Total Variation Minimization via Dual-Based Methods and its Discretization Aspects

Von der Fakultät Mathematik und Physik der Universität Stuttgart zur Erlangung der Würde eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigte Abhandlung

Vorgelegt von

Stephan Hilb

aus Stuttgart

Hauptberichter:	Prof. Dr. Bernard Haasdonk
Mitberichter:	Dr. Andreas Langer
	Prof. Dr. Carola-Bibiane Schönlieb
Tag der mündlichen Prüfung:	18.07.2022

Institut für Angewandte Analysis und Numerische Simulation
der Universität Stuttgart
2023

Contents

Abstract

The total variation has been widely used as a regularizing term in variational image processing methods for its ability to preserve sharp edges. Corresponding denoising models with L^1 or L^2 data terms have previously been analyzed and discretized in various ways, as well as extended to incorporate a linear operator to allow for a wider range of applications, including image reconstruction and analysis.

In this work we aim to analyze, discretize and evaluate one particular, previously proposed, widely applicable total variation model with a combined L^1-L^2 data term using the convex duality principle and thereby extend previous results, as well as improve upon them selectively. In particular, we derive suitable optimization algorithms, adaptive finite element and domain decomposition methods, and apply them to various image processing tasks, including denoising, inpainting and optical flow estimation.

Zusammenfassung

Die Totale Variation wird als Regularisierungsterm oft in der variationellen Bildverarbeitung genutzt, da sie scharfe Kanten zu erhalten vermag. Entsprechende Modelle zur Entrauschung von Bildern mittels eines L^1- oder L^2-Datenterms wurden bereits mit verschiedenen Methoden analysiert und diskretisiert. Die Erweiterung um einen linearen Operator erlaubt darüber hinaus den Einsatz in vielen weiteren Anwendungsgebieten, einschließlich der Bilderkennung und -rekonstruktion.

In der vorliegenden Arbeit wird ein vorgeschlagenes, spezielles, kombiniertes L^1-L^2-Modell, das weitreichende Anwendungen erlaubt, mit Hilfe des Prinzips der konvexen Dualität analysiert, diskretisiert und ausgewertet. Im Zuge dessen erweitern wir bisherige Resultate und verbessern diese stellenweise. Insbesondere leiten wir sowohl geeignete Optimierungsalgorithmen, als auch Finite-Elemente- und Gebietszerlegungsmethoden her und wenden diese in der Bildverarbeitung unter anderem zur Rauschreduktion, Bildrekonstruktion und zur Bestimmung des optischen Flusses an.

Acknowledgements

First and foremost, I thank Andreas Langer for his patience and rigour in guiding me throughout my studies. Without his professional support and encouragement, this work would not have taken place. He has provided grounding and perspective in all of our discussions and thus shaped the way I think about mathematics and the world. I further thank the Institute of Applied Analysis and Numerical Simulation at the University of Stuttgart for my employment, most notably Kunibert Siebert for initially supervising me and Bernard Haasdonk for officially taking over this role.

I am very grateful to Martin Alkämper who showed me the ropes in Stuttgart, kindly assisted me in various tasks, collaborated with me and teached me about the pareto principle for working efficiently, which I still find difficult to apply up to this day. I thank Carola-Bibiane Schönlieb for hosting my research stay in Cambridge and Robert Tovey for invaluable interesting discussions and collaborative work during that time, which I, sadly, did not quite manage to fit into this thesis.

Furthermore, I have received feedback and support from various people, including Michael Eisermann, Fernando Gaspoz, Mišo Gavrilović, Arthur Günthner, Claus-Justus Heine, Birane Kane, Jim Magiera, Björn de Rijk, Stephan Schmid and Alexander Thumm. Discussions with them, no matter how minor, have helped to shape bits and pieces of this work to various degrees and I am very grateful for that.

Finally, I would like to thank the university staff for their help with organizational concerns, friends and family for their emotional support and everyone else who inspired me to stay motivated.

Funding

Part of this work was supported by the Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg (Az: 7533.-30-10/56/1) through the RISC-project ”Automatische Erkennung von bewegten Objekten in hochauflösenden Bildsequenzen mittels neuer Gebietszerlegungsverfahren”.

1 Introduction

1.1 Background and Motivation

Without being constrained by excessive rigour and detail, we would like to take the opportunity in this section to try and provide a generally intelligible introduction to and motivation for our topic of research, namely the total variation and its dual characterization in the context of variational image processing.

1.1.1 Digital Image Processing

Images are widely used to record static visual data. Captured with a camera or created in some other way, they make archiving, processing or distribution of this data possible. Especially with the advent of digital cameras in smartphones, being able to capture and share digital images has today become something which is mostly taken for granted. Compared to analogue images, they may readily become subject to various forms of image processing. Well-known processing tasks include e.g. removal of the flash-induced red-eye effect in photographs and lightning correction of images. For smartphones, in fact, advanced image post processing features have become standard [38].

More specialized processing tasks may include noise removal (*denoising*) or the reconstruction of missing parts of an image (*inpainting*). Even motion detection from an image sequence (*optical flow* estimation) or reconstructing visual data from certain measured signals (*tomography* or *imaging*) may be viewed as image processing tasks. They find application in various fields, including medicine, art conservation and forensics [9].

Recently, machine learning approaches have made promising advances in the field of image processing due to their outstanding ability to learn and generalize from trained data [64]. While their good practical performance is well-recognized, they are generally sensitive to bias in training data, lack a transparent model to explain their output and are often expensive to train. More traditional mathematical modeling, on the other hand, features a transparent pre-defined objective, does not require an expensive training process and can build upon a vast literature of existing mathematical analysis tools. Thus, arguably, mathematical modeling of image processing tasks is still a valuable approach.

1.1.2 The Variational Principle as a Modeling Tool

It is often convenient to describe target solutions of a problem as minimizers of a certain quantity, a property known as the *variational principle* [27]. Just as e.g. the shape of a water droplet on a flat surface is given by minimizing the sum of its potential and surface energy [27], in variational image processing we set out to find a target image, which minimizes a certain quantity of interest, often called *energy*.

For image processing the energy functional of an image u is traditionally composed by a sum of two quantities, the *image term* and the *data term*. The data term describes how well the image corresponds to given input data g, while the image term describes how probable the image is, irrespective of input data. This formulation is often motivated by viewing the image processing task in reverse as a two-step random process, where first, some original image is chosen at random and then a certain random process is applied, to arrive at some observed data corresponding to the image processing task input data. One may then ask for the most probable original image given this observed data. Due to the Bayesian principle, this *posterior probability* $P(u|g)$ to be maximized is proportional to the product of the *prior probability* $P(u)$ of the original image and the *likelihood* $P(g|u)$ of the observation given that original image. Maximizing this product can equivalently be seen as minimizing its negative logarithm, which yields the sum formulation mentioned above.

For denoising in particular, this approach may be used to explicitly derive the data term, given a probabilistic noise model. It is well-known, for example, that the sum of squared errors corresponds to additive Gaussian noise [9, 29], while the sum of absolute errors is to be used for impulse noise [6, 25, 73]. For other applications more sophisticated data terms are used, e.g. [9, 25]. These are not necessarily always explicitly derived from a probabilistic model, but may as well follow some heuristic.

The image term, on the other hand, captures some a-priori knowledge about how the original image is conditioned and thus, in a way, guides the image processing task towards solutions which are favored by it. Since the range of all possible original images is usually unmanageable, instead of trying to capture the full probabilistic model as an image term, again, heuristics are employed. Another role of the image term is to act as a regularizer for the minimization problem, i.e. to supplement the problem with additional information when the data term is not sufficient, e.g. due to missing or uncertain data. As image term one may, for example, choose to penalize image gradients quadratically, which readily yields to minimization techniques and provides smooth solutions [9]. Seeing, however, that images often contain sharp edges and therefore, smoothened solutions are not preferred, the so-called *total variation* as an alternative measure of change for an image over its domain has established itself as a suitable alternative [21].

1.1.3 The Role of Total Variation

Originally used in denoising audio signals (which may be viewed as one-dimensional images), the total variation has found its way into image processing. Its distinguishing effect as an image term is to promote coherent regions in an image, while preserving sharp edges between them.

The total variation of an image, viewed as a real-valued brightness function over a two-dimensional domain, may loosely be defined as the total sum of absolute gradient lengths. Functions with finite total variation may form a complete normed vector space with interesting properties and therefore, have been an object of interest in functional analysis.

Compared to the sum of squared gradient lengths, the total variation, as a function of its input image, is non-smooth, which makes it hard to design minimization algorithms for it. One approach is to consider a related, so-called *dual problem*, which under certain conditions can be shown to be equivalent to the original minimization problem in question.

1.1.4 Duality in Convex Minimization

Loosely speaking, a dual formulation of some problem is given by a changed perspective. For example, instead of maximizing profits in economics, one may instead take the viewpoint of minimizing total production cost. This duality principle is well-known e.g. for linear programs, where in general a linear objective with linear inequality constraints is considered, and can be extended to more general, convex minimization objectives.

For the problem of minimizing a convex, not necessarily smooth function, the field of convex optimization has celebrated the so-called Fenchel duality, which defines a corresponding dual problem and provides sufficient conditions as to when its solution agrees with the primal one [43]. Here, convex constraints may be incorporated directly into the convex objective function by use of a characteristic function, which evaluates to 0 in feasible regions and to positive infinity otherwise.

1.2 Outline and Contributions

We will now provide the reader with an overview of the following chapters and highlight the major contributions of this thesis. Its focus will lean towards the theoretical side and numerical examples are mostly provided as a proof-of-concept.

Beginning in Chapter 2, we cover some well known results from functional analysis and convex optimization, which are essential for our presentation.

In Chapter 3 we extend a variational scalar model originally proposed for image restoration to support vector-valued functions for application to a wider class of problems, including optical flow estimation. A dual formulation as well as optimality conditions for an optionally regularized extended version of this model are derived and existence and uniqueness of both the primal and dual formulation are analyzed in the infinite-dimensional setting. Finally we establish Γ-convergence for the regularized functional to the original one.

In Chapter 4 we extend parallel and sequential versions of an existing decomposition algorithm to a more general pointwise constrained quadratic problem and modify it to incorporate approximate local minimization. For these methods we derive improved theoretical convergence results using a different proof strategy and compare these to existing work.

Chapter 5 considers discretization of the continuous models from Chapters 3 and 4 in the context of image processing. First we propose a new pixel-adapted method to interpolate an image onto an unstructured finite element grid and compare it to other alternatives. For the regularized model from Chapter 3, we derive two different a-posteriori error estimates, formulate a semi-smooth Newton algorithm and evaluate their performance numerically in an adaptive finite element setting for various applications. Lastly, we apply the decomposition algorithm given in Chapter 4 to various image processing tasks, numerically verify the theoretical convergence bounds and evaluate the performance of a parallel implementation.

Chapter 6 gives an overview of the newly developed numerical software, which is available publicly under a permissive license. Apart from being able to reproduce the numerical examples present in Chapter 5, it includes a utility package for handling the optical flow file format, a package for defining custom kernel operations on arrays of arbitrary dimension, a tiny finite element framework and a helper library for managing stateful parallelism.

Finally, Chapter 7 provides a short overview of some interesting areas for further research, mainly concerning discretization and software development.

2 Fundamentals

We assume the reader to be at least somewhat familiar with the basics of functional analysis and finite elements and recommend [4, 34] for further reference. Before analyzing the variational model central to our work in Chapter 3, this chapter will fix notation, cover necessary results from functional analysis and convex optimization, as well as introduce the total variation and its basic properties. The contents here are considered to be common knowledge within their respective domains and contain no particularly new results. We do, however, try to give proofs for statements in this chapter whenever they are reasonably easy, self-contained or educational.

2.1 Functional Analysis

We will denote by $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$ the extended real number line, equipped with its order topology. For a Banach space V the expression V^* denotes the continuous dual space, i.e. the space of bounded linear functionals $V \to \mathbb{R}$, while we use $\langle \cdot \,, \cdot \rangle_{V,V^*}$ for the duality pairing. A sequence $(v_j)_{j\in\mathbb{N}} \subseteq V$ is called *(weakly) V-convergent*, if it converges (weakly) in the space V. For a bounded linear operator $A : V \to W$ between two Banach spaces V and W we use $\|A\|$ for the operator norm and denote the adjoint operator by $A^* : W^* \to V^*$. The inner product of an inner product space V is generally written as $\langle \cdot \,, \cdot \rangle_V$ and correspondingly $\| \cdot \|_V$ will denote its induced norm. For finite dimensional V specifically, we use single strokes for the norm: $| \cdot |_V$. For standard Euclidean spaces $\mathbb{R}^n, n \in \mathbb{N}$, we may omit the norm subscript, $| \cdot |$, to denote the standard Euclidean norm.

Throughout this work, $\Omega \subseteq \mathbb{R}^d$ with domain dimension $d \in \mathbb{N}$ will denote a connected bounded open set with Lipschitz boundary as in e.g. [34, Ch. 1, §1.2]. The expression "a.e. in Ω" denotes a condition that holds pointwise almost everywhere in Ω, i.e. everywhere up to a subset of Lebesgue measure zero. We use the notation $L^2(\Omega)^n$, $n \in \mathbb{N}$ to denote the space of square-integrable vector-valued functions, accompanied with the inner product $\langle \cdot, \cdot \rangle : \big((u_k)_{k=1}^n, (v_k)_{k=1}^n\big) \mapsto \sum_{k=1}^n \langle u_k, v_k \rangle_{L^2(\Omega)}$. Apart from notational convenience, we treat a matrix-valued space $L^2(\Omega)^{d\times m}$, $m \in \mathbb{N}$ as equivalent to $L^2(\Omega)^{dm}$ using any fixed isomorphism. Finally, we may use the shorthand $\langle \cdot, \cdot \rangle := \langle \cdot, \cdot \rangle_{L^2} := \langle \cdot, \cdot \rangle_V$ for any L^2 function space V and similarly $\| \cdot \| := \| \cdot \|_{L^2} := \| \cdot \|_V$ for the norm. Often, operations are applied in a pointwise sense, such that for a vector-valued function $u : \Omega \to \mathbb{R}^m$, $m \in \mathbb{N}$ the expression $|u|$ denotes the function $|u| : \Omega \to \mathbb{R}$, $x \mapsto |u(x)|$. Naturally, we extend the definition such that e.g. $u \in L^p(\Omega)^m$ implies $|u| \in L^p(\Omega)$ for $1 \leq p \leq \infty$.

2.1.1 Banach Spaces

Since the norm and weak topologies for infinite dimensional spaces are different [4, Theorem 6.26], we collect some convenient tools to help us, which are mainly related to weak convergence and convex closed sets.

Lemma 2.1 ([4, Theorem 6.25])**.** *A Banach space V is reflexive if and only if the closed unit ball of V is weakly compact.*

Lemma 2.2. *Let V be a reflexive Banach space and $F : V \to \overline{\mathbb{R}}$ coercive, i.e. for any sequence $(v_n)_{n\in\mathbb{N}} \subseteq V$ we have*

$$\|v_n\|_V \to \infty \implies F(v_n) \to \infty.$$

Then F is weakly coercive with regard to weak convergence, i.e. $\inf_{v\in V} F(v) = \inf_{v\in K} F(v)$ for some sequentially weakly compact set K.

Proof. Since F is coercive, we have $\inf_X F = \inf_K F$ for some sufficiently large ball $\emptyset \neq K \subseteq X$. Then because V is reflexive, K is weakly compact due to Lemma 2.1. □

Lemma 2.3 ([42, Corollary 8.74]). *Let V be a Banach space. A convex set $K \subseteq V$ is closed if and only if it is weakly closed.*

Lemma 2.4 ([4, Theorem 13.6]). *Let $(f_n)_{n\in\mathbb{N}} \subseteq L^p(\Omega)$, $1 \le p \le \infty$, $f_n \to f$ be an $L^p(\Omega)$-convergent sequence. Then there exists $g \in L^p(\Omega)$ and a subsequence $(g_n)_{n\in\mathbb{N}} \subseteq (f_n)_{n\in\mathbb{N}}$ with $|g_n| \le g$ for all $n \in \mathbb{N}$ and $g_n \to f$ pointwise almost everywhere.*

Lemma 2.5. *The set $A := \{f \in L^p(\Omega) : |f| \le \alpha\} \subseteq L^p(\Omega)$, $1 \le p \le \infty$ is (weakly) closed, convex and bounded for any $\alpha \in L^p(\Omega)$.*

Proof. It is easy to see that A is convex by a pointwise consideration of the constraint and bounded in $L^p(\Omega)$ by α. For showing closedness let $(p_n)_{n\in\mathbb{N}} \subseteq A$, $p_n \to p \in L^p(\Omega)$ be a convergent sequence in A. Due to Lemma 2.4 there exists a subsequence $(q_n)_{n\in\mathbb{N}} \subseteq (p_n)_{n\in\mathbb{N}}$ with $q_n \to p$ pointwise almost everywhere. In particular we have $|p| \le \sup_{n\in\mathbb{N}} |q_n| \le \alpha$ almost everywhere and therefore conclude $p \in A$. Finally, using Lemma 2.3 we find that A is weakly closed as well. □

2.1.2 Mollifiers

We briefly review the central properties of *mollifiers*, which are useful to construct smooth function approximants.

Definition 2.6 (Convolution, c.f. [2, page 38]). *For $f, g \in L^1(\mathbb{R}^d)$ we define the* convolution *$f * g \in L^1(\mathbb{R}^d)$ by*

$$(f * g)(x) := \int_{\mathbb{R}^d} f(x-y)g(y)\,\mathrm{d}y.$$

If f or g in Definition 2.6 are vector-valued, then $f * g$ is defined in a component-wise manner. Further, if $f, g \in L^1(\Omega)$ are defined on Ω, then one defines the convolution $f * g \in L^1(\mathbb{R}^d)$ by implicitly extending f and g to $\mathbb{R}^d \supseteq \Omega$ with zero.

