THEORETICAL AND NUMERICAL APPROXIMATION OF THE RUDIN-OSHER-FATEMI MODEL FOR IMAGE DENOISING IN THE CONTINUOUS SETTING

by

MATAMBA MESSI, LEOPOLD

(Under the Direction of Lai, Ming-Jun)

ABSTRACT

This dissertation studies the approximation of the continuous total variation based model for image denoising by piecewise polynomial functions on polygonal domains. Our main contributions are the explicit construction of a continuous piecewise linear approximation on rectangular domains, and the construction of a minimizing sequence of bivariate splines of arbitrary degree for a general polygonal domain. For rectangular domains, we propose an alternate discretization of the ROF model and construct the continuous piecewise linear function as the piecewise linear interpolation of the minimizer of the new discrete model. Whereas on general polygonal domains, we use the Galerkin method to define the spline approximation as the minimizer of the ROF functional over a spline space. We then show that when given a suitable family of triangulations, our approach generates a minimizing sequence for the total variation model. In each case, we use an extension argument to show that the approximation converges to the ROF minimizer in the strict topology of the space of bounded variation functions.

INDEX WORDS: Image denoising; Total variation regularization; Projected-gradient algorithm; Fixed-point algorithm; Bivariate spline approximation; Polynomial interpolation.

THEORETICAL AND NUMERICAL APPROXIMATION OF THE RUDIN-OSHER-FATEMI MODEL FOR IMAGE DENOISING IN THE CONTINUOUS SETTING

by

MATAMBA MESSI, LEOPOLD

FORB, University of Yaounde I, Cameroon, 2001

FORM, University of Yaounde I, Cameroon, 2004

MS, University of Georgia, 2011

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2012

THEORETICAL AND NUMERICAL APPROXIMATION OF THE RUDIN-OSHER-FATEMI MODEL FOR IMAGE DENOISING IN THE CONTINUOUS SETTING

by

MATAMBA MESSI, LEOPOLD

Major Professor:	Lai, Ming-Jun
Committee:	Varley, Robert
	Kazanci, Caner
	Petukhov, Alexander

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2012

To the memory of my Dear Mother Hedwige Ngoh Tsanga

ACKNOWLEDGEMENTS

First and foremost, I would like to express my immense gratitude to Dr. Ming-Jun Lai for accepting to direct this dissertation and for his great generosity and infinite patience. Thank you for putting up with me for so long and making this day come to pass. Thank you for all the thanksgiving dinners at your house, I enjoyed every minute of it, away from mathematics.

To all the members of my graduate committee, I would like to say that I am very appreciative of your accepting to serve on this committee and seeing to it that I complete the program. Each of you has contributed to my academic and/or professional development in your own unique way; I will forever be grateful for your efforts. I am very thankful to Dr. Caner Kazanci for his invaluable counsels. He encouraged me to plan my program of study with an eye beyond the degree and to get some publications in print prior to graduation. I heeded his advice and with his collaboration I succeeded in that regards; I am now looking forward to greener pastures as I complete this degree.

The graduate faculty of the mathematics department has been fantastic. Your love and dedication to your work have kindled a fire that I will now strive to sustain as I carry on with my career. Many of you have marked my stay at The University of Georgia in a unique way. I will forever be grateful to Dr. Ted Shifrin; my taking a class from you has been the most positive classroom experience of my college life. It has reinforced my desire to become a teacher for life. Thank you for telling me of my potential; your voice kept me going when I felt burnt out. Dr. Robert Varley, your humanity and kindheart are unique; I may never come across such a caring and encouraging person again in my life. I am glad I choose this department for my graduate studies.

While I am earning my doctorate degree in the US, it would not have been possible without the mathematical foundation given to me by the mathematics faculty at the University of Yaounde I, Cameroon. The most prominent among them are: Dr. Nicolas G. Andjiga from whom I learned to write rigorous mathematical proofs, and Dr. Mama Foupouagnigni who insisted that I pursue graduate education and provided me with employment opportunities so that I may sustain myself. These two persons are father figures to me and have deeply shaped my academic career. I will never be able to repay them, but will strive to mirror their deep care for their students in my career.

My transition to the United States was a rocky one; I would not have made it had it not been for the many kind persons that crossed my path here in Athens. To my dear friends Michele and Ryan Ross, I say thank you from the bottom of my heart. You sacrificed a lot of your valuable time and belonging to make sure that I felt at home. Your attentive ears and encouragements kept me in at the most difficult period: my first year in Athens. I would like show my deep appreciation to Jerry and Philomene Saliki for recreating a sense of our homeland once a month at their family home.

Many persons have made my academic and family life very pleasant, as a thank you note, I will list their names here: Phil and Linda Smith, Greg and Amy Wheeler, Herbert and Christine Ssegane, the church family at Green-Acres Baptist Church, and all the graduate students of the mathematics department, especially, Yang Liu, Weinjing Li, Nham Ngo, and Jane Hong.

I am infinitely grateful to my parents, Laurent Messi Messi and Hedwige Ngoh Tsanga, for making the right choice and giving me an education. *All the praises to my late mother; she sacrificed her entire life just for me. Although you did not live long enough to see this achievement, I hope that I have made you proud, mama*. Many members of my extended family accompanied my parents in the arduous task of raising a son. Most prominently among them are my uncles: Jean Paul Ngah Messi, Alphonse Victor Messi, and the lates Wenceslas Nkada and Cyrille Ntede. They were my guidance counselors and constantly reminded me that education is my only shot at a decent life. I am glad that I listened to your counsels and know you are very proud of me.

More importantly, I would like to mention my lovely wife, Marie Matamba, and my daughters Danielle and Esther Eliora. I can never repay you, mommy, for your sacrifices over the last five years. Thank you for your unrelenting support and for taking care of me and the girls; my apologies for all the frustration and the long lonely days. Last but not the least, all Glory and Honor to God; You have made all things possible and I shall continue to Glorify Your Name.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1

INTRODUCTION

Rapid technological advances and the increasing miniaturization of electronic components have put in the hands of the average citizen electronic gadgets of exceptional quality. In the medical field, imaging devices have revolutionized the medical practice. It is now possible to carry high precision surgical procedures with minimal damage to the muscular apparatus, resulting in reduced recovery time. We have also gained the ability to diagnose most of our internal ailments using imaging technologies.

Unfortunately, the images that we capture are not always of good quality for various reasons ranging from defects of the imaging sensors to the lack of dexterity by the user. Some “noise” is introduced, and it is necessary to remove as much of it as possible prior to using the image for its intended purpose, making image cleaning a necessary preprocessing step to any imaging application. However, the notion of noise is vague, and its origin and structure are very hard to model.

The image processing field is a multidisciplinary endeavor ranging from filtering theory to partial differential equations (PDE), and remains an active area of research. Image denoising models come in two broad classes: the PDE approach pioneered by J.-M. Morel [5–7, 30], and the variational approach in which images are modeled as oscillatory spaces and the cleaning process is reduced to functionals minimization over a suitable functional space [39, 58, 71].

One of the most “efficient” models for image denoising is the total variation minimization model of Rudin, Osher and Fatemi (ROF) [71]. These scholars proposed to recover a cleaner image as the function with minimum total variation subject to constraints corresponding to a priori knowledge of the statistics of the noise. The striking thing about the

ROF model is its capability to remove noise while enhancing the edges in the image. The study of the ROF model has generated a sizeable literature, the bulk of which deals with algorithmic considerations.

In their seminal paper [71] introducing the total variation model, Rudin et al. enforced the constraints using the Lagrange multipliers method, and used the gradient projection method of Rosen [69, 70] with a Lagrange multiplier update to compute approximations of the minimizer of the associated Lagrangian. A big drawback of the ROF Lagrangian is its nondifferentiability; Acar and Vogel [1] circumvented this with a relaxation of the total variation functional. Using a duality argument, they constructed a family of functionals that approximate the total variation functional uniformly, and showed that the minimizers of these functionals yielded an approximation of the clean image recovered with the ROF model.

Chambolle and Lions [35] also used a functional relaxation technique to construct a minimizing sequence of the ROF model in the Hilbert space $H^1(\Omega)$, using Γ-convergence arguments. Exploiting the functionals in [1], Dobson and Vogel [41] studied finite elements approximation of the ROF functional and developed an algorithm for computing approximations of the ROF minimizer. Hong [48] followed a similar line of reasoning to study the so called minimal surface bivariate spline image enhancement approach. The Galerkin method is used on spaces of smooth bivariate splines to construct a minimizing sequence of the Acar-Vogel relaxation of the ROF model when its minimizer belongs to $\mathrm{W}^{1,1}(\Omega)$.

The gradient descent algorithm studied by Rudin et al. [71] may be regarded as a fully explicit numerical scheme for the total variation flow. Feng and Prohl [45] studied the approximation of the total variation flow and established the optimal convergence rate of a fully explicit finite elements approximation of the latter. Wang and Lucier [74] established error rates of a piecewise constant approximation of the ROF model under the assumptions that the image being denoised is L^2-Hölder continuous and the domain of the image is rectangular; their arguments do not extend to general polygonal domains. Their approach did not use the functional relaxation method, instead they exploited the stability property of the ROF model to give a direct proof of convergence in $L^2(\Omega)$.

Recently, the regularity property of the ROF model has been studied [29] and it was shown that the model preserves the modulus of continuity of the data when the domain is convex. The research carried in this dissertation is motivated by the latter regularity property, and the need of a computational scheme that could be used to simulate the smooth solutions of the ROF model.

This dissertation is the first time an interpolation argument is used in conjunction with a finite difference scheme to construct a continuous approximation of the ROF model. It also contains a substantial improvement on the approximation of the ROF model on arbitrary polygonal domain. Unlike in most of the works mentioned above, relaxation methods are not used in the construction of the minimizing sequence. This work also improves Hong's [48] work significantly by producing a direct approximation of the ROF model with bivariate splines that converges in the space of functions of bounded variation. The dissertation is organized as follows.

The second chapter is devoted to the mathematical preliminaries and the total variation denoising model. We set up the mathematical framework in which the total variation model of image denoising is posed, then introduce the ROF model and review its properties as relevant to this work.

The third chapter contains the first contribution of this dissertation. In the specific case where the image domain is rectangular, we construct a piecewise linear interpolation function and prove its convergence to the ROF minimizer when the data are bounded and Hölder continuous with respect to the L^2-norm. Along the way, we also introduce a novel discretization of the total variation functional, which enforces more of the desired Neumann boundary conditions.

In chapter 4, we develop three algorithms for computing the minimizer of the new discrete total variation model introduced in chapter 3. These are: the dual projected-gradient algorithm, the alternating dual projected-gradient algorithm, and the alternating dual fixed point algorithm. The dual projected-gradient algorithm fits in the class of proximal gradient algorithms so its convergence could be derived from the general theory of such algorithms, however, we provide a direct proof. We also establish the convergence of the alternating fixed point algorithm, and conjecture the convergence of the alternating dual projected gradient algorithm.

Chapter 5 deals with the approximation of the ROF model on arbitrary polygonal domains. Unlike in chapter 3 where we used an interpolation scheme, here we use the Galerkin method to construct a bivariate spline minimizing sequence. The convergence to the minimizer is established for any data function in the space of square integrable functions. For numerical simulation purposes, we study a relaxation of the ROF model from which an algorithm is derived. A proof of the convergence of the algorithm is derived and the details of its implementation provided.

A key tool in our analysis is the existence of an extension operator T from $BV(\Omega)$ into $BV(\mathbb{R}^n)$ such that for all $u \in BV(\Omega)$, $T(u)$ is compactly supported and the total variation of $T(u)$ on the boundary of Ω is zero. For the type of domains that we are concerned with, the result already existed in the literature and we reviewed it in the appendix A.

2

PRELIMINARIES AND THE IMAGE DENOISING PROBLEM

In this chapter, we present the total variation based image denoising model and give an overview of the research efforts that have contributed significantly to the understanding of the model. We also present a general framework on which the total variation model is founded.

2.1 FUNCTIONAL MINIMIZATION OVER BANACH SPACES

In this section, we review sufficient conditions on the functional F that guarantee the existence of a solution to (2.1), and discuss the functional framework in which the denoising problem is posed to guarantee the existence of the solution. The image denoising model that is the object of this dissertation falls in the category of problems of the form

$$\text{Find } u \in X \text{ such that } F(u) = \inf_{x \in X} F(x), \tag{2.1}$$

where X is a Banach space and F is a functional defined on X that takes values in $[-\infty, +\infty]$. We remark that problem (2.1) is of interest only if F is not identically $-\infty$ or $+\infty$; this leads us to the following definition.

DEFINITION 2.1. A function $F : X \to [-\infty, \infty]$ is said to be proper if for any $x \in X$, either $F(x) > -\infty$ or $F(x)$ is not defined, and there exists $x_0 \in X$ such that $F(x_0) < \infty$. The domain of F, $\mathrm{Dom}(F)$, is defined by

$$\mathrm{Dom}(F) = \{x \in X \colon F(x) \in \mathbb{R}\}. \tag{2.2}$$

Let us first describe a common strategy used in proving the existence of a solution of (2.1). The argument is carried out in the following three steps.

(1) We construct a minimizing sequence $x_n \in X$ such that

$$\lim_{n\to\infty} F(x_n) = \inf_{x\in X} F(x).$$

(2) If F is *coercive*, *i.e.* $\lim_{\|x\|\uparrow\infty} F(x) = \infty$, we deduce that the minimizing sequence x_n is bounded. Then using the topological properties of X, we may extract a subsequence x_{n_j} of x_n that converges to u in X, and show that u is a solution.

(3) To show that u is a solution, it suffices to argue that $\liminf_{j\to\infty} F(x_{n_j}) \geq F(u)$ from which it follows easily that $F(u) = \min_{x\in X} F(x)$.

The strategy outlined above requires that both the Banach space X and the functional F have some favorable properties. We now discuss properties of X and F that are sufficient for the procedure to be carried successfully.

Let $(X, \|\cdot\|)$ be a Banach space over $\mathbb{R}$ and X' its topological dual, *i.e.* the space of all continuous linear functionals defined of X and taking values in $\mathbb{R}$.

DEFINITION 2.2. Let $(x_n)_{n\geq 1}$ be a sequence of elements of X, and $x \in X$.

(a) We say that x_n converges strongly to x if $\|x_n - x\| \to 0$ as $n \to \infty$.

(b) We say that x_n converges weakly to x, and write $x_n \rightharpoonup x$, if

$$\ell(x_n) \to \ell(x) \text{ for all } \ell \in X'.$$

The dual X' of X is a Banach space when endowed with its natural dual norm

$$\|\ell\|_* = \sup_{x\in X,\, x\neq 0} \frac{\ell(x)}{\|x\|}$$

so that the concepts of strong and weak convergence are well defined on X'. A third and often more convenient (than the weak convergence) notion of convergence may be defined on X'.

DEFINITION 2.3. We say that a sequence of linear functional (ℓ_n) converges weakly-* to ℓ in X', and write $\ell_n \overset{*}{\rightharpoonup} \ell$, if $\ell_n(x) \to \ell(x)$ for all $x \in X$.

We note that for any $x \in X$, the mapping $\ell \mapsto \ell(x)$ is a bounded linear functional on X'. Moreover the map $E : X \to (X')'$ defined by the identity

$$\langle Ex, \ell \rangle_{X'',X'} = \langle \ell, x \rangle_{X',X} := \ell(x), \quad \forall x \in X,\ \forall \ell \in X'$$

is one-to-one, so that X is identified to a subspace of $(X')'$. E is often referred to as the canonical embedding of X into $(X')'$.

DEFINITION 2.4. We say that the space X is reflexive if the canonical embedding E is surjective, *i.e* $E(X) = (X')'$.

As a direct consequence of the Riesz representation theorem, every Hilbert space is reflexive. The following result is very useful in the study of problems of the type (2.1).

THEOREM 2.5 (Brezis [27]). *A Banach space X is reflexive if and only if every bounded sequence in X has a weakly convergent subsequence.*

Now, we turn to the properties of the functional F sufficient to the successful completion of the strategy for showing the existence of a solution of (2.1).

DEFINITION 2.6. Let $(X, \|\cdot\|)$ be a Banach space and $F : X \to \mathbb{R}$ be given.

(a) F is said to be strongly lower semicontinuous at the point x_0 if for any sequence x_n such that $\|x_n - x_0\| \to 0$, we have
$$\liminf_{n\to\infty} F(x_n) \geq F(x_0). \tag{2.3}$$

(b) F is said to be weakly lower semicontinuous at x_0 if for any sequence (x_n) such that $x_n \rightharpoonup x_0$, we have
$$\liminf_{n\to\infty} F(x_n) \geq F(x_0). \tag{2.4}$$

(c) F is said to be strongly (resp. weakly) upper semicontinuous if $-F$ is strongly(resp. weakly) lower semicontinuous.

In general, it is very cumbersome to prove that a functional is weakly lower semicontinuous. However, a simpler property of F, sufficient to achieving lower semicontinuity, is convexity which we now define.

DEFINITION 2.7. Let $F : X \to \mathbb{R}$ be a functional defined from X into $\mathbb{R}$

(a) The functional F is said to be convex on X if F is said to be convex if

$$F(tx + (1-t)y) \leq tF(x) + (1-t)F(y) \quad \forall\, x, y \in X \text{ and } \forall\, t \in [0,\, 1].$$

We say that F is strictly convex if for any pair of distinct elements x, y in X,

$$F(tx + (1-t)y) < tF(x) + (1-t)F(y) \quad \forall\, t \in (0,\, 1).$$

(b) F will be termed (strictly) concave if $-F$ is (strictly) convex.

We now present a result that establishes the equivalence between strong and weak semicontinuity in the class of convex functionals.

THEOREM 2.8. *Suppose that F is convex on X. Then F is strongly lower semicontinuous if and only if F is weakly lower semicontinuous.*

The next result will be used later to prove the existence and uniqueness of the solution of the model that we study in this dissertation.

PROPOSITION 2.9. *Suppose that X is a reflexive Banach space and the functional F is*

convex, lower semicontinuous, and proper, (A1)

and

$$\textit{coercive, i.e} \lim_{\|x\| \to \infty} F(x) = \infty. \tag{A2}$$

Then, the minimization problem (2.1) *has at least one solution. The solution is unique if* F *is strictly convex.*

Finally, we would like to characterize the solution of problem (2.1) when F is the sum of two functionals one of which is Gâteaux differentiable. Let us begin by clarifying what is meant by Gâteaux differentiable.

DEFINITION 2.10. A functional $F : X \to \mathbb{R}$ is said to be Gâteaux differentiable at a point $u \in X$ if there exists $F'(u) \in X'$ such that

$$\lim_{t \to 0^+} \frac{F(u+tv) - F(u)}{t} = \langle F'(u), v \rangle_{X',X} \quad \forall v \in X. \tag{2.5}$$

F is Gâteaux differentiable if it is Gâteaux differentiable at every point u in its domain.

The following result is a characterization of convexity for Gâteaux differentiable functions.

PROPOSITION 2.11. *Suppose that* $F : X \to \mathbb{R}$ *is Gâteaux differentiable on a convex subset* $\mathcal{C}$ *of* X*. Then the following are equivalent*

$$F \text{ is convex on } \mathcal{C}, \tag{2.6}$$

$$F(v) \geq F(u) + \langle F'(u), v-u \rangle \quad \forall\, u, v \in \mathcal{C}. \tag{2.7}$$

PROOF. Suppose that F is convex on $\mathcal{C}$. Then

$$F(v) \geq F(u) + \frac{F(u + t(v-u)) - F(u)}{t}, \quad \forall\, t \in (0,1),\ \forall\, u, v \in \mathcal{C}.$$

Since F is Gâteaux differentiable on $\mathcal{C}$, letting $t \to 0^+$ in the latter inequality, we obtain (2.7) outright.

Conversely, assume that (2.7) holds. Then for any pair $u, v \in \mathcal{C}$

$$F(u) \geq F((1-t)u + tv) - t\langle F'((1-t)u + tv), v-u \rangle, \quad \forall\, t \in [0,1], \tag{2.8}$$

and

$$F(v) \geq F((1-t)u + tv) + (1-t)\langle F'((1-t)u + tv), v - u\rangle, \quad \forall\, t \in [0,1]. \tag{2.9}$$

Multiplying (2.8) by $1-t$ and (2.9) by t, and adding the resulting inequalities, we get

$$(1-t)F(u) + tF(v) \geq F((1-t)u + tv), \quad \forall\, t \in [0,1].$$

Since the pair $u, v \in \mathcal{C}$ was arbitrary, we infer that F is convex over $\mathcal{C}$. □

PROPOSITION 2.12. *Let $\mathcal{C}$ be a closed convex subset of X. Suppose that F_1 and F_2 are convex lower semicontinuous, and F_1 is Gâteaux differentiable. Then, an element $u \in \mathcal{C}$ is a global minimizer of $F_1 + F_2$ over $\mathcal{C}$ if and only if*

$$\langle F_1'(u), v - u\rangle + F_2(v) - F_2(u) \geq 0 \quad \forall\, v \in \mathcal{C}. \tag{2.10}$$

PROOF. Suppose that $u \in \mathcal{C}$ is a minimizer of $F_1 + F_2$ over $\mathcal{C}$. Then, we have

$$F_1(u) + F_2(u) \leq F_1(u + t(v-u)) + F_2(u + t(v-u)), \quad \forall\, t \in [0,1], \quad \forall\, v \in \mathcal{C}$$

and by the convexity of F_2

$$F_1(u) + F_2(u) \leq F_1(u + t(v-u)) + (1-t)F_2(u) + tF_2(v), \quad \forall\, t \in [0,1], \quad \forall\, v \in \mathcal{C},$$

so that

$$\frac{F_1(u + t(v-u)) - F_1(u)}{t} + F_2(v) - F_2(u) \geq 0 \quad \forall\, t \in (0,1), \quad \forall\, v \in \mathcal{C}.$$

Since F_1 is Gâteaux differentiable at u, taking the limit as $t \to 0^+$ in the latter inequality yields precisely (2.10).

Conversely, suppose that $u \in \mathcal{C}$ satisfies (2.10). Then since F_1 is differentiable and convex, it follows from Proposition 2.11 that

$$F_1(v) - F_1(u) - \langle F_1'(u), v - u\rangle \geq 0, \quad \forall\, v \in \mathcal{C}.$$

Adding the latter inequality to (2.10) yields

$$F_1(v) + F_2(v) \geq F_1(u) + F_2(u), \quad \forall\, v \in \mathcal{C};$$

hence u is a minimizer of $F_1(v) + F_2(v)$ over $\mathcal{C}$ and the proof is complete. □

2.2 FUNCTIONS OF BOUNDED VARIATION

In this section, we collect relevant results on functions of bounded variation. Throughout the section, Ω will stand for an open connected subset of $\mathbb{R}^N$ $(N \geq 1)$, unless otherwise noted. All the results given below are found in the two books [46] and [8].

2.2.1 Total Variation

We begin with the definition and properties of the concept of total variation. Let u be a locally integrable function on Ω, *i.e*, u is integrable on every compact subset of Ω.

DEFINITION 2.13. The total variation of u over Ω is given by

$$J(u) := \sup \left\{ -\int_\Omega u \operatorname{div}(\boldsymbol{g}) dx : \boldsymbol{g} \in C_c^1(\Omega, \mathbb{R}^N) \text{ and } |\boldsymbol{g}(x)| \leq 1 \ \ \forall\, x \in \Omega \right\}, \tag{2.11}$$

where $\operatorname{div}(\boldsymbol{g}) = \displaystyle\sum_{i=1}^N \frac{\partial g_i}{\partial x_i}$, and $C_c^1(\Omega, \mathbb{R}^N)$ is the space of continuously differentiable functions $u : \Omega \to \mathbb{R}^N$ with compact support in Ω.

EXAMPLE 2.14. We give two examples of functions with finite total variation.

(a) Let $W^{1,1}(\Omega)$ be the space of weakly differentiable functions $u \in L^1(\Omega)$ such that $\dfrac{\partial u}{\partial x_i} \in L^1(\Omega)$ for all $1 \leq i \leq N$. Using Gauss-Green Theorem and a regularization argument, one can show that if $u \in W^{1,1}(\Omega)$, then

$$J(u) = \int_\Omega |\nabla u| dx.$$

However, $W^{1,1}(\Omega)$ is a proper subset of the set of functions of bounded variations as will be illustrated by the next example.

(b) Let E be a ball of radius ρ such that $E \subset \Omega$, and φ_E the characteristic function of E. Given $\boldsymbol{g} \in C_c^1(\Omega, \mathbb{R}^N)$, by Gauss-Green theorem we have

$$-\int_\Omega \varphi_E \operatorname{div}(\boldsymbol{g}) dx = -\int_E \varphi_E \operatorname{div}(\boldsymbol{g}) dx = \int_{\partial E \cap \Omega} \boldsymbol{g} \cdot \boldsymbol{\nu} d\mathcal{H}^{N-1},$$

where $\boldsymbol{\nu}$ is the outer unit normal vector field to ∂E, and $\mathcal{H}^{N-1}$ is the $N-1$-Hausdorff measure on Ω, see for example [8, section 2.8, p. 72] for a definition of Hausdorff measures. Thus,

$$J(\varphi_E) \leq \mathcal{H}^{N-1}(\partial E \cap \Omega) \leq \mathcal{H}^{N-1}(\partial E) < \infty.$$

We remark that φ_E does not belong to $W^{1,1}(\Omega)$.

DEFINITION 2.15. A function $u \in L^1(\Omega)$ is said to be of bounded variation (BV) if $J(u)$ is finite. We shall denote by $BV(\Omega)$ the subset of $L^1(\Omega)$ made of functions of bounded variation.

We now highlight some properties of the total variation that make functions of bounded variation a viable model for images.

PROPOSITION 2.16. *The total variation functional* $J : L^1(\Omega) \to [0, +\infty]$ *satisfies the following properties:*

(a) $J(tu) = tJ(u)$ *for any* $t \in [0, \infty)$ *and any* $u \in BV(\Omega)$*;*

(b) $J(tu + (1-t)v) \leq tJ(u) + (1-t)J(v)$ *for any* $t \in [0, 1]$ *and any* $u, v \in L^1(\Omega)$*;*

(c) if (u_n) *is a sequence which converges in* $L^1(\Omega)$ *to* u*, then*

$$J(u) \leq \liminf_{n \to \infty} J(u_n). \tag{2.12}$$

PROOF. Establishing properties (a) and (b) is a straightforward computational exercise. Property (c) on the other hand follows from the Dominated Convergence Theorem and the definition of J. In fact, let $\boldsymbol{g} \in C_c^1(\Omega, \mathbb{R}^N)$ be such that $|\boldsymbol{g}(x)| \leq 1$ for any $x \in \Omega$.

Then by the dominated convergence theorem, we have

$$\int_\Omega u \operatorname{div}(\boldsymbol{g})dx = \lim_{n\to\infty} \int_\Omega u_n \operatorname{div}(\boldsymbol{g})dx \leq \liminf_{n\to\infty} J(u_n).$$

Now taking the supremum on the latter inequality over all such $\boldsymbol{g}$, we obtain (2.12) and the proof is complete. □

REMARK 2.17. The equality in (2.12) above need not be achieved as illustrated by the sequence

$$u_n(x) = \frac{1}{n}\sin(nx),\ x \in \Omega = (0, 2\pi).$$

It is easily seen using the dominated convergence theorem that $u_n \xrightarrow{L^1(\Omega)} u = 0$ as $n \to \infty$. However, since the u_n are smooth, we have

$$J(u_n) = \int_0^{2\pi} |u_n'(x)|dx = \int_0^{2\pi} |\cos(nx)|dx = 4\int_0^{\pi/2} \cos(u)du = 4.$$

So, $\liminf_n J(u_n) = 4 > 0 = J(u)$.

We now explain how one extends the total variation of a function $u \in BV(\Omega)$ into a finite positive Borel measure over Ω. Let $u \in BV(\Omega)$ be fixed. The total variation of u with respect to an open subset $A \subset \Omega$ is naturally given by

$$|Du|(A) = \sup\left\{\int_\Omega u \operatorname{div}(\varphi)\, dx : \varphi \in C_c^1(A, \mathbb{R}^n),\ |\varphi(x)| \leq 1,\ \forall\, x \in A\right\}. \tag{2.13}$$

Furthermore, if B is a general Borel subset of Ω, then we define the total variation of u over B by

$$|Du|(B) := \inf\{|Du|(O)\colon O \supset B \text{ and } O \text{ open}\}. \tag{2.14}$$

It can be shown that under the definition 2.14, $|Du|$ is a positive Borel measure on Ω which will be called *the total variation measure of* u. Consequently, by additivity of measures, the following identity holds for all Borel subset $K \subseteq \Omega$

$$|Du|(\Omega) = |Du|(\Omega \setminus K) + |Du|(K). \tag{2.15}$$

Although, generally, we cannot expect equality in (2.12), we can provide it in some cases as demonstrated by the following lemma. Let η be the radially symmetric function defined by

$$\eta(x) = \begin{cases} c\exp\left(\dfrac{1}{|x|^2-1}\right), & |x|<1, \\ 0, & \text{otherwise}, \end{cases} \tag{2.16}$$

where the constant c is chosen such that $\displaystyle\int_{\mathbb{R}^2} \eta(x)dx = 1$. Let $\left\{\eta_\varepsilon(x) = \varepsilon^{-2}\eta\left(\dfrac{x}{\varepsilon}\right) : \epsilon > 0\right\}$ be the corresponding family of mollifiers. We observe that for each $\epsilon > 0$ the function η_ϵ is supported on the $\{x \in \mathbb{R}^2 \colon |x| \le \epsilon\}$ and $\displaystyle\int_{\mathbb{R}^2} \eta_\epsilon(x)dx = 1$. Hereafter, we shall refer to the family of mollifiers $\{\eta_\epsilon \colon \epsilon > 0\}$ as the standard family of mollifiers.

LEMMA 2.18. *Suppose $u \in BV(\Omega)$. If $A \subset\subset \Omega$ is an open set such that*

$$\int_{\partial A} |Du| = 0, \tag{2.17}$$

then

$$\int_A |Du| = \lim_{\epsilon\to 0} \int_A |D(u*\eta_\epsilon)|, \tag{2.18}$$

PROOF. Since $u*\eta_\epsilon \xrightarrow{L^1(\Omega)} u$ as $\epsilon \searrow 0$, we already have by lower semicontinuity of the total variation, the inequality

$$\int_A |Du| \le \liminf_{\epsilon\to 0} \int_A |D(u*\eta_\epsilon)|;$$

so it remains to show that

$$\limsup_{\epsilon\to 0} |D(u*\eta_\epsilon)| \le \int_A |Du|.$$

Let $\boldsymbol{g} \in C_c^1(A;\mathbb{R}^2)$ be such that $\boldsymbol{g}(x)| \le 1$ for all $x \in A$ and $0 < \epsilon < \operatorname{dist}(A, \partial\Omega)$. Then, by Fubini Theorem, we have

$$\int_\Omega (u*\eta_\epsilon)\operatorname{div}(\boldsymbol{g})dx = \int_\Omega u(\operatorname{div}(\boldsymbol{g})*\eta_\epsilon)dx = \int_\Omega u\operatorname{div}(\boldsymbol{g}*\boldsymbol{\eta_\epsilon})dx.$$

Now a simple computation verifies that

$$|\boldsymbol{g}(x)| \leq 1 \; \forall x \in A \Rightarrow |g * \eta_\epsilon(x)| \leq 1 \; \forall x \in A_\epsilon = \{x \in \Omega \colon \operatorname{dist}(x, A) \leq \epsilon\},$$

and

$$\operatorname{spt}(\boldsymbol{g}) \subset A \Rightarrow \operatorname{spt}(g * \eta_\epsilon) \subset A_\epsilon$$

so that

$$\int_\Omega (u * \eta_\epsilon) \operatorname{div}(\boldsymbol{g}) dx \leq \int_{A_\epsilon} |Du|.$$

Taking the supremum with respect to $\boldsymbol{g} \in C_c^1(A; \mathbb{R}^2)$ in the inequality above, we get

$$\int_A (u * \eta_\epsilon) \operatorname{div}(\boldsymbol{g}) dx \leq \int_{A_\epsilon} |Du|.$$

Thus

$$\limsup_{\epsilon \to 0} |D(u * \eta_\epsilon)| \leq \lim_{\epsilon \to 0} \int_{A_\epsilon} |Du| = \int_{\bar{A}} |Du| = \int_A |Du| + \int_{\partial A} |Du|,$$

from which it follows under assumption (2.17) that

$$\limsup_{\epsilon \to 0} |D(u * \eta_\epsilon)| \leq \int_A |Du|.$$

□

Finally, we recall an alternate formula for the total variation of u over Ω that shall be instrumental in establishing a maximum principle like property for the minimizer of the ROF functional.

THEOREM 2.19 (Coarea formula). *Let a function $u \in BV(\Omega)$ be given and define the sublevel set of u at level $t \in \mathbb{R}$ by*

$$U_t := \{x \in \Omega \colon u(x) < t\}. \tag{2.19}$$

Then, the following identity holds

$$J(u) = \int_{-\infty}^{\infty} J(\mathbb{1}_{U_t}) dt. \tag{2.20}$$

2.2.2 PROPERTIES OF FUNCTIONS OF BOUNDED VARIATION

We begin with a direct consequence of property (c) in Proposition 2.16. In fact is not hard to show that the space of functions of bounded variation, $BV(\Omega)$, is Banach space under the norm

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

A property of functions of bounded variation that is central to our contributions in this dissertation is the existence of an extension operator over $BV(\Omega)$ that does not turn the boundary of Ω into a singular hypersurface under the total variation measure. The next result may be used to construct such an extension for rectangular and polygonal domains.

