}^{\infty} \) since \( \alpha \in \mathbb{R}^+ \).

The liminf inequality
\[ \liminf_{n \to \infty} TV_{p_n}(u) \geq TV_p(u) \]
is a direct consequence of Proposition 2.4 by choosing \( u_n := u \). Next, by (2.3) we have that
\[ TV_{p_n}(u) \leq N^{1/p_n - 1/p} TV_p(u). \]
and the limsup inequality
\[ \limsup_{n \to \infty} TV_{p_n}(u) \leq TV_p(u). \]
is asserted by sending \( p_n \to p \). \( \Box \)

**Proof of Theorem 2.3.** Let \( (\alpha_n, p_n) \to (\alpha_0, p_0) \in \mathbb{R}^+ \times [1, +\infty] \) be given. We obtain Property (LI) in view of Proposition 2.4. Property (RS) follows by Proposition 2.6, choosing \( u_n = u_0 \) for every \( n \in \mathbb{N} \). \( \Box \)

### 2.3. Bilevel training scheme \( \mathcal{T} \) and existence of solutions.

We recall the training ground \( \mathcal{T} \) from (1.2) and two levels of the scheme \( \mathcal{T} \) are

**Level 1.** \( (\alpha_\mathcal{T}, p_\mathcal{L}) \in \mathcal{A}[\mathcal{T}] := \arg \min \left\{ \|u_{\alpha,p} - u_\mathcal{L}\|^2_{L^2(Q)} : (\alpha, p) \in \mathcal{T} \right\}, \) \( (\mathcal{T}-\text{LI}) \)

**Level 2.** \( u_{n,p} := \arg \min \left\{ \|u - u_\mathcal{L}\|^2_{L^2(Q)} + \alpha TV_p(u) : u \in BV(Q) \right\}. \) \( (\mathcal{T}-\text{L2}) \)

The following theorem is the main result of Section 2.3.

**Theorem 2.7.** Let \( u_\mathcal{L} \in BV(Q) \) and \( u_\eta \in L^2(Q) \) be given such that
\[ TV_\infty(u_\eta) > TV_1(u_\mathcal{L}). \] (2.9)
Then, the training scheme \( \mathcal{T} \) \((\mathcal{T}-\text{L1})-(\mathcal{T}-\text{L2})\) admits at least one solution \((\alpha_\mathcal{T}, p_\mathcal{L}) \in (0, \alpha_{\mathcal{L}}] \times [1, +\infty] \), where the upper bound \( \alpha_{\mathcal{L}} \in \mathbb{R}^+ \) is determined in Proposition 2.10.

**Proposition 2.8.** Let \( u_\mathcal{L}, u_\eta \in L^2(Q) \) be given such that (2.9) holds. Then, there exists an \( \alpha_L \in \mathbb{R}^+ \) such that
\[ \sup \left\{ \|u_{\alpha_L,p} - u_\mathcal{L}\|^2_{L^2(Q)} : p \in [1, +\infty] \right\} < \|u_\eta - u_\mathcal{L}\|^2_{L^2(Q)}. \] (2.10)
Proof. Fix $\alpha > 0$ and let $\partial TV_p(u)$ denotes the sub-differential of $TV_p$ at $u$, we observe that
\[
\|u_\eta - u_c\|_{L^2(Q)}^2 - \|u_{\alpha,p} - u_c\|_{L^2(Q)}^2 = 2 \langle u_\eta - u_{\alpha,p}, u_\eta - u_c \rangle_{L^2} + \|u_\eta - u_{\alpha,p}\|_{L^2(Q)}^2 \\
= 2\alpha \langle \partial TV_p(u_{\alpha,p}), u_\eta - u_{\alpha,p} \rangle_{L^2} + \|u_\eta - u_{\alpha,p}\|_{L^2(Q)}^2 \\
\geq 2\alpha (TV_p(u_{\alpha,p}) - TV_p(u_c)) + \|u_\eta - u_{\alpha,p}\|_{L^2(Q)}^2,
\]
where at the last inequality we used the property of sub-gradient operator (see [12, Proposition 5.4, page 24]).

Recall from Lemma 2.5 that $TV_p(u_{\alpha,p})$ is continuously decreasing with respect to $\alpha$, and thus we can find $\alpha_p > 0$, might depend on $p$, such that
\[
TV_p(u_c) + [TV_{\infty}(u_\eta) - TV_1(u_c)]/4 > TV_p(u_{\alpha,p}) > TV_p(u_c) + [TV_{\infty}(u_\eta) - TV_1(u_c)]/8,
\]
provided that (2.9) holds. Therefore, by (2.11) we have
\[
\|u_\eta - u_c\|_{L^2(Q)}^2 - \|u_{\alpha,p} - u_c\|_{L^2(Q)}^2 \\
\geq 2\alpha (TV_p(u_{\alpha,p}) - TV_p(u_c)) + \|u_\eta - u_{\alpha,p}\|_{L^2(Q)}^2 \\
\geq 2\alpha_p [TV_{\infty}(u_\eta) - TV_1(u_c)]/8 + \|u_\eta - u_{\alpha,p}\|_{L^2(Q)}^2 > 0.
\]

We next claim that
\[
\alpha_L := \inf \{ \alpha_p : p \in [1, +\infty) \} > 0.
\]
Assume that not, that is, there exists sequence $p_n \to p \in [1, +\infty)$ such that
\[
\lim_{n \to \infty} \alpha_{p_n} \searrow 0.
\]

We claim that $u_{\alpha_{p_n}, p_n} \to u_\eta$ strongly in $L^2$. Let $\{u_{\eta,k}\}_{k=1}^\infty \subset C^\infty(Q)$ be such that $u_{\eta,k} \to u_\eta$ strongly in $L^2$, and by the optimality condition of $u_{\alpha_{p_n}, p_n}$, we deduce that
\[
\|u_\eta - u_{\alpha_{p_n}, p_n}\|_{L^2(Q)}^2 \leq \alpha_{p_n} TV_{p_n}(u_{\alpha_{p_n}, p_n}) \\
\leq \alpha_{p_n} TV_{p_n}(u_{\eta,k}) \\
\leq \alpha_{p_n} TV_1(u_{\eta,k}) \quad \text{by the property of sub-gradient operator (see [12, Proposition 5.4, page 24])}.
\]

Thus, by (2.15) and letting $n \to \infty$ first and $k \to \infty$ second, we conclude that
\[
\lim_{k,n \to \infty} \sup \|u_\eta - u_{\alpha_{p_n}, p_n}\|_{L^2(Q)}^2 + \alpha_{p_n} TV_{p_n}(u_{\alpha_{p_n}, p_n}) \\
\leq \lim_{k \to \infty} \limsup \|u_\eta - u_{\eta,k}\|_{L^2(Q)}^2 = 0.
\]

That is, we have $u_{\alpha_{p_n}, p_n} \to u_\eta$ strongly in $L^2(Q)$ and, upon extracting a further subsequence (not relabeled), there holds $p_n \to p$ and
\[
\liminf_{u \to \infty} TV_{p_n}(u_{\alpha_{p_n}, p_n}) \geq TV_p(u_\eta) > TV_p(u_c) + [TV_{\infty}(u_\eta) - TV_1(u_c)]/4,
\]
which contradicts (2.12). This completes the proof of (2.14).

Now we prove (2.10). In view of (2.13) and (2.14) we have
\[
\|u_\eta - u_{\alpha,p}\|_{L^2(Q)}^2 \\
\geq 2\alpha_p [TV_{\infty}(u_\eta) - TV_1(u_c)]/8 + \|u_\eta - u_{\alpha,p}\|_{L^2(Q)}^2 > 0.
\]
of the total variation semi-norm is the space of constant functions (see, e.g., [1]), which is

and thus we conclude (2.10) since the right hand side of above inequality does not depend on $p$.

Next, we determine a uniform upper bound on the optimal regularization parameter $\alpha_T$.

We start with the following lemma, where $(u_\eta)_Q$ denotes the average of $u_\eta$ over $Q$, i.e.

$$
(u_\eta)_Q := \frac{\int_Q u_\eta \, dx}{\int_Q 1 \, dx}.
$$

Lemma 2.9. Let $p \in [1, +\infty]$ be fixed and $u_{\alpha,p}$ the minimiser of (T-L2). Then there exists $\alpha_U = \alpha_U(p) < +\infty$ such that

$$
TV_p(u_{\alpha,p}) > 0 \text{ for all } \alpha < \alpha_U,
$$

and

$$
\alpha_{U,p} = (u_\eta)_Q \text{ for all } \alpha \geq \alpha_U. \quad (2.17)
$$

Proof. Since $p \in [1, +\infty]$ is fixed, we abbreviate $TV_p$, $u_{\alpha,p}$, and $\partial TV_p$, by $TV$, $u_\alpha$, and $\partial TV$, respectively, in this proof. We note that the null space

$$
\mathcal{N}(TV) = \{ u \in L^1(Q), \ TV(u) = 0 \},
$$

of the total variation semi-norm is the space of constant functions (see, e.g., [1]), which is a linear subspace of $L^1(Q)$. Let $\mathbb{P}[\cdot]$ denote the projection operator onto $\mathcal{N}(TV)$, and thus $\mathbb{P}[u_\eta]$ is a constant by (2.18). We claim that

$$
\frac{1}{\alpha}(u_\eta - \mathbb{P}[u_\eta]) \in \partial TV(0) \quad (2.19)
$$

for $\alpha > 0$ large enough. Indeed, since $\partial TV(0)$ has nonempty relative interior in $\mathcal{N}(TV)$ (see, e.g., [24]), we have that (2.19) holds for $\alpha \in \mathbb{R}^+$ sufficiently large since $u_\eta \in L^2(Q)$ and $\mathbb{P}[u_\eta]$ is a constant. Let $\alpha_0 > 0$ be large enough such that (2.19) hold. Then we have

$$
\frac{1}{\alpha_0}(u_\eta - \mathbb{P}[u_\eta]) \in \partial TV(0) = \partial TV(\mathbb{P}[u_\eta]),
$$

where in the last inequality we used again the fact that $\mathbb{P}[u_\eta]$ is a constant. That is, we have

$$
\frac{1}{\alpha_0}(u_\eta - \mathbb{P}[u_\eta]) \in \partial TV(\mathbb{P}[u_\eta]),
$$

and hence $\mathbb{P}[u_\eta]$ satisfies optimal condition of (T-L2) and we conclude that $\mathbb{P}[u_\eta] = u_\alpha_0$.