Proposition 2.7 (Mollifier, c.f. [2, Theorem 2.29])**.** *Let $\varrho \in C_0^\infty(\mathbb{R}^d)$ be given by*

$$\varrho(x) := \begin{cases} e^{-\frac{1}{1-|x|^2}} & \textit{if } |x| < 1, \\ 0 & \textit{else.} \end{cases}$$

For $\varepsilon > 0$ the mollifier *$\varrho_\varepsilon \in C_0^\infty(\mathbb{R}^d)$ is defined by $\varrho_\varepsilon(x) := \frac{1}{c_\varrho \varepsilon^d} \varrho(\frac{x}{\varepsilon})$, $c_\varrho := \int_{\mathbb{R}^d} \varrho(x) \, \mathrm{d}x$, such that $\int_{\mathbb{R}^d} \varrho_\varepsilon(x) \, \mathrm{d}x = 1$.*

If $u \in L^1_{loc}(\mathbb{R}^d)$, then the mollified function *$u * \varrho_\varepsilon$ satisfies the following properties:*

(i) $\operatorname{supp}(u * \varrho_\varepsilon) \subseteq \operatorname{supp} u + \overline{B_\varepsilon(0)}$,

(ii) $u * \varrho_\varepsilon \in C^\infty(\mathbb{R}^d)$,

(iii) if $u \in C^\infty(\Omega)$, then $\partial^\alpha(u * \varrho_\varepsilon) = (\partial^\alpha u) * \varrho_\varepsilon$ on Ω for $\alpha \in \mathbb{N}_0^d$,

(iv) if $u \in C(\Omega)$, then $u * \varrho_\varepsilon \to u$ pointwise on Ω and uniformly on any compact subset $K \subseteq \Omega$ for $\varepsilon \to 0$,

(v) if $u \in C(\overline{\Omega})$, then $u * \varrho_\varepsilon \to u$ uniformly on Ω for $\varepsilon \to 0$.

(vi) if $u \in L^p(\Omega)$ with $1 \leq p < \infty$, then $\|u * \varrho_\varepsilon\|_{L^p(\Omega)} \leq \|u\|_{L^p(\Omega)}$ and $\|u * \varrho_\varepsilon - u\|_{L^p(\Omega)} \to 0$ for $\varepsilon \to 0$,

(vii) if $u \leq v$ a.e. for some $v \in L^1(\mathbb{R}^d)$, then $u * \varrho_\varepsilon \leq v * \varrho_\varepsilon$.

Proof. It is straightforward to see that ϱ and ϱ_ε are infinitely differentiable and have compact support. We now prove the properties in order, following along the lines of [2, Theorem 2.29].

(i) Having $x \notin \operatorname{supp} u + \overline{B_\varepsilon(0)}$ implies $x - y \notin \operatorname{supp} u$ for all $y \in \overline{B_\varepsilon(0)} = \operatorname{supp} \varrho_\varepsilon$. Therefore

$$\begin{aligned} (u * \varrho_\varepsilon)(x) &= \int_{\mathbb{R}^d} u(x-y) \varrho_\varepsilon(y) \, \mathrm{d}y \\ &= \int_{\operatorname{supp} \varrho_\varepsilon} u(x-y) \varrho_\varepsilon(y) \, \mathrm{d}y = 0, \end{aligned}$$

i.e. $x \notin \operatorname{supp}(u * \varrho_\varepsilon)$.

(ii) Since $\varrho_\varepsilon \in C_0^\infty(\mathbb{R}^d)$ has compact support and $u \in L^1_{\mathrm{loc}}(\Omega)$ we have for $\alpha \in \mathbb{N}_0^d$, $x \in \Omega$ a well-defined derivative

$$\begin{aligned}\partial^\alpha(u * \varrho_\varepsilon)(x) &= \partial_x^\alpha \int_{\mathbb{R}^d} u(x-y)\varrho_\varepsilon(y)\,\mathrm{d}y \\ &= \partial_x^\alpha \int_{\mathbb{R}^d} u(z)\varrho_\varepsilon(x-z)\,\mathrm{d}z \\ &= \int_{\mathbb{R}^d} u(z)\partial_x^\alpha \varrho_\varepsilon(x-z)\,\mathrm{d}z,\end{aligned}$$

where we transformed with $z = x - y$.

(iii) Since $u \in C^\infty(\Omega)$ we have

$$\begin{aligned}\partial^\alpha(u * \varrho_\varepsilon)(x) &= \int_{\mathbb{R}^d} \partial_x^\alpha u(x-y)\varrho_\varepsilon(y)\,\mathrm{d}y \\ &= \int_{\mathbb{R}^d} (\partial^\alpha u)(x-y)\varrho_\varepsilon(y)\,\mathrm{d}y = (\partial^\alpha u * \varrho_\varepsilon)(x).\end{aligned}$$

(iv) Let $\delta > 0$. Since $\int_{\mathbb{R}^d} \varrho_\varepsilon(x)\,\mathrm{d}x = 1$ for any $\varepsilon > 0$, we have for $x \in \Omega$:

$$\begin{aligned}(u * \varrho_\varepsilon - u)(x) &= \int_{\mathbb{R}^d} \big(u(x-y) - u(x)\big)\varrho_\varepsilon(y)\,\mathrm{d}y \\ &\le \sup_{y \in B_\varepsilon(0)} |u(x-y) - u(x)| < \delta \qquad (2.1)\end{aligned}$$

whenever we choose $\varepsilon > 0$ small enough, because u is continuous on Ω. Additionally ε in (2.1) may be chosen uniformly, i.e. independent of x, for all $x \in K$ if K is compact by selecting the smallest such ε for a finite subcover of $\bigcup_{z \in K} B_\varepsilon(z) \supseteq K$.

(v) Since Ω is bounded, $\overline{\Omega} \subseteq \mathbb{R}^d$ is compact and u on $\overline{\Omega}$ uniformly continuous. The statement then follows from (2.1).

(vi) For $1 < p < \infty$ and $1 < p' < \infty$ with $\frac{1}{p} + \frac{1}{p'} = 1$ we get using

Hölder's inequality

$$\begin{aligned} |(u * \varrho_\varepsilon)(x)| &\leq \int_{\mathbb{R}^d} |u(x-y)| \varrho_\varepsilon(y) \, \mathrm{d}y \\ &= \int_{\mathbb{R}^d} |u(x-y)| \varrho_\varepsilon(y)^{\frac{1}{p}} \varrho_\varepsilon(y)^{\frac{1}{p'}} \, \mathrm{d}y \\ &\leq \Big(\int_{\mathbb{R}^d} |u(x-y)|^p \varrho_\varepsilon(y) \Big)^{\frac{1}{p}} \Big(\int_{\mathbb{R}^d} \varrho_\varepsilon(y) \, \mathrm{d}y \Big)^{\frac{1}{p'}} \\ &= \Big(\int_{\mathbb{R}^d} |u(x-y)|^p \varrho_\varepsilon(y) \Big)^{\frac{1}{p}}. \end{aligned} \tag{2.2}$$

Making use of this it follows for $1 \leq p < \infty$:

$$\begin{aligned} \|u * \varrho_\varepsilon\|_{L^p(\Omega)}^p &\leq \int_{\mathbb{R}^d} |(u * \varrho_\varepsilon)(x)|^p \, \mathrm{d}x \\ &\leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |u(x-y)|^p \varrho_\varepsilon(y) \, \mathrm{d}y \, \mathrm{d}x \\ &= \int_{\mathbb{R}^d} |u(z)|^p \int_{\mathbb{R}^d} \varrho_\varepsilon(x-z) \, \mathrm{d}x \, \mathrm{d}z \\ &= \|u\|_{L^p(\Omega)}^p, \end{aligned} \tag{2.3}$$

where for $p = 1$ the inequality (2.2) is not required.

Let now $\delta > 0$. Since $C_0(\Omega) \subseteq L^p(\Omega)$ is dense, we may choose $\varphi \in C_0(\Omega)$ with $\|\varphi - u\|_{L^p(\Omega)} < \frac{\delta}{3}$ and due to (2.3) also $\|\varphi * \varrho_\varepsilon - u * \varrho_\varepsilon\|_{L^p(\Omega)} < \frac{\delta}{3}$ for any $\varepsilon > 0$. Because φ is continuous with compact support, it is uniformly continuous and we may thus choose ε uniformly in (2.1) to get $\|\varphi * \varrho_\varepsilon - \varphi\|_{L^p(\Omega)} < \frac{\delta}{3}$, resulting in

$$\begin{aligned} \|u * \varrho_\varepsilon - u\|_{L^p(\Omega)} &\leq \|u * \varrho_\varepsilon - \varphi * \varrho_\varepsilon\|_{L^p(\Omega)} \\ &\qquad + \|\varphi * \varrho_\varepsilon - \varphi\|_{L^p(\Omega)} + \|\varphi - u\|_{L^p(\Omega)} \\ &< \delta. \end{aligned}$$

(vii) Since ϱ_ε is non-negative one gets $u * \varrho_\varepsilon \leq v * \varrho_\varepsilon$ immediately from the monotonicity of the integral. □

Note that Proposition 2.7 directly generalizes to vector-valued functions in a component-wise manner.

2.1.3 Hilbert Spaces

We review some central statements for real Hilbert spaces, namely the metric projection map onto closed convex sets and the Lax-Milgram theorem to show existence and uniqueness of weak solutions. Further, we define the Sobolev space of bounded weak divergence functions. In this work we only consider Hilbert spaces which are real, as opposed to complex ones, and any use of the term shall implicitly contain this assumption.

Lemma 2.8 (Metric projection, c.f. [4, Theorem 6.53, Lemma 6.54])**.** *Let H be a Hilbert space and $K \subseteq H$ a closed, convex set. Then the projection map onto K, i.e. $\pi_K : H \to K$ such that $\|\pi_K(u) - u\|_H = \inf_{v \in K} \|v - u\|_H$ for all $u \in H$ is well-defined with*

$$\|\pi_K(u) - \pi_K(v)\|_H \leq \|u - v\|_H.$$

In particular, π_K is uniformly continuous.

Theorem 2.9 (Fréchet-Riesz representation, [42, Corollary 3.19])**.** *Let V be a Hilbert space, then the map $\varphi : V \to V^*$, $v \mapsto (w \mapsto \langle v, w\rangle_V)$ is a linear isometric isomorphism.*

Theorem 2.10 (Lax-Milgram, c.f. [34, Theorem 1.1.3, Remark 1.1.3])**.** *Let V be a Hilbert space, $a_B : V \times V \to \mathbb{R}$ a bilinear function and $l \in V^*$. If a_B is both*

(i) continuous, i.e. $|a(v, w)| \leq C_B \|v\|_V \|w\|_V$ for all $v, w \in V$, and

(ii) coercive, i.e. $|a(v, v)| \geq c_B \|v\|_V^2$ for all $v \in V$,

then the problem

$$a(u, v) = l(v) \qquad \forall v \in V,$$

has a unique solution $u \in V$ which depends continuously on l owing to $\|u\|_V \leq c_B^{-1} \|l\|_{V^*}$.

A continuous bilinear function $a_B : V \times V \to \mathbb{R}$ may equivalently be described using the linear operator $B : V \to V^*$ defined by $\langle Bv, w\rangle_{V^*,V} := a_B(v, w)$. Theorem 2.10 then asserts that the inverse $B^{-1} : V^* \to V$ exists and is bounded through $\|B^{-1}v^*\|_V \le c_B^{-1}\|v^*\|_{V^*}$ for all $v^* \in V^*$ where c_B denotes the coercivity constant of a. This is a particularly useful point of view which we will make use of in Chapter 3.

Lemma 2.11 (Poincaré-Wirtinger inequality, c.f. [8, Corollary 5.4.1])**.** *There exists $C > 0$ such that*

$$\|v - \overline{v}\|_{L^2(\Omega)} \le C\|\nabla v\|_{L^2(\Omega)^d}$$

for any $v \in H^1(\Omega)$ and $\overline{v} := \frac{1}{|\Omega|}\int_\Omega v(x)\,\mathrm{d}x$ denoting the mean value of v.

Definition 2.12 ($H^{\operatorname{div}}(\Omega)^m$, c.f. [37])**.** *Let the space of bounded weak divergence functions be defined by $H^{\operatorname{div}}(\Omega)^m := \{v \in L^2(\Omega)^{d\times m} : \operatorname{div} v \in L^2(\Omega)^m\}$, $m \in \mathbb{N}$ with norm*

$$\|u\|^2_{H^{\operatorname{div}}(\Omega)^m} := \|u\|^2_{L^2(\Omega)^{d\times m}} + \|\operatorname{div} u\|^2_{L^2(\Omega)^m},$$

where the operator div *denotes the row-wise weak divergence (in d).*

The space of bounded weak divergence functions with zero normal boundary $H_0^{\operatorname{div}}(\Omega)^m$ is then defined as the closure of $C_0^\infty(\Omega)^{d\times m}$ with regard to $\|\cdot\|_{H^{\operatorname{div}}(\Omega)^m}$.

2.1.4 Γ-Convergence

We may equip functionals with a certain weak notion of convergence, called Γ-convergence [19], which still allows us to make statements about how sequences of minimizers of those functionals behave in the limit. The notion of Γ-convergence and Γ-limit is made precise in the following and we refer to [19] for further reference.

Definition 2.13 (Gamma-convergence, [19, Definition 1.5])**.** *Let V be a metric space. A sequence $(F_j)_{j\in\mathbb{N}}$ of functions $F_j : V \to \overline{\mathbb{R}}$ is said to Γ-converge in V to its Γ-limit $F : V \to \overline{\mathbb{R}}$ (also written $F = \Gamma\text{-}\lim_{j\to\infty} F_j$), if for all $v \in V$ we have* *(i)* $F(v) \leq \liminf_{j\to\infty} F_j(v_j)$ *for every sequence* $(v_j)_{j\in\mathbb{N}} \subseteq V$ *converging to* v*,*

(ii) $F(v) \geq \limsup_{j\to\infty} F_j(v_j)$ *for some sequence* $(v_j)_{j\in\mathbb{N}} \subseteq V$ *converging to* v*.*

Note that the choice of metric (or more generally the notion of convergence) in Definition 2.13 directly affects the Γ-limit. For a constant sequence $(F_j)_{j\in\mathbb{N}}$ of functions $F_j = F : V \to \overline{\mathbb{R}}$ it holds $\Gamma\text{-}\lim_{j\to\infty} F_j = F$ if and only if F is lower semi-continuous [19, Remark 1.8], while in general one has $\Gamma\text{-}\lim_{j\to\infty} F_j \leq \lim_{j\to\infty} F_j$ pointwise [19, Remark 1.10]. If the functions F_j are lower semi-continuous and increasing there is a simple characterization given by the pointwise limit.

Lemma 2.14 (Gamma-limit of increasing sequences, [19, Remark 1.40 (ii)])**.** *Let* V *be a metric space and* $(F_j)_{j\in\mathbb{N}}$ *a sequence of lower semi-continuous functions* $F_j : V \to \overline{\mathbb{R}}$*. If* $(F_j)_{j\in\mathbb{N}}$ *is increasing, i.e.* $F_j(v) \leq F_{j+1}(v)$ *for all* $v \in V$ *and* $j \in \mathbb{N}$*, then the* Γ*-limit is given by the pointwise limit*

$$\big(\Gamma\text{-}\lim_{j\to\infty} F_j\big)(v) = \lim_{j\to\infty} F_j(v)$$

for all $v \in V$*.*

If the sequence of functions F_j are minimized over a common compact subset, then Γ-convergence implies the convergence of those minimizers to the minimum of the Γ-limit.

Theorem 2.15 (Convergence of minimizers, c.f. [19, Theorem 1.21])**.** *Let* V *be a metric space and* $(F_j)_{j\in\mathbb{N}}$ *a sequence of uniformly mildly coercive functions* $F_j : V \to \overline{\mathbb{R}}$*, i.e. there exists a non-empty compact* $K \subseteq X$ *such that* $\inf_{v\in K} F_j = \inf_{v\in V} F_j$ *for all* $j \in \mathbb{N}$*. If* $(F_j)_{j\in\mathbb{N}}$ Γ*-converges to* $F_\infty = \Gamma\text{-}\lim_{j\to\infty} F_j$*, then there exists a minimizer* $\hat{v} \in V$ *of* F_∞ *such that*

$$F_\infty(\hat{v}) = \lim_{j\to\infty} \inf_{v\in V} F_j(v).$$

Further if $(v_j)_{j\in\mathbb{N}}$ *is any sequence satisfying* $\lim_{j\to\infty} F_j(v_j) = \lim_{j\to\infty}\inf_{v\in V} F_j(v)$*, then every limit point of* $(v_j)_{j\in\mathbb{N}}$ *is a minimizer of* F_∞*. In particular if* $(v_j)_{j\in\mathbb{N}}$ *is a sequence with* $F_j(v_j) = \inf_{v\in V} F_j$*, then* $\lim_{j\to\infty} v_j \to \hat{v} \in V$*.*

2.2 Convex Optimization

Minimizing a convex function $F : V \to \overline{\mathbb{R}}$ over some Banach space V will be a recurring theme throughout this thesis. The field of *convex optimization* and, more generally, *convex analysis* provides tools and structure to make statements about existence and uniqueness of such minimizers as well as to formalize optimality conditions and equivalent dual problems. One may as well see convex optimization as a generalization of *smooth* convex optimization which generally assumes differentiable F over finite-dimensional V. We are going to review the basic tools of convex optimization in this section.

The generalization to non-smooth optimization allows to incorporate domain constraints directly within the minimization objective as so-called *indicator functionals* $\chi : V \to \overline{\mathbb{R}}$ which evaluate to ∞ for infeasible $v \in V$ and to zero otherwise.

Definition 2.16 (Indicator functional)**.** *For a boolean value* $w \in \{\text{true}, \text{false}\}$ *we define the indicator* $\chi_w \in \overline{\mathbb{R}}$ *by*

$$\chi_w := \begin{cases} 0 & \textit{if } w \textit{ is true,} \\ \infty & \textit{if } w \textit{ is false,} \end{cases}$$

while for a predicate $w : \Omega \to \{\text{true}, \text{false}\}$ *(e.g.* $|u| \le 1$ *in a pointwise sense) we define* $\chi_w \in \overline{\mathbb{R}}$ *by*

$$\chi_w := \begin{cases} 0 & \textit{if } w(x) \textit{ is true for almost every } x \in \Omega, \\ \infty & \textit{else.} \end{cases}$$

Thus $\chi_{|w|\le 1}$ would evaluate to ∞ if and only if $|w|$ is greater than 1 on a set of non-zero measure.

Allowing convex functions to assume values within the extended real numbers $\overline{\mathbb{R}}$ does come with some peculiarities. Indeed, if $F : V \to \overline{\mathbb{R}}$ assumes $F(u) = -\infty$ at a single point $u \in V$, then on each ray starting in u, F may assume only one single finite value without violating convexity [43, Section I.2]. For convenience, we distinguish these special cases with the following definition.

Definition 2.17 (Proper function)**.** *Let V be a set. A function $F : V \to \overline{\mathbb{R}}$ is called* proper *if $F(u) < \infty$ for at least one $u \in V$ and $F(v) > -\infty$ for all $v \in V$.*

With the following property we want to ensure that discontinuities of functions are well-behaved for minimization in the sense that in particular for function values of some convergent sequence of points, the value does not jump upwards at the limit.

