NTNU
Norwegian University of
Science and Technology

Variational Methods for Coherence Enhancing Image Denoising

Lasse Åsmot

Master of Science in Physics and Mathematics
Submission date: June 2017
Supervisor: Markus Grasmair, IMF

Norwegian University of Science and Technology
Department of Mathematical Sciences

Abstract

Image denoising by regularization of coherence enhancing functionals have become increasingly standard due to their structural preservation properties. One example of coherence enhancing regularization is formulated by Weickert (1999) as a PDE. A typical drawback of such methods, however, is the creation of artifacts – structures created from random noise by the denoiser. In this paper, we use coherence enhancing regularization approaches to create and test low-level denoising algorithms based on the regularization method of Weickert. To combat the artifact generation of the method, we propose as an alternative a non-local functional, with the goal of inheriting the anisotropical enhancement properties of the Weickert functional, while at the same time suppressing artifact generation. Mathematically, the paper compares the two coherence enhancing functionals and covers in detail both theory on existence of minimizers and how to find these minima. Concluding the paper, we present some numerical results demonstrating the denoising properties of both the Weickert functional and the proposed non-local functional.

Sammendrag

Støyfjerning i bilder ved regularisering av koherensfremmende funksjonaler har i større grad blitt standard takket være de strukturbevarende egenskapene de har. Et eksempel på en slik metode ble formulert som en partiell differensialligning i Weickert (1999). En typisk ulempe ved bruk av slike metoder er forekommelser av artefakter – gjenstander som genereres fra tilfeldig støy gjennom støyfjerneren. Vi tar i denne avhandlingen i bruk koherensefremmende metoder til å lage og teste lavnivåalgoritmer for strøyfjerning, basert på metoden Weickert introduserte. For å redusere mengden artefakter som genereres, foreslår vi en alternativ ikke-lokal funksjonal. Målet ved dette alternativet er å beholde de anisotropifremmende egenskapene til Weickert-funksjonalen, samtidig som vi demper forekomster av artefakter. Matematisk sammenlignes de to koherensefremmende funksjonalene, og det gis en detaljert utledning av eksistensteori for minimum, samt en måte å finne dem på. Mot slutten av avhandlingen presenterer vi noen numeriske resultater som viser støyfjerningsegenskapene for både Weickert-funksjonalen og den foreslåtte ikke-lokale funksjonalen

Preface

The field of image processing is an interesting way of visualizing rather abstract mathematical concepts, while also being a highly applicable field of study. Image processing covers a lot of different subjects, ranging from scientific purposes like forensic evidence and seismological readings, to more everyday functionality like Snapchat® filters enjoyed by many people all over the world.

I have spent this year looking into image denoising. Throughout the research process, I have had to both brush up on old knowledge gained throughout my years as a student, as well as attain new knowledge to better understand the problems faced in the paper. I have been through the field of measure theory, functional analysis, optimization and numerics, and at the end of the process, I am left with deeper understandings of functionals, topologies and more – and I have some really cool images to show for it.

In this paper, the reader will be taken through mathematical considerations, where some knowledge of measure theory and functional analysis will be required. The discussions of the results are qualitative and more tangible even for those who lacks said knowledge.

I would like to express my deepest gratitude to my supervisor, Associate Professor Markus Grasmair, for his invaluable aid throughout the research process, without which the paper would not have come to be. A word of thanks also go to my fellow students, close friends and family, for helping me vent my frustrations, heightening my morale and simple reassurance.

Table of Contents

Abbreviations

Symbol	=	definition
(w)slsc	=	(weakly) sequentially lower semi-continuous

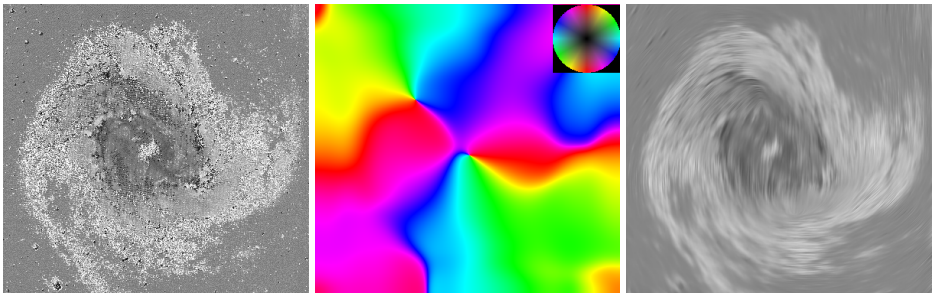
Chapter 1

Introduction

The denoising of images has applications in many fields, for instance in medical sciences and space exploration and discovery. In the field of digital signal analysis, signals are bound to be damaged and the data they carry is consequently distorted or lost. Such problems can arise from things like errant signal reading and other hardware defects. It could also arise from physical limitations like blurring because of finite lens sizes Jähne (2005) and random noise due to low photon counts.

This paper covers a mathematical and computational study of image denoising through variational methods, and the ultimate goal is to create software that can remove noise in an image, while at the same time enhance its anisotropic structures through coherence enhancing. A standard approach for regularization in image processing is to let images be described as functions $u : \Omega \to \mathbb{R}$ over some rectangle $\Omega \subset \mathbb{R}^2$, and define a functional $\mathcal{F}(u)$ penalizing unwanted properties in these functions. A regularized image $\bar{u}$ is thereafter defined as the minimizer of $\mathcal{F}$ in some function space. Usual choices of $\mathcal{F}$ are integrals of some density $f(x, u(x), \nabla u(x); u_0(x))$, or a non-local density $f(x, y, u(x), u(y), \nabla u(x), \nabla u(y); u_0(x))$. In the field of image processing, such a non-local density integral approach is called neighbourhood filtering, or more generally, patch-based filtering, and has been widely applied. Some examples of this can be found in the works of Buades et al. (2005) and Weickert (1999). Total variation filtering approaches are closely connected to regularization through partial differential equations Scherzer et al. (2009), and in this paper, we will focus on the former.

The method of anisotropic coherence enhancement seeks enhance internal image structures. To this end, one may define a *structure tensor* A and use it to recognize structures in an image. These structures are characterized by profoundly distinguished eigenvectors in some locale in an image. These eigenvectors are again related to their corresponding eigenvalues, the difference between which are called the coherence. Coherence enhancement thereby seeks to create clear anisotropic structures by increasing the difference of eigenvalues, and consequently the eigenvectors, of A. In figures 1.1a, 1.1b and 1.1c, one



(a) Original data. **(b)** Structure tensor. **(c)** Denoised image.

Figure 1.1: Example of how data can be denoised through structure tensor. In the top right corner of fig. 1.1b we can see the connection between color and direction of flow.

can see one example of data, its structure tensor, and the denoised data. As can be seen clearly, the denoiser has enhance flow-like structures in the image.

Figure 1.1 shows one result from the denoiser introduced in Weickert (1999). This article proposes a denoising approach that, while proposed as a PDE in the article, can be equivalently formulated as a regularization of the functional

$$\mathcal{F}_W(u) = \frac{1}{2}\|u - u_0\|_2^2 + \frac{\alpha}{p}\int_\Omega \left|A(u,x)\nabla u(x)\right|_E^p dx,$$

where the largest eigenvector of the local structure tensor A is pointing in a locally dominant direction of ∇u. Here and troughout the paper, $\|\cdot\|_E$ denotes the euclidean $\mathbb{R}^n$ norm. The exact definition of this A will be given in section 2.3.

The abovementioned functional effectively enhances anisotropic structures in an image, but it also poses a problem, namely that of *artifact* generation. Artifacts are structures which the denoiser creates from random noise, or other structural distortions that may enhances coherence. As in Grasmair and Lenzen (2010), we propose an alternative coherence enhancing functional, defined by

$$\mathcal{F}(u) = \frac{1}{2}\|u - u_0\|_2^2 + \frac{\alpha}{2}\|\nabla u\|_2^2 + \frac{\beta}{p}\int_\Omega\int_\Omega w(|x-y|)\left|\left\langle \nabla u(x), \nabla u(y)^\perp\right\rangle\right|^p dy\, dx.$$

With this functional we hope to reproduce the coherence enhancing results of $\mathcal{F}_W$ while at the same time suppress artifact generation. In the case when $p = 2$ the minimization of $\mathcal{F}$ is also significantly easier than the minimization of $\mathcal{F}_W$.

The goal of this paper is to denoise images through the functionals $\mathcal{F}_W$ and $\mathcal{F}$, prepare data for post-processing, and compare numerical performance inbetween these functionals with respect to the resulting denoised images. The functionals $\mathcal{F}_W$ and $\mathcal{F}$ are uniquely determined by the parameters $\alpha > 0$, $\beta > 0$ and $p > 1$, as well as the function $w : \mathbb{R}_{\geq 0} \to$

$\mathbb{R}_{\geq 0}$, and the choices of these affect the denoised images. Each term of the densities of $\mathcal{F}$ and $\mathcal{F}_W$ has specific properties of interest; the first term of both functionals ensures that the denoised image $\bar{u}$ resembles the data u_0. The second term of $\mathcal{F}$ is a standard regularization term, while the third term seeks to create contrasts, draw straight lines, and enhance non-local coherence to preserve the structures of the image. The second term of $\mathcal{F}_W$ tries to incorporate both regularization and coherence enhancement at once. Parameter choices are discussed in detail in chapter 4.

Non-local function classes like the functional $\mathcal{F}$ have been studied in (Boulanger et al., 2011, 131-154), and much of the theory needed in our study of $\mathcal{F}$ can directly or indirectly use the results from this work. Beyond the theory, this paper focuses on $p = 2$, which will significantly simplify all formulas, and a normalized gaussian kernel w.

The second chapter deals with existence of solutions to the minimization problems of $\mathcal{F}_W$ and $\mathcal{F}$, and derivation of formal gradients and the minimizing sequences. The third chapter describe a discretizations used in our numerical approach to find approximate solutions, closely following the steepest descent approach described in Nocedal and Wright (2006). The remainder of the paper discusses some numerical results, clearly demonstrating the coherence enhancing properties of both functionals, and how the denoisers are affected by parameter choices.

Chapter 2

Basic Theory

In this chapter we look in detail at the regularization of the functionals mentioned in the introduction and define them as minimising problems. Specifically, we look at the existence of minimizers through Tonelli's direct method and the derivation of formal gradients for both functionals, and prove explicitly that $\mathcal{F}$ is Gâteaux differentiable. We also find the function spaces in which the proposed denoising problem makes sense. Key concepts that are essential to create the denoising software, such as boundary conditions and convolutions over finite domains, are presented explicitly. We begin by introducing notation for some basic concepts that will be used throughout the chapter.

2.1 Diagonalization of real symmetric matrices

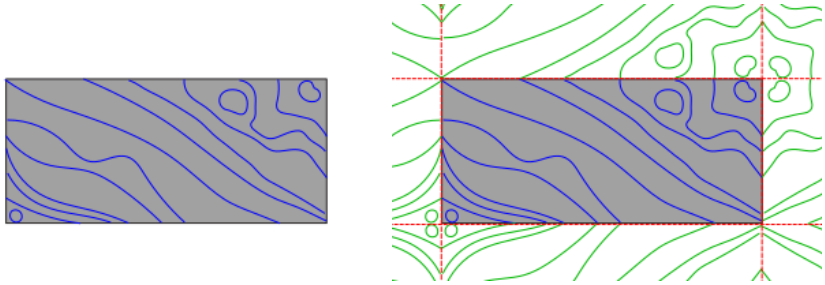
Let $A \in \mathbb{R}^{2\times 2}$ be symmetric i.e. there are real numbers a, b and c such that

$$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}.$$

Then there exists a unique diagonal matrix Σ of ordered eigenvalues of A, and a unique(up to signs) orthogonal matrix Q whose columns are eigenvectors corresponding to each of these eigenvalues, such that

$$A = Q\Sigma Q^T.$$

The matrix Q is not unique if the eigenvalues of A coincide, in which case $Q = \pm\mathcal{I}\delta$. We will revisit the consequences of this in section 4.1. For further details on the diagonalization, see appendix A.1.



(a) The level curves of u.

(b) The level curves of u_e and the mirror axes used in obtaining them.

Figure 2.1: The symmetric extension of u_e of u to $\mathbb{R}^2$.

2.2 Convolution

In this section we define what we mean by convolution. In short, the convolution of two functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ is defined by

$$\big(f * g\big)(x) = \int_{\mathbb{R}} f(y)g(x-y)\,dy = \int_{\mathbb{R}} g(y)f(x-y)\,dy.$$

Note that this definition of convolution only applies to functions f and g that are defined over all of $\mathbb{R}$. However, throughout the paper, we will only consider convolutions of functions $f : \Omega \to \mathbb{R}$ for some rectangle Ω in $\mathbb{R}^2$, and kernels $g : \mathbb{R}^2 \to \mathbb{R}$, either rapidly decreasing or with compact support. We therefore need to define what a convolution should look like in such cases.

Let Ω be a rectangle in $\mathbb{R}^2$, and let $u : \Omega \to \mathbb{R}$. We define the *symmetric extension* $u_e : \mathbb{R} \to \mathbb{R}$ of u by mirroring $u(\Omega)$ repeatedly across its borders. For a geometric depiction of this extension, see figure 2.1.

With this, we can define the convolution $u * g : \Omega \to \mathbb{R}$ by letting

$$\big(u * g\big)(x) := \big(u_e * g\big)(x). \tag{2.1}$$

With this, we are ready to define the minimization problems.

2.3 Minimizing problem definitions

Let Ω be a rectangle in $\mathbb{R}^2$ and $u_0 \in L^\infty(\Omega)$ be an image.

First, let $\mathcal{F} : W^{1,q}(\Omega) \to \mathbb{R}$ be defined as

$$\mathcal{F}(u) = \frac{1}{2}\|u - u_0\|_2^2 + \frac{\alpha}{2}\|\nabla u\|_2^2 + \beta\mathcal{R}(\nabla u), \tag{2.2}$$

where $\alpha, \beta > 0$, and where $\mathcal{R} : L^q(\Omega, \mathbb{R}^2) \to \mathbb{R}$ is defined as

$$\mathcal{R}(\xi) = \frac{1}{2p}\int_{\Omega\times\Omega} w(\|x - y\|_E)\left|\langle \xi(x), \xi(y)^\perp\rangle\right|^p dy\, dx, \tag{2.3}$$

for some bounded Borel measurable function $w : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ such that $w(0) > 0$, and $p \geq 1$. Here, $\xi^\perp$ denotes a vector in $\mathbb{R}^2$ that is orthogonal to ξ and of the same length. This is a unique definition up to signs, which in this functional do not matter due to the absolute value in the integral above. The function w serves as a definition of the non-locality of the functional.

Note that $\mathcal{F}(u)$ is finite whenever $u \in W^{1,q}$, $q \geq \max\{2, p\}$. This should be clear for the first two terms, and by the Cauchy-Schwarz inequality, $\left|\langle \nabla u(x), \nabla u(y)^\perp\rangle\right|^p \leq \|\nabla u(x)\|_E^p \|\nabla u(y)\|_E^p$, thus it also holds for $\mathcal{R}$. In section 2.5.2 we will show that $\mathcal{F}$ is Gâteaux differentiable when $q \geq \max\{2p, 2\}$.

Secondly, let $\mathcal{F}_W : W^{1,p}(\Omega) \to \mathbb{R}$ be defined by

$$\begin{aligned} \mathcal{F}_W(u) &= \int_\Omega f_W\big(x, u(x), A(u,x), \nabla u(x)\big), \\ f_W(x, v, B, \xi) &= \frac{1}{2}\big(v - u_0(x)\big)^2 + \frac{\alpha}{p}\|\langle \xi, B\xi\rangle\|_E^p. \end{aligned} \tag{2.4}$$

We define the operator $A(u, x) : \mathbb{R}^2 \to \mathbb{R}^2$. First, let

$$\begin{aligned} K_\sigma(x) &= \frac{1}{(2\pi\sigma^2)^{m/2}}\exp\left(-\frac{|x|^2}{2\sigma^2}\right) \\ u_\sigma &= \big(K_\sigma * u\big) \\ \mathcal{A}(u, x) &= \big(K_\rho * (\nabla u_\sigma \otimes \nabla u_\sigma)\big)(x). \end{aligned} \tag{2.5}$$

Here, and for the remainder of the paper, $\otimes$ denotes the $\mathbb{R}^2$ tensor product, that is, $\xi \otimes \zeta = \xi\zeta^T \in \mathbb{R}^{2\times 2}$. It should be clear that $\mathcal{A}(u, x) \in \mathbb{R}^{N\times N}$ is a symmetric positive semi-definite matrix. It can therefore be diagonalized as

$$\mathcal{A}(u, x) = (Q\Lambda Q^T)(x),$$

with Λ the diagonal matrix of decreasing eigenvalues of $\mathcal{A}(u, x)$, and Q the matrix of orthonormal eigenvectors corresponding to these eigenvalues. Since $\mathcal{A}(u, x)$ is positive semi-definite, we know that its eigenvalues are non-negative.

For a diagonal matrix Λ, define $h_\gamma(\Lambda)$ by

$$\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \to \begin{bmatrix} \frac{1}{1+\frac{(\lambda_1-\lambda_2)^2}{\gamma^2}} & 0 \\ 0 & 1 \end{bmatrix}.$$

Next, for diagonalizable matrices B, we define $g(B) = Qh_\gamma(\Lambda)Q^T$, where $B = Q\Lambda Q^T$. We then define the operator $A(u, x)$ by

$$A(u, x) = g(\mathcal{A}(u, x)). \tag{2.6}$$

In line with the introduced denoising approach, we now define the problems

$$\min_{u \in W^{1,2p}} \mathcal{F}(u), \tag{2.7}$$

and

$$\min_{u \in W^{1,p}} \mathcal{F}_W(u). \tag{2.8}$$

The solution to (2.7) and (2.8) will then be our denoised images. These problems are parameter dependant, the parameters being the numbers α, β, γ, σ, ρ and p, as well as the original data u_0.

In the above defined functionals, each term has its own purpose. The term $\frac{1}{2}\|u - u_0\|_2^2$ ensures the denoised image should not be too different from the original data. The term $\frac{\alpha}{2}\|\nabla u\|_2^2$ is a standard tool occuring in most regularization problems, as it grows large whenever u is irregular. In the field of image processing, however, large ∇u is often an indicator of edges, thus this simple term causes blurring that can ultimately remove important structures in the data. In a similar way, the term $\int_\Omega \frac{alpha}{p}\|A(u, x)\nabla u(x)\|_E^2 \, dx$ penalizes irregularities parallell to local dominating flow lines in u. Finally, the term $\frac{\beta}{p}\mathcal{R}(\nabla u)$ grows large whenever the directions of ∇u changes rapidly in areas weighted by the function w, and as a consequence, the term tries to remove corners and create flow lines.

2.4 Existence of minimizers by Tonelli's direct method

We now proceed to show the existence of minimizers of $\mathcal{F}_W$ and $\mathcal{F}$. In the variation of calculus, there is a well-established method of proving existence of minimizers of functions called the Tonelli direct method [Fonseca and Leoni (2007)]. Let U be a normed space, and let $\mathcal{F} : U \to [-\infty, \infty]$. The direct method provides conditions on $\mathcal{F}$ and U to ensure the existence of minimizers of $\mathcal{F}$. In short, the method roughly reduces to four steps, as seen in section 3.2 of Fonseca and Leoni (2007):

Step 1 Consider some minimizing sequence $\{u_n\} \in U$, i.e. a sequence such that

$$\mathcal{F}(u_n) \overset{n\to\infty}{\to} \inf_{u\in U} \mathcal{F}(u).$$

Step 2 Prove that $\{u_n\}$ admits a subsequence $\{u_{nk}\}$ converging with respect to some (possibly weaker) topology τ to some point $u_0 \in U$.

Step 3 Establish the sequential lower semi-continuity of $\mathcal{F}$ with respect to τ.