Therefore, we have $u_{\alpha_0}$ is a constant.

We claim next that $u_{\alpha_0} = (u_\eta)_Q$. Again by optimality condition we have

$$
\|u_{\alpha_0} - u_\eta\|_{L^2(Q)}^2 + \alpha TV(u_{\alpha_0}) \leq \|(u_\eta)_Q - u_\eta\|_{L^2(Q)}^2,
$$

that is

$$
\int_Q |u_{\alpha_0} - u_\eta|^2 \, dx \leq \int_Q |(u_\eta)_Q - u_\eta|^2 \, dx. \quad (2.20)
$$

Note that for $\lambda \in \mathbb{R}$,

$$
\frac{d}{d\lambda} \int_Q |\lambda - u_\eta|^2 \, dx = 2 \int_Q (\lambda - u_\eta) dx,
$$

which implies that the left hand side of (2.20) reaches the minimum value at $\lambda = (u_\eta)_Q$. Thus, we have $u_{\alpha_0} = (u_\eta)_Q$ and we deduce that $u_\alpha = (u_\eta)_Q$ for all $\alpha \geq \alpha_0$. 

Define
\[ \alpha_{U_p} := \inf \{ \alpha > 0, \; u_\alpha = (u_\eta)_Q \}, \]
and let \( \{ \alpha_n \}_{n=1}^\infty \subset \{ \alpha > 0, \; u_\alpha = (u_\eta)_Q \} \) be such that \( \alpha_n \searrow \alpha_{U_p} \). Thus, in view of Theorem 2.3, we conclude that \( u_{\alpha_{U_p}} = (u_\eta)_Q \), and hence the claim is true. \( \square \)

**Proposition 2.10.** Let \( u_\alpha, u_\eta \in L^2(Q) \) be given such that (2.9) hold. Then, there exists \( \alpha_U \in \mathbb{R}^+ \) such that the following assertions hold.

1. For all \( \alpha \geq \alpha_U/2 \) and \( p \in [1, +\infty] \), we have

   \[ TV_p(u_{\alpha,p}) = 0 \text{ and } u_{\alpha,p} = (u_\eta)_Q. \] (2.21)

2. The value of \( \alpha_U \) can be determined numerically.

**Proof.** For each \( p \in [1, +\infty] \), let \( \alpha_{U_p} > 0 \) be obtained from Lemma 2.9. We claim that

\[ \sup \{ \alpha_{U,p} : p \in [1, +\infty] \} < +\infty. \]

Take two arbitrary \( p_1 \) and \( p_2 \) such that \( 1 \leq p_1 \leq p_2 \leq +\infty \). For \( \alpha > 0 \) fixed, we have by optimality condition of (7-L2) that

\[ \|u_\eta - u_{\alpha,p_1}\|_{L^2(Q)}^2 + \alpha TV_{p_1}(u_{\alpha,p_1}) \leq \|u_\eta - u_{\alpha,p_2}\|_{L^2(Q)}^2 + \alpha TV_{p_1}(u_{\alpha,p_2}) \]
and

\[ \|u_\eta - u_{\alpha,p_2}\|_{L^2(Q)}^2 + \alpha TV_{p_2}(u_{\alpha,p_2}) \leq \|u_\eta - u_{\alpha,p_1}\|_{L^2(Q)}^2 + \alpha TV_{p_2}(u_{\alpha,p_1}). \]

Summing the above two inequalities yields

\[ 0 \leq TV_{p_1}(u_{\alpha,p_1}) - TV_{p_2}(u_{\alpha,p_1}) \]
\[ \leq TV_{p_1}(u_{\alpha,p_2}) - TV_{p_2}(u_{\alpha,p_2}) \]
\[ \leq (\mathcal{N}^{1/p_2 - 1/p_1} - 1)TV_{p_2}(u_{\alpha,p_2}), \]

where at the first and last inequality we used Remark 2.2. Thus, by (2.17) and letting \( \alpha = \alpha_{U_{p_2}} \), we infer that

\[ 0 \leq TV_{p_1}(u_{\alpha_{U_{p_2}},p_1}) - TV_{p_2}(u_{\alpha_{U_{p_2}},p_1}) \leq 0, \]

which, in turn, yields

\[ TV_{p_1}(u_{\alpha_{U_{p_2}},p_1}) = TV_{p_2}(u_{\alpha_{U_{p_2}},p_1}). \] (2.22)

By Remark 2.2 again, we have (2.22) holds unless \( u_{\alpha_{U_{p_2}},p_1} \in \mathcal{N}(TV_{p_2}) \), which implies that \( u_{\alpha_{U_{p_2}},p_1} \) must be a constant. Hence, by the argument used in Lemma 2.9 we conclude that

\[ u_{\alpha_{U_{p_2}},p_1} = (u_\eta)_Q \text{ and } \alpha_{U_{p_2}} \geq \alpha_{U_{p_1}}. \]

Therefore, we have

\[ \sup \{ \alpha_{U,p} : p \in [1, +\infty] \} \leq \alpha_{U,\infty} < +\infty, \]

and we conclude Assertion 1 by letting \( \alpha_U := 2\alpha_{U,\infty}. \)

We notice that, by Lemma 2.5 again, the function

\[ f(\alpha) = TV_{+\infty}(u_{\alpha,+,\infty}) \]

is continuous monotone decreasing and \( f(\alpha_{U,+,\infty}) = 0. \) Hence, we can apply Newton descent to compute \( \alpha_{U,\infty} \) numerically, which concludes Assertion 2. \( \square \)

We are now ready to proof Theorem 2.7.
Proof of Theorem 2.7. Let \( u_\varepsilon \in BV(Q) \) and \( u_\eta \in L^2(Q) \) be given such that (2.9) holds, and recall the definition of the training ground \( \mathcal{T} \), the assessment function \( \mathcal{A}(\alpha, p) \), and the optimal set \( \mathcal{A}[\mathcal{T}] \) from (1.2), (1.3), and (T-L1). Let
\[
\begin{align*}
m_\mathcal{T} := \inf \left\{ \| u_{\alpha,p} - u_\varepsilon \|_{L^2(Q)} : (\alpha, p) \in \mathcal{T} \right\}.
\end{align*}
\]
We claim first that \( \mathcal{A}[\mathcal{T}] \) is not empty. Let \( \{(\alpha_n, p_n)\}_{n=1}^\infty \subset \mathcal{T} \) be a minimizing sequence obtained from (T-L1) such that
\[
\lim_{n \to \infty} \| u_{\alpha_n,p_n} - u_\varepsilon \|_{L^2(Q)} = m_\mathcal{T}.
\]
Then, up to a subsequence, there exists \((\tilde{\alpha}, \tilde{p}) \in [0, +\infty] \times [1, +\infty] \) such that \((\alpha_n, p_n) \to (\tilde{\alpha}, \tilde{p})\). Suppose for a moment that \( \tilde{\alpha} \in (0, +\infty] \). Then, in view of Theorem 2.3 and the properties of \( \Gamma \)-convergence, we have
\[
\begin{align*}
u_{\alpha_n,p_n} \Rightarrow & \ u_{\tilde{\alpha},\tilde{p}} \text{ weakly* in } BV(Q) \text{ and strongly in } L^1(Q).
\end{align*}
\]
Thus, we conclude that
\[
\begin{align*}
\| u_{\tilde{\alpha},\tilde{p}} - u_\varepsilon \|_{L^2(Q)} \leq \liminf_{n \to \infty} \| u_{\alpha_n,p_n} - u_\varepsilon \|_{L^2(Q)} = m_\mathcal{T},
\end{align*}
\]
which implies \((\tilde{\alpha}, \tilde{p}) \in \mathcal{A}[\mathcal{T}]\).

Now we claim that \( \inf \alpha_n > 0 \). Indeed, assume by contradiction that \( \alpha_n \nearrow 0 \), and in this case we already showed in (2.16) that \( u_{\alpha_n,p_n} \to u_\eta \) in \( L^2 \) strong. Therefore, we have that
\[
\begin{align*}
m_\mathcal{T} = \liminf_{n \to \infty} \| u_{\alpha_n,p_n} - u_\varepsilon \|_{L^2(Q)} \geq \| u_\eta - u_\varepsilon \|_{L^2(Q)},
\end{align*}
\]
which contradicts Proposition 2.8. Thus, we conclude that \( \inf \alpha_n > 0 \), which implies \( \tilde{\alpha} > 0 \).

We next claim that there exists at least one optimal solution \((\alpha_T, p_T) \in \mathcal{A}[\mathcal{T}] \) such that \( \alpha_T \leq \alpha_U < +\infty \), where \( \alpha_U \) is obtained from Proposition 2.10. Suppose for all \((\alpha_T, p_T) \in \mathcal{A}[\mathcal{T}] \) such that \( \alpha_U < \alpha_T \leq +\infty \). Then, take arbitrary \((\alpha_0, p_0) \in \mathcal{A}[\mathcal{T}], (2.21) \) implies that
\[
\begin{align*}
\mathcal{A}(\alpha_0, p_0) = \mathcal{A}(\alpha_0, p_0) = m_\mathcal{T}, \text{ for all } \alpha \geq \alpha_U/2.
\end{align*}
\]
In another word, we have \((\alpha_U/2, p_0) \in \mathcal{A}[\mathcal{T}] \) as desired.