Definition 2.18 (Lower semi-continuity)**.** *Let V be a topological space. A function $F : V \to \overline{\mathbb{R}}$ is called (sequentially) lower semi-continuous at $v \in V$ if for any sequence $(v_k)_{k\in\mathbb{N}} \subseteq V$ with $\lim_{k\to\infty} v_k = v$ we have*

$$F(v) \leq \liminf_{k\in\mathbb{N}} F(v_k).$$

It is important to note, that the definition above depends on the notion of convergence used. Indeed, since strong convergence of a sequence implies weak convergence, a weakly lower semi-continuous function is also strongly lower semi-continuous.

We have the following important set-based characterization of lower semi-continuity.

Proposition 2.19. *A function $F : V \to \overline{\mathbb{R}}$ is lower semi-continuous if and only if all level sets $L_a := \{v \in V : F(v) \leq a\}$, $a \in \mathbb{R}$ are sequentially closed.*

Proof. Assume F is lower semi-continuous and let $a \in \mathbb{R}$. If $L_a = \emptyset$ then it is closed. Otherwise let $(u_k)_{k\in\mathbb{N}} \subseteq L_a \neq \emptyset$ with $u_k \to u \in V$. Then $F(u) \leq \liminf_{k\in\mathbb{N}} F(u_k) \leq a$ and therefore $u \in L_a$.

Assume on the other hand that L_a is sequentially closed for any $a \in \mathbb{R}$ and let $(u_k)_{k\in\mathbb{N}} \subseteq V$ with $u_k \to u \in V$. If $\liminf_{k\in\mathbb{N}} F(u_k) \in \mathbb{R}$ define $a := \liminf_{k\in\mathbb{N}} F(u_k) + \varepsilon$ for some $\varepsilon > 0$. Then $(u_k)_{k\geq\hat{k}} \subseteq L_a$ for some $\hat{k} \in \mathbb{N}$. Thus since L_a is sequentially closed we have $u \in L_a$, i.e.

$$F(u) \leq a = \liminf_{k\in\mathbb{N}} F(u_k) + \varepsilon.$$

Letting $\varepsilon \to 0$ concludes lower semi-continuity. If $\liminf_{k\in\mathbb{N}} F(u_k) = -\infty$ then the above argument may be carried out for any $a \in \mathbb{R}$ thus showing $F(u) = -\infty$. If finally $\liminf_{k\in\mathbb{N}} F(u_k) = \infty$ then $F(u) \leq \liminf_{k\in\mathbb{N}} F(u_k)$ holds immediately. □

Using this characterization, one may easily show the following statement about pointwise suprema of functions.

Lemma 2.20. *Let $F_k : V \to \overline{\mathbb{R}}$, $k \in \mathbb{N}$, be lower semi-continuous functions. Then the pointwise supremum $F : V \to \overline{\mathbb{R}}$, $F(v) := \sup_{k\in\mathbb{N}} F_k(v)$ is lower semi-continuous.*

Proof. The level sets L_a, $a \in \mathbb{R}$ of F are given by

$$L_a = \big\{v \in V : \sup_{k\in\mathbb{N}} F_k(v) \leq a\big\} = \bigcap_{k\in\mathbb{N}} \{v \in V : F_k(v) \leq a\}.$$

Since the countable intersection of closed sets is closed, the statement follows from the characterization in Proposition 2.19. □

For convex functions, lower semi-continuity with regard to weak convergence agrees with lower semi-continuity.

Lemma 2.21. *Let $F : V \to \overline{\mathbb{R}}$ be a convex function. Then F is lower semi-continuous if and only if it is weakly lower semi-continuous.*

Proof. Since F is convex, the level sets L_a, $a \in \mathbb{R}$ of F are convex. Then due to Lemma 2.3, all L_a, $a \in \mathbb{R}$ are closed if and only if they are weakly closed. □

The introduced properties, together with a compactness condition, allow to show existence of minimizers by directly analyzing a potential minimizing sequence.

Theorem 2.22 (Direct Method, c.f. [19, Remark 1.36])**.** *Let V be a metric space and $F : V \to \overline{\mathbb{R}}$ be*

(i) weakly coercive, i.e. $\inf_{v\in V} F(v) = \inf_{v\in K} F(v)$ *for some (sequentially) compact set $\emptyset \neq K \subseteq V$,*

(ii) (sequentially) lower semi-continuous and

(iii) proper.

Then F has at least one minimizer $\hat{v} \in V$ with finite $F(\hat{v})$.

Proof. Since F is weakly coercive, we have

$$\inf_{v\in V} F(v) = \inf_{v\in K} F(v) = \lim_{n\to\infty} F(w_n),$$

for some compact set $K \subseteq V$ and some sequence $(w_n)_{n\in\mathbb{N}} \subseteq K \neq \emptyset$. Since K is compact, there exists a convergent subsequence $(v_n)_{n\in\mathbb{N}} \subseteq (w_n)_{n\in\mathbb{N}}$, $v_n \to \hat{v} \in K$. Due to F being lower semi-continuous at $\hat{v} \in K$, we have

$$\inf_{v\in V} F(v) = \lim_{n\to\infty} F(v_n) \geq F(\hat{v}).$$

Consequently, $\hat{v} \in K$ is a minimizer of F with finite $F(\hat{v})$ since F is proper. □

The topology or, more generally, the notion of convergence in Theorem 2.22 for lower semi-continuity and weak coercivity need to agree, but may be chosen at will. In practice it should be chosen as weak as necessary to ensure weak coercivity, while at the same time being as strong as possible to relax the precondition on lower semi-continuity. Often weak coercivity of the functional F in Theorem 2.22 is ensured by checking for coercivity of F (note the unfortunate naming) thanks to Lemma 2.2.

Convex functions allow for a set-valued generalization of the derivative, which, roughly speaking, at a given point represents the set of all tangent hyperplanes bounding the function from below.

Definition 2.23 (Subdifferential)**.** *For a convex function $F : V \to \overline{\mathbb{R}}$ the* subdifferential $\partial F(u) \subseteq V^*$ *of F at $u \in V$ is defined by*

$$\begin{aligned} &u^* \in \partial F(u) \\ &\qquad \iff \quad F(u) < \infty \;\wedge\; \forall v \in V : \langle u^*, v - u\rangle_{V^*,V} \leq F(v) - F(u). \end{aligned}$$

For a function $F : V \to \overline{\mathbb{R}}$, $v \in V$ we will make use of the shorthand $\langle \partial F, v\rangle_{V^*,V} = \{\langle u^*, v\rangle_{V^*,V} : u^* \in \partial F\}$. Also, statements containing $\partial F(u)$ are to be understood as being valid for all elements of $\partial F(u)$, e.g. $\langle \partial F(u), v\rangle_{V^*,V} \leq 0$ is a shorter way of writing $\forall u^* \in \partial F(u) : \langle u^*, v\rangle_{V^*,V} \leq 0$.

The upcoming notion sits at the heart of convex duality theory and defines a mapping between functions $V \to \overline{\mathbb{R}}$ and certain conjugate functions $V^* \to \overline{\mathbb{R}}$.

Definition 2.24 (Convex conjugate function)**.** *Let $F : V \to \overline{\mathbb{R}}$ be a function. Then $F^* : V^* \to \overline{\mathbb{R}}$ denotes the* convex conjugate *(also called* polar function *or* Legendre transform*), defined by*

$$F^*(v^*) := \sup_{v \in V} \big\{ \langle v^*, v\rangle_{V^*,V} - F(v) \big\}.$$

With the following propositions we collect some important properties of the convex conjugate.

Proposition 2.25 (Convex conjugate identities)**.** *Let $F : V \to \overline{\mathbb{R}}$ be proper, lower semi-continuous and convex. Then the following identities apply:*

(i) $(F^*)^* = F$,

(ii) $(\lambda F + c)^* = \lambda F^*(\frac{\cdot}{\lambda}) - c$, *where* $\lambda > 0$, $c \in \mathbb{R}$,

(iii) $F(\lambda \cdot + u)^* = F^*(\frac{\cdot}{\lambda}) - \langle \cdot\,, \frac{u}{\lambda}\rangle_{V^*,V}$, *where* $\lambda \in \mathbb{R} \setminus \{0\}$, $u \in V$,

(iv) *Let $F : V \to \mathbb{R}$ be an absolutely homogeneous function, i.e. $F(\alpha v) = |\alpha| F(v)$ for all $\alpha \in \mathbb{R}$, $v \in V$. Then*

$$F^*(v^*) = \chi_{g(v^*)\leq 1},$$

where $v^* \in V^*$ and $g(v^*) := \sup_{v\in V, F(v)\leq 1}\langle v^*, v\rangle_{V^*,V}$. In particular $\|\cdot\|^* = \chi_{\|\cdot\|_*\leq 1}$ for any norm $\|\cdot\| : V \to \mathbb{R}$ and corresponding dual norm $\|\cdot\|_* : V^* \to \mathbb{R}$.

(v) *If V is a Hilbert space and $A : V \to V$ is an invertible symmetric bounded linear operator, then*

$$\Big(\tfrac{1}{2}\langle\,\cdot\,, A\,\cdot\,\rangle_V\Big)^* = \tfrac{1}{2}\langle\,\cdot\,, A^{-1}\,\cdot\,\rangle_V.$$

Proof. (i) We refer to [43, Proposition II.4.1].

(ii) We calculate

$$\begin{aligned}(\lambda F + c)^* &= v^* \mapsto \sup_{v\in V}\big\{\langle v^*, v\rangle_{V^*,V} - (\lambda F(v) + c)\big\}\\ &= v^* \mapsto \lambda \sup_{v\in V}\big\{\langle \tfrac{v^*}{\lambda}, v\rangle_{V^*,V} - F(v)\big\} - c\\ &= \lambda F^*(\tfrac{v^*}{\lambda}) - c.\end{aligned}$$

(iii) We calculate

$$\begin{aligned}F(\lambda\cdot + u)^* &= v^* \mapsto \sup_{v\in V}\big\{\langle v^*, v\rangle_{V^*,V} - F(\lambda v + u)\big\}\\ &= v^* \mapsto \sup_{v'\in V}\big\{\langle v^*, \tfrac{v'-u}{\lambda}\rangle_{V^*,V} - F(v')\big\}\\ &= v^* \mapsto \sup_{v'\in V}\big\{\langle \tfrac{v^*}{\lambda}, v'\rangle_{V^*,V} - F(v')\big\} - \langle v^*, \tfrac{u}{\lambda}\rangle_{V^*,V}\\ &= F^*(\tfrac{\cdot}{\lambda}) - \langle\,\cdot\,, \tfrac{u}{\lambda}\rangle_{V^*,V}.\end{aligned}$$

(iv) Due to absolute homogeneity of F, we see that for the convex conjugate

$$\begin{aligned} F^*(v^*) &= \sup_{v\in V} \langle v^*, v\rangle_{V^*,V} - F(v) \\ &\le \sup_{\alpha\ge 0} \alpha\Big(\sup_{\substack{w\in V \\ F(w)\le 1}} \langle v^*, w\rangle_{V^*,V} - F(w)\Big) \in \{0,\infty\}, \end{aligned}$$

depending on whether $\langle v^*, w\rangle - F(w) > 0$ for any $w \in W$ with $F(w) \le 1$. Per definition of g, this is the case if $g(v^*) > 1$. On the other hand, if $\langle v^*, w\rangle > F(w)$ for some $w \in V$, $F(w) \le 1$, then due to homogeneity $F(w) \ne 0$ and $\langle v^*, \alpha w\rangle > F(\alpha w) = 1$ for $\alpha = |F(w)|^{-1}$, which implies $g(v^*) > 1$. Thus, by full case distinction we get

$$F^*(v^*) = \begin{cases} 0 & \text{if } g(v^*) \le 1, \\ \infty & \text{if } g(v^*) > 1. \end{cases}$$

(v) We have $(\frac{1}{2}\langle\,\cdot\,, A\,\cdot\,\rangle_V)^* = v^* \mapsto \sup_{v\in V}\langle v^*, v\rangle_{V^*,V} - \frac{1}{2}\langle v, Av\rangle_V$. Identifying $v^* \in V^*$ with its Riesz representative in V due to Theorem 2.9, the supremum is attained for

$$0 = v^* - \tfrac{1}{2}(A + A^*)v,$$

i.e. $v = A^{-1}v^*$, since A is symmetric. Consequently, again using the Riesz representation from Theorem 2.9, we get

$$\begin{aligned} (\tfrac{1}{2}\langle\,\cdot\,, A\,\cdot\,\rangle_V)^* &= v^* \mapsto \langle v^*, A^{-1}v^*\rangle_{V^*,V} - \tfrac{1}{2}\langle A^{-1}v^*, AA^{-1}v^*\rangle_V \\ &= v^* \mapsto \tfrac{1}{2}\langle v^*, A^{-1}v^*\rangle_V. \end{aligned}$$ □

Proposition 2.26 (Convex conjugate separability, [43, III, Remark 4.3])**.** *Let $V = V_1 \times V_2$ be a Banach space. If $F : V \to \overline{\mathbb{R}}$ is separable, i.e. $F(v_1, v_2) = F_1(v_1) + F_2(v_2)$ for functions $F_1 : V_1 \to \overline{\mathbb{R}}$ and $F_2 : V_2 \to \overline{\mathbb{R}}$, then so is F^*:*

$$F^*(v_1^*, v_2^*) = F_1^*(v_1^*) + F_2^*(v_2^*).$$

Proof. From the definition we see immediately that for $v^* = (v_1^*, v_2^*) \in V^*$:

$$\begin{aligned}
F^*(v^*) &= \sup_{v \in V} \langle v^*, v \rangle_{V^*,V} - F(v) \\
&= \sup_{(v_1,v_2) \in V} \langle v_1^*, v_1 \rangle_{V_1^*,V_1} + \langle v_2^*, v_2 \rangle_{V_2^*,V_2} - F_1(v_1) - F_2(v_2) \\
&= F_1^*(v_1^*) + F_2^*(v_2^*). \qquad \square
\end{aligned}$$

Proposition 2.27. *For a proper function $F : V \to \overline{\mathbb{R}}$, $u \in V$, $u^* \in V^*$ the following statements are equivalent:*

(i) $u^* \in \partial F(u)$,

(ii) $u \in \partial F^*(u^*)$,

(iii) $\langle u^*, u \rangle_{V^*,V} = F(u) + F^*(u^*)$.

Proof. Observe that F^* is proper. Thus, the statement of (iii) is symmetric in the terms $F(u)$, $F^*(u^*)$ and it suffices to show the equivalence between (i) and (iii) since the equivalence with (ii) can be established analogously.

Let now (i) be true. Rearranging gives $\forall v \in V : \langle u^*, v \rangle_{V^*,V} - F(v) \leq \langle u^*, u \rangle_{V^*,V} - F(u)$. Since F is proper, $F(u) > -\infty$ holds and from the definition of the subdifferential we infer

$$F^*(u^*) = \sup_{v \in V} \{ \langle u^*, v \rangle_{V^*,V} - F(v) \} = \langle u^*, u \rangle_{V^*,V} - F(u) < \infty,$$

which shows (iii). From (iii), on the other hand, we see that $F(u) < \infty$ since F^* is proper, and $\langle u^*, u \rangle_{V^*,V} - F(u) = F^*(u^*) \geq \langle u^*, v \rangle_{V^*,V} - F(v)$ for all $v \in V$, which yields (i). □

Proposition 2.27 shows that F and F^* are closely related through their subdifferentials ∂F and ∂F^*. For minimization problems a duality theory due to Fenchel (see [43] for details) then allows to formulate equivalent dual problems. We present a specific version of the duality theorem for the type of minimization problem that will be relevant in Chapter 3.

Theorem 2.28 (Fenchel duality, [43, Remark III.4.2])**.** *Let V and W be reflexive Banach spaces, $A : V \to W$ be a continuous linear operator and $F : V \to \overline{\mathbb{R}}$, $G : W \to \overline{\mathbb{R}}$ be proper, convex, lower semi-continuous functions such that there exists $v_0 \in V$ with $F(v_0) + G(Av_0) < \infty$ and G continuous at Av_0. Then the following holds:*

$$\inf_{v \in V} F(v) + G(Av) = \sup_{w^* \in W^*} -F^*(A^* w^*) - G^*(-w^*). \tag{2.4}$$

The problem on the right hand side in (2.4) *has at least one solution. In addition $\hat{v} \in V$, $\hat{w}^* \in W^*$ are solutions to both optimization problems if and only if*

$$\begin{aligned} A^* \hat{w}^* &\in \partial F(\hat{v}), \\ -\hat{w}^* &\in \partial G(A\hat{v}). \end{aligned}$$

2.3 Total Variation

The total variation of some real-valued function is a measure for its change or oscillation over its whole domain. For weakly differentiable f one may think of it as the Sobolev $W^{1,1}$ semi-norm, as will become clear through Proposition 2.33. The following definition applies more generally to non-smooth, vector-valued functions.

Definition 2.29 (Total variation)**.** *For $u \in L^1(\Omega)^m$ let the* total variation *be defined as*

$$\begin{aligned} \mathrm{TV}(u) := \int_\Omega |Du|_F &:= \sup_{|\boldsymbol{p}|_F \le 1} \langle u, \mathrm{div}\boldsymbol{p} \rangle \\ &:= \sup \Big\{ \int_\Omega u \cdot \mathrm{div}\boldsymbol{p} \,\mathrm{d}x : \boldsymbol{p} \in C_0^\infty(\Omega)^{d \times m}, \\ &\qquad\qquad |\boldsymbol{p}(x)|_F \le 1 \textit{ for almost every } x \in \Omega \Big\}, \end{aligned} \tag{2.5}$$

where the operator div : $C_0^\infty(\Omega)^{d \times m} \to C_0^\infty(\Omega)^m$ *is the column-wise divergence (in d), while $|\cdot|_F : \mathbb{R}^{d \times m} \to \mathbb{R}$ denotes the Frobenius norm.*

Note that in Definition 2.29 we defined the whole expression $\int_\Omega |Du|_F$ in u *as is* without specifying its components. The integral notation is natural due to a characterization of the total variation as a measure $|Du|_F$ [7]. We record the following two central properties of the total variation.

Proposition 2.30 (c.f. [7, Remark 3.5])**.** *The total variation* TV : $L^1(\Omega)^m \to \overline{\mathbb{R}}$ *is*

(i) lower semi-continuous

(ii) convex

Proof. (i) Since $L^1(\Omega)^m \to \mathbb{R}$, $u \mapsto \int_\Omega u \cdot \operatorname{div} p \, \mathrm{d}x$ is continuous for any fixed $p \in C_0^\infty(\Omega)^{d\times m}$, we conclude lower semi-continuity of the supremum using Lemma 2.20.

(ii) As a supremum of affine functions, TV is convex. □

Remarkably, the set of all L^1-functions with bounded total variation forms a Banach space in the following way.