Step 4 Conclude that u_0 is a minimum of $\mathcal{F}$ because

$$\inf_{u\in U} \mathcal{F}(u) \leq \mathcal{F}(u_0) \leq \lim_{k\to\infty} \mathcal{F}(u_{nk}) = \lim_{n\to\infty} \mathcal{F}(u_n) = \inf_{u\in U} \mathcal{F}(u)$$

The two inequalities in step 4 follow from the definition of the infimum and the sequential lower semi-continuity of $\mathcal{F}$, while the two equalities follow from step 2 and step 1, respectively. Thus proving the existence of minimizers of (2.7) and (2.8) amounts to applying Tonelli's direct method to $\mathcal{F}$ of (2.2) and to $\mathcal{F}_W$ of (2.4).

Definition 1 (Sequentially lower semi-continuous)**.** Let U be a topological space. A function $\mathscr{J} : U \to \mathbb{R} \cup \{\infty\}$ is called sequentially lower semi-continuous(abbreviated slsc) if for every sequence $\{u_k\}_{k\in\mathbb{N}} \in U$ converging to $u \in U$ we have that

$$\liminf_{k\to\infty} \mathscr{J}(u_k) \geq \mathscr{J}(u). \tag{2.9}$$

Definition 2 (Coercive)**.** Let $(U, \|\cdot\|_U)$ be a normed space. A function $\mathscr{J} : U \to \mathbb{R} \cup \{\infty\}$ is called coercive if

$$\lim_{k\to\infty} \mathscr{J}(u_k) = \infty \tag{2.10}$$

for all sequences $\{u_k\}$ such that $\lim_{k\to\infty} \|u_k\|_U = \infty$.

Sequential lower semi-continuity of $\mathcal{F}$

The goal of this section is to show that $\mathcal{F}$ is slsc. Let $\mathcal{F}$ and $\mathcal{R}$ be as defined in (2.2) and (2.3), respectively. We note that both $\|u - u_0\|^2_{W^{1,q}}$ and $\|\nabla u\|^2_{W^{1,q}}$ are continuous and convex on $W^{1,q}$. By lemma 10.6 of Scherzer et al. (2009) they are consequently also weakly slsc(abbreviated wslsc) on $W^{1,q}(\Omega)$, that is, they are slsc with respect to the weak convergence $u_k \rightharpoonup u$ on $W^{1,q}$. Furthermore, for $q \geq 2$ and bounded Ω, we have that $W^{1,q}$ is embedded in $W^{1,2}$, and it follows that both $\|u - u_0\|^2$ and $\|\nabla u\|^2$ are wslsc on $W^{1,2}(\Omega)$.

If we can prove that $\mathcal{R}(\nabla u) = \mathcal{R} \circ \nabla(u)$ is also wslsc on $W^{1,q}(\Omega)$, $q \geq \max\{p, 2\}$, then also $\mathcal{F}$ must be wslsc on $W^{1,q}(\Omega)$.

We note that $\nabla : W^{1,q} \to L^q$, the gradient operator, is a bounded linear operator, and consequently the mapping $u \to \nabla u$ is weakly continuous. Therefore, the mapping $\mathcal{R} \circ \nabla : W^{1,q} \to \mathbb{R}$ is wslsc if the mapping $\mathcal{R} : L^q \to \mathbb{R}$ is wslsc. We will show that $\mathcal{R}$ is wslsc using the results of prop. 1. Before we show this, however, we must introduce some terminology.

Definition 3. Carathéodory function(Kubińska (2004/05))

Let X,Y be topological spaces and $(\mathcal{T}, \mathcal{M}, \mu)$ be a measureable space. We say that $f : \mathcal{T} \times X \to Y$ is a *Carathéodory function* if

- $f(\cdot, x)$ is measurable for all $x \in X$.
- $f(t, \cdot)$ is continuous for all $t \in \mathcal{T}$.

Definition 4. Non-local functional(Boulanger et al. (2011)) Let $n \in \mathbb{N}$ and $1 \leq q < \infty$. We call a mapping $\mathscr{J} : L^q(\Omega; \mathbb{R}^n) \to \mathbb{R} \cup \{+\infty\}$, which is not constantly equal to infinity, a non-local functional on $L^q(\Omega; \mathbb{R}^n)$ if there exists a function $f : \Omega \times \Omega \times \mathbb{R}^n \times \mathbb{R}^n$ such that

$$\mathscr{J}(\xi) = \int_\Omega \int_\Omega f(x, y, \xi(x), \xi(y)) dx\, dy \text{ for all } \xi \in L^q(\Omega; \mathbb{R}^n),$$

and such that

- f is Carathéodory,
- f has the symmetry property $f(x, y, s, t) = f(y, x, t, s)$ for all $x, y \in \Omega$ and all $s, t \in \mathbb{R}^n$,
- f is bounded below.

We say such a functional is defined by f and we denote it $\mathscr{J}_f^q(u)$.

From the above definition, we see that the functional $\mathcal{R}(\xi)$ is non-local and defined by

$$r(x, y, s, t) = w(|x - y|) \left| \langle s, t^\perp \rangle \right|^p. \tag{2.11}$$

Definition 5. Let $f : \Omega \times \Omega \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$. For fixed $x, y \in \Omega$, we define the mapping $f_{(x,y)} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ by

$$f_{(x,y)}(s, t) = f(x, y, s, t).$$

Proposition 1 (Proposition 8.8 of Boulanger et al. (2011))**.** Let $\mathscr{J}_f^q : L^q(\Omega; \mathbb{R}^n) \to \mathbb{R} \cup \{+\infty\}$, $1 \leq q < \infty$, be a non-local functional. If there exist $C \in \mathbb{R}$, $\gamma \in L^1(\Omega \times \Omega)$ and $\lambda \in L^1(\Omega)$ such that

$$|f(x, y, s, t)| \leq \gamma(x, y) + \lambda(x)|t|^q + \lambda(y)|s|^q + C|s|^q|t|^q \tag{2.12}$$

for a.e. $(x, y) \in \Omega \times \Omega$ and all $s, t \in \mathbb{R}^n$, then $\mathscr{J}_f^q$ is sequentially lower semi-continuous with respect to the weak topology on $L^q(\Omega; \mathbb{R}^n)$.

We are now ready to show the following proposition:

Proposition 2. $\mathcal{R}$ is wslsc

The functional $\mathcal{R} : W^{1,q} \to \mathbb{R}$ defined in (2.3) is wslsc.

Proof. Since we know that w is bounded on Ω, we can define $W := \sup_{x,y\in\Omega} w(|x-y|) < \infty$. It follows that

$$\begin{aligned} |r(x, y, s, t)| &= w(|x - y|) \left|\left\langle s, t^{\perp}\right\rangle\right|^p \leq W|s|^p|t|^p \\ &\leq W(1 + |s|^q)(1 + |t|^q) = W\left(1 + |t|^q + |s|^q + |t|^q|s|^q\right), \end{aligned} \tag{2.13}$$

for $q \geq \max\{2, p\}$. That is, r satisfies condition (2.12) with $\gamma = \lambda = C = W$, and seeing as Ω is bounded, we know that $\gamma \in L^1(\Omega)$and $\lambda \in L^1(\Omega \times \Omega)$. We therefore know that $\mathcal{R}$ is sequentially lower semi-continuous with respect to the weak topology on $L^q(\Omega; \mathbb{R}^n)$. □

By proposition 2, $\mathcal{R}$ is wslsc, and following the arguments at the beginning of this section, we conclude that $\mathcal{F}$ is indeed wslsc.

Coercivity of $\mathcal{F}$

We show coercivity of $\mathcal{F}$ as defined in (2.2).

From its definition, we know $\mathcal{R}(\xi) \geq 0$. Hence,

$$\mathcal{F}(u) \geq \frac{1}{2}\|u - u_0\|_2^2 + \frac{\alpha}{2}\|\nabla u\|_2^2.$$

Further, we know that

$$\begin{aligned} \|u - u_0\|_2^2 \geq \left(\|u\|_2 - \|u_0\|_2\right)^2 &= \|u\|_2^2 - 2\|u\|_2\|u_0\|_2 + \|u_0\|_2^2 \\ &= \|u\|_2\left(\|u\|_2 - 2\|u_0\|_2\right) + C, \end{aligned}$$

with $C = \|u_0\|_2^2 \geq 0$. Denoting $\|u\|_2 - 2\|u_0\|_2 =: M(u)$, we see that for all u such that $\|u\|_2 \geq 3\|u_0\|_2$, $M(u) \geq \|u\|_2$. We then get that $\|u - u_0\|_2^2 \geq \|u\|_2 M(u)$ is coercive in L^2. In other words, for sufficiently large u there exists C_1 such that $\|u - u_0\|_2 \geq \|u\|_2$. It follows that

$$\mathcal{F}(u) \geq \frac{C_1}{2}\|u\|_2^2 + \frac{C}{2} + \frac{\alpha}{2}\|\nabla u\|_2^2 \geq C_0\|u\|_{W^{1,2}}^2 + C_2 \geq C_0\|u\|_{W^{1,2}}^2,$$

where $C_0 = \frac{1}{2}\min\{C_1, \alpha\}$, $C_2 = \frac{C}{2}$. As $\mathcal{F} : W^{1,q} \to \mathbb{R}$, we get that it is coercive when $q = 2$.

As we have proven that $\mathcal{F}$ is coercive in $W^{1,2}$ and sequentially lower semi-continuous on $W^{1,q}$ for any $q \geq 1$, it follows from proposition 1 that the problem (2.7) has a solution $\bar{u} \in W^{1,2}$. □

Sequential lower semi-continuity of $\mathcal{F}_W(u)$

In this section, we wish to show that the functional $\mathcal{F}_W(u)$ is slsc with respect to the weak topology on $L^p(\Omega)$. We will do this by applying theorem 7.5 of Fonseca and Leoni (2007) to the functional. Before we do this, however, we need to introduce the notion a *normal integrand* and a few results about them. We define a normal integrand as in Fonseca and Leoni (2007):

Definition 6. Normal integrand(Fonseca and Leoni (2007))

Let $E \subset \mathbb{R}^N$ be a Lebesgue measurable set and let $B \subset \mathbb{R}^m$ be a Borel set. A function $f : E \times B \to [-\infty, \infty]$ is said to be a *normal integrand* if:

- For $\mathcal{L}^N$-a.e. $x \in E$ the function $f(x, \cdot)$ is lsc on B.
- There exists a Borel function $g : E \times B \to [-\infty, \infty]$ such that

$$f(x, \cdot) = g(x, \cdot)$$

for $\mathcal{L}^N$-a.e. $x \in E$.

The following results are found in Fonseca and Leoni (2007), and are stated without proof.

Theorem 1. (6.31 of Fonseca and Leoni (2007))

Let $\Omega \subset \mathbb{R}^N$ be Lebesgue measurable. Let $f : \Omega \times \mathbb{R}^m \to [-\infty, \infty]$ be lsc for $\mathcal{L}^N$-a.e. $x \in \Omega$. Then f is a normal integrand if and only if f is $\mathcal{L}^N \times B$ measurable.

Theorem 2. (6.34 of Fonseca and Leoni (2007))

Let $\Omega \subset \mathbb{R}^N$ be Lebesgue measurable, and let $B \subset \mathbb{R}^m$ be a Borel set. If $f : \Omega \times B \to [-\infty, \infty]$ is a Carathéodory function, then f is a normal integrand.

Theorem 3. (7.5 of Fonseca and Leoni (2007))

Let $\Omega \subset \mathbb{R}^N$ be a Lebesgue measurable set of finite measure, $1 \leq p$, $q < \infty$, and let $f : \Omega \times \mathbb{R}^d \times \mathbb{R}^m \to (-\infty, \infty]$ be an $\mathcal{L}^N \times \mathcal{B}$ measurable function, such that

$$f(x, z, \xi) \geq 0$$

for $\mathcal{L}^N$-a.e. $x \in \Omega$ and all $(z, \xi) \in \mathbb{R}^d \times \mathbb{R}^m$. Assume $f(x, \cdot, \cdot)$ is lsc in $\mathbb{R}^d \times \mathbb{R}^m$ for $\mathcal{L}^N$-a.e. $x \in \Omega$. Then the functional

$$(z,\xi) \in L^q(\Omega;\mathbb{R}^d) \times L^p(\Omega;\mathbb{R}^m) \to \int_\Omega f(x,z(x),\xi(x))dx$$

is slsc with respect to strong convergence in $L^q(\Omega;\mathbb{R}^d)$ and to weak convergence in $L^p(\Omega;\mathbb{R}^m)$, if and only if(up to equivalent integrands) $f(x,z,\cdot)$ is convex in $\mathbb{R}^m$ for $\mathcal{L}^N$ a.e. $x \in \Omega$ and all $z \in \mathbb{R}^d$.

Theorem 3 holds true even for integrands that allow negative values(on sets of non-zero $\mathcal{L}^N$ measure), however the conditions listed above become more complicated, and considering such functions is unneccessary for our purposes since the integrand of $\mathcal{F}_W$ is non-negative anyway.

Proving that the functional $\mathcal{F}_W$ is slsc amounts to showing that theorem 3 is applicable to it. This requires us to verify that all assumptions are valid and that the conditions listed above are satisfied for the integrand f_W. In other words, for $B = (\mathbb{R} \times \mathbb{R}^{2\times 2}) \times \mathbb{R}^2$ and $\Omega \subset \mathbb{R}^2$ a rectangle, we need to prove the following:

- f_W is $\mathcal{L}^2 \times \mathcal{B}$ measurable.
- $f_W(x,z,\xi) \geq 0$ for $\mathcal{L}^2$-a.e. $x \in \Omega$ and all $(z,\xi) \in B$.
- $f_W(x,\cdot,\cdot)$ is lsc in B for $\mathcal{L}^2$-a.e. $x \in \Omega$.
- $f_W(x,z,\cdot)$ is convex for $\mathcal{L}^2$-a.e. $x \in \Omega$ and all $z \in \mathbb{R} \times \mathbb{R}^{2\times 2}$.

We verify the list top-to-bottom.

i) f is $\mathcal{L}^2 \times \mathcal{B}$ measurable

Following the results of theorem 1, we can conclude that f_W is $\mathcal{L}^N \times \mathcal{B}$ measurable by showing that f_W is a normal integrand. We show normality of f_W by showing that f_W is Carathéodory and applying the results of theorem 2.

For fixed $((u,A),\xi) \in B$, $f(\cdot,(u,A),\xi) = \frac{1}{2}(u-u_0(\cdot))^2 + \frac{\alpha}{p}|A\xi|^p$ is measurable whenever $u_0 \in L^2(\Omega)$. We have assumed our data $u_0 \in L^\infty(\Omega) \subset L^2(\Omega)$, thus $f(\cdot,(u,A),\xi)$ is indeed measurable.

We now fix $x \in \Omega$. Define

$$g_p : \mathbb{R}^2 \to \mathbb{R} \text{ by } t \to |t|^p,$$

$$h : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^2 \text{ by } (s,t) \to s-t, \text{ and}$$

$$i : \mathbb{R}^{2\times 2} \times \mathbb{R}^2 \to \mathbb{R}^2 \text{ by } (A,\xi) \to A\xi.$$

Observe that these functions are all continuous. Further, note that $f_W(x,(u,A),\xi) = \frac{1}{2}g_2 \circ h(u,u_0(x)) + \frac{\alpha}{p}g_p \circ i(A,\xi)$, i.e. $f(x,\cdot,\cdot)$ is a composition of continuous functions. It follows that $f_W(x,\cdot,\cdot)$ is itself a continuous function.

Since the set B is a cartesian product of real vector spaces, it is a Borel set. Therefore f is by definition a Carathéodory function. Theorem 2 then tells us that f is normal, and consequently, theorem 1 tells us that f is $\mathcal{L}^N \times \mathcal{B}$ measurable.

ii) $f_W(x, z, \xi) \geq 0$

Obvious.

iii) $f_W(x, \cdot, \cdot)$ **is lsc for** $\mathcal{L}^2$**-a.e.** $x \in \Omega$

In step i) we proved that f_W is Carathéodory, hence $f_W(x, \cdot, \cdot)$ is continuous, and in conclusion f_W is lsc.

iv) $f_W(x, z, \cdot)$ **is convex for** $\mathcal{L}^2$**-a.e.** $x \in \Omega$ **and all** $z \in \mathbb{R}^2 \times (\mathbb{R}^2 \times \mathbb{R}^2)$

Let $x \in \Omega$ and $z = (u, A) \in \mathbb{R}^2 \times (\mathbb{R}^2 \times \mathbb{R}^2)$ be fixed. Recall from i) that $f(x, (u, A), \xi) = \frac{1}{2} g_2 \circ h(u, u_0(x)) + \frac{\alpha}{p} g_p \circ i(A, \xi)$. Clearly convexity of f in ξ is only dependant on $g_p \circ i(A, \cdot)$, hence $f(x, z, \cdot)$ is convex if and only if $g_p \circ i(A, \cdot)$ is convex.

Note that $i(A, \cdot) : \mathbb{R}^2 \to \mathbb{R}^2$ is simply a linear transformation, and in particular, it is convex. We calculate the Hessian of g_p:

$$
\begin{aligned}
g_p(\mathbf{x}) &= \|\mathbf{x}\|_E^p = (x_1^2 + x_2^2)^{p/2}, \\
\nabla g_p &= p\|\mathbf{x}\| E^{p-2}\mathbf{x}, \\
\nabla^2 g_p &= p\|\mathbf{x}\|_E^{p-2}\mathcal{I}\delta + p(p-2)\|\mathbf{x}\|_E^{p-4}\mathbf{x}\mathbf{x}^T \text{ for } \mathbb{X} \neq 0.
\end{aligned}
$$

We see that $\nabla^2 g_p$ is positive definite, hence g_p is (strictly) convex. It follows from these facts that $g_p \circ i(A, \cdot)$ is convex, and consequently, so is $f_W(x, (u, A), \cdot)$.

