THE SPATIALLY DEPENDENT BI-LEVEL LEARNING SCHEME

PAN LIU

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

November 3, 2016

ABSTRACT. An improved version of bilevel learning scheme (BLS) is introduced by utilizing the original BLS in each subdomain, and by searching for the best combination of different subdomains to reach a recovered image which best fits the given training data. Numerical experiments are carried out to illustrate the improved performance of the proposed learning scheme.

CONTENTS

1. INTRODUCTION

The variational formulation of problems in image processing often have an underlying functional

$$\mathcal{I}(u) := \|\mathcal{K}u - \mathscr{I}\|_{L^2(Q)} + \alpha R(u,Q), \tag{1.1}$$

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

Key words and phrases. Calculus of Variations, Computer Vision and Pattern Recognition, Machine Learning.

where $\mathscr{I}$ is a given corrupted image, $Q := (0,1) \times (0,1)$ represents the domain of a square image, $\|\mathcal{K}u - \mathscr{I}\|_{L^2(Q)}$ represents the *fidelity term*, R represents the *regularization term*, and $\mathcal{K}$ is usually a linear operator.

Image denoising is a fundamental task in image processing, as it is always a necessary step prior to higher level image processing such as reconstruction and segmentation. Regarding image denoising, we set $\mathcal{K}$ to be the identity operator, and we write $\mathscr{I} = \mathscr{C} + \eta$ where $\mathscr{C}$ represents a noise-free clean image, and η encodes noise. The key task of image denoising is to remove the noise η from the given corrupted image $\mathscr{I}$ while preserving the clean image $\mathscr{C}$ as much as possible. The *ROF* total variational functional

$$\frac{1}{2}\int_Q |u - \mathscr{I}|^2\,dx + \alpha TV(u,Q), \tag{1.2}$$

introduced in [31], is one of the most popular choices for such task, where the fidelity term in (1.1) is taken to be the L^2-distance, the regularization term the total variation $TV(u,Q)$, and $\alpha \in \mathbb{R}^+$. The parameter $\alpha > 0$ in (1.2) is used to control the balance between the regularization term and the fidelity term, and the choice of the "best" parameter α then becomes an important task. One way to choose the "best" α is using a bilevel learning optimization scheme, which can optimally adapt itself to the given "perfect data", defined in machine learning (see [10, 11, 18, 35]). This learning scheme searches for $\alpha > 0$ such that the recovered image u_α, which is defined to be the minimizer of (1.2) for each $\alpha > 0$ fixed, best fits the given clean image $\mathscr{C}$, measured in terms of the L^2-distance. An example of a bilevel learning scheme $(\mathcal{B})$ equipped with the *ROF* functional is the following:

Level 1.

$$\alpha_m := \operatorname*{arg\,min}_{\alpha>0} \frac{1}{2}\int_Q |u_\alpha - \mathscr{C}|^2\,dx, \tag{1.3}$$

Level 2.

$$u_\alpha := \operatorname*{arg\,min}_{u\in SBV(Q)} \left\{\frac{1}{2}\int_Q |u - \mathscr{I}|^2\,dx + \alpha TV(u,Q)\right\}. \tag{1.4}$$

The scheme $(\mathcal{B})$ has been proved to have at least one solution $\alpha_m \in (0,+\infty]$ provide $TV(\mathscr{I},Q) > TV(\mathscr{C},Q)$ (see [17]), and a small modification rules out the possibility of $\alpha_m = +\infty$.

To simplify our presentation, we define the reconstruction operator $\mathscr{L}\colon \mathbb{R}^+ \times L^2(Q) \times \mathcal{B}(Q)$ by

$$\mathscr{L}(\alpha, v, Q') := \operatorname*{arg\,min}_{u\in SBV(Q')} \left\{\frac{1}{2}\int_{Q'} |u - v|^2\,dx + \alpha TV(u,Q')\right\}. \tag{1.5}$$

where $\mathcal{B}(Q) := \{Q' \subset Q :\ Q' \text{ is open in } Q\}$. We note that $\mathscr{L}(\alpha, v, Q')$ is uniquely defined due to the strictly convexity of TV.

Although in [17] the existence of a minimizer α_m of the error function

$$\mathcal{E}(\alpha) := \frac{1}{2}\int_Q |\mathscr{L}(\alpha,\mathscr{I},Q) - \mathscr{C}|^2\,dx \tag{1.6}$$

has been established, an executable numerical scheme to find such a minimizer is also in need. Moreover, it is well known that the ROF model in (1.2) suffers drawbacks like the staircasing effect, and, unfortunately, scheme $(\mathcal{B})$ inherits that feature. That is, the optimized reconstruction function $\mathscr{L}(\alpha_m, \mathscr{I}, Q)$ also exhibits the staircasing effect.

In this work we propose a new learning scheme aiming at further reducing the value of $\min_{\alpha>0}\mathcal{E}(\alpha)$ and also at mitigating the staircasing effect. Our paper is organized as follows. In Section 2 we introduce a new way to represent the clean image and the noise, which is compatible with a discrete computer image data in domain Q, and hence we may apply our PDEs and functional analysis tools to it. To be precise, we postulate an ideal clean image $\mathscr{C}\in BV(Q)$ can only be captured by a "super" camera which has infinite resolution, and we postulate that a finite $N\in\mathbb{N}$ resolution level image captured by a real world digital camera is a piecewise constant function $\mathscr{C}_N$ which is related to $\mathscr{C}$ via its averages

$$\mathscr{C}_N(x) := \fint_{Q_N(i,j)} \mathscr{C}\,dy \text{ for } x\in Q_N(i,j),$$

where $Q_N(i,j) := ((i-1)/N, i/N)\times((j-1)/N, j/N)$, $1\le i,j\le N$, and

$$\mathcal{Q}_N := \{Q_N(i,j),\ 1\le i,j\le N\}. \tag{1.7}$$

That is, each $Q_N(i,j)$ represents a pixel in discrete computer image data. We build our analysis mostly based on $\mathscr{C}_N$ rather than on $\mathscr{C}$. Furthermore, we define the corrupted image data to be, at resolution level N,

$$\mathscr{I}_N := \mathscr{C}_N + \eta_N,$$

where by η_N we denote a piecewise noise constant function over $\mathcal{Q}_N$. That is, η_N is a constant in each $Q_N\in\mathcal{Q}_N$, $1\le i,j\le N$.

Moreover, we define the *stopping time* $\alpha_s(v)$ of a function $v\in L^\infty(Q)$ via the following definition.

Definition 1.1. *Let $v\in L^\infty(Q)$ be given. We say that $\alpha_s(v)\in[0,+\infty)$ is the* stopping time *for v if*

$$\mathscr{L}(\alpha_s, v, Q) = \mathscr{L}(\alpha_s+\alpha, v, Q) =: C(v) \text{ and } \mathscr{L}(\alpha_s, v, Q) \neq \mathscr{L}(\alpha_s-\alpha, v, Q) \tag{1.8}$$

for all $\alpha>0$, where $C(v)$ is a constant depends on v.

By its definition, if it exists then the stopping time is unique. In Section 3.1 we show that the minimizer of (1.6) exists in the one dimensional setting. To do so, we first show that the stopping time $\alpha_s(\mathscr{I}_N)$ exists where $\alpha_s(\mathscr{I}_N)$ is defined in (1.8) with Q replaced by $I := (0,1)$. Next, in Proposition 3.2, using Theorem 2.2 repeatedly, we show that the *level N error function*

$$\mathcal{E}_N(\alpha) := \frac{1}{2}\int_I \left|\mathscr{L}(\alpha,\mathscr{I}_N, I) - \mathscr{C}_N\right|^2 dx \tag{1.9}$$

is continuous, and there exist finitely many $0<\alpha_1<\alpha_2<\cdots<\alpha_M=\alpha_s(\mathscr{I}_N)<+\infty$ such that in each interval $[\alpha_i,\alpha_{i+1})$, $\mathcal{E}_N(\cdot)$ is convex and $\mathcal{E}_N'(\cdot)$ is linearly increasing. Hence, a direct search, which we detail in Section 3.1.2, of a minimizer α_m of (1.9) inside the finite interval $[0,\alpha_s(\mathscr{I}_N)]$ can be executed numerically and terminated within a finite time, although it may take a long CPU time, in the order of $O(N)$.

In Section 3.2 we discuss the behavior of (1.9) in the two dimensional setting. We present a two dimensional version of Proposition 3.2 in Proposition 3.6. Although the statement of Proposition 3.6 is weaker compared with Proposition 3.2, due to the lack of the two dimensional version of Theorem 2.2, it is still sufficient to allow us to perform the same direct search to locate α_m within a finite time.

Although the direct search of a minimizer α_m is numerically executable, it takes a long CPU time, and hence an efficient numerical scheme to locate α_m is in great need. In Section 3.3 we provide an easy way to accelerate our direct search by determining a lower bound of α_m and hence save some CPU time.

To really make an improvement on reducing the CPU time needed to locate α_m, we observe from numerical simulations that $\mathcal{E}_N(\cdot)$ defined in (1.9) is likely to be *strictly quasi-convex* ([5], Section 3.4), i.e., for any α_1, $\alpha_2 \in \mathbb{R}^+$ and $\lambda \in (0,1)$,

$$\mathcal{E}_N(\lambda\alpha_1 + (1-\lambda)\alpha_2) < \max\{\mathcal{E}_N(\alpha_1), \mathcal{E}_N(\alpha_2)\}. \tag{1.10}$$

If (1.10) holds, then locating α_m could be done much faster and efficiently, compared to the direct search, by using quasi-convex programming method (see, e.g., [24]). However, (1.10) only holds if $\mathscr{I}_N$ satisfies some very restrictive assumptions which are unlikely to be satisfied in concrete setting (see, e.g. [25]). To overcome this drawback, we propose in [25] an alternative searching algorithm which could locate α_m, or at least a good approximation of α_m with a controllable error, efficiently without requiring $\mathcal{E}_N(\cdot)$ to be strictly quasi-convex. However, that algorithm is complex and its theoretical validation would require several auxiliary results beyond the scope of this paper. We refer readers to our upcoming work [25] for this acceleration algorithm, and also the study of the quasi-convexity of the level N error function (1.9).

In Section 4.1 we introduce our new learning scheme $(\mathcal{P})$. We recall that $Q := (0,1)\times(0,1)$, and for $K \in \mathbb{N}$, $Q_K \subset \mathbb{R}^2$ denotes a cube with its faces normal to the orthonormal basis of $\mathbb{R}^2$, and with side-length greater than or equal to $1/K$. $\mathcal{L}_K$ will be a collection of finitely many Q_K such that

$$\mathcal{L}_K := \left\{Q_K \subset Q: \ Q_K \text{ are mutually disjoint}, \ Q \subset \bigcup \overline{Q_K}\right\},$$

and $\mathcal{V}_K$ denotes the collection of all possible $\mathcal{L}_K$. For $K=0$ we set $Q_0 := Q$, hence $\mathcal{L}_0 = \{Q\}$. We define our improved learning scheme $(\mathcal{P})$ in resolution level N as:

Level 1.

$$u_{\mathcal{P},N} := \arg\min\left\{\int_Q \left|\mathscr{C}_N - u_{\mathcal{L}_K}\right|^2 dx, \ K \geq 0, \ \mathcal{L}_K \in \mathcal{V}_K\right\} \tag{1.11}$$

Level 2.

$$u_{\mathcal{L}_K}(x) := \mathscr{L}(\alpha_{Q_K}, \mathscr{I}_N, Q_K) \text{ for } x \in Q_K \text{ and } Q_K \in \mathcal{L}_K,$$

$$\text{where } \alpha_{Q_K} := \operatorname*{arg\,min}_{\alpha>0} \int_{Q_K} \left|\mathscr{L}(\alpha, \mathscr{I}_N, Q_K) - \mathscr{C}_N\right|^2 dx.$$

The new bilevel learning scheme $(\mathcal{P})$ utilizes the scheme $(\mathcal{B})$ in each subdomain of Q, and searches for the best combination of different subdomains from which a recovered image $u_{\mathcal{P},N}$, which best fits $\mathscr{C}_N$, might be obtained, and hence the name "spatially dependent".