Theorem 2.31 ([8, Theorem 10.1.1])**.** *The vector space* $BV(\Omega)^m := \{u \in L^1(\Omega)^m : \mathrm{TV}(u) < \infty\}$ *together with the norm*

$$\|u\|_{BV(\Omega)^m} := \|u\|_{L^1(\Omega)^m} + \mathrm{TV}(u)$$

is a Banach space, called the space of m-vector-valued bounded variation functions $BV(\Omega)^m$*.*

We note that there are different ways to define the total variation for vector-valued functions and refer to [47] for a short overview. Nevertheless, the topological properties remain the same, as we record with the following remark.

Remark 2.32. *If one replaces the pointwise norm $|\cdot|_F$ in Definition 2.29 with any other norm, the defined total variation $TV(u)$ may be different but the resulting space $BV(\Omega)^m$ will be topologically equivalent.*

Indeed, since any two norms $|\cdot|_a,\ |\cdot|_b : \mathbb{R}^m \to [0,\infty)$ *are equivalent, i.e.* $c|x|_b \le |x|_a \le C|x|_b$ *for all* $x \in \mathbb{R}^m$ *for constants* $c, C > 0$*, we observe for any homogeneous functional* $F : C_c^1(\Omega)^{d\times m} \to \mathbb{R}$ *that*

$$\begin{aligned}\tfrac{1}{C} \sup_{|\boldsymbol{p}(x)|_b \le 1} F(\boldsymbol{p}) \le \tfrac{1}{C} \sup_{\frac{1}{C}|\boldsymbol{p}(x)|_a \le 1} F(\boldsymbol{p}) &= \sup_{|\boldsymbol{p}(x)|_a \le 1} F(\boldsymbol{p}) \\ &\le \sup_{c|\boldsymbol{p}(x)|_b \le 1} F(\boldsymbol{p}) = \tfrac{1}{c} \sup_{|\boldsymbol{p}(x)|_b \le 1} F(\boldsymbol{p}).\end{aligned}$$

Consequently, the corresponding norms $\|\cdot\|_a$ *and* $\|\cdot\|_b$ *on* $BV(\Omega)^m$ *are equivalent and* $BV(\Omega)^m$ *carries the same topology as e.g. the space of bounded variation from the extensive work [7].*

The total variation displays some remarkable elementary properties which we try to summarize with the following proposition. The choice of $|\cdot|_F$ in particular allows for rotational invariance of the total variation in both the domain (change of coordinates) and the range (global rotation of vector field) of u.

Proposition 2.33. *Let* $u, v \in L^1(\Omega)^m$*. The total variation has the following basic properties:*

(i) $\mathrm{TV}(u + c) = \mathrm{TV}(u)$ *for* $c \in \mathbb{R}$*,*

(ii) $\mathrm{TV}(u + v) \le \mathrm{TV}(u) + \mathrm{TV}(v)$*,*

(iii) $\mathrm{TV}(\lambda u) = |\lambda|\,\mathrm{TV}(u)$ *for* $\lambda \in \mathbb{R}$*,*

(iv) $\mathrm{TV}(R \circ u \circ Q) = \mathrm{TV}(u)$ *for any rotation* $Q \in \mathbb{R}^{d\times d}$*,* $Q^TQ = I$ *of the domain and any rotation* $R \in \mathbb{R}^{m\times m}$*,* $R^TR = I$ *of the codomain,*

(v) If $u \in H^1(\Omega)^m$ *then* $\mathrm{TV}(u) = \int_\Omega |\nabla u|_F$*,*

Proof. We prove the statements by extensive use of the divergence theorem.

(i) Due to linearity of the inner product and since $\boldsymbol{p}$ has compact support, we deduce

$$\begin{aligned}\mathrm{TV}(u+c) &= \sup_{|\boldsymbol{p}|_F \leq 1} \langle u+c, \operatorname{div} \boldsymbol{p}\rangle \\ &= \sup_{|\boldsymbol{p}|_F \leq 1} \langle u, \operatorname{div} \boldsymbol{p}\rangle + c \int_\Omega \operatorname{div} \boldsymbol{p} = \mathrm{TV}(u).\end{aligned}$$

(ii) Splitting the supremum for u and v individually yields the inequality

$$\begin{aligned}\mathrm{TV}(u+v) &= \sup_{|\boldsymbol{p}|_F \leq 1} \langle u+v, \operatorname{div} \boldsymbol{p}\rangle \\ &\leq \sup_{|\boldsymbol{p}|_F \leq 1} \langle u, \operatorname{div} \boldsymbol{p}\rangle + \sup_{|\boldsymbol{q}|_F \leq 1} \langle u, \operatorname{div} \boldsymbol{q}\rangle \\ &= \mathrm{TV}(u) + \mathrm{TV}(v).\end{aligned}$$

(iii) By moving the sign $\operatorname{sgn}(\lambda)$ into the dual variables, we get

$$\begin{aligned}\mathrm{TV}(\lambda u) &= \sup_{|\boldsymbol{p}|_F \leq 1} \langle \lambda u, \operatorname{div} \boldsymbol{p}\rangle \\ &= |\lambda| \sup_{|\operatorname{sgn}(\lambda)\boldsymbol{p}| \leq 1} \langle u, \operatorname{div} \boldsymbol{p}\rangle = |\lambda| \, \mathrm{TV}(u).\end{aligned}$$

(iv) For $\boldsymbol{p} \in C_0^\infty(\Omega)^{d\times m}$, let $\boldsymbol{p}_k \in C_0^\infty(\Omega)^m$, $k = 1, \dots, d$ denote its k-th row component. We then evaluate

$$\begin{aligned}\mathrm{TV}(R \circ u \circ Q) &= \sup_{|\boldsymbol{p}|_F \leq 1} \langle R \circ u \circ Q, \operatorname{div} \boldsymbol{p}\rangle_{L^2(Q^T\Omega)^m} \\ &= \sup_{|\boldsymbol{p}|_F \leq 1} \langle R \circ u, \operatorname{div}(\boldsymbol{p} \circ Q^T)\rangle_{L^2(\Omega)} \\ &= \sup_{|\boldsymbol{p}|_F \leq 1} \langle u, R^T \circ \operatorname{div}(\boldsymbol{p} \circ Q^T)\rangle_{L^2(\Omega)} \\ &= \sup_{|\boldsymbol{p}|_F \leq 1} \langle u, \operatorname{div}((R^T \circ \boldsymbol{p}_k \circ Q^T)_{k=1}^d)\rangle_{L^2(\Omega)} \\ &= \mathrm{TV}(u).\end{aligned}$$

(v) We refer to [8, Section 10.1] at this point and note that we will be able to prove this more generally in Proposition 3.14 ourselves as soon as a certain density argument is established by Theorem 3.12. □

We remark that while rotational invariance in the domain space in particular is a desirable property, it is difficult to achieve in a discrete setting, as we will point out in Chapter 5.

3 The L^1-L^2-TV-Functional and Duality

After having laid some of the groundwork for total variation functionals in Chapter 2 the current chapter will introduce the optimization model which will be the focus of both our decomposition methods in Chapter 4 and our discretization efforts in Chapter 5. In this chapter we cover basic applications of the model, primal and dual formulations, existence and uniqueness of their solutions as well as regularization.

The meticulous dualization of the combined L^1-L^2-TV model and its regularized variant for vector-valued functions in a general Hilbert space setting in particular may be considered a new contribution. Compared to the primal-dual methods in [40, 62, 65, 67], where the dualization is performed either on smooth or on discrete function spaces, our approach applies to non-smooth, vector-valued functions in $BV(\Omega)^m \cap L^2(\Omega)^m$, $m \in \mathbb{N}$ as well. Due to this vector-valued setting, dualization results for the scalar case derived in [58] and [60] need to be adjusted accordingly. Compared to [60] we also explore a new alternative proof of the density argument necessary for the dual characterization of the total variation.

A preliminary primal-dual formulation for the corresponding scalar version was kindly contributed by Andreas Langer and used as a basis for the derivation. Results of this chapter paired with corresponding numerical examples from Chapter 5 are in preparation to be published separately [56].

3.1 Model and Motivation

As in Chapter 2, let $\Omega \subseteq \mathbb{R}^d$ be an open, bounded and simply connected domain with Lipschitz boundary, where $d \in \mathbb{N}$ denotes the spatial dimension, e.g. $d = 1$ for signals or $d = 2$ for typical images. Functions $\Omega \to \mathbb{R}^m$ may be viewed as general images, where $m \in \mathbb{N}$ denotes the number of output channels, e.g. $m = 1$ for grey-scale images or $m = d$ for motion fields. We will concern ourselves with minimizing a non-smooth functional consisting of a combined L^1-L^2 data fidelity term and a total variation term. More precisely, letting $g \in L^2(\Omega)$ be the given data, $T : L^2(\Omega)^m \mapsto L^2(\Omega)$ be a bounded linear operator and $\alpha_1, \alpha_2, \lambda \geq 0$ be adjustable weighting parameters, we consider the so-called L^1-L^2-TV model

$$\inf_{\substack{u \in L^2(\Omega)^m \\ \cap BV(\Omega)^m}} \alpha_1 \|Tu - g\|_{L^1(\Omega)} + \tfrac{\alpha_2}{2} \|Tu - g\|^2_{L^2(\Omega)} + \lambda \int_\Omega |Du|_F, \tag{3.1}$$

which was first proposed in a slightly more general way in [59] for the scalar-valued case $m = 1$. Modifications of the L^1-L^2-TV model have been presented in [48], where the total variation is replaced by $\|Wu\|_{L^1}$ with W being a wavelet tight frame transform, and in [70], where the second order total generalized variation [20] has been used as regularization term and box-constraints were incorporated, which assure that the reconstruction lies in the respective dynamic range.

Before we proceed to analyze (3.1) through its upcoming extensions (3.5) and (3.13), we will first motivate its study by giving some example applications.

3.1.1 Applications

The model (3.1) may be applied to imaging problems in various ways by choosing the operator T appropriately. We highlight some of these approaches to give a rough sense of the practical applicability of the model. These will also come up later as numerical examples in Chapter 5.

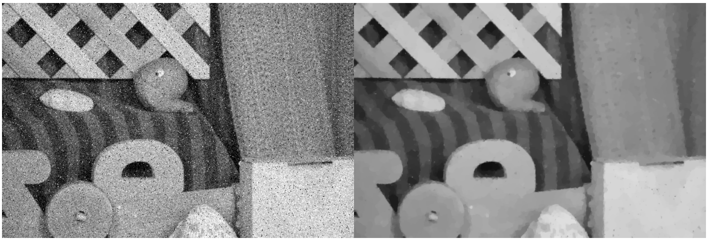


Figure 3.1: left: input image g with mixed noise (original image from [11]), right: denoised output image u, see Section 5.5.1

Denoising

Given a noisy image $g \in L^2(\Omega)$, the removal of noise in order to obtain a clear image $u \in L^2(\Omega)$ is called *denoising*, see Figure 3.1. This operation may be performed by using the identity operator $T := I$ in (3.1). The usage of total variation as a regularization term for noise removal is well-known to preserve edges. In particular, it has been demonstrated [59, 66, 68] that the optimization problem (3.1) is well suited for removing a mixture of Gaussian and impulse noise, which is relevant when the input data has been affected by both noise types, possibly by separate processes. Moreover it is easy to see that (3.1) is a generalization of two well-known total variation models. Namely, for $\alpha_1 = 0$ we obtain the so-called L^2-TV model, which has been successfully used to remove additive Gaussian noise in images [29], while for $\alpha_2 = 0$ we get the so-called L^1-TV model which is used to remove impulse noise [6, 72, 73].

Inpainting

Restoring a given defective image $g \in L^2(\Omega \setminus D)$ to obtain a reconstruction $u \in L^2(\Omega)$ covering the defective region $D \subseteq \Omega$ is called *inpainting*, see Figure 3.2. We may perform inpainting using the model

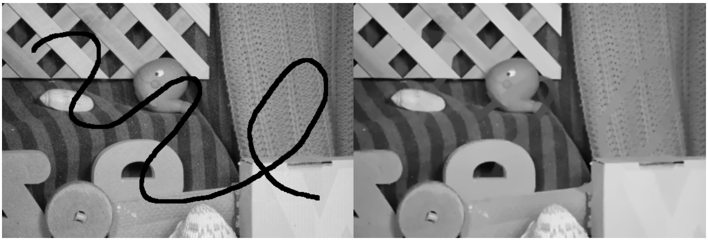


Figure 3.2: left: input image g [11] with corrupted regions, right: inpainted output image u, see Section 5.5.2

(3.1) by setting $T := \mathrm{Id}_{\Omega \setminus D}$ as a masking operator defined by

$$(\mathrm{Id}_{\Omega\setminus D} u)(x) := \begin{cases} u(x) & x \in \Omega \setminus D, \\ 0 & x \in D, \end{cases} \tag{3.2}$$

which may be interpreted as a restriction of the data functional to the known area $\Omega \setminus D$. A solution u to (3.1) will thus intuitively, try to match g on $\Omega \setminus D$ where data is given, while having small total variation overall. In particular, this formulation matches the denoising setting for $D = \emptyset$.

Optical Flow

The problem of optical flow is to compute the apparent motion field of an image sequence. One approach, given two grey-scale images $f_0, f_1 : \Omega \to [0,1]$, is to estimate a displacement field $u : \Omega \to \mathbb{R}^m$, $m = d$ which maps points of similar brightness, see Figure 3.3, i.e. for all $x \in \Omega$

$$f_0(x) = f_1(x + u(x)). \tag{3.3}$$

Here exceeding displacements $x + u(x) \notin \Omega$ are ignored. Equation (3.3) is called the *brightness constancy assumption.* It is usually underdetermined since u is vector-valued while (3.3) is scalar, and depending on f_0, f_1 there might not even exist a solution, e.g. due to occlusion or brightness change. Nevertheless, (3.3) may be still applied

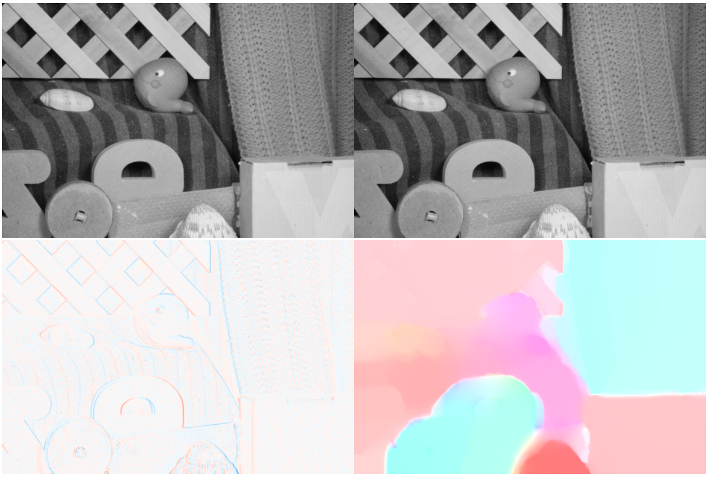


Figure 3.3: from left to right, top to bottom: input frames f_0, f_1 [11], image difference $f_1 - f_0$, estimated optical flow displacment field u, see Section 5.5.3

in a minimization setting as a data term, e.g. using the L^2 residual, together with suitable regularization to arrive at an approximate motion field u [9].

Assuming smooth f_0, f_1 and expanding the right hand side of (3.3) at $x + u_0(x)$ for some smooth initial guess $u_0 : \Omega \to \mathbb{R}^2$ one arrives at

$$\begin{aligned} f_0(x) &= f_1(x + u_0(x) + (u - u_0)(x)) \\ &\approx f_1(x + u_0(x)) + \nabla f_1(x + u_0(x)) \cdot (u - u_0)(x) \\ &\approx f_w(x) + \nabla f_w(x) \cdot (u - u_0)(x) \\ &= f_w(x) + \nabla f_w(x) \cdot u(x) - \nabla f_w(x) \cdot u_0(x), \end{aligned} \tag{3.4}$$

where f_w, defined as $f_w(x) := f_1(x + u_0(x))$ is a (backwards-)warped version of f_1. Note that in the derivation sketched above we generally have

$$\nabla f_w(x) = (I + u_0'(x)^T)\nabla f_1(x + u_0(x)) \neq \nabla f_1(x + u_0(x)).$$

We call (3.4) the *optical flow equation linearized at the initial guess* u_0. Note that for any solution u to (3.4), $u + v$ with $\nabla f_w \cdot v = 0$ is a solution as well, i.e. the linearized optical flow equation provides flow information only in the image gradient direction, a phenomenon also known as *aperture problem.*

We use model (3.1) to estimate a solution to (3.4) by setting

$$\begin{aligned} Tu &:= \nabla f_w \cdot u, \\ g &:= \nabla f_w \cdot u_0 - (f_w - f_0). \end{aligned}$$

The parameters α_1, α_2, λ in (3.1) allow us to control the optical flow model and special cases have been used in their discrete forms for calculating the optical flow of image sequences, e.g. discrete L1-TV optical flow in [82] and a comparison of L1-TV and L2-TV in [39].

Since (3.4) is linearized, it is intuitively clear that we cannot expect large displacements to be resolved without providing a close initial guess u_0. Nevertheless, we will see in Chapter 5 that an iterative warping algorithm, namely Algorithm 5.18, can alleviate this problem and produce adequate results.

3.1.2 Primal Formulation

We return to study our L^1-L^2-TV model (3.1). In its general form, it exhibits undesirable properties: namely, existence and uniqueness of the solution may not be guaranteed, as we show by the following simple examples.

Example 3.1. *Let* $\Omega := (0,1) \subseteq \mathbb{R}^1$*. Counter-examples to existence and uniqueness of solutions of* (3.1) *are given as follows.*

(i) Let $\alpha_1 = 1$*,* $\alpha_2 = 0$*,* $\lambda = 0$*,* $(Tu)(x) := xu(x)$ *and* $g(x) := 1$ *for* $x \in \Omega$*. Observe the sequence* $(v_k)_{k\in\mathbb{N}} \in BV(\Omega)$ *with* $v_k(x) :=$

$\min\{k, \frac{1}{x}\}$. *Since for $k \in \mathbb{N}$ the functional in* (3.1) *becomes*

$$0 \leq \|Tv_k - g\|_{L^1} = \int_0^1 |x \min\{k, \tfrac{1}{x}\} - 1| \,\mathrm{d}x$$
$$= \int_0^{\frac{1}{k}} 1 - kx \,\mathrm{d}x = \tfrac{1}{2k} \to 0,$$

we infer that $\|Tu_k - g\|_{L^1} \to 0$ must hold for any minimizing sequence $(u_k)_{k\in\mathbb{N}} \subseteq BV(\Omega)$. Further, by restricting (u_k) to some subsequence, we have $u_k(x) \to \frac{1}{x} =: \hat{u}(x)$ for a.e. $x \in \Omega$. But $\hat{u} \notin L^1(\Omega)$ and therefore, no solution to (3.1) *exists in $BV(\Omega) \subseteq L^1(\Omega)$.*

(ii) Let $T = 0$. Then any constant function $u \in BV(\Omega)$ minimizes (3.1) *since the total variation vanishes:*

$$0 \leq \lambda \int_\Omega |Du|_F = 0.$$

We will address these issues by formulating a coercivity condition in Proposition 3.2 for the following extended version of (3.1).