As f_W and Ω fulfills conditions i)-iv), theorem 3 then tells us that $\mathcal{F}_W(u)$ is in fact slsc.□

Coercivity of $\mathcal{F}_W$

Next, we look into coercivity of $\mathcal{F}_W$. Recall from (2.6) that

$$A(u, x) = g(\mathcal{A}(u, x)) = Q h_\gamma(\Sigma) Q^T.$$

For any $x \in \Omega$, we have that

$$\|A(u, x)\xi\|_E^2 \geq \left|\min\{\text{eig}(A(u, x))\}\right| \cdot \|\xi\|_E^2 = \frac{1}{1 + \left(\frac{\lambda_1 - \lambda_2}{\gamma}\right)^2} \|\xi\|_E^2,$$

where λ_i are the eigenvalues of $\mathcal{A}(u,x)$, defined in section 2.3. To find a lower bound of $\mathrm{eig}(A(u,x))$ for any $x \in \Omega$, we will need to find an upper bound on $\lambda_1 - \lambda_2$ over Ω. Since $A(u,x)$ is a symmetric positive semi-definite matrix by construction, we know that $\big|\lambda_i\big| \geq 0$, hence it is bounded from below. Seeking an upper bound, we know that

$$\big|\max_i \lambda_i\big| = \|A(u,x)\|_2 \leq \|A(u,x)\|_F$$

where $\|\cdot\|_2$ and $\|\cdot\|_F$ the 2-norm and Frobenius norm, respectively. Further, we know that

$$\|A(u,x)\|_F \leq 4 \max_{i,j} \|K_\rho * (\partial_i u_\sigma \cdot \partial_j u_\sigma)\|_\infty \leq 4 \max_{y\in\Omega} K_\rho(y) \max_{i,j} \|\partial_i u_\sigma \partial_j u_\sigma\|_1.$$

Since K_ρ is bounded, we know $\max_{y\in\Omega} K_\rho(y)$ exists. Denote $C := \max_{y\in\Omega} K_\rho(y)$. By the Cauchy-Schwarz inequality, we now get

$$\|A(u,x)\|_F \leq 4c \max_{i,j} \|\partial_i u_\sigma\|_2 \|\partial_j u_\sigma\|_2.$$

Furthermore, we have that

$$\partial_i u_\sigma = \partial_i (K_\sigma * u) = (\partial_i K_\sigma) * u,$$

hence we find that

$$\|A(u,x)\|_F \leq 4c \max_{i,j} \|u * \partial_i K_\sigma\|_2 \|u * \partial_j K_\sigma\|_2 \leq \tilde{c} \|u\|_{L^2}^2.$$

Here, $\tilde{c} = 4c\big(\max_{y\in\Omega} K_\sigma(y)\big)^2$. Note that $\tilde{c}$ only exists when K_σ is a bounded function. Given this, we now have

$$\big|\max_i \lambda_i\big| \leq \tilde{c} \|u\|_{L^2}^2,$$

thus providing the needed upper bound. Using the non-negativity of $\min_i \lambda_i$, we then bound the difference

$$\big|\lambda_1 - \lambda_2\big| \leq \big|\max_i \lambda_i\big| = \tilde{c} \|u\|_{L^2}^2.$$

We now return to finding $\min\{\mathrm{eig}(A(u,x))\}$. We have

$$\begin{aligned}\min\{\mathrm{eig}(A(u,x))\} &\geq \inf_{x\in\Omega} \min\{\mathrm{eig}(A(u,x))\} \\ &= \inf_{x\in\Omega} \left\{ \frac{1}{1+\left(\frac{\lambda_1(x)-\lambda_2(x)}{\gamma}\right)^2} \right\} \\ &\geq \frac{1}{1+\left(\frac{\max\lambda_i - \min\lambda_i}{\gamma}\right)^2} \\ &= \frac{1}{1+\tilde{c}\left(\frac{\|u\|_{L^2}^2}{\gamma}\right)^2} =: h(\|u\|_{L^2}).\end{aligned}$$

Notice that $h : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, $h(s) > 0$ for all $s < \infty$ and $h(s) \leq 1$ for all s. This means, in the end, that

$$\|A(u,x)\xi\|_E \geq h(\|u\|_{L^2})\|\xi\|_E > 0 \text{ for } \xi \neq 0.$$

Now, let $u \in W^{1,p}(\Omega)$, and let $\|u\|_{1,p} \to \infty$. Then, either $\|u\|_{L^2} \to \infty$, or $\|u\|_{L^2} \leq C < \infty$ and $\|\nabla u\|_{L^p} \to \infty$. If $\|u\|_{L^2} \to \infty$, we have

$$\|u\|_{1,p} \to \infty \Rightarrow \|u - u_0\|_{L_2} \to \infty,$$

and consequently $\mathcal{F}_W \to \infty$. If $\|\nabla u\|_{L^p} \to \infty$, then we have

$$\int_\Omega \|A(u,x)\nabla u(x)\|_E^p \geq h(\|u\|_{L^2})\|\nabla u(x)\|_{L_p}\, dx \to \infty,$$

and thus $\mathcal{F}_W \to \infty$. Either way, as $\|u\|_{1,p} \to \infty$, we consequently know that $\mathcal{F}_W(u) \to \infty$, hence the functional is indeed coercive. □

Conclusions on existence of minimizers

In the previous section, we established sequential lower semi-continuity with respect to the weak topology on $W^{1,p}$ for both $\mathcal{F}_W$ and $\mathcal{F}$, coercivity in $W^{1,2}$ for $\mathcal{F}$ and coercivity in $W^{1,p}$ for $\mathcal{F}_W$. By the Tonelli direct method, we can then conclude with the existence of minimizers of $\mathcal{F}$ and $\mathcal{F}_W$ by creating any minimizing sequence u_n, for instance by the steepest descent method of Nocedal and Wright (2006).

2.5 Steepest Descent Algorithm

The steepest descent method, also known as the *gradient descent* method is a simple and logically straightforward optimization method. This method seeks to minimize a functional $\mathcal{G}$ by iteratively minimizing $\mathcal{G}$ along its negative gradient direction at iterates v_k – hence the name of the algorithm. The method yields the minimizing sequence

$$v_{k+1} = v_k - c_k \text{grad}\, \mathcal{G}(v_k), \tag{2.14}$$

where $0 < c_k < \infty$ is the *step length* of iteration k. This number has to be found by some *line search*, which will be further explained in section 2.5.1.

The gradient descent method is simple to implement and logically sound, but in practice it is often very slow, having only a linear convergence rate. It is convergent, but not necessarily to a minimizer of $\mathcal{G}$ – only to a critical point. Since neither $\mathcal{F}$ nor $\mathcal{F}_W$ of eqs. (2.7) and (2.8) are convex, a critical point might not necessarily be a minimum, global or local. This is problematic for our solution method, but for now we ignore this fact and move on.

2.5.1 Line search

The idea of a line search is to find a step length s that minimizes the functional $\mathcal{G}(v_k + sh_k)$ in some search direction h_k.

Most line searches are at a minimum required to fulfill an Armijo-Goldstein condition, which ensures the functional values decrease from one iterate to the next, but other conditions can also be applied [Nocedal and Wright (2006)].

A simple *backtracking* line search tries to find a suitable s_k such that

$$\mathcal{G}(v_{k+1}) = \mathcal{G}(v_k - s_k \text{grad}\, \mathcal{G}(v_k)) < \mathcal{G}(v_k) - \tau \langle s_k h_k, \text{grad}\, \mathcal{G}(v_k; h_k) \rangle$$

for some threshold $0 < \tau < 1$ and a search direction h_k. For a general functional, our algorithm will implement a backtracking line search as described in Nocedal and Wright (2006).

Note, however, that for any quadratic positive definite functional, it is possible to perform an exact line search. The idea is to consider v_{k+1} as a function of the step length s with the parameters v_k and $\nabla \mathcal{F}(v_k)$. Then one solves $\partial_s(v_{k+1}) = 0$. This could potentially reduce running times and increase accuracy of the method.

In the gradient descent method, the search direction $h_k = \text{grad}\, \mathcal{G}(v_k)$, where it is assumed that the formal gradient exists, i.e., the natural boundary conditions hold. For each functional $\mathcal{F}$ and $\mathcal{F}_W$, these conditions are specified in lemmas 2 and 3, respectively.

To create the minimizing sequence of the gradient descent method, we need gradients of our functionals. We devote this next section to find these.

2.5.2 Formal gradients

Let $\mathcal{F}$, $\mathcal{R}$, $\mathcal{F}_W$ and $\hat{\mathcal{F}}_W$ be as defined in (2.2), (2.3), (2.4) and (2.27), respectively. In this section, we seek the formal gradients of $\mathcal{F}$ and $\mathcal{F}_W$.

Formal gradient of $\mathcal{F}$

We first look for a directional derivative of $\mathcal{F}$. To this end we compute the limit

$$\lim_{\epsilon \to 0} \frac{\mathcal{F}(u + \epsilon h) - \mathcal{F}(u)}{\epsilon} =: \mathrm{D}\mathcal{F}(u, h). \tag{2.15}$$

Here, $h \in W^{1,q}(\Omega)$ indicates a direction in the space of admissible functions. If the limit exists, and it defines a bounded linear operator $D\mathcal{F}(u) : W^{1,q}(\Omega) \to \mathbb{R}$, we say that the functional $\mathcal{F}$ is Gâteaux differentiable, and that $D\mathcal{F}(u, h)$ is the directional derivative of $\mathcal{F}$ at u in direction h.

We note that $\mathcal{F}$ consists three terms, and if each of these terms is Gâteaux differentiable, then $\mathcal{F}$ is Gâteaux differentiable too. Let $\mathcal{F}_1(u) = \frac{1}{2}\|u - u_0\|_2^2$ and $\mathcal{F}_2(u) = \frac{\alpha}{2}\|\nabla u\|_2^2$, such that $\mathcal{F}(u) = \mathcal{F}_1(u) + \mathcal{F}_2(u) + \beta\mathcal{R}(\nabla u)$.

Lemma 1. Gâteaux differentiability of $\mathcal{R}$

The functional $\mathcal{R}$ defined in (2.3) is Gâteaux differentiable when $q \geq 2p$. Its directional derivative is

$$D\mathcal{R}(\xi; \tau) = \int_{\Omega \times \Omega} w(|x - y|) \left\langle \xi(x), \xi(y)^{\perp} \right\rangle | \left\langle \xi(x), \xi(y)^{\perp} \right\rangle |^{p-2} \left\langle \tau(x), \xi(y)^{\perp} \right\rangle dx\, dy, \tag{2.16}$$

Proof. See Appendix A.2.3. □

We state without proof that

$$D\mathcal{F}_1(u, h) = \langle u - u_0, h \rangle_{L^2(\Omega)},$$
$$D\mathcal{F}_2(u, h) = \alpha \langle \nabla u, \nabla h \rangle_{L^2(\Omega)}.$$

From this and the above lemma we can find that

$$D\mathcal{F}(u, h) = \langle u - u_0, h \rangle + \alpha \langle \nabla u, \nabla h \rangle + \beta D\mathcal{R}(\nabla u, \nabla h). \tag{2.17}$$

This is the general form of the directional derivative of $\mathcal{F}$ at u in direction h.

Lemma 2. Formal gradient of $\mathcal{F}$

Let $D\mathcal{F}(u;h)$ be as defined in (2.17). If u satisfies, for all $x \in \partial\Omega$, the natural boundary condition

$$\left\langle \nabla u(x) + 2\beta V(x)\nabla u(x), \mathbf{n}(x) \right\rangle = 0, \tag{2.18}$$

where $\mathbf{n}(x)$ is the normal vector on Ω at $x \in \partial\Omega$ and

$$V(x) = \left(\int_\Omega w(|x-y|) \left| \langle \nabla u(x), \nabla u(y)^\perp \rangle \right|^{p-2} \nabla u(y)^\perp \otimes \nabla u(y)^\perp dy \right),$$

then $D\mathcal{F}(u,h)$ can be represented as

$$D\mathcal{F}(u;h) = \langle \operatorname{grad} \mathcal{F}, h \rangle = \int_\Omega \operatorname{grad} \mathcal{F}(u,x) h(x) dx, \tag{2.19}$$

where

$$\begin{aligned} \operatorname{grad} \mathcal{F}(u,x) &= u(x) - u_0(x) + \alpha \Delta u(x) \\ &\quad + \int_\Omega \operatorname{div} \left(w(|x-y|) \Upsilon(\nabla u(x), \nabla u(y)^\perp) \nabla u(y)^\perp \right) dy \end{aligned} \tag{2.20}$$

is the formal gradient of $\mathcal{F}$. Here, div denotes the $\mathbb{R}^2$ divergence operator, and $\Upsilon(\xi,\zeta) = \langle \xi, \zeta \rangle |\langle \xi, \zeta \rangle|^{p-2}$.

When $p = 2$ the formal gradient simplifies significantly to

$$\begin{aligned} \operatorname{grad} \mathcal{F}(u,x) &= u(x) - u_0(x) + \alpha \Delta u(x) \\ &\quad + \operatorname{div} \Big(\big(w * (\nabla u^\perp \otimes \nabla u^\perp) \big)(x) \nabla u(x) \Big). \end{aligned} \tag{2.21}$$

Proof. See appendix A.3. □

Formal gradient of $\mathcal{F}_W$

Next, we seek the formal gradient of $\mathcal{F}_W$. We recall some notation from section 2.3. First, define the anisotropic structure tensor $A(u,x)$ by

$$\begin{aligned} A(u,x) &= g\Big(\big(K_\rho * (\nabla u_\sigma \otimes \nabla u_\sigma) \big)(x) \Big) = g \circ B(u,x), \\ B(u,x) &= \big(K_\rho * (\nabla u_\sigma \otimes \nabla u_\sigma) \big)(x) \end{aligned} \tag{2.22}$$

and let Q and Σ define the diagonalization of $B(u,x)$, i.e.

$$B(u,x) = Q\Sigma Q^T.$$

Consequently both Q and Σ are functions of u and x, however we will use the notation above throughout the paper. Since Q is a 2×2 matrix, its columns are orthonormal, thereby yielding the form

$$Q = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

where θ is the phase angle. This means that Q is simply a rotation matrix. We will in the following use the notation

$$\cos(\theta) = a,\ \sin(\theta) = b,$$

so that

$$Q = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}. \tag{2.23}$$

Again, remember that a and b are functions of u and x. Moving on, we now introduce the functions

$$\begin{aligned} T(u,x) &= K_\sigma^* * \Big[H_2 \nabla u_\sigma K_\rho^* * \Big(N(u,\cdot)(a^2-b^2) \Big) \\ &\qquad - 2H_3 \nabla u_\sigma K_\rho^* * \Big(N(u,\cdot)(ab) \Big) \Big](x), \\ f(u,x) &= A(u,x)\nabla u(x), \\ M(u,x) &= \Big(\|f(u,x)\|_E^2 \Big)^{p/2-1}, \end{aligned} \tag{2.24}$$

where

$$
\begin{aligned}
N(u,x) &= 2M(u,x)\beta(u,x)\big\langle H_1(u,x)\nabla u(x), f(u,x)\big\rangle,\\
\beta(u,x;\gamma) &= \frac{2\Big(\frac{\sigma_1-\sigma_2}{\gamma}\Big)}{\Big[1+\Big(\frac{\sigma_1-\sigma_2}{\gamma}\Big)^2\Big]^2},\\
H_1(u,x) &= Q\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}Q^T = \begin{bmatrix}a^2 & ab\\ ab & b^2\end{bmatrix},\\
H_2 &= \begin{bmatrix}-1 & 0\\ 0 & 1\end{bmatrix},\\
H_3 &= \begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix},
\end{aligned}
\tag{2.25}
$$

and the adjoints K_σ^* and K_ρ^* are defined by

$$K_i^*(x) = -K_i(-x),\ \ i \in \{\sigma,\rho\}.$$

Lemma 3. Formal gradient of $\mathcal{F}_W$.

With the notation introduced above, the formal gradient of $\mathcal{F}_W$ has the form

$$\operatorname{grad}\mathcal{F}_W(u,x) = u(x) - u_0(x) - \alpha\operatorname{div}\big(T(u,x) + M(u,x)A(u,x)^T f(u,x)\big) \tag{2.26}$$

where $A(u,x)^T$ is the transpose of $A(u,x)$.

Proof. See appendix A.4. □

With this we have all necessary theoretical tools to solve the minimising problems introduced in section 2.3.

Remark. An alternative definition of $\mathcal{F}_W$ is

$$\hat{\mathcal{F}}_W(u,v) = \int_\Omega f_W(x,u(x),A(v,x),\nabla u(x))\,dx. \tag{2.27}$$

Notice that $\mathcal{F}_W(u) = \hat{\mathcal{F}}_W(u,u)$. The main purpose of this particular definition is that $\hat{\mathcal{F}}_W$ is a quadratic positive definite functional in u when $p = 2$, consequently enabling exact line search. We will not go into great detail of this functional in this paper. The inexact u-gradient of $\hat{\mathcal{F}}_W$ and the corresponding boundary condition it imposes can then be found by simply removing the T-term from $\operatorname{grad}\mathcal{F}$, that is,

$$\operatorname{grad}_u\hat{\mathcal{F}}(u,v) = \int_\Omega u(x) - u_0(x) + \alpha\operatorname{div}\big(A(v,x)^T A(v,x)\nabla u(x)\big)\|A(v,x)\nabla u(x)\|_E^{p/2-1},$$

whenever

$$\|A(v,x)\nabla u(x)\|_E^{p/2-1}\langle \mathbf{n}(x), A(v,x)^T A(v,x)\nabla u(x)\rangle = 0 \text{ on } \partial\Omega.$$

With this it is possible to define a slightly different numerical scheme. Instead of using the search direction $\text{grad}\,\hat{\mathcal{F}}_W(v_k, v_k)$, we use only the non-tensor gradient $\text{grad}_u\,\hat{\mathcal{F}}(v_k, v_k)$. This sequence simply ignores the contribution of the structure tensor to the gradient. We then obtain the minimizing sequence

$$v_{k+1} = v_k - c_k \text{grad}_u\,\hat{\mathcal{F}}(v_k, v), \tag{2.28}$$

where the variable v is the image seen by the structure tensor A.

This implementation has been tested parallel to experiments of chapter 4, but depending on parameter choices, produced results either identical to that of $\mathcal{F}$, or were highly unstable, likely due to stability issues of the diagonalization of A (for details, see appendix A.1). We have therefore not included these experiments in the paper.

Chapter 3

Experiment

For the remainder of the paper, we fix $p = 2$ and the function $w : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ as a normalized Gaussian kernel

$$w(r) = C \exp\Big(\frac{-r^2}{2\rho^2}\Big),$$

where C is the normalizing constant. Notice that this definition of w coincides with the kernel K_ρ defined in (2.3).

3.1 Discretizations

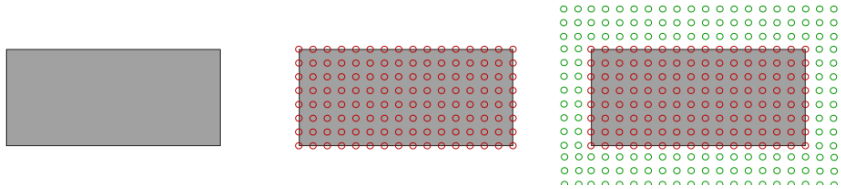
This section will define all discretizations used for implementing our numerical scheme solving (2.7) and (2.8). This requires us to approximate Ω, u, ∇ and $\int_\Omega dx$.

3.1.1 Discretization of Ω

Recall that that our data u_0 is an $M \times N$ regular grid of pixel values. The simplest and most natural choice for discretising Ω is therefore a 2D mesh whose points coincide with the pixels. This results in a set of points

$$\hat{\Omega} := \{\omega_{i,j}\}_{i,j=1,1}^{M,N} \subset \Omega \subset \mathbb{R}^2.$$

As we will need access to points outside Ω for computing convolutions, it is necessary to extend $\hat{\Omega}$ to $\mathbb{R}^2$. This extension is denoted by $\hat{\Omega}_e$, and is obtained recursively by adding



(a) The rectangle $\Omega \subset \mathbb{R}^2$.

(b) The discretization $\hat{\Omega}$ of Ω indicated by red circles as a covering set of points in $\bar{\Omega}$

(c) $\hat{\Omega}_e$ is the extension of $\hat{\Omega}$ as a covering all of $\mathbb{R}^2$

Figure 3.1: The extension $\hat{\Omega}_e$ of $\hat{\Omega}$ to $\mathbb{R}^2$.

mesh points in an appropriate distance and location away from the remainder of the grid. For a geometric depiction of the descretization $\hat{\Omega}$ and its extension $\hat{\Omega}_e$, see figure 3.1.

3.1.2 Discretization of integrals and convolution

We use a simple column approximation for integrals, that is, for any functional $F(f) = \int_\Omega f(x)dx$, with $f : \Omega \to \mathbb{R}$, we define

$$F(f) = \int_\Omega f(x)\,dx \approx \sum_{\omega\in\hat{\Omega}} f(\omega) = \sum_{\omega\in\hat{\Omega}} f(\omega) =: \hat{F}(f).$$

This approximation implicitly introduces a scaling of our problem such that the size of one pixel is exactly 1. Bearing this in mind, we can now discretize the convolution operation. Let $g : \mathbb{R}^2 \to \mathbb{R}$ have compact support, and let f_e be the symmetric extension of f, as defined in section 2.2. Then for any $\omega \in \hat{\Omega}$, we have

$$\big(f * g\big)(\omega) = \int_{\mathbb{R}} g(\omega) f_e(y-\omega)dy \approx \sum_{\alpha\in\hat{\Omega}_e} g(\omega) f_e(\alpha-\omega) =: \big[f * g\big](\omega). \tag{3.1}$$

We call $[f * g] : \hat{\Omega}_e \to \mathbb{R}$ the discrete convolution of f and g over $\hat{\Omega}$.