In Section 4.2 we provide an example to demonstrate that the proposed scheme $(\mathcal{P})$ can avoid the staircasing effect in certain situations. Moreover, we compare our results with the recent results in [22], in which the authors propose a different way to avoid the staircasing effect.

In Section 4.3 we show that under a mild assumption on the noise η_N, the scheme $(\mathcal{P})$ is able to fully recover the clean image $\mathscr{C}$ as the resolution level N goes to infinite. Let

$$\mathcal{P}_N(K) := \min_{\mathcal{L}_K \in \mathcal{V}_K} \int_Q |\mathscr{C}_N - u_{\mathcal{L}_K}|^2 \, dx \text{ and } \mathcal{P}(N) := \mathcal{P}_N(N) = \int_Q |\mathscr{C}_N - u_{\mathcal{P},N}|^2 \, dx$$

where $u_{\mathcal{P},N}$ is defined in (1.11). In Theorem 4.4 we prove the following result.

Theorem 1.2. *Assume that the noise η_{K^2} has locally average 0, that is,*

$$⨍_{Q_K} \eta_{K^2} = 0 \tag{1.12}$$

for any $Q_K \in \mathcal{Q}_K$ where $\mathcal{Q}_K$ is defined in (1.7). Then

$$\lim_{K\to\infty} \mathcal{P}(K^2) = 0.$$

Next, we propose a simplified version of scheme $(\mathcal{P})$, the scheme $(\mathcal{P}')$, as following (recall $\mathcal{Q}_K$ from (1.7)):

Level 1.

$$u_{\mathcal{P}',N} := \arg\min \left\{ \int_Q |\mathscr{C}_N - u_{\mathcal{Q}_K}|^2 \, dx, \; K \geq 0 \right\}$$

Level 2.

$$u_{\mathcal{Q}_K}(x) := \mathscr{L}(\alpha_{Q_K}, \mathscr{I}_N, Q_K) \text{ for } x \in Q_K \text{ and } Q_K \in \mathcal{Q}_K,$$

$$\text{where } \alpha_{Q_K} := \arg\min_{\alpha>0} \int_{Q_K} |\mathscr{L}(\alpha, \mathscr{I}_N, Q_K) - \mathscr{C}_N|^2 \, dx.$$

Note that $\mathcal{Q}_K \in \mathcal{V}_K$ and hence we have $\mathcal{P}(N) \leq \mathcal{P}'(N)$ where $\mathcal{P}'(N) := \|\mathscr{C}_N - u_{\mathcal{P}',N}\|_{L^2(Q)}$. That is, the optimized reconstructed image $u_{\mathcal{P}',N}$ produced by scheme $(\mathcal{P}')$ might result in a higher error compare with the optimized reconstructed image $u_{\mathcal{P},N}$ produced by scheme $(\mathcal{P})$. However, Theorem 1.2 holds for scheme $(\mathcal{P}')$ too. That is, $\lim_{K\to\infty} \mathcal{P}'(K^2) = 0$ if (1.12) is satisfied, and that implies $\lim_{N\to\infty} \|u_{\mathcal{P}',N} - u_{\mathcal{P},N}\|_{L^2} = 0$. Hence, scheme $(\mathcal{P}')$ could be used as a "risky" replacement of scheme $(\mathcal{P})$ since it has the same result as resolution level N goes to infinity, and more importantly, it requires much less CPU time needed as required by scheme $(\mathcal{P})$.

Lastly, in Section 5 we propose a comprehensive learning scheme $(\mathcal{CT})$ which generalizes scheme $(\mathcal{P})$ by allowing more turning options in Level 2 above. To be precise, we introduce the following scheme:

Level 1.

$$u_{\mathcal{CT},N} := \arg\min \{\mathcal{F}(\mathscr{C}_N - u, Q) : \; K \geq 0, \; u \in \mathcal{T}_K\}$$

Level 2.

$$\mathcal{T}_K := \{u_{\mathcal{CT}_K,N}, \; \mathcal{L}_K \in \mathcal{V}_K\}$$

$$u_{\mathcal{CT}_K,N}(x) \text{ is constructed upon the information obtained by } u_{Q_K,N} \tag{1.13}$$

Level 3.

$$u_{Q_K,N} := \arg\min \{\mathcal{F}(\mathscr{I}_N - u, Q_K) + R(u, \alpha, Q_K), \; u \in SBV(Q), \; R \in \mathcal{R}\}, \tag{1.14}$$

where $\mathcal{F}$ is the fidelity term which is quasiconvex in the sense of [12], and $\mathcal{R}$ is a collection of regularizers. Note that by letting $R(u, \alpha, Q_K) = \alpha TV(u, Q_K)$ in (1.14) and $u_{\mathcal{CT}_K,N}(x) := u_{Q_K,N}$ for $x \in Q_K$ in (1.13), scheme $(\mathcal{CT})$ reduces to scheme $(\mathcal{P})$. In the end, we provide examples of new regularizers which can be inserted into (1.14) and we also comment on possible options regarding to construction of $u_{Q_K,N}$ is (1.13).

2. The finite resolution image and some preliminary results

2.1. **The finite resolution clean image and the unavoidable noise during acquisition.** As we stated in the introduction, in the two dimensional setting we represent a finite $N \times N$ resolution clean image $\mathscr{C}_N$ of an infinite resolution ideal clean image $\mathscr{C}$, which is assumed to be represented by a BV function, via its average

$$\mathscr{C}_N(x) := \fint_{Q_N(i,j)} \mathscr{C}\, dx \text{ for } x \in Q_N(i,j),$$

where $Q_N(i,j) := ((i-1)/N, i/N) \times ((j-1/N), j/N)$, for $1 \le i, j \le N$, and we define the collection

$$\mathcal{Q}_N := \{Q_N(i,j),\ 1 \le i, j \le N\}.$$

Similarly, in one dimension, we define

$$\mathscr{C}_N(x) := \fint_{I_N(k)} \mathscr{C}\, dx \text{ for } x \in I_N(k), \tag{2.1}$$

where $I_N(k) := ((k-1)/N, k/N)$, for $1 \le k \le N$, and introduce the collection

$$\mathcal{I}_N := \{I_N(k),\ 1 \le k \le N\}.$$

The principal sources of noise in digital images are introduced during acquisition, for example, the sensor noise caused by poor illumination, high temperature, and circuity of a scanner. Other possible sources could be digital error during the transmission, and the unavoidable shot noise of an photon detector. The noise is only generated during the acquiring of image, i.e., it is only added to $\mathscr{C}_N$, and each time we acquire an image, we would produce a different noise η_N. Therefore, we propose to use a piecewise constant function η_N over $\mathcal{Q}_N$ to represent the noise on the resolution level $N \in \mathbb{N}$, and we write

$$\mathscr{I}_N := \mathscr{C}_N + \eta_N.$$

That is, when a image is taken with resolution $N \in \mathbb{N}$, although we only wish to observe $\mathscr{C}_N$, the noise η_N is an unavoidable by-product, and hence the corrupted image $\mathscr{I}_N$ is produced.

Since $\mathscr{I}_N$ represents an image data, we may assume that

$$\|\mathscr{I}_N\|_{L^\infty} \le 1.$$

When $N \to \infty$, $\mathscr{C}_N \to \mathscr{C}$ in L^2, but since η_N is randomly generated, so although for a fixed N, $\mathscr{C}_N$ is fixed, η_N would vary. As it often assumed in image reconstruction papers (see, e.g., [9]), we also assume that

$$\int_Q \eta_N\, dx = 0.$$

2.2. **The total variation and some preliminary results.** We start by introducing notations that will be used in the sequel.

Notation 2.1. Recall that $I := (0,1) \subset \mathbb{R}$ and $Q := (0,1) \times (0,1) \subset \mathbb{R}^2$. Here $M \in \mathbb{N}$ is a positive integer.

1. we say a function w is a *piecewise constant function with M pieces* if there exist M intervals $I_M(j) := (x_j, x_{j+1})$, where $0 = x_1 < \cdots < x_j < \cdots < x_M = 1$, such that w is a constant in each $I_M(j)$. Moreover, we use $w(I_M(k))$ to denote the value of $w(x)$ for $x \in I_M(j)$, $1 \leq j \leq M$;
2. given a piecewise constant function ω with M pieces, we say that $I_M(j)$, $1 < j < M$, is a *step region* of w if
$$w(I_M(j-1)) \leq w(I_M(j)) \leq w(I_M(j+1)) \text{ or } w(I_M(j-1)) \geq w(I_M(j)) \geq w(I_M(j+1));$$
and $(I_M(j))$ is a *high extreme region* of ω if
$$w(I_M(j)) > \max\{w(I_M(j-1)), w(I_M(j+1))\}$$
and $(I_M(j))$ is a *low extreme region* of ω if
$$w(I_M(j)) < \min\{w(I_M(j-1)), w(I_M(j+1))\}.$$
3. we say $I_M(1)$ is a *high (low) boundary regions* of ω if $w(I_M(1)) > (<) w(I_M(2))$, and $I_M(M)$ is a *high (low) boundary regions* of ω if $w(I_M(M)) > (<) w(I_M(M-1))$, respectively.

We recall that in [34] it is shown that, in one dimension, if u_0 is a piecewise constant function, then the solution $u_\alpha := \mathscr{L}(\alpha, u_0, I)$ defined in (1.5) is piecewise constant also.

Theorem 2.2 ([34], Theorem 2)**.** *Suppose that the function u_0 is piecewise constant with M pieces, and let α be small enough. Then the unique solution $u_\alpha := \mathcal{L}(\alpha, u_0, I)$ is also piecewise constant with the same piece of u_0, and we have*
$$\begin{aligned}
&u_\alpha(I_M(j)) = u_0(I_M(j)) \mp \frac{2}{|I_M(j)|}\alpha, \ \text{ if } I_M(j) \text{ is a high (low) extremum region,}\\
&u_\alpha(I_M(j)) = u_0(I_M(j)), \ \text{ if } I_M(j) \text{ is a step region,}\\
&u_\alpha(I_M(j)) = u_0(I_M(j)) \mp \frac{1}{|I_M(j)|}\alpha, \ \text{ if } I_M(j) \text{ is a high (low) boundary region.}
\end{aligned}$$
Moreover, for α is large enough, the function u_α is a constant.

Theorem 2.3 ([32], Theorem 10.10)**.** *Let $v \in L^\infty(I)$ and $\alpha_1, \alpha_2 \in \mathbb{R}^+$ be given. Then the semigroup property*
$$\mathscr{L}(\alpha_1 + \alpha_2, v, I) = \mathscr{L}(\alpha_2, \mathscr{L}(\alpha_1, v, I), I) = \mathscr{L}(\alpha_1, \mathscr{L}(\alpha_2, v, I), I) \tag{2.2}$$
holds for the one dimensional scalar total variation problem.

Theorem 2.4 ([8], Theorem 3.4)**.** *Let $v \in BV(Q)$ be given. Then*
$$J_{\mathscr{L}(\alpha, v, Q)} \subset J_v$$
for any $\alpha > 0$, where J_v denotes the jump set of v. Moreover, the same result holds if we replace Q by I.