THEOREM 2.20 (Trace on the boundary, [46, Theorem 2.10]). *Let $\Omega \subset \mathbb{R}^2$ be a bounded Lipschitz domain. Then for any $u \in BV(\Omega)$ there exists a function $\gamma_0(u) \in L^1(\partial\Omega)$ such that for $\mathcal{H}^1$-almost all $x \in \partial\Omega$,*

$$\lim_{r\to 0} \frac{1}{r^2} \int_{\{z\in\Omega:\ |z-x|<r\}} |u(z) - \gamma_0(u)(x)| dz = 0. \tag{2.21}$$

Furthermore, for every $\boldsymbol{g} \in C^1(\bar{\Omega}, \mathbb{R}^2)$

$$\int_\Omega u \operatorname{div}(\boldsymbol{g}) dx = -\int_\Omega \langle \boldsymbol{g}, Du\rangle + \int_{\partial\Omega} \gamma_0(u)\langle \boldsymbol{g}, \boldsymbol{\nu}\rangle d\mathcal{H}^1, \tag{2.22}$$

where $\boldsymbol{\nu}$ is the unit outer normal to $\partial\Omega$, and $\mathcal{H}^1$ is the $1-$dimensional Hausdorff measure on $\mathbb{R}^2$.

The trace $\gamma_0(u)$ of a function $u \in BV(\Omega)$ is uniquely defined by the equation (2.21) and for $\mathcal{H}^1$-almost every $x \in \partial\Omega$

$$\gamma_0(u)(x) = \lim_{r\to 0} \frac{1}{|C(x,r)|} \int_{C(x,r)} u(z) dz, \tag{2.23}$$

where $C(x,r) = \{z \in \Omega\colon |z-x| < r\}$ and $|C(x,r)|$ is the Lebesgue measure of $C(x,r)$.

The next result allows us to define extensions beyond Ω of functions of bounded variation on Ω. We will use it later in our work to define an extension via successive reflections of a function of bounded variation without creating new oscillations at the boundary of Ω.

LEMMA 2.21 (Pasting Lemma [46, Proposition 2.8]). *Let O be an open set such that $\Omega \subset\subset O$. Let $u_1 \in BV(\Omega)$, and $u_2 \in BV(O \setminus \bar{\Omega})$ be given. Then the function $u : O \to \mathbb{R}$ defined by*

$$u(x) = \begin{cases} u_1(x), & x \in \Omega \\ u_2(x), & x \notin \bar{\Omega} \end{cases}$$

is an element of $BV(O)$ and

$$|Du|(O) = |Du_1|(\Omega) + |Du_2|(O \setminus \bar{\Omega}) + \int_{\partial\Omega} |\gamma_0(u_1) - \gamma_0(u_2)| d\mathcal{H}^1. \tag{2.24}$$

Moreover, the total variation of u over the boundary of Ω is given by

$$|Du|(\partial\Omega) = \int_{\partial\Omega} |\gamma_0(u_1) - \gamma_0(u_2)| d\mathcal{H}^1. \tag{2.25}$$

We close this section with a result that shows that BV functions are well approximated by smooth functions and the statement of the Sobolev inequality for BV functions.

THEOREM 2.22. *Let $u \in BV(\Omega)$ be given. Then, there exists a sequence $\{u_n\}_n \in C^\infty(\Omega) \cap W^{1,1}(\Omega)$ such that $u_n \xrightarrow{L^1(\Omega)} u$ and $J(u_n) = \int_\Omega |\nabla u_n| dx \to J(u) = \int_\Omega |Du|$.*

PROOF. When $\Omega = \mathbb{R}^2$, the proof is simple and uses the standard smoothing through convolution with mollifiers argument. For $\Omega \neq \mathbb{R}^2$, the proof is classical and can be found in the monographs [8, 46]. □

For $u \in L^1(\Omega)$, we denote the average value of u over Ω by

$$u_\Omega = \frac{1}{|\Omega|} \int_\Omega u(x)\, dx. \tag{2.26}$$

THEOREM 2.23 (Sobolev's Inequality). *Suppose that Ω is a bounded Lipschitz domain in $\mathbb{R}^2$. Then there exists a constant C depending only on Ω such that*

$$\|u - u_\Omega\|_{L^2(\Omega)} \leq C \int_\Omega |Du|, \quad \forall\, u \in BV(\Omega). \tag{2.27}$$

If $\Omega = \mathbb{R}^2$, then there exists $C > 0$ such that

$$\|u\|_{L^2(\mathbb{R}^2)} \leq C \int_{\mathbb{R}^2} |Du|, \quad \forall u \in BV(\mathbb{R}^2). \tag{2.28}$$

2.2.3 SETS OF FINITE PERIMETER

Let $E \subset \mathbb{R}^2$ be a Borel subset of $\mathbb{R}^2$, and Ω a domain of $\mathbb{R}^2$. We say that E is of finite perimeter in Ω if its characteristic function

$$\mathbb{1}_E(x) = \begin{cases} 1 & x \in E, \\ 0 & x \in \Omega \setminus E, \end{cases}$$

is of bounded variation on Ω. The total variation $J(\mathbb{1}_E)$ is the perimeter of E in Ω and shall be denoted by $\mathrm{Per}(E, \Omega)$. If $\Omega = \mathbb{R}^2$, then we shall simply write $\mathrm{Per}(E)$.

Sets of finite perimeter play a crucial role in the analysis of the total variation based model that we introduce in the next section. We highlight some of the properties of these sets.

PROPOSITION 2.24. *Suppose that Ω is an open domain in $\mathbb{R}^2$. The following are true:*

(a) $\mathrm{Per}(E, \Omega) = \mathrm{Per}(F, \Omega)$ *whenever the set $\Omega \cap (E \Delta F)$ is Lebesgue negligible.*

(b) $\mathrm{Per}(E, \Omega) = \mathrm{Per}(E, \mathbb{R}^2 \setminus \Omega)$.

(c) $\mathrm{Per}(E \cup F, \Omega) + \mathrm{Per}(E \cap F, \Omega) \leq \mathrm{Per}(E, \Omega) + \mathrm{Per}(F, \Omega)$.

PROOF. Properties (a) and (b) follow from the definition of the total variation. The proof of property (c) is a consequence of the approximation theorem 2.22 and the lower semicontinuity of the total variation. The details may be found in [8, 33, 46]. □

The following result is a direct consequence of the Sobolev's inequality for sets of finite perimeter.

THEOREM 2.25 (Isoperimetric inequality [8]). *If E is a set of finite perimeter in $\mathbb{R}^2$, then either E or $\mathbb{R}^2 \setminus E$ has finite Lebesgue measure and*

$$\min(|E|, |\mathbb{R}^2 \setminus E|) \leq \frac{\operatorname{Per}(E)^2}{4\pi}. \tag{2.29}$$

2.3 THE IMAGE DENOISING PROBLEM

Any attempt to reconstruct an image from degraded measurements must first account for the source of the degradation, and choose a model of the degradation that is as close to reality as possible. A common model of image degradation is the following. Let $u : \Omega \to \mathbb{R}^N (N \geq 1)$ be the perfect description – as sensed by a healthy human eye – of a natural scene, and f the same scene as captured by an imaging device. It is generally assumed that

$$f = Au + n, \tag{2.30}$$

where n is a realization of the noise, and A is a deterministic acquisition procedure that may also contribute to degrading the image.

In general, we do not have the exact model of the noise, and we do not have a complete understanding of the deterministic degrading process A. Thus, the task of recovering u exactly from f is a daunting one; we are only able to carry out an approximation of u, conditioned on a-priori models for A and the noise n. The noise n is understood as a perturbation causing spurious and unstructured oscillations in the measurements f.

The goal of the denoising problem is to remove as much of these oscillations as possible while preserving key features of the available measurements f that are discernible to the human eye. In the special case where A is the identity operator, we get the "pure" denoising problem; meaning that the only degradation contributed in f is the random noise and the acquisition procedure is flawless.

2.3.1 THE ROF DENOISING MODEL

Rudin, Osher and Fatemi [71] proposed to recover an approximation of the ideal image u, from its corrupt measurements f, by solving the minimization problem

$$\begin{aligned}&\text{Find } u \in L^2(\Omega) \text{ such that}\\ &u \in \underset{v\in L^2(\Omega)}{\arg\min}\left\{E_\lambda^f(v) := \lambda J(v) + \frac{1}{2}\int_\Omega |v-f|^2 dx\right\},\end{aligned} \tag{ROF}$$

where $\lambda > 0$ is the threshold of the preserved scale of oscillations, and $J(v) := |Dv|(\Omega)$, the total variation of v on Ω, quantifies the oscillations in v on Ω .

We note that for any $\lambda \in \mathbb{R}_+$, the objective functional $E_\lambda^f(u)$ is coercive, lower semicontinuous, and strictly convex. Therefore, the existence and uniqueness of the solution to (ROF) are guaranteed by Proposition 2.9. Moreover, since $E_\lambda^f(u)$ is the sum of two convex and lower semicontinuous functionals one of which is differentiable, it follows from Proposition 2.12 that the solution u_λ^f of (ROF) is characterized by

$$\int_\Omega (u_\lambda^f - f)(v - u_\lambda^f)dx + \lambda(J(v) - J(u_\lambda^f)) \geq 0, \quad \forall\, v \in L^2(\Omega). \tag{2.31}$$

Furthermore, since $J(u)$ is positively 1-homogeneous, we infer from the above characterization of u_λ^f that

$$J(u_\lambda^f) = \frac{1}{\lambda}\int_\Omega (f - u_\lambda^f)u_\lambda^f\, dx. \tag{2.32}$$

Under the assumption that the variance, σ^2, of the noise contributed in f is such that $|\Omega|\sigma^2 \leq \displaystyle\int_\Omega |f - f_\Omega|^2 dx$, where f_Ω is the average value of f over Ω, Chambolle and Lions [35] showed that there exists a Lagrange multiplier λ^* for which the objective function in (ROF) with scale $\dfrac{1}{2\lambda^*}$ is the Lagrangian functional associated to the constrained minimization problem

$$\underset{u\in L^2(\Omega)}{\arg\min}\, J(u) \text{ subject to } \int_\Omega |u-f|^2 dx \leq |\Omega|\sigma^2. \tag{TVD}$$

Consequently, assuming that we can estimate the Lagrange multiplier λ^*, the ROF model with parameter λ^* achieves a cleanup of the image f by decomposing it into

$f = u + v$ such that the L^2-energy of v is proportional to the variance of the additive noise in the measurements f.

The question of estimating the Lagrange multiplier λ, *i.e*,

$$\text{Find } \lambda^* \text{ such that } \|u^f_{\lambda^*} - f\|_2^2 = |\Omega|\sigma^2,$$

requires a good understanding of the role of λ in the ROF model. For a fixed function $f \in L^2(\Omega)$, we define the function $\Sigma : \mathbb{R}_+ \to \mathbb{R}_+$ by $\Sigma(\lambda) := \|u^f_\lambda - f\|_2$, where u^f_λ is the solution of (ROF) with parameter λ. The following result giving an insight on the role played by λ was first proved in [35, Lemma 2.3], and was completed by Chambolle in [31].

THEOREM 2.26 (Chambolle and Lions [35], Chambolle [31]). *The function Σ maps $\mathbb{R}_+$ into $[0, \|f - f_\Omega\|_2]$, is continuous, and monotone non-decreasing. Furthermore, the function $\lambda \mapsto \Sigma(\lambda)/\lambda$ is monotone non-increasing.*

Several scholars have studied the Lagrange multiplier estimation problem. For example, Chambolle [31] exploited Theorem 2.26 above in designing an iterative method that simultaneously solves the total variation denoising problem (TVD) and computes the Lagrange multiplier λ^*. Aujol and Gilboa [15] proposed a signal-to-noise-ratio (SNR) parameter selection method for approximating λ^* as a value of λ that maximizes to SNR of the image recovered with the ROF model.

The ROF model is very efficient on images dominated by geometric structures, while its performance decays on images that contains significant oscillatory components such as textures and fine structures. This fact was justified in theory by Nikolova, [64]. She proved that total variation based models favor flat regions, that is, the image recovered by this model will have patches where it looks like a piecewise constant function. This is known in the image processing community as the staircasing effect.

While the ROF model is generally efficient at detecting the geometric structure of images, it uses limited amount of information about the noise (first and second moments only) itself. Moreover, even in the simplest case of piecewise constant images, we observe a reduction of contrast in the recovered image. As a consequence, it may not be suitable for most noise priors. For example, Nikolova [65] showed that by replacing the L^2-norm in the model by the L^1-norm,

$$\underset{u \in L^1(\Omega)}{\arg\min}\ \lambda J(u) + |u - f|_{L^1} \tag{TVL$_1$}$$

one obtains a new variational model that is superior to the ROF model for images corrupted with impulse noise and outliers. A thorough analysis of the model (TVL$_1$) is done by Chan and Esedoḡlu, [37]; they prove among other things that this model is capable of recovering the characteristic function of the disc unlike the ROF model (see section 2.3.3), without loss of contrast.

REMARK 2.27. The fundamental assumption of the ROF model is that images are functions of bounded variation. However, total variation based models perform poorly when handed an image with textures and fine structures. The scholars in [47] showed experimentally that natural images are overwhelmingly not of bounded variation. Nonetheless, total variation image enhancement model are still popular in the community and remain competitive with filtering methods.

2.3.2 SOME PROPERTIES OF THE ROF MODEL

In this section, we review some of the properties of the minimization problem (ROF). We start with two important properties of this model that are the foundation of the convergence analysis carried in this dissertation.

THEOREM 2.28. *Let $u_\lambda^f \in BV(\Omega)$ be the minimizer of the ROF functional $E_\lambda^f(u)$. Then, for any $v \in BV(\Omega)$, there holds*

$$\left\| v - u_\lambda^f \right\|_{L^2}^2 \leq 2 \left(E_\lambda^f(v) - E_\lambda^f(u_\lambda^f) \right). \tag{2.33}$$

Moreover, if u_λ^g is the minimizer of $E_\lambda^g(u)$, then

$$\|u_\lambda^f - u_\lambda^g\|_{L^2} \leq \|f - g\|_{L^2}. \tag{2.34}$$

PROOF. Let $v \in BV(\Omega)$ be fixed and u_λ^f be the minimizer of $E_\lambda^f(u)$. Then

$$\begin{aligned} E_\lambda^f(v) - E_\lambda^f(u_\lambda^f) &= \lambda(J(v) - J(u_\lambda^f)) + \frac{1}{2}\left(\|v - f\|_{L^2}^2 - \|u_\lambda^f - f\|_{L^2}^2\right) \\ &= \underbrace{\lambda(J(v) - J(u_\lambda^f)) + \int_\Omega (v - u_\lambda^f)(u_\lambda^f - f)\,dx}_{\geq 0 \text{ by (2.31)}} + \frac{1}{2}\|v - u_\lambda^f\|_{L^2}^2 \\ &\geq \frac{1}{2}\|v - u_\lambda^f\|_{L^2}^2. \end{aligned}$$

Since v was arbitrarily chosen, we obtain (2.33).

On the other hand, if u_λ^g is the minimizer of $E_\lambda^g(u)$, then by the characterizing equation (2.31), we have

$$\lambda(J(u_\lambda^f) - J(u_\lambda^g)) \leq \int_\Omega (u_\lambda^f - f)(u_\lambda^g - u_\lambda^f)\,dx,$$

so that

$$\begin{aligned} 2\left(E_\lambda^g(u_\lambda^f) - E_\lambda^g(u_\lambda^g)\right) &\leq 2\int_\Omega (u_\lambda^f - f)(u_\lambda^g - u_\lambda^f)\,dx + \int_\Omega (u_\lambda^f - u_\lambda^g)(u_\lambda^f + u_\lambda^g - 2g)\,dx \\ &= 2\int_\Omega (u_\lambda^g - u_\lambda^f)(g - f)\,dx - \int_\Omega (u_\lambda^g - u_\lambda^f)^2\,dx \\ &\leq \|f - g\|_{L^2}^2, \end{aligned}$$

where we have used the inequality $2ab \leq a^2 + b^2$. Hence, using (2.33), we obtain $\left\|u_\lambda^f - u_\lambda^g\right\|_{L^2} \leq \|f - g\|_{L^2}$ and the proof is complete. □

We now prove a maximum principle result for the total variation based image denoising model (ROF).

THEOREM 2.29 (Maximum principle). *Suppose that $f \in L^\infty(\Omega)$ and let u_λ^f be the minimizer of $E_\lambda^f(u)$ on $BV(\Omega)$. Then, $u \in L^\infty(\Omega)$ and $\|u_\lambda^f\|_\infty \leq \|f\|_\infty$. More precisely,*

$$\inf_{x\in\Omega} f(x) \leq u(x) \leq \sup_{x\in\Omega} f(x) \quad \text{for a.e } x \in \Omega. \tag{2.35}$$

PROOF. Let $u \in BV(\Omega)$ be fixed. Let $M = \|f\|_\infty$ and set

$$u^M(x) = \begin{cases} u(x), & |u(x)| \le M \\ \text{sign}(u(x))M, & |u(x)| > M. \end{cases}$$

The sub-level sets of the function u^M are as follows

$$U_t^M = \begin{cases} \Omega & t > M, \\ U_t & |t| \le M, \\ \emptyset, & t < -M, \end{cases}$$

where U_t is the sublevel set of u at level t.

On the one hand, since $|D\mathbb{1}_\Omega|(\Omega) = 0$ and $|D\mathbb{1}_\emptyset|(\Omega) = 0$, it follows from the coarea formula that

$$|Du^M|(\Omega) = \int_{-M}^{M} |D\mathbb{1}_{U_t}|(\Omega)dt \le |Du|(\Omega).$$

On the other hand, it is easy to check that $|u(x) - f(x)| \ge |f(x) - u^M(x)|$ for a.e. $x \in \Omega$. Thus, we have $E_\lambda^f(u) \ge E_\lambda^f(u^M)$ for all $u \in BV(\Omega)$ and it follows by uniqueness of the minimizer of $E_\lambda^f(u)$ that $\|u_\lambda^f\|_\infty \le M$. A similar truncation argument shows that (2.35) holds. □

2.3.3 AN EXPLICIT SOLUTION OF THE ROF MODEL

Sets of finite perimeter have been instrumental in the study of the ROF model, and have led to explicit solutions of the ROF model [3] in some special cases. More precisely, the solution of the ROF model has been characterized through its level sets as follows.

THEOREM 2.30 (Chambolle et al. [33]). *A function $u \in BV(\Omega)$ solves* (ROF) *if and only if for any $s \in \mathbb{R}$, the set $\{x \in \Omega \colon u(x) > s\}$ solves the perimeter minimization problem*

$$\arg\min_E \lambda \operatorname{Per}(E, \Omega) + \int_E s - f(x)\, dx. \qquad \text{(ROFs)}$$

PROOF. A detailed proof of the above result is found in [33]. □

The level sets characterization theorem above is used to compute the solution of the ROF model when f is the characteristic function of some convex sets [3, 4]. For example, when f is the characteristic function of a ball of radius R, B_R, the solution of (ROF) is

$$u = \max(0, 1 - 2\lambda/R)\, B_R. \tag{2.36}$$

2.3.4 REGULARITY OF THE SOLUTION OF THE ROF MODEL

Assuming that Ω is a convex domain in $\mathbb{R}^2$, the level sets characterization theorem above allows for a delicate control of the curvature of the level sets of a solution. Thus, one can establish some regularity of the solutions of (ROF). In fact, the following regularity result was proved in [29]:

THEOREM 2.31 ([29]). *If Ω is convex and f is uniformly continuous with modulus ω, then the solution, u, of* (ROF) *is uniformly continuous with modulus of continuity ω as well.*

This result is similar to the stability property established earlier in Theorem 2.28 and suggest that the inequality (2.34) may hold in the $L^\infty(\Omega)$ norm as well. In fact, a much stronger result was established in [29], where the authors proved that if the data f of the ROF model is already a bounded function of bounded variation, then the discontinuity set of the minimizer is a subset of that of f. Therefore, the ROF model does not produce new edges in the recovered image.

2.3.5 ALGORITHMIC CONSIDERATIONS

Any algorithmic consideration for the ROF model start with a viable discretization of the energy functional $E_\lambda^f(u)$. When dealing with digital images, the most popular discretization of total variation is based on the following finite difference approximation of the gradient. For a digital image $u = (u_{i,j})_{1\leq i,j\leq N}$, the gradient of u is defined by

$$(\nabla u)_{i,j} := \begin{pmatrix} (\partial_x^+ u)_{i,j} \\ (\partial_y^+ u)_{i,j} \end{pmatrix} = \begin{pmatrix} \dfrac{u_{i+1,j} - u_{i,j}}{h} \\ \dfrac{u_{i,j+1} - u_{i,j}}{h} \end{pmatrix}, \ \ 1 \leq i, j \leq N, \tag{2.37}$$

with $(\partial_x^+ u)_{i,j} = 0$ if $i = N$, and $(\partial_y^+ u)_{i,j} = 0$ if $j = N$. The corresponding discrete model, obtained using quadrature approximations of the integrals, then reads

$$\underset{u \in \mathbb{R}^{N \times N}}{\arg\min}\, \lambda \sum_{1 \leq i,j \leq N} |(\nabla u)_{i,j}|_2 + \frac{1}{2} \sum_{1 \leq i,j \leq N} |u_{i,j} - f_{i,j}|^2, \tag{2.38}$$

where $|(\nabla u)_{i,j}|_2$ is the Euclidean norm of the vector $(\nabla u)_{i,j}$.

Over the last two decades, several efficient algorithms for approximating the solution of (2.38) have been proposed and analyzed. The plethora of algorithms available in the literature for computing a solution of (2.38) are based on three main methods.

The dual method

This approach is based on the observation that problem (2.38) is equivalent to the minimization problem

$$\underset{|p|_{2,\infty} \leq 1}{\arg\min} \|\lambda \operatorname{div}(p) + f\|^2, \tag{2.39}$$

where $\operatorname{div}\colon\ Y := \mathbb{R}^{N \times N} \times \mathbb{R}^{N \times N} \to X := \mathbb{R}^{N \times N}$ is the negative adjoint of the discrete gradient ∇, and $|p|_{2,\infty} = \max\limits_{1 \leq i,j \leq N} |p_{i,j}|_2$ with $p_{i,j} \in \mathbb{R}^2$ and $p = (p_{i,j}) \in Y$.

In this framework, the primal problem (2.38) is reduce to the computation of an orthogonal projection onto a closed convex subset which is a smooth quadratic program with convex constraints. Several algorithms for computing a solution of (2.39) are found in the literature. Carter [28] studied interior point and coordinate descent algorithms. Chambolle computed the closed form of the Kuhn-Tucker vector associated to (2.39) to come up with a very efficient fixed-point algorithm [31], and formulated a projected gradient algorithm in [32]. Duval et al. [42] then followed with a direct proof of convergence of the projected gradient algorithm formulated by Chambolle [32]. An optimal first order algorithm is obtained by following Nesterov's framework in [63] or its generalization to objective functions that are sum of two functionals developed by Beck and Teboulle [19, 20].

The primal-dual method

The primal-dual method is based on the simple observation that the minimization problem (2.38) may also be written as the saddle point problem

$$\arg\min_{u\in\mathbb{R}^{N\times N}} \max_{|p|_{2,\infty}\leq 1} \langle u, -\operatorname{div}(p)\rangle + \frac{1}{2\lambda}\sum_{1\leq i,j\leq N} |u_{i,j} - f_{i,j}|^2. \tag{2.40}$$

In fact, this is the first step in deriving the dual approach above. Therefore, the minimizer of (2.38) yield a saddle point of the above objective functional. Algorithms based on the primal-dual approach aim at computing a solution of the above saddle point problem. A typical algorithm in this category alternates between a gradient descent in the primal variable u and a gradient ascent in the dual variable p. The first primal-dual algorithm for computing the solution of (2.38), using the approach described above, was proposed by Zhu et al. [76], and a variant of their algorithm has recently been studied by Chambolle and Pock [36]. A general framework for primal-dual algorithms in image processing was investigated by Esser et al. [44].

The augmented Lagrangian method

The basic idea of the augmented Lagrangian approach is that in lieu of the problem (2.38), one solves the following constrained minimization problem

$$\min_{p=\nabla u} \sum_{1\leq i,j\leq N} |p_{i,j}|_2 + \frac{1}{2\lambda}\sum_{1\leq i,j\leq N} |u_{i,j} - f_{i,j}|^2. \tag{2.41}$$

The constraint is then enforced using the augmented Lagrangian method, which consists in solving an unconstrained minimization problem with objective

$$L_\beta(u,p;\mu) = \sum_{1\leq i,j\leq N} |p_{i,j}|_2 + \frac{1}{2\lambda}\|u-f\|_2^2 + \langle \mu, p-\nabla u\rangle + \frac{\beta}{2}\|p-\nabla u\|_2^2, \tag{2.42}$$

where β is a large positive number. The algorithm is then obtained by minimizing $L_\beta(u,p,\mu)$

with respect to (u, p) and updating the Lagrange multiplier μ in a gradient ascent scheme. The corresponding algorithm and its convergence are studied by Wu and Tai [75]. Moreover, these authors showed that u is the solution of (2.38) if and only if there exists $(p, \lambda) \in Y^2$ such that

$$\forall\,(v, q, \mu) \in X \times Y \times Y, \quad L_\beta(u, p; \mu) \leq L_\beta(u, p; \lambda) \leq L_\beta(v, q; \lambda). \tag{2.43}$$

Since, we know that (2.38) has exactly one solution, it follows that L_β has at leat one saddle point on $(X \times Y) \times Y$.

REMARK 2.32. Chambolle et al. [34] have shown that one can improve the performance of total variation based imaging model by choosing a discretization of the total variation that is inherently capable of capturing the big jump in pixel values. They illustrate this by studying an upwind finite difference approximation of the total variation and observe that upwind schemes deliver sharper edges than the discrete model (2.38).

3

PIECEWISE LINEAR APPROXIMATION OF THE ROF MODEL ON RECTANGULAR DOMAINS

In this chapter, we study the approximation of the ROF model

$$u^f = \underset{u \in L^2(\Omega)}{\arg\min} \left\{ E_\lambda^f(u) := \lambda J(u) + \frac{1}{2}\int_\Omega |u-f|^2 dx \right\} \tag{ROF}$$

when Ω is rectangular. As mentioned in the introduction, the approximation of the ROF model in the continuous setting has received some attention over the last two decades, with most of the effort [35, 41] using a relaxation technique on the total variation part of the functional $E_\lambda^f(u)$ to construct a minimizing sequence. In a departure with tradition, Wang and Lucier [74] did not use a relaxation approach to construct their approximation, instead they exploited the well posedness property (see Theorem 2.28) of the ROF model and construct discretizations of $E_\lambda^f(u)$ whose minimum values converge to the minimum value of $E_\lambda^f(u)$.

We construct a continuous piecewise linear approximation of u^f as a linear interpolation of the minimizer of a suitable discrete counterpart of $E_\lambda^f(u)$, and obtain convergence when the data function f is bounded and L^2-Hölder continuous in a sense that will be specified below. The interpolatory method used in this chapter fully exploits the geometry of Ω and the arguments do not extend to general polygonal domains. The present chapter has been submitted for publication [53], jointly with Ming-Jun Lai, in SIAM Journal of Numerical Analysis.

3.1 PRELIMINARIES AND NOTATIONS

In this section we give preliminary results and introduce the notations that we shall use in this chapter. Throughout the chapter, Ω shall denote the open set $(0,1)\times(0,1)$ unless otherwise noted, and Ω_m the open set $(-m,m)\times(-m,m)$, where m is a natural number.

3.1.1 BASIC NOTATIONS

For any $\nu\in\mathbb{R}^2$, we shall denote by $\tau_\nu\Omega$ the image of the set Ω under the translation with the vector η, *i.e*

$$\tau_\nu\Omega := \{x+\nu\colon x\in\Omega\}.$$

For a function $u:\Omega\to\mathbb{R}$, we denote by $\tau_\nu u$ the function whose domain is $\tau_{-\nu}\Omega$ and is defined by

$$\tau_\nu u(x)=u(x+\nu),\quad x\in\tau_{-\nu}\Omega.$$

It is well known that the translation operator τ_η is a bounded linear operator from $L^p(\Omega)$ into $L^p(\tau_{-\nu}\Omega)$.

Let $h>0$ be given. The $p-$modulus of continuity of order h, of a function $u\in L^p(\Omega)$, is defined by

$$\omega(u,h)_p=\sup_{|\nu|\le h}\|\tau_\nu u-u\|_{L^p(\Omega\cap\tau_{-\nu}\Omega)}, \tag{3.1}$$

where $|\nu|$ stands for the Euclidean norm of ν.

Let $u\in L^p_{\text{loc}}(\mathbb{R}^2)$ and $A\subset\subset\mathbb{R}^2$ a relatively compact open subset of $\mathbb{R}^2$. The $p-$modulus of continuity of u of order h with respect to A, denoted $\omega(u,h)_{p,A}$, is defined by

$$\omega(u,h)_{p,A}=\omega(u\mathbb{1}_A,h)_p. \tag{3.2}$$

Let $0<\alpha\le 1$, we denote Lip$(\alpha,L^p(\Omega))$ the subspace of $L^p(\Omega)$ defined by

$$\text{Lip}(\alpha,L^p(\Omega)) := \left\{u\in L^p(\Omega):\ \sup_{0<h<1} h^{-\alpha}\omega(u,h)_p<\infty\right\}.$$

One easily checks that the space Lip$(\alpha, L^p(\Omega))$ is a Banach space when endowed with the norm

$$\|u\|_{p,\alpha} = \|u\|_{L^p} + \sup_{0<h<1} \frac{\omega(u,h)_p}{h^\alpha}.$$

3.1.2 An extension operator for BV functions

We already know that a function of bounded variation, u, on a Lipschitz domain Ω admits a compactly supported extension, $Tu \in BV(\mathbb{R}^2)$, such that the total variation of Tu over the boundary $\partial\Omega$ is zero. In this section, we detail the construction of such an extension in the special case of rectangular domains.

Let $u \in BV(\Omega)$ be given. The extension of u to all of $\mathbb{R}^2$, $\boldsymbol{X}[u]$, is defined in two steps as follows:

1) First, define $\boldsymbol{X}[u]$ on the open set $\Omega_0 := \{x \in \mathbb{R}^2 : -1 < x_1, x_2 < 3\}$ using four successive reflections of the function u across the four sides of Ω as illustrated in Figure 3.1.

2) Finally, set $\boldsymbol{X}[u] = 0$ outside the closed rectangle $\bar{\Omega}_0$.

Clearly, for any $u \in BV(\Omega)$, the function $\boldsymbol{X}[u]$ is compactly supported on the open set $\Omega_4 := (-4,4) \times (-4,4)$. We also note – thanks to Lemma A.4 – that each reflection in step 1 above yields a function of bounded variation whose total variation is exactly twice that of the function that was reflected. Thus, we have

$$|D\boldsymbol{X}[u]|(\Omega_0) = 16J(u) \text{ and } \|\boldsymbol{X}[u]\|_{L^1(\Omega_0)} = 16\|u\|_{L^1(\Omega)}. \tag{3.3}$$

Proposition 3.1. *The operator $\boldsymbol{X} : BV(\Omega) \to BV(\mathbb{R}^2)$ defined above is a bounded linear operator. Moreover, for any $u \in BV(\Omega)$*

$$\big|D\boldsymbol{X}[u]\big|(\partial\Omega) = 0, \tag{3.4}$$

$$\lim_{\varepsilon\to 0} |D(\boldsymbol{X}[u] * \eta_\varepsilon)|(\Omega) = J(u). \tag{3.5}$$

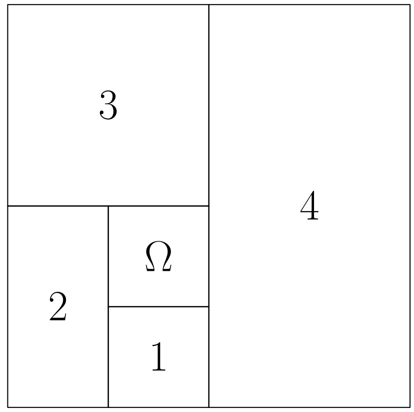

FIGURE 3.1: Schematic of the extension of u by successive reflections across the sides of Ω.

PROOF. That $\boldsymbol{X}$ is a linear operator is obvious and the boundedness follows from (3.3). It remains to show that $|D\boldsymbol{X}[u]|(\partial\Omega) = 0$.

Let $u \in BV(\Omega)$ be given, and u_0 the restriction of $\boldsymbol{X}[u]$ to $O = \mathbb{R}^2 \setminus \bar{\Omega}$. Clearly $\partial O = \partial\Omega$ and it is easy to check that the trace of $\gamma_0(u_0) = \gamma_0(u)$. Since $\boldsymbol{X}[u]$ is obtained by pasting u and u_0, it follows from the pasting Lemma 2.21 that $\big|D\boldsymbol{X}[u]\big|(\partial\Omega) = 0$. Finally, we observe that Ω is relatively compact in Ω_0 and since $|D\boldsymbol{X}[u]|(\partial\Omega) = 0$, it follows from Lemma 2.18 that $|D(\boldsymbol{X}[u] * \eta_\varepsilon)|(\Omega) \to |Du|(\Omega)$ as $\epsilon \to 0$. □

PROPOSITION 3.2. *Let $f \in L^2(\Omega)$ be fixed. Then for any $0 < h \ll 1$, we have*

$$\omega(\boldsymbol{X}[f], h)_{2,\Omega_{1,2}} \leq 4\sqrt{2}\,\omega(f,h)_2, \tag{3.6}$$

where $\Omega_{1,2} = (-1,2) \times (-1,2)$.

PROOF. Let $f \in L^2(\Omega)$ be given, and $\eta \in \mathbb{R}^2$ be fixed with $|\eta| \leq h$.

$$\begin{aligned}
\|\tau_\eta(\boldsymbol{X}[f]) - \boldsymbol{X}[f]\|^2_{L^2(\Omega_{1,2}\cap\tau_{-\eta}\Omega_{1,2})} &= \int_{\Omega_{1,2}\cap\tau_{-\eta}\Omega_{1,2}} |\boldsymbol{X}[f](x+\eta) - \boldsymbol{X}[f](x)|^2 dx \\
&\leq \sum_{\substack{-1\leq m,n\leq 2 \\ -1\leq i,j\leq 2}} \sum_{\substack{|m-i|=1 \\ |n-j|=1}} \int_{\tau_{m,n}\Omega\cap\tau_{(i,j)-\eta}\Omega} |\boldsymbol{X}[f](x+\eta) - \boldsymbol{X}[f](x)|^2 dx \\
&\leq 2 \sum_{-1\leq i,j\leq 2} \int_{\tau_{i,j}(\Omega\cap\tau_{-\eta}\Omega)} |\boldsymbol{X}[f](x+\eta) - \boldsymbol{X}[f](x)|^2 dx
\end{aligned}$$

$$\leq 32 \int_{\Omega \cap \tau_{-\eta}\Omega} |f(x+\eta) - f(x)|^2 dx = 32\|\tau_\eta f - f\|_2^2.$$

Thus, we have $\omega(\boldsymbol{X}[f], h)_{2,\Omega_{1,2}} \leq 4\sqrt{2}\,\omega(f,h)_2$. □

3.1.3 A discretization of the ROF functional.

We assume that Ω is endowed with a triangulation Δ_h constructed as follows: First, Ω is subdivided into N^2 square sub-domains of side length h; each rectangle is then divided into two triangles using the Northwest-Southeast diagonal as shown in Figure 3.2. We denote the set of vertices of the triangulation Δ_h by $\mathcal{V}_h := \{\omega_{i,j} : 1 \leq i, j \leq N\}$, and associated to each vertex $\omega_{i,j}$ a rectangle $\Omega_{i,j} := \Omega \cap \left(\omega_{i,j} + (-h/2, h/2)^2\right)$.