Hilbert Space Setting

As in (3.1), let $\Omega \subseteq \mathbb{R}^d$, $d \in \mathbb{N}$ be an open, bounded and simply connected domain with Lipschitz boundary. Let $V \subseteq L^2(\Omega)^m$ be a continuously embedded Hilbert space, $T : V \to L^2(\Omega)$ a bounded linear operator, $S : V \to V_S$ a bounded linear operator for some Hilbert space V_S and $\alpha_1, \alpha_2, \lambda, \beta \geq 0$. Then the penalized version of (3.1) in a Hilbert space setting reads

$$\inf_{u\in V} \alpha_1\|Tu - g\|_{L^1} + \tfrac{\alpha_2}{2}\|Tu - g\|^2_{L^2} + \tfrac{\beta}{2}\|Su\|^2_{V_S} + \lambda \int_\Omega |Du|_F. \quad (3.5)$$

Here $\beta \geq 0$ is an optional penalization parameter. We do want to note that searching for solutions $u \in V$ in (3.5) instead of in the space $V \cap BV(\Omega)^m$ as in (3.1) does not affect the orginal problem in its intended purpose. Indeed, for $\lambda > 0$ any $u \in V \subseteq L^2(\Omega)^m \subseteq L^1(\Omega)^m$

with finite energy (3.5) will have finite total variation and therefore be an element of $BV(\Omega)^m$.

For the operator S and its related spaces we will restrict ourselves to the choices

(S.i) $S = I : V \to V_S$ where $V \subseteq L^2(\Omega)^m$ and $V_S \subseteq L^2(\Omega)^m$, and

(S.ii) $S = \nabla : V \to V_S$ where $V \subseteq H^1(\Omega)^m$ and $V_S \subseteq L^2(\Omega)^{d\times m}$,

which we will refer to as Setting (S.i) and Setting (S.ii), respectively. Note that Setting (S.ii) has $V \subseteq H^1(\Omega)^m$, which restricts $u \in V$ to allow for weak derivatives, while Setting (S.i) does not.

3.1.3 The Bilinear Form a_B

To describe the differentiable part of (3.5) it is convenient to define the symmetric bilinear form $a_B : V \times V \to \mathbb{R}$ by

$$a_B(u, w) := \alpha_2\langle Tu, Tw\rangle + \beta\langle Su, Sw\rangle = \langle Bu, w\rangle_{V^*,V} \tag{3.6}$$

with $B : V \to V^*$ denoting the operator $B := \alpha_2 T^*T + \beta S^*S$. Thus $Bu = v$ for $u \in V$, $v \in V^*$ if and only if

$$a_B(u, w) = \langle v, w\rangle_{V^*,V} \tag{3.7}$$

for all $w \in V$. The bilinear form $a_B(\cdot,\cdot)$ induces a respective energy norm defined by $\|u\|_B^2 := a_B(u, u)$ for $u \in V$. Since T and S are bounded linear operators, it is easy to see that a_B is bounded as well. The definition of a_B allows us to give the following simple condition for existence and uniqueness of solutions to (3.5).

Proposition 3.2. *If a_B is coercive, then* (3.5) *has a unique solution $\hat{u} \in V$. If additionally $\lambda > 0$, then $\hat{u} \in V \cap BV(\Omega)$.*

Proof. We denote by F the functional from (3.5) and aim to apply the direct method from Theorem 2.22. Since it is clear that F is proper by having a lower bound of 0 and satisfying $F(0) < \infty$, it remains to check that F is weakly coercive and lower semi-continuous.

Since $T : V \to L^2(\Omega)$ is bounded, $V \to \mathbb{R}, u \mapsto \alpha_1 \|Tu - g\|_{L^1} + \frac{\alpha_2}{2}\|Tu - g\|_{L^2}^2$ is continuous and due to convexity weakly lower semi-continuous, see Lemma 2.21. By the same argument, since $S : V \to V_S \in \{L^2(\Omega)^m, L^2(\Omega)^{d\times m}\}$ is bounded, $V \to \mathbb{R}, u \mapsto \frac{\beta}{2}\|Su\|_{L^2}^2$ is weakly lower semi-continuous. The total variation is weakly lower semi-continuous on $L^1(\Omega)^m$, see Proposition 2.30, and in particular on V because of the continuous embeddings $V \subseteq L^2(\Omega)^m \subseteq L^1(\Omega)^m$. In total, $F : V \to \overline{\mathbb{R}}$ is weakly lower semi-continuous.

Since a_B is coercive we know that for $\|u\|_V \to \infty$ we have $F(u) \geq \frac{1}{2}a_B(u,u) \to \infty$. Therefore $\inf_{u\in K} F(u) = \inf_{u\in V} F(u)$ for some sufficiently large bounded closed convex set $K \subseteq V$. Since V is reflexive, K is weakly compact and the existence of a minimizer $\hat{u} \in V$ now follows from Theorem 2.22.

For uniqueness, we note that F is strongly convex since a_B is coercive and we may write $F(u) = \frac{1}{2}a_B(u,u) + \alpha_1\|Tu - g\|_{L^1} - \alpha_2\langle Tu, g\rangle_{L^2} + \|g\|_{L^2}^2 + \lambda\,\mathrm{TV}(u)$ with all other terms being convex. Following the standard argument, assuming there are two different minimizers $u, v \in V$, $u \neq v$ of F with minimum $\hat{F}$ we have in particular $\frac{u+v}{2} \in V$ with

$$F(\tfrac{u+v}{2}) < \tfrac{1}{2}F(u) + \tfrac{1}{2}F(v) = \hat{F},$$

which contradicts the assumption. Therefore the minimizer $\hat{u}$ of F must be unique.

Since $0 \in V$ has finite energy $F(0)$, the minimizer $\hat{u}$ will have finite energy as well and in particular finite total variation $\int_\Omega |D\hat{u}|_F$ if $\lambda > 0$. In this case we conclude $\hat{u} \in V \cap BV(\Omega)^m$ since $\hat{u} \in V \subseteq L^1(\Omega)^m$. □

Specifically for our two main choices $S \in \{I, \nabla\}$ we can describe coercivity of the bilinear form a_B in slightly more explicit terms as given by the following proposition.

Proposition 3.3. *The bilinear form $a_B : V \times V \to \mathbb{R}$ is coercive in any of the following cases:*

(i) $\alpha_2 > 0$ *and* $T = I$,

(ii) $\beta > 0$ *and* $S = I$, *or*

(iii) $\beta > 0$, $S = \nabla$ *and* $1 \notin \ker T$.

Proof. If $T = I$ with $\alpha_2 > 0$ or $S = I$ with $\beta > 0$ we see directly

$$a_B(v, v) = \alpha_2 \|Tv\|_{L^2}^2 + \beta \|Sv\|_{L^2}^2 \geq \max\{\alpha_2, \beta\} \|v\|_V^2$$

for all $v \in V$, which shows the statement for items (i) and (ii).

For item (iii) with $S = \nabla$ we show coercivity in $H^1(\Omega)^m$ from which coercivity in the subspace V follows immediately. Split $u \in H^1(\Omega)^m$ into $u = v + w$ with $w_i := \frac{1}{|\Omega|} \int_\Omega u_i(y)\, \mathrm{d}y$ being the componentwise mean and $v \in H^1(\Omega)^m$ such that $\int_\Omega v_i(x)\, \mathrm{d}x = 0$ for $i = 1, \dots, m$. Due to the Poincaré-Wirtinger inequality, see Lemma 2.11 we have

$$\begin{aligned} \|u\|_{H^1(\Omega)^m}^2 &= \|v + w\|_{L^2}^2 + \|\nabla v\|_{L^2}^2 \\ &\leq \|v\|_{L^2}^2 + 2\|v\|_{L^2} \|w\|_{L^2} + \|w\|_{L^2}^2 + \|\nabla v\|_{L^2}^2 \\ &\leq 2\|w\|_{L^2}^2 + 2\|v\|_{L^2}^2 + \|\nabla v\|_{L^2}^2 \\ &\leq 2\|w\|_{L^2}^2 + c_1 \|\nabla v\|_{L^2}^2 \end{aligned} \tag{3.8}$$

for a constant $c_1 > 0$, where we used $0 \leq (a - b)^2 = a^2 - 2ab + b^2$, $a, b \geq 0$ to obtain the second inequality. Because the operator T cannot annihilate constant functions, there is $c_T > 0$ independent of w such that $\|Tw\|_{L^2} \geq c_T \|w\|_{L^2}$. This means that if $\|w\|_{L^2} \geq 2c_T^{-1} \|T\|_{\mathcal{L}(L^2, L^2)} \|v\|_{L^2}$, then

$$\|Tu\|_{L^2} = \|Tw + Tv\|_{L^2} \geq c_T \|w\|_{L^2} - \|T\|_{\mathcal{L}(L^2,L^2)} \|v\|_{L^2} \geq \tfrac{c_T}{2} \|w\|_{L^2}.$$

This together with (3.8) yields

$$\begin{aligned} \|u\|_{H^1(\Omega)^m}^2 &\leq 2\|w\|_{L^2}^2 + c_1 \|\nabla v\|_{L^2}^2 \\ &\leq \tfrac{8}{c_T^2} \|Tu\|_{L^2}^2 + c_1 \|\nabla u\|_{L^2}^2 \leq c_2 a_B(u, u) \end{aligned}$$

for some constant $c_2 > 0$.

If on the other hand $\|w\|_{L^2} < 2c_T^{-1} \|T\|_{\mathcal{L}(L^2,L^2)} \|v\|_{L^2}$ then (again using the Poincaré-Wirtinger inequality, see Lemma 2.11) we have

$$\|w\|_{L^2} < 2c_T^{-1} \|T\|_{\mathcal{L}(L^2,L^2)} \|v\|_{L^2} \leq c_3 \|\nabla v\|_{L^2}$$

for some constant $c_3 > 0$ and hence

$$\begin{aligned}\|u\|^2_{H^1(\Omega)^m} &\le 2\|w\|^2_{L^2} + c_1\|\nabla v\|^2_{L^2}\\ &\le (c_1 + 2c_3^2)\|\nabla u\|^2_{L^2} \le \tfrac{c_1+2c_3^2}{\beta} a_B(u,u)\end{aligned}$$

which concludes coercivity of a_B for item (iii). □

Note that injectivity of $\alpha_2 T$ suffices to guarantee that B is invertible but does not necessarily imply coercivity of a_B.

From now on we will assume coercivity of a_B and due to Theorem 2.10 invertibility of $B = \alpha_2 T^* T + \beta S^* S : V \to V^*$ in particular.

(A1) The bilinear form $a_B : V \times V \to \mathbb{R}$ is coercive.

While this is not required for dualization in itself, it will allow us to state the dual problem to (3.5) in a more explicit form in Theorem 3.4 and (3.12) using the inverse of B. Namely, we introduce the dual norm on V^* by $\|u^*\|^2_{B^{-1}} := \langle u^*, B^{-1}u^*\rangle_{V^*,V}$ for $u^* \in V^*$. Coercivity of a_B will also be useful later in showing other uniqueness properties such as Theorems 3.8 and 5.15.

3.1.4 Dualization in $H^1(\Omega)^m$

In this subsection we fix Setting $(S.\text{ii})$ with $V = H^1(\Omega)^m$ and aim to derive the dual problem to (3.5) which will later motivate the regularized predual formulation (3.12) in a more general setting. We recall from Proposition 2.33 that for this choice of V the total variation reduces to $\int_\Omega |Du|_F = \int_\Omega |\nabla u|_F \,\mathrm{d}x$.

Theorem 3.4. *Let $V = H^1(\Omega)^m$ and $W = W_1 \times W_2 = L^2(\Omega) \times L^2(\Omega)^{d\times m}$. Then the problem*

$$\begin{aligned}\inf_{\boldsymbol{p}=(p_1,\boldsymbol{p}_2)\in W^*} \ &\tfrac{1}{2}\|T^*p_1 + \nabla^*\boldsymbol{p}_2 + \alpha_2 T^* g\|^2_{B^{-1}} - \tfrac{\alpha_2}{2}\|g\|^2_{L^2}\\ &- \langle g, p_1\rangle + \chi_{|p_1|\le\alpha_1} + \chi_{|\boldsymbol{p}_2|_F\le\lambda},\end{aligned} \tag{3.9}$$

is dual to (3.5)*. Furthermore, solutions $u \in V$ and $\boldsymbol{p} \in W^*$ to* (3.5) *and* (3.9) *respectively are characterized by*

$$\begin{aligned} T^*p_1 + \nabla^* p_2 &= Bu - \alpha_2 T^* g, \\ |Tu - g|\, p_1 &= -\alpha_1 (Tu - g), \qquad |p_1| \le \alpha_1, \\ |\nabla u|_F\, \boldsymbol{p}_2 &= -\lambda \nabla u, \qquad\qquad |\boldsymbol{p}_2|_F \le \lambda. \end{aligned} \tag{3.10}$$

Proof. The proper, convex and lower semicontinuous functions $\mathcal{F} : V \to \overline{\mathbb{R}}$, $\mathcal{G} : W \to \overline{\mathbb{R}}$ and the linear operator $A : V \to W$ are set as follows:

$$\begin{aligned} \mathcal{F}(u) &:= \tfrac{\alpha_2}{2} \|Tu - g\|_{L^2}^2 + \tfrac{\beta}{2} \|Su\|_{L^2}^2, \\ \mathcal{G}(Au) &:= \alpha_1 \|Tu - g\|_{L^1} + \lambda \int_\Omega |\nabla u|_F \,\mathrm{d}x, \quad A := (T, \nabla). \end{aligned}$$

Using the definition of the conjugate function, we compute $\mathcal{F}^*$ and $\mathcal{G}^*$. We have:

$$\begin{aligned} \mathcal{F}^*(u^*) &= \sup_{u \in V} \left\{ \langle u^*, u \rangle_{V^*,V} - \tfrac{\alpha_2}{2} \|Tu - g\|_{L^2}^2 - \tfrac{\beta}{2} \|Su\|_{L^2}^2 \right\} \\ &= \sup_{u \in V} \Big\{ \langle u^* + \alpha_2 T^* g, u \rangle_{V^*,V} \\ &\qquad\qquad - \tfrac{1}{2} \langle (\alpha_2 T^*T + \beta S^*S) u, u \rangle_{V^*,V} - \tfrac{\alpha_2}{2} \|g\|_{L^2}^2 \Big\} \end{aligned}$$

A function $u \in V$ is a supremum of the above set if

$$0 = \partial_u \{ \langle u^*, u \rangle_{V^*,V} - \mathcal{F}(u) \} = u^* + \alpha_2 T^* g - (\alpha_2 T^*T + \beta S^*S) u.$$

and hence the supremum is obtained at

$$u = (\alpha_2 T^*T + \beta S^*S)^{-1} (u^* + \alpha_2 T^* g) = B^{-1} (u^* + \alpha_2 T^* g).$$

Thus we obtain an explicit formulation for $\mathcal{F}^* : V^* \to \overline{\mathbb{R}}$ as

$$\begin{aligned} \mathcal{F}^*(u^*) &= \tfrac{1}{2} \big\langle u^* + \alpha_2 T^* g, B^{-1} (u^* + \alpha_2 T^* g) \big\rangle_{V^*,V} - \tfrac{\alpha_2}{2} \|g\|_{L^2}^2 \\ &= \tfrac{1}{2} \|u^* + \alpha_2 T^* g\|_{B^{-1}}^2 - \tfrac{\alpha_2}{2} \|g\|_{L^2}^2. \end{aligned}$$

For the computation of $\mathcal{G}^*$ we split according to Proposition 2.26:

$$\mathcal{G}^*(\boldsymbol{v}^*) = \mathcal{G}_1^*(v_1^*) + \mathcal{G}_2^*(\boldsymbol{v}_2^*),$$

with $\mathcal{G}_1(v_1) := \alpha_1 \|v_1 - g\|_{L^1}$, $\mathcal{G}_2(\boldsymbol{v}_2) := \lambda \|\boldsymbol{v}_2\|_{L^1}$. Then we have

$$\begin{aligned} \mathcal{G}_1^*(v_1^*) &= \sup_{v_1 \in L^2(\Omega)} \{\langle v_1, v_1^* \rangle - \alpha_1 \|v_1 - g\|_{L^1}\} \\ &= \sup_{v_1' = v_1 - g \in L^2(\Omega)} \Big\{ \int_\Omega v_1^* \cdot v_1' - \alpha_1 |v_1'| + v_1^* \cdot g \, \mathrm{d}x \Big\} \\ &= \langle g, v_1^* \rangle + \chi_{|v_1^*| \leq \alpha_1}. \end{aligned}$$

Analogously we find

$$\begin{aligned} \mathcal{G}_2^*(\boldsymbol{v}_2^*) &= \sup_{\boldsymbol{v}_2 \in L^2(\Omega)^{d \times m}} \Big\{ \langle \boldsymbol{v}_2, \boldsymbol{v}_2^* \rangle - \lambda \int_\Omega |\boldsymbol{v}_2(x)|_F \, \mathrm{d}x \Big\} \\ &= \begin{cases} 0 & \text{if } |\boldsymbol{v}_2^*(x)|_F \leq \lambda, \\ \infty & \text{if } |\boldsymbol{v}_2^*(x)|_F > \lambda. \end{cases} \end{aligned}$$

Combining these calculations we obtain

$$\mathcal{G}^*(\boldsymbol{v}) = \langle g, v_1^* \rangle + \chi_{|v_1^*| \leq \alpha} + \chi_{|v_2^*|_F \leq \lambda}.$$

Applying the Fenchel duality from Theorem 2.28 yields the corresponding dual formulation and the optimality conditions $A^* p \in \partial \mathcal{F}(u)$ and $-p \in \partial \mathcal{G}(Au)$. The former reads

$$T^* p_1 + \nabla^* p_2 = \alpha_2 T^*(Tu - g) + \beta S^* S u = Bu - \alpha_2 T^* g.$$

The latter resolves pointwise to

$$\begin{aligned} -p_1 &= \alpha_1 \tfrac{Tu - g}{|Tu - g|}, \\ -\boldsymbol{p}_2 &= \lambda \tfrac{\nabla u}{|\nabla u|_F}, \end{aligned}$$

whenever $Tu - g \neq 0$ or $\nabla u \neq 0$ respectively and $|p_1| \leq \alpha_1$ or $|\boldsymbol{p}_2|_F \leq \lambda$ otherwise. Equivalently one has $|p_1| \leq \alpha_1$, $|\boldsymbol{p}_2|_F \leq \lambda$ a.e. on Ω and

$$\begin{aligned} |Tu - g| \, p_1 &= -\alpha_1 (Tu - g), \\ |\nabla u|_F \, \boldsymbol{p}_2 &= -\lambda \nabla u. \end{aligned}$$

□

Note that (3.10) is a relation in the dual space V^* and the term ∇^* may be understood as $\nabla^* : L^2(\Omega)^{d\times m} \to V^*, \boldsymbol{p} \mapsto (w \mapsto \langle \boldsymbol{p}, \nabla w\rangle)$. Further, equation (3.10) can be rewritten using the bilinear form a_B from equation (3.6) as

$$-\langle p_1, Tv\rangle - \langle \boldsymbol{p}_2, \nabla v\rangle = a_B(u,v) - l(v) \qquad \forall v \in V, \tag{3.11}$$

where $l(v) := \alpha_2\langle g, Tv\rangle$.

