Novel higher order regularisation methods for image reconstruction

Konstantinos Papafitsoros

St Edmunds College

University of Cambridge

A thesis submitted for the degree of

Doctor of Philosophy

June 2014

Abstract

In this thesis we study novel higher order total variation-based variational methods for digital image reconstruction. These methods are formulated in the context of Tikhonov regularisation. We focus on regularisation techniques in which the regulariser incorporates second order derivatives or a sophisticated combination of first and second order derivatives. The introduction of higher order derivatives in the regularisation process has been shown to be an advantage over the classical first order case, i.e., total variation regularisation, as classical artifacts such as the staircasing effect are significantly reduced or totally eliminated. Also in image inpainting the introduction of higher order derivatives in the regulariser turns out to be crucial to achieve interpolation across large gaps.

First, we introduce, analyse and implement a combined first and second order regularisation method with applications in image denoising, deblurring and inpainting. The method, numerically realised by the split Bregman algorithm, is computationally efficient and capable of giving comparable results with total generalised variation (TGV), a state of the art higher order method. An additional experimental analysis is performed for image inpainting and an online demo is provided on the IPOL website (Image Processing Online).

We also compute and study properties of exact solutions of the one dimensional total generalised variation problem with L^2 data fitting term, for simple piecewise affine data functions, with or without jumps . This gives an insight on how this type of regularisation behaves and unravels the role of the TGV parameters.

Finally, we introduce, study and analyse a novel non-local Hessian functional. We prove localisations of the non-local Hessian to the local analogue in several topologies and our analysis results in derivative-free characterisations of higher order Sobolev and BV spaces. An alternative formulation of a non-local Hessian functional is also introduced which is able to produce piecewise affine reconstructions in image denoising, outperforming TGV.

Keywords: Higher order total variation, functions of bounded Hessian, total generalised variation, denoising, deblurring, inpainting, staircasing effect, split Bregman, exact TGV solutions, non-local Hessian, characterisation of higher order Sobolev and BV spaces.

Acknowledgements

I would like to thank at this point all the people that made this journey towards the completion of this Ph.D. thesis a pleasant one and made the last four years truly memorable.

Firstly, I would like to thank Carola Schönlieb for being an excellent supervisor from every perspective. Her continuous help and encouragement played a great role to my academic evolution. She was the one that introduced me to the field of mathematical imaging and showed me how a modern professional mathematician should be.

The people I lived with in Cambridge played an important role during my Ph.D.. A big thanks goes to Sara Merino for being the best flatmate and friend possible for the last four years. The same amount of gratitude goes as well to Nayia Constantinou and Luca Calatroni. I could not have asked for better flatmates and good friends.

A stimulating environment in the office makes the working hours pleasant and never boring. Bati Sengul and Spencer Hughes were an excellent pair of officemates and friends and I thank them a lot for that.

I would like to thank Thalia Sotirchou for her love and support during the demanding last year of the Ph.D..

Many thanks to all my friends in CCA and CMS, Marc Briant, Kolyan Ray (my gym-mate), Damon Civin, Ed Mottram, Vaggelis Papoutsellis, Julio Brau, Ioan Manolescu and also Panagiotis Kosteletos for all the nice moments we had together.

Moreover, I would like to thank my collaborators Jan Lellmann, Kristian Bredies and Daniel Spector as well as Martin Benning and Tuomo Valkonen for useful discussions.

I am grateful to the CCA directors Arieh Iserles and James Norris for giving me the opportunity to be part of it and to Emma Hacking for providing excellent administrative support. I also acknowledge the financial support provided by the Engineering and Physical Sciences Research Council, the Cambridge Centre for Analysis, the department of Pure Mathematics and Mathematical

Statement of Originality

I hereby declare that my dissertation entitled "Novel higher order regularisation methods for image reconstruction" is not substantially the same as any that I have submitted for a degree or diploma or other qualification at any other university. I further declare that it contains nothing which is the outcome of work done in collaboration with others, except where specifically indicated in the text.

Chapter 1 is a review of the existing literature on variational methods for image reconstruction. It places the dissertation into context and summarises its results. It is a result of my own personal work.

Chapter 2 introduces some necessary mathematical preliminaries for the thesis. It consists of a review of some of the known results of measure theory, functions of bounded variation and convex analysis. It is not original material.

Chapter 3 is dedicated to the introduction, analysis and implementation of the $TV–TV^2$ method for image reconstruction and it is novel material. It is a result of collaboration and it is contained in [PS14, PSS13]. The proofs as well as the implementation of the method were written by me, apart from Theorem 3.4.7 proved by my supervisor Carola Schönlieb from whom I also greatly benefited through several discussions. The C code implementation for $TV–TV^2$ inpainting that was used in the IPOL online demonstration was written by my collaborator Bati Sengul, University of Cambridge.

Chapter 4 contains an analysis of the one dimensional total generalised variation denoising problem. It is novel material and it has been done in collaboration with Kristian Bredies, University of Graz, Austria, and follows [PB13]. All the proofs were written by me apart from Proposition 4.4.3 while I greatly benefited from my two visits in Graz.

Chapter 5 studies two forms of non-local Hessian functionals following [LPSS]. It is novel material and it has been done in collaboration with Jan Lellmann, Carola Schönlieb, University of Cambridge and Daniel Spector, Technion, Israel. All the proofs are the result of my own personal work while I benefited from discussions with all the three aforementioned collaborators. The numerical implementation in Section 5.7 is due to my collaborator Jan Lellmann.

Contents

List of Notation

Against each entry is the page at which the notation is introduced.

Sets and Measures:

$\mathbb{R}^+$	41	The set of positive real numbers.
$\overline{\mathbb{R}}$	49	The set of extended real numbers, $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$.
$\mathcal{S}^{d \times d}$	30	The set of $d \times d$ symmetric matrices.
Ω	42	An open subset of $\mathbb{R}^d$.
$\mathcal{B}(X)$	39	The Borel σ-algebra of X.
$|\mu|$	41	The total variation measure of the finite Radon measure μ.
$\mathcal{M}(X)$	41	The set of real valued finite Radon measures on X. Endowed with the norm $\|\mu\|_{\mathcal{M}(X)} = |\mu|(X)$.
$\mathcal{M}(X, \mathbb{R}^\ell)$	41	The set of $\mathbb{R}^\ell$–valued finite Radon measures on X. Endowed with the norm $\|\mu\|_{\mathcal{M}(X)} = |\mu|(X)$.
$\mathcal{M}^+(X)$	41	The set of finite positive Radon measures on X.
$\mathcal{M}^+_{loc}(X)$	41	The set of positive Radon measures on X.
$\|\mathfrak{T}\|_{\mathcal{M}}$	42	The Radon norm of the distribution $\mathfrak{T}$ on a set Ω: $\|\mathfrak{T}\|_{\mathcal{M}} = \sup\{\langle \mathfrak{T}, v \rangle : \ v \in \mathcal{C}_c^\infty(\Omega), \ \|v\|_\infty \leq 1\}$. If $\mathfrak{T}$ is a finite Radon measure μ then $\|\mu\|_{\mathcal{M}} = |\mu|(\Omega)$.
$\mathcal{L}$	39	The Lebesgue measure on $\mathbb{R}$.
$\mathcal{L}^d$	39	The Lebesgue measure on $\mathbb{R}^d$.
$\mathcal{H}$	26	The one dimensional Hausdorff measure.
$\mathcal{H}^d$	46	The d–dimensional Hausdorff measure.
$\frac{\mu}{\nu}$	42	The Radon–Nikodým density of μ with respect to ν.
$\mathrm{sgn}(\mu)$	42	Alternative notation for $\frac{\mu}{|\mu|}$.
μ^α	42	The absolutely continuous part of μ with respect to Lebesgue measure, i.e., $\mu^\alpha = \frac{\mu}{\mathcal{L}^d}$.
μ^s	42	The singular part of μ with respect to Lebesgue measure.

Spaces of continuous and differentiable functions:

Spaces of Lebesgue integrable functions:

Sobolev spaces:

$W^{k,p}(\Omega)$	40	Sobolev space of functions $u \in L^p(\Omega)$, such that the distributional derivatives up to order k are also L^p functions (weak derivatives). The corresponding norm is $\|u\|_{W^{k,p}(\Omega)} = \left(\sum_{|a| \le k} \int_\Omega |D^a u|^p dx \right)^{1/p}$, where D^a denotes the a–th distributional derivative of u and $|a| = a_1 + \cdots + a_d$ is the order of the multiindex $a = (a_1, \ldots, a_d)$.
$H^k(\Omega)$	40	The Sobolev space $W^{k,2}(\Omega)$.
$H_0^k(\Omega)$	40	The completion of $\mathcal{C}_c^\infty(\Omega)$ under the $\|\cdot\|_{H^k(\Omega)}$ norm.

Functions of bounded variation, bounded Hessian and bounded deformation:

$\mathrm{BV}(\Omega)$	44	The space of functions of bounded variation on Ω, i.e., all the functions $u \in L^1(\Omega)$, such that Du is a $\mathbb{R}^d$–valued finite Radon measure. The corresponding norm is $\|u\|_{\mathrm{BV}(\Omega)} = \|u\|_{L^1(\Omega)} + |Du|(\Omega)$.
$\mathrm{BV}(\Omega, \mathbb{R}^\ell)$	44	The space of $\mathbb{R}^\ell$–valued functions of bounded variation on Ω, i.e., all the functions $u \in L^1(\Omega, \mathbb{R}^\ell)$, such that Du is a $\mathbb{R}^{\ell \times d}$–valued finite Radon measure. The corresponding norm is $\|u\|_{\mathrm{BV}(\Omega, \mathbb{R}^\ell)} = \|u\|_{L^1(\Omega, \mathbb{R}^\ell)} + |Du|(\Omega)$.
$\mathrm{TV}(u)$	44	The total variation of $u \in L^1(\Omega, \mathbb{R}^\ell)$, equal to $|Du|(\Omega)$ for a function $u \in \mathrm{BV}(\Omega, \mathbb{R}^\ell)$.
$D^2 u$	60	The second order distributional derivative of u.
$\mathrm{BV}^2(\Omega)$	29	The space of functions of bounded Hessian on Ω, i.e., all the functions $u \in W^{1,1}(\Omega)$ such that $\nabla u \in \mathrm{BV}(\Omega, \mathbb{R}^d)$. The corresponding norm is $\|u\|_{\mathrm{BV}^2(\Omega)} = \|u\|_{\mathrm{BV}(\Omega)} + |D^2 u|(\Omega)$.
$\mathrm{TV}^2(u)$	29	The second order total variation of u, equal to $|D^2 u|(\Omega)$ for a function $u \in \mathrm{BV}^2(\Omega)$.
$\mathcal{E}u$	30	The distributional symmetrised gradient of u.
$\mathrm{ICTV}_{\beta,\alpha}(u)$	29	The TV–TV^2 infimal convolution of u with weights α and β, i.e., $\mathrm{ICTV}_{\beta,\alpha}(u) = \inf_{v \in \mathrm{BV}^2(\Omega)} \alpha \, \mathrm{TV}(u - v) + \beta \, \mathrm{TV}^2(v)$.
$\mathrm{TGV}^2_{\beta,\alpha}(u)$	30	The second order total generalised variation of u with weights $\alpha, \beta > 0$.
$\mathrm{BD}(\Omega)$	30	The space of functions of bounded deformation on Ω, i.e., all the functions $u \in L^1(\Omega, \mathbb{R}^d)$ such that $\mathcal{E}u$ is a $\mathbb{R}^{d \times d}$–valued finite Radon measure.
$D^\alpha u$ or ∇u	44	The absolutely continuous part of Du with respect to Lebesgue measure, i.e., shorter version of $(Du)^\alpha = \frac{Du}{\mathcal{L}^d}$.
$D^s u$	44	The singular part of Du with respect to Lebesgue measure, i.e., shorter version of $(Du)^s$.
u'	44	Short version of ∇u, when u is an one dimensional BV function.
$\mathrm{pV}(u, (a, b))$	47	The pointwise variation of u in $(a, b) \subseteq \mathbb{R}$.
$\mathrm{eV}(u, (a, b))$	47	The essential variation of u in $(a, b) \subseteq \mathbb{R}$.

J_u	48	The jump set of u.
$u(x_-),\ u(x_+)$	48	Left and right limits of u at the point x.
$u^l,\ u^r$	48	Left and right continuous representatives of u.
$u^{up},\ u_{low}$	48	Good representatives of u, defined as $u^{up}(x) = \max\{u^l(x),\ u^r(x)\}$ and $u_{low}(x) = \min\{u^l(x),\ u^r(x)\}$.
u^Ω	46	The trace of u in $\partial\Omega$ for a function $u \in \mathrm{BV}(\Omega)$.

Miscellaneous notation:

u_Ω	46	The mean value of u in Ω, i.e., $u_\Omega = \frac{1}{\mathcal{L}^d(\Omega)}\int_\Omega u\,dx$.
$\mathcal{X}_A$	21	The characteristic function of the set A, i.e., $$\mathcal{X}_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases}$$
$\mathcal{I}_A$	51	The indicator function of the set A, i.e., $$\mathcal{I}_A(x) = \begin{cases} 0 & \text{if } x \in A, \\ \infty & \text{if } x \notin A. \end{cases}$$
φ_∞	59	The recession function of $\varphi : \mathbb{R}^\ell \to \mathbb{R}$, defined for every $x \in \mathbb{R}^\ell$: $\varphi_\infty(x) := \lim_{t\to\infty}\frac{\varphi(tx)}{t}$.
$\mathrm{sc}_\tau F$	50	The lower semicontinuous envelope (relaxed functional) of $F : X \to \mathbb{R}$, with respect to the topology τ.
X^*	50	The analytic dual of the space X. Its elements are denoted by $x^*, y^* \ldots$.
F^*	51	The convex conjugate of a function F. If $F : X \to \overline{\mathbb{R}}$, then $F^* : X^* \to \overline{\mathbb{R}}$ is defined as $F^*(x^*) = \sup_{x\in X}\langle x^*, x\rangle - F(x)$.
Λ^*	52	The adjoint operator $\Lambda^* : Y^* \to X^*$ of the bounded linear functional $\Lambda : X \to Y$, where X, Y are Banach spaces.
∂F	51	The subdifferential of F.
$\mathrm{Sgn}(\mu)$	114	The set valued sign of the finite Radon measure μ on Ω, defined as: $\mathrm{Sgn}(\mu) = \{v \in L^\infty(\Omega) \cap L^\infty(\Omega,\lvert\mu\rvert) :\ \lVert v\rVert_{L^\infty(\Omega)} \le 1,\ v = \mathrm{sgn}(\mu),\ \lvert\mu\rvert - a.e.\}$.
$f^\diamond$	120	The L^1–linear regression of f, $f^\diamond := \underset{v \text{ affine}}{\mathrm{argmin}}\ \lVert f - v\rVert_{L^1(\Omega)}$.
$f^\star$	121	The L^2–linear regression of f, $f^\star := \underset{v \text{ affine}}{\mathrm{argmin}}\ \lVert f - v\rVert^2_{L^2(\Omega)}$.
$\mathcal{G}_n u$	37	Non-local gradient of u (explicit formulation).
$\mathcal{H}_n u$	37	Non-local Hessian of u (explicit formulation).
$\mathcal{G}_{\sigma_x} u(x)$	37	Non-local gradient of u at x with weighting function σ_x (implicit formulation).
$\mathcal{H}_{\sigma_x} u(x)$	37	Non-local Hessian of u at x with weighting function σ_x (implicit formulation).
$d^2 u(x,y)$	37	Second order finite difference scheme: $d^2 u(x,y) = u(y) - 2u(x) + u(x + (x - y))$.

Assumed notation:

$\mathbb{N}$	-	The set of natural numbers.
$\mathbb{R}$, $\mathbb{R}^d$	-	The set of real numbers and its d–Cartesian product. In both cases the Euclidean norm is denoted by $|\cdot|$.
$\mathbb{R}^{\ell \times d}$	-	The set of $\ell \times d$ real matrices, equipped with the Frobenious norm denoted by $|\cdot|$.
$D \Subset \Omega$	-	$\overline{D} \subseteq \Omega$ with $\overline{D}$ being compact.
$\mu \lfloor A$	-	The restriction of the measure μ to the set A.
δ_x	-	The Dirac atomic measure concentrated on $\{x\}$.
$B(x, \delta)$	-	The open ball in $\mathbb{R}^d$ with center x and radius δ.
$a \otimes b$	-	The $\ell \times d$ matrix with (ij)–th entry $a_i b_j$, for $a \in \mathbb{R}^\ell$, $b \in \mathbb{R}^d$.
$\mathrm{div}\, u$	-	Divergence of $u : \Omega \to \mathbb{R}^d$, i.e., $\mathrm{div}\, u = \sum_{i=1}^d \frac{\partial u_i}{\partial x_i}$.
$\mathrm{div}^2 u$	-	Second order divergence of $u : \Omega \to \mathbb{R}^{d \times d}$, i.e., $\sum_{i,j=1}^d \frac{\partial^2 u_{ij}}{\partial x_i \partial x_j}$.
Δu	-	The Laplacian of $u : \Omega \to \mathbb{R}$, i.e., $\Delta u = \sum_{i=1}^d \frac{\partial^2 u_i}{\partial x_i^2}$.

Chapter 1

Introduction

1.1 Digital images, image processing and their mathematics

This is "the earliest surviving camera photograph" [Hir08]:

Figure 1.1: Joseph Nicéphore Niépce: Point de vue du Gras, 1826 or 1827. Source: http://en.wikipedia.org/wiki/History_of_photography.

Even though undoubtedly a photograph like the above represents a real breakthrough for humanity, it would be considered an image of very low quality in the modern world. High quality digital cameras dominate the photography world of our time and people always look for the technically and visually perfect photo. Joseph Nicéphore Niépce had no hope to improve his photograph and that was not only because he was lacking the proper – by today's standards – camera equipment. Even a few decades ago, when film cameras were the photographer's only choice, the post-processing techniques were only accessible to the selected few that possessed the secrets of the darkroom. Yet, those techniques were fairly limited, allowing only for some simple colour, brightness and contrast adjustment. However, the development of computers and their unavoidable use in our everyday life arrived soon and all the information started to be encoded in bits and bytes. Photographs were no exception and digital photography was born. An image is no longer seen as a

chemical mark left on a film but as a collection of numbers. Even though admittedly this sounds less romantic, it provides a convenient framework for us mathematicians, to see digital images as functions: A typical digital image of resolution $N \times M$, that is to say $N \cdot M$ pixels, can be modelled as a function $u = (u_{\mathrm{red}}, u_{\mathrm{green}}, u_{\mathrm{blue}})$

$$u : \{1, \ldots, N\} \times \{1, \ldots, M\} \to \{0, \ldots, 255\}^3, \qquad (1.1)$$

where each component measures the intensity of the corresponding channel (RGB images). This representation gives now a new meaning to image processing. For example, the sentence "Given a digital image which has some undesirable characteristics, like noise, blur etc., we would like to obtain a visually and aesthetically better version of it" can be interpreted as "Given a function that has certain characteristics and some local or global regularity properties, we would like to obtain another function, close to the initial one in some sense, with potentially different regularity properties and with some desirable features". Mathematicians use the term *functional* to denote an object that takes a function as an input and gives another function as an output. Thus, past complicated processing procedures in the mini lab, like contrast and colour adjustment, are just simple linear operations in the new language. However, more complicated image processing tasks like noise removal require more sophisticated "functionals" or in other words, mathematical methods and algorithms. If today we can perform these tasks it is not only due to the introduction of computers but mainly because of the development of an advanced mathematical technology both at a theoretical and an applied level. Therefore, there is a duality between the increasingly challenging image processing problems and the progress of novel mathematical theories. While a new mathematical method is most of the times introduced to tackle a specific image processing task, it often turns out that this method leads to fruitful mathematical theories that are interesting in their own right. A mathematician working in mathematical imaging likes to prove things about a method, not only in order to provide rigorously justified insights about the behaviour of the method but also because he/she is intrigued by the mathematical beauty of the theory behind it.

We do not claim that the present thesis is solely about image processing. Even though, we are motivated by complicated image processing problems and we introduce novel methods, most of the results can find their own autonomous place in the mathematical literature (especially Chapter 5).

1.2 The variational approach for image reconstruction

We have already mentioned that a digital image can be considered as a vector valued function on a $N \times M$ grid. However, in order to have a convenient and familiar mathematical framework, the analysis is usually done in the continuous setting (returning to the discrete

setting for the numerical implementation). Thus, an RGB colour image u is regarded as a function

$$u : \Omega \to \mathbb{R}^3, \tag{1.2}$$

where $\Omega \subseteq \mathbb{R}^2$ is typically a rectangular domain. Analogously greyscale images are modelled as real valued functions on Ω, i.e., $u : \Omega \to \mathbb{R}$.

Suppose now that we have some image acquisition device. In an ideal world that device should provide us with a *perfect* image u. Of course, the world perfect is subjective but we ignore any artistic views on the matter. Thus for us, features like noise, blur, artificial artifacts are always undesirable. Quite often, due to imperfections and limitations in the acquisition process we obtain a corrupted version f of the image u. Corruptions could be noise or blur for instance. A typical model in imaging assumes that u has been transformed through a continuous and linear operation T with the additional presence of random noise. In particular we assume that the two images fulfil the model

$$f = Tu + \eta. \tag{1.3}$$

Here η denotes a random noise component that follows a certain distribution. Typically, noise is modelled by Gaussian, Poisson, uniform, "salt and pepper" (impulse) distributions or even a combination of these. The operator T is called the *forward operator* of the problem. Examples are the identity operator (in that case the only undesirable feature is noise), blurring operators (in which case Tu denotes the convolution with a blurring kernel) and pointwise multiplication with the characteristic function of $\Omega \setminus D$, $\mathcal{X}_{\Omega \setminus D}$, where $D \subseteq \Omega$ is a subdomain of Ω (inpainting domain) where no data are known at all and in which information from the known part should be interpolated (image inpainting).

In Figure 1.2 we give some examples of corrupted images for different forward operators T and noise distributions. Depending on the nature of the operator T the reconstruction process is given a specific name. The classical imaging tasks considered in this thesis are the following:

(i) **Image denoising**: T is the identity operator, i.e., only noise is present, see Figures 1.2(b) and 1.2(d).

(ii) **Image deblurring**: T represents a convolution (blurring) operator and noise might or might not be present, see Figure 1.2(c).

(iii) **Image inpainting**: In this case $Tu = \mathcal{X}_{\Omega \setminus D} u$, where $D \subseteq \Omega$ is a domain of missing information and noise might or might not be present, see Figures 1.2(e) and 1.2(f).

In order to reconstruct the initial image u one has to invert the operator T. However, this is not always possible. For example, in the case of inpainting the operator T is not invertible at all while in the case of deblurring, even though the operator could be

(a) The original image u.

(b) Image f corrupted with Gaussian noise: $f = u + \eta$, where η follows a Gaussian distribution.

(c) Image f corrupted by Gaussian blur and noise: $f = Tu + \eta$, where T denotes a convolution with a Gaussian kernel and η follows a Gaussian distribution.

(d) Image f corrupted by impulse noise: $f = u + \eta$, where η follows a "salt and pepper" distribution (only a percentage of pixels are corrupted).

(e) Image f with missing information (white letters): $f = \mathcal{X}_{\Omega \setminus D} u$, where D denotes the area covered by the letters.

(f) Image f with missing information (white letters) and corrupted by noise: $f = \mathcal{X}_{\Omega \setminus D} u + \eta$, where D denotes the area covered by the letters and η follows a Gaussian distribution.

Figure 1.2: Examples of corrupted images for different operators T and noise distributions.

invertible in theory, its inversion involves an ill-conditioned problem even in the absence of noise. The presence of noise further complicates the problem. In this case, it is a common procedure to add some a priori information to the model, which in general is given by a certain regularity assumption on the image u. This is the so-called regularisation process. In its most general setting one tries to recover a reconstructed version of f as a minimiser of a functional which has the following form:

$$\mathcal{J}(u) = \Phi(Tu, f) + \Psi(u). \tag{1.4}$$

The function Φ is called the *data fidelity* or *data fitting term* and measures the distance between the data f and the reconstruction u after the forward operator has acted on it. Small values of this term ensure that the transformed image Tu is close to the data f in a suitable sense. The function Ψ is called the *regulariser* or *regularising term* and it imposes extra regularity on u. It is expected that small values of Ψ will lead, up to a

certain extent, to the elimination of the undesirable features in the corrupted data f. The two terms are typically balanced by one or more positive parameters within Ψ. The choice of regulariser also decides upon the space in which the minimisation of $\mathcal{J}$ is well-posed. It is normally chosen to be a Banach space X on which Ψ takes finite values.

As one can speculate, given the problem (1.3), a bad/good choice of the data fidelity term can lead to a reconstruction of high/poor quality. It is well known in the mathematical imaging community that the correct choice of the fidelity term depends on the noise distribution. For example if the noise follows a Gaussian distribution, the appropriate choice for Φ is the squared L^2 norm of the difference $Tu - f$, i.e., $\Phi(Tu, f) = \frac{1}{2}\int_\Omega (Tu - f)^2 dx$. In the case of impulse noise, it has been shown in various contexts [Nik04, Nik02, CE05, DAG09, WGO07] that the L^1 norm of the difference $Tu - f$, $\int_\Omega |Tu - f| dx$ leads to more efficient restorations. In the case of Poisson noise, the Kullback–Leibler divergence of Tu and f, $\int_\Omega (Tu - f \log Tu) dx$ is the most suitable choice [LCA07, BSW$^+$13, SBJM09] while the supremum norm $\|Tu - f\|_{L^\infty(\Omega)}$ should be preferred in the case of uniform noise [Cla12]. Table (1.1) provides a summary of the above. Noise of mixed distributions can also be considered. In that case a combination of

Type of noise	$\Phi(Tu, f)$
Gaussian	$\frac{1}{2}\int_\Omega (Tu - f)^2 dx$
Impulse	$\int_\Omega |Tu - f| dx$
Poisson	$\int_\Omega (Tu - f \log Tu) dx$
Uniform	$\|Tu - f\|_{L^\infty(\Omega)}$

Table 1.1: The correct choices of data fidelity terms for different noise distributions.

the corresponding fidelities is used [NWC13, DlRS12, CDlRS14, HL13].

While the statistics of the noise give rise to a particular choice of the fidelity term, the choice of the regulariser is determined by our a priori knowledge of the properties of the desired image u. Different regularisers impose different regularity and qualitative characteristics on the image u. Squared Hilbert space norms have a long tradition in inverse problems. The most prominent example is *Tikhonov* regularisation [Tik63, Whi23] with

$$\Psi(u) = \alpha \int_\Omega |\nabla u|^2 dx, \tag{1.5}$$

where $\alpha > 0$ and ∇u is a weak gradient of u, which – for Gaussian noise – leads to the following minimisation of (1.4):

$$\min_{u \in H^1(\Omega)} \frac{1}{2} \int_\Omega (Tu - f)^2 dx + \alpha \int_\Omega |\nabla u|^2 dx. \tag{1.6}$$

The well-posedness of (1.6) follows from a standard application of the direct method of calculus of variations, exploiting the fact that $H^1(\Omega)$ is reflexive and thus bounded sets in that space are weakly pre-compact. In the denoising case, i.e., when T is the identity, the gradient flow of the corresponding Euler–Lagrange equation of (1.6) reads $u_t = \alpha\Delta u - u + f$. The result of such a regularisation technique is a linearly, i.e., isotropically, smoothed image u for which the smoothing strength does not depend on f. Hence, while eliminating the disruptive noise in the given data, also desirable structures like edges in the reconstructed image are blurred, see Figure 1.3. This observation gave way to a new class of non-smooth norm regularisers, which aim to eliminate noise and smooth the image in homogeneous areas, while preserving the relevant structures such as objects boundaries and edges.

(a) The original image u. (b) Image f corrupted with Gaussian noise (c) Denoised image with Tikhonov regularisation.
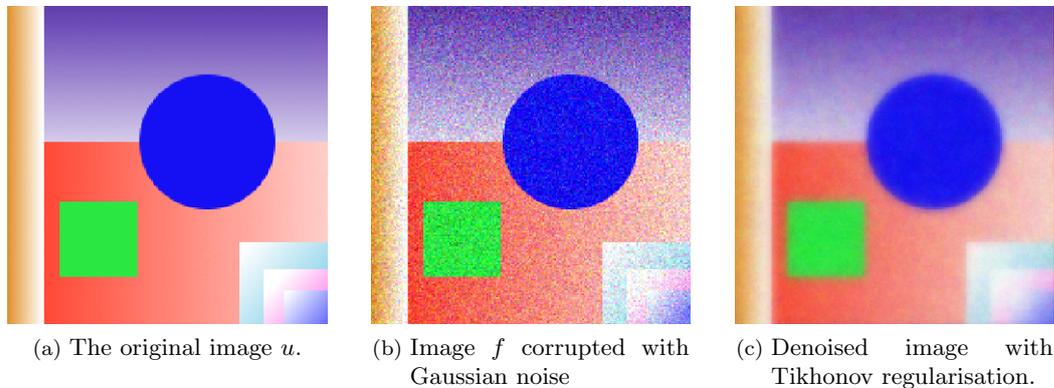
Figure 1.3: Illustration of the isotropically smoothed reconstruction using Tikhonov regularisation, i.e., when $\Psi(u) = \alpha\|\nabla u\|^2_{L^2(\Omega)}$.

1.3 Total variation based methods

In their pioneering work, Rudin, Osher and Fatemi [ROF92] proposed the use of total variation of u, TV(u), as regulariser. Recall that for a function $u \in L^1(\Omega)$, the total variation of u is defined as

$$\mathrm{TV}(u) = \sup\left\{\int_\Omega u\,\mathrm{div}v\,dx : \ v \in \mathcal{C}^1_c(\Omega,\mathbb{R}^2),\ \|v\|_\infty \leq 1\right\}, \qquad (1.7)$$

where $\mathcal{C}^1_c(\Omega,\mathbb{R}^2)$ is the space of $\mathbb{R}^2$–valued continuously differentiable functions with compact support in Ω. Functions that have finite total variation are exactly those whose distributional derivative Du can be represented by a finite Radon measure (space of functions of bounded variation, BV(Ω)), see Chapter 2 for a complete account on this subject. Recall also that if $u \in W^{1,1}(\Omega)$, one can easily show that $\mathrm{TV}(u) = \int_\Omega |\nabla u| dx$. The

corresponding L^2 fidelity minimisation problem now reads

$$\min_{u \in BV(\Omega)} \frac{1}{2} \int_\Omega (Tu - f)^2 dx + \alpha \, TV(u). \tag{1.8}$$

The well-posedness of (1.8) is shown again using the direct method of calculus of variations together with some compactness properties of the space $BV(\Omega)$ (Theorem 2.3.5). When $T = Id$, the above problem is commonly known as the ROF problem. Here, less regularity than in (1.6) is imposed in the image u, as BV functions are allowed to have jump discontinuities. This non-smooth regularisation procedure results in a nonlinear smoothing of the image, smoothing more in homogeneous areas and preserving characteristic structures such as edges. In particular, total variation regularisation is tuned towards the preservation of edges and performs very well for piecewise constant images, where typically almost a perfect reconstruction is obtained, up to a slight loss of contrast, see for example Figure 1.4.

(a) Piecewise constant image.   (b) Image   corrupted   with   (c) TV restoration.
                                  Gaussian noise
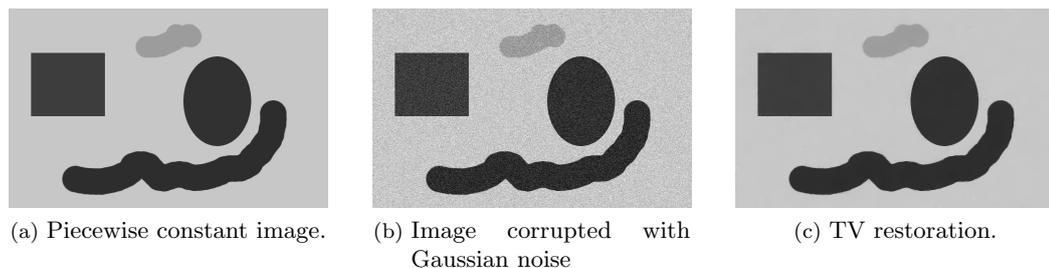
Figure 1.4: Illustration of the ability of TV regularisation to preserve edges in the reconstructed image and its efficient restoration of piecewise constant images.

Since its introduction in image processing in 1992, total variation has found numerous applications in the field, as the edge preservation ability made it quite an appealing tool. These applications cover a wide area, from image denoising [ROF92, CL97, Nik04, LCA07, BC98, CKS01, DHKS10, Pas12a], image deblurring [DS96, CW98, WYZ07, BT09, Pas12b], image inpainting [SC02, CS05, Pas12c] and image zooming [ACHR06] to JPEG decompression [ADF05, BH12b] MRI and PET reconstruction [BGH$^+$14, BMPS14, Mül13] to name a few.

Properties of total variation regularisation have been extensively studied at the theoretical level as well. Acar and Vogel [AV94] studied the well-posedness of problem (1.8). Meyer [Mey01] showed that in the L^2–TV denoising problem with f being the characteristic function of a disk, the solution is a characteristic function of the same disk with a loss of contrast, whose amount depends on the parameter α, see Figure 1.5. Alter, Caselles and Chambolle [ACC05] extended these results for characteristic functions of certain convex bounded sets with a bound on the mean curvature of their boundaries (calibrable
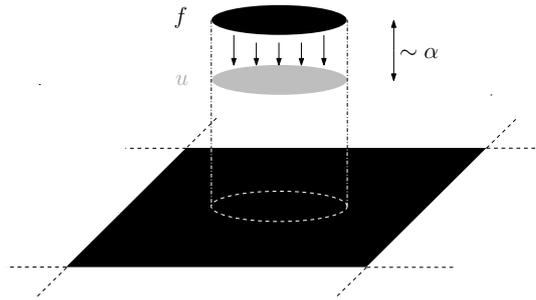
Figure 1.5: Illustration of Meyer's result: Total variation regularisation on a characteristic function of a disk in $\mathbb{R}^2$ results only to a loss of contrast which is proportional to the parameter α.

sets). For the same problem, Strong and Chan [SC03] computed exact solutions for 1D and rotationally symmetric 2D data and similar results were obtained by Ring [Rin00]. In [Gra07] Grasmair proved the equivalence of one dimensional TV regularisation and the taut–string algorithm and an extension to higher dimensions was done by Hinterberger et al. [HHK$^+$03]. Properties of the L^1–TV model were studied in [Nik02, DAG09, CE05]. Finally, Caselles, Chambolle and Novaga [CCN07] showed that the discontinuity set of the solution u of the L^2–TV denoising problem in $\Omega \subseteq \mathbb{R}^d$, is contained, up to a $\mathcal{H}^{d-1}$–null set, in the discontinuity set of the data function f. Here $\mathcal{H}^{d-1}$ denotes the $(d-1)$–dimensional Hausdorff measure in $\mathbb{R}^d$. Further regularity properties of the solution u of total variation denoising were proved by Allard [All08a, All08b, All09].

As it is observed in all the aforementioned applications and rigorously justified analytically, total variation regularisation promotes piecewise constant reconstructions leading to images with blocky like structures. This generally becomes apparent in images that do not only consist of flat regions and jumps but also possess slanted regions, i.e., piecewise affine parts, see Figure 1.6. This is called the *staircasing effect*. In one dimension, this roughly
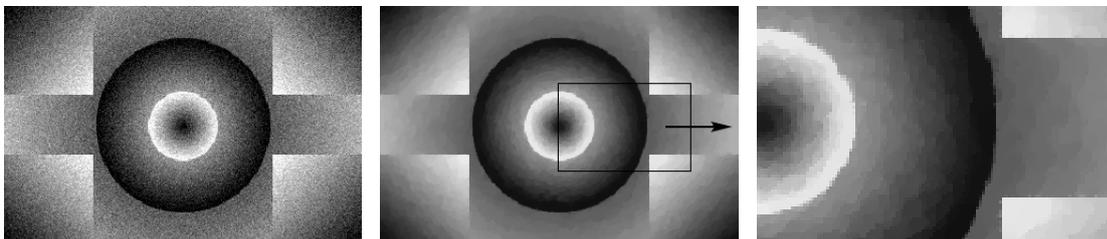
(a) Image corrupted with Gaussian noise

(b) Total variation denoised image with blocky-like artifacts.

(c) Detail of total variation denoised image.

Figure 1.6: Illustration of the staircasing effect in total variation denoising in two dimensions.

means that the total variation regularisation of a noisy signal f, is a staircase u, whose L^2 norm is close to f, see Figure 1.7. In fact it can be rigorously proved that if in an open in-
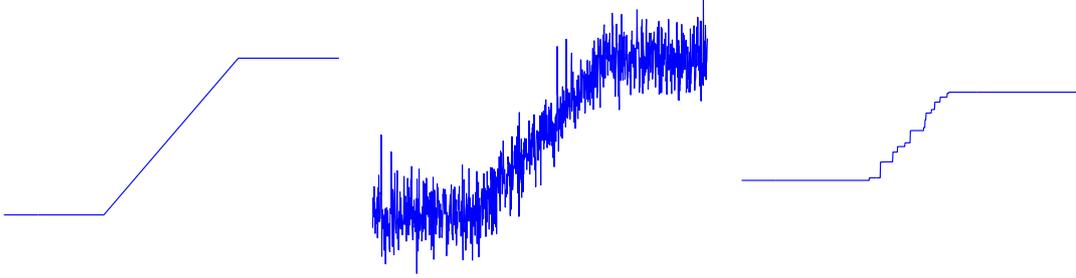
Figure 1.7: The staircasing effect in total variation denoising in one dimension.

terval I we have $u > f$ (or $u < f$), then u is constant on I [CCN07, Rin00, BKV13]. Since u and f are BV functions defined a.e., the inequalities $u > f$ and $u < f$ are interpreted through inequalities of certain good representatives of them, see Chapter 2 for the relevant definitions. Analytical studies of the staircasing effect are also done in [DMFLM09, DS96]. Since it leads to unnaturally looking images, the staircasing effect is considered an undesirable artifact and many methods have been proposed in order to reduce or eliminate it, see Section 1.4 for a detailed discussion.

The ability of total variation regularisation to preserve and promote discontinuities on the image u has also been exploited in image inpainting [SC02, CS05, Pas12c]. There, the corresponding minimisation problem reads

$$\min_{u \in \mathrm{BV}(\Omega)} \frac{1}{2} \int_{\Omega \setminus D} (u - f)^2 dx + \alpha \,\mathrm{TV}(u), \tag{1.9}$$

where D is the inpainting domain, i.e., the missing part of the image. A desirable feature in image inpainting is the interpolation of the image along large gaps (*connectivity principle*). Unfortunately, it turns out that total variation is able to provide interpolation results with sharp edges only if the gap is small enough [CKS02, Sch09]. In fact total variation minimisation can be interpreted as penalisation of the length of level lines interpolating them via the shortest path. In Figure 1.8 for example, we see that interpolation of a broken stripe is achieved only if the gap length is smaller than the width of the stripe, while Tikhonov regularisation (harmonic inpainting) leads to blurry reconstructions as expected, with no connection occurring either. Total variation achieves perfect sharp reconstruction of the stripe for small gaps but fails completely for large gaps. Also, following its general characteristic, total variation promotes piecewise constant reconstructions inside the inpainting domain, resulting again in images with blocky, unnaturally looking features, see Figure 1.9. In order to overcome these issues, several inpainting methods based on higher order derivatives have been suggested, see Section 1.4.

(a) Small gap.

(b) TV inpainting.

(c) Harmonic inpainting.

(d) Large gap.

(e) TV inpainting.

(f) Harmonic inpainting.
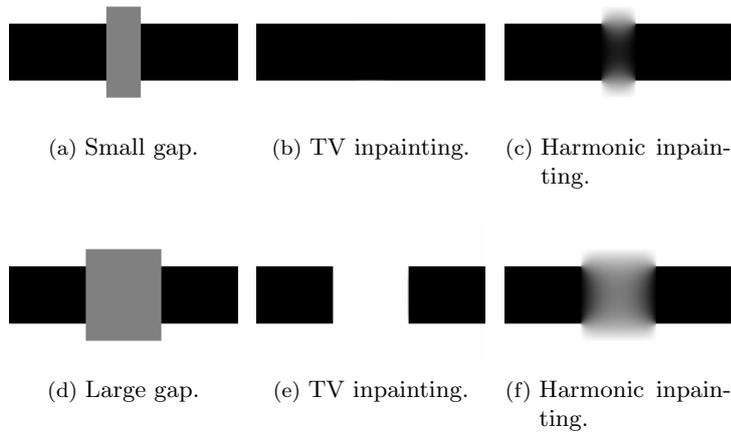
Figure 1.8: Illustration of the dependence of the connection ability of total variation inpainting on the gap size and comparison with harmonic inpainting ($\Psi(u) = \alpha\|\nabla u\|^2_{L^2(\Omega)}$). The inpainting domain D is denoted with grey colour.

(a) Destroyed image.

(b) TV inpainted image.

(c) Detail of TV inpainted image.
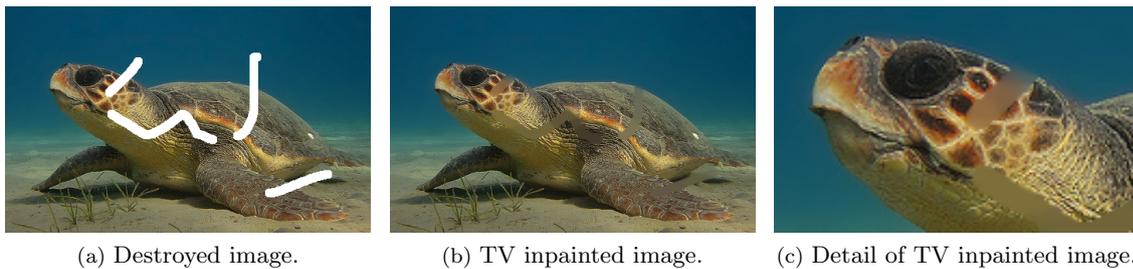
Figure 1.9: Total variation inpainting promotes piecewise constant reconstructions inside the inpainting domain which is unnatural for real world images.

1.4 Higher order variational models

In recent years, higher order versions of non-smooth image reconstructions methods have been considered in order to reduce the staircasing effect and hence to improve the quality of the restored image, as well as to fulfil the connectivity principle in inpainting. Some incorporate a combination of first and higher order derivatives – mainly second order – and some others are purely higher order. Intuitively, minimising second order derivative–based energies, one is trying to recover a piecewise affine version of the data in contrast to a piecewise constant version obtained by total variation regularisation. This is expected, roughly because, due to the introduction of the second order derivatives in the regulariser, its kernel now contains affine functions and not only constants.

Already in the paper of Chambolle and Lions [CL97], the authors proposed a higher order method for denoising by means of an infimal convolution of two convex regularisers.

Here, a noisy image is splitted into $u = u_1 + u_2$ by solving

$$\min_{\substack{u_1 \in \mathrm{BV}(\Omega) \\ u_2 \in \mathrm{BV}^2(\Omega)}} \frac{1}{2} \int_\Omega (u_1 + u_2 - f)^2 dx + \alpha \, \mathrm{TV}(u_1) + \beta \, \mathrm{TV}^2(u_2). \qquad (1.10)$$

where $\mathrm{TV}^2(u_2) := \mathrm{TV}(\nabla u_2)$ and $\mathrm{BV}^2(\Omega) = \{u \in W^{1,1}(\Omega) : \ \nabla u \in \mathrm{BV}(\Omega)\}$, the space of *functions of bounded Hessian* [Dem85]. The idea is to decompose the image u into a piecewise constant part u_1 and a piecewise affine part u_2. An alternative formulation for the minimisation problem (1.10) is

$$\min_{u \in \mathrm{BV}(\Omega)} \frac{1}{2} \int_\Omega (u - f)^2 dx + \mathrm{ICTV}_{\beta,\alpha}(u), \qquad (1.11)$$

with

$$
\begin{aligned}
\mathrm{ICTV}_{\beta,\alpha}(u) \quad &:= \quad \min_{v \in \mathrm{BV}^2(\Omega)} \alpha \, \mathrm{TV}(u - v) + \beta \, \mathrm{TV}^2(v) \\
&= \quad \min_{\substack{w \in \mathrm{BV}(\Omega, \mathbb{R}^2) \\ w \in \mathcal{R}(\nabla, \mathbb{R}^2)}} \alpha \|Du - w\|_{\mathcal{M}} + \beta \|Dw\|_{\mathcal{M}}.
\end{aligned} \qquad (1.12)
$$

Here $\mathrm{BV}(\Omega, \mathbb{R}^2)$ is the space of $\mathbb{R}^2$–valued functions of bounded variation, $\mathcal{R}(\nabla, \mathbb{R}^2) = \{w \in L^1(\Omega, \mathbb{R}^2) : \ \exists v \in W^{1,1}(\Omega) \text{ such that } w = \nabla v\}$ is the range of the operator ∇ and $\|\mathfrak{T}\|_{\mathcal{M}}$ is the Radon norm of the distribution $\mathfrak{T}$ in Ω, $\|\mathfrak{T}\|_{\mathcal{M}} = \sup\{\langle \mathfrak{T}, v \rangle : \ v \in \mathcal{C}_c^\infty(\Omega), \ \|v\|_\infty \le 1\}$. Note that for a finite Radon measure μ on Ω, we have $\|\mu\|_{\mathcal{M}} = |\mu|(\Omega)$. Again, the well-posedness of (1.10) and hence (1.11) is shown using the direct method of calculus of variations. Note that while $u_1 + u_2$ is unique in (1.10), one can always add a constant to u_1 and subtract the same constant from u_2, thus constructing another solution. In that sense, u_1 and u_2 are not unique but their sum is.

Another attempt to combine first and second order regularisation originates from Chan, Marquina, and Mulet [CMM01], who considered total variation minimisation together with weighted versions of the Laplacian. More precisely, they considered a regularising term of the form

$$\alpha \int_\Omega |\nabla u| dx + \beta \int_\Omega \varphi(|\nabla u|)(\Delta u)^2 dx, \qquad (1.13)$$

where φ must be a function with certain growth conditions at infinity in order to allow jumps. The well-posedness of (1.13) in one dimension was rigorously analysed by Dal Maso et al. in [DMFLM09]. In [CEP07] the term $\|\Delta u_2\|_{L^2(\Omega)}^2$ was used instead of $\mathrm{TV}^2(u_2)$ in (1.10), also in combination with the use of H^{-1} norm in the fidelity term.

Regularising with second order derivatives only, taking $\Psi(u) = \alpha \, \mathrm{TV}^2(u)$ has also been considered by Lysaker et al., [LLT03], Scherzer [Sch98], Hinterberger and Scherzer [HS06a], Bergounioux and Piffet [BP10], Lai et al. [LTC13]. In Lefkimmiatis et al. [LBU12], the use of spectral norm of the Hessian matrix was proposed. Further, Pöschl and Scherzer

[PS08] studied minimisers of functionals which are regularised by the total variation of the k–th derivative. Finally, a combination of total variation with a fourth order PDE was considered by Lysaker and Tai in [LT06] while a higher order nonlinear diffusion filtering denoising technique was studied by Didas at al. in [DWB09].

All the above methods manage to reduce the staircasing effect up to one degree but not totally eliminate it. For example the decomposition obtained by the infimal convolution, is not the desirable one since the effect of the TV^2 term is not strong enough to make the reconstruction piecewise smooth. On the other hand regularising only with TV^2 introduces a fair amount of blur in the image. Moreover, some of the problems are difficult to be implemented and optimised – problem (1.13) is not even convex – while others are only tuned for image denoising, lacking the wide applicability of total variation regularisation.

Total Generalised Variation

A high quality regulariser that eliminates the common staircasing artifacts is the total generalised variation (TGV) introduced recently by Bredies, Kunisch and Pock [BKP10]. The second order total generalised variation of a function $u \in L^1(\Omega)$ with positive parameters α and β is defined as

$$\mathrm{TGV}_{\beta,\alpha}^2(u) = \sup \left\{ \int_\Omega u \operatorname{div}^2 v \, dx : \ v \in \mathcal{C}_c^2(\Omega, \mathcal{S}^{2\times2}), \ \|v\|_\infty \le \beta, \ \|\operatorname{div} v\|_\infty \le \alpha \right\}, \quad (1.14)$$

where $\mathcal{C}_c^2(\Omega, \mathcal{S}^{2\times2})$ is the space of $\mathcal{S}^{2\times2}$–valued twice continuously differentiable functions with compact support in Ω, with $\mathcal{S}^{2\times2}$ being the space of 2×2 symmetric matrices. Definition (1.14) is called the *supremum* or the *predual* definition of TGV. An equivalent formulation, the *minimum* or *primal* definition, was shown in [BV11]

$$\mathrm{TGV}_{\beta,\alpha}^2(u) = \min_{w \in \mathrm{BD}(\Omega)} \alpha \|Du - w\|_\mathcal{M} + \beta \|\mathcal{E}w\|_\mathcal{M}, \quad (1.15)$$

where $\mathcal{E}w$ is the distributional symmetrised gradient of w and $\mathrm{BD}(\Omega)$ is the space of functions of bounded deformation, i.e., the functions whose symmetrised distributional gradient can be represented by a finite Radon measure, see [TS80, Tem85]. One should note the difference between $\mathrm{ICTV}_{\beta,\alpha}$ and $\mathrm{TGV}_{\beta,\alpha}^2$ by observing the two definitions (1.12) and (1.15). While the minimisation in the $\mathrm{ICTV}_{\beta,\alpha}$ definition (1.12) is restricted over the functions that are gradients, the minimisation in (1.15) is done over the wider space $\mathrm{BD}(\Omega)$ leading eventually for a more optimal decomposition in piecewise constant and piecewise affine parts [Mül13]. As a result, the total generalised variation–based regularisation, i.e., minimisation of (1.4) with $\Psi(u) = \mathrm{TGV}_{\beta,\alpha}^2(u)$, has the ability to adapt to the regularity of the data and images restored with this method are typically piecewise smooth that is to say, not only the discontinuities but also the affine structures are preserved, see Figure 1.10. Let us note here that a modified infimal convolution approach that has been recently

proposed in the discrete setting by Setzer et al. in [SST11] has very similar results with
TGV and in fact in some if its variants the two methods are equivalent.

(a) Noisy image. (b) Total variation denoising. (c) Total generalised variation denoising.
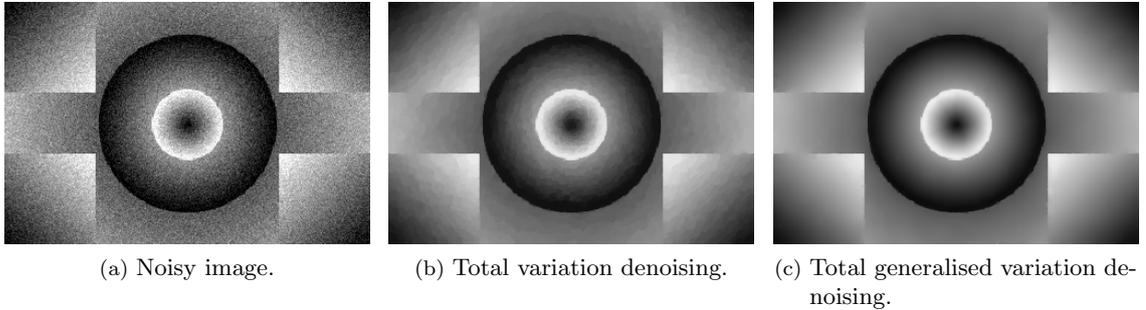
Figure 1.10: Illustration of the ability of total generalised variation to preserve edges and
affine structures in the same time, thus eliminating the staircasing artifacts.

Some successful applications of total generalised variation in image restoration and related tasks have been done in image denoising [BKP10, Bre14, BDH13], image deblurring
[BV11, Bre14], image zooming [Bre14, BH13b], MRI reconstruction [KBPS11, BGH+14],
diffusion tensor imaging [VBK13] and JPEG decompression [BH12a, Hol13]. Moreover
there have already been a few contributions to the analysis of the model. Properties of
the one dimensional L^1–TGV model were studied by Bredies, Kunisch and Valkonen in
[BKV13], while in [BH13a] Bredies and Holler studied the well-posedness of the corresponding Tikhonov regularisation problem established in the BD space. A comparison
among TV, TV^2, ICTV and TGV regularisation is done in [BBBM13] by Benning et al.
in the context of eigenfunctions, see also Benning's Ph.D. thesis [Ben11]. In particular, the
authors investigate the capability of the three methods to recover certain one dimensional
data exactly apart from a loss of contrast (almost exact recovery). They compute exact
solutions of the corresponding minimisation problems for data that can be recovered almost exactly by one method but not the others. Additionally in his Ph.D. thesis [Mül13],
Müller constructed a 2D function that can be recovered almost exactly by TGV but not by
ICTV, emphasising the differences between the two functionals in two dimensions. Note
that ICTV and TGV regularisations coincide in dimension one since in that case every L^1
function can be written as a weak derivative of another function, compare again definitions
(1.12) and (1.15).

The main disadvantage of TGV regularisation problem is the increased computational
time involved to its solution. The most popular numerical method that is used for this
purpose is the primal-dual algorithm of Chambolle–Pock [CP11]. Even though TGV minimisation is solved fairly fast with this algorithm, it is much slower than TV minimisation
solved with the state of the art methods that have been developed for it, e.g. the split
Bregman algorithm [GO09].

Inpainting methods

Higher order inpainting methods perform in general better than first order ones, like total variation inpainting, because of the additional directional information used for the interpolation process. Euler's elastica is a popular higher order variational method of this kind which is able to achieve large gap connectivity, see Nitzberg et al. [NMS93], Masnou and Morel [MM98], Chan, Kang and Shen [CKS02] and also Tai, Hahn and Chung [THC11]. In Euler's elastica, the regularising term reads

$$\int_{\Omega} \left(\alpha + \beta \left(\nabla \cdot \frac{\nabla u}{|\nabla u|} \right)^2 \right) |\nabla u| \, dx, \tag{1.16}$$

i.e., is a combination of the total variation and the curvature of the level lines of u. However, one disadvantage of the Euler's elastica inpainting model for analytic and numerical issues is that it is a non-convex minimisation problem.

Convex approximations to the elastica energy have been studied in [BPW13]. Other examples of higher order inpainting are the Cahn–Hilliard inpainting by Bertozzi, Esedoglu and Gillette [BEG07], TV–H^{-1} inpainting by Burger et al. [BHS09], Schönlieb et al. [SBBH09] and Hessian–based surface restoration by Lai et al. [LTC13]. Moreover, curvature driven diffusion inpainting has been considered by Chan and Shen [CS01] while Mumford–Shah based inpainting has been proposed by Esedoglu and Shen [ES02]. Finally, inpainting via transport (first order inpainting) has also been considered, see for instance Bertalmío et al. [BSCB00] and Bornemann and März [BM07].

All the aforementioned inpainting methods are *local* methods, that is to say the information that is needed to fill in the inpainting domain D is only taken from points neighboring the boundary of D. *Non-local* or *global* methods take into account all the information from the known part of the image, usually weighted by its distance to the point that is to be filled in. This class of methods is very powerful, allowing to fill in structures and textures almost equally well. However, they still have some disadvantages, for instance high computational cost is involved in their solution. Moreover local methods are sometimes more desirable, as they tend to preserve geometrical features better, especially in relatively small inpainting domains. For non-local inpainting methods the reader is referred to e.g. Cao at al. [CGMP09], Criminisi et al. [CPT03] and Arias, Caselles and Sapiro [ACS09].

1.5   Contribution

This thesis consists of three main distinguishable parts as these also result from the associated papers that have been produced by the author as part of his Ph.D. [PS14, PSS13], [PB13], [LPSS]. In [PS14], we introduce, analyse and implement an efficient combined first

and second order regularisation method with applications in image denoising and deblurring, also providing an additional experimental analysis for image inpainting accompanied by an online demonstration in [PSS13]. We contribute to the analysis of TGV in [PB13] by analytically solving the one dimensional L^2–TGV denoising problem for simple data functions and studying further properties of the model. Finally in [LPSS], we introduce and study a non-local Hessian functional, proving its localisations to the classical Hessian in several regularity levels, a work that also leads to some novel characterisations of higher order Sobolev and BV spaces in the spirit of Bourgain, Brezis, Mironescu [BBM01]. We also introduce an alternative definition of non-local Hessian, that is able to produce high quality piecewise affine reconstructions in image denoising as this is confirmed by numerical examples.

In what follows, let us give an outline for each of these contributions which will be presented in more details in Chapters 3, 4 and 5.

1.5.1   The TV–TV^2 approach for image reconstruction

In Chapter 3, we introduce the following combined first and second order regulariser (TV–TV^2) for a function $u \in \mathrm{BV}^2(\Omega)$:

$$\alpha \int_\Omega \varphi_1(\nabla u)\, dx + \beta\, \varphi_2(D^2 u)(\Omega), \tag{1.17}$$

where φ_1, φ_2 are two convex functions $\varphi_1 : \mathbb{R}^2 \to \mathbb{R}^+$, $\varphi_2 : \mathbb{R}^4 \to \mathbb{R}^+$ with at most linear growth at infinity. As usual α, β are positive regularisation parameters that balance the two terms. The definition is set up in the context of convex functions of measures – the term $D^2 u$ is generally a finite Radon measure. This is done in order to define rigorously expressions like $\sqrt{|D^2 u|^2 + \epsilon}$ (where $\phi_2(x) = \sqrt{|x|^2 + \epsilon}$) making them equal to $\sqrt{|\nabla u|^2 + \epsilon}$ in the special case when $D^2 u = \nabla^2 u \, \mathcal{L}^2$, i.e., when the measure $D^2 u$ is representable by a function $\nabla^2 u$. In the case when φ_2 is the Euclidean norm $|\cdot|$ the total variation of the measure $|D^2 u|(\Omega)$ is recovered. We study the corresponding variational problem

$$\min_{u \in \mathrm{BV}^2(\Omega)} \frac{1}{2} \int_\Omega (Tu - f)^2 dx + \alpha \int_\Omega \varphi_1(\nabla u)\, dx + \beta\, \varphi_2(D^2 u)(\Omega). \tag{1.18}$$

Under some mild additional assumptions, we prove the existence and uniqueness for solutions of (1.18), via the method of relaxation in the spirit of [Ves01], i.e., by identifying the minimising functional's lower semicontinuous envelope with respect to the naturally defined weak* and strict topologies in $\mathrm{BV}^2(\Omega)$. The method is implemented with the split Bregman algorithm [GO09] to whose convergence theory we also contribute and we present applications in image denoising and deblurring.

The idea of this combination of first and second order dynamics is to regularise with a fairly large weight α in the first order term preserving the jumps as well as possible and

using a not too large weight β for the second order term, in a way that staircasing artifacts created by the first order regulariser are reduced without introducing any serious blur in the reconstructed image. We show that for image denoising and deblurring the model (1.18) offers solutions whose quality (accessed by the structural similarity index SSIM [WBSS04, WB09]) is not far off from the ones produced by the TGV. Moreover, the computational effort needed for its numerical solution is not much more than the one needed for solving the standard total variation model. For comparison the numerical solution for TGV regularisation is in general about ten times slower than this, see corresponding table in Section 3.6.

We proceed in applying the TV–TV^2 method in image inpainting, where we observe that we can achieve connectivity over larger gaps than total variation. Thus, the method results in more naturally looking inpainted images see Figure 1.11. The minimisation of

| (a) Destroyed image. | (b) TV inpainted image. | (c) TV^2 inpainted image. |

| (d) Detail of destroyed image. | (e) Detail of TV inpainted image. | (f) Detail of TV^2 inpainted image |
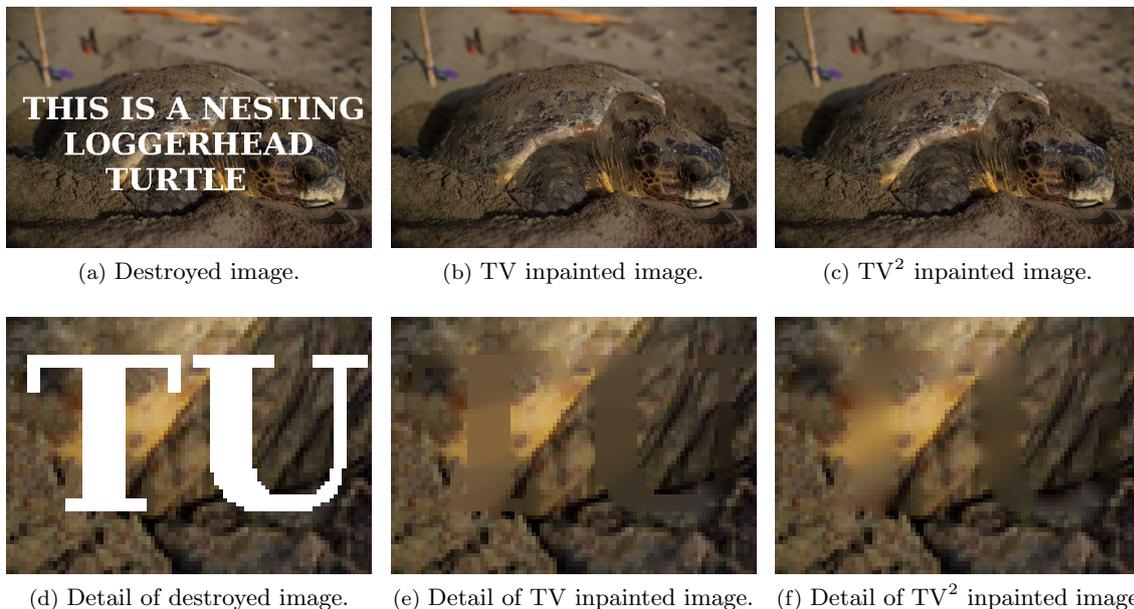
Figure 1.11: Illustration of the differences between pure TV and pure TV^2 inpainting reconstructions.

the TV–TV^2 functional can be seen as convex simplification of the Euler's elastica idea, where we have replaced the non-convex curvature by the convex total variation of the gradient of u. In an experimental analysis we indicate the differences between pure first and pure second order total variation inpainting, we study the influence of the inpainting domain on large gap connection and we propose a numerically justified rule for optimal tuning of the parameters within the split Bregman iteration. We thus end up with a fast implementation of TV–TV^2 inpainting. Finally, we provide a source code for the algorithm written in C and an online demonstration on *Image Processing Online* (IPOL) platform,

accessible on the article web page http://dx.doi.org/10.5201/ipol.2013.40.

1.5.2   Exact solutions of TGV minimisation in one dimension

In Chapter 4, we study the one dimensional L^2–TGV denoising model:

$$\min_{u \in \mathrm{BV}(\Omega)} \frac{1}{2} \int_\Omega (u - f)^2 dx + \mathrm{TGV}^2_{\beta,\alpha}(u), \tag{1.19}$$

where $\Omega = (a, b)$ is an open interval of $\mathbb{R}$. The motivation is to understand deeper how this kind of regularisation behaves and to unravel the role of the parameters α and β. We formulate the predual problem of (1.19) and we derive the corresponding optimality conditions in the spirit of [Rin00, BKV13] using Fenchel–Rockafellar duality. We examine some of the basic properties of the solutions, e.g., behaviour near and away from the boundary, preservation of discontinuities and facts about the L^2–linear regression. We also show that at least for even data functions, TGV and TV regularisations coincide for large enough ratio β/α and we extend this result in dimension two. Moreover we compute exact solutions of (1.19) for three different data functions f: a piecewise constant function with a single jump, a piecewise affine function with a single jump and a hat function, see for instance Figures 1.12 and 1.13. The length of the interval (a, b), the size of the jump and the slope of the hat function appear as parameters in the solutions formulae. Emphasis is given on how the characteristic features of the solutions (discontinuities, piecewise affinity) are affected by the parameters α and β. The analysis is furnished with some numerical experiments that verify our computations.

Note: We should mention here that a paper by Pöschl and Scherzer [PS13], has been prepared at the same time, independently of our work which also addresses exact solutions of L^2–TGV in dimension one. Their focus is on determining how to choose the parameters α and β such that the solutions do not coincide with the respective L^2–TV and L^2–TV^2 regularisations. There, the authors compute exact solutions for a characteristic function of an interval $I \subseteq (a, b)$ and for a hat function, with a fixed slope.

1.5.3   Non-local Hessian

In a pioneering work, Bourgain, Brezis and Mironescu [BBM01] examined functionals of the type

$$\int_\Omega \int_\Omega \frac{|u(x) - u(y)|^p}{|x - y|^p} \rho_n(x - y) dx dy, \tag{1.20}$$

where ρ_n is a sequence of radial functions that concentrates at the origin as n tends to infinity and $1 \le p < \infty$. They characterised the spaces $W^{1,p}(\Omega)$ for $p > 1$ and $\mathrm{BV}(\Omega)$ by
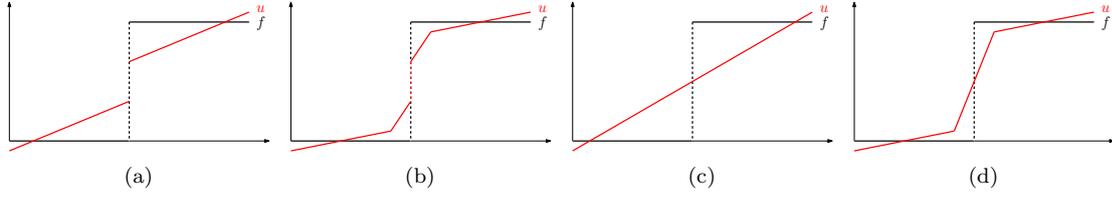
Figure 1.12: All the possible types of solutions (red) of the L^2–TGV denoising problem for a single jump data function.
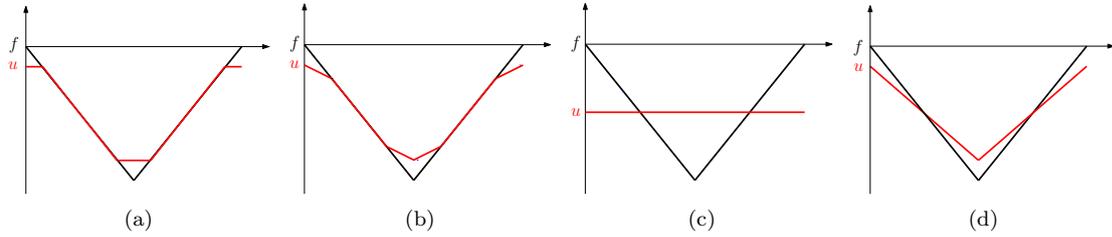
Figure 1.13: All the possible types of solutions (red) of the L^2–TGV denoising problem for a hat function.

showing that

$$u \in W^{1,p}(\Omega) \iff u \in L^p(\Omega) \text{ and } \liminf_{n \to \infty} \int_\Omega \int_\Omega \frac{|u(x) - u(y)|^p}{|x - y|^p} \rho_n(x - y) dx dy < \infty,$$

$$u \in \mathrm{BV}(\Omega) \iff u \in L^1(\Omega) \text{ and } \liminf_{n \to \infty} \int_\Omega \int_\Omega \frac{|u(x) - u(y)|}{|x - y|} \rho_n(x - y) dx dy < \infty.$$

The above characterisations were used by Aubert and Kornprobst [AK09] to solve some non-local variational problems in image denoising. In fact, there exists a fairly large amount of literature regarding non-local models for image reconstruction – we have already mentioned some non-local inpainting methods in Section 1.4. We mention as well the work by Gilboa and Osher [GO07, GO08], Buades, Bartomeu, Morel (non-local means) [BCM05] and also the Ph.D. thesis of Sawatzky [Saw11].

In [GM01], Gobbino and Mora examined the approximation of functionals depending on the gradient of u and on the behaviour of u near the discontinuity points by families of non-local functionals where the gradient is replaced by finite differences.

Ponce [Pon04] derived similar characterisations with [BBM01] by studying functionals of the type

$$\int_\Omega \int_\Omega \omega \left(\frac{|u(x) - u(y)|}{|x - y|} \right) \rho_\epsilon(x - y) dx dy, \tag{1.21}$$

where ω is a continuous function and ρ_ϵ are not necessarily radial.

Mengesha and Spector in [MS13] introduced the following non-local gradient operator:

$$\mathcal{G}_n u(x) = N \int_\Omega \frac{u(x) - u(y)}{|x-y|} \frac{x-y}{|x-y|} \rho_n(x-y) dy, \quad x \in \Omega. \tag{1.22}$$

Functional (1.22) is defined rigorously as a distribution. The authors prove the localisation of the functionals (1.22) to their classical analogue ∇u, in various topologies and they obtain yet another characterisation of the spaces $W^{1,1}(\Omega)$ and $\mathrm{BV}(\Omega)$:

$$u \in W^{1,p}(\Omega) \iff u \in L^p(\Omega) \quad \text{and} \quad \liminf_{n \to \infty} \|\mathcal{G}_n u\|_{L^p(\Omega)} < \infty, \quad \text{for } p > 1,$$

$$u \in \mathrm{BV}(\Omega) \iff u \in L^1(\Omega) \quad \text{and} \quad \liminf_{n \to \infty} \|\mathcal{G}_n u\|_{L^1(\Omega)} < \infty, \quad \text{for } p = 1.$$

In Chapter 5 which constitutes the final part of the thesis, we introduce and study the following higher order non-local functional

$$\mathcal{H}_n u(x) = \frac{d(d+2)}{2} \int_{\mathbb{R}^N} \frac{d^2 u(x,y)}{|x-y|^2} \frac{\left(((x-y)) \otimes (x-y) - \frac{|x-y|^2}{d+2} I_d \right)}{|x-y|^2} \rho_n(x-y) dy, \quad x \in \mathbb{R}^d, \tag{1.23}$$

where $d^2 u(x,y) := u(y) - 2u(x) + u(x + (x-y))$ and the functions ρ_n are radial. We call (1.23) the *explicit* formulation of non-local Hessian. We show that the functional (1.23) localises as n tends to infinity to the continuous analogue $\nabla^2 u$, in the topology that corresponds to the regularity of u. In particular we show that if u is smooth, this convergence is uniform, while if $u \in W^{2,p}(\mathbb{R}^d)$, $1 \le p < \infty$ then we have L^p convergence. Finally, if $u \in \mathrm{BV}^2(\mathbb{R}^d)$ then we show that the sequence of measures $\mathcal{H}_n u \mathcal{L}^d$ converges weakly* to $D^2 u$. Finally after introducing of a second order non-local integration by parts formula we are able to derive non-local characterisations of the spaces $W^{2,p}(\Omega)$ for $1 \le p < \infty$ and $\mathrm{BV}^2(\mathbb{R}^d)$. In particular, we prove the following

$$u \in W^{2,p}(\mathbb{R}^d) \iff u \in L^p(\mathbb{R}^d) \quad \text{and} \quad \liminf_{n \to \infty} \|\mathcal{H}_n u\|_{L^p(\mathbb{R}^d)} < \infty, \quad p > 1,$$

$$u \in \mathrm{BV}^2(\mathbb{R}^d) \iff u \in L^1(\mathbb{R}^d) \quad \text{and} \quad \liminf_{n \to \infty} \|\mathcal{H}_n u\|_{L^1(\mathbb{R}^d)} < \infty, \quad p = 1.$$

We then proceed to an alternative, *implicit* formulation of a non-local Hessian functional which is more suitable for variational problems in imaging. We define the non-local gradient and non-local Hessian of u at a point x with a positive weighting function σ_x, denoted by $\mathcal{G}_{\sigma_x} u(x)$ and $\mathcal{H}_{\sigma_x} u(x)$ respectively, as the minimisers of

$$(\mathcal{H}_{\sigma_x} u(x), \mathcal{G}_{\sigma_x} u(x)) := \operatorname*{argmin}_{\mathcal{H}' \in \mathbb{R}^{d \times d}, \mathcal{G}' \in \mathbb{R}^d} \frac{1}{2} \int_{\Omega - \{x\}} R_{u, \mathcal{G}', \mathcal{H}'}(x,z)^2 \sigma_x(z) \, dz, \tag{1.24}$$

where

$$R_{u, \mathcal{G}', \mathcal{H}'}(x,z) = u(x+z) - u(x) - G'^\top z - \frac{1}{2} z^\top H' z. \tag{1.25}$$

Here $\Omega \subseteq \mathbb{R}^d$ is an open and bounded domain. The weighting function σ_x is determined

by solving a weighted Eikonal equation, thus making it aware of edges in the image u. We provide numerical examples in image denoising where we use the L^1 norm of this implicit form of non-local Hessian as a regulariser, i.e., $\Psi(u) = \alpha \left\| \mathcal{H}_\sigma u \right\|_{L^1(\Omega)}$, for $\alpha > 0$. Our results indicate that this model is able to preserve edges and promote piecewise affine reconstructions, achieving results of better quality than TGV does in that context.

1.6   Organisation of the thesis

In Chapter 2 we provide some mathematical preliminaries mostly on Radon measures, functions of bounded variation, lower semicontinuous envelopes and some results from convex analysis, following mainly [AFP00] [DM93] and [ET76].

In Chapter 3 we discuss the TV–TV^2 approach for image reconstruction following [PS14] and [PSS13]. The well-posedness of the method is shown and we discuss its applications to image denoising, deblurring and inpainting by implementing it with the split Bregman algorithm.