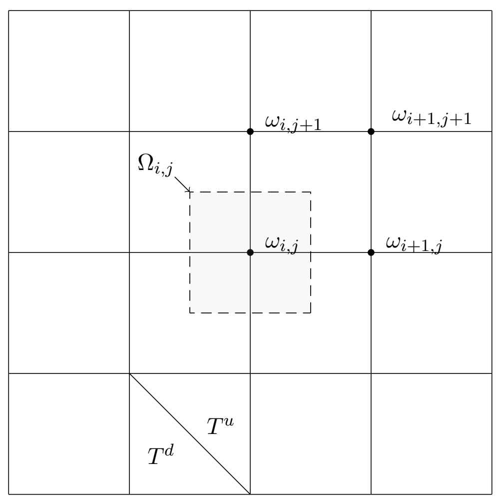

FIGURE 3.2: A type I triangulation of Ω: $T^u_{i,j}$ is the triangle with vertexes $\langle \omega_{i+1,j},\, \omega_{i+1,j+1},\, \omega_{i,j+1} \rangle$ and $T^d_{i,j}$ is the triangle with vertexes $\langle \omega_{i,j},\, \omega_{i+1,j},\, \omega_{i,j+1} \rangle$. $\Omega_{i,j}$ is used to discretize functions in $L^1(\Omega)$.

We are interested in constructing a continuous piecewise linear function on Δ_h that approximates the minimizer u^f. To this end, we first construct a discrete approximation of the functional $E^f_\lambda(u)$ on the space $\mathcal{P}_1(\Delta_h)$ of continuous piecewise linear functions on Δ_h.

Let u be a continuous piecewise linear function on Δ_h. It is well known that u is an element of the Sobolev space $W^{1,1}(\Omega)$, and u is uniquely defined by its values at the vertices of Δ_h. Therefore, the space $\mathcal{P}_1(\Delta_h)$ is a subspace of $W^{1,1}(\Omega)$ that is isomorphic to $\mathbb{R}^{N\times N}$. The total variation of an element u of $\mathcal{P}_1(\Delta_h)$ is given by

$$J_h(u) = \frac{h^2}{2} \sum_{1\le i,j\le N} |(\nabla_+ u)_{i,j}| + \frac{h^2}{2} \sum_{1\le i,j\le N} |\nabla_-(u)_{i,j})|, \tag{3.7}$$

where $|\cdot|$ is the Euclidean norm in $\mathbb{R}^2$, and the operators $\nabla_+ = (\nabla_+^x, \nabla_+^y)$ and $\nabla_- = (\nabla_-^x, \nabla_-^y)$ are defined by

$$\begin{aligned} (\nabla_+^x u)_{i,j} &= \begin{cases} 0, & \text{if } i = N \text{ or } j = N \\ \dfrac{u_{i+1,j} - u_{i,j}}{h} & \text{otherwise;} \end{cases} \\ (\nabla_+^y u)_{i,j} &= \begin{cases} 0, & \text{if } i = N \text{ or } j = N \\ \dfrac{u_{i,j+1} - u_{i,j}}{h} & \text{otherwise;} \end{cases} \end{aligned} \tag{3.8}$$

and

$$\begin{aligned} (\nabla_-^x u)_{i,j} &= \begin{cases} 0, & \text{if } i = 1 \text{ or } j = 1 \\ \dfrac{u_{i,j} - u_{i-1,j}}{h} & \text{otherwise;} \end{cases} \\ (\nabla_-^y u)_{i,j} &= \begin{cases} 0, & \text{if } i = 1 \text{ or } j = 1 \\ \dfrac{u_{i,j} - u_{i,j-1}}{h} & \text{otherwise.} \end{cases} \end{aligned} \tag{3.9}$$

Finally, assuming that a suitable discrete approximation $(f_{i,j})_{1\le i,j\le N}$ of f with respect to $\mathcal{V}_h$ is available, we approximate the energy functional $E_\lambda^f(u)$ by

$$E_h^f(u) = \lambda J_h(u) + \frac{h^2}{2} \sum_{1\le i,j\le N} |u_{i,j} - f_{i,j}|^2 . \tag{3.10}$$

REMARK 3.3. We note that since $f \in L^2(\Omega)$ can be changed arbitrarily on any set of measure zero without changing the value of the energy $E_\lambda^f(u)$, we would also want the discrete model $E_h^f(u)$ to have the same property and that will require a well defined discretization operator from $L^2(\Omega)$ into $\mathbb{R}^{N\times N}$. Consequently, we cannot use the evaluation map on $\mathcal{V}_h$ to obtain the discretization $(f_{i,j})_{1\le i,j\le N}$.

REMARK 3.4. In a departure with tradition, the discrete gradient operators ∇_+ and ∇_- are set to zero at every grid point where the finite differences are not defined in both directions. Also, we compute the discrete total variation as an average of the pointwise norm of the forward (∇_+) and backward (∇_-) gradients. We note that although Wang and Lucier [74] use an average to define a discrete total variation, their gradient operators are the classical ones.

3.2 EMBEDDING AND PROJECTION OPERATORS

In the previous section, we proposed a discrete approximation of the ROF functional that hinged on a hypothetical discretization of the data f. We now clarify how such a discretization may be obtained.

Projection operators. Short of just using the function values on $\mathcal{V}_h$ and inspired by the Lebesgue's theorem, we propose to use the local-averaging discretization operator Q_h associated to the quadrangulation $\{\Omega_{i,j} \colon 1 \leq i, j \leq N\}$. So Q_h maps $L^2(\Omega)$ into $\mathbb{R}^{N\times N}$ and is defined by

$$(Q_h f)_{i,j} := \frac{1}{|\Omega_{i,j}|} \int_{\Omega_{i,j}} f(x)dx, \;\; 1 \leq i, j \leq N \text{ and } f \in L^2(\Omega). \tag{3.11}$$

With a slight abuse of notation, Q_h shall also denote the projection of $L^2(\Omega)$ onto the space of piecewise constant function with respect to the partition $\{\Omega_{i,j} \colon 1 \leq i, j \leq N\}$ of Ω, in which case Q_h is defined by

$$Q_h f(x) = \frac{1}{|\Omega_{i,j}|} \int_{\Omega_{i,j}} f(y)dy \quad \text{ for all } x \in \Omega_{i,j}. \tag{3.12}$$

Embedding operators. One the other hand, an element $u \in \mathbb{R}^{N\times N}$ is extended into a function $C_h u \in L^p(\Omega)$, $1 \leq p \leq \infty$, as a piecewise constant function with respect to $\{\Omega_{i,j} \colon 1 \leq i, j \leq N\}$ as follows:

$$C_h u(x) = u_{i,j}, \text{ if } x \in \Omega_{i,j}. \tag{3.13}$$

We shall also need the continuous interpolation operator $P_h \colon \mathbb{R}^{N\times N} \to L^p(\Omega)$ defined by

$$P_h u(y) = \sum_{1\le i,j\le N} u_{i,j}\phi_{i,j}(y), \tag{3.14}$$

where for any $1 \le i, j \le N$, $\phi_{i,j} : \Omega \to \mathbb{R}$ is the continuous piecewise linear function such that

$$\phi_{i,j}(\omega_{i,j}) = 1, \text{ and } \phi_{i,j}(\omega) = 0,\ \omega \in \mathcal{V}_h \setminus \{\omega_{i,j}\}. \tag{3.15}$$

We note that P_h is an isomorphism between $\mathbb{R}^{N\times N}$ and $\mathcal{P}_1(\Delta_h)$. We conclude the section with a result that we will need later when studying the convergence of our approximations.

LEMMA 3.5. *Suppose that Ω is endowed with the triangulation Δ_h. Then for all $u \in \mathbb{R}^{N\times N}$, there holds*

$$\|P_h u\|_{L^2}^2 \le \|C_h u\|_{L^2}^2 + \frac{h^2}{12}(u_{1,N}^2 + u_{N,1}^2). \tag{3.16}$$

PROOF. Let $u \in \mathbb{R}^{N\times N}$ be fixed. We first observe that $P_h u$ is the continuous bivariate spline of degree 1 over the triangulation Δ_h whose coefficients in the Bernstein-Bézier representation are $\{u_{i,j},\ 1 \le i, j \le N\}$. Therefore, using the closed form formula for the inner product of splines in Bernstein-Bézier form [54, Theorem 2.34], we get

$$\int_{T_{i,j}^u} P_h u(y)^2 dy = \frac{h^2}{24}\left(u_{i+1,j}^2 + u_{i+1,j+1}^2 + u_{i,j+1}^2 + (u_{i+1,j} + u_{i+1,j+1} + u_{i,j+1})^2\right)$$

and

$$\int_{T_{i,j}^d} P_h u(y)^2 dy = \frac{h^2}{24}\left(u_{i,j+1}^2 + u_{i,j}^2 + u_{i+1,j}^2 + (u_{i,j+1} + u_{i,j} + u_{i+1,j})^2\right).$$

Consequently, by the multinomial theorem and the elementary inequality $2ab \le a^2 + b^2$, we have

$$\int_{T_{i,j}^u} P_h u(y)^2 dy \le \frac{h^2}{6}\left(u_{i+1,j}^2 + u_{i+1,j+1}^2 + u_{i,j+1}^2\right) \tag{3.17}$$

and

$$\int_{T_{i,j}^d} P_h u(y)^2 dy \le \frac{h^2}{6}\left(u_{i,j+1}^2 + u_{i,j}^2 + u_{i+1,j}^2\right). \tag{3.18}$$

Furthermore, a direct computation gives

$$\|C_h u\|_{L^2}^2 = h^2 \sum_{i,j=2}^{N-1} u_{i,j}^2 + \frac{h^2}{2} \sum_{\substack{i=2\\ j\in\{1,N\}}}^{N-1} (u_{j,i}^2 + u_{i,j}^2) + \frac{h^2}{4} \sum_{i,j\in\{1,N\}} u_{i,j}^2. \tag{3.19}$$

Thus, using (3.17) and (3.18) we obtain

$$\begin{aligned}
\|P_h u\|_{L^2}^2 &= \sum_{1\le i,j\le N-1} \int_{T_{i,j}^u} P_h u(y)^2 dy + \int_{T_{i,j}^d} P_h u(y)^2 dy \\
&\le \frac{h^2}{3} \sum_{1\le i,j<N} \left(u_{i,j+1}^2 + u_{i+1,j}^2\right) + \frac{h^2}{6} \sum_{1\le i,j<N} \left(u_{i,j}^2 + u_{i+1,j+1}^2\right) \\
&= h^2 \sum_{i,j=2}^{N-1} u_{i,j}^2 + \frac{h^2}{2} \sum_{\substack{i=2\\ j\in\{1,N\}}}^{N-1} (u_{j,i}^2 + u_{i,j}^2) + \frac{h^2}{6} \sum_{i,j\in\{1,N\}} u_{i,j}^2 \\
&\le \|C_h u\|_{L^2}^2 + \frac{h^2}{12}(u_{1,N}^2 + u_{N,1}^2).
\end{aligned}$$

□

LEMMA 3.6. *For any $f \in L^2(\Omega)$ and $0 < h \ll 1$, there holds*

$$\|f - C_h Q_h f\|_2 \le K_1\, \omega(f,h)_2 \tag{3.20}$$

and

$$\|P_h Q_h f - C_h Q_h f\|_2 \le K_2 \omega(f,h)_2, \tag{3.21}$$

where K_1 and K_2 are positive constants independent of h.

PROOF. By definition of the operators Q_h (see (3.11)) and C_h (see (3.13)), we have

$$\|f - C_h Q_h f\|_2^2 = \sum_{1\le i,j\le N} \int_{\Omega_{i,j}} \left| f(x) - \frac{1}{|\Omega_{i,j}|} \int_{\Omega_{i,j}} f(y) dy \right|^2 dx$$

$$
\begin{aligned}
&\le \sum_{1\le i,j\le N} \int_{\Omega_{i,j}} \left(\frac{1}{|\Omega_{i,j}|} \int_{\Omega_{i,j}} |f(x)-f(y)|\, dy \right)^2 dx \\
&\le \sum_{1\le i,j\le N} \int_{\Omega_{i,j}} \left(\frac{4}{h^2} \int_{\{z:\ |z|\le\sqrt{2}h\}} |\boldsymbol{X}[f](x)-\boldsymbol{X}[f](x+z)|\, dz \right)^2 dx \\
&= \int_{\Omega} \left(\frac{4}{h^2} \int_{\{z:\ |z|\le\sqrt{2}h\}} |\boldsymbol{X}[f](x)-\boldsymbol{X}[f](x+z)|\, dz \right)^2 dx \\
&\le \frac{4}{h^2} \int_{\{z:\ |z|\le\sqrt{2}h\}} \int_{\Omega} |\boldsymbol{X}[f](x)-\boldsymbol{X}[f](x+z)|^2\, dxdz,
\end{aligned}
$$

where we have used Cauchy-Schwarz inequality, and Fubini Theorem to swap the order of integration. Now, we observe that for $h \ll 1$, for any $x \in \Omega$ and any $z \in \mathbb{R}^2$ such that $|z| \le \sqrt{2}\, h$, we have $\{x,\ x+z\} \subset \Omega_{1,2}$; so that

$$
\int_{\Omega} |\boldsymbol{X}[f](x)-\boldsymbol{X}[f](x+z)|^2\, dx \le (\omega(\boldsymbol{X}[f], \sqrt{2}\, h)_{2,\Omega_{1,2}})^2.
$$

Therefore,

$$
\begin{aligned}
\|f - C_h Q_h f\|_2^2 &\le \frac{4}{h^2} \int_{\{z:\ |z|\le\sqrt{2}h\}} \int_{\Omega} |\boldsymbol{X}[f](x)-\boldsymbol{X}[f](x+z)|^2\, dxdz \\
&\le (\omega(\boldsymbol{X}[f], h)_{2,\Omega_{1,2}})^2 \frac{4}{h^2} \int_{\{z:\ |z|\le\sqrt{2}h\}} dz \\
&\le 8\pi\, (\omega(\boldsymbol{X}[f], \sqrt{2}\, h)_{2,\Omega_{1,2}})^2 \\
&\le 32\pi\, (\omega(\boldsymbol{X}[f], h)_{2,\Omega_{1,2}})^2 \text{ since } \omega(\boldsymbol{X}[f], \sqrt{2}h)_{2,\Omega_{1,2}} \le 2\, \omega(\boldsymbol{X}[f], h)_{2,\Omega_{1,2}} \\
&\le \pi\, (32\, \omega(f,h)_2)^2 \text{ by (3.6);}
\end{aligned}
$$

hence the inequality (3.20) holds with $K_1 = 32\sqrt{\pi}$.

We now prove the inequality (3.21). By definition of the operators P_h, Q_h, and C_h, we have

$$
\begin{aligned}
\|P_h Q_h f - C_h Q_h f\|_2^2 &= \sum_{1\le i,j\le N} \int_{\Omega_{i,j}} |P_h Q_h f(x) - (Q_h f)_{i,j}|^2 dx \\
&\le 2 \sum_{1\le i,j\le N} \int_{\Omega_{i,j}} \sum_{-1\le k,l\le 1} \left| \left((Q_h f)_{i+l,j+k} - (Q_h f)_{i,j}\right) \phi_{i+l,j+k}(x) \right|^2 dx
\end{aligned}
$$

$$\leq 2 \sum_{-1\leq l,k\leq 1} \sum_{\substack{1\leq i+l\leq N\\ 1\leq j+k\leq N}} h^2\big|(Q_hf)_{i+l,j+k} - (Q_hf)_{i,j}\big|^2$$

$$\leq 2 \sum_{-1\leq l,k\leq 1} \sum_{\substack{1\leq i+l\leq N\\ 1\leq j+k\leq N}} \int_{\Omega_{i,j}} |f(x+(lh,kh)) - f(x)|^2 dx$$

$$\leq 18(\omega(f,\sqrt{2}\,h)_{2,\Omega_{1,2}})^2.$$

Thus,

$$\begin{aligned}\|P_hQ_hf - C_hQ_hf\|_2 &\leq 3\sqrt{2}\omega(f,\sqrt{2}\,h)_{2,\Omega_{1,2}}\\ &\leq 6\sqrt{2}\omega(f,h)_{2,\Omega_{1,2}} \text{ since } \omega(f,\sqrt{2}\,h)_{2,\Omega_{1,2}} \leq 2\omega(f,h)_{2,\Omega_{1,2}}\\ &\leq 48\,\omega(f,h)_2 \text{ by (3.6)}.\end{aligned}$$

Hence (3.21) holds with $K_2 = 48$, and the proof is complete. □

3.3 A PIECEWISE LINEAR APPROXIMATION OF THE ROF MODEL

In this section, we construct continuous piecewise linear functions and prove their convergence to the minimizer of the ROF model. Let $f \in L^2(\Omega)$ be fixed and Q_hf the discretization of f with respect to the quadrangulation $\{\Omega_{i,j} : 1 \leq i,j \leq N\}$. Let $z^{f,h}$ be the minimizer of the functional

$$E_h^f(u) = \lambda J_h(u) + \frac{h^2}{2} \sum_{1\leq i,j\leq N} |u_{i,j} - (Q_hf)_{i,j}|^2, \tag{3.22}$$

over $\mathbb{R}^{N\times N}$ with $J_h(u)$ defined in (3.7). We denote the minimizer of the ROF model in the continuous setting by

$$u^f = \underset{u\in BV(\Omega)}{\arg\min}\, E_\lambda^f(u), \tag{3.23}$$

where $E_\lambda^f(u)$ is defined in (ROF).

We now construct a continuous piecewise linear function by interpolating the discrete minimizer $z^{f,h}$ and show that it converges to u^f for a special class of functions f. Let $P_hz^{f,h}$ be the continuous piecewise linear interpolation of the discrete minimizer $z^{f,h}$ over the triangulation Δ_h. By the estimate (2.33), we have

$$\left\|P_h z^{f,h} - u^f\right\|_2^2 \leq 2\left(E_\lambda^f(P_h z^{f,h}) - E_\lambda^f(u^f)\right).$$

Therefore, it suffices to show that $\left(E_\lambda^f(P_h z^{f,h}) - E_\lambda^f(u^f)\right) \to 0$ as $h \to 0$, to infer that the continuous piecewise linear functions $P_h z^{f,h}$ approximate the solution of the ROF model. To this aim, we shall compare both $E_\lambda^f(P_h z^{f,h})$ and $E_\lambda^f(u^f)$ to the discrete energy $E_h^f(z^{f,h})$.

LEMMA 3.7. *Let $z^{f,h}$ be the minimizer of the functional $E_h^f(u)$ with respect to $\mathbb{R}^{N\times N}$. Then*

$$E_\lambda^f(P_h z^{f,h}) \leq E_h^f(z^{f,h}) + \frac{1}{2}C\omega(f,h)_2\left(C\omega(f,h)_2 + 4\|f\|_2\right) \tag{3.24}$$

where C is a positive constant depending on f.

PROOF. Since $J_h(z^{f,h}) = |DP_h z^{f,h}|(\Omega)$, we have

$$\begin{aligned}
2(E_\lambda^f(P_h z^{f,h}) - E_h^f(z^{f,h})) &= \|P_h z^{f,h} - f\|_2^2 - \sum_{1\leq i,j\leq N} h^2|z_{i,j}^{f,h} - Q_h f_{i,j}|^2 \\
&\leq \|P_h Q_h f - f\|_2(\|P_h Q_h f - f\|_2 + 2\|P_h(z^{f,h} - Q_h f)\|_2) + \\
&\quad + \underbrace{\|P_h(z^{f,h} - Q_h f)\|_2^2 - \sum_{1\leq i,j\leq N} h^2|z_{i,j}^{f,h} - Q_h f_{i,j}|^2}_{\leq 0 \text{ by Lemma 3.5}} \\
&\leq \|P_h Q_h f - f\|_2(\|P_h Q_h f - f\|_2 + 2\|P_h(z^{f,h} - Q_h f)\|_2),
\end{aligned} \tag{3.25}$$

To finish the proof, it suffices to show that

$$\|P_h Q_h f - f\|_2 \leq C\omega(f,h)_2 \text{ and } \|P_h(z^{f,h} - Q_h f)\|_2 \leq 2\|f\|_2. \tag{3.26}$$

First, as a byproduct of the proof of Lemma 3.5 we have

$$\begin{aligned}
\|P_h(z^{f,h} - Q_h f)\|_2^2 &\leq \sum_{1\leq i,j\leq N} h^2|z_{i,j}^{f,h} - (Q_h f)_{i,j}|^2 \leq 2E_h^f(0) \\
&= \sum_{1\leq i,j\leq N} h^2|(Q_h f)_{i,j}|^2 \leq 4\|f\|_{L^2}^2.
\end{aligned}$$

Next, by equation (3.20) in Lemma 3.6 we know that

$$\|P_h Q_h f - f\|_2 \leq K\omega(f,h)_2.$$

Therefore the inequalities in (3.26) hold with $C = K$, and the proof is complete. $\square$

LEMMA 3.8. *Let $z^{f,h}$ be the solution of* (3.22). *For $0 < \varepsilon \ll 1$, set $u_\varepsilon^f = \boldsymbol{X}[u^f] * \eta_\epsilon$. If $f \in L^\infty(\Omega)$, then*

$$E_h^f(z^{f,h}) \leq E_\lambda^f(u_\varepsilon^f) + 16\|f\|_\infty^2 h + \mathcal{O}(h/\varepsilon). \tag{3.27}$$

PROOF. With a slight abuse of notation, we let u_ε^f be the element of $\mathbb{R}^{N\times N}$ obtained by evaluating u_ε^f at the grid points $\omega_{i,j}$. By definition of $z^{f,h}$, we have

$$\begin{aligned} E_h^f(z^{f,h}) &\leq E_h^f(u_\varepsilon^f) = \lambda\, J_h(P_h u_\varepsilon^f) + \frac{1}{2}\sum_{i,j=1}^N h^2 |u_\varepsilon^f(\omega_{i,j}) - (Q_h f)_{i,j}|^2 \\ &\leq \lambda \int_\Omega |\nabla(P_h u_\varepsilon^f)| dx + \frac{1}{2}\sum_{i,j=1}^N h^2 |u_\varepsilon^f(\omega_{i,j}) - (Q_h f)_{i,j}|^2 \\ &\leq \lambda \int_\Omega |\nabla u_\varepsilon^f| dx + \lambda \int_\Omega |\nabla(P_h u_\varepsilon^f - u_\varepsilon^f)| dx + \frac{1}{2}\sum_{i,j=1}^N h^2 |u_\varepsilon^f(\omega_{i,j}) - (Q_h f)_{i,j}|^2. \end{aligned} \tag{3.28}$$

Next, for each $1 \leq i,j \leq N$, we have

$$\begin{aligned} |u_\varepsilon^f(\omega_{i,j}) - (Q_h f)_{i,j}|^2 = |u_\varepsilon^f(\omega_{i,j}) - (Q_h u_\varepsilon^f)_{i,j}|^2 + |(Q_h u_\varepsilon^f - Q_h f)_{i,j}|^2 + \\ + 2|u_\varepsilon^f(\omega_{i,j}) - (Q_h u_\varepsilon^f)_{i,j}| \cdot |(Q_h u_\varepsilon^f - Q_h f)_{i,j}| \end{aligned} \tag{3.29}$$

and by the mean value theorem

$$\begin{aligned} |u_\varepsilon^f(\omega_{i,j}) - (Q_h u_\varepsilon^f)_{i,j}|^2 &\leq \frac{1}{|\Omega_{i,j}|}\int_{\Omega_{i,j}} |u_\varepsilon^f(\omega_{i,j}) - u_\varepsilon^f(x)|^2 dx \\ &\leq \frac{1}{|\Omega_{i,j}|}\sup_{x\in\bar{\Omega}_{i,j}} |\nabla u_\varepsilon^f(x)|^2 \int_{\Omega_{i,j}} |x - \omega_{i,j}|^2 dx \\ &\leq \frac{C}{\varepsilon^2}|\Omega_{i,j}|, \end{aligned} \tag{3.30}$$

where C is a positive constant depending only on u through its L^1- norm. Thus

$$\sum_{i,j=1}^{N} h^2|u_\varepsilon^f(\omega_{i,j}) - (Q_h f)_{i,j}|^2 \leq C|\Omega|\frac{h^2}{\varepsilon^2} + \sum_{i,j=1}^{N} h^2|(Q_h u_\varepsilon^f - Q_h f)_{i,j}|^2 + C'\frac{h}{\varepsilon}, \quad (3.31)$$

where C, C' are positive constants depending on f, u^f, and Ω. Now, we establish an upper bound for the second term on the right in the inequality (3.31). By definition of the operator Q_h, the Cauchy-Schwarz inequality and Theorem 2.29, we have

$$\sum_{i,j=1}^{N} h^2|Q_h(u_\varepsilon^f - f)_{i,j}|^2 \leq \|u_\varepsilon^f - f\|_{L^2(\Omega)}^2 + 16\|f\|_\infty^2 h. \quad (3.32)$$

Taking into account (3.32) and (3.31) in the inequality (3.28), we obtain

$$E_h^f(z^{f,h}) \leq E_\lambda^f(u_\varepsilon^f) + 16\|f\|_\infty^2 h + C|\Omega|\frac{h^2}{\varepsilon^2} + C'\frac{h}{\varepsilon} + \|P_h u_\varepsilon^f - u_\varepsilon^f\|_{W^{1,1}(\Omega)}. \quad (3.33)$$

Since the rectangular domain is endowed with a type I triangulation, we have [see 26, Theorem 4.4.20, p. 108]

$$\|P_h u_\varepsilon^f - u_\varepsilon^f\|_{W^{1,1}(\Omega)} \leq Ch \sum_{|\alpha|=2} \|D^\alpha u_\varepsilon^f\|_{L^1(\Omega)} \leq C''\frac{h}{\varepsilon^2}, \quad (3.34)$$

where C'' is a constant that depends on $\|u\|_{L^1(\Omega)}$. Thus, the estimate (3.33) becomes

$$E_h^f(z^{f,h}) \leq E_\lambda^f(u_\varepsilon^f) + 16\|f\|_\infty^2 h + C\frac{h}{\varepsilon^2},$$

where we have used the fact that $x^2 < x$ for any $0 < x < 1$. □

We now state and prove the main result of this chapter.

THEOREM 3.9. *Suppose that $f \in \mathrm{Lip}\,(\alpha, L^2(\Omega)) \cap L^\infty(\Omega)$ for some $\alpha \in (0,1]$. Let $z^{f,h}$ be the minimizer of the functional $E_h^f(u)$ in $\mathbb{R}^{N\times N}$ and u^f be defined by* (3.23). *Then $P_h z^{f,h}$ converges in $L^2(\Omega)$ to u^f as $h \to 0$.*

PROOF. For any $0 < h \ll 1$ and any $\varepsilon > 0$, we have

$$\|P_h z^{f,h} - u^f\|_{L^2(\Omega)}^2 \leq 2\left[E_\lambda^f(P_h z^{f,h}) - E_h^f(z^{f,h}) + E_h^f(z^{f,h}) - E_\lambda^f(u^f)\right].$$

Next, by Lemma 3.7 we have

$$E_\lambda^f(P_h z^{f,h}) - E_h^f(z^{f,h}) \le \frac{1}{2} C_1\, \omega(f,h)_2 \left(C_1\, \omega(f,h)_2 + 4\|f\|_2\right),$$

while Lemma 3.8 yields

$$E_h^f(z^{f,h}) - E_\lambda^f(u^f) \le E_\lambda^f(u_\varepsilon^f) - E_\lambda^f(u^f) + 16\|f\|_\infty^2 h + C_2 \frac{h}{\varepsilon^2}.$$

Thus,

$$\|P_h z^{f,h} - u^f\|_{L^2(\Omega)}^2 \le C_1\, \omega(f,h)_2 \left(C_1\, \omega(f,h)_2 + 4\|f\|_2\right) + \\ + 32\|f\|_\infty^2 h + 2C_2 \frac{h}{\varepsilon^2} + 2\lambda(E_\lambda^f(u_\varepsilon^f) - E_\lambda^f(u^f)). \quad (3.35)$$

Now, since $f \in \mathrm{Lip}\,(\alpha, L^\alpha(\Omega))$ we have $\omega(f,h)_2 \le \mathcal{O}(h^\alpha)$. Letting $\varepsilon = h^{1/2(\alpha+1)}$, we infer from inequality (3.35) that

$$\|P_h z^{f,h} - u^f\|_{L^2(\Omega)}^2 \le C h^{\alpha/(\alpha+1)} + 2\lambda(E_\lambda^f(u_\varepsilon^f) - E_\lambda^f(u^f)), \quad (3.36)$$

where we have used the fact that the function $x \mapsto a^x$ is decreasing when $0 < a < 1$.

Since $u_\varepsilon^f \xrightarrow{\epsilon \to 0} u^f$ in $L^2(\Omega)$ and $Du_\epsilon^f|(\Omega) \xrightarrow{\epsilon \to 0} Du|(\Omega)$, it follows that for our choice of $\varepsilon = h^{1/2(\alpha+1)}$, $E_\lambda^f(u_\varepsilon^f) - E_\lambda^f(u^f) \to 0$ as $h \to 0$. Thus taking the limit as $h \to 0$ in (3.36), we conclude that $\|P_h z^{f,h} - u^f\|_{L^2(\Omega)} \to 0$ as $h \to 0$ and the proof is complete. □

COROLLARY 3.10. *Under the assumptions of Theorem 3.9, we have*

$$J_h(P_h z^{f,h}) \to J(u^f), \quad \text{when } h \to 0.$$

PROOF. This is a direct consequence of the convergence of $E_h^f(P_h z^{f,h})$ to $E_\lambda^f(u^f)$ as $h \to 0$. □

REMARK 3.11. It transpires from the proof above that to establish a convergence rate of the proposed piecewise linear approximation, one will need a convergence rate of $E_\lambda^f(u_\varepsilon^f)$ to $E_\lambda^f(u^f)$ which we have not been able to establish at this point. Moreover the optimal convergence rate, if one could be derived, should be of the order of $\mathcal{O}(h^{\beta})$ with $0 < \beta \leq 1/2$. The convergence is slower for smaller values of β and one would need very small values of h to get significant evidence of the convergence when doing numerical simulations.

3.4 A NUMERICAL EXPERIMENT

We report here the results of a numerical test carried to confirm our theoretical result. We use the algorithms to be introduced in the next chapter to simulate our approximation theory for the data function

$$f = 255\mathbb{1}_B$$

where B is the disk centered at $(1/2, 1/2)$ with radius $R = 1/4$. Our choice of this function is supported in the fact that it is one of the few functions for which an explicit formula of the ROF minimizer is know. In fact, we have seen in section 2.3.3 that in this case the minimizer u^f is given by

$$u^f = 255\max(1 - 2\lambda/R, 0)\mathbb{1}_B, \quad \forall\,\lambda > 0.$$

Note that the size of the discrete data $Q_h f$ grows as $1/h^2$ as $h \to 0$, therefore we will only show the result of moderate size data. The table below shows the distance between u^f and $P_h u_{100}$ where u_{100} is the approximation of $z^{f,h}$ computed using Algorithm 4.12. It is already apparent that the distance is decreasing with h even though we are only using an approximation of the discrete minimizer $z^{f,h}$.

	λ/R			
h	2^{-3}	2^{-5}	2^{-7}	2^{-9}
2^{-5}	25.0682	18.4406	17.9938	17.9808
2^{-6}	26.1967	13.8377	11.5935	11.3495
2^{-7}	21.0148	14.1954	9.1324	8.5836
2^{-8}	17.8916	14.1036	7.3424	6.0095
2^{-9}	16.1267	10.2853	7.3082	4.5298
2^{-10}	15.1462	7.6813	7.1739	3.6942

TABLE 3.1: The $L^2(\Omega)$ distance between u^f and $P_h u_{100}$ where u_{100} is the approximation of $z^{f,h}$ computed using Algorithm 4.12.

4

ALGORITHMS FOR COMPUTING THE PIECEWISE LINEAR APPROXIMATION

In this chapter we present a fixed point and proximal gradient algorithms for computing approximation of the discrete minimizer used in constructing the piecewise linear approximation studied in the previous chapter. We recall that to compute the piecewise linear approximation, it suffices to compute the discrete minimizer

$$z^{f,h} := \arg\min_{u \in \mathbb{R}^{N\times N}} \lambda J_h(u) + \frac{h^2}{2} \sum_{1\leq i,j\leq N} |u_{ij} - f_{ij}|^2, \tag{PM}$$

where $J_h(u)$ is defined in (3.7). We note that the existence and uniqueness of $z^{f,h}$ follows from Proposition 2.9. However, since the objective function is not differentiable, computing $z^{f,h}$ is hard and only iterative approximations are possible.

For the purpose of developing algorithms for computing $z^{f,h}$, we will show that the minimization problem (PM) is equivalent to a saddle point problem from which a more tractable counterpart of (PM) will be derived. Zhu et al. [76] used a similar approach for the standard discrete total variation based digital image denoising model. However, since we are using an averaged discrete total variation, we will deal with a pair of dual variables, paving the way for alternating dual algorithms. This work is the first time that such algorithms are proposed for the total variation based image denoising problem.