3.2 Regularized Model

The dual problem (3.9) is convex but does not necessarily have a unique solution due to the nontrivial kernel of ∇^*. For us to be able to enforce a unique solution, we slightly modify the objective function in (3.9) by adding terms $\frac{\gamma_1}{2\alpha_1}\|p_1\|_{L^2}^2$ and $\frac{\gamma_2}{2\lambda}\|\boldsymbol{p}_2\|_{L^2}^2$ with $\gamma_1, \gamma_2 \geq 0$. Setting $\gamma_1, \gamma_2 > 0$ will then guarantee strong convexity of the dual energy (see Lemma 3.7 below), which will become essential for solution algorithms taking advantage of that property, e.g. Theorem 5.15. Similar to [62], adding these two terms corresponds to using a Huber-type function in the primal problem, which will be made explicit by Theorem 3.5 and Proposition 3.9.Additionally, compared to the motivation by Theorem 3.4 in a smooth setting, we will generalize the space V to allow for discontinuous functions as originally intended by (3.5).

3.2.1 Predual Problem and Dualization

Let $V \subseteq L^2(\Omega)^m$ be as in (3.5). We aim to choose W as a Hilbert space such that the linear operator $\Lambda := (T, \nabla) : V \to W = (W_1, W_2)$, corresponding to A in the proof of Theorem 3.4, remains bounded. In particular we restrict ourselves to $W_1 \subseteq L^2(\Omega)$ and the following choices for $\nabla : V \to W_2$ and its corresponding spaces:

(∇.i) $V \subseteq H^1(\Omega)^m$ allowing for Settings (S.i) and (S.ii) and $W_2 \subseteq L^2(\Omega)^{d\times m}$,

(∇.ii) $V \subseteq H_0^1(\Omega)^m$ allowing for Settings (S.i) and (S.ii) and $W_2 \subseteq L^2(\Omega)^{d\times m}$,

(∇.iii) $V \subseteq L^2(\Omega^m)$ with Setting (S.i), and $W_2 \subseteq (H_0^{\mathrm{div}}(\Omega)^m)^*$, by defining $\nabla : u \mapsto (p \mapsto \langle u, -\operatorname{div} p\rangle)$.

Note that for Settings (∇.ii) and (∇.iii) we have $\nabla^* = -\operatorname{div}$ due to vanishing boundary terms, while for Setting (∇.i) this is not necessarily true.

Using $\gamma_1, \gamma_2 \geq 0$ we propose the following regularized dual problem:

$$\inf_{\boldsymbol{p}=(p_1,\boldsymbol{p}_2)\in W^*} \Big\{ \tfrac{1}{2}\big\|\Lambda^*\boldsymbol{p} - \alpha_2 T^* g\big\|_{B^{-1}}^2 - \tfrac{\alpha_2}{2}\|g\|_{L^2}^2 \\ + \langle g, p_1\rangle + \chi_{|p_1|\leq\alpha_1} + \tfrac{\gamma_1}{2\alpha_1}\|p_1\|_{L^2}^2 \\ + \chi_{|\boldsymbol{p}_2|_F\leq\lambda} + \tfrac{\gamma_2}{2\lambda}\|\boldsymbol{p}_2\|_{L^2}^2 =: E^*(\boldsymbol{p}) \Big\}, \tag{3.12}$$

Note that if $\alpha_1 = 0$, then it follows immediately that $p_1 = 0$ due to the box-constraint $\chi_{|p_1|\leq\alpha_1}$. Analogously if $\lambda = 0$, then $\boldsymbol{p}_2 = 0$. In these cases we use the convention that the terms $\frac{\gamma_1}{2\alpha_1}\|p_1\|_{L^2}^2$ and $\frac{\gamma_2}{2\lambda}\|\boldsymbol{p}_2\|_{L^2}^2$ vanish respectively. This convention both makes sense as a continuous extension of the limit process $\alpha_1, \lambda \to 0$ and agrees with setting $\alpha_1, \lambda = 0$ prior to dualization.

Theorem 3.5. *The dual problem to* (3.12) *reads*

$$\inf_{u\in V} \Big\{ \tfrac{\alpha_2}{2}\|Tu - g\|_{L^2}^2 + \tfrac{\beta}{2}\|Su\|_{L^2}^2 \\ + F_1^*(Tu) + F_2^*(\nabla u) =: E(u) \Big\} \tag{3.13}$$

where F_1^*, F_2^* are the convex conjugates to $F_1 : W_1^* \to \overline{\mathbb{R}}$, $F_2 : W_2^* \to \overline{\mathbb{R}}$ given by

$$F_1(p_1) := \langle g, p_1\rangle + \chi_{|p_1|\leq\alpha_1} + \tfrac{\gamma_1}{2\alpha_1}\|p_1\|_{L^2}^2,$$
$$F_2(\boldsymbol{p}_2) := \chi_{|\boldsymbol{p}_2|_F\leq\lambda} + \tfrac{\gamma_2}{2\lambda}\|\boldsymbol{p}_2\|_{L^2}^2.$$

Furthermore solutions $p = (p_1, \boldsymbol{p}_2) \in W^*$, $u \in V$ of (3.12) *and* (3.13) *respectively are characterized by*

$$\begin{aligned} 0 &= \Lambda^*\boldsymbol{p} - \alpha_2 T^* g + Bu, \\ Tu &\in \partial F_1(p_1), \\ \nabla u &\in \partial F_2(\boldsymbol{p}_2). \end{aligned} \tag{3.14}$$

Proof. We again use the Fenchel duality from Theorem 2.28, choosing $\mathcal{F} : W^* \to \overline{\mathbb{R}}$, $\mathcal{G} : V^* \to \overline{\mathbb{R}}$ and $A : W^* \to V^*$ as follows

$$\begin{aligned}
\mathcal{F}(\boldsymbol{p}) &:= F_1(p_1) + F_2(\boldsymbol{p}_2) \\
&= \langle g, p_1\rangle + \chi_{|p_1|\le\alpha_1} + \chi_{|\boldsymbol{p}_2|_F\le\lambda} + \tfrac{\gamma_1}{2\alpha_1}\|p_1\|_{L^2}^2 + \tfrac{\gamma_2}{2\lambda}\|\boldsymbol{p}_2\|_{L^2}^2 \\
\mathcal{G}(A\boldsymbol{p}) &:= \tfrac{1}{2}\|A\boldsymbol{p} - \alpha_2 T^* g\|_{B^{-1}}^2 - \tfrac{\alpha_2}{2}\|g\|_{L^2}^2, \\
A\boldsymbol{p} &:= \Lambda^*\boldsymbol{p} = T^* p_1 + \nabla^*\boldsymbol{p}_2.
\end{aligned}$$

For $\mathcal{G}^*$ we get by the definition of the convex conjugate

$$\begin{aligned}
\mathcal{G}^*(u) = \sup_{v\in V^*} \Big\{ &\langle v, u\rangle_{V^*,V} - \tfrac{1}{2}\big\langle v - \alpha_2 T^* g, B^{-1}(v - \alpha_2 T^* g)\big\rangle_{V^*,V} \\
&+ \tfrac{\alpha_2}{2}\|g\|_{L^2}^2 \Big\}
\end{aligned}$$

where the supremum is attained whenever

$$0 = \partial_v\big(\langle v, u\rangle_{V^*,V} - \mathcal{G}(v)\big) = u - B^{-1}(v - \alpha_2 T^* g),$$

which implies $v = Bu + \alpha_2 T^* g$. Hence we have

$$\begin{aligned}
\mathcal{G}^*(u) &= \langle Bu + \alpha_2 T^* g, u\rangle_{\hat{W},\hat{W}^*} - \tfrac{1}{2}\langle B^{-1}Bu, Bu\rangle_{\hat{W}^*,\hat{W}} + \tfrac{\alpha_2}{2}\|g\|_{L^2}^2 \\
&= \langle u, Bu\rangle_{\hat{W},\hat{W}^*} + \langle u, \alpha_2 T^* g\rangle - \tfrac{1}{2}\langle u, Bu\rangle_{\hat{W},\hat{W}^*} + \tfrac{\alpha_2}{2}\|g\|_{L^2}^2 \\
&= \tfrac{1}{2}\langle u, (\alpha_2 T^*T + \beta S^*S)u\rangle_{V,V*} + \langle u, \alpha_2 T^* g\rangle + \tfrac{\alpha_2}{2}\|g\|_{L^2}^2 \\
&= \tfrac{\alpha_2}{2}\langle Tu, Tu\rangle + \tfrac{\beta}{2}\langle Su, Su\rangle + \alpha_2\langle Tu, g\rangle + \tfrac{\alpha_2}{2}\|g\|_{L^2}^2 \\
&= \tfrac{\alpha_2}{2}\|Tu + g\|_{L^2}^2 + \tfrac{\beta}{2}\|Su\|_{L^2}^2.
\end{aligned}$$

For F^*, since F is separable in p_1 and $\boldsymbol{p}_2$, we only apply Proposition 2.26 without resolving F_1^* and F_2^* explicitly. The optimality conditions in Theorem 2.28 correspond to $\Lambda u \in \partial F(\boldsymbol{p})$ and $-u = B^{-1}(\Lambda^*\boldsymbol{p} - \alpha_2 T^* g)$ which yield (3.14). □

Theorem 3.5 established the duality of (3.12) and (3.13) based on the predual formulation (3.12) similar to the approach of [58]. It is, however, interesting to note that the spaces V and W used for dualization are reflexive and thus the Fenchel duality from Theorem 2.28 may be used to equivalently establish the duality of (3.13) and (3.12) based on the primal formulation (3.13) (the only difference being a change in sign as can be seen when comparing (3.9) with (3.12)).

Proposition 3.6. *If a_B is coercive, then* (3.13) *has a unique solution $\hat{u} \in V$. If additionally $\lambda > 0$, then $\hat{u} \in V \cap BV(\Omega)$.*

Proof. The proof is similar to the one of Proposition 3.2. To see that $\hat{u} \in BV(\Omega)$ if $\lambda > 0$, we additionally refer to Proposition 3.14. □

The primal functional E from (3.13) satisfies the following strong convexity property.

Lemma 3.7. *If $\hat{u} \in V$ is a minimizer of E, then we have*

$$\tfrac{1}{2}\|u - \hat{u}\|_B^2 \leq E(u) - E(\hat{u})$$

for all $u \in V$.

Proof. We apply the same method as in [13, Lemma 10.2]. For the energy E from (3.13) we write $E(u) = F(u) + G(u)$ with

$$\begin{aligned} G(u) &:= F_1^*(Tu) + F_2^*(\nabla u), \\ F(u) &:= \tfrac{\alpha_2}{2}\|Tu - g\|_{L^2}^2 + \tfrac{\beta}{2}\|Su\|_{L^2}^2. \end{aligned}$$

Note, that F is Frechet-differentiable with

$$\langle F'(u), w\rangle_V = \alpha_2\langle Tu - g, Tw\rangle + \beta\langle Su, Sw\rangle$$

for all $w \in V$. Expanding $F(u)$ at $\hat{u}$ then yields

$$F(u) = F(\hat{u}) + \langle F'(\hat{u}), u - \hat{u}\rangle_V + \tfrac{1}{2}\|u - \hat{u}\|_B^2.$$

Since $\hat{u} \in V$ is a minimizer, we have $0 \in \partial E(\hat{u}) = F'(\hat{u}) + \partial G(\hat{u})$. Hence $-F'(\hat{u}) \in \partial G(\hat{u})$, i.e.

$$\begin{aligned} & \langle -F'(\hat{u}), u - \hat{u}\rangle_V \leq G(u) - G(\hat{u}) \\ \iff \quad & \tfrac{1}{2}\|u - \hat{u}\|_B^2 \leq F(u) - F(\hat{u}) + G(u) - G(\hat{u}), \end{aligned}$$

which proves the assertion. □

Putting Lemma 3.7 and coercivity of a_B from Assumption (A1) together we obtain that for a minimizer u of E there is a constant $c > 0$ such that

$$E(v) - E(u) \geq \tfrac{1}{2}\|u - v\|_B^2 \geq \tfrac{c}{2}\|u - v\|_V^2$$

for all $v \in V$ as long as a_B is coercive.

Theorem 3.8. *Problem* (3.12) *has at least one solution* $\boldsymbol{p} \in W^*$*, which is unique if* $\gamma_1, \gamma_2 > 0$*.*

Proof. We aim to apply the direct method from Theorem 2.22 using the weak topology on W^*.

The functional $E^* : W^* \to \overline{\mathbb{R}}$ is proper since it is bounded from below and e.g. $E^*(0) < \infty$.

Further, since the linear operator $\Lambda^* : W^* \to V^*$ is bounded and $B^{-1} : V^* \to V$ is bounded as well due to coercivity of a_B, see Theorem 2.10, the term $p \mapsto \|\Lambda^* p - \alpha_2 T^* g\|_{B^{-1}}^2$ is continuous and due to convexity also weakly lower semi-continuous, see Lemma 2.21. Similarly, the terms $p \mapsto -\frac{\alpha_2}{2}\|g\|_{L^2}^2 + \langle g, p_1\rangle + \frac{\gamma_1}{2\alpha_1}\|p_1\|_{L^2}^2 + \frac{\gamma_2}{2\lambda}\|p_2\|_{L^2}^2$ are weakly lower semi-continuous. For the box-constraints the set $\tilde{K} := \{p \in L^2(\Omega) \times L^2(\Omega)^{d\times m} : |p_1| \leq \alpha_1, |\boldsymbol{p}_2|_F \leq \lambda\}$ is weakly closed in $L^2(\Omega) \times L^2(\Omega)^{d\times m}$ by an application of Lemma 2.5. With the continuous embedding $W^* \subseteq L^2(\Omega) \times L^2(\Omega)^{d\times m}$ we follow that $K := \tilde{K} \cap W^*$ must be weakly closed in W^*. Noticing that K defines the only non-trivial levelset of $p \mapsto \chi_{|p_1| \leq \alpha_1} + \chi_{|\boldsymbol{p}_2|_F \leq \lambda}$ we conclude by Proposition 2.19 that this term is weakly lower semi-continuous as well and as such E^* in total.

Since W^* is reflexive, to show that $E^* : W^* \to \overline{\mathbb{R}}$ is weakly coercive with regard to weak convergence, it is sufficient to show $\|\boldsymbol{p}\|_{W^*} \to \infty \implies E^*(\boldsymbol{p}) \to \infty$, see Lemma 2.2. Due to the box-constraints $\chi_{|p_1| \leq \alpha_1} + \chi_{|\boldsymbol{p}_2|_F \leq \lambda}$ it is easy to see that $\|\boldsymbol{p}\|_{L^2} \to \infty$ implies $E^*(\boldsymbol{p}) \to \infty$. It therefore remains to check the case when $W_2 \subseteq H_0^{\mathrm{div}}(\Omega)^m$ and $\|\operatorname{div} \boldsymbol{p}_2\|_{L^2} \to \infty$. Since a_B is coercive with coercivity constant $c_B > 0$, we have

$$\begin{aligned}\|v\|_{V^*}^2 &\leq \|B\|^2\|B^{-1}v\|_V^2\\ &\leq \tfrac{\|B\|^2}{c_B} a_B(B^{-1}v, B^{-1}v)\\ &\leq \tfrac{\|B\|^2}{c_B}\langle BB^{-1}v, B^{-1}v\rangle_{V^*,V} = \tfrac{\|B\|^2}{c_B}\|v\|_{B^{-1}}^2,\end{aligned}$$

which allows us to bound

$$\begin{aligned}E^*(\boldsymbol{p}) &\geq \|T^*p_1 - \operatorname{div}\boldsymbol{p}_2 - \alpha_2 T^* g\|_{B^{-1}}\\ &\geq \|\operatorname{div}\boldsymbol{p}_2\|_{B^{-1}} - \|T^*p_1 - \alpha_2 T^*g\|_{B^{-1}}\\ &\geq c_2\|\operatorname{div}\boldsymbol{p}_2\|_{V^*} - c_3 \to \infty\end{aligned}$$

for some constant $c > 0$ independent of $\boldsymbol{p}_2$, which shows coercivity of the functional E^*. The direct method from Theorem 2.22 then concludes the existence of a solution $\boldsymbol{p} \in W^*$.

Uniqueness in case $\gamma_1, \gamma_2 > 0$ follows from strict convexity in the terms $\frac{\gamma_1}{2\alpha_1}\|p_1\|_{L^2}^2$ and $\frac{\gamma_2}{2\lambda}\|\boldsymbol{p}_2\|_{L^2}^2$ similar to the proof of Proposition 3.2.

□

For special choices of V the regularized terms in the primal problem (3.13) may be formulated in a more explicit way. They form integral expressions similar to those of the non-regularized primal problem (3.5) but including a pointwise so-called Huber-smoothing of the integrand.