In Chapter 4 we study the one dimensional total generalised variation denoising problem. We examine the basic properties of the model and we compute exact solutions for simple data functions. All the material is contained in [PB13] apart from Section 4.5 in where we state a result on TV–TGV relationship in dimension two that was obtained after the submission of [PB13].

In Chapter 5, we introduce and study the explicit form of non-local Hessian functional following [LPSS]. We prove its localisation to the classical Hessian and we provide some non-local characterisations of $W^{2,p}(\mathbb{R}^d)$ and $\text{BV}^2(\mathbb{R}^d)$ spaces. We also introduce the implicit non-local Hessian formulation and apply it in image denoising.

Some Lemmas that require simple but extensive computations were moved to Appendix A for the sake of the nice presentation of the thesis.

Chapter 2

Mathematical preliminaries

In this chapter we summarise some mathematical tools and notions that are necessary for the rest of the thesis. In particular, after recalling some basic facts about Radon measures, we proceed to a review of the theory of functions of bounded variation which is central for our work. After that, we state some basic facts about lower semicontinuous envelopes or otherwise called relaxed functionals. Finally, we mention some results from convex analysis mainly from duality theory.

We assume that the reader has a basic knowledge of real and functional analysis, measure theory and Sobolev space theory. For these subjects we refer the reader to [DiB02, Bol90, LL01, Eva10].

2.1 Notation on function spaces

We start with a few remarks regarding our notation on functions spaces. Let μ be a positive measure on a metric measure space $(X, \mathcal{B}(X))$, where $\mathcal{B}(X)$ denotes the σ-algebra of X. We denote by $L^p(X, \mathbb{R}^\ell; \mu)$ the space of $\mathbb{R}^\ell$–valued, μ–measurable functions $u : X \to \mathbb{R}^\ell$, such that $\int_X |u|^p d\mu < \infty$, for $1 \le p < \infty$ and the space of μ–essentially bounded functions for $p = \infty$. These spaces are Banach spaces with the norms

$$\|u\|_{L^p(X,\mathbb{R}^\ell;\mu)} = \left(\int_X |u|^p d\mu \right)^{1/p}, \quad 1 \le p < \infty,$$

$$\|u\|_{L^\infty(X,\mathbb{R}^\ell;\mu)} = \operatorname*{ess\,sup}_{x \in X} |u(x)|, \quad p = \infty.$$

When the measure μ is omitted then it is assumed to be the Lebesgue measure $\mathcal{L}^d$ in $\mathbb{R}^d$ ($\mathcal{L}$ in $\mathbb{R}$) and of course in that case X is a subset of $\mathbb{R}^d$. Thus, $L^p(X, \mathbb{R}^\ell)$ is a short version of $L^p(X, \mathbb{R}^\ell; \mathcal{L}^d)$. When the range of the functions, $\mathbb{R}^\ell$, is omitted then it is assumed that $\ell = 1$. Hence, $L^p(X)$ is a short version of $L^p(X, \mathbb{R}; \mathcal{L}^d)$. This notation rule applies in all the function spaces.

The space of $\mathbb{R}^\ell$–valued continuous functions with compact support in X is denoted by $\mathcal{C}_c(X, \mathbb{R}^\ell)$ and it is endowed with the supremum norm $\|u\|_\infty = sup_{x \in X}|u(x)|$. The completion of $\mathcal{C}_c(X, \mathbb{R}^\ell)$ under the supremum norm is denoted by $\mathcal{C}_0(X, \mathbb{R}^\ell)$. Consistently, $\mathcal{C}_c(X)$ and $\mathcal{C}_0(X)$ are short versions of $\mathcal{C}_c(X, \mathbb{R})$ and $\mathcal{C}_0(X, \mathbb{R})$ respectively.

The space of $\mathbb{R}^\ell$–valued, k–times continuously differentiable functions with compact support in Ω, where Ω is an open domain in $\mathbb{R}^d$, is denoted by $\mathcal{C}_c^k(\Omega, \mathbb{R}^\ell)$. When $k = \infty$, we have the space of smooth functions of compact support in Ω. Again, $\mathcal{C}_c^k(\Omega)$ and $\mathcal{C}_c^\infty(\Omega)$ stand for $\mathcal{C}_c^k(\Omega, \mathbb{R})$ and $\mathcal{C}_c^\infty(\Omega, \mathbb{R})$ respectively. Similarly, the corresponding spaces of differentiable functions with not necessarily compact support are denoted by $\mathcal{C}^k(\Omega, \mathbb{R}^\ell)$, $\mathcal{C}^\infty(\Omega, \mathbb{R}^\ell)$, $\mathcal{C}^k(\Omega)$ and $\mathcal{C}^\infty(\Omega)$.

As usual, $W^{k,p}(\Omega)$ denotes the Sobolev space of functions whose distributional derivatives from order zero up to k are L^p functions. The corresponding norm is

$$\|u\|_{W^{k,p}(\Omega)} = \left(\sum_{|a| \leq k} \int_\Omega |D^a u|^p dx \right)^{1/p},$$

where $D^a u$ denotes the a–th distributional derivative of u and $|a| = a_1 + \cdots + a_d$ is the order of the multiindex $a = (a_1, \ldots, a_d)$. Furthermore, $H^k(\Omega)$ denotes the Hilbert space $W^{k,2}(\Omega)$ and $H_0^k(\Omega)$ is the completion of $\mathcal{C}_c^\infty(\Omega)$ under the $\|\cdot\|_{H^k(\Omega)}$ norm. As expected, the corresponding $\mathbb{R}^\ell$–valued Sobolev spaces are denoted by $W^{k,p}(\Omega, \mathbb{R}^\ell)$, $H^k(\Omega, \mathbb{R}^\ell)$ and $H_0^k(\Omega, \mathbb{R}^\ell)$.

We denote with $\mathcal{C}^k(\overline{\Omega})$ the space of k–times differentiable functions, with continuous derivatives up to the boundary of Ω.

Finally, matrix–valued functions are also considered. For instance $\mathcal{C}_c^\infty(X, \mathbb{R}^{\ell \times d})$ denotes the space of $\mathbb{R}^{\ell \times d}$ matrix–valued, smooth functions of compact support in X and similar notations hold for other spaces. Note that the norm used for matrices will always be the Frobenious norm, unless stated otherwise.

2.2 Radon measures

Since standard textbooks are mainly concerned with positive measures, we give here a condensed account of finite Radon measures. For this and the following section on functions of bounded variation, we are following [AFP00] but we also refer the reader to [EG92, Giu84, Rud87].

Definition 2.2.1 (Finite Radon measures)**.** *Let X be a locally compact, separable metric space and let $\mathcal{B}(X)$, denote the Borel σ-algebra of X. We say that $\mu : \mathcal{B}(X) \to \mathbb{R}^\ell$ is an $\mathbb{R}^\ell$–valued finite Radon measure if $\mu(\emptyset) = 0$ and for any sequence $(A_n)_{n \in \mathbb{N}}$ of pairwise*

disjoint elements of $\mathcal{B}(X)$ we have

$$\mu\left(\bigcup_{n=0}^{\infty} A_n\right) = \sum_{n=0}^{\infty} \mu(A_n).$$

The space of $\mathbb{R}^{\ell}$–valued finite Radon measures is denoted by $\mathcal{M}(X, \mathbb{R}^{\ell})$, while in the case $\ell = 1$ (real valued measures), we simply write $\mathcal{M}(X)$.

Note that every $\mu \in \mathcal{M}(X, \mathbb{R}^{\ell})$ can be written as $\mu = (\mu_1, \ldots, \mu_{\ell})$ where $\mu_i \in \mathcal{M}(X)$ for every $i = 1, \ldots, \ell$.

Definition 2.2.2 (Total variation measure). *For a measure $\mu \in \mathcal{M}(X, \mathbb{R}^{\ell})$ we define the total variation measure of $|\mu| : \mathcal{B}(X) \to \mathbb{R}^{+}$, where $\mathbb{R}^{+}$ is the set of positive real numbers, as follows*

$$|\mu|(A) = \sup\left\{\sum_{n=1}^{\infty} |\mu(A_n)| : A_n \in \mathcal{B}(X) \text{ pairwise disjoint}, A = \bigcup_{n=1}^{\infty} A_n\right\}, \quad A \in \mathcal{B}(X).$$

It can be shown that if $\mu \in \mathcal{M}(X, \mathbb{R}^{\ell})$ then $|\mu|$ is a finite positive measure. It is also useful to introduce the notation $\mathcal{M}^{+}(X)$ for the set of finite positive Radon measures on X and $\mathcal{M}^{+}_{loc}(X)$ for the set of positive Radon measures on X, i.e., all the positive measures in $(X, \mathcal{B}(X))$ that are finite on every compact subset of X.

The space $\mathcal{M}(X, \mathbb{R}^{\ell})$ is a Banach space under the norm $\|\mu\|_{\mathcal{M}(X, \mathbb{R}^{\ell})} = |\mu|(X)$. In fact, as the next theorem shows, the space $\mathcal{M}(X, \mathbb{R}^{\ell})$ can be regarded as the dual space of $(\mathcal{C}_0(X, \mathbb{R}^{\ell}), \|\cdot\|_{\infty})$.

Theorem 2.2.3 (Riesz representation theorem). *Let X be a locally compact, separable metric space and let T be a bounded linear functional on $(\mathcal{C}_0(X, \mathbb{R}^{\ell}), \|\cdot\|_{\infty})$. Then there exists a unique $\mu \in \mathcal{M}(X, \mathbb{R}^{\ell})$ such that*

$$T(u) = \sum_{i=1}^{\ell} \int_X u_i \, d\mu_i, \quad \forall u \in \mathcal{C}_0(X, \mathbb{R}^{\ell}),$$

i.e., T can be represented by the measure μ. Moreover, we also have

$$\|T\| = |\mu|(X),$$

where $\|\cdot\|$ denotes the operator norm in the dual space of $\mathcal{C}_0(X, \mathbb{R}^{\ell})$.

From the definition of the operator norm in a dual space and the fact that $\mathcal{C}_c^{\infty}(X, \mathbb{R}^{\ell})$ is dense in $\mathcal{C}_0(X, \mathbb{R}^{\ell})$ under the supremum norm, we have that for every $\mu \in \mathcal{M}(X, \mathbb{R}^{\ell})$

$$\|\mu\|_{\mathcal{M}(\Omega, \mathbb{R}^{\ell})} = \sup\left\{\langle \mu, v \rangle : v \in \mathcal{C}_c^{\infty}(X, \mathbb{R}^{\ell}), \|v\|_{\infty} \leq 1\right\}, \tag{2.1}$$

where here $\langle \mu, v \rangle$ denotes the duality pairing $\langle \mu, v \rangle = \sum_{i=1}^{\ell} \int_X u_i \, d\mu_i$. Thus (2.1) gives an alternative definition for $\|\mu\|_{\mathcal{M}(X,\mathbb{R}^\ell)}$ and provides a nice framework to extend its use in distributions: Given an $\mathbb{R}^\ell$–valued distribution $\mathfrak{T}$ in an open domain $\Omega \subseteq \mathbb{R}^d$ we define the *Radon norm* of $\mathfrak{T}$ to be

$$\|\mathfrak{T}\|_{\mathcal{M}(\Omega,\mathbb{R}^\ell)} = \sup \left\{ \langle \mathfrak{T}, v \rangle : \ v \in \mathcal{C}_c^\infty(\Omega, \mathbb{R}^\ell), \ \|v\|_\infty \leq 1 \right\}. \tag{2.2}$$

In fact, as a consequence of Theorem 2.2.3 we have that $\|\mathfrak{T}\|_{\mathcal{M}(\Omega,\mathbb{R}^\ell)} < \infty$ if and only if $\mathfrak{T}$ is a finite Radon measure μ and in that case $\|\mathfrak{T}\|_{\mathcal{M}(\Omega,\mathbb{R}^\ell)} = |\mu|(\Omega)$. When there is no ambiguity for the domain and the range of $\mathfrak{T}$ we will simply write $\|\mathfrak{T}\|_{\mathcal{M}}$.

We will often consider the Lebesgue decomposition of a finite Radon measure with respect to a positive measure. We first recall the notion of absolute continuity and mutual singularity for measures.

Definition 2.2.4. *Let $(X, \mathcal{B}(X))$ be a measure space.*

(i) *Let $\mu \in \mathcal{M}_{loc}^+(X)$ and $\nu \in \mathcal{M}(X, \mathbb{R}^\ell)$. We say that ν is absolutely continuous with respect to μ and we write $\nu \ll \mu$, if whenever $\mu(A) = 0$ then $|\nu|(A) = 0$.*

(ii) *Let $\mu, \nu \in \mathcal{M}(X, \mathbb{R}^\ell)$. We say that μ and ν are mutually singular and we write $\mu \perp \nu$ if there exists an $A \in \mathcal{B}(X)$ such that $|\mu|(E) = 0$ and $|\mu|(X \setminus E) = 0$.*

We note here that if $\mu \in \mathcal{M}_{loc}^+(X)$ and $f \in L^1(X, \mathbb{R}^\ell; \mu)$ then $f\mu$ denotes the measure in $\mathcal{M}(X, \mathbb{R}^\ell)$ defined by

$$f\mu(A) = \int_A f \, d\mu, \quad A \in \mathcal{B}(X).$$

Theorem 2.2.5 (Lebesgue decomposition)**.** *Let μ be a σ-finite measure in $\mathcal{M}_{loc}^+(X)$ and $\nu \in \mathcal{M}(X, \mathbb{R}^\ell)$. Then there exists a unique pair ν^α, $\nu^s \in \mathcal{M}(X, \mathbb{R}^\ell)$ such that*

$$\nu = \nu^\alpha + \nu^s, \quad \nu^\alpha \ll \mu, \quad \nu^s \perp \mu.$$

Moreover there exists a unique $f \in L^1(X, \mathbb{R}^\ell; \mu)$ such that $\nu^\alpha = f\mu$. The function f is called the Radon–Nikodým density of ν^α with respect to μ and is denoted by $\frac{\nu^\alpha}{\mu}$.

From now on, unless otherwise stated, all the Lebesgue decompositions will be considered with respect to the Lebesgue measure $\mathcal{L}^d$. Thus, notation-wise μ^α and μ^s denote the absolutely continuous and singular part of μ with respect to the Lebesgue measure.

It is very easy to check that a finite Radon measure μ is always absolutely continuous with respect to its total variation measure $|\mu|$ and hence it has a density with respect to it. This leads to the following theorem.

Theorem 2.2.6 (Polar decomposition)**.** *Let $\mu \in \mathcal{M}(X, \mathbb{R}^\ell)$. Then there exists a unique function $\mathrm{sgn}(\mu) \in L^1(X, \mathbb{R}^\ell; |\mu|)$, equal to 1 μ-almost everywhere such that $\mu = \mathrm{sgn}(\mu)|\mu|$, i.e, $\mathrm{sgn}(\mu) = \frac{\mu}{|\mu|}$.*

We finish this section mentioning two important notions of convergence in $\mathcal{M}(X, \mathbb{R}^\ell)$, namely the weak* and the strict convergence.

Definition 2.2.7 (Weak* convergence of measures)**.** *Let* $\mu \in \mathcal{M}(X, \mathbb{R}^\ell)$, $(\mu_n)_{n \in \mathbb{N}} \subseteq \mathcal{M}(X, \mathbb{R}^\ell)$. *We say that* $(\mu_n)_{n \in \mathbb{N}}$ *converges weakly* to μ if for every $u \in \mathcal{C}_0(X)$ we have*

$$\lim_{n \to \infty} \int_X u \, d\mu_n = \int_X u \, d\mu,$$

where the above integrals are computed component-wise, i.e,

$$\int_X u \, d\mu = \left(\int_X u \, d\mu_1, \dots, \int_X u \, d\mu_\ell \right).$$

Observe that the above notion of weak* convergence corresponds exactly to the usual notion of weak* convergence in duals of Banach spaces since from the Riesz representation theorem $\mathcal{M}(X, \mathbb{R}^\ell)$ is the dual space of $\mathcal{C}_0(X, \mathbb{R}^\ell)$.

Definition 2.2.8 (Strict convergence of measures)**.** *Let* $\mu \in \mathcal{M}(X, \mathbb{R}^\ell)$, $(\mu_n)_{n \in \mathbb{N}} \subseteq \mathcal{M}(X, \mathbb{R}^\ell)$. *We say that* $(\mu_n)_{n \in \mathbb{N}}$ *converges strictly to* μ *if*

$$\mu_n \to \mu, \quad \textit{weakly* in measures} \quad \textit{and} \quad \lim_{n \to \infty} \|\mu_n\|_{\mathcal{M}(X, \mathbb{R}^\ell)} = \|\mu\|_{\mathcal{M}(X, \mathbb{R}^\ell)}.$$

Recall that from the general Banach space theory, if $\mu_n \to \mu$ weakly* then

$$\|\mu\|_{\mathcal{M}(X, \mathbb{R}^\ell)} \leq \liminf_{n \to \infty} \|\mu_n\|_{\mathcal{M}(X, \mathbb{R}^\ell)},$$

i.e., "mass" can still "escape" at infinity. Thus, strict convergence is a stronger notion of convergence than weak*.

2.3 Functions of bounded variation

2.3.1 Definitions and properties

The space of functions of bounded variation is a central notion in mathematical imaging as functions that belong to that space are allowed to have jump discontinuities which are interpreted as sharp edges in images. Here, we provide a short summary of the theory of this space, giving emphasis to theorems that we are going to use. As usual Ω denotes an open domain in $\mathbb{R}^d$.

Definition 2.3.1 (Space of functions of bounded variation)**.** *A function* $u \in L^1(\Omega, \mathbb{R}^\ell)$ *is said to be a function of bounded variation or else* $u \in \mathrm{BV}(\Omega, \mathbb{R}^\ell)$ *if its distributional derivative* Du *can be represented by a* $\mathbb{R}^{\ell \times d}$–*valued finite Radon measure (also denoted by*

Du). That is to say, $Du \in \mathcal{M}(\Omega, \mathbb{R}^{\ell \times d})$ and if $u = (u^1, \ldots, u^\ell)$ then

$$\int_\Omega u^a \frac{\partial v}{\partial x_i} dx = -\int_\Omega v\, dD_i u^a, \quad \forall v \in \mathcal{C}_c^\infty(\Omega), \quad i = 1, \ldots, d,\ a = 1, \ldots, \ell.$$

Consistently with our notation, $\mathrm{BV}(\Omega)$ denotes the space $\mathrm{BV}(\Omega, \mathbb{R})$.

According to the Lebesgue decomposition theorem Du can be decomposed to the absolutely continuous and singular part with respect to the Lebesgue measure

$$Du = D^\alpha u + D^s u,$$

where $D^\alpha u$ and $D^s u$ are short versions of $(Du)^\alpha$ and $(Du)^s$ respectively. We will also use the traditional notation ∇u for the absolutely continuous part of Du, that is $\nabla u = D^\alpha u$. When $d = \ell = 1$, ∇u will be also denoted by u'. It is immediate from the Definition 2.3.1 that $W^{1,1}(\Omega, \mathbb{R}^\ell) \subseteq \mathrm{BV}(\Omega, \mathbb{R}^\ell)$ since if $u \in W^{1,1}(\Omega, \mathbb{R}^\ell)$ then $Du = \nabla u \mathcal{L}^d$.

For a function $u \in L^1(\Omega, \mathbb{R}^\ell)$ the total variation of u, $\mathrm{TV}(u)$, is defined as

$$\mathrm{TV}(u) = \sup \left\{ \sum_{a=1}^\ell \int_\Omega u^a \mathrm{div} v^a dx : \ v \in \mathcal{C}_c^1(\Omega, \mathbb{R}^{\ell \times d}),\ \|v\|_\infty \leq 1 \right\}.$$

It can be proven that $\mathrm{TV}(u) < \infty$ if and only if $u \in \mathrm{BV}(\Omega, \mathbb{R}^\ell)$ and in that case $\mathrm{TV}(u) = |Du|(\Omega)$. If $u \in W^{1,1}(\Omega, \mathbb{R}^\ell)$ then $|Du|(\Omega)$ is simply the L^1 norm of ∇u, $\int_\Omega |\nabla u|\, dx$. It can be easily checked that the total variation is lower semicontinuous with respect to the strong L^1 convergence.

The space $\mathrm{BV}(\Omega, \mathbb{R}^\ell)$ endowed with the norm

$$\|u\|_{\mathrm{BV}(\Omega, \mathbb{R}^\ell)} = \|u\|_{L^1(\Omega, \mathbb{R}^\ell)} + |Du|(\Omega),$$

is a Banach space. However the topology induced by that norm is very strong, in fact smooth functions are not dense with respect to that topology. To that scope, two weaker notions of convergence are introduced, namely the weak* and the strict convergence.

Definition 2.3.2 (Weak* convergence in BV). *Let $u \in \mathrm{BV}(\Omega, \mathbb{R}^\ell)$, $(u_n)_{n \in \mathbb{N}} \subseteq \mathrm{BV}(\Omega, \mathbb{R}^\ell)$. We say that the sequence $(u_n)_{n \in \mathbb{N}}$ converges to u weakly* in $\mathrm{BV}(\Omega, \mathbb{R}^\ell)$ if $(u_n)_{n \in \mathbb{N}}$ converges to u in $L^1(\Omega, \mathbb{R}^\ell)$ and $(Du_n)_{n \in \mathbb{N}}$ converges to Du weakly* in $\mathcal{M}(\Omega, \mathbb{R}^{\ell \times d})$, i.e.,*

$$\lim_{n \to \infty} \|u_n - u\|_{L^1(\Omega, \mathbb{R}^\ell)} = 0 \quad and \quad \lim_{n \to \infty} \int_\Omega v\, dDu_n = \int_\Omega v\, dDu, \quad \forall v \in \mathcal{C}_0(\Omega).$$

Definition 2.3.3 (Strict convergence in BV). *Let $u \in \mathrm{BV}(\Omega, \mathbb{R}^\ell)$, $(u_n)_{n \in \mathbb{N}} \subseteq \mathrm{BV}(\Omega, \mathbb{R}^\ell)$. We say that the sequence $(u_n)_{n \in \mathbb{N}}$ converges to u strictly in $\mathrm{BV}(\Omega, \mathbb{R}^\ell)$ if $(u_n)_{n \in \mathbb{N}}$ converges*

to u in $L^1(\Omega, \mathbb{R}^\ell)$ and the total variations $(|Du_n|(\Omega))_{n \in \mathbb{N}}$ converge to $|Du|(\Omega)$, i.e.,

$$\lim_{n \to \infty} \|u_n - u\|_{L^1(\Omega, \mathbb{R}^\ell)} = 0 \quad and \quad \lim_{n \to \infty} |Du_n|(\Omega) = |Du|(\Omega).$$

It can be easily verified that strict convergence is induced by the following metric

$$d(u, v) = \|u - v\|_{L^1(\Omega, \mathbb{R}^\ell)} + \big||Du|(\Omega)| - |Dv|(\Omega)\big|.$$

Moreover it can be also shown that strict convergence implies weak* convergence while the opposite implication is not true in general. The usefulness of the introduction of these weaker notions of convergence can be seen in the following theorem.

Theorem 2.3.4 (Strict approximation by smooth functions in BV)**.** *Let* $u \in L^1(\Omega, \mathbb{R}^\ell)$. *Then* $u \in \mathrm{BV}(\Omega, \mathbb{R}^\ell)$ *if and only if there exists a sequence* $(u_n)_{n \in \mathbb{N}} \in C^\infty(\Omega, \mathbb{R}^\ell) \cap W^{1,1}(\Omega, \mathbb{R}^\ell)$ *such that*

$$\lim_{n \to \infty} \|u_n - u\|_{L^1(\Omega, \mathbb{R}^\ell)} = 0 \quad and \quad \lim_{n \to \infty} \int_\Omega |\nabla u_n| \, dx < \infty$$

In fact, if $u \in \mathrm{BV}(\Omega, \mathbb{R}^\ell)$ *then the sequence* $(u_n)_{n \in \mathbb{N}}$ *can be chosen such that*

$$\lim_{n \to \infty} \int_\Omega |\nabla u_n| \, dx = |Du|(\Omega),$$

i.e., u *can be approximated strictly in* $\mathrm{BV}(\Omega, \mathbb{R}^\ell)$ *by a sequence of smooth functions.*

The following compactness theorem is a useful tool for proving the well-posedness of certain variational problems through the direct method of calculus of variations. Recall that the lack of reflexivity of the Sobolev space $W^{1,1}(\Omega)$ is translated to lack of weak compactness. Thus, as we also see in the next sections, minimisation problems originally posed in $W^{1,1}(\Omega)$ and are often embedded in a suitable way in $\mathrm{BV}(\Omega)$ whose compactness properties guarantee a solution.

Theorem 2.3.5 (Compactness in BV)**.** *Let* Ω *be an open bounded domain in* $\mathbb{R}^d$ *with Lipschitz boundary and let* $(u_n)_{n \in \mathbb{N}}$ *be a sequence in* $\mathrm{BV}(\Omega, \mathbb{R}^\ell)$ *such that*

$$\sup_{n \in \mathbb{N}} \|u_n\|_{\mathrm{BV}(\Omega, \mathbb{R}^\ell)} < \infty.$$

Then there exists a subsequence $(u_{n_k})_{k \in \mathbb{N}}$ *that converges to some* $u \in \mathrm{BV}(\Omega, \mathbb{R}^\ell)$, *weakly* * *in* $\mathrm{BV}(\Omega, \mathbb{R}^\ell)$.

We now state a useful embedding theorem and the version of Poincaré inequality in BV. We define $1^* = d/(d-1)$ when $d > 1$ and $1^* = \infty$ when $d = 1$. Note also that u_Ω

denotes the mean value of u in Ω, i.e.,

$$u_\Omega := \frac{1}{\mathcal{L}^d(\Omega)} \int_\Omega u \, dx. \tag{2.3}$$

Theorem 2.3.6. *(BV embedding theorem / Poincaré inequality) Let Ω be an open bounded domain in $\mathbb{R}^d$ with Lipschitz boundary. Then $\mathrm{BV}(\Omega) \subseteq L^{1^*}(\Omega)$ with continuous embedding. Moreover if Ω is connected, the following Poincaré type inequality holds*

$$\|u - u_\Omega\|_{L^p(\Omega)} \le C|Du|(\Omega), \quad \forall u \in \mathrm{BV}(\Omega), \quad 1 \le p \le 1^*,$$

for some constant C depending only on Ω.

We finish this section mentioning some facts about traces in BV.

Theorem 2.3.7 (Boundary trace theorem in BV)**.** *Let Ω be an open bounded domain in $\mathbb{R}^d$ with Lipschitz boundary and let $u \in \mathrm{BV}(\Omega, \mathbb{R}^\ell)$. Then for $\mathcal{H}^{d-1}$–almost every $x \in \partial\Omega$, there exists $u^\Omega(x) \in \mathbb{R}^\ell$ such that*

$$\lim_{\delta \to 0} \frac{1}{\delta^d} \int_{\Omega \cap B(x,\delta)} |u(y) - u^\Omega(x)| \, dy = 0.$$

Moreover,

$$\|u^\Omega\|_{L^1(\partial\Omega, \mathbb{R}^\ell; \mathcal{H}^{d-1})} \le C\|u\|_{\mathrm{BV}(\Omega, \mathbb{R}^\ell)}, \tag{2.4}$$

for some constant C that depends only on Ω. The function $u^\Omega \in L^1(\partial\Omega, \mathbb{R}^\ell; \mathcal{H}^{d-1})$ is called the trace of u on $\partial\Omega$. Moreover the extension $\overline{u}$ of u to 0 out of Ω belongs to $\mathrm{BV}(\mathbb{R}^d, \mathbb{R}^\ell)$ and viewing Du as a measure on the whole of $\mathbb{R}^d$ and concentrated on Ω, $D\overline{u}$ is given by

$$D\overline{u} = Du + (u^\Omega \otimes \nu_\Omega)\mathcal{H}^{d-1}\lfloor \partial\Omega,$$

where ν_Ω is the outward unit normal vector on $\partial\Omega$ and $\mathcal{H}^{d-1}\lfloor \partial\Omega$ is the $(d-1)$–dimensional Hausdorff measure restricted on $\partial\Omega$.

As a consequence of the boundary trace theorem, we also get the following results:

Theorem 2.3.8 (Integration by parts)**.** *Let Ω be an open bounded domain in $\mathbb{R}^d$ with Lipschitz boundary and let $u \in \mathrm{BV}(\Omega, \mathbb{R}^\ell)$. Then for every $i = 1, \dots, d$, $a = 1, \dots, \ell$ and every function $v \in \mathcal{C}^1(\overline{\Omega})$ we have*

$$\int_\Omega u^a \frac{\partial v}{\partial x_i} dx = -\int_\Omega v \, dD_i u^a + \int_{\partial\Omega} (u^\Omega)^a (\nu_\Omega)_i v \, d\mathcal{H}^{d-1} \tag{2.5}$$

Note that by summing (2.5) over i and by using an appropriate affine function v we

get the following estimate

$$\left| \int_{\Omega} u \, dx \right| \leq K \left(|Du|(\Omega) + \|u^{\Omega}\|_{L^1(\partial\Omega, \mathbb{R}^{\ell}; \mathcal{H}^{d-1})} \right), \quad \forall u \in \mathrm{BV}(\Omega), \tag{2.6}$$

where the constant $K > 0$, depends only on d and Ω. If Ω is connected, we can combine further the above estimate with the Poincaré inequality (2.3) and get

$$\|u\|_{L^2(\Omega)} \leq C \left(|Du|(\Omega) + \|u^{\Omega}\|_{L^1(\partial\Omega, \mathbb{R}^{\ell}; \mathcal{H}^{d-1})} \right), \quad \forall u \in \mathrm{BV}(\Omega), \tag{2.7}$$

where again, the constant $C > 0$, depends only on d and Ω.

Proposition 2.3.9. *Let Ω be an open bounded domain in $\mathbb{R}^d$ with Lipschitz boundary, $u \in \mathrm{BV}(\Omega, \mathbb{R}^{\ell})$ and $v \in \mathrm{BV}(\mathbb{R}^d \setminus \Omega, \mathbb{R}^{\ell})$. Then the function*

$$w(x) = \begin{cases} u(x) & \text{if } x \in \Omega, \\ v(x) & \text{if } x \in \mathbb{R}^d \setminus \Omega, \end{cases}$$

belongs to $\mathrm{BV}(\mathbb{R}^d, \mathbb{R}^{\ell})$ and viewing Du (respectively Dv) as a measure on the whole of $\mathbb{R}^d$ and concentrated on Ω (respectively $\mathbb{R}^d \setminus \Omega$), we have

$$Dw = Du + Dv + \left(u^{\Omega} - v^{\mathbb{R}^d \setminus \Omega} \right) \otimes \nu_{\Omega} \mathcal{H}^{d-1} \lfloor \partial\Omega.$$

2.3.2 Good representatives

For this section, Ω will be a bounded open interval $(a, b) \subseteq \mathbb{R}$. Since every function $u \in L^1(a, b)$ is defined as an equivalence class of functions that are equal almost everywhere, it is not possible to make sense of pointwise values of u. In $\mathrm{BV}(a, b)$ this problem can be overcome through the introduction of *good representatives*. This is a notion that we are going to use often in Chapter 4.

Definition 2.3.10 (Pointwise variation). *For any function $u : (a, b) \to \mathbb{R}$, the pointwise variation $\mathrm{pV}(u, (a, b))$ of u in (a, b) is defined as*

$$\mathrm{pV}(u, (a, b)) = \sup \left\{ \sum_{i=1}^{n-1} |u(x_{i+1}) - u(x_i)| : \ n \geq 2, \ a < x_1 < \cdots < x_n < b \right\}.$$

As one can observe the pointwise variation is extremely sensitive on the choice of the representative of u. This motivates the following definition where the pointwise variation in minimised among elements in the equivalence class of u.

Definition 2.3.11 (Essential variation). *For any function $u : (a, b) \to \mathbb{R}$, the essential*

variation $\mathrm{eV}(u, (a, b))$ *of u in* (a, b) *is defined as*

$$\mathrm{eV}(u, (a, b)) = \inf \left\{ \mathrm{pV}(v, (a, b)) : \ v = u, \ \mathcal{L}\text{--}a.e. \ in \ (a, b) \right\}.$$

It can be proved that for every function $u \in L^1(a, b)$ the infimum in the definition of essential variation is achieved and it is equal to the total variation of u. Thus, if $u \in \mathrm{BV}(a, b)$ then $\mathrm{eV}(u, (a, b)) = |Du|(a, b)$. Any function $\tilde{u}$ in the equivalence class of u with the property $\mathrm{pV}(\tilde{u}, (a, b)) = |Du|(a, b)$ is called *good* or *precise representative* of u. The next theorem shows that these representatives can be characterised in a more explicit way and they have some good continuity and differentiability properties. We define the jump set of u, J_u, to be the set of atoms of Du, i.e., $J_u = \{x \in (a, b) : \ Du(\{x\}) \neq 0\}$.

Theorem 2.3.12 (Good representatives). *Let* $u \in \mathrm{BV}(a, b)$. *Then the following statements hold:*

(i) *There exists a unique* $c \in \mathbb{R}$ *such that*

$$u^l(x) := c + Du((a, x)), \quad u^r(x) := c + Du((a, x]), \quad \forall x \in (a, b),$$

are good representatives of u, the left continuous and the right continuous one. Any other function $\tilde{u} : (a, b) \to \mathbb{R}$ *is a good representative of u if and only if*

$$\tilde{u}(x) \in \left\{ \lambda u^l(x) + (1 - \lambda) u^r(x) : \ \lambda \in [0, 1] \right\}, \quad \forall x \in (a, b).$$

(ii) *Any good representative* $\tilde{u}$ *is continuous in* $(a, b) \setminus J_u$ *and has a jump discontinuity at any point of* J_u:

$$\tilde{u}(x_-) = u^l(x) = u^r(x_-), \quad \tilde{u}(x_+) = u^l(x_+) = u^r(x), \quad \forall x \in J_u,$$

where x_- *and* x_+ *denote left and right limits at x.*

(iii) *Any good representative* $\tilde{u}$ *is differentiable at* $\mathcal{L}$–*a.e. point of* (a, b) *and the derivative* $\tilde{u}'$ *is the density of* Du *with respect to* $\mathcal{L}$.

It is clear from the above that if $u \in \mathrm{BV}(a, b)$ then the functions u^{up} and u_{low} are also good representatives of u where

$$u^{up}(x) := \max\{u^l(x), \ u^r(x)\} \quad \text{and} \quad u_{low}(x) := \min\{u^l(x), \ u^r(x)\}, \quad \forall x \in (a, b).$$

see Figure 2.1 for an illustration.
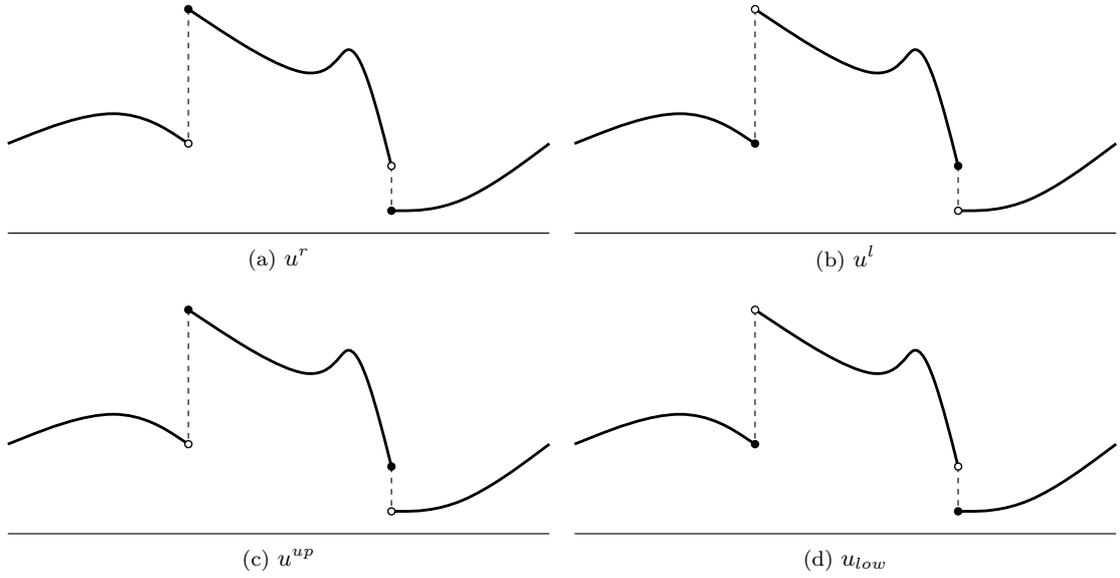
Figure 2.1: Good representatives u^r, u^l, u^{up} and u_{low} of a BV function u.

2.4 Lower semicontinuous envelopes

This section concerns lower semicontinuous envelopes of functionals or otherwise called relaxed functionals. We essentially follow [DM93] but [Bra02] is a good reference as well. We denote by $\overline{\overline{\mathbb{R}}}$ the set of extended real numbers, i.e., $\overline{\overline{\mathbb{R}}} = \mathbb{R} \cup \{+\infty\}$. Let X be a set endowed with some topology $\mathcal{T}$. For every $x \in X$ we denote by $\mathcal{N}(x)$ the set of all the open neighbourhoods of x in X.

Definition 2.4.1 (Lower semicontinuity)**.** *A function* $F : X \to \overline{\overline{\mathbb{R}}}$ *is called* $\mathcal{T}$*–lower semicontinuous at a point* $x \in X$ *if for every* $t \in \mathbb{R}$ *with* $t < F(x)$*, there exists a* $U \in \mathcal{N}(x)$ *such that* $t < F(y)$ *for every* $y \in U$*.*

For first countable topological spaces (hence for metric spaces as well) lower semicontinuity is translated to a more familiar expression:

Proposition 2.4.2. *Suppose that* $(X, \mathcal{T})$ *is a first countable topological space. Then the following are equivalent:*

(i) F *is* $\mathcal{T}$*–lower semicontinuous at* x*.*

(ii) *For every sequence* $(x_n)_{n \in \mathbb{N}}$ *converging to* x*, we have*

$$F(x) \leq \liminf_{n \to \infty} F(x_n).$$

Definition 2.4.3 (Lower semicontinuous envelope)**.** *Let* $F : X \to \overline{\overline{\mathbb{R}}}$*. The lower semicontinuous envelope or otherwise called the relaxed functional of* F *with respect to the topology*

$\mathcal{T}$, $\mathrm{sc}_{\mathcal{T}}F : X \to \overline{\mathbb{R}}$ is defined as

$$\mathrm{sc}_{\mathcal{T}}F(x) = \sup\left\{G(x) : \ G : X \to \overline{\mathbb{R}}, \ \mathcal{T}\text{--lower semicontinuous}, \ G(y) \leq F(y), \ \forall y \in X\right\},$$

for every $x \in \mathbb{R}$.

It easy to check that $\mathrm{sc}_{\mathcal{T}}F$ is the greatest $\mathcal{T}$--lower semicontinuous functional which is smaller or equal than F. It can be also checked that

$$\mathrm{sc}_{\mathcal{T}}F(x) = \sup_{U \in \mathcal{N}(x)} \inf_{y \in U} F(y).$$

Similarly to lower semicontinuity, if X is first countable then lower semicontinuous envelopes can be easily characterised.

Proposition 2.4.4. *Suppose that $(X, \mathcal{T})$ is a first countable topological space. Then $\mathrm{sc}_{\mathcal{T}}F$ is characterised by the following two properties:*

(i) For every sequence $(x_n)_{n\in\mathbb{N}}$ converging to x, we have

$$\mathrm{sc}_{\mathcal{T}}F(x) \leq \liminf_{n\to\infty} F(x_n).$$

(ii) There exists a sequence $(x_n)_{n\in\mathbb{N}}$ converging to x, such that

$$\mathrm{sc}_{\mathcal{T}}F(x) \geq \limsup_{n\to\infty} F(x_n).$$

The following proposition shows the usefulness of lower semicontinuous envelopes regarding the existence of minimum points.

Proposition 2.4.5. *Let $F : X \to \overline{\mathbb{R}}$ and suppose that $\inf_{x\in X} F(x) \in \mathbb{R}$. Then if $\mathrm{sc}_{\mathcal{T}}F$ has minimum point we have that*

$$\min_{x\in X} \mathrm{sc}_{\mathcal{T}}F(x) = \inf_{x\in X} F(x).$$

2.5 Convex analysis

In this section we recall some useful tools from convex analysis. In the first part we introduce the basic concepts while in the second part we provide a short summary of the Fenchel–Rockafellar duality theory. For further study we refer the reader to [ET76] which we also follow here.

2.5.1 Basic concepts

We start with a few definitions and remarks on notation. If $(X, \|\cdot\|_X)$ is a Banach space then X^* denotes its dual space. The elements of X^* are denoted with x^*, $y^*, \dots$ etc.. The

indicator function of a set $A \subseteq X$ is defined as

$$\mathcal{I}_A(x) = \begin{cases} 0 & \text{if } x \in A, \\ \infty & \text{if } x \notin A. \end{cases}$$

For a function $F : X \to \overline{\mathbb{R}}$ we define the domain of F, $\text{dom}(F)$, to be the set of points where F takes finite values, i.e.,

$$\text{dom}(F) = \{x \in X : F(x) < \infty\}.$$

A function F is called proper if $\text{dom}(F) \neq \emptyset$. We now define the convex conjugate F^* of a function F:

Definition 2.5.1 (Convex conjugate)**.** *Let $F : X \to \overline{\mathbb{R}}$. Then the convex conjugate of F, $F^* : X^* \to \overline{\mathbb{R}}$ is defined as*

$$F^*(x^*) = \sup_{x \in X} \langle x^*, x \rangle - F(x), \quad \forall x^* \in X^*.$$

The following, easy to verify result, will be useful to us.

Proposition 2.5.2. *Let $(X, \|\cdot\|_X)$ be a Banach space $1 < p < \infty$ and $F : X \to \mathbb{R}$ with*

$$F(x) = \frac{1}{p}\|x\|_X^p, \quad \forall x \in X.$$

Then the convex conjugate of F is

$$F^*(x^*) = \frac{1}{q}\|x^*\|_{X^*}^q, \quad x^* \in X^*.$$

where $1/p + 1/q = 1$.

We finally recall the definition of the subdifferential of a function.

Definition 2.5.3. *Let $(X, \|\cdot\|_X)$ be a Banach space and $F : X \to \mathbb{R}$. We say that $x^* \in X^*$ belongs to the subdifferential of F at $x \in X$ and we write $x^* \in \partial F(x)$ if*

$$F(x) < \infty \quad \text{and} \quad \langle x^*, y - x \rangle + F(x) \leq F(y), \quad \forall y \in X.$$

If $\partial F(x) \neq \emptyset$ we say that F is subdifferentiable in x.

The subdifferential of F at a given point can be also characterised as follows:

Proposition 2.5.4. *Let $F : X \to \overline{\mathbb{R}}$. Then $x^* \in \partial F(x)$ if and only if*

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

2.5.2 Fenchel–Rockafellar duality

Fenchel–Rockafellar duality theory provides a convenient framework for characterising solutions of variational problems. In this section, we present a basic review for variational problems that have a particular form, i.e.,

$$\inf_{x \in X} F_1(x) + F_2(\Lambda(x)), \tag{2.8}$$

where X, Y are Banach spaces, $\Lambda : X \to Y$ is a bounded linear operator and $F_1 : X \to \overline{\mathbb{R}}$, $F_2 : Y \to \overline{\mathbb{R}}$ are proper, convex, lower semicontinuous functions. The minimisation problem (2.8) is called the primal problem and it is denoted by $\mathcal{P}_{primal}$. The dual problem of (2.8) is denoted by $\mathcal{P}_{dual}$ and is defined as

$$\sup_{y^* \in Y^*} -F_1^*(\Lambda^*(y^*)) - F_2^*(-y^*), \tag{2.9}$$

where $\Lambda^* : Y^* \to X^*$ is the adjoint operator of Λ. We denote with $\inf \mathcal{P}_{primal}$ and $\sup \mathcal{P}_{dual}$ the infimum of (2.8) and the supremum of (2.9) respectively. It can be shown that it always holds $\sup \mathcal{P}_{dual} \leq \inf \mathcal{P}_{primal}$. If $\sup \mathcal{P}_{dual} = \inf \mathcal{P}_{primal}$ we say that no duality gap occurs. The following theorem shows how the solutions of $\mathcal{P}_{primal}$ and $\mathcal{P}_{dual}$ are related when no duality gap occurs.

Theorem 2.5.5 (Optimality conditions)**.** *Suppose that both problems $\mathcal{P}_{primal}$ and $\mathcal{P}_{dual}$ have solutions and that*

$$\sup \mathcal{P}_{dual} = \inf \mathcal{P}_{primal},$$

with that number being finite. Then all the solutions x of $\mathcal{P}_{primal}$ and y^* of $\mathcal{P}_{dual}$ are related through the following optimality conditions:

$$\Lambda^*(y^*) \quad \in \quad \partial F_1(x), \tag{2.10}$$

$$-y^* \quad \in \quad \partial F_2(\Lambda(x)). \tag{2.11}$$

Conversely, if $x \in X$, $y^* \in Y^*$ satisfy (2.10)–(2.11) then x is a solution of $\mathcal{P}_{primal}$, y^* is a solution of $\mathcal{P}_{dual}$ and $\sup \mathcal{P}_{dual} = \inf \mathcal{P}_{primal}$ with that number being finite.

We note here that using Proposition 2.5.4 and having in mind that $F^{**} = F$ for convex lower semicontinuous functions, we can easily check that the optimality conditions (2.10)-(2.11) are equivalent to

$$x \quad \in \quad \partial F_1^*(\Lambda^*(y^*)), \tag{2.12}$$

$$\Lambda(x) \quad \in \quad \partial F_2^*(-y^*). \tag{2.13}$$

There have been quite a few results regarding necessary conditions for the absence of

duality gap. We mention here a condition due to Attouch and Brezis [AB86], which is useful for our purposes.

Theorem 2.5.6 (Attouch–Brezis). *Suppose that X, Y are Banach spaces, $\Lambda : X \to Y$ is a bounded linear operator and $F_1 : X \to \overline{\mathbb{R}}$, $F_2 : Y \to \overline{\mathbb{R}}$ are proper, convex, lower semicontinuous functionals. If the set*

$$\bigcup_{\lambda \geq 0} \lambda(\mathrm{dom}(F_2) - \Lambda(\mathrm{dom}(F_1))),$$

is a closed subspace of Y, then

$$\inf_{x \in X} F_1(x) + F_2(\Lambda(x)) = \max_{y^* \in Y^*} -F_1^*(\Lambda^*(y^*)) - F_2^*(-y^*),$$

i.e., the dual problem $\mathcal{P}_{dual}$ has a solution and no duality gap occurs.

Chapter 3

The combined TV–TV^2 approach for image reconstruction

3.1 Introduction

As we have already discussed in the introduction, one way to reduce the staircasing effect in total variation-based image reconstruction methods, is the incorporation of higher order derivatives in the regularisation process. Infimal convolution type of methods have been proven to be very successful towards this task with total generalised variation being the current state of the art. However, one of the disadvantages of higher order regularisation methods is the increased computational cost involved in their implementation. TGV regularisation is no exception even though sophisticated algorithms can be used for its solution. Currently, TGV minimisation problems are commonly solved with the primal dual hybrid method of Chambolle–Pock [CP11], see [Bre14]. We will show that the method we propose here can be solved even faster with the split Bregman algorithm [GO09] while still reducing the staircasing effect in image denoising and deblurring. Moreover, unlike TGV, our method can be also used for image inpainting, having the advantage of being able to interpolate images along large gaps in the inpainting domain, which is an improvement to first order total variation inpainting.

In summary the model we are studying here is the following:

$$\min_{u \in \mathrm{BV}^2(\Omega)} \frac{1}{s} \int_\Omega |Tu - f|^s dx + \alpha \int_\Omega \varphi_1(\nabla u)\, dx + \beta\, \varphi_2(D^2 u)(\Omega), \quad s = 1, 2, \qquad (3.1)$$

where α, β are positive parameters, φ_1, φ_2 are two convex functions $\varphi_1 : \mathbb{R}^2 \to \mathbb{R}^+$, $\varphi_2 : \mathbb{R}^4 \to \mathbb{R}^+$ with at most linear growth at infinity and T is suitable bounded linear operator. The minimisation (3.1) is done over $\mathrm{BV}^2(\Omega)$, the space of functions of bounded Hessian, see Section 3.3 for details. For the time being, we note that for a function $u \in \mathrm{BV}^2(\Omega)$, the second order distributional derivative $D^2 u$, is a finite Radon measure.

This means that problem (3.1) is posed in the framework of convex functions of measures. The motivation for that is to make sense of expressions like $\int_\Omega \sqrt{|\nabla^2 u|^2 + \epsilon}$, where $\epsilon \ll 1$ and $\nabla^2 u$ is not necessarily an L^1 function but it can also be a measure. This concept has already been considered in the first order case, e.g., in the study of the following "smoothed" total variation denoising problem [Ves01]

$$\min_{u \in \mathrm{BV}(\Omega)} \frac{1}{2} \int_\Omega (u - f)^2 dx + \int_\Omega \sqrt{|\nabla u|^2 + \epsilon}\, dx, \tag{3.2}$$

where the term $\int_\Omega \sqrt{|\nabla u|^2 + \epsilon}\, dx$ has to be defined appropriately in order to make sense for measures. Such "rounding offs" of non-smooth regularisers like the total variation become necessary for the numerical implementation by means of time-stepping [Ves01], multigrid methods [FSHW04, Vog95, VW97] or semi-smooth Newton methods [HS06b] for instance.

Well-posedness of (3.1) is proved by identifying the minimising functional as a lower semicontinuous envelope of another naturally defined functional. The underlying topology under which the lower semicontinuous envelope is considered is a suitably defined topology in $\mathrm{BV}^2(\Omega)$ which is analogous to the strict topology in $\mathrm{BV}(\Omega)$. Uniqueness and stability results are also shown. We proceed to the numerical implementation of our method using the split Bregman algorithm. To do so, we leave the general framework of convex functions of measures by choosing $\varphi_1 = |\cdot|$ in $\mathbb{R}^2$ and $\varphi_2 = |\cdot|$ in $\mathbb{R}^4$. As a result, the minimisation we treat numerically is

$$\min_{u \in \mathrm{BV}^2(\Omega)} \frac{1}{s} \int_\Omega |Tu - f|^s dx + \alpha \, \mathrm{TV}(u) + \beta \, \mathrm{TV}^2(u), \quad s = 1, 2, \tag{3.3}$$

where $\mathrm{TV}^2(u)$ denotes the second order total variation of u, $|D^2 u|(\Omega)$. Note that the case $\beta = 0$, is the classical ROF model while the case $\alpha = 0$, i.e., pure second order regularisation, has been studied in [BP10]. We show that by choosing $\alpha \neq 0$, $\beta \neq 0$ one can achieve better results (as these are assessed by image quality measures) by both these special cases. The idea in our approach is to weight TV with a fairly large α and in the same time keeping β small enough such that no serious blur is introduced in the image while staircasing is reduced. Even though the introduction of TV^2 does not allow the preservation of edges like TGV, by keeping the value of β small, the quality of the results is still comparable with the TGV ones. Moreover, since our method do not decompose the image á la infimal convolution, it can solved very efficiently with the split Bregman algorithm. Thus, we end up with a very fast method which is ideal as a pre-processing step and provides a solid basis for some simple and fast further improvements, e.g., contrast adjustment and sharpening.

Let us stress finally one more time that the image model we are based on is $f = Tu + \eta$, where T is a bounded, linear operator and η is random noise. This model is well suited for a

variety of image reconstruction tasks e.g. denoising, deconvolution, inpainting, zooming, to name a few. Even though this model and its corresponding variational problem (3.1) is not suitable for some advanced medical imaging tasks like MRI and tomography, it introduces a new regulariser, namely the weighted sum of first and second order total variation which can be exploited in future research focusing on these tasks. In that case extension of our results for complex valued images will be necessary as for instance the images that one typically obtains from an MRI machine are of that kind, as a result of the use of Fourier transform. However, as the present thesis deals with more "traditional" image processing tasks we prefer to focus on the real valued case. We only use the fast Fourier transform in Section 3.5.3 as a tool for the efficient solution of a certain minimisation problem but the data obtained in the end are real valued.

3.1.1 Organisation of the chapter

In Section 3.2 we state some preliminary facts concerning convex functions of measures. The definition of a convex function of a measure is given, along with a lower semicontinuity result due to Buttazzo and Freddi.

In Section 3.3 we introduce the space of functions of bounded Hessian $\mathrm{BV}^2(\Omega)$. More specifically, the definition of $\mathrm{BV}^2(\Omega)$ together with its basic properties is stated in Section 3.3.1. In Section 3.3.2 we introduce two naturally defined topologies on $\mathrm{BV}^2(\Omega)$, namely the weak* and the strict topology.

The well-posedness of the TV–TV^2 model is shown in Section 3.4. There, the minimising functional in (3.1) is identified as the lower semicontinuous envelope of a naturally defined functional (defined in the context of convex functions instead of measures) with respect to the aforementioned topologies of $\mathrm{BV}^2(\Omega)$. Existence, uniqueness as well as stability results are shown.

Section 3.5 discusses the numerical realisation of the model using the split Bregman iteration. In particular, Section 3.5.1 is concerned with the discretisation of the model. The corresponding finite dimensional norms along with the discrete differential operators are introduced. In Section 3.5.2 we recall the Bregman iteration and its use for solving constrained optimisation problems and we contribute to its convergence theory. In Section 3.5.3, we describe how the Bregman iteration can be adapted to our model and we propose a splitting strategy for it (split Bregman) as a numerical method for the solution of our problem.

Applications in denoising and deblurring of the TV–TV^2 approach are presented in Section 3.6. We present numerical experiments for images that have been corrupted by Gaussian noise, comparing the TV–TV^2 method with TV, TV^2, TGV and infimal convolution denoising and we do the same for image deblurring.

The application of the TV–TV^2 method in image inpainting is presented in detail in Section 3.7. In Section 3.7.1, we provide a motivation for the use of our method for

inpainting tasks (large gap connectivity) and we present the concept and philosophy of the online journal *Image Processing Online* in whose platform we provide an online, user friendly, demonstration of TV–TV^2 inpainting. The numerical realisation using the split Bregman algorithm is described in Section 3.7.2, while the extension of the algorithm to colour images is done in Section 3.7.3. Stopping criteria and optimal selection of the split Bregman parameters are proposed and numerically justified in Section 3.7.4. Finally, in Section 3.7.5 we provide some inpainting examples for both real life and synthetic images. We stress the visual improvement of the TV–TV^2 method over pure TV inpainting and we discuss experimentally the role of the parameters α, β as well as the role of the geometry of the inpainting domain, as far as connectivity along large gaps is concerned.

3.2 Convex functions of measures

In this section we introduce the notion of a convex function of a measure, for functions with at most linear growth, following essentially [DT84]. Let $\varphi : \mathbb{R}^\ell \to \mathbb{R}$ be a continuous function, positively one homogeneous, i.e., for every $x \in \mathbb{R}^\ell$

$$\varphi(tx) = t\varphi(x), \quad \forall t \geq 0.$$

Given a measure in $\mu \in \mathcal{M}(\Omega, \mathbb{R}^\ell)$, where Ω is an open subset of $\mathbb{R}^d$, we define the measure $\varphi(\mu) \in \mathcal{M}(\Omega)$ as follows:

$$\varphi(\mu) := \varphi\left(\frac{\mu}{|\mu|}\right) |\mu|. \tag{3.4}$$

Notice that $\varphi(\mu)$ is well defined as $\mu/|\mu| = 1$, $|\mu|$–a.e., and φ is bounded on $\mathbb{S}^{\ell-1} := \{x \in \mathbb{R}^\ell : |x| = 1\}$. Moreover the following proposition, that was mentioned in [DT84] and we give here a proof, holds:

Proposition 3.2.1. *Suppose that $\varphi : \mathbb{R}^\ell \to \mathbb{R}$ is a continuous function, positively one homogeneous and $\mu \in \mathcal{M}(\Omega, \mathbb{R}^\ell)$. Then for every measure $\nu \in \mathcal{M}^+(\Omega)$ such that μ is absolutely continuous with respect to ν, we have*

$$\varphi(\mu) = \varphi\left(\frac{\mu}{|\nu|}\right) |\nu|.$$

Moreover, if φ is a convex function from $\mathbb{R}^\ell$ to $\mathbb{R}$, then $\varphi : \mathcal{M}(\Omega, \mathbb{R}^\ell) \to \mathcal{M}(\Omega)$ is a convex function is the space of measures as well.

Proof. Since $\mu \ll \nu$, we also have that $|\mu| \ll \nu$. Using the fact that φ is positively one homogeneous and the fact that $|\mu|/\nu$ is a positive function ν–a.e., we get

$$\varphi(\mu) = \varphi\left(\frac{\mu}{|\mu|}\right) |\mu| = \varphi\left(\frac{\mu}{|\mu|}\right) \frac{|\mu|}{\nu}\nu = \varphi\left(\frac{\mu}{|\mu|}\frac{|\mu|}{\nu}\right)\nu = \varphi\left(\frac{\mu}{\nu}\right)\nu.$$

Assuming now that φ is convex and using the first part of the proposition we get for $0 \leq \lambda \leq 1$ and $\mu, \nu \in \mathcal{M}(\Omega, \mathbb{R}^\ell)$:

$$
\begin{aligned}
\varphi(\lambda\mu + (1-\lambda)\nu) &= \varphi\left(\frac{\lambda\mu + (1-\lambda)\nu}{|\lambda\mu + (1-\lambda)\nu|}\right)|\lambda\mu + (1-\lambda)\nu| \\
&= \varphi\left(\frac{\lambda\mu + (1-\lambda)\nu}{|\mu| + |\nu|}\right)(|\mu| + |\nu|) \\
&= \varphi\left(\lambda\frac{\mu}{|\mu| + |\nu|} + (1-\lambda)\frac{\nu}{|\mu| + |\nu|}\right)(|\mu| + |\nu|) \\
&\leq \lambda\varphi\left(\frac{\mu}{|\mu| + |\nu|}\right)(|\mu| + |\nu|) + (1-\lambda)\varphi\left(\frac{\nu}{|\mu| + |\nu|}\right)(|\mu| + |\nu|) \\
&= \lambda\varphi\left(\frac{\mu}{|\mu|}\right)|\mu| + (1-\lambda)\varphi\left(\frac{\nu}{|\nu|}\right)|\nu| \\
&= \lambda\varphi(\mu) + (1-\lambda)\varphi(\nu).
\end{aligned}
$$

$\square$

Suppose now that φ is not necessarily positively one homogeneous but a convex function $\varphi : \mathbb{R}^\ell \to \mathbb{R}$ which has at most linear growth at infinity, i.e., there exists a positive constant K such that

$$
\varphi(x) \leq K(1 + |x|), \quad \forall x \in \mathbb{R}^\ell.
$$

In that case the *recession function* φ_∞ of φ is well defined everywhere, where

$$
\varphi_\infty(x) := \lim_{t \to \infty} \frac{\varphi(tx)}{t}, \quad \forall x \in \mathbb{R}^\ell.
$$

It can be proved in this case, see for instance [AFP00], that φ_∞ is a convex, positively one homogeneous function. We are now ready to define convex functions of measures for functions of at most linear growth.

Definition 3.2.2 (Convex function of a measure). *Let $\varphi : \mathbb{R}^\ell \to \mathbb{R}$ be a convex function of at most linear growth at infinity and let $\mu \in \mathcal{M}(\Omega, \mathbb{R}^\ell)$. Moreover, consider the Lebesgue decomposition of μ with respect to the Lebesgue measure $\mathcal{L}^d$ in Ω:*

$$
\mu = \left(\frac{\mu}{\mathcal{L}^d}\right)\mathcal{L}^d + \mu^s.
$$

Then we define the measure $\varphi(\mu) \in \mathcal{M}(\Omega)$ as follows:

$$
\varphi(\mu) = \varphi\left(\frac{\mu}{\mathcal{L}^d}\right)\mathcal{L}^d + \varphi_\infty\left(\frac{\mu^s}{|\mu^s|}\right)|\mu^s|. \tag{3.5}
$$

It can be proved in a similar way with Proposition 3.2.1 that $\varphi : \mathcal{M}(\Omega, \mathbb{R}^\ell) \to \mathcal{M}(\Omega, \mathbb{R})$ is

also a convex function. The following theorem, proved in [BF91] and can also be found in [AFP00] establishes the lower semicontinuity of φ with respect to the weak* convergence in $\mathcal{M}(\Omega, \mathbb{R}^\ell)$.

Theorem 3.2.3 (Buttazzo–Freddi, 1991). *Let Ω be an open subset of $\mathbb{R}^d$, $\nu \in \mathcal{M}(\Omega, \mathbb{R}^\ell)$, $(\nu_n)_{n \in \mathbb{N}} \subseteq \mathcal{M}(\Omega, \mathbb{R}^\ell)$ and let $\mu \in \mathcal{M}^+(\Omega)$, $(\mu_n)_{n \in \mathbb{N}} \subseteq \mathcal{M}^+(\Omega)$. Let $\varphi : \mathbb{R}^\ell \to \mathbb{R}$ be a convex function with at most linear growth at infinity and suppose that $\nu_n \to \nu$ and $\mu_n \to \mu$ weakly* in measures. If $\nu = (\nu/\mu)\mu + \nu^s$ and $\nu_n = (\nu_n/\mu_n)\mu_n + \nu_n^s$ are the Lebesgue decompositions of ν and ν_n with respect to μ and μ_n respectively, then*

$$\int_\Omega \varphi\left(\frac{\nu}{\mu}\right) d\mu + \int_\Omega \varphi_\infty\left(\frac{\nu^s}{|\nu^s|}\right) d|\nu^s| \le \liminf_{n \to \infty} \int_\Omega \varphi\left(\frac{\nu_n}{\mu_n}\right) d\mu_n + \int_\Omega \varphi_\infty\left(\frac{\nu_n^s}{|\nu_n^s|}\right) d|\nu_n^s|.$$

In particular, if $\mu = \mu_n = \mathcal{L}^d$ for all $n \in \mathbb{N}$ then the above inequality can be written as

$$\varphi(\nu)(\Omega) \le \liminf_{n \to \infty} \varphi(\nu_n)(\Omega).$$

3.3 The space of functions of bounded Hessian BV$^2(\Omega)$

3.3.1 Definition and basic properties

In this section we recall the definition and the basic properties of the space of functions of bounded Hessian. It was firstly introduced and studied by Demengel in [Dem85].

Definition 3.3.1 (Space of functions of bounded Hessian). *Let $\Omega \subseteq \mathbb{R}^d$ be open. A function $u \in W^{1,1}(\Omega)$ is said to be a function of bounded Hessian or else $u \in \mathrm{BV}^2(\Omega)$ if its second order distributional derivative D^2u can be represented by a $\mathbb{R}^{d \times d}$–valued finite Radon measure which is still denoted by D^2u. In other words*

$$\mathrm{BV}^2(\Omega) = \left\{ u \in W^{1,1}(\Omega) : \ \nabla u \in \mathrm{BV}(\Omega, \mathbb{R}^d) \right\}.$$

It can proved, in a similar way with the BV case that a function $u \in L^1(\Omega)$ belongs to $\mathrm{BV}^2(\Omega)$ if and only if its second order total variation $\mathrm{TV}^2(u)$ is finite, where

$$\mathrm{TV}^2(u) = \sup\left\{ \int_\Omega u \operatorname{div}^2 v \, dx : \ v \in \mathcal{C}_c^2(\Omega, \mathbb{R}^{d \times d}), \ \|v\|_\infty \le 1 \right\},$$

and in that case we have $\mathrm{TV}^2(u) = |D^2u|(\Omega)$. The space $\mathrm{BV}^2(\Omega)$ is equipped with the norm

$$\|u\|_{\mathrm{BV}^2(\Omega)} := \|u\|_{\mathrm{BV}(\Omega)} + |D^2u|(\Omega),$$

which makes it a Banach space. If Ω is connected and has a Lipschitz boundary then it can be shown [Dem85] that there exist positive constants C_1, C_2 that depend only on Ω

such that

$$\|\nabla u\|_{L^1(\Omega,\mathbb{R}^d)} \le C_1 |D^2 u|(\Omega) + C_2 \|u\|_{L^1(\Omega)}, \quad \forall u \in \mathrm{BV}^2(\Omega). \tag{3.6}$$

It is obvious that in that case the norm $\|u\| = \|u\|_{L^1(\Omega)} + |D^2 u|(\Omega)$ is equivalent with $\|\cdot\|_{\mathrm{BV}^2(\Omega)}$.

It is immediate that $W^{2,1}(\Omega) \subseteq \mathrm{BV}^2(\Omega)$ since if $u \in W^{2,1}(\Omega)$ then $D^2 u = \nabla^2 u \mathcal{L}^d$, i.e., the second order distributional derivative is a function in $L^1(\Omega, \mathbb{R}^{d\times d})$. Generally, for a function $u \in \mathrm{BV}^2(\Omega)$, we denote by $\nabla^2 u$, the absolutely continuous part of $D^2 u$ with respect to $\mathcal{L}^d$. Finally, it can also be proved, see also [Dem85] that the embedding of $\mathrm{BV}^2(\Omega)$ into $W^{1,1}(\Omega)$ is compact.

3.3.2 Weak* and strict convergence in $\mathrm{BV}^2(\Omega)$

Similarly with $\mathrm{BV}(\Omega)$ we can define some weaker notions of convergence in $\mathrm{BV}^2(\Omega)$ namely the weak* and the strict convergence. We also prove a weak* compactness theorem in $\mathrm{BV}^2(\Omega)$ which is analogous to Theorem 2.3.5.

Definition 3.3.2 (Weak* convergence in BV^2)**.** *Let $u \in \mathrm{BV}^2(\Omega)$, $(u_n)_{n\in\mathbb{N}} \subseteq \mathrm{BV}^2(\Omega)$. We say that the sequence $(u_n)_{n\in\mathbb{N}}$ converges to u weakly* in $\mathrm{BV}^2(\Omega)$ if $(u_n)_{n\in\mathbb{N}}$ converges to u in $L^1(\Omega, \mathbb{R}^d)$ and $(\nabla u_n)_{n\in\mathbb{N}}$ converges to ∇u weakly* in $\mathrm{BV}(\Omega, \mathbb{R}^d)$, i.e.,*

$$\lim_{n\to\infty} \|u_n - u\|_{L^1(\Omega)} = 0, \quad \lim_{n\to\infty} \|\nabla u_n - \nabla u\|_{L^1(\Omega,\mathbb{R}^d)} = 0,$$

$$\text{and} \quad \lim_{n\to\infty} \int_\Omega v \, dD^2 u_n = \int_\Omega v \, dD^2 u, \quad \forall v \in \mathcal{C}_0(\Omega).$$

It is not hard to check that weak* convergence is induced by a topology which has a basis consisting of the following sets:

$$U(u_0, F, \epsilon) = \left\{ u \in \mathrm{BV}^2(\Omega) : \; \|u_0 - u\|_{L^1(\Omega)} + \|\nabla u_0 - \nabla u\|_{L^1(\Omega,\mathbb{R}^d)} \right.$$
$$\left. + \left| \int_\Omega v_i \, dD^2 u_0 - \int_\Omega v_i \, dD^2 u \right| < \epsilon, \; i \in F \right\},$$

where $u_0 \in \mathrm{BV}^2(\Omega)$, $F \subseteq \mathbb{N}$ finite, $\epsilon > 0$ and $v_i \in \mathcal{C}_0(\Omega)$ for every $i \in F$. Using now Theorem 2.3.5 we can easily prove the following compactness result:

Theorem 3.3.3 (Compactness in BV^2)**.** *Let Ω be an open bounded domain in $\mathbb{R}^d$ with Lipschitz boundary and let $(u_n)_{n\in\mathbb{N}}$ be a sequence in $\mathrm{BV}^2(\Omega)$ such that*

$$\sup_{n\in\mathbb{N}} \|u_n\|_{\mathrm{BV}^2(\Omega)} < \infty.$$

Then there exists a subsequence $(u_{n_k})_{k\in\mathbb{N}}$ that converges to some u, weakly* in $\mathrm{BV}^2(\Omega)$.

Proof. From the compact embedding of $\mathrm{BV}^2(\Omega)$ into $W^{1,1}(\Omega)$ and the fact that the sequence $(\nabla u_n)_{n\in\mathbb{N}}$ is bounded in $\mathrm{BV}(\Omega,\mathbb{R}^d)$ we have that there exists a subsequence $(u_{n_k})_{k\in\mathbb{N}}$, a function $u \in W^{1,1}(\Omega)$ and a function $v \in \mathrm{BV}(\Omega,\mathbb{R}^d)$ such that $(u_{n_k})_{k\in\mathbb{N}}$ converges to u strongly in $W^{1,1}(\Omega)$ and $(\nabla u_{n_k})_{k\in\mathbb{N}}$ converges to v weakly* in $\mathrm{BV}(\Omega,\mathbb{R}^d)$. Then, it is obvious that $\nabla u = v$ a.e., $u \in \mathrm{BV}^2(\Omega)$ and $(u_{n_k})_{k\in\mathbb{N}}$ converges to u weakly* in $\mathrm{BV}^2(\Omega)$. $\qquad\square$

Analogously with $\mathrm{BV}(\Omega)$ we define the strict convergence in $\mathrm{BV}^2(\Omega)$.

Definition 3.3.4 (Strict convergence in BV^2). *Let $u \in \mathrm{BV}^2(\Omega)$, $(u_n)_{n\in\mathbb{N}} \subseteq \mathrm{BV}^2(\Omega)$. We say that the sequence $(u_n)_{n\in\mathbb{N}}$ converges to u strictly in $\mathrm{BV}^2(\Omega)$ if $(u_n)_{n\in\mathbb{N}}$ converges to u in $L^1(\Omega)$ and the sequence $(|D^2u_n|(\Omega))_{n\in\mathbb{N}}$ converges to $|D^2u|(\Omega)$, i.e.,*

$$\lim_{n\to\infty} \|u_n - u\|_{L^1(\Omega} = 0 \quad \text{and} \quad \lim_{n\to\infty} |D^2u_n|(\Omega) = |D^2u|(\Omega).$$

It can be easily verified that the strict convergence in $\mathrm{BV}^2(\Omega)$ is induced by the following metric:

$$d(u,v) = \|u - v\|_{L^1(\Omega)} + \big||D^2u|(\Omega)| - |D^2v|(\Omega)\big|, \quad u,v \in \mathrm{BV}^2(\Omega).$$

The following lemma can be used to compare the weak* and the strict convergence in $\mathrm{BV}^2(\Omega)$.

Lemma 3.3.5. *Suppose that $\Omega \subseteq \mathbb{R}^d$ is an open, bounded, connected set with Lipschitz boundary. Let $u \in \mathrm{BV}^2(\Omega)$, $(u_n)_{n\in\mathbb{N}} \subseteq \mathrm{BV}^2(\Omega)$ and suppose that $(u_n)_{n\in\mathbb{N}}$ converges to u strictly in $\mathrm{BV}^2(\Omega)$. Then $(u_n)_{n\in\mathbb{N}}$ converges to u strongly in $W^{1,1}(\Omega)$, i.e.,*

$$\lim_{n\to\infty} \|u_n - u\|_{W^{1,1}(\Omega)} = 0.$$

Proof. Since the sequence $(u_n)_{n\in\mathbb{N}}$ is strictly convergent in $\mathrm{BV}^2(\Omega)$, we have that the sequences $(\|u_n\|_{L^1(\Omega)})_{n\in\mathbb{N}}$ and $(|D^2u_n|(\Omega))_{n\in\mathbb{N}}$ are bounded. Using the estimate (3.6) we deduce that the sequence $(\|\nabla u_n\|_{L^1(\Omega,\mathbb{R}^d)})_{n\in\mathbb{N}}$ is bounded as well, which implies that the sequence $(u_n)_{n\in\mathbb{N}}$ is bounded in $\mathrm{BV}^2(\Omega)$. From the compact embedding of $\mathrm{BV}^2(\Omega)$ into $W^{1,1}(\Omega)$, we get that there exists a subsequence $(u_{n_k})_{k\in\mathbb{N}}$ and a function $v \in W^{1,1}(\Omega)$ such that $(u_{n_k})_{k\in\mathbb{N}}$ converges to v strongly in $W^{1,1}(\Omega)$. In particular $(u_{n_k})_{k\in\mathbb{N}}$ converges to v in $L^1(\Omega)$, thus $v = u$ a.e. and hence, $(u_{n_k})_{k\in\mathbb{N}}$ converges to u strongly in $W^{1,1}(\Omega)$. However, since every subsequence $(u_n)_{n\in\mathbb{N}}$ is bounded in $\mathrm{BV}^2(\Omega)$ we can repeat the same argument and deduce that for every subsequence $(u_{n_k})_{k\in\mathbb{N}}$ there exists a further subsequence $(u_{n_{k_h}})_{h\in\mathbb{N}}$ that converges to u strongly in $W^{1,1}(\Omega)$. This proves that the initial sequence $(u_n)_{n\in\mathbb{N}}$ converges to u strongly in $W^{1,1}(\Omega)$. $\qquad\square$

Corollary 3.3.6. *Strict convergence implies weak* convergence in $\mathrm{BV}^2(\Omega)$.*

Proof. The proof is straightforward using Lemma (3.3.5) and the fact that strict convergence implies weak* convergence in $\mathrm{BV}(\Omega, \mathbb{R}^d)$. □

As in the BV case, functions in $\mathrm{BV}^2(\Omega)$ can also be approximated strictly by smooth functions. In fact a stronger result was proved in [DT84] were in addition to strict approximation by a sequence of smooth functions $(u_n)_{n\in\mathbb{N}}$, also convergence of convex functions of the second order total variations $(\varphi(D^2 u_n))_{n\in\mathbb{N}}$ to $\varphi(D^2 u)$ was shown.