Therefore, we conclude that there exists at least one \((\alpha_T, p_T) \in \mathcal{A}[\mathcal{T}] \) such that
\[
0 < \alpha_T < \alpha_U < +\infty,
\]
which completes the proof of Theorem 2.7. \( \square \)

2.4. Extension of \( \ell^p \)-anisotropic total variation via Finsler metrics. We can further extend the \( \ell^p \)-(an)-isotropic total variational by using the Finsler metric (see [2] and Definition 2.11). Let \( |\cdot|_\omega : \mathbb{R}^N \to [0, +\infty) \) be a Finsler metric. That is, we assume that the function \( |\cdot| \) is convex and satisfies the properties
\[
|\alpha x|_* \leq |\alpha|_* |x|_* \text{ for } x \in \mathbb{R}^N, \ a \geq 0,
\]
where \( C \in \mathbb{R}^+ \) is a positive constant. Then, we define the \( \omega \)-total variation by
\[
TV_\omega(u) := \sup \left\{ \int_Q u \operatorname{div} \varphi \, dx : \varphi \in C_c^\infty(Q; \mathbb{R}^N), \ |\varphi|_* < 1 \right\} < +\infty.
\]

Definition 2.11. We say a collection \( \mathcal{F} \) of Finsler metrics is training compatible if the following assertions hold.
1. For any \( \omega \in \mathbb{F} \), \( |\cdot|_\omega : \mathbb{R}^N \to [0, +\infty) \) is a convex, positively 1-homogeneous function, and \( |x|_\omega > 0 \) if \( x \neq 0 \).

2. We denote the unit sphere of \( \omega \) by

\[
S_\omega := \left\{ x \in \mathbb{R}^N : |x|_\omega^* = 1 \right\}.
\]

Then, we say \( \omega_n \xrightarrow{p} \omega \) in \( \mathbb{F} \) if

\[
\text{dist}(S_{\omega_n}, S_\omega) \to 0.
\]

3. (compactness) For any sequence \( \{\omega_n\}_{n=1}^\infty \subset \mathbb{F} \), there exists a subsequence, still denote by \( \omega_n \), such that \( \omega_n \xrightarrow{p} \omega \) in \( \mathbb{F} \).

We present a similar version of Theorem 2.7 but with \( TV_\omega \) variation. First, we introduce the training scheme \( (T_F) \) by

Level 1. \( (\alpha_T, \omega_T) \in A[T_F] := \operatorname{arg min} \left\{ \|u_{\alpha, \omega} - u_c\|_{L^2(Q)}^2 : (\alpha, \omega) \in T_F \right\} \), \( (T_F\text{-L1}) \)

Level 2. \( u_{\alpha, \omega} := \operatorname{arg min} \left\{ \|u - u_\eta\|_{L^2(Q)}^2 + \alpha TV_\omega(u) : u \in BV_\omega(Q) \right\} \). \( (T_F\text{-L2}) \)

with the training ground

\[
T_F := \mathbb{R}^+ \times \mathbb{F}.
\]

**Theorem 2.12** (Existence of solutions of scheme \( (T_F) \)). Let \( u_c \) and \( u_\eta \in BV(Q) \) be given such that

\[
\inf \{ TV_\omega(u_\eta) : \omega \in \mathcal{W} \} > \sup \{ TV_\omega(u_c) : \omega \in \mathbb{F} \}.
\]

Then, the training scheme \( T_F (T_F\text{-L1})-(T_F\text{-L2}) \) admits at least one pair of solution \( (\alpha_T, \omega_T) \in T_F \), provided that \( \mathbb{F} \) is training compatible.

Proof. The proof can be obtained by following line by line of the argument presented in the proof of Theorem 2.7. In particular, the equivalent property (2.3), which used extensively in the proof of Theorem 2.3, can be replaced by assertions 2 and 3 in Definition 2.11. \( \square \)

We conclude this section by presenting several examples of training compatible \( \mathbb{F} \).

1. The \( \ell^p \text{- Euclidean norm defined in (2.1)} \). That is, we define

\[
\mathbb{F} := \left\{ |x|_p := \left( \sum_{i=1}^N |x_i|^p \right)^{1/p} : p \in [1, +\infty] \right\}.
\]

2. The skewed \( \ell^p \text{- Euclidean norm.} \)

\[
\mathbb{F} := \left\{ |x|_{ap} := \left( \sum_{i=1}^N a_i |x_i|^p \right)^{1/p} : p \in [1, +\infty], \sum_{i=1}^N a_i = 1, a_i > 0 \right\}.
\]

3. **Learning of acceptable optimal solutions**

3.1. **Non-convexity of the assessment function and counterexamples.** We present an explicit counterexample in one dimension \( (N = 1) \) to show that the assessment function \( \mathcal{A}(\alpha, p) \) is not quasi-convex. Note that as \( N = 1 \), we have \( |x|_p = |x|_2 \) for all \( p \in [1, +\infty] \). Thus, we only need to consider the case in which \( p = 2 \) and we abbreviate \( \mathcal{A}(\alpha, p) \) by \( \mathcal{A}(\alpha) \) in Section 3.1.
We define the corrupted signal $u_\eta$ (red line in Figure 1A) and the clean signal $u_c$ to be (blue line in Figure 1A)

$$u_\eta(x) := \begin{cases} -10 & \text{if } x \in (0, 1/4) \\ 2 & \text{if } x \in (1/4, 1/2) \\ 98 & \text{if } x \in (1/2, 3/4) \\ 110 & \text{if } x \in (3/4, 1), \end{cases}$$

and

$$u_c(x) := \begin{cases} 0 & \text{if } x \in (0, 1/4) \\ 20 & \text{if } x \in (1/4, 1/2) \\ 80 & \text{if } x \in (1/2, 3/4) \\ 100 & \text{if } x \in (3/4, 1). \end{cases}$$

Figure 1. Figure 1B shows that $A(\alpha)$ is not convex at $\alpha = 1.5$.

By [28] we can explicitly compute that

$$u_\alpha(x) = \begin{cases} -10 + 8\alpha & \text{if } x \in (0, 1/4) \\ 2 & \text{if } x \in (1/4, 1/2) \\ 98 & \text{if } x \in (1/2, 3/4) \\ 110 - 8\alpha & \text{if } x \in (3/4, 1) \end{cases},$$

for $0 \leq \alpha \leq 3/2$, and consequently

$$A(\alpha) = \frac{1}{4} |8\alpha - 10|^2 + \frac{1}{4} |110 - 8\alpha - 100|^2 + 18^2/2 = \frac{1}{4} |8\alpha - 10|^2 + \frac{1}{4} |10 - 8\alpha|^2 + 18^2/2,$$

for $0 \leq \alpha \leq 3/2$. Hence, we have

$$A'(\alpha) := \frac{d}{d\alpha} A(\alpha) = 8(8\alpha - 10) \text{ and } A''(\alpha) = 64 > 0$$

for $0 \leq \alpha \leq 3/2$. That is, $A(\alpha)$ is convex and decreasing for $0 \leq \alpha \leq 5/4$ and increasing for $5/4 \leq \alpha \leq 3/2$ (see the first convex part in Figure 1B).

Next, by applying [28] again, we have, for $3/2 \leq \alpha \leq 12$, that

$$u_\alpha(x) = \begin{cases} 2 + 4(\alpha - 1.5) & \text{if } x \in (0, 1/2) \\ 98 - 4(\alpha - 1.5) & \text{if } x \in (1/2, 1). \end{cases}$$

Hence, we have, for $3/2 \leq \alpha \leq 12$,

$$A(\alpha) = \frac{1}{4} |2 + 4(\alpha - 1.5)|^2 + \frac{1}{4} |20 - (2 + 4(\alpha - 1.5))|^2.$$
1. As obtained from Proposition 2.10. Then the following assertions hold.

Theorem 3.3. Let $\alpha T$ be given such that (2.9) is satisfied, and let $\alpha U$ be obtained from Proposition 2.10. Then the following assertions hold.

1. As $l \to \infty$, we have that

$$\text{dist}(A[T], A[T_l]) \to 0.$$
2. For each \( l \in \mathbb{N} \), there holds
\[
|A(\alpha T_l, p_T) - A(\alpha T, p_T)| \leq \sqrt{\alpha U}
\left[
\frac{1}{l^{1/2}} + 2 \left(1 - N^{-1/\sqrt{l}}\right)^{1/2}\right]
[TV_1(u_\eta)]^{1/2}.
\] (3.3)

This theorem will be proved in several propositions.

**Proposition 3.4.** Recall \( T_l \) from Definition 3.1 and the optimal set \( A[T_l] \) defined by
\[
A[T_l] = \arg\min \{A(\alpha, p) : (\alpha, p) \in T_l\}, \quad \text{for each } l \in \mathbb{N}.
\]
Then, all cluster points of sequence of sets \( \{A[T_l]\}_{l=1}^{\infty} \) belongs to the collection \( A[T] \).

**Proof.** We claim first that
\[
\min \{A(\alpha, p) : (\alpha, p) \in T_l\} := mL := \min \{A(\alpha, p) : (\alpha, p) \in T\}. \tag{3.4}
\]
Not that, for each \( l \in \mathbb{N} \) that \( T_l \subseteq T_{l+1} \) by Remark 3.2. Thus, we have
\[
m_l \geq m_{l+1} \geq \cdots \geq 0.
\] (3.5)

In view of Monotone Convergence Theorem, with no further subsequence extracted, there exists \( \bar{m} \geq 0 \) such that
\[
m_l \searrow \bar{m}.
\] (3.6)

We show that \( \bar{m} = m \). Suppose \( \bar{m} > m \). Then there exists \( \bar{l} \in \mathbb{N} \) such that
\[
m_l > m + \frac{1}{2} (\bar{m} - m), \quad \text{for all } l \geq \bar{l}.
\] (3.7)

On the other hand, by (3.2), for any \( (\alpha_l, p_{T_l}) \in A[T] \) we can extract a sequence \( \{(\alpha_i, p_i)\}_{i=1}^{\infty} \subset T \), in which \( (\alpha_i, p_i) \in T_i \) for each \( l \in \mathbb{N} \), such that \( (\alpha_i, p_i) \to (\alpha T_l, p_T) \). Thus, by (3.4) we have
\[
A(\alpha_l, p_l) \geq m_l.
\] (3.8)

Next, by Theorem 2.3 we have
\[
A(\alpha_l, p_l) \to A(\alpha T_l, p_T) = m,
\]
which implies that there exists \( \bar{l} \in \mathbb{N} \) such that
\[
A(\alpha_l, p_l) \leq m + \frac{1}{4} (\bar{m} - m), \quad \text{for all } l \geq \bar{l}.
\] (3.8)

Hence, by (3.6), (3.7), and (3.8), we must have
\[
m_l \leq m + \frac{1}{4} (\bar{m} - m) < m + \frac{1}{2} (\bar{m} - m) < m_l,
\]
which is a contradiction.