Remark 2.5. It follows from Theorem 2.3 and Theorem 2.4 that
$$J_{\mathscr{L}(\alpha_2, v, I)} \subset J_{\mathscr{L}(\alpha_1, v, I)}$$

for any $\alpha_1 \leq \alpha_2$. Indeed, by Theorem 2.3

$$\mathscr{L}(\alpha_2, v, I) = \mathscr{L}(\alpha_2 - \alpha_1 + \alpha_1, v, I) = \mathscr{L}(\alpha_2 - \alpha_1, \mathscr{L}(\alpha_1, v, I), I),$$

and hence by Theorem 2.4 with $\alpha := \alpha_2 - \alpha_1$, we obtain the result.

Proposition 2.6 ([3], Theorem 3)**.** *Let $v \in L^2(Q)$ be given. Then $\mathscr{L}(\cdot, v, Q) \in C([0, +\infty); L^2(Q))$. The same result holds for one dimension case, i.e., $\mathscr{L}(\cdot, v, I) \in C([0, +\infty); L^2(I))$.*

3. A direct search for a minimizer α_m of error function

3.1. **The one dimensional case.** In Section 3.1 we will abbreviate $\mathscr{L}(\alpha, v, I)$ as $\mathscr{L}(\alpha, v)$ and $TV(v, I)$ as $TV(v)$.

3.1.1. *Some properties of piecewise constant functions.*

Lemma 3.1. *Let w be a piecewise constant function with M pieces where $M > 1$ large is a positive integer, then there exists a positive integer $M' \leq M$ and*

$$0 = \alpha_0 < \alpha_1 < \alpha_2 < \cdots < \alpha_{M'} < +\infty \tag{3.1}$$

such that

1. *$\mathscr{L}(\alpha_i, w)$ has at least one more constant piece than $\mathscr{L}(\alpha_{i+1}, w)$ for $i = 0, 1, \ldots, M' - 1$;*
2. *$\mathscr{L}(\alpha_i + \alpha, w)$ has the same number of constant pieces of $\mathscr{L}(\alpha_i, w)$, for any $0 \leq \alpha < \alpha_{i+1} - \alpha_i$ where $0 \leq i \leq M' - 1$;*
3. *$\mathscr{L}(\alpha, w) =: C(v)$ for all $\alpha \geq \alpha_{M'}$, where $C(v)$ is a constant depends on v.*

Moreover, the function t: $[0, +\infty) \to [0, +\infty)$ defined as

$$t(\alpha) := \|\mathscr{L}(\alpha, w)\|^2_{L^2(I)}$$

is continuous, and in each interval $[\alpha_j, \alpha_{j+1})$, t' is linearly increasing and t is convex.

Proof. According to Theorem 2.2, for each $1 < j < M$ and $\alpha > 0$ small enough, we have

$$(\mathscr{L}(\alpha, w))(I_M(j)) = w(I_M(j)) \mp \frac{2}{|I_M(j)|}\alpha, \text{ if } I_M(j) \text{ is a high (low) extremum region of } w,$$
$$(\mathscr{L}(\alpha, w))(I_M(j)) = w(I_M(j)) \mp \frac{1}{|I_M(j)|}\alpha, \text{ if } I_M(j) \text{ is a high (low) boundary region of } w.$$

Therefore, we have

$$\|\mathscr{L}(\alpha, w)\|^2_{L^2(I_M(j))} = \left\| w(I_M(j)) \mp \frac{2}{|I_M(j)|}\alpha \right\|^2_{L^2(I_M(j))}$$

provided that $I_M(j)$ is a high (low) extremum region of ω. We obtain

$$\frac{1}{2}\frac{d}{d\alpha}\|\mathscr{L}(\alpha, w)\|^2_{L^2(I_M(j))} = 2\left(\frac{2}{|I_M(j)|}\alpha \mp w(I_M(j))\right),$$

which is continuous and linearly increasing in α, and

$$\frac{1}{2}\frac{d^2}{d\alpha^2}\left(\|\mathscr{L}(\alpha, w)\|^2_{L^2(I_M(j))}\right) = \frac{4}{|I_M(j)|}$$

which is strictly positive. A similar result holds if $I_M(j)$ is a boundary region. Moreover, since

$$t(\alpha) := \|\mathscr{L}(\alpha,w)\|^2_{L^2(I)} = \sum_{j=1}^{M} \|\mathscr{L}(\alpha,w)\|^2_{L^2(I_M(j))},$$

which is a finite summation of $\|\mathscr{L}(\alpha,w)\|^2_{L^2(I_M(j))}$, we conclude that $t'(\alpha)$ is continuous increasing and $t''(\alpha) > 0$ for $\alpha > 0$ small.

We claim that there exists an unique $\alpha_1 > 0$ such that for all $\alpha \in (0, \alpha_1)$

$$\mathscr{L}(\alpha_1 - \alpha, w) \text{ has } M \text{ pieces, but } \mathscr{L}(\beta, w) \text{ have at most } M-1 \text{ pieces,} \tag{3.2}$$

for all $\beta \geq \alpha_1$.

We first show the uniqueness. Assume there exist distinct α_1 and $\alpha'_1 > 0$ such that (3.2) holds for both α_1 and α'_1. Without lose of generality we assume that $\alpha_1 < \alpha'_1$. Let $\alpha''_1 > 0$ be such that $\alpha_1 < \alpha''_1 < \alpha'_1$. Then, on the one hand, by (3.2) and Remark 2.5 we have

$$\mathscr{L}(\alpha''_1, w) = \mathscr{L}(\alpha_1 + (\alpha''_1 - \alpha_1), w) = \mathscr{L}(\alpha''_1 - \alpha_1, \mathscr{L}(\alpha_1, w))$$

has at most $M-1$ pieces, on the anther hand we have, again by (3.2), that $\mathscr{L}(\alpha'_1 - (\alpha'_1 - \alpha''_1), w)$ has M pieces since $\alpha'_1 - \alpha''_1 > 0$, and we have a contradiction.

We define the set

$$\mathcal{A} := \{\alpha' > 0,\ \mathscr{L}(\alpha', w) \text{ has at most } M-1 \text{ pieces}\}$$

and we claim that

$$\beta := \inf_{\alpha>0} \{\alpha \in \mathcal{A}\} \tag{3.3}$$

has the properties required by (3.2). First, we have that $\beta < +\infty$ since by Theorem 2.2 there exists $\alpha' > 0$ large enough such that $\mathscr{L}(\alpha, w)$ is a constant, i.e., it has only one constant piece, and hence $\mathcal{A} \neq \varnothing$. Next, let $\{\alpha_n\}_{n=1}^{\infty} \subset \mathcal{A}$ be such that $\alpha_n \searrow \beta$. We have $\mathscr{L}(\beta, w) = \lim_{n\to\infty} \mathscr{L}(\alpha_n, w)$ by Proposition, 2.6 and hence $\mathscr{L}(\beta, w)$ has at most $M-1$ pieces. Finally, we claim that $\mathscr{L}(\beta - \alpha, w)$ has M constant pieces for any $\alpha > 0$. If not, then there would be $\alpha'' > 0$ such that $\mathscr{L}(\beta - \alpha'', w)$ has at most $M-1$ constant pieces, but this contradicts (3.3).

We have shown that the function t has the required properties for $0 \leq \alpha < \alpha_1$ where α_1 is obtain via (3.3) ($\alpha_1 := \beta$), and α_1 satisfies items 1 and 2 in Lemma 3.1. Next, by (2.2) we may write, for $\alpha \geq \alpha_1$, that

$$\mathscr{L}(\alpha, w) = \mathscr{L}(\alpha_1 + \alpha - \alpha_1, w) = \mathscr{L}(\alpha - \alpha_1, w_1)$$

where $w_1 := \mathscr{L}(\alpha_1, w)$ is a piecewise constant function with M_1 pieces and $M_1 \leq M-1$. We can repeat the above argument to obtain α'_2 such that $w_2 := \mathscr{L}(\alpha'_2, \omega_1)$ has at most M_2 constant pieces where $M_2 \leq M_1 - 1$, and we define $\alpha_2 := \alpha_1 + \alpha'_2$. A recursive argument will lead to $w_{M'}$ a constant for M' sufficiently large. Since w only has M pieces, $M' \in \mathbb{N}$ is finite and $\alpha_{M'} < +\infty$ and so we obtain (3.1) as desired. Finally, since $w_{M'} := \mathscr{L}(\alpha_{M'}, w)$ has only one piece, $w_{M'}(x) =: C$ for all $x \in I$ and C is a constant. We conclude that for all $\alpha > \alpha_{M'}$

$$\mathscr{L}(\alpha, w) = \mathscr{L}(\alpha - \alpha_{M'}, w_{M'}) = w_{M'}.$$

□

Proposition 3.2. *For any given corrupted image $\mathscr{I}_N$ and clean image $\mathscr{C}_N$, there exists an integer $N' < N$ and*

$$0 = \alpha_0 < \alpha_1 < \alpha_2 < \cdots < \alpha_{N'} = \alpha_s(\mathscr{I}_N) < +\infty \tag{3.4}$$

such that item 1, 2, and 3 of Lemma 3.1 holds. Moreover, in each interval (α_i, α_{i+1}) $\mathcal{E}_N(\cdot)$ is convex and $\mathcal{E}'_N(\cdot)$ is linearly increasing, where $\mathcal{E}_N$ is defined in (1.9)*.*

Proof. Since $\mathscr{C}_N$ is a fixed piecewise constant function, we may apply Lemma 3.1 to $\mathscr{I}_N$ to obtain (3.4), and that $\mathcal{E}_N(\cdot)$ is convex and $\mathcal{E}'_N(\cdot)$ linearly increasing within each interval (α_i, α_{i+1}). Moreover, we conclude that $\alpha_{N'} = \alpha_s(\mathscr{I}_N)$ by applying items 1, 2, and 3 in Lemma 3.1 with $i = N'$. □

3.1.2. *The direct search for a minimizer α_m of level N error function.* Proposition 3.2 allows us to perform a direct search to find a minimizer α_m of (1.9). Indeed, recall that in each interval $[\alpha_i, \alpha_{i+1})$, $\mathcal{E}_N(\cdot)$ is convex and $\mathcal{E}'_N(\cdot)$ is linearly increasing. Hence, we may apply *Newton descent* (see, e.g., [2]) algorithm to locate the unique local minimizer $\alpha_{i,m}$ for $\mathcal{E}_N(\alpha)$ in $[\alpha_i, \alpha_{i+1}]$, and repeat over all intervals provided by (3.4). Since there are only finitely many intervals $[\alpha_i, \alpha_{i+1})$, we can locate all possible local minimizers $\alpha_{i,m}$ within a finite time. Finally, the finite stopping time $\alpha_s(\mathscr{I}_N)$ provides a natural stopping criterion for our searching algorithm. That is, we terminate our searching progress once we reach the point when $\mathcal{E}_N(\cdot)$ is a constant. After we terminate our searching progress, we only need to find the smallest local minimizer $\alpha_{i,m}$ and that is our α_m as desired. Lastly, if there is a tie, i.e., two local minimizer $\alpha_{i,m} < \alpha_{i',m}$ such that both gave the smallest value of $\mathcal{E}_N(\cdot)$, we choose $\alpha_{i,m}$ as our minimizer α_m and ignore $\alpha_{i'm}$.

3.2. **The two dimensional case.** In this section we present a two dimensional (weaker) version of Lemma 3.1 and Proposition 3.2 in Proposition 3.6. In particular, items 1 and 2 in Lemma 3.1 will be absent due to the lack of a two dimensional version of Theorem 2.2. We remark that so far we only have a weaker version of Theorem 2.2 in two dimensions and we refer readers to our follow up work [25].

We start by recalling the following theorem in [7].