4.1 SADDLE POINTS AND CONSTRAINED OPTIMIZATION

In this section, we recall some classical results in convex analysis and optimization that we will use in formulating algorithms for the approximation of the solution of (PM). Our exposition follows the monographs [40, 43].

4.1.1 THE SADDLE POINT PROBLEM

We recall the relevant fact about saddle point problem that we will need to establish the equivalence of (PM) to a saddle point problem from which the algorithms will be developed. All the results presented hereafter are taken for the textbook [43].

Let $L(u,p)$ be a function defined on the product space $\mathbb{R}^n \times \mathbb{R}^m$ with values in $\mathbb{R}$, $\mathcal{A}$ a nonempty subset of $\mathbb{R}^n$, and $\mathcal{B}$ a nonempty subset of $\mathbb{R}^m$. We start with the definition of a saddle point.

DEFINITION 4.1. We say that a pair $(\bar{u},\bar{p}) \in \mathcal{A}\times\mathcal{B}$ is a saddle point of L over $\mathcal{A}\times\mathcal{B}$ if

$$L(\bar{u},p) \le L(\bar{u},\bar{p}) \le L(u,\bar{p}), \quad \forall (u,p) \in \mathcal{A}\times\mathcal{B}. \tag{4.1}$$

REMARK 4.2. We observe that if $L : \mathcal{A}\times\mathcal{B} \to \mathbb{R}$, then

$$\sup_{p\in\mathcal{B}} \inf_{u\in\mathcal{A}} L(u,p) \le \inf_{u\in\mathcal{A}} \sup_{p\in\mathcal{B}} L(u,p) \tag{4.2}$$

The next result gives a necessary and sufficient condition for the existence of a saddle point. Though not very practical for showing the existence of a saddle point, it is very instrumental in our characterization of $z^{f,h}$.

PROPOSITION 4.3 ([43]). *The function $L : \mathcal{A}\times\mathcal{B} \to \mathbb{R}$ has a saddle point at $(\bar{u},\bar{p}) \in \mathcal{A}\times\mathcal{B}$ if and only if*

$$L(\bar{u},\bar{p}) = \max_{p\in\mathcal{B}} \inf_{u\in\mathcal{A}} L(u,p) = \min_{u\in\mathcal{A}} \sup_{p\in\mathcal{B}} L(u,p). \tag{4.3}$$

We also have a characterization for the saddle points of a Gâteaux differentiable functional L.

THEOREM 4.4 ([43]). *Let L be a functional from $\mathcal{A}\times\mathcal{B}$ into $\mathbb{R}$. Suppose that*

$$\mathcal{A} \text{ and } \mathcal{B} \text{ are convex and closed,} \tag{4.4}$$

$$\forall\, u \in \mathcal{A},\ p \mapsto L(u,p) \textit{ is concave and upper semicontinuous}, \tag{4.5}$$

$$\forall\, p \in \mathcal{B},\ u \mapsto L(u,p) \textit{ is convex and lower semicontinuous}. \tag{4.6}$$

Assume in addition that

$$\forall\, u \in \mathcal{A},\ p \mapsto L(u,p) \textit{ is Gâteaux differentiable}, \tag{4.7}$$

and

$$\forall\, p \in \mathcal{B},\ u \mapsto L(u,p) \textit{ is Gâteaux differentiable}. \tag{4.8}$$

Then $(\bar{u}, \bar{p}) \in \mathcal{A} \times \mathcal{B}$ *is a saddle point of* L *if and only if*

$$\langle \frac{\partial L}{\partial u}(\bar{u}, \bar{p}), u - \bar{u} \rangle \geq 0, \quad \forall\, u \in \mathcal{A}, \tag{4.9a}$$

$$\langle \frac{\partial L}{\partial p}(\bar{u}, \bar{p}), p - \bar{p} \rangle \leq 0, \quad \forall\, p \in \mathcal{B}. \tag{4.9b}$$

4.1.2 Constrained minimization problems

We now recall a fundamental result in constrained optimization and use it later to derive a fixed-point iterative algorithm for computing $z^{f,h}$. Let f and $\{g_i\}_{1 \leq i \leq m}$ be functions defined from $\mathbb{R}^n$ into $\mathbb{R}$. The problem of interest is the following:

$$\begin{aligned} &\text{Minimize } f(x) &&\text{over } \mathbb{R}^n \\ &\text{subject to: } g_i(x) \leq 0, && 1 \leq i \leq m. \end{aligned} \tag{4.10}$$

Of course problem (4.10) is of interest only if the feasible set $\mathcal{F} := \bigcap_{i=1}^{m} \{g_i \leq 0\}$ is nonempty. For our purposes in this chapter, we will assume that the constraints g_i are convex and introduce the following definition.

Definition 4.5. We will say that the convex constraints $g_i,\ 1 \leq i \leq m,$ are *qualified* if either all the functions g_i are affine and $\mathcal{F} \neq \emptyset$ or there exists a point v such that for any $1 \leq i \leq m$, $g_i(v) \leq 0$ and $g_i(v) < 0$ if g_i is not affine.

We now state a classical result giving a necessary and sufficient condition for the existence and a solution of the solution to problem (4.10).

THEOREM 4.6 (Kuhn-Tucker conditions,[40]). *Suppose that the functions f and $g_i,\ 1 \le i \le m$ are differentiable and convex. Assume further that the constraints are qualified. Then a point $u \in \mathcal{F}$ is a solution of* (4.10) *if and only if there exists $\lambda(u) \in \mathbb{R}^m_+$ such that*

$$\begin{aligned} f'(u) + \sum_{1\le i\le m} \lambda_i(u) g_i'(u) &= 0, \\ \sum_{1\le i\le m} \lambda_i(u) g_i(u) &= 0. \end{aligned} \tag{4.11}$$

The vector $\lambda(u)$ is a Kuhn-Tucker vector associated to problem (4.10).

The theorem above states that under the assumptions of convexity and differentiability, the constrained minimization problem (4.10) with qualified constraints has a solution if and only if the Lagrangian

$$L(v,\lambda) = f(v) + \sum_{1\le i\le m} \lambda_i g_i(v), \quad (v,\lambda) \in \mathbb{R}^n \times \mathbb{R}^m$$

has at least one saddle point $(u, \lambda(u)) \in \mathbb{R}^n \times \mathbb{R}^m_+$ and $L(u,\lambda(u)) = f(u)$.

4.2 CHARACTERIZATION OF THE DISCRETE MINIMIZER

In this section, we establish that the discrete minimization problem (PM) is equivalent to a saddle point problem and derive a dual formulation of (PM) that we will use in developing the algorithms.

We begin with some notations. Let $X := \mathbb{R}^{N\times N}$ and $Y := X \times X$. An element $p \in Y$ will be represented by

$$p = (p_{i,j})_{1\le i,j\le N} = (p^1_{i,j}, p^2_{i,j})_{1\le i,j\le N}.$$

The vector space Y is naturally endowed with an inner product inherited from X and defined by

$$\langle p, q \rangle = \sum_{1 \le i,j \le N} p^1_{ij} q^1_{i,j} + p^2_{i,j} q^2_{i,j}, \quad p, q \in Y.$$

The norm associated to the above inner product will be denoted

$$\|p\|_2 := \sqrt{\langle p, p \rangle}, \quad p \in Y.$$

We also endow Y with the norm $\|\cdot\|_{2,\infty}$ defined by

$$\|p\|_{2,\infty} := \max_{1 \le i,j \le N} |p_{i,j}|_2,$$

where $|p_{i,j}|_2 = \sqrt{(p^1_{i,j})^2 + (p^2_{i,j})^2}$ is the Euclidean norm of $p_{i,j}$ in $\mathbb{R}^2$.

Next, we associate to the gradient operators $\nabla_+ : X \to Y$ and $\nabla_- : X \to Y$ defined in (3.8) and (3.9), the discrete divergence operators $\mathrm{div}_+ := -\nabla_+^* : Y \to X$ and $\mathrm{div}_- := -\nabla_-^* : Y \to X$, defined as the negative adjoint of ∇_+ and ∇_-, respectively:

$$\begin{aligned} \mathrm{div}_+(p)_{i,j} = &\begin{cases} 0 & \text{if } i = N \text{ or } j = N \\ \dfrac{p^1_{i,j}}{h} & \text{otherwise} \end{cases} - \begin{cases} 0 & \text{if } i = 1 \text{ or } j = N \\ \dfrac{p^1_{i-1,j}}{h} & \text{otherwise} \end{cases} \\ &+ \begin{cases} 0 & \text{if } i = N \text{ or } j = N \\ \dfrac{p^2_{i,j}}{h} & \text{otherwise} \end{cases} - \begin{cases} 0 & \text{if } i = N \text{ or } j = 1 \\ \dfrac{p^2_{i,j-1}}{h} & \text{otherwise,} \end{cases} \end{aligned} \tag{4.12}$$

and

$$\begin{aligned} \mathrm{div}_-(p)_{i,j} = &\begin{cases} 0 & \text{if } i = N \text{ or } j = 1 \\ \dfrac{p^1_{i+1,j}}{h} & \text{otherwise} \end{cases} - \begin{cases} 0 & \text{if } i = 1 \text{ or } j = 1 \\ \dfrac{p^1_{i,j}}{h} & \text{otherwise} \end{cases} \\ &+ \begin{cases} 0 & \text{if } i = 1 \text{ or } j = N \\ \dfrac{p^2_{i,j+1}}{h} & \text{otherwise} \end{cases} - \begin{cases} 0 & \text{if } i = 1 \text{ or } j = 1 \\ \dfrac{p^2_{i,j}}{h} & \text{otherwise.} \end{cases} \end{aligned} \tag{4.13}$$

Under our new notations, we have

$$J_h(u) = \frac{h^2}{2} \left(\|\nabla_+(u)\|_2 + \|\nabla_-(u)\|_2 \right).$$

Furthermore, by the Riesz representation theorem and the definition of the divergence operators, we also have

$$J_h(u) = h^2 \sup_{p,q\in B_Y} -\langle u, \frac{1}{2}\operatorname{div}_+(p) + \frac{1}{2}\operatorname{div}_-(q)\rangle_X, \tag{4.14}$$

where

$$B_Y := \{p \in Y : \|p\|_{2,\infty} \le 1\}$$

is be the closed unit ball of Y in the infinity norm $\|\cdot\|_{2,\infty}$. Therefore, the minimization problem (PM) is equivalent to the saddle-point problem

$$\arg\min_{u\in X} \sup_{p,q\in B_Y} -\lambda h^2\langle u, \frac{1}{2}\operatorname{div}_+(p) + \frac{1}{2}\operatorname{div}_-(q)\rangle_X + \frac{h^2}{2}\sum_{1\le i,j\le N} |u_{i,j} - f_{i,j}|^2. \tag{PDM}$$

We will refer to (PDM) as the *primal-dual total variation model*.

We now show that (PDM) has a solution. Let $\mathcal{L}$ be the functional defined on $X \times Y^2$ by

$$\mathcal{L}(u; p, q) := -\frac{\lambda h^2}{2}\langle u, \operatorname{div}_+(p) + \operatorname{div}_-(q)\rangle_X + \frac{h^2}{2}\|u - f\|_X^2. \tag{4.15}$$

We note that $\mathcal{L}$ is quadratic in the variable u and linear in the dual pair (p, q); thus we can apply Theorem 4.4 to show the existence of a saddle point of $\mathcal{L}$. We have the following

THEOREM 4.7. *A point $(\bar{u}; \bar{p}, \bar{q}) \in X \times B_Y^2$ is a saddle point of $\mathcal{L}$ over the set $X \times B_Y^2$ if and only if*

$$(\bar{p}, \bar{q}) \in \arg\min_{p,q\in B_Y} \|\lambda(\operatorname{div}_+(p) + \operatorname{div}_-(q)) + 2f\|_X^2. \tag{4.16}$$

and

$$\bar{u} = f + \frac{\lambda}{2}\left(\operatorname{div}_+(\bar{p}) + \operatorname{div}_-(\bar{q})\right) \tag{4.17}$$

Furthermore, if $(\bar{u}; \bar{p}, \bar{q})$ is a saddle point of $\mathcal{L}$ with respect to $X \times B_Y^2$, then

$$\bar{u} = \arg\min_{u\in\mathbb{R}^{N\times N}} \lambda J_h(u) + \frac{h^2}{2}\sum_{1\le i,j\le N} |u_{ij} - f_{ij}|^2.$$

PROOF. We note that $\mathcal{L}$ is Gâteaux differentiable with partial differentials

$$\nabla_u \mathcal{L}(u;p,q) = h^2(u-f) - \frac{\lambda h^2}{2}\left(\operatorname{div}_+(p) + \operatorname{div}_-(q)\right),$$
$$\nabla_{p,q}\mathcal{L}(u;p,q) = \frac{\lambda h^2}{2}\left[\nabla_+(u), \nabla_-(u)\right].$$

Suppose that $(\bar{u};\bar{p},\bar{q}) \in X \times B_Y^2$ is a saddle point of $\mathcal{L}$ over the set $X \times B_Y^2$. Then, by Theorem 4.4 we have

$$\langle \nabla_u \mathcal{L}(\bar{u};\bar{p},\bar{q}), u - \bar{u}\rangle \geq 0, \ \ \forall u \in X;$$

so that taking $u = \bar{u} \pm \nabla_u \mathcal{L}(\bar{u};\bar{p},\bar{q})$ in the inequality above yields $\nabla_u \mathcal{L}(\bar{u};\bar{p},\bar{q}) = 0$. Thus the point $(\bar{u};\bar{p},\bar{q})$ satisfies the equation (4.17). To show that (4.16) holds, we use the characterization of saddle points in Proposition 4.3 to obtain

$$\mathcal{L}(\bar{u},\bar{p},\bar{q}) = \max_{p,q\in B_Y}\min_{u\in X}\mathcal{L}(u;p,q).$$

Now, since $v \mapsto \mathcal{L}(v;p,q)$ is differentiable for any $(p,q) \in Y^2$, it follows that for each point $(p,q) \in B_Y^2$, the minimum of $v \mapsto \mathcal{L}(v;p,q)$ over X is achieved at a point u such that $\nabla_u \mathcal{L}(u;p,q) = 0$, *i.e*,

$$u = f + \frac{\lambda}{2}\left(\operatorname{div}_+(p) + \operatorname{div}_-(q)\right).$$

So,

$$\begin{aligned}
\mathcal{L}(\bar{u},\bar{p},\bar{q}) &= \max_{p,q\in B_Y}\min_{u\in X}\mathcal{L}(u;p,q) \\
&= \frac{h^2}{2}\max_{p,q\in B_Y} -2\langle f, \frac{\lambda}{2}(\operatorname{div}_+(p)+\operatorname{div}_-(q))\rangle_X - \frac{\lambda^2}{4}\|\operatorname{div}_+(p)+\operatorname{div}_-(q)\|_X^2 \\
&= \max_{p,q\in B_Y}\frac{h^2}{2}\left(\|f\|_X^2 - \frac{1}{4}\|\lambda(\operatorname{div}_+(p)+\operatorname{div}_-(q)) + 2f\|_X^2\right).
\end{aligned}$$

Thus,

$$(\bar{p},\bar{q}) \in \arg\min_{p,q\in B_Y}|\lambda(\operatorname{div}_+(p)+\operatorname{div}_-(q)) + 2f|^2.$$

Conversely, if $(\bar{u};\bar{p},\bar{q}) \in X \times B_Y^2$ satisfies (4.17) and (4.16), then it is easy to see from (4.16) that (4.9a) holds. Now, we notice that (4.17) means that $\operatorname{div}_+(\bar{p}) + \operatorname{div}_-(\bar{q})$ is the orthogonal projection of $-2f/\lambda$ onto the closed convex subset of X given by

$$K = \{\mathrm{div}_+(p) + \mathrm{div}_-(q) \colon p, q \in B_Y\}. \tag{4.18}$$

Thus, by the characterization of the orthogonal projection, we have

$$\langle -2f/\lambda - \mathrm{div}_+(\bar{p}) - \mathrm{div}_-(\bar{q}), \mathrm{div}_+(p - \bar{p}) + \mathrm{div}_-(q - \bar{q})\rangle_X \leq 0, \quad \forall\, p, q \in B_Y.$$

But by definition of the divergence operators, the latter inequality is equivalent to

$$\langle \nabla_{p,q}\mathcal{L}(\bar{u}; \bar{p}, \bar{q}), (p - \bar{p}, q - \bar{q}\rangle_{Y\times Y} \leq 0, \quad \forall\, p, q \in B_Y,$$

where we have used (4.16) again. Hence (4.9b) holds, and by Theorem 4.4, $(\bar{u}; \bar{p}, \bar{q})$ is a saddle point of $\mathcal{L}$ over $X \times B_Y^2$.

Finally, if $(\bar{u}; \bar{p}, \bar{q})$ is a saddle-point of $\mathcal{L}$ with respect to $X \times B_Y^2$, then

$$\begin{aligned} \mathcal{L}(\bar{u}; \bar{p}, \bar{q}) &= \min_{u\in X} \max_{p,q\in B_Y} \mathcal{L}(u; p, q) \\ &= \min_{u\in X} \lambda J_h(u) + \frac{1}{2\lambda}\|u - f\|_X^2 \text{ by (4.14).} \end{aligned}$$

Hence $\bar{u} = \arg\min_{u\in X} \lambda J_h(u) + \frac{1}{2\lambda}\|u - f\|_X^2$; and the proof is complete. □

REMARK 4.8. Theorem 4.7 asserts that solving the primal total variation model (PM) is equivalent to solving the *dual total variation model*

$$\text{Find } (\bar{p}, \bar{q}) \in \arg\min_{p,q\in B_Y} \|\lambda(\mathrm{div}_+(p) + \mathrm{div}_-(q)) + 2f\|_X^2. \tag{DM}$$

Moreover, an element $u \in X$ is the solution of (PM) if and only if $2\,(u - f)\,/\lambda$ is the orthogonal projection of $-2f/\lambda$ onto the closed convex set K defined in 4.18.

4.3 THE ALGORITHMS

We have established a primal-dual and a dual formulation of the original minimization (PM). In this section, we will develop three algorithms for computing a solution of (PM) via the dual problem (DM) and the identity (4.17).

4.3.1 Projected-gradient algorithm

The dual problem (DM) is a constrained quadratic program with convex constraints. Consequently, reasoning as [40, p. 323], we will develop a projected-gradient algorithm for computing its solution. However, since the constraints contains an open set, and the gradient of the objective is not coercive, we cannot use Theorem 8.6.2 of [40] to obtain the convergence regime of the algorithm. A pointed analysis is required in this case. We note that such an analysis has been done by Duval et al. [42] for the standard discrete ROF model (2.38). In a more recent paper, Aujol [14] demonstrated that the projected-gradient algorithm that they studied in [42], is a special class of the Bermùdez-Moreno algorithm [23, 1981].

The next result gives a necessary and sufficient condition satisfied by any solution $(\bar{p}, \bar{q})$ of the dual problem (DM), and is the motivation of the proposed projected-gradient algorithm.

Proposition 4.9. *Let $f \in X$ be given. The following are equivalent*

$$(\bar{p}, \bar{q}) \text{ is a solution of } (\mathrm{DM}), \tag{4.19}$$

$$\begin{cases} \bar{p} = P_{B_Y}\left(\bar{p} + \tau\nabla_+\left[\operatorname{div}_+(\bar{p}) + \operatorname{div}_-(\bar{q}) + 2f/\lambda\right]\right) \\ \bar{q} = P_{B_Y}\left(\bar{q} + \tau\nabla_-\left[\operatorname{div}_+(\bar{p}) + \operatorname{div}_-(\bar{q}) + 2f/\lambda\right]\right) \end{cases}, \quad \forall\, \tau > 0. \tag{4.20}$$

Proof. Let $(\bar{p}, \bar{q}) \in B_Y^2$ be such that (4.19) holds. It suffices to proof the first identity in (4.20); the proof of the second is identical.

If $(\bar{p}, \bar{q}) \in \underset{p,q \in B_Y}{\arg\min} |\lambda(\operatorname{div}_+(p) + \operatorname{div}_-(q) + 2f)|^2$, then

$$\bar{p} \in \underset{p \in B_Y}{\arg\min}\, |\operatorname{div}_+(p) + \operatorname{div}_-(\bar{q}) + 2f/\lambda|^2,$$

so that $\lambda \operatorname{div}_+(\bar{p})$ is the orthogonal projection of $-\lambda \operatorname{div}_-(\bar{q}) - 2f$ onto the closed convex set

$$\lambda K_+ := \{\lambda \operatorname{div}_+(p) \colon p \in B_Y\}.$$

Therefore, we have

$$\langle \lambda(\operatorname{div}_+(\bar{p}) + \operatorname{div}_-(\bar{q})) + 2f, \lambda \operatorname{div}_+(p - \bar{p})\rangle \geq 0, \quad \forall p \in B_Y,$$

or equivalently

$$\langle -\nabla_+(\operatorname{div}_+(\bar{p}) + \operatorname{div}_-(\bar{q}) + 2f/\lambda), p - \bar{p}\rangle \geq 0, \quad \forall p \in B_Y,$$

where we have used the definition of $-\operatorname{div}_+$ as the adjoint of ∇_+. The latter inequality is equivalent to

$$\langle [\bar{p} + \tau\nabla_+(\operatorname{div}_+(\bar{p}) + \operatorname{div}_-(\bar{q}) + 2f/\lambda)] - \bar{p}, p - \bar{p}\rangle \leq 0, \quad \forall\, p \in B_Y,\ \forall \tau > 0;$$

hence

$$\bar{p} = P_{B_Y}\left(\bar{p} + \tau\nabla_+\left[\operatorname{div}_+(\bar{p}) + \operatorname{div}_-(\bar{q}) + 2f/\lambda\right]\right), \quad \forall\, \tau > 0.$$

The proof of the second identity in (4.20) is identical to the one above, changing p to q, and div_+ to div_-.

Conversely, suppose there is a point $(\bar{p}, \bar{q}) \in B_Y^2$ such that (4.20) holds. Then, by the characterization of the orthogonal projection and for $\tau = 1$, we have

$$\langle -2f/\lambda - (\operatorname{div}_+(\bar{p}) + \operatorname{div}_-(\bar{q})), \operatorname{div}_+(p - \bar{p})\rangle_X \leq 0 \quad \forall\, p \in B_Y,$$
$$\langle -2f/\lambda - (\operatorname{div}_-(\bar{p}) + \operatorname{div}_-(\bar{q})), \operatorname{div}_-(q - \bar{q})\rangle_X \leq 0 \quad \forall\, q \in B_Y,$$

where we have used the definition of the divergence operators as the negative adjoint of the corresponding gradient operators. Adding the two inequalities above, we obtain

$$\langle -2f/\lambda - (\operatorname{div}_+(\bar{p}) + \operatorname{div}_-(\bar{q})), \operatorname{div}_+(p - \bar{p}) + \operatorname{div}_-(q - \bar{q})\rangle_X \leq 0, \ \forall\, p, q \in B_Y.$$

Thus $\operatorname{div}_+(\bar{p}) + \operatorname{div}_-(\bar{q})$ is the orthogonal projection of $-2f/\lambda$ on to the subset K of X, or equivalently $(\bar{p}, \bar{q})$ satisfies (DM). □

REMARK 4.10. Let $p \in Y$ be fixed. Then, it can easily be shown that

$$P_{B_Y}(p)_{i,j} = \left(\frac{p^1_{i,j}}{\max(1, |p_{i,j}|)}, \frac{p^2_{i,j}}{\max(1, |p_{i,j}|)} \right), \quad 1 \leq i, j \leq N. \tag{4.21}$$

In fact, since B_Y is the Cartesian product of closed unit disks of $\mathbb{R}^2$, and $\mathbb{R}^2$ is isometrically embedded in Y, $P_{B_Y}(p)_{i,j}$ is simply the orthogonal projection of $p_{i,j} = (p^1_{i,j}, p^2_{i,j})$ onto the closed unit ball of $\mathbb{R}^2$, which is given by (4.21).

Proposition 4.9 defines the solutions of the dual problem (DM) as fixed points of a one parameter families of Lipschitz continuous operators. Consequently, if we can show that for some values of the parameter τ, these operators are contractions, then the corresponding fixed point algorithm will certainly converge. Indeed, we have the following Lemma.

LEMMA 4.11. *Let I be the identity operator on $Y \times Y$, and $A : Y \times Y \to Y \times Y$ the linear operator defined by*

$$A := \begin{pmatrix} -\nabla_+ \operatorname{div}_+ & -\nabla_+ \operatorname{div}_- \\ -\nabla_- \operatorname{div}_+ & -\nabla_- \operatorname{div}_- \end{pmatrix}. \tag{4.22}$$

The following are true

The operator A is Hermitian and nonnegative definite. (4.23)

The norm of A is bounded above as follows: $\|A\| \leq 16/h^2$. (4.24)

The operator $I - \tau A$ is non expansive for any $\tau \in [0, h^2/8]$. (4.25)

The restriction of $I - \tau A$ to the range of A is a contraction, for any $0 < \tau < h^2/8$. (4.26)

PROOF. First, we show that A is nonnegative definite. Indeed, for any $p, q \in Y$

$$\begin{aligned} \langle A \begin{pmatrix} p \\ q \end{pmatrix}, \begin{pmatrix} p \\ q \end{pmatrix} \rangle_{Y \times Y} &= -\langle \nabla_+(\operatorname{div}_+(p) + \operatorname{div}_-(q)), p \rangle_Y - \langle \nabla_-(\operatorname{div}_+(p) + \operatorname{div}_-(q)), q \rangle_Y \\ &= \langle \operatorname{div}_+(p) + \operatorname{div}_-(q), \operatorname{div}_+(p) + \operatorname{div}_-(q) \rangle_X \\ &= \| \operatorname{div}_+(p) + \operatorname{div}_-(q) \|_X^2 \geq 0; \end{aligned}$$

hence A is nonnegative definite. Now let $s, r \in Y$ be given. Then,

$$\begin{aligned}
\langle A \begin{pmatrix} p \\ q \end{pmatrix}, \begin{pmatrix} s \\ r \end{pmatrix} \rangle_{Y \times Y} &= \langle \mathrm{div}_+(p) + \mathrm{div}_-(q), \mathrm{div}_+(s) + \mathrm{div}_-(r) \rangle_X \\
&= \langle p, -\nabla_+(\mathrm{div}_+(s) + \mathrm{div}_-(r)) \rangle_Y + \langle q, -\nabla_-(\mathrm{div}_+(s) + \mathrm{div}_-(r)) \rangle_Y \\
&= \langle \begin{pmatrix} p \\ q \end{pmatrix}, A \begin{pmatrix} s \\ r \end{pmatrix} \rangle_{Y \times Y};
\end{aligned}$$

thus A is clearly Hermitian, and (4.23) is true.

We now show that (4.24) holds. For any $p, q \in Y$, we have

$$\begin{aligned}
\left\| A \begin{pmatrix} p \\ q \end{pmatrix} \right\|_{Y^2}^2 &:= \|\nabla_+(\mathrm{div}_+(p) + \mathrm{div}_-(q))\|_X^2 + \|\nabla_-(\mathrm{div}_+(p) + \mathrm{div}_-(q))\|_X^2 \\
&\leq (\|\nabla_+\|^2 + \|\nabla_-\|^2) \| \|(\mathrm{div}_+(p) + \mathrm{div}_-(q))\|_Y^2 \\
&\leq 2(\|\nabla_+\|^2 + \|\nabla_-\|^2) \max(\|\mathrm{div}_+\|^2, \|\mathrm{div}_-\|^2)(\|p\|_Y^2 + \|q\|_Y^2)
\end{aligned}$$

Therefore,

$$\|A\| \leq \sqrt{2(\|\nabla_+\|^2 + \|\nabla_-\|^2) \max(\|\mathrm{div}_+\|^2, \|\mathrm{div}_-\|^2)}. \tag{4.27}$$

It now suffices to established that $\|\nabla_+\| = \|\mathrm{div}_+\| \leq 8/h^2$; the proof of $\|\nabla_-\| = \|\mathrm{div}_-\| \leq 8/h^2$ follows mutatis mutandis from that of $\|\nabla_+\| = \|\mathrm{div}_+\| \leq 8/h^2$. By definition of $\mathrm{div}_+ = -\nabla_+^*$, we have

$$\begin{aligned}
\langle \nabla_+ u, \nabla_+ u \rangle_Y &= \langle u, -\mathrm{div}_+(\nabla_+ u) \rangle_X, \quad \forall u \in X, \\
\langle \mathrm{div}_+(p), \mathrm{div}_+(p) \rangle_X &= \langle p, -\nabla_+(\mathrm{div}_+(p)) \rangle_Y, \quad \forall p \in Y.
\end{aligned}$$

Consequently, by Cauchy-Schwarz inequality,

$$\begin{aligned}
\|\nabla_+(u)\|_Y^2 &\leq \|\mathrm{div}_+\| \cdot \|\nabla_+\| \cdot \|u\|_X^2, \quad \forall\, u \in X, \\
\|\mathrm{div}_+(p)\|_X^2 &\leq \|\nabla_+\| \cdot \|\mathrm{div}_+\| \cdot \|p\|_Y^2, \quad \forall\, p \in Y,
\end{aligned}$$

so that taking the supremum over all $u \in X$ such that $\|u\|_X = 1$ and all $p \in Y$ such $\|p\|_Y = 1$, we obtain that $\|\nabla_+\| = \|\mathrm{div}_+\|$.

Furthermore, for any $u \in X$,

$$\|\nabla_+ u\|_Y^2 = \frac{1}{h^2} \sum_{1 \le i,j \le N-1} (u_{i+1,j} - u_{i,j})^2 + (u_{i,j+1} - u_{i,j})^2$$
$$\le \frac{2}{h^2} \sum_{1 \le i,j \le N-1} u_{i+1,j}^2 + 2u_{i,j}^2 + u_{i,j+1}^2 \le \frac{8}{h^2}\|u\|^2,$$

where we have used the estimate $2ab \le a^2 + b^2$ to obtain the last inequality above. Hence, $\|\nabla_+\|^2 = \|\operatorname{div}_+\|^2 \le 8/h^2$. Likewise, we show that $\|\nabla_-\|^2 = \|\operatorname{div}_-\|^2 \le 8/h^2$. Consequently, it follows from (4.27) that $\|A\| \le h^2/16$; thus (4.24) is true.

We now prove the last two properties of A. We note to begin that since A is Hermitian and nonnegative definite, we have

$$Y \times Y = \ker(A) \overset{\perp}{\oplus} F, \tag{4.28}$$

where $F = \ker(A)^\perp =$ Range(A). Moreover, all the eigenvalues of A are nonnegative and

$$\kappa := \|I - \tau A\| = \max(1, |1 - \tau\|A\||),$$

where $\|A\|$ denotes the spectral norm of A (the largest eigenvalue of A). So, $I - \tau A$ is non expansive if and only if $0 < \tau \le 2/\|A\|$. Since $\|A\| \le h^2/16$, it follows that $I - \tau A$ is indeed non expansive for $\tau \in [0, h^2/8]$.

Finally, since F has an orthogonal basis made of the eigenvectors of A associated to nonzero eigenvalues, it follows that the norm of the restriction of $I - \tau A$ to F is

$$\kappa_F = \max(|1 - \tau\lambda_2(A)|, |1 - \tau\|A\||),$$

where $\lambda_2(A)$, the smallest nonzero eigenvalue of A, satisfies $0 < \lambda_2(A) \le \|A\| \le 16/h^2$. Hence, if $0 < \tau < h^2/8$, we get $-1 < 1 - \tau\lambda_2(A) < 1$ and $-1 < 1 - \tau\|A\| < 1$, so that $\kappa_F < 1$. Thus $I - \tau A$ is a contraction on the range of A for all $0 < \tau < h^2/8$. □

ALGORITHM 4.12 (Dual Projected-Gradient). *Choose* $\tau > 0$ *and* $p_0, q_0 \in B_Y$.

1. *For any* $n \geq 0$, *compute* u_n

$$u_n = f + \frac{\lambda}{2}\left[\mathrm{div}_+(p_n) + \mathrm{div}_-(q_n)\right]. \tag{4.29}$$

2. *Update the dual variables* p *and* q *as follows*

$$p_{n+1} = P_{B_Y}\left(p_n + 2\tau/\lambda \nabla_+(u_n)\right) \tag{4.30a}$$

$$q_{n+1} = P_{B_Y}\left(q_n + 2\tau/\lambda \nabla_-(u_n)\right). \tag{4.30b}$$

3. *Until stopping criterion, increment* $n \leftarrow n+1$ *and return to* ***1***.

A heuristic for the appellation of the above algorithm is as follows. The algorithm updates the dual variables p and q in two stages: (1) we update the current state of p and q with a gradient descent step of size τ; (2) we project the resulting updates onto the feasible set of the dual problem (DM).

We are now ready to prove the convergence of the dual projected gradient algorithm. Our proof follows the argument in [42] and completes it by also showing the convergence of the dual sequence (p_n, q_n).

PROPOSITION 4.13. *If* $0 < \tau < h^2/8$, *then Algorithm 4.12 converges. More precisely, given* p_0, $q_0 \in B_Y$, *there exists* $(\bar{p}_0, \bar{q}_0) \in B_Y^2$ *satisfying* (4.20) *such that the sequence* (p_n, q_n) *defined by* (4.30) *converges to* $(\bar{p}_0, \bar{q}_0)$ *and the sequence* u_n *defined by* (4.29) *converges to* $z^{f,h}$, *the solution of* (PM).