To finish, we point out that all cluster points of the sequence \( \{A[T_l]\}_{l=1}^{\infty} \) satisfy (3.4) since there is no subsequence extracted from (3.5) due to the property of the monotone convergence theorem, and hence we conclude this proposition. \( \square \)

**Proposition 3.5.** Let \( \alpha > 0, \varepsilon > 0, p \in [1, +\infty] \). Then we have
\[
\|u_{\alpha + \varepsilon, p} - u_{\alpha, p}\|_{L^2(Q)}^2 \leq \varepsilon TV_p(u_{\alpha, p}).
\]
Proof. By the minimality of \( u_{\alpha,p} \), there holds
\[
u_{\alpha,p} - u_q = -\alpha \partial TV_p(\mathcal{H})
\]
and
\[
u_{\alpha+\varepsilon,p} - u_q = -(\alpha + \varepsilon) \partial TV_p(\mathcal{H}).
\]
Subtracting one from another, we have that
\[
u_{\alpha,p} - \nu_{\alpha+\varepsilon,p} = \alpha [\partial TV_p(u_{\alpha,p}) - \partial TV_p(u_{\alpha+\varepsilon,p})] - \varepsilon \partial TV_p(u_{\alpha+\varepsilon,p}) = 0.
\]
Next, by multiplying with \( u_{\alpha,p} - u_{\alpha+\varepsilon,p} \) and integrating over \( Q \), we deduce that
\[
\|u_{\alpha+\varepsilon,p} - u_{\alpha,p}\|_{L^2(Q)}^2 + \alpha \langle u_{\alpha,p} - u_{\alpha+\varepsilon,p}, \partial TV_p(u_{\alpha,p}) - \partial TV_p(u_{\alpha+\varepsilon,p}) \rangle_{L^2} - \varepsilon \langle \partial TV_p(u_{\alpha+\varepsilon,p}), u_{\alpha,p} - u_{\alpha+\varepsilon,p} \rangle_{L^2} = 0.
\]
Since \( \partial TV_p \) is a maximal monotone operator (see [12, Proposition 5.5, page 25]), we obtain that
\[
\langle u_{\alpha,p} - u_{\alpha+\varepsilon,p}, \partial TV_p(u_{\alpha,p}) - \partial TV_p(u_{\alpha+\varepsilon,p}) \rangle_{L^2} \geq 0.
\]
This, and together with (3.9), we have that
\[
\|u_{\alpha+\varepsilon,p} - u_{\alpha,p}\|_{L^2(Q)}^2 \leq \varepsilon TV(u_{\alpha,p}),
\]
which concludes the proof. \( \Box \)

Proposition 3.6. Let \( p \geq 1, \varepsilon > 0, \alpha \in \mathbb{R}^+ \). Then we have
\[
\|u_{\alpha,p+\varepsilon} - u_{\alpha,p}\|_{L^2(Q)}^2 \leq \alpha \left[ 1 - N^{1/(p+\varepsilon)-1/p} \right] [TV_{p+\varepsilon}(u_{\alpha,p}) + TV_{p+\varepsilon}(u_{\alpha+\varepsilon,p})].
\]

Proof. Instead of using the sub-gradient operator as used in Proposition 3.5, we proceed with the first variation of \((\mathcal{T}_Q,L^2)\). Since in this argument \( \alpha \in \mathbb{R}^+ \) is fixed, we abbreviate \( u_{\alpha,p} \) by \( u_p \).

As first suggested in [27], we can regularize the \( TV_p \) seminorm by a factor \( \delta > 0 \) and consider
\[
u^\delta_p := \arg \min \{ \|u - u\|_{L^2(Q)}^2 + \alpha TV^\delta_p(u) : u \in BV(Q) \}
\]
where
\[
TV^\delta_p(u) := \int_Q \sqrt{|\nabla u^\delta_p|^2 + \delta dx}.
\]
The new \( TV^\delta_p \) seminorm is differentiable in \( u \). For arbitrary \( v \in BV(Q) \), we write the first variation of (3.10) as follows
\[
\int_Q (\nu^\delta_{p+\varepsilon} - u_q) v dx + \alpha \int_Q \frac{\nabla \nu^\delta_{p+\varepsilon}}{\sqrt{\left|\nabla \nu^\delta_{p+\varepsilon}\right|^2 + \delta}} \nabla v dx = 0
\]
and
\[
\int_Q (\nu^\delta_p - u_q) v dx + \alpha \int_Q \frac{\nabla \nu^\delta_p}{\sqrt{\left|\nabla \nu^\delta_p\right|^2 + \delta}} \nabla v dx = 0.
\]
Subtracting one from another, we have that

$$\frac{1}{\alpha} \int_Q (u^\delta_{p+\varepsilon} - u^\delta_p) v dx = \int_Q \left[ \frac{\nabla u^\delta_p}{\sqrt{|\nabla u^\delta_p|^2 + \delta}} - \frac{\nabla u^\delta_{p+\varepsilon}}{\sqrt{|\nabla u^\delta_{p+\varepsilon}|^2 + \delta}} \right] \nabla v dx$$

(3.11)

$$= \int_Q \left[ \frac{\nabla u^\delta_p}{\sqrt{|\nabla u^\delta_p|^2 + \delta}} - \frac{\nabla u^\delta_{p+\varepsilon}}{\sqrt{|\nabla u^\delta_{p+\varepsilon}|^2 + \delta}} \right] \nabla v dx$$

$$+ \int_Q \left[ \frac{\nabla u^\delta_{p+\varepsilon}}{\sqrt{|\nabla u^\delta_{p+\varepsilon}|^2 + \delta}} - \frac{\nabla u^\delta_p}{\sqrt{|\nabla u^\delta_p|^2 + \delta}} \right] \nabla v dx.$$ 

Set \( v := u^\delta_{p+\varepsilon} - u^\delta_p \). We compute that

$$\int_Q \left[ \frac{\nabla u^\delta_p}{\sqrt{|\nabla u^\delta_p|^2 + \delta}} - \frac{\nabla u^\delta_{p+\varepsilon}}{\sqrt{|\nabla u^\delta_{p+\varepsilon}|^2 + \delta}} \right] \nabla v dx$$

(3.12)

$$= \int_Q \left[ -\text{div} \frac{\nabla u^\delta_p}{\sqrt{|\nabla u^\delta_p|^2 + \delta}} + \text{div} \frac{\nabla u^\delta_{p+\varepsilon}}{\sqrt{|\nabla u^\delta_{p+\varepsilon}|^2 + \delta}} \right] (u^\delta_{p+\varepsilon} - u^\delta_p) dx$$

$$\leq 0,$$

where at the last inequality we used the fact that

$$-\text{div} \frac{\nabla u}{\sqrt{|\nabla u|^2 + \delta}} \in \partial TV^\delta_p(u)$$

is a maximal monotone operator.

Next, we compute that

$$\int_Q \left[ \frac{\nabla u^\delta_{p+\varepsilon}}{\sqrt{|\nabla u^\delta_{p+\varepsilon}|^2 + \delta}} - \frac{\nabla u^\delta_p}{\sqrt{|\nabla u^\delta_p|^2 + \delta}} \right] \nabla u^\delta_{p+\varepsilon} dx$$

(3.13)

$$= \int_Q \left[ \frac{1}{\sqrt{|\nabla u^\delta_{p+\varepsilon}|^2 + \delta}} - \frac{1}{\sqrt{|\nabla u^\delta_p|^2 + \delta}} \right] (\nabla u^\delta_{p+\varepsilon} \cdot \nabla u^\delta_p) dx$$

$$\leq \int_Q \left[ \frac{1}{\sqrt{|\nabla u^\delta_{p+\varepsilon}|^2 + \delta}} - \frac{1}{\sqrt{|\nabla u^\delta_p|^2 + \delta}} \right] |\nabla u^\delta_{p+\varepsilon} \cdot \nabla u^\delta_p| \ dx$$

$$\leq \left[ 1 - \frac{1}{N^{1/p-1/p+\varepsilon}} \right] \int_Q \left[ \frac{1}{\sqrt{|\nabla u^\delta_{p+\varepsilon}|^2 + \delta}} \right] |\nabla u^\delta_{p+\varepsilon} \cdot \nabla u^\delta_p| \ dx$$

$$\leq \left[ 1 - \frac{1}{N^{1/p-1/p+\varepsilon}} \right] TV^\delta_{p+\varepsilon}(u^\delta_{p+\varepsilon}),$$
where at the second inequality we used (2.2) and Hölder inequality. We could similarly estimate that

\[
\int_Q \left[ \frac{\nabla u_{p+\varepsilon}^\delta}{\sqrt{|\nabla u_{p+\varepsilon}^\delta|^2 + \delta}} - \frac{\nabla u_{p+\varepsilon}^\delta}{\sqrt{|\nabla u_{p+\varepsilon}^\delta|^2 + \delta}} \right] (-\nabla u_p^\delta) \, dx
\]

\[
\leq \left[ 1 - \frac{1}{N^{1/(p+\varepsilon)}} \right] \int_Q \frac{1}{\sqrt{|\nabla u_{p+\varepsilon}^\delta|^2 + \delta}} \left| \nabla u_{p+\varepsilon}^\delta \right|_{L^{p+\varepsilon}} \, dx
\]