Theorem 3.3 ([7], Theorem 4 and 5)**.** *Let $v \in L^\infty(Q)$ be given and let ∂TV denote the subgradient of the TV seminorm. Considering the* gradient flow *defined as*

$$\begin{cases} -\partial_t \mathcal{G}(t,v) \in \partial TV(\mathcal{G}(t,v)), \\ \mathcal{G}(0,v) := v. \end{cases} \tag{3.5}$$

Then following hold:

1. *the solution $\mathcal{G}(t,v)$ is uniquely defined;*
2. *the solution $\mathcal{G}(t,v)$ satisfies $\mathcal{G}(t,v) = \mathscr{L}(\alpha, v)$ for $t = \alpha$;*
3. *there exist finitely many $0 = t_0 < t_1 < t_2 < \cdots < t_K \le \infty$ such that the solution of* (3.5) *is given by*

$$\mathcal{G}(t,v) = \mathcal{G}(t_i,v) - (t - t_i) SG(t_{i+1})$$

for $t \in [t_i, t_{i+1})$, where $SG(t_{i+1}) \in \partial TV(\mathcal{G}(t_i, v))$.

We now prove the following two dimensional "semi-group" property.

Proposition 3.4. *Let $v \in L^\infty(Q)$ and let $0 < \alpha_1 < \alpha_2 < +\infty$ be given. Then*

$$\mathscr{L}(\alpha_2, v) = \mathscr{L}(\alpha_2 - \alpha_1, \mathscr{L}(\alpha_1, v)). \tag{3.6}$$

Proof. Let $v_1 := \mathscr{L}(\alpha_1, v)$, and define a new gradient flow by

$$-\partial_t \mathcal{G}^1(t, v_1) \in \partial TV(\mathcal{G}^1(t, v_1)),\ \mathcal{G}^1(0, v_1) := v_1,$$

and we have $\mathcal{G}^1(t, v_1)$ is uniquely defined. By Theorem 3.3 we have that

$$\mathcal{G}^1(\alpha_2 - \alpha_1, v_1) = \mathscr{L}(\alpha_2 - \alpha_1, \mathscr{L}(\alpha_1, v)),$$

and

$$\mathcal{G}(\alpha_2, v) = \mathscr{L}(\alpha_2, v).$$

Moreover, by the property of gradient flow, we have

$$\mathcal{G}(\alpha_2, v) = \mathcal{G}^1(\alpha_2 - \alpha_1, v_1),$$

and hence (3.6) hold. □

We recall that the stopping time α_s was defined in Definition 1.1.

Lemma 3.5. *Let $v \in L^\infty(Q)$ be given. Then $\alpha_s(v) < +\infty$ and $\mathscr{L}(\alpha_s(v), v)$ is a constant.*

Proof. We note that the null space of total variation seminorm

$$\mathcal{N}(TV) = \left\{v \in L^1(Q),\ TV(v) = 0\right\}, \tag{3.7}$$

is the space of constant function (see, e.g., [1]), and hence a linear subspace of $L^1(Q)$.

By Proposition 2.1 in [9], the optimality condition of (1.5), with v in place of $\mathscr{I}$, is

$$\frac{1}{\alpha}\left(\mathscr{L}(\alpha, v) - v\right) \in \partial TV(\mathscr{L}(\alpha, v)).$$

Let P_{TV} denote the projection operator onto $\mathcal{N}(TV)$. Hence $P_{TV}(v)$ is a constant by (3.7). We claim that

$$\frac{1}{\alpha}\left(v - P_{TV}(v)\right) \in \partial TV(0) \tag{3.8}$$

for $\alpha > 0$ large enough. Indeed, since $\partial TV(0)$ has nonempty relative interior in $\mathcal{N}(TV)$ (see, e.g., [28]), we have that (3.8) holds for $\alpha > 0$ sufficient large since $v \in L^\infty(Q)$ and $P_{TV}(v)$ is a constant. Let $\alpha_S > 0$ be large enough such that (3.8) hold. Then we have

$$\frac{1}{\alpha_S}\left(v - P_{TV}(v)\right) \in \partial TV(0) = \partial TV(P_{TV}(v))$$

where in the last inequality we used again the fact that $P_{TV}(v)$ is a constant. That is, we have

$$\frac{1}{\alpha_S}\left(v - P_{TV}(v)\right) \in \partial TV(P_{TV}(v)),$$

and hence $P_{TV}(v)$ is a solution of (1.5). Since the minimizer of (1.5) is unique, we conclude that

$$P_{TV}(v) = \mathscr{L}(\alpha_S, v) \tag{3.9}$$

and thus $\mathscr{L}(\alpha_S, v)$ is a constant.

Define

$$\alpha_s := \inf\left\{\alpha > 0,\ \mathscr{L}(\alpha, v) = P_{TV}(v)\right\}.$$

Let $\{\alpha_n\}_{n=1}^{\infty} \subset \{\alpha > 0,\ \mathscr{L}(\alpha, v) = P_{TV}(v)\}$ and $\alpha_n \searrow \alpha_s$. We claim that α_s is indeed the stopping time of v. First, α_s is unique by its definition, and α_s is finite since there exists at least one $\alpha_S < +\infty$ such that (3.9) hold. Next, by Proposition 2.6 we have

$$\mathscr{L}(\alpha_s, v) = \lim_{n\to\infty} \mathscr{L}(\alpha_n, v) = P_{TV}(v).$$

Therefore, for all $\alpha > 0$, we have

$$\mathscr{L}(\alpha_s + \alpha, v) = \mathscr{L}(\alpha, \mathscr{L}(\alpha_s, v)) = \mathscr{L}(\alpha, P_{TV}(v)) = P_{TV}(v),$$

where in the first equality we used Proposition 3.4. This concludes the proof. □

Proposition 3.6. *For any given corrupted image $\mathscr{I}_N$ and clean image $\mathscr{C}_N$, there exists an integer $N' \in \mathbb{N}$ and*

$$0 = \alpha_0 < \alpha_1 < \alpha_2 < \cdots < \alpha_{N'} = \alpha_s(\mathscr{I}_N) < +\infty$$

such that, in each interval (α_i, α_{i+1}), $\mathcal{E}_N(\cdot)$ is convex and $\mathcal{E}_N'(\cdot)$ is linearly increasing, where

$$\mathcal{E}_N(\alpha) := \frac{1}{2}\int_Q |\mathscr{L}(\alpha, \mathscr{I}_N) - \mathscr{C}_N|^2\, dx. \tag{3.10}$$

Proof. Applying Theorem 3.3 to $\mathscr{I}_N$, we obtain finitely many

$$0 := \alpha_0 < \alpha_1 < \alpha_2 < \cdots < \alpha_{N'} \le +\infty$$

such that

$$\mathscr{L}(\alpha, \mathscr{I}_N) = \mathscr{L}(\alpha_i, \mathscr{I}_N) - (\alpha - \alpha_i) SG(\alpha_{i+1}) \tag{3.11}$$

for $\alpha \in (\alpha_i, \alpha_{i+1})$, where $SG(\alpha_{i+1}) \in \partial TV(\mathscr{L}(\alpha_i, v))$. By Lemma 3.5 we have $\mathscr{L}(\alpha_s(\mathscr{I}_N), \mathscr{I}_N)$ is a constant and hence $SG(\alpha_s(\mathscr{I}_N)) = 0$. Therefore, invoking (3.11) we deduce that $\alpha_{N'} \le \alpha_s(\mathscr{I}_N) < +\infty$ and $\mathscr{L}(\alpha, \mathscr{I}_N) = \mathscr{L}(\alpha_s(\mathscr{I}_N), \mathscr{I}_N)$ for all $\alpha \ge \alpha_s(\mathscr{I}_N)$. Moreover, by (3.11) and the fact that $\mathscr{C}_N$ is a fixed function, we conclude that in each interval (α_i, α_{i+1}), $\mathcal{E}_N(\cdot)$ is convex and $\mathcal{E}_N'(\cdot)$ is linearly increasing, as desired. □

3.3. **An easy acceleration of the direct search scheme.** In this section we provide a lower bound, which can be located efficiently, of α_m. Thus, we may only need to search for α_m above such lower bound and hence reduce the CPU time needed to locate α_m.

In [9], it is shown that the function $TV(\mathscr{L}(\cdot, \mathscr{I}_N))$ is decreasing to 0 and hence the function

$$\mathcal{T}_N(\alpha) := |TV(\mathscr{L}(\alpha, \mathscr{I}_N)) - TV(\mathscr{C}_N)|$$

is quasi-convex. Thus, solving $\alpha_p := \min_{\alpha>0} \mathcal{T}_N(\alpha)$ can be efficiently done by using quasi-convex programming.

We next claim that, if we know apriori that $TV(\mathscr{I}_N) > TV(\mathscr{C}_N)$, then we have $TV(\mathscr{L}(\alpha_p, \mathscr{I}_N)) = TV(\mathscr{C}_N)$ and

$$\alpha_p \le \alpha_m.$$

Indeed, if $\alpha_m < \alpha_p$, then, since $TV(\mathscr{L}(\cdot, \mathscr{I}_N))$ is strictly decreasing, we have that

$$TV(\mathscr{L}(\alpha_m, \mathscr{I}_N)) > TV(\mathscr{L}(\alpha_p, \mathscr{I}_N)) = TV(\mathscr{C}_N). \tag{3.12}$$

where at the last equality we used the assumption that $TV(\mathscr{I}_N) > TV(\mathscr{C}_N)$ and $TV(\mathscr{L}(\cdot, \mathscr{I}_N))$ is strictly decreasing to 0. Therefore, by letting ${\mathscr{I}_N}' := \mathscr{L}(\alpha_m, \mathscr{I}_N)$, and using (3.12) and the result from scheme $(\mathcal{B})$, there exists a $\alpha' > 0$ such that

$$\left\| \mathscr{L}(\alpha', {\mathscr{I}_N}') - \mathscr{C}_N \right\|_{L^2} < \left\| {\mathscr{I}_N}' - \mathscr{C}_N \right\|_{L^2},$$

which is equivalent to

$$\|\mathscr{L}(\alpha', \mathscr{L}(\alpha_m, \mathscr{I}_N)) - \mathscr{C}_N\|_{L^2} < \|\mathscr{L}(\alpha_m, \mathscr{I}_N) - \mathscr{C}_N\|_{L^2}.$$

Invoking Theorem 3.4, we have

$$\|\mathscr{L}(\alpha_m + \alpha', \mathscr{I}_N) - \mathscr{C}_N\|_{L^2} = \|\mathscr{L}(\alpha', \mathscr{L}(\alpha_m, \mathscr{I}_N)) - \mathscr{C}_N\|_{L^2} < \|\mathscr{L}(\alpha_m, \mathscr{I}_N) - \mathscr{C}_N\|_{L^2},$$

and so

$$\mathcal{E}_N(\alpha_m + \alpha') < \mathcal{E}_N(\alpha_m),$$

contradicting the definition of α_m. Thus, we only need to search α_m for $\alpha \in [\alpha_p, \alpha_s(\mathscr{I}_N)]$ and hence save some CPU time. Moreover, in the follow-up work [25] we shall show that α_p not only is a lower bound of α_m, but we can also obtain the estimate α_m by $\alpha_m \leq \alpha_p + O(1/N)$, where N is the resolution level of given image, and $\mathcal{E}_N(\alpha_p) \leq \mathcal{E}_N(\alpha_m) + O(1/N)$. The proofs of these statements require several auxiliary results which are beyond the scope of this paper, and we refer readers to the follow-up work ([25]).

In the end, we invoke the *stochastic optimization methods* to reduce further the CPU time needed to locate the minimizer of $\mathcal{E}_N(\alpha)$. Roughly speaking, the stochastic optimization methods leverage random search direction to efficiently explore the landscape of $\mathcal{E}_N(\alpha)$. The injected randomness may enable the scheme to escape a local optimum and eventually to approach a global optimum ([33]). Moreover, a good initial guess could efficiently decrease the search steps needed in SDE scheme, and we may use α_p as the initial guess since it is shown to be close to α_m.