PROOF. Let $(\bar{p}, \bar{q})$ be any solution of (4.20), and $\tau \in (0, h^2/8)$ be given. Since the orthogonal projection onto $B_Y \times B_Y$ is equivalent to projecting each copy of Y in $Y \times Y$ onto the corresponding copy of B_Y, it follows from (4.20) and (4.30) that

$$\begin{pmatrix} p_{n+1} \\ q_{n+1} \end{pmatrix} = P_{B_Y \times B_Y}\left((I - \tau A)\begin{pmatrix} p_n \\ q_n \end{pmatrix} + \begin{pmatrix} 2f/\lambda \\ 2f/\lambda \end{pmatrix} \right),$$

and

$$\begin{pmatrix} \bar{p} \\ \bar{q} \end{pmatrix} = P_{B_Y \times B_Y} \left((I - \tau\, A) \begin{pmatrix} \bar{p} \\ \bar{q} \end{pmatrix} + \begin{pmatrix} 2f/\lambda \\ 2f/\lambda \end{pmatrix} \right),$$

where A is the linear operator defined in (4.22). Since orthogonal projections are non expansive, we infer from the latter identities that

$$\left\| \begin{pmatrix} p_{n+1} - \bar{p} \\ q_{n+1} - \bar{q} \end{pmatrix} \right\|_{Y^2} \leq \|I - \tau A\| \left\| \begin{pmatrix} p_n - \bar{p} \\ q_n - \bar{q} \end{pmatrix} \right\|_{Y^2}$$

As a consequence, by Lemma 4.11, we have for any $\tau \in [0, h^2/8]$ and for all $(\bar{p}, \bar{q})$ satisfying (4.20)

$$\left\| \begin{pmatrix} p_{n+1} - \bar{p} \\ q_{n+1} - \bar{q} \end{pmatrix} \right\|_{Y^2} \leq \left\| \begin{pmatrix} p_n - \bar{p} \\ q_n - \bar{q} \end{pmatrix} \right\|_{Y^2}, \quad \forall\, n \geq 0. \tag{4.31}$$

We now show that the sequence $(p_n, q_n)_{n \geq 0}$ converges by showing that all of its convergent subsequences have the same limit. Let $(p_{n_k}, q_{n_k})_{k \geq 0}$ be a convergent subsequence with limit $(\tilde{p}, \tilde{q})$. Then by (4.29) and (4.30), the subsequence (p_{n_k+1}, q_{n_k+1}) converges to $(\hat{p}, \hat{q})$ and by (4.31), for all $(\bar{p}, \bar{q})$ solution of (DM)

$$\left\| \begin{pmatrix} \hat{p} - \bar{p} \\ \hat{q} - \bar{q} \end{pmatrix} \right\|_{Y^2} = \left\| \begin{pmatrix} \tilde{p} - \bar{p} \\ \tilde{q} - \bar{q} \end{pmatrix} \right\|_{Y^2}, \tag{4.32}$$

so that using the equations (4.20) and (4.28), we have

$$\begin{aligned} \left\| \begin{pmatrix} \tilde{p} - \bar{p} \\ \tilde{q} - \bar{q} \end{pmatrix} \right\|_{Y^2}^2 &\leq \left\| (I - \tau A) \begin{pmatrix} \tilde{p} - \bar{p} \\ \tilde{q} - \bar{q} \end{pmatrix} \right\|_{Y^2}^2 \\ &\leq \left\| \begin{pmatrix} \tilde{p} - \bar{p} \\ \tilde{q} - \bar{q} \end{pmatrix}_{\ker(A)} \right\|_{Y^2}^2 + \kappa_F^2 \left\| \begin{pmatrix} \tilde{p} - \bar{p} \\ \tilde{q} - \bar{q} \end{pmatrix}_F \right\|_{Y^2}^2, \end{aligned} \tag{4.33}$$

where κ_F is the norm of $I - \tau A$ with respect to F. But we know from Lemma 4.11 that $\kappa_F < 1$, thus (4.33) implies that we must have $\begin{pmatrix} \tilde{p} - \bar{p} \\ \tilde{q} - \bar{q} \end{pmatrix} \in \ker(A)$; hence

$$\operatorname{div}_+(\tilde{p}) + \operatorname{div}_-(\tilde{q}) = \operatorname{div}_+(\bar{p}) + \operatorname{div}_-(\bar{q}). \tag{4.34}$$

Since $\operatorname{div}_+(\bar{p}) + \operatorname{div}_-(\bar{q})$ is the orthogonal projection of $-2f/\lambda$ onto the closed convex set $K = \{\operatorname{div}_+(p) + \operatorname{div}_-(q) \colon p, q \in B_Y\}$, so is $\operatorname{div}_+(\tilde{p}) + \operatorname{div}_-(\tilde{q})$ and it follows that $(\tilde{p}, \tilde{q})$ is a solution of (DM).

Finally, rewriting (4.32) for the solution $(\tilde{p}, \tilde{q})$, we obtain that the subsequence (p_{n_k+1}, q_{n_k+1}) converges to $(\tilde{p}, \tilde{q})$. Moreover, by continuity of the divergence operators, equation (4.34), and Theorem 4.7, the subsequence u_{n_k} defined by (4.29) converges to $z^{f,h}$ the solution of (PM). To finish the proof, we show that any two convergent subsequences of (p_n, q_n) have the same limit. Let $(p_{n_k}, q_{n_k}) \to (\tilde{p}, \tilde{q})$ and $(p_{m_k}, q_{m_k}) \to (\hat{p}, \hat{q})$ be two such subsequences. We may assume without loss of generality that $n_k \leq m_k$, otherwise we can always extract a further subsequence of (p_{n_k}, q_{n_k}) for which the property holds. By the monotonicity property of (p_n, q_n) given in (4.31) and the fact that $(\hat{p}, \hat{q})$ solves (DM), we obtain

$$\left\| \begin{pmatrix} p_{n_k} - \hat{p} \\ q_{n_k} - \hat{q} \end{pmatrix} \right\| \leq \left\| \begin{pmatrix} p_{m_k} - \hat{p} \\ q_{m_k} - \hat{q} \end{pmatrix} \right\|. \tag{4.35}$$

Passing to the limit in the latter inequality yields $\tilde{p} = \hat{p}$ and $\tilde{q} = \hat{q}$. Thus, the sequence (p_n, q_n) converges to a point $(\bar{p}_0, \bar{q}_0)$ solution of (DM). □

REMARK 4.14. Paralleling the argument in [14], we can show that Algorithm 4.12 is a special case of Bermùdez-Moreno algorithm.

We obtain an alternating version of the above algorithm by using a Gauss-Seidel type update on the dual variables p and q. The resulting algorithm reads as follows:

ALGORITHM 4.15 (Alternating Dual Projected-Gradient). *Choose $\tau > 0$ and $p_0, q_0 \in B_Y$.*

1. *For any $n \geq 0$, compute u_n*

$$u_n = f + \frac{\lambda}{2} \left[\operatorname{div}_+(p_n) + \operatorname{div}_-(q_n) \right]. \tag{4.36}$$

2. *Update p, do a half step update of u, then update q*

$$p_{n+1} = P_{B_Y}\left(p_n + 2\tau/\lambda \nabla_+(u_n)\right), \tag{4.37a}$$

$$u_{n+\frac{1}{2}} = f + \frac{\lambda}{2}\left[\mathrm{div}_+(p_{n+1}) + \mathrm{div}_-(q_n)\right], \tag{4.37b}$$

$$q_{n+1} = P_{B_Y}\left(q_n + 2\tau/\lambda \nabla_-(u_{n+\frac{1}{2}})\right). \tag{4.37c}$$

3. *Until stopping criterion, increment $n \leftarrow n+1$ and go to **1**.*

While the proof of convergence of the alternating projected-gradient algorithm above is still eluding us, the numerical experiments suggest that one should be able to prove the following conjecture

CONJECTURE 1. *If $0 < \tau \leq h^2/4$, then Algorithm 4.15 converges. More precisely, given p_0, $q_0 \in B_Y$, there exists $(\bar{p}_0, \bar{q}_0) \in B_Y^2$ satisfying* (4.20) *such that the sequence (p_n, q_n) defined by* (4.30) *converges to $(\bar{p}_0, \bar{q}_0)$ and the sequence u_n defined by* (4.29) *converges to $z^{f,h}$, the solution of* (PM).

4.3.2 A FIXED-POINT ITERATIVE ALGORITHM

Since the constraint of (DM) is qualified in the sense of definition 4.5, we use Theorem 4.6 to derive an alternate one parameter family of functionals for which the solutions of the dual problem (DM) arise as fixed points. This approach was first used by Chambolle [31] to construct a breakthrough algorithm for the standard discrete ROF model (2.38). The peculiarity of the algorithm to follow is that we used and alternating scheme to update the dual variables (p, q).

We observe that the dual problem is equivalent to the constrained minimization problem with quadratic objective

$$F(p,q) = \langle A \begin{pmatrix} p \\ q \end{pmatrix}, \begin{pmatrix} p \\ q \end{pmatrix} \rangle_X - 2\langle \begin{pmatrix} \nabla_+(2f/\lambda) \\ \nabla_-(2f/\lambda) \end{pmatrix}, \begin{pmatrix} p \\ q \end{pmatrix} \rangle_X + \|2f/\lambda\|_X^2 \tag{4.38}$$

where A is the operator defined in (4.22), and convex quadratic constraints:

$$g_{i,j}(p,q) := (p^1_{i,j})^2 + (p^2_{i,j})^2 - 1 \leq 0, \quad 1 \leq i,j \leq N, \tag{4.39a}$$

$$h_{i,j}(p,q) := (q^1_{i,j})^2 + (q^2_{i,j})^2 - 1 \leq 0, \quad 1 \leq i,j \leq N. \tag{4.39b}$$

Clearly, the objective function F is differentiable with gradient field

$$\operatorname{grad}(F)(p,q) = -2 \begin{pmatrix} \nabla_+(\operatorname{div}_+(p) + \operatorname{div}_-(q) + 2f/\lambda) \\ \nabla_+(\operatorname{div}_-(p) + \operatorname{div}_-(q) + 2f/\lambda) \end{pmatrix}, \tag{4.40}$$

and the constraints are also differentiable with gradient

$$\operatorname{grad}(g_{i,j})(p,q) = 2 \begin{pmatrix} p_{i,j} \\ 0 \end{pmatrix} \text{ and } \operatorname{grad}(h_{i,j})(p,q) = 2 \begin{pmatrix} 0 \\ q_{i,j} \end{pmatrix}, \quad 1 \leq i,j \leq N. \tag{4.41}$$

Moreover, the constraints are qualified in the sense of Definition 4.5 with $p = q = 0$. Therefore, the solutions of (DM) are also characterized by the Kuhn-Tucker conditions in Theorem 4.6.

The next result which follows directly from Theorem 4.6, gives another characterization of a solution of the dual problem (DM). Let $(p,q) \in Y^2$ be fixed and define $\alpha \in Y$ and $\beta \in Y$ in terms of p and q as follows:

$$\alpha_{i,j}(p,q) = \left| \nabla_+ \left(\operatorname{div}_+(p) + \operatorname{div}_-(q) + 2f/\lambda\right)_{i,j} \right|, \;\; 1 \leq i,j \leq N, \tag{4.42a}$$

$$\beta_{i,j}(p,q) = \left| \nabla_- \left(\operatorname{div}_+(p) + \operatorname{div}_-(q) + 2f/\lambda\right)_{i,j} \right| \;\; 1 \leq i,j \leq N. \tag{4.42b}$$

THEOREM 4.16. *A point $(p,q) \in B_Y \times B_Y$ is a solution of the dual problem* (DM) *if and only if for any $\tau > 0$*

$$p_{i,j} = \frac{p_{i,j} + \tau \nabla_+ \left(\operatorname{div}_+(p) + \operatorname{div}_-(q) + 2f/\lambda\right)_{i,j}}{1 + \tau \alpha_{i,j}(p,q)}, \;\; 1 \leq i,j \leq N, \tag{4.43a}$$

$$q_{i,j} = \frac{q_{i,j} + \tau \nabla_- \left(\operatorname{div}_+(p) + \operatorname{div}_-(q) + 2f/\lambda\right)_{i,j}}{1 + \tau \beta_{i,j}(p,q)}, \;\; 1 \leq i,j \leq N, \tag{4.43b}$$

where α and β are given by (4.42).

PROOF. First we observe that the equations (4.43) are equivalent to

$$\operatorname{grad}(F)(p,q) + \sum_{1\le i,j\le N} \alpha_{i,j}(p,q)\operatorname{grad}(g_{i,j}(p) + \beta_{i,j}(p,q)\operatorname{grad}(h_{i,j})(q) = 0, \tag{4.44}$$

and it follows from the latter that

$$\sum_{1\le i,j\le N} \alpha_{i,j}(p,q)g_{i,j}(p) + \beta_{i,j}(p,q)h_{i,j}(q) = 0; \tag{4.45}$$

hence we obtain the sufficient condition for the existence of a solution.

Conversely, if $(p,q) \in B_Y$ is a solution of (DM), then (p,q) is a solution of the constrained minimization problem

$$\begin{aligned} &\text{Minimize } F(p,q), \\ &\text{subject to: } g_{i,j}(p) \le 0,\ 1 \le i,j \le N, \\ &\qquad\qquad h_{i,j}(q) \le 0,\ 1 \le i,j \le N. \end{aligned} \tag{4.46}$$

Thus by Theorem 4.6, there exists $\alpha(p,q), \beta(p,q) \in \mathbb{R}_+^{N\times N}$ such that (4.45) and (4.44) hold. But under the condition $p,q \in B_Y$, (4.45) is equivalent to

$$\alpha_{i,j}(p,q)g_{i,j}(p) = 0 \text{ and } \beta_{i,j}(p,q)h_{i,j}(q) = 0,\ \ 1 \le i,j \le N. \tag{4.47}$$

Finally, combining the latter equation with (4.44), we obtain that $\alpha(p,q)$ and $\beta(p,q)$ are defined by the equations (4.42); and the proof is complete. □

The following fixed point algorithm is a direct consequence of the equations (4.44) and uses a Gauss-Seidel update technique.

ALGORITHM 4.17. *Choose $\tau > 0$ and $p^0, q^0 \in B_Y$.*

1. *For any $n \ge 0$, update the primal variable u:*

$$u^n = f + \frac{\lambda}{2}\left(\operatorname{div}_+(p^n) + \operatorname{div}_-(q^n)\right). \tag{4.48}$$

2. *Update the variables p, u, and q as follows:*

$$p_{i,j}^{n+1} = \frac{p_{i,j}^n + 2\tau/\lambda \nabla_+(u^n)_{i,j}}{1 + 2\tau/\lambda \left|\nabla_+(u^n)_{i,j}\right|}, \quad 1 \le i, j \le N, \tag{4.49a}$$

$$u^{n+\frac{1}{2}} = f + \frac{\lambda}{2}\left(\operatorname{div}_+(p^{n+1}) + \operatorname{div}_-(q^n)\right), \tag{4.49b}$$

$$q_{i,j}^{n+1} = \frac{q_{i,j}^n + 2\tau/\lambda \nabla_-(u^{n+\frac{1}{2}})_{i,j}}{1 + 2\tau/\lambda \left|\nabla_-(u^{n+\frac{1}{2}})_{i,j}\right|}, \quad 1 \le i, j \le N. \tag{4.49c}$$

3. *Until stopping criterion, increment $n \leftarrow n+1$ and return to* **1**.

REMARK 4.18. A similar algorithm was proposed in the literature [31] for the standard discrete ROF model (2.38). In this case, the only dual variable is p, so the second step in the above algorithm reduces to one update on the variable p.

LEMMA 4.19. *Let $0 < \tau \le h^2/8$ and $p^0, q^0 \in B_Y$ be given. Let (p^n, q^n) be the sequence defined in Algorithm 4.17. Then, the sequence $\{\| \operatorname{div}_+(p^n) + \operatorname{div}_-(q^n) + 2f/\lambda\|_X\}_n$ is monotonic nonincreasing.*

PROOF. Let $p^0, q^0 \in B_Y$ be fixed. An easy induction shows that $p^n, q^n \in B_Y$ for all $n \ge 0$. We now show that for $\tau \le h^2/8$, the sequence $\| \operatorname{div}_+(p^n) + \operatorname{div}_-(q^n) + 2f/\lambda\|_X$ is monotonic nonincreasing. Let n be fixed and define

$$\delta^n(p) = \frac{p^{n+1} - p^n}{\tau} \text{ and } \delta^n(q) = \frac{q^{n+1} - q^n}{\tau}.$$

Then, from (4.29) and (4.49), we have

$$\begin{aligned}\|\operatorname{div}_+(p^{n+1}) + \operatorname{div}_-(q^{n+1}) + \frac{2f}{\lambda}\|_X^2 &= \| \operatorname{div}_+(p^n) + \operatorname{div}_-(q^n) + \frac{2f}{\lambda}\|_X^2 + \frac{4\tau}{\lambda}\langle \operatorname{div}_+(\delta^n(p)), u^n\rangle \\ &+ \frac{4\tau}{\lambda}\langle \operatorname{div}_-(\delta^n(q)), u^{n+\frac{1}{2}}\rangle_X + \tau^2(\| \operatorname{div}_+(\delta^n(p))\|_X^2 + \| \operatorname{div}_-(\delta^n(q))\|_X^2) \\ &\le \| \operatorname{div}_+(p^n) + \operatorname{div}_-(q^n) + 2f/\lambda\|_X^2 \underbrace{-\frac{4\tau}{\lambda}\langle \delta^n(p), \nabla_+(u^n)\rangle_X + \tau^2 \| \operatorname{div}_+\|^2 \|\delta^n(p)\|_X^2}_{(I)} \\ &\underbrace{-\frac{4\tau}{\lambda}\langle \delta^n(q), \nabla_-(u^{n+\frac{1}{2}})\rangle_X + \| \operatorname{div}_-\|^2 \|\delta^n(q)\|_X^2}_{(II)}.\end{aligned}$$

Furthermore, we also have for all $1 \leq i, j \leq N$,

$$\delta^n(p)_{i,j} = \frac{2}{\lambda}\left(\nabla_+(u^n)_{i,j} - |\nabla_+(u^n)_{i,j}|p^{n+1}_{i,j}\right), \tag{4.50a}$$

$$\delta^n(q)_{i,j} = \frac{2}{\lambda}\left(\nabla_-(u^{n+\frac{1}{2}})_{i,j} - |\nabla_-(u^{n+\frac{1}{2}})_{i,j}|q^{n+1}_{i,j}\right), \tag{4.50b}$$

so that

$$\begin{aligned}
(I) &= -\sum_{1\leq i,j\leq N} \frac{4\tau}{\lambda}\delta^n(p)_{i,j}\nabla_+(u^n)_{i,j} - \tau^2\|\operatorname{div}_+\|^2|\delta^n(p)_{i,j}|^2 \\
&= \tau\sum_{1\leq i,j\leq N} |\delta^n(p)_{i,j} - \frac{2}{\lambda}\nabla_+(u^n)_{i,j}|^2 - |\frac{2}{\lambda}\nabla_+(u^n)_{i,j}|^2 - |\delta^n(p)_{i,j}|^2 + \tau\|\operatorname{div}_+\|^2|\delta^n(p)_{i,j}|^2 \\
&= \tau\sum_{1\leq i,j\leq N} (\tau\|\operatorname{div}_+\|^2 - 1)|\delta^n(p)_{i,j}|^2 + (\underbrace{|p^{n+1}_{i,j}|^2 - 1}_{\leq 0})|\frac{2}{\lambda}\nabla_+(u^n)_{i,j}|^2.
\end{aligned}$$

Therefore, using the estimate $\|\operatorname{div}_+\|^2 \leq 8/h^2$ obtained in the proof of Lemma 4.11, we get

$$-\frac{4\tau}{\lambda}\langle\delta^n(p), \nabla_+(u^n)\rangle_X + \tau^2\|\operatorname{div}_+\|^2\|\delta^n(p)\|_X^2 \leq 0 \quad \text{if } \tau \leq h^2/8.$$

Likewise, we show that

$$-\frac{4\tau}{\lambda}\langle\delta^n(q), \nabla_-(u^{n\frac{1}{2}})\rangle_X + \tau^2\|\operatorname{div}_-\|^2\|\delta^n(q)\|_X^2 \leq 0 \quad \text{if } \tau \leq h^2/8.$$

Hence, assuming that $\tau \leq h^2/8$, we get that for all $n \in \mathbb{N}$

$$\|\operatorname{div}_+(p^{n+1}) + \operatorname{div}_-(q^{n+1}) + 2f/\lambda\|_X \leq \|\operatorname{div}_+(p^n) + \operatorname{div}_-(q^n) + 2f/\lambda\|_X. \tag{4.51}$$

□

We are ready to prove the convergence of the proposed algorithm. More precisely, we have the following result.

THEOREM 4.20. *Let $0 < \tau \leq h^2/8$ and $p^0, q^0 \in B_Y$ be given. Then, every cluster point of the sequence $\{(p^n, q^n)\}_{n\geq 1}$ is a solution of the dual problem* (DM)*, and the sequence $\{u^n\}_{n\geq 1}$ converges to the solution $z^{f,h}$ of the primal problem* (PM).

PROOF. Let $0 < \tau \le h^2/8$ be given. Let (p^{n_k}, q^{n_k}) be a subsequence of (p^n, q^n) that converges to $(\bar{p}, \bar{q})$ Then, by the equations (4.49), the sequence (p^{n_k+1}, q^{n_k+1}) converges to the point $(\tilde{p}, \tilde{q})$ defined as follows:

$$\begin{aligned} \tilde{p}_{i,j} &= \frac{\bar{p}_{i,j} + \tau \nabla_+(\operatorname{div}_+(\bar{p}) + \operatorname{div}_-(\bar{q}) + 2f/\lambda)_{i,j}}{1 + \tau |\nabla_+(\operatorname{div}_+(\bar{p}) + \operatorname{div}_-(\bar{q}) + 2f/\lambda)_{i,j}|}, \quad 1 \le i,j \le N, \\ \tilde{q}_{i,j} &= \frac{\bar{q}_{i,j} + \tau \nabla_-(\operatorname{div}_+(\bar{p}) + \operatorname{div}_-(\bar{q}) + 2f/\lambda)_{i,j}}{1 + \tau |\nabla_-(\operatorname{div}_+(\bar{p}) + \operatorname{div}_-(\bar{q}) + 2f/\lambda)_{i,j}|}, \quad 1 \le i,j \le N. \end{aligned} \tag{4.52}$$

Furthermore, by Lemma 4.19, we know that

$$\|\operatorname{div}_+(\tilde{p}) + \operatorname{div}_-(\tilde{q}) + 2f/\lambda\|_X = \|\operatorname{div}_+(\bar{p}) + \operatorname{div}_-(\bar{q}) + 2f/\lambda\|_X. \tag{4.53}$$

Now, repeating the core computation in the proof of Lemma 4.19 with the pairs $(\tilde{p}, \tilde{q})$ and $(\bar{p}, \bar{q})$, we get that for any $\tau < \min(1/\|\operatorname{div}_+\|^2, 1/\|\operatorname{div}_-\|^2)$

$$\tilde{p}_{i,j} = \bar{p}_{i,j} \text{ and } \tilde{q}_{i,j} = \bar{q}_{i,j}, \quad \forall\, 1 \le i,j \le N. \tag{4.54}$$

The case $\tau = 1/\|\operatorname{div}_+\|^2 = 1/\|\operatorname{div}_-\|^2$ requires a more delicate analysis that we now undertake. We have

$$\begin{aligned} &|\tilde{p}_{i,j}| = 1 \text{ or } |\nabla_+(\operatorname{div}_+(\bar{p}) + \operatorname{div}_-(\bar{q}) + 2f/\lambda)_{i,j}| = 0, \quad \forall\, 1 \le i,j \le N, \\ &|\tilde{q}_{i,j}| = 1 \text{ or } |\nabla_-(\operatorname{div}_+(\bar{p}) + \operatorname{div}_-(\bar{q}) + 2f/\lambda)_{i,j}| = 0, \quad \forall\, 1 \le i,j \le N. \end{aligned} \tag{4.55}$$

Clearly, only the pairs (i,j) for which we have exclusively $|\tilde{p}_{i,j}| = 1$ or $|\tilde{q}_{i,j}| = 1$ are worth pursuing further. If $|\tilde{p}_{i,j}| = 1$, then by (4.52) and the Cauchy-Schwarz inequality, we have

$$\begin{aligned} 1 + 2\tau|\nabla_+(\operatorname{div}_+(\bar{p}) + \operatorname{div}_-(\bar{q}) + \frac{2f}{\lambda})_{i,j}| &= |\bar{p}_{i,j}|^2 + 2\tau\langle \bar{p}_{ij}, \nabla_+(\operatorname{div}_+(\bar{p}) + \operatorname{div}_-(\bar{q}) + \frac{2f}{\lambda})_{i,j}\rangle \\ &\le |\bar{p}_{i,j}|^2 + 2\tau|\bar{p}_{ij}||\nabla_+(\operatorname{div}_+(\bar{p}) + \operatorname{div}_-(\bar{q}) + 2f/\lambda)_{i,j}|, \end{aligned}$$

Since we assumed that $|\nabla_+(\operatorname{div}_+(\bar{p}) + \operatorname{div}_-(\bar{q}) + 2f/\lambda)_{i,j}| \ne 0$, it now follows from the latter inequality that $|\bar{p}_{ij}| = 1$, so that $|\tilde{p}_{ij}| = |\bar{p}_{ij}| = 1$ and (4.52) imply

$$\tilde{p}_{i,j} = \nabla_+(\operatorname{div}_+(\bar{p}) + \operatorname{div}_-(\bar{q}) + 2f/\lambda)_{i,j};$$

thus by (4.52) again, we get

$$\tilde{p}_{i,j} = \frac{\bar{p}_{ij}(1+\tau)}{1+\tau} = \bar{p}_{ij}.$$

In conclusion, we have shown that for $0 < \tau \leq 1/\| \operatorname{div}_+ \|$, the sequence (p^{n_k+1}, q^{n_k+1}) converges to $(\bar{p}, \bar{q})$ so that the equation (4.52) now reads

$$\begin{aligned} \bar{p}_{i,j} &= \frac{\bar{p}_{i,j} + \tau \nabla_+ (\operatorname{div}_+(\bar{p}) + \operatorname{div}_-(\bar{q}) + 2f/\lambda)_{i,j}}{1 + \tau |\nabla_+ (\operatorname{div}_+(\bar{p}) + \operatorname{div}_-(\bar{q}) + 2f/\lambda)_{i,j}|}, \quad 1 \leq i,j \leq N, \\ \bar{q}_{i,j} &= \frac{\bar{q}_{i,j} + \tau \nabla_- (\operatorname{div}_+(\bar{p}) + \operatorname{div}_-(\bar{q}) + 2f/\lambda)_{i,j}}{1 + \tau |\nabla_- (\operatorname{div}_+(\bar{p}) + \operatorname{div}_-(\bar{q}) + 2f/\lambda)_{i,j}|}, \quad 1 \leq i,j \leq N. \end{aligned} \tag{4.56}$$

Therefore by Theorem 4.16, $(\bar{p}, \bar{q})$ is a solution of (DM); and by Theorem 4.7, the sequence $\{u^n\}$ converges to the solution $z^{f,h}$ of (PM). Since $\| \operatorname{div}_+ \| \leq 8/h^2$, the result remains true for $0 < \tau \leq h^2/8$ and the proof is complete □

4.4 NUMERICAL EXPERIMENTS

In this section, we report the results of numerical experiments with the three algorithms presented in the previous section. The test images that we used are found in Figure 4.1. We use the following abbreviations to identify the three algorithms under consideration here.

ALG1: The projected-gradient algorithm 4.12

ALG2: The alternating projected-gradient algorithm 4.15.

ALG3: The fixed point algorithm 4.17.

It should be noted that in our tests, we did not attempt to choose the parameters τ and λ for optimal performance of the algorithms. The algorithms are implemented in the MATLAB® [72] programming language.

Table 4.1 through Table 4.4 below show the capability of Algorithm 4.12 to remove noise for various noise levels. The inputs for all four algorithms are obtained by adding a zero mean Gaussian noise with standard deviation σ to the images in Figure 4.1. The parameter τ is set to $1/8$ in all three algorithms and the maximum number of iterations is set to 1000. Each algorithm is terminated when the change in mean square error at consecutive steps is below 10^{-8}. The numbers in each column identifying an algorithm are the Peak-Signal-to-Noise-Ratios, measured relative to the ground truth images in Figure 4.1, with the number of iterations and the CPU time used to reach that value in parenthesis.

(A) Lena

(B) Peppers

(C) Bank.

(D) Boats

FIGURE 4.1: The images used in the numerical experiments. The images in the top row are of size 256×256, while those in the bottom row have resolution 512×512.

λ	σ	ALG1	ALG2	ALG3
$\frac{1}{24}$	15	$31.1832(1000,\ 31s)$	$31.1951(72,\ 2.5s)$	$31.1872(1000,\ 32s)$
$\frac{1}{16}$	20	$29.6665(1000,\ 31s)$	$29.6820(101,\ 3.5s)$	$29.6742(1000,\ 32s)$
$\frac{1}{8}$	25	$27.5282(1000,\ 30s)$	$27.5323(806,\ 27.5s)$	$27.5681(1000,\ 32s)$

TABLE 4.1: Comparison of the algorithms using the image in Figure 4.1a. The numbers are PSNR(number of iterations, CPU time in seconds).

λ	σ	ALG1	ALG2	ALG3
$\frac{1}{24}$	15	$31.8105(1000,\ 31s)$	$31.8653(36, 1.5s)$	$31.8157(1000,\ 32s)$
$\frac{1}{16}$	20	$30.3494(1000,\ 34s)$	$30.3587(345,\ 12s)$	$30.3603(1000,\ 31s)$
$\frac{1}{8}$	25	$28.3146(1000,\ 30s)$	$28.3206(696,\ 24s)$	28.3606(1000, 32s)

TABLE 4.2: Comparison of the algorithms using the image in Figure 4.1b. The numbers are PSNR(number of iterations, CPU time in seconds).

λ	σ	ALG1	ALG2	ALG3
$\frac{1}{24}$	15	$31.4563\,(635,\ 103s)$	$31.4520(209,\ 43s)$	$31.4571(1000,\ 164s)$
$\frac{1}{16}$	20	$29.9481(620,\ 101s)$	$29.9446(292,\ 61s)$	$29.9496(1000,\ 166s)$
$\frac{1}{8}$	25	$27.8264(716,\ 115s)$	$27.8239(477,\ 95s)$	$27.8384(1000,\ 114s)$

TABLE 4.3: Comparison of the algorithms using the image in Figure 4.1c. The numbers are PSNR(number of iterations, CPU time in seconds).

λ	σ	ALG1	ALG2	ALG3
$\frac{1}{24}$	15	$30.5134(1000,\ 158s)$	$30.5148(102,\ 22s)$	$30.5166(1000,\ 164s)$
$\frac{1}{16}$	20	$29.1286(1000,\ 169s)$	$29.1311(127,\ 27s)$	$29.1350(1000,\ 167s)$
$\frac{1}{8}$	25	$26.9855(1000,\ 158s)$	$26.9904(663,\ 122s)$	$27.0197(1000,\ 165s)$

TABLE 4.4: Comparison of the algorithms using the image in Figure 4.1d. The numbers are PSNR(number of iterations, CPU time in seconds).

Our numerical experiments suggest that the alternating projected-gradient algorithm is the most of efficient of the three algorithms. For moderate noise levels, the speed up is a few order of magnitudes. To further confirm this observation, we compared ALG2 and ALG3 to ALG1, see Figure 4.2. The comparison is done as follows: (1) First we generate a ground truth by running 10^5 iterations of ALG1. (2) We find the number of iterations that each of ALG2 and ALG3 is going to use to get to within 10^{-13} of the ground truth computed using ALG1.

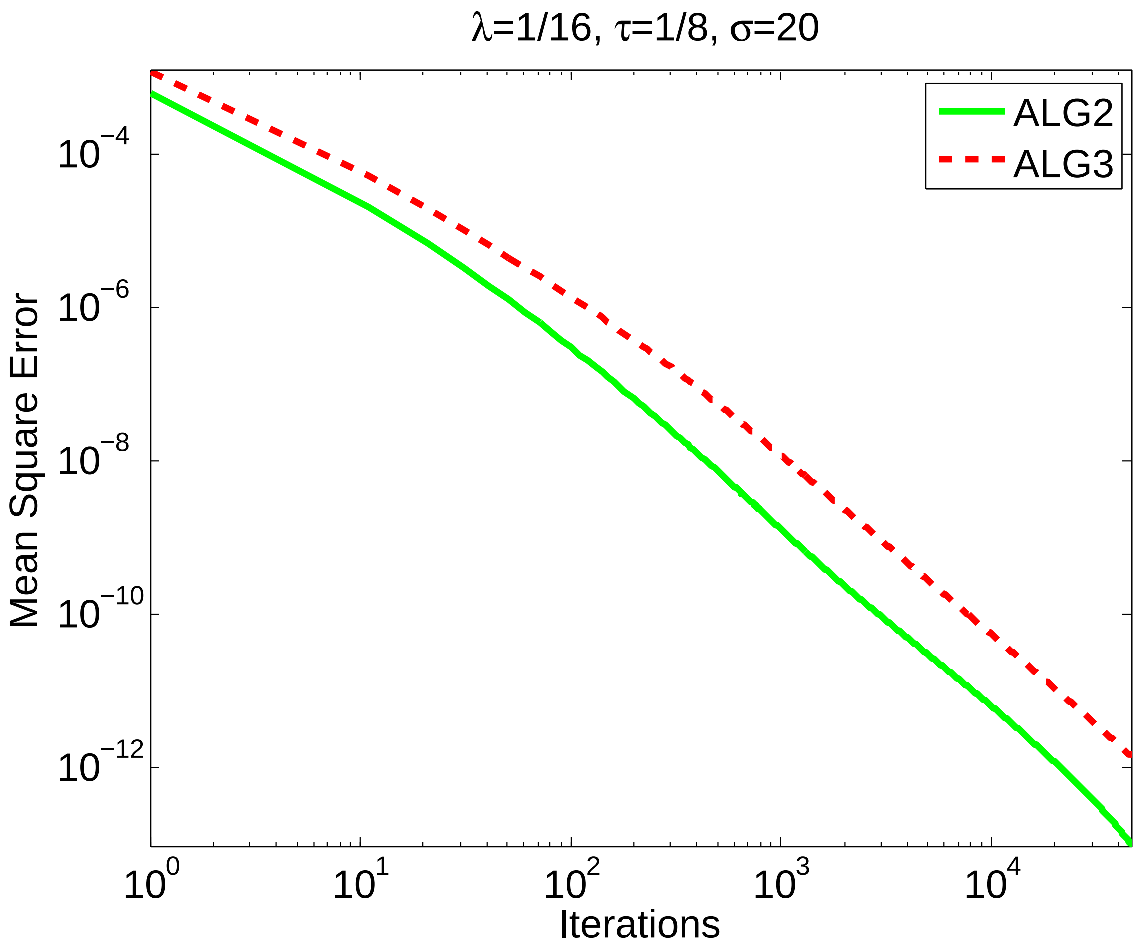

FIGURE 4.2: The alternating projected gradient algorithm (ALG2) takes about 46000 iterations to get to within 10^{-13} of the solution of ALG1 generated with 10^5 iterations. ALG2 is consistently faster than ALG3 and ALG1.