Theorem 3.3.7 (Demengel–Temam, 1984). *Suppose that $\Omega \subseteq \mathbb{R}^d$ is an open set with Lipschitz boundary and let $\varphi : \mathbb{R}^{d\times d} \to \mathbb{R}$ be a convex function with at most linear growth at infinity. Then for every $u \in \mathrm{BV}^2(\Omega)$ there exists a sequence $(u_n)_{n\in\mathbb{N}} \in \mathcal{C}^\infty(\Omega) \cap W^{2,1}(\Omega)$ such that $(u_n)_{n\in\mathbb{N}}$ converges to u strictly in $\mathrm{BV}^2(\Omega)$ and*

$$\lim_{n\to\infty} \varphi(D^2 u_n)(\Omega) = \varphi(D^2 u)(\Omega).$$

Note that for a function $u \in W^{2,1}(\Omega)$ the term $\varphi(D^2 u)(\Omega)$ is equal to $\int_\Omega \varphi(\nabla^2 u)\, dx$.

3.4 Well-posedness of the model

3.4.1 Existence and uniqueness

In this section we prove existence and uniqueness for the solutions of the minimisation problem (3.1). We suppose that Ω is an open, bounded, connected subset of $\mathbb{R}^2$ with Lipschitz boundary so that Theorems 2.3.5, 2.3.6, 3.3.3, 3.3.7 and Lemma 3.3.5 hold. We remark that the analysis is done in $\mathbb{R}^2$, i.e., $d = 2$, not only because we are planning to implement the model in two dimensional images but also because we are going to use the Poincaré inequality

$$\|u - u_\Omega\|_{L^2(\Omega)} \le C|Du|(\Omega), \quad \forall u \in \mathrm{BV}(\Omega),$$

which is not true for $d > 2$. Also for convenience, we take $s = 2$ in (3.1) as the case $s = 1$ is treated similarly. We assume that $T : L^2(\Omega) \to L^2(\Omega)$ is a bounded linear operator and $f \in L^2(\Omega)$.

We also assume that $\varphi_1 : \mathbb{R}^2 \to \mathbb{R}^+$ and $\varphi_2 : \mathbb{R}^4 \to \mathbb{R}^+$ are convex functions with at most linear growth at infinity and they both satisfy a coercivity condition, i.e., there exist constants $K_1, K_2 > 0$ such that

$$K_1|x| \le \varphi_1(x) \le K_2(1 + |x|), \quad \forall x \in \mathbb{R}^2, \tag{3.7}$$

$$K_1|x| \le \varphi_2(x) \le K_2(1 + |x|), \quad \forall x \in \mathbb{R}^4. \tag{3.8}$$

Thus, we are interested in the following minimisation problem

$$\inf_{u\in\mathrm{BV}^2(\Omega)} H(u), \tag{3.9}$$

where

$$H(u) := \frac{1}{2}\int_\Omega (Tu - f)^2 dx + \alpha\int_\Omega \varphi_1(\nabla u)\, dx + \beta\,\varphi_2(D^2 u)(\Omega). \tag{3.10}$$

Note that according to (3.5), H can also be written as

$$H(u) = \frac{1}{2}\int_\Omega (Tu - f)^2 dx + \alpha\int_\Omega \varphi_1(\nabla u)\, dx + \beta\int_\Omega \varphi_2(\nabla^2 u) \tag{3.11}$$
$$+ \beta\int_\Omega (\varphi_2)_\infty\left(\frac{D^s\nabla u}{|D^s\nabla u|}\right) d|D^s\nabla u|.$$

Observe that $\mathrm{BV}^2(\Omega)$ is the right space in which one can try to prove existence of solutions for (3.9). Indeed, if one tries to pose the problem is a Sobolev space setting, he/she would have to minimise the following functional:

$$F(u) = \frac{1}{2}\int_\Omega (Tu - f)^2 dx + \alpha\int_\Omega \varphi_1(\nabla u)\, dx + \beta\int_\Omega \varphi_2(\nabla^2 u)\, dx, \tag{3.12}$$

where now $\nabla^2 u$ is a function in $L^1(\Omega, \mathbb{R}^2)$. The natural space for the functional F to be defined in, is $W^{2,1}(\Omega)$. However, since this space is not reflexive we cannot prove existence of solutions for the minimisation of F using the direct method of calculus of variations. However, we will show that there is a connection between the F and H. We can naturally extend the functional F in $\mathrm{BV}^2(\Omega)$ (also denoted by F) as follows:

$$F(u) = \begin{cases} \frac{1}{2}\int_\Omega (Tu - f)^2 dx + \alpha\int_\Omega \varphi_1(\nabla u)\, dx \\ +\beta\int_\Omega \varphi_2(\nabla^2 u)\, dx, & \text{if } u \in W^{2,1}(\Omega) \\ +\infty, & \text{if } u \in \mathrm{BV}^2(\Omega)\setminus W^{2,1}(\Omega). \end{cases} \tag{3.13}$$

Even though F is now defined in $\mathrm{BV}^2(\Omega)$, which, as we have seen, has the useful weak* compactness property, one can still not prove existence of solutions of the minimisation of F using the direct method as the next proposition shows:

Proposition 3.4.1. *Consider the functional F defined in (3.13). Then F is not lower semicontinuous with respect to the strict topology in $\mathrm{BV}^2(\Omega)$ and hence it is neither with respect to the weak* topology in $\mathrm{BV}^2(\Omega)$.*

Proof. Note first that we can find a function $u \in \mathrm{BV}^2(\Omega)\setminus W^{2,1}(\Omega)$, for instance, see [Dem85] for such an example. Hence, from the definition of F, we have $F(u) = \infty$. However, according to the Theorem 3.3.7 we can find a sequence $(u_n)_{n\in\mathbb{N}}$ in $W^{2,1}(\Omega)$ that converges to u strictly in $\mathrm{BV}^2(\Omega)$. It follows that the sequences $(\|u_n\|_{L^1(\Omega)})_{n\in\mathbb{N}}$,

$(\|\nabla^2 u_n\|_{L^1(\Omega,\mathbb{R}^4)})_{n\in\mathbb{N}}$, as well as $(\|\nabla u_n\|_{L^1(\Omega,\mathbb{R}^2)})_{n\in\mathbb{N}}$ (using estimate (3.6)) are bounded. Using the Poincaré inequality we can easily check that $(u_n)_{n\in\mathbb{N}}$ is also bounded in $L^2(\Omega)$. From the fact that T is a bounded, linear operator and from conditions (3.7)–(3.8) on φ_1 and φ_2 we deduce that the sequence $(F(u_n))_{n\in\mathbb{N}}$ is bounded as well. Hence, we get

$$F(u) > \liminf_{n\to\infty} F(u_n),$$

which proves that F is not lower semicontinuous with respect to the strict topology in $\mathrm{BV}^2(\Omega)$. □

We are going to prove that H is the lower semicontinuous envelope of F with respect to the weak* topology in $\mathrm{BV}^2(\Omega)$, i.e.,

$$H = \mathrm{sc}_{(\mathrm{BV}^2(\Omega),w^*)}(F).$$

The proof is done in two steps. Firstly, we show that H is indeed lower semicontinuous with respect to the weak* topology.

Theorem 3.4.2. *The functional H is lower semicontinuous with respect to the weak* topology in $\mathrm{BV}^2(\Omega)$.*

Proof. Let $u \in \mathrm{BV}^2(\Omega)$, $(u_n)_{n\in\mathbb{N}} \subseteq \mathrm{BV}^2(\Omega)$ such that $(u_n)_{n\in\mathbb{N}}$ converges to u weakly* in $\mathrm{BV}^2(\Omega)$. We have to show that

$$H(u) \leq \liminf_{n\to\infty} H(u_n). \tag{3.14}$$

From the weak* convergence we have that $(u_n)_{n\in\mathbb{N}}$ converges to u in $W^{1,1}(\Omega)$. Using the standard Sobolev inequality, see for instance [Eva10],

$$\|v\|_{L^2(\Omega)} \leq C\|v\|_{W^{1,1}(\Omega)}, \quad \forall v \in W^{1,1}(\Omega),$$

we deduce that $(u_n)_{n\in\mathbb{N}}$ converges to u in $L^2(\Omega)$. Since $T : L^2(\Omega) \to L^2(\Omega)$ is continuous we get that the map $u \mapsto \frac{1}{2}\int_\Omega (Tu - f)^2 dx$ is continuous and hence we have

$$\frac{1}{2}\int_\Omega (Tu_n - f)^2 dx \to \frac{1}{2}\int_\Omega (Tu - f)^2 dx, \quad \text{as } n \to \infty. \tag{3.15}$$

It is not hard to check that since φ_1 is convex with at most linear growth at infinity, then is Lipschitz, say with constant $L > 0$. Using that fact we get

$$\left|\int_\Omega \varphi_1(\nabla u_n)\,dx - \int_\Omega \varphi_1(\nabla u)\,dx\right| \leq \int_\Omega |\varphi_1(\nabla u_n) - \varphi_1(\nabla u)|\,dx$$
$$\leq L\int_\Omega |\nabla u_n - \nabla u|\,dx. \tag{3.16}$$

From the estimate (3.16) and the fact that $(\nabla u_n)_{n\in\mathbb{N}}$ converges to ∇u in $L^1(\Omega, \mathbb{R}^2)$ we eventually get that

$$\int_\Omega \varphi_1(\nabla u_n)\,dx \to \int_\Omega \varphi_1(\nabla u)\,dx, \quad \text{as } n \to \infty. \tag{3.17}$$

Finally from the weak* convergence we have that $(D^2 u_n)_{n\in\mathbb{N}}$ converges to $D^2 u$ weakly* in $\mathcal{M}(\Omega, \mathbb{R}^{2\times 2})$. We can then apply Theorem 3.2.3 for $\mu_n = \mu = \mathcal{L}^2$, $\nu = D^2 u$, $\nu_n = D^2 u_n$ and get

$$\varphi_2(D^2 u)(\Omega) \le \liminf_{n\to\infty} \varphi_2(D^2 u_n)(\Omega). \tag{3.18}$$

Combining (3.15), (3.17) and (3.18) we derive (3.14). □

Theorem 3.4.3. *The functional H is the lower semicontinuous envelope of F with respect to the weak* topology in $\mathrm{BV}^2(\Omega)$.*

Proof. We are going to use the characterisation of lower semicontinuous envelopes as that was given in Proposition 2.4.4. Condition (i) of Proposition 2.4.4 follows from the lower semicontinuity of H and the fact that $H \le F$. Thus it suffices to prove condition (ii), i.e., to prove that for every $u \in \mathrm{BV}^2(\Omega)$, there exists a sequence $(u_n)_{n\in\mathbb{N}} \in \mathrm{BV}^2(\Omega)$ converging to u weakly* and

$$H(u) = \lim_{n\to\infty} F(u_n). \tag{3.19}$$

From Theorem 3.3.7 we get the existence of a sequence $(u_n)_{n\in\mathbb{N}} \in \mathcal{C}^\infty(\Omega) \cap W^{2,1}(\Omega)$ such that $(u_n)_{n\in\mathbb{N}}$ converges strictly, and thus also weakly* in $\mathrm{BV}^2(\Omega)$. Moreover, from the same theorem we also get that

$$\lim_{n\to\infty} \varphi_2(D^2 u_n)(\Omega) = \varphi_2(D^2 u)(\Omega). \tag{3.20}$$

We also have that $(u_n)_{n\in\mathbb{N}}$ converges to u in $W^{1,1}(\Omega)$ which, according to the proof of Theorem 3.4.2 implies that

$$\frac{1}{2}\int_\Omega (Tu-f)^2 dx + \alpha \int_\Omega \varphi_1(\nabla u_n)\,dx \to \frac{1}{2}\int_\Omega (Tu-f)^2 dx + \alpha \int_\Omega \varphi_1(\nabla u_n)\,dx, \quad \text{as } n \to \infty. \tag{3.21}$$

Combining (3.20), (3.21) and that fact that $H = F$ on $\mathcal{C}^\infty(\Omega) \cap W^{2,1}(\Omega)$ we get (3.19). □

Observe that since strict convergence induces weak* convergence, we get that H is also the lower semicontinuous envelope of F with respect to the strict convergence in $\mathrm{BV}^2(\Omega)$, i.e.,

$$H = \mathrm{sc}_{(\mathrm{BV}^2(\Omega), w^*)}(F) = \mathrm{sc}_{(\mathrm{BV}^2(\Omega), \mathrm{strict})}(F).$$

Let us note here that Theorem 3.4.3 can also be proved by incorporating a general relaxation result in [AC94], where the authors identify lower semicontinuous envelopes of functionals of the type $\int_\Omega \varphi(\nabla^k u)\,dx$ in higher order BV spaces.

We are now ready to prove existence of solutions for the minimisation problem (3.9), where we essentially use the lower semicontinuity of H, combined with the weak* compactness in $\text{BV}^2(\Omega)$. The proof of the following theorem follows the proof of the corresponding theorem in [Ves01] which deals with the minimisation of the analogue first order functional ($\beta = 0$).

Theorem 3.4.4. *In addition to the already mentioned conditions on Ω, T, φ_1 and φ_2, assume also that $\alpha > 0$, $\beta > 0$ and $T(\mathcal{X}_\Omega) \neq 0$. Then the minimisation problem $\min_{u \in \text{BV}^2(\Omega)} H(u)$, i.e.,*

$$\min_{u \in \text{BV}^2(\Omega)} \frac{1}{2} \int_\Omega (Tu - f)^2 dx + \alpha \int_\Omega \varphi_1(\nabla u)\, dx + \beta\, \varphi_2(D^2 u)(\Omega), \qquad (3.22)$$

has a solution $u^\star \in \text{BV}^2(\Omega)$.

Proof. Let $(u_n)_{n \in \mathbb{N}}$ be a minimising sequence for (3.22), i.e., $\lim_{n \to \infty} H(u_n) = \inf H$ and let $M > 0$ be an upper bound for $(H(u_n))_{n \in \mathbb{N}}$. We have that

$$\sup_{n \in \mathbb{N}} \varphi_2(D^2 u_n)(\Omega) < M, \quad \sup_{n \in \mathbb{N}} \int_\Omega \varphi_1(\nabla u_n)\, dx < M \quad \text{and} \quad \sup_{n \in \mathbb{N}} \frac{1}{2} \int_\Omega (Tu_n - f)^2 dx < M. \tag{3.23}$$

The estimates (3.23) together with the coercivity assumptions (3.7)–(3.8) give that

$$\sup_{n \in \mathbb{N}} |Du_n|(\Omega) = \sup_{n \in \mathbb{N}} \int_\Omega |\nabla u_n|\, dx < \frac{M}{K_1}, \tag{3.24}$$

$$\sup_{n \in \mathbb{N}} |D^2 u_n|(\Omega) < \frac{M}{K_1}. \tag{3.25}$$

We now show that the sequence $(u_n)_{n \in \mathbb{N}}$ is bounded in $L^2(\Omega)$, following essentially [Ves01]. By the Poincaré inequality, there exists a constant $C > 0$ such that for every $n \in \mathbb{N}$

$$\begin{aligned}
\|u_n\|_{L^2(\Omega)} &= \left\| u_n - \mathcal{X}_\Omega \frac{1}{\mathcal{L}^2(\Omega)} \int_\Omega u_n\, dx + \mathcal{X}_\Omega \frac{1}{\mathcal{L}^2(\Omega)} \int_\Omega u_n\, dx \right\|_{L^2(\Omega)} \\
&\leq C|Du_n|(\Omega) + \frac{1}{\mathcal{L}^2(\Omega)} \left| \int_\Omega u_n\, dx \right| \\
&\leq \frac{CM}{K_1} + \frac{1}{\mathcal{L}^2(\Omega)} \left| \int_\Omega u_n\, dx \right|.
\end{aligned}$$

Thus in order to bound $(u_n)_{n \in \mathbb{N}}$ in $L^2(\Omega)$ it suffices to bound $|\int_\Omega u_n\, dx|$ uniformly in n. We have for every $n \in \mathbb{N}$

$$\left\| T\left(\mathcal{X}_\Omega \frac{1}{\mathcal{L}^2(\Omega)} \int_\Omega u_n\, dx \right) \right\|_{L^2(\Omega)} \leq \left\| T\left(\mathcal{X}_\Omega \frac{1}{\mathcal{L}^2(\Omega)} \int_\Omega u_n\, dx \right) - Tu_n \right\|_{L^2(\Omega)}$$
$$+ \|Tu_n - f\|_{L^2(\Omega)} + \|f\|_{L^2(\Omega)}$$

67

$$\leq \|T\| \left\| u_n - \mathcal{X}_\Omega \frac{1}{\mathcal{L}^2(\Omega)} \int_\Omega u_n \, dx \right\|_{L^2(\Omega)}$$

$$+ \|Tu_n - f\|_{L^2(\Omega)} + \|f\|_{L^2(\Omega)}$$

$$\leq C\|T\| |Du_n|(\Omega) + \sqrt{2M} + \|f\|_{L^2(\Omega)}$$

$$\leq \frac{CM}{K_1} \|T\| + \sqrt{2M} + \|f\|_{L^2(\Omega)}. \tag{3.26}$$

Thus, setting $K = \frac{CM}{K_1}\|T\| + \sqrt{2M} + \|f\|_{L^2(\Omega)}$ and using the fact that $T(\mathcal{X}_\Omega) \neq 0$, it follows from (3.26)

$$\left| \int_\Omega u_n \, dx \right| \|T(\mathcal{X}_\Omega)\|_{L^2(\Omega)} \leq K\mathcal{L}^2(\Omega),$$

and since $T(\mathcal{X}_\Omega) \neq 0$ we get

$$\left| \int_\Omega u_n \, dx \right| \leq \frac{K\mathcal{L}^2(\Omega)}{\|T(\mathcal{X}_\Omega)\|_{L^2(\Omega)}}.$$

Since the sequence $(u_n)_{n\in\mathbb{N}}$ is bounded in $L^2(\Omega)$ and Ω is a bounded domain, we have that it is also bounded in $L^1(\Omega)$, and from estimates (3.24)–(3.25) we deduce that it is bounded in $\mathrm{BV}^2(\Omega)$ as well. From Theorem 3.3.3 we obtain the existence of a subsequence $(u_k)_{k\in\mathbb{N}}$ that converges weakly* to a function $u^\star \in \mathrm{BV}^2(\Omega)$. Since the functional H is weakly* lower semicontinuous we have that

$$\inf F \leq H(u^\star) \leq \liminf_{n\to\infty} H(u_n) = \inf F,$$

which implies that

$$H(u^\star) = \min_{u\in\mathrm{BV}^2(\Omega)} H(u).$$

$\square$

Let us note here that in the above proof we used the fact that $\alpha > 0$ in order to obtain an $L^1(\Omega)$ bound in the gradient of the minimising sequence $(u_n)_{n\in\mathbb{N}}$ that was further used to bound $(u_n)_{n\in\mathbb{N}}$ in $L^2(\Omega)$ (for the case $\beta = 0$ see [Ves01]). However even in the case $\alpha = 0$, we can also bound $(u_n)_{n\in\mathbb{N}}$ in $L^2(\Omega)$, and hence in $\mathrm{BV}^2(\Omega)$ using (3.6), if T satisfies a condition of the type

$$M\|u\|_{L^2(\Omega)} \leq \|Tu\|_{L^2(\Omega)}, \quad \forall u \in L^2(\Omega), \tag{3.27}$$

for a constant $M > 0$. For example (3.27) is true if T has a bounded inverse by setting $M = 1/\|T^{-1}\|$. If T does not satisfy a condition of the type (3.27) then it is not clear how to get existence. However, we are still able to prove existence when $\alpha = 0$, for operators T that correspond to image inpainting. i.e., projections on subsets D of Ω. The proof is more involved than the one of Theorem 3.4.4 and makes use of theorems on BV traces.
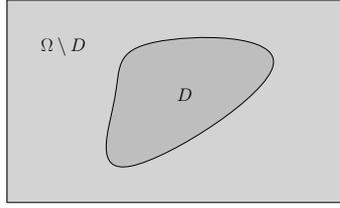
Figure 3.1: Inpainting domain D.

Theorem 3.4.5. *If $\beta > 0$, then the minimisation problem*

$$\inf_{u \in \mathrm{BV}^2(\Omega)} \int_{\Omega \backslash D} (u - f)^2 dx + \beta \, \varphi_2(D^2 u), \tag{3.28}$$

has a solution, where $D \Subset \Omega$ is an open, connected set with Lipschitz boundary, see also Figure 3.1.

Proof. Consider a minimising sequence $(u_n)_{n \in \mathbb{N}}$ for (3.28). From the coercivity assumptions (3.8) we get that

$$\sup_{n \in \mathbb{N}} |D^2 u_n|(\Omega) < \infty. \tag{3.29}$$

In order to prove the theorem it suffices to bound $(u_n)_{n \in \mathbb{N}}$ and $(\nabla u_n)_{n \in \mathbb{N}}$ in $L^1(\Omega)$ and $L^1(\Omega, \mathbb{R}^2)$ respectively. Then the proof can be finished using a straightforward application of the direct method as was done in the proof of Theorem 3.4.4. Since $(u_n)_{n \in \mathbb{N}}$ is a minimising sequence for (3.28) and Ω is bounded we have that

$$\sup_{n \in \mathbb{N}} \|u_n\|_{L^1(\Omega \backslash D)} < \infty. \tag{3.30}$$

Estimates (3.29), (3.30) in combination with (3.6) give that

$$\sup_{n \in \mathbb{N}} |D u_n|(\Omega \backslash D) < \infty. \tag{3.31}$$

Denote with $u_n^{(1)}$ and $u_n^{(2)}$ the restrictions of u_n in $\Omega \backslash D$ and D respectively. Then according to Theorems 2.3.7 and 2.3.9 we have for every $n \in \mathbb{N}$

$$|D u_n|(\Omega) = |D u_n|(\Omega \backslash D) + |D u_n|(D) + \|(u_n^{(1)})^{\Omega \backslash D} - (u_n^{(2)})^D\|_{L^1(\partial D; \mathcal{H})}. \tag{3.32}$$

Since $u_n \in W^{1,1}(\Omega)$ we have for every $n \in \mathbb{N}$

$$|D u_n|(\Omega) = \int_\Omega |\nabla u| \, dx = \int_{\Omega \backslash D} |\nabla u| \, dx + \int_D |\nabla u| \, dx = |D u_n|(\Omega \backslash D) + |D u_n|(D). \tag{3.33}$$

From (3.32) and (3.33) we deduce that

$$\|(u_n^{(1)})^{\Omega \backslash D}\|_{L^1(\partial D; \mathcal{H})} = \|(u_n^{(2)})^D\|_{L^1(\partial D; \mathcal{H})}. \tag{3.34}$$

From the estimate (2.4) of Theorem 2.3.7 and from (3.30), (3.31) and (3.34) we get that

$$\sup_{n\in\mathbb{N}} \|(u_n^{(2)})^D\|_{L^1(\partial D;\mathcal{H})} < \infty. \tag{3.35}$$

Suppose now that $\sup_{n\in\mathbb{N}} \|u_n\|_{L^1(D)} = \infty$. Then from the estimate (2.7) and (3.35), that would mean that

$$\sup_{n\in\mathbb{N}} |Du_n|(D) = \infty, \tag{3.36}$$

i.e, there exists an $i_0 \in \{1,\dots,d\}$ such that

$$\sup_{n\in\mathbb{N}} \left\| \frac{\partial u_n}{\partial x_{i_0}} \right\|_{L^1(D)} = \infty, \tag{3.37}$$

but this is a contradiction. In order to see that, observe first that again from Theorems 2.3.7 and 2.3.9 we get for every $n \in \mathbb{N}$

$$|D^2 u_n|(\Omega) = |D^2 u_n|(\Omega \setminus D) + |D^2 u_n|(D) + \|(\nabla u_n^{(1)})^{\Omega\setminus D} - (\nabla u_n^{(2)})^D\|_{L^1(\partial D,\mathbb{R}^d;\mathcal{H})}. \tag{3.38}$$

From (3.29), (3.31) and the estimate (2.4) we get that

$$\sup_{n\in\mathbb{N}} \|(\nabla u_n^{(1)})^{\Omega\setminus D}\|_{L^1(\partial D,\mathbb{R}^d;\mathcal{H})} < \infty, \tag{3.39}$$

which in combination with (3.29) and (3.38) gives

$$\sup_{n\in\mathbb{N}} \|(\nabla u_n^{(2)})^D\|_{L^1(\partial D,\mathbb{R}^d;\mathcal{H})} < \infty \Rightarrow \sup_{n\in\mathbb{N}} \left\| \left(\frac{\partial u_n^{(2)}}{\partial x_{i_0}}\right)^D \right\|_{L^1(\partial D;\mathcal{H})} < \infty. \tag{3.40}$$

But again estimate (2.7) gives

$$\left\| \frac{\partial u_n}{\partial x_{i_0}} \right\|_{L^1(D)} \leq C \left(\left| D\left(\frac{\partial u_n}{\partial x_{i_0}}\right) \right|(D) + \left\| \left(\frac{\partial u_n^{(2)}}{\partial x_{i_0}}\right)^D \right\|_{L^1(\partial D;\mathcal{H})} \right). \tag{3.41}$$

From (3.37), (3.40) and (3.41) we deduce

$$\sup_{n\in\mathbb{N}} \left| D\left(\frac{\partial u_n}{\partial x_{i_0}}\right) \right|(D) = \infty,$$

which cannot happen due to (3.29). $\qquad\square$

Observe that the above proof works also in the case where the inpainting domain consists of finitely many open connected sets $(D_i)_{i=1,\dots,N} \Subset \Omega$ with Lipschitz boundaries such that $\mathrm{dist}(D_i, D_j) > 0$ for every $i \neq j$.

We now prove uniqueness of solutions for (3.9). The following proof also follows the proof of the corresponding theorem for the first order analogue in [Ves01].

Theorem 3.4.6. *If, in addition to $T(\mathcal{X}_\Omega)$, T is injective or if φ_1 is strictly convex, then the solution of the minimisation problem (3.9) is unique.*

Proof. Suppose that u_1 and u_2 are two minimisers for (3.9). Using Proposition 3.2.1, it can be easily checked that H is convex. If $Tu_1 \neq Tu_2$ then from the strict convexity of the fidelity term in H we have

$$H\left(\frac{1}{2}u_1 + \frac{1}{2}u_2\right) < \frac{1}{2}H(u_1) + \frac{1}{2}H(u_2) = \inf H,$$

which is a contradiction. Thus we must have $Tu_1 = Tu_2$. If T is injective, we have $u_1 = u_2$. If T is not injective but φ_1 is strictly convex, we must have $\nabla u_1 = \nabla u_2$ otherwise we get the same contradiction as before. In that case, since Ω is connected, there exists a constant c such that $u_1 = u_2 + c\mathcal{X}_\Omega$ and since $T(\mathcal{X}_\Omega) \neq 0$, we get $c = 0$. $\qquad\square$

We note that in general we cannot obtain uniqueness results when the L^1 norm is used in the fidelity term, due to the lack of strict convexity.

3.4.2 Stability

In order to complete the well-posedness picture for the problem

$$\min_{u \in \mathrm{BV}^2(\Omega)} \frac{1}{2}\int_\Omega (Tu - f)^2 dx + \alpha \int_\Omega \varphi_1(\nabla u)\, dx + \beta\, \varphi_2(D^2 u)(\Omega), \qquad (3.42)$$

it remains to analyse its stability. More precisely, we want to know which effect deviations in the data f have on a corresponding minimiser of (3.42). Ideally the deviation in the minimisers for different input data should be bounded by the deviation of the data. Let Ψ be the regularising functional in (3.42), i.e.,

$$\Psi(u) = \alpha \int_\Omega \varphi_1(\nabla u)\, dx + \beta\, \varphi_2(D^2 u)(\Omega).$$

It has been demonstrated by many authors [BO04, BRH07, Pös08, BB11] that Bregman distances related to the regularisation functional Ψ are natural error measures for variational regularisation methods with Ψ convex. In particular Pöschl [Pös08] has derived estimates for variational regularisation methods for power of metrics, which apply to the functional we consider here. However, for demonstration issues and in order to make the constants in the estimates more explicit we state and prove here the result for a special case of (3.42).

For what we are going to do we assume that one of the regularisers is differentiable. Without use of generality, we assume that the convex function ϕ_1 is differentiable. Analo-

gous analysis can be done if ϕ_2 is differentiable or even under weaker continuity solutions, see [BL11]. Let $\tilde{u}$ be the original image and $\tilde{f}$ the exact data (without noise), i.e., $\tilde{f}$ is a solution of $T\tilde{u} = \tilde{f}$. We assume that the noisy data deviate from the exact data by $\|\tilde{f} - f\|_{L^2(\Omega)} \leq \delta$, for a small $\delta > 0$. For the original image $\tilde{u}$ we assume that the following condition, called source condition, holds

$$\text{There exists a } \xi \in \partial\Psi(\tilde{u}) \text{ such that } \xi = T^*q \text{ for a source element } q \in L^2(\Omega), \quad \text{(SC)}$$

where T^* is the adjoint of T. As it was shown in [BO04] for the general setting, the elements that satisfy SC are exactly the minimisers of (3.42) for arbitrary f, α and β. Since ϕ_1 is differentiable and both ϕ_1 and ϕ_2 are convex, the subdifferential of Ψ can be written as a sum of subdifferentials [ET76]

$$\partial\Psi(u) = \alpha\,\partial\left(\int_\Omega \phi_1(\nabla(\cdot))dx\right)(u) + \beta\,\partial\left(\phi_2(D^2(\cdot))(\Omega)\right)(u)$$
$$= -\alpha\,\text{div}(\phi_1'(\nabla u)) + \beta\,\partial\left(\phi_2(D^2(\cdot))(\Omega)\right)(u).$$

We also define the symmetric Bregman distance for the regularising functional Ψ as

$$D_\Psi^{symm}(u_1, u_2) := \langle p_1 - p_2, u_1 - u_2 \rangle, \quad p_1 \in \partial\Psi(u_1),\ p_2 \in \partial\Psi(u_2).$$

Theorem 3.4.7. *Let $\tilde{u}$ be the original image with source condition SC satisfying $T\tilde{u} = \tilde{f}$. Let $f \in L^2(\Omega)$ be the noisy data with $\|\tilde{f} - f\|_{L^2(\Omega)} \leq \delta$. Then a minimiser u of (3.42) satisfies*

$$\alpha D_{\int_\Omega \phi_1(\nabla(\cdot))}^{symm}(u, \tilde{u}) + \beta D_{\phi_2(D^2(\cdot))(\Omega)}^{symm}(u, \tilde{u}) + \frac{1}{2}\|Tu - \tilde{f}\|_{L^2(\Omega)} \leq \|q\|_{L^2(\Omega)}^2 + \delta^2, \quad (3.43)$$

where q is the source element in SC. Moreover if u_1 and u_2 are minimisers for (3.42) with data f_1 and f_2 then we have the following estimate

$$\alpha D_{\int_\Omega \phi_1(\nabla\cdot)}^{symm}(u_1, u_2) + \beta D_{\phi_2(D^2(\cdot))(\Omega)}^{symm}(u_1, u_2) + \frac{1}{2}\|Tu_1 - Tu_2\|_{L^2(\Omega)} \leq \frac{1}{2}\|f_1 - f_2\|_{L^2(\Omega)}^2. \quad (3.44)$$

Proof. The optimality condition for (3.42) reads

$$\alpha p_1 + \beta p_2 + T^*(Tu - f) = 0, \quad \text{where} \quad p_1 = -\text{div}(\phi_1'(\nabla u)),\ p_2 \in \partial\left(\phi_2(D^2(\cdot))(\Omega)\right)(u) \quad (3.45)$$

Adding the element $\xi = T^*q$ from SC and $T^*\tilde{f}$ to (3.45), we get

$$\alpha p_1 + \beta p_2 - \xi + T^*(Tu - \tilde{f}) = T^*((f - \tilde{f}) - q). \quad (3.46)$$

Note now that $\xi = \alpha\xi_1 + \beta\xi_2$ for some ξ_1 and ξ_2 that belong to the subdifferential of $\int_\Omega \phi_1(\nabla\cdot)$ and $\phi_2(D^2(\cdot))(\Omega)$ respectively. Using that fact and taking the duality product

of (3.46) with $u - \tilde{u}$ we get

$$\alpha\langle p_1 - x_1, u - \tilde{u}\rangle + \beta\langle p_2 - \xi_2, u - \tilde{u}\rangle + \|Tu - \tilde{f}\|_{L^2} = \langle(f - \tilde{f}) - q, Tu - \tilde{f}\rangle. \quad (3.47)$$

where we used $T\tilde{u} = \tilde{f}$. By Young's inequality, we eventually get

$$\alpha D^{symm}_{\int_\Omega \phi_1(\nabla(\cdot))}(u, \tilde{u}) + \beta D^{symm}_{\phi_2(D^2(\cdot))(\Omega)}(u, \tilde{u}) + \frac{1}{2}\|Tu - \tilde{f}\|_{L^2(\Omega)} \leq \|q\|^2_{L^2(\Omega)} + \|\tilde{f} - f\|^2_{L^2(\Omega)}, \quad (3.48)$$

and in view of $\|\tilde{f} - f\|_{L^2(\Omega)} \leq \delta$ we get (3.43). Similarly, if u_1 and u_2 are minimisers for (3.42) with data f_1 and f_2 we can derive estimate (3.44) by subtracting the optimality conditions for u_1 and u_2 and applying Young's inequality once again. $\qquad\square$

3.5   The numerical implementation

It is a common practice in mathematical imaging and inverse problems in general, to introduce a model in the continuous set up and after analysing it, to discretise functions and operators in order to implement it in a computer. Here we make no exceptions. In order to be completely rigorous, the link between the continuous and the discretised model should be provided. Such a link is normally given in terms of Γ-convergence [Bra02, DM93] of the discrete functional to the continuous one as the resolution of the image tends to infinity. However, as this is not our purpose here we leave this issue for future research. We use standard discretisations for the norms, differential operators and the operator T. The discretisation of the latter is the obvious one as we only consider T to be the identity function (denoising), a convolution operator (deblurring) or a projection operator (inpainting). However, let us just mention here that in some advanced medical imaging tasks e.g. MRI imaging, the way one discretises the problem is not trivial and it is crucial for the quality of the reconstruction results.

In this section we work with the discretised version of the problem (3.1) and we discuss its numerical realisation by the so-called split Bregman technique [GO09]. We start by defining the discrete versions of L^1 and L^2 norms and we also introduce the appropriate discrete differential operators. We proceed with an introduction to the Bregman iteration which is used to solve constrained optimisation problems, an idea originated in [OBG$^+$]. In [GO09] the Bregman iteration and an operator splitting technique (split Bregman) was used to solve the total variation minimisation problem after reformulating the problem as a constrained one. In the latter paper it was also also proved that the iterates of the Bregman iteration converge to the solution of the constrained problem assuming that the iterates satisfy the constraint in a finite number of iterations. Here, we give a more general convergence result where we do not use that assumption, see Theorem 3.5.1. We proceed with describing how our problem can be solved with the Bregman iteration, using the splitting procedure mentioned above. We finish by applying our method to image denoising

and deblurring comparing it with TGV, TV^2 and infimal convolution minimisation, with respect both to reconstruction quality and computational time. An extensive study of the application of our method to image inpainting is presented in Section 3.7.

3.5.1 Discretisation of the model

In this section we study the discretisation of the problem (3.1). In our numerical examples we set $\varphi_1 = |\cdot|$ and $\varphi_2 = |\cdot|$, the Euclidean norms in $\mathbb{R}^2$ and $\mathbb{R}^4$ respectively. That is to say we discretise the following problem

$$\min_{u \in \mathrm{BV}^2(\Omega)} \frac{1}{s} \int_\Omega |Tu - f|^s dx + \alpha \,\mathrm{TV}(u) + \beta \,\mathrm{TV}^2(u), \quad s = 1, 2. \tag{3.49}$$

In order to do so, we specify the corresponding discrete operators and norms that appear in the continuous functional. In the discrete setting u and f are elements of $\mathbb{R}^{N \times M}$ where $N \times M$ is the resolution of the image and $T : \mathbb{R}^{N \times M} \to \mathbb{R}^{N \times M}$, is a linear operator. Here, we consider only greyscale images and we describe an extension to colour images for the application of (3.49) to image inpainting in Section 3.7.3. We define the finite dimensional L^1 and L^2 norms of a discrete vector field $U : \{1, \ldots, N\} \times \{1, \ldots, M\} \to \mathbb{R}^K$ with $U = (U_1, \ldots, U_K)$ as follows:

$$\|U\|_2 = \left(\sum_{i=1}^N \sum_{j=1}^M \sum_{k=1}^K U_k(i,j)^2 \right)^{1/2},$$

$$\|U\|_1 = \sum_{i=1}^N \sum_{j=1}^M \left(\sum_{k=1}^K U_k(i,j)^2 \right)^{1/2}.$$

We also define the discrete first and second order differential operators D_x, D_y, D_{xx}, D_{yy} and D_{xy} which are all operators from $\mathbb{R}^{N \times M}$ to $\mathbb{R}^{N \times M}$. We note here that we assume periodic boundary conditions for u. By choosing periodic boundary conditions, the action of each of the discrete differential operators can be regarded as a circular convolution of u and allows the use of Fast Fourier Transform (FFT), see also [WT10]. Thus, even though the use of periodic boundary conditions is not necessary, it has been chosen in order to speed up the computational performance of our implementation, especially as far as deblurring and inpainting are concerned.

In our discrete implementation, the coordinates x and y are oriented along columns and rows respectively. We define the discrete gradient $\nabla u = (D_x u, D_y u)$ as a forward finite difference operator

$$D_x u(i,j) = \begin{cases} u(i, j+1) - u(i,j) & \text{if } 1 \leq i \leq N, \ 1 \leq j < M, \\ u(i, 1) - u(i,j) & \text{if } 1 \leq i \leq N, \ j = M, \end{cases}$$

$$D_y u(i,j) = \begin{cases} u(i+1,j) - u(i,j) & \text{if } 1 \le i < N,\ 1 \le j \le M, \\ u(1,j) - u(i,j) & \text{if } i = N,\ 1 \le j \le M. \end{cases}$$

We also need the discrete divergence operator div : $\left(\mathbb{R}^{N \times M}\right)^2 \to \mathbb{R}^{N \times M}$ that has the adjointness property

$$-\text{div} p \cdot u = p \cdot \nabla u, \quad \forall u \in \mathbb{R}^{N \times M},\ p \in \left(\mathbb{R}^{N \times M}\right)^2.$$

This property is essentially the discrete analogue of integration by parts. For a $p = (p_1, p_2) \in \left(\mathbb{R}^{N \times M}\right)^2$ we define

$$(\text{div} p)(i,j) = \overleftarrow{D_x} p_1(i,j) + \overleftarrow{D_y} p_2(i,j),$$

where $\overleftarrow{D_x}$ and $\overleftarrow{D_y}$ are the following backward finite difference operators:

$$\overleftarrow{D_x} u(i,j) = \begin{cases} u(i,j) - u(i,M) & \text{if } 1 \le i \le N,\ j = 1, \\ u(i,j) - u(i,j-1) & \text{if } 1 \le i \le N,\ 1 < j \le M, \end{cases}$$

$$\overleftarrow{D_y} u(i,j) = \begin{cases} u(i,j) - u(N,j) & \text{if } i = 1,\ 1 \le j \le M, \\ u(i,j) - u(i-1,j) & \text{if } 1 < i \le N,\ 1 \le j \le M. \end{cases}$$

Analogously we define the discrete Hessian $\nabla^2 u = (D_{xx} u, D_{yy} u, D_{xy} u, D_{yx} u)$. This consists of the corresponding compositions of the first order operators,

$$D_{xx} u(i,j) = \begin{cases} u(i,m) - 2u(i,j) + u(i,j+1) & \text{if } 1 \le i \le N,\ j = 1, \\ u(i,j-1) - 2u(i,j) + u(i,j+1) & \text{if } 1 \le i \le N,\ 1 < j < M, \\ u(i,j-1) - 2u(i,j) + u(i,1) & \text{if } 1 \le i \le n,\ j = M, \end{cases}$$

$$D_{yy} u(i,j) = \begin{cases} u(n,j) - 2u(i,j) + u(i+1,j) & \text{if } i = 1,\ 1 \le j \le M, \\ u(i-1,j) - 2u(i,j) + u(i+1,j) & \text{if } 1 < i < N,\ 1 \le j \le M, \\ u(i-1,j) - 2u(i,j) + u(1,i) & \text{if } i = N,\ 1 \le j \le M, \end{cases}$$

$$D_{xy} u(i,j) = \begin{cases} u(i,j) - u(i+1,j) - u(i,j+1) + u(i+1,j+1) & \text{if } 1 \le i < N,\ 1 \le j < M, \\ u(i,j) - u(1,j) - u(i,j+1) + u(1,j+1) & \text{if } i = N,\ 1 \le j < M, \\ u(i,j) - u(i+1,j) - u(i,1) + u(i+1,1) & \text{if } 1 \le i < N,\ j = M, \\ u(i,j) - u(1,j) - u(i,1) + u(1,1) & \text{if } i = N,\ j = M. \end{cases}$$

One can easily check that $D_{xy} = D_x(D_y) = D_y(D_x) = D_{yx}$. As before, we also need the discrete second order divergence operator $\text{div}^2 : \left(\mathbb{R}^{N\times M}\right)^4 \to \mathbb{R}^{N\times M}$ with the adjointness property

$$\text{div}^2 q \cdot u = q \cdot \nabla^2 u, \quad \forall u \in \mathbb{R}^{n\times m}, \ q \in \left(\mathbb{R}^{N\times M}\right)^4.$$

For a $q = (q_{11}, q_{22}, q_{12}, q_{21}) \in \left(\mathbb{R}^{N\times M}\right)^4$ we define

$$(\text{div}^2 q)(i,j) = \overleftarrow{D_{xx}}q_{11}(i,j) + \overleftarrow{D_{yy}}q_{22}(i,j) + \overleftarrow{D_{xy}}q_{12}(i,j) + \overleftarrow{D_{yx}}q_{21}(i,j),$$

where

$$\overleftarrow{D_{xx}} = D_{xx}, \ \ \overleftarrow{D_{yy}} = D_{yy}, \ \ \overleftarrow{D_{xy}} = \overleftarrow{D_{yx}} \quad \text{with}$$

$$\overleftarrow{D_{xy}}u(i,j) = \begin{cases} u(i,j) - u(i,M) - u(N,i) + u(N,M) & \text{if } i=1, \ j=1, \\ u(i,j) - u(i,j-1) - u(N,j) + u(N,j-1) & \text{if } i=1, \ 1 < j \le M, \\ u(i,j) - u(i-1,j) - u(i,M) + u(i-1,M) & \text{if } 1 < i \le M \ j=1, \\ u(i,j) - u(i,j-1) - u(i-1,j) + u(i-1,j-1) & \text{if } 1 < i \le N, \ 1 < j \le M. \end{cases}$$

Figure 3.2 describes schematically the action of all the above discrete differential operators. According to all the definitions above the discretised version of problem (3.49) can be
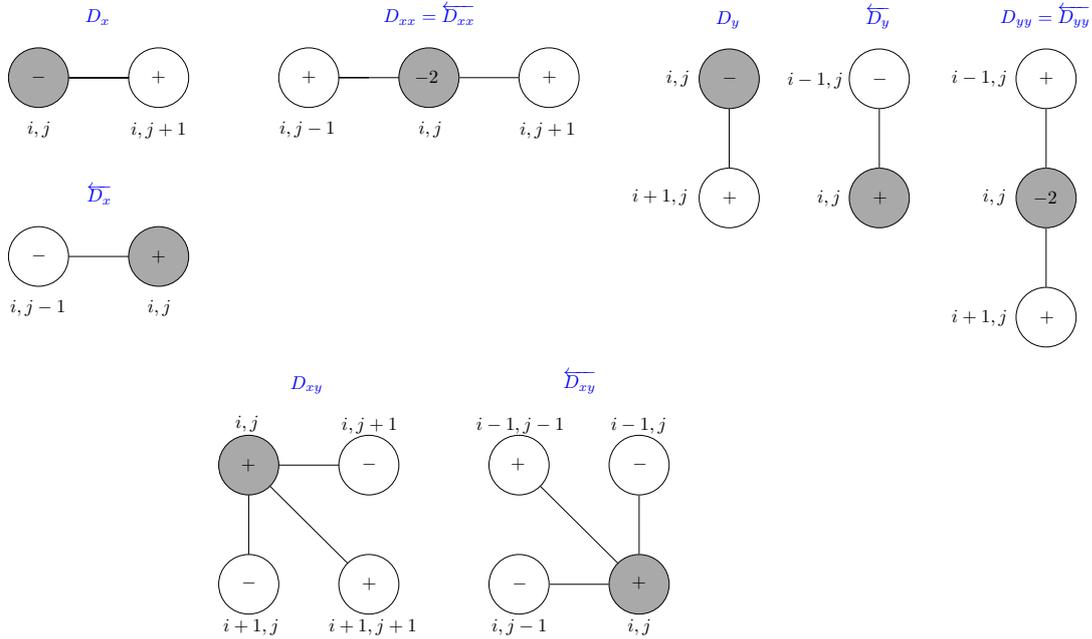
Figure 3.2: Illustration of the discrete derivative approximations.

written as

$$\min_{u\in\mathbb{R}^{N\times M}} \frac{1}{s}\|Tu - f\|_s^s + \alpha\|\nabla u\|_1 + \beta\|\nabla^2 u\|_1, \quad s = 1, 2. \tag{3.50}$$

3.5.2 Bregman iteration

We now introduce the Bregman and split Bregman iterations as suitable numerical methods for the solution of (3.50). We would like to recall some basic aspects of the general theory of Bregman iteration before finishing this discussion with a novel convergence result in Theorem 3.5.1.

Suppose we want to solve the following constrained minimisation problem

$$\min_{u \in \mathbb{R}^L} E(u), \quad \text{such that } Au = b, \tag{3.51}$$

where E is a convex function and $A : \mathbb{R}^L \to \mathbb{R}^D$ a linear map. We can transform the constrained minimisation problem (3.51) into an unconstrained one, introducing a parameter λ:

$$\sup_{\lambda} \min_{u \in \mathbb{R}^L} E(u) + \frac{\lambda}{2} \|Au - b\|_2^2, \tag{3.52}$$

where in order to satisfy the constraint $Au = b$, one has to let λ increase to infinity. Instead of doing that we perform the Bregman iteration as it was proposed in [OBG$^+$] and [YOGD08]:

_____

Bregman Iteration

$$
\begin{aligned}
u_{n+1} &= \operatorname*{argmin}_{u \in \mathbb{R}^L} E(u) + \frac{\lambda}{2} \|Au - b_n\|_2^2, &\tag{3.53}\\
b_{n+1} &= b_n + b - Au_{n+1}. &\tag{3.54}
\end{aligned}
$$

_____

In [OBG$^+$], assuming that (3.53) has a unique solution, the authors derived among others, the following facts about the iterates u_n:

(i) The constraint is satisfied at infinity: $\|Au_n - b\|_2^2 < \dfrac{C_1}{n-1}, \quad C_1 > 0, \ \forall n \geq 2.$ (3.55)

(ii) Monotonic decrease of the residuals: $\|Au_{n+1} - b\|_2^2 \leq \|Au_n - b\|_2^2, \quad \forall n \in \mathbb{N}.$ (3.56)

(iii) Summability of the residuals: $\displaystyle\sum_{n=0}^{\infty} \|Au_n - b\|_2^2 < \infty,$ (3.57)

(iv) Uniform bound on $E(u_n)$: $E(u_n) < C_2, \quad C_2 > 0, \ \forall n \in \mathbb{N}.$ (3.58)

Using these results, we are able to prove convergence of the Bregman iterates to a solution of the original constrained problem (3.51). This result was proved in [GO09] in the case when the constraint is satisfied in a finite number of iterations, i.e., $Au_{n_0} = b$ for some iterate u_{n_0}. Here we do not use that assumption, something that makes our result a

genuine contribution to the convergence theory of Bregman iteration.

Theorem 3.5.1. *Suppose that the constrained minimisation problem* (3.51) *has a unique solution $u^\star$. Moreover suppose that the convex function E is coercive and that* (3.53) *has a unique solution for every $n \in \mathbb{N}$. Then the sequence of the iterates $(u_n)_{n \in \mathbb{N}}$ of Bregman iteration converges to $u^\star$.*

Proof. Since (3.53) has a unique solution for every $n \in \mathbb{N}$ we have that the statements (3.55)–(3.58) hold. Moreover we have for every $n \geq 1$

$$b_n = b_0 + \sum_{\ell=1}^{n}(b - Au_\ell) \Rightarrow \|b_n\|_2 \leq \|b_0\|_2 + \sum_{\ell=1}^{n} \|b - Au_\ell\|_2. \tag{3.59}$$

From (3.58) and the coercivity of E we have that the sequence $(\|u_n\|_2)_{n \in \mathbb{N}}$ is bounded, say by a constant $C > 0$. Thus, it suffices to show that every accumulation point of $(u_n)_{n \in \mathbb{N}}$ is equal to $u^\star$. Using the fact that $Au^\star = b$, together with (3.55), we have for every increasing sequence of naturals $(n_k)_{k \in \mathbb{N}}$, $k > 2$,

$$
\begin{aligned}
\frac{\lambda}{2}\|Au^\star - b_{n_k-1}\|_2^2 &\leq \frac{\lambda}{2}(\|Au^\star - Au_{n_k}\|_2 + \|Au_{n_k} - b_{n_k-1}\|_2)^2 \\
&= \frac{\lambda}{2}(\|Au_{n_k} - b\|_2 + \|Au_{n_k} - b_{n_k-1}\|_2)^2 \\
&\leq \frac{\lambda C_1}{2(n_k - 1)} + \lambda\|Au_{n_k} - b\|_2\|Au_{n_k} - b_{n_k-1}\|_2 + \frac{\lambda}{2}\|Au_{n_k} - b_{n_k-1}\|_2^2.
\end{aligned}
\tag{3.60}
$$

Since u_{n_k} is a solution to (3.53) we have

$$E(u_{n_k}) + \frac{\lambda}{2}\|Au_{n_k} - b_{n_k-1}\|_2^2 \leq E(u^\star) + \frac{\lambda}{2}\|Au^\star - b_{n_k-1}\|_2^2, \quad \forall k \geq 1. \tag{3.61}$$

Combining (3.60) and (3.61) with (3.55) and (3.59) we get for all $k \geq 2$,

$$
\begin{aligned}
E(u_{n_k}) &\leq E(u^\star) + \frac{\lambda C_1}{2(n_k - 1)} + \lambda\|Au_{n_k} - b\|_2\|Au_{n_k} - b_{n_k-1}\|_2 \\
&\leq E(u^\star) + \frac{\lambda C_1}{2(n_k - 1)} + \lambda\|Au_{n_k} - b\|_2\|Au_{n_k}\|_2 + \lambda\|Au_{n_k} - b\|_2\|b_{n_k-1}\|_2 \\
&\leq E(u^\star) + \frac{\lambda C_1}{2(n_k - 1)} + \frac{\lambda\sqrt{C_1}\|A\|\|u_{n_k}\|_2}{\sqrt{n_k - 1}} + \lambda\|Au_{n_k} - b\|_2\|b_{n_k-1}\|_2 \\
&\leq E(u^\star) + \frac{\lambda C_1}{2(n_k - 1)} + \frac{\lambda\sqrt{C_1}\|A\|C}{\sqrt{n_k - 1}} + \lambda\|Au_{n_k} - b\|_2\left(\|b_0\|_2 + \sum_{\ell=1}^{n_k} \|Au_\ell - b\|_2\right) \\
&\leq E(u^\star) + \frac{\lambda C_1}{2(n_k - 1)} + \frac{\lambda\sqrt{C_1}(\|A\|C + \|b_0\|_2)}{\sqrt{n_k - 1}} + \lambda\|Au_{n_k} - b\|_2\sum_{\ell=1}^{n_k} \|Au_\ell - b\|_2.
\end{aligned}
\tag{3.62}
$$

Suppose now that $(u_{n_k})_{k\in\mathbb{N}}$ converges to some $\tilde{u}$ as ℓ goes to infinity. Taking (3.55) into account, we have that $\|A\tilde{u} - b\|_2 = 0$, i.e., $\tilde{u}$ satisfies the constraint $A\tilde{u} = b$. On the other hand, taking limits in (3.62) and using Kronecker's Lemma, see Lemma A.2.1 in Appendix A, we have that the limit in the right hand side of (3.62) is $E(u^\star)$. Thus we have

$$E(\tilde{u}) \leq E(u^\star),$$

which means that $\tilde{u}$ is a solution of the constrained problem (3.51) and since the solution is unique then $\tilde{u} = u^\star$. Since every accumulation point of the bounded sequence $(u_n)_{n\in\mathbb{N}}$ is equal to $u^\star$, we conclude that the whole sequence converges to $u^\star$. $\qquad\square$

3.5.3 Numerical solution via the split Bregman algorithm

In this section, we explain how the Bregman iteration (3.53)–(3.54) together with a splitting technique can be used to implement numerically the minimisation of the functional (3.50). The idea originates from [GO09], where such a procedure was applied to total variation minimisation and was given the name split Bregman algorithm. This iterative technique is equivalent to certain instances of combinations of the augmented Lagrangian method with classical operator splitting such as Douglas–Rachford, see [Set09].

Recall that we want to solve the following unconstrained minimisation problem:

$$\min_{u\in\mathbb{R}^{N\times M}} \frac{1}{2}\|Tu - f\|_2^2 + \alpha\|\nabla u\|_1 + \beta\|\nabla^2 u\|_1, \tag{3.63}$$

where here for simplicity, we consider only the L^2 fidelity case. The derivation of the split Bregman algorithm for solving (3.63) starts with the observation that the above minimisation problem is equivalent to the following constrained minimisation problem

$$\min_{\substack{u\in\mathbb{R}^{N\times M} \\ v\in(\mathbb{R}^{N\times M})^2 \\ w\in(\mathbb{R}^{N\times M})^4}} \frac{1}{2}\|Tu - f\|_2^2 + \alpha\|v\|_1 + \beta\|w\|_1, \quad \text{such that } v = \nabla u, \ w = \nabla^2 u. \tag{3.64}$$

It is clear that since the discrete gradient and Hessian are linear operations, the minimisation problem (3.64) can be reformulated into the more general problem

$$\min_{z\in\mathbb{R}^L} E(z), \quad \text{such that } Az = b,$$

where $E : \mathbb{R}^L \to \mathbb{R}^+$ is convex, A is a $L \times L$ matrix and b is a vector of length L, where $L = 7NM$. It is also easy to see that the iterative scheme of the type (3.53)–(3.54) that corresponds to the constrained minimisation problem (3.64) is:

$$(u^{n+1}, v^{n+1}, w^{n+1}) = \operatorname*{argmin}_{u,v,w} \frac{1}{2}\|Tu - f\|_2^2 + \alpha\|v\|_1 + \beta\|w\|_1 + \frac{\lambda}{2}\|b_1^n + \nabla u - v\|_2^2 \tag{3.65}$$

$$+ \frac{\lambda}{2} \|b_2^n + \nabla^2 u - w\|_2^2,$$

$$b_1^{n+1} = b_1^n + \nabla u^{n+1} - v^{n+1}, \tag{3.66}$$

$$b_2^{n+1} = b_2^n + \nabla^2 u^{n+1} - w^{n+1}, \tag{3.67}$$

where $b_1^{n+1} = (b_{1,1}^{n+1}, b_{1,2}^{n+1}) \in \left(\mathbb{R}^{N \times M}\right)^2$ and $b_2^{n+1} = (b_{2,11}^{n+1}, b_{2,22}^{n+1}, b_{2,12}^{n+1}, b_{2,21}^{n+1}) \in \left(\mathbb{R}^{N \times M}\right)^4$. Notice that at least in the case where T is injective (e.g. denoising, deblurring), the minimisation problem (3.65) has a unique solution. Moreover, in that case, the functional E is coercive and the constrained minimisation problem (3.64) has a unique solution, thus, Theorem 3.5.1 holds. We also note that one can consider having two parameters λ_1 and λ_2 for the first and second order term in (3.65) respectively, as it is easily checked that this does not affect the convergence of the Bregman iteration.

Our next concern is the efficient numerical solution of the minimisation problem (3.65). We follow [GO09] and minimise with respect to u, v and w alternatingly:

Split Bregman for $\mathbf{L^2 - TV - TV^2}$ denoising, deblurring

Subproblem 1: $\displaystyle u^{n+1} = \operatorname*{argmin}_{u \in \mathbb{R}^{N \times M}} \ \frac{1}{2}\|Tu - f\|_2^2 + \frac{\lambda_1}{2}\|b_1^n + \nabla u - v^n\|_2^2$ (3.68)

$$+ \frac{\lambda_2}{2}\|b_2^n + \nabla^2 u - w^n\|_2^2,$$

Subproblem 2: $\displaystyle v^{n+1} = \operatorname*{argmin}_{v \in (\mathbb{R}^{N \times M})^2} \ \alpha\|v\|_1 + \frac{\lambda_1}{2}\|b_1^n + \nabla u^{n+1} - v\|_2^2,$ (3.69)

Subproblem 3: $\displaystyle w^{n+1} = \operatorname*{argmin}_{w \in (\mathbb{R}^{N \times M})^4} \ \beta\|w\|_1 + \frac{\lambda_2}{2}\|b_2^n + \nabla^2 u^{n+1} - w\|_2^2,$ (3.70)

Update on b_1: $b_1^{n+1} = b_1^n + \nabla u^{n+1} - v^{n+1},$ (3.71)

Update on b_2: $b_2^{n+1} = b_2^n + \nabla^2 u^{n+1} - w^{n+1}.$ (3.72)

We note that the above splitting strategy is tuned for image denoising and deblurring. As far as image inpainting is concerned an alternative splitting is required for efficient implementation. We discuss that in Section 3.7.2. The above alternating minimisation scheme, make up the split Bregman iteration that is proposed in [GO09] to solve the total variation minimisation problem as well as problems related to compress sensing. For convergence properties of the split Bregman iteration and also other splitting techniques we refer the reader to [CW06, EZC10, Set09]. In [WT10] and [YOGD08], it is noted that the Bregman iteration coincides with the augmented Lagrangian method. Minimising

alternatingly with respect to the variables in the augmented Lagrangian method results to the alternating direction method of multipliers (ADMM), see [Gab83]. Thus, split Bregman is equivalent to ADMM. In [Eck89] and [Gab83] it is shown that ADMM is equivalent to the Douglas–Rachford splitting algorithm whose convergence is guaranteed. We refer the reader to [Set09] for an interesting study in this subject.

We now discuss how we solve each of the minimisation problems (3.68)–(3.70) for the case of denoising and deblurring. Thus, we assume that T is either the identity or represents a circular convolution. We denote by $\mathcal{F}$ the two dimensional discrete Fourier transform.

Subproblem 1: We want to solve (3.68), i.e.,

$$u^{n+1} = \operatorname*{argmin}_{u \in \mathbb{R}^{N \times M}} \frac{1}{2}\|Tu - f\|_2^2 + \frac{\lambda_1}{2}\|b_1^n + \nabla u - v^n\|_2^2 + \frac{\lambda_2}{2}\|b_2^n + \nabla^2 u - w^n\|_2^2.$$

This is solved through its optimality condition. In the continuous setting that results in a fourth order linear PDE. In the discrete setting the optimality condition reads as follows:

$$
\begin{aligned}
T^*Tu^{n+1} &- \lambda_1 \left(\overleftarrow{D_x}(D_x(u^{n+1})) + \overleftarrow{D_y}(D_y(u^{n+1})) \right) \\
&+ \lambda_2 \left(\overleftarrow{D_{xx}}(D_{xx}(u^{n+1})) + \overleftarrow{D_{yy}}(D_{yy}(u^{n+1})) + 2\overleftarrow{D_{xy}}(D_{xy}(u^{n+1})) \right) \\
&= f + \lambda_1 \left(\overleftarrow{D_x}(b_{1,1}^n - v_1^n) + \overleftarrow{D_y}(b_{1,2}^n - v_2^n) \right) \\
&- \lambda_2 \left(\overleftarrow{D_{xx}}(b_{2,11}^n - w_{11}^n) + \overleftarrow{D_{yy}}(b_{2,22}^n - w_{22}^n) + 2\overleftarrow{D_{xy}}(b_{2,12}^n - w_{12}^n) \right),
\end{aligned}
\tag{3.73}
$$

where here T^* is the adjoint operator of T. Note that for the special cases we are considering here we have in fact $T = T^*$. Due to the periodic boundary conditions, the discrete differential operators act like circular convolution. The same holds for the operator T from our assumptions. Thus taking Fourier transforms in (3.73) gives

$$FD \cdot \mathcal{F}(u^{n+1}) = \mathcal{F}(R^n),$$

where

$$
\begin{aligned}
FD := \mathcal{F}(T^*) \cdot \mathcal{F}(T) &- \lambda_1 \left(\mathcal{F}(\overleftarrow{D_x}(D_x)) + \mathcal{F}(\overleftarrow{D_y}(D_y)) \right) \\
&+ \lambda_2 \left(\mathcal{F}(\overleftarrow{D_{xx}}(D_{xx})) + \mathcal{F}(\overleftarrow{D_{yy}}(D_{yy})) + \mathcal{F}(\overleftarrow{D_{xy}}(D_{xy})) \right),
\end{aligned}
$$

and R^n is the righthand side of (3.73). Here "·" denotes pointwise multiplication of matrices. Thus we have

$$u^{n+1} = \mathcal{F}^{-1}\left(\mathcal{F}(R^n)/FD\right),$$

where "/" denotes pointwise division. Note that the term FD needs to be calculated only

once at the beginning of the algorithm.

Subproblem 2: Problem (3.68) has a closed form solution too. We have

$$
\begin{aligned}
v^{n+1} &= \operatorname*{argmin}_{v\in(\mathbb{R}^{N\times M})^2} \alpha\|v\|_1 + \frac{\lambda_1}{2}\|b_1^n + \nabla u^{n+1} - v\|_2^2 \\
&= \operatorname*{argmin}_{v\in(\mathbb{R}^{N\times M})^2} \sum_{i,j} \alpha\sqrt{v_1(i,j)^2 + v_2(i,j)^2} + \frac{\lambda_1}{2}\left((b_{1,1}^n(i,j) + D_x u^{n+1}(i,j) - v_1(i,j))^2\right. \\
&\qquad\qquad \left. +(b_{1,2}^n(i,j) + D_y u^{n+1}(i,j) - v_2(i,j))^2\right) \Rightarrow \\
v^{n+1}(i,j) &= \operatorname*{argmin}_{(z_1,z_2)\in\mathbb{R}^2} \sum_{i,j} \alpha\sqrt{z_1^2 + z_2^2} + \frac{\lambda_1}{2}\left((b_{1,1}^n(i,j) + D_x u^{n+1}(i,j) - z_1)^2\right. \\
&\qquad\qquad \left. +(b_{1,2}^n(i,j) + D_y u^{n+1}(i,j) - z_2)^2\right).
\end{aligned}
$$

Setting $s^n(i,j) = (s_1^n(i,j), s_2^n(i,j)) = \left(b_{1,1}^n(i,j) + D_x u^{n+1}(i,j), b_{1,2}^n(i,j) + D_y u^{n+1}(i,j)\right) \in \mathbb{R}^2$ it is easy to check that

$$
\begin{aligned}
v_1^{n+1}(i,j) &= \max\left(|s^n(i,j)| - \frac{\alpha}{\lambda_1}, 0\right)\frac{s_1^n(i,j)}{|s^k(i,j)|}, \\
v_1^{n+1}(i,j) &= \max\left(|s^n(i,j)| - \frac{\alpha}{\lambda_1}, 0\right)\frac{s_2^n(i,j)}{|s^k(i,j)|},
\end{aligned}
$$

with the convention that $0/0 = 0$. The above operation on v is known as *shrinkage*.

Subproblem 3: We have

$$
w^{n+1} = \operatorname*{argmin}_{w\in(\mathbb{R}^{N\times M})^4} \beta\|w\|_1 + \frac{\lambda_2}{2}\|b_2^n + \nabla^2 u^{n+1} - w\|_2^2.
$$

Working in the same way as in subproblem 2 and setting

$$
\begin{aligned}
t^n(i,j) &= (t_1^n(i,j), t_2^n(i,j), t_3^n(i,j), t_3^n(i,j)) \\
&= \left(b_{2,11}^n(i,j) + D_{xx} u^{n+1}(i,j), b_{2,22}^n(i,j) + D_{yy} u^{n+1}(i,j),\right. \\
&\qquad \left. b_{2,12}^n(i,j) + D_{12} u^{n+1}(i,j), b_{2,12}^n(i,j) + D_{xy} u^{n+1}(i,j)\right) \in \mathbb{R}^4,
\end{aligned}
$$

we obtain

$$
\begin{aligned}
w_{11}^{n+1}(i,j) &= \max\left(|t^n(i,j)| - \frac{\beta}{\lambda_2}, 0\right)\frac{t_1^n(i,j)}{|t^n(i,j)|}, \\
w_{22}^{n+1}(i,j) &= \max\left(|t^n(i,j)| - \frac{\beta}{\lambda_2}, 0\right)\frac{t_2^n(i,j)}{|t^n(i,j)|}, \\
w_{12}^{n+1}(i,j) &= \max\left(|t^n(i,j)| - \frac{\beta}{\lambda_2}, 0\right)\frac{t_3^n(i,j)}{|t^n(i,j)|}.
\end{aligned}
$$

We remark here that all three subproblems (3.68)–(3.70) are solved exactly. This is in contrast with several previous works on total variation minimisation using split Bregman [GO09, Pas12a, Pas12c, Pas12b] where the subproblem of the type (3.68) is solved approximately with one iteration of Gauss–Seidel to the linear system that results from the optimality condition. There, the optimality condition leads to a second order linear PDE instead of fourth order. Even though, convergence is achieved using both methods, the FFT approach leads to a faster and a more robust convergence, see also the corresponding experimental discussion in Section 3.7.2.

Note finally that different values of λ_1 and λ_2 can result in different convergence speeds. Even though, it is not obvious a priori how to choose λ_1 and λ_2 is order to get optimal performance, this choice has to be done only once and a potential user does not have to worry about it. For the cases of denoising and deblurring we have found empirically that $\lambda_1 = \lambda_2 = 1$ give good results. For the case of inpainting, see the corresponding discussion in Section 3.7.4.

3.6 Applications to image denoising and deblurring

Image denoising

In this section we discuss the application of the TV–TV^2 approach (3.50) to image denoising, i.e., when the operator T is the identity. We have performed experiments to images that have been corrupted by Gaussian noise, thus the L^2 norm in the fidelity term is the most suitable. We compare our method with infimal convolution [CL97] solved also with a split Bregman scheme and the total generalised variation [BKP10] solved with the primal-dual method of Chambolle and Pock [CP11] as it is described in [Bre14]. We present among others, examples for $a = 0$, $\beta > 0$ and $\alpha > 0$, $\beta = 0$. Recall that for $\beta = 0$, our model corresponds to the classical ROF denoising model, while for $\alpha = 0$, it corresponds to the pure TV^2 restoration [BP10].

Our main assessment tool for the quality of the reconstructions is the *structural similarity index* SSIM [WBSS04, WB09]. The reason for this choice is that in contrast to traditional quality measures like the peak signal-to-noise ratio PSNR, the SSIM index also assesses the conservation of the structural information of the reconstructed image. A justification for the choice of SSIM as a good fidelity measure instead of the traditional PSNR can be seen in Figure 3.3. Note that a perfect reconstruction has SSIM=1. Figures 3.3(c) and 3.3(d) show the denoising results of pure TV minimisation ($\beta = 0$). Figure 3.3(c) is the one with the highest SSIM value (0.6595) while the SSIM value of Figure 3.3(d) is significantly lower (0.4955). This assessment comes in agreement with the human point of view since, even though this is subjective, one would consider Figure 3.3(c) as a better reconstruction. On the other hand, Figure 3.3(c) has slightly smaller PSNR value (14.02) than Figure 3.3(d) (14.63), which is the one with the highest PSNR value. Similar results

| (a) Clean image | (b) Noisy image | (c) Best SSIM assessment | (d) Best PSNR assessmet |
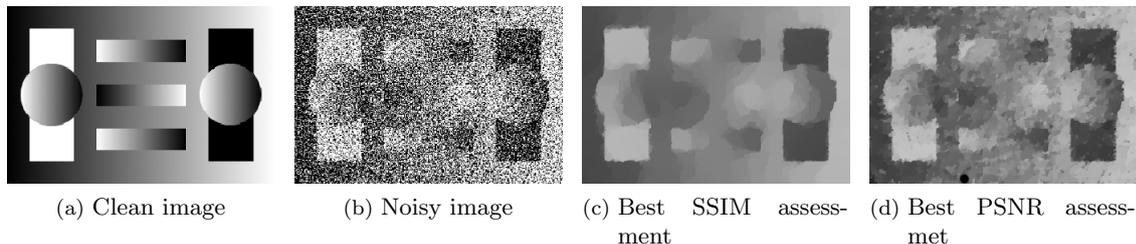
Figure 3.3: Justification for the use of SSIM as an appropriate image quality assessment tool. The initial image, Figure (a) is contaminated with Gaussian noise of variance 0.5, Figure (b). We show the best SSIM valued, Figure (c), (SSIM=0.6595, PSNR=14.02) and the best PSNR valued, Figure (d), (SSIM=0.4955, PSNR=14.63), recontructions among reconstructed images with pure TV denoising ($\beta = 0$). The better SSIM assessment agrees more with the human perception.

are obtained for $\beta > 0$.

For this section, we use a predefined number of iterations as a stopping criterium for our algorithm which in most examples is 300. In fact, we observe that after 80-100 iterations the relative residual of successive iterates is of the order of 10^{-3} or lower (see also Table 3.1) and hence no noticeable change in the iterates is observed after that. Figure 3.4 depicts one denoising example, where the original image has been corrupted by Gaussian noise of variance 0.005. For better visualisation, we include the middle row slices of all reconstructions in Figure 3.5. The highest SSIM value for TV denoising is achieved for $\alpha = 0.12$ (SSIM=0.8979) while the highest one for TV–TV^2 is achieved for $\alpha = 0.06$, $\beta = 0.03$ (SSIM=0.9081). This is slightly better than infimal convolution (SSIM=0.9053). Note, however, that this optimal combination of α and β in terms of SSIM does not always correspond to the best visual result. In general, the latter corresponds to a slightly larger β than the one chosen by SSIM see Figure 3.4(h). Still, for proof of concept, we prefer to stick with an objective quality measure and SSIM, in our opinion, is the most reliable choice for that matter. In the image of Figure 3.4(h) the staircasing effect has almost disappeared and the image is still *pleasant* to the human eye despite being slightly more blurry. This slight blur, which is the price that the method pays for the removal of the staircasing effect, can be easily and efficiently removed in post processing using simple sharpening filters, e.g. in GIMP. We did that in Figure 3.4(i), also adjusting the constrast, achieving a very good result both visually and SSIM-wise (0.9463). The overall highest SSIM value is achieved by TGV (0.9249), producing a reconstruction of very good quality. However, TGV converges slowly to the true solution. In order to check that, we compute the ground truth solution (denoted by GT) for the parameters of the TGV problem that correspond to Figure 3.4(d), by taking a large amount of iterations (2000). We check the

(a) Clean image, SSIM=1

(b) Noisy image, Gaussian noise, variance=0.005, SSIM=0.3261

(c) TV denoising, α=0.12, SSIM= 0.8979

(d) TGV denoising, SSIM=0.9249

(e) Infimal convolution denoising, SSIM=0.9053

(f) TV–TV^2 denoising, α=0.06, β=0.03, SSIM=0.9081

(g) TV^2 denoising, β=0.07, SSIM=0.8988

(h) TV–TV^2 denoising, α=0.06, β=0.06, SSIM=0.8989

(i) TV–TV^2 denoising, α=0.06, β=0.06, SSIM=0.9463, Post-processing: GIMP sharpening & contrast

Figure 3.4: Denoising of a synthetic image that has been corrupted by Gaussian noise of variance 0.005.

GPU time that is needed for the iterates to have a relative residual

$$\frac{\|u^n - \mathrm{GT}\|_2}{\|\mathrm{GT}\|_2} \leq 10^{-3},$$

and we do the same for the TV–TV^2 example of Figure 3.4(f). For TGV, it takes 1297 iterations (primal-dual method [CP11]) and 36.25 seconds while for TV–TV^2 it takes 86 split Bregman iterations and 4.05 seconds, see Table 3.1. That makes our method more suitable for cases where fast but not necessarily optimal results are needed, e.g. video processing.