Proposition 3.9. *The terms $F_1^*(Tu)$ and $F_2^*(\nabla u)$ from Theorem 3.5 are called* Huber-regularized L^1 *and* Huber-regularized total variation *respectively and may take on the following explicit form:*

(i) $F_1^*(Tu) = \alpha_1 \int_\Omega \varphi_{\gamma_1}(|Tu - g|)\,\mathrm{d}x$ *if* $V \in \{H^1(\Omega)^m, L^2(\Omega)^m\}$,

(ii) $F_2^*(\nabla u) = \lambda \int_\Omega \varphi_{\gamma_2}(|\nabla u|_F)\,\mathrm{d}x$ *if* $V = H^1(\Omega)^m$,

where the Huber-function $\varphi_\gamma : \mathbb{R} \to [0,\infty)$ *for* $\gamma \geq 0$ *is defined by*

$$\varphi_\gamma(x) := \begin{cases} \frac{1}{2\gamma}x^2 & \text{if } |x| \leq \gamma,\\ |x| - \frac{\gamma}{2} & \text{if } |x| > \gamma.\end{cases} \tag{3.15}$$

In particular, if $V = H^1(\Omega)^m$, *then the optimality conditions* (3.14) *may be written as*

$$\begin{aligned}
&0 = \Lambda^* \boldsymbol{p} - \alpha_2 T^* g + Bu, && \\
&0 = p_1 \max\{\gamma_1, |Tu - g|\} - \alpha_1(Tu - g), && |p_1| \le \alpha_1 \qquad (3.16)\\
&0 = \boldsymbol{p}_2 \max\{\gamma_2, |\nabla u|_F\} - \lambda \nabla u, && |\boldsymbol{p}_2|_F \le \lambda
\end{aligned}$$

where max *denotes the pointwise maximum.*

Proof. We have

$$F_1^*(q_1) = \sup_{p_1 \in W_1^*} \Big\{ \langle p_1, q_1 \rangle_{W_1^*, W_1} - \langle p_1, g \rangle - \chi_{|p_1| \le \alpha_1} - \tfrac{\gamma_1}{2\alpha_1} \|p_1\|_{L^2}^2 \Big\}.$$

A function p_1 is a supremum of this set if $|p_1| \le \alpha_1$ with $q_1 - g - \frac{\gamma_1}{\alpha_1} p_1 = 0$ and hence $p_1 = \frac{\alpha_1}{\gamma_1}(q_1 - g)$. Thus we have $|q_1 - g| = \gamma_1 \frac{|p_1|}{|\alpha_1|} \le \gamma_1$ and we deduce

$$\begin{aligned}
F_1^*(q_1) &= \int_{|q_1 - g| \le \gamma_1} \tfrac{\alpha_1}{2\gamma_1} |q_1 - g|^2 \,\mathrm{d}x + \int_{|q_1 - g| > \gamma_1} \alpha_1 |q_1 - g| - \tfrac{\alpha_1 \gamma_1}{2} \,\mathrm{d}x \\
&= \alpha_1 \int_\Omega \varphi_{\gamma_1}(|q_1(x) - g(x)|) \,\mathrm{d}x.
\end{aligned}$$

For the conjugate F_2^* of F_2 we get

$$F_2^*(\boldsymbol{q}_2) = \sup_{\boldsymbol{p}_2 \in W_2^*} \Big\{ \langle \boldsymbol{p}_2, \boldsymbol{q}_2 \rangle_{W_2^*, W_2} - \chi_{|\boldsymbol{p}_2|_F \le \lambda} - \tfrac{\gamma_2}{2\lambda} \|\boldsymbol{p}_2\|_{L^2}^2 \Big\}.$$

After scaling with $\frac{1}{\lambda}$, i.e., substituting $\boldsymbol{w} := \frac{\boldsymbol{p}_2}{\lambda}$, we obtain

$$F_2^*((\nabla^*)^* u) = F_2^*(\nabla u) = \lambda \sup_{\substack{\boldsymbol{w} \in W_2^* \\ |\boldsymbol{w}|_F \le 1}} \Big\{ \int_\Omega \nabla u \cdot \boldsymbol{w} - \tfrac{\gamma_2}{2} |\boldsymbol{w}|_F^2 \,\mathrm{d}x \Big\}. \qquad (3.17)$$

The pointwise constrained maximization problem on the right hand side yields the Karush-Kuhn-Tucker (KKT) conditions

$$\begin{aligned}
\nabla u - \gamma_2 \boldsymbol{w} - 2\mu \boldsymbol{w} &= 0, & |\boldsymbol{w}|_F^2 - 1 &\le 0, \\
\mu(|\boldsymbol{w}|_F^2 - 1) &= 0, & \mu &\ge 0.
\end{aligned}$$

Assuming $\gamma_2 > 0$ implies $\gamma_2 + 2\mu > 0$ and hence we have $\boldsymbol{w} = \frac{\nabla u}{\gamma_2+2\mu}$. If $|\boldsymbol{w}|_F < 1$ then $\mu = 0$ and hence we obtain $\boldsymbol{w} = \frac{\nabla u}{\gamma_2}$. Inserting this in (3.17) yields the integrand $\frac{1}{2\gamma_2}|\nabla u|_F^2$. If $|\boldsymbol{w}|_F = 1$ then we observe that $1 = |\boldsymbol{w}|_F = \frac{1}{\gamma_2+2\mu}|\nabla u|_F$ which leads to $\gamma_2 + 2\mu = |\nabla u|_F$ and thus $\boldsymbol{w} = \frac{\nabla u}{|\nabla u|_F}$. Inserting in (3.17) yields the integrand $|\nabla u|_F - \frac{\gamma_2}{2}$. Summarizing our findings we arrive at the integrand

$$\varphi_{\gamma_2}(|\nabla u|_F) = \begin{cases} \frac{1}{2\gamma_2}|\nabla u|_F^2 & \text{if } |\nabla u|_F < \gamma_2, \\ |\nabla u|_F - \frac{\gamma_2}{2} & \text{else} \end{cases}$$

and thus

$$F_2^*(\nabla u) = \lambda \int_\Omega \varphi_{\gamma_2}(|\nabla u|_F)\,\mathrm{d}x.$$

If $\gamma_2 = 0$, a similar argument shows that

$$F_2^*(\nabla u) = \lambda \int_\Omega |\nabla u|_F\,\mathrm{d}x = \lambda \int_\Omega \varphi_0(|\nabla u|_F)\,\mathrm{d}x.$$

To show that (3.14) can be written as in (3.16) if $V = H^1(\Omega)^m$, we derive from $L^2(\Omega)^{d\times m} \ni \nabla u \in \partial F_2(\boldsymbol{p}_2)$ for $\gamma_2 > 0$ that necessarily $|\boldsymbol{p}_2|_F \leq \lambda$ and pointwise

$$\begin{aligned} \nabla u &\in \begin{cases} \{\frac{\gamma_2}{\lambda}\boldsymbol{p}_2\} & \text{if } |\boldsymbol{p}_2|_F < \lambda, \\ \{\mu\boldsymbol{p}_2 : \mu \geq 0\} & \text{if } |\boldsymbol{p}_2|_F = \lambda, \end{cases} \\ \iff \qquad p_2 &= \begin{cases} \lambda\frac{\nabla u}{\gamma_2} & \text{if } |\nabla u|_F < \gamma_2, \\ \lambda\frac{\nabla u}{|\nabla u|_F} & \text{if } |\nabla u|_F \geq \gamma_2, \end{cases} \\ &= \lambda\frac{\nabla u}{\max\{\gamma_2, |\nabla u|_F\}}. \end{aligned}$$

For $\gamma_2 = 0$ the same argument applies, except for $\nabla u = 0$, in which case only $|\boldsymbol{p}_2|_F \leq \lambda$ holds. In any case, we can summarize for $\gamma_2 \geq 0$ that $\nabla u \in \partial F_2(\boldsymbol{p}_2)$ is indeed equivalent to

$$0 = p_2 \max\{\gamma_2, |\nabla u|_F\} - \lambda\nabla u, \qquad |\boldsymbol{p}_2|_F \leq \lambda.$$

For the representation $Tu \in \partial F_1(p_1)$ one may proceed analogously. □

3.2.2 Dual Characterization of the Huber-TV-Functional

In Proposition 3.9 we have seen the pointwise representation of the regularized primal total variation term $F_2^*(\nabla u)$ for $V = H^1(\Omega)^m$ by utilizing the Huber-function (3.15). We will now extend this representation to $V = L^2(\Omega)$ by means of a more generally defined *Huber-TV functional*, c.f. Definition 3.11. This functional has been subject of analysis in e.g. [22, 76] and is recognized to reduce the staircasing effect of the total variation [22].

Proposition 3.10. *The Huber-function* (3.15) *satisfies the following properties:*

(i) $0 \leq \gamma_- \leq \gamma_+ \implies \forall x \in \mathbb{R} : \varphi_{\gamma_-}(x) \geq \varphi_{\gamma_+}(x)$*,*

(ii) $\forall x \in \mathbb{R} : \lim_{\gamma \to 0^+} \varphi_\gamma(x) = \varphi_0(x) = |x|$*.*

(iii) $\forall x \in \mathbb{R} : |\varphi_\gamma'(x)| \leq 1$*,*

(iv) $\lim_{\gamma \to 0^+} \int_\Omega \varphi_\gamma(f(x))\,\mathrm{d}x = \int_\Omega |f(x)|\,\mathrm{d}x$ *for any* $f \in L^2(\Omega)$*.*

Proof. (i) We distinguish depending on $x \in \mathbb{R}$ the cases

$$\begin{aligned} |x| \leq \gamma_- \leq \gamma_+ :&\quad \tfrac{1}{2\gamma_+}x^2 \leq \tfrac{1}{2\gamma_-}x^2,\\ \gamma_- \leq |x| \leq \gamma_+ :&\quad |x| - \tfrac{\gamma_+}{2} \leq \tfrac{1}{2}|x| \leq \tfrac{1}{2\gamma_-}x^2,\\ \gamma_- \leq \gamma_+ \leq |x| :&\quad |x| - \tfrac{\gamma_+}{2} \leq |x| - \tfrac{\gamma_-}{2}. \end{aligned}$$

(ii) For $x = 0$ it is clear that $\varphi_\gamma(x) = |0|$. Otherwise one has $\varphi_\gamma(x) = |x| - \frac{\gamma}{2} \to |x|$ for any small $0 < \gamma < |x|$ and $x \in \mathbb{R}$.

(iii) We derive for $x \in \mathbb{R}$ directly

$$\varphi_\gamma'(x) = \begin{cases} \frac{1}{\gamma}x & \text{if } |x| \leq \gamma,\\ \operatorname{sgn}(x) & \text{if } |x| > \gamma. \end{cases}$$

In any case $|\varphi_\gamma'(x)| \leq 1$ for every $x \in \mathbb{R}$.

(iv) This is a direct consequence of items (i) and (ii). □

Definition 3.11 (Huber-TV-Functional, c.f. [76]). *For $u \in L^2(\Omega)^m$, $\gamma \geq 0$, $\lambda \in C(\overline{\Omega})$, $\lambda \geq 0$ we denote by*

$$\int_\Omega \lambda \varphi_\gamma(|Du|) := \sup_{\substack{\boldsymbol{w} \in C_0^\infty(\Omega)^{d\times m} \\ |\boldsymbol{w}|_F \leq \lambda}} \left\{ \langle u, -\operatorname{div} \boldsymbol{w} \rangle - \tfrac{\gamma}{2} \|\boldsymbol{w}\|_{L^2}^2 \right\} \tag{3.18}$$

the λ-weighted γ-regularized Huber-TV functional.

Definition 3.11, similarly to the total variation from Definition 2.29, supremizes over pointwise constrained functions in $C_0^\infty(\Omega)^{d\times m}$, while F_2^* does so over $H_0^{\text{div}}(\Omega)^m$. Though $C_0^\infty(\Omega)^{d\times m}$ is a dense subset of $H_0^{\text{div}}(\Omega)^m$, the equivalence of F_2^* and (3.18) is non-trivial in view of the pointwise constraints, c.f. [60]. This kind of equivalence was first claimed in [58], while the necessary argument was only sufficiently established later in [60].

We now aim to show that Definition 3.11 indeed matches up to F_2^* from Theorem 3.5 using a construction inspired from [60], but instead making use of the continuity of the projection operator in Hilbert spaces onto closed convex subsets.

Theorem 3.12. *Let $W_2^* \in \{H_0^{\text{div}}(\Omega)^m, L^2(\Omega)^{d\times m}\}$, $\lambda \geq 0$ with $\lambda_{min} := \inf_{x\in\Omega} \lambda(x) > 0$ and denote*

$$K_\lambda := \{\boldsymbol{p} \in W_2^* : |\boldsymbol{p}|_F \leq \lambda\}.$$

Then $\overline{K_\lambda \cap C_0^\infty(\Omega)^{d\times m}}^{\|\cdot\|_{W_2^*}} = K_\lambda$.

Proof. We show the statement for $W_2^* = H_0^{\text{div}}(\Omega)^m$, since the proof for $L^2(\Omega)^{d\times m}$ works analogously. Let $\boldsymbol{p}_2 \in K_\lambda \subseteq H_0^{\text{div}}(\Omega)^m$ and $(\boldsymbol{p}_{2,n})_{n\in\mathbb{N}} \subseteq C_0^\infty(\Omega)^{d\times m}$ with $\|\boldsymbol{p}_{2,n} - \boldsymbol{p}_2\|_{H_0^{\text{div}}(\Omega)^m} \to 0$ by making use of density.

The projection π_{K_λ} is closed and convex due to Lemma 2.5 and the continuous embedding $H_0^{\text{div}}(\Omega)^m \subseteq L^2(\Omega)^{d\times m}$. The projection π_{K_λ} from Lemma 2.8 then yields $\pi_{K_\lambda}(\boldsymbol{p}_{2,n}) \in K_\lambda$ with compact support $\operatorname{supp} \pi_{K_\lambda}(\boldsymbol{p}_{2,n})$ for every $n \in \mathbb{N}$. Let $\varepsilon > 0$. Since π_{K_λ} is continuous, choose $N \in \mathbb{N}$ such that $\|\pi_{K_\lambda}(\boldsymbol{p}_{2,n}) - \boldsymbol{p}_2\|_{H_0^{\text{div}}(\Omega)^m} < \frac{\varepsilon}{3}$ for all $n \geq N$.

Using the mollifiers from Proposition 2.7, since $\lambda \in C(\overline{\Omega})$, we have $\lambda * \varrho_{\delta'} \to \lambda$ uniformly on Ω for $\delta' \to 0$. Let $\delta > 0$ and choose $\delta' > 0$ sufficiently small such that $\sup_{x\in\Omega} |(\lambda * \varrho_{\delta'})(x) - \lambda(x)| < \delta$, $\operatorname{supp} \pi_{K_\lambda}(\boldsymbol{p}_{2,n}) \subseteq \Omega$ and $\|\pi_{K_\lambda}(\boldsymbol{p}_{2,n}) * \varrho_{\delta'} - \pi_{K_\lambda}(\boldsymbol{p}_{2,n})\|_{H_0^{\mathrm{div}}(\Omega)^m} < \frac{\varepsilon}{3}$.

Now, similarly to [60] by scaling with

$$\begin{aligned}\eta_\delta := (1 + \tfrac{\delta}{\lambda_{\min}})^{-1} \leq (1 + \tfrac{\delta}{\lambda})^{-1} &= \frac{\lambda}{\lambda + \delta} \\ \leq \frac{\lambda}{\lambda * \varrho_{\delta'}} \leq \frac{\lambda}{|\pi_{K_\lambda}(\boldsymbol{p}_{2,n}) * \varrho_{\delta'}|_F}\end{aligned}$$

we guarantee $\boldsymbol{q} := \eta_\delta \pi_{K_\lambda}(\boldsymbol{p}_{2,n}) * \varrho_{\delta'} \in K_\lambda \cap C_0^\infty(\Omega)^{d\times m}$, where $\eta_\delta \to 1$ as $\delta \to 0$. In total, by choosing δ sufficiently small we thus achieve

$$\begin{aligned}\|\boldsymbol{q} - \boldsymbol{p}_2\|_{H_0^{\mathrm{div}}(\Omega)^m} &\leq \|\eta_\delta \pi_{K_\lambda}(\boldsymbol{p}_{2,n}) * \varrho_{\delta'} - \pi_{K_\lambda}(\boldsymbol{p}_{2,n}) * \varrho_{\delta'}\|_{H_0^{\mathrm{div}}(\Omega)^m} \\ &\quad + \|\pi_{K_\lambda}(\boldsymbol{p}_{2,n}) * \varrho_{\delta'} - \pi_{K_\lambda}(\boldsymbol{p}_{2,n})\|_{H_0^{\mathrm{div}}(\Omega)^m} \\ &\quad + \|\pi_{K_\lambda}(\boldsymbol{p}_{2,n}) - \boldsymbol{p}_2\|_{H_0^{\mathrm{div}}(\Omega)^m} \\ &< \tfrac{\varepsilon}{3} + \tfrac{\varepsilon}{3} + \tfrac{\varepsilon}{3} = \varepsilon.\end{aligned}$$

This shows that any $\boldsymbol{p}_2 \in K_\lambda$ may be approximated by smooth functions in $K_\lambda \cap C_0^\infty(\Omega)^{d\times m}$, respecting the box-constraint, which concludes the proof. □

Note that compared to [60] where a domain scaling argument for star-shaped domains is used, which requires some regularity of Ω, the proof of Theorem 3.12 does not and may therefore generalize to domains with less regular boundary.

Corollary 3.13. *The term $F_2^*(\nabla u)$ from Theorem 3.5 is called* Huber-regularized total variation *and may take on the following explicit form:*

$$F_2^*(\nabla u) = \lambda \int_\Omega \varphi_{\gamma_2}(|Du|_F)$$

if $V \in \{H^1(\Omega)^m, L^2(\Omega)^m\}$.

Proof. The statement follows directly by applying Theorem 3.12. □

If $u \in H^1(\Omega)^m$, the Huber-TV functional degrades to the Lebesgue integral over Ω of the Huber function term $\varphi_{\gamma_2}(|\nabla u|_F)$ as we see in the following proposition.

Proposition 3.14. *The Huber-TV functional* (3.18) *satisfies the following properties*

(i) $u \in BV(\Omega)^m \iff \int_\Omega \varphi_\gamma(|Du|_F) < \infty$ *for any* $\gamma \geq 0$*,*

(ii) If $u \in H^1(\Omega)^m$ *then*

$$\int_\Omega \varphi_\gamma(|Du|_F) = \int_\Omega \varphi_\gamma(|\nabla u|_F)\,\mathrm{d}x,$$

where φ_γ*,* $\gamma \geq 0$ *in the second integral is the Huber-function* (3.15)*.*

(iii) $0 \leq \gamma_- \leq \gamma_+ \implies \int_\Omega \varphi_{\gamma_-}(|Du|_F) \geq \int_\Omega \varphi_{\gamma_+}(|Du|_F)$*,*

(iv) $\lim_{\gamma\to 0} \int_\Omega \varphi_\gamma(|Du|_F) = \int_\Omega |Du|_F$*.*

Proof. (i) Since $\boldsymbol{w}$ is box-constrained in the supremum from Definition 3.11 we can bound $\int_\Omega \varphi_\gamma(|Du|)$ from above and below:

$$\begin{aligned}\int_\Omega |Du|_F - c &\leq \sup_{\substack{\boldsymbol{w}\in C_0^\infty(\Omega)^{d\times m}\\ |\boldsymbol{w}|_F\leq 1}} \Big\{\langle u, \operatorname{div}\boldsymbol{w}\rangle - \tfrac{\gamma}{2}\|\boldsymbol{w}\|_{L^2}^2\Big\},\\ &\leq \int_\Omega |Du|_F\end{aligned}$$

where $c := \frac{\gamma}{2}|\Omega| < \infty$.