3.1.3 Discretization of ∇

We now look for an operator $G \approx \nabla$ on the grid $\hat{\Omega}_e$. More precisely, we look for operators $G_x \approx \partial_x$ and $G_y \approx \partial_y$, and define $G = [G_x, G_y]^T$. For simplicity, we let G_x and G_y be the *forward difference* operators. That is, for some function $f : \Omega \to \mathbb{R}$,

$$(G_x)_{i,j}\hat{f} = G_x\hat{f}_e(\omega_{i,j}) = \frac{1}{\delta}\Big(\hat{f}_e(\omega_{i+1,j}) - \hat{f}_e(\omega_{i,j})\Big),$$
$$(G_y)_{i,j}\hat{f} = G_y\hat{f}_e(\omega_{i,j}) = \frac{1}{\delta}\Big(\hat{f}_e(\omega_{i,j+1}) - \hat{f}_(\omega_{i,j})\Big),$$

where $\omega_{i,j} \in \hat{\Omega}$ and δ is the euclidean distance between neighbouring pixels. In line with the problem scaling implicitly introduced in section 3.1.2, we define $\delta = 1$, simplifying the above equation to

$$\begin{aligned}(G_x)_{i,j}\hat{f} &= \hat{f}_e(\omega_{i+1,j}) - \hat{f}_e(\omega_{i,j}),\\ (G_y)_{i,j}\hat{f} &= \hat{f}_e(\omega_{i,j+1}) - \hat{f}_e(\omega_{i,j}).\end{aligned} \tag{3.2}$$

By the above definition, calculating $G\hat{f}$ on the upper index boundaries of Ω, we find

$$\begin{aligned}(G_x)_{M,j}\hat{f} &= G_x\hat{f}_e(\omega_{M,j}) = \hat{f}(\omega_{M-1,j}) - \hat{f}(\omega_{M,j}),\\ (G_y)_{i,N}\hat{f} &= G_y\hat{f}_e(\omega_{i,N+1}) = \hat{f}(\omega_{i,N-1}) - \hat{f}(\omega_{i,N}).\end{aligned} \tag{3.3}$$

Notice that these are the negative backward difference operators.

It is possible to define G_x and G_y as *discrete convolutions* of $\hat{f}$ with some kernels $k_d : \hat{\Omega}_e \to \mathbb{R}$, $d \in \{x, y\}$ using the definition of discrete convolution in (3.1). In such notation, we can write

$$\begin{aligned}G_x\hat{f}_{i,j} &= [k_x * \hat{f}](\omega_{i,j}),\\ G_y\hat{f}_{i,j} &= [k_y * \hat{f}](\omega_{i,j}),\end{aligned} \tag{3.4}$$

where

$$k_x(r) = \begin{cases}-1, & \text{if } r = [0,0]^T,\\ 1, & \text{if } r = [1,0]^T,\\ 0, & \text{otherwise.}\end{cases} \quad k_y(r) = \begin{cases}-1, & \text{if } r = [0,0]^T,\\ 1, & \text{if } r = [0,1]^T,\\ 0, & \text{otherwise.}\end{cases} \tag{3.5}$$

In conclusion, we define

$$G\hat{f} = [G_x\hat{f}, G_y\hat{f}]^T.$$

Discretization of div

Directly consequential from the definition of G, we can now approximate the $\mathbb{R}^2$ divergence operator. Let $\hat{f}$ approximate $f : \Omega \to \mathbb{R}$ on the grid $\hat{\Omega}$. We then approximate

$$\operatorname{div} f = \langle \nabla^*, f \rangle^2_{\mathbb{R}} \approx \langle G^*, \hat{f} \rangle^2_{\mathbb{R}} = \left(G^*_x + G^*_y\right) \hat{f} =: \operatorname{div} \hat{f}, \tag{3.6}$$

with G^*_x and G^*_y being the adjoint operators of G_x and G_y, respectively, and $G^* = [G^*_x, G^*_y]^T$. Therefore finding the divergence operator requires finding the adjoint gradient operator. With G_x and G_y as defined by (3.4), we now seek the adjoint operators G^*_x and G^*_y relative to the $L_2(\hat{\Omega})$ inner product, which we define by approximation of the $L^2(\Omega)$ inner product:

$$\langle u, v \rangle_{L^2(\Omega)} = \int_\Omega u(x)v(x)dx \approx \sum_{\omega \in \hat{\Omega}} u(\omega)v(\omega) =: \langle u, v \rangle_{L^2(\hat{\Omega})}. \tag{3.7}$$

In this notation, we want to find G^*_x and G^*_y such that

$$\langle G_x u, v \rangle_{L^2(\hat{\Omega})} = \langle u, G^*_x v \rangle_{L^2(\hat{\Omega})}, \tag{3.8}$$

$$\langle G_y u, v \rangle_{L^2(\hat{\Omega})} = \left\langle u, G^*_y v \right\rangle_{L^2(\hat{\Omega})}. \tag{3.9}$$

We first look for G^*_x. From (3.4), (3.1) and (3.9), we find that

$$\begin{aligned} \langle G_x u, v \rangle_{L^2(\hat{\Omega})} &= \sum_{\omega \in \hat{\Omega}} [k_x * u](\omega)v(\omega) \\ &= \sum_{\alpha \in \hat{\Omega}} \sum_{\omega \in \hat{\Omega}} k_x(\omega - \alpha)u(\alpha)v(\omega). \end{aligned} \tag{3.10}$$

Letting $k^*_x(x) = k_x(-x)$,

$$\begin{aligned} \langle G_x u, v \rangle_{L^2(\Pi)} &= \sum_{\omega \in \hat{\Omega}} \sum_{\alpha \in \hat{\Omega}} k^*_x(\alpha - \omega)u(\alpha)v(\omega), \\ &= \sum_{\alpha \in \hat{\Omega}} \sum_{\omega \in \hat{\Omega}} u(\alpha)k^*_x(\alpha - \omega)v(\omega), \\ &= \sum_{\alpha \in \hat{\Omega}} u(\alpha)[k^*_x * v](\alpha), \\ &\langle u, G^*_x v \rangle, \end{aligned} \tag{3.11}$$

thus

$$G_x^* = [k_x^* * \cdot]. \tag{3.12}$$

An almost identical approach is used to find the adjoint operator G_y^* of G_y, yielding the result

$$G_y^* = (k_y^* * \cdot). \tag{3.13}$$

This concludes the search for G^* of (3.6). With this, we define

$$\operatorname{div} \hat{f} := \langle G^*, \hat{f} \rangle \tag{3.14}$$

Discretization of Δ

The discrete Laplacian operator is further derived in a similar way. We recall that $\Delta = \nabla^* \cdot \nabla = \operatorname{div} \nabla$. Applying $\nabla \approx G$, we find that $\Delta \approx G^*G = G_x^*G_x + G_y^*G_y := L$. This explicitly gives

$$G_x^*G_x = \big[k_x^* * [k_x * \cdot]\big], \qquad G_y^*G_y = \big[k_y^* * [k_y * \cdot]\big]$$

By extending our definition of $[f * g]$ to $\hat{\Omega}_e$, one can define

$$\big[k_x^* * [k_x * \cdot]\big] = [K_x * \cdot], \qquad \big[k_y^* * [k_y * \cdot]\big] = [K_y * \cdot]$$

with $[k_x^* * k_x] =: K_x$ and $[k_y^* * k_y] =: K_y$. In this notation, we conclude that

$$\Delta \approx L = G^*G = [K_x * \cdot] + [K_y * \cdot] = \big[(K_x + K_y) * \cdot\big] \tag{3.15}$$

by linearity of the discrete convolution.

3.1.4 Discretization of functions on Ω

Next, we look at how we discretize functions defined on Ω. Let $f : \Omega \to \mathbb{R}$. For any such function, we denote its discretization $\hat{f} : \hat{\Omega} \to \mathbb{R}$ by letting

$$\hat{f} = \{f(\omega)\}_{\omega \in \hat{\Omega}}.$$

Note that this discretization structures $\hat{f}$ in the same way as $\hat{\Omega}$. That means $\hat{f}$ is an $M \times N$ array of function values.

With the discretizations introduced in sections 3.1.1-3.1.4, we are fully equipped to create our denoisers.

Chapter 4

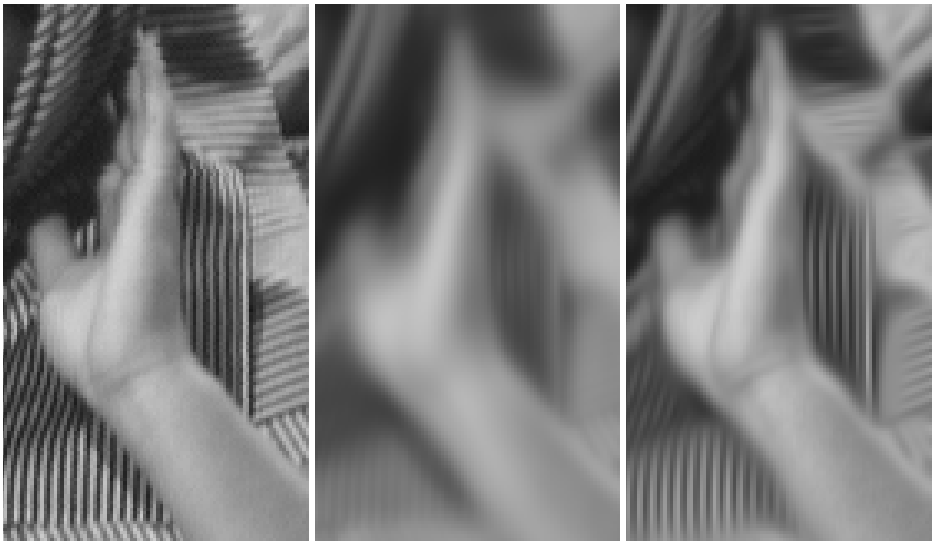
Analysis

4.1 Parameter effects

Before we go in detail into the analysis, it is worth our time to recap what behaviour we expect from our functionals, and how our parameter choices affect this behaviour.

For the functional $\mathcal{F}$, we must choose parameters α, β and the standard deviation ρ of the normalized gaussian kernel w. The α parameter is included in the functional mostly to guarantee coercivity, and is therefore mainly of theoretical interest. The effect of the α term is isotropic denoising, and because we wish our denoiser to enhance anisotropy, we wish to have as little isotropic denoising as possible. Hence we will opt to choose small values of α. The β and ρ parameters are both key components of the non-local anisotropy enhancer. Qualitatively, ρ describes the size of the structures we wish to enhance, or in other words, the size of areas in which anisotropies in the image are roughly the same. We should therefore expect better results when ρ coincides with the size of structures we wish to enhance. In our case, ρ translates directly to pixel distances, that is, for all grid points, mostly pixels no further than ρ away from the grid point are considered. Lastly, the value of β should reflect the degree of anisotropy enhancement we would like. In general, large values of β yield good denoising results. We note, however, that it is important to carefully inspect the data u_0 before choosing any parameters.

For the other functional, $\mathcal{F}_W$, it is required to define the parameters α, σ, ρ and γ. The parameter ρ functions largely the same way as it did for $\mathcal{F}$, and the values of it will be chosen by the same means. As stated in Weickert (1999), the parameter σ reduces local noise, and makes the structure tensor ignore changes less than $\mathcal{O}(\sigma)$. Qualitatively, σ is interpreted as a noise scale in the data. Note also that it has similar properties as ρ in the sense that it will consider pixels in a σ-sized neighbourhood of any grid point, hence to reduce isotropic blurring, σ should not be chosen too large. The best way to describe γ is that it acts as an "inverse" tolerance of anisotropy. That is, small γ allows a higher degree



(a) Original data **(b)** $\beta = 1$, $\rho = 20$ **(c)** $\beta = 1$, $\rho = 3$.

Figure 4.1: The figure shows how the choices of ρ affect the minimization results of $\mathcal{F}$. Here $\alpha = 0.001$ and $\beta = 1$, and ρ are chosen as seen in the respective figure texts. Image size: 92×172.

of anisotropic enhancement. Lastly, α here is a value reflecting the amount of anisotropic enhancement we seek, in many ways similar to what β was in $\mathcal{F}$.

Due to our construction of $\hat{\Omega}$, both σ and ρ will correspond to unit lengths in amounts of pixels. In other words, a value of $\sigma = 1$ means that around any pixel ω in $\hat{\Omega}$, we take into account mainly values in pixels with distance not much further than 1 pixel away from ω.

As an added experimental parameter, we also fix the amount of iterations the steepest descent method performs before terminating the program. We do this because our functionals $\mathcal{F}$ and $\mathcal{F}_W$ are not convex, and as such there are no simple way of determining whether or not the limit the program approaches is in fact a minimum. We do expect a higher number of iterates to yield more "correct" solutions, as seen by our functionals.

In the in the remainder of this section, the images presented are chosen such that the features of each functional as a denoiser is presented clearly, and their effects may differ from image to image even for the same parameter choices. Also, for each parameter choice, we interrupt the denoiser after exactly 100 iterations unless otherwise specified.

4.1.1 Parameter effects on $\mathcal{F}$

Figures 4.1 show how the parameter ρ affects the results, with $\alpha = 0.001$ and $\beta = 1$. With the stripes in figure 4.1a being very slim, it follows that enhancing them would require ρ to be sufficiently small. Comparing figures 4.1b and 4.1c, the improvement from considering

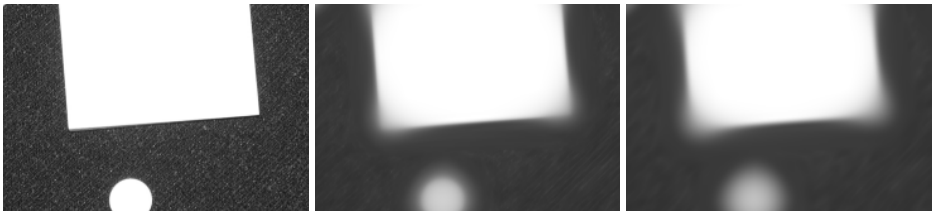


Figure 4.2: Example of effects of the β parameter, with $\alpha = 0.001$ and $\rho = 20$. The original image on the left, and results with $\beta = 1$ in the middle and $\beta = 100$ on the right. Image size: 242×164.

rather large anisotropy areas with $\rho = 20$ to considering smaller areas with $\rho = 3$ are drastic.

In figure 4.2, we present the functionality of the β parameter. With changes in β, the differences between results after 100 iterations were minuscule, and therefore we have in this particular figure performed 1000 iterations. The figure shows how blurring occurs near the corners of the white sheet, which are areas in which the β term should be large. Similar behaviour is seen near the curved string in the top left corner of figure 4.2. These effects demonstrate that the functional tries to draw straight lines. While the blurring of edges occur throughout the entire result images in figure 4.2, it is clear that its effect are greatest near corners, which makes perfect sense from a mathematical point of view. We also see that the changes in the results appear small when compared to the huge change in β, though revelations from experiments show that for $\beta = 500$, the effects actually seem to diminish. This behaviour, along with the large amount of iterations needed to spot the parameter effects, likely stems from the nature of the steepest descent method, as this behaviour resembles how the method acts on a simple Rosenbrock function. This tells us that the Hessian of the problem is ill-conditioned. It is therefore likely that the result images in figure 4.2 – and probably most other results from $\mathcal{F}$ presented in this paper – are not necessarily close to an actual minimum.

4.1.2 Parameter effects on $\mathcal{F}_W$

From figure 4.3 we can see some interesting effects. First of all, we notice that the engravings on each coin have been distorted near the edges. This distortion is a result of the functional trying to preserve the outline of the coins. The reason for prioritizing the outline is due to the large contrasts between the coins and the background, which opts the functional to leave the outlines untouched, rather than focus on preserving the engravings.

An unwanted side effect of image denoising through minimizing functionals is artifact production. While mathematically explicable, the functional $\mathcal{F}_W$ seems to find flow structures where there are supposed to be none, and they could be generated from noise or from non-flowing structures in the image. In figure 4.4, we compare how the denoiser behaviour changes with noise. The results seen in the bottom right image show that the background has been littered with artifacts that are clearly not there in the original image. While a value of $\alpha = 100$ might be a bad parameter choice for decent denoising, it clearly

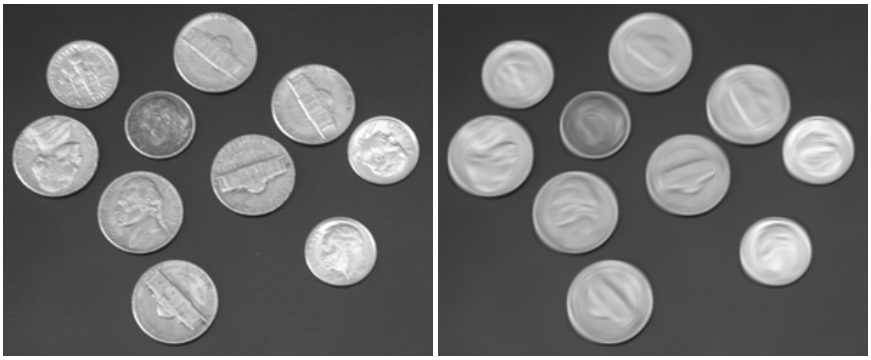


Figure 4.3: Example of how some structures can dominate others when minimizing $\mathcal{F}_W$, here with parameters $\alpha = 100$, $\gamma = 1$ $\sigma = 1$ and $\rho = 10$. The original data is on the left, and the denoised data is on the right. Image size: 300×246

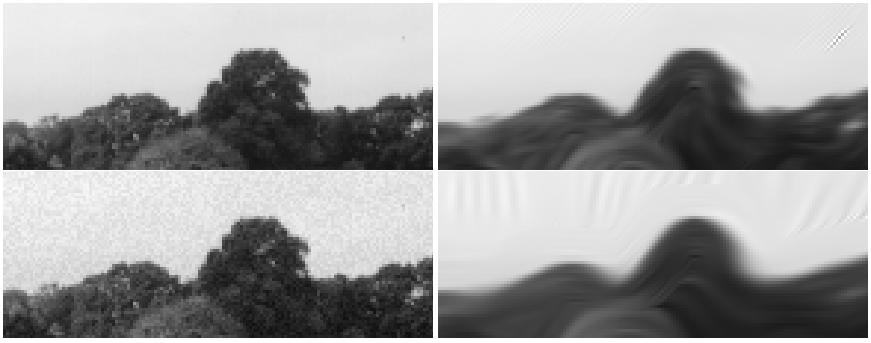


Figure 4.4: An example of how noise can lead to artifact generation. In both cases, $\alpha = 100$, $\gamma = 1$ and $\rho = 10$. The left column contains non-noisy and noisy data, and the right column are their respective denoised results. Image size: 145×56.

demonstrates the problem of artifact generation. Similar effects are seen in figure 4.5. Although these artifacts are not generated from noise, but rather from spots, such structural degradation should still be regarded as a defect.

Remark. The keen observer may also notice the tendencies of instability in the upper right corner of the top right image of figure 4.4. Comparing this area to the original image on the top left, what could appear to simply be dust on the camera lens has been severely distorted. These issues could have risen from theoretical problems of uniqueness of diagonalization whenever the eigenvalues of the structure tensor coincide, mentioned in appendix A.1. The fact that experiments reveal that this particular artifact completely dissappears by lowering γ, further increasing suspicion towards these theoretical issues.