In order to examine the quality of the reconstructions that are produced from each

(a) Clean image, SSIM=1

(b) Noisy image, Gaussian noise, variance=0.005, SSIM=0.3261

(c) TV denoising, α=0.12, SSIM= 0.8979

(d) TGV denoising, SSIM=0.9249

(e) Infimal convolution denoising, SSIM=0.9053

(f) TV–TV^2 denoising, α=0.06, β=0.03, SSIM=0.9081

(g) TV^2 denoising, β=0.07, SSIM=0.8988

(h) TV–TV^2 denoising, α=0.06, β=0.06, SSIM=0.8989

(i) TV–TV^2 denoising, α=0.06, β=0.06, SSIM=0.9463, Post-processing: GIMP sharpening & contrast

Figure 3.5: Corresponding middle row slices of images in Figure 3.4.

method as the number of iteration increases, we have plotted the evolution of the SSIM values in Figure 3.6. In the horizontal axis, instead of the number of iterations, we put the absolute CPU time calculated by the product of the number of iterations times the CPU time per iteration as it is seen in Table 3.1. We observe that for TGV, the SSIM value increases gradually with time, while for the methods solved with split Bregman, the image quality peaks very quickly and then remains almost constant except for TV, where the staircasing appears in the later iterations, leading eventually to a decrease of the SSIM value.

We also examine how the SSIM and PSNR values of the restored images behave as
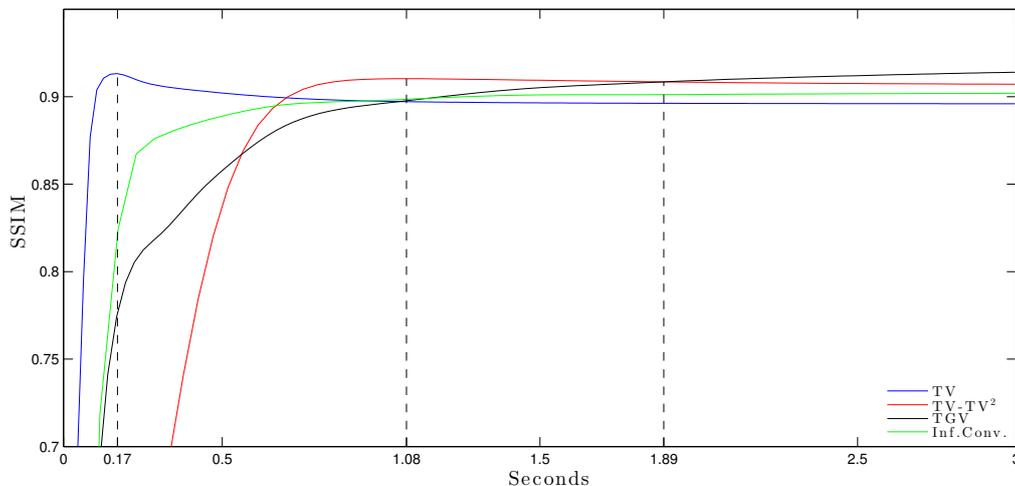
Figure 3.6: Evolution of the SSIM index with absolute CPU time for the examples of Figure 3.4. For TV denoising the SSIM value peaks after 0.17 seconds (0.9130) and after it drops when the staircasing appears, see corresponding comments on [GO09]. For TV–TV^2 the peak appears after 1.08 seconds (0.9103) and remains essentially constant. The TGV iteration starts to outperform the methods after 1.89 seconds. This shows the potential of split Bregman to produce visually satisfactory results before convergence has occurred, in contrast with the primal-dual method.

(a) SSIM values   (b) PSNR values
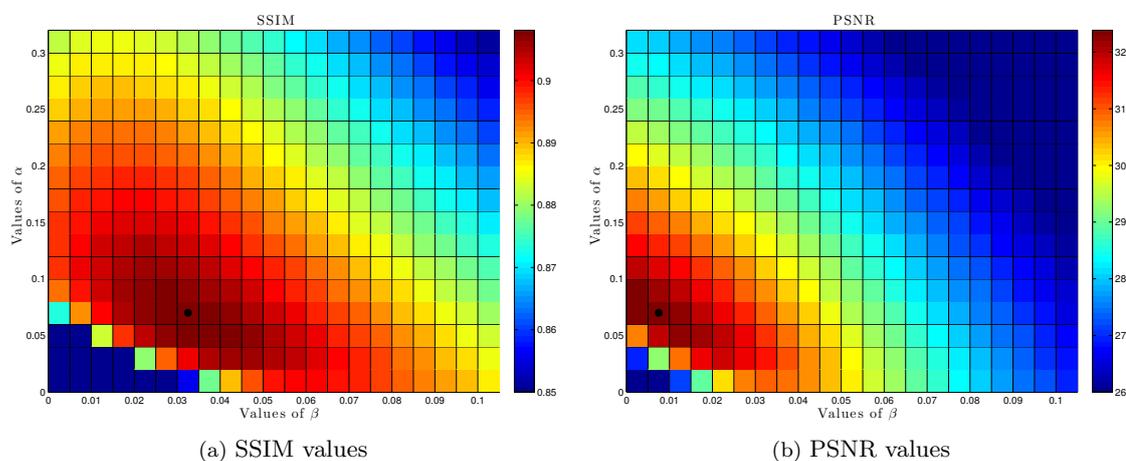
Figure 3.7: Plot of the SSIM and PSNR values of the restored image as functions of α and β, for the example of Figure 3.4. For display convenience all the values under 0.85 (SSIM) and 26 (PSNR) are colour with dark blue. The dotted cells correspond to the highest SSIM (0.9081) and PSNR (32.39) value that was achieved achieved for $\alpha = 0.06$, $\beta = 0.03$ and $\alpha = 0.06$, $\beta = 0.005$ respectively. Note that the first column in both plots corresponds to pure TV denoising, ($\beta = 0$).

87

a function of the weighting parameters α and β. In Figure 3.7 we plot these values for $\alpha = 0, 0.02, 0.04, \ldots, 0.3$ and $\beta = 0, 0.005, 0.01, \ldots, 0.1$. The plot suggests that both quality measures behave in a continuous way and they have a global maximum. However, PSNR tends to rate higher those images that have been processed with a small value of β or even with $\beta = 0$ which is not the case for SSIM. An explanation for this is that higher values of β result to a further loss of contrast, see Figure 3.8, something that is penalised by the PSNR. Note however, that the contrast can be recovered easily in a post-prosessing stage while it is not an easy task to reduce the staircasing effect using conventional processing methods.
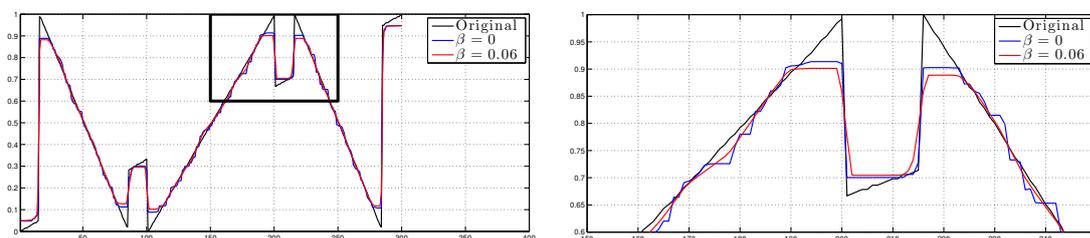
Figure 3.8: Middle row slices of reconstructed images (left) with $\alpha = 0.12$, $\beta = 0$ (blue colour) and $\alpha = 0.12$, $\beta = 0.06$ (red colour). Slices of the original image are plotted with black colour. A detail of the plot is provided on the right image. Even though the TV–TV^2 method eliminates the staircasing effect, it also results to further slight loss of contrast.

Finally in Figure 3.9 we perform denoising in an natural image which has been corrupted by Gaussian noise of variance 0.005. The staircasing of TV denoising (SSIM=0.8168) is obvious in the Figures 3.9(c) and (i). The overall best performance of the TV–TV^2 method (SSIM=0.8319) is achieved by choosing $\alpha = \beta = 0.017$, Figure 3.9(d). However, one can get satisfactory results by choosing $\alpha = \beta = 0.023$, eliminating further the staircasing without blurring the image too much, compare for example the details in Figures 3.9(j) and 3.9(k).

Image deblurring

In our deblurring implementation T denotes a circular convolution with a discrete Gaussian kernel ($\sigma = 2$, size: 11×11 pixels). The blurred image is also corrupted by additive Gaussian noise of variance 10^{-4}. Deblurring results are shown in Figure 3.10 and the corresponding middle row slices in Figure 3.11.

As in the denoising case the introduction of the second order term with a small weight β decreases noticeably the staircasing effect, compare Figures 3.10(c) and 3.10(f). Moreover, we can achieve better visual results if we increase further the value of β without blurring the image significantly, Figure 3.10(g). Infimal convolution, does not give a satisfactory

(a) Clean image, SSIM=1

(b) Noisy image, Gaussian noise, variance=0.005, SSIM=0.4436

(c) TV denoising, α=0.05, SSIM=0.8168

(d) TV–TV^2 denoising, α=0.017, β=0.017, SSIM=0.8319

(e) TV–TV^2 denoising, α=0.023, β=0.023, SSIM=0.8185

(f) TV^2 denoising, β=0.05, SSIM=0.8171

(g) Clean image, SSIM=1

(h) Noisy image, Gaussian noise, variance=0.005, SSIM=0.4436

(i) TV denoising, α=0.05, SSIM=0.8168

(j) TV–TV^2 denoising, α=0.017, β=0.017, SSIM=0.8319

(k) TV–TV^2 denoising, α=0.023, β=0.023, SSIM=0.8185

(l) TV^2 denoising, β=0.05, SSIM=0.8171

Figure 3.9: Denoising of a natural image that has been corrupted by Gaussian noise of variance 0.005.

(a) Clean image , SSIM=1

(b) Blurred and noisy image, SSIM=0.8003

(c) TV deblurring, α=0.006, SSIM=0.9680

(d) TGV deblurring, SSIM=0.9806

(e) Infimal convolution deblurring, SSIM=0.9466

(f) TV–TV^2 deblurring, α=0.004, β=0.0001, SSIM=0.9739

(g) TV–TV^2 deblurring, α=0.004, β=0.0002, SSIM=0.9710

(h) TV^2 deblurring, β=0.0012, SSIM=0.9199
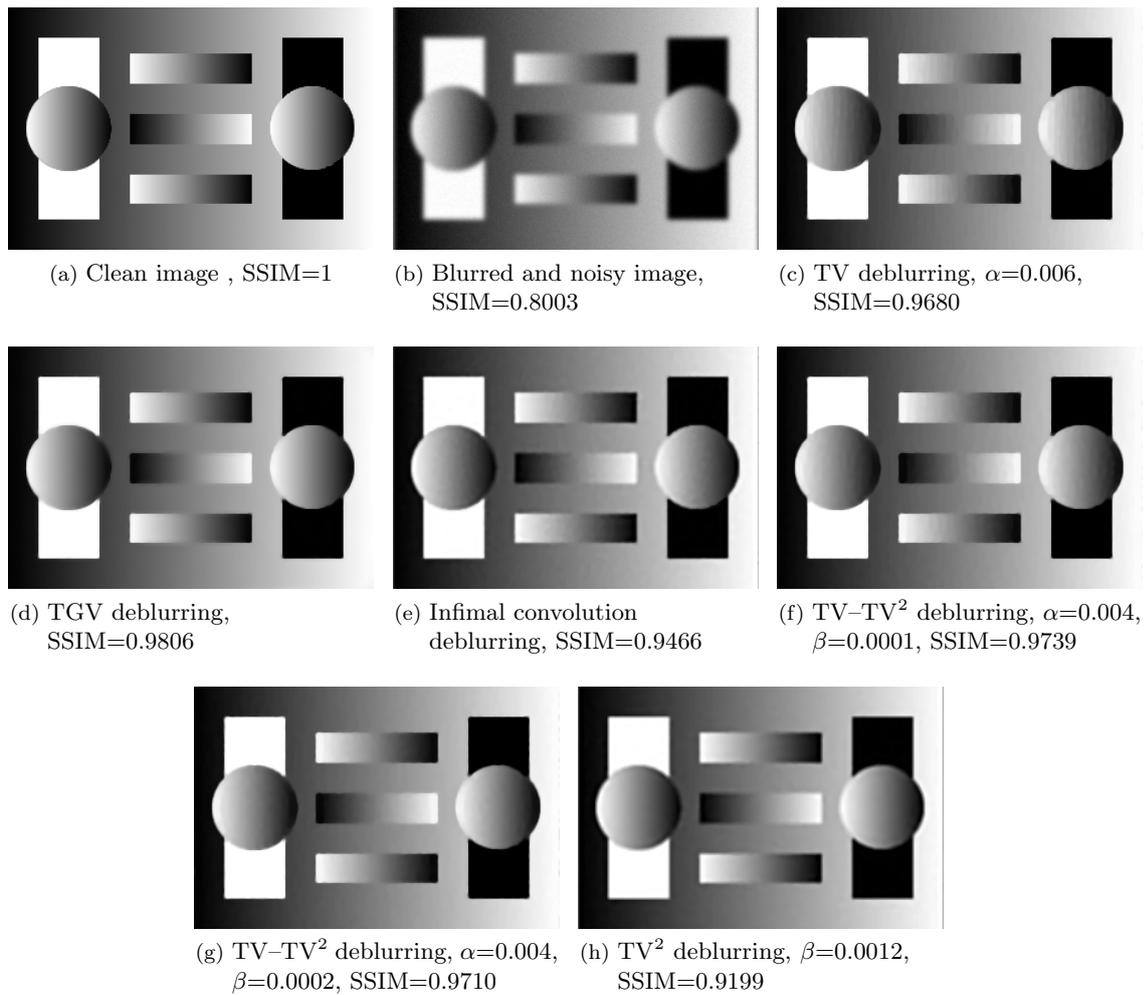
Figure 3.10: Deblurring of a blurred (Gaussian kernel of variance $\sigma = 2$) and noisy (additive Gaussian noise, variance 10^{-4}) synthetic image.

(a) Clean image , SSIM=1

(b) Blurred and noisy image, SSIM=0.8003

(c) TV deblurring, α=0.006, SSIM=0.9680

(d) TGV deblurring, SSIM=0.9806

(e) Infimal convolution deblurring, SSIM=0.9466

(f) TV–TV^2 deblurring, α=0.004, β=0.0001, SSIM=0.9739

(g) TV–TV^2 deblurring, α=0.004, β=0.0002, SSIM=0.9710

(h) TV^2 deblurring, β=0.0012, SSIM=0.9199
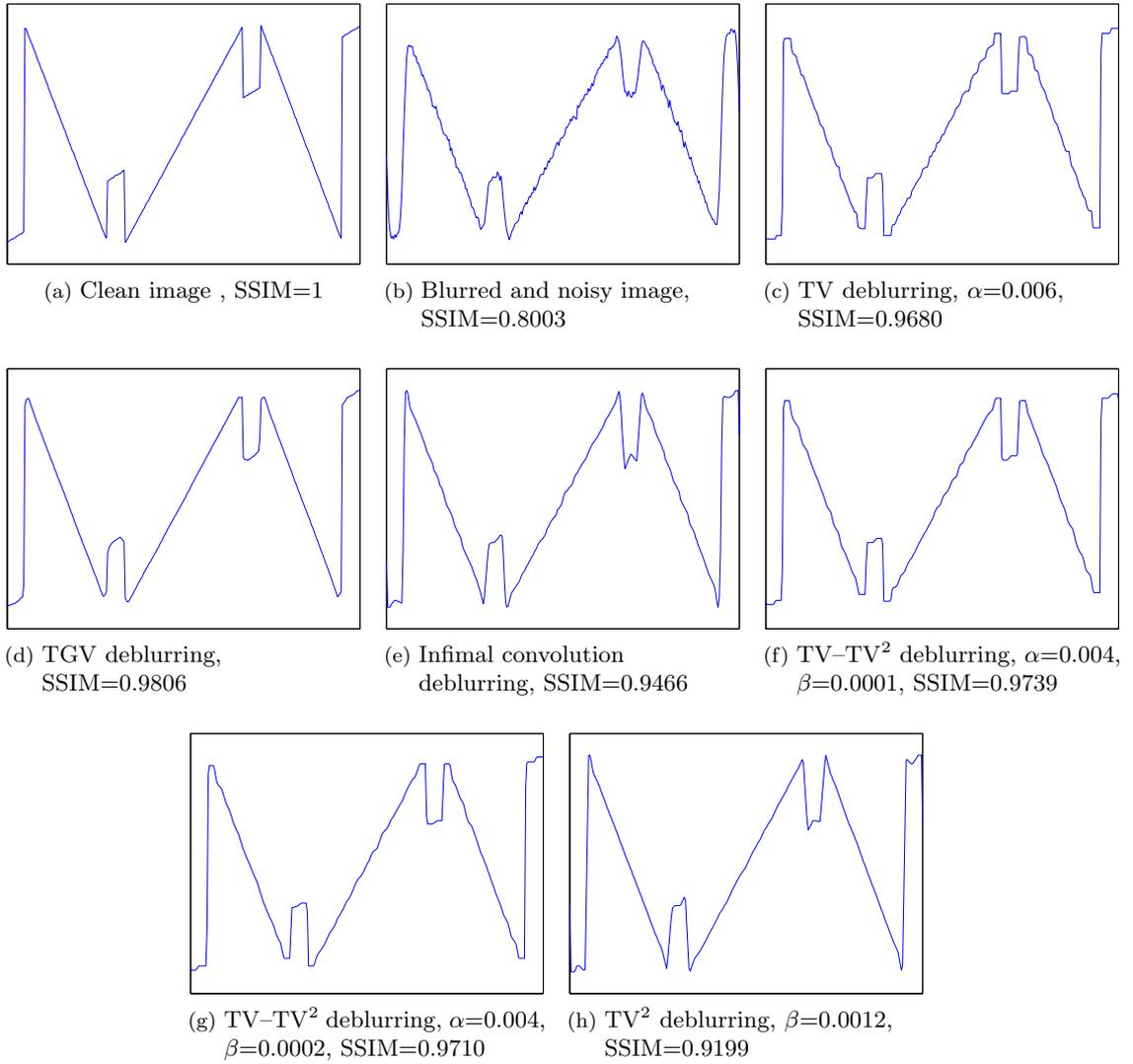
Figure 3.11: Corresponding middle row slices of images in Figure 3.10.

(a) Clean image, SSIM=1

(b) Blurred and noisy image, SSIM=0.7149

(c) TV deblurring, α=0.0007, SSIM=0.8293

(d) TV–TV^2 deblurring, α=0.0005, β=0.0001, SSIM=0.8361

(e) TV–TV^2 deblurring, α=0.0005, β=0.0003, SSIM=0.8307

(f) TV–TV^2 deblurring, α=0.0005, β=0.0003, SSIM=0.8330, Post-processing: GIMP sharpening

(g) Clean image, SSIM=1

(h) Blurred and noisy image, SSIM=0.7149

(i) TV deblurring, α=0.0007, SSIM=0.8293

(j) TV–TV^2 deblurring, α=0.0005, β=0.0001, SSIM=0.8361

(k) TV–TV^2 deblurring, α=0.0005, β=0.0003, SSIM=0.8307

(l) TV–TV^2 deblurring, α=0.0005, β=0.0003, SSIM=0.8330, Post-processing: GIMP sharpening

Figure 3.12: Deblurring of a blurred (Gaussian kernel of variance $\sigma = 2$) and noisy (additive Gaussian noise, variance 10^{-4}) natural image.

result here, Figure 3.10(e). TGV gives again the best qualitative result, Figure 3.10(d), but the computation takes about 10 minutes. Even though the time comparison is not completely fair here (the implementation described in [Bre14] does not use FFT) it takes a few thousands of iterations for TGV to deblur the image satisfactorily, in comparison with a few hundreds for the TV–TV^2 method.

In Figure 3.12 we show the performance of TV–TV^2 method for deblurring a natural image. The best result for TV–TV^2 (SSIM=0.8361) is achieved with $\alpha = 0.0005$ and $\beta = 0.0001$, Figure 3.12(d). As in the case of denoising, one can increase the value of β slightly, reducing further the staircasing effect, Figures 3.12(e) and 3.12(k). The additional blur which is a result of the larger β can be controlled using a sharpening filter, Figures 3.12(f) and 3.12(l).

Discussion

In the case of denoising and deblurring we compared our method with TGV, which is considered a state of the art method in the area of higher order image reconstruction methods in the variational context. Indeed, in both image reconstruction tasks, TGV gives better qualitative results, in terms of the SSIM index. However, the computational time that was needed to obtain the TGV result solved with the primal-dual method is significantly more than the one needed to compute the TV–TV^2 method using split Bregman, see Table 3.1. We also show that with simple and fast post-processing techniques we can obtain results comparable with TGV. For these reasons, we think that the TV–TV^2 approach is particularly interesting for applications in which the speed of computation matters. Regarding the comparison with infimal convolution, our method is slightly faster and results in better reconstructions in deblurring while in denoising the results are comparable.

3.7 Applications to image inpainting and online demo in IPOL

3.7.1 Motivation and IPOL's philosophy

In this section we present examples of the application of our TV–TV^2 approach to image inpainting. Recall that in image inpainting, the goal is to reconstruct an image inside a missing part $D \subseteq \Omega$, the inpainting domain, using the information from the intact part. The corresponding minimisation problem in the discrete setting reads as follows

$$\min_{u \in \mathbb{R}^{N \times M}} \frac{1}{2} \| \mathcal{X}_{\Omega \setminus D} (u - f) \|_2^2 + \alpha \| \nabla u \|_1 + \beta \| \nabla^2 u \|_1. \tag{3.74}$$

We note that in the inpainting task one wants to keep both values of α and β small, such that more weight is put on the fidelity term so it remains close to zero and hence

Denoising (B&W) – Image size: 200×300				
	No of iterations for GT	No of iterations for $\|u_k - GT\|_2/\|GT\|_2 \leq 10^{-3}$	CPU time (secs)	time per iteration (secs)
TV	2000	136	2.86	0.0210
TV^2	2000	107	3.62	0.0338
TV–TV^2	2000	86	4.05	0.0471
Inf.-Conv.	2000	58	3.33	0.0574
TGV	2000	1297	36.25	0.0279
Deblurring (B&W) – Image size: 200×300				
	No of iterations for GT	No of iterations for $\|u_k - GT\|_2/\|GT\|_2 \leq 10^{-2}$	CPU time (secs)	time per iteration (secs)
TV	1000	478	10.72	0.0257
TV^2	1000	108	3.64	0.0337
TV–TV^2	1000	517	25.47	0.0493
Inf.-Conv.	1000	108	7.47	0.0692
TGV	CPU time more than 10 minutes			1.22

Table 3.1: Computational times for the examples of Figures 3.4 and 3.10. We computed a ground true solution (GT) for every method by taking a large number of iterations and record the number of iterations and CPU time it takes for the relative residual between the iterates and the ground true solution to fall below a certain threshold. The TGV examples were computed using $\sigma = \tau = 0.25$ in the primal-dual method described in [Bre14]. The implementation was done using MATLAB in a Macbook 10.7.3, 2.4 GHz Intel Core 2 Duo and 2 GB of memory.

it essentially holds $u = f$ in $\Omega \setminus D$. As we are going to see later on, connectivity along large gaps in the inpainting domain is essentially achieved only with pure TV^2 inpainting ($\alpha = 0$). However, we keep on using this TV–TV^2 combination in order to keep the method flexible, regarded as a superset of pure TV and pure TV^2 inpainting. In this way, we can study the effect that each term has on the inpainting image.

As we have already mentioned, our motivation here is that the kind of regularisation given in (3.74) has the ability to connect large gaps in the inpainting domain with the price of some blur, see Figure 3.13. There, the task is to inpaint a large gap in a black stripe.

(a) Inpainting domain: grey area   (b) Harmonic inpainting   (c) TV inpainting   (d) TV^2 inpainting
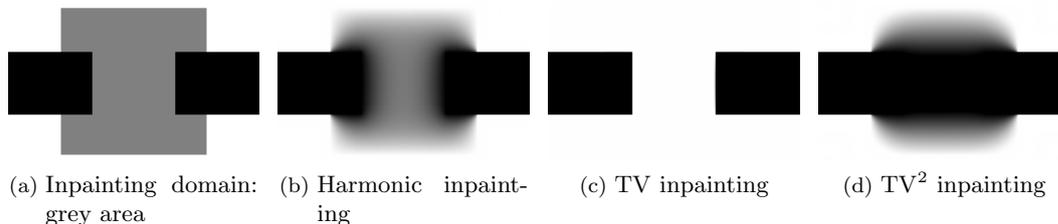
Figure 3.13: Illustration of the connectivity property of TV^2 inpainting.

Harmonic inpainting, i.e., regularising with $\|\nabla u\|_2^2$, achieves no connectivity, producing a

rather smooth result, Figure 3.13(b). Pure TV inpainting is not able to make a connection at all, Figure 3.13(c) but pure TV^2 inpainting is able to connect the large gap with the price of some blur. As we will see in Section 3.7.5, this connectivity depends also on the size and geometry of the inpainting domain.

We will next present an algorithm in order to solve (3.74). We also provide a source code for that algorithm written in C and an online demonstration in *Image Processing Online* (IPOL) accessible on the online version of [PSS13], `http://dx.doi.org/10.5201/ipol.2013.40`. IPOL is a peer reviewed online research journal specialising on image processing algorithms. Its philosophy is that each article consists of a standard text describing an image processing algorithm and also its source code, along with an online demonstration in a provided platform. Everyone has access to the code and the online demonstration and can test by themselves what each method does (open access software). In Figure 3.14 we provide two screenshots from the IPOL website that show the structure
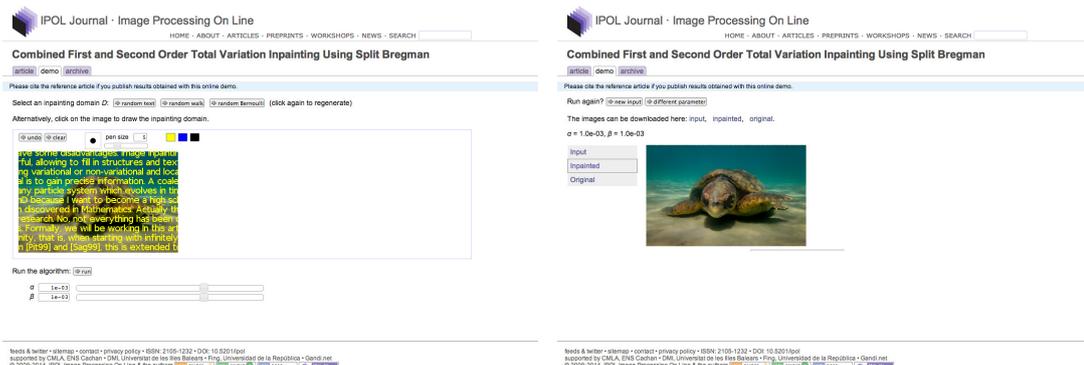
(a) Setting D and the parameters α, β.   (b) Showing the inpainting result.

Figure 3.14: IPOL's online demonstration platform.

of the online demonstration. The user can choose the parameters α and β as well as the inpainting domain D. There is also an opportunity for the users to upload their own image. Every test of the method is stored in the "Archive" section of the platform. Each algorithm should be able to compute the result in at most 30 seconds for images of resolution up to 500×500 pixels. Thus, we need a fast and efficient algorithm in order to solve (3.74) which we present in the next section.

3.7.2 Split Bregman for TV–TV^2 inpainting

The algorithm for TV–TV^2 inpainting is also based on a splitting technique but it is slightly different than the one used for denoising and deblurring. As in the latter cases, we first transform the unconstrained minimisation problem (3.74) into a constrained one, formulate the corresponding Bregman iteration algorithm and then we minimise alternatingly with

respect to all the different variables.

Note that (3.74) is equivalent to

$$\min_{\substack{u\in\mathbb{R}^{N\times M},\,\tilde{u}\in\mathbb{R}^{N\times M} \\ v\in(\mathbb{R}^{N\times M})^2,\,w\in(\mathbb{R}^{N\times M})^4}} \|\mathcal{X}_{\Omega\setminus D}(u-f)\|_2^2+\alpha\|v\|_1+\beta\|w\|_1,\text{ such that } u=\tilde{u},\ v=\nabla\tilde{u},\ w=\nabla^2\tilde{u}.$$

$$(3.75)$$

The introduction of the auxiliary variable $\tilde{u}$ is crucial since it allows the use of fast Fourier transform, leading to a fast implementation see Remark in page 97. The Bregman iteration that corresponds to (3.75) is

$$(u^{n+1},\tilde{u}^{n+1},v^{n+1},w^{n+1}) = \operatorname*{argmin}_{u,\tilde{u},v,w}\ \|\mathcal{X}_{\Omega\setminus D}(u-f)\|_2^2 + \alpha\|v\|_1 + \beta\|w\|_1$$

$$+ \frac{\lambda_0}{2}\|b_0^n + \tilde{u} - u\|_2^2 + \frac{\lambda_1}{2}\|b_1^n + \nabla\tilde{u} - v\|_2^2$$

$$+ \frac{\lambda_2}{2}\|b_2^n + \nabla^2\tilde{u} - w\|_2^2, \qquad (3.76)$$

$$b_0^{n+1} = b_0^n + \tilde{u}^{n+1} - u^{n+1},$$

$$b_1^{n+1} = b_1^n + \nabla\tilde{u}^{n+1} - v^{n+1},$$

$$b_2^{n+1} = b_2^n + \nabla^2\tilde{u}^{n+1} - w^{n+1}.$$

Solving (3.76) approximately by minimising alternatingly with respect to u, $\tilde{u}$, v and w, leads to the split Bregman algorithm for TV–TV^2 inpainting. We also refer the reader to the works by Getreuer in IPOL [Pas12a, Pas12b, Pas12c] which use the split Bregman method for TV denoising, deblurring and inpainting respectively. In our case the split Bregman formulation reads as follows:

Split Bregman for $\mathbf{L^2 - TV - TV^2}$ inpainting – Greyscale images

Subproblem 1: $\quad u^{n+1} = \operatorname*{argmin}_{u\in\mathbb{R}^{N\times M}}\ \dfrac{1}{2}\|\mathcal{X}_{\Omega\setminus D}(u-f)\|_2^2 + \dfrac{\lambda_0}{2}\|b_0^n + \tilde{u}^n - u\|_2^2$ $\qquad(3.77)$

Subproblem 2: $\quad \tilde{u}^{n+1} = \operatorname*{argmin}_{\tilde{u}\in\mathbb{R}^{N\times M}}\ \dfrac{\lambda_0}{2}\|b_0^n + \tilde{u} - u^{n+1}\|_2^2 + \dfrac{\lambda_1}{2}\|b_1^n + \nabla\tilde{u} - v^n\|_2^2$ $\quad(3.78)$

$$+ \frac{\lambda_2}{2}\|b_2^n + \nabla^2\tilde{u} - w^n\|_2^2,$$

Subproblem 3: $\quad v^{n+1} = \operatorname*{argmin}_{v\in(\mathbb{R}^{N\times M})^2}\ \alpha\|v\|_1 + \dfrac{\lambda_1}{2}\|b_1^n + \nabla\tilde{u}^{n+1} - v\|_2^2,$ $\qquad(3.79)$

Subproblem 4: $\quad w^{n+1} = \operatorname*{argmin}_{w\in(\mathbb{R}^{N\times M})^4}\ \beta\|w\|_1 + \dfrac{\lambda_2}{2}\|b_2^n + \nabla^2\tilde{u}^{n+1} - w\|_2^2,$ $\qquad(3.80)$

Update on b_0: $\quad b_0^{n+1} = b_0^n + \tilde{u}^{n+1} - u^{n+1},$ $\qquad(3.81)$

Update on b_1: $\quad b_1^{n+1} = b_1^n + \nabla\tilde{u}^{n+1} - v^{n+1},$ $\qquad(3.82)$

Update on b₂: $b_2^{n+1} = b_2^n + \nabla^2 \tilde{u}^{n+1} - w^{n+1}.$ (3.83)

Note that problem (3.78) is solved through its optimality condition, using FFT, exactly like the corresponding problem (3.68). Similarly, problems (3.79) and (3.80) are solved via shrinkage, like problems (3.69) and (3.70). Lastly, it is easy to show that problem (3.77) has a closed form solution, so it can be solved very fast as well. Indeed, we have,

$$
\begin{aligned}
u^{n+1} &= \underset{u\in\mathbb{R}^{N\times M}}{\operatorname{argmin}} \ \frac{1}{2}\|\mathcal{X}_{\Omega\backslash D}(u-f)\|_2^2 + \frac{\lambda_0}{2}\|b_0^n + \tilde{u}^n - u\|_2^2 \\
&= \underset{u\in\mathbb{R}^{N\times M}}{\operatorname{argmin}} \ \sum_{i,j}\mathcal{X}_{\Omega\backslash D}(u(i,j)-f(i,j))^2 + \frac{\lambda_0}{2}\left(b_0^n(i,j)+\tilde{u}^n(i,j)-u(i,j)\right)^2 \Rightarrow
\end{aligned}
$$

$$
u^{n+1}(i,j) = \underset{z\in\mathbb{R}}{\operatorname{argmin}} \ \mathcal{X}_{\Omega\backslash D}(z-f(i,j))^2 + \frac{\lambda_0}{2}\left(b_0^n(i,j)+\tilde{u}^n(i,j)-z\right)^2.
$$

It is trivial to calculate the solution of the last minimisation problem:

$$
u^{n+1}(i,j) = \left(\frac{2\mathcal{X}_{\Omega\backslash D}(i,j)}{\lambda_0+2}\right)f(i,j) + \left(\frac{\lambda_0+2(1-\mathcal{X}_{\Omega\backslash D}(i,j))}{\lambda_0+2}\right)(b_0^n(i,j)+\tilde{u}^n(i,j)). \ (3.84)
$$

Remark: One can notice here the utility of the auxiliary variable $\tilde{u}$. Without this variable, the minimisation problem (3.77) would be absent and the problem (3.78) would have the form

$$
u^{n+1} = \underset{u\in\mathbb{R}^{N\times M}}{\operatorname{argmin}} \ \|\mathcal{X}_{\Omega\backslash D}(u-f)\|_2^2 + \frac{\lambda_1}{2}\|b_1^n + \nabla u - v^n\|_2^2 + \frac{\lambda_2}{2}\|b_2^n + \nabla^2 u - w^n\|_2^2. \ (3.85)
$$

However, we cannot take advantage of the fast Fourier transform in order to solve the optimality condition of (3.85) since in general $\mathcal{F}(\mathcal{X}_{\Omega\backslash D}) \neq \mathcal{X}_{\Omega\backslash D}\mathcal{F}(u)$. An alternative approach would be to solve (3.85) approximately, using one or more iterations of Gauss–Seidel to the linear system that corresponds to its optimality condition as it is done in [GO09, Pas12a, Pas12c, Pas12b]. That would reduce the number of subproblems to be solved but it would have the disadvantage that one of them is not solved exactly. However, we prefer to introduce this auxiliary variable $\tilde{u}$ increasing the number of subproblems to four but being able to solve them exactly. In order to justify that, see Figure 3.15 for a quantitative comparison of the two approaches. There, we perform TV–TV² inpainting on the image of Figure 3.15(a) using both algorithms i.e., (3.68)–(3.72) using one iteration of Gauss-Seidel for (3.68) and (3.77)–(3.83) using FFT in (3.77). For each approach, we plot the relative error

$$
\frac{\|u^n - \text{GT}\|_2}{\|\text{GT}\|_2},
$$

where GT denotes a ground solution which is taken to be an iterate u^n for a large enough

(a) Original image. (b) Inpainting domain (c) Inpainted image with $\alpha = \beta = 0.001$.
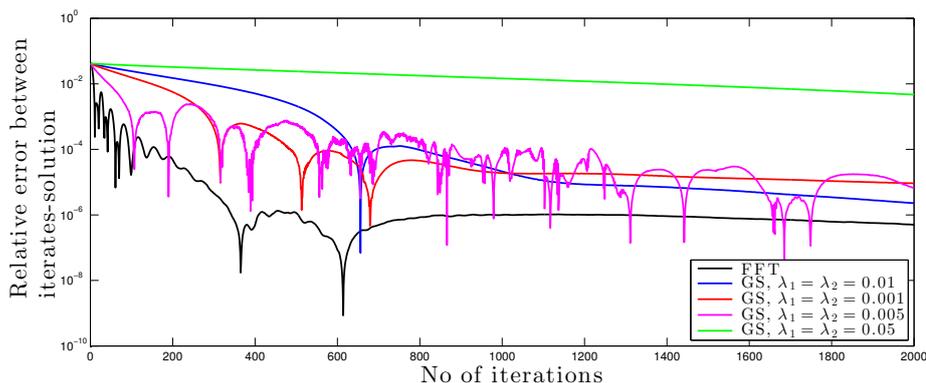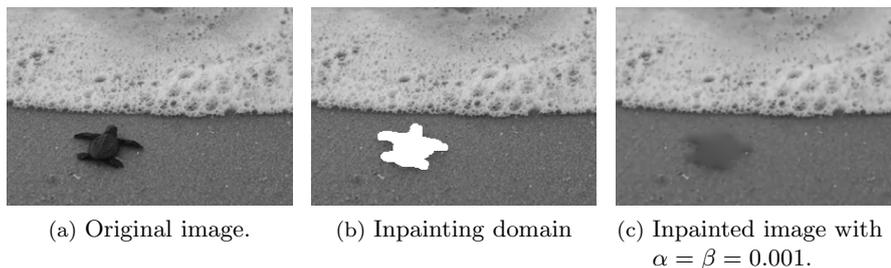
(d) Plot of the relative error $\|u_k - \mathrm{GT}\|_2/\|\mathrm{GT}\|_2$ versus the number of iterations. Here *sol* denotes a ground solution which is taken to be the solution u^n for a large enough n (3000).

Figure 3.15: Comparison of the efficiency of the use of fast Fourier transform for the solution of (3.78) over one iteration of Gauss–Seidel for the solution of (3.85).

n (3000). For the FFT approach, the choice of λ's is done according to the rule (3.105) presented in Section 3.7.4, while for the Gauss–Seidel the selection of λ's is shown in Figure 3.15(d). As Figure 3.15(d) suggests (and also observed empirically) solving the linear subproblem with FFT leads to a faster convergence. Using more than one iterations of Gauss-Seidel does not improve the situation as it increases further the computational time, see also the corresponding discussion in [GO09].

3.7.3 The colour image case

So far we have only considered greyscale images, i.e., $u \in \mathbb{R}^{n \times m}$. However it is easy to derive a version of the algorithm (3.77)–(3.83) for colour images, that is to say when $\mathbf{u} \in (\mathbb{R}^3)^{N \times M}$, where for simplicity we only consider RGB images. Starting from the continuous setting, if $\mathbf{u} = (u^1, u^2, u^3) = (u, u, u) \in \mathrm{BV}(\Omega, \mathbb{R}^3)$, then recall that the total variation of u is defined as

$$\mathrm{TV}(\mathbf{u}) = \sup \left\{ \sum_{a=1}^{3} \int_{\Omega} u^a \mathrm{div} v^a \, dx : \ v \in \mathcal{C}_c^1(\Omega, \mathbb{R}^{3 \times 2}), \ \|v\|_\infty \leq 1 \right\}. \qquad (3.86)$$

As in the scalar case, if $\mathbf{u} \in W^{1,1}(\Omega, \mathbb{R}^3)$, then

$$\mathrm{TV}(\mathbf{u}) = \int_\Omega |\nabla \mathbf{u}| \, dx. \tag{3.87}$$

Analogously we define $\mathrm{TV}^2(\mathbf{u})$. In the discrete setting the first and second order total variation have the form

$$\mathrm{TV}(\mathbf{u}) = \|(\nabla u, \nabla u, \nabla u)\|_1, \tag{3.88}$$

$$\mathrm{TV}^2(\mathbf{u}) = \|(\nabla^2 u, \nabla^2 u, \nabla^2 u)\|_1. \tag{3.89}$$

The minimisation problem corresponding to (3.74) now reads

$$\min_{\mathbf{u} \in (\mathbb{R}^3)^{N \times M}} \|\mathcal{X}_{\Omega \setminus D}(\mathbf{u} - \mathbf{f})\|_2^2 + \alpha\|\nabla \mathbf{u}\|_1 + \beta\|\nabla^2 \mathbf{u}\|_1. \tag{3.90}$$

It is not hard to check that the only difference to the minimisation algorithm of the greyscale case lies in the subproblems (3.79) and (3.80). This is the only point where the colours are "mixed". Subproblems (3.77), (3.78) and the updates (3.81)–(3.83) are done separately for each channel. This favours parallelisation of the algorithm. Thus the split Bregman algorithm for colour image inpainting reads as follows:

Split Bregman for $\mathbf{L^2 - TV - TV^2}$ inpainting – colour images

Subproblem 1: Solve (3.77) separately for $u^{n+1}, u^{n+1}, u^{n+1}$, $\tag{3.91}$

Subproblem 2: Solve (3.77) separately for $\tilde{u}^{n+1}, \tilde{u}^{n+1}, \tilde{u}^{n+1}$, $\tag{3.92}$

Subproblem 3: $\mathbf{v}^{n+1} = \underset{\mathbf{v} \in \left((\mathbb{R}^{N \times M})^2\right)^3}{\mathrm{argmin}} \alpha\|\mathbf{v}\|_1 + \dfrac{\lambda_1}{2}\|\mathbf{b}_1^n + \nabla\tilde{\mathbf{u}}^{n+1} - \mathbf{v}\|_2^2$, $\tag{3.93}$

Subproblem 4: $\mathbf{w}^{n+1} = \underset{\mathbf{w} \in \left((\mathbb{R}^{N \times M})^4\right)^3}{\mathrm{argmin}} \beta\|\mathbf{w}\|_1 + \dfrac{\lambda_2}{2}\|\mathbf{b}_2^n + \nabla^2\tilde{\mathbf{u}}^{n+1} - \mathbf{w}\|_2^2$, $\tag{3.94}$

Update on $\mathbf{b_0}$: Do update (3.81) separately for $b_0^{n+1}, b_0^{n+1}, b_0^{n+1}$, $\tag{3.95}$

Update on $\mathbf{b_1}$: Do update (3.82) separately for $b_1^{n+1}, b_1^{n+1}, b_1^{n+1}$, $\tag{3.96}$

Update on $\mathbf{b_2}$: Do update (3.83) separately for $b_2^{n+1}, b_2^{n+1}, b_2^{n+1}$. $\tag{3.97}$

As far as the solution of (3.93) is concerned, similarly as before we set

$$
\begin{aligned}
\mathbf{s}^n(i,j) &= (s_1^n(i,j), s_2^n(i,j), s_1^n(i,j), s_2^n(i,j), s_1^n(i,j), s_2^n(i,j)) \\
&= \Big(b_{1,1}^n(i,j) + D_1 \tilde{u}^{n+1}(i,j), b_{1,2}^n(i,j) + D_2 \tilde{u}^{n+1}(i,j), \\
&\qquad b_{1,1}^n(i,j) + D_1 \tilde{u}^{n+1}(i,j), b_{1,2}^n(i,j) + D_2 \tilde{u}^{k+1}(i,j), \\
&\qquad b_{1,1}^n(i,j) + D_1 \tilde{u}^{n+1}(i,j), b_{1,2}^n(i,j) + D_2 \tilde{u}^{n+1}(i,j) \Big),
\end{aligned}
$$

and we can easily compute

$$
v_x^{n+1}(i,j) = \max\left(|\mathbf{s}^n(i,j)| - \frac{\alpha}{\lambda_1}, 0 \right) \frac{s_1^n(i,j)}{|\mathbf{s}^n(i,j)|}, \tag{3.98}
$$

$$
v_y^{n+1}(i,j) = \max\left(|\mathbf{s}^n(i,j)| - \frac{\alpha}{\lambda_1}, 0 \right) \frac{s_2^n(i,j)}{|\mathbf{s}^n(i,j)|}, \tag{3.99}
$$

$$
v_x^{n+1}(i,j) = \max\left(|\mathbf{s}^n(i,j)| - \frac{\alpha}{\lambda_1}, 0 \right) \frac{s_1^n(i,j)}{|\mathbf{s}^n(i,j)|}, \tag{3.100}
$$

$$
v_y^{n+1}(i,j) = \max\left(|\mathbf{s}^n(i,j)| - \frac{\alpha}{\lambda_1}, 0 \right) \frac{s_2^n(i,j)}{|\mathbf{s}^n(i,j)|}, \tag{3.101}
$$

$$
v_x^{n+1}(i,j) = \max\left(|\mathbf{s}^n(i,j)| - \frac{\alpha}{\lambda_1}, 0 \right) \frac{s_1^n(i,j)}{|\mathbf{s}^n(i,j)|}, \tag{3.102}
$$

$$
v_y^{n+1}(i,j) = \max\left(|\mathbf{s}^n(i,j)| - \frac{\alpha}{\lambda_1}, 0 \right) \frac{s_2^n(i,j)}{|\mathbf{s}^n(i,j)|}. \tag{3.103}
$$

Analogously, we compute the solution of (3.94). We finally note that an analogous generalisation to the colour case can be done for the denoising and deblurring algorithm (3.68)–(3.72).

3.7.4 Stopping criteria and the selection of λ's

Establishing stopping criteria for our algorithm is necessary for the online implementation. After extensive experimentation, we have found empirically that a stopping criterion depending on the relative residue of the iterates u^n is an appropriate one for the termination of the algorithm. Thus, the iterations are terminated when

$$
\max\left\{ \frac{|u_k - u_{k-1}|}{|u_{k-1}|}, \frac{|u_k - u_{k-1}|}{|u_{k-1}|}, \frac{|u_k - u_{k-1}|}{|u_{k-1}|} \right\} \le 8 \cdot 10^{-5}. \tag{3.104}
$$

The values of the parameters λ_0, λ_1 and λ_2 have to be chosen carefully. This selection is done automatically in the online demo. All the combinations eventually lead to convergence but the number of iterations that are needed for convergence differ dramatically for different combinations. Note that in the case of denoising and deblurring, only two parameters have to be specified, λ_1 and λ_2 instead of three here. This makes the choice a more difficult task as they have to be chosen such that they balance each other in an

optimal way within the problems (3.77)–(3.80). Empirically we have found that λ_1 and λ_2 have to be one or two orders of magnitude larger than α and β respectively. The value of λ_0 should not be much larger than α and β as that leads to a slow convergence. Here we choose the following values of these parameters, a combination that has shown to ensure fast convergence:

$$\lambda_1 = 10\alpha, \quad \lambda_2 = 10\beta, \quad \lambda_0 = \max\{\alpha, \beta\}. \tag{3.105}$$

Table 3.2 and Figure 3.16 justify this empirical choice. In Figure 3.16, we perform pure TV^2 inpainting to the image of Figure 3.15 for different choices of λ_0 and λ_2 and fixed $\beta = 0.001$. In each case we plot the relative error between the iterates and the solution. The solution is taken to be a large iterate ($n = 4000$) that is obtained with the choice (3.105). What we generally observe can be seen in Table 3.2. The choice (3.105) leads to a fast convergence with small oscillatory behaviour in the iterates, see black line in Figure 3.16.

Choice of λ's	Convergence to solution	Qualitative behaviour
$\lambda_0 = \beta, \lambda_2 = 10\beta$	Fast	Small oscillations
$\lambda_0 = \beta, \lambda_2 \gg 10\beta$	Fairly fast	Smooth transition
$\lambda_0 = \beta, \lambda_2 \ll 10\beta$	Very slow	Wild oscillations
$\lambda_0 \gg \beta, \lambda_2 = 10\beta$	Very slow	Smooth transition
$\lambda_0 \ll \beta, \lambda_2 = 10\beta$	Slow	Smooth transition

Table 3.2: Different behaviours of the iterates for different choices of the λ parameters.

We refer the reader to the Appendix of [PSS13] for some further experimental discussion on the effect of different choices of λ's. We should note here that the stopping criterion (3.104) works well only in combination with the choices (3.105). Some choices of λ's lead to such a slow convergence that even though criterion (3.104) might be satisfied, the gap in the inpainting domain may not be filled in yet.

3.7.5 Inpainting examples

In this section we give some inpainting examples of the TV–TV^2 method. Mainly we want to emphasise the differences in practice between pure first order ($\beta = 0$) and pure second order ($\alpha = 0$) total variation inpainting.

In Figure 3.17 we see a first example where we try to remove a large font text from a natural image. The TV^2 result seems more pleasant to the human eye (subjective) as TV inpainting produces a piecewise constant reconstruction in the inpainting domain. This is confirmed quantitatively by the PSNR values inside the inpainting domain, as these are
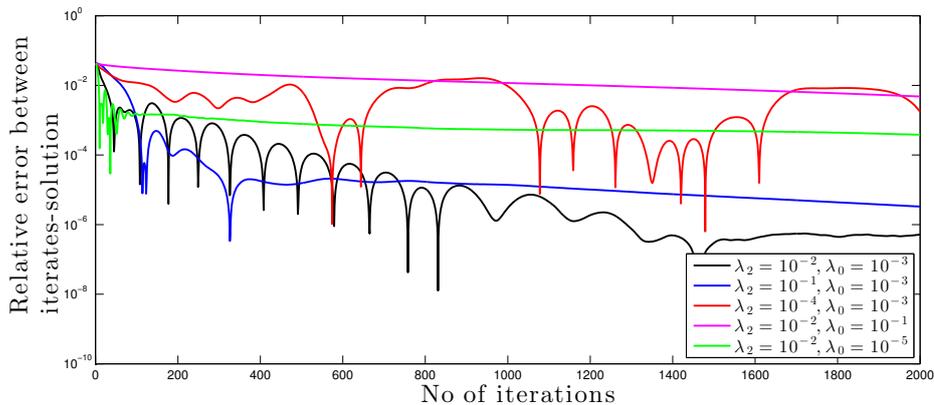
Figure 3.16: Plot of the relative error of the iterates and the solution (pure TV^2 inpainting of the image in Figure 3.15) for different choices of the parameters λ_0 and λ_2. The value of β is 0.001 and the solution is taken to be the 4000–th iterate obtained with $\lambda_0 = 0.001$, $\lambda_2 = 0.01$, see black line. We observe that other choices of λ's generally lead to a slower convergence.

(a) Original image with text   (b) Pure TV inpainting, α=0.001   (c) Pure TV^2 inpainting, β=0.001

(d) Detail of original image   (e) Detail of TV inpainting   (f) Detail of TV^2 inpainting
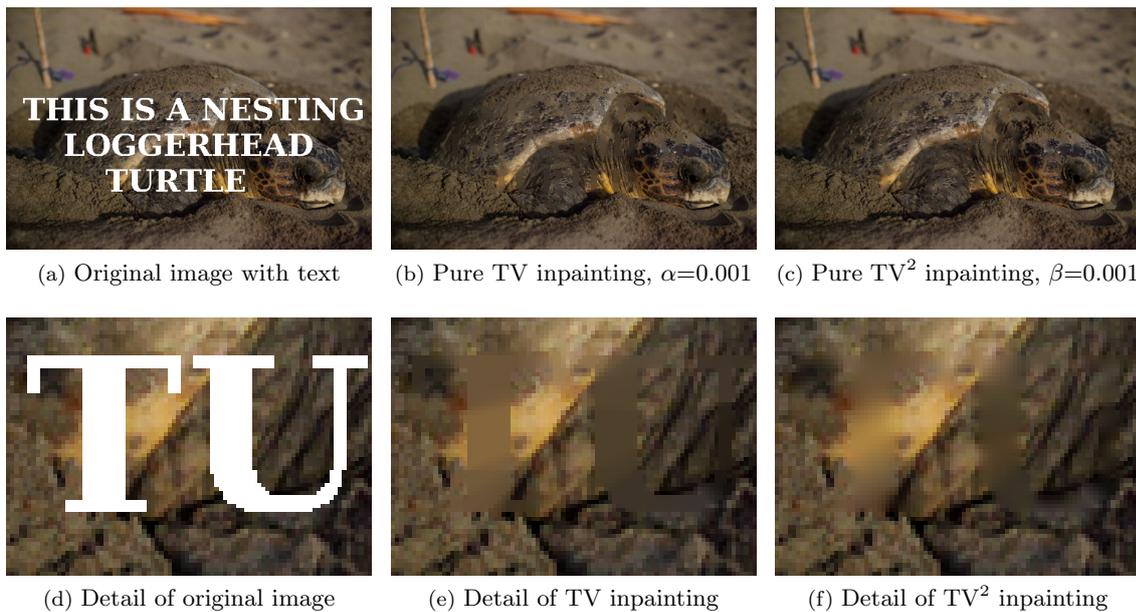
Figure 3.17: Removing large font text from an image. Even though this is subjective, we believe that the pure TV^2 result looks more realistic since the piecewise constant TV reconstruction is not desirable for natural images, see for example the inpainting of the letter "T". The PSNR inside the inpainting domain for the three channels is $(34.19, 34.97, 36.99)$ for the TV inpainting and $(34.74, 35.34, 37.20)$ for the TV^2 inpainting.

higher for the TV^2 inpainting, see the caption of Figure 3.17.

In Figure 3.18 we see a second example with colour stripes. Here we include also

(a) Original image.

(b) Harmonic inpainting, α=0.001

(c) Pure TV inpainting, α=0.001

(d) Pure TV^2 inpainting, β=0.001

(e) Detail of TV inpainting

(f) Detail of TV^2 inpainting

(g) Detail of harmonic inpainting

(h) Detail of TV^2 inpainting
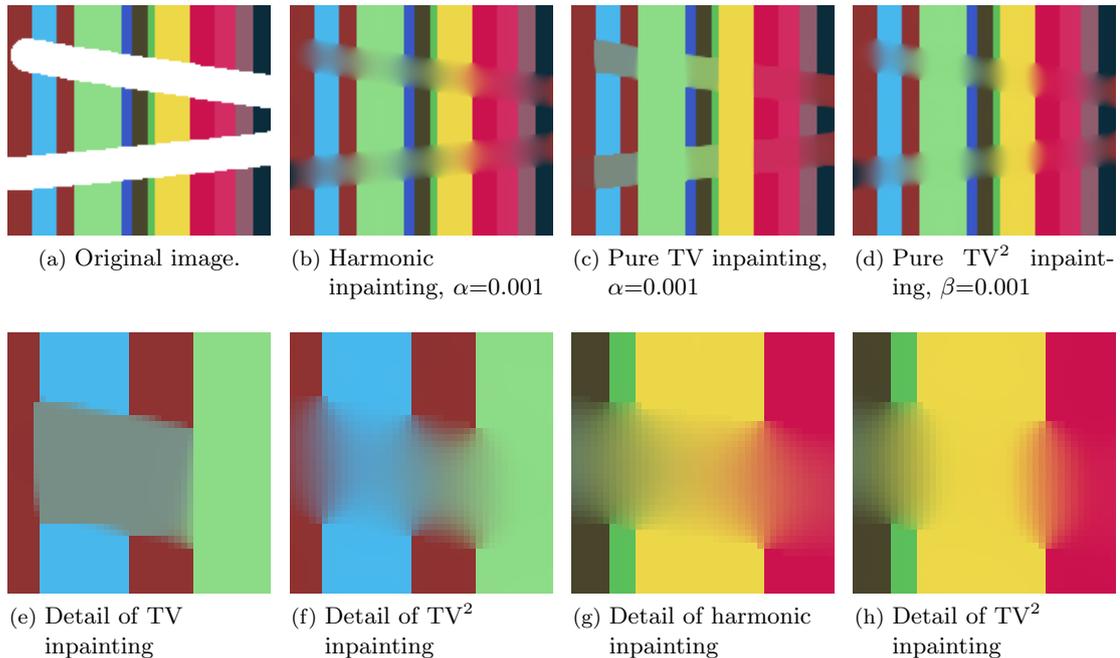
Figure 3.18: colour stripes inpainting. Observe the difference between TV and TV^2 inpainting in the light blue stripe at the left, Figures (e) and (f), where in the TV^2 case the blue colour is propagated inside the inpainting domain.

a harmonic inpainting example. As expected, in the case of TV inpainting the image is piecewise constant inside the inpainting domain, see Figure 3.18(c). This produces a desirable result in stripes whose width is larger than the size of the gap, connecting the gap and preserving the edges there, while TV^2 inpainting, Figure 3.18(d), adds some additional blur at the edge of the stripe that belongs to the inpainting domain. However, TV inpainting fails to connect the thin stripes while TV^2 propagates the right colour more reliably, see Figures 3.18(e) and 3.18(f). Finally, notice the difference between harmonic and TV^2 inpainting in terms of connectivity, as the example of the yellow stripe of Figures 3.18(g) and 3.18(h) indicates.

Let us note here that Euler's elastica is also able of achieving connection along large gaps but even its fast implementation in [THC11] is more than ten times slower than our method, see the corresponding tables in that paper.

We also want to examine how the inpainting result behaves while we are moving from pure TV to pure TV^2 inpainting in a continuous way. We can see this in Figure 3.19, where, keeping the geometry of the inpainting domain fixed, we see how connectivity is achieved while we are making the transition from pure TV to pure TV^2. The parameters α and β vary as follows:

$$\alpha : 0.01 \to 0.008 \to 0.006 \to 0.005 \to 0.004 \to 0.002 \to 0,$$

(a) Inpainting domain   (b) α=0, β=0.01   (c) α=0.008, β=0.002   (d) α=0.006, β=0.004

(e) α=0.005, β=0.005   (f) α=0.004, β=0.006   (g) α=0.002, β=0.008   (h) α=0, β=0.01
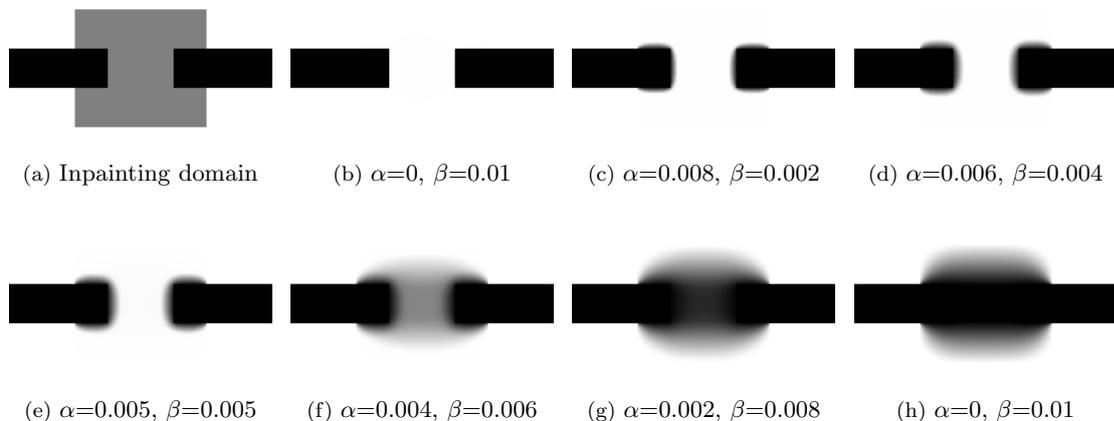
Figure 3.19: Inpainting of a stripe with a large gap. Transition from pure TV to pure TV^2. Connectivity is achieved only for large ratio β/α, see Figures (g) and (h). No connectivity is obtained when the weights α and β are equal, see Figure (e).

$$\beta : 0 \quad \rightarrow 0.002 \rightarrow 0.004 \rightarrow 0.005 \rightarrow 0.006 \rightarrow 0.008 \rightarrow 0.01.$$

As we see in Figure 3.19, the large gap in the stripe is connected only for large values of β and small values of α. In the case where both weights α and β are equal we observe no connection.

Finally we want to show the influence that the inpainting domain has on large gap connectivity. As we saw, in Figures 3.13(d) and 3.19(h), pure TV^2 inpainting has the ability to connect large gaps along the inpainting domain. However, as numerical experiments have shown, the quality of the connection depends on the geometry of the inpainting domain D, see Figures 3.20 and 3.21. In particular, for the broken stripe example of Figures 3.20 and 3.21, the domain must extend above, underneath, left and right of the gap between the two parts of the stripe. For example in Figure 3.20(e), where the inpainting domain is just the area between the two parts, we do not have connection, see Figure 3.20(j), nor we have when we extend the domain only in the vertical direction, Figures 3.21(e) and (j).

In conclusion, TV–TV^2 inpainting, especially its pure TV^2 version, is a fast inpainting method that is capable of producing visually pleasing results. Its computational cost is not much higher than the one needed for TV inpainting and its connectivity properties make it in our opinion an appealing method to use, as far as local inpainting is concerned.

(a) Domain 1   (b) Domain 2   (c) Domain 3   (d) Domain 4   (e) Domain 5

(f) Domain 1, TV^2 (g) Domain 2, TV^2 (h) Domain 3, TV^2 (i) Domain 4, TV^2 (j) Domain 5, TV^2
   inpainting   inpainting   inpainting   inpainting   inpainting
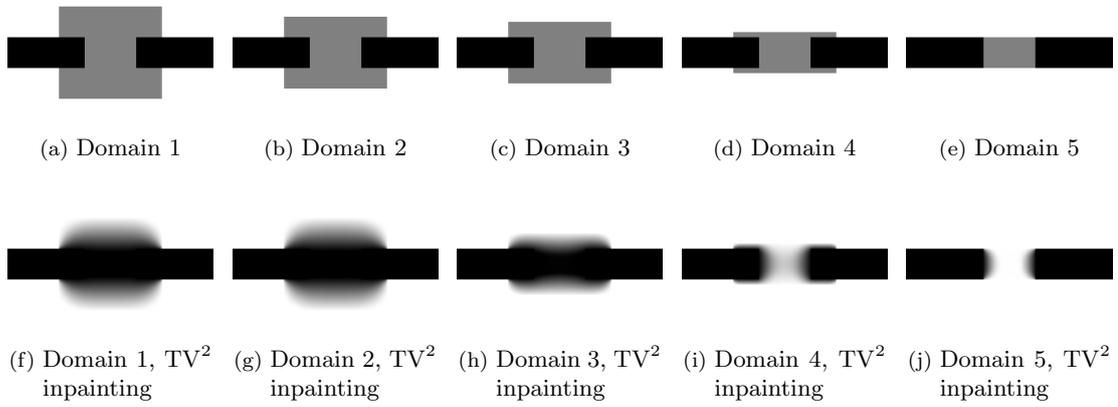
Figure 3.20: Different pure TV^2 inpainting results for different inpainting domains of decreasing height. In all computations we set $\beta = 0.001$.

(a) Domain 1   (b) Domain 6   (c) Domain 7   (d) Domain 8   (e) Domain 9

(f) Domain 1, TV^2 (g) Domain 6, TV^2 (h) Domain 7, TV^2 (i) Domain 8, TV^2 (j) Domain 9, TV^2
   inpainting   inpainting   inpainting   inpainting   inpainting
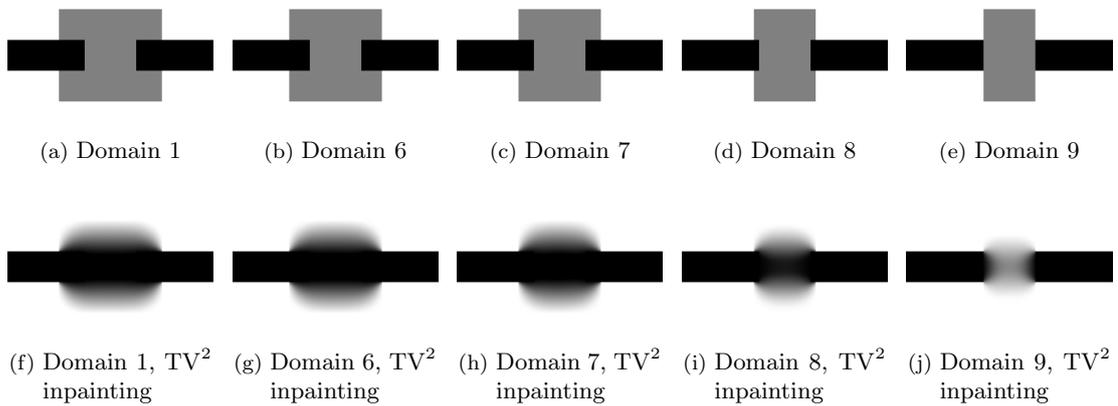
Figure 3.21: Different pure TV^2 inpainting results for different inpainting domains of decreasing width. In all computations we set $\beta = 0.001$.

Chapter 4

Exact solutions of the one dimensional TGV regularisation problem

4.1 Introduction

As we have already seen in the introduction of the thesis, the total generalised variation of second order is a high quality regulariser that incorporates second order derivatives. In its more general definition, it incorporates derivatives up to k–th order, see [BKP10]. Given a function $u \in L^1(\Omega)$ with $\Omega \subseteq \mathbb{R}^2$ being open and bounded, its total generalised variation of order k is defined as:

$$\mathrm{TGV}^k_{\boldsymbol{\alpha}}(u) = \sup \left\{ \int_\Omega u \operatorname{div}^k v \, dx : \ v \in \mathcal{C}^k_c(\Omega, \mathrm{Sym}^k(\mathbb{R}^d)), \ \|\operatorname{div}^i v\|_\infty \le \alpha_i, \ i = 0, \dots, k-1 \right\},$$

$$(4.1)$$

where $\boldsymbol{\alpha} = (\alpha_0, \dots, \alpha_{k-1})$ is a multi-index of strictly positive entries and $\mathrm{Sym}^k(\mathbb{R}^d)$ denotes the space of symmetric tensors of order k with arguments in $\mathbb{R}^d$. However, regarding its use in regularisation problems of the type

$$\min_u \Phi(Tu, f) + \mathrm{TGV}^k_{\boldsymbol{\alpha}}(u),$$

only the case $k = 2$ is typically considered. This is because of the increased computational cost to minimise $\mathrm{TGV}^k_{\boldsymbol{\alpha}}(u)$, for $k > 2$ and also due to the fact that no significant improvement to the reconstructed image is observed to justify this additional cost. Note also that, as it follows easily from the definition,

$$\mathrm{TGV}^1_\alpha(u) = \alpha \,\mathrm{TV}(u).$$

We are thus focusing on the second order total generalised variation, which can be written in a more familiar way as

$$\mathrm{TGV}^2_{\beta,\alpha}(u) = \sup \left\{ \int_\Omega u \, \mathrm{div}^2 v \, dx : \ v \in \mathcal{C}^2_c(\Omega, \mathcal{S}^{d \times d}), \ \|v\|_\infty \leq \beta, \ \|\mathrm{div} v\|_\infty \leq \alpha \right\}, \quad (4.2)$$

where α, $\beta > 0$ and $\mathcal{S}^{d \times d}$ is the space of $d \times d$ symmetric matrices.

As we mentioned in the introduction, the use of TGV in variational image reconstruction problems is becoming increasingly popular, due to the high quality of the restoration results. Like TV, it has the ability to preserve edges in the reconstructed images and in the same time it avoids the creation of the staircasing effect. Our motivation for this chapter is to gain more understanding of these properties of TGV by computing exact solutions for the one dimensional TGV denoising problem with L^2 fidelity term

$$\min_u \frac{1}{2} \int_\Omega (u - f)^2 dx + \mathrm{TGV}^2_{\beta,\alpha}(u), \quad (4.3)$$

for simple data functions f. We examine under which conditions discontinuities are preserved and what is the role of the parameters α and β. The data functions we are considering is a piecewise constant function $f_{\mathrm{p.c.}}$, a piecewise affine function $f_{\mathrm{p.a.}}$ and a hat function f_{hat}, see Figure 4.1. The rationale is that by choosing as data the functions $f_{\mathrm{p.c.}}$
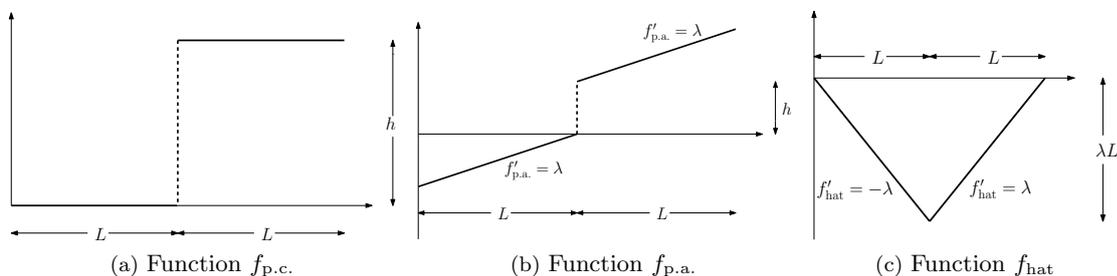
Figure 4.1: Data functions for which we compute exact solutions of the L^2–$\mathrm{TGV}^2_{\beta,\alpha}$ denoising problem.

and $f_{\mathrm{p.a.}}$ we can study the preservation of jumps and piecewise affinity in the solutions while by choosing f_{hat} we examine the effect of TGV on local extrema.

4.1.1 Organisation of the chapter

We start by describing some basic properties of the TGV functional in Section 4.2, including a useful equivalent definition that we are going to use for the rest of the chapter.

In Section 4.3 we formulate the one dimensional L^2–$\mathrm{TGV}^2_{\beta,\alpha}$ denoising problem. We identify its predual problem for which we prove well-posedness and derive the corresponding optimality conditions using Fenchel–Rockafellar duality theory, as this was described in Section 2.5.2.

Section 4.4 deals with a description of the properties of the solutions. More specifically, in Section 4.4.1, the structure of solutions is examined with emphasis on piecewise affinity, the preservation of discontinuities, behaviour near and away of the boundary of Ω and near jump points. In Section 4.4.2 we prove some interesting facts about the L^2–linear regression of the data. Among other results, we provide certain thresholds for the parameters α and β, above which, the resulting solution is the L^2–linear regression of the data f. Section 4.4.3 starts with some remarks on the inheritance of the data's symmetry to the solution. We proceed with proving the equivalence of TGV and TV regularisation, at least for even data, under some assumptions on α and β.

In Section 4.5 we extend the latter result to dimension two. In particular, we prove that for symmetric enough data functions f, the two dimensional TGV and TV regularisations coincide for large enough ratio β/α and we also provide some numerical examples that justify this result.

The computation of exact solutions for simple data functions f is done in Section 4.6. In Section 4.6.1 exact solutions are computed, taken as data a piecewise constant function with a single jump. We show that only four types of solutions are possible and we show for which combinations of the parameters α and β we have each kind of solution. Having these results as a basis, we determine straightforwardly the solutions for a piecewise affine function with a single jump, in Section 4.6.2. In Section 4.6.3 we compute exact solutions for a hat (absolute value–type of) function and also provide the corresponding combinations for α and β that lead to each type of solutions. As in the previous case, only four different types of solutions can occur.

Finally, in Section 4.6.4 we provide some numerical experiments. We show that the exact solutions coincide with the numerical ones in the absence of noise, as expected. We also numerically compute some solutions with noisy data, in which we observe only a slight deviation from the corresponding solutions with clean data.

4.2   Basic properties of TGV

In this section we recall some of the basic properties of the TGV functional as these were proved in [BKP10, BV11, BKV13].

Proposition 4.2.1. *Let* $\Omega \subseteq \mathbb{R}^d$ *be open and bounded. Then the following facts hold:*

(i) $\mathrm{TGV}^2_{\beta,\alpha}$ *is a seminorm on the Banach space*

$$\mathrm{BGV}^2_{\beta,\alpha}(\Omega) = \left\{ u \in L^1(\Omega) : \ \mathrm{TGV}^2_{\beta,\alpha}(u) < \infty \right\},$$

where $\|u\|_{\mathrm{BGV}^2_{\beta,\alpha}(\Omega)} := \|u\|_{L^1(\Omega)} + \mathrm{TGV}^2_{\beta,\alpha}(u).$

(ii) $\mathrm{TGV}^2_{\beta,\alpha}(u) = 0$ *if and only of* u *is a polynomial of degree less than* 2.

(iii) $\text{TGV}^2_{\beta,\alpha}$ *is rotationally invariant.*

(iv) $\text{TGV}^2_{\beta,\alpha}$ *is a proper, convex, lower semicontinuous functional on every $L^p(\Omega)$, $1 \le p < \infty$.*

(v) *If Ω has Lipschitz boundary then there exist constants $0 < c < C$ that depend only on Ω, such that for every u with $\text{TGV}^2_{\beta,\alpha}(u) < \infty$ we have*

$$c\|u\|_{\text{BV}(\Omega)} \le \|u\|_{L^1(\Omega)} + \text{TGV}^2_{\beta,\alpha}(u) \le C\|u\|_{\text{BV}(\Omega)}.$$

The statement (v) of Proposition 4.2.1 actually says that $\text{BGV}^2_{\beta,\alpha}(\Omega)$ and $\text{BV}(\Omega)$ are topologically equivalent Banach spaces. Thus, the minimisation of the L^2–$\text{TGV}^2_{\beta,\alpha}$ problem is done over $\text{BV}(\Omega)$. We now present an alternative definition of $\text{TGV}^2_{\beta,\alpha}$ that was proved in [BKV13].