4. The spatially dependent bilevel learning scheme with respect to TV

One significant drawback of TV denoising is the staircasing effect, and many attempts have been made to avoid such effect by, for example, introducing a higher level of derivative [13, 6], or by introducing a spatially dependent denoising parameter $\alpha(x)$ (see, e.g., [22]). In this section we present a new learning scheme which is adapted from the bilevel learning scheme $(\mathcal{B})$ (see (1.3), (1.4)).

Before we introduce our new learning scheme, we prove a useful lemma.

Lemma 4.1. *Let $v \in L^\infty(Q)$ be given. Then*

$$\mathscr{L}(\alpha, v) =: u_\alpha \to (v)_Q := \fint_Q v\,dx \text{ a.e.}.$$

Proof. Recalling the definition of $\mathscr{L}(\alpha, v)$ from (1.5) and using $(v)_Q$ as test function, we have

$$\int_Q |u_\alpha - v|^2\,dx + \alpha TV(u_\alpha) \leq \int_Q |(v)_Q - v_0|^2\,dx < +\infty.$$

Hence, $\{u_\alpha\}_{\alpha>0}$ is bounded in L^2, and (up to a not relabeled subsequence) there exists a $u_\infty \in L^2$ such that $u_\alpha \rightharpoonup u_\infty$ in L^2 as $\alpha \to \infty$. In turn, $TV(u_\alpha)$ is bounded, i.e., $\{u_\alpha\}_{\alpha>0}$ is bounded in BV. Hence $u_\infty \in BV$ and

$$TV(u_\infty) \leq \liminf_{\alpha\to\infty} TV(u_\alpha) \leq \lim_{\alpha\to\infty} \frac{1}{\alpha}\int_Q |(v)_Q - v_0|^2\,dx = 0,$$

which implies that $u_\infty =: c$ is a constant. Invoking the compactness embedding in BV space, we have $u_\alpha \to c$ in L^1, and we have (up to a not relabeled subsequence) $u_\alpha \to c$ a.e.. Moreover, by Fatou's Lemma,

$$\int_Q |v-c|^2\,dx \le \int_Q \left|v-(v)_Q\right|^2 dx. \tag{4.1}$$

Note that

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

and hence the left hand side of (4.1) reaches the minimum value at $\lambda = (v)_Q$. We conclude that $c = (v)_Q$, and the proof is completed. □

Remark 4.2. Combining the results from Lemma 4.1 and Lemma 3.5, we deduce that for $\alpha > \alpha_s(v)$, $\mathscr{L}(\alpha, v) = (v)_Q$, which is in agreement with Theorem 2.2.

4.1. **A spatially dependent construction.** Let $N \in \mathbb{N}$, $\mathscr{C}_N$, and η_N be given. For $K \in \mathbb{N}$, $Q_K \subset \mathbb{R}^2$ denotes a cube with its faces normal to the orthonormal basis of $\mathbb{R}^2$, and with side-length greater than or equal to $1/K$. $\mathcal{L}_K$ will be a collection of finitely many Q_K such that

$$\mathcal{L}_K := \left\{Q_K \subset Q:\ Q_K \text{ are mutually disjoint},\ Q \subset \bigcup \overline{Q_K}\right\}, \tag{4.2}$$

and $\mathcal{V}_K$ denotes the collection of all possible $\mathcal{L}_K$. For $K = 0$ we set $Q_0 := Q$, hence $\mathcal{L}_0 = \{Q\}$. We define our improved learning scheme $(\mathcal{P})$ in resolution level N as:

Level 1.

$$u_{\mathcal{P},N} := \arg\min\left\{\int_Q |\mathscr{C}_N - u_{\mathcal{L}_K}|^2\,dx,\ K \ge 0,\, \mathcal{L}_K \in \mathcal{V}_K\right\} \tag{4.3}$$

Level 2.

$$u_{\mathcal{L}_K}(x) := \mathscr{L}(\alpha_{Q_K}, \mathscr{I}_N, Q_K) \text{ for } x \in Q_K \text{ and } Q_K \in \mathcal{L}_K, \tag{4.4}$$

$$\text{where }\ \alpha_{Q_K} := \operatorname*{arg\,min}_{\alpha > 0} \int_{Q_K} |\mathscr{L}(\alpha, \mathscr{I}_N, Q_K) - \mathscr{C}_N|^2\,dx. \tag{4.5}$$

The learning scheme $(\mathcal{P})$ performs the learning scheme $(\mathcal{B})$ in each subdomain and combines it all together to achieve an improved global result. Let

$$\mathcal{P}_N(K) := \inf_{\mathcal{L}_K \in \mathcal{V}_K}\left\{\int_Q |\mathscr{C}_N - u_{\mathcal{L}_K}|^2\,dx\right\}$$

where $u_{\mathcal{L}_K}$ is defined in (4.4), and

$$\mathcal{P}(N) := \int_Q |\mathscr{C}_N - u_{\mathcal{P},N}|^2\,dx \tag{4.6}$$

where $u_{\mathcal{P},N}$ is obtained from (4.3). Since $\mathcal{V}_K \subset \mathcal{V}_{K+1}$, we have $\mathcal{P}_N(K) \ge \mathcal{P}_N(K+1)$ and hence

$$\lim_{K\to\infty} \mathcal{P}_N(K) \text{ exists}$$

and is equal to $\inf_{K\in\mathbb{N}_0} \mathcal{P}_N(K)$. Note that when $K = 0$, $\mathcal{P}_N(0) = \mathcal{E}_N(\alpha_m)$ where $\mathcal{E}_N(\cdot)$ is defined in (3.10) and α_m is the minimizer. That is, the improved scheme $(\mathcal{P})$ does make an improvement since $\mathcal{P}(N) \le \mathcal{P}_N(0) = \mathcal{E}_N(\alpha_m)$.

The assumption that $\mathscr{I}_N$ is a piecewise constant function attaining finitely many values yields a natural stop criterion of scheme $(\mathcal{P})$ and prevents us from letting $K \to \infty$. Indeed, since $\mathscr{I}_N$ is constant in each $Q_N \in \mathcal{Q}_N$ where $\mathcal{Q}_N$ is defined in (1.7), searching in cubes Q_K such that $K > N$ would not benefit us anymore since $\mathscr{L}(\alpha, v, Q_K) = v$ for any $\alpha \geq 0$ if v is constant in Q_K.

4.2. **The staircasing effect.** In this section we first illustrate with a simple example how $(\mathcal{P})$ avoids the staircasing effect. Figure 1A shows the given corrupted image $\mathscr{I}_N$ and the clean image $\mathscr{C}_N$, with $N = 4$. Scheme $(\mathcal{B})$ results in $\mathscr{L}(\alpha_m, \mathscr{I}_N, I)(I_4(2)) = \mathscr{L}(\alpha_m, \mathscr{I}_N, I)(I_4(3))$ and hence the staircasing effect occurs, as Figure 1B indicates. Scheme $(\mathcal{P})$ operates in the subintervals $I' := (0, 0.5)$ and $I'' := (0.5, 1)$ separately, and hence $\mathscr{L}(\alpha, \mathscr{I}_N, I')(I_4(2))$ and $\mathscr{L}(\alpha, \mathscr{I}_N, I'')(I_4(3))$ are able to break up the staircase produced in Figure 1B and go across each other, as shown in Figure 1C, and finally achieve a better result, as Figure 1D indicates. Moreover, as shown in the end of this paper for the two dimensional case, where Figure 4 and 5 represents the clean image $\mathscr{C}_N$ and corrupted image $\mathscr{I}_N$, respectively. We see in Figure 7, the reconstructed image by scheme $(\mathcal{P})$ results in smaller error value, mitigated staircasing effect (upper right corner), and sharped edge (around the middle area), compare with the reconstructed image by scheme $(\mathcal{B})$ in Figure 6.

We remark that the ability to create a new jump point in $\mathscr{L}(\alpha, \mathscr{I}_N)$, as shown in Figure 1C, is key to avoid the staircasing effect. In [22], the authors proposed a method to avoid the staircasing effect by letting $\alpha = 0$ in certain points and hence at those points new jump points could be created in $\mathscr{L}(\alpha, \mathscr{I}_N)$. In Section 5.2 in [22] they showed that if η_N has average 0 in each subinterval I_i, where $I = \bigcup_{i=1}^M I_i$, and if $\mathscr{C}_N$ is constant in each I_i, then they can achieve a perfect recovery (See Figure 2A to 2C). We remark that our scheme $(\mathcal{P})$ can produce the same perfect recovery result by choosing K large enough such that $\{I_1, \ldots, I_M\} \subset \mathcal{L}_K$. Indeed, invoking Lemma 4.1 we have that, for $\alpha > 0$ large enough,

$$\mathscr{L}(\alpha, \mathscr{I}_N, I_i) = \fint_{I_i} \mathscr{I}_N \, dx = \fint_{I_i} (\mathscr{C}_N + \eta_N) \, dx = \fint_{I_i} \mathscr{C}_N \, dx = \mathscr{C}_N(I_i)$$

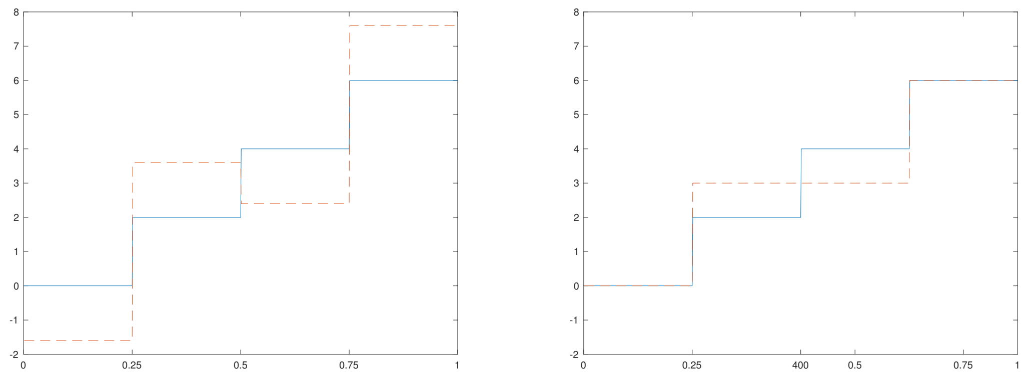
for any $1 \leq i \leq M$, where in the last two equalities we used the assumptions that η_N has average 0 in I_i and that $\mathscr{C}_N$ is constant in I_i.

Finally, we remark that scheme $(\mathcal{P})$ can deal with more generalized situations which cannot be dealt by the method proposed in [22]. For example, in Figure 3A, $\mathscr{C}(x) := x$ and hence $\mathscr{C}_N(I_N(i)) = i/N$ for $x \in I_N(i)$, $1 \leq i \leq N$ (recall $I_N(i)$ from (2.1)). We define $\eta_N(2i-1) = -\eta_N(2i)$, $1 \leq i \leq N/2$. That is, η_N does not have average 0 in each subinterval $I_N(i)$ and so Proposition 5.5 in [22] can not be applied. However, scheme $(\mathcal{P})$ can still provide a perfect recovery result, as shown in Figure 1D, by choosing K large enough such that $\{I_{2i-1} \cup I_{2i},\ 1 \leq i \leq N/2\} \subset \mathcal{L}_K$. Moreover, we observe that scheme $(\mathcal{B})$ produces, again, the staircasing effect, as shown in Fig 3B.