Figure 4.6 captures the effects of ρ, σ and γ – three parameters which for this functional all occur in the same term. The figure shows a satellite image of a hurricane. Compare first figures 4.6c–4.6e to figure 4.6b. When ρ is small, every pixel focuses only on a small area,

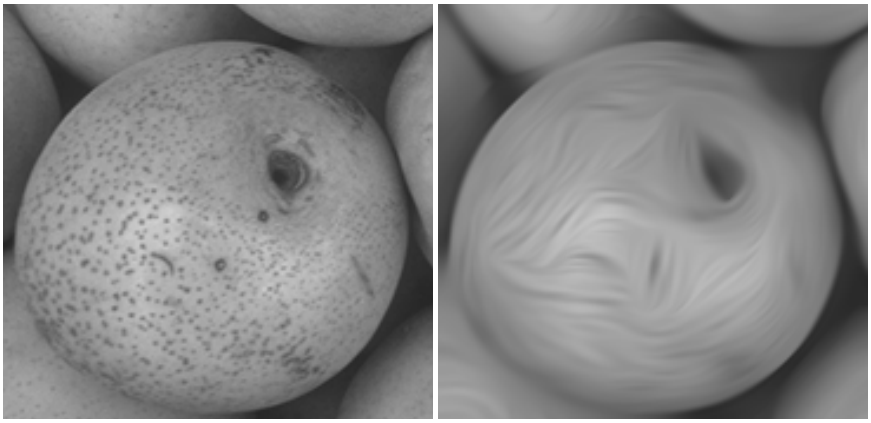
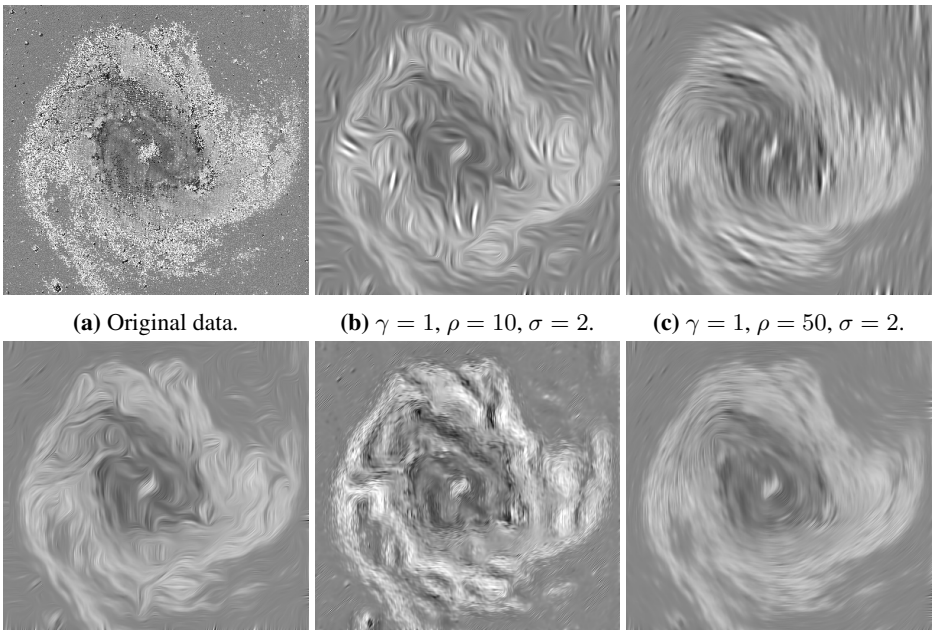


Figure 4.5: An example of structural distortions when minimizing $\mathcal{F}_W$. Original data on the left and denoised data on the right. Here $\alpha = 100$, $\gamma = 1$, $\sigma = 1$ and $\rho = 10$.

and the large amounts of noise makes it difficult for the functional to correctly determine the flow directions. In this case, the entire image contains a single structure of interest, thus by increasing ρ, more data is accounted for and the superstructure of the hurricane can eventually be found, as seen in figure 4.6c. Increasing σ immediately reduces local noise, thereby significantly reducing artifact generation, which can be seen by comparing figures 4.6b and 4.6e. Comparing figures 4.6b and 4.6d, we can see that by decreasing γ, the flow lines have become more distinguished. Finally, figure 4.6f shows the result of a smart choices of parameters. In its entirety, figure 4.6 again demonstrates how parameter choices should reflect the data we wish to denoise.

4.2 Verdict

At the epitaph of this paper, we now wish to compare some results in which both functionals excel. Figure 4.7 shows the results of the denoiser on both non-noisy, noisy and very noisy data, with the results of $\mathcal{F}$ in the middle column and $\mathcal{F}_W$ in the right column, respectively. The top row is the results of non-nosiy data, while to the middle and bottom row we have added random Gaussian noise of intensity 50 and 100, respectively. By comparing the different rows of figure 4.7, we can see for the non-local functional more apparent blurring, particularly in the bottom parts of the pictures, while for the functional $\mathcal{F}_W$, these same areas are more prone to artifact generation. The main features of the fingerprint, however, have been clearly reproduced by the denoisers. The conclusion from these results then has to be that the denoisers indeed removes noise, and quite effectively so, from images with clear anisotropies. Comparing the results in the middle and right columns, we can see that the prints are slightly more contrasted for $\mathcal{F}_W$. At the same time, looking at the bottom parts of the images, the functional $\mathcal{F}_W$ tends to create more apparent artifacts than the non-local $\mathcal{F}$. Other experiments on this very image also reveal that for

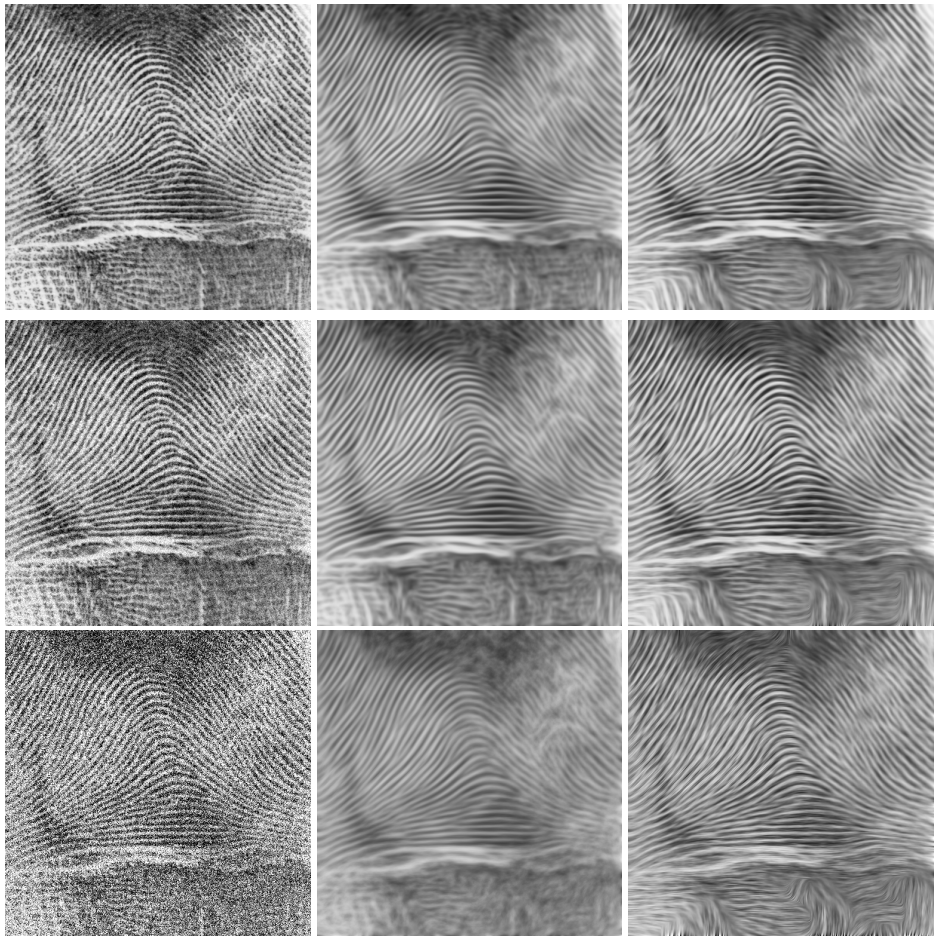


(a) Original data. **(b)** $\gamma = 1$, $\rho = 10$, $\sigma = 2$. **(c)** $\gamma = 1$, $\rho = 50$, $\sigma = 2$.

(d) $\gamma = 0.001$, $\rho = 10$, $\sigma = 2$. **(e)** $\gamma = 1$, $\rho = 10$, $\sigma = 10$. **(f)** $\gamma = 0.001$, $\rho = 100$, $\sigma = 10$.

Figure 4.6: Examples signifying the importance of structure sizes. In all cases, $\alpha = 50$. Remaining parameter selection is shown in figure texts. Image size: 400×378.

other parameter choices for $\mathcal{F}_W$, artifact generation is more apparent and occur in other parts of the image as well.



(a) Original data with increasing amounts of noise.

(b) $\alpha = 0.001$, $\beta = 1$, $\rho = 20$.

(c) $\alpha = 50$, $\gamma = 0.01$, $\sigma = 2$, $\rho = 20$.

Figure 4.7: Results of the denoisers. Row-wise the data become increasingly noisy, with the top image having no added noise, the middle row has an added Gaussian noise of intensity 50, and the bottom image has an added Gaussian noise of intesity 100. The middle column show results of denoising through the non-local $\mathcal{F}$, while the right column show results of denoising through $\mathcal{F}_W$. Parameter choices are seen below their respective columns. Image size: 512×512.

Chapter 5

Conclusion

In this paper, we have proposed a non-local coherence enhancing functional $\mathcal{F}$ for denoising images. We have also compared its behaviour to that of the functional $\mathcal{F}_W$, formulated as a PDE in Weickert (1999). For both functionals, we have proven the existence of minimizers and derived their formal gradients. The denoising properties of the proposed non-local functional $\mathcal{F}$ and the functional $\mathcal{F}_W$ are demonstrated through numerical experiments. These experiments reveal that both functionals preserve anisotropy and reduce local noise, and work very well for images of distinct and uniformly sized structures, sharp contrasts and protruding anisotropies. By choosing parameters to align with the given data and the amount of enhancement we seek, noise is removed while important features like edges are preserved.

However, neither method is without drawbacks. The functional $\mathcal{F}_W$ produced artifacts and structural distortions in areas where no dominating anisotropy were found or when conflicting flows met, while the proposed $\mathcal{F}$ instead blurred such areas. As for numerical results, the denoisers require many iterations to converge, and do not necessarily converge to a minimum. For the non-local $\mathcal{F}_W$ we have seen that for large parameter choices, the number of iterations required to reach a minimum increased. The large amount of iterations required likely stem from an ill-conditioned Hessian of the functional, and due to this, the steepest descent method may be not optimal for solving the minimising problems.

Appendix A

Appendix

A.1 Diagonalization of real symmetric matrices

Let $A \in \mathbb{R}^{2\times 2}$ be symmetric i.e. there are real numbers a, b and c such that

$$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}.$$

Then there exists a unique diagonal matrix Σ of ordered eigenvalues of A, and an orthogonal matrix Q whose columns are eigenvectors corresponding to each of these eigenvalues, such that

$$A = Q\Sigma Q^T.$$

Note that Q is unique if and only if the eigenvalues of A are different. The eigenvalues of A are the solutions of

$$\det(\sigma I - A) = 0, \tag{A.1}$$

and it is simple to show that these solutions are

$$\sigma = \frac{a + c \pm \sqrt{(a-c)^2 + 4b^2}}{2}. \tag{A.2}$$

From this it is clear that the eigenvalues coincide only when $a = c$ and $b = 0$, and thus, stability issues are bound to arise whenever A is almost equal to I. More details on this

can be found in section A.1.1. As Q is orthogonal, $QQ^T = Q^TQ = I$. As such, we find that Q has to be on the form

$$Q = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}.$$

The phase angle α can be found by solving $[(\sigma I - A)]_i Q_i = 0$ for either column vector Q_i, that is

$$\begin{aligned} (\sigma_1 - a)\cos(\alpha) + b\sin(\alpha) = 0, \text{ or} \\ -b\sin(\alpha) + (\sigma_2 - c)\cos(\alpha) = 0. \end{aligned}$$

Either way, for $b \neq 0$, we find the phase angle to be

$$\alpha = \arctan\Big(\frac{\sigma_1 - a}{b}\Big) = \arctan\Big(\frac{c - \sigma_2}{b}\Big). \tag{A.3}$$

The last equality in the equation above is easily verified by inserting σ_1 and σ_2 from (A.2).

In the case when $b = 0$, the matrix A is already diagonal. If $a > c$, we see immediately that $A = \Sigma$ and thus $Q = \pm I$. If $a < c$, we can find, by calculating $QAQ^T = \Sigma$, that

$$Q = \pm \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

This corresponds to the phase angle $\alpha = \pm\pi/2$.

A.1.1 Stability issues for small b

When b is small relative to $a - c$, very tiny perturbations of b will lead to large changes in α. This will likely lead to numerical issues when attempting to diagonalize matrices that are already almost diagonal on a computer. A very simple quick fix is simply to use the approxamtion $b = 0$ for a certain threshold $|b| < \epsilon$. In other words, for $|b| < \epsilon$, we approximate

$$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \approx \begin{bmatrix} a & 0 \\ 0 & c \end{bmatrix}.$$

A.2 Directional derivative of $\mathcal{R}$

In this section, we seek to derive a formula for directional derivative of $\mathcal{R}$ as defined in section 2.3 and find the criteria needed for a formal gradient to exist. That is, for a fixed $\xi \in \mathbb{R}^2$, we wish to find the functional

$$D\mathcal{R}(\xi;\tau) = \lim_{\epsilon\to 0} \frac{\mathcal{R}(\xi+\epsilon\tau) - \mathcal{R}(\xi)}{\epsilon}$$

and find a function $\operatorname{grad}\mathcal{R}(\xi, x)$ such that

$$D\mathcal{R}(\xi;\tau) = \left\langle \operatorname{grad}\mathcal{R}(\xi), \tau \right\rangle_{L^2(\Omega)}.$$

A.2.1 First variation of $\mathcal{R}$, part one

Let $\mathcal{R} : L^q(\Omega;\mathbb{R}^2) \to \mathbb{R} \cup \{+\infty\}$ be the functional

$$\mathcal{R}(\xi) = \frac{1}{2p}\int_{\Omega\times\Omega} w(|x-y|)\left|\left\langle \xi(x), \xi(y)^\perp \right\rangle_{\mathbb{R}^2}\right|^p dx\, dy \tag{A.4}$$

where $\Omega \subset \mathbb{R}^2$, $p > 1$, and $w : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is a bounded Borel measureable function, with $w(0) > 0$. The first variation of $\mathcal{R}$ in direction τ is then defined as the limit

$$D\mathcal{R}(\xi;\tau) := \lim_{\epsilon\to 0} \frac{\mathcal{R}(\xi+\epsilon\tau) - \mathcal{R}(\xi)}{\epsilon} \tag{A.5}$$

for some direction $\tau : L^q(\Omega;\mathbb{R}^2) \to \mathbb{R}^2$. If the limit exists for any $\tau \in L^q(\Omega;\mathbb{R}^2)$ and defines a bounded linear mapping $D\mathcal{R}(\xi) : L^q \to \mathbb{R}$, we say the functional $\mathcal{R}(\xi)$ is Gâteaux differentiable, and that $D\mathcal{R}(\xi;\tau)$ is the directional derivative of $\mathcal{R}(\xi)$ in direction τ.

Define now the functions $g : \mathbb{R} \to \mathbb{R}$ and $f_1, f_2 : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ as $f_1(x,y) = \langle x, y\rangle$, $g(t) = |t|^p$ and $f_2(x,y) = f_1 \circ g(x,y)$. Then

$$\mathcal{R}(\xi) = \frac{1}{2p}\int_{\Omega\times\Omega} w(|x-y|) f_2(\xi(x), \xi(y)^\perp) dx\, dy, \tag{A.6}$$

and consequently,

$$\begin{aligned} D\mathcal{R}(\xi;\tau) = \lim_{\epsilon\to 0}\frac{1}{\epsilon}\Bigg[\frac{1}{2p}\int_{\Omega\times\Omega} w(|x-y|) f_2\left((\xi+\epsilon\tau)(x), (\xi+\epsilon\tau)(y)^\perp\right) dx\, dy \\ - \frac{1}{2p}\int_{\Omega\times\Omega} w(|x-y|) f_2(\xi(x), \xi(y)^\perp) dx\, dy\Bigg]. \end{aligned} \tag{A.7}$$

One can prove by applying the Lebesgue dominated convergence theorem that the limit and integral of (A.7) can switch places, under sufficient criteria. Proving this is quite a bit of work, so for now, we focus on that, and come back to the first variation later.

A.2.2 Convergence of the directional derivative

Our goal in this section is to derive sufficient criteria such that we can apply the Lebesgue dominated convergence theorem to the functional $\mathcal{R}(\xi)$. Similar work has already been done in theorem 3.37 of Dacorogna (2008), where the author finds exactly such criteria, and moreover, some interesting consequences of these. To lessen confusion, we first introduce some notation similar to that of Dacorogna (2008).

First we define the density

$$s\big(x,y;\xi(x),\zeta(y)\big) = \frac{1}{2p}w(|x-y|)\Big|\big\langle \xi(x),\zeta(y)^{\perp}\big\rangle\Big|^{p}.$$

such that $(x,y)\in\Omega\times\Omega$ and $\big(\xi(x),\zeta(y)\big)\in\mathbb{R}^2\times\mathbb{R}^2$. We define

$$S(\xi,\zeta) = \int_{\Omega\times\Omega} s\big(x,y;\xi(x),\zeta(y)\big)dy\,dx$$

and notice that $S(\xi,\xi)=\mathcal{R}(\xi)$. Let $\phi,\psi:\Omega\times\Omega\to\mathbb{R}$. Define the functional

$$\begin{aligned}\mathcal{L}\big((u,v),(\phi,\psi)\big) := \int_{\Omega\times\Omega}\Big(&\big\langle D_{\xi}s\big(x,y;\nabla u(x),\nabla v(y)\big),\nabla\phi\big\rangle\\ &+\big\langle D_{\zeta}s\big(x,y;\nabla u(x),\nabla v(y)\big),\nabla\psi\big\rangle\Big)dy\,dx\end{aligned}$$

which is the directional derivative of $S(\nabla u,\nabla v)$ in direction $(\nabla\phi,\nabla\psi)$.

Theorem 3.37 of Dacorogna (2008) now states that if s is a Carathéodory function and

$$\begin{aligned}\Big|s\big(x,;\xi(x),\zeta(y)\big)\Big| &\le \alpha_0(x,y)+\beta\big(|\xi(x)|^q+|\zeta(y)|^q\big)\\ &\qquad\forall\,(\xi(x),\zeta(y))\in\mathbb{R}^2\times\mathbb{R}^2,\\ \Big|D_{\xi}s\big(x,y;\xi(x),\zeta(y)\big)\Big| &\le \alpha_1(x,y)+\beta\left(|\xi(x)|^{q-1}+|\zeta(y)|^{q-1}\right),\\ \Big|D_{\zeta}s\big(x,y;\xi(x),\zeta(y))\big)\Big| &\le \alpha_2(x,y)+\beta\left(|\xi(x)|^{q-1}+|\zeta(y)|^{q-1}\right),\end{aligned}$$

for a.e. $(x,y)\in\Omega\times\Omega$, $\alpha_0\in L^1(\Omega\times\Omega)$, $\alpha_1,\alpha_2\in L^{q/(q-1)}$ and $\beta>0$, then for any $\bar{\phi}$,

$$\mathcal{L}\big((\bar{u},\bar{u},(\bar{\phi},\bar{\phi})\big)=0$$

for a solution $\bar{u}$ to the minimising problem (2.7). In the proof of the theorem it is also shown that $\mathcal{L}$ is the Gâteaux derivative of S.

We plan to utilize this theorem to show that the Gâteaux derivatives $DS(\xi,\zeta;\phi,\psi)$ of $S(\xi,\zeta)$ exists — i,e, that the dominated Lebesgue convergence theorem holds for $DS(\xi,\zeta)$. We will prove that condition (A.2.2) holds for our functional in three parts.

i) s is Carathéodory and the estimate for s.