This, and together with (3.13), we conclude that

\[
\int_Q \left[ \frac{\nabla u_{p+\varepsilon}^\delta}{\sqrt{|\nabla u_{p+\varepsilon}^\delta|^2 + \delta}} - \frac{\nabla u_{p+\varepsilon}^\delta}{\sqrt{|\nabla u_{p+\varepsilon}^\delta|^2 + \delta}} \right] \nabla (u_{p+\varepsilon}^\delta - u_p^\delta) \, dx
\]

\[
\leq \left[ 1 - \frac{1}{N^{1/(p+\varepsilon)}} \right] \left[ TV_{p+\varepsilon}(u_p^\delta) + TV_{p+\varepsilon}(u_{p+\varepsilon}^\delta) \right].
\]

Hence, by (3.11), (3.12), and (3.14), we have

\[
\|u_p^\delta - u_{p+\varepsilon}^\delta\|_{L^2(Q)}^2 \leq \alpha \left[ 1 - N^{1/(p+\varepsilon)-1/p} \right] \left[ TV_{p+\varepsilon}(u_p^\delta) + TV_{p+\varepsilon}(u_{p+\varepsilon}^\delta) \right].
\]

Moreover, since \(u_p^\delta \to u_p\) in the strict topology of BV (see [27]), it follows that

\[
\|u_p - u_{p+\varepsilon}\|_{L^2(Q)}^2 = \lim_{\delta \to 0} \|u_p^\delta - u_{p+\varepsilon}^\delta\|_{L^2(Q)}^2
\]

\[
\leq \alpha \left[ 1 - N^{1/(p+\varepsilon)-1/p} \right] \limsup_{\delta \to 0} \left[ TV_{p+\varepsilon}(u_p^\delta) + TV_{p+\varepsilon}(u_{p+\varepsilon}^\delta) \right]
\]

\[
= \alpha \left[ 1 - N^{1/(p+\varepsilon)-1/p} \right] \left[ TV_{p+\varepsilon}(u_p) + TV_{p+\varepsilon}(u_{p+\varepsilon}) \right],
\]

which completes the proof of the proposition.

\[\square\]

We recall the reduced training ground \(\mathbb{T}\) from Definition 3.1.

**Corollary 3.7.** Let \((\alpha_1, p_1), (\alpha_2, p_2) \in \mathbb{T}\). Then we have

\[
\|u_{\alpha_2, p_2} - u_{\alpha_1, p_1}\|_{L^2(Q)} \leq \left[ \alpha_1 - \alpha_2 \right]^{1/2} + 2 \left( \alpha_U \left( 1 - N^{-[1/p_2-1/p_1]} \right) \right)^{1/2} \left[ TV_1(u_{p_1}) \right]^{1/2}.
\]

**Proof.** By Proposition 3.5 and Proposition 3.6, we observe that

\[
\|u_{\alpha_2, p_2} - u_{\alpha_1, p_1}\|_{L^2(Q)}
\]

\[
\leq \|u_{\alpha_2, p_2} - u_{\alpha_1, p_2}\|_{L^2(Q)} + \|u_{\alpha_1, p_2} - u_{\alpha_1, p_1}\|_{L^2(Q)}
\]

\[
\leq |\alpha_1 - \alpha_2| \left[ TV_{p_2}(u_{\alpha_1, p_2}) \right]^{1/2} + \left[ \alpha_1 \left( 1 - N^{-[1/p_2-1/p_1]} \right) \right] \left[ TV_{p_2}(u_{\alpha_1, p_2}) \right]^{1/2}.
\]

Next, in view of Lemma (2.5), we have \(TV_{p}(u_{\alpha, p}) \leq TV_{p}(u_{\eta}) \leq TV_{1}(u_{\eta})\) and hence we conclude that

\[
\|u_{\alpha_2, p_2} - u_{\alpha_1, p_1}\|_{L^2(Q)} \leq \left[ \alpha_1 - \alpha_2 \right]^{1/2} + 2 \left( \alpha_U \left( 1 - N^{-[1/p_2-1/p_1]} \right) \right)^{1/2} \left[ TV_1(u_{\eta}) \right]^{1/2}
\]

as desired. \[\square\]
We are now ready to prove Theorem 3.3.

**Proof of Theorem 3.3.** The Assertion 1 can be deduced from Proposition 3.4 directly.

We next prove Assertion 2. Let \((\alpha_1, p_1), (\alpha_2, p_2) \in \mathbb{T}\). By Corollary 3.7 we have that

\[
|\mathcal{A}(\alpha_1, p_1) - \mathcal{A}(\alpha_2, p_2)| 
= \left| \|u_{\alpha_1, p_1} - u_{e}\|_{L^2(Q)} - \|u_{\alpha_2, p_2} - u_{e}\|_{L^2(Q)} \right| 
\leq \left| \alpha_1 - \alpha_2 \right|^{1/2} \left( \alpha_U \left( 1 - N^{-1/2p_2 - 1/p_1} \right) \right)^{1/2} \left[ TV_1(u_p) \right]^{1/2}. 
\]

Let \((\alpha_T, p_T) \in \mathcal{A}[\mathbb{T}]\) and a minimizing sequence \((\alpha_{T_l}, p_{T_l}) \in \mathcal{A}[\mathbb{T}_l]\) such that \((\alpha_{T_l}, p_{T_l}) \rightarrow (\alpha_T, p_T)\) as \(l \rightarrow \infty\). Also, for each \(l \in \mathbb{N}\), we fix a finite grid \(G_l(i_l, j_l)\) (recall \(3.1\)) such that \((\alpha_T, p_T) \in G_l(i_l, j_l) \subset \mathbb{R}^2\). (3.16)

Thus, since \(G_l(i_l, j_l)\) is closed, we have

\[
\mathcal{A}(\alpha_T, p_T) = \min \{\mathcal{A}(\alpha, p) : (\alpha, p) \in G_l(i_l, j_l)\}. 
\]

Also, in view of (3.15) and Theorem 2.3, there holds

\[
\max \{\mathcal{A}(\alpha, p) : (\alpha, p) \in G_l(i_l, j_l)\} - \min \{\mathcal{A}(\alpha, p) : (\alpha, p) \in G_l(i_l, j_l)\} 
\leq \left[ \delta_l^{1/2} + 2(\alpha_U \left( 1 - N^{-\delta_l} \right))^{1/2} \right] \left[ TV_1(u_p) \right]^{1/2}. 
\]

Next, if at step \(l\) that \((\alpha_{T_l}, p_{T_l}) \in G_l(i_l, j_l)\), we can immediately deduce that

\[
\mathcal{A}(\alpha_{T_l}, p_{T_l}) - \mathcal{A}(\alpha_T, p_T) 
\leq \max \{\mathcal{A}(\alpha, p) : (\alpha, p) \in G_l(i_l, j_l)\} - \mathcal{A}(\alpha_{T_l}, p_{T_l}) 
\leq \max \{\mathcal{A}(\alpha, p) : (\alpha, p) \in G_l(i_l, j_l)\} - \min \{\mathcal{A}(\alpha, p) : (\alpha, p) \in G_l(i_l, j_l)\}, 
\]

where for the last inequality we used (3.17).

If \((\alpha_{T_l}, p_{T_l}) \notin G_l(i_l, j_l)\), then in view of the definition of \((\alpha_{T_l}, p_{T_l})\), we must have

\[
\max \{\mathcal{A}(\alpha, p) : (\alpha, p) \in G_l(i_l, j_l) \cap \mathbb{T}_l\} \geq \mathcal{A}(\alpha_{T_l}, p_{T_l}). 
\]

Hence, by (3.19) we again obtain that

\[
\mathcal{A}(\alpha_{T_l}, p_{T_l}) - \mathcal{A}(\alpha_T, p_T) 
\leq \max \{\mathcal{A}(\alpha, p) : (\alpha, p) \in G_l(i_l, j_l) \cap \mathbb{T}_l\} - \mathcal{A}(\alpha_{T_l}, p_{T_l}) 
\leq \max \{\mathcal{A}(\alpha, p) : (\alpha, p) \in G_l(i_l, j_l)\} - \min \{\mathcal{A}(\alpha, p) : (\alpha, p) \in G_l(i_l, j_l)\}, 
\]

where at the last inequality we used the assumption (3.16). In the end, by (3.15), (3.18), and (3.20), we observe that

\[
\mathcal{A}(\alpha_{T_l}, p_{T_l}) - \mathcal{A}(\alpha_T, p_T) 
\leq \max \{\mathcal{A}(\alpha, p) : (\alpha, p) \in \partial G_l(i_l, j_l)\} - \min \{\mathcal{A}(\alpha, p) : (\alpha, p) \in G_l(i_l, j_l)\} 
\leq \left[ \delta_l^{1/2} + 2(\alpha_U \left( 1 - N^{-\delta_l} \right))^{1/2} \right] \left[ TV_1(u_p) \right]^{1/2} 
\leq \sqrt{\alpha_U} \left[ 1/2 + 2 \left( 1 - N^{-1/\sqrt{1}} \right) \right] \left[ TV_1(u_p) \right]^{1/2} 
\]

and hence the thesis. \(\square\)
Remark 3.8. We point out that the optimal solutions \((\alpha_T, p_T) \in \mathbb{A}[T]\) can be determined precisely since \(\# \{ T_I \} < +\infty\) at each \(l \in \mathbb{N}\). Then, for any given acceptable error, we can determine the required approximation step \(l\) by using (3.3).