5

APPROXIMATION BY BIVARIATE SPLINES ON ARBITRARY POLYGONAL DOMAINS

In this chapter, we investigate the approximation of the total variation based denoising model on non rectangular polygonal domains. We assume that Ω is a polygonal domain endowed with a quasi-regular family of triangulations $\{\Delta_h\}_{h>0}$ and construct a bivariate spline approximation of the minimizer of the ROF functional.

Until now, the preferred approach in the approximation of the ROF model has been to used relaxations [1, 35] of the functional $E_\lambda^f(u)$. In a departure with that tradition, we use the Galerkin method on a suitable lattice of bivariate spline spaces to directly generate a minimizing sequence for $E_\lambda^f(u)$.

Motivated by the work of Acar and Vogel [1], Hong [48] investigated the use of bivariate spline spaces for image enhancement using the minimal surface functional

$$\mathcal{E}_\lambda^f(u) := \int_\Omega \sqrt{1+|\nabla u|^2}dx + \frac{1}{2\lambda}\int_\Omega |u-f|^2 dx. \tag{5.1}$$

Although the latter functional is defined on $BV(\Omega)$, she only constructed a minimizing spline sequence in the case where the minimizer of $\mathcal{E}_\lambda^f(u)$ belong to the Sobolev space $\mathrm{W}^{1,1}(\Omega)$. The main result of this chapter is a significant improvement of her work. We apply the Galerkin method with continuous bivariate spline space, directly on the ROF functional,

$$E_\lambda^f(u) := \lambda J(u) + \frac{1}{2}\int_\Omega |u-f|^2 dx,$$

to obtain a minimizing sequence of $E_\lambda^f(u)$ for any $f \in L^2(\Omega)$.

5.1 SPLINE FUNCTIONS ON TRIANGULATIONS

In this section, we review the concept of bivariate spline function and cover the properties of spline functions that we will use in the chapter. Throughout the section and the chapter, we assume that Ω is a polygonal domain, possibly non-rectangular, endowed with a triangulation $\Delta_h,\ 0 < h \ll 1$ such that no triangle in Δ_h has an edge with length greater than h. The presentation follows the monograph [54].

DEFINITION 5.1. Suppose that Δ_h is a triangulation on Ω and d is a fixed natural number.

(a) A spline function on the triangulation Δ_h is a function s defined on Ω such that for any triangle $T \in \Delta_h$, $s|_T$ is a polynomial.

(b) We say that a spline function s is of degree d, if $s|_T$ is a polynomial of degree less than or equal to d for any $T \in \Delta_h$. We denote the set of spline functions of degree d by

$$\mathcal{S}_d^{-1}(\Delta_h) := \{s : \Omega \to \mathbb{R} \colon s|_T \in \mathbb{P}_d \ \ \forall T \in \Delta_h\},$$

where $\mathbb{P}_d$ is the vector space of bivariate polynomials of degree less than or equal to d.

(c) The space of smooth spline functions of order r and degree d is defined by

$$\mathcal{S}_d^r(\Delta_h) = C^r(\Omega) \cap \mathcal{S}_d^{-1}(\Delta_h) = \{s \in C^r(\Omega) \colon s|_T \in \mathbb{P}_d, \ \ \forall T \in \Delta_h\}.$$

The space of bivariate spline functions $\mathcal{S}_d^{-1}(\Delta_h)$ is isomorphic to $\mathbb{R}^N$ where N depends on d and the number of triangles in Δ_h. The space of smooth splines of order r is characterized as the solution set of a rectangular system of linear equations enforcing the smoothness at the interior edges of the triangulation. A Convenient representation of spline spaces as subspaces of an $\mathbb{R}^N$ are obtained through the so called Bernstein-Bezier representation.

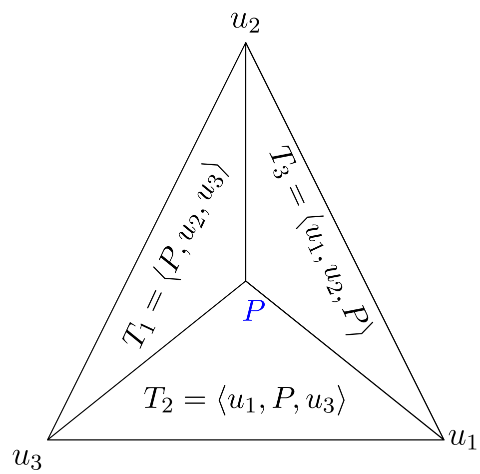

FIGURE 5.1: The i-th barycentric coordinates of P is the ratio of the area of the triangle T_i and that of the triangle $T = \langle u_1, u_2, u_3 \rangle$.

5.1.1 BERNSTEIN-BEZIER REPRESENTATION

The Bernstein-Bezier representation of bivariate spline functions of degree d is built upon the Bernstein-Bezier basis, hereafter B-basis, of $\mathbb{P}_d$, the space of polynomials of degree less than or equal to d. To define the B-basis, we need the concept of barycentric coordinates. Suppose $T = \langle u_1(x_1, y_1), u_2(x_2, y_2), u_3(x_3, y_3) \rangle$ is a triangle of the affine plane $\mathbb{R}^2$. Let $P(x, y)$ be an arbitrary point in the plane.

DEFINITION 5.2. The barycentric coordinates of P relative to T are the triplet $(\lambda_1, \lambda_2, \lambda_3)$ solution of the linear system

$$\begin{cases} \lambda_1 & + \lambda_2 & + \lambda_3 & = 1 \\ \lambda_1 x_1 & + \lambda_2 x_2 & + \lambda_3 x_3 & = x \\ \lambda_1 y_1 & + \lambda_2 y_2 & + \lambda_3 y_3 & = y. \end{cases} \tag{5.2}$$

Let $d \in \mathbb{N}$ be fixed. The points $\{\xi_{ijk}^T \colon i + j + k = d\} \subset \mathbb{R}^2$ defined by

$$\xi_{ijk}^T = \frac{iu_1 + ju_2 + ku_3}{d}, \quad i + j + k = d$$

are called the domain points of T; their barycentric coordinates with respect to T are given by $\{(i/d, j/d, k/d) \colon i+j+k=d\}$.

DEFINITION 5.3. A Bernstein-Bezier bivariate polynomial of degree d relative to T is defined by

$$B_{ijk}^{T,d}(x,y) = \frac{d!}{i!j!k!}\lambda_1^i\lambda_2^j\lambda_3^k \qquad i+j+k=d, \tag{5.3}$$

where $(\lambda_1, \lambda_1, \lambda_3)$ are the barycentric coordinates of $P(x,y)$ relative to T.

It is easy to show using the multinomial theorem and the equations (5.2) that the collection of Bernstein-Bezier polynomials $\{B_{ijk}^{T,d} \colon i+j+k=d\}$ form a basis of $\mathbb{P}_d$ and the following partition of unity holds

$$\sum_{i+j+k=d} B_{ijk}^{T,d}(x,y) = 1, \quad \forall (x,y) \in \mathbb{R}^2. \tag{5.4}$$

As a consequence, any bivariate spline function $s \in \mathcal{S}_d^{-1}(\Delta_h)$ is uniquely represented by a tuple $(c_{ijk}^T \colon T \in \Delta_h,\ i+j+k=d)$ such that

$$s|_T = \sum_{i+j+k=d} c_{ijk}^T B_{ijk}^{T,d}, \quad \forall\, T \in \Delta_h. \tag{5.5}$$

Thus, the space of bivariate spline functions on Ω with respect to Δ_h, $\mathcal{S}_d^{-1}(\Delta_h)$, is identified to an $\mathbb{R}^N$ where $N = \#(\Delta_h)\binom{d+2}{2}$ and $\#(\Delta_h)$ is the number of triangles in Δ_h. Moreover, the space of smooth bivariate splines, $\mathcal{S}_d^r(\Delta_h)$, is a subspace of $\mathbb{R}^N$ of the form

$$\mathcal{S}_d^r(\Delta_h) = \{\boldsymbol{c} \in \mathbb{R}^N \colon A(r)\boldsymbol{c} = \boldsymbol{0}\}, \tag{5.6}$$

where $A(r)$ is an $(r+1)(d+1)E \times N$ matrix encoding the smoothness condition across the interior edges of the triangulation Δ_h, and E is the number of interior edges of Δ_h.

When solving variational equation on bivariate spline spaces, we often need to compute the inner product between two bivariate spline functions. The next result gives a simple formula for evaluating the inner product when the spline functions are expressed in B-form.

THEOREM 5.4 ([54]). *Let T be a triangle in $\mathbb{R}^2$, and $d \in \mathbb{N}$ be given. If $p = \sum_{i+j+k} c_{ijk} B_{ijk}^{T,d}$ and $p = \sum_{i+j+k} \tilde{c}_{ijk} B_{ijk}^{T,d}$ are bivariate polynomials of degree $\leq d$ written in B-form, then*

$$\int_T p(x,y)q(x,y)\,dxdy = \frac{A_T}{\binom{2d}{d}\binom{2d+2}{2}} \sum_{\substack{i+j+k=d \\ \nu+\mu+\kappa=q}} c_{ijk}\tilde{c}_{\nu\mu\kappa} \binom{i+\nu}{i}\binom{j+\mu}{j}\binom{k+\kappa}{k}, \quad (5.7)$$

where A_T is the area of T and $\binom{a}{b}$ is the number of combinations of a objects chosen b at a time.

For computational purposes, the latter inner product may be written in condensed form as

$$\int_T p(x,y)q(x,y)\,dxdy = \frac{A_T}{\binom{2d}{d}\binom{2d+2}{2}} \tilde{\boldsymbol{c}}^T G(d)\boldsymbol{c},$$

where $\tilde{\boldsymbol{c}}$ and $\boldsymbol{c}$ as the coefficients of q and p in the Bernstein-Bezier basis and G is a $\binom{d+2}{2}$ square matrix that may be preassembled in a computational library.

Later when we construct a spline minimizing sequence of the ROF model. we will need a convenient family of triangulation that ensure the convergence of our sequence. We now introduce a definition of such a convenient family.

DEFINITION 5.5. We say that a family of triangulations $\{\Delta_h \colon h \in I \subset \mathbb{R}_+\}$ of Ω is quasi-regular if there exists $\beta > 0$ such that

$$\frac{\operatorname{diam}(T)}{\rho_T} < \beta, \quad \forall\, T \in \Delta_h,\ \forall\, h \in I,$$

where $\operatorname{diam}(T)$ is the longest edge of T and ρ_T is the radius of the incircle of T.

Constructing such families of triangulations is not too hard. It is sufficient to ensure that the smallest angle in the triangulations remains bounded away form zero as the triangulation size goes to zero.

EXAMPLE 5.6. Let $h_0 > 0$ fixed and Δ_{h_0} a triangulation with mesh size h_0 and smallest angle θ_0. We construct a family of triangulations Δ_n with mesh size $h_0/2^n$ iteratively as follows: Given Δ_n, we obtain Δ_{n+1} by subdividing each triangle $T \in \Delta_n$ into four triangles by connecting the midpoints of the edges of T as illustrated in the figure below. The resulting family of triangulations $\{\Delta_n : n \in \mathbb{N}\}$ is quasi-regular with constant

$$\beta = \frac{2}{\sin(\theta_0/2)}.$$

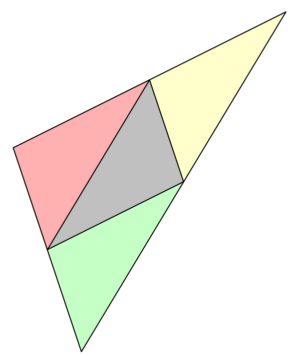

FIGURE 5.2: Midpoints refinement of a triangle into four smaller triangles.

5.1.2 RELEVANT PROPERTIES OF BIVARIATE SPLINES

Spline functions have been used with much success in the numerical computation of partial differential equations [51, 52, 55–57] and more recently for the numerical simulation of the Darcy-Stokes equation [16]. In general, splines function may be used under Galerkin methods to approximate variational equations over function spaces that are well approximated by spline functions. Their appeal to us in this work is twofold:

(a) Bivariate spline functions yield good approximation of Sobolev functions, *i.e*, functions that are elements of the Sobolev spaces $\mathrm{W}^{m,p}(\Omega)$, $m \in \mathbb{N}$ and $p \geq 1$.

(b) The derivative operators $D_1^{\alpha} D_2^{\beta}$ are bounded linear operators between the spaces $\mathcal{S}_d^{-1}(\Delta_h)$ and $\mathcal{S}_{d-\alpha-\beta}^{-1}(\Delta_h)$. We will refer to this property as Markov Inequality.

THEOREM 5.7 (Markov inequality [54, Theorem 2.32]). *Suppose a quasi-regular triangulation, Δ_h, of Ω is given, and $p \in [1, \infty)$ and $d \in \mathbb{N}$ be fixed. There exists a constant K depending only on d such that for all nonnegative integers α and β with $0 \le \alpha + \beta \le d$, we have*

$$\|D_1^\alpha D_2^\beta s\|_{L^p} \le \frac{K}{\rho^{\alpha+\beta}} \|s\|_{L^p}, \quad \forall s \in \mathcal{S}_d^{-1}(\Delta_h), \tag{5.8}$$

where $\rho = \min\{\rho_T : T \in \Delta_h\}$ and ρ_T is the radius of the incircle to the triangle T.

The next result gives the approximation power of the space of continuous spline functions in Sobolev spaces; we use this result in our construction of a spline minimizing sequence for the ROF functional.

THEOREM 5.8 ([54, Theorem 10.2, p. 277]). *Suppose that Δ_h is a quasi-regular triangulation of Ω, and let $p \in [1, \infty]$ and $d \in \mathbb{N}$ be given. Then for every $u \in W^{d+1,p}(\Omega)$, there exists a spline $s_u \in \mathcal{S}_d^0(\Delta_h)$ such that*

$$\|D_1^\alpha D_2^\beta (u - s_u)\|_{L^p} \le K h^{d+1-\alpha-\beta} |u|_{d+1,p} \ \forall\, 0 \le \alpha + \beta \le d, \tag{5.9}$$

where K depends only on d and the smallest angle in Δ_h, and

$$|u|_{d+1,p} = \sum_{\alpha+\beta=d+1} \|D_1^\alpha D_2^\beta u\|_{L^p}.$$

5.2 APPROXIMATION OF THE ROF MODEL BY CONTINUOUS SPLINES

In this section, we describe how we arrive at a family of continuous bivariate splines that approximate the minimizer of the functional

$$E_\lambda^f(u) := \lambda J(u) + \frac{1}{2} \int_\Omega |u - f|^2 dx, \quad u \in BV(\Omega). \tag{5.10}$$

The approximation of the minimizer of the above functional by continuous splines is possible because the space $\mathcal{S}_d^0(\Delta_h)$ possesses very good approximation power in high order Sobolev spaces as illustrated by Theorem 5.8. In using the approximation power of spline functions exhibited by Theorem 5.8, we will need to control the norm of high order derivatives of the mollification of a BV function. This is done as in the lemma below.

LEMMA 5.9. *Suppose that Ω is a bounded polygonal domain. Let $u \in BV(\Omega)$ be fixed. Then for any integer $m \geq 0$, any pair of nonnegative integer (α, β) such that $\alpha+\beta = m+1$, and any $\epsilon \in (0,1)$, we have*

$$\left\| D_1^{\alpha} D_2^{\beta}(\eta_\epsilon * Tu) \right\|_{L^1(\Omega)} \leq \frac{C}{\epsilon^m} |DTu|(\mathbb{R}^2), \tag{5.11}$$

where C is a constant depending only on α, β and Ω, and $T : BV(\Omega) \to BV(\mathbb{R}^2)$ is the extension operator guaranteed by Theorem A.8.

PROOF. Let m a nonnegative integer be fixed, and $\epsilon \in (0,1)$ be given. Let $u \in BV(\Omega)$ and $\varphi \in C_c^1(\Omega)$ be given. Let α and β be two nonnegative integers such that $\alpha + \beta = m + 1$; we may assume without loss of generality that $\alpha \geq 1$.Then, we have

$$\begin{aligned}
\int_\Omega D_1^{\alpha} D_2^{\beta}(\eta_\epsilon * Tu)\varphi dx &= -\int_{\mathbb{R}^2} D_1^{\alpha-1} D_2^{\beta}(\eta_\epsilon * Tu)\frac{\partial \varphi}{\partial x_1} dx \\
&= -\int_{\mathbb{R}^2} D_1^{\alpha-1} D_2^{\beta}\eta_\epsilon * Tu \, \frac{\partial \varphi}{\partial x_1} dx \\
&= -\int_{\mathbb{R}^2} Tu \, \check{\eta}_\epsilon^m * \frac{\partial \varphi}{\partial x_1} dx \text{ with } \check{\eta}_\epsilon^m(x) = D_1^{\alpha-1} D_2^{\beta}\eta_\epsilon(-x) \\
&= -\int_{\mathbb{R}^2} Tu \, \frac{\partial}{\partial x_1}[\check{\eta}_\epsilon^m * \varphi] dx.
\end{aligned}$$

Thus

$$\int_\Omega D_1^{\alpha} D_2^{\beta}(\eta_\epsilon * Tu) \, \varphi dx \leq \|\check{\eta}_\epsilon^m * \varphi\|_\infty |DTu|(\mathbb{R}^2).$$

Now, by Young's inequality, we have

$$\|\check{\eta}_\epsilon^m * \varphi\|_\infty \leq \|\check{\eta}_\epsilon^m\|_{L^2(\mathbb{R}^2)} \|\varphi\|_{L^2(\Omega)},$$

and a simple computation shows that

$$\|\check{\eta}_\epsilon^m\|_{L^2(\mathbb{R}^2)}^2 \leq \frac{\sqrt{\pi}}{\epsilon^m} \left\| D_1^{\alpha-1} D_2^{\beta}\eta \right\|_\infty^{1/2} \text{ and } \|\varphi\|_{L^2(\Omega)} \leq \sqrt{|\Omega|}\|\varphi\|_\infty,$$

where $|\Omega|$ denotes the area of Ω. Consequently, we obtain

$$\int_\Omega D_1^\alpha D_2^\beta(\eta_\epsilon * Tu)\varphi dx \leq \frac{C(\alpha,\beta,\rho)}{\epsilon^m}\|\varphi\|_\infty |DTu|(\mathbb{R}^2) \tag{5.12}$$

where

$$C(\alpha,\beta,\eta) = \sqrt{\pi|\Omega|}\left\|D_1^{\alpha-1}D_2^\beta\eta\right\|_\infty^{1/2}. \tag{5.13}$$

On taking the supremum in (5.12) over all $\varphi \in C_c^1(\Omega)$ such that $\|\varphi\|_\infty \leq 1$, we obtain by duality and a denseness argument that

$$\left\|D_1^\alpha D_2^\beta(\eta_\epsilon * Tu)\right\|_{L^1(\Omega)} \leq \frac{C(\alpha,\beta,\eta)}{\epsilon^m}|DTu|(\mathbb{R}^2).$$

□

Suppose that Ω is endowed with a quasi-regular triangulation Δ_h, and let $d \in \mathbb{N}$ be given. As a finite dimensional space, $\mathcal{S}_d^0(\Delta_h)$ is a closed and convex subset of $L^2(\Omega)$. Thus, the ROF functional has a unique minimizer in $\mathcal{S}_d^0(\Delta_h)$. Let $s_h^d(f)$ be the spline function defined by

$$s_h^d(f) = \underset{u\in\mathcal{S}_d^0(\Delta_h)}{\arg\min}\, \lambda J(u) + \frac{1}{2}\int_\Omega |u-f|^2 dx. \tag{5.14}$$

We are ready to prove that our construction of minimum splines above yields a minimizing sequence for the ROF functional. Let h_n be a monotonically decreasing sequence of real numbers such that $h_n \searrow 0$. Let Δ_n be a quasi-regular triangulation with mesh size h_n and smallest angle θ_n. We have the following result:

THEOREM 5.10. *Suppose that the sequence of triangulations $\{\Delta_n\}_n$ is such that*

$$\inf_{n\in\mathbb{N}} \theta_n > \theta > 0. \tag{5.15}$$

Given $d \in \mathbb{N}$, the sequence $\{s_n^d(f)\}_n$ defined by

$$s_n^d(f) = \underset{u\in\mathcal{S}_d^0(\Delta_n)}{\arg\min}\, \lambda J(u) + \frac{1}{2}\int_\Omega |u-f|^2 dx \tag{5.16}$$

is minimizing for the ROF functional $E_\lambda^f(u)$.

PROOF. To begin with, choose a finite rectangular covering, $\{R_i \colon i = 1, 2, \dots, N\}$, of the boundary of Ω, and let $T \colon BV(\Omega) \to BV(\mathbb{R}^2)$ be the corresponding extension operator, the existence of which is guaranteed by Theorem A.8. We recall that T is also a bounded linear operator from $W^{1,1}(\Omega)$ into $W^{1,1}(\mathbb{R}^2)$, and for any $u \in BV(\Omega)$, Tu is supported on the relatively compact open set $\Omega \cup \bigcup_{i=1}^{N} R_i$.

Let $0 < \epsilon < 1$ and $d \in \mathbb{N}$ be fixed. Let $u_\epsilon^f = \eta_\epsilon * Tu^f$ and $s_\epsilon^f \in \mathcal{S}_d^0(\Delta_n)$ be as in Theorem 5.8. Then by Lemma 5.9, we have

$$\|u_\epsilon^f - s_\epsilon^f\|_{W^{1,1}(\Omega)} \le C(d, \theta) \left(\frac{h_n}{\epsilon}\right)^d, \tag{5.17}$$

where C depends solely on d and θ. Moreover, since $T : W^{1,1}(\Omega) \to W^{1,1}(\mathbb{R}^2)$ is linear and bounded, and Tu is compactly supported for every u, it follows from the BV version of the Sobolev's inequality [see 46, Theorem 1.28, p.24] that

$$\begin{aligned}\|u_\epsilon^f - s_\epsilon^f\|_{L^2(\Omega)} &\le \|T(u_\epsilon^f - s_\epsilon^f)\|_{L^2(\mathbb{R}^2)} \le C \int_{\mathbb{R}^2} |\nabla(T(u_\epsilon^f - s_\epsilon^f))| dx \\ &\le C\|T(u_\epsilon^f - s_\epsilon^f)\|_{W^{1,1}(\mathbb{R}^2)} \le C\|T\|_* \|u_\epsilon^f - s_\epsilon^f\|_{W^{1,1}(\Omega)},\end{aligned} \tag{5.18}$$

with C a universal constant depending only on Ω and the covering $\{R_i \colon i = 1, 2, \dots, N\}$, and $\|T\|_*$ is the operator norm of T.

We now proceed to show that by choosing a suitable regularization scale ϵ, we achieve the convergence of $E_\lambda^f(s_n^d(f))$ to $E_\lambda^f(u^f)$ as $n \to \infty$. In fact, we have

$$\begin{aligned}E_\lambda^f(s_n^d(f)) - E_\lambda^f(u^f) &= \underbrace{E_\lambda^f(s_n^d(f)) - E_\lambda^f(s_\epsilon^f)}_{\le 0} + E_\lambda^f(s_\epsilon^f) - E_\lambda^f(u_\epsilon^f) + E_\lambda^f(u_\epsilon^f) - E_\lambda^f(u^f) \\ &\le E_\lambda^f(s_\epsilon^f) - E_\lambda^f(u_\epsilon^f) + E_\lambda^f(u_\epsilon^f) - E_\lambda^f(u^f).\end{aligned}$$

So to finish the proof, it suffices to show that $E_\lambda^f(u_\epsilon^f) \to E_\lambda^f(u^f)$ and $E_\lambda^f(s_\epsilon^f) \to E_\lambda^f(u_\epsilon^f)$ as $n \to \infty$. First, we observe that the convergence of $E_\lambda^f(u_\epsilon^f)$ to $E_\lambda^f(u^f)$ follows from the fact that $u_\epsilon^f \xrightarrow{L^2(\Omega)} u^f$ as $\epsilon \to 0$, and (by Lemma 2.18),

$$|Du_\epsilon^f|(\Omega) \to |DTu^f|(\bar{\Omega}) = J(u) \text{ as } \epsilon \to 0.$$

We will be done if we can show that for our choice of ϵ, $E_\lambda^f(s_\epsilon^f) \to E_\lambda^f(u_\epsilon^f)$ as $n \to \infty$. Indeed, we have

$$\begin{aligned}
|E_\lambda^f(s_\epsilon^f) - E_\lambda^f(u_\epsilon^f)| &= \left|\lambda \left[\int_\Omega |\nabla s_\epsilon^f| dx - \int_\Omega |\nabla u_\epsilon^f| dx\right] + \frac{1}{2}\left[\|s_\epsilon^f - f\|_{L^2}^2 - \|u_\epsilon^f - f\|_{L^2}^2\right]\right| \\
&\leq \lambda \int_\Omega |\nabla(s_\epsilon^f - u_\epsilon^f)| dx + \frac{1}{2}\left[\|s_\epsilon^f - u_\epsilon^f\|_{L^2(\Omega)}^2 + 2\|u_\epsilon^f - f\|_{L^2(\Omega)}\|u_\epsilon^f - s_\epsilon^f\|_{L^2(\Omega)}\right] \\
&\leq \lambda \int_\Omega \left|\nabla(s_\epsilon^f - u_\epsilon^f)\right| dx + \frac{1}{2}\|s_\epsilon^f - u_\epsilon^f\|_{L^2(\Omega)}(\|u_\epsilon^f - s_\epsilon^f\|_{L^2(\Omega)} + 2\|u_\epsilon^f - f\|_{L^2(\Omega)}) \\
&\leq \left[\lambda + \frac{1}{2}\|u_\epsilon^f - s_\epsilon^f\|_{L^2(\Omega)} + \|u_\epsilon^f - f\|_{L^2(\Omega)}\right]\left[\|u_\epsilon^f - s_\epsilon^f\|_{W^{1,1}(\Omega)} + \|u_\epsilon^f - s_\epsilon^f\|_{L^2(\Omega)}\right] \\
&\leq (1 + C\|T\|_*)\left[\lambda + \frac{C\|T\|_*}{2}\|u_\epsilon^f - s_\epsilon^f\|_{W^{1,1}(\Omega)} + \|u_\epsilon^f - f\|_{L^2}\right]\|u_\epsilon^f - s_\epsilon^f\|_{W^{1,1}(\Omega)},
\end{aligned}$$

where we have used the estimate (5.18). Now, using the estimate (5.17) and letting $\epsilon = h_n^{1/4d}$, we infer from the latter inequality that

$$|E_\lambda^f(s_\epsilon^f) - E_\lambda^f(u_\epsilon^f)| \leq (1 + C\|T\|_*)C(d,\theta)\left[\lambda + C(d,\theta,T)h_n^{d-1/4} + C(f,u^f)\right]h_n^{d-1/4},$$

where

$$C(f,u^f) = \|f\|_{L^2(\Omega)} \sup_{0<\epsilon<1} \|u_\epsilon^f\|_{L^2(\Omega)} \text{ and } C(d,\theta,T) := \frac{C\|T\|_* C(d,\theta)}{2}.$$

Thus, $E_\lambda^f(s_n(f)) \to E_\lambda^f(u^f)$ as $h_n \to 0$, and the proof is complete. □

COROLLARY 5.11. *Under the assumptions of Theorem 5.10, the sequence $\{s_n^d(f)\}_n$ satisfies the following two properties:*

$$s_n^d(f) \xrightarrow{L^p(\Omega)} u^f \text{ as } n \to \infty, \text{ for any } p \in [1,2], \tag{5.19}$$

and

$$J(s_n^d(f)) \to J(u) \text{ as } n \to \infty. \tag{5.20}$$

PROOF. We recall that Ω is assumed to be a bounded domain; therefore it suffices to establish (5.19) for $p = 2$. The result for $1 \leq p < 2$ follows from the fact that $L^2(\Omega)$ is canonically embedded into $L^p(\Omega)$. The case $p = 2$ follows easily from Theorem 2.28 and

Theorem 5.10. In fact, by equation (2.33) we have

$$\forall n \in \mathbb{N}, \quad \|s_n(f) - u^f\|^2_{L^2(\Omega)} \le 2\left(E^f_\lambda(s_n(f)) - E^f_\lambda(u^f)\right);$$

thus by Theorem 5.10 above, we have $\|s_n(f) - u^f\|^2_{L^2(\Omega)} \to 0$ as $n \to \infty$.

Finally, we observe that

$$J(s_n(f)) - J(u) = \frac{1}{\lambda}\left[E^f_\lambda(s_n(f)) - E^f_\lambda(u^f) + \frac{1}{2}\|u^f - f\|^2_{L^2} - \frac{1}{2}\|s_n(f) - f\|^2_{L^2}\right].$$

Thus, by Theorem 5.10 and Corollary 5.11, taking the limit of the latter identity as $n \to \infty$ yields (5.20) and the proof is complete. □

5.3 COMPUTATION OF THE SPLINE APPROXIMATION

Suppose that we have set a mesh-size h and chosen a corresponding triangulation Δ_h of Ω. Let $s^d_h(f)$ be the minimizer of $E^f_\lambda(f)$ with respect to $\mathcal{S}^0_d(\Delta_h)$. Although we have reduced the problem to a very tractable function space, the computation of the spline minimizer remains as challenging as the original minimization problem in the BV space. Nonetheless, we are now able to derive the Euler-Lagrange equation. In fact, it is easy to show that $s^d_h(f)$ necessarily satisfies the nonlinear equation

$$\lambda \int_{\Omega\setminus F(u)} \frac{\nabla u \cdot \nabla s}{|\nabla u|} dx + \lambda \int_{F(u)} |\nabla s| dx + \int_\Omega (u - f)s\, dx = 0 \;\; \forall\, s \in \mathcal{S}^0_d(\Delta_h), \tag{5.21}$$

where $F(u) := \{x \in \Omega\colon \nabla u = 0\}$ is made of the flat regions of u.

In particular, if $F(s^d_h(f))$ is negligible, the Euler-Lagrange equation (5.21) reduces to

$$\lambda \int_\Omega \frac{\nabla u \cdot \nabla s}{|\nabla u|} dx + \int_\Omega (u - f)s\, dx = 0, \;\; \forall\, s \in \mathcal{S}^0_d(\Delta_h) \tag{5.22}$$

which is now amenable to variational techniques.

REMARK 5.12. In practice, the mesh size of an admissible triangulation is dictated by the structure of the image f. An image with a significant amount of textures will require a finer triangulation to preserve textures in the recovered image. The choice of the parameter λ

will also be influenced by the textures and the noise information contained in f.

For the purpose of this section and the necessity of a stable numerical scheme, we do not solve the minimization problem (5.14) or its Euler-Lagrange equation (5.21) directly. Instead, we solve a perturbation of this problem. We recall that the difficulty in dealing with (5.21) is due to the fact that the associated Lagrangian

$$L(\boldsymbol{p}, z, x) = \lambda|\boldsymbol{p}| + \frac{1}{2}(z-f)^2, \quad \forall(\boldsymbol{p}, z, x) \in \mathbb{R}^2 \times \mathbb{R} \times \mathbb{R}^2$$

is not differentiable with respect to $\boldsymbol{p}$ at $\boldsymbol{p} = \mathbf{0}$. As mentioned above, one way to mitigate this difficulty is to find a differentiable relaxation of the Lagrangian L such that the corresponding energy is a perturbation of $E^f_\lambda(u)$. This approach has been successfully used in two occasions in the literature [1, 35].

We will use a similar technique to construct an algorithm for computing a numerical approximation of the spline minimizer $s^d_h(f)$. In fact the relaxation that we use has already been used in the literature for the discrete ROF model under the appellation Huber-ROF model [36], and was shown to produce smoother images than the original ROF model with no staircase effect as is customary for the ROF model.

Let $\epsilon > 0$ be fixed and Φ_ϵ the continuously differentiable function defined by

$$\Phi_\epsilon(x) = \begin{cases} \frac{1}{2\epsilon}x^2 + \frac{\epsilon}{2} & \text{if } 0 \leq |x| \leq \epsilon \\ |x| & \text{if } |x| > \epsilon \end{cases} \tag{5.23}$$

and consider the problem

$$\underset{u \in \mathcal{S}^0_d(\Delta_h)}{\arg\min} \left\{ E^f_{\lambda,\epsilon}(u) := \lambda \int_\Omega \Phi_\epsilon(|\nabla u|)dx + \frac{1}{2}\int_\Omega |u-f|^2 dx \right\}. \tag{5.24}$$

We observe that the relaxation functional $E^f_{\lambda,\epsilon}$ is strictly convex and lower semicontinuous on $\mathcal{S}^0_d(\Delta_h)$ with respect to the L^2-norm. Consequently, the minimization problem (5.24) has a unique minimizer that we denote by $s^d_h(f,\epsilon)$. We have the following characterization of the minimizer $s^d_h(f,\epsilon)$.

PROPOSITION 5.13. *A function $u \in \mathcal{S}_d^0(\Delta_h)$ is a minimizer of the functional $E^f_{\lambda,\epsilon}$ in $\mathcal{S}_d^0(\Delta_h)$ if and only if u satisfies the variational equation*

$$\lambda \int_\Omega \frac{1}{\epsilon \vee |\nabla u|} \nabla u \cdot \nabla s \, dx + \int_\Omega (u - f) s \, dx = 0 \quad \forall\, s \in \mathcal{S}_d^0(\Delta_h), \tag{5.25}$$

where

$$a \vee b := \max(a, b) = \frac{1}{2}(a + b + |a - b|). \tag{5.26}$$

PROOF. First, we observe that $E^f_{\lambda,\epsilon}(u)$ is Gâteaux differentiable with directional derivatives at any point $u \in \mathcal{S}_d^0(\Delta_h)$ given by

$$\langle dE^f_{\lambda,\epsilon}(u), s\rangle = \lambda \int_\Omega \frac{1}{\epsilon \vee |\nabla u|} \nabla u \cdot \nabla s \, dx + \int_\Omega (u - f) s \, dx \quad \forall s \in \mathcal{S}_d^0(\Delta_h). \tag{5.27}$$

Therefore, u is a minimizer of $E^f_{\lambda,\epsilon}(u)$ in $\mathcal{S}_d^0(\Delta_h)$ if and only if $dE^f_{\lambda,\epsilon}(u) = 0$, *i.e* (5.25) holds. □

The following result holds and is at the core of the relaxation algorithm that we will propose for computing a numerical approximation of $s_h^d(f)$.