(ii) Using partial integration we get

$$\begin{aligned}\int_\Omega \varphi_\gamma(|Du|_F) &= \sup_{\substack{\boldsymbol{w}\in C_0^\infty(\Omega)^{d\times m}\\ |\boldsymbol{w}|_F\leq 1}} \Big\{\langle u, -\operatorname{div}\boldsymbol{w}\rangle - \tfrac{\gamma}{2}\|\boldsymbol{w}\|_{L^2}^2\Big\}\\ &= \sup_{\substack{\boldsymbol{w}\in C_0^\infty(\Omega)^{d\times m}\\ |\boldsymbol{w}|_F\leq 1}} \Big\{\int_\Omega \nabla u\cdot\boldsymbol{w} - \tfrac{\gamma}{2}|\boldsymbol{w}|_F^2\,\mathrm{d}x\Big\}.\end{aligned}$$

We may replace $C_0^\infty(\Omega)^{d\times m}$ by $L^2(\Omega)^{d\times m}$ due to Theorem 3.12. By the same pointwise consideration as in the proof of Proposition 3.9, we see that the supremum is attained for the Huber function integrand $\varphi_\gamma(|\nabla u|_F)$.

(iii) There exists a sequence $(w_n)_{n\in\mathbb{N}} \subseteq C_0^\infty(\Omega)^{d\times m}$, $|w_n|_F \leq 1$ such that

$$\begin{aligned}\int_\Omega \varphi_{\gamma_+}(|Du|_F) &= \lim_{n\to\infty}\Big(-\int_\Omega u\cdot \operatorname{div} w_n \,\mathrm{d}x - \tfrac{\gamma_+}{2}\|w_n\|_{L^2}^2\Big)\\ &\leq \lim_{n\to\infty}\Big(-\int_\Omega u\cdot \operatorname{div} w_n \,\mathrm{d}x - \tfrac{\gamma_-}{2}\|w_n\|_{L^2}^2\Big)\\ &\leq \int_\Omega \varphi_{\gamma_-}(|Du|_F).\end{aligned}$$

(iv) Because of strict monotonicity from (iii), the limit is achieved by the supremum

$$\begin{aligned}\lim_{\gamma\to 0}\int_\Omega \varphi_\gamma(|Du|_F) &= \sup_{\gamma>0}\sup_{|\boldsymbol{w}|_F\leq 1}\Big\{-\int_\Omega u\cdot\operatorname{div}\boldsymbol{w}\,\mathrm{d}x - \tfrac{\gamma}{2}\|\boldsymbol{w}\|_{L^2}^2\Big\}\\ &= \sup_{|\boldsymbol{w}|_F\leq 1}\sup_{\gamma>0}\Big\{-\int_\Omega u\cdot\operatorname{div}\boldsymbol{w}\,\mathrm{d}x - \tfrac{\gamma}{2}\|\boldsymbol{w}\|_{L^2}^2\Big\}\\ &= \int_\Omega |Du|_F.\end{aligned}$$ □

3.2.3 Γ-Convergence

We will now analyze how the minimizers of (3.13) behave for $\gamma := (\gamma_1, \gamma_2) \to 0$ by making use of Γ-convergence.

Lemma 3.15 (Sequential lower semi-continuity)**.** *The functional E defined in* (3.13) *is lower semi-continuous with regards to weak V-convergence.*

Proof. We show lower semi-continuity of each summand of E:

(i) The term $F_1^*(Tu)$ is per definition given by the supremum

$$F_1^*(Tu) = \sup_{\substack{p_1 \in L^2(\Omega) \\ |p_1| \le \alpha_1}} \Big\{ \langle Tu - g, p_1 \rangle_{L^2} - \tfrac{\gamma_1}{2\alpha_1} \|p_1\|_{L^2}^2 \Big\}.$$

Since the supremum of lower semi-continuous functions is lower semi-continuous due to Lemma 2.20, it suffices to show that $\tilde{F}_1 : V \to \mathbb{R}$, $\tilde{F}_1(u) := \langle Tu - g, p_1 \rangle_{L^2} - \frac{\gamma_1}{2\alpha_1} \|p_1\|_{L^2}^2$ is V-weakly lower semi-continuous for every fixed $p_1 \in L^2(\Omega)$, $|p_1| \le \alpha_1$. This, however, is imminent since both $T : V \to L^2(\Omega)$ and the inner product are V-weakly continuous.

(ii) Similarly, the term $F_2^*(\nabla u)$ is given by the supremum

$$F_2^*(\nabla u) = \sup_{\substack{\boldsymbol{p}_2 \in W_2^* \\ |\boldsymbol{p}_2|_F \le \lambda}} \Big\{ \langle u, -\operatorname{div} \boldsymbol{p}_2 \rangle_{L^2} - \tfrac{\gamma_2}{2\lambda} \|\boldsymbol{p}_2\|_{L^2}^2 \Big\}$$

and we conclude by the same argument.

(iii) Since the terms $\|Tu - g\|_{L^2}^2$ and $\|u\|_{L^2}^2$ are both convex and continuous in $u \in V$, they are also weakly lower semi-continuous.

(iv) For the term $\|Su\|_{L^2}^2$ we distinguish both possible choices of S. If $S = I : V \to V_S$, $V = V_S \subseteq L^2(\Omega)^m$, then $u \mapsto \|u\|_{L^2}^2$ is weakly continuous since it is both convex and continuous. If $S = \nabla : V \to V_S$, $V \subseteq H^1(\Omega)^m$, then $u \mapsto \|\nabla u\|_{L^2}^2$ is weakly continuous with the same argument since $\nabla : H^1(\Omega)^m \to L^2(\Omega)^{d \times m}$ is a continuous operator. □

The previous lemma together with the properties of the Huber-TV functional allows us to prove a Γ-convergence result for the functional E.

Lemma 3.16 (Gamma-convergence)**.** *Let* $(\gamma_1^j)_{j\in\mathbb{N}}, (\gamma_2^j)_{j\in\mathbb{N}} > 0$ *be monotonically decreasing sequences with* $\lim_{j\to\infty} \gamma_1^j = \lim_{j\to\infty} \gamma_2^j = 0$. *Denote by* $E^j : V \to \overline{\mathbb{R}}$ *the energy functional in* (3.13) *for* $(\gamma_1, \gamma_2) = (\gamma_1^j, \gamma_2^j)$ *for* $j \in \mathbb{N}$ *and* E^∞ *the functional in* (3.5). *Then* $\Gamma\text{-}\lim_{j\to\infty} E^j = E^\infty$ *with respect to weak* V*-convergence.*

Proof. By the monotonicity property of the Huber-TV-functional from Proposition 3.14 (iii), we observe that $E^j(u) \leq E^{j+1}(u)$ and $E^j(u) \to E^\infty(u)$ pointwise for every fixed $u \in V$. Further for every $j \in \mathbb{N}$ we have that E^j is (sequentially) weakly lower semi-continuous in V due to Lemma 3.15. According to Lemma 2.14 we thus have $\Gamma\text{-}\lim_{j\to\infty} E^j = \lim_j E^j = E^\infty$ with respect to weak V-convergence. □

Lemma 3.17 (Equi-coercivity)**.** *Let $\lambda > 0$ and $(E^j)_{j\in\mathbb{N}}$, $(\gamma_1^j)_{j\in\mathbb{N}}$, $(\gamma_2^j)_{j\in\mathbb{N}}$ as in Lemma 3.16. Then the sequence $(E^j)_j$ is equi-mildly coercive with regard to weak V-convergence, i.e. there exists a non-empty sequentially (with regard to weak V-convergence) compact set $K \subseteq V$ such that $\inf_V E^j = \inf_K E^j$ for all $j \in \mathbb{N}$.*

Proof. As E^j is proper for any $j \in \mathbb{N}$, i.e., there exist $u \in V$ such that $E^j(u) < \infty$, by coercivity of a_B from Assumption (A1) we obtain the coercivity of E^j in V for all $j \in \mathbb{N}$.

Denote by $L_a^j := \{u \in V : E^j(u) \leq a\}$, $a \in \mathbb{R}$ the lower level sets of E^j for $j \in \mathbb{N}$. The level sets L_a^j, $j \in \mathbb{N}$, are bounded due to coercivity of E^j shown above.

Since $E^j \leq E^{j+1}$ due to Proposition 3.14, the level sets L_a^j are nested for any fixed $a \in \mathbb{R}$, i.e. $L_a^j \supseteq L_a^{j+1}$, for $j \in \mathbb{N}$. Consequently $E^j \leq E^\infty$ and since $E^\infty(0) < \infty$ we may chose $a := E^\infty(0)$ to ensure $L_a^j \neq \emptyset$ for all $j \in \mathbb{N}$.

For all $j \in \mathbb{N}$ the minimizers of E^j exist in V (see Proposition 3.6) and are contained within some non-empty weakly closed ball $K \supseteq \overline{L_a^j}$ in V centred at the origin. Since V is reflexive K is weakly compact, concluding the proof. □

We are now ready to show our final main result, namely that for $\gamma \to 0$ minimizers of (3.13) approach the minimizer of (3.5).

Theorem 3.18. *Let $\lambda > 0$ and u^j, u^∞ denote the unique minimizers of E^j and E^∞ as given in Lemma 3.16 respectively for $j \in \mathbb{N}$. Then $u^j \rightharpoonup u^\infty$ for $j \to \infty$ with respect to weak V-convergence.*

Proof. As shown in the proof of Lemma 3.17 the minimizers $(u^j)_{j\in\mathbb{N}}$ are contained within a sequentially compact (with regard to weak V-convergence) set K. Then, according to Theorem 2.15, every weak limit of a subsequence of $(u^j)_{j\in\mathbb{N}}$ is a minimum point of E^∞. Since the minimum u^∞ of E^∞ is unique we have $u^j \rightharpoonup u^\infty$ for $j \to \infty$. □

4 Decomposition

In order to work with constrained main memory and make use of parallel computation, it is a standard technique to decompose a given optimization problem into smaller parts. In this chapter we explore two decomposition techniques, strikingly similar to the additive and multiplicative Schwarz methods, which are well known for partial differential equations [79].

Our contribution involves the extension of the algorithms found in [33] to our model, i.e. allowing a non-trivial operator B and vector-valued u, and an improved theoretical convergence proof. Namely, we only require approximate solutions of the local problems instead of exact ones and arrive at the same asymptotic convergence guarantee. Our convergence proof differs from [33] and in the special case of exact solvers and parallel decomposition our statement reduces to a known result from [75].

An initial extension proposal and preliminary work are due to Andreas Langer and have been used as a basis. Results of this chapter paired with corresponding numerical examples from Chapter 5 are in preparation to be published separately [55].

4.1 Introduction

Recall the regularized L^1-L^2-TV model from (3.13) in the special case $S = I$, $\alpha_1 = 0$, $\alpha_2 = 1$, $\gamma_1, \gamma_2 = 0$:

$$\inf_{u \in L^2(\Omega)^m \cap BV(\Omega)^m} \tfrac{1}{2}\|Tu - g\|_{L^2(\Omega)}^2 + \tfrac{\beta}{2}\|u\|_{L^2(\Omega)}^2 + \lambda \operatorname{TV}(u). \tag{4.1}$$

While (4.1) is convex, the total variation term makes the functional both non-smooth and non-additive with regard to spatial decomposition.

Decomposition algorithms which take advantage of these properties may thus not apply directly or only with limitations.

As already inferred from Chapter 3, there exists a predual problem which involves constrained minimization of a smooth functional. We will see that the pointwise constrained smooth formulation allows for an additive spatial decomposition, around which we can construct a corresponding decomposition algorithm.

Corollary 4.1 (c.f. Theorem 3.5)**.** *Problem* (4.1) *is dual to*

$$\inf_{p\in K} \left\{ D(p) := \tfrac{1}{2}\|\Lambda^* p - T^* g\|_{B^{-1}}^2 \right\}, \tag{4.2}$$

where $K := \{p \in H_0^{\operatorname{div}}(\Omega)^m : |p(x)|_F \le \lambda \text{ a.e.}\}$*,* $\Lambda^* : H_0^{\operatorname{div}}(\Omega)^m \to L^2(\Omega)^m$*,* $\Lambda^* p = \operatorname{div} p$*,* $B : L^2(\Omega)^m \to L^2(\Omega)^m$ *denotes the operator* $B := \alpha_2 T^* T + \beta I$ *and the norm is given by* $\|u^*\|_{B^{-1}}^2 := \langle u^*, B^{-1} u^*\rangle_{L^2}$ *for* $u^* \in L^2(\Omega)^m$*.*

The unique solution $\hat{u}$ *of* (4.1) *is related to any solution* $\hat{p}$ *of* (4.2) *by*

$$\hat{u} = B^{-1}(-\Lambda^*\hat{p} + T^* g) \quad \wedge \quad \forall p \in K : \langle \Lambda\hat{u}, p - \hat{p}\rangle_{V^*,V} \le 0. \tag{4.3}$$

Proof. We apply Theorem 3.5, omit the first dual variable p_1, since $\alpha_1 = 0$ and thus $p_1 = 0$ is fixed and discard the constant additive term $-\frac{\alpha_2}{2}\|g\|_{L^2}^2$. □

Corollary 4.1 allows one to solve for $\hat{p}$ in the predual domain of (4.2) and to later assemble the original solution $\hat{u}$ using the optimality relation (4.3).

Similar to Lemma 3.7 the error with respect to the L^2 norm can be related to the difference in the predual energy as follows.

Proposition 4.2. *Let* $\hat{p} \in H_0^{\operatorname{div}}(\Omega)^m$ *be a minimizer of* (4.2) *and* $\hat{u} \in L^2(\Omega)^m$ *be the minimizer of* (4.1)*. If the bilinear form* $a_B : L^2(\Omega)^m \times L^2(\Omega)^m \to \mathbb{R}$*,* $a_B(u, v) = \alpha_2\langle Tu, Tv\rangle + \beta\langle u, v\rangle$ *from* (3.6) *is coercive, i.e.* $a_B(u, u) \ge c_B\|u\|^2$ *with coercivity constant* $c_B > 0$ *(e.g. due to Proposition 3.3 with* $c_B = \beta > 0$*) then for all* $p \in H_0^{\operatorname{div}}(\Omega)$ *and* $u := B^{-1}(-\Lambda^* p + T^* g)$ *we have*

$$\tfrac{c_B}{2}\|u - \hat{u}\|^2 \le D(p) - D(\hat{p}).$$

Proof. Due to coercivity of a we have for $v \in L^2(\Omega)^m$

$$\begin{aligned} c_B \|B^{-1}v\|^2 &\leq a_B(B^{-1}v, B^{-1}v) \\ &= \big\langle (T^*T + \beta I)B^{-1}v, B^{-1}v \big\rangle_V \\ &= \langle v, B^{-1}v \rangle = \|v\|_{B^{-1}}^2. \end{aligned}$$

By expanding the quadratic functional D at $\hat{p}$ and using optimality of $\hat{p}$, i.e. $\langle D'(\hat{p}), p - \hat{p} \rangle_V \geq 0$, we then see that

$$\begin{aligned} D(p) - D(\hat{p}) &= \langle D'(\hat{p}), p - \hat{p} \rangle_V + \tfrac{1}{2} \langle \Lambda B^{-1} \Lambda^* (p - \hat{p}), p - \hat{p} \rangle_V \\ &\geq \tfrac{1}{2} \|\Lambda^*(p - \hat{p})\|_{B^{-1}}^2 \\ &\geq \tfrac{c_B}{2} \|B^{-1}\Lambda^*(p - \hat{p})\|^2 = \tfrac{c_B}{2} \|u - \hat{u}\|^2, \end{aligned}$$

since due to Corollary 4.1 $\hat{u}$ is given by $\hat{u} = B^{-1}(-\Lambda^*\hat{p} + T^*g)$. □

To numerically solve (4.1) or (4.2) in a distributed parallel or memory-constrained setting, decomposition methods dissect the problem into smaller subproblems that can be solved independently of each other while still approaching the original solution by means of an iterative algorithm. One particular approach are domain decomposition algorithms which subdivide the problem domain.

While we are interested in solving (4.1) and (4.2), we opt to formulate the decomposition in a slightly more general way. More precisely, in the rest of this chapter we consider the following general problem.

Decomposition Setting

Let V, W be real Hilbert spaces, $\Lambda^* : V \to W$ a bounded linear operator, $B^{-1} : W \to W$ a positive definite self-adjoint bounded linear operator, $K \subseteq V$ a closed convex set and $f \in W$. We then consider the minimization problem

$$\inf_{p \in K} \Big\{ D(p) := \tfrac{1}{2} \|\Lambda^* p - T^* g\|_{B^{-1}}^2 \Big\}, \tag{4.4}$$

where $\|q\|_{B^{-1}}^2 := \langle B^{-1}q, q \rangle_W$, $q \in W$.

To ensure existence of a solution to (4.4) we will assume coercivity in the sense that for any feasible sequence $(p^n)_{n\in\mathbb{N}} \subseteq K$

$$\|p^n\|_V \to \infty \implies D(p^n) \to \infty. \tag{4.5}$$

Note that, while $\|\cdot\|^2_{B^{-1}}$ is strictly convex, Λ^* might not be injective and thus the solution to (4.4) does not necessarily need to be unique.

We will analyze a decomposition algorithm for problem (4.4) that requires a suitable partition of unity respecting the closed convex set K. More precisely we use bounded linear operators $\theta_i : V \to V$, $i = 1, \dots, M$, $M \in \mathbb{N}$ such that

$$I = \sum_{i=1}^{M} \theta_i \qquad \text{and} \qquad K = \sum_{i=1}^{M} \theta_i K. \tag{4.6}$$

Note here that, since θ_i is a bounded linear operator, $\theta_i K = \{\theta_i p : p \in K\} \subseteq V$ stays closed and convex.

The requirements for the partition given in (4.6) are in particular fulfilled by the following domain decomposition formulation. Let Ω_i, $i = 1, \dots, M$, $M \in \mathbb{N}$ be bounded open sets with Lipschitz boundary such that $\bigcup_{i=1}^{M} \Omega_i = \Omega$. Denote by $\tilde{\theta}_i : \Omega \to [0, 1]$, $i = 1, \dots, M$ a partition of unity satisfying

(i) $\tilde{\theta}_i \in W^{1,\infty}(\Omega)$,

(ii) $1 = \sum_{i=1}^{M} \tilde{\theta}_i$,

(iii) $\operatorname{supp} \tilde{\theta}_i \subseteq \overline{\Omega_i}$.

We then define the partition of unity operator $\theta_i : V \to V$ by pointwise multiplication

$$(\theta_i p)(x) := \tilde{\theta}_i(x) p(x), \tag{4.7}$$

for all $p \in V$.

Lemma 4.3. *The partition of unity operators* $(\theta_i)_{i=1}^M$ *defined in* (4.7) *satisfy the requirements of* (