3.2.1. A relaxation of corrupted image \(u_\eta\). The assumption in Theorem 3.3 that \(u_\eta \in BV(Q)\) is a rather strong one and, in fact, not realistic for image denoising. We argue, however, that the error bound \((T-L)\) can still be used in practice by replacing the noisy image \(u_\eta\) with an approximation that has bounded variation. To be precise, we consider a sequence \(\{ u^K_\eta \}_{K=1}^\infty \subset BV(Q)\) such that

\[ u^K_\eta \to u_\eta \text{ strongly in } L^2(Q) \text{ and } TV(u^K_\eta) < +\infty \]  

(3.21)

and introduce the training scheme \(T^K \) \((T^K-(T-L))\) as follows.

Level 1. \( (\alpha^K_T, p^K_L) \in A^K[T] := \arg \min \left\{ \| u^K_{\alpha,p} - u_c \|_{L^2(Q)} : (\alpha, p) \in T \right\} \) . \( (T^K-(T-L))\)

Level 2. \( u^K_{\alpha,p} := \arg \min \left\{ \| u - u^K_\eta \|_{L^2(Q)} + \alpha TV_p(u) : u \in BV(Q) \right\} \). \( (T^K-(T-L))\)

We also define the assessment function with respect to \(u^K_\eta\) by

\[ \mathcal{A}^K(\alpha, p) := \| u^K_{\alpha,p} - u_c \|_{L^2(Q)} \]  

\[ \alpha \]

Theorem 3.9. Let \( K \in \mathbb{N}, u_c \in BV(Q) \) and \( u_\eta \in L^2(Q) \) be given. Let \( u^K_\eta \) be defined as in (3.21). Then the following assertions hold.

1. As \( K \to \infty \), we have

\[ \text{dist}(A^K[T], \mathbb{A}[T]) \to 0. \]

2. For each \( K \in \mathbb{N} \), there holds

\[ |A^K(\alpha^K, p^K) - A(\alpha_T, p_T)| \leq 2 \| u^K_\eta - u_\eta \|_{L^2(Q)}. \]

Before we prove Theorem 3.9, we first prove an enhanced version of Proposition 2.4.

Proposition 3.10. Let \( \{ u^K_\eta \}_{K=1}^\infty \subset BV(Q) \) and \( \{ \alpha^K, p^K \}_{K=1}^\infty \subset T \) be such that

\[ u^K_\eta \to u_\eta \text{ strongly in } L^2 \]  

(3.22)

and

\[ (\alpha^K, p^K) \to (\alpha, p) \in T. \]  

(3.23)

Then we have

\[ u^K_{\alpha^K, p^K} \to u_{\alpha,p} \text{ in } BV \]  

(3.24)

and

\[ \lim_{K \to \infty} TV_{p,K}(u^K_{\alpha^K, p^K}) = TV_p(u_{\alpha,p}). \]  

(3.25)

Proof. We assume first that \( \alpha > 0 \). By (3.22) and (3.23), there exist \( N^* \in \mathbb{N} \) such that

\[ \frac{1}{2} \alpha \leq \alpha_K \leq \alpha + 1 \text{ and } \| u^K_\eta \|^2_{L^2(Q)} \leq \| u_\eta \|^2_{L^2(Q)} + 1, \text{ for all } n \geq N^*. \]

It follows that

\[ \frac{1}{2} \alpha TV_2(u^K_{\alpha^K, p^K}) \leq \max \left\{ 1, N^{1/p_K-1/2} \right\} \frac{1}{2} \alpha TV_{p,K}(u^K_{\alpha^K, p^K}) \]

\[ \leq \alpha_K TV_{p,K}(u^K_{\alpha^K, p^K}) \leq \| u^K_{\alpha^K, p^K} - u^K_\eta \|^2_{L^2(Q)} + \alpha_K TV_{p,K}(u^K_{\alpha^K, p^K}) \]

\[ \leq \| u^K_\eta \|^2_{L^2(Q)} \leq \| u_\eta \|^2_{L^2(Q)} + 1 < +\infty. \]
Thus, we have
\[
\sup \left\{ \| u_{\alpha K,pK} \|_{BV(Q)} : n \in \mathbb{N} \right\} < +\infty,
\]
and, up to a (not-relabeled) subsequence, there exists \( w \in BV(Q) \) such that
\[
u_{\alpha K,pK}^K \rightharpoonup w \quad \text{in} \ BV \quad \text{and} \quad u_{\alpha K,pK}^K \to w \quad \text{in} \ L^1 \quad \text{and} \quad \text{a.e.}
\]
(3.26)
We claim that \( w = u_{\alpha,p} \) a.e.. Indeed, since \( u_{\alpha K,pK}^K \) is the unique minimizer of \((T^K-L2)\), we have that
\[
\| u_{\alpha,p} - u_{\eta}^K \|_{L^2(Q)}^2 + \alpha_K TV_{pK}(u_{\alpha,p}) \geq \| u_{\alpha K,pK} - u_{\eta}^K \|_{L^2(Q)}^2 + \alpha_K TV_{pK}(u_{\alpha K,pK}),
\]
and hence
\[
\liminf_{K \to \infty} \| u_{\alpha,p} - u_{\eta}^K \|_{L^2(Q)}^2 + \alpha_K TV_{pK}(u_{\alpha,p}) \geq \liminf_{K \to \infty} \| u_{\alpha K,pK} - u_{\eta}^K \|_{L^2(Q)}^2 + \alpha_K TV_{pK}(u_{\alpha K,pK}) \geq \| w - u_{\eta} \|_{L^2(Q)}^2 + \alpha TV_p(w),
\]
where at the last inequality we used Fatou’s lemma and (3.26). On the other hand, we have
\[
\limsup_{K \to \infty} \| u_{\alpha,p} - u_{\eta}^K \|_{L^2(Q)}^2 + \alpha_K TV_{pK}(u_{\alpha,p}) = \| u_{\alpha,p} - u_{\eta} \|_{L^2(Q)}^2 + \alpha TV_p(u_{\alpha,p}),
\]
where we used the fact that \( u_{\eta}^K \to u_{\eta} \) in \( L^2 \) and Proposition 2.6. Hence, by (3.27) and (3.28) we obtain that
\[
\| u_{\alpha,p} - u_{\eta} \|_{L^2(Q)}^2 + \alpha TV_p(u_{\alpha,p}) \geq \| w - u_{\eta} \|_{L^2(Q)}^2 + \alpha TV_p(w).
\]
Therefore, we must have \( u_{\alpha,p} = w \) since the minimizer of \((T-L2)\) is unique, which concludes (3.24) as desired.

We next claim (3.25). Indeed, the \( \liminf \) inequality
\[
\liminf_{K \to \infty} TV_{pK}(u_{\alpha K,pK}) \geq TV_p(u_{\alpha,p}) \tag{3.29}
\]
can be directly obtained from (2.6). Again, since \( u_{\alpha K,pK}^K \) is the unique minimizer, we observe that
\[
\| u_{\alpha K,pK}^K - u_{\eta}^K \|_{L^2(Q)}^2 + \alpha_K TV_{pK}(u_{\alpha K,pK}) \leq \| u_{\alpha,p} - u_{\eta} \|_{L^2(Q)}^2 + \alpha_K TV_{pK}(u_{\alpha,p}).
\]
By (3.24) we infer that
\[
\lim_{K \to \infty} \| u_{\alpha K,pK}^K - u_{\eta}^K \|_{L^2(Q)}^2 = \| u_{\alpha,p} - u_{\eta} \|_{L^2(Q)}^2,
\]
which, in turn, yields
\[
\| u_{\alpha,p} - u_{\eta} \|_{L^2(Q)}^2 + \alpha TV_p(u_{\alpha,p}) \leq \lim_{K \to \infty} \| u_{\alpha K,pK}^K - u_{\eta}^K \|_{L^2(Q)}^2 + \limsup_{K \to \infty} \alpha_K TV_{pK}(u_{\alpha K,pK}) \leq \limsup_{K \to \infty} \left[ \| u_{\alpha,p} - u_{\eta} \|_{L^2(Q)}^2 + \alpha_K TV_{pK}(u_{\alpha,p}) \right] \leq \| u_{\alpha,p} - u_{\eta} \|_{L^2(Q)}^2 + \alpha TV_p(u_{\alpha,p}),
\]
where at the last inequality we used Proposition 2.6 again. Thus, we conclude that
\[
\limsup_{K \to \infty} \alpha_K TV_{pK}(u_{\alpha K,pK}) = \alpha TV_p(u_{\alpha,p}).
\]
This, together with (3.29), we conclude (3.25) and hence the thesis for the case $\alpha > 0$.

Lastly, we assume $\alpha = 0$. In this case we have $u_{\alpha,p} = u_\eta$, and we could refer to the proof used in Theorem 2.7 to conclude our thesis. □

**Proof of Theorem 3.9.** The Assertion 1 can be directly deduced from Proposition 3.10. We focus on claiming Assertion 2. Let $u_{\alpha,p}$ and $u^K_{\alpha,p}$ be obtained from $(T\!-\!L_2)$ and $(T^K\!-\!L_2)$, respectively. By the optimality condition, we have

$$u_{\alpha,p} - u_\eta = -\alpha \partial TV_p(u_{\alpha,p})$$

and

$$u^K_{\alpha,p} - u^K_\eta = -\alpha \partial TV_p(u^K_{\alpha,p}).$$

Subtracting one from another, we deduce that

$$u_{\alpha,p} - u^K_{\alpha,p} + u^K_\eta - u_\eta = -\alpha \left[ \partial TV_p(u_{\alpha,p}) - \partial TV_p(u^K_{\alpha,p}) \right].$$

Hence, by multiplying $u_{\alpha,p} - u^K_{\alpha,p}$ on the both hand side and integrating over $Q$, we have that

$$\left\| u_{\alpha,p} - u^K_{\alpha,p} \right\|_{L^2(Q)}^2 + \langle u_{\alpha,p} - u^K_{\alpha,p}, u^K_\eta - u_\eta \rangle_{L^2}$$

$$= -\alpha \langle \partial TV_p(u_{\alpha,p}) - \partial TV_p(u^K_{\alpha,p}), u_{\alpha,p} - u^K_{\alpha,p} \rangle_{L^2}$$

$$\leq 0,$$

which yields that

$$\left\| u_{\alpha,p} - u^K_{\alpha,p} \right\|_{L^2(Q)} \leq \left\| u^K_\eta - u_\eta \right\|_{L^2(Q)}.$$

We conclude our thesis by following the argument used in the proof of Theorem 3.3. □

In [23], it is shown that if the noisy image $u_\eta$ is a piece-wise constant function, we can take advantage of this when numerically computing the solution $u_{\alpha,p}$. One good choice for the approximation sequence $u^K_\eta$, therefore, could be the piece-wise average of $u_\eta$ introduced as follows.