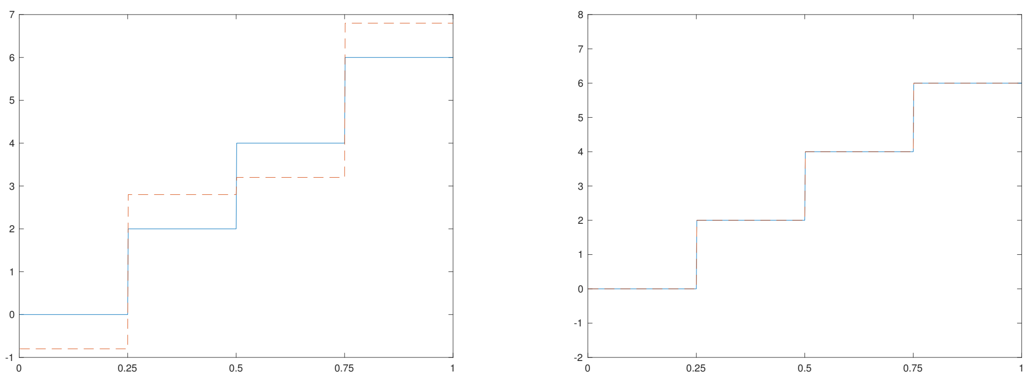
4.3. **Approximation of the clean image and a risk acceleration of scheme $(\mathcal{P})$.** In the last section of this paper, we show that, under mild assumptions on the noise η_N, the scheme $(\mathcal{P})$ can produce a perfect recovery result for an arbitrary clean image $\mathscr{C}$, as the resolution level N goes to ∞.

We recall that $\mathcal{Q}_K$ is defined in (1.7) and $\mathcal{P}(K)$ was introduced in (4.6). Also, we recall a useful corollary for Lusin's Theorem.

Corollary 4.3 ([19], Corollary 1, page 16. Also see [20], 7.10)**.** *Let μ be a Borel regular measure on $\mathbb{R}^N$ and let $f\colon \mathbb{R}^N \to \mathbb{R}^M$ be μ-measurable and bounded. Assume $A \subset \mathbb{R}^N$ is μ-measurable and $\mu(A) < +\infty$. Fix $\varepsilon > 0$. Then there exists a continuous function $\bar{f}\colon \mathbb{R}^N \to \mathbb{R}^M$ such that $\|\bar{f}\|_{L^\infty} \leq \|f\|_{L^\infty}$ and $\mu\left\{x \in A :\ \bar{f}(x) \neq f(x)\right\} < \varepsilon$.*

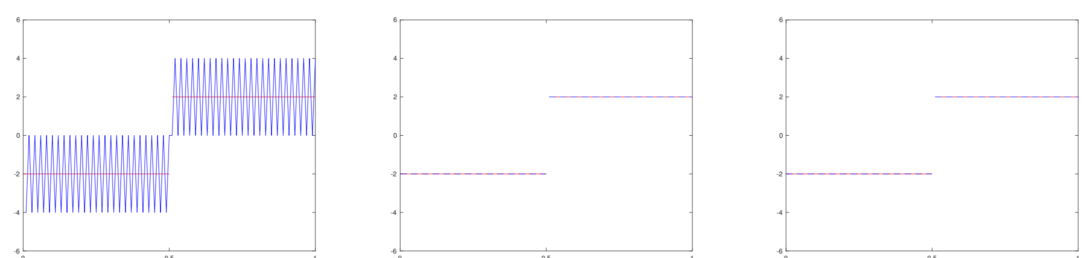


(A) $\mathscr{I}_N$ and $\mathscr{C}_N$ (B) $\mathscr{L}(\alpha_m, \mathscr{I}_N, I)$ produces staircase



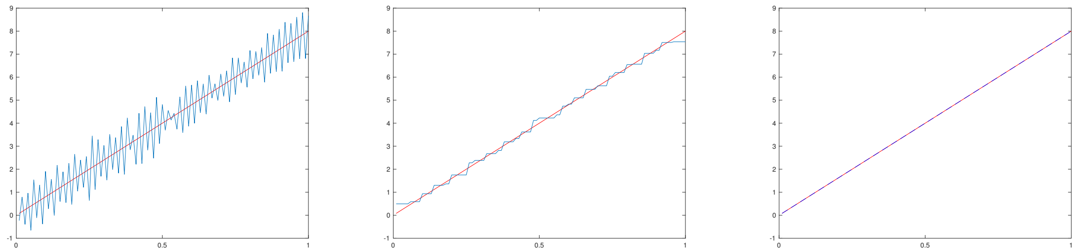
(C) Scheme $(\mathcal{P})$ avoids staircasing (D) $u_{\mathcal{P},N}$ and $\mathscr{C}_N$ overlap, a perfect recovery

FIGURE 1. $I_4(1) = (0, 0.25)$, $I_4(2) = (0.25, 0.5)$, $I_4(4) = (0.5, 0.75)$, $I_4(4) = (0.75, 1)$



(A) η_N has average 0 in I_1 and I_2 (B) perfect recovery by [22] (C) perfect recovery by scheme $(\mathcal{P})$

FIGURE 2. $M = 2$. $I_1 = (0, 0.5)$, $I_2 = (0.5, 1)$



(A) $\mathscr{I}_N$ and $\mathscr{C}_N$ (B) $\mathscr{L}(\alpha_m, \mathscr{I}_N)$ by scheme $(\mathcal{B})$ (C) $u_{\mathcal{P},N}$ by scheme $(\mathcal{P})$

FIGURE 3. $N = 100$. The noise η_N is designed such that $\eta_N(i) = -\eta_N(i+1)$. Note that in Figure 3B, scheme $(\mathcal{B})$ produces staircasing; in Figure 3C, scheme $(\mathcal{P})$ produces an almost perfect recovery

The main theorem of this section is as follows.

Theorem 4.4. *Assume that the noise η_{K^2} has locally average 0, that is*

$$\fint_{Q_K} \eta_{K^2} = 0 \tag{4.7}$$

for any $Q_K \in \mathcal{Q}_K$ and all $k \in \mathbb{N}$. Then

$$\lim_{K\to\infty} \mathcal{P}(K^2) = 0.$$

Proof. Let $K \in \mathbb{N}$ be fixed. Note that $\mathcal{Q}_K \in \mathcal{V}_K$. Then, according to (4.4) and invoking Lemma 4.1, for each $Q_K \in \mathcal{Q}_K$ we have

$$\|\mathscr{L}(\alpha_{Q_K}, \mathscr{I}_{K^2}, Q_K) - \mathscr{C}_{K^2}\|^2_{L^2(Q_K)} \le \left\|\fint_{Q_K} \mathscr{I}_{K^2} dx - \mathscr{C}_{K^2}\right\|^2_{L^2(Q_K)} = \left\|\fint_{Q_K} \mathscr{C}_{K^2}\, dx - \mathscr{C}_{K^2}\right\|^2_{L^2(Q_K)},$$

where in the last equality we used (4.7).

Hence, we have

$$\mathcal{P}(K^2) \le \sum_{Q_K\in\mathcal{Q}_K} \|\mathscr{L}(\alpha_{Q_K}, \mathscr{I}_{K^2}, Q_K) - \mathscr{C}_{K^2}\|^2_{L^2(Q_K)} \le \sum_{Q_K\in\mathcal{Q}_K} \left\|\fint_{Q_K} \mathscr{C}_{K^2}(x) dx - \mathscr{C}_{K^2}\right\|^2_{L^2(Q_K)}.$$

We claim that

$$\lim_{K\to\infty} \sum_{Q_K\in\mathcal{Q}_K} \|\mathscr{C}_{K^2} - \mathscr{C}\|^2_{L^2(Q_K)} = \lim_{K\to\infty} \|\mathscr{C}_{K^2} - \mathscr{C}\|^2_{L^2(Q)} = 0. \tag{4.8}$$

It is clear that (4.8) holds if $\mathscr{C}$ is continuous and using Lebesgue Dominated Convergence Theorem. We prove that (4.8) still hold if $\mathscr{C} \in L^\infty(Q)$. For simplicity, assume that $\|\mathscr{C}\|_{L^\infty(Q)} \le 1$. Fix $\varepsilon > 0$. By Corollary 4.3 there exists a compact set $W \subset\subset Q$ and a continuous function v such that $v\lfloor W = \mathscr{C}\lfloor W$, $\|v\|_{L^\infty} \le \|\mathscr{C}\|_{L^\infty}$, and $\mathcal{L}^2(W) \ge \mathcal{L}^2(Q) - \varepsilon = 1 - \varepsilon$, where $\mathcal{L}^2$ stands for the two dimensional Lebesgure measure. Then we immediately have

$$\int_Q |v - \mathscr{C}|^2\, dx < \varepsilon. \tag{4.9}$$

Let v_{K^2} be defined similarly to $\mathscr{C}_{K^2}$ and we observe that $v_{K^2} \to v$ in $L^2(Q)$. That is,

$$\lim_{K\to\infty} \|v_{K^2} - v\|_{L^2(Q)}^2 = 0. \tag{4.10}$$

We obtain

$$\begin{aligned}\int_Q |v_{K^2} - \mathscr{C}_{K^2}|\, dx &= \sum_{1\le i,j\le K^2} \int_{Q_{K^2}(i,j)} \left| K^2 \int_{Q_{K^2}(i,j)} (v - \mathscr{C})\, dx \right| dy \\ &\le \sum_{1\le i,j\le K^2} K^2 \int_{Q_{K^2}(i,j)} \int_{Q_{K^2}(i,j)} |v - \mathscr{C}|\, dxdy \\ &= \sum_{1\le i,j\le K^2} \int_{Q_{K^2}(i,j)} |v - \mathscr{C}|\, dx = \|v - \mathscr{C}\|_{L^1(Q)} \le 2\varepsilon.\end{aligned}$$

Since $|v_{K^2} - \mathscr{C}_{K^2}| \le 2$ uniformly in W we deduce that

$$\int_Q |v_{K^2} - \mathscr{C}_{K^2}|^2\, dx \le 2\int_Q |v_{K^2} - \mathscr{C}_{K^2}|\, dx \le 4\varepsilon. \tag{4.11}$$

Hence, for $K \in \mathbb{N}$ large enough, and in view of (4.11), (4.10), and (4.9), in this order, we observe that

$$\begin{aligned}\|\mathscr{C}_{K^2} - \mathscr{C}\|_{L^2(Q)}^2 &= \|\mathscr{C}_{K^2} - v_{K^2} + v_{K^2} - v + v - \mathscr{C}\|_{L^2(Q)}^2 \\ &\le 3\,\|\mathscr{C}_{K^2} - v_{K^2}\|_{L^2(Q_K)}^2 + 3\,\|v_{K^2} - v\|_{L^2(Q_K)}^2 + 3\,\|v - \mathscr{C}\|_{L^2(Q_K)}^2 \\ &\le 12\varepsilon + 3\varepsilon + 3\varepsilon = 18\varepsilon\end{aligned}$$

and (4.8) is verified.

Similarly, we could show that (note below we have $\mathscr{C}_K$, but in (4.8) we have $\mathscr{C}_{K^2}$)

$$\lim_{K\to\infty} \sum_{Q_K} \|\mathscr{C}_K - \mathscr{C}\|_{L^2(Q_K)}^2 = 0. \tag{4.12}$$

Note that

$$\fint_{Q_K} \mathscr{C}_{K^2}(y)dy = \mathscr{C}_K(x) \text{ for } x \in Q_K.$$

Then, in view of (4.8) and (4.12),

$$\begin{aligned}&\sum_{Q_K\in\mathcal{Q}_K} \left\| \fint_{Q_K} \mathscr{C}_{K^2}(x)dx - \mathscr{C}_{K^2} \right\|_{L^2(Q_K)}^2 \\ &= \sum_{Q_K\in\mathcal{Q}_K} \|\mathscr{C}_K - \mathscr{C} + \mathscr{C} - \mathscr{C}_{K^2}\|_{L^2(Q_K)}^2 \\ &\le \sum_{Q_K\in\mathcal{Q}_K} \|\mathscr{C}_K - \mathscr{C}\|_{L^2(Q_K)}^2 + \sum_{Q_K\in\mathcal{Q}_K} \|\mathscr{C}_{K^2} - \mathscr{C}\|_{L^2(Q_K)}^2 \to 0\end{aligned}$$

as $K \to \infty$.