Proposition 4.2.2. *Let $u \in L^1(\Omega)$. Then*

$$\text{TGV}^2_{\beta,\alpha}(u) = \min_{w\in\text{BD}(\Omega)} \alpha\|Du - w\|_{\mathcal{M}} + \beta\|\mathcal{E}u\|_{\mathcal{M}}, \tag{4.4}$$

where $\text{BD}(\Omega)$ is the space of functions of bounded deformation, i.e., the space of all functions in $L^1(\Omega, \mathbb{R}^d)$ such that the symmetrised gradient $\mathcal{E}u$ can be represented by a finite Radon measure. In one dimension we have $\text{BD}(\Omega) = \text{BV}(\Omega)$ and $\mathcal{E}(u) = Du$, thus (4.4) can be written as

$$\text{TGV}^2_{\beta,\alpha}(u) = \min_{w\in\text{BV}(\Omega)} \alpha\|Du - w\|_{\mathcal{M}} + \beta\|Du\|_{\mathcal{M}}. \tag{4.5}$$

As we have mentioned in the introduction, (4.4) is called the *minimum* or *primal* definition of $\text{TGV}^2_{\beta,\alpha}$ and it is the one that we are going to use for our purposes, while (4.2) is called the *supremum* or *predual* definition.

4.3 Formulation of the problem and optimality conditions

We are interested in studying the one dimensional second order TGV denoising problem with L^2 fidelity term, i.e.,

$$\min_{u\in\text{BV}(\Omega)} \frac{1}{2}\int_\Omega (u-f)^2 dx + \text{TGV}^2_{\beta,\alpha}(u), \tag{4.6}$$

where $\Omega = (a, b)$ is an open interval of $\mathbb{R}$ and $f \in L^2(\Omega)$. Note that existence of solution for (4.6) follows from a simple application of the direct method of calculus of variations using the weak* compactness in $\text{BV}(\Omega)$ and the lower semicontinuity of TGV while uniqueness follows from the strict convexity of (4.6). Using the formulation (4.5), the problem (4.6)

is equivalent to

$$\min_{\substack{u \in \mathrm{BV}(\Omega) \\ w \in \mathrm{BV}(\Omega)}} \frac{1}{2} \int_\Omega (u-f)^2 dx + \alpha \|Du - w\|_{\mathcal{M}} + \beta \|Dw\|_{\mathcal{M}}. \tag{$\mathcal{P}$}$$

Let us note here that while a solution (w, u) for $\mathcal{P}$ exists and uniqueness is guaranteed for u, the same is not true for w. In fact w is a solution to an L^1–TV problem, which is not strictly convex:

$$w \in \operatorname*{argmin}_{w \in \mathrm{BV}(\Omega)} \alpha \|Du - w\|_{\mathcal{M}} + \beta \|Dw\|_{\mathcal{M}} \iff$$

$$w \in \operatorname*{argmin}_{w \in \mathrm{BV}(\Omega)} \|D^a u + D^s u - w\|_{\mathcal{M}} + \frac{\beta}{\alpha} |Dw|(\Omega) \iff$$

$$w \in \operatorname*{argmin}_{w \in \mathrm{BV}(\Omega)} \int_\Omega |u' - w| \, dx + \frac{\beta}{\alpha} \mathrm{TV}(w).$$

In order to study exact solutions of the problem $\mathcal{P}$, we essentially follow [BKV13] where the corresponding L^1 fidelity term case is studied. We identify the predual problem of $\mathcal{P}$ and derive the optimality conditions using Fenchel–Rockafellar duality.

Consider the following problem

$$\sup \left\{ \int_\Omega f v'' dx - \frac{1}{2} \int_\Omega (v'')^2 dx : \ v \in H_0^2(\Omega), \ \|v\|_\infty \le \beta, \ \|v'\|_\infty \le \alpha \right\}. \tag{$\mathcal{P}'$}$$

We shall prove that $\mathcal{P}'$ is the predual problem of $\mathcal{P}$. Firstly, we show that $\mathcal{P}'$ has a solution indeed.

Proposition 4.3.1. *The problem $\mathcal{P}'$ admits a unique solution in $H_0^2(\Omega)$.*

Proof. Because of the estimate

$$\|v\|_\infty + \|v'\|_\infty \le C \|v\|_{H_0^2(\Omega)}, \quad \forall v \in H_0^2(\Omega),$$

see [Eva10], we have that the set

$$K = \left\{ v \in H_0^2(\Omega) : \ \|v\|_\infty \le \beta, \ \|v'\|_\infty \le \alpha \right\},$$

is norm-closed and since it is convex, it is also weakly closed from Mazur's theorem. Since, according to Proposition 2.5.2, $(\frac{1}{2}\| \cdot \|_{L^2(\Omega)}^2)^* = \frac{1}{2}\| \cdot \|_{L^2(\Omega)}^2$, we have that

$$\sup_{v \in L^2(\Omega)} \int_\Omega f v \, dx - \frac{1}{2} \int_\Omega v^2 \, dx = \left(\frac{1}{2}\| \cdot \|_{L^2(\Omega)}^2 \right)^* (f) = \frac{1}{2}\|f\|_{L^2(\Omega)}^2.$$

Thus the supremum in $\mathcal{P}'$ is finite and we denote it with $\sup \mathcal{P}'$. Consider now a maximising

sequence $(v_n)_{n \in \mathbb{N}}$ such that

$$\int_\Omega f v_n'' dx - \frac{1}{2} \int_\Omega (v_n'')^2 dx \to \sup \mathcal{P}', \quad \text{as } n \to \infty.$$

Then there exists a positive constant M such that

$$\left| \int_\Omega f v_n'' \, dx - \frac{1}{2} \int_\Omega (v_n'')^2 \, dx \right| \leq M, \quad \forall n \in \mathbb{N}. \tag{4.7}$$

We will show that the sequence $\left(\|v_n''\|_{L^2(\Omega)} \right)_{n \in \mathbb{N}}$ is bounded as well. Suppose not, then $\limsup_{n \to \infty} \int_\Omega (v_n'')^2 \, dx = \infty$. Thus, passing to a subsequence if necessary we have

$$\int_\Omega f v_n'' \, dx - \frac{1}{2} \int_\Omega (v_n'')^2 \, dx \leq \|f\|_{L^2(\Omega)} \|v_n''\|_{L^2(\Omega)} - \frac{1}{2} \|v_n''\|_{L^2(\Omega)}^2 \to -\infty, \quad \text{as } n \to \infty.$$

which is a contradiction from (4.7). Hence, the sequence $(v_n)_{n \in \mathbb{N}}$ is bounded in $H_0^2(\Omega)$ and since that space is reflexive we get the existence of a subsequence $(v_{n_k})_{k \in \mathbb{N}}$ converging weakly to a function $v \in H_0^2(\Omega)$. Since K is weakly closed we have that $v \in K$. Moreover the maximising functional is weakly upper semicontinuous since $\int_\Omega f(\cdot)'' \, dx$ and $-\frac{1}{2} \|\cdot\|_{L^2(\Omega)}^2$ are weakly continuous and weakly upper semicontinuous in $H_0^2(\Omega)$ respectively. Thus, we have

$$\sup \mathcal{P}' \geq \int_\Omega f v'' \, dx - \frac{1}{2} \int_\Omega (v'')^2 dx \geq \limsup_{k \to \infty} \int_\Omega f v_{n_k}'' dx - \frac{1}{2} \int_\Omega (v_{n_k}'')^2 dx = \sup \mathcal{P}',$$

which means that v is a solution to $\mathcal{P}'$. This solution is unique as the maximising functional is strictly concave, defined on a convex domain. $\qquad \square$

We define now the following spaces and operators:

- $X = H_0^2(\Omega) \times H_0^1(\Omega), \quad Y = H_0^1(\Omega) \times L^2(\Omega),$

- $\Lambda : X \to Y, \quad$ linear, bounded,

- $F_1 : X \to \overline{\mathbb{R}}, \quad F_2 : Y \to \overline{\mathbb{R}},$

with

$$\Lambda(v, \omega) = (\omega + v', \omega'),$$
$$F_1(v, \omega) = \mathcal{I}_{\{\|\cdot\|_\infty \leq \beta\}}(v) + \mathcal{I}_{\{\|\cdot\|_\infty \leq \alpha\}}(\omega),$$
$$F_2(\phi, \psi) = \mathcal{I}_{\{0\}}(\phi) + \int_\Omega f\psi \, dx + \frac{1}{2} \int_\Omega \psi^2 \, dx.$$

It is easy to check that under these definitions the problem $\mathcal{P}'$ is a equivalent to

$$- \min_{(v, \omega) \in X} F_1(v, \omega) + F_2(\Lambda(v, \omega)). \tag{4.8}$$

The dual problem of (4.8), see Section 2.5.2, is

$$\min_{(w,u)\in Y^*} F_1^*(-\Lambda^*(w,u)) + F_2^*(w,u). \tag{4.9}$$

Moreover, X, Y are Banach spaces, F_1, F_2 are proper, lower semicontinuous functions and the condition

$$Y = \bigcup_{\lambda \geq 0} \lambda(\mathrm{dom}(F_2) - \Lambda(\mathrm{dom}(F_1))), \tag{4.10}$$

holds. Thus we have that Theorem 2.5.6 holds, i.e., no duality gap occurs:

$$\left(\min_{(v,\omega)\in X} F_1(v,\omega) + F_2(\Lambda(v,\omega))\right) + \left(\min_{(w,u)\in Y^*} F_1^*(-\Lambda^*(w,u)) + F_2^*(w,u)\right) = 0.$$

The fact that condition (4.10) holds follows from [BKV13, Proposition 3.6], where the same condition is proved for the same X, Y, F_1, Λ and for a F_2 with a smaller domain.

Our next step is to identify the dual problem (4.9) with the problem $\mathcal{P}$.

Proposition 4.3.2. *The problem*

$$\min_{(w,u)\in Y^*} F_1^*(-\Lambda^*(w,u)) + F_2^*(w,u), \tag{4.11}$$

is equivalent to $\mathcal{P}$ in the sense that (w,u) solve (4.11) if and only if $(w,u) \in \mathrm{BV}(\Omega) \times \mathrm{BV}(\Omega)$ and solve $\mathcal{P}$.

Proof. The proof follows closely the proof of the corresponding theorem in [BKV13]. Firstly, we compute F_1^*. For a pair $(\sigma, \tau) \in H_0^2(\Omega)^* \times H_0^1(\Omega)^*$ we have that

$$F_1^*(\sigma, \tau) = \sup_{\substack{(v,\omega)\in H_0^2(\Omega)\times H_0^1(\Omega) \\ \|v\|_\infty \leq \beta \\ \|\omega\|_\infty \leq \alpha}} \langle \sigma, v \rangle + \langle \tau, \omega \rangle = \beta \sup_{\substack{v\in H_0^2(\Omega) \\ \|v\|_\infty \leq 1}} \langle \sigma, v \rangle + \alpha \sup_{\substack{\omega\in H_0^1(\Omega) \\ \|\omega\|_\infty \leq 1}} \langle \tau, \omega \rangle.$$

Using density arguments, it can be checked that

$$F_1^*(\sigma, \tau) = \beta \sup_{\substack{v\in \mathcal{C}_c^\infty(\Omega) \\ \|v\|_\infty \leq 1}} \langle \sigma, v \rangle + \alpha \sup_{\substack{\omega\in \mathcal{C}_c^\infty(\Omega) \\ \|\omega\|_\infty \leq 1}} \langle \tau, \omega \rangle. \tag{4.12}$$

We also have for every $(w,u) \in Y^*$ and $(v,\omega) \in X$,

$$-\langle \underbrace{\Lambda^*(w,u)}_{\in Y^*}, \underbrace{(v,\omega)}_{\in X} \rangle = -\langle (w,u), \Lambda(v,\omega) \rangle = -\langle w, v' + \omega \rangle - \langle u, \omega' \rangle. \tag{4.13}$$

$\underbrace{}_{\in X^*}$

Combining (4.12) and (4.13) we have that for every $(w, u) \in Y^*$

$$F_1^*(-(\Lambda^*(w, u))) = \beta \sup_{\substack{v \in \mathcal{C}_c^\infty(\Omega) \\ \|v\|_\infty \leq 1}} -\langle w, v' \rangle + \alpha \sup_{\substack{\omega \in \mathcal{C}_c^\infty(\Omega) \\ \|\omega\|_\infty \leq 1}} -\langle w, \omega \rangle - \langle u, \omega' \rangle. \tag{4.14}$$

Since $w \in H_0^1(\Omega)^*$ and $u \in L^2(\Omega)$, they can be regarded as distributions and thus (4.14) can be written as

$$F_1^*(-(\Lambda^*(w, u))) = \beta \sup_{\substack{v \in \mathcal{C}_c^\infty(\Omega) \\ \|v\|_\infty \leq 1}} \langle Dw, v \rangle + \alpha \sup_{\substack{\omega \in \mathcal{C}_c^\infty(\Omega) \\ \|\omega\|_\infty \leq 1}} \langle Du - w, \omega \rangle. \tag{4.15}$$

Since we consider the minimisation problem (4.11), both terms in (4.15) are finite. Thus, see Theorem 2.2.3 and the discussion after that, we have that Dw is a Radon measure, w is an L^1 function (thus a BV function) and Du is a Radon measure, i.e., u is a BV function as well. Thus, if (w, u) is a pair of minimisers for (4.11) we have that $(w, u) \in \mathrm{BV}(\Omega) \times \mathrm{BV}(\Omega)$ and

$$F_1^*(-(\Lambda^*(w, u))) = \beta \|Dw\|_{\mathcal{M}} + \alpha \|Du - w\|_{\mathcal{M}}.$$

We now compute F_2^*. We have

$$\begin{aligned} F_2^*(w, u) &= \sup_{\substack{(\phi, \psi) \in Y \\ \phi = 0}} \langle w, \phi \rangle + \int_\Omega u \psi \, dx - \int_\Omega f \psi \, dx - \frac{1}{2} \int_\Omega \psi^2 \, dx \\ &= \sup_{\psi \in L^2(\Omega)} \int_\Omega (u - f) \psi \, dx - \frac{1}{2} \int_\Omega \psi^2 \, dx \\ &= \left(\frac{1}{2} \| \cdot \|_{L^2(\Omega)}^2 \right)^* (u - f) \\ &= \frac{1}{2} \int_\Omega (u - f)^2 \, dx, \end{aligned}$$

and the proof is complete. $\qquad\qquad\square$

We are now ready to derive the optimality conditions that link the solutions of the problems $\mathcal{P}$ and $\mathcal{P}'$. We first need the following definition.

Definition 4.3.3. *Let $\mu \in \mathcal{M}(\Omega)$. We define the set-valued sign, $\mathrm{Sgn}(\mu)$ as*

$$\mathrm{Sgn}(\mu) = \{v \in L^\infty(\Omega) \cap L^\infty(\Omega, |\mu|) : \|v\|_{L^\infty(\Omega)} \leq 1, \ v = \mathrm{sgn}(\mu), \ |\mu| - a.e.\}$$

In other words $\mathrm{Sgn}(\mu)$ is the set of all functions v that are equal to the Radon–Nikodým density $\mu/|\mu|$, μ–almost everywhere with the additional property that $|v(x)| \leq 1$, Lebesgue–almost everywhere. The following lemma was proved in [BKV13].

Lemma 4.3.4. *If $\mu \in \mathcal{M}(\Omega)$ then*

$$\mathrm{Sgn}(\mu) \cap \mathcal{C}_0(\Omega) = \partial \|\cdot\|_{\mathcal{M}}(\mu) \cap \mathcal{C}_0(\Omega).$$

The following proposition characterises the minimisers of the problem $\mathcal{P}$.

Proposition 4.3.5 (Optimality conditions for L^2–$\mathrm{TGV}^2_{\beta,\alpha}$). *A pair $(w, u) \in \mathrm{BV}(\Omega) \times \mathrm{BV}(\Omega)$ is a minimiser for $\mathcal{P}$ if and only if there exists a function $v \in H^2_0(\Omega)$, such that*

$$v'' = f - u, \tag{C_f}$$

$$-v' \in \alpha \, \mathrm{Sgn}(Du - w), \tag{C_α}$$

$$v \in \beta \, \mathrm{Sgn}(Dw). \tag{C_β}$$

Proof. Since there is no duality gap, from Theorem 2.5.5 we get that the solutions (v, ω) and (w, u) of $\mathcal{P}'$ and $\mathcal{P}$ respectively (or equivalently (4.8) and (4.9)) are linked through the optimality conditions

$$(v, \omega) \in \partial F_1^*(-\Lambda^*(w, u)), \tag{4.16}$$

$$\Lambda(v, \omega) \in \partial F_2^*(w, u). \tag{4.17}$$

Conversely, (w, u) is a solution for $\mathcal{P}$ if there exists a $v \in H^2_0(\Omega) \subseteq \mathcal{C}_0(\Omega)$ such that (4.16)–(4.17) hold with $\omega = -v'$. We have that the first optimality condition (4.16) is equivalent to

$$F_1^*(-\Lambda^*(w, u)) + \langle (v, \omega), (\sigma, \tau) + \Lambda^*(w, u) \rangle \leq F_1^*(\sigma, \tau), \quad \forall (\sigma, \tau) \in X^*$$

$$\Leftrightarrow \alpha \|Du - w\|_{\mathcal{M}} + \beta \|Dw\|_{\mathcal{M}} + \langle (v, \omega), (\sigma - Dw, \tau - (Du - w)) \rangle \leq \alpha \|\tau\|_{\mathcal{M}} + \beta \|\sigma\|_{\mathcal{M}},$$
$$\forall (\sigma, \tau) \in X^*$$

$$\Leftrightarrow \alpha \|Du - w\|_{\mathcal{M}} + \beta \|Dw\|_{\mathcal{M}} + \langle v, \sigma - Dw \rangle + \langle \omega, \tau - (Du - w) \rangle \leq \alpha \|\tau\|_{\mathcal{M}} + \beta \|\sigma\|_{\mathcal{M}},$$
$$\forall (\sigma, \tau) \in X^*$$

$$\Leftrightarrow \begin{cases} \alpha \|Du - w\|_{\mathcal{M}} + \langle \omega, \tau - (Du - w) \rangle \leq \alpha \|\tau\|_{\mathcal{M}}, \quad \forall \tau \in H^1_0(\Omega)^* \\ \beta \|Dw\|_{\mathcal{M}} + \langle v, \sigma - Dw \rangle \leq \beta \|\sigma\|_{\mathcal{M}}, \quad \forall \sigma \in H^2_0(\Omega)^* \end{cases}$$

$$\Leftrightarrow \begin{cases} \alpha \|Du - w\|_{\mathcal{M}} + \langle \omega, \psi - (Du - w) \rangle \leq \alpha \|\tau\|_{\mathcal{M}}, \quad \forall \tau \in \mathcal{M}(\Omega) \\ \beta \|Dw\|_{\mathcal{M}} + \langle v, \sigma - Dw \rangle \leq \beta \|\sigma\|_{\mathcal{M}}, \quad \forall \tau \in \mathcal{M}(\Omega) \end{cases}$$

$$\Leftrightarrow \begin{cases} \omega \in \alpha \, \partial \|\cdot\|_{\mathcal{M}}(Du - w) \\ v \in \beta \, \partial \|\cdot\|_{\mathcal{M}}(Dw) \end{cases} \tag{4.18}$$

Using Lemma 4.3.4, the relations (4.18) are equivalent to

$$\omega \in \alpha \operatorname{Sgn}(Du - w), \tag{4.19}$$

$$v \in \beta \operatorname{Sgn}(Dw). \tag{4.20}$$

Now the second optimality condition (4.17) is equivalent to

$$F_2^*(w, u) + \langle \Lambda(v, \omega), (\hat{w}, \hat{u}) - (w, u) \rangle \leq F_2^*(\hat{w}, \hat{u}), \ \ \forall (\hat{w}, \hat{u}) \in Y^*$$

$$\Leftrightarrow \frac{1}{2} \int_\Omega (u - f)^2 dx + \langle \omega + v', \hat{w} - w \rangle + \langle \omega', \hat{u} - u \rangle \leq \frac{1}{2} \int_\Omega (\hat{u} - f)^2 dx, \ \ \forall (\hat{w}, \hat{u}) \in Y^*$$

$$\Leftrightarrow \begin{cases} 0 + \langle \omega + v', \hat{w} - w \rangle \leq 0, \ \ \forall \hat{w} \in H_0^1(\Omega)^* \\ \frac{1}{2} \int_\Omega (u - f)^2 dx + \langle \omega', \hat{u} - u \rangle \leq \frac{1}{2} \int_\Omega (\hat{u} - f)^2 dx, \ \ \forall \hat{u} \in L^2(\Omega)^* \end{cases}$$

$$\Leftrightarrow \begin{cases} \omega = -v' \\ \omega' \in \partial \left(\frac{1}{2} \| \cdot - f \|_{L^2(\Omega)}^2 \right)(u) \end{cases}$$

$$\Leftrightarrow \begin{cases} \omega = -v' \\ \omega' = u - f \end{cases} \tag{4.21}$$

Combining (4.21) with (4.19) and (4.20) we deduce (C_f)–(C_β).

$\square$

We note here that analogous optimality conditions hold for the L^1–$\mathrm{TGV}_{\beta,\alpha}^2$ problem as well as for the corresponding L^1–TV, L^2–TV problems, see [BKV13] and [Rin00]. We provide a summary of these in Table 4.1.

Optimality conditions – Dimension one			
L^1–αTV	L^2–αTV	L^1–$\mathrm{TGV}_{\beta,\alpha}^2$	L^2–$\mathrm{TGV}_{\beta,\alpha}^2$
$u \in \mathrm{BV}(\Omega), \ v \in H_0^1(\Omega)$	$u \in \mathrm{BV}(\Omega), \ v \in H_0^1(\Omega)$	$u, w \in \mathrm{BV}(\Omega), \ v \in H_0^2(\Omega)$	$u, w \in \mathrm{BV}(\Omega), \ v \in H_0^2(\Omega)$
$v' \in \operatorname{Sgn}(f - u),$ $-v \in \alpha \operatorname{Sgn}(Du)$	$v' = f - u,$ $-v \in \alpha \operatorname{Sgn}(Du)$	$v'' \in \operatorname{Sgn}(f - u),$ $-v' \in \alpha \operatorname{Sgn}(Du - w),$ $v \in \beta \operatorname{Sgn}(Dw)$	$v'' = f - u,$ $-v' \in \alpha \operatorname{Sgn}(Du - w),$ $v \in \beta \operatorname{Sgn}(Dw)$

Table 4.1: Summary of the optimality conditions for the one dimensional TV and TGV denoising problems with L^1 and L^2 fidelity terms.

4.4 Properties of solutions

4.4.1 Basic properties

Before we proceed to our computations of exact solutions for simple data functions, we firstly show some properties of the solutions of the one dimensional L^2–$\mathrm{TGV}^2_{\beta,\alpha}$ regularisation. The following two propositions were proved in [BKV13] for the one dimensional L^1–$\mathrm{TGV}^2_{\beta,\alpha}$ model and they hold in the L^2 fidelity case as well. Their proofs are minor adjustments of the corresponding proofs for the L^1 fidelity case and we omit them. Recall also the definitions of the good representatives u^{up} and u_{low} of a function $u \in \mathrm{BV}(\Omega)$, defined after Theorem 2.3.12.

Proposition 4.4.1. *Let $f \in \mathrm{BV}(\Omega)$ and suppose that $u, w \in \mathrm{BV}(\Omega)$ solve $\mathcal{P}$. Suppose that $u^{up} < f_{low}$ on an open interval $I \subseteq \Omega$. Then the following hold:*

(i) $(Du - w)\lfloor I = 0$, that is to say $u' = w$ on I and $|D^s u|(I) = 0$.

(ii) $w' = 0$ on I and $0 \le -Dw\lfloor I \ll \delta_x$ for some $x \in I$.

(iii) The function $w = u'$ is non-increasing on I.

Similarly, if $u_{low} > f^{up}$ on I then we have

(i) $(Du - w)\lfloor I = 0$, that is to say $u' = w$ on I and $|D^s u|(I) = 0$.

(ii) $w' = 0$ on I and $0 \le Dw\lfloor I \ll \delta_x$ for some $x \in I$.

(iii) The function $w = u'$ is non-decreasing on I.

The following proposition states that the jump set of the solution u is contained in the jump set of the solution data f.

Proposition 4.4.2. *Let $f \in \mathrm{BV}(\Omega)$ and let*

$$G_f := \{(x, t) \in \Omega \times \mathbb{R} : \ x \in J_f, \ t \in [f_{low}(x), f^{up}(x)]\} \,.$$

If u solves the L^2–$\mathrm{TGV}^2_{\beta,\alpha}$ minimisation problem with data function f, then

$$G_u \subseteq G_f. \tag{4.22}$$

In particular $J_u \subseteq J_f$.

Let us note that in higher dimensions ($\Omega \subseteq \mathbb{R}^d$), an analogue inclusion holds (up to an $\mathcal{H}^{d-1}$–null set) for the jump sets of the data and the solution of the L^2–TV problem, see [CCN07] and also [Jal12, Jal14]. The corresponding problem for the L^2–$\mathrm{TGV}^2_{\beta,\alpha}$ model in higher dimensions still remains open, however, significant advances have been made recently by Valkonen [Val14a, Val14b].

The following proposition states that at least away from the boundary of Ω, the solution u can be bounded pointwise by the data f.

Proposition 4.4.3 (Behaviour away from the boundary)**.** *Let u be the solution to the problem $\mathcal{P}$ with data f and let $(c,d) \subseteq \{x \in (a,b) : f^{up}(x) < u_{low}(x)\}$ be a maximal interval with that property and $a < c < d < b$. Then*

$$
\begin{aligned}
u_{low}(c) &\in [f_{low}(c), f^{up}(c)], & (4.23)\\
u^{up}(d) &\in [f_{low}(d), f^{up}(d)], & (4.24)
\end{aligned}
$$

and

$$
\inf_{x \in (c,d)} f^{up}(x) \le \min_{x \in [c,d]} u_{low}(x) \le \max_{x \in [c,d]} u^{up}(x) \le \max\{f^{up}(c), f^{up}(d)\}. \qquad (4.25)
$$

Analogue results hold for the $u^{up} < f_{low}$ case.

Proof. Let us note first that the set $\{x \in (a,b) : f^{up}(x) < u_{low}(x)\}$ is open, see [BKV13]. We prove here (4.23) while (4.24) can be shown similarly.

Notice that in the case where $u_{low}(c) < u^{up}(c)$, i.e., $c \in J_u$, (4.23) follows directly from (4.22), since in that case we have $[u_{low}(c), u^{up}(c)] \subseteq [f_{low}(c), f^{up}(c)]$. Thus, we suppose that $u_{low}(c) = u^{up}(c)$. We have two cases now, either $f_{low}(c) = f^{up}(c)$ or $f_{low}(c) < f^{up}(c)$. Consider the first case, $f_{low}(c) = f^{up}(c)$, and suppose towards contradiction that $f^{up}(c) < u_{low}(c)$. In that case we claim that there exists a $\delta > 0$ such that for every $x \in (c - \delta, c]$ we have $f^{up}(x) < u_{low}(x)$ which is a contradiction by the maximality of (c,d). Indeed, if the claim is not true, there exists a sequence $(x_n)_{n \in \mathbb{N}}$ with $x_n < c$ for every $n \in \mathbb{N}$ such that $x_n \to c$ and $u_{low}(x_n) \le f^{up}(x_n)$ for every $n \in \mathbb{N}$. Since both u_{low} and f^{up} are continuous at c that would mean that $u_{low}(c) \le f^{up}(c)$. If $f^{up}(c) > u_{low}(c)$ we arrive again in contradiction since from the continuity of f^{up} and u_{low} at c there must be points in (c,d) where $f^{up} > u_{low}$ which contradicts our assumption. We work similarly for the case $f^{up}(c) < u_{low}(c)$.

In order to prove (4.25), notice first that Proposition 4.4.1 implies that u is continuous and convex in (c,d), thus $u^{up} = u_{low}$ there. From the convexity of u we have for every $0 \le \lambda \le 1$

$$
\begin{aligned}
u^{up}(\lambda c + (1-\lambda)d) &\le \lambda u^{up}(c) + (1-\lambda)u^{up}(d) \\
&\le \lambda \max\{f^{up}(c), f^{up}(d)\} + (1-\lambda)\max\{f^{up}(c), f^{up}(d)\} \\
&= \max\{f^{up}(c), f^{up}(d)\}. \qquad (4.26)
\end{aligned}
$$

Finally, we have that $f^{up}(x) < u_{low}(x)$ for all $x \in (c,d)$. From the existence of the side limits we have that u_{low} can be extended continuously to $[c,d]$ and thus

$$
\inf_{x \in (c,d)} f^{up}(x) \le \min_{x \in [c,d]} u_{low}(x). \qquad (4.27)
$$

Combining (4.26) and (4.27) we get (4.25). □

We note that the above pointwise estimates hold only away from the boundary of Ω. Indeed as we will see in the next sections, there are examples where $u < \inf f$ and $u > \sup f$ near the boundary (in the sense of good representatives). This is in fact in contrast to TV regularisation, where one can show that $\inf f < u_{low} < u^{up} < \sup f$ in the whole domain Ω. In the following proposition we investigate further the structure of the solution u close to the boundary of Ω.

Proposition 4.4.4 (Behaviour near the boundary)**.** *The following two statements hold:*

  (i) If (a, c) is a maximal interval subset of $\{x \in (a, b) : u^{up} < f_{low}\}$, then u^{up} is an affine function there.

  (ii) Suppose that $u = f$ a.e. in a maximal set of a form (a, c) and suppose that $u^{up} < f_{low}$ in a set of the form (c, d). Then u is affine in (a, d).

Analogue results hold for the case $u_{low} > f^{up}$ and also near b.

Proof. (*i*) We know from from Proposition 4.4.1 that u is continuous and concave in (a, c). We claim that $Dw = 0$, i.e., w is a constant function there, and thus from the fact that $u' = w$ we get that u is affine. Indeed, we have that v is strictly convex on (a, c) with $v(a) = v'(a_+) = 0$. This means that v is strictly increasing in (a, c). If $Dw\lfloor(a, c) = -\delta_x$ for some $x \in (a, c)$ we would get from (C_β) that $v(x) = -\beta$ and $v(x') < -\beta$ for every $a < x' < x$ which is a contradition.

(*ii*) Because of the fact that $u = f$ a.e. in (a, c) we have from (C_f) that v is an affine function there and since $v \in H_0^2(a, b)$ we have that $v = 0$ in (a, c). Condition (C_β) forces w to be constant and condition (C_α) forces $Du = w$ and thus u is an affine function in (a, c) say with derivative equal to λ_1. Since $u^{up} < f_{low}$ in (c, d) we have that u is also an affine function there, say with derivative equal to λ_2. Moreover, u must be continuous in c because otherwise from condition (C_α) we would have that $|v'(c)| = \alpha$ and then v' would be discontinuous at c. Also, we have $\lambda_1 = \lambda_2$. This is because of the fact that $w\lfloor(a, c) = \lambda_1$, $w\lfloor(c, d) = \lambda_2$ and if $\lambda_1 \ne \lambda_2$, from condition (C_β) we would have that $|v(c)| = \beta$ and thus v would be discontinuous at c. Note finally, that u cannot have a gradient change in (c, d) because then v would have a local minimum in (c, d) which is again impossible. □

The second part of Proposition 4.4.4 tells us that the case $u = f$ near the boundary can possibly happen only if f is affine there. In the next proposition we point out that the derivative of u cannot change across a jump point.

Proposition 4.4.5 (Behaviour at a jump point)**.** *Let u be a solution to the problem $\mathcal{P}$ with data f. Suppose that u has a jump discontinuity at a point x_0 and that there exists*

an $\epsilon > 0$ such that $u^{up} < f_{low}$ or $u_{low} > f^{up}$ in $(x_0 - \epsilon, x_0 + \epsilon)$ where also f is continuous. Then there exists $0 < \delta < \epsilon$ such that u' is constant on $(x_0 - \delta, x_0 + \delta)$, i.e., the derivative of u does not jump on x_0.

Proof. Since u has a jump discontinuity at x_0, from condition (C_α) we have that

$$|v'(x_0)| = \alpha. \tag{4.28}$$

Since $u^{up} < f_{low}$ or $u_{low} > f^{up}$ in $(x_0 - \epsilon, x_0 + \epsilon)$ we have from Proposition 4.4.1 that $Du = w$ in $(x_0 - \epsilon, x_0)$ and also in $(x_0, x_0 + \epsilon)$. Moreover, there exists $\delta \leq \epsilon$ such that Du is constant in each one of the intervals $(x_0 - \delta, x_0)$ and $(x_0, x_0 + \delta)$. Thus if Du has a jump discontinuity then the same is true for w and thus from condition (C_β) we have that

$$|v(x_0)| = \beta. \tag{4.29}$$

However its is easy to check that (4.28) and (4.29) cannot hold simultaneously, since v has an extremum in x_0 and hence its derivative must be zero there. Thus we have a contradiction. $\qquad\square$

In Figure 4.2 we provide an illustration about how the possible solutions can look like. For example, in Figure 4.2(a) solution of the type u_1 cannot occur as according to Proposition 4.4.1 u must be piecewise affine in (c, d). Similarly, u_3 violates the fact that u' must be non-increasing in (c, d) and also contradicts Proposition 4.4.5 since its derivative changes at the jump point d. Solution u_5 comes in contrast to Proposition 4.4.2, as it introduces a new jump. In Figure 4.2(b), solution u_1 cannot occur since f is not affine near the boundary, see Proposition 4.4.4, while u_3 violates the (i) part of the same proposition. Finally, in Figures 4.2(c) and 4.2(d) we describe two examples where the solution can possibly coincide with the data near the boundary. Note that the solution u_1 in Figure 4.2(a) cannot occur as again the (ii) part of Proposition 4.4.4 is violated.

4.4.2 L^2–linear regressions

We continue with some results concerning the L^2–linear regression of the data f. In [BKV13], it was proved that in the L^1 fidelity case, there exist some thresholds $\alpha^*, \beta^* > 0$ such that if $\alpha > \alpha^*$, $\beta > \beta^*$ then the solution u of the one dimensional L^1–$\text{TGV}^2_{\beta,\alpha}$ problem is the L^1–linear regression $f^\diamond$ of f, where

$$f^\diamond \in \operatorname*{argmin}_{v \text{ affine}} \|f - v\|_{L^1(\Omega)}.$$

There, the values of α^* and β^* are independent of the function f and depend only on the domain Ω. Here we show that these thresholds exist for the L^2 fidelity case as well but they depend on the data f. For a function $f \in \text{BV}(\Omega)$, we define $f^\star$ to be the L^2–linear

(a) Away from the boundary

(b) Near the boundary

(c) Solution u equal to the data f near the boundary

(d) Solution u equal to the data f near the boundary
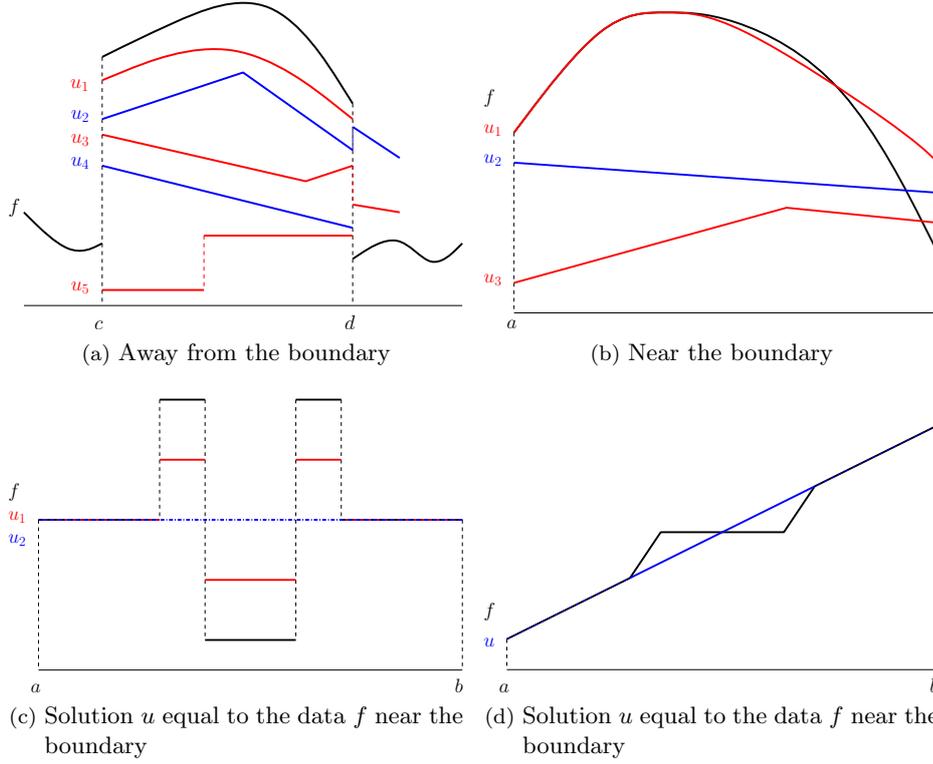
Figure 4.2: Possible (blue) and not possible (red) solutions of the L^2–$\mathrm{TGV}^2_{\beta,\alpha}$ problem as these are dictated by Propositions 4.4.1–4.4.5. These are not actual computed solutions but they illustrate their qualitative properties.

regression of f, i.e.,

$$f^\star := \operatorname*{argmin}_{v \text{ affine}} \|f - v\|^2_{L^2(\Omega)}.$$

We derive the following result:

Proposition 4.4.6 (Thresholds for L^2–linear regression)**.** *Let u be the solution to $\mathcal{P}$ with data function f. Then u is an affine function if and only if $u = f^\star$. Moreover, if*

$$\alpha \geq \frac{b-a}{2}\|f - f^\star\|_\infty, \tag{4.30}$$

$$\beta \geq \frac{(b-a)^2}{4}\|f - f^\star\|_\infty, \tag{4.31}$$

then u is equal to $f^\star$.

Proof. We firstly show that if u is an affine function and solves $\mathcal{P}$, then $u = f^\star$. Since u is affine we have that $D^2 u = 0$, thus the optimum w in $\mathcal{P}$ is $w = Du$. Then, we obviously have

$$u = \operatorname*{argmin}_{\tilde{u} \text{ affine}} \frac{1}{2}\|f - \tilde{u}\|^2_{L^2(\Omega)},$$

i.e., $u = f^\star$. For the second part, since $Df^\star - w = 0$ and $Dw = 0$, from (C_f)–(C_β) we get that in order for the solution to be equal to $f^\star$ it suffices to find a function $v \in H_0^2(\Omega)$, such that

$$v'' = f - f^\star, \quad \|v\|_\infty \leq \beta \quad \text{and} \quad \|v'\|_\infty \leq \alpha. \tag{4.32}$$

It is easy to see that for every function $v \in W^{1,\infty}(\Omega)$ with $v(a) = v(b) = 0$ (recall that v is necessarily continuous) we have the estimate

$$\|v\|_\infty \leq \frac{b-a}{2}\|v'\|_\infty.$$

Hence, a function $v \in H_0^2(\Omega)$ that satisfies $v'' = f - f^\star$, it satisfies

$$\|v'\|_\infty \quad \leq \quad \frac{b-a}{2}\|f - f^\star\|_\infty, \tag{4.33}$$

$$\|v\|_\infty \quad \leq \quad \frac{(b-a)^2}{4}\|f - f^\star\|_\infty. \tag{4.34}$$

Thus, in order to have this kind of solution, it suffices to choose α and β as in (4.30)–(4.31) and to find a function $v \in H_0^2(\Omega)$ that satisfies $v'' = f - f^\star$. It can be easily checked that the function

$$v(x) = \int_a^x \int_a^t (f - f^*)\, ds dt,$$

satisfies these conditions, baring also in mind that the conditions

$$\int_a^b (f - f^\star)'\, dx = 0 \quad \text{and} \quad \int_a^b x(f - f^\star)\, dx = 0,$$

hold for the L^2–linear regression.

$\square$

We note that in Sections 4.6.1 and 4.6.3, we provide some sharper L^2–linear regression thresholds for the data functions for which we are computing exact solutions. The next proposition states that the solution u to the L^2–TGV$_{\beta,\alpha}^2$ problem, has the same L^2–linear regression with the data f.

Proposition 4.4.7. *Let $u_{f,\beta,\alpha}$ be the solution to L^2–TGV$_{\beta,\alpha}^2$ minimisation problem with data f. Then $f^\star = u_{f,\beta,\alpha}^\star$, i.e.,*

$$\operatorname*{argmin}_{v\ affine} \|f - v\|_{L^2(\Omega)}^2 = \operatorname*{argmin}_{v\ affine} \|u_{f,\beta,\alpha} - v\|_{L^2(\Omega)}^2.$$

Proof. Since the L^2–linear regression is a geometric notion, we can assume without loss of generality that Ω is an interval of the form $(0, b)$. From (C_f), we have that $v'' = f - u_{f,\beta,\alpha}$ with $v \in H_0^2(\Omega)$. Thus, by integrating by parts, we have that $\int_\Omega v''\, dx = 0$ and $\int_\Omega xv''\, dx =$

0 which implies

$$\int_\Omega f\,dx = \int_\Omega u_{f,\beta,\alpha}\,dx \quad \text{and} \quad \int_\Omega xf\,dx = \int_\Omega xu_{f,\beta,\alpha}\,dx.$$

Then the proof is finished by simply observing that for a function g we have

$$\underset{v \text{ affine}}{\operatorname{argmin}} \|g - v\|_{L^2(\Omega)}^2 = \underset{v=\lambda x+\mu}{\operatorname{argmin}} \|g - \lambda x - \mu\|_{L^2(\Omega)}^2$$

$$= \underset{v=\lambda x+\mu}{\operatorname{argmin}} \frac{1}{2}\lambda^2 b^2 + \lambda b^3 + b\mu^2 - 2\lambda \int_\Omega xg\,dx - 2\mu \int_\Omega g\,dx,$$

which means that $g^\star$ depends only on the quantities $\int_\Omega g\,dx$ and $\int_\Omega xg\,dx$. $\qquad\square$

4.4.3 Even and odd functions

Before we proceed with the computation of exact solutions, we point out some facts concerning odd and even data functions. In particular, at the end of this section we prove that in some symmetric cases, TV and TGV regularisations coincide, given that α and β satisfy a very simple condition. It is convenient for this section to assume that Ω is an interval of the type $(-L, L)$. We first prove the following two propositions, which essentially state that the symmetry of the data f is inherited to the solutions u and v of the problems $\mathcal{P}$ and $\mathcal{P}'$ respectively.

Proposition 4.4.8. *Let $f \in \mathrm{BV}(-L, L)$ be an odd (even) function. Then the solution $v \in H_0^2(-L, L)$ to the problem $\mathcal{P}'$ is also odd (even).*

Proof. We prove the odd case, the even case is proved analogously. Suppose that $f(x) = -f(-x)$ a.e.. We will show that $v(x) = -v(-x)$. Let $\hat{f}(x) = -f(-x)$ and $\hat{v}$ to be the solution to $\mathcal{P}'$ with data function $\hat{f}$ and set $\tilde{v}(x) = -v(-x)$. Obviously

$$\|v\|_\infty \le \beta, \ \|v'\|_\infty \le \alpha \iff \|\tilde{v}\|_\infty \le \beta, \ \|\tilde{v}'\|_\infty \le \alpha.$$

Since $\tilde{v}(x) = -v(-x)$, we have $\tilde{v}''(x) = -v''(-x)$ a.e. and thus

$$\int_{-L}^{L} \hat{f}(x)\tilde{v}''(x)dx - \frac{1}{2}\int_{-L}^{L}(\tilde{v}''(x))^2dx \ \le \ \int_{-L}^{L}\hat{f}(x)\hat{v}''(x)dx - \frac{1}{2}\int_{-L}^{L}(\hat{v}''(x))^2dx \Longrightarrow$$

$$\int_{-L}^{L} f(x)v''(x)dx - \frac{1}{2}\int_{-L}^{L}v''(x)^2dx \ \le \ \int_{-L}^{L}f(x)(-\hat{v}''(-x))dx - \frac{1}{2}\int_{-L}^{L}((-\hat{v}(-x))'')^2dx,$$

thus $v(x) = -\hat{v}(-x)$ but since $\hat{f} = f$ we have that $\hat{v} = v$. $\qquad\square$

Proposition 4.4.9. *Let $f \in \mathrm{BV}(-L, L)$ be an odd (even) function. Then the solution u to the problem $\mathcal{P}$ is also odd (even).*

Proof. Again we only prove the odd case as the even case can be proved in a similar fashion. Set $\hat{f}(x) = -f(-x)$, $\hat{u}$ the solution of $\mathcal{P}$ with data function $\hat{f}$ and set $\tilde{u}(x) = -u(-x)$. One can easily notice from the rotational invariance of TGV that $\mathrm{TGV}^2_{\beta,\alpha}(u) = \mathrm{TGV}^2_{\beta.\alpha}(\tilde{u})$. We have

$$\frac{1}{2}\int_{-L}^{L}(\hat{u}(x) - \hat{f}(x))^2 dx + \mathrm{TGV}^2_{\beta,\alpha}(\hat{u}) \quad \leq \quad \frac{1}{2}\int_{-L}^{L}(\tilde{u}(x) - \hat{f}(x))^2 dx + \mathrm{TGV}^2_{\beta,\alpha}(\tilde{u})$$

$$\Longrightarrow$$

$$\frac{1}{2}\int_{-L}^{L}(-\hat{u}(-x) - f(x))^2 dx + \mathrm{TGV}^2_{\beta,\alpha}(-\hat{u}(-\cdot)) \quad \leq \quad \frac{1}{2}\int_{-L}^{L}(u(x) - f(x))^2 dx + \mathrm{TGV}^2_{\beta,\alpha}(u).$$

Thus we have $u(x) = -\hat{u}(-x)$. Since $\hat{f} = f$ we have $\hat{u} = u$ and thus $u(x) = -u(-x)$. $\qquad\square$

We are going to prove that at least for even data functions f, if the ratio β/α of the $\mathrm{TGV}^2_{\beta,\alpha}$ parameters is large enough, then TGV regularisation is equivalent to TV regularisation with weight α. We firstly need the following lemma concerning the total variation of even functions. It states that the total variation of an even function cannot be decreased by adding an affine function to it.

Lemma 4.4.10. *Let $u \in \mathrm{BV}(-L, L)$ be an even function. Then for every $c \in \mathbb{R}$ we have*

$$\|Du\|_{\mathcal{M}} \leq \|Du + c\|_{\mathcal{M}}. \tag{4.35}$$

Proof. We can assume without loss of generality that $c > 0$. In order to show (4.35), it suffices to show that

$$\int_{(-L,L)} |u'|\, dx \leq \int_{(-L,L)} |u' + c|\, dx. \tag{4.36}$$

Since u is even we have that u' is odd and thus up to null sets we have

$$\{u' > 0\} \cap (-L, 0) \quad = \quad -\{u' < 0\} \cap (0, L),$$
$$\{u' < 0\} \cap (-L, 0) \quad = \quad -\{u' > 0\} \cap (0, L).$$

Then we estimate

$$\int_{(-L,L)} |u'|\, dx = \int_{\{u'>0\}\cap(-L,0)} u'\, dx \; - \int_{\{u'<0\}\cap(0,L)} u'\, dx \; + \int_{\{u'>0\}\cap(0,L)} u'\, dx \; - \int_{\{u'<0\}\cap(-L,0)} u'\, dx$$

$$= \int_{\{u'>0\}\cap(-L,0)} c + u'\, dx \; - \int_{\{u'<0\}\cap(0,L)} c + u'\, dx \; + \int_{\{u'>0\}\cap(0,L)} c + u'\, dx \; - \int_{\{u'<0\}\cap(-L,0)} c + u'\, dx$$

$$\leq \int_{\{u'>0\}\cap(-L,0)} c + u'\, dx \; + \int_{\{u'<0\}\cap(0,L)} |c + u'|\, dx \; + \int_{\{u'>0\}\cap(0,L)} c + u'\, dx \; + \int_{\{u'<0\}\cap(-L,0)} |c + u'|\, dx$$

$$+ \int\limits_{\{u'=0\}\cap(-L,L)} |c|\, dx$$

$$= \int\limits_{(-L,L)} |u' + c|\, dx.$$

$\square$

Theorem 4.4.11. *Let $f \in \mathrm{BV}(-L, L)$ be an even function. If*

$$\frac{\beta}{\alpha} \geq L, \tag{4.37}$$

then the solution of the L^2–$\mathrm{TGV}^2_{\beta,\alpha}$ regularisation problem is the same with the solution of the following TV *regularisation problem:*

$$\min_{u \in \mathrm{BV}(-L,L)} \frac{1}{2} \int_L^L (u - f)^2 dx + \alpha \, \mathrm{TV}(u).$$

Proof. Recall that for every $u \in W^{1,1}(-L, L)$ the following Poincaré inequality holds,

$$\|u - u_\Omega\|_{L^1(-L,L)} \leq L \|\nabla u\|_{L^1(-L,L)},$$

i.e., the Poincaré constant is equal to L, see for example [AFP00]. Using density arguments, we can prove that the same inequality holds for the space $\mathrm{BV}(-L, L)$ with the same constant, i.e., for every $u \in \mathrm{BV}(-L, L)$

$$\|u - u_\Omega\|_{L^1(-L,L)} \leq L \|Du\|_{\mathcal{M}}. \tag{4.38}$$

Since f is even, from Proposition 4.4.9 we have that the solution of

$$\min_{u \in \mathrm{BV}(-L,L)} \frac{1}{2} \int_\Omega (u - f)^2 dx + \mathrm{TGV}^2_{\beta,\alpha}(u), \tag{4.39}$$

is an even function as well. Thus, the problem (4.39) is equivalent to

$$\min_{\substack{u \in \mathrm{BV}(-L,L) \\ u \text{ even}}} \frac{1}{2} \int_{-L}^L (u - f)^2 dx + \mathrm{TGV}^2_{\beta,\alpha}(u). \tag{4.40}$$

Similarly we can prove that the outcome of TV regularisation for even data, is even as well. Thus, it suffices to prove that for an even function u, we have $\mathrm{TGV}^2_{\beta,\alpha}(u) = \alpha \, \mathrm{TV}(u) := \alpha \|Du\|_{\mathcal{M}}$, provided (4.37) holds. We calculate successively:

$$\mathrm{TGV}^2_{\beta,\alpha}(u) = \min_{w \in \mathrm{BV}(\Omega)} \alpha \|Du - w\|_M + \beta \|Dw\|_{\mathcal{M}}$$

$$
\begin{aligned}
&\leq \alpha\|Du\|_{\mathcal{M}} && \text{(choosing } w = 0\text{)},\\
&\leq \alpha\|Du - w_\Omega\|_{\mathcal{M}} && \forall w \in \mathrm{BV}(\Omega), \text{(from Lemma 4.4.10)},\\
&\leq \alpha\|Du - w\|_{\mathcal{M}} + \alpha\|w - w_\Omega\|_{L^1(\Omega)}\\
&\leq \alpha\|Du - w\|_{\mathcal{M}} + \frac{\beta\alpha L}{\beta}\|Dw\|_{\mathcal{M}} && \text{(Poincaré inequality)},\\
&\leq \max\left\{1, \frac{aL}{\beta}\right\}\left(\alpha\|Du - w\|_{\mathcal{M}} + \beta\|Dw\|_{\mathcal{M}}\right)\\
&= \alpha\|Du - w\|_{\mathcal{M}} + \beta\|Dw\|_{\mathcal{M}} && \text{(since (4.37) holds)}.
\end{aligned}
$$

Thus, we have proved above that

$$
\alpha\|Du\|_{\mathcal{M}} \leq \alpha\|Du - w\|_{\mathcal{M}} + \beta\|Dw\|_{\mathcal{M}} \quad \forall w \in \mathrm{BV}(\Omega)
$$

which implies that

$$
\begin{aligned}
\mathrm{TGV}^2_{\beta,\alpha}(u) &\leq \alpha\|Du\|_{\mathcal{M}}\\
&\leq \min_{w\in\mathrm{BV}(\Omega)} \alpha\|Du - w\|_{\mathcal{M}} + \beta\|Dw\|_{\mathcal{M}}\\
&= \mathrm{TGV}^2_{\beta,\alpha}(u),
\end{aligned}
$$

and the proof is complete. $\qquad\square$

Let us emphasise the fact that the condition (4.37) is independent of the function f. Moreover, as we will see in Section 4.6.1 there exist simple examples of odd data functions (function $f_{\mathrm{p.c.}}$) where the TGV and TV regularisations do not coincide for any choice of α and β. In Figure 4.3, we verify numerically the above result. Both TGV and TV minimisation problems are solved with the Chambolle–Pock primal dual method [CP11] as it is described in [Bre14].

4.5 A note on the relationship of TV and TGV in dimension two

In this section we want to point out that a version of Theorem 4.4.11 holds in dimension two as well, i.e., if the ratio β/α is large enough and the data function f is symmetric enough then TV and TGV regularisations coincide. To do so, we first recall some results from the theory of functions of bounded deformation, see [TS80, Tem85]. We state them in dimension two but higher dimensional analogues also hold.

Proposition 4.5.1. *Let $\Omega \subseteq \mathbb{R}^2$ be an open, bounded set with Lipschitz boundary. Then*
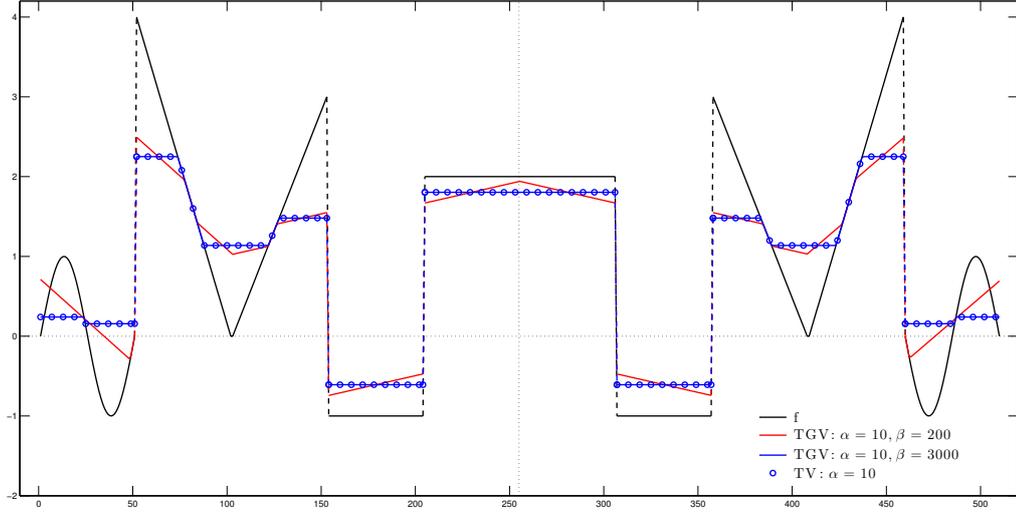
Figure 4.3: Numerical verification of Theorem 4.4.11. Using the even function f shown above as a data function and setting $\beta/\alpha \geq L$, the solutions of the $\text{TGV}^2_{\beta,\alpha}$ and $\alpha\,\text{TV}$ regularisation problems coincide. If we decrease the ratio β/α violating condition (4.37), the two regularisations do not coincide any more with the TGV solution (red) being different than the TV one (blue).

for every function $w \in \text{BD}(\Omega)$ there exists an element $r_w \in \text{Ker}\mathcal{E}$ such that

$$\|w - r_w\|_{L^1(\Omega,\mathbb{R}^2)} \leq C\|\mathcal{E}w\|_{\mathcal{M}}, \tag{4.41}$$

where the constant C depends only on the domain Ω. Note moreover that

$\text{Ker}\mathcal{E} = \{r \in L^1(\Omega,\mathbb{R}^2): r(x) = Ax+B,\ B \in \mathbb{R}^2,\ A \in \mathbb{R}^{2\times 2}$ *is a skew symmetric matrix*$\}$.

Recall that in dimension one the variable w is a solution to an L^1–TV problem. In dimension two, w is a solution to the following L^1–$\mathcal{E}$ problem:

$$w \in \operatorname*{argmin}_{\tilde{w}\in\text{BD}(\Omega)} \int_\Omega |D^\alpha u - \tilde{w}|\ dx + \frac{\beta}{\alpha}\|\mathcal{E}\tilde{w}\|_{\mathcal{M}}. \tag{4.42}$$

We are thus turning our attention to the more generic L^1–$\mathcal{E}$ problem

$$\min_{w\in\text{BD}(\Omega)} \|f - w\|_{L^1(\Omega,\mathbb{R}^2)} + \lambda\|\mathcal{E}w\|_{\mathcal{M}}, \quad f \in L^1(\Omega,\mathbb{R}^2),\ \lambda > 0. \tag{4.43}$$

Notice that (4.43) does not have necessarily unique solutions as it is not strictly convex. The next theorem states that that if the parameter λ is large enough then a solution w of (4.43) belongs to $\text{Ker}\mathcal{E}$. This is analogous to the L^1–TV problem, where for large enough value of the parameter λ we have that the solution is constant, i.e., it belongs to the kernel

of TV.

Proposition 4.5.2. *Let $\Omega \subseteq \mathbb{R}^2$ be an open, bounded set with Lipschitz boundary and let $f \in L^1(\Omega, \mathbb{R}^2)$. Then there exists a $\lambda^* > 0$ such that for every $\lambda > \lambda^*$, if w_λ is a solution of (4.43) with parameter λ then*

$$w_\lambda \in \operatorname*{argmin}_{w \in \operatorname{Ker}\mathcal{E}} \|f - w\|_{L^1(\Omega, \mathbb{R}^2)}.$$

Proof. Since w_λ is a solution of (4.43) we have that

$$\|f - w_\lambda\|_{L^1(\Omega, \mathbb{R}^2)} + \lambda\|\mathcal{E}w_\lambda\|_{\mathcal{M}} \le \|f\|_{L^1(\Omega, \mathbb{R}^2)}. \tag{4.44}$$

From inequality (4.41) we get

$$\|f - w_\lambda\|_{L^1(\Omega, \mathbb{R}^2)} + \frac{\lambda}{C}\|w_\lambda - r_{w_\lambda}\|_{L^1(\Omega, \mathbb{R}^2)} \le \|f\|_{L^1(\Omega, \mathbb{R}^2)}. \tag{4.45}$$

It is now easy to see that $W_\lambda := w_\lambda - r_{w_\lambda}$ solves the following problem:

$$\min_{w \in \operatorname{BD}(\Omega)} \|(f - r_{w_\lambda}) - w\|_{L^1(\Omega, \mathbb{R}^2)} + \lambda\|\mathcal{E}w\|_{\mathcal{M}}. \tag{4.46}$$

Indeed, one has to check that

$$\|(f - r_{w_\lambda}) - W_\lambda\|_{L^1(\Omega, \mathbb{R}^2)} + \lambda\|\mathcal{E}W_\lambda\|_{\mathcal{M}} \le \|(f - r_{w_\lambda}) - w\|_{L^1(\Omega, \mathbb{R}^2)} + \lambda\|\mathcal{E}w\|_{\mathcal{M}}, \ \forall w \in \operatorname{BD}(\Omega)$$

$$\Longleftrightarrow$$

$$\|(f - w_\lambda\|_{L^1(\Omega, \mathbb{R}^2)} + \lambda\|\mathcal{E}w_\lambda\|_{\mathcal{M}} \le \|f - (w + r_{w_\lambda})\|_{L^1(\Omega, \mathbb{R}^2)} + \lambda\|\mathcal{E}(w + r_{w_\lambda})\|_{\mathcal{M}}, \ \forall w \in \operatorname{BD}(\Omega)$$

$$\Longleftrightarrow$$

$$\|(f - w_\lambda\|_{L^1(\Omega, \mathbb{R}^2)} + \lambda\|\mathcal{E}w_\lambda\|_{\mathcal{M}} \le \|f - w\|_{L^1(\Omega, \mathbb{R}^2)} + \lambda\|\mathcal{E}w\|_{\mathcal{M}}, \ \forall w \in \operatorname{BD}(\Omega),$$

with the last inequality being true since w_λ is a solution of (4.43). Setting $F_\lambda := f - r_{w_\lambda}$, using the fact that W_λ solves (4.46) and the inequality $\|W_\lambda\|_{L^1(\Omega, \mathbb{R}^2)} \le C\|\mathcal{E}W_\lambda\|_{\mathcal{M}}$ we get

$$\|F_\lambda - W_\lambda\|_{L^1(\Omega, \mathbb{R}^2)} + \frac{\lambda}{C}\|W_\lambda\|_{L^1(\Omega, \mathbb{R}^2)} \le \|F_\lambda\|_{L^1(\Omega, \mathbb{R}^2)}. \tag{4.47}$$

A simple application of triangle inequality in (4.47) yields that if $\lambda > C$, then we must have $W_\lambda = 0$, i.e., $w_\lambda = r_{w_\lambda}$. We thus set $\lambda^* = C$ and it is obvious that if $\lambda > \lambda^*$ then

$$w_\lambda \in \operatorname*{argmin}_{w \in \operatorname{Ker}\mathcal{E}} \|f - w\|_{L^1(\Omega, \mathbb{R}^2)}.$$

$\square$

We are now ready to prove that under some symmetry assumptions for Ω and f and for a large enough ratio β/α, TV and TGV regularisations coincide in dimension two as

well. For simplicity we will assume that Ω is a square but other symmetric domains can be considered as well.

Theorem 4.5.3. *Suppose that $\Omega \subseteq \mathbb{R}^2$ is a bounded square, centred at the origin. Let $f \in \mathrm{BV}(\Omega)$ satisfy the following symmetry properties*

    (i) f is symmetric with respect to both x and y axis, i.e.,

$$f(x_1, x_2) = f(-x_1, x_2), \quad f(x_1, x_2) = f(x_1, -x_2), \quad \text{for a.e. } (x_1, x_2) \in \Omega. \tag{4.48}$$

    (ii) f is invariant under $\pi/2$ rotations, i.e., $f(O_{\pi/2}x) = f(x)$, where $O_{\pi/2}$ denotes

    counterclockwise rotation by $\pi/2$ degrees. $\tag{4.49}$

Then if $\beta/\alpha > C$, where C is the constant appearing in (4.41), the solution of the L^2–$\mathrm{TGV}^2_{\beta,\alpha}$ regularisation problem with data f is the same with the solution of the following TV regularisation problem:

$$\min_{u \in \mathrm{BV}(\Omega)} \frac{1}{2} \int_\Omega (u-f)^2 dx + \alpha \, \mathrm{TV}(u).$$

Proof. Recall that the L^2–$\mathrm{TGV}^2_{\beta,\alpha}$ reads

$$\min_{\substack{u \in \mathrm{BV}(\Omega) \\ w \in \mathrm{BD}(\Omega)}} \frac{1}{2} \int_\Omega (u-f)^2 dx + \alpha \|Du - w\|_{\mathcal{M}} + \beta \|\mathcal{E}w\|_{\mathcal{M}}, \tag{4.50}$$

where as we have mentioned the variable w satisfies

$$w \in \operatorname*{argmin}_{w \in \mathrm{BD}(\Omega)} \int_\Omega |D^\alpha u - w| \, dx + \frac{\beta}{\alpha} \|\mathcal{E}w\|_{\mathcal{M}}.$$

Since $\beta/\alpha > C$ is satisfied from Proposition 4.5.2 we have that w is of the form $w(x) = Ax + B$ for some $B \in \mathbb{R}^2$ and a skew symmetric matrix A. Since f satisfies the symmetry properties (4.48)–(4.49) from the rotational invariance of TGV we have the same holds for the solution u. That means that $D^\alpha u$ has the following properties

$$D^\alpha u(x) = -D^\alpha u(-x), \quad D^\alpha_1 u(x_1, x_2) = D^\alpha_1 u(x_1, -x_2), \quad D^\alpha_2 u(x_1, x_2) = D^\alpha_2 u(-x_1, x_2) \tag{4.51}$$

$$D_1 u(O_{\pi/2}x) = D_2 u(x), \tag{4.52}$$

for almost every $x = (x_1, x_2) \in \Omega$. Since w satisfies

$$w \in \operatorname*{argmin}_{\tilde{w} \in \mathrm{Ker}\mathcal{E}} \|D^\alpha u - \tilde{w}\|_{L^1(\Omega, \mathbb{R}^2)},$$

and w is of the form $w(x) = Ax + B$ it is easy to check (see Lemma A.2.3 in Appendix A)

that w is zero.

$\square$

In Figure 4.5 we provide two numerical examples that verify Theorem 4.5.3. There, we apply TV and TGV denoising to a characteristic function of a circle centred both at the origin, Figure 4.5.3(a), and away from it, Figure 4.5.3(f). Note that the symmetry properties (4.48)–(4.49) are satisfied for the first case. There, we observe that by choosing the ratio β/α large enough $\text{TGV}^2_{\beta,\alpha}$ and αTV regularisations produce the same results, Figures 4.5.3(b) and 4.5.3(c). However, TV and TGV do not coincide for small ratio β/α, Figure 4.5.3(d). Finally note that when the symmetry is broken, Figure 4.5.3(f), the TV and TGV solutions do not coincide even for large ratio β/α, Figures 4.5.3(g) and 4.5.3(h).

4.6 Computation of exact solutions

In this section we compute exact solutions for the one dimensional L^2–$\text{TGV}^2_{\beta,\alpha}$ regularisation problem for several data functions f. In particular we calculate exact solutions for simple piecewise constant, piecewise affine and hat functions. We focus on the structure of the solutions and their relationship with the parameters α and β.

4.6.1 Piecewise constant function with a single jump

For convenience in the calculations, for this sections, Ω will be an interval of the form $(0, 2L)$. We define $f_{\text{p.c.}}$ to be a piecewise constant function with a single jump, i.e.,

$$f_{\text{p.c.}}(x) = \begin{cases} 0 & \text{if } x \in (0, L), \\ h & \text{if } x \in [L, 2L), \end{cases} \tag{4.53}$$

for some $h > 0$, see Figure 4.4.
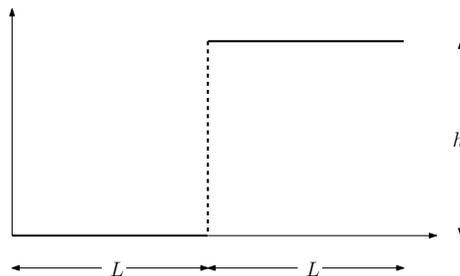
Figure 4.4: Piecewise constant function $f_{\text{p.c.}}$ with a jump discontinuity at $x = L$.

In order to compute exact solutions for $f_{\text{p.c.}}$ our strategy is as follows: Firstly, we investigate what are the possible types of solutions, describing then in a qualitative way and then we calculate them explicitly.
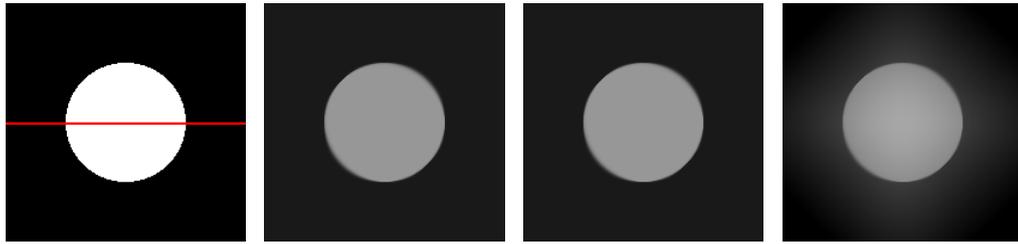
(a) Original image,
200×200 pixels

(b) TV denoising,
$\alpha = 10$

(c) TGV denoising,
$\alpha = 10$, $\beta = 10^6$

(d) TGV denoising,
$\alpha = 10$, $\beta = 200$
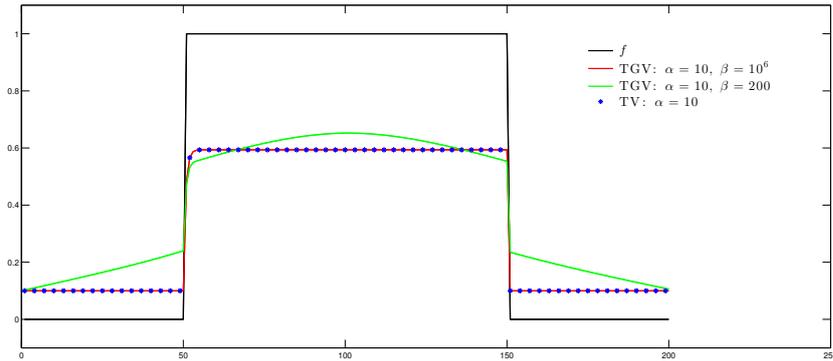
(e) Corresponding middle row slices
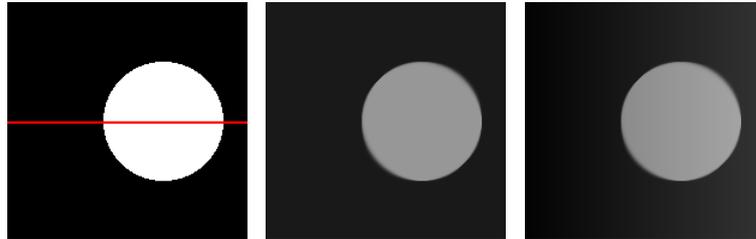
(f) Original image,
200×200 pixels

(g) TV denoising,
$\alpha = 10$

(h) TGV denoising,
$\alpha = 10$, $\beta = 10^6$
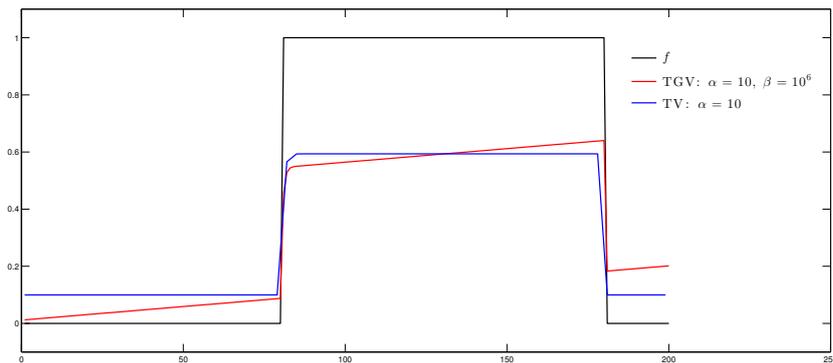
(i) Corresponding middle row slices

Figure 4.5: Illustration of the two dimensional TV–TGV equivalence for symmetric data when β/α is large enough. Notice that the equivalence does not hold once the symmetry is broken, Figures (f)–(i). The red lines to the original images indicate the slices in Figures (d) and (f).

Since after a simple translation $f_{\text{p.c.}}$ is an odd function, from Proposition 4.4.9 we have that the solution u is symmetric with respect to the point $(L, h/2)$. That is to say

$$u(x) = -u(2L - x) + h, \quad \forall x \in (L, 2L).$$

Thus, we only need to describe the solution in $(0, L)$. The following theorem states which kinds of solutions are allowed to occur.

Theorem 4.6.1. *Let u be the solution to the L^2–$\text{TGV}^2_{\beta,\alpha}$ regularisation problem with data $f_{\text{p.c.}}$. Then u can only be of the following form on $(0, L)$:*

- *There exists $0 < x_1 < L$ such that u is strictly negative and affine in $(0, x_1)$ with $u(x_1) = 0$.*

- *There exists potentially $0 < x_1 \leq \tilde{x}_1 < L$ such that $u = 0$ in $[x_1, \tilde{x}_1]$.*

- *$u > 0$ in $(\tilde{x}_1, L)$ consisting of at most two affine parts of increasing gradient in $(\tilde{x}_1, x_2)$ and (x_2, L) respectively with $\tilde{x}_1 < x_2 < L$. Moreover in the case $\tilde{x}_1 = x_1$, u has the same gradient in $(0, x_1)$ and (x_1, x_2).*

Note: In fact, as we will see later, $\tilde{x}_1$ must be equal to x_1, i.e., the allowed solutions are strictly negative in $(0, x_1)$ and strictly positive in (x_1, L). However this is not obvious by simply interpreting the optimality conditions (C_f), (C_α) and (C_β) (which is how Theorem 4.6.1 is proved) but it is a result of subsequent calculations, see Proposition 4.6.2 and Propositions A.1.1–A.1.3 in Appendix A.

Proof of Theorem 4.6.1. Let us note first that from Propositions 4.4.2 and 4.4.3, we have that u is continuous in $(0, L)$ with $u(L_-) \geq 0$. Also, from Proposition 4.4.4 we have that if $u < 0$ on a set of the form $(0, x_1)$ then u must be affine there. Moreover, Proposition 4.4.1 implies that there is no interval $(c, d) \subseteq (0, L)$ such that $u < 0$ ($u > 0$) in (c, d) with $u(c_+) = u(d_-) = 0$. Indeed, for the $u < 0$ case, there would be a point $x \in (c, d)$ such that u is affine on (c, x) and (x, d) with strictly negative and strictly positive gradient respectively on these intervals. But then u' would be increasing in an interval where $u^{up} < (f_{\text{p.c.}})_{low}$, which is a contradiction. We work similarly for the $u > 0$ case. We conclude that u can be strictly negative only in an interval of the type $(0, c)$, $0 \leq c \leq L$ and strictly positive only in an interval of the type (d, L), $0 \leq d \leq L$, in the latter possibly consisting of two affine parts of increasing gradient. In particular, u is increasing in $(0, L)$. The proof will be complete if we prove that the following situations cannot happen:

(i) $u > 0$ in $(0, L)$.