**Definition 3.11** (Piece-wise approximation of the noisy image). Let $Q = (0,1)^N$ and the corrupted image $u_\eta \in L^2(Q)$ be given. We define the $K$-resolution approximation $u^K_\eta$ of $u_\eta$ via its average

$$u^K_\eta(x) := \int_{Q_K(k_1,\ldots,k_N)} u_\eta \, dx \text{ for } x \in Q_K(k_1,\ldots,k_N), \quad (3.30)$$

where

$$Q_K(k_1,\ldots,k_N) := [k_1 - 1/K, k_1/K] \times \cdots \times [k_N - 1/K, k_N/K],$$

for $1 \leq k_1,\ldots,k_N \leq K$.

We note that $u^K_\eta$ defined in (3.30) satisfies (3.21).

As a result of Theorem 3.3 and Theorem 3.9, the following corollary can be established.

**Corollary 3.12.** Let $u_c \in BV(Q)$ and $u_\eta \in L^2(Q)$ be given. Then, for arbitrary $\varepsilon > 0$, there exists $l_\varepsilon \in \mathbb{N}$ large enough such that

$$|\mathcal{A}(\alpha_{T_{l_\varepsilon}}, p_{T_{l_\varepsilon}}) - \mathcal{A}(\alpha_T, p_T)| \leq \varepsilon.$$
Proof. Let \( \varepsilon > 0 \) be given. Then by Theorem 3.9, we could choose \( K_\varepsilon \in \mathbb{N} \) large enough so that
\[
\left| A^{K_\varepsilon}(\alpha^K_{T^\varepsilon}, p^K_{T^\varepsilon}) - A(\alpha_T, p_T) \right| \leq 2 \| u^K_{\eta} - u_\eta \|_{L^2(Q)} \leq \frac{1}{4} \varepsilon. \tag{3.31}
\]
On the other hand, by Theorem 3.3, we have that
\[
\left| A^{K_\varepsilon}(\alpha^K_{T^\varepsilon}, p^K_{T^\varepsilon}) - A^{K_\varepsilon}(\alpha^K_{T^\varepsilon}, p^K_{T^\varepsilon}) \right| \leq \sqrt{\frac{K}{U}} \left[ 1/4^{1/2} + 2 \left( 1 - N^{-1/\sqrt{T}} \right)^{1/2} \right] |TV_i(u^K_{\eta})|^{1/2},
\]
where by (3.30) we observe that \( TV_i(u^K_{\eta}) < +\infty \) even if \( u_\eta \notin BV(Q) \).

Hence, by taking \( l_\varepsilon \in \mathbb{N} \) large enough, and together with (3.31), we conclude
\[
\begin{align*}
&\left| A(\alpha_{T^\varepsilon}, p_{T^\varepsilon}) - A(\alpha_T, p_T) \right| \\
&\leq \left| A(\alpha_{T^\varepsilon}, p_{T^\varepsilon}) - A^{K_\varepsilon}(\alpha^K_{T^\varepsilon}, p^K_{T^\varepsilon}) \right| + \left| A^{K_\varepsilon}(\alpha^K_{T^\varepsilon}, p^K_{T^\varepsilon}) - A^{K_\varepsilon}(\alpha^K_{T^\varepsilon}, p^K_{T^\varepsilon}) \right| \\
&\leq \frac{1}{2} \varepsilon + \frac{1}{2} \varepsilon \leq \varepsilon,
\end{align*}
\]
as desired. \( \square \)

Then, based on Corollary 3.12, we suggest the following practical strategy for computing an acceptable solution \( u_{\alpha_{T^\varepsilon}, p_{T^\varepsilon}} \).

Let \( u_\varepsilon \in BV(Q) \) and \( u_\eta \in L^2(Q) \) be given. Let an acceptable error \( \varepsilon > 0 \) be given.

- **Initialization:** Compute \( \alpha_U \) defined in Proposition 2.10 and construct piece-wise constant function \( u^K_\eta \), defined in (3.30), such that
  \[
  \| u^K_\eta - u_\eta \|_{L^2(Q)} \leq \frac{1}{4} \varepsilon
  \]
is satisfied.

- **Step 1:** Submit \( u^K_\eta \) into Theorem 3.3, and determine step \( l \in \mathbb{N} \) so that the right hand side of (3.3) is less than \( \varepsilon/4 \).

- **Step 2:** Determine one optimal solution \( (\alpha_{T^\varepsilon}, p_{T^\varepsilon}) \in A[T^\varepsilon] \). By Theorem 3.3 and Theorem 3.9 we have that
  \[
  |A(\alpha_{T^\varepsilon}, p_{T^\varepsilon}) - A(\alpha_T, p_T)| \leq \varepsilon.
  \]

- **Step 3:** The reconstructed image \( u_{\alpha_{T^\varepsilon}, p_{T^\varepsilon}} \) is then an acceptable optimal solution defined in (1.4).

4. **Numerical simulations and conclusions**

4.1. **Simulations and insights.** We perform numerical simulations of the bilevel scheme \( T \) using the clean image \( u_\varepsilon \) and the noisy image \( u_\eta \) shown in the first and second picture in Figure 2, respectively, and we report that their total variations \( TV_i(u_\eta) = 264.5255 \) and \( TV_i(u_\varepsilon) = 98.4627 \). Note, that we are reporting the TV values of finite resolution digital images which coincide with their piece-wise constant approximation mentioned in Section 3.2.1. Ideally, a clean image \( u_\varepsilon \in BV(Q) \) can only be captured by a “super” camera which has infinite resolution. However, in the real world, such “super” camera, with infinite resolution, does not exist, and hence, in the numerical section, we assume that a finite \( K \in \mathbb{N} \) resolution clean image \( u^K_\varepsilon \) that we wish to capture by a real world digital camera is a piece-wise constant function, which is related to \( u_\varepsilon \) via its averages and defined in the way of
(3.30), with \( u_c \) in place of \( u_\eta \).

The principal sources of noise in digital images are introduced during acquisition, for example, the sensor noise caused by poor illumination, circuitry of a scanner, and the unavoidable shot noise of a photon detector. The noise is only generated during the acquiring of the image, i.e., it is only added to \( u^K \); and each time we acquire an image, we produce a different noise \( \eta^K \). Therefore, we propose to use a piecewise constant function \( \eta^K \) over \( Q^K \) to represent the noise at the resolution level \( K \in \mathbb{N} \), and we write

\[
 u^K_\eta := u^K_c + \eta^K.
\]

That is, when an image is taken with resolution \( K \in \mathbb{N} \), although we only wish to observe \( u^K_c \), the noise \( \eta^K \) is an unavoidable by-product, and hence the corrupted image \( u^K_\eta \) is produced. Therefore, we assume that all (corrupted) images captured by digital camera with finite resolution is already a piece-wise constant function, which implies that \( u_\eta \in BV(Q) \), and hence no relaxation is needed as studied in Section 3.2.1.

![Figure 2](image.png)

Figure 2. L-R: Clean image; noisy version (with artificial Gaussian noise); optimally reconstructed image \( u_{\alpha,p_T} \) provided by the Scheme \( \mathcal{T} \) (\( \mathcal{T}-\mathcal{L}1 \))- (\( \mathcal{T}-\mathcal{L}2 \)) where \( p_T = 2.6622 \); and the contour plot of the assessment function \( A(\alpha,p) \) at \( l = 10000 \). We note that \( A(\alpha,p) \) is indeed not quasi-convex.

The Level 2 problem (\( \mathcal{T}-\mathcal{L}2 \)) is solved via the primal-dual algorithm studied in [5, 4], as we can recast (\( \mathcal{T}-\mathcal{L}2 \)) as

\[
 \min_{u} \max_{v} \left\{ -\alpha \langle u, \text{div} \varphi \rangle + \|u - u_\eta\|_{L^2}^2 - \delta_{V_q}(v) \right\}.
\]

Here \( \delta_{V_q} \) denotes the indicator function of the set \( V_q \)

\[
 V_q := \{ v \in C_0^\infty(Q; \mathbb{R}^2) : |v|_{L^2} \leq 1 \}.
\]

For the sake of appropriate comparison, we use the finite training ground \( \mathbb{T}_l \) at \( l = 10000 \) to simulate the continuous training ground \( \mathbb{T} \), and we plot the contour image of \( A(\alpha,p) \) at \( \mathbb{T}_l \), \( l = 10000 \), in the last column in Figure 2. We summarize our simulation results and computed \( l \)'s from Theorem 3.3 in Table 1. Note that the step \( l \) predicted by Theorem 3.3 (shown in column 2) is rounded up to the nearest integer.