Therefore, we deduce that

$$\lim_{K\to\infty} \mathcal{P}(K^2) \le \lim_{K\to\infty} \sum_{Q_K\in\mathcal{Q}_K} \left\| \fint_{Q_K} \mathscr{C}_{K^2}(x)dx - \mathscr{C}_{K^2} \right\|_{L^2(Q_K)}^2 = 0,$$

and the proof is concluded. □

Remark 4.5. The noise η_{K^2} in Theorem 4.4, which has locally zero average, can be produced by using the *compound camera* which is the leading technology in robotic vision. Roughly speaking, the compound camera captures a corrupted image $\mathscr{I}_{K^2}$ with resolution K^2 by capturing with K^2 number of small cameras, each has resolution level K and captures a part of $\mathscr{C}$ in the subdomain Q_K, and these put together yield $\mathscr{I}_{K^2}$. It is usually assumed that each individual camera produces noise with zero average (see, e.g., [9]), which implies that the nose η_{K^2} has average zero in each Q_K as required.

We remark that the CPU time needed to produce the best reconstructed image $u_{\mathcal{P},N}$ by using scheme $(\mathcal{P})$ is relatively long. Typically, for a 1024×1024 resolution image $\mathscr{I}_N$, it takes around 2 hours to produce $u_{\mathcal{P},N}$, and we note that the CPU time needed increased exponentially as resolution level increase. One reason for such long CPU time consumed is because $\mathcal{V}_K$ is nested, i.e., $\mathcal{V}_K \subset \mathcal{V}_{K+1}$, and thus at each layer $\mathcal{V}_K$, we have to search within, not only $\mathcal{V}_K$, but all $\mathcal{V}_i$, $i \le K$. However, such design in scheme $(\mathcal{P})$ is important since it guarantees that

$$\mathcal{P}_N(K) \ge \mathcal{P}_N(K+1)$$

and, in turn, it ensures that we are approaching to minimizer $u_{\mathcal{P},N}$ as $K \to \infty$. In practice we may pre-fix a small acceptable error $\varepsilon > 0$ such that once we reach at K_0 large such that

$$\mathcal{P}_N(K_0) - \mathcal{P}_N(K_0+1) < \varepsilon,$$

we stop scheme $(\mathcal{P})$ and use

$$u_{\mathcal{P},K_0} := \arg\min \left\{ \int_Q \left|\mathscr{C}_N - u_{\mathcal{L}_{K_0}}\right|^2 dx,\ \mathcal{L}_{K_0} \in \mathcal{V}_{K_0} \right\}$$

as an approximation of the minimizer produced by scheme $(\mathcal{P})$.

To further accelerate $(\mathcal{P})$, we propose a simplified version of scheme $(\mathcal{P})$, the scheme $(\mathcal{P}')$ as follows (recall $\mathcal{Q}_K$ from (1.7)):

Level 1.

$$u_{\mathcal{P}',N} := \arg\min \left\{ \int_Q |\mathscr{C}_N - u_{\mathcal{Q}_K}|^2 dx :\ K \ge 0 \right\}$$

Level 2.

$$u_{\mathcal{Q}_K}(x) := \mathscr{L}(\alpha_{Q_K}, \mathscr{I}_N, Q_K) \text{ for } x \in Q_K \text{ and } Q_K \in \mathcal{Q}_K,$$

$$\text{where } \alpha_{Q_K} := \arg\min \left\{ \int_{Q_K} |\mathscr{L}(\alpha, \mathscr{I}_N, Q_K) - \mathscr{C}_N|^2 dx :\ \alpha > 0 \right\}.$$

Note that $\mathcal{Q}_K \in \mathcal{V}_K$ and so $\mathcal{P}(N) \le \mathcal{P}'(N)$ where

$$\mathcal{P}'(N) := \int_Q |\mathscr{C}_N - u_{\mathcal{P}',N}|^2 dx.$$

That is, the optimized reconstructed image $u_{\mathcal{P}',N}$ produced by scheme $(\mathcal{P}')$ might result in a higher error compare with the optimized reconstructed image $u_{\mathcal{P},N}$ produced by scheme $(\mathcal{P})$, and in turn, $u_{\mathcal{P}',N}$ is worse than $u_{\mathcal{P},N}$. However, as the resolution level N gets large, $u_{\mathcal{P}',N}$ becomes close to $u_{\mathcal{P},N}$ as the following theorem asserts.

Theorem 4.6. *Assume that the noise η_{K^2} has locally average* 0*, that is,*

$$\fint_{Q_K} \eta_{K^2} = 0$$

for any $Q_K \in \mathcal{Q}_K$ where $\mathcal{Q}_K$ is defined in (1.7)*. Then*

$$\lim_{K\to\infty} \mathcal{P}'(K^2) = 0.$$

Proof. The proof is same to the proof of Theorem 4.4. □

Theorem 4.4 and Theorem 4.6 imply that

$$\lim_{N\to\infty} \int_Q |u_{\mathcal{P}',N} - u_{\mathcal{P},N}|^2 \, dx = 0,$$

and hence scheme $(\mathcal{P}')$ could be used as a "risky" replacement of scheme $(\mathcal{P})$ if the resolution level N is large enough, although

$$\mathcal{P}(N) < \mathcal{P}'(N)$$

might happen.

The advantage of scheme $(\mathcal{P}')$ is that, not only it can reduce the staircasing effect as the example proposed in Figure 1c to 1d applies to $(\mathcal{P}')$ as well, but also, and more importantly, the scheme $(\mathcal{P}')$ does require much less CPU time as compared to scheme $(\mathcal{P})$, since $(\mathcal{P}')$ only searchs within $\mathcal{Q}_K$, which is a much smaller collection than $\mathcal{V}_K$ in which $(\mathcal{P})$ searches. In practice, for a 1024×1024 resolution level image, $(\mathcal{P}')$ only need around 10 minutes to reach its optimal solution, but $(\mathcal{P})$ usually takes around 2 hours to reach its.

5. Adaptation of generalized regularizers - a comprehensive learning scheme

We recall that the level 2 of scheme $(\mathcal{P})$ utilizes the scheme $(\mathcal{B})$ in each subdomain of Q, and $(\mathcal{B})$ uses TV as its regularizer. There are many choice of possible regularizers. For example, $ICTV^k$ in [23], TV^k in [30], and TGV^k in [29]. Therefore, to consider a broad spectrum of regularizers at the same time, we introduce the following comprehensive learning scheme $(\mathcal{CT})$:

Level 1.

$$u_{\mathcal{CT},N} := \arg\min \left\{ \mathcal{F}\left(\mathscr{C}_N - u, Q\right) : \; K \geq 0, \; u \in \mathcal{T}_K \right\}$$

Level 2.

$$\mathcal{T}_K := \left\{ u_{\mathcal{CT}_K,N}, \; \mathcal{L}_K \in \mathcal{V}_K \right\}$$

$$u_{\mathcal{CT}_K,N}(x) \text{ is constructed upon the information obtained by } u_{Q_K,N} \tag{5.1}$$

Level 3.

$$u_{Q_K,N} := \arg\min \left\{ \mathcal{F}(\mathscr{I}_N - u, Q_K) + R(u, \alpha, Q_K), \; u \in SBV(Q), \; R \in \mathcal{R} \right\},$$

where $\mathcal{F}$ is the fidelity term which is quasiconvex in the sense of [12], and $\mathcal{R}$ is a collection of regularizers. Indeed, letting $\mathcal{F}$ be L^2 norm, $R(u, \alpha, Q_K) = \alpha TV(u, Q_K)$, and in (5.1) using the direct construction, i.e., $u_{\mathcal{CT}_K,N}(x) := u_{Q_K,N}$ for $x \in Q_K$, we recover scheme $(\mathcal{P})$.

We remark that by choosing the right regularizers, scheme $(\mathcal{CT})$ can, in addition to learn the parameter of the regularizers, also learn the order of the regularizers. In my recent work with Elisa Davoli (see [15]) , a new fractional order seminorm, $ICTV^r_{\tilde{\alpha}}$, $r \in \mathbb{R}$, $r \geq 1$, $\tilde{\alpha} \in \mathbb{R}_+^{\lfloor r \rfloor + 1}$ is proposed in the one-dimensional setting, as a generalization of the standard $ICTV^k_{\tilde{\alpha}}$-seminorms, $k \in \mathbb{N}$, $\tilde{\alpha} \in \mathbb{R}^{k+1}_+$. To be precise, we recall that the *Gagliardo seminorm* (see [21, 27]) for $u \in L^p$ is defined as

$$|u|_{W^{s,p}(I)} := \left(\int_I \int_I \frac{|u(x) - u(y)|^p}{|x - y|^{1+sp}}\, dx\, dy \right)^{\frac{1}{p}}.$$

where $0 < s < 1$ and $p \geq 1$. We define the *fractional ICTV* seminorm below:

Definition 5.1. *Let $0 < s < 1$, $k \in \mathbb{N}$, and let $\tilde{\alpha} = (\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_{k+1}) \in \mathbb{R}^{k+1}_+$. For every $u \in L^1(I)$, we define its fractional $ICTV^{k+s}$ seminorm as follows:*

1. for $k = 1$

$$|u|_{ICTV^{1+s}_{\tilde{\alpha}}(I)} := \inf \Big\{ \alpha_0 \left| u' - s v_0 \right|_{\mathcal{M}_b(I)} + \alpha_1 s(1-s) \left| v_0 \right|_{W^{s,1+s(1-s)}(I)} : \\ v_0 \in W^{s,1+s(1-s)}(I),\ \int_I v_0(x)\, dx = 0 \Big\}.$$

2. for $k > 1$

$$|u|_{ICTV^{k+s}_{\tilde{\alpha}}(I)} := \inf \Big\{ \alpha_0 \left| u' - v_0 \right|_{\mathcal{M}_b(I)} + \alpha_1 \left| v_0' - v_1 \right|_{\mathcal{M}_b(I)} + \\ \cdots + \alpha_{k-1} \left| v_{k-2}' - s v_{k-1} \right|_{\mathcal{M}_b(I)} + \alpha_k s(1-s) \left| v_{k-1} \right|_{W^{s,1+s(1-s)}(I)} : \\ v_i \in BV(I) \text{ for } 0 \leq i \leq k-2,\ v_{k-1} \in W^{s,1+s(1-s)}(I),\ \int_I v_{k-1}(x)\, dx = 0 \Big\}.$$

We say that $u \in BCV^{k+s}_{\tilde{\alpha}}(I)$ if

$$\|u\|_{BCV^{k+s}_{\tilde{\alpha}}(I)} := \|u\|_{L^1(I)} + |u|_{ICTV^{k+s}_{\tilde{\alpha}}(I)} < +\infty.$$

We restrict our analysis to the case in which $\tilde{\alpha}$ and r satisfy the *box constraint* (see, e.g. [4, 16])

$$(\tilde{\alpha}, r) \in [\hat{\alpha}, \hat{\beta}]^{\lfloor r \rfloor + 1} \times [1, \hat{r}] \tag{5.2}$$

where $\hat{\alpha} > 0$, $\hat{\beta} > 0$, and $\hat{r} > 1$ are fixed real numbers. Then, using our notation, we propose the following learning scheme $(\mathcal{D})$ which takes into account the order of the regularizer and the parameter $\tilde{\alpha} \in \mathbb{R}^{\lfloor r \rfloor + 1}_+$ simultaneously, where $r \geq 1$ is given and $\lfloor r \rfloor$ denotes the largest integer smaller than or equal to r.