LEMMA 5.14. *The family of functionals $E^f_{\lambda,\epsilon}(u)$ converges uniformly in $\mathcal{S}_d^0(\Delta_h)$ to $E^f_\lambda(u)$ as $\epsilon \searrow 0$. Moreover, we have $s_h^d(f, \epsilon) \xrightarrow[\epsilon \to 0]{L^2(\Omega)} s_h^d(f)$.*

PROOF. Let Φ be the continuous function defined by $\Phi(x) = |x|$. It is easy to show that

$$\sup_{x \in \mathbb{R}} |\Phi_\epsilon(x) - \Phi(x)| \le \epsilon.$$

Therefore, for any $u \in \mathcal{S}_d^0(\Delta_h)$ we have the estimate

$$|E^f_{\lambda,\epsilon}(u) - E^f_\lambda(u)| \le \lambda \int_\Omega |\Phi_\epsilon(|\nabla u|) - \Phi(|\nabla u|)| \, dx \le \lambda |\Omega| \epsilon,$$

and it follows that $E^f_{\lambda,\epsilon}$ converges uniformly in $\mathcal{S}_d^0(\Delta_h)$ to E^f_λ.

Next, we note that Theorem 2.28 remains true on $\mathcal{S}_d^0(\Delta_h)$. Therefore, rewriting equation (2.33) in $\mathcal{S}_d^0(\Delta_h)$ for $s_h^d(f)$, we obtain

$$\begin{aligned}
\|s_h^d(f,\epsilon) - s_h^d(f)\|^2_{L^2(\Omega)} &\le 2(E_\lambda^f(s_h^d(f,\epsilon)) - E_\lambda^f(s_h^d(f))) \\
&\le 2(E_\lambda^f(s_h^d(f,\epsilon)) - E_{\lambda,\epsilon}^f(s_h^d(f,\epsilon))) + 2(E_{\lambda,\epsilon}^f(s_h^d(f,\epsilon)) - E_\lambda^f(s_h^d(f))) \\
&\le 2\lambda|\Omega|\epsilon + \underbrace{E_{\lambda,\epsilon}^f(s_h^d(f,\epsilon)) - E_{\lambda,\epsilon}^f(s_h^d(f))}_{\le 0} + E_{\lambda,\epsilon}^f(s_h^d(f)) - E_\lambda^f(s_h^d(f)) \\
&\le 4\lambda\epsilon|\Omega|.
\end{aligned}$$

Thus, $\|s_f^d(f,\epsilon) - s_h^d(f)\|_{L^2(\Omega)} \le 2\sqrt{\lambda|\Omega|\epsilon}$, and it follows that $s_h^d(f,\epsilon)$ converges to $s_h^d(f)$ in $L^2(\Omega)$ as ϵ goes to 0. □

We now develop an algorithm for computing approximations of the minimizer $s_h^d(f,\epsilon)$ and study its convergence.

ALGORITHM 5.15. *Start from any bounded nonnegative function $v^0 \in \mathcal{S}_d^0(\Delta_h)$ and let*

$$u^{n+1} = \arg\min_{u \in S_d^0(\Delta_h)} \lambda \int_\Omega v^n |\nabla u|^2\, dx + \frac{1}{2}\int_\Omega |u - f|^2\, dx \quad \forall n \ge 0, \tag{5.28a}$$

$$v^{n+1} := \arg\min_{0 < v \le 1/\epsilon} \int_\Omega v|\nabla u^{n+1}|^2 + \frac{1}{v} dx = \frac{1}{\epsilon \vee |\nabla u^{n+1}|}. \tag{5.28b}$$

A standard argument using Lax-Milgram Theorem (see [27, Corollary 5.8 p. 140]) shows that u^{n+1} is characterized by the variational equation

$$2\lambda \int_\Omega v^n \nabla u^{n+1} \cdot \nabla s\, dx + \int_\Omega (u^{n+1} - f)s\, dx = 0, \quad \forall\, s \in \mathcal{S}_d^0(\Delta_h). \tag{5.29}$$

The existence and uniqueness of u^{n+1} follows by observing that the bilinear form

$$a_n(u,v) := \int_\Omega 2\lambda v^n \nabla u \cdot \nabla v + uv\, dx$$

is continuous – thanks to Theorem 5.7– and coercive on $\mathcal{S}_d^0(\Delta_h) \times \mathcal{S}_d^0(\Delta_h)$ with respect to the L^2-norm. Consequently, the equation (5.29) has a unique solution.

We fix $\epsilon > 0$ and for the sake of notation conciseness, consider the functional E defined by

$$E(u,v) = \int_\Omega \lambda(v|\nabla u|^2 + \frac{1}{v})\, dx + \frac{1}{2}\int_\Omega |u - f|^2\, dx. \tag{5.30}$$

It is easy to check that

$$u^{n+1} = \underset{u \in \mathcal{S}_d^0(\Delta_h)}{\arg\min} E(u, v^n) \text{ and } v^{n+1} = \underset{0<v\leq 1/\epsilon}{\arg\min} E(u^{n+1}, v). \tag{5.31}$$

LEMMA 5.16. *The sequence $\{u^n\}_n$ is bounded in $H^1(\Omega)$ and satisfies*

$$\forall\, n \in \mathbb{N},\ \forall s \in \mathcal{S}_d^0(\Delta_h),\ \ \|s - u^n\|^2_{L^2(\Omega)} \leq 2(E(s, v^{n-1}) - E(u^n, v^{n-1})). \tag{5.32}$$

In particular, we have

$$\forall n \in \mathbb{N}, \quad \|u^{n+1} - u^n\|^2_{L^2(\Omega)} \leq 2(E(u^n, v^n) - E(u^{n+1}, v^{n+1})). \tag{5.33}$$

PROOF. First, we observe that in view of Theorem 5.7, proving the boundedness of $\{u^n\}$ in $H^1(\Omega)$ is equivalent to proving its boundedness in $L^2(\Omega)$. Let $n \in \mathbb{N}$ be given. Then by definition of u^n, we have

$$E(u^n, v^{n-1}) \leq E(0, v^{n-1}) = \frac{1}{2}\|f\|^2_{L^2} + \int_\Omega \frac{1}{v^{n-1}}\, dx \leq \frac{1}{2}\|f\|^2_{L^2} + \frac{|\Omega|}{\epsilon}.$$

Consequently, we get $\|u^n - f\|^2_{L^2(\Omega)} \leq \|f\|^2_{L^2(\Omega)} + \dfrac{2|\Omega|}{\epsilon}$, and deduce by the triangle inequality that

$$\|u^n\|_{L^2(\Omega)} \leq 2\|f\|_{L^2(\Omega)} + \sqrt{\frac{2|\Omega|}{\epsilon}}.$$

We now show that $\|u^n - s\|^2_{L^2(\Omega)} \leq 2(E(s, v^{n-1}) - E(u^n, v^{n-1}))$. In fact, for any $s \in \mathcal{S}_d^0(\Delta_h)$, we have

$$\begin{aligned}
E(s, v^{n-1}) - E(u^n, v^{n-1}) &= \int_\Omega \lambda v^{n-1}(|\nabla s|^2 - |\nabla u^n|^2) + \frac{1}{2}(|s-f|^2 - |u^n - f|^2)\, dx \\
&= \int_\Omega \lambda v^{n-1}|\nabla(s - u^n)|^2 + \frac{1}{2}|s - u^n|^2\, dx + \\
&\qquad \underbrace{\int_\Omega 2\lambda v^{n-1}\nabla u^n \cdot \nabla(s - u^n) + (u^n - f)(s - u^n)\, dx}_{=0 \text{ by } (5.29)} \\
&= \int_\Omega \lambda v^{n-1}|\nabla(s - u^n)|^2 + \frac{1}{2}|s - u^n|^2\, dx \\
&\geq \frac{1}{2}\|s - u^n\|^2_{L^2(\Omega)} \quad \text{since } v^{n-1} \geq 0.
\end{aligned}$$

In particular, for any $n \in \mathbb{N}$,

$$
\begin{aligned}
\|u^{n+1} - u^n\|^2_{L^2(\Omega)} &\leq 2(E(u^n, v^n) - E(u^{n+1}, v^n)) \\
&= 2(E(u^n, v^n) - E(u^{n+1}, v^{n+1})) + 2\underbrace{(E(u^{n+1}, v^{n+1}) - E(u^{n+1}, v^n))}_{\leq 0 \text{ by (5.31)}} \\
&\leq 2(E(u^n, v^n) - E(u^{n+1}, v^{n+1})).
\end{aligned}
$$

Thus, the sequence $\{E(u^n, v^n)\}_n$ is monotone nonincreasing and $\|u^n - u^{n+1}\|_{L^2(\Omega)} \to 0$.
□

We are now ready to prove the convergence of the sequence $\{u^n\}_n$ to the minimizer $s^d_h(f, \epsilon)$.

THEOREM 5.17. *The sequence $\{u_n\}_n$ constructed in Algorithm 5.15 converges in $L^2(\Omega)$ to the minimizer $s^d_h(f, \epsilon)$ of $E^f_{\lambda,\epsilon}(u)$.*

PROOF. In view of Proposition 5.13, it suffices to show that any cluster point u of the sequence $\{u^n\}_n$ with respect to the L^2-norm satisfies the Euler-Lagrange equation (5.25). To begin, we note that the sequence $\{u^n\}_n$ has at least one cluster point as a bounded sequence in a finite dimensional normed vector space.

Let u be any cluster point of $\{u^n\}_n$ in $L^2(\Omega)$ and $\{u^{n_k}\}_k$ a subsequence such that $u^{n_k} \xrightarrow{L^2(\Omega)} u$. Since $\|u^{n_k+1} - u^{n_k}\|_{L^2(\Omega)} \to 0$, it follows that $u^{n_k+1} \xrightarrow{L^2(\Omega)} u$ as well. By Markov inequality – Theorem 5.7 – we also have

$$u^{n_k} \xrightarrow[k\to\infty]{H^1(\Omega)} u \text{ and } u^{n_k+1} \xrightarrow[k\to\infty]{H^1(\Omega)} u.$$

Therefore, by Lebesgue dominated convergence theorem, we get

$$v^{n_k} = \frac{1}{|\nabla u^{n_k}|} \wedge \frac{1}{\epsilon} \xrightarrow[k\to\infty]{L^2(\Omega)} \frac{1}{|\nabla u|} \wedge \frac{1}{\epsilon} = \frac{1}{\epsilon \vee |\nabla u|}.$$

Next, we establish that u satisfies the variational equation

$$2\lambda \int_\Omega \frac{1}{\epsilon \vee |\nabla u|} \nabla u \cdot \nabla s \, dx + \int_\Omega (u - f) s \, dx = 0, \quad \forall s \in \mathcal{S}^0_d(\Delta_h). \tag{5.34}$$

Indeed by definition of u^{n_k+1}, for any $s \in \mathcal{S}_d^0(\Delta_h)$, there holds

$$2\lambda \int_\Omega v^{n_k} \nabla s \cdot \nabla u^{n_k+1}\, dx + \int_\Omega (u^{n_k+1} - f) s\, dx = 0, \quad \forall k \in \mathbb{N}. \tag{5.35}$$

Since ∇u^{n_k+1} converges strongly to ∇u in $L^2(\Omega) \times L^2(\Omega)$ and $v^{n_k} \nabla s$ converges strongly to $\dfrac{\nabla s}{\epsilon \vee |\nabla u|}$, it follows that

$$\int_\Omega v^{n_k} \nabla s \cdot \nabla u^{n_k+1}\, dx \longrightarrow \int_\Omega \frac{1}{\epsilon \vee |\nabla u|} \nabla u \cdot \nabla s\, dx \text{ as } k \to \infty. \tag{5.36}$$

Similarly, as u^{n_k+1} converges strongly to u in $L^2(\Omega)$, we infer that

$$\int_\Omega (u^{n_k+1} - f) s\, dx \longrightarrow \int_\Omega (u - f) s\, dx \text{ as } k \to \infty. \tag{5.37}$$

On passing to the limit as $k \to \infty$ in (5.35) and taking into account (5.36) and (5.37), we obtain (5.34) and the proof is complete. □

REMARK 5.18. A similar relaxation functional was used in [35] to derive a minimizing sequence of the ROF functional in the Hilbert space $H^1(\Omega)$. More specifically, these authors used the following C^1 function to obtain a relaxation of the ROF model

$$\Psi_\epsilon(x) = \begin{cases} \dfrac{x^2}{2\epsilon} & \text{if } |x| \leq \epsilon \\ |x| - \dfrac{\epsilon}{2} & \text{if } \epsilon \leq |x| \leq \dfrac{1}{\epsilon} \\ \dfrac{\epsilon}{2} x^2 + \dfrac{1}{2}\left(\dfrac{1}{\epsilon} - \epsilon\right) & \text{if } |x| \geq \dfrac{1}{\epsilon}. \end{cases}$$

The algorithm that we studied above is a minor modification of the one they proposed for constructing a minimizing sequence of the ROF model in the Hilbert space $H^1(\Omega)$. The difference is the way we update v^n. Their update is given by

$$v^{n+1} = \underset{\epsilon \leq v \leq 1/\epsilon}{\arg\min}\, E(u^{n+1}, v).$$

This choice is made so as to guarantee that $E(u, v^n)$ used in defining u^{n+1} is the energy functional associated to a continuous and coercive bilinear functional on $H^1(\Omega)$.

In our case, since we are working on a finite dimensional space, relaxing the lower bound on v does not pause any problem and we get a continuous (thanks to Markov Inequality) and coercive bilinear functional on the spline space $\mathcal{S}_d^0(\Delta_h)$ with respect to the L^2-norm.

5.4 IMPLEMENTATION OF THE ALGORITHM

In this section, we explain how to compute the sequence u_n in Algorithm 5.15. We exploit the B-form representation of bivariate splines and solve the following constrained variant of (5.29):

$$\begin{aligned}&\text{Find } u^{n+1} \in \mathcal{S}_d^0(\Delta_h) \text{ such that for any } v \in \mathcal{S}_d^{-1}(\Delta_h)\\ &2\lambda \int_\Omega v^n \nabla u^{n+1} \cdot \nabla u \, dx + \int_\Omega u^{n+1} v \, dx = \int_\Omega f v \, dx.\end{aligned} \tag{5.38}$$

Clearly, any solution of (5.38) is a solution of (5.29).

Let $\mathcal{Q}_d$ be the orthogonal projection operator from $L^2(\Omega)$ on the spline space $\mathcal{S}_d^0(\Delta_h)$. Then, it is easy to see that the constrained variational system (5.38) is equivalent to

$$\begin{aligned}&\text{Find } u^{n+1} \in \mathcal{S}_d^0(\Delta_h) \text{ such that for any } T \in \Delta_h \text{ and any } i+j+k=d,\\ &2\lambda \int_T \mathcal{Q}_{2d-2}(v^n) \nabla u^{n+1} \cdot \nabla B_{ijk}^{T,d} \, dx + \int_T u^{n+1} B_{ijk}^{T,d} \, dx = \int_T \mathcal{Q}_d(f) B_{ijk}^{T,d} \, dx.\end{aligned} \tag{5.39}$$

The latter variational problem is more amenable to computation on a computer as all the data are now in spline spaces, and the integrals are easily computed using Theorem 5.4.

REMARK 5.19. Although, the computation of the operator $\mathcal{Q}_d$ on a typical L^2 function is hard, in practice when we deal with digital images, we use a (penalized) least-square fit to evaluate $\mathcal{Q}_d(f)$ based on the pixel values of the image.

We now derive the linear system associated with the variational problem (5.39) and discuss the existence of a solution as well as an iterative algorithm for computing it. Let E be the number of interior edges of Δ_h and N the number of triangles in Δ_h. First, we recall that there exists a matrix A_0 of dimension $E(d+1) \times \binom{d+2}{2} N$ such that

$$A_0 \boldsymbol{c}^{n+1} = \mathbf{0}, \tag{5.40}$$

where $\boldsymbol{c}^{n+1}$ is a length $\binom{d+2}{2} N$ vector representing the B-net of u^{n+1}.

Next, given a listing $\{T_1, T_2, \ldots, T_N\}$ of the triangles in Δ_h, we write $\boldsymbol{c}^{n+1}$ in block form as follows

$$\boldsymbol{c}^{n+1} = \left(\boldsymbol{c}^{n+1,1}, \boldsymbol{c}^{n+1,2}, \ldots, \boldsymbol{c}^{n+1,N}\right),$$

where for each $i = 1, 2, \ldots, N$, $\boldsymbol{c}^{n+1,i}$ are the coefficients of $u^{n+1}|_{T_i}$ in the B-basis of $\mathbb{P}_d$, *i.e*

$$u^{n+1}|_{T_i} = \sum_{j+k+\ell=d} \boldsymbol{c}^{n+1,i}_{jk\ell} B^{T_i,d}_{jk\ell}.$$

Furthermore, for the triangle T_i and using the lexicographical ordering of the index set $\{j+k+\ell = d, 0 \leq j \leq k \leq \ell \leq d\}$, we define the local stiffness, $S^{n,i}$ by

$$S^{n,i}_{p,q}(\lambda) = \int_{T_i} \mathcal{Q}_{2d-2}(2\lambda v^n) \nabla B^{T_i,d}_p \cdot \nabla B^{T_i,d}_q \, dx, \quad 1 \leq p, q \leq \binom{d+2}{2}, \tag{5.41}$$

the local mass matrix M^i by

$$M^i_{pq} = \int_{T_i} B^{T,d}_q B^{T,d}_q \, dx, \quad 1 \leq p, q \leq \binom{d+2}{2}, \tag{5.42}$$

and the local load vector $\boldsymbol{F}^i$ by

$$\boldsymbol{F}^i_p = \int_{T_i} \mathcal{Q}_d(f) B^{T,d}_q \, dx, \quad 1 \leq p \leq \binom{d+2}{2}. \tag{5.43}$$

LEMMA 5.20. *For each $i = 1, 2, \ldots, N$, the local stiffness matrix $S^{n,i}(\lambda)$ is symmetric nonnegative definite, while the local mass matrix M^i is symmetric positive definite. Consequently, the matrix $S^{n,i}(\lambda) + M^i$ is symmetric positive definite.*

PROOF. Indeed, for any $c \in \mathbb{R}^D$ with $D = \binom{d+2}{2}$, we have

$$c^T S^{n,i}(\lambda) c = \int_{T_i} 2\lambda v_n |\nabla p|^2 dx \text{ and } c^T M^i c = \int_{T_i} |p|^2 dx,$$

where p is the polynomial with coefficients $\boldsymbol{c}$ in the Bernstein-Bezier basis of $\mathbb{P}_d$ with respect to T_i. Thus, $S^{n,i}(\lambda)$ is clearly nonnegative definite, and M^i is positive definite. Both matrices are symmetric by construction, and the matrix $S^{n,i}(\lambda)$ inherits its properties from $S^{n,i}(\lambda)$ and M^i. □

Consequently, the variational equation (5.39) is equivalent to the linear system

$$\begin{bmatrix} S^{n,1}(\lambda)+M^1 & 0 & \cdot\ \cdot & 0 \\ 0 & S^{n,2}(\lambda)+M^2 & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & S^{n,N}(\lambda)+M^N \\ A_0^1 & A_0^2 & \cdot\ \cdot & A_0^N \end{bmatrix} \begin{pmatrix} \boldsymbol{c}^{n+1,1} \\ \boldsymbol{c}^{n+1,2} \\ \vdots \\ \boldsymbol{c}^{n+1,N} \end{pmatrix} = \begin{pmatrix} \boldsymbol{F}^1 \\ \boldsymbol{F}^2 \\ \vdots \\ \boldsymbol{F}^N \\ \boldsymbol{0} \end{pmatrix}, \quad (5.44)$$

where we have written the continuity matrix A_0 in block form according to the listing $\{T_1, T_2, \ldots, T_N\}$ of the triangles in Δ_h. Since, we know that the variational equation (5.38) has a unique solution, the system of equations (5.44) must be consistent. Also by Lemma 5.20, the coefficient matrix of the above overdetermined system has full rank. Therefore, there is a unique solution which can be recovered using the best linear system solver that may be found or developed. However, this approach poses some practical issues, such as the supplementary storage required to assemble the system's coefficient matrix.

Alternatively, we may use Lax-Milgram Theorem [27, Corollary 5.8, p. 140] to see that the variational equation (5.38) is equivalent to the constrained minimization problem

$$\begin{aligned} &\text{Minimize} \quad \frac{1}{2}\boldsymbol{c}^T(S^n(\lambda)+M)\boldsymbol{c} - \boldsymbol{F}^T\boldsymbol{c} \\ &\text{subject to:} \quad A_0\boldsymbol{c} = \boldsymbol{0}, \end{aligned} \quad (5.45)$$

where $S^n(\lambda)$ is the block diagonal matrix

$$S^n(\lambda) = \begin{bmatrix} S^{n,1}(\lambda) & 0 & \cdot\ \cdot & 0 \\ 0 & S^{n,2}(\lambda) & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & S^{n,N}(\lambda) \end{bmatrix},$$

with M and $\boldsymbol{F}$ given by

$$M = \begin{pmatrix} M^1 & 0 & \cdot\ \cdot & 0 \\ 0 & M^2 & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & M^N \end{pmatrix}, \quad \boldsymbol{F} = \begin{pmatrix} \boldsymbol{F}^1 \\ \boldsymbol{F}^2 \\ \vdots \\ \boldsymbol{F}^N \end{pmatrix},$$

and A_0 has full rank. The full rank assumption on A_0 is equivalent to saying that A_0 enforces a minimal set of conditions for continuity across the edges of the triangulation.

Let $\boldsymbol{c}^*$ be a solution of the constrained minimization (5.45). Then, there exists a vector of Lagrange multipliers $\boldsymbol{\mu}^*$ such that the pair $(\boldsymbol{c}^*, \boldsymbol{\mu}^*)$ solves the saddle point system

$$\begin{bmatrix} S^n(\lambda) + M & A_0^T \\ A_0 & \mathbf{0} \end{bmatrix} \begin{pmatrix} \boldsymbol{c} \\ \boldsymbol{\mu} \end{pmatrix} = \begin{pmatrix} \boldsymbol{F} \\ \mathbf{0} \end{pmatrix}. \tag{5.46}$$

Moreover, since $S^n(\lambda) + M$ is positive definite and A_0 has full row rank, the pair $(\boldsymbol{c}^*, \boldsymbol{\mu}^*)$ is the unique solution of the saddle point system (5.46). We now summarize the spline algorithm

ALGORITHM 5.21. *Given a function $f \in L^2(\Omega)$, choose a triangulation Δ_h of Ω, the degree d of the spline, and the flat region parameter $\epsilon > 0$. Let $\mathbf{c}^0$ be the B-net of a nonnegative constant function defined on Ω.*

1. *Assemble the smoothness matrix A_0, the total variation matrix $S^0(\lambda)$ and the mass matrix M, compute the orthogonal projection $\mathcal{Q}_d(f)$, and the load vector $\boldsymbol{F}$.*

2. *For any $n \geq 0$, compute the B-net $\mathbf{c}^{n+1}$ by solving the saddle point system* (5.46).

3. *Compute the orthogonal projection $\mathcal{Q}_{2d-2}(1/\epsilon \vee |\nabla u^{n+1}|)$, where u^{n+1} is the spline with B-net $\mathbf{c}^{n+1}$, and update the total variation matrix $S^{n+1}(\lambda)$.*

4. *Until a stopping criterion is met, increment $n \leftarrow n+1$ and go to step* **2**.

The storage cost for using a direct solver to solve the above saddle point system (5.46) quickly become prohibitive as the degree of the spline function is increased, and an iterative method with low storage cost is desirable. One such method is the method of multipliers [24] which can be described as follows.

Let $\gamma > 0$ be fixed and consider the augmented Lagrangian functional

$$\mathcal{L}(\boldsymbol{c}, \boldsymbol{\mu}) := \frac{1}{2}\,\boldsymbol{c}^T(S^n(\lambda) + M)\boldsymbol{c} - \boldsymbol{F}^T\boldsymbol{c} + \boldsymbol{\mu}^T A_0\boldsymbol{c} + \frac{\gamma}{2}\|A_0\boldsymbol{c}\|^2 \tag{5.47}$$

Given an approximation $\boldsymbol{\mu}_k$ of the vector of multipliers $\boldsymbol{\mu}^*$, we compute an approximation $\boldsymbol{c}_{k+1}$ of the solution $\boldsymbol{c}^*$ as the minimizer of the functional $\mathcal{L}_k(\boldsymbol{c}) = \mathcal{L}(\boldsymbol{c}, \boldsymbol{\mu}_k)$ and update the multipliers' estimate according to the rule $\boldsymbol{\mu}_{k+1} = \boldsymbol{\mu}_k + \rho\, A_0\boldsymbol{c}_{k+1}$. Thus, we obtain the following iterative algorithm.

ALGORITHM 5.22 (Method of Multipliers). *Choose $\gamma > 0$ and $\rho > 0$. Pick an initial vector of multipliers $\boldsymbol{\mu}_0$. For each $k = 0, 1, 2, \ldots$. Solve the linear system*

$$(S^n(\lambda) + M + \gamma A_0^T A_0)\boldsymbol{c}_{k+1} = \boldsymbol{F} - A_0^T\boldsymbol{\mu}_k, \tag{5.48}$$

and update the multiplier by the rule

$$\boldsymbol{\mu}_{k+1} = \boldsymbol{\mu}_k + \rho A_0\boldsymbol{c}_{k+1}. \tag{5.49}$$

The method of multipliers algorithm above is a special case of the augmented Lagrangian algorithm studied in [18]. Consequently, since $S^n(\lambda) + M$ is symmetric positive definite, the sequence $\boldsymbol{c}_k$ converges to the solution $\boldsymbol{c}^*$ provided that $2\gamma - \rho \geq 0$ [18, Theorem 4, p.128]. Moreover, if A_0 is of full rank, then under the same condition, the sequence of Lagrange multipliers $\boldsymbol{\mu}_k$ also converges to $\boldsymbol{\mu}^*$ [40, Theorem 9.4-1, p. 362]. In practice when using the method of multipliers algorithm, care should be taken not to choose ρ so large as to make the matrix $S^n(\lambda) + M + \gamma A_0^T A_0$ ill conditioned.

5.5 NUMERICAL EXPERIMENTS

In this section, we report the results of some numerical experiments done using the algorithm described above for denoising digital images. It is well known (some of these observations have been confirmed by theory) that the ROF model : **(1)** is excellent on piecewise constant images up to a reduction in contrast; **(2)** finite difference algorithms are vulnerable to the staircase effect, whereby smooth regions are recovered decomposed into piecewise constant subregions; **(3)** is ineffective at discriminating textures and noise. Two examples illustrating the issues raised above are provided, using both the finite difference and spline methods. However, we will see that the staircase effect is tamed by the spline algorithm method.

The semidiscrete spline model and algorithm. The algorithm described in the previous section assumes that the data f is a function on a continuum domain, however, digital images are merely samples of such function, the size of which may not be sufficient to estimate the orthogonal projection $\mathcal{Q}_d(f)$. Therefore, we solve the following variant of the model (5.45).

For a digital image $\boldsymbol{f}$ of size $m \times n$, we identify $\boldsymbol{f}$ to the piecewise constant function, f, defined on $\Omega = (1/2,\ m + 1/2) \times (1/2,\ n + 1/2)$ by

$$f(x) = \boldsymbol{f}_{i,j}, \quad x \in (j - 1/2,\ j + 1/2) \times (i - 1/2,\ i + 1/2),$$

for any $1 \leq i \leq n$ and $1 \leq j \leq m$. Let a triangulation $\{T_1, T_2, \ldots, T_N\}$ of Ω be fixed, we compute the spline approximation using Algorithm 5.21 with a minor twist to step 2. Instead of solving the saddle point system (5.46), we solve the following system

$$\begin{bmatrix} S^n(\lambda) + O^T O & A_0^T \\ A_0 & \mathbf{0} \end{bmatrix} \begin{pmatrix} \boldsymbol{c} \\ \boldsymbol{\mu} \end{pmatrix} = \begin{pmatrix} O^T \boldsymbol{f} \\ \mathbf{0} \end{pmatrix}, \tag{5.50}$$

where $\boldsymbol{f}$ is the column vector representing the image $\boldsymbol{f}$, and O is a $N \times N$ block diagonal observation matrix. The $i-$th block of O, O_i with dimensions $n_i \times \binom{d+2}{2}$, is such that each row is obtained by evaluating the B-basis functions, $B_{ijk}^{T,d}$, at a pixel location that falls within triangle T. Moreover, we must have $\sum_{i=1}^{N} n_i = nm$, so that O has dimensions $nm \times N\binom{d+2}{2}$.

How do we obtain the triangulation? We compute a mesh of the domain Ω using the mesh generating MATLAB function *distmesh* developed by Persson and Strang [66]. The distmesh function aims at generating a triangulation with maximum smallest angle; thus, produces meshes in which most triangles are close to being equilateral.

Example 1: piecewise constant image. In this test, we use the ROF model to clean up realization of a Gaussian noise added to the a binary image made of five geometric shapes. For comparison purposes, we ran the spline algorithm 5.21 and the finite difference alternating projected gradient algorithm 4.15. We used $\tau = 1/8$ for the finite difference algorithm. For the spline algorithm, we computed a continuous cubic spline with $\epsilon = 1/4$. In both cases, we use $\lambda = 1/8$. The spline algorithm is less capable to accurately resolve the edges than the finite difference alternating projected gradient algorithm, as seen by comparing panel (D) and panel (F) in Figure 5.3. The performance of the spline algorithm may be improved by choosing a triangulation that is adapted to the edges in the image. However, generating such triangulations augment the computational cost of the algorithm as we would have to identify the edges in a preprocessing step.

Example 2: Piecewise smooth image. We now show the performance of the spline algorithm on a natural image with minor textures. The parameters for this test are $\lambda = 1/8$, $\epsilon = 1/20$, and $\tau = 1/8$. Both the projected gradient algorithm and the spline method effectively reduce the noise. The finite difference method produces shaper edges than the spline method, see panel (D) and panel (F) in Figure 5.5. However, the finite difference method results in an image with more blocky regions than the one recovered by the spline method, see panel (B) and panel (C) in Figure 5.4.

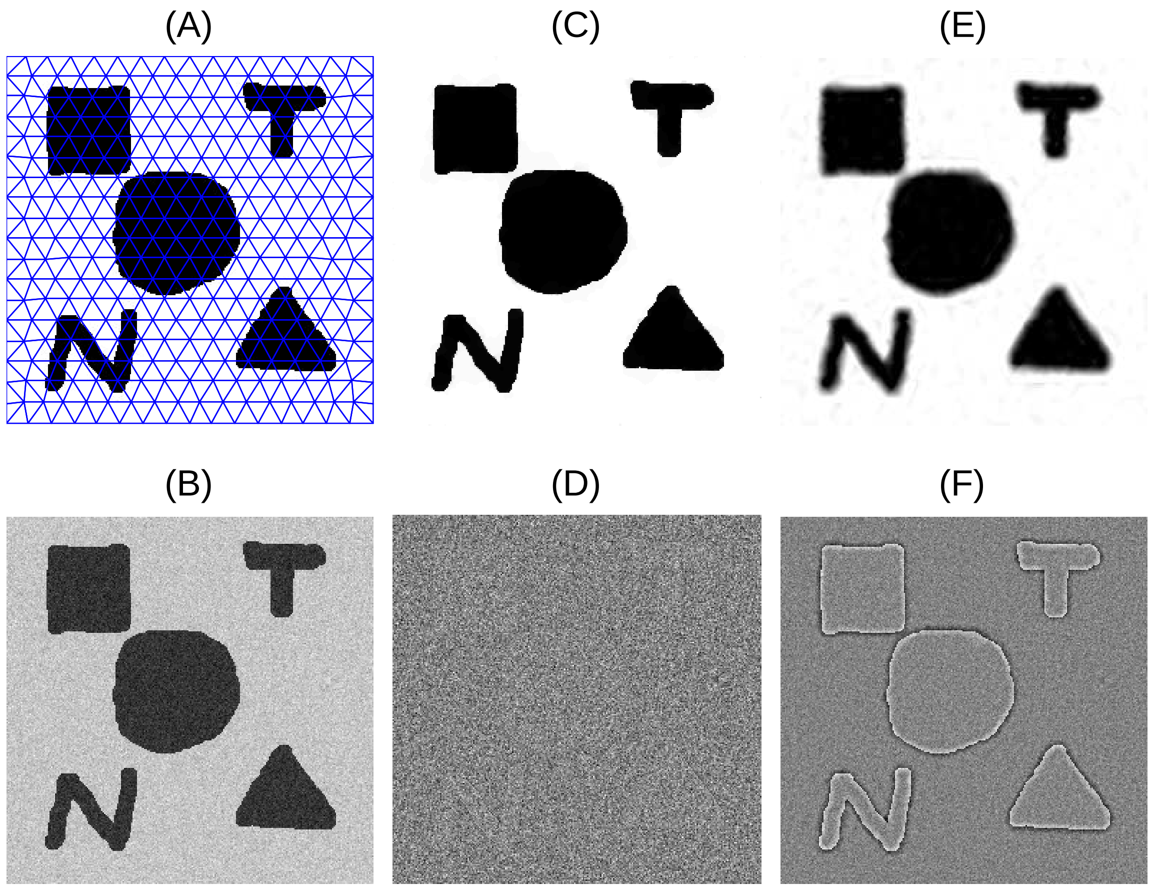

FIGURE 5.3: (A) The original cartoon image overlayed with a triangulation made of 562 triangles and 316 vertices. (B) The noised image obtained by adding a white noise with $\sigma = 25$ to the cartoon image. (C) The image recovered with the projected gradient algorithm. (D) The difference between the noised and recovered images. (E) The image recovered by fitting a continuous cubic spline over the triangulation in image (A). (F) The difference between the spline and noised images.