We also applied the estimate in Theorem 3.3 for the bilevel scheme \( \mathcal{B} \) and we observe for the example in Figure 2 that

\[
 \inf A(\alpha,p) < \inf A(\alpha,2),
\]

which indicates that the scheme \( \mathcal{T} \) in which we optimise over the parameter \( p \) indeed provides an improved reconstruction result compared with the \( TV_2 \) scheme \( \mathcal{B} \).
Acceptable error | $l$ estimated by Theorem 3.3 | Numerical error | Optimal $(\alpha_T, p_T)$
---|---|---|---
$\varepsilon = 0.2$ & $l = 37$ & 0.1018 & (0.04, 2.26)
$\varepsilon = 0.1$ & $l = 472$ & 0.0217 & (0.042, 2.263)
$\varepsilon = 0.05$ & $l = 7319$ & 0.0153 & (0.0488, 2.2622)

Table 1. Simulation for the example in Figure 2 for different discretisation levels $l$ selected from Theorem 3.3 with three acceptable errors. We also report that $TV_{2.6622}(u_c) = 14.6654$ while $TV_{2.2622}(u_{0.0488, 2.6622}) = 11.0291$.

Figure 3. Left column to right column: test images; noisy version (with artificial Gaussian noisy); optimal reconstructed image $u_{\alpha_T, 2}$ provided by Scheme $B$ ($B$-L1)-(B-L2) and optimal reconstructed image $u_{\alpha_T, p_T}$ provided by Scheme $T$ ($T$-L1)-(T-L2), respectively. We report that the optimal $p$ are achieved at $p_T = 1.0197$ and $p_T = 1.1359$ for test set in row one and two, respectively.

4.2. Conclusions and future works. In this work, we first constructed $\ell^p$-(an)-isotropic total variation $TV_p$ semi-norms and applied it into the imaging processing problems. This class of semi-norms can be viewed as a generalization of the standard total variation. Then, we introduce a semi-supervised learning scheme to optimize the underlying Euclidean parameter $p$ in $TV_p$. A further finite approximation method of such learning scheme allows us not only conclude the existence of global optimization but also allows us to numerically compute it, especially in the situation that the convex condition is missing.

We also want to remark a few words about the inefficiency of the finite approximation scheme studied above, and provide several ways to mitigate such inefficiency. The finite approximation scheme searches the global optimizer by walking through every grid point. Although the number of grid points is finite, the massive amount of them would inevitably cause long CPU time. One way to mitigate such problem is to implement a parallel computation method as the construction and searching procedure used in our finite approximation is in particular suitable for such acceleration method.
On the other hand, we observe from Table 1 that the numerical error (column 3) (which is what we actually obtained) is much smaller than the given acceptable error (column 1) (which is what we expect to obtain), which likely causes over-computing and hence wait of CPU time. This phenomenon is partially due to the large value of $TV(u_\eta)$ and $\alpha_U$. To mitigate this drawback, we observe from the numerical simulation, reported in Table 2, that the optimal solution usually has the property that

$$TV_{p_T}(u_{\alpha_T,p_T}) \approx TV_{p_T}(u_c). \tag{4.1}$$

| Test image       | optimal $p_T$ | $TV_{p_T}(u_\eta)$ | $TV_{p_T}(u_c)$ | $TV_{p_T}(u_{\alpha_T,p_T})$ |
|------------------|---------------|---------------------|-----------------|-------------------------------|
| Figure 2         | 2.6622        | 29.7651             | 14.6654         | 11.0291                       |
| Figure 3, row 1  | 1.0197        | 142.0373            | 98.4627         | 108.0941                      |
| Figure 3, row 2  | 1.1359        | 102.3078            | 48.7725         | 33.9762                       |

Table 2. Total variation of clean image, corrupted image, and optimal reconstructed image.

Thus, if we take (4.1) for guaranteed, we could reduce the range of regularization parameters $\alpha$ to be $[\alpha_L, \alpha_R] \subset [0, \alpha_U]$ such that

$$TV_p(u_{\alpha_L,p}) \geq TV_p(u_c) + \frac{1}{2}(TV_p(u_\eta) - TV_p(u_c)) > \frac{1}{2}TV_p(u_c) \geq TV_p(u_{\alpha_R,p}).$$

Then, we can reduce estimate (3.3) in Theorem 3.3 to

$$|A(\alpha_T,p_T) - A(\alpha_T,p_T)| \leq \sqrt{\alpha_R - \alpha_L} \left[1/l \right]^{1/2} + 2 \left[(1 - N^{-1/\sqrt{2}})^{1/2} (TV_4(u_\eta))^1/2\right].$$

We notice that this new estimate uses only the total variation of the clean image $u_c$, which is assumed to be much smaller than the variation of the corrupted image and the value of $\alpha_R - \alpha_L$ is also much smaller than $\alpha_U$. However, due to the length of this article and the fact that the estimate (4.1) already fully satisfies our purpose, we decide not to pursue on how to prove (4.1) but leave it for future work.

Another interesting direction is to understand which properties of a given image influence the value of optimal tuning parameter $\alpha$ and underlying Euclidean norm $p$ the most. The tuning parameter, by its definition, decides the regularization strength, and hence, higher noise usually requires a larger $\alpha$ value (as an extreme example, for an image with zero noise the optimal $\alpha$ is 0). However, what properties of a given image decide the optimal value for the underlying Euclidean norm $p$ is unclear so far. As we can see from Table 2 for the 3 test images (with the exact same level of Gaussian noise) the optimal value $p$ ranges from almost 1 to almost 3. The current guess is that the optimal $p$ is partially decided by the properties of edges of the given image but a detailed theoretical explanation is still missing.

As a final remark of the training scheme $\mathcal{T}$ with $\ell^p$-(an)-isotropic total variation, the introduction of Euclidean order $p \in [1, +\infty]$ into training scheme only meant to expand the training choices, but not to provide a superior seminorm to the popular choice $TV_2$ or $TV_1$. The optimal order $\hat{p} \in \{1, 2\}$ or not, is completely up to the given training image $u_\eta = u_c + \eta$.

Acknowledgments. PL acknowledges support from the EPSRC Centre Nr. EP/N014588/1
and the Leverhulme Trust project on Breaking the non-convexity barrier. CBS acknowledges support from the Leverhulme Trust project on Breaking the non-convexity barrier, the Philip Leverhulme Prize, the EPSRC grant Nr. EP/M00483X/1, the EPSRC Centre Nr. EP/N014588/1, the RISE projects CHiPS and NoMADS, the Cantab Capital Institute for the Mathematics of Information and the Alan Turing Institute. We gratefully acknowledge the support of NVIDIA Corporation with the donation of a Quadro P6000 GPU used for this research.

References

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

[2] P. L. Antonelli, editor. Handbook of Finsler geometry. Vol. 1, 2. Kluwer Academic Publishers, Dordrecht, 2003. With 1 CD-ROM containing the software package FINSLER.

[3] A. Braides. Γ-convergence for beginners, volume 22 of Oxford Lecture Series in Mathematics and its Applications. Oxford University Press, Oxford, 2002.

[4] A. Chambolle, M. J. Ehrhardt, P. Richtárik, and C.-B. Schönlieb. Stochastic primal-dual hybrid gradient algorithm with arbitrary sampling and imaging application. arXiv preprint arXiv:1706.04957, 2017.

[5] 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.

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

[7] Y. Chen, R. Ranftl, and T. Pock. Insights into analysis operator learning: From patch-based sparse models to higher order mrfs. IEEE Transactions on Image Processing, 23(3):1060–1072, March 2014.

[8] G. Dal Maso. An introduction to Γ-convergence, volume 8 of Progress in Nonlinear Differential Equations and their Applications. Birkhäuser Boston, Inc., Boston, MA, 1993.

[9] 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.

[10] J. C. De Los Reyes, C.-B. Schönlieb, and T. Valkonen. The structure of optimal parameters for image restoration problems. J. Math. Anal. Appl., 434(1):464–500, 2016.

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

[12] I. Ekeland and R. Témam. Convex analysis and variational problems, volume 28 of Classics in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, English edition, 1999. Translated from the French.

[13] H. W. Engl, M. Hanke, and A. Neubauer. Regularization of Inverse Problems, volume 375 of Mathematics and Its Applications. Springer Verlag, 2000.

[14] L. C. Evans and R. F. Gariepy. Measure theory and fine properties of functions. Textbooks in Mathematics. CRC Press, Boca Raton, FL, revised edition, 2015.

[15] T. Goldstein and S. Osher. The split Bregman method for L1-regularized problems. SIAM J. Imaging Sci., 2(2):323–343, 2009.

[16] G. H. Golub, M. Heat, and G. Wahba. Generalized cross validation as a method for choosing a good ridge parameter. Technometrics, 21:215–223, 1979.

[17] M. Grasmair and F. Lenzen. Anisotropic total variation filtering. Appl. Math. Optim., 62(3):323–339, 2010.

[18] E. Haber and L. Tenorio. Learning regularization functionals—a supervised training approach. Inverse Problems, 19(3):611–626, 2003.

[19] P. C. Hansen. Analysis of discrete ill-posed problems by means of the L-curve. SIAM Review, 34:561–580, 1992.

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

[21] K. Kunisch and T. Pock. A bilevel optimization approach for parameter learning in variational models. SIAM J. Imaging Sci., 6(2):908–983, 2013.

[22] M. Lustica, S. Moll, and P. B. Mucha. Total variation denoising in l1 anisotropy. SIAM J. Imaging Sci., 10(4):1691–1723, 2017.

[23] P. Liu and C.-B. Schönlieb. Anisotropic total variation and piecewise constant solutions. in preparation, 2019.
[24] Y. Meyer. Oscillating patterns in image processing and nonlinear evolution equations: the fifteenth Dean Jacqueline B. Lewis memorial lectures, volume 22. American Mathematical Soc., 2001.

[25] J. Moll. The anisotropic total variation flow. *Mathematische Annalen*, 332(1):177–218, 2005.

[26] V. A. Morozov. On the solution of functional equations by the method of regularization. *Soviet mathematics – Doklady*, 7:414–417, 1966.

[27] 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).

[28] D. M. Strong, T. F. Chan, et al. Exact solutions to total variation regularization problems. In *UCLA CAM Report*. Citeseer, 1996.

[29] 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.

[30] J. Weickert. *Anisotropic diffusion in image processing*. European Consortium for Mathematics in Industry. B. G. Teubner, Stuttgart, 1998.