Recall

$$s\big(x,y;\xi(x),\zeta(y)\big) = \frac{1}{2p}w(|x-y|)\Big|\big\langle \xi(x),\zeta(y)^{\perp}\big\rangle\Big|^{p}.$$

Since $w(\cdot)$ is a Borel function and $\langle\cdot,\cdot\rangle$ is a continuous functional, it follows that s is Carathéodory .

Further, we know that

$$\begin{aligned}\Big|s\big(x,y;\xi(x),\zeta(y)\big)\Big| &\le \frac{1}{2p}\sup_{(x,y)\in\Omega\times\Omega}\big(w(|x-y|)\big)\Big|\big\langle \xi(x),\zeta(y)^{\perp}\big\rangle\Big|^{p}\\ &\le \frac{1}{2p}\sup_{(x,y)\in\Omega\times\Omega}\big(w(|x-y|)\big)\big|\xi(x)\big|^{p}\big|\zeta(y)\big|^{p}.\end{aligned}$$

We denote $\sup_{(x,y)\in\Omega\times\Omega} w(|x-y|) := W$. Noting that for any non-negative numbers a, b and n, we have $a^n b^n \le a^{2n} + b^{2n}$. It follows that

$$\begin{aligned}\Big|s\Big(\big(x,y\big),\big(\xi(x),\zeta(y)\big)\Big)\Big| &\le \frac{W}{2p}\big(|\xi(x)|^{2p} + |\zeta(y)|^{2p}\big)\\ &\le \alpha_0(x) + \beta\big(|\xi(x)|^{q'} + |\zeta(y)|^{q'}\big)\end{aligned} \tag{A.8}$$

with $\beta = W/2p > 0$, $\alpha_0(x) = 0$ and $q' = 2p$. Obviously $\alpha_0 \in L^1(\Omega)$.

Hence s is Carathéodory, and

$$\Big|s\Big(\big(x,y\big),\big(\xi(x),\zeta(y)\big)\Big)\Big| \le \alpha_0(x) + \beta\left(|\xi(x)|^{q'} + |\zeta(y)|^{q'}\right)$$

for some $0 < \alpha_0(x) \in L^1(\Omega)$, proving part i).

ii) Estimate for $|D_\xi s|$.

From the definition of s, we get immediately

$$\left|D_\xi s\big(x,y;\xi(x),\zeta(y)\big)\right| = \frac{1}{2p} w\big(|x-y|\big)\,\big\langle \xi(x),\zeta(y)^\perp\big\rangle \left|\big\langle \xi(x),\zeta(y)^\perp\big\rangle\right|^{p-2} dy \tag{A.9}$$

Consequently, as $p \geq 1 > 0$ and $w(\cdot)$ is non-negative,

$$\begin{aligned}\left|D_\xi s\big(x,y;\xi(x),\zeta(y)\big)\right| \leq& \frac{1}{2p}\sup_{(x,y)\in\Omega}\big(w(|x-y|)\big)\left|\big\langle \xi(x),\zeta(y)^\perp\big\rangle\right|^{p-1} \\ \leq& \frac{1}{2p}\sup_{(x,y)\in\Omega}\big(w(|x-y|)|\big)\xi(x)|^{p-1}|\zeta(y)|^{p-1}\end{aligned} \tag{A.10}$$

Again, denote $W := \sup_{(x,y)\in\Omega\times\Omega} w(|x-y|)$ and exploit the fact that $a^n b^n \leq a^{2n} + b^{2n}$ for non-negative a, b and n to find that

$$\begin{aligned}\left|D_\xi s\big(x,y;\xi(x),\zeta(y)\big)\right| &\leq \frac{1}{2p} W|\xi(x)|^{p-1}|\zeta(y)|^{p-1} \\ &\leq \frac{1}{2p} W\Big(|\xi(x)|^{2(p-1)} + |\zeta(y)|^{2(p-1)}\Big) \\ &\leq \alpha_1(x,y) + \beta\Big(\big|\xi(x)\big|^{2(p-1)} + |\zeta(y)|^{2(p-1)}\Big),\end{aligned} \tag{A.11}$$

where $\beta = W/2p$ and $\alpha_1(x,y) = 0$. Defining $2(p-1) = q-1$, we find that

$$\left|D_\xi s\Big(\big(x,y\big),\big(\xi(x),\zeta(y)\big)\Big)\right| \leq \alpha_1(x,y) + \beta\Big(\big|\xi(x)\big|^{q-1} + |\zeta(y)|^{q-1}\Big), \tag{A.12}$$

thus proving part ii).

iii) Estimate for $|D_\zeta s|$.

This is completely analogous to part ii).

We have seen that (A.8) holds whenever $q \geq 2p$, and that (A.12) and (A.12) hold for $q \geq 2p-1$. Theorem 3.37 of Dacorogna (2008) thus holds true for $S(\nabla u, \nabla v)$ for any $u, v \in W^{1,2p}(\Omega)$, and its Gâteaux derivatives $DS(\nabla u, \nabla v; \nabla\tau, \nabla\eta)$ exist for any direction $(\tau,\eta) \in W^{1,2p}$. Consequently we know that as long as $u, h \in W^{1,2p}$, the Lebesgue dominated convergence theorem is applicable on $D\mathcal{R}(\nabla u)(\nabla h) = DS(\nabla u, \nabla u; \nabla h, \nabla h)$ for any $u_k \to u, h_k \to h$. □

The criteria we were looking for in section A.2.1 are therefore that $u, h \in W^{1,2p}$, and are also the criteria required mentioned in section 2.5.2 for $\mathcal{F}$ to be differentiable, i.e. that the Lebesgue dominated convergence theorem can be applied.

A.2.3 First variation of $\mathcal{R}$, part two

Having proven that the Lebesgue dominated convergence theorem is indeed applicable to the functional $\mathcal{R}$, we continue where we left at the end of section A.2.1. We now know that we can switch the order of limit and integration of (A.7) whenever $u, h \in W^{1,2p}$. We thus find that

$$\begin{aligned} D\mathcal{R}(\xi;\tau) = \frac{1}{2p}\int_{\Omega\times\Omega} w(|x-y|)\lim_{\epsilon\to 0}\frac{1}{\epsilon}\Big[f_2\left((\xi+\epsilon\tau)(x),(\xi+\epsilon\tau)(y)^\perp\right) \\ - f_2(\xi(x),\xi(y)^\perp)\Big]dx\,dy. \end{aligned} \tag{A.13}$$

Recalling $f_1(x,y) = \langle x, y\rangle$, $g(t) = |t|^p$ and $f_2(x,y) = f_1 \circ g(x,y)$, we recognize the limit in (A.13) as

$$\begin{aligned} &\lim_{\epsilon\to 0}\frac{1}{\epsilon}\Big[f_2\left((\xi+\epsilon\tau)(x),(\xi+\epsilon\tau)(y)^\perp\right) - f_2(\xi(x),\xi(y)^\perp)\Big] \\ &\qquad =: Df_2\big(\xi(x),\xi(y)^\perp;\tau(x),\tau(y)^\perp\big). \end{aligned}$$

Thus finding $D\mathcal{R}(\xi;\tau)$ reduces to finding $Df_2(\xi(x),\xi(y)^\perp;\tau(x),\tau(y)^\perp)$. By the definition of f_2 and using the chain rule, one obtains that

$$Df_2(\xi(x),\xi(y)^\perp;\tau(x),\tau(y)^\perp) = g'\big(f_1(\xi(x),\xi(y)^\perp)\big)\cdot Df_1\big(\xi(x),\xi(y)^\perp\big)(\tau(x),\tau(y)^\perp),$$

with $g'(t) = p|t|^{p-1}\mathrm{sgn}(t) = pt|t|^{p-2}$, and

$$\begin{aligned} Df_1(\xi(x),\xi(y)^\perp;\tau(x),\tau(y)^\perp) &= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\Big[f_1\left((\xi+\epsilon\tau)(x),(\xi+\epsilon\tau)(y)^\perp\right) \\ &\qquad - f_1(\xi(x),\xi(y)^\perp)\Big] \\ &= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\Big[\left\langle\xi(x),\xi(y)^\perp\right\rangle \\ &\qquad + \epsilon\left\langle\nabla+h(x),\xi(y)^\perp\right\rangle \\ &\qquad + \epsilon\left\langle\xi(x),\tau(y)^\perp\right\rangle + \mathcal{O}(\epsilon^2) \\ &\qquad - \left\langle\xi(x),\xi(y)^\perp\right\rangle\Big] \\ &= \lim_{\epsilon\to 0}\Big[\left\langle\tau(x),\xi(y)^\perp\right\rangle\left\langle\xi(x),\tau(y)^\perp\right\rangle + \mathcal{O}(\epsilon)\Big] \\ &= \left\langle\tau(x),\xi(y)^\perp\right\rangle + \left\langle\xi(x),\tau(y)^\perp\right\rangle \\ &= \left\langle\tau(x),\xi(y)^\perp\right\rangle - \left\langle\tau(y),\xi(x)^\perp\right\rangle, \end{aligned}$$

where we have used that $\left\langle a, b^\perp\right\rangle = -\left\langle a^\perp, b\right\rangle$. This gives us that

$$Df_2(\xi(x),\xi(y)^\perp;\tau(x),\tau(y)^\perp) = p\langle\xi(x),\xi(y)^\perp\rangle\Big|\langle\xi(x),\xi(y)^\perp\rangle\Big|^{p-2} \cdot\Big(\big\langle\tau(x),\xi(y)^\perp\big\rangle - \big\langle\tau(y),\xi(x)^\perp\big\rangle\Big),$$

which in turn yields

$$D\mathcal{R}(\xi;\tau) = p\int_{\Omega\times\Omega} w(|x-y|)\langle\xi(x),\xi(y)^\perp\rangle\Big|\langle\xi(x),\xi(y)^\perp\rangle\Big|^{p-2} \cdot\Big(\big\langle\tau(x),\xi(y)^\perp\big\rangle - \big\langle\tau(y),\xi(x)^\perp\big\rangle\Big)dx\,dy.$$

We split the integrand to get

$$D\mathcal{R}(\xi;\tau) = p\frac{1}{2p}\int_{\Omega\times\Omega} w(|x-y|)\left\langle\xi(x),\xi(y)^\perp\right\rangle\left|\left\langle\xi(x),\xi(y)^\perp\right\rangle\right|^{p-2}\left\langle\tau(x),\xi(y)^\perp\right\rangle dx\,dy - p\frac{1}{2p}\int_{\Omega\times\Omega} w(|x-y|)\left\langle\xi(x),\xi(y)^\perp\right\rangle\left|\left\langle\xi(x),\xi(y)^\perp\right\rangle\right|^{p-2}\left\langle\tau(y),\xi(x)^\perp\right\rangle dx\,dy.$$

Next, we again use the fact that $\langle a,b^\perp\rangle = -\langle a^\perp,b\rangle$ and $\langle a,b\rangle = \langle b,a\rangle$ to find

$$\begin{aligned} D\mathcal{R}(\xi;\tau) = \frac{1}{2}\int_{\Omega\times\Omega} & w(|x-y|)\left\langle\xi(x),\xi(y)^\perp\right\rangle\left|\left\langle\xi(x),\xi(y)^\perp\right\rangle\right|^{p-2} \\ & \cdot\left\langle\tau(x),\xi(y)^\perp\right\rangle dx\,dy \\ + \frac{1}{2}\int_{\Omega\times\Omega} & w(|x-y|)\left\langle\xi(y),\xi(x)^\perp\right\rangle\left|-\left\langle\xi(y),\xi(x)^\perp\right\rangle\right|^{p-2} \\ & \cdot\left\langle\tau(y),\xi(x)^\perp\right\rangle dx\,dy \end{aligned}$$

Noting that $|-a| = |a|$ for all $a\in\mathbb{R}$ and that the order of integration can be switched, one can see by relabelling the variables of the second integral that

$$\begin{aligned} D\mathcal{R}(\xi;\tau) &= \int_{\Omega\times\Omega} w(|x-y|)\left\langle\xi(x),\xi(y)^\perp\right\rangle|\left\langle\xi(x),\xi(y)^\perp\right\rangle|^{p-2}\left\langle\tau(x),\xi(y)^\perp\right\rangle dx\,dy \\ &= \int_\Omega\left\langle\int_\Omega \tau(x)w(|x-y|)\left\langle\xi(x),\xi(y)^\perp\right\rangle|\left\langle\xi(x),\xi(y)^\perp\right\rangle|^{p-2}dx,\xi(y)^\perp\right\rangle dy \end{aligned}$$

where in the last equality we have moved the integral over x into the inner product involving $\tau(x)$. Define

$$\Upsilon(\xi,\zeta) = \langle\xi,\zeta\rangle\Big|\langle\xi,\zeta\rangle\Big|^{p-2},$$

such that

$$D\mathcal{R}(\xi;\tau) = \int_\Omega \left\langle \int_\Omega \tau(x) w(|x-y|) \Upsilon(\xi(x), \xi(y)^\perp)\, dx, \xi(y)^\perp \right\rangle dy. \tag{A.14}$$

This is the general form of the directional derivative $D\mathcal{R}(\xi;\tau)$, and is exactly (2.16) from lemma 1.

A.3 The formal gradient of $\mathcal{F}$

For a series of practical applications, for instance in the gradient descent method, it is of interest to derive a gradient of a functional. We now want to derive the formal gradient of $\mathcal{F}$. From eq. (2.17) and lemma 1, we have that

$$\begin{aligned} D\mathcal{F}(u;h) = \int_\Omega \Big[&(u(x) - u_0(x))h(x) + \alpha \nabla u(x)^T \nabla h(x) \\ &+ \left\langle \beta \int_\Omega w(|x-y|) \Upsilon(\nabla u(x), \nabla u(y)^\perp) \nabla u(y)^\perp dy, \nabla h(x) \right\rangle \Big] dx. \end{aligned} \tag{A.15}$$

We integrate by parts and find that

$$\begin{aligned} D\mathcal{F}(u;h) = \int_\Omega \Big[&\big(u(x) - u_0(x)\big)h(x) + \alpha \Delta u(x) h(x) \\ &+ 2\beta \int_\Omega \operatorname{div}\Big(w(|x-y|) \Upsilon(\nabla u(x), \nabla u(y)^\perp) \nabla u(y)^\perp \Big) dy\, h(x) \Big] dx \\ &+ \int_{\partial\Omega} h(x) \Big\langle \nabla u(x) \\ &+ \beta \int_\Omega w(|x-y|) \Upsilon(\nabla u(x), \nabla u(y)^\perp) \nabla u(y)^\perp dy, \mathbf{n}(x) \Big\rangle dx \end{aligned}$$

where $\mathbf{n}(x)$ is the normal vector on Ω at $x \in \partial\Omega$. For a formal gradient to exist, this boundary integral has to vanish. By a variational argument in h, this imposes the following natural boundary condition on u:

$$\left\langle \nabla u(x) + \beta \int_\Omega w(|x-y|) \Upsilon(\nabla u(x), \nabla u(y)^\perp) \nabla u(y)^\perp dy, \mathbf{n}(x) \right\rangle = 0,\ x \in \partial\Omega. \tag{A.16}$$

When the above condition holds, it follows that

$$D\mathcal{F}(u;h) = \int_\Omega \Big[(u(x) - u_0(x))h(x) + \alpha\Delta u(x)h(x) + \mathcal{J}(x,\nabla u)h(x)\Big]\,dx$$
$$= \int_\Omega \Big[u(x) - u_0(x) + \alpha\Delta u(x) + \mathcal{J}(x,\nabla u)\Big]h(x)\,dx$$

where

$$\mathcal{J}(x,\xi) = \int_\Omega \operatorname{div}\Big(w(|x-y|)\Upsilon(\nabla u(x), \nabla u(y)^\perp)\,dx, \xi(y)^\perp dy\Big).$$

This yields the formal gradient of $\mathcal{F}$ as

$$\operatorname{grad}\mathcal{F}(u,x) = u(x) - u_0(x) + \alpha\Delta u(x) + 2\beta\mathcal{J}(x,\nabla u(x)),$$

leading immediately to the following representation of $D\mathcal{F}(u;h)$:

$$D\mathcal{F}(u;h) = \langle \operatorname{grad}\mathcal{F}, h\rangle_{L^2} = \int_\Omega \operatorname{grad}\mathcal{F}(u,x)h(x)dx \tag{A.17}$$

□

A.3.1 Simplification for $p = 2$

The main interest of this paper will be for the case when $p = 2$. For the functional $\mathcal{F}$, defined in (2.2), the main results to draw from this special case is that $\Upsilon(\xi,\zeta)$ reduces to $\langle\xi,\zeta\rangle$, and that we can reformulate $\mathcal{R}(\xi)$ using *convolution* – both which significantly simplify the computations. These reformulations also have interesting consequences for the formal gradient $\operatorname{grad}\mathcal{F}$.

When $p = 2$, the functional $\mathcal{R}(\xi)$ becomes

$$\mathcal{R}(\xi) = \int_{\Omega\times\Omega} w(|x-y|)\left\langle \xi(x), \xi(y)^\perp\right\rangle_{\mathbb{R}^2}^2 dy\,dx$$
$$= \int_{\Omega\times\Omega} w(|x-y|)\left(\xi(x)_1\xi(y)_2 - \xi(x)_2\xi(y)_1\right)^2 dy\,dx$$
$$= \int_\Omega \Big[\xi_1(x)^2\int_\Omega w(|x-y|)\xi_2(y)^2dy + \xi_2(x)^2\int_\Omega w(|x-y|)\xi_1(y)^2dy$$
$$- 2\xi_1(x)\xi_2(x)\int_\Omega w(|x-y|)\xi_1(y)\xi_2(y)dy\Big]\,dx.$$
$$= \int_\Omega \big[\xi_1(x)^2(\tilde{w}_x * \xi_2)(x) + \xi_2(x)^2(\tilde{w}_x * \xi_1)(x)$$
$$- 2\xi_1(x)\xi_2(x)(\tilde{w}_x * \xi_1\xi_2)(x)\big]dx.$$

On the other hand, $D\mathcal{R}(\xi,\tau)$ becomes

$$\begin{aligned} D\mathcal{R}(\xi;\tau) &= \int_\Omega \int_\Omega w(|x-y|)\langle \xi(x), \xi(y)^\perp\rangle\langle \tau(x), \xi(y)^\perp\rangle dy\, dx \\ &= \int_\Omega \int_\Omega \Big\langle \tau(x), w(|x-y|)\langle \xi(x), \xi(y)^\perp\rangle \xi(y)^\perp \Big\rangle dy\, dx \\ &= \int_\Omega \Big\langle \tau(x), \int_\Omega w(|x-y|)\langle \xi(x), \xi(y)^\perp\rangle \xi(y)^\perp dy \Big\rangle dx \\ &= \int_\Omega \langle \tau(x), \big(w * (\xi^\perp \otimes \xi^\perp)\big)(x)\xi(x)\rangle dx, \end{aligned}$$

where $\otimes$ is the $\mathbb{R}^2$ tensor product. Let now $\nabla h = \tau$, as is the case in our problem. Inserting the above into $D\mathcal{F}(u;h)$ and integrating by parts then yield

$$\begin{aligned} D\mathcal{F}(u;h) &= \int_\Omega \big(u(x) - u_0(x)\big)h(x) + \alpha\langle \nabla u(x), \nabla h(x)\rangle \\ &\qquad + \beta\langle \nabla h(x), \big(w * (\xi^\perp \otimes \xi^\perp)\big)(x)\xi(x)\rangle dx \\ &= -\int_\Omega h(x)\big(u(x) - u_0(x)\big)dx + \int_\Omega \alpha\Delta u(x) \\ &\qquad + \beta \mathrm{div}\Big(\big(w * (\xi^\perp \otimes \xi^\perp)\big)(x)\xi(x)\Big)\Big) dx \\ &\qquad + \int_{\partial\Omega} \Big\langle h(x)\Big(\nabla u(x) + \big(w * (\xi^\perp \otimes \xi^\perp)\big)(x)\xi(x)\Big), \mathbf{n}(x)\Big\rangle dx. \end{aligned}$$