Level 1.

$$(\bar{\alpha}, \bar{r}) := \arg\min \left\{ \int_I \left| u_{\tilde{\alpha},r} - \mathscr{C}_N \right|^2 dx,\ (\tilde{\alpha}, r) \in [\hat{\alpha}, \hat{\beta}]^{\lfloor r \rfloor + 1} \times [1, \hat{r}] \right\},$$

Level 2.

$$u_{\tilde{\alpha},r} := \underset{u \in BCV^r_{\tilde{\alpha}}(I)}{\arg\min} \left\{ \int_I \left| u - \mathscr{I}_N \right|^2 dx + |u|_{ICTV^r_{\tilde{\alpha}}(I)} \right\}.$$

Then the following theorem is established.

Theorem 5.2. *Under the box constraint (5.2), the learning scheme $(\mathcal{D})$ admits a unique solution $(\bar{\alpha}, \bar{r}) \in [\hat{\alpha}, \hat{\beta}]^{\lfloor \bar{r} \rfloor + 1} \times [1, \hat{r}]$ and provides an associated optimally reconstructed image $u_{\bar{\alpha},\bar{r}} \in BCV^{\bar{r}}_{\bar{\alpha}}(I)$.*

Then, inserting scheme $(\mathcal{D})$ into the level 3 of scheme $(\mathcal{CT})$, we obtain a learning scheme which learns a spatially dependent parameters as well as a spatially dependent regularizer, where I_K, $\mathcal{L}_K$, and $\mathcal{V}_K$ defined in one dimension similar to (4.2).

Level 1.

$$u_{\mathcal{CT},N} := \arg\min\left\{\int_I |\mathscr{C}_N - u_{\mathcal{L}_K}|^2\,dx,\ K \geq 0,\ \mathcal{L}_K \in \mathcal{V}_K\right\}$$

Level 2.

$$u_{\mathcal{L}_K}(x) := u_{I_K,N} \text{ for } x \in I_K \text{ and } I_K \in \mathcal{L}_K,$$

Level 3. $u_{I_K,N} := u_{\bar{\alpha},\bar{r}}$ for $x \in I_K$, where

$$(\bar{\alpha},\bar{r}) := \arg\min\left\{\int_{I_K} |u_{\tilde{\alpha},r} - \mathscr{C}_N|^2\,dx,\ (\tilde{\alpha},r) \in [\hat{\alpha},\hat{\beta}]^{\lfloor r\rfloor+1} \times [1,\hat{r}]\right\}$$

and

$$u_{\tilde{\alpha},r} := \underset{u\in BCV^r_{\tilde{\alpha}}(I_K)}{\arg\min}\left\{\int_{I_K} |u - \mathscr{I}_N|^2\,dx + |u|_{ICTV^r_{\tilde{\alpha}}(I_K)}\right\}.$$

We conclude this paper with an outlook on current and future works providing more options of the regularizers and construction in (5.1).

Regarding new regularizers: in [14] a $\mathscr{A}$-$\mathscr{B}$ quasiconvex seminorm is proposed as a further extension of fractional order $ICTV^r$ seminorm.

Regarding the new construction in (5.1): using scheme ($\mathcal{P}$) as an example, in [26] we introduce a spatially dependent parameter $\alpha_{\mathcal{P}_K,N}(\cdot)$ defined by $\alpha_{\mathcal{P}_K,N}(x) := \alpha_{Q_K,N}$, for $x \in Q_K \in \mathcal{P}_K$, where $\alpha_{Q_K,N}$ is given in (4.5), and we replace $u_{\mathcal{P}_K,N}$ defined in (4.4) by setting

$$u_{\mathcal{P}_K,N} := \mathcal{L}(\alpha_{\mathcal{P}_K,N}, \mathscr{I}_N, Q).$$

This new construction for $u_{\mathcal{P}_K,N}$ brings in several advantages. For example, $u_{\mathcal{P}_K,N}$ is expected to be "close" to $\mathscr{C}_N$ locally in Q_K since $\alpha_{\mathcal{P}_K,N}$ is defined by locally optimizing parameter $\alpha_{Q_K,N}$, and, at the same time, it is expected to have a good balance between local optimization and global optimization since it is reconstructed over the entire domain Q.

Acknowledgements

P. Liu is partially funded by the National Science Foundation under Grant No. DMS - 1411646. The author wishes to thank Professor Irene Fonseca and Professor Noel J. Walkington for many helpful discussions and comments.

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] M. Avriel. *Nonlinear programming: analysis and methods*. Courier Corporation, 2003.

[3] G. Bellettini, V. Caselles, and M. Novaga. The total variation flow in $\mathbb{R}^N$. *J. Differential Equations*, 184(2):475–525, 2002.

[4] 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.

[5] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, Cambridge, 2004.

FIGURE 4. Clean image $\mathscr{C}_N$

[6] K. Bredies and T. Valkonen. Inverse problems with second-order total generalized variation constraints. *Proceedings of SampTA*, 201, 2011.

[7] M. Burger, G. Gilboa, M. Moeller, L. Eckardt, and D. Cremers. Spectral Decompositions using One-Homogeneous Functionals. *ArXiv e-prints*, Jan. 2016.

[8] V. Caselles, A. Chambolle, and M. Novaga. The discontinuity set of solutions of the TV denoising problem and some extensions. *Multiscale Model. Simul.*, 6(3):879–894, 2007.

[9] A. Chambolle and P.-L. Lions. Image recovery via total variation minimization and related problems. *Numer. Math.*, 76(2):167–188, 1997.

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

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

[12] B. Dacorogna. *Direct methods in the calculus of variations*, volume 78 of *Applied Mathematical Sciences*. Springer, New York, second edition, 2008.

[13] G. Dal Maso, I. Fonseca, G. Leoni, and M. Morini. A higher order model for image restoration: the one-dimensional case. *SIAM J. Math. Anal.*, 40(6):2351–2391, 2009.

[14] E. Davoli, I. Fonseca, and P. Liu. Image reconstruction and $\mathscr{A}$-$\mathscr{B}$ quasiconvexity. *In preparation.*

[15] E. Davoli and P. Liu. The optimal parameters and regularizers for image reconstruction problems. *Preprint.*

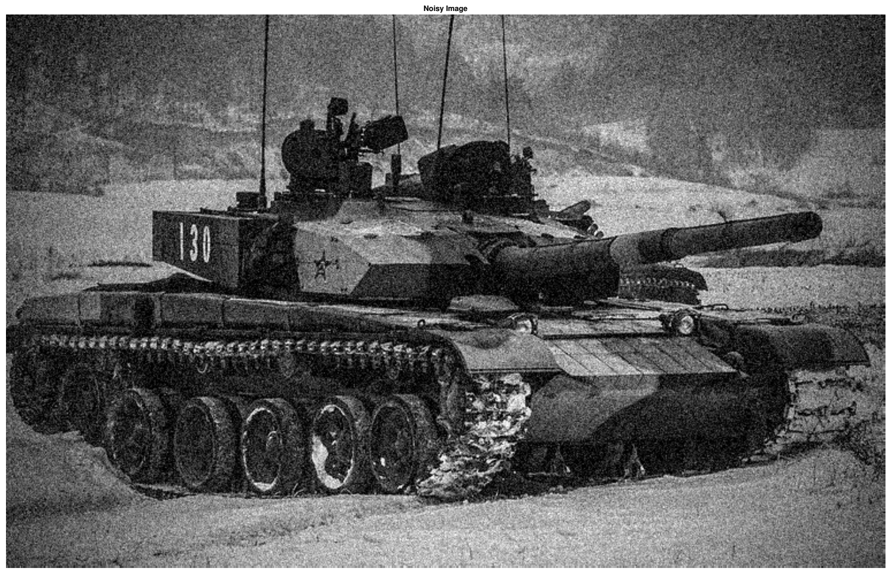


FIGURE 5. Corrupted image $\mathscr{I}_N$, where the artificial noise is added by using a Gaussian noise distribution

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

[17] 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.

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

[19] L. C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2010.

[20] G. B. Folland. *Real analysis: modern techniques and their applications*. John Wiley & Sons, 2013.

[21] E. Gagliardo. Proprietà di alcune classi di funzioni in più variabili. *Ricerche Mat.*, 7:102–137, 1958.

[22] M. Hintermüller, K. Papafitsoros, and C. N. Rautenberg. Analytical aspects of spatially adapted total variation regularisation. *ArXiv e-prints*, Sept. 2016.

[23] M. Holler and K. Kunisch. On infimal convolution of TV-type functionals and applications to video and image reconstruction. *SIAM J. Imaging Sci.*, 7(4):2258–2300, 2014.

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

[25] P. Liu. The quasi-convexity of bi-level learning scheme. *In preparation.*

[26] P. Liu. The spatially dependent bi-level learning scheme. *Preprint.*

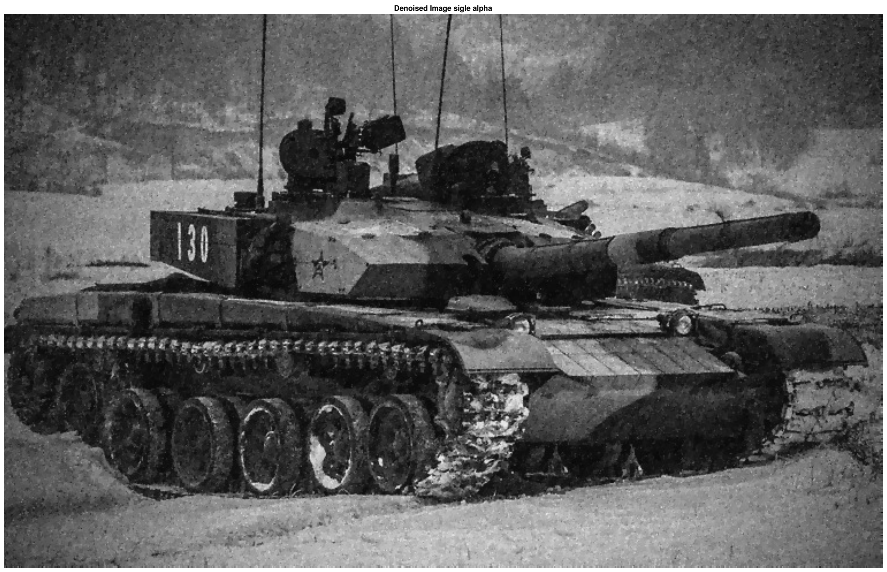


FIGURE 6. The reconstructed image by scheme ($\mathcal{B}$). The training error is 931.667. Note that the staircasing effect is observed, upper left and right corner.

[27] V. Maz′ya and T. Shaposhnikova. On the Bourgain, Brezis, and Mironescu theorem concerning limiting embeddings of fractional Sobolev spaces. *J. Funct. Anal.*, 195(2):230–238, 2002.
[28] 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.
[29] K. Papafitsoros and K. Bredies. A study of the one dimensional total generalised variation regularisation problem. *arXiv preprint arXiv:1309.5900*, 2013.
[30] K. Papafitsoros and C.-B. Schönlieb. A combined first and second order variational approach for image reconstruction. *Journal of mathematical imaging and vision*, 48(2):308–338, 2014.
[31] L. Rudin. Segmentation and restoration using local constraints. *Technical Repport*, 36, 1993.
[32] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational methods in imaging*, volume 167 of *Applied Mathematical Sciences*. Springer, New York, 2009.
[33] J. C. Spall. *Introduction to stochastic search and optimization*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, 2003. Estimation, simulation, and control.
[34] D. M. Strong, T. F. Chan, et al. Exact solutions to total variation regularization problems. In *UCLA CAM Report*. Citeseer, 1996.
[35] 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.

Figure 7. The reconstructed image by scheme ($\mathcal{P}$). The training error is 900.325. Note that the staircasing effect is reduced, and edges are sharper