(ii) $u = 0$ in $(0, x_1)$, $0 < x_1 < L$ and $u > 0$ in (x_1, L).

(iii) $u = 0$ in $(0, L)$.

(iv) $u < 0$ in $(0, x_1)$, $0 < x_1 < L$ and $u = 0$ in (x_1, L).

(v) $u < 0$ in $(0, L)$.

For (i) suppose that $u > 0$ in $(0, L)$. By symmetry we have that $u < h$ in $(L, 2L)$. This means from (C_f) that v is strictly convex in $(0, L)$, strictly concave in $(L, 2L)$ with $v(0_+) = v'(0_+) = v(2L_-) = v'(2l_-) = 0$ something that cannon happen.

For (ii) suppose that there exists a point $0 < x_1 < L$ such that $u = 0$ in $(0, x_1)$ and $u > 0$ in (x_1, L). Then from condition (C_f) v will be an affine function in $(0, x_1)$ but since $v \in H_0^2(0, 2L)$, we have $v(0_+) = v'(0_+) = 0$ and thus $v = 0$ in $(0, x_1)$. Moreover $Du = 0$ in $(0, x_1)$ forcing $w = 0$ a.e. there as well, because otherwise from condition (C_α) we would have that $|v'| = \alpha$ on a set of positive measure in $(0, x_1)$. Now, since $u > 0$ in (x_1, L) (assume without loss of generality that u is affine there) from Proposition 4.4.1 we have that $w = Du = c > 0$ there, something that makes w to have jump discontinuity at x_1 and thus from (C_β) we get that $v(x_1) = \beta$. However, this contradicts to the continuity of v.

For (iii) suppose that $u = 0$ in $(0, L)$. Then by symmetry we have that $u = h$ in $(L, 2L)$. Thus, from condition (C_f) again, we have that v will be affine in those intervals and from its boundary conditions and continuity we get that $v = 0$ in $(0, L)$. But since u has a jump discontinuity at $x = L$, according to condition (C_α), we must have $v'(L) = -\alpha$, a contradiction.

For (iv) suppose that there exists a point $0 < x_1 < L$ such that $u < 0$ in $(0, x_1)$ and $u = 0$ in (x_1, L). Then, from (C_f) v will be strictly convex and thus have a strictly increasing derivative in $(0, x_1)$. Since $v'(0_+) = 0$ we have that $v'(x_1) > 0$. Moreover again from (C_f), v will be affine in (x_1, L) something that forces $v(L) > 0$. However, since after a simple translation $f_{\text{p.c.}}$ is an odd function and according to Proposition 4.4.8 the same is true for the continuous function v, we must have $v(L) = 0$. The case (v) is excluded in a similar way.

Finally, for the last statement of the proposition, suppose that $u < 0$ in $(0, x_1)$, $u > 0$ in (x_1, L) and suppose towards contradiction that the derivative of u has a jump discontinuity at x_1. Since $Du = w$ in $(0, x_1)$ and (x_1, L) this means that w will have a jump discontinuity at x_1 and condition (C_β) imposes that $|v(x_1)| = \beta$. The case $v(x_1) = -\beta$ is impossible since v is strictly increasing in $(0, x_1)$, and also if $v(x_1) = \beta$ from the fact that $v'(x_1) > 0$ and v' is continuous at x_1, we have that v is increasing in $(x_1, x_1 + \delta)$ for some $\delta > 0$, thus $v(x_1 + \delta) > \beta$, contradicting (C_β). $\qquad\square$

Notice how we take advantage of the symmetries of u and v in order to prove Theorem 4.6.1. It remains to rule out the possibility that $x_1 < \tilde{x}_1$. In particular, we have to show

that the four following situations cannot occur [1]:

   (i) $u < 0$ in $(0, x_1)$, $u = 0$ in $(x_1, \tilde{x}_1)$, $u > 0$ with one affine part in   (N-0-P1-J)
      $(\tilde{x}_1, L)$, jump discontinuity at $x = L$,

   (ii) $u < 0$ in $(0, x_1)$, $u = 0$ in $(x_1, \tilde{x}_1)$, $u > 0$ with two affine parts in   (N-0-P2-J)
      $(\tilde{x}_1, L)$, jump discontinuity at $x = L$,

   (iii) $u < 0$ in $(0, x_1)$, $u = 0$ in $(x_1, \tilde{x}_1)$, $u > 0$ with one affine part in   (N-0-P1-C)
      $(\tilde{x}_1, L)$, continuous at $x = L$,

   (iv) $u < 0$ in $(0, x_1)$, $u = 0$ in $(x_1, \tilde{x}_1)$, $u > 0$ with two affine parts in   (N-0-P2-C)
      $(\tilde{x}_1, L)$, continuous at $x = L$,

where $0 < x_1 < \tilde{x}_1 < L$, see also Figure 4.6.

    (a) N-0-P1-J   (b) N-0-P2-J   (c) N-0-P1-C   (d) N-0-P2-C

Figure 4.6: Not allowed type of solutions for the data function $f_{\text{p.c.}}$.

In the following we show that N-0-P1-J cannot occur. The exclusions of the cases N-0-P2-J, N-0-P1-C, N-0-P2-C are done in a similar way and they can be found in Appendix A.

Proposition 4.6.2. *The case N-0-P1-J cannot occur.*

Proof. Suppose that $u < 0$ in $(0, x_1)$, $u = 0$ in $(x_1, \tilde{x}_1)$ and $u > 0$ and affine in $(\tilde{x}_1, L)$ with $0 < x_1 < \tilde{x}_1 < L$. We claim that in that case, u' is constant, say equal to $c > 0$, in $(0, x_1) \cup (\tilde{x}_1, L)$, $w = c$ in $(0, L)$ and v is affine in $(x_1, \tilde{x}_1)$ with gradient equal to α there. Indeed, let c be the value of the derivative of u in $(0, x_1)$. We know already from Proposition 4.4.1 that $w = u'$ there. Now, if $w \neq c$ on a set of positive measure in $(x_1, \tilde{x}_1)$, that will mean that $Dw \neq 0$ somewhere in $(x_1, \tilde{x}_1)$ and thus from (C_β), v must be $\pm\beta$ somewhere in $(x_1, \tilde{x}_1)$ and also $|v| \leq \beta$ there. But this is not possible since v is strictly convex in $(0, x_1)$ with $v'(0_+) = 0$ and it is affine in $(x_1, \tilde{x}_1)$ and thus v is strictly increasing in $(0, \tilde{x}_1)$. Moreover, $w = v' = c$ in $(\tilde{x}_1, L)$ because otherwise w would have a jump discontinuity at $\tilde{x}_1$, something that forces $v(\tilde{x}_1) = \beta$ but this cannot happen since

---

[1] N: negative, 0: zero, P1: positive one affine part, P2: positive two affine parts, J: jump discontinuity, C: continuous.

$v'(\tilde{x}_1) > 0$. Finally, since $Du - w = -c < 0$ in $(x_1, \tilde{x}_1)$, from condition (C_α) we have that $v' = \alpha$ there.

Also note that since u has a jump discontinuity at $x = L$, condition (C_α) forces $v'(L) = -\alpha$ and from the symmetry of v we must have $v(L) = 0$.

For convenience we set

$$\ell_1 = x_1 > 0, \quad \ell_2 = \tilde{x}_1 - x_1 > 0, \quad \ell_3 = L - \tilde{x}_1 > 0 \quad \text{with} \quad \ell_1 + \ell_2 + \ell_3 = L.$$

and

$$v_1 = v \lfloor (0, x_1), \quad v_2 = v \lfloor (x_1, \tilde{x}_1), \quad v_3 = v \lfloor (\tilde{x}_1, L).$$

Since u is an affine function on $(0, x_1)$ we have that v_1 is a cubic function

$$v_1(x) = a_1 x^3 + b_1 x^2 + c_1 x + d_1, \quad x \in (0, x_1),$$

that satisfies the conditions

$$v_1(0) = 0, \quad v_1'(0) = 0, \quad v_1'(\ell_1) = \alpha, \quad v_1''(\ell_1) = 0. \tag{4.54}$$

Thus v_1 will have the form

$$v_1(x) = -\frac{\alpha}{3\ell_1^2} x^3 + \frac{\alpha}{\ell_1} x^2, \quad x \in (0, x_1). \tag{4.55}$$

Since v_2 is an affine function in (x_1, x_2) with gradient α and v is continuous at x_1, we have that v_2 will be of the form

$$v_2(x) = \alpha(x - \ell_1) + \frac{2\alpha\ell_1}{3}, \quad x \in (x_1, \tilde{x}_1). \tag{4.56}$$

Since u is a strictly positive affine function on $(\tilde{x}_1, L)$ (with the same gradient as in $(0, x_1)$), we have that v_3 must be a cubic function

$$v_3(x) = a_3 x^3 + b_3 x^2 + c_3 x + d_3, \quad x \in (\tilde{x}_1, L),$$

that satisfies the following conditions at $\tilde{x}_1 = \ell_1 + \ell_2$:

$$v_3(\ell_1 + \ell_2) = \alpha\ell_2 + \frac{2\alpha\ell_1}{3}, \quad v_3'(\ell_1 + \ell_2) = \alpha, \quad v_3''(\ell_1 + \ell_2) = 0, \quad v_3'''(\ell_1 + \ell_2) = -\frac{2\alpha}{\ell_1^2}, \tag{4.57}$$

and also the following conditions at $L = \ell_1 + \ell_2 + \ell_3$:

$$v_3(\ell_1 + \ell_2 + \ell_3) = 0, \tag{4.58}$$

$$v_3'(\ell_1 + \ell_2 + \ell_3) = -\alpha. \tag{4.59}$$

Conditions (4.57) give

$$v_3(x) = -\frac{\alpha}{3\ell_1^2}(x - \ell_1 - \ell_2)^3 + \alpha(x - \ell_1 - \ell_2) + \alpha\ell_2 + \frac{2\alpha\ell_1}{3}, \quad x \in (\tilde{x}_1, L). \qquad (4.60)$$

What is left is to find the relationship among ℓ_1, ℓ_2 and ℓ_3. Condition (4.59) gives

$$\ell_3 = \sqrt{2}\ell_1, \qquad (4.61)$$

and the condition (4.58) together with (4.61) gives

$$\frac{2 + \sqrt{2}}{3}\ell_1 + \ell_2 = 0, \qquad (4.62)$$

which is a contradiction since both ℓ_1 and ℓ_2 are strictly positive. Thus, this kind of solution cannot occur. $\qquad\square$

From all the previous results, it follows that only the following situations can occur:

 (i) $u < 0$ in $(0, x_1)$, $u > 0$ with one affine part in (x_1, L), jump   (N-P1-J)
 discontinuity at $x = L$,

 (ii) $u < 0$ in $(0, x_1)$, $u > 0$ with two affine parts in (x_1, L), jump   (N-P2-J)
 discontinuity at $x = L$,

 (iii) $u < 0$ in $(0, x_1)$, $u > 0$ with one affine part in (x_1, L), continuous   (N-P1-C)
 at $x = L$,

 (iv) $u < 0$ in $(0, x_1)$, $u > 0$ with two affine parts in (x_1, L), continuous   (N-P2-C)
 at $x = L$,

where $0 < x_1 < L$. In Figure 4.7 we provide a qualitative description of how these four allowed solutions look like, along with a description of the corresponding variables w and v. Notice that from Theorem 4.4.6, the solution of the type N-P1-C is in fact the L^2–linear regression of $f_{\mathrm{p.c.}}$. Note finally that it can easily checked that the direction of the jump of u at $x = L$, is the same with that of the data $f_{\mathrm{p.c.}}$. Otherwise, from (C_α) we would have that $v'(L) = \alpha$ and that would lead again to contradictions. Our next step is to identify the combinations of the parameters α and β that lead to each type of solution.

Proposition 4.6.3 (N-P1-J). *The solution u of the problem $\mathcal{P}$ with data function $f_{\mathrm{p.c.}}$, is of the type N-P1-J, see Figure 4.7(a), if and only if α and β satisfy the conditions*

$$\frac{\beta}{\alpha} \geq \frac{4L}{27} \quad and \quad \alpha < \frac{hL}{8}. \qquad (4.63)$$

(a) Solution of the type N-P1-J.

(b) Solution of the type N-P2-J.

(c) Solution of the type N-P1-C.

(d) Solution of the type N-P2-C.
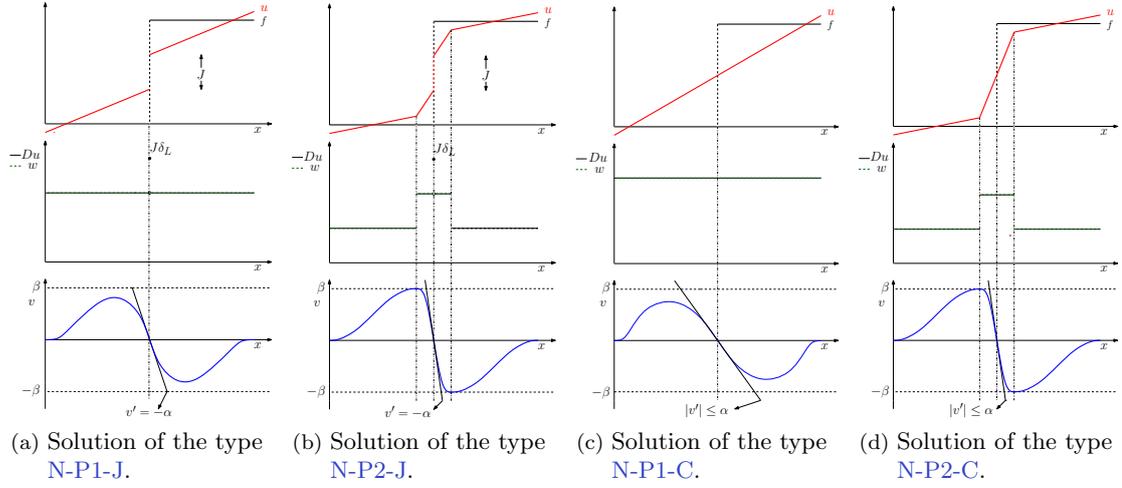
Figure 4.7: All the possible types of solution u to the minimisation problem $\mathcal{P}$ for the data function $f_{\text{p.c.}}$. We also show the forms of the corresponding variables w and v.

More specifically in that case:

$$u(x) = \begin{cases} \frac{6\alpha}{L^2}x - \frac{2\alpha}{L}, & if \quad x \in (0, L), \\ \frac{6\alpha}{L^2}(x - L) + h - \frac{4\alpha}{L}, & if \quad x \in (L, 2L), \end{cases} \tag{4.64}$$

i.e., u has a jump discontinuity of size $h - 8\alpha/L$ at $x = L$.

Proof. It is easy to check that the solution u will be of the type N-P1-J if and only if we can find a function v in $[0, L]$ that satisfies the following conditions:

$$v \text{ is a cubic polynomial,} \quad \text{(from condition } (\text{C}_f), v'' = -u \text{ is an affine function),} \tag{4.65}$$

$$v(0) = 0, \ v'(0) = 0, \quad \text{(boundary conditions for } v), \tag{4.66}$$

$$v(L) = 0, \quad (v \text{ is an odd function),} \tag{4.67}$$

$$v'(L) = -\alpha, \quad (u \text{ jumps at } L, \text{ condition } (\text{C}_\alpha)), \tag{4.68}$$

$$0 < -v''(L) < \frac{h}{2}, \quad (u \text{ jumps at } L \ \& \text{ symmetry),} \tag{4.69}$$

$$|v'| \le \alpha, \quad \text{(condition } (\text{C}_\alpha)), \tag{4.70}$$

$$|v| \le \beta, \quad \text{(condition } (\text{C}_\beta)). \tag{4.71}$$

It is now easy to check that the function

$$v(x) = -\frac{\alpha}{L^2}x^3 + \frac{\alpha}{L}x^2, \quad x \in [0, L],$$

satisfies conditions (4.65)–(4.68) and also the condition (4.70). Observe that v obtains its

137

maximum in $[0, L]$ at $x = 2L/3$, with $v(2L/3) = 4\alpha L/27$, thus (4.71) can be written as

$$\frac{4\alpha L}{27} \leq \beta.$$

Finally, condition (4.69) can be written as

$$0 < \frac{4\alpha}{L} < \frac{h}{2}.$$

The last two inequalities form the two conditions in (4.63). Using the fact that $u = -v''$ in $(0, L)$ and taking advantage of the symmetry of u, we compute (4.64). $\qquad \square$

We note here that the computation of the above type of exact solution for $f_{\text{p.c.}}$ (for a fixed jump $h = 2$) is also done in [Ben11] in order to indicate that $f_{\text{p.c.}}$ cannot be recovered by TGV regularisation up to a loss of contrast, as it is done in TV regularisation

We now proceed to the study of the case N-P2-J. As we will see, it is computationally inaccessible to find the exact conditions on α and β that result to this kind of solution but we are able to provide some sufficient conditions that are not far away from being necessary as well.

Proposition 4.6.4 (N-P2-J). *If α and β satisfy the conditions*

$$\frac{\beta}{\alpha} < \frac{4L}{27} \quad and \quad \beta > \frac{4\alpha^2}{3h}, \tag{4.72}$$

then the solution u of the problem $\mathcal{P}$ with data function $f_{\text{p.c.}}$ is of the type N-P2-J, see Figure 4.7(b). More specifically, there exist points $0 < x_1 < x_2 < 0$ with

$$x_1 = \frac{x_2}{2} \quad and \quad L - \frac{3\beta}{\alpha} < x_2 < L - \frac{9\beta}{4\alpha},$$

such that $u < 0$ and $u > 0$ in $(0, x_1)$ and (x_1, L) respectively with u' having a positive jump at x_2.

Proof. Again, it is easily verified that u will be a solution of the type N-P2-J if and only if we can find $\ell_1 > 0$, $\ell_2 > 0$ with $\ell_1 + \ell_2 = L$ and two functions v_1, v_2 defined on $[0, \ell_1]$ and $[\ell_1, \ell_1 + \ell_2]$ respectively, such that the following conditions are satisfied:

$$v_1, v_2 \text{ are cubic polynomials,} \quad \text{(from condition } (C_f)), \tag{4.73}$$

$$v_1(0) = 0,\ v_1'(0) = 0, \quad \text{(boundary conditions for } v), \tag{4.74}$$

$$v_1(\ell_1) = \beta,\ v_1'(\ell_1) = 0, \quad (v \text{ has a max at } \ell_1, \text{ condition } (C_\alpha)), \tag{4.75}$$

$$v_2(\ell_1) = \beta,\ v_2'(\ell_1) = 0,\ v_2''(\ell_1) = v_1''(\ell_1), \quad \text{(continuity of } v,\ v' \text{ and } u), \tag{4.76}$$

$$v_2'''(\ell_1) < v_1'''(\ell_1), \quad (u' \text{ has a positive jump at } \ell_1) \tag{4.77}$$

$$v_2(L) = 0, \quad (v \text{ is an odd function}), \tag{4.78}$$

$$v_2'(L) = -\alpha, \quad (u \text{ jumps at } L, \text{ condition } (C_\alpha)), \tag{4.79}$$

$$0 < -v_2''(L) < \frac{h}{2}, \quad (u \text{ jumps at } L \text{ \& symmetry}), \tag{4.80}$$

$$|v_i'| \leq \alpha, \ i = 1, 2, \quad (\text{condition } (C_\alpha)), \tag{4.81}$$

$$|v_i| \leq \beta, \ i = 1, 2, \quad (\text{condition } (C_\beta)). \tag{4.82}$$

In that case $\ell_1 = x_2$ and $\ell_2 = L - \ell_1$. From conditions (4.73)–(4.75), we get that

$$v_1(x) = -\frac{2\beta}{\ell_1^3}x^3 + \frac{3\beta}{\ell_1^2}x^2, \quad x \in [0, \ell_1]. \tag{4.83}$$

One can easily check from the fact that $u = -v_1''$ in $(0, \ell_1)$, that $x_1 = \ell_1/2$. We can also check that condition (4.81) for v_2 is equivalent to

$$\frac{3\beta}{2\alpha} \leq \ell_1, \tag{4.84}$$

while condition (4.82) is also satisfied in that case. After some computations, we have that conditions (4.76), (4.78), (4.79) give that v_2 will be of the form

$$v_2(x) = \left(\frac{2\beta}{\ell_1^2\ell_2} - \frac{\alpha}{3\ell_2^2}\right)(x - \ell_1)^3 - \frac{3\beta}{\ell_1^2}(x - \ell_1)^2 + \beta. \tag{4.85}$$

We also get the following relationship between ℓ_1 and ℓ_2:

$$\gamma\ell_2^2 + \ell_2\ell_1^2 - \gamma\ell_1^2 = 0, \quad \text{with} \quad \gamma := \frac{3\beta}{\alpha}, \tag{4.86}$$

where one can check that $\ell_2 < \gamma$ independently of the value of ℓ_1. Notice now, that if the condition (4.77) is satisfied, then v_2 will be decreasing, with decreasing derivative as well something that would imply that conditions (4.81)–(4.82) hold for v_2. With the help of (4.85)–(4.86) and condition $\ell_1 + \ell_2 = L$, we get that condition (4.77) is equivalent to

$$\ell_1^2 - 2L\ell_1 + 2\gamma L < 0. \tag{4.87}$$

The inequality (4.87) is true if and only if

$$\gamma < \frac{L}{2} \quad \text{and} \quad L - \sqrt{L^2 - 2\gamma L} < \ell_1 \iff \ell_2 < \sqrt{L^2 - 2\gamma L}. \tag{4.88}$$

From the fact that $\ell_1 + \ell_2 = L$ and (4.86) we get that ℓ_2 is the unique solution of

$$\phi(\ell_2) := \sqrt{\frac{\gamma\ell_2^2}{\gamma - \ell_2}} + \ell_2 - L = 0. \tag{4.89}$$

In view of (4.89) we have that $\ell_2 = \sqrt{L^2 - 2\gamma L}$ if and only if $\gamma = 4L/9$. Since ϕ is strictly

increasing, one can check that (4.88) is equivalent to the very simple expression

$$\frac{\beta}{\alpha} < \frac{4L}{27}. \tag{4.90}$$

Notice that in that case (4.84) is also satisfied. Finally, the last condition that has to be satisfied is (4.80). Using (4.85) and (4.86) it can be checked that it is equivalent to

$$\frac{4\ell_2 - 2\gamma}{\ell_2^2} < \frac{h}{2\alpha}. \tag{4.91}$$

Ideally, one would like to obtain an explicit expression for ℓ_2 from (4.89) and obtain an inequality involving α and β using (4.91). However, this is practically impossible as one would have to solve a cubic equation for ℓ_2. This is why we are giving some estimates instead. One can check that from (4.89) and (4.90) we can derive that

$$\frac{3}{4}\gamma < \ell_2 < \gamma. \tag{4.92}$$

We have that the expression $(4\ell_2 - 2\gamma)/\ell_2^2$ in (4.91) is a strictly increasing function of ℓ_2 provided that $\ell_2 < \gamma$ which is true in our case. Thus, in order to satisfy (4.91) a sufficient (but not necessary) condition is

$$\frac{4\gamma - 2\gamma}{\gamma^2} < \frac{h}{2\alpha} \iff \beta > \frac{4\alpha^2}{3h}.$$

$$\square$$

Remark: Let us also point out the following fact: Suppose that α^* and β^* are such, so that conditions (4.72) hold, i.e., the solution is of the type N-P2-J. Then there exists a $0 < c < \beta^*$ such that for every $\beta \le c$ the condition (4.91) is violated, so we do not have this type of solution. Indeed from (4.91) and (4.92) we have that

$$\frac{16}{9\gamma} < \frac{4\ell_2 - 2\gamma}{\ell_2^2} < \frac{h}{2\alpha}.$$

Thus keeping α fixed and choosing β such that

$$\frac{16}{9\gamma} > \frac{h}{2\alpha} \iff \beta < \frac{32\alpha^2}{27h}, \tag{4.93}$$

we cannot have the solution of the type N-P2-J any more.

We now turn our attention to the solution of the type N-P1-C. As we mentioned earlier, in that case u is an affine function and thus also the L^2–linear regression of $f_{\text{p.c.}}$, see Theorem 4.4.6. In the same theorem we gave some thresholds for α and β in the general case but we can be more explicit in this specific example.

Proposition 4.6.5 (**N-P1-C**)**.** *The solution u of the problem $\mathcal{P}$ with data function $f_{\text{p.c.}}$ is of the type* N-P1-C, *see Figure 4.7(c), if and only if α and β satisfy the conditions*

$$\alpha \geq \frac{hL}{8} \quad and \quad \beta \geq \frac{hL^2}{54}. \tag{4.94}$$

Moreover, in that case u will be the L^2–linear regression of $f_{\text{p.c.}}$ and equal to

$$u(x) = \frac{3h}{4L}x - \frac{h}{4}, \quad x \in (0, 2L). \tag{4.95}$$

Proof. Again we can check that the solution will be of the type N-P1-C if and only if we can find a function v defined on $[0, L]$ such that the following conditions hold:

$$v \text{ is a cubic polynomial}, \tag{4.96}$$
$$v(0) = 0, \ v'(0) = 0, \tag{4.97}$$
$$v(L) = 0, \tag{4.98}$$
$$-v''(L) = \frac{h}{2}, \quad (\text{condition } (C_f), \text{ continuity at L \& symmetry}), \tag{4.99}$$
$$|v'| \leq \alpha, \tag{4.100}$$
$$|v| \leq \beta. \tag{4.101}$$

We can easily check that conditions (4.96)–(4.99) give

$$v(x) = -\frac{h}{8L}x^3 + \frac{h}{8}x^2, \quad x \in [0, L]. \tag{4.102}$$

Moreover, the maximum values of $|v|$ and $|v'|$ are $\frac{hL^2}{54}$ and $\frac{hL}{8}$ respectively. Thus, conditions (4.100)–(4.101) will be satisfied if and only if (4.94) holds. $\qquad\square$

Finally, we investigate under which combinations of α and β we have solutions of the type N-P2-C. As in the case N-P2-J, it is computationally inaccessible to provide explicit conditions, thus we provide again some sufficient but not necessary conditions.

Proposition 4.6.6 (**N-P2-C**)**.** *If α and β satisfy the conditions*

$$\beta < \frac{hL^2}{54} \quad and \quad \beta \leq \frac{4L\alpha}{9} - \frac{L^2}{27}, \tag{4.103}$$

then the solution u of the problem $\mathcal{P}$ with data function $f_{\text{p.c.}}$ is of the type N-P2-C, *see Figure 4.7(d), and u' has a positive jump at a point x_2 where $\frac{2L}{3} < x_2 < L$.*

Proof. As in the N-P2-J case, u will be a solution of the type N-P2-C if and only if we can find $\ell_1 > 0$, $\ell_2 > 0$ with $\ell_1 + \ell_2 = L$ and two functions v_1, v_2 defined on $[0, \ell_1]$ and

$[\ell_1, \ell_1 + \ell_2]$ respectively, such that the following conditions are satisfied:

$$v_1, v_2 \text{ are cubic polynomials,} \tag{4.104}$$

$$v_1(0) = 0, \ v_1'(0) = 0, \tag{4.105}$$

$$v_1(\ell_1) = \beta, \ v_1'(\ell_1) = 0, \tag{4.106}$$

$$v_2(\ell_1) = \beta, \ v_2'(\ell_1) = 0, \ v_2''(\ell_1) = v_1''(\ell_1), \tag{4.107}$$

$$v_2'''(\ell_1) < v_1'''(\ell_1), \tag{4.108}$$

$$v_2(L) = 0, \tag{4.109}$$

$$-v''(L) = \frac{h}{2}, \quad ((\mathrm{C}_f), \text{ continuity at L \& symmetry)}, \tag{4.110}$$

$$|v_i'| \leq \alpha, \ i = 1, 2, \tag{4.111}$$

$$|v_i| \leq \beta, \ i = 1, 2. \tag{4.112}$$

Conditions (4.104)–(4.106) together with (4.111)–(4.112) yield

$$v_1(x) = -\frac{2\beta}{\ell_1^3} x^3 + \frac{3\beta}{\ell_1^2} x^2, \quad x \in [0, \ell_1], \tag{4.113}$$

with

$$\frac{3\beta}{2\alpha} \leq \ell_1. \tag{4.114}$$

Supposing that v_2 is of the form

$$v_2(x) = a_3(x - \ell_1)^3 + a_2(x - \ell_1)^2 + a_1(x - \ell_1) + a_0, \quad x \in [\ell_1, \ell_1 + \ell_2],$$

conditions (4.107), (4.109), (4.110) give

$$v_2(x) = a_3(x - \ell_1)^3 - \frac{3\beta}{\ell_1^2}(x - \ell_1)^2 + \beta, \quad x \in [\ell_1, \ell_1 + \ell_2], \tag{4.115}$$

and

$$a_3 \ell_2^3 - \frac{3\beta \ell_2^2}{\ell_1^2} + \beta = 0, \tag{4.116}$$

$$-6a_3 \ell_2 + \frac{6\beta}{\ell_1^2} = \frac{h}{2}. \tag{4.117}$$

One can check that conditions (4.107) and (4.110) impose v_2' to be decreasing , so in order to satisfy (4.111) and (4.112) it suffices to impose $|v_2'(\ell_1 + \ell_2)| \leq \alpha$, which, with the help of (4.115) and (4.117), is equivalent to

$$\frac{3\beta \ell_2}{\ell_1^2} + \frac{h \ell_2}{4} \leq \alpha. \tag{4.118}$$

Moreover combining (4.116), (4.117) and the fact that $\ell_1 + \ell_2 = L$, we can get a relationship between ℓ_1 and ℓ_2 and an equation for ℓ_1:

$$\ell_2 = \sqrt{\frac{12\beta}{h + \frac{24\beta}{\ell_1^2}}}, \qquad (4.119)$$

$$\phi(\ell_1) := \ell_1 + \sqrt{\frac{12\beta}{h + \frac{24\beta}{\ell_1^2}}} - L = 0. \qquad (4.120)$$

It is easy to see that the equation (4.120) has a unique solution in $(0, L)$ but one has to solve a quartic in order to express it explicitly. Thus, as in the case N-P2-J we give some estimates. Observe firstly that using (4.113), (4.115), (4.117) and (4.119) condition (4.108) is satisfied if and only if

$$\frac{18\beta\ell_1}{\ell_2} - \frac{6\beta\ell_1^3}{\ell_2^3} < -12\beta, \qquad (4.121)$$

which is satisfied if and only if $\ell_2 < \frac{\ell_1}{2}$. However, from (4.119), this is true if and only if $\ell_1 > \sqrt{\frac{24\beta}{h}}$. From the fact that the function ϕ is strictly increasing, the last inequality is true if and only if

$$\beta < \frac{hL^2}{54}. \qquad (4.122)$$

It remains to identify some sufficient conditions for (4.118). Observe that under (4.122) we have $\phi(2L/3) < 0$ and $\phi(L > 0)$, which means that

$$\frac{2L}{3} < \ell_1 < L \quad \text{and} \quad 0 < \ell_2 < \frac{L}{3}. \qquad (4.123)$$

Using (4.123), we find that a sufficient (but not necessary) condition for (4.118) is

$$\frac{3\beta}{(\frac{2L}{3})^2}\frac{L}{3} + \frac{hL}{12} < \alpha \iff \beta \leq \frac{4L\alpha}{9} - \frac{L^2 h}{27}. \qquad (4.124)$$

We can easily verify that under (4.124) the condition (4.114) is satisfied as well. □

We now summarise how the solutions of the problem $\mathcal{P}$ with data function $f_{\text{p.c.}}$ are affected by the different choices of the parameters α and β. In Figure 4.8 we have partitioned the set $\{\alpha > 0, \ \beta > 0\}$ into different areas that correspond to the four different possible solutions. These areas are:

(1) Blue colour: $\left\{\alpha < \frac{hL}{8}, \ \beta \geq \frac{4L}{27}\alpha\right\}$, (necessary and sufficient condition), **N-P1-J**, see Figure 4.7(a).

(2) Purple colour: $\left\{\beta < \frac{4L}{27}\alpha, \ \beta \geq \frac{36}{27h}\alpha^2\right\}$, (sufficient condition), **N-P2-J**, see Figure
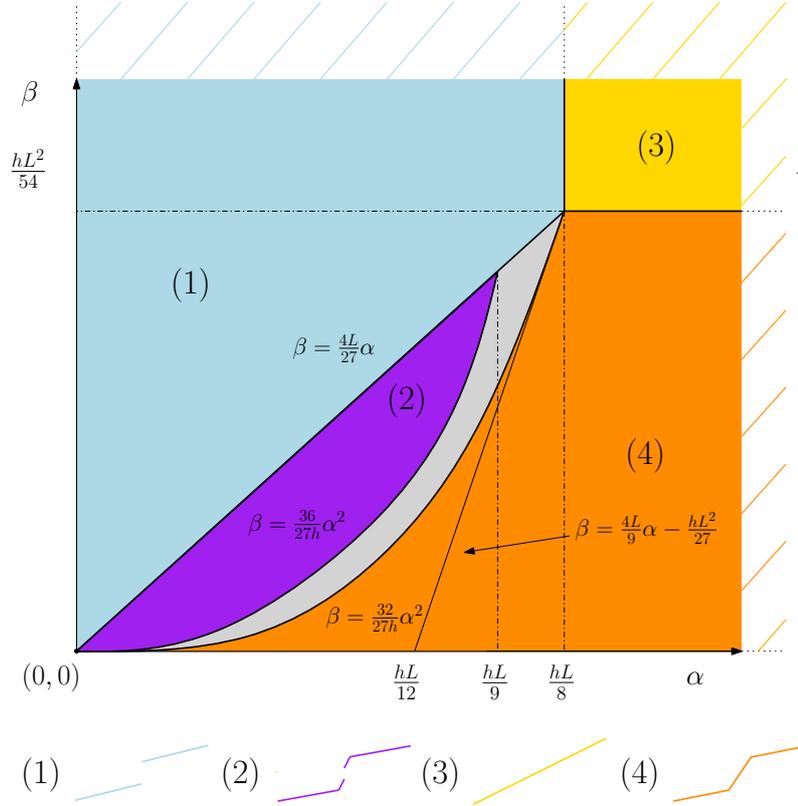
Figure 4.8: Illustration of the four different types of solutions of $\mathcal{P}$ for the piecewise constant function $f_{\text{p.c.}}$ that result for different combinations of the parameters α and β.

    4.7(b).

(3) Yellow colour: $\left\{\alpha \geq \frac{hL}{8},\ \beta \geq \frac{hL^2}{54}\right\}$, (necessary and sufficient condition), **N-P1-C**, see Figure 4.7(c).

(4) Orange colour: $\left\{\beta \leq \frac{32}{27h}\alpha^2,\ \beta < \frac{hL^2}{54}\right\}$, (sufficient condition), **N-P2-C**, see Figure 4.7(d).

    Notice that according to Proposition 4.6.6 our initial sufficient conditions for the solution of the type N-P2-C to happen were $\beta < \frac{hL^2}{54}$ and $\beta \leq \frac{4L}{9}\alpha - \frac{L^2h}{27}$. However, according to the remark after Proposition 4.6.4, page 140, when condition $\beta \leq \frac{32}{27h}\alpha^2$ holds then the solution of the type N-P2-J cannot happen and since the conditions for N-P1-C and N-P2-C are necessary and sufficient we conclude that when $\beta \leq \frac{32}{27h}\alpha^2$ holds (together with $\alpha \leq \frac{hL}{8}$) then the solution of the type N-P2-C occur. Notice that

$$\left\{\alpha > 0,\ \beta > 0:\ \beta \leq \frac{4L}{9}\alpha - \frac{L^2h}{27},\ \alpha \leq \frac{hL}{8}\right\} \subseteq \left\{\alpha > 0,\ \beta > 0:\ \beta \leq \frac{32}{27h}\alpha^2,\ \alpha \leq \frac{hL}{8}\right\},$$

see Figure 4.8.

    As we have mentioned, due to computational issues it is difficult to derive necessary

and sufficient conditions for the solutions of the type N-P2-J and N-P2-C. However, the estimates we have provided are not far away from being sharp as the unknown grey area in Figure 4.8, $\{\beta > \frac{32}{27h}\alpha^2,\ \beta < \frac{36}{27h}\alpha^2,\ \beta < \frac{4L}{27}\alpha\}$ is relatively small.

4.6.2 Piecewise affine function with a single jump

In this section we choose the data function to be a simple piecewise affine function $f_{\text{p.a.}}$, defined as

$$f_{\text{p.a.}}(x) = \begin{cases} \lambda(x - L) & \text{if } x \in (0, L), \\ \lambda(x - L) + h & \text{if } x \in [L, 2L), \end{cases} \tag{4.125}$$

for some $h > 0$ and $\lambda \in \mathbb{R}$, see also Figure 4.9.
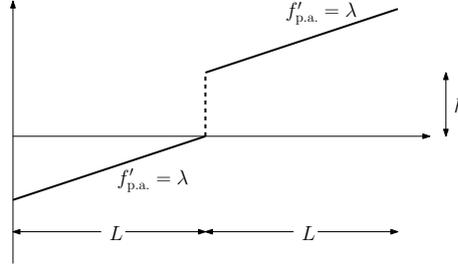
Figure 4.9: Piecewise affine function $f_{\text{p.c.}}$ with a jump discontinuity at $x = L$.

Note that $f_{\text{p.c.}}$ is a special case of $f_{\text{p.a.}}$ where $\lambda = 0$. However, we decided to treat the $f_{\text{p.c.}}$ case as a separate one since as the next proposition shows the solutions of $\mathcal{P}$ with data $f_{\text{p.a.}}$ can be obtained from the ones that correspond to $f_{\text{p.c.}}$. The latter ones are more easy to compute since $f_{\text{p.c.}}$ has a simpler form. The following Proposition essentially states that if an affine function is added to the data f then the solution u of $\mathcal{P}$ is also shifted by the same affine function.

Proposition 4.6.7. *The function $u_{f_{\text{p.c.}}}$ is a solution to the minimisation problem $\mathcal{P}$ with data function $f_{\text{p.c.}}$ if and only if $u_{f_{\text{p.a.}}}$ is a solution to $\mathcal{P}$ with data function $f_{\text{p.a.}}$, where*

$$u_{f_{\text{p.a.}}}(x) = u_{f_{\text{p.c.}}}(x) + \lambda(x - L).$$

Proof. Suppose that $u_{f_{\text{p.c.}}}$ is a solution of $\mathcal{P}$ with data $f_{\text{p.c.}}$, for some combination of the parameters α and β and let $v_{f_{\text{p.c.}}}$, $w_{f_{\text{p.c.}}}$ be the corresponding v and w variables. We will show that $u_{f_{\text{p.a.}}}(x) = u_{f_{\text{p.c.}}}(x) + \lambda(x - L)$ is a solution for data $f_{\text{p.a.}}$ for the same

combination of α and β and vice versa. The optimality conditions (C_f), (C_α), (C_β) read:

$$v''_{f_{\text{p.c.}}} = f - u_{f_{\text{p.c.}}},$$
$$-v'_{f_{\text{p.c.}}} \in \alpha \operatorname{Sgn}(Du_{f_{\text{p.c.}}} - w_{f_{\text{p.c.}}}),$$
$$v_{f_{\text{p.c.}}} \in \beta \operatorname{Sgn}(Dw_{f_{\text{p.c.}}}).$$

Observing that $f_{\text{p.a.}}(x) = f_{\text{p.c.}}(x) + \lambda(x - L)$ we set $v_{f_{\text{p.a.}}} = v_{f_{\text{p.c.}}}$ and $w_{f_{\text{p.a.}}}(x) = w_{f_{\text{p.c.}}}(x) + \lambda$. Then we have

$$v''_{f_{\text{p.a.}}} = v''_{f_{\text{p.c.}}} = f_{\text{p.c.}} - u_{f_{\text{p.c.}}} = f_{\text{p.a.}} - u_{f_{\text{p.a.}}},$$

and

$$-v'_{f_{\text{p.a.}}} \in \alpha \operatorname{Sgn}(Du_{f_{\text{p.a.}}} - w_{f_{\text{p.a.}}}) \iff -v'_{f_{\text{p.a.}}} \in \alpha \operatorname{Sgn}(Du_{f_{\text{p.c.}}} + \lambda - w_{f_{\text{p.c.}}} - \lambda)$$
$$\iff -v'_{f_{\text{p.c.}}} \in \alpha \operatorname{Sgn}(Du_{f_{\text{p.c.}}} - w_{f_{\text{p.c.}}}).$$

Finally,

$$v_{f_{\text{p.a.}}} \in \beta \operatorname{Sgn}(Dw_{f_{\text{p.a.}}}) \iff v_{f_{\text{p.a.}}} \in \beta \operatorname{Sgn}(Dw_{f_{\text{p.c.}}}) \iff v_{f_{\text{p.c.}}} \in \beta \operatorname{Sgn}(Dw_{f_{\text{p.c.}}}),$$

thus (C_f), (C_α), (C_β) hold for $u_{f_{\text{p.a.}}}$, $v_{f_{\text{p.a.}}}$ and $w_{f_{\text{p.a.}}}$. We can similarly show that if $u_{f_{\text{p.a.}}}$ is a solution for data $f_{\text{p.a.}}$ then $u_{f_{\text{p.c.}}}(x) = u_{f_{\text{p.a.}}}(x) - \lambda(x - L)$ is a solution for data $f_{\text{p.c.}}$. $\quad\square$

(a) Solution of the type N-P1-J.

(b) Solution of the type N-P2-J.

(c) Solution of the type N-P1-C.

(d) Solution of the type N-P2-C.
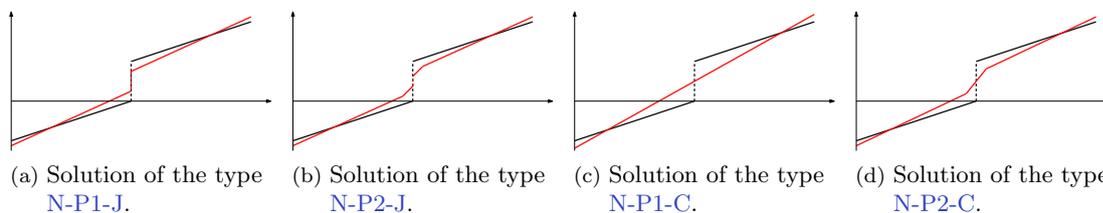
Figure 4.10: All the possible types of solution u to the minimisation problem $\mathcal{P}$ for the data function $f_{\text{p.a.}}$.

In Figure 4.10 we show all the possible type of solutions for the data $f_{\text{p.a.}}$. These solutions correspond to the same combinations of α and β that are shown in Figure 4.8. One can observe here the capability of TGV to preserve piecewise affine structures. This is in contrast to TV regularisation which promotes piecewise constant reconstructions. In fact, one can easily check using the optimality conditions of the L^2–TV problem in Table 4.1 that all TV solutions have the form shown in Figure 4.11.
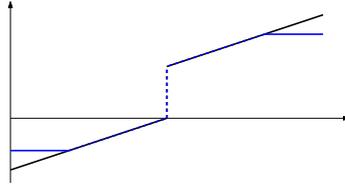
Figure 4.11: Type of solutions for the L^2–TV problem with data $f_{\mathrm{p.a.}}$.

4.6.3   Hat function

In this section we compute exact solutions for the hat function

$$f_{\mathrm{hat}}(x) = \lambda|x - L| - \lambda L,$$

where $\lambda > 0$, see Figure 4.12. The study of this case gives an insight about how TGV
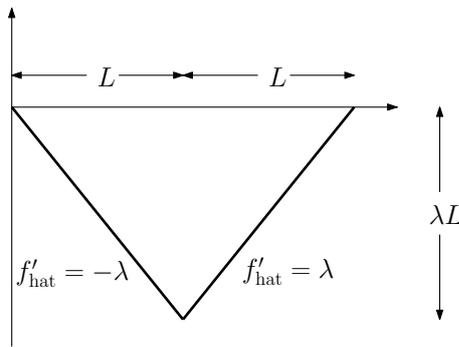
Figure 4.12: The hat function f_{hat}.

affects local extrema. Note that since f_{hat} is an even function, the solutions will be even as well. Working similarly as we did for the function $f_{\mathrm{p.c.}}$, we conclude that the only types of solutions u are the following[2]:

(i) u is constant in $(0, x_1)$, equal to f_{hat} in (x_1, x_2) and constant in (x_1, L),   (C-E-C)

(ii) u is affine in $(0, x_1)$, equal to f_{hat} in (x_1, x_2) and affine in (x_1, L),   (A-E-A)

(iii) u is constant in $(0, L)$,   (C)

(iv) u is affine in $(0, L)$,   (A)

where $0 < x_1 < x_2 < L$. In Figure 4.13 we show how these four types of solutions look like, along with the corresponding w and v variables. Notice that the solution of the type C is the L^2–linear regression of f_{hat}. Note also, that the solutions of the type C-E-C correspond to the TV solutions predicted by Theorem 4.4.11. We finally note that the

---

[2]C: constant, E: equal to f_{hat}, A: affine.

solution of the type **A** is also computed in [Ben11, Mül13, BBBM13] in order to indicate that f_{hat} can be recovered by TGV regularisation up to a loss of contrast.

(a) Solution of the type C-E-C.

(b) Solution of the type A-E-A.

(c) Solution of the type C.

(d) Solution of the type A.
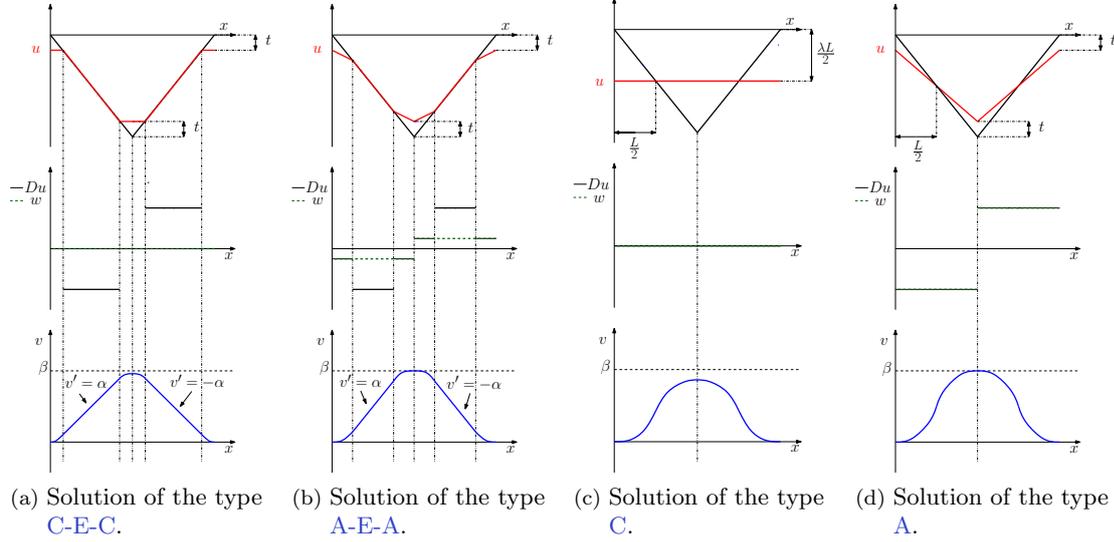
Figure 4.13: All the possible types of solution u to the minimisation problem $\mathcal{P}$ for the data function f_{hat}. We also show the forms of the corresponding variables w and v.

In the following, we investigate which combinations of the parameters α and β correspond to each type of solution. Unlike the case with the function $f_{\text{p.c.}}$, here we are able to give necessary and sufficient conditions for all the types of solutions. We summarise our results in the following proposition.

Proposition 4.6.8. *Consider the problem $\mathcal{P}$ with data function f_{hat}. Then we have the following four cases:*

(i) The solution u is of the type **C-E-C**, *Figure 4.13(a), if and only if*

$$\alpha < \frac{\lambda L^2}{8} \quad and \quad \beta \geq \alpha L - \frac{2\alpha}{3}\sqrt{\frac{2\alpha}{\lambda}}. \tag{4.126}$$

In that case

$$u(x) = \begin{cases} -\sqrt{2\alpha\lambda} & if \ x \in \left(0, \sqrt{\frac{2\alpha}{\lambda}}\right), \\ -\lambda x & if \ x \in \left(\sqrt{\frac{2\alpha}{\lambda}}, L - \sqrt{\frac{2\alpha}{\lambda}}\right), \\ -\lambda L + \sqrt{2\alpha\lambda} & if \ x \in \left(L - \sqrt{\frac{2\alpha}{\lambda}}), L\right). \end{cases}$$

(ii) The solution u is of the type **A-E-A**, *Figure 4.13(b), if and only if*

$$\beta > \frac{2L\alpha}{3} \quad and \quad \beta < \alpha L - \frac{2\alpha}{3}\sqrt{\frac{2\alpha}{\lambda}}. \tag{4.127}$$

In that case

$$u(x) = \begin{cases} -\mu x - \sqrt{2\alpha(\lambda - \mu)} & \text{if } x \in \left(0, \frac{3}{2}\left(L - \frac{\beta}{\alpha}\right)\right), \\ -\lambda x & \text{if } x \in \left(\frac{3}{2}\left(L - \frac{\beta}{\alpha}\right), L - \frac{3}{2}\left(L - \frac{\beta}{\alpha}\right)\right), \\ -\mu x - (\lambda - \mu)L + 2\sqrt{2\alpha(\lambda - \mu)} & \text{if } x \in \left(L - \frac{3}{2}\left(L - \frac{\beta}{\alpha}\right), L\right), \end{cases}$$

with

$$\mu = \lambda - \frac{8\alpha}{9\left(L - \frac{\beta}{\alpha}\right)^2}.$$

(iii) The solution u is of the type **C***, Figure 4.13(c), if and only if*

$$\beta \geq \frac{\lambda L^3}{12} \quad \text{and} \quad \alpha \geq \frac{\lambda L^2}{8}. \tag{4.128}$$

In that case

$$u(x) = \frac{\lambda L}{2}, \quad x \in (0, 2L).$$

(iv) The solution u is of the type **A***, Figure 4.13(d), if and only if*

$$\beta < \frac{\lambda L^3}{12} \quad \text{and} \quad \beta \leq \frac{2L\alpha}{3}. \tag{4.129}$$

In that case

$$u(x) = \mu|x - L| - (\lambda L - t), \quad x \in (0, 2L)$$

with

$$\mu = \lambda - \frac{12\beta}{L^3} \quad \text{and} \quad t = \frac{6\beta}{L^2}.$$

Proof. (i) Since we are looking for solutions of the type C-E-C, we must find $0 < x_1 < x_2 < L$ such that $u = t$ for a constant t in $(0, x_1)$, $u = f_{\text{hat}}$ in (x_1, x_2) and $u = -t - \lambda(x_2 - x_1)$ in (x_2, L). As far as the variable w is concerned, we have that $w = Du = 0$ in $(0, x_1) \cup (x_2, L)$. However since $u = f_{\text{hat}}$ in (x_1, x_2), we have that v is affine in that interval, thus it cannot have an extremum there. Thus, condition (C_β) forces $w = 0$ in (x_1, x_2) as well and (C_α) forces $v' = \alpha$ there. We conclude that the solution u will be of the type C-E-C if and only if we can find $0 < x_1 < x_2 < L$ and a function v with the following properties:

$$v \text{ and } v' \text{ are continuous}, \quad (v \in H_0^2), \tag{4.130}$$

$$v \text{ is a cubic polynomial on } (0, x_1) \text{ and } (x_2, L), \quad (\text{condition } (C_f)), \tag{4.131}$$

$$v(0) = 0, \; v'(0) = 0, \quad (\text{boundary conditions for } v), \tag{4.132}$$

$$v \text{ is affine and } v' = \alpha \text{ on } (x_1, x_2), \quad (\text{as explained above}), \tag{4.133}$$

$$v'(L) = 0, \quad (v \text{ is an even function}), \tag{4.134}$$

$$|v| \leq \beta, \quad \text{(condition } (C_\beta)), \tag{4.135}$$

$$|v'| \leq \alpha, \quad \text{(condition } (C_\alpha)). \tag{4.136}$$

After some computations we find that

$$x_1 = L - x_2 = \sqrt{\frac{2\alpha}{\lambda}}, \quad t = \sqrt{2\alpha\lambda},$$

$$v(x) = \begin{cases} -\frac{\lambda}{6}x^3 + \sqrt{\frac{\alpha\lambda}{2}}x^2 & \text{if } x \in \left(0, \sqrt{\frac{2\alpha}{\lambda}}\right), \\ \alpha(x - x_1) + \frac{\lambda x_1^2}{3} & \text{if } x \in \left(\sqrt{\frac{2\alpha}{\lambda}}, L - \sqrt{\frac{2\alpha}{\lambda}}\right), \\ -\frac{\lambda}{6}(x + x_1 - L)^3 + \alpha(x + x_1 - L) \\ +\alpha(L - 2x_1) + \frac{\lambda x_1^3}{3} & \text{if } x \in \left(L - \sqrt{\frac{2\alpha}{\lambda}}, L\right). \end{cases}$$

From the equation $v'' = f_{\text{hat}} - u$ we can get an expression for u. Moreover $x_1 < x_2$ holds if and only if

$$\alpha < \frac{\lambda L^2}{8}.$$

Finally one can check that (4.136) holds and in order to satisfy (4.135), since v is increasing in $(0, L)$, it suffices to have $v(L) \leq \beta$ something that translates to

$$\beta \geq \alpha L - \frac{2\alpha}{3}\sqrt{\frac{2\alpha}{\lambda}}.$$

(ii) The proof follows essentially the proof of (i). Here we are looking for solutions of the type $u = -\mu x - t$ instead of $u = t$ on $(0, x_1)$. In addition to the conditions (4.130)–(4.136) for v here we also have

$$v(L) = \beta, \qquad \text{(condition } (C_\beta)), \tag{4.137}$$

because w makes a positive jump at $x = L$, see also Figure 4.13(b). Again after some computations we find

$$x_1 = L - x_2 = \sqrt{\frac{2\alpha}{\lambda - \mu}}, \quad t = \sqrt{2\alpha(\lambda - \mu)},$$

$$v(x) = \begin{cases} -\frac{\lambda-\mu}{6}x^3 + \sqrt{\frac{\alpha(\lambda-\mu)}{2}}x^2 & \text{if } x \in \left(0, \sqrt{\frac{2\alpha}{\lambda-\mu}}\right), \\ \alpha(x - x_1) + \frac{(\lambda-\mu)x_1^2}{3} & \text{if } x \in \left(\sqrt{\frac{2\alpha}{\lambda-\mu}}, L - \sqrt{\frac{2\alpha}{\lambda-\mu}}\right), \\ -\frac{\lambda-\mu}{6}(x + x_1 - L)^3 + \alpha(x + x_1 - L) \\ +\alpha(L - 2x_1) + \frac{(\lambda-\mu)x_1^3}{3} & \text{if } x \in \left(L - \sqrt{\frac{2\alpha}{\lambda-\mu}}, L\right). \end{cases}$$

Again one can check that $|v'| < \alpha$. Moreover the condition $v(L) = \beta$ gives

$$\alpha L - \frac{2\alpha}{3}\sqrt{\frac{2\alpha}{\lambda - -\mu}} = \beta \iff \mu = \lambda - \frac{8\alpha}{9\left(L - \frac{\beta}{\alpha}\right)^2}. \tag{4.138}$$

Since we are looking into cases where $0 < \mu < \lambda$ we must have $\sqrt{\frac{2\alpha}{\lambda - \mu}} > \sqrt{\frac{2\alpha}{\lambda}}$ and using (4.138) this translates to

$$\beta < \alpha L - \frac{2\alpha}{3}\sqrt{\frac{2\alpha}{\lambda}}. \tag{4.139}$$

Finally, it is easily checked that in order to impose $x_1 < x_2$, we must have

$$\beta > \frac{2L\alpha}{3}. \tag{4.140}$$

(iii) In this case, we are looking for solutions of the type $u = -t$, $t > 0$. The function v will be a cubic polynomial that satisfy the conditions:

$$v(0) = 0, \quad v'(0) = 0, \quad v'(L) = 0, \quad v'' = h + t, \quad |v| \leq \beta, \quad |v'| \leq \alpha.$$

We easily compute

$$t = \frac{\lambda L}{2}, \quad v(x) = -\frac{\lambda}{6}x^3 + \frac{1}{2}tx^2, \quad x \in (0, L).$$

We can also check that the conditions $|v| \leq \beta$, $|v'| \leq \alpha$ are equivalent to

$$\beta \geq \frac{\lambda L^3}{12}, \quad \alpha \geq \frac{\lambda L^2}{8},$$

respectively.

(iv) The proof is similar to (iii). We are looking for a solution of the type $u(x) = -\mu x - t$, $x \in (0, L)$, $0 < \mu < \lambda$ and $t > 0$. As before, the function v will be a cubic polynomial satisfying the conditions:

$$v(0) = 0, \quad v'(0) = 0, \quad v(L) = \beta, \quad v'(L) = 0, \quad v'' = h + t, \quad |v| \leq \beta, \quad |v'| \leq \alpha.$$

We get that

$$v(x) = -\frac{\lambda - \mu}{6}x^3 + \frac{t}{2}x^2, \quad x \in (0, L), \quad \text{with} \quad t = \frac{6\beta}{L^2}, \quad \mu = \lambda - \frac{12\beta}{L^3}.$$

Thus $0 < \mu < \lambda$ is equivalent to

$$\beta < \frac{\lambda L^3}{12}.$$

Finally, we check easily that $|v| \leq \beta$ holds and $|v'| \leq \alpha$ is equivalent to

$$\beta \leq \frac{2L\alpha}{3}.$$

$\square$
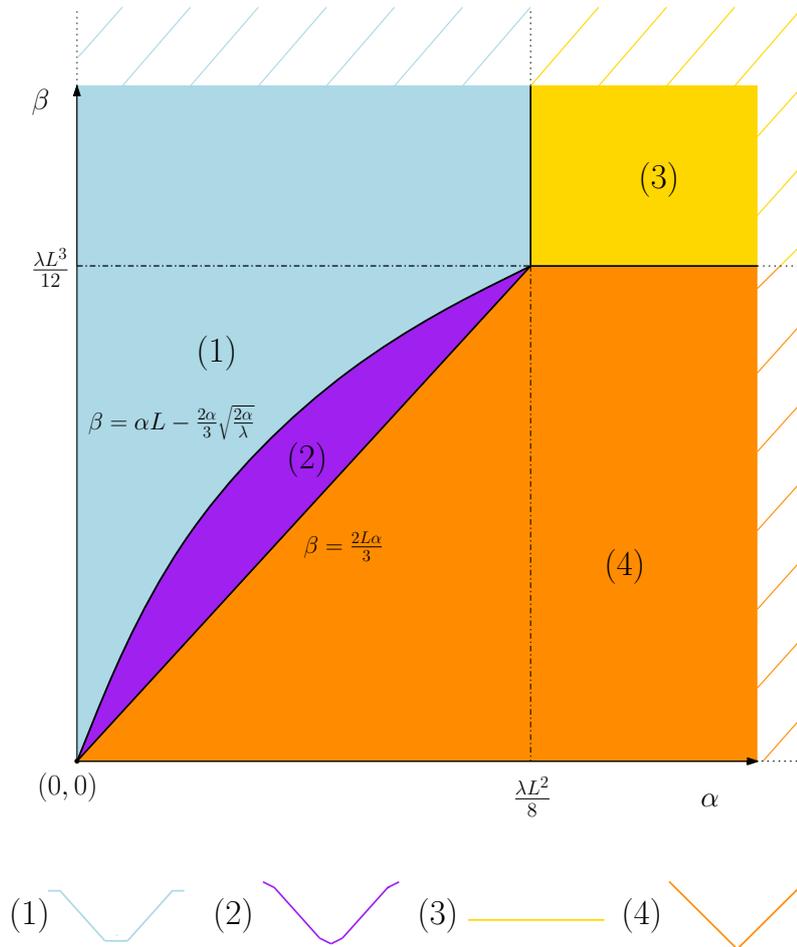
Figure 4.14: Illustration of the four different types of solutions of $\mathcal{P}$ for the hat function f_{hat} that result for different combinations of the parameters α and β.

As we did for the function $f_{\text{p.c.}}$ we summarise how the solutions of the problem $\mathcal{P}$ with data f_{hat} are determined by α and β. In Figure 4.14 we have partitioned again the set $\{\alpha > 0, \ \beta > 0\}$ into different areas that correspond to the four different possible solutions. We note again that in contrast to the cases of the functions $f_{\text{p.c.}}$ and $f_{\text{p.a.}}$, here we provide both necessary and sufficient conditions for all the four different types of solutions. The corresponding areas are:

(1) Blue colour: $\left\{ \alpha < \frac{\lambda L^2}{8}, \ \beta \geq \alpha L - \frac{2\alpha}{3}\sqrt{\frac{2\alpha}{\lambda}} \right\}$, **C-E-C**, see Figure 4.13(a).

(2) Purple colour: $\left\{ \beta > \frac{2L\alpha}{3}, \ \beta < \alpha L - \frac{2\alpha}{3}\sqrt{\frac{2\alpha}{\lambda}} \right\}$, **A-E-A**, see Figure 4.13(b).

(3) Yellow colour: $\left\{ \alpha \geq \frac{\lambda L^2}{8}, \ \beta \geq \frac{\lambda L^3}{12} \right\}$, **C**, see Figure 4.13(c).

(4) Orange colour: $\left\{ \beta < \frac{\lambda L^3}{12}, \ \beta \leq \frac{2L\alpha}{3} \right\}$, **A**, see Figure 4.13(d).

4.6.4 Numerical experiments

In this final section, we compare our theoretical results with numerical ones obtained by solving the discrete version of $\mathcal{P}$ with the primal-dual algorithm of Chambolle-Pock, [CP11]. A description of the algorithm for TGV minimisation can be found in [Bre14]. We also compute some numerical results with the presence of Gaussian noise that show that TGV regularisation is quite robust for noisy data. Let us note here that even though some sensitivity analysis can be done [BV11], it is not an easy task to prove that the corresponding solutions of $\mathcal{P}$ with clean and corrupted data have the same structure provided the noise is sufficiently small. Some relevant work has been done in [BB13] in terms of ground states of regularisation functionals.

In Figure 4.15, we plot the exact and the numerical solutions of $\mathcal{P}$ using the function $f_{\text{p.c.}}$ as a data function. We choose the parameters α and β so that we have the solution of the type N-P1-J. We observe that for clean data, the exact and the numerical solutions coincide, see Figure 4.15(a). Moreover, even under the presence of noise, the numerical solution is not far away from the corresponding solution without the noise, Figure 4.15(b).

We observe similar results in Figures 4.16 and 4.17 where we choose the parameters so that we have solutions of the type N-P2-J and N-P1-C respectively.

In Figure 4.18 we use the function f_{hat} as a data function. Again, without the presence of noise, the exact solution agrees with the numerical one, Figure 4.18(a). However, when noise is added, even though the numerical solution is close to the one that corresponds to the clean data, some staircasing is observed, Figure 4.18(b). This is not surprising as with these combinations of α and β, TGV behaves like TV as it was shown in Theorem 4.4.11.

In Figure 4.19 the parameters are chosen so that the solution of the type A-E-A occurs. In the clean data case we have agreement between the exact and the numerical solution, Figure 4.19(a), but in the noisy case a kind of "affine" staircasing effect appears in the area where the exact solution equals with the data, see detail of Figure 4.19(b).

Finally, in Figure 4.20, the parameters are chosen so the solution of the type A occurs. Again, the numerical solution agrees with the exact one and deviates from it slightly in the presence of noise, while no staircasing of any kind appears this time.

(a) Clean data, the numerical solution agrees with the exact one.

(b) Noisy data, the numerical solution deviates slightly from the corresponding exact solution with clean data.
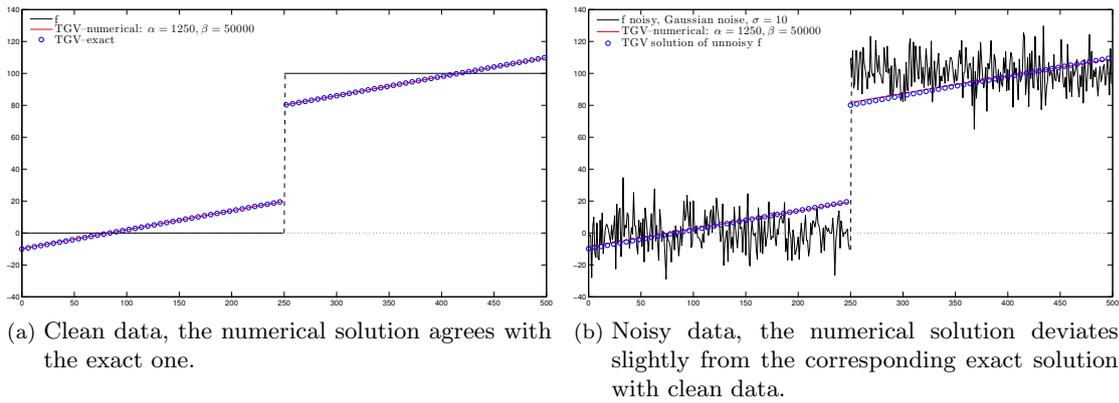
Figure 4.15: Piecewise constant function $f_{\mathrm{p.c.}}$. The parameters α and β satisfy the conditions (4.63), thus the solution is of the type N-P1-J.

(a) Clean data, the numerical solution agrees with the exact one.

(b) Noisy data, the numerical solution deviates slightly from the corresponding exact solution with clean data.
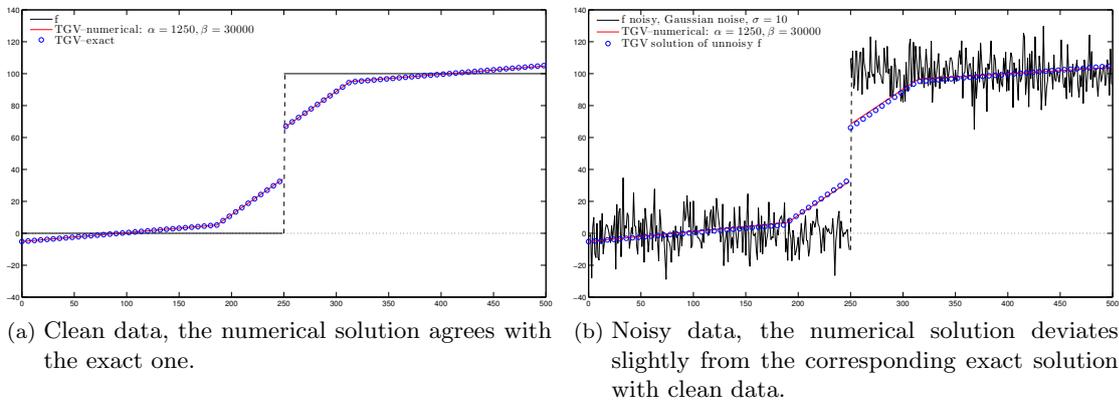
Figure 4.16: Piecewise constant function $f_{\mathrm{p.c.}}$. The parameters α and β satisfy the conditions (4.72), thus the solution is of the type N-P2-J.

(a) Clean data, the numerical solution agrees with the exact one.

(b) Noisy data, the numerical solution deviates slightly from the corresponding exact solution with clean data.
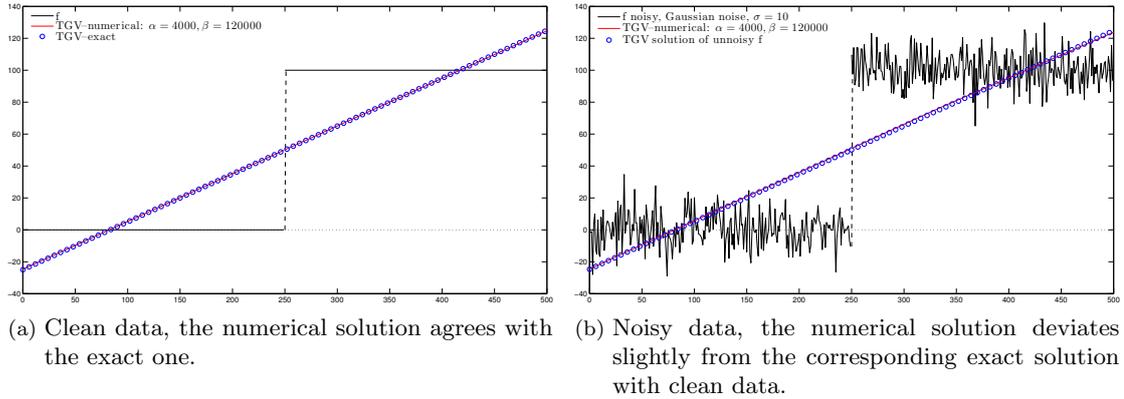
Figure 4.17: Piecewise constant function $f_{\text{p.c.}}$. The parameters α and β satisfy the conditions (4.94), thus the solution is of the type N-P1-C (L^2–linear regression).

(a) Clean data, the numerical solution agrees with the exact one.

(b) Noisy data, appearance of staircasing effect.
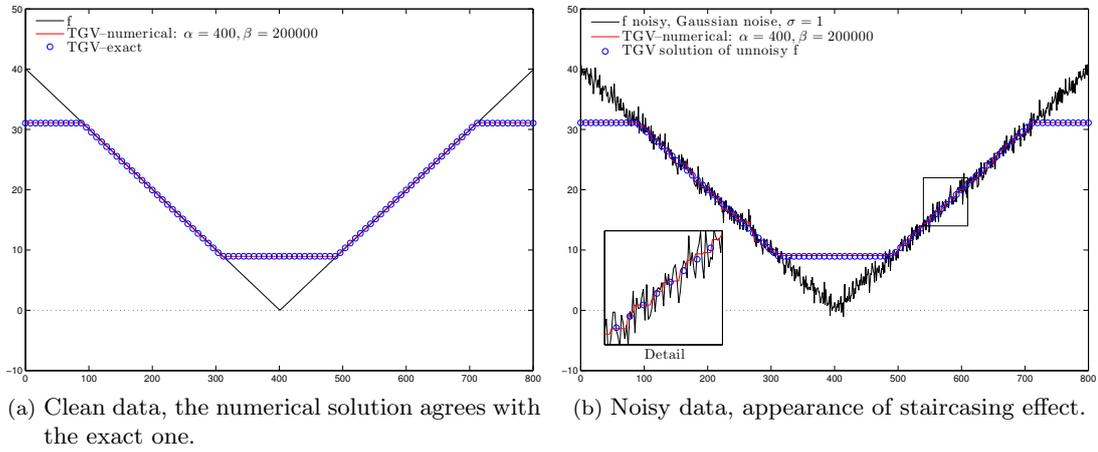
Figure 4.18: Hat function f_{hat}. The parameters α and β satisfy the conditions (4.126), thus the solution is of the type C-E-C.

(a) Clean data, the numerical solution agrees with the exact one.

(b) Noisy data, appearance of an "affine" staircasing effect.
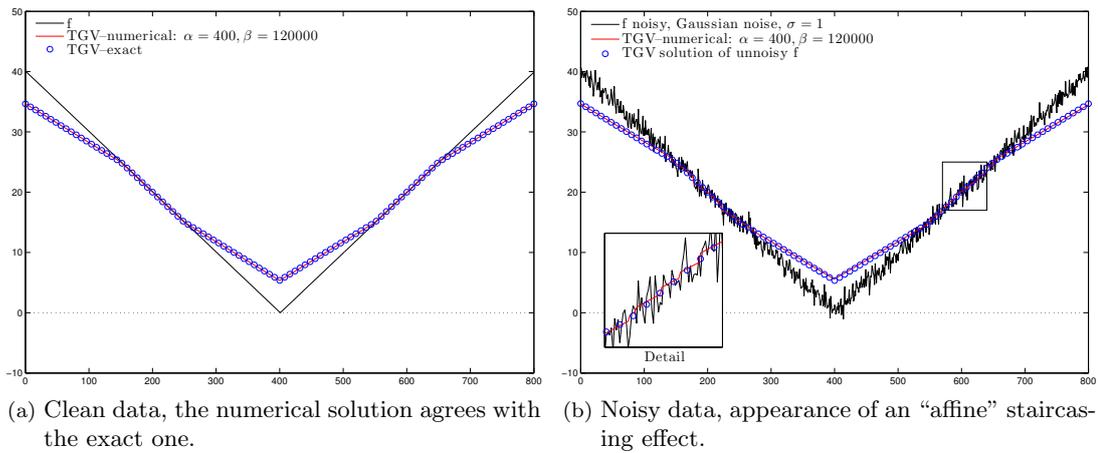
Figure 4.19: Hat function f_{hat}. The parameters α and β satisfy the conditions (4.127), thus the solution is of the type A-E-A.

(a) Clean data, the numerical solution agrees with the exact one.