The imposed boundary integral is the same as in (2.18), only reformulated for $p = 2$. Imposing the boundary condition

$$\langle \nabla u(x) + \big(w * (\xi^\perp \otimes \xi^\perp)\big)(x)\xi(x), \mathbf{n}(x)\rangle = 0 \text{ on } \partial\Omega, \tag{A.18}$$

yields the directional derivative

$$\begin{aligned} D\mathcal{F}(u;h) &= \Big\langle u - u_0 - \alpha\Delta u - \beta\,\mathrm{div}\Big(\big(w * (\nabla u^\perp \otimes \nabla u^\perp)\big)\nabla u\Big), h\Big\rangle_{L^2} \\ &= \langle \mathrm{grad}\,\mathcal{F}(u,\cdot), h\rangle_{L^2}. \end{aligned} \tag{A.19}$$

where the formal gradient $\mathrm{grad}\,\mathcal{F}$ has the form

$$\begin{aligned} \mathrm{grad}\,\mathcal{F}(u,x) &= u(x) - u_0(x) - \alpha\Delta u(x) \\ &\qquad - \beta\,\mathrm{div}\Big(\big(w * (\nabla u^\perp \otimes \nabla u^\perp)\big)(x)\nabla u(x)\Big). \end{aligned} \tag{A.20}$$

A.4 Formal gradient of $\mathcal{F}_W$

Let $\mathcal{F}_W : W^{1,p} \to \mathbb{R}$ be defined by,

$$\mathcal{F}_W(u) = \int_\Omega \frac{1}{2}\big(u(x) - u_0(x)\big)^2 + \frac{\alpha}{p}\|A(u,x)\nabla u(x)\|_E^p \, dx \tag{A.21}$$

where $\alpha > 0$ and $u_0 \in L^\infty$, and $\|\cdot\|_E$ is the $\mathbb{R}^2$ euclidean norm. The directional derivative of $\mathcal{F}$ in direction ϕ is then

$$D\mathcal{F}_W(u;\phi) = \big\langle u - u_0, \phi \big\rangle_{L^2(\Omega)} + \alpha D\mathcal{G}(u;\phi) \tag{A.22}$$

where

$$\mathcal{G}(u) = \frac{1}{p}\int_\Omega \big\|A(u,x)\nabla u(x)\big\|_E^p \, dx.$$

From (2.22), recall that

$$\begin{aligned} A(u,x) &= g\Big(\big(K_\rho * (\nabla u_\sigma \otimes \nabla u_\sigma)\big)(x)\Big) = g \circ B(u,x), \\ B(u,x) &= \big(K_\rho * (\nabla u_\sigma \otimes \nabla u_\sigma)\big)(x) \end{aligned}$$

with g defined on diagonalizable matrices A as

$$\begin{aligned} A &= Q\Sigma Q^T, \\ g(A) &= Q h_\gamma(\Sigma) Q^T, \\ h_\gamma(\Sigma) &= \begin{bmatrix} \frac{1}{1+\frac{(\sigma_1-\sigma_2)^2}{\gamma^2}} & 0 \\ 0 & 1 \end{bmatrix}, \end{aligned}$$

where the eigenvalues of Σ are ordered in descending order. We can find that

$$D\mathcal{G}(u;\phi) = \frac{1}{p}\frac{p}{2}\int_\Omega \Big(\big\|A(u,x)\nabla u(x)\big\|_E^2\Big)^{p/2-1} t'(u,x;\phi)\, dx$$

with

$$
\begin{aligned}
t'(u,x;\phi) &= \Big\langle A'(u,x;\phi)\nabla u(x), A(u,x)\nabla u(x)\Big\rangle_{\mathbb{R}^2} \\
&\quad + \Big\langle A(u,x)\nabla\phi(x), A(u,x)\nabla u(x)\Big\rangle_{\mathbb{R}^2} \\
&\quad + \Big\langle A(u,x)\nabla u(x), A'(u,x;\phi)\nabla u(x)\Big\rangle_{\mathbb{R}^2} \\
&\quad + \Big\langle A(u,x)\nabla u(x), A(u,x)\nabla\phi(x)\Big\rangle_{\mathbb{R}^2} \\
&= 2\Big(\Big\langle A'(u,x;\phi)\nabla u(x), A(u,x)\nabla u(x)\Big\rangle_{\mathbb{R}^2} \\
&\quad + \Big\langle A(u,x)\nabla\phi(x), A(u,x)\nabla u(x)\Big\rangle_{\mathbb{R}^2}\Big) \\
&=: 2h(u,x;\phi),
\end{aligned}
$$

which when put into $D\mathcal{G}(u;\phi)$ gives

$$
D\mathcal{G}(u;\phi) = \int_\Omega \Big(\|A(u,x)\nabla u(x)\|_E^2\Big)^{p/2-1} h(u,x;\phi)\,dx. \tag{A.23}
$$

The next step is to make sense of all terms in $h(u,x;\phi)$.

A.4.1 Computation of $h(u,x;\phi)$

We have

$$
\begin{aligned}
h(u,x,\phi) &= \Big\langle A'(u,x;\phi)\nabla u(x), A(u,x)\nabla u(x)\Big\rangle_{\mathbb{R}^2} \\
&\quad + \Big\langle A(u,x)\nabla\phi(x), A(u,x)\nabla u(x)\Big\rangle_{\mathbb{R}^2}.
\end{aligned} \tag{A.24}
$$

Both $\nabla u(x)$ and $\nabla\phi(x)$ are sensible, and $A(u,x)$ was described section 2.3 and recalled in eq. (2.22). It remains to make sense of $A'(u,x;\phi)$, and we now seek to do so. Applying standard differentiation rules,

$$
\begin{aligned}
A'(u,x;\phi) &= \big(g'\circ B(u,x)\big)\cdot B'(u,x;\phi) \\
&= \big(g'\circ B(u,x)\big)\cdot K_\rho * (\nabla u_\sigma\otimes\nabla u_\sigma)'_\phi(x) \\
&= \big(g'\circ B(u,x)\big)\cdot \Big(K_\rho * \big((\nabla u_\sigma\otimes\nabla\phi_\sigma) + (\nabla\phi_\sigma\otimes\nabla u_\sigma)\big)\Big)(x) \\
&= \big(g'\circ B(u,x)\big)\cdot C(u,x;\phi),
\end{aligned} \tag{A.25}
$$

with $B(u,x)$ as defined in (2.22), and

$$
C(u,x;\phi) = \Big(K_\rho * \big((\nabla u_\sigma\otimes\nabla\phi_\sigma) + (\nabla\phi_\sigma\otimes\nabla u_\sigma)\big)\Big)(x). \tag{A.26}
$$

In (A.25), $(g' \circ B) \cdot C$ is interpreted as the directional derivative of g – defined in (A.4) – in direction of the matrix C, evaluated at the matrix B. That is,

$$\big(g' \circ B\big) \cdot C := \lim_{\epsilon \to 0} \frac{1}{\epsilon} \big[g(B + \epsilon C) - g(B)\big].$$

For the above definition to make sense, both B and C must be diagonalizable, otherwise $g(B + \epsilon C)$ is undefined. For our purposes, both $B(u, x)$ and $C(u, x)$ will always be symmetric by construction, hence they will also be diagonalizable.

Remark. The diagonalization of $B(u, x)$ is not unique whenever $\sigma_1 = \sigma_2$. From appendix A.1, we find that σ_1 and σ_2, the eigenvalues of $B(u, x)$ are

$$\sigma_i = K_\rho * (\partial_1 u)^2 + K_\rho * (\partial_2 u)^2 \pm \sqrt{[K_\rho * (\partial_1 u)^2 - K_\rho * (\partial_2 u)^2]^2 + 4[K_\rho * (\partial_1 u \partial_2 u)]^2}.$$

Note that this equation is a function of x. We see that $\sigma_1 = \sigma_2$ when the root vanishes, that is, when

$$[K_\rho * (\partial_1 u)^2 - K_\rho * (\partial_2 u)^2]^2 + 4[K_\rho * (\partial_1 u \partial_2 u)]^2 = 0.$$

The only way this can be true is if $K_\rho * (\partial_1 u \partial_2 u) = 0$ and $K_\rho * (\partial_1 u)^2 - K_\rho * (\partial_2 u)^2 = 0$ simultaneously. Qualitatively speaking, this this happens in areas where $\partial_1 u \cdot \partial_2 u$ averages to 0 relative to the kernel K_ρ, and $(\partial_1 u)^2$ on average equals $(\partial_2)^2$, also relative to the kernel K_ρ. These issues occur when there are either multiple conflicting dominant directions, for instance in a corner, or no dominant direction at all, for near-constant locales, throughout an area of radius ρ.

Define $J(\epsilon) = g(B + \epsilon C)$ and define $P(\epsilon)$ and $\Lambda(\epsilon)$ by $B + \epsilon C := P(\epsilon)\Lambda(\epsilon)P(\epsilon)^T$. Then

$$\big(g' \circ B\big) \cdot C = J'(0),$$

where

$$\begin{aligned} J'(\epsilon) &= \frac{\partial}{\partial \epsilon} g\Big(P(\epsilon)\Lambda(\epsilon)P(\epsilon)^T\Big) \\ \Rightarrow J'(0) &= \partial_{\epsilon=0} P(\epsilon) h_\gamma\big(\Lambda(0)\big) P(0)^T \\ &\quad + P(0) \partial_{\epsilon=0} h_\gamma\big(\Lambda(\epsilon)\big) P(0)^T \\ &\quad + P(0) h_\gamma\big(\Lambda(0)\big) \partial_{\epsilon=0} P(\epsilon)^T. \end{aligned} \tag{A.27}$$

It should be clear that $\Lambda(0) = \Sigma$ and $P(0) = Q$, hence the only unknowns in the equation above are $\partial_{\epsilon=0} P(\epsilon)$ and $\partial_{\epsilon=0} h_\gamma\big(\Lambda(\epsilon)\big)$. We will find these matrices step-by-step, first in the case where B is already diagonal, and then for arbitrary symmetric matrices B.

Case I) B **is diagonal**

When B is diagonal, the diagonalization of B is semi-trivial:

$$B = \begin{bmatrix} b_1 & 0 \\ 0 & b_2 \end{bmatrix} = Q \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} Q^T = \Sigma = Q\Sigma Q^T,$$

with

$$Q = \begin{cases} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \text{ if } b_1 \geq b_2 \\ \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \text{ if } b_2 < b_1 \end{cases}.$$

These matrices are equivalent to no rotation, or rotation by $\pi/2$. With this, we guarantee that $\sigma_1 \geq \sigma_2$, and thus that Σ is indeed in descending order. Next, we look to diagonalize the matrix

$$B + \epsilon C = \begin{bmatrix} \sigma_1 + \epsilon c_{11} & \epsilon c_{12} \\ \epsilon c_{12} & \sigma_2 + \epsilon c_{22} \end{bmatrix},$$

whose eigenvalues λ are the solutions of

$$\begin{aligned} &\det\big(\lambda I - B - \epsilon C\big) = 0 \\ &\qquad \Rightarrow \lambda(\epsilon) = \frac{1}{2}\Bigg((\sigma_1 + \epsilon c_{11}) + (\sigma_2 + \epsilon c_{22}) \\ &\qquad\qquad \pm \sqrt{\big[(\sigma_1 + \epsilon c_{11}) - (\sigma_2 + \epsilon c_{22})\big]^2 - 4\epsilon^2 c_{12}^2} \Bigg). \end{aligned}$$

Differentiating λ with respect to ϵ yields

$$\lambda'(\epsilon) = \frac{1}{2}(c_{11} + c_{22}) \pm \frac{\Big[\big((\sigma_1 + \epsilon c_{11}) - (\sigma_2 + \epsilon c_{22})\big)(c_{11} - c_{22}) - 4\epsilon c_{12}^2\Big]}{2\sqrt{\big[(\sigma_1 + \epsilon c_{11}) - (\sigma_2 + \epsilon c_{22}\big)^2 - 4\epsilon^2 c_{12}^2}}$$

which evaluated at $\epsilon = 0$, whenever $\sigma_1 \neq \sigma_2$, reduces to

$$\begin{aligned} \lambda'(0) &= \frac{1}{2}\big(c_{11} + c_{22} \pm (c_{11} - c_{22})\big) \\ &= \begin{cases} c_{11}, \\ c_{22}. \end{cases} \end{aligned} \tag{A.28}$$

Equation (A.28) tells us that the immediate changes in eigenvalues of B in direction of the symmetric matrix C depend only in the diagonal entries of C, and also that $\lambda_1'(0) = \partial_C \sigma_1 = c_{11}$ and $\lambda_2'(0) = \partial_C \sigma_2 = c_{22}$. Furthermore, we know that

$$\begin{aligned} P(\epsilon) &= \begin{bmatrix} \cos\big(\alpha(\epsilon)\big) & -\sin\big(\alpha(\epsilon)\big) \\ \sin\big(\alpha(\epsilon)\big) & \cos\big(\alpha(\epsilon)\big) \end{bmatrix}, \\ \alpha(\epsilon) &= \arctan\left(\frac{\lambda_1(\epsilon) - (\sigma_1 + \epsilon c_{11})}{\epsilon c_{12}}\right). \end{aligned} \tag{A.29}$$

As it turns out, $\partial_{\epsilon=0} Q(\epsilon)$ vanishes at 0. We show this by simple calculus rules and expanding $\alpha(\epsilon)$ around 0, giving that

$$\partial_{\epsilon=0} P(\epsilon) = \partial_\alpha P(0) \partial_{\epsilon=0} \alpha(\epsilon)$$

with the Taylor expansion of α being

$$\begin{aligned} \alpha(\epsilon) &= \alpha(0) + \alpha'(0)\epsilon + \mathcal{O}(\epsilon^2) \\ &= \alpha(0) + \partial_{\epsilon=0}\Big[\arctan\Big(\frac{\lambda_1(\epsilon) - (\sigma_1 + \epsilon c_{11})}{\epsilon c_{12}}\Big)\Big]\epsilon + \mathcal{O}(\epsilon^2) \\ &= \alpha(0) + \partial_{\epsilon=0}\big[\frac{\lambda_1(\epsilon) - (\sigma_1 + \epsilon c_{11})}{\epsilon c_{12}}\big]\epsilon + \mathcal{O}(\epsilon^2) \\ &= \alpha(0) + \frac{\epsilon\Big(\partial_{\epsilon=0}\lambda_1(\epsilon) - c_{11}\Big) - \big(\lambda_1(0) - \sigma_1\big)}{\epsilon^2 c_{12}} + \mathcal{O}(\epsilon^2) \\ &= \alpha(0) + \mathcal{O}(\epsilon^2), \end{aligned}$$

where in the last equality we have used that $\lambda_1'(0) = c_{11}$(as was the result of (A.28)) and $\lambda_1(0) = \sigma_1$. Differentiating α with respect to ϵ now easily shows us that

$$\partial_\epsilon \alpha(\epsilon) = \mathcal{O}(\epsilon) \overset{\epsilon=0}{=} 0,$$

ultimately telling us that $\partial_{\epsilon=0} P(\epsilon) = 0$. Recalling (A.27), this simplifies the equation significantly, reducing $J'(0)$ to

$$J'(0) = P(0)\partial_{\epsilon=0} g\big(\Lambda(\epsilon)\big) P(0)^T \overset{P(0)=Q}{=} Q \partial_{\epsilon=0} h_\gamma\big(\Lambda(\epsilon)\big) Q^T.$$

Next, we observe that

$$\begin{aligned}
\partial_{\epsilon=0} h_\gamma(\Lambda(\epsilon)) &= \lim_{\epsilon\to 0} \frac{1}{\epsilon}\Big[h_\gamma(\Lambda(\epsilon)) - h_\gamma(\Lambda(0))\Big] \\
&= \lim_{\epsilon\to 0} \frac{1}{\epsilon}\left(\begin{bmatrix} \frac{1}{1+\left(\frac{(\lambda_1(\epsilon)-\lambda_2(\epsilon)}{\gamma}\right)^2} & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} \frac{1}{1+\left(\frac{(\lambda_1(0)-\lambda_2(0)}{\gamma}\right)^2} & 0 \\ 0 & 1 \end{bmatrix}\right) \\
&= \lim_{\epsilon\to 0} \frac{1}{\epsilon}\begin{bmatrix} \left(1+\left(\frac{(\lambda_1(\epsilon)-\lambda_2(\epsilon)}{\gamma}\right)^2\right)^{-1} - \left(1+\left(\frac{(\lambda_1(0)-\lambda_2(0)}{\gamma}\right)^2\right)^{-1} & 0 \\ 0 & 0 \end{bmatrix} \\
&= \begin{bmatrix} \partial_{\epsilon=0}\tau(\epsilon) & 0 \\ 0 & 0 \end{bmatrix},
\end{aligned}$$

where

$$\tau(\epsilon) = \left[1 + \Big(\frac{\lambda_1(\epsilon) - \lambda_2(\epsilon)}{\gamma}\Big)^2\right]^{-1}.$$

Differentiating and evaluating $\tau(\epsilon)$ at $\epsilon = 0$ gives

$$\begin{aligned}
\partial_{\epsilon=0}\tau(\epsilon) &= -\left[1 + \Big(\frac{\lambda_1(0) - \lambda_2(0)}{\gamma}\Big)^2\right]^{-2} \cdot 2\Big(\frac{\lambda_1(0) - \lambda_2(0)}{\gamma}\Big)\big(\lambda_1'(0) - \lambda_2'(0)\big) \\
&= -\left[1 + \Big(\frac{\sigma_1 - \sigma_2}{\gamma}\Big)^2\right]^{-2} \cdot 2\Big(\frac{\sigma_1 - \sigma_2}{\gamma}\Big)(c_{11} - c_{22}) \\
&= \frac{2(c_{22} - c_{11})\Big(\frac{\sigma_1-\sigma_2}{\gamma}\Big)}{\left[1 + \Big(\frac{\sigma_1-\sigma_2}{\gamma}\Big)^2\right]^2}.
\end{aligned}$$

With this, we can compute $J'(0)$ with only the eigenvalues of B, the parameter γ and the direction C:

$$\begin{aligned}
J'(0) &= (g' \circ B) \cdot C = Q\Sigma_\tau Q^T, \\
\Sigma_\tau &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \frac{2(c_{22} - c_{11})\Big(\frac{\sigma_1-\sigma_2}{\gamma}\Big)}{\left[1 + \Big(\frac{\sigma_1-\sigma_2}{\gamma}\Big)^2\right]^2} \\
&= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \beta(u, x; \gamma)(c_{22} - c_{11}),
\end{aligned} \tag{A.30}$$

where

$$\beta(u, x; \gamma) = \frac{2\left(\frac{\sigma_1 - \sigma_2}{\gamma}\right)}{\left[1 + \left(\frac{\sigma_1 - \sigma_2}{\gamma}\right)^2\right]^2}.$$

Recall at this moment that B is diagonal and that C is an arbitrary symmetric matrix.