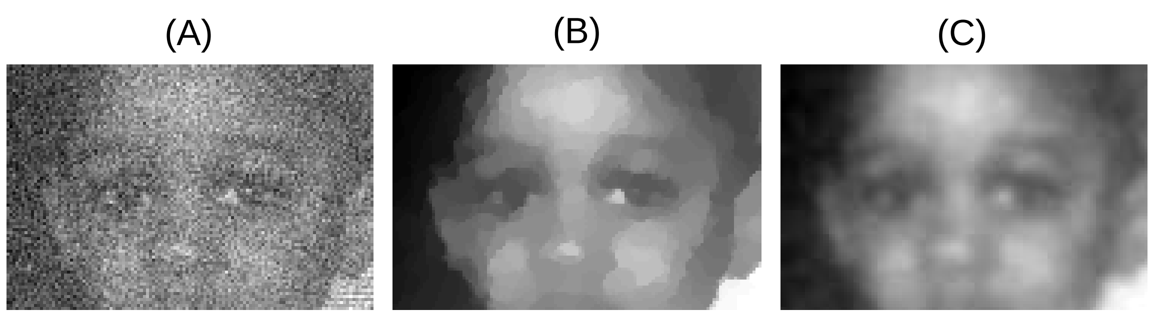

FIGURE 5.4: The spline method is less vulnerable to the staircase effect. (A) Portion of the noised image in Figure 5.3(B). (B) The same portion from the image recovered using the projected gradient algorithm. (C) The same portion form the image recovered with the spline method.

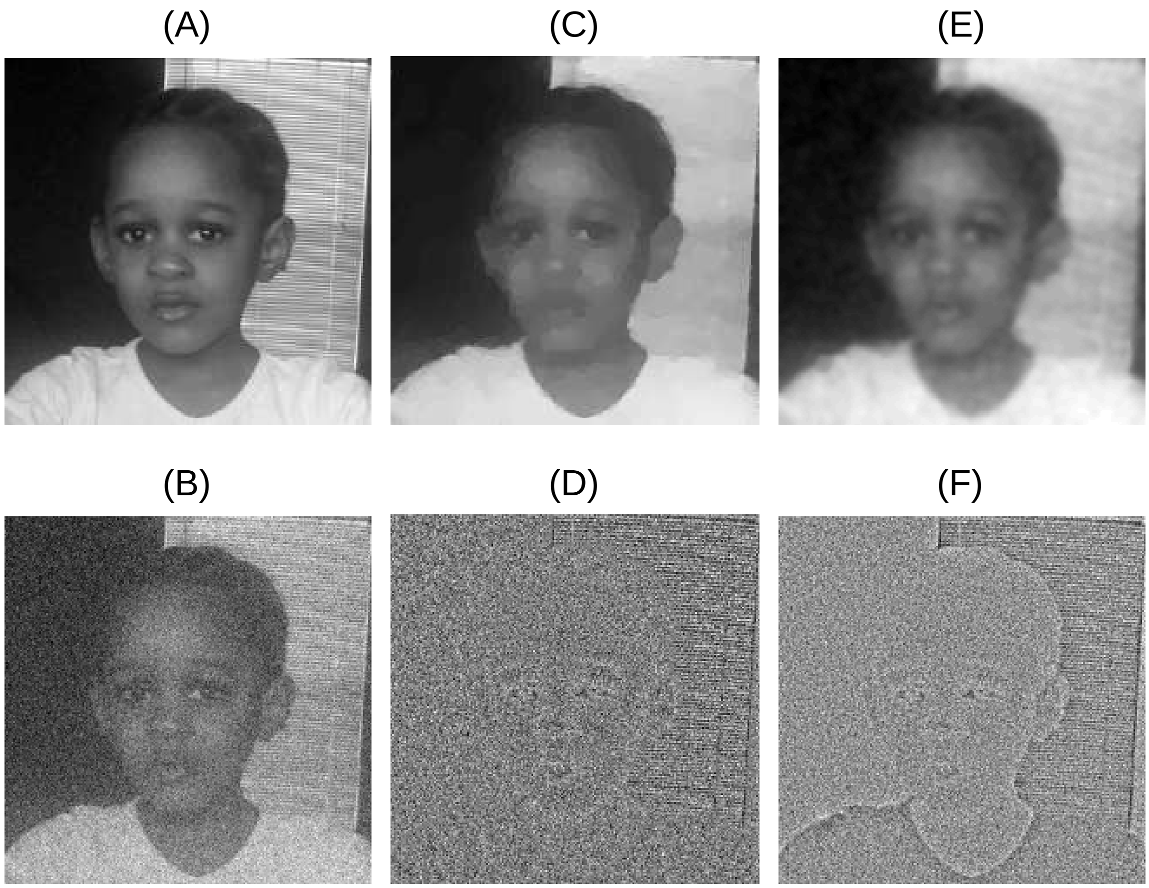

FIGURE 5.5: (A) The clean portrait of a toddler. (B) The noised image obtained by adding a white noise with $\sigma = 25$ to the image in (A). (C) The image recovered with the projected gradient algorithm. (D) The difference between the noised and recovered images. (E) The image recovered by fitting a continuous cubic spline over a mesh made of 1271 triangles and 688 vertices. (F) The residual image of the spline fitting algorithm.

A

EXTENSION OF BOUNDED VARIATION FUNCTIONS

In this chapter, we cover the construction of an extension operator over the space $BV(\Omega)$. Our exposition follows the monographs [46] and [8]. In the sequel, Ω will denote a bounded domain of $\mathbb{R}^2$, unless otherwise noted. We define the sets

$$S = \{x = (x_1, x_2)\colon \, -1 < x_1, x_2 < 1\},$$

$$S_+ = \{x = (x_1, x_2) \in S\colon 0 < x_2 < 1\},$$

$$S_0 = \{x = (x_1, 0)\colon \, -1 < x_1 < 1\}$$

$$B_\rho(x) = \{y \in \mathbb{R}^2\colon |y - x| < \rho\}.$$

An essential tool in the definition of the trace of a function of bounded variation is the Lebesgue theorem.

THEOREM A.1 (Lebesgue). *If $u \in L^1(\mathbb{R}^2)$, then for almost all $x \in \mathbb{R}^2$*

$$\lim_{\rho \to 0} \frac{1}{\rho^2} \int_{B_\rho(x)} |f(z) - f(x)| dt = 0. \tag{A.1}$$

The following special properties of finite Radon measure is central in the construction of the trace operator; a proof is found in [46].

LEMMA A.2 ([46]). *Suppose that Ω is an open domain on $\mathbb{R}^2$ that lies on one side of its boundary. Suppose that μ is a finite Radon measure on $\Omega \subset \mathbb{R}^2$. Then, for $\mathcal{H}_1$-almost every $x \in \partial\Omega$*

$$\lim_{\rho \to 0} \frac{\mu(B_\rho(x) \cap \Omega)}{\rho} = 0. \tag{A.2}$$

A.1 EXTENSION BY REFLECTION

We first study the case of a rectangular domain showing that we can define an extension operator through reflection across the sides of the domain. We begin with an important lemma defining the trace on one side of the rectangle and extending the extension of the Gauss-Green Theorem to functions of bounded variation. We note that it is not always possible to define the trace of a function of bounded variation on the boundary of Ω. In fact, like with Sobolev spaces, the trace operator is closely linked to the geometry of the domain Ω.

LEMMA A.3. *Suppose that $u \in BV(S_+)$. Then there exists a function $u^+ \in L^1(-1,1)$ such that for $\mathcal{H}_1$–almost every $x \in (-1,1)$*

$$\lim_{\rho \to 0} \frac{1}{\rho^2} \int_{S_\rho^+(x)} |u(z) - u^+(x)| dz = 0, \tag{A.3}$$

where $S_\rho^+(x) = (x-\rho, x+\rho) \times (0,\rho)$. Furthermore, for every $\boldsymbol{g} \in C_c^1(S, \mathbb{R}^2)$

$$\int_{S_+} u \operatorname{div}(\boldsymbol{g}) dx = -\int_{S_+} \langle Du, \boldsymbol{g} \rangle + \int_{-1}^{1} u^+ \boldsymbol{g} \cdot \boldsymbol{\nu} d\mathcal{H}^1, \tag{A.4}$$

where $\boldsymbol{\nu} = (0,-1)$, and $\mathcal{H}^1$ is the 1-dimensional Hausdorff measure on $\mathbb{R}$.

PROOF. The proof is done in two steps. First suppose that $u \in C^\infty(\Omega) \cap BV(\Omega)$, and for $\epsilon > 0$ define $u_\epsilon : (-1,1) \to \mathbb{R}$ by

$$u_\epsilon(y) = u(y, \epsilon). \tag{A.5}$$

We have for any $\rho \in (0,1]$ and for every $0 < \epsilon' < \epsilon < \rho$

$$\int_{-\rho}^{\rho} |u_\epsilon - u_{\epsilon'}| d\mathcal{H}^1 \le \int_{-\rho}^{\rho} \int_{-\epsilon'}^{\epsilon} \left| \frac{\partial u}{\partial x_2} \right| dx \le \int_{-\rho}^{\rho} \int_{-\epsilon'}^{\epsilon} |\nabla u| \, dx. \tag{A.6}$$

Letting $\epsilon \to 0$ in the above inequality with $\rho = 1$ shows that u_ϵ is Cauchy in $L^1(-1,1)$; hence converges to $u^+ \in L^1(-1,1)$. On the other hand by Gauss-Green theorem, we have for any $\boldsymbol{g} \in C_c^1(S, \mathbb{R}^2)$

$$\int_{-1}^{1} \int_{\epsilon}^{1} u \operatorname{div}(g) dx = -\int_{-1}^{1} \int_{\epsilon}^{1} \nabla u \cdot \boldsymbol{g} dx + \int_{-1}^{1} u_\epsilon \boldsymbol{g}_\epsilon \cdot \boldsymbol{\nu} dH_1. \tag{A.7}$$

Letting $\epsilon \to 0$ in the latter identity yields (A.4) at once for any smooth u. To obtain (A.3) we notice that for any $x \in (-1, 1)$ and any $0 < \rho < \min(1 - x, 1 + x)$, we have

$$\begin{aligned} \int_{S^+_{\rho(x)}} |u(z) - u^+(x)| dz &= \int_{-\rho}^{\rho} d\eta \int_0^{\rho} |u(x + \eta, t) - u^+(x)| dt \\ &\leq \int_{-\rho}^{\rho} d\eta \int_0^{\rho} |u(x + \eta, t) - u^+(x + \eta)| dt + \\ &\quad + \rho \int_{-\rho}^{\rho} |u^+(x + \eta) - u^+(x)| d\eta. \end{aligned}$$

Now we infer from (A.6) and Fubini Theorem that

$$\int_{-\rho}^{\rho} d\eta \int_0^{\rho} |u(x + \eta, t) - u^+(x + \eta)| dt \leq \rho \int_{S^+_\rho(x)} |\nabla u| dx.$$

Thus

$$\frac{1}{\rho^2} \int_{S^+_{\rho(x)}} |u(z) - u^+(x)| dz \leq \frac{1}{\rho} \int_{S^+_\rho(x)} |\nabla u| dx + \frac{1}{\rho} \int_{-\rho}^{\rho} |u^+(x + \eta) - u^+(x)| d\eta.$$

But by Lemma A.2, we have for $\mathcal{H}_1$-almost every $x \in (-1, 1)$

$$\lim_{\rho \to 0} \frac{1}{\rho} \int_{S^+_\rho(x)} |\nabla u| dx = 0$$

and so by Lebesgue's Theorem applied to $u^+ \in L^1(-1, 1)$, we obtain inequality (A.3) when u is smooth.

For $u \in BV(\Omega)$, let $u_n \in C^\infty(\Omega) \cap BV(\Omega)$ be such that $u_n \xrightarrow{L^1(\Omega)} u$ and $\int_\Omega |\nabla u_n| dx \to |Du|(\Omega)$. Then, from (A.3), we easily get that for $\mathcal{H}_1$-almost every $x \in \partial\Omega$

$$\lim_{\rho \to 0} \frac{1}{\rho^2} \int_{S^+_\rho(x)} |u(z) - u_n^+(x)| dz = 0, \ \ \forall n \in \mathbb{N}.$$

So all the traces u_n^+ are equal and defining $u^+ = u_n^+$, we obtain (A.3) at once from the above inequality. Finally, writing the Gauss-Green identity (A.4) for each u_n, then taking the limit as $n \to \infty$ yields (A.4) for u and the proof is complete. □

We have the following result showing that one can extend a BV function u over a rectangular domain by reflection across one side of the domain without making that side into a singular curve for u.

LEMMA A.4. *The linear map $\mathcal{R} : BV(S_+) \to BV(S)$ defined by $\mathcal{R}u(x_1, x_2) = u(x_1, |x_2|)$ is bounded and satisfies*

$$|D\mathcal{R}u|\left((-1,1)\times(-r,r)\right) \leq 2|Du|\left((-1,1)\times(0,r)\right) \quad \forall\, r \in (0,1]. \tag{A.8}$$

In particular, we have

$$|D\mathcal{R}u|(S_0) = 0. \tag{A.9}$$

PROOF. Let $u \in BV(S_+)$ be fixed. The restriction of $\mathcal{R}u$ to $S \setminus \bar{S}_+$ belongs to $BV(S \setminus \bar{S_+})$ and its trace on S_0 is equal to the trace of u on S_0. Consequently, by Gauss-Green theorem we have

$$\int_S \mathcal{R}u \,\mathrm{div}(\boldsymbol{g})dx = -\int_{S_+} \langle Du, \boldsymbol{g}\rangle - \int_{S_+} \langle Du, \tilde{\boldsymbol{g}}\rangle, \;\; \forall \boldsymbol{g} \in C_c^1(S, \mathbb{R}^2),$$

where

$$\tilde{\boldsymbol{g}}(x_1, x_2) = (g_1(x_1, -x_2), -g_2(x_1, -x_2)).$$

On taking the supremum over all $\boldsymbol{g}$ such that $|\boldsymbol{g}(x)| \leq 1, \forall x \in S$, we get that $\mathcal{R}u \in BV(S)$ and $J(\mathcal{R}(u)) \leq 2J(u)$. In fact, we have

$$J(\mathcal{R}(u)) = 2J(u), \tag{A.10}$$

since it is easy to check that $J(\mathcal{R}(u)) \geq 2J(u)$.

A similar argument with $\boldsymbol{g} \in C_c^1((-1,1)\times(-r,r), \mathbb{R}^2)$ yields (A.8) and the proof is complete. On taking the limit as r goes to 0 in inequality (A.8) above, we see that $|D\mathcal{R}u|(S_0) = 0$. □

A.2 GENERALIZATION TO LIPSCHITZ DOMAINS

In this section, we construct an extension operator of the space $BV(\Omega)$ into the space $BV(\mathbb{R}^2)$ when Ω is a Lipschitz domain. We begin with the definition of a Lipschitz domain.

DEFINITION A.5. We say that an open set Ω is Lipschitz continuous if for every $x \in \partial\Omega =: \Gamma$ there exist a rectangular neighborhood R of x in $\mathbb{R}^2$ and a bijective map $H : S \to R$ such that

(L1) H and H^{-1} are Lipschitz continuous, and H^{-1} maps negligible sets to negligible sets;

(L2) $H(S_+) = \Omega \cap R$;

(L3) $H(S_0) = R \cap \Gamma$.

The map H is called a local chart.

We will need the following result on the right-composition of a BV function with an invertible Lipschitz function satisfying property (L1) in Definition A.5.

LEMMA A.6. *Let Ω be a bounded domain in $\mathbb{R}^2$ and $\varphi : \Omega \to \Omega'$ be a Lipschitz invertible map satisfying property (L1). Then the mapping $\#_\varphi \colon u \mapsto u \circ \varphi^{-1}$ is a bounded linear map between the spaces $BV(\Omega)$ and $BV(\Omega')$ and*

$$J(\#_\varphi u) \leq \textit{Lip}(\varphi) J(u), \;\; \forall u \in BV(\Omega), \tag{A.11}$$

where Lip(φ) is the Lipschitz constant of φ. Furthermore $\#_\varphi$ maps $W^{1,1}(\Omega)$ in $W^{1,1}(\Omega')$ and

$$\nabla(\#_\varphi u)(y) = \nabla u(\varphi^{-1}(y)) \left[\nabla\varphi(\varphi^{-1}(y))\right]^{-1} \;\; \textit{for a.e. } y \in \Omega'. \tag{A.12}$$

PROOF. See the proofs of Theorem 3.16 and Corollary 3.19 in [8, pp. 127–130]. □

We will also need the following standard result.

LEMMA A.7 (Partittion of unity). *Suppose* Ω *is bounded and let* $\{U_i \colon i = 1, 2, \ldots, n\}$ *be an open cover on its boundary* $\partial\Omega$*. Then there exist a family of smooth function* $\{\theta_i, i = 0, 1, \ldots, n\} \subset C^\infty(\mathbb{R}^2)$ *such that*

(a) $\sum_{i=0}^{n} \theta_i = 1$ *on* $\mathbb{R}^2$ *and* $0 \leq \theta_i \leq 1 \;\; \forall\, i = 0, 1, \ldots, n$*;*

(b) $\theta_i \in C_c^\infty(U_i) \;\; \forall\, i = 1, 2, \ldots, n$ *and* $\theta_0 \in C_c^\infty(\Omega)$.

In our study of the approximation of the image denoising problem, we needed to extend the solution in $BV(\Omega)$ to a function in $BV(\mathbb{R}^2)$. We will see below that in the specific case of Lipschitz domains, we can construct an extension operator that coincide with the standard extension of $W^{1,p}(\Omega)$ to $W^{1,p}(\mathbb{R}^2)$ for any $p \in [1, \infty]$.

THEOREM A.8. *Suppose that* Ω *is bounded Lipschitz domain. Then there exists a linear and continuous extension operator* $T : BV(\Omega) \to BV(\mathbb{R}^2)$ *such that*

(a) $|DTu|(\Gamma) = 0$ *for any* $u \in BV(\Omega)$*;*

(b) the restriction of T *to* $W^{1,1}(\Omega)$ *is a bounded linear operator into* $W^{1,1}(\mathbb{R}^2)$.

PROOF. Since $\Gamma := \partial\Omega$ is compact and Lipschitz continuous, there exists an rectangular open cover $\{R_i, i = 1, 2 \ldots, n\}$ and bijective Lipschitz maps $\varphi_i \colon S \to R_i$ such that the properties (L1), (L2) and (L3) hold for each $1 \leq i \leq n$. Let $\{\theta_i\}_{1 \leq i \leq n}$ the a partition of unity subordinate to the open cover $\{R_i\}_{1 \leq i \leq n}$. Given $u \in BV(\Omega)$, we write

$$u = \sum_{i=0}^{n} \theta_i u = \sum_{i=0}^{n} u_i, \;\; \text{with } u_i = \theta_i u.$$

We observe that for any $i = 1, 2, \ldots, n$, $u_i \in BV(R_i \cap \Omega)$, and $u_0 \in BV(\Omega)$ with compact support. Now, we extend each u_i to $\mathbb{R}^2$, distinguishing between u_0 and $u_i,\ 1 \leq i \leq n$.

Extension of u_0. Since u_0 has compact support in Ω a natural extension is the zero extension defined by

$$\bar{u}_0(x) = \begin{cases} u_0(x) & \text{it } x \in \Omega, \\ 0 & \text{if } x \notin \Omega. \end{cases}$$

Given $\boldsymbol{g} \in C_c^1(\mathbb{R}^2, \mathbb{R}^2)$, we have

$$\int_{\mathbb{R}^2} \bar{u}_0 \operatorname{div}(\boldsymbol{g}) dx = -\int_{\Omega} u \operatorname{div}(\theta_0 \boldsymbol{g}) dx + \int_{\Omega} u \nabla(\theta_0) \cdot \boldsymbol{g} dx$$

so that on taking the supremum on all $\boldsymbol{g}$ with $\||\boldsymbol{g}|\|_\infty := \sup_{x \in \mathbb{R}^2} |\boldsymbol{g}(x)| \leq 1$, we obtain

$$J(\bar{u}_0) \leq C(\theta_0) \|u\|_{BV} \text{ with } C(\theta_0) = \|\theta_0\|_\infty + \||\nabla(\theta_0)|\|_\infty,$$

and it follows that $\|u_0\|_{BV} \leq C\|u\|_{BV}$. Moreover the support of $\bar{u}_0$ is relatively compact in Ω.

Extension of u_i, $i = 1, 2, \ldots, k$. Let v_i be the restriction of u to $R_i \cap \Omega$ transfered to S_+ via the local chart φ_i, *i.e.* $v_i(y) = u \circ \varphi(y)$ for $y \in S_+$. By Lemma A.6 $v_i \in BV(S_+)$, and Lemma A.4 guarantees that $\mathcal{R}v_i \in BV(S)$ and $|D\mathcal{R}v_i|(S_0) = 0$. We know from Lemma A.6 that $w_i = \#_{\varphi_i}\mathcal{R}v_i$ belongs to $BV(R_i)$ and $u_i = u$ on $R_i \cap \Omega$. Now, define

$$\bar{u}_i = \begin{cases} \theta_i(x) w_i(x), & \text{if } x \in R_i, \\ 0 & \text{if } x \notin R_i, \end{cases}$$

so that $\bar{u}_i \in BV(\mathbb{R}^2)$ and $\|\bar{u}_i\|_{BV} \leq C_i \|u\|_{BV}$ where C_i depends only on θ_i and φ_i. Moreover, the support of $\bar{u}_i$ is relatively compact in R_i and we have

$$|D\bar{u}_i|(\partial\Omega) = 0.$$

The operator T. The extension operator is then defined by $Tu = \sum_{i=0}^{n} \bar{u}_i$ and possesses all the desired properties by construction. □

BIBLIOGRAPHY

[1] R. Acar and C. R. Vogel. Analysis of bounded variation penalty methods for ill-posed problems. *Inverse Problems*, 10(6):1217–1229, 1994.

[2] W. K. Allard. Total variation regularization for image denoising. I. Geometric theory. *SIAM J. Math. Anal.*, 39(4):1150–1190, 2007.

[3] F. Alter, V. Caselles, and A. Chambolle. A characterization of convex calibrable sets in $\mathbb{R}^N$. *Math. Ann.*, 332(2):329–366, 2005.

[4] F. Alter, V. Caselles, and A. Chambolle. Evolution of characteristic functions of convex sets in the plane by the minimizing total variation flow. *Interfaces Free Bound.*, 7(1):29–53, 2005.

[5] L. Alvarez, F. Guichard, P.-L. Lions, and J.-M. Morel. Axioms and fundamental equations of image processing. *Arch. Rational Mech. Anal.*, 123(3):199–257, 1993.

[6] L. Alvarez, P.-L. Lions, and J.-M. Morel. Image selective smoothing and edge detection by nonlinear diffusion. II. *SIAM J. Numer. Anal.*, 29(3):845–866, 1992.

[7] L. Alvarez and J.-M. Morel. Formalization and computational aspects of image analysis. In *Acta numerica, 1994*, Acta Numer., pages 1–59. Cambridge Univ. Press, Cambridge, 1994.

[8] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of bounded variation and free discontinuity problems*. Oxford Mathematical Monographs, 2000.

[9] F. Andreu, J. M. Mazón, J. S. Moll, and V. Caselles. The minimizing total variation flow with measure initial conditions. *Commun. Contemp. Math.*, 6(3):431–494, 2004.

[10] G. Aubert and P. Kornprobst. *Mathematical problems in image processing: partial differential equations and the calculus of variations*, volume 147 of *Applied Mathematical Sciences*. Springer-Verlag, New York, Second edition, 2006.

[11] J. Aujol. *Contribution à l'Analyse de Textures en Traitement d'Images par Méthodes Variationnelles et Equations aux Dérivées Partielles*. PhD thesis, Université de Nice Sophia Antipolis, June 2004.

[12] J. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle. Image decomposition into a bounded variation component and an oscillating component. *Journal of Mathematical Imaging and Vision*, 22(1):71–88, January 2005.

[13] J. Aujol and A. Chambolle. Dual norms and image decomposition models. *International Journal of Computer Vision*, 63(1):85–104, June 2005.

[14] J.-F. Aujol. Some first-order algorithms for total variation based image restoration. *J. Math. Imaging Vision*, 34(3):307–327, 2009.

[15] J.-F. Aujol and G. Gilboa. Constrained and SNR-based solutions for TV-Hilbert space image denoising. *J. Math. Imaging Vision*, 26(1-2):217–237, 2006.

[16] G. Awanou. Robustness of a spline element method with constraints. *J. Sci. Comput.*, 36(3):421–432, 2008.

[17] G. Awanou, M.-J. Lai, and P. Wenston. The multivariate spline method for scattered data fitting and numerical solution of partial differential equations. In *Wavelets and splines: Athens 2005*, Mod. Methods Math., pages 24–74. Nashboro Press, Brentwood, TN, 2006.

[18] G. M. Awanou and M. J. Lai. On convergence rate of the augmented Lagrangian algorithm for nonsymmetric saddle point problems. *Appl. Numer. Math.*, 54(2):122–134, 2005.

[19] A. Beck and M. Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Trans. Image Process.*, 18(11):2419–2434, 2009.

[20] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.*, 2(1):183–202, 2009.

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

[22] M. Benzi, G. H. Golub, and J. Liesen. Numerical solution of saddle point problems. *Acta Numer.*, 14:1–137, 2005.

[23] A. Bermúdez and C. Moreno. Duality methods for solving variational inequalities. *Comput. Math. Appl.*, 7(1):43–58, 1981.

[24] D. P. Bertsekas. *Constrained optimization and Lagrange multiplier methods*. Computer Science and Applied Mathematics. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1982.

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

[26] S. C. Brenner and L. R. Scott. *The mathematical theory of finite element methods*, volume 15 of *Texts in Applied Mathematics*. Springer, New York, third edition, 2008.

[27] H. Brezis. *Functional analysis, Sobolev spaces and partial differential equations*. Universitext. Springer, New York, 2011.

[28] J. L. Carter. *Dual methods for total variation-based image restoration*. PhD thesis, University of California, Los Angeles, 2001.

[29] V. Caselles, A. Chambolle, and M. Novaga. Regularity for solutions of the total variation denoising problem. *Rev. Mat. Iberoam.*, 27(1):233–252, 2010.

[30] F. Catté, P.-L. Lions, J.-M. Morel, and T. Coll. Image selective smoothing and edge detection by nonlinear diffusion. *SIAM J. Numer. Anal.*, 29(1):182–193, 1992.

[31] A. Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imaging Vision*, 20(1-2):89–97, 2004. Special issue on mathematics and image analysis.

[32] A. Chambolle. Total variation minimization and a class of binary MRF models. In A. Rangarajan, B. Vemuri, and A. Yuille, editors, *Energy Minimization Methods in Computer Vision and Pattern Recognition*, volume 3757 of *Lecture Notes in Computer Science*, pages 136–152. Springer Berlin / Heidelberg, 2005.

[33] A. Chambolle, V. Caselles, D. Cremers, M. Novaga, and T. Pock. An introduction to total variation for image analysis. In *Theoretical foundations and numerical methods for sparse recovery*, volume 9 of *Radon Ser. Comput. Appl. Math.*, pages 263–340. Walter de Gruyter, Berlin, 2010.

[34] A. Chambolle, S. Levine, and B. J. Lucier. An upwind finite-difference method for total variation-based image smoothing. *SIAM J. Imaging Sciences*, 4(1):277–299, 2011.

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

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

[37] T. F. Chan and S. Esedoḡlu. Aspects of total variation regularized L^1 function approximation. *SIAM J. Appl. Math.*, 65(5):1817–1837 (electronic), 2005.

[38] T. F. Chan, G. H. Golub, and P. Mulet. A nonlinear primal-dual method for total variation-based image restoration. *SIAM J. Sci. Comput.*, 20(6):1964–1977, 1999.

[39] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Transactions on Image Processing*, 6(2):298–310, 1997.

[40] P. G. Ciarlet. *Introduction to numerical linear algebra and optimisation*. Cambridge Texts in Applied Mathematics. Cambridge University Press, Cambridge, 1989. With the assistance of Bernadette Miara and Jean-Marie Thomas, Translated from the French by A. Buttigieg.

[41] D. C. Dobson and C. R. Vogel. Convergence of an iterative method for total variation denoising. *SIAM J. Numer. Anal.*, 34(5):1779–1791, 1997.

[42] V. Duval, J.-F. Aujol, and L. A. Vese. Mathematical modeling of textures: Application to color image decomposition with a projected gradient algorithm. *J Math Imaging Vis*, 37:232–248, 2010.

[43] I. Ekeland and R. Temam. *Convex analysis and variational problems*. Society for Industrial Mathematics, 1987.

[44] E. Esser, X. Zhang, and T. F. Chan. A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM J. Imaging Sci.*, 3(4):1015–1046, 2010.

[45] X. Feng and A. Prohl. Analysis of total variation flow and its finite element approximations. *M2AN Math. Model. Numer. Anal.*, 37(3):533–556, 2003.

[46] E. Giusti. *Minimal Surfaces and Functions of Bounded Variation*, volume 80 of *Monographs in Mathematics*. Birkhäuser, 1984.

[47] Y. Gousseau and J.-M. Morel. Are natural images of bounded variation? *SIAM J. Math. Anal.*, 33(3):634–648 (electronic), 2001.

[48] Q. Hong. *Bivariate splines for image enhancements based on variational models*. PhD thesis, The University of Georgia, 2011.

[49] H. W. Kuhn and A. W. Tucker. Nonlinear programming. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950*, pages 481–492, Berkeley and Los Angeles, 1951. University of California Press.

[50] M. Lai, B. Lucier, and J. Wang. The convergence of a central-difference discretization of rudin-osher-fatemi model for image denoising. *Scale Space and Variational Methods in Computer Vision*, pages 514–526, 2009.

[51] M.-J. Lai, C. Liu, and P. Wenston. Bivariate spline method for numerical solution of time evolution Navier-Stokes equations over polygons in stream function formulation. *Numer. Methods Partial Differential Equations*, 19(6):776–827, 2003.

[52] M.-J. Lai, C. Liu, and P. Wenston. Numerical simulations on two nonlinear biharmonic evolution equations. *Appl. Anal.*, 83(6):563–577, 2004.

[53] M.-J. Lai and L. M. Messi. Piecewise linear approximation of the continuous Rudin-Osher-Fatemi model for image denoising. *SIAM J. Numer. Anal.*, 2012. Revised.

[54] M.-J. Lai and L. L. Schumaker. *Spline functions on triangulations*, volume 110 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2007.

[55] M.-J. Lai and P. Wenston. Bivariate spline method for numerical solution of steady state Navier-Stokes equations over polygons in stream function formulation. *Numer. Methods Partial Differential Equations*, 16(2):147–183, 2000.

[56] M.-J. Lai and P. Wenston. Trivariate C^1 cubic splines for numerical solution of biharmonic equations. In *Trends in approximation theory (Nashville, TN, 2000)*, Innov. Appl. Math., pages 225–234. Vanderbilt Univ. Press, 2001.

[57] M.-J. Lai, P. Wenston, and L. an Ying. Bivariate C^1 cubic splines for exterior biharmonic equations. In *Approximation theory, X (St. Louis, MO, 2001)*, Innov. Appl. Math., pages 385–404. Vanderbilt Univ. Press, Nashville, TN, 2002.

[58] M. Lysaker, S. Osher, and X.-C. Tai. Noise removal using smoothed normals and surface fitting. *IEEE Trans. Image Process.*, 13(10):1345–1357, 2004.

[59] Y. Meyer. *Oscillating patterns in image processing and nonlinear evolution equations*, volume 22 of *University Lecture Series*. American Mathematical Society, Providence, RI, 2001. The fifteenth Dean Jacqueline B. Lewis memorial lectures.

[60] Y. Nesterov. *Introductory lectures on convex optimization*, volume 87 of *Applied Optimization*. Kluwer Academic Publishers, Boston, MA, 2004. A basic course.

[61] Y. Nesterov. Smooth minimization of non-smooth functions. *Math. Program.*, 103(1, Ser. A):127–152, 2005.

[62] Y. Nesterov. Primal-dual subgradient methods for convex problems. *Math. Program.*, 120(1, Ser. B):221–259, 2009.

[63] Y. E. Nesterov. A method for solving the convex programming problem with convergence rate $O(1/k^2)$. *Dokl. Akad. Nauk SSSR*, 269(3):543–547, 1983.

[64] M. Nikolova. Local strong homogeneity of a regularized estimator. *SIAM J. Appl. Math.*, 61(2):633–658 (electronic), 2000.

[65] M. Nikolova. A variational approach to remove outliers and impulse noise. *J. Math. Imaging Vision*, 20(1-2):99–120, 2004. Special issue on mathematics and image analysis.

[66] P.-O. Persson and G. Strang. A simple mesh generator in Matlab. *SIAM Rev.*, 46(2):329–345 (electronic), 2004.

[67] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. Global solutions of variational models with convex regularization. *SIAM J. Imaging Sciences*, 3(4):1122–1145, 2010.

[68] R. T. Rockafellar. *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.

[69] J. B. Rosen. The gradient projection method for nonlinear programming. I. Linear constraints. *J. Soc. Indust. Appl. Math.*, 8:181–217, 1960.

[70] J. B. Rosen. The gradient projection method for nonlinear programming. II. Nonlinear constraints. *J. Soc. Indust. Appl. Math.*, 9:514–532, 1961.

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

[72] The MathWorks, Inc. *MATLAB® (R2011a)*. The MathWorks, Inc., Natick, Massachusetts, April 2011.

[73] C. R. Vogel and M. E. Oman. Iterative methods for total variation denoising. *SIAM J. Sci. Comput.*, 17(1):227–238, 1996.

[74] J. Wang and B. J. Lucier. Error bounds for finite-difference methods for Rudin-Osher-Fatemi image smoothing. *SIAM J. Numer. Anal.*, 49(2):845–868, 2011.

[75] C. Wu and X.-C. Tai. Augmented Lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models. *SIAM J. Imaging Sci.*, 3(3):300–339, 2010.

[76] M. Zhu, S. J. Wright, and T. F. Chan. Duality-based algorithms for total-variation-regularized image restoration. *Comput. Optim. Appl.*, 47(3):377–400, 2010. Published online: 6 December 2008.