(b) Noisy data, the numerical solution deviates slightly from the corresponding exact solution with clean data.
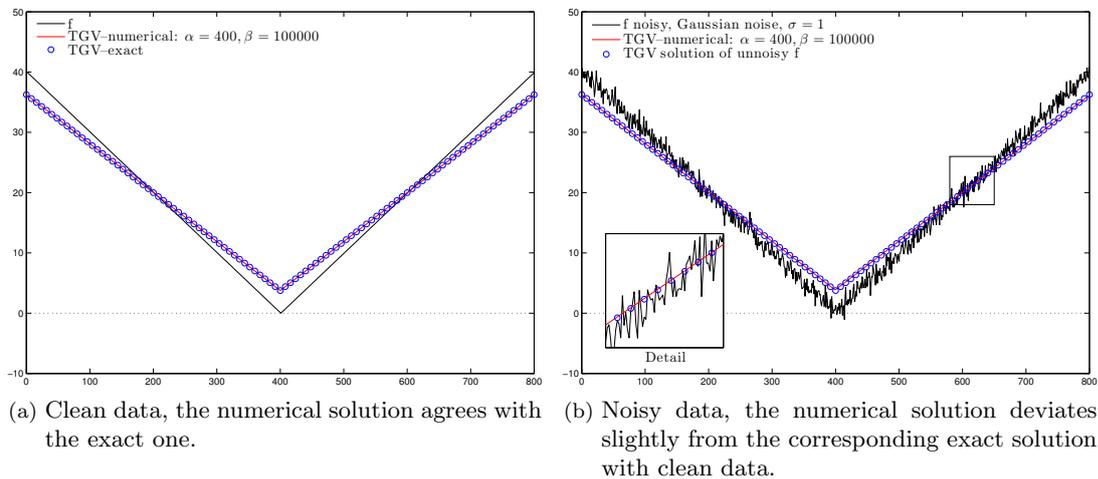
Figure 4.20: Hat function f_{hat}. The parameters α and β satisfy the conditions (4.129), thus the solution is of the type A.

156

Chapter 5

Non-local Hessian: Localisation results, characterisation of higher order Sobolev and BV spaces and applications

5.1   Introduction

In this chapter we are concerned with two different forms of non-local Hessian functionals. For the first formulation (explicit formulation) we prove its localisation to the classical Hessian as the non-locality vanishes and we use it to characterise some higher order Sobolev and BV spaces. The second formulation (implicit formulation) that we introduce, is better tuned for use in variational problems in imaging. We provide some numerical examples for image denoising where we show that the model is capable of outperforming TGV as far as the restoration of piecewise affine data is concerned.

Explicit formulation

We are initially interested in studying the following non-local Hessian functional (explicit formulation):

$$\mathcal{H}_n u(x) = \frac{d(d+2)}{2} \int_{\mathbb{R}^N} \frac{d^2 u(x,y)}{|x-y|^2} \frac{\left((x-y) \otimes (x-y) - \frac{|x-y|^2}{d+2} I_d\right)}{|x-y|^2} \rho_n(x-y) dy, \quad x \in \mathbb{R}^d, \tag{5.1}$$

with

$$d^2 u(x,y) := u(y) - 2u(x) + u(x + (x-y)),$$

157

see Figure 5.1, where here I_d is the $d \times d$ identity matrix and ρ_n is a sequence of $L^1(\mathbb{R}^d)$ radial functions, i.e., $\rho_n(x) = \rho_n(|x|)$, that satisfy the properties

$$\rho_n \geq 0, \quad \int_{\mathbb{R}^d} \rho_n dx = 1, \quad \lim_{n \to \infty} \int_{|x| > \delta} \rho_n dx = 0, \;\; \forall \delta > 0. \qquad (5.2)$$

This means that eventually all the mass of ρ_n is concentrating to the origin, leading finally to the localisation of (5.1).
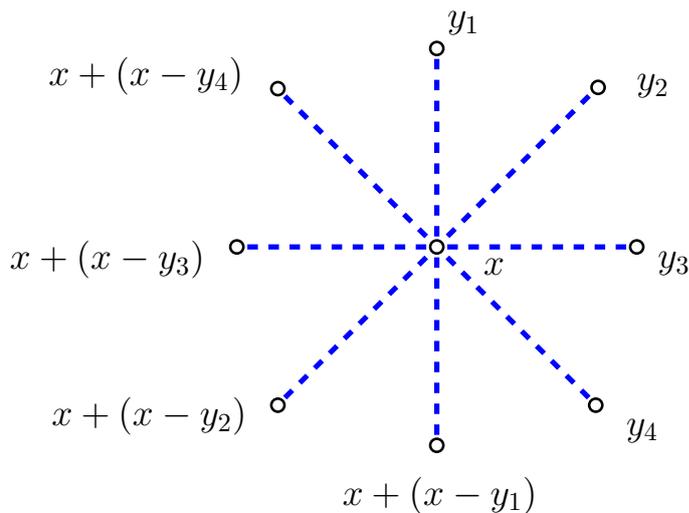
Figure 5.1: Configuration of the second order finite difference scheme $d^2 u(x, y) = u(y) - 2u(x) + u(x + (x - y))$.

As we have already mentioned in the introduction, analogous first order non-local functionals have been introduced and studied in the literature. More particularly, Bourgain, Brezis and Mironescu [BBM01] examined functionals of the type

$$\int_\Omega \int_\Omega \frac{|u(x) - u(y)|^p}{|x - y|^p} \rho_n(x - y) dx dy, \qquad (5.3)$$

where Ω is a smooth bounded domain in $\mathbb{R}^d$ and $1 \leq p < \infty$. They characterised $W^{1,p}(\Omega)$ and $BV(\Omega)$ by proving that,

$$u \in W^{1,p}(\Omega) \iff u \in L^p(\Omega) \text{ and } \liminf_{n \to \infty} \int_\Omega \int_\Omega \frac{|u(x) - u(y)|^p}{|x - y|^p} \rho_n(x - y) dx dy < \infty, \;\; p > 1,$$

$$u \in BV(\Omega) \iff u \in L^1(\Omega) \text{ and } \liminf_{n \to \infty} \int_\Omega \int_\Omega \frac{|u(x) - u(y)|}{|x - y|} \rho_n(x - y) dx dy < \infty, \;\; p = 1.$$

Ponce in [Pon04] derived similar characterisations by studying functionals of the type

$$\int_\Omega \int_\Omega \omega \left(\frac{|u(x) - u(y)|}{|x - y|} \right) \rho_\epsilon(x - y) dx dy, \qquad (5.4)$$

where ω is a continuous function and ρ_ϵ are not necessarily radial. Analogous were the results by Gobbino and Mora in [GM01].

Mengesha and Spector in [MS13] introduced the following non-local gradient operator:

$$\mathcal{G}_n u(x) = d \int_\Omega \frac{u(x) - u(y)}{|x-y|} \frac{x-y}{|x-y|} \rho_n(x-y) dy, \quad x \in \Omega. \tag{5.5}$$

Functional (5.5) is defined rigorously as a distribution. The authors proved the localisation of the functionals (5.5) to their classical analogue, ∇u, in the topology that corresponds to the regularity of the function u and they obtained yet another characterisation of the spaces $W^{1,p}(\Omega)$ and $\mathrm{BV}(\Omega)$:

$$u \in W^{1,p}(\Omega) \iff u \in L^p(\Omega) \quad \text{and} \quad \liminf_{n \to \infty} \|\mathcal{G}_n u\|_{L^p(\Omega)} < \infty, \quad p > 1$$

$$u \in \mathrm{BV}(\Omega) \iff u \in L^1(\Omega) \quad \text{and} \quad \liminf_{n \to \infty} \|\mathcal{G}_n u\|_{L^1(\Omega)} < \infty, \quad p = 1.$$

In this chapter, we prove the localisation of functional (5.1) to the classical Hessian $\nabla^2 u$ in various topologies and we obtain some novel characterisations of the spaces $W^{2,p}(\mathbb{R}^d)$ and $\mathrm{BV}^2(\mathbb{R}^d)$. In summary, we prove the following localisations:

$$\text{if } u \in \mathcal{C}_c^2(\mathbb{R}^d) \quad \implies \quad \mathcal{H}_n u \to \nabla^2 u \qquad\qquad\qquad \text{uniformly,}$$

$$\text{if } u \in W^{2,p}(\mathbb{R}^d) \quad \implies \quad \mathcal{H}_n u \to \nabla^2 u \qquad \text{in } L^p(\mathbb{R}^d, \mathbb{R}^{d \times d}), \quad 1 \leq p < \infty,$$

$$\text{if } u \in \mathrm{BV}^2(\mathbb{R}^d) \quad \implies \quad \mathcal{H}_n u \, \mathcal{L}^d \to D^2 u \qquad\qquad \text{weakly}^* \text{ in measures.}$$

We also prove the following derivative-free characterisations of $W^{2,p}(\mathbb{R}^d)$ and $\mathrm{BV}^2(\mathbb{R}^d)$ in the spirit of [BBM01] and [MS13]:

$$u \in W^{2,p}(\mathbb{R}^d) \iff u \in L^p(\mathbb{R}^d) \quad \text{and} \quad \liminf_{n \to \infty} \|\mathcal{H}_n u\|_{L^p(\mathbb{R}^d, \mathbb{R}^{d \times d})} < \infty, \quad p > 1, \tag{5.6}$$

$$u \in \mathrm{BV}^2(\mathbb{R}^d) \iff u \in L^1(\mathbb{R}^d) \quad \text{and} \quad \liminf_{n \to \infty} \|\mathcal{H}_n u\|_{L^1(\mathbb{R}^d, \mathbb{R}^{d \times d})} < \infty, \quad p = 1. \tag{5.7}$$

We must note here that the results in [MS13] hold for the vector value case as well, that is to say one can define a form of non-local Hessian as

$$\mathcal{G}_n(\nabla u)(x) = d \int_\Omega \frac{\nabla u(x) - \nabla u(y)}{|x-y|} \otimes \frac{x-y}{|x-y|} \rho_n(x-y) dy, \tag{5.8}$$

and obtain a straitforward characterisation of $W^{2,p}(\mathbb{R}^d)$ and $\mathrm{BV}^2(\mathbb{R}^d)$:

$$u \in W^{2,p}(\mathbb{R}^d) \iff u \in W^{1,p}(\mathbb{R}^d) \quad \text{and} \quad \liminf_{n \to \infty} \|\mathcal{G}_n \nabla u\|_{L^p(\mathbb{R}^d, \mathbb{R}^{d \times d})} < \infty, \ p > 1 \tag{5.9}$$

$$u \in \mathrm{BV}^2(\mathbb{R}^d) \iff u \in W^{1,1}(\mathbb{R}^d) \quad \text{and} \quad \liminf_{n \to \infty} \|\mathcal{G}_n \nabla u\|_{L^1(\mathbb{R}^d, \mathbb{R}^{d \times d})} < \infty, \ p = 1. \tag{5.10}$$

However the advantage of the characterisations (5.6)–(5.7) over (5.9)–(5.10) is that they

are derivative-free and require the function u to belong to an L^p space instead of $W^{1,p}$.

Implicit formulation

We then proceed to the definition of the implicit formulation of a non-local Hessian functional, see Definition 5.7.1. The non-local Hessian $\mathcal{H}_{\sigma_x} u(x)$ of a function u at a point x, is defined implicitly through a minimisation problem. We note that in this formulation the weighting function σ_x is not necessarily radial and depends heavily on the point x. It is constructed in such a way (by solving a weighted Eikonal equation) so that when x lies near an edge, points across the edge are weighted with small or zero weight something that eventually leads to an edge preservation scheme.

We solve the corresponding regularisation problem for the case of image denoising, where we observe that the model is able to recover to a very large degree piecewise affine data, outperforming TGV.

5.1.1   Organisation of the chapter

In Section 5.2 we prove the localisation of the sequence $(\mathcal{H}_n u)_{n \in \mathbb{N}}$ to the classical Hessian $\nabla^2 u$ for smooth, compactly supported functions u. This localisation occurs in the uniform topology.

In Section 5.3, after proving that $\mathcal{H}_n u$ is well defined for functions $u \in W^{2,p}(\mathbb{R}^d)$, we show that $(\mathcal{H}_n u)_{n \in \mathbb{N}}$ localises to the second order weak derivative $\nabla^2 u$.

A second order non-local integration by parts formula is shown in Section 5.4 where also a second order non-local divergence is introduced. This formula is an essential tool for proving the localisation of $(\mathcal{H}_n u)_{n \in \mathbb{N}}$ for the BV^2 case as well as for the non-local characterisation of BV$^2(\mathbb{R}^d)$.

In Section 5.6 we derive the non-local characterisations of $W^{2,p}(\mathbb{R}^d)$ and BV$^2(\mathbb{R}^d)$, after redefining the non-local Hessian functional as a distribution.

Finally, in Section 5.7 we introduce the implicit formulation of non-local Hessian and we provide some numerical examples in image denoising.

5.2   Localisation – The smooth case

Our target for this section is to show that at least for sufficiently smooth functions u, the sequence $(\mathcal{H}_n u)_{n \in \mathbb{N}}$ converges to the classical Hessian $\nabla^2 u$ uniformly. We first need the following lemma:

Lemma 5.2.1. *Let $c_n^{i,j}(x)$ be the following quantities for $i, j = 1, \ldots, d$:*

$$c_n^{i,j}(x) = (A(d))_{i,j} \int_{\mathbb{R}^d} \frac{(x_i - y_i)^2 (x_j - y_j)^2}{|x - y|^4} \rho_n(x - y) dy,$$

where

$$(A(d))_{i,j} = \begin{cases} d^2 + 2d & if \quad i \neq j, \\ \frac{d^2 + 2d}{3} & if \quad i = j. \end{cases}$$

Then $c_n^{i,j} = 1$ for every $n \in \mathbb{N}$ and $i,j = 1, \ldots, d$.

Proof. The proof follows directly from Lemma A.2.2 in Appendix A. □

We are now ready to prove the localisation of $\mathcal{H}_n u$ to $\nabla^2 u$ for sufficiently smooth functions:

Theorem 5.2.2. *Suppose that $u \in \mathcal{C}_c^2(\mathbb{R}^d)$. Then*

$$\mathcal{H}_n u \to \nabla^2 u, \quad uniformly \ as \ n \to \infty.$$

Proof. Note first that by using the mean value theorem for one dimensional restrictions of u we have

$$d^2 u(x,y) = (x-y)^T \left(\int_0^1 \int_0^1 \nabla^2 u(x + (t+s-1)(y-x)) ds dt \right) (x-y). \quad (5.11)$$

Define now the following quantity for $\delta > 0$:

$$Q_\delta u(x) = \left| \int_{B(x,\delta)} \frac{d^2 u(x,y) - (x-y)^T \nabla^2 u(x)(x-y)}{|x-y|^2} \frac{(x-y) \otimes (x-y)}{|x-y|^2} \rho_n(x-y) dy \right|. \tag{5.12}$$

We claim that we can make $Q_\delta u(x)$ as small as possible, independently of x, by choosing sufficiently small $\delta > 0$. Indeed, given $\epsilon > 0$, using the uniform continuity of the partial derivatives of u, we can choose $\delta > 0$ such that for every $i,j = 1, \ldots, d$ we have

$$\left| \frac{\partial u}{\partial x_i \partial x_j}(x) - \frac{\partial u}{\partial x_i \partial x_j}(y) \right| < \epsilon, \quad \text{whenever } |x-y| < \delta.$$

Using (5.11) we can estimate

$$
\begin{aligned}
Q_\delta u(x) &= \left| \int_{B(x,\delta)} \frac{d^2 u(x,y) - (x-y)^T \nabla^2 u(x)(x-y)}{|x-y|^2} \frac{(x-y) \otimes (x-y)}{|x-y|^2} \rho_n(x-y) dy \right| \\
&= \left| \int_{B(0,\delta)} \frac{z^T \left(\int_0^1 \int_0^1 \nabla^2 u(x + (1-s-t)z) - \nabla^2 u(x) ds dt \right) z}{|z|^2} \frac{z \otimes z}{|z|^2} \rho_n(z) dz \right| \\
&\leq d \int_{B(0,\delta)} \frac{|x-y| \epsilon |z|}{|z|^2} \frac{|z||z|}{|z|^2} \rho_n(z) dz \\
&\leq \epsilon d. \tag{5.13}
\end{aligned}
$$

On the other hand we have that $Q_\delta u(x)$ is also equal to

$$Q_\delta u(x) = \left| \int_{B(x,\delta)} \frac{d^2 u(x,y)}{|x-y|^2} \frac{(x-y) \otimes (x-y)}{|x-y|^2} \rho_n(x-y) dy \right.$$

$$\left. - \underbrace{\int_{B(x,\delta)} \frac{(x-y)^T \nabla^2 u(x)(x-y)}{|x-y|^2} \frac{(x-y) \otimes (x-y)}{|x-y|^2} \rho_n(x-y) dy}_{C} \right|, \qquad (5.14)$$

where C is an $d \times d$ matrix whose (i_0, j_0)–th element is equal to

$$C(i_0, j_0) = \sum_{i,j=1}^d \frac{\partial u}{\partial x_i \partial x_j}(x) \int_{B(0,\delta)} \frac{z_i z_j z_{i_0} z_{j_0}}{|z|^4} \rho_n(z) dz. \qquad (5.15)$$

However due to radial symmetry we have that the following integrals vanish:

$$\mathbf{i_0 = j_0} : \int_{B(0,\delta)} \frac{z_i z_{j_0}^3}{|z|^4} \rho_n(z) dz = 0, \quad \text{for } i \neq j_0,$$

$$\mathbf{i_0 = j_0} : \int_{B(0,\delta)} \frac{z_i z_j z_{j_0}^2}{|z|^4} \rho_n(z) dz = 0, \quad \text{for } i \neq j_0, \; j \neq j_0, \; i \neq j,$$

$$\mathbf{i_0 \neq j_0} : \int_{B(0,\delta)} \frac{z_i z_j z_{i_0} z_{j_0}}{|z|^4} \rho_n(z) dz = 0, \quad \text{for } i \neq i_0, \; j \neq j_0.$$

This means that when $i_0 \neq j_0$, the only integrals that are not zero in (5.15) correspond to the cases $i = i_0, \; j = j_0$ or $i = j_0, \; j = i_0$. Thus, using also the symmetry of the mixed derivatives, we have that

$$C(i_0, j_0) = 2 \frac{\partial u}{\partial x_{i_0} \partial x_{j_0}}(x) \int_{B(0,\delta)} \frac{z_{i_0}^2 z_{j_0}^2}{|z|^4} \rho_n(z) dz, \quad i_0 \neq j_0. \qquad (5.16)$$

Now, when $i_0 = j_0$ we have that the only integrals that are not zero in (5.15) correspond to the case $i = j$. This means that

$$C(i_0, j_0) = \sum_{i=1}^d \frac{\partial^2 u}{\partial x_i^2}(x) \int_{B(0,\delta)} \frac{z_i^2 z_{i_0}^2}{|z|^4} \rho_n(z) dz, \quad i_0 = j_0. \qquad (5.17)$$

We now proceed to the final steps of the proof. Since $c_n^{i,j} = 1$, in order to prove the theorem it suffices to prove that

$$\left| \mathcal{H}_n u - c_n \times \nabla^2 u \right| \to 0, \quad \text{uniformly as } n \to \infty.$$

where here "$\times$" denotes pointwise multiplication between matrices. We prove that the above quantity converges to zero uniformly, component-wise, i.e., $\left| \left(\mathcal{H}_n u - c_n \times \nabla^2 u \right)_{(i_0,j_0)} \right| \to$

0 by considering two cases $i_0 \neq j_0$ and $i_0 = j_0$.

Case $i_0 \neq j_0$: Combining, (5.13), (5.14) and (5.16) we have that for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$\left| \int_{B(0,\delta)} \frac{d^2 u(x, x+z)}{|z|^2} \frac{z_{i_0} z_{j_0}}{|z|^2} \rho_n(z) dz - 2 \frac{\partial u}{\partial x_{i_0} \partial x_{j_0}}(x) \int_{B(0,\delta)} \frac{z_{i_0}^2 z_{j_0}^2}{|z|^4} \rho_n(z) dz \right| \leq \epsilon. \quad (5.18)$$

Using (5.18), we have for large enough n

$$
\begin{aligned}
\left| \left(\mathcal{H}_n u - c_n \times \nabla^2 u \right)_{(i_0, j_0)} \right| &= \left| \frac{d(d+2)}{2} \int_{\mathbb{R}^d} \frac{d^2 u(x,y)}{|x-y|^2} \frac{(x_{i_0} - y_{i_0})(x_{j_0} - y_{j_0})}{|x-y|^2} \rho_n(x-y) dy \right. \\
&\qquad \left. - \frac{\partial u}{\partial x_{i_0} \partial x_{j_0}}(x) d(d+2) \int_{\mathbb{R}^d} \frac{z_{i_0}^2 z_{j_0}^2}{|z|^4} \rho_n(z) dz \right| \\
&\leq \frac{d(d+2)}{2} \epsilon + \frac{d(d+2)}{2} \left| \int_{|z| \geq \delta} \frac{d^2 u(x, x+z)}{|z|^2} \frac{z_{i_0} z_{j_0}}{|z|^2} \rho_n(z) dz \right| \\
&\qquad + \frac{d(d+2)}{2} \left| \frac{\partial u}{\partial x_i \partial x_j}(x) \right| \left| \int_{|z| \geq \delta} \frac{z_{i_0}^2 z_{j_0}^2}{|z|^4} \rho_n(z) dz \right| \\
&\leq \frac{d(d+2)}{2} \epsilon + \frac{d(d+2)}{2} \frac{4\|u\|_\infty}{\delta^2} \int_{|z| \geq \delta} \rho_n(z) dz \\
&\qquad + \frac{d(d+2)}{2} \|\nabla^2 u\|_\infty \int_{|z| \geq \delta} \rho_n(z) dz \\
&< d(d+2)\epsilon.
\end{aligned}
$$

Case $i_0 = j_0$:

This case is technically more difficult. Note that we cannot follow the corresponding proof for the $i_0 \neq j_0$ case as all the pure second derivatives appear in (5.17). However we can take advantage of the fact that $\frac{\partial^2 u}{\partial x_{i_0}^2}(x)$ appears with different weight than $\frac{\partial^2 u}{\partial x_i^2}(x)$, for $i \neq i_0$. Note that this is the reason for the correction $\frac{|x-y|^2}{d+2}$ in the diagonal elements of the non-local Hessian $\mathcal{H}_n u$. As before, combining, (5.13), (5.14) and (5.17) we have that for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$\left| \int_{B(0,\delta)} \frac{d^2 u(x, x+z)}{|z|^2} \frac{z_{i_0}^2}{|z|^2} \rho_n(z) dz - \sum_{i=1}^d \frac{\partial^2 u}{\partial x_i^2}(x) \int_{B(0,\delta)} \frac{z_i^2 z_{i_0}^2}{|z|^4} \rho_n(z) dz \right| \leq \epsilon. \quad (5.19)$$

By considering the inequality (5.19) for every $i \neq i_0$ and summing over all $i \neq i_0$ we get

$$\left| \sum_{\substack{i=1 \\ i \neq i_0}}^d \int_{B(0,\delta)} \frac{d^2 u(x, x+z)}{|z|^2} \frac{z_i^2}{|z|^2} \rho_n(z) dz - \sum_{\substack{i=1 \\ i \neq i_0}}^d \sum_{j=1}^d \frac{\partial^2 u}{\partial x_j^2}(x) \int_{B(0,\delta)} \frac{z_j^2 z_i^2}{|z|^4} \rho_n(z) dz \right| \leq (d-1)\epsilon. $$

$$(5.20)$$

Observe now that using Lemma 5.2.1, the second term in the lefthand side of (5.20) can be written as follows:

$$\sum_{\substack{i=1\\i\neq i_0}}^{d}\sum_{j=1}^{d}\frac{\partial^2 u}{\partial x_j^2}(x)\int_{B(0,\delta)}\frac{z_j^2 z_i^2}{|z|^4}\rho_n(z)dz = (d-1)\frac{\partial^2 u}{\partial x_{i_0}^2}(x)\int_{B(0,\delta)}\underbrace{\frac{z_i^2 z_{i_0}^2}{|z|^4}}_{i\neq i_0}\rho_n(z)dz$$

$$+\sum_{\substack{i=1\\i\neq i_0}}^{d}\frac{\partial^2 u}{\partial x_i^2}(x)\int_{B(0,\delta)}\frac{z_i^4}{|z|^4}\rho_n(z)dz$$

$$+\sum_{\substack{i=1\\i\neq i_0}}^{d}(d-2)\frac{\partial^2 u}{\partial x_i^2}(x)\int_{B(0,\delta)}\underbrace{\frac{z_j^2 z_i^2}{|z|^4}}_{j\neq i}\rho_n(z)dz$$

$$=\frac{d-1}{3}\frac{\partial^2 u}{\partial x_{i_0}^2}(x)\int_{B(0,\delta)}\frac{z_{i_0}^4}{|z|^4}\rho_n(z)dx$$

$$+\sum_{\substack{i=1\\i\neq i_0}}^{d}(d+1)\frac{\partial^2 u}{\partial x_i^2}(x)\int_{B(0,\delta)}\underbrace{\frac{z_j^2 z_i^2}{|z|^4}}_{j\neq i}\rho_n(z)dz. \quad (5.21)$$

Now, (5.20) combined with (5.21) give

$$\left|\frac{1}{d+1}\sum_{\substack{i=1\\i\neq i_0}}^{d}\int_{B(0,\delta)}\frac{d^2 u(x,x+z)}{|z|^2}\frac{z_i^2}{|z|^2}\rho_n(z)dz\right. \quad (5.22)$$

$$\left.-\frac{d-1}{3(d+1)}\frac{\partial^2 u}{\partial x_{i_0}^2}(x)\int_{B(0,\delta)}\frac{z_{i_0}^4}{|z|^4}\rho_n(z)dx-\sum_{\substack{i=1\\i\neq i_0}}^{d}\frac{\partial^2 u}{\partial x_i^2}(x)\int_{B(0,\delta)}\frac{z_{i_0}^2 z_i^2}{|z|^4}\rho_n(z)dz\right|\leq\frac{d-1}{d+1}\epsilon.$$

Using (5.19), (5.22) and the triangle inequality we can now vanish the terms $\frac{\partial^2 u}{\partial x_i^2}(x)$, for $i\neq i_0$:

$$\left|\int_{B(0,\delta)}\frac{d^2 u(x,x+z)}{|z|^2}\frac{z_{i_0}^2-\frac{1}{d+1}\sum_{\substack{i=1\\i\neq i_0}}^{d}z_i^2}{|z|^2}\rho_n(z)dy\right.$$

$$\left.-\frac{2d+4}{3(d+1)}\frac{\partial^2 u}{\partial x_{i_0}^2}(x)\int_{B(z,\delta)}\frac{z_{i_0}^4}{|z|^4}\rho_n(z)dz\right|\leq\frac{2d}{d+1}\epsilon. \quad (5.23)$$

which is equivalent to

$$\left|\frac{d(d+2)}{2}\int_{B(0,\delta)}\frac{d^2 u(x,x+z)}{|z|^2}\frac{z_{i_0}^2-\frac{1}{d+2}|z|^2}{|z|^2}\rho_n(z)dz\right.$$

$$\left. -\frac{\partial^2 u}{\partial x_{i_0}^2}(x)(A(d))_{i_0,j_0}\int_{B(0,\delta)}\frac{z_{i_0}^4}{|z|^4}\rho_n(z)dz\right| \le d^2\epsilon. \tag{5.24}$$

Notice that the last inequality (5.24) corresponds to inequality (5.18) for the $i_0 \ne j_0$ case. Using (5.24) we can now finish the proof in a similar fashion. For large enough n we have

$$\begin{aligned}
\left|(\mathcal{H}_n u - c_n \times \nabla^2 u)_{(i_0,i_0)}\right| &= \left|\frac{d(d+2)}{2}\int_{\mathbb{R}^d}\frac{d^2 u(x,y)}{|y-x|^2}\frac{(y_{i_0}-x_{i_0})^2 - \frac{1}{d+2}|y-x|^2}{|y-x|^2}\rho_n(x-y)dy\right.\\
&\qquad \left. -\frac{\partial^2 u}{\partial x_{i_0}^2}(x)(A(d))_{i_0,j_0}\int_{\mathbb{R}^d}\frac{z_{i_0}^4}{|z|^4}\rho_n(z)dz\right|\\
&\le d^2\epsilon + \frac{d(d+2)}{2}\int_{|z|\ge\delta}\frac{|d^2 u(x,x+z)|}{|z|^2}\underbrace{\left|\frac{z_{i_0}^2 - \frac{1}{d+2}|z|^2}{|z|^2}\right|}_{\le 1}\rho_n(z)dz\\
&\qquad + \left|\frac{\partial^2 u}{\partial x_{i_0}^2}(x)\right|(A(d))_{i_0,j_0}\int_{|z|\ge\delta}\underbrace{\frac{z_{i_0}^4}{|z|^4}}_{\le 1}\rho_n(z)dz\\
&\le d^2\epsilon + \frac{d(d+2)}{2}\frac{4\|u\|_\infty}{\delta^2}\int_{|z|\ge\delta}\rho_n(z)dz\\
&\qquad + \|\nabla^2 u\|_\infty(A(d))_{i_0,j_0}\int_{|z|\ge\delta}\rho_n(z)dz\\
&\le 2d^2\epsilon.
\end{aligned}$$

This completes the proof of the Theorem. $\qquad\square$

5.3 Localisation – The $W^{2,p}(\mathbb{R}^d)$ case

The objective of this section is to show that if $u \in W^{2,p}(\mathbb{R}^d)$, $1 \le p < \infty$ then the non-local Hessian $\mathcal{H}_n u$ converges to $\nabla^2 u$ in L^p. The first step is to show that in that case, $\mathcal{H}_n u$ is indeed an L^p function. This follows from the estimate in the following lemma.

Lemma 5.3.1. *Suppose that $u \in W^{2,p}(\mathbb{R}^d)$, where $1 \le p < \infty$. Then $\mathcal{H}_n u \in L^p(\mathbb{R}^d, \mathbb{R}^{d\times d})$ with*

$$\int_{\mathbb{R}^d}|\mathcal{H}_n u(x)|^p dx \le M\|\nabla^2 u\|_{L^p(\mathbb{R}^d,\mathbb{R}^{d\times d})}^p, \tag{5.25}$$

where the constant M depends only on d and p.

Proof. Recall, see [Bre83], that there exists a constant $M = M(p,d)$ such that for every $v \in W^{1,p}(\mathbb{R}^d)$

$$\int_{\mathbb{R}^d}|v(x-y) - v(x)|^p dx \le M|y|^p\|\nabla v\|_{L^p(\mathbb{R}^d,\mathbb{R}^d)}^p. \tag{5.26}$$

We first prove the result for functions $v \in \mathcal{C}^\infty(\mathbb{R}^d) \cap W^{2,p}(\mathbb{R}^d)$ and then we complete the

proof using a density argument. Using (5.26), Hölder's and Jensen's inequality, equation (5.11) as well as Fubini's theorem we have the following successive estimates (the constant is always denoted with M and everytime it depends only on d and p):

$$\int_{\mathbb{R}^d} |\mathcal{H}_n v(x)|_{\mathbb{R}^{d\times d}}^p dx$$

$$= \left(\frac{d(d+2)}{2}\right)^p \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} \frac{d^2 v(x,y)}{|x-y|^2} \frac{\left((x-y)\otimes(x-y) - \frac{|x-y|^2}{d+2}I_d\right)}{|x-y|^2} \rho_n(x-y)dy \right|^p dx$$

$$\leq M \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \frac{|d^2 v(x,y)|}{|x-y|^2}\rho_n(x-y)dy \right)^p dx$$

$$\leq M \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \frac{|d^2 v(x,y)|^p}{|x-y|^{2p}}\rho_n(x-y)dy \right) \left(\underbrace{\int_{\mathbb{R}^d} \rho(x-y)dy}_{=1} \right)^{p/p'} dx \qquad (5.27)$$

$$\leq M \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \frac{|\int_0^1 \nabla v(x+t(y-x)) - \nabla v(x+(t-1)(y-x))dt|^p |x-y|^p}{|x-y|^{2p}}\rho_n(x-y)dy \right) dx$$

$$\leq M \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \frac{\int_0^1 |\nabla v(x+t(y-x)) - \nabla v(x+(t-1)(y-x))|^p dt}{|x-y|^p}\rho_n(x-y)dy \right) dx$$

$$\leq M \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \left(\int_0^1 \int_0^1 |\nabla^2 v(x+(t+s-1)(y-x))|^p dsdt \right) \rho_n(x-y)dy \right) dx$$

$$= M \int_0^1 \int_0^1 \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \left(|\nabla^2 v(x+(t+s-1)(y-x))|^p \right) \rho_n(x-y)dy \right) dxdsdt$$

$$= M \int_0^1 \int_0^1 \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \left(|\nabla^2 v(x+(t+s-1)\xi)|^p \right) \rho_n(\xi)d\xi \right) dxdsdt,$$

$$= M \int_0^1 \int_0^1 \int_{\mathbb{R}^d} \left(\rho_n(\xi) \int_{\mathbb{R}^d} \left(|\nabla^2 v(x+(t+s-1)\xi)|^p \right) dx \right) d\xi dsdt,$$

$$= M \int_0^1 \int_0^1 \int_{\mathbb{R}^d} \rho_n(\xi) \|\nabla^2 v\|_{L^p(\mathbb{R}^d, \mathbb{R}^{d\times d})}^p d\xi dsdt$$

$$= M \|\nabla^2 v\|_{L^p(\mathbb{R}^d, \mathbb{R}^{d\times d})}^p.$$

Consider now a sequence $(v_k)_{k\in\mathbb{N}}$ in $\mathcal{C}^\infty(\mathbb{R}^d) \cap W^{2,p}(\mathbb{R}^d)$ approximating u in $W^{2,p}(\mathbb{R}^d)$. We already have from above that

$$\int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \frac{|d^2 v_k(x,y)|^p}{|x-y|^{2p}}\rho_n(x-y)dy \right) dx \leq M\|\nabla^2 v_k\|_{L^p(\mathbb{R}^d, \mathbb{R}^{d\times d})}, \quad \forall k\in\mathbb{N}, \qquad (5.28)$$

or

$$\int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \frac{|d^2 v_k(x,x+z)|^p}{|z|^{2p}}\rho_n(z)dz \right) dx \leq M\|\nabla^2 v_k\|_{L^p(\mathbb{R}^d, \mathbb{R}^{d\times d})}, \quad \forall k\in\mathbb{N}. \qquad (5.29)$$

Since v_k converges to u in $L^p(\mathbb{R}^d)$ we have that there exists a subsequence $(v_{k_\ell})_{\ell\in\mathbb{N}}$ con-

verging to u almost everywhere. From an application of Fatou's lemma we get that for every $x \in \mathbb{R}^d$

$$\underbrace{\int_{\mathbb{R}^d} \frac{|d^2u(x, x+z))|^p}{|z|^{2p}} \rho_n(z)dz}_{F(x)} \leq M \liminf_{\ell \to \infty} \underbrace{\int_{\mathbb{R}^d} \frac{|d^2 v_{k_\ell}(x, x+z)|^p}{|z|^{2p}} \rho_n(z)dz}_{F_{k_\ell}(x)}.$$

Applying one more time Fatou's Lemma to $(F_{k_\ell})_{\ell \in \mathbb{N}}$ and F we get that

$$\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{|d^2u(x, x+z))|^p}{|z|^{2p}} \rho_n(z)dzdx \leq M \liminf_{\ell \to \infty} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{|d^2 v_{k_\ell}(x, x+z))|^p}{|z|^{2p}} \rho_n(z)dzdx$$

$$\leq M \liminf_{\ell \to \infty} \|\nabla^2 v_{k_\ell}\|_{L^p(\mathbb{R}^d, \mathbb{R}^{d \times d})}$$

$$= M \|\nabla^2 u\|_{L^p(\mathbb{R}^d, \mathbb{R}^{d \times d})}.$$

From the last inequality and the fact that (5.27) holds for $W^{2,p}$ functions as well, the proof is complete.

$\square$

We now have the necessary tools to prove the localisation for $W^{2,p}$ functions.

Theorem 5.3.2. *Let $1 \leq p < \infty$. Then for every $u \in W^{2,p}(\mathbb{R}^d)$ we have that*

$$\mathcal{H}_n u \to \nabla^2 u \quad in \ L^p(\mathbb{R}^d, \mathbb{R}^{d \times d}) \quad as \ n \to \infty.$$

Proof. We notice first that the result holds for functions $v \in \mathcal{C}_c^2(\mathbb{R}^d)$. In order to see that observe that it suffices to show that for every $\epsilon > 0$, there exists a constant $L > 1$ such that

$$\sup_{n \in \mathbb{N}} \int_{B(0, Lr)^c} |\mathcal{H}_n v(x)|^p dx \leq \epsilon, \tag{5.30}$$

where $B(0, r)$ is a ball containing the support of v. In that case we can estimate as follows

$$\int_{\mathbb{R}^d} |\mathcal{H}_n v(x) - \nabla^2 v(x)|^p dx \leq \int_{B(0, Lr)} |\mathcal{H}_n v(x) - \nabla^2 v(x)|^p dx + \epsilon, \tag{5.31}$$

and using Theorem 5.2.2 we derive

$$\limsup_{n \to \infty} \int_{\mathbb{R}^d} |\mathcal{H}_n v(x) - \nabla^2 v(x)|^p dx \leq \epsilon,$$

from where the result follows since ϵ is arbitrary. In order to get (5.30) we apply Jensen's inequality with respect to the measure $\rho_n \mathcal{L}^d$

$$\int_{B(0, Lr)} |\mathcal{H}_n v(x)|^p dx \leq \int_{B(0, Lr)^c} \int_{\mathbb{R}^d} \frac{|v(y) - 2v(x) + v(x + (x-y))|^p}{|x - y|^{2p}} \rho_n(x - y)dydx$$

$$= \int_{B(0,Lr)^c} \int_{B(0,r)} \frac{|v(y)|^p}{|x-y|^{2p}} \rho_n(x-y)dydx$$

$$+ \int_{B(0,Lr)^c} \int_{\{y:\, x+(x+y)\in B(0,r)\}} \frac{|v(x+(x-y))|^p}{|x-y|^{2p}} \rho_n(x-y)dydx.$$

Letting $z = x + (x-y)$, we obtain

$$\int_{B(0,Lr)^c} \int_{\{y:\, x+(x+y)\in B(0,r)\}} \frac{|v(x+(x-y))|^p}{|x-y|^p} \rho_n(x-y)dydx =$$

$$\int_{B(0,Lr)^c} \int_{B(0,r)} \frac{|v(z)|^p}{|z-x|^{2p}} \rho_n(z-x)dzdx,$$

and therefore by the symmetry of ρ_n we have

$$\int_{B(0,Lr)^c} |\mathcal{H}_n v(x)|^p dx \leq 2 \int_{B(0,Lr)^c} \int_{B(0,r)} \frac{|v(y)|^p}{|x-y|^{2p}} \rho_n(x-y)dydx$$

$$\leq \frac{2}{(L-1)^{2p}} \int_{B(0,Lr)^c} \int_{B(0,r)} |v(y)|^p \rho_n(x-y)dydx$$

$$\leq \frac{2}{(L-1)^{2p}} \|\rho_n\|_{L^1(\mathbb{R}^d)} \|v\|_{L^p(\mathbb{R}^d)}^p.$$

Thus (5.30) follows by choosing L to be large enough.

We use now the fact that that $\mathcal{C}_c^\infty(\mathbb{R}^d)$ and hence $\mathcal{C}_c^2(\mathbb{R}^d)$ is dense in $W^{2,p}(\mathbb{R}^d)$, see for example [Bre83]. Let $\epsilon > 0$, then from density we have that there exists a function $v \in \mathcal{C}_c^2(\mathbb{R}^d)$ such that

$$\|\nabla^2 u - \nabla^2 v\|_{L^p(\mathbb{R}^d, \mathbb{R}^{d\times d})} \leq \epsilon.$$

Thus using also Lemma 5.3.1 we have

$$\|\mathcal{H}_n u - \nabla^2 u\|_{L^p(\mathbb{R}^d, \mathbb{R}^{d\times d})} \leq \|\mathcal{H}_n u - \mathcal{H}_n v\|_{L^p(\mathbb{R}^d, \mathbb{R}^{d\times d})} + \|\mathcal{H}_n v - \nabla^2 v\|_{L^p(\mathbb{R}^d, \mathbb{R}^{d\times d})}$$

$$+ \|\nabla^2 v - \nabla^2 u\|_{L^p(\mathbb{R}^d, \mathbb{R}^{d\times d})}$$

$$\leq C\epsilon + \|\mathcal{H}_n v - \nabla^2 v\|_{L^p(\mathbb{R}^d, \mathbb{R}^{d\times d})} + \epsilon,$$

Taking limits as $n \to \infty$ we get

$$\limsup_{n\to\infty} \|\mathcal{H}_n u - \nabla^2 u\|_{L^p(\mathbb{R}^d, \mathbb{R}^{d\times d})} \leq (C+1)\epsilon,$$

and thus we conclude that

$$\lim_{n\to\infty} \|\mathcal{H}_n u - \nabla^2 u\|_{L^p(\mathbb{R}^d, \mathbb{R}^{d\times d})} = 0.$$

$\square$

5.4   Second order non-local integration by parts

Before we proceed to the localisation of $\mathcal{H}_n$ for the BV^2 case we first need to introduce a second order non-local integration by parts formula which is an essential tool for the proofs of the next section.

We define the second order non-local divergence of a function $\phi = (\phi_{ij})_{i,j=1}^d$ as

$$\mathcal{D}_n^2 \phi(x) = \frac{d(d+2)}{2} \int_{\mathbb{R}^d} \frac{d^2\phi(x,y)}{|x-y|^2} \cdot \frac{\left((x-y) \otimes (x-y) - \frac{|x-y|^2}{d+2} I_d\right)}{|x-y|^2} \rho_n(x-y)dy. \quad (5.32)$$

where here $A \cdot B = \sum_{i,j=1}^d A_{ij}B_{ij}$ for two $d \times d$ matrices A and B. Notice that (5.32) is well defined for $\phi \in \mathcal{C}_c^2(\mathbb{R}^d, \mathbb{R}^{d \times d})$.

Theorem 5.4.1 (Second order non-local integration by parts formula). *Suppose that* $u \in L^p(\mathbb{R}^d)$, $1 \leq p < \infty$, $\frac{|d^2 u(x,y)|}{|x-y|^2}\rho_n(x-y) \in L^1(\mathbb{R}^d \times \mathbb{R}^d)$ *and let* $\phi \in \mathcal{C}_c^2(\mathbb{R}^d, \mathbb{R}^{d \times d})$. *Then*

$$\int_{\mathbb{R}^d} \mathcal{H}_\rho u(x) \cdot \phi(x)dx = \int_{\mathbb{R}^d} u(x)\mathcal{D}_\rho^2 \phi(x)dx. \quad (5.33)$$

Proof. The proof is very similar with the corresponding proof in [MS13]. Using Fubini's and dominated convergence theorem we have

$$\int_{\mathbb{R}^N} \mathcal{H}_\rho u(x)\phi(x)dx$$

$$= \frac{d(d+2)}{2} \lim_{\epsilon \to 0} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d \backslash B(x,\epsilon)} \frac{d^2 u(x,y)}{|x-y|^2} \frac{\left((x-y) \otimes (x-y) - \frac{|x-y|^2}{d+2}I_d\right)}{|x-y|^2} \rho(x-y) \cdot \phi(x)dydx$$

$$= \frac{d(d+2)}{2} \lim_{\epsilon \to 0} \int_{D_\epsilon^d} \frac{d^2 u(x,y)}{|x-y|^2} \frac{\left((x-y) \otimes (x-y) - \frac{|x-y|^2}{d+2}I_d\right)}{|x-y|^2} \rho(x-y) \cdot \phi(x)d(\mathcal{L}^d)^2(x,y),$$

where $D_\epsilon^d := \mathbb{R}^d \times \mathbb{R}^d \setminus \{|x-y| < \epsilon\}$. Similarly we have

$$\int_{\mathbb{R}^d} u(x)\mathcal{D}_\rho^2 \phi(x)dx$$

$$= \frac{d(d+2)}{2} \lim_{\epsilon \to 0} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d \backslash B(x,\epsilon)} u(x) \sum_{i,j=1}^d \frac{d^2 \phi_{ij}(x,y)}{|x-y|^2} \frac{(x_i - y_i)(x_j - y_j) - \delta_{ij}\frac{|x-y|^2}{d+2}}{|x-y|^2} \rho(x-y)dydx$$

$$= \frac{d(d+2)}{2} \lim_{\epsilon \to 0} \int_{D_\epsilon^d} u(x) \sum_{i,j=1}^d \frac{d^2 \phi_{ij}(x,y)}{|x-y|^2} \frac{(x_i - y_i)(x_j - y_j) - \delta_{ij}\frac{|x-y|^2}{d+2}}{|x-y|^2} \rho(x-y)d(\mathcal{L}^d)^2(x,y),$$

where, for notational convenience, we used the standard convention

$$\delta_{ij} = \begin{cases} 1 & \text{if} \quad i = j, \\ 0 & \text{if} \quad i \neq j. \end{cases}$$

169

Thus, it suffices to show that for every i, j and $\epsilon > 0$ we have

$$\int_{D_\epsilon^d} \frac{d^2 u(x,y)}{|x-y|^2} \frac{(x_i - y_i)(x_j - y_j) - \delta_{ij} \frac{|x-y|^2}{d+2}}{|x-y|^2} \rho(x-y) \phi_{ij}(x) d(\mathcal{L}^d)^2(x,y) =$$
$$\int_{D_\epsilon^d} u(x) \frac{d^2 \phi_{ij}(x,y)}{|x-y|^2} \frac{(x_i - y_i)(x_j - y_j) - \delta_{ij} \frac{|x-y|^2}{d+2}}{|x-y|^2} \rho(x-y) d(\mathcal{L}^d)^2(x,y). \qquad (5.34)$$

In order to show (5.34), it suffices to prove

$$\int_{D_\epsilon^d} \frac{u(y) \phi_{ij}(x)}{|x-y|^2} \frac{(x_i - y_i)(x_j - y_j) - \delta_{ij} \frac{|x-y|^2}{d+2}}{|x-y|^2} \rho(x-y) d(\mathcal{L}^d)^2(x,y) =$$
$$\int_{D_\epsilon^d} \frac{u(x) \phi_{ij}(y)}{|x-y|^2} \frac{(x_i - y_i)(x_j - y_j) - \delta_{ij} \frac{|x-y|^2}{d+2}}{|x-y|^2} \rho(x-y) d(\mathcal{L}^d)^2(x,y), \qquad (5.35)$$

and

$$\int_{D_\epsilon^d} \frac{u(x + (x-y)) \phi_{ij}(x)}{|x-y|^2} \frac{(x_i - y_i)(x_j - y_j) - \delta_{ij} \frac{|x-y|^2}{d+2}}{|x-y|^2} \rho(x-y) d(\mathcal{L}^d)^2(x,y) =$$
$$\int_{D_\epsilon^d} \frac{u(x) \phi_{ij}(x + (x-y))}{|x-y|^2} \frac{(x_i - y_i)(x_j - y_j) - \delta_{ij} \frac{|x-y|^2}{N+2}}{|x-y|^2} \rho(x-y) d(\mathcal{L}^d)^2(x,y). \qquad (5.36)$$

Equation (5.35) can be easily showed by alternating x and y and using the symmetry of the domain. Finally equation (5.36) can be proved by employing the substitution $u = 2x - y$, $v = 3x - 2y$, noting that $x - y = v - u$ and that the determinant of the Jacobian of this substitution is -1. $\qquad \square$

The following lemma shows the convergence of the second order non-local divergence to the continuous analogue $\mathrm{div}^2 \phi$, where $\phi \in \mathcal{C}_c^2(\mathbb{R}^d, \mathbb{R}^{d \times d})$ and

$$\mathrm{div}^2 \phi := \sum_{i,j=1}^d \frac{\partial \phi_{ij}}{\partial x_i \partial x_j}.$$

Lemma 5.4.2. *Let $\phi \in \mathcal{C}_c^2(\mathbb{R}^d, \mathbb{R}^{d \times d})$. Then for every $1 \le p \le \infty$ we have*

$$\lim_{n \to \infty} \|\mathcal{D}_n^2 \phi - \mathrm{div}^2 \phi\|_{L^p(\mathbb{R}^d)} = 0. \qquad (5.37)$$

Proof. The convergence for $p = \infty$, follows immediately from Theorem 5.2.2 and it also holds for $1 \le p < \infty$ from Theorem 5.3.2. $\qquad \square$

5.5 Localisation – The $\mathrm{BV}^2(\mathbb{R}^d)$ case

In this section we prove that for functions $u \in \mathrm{BV}^2(\mathbb{R}^d)$ the sequence $\mathcal{H}_n$ localises to D^2u weakly* in the sense of measures as n tends to infinity. In order to show that, we first need to prove that the non-local Hessian is well defined for BV^2 functions. We do that in the following lemma.

Lemma 5.5.1. *Suppose that $u \in \mathrm{BV}^2(\mathbb{R}^d)$. Then $\mathcal{H}_n u \in L^1(\mathbb{R}^d, \mathbb{R}^{d \times d})$ with*

$$\int_{\mathbb{R}^N} |\mathcal{H}_n u(x)| dx \leq M |D^2 u|(\mathbb{R}^d), \tag{5.38}$$

where the constant M depends only on d.

Proof. Let $(u_k)_{k \in \mathbb{N}}$ be a sequence of functions in $C^\infty(\mathbb{R}^d)$ that converges strictly to u in $\mathrm{BV}^2(\mathbb{R}^d)$. From the same calculations as in the proof of Lemma 5.3.1 we have for every $k \in \mathbb{N}$

$$\int_{\mathbb{R}^d} |\mathcal{H}_n u_k(x)| dx \leq M \|\nabla^2 u_k\|_{L^1(\mathbb{R}^d, \mathbb{R}^{d \times d})}.$$

Using Fatou's Lemma in a similar way as in Lemma 5.3.1 we get

$$\int_{\mathbb{R}^d} |\mathcal{H}_n u(x)| dx \leq M \liminf_{k \to \infty} |D^2 u_k|(\mathbb{R}^d)$$
$$= |D^2 u|(\mathbb{R}^d).$$

$\square$

We can now proceed to the proof of the localisation result for BV^2 functions. We define $(\mu_n)_{n \in \mathbb{N}}$ to be the sequence of $\mathbb{R}^{d \times d}$–valued finite Radon measures with $\mu_n := \mathcal{H}_n u \, \mathcal{L}^d$.

Theorem 5.5.2. *Let $u \in \mathrm{BV}^2(\mathbb{R}^d)$. Then*

$$\mu_n \to D^2 u, \quad \text{weakly* in measures,}$$

i.e., for every $\phi \in \mathcal{C}_0(\mathbb{R}^d, \mathbb{R}^{d \times d})$

$$\lim_{n \to \infty} \int_{\mathbb{R}^d} \mathcal{H}_n u(x) \cdot \phi(x) dx = \int_{\mathbb{R}^d} \phi(x) \, dD^2 u. \tag{5.39}$$

Proof. Since $\mathcal{C}_c^\infty(\mathbb{R}^d, \mathbb{R}^{d \times d})$ is dense in $\mathcal{C}_0(\mathbb{R}^d, \mathbb{R}^{d \times d})$ it suffices to prove (5.39) for every $\psi \in \mathcal{C}_c^\infty(\mathbb{R}^d, \mathbb{R}^{d \times d})$. Indeed, suppose we have done this, let $\epsilon > 0$ and let $\phi \in \mathcal{C}_0(\mathbb{R}^d, \mathbb{R}^{d \times d})$ and $\psi \in \mathcal{C}_c^\infty(\mathbb{R}^d, \mathbb{R}^{d \times d})$ such that $\|\phi - \psi\|_\infty < \epsilon$. Then, using also the estimate (5.38), we

have

$$
\begin{aligned}
\left| \int_{\mathbb{R}^d} \mathcal{H}_n(x) \cdot \phi(x) dx - \int_{\mathbb{R}^d} \phi(x) \, dD^2 u \right| &\leq \left| \int_{\mathbb{R}^d} \mathcal{H}_n(x) \cdot (\phi(x) - \psi(x)) dx \right| \\
&\quad + \left| \int_{\mathbb{R}^d} \mathcal{H}_n(x) \cdot \psi(x) dx - \int_{\mathbb{R}^d} \psi(x) \, dD^2 u \right| \\
&\quad + \left| \int_{\mathbb{R}^d} (\phi(x) - \psi(x)) \, dD^2 u \right| \\
&\leq \epsilon \int_{\mathbb{R}} |\mathcal{H}_n u(x)| dx \\
&\quad + \left| \int_{\mathbb{R}^d} \mathcal{H}_n(x) \cdot \psi(x) dx - \int_{\mathbb{R}^d} \psi(x) \, dD^2 u \right| \\
&\quad + \epsilon |D^2 u|(\mathbb{R}^d) \\
&\leq M\epsilon |D^2 u|(\mathbb{R}^d) \\
&\quad + \left| \int_{\mathbb{R}^d} \mathcal{H}_n(x) \cdot \psi(x) dx - \int_{\mathbb{R}^d} \psi(x) \, dD^2 u \right| \\
&\quad + \epsilon |D^2 u|(\mathbb{R}^d).
\end{aligned}
$$

Taking the limit $n \to \infty$ from both sides of the above inequality we get

$$
\limsup_{n \to \infty} \left| \int_{\mathbb{R}^d} \mathcal{H}_n(x) \cdot \phi(x) dx - \int_{\mathbb{R}^d} \phi(x) \, dD^2 u \right| \leq \tilde{M}\epsilon,
$$

for a constant $\tilde{M}$ and since ϵ is arbitrary we have (5.39). We thus proceed to prove (5.39) for compactly supported smooth functions. From the estimate (5.38) we have that $(|\mu_n|)_{n \in \mathbb{N}}$ is bounded, thus there exists a subsequence $(\mu_{n_k})_{k \in \mathbb{N}}$ and a $\mathbb{R}^{d \times d}$–valued Radon measure μ such that μ_{n_k} converges to μ weakly*. This means that for every $\psi \in \mathcal{C}_c^\infty(\mathbb{R}^d, \mathbb{R}^{d \times d})$ we have

$$
\lim_{k \to \infty} \int_{\mathbb{R}^d} \mathcal{H}_{n_k}(x) \cdot \psi(x) dx = \int_{\mathbb{R}^d} \psi(x) \cdot d\mu.
$$

On the other hand from the integration by parts formula (5.33) and Lemma 5.4.2 we get

$$
\begin{aligned}
\lim_{k \to \infty} \int_{\mathbb{R}^d} \mathcal{H}_{n_k}(x) \cdot \psi(x) dx &= \lim_{k \to \infty} \int_{\mathbb{R}^d} u(x) \mathcal{D}_{n_k}^2 \psi(x) dx \\
&= \int_{\mathbb{R}^d} u(x) \mathrm{div}^2 \psi(x) dx \\
&= \int_{\mathbb{R}^d} \psi(x) \cdot dD^2 u.
\end{aligned}
$$

This means that $\mu = D^2 u$. Observe now that since we actually deduce that every subsequence of $(\mu_n)_{n \in \mathbb{N}}$ has a further subsequence that converges to $D^2 u$ weakly*, then the initial sequence $(\mu_n)_{n \in \mathbb{N}}$ converges to $D^2 u$ weakly*. $\qquad \square$

Let us note here that in the case $d = 1$, we can also prove strict convergence of the

measures μ_n to $D^2 u$, that is, in addition to (5.39) we also have

$$|\mu_n|(\mathbb{R}) \to |D^2 u|(\mathbb{R}).$$

Theorem 5.5.3. *Let $d = 1$. Then the sequence $(\mu_n)_{n \in \mathbb{N}}$ converges to $D^2 u$ strictly as measures, i.e,*

$$\mu_n \to D^2 u, \quad weakly^* \ in \ measures, \tag{5.40}$$

and

$$|\mu_n|(\mathbb{R}) \to |D^2 u|(\mathbb{R}). \tag{5.41}$$

Proof. The weak* convergence was proved in Theorem 5.5.2. Since the total variation norm is lower semicontinuous with respect to the weak* convergence, we also have

$$|D^2 u|(\mathbb{R}) \le \liminf_{n \to \infty} |\mu_n|(\mathbb{R}). \tag{5.42}$$

Thus it suffices to show that

$$\limsup_{n \to \infty} |\mu_n|(\mathbb{R}) \le |D^2 u|(\mathbb{R}). \tag{5.43}$$

Note that in dimension one, the non-local Hessian formula is written as

$$\mathcal{H}_n u(x) = \int_{\mathbb{R}} \frac{u(y) - 2u(x) + u(x + (x - y))}{|x - y|^2} \rho_n(x - y) dy. \tag{5.44}$$

Following the proof of Lemma 5.3.1, we can easily verify that for $v \in \mathcal{C}^\infty(\mathbb{R}) \cap \mathrm{BV}^2(\mathbb{R})$ we have

$$\int_{\mathbb{R}} |\mathcal{H}_n v(x)| dx \le \|\nabla^2 v\|_{L^1(\mathbb{R})},$$

i.e., the constant M that appears in the estimate (5.25) is equal to one. Using Fatou's Lemma and the BV^2 strict approximation of u by smooth functions we get that

$$|\mu_n|(\mathbb{R}) = \int_{\mathbb{R}} |\mathcal{H}_n u(x)| dx \le |D^2 u|(\mathbb{R}),$$

from where (5.43) follows straightforwardly. $\qquad \square$

5.6 Non-local characterisation of $W^{2,p}(\mathbb{R}^d)$ and $\mathrm{BV}^2(\mathbb{R}^d)$

Characterisation of Sobolev and BV spaces in terms of non-local, derivative-free energies has been done so far only in the first order case, see [BBM01, Pon04, Men12, MS13]. Here we characterise the spaces $W^{2,p}(\mathbb{R}^N)$, $1 < p < \infty$ and $\mathrm{BV}^2(\mathbb{R}^N)$ using our definition of non-local Hessian. For that we need to define the non-local Hessian as a distribution. Note that so far we have used the non-local Hessian operator $\mathcal{H}_n$ for regular enough functions so

that formula (5.1) is well-defined. Indeed, the estimates (5.25) and (5.38) show that (5.1) is well defined for $W^{2,p}(\mathbb{R}^d)$, $1 \leq p < \infty$, and $\mathrm{BV}^2(\mathbb{R}^d)$ functions. One can easily check that it is also well defined for $W^{2,\infty}(\mathbb{R}^d)$ functions. Moreover, it is very easy to check that if $u \in \mathcal{C}_c^2(\mathbb{R}^d)$ then the following estimate hold

$$|\mathcal{H}_n u(x)| \leq \frac{d^2(d+2)}{2} \|\nabla^2 u\|_{L^\infty(\mathbb{R}^d)}, \quad \forall x \in \mathbb{R}^d. \tag{5.45}$$

In fact as the proof of Theorem 5.4.1 shows, a sufficient condition for the well-posedness of (5.1) is

$$\frac{|d^2 u(x,y)|}{|x-y|^2} \rho_n(x-y) \in L^1(\mathbb{R}^d \times \mathbb{R}^d), \tag{5.46}$$

and in that case we have that $\mathcal{H}_n u \in L^1(\mathbb{R}^d)$. However, we define $\mathcal{H}_n$ as a distribution, extending thus the definition for arbitrary L^p functions $1 \leq p < \infty$.

Definition 5.6.1 (Distributional definition of non-local Hessian). *Let $u \in L^p(\mathbb{R}^d)$ for some $p \in [1, \infty)$. We define the distributional non-local Hessian $\mathfrak{H}_n u$ as*

$$\langle \mathfrak{H}_n u, \phi \rangle = \int_{\mathbb{R}^d} u(x) \mathcal{D}_n^2 \phi(x) dx, \quad \forall \phi \in \mathcal{C}_c^2(\Omega, \mathbb{R}^{d \times d}). \tag{5.47}$$

Observe that since $\phi \in W^{2,p}(\mathbb{R}^d)$ for every $p \in [1, \infty]$, estimates (5.25) and (5.45) imply that $\mathcal{D}_n^2 \phi \in L^p(\mathbb{R}^d)$ for every $p \in [1, \infty]$ and that (5.47) is well defined from Hölder's inequality. Note also that $\mathfrak{H}_n u$ is indeed a distribution. In order to see that observe that if $u \in L^1(\mathbb{R}^d)$ then estimate (5.45) gives

$$\langle \mathfrak{H}_n u, \phi \rangle \leq M \|u\|_{L^1(\mathbb{R}^d)} \|\nabla^2 \phi\|_\infty, \tag{5.48}$$

where M depends only on d. If $u \in L^p(\mathbb{R}^d)$ with $1 < p < \infty$ then from the estimate 5.3.1 and the fact that $\nabla^2 \phi$ is of compact support we have

$$\langle \mathfrak{H}_n u, \phi \rangle \leq M \|u\|_{L^p(\mathbb{R}^d)} \|\nabla^2 \phi\|_{L^q(\mathbb{R}^d, \mathbb{R}^{d \times d})} \leq M \|u\|_{L^p(\mathbb{R}^d)} \|\nabla^2 \phi\|_\infty, \tag{5.49}$$

where $1/p + 1/q = 1$ and the constant M depends only on p and d. Finally, the integration by parts formula (5.33) shows that if (5.46) holds (for example in the case of $W^{2,p}(\mathbb{R}^d)$, $\mathrm{BV}^2(\mathbb{R}^d)$ and $\mathcal{C}_c^2(\Omega, \mathbb{R}^{d \times d})$ functions) then the distribution $\mathfrak{H}_n u$ can be represented by the function $\mathcal{H}_n u$. We are now ready to prove our characterisations.

Theorem 5.6.2. *Let $u \in L^p(\mathbb{R}^d)$ for some $1 < p < \infty$. Then*

$$u \in W^{2,p}(\mathbb{R}^d) \quad \Longleftrightarrow \quad \mathfrak{H}_n u \in L^p(\mathbb{R}^d, \mathbb{R}^{d \times d}) \, \forall n \in \mathbb{N} \text{ and } \liminf_{n \to \infty} \int_{\mathbb{R}^d} |\mathcal{H}_n u(x)|^p dx < \infty. \tag{5.50}$$

Let $u \in L^1(\mathbb{R}^d)$. Then

$$u \in \mathrm{BV}^2(\mathbb{R}^d) \quad \Longleftrightarrow \quad \mathfrak{H}_n u \in L^1(\mathbb{R}^d, \mathbb{R}^{d \times d}) \; \forall n \in \mathbb{N} \text{ and } \liminf_{n \to \infty} \int_{\mathbb{R}^d} |\mathcal{H}_n u(x)| dx < \infty.$$
(5.51)

Proof. Firstly, we prove (5.50). Suppose that $u \in W^{2,p}(\mathbb{R}^d)$. Then, Lemma 5.3.1 gives

$$\liminf_{n \to \infty} \int_{\mathbb{R}^d} |\mathcal{H}_n u(x)|^p dx \leq M \|\nabla^2 u\|_{L^p(\mathbb{R}^d)}^p < \infty.$$

Suppose now conversely that

$$\liminf_{n \to \infty} \int_{\mathbb{R}^d} |\mathcal{H}_n u(x)|^p dx < \infty.$$

This means that up to a subsequence, $\mathcal{H}_n u$ is bounded in $L^p(\mathbb{R}^d, \mathbb{R}^{d \times d})$, thus there exists a subsequence $(\mathcal{H}_{n_k} u)_{k \in \mathbb{N}}$ and $v \in L^p(\mathbb{R}^d, \mathbb{R}^{d \times d})$ such that $\mathcal{H}_{n_k} u \to v$ weakly in $L^p(\mathbb{R}^d, \mathbb{R}^{d \times d})$. Thus, using the definition of L^p weak convergence together with the non-local integration by parts formula and Lemma 5.4.2, we have for every $\psi \in \mathcal{C}_c^\infty(\mathbb{R}^d, \mathbb{R}^{d \times d})$,

$$\begin{aligned}
\int_{\mathbb{R}^d} v(x) \cdot \psi(x) dx &= \lim_{k \to \infty} \int_{\mathbb{R}^d} \mathcal{H}_{n_k} u(x) \cdot \psi(x) dx \\
&= \lim_{k \to \infty} \int_{\mathbb{R}^d} u(x) \mathcal{D}_{n_k}^2 \psi(x) dx \\
&= \int_{\mathbb{R}^d} u(x) \mathrm{div}^2 \psi(x) dx,
\end{aligned}$$

something that shows that $v = \nabla^2 u \in L^p(\mathbb{R}^d, \mathbb{R}^{d \times d})$ is the second order weak derivative of u. Now since the second order distributional derivative is a function, the same is true for the first distributional derivatives and they belong to L^p and thus $u \in W^{2,p}(\mathbb{R}^d)$. This is a consequence of the Gagliardo–Nirenberg interpolation inequality [Nir59]

$$\|\nabla u\|_{L^p(\mathbb{R}^d, \mathbb{R}^d)} \leq C \|\nabla^2 u\|_{L^p(\mathbb{R}^d, \mathbb{R}^{d \times d})}^{\frac{1}{2}} \|u\|_{L^p(\mathbb{R}^d)}^{\frac{1}{2}},$$
(5.52)

where the constant C depends only on d.

We now proceed in proving (5.51). Again supposing that $u \in \mathrm{BV}^2(\mathbb{R}^d)$ we have that Lemma 5.5.1 gives us

$$\liminf_{n \to \infty} \int_{\mathbb{R}^d} |\mathcal{H}_n u(x)| dx \leq C |D^2 u|(\mathbb{R}^d).$$

Suppose now that

$$\liminf_{n \to \infty} \int_{\mathbb{R}^N} |\mathcal{H}_n u(x)| dx < \infty.$$

Considering again the measures $\mu_n = \mathcal{H}_n u \, \mathcal{L}^d$ we have that there exists a subsequence

$(\mu_{n_k})_{k\in\mathbb{N}}$ and a finite Radon measure μ such that $\mu_{n_k} \to \mu$ weakly*. Then for every $\psi \in \mathcal{C}_c^\infty(\mathbb{R}^d, \mathbb{R}^{d\times d})$ we have, similarly as before,

$$
\begin{aligned}
\int_{\mathbb{R}^d} \psi \cdot d\mu &= \lim_{k\to\infty} \int_{\mathbb{R}^d} \mathcal{H}_{n_k} u(x) \cdot \psi(x) dx \\
&= \lim_{k\to\infty} \int_{\mathbb{R}^d} u(x) D_{n_k}^2 \psi(x) dx \\
&= \int_{\mathbb{R}^d} u(x) \operatorname{div}^2 \psi(x) dx,
\end{aligned}
$$

something that shows that $\mu = D^2 u$. Thus, $u \in \mathrm{BV}^2(\mathbb{R}^d)$, since if the second order distributional derivative is a finite Radon measure then the first order one is an L^1 function. This again due to a special form of the Gagliardo–Nirenberg interpolation inequality for $W^{2,1}$ functions this time. In order to see that, notice first that from [TS80] we have first that $Du \in L_{loc}^1(\mathbb{R}^d)$. Fix an $\Omega \Subset \mathbb{R}^d$, then we have that $u\lfloor\Omega \in \mathrm{BV}^2(\Omega)$. From Theorem 3.3.7 we can find a sequence $(u_n)_{n\in\mathbb{N}} \in W^{2,1}(\Omega)$ that converges to $u\lfloor\Omega$ strictly in $\mathrm{BV}^2(\Omega)$. From the definition of strict convergence together with the lower semicontinuity of total variation with respect to the strong L^1 convergence we have

$$
\begin{aligned}
\|\nabla(u\lfloor\Omega)\|_{L^1(\Omega)} &\le \liminf_{n\to\infty} \|\nabla u_n\|_{L^1(\Omega)} \\
&\le \liminf_{n\to\infty} C\|\nabla^2 u_n\|_{L^1(\Omega,\mathbb{R}^{d\times d})}^{\frac{1}{2}} \|u_n\|_{L^1(\Omega)}^{\frac{1}{2}} \\
&= C\|\nabla^2(u\lfloor\Omega)\|_{L^1(\Omega,\mathbb{R}^{d\times d})}^{\frac{1}{2}} \|u\lfloor\Omega\|_{L^1(\Omega)}^{\frac{1}{2}} \\
&\le C\|\nabla^2 u\|_{L^1(\mathbb{R}^d,\mathbb{R}^{d\times d})}^{\frac{1}{2}} \|u\|_{L^1(\mathbb{R}^d)}^{\frac{1}{2}},
\end{aligned}
$$

from where we deduce that $\nabla u \in L^1(\mathbb{R}^d)$ with

$$
\|\nabla u\|_{L^1(\mathbb{R}^d)} \le C\|\nabla^2 u\|_{L^1(\mathbb{R}^d,\mathbb{R}^{d\times d})}^{\frac{1}{2}} \|u\|_{L^1(\mathbb{R}^d)}^{\frac{1}{2}},
$$

and hence $u \in \mathrm{BV}^2(\mathbb{R}^d)$. □

5.7 An alternative non-local Hessian approach and an application to image denoising

In the previous sections we saw that the explicit formulation of non-local Hessian defined in (5.1) has some good analytical properties, i.e., the sequence $(\mathcal{H}_n u)_{n\in\mathbb{N}}$ converges to the local analogue $\nabla^2 u$ as the non-locality vanishes. It also led to novel non-local characterisations of the spaces $W^{2,p}(\mathbb{R}^d)$ and $\mathrm{BV}^2(\mathbb{R}^d)$. However the formulation (5.1) is not so convenient for implementation in variational problems in imaging. One reason is that it cannot be defined straightforwardly for functions $u : \Omega \to \mathbb{R}$, where Ω is a bounded do-

main as the point $x + (x - y)$ does not necessarily belong to Ω whenever $x \in \Omega$. Moreover, the radial symmetry of the weighting functions ρ_n is in practice not desirable for imaging purposes. For example, given that a point x of an image u lies next to an edge, we would like to have the freedom to define a form of non-local Hessian of u at x using a weighting function that puts more weight on points that lie on the same side of x, see Figure 5.2. That would eventually lead to an edge preservation regularisation model.

(a) A point near an edge.

(b) Weighting of the neighbouring pixels that correspond to a radially symmetric weighting function.

(c) Weighting of the neighbouring pixels that correspond to weighting function aware of discontinuities.
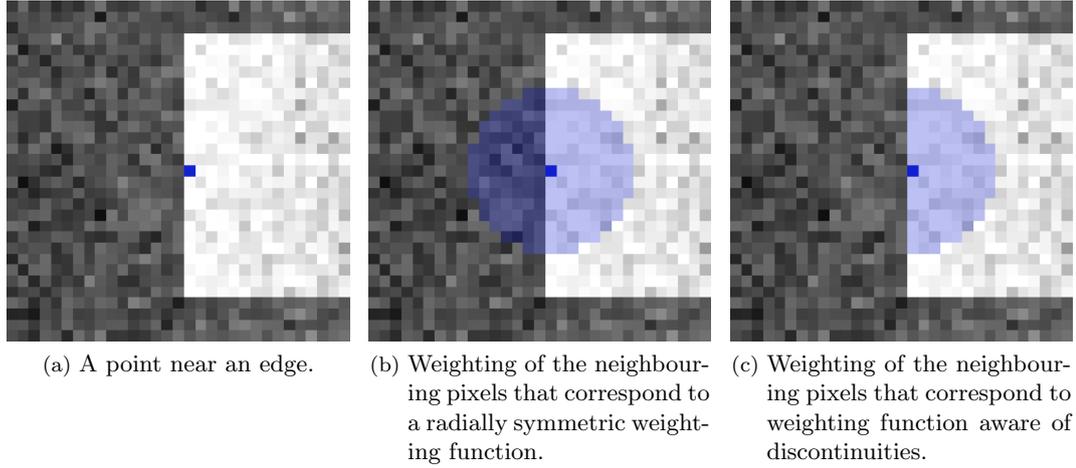
Figure 5.2: Radially symmetric weighting function versus a weighting function aware of discontinuities.

To that scope we define an *implicit* formulation of a non-local Hessian $\mathcal{H}_{\sigma_x} u(x)$ and non-local gradient $\mathcal{G}_{\sigma_x} u(x)$ of a function u at a point x with a weighting function σ_x as follows:

Definition 5.7.1 (Implicit definition of non-local Hessian)**.** *Let* $\Omega \subseteq \mathbb{R}^d$ *be an open bounded domain and* $u \in L^2(\Omega)$*. We define* $\mathcal{G}_{\sigma_x} u(x)$ *and* $\mathcal{H}_{\sigma_x} u(x)$ *to be the non-local gradient and Hessian respectively, of* u *at a point* $x \in \Omega$ *with a positive weight function* $\sigma_x \in L^\infty(\Omega \setminus \{x\})$ *as the minimisers of*

$$(\mathcal{H}_{\sigma_x} u(x), \mathcal{G}_{\sigma_x} u(x)) := \operatorname*{argmin}_{\mathcal{H}' \in \mathbb{R}^{d \times d}, \mathcal{G}' \in \mathbb{R}^d} \frac{1}{2} \int_{\Omega - \{x\}} R_{u, \mathcal{G}', \mathcal{H}'}(x, z)^2 \sigma_x(z) \, dz, \qquad (5.53)$$

where

$$R_{u, \mathcal{G}', \mathcal{H}'}(x, z) = u(x + z) - u(x) - G'^\top z - \frac{1}{2} z^\top H' z. \qquad (5.54)$$

The idea behind Definition 5.7.1 is that at a given point x, we choose $\mathcal{H}_{\sigma_x} u(x)$ and $\mathcal{G}_{\sigma_x} u(x)$ so that the remaining values of u best fit a quadratic model with gradient $\mathcal{G}_{\sigma_x} u(x)$ and Hessian $\mathcal{H}_{\sigma_x} u(x)$ located at x.