Case II) B **is arbitrary symmetric**

In this case, let $B = Q\Sigma Q^T$. Note that since Q is orthogonal, we have for any matrix C the identity $C = Q(Q^T C Q)Q^T = QDQ^T$, where $D := Q^T C Q$. In words, D is the matrix C presented in the eigenbasis of B. Let us now define

$$\begin{aligned} G(\epsilon) &= g(B + \epsilon C) \\ &= g(Q\Sigma Q^T + \epsilon Q(Q^T C Q)Q^T) \\ &\overset{!}{=} Qg(\Sigma + \epsilon D)Q^T \\ &= Q\hat{J}(\epsilon)Q^T, \end{aligned} \tag{A.31}$$

with $\hat{J}(\epsilon) = g(\Sigma + \epsilon D)$. In the third equality of (A.31), we have used that g is invariant under orthogonal similarity transformations. It can now be shown, completely analogously to eqs. (A.29) through (A.30), that

$$G'(0) = Q\partial_{\epsilon=0}\hat{J}'(0)Q^T,$$

where $\hat{J}'(0) = (g' \circ \Sigma) \cdot D$. This yields the directional derivative

$$G'(0) = Q\big[(g' \circ \Sigma) \cdot D\big]Q^T. \tag{A.32}$$

Inserting

$$B = B(u, x) = \big(K_\rho * (\nabla u_\sigma \otimes \nabla u_\sigma)\big)(x) = Q\Sigma Q^T$$

and

$$D = Q^T C(u, x; \phi)Q = Q^T\Big(K_\rho * \big((\nabla u_\sigma \otimes \nabla \phi_\sigma) + (\nabla \phi_\sigma \otimes \nabla u_\sigma)\big)\Big)(x)Q$$

into (A.32) gives us $A'(u, x; \phi)$ from (A.25), which gives us everything we need to compute the directional derivative $D\mathcal{G}(u; \phi)$. Noting that $A(u, x) = Q\Sigma Q^T$, this explicitly yields

$$
\begin{aligned}
D\mathcal{G}(u;\phi) &= \int_\Omega M(u,x)h(u,x;\phi)\,dx,\\
M(u,x) &= \left(\|A(u,x)\nabla u(x)\|_E^2\right)^{\frac{p}{2}-1},\\
h(u,x;\phi) &= \Big\langle g'\circ B(u,x)\cdot D(u,x;\phi), f(u,x)\Big\rangle_{\mathbb{R}^2}\\
&\qquad + \Big\langle A(u,x)\nabla\phi(x), f(u,x)\Big\rangle_{\mathbb{R}^2},\\
&= \Big\langle g'\circ B(u,x)\cdot D(u,x;\phi), f(u,x)\Big\rangle_{\mathbb{R}^2}\\
&\qquad + \Big\langle \nabla\phi(x), A(u,x)^T f(u,x)\Big\rangle_{\mathbb{R}^2},\\
f(u,x) &= A(u,x)\nabla u(x),\\
D(u,x;\phi) &= Q^T C(u,x;\phi)Q,
\end{aligned}
\tag{A.33}
$$

with $B(u,x)$ and $C(u,x;\phi)$ as defined in (2.22) and (A.26).

A.4.2 Formal gradient

The directional derivative $D\mathcal{F}_W(u;\phi)$ described in the previous section can under certain conditions be interpreted as the $L^2(\Omega)$ inner product of ϕ and some function grad $\mathcal{F}_W$. In this section, we seek to find the function grad $\mathcal{F}_W$, called the gradient of $\mathcal{F}_W$, and the conditions under which this interpretation is valid.

Recall eqs. (A.22) and (A.33), that is,

$$
\begin{aligned}
D\mathcal{F}_W(u;\phi) &= \int_\Omega \Big[\big(u(x)-u_0(x)\big)\phi(x)\\
&\qquad + \alpha M(u,x)h(u,x;\phi)\Big]dx\\
&= \big\langle u-u_0,\phi\big\rangle_{L^2(\Omega)}\\
&\qquad + \alpha\int_\Omega M(u,x)\langle(\nabla\phi(x), A(u,x)^T f(u,x)\rangle_{\mathbb{R}^2}dx\\
&\qquad + \alpha\int_\Omega M(u,x)\langle A'(u,x;\phi)\nabla u(x), f(u,x)\rangle_{\mathbb{R}^2}dx.
\end{aligned}
\tag{A.34}
$$

In both the first and second term of the last equality, we have managed to get the dependency on ϕ expressed explicitly. This is not so straightforward with the last integral, which we now take a closer look at. Beginning with $A('(u,x;\phi)$, following the results of section 2.5.2, we find that

$$
\begin{aligned}
A'(u,x;\phi) &= g' \circ B(u,x) \cdot D(u,x;\phi),\\
B(u,x) &= K_\rho * \Big(\nabla u_\sigma \otimes \nabla u_\sigma\Big)(x),\\
&= Q\Sigma Q^T,\\
Q &= \begin{bmatrix}\cos(\theta) & -\sin(\theta)\\ \sin(\theta) & \cos(\theta)\end{bmatrix} =: \begin{bmatrix}a & -b\\ b & a\end{bmatrix},\\
D(u,x;\phi) &= Q^T C(u,x;\phi) Q,\\
C(u,x;\phi) &= \nabla u_\sigma \otimes \nabla \phi_\sigma + \nabla \phi_\sigma \otimes \nabla u_\sigma.
\end{aligned}
$$

As we progress, note in the above that Q and Σ are both functions of u and x. From (A.25), we have

$$
g' \circ B(u,x) \cdot D(u,x,\phi) = H_1(u,x)\beta(u,x;\gamma)\big(D_{22}(u,x) - D_{11}(u,x)\big)
$$

with

$$
\begin{aligned}
H_1(u,x) &= Q^T \begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix} Q^T\\
&= \begin{bmatrix}a^2 & ab\\ ab & b^2\end{bmatrix},\\
D(u,x;\phi) &= \begin{bmatrix}D_{11} & D_{12}\\ D_{21} & D_{22}\end{bmatrix}(u,x).
\end{aligned}
$$

From the definition, we have that $D(u,x;\phi) = Q^T C(u,x;\phi) Q$. We express $C(u,x;\phi)$ and Q in matrix form as

$$
\begin{aligned}
C(u,x;\phi) &= K_\rho * \begin{bmatrix}2\partial_1 u_\sigma \partial_1 \phi_\sigma & \partial_1 u_\sigma \partial_2 \phi_\sigma + \partial_2 u_\sigma \partial_1 \phi_\sigma\\ \partial_2 u_\sigma \partial_1 \phi_\sigma + \partial_1 u_\sigma \partial_2 \phi_\sigma & 2\partial_2 u_\sigma \partial_2 \phi_\sigma\end{bmatrix}(x)\\
&= \begin{bmatrix}K_\rho * [2\partial_1 u_\sigma \partial_1 \phi_\sigma](x) & K_\rho * [\partial_1 u_\sigma \partial_2 \phi_\sigma + \partial_2 u_\sigma \partial_1 \phi_\sigma](x)\\ K_\rho * [\partial_2 u_\sigma \partial_1 \phi_\sigma + \partial_1 u_\sigma \partial_2 \phi_\sigma](x) & K_\rho * [2\partial_2 u_\sigma \partial_2 \phi_\sigma](x)\end{bmatrix}\\
&:= \begin{bmatrix}c_1 & c_2\\ c_2 & c_3\end{bmatrix},
\end{aligned}
$$

Note at this point that a, b and c_i are all functions of u and x (and ϕ), though for the sake of readability, we will for now not explicitly write this down. Continuing from above, straightforward matrix multiplications then yield

$$\begin{aligned}
D(u,x;\phi) = Q^T C(u,x;\phi) Q &= \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \begin{bmatrix} c_1 & c_2 \\ c_2 & c_3 \end{bmatrix} \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \\
&= \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} ac_1 + bc_2 & -bc_1 + ac_2 \\ ac_2 + bc_3 & -bc_2 + ac_3 \end{bmatrix} \\
&= \begin{bmatrix} a^2c_1 + abc_2 + abc_2 + b^2c_3 & -abc_1 + a^2c_2 - b^2c_2 + abc_3 \\ -abc_1 - b^2c_2 + a^2c_2 + abc_3 & b^2c_1 - abc_2 - abc_2 + a^2c_3 \end{bmatrix} \\
&= \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}.
\end{aligned}$$

This means that

$$D_{22} - D_{11} = (a^2 - b^2)(c_3 - c_1) - 4abc_2.$$

Furthermore,

$$\begin{aligned}
c_3 - c_1 &= 2K_\rho * [\partial_2 u_\sigma \partial_2 \phi_\sigma - \partial_1 u_\sigma \partial_1 \phi_\sigma](x) \\
&= 2K_\rho * \left\langle (H_2 \nabla u_\sigma), \nabla \phi_\sigma \right\rangle_{\mathbb{R}^2}(x) \\
&= 2K_\rho * \left\langle (H_2 \nabla u_\sigma), K_\sigma * \nabla \phi \right\rangle_{\mathbb{R}^2}(x),
\end{aligned}$$

where

$$H_2 = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix},$$

and

$$\begin{aligned}
c_2 &= K_\rho * [\partial_1 u_\sigma \partial_2 \phi_\sigma + \partial_2 u_\sigma \partial_1 \phi_\sigma](x) \\
&= K_\rho * \left\langle (H_3 \nabla u_\sigma), K_\sigma * \nabla \phi \right\rangle_{\mathbb{R}^2}(x),
\end{aligned}$$

where

$$H_3 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

In this notation, we can currently conclude that

$$\begin{aligned}
D_{22} - D_{11} = 2(a^2 - b^2) K_\rho * \left\langle (H_2 \nabla u_\sigma), K_\sigma * \nabla \phi \right\rangle_{\mathbb{R}^2}(x) \\
- 4ab K_\rho * \left\langle (H_3 \nabla u_\sigma), K_\sigma * \nabla \phi \right\rangle_{\mathbb{R}^2}(x).
\end{aligned}$$

This, in turn, tells us that

$$\begin{aligned} A'(u,x;\phi) = 2H_1(u,x)\beta(u,x)\Big((a^2-b^2)K_\rho * \Big\langle (H_2\nabla u_\sigma), K_\sigma * \nabla\phi\Big\rangle_{\mathbb{R}^2}(x) \\ - 2abK_\rho * \Big\langle (H_3\nabla u_\sigma), K_\sigma * \nabla\phi\Big\rangle_{\mathbb{R}^2}\Big). \end{aligned} \tag{A.35}$$

This form of $A'(u,x;\phi)$ is explicit in ϕ. We now put this result back into the last term of (A.34), giving the full integral

$$\begin{aligned} &\alpha\int_\Omega M(u,x)\big\langle A'(u,x;\phi)\nabla u(x), f(u,x)\big\rangle_{\mathbb{R}^2}dx \\ =&\alpha\int_\Omega M(u,x)\Big\langle H_1(u,x)\beta(u,x)\Big[2(a^2-b^2)\Big(K_\rho * \Big\langle H_2\nabla u_\sigma, K_\sigma * \nabla\phi\Big\rangle_{\mathbb{R}^2}\Big)(x) \\ &\quad - 4abK_\rho * \Big\langle (H_3\nabla u_\sigma), K_\sigma * \nabla\phi\Big\rangle_{\mathbb{R}^2}(x)\Big]\nabla u(x), f(u,x)\Big\rangle_{\mathbb{R}^2}dx \\ =&\alpha\int_\Omega N(u,x)\Big[(a^2-b^2)\big(K_\rho * \Big\langle H_2\nabla u_\sigma, K_\sigma * \nabla\phi\Big\rangle_{\mathbb{R}^2}\big)(x) \\ &\quad - 2abK_\rho * \Big\langle (H_3\nabla u_\sigma), K_\sigma * \nabla\phi\Big\rangle_{\mathbb{R}^2}(x)\Big]dx \end{aligned}$$

where

$$N(u,x) = 2M(u,x)\beta(u,x)\big\langle H_1(u,x)\nabla u(x), f(u,x)\big\rangle.$$

In the integrals above, $\nabla\phi$ does not show up outside convolutions. As such, to proceed towards our goal, we must now find some adjoint form of the integrand. Proceeding, we again remind the reader that a and b are functions of u and x. We write out convolutions in full, giving

$$\begin{aligned} &\alpha\int_\Omega M(u,x)\big\langle A'(u,x;\phi)\nabla u(x), f(u,x)\big\rangle_{\mathbb{R}^2}dx \\ =&\alpha\int_\Omega\int_\Omega\int_\Omega N(u,x)\Big[(a^2-b^2)(x)\Big(K_\rho(y)K_\sigma(z)\Big\langle H_2\nabla u_\sigma(x-y), \nabla\phi(x-y-z)\Big\rangle_{\mathbb{R}^2}\Big) \\ &\quad - 2ab(x)K_\rho(y)K_\sigma(z)\Big\langle (H_3\nabla u_\sigma(x-y)), \nabla\phi(x-y-z)\Big\rangle_{\mathbb{R}^2}\Big]dz\,dy\,dx. \end{aligned}$$

Substituting $\tilde{x} = x-y-z$, $\tilde{y} = -y$ and $\tilde{z} = -z$, we arrive at

$$
\begin{aligned}
&\alpha \int_\Omega M(u,x)\big\langle A'(u,x;\phi)\nabla u(x), f(u,x)\big\rangle_{\mathbb{R}^2} dx \\
=&\alpha \int_\Omega \int_\Omega \int_\Omega N(u,\tilde{x}-\tilde{y}-\tilde{z})\Big[(a^2-b^2)(\tilde{x}-\tilde{y}-\tilde{z}) \\
&\quad \cdot \Big(K_\rho^*(\tilde{y})K_\sigma^*(\tilde{z})\Big\langle H_2\nabla u_\sigma(\tilde{x}-\tilde{z}), \nabla\phi(\tilde{x})\Big\rangle_{\mathbb{R}^2}\Big) \\
&\quad - 2ab(\tilde{x}-\tilde{y}-\tilde{z})K_\rho^*(\tilde{y})K_\sigma^*(\tilde{z})\Big\langle (H_3\nabla u_\sigma(\tilde{x}-\tilde{z})), \nabla\phi(\tilde{x})\Big\rangle_{\mathbb{R}^2}\Big] d\tilde{z}\, d\tilde{y}\, d\tilde{x},
\end{aligned}
$$

where we have introduced the adjoint convolution kernels K_ρ^* and K_σ^*, defined by

$$K_i^*(x) = -K_i(-x).$$

After some consideration of the integral above, we can recognize it as

$$
\begin{aligned}
&\alpha \int_\Omega M(u,x)\big\langle A'(u,x;\phi)\nabla u(x), f(u,x)\big\rangle_{\mathbb{R}^2} dx \\
=&\alpha \int_\Omega \Big\langle K_\sigma^* * \Big[H_2\nabla u_\sigma K_\rho^* * \Big(N(u,\cdot)(a^2-b^2)\Big) \\
&\qquad - 2H_3\nabla u_\sigma K_\rho^*\Big(N(u,\cdot)(ab)\Big)\Big](x), \nabla\phi(x)\Big\rangle_{\mathbb{R}^2} dx.
\end{aligned}
$$

We now define

$$
\begin{aligned}
T(u,x) = K_\sigma^* * \Big[&H_2\nabla u_\sigma K_\rho^* * \Big(N(u,\cdot)(a^2-b^2)\Big) \\
&- 2H_3\nabla u_\sigma K_\rho^* * \Big(N(u,\cdot)(ab)\Big)\Big](x)
\end{aligned}
\tag{A.36}
$$

yielding the condensed

$$
\begin{aligned}
&\alpha \int_\Omega M(u,x)\big\langle A'(u,x;\phi)\nabla u(x), f(u,x)\big\rangle_{\mathbb{R}^2} dx \\
=&\alpha \int_\Omega \Big\langle T(u,x), \nabla\phi(x)\Big\rangle_{\mathbb{R}^2} dx \\
=&\alpha \Big\langle \big\langle T(u,\cdot), \nabla\phi\big\rangle_{\mathbb{R}^2}\Big\rangle_{L^2(\Omega)}
\end{aligned}
$$

At this point, we recall the directional derivative

$$
\begin{aligned}
D\mathcal{F}_W(u;\phi) = \big\langle u - u_0, \phi \big\rangle_{L^2(\Omega)} \\
&+ \alpha \int_\Omega M(u,x) \langle (\nabla\phi(x), A(u,x)^T f(u,x) \rangle_{\mathbb{R}^2} dx \\
&+ \alpha \int_\Omega M(u,x) \langle A'(u,x;\phi)\nabla u(x), f(u,x) \rangle_{\mathbb{R}^2} dx \\
&= \big\langle u - u_0, \phi \big\rangle_{L^2(\Omega)} + \alpha \int_\Omega \big\langle T(u,\cdot) + M(u,x)A(u,\cdot)^T f(u,\cdot), \nabla\phi \big\rangle_{\mathbb{R}^2} dx.
\end{aligned}
$$

Integration by parts now yield

$$
\begin{aligned}
D\mathcal{F}_W(u;\phi) = \big\langle u - u_0 - \alpha \mathrm{div}\big(T(u,\cdot) + M(u,x)A(u,\cdot)^T f(u,\cdot)\big), \phi \big\rangle_{L^2(\Omega)} \\
+ \int_{\partial\Omega} \big\langle \mathbf{n}(x), T(u,\cdot) + M(u,x)A(u,x)^T f(u,x) \big\rangle_{\mathbb{R}^2}(x)\phi(x).
\end{aligned}
$$

Above, $\mathbf{n}(x)$ denotes the outward-pointing normal vector to Ω at $x \in \partial\Omega$.By variational arguments, the integral over the boundary vanishes whenever

$$
\big\langle \mathbf{n}(x), T(u,\cdot) + M(u,x)A(u,x)^T f(u,x) \big\rangle_{\mathbb{R}^2} = 0 \text{ on } \partial\Omega \tag{A.37}
$$

Whenever (A.37) holds, $D\mathcal{F}_W(u;\phi)$ reduces to

$$
\begin{aligned}
D\mathcal{F}_W(u;\phi) &= \big\langle u - u_0 - \alpha \mathrm{div}\big(T(u,\cdot) + M(u,\cdot)A(u,\cdot)^T f(u,\cdot)\big), \phi \big\rangle_{L^2(\Omega)} \\
&=: \big\langle \text{grad } \mathcal{F}_W, \phi \big\rangle_{L^2(\Omega)}.
\end{aligned} \tag{A.38}
$$

where we recognize

$$
\text{grad } \mathcal{F}_W(u,\cdot) = u - u_0 - \alpha \mathrm{div}\big(T(u,\cdot) + M(u,\cdot)A(u,\cdot)^T f(u,\cdot)\big) \tag{A.39}
$$

as the formal gradient of $\mathcal{F}_W$. □

Bibliography

Boulanger, J., Elbau, P., Pontow, C., Scherzer, O., 2011. Non-local functionals for imaging. In: Fixed-point algorithms for inverse problems in science and engineering. Vol. 49 of Springer Optim. Appl. Springer, New York, pp. 131–154.
URL http://dx.doi.org/10.1007/978-1-4419-9569-8_8

Buades, A., Coll, B., Morel, J. M., 2005. A review of image denoising algorithms, with a new one. Multiscale Model. Simul. 4 (2), 490–530.
URL http://dx.doi.org/10.1137/040616024

Dacorogna, B., 2008. Direct methods in the calculus of variations, 2nd Edition. Vol. 78 of Applied Mathematical Sciences. Springer, New York.

Fonseca, I., Leoni, G., 2007. Modern methods in the calculus of variations: L^p spaces. Springer Monographs in Mathematics. Springer, New York.

Grasmair, M., Lenzen, F., 2010. Anisotropic total variation filtering. Appl. Math. Optim. 62 (3), 323–339.
URL http://dx.doi.org/10.1007/s00245-010-9105-x

Jähne, B., 2005. Digital Image Processing, 6th Edition. Springer, New York.

Kubińska, E., 2004/05. Approximation of Carathéodory functions and multifunctions. Real Anal. Exchange 30 (1), 351–359.
URL http://projecteuclid.org/euclid.rae/1122482142

Nocedal, J., Wright, S. J., 2006. Numerical optimization, 2nd Edition. Springer Series in Operations Research and Financial Engineering. Springer, New York.

Scherzer, O., Grasmair, M., Grossauer, H., Haltmeier, M., Lenzen, F., 2009. Variational methods in imaging. Vol. 167 of Applied Mathematical Sciences. Springer, New York.

Weickert, J., 1999. Coherence-enhancing diffusion filtering. IJCV 31, 111–127.