In this section we work with the discretised version of (5.53)–(5.54) in dimension 2, i.e., the integral in (5.53) is interpreted as a sum and Ω as an $N \times M$ grid. We can now

define the corresponding non-local Hessian regularised minimisation problems as follows:

L^s–$\mathcal{H}$ regularisation

$$\min_{u \in \mathbb{R}^{N \times M}} \frac{1}{s} \|u - f\|_s^s + \alpha \|\mathcal{H}_\sigma u\|_1, \quad \alpha > 0, \ \ s = 1, 2, \tag{5.55}$$

subject to

$$(\mathcal{H}_{\sigma_x} u(x), \mathcal{G}_{\sigma_x} u(x)) = \underset{\mathcal{H}' \in \mathbb{R}^{2 \times 2}, \mathcal{G}' \in \mathbb{R}^2}{\operatorname{argmin}} \frac{1}{2} \int_{\Omega - \{x\}} R_{u, \mathcal{G}', \mathcal{H}'}(x, z)^2 \sigma_x(z) \, dz, \tag{5.56}$$

$$R_{u, \mathcal{G}', \mathcal{H}'}(x, z) = u(x + z) - u(x) - G'^\top z - \frac{1}{2} z^\top H' z. \tag{5.57}$$

Note that the optimality conditions of the quadratic minimisation problem (5.56)–(5.57) are linear. Thus essentially, problem (5.55)–(5.57) is an optimisation problem with convex objective and linear constraints. We can solve this problem for instance, using CVX in MATLAB, a package for specifying and solving convex programs [GB14, GB08].

Our next concern is how to choose the weighting function σ_x at a point x. As we have discussed previously ideally this function has to be chosen in a way such that points that lie across an edge have low or zero weights. We do that by defining first a positive cost function $c_x : \Omega \to \mathbb{R}$, where $c_x(y)$ must be high when we cannot move from x to y without passing from points with large gradient (edge). A function like that can be defined by solving a weighted Eikonal equation

$$\|\nabla c_x\|_2 = \phi(\nabla f), \quad c_x(x) = 0, \tag{5.58}$$

where $\phi(\nabla f) = \|\nabla f\|_2^2 + \epsilon$, for a small $\epsilon > 0$. Equation (5.58) can be solved efficiently using a fast marching method [Set99, OF03]. Defining c_x as above we have that $c_x(y)$ is high if x and y are separated by a strong edge.

For efficiency in the numerical implementation we do not use all the points $y \in \Omega$ for the formulation of $\mathcal{H}_{\sigma_x} u(x)$ but only those K ones with the shortest distance $c_x(y)$ to x. That is to say, we sort the neighbours of x, $y_1, y_2, \ldots$ such that $c_x(y_1) \le c_x(y_2), \ldots, c_x(y_K) \le \ldots$ and we define the weighting function σ_x we follows:

$$\sigma_x(y_i) = \begin{cases} \frac{1}{(c_x(y_i))^2} & \text{if } i \le K, \\ 0, & \text{if } i > K. \end{cases} \tag{5.59}$$

This number K can be predefined and for our numerical implementations we use $K = 12$, while we also set $\epsilon = 0.01$.

(a) Clean image

(b) Noisy image,
SSIM=0.3261. PSNR=22.82

(c) TV denoised image,
SSIM=0.8979, PSNR=31.82

(d) TGV denoised image,
SSIM=0.9249, PSNR=33.29

(e) L^2–$\mathcal{H}$ denoised image, α=0.2,
SSIM=0.9262, PSNR= 34.39

(f) L^1–$\mathcal{H}$ denoised image, α=1.2,
SSIM=0.9766, PSNR= 35.97
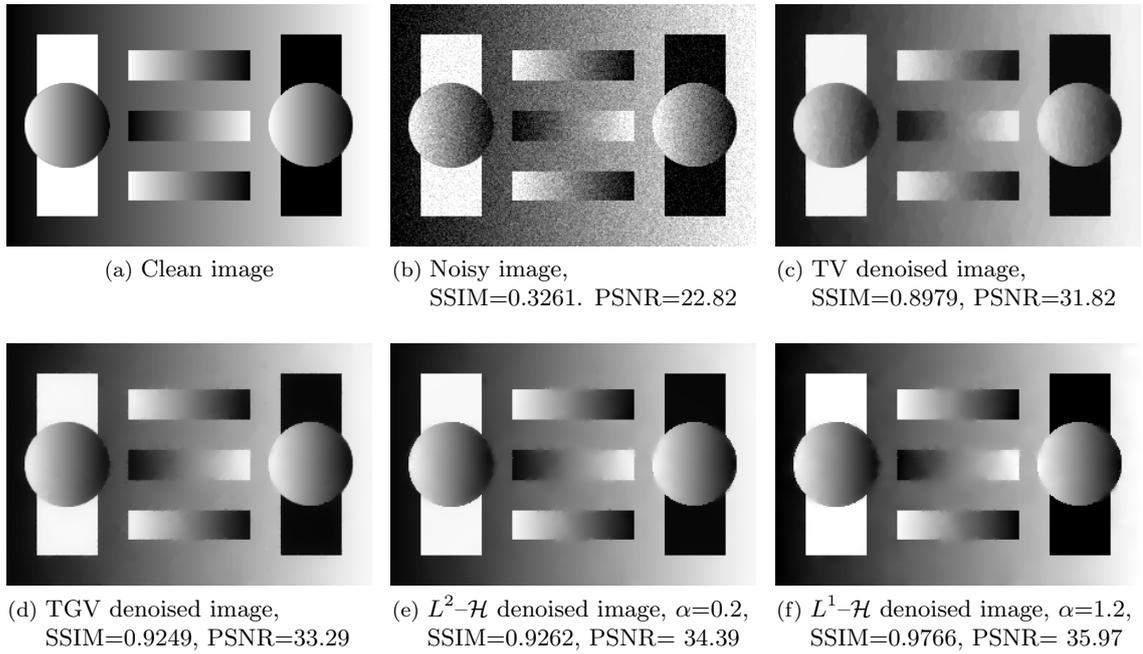
Figure 5.3: Denoising example for Gaussian noise of variance 0.005.

(a) Clean image

(b) Noisy image,
SSIM=0.3261. PSNR=22.82

(c) TV denoised image,
SSIM=0.8979, PSNR=31.82

(d) TGV denoised image,
SSIM=0.9249, PSNR=33.29

(e) L^2–$\mathcal{H}$ denoised image,
SSIM=0.9262, PSNR= 34.39

(f) L^1–$\mathcal{H}$ denoised image,
SSIM=0.9766, PSNR= 35.97
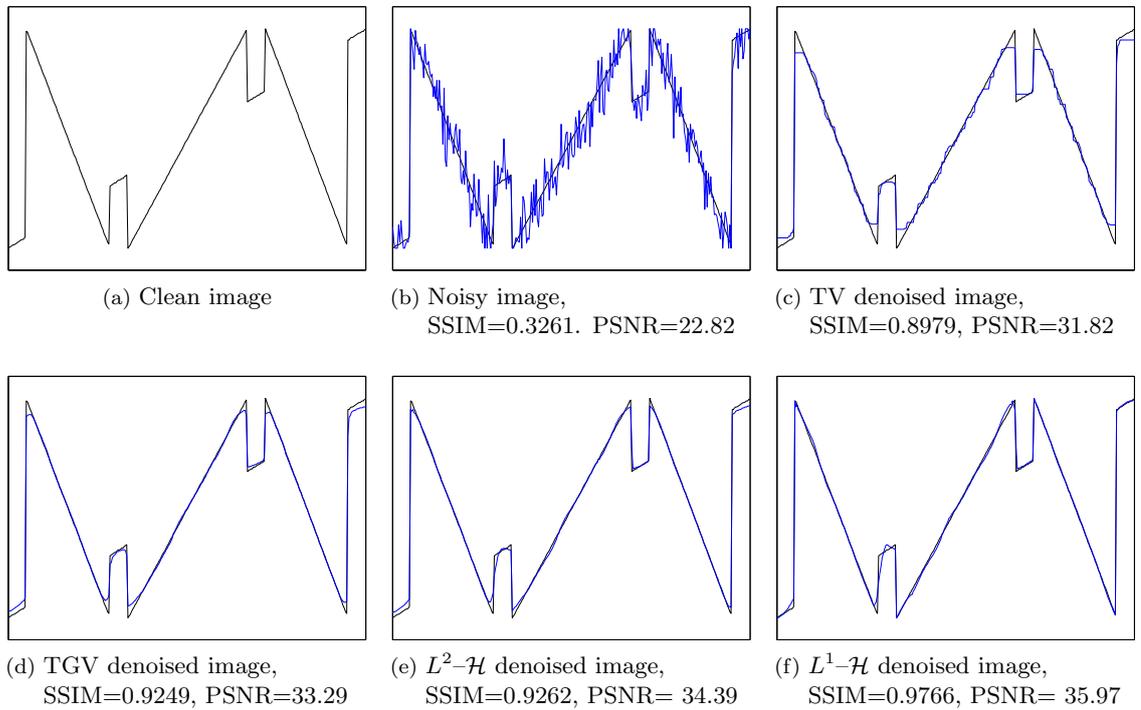
Figure 5.4: Corresponding middle row slices for the results of Figure 5.3.

Figure 5.3 depicts a denoising example for the same synthetic example we used in Section 3.6 (Gaussian noise of variance 0.005). In Figure 5.4 we show the corresponding middle row slices. We observe that the L^2 fidelity based, non-local Hessian denoised image, is of higher quality then the TGV denoised one, as this is confirmed by both PSNR and SSIM quality measures, see Figure 5.3(e). Notice however that we can achieve a far better result using the L^1 norm in the fidelity term, Figure 5.3(f), despite the fact that the noise is Gaussian. This is because this special selection of the weights σ_x allows for a use of a quite high regularisation parameter, ($\alpha = 1.2$ in that example), while still preserving the edges. In the same time the L^1 fidelity term is capable of preserving the contrast better than the L^2 case, compare the slices in Figures 5.4(e) and 5.4(f).

Chapter 6

Conclusions

Higher order variational regularisation methods have been an active field of research among the mathematical imaging community. These kind of methods have been proposed to tackle both classical and modern image processing tasks. The high quality of the reconstructions make them appealing for a variety of real world applications. In the same time, novel mathematical techniques are developed, aimed to implement these methods as well as to analyse their properties. We provide here a short summary about how the present thesis contributed to this field and we give some directions for future research work.

In Chapter 3 we introduced a combined first and second order method for image restoration. After a rigorous analysis of the well-posedness of the model, we applied the method for image denoising, deblurring and inpainting, providing also an online user-friendly implementation for the latter task in IPOL's website. The simplicity of the model combined with the split Bregman algorithm leads to a very efficient implementation, indicating the suitability of the method as a fast pre-processing technique that approaches the state of the art results as far as restoration quality in concerned. We think that a fully automatisation of the method, i.e., automatic selection of the optimal parameters α and β, is a usefull research direction that will enhance further the applicability of the method. Among other techniques, this optimal parameter learning can be achieved for instance through bilevel optimisation approaches [DlRS12, KP13, CDlRS14].

In Chapter 4 we contributed to the analysis of the second order total generalised variation model, a state of the art higher order regulariser, focusing mainly in the one dimensional case. We exhibited results concerning the structure of solutions of the L^2–TGV problem as well as the relationship of TV and TGV both in dimension one and two. We computed exact solutions for simple data functions, providing further insights into the regularising mechanism of TGV. A future research direction involves extension of these results in higher dimensions and investigation of the analytical properties of the model for other imaging tasks, e.g., deblurring. Identifying ground states for TGV in the spirit of [BB13] is another possibility.

Finally in Chapter 5 we introduced and analysed a non-local Hessian functional. As the non-locality vanishes, we proved the convergence of the non-local Hessian of a function u to the continuous analogue with respect to the natural topology that corresponds to the regularity of u. This analysis led to novel characterisations of some higher order Sobolev and BV spaces, namely, $W^{2,p}(\mathbb{R}^d)$ and $\mathrm{BV}^2(\mathbb{R}^d)$. Defining a non-local Hessian in an alternative, implicit way, resulted in a model that can be easily handled numerically for image processing purposes. We gave a flavour of this application in image denoising, showing that the model is capable of obtaining true piecewise affine reconstructions, being in that sense, superior to TGV. Future work involves the investigation of the application of the method to other imaging tasks as well as a study of the relationship between the explicit and the implicit formulation of the non-local Hessian functional.

Appendix A

A.1 Exclusion of the cases N-0-P2-J, N-0-P1-C, N-0-P2-C

In Section 4.6.1 we mentioned that solutions of the type N-0-P1-J, N-0-P2-J, N-0-P1-C and N-0-P2-C cannot occur when we consider the one dimensional L^2–$\mathrm{TGV}^2_{\beta,\alpha}$ minimisation problem with data $f_{\mathrm{p.c.}}$. In Proposition 4.6.2 we proved that fact for solutions of the type N-0-P1-J. Here, we do the same for the cases N-0-P2-J, N-0-P1-C and N-0-P2-C in Propositions A.1.1, A.1.2 and A.1.3 respectively.

Proposition A.1.1. *The case* N-0-P2-J *cannot occur.*

Proof. Suppose that there exist points $0 < x_1 < \tilde{x}_1 < x_2 < L$ such that $u < 0$ in $(0, x_1)$, $u = 0$ in $(x_1, \tilde{x}_1)$ and $u > 0$ in $(\tilde{x}_1, L)$. Suppose further that u' has a positive jump at x_2 and u has a jump discontinuity at $x = L$. Similarly to the proof of Proposition 4.6.2 we can show that the gradient of u in $(0, x_1)$ and in $(\tilde{x}_1, x_2)$ is the same, say equal to c, and that $w = c$ in $(0, x_2)$. As in Proposition 4.6.2 we set

$$\ell_1 = x_1, \quad \ell_2 = \tilde{x}_1 - x_1, \quad \ell_3 = x_2 - \tilde{x}_1, \quad \ell_4 = L - x_3 \quad \text{with} \quad \ell_1 + \ell_2 + \ell_3 + \ell_4 = L,$$

and

$$v_1 = v\lfloor(0, x_1), \quad v_2 = v\lfloor(x_1, \tilde{x}_1), \quad v_3 = \lfloor(\tilde{x}_1, x_2), \quad v_4 = \lfloor(x_2, L).$$

As in Proposition 4.6.2 we have

$$v_1(x) = -\frac{\alpha}{3\ell_1^2}x^3 + \frac{\alpha}{\ell_1}x^2, \quad x \in (0, x_1), \tag{A.1}$$

$$v_2(x) = \alpha(x - \ell_1) + \frac{2\alpha\ell_1}{3}, \quad x \in (x_1, \tilde{x}_1), \tag{A.2}$$

$$v_3(x) = -\frac{\alpha}{3\ell_1^2}(x - \ell_1 - \ell_2)^3 + \alpha(x - \ell_1 - \ell_2) + \alpha\ell_2 + \frac{2\alpha\ell_1}{3}, \quad x \in (\tilde{x}_1, x_2). \tag{A.3}$$

Now the extra conditions for v_3 are

$$v_3'(\ell_1 + \ell_2 + \ell_3) = 0, \tag{A.4}$$

and

$$v_3(\ell_1 + \ell_2 + \ell_3) = \beta. \tag{A.5}$$

Condition (A.4) gives

$$\ell_3 = \ell_1, \tag{A.6}$$

while condition (A.5) in combination with (A.6) gives

$$\frac{4}{3}\ell_1 + \ell_2 = \frac{\beta}{\alpha}. \tag{A.7}$$

The function v_4 will be also a cubic function

$$v_4(x) = a_4 x^3 + b_4 x^2 + c_4 x + d_4, \quad x \in (x_2, L).$$

We have v_4 must satisfy the conditions at $x = \ell_1 + \ell_2 + \ell_3$:

$$v_4(\ell_1+\ell_2+\ell_3) = \beta, \quad v_4'(\ell_1+\ell_2+\ell_3) = 0, \quad v_4''(\ell_1+\ell_2+\ell_3) = v_3''(\ell_1+\ell_2+\ell_3) = -\frac{2\alpha}{\ell_1}, \tag{A.8}$$

from which we get that

$$v_4 = \alpha_4(x - \ell_1 - \ell_2 - \ell_3)^3 - \frac{\alpha}{\ell_1}(x - \ell_1 - \ell_2 - \ell_3)^2 + \beta, \tag{A.9}$$

plus the following conditions on $x = \ell_1 + \ell_2 + \ell_3 + \ell_4$:

$$v_4(\ell_1 + \ell_2 + \ell_3 + \ell_4) = 0, \quad v_4'(\ell_1 + \ell_2 + \ell_3 + \ell_4) = -\alpha, \tag{A.10}$$

from which, in combination with (A.9) we get

$$a_4\ell_4^3 - \frac{\alpha}{\ell_1}\ell_4^2 + \beta = 0, \tag{A.11}$$

$$3a_4\ell_4^2 - \frac{2\alpha}{\ell_1}\ell_4 = -\alpha. \tag{A.12}$$

Multiplying (A.12) by ℓ_4 and using (A.11) we end up to

$$\ell_4^2 + \ell_1\ell_4 - \gamma\ell_1 = 0, \tag{A.13}$$

where we define

$$\gamma := 3\frac{\beta}{\alpha}.$$

Solving (A.13) with respect to ℓ_4 we get the (positive solution):

$$\ell_4 = \frac{\sqrt{\ell_1^2 + 4\gamma\ell_1} - \ell_1}{2}. \tag{A.14}$$

We now use the fact that $\ell_1 + \ell_2 + \ell_3 + \ell_4 = L$ in combination with (A.6), (A.7) and (A.14) and we get

$$\frac{\sqrt{\ell_1^2 + 4\gamma\ell_1}}{2} = L - \frac{\gamma}{3} - \frac{\ell_1}{6}. \tag{A.15}$$

Squaring up both sides of (A.15) after a bit of algebra we end up to the following quadratic equation for ℓ_1

$$2\ell_1^2 + \ell_1(8\gamma + 3L) - (3L - \gamma)^2 = 0. \tag{A.16}$$

whose only possible positive root is

$$\ell_1 = \frac{3\sqrt{9L^2 + 8\gamma^2} - (8\gamma + 3L)}{4}. \tag{A.17}$$

However we must make sure that $\ell_2 > 0$ thus from (A.7) we must impose

$$0 < \ell_1 < \gamma/4. \tag{A.18}$$

We claim now that the condition $|v_4'| \le \alpha$ is violated. Notice firstly that $a_4 > 0$. Indeed from (A.12) we have that $a_4 > 0$ if and only if $\ell_4 > \ell_1/2$ and using (A.14) that is if and only if $\ell_1 < \frac{4}{3}\gamma$ which is true from (A.18). Using that and the fact that $v_4'(\ell_1 + \ell_2 + \ell_3) = 0$, $v_4'(L) = -\alpha$ if $v_4''(x) = 0$ for some $x \in (\ell_1 + \ell_2 + \ell_3, L)$ then $v_4'(x) < v_4'(L) = -\alpha$ and we are done. Indeed we have that v_4'' vanishes at $x_0 = \ell_1 + \ell_2 + \ell_3 + \frac{\alpha}{3a_4\ell_1}$ and we have that $x_0 < L$ if and only if $\ell_4 > \ell_1$ which again from (A.14) that is if and only if $\ell_1 < \frac{\gamma}{2}$ which is true from (A.18).

We conclude that neither this type of solution is possible.

$\square$

Proposition A.1.2. *The case* N-0-P1-C *cannot occur.*

Proof. Suppose that there exist points $0 < x_1 < \tilde{x}_1 < L$ such that $u < 0$ in $(0, x_1)$, $u = 0$ in $(x_1, \tilde{x}_1)$ and $u > 0$ in $(\tilde{x}_1, L)$. Suppose further that u' is constant in $(\tilde{x}_1, L)$ and u is continuous at $x = L$. Similarly to the proof of Propositions 4.6.2 and A.1.1 we can show that the gradient of u in $(0, x_1)$ and in $(\tilde{x}_1, L)$ is the same, say equal to c, and that $w = c$ in $(0, L)$. Note that due to symmetry we have that condition $u(L) = h/2$ holds. As in the Propositions 4.6.2 and A.1.1 we have that, setting $\ell_1 = x_1$, $\ell_2 = \tilde{x}_1 - x_1$, $\ell_3 = L - \tilde{x}_1$,

$$v_1(x) = -\frac{\alpha}{3\ell_1^2}x^3 + \frac{\alpha}{\ell_1}x^2, \quad x \in (0, x_1), \tag{A.19}$$

$$v_2(x) = \alpha(x - \ell_1) + \frac{2\alpha\ell_1}{3}, \quad x \in (x_1, \tilde{x}_1), \tag{A.20}$$

$$v_3(x) = -\frac{\alpha}{3\ell_1^2}(x - \ell_1 - \ell_2)^3 + \alpha(x - \ell_1 - \ell_2) + \alpha\ell_2 + \frac{2\alpha\ell_1}{3}, \quad x \in (\tilde{x}_1, L), \tag{A.21}$$

and v_3 has the following conditions on $x = L = \ell_1 + \ell_2 + \ell_3$:

$$v_3(\ell_1 + \ell_2 + \ell_3) = 0 \iff -\frac{\ell_3^3}{3\ell_1^2} + \ell_3 + \ell_2 + \frac{2\ell_1}{3} = 0, \tag{A.22}$$

$$|v_3'(\ell_1 + \ell_2 + \ell_3)| \le \alpha \iff 0 \le \frac{\ell_3^2}{\ell_1^2} \le 2, \tag{A.23}$$

$$-v_3''(\ell_1 + \ell_2 + \ell_3) = \frac{h}{2} \iff \ell_1^2 = \frac{4\alpha\ell_3}{h}. \tag{A.24}$$

Combining (A.23) and (A.24) we get

$$0 \le \ell_3 \le \frac{8\alpha}{h}. \tag{A.25}$$

Moreover combining (A.22) and (A.24) we get

$$\frac{-h\ell_3^2}{12\alpha} + \ell_3 + \ell_2 + \frac{2\ell_1}{3} = 0. \tag{A.26}$$

However one can easily check that the term $\frac{-h\ell_3^2}{12\alpha} + \ell_3$ is positive if and only if $0 \le \ell_3 \le \frac{12\alpha}{h}$, which is the case from condition (A.25) thus (A.26) forms a contradiction.

We conclude that neither this solution can happen. $\qquad\square$

Proposition A.1.3. *The case* N-0-P2-C *cannot occur.*

Proof. Suppose that there exist points $0 < x_1 < \tilde{x}_1 < x_2 < L$ such that $u < 0$ in $(0, x_1)$, $u = 0$ in $(x_1, \tilde{x}_1)$ and $u > 0$ in $(\tilde{x}_1, L)$. Suppose further that u' has a positive jump at x_2 and u is continuous at $x = L$. As in the previous cases we have

$$v_1(x) = -\frac{\alpha}{3\ell_1^2}x^3 + \frac{\alpha}{\ell_1}x^2, \quad x \in (0, x_1), \tag{A.27}$$

$$v_2(x) = \alpha(x - \ell_1) + \frac{2\alpha\ell_1}{3}, \quad x \in (x_1, \tilde{x}_1), \tag{A.28}$$

$$v_3(x) = -\frac{\alpha}{3\ell_1^2}(x - \ell_1 - \ell_2)^3 + \alpha(x - \ell_1 - \ell_2) + \alpha\ell_2 + \frac{2\alpha\ell_1}{3}, \quad x \in (\tilde{x}_1, x_2), \tag{A.29}$$

with v_3 having the following conditions

$$v_3'(\ell_1 + \ell_2 + \ell_3) = 0 \iff \ell_3 = \ell_1, \tag{A.30}$$

$$v_3(\ell_1 + \ell_2 + \ell_3) = \beta \iff \frac{4}{3}\ell_1 + \ell_2 = \frac{\beta}{\alpha}. \tag{A.31}$$

Finally as before, v_4 will be of the form

$$v_4 = a_4(x - \ell_1 - \ell_2 - \ell_3)^3 - \frac{\alpha}{\ell_1}(x - \ell_1 - \ell_2 - \ell_3)^2 + \beta, \quad x \in (x_2, L), \tag{A.32}$$

with the following conditions on $x = L = \ell_1 + \ell_2 + \ell_3 + \ell_4$:

$$v_4(\ell_1 + \ell_2 + \ell_3 + \ell_4) = 0 \iff a_4\ell_4^3 - \frac{\alpha\ell_4^2}{\ell_1} + \beta = 0, \tag{A.33}$$

$$|v_4'(\ell_1 + \ell_2 + \ell_3 + \ell_4)| \le \alpha \iff \left| 3a_4\ell_4^2 - \frac{2\alpha\ell_4}{\ell_1} \right| \le \alpha, \tag{A.34}$$

$$-v_4'(\ell_1 + \ell_2 + \ell_3 + \ell_4) = \frac{h}{2} \iff -6a_4\ell_4 + \frac{2\alpha}{\ell_1} = \frac{h}{2}. \tag{A.35}$$

Combining (A.34) and (A.35) we end up to

$$\frac{\ell_4}{\ell_1} + \frac{h\ell_4}{4\alpha} \le 1, \tag{A.36}$$

which in particular implies that

$$\ell_4 < \ell_1. \tag{A.37}$$

Observe now that because of the fact that $\ell_2 > 0$ from (A.31) we get that $\ell_1 < \gamma/4$. But now, since $v_4(\ell_1 + \ell_2 + \ell_3) = \beta$, $v_4(\ell_1 + \ell_2 + \ell_3 + \ell_4) = 0$, from the mean value theorem and the fact that the derivative of v_4 is decreasing we have that

$$v_4'(\ell_1 + \ell_2 + \ell_3 + \ell_4) < -\frac{\beta}{\ell_4}. \tag{A.38}$$

Moreover

$$\ell_4 < \ell_1 < \frac{\gamma}{4} \Rightarrow -\frac{\beta}{\ell_4} < -\frac{4\alpha}{3} < -\alpha,$$

and thus from (A.38) we have that $v_4'(L) < -\alpha$, which is a contradiction.

Thus neither this kind of solution is possible.

$\square$

A.2 Some useful Lemmas

The following lemma (pointed out by Luca Calatroni) was used in the proof of Theorem 3.5.1.

Lemma A.2.1 (Kronecker's lemma). *Suppose that $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ are two sequences of real numbers such that $\sum_{n=1}^{\infty} a_n < \infty$ and $0 < b_1 \leq b_2 \leq \ldots$ with $b_n \to \infty$. Then*

$$\frac{1}{b_n} \sum_{k=1}^{n} b_k a_k \to 0, \quad as\ n \to \infty.$$

In particular, if $(c_n)_{n\in\mathbb{N}}$ is a decreasing positive real sequence such that $\sum_{n=1}^{\infty} c_n^2 < \infty$, then

$$c_n \sum_{k=1}^{n} c_k \to 0, \quad as\ n \to \infty.$$

The following Lemma was used in the proof of Lemma 5.2.1.

Lemma A.2.2. *Let $d \geq 2$. Then for every $i, j = 1, \ldots, d$ with $i \neq j$ and for every $R > 0$ we have*

$$\int_{|z|<R} \frac{z_i^2 z_j^2}{|z|^4} \rho_n(z) dz = \frac{1}{3} \int_{|z|<R} \frac{z_i^4}{|z|^4} \rho_n(z) dz. \tag{A.39}$$

In particular this means that

$$\int_{\mathbb{R}^d} \frac{z_i^2 z_j^2}{|z|^4} \rho_n(z) dz = \frac{1}{3} \int_{\mathbb{R}^d} \frac{z_i^4}{|z|^4} \rho_n(z) dz, \tag{A.40}$$

and

$$\int_{\mathbb{R}^d} \frac{z_i^2 z_j^2}{|z|^4} \rho_n(z) dz = \begin{cases} \frac{1}{N(N+2)} & \text{if } i \neq j, \\ \frac{3}{N(N+2)} & \text{if } i = j. \end{cases} \tag{A.41}$$

Proof. Note first that due to symmetry, it suffices to prove (A.39) for $i = 1$ and $j = 2$. We will consider three different cases depending on whether $d = 2$, $d = 3$ or $d > 3$.

$d = 2$: Employing the transformation $z_1 = r\cos\theta$, $z_2 = r\sin\theta$, we have that

$$\int_{|z|<R} \frac{z_1^2 z_2^2}{|z|^4} \rho_n(z) dz = \int_{r=0}^{R} r\hat{\rho}_n(r) dr \int_{0}^{2\pi} \cos^2\theta \sin^2\theta\, d\theta = \frac{\pi}{4} \int_{r=0}^{R} r\hat{\rho}_n(r) dr,$$

$$\int_{|z|<R} \frac{z_1^4}{|z|^4} \rho_n(z) dz = \int_{r=0}^{R} r\hat{\rho}_n(r) dr \int_{0}^{2\pi} \cos^4\theta\, d\theta = \frac{3\pi}{4} \int_{r=0}^{R} r\hat{\rho}_n(r) dr,$$

and thus, (A.39) holds.

d = 3: Here we employ the substitution $z_1 = r\cos\theta_1$, $z_2 = r\sin\theta_1\cos\theta_2$. Taking into account that the determinant of Jacobian of this substitution is $r^2\sin\theta_1$ we have that

$$\int_{|z|<R} \frac{z_1^2 z_2^2}{|z|^4}\rho_n(z)dz = \int_{r=0}^R r^2\hat\rho_n(r)dr \int_0^\pi \cos^2\theta_1\sin^3\theta_1\,d\theta_1 \int_{\theta_2=0}^{2\pi}\cos^2\theta_2 d\theta_2$$

$$= \frac{4\pi}{15}\int_{r=0}^R r^2\hat\rho_n(r)dr,$$

$$\int_{|z|<R} \frac{z_1^4}{|z|^4}\rho_n(z)dz = \int_{r=0}^R r^2\hat\rho_n(r)dr \int_0^\pi \cos^4\theta_1\sin\theta_1\,d\theta_1 \int_0^{2\pi} d\theta_2$$

$$= \frac{4\pi}{5}\int_{r=0}^R r^2\hat\rho_n(r)dr,$$

hence, (A.39) holds again.

d > 3: For this case, we employ the hyperspherical coordinate transformation:

$$z_1 = r\cos\theta_1,$$
$$z_2 = r\sin\theta_1\cos\theta_2,$$
$$z_3 = r\sin\theta_1\sin\theta_2\cos\theta_3,$$
$$\vdots$$
$$z_{d-1} = r\sin\theta_1\ldots\sin\theta_{d-2}\cos\theta_{d-1},$$
$$z_d = r\sin\theta_1\ldots\sin\theta_{d-2}\sin\theta_{d-1}.$$

Here $\theta_1,\theta_2,\ldots,\theta_{d-2}$ range over $[0,\pi)$ while θ_{d-1} ranges over $[0,2\pi)$. The determinant of the Jacobian of theis transformation is $r^{d-1}\sin^{d-2}\theta_1\sin^{d-3}\theta_2\ldots\sin\theta_{d-2}$, thus we can calculate

$$\int_{|z|<R}\frac{z_1^2 z_2^2}{|z|^4}\rho_n(z)dz = \underbrace{\int_0^R r^{d-1}\hat\rho_n(r)dr\int_0^\pi \sin^{d-4}\theta_3\,d\theta_3\cdots\int_0^\pi \sin\theta_{d-2}\,d\theta_{d-2}\int_0^{2\pi} d\theta_{d-1}}_{C}$$

$$\cdot\int_0^\pi \cos^2\theta_1\sin^d\theta_1\,d\theta_1\int_0^\pi \cos^2\theta_2\sin^{d-3}\theta_2\,d\theta_2$$

$$= C\frac{\pi\Gamma\left(\frac{d}{2}-1\right)}{4\Gamma\left(\frac{d}{2}+2\right)},$$

$$\int_{|z|<R}\frac{z_1^4}{|z|^4}\rho_n(z)dz = \underbrace{\int_0^R r^{d-1}\hat\rho_n(r)dr\int_0^\pi \sin^{d-4}\theta_3\,d\theta_3\cdots\int_0^\pi \sin\theta_{d-2}\,d\theta_{d-2}\int_0^{2\pi} d\theta_{d-1}}_{C}$$

$$\cdot\int_0^\pi \cos^4\theta_1\sin^{d-2}\theta_1\,d\theta_1\int_0^\pi \sin^{d-3}\theta_2\,d\theta_2$$

$$= C\frac{3\pi\Gamma\left(\frac{d}{2}-1\right)}{4\Gamma\left(\frac{d}{2}+2\right)}.$$

Hence, (A.39) holds in this case as well and the proof is complete. □

Finally, the following lemma was used in the proof of Theorem 4.5.3.

Lemma A.2.3. *Let Ω be a square centred at the origin and suppose that $f = (f_1, f_2) \in L^1(\Omega, \mathbb{R}^2)$ satisfies the symmetry properties*

$$f(x) = -f(-x), \quad f_1(x_1, x_2) = f(x_1, -x_2), \quad f_2(x_1, x_2) = f_2(-x_1, x_2), \tag{A.42}$$

$$f_1(O_{\pi/2}x) = f_2(x), \tag{A.43}$$

for almost every $x = (x_1, x_2) \in \Omega$. Then the minimisation problem

$$\min_{w \in \mathrm{Ker}\mathcal{E}} \|f - w\|_{L^1(\Omega, \mathbb{R}^2)}, \tag{A.44}$$

admits $w = 0$, as a solution.

Proof. Since

$$\mathrm{Ker}\mathcal{E} = \{r \in L^1(\Omega, \mathbb{R}^2) : \ r(x) = Ax + B, \ B \in \mathbb{R}^2, \ A \in \mathbb{R}^{2\times 2} \text{ is a skew symmetric matrix}\},$$

we have that the minimisation problem (A.44) is equivalent to

$$\min_{a, b_1, b_2 \in \mathbb{R}} \int_\Omega \left((f_1(x) - ax_2 - b_1)^2 + (f_2(x) + ax_1 - b_2)^2 \right)^{1/2} dx. \tag{A.45}$$

Since f_1 and f_2 coincide after a $\pi/2$ rotation, the problem (A.45) is now equivalent to

$$\min_{a, b \in \mathbb{R}} \sqrt{2} \int_\Omega |f_2(x) + ax_1 - b| \, dx. \tag{A.46}$$

From the symmetries (A.43) we have that if a minimises (A.45) the same is true for $-a$. Thus from convexity of the problem (A.46) we have that also $a = 0$ is an admissible solution. Thus finally we have that the problem (A.46) is equivalent to

$$\min_{b \in \mathbb{R}} \sqrt{2} \int_\Omega |f_2(x) - b| \, dx. \tag{A.47}$$

Now we take advantage of the fact that $f_2(x_1, -x_2) = -f_2(x_1, x_2)$ for almost every $(x_1, x_2) \in \Omega$. In that case it is obvious that if b solves (A.47) then the same is true for $-b$, thus again from convexity we have that $b = 0$ is an admissible solution. □

Bibliography

[AB86]   H. Attouch and H. Brezis, *Duality for the sum of convex functions in general Banach spaces*, North-Holland Mathematical Library **34** (1986), 125–133. (Cited on page 53).

[AC94]   M. Amar and V.D. Cicco, *Relaxation of quasi-convex integrals of arbitrary order*, Proceedings of the Royal Society of Edinburgh: Section A Mathematics **124** (1994), no. 5, 927–946, http://dx.doi.org/10.1017/S0308210500022423. (Cited on page 66).

[ACC05]   F. Alter, V. Caselles, and A. Chambolle, *A characterization of convex calibrable sets in* $\mathbb{R}^N$, Mathematische Annalen **332** (2005), no. 2, 329–366, http://dx.doi.org/10.1007/s00208-004-0628-9. (Cited on page 25).

[ACHR06]   A. Almansa, V. Caselles, G. Haro, and B. Rougé, *Restoration and zoom of irregularly sampled, blurred, and noisy images by accurate total variation minimization with local constraints*, Multiscale Modeling & Simulation **5** (2006), no. 1, 235–272, http://dx.doi.org/10.1137/050634086. (Cited on page 25).

[ACS09]   P. Arias, V. Caselles, and G. Sapiro, *A variational framework for non-local image inpainting*, Energy Minimization Methods in Computer Vision and Pattern Recognition, 2009, http://dx.doi.org/10.1007/978-3-642-03641-5_26, pp. 345–358. (Cited on page 32).

[ADF05]   F. Alter, S. Durand, and J. Froment, *Adapted total variation for artifact free decompression of JPEG images*, Journal of Mathematical Imaging and Vision **23** (2005), no. 2, 199–211, http://dx.doi.org/10.1007/s10851-005-6467-9. (Cited on page 25).

[AFP00]   L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*, Oxford University Press, USA, 2000. (Cited on pages 38, 40, 59, 60, and 125).

[AK09]   G. Aubert and P. Kornprobst, *Can the nonlocal characterization of Sobolev spaces by Bourgain et al. be useful for solving variational problems?*, SIAM Journal on Numerical Analysis **47** (2009), no. 2, 844–860, http://dx.doi.org/10.1137/070696751. (Cited on page 36).

[All08a]   W. Allard, *Total variation regularization for image denoising, I. Geometric theory*, SIAM Journal on Mathematical Analysis **39** (2008), no. 4, 1150–1190, http://dx.doi.org/10.1137/060662617. (Cited on page 26).

[All08b]   _____, *Total variation regularization for image denoising, II. Examples*, SIAM Journal on Imaging Sciences **1** (2008), no. 4, 400–417, http://dx.doi.org/10.1137/070698749. (Cited on page 26).

[All09]   _____, *Total variation regularization for image denoising, III. Examples.*, SIAM Journal on Imaging Sciences **2** (2009), no. 2, 532–568, http://dx.doi.org/10.1137/070711128. (Cited on page 26).

[AV94]   R. Acar and C.R. Vogel, *Analysis of bounded variation penalty methods for ill-posed problems*, Inverse Problems **10** (1994), no. 6, 1217–1229, http://dx.doi.org/10.1088/0266-5611/10/6/003. (Cited on page 25).

[BB11]   M. Benning and M. Burger, *Error estimates for general fidelities*, Electronic Transactions on Numerical Analysis **38** (2011), 44–68. (Cited on page 71).

[BB13]   _____, *Ground states and singular vectors of convex variational regularization methods*, Methods and Applications of Analysis **20** (2013), no. 4, 295–334, http://dx.doi.org/10.4310/MAA.2013.v20.n4.a1. (Cited on pages 153 and 181).

[BBBM13]   M. Benning, C. Brune, M. Burger, and J. Müller, *Higher-order TV methods – Enhancement via Bregman iteration*, Journal of Scientific Computing **54** (2013), no. 2-3, 269–310, http://dx.doi.org/10.1007/s10915-012-9650-3. (Cited on pages 31 and 148).

[BBM01]   J. Bourgain, H. Brezis, and P. Mironescu, *Another look at Sobolev spaces*, Optimal Control and Partial Differential Equations. A Volume in Honour of A. Bensoussan's 60th Birthday (2001), 439–455. (Cited on pages 33, 35, 36, 158, 159, and 173).

[BC98]   P. Blomgren and T. F. Chan, *Color TV: total variation methods for restoration of vector-valued images*, IEEE Transactions on Image Processing **7** (1998), no. 3, 304–309, http://dx.doi.org/10.1109/83.661180. (Cited on page 25).

[BCM05]   A. Buades, B. Coll, and J.M. Morel, *A review of image denoising algorithms, with a new one*, Multiscale Modeling & Simulation **4** (2005), no. 2, 490–530, http://dx.doi.org/10.1137/040616024. (Cited on page 36).

[BDH13]   K. Bredies, Y. Dong, and M. Hintermüller, *Spatially dependent regularization parameter selection in total generalized variation models for image restoration*, International Journal of Computer Mathematics **90** (2013), no. 1, 109–123, http://dx.doi.org/10.1080/00207160.2012.700400. (Cited on page 31).

[BEG07]   A. Bertozzi, S. Esedoglu, and A. Gillette, *Analysis of a two-scale Cahn-Hilliard model for binary image inpainting*, Multiscale Modeling & Simulation **6** (2007), no. 3, 913–936, http://dx.doi.org/10.1137/060660631. (Cited on page 32).

[Ben11]   M. Benning, *Singular regularization of inverse problems. Bregman distances and applications to variational frameworks with singular regularization energies*, Ph.D. thesis, University of Münster, 2011. (Cited on pages 31, 138, and 148).

[BF91]   G. Buttazzo and L. Freddi, *Functionals defined on measures and applications to non equi-uniformly elliptic problems*, Annali di Matematica Pura ed Applicata **159** (1991), no. 1, 133–149, http://dx.doi.org/10.1007/BF01766298. (Cited on page 60).

[BGH⁺14]   M. Benning, L. Gladden, D. Holland, C.B. Schönlieb, and T. Valkonen, *Phase reconstruction from velocity-encoded MRI measurements – A survey of sparsity-promoting variational approaches*, Journal of Magnetic Resonance **238** (2014), 26–43, http://dx.doi.org/10.1016/j.jmr.2013.10.003. (Cited on pages 25 and 31).

[BH12a]   K. Bredies and M. Holler, *Artifact-free JPEG decompression with total generalized variation*, VISAP 2012: Proceedings of the International Conference on Computer Vision and Applications, 2012. (Cited on page 31).

[BH12b]   _____, *A total variation-based JPEG decompression model*, SIAM Journal on Imaging Sciences **5** (2012), no. 1, 366–393, http://dx.doi.org/10.1137/110833531. (Cited on page 25).

[BH13a]   _____, *Regularization of linear inverse problems with total generalized variation*, preprint (2013). (Cited on page 31).

[BH13b]   _____ , *A TGV regularized wavelet based zooming model*, Scale Space and Variational Methods in Computer Vision, Springer, 2013, `http://dx.doi.org/10.1007/978-3-642-38267-3_13`, pp. 149–160. (Cited on page 31).

[BHS09]   M. Burger, L. He, and C.B. Schönlieb, *Cahn-Hilliard inpainting and a generalization for grayvalue images*, SIAM Journal on Imaging Sciences **2** (2009), no. 4, 1129–1167, `http://dx.doi.org/10.1137/080728548`. (Cited on page 32).

[BKP10]   K. Bredies, K. Kunisch, and T. Pock, *Total generalized variation*, SIAM Journal on Imaging Sciences **3** (2010), no. 3, 492–526, `http://dx.doi.org/10.1137/090769521`. (Cited on pages 30, 31, 83, 107, and 109).

[BKV13]   K. Bredies, K. Kunisch, and T. Valkonen, *Properties of L^1-TGV^2 : The one-dimensional case*, Journal of Mathematical Analysis and Applications **398** (2013), no. 1, 438 – 454, `http://dx.doi.org/10.1016/j.jmaa.2012.08.053`. (Cited on pages 27, 31, 35, 109, 110, 111, 113, 114, 116, 117, 118, and 120).

[BL11]   K. Bredies and D. Lorenz, *Mathematische bildverarbeitung: Einführung in grundlagen und moderne theorie*, Vieweg+Teubner Verlag, 2011. (Cited on page 72).

[BM07]   F. Bornemann and T. März, *Fast image inpainting based on coherence transport*, Journal of Mathematical Imaging and Vision **28** (2007), no. 3, 259–278, `http://dx.doi.org/10.1007/s10851-007-0017-6`. (Cited on page 32).

[BMPS14]   M. Burger, J. Müller, E. Papoutsellis, and C.B. Schönlieb, *Total variation regularisation in measurement and image space for PET reconstruction*, arXiv preprint (2014), `http://arxiv.org/abs/1403.1272`. (Cited on page 25).

[BO04]   M. Burger and S. Osher, *Convergence rates of convex variational regularization*, Inverse problems **20** (2004), no. 5, 1411, `http://dx.doi.org/10.1088/0266-5611/20/5/005`. (Cited on pages 71 and 72).

[Bol90]   Béla Bollobás, *Linear analysis*, vol. 199, Cambridge University Press Cambridge, 1990. (Cited on page 39).

[BP10]   M. Bergounioux and L. Piffet, *A second-order model for image denoising*, Set-Valued and Variational Analysis **18** (2010), no. 3-4, 277–306, `http://dx.doi.org/10.1007/s11228-010-0156-6`. (Cited on pages 29, 56, and 83).

[BPW13]   K. Bredies, T. Pock, and B. Wirth, *A convex, lower semi-continuous approximation of Euler's elastica energy*, preprint (2013). (Cited on page 32).

[Bra02]   A. Braides, Γ-*convergence for beginners*, Oxford lecture series in mathematics and its applications, 2002. (Cited on pages 49 and 73).

[Bre83]   Haïm Brezis, *Analyse fonctionelle*, vol. 5, Masson, 1983. (Cited on pages 165 and 168).

[Bre14]   K. Bredies, *Recovering piecewise smooth multichannel images by minimization of convex functionals with total generalized variation penalty*, Efficient Algorithms for Global Optimization Methods in Computer Vision, Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2014, `http://dx.doi.org/10.1007/978-3-642-54774-4_3`, pp. 44–77. (Cited on pages 31, 55, 83, 93, 94, 126, and 153).

[BRH07]   M. Burger, E. Resmerita, and L. He, *Error estimation for Bregman iterations and inverse scale space methods in image restoration*, Computing **81** (2007), no. 2, 109–135, `http://dx.doi.org/10.1007/s00607-007-0245-z`. (Cited on page 71).

[BSCB00]   M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, *Image inpainting*, Proceedings of the 27th annual conference on Computer graphics and interactive techniques, 2000, `http://dx.doi.org/10.1145/344779.344972`, pp. 417–424. (Cited on page 32).

[BSW$^+$13]   C. Brune, A. Sawatzky, F. Wübbeling, T. Kösters, and M. Burger, *EM-TV methods for inverse problems with poisson noise*, Level Set and PDE Based Reconstruction Methods in Imaging, Lecture Notes in Mathematics, Springer International Publishing, 2013, `http://dx.doi.org/10.1007/978-3-319-01712-9_2`, pp. 71–142. (Cited on page 23).

[BT09]   A. Beck and M. Teboulle, *Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems*, IEEE Transactions on Image Processing **18** (2009), no. 11, 2419–2434, `http://dx.doi.org/10.1109/TIP.2009.2028250`. (Cited on page 25).

[BV11]   K. Bredies and T. Valkonen, *Inverse problems with second-order total generalized variation constraints*, Proceedings of SampTA 2011 - 9th International Conference on Sampling Theory and Applications, Singapore, 2011. (Cited on pages 30, 31, 109, and 153).

[CCN07]   Vicent Caselles, Antonin Chambolle, and Matteo Novaga, *The discontinuity set of solutions of the TV denoising problem and some extensions*, Multiscale

Modeling & Simulation **6** (2007), no. 3, 879–894, http://dx.doi.org/10.1137/070683003. (Cited on pages 26, 27, and 117).

[CDlRS14] L. Calatroni, J.C. De los Reyes, and C.B. Schönlieb, *Dynamic sampling schemes for optimal noise learning under multiple nonsmooth constraints*, arXiv preprint (2014), http://arxiv.org/abs/1403.1278. (Cited on pages 23 and 181).

[CE05] T.F. Chan and S. Esedoglu, *Aspects of total variation regularized L^1 function approximation*, SIAM Journal on Applied Mathematics (2005), 1817–1837, http://dx.doi.org/10.1137/040604297. (Cited on pages 23 and 26).

[CEP07] T.F. Chan, S. Esedoglu, and F.E. Park, *Image decomposition combining staircase reduction and texture extraction*, Journal of Visual Communication and Image Representation **18** (2007), no. 6, 464–486, http://dx.doi.org/10.1016/j.jvcir.2006.12.004. (Cited on page 29).

[CGMP09] F. Cao, Y. Gousseau, S. Masnou, and P. Pérez, *Geometrically guided exemplar-based inpainting*, http://hal.archives-ouvertes.fr/hal-00392018. (Cited on page 32).

[CKS01] T.F. Chan, S.H. Kang, and J. Shen, *Total variation denoising and enhancement of color images based on the CB and HSV color models*, Journal of Visual Communication and Image Representation **12** (2001), no. 4, 422 – 435, http://dx.doi.org/10.1006/jvci.2001.0491. (Cited on page 25).

[CKS02] _____, *Euler's elastica and curvature-based inpainting*, SIAM Journal on Applied Mathematics **63** (2002), no. 2, 564–592, http://dx.doi.org/10.1137/S0036139901390088. (Cited on pages 27 and 32).

[CL97] A. Chambolle and P.L. Lions, *Image recovery via total variation minimization and related problems*, Numerische Mathematik **76** (1997), 167–188, http://dx.doi.org/10.1007/s002110050258. (Cited on pages 25, 28, and 83).

[Cla12] C. Clason, *L^∞ fitting for inverse problems with uniform noise*, Inverse Problems **28** (2012), no. 10, 104007, http://dx.doi.org/10.1088/0266-5611/28/10/104007. (Cited on page 23).

[CMM01] T. Chan, A. Marquina, and P. Mulet, *High-order total variation-based image restoration*, SIAM Journal on Scientific Computing **22** (2001), no. 2, 503–516, http://dx.doi.org/10.1137/S1064827598344169. (Cited on page 29).

[CP11]   A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, Journal of Mathematical Imaging and Vision **40** (2011), no. 1, 120–145, http://dx.doi.org/10.1007/s10851-010-0251-1. (Cited on pages 31, 55, 83, 85, 126, and 153).

[CPT03]   A. Criminisi, P. Perez, and K. Toyama, *Object removal by exemplar-based inpainting*, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, 2003, http://dx.doi.org/10.1109/CVPR.2003.1211538, pp. II–721. (Cited on page 32).

[CS01]   T.F. Chan and J. Shen, *Nontexture inpainting by curvature-driven diffusions*, Journal of Visual Communication and Image Representation **12** (2001), no. 4, 436–449, http://dx.doi.org/10.1006/jvci.2001.0487. (Cited on page 32).

[CS05]   ———, *Variational image inpainting*, Communications on Pure and Applied Mathematics **58** (2005), no. 5, 579–619, http://dx.doi.org/10.1002/cpa.20075. (Cited on pages 25 and 27).

[CW98]   T.F. Chan and C.K. Wong, *Total variation blind deconvolution*, IEEE Transactions on Image Processing **7** (1998), no. 3, 370–375, http://dx.doi.org/10.1109/83.661187. (Cited on page 25).

[CW06]   P.L. Combettes and V.R. Wajs, *Signal recovery by proximal forward-backward splitting*, Multiscale Modeling and Simulation **4** (2006), no. 4, 1168–1200, http://dx.doi.org/10.1137/050626090. (Cited on page 80).

[DAG09]   V. Duval, J.F. Aujol, and Y. Gousseau, *The TVL1 model: a geometric point of view*, SIAM Journal on Multiscale Modeling and Simulation **8** (2009), no. 1, 154–189, http://dx.doi.org/10.1137/090757083. (Cited on pages 23 and 26).

[Dem85]   F. Demengel, *Fonctions à Hessien borné*, Annales de l'Institut Fourier **34** (1985), 155–190. (Cited on pages 29, 60, 61, and 64).

[DHKS10]   Y. Dong, M. Hintermüller, F. Knoll, and R. Stollberger, *Total variation denoising with spatially dependent regularization*, ISMRM 18th Annual Scientific Meeting and Exhibition Proceedings, 2010, p. 5088. (Cited on page 25).

[DiB02]   E. DiBenedetto, *Real analysis*, Springer, 2002. (Cited on page 39).

[DlRS12]   J.C. De los Reyes and C.B. Schönlieb, *Image denoising: learning noise distribution via PDE-constrained optimization*, to appear in Inverse Problems and Imaging (2012), http://arxiv.org/abs/1207.3425. (Cited on pages 23 and 181).

[DM93]   G. Dal Maso, *Introduction to Γ-convergence*, Birkhauser, 1993. (Cited on pages 38, 49, and 73).

[DMFLM09] G. Dal Maso, I. Fonseca, G. Leoni, and M. Morini, *A higher order model for image restoration: the one dimensional case*, SIAM Journal on Mathematical Analysis **40** (2009), no. 6, 2351–2391, http://dx.doi.org/10.1137/070697823. (Cited on pages 27 and 29).

[DS96]   D. Dobson and F. Santosa, *Recovery of blocky images from noisy and blurred data*, SIAM Journal on Applied Mathematics **56** (1996), no. 4, 1181–1198, http://dx.doi.org/10.1137/S003613999427560X. (Cited on pages 25 and 27).

[DT84]   F. Demengel and R. Temam, *Convex functions of a measure and applications*, Indiana University Mathematics Journal **33** (1984), 673–709. (Cited on pages 58 and 63).

[DWB09]   S. Didas, J. Weickert, and B. Burgeth, *Properties of higher order nonlinear diffusion filtering*, Journal of Mathematical Imaging and Vision **35** (2009), no. 3, 208–226, http://dx.doi.org/10.1007/s10851-009-0166-x. (Cited on page 30).

[Eck89]   J. Eckstein, *Splitting methods for monotone operators with applications to parallel optimisation*, Ph.D. thesis, Massachusetts Institute of Technology, 1989. (Cited on page 81).

[EG92]   L.C. Evans and R.F. Gariepy, *Measure theory and fine properties of functions*, CRC Press, Boca Raton, FL, 1992. (Cited on page 40).

[ES02]   S. Esedoglu and J. Shen, *Digital inpainting based on the Mumford-Shah-Euler image model*, European Journal of Applied Mathematics **13** (2002), no. 4, 353–370, http://dx.doi.org/10.1017/S0956792502004904. (Cited on page 32).

[ET76]   I. Ekeland and R. Temam, *Convex analysis and variational problems*, vol. 1, North Holland, 1976. (Cited on pages 38, 50, and 72).

[Eva10]   L.C. Evans, *Partial differential equations, volume 19 of graduate studies in mathematics, second edition*, American Mathematical Society, 2010. (Cited on pages 39, 65, and 111).

[EZC10]   E. Esser, X. Zhang, and T. Chan, *A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science*, SIAM Journal on Imaging Sciences **3** (2010), no. 4, 1015–1046, http://dx.doi.org/10.1137/09076934X. (Cited on page 80).

[FSHW04] C. Frohn-Schauf, S. Henn, and K. Witsch, *Nonlinear multigrid methods for total variation image denoising*, Computing and Visualization in Science **7** (2004), no. 3, 199–206, http://dx.doi.org/10.1007/s00791-004-0150-3. (Cited on page 56).

[Gab83] D. Gabay, *Chapter IX applications of the method of multipliers to variational inequalities*, Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems (Michel Fortin and Roland Glowinski, eds.), Studies in Mathematics and its Applications, vol. 15, Elsevier, 1983, pp. 299 – 331. (Cited on page 81).

[GB08] M. Grant and S. Boyd, *Graph implementations for nonsmooth convex programs*, Recent Advances in Learning and Control (V. Blondel, S. Boyd, and H. Kimura, eds.), Lecture Notes in Control and Information Sciences, Springer-Verlag Limited, 2008, http://stanford.edu/~boyd/graph_dcp.html, pp. 95–110. (Cited on page 178).

[GB14] ———, *CVX: Matlab software for disciplined convex programming, version 2.1*, http://cvxr.com/cvx, March 2014. (Cited on page 178).

[Giu84] E. Giusti, *Minimal surfaces and functions of bounded variation*, Birkhäuser, 1984. (Cited on page 40).

[GM01] M. Gobbino and M. G. Mora, *Finite-difference approximation of free-discontinuity problems*, Proceedings of the Royal Society of Edinburgh: Section A Mathematics **131** (2001), 567–595, http://dx.doi.org/10.1017/S0308210501000257. (Cited on pages 36 and 159).

[GO07] G. Gilboa and S. Osher, *Nonlocal linear image regularization and supervised segmentation*, Multiscale Modeling & Simulation **6** (2007), no. 2, 595–630, http://dx.doi.org/10.1137/060669358. (Cited on page 36).

[GO08] ———, *Nonlocal operators with applications to image processing*, Multiscale Modeling & Simulation **7** (2008), no. 3, 1005–1028, http://dx.doi.org/10.1137/070698592. (Cited on page 36).

[GO09] T. Goldstein and S. Osher, *The split Bregman method for L1 regularized problems*, SIAM Journal on Imaging Sciences **2** (2009), no. 2, 323–343, http://dx.doi.org/10.1137/080725891. (Cited on pages 31, 33, 55, 73, 77, 79, 80, 83, 87, 97, and 98).

[Gra07] M. Grasmair, *The equivalence of the taut string algorithm and BV-regularization*, Journal of Mathematical Imaging and Vision **27** (2007),

no. 1, 59–66, http://dx.doi.org/10.1007/s10851-006-9796-4. (Cited on page 26).

[HHK⁺03] W. Hinterberger, M. Hintermüller, K. Kunisch, M. Von Oehsen, and O. Scherzer, *Tube methods for BV regularization*, Journal of Mathematical Imaging and Vision **19** (2003), no. 3, 219–235, http://dx.doi.org/10.1023/A:1026276804745. (Cited on page 26).

[Hir08] R. Hirsch, *Seizing the light*, A history of photography (2008). (Cited on page 19).

[HL13] M. Hintermüller and A. Langer, *Subspace correction methods for a class of nonsmooth and nonadditive convex variational problems with mixed L^1/L^2 data-fidelity in image processing*, SIAM Journal on Imaging Sciences **6** (2013), no. 4, 2134–2173, http://dx.doi.org/10.1137/120894130. (Cited on page 23).

[Hol13] M. Holler, *Higher order regularization for model based data decompression*, Ph.D. thesis, University of Graz, 2013. (Cited on page 31).

[HS06a] W. Hinterberger and O. Scherzer, *Variational methods on the space of functions of bounded Hessian for convexification and denoising*, Computing **76** (2006), no. 1-2, 109–133, http://dx.doi.org/10.1007/s00607-005-0119-1. (Cited on page 29).

[HS06b] M. Hintermüler and G. Stadler, *An infeasible primal-dual algorithm for total bounded variation–based inf-convolution-type image restoration*, SIAM Journal on Scientific Computing **28** (2006), no. 1, 1–23, http://dx.doi.org/10.1137/040613263. (Cited on page 56).

[Jal12] K. Jalalzai, *Regularization of inverse problems in image processing*, Ph.D. thesis, École polytechnique, 2012. (Cited on page 117).

[Jal14] _____, *Discontinuities of the minimizers of the weighted or anisotropic total variation for image reconstruction*, arXiv preprint (2014), http://arxiv.org/abs/1402.0026. (Cited on page 117).

[KBPS11] F. Knoll, K. Bredies, T. Pock, and R. Stollberger, *Second order total generalized variation (TGV) for MRI*, Magnetic Resonance in Medicine **65** (2011), no. 2, 480–491, http://dx.doi.org/10.1002/mrm.22595. (Cited on page 31).

[KP13] K. Kunisch and T. Pock, *A bilevel optimization approach for parameter learning in variational models*, SIAM Journal on Imaging Sciences **6** (2013), no. 2, 938–983, http://dx.doi.org/10.1137/120882706. (Cited on page 181).

[LBU12]   S. Lefkimmiatis, A. Bourquard, and M. Unser, *Hessian-based norm regularization for image restoration with biomedical applications*, IEEE Transactions on Image Processing **21** (2012), 983–995, `http://dx.doi.org/10.1109/TIP.2011.2168232`. (Cited on page 29).

[LCA07]   T. Le, R. Chartrand, and T.J. Asaki, *A variational approach to reconstructing images corrupted by Poisson noise*, Journal of Mathematical Imaging and Vision **27** (2007), no. 3, 257–263, `http://dx.doi.org/10.1007/s10851-007-0652-y`. (Cited on pages 23 and 25).

[LL01]   E.H. Lieb and M. Loss, *Analysis, volume 14 of graduate studies in mathematics*, American Mathematical Society, Providence, RI, **4** (2001). (Cited on page 39).

[LLT03]   M. Lysaker, A. Lundervold, and X.C. Tai, *Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time*, IEEE Transactions on Image Processing **12** (2003), no. 12, 1579–1590, `http://dx.doi.org/10.1109/TIP.2003.819229`. (Cited on page 29).

[LPSS]   J. Lellmann, K. Papafitsoros, C.B. Schönlieb, and D. Spector, *Non-local higher-order regularisation*, in preparation. (Cited on pages 7, 32, 33, and 38).

[LT06]   M. Lysaker and X.C. Tai, *Iterative image restoration combining total variation minimization and a second-order functional*, International Journal of Computer Vision **66** (2006), no. 1, 5–18, `http://dx.doi.org/10.1007/s11263-005-3219-7`. (Cited on page 30).

[LTC13]   R. Lai, X. Tai, and T. Chan, *A ridge and corner preserving model for surface restoration*, SIAM Journal on Scientific Computing **35** (2013), no. 2, A675–A695, `http://dx.doi.org/10.1137/110846634`. (Cited on pages 29 and 32).

[Men12]   T. Mengesha, *Nonlocal Korn-type characterization of Sobolev vector fields*, Communications in Contemporary Mathematics **14** (2012), no. 4, `http://dx.doi.org/10.1142/S0219199712500289`. (Cited on page 173).

[Mey01]   Y. Meyer, *Oscillating patterns in image processing and nonlinear evolution equations: the fifteenth Dean Jacqueline B. Lewis memorial lectures*, vol. 22, American Mathematical Society, 2001. (Cited on page 25).

[MM98]   S. Masnou and J.M. Morel, *Level lines based disocclusion*, IEEE Proceedings of International Conference on Image Processing, 1998, `http://dx.doi.org/10.1109/ICIP.1998.999016`, pp. 259–263. (Cited on page 32).

[MS13]   T. Mengesha and D. Spector, *Localization of nonlocal gradients in various topologies*, CVGMT preprint (2013). (Cited on pages 36, 159, 169, and 173).

[Mül13]   J. Müller, *Advanced image reconstruction and denoising – Bregmanized (higher order) total variation and application in PET*, Ph.D. thesis, University of Münster, 2013. (Cited on pages 25, 30, 31, and 148).

[Nik02]   M. Nikolova, *Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers*, SIAM Journal on Numerical Analysis **40** (2002), no. 3, 965–994, `http://dx.doi.org/10.1137/S0036142901389165`. (Cited on pages 23 and 26).

[Nik04]   ――――, *A variational approach to remove outliers and impulse noise*, Journal of Mathematical Imaging and Vision **20** (2004), no. 1, 99–120, `http://dx.doi.org/10.1023/B:JMIV.0000011326.88682.e5`. (Cited on pages 23 and 25).

[Nir59]   L. Nirenberg, *On elliptic partial differential equations*, Annali della Scuola Normale Superiore di Pisa, 3^e serie **13** (1959), no. 2, 115–162. (Cited on page 175).

[NMS93]   M. Nitzberg, D. Mumford, and T. Shiota, *Filtering, segmentation, and depth*, Springer-Verlag New York, Inc., 1993. (Cited on page 32).

[NWC13]   M. Nikolova, Y.W. Wen, and R. Chan, *Exact histogram specification for digital images using a variational approach*, Journal of Mathematical Imaging and Vision **46** (2013), no. 3, 309–325, `http://dx.doi.org/10.1007/s10851-012-0401-8`. (Cited on page 23).

[OBG⁺]   S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, *An iterative regularization method for total variation-based image restoration*, Multiscale Modeling & Simulation **4**, 460–489, `http://dx.doi.org/10.1137/040605412`. (Cited on pages 73 and 77).

[OF03]   S. Osher and R. Fedkiw, *Level set methods and dynamic implicit surfaces*, Springer, 2003. (Cited on page 178).

[Pas12a]   Pascal Getreuer, *Rudin-Osher-Fatemi Total Variation Denoising using Split Bregman*, Image Processing On Line (2012), `http://dx.doi.org/10.5201/ipol.2012.g-tvd`. (Cited on pages 25, 83, 96, and 97).

[Pas12b]   _____, *Total Variation Deconvolution using Split Bregman*, Image Processing On Line (2012), http://dx.doi.org/10.5201/ipol.2012.g-tvdc. (Cited on pages 25, 83, 96, and 97).

[Pas12c]   _____, *Total Variation Inpainting using Split Bregman*, Image Processing On Line (2012), http://dx.doi.org/10.5201/ipol.2012.g-tvi. (Cited on pages 25, 27, 83, 96, and 97).

[PB13]   K. Papafitsoros and K. Bredies, *A study of the one dimensional total generalised variation regularisation problem*, submitted (2013), http://arxiv.org/abs/1309.5900. (Cited on pages 7, 32, 33, and 38).

[Pon04]   A.C. Ponce, *A new approach to Sobolev spaces and connections to Γ-convergence*, Calculus of Variations and Partial Differential Equations **19** (2004), no. 3, 229–255, http://dx.doi.org/10.1007/s00526-003-0195-z. (Cited on pages 36, 158, and 173).

[Pös08]   C. Pöschl, *Tikhonov regularization with general residual term*, Ph.D. thesis, Leopold-Franzens-Universität Innsbruck, Austria, 2008. (Cited on page 71).

[PS08]   C. Pöschl and O. Scherzer, *Characterization of minimisers of convex regularization functionals*, Contemporary mathematics **451** (2008), 219–248. (Cited on page 30).

[PS13]   _____, *Exact solutions of one-dimensional TGV*, arXiv preprint (2013), http://arxiv.org/abs/1309.7152. (Cited on page 35).

[PS14]   K. Papafitsoros and C.B. Schönlieb, *A combined first and second order variational approach for image reconstruction*, Journal of Mathematical Imaging and Vision **48** (2014), no. 2, 308–338, http://dx.doi.org/10.1007/s10851-013-0445-4. (Cited on pages 7, 32, and 38).

[PSS13]   K. Papafitsoros, C.B. Schönlieb, and B. Sengul, *Combined First and Second Order Total Variation Inpainting using Split Bregman*, Image Processing On Line **2013** (2013), 112–136, http://dx.doi.org/10.5201/ipol.2013.40. (Cited on pages 7, 32, 33, 38, 95, and 101).

[Rin00]   W. Ring, *Structural properties of solutions to total variation regularization problems*, ESAIM: Mathematical Modelling and Numerical Analysis **34** (2000), no. 4, 799–810, http://dx.doi.org/10.1051/m2an:2000104. (Cited on pages 26, 27, 35, and 116).

[ROF92]   L.I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D: Nonlinear Phenomena **60** (1992), no. 1-4,

259–268, http://dx.doi.org/10.1016/0167-2789(92)90242-F. (Cited on pages 24 and 25).

[Rud87]   W. Rudin, *Real and complex analysis, WCB*, McGraw-Hill, Boston, 1987. (Cited on page 40).

[Saw11]   A. Sawatzky, *(Nonlocal) total variation in medical imaging*, Ph.D. thesis, University of Münster, 2011. (Cited on page 36).

[SBBH09]   C.B. Schönlieb, A. Bertozzi, M. Burger, and L. He, *Image inpainting using a fourth-order total variation flow*, SAMPTA'09, International Conference on Sampling Theory and Applications, Marseilles (2009). (Cited on page 32).

[SBJM09]   A. Sawatzky, C. Brune, Müller J., and Burger M., *Total variation processing of images with Poisson statistics*, Computer Analysis of Images and Patterns, Springer, 2009, http://dx.doi.org/10.1007/978-3-642-03767-2_65, pp. 533–540. (Cited on page 23).

[SC02]   J. Shen and T.F. Chan, *Mathematical models for local nontexture inpaintings*, SIAM Journal on Applied Mathematics **62** (2002), no. 3, 1019–1043, http://dx.doi.org/10.1137/S0036139900368844. (Cited on pages 25 and 27).

[SC03]   D. Strong and T. Chan, *Edge-preserving and scale-dependent properties of total variation regularization*, Inverse Problems **19** (2003), no. 6, S165, http://dx.doi.org/10.1088/0266-5611/19/6/059. (Cited on page 26).

[Sch98]   O. Scherzer, *Denoising with higher order derivatives of bounded variation and an application to parameter estimation*, Computing **60** (1998), no. 1, 1–27, http://dx.doi.org/10.1007/BF02684327. (Cited on page 29).

[Sch09]   C.B. Schönlieb, *Modern PDE techniques for image inpainting*, Ph.D. thesis, University of Cambridge, 2009. (Cited on page 27).

[Set99]   J.A. Sethian, *Level set methods and fast marching methods: Evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science*, vol. 3, Cambridge University Press, 1999. (Cited on page 178).

[Set09]   S. Setzer, *Split Bregman algorithm, Douglas-Rachford splitting and frame shrinkage*, Scale Space and Variational Methods in Computer Vision **5567** (2009), 464–476, http://dx.doi.org/10.1007/978-3-642-02256-2_39. (Cited on pages 79, 80, and 81).

[SST11]   S. Setzer, G. Steidl, and T. Teuber, *Infimal convolution regularizations with discrete ℓ_1-type functionals*, Communications in Mathematical Sciences **9** (2011), 797–872, http://dx.doi.org/10.4310/CMS.2011.v9.n3.a7. (Cited on page 31).

[Tem85]   R. Temam, *Mathematical problems in plasticity*, vol. 15, Gauthier-Villars Paris, 1985. (Cited on pages 30 and 126).

[THC11]   X.C. Tai, J. Hahn, and G.J. Chung, *A fast algorithm for Euler's elastica model using augmented Lagrangian method*, SIAM Journal on Imaging Sciences **4** (2011), no. 1, 313–344, http://dx.doi.org/10.1137/100803730. (Cited on pages 32 and 103).

[Tik63]   A. Tikhonov, *Solution of incorrectly formulated problems and the regularization method*, Soviet Math. Dokl., vol. 5, 1963, p. 1035. (Cited on page 23).

[TS80]   R. Temam and G. Strang, *Functions of bounded deformation*, Archive for Rational Mechanics and Analysis **75** (1980), no. 1, 7–21, http://dx.doi.org/10.1007/BF00284617. (Cited on pages 30, 126, and 176).

[Val14a]   T. Valkonen, *The jump set under geometric regularisation. Part 1: Basic technique and first-order denoising*, arXiv preprint (2014), http://arxiv.org/abs/1407.1531. (Cited on page 117).

[Val14b]   _____, *The jump set under geometric regularisation. Part 2: Higher-order approaches*, arXiv preprint (2014), http://arxiv.org/abs/1407.2334. (Cited on page 117).

[VBK13]   T. Valkonen, K. Bredies, and F. Knoll, *Total generalized variation in diffusion tensor imaging*, SIAM Journal on Imaging Sciences **6** (2013), no. 1, 487–525, http://dx.doi.org/10.1137/120867172. (Cited on page 31).

[Ves01]   L. Vese, *A study in the BV space of a denoising-deblurring variational problem*, Applied Mathematics and Optimization **44** (2001), no. 2, 131–161. (Cited on pages 33, 56, 67, 68, and 71).

[Vog95]   C.R. Vogel, *A multigrid method for total variation-based image denoising*, Progress in Systems and Control Theory **20** (1995), 323–323, http://dx.doi.org/10.1007/978-1-4612-2574-4_21. (Cited on page 56).

[VW97]   P.S. Vassilevski and J.G. Wade, *A comparison of multilevel methods for total variation regularization*, Electronic Transactions on Numerical Analysis (1997), 255–270. (Cited on page 56).

[WB09]   Z. Wang and A.C. Bovik, *Mean squared error: Love it or leave it? a new look at signal fidelity measures*, IEEE Signal Processing Magazine **26** (2009), no. 1, 98–117, http://dx.doi.org/10.1109/MSP.2008.930649. (Cited on pages 34 and 83).

[WBSS04]   Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, *Image quality assessment: From error visibility to structural similarity*, IEEE Transactions on Image Processing **13** (2004), no. 4, 600–612, http://dx.doi.org/10.1109/TIP.2003.819861. (Cited on pages 34 and 83).

[WGO07]   Y. Wotao, D. Goldfarb, and S. Osher, *The total variation regularized L^1 model for multiscale decomposition*, Multiscale Modeling & Simulation **6** (2007), no. 1, 190–211, http://dx.doi.org/10.1137/060663027. (Cited on page 23).

[Whi23]   E. T. Whittaker, *On a new method of graduation*, Proceedings of the Edinburgh Mathematical Society **41** (1923), 63–75. (Cited on page 23).

[WT10]   C. Wu and X.C. Tai, *Augmented Lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models*, SIAM Journal on Imaging Sciences **3** (2010), no. 3, 300–339, http://dx.doi.org/10.1137/090767558. (Cited on pages 74 and 80).

[WYZ07]   Y. Wang, W. Yin, and Y. Zhang, *A fast algorithm for image deblurring with total variation regularization*. (Cited on page 25).

[YOGD08]   W. Yin, S. Osher, D. Goldfarb, and J. Darbon, *Bregman iterative algorithms for ℓ_1-minimisation with applications to compressed sensing*, SIAM Journal on Imaging Sciences **1** (2008), 142–168, http://dx.doi.org/10.1137/070703983. (Cited on pages 77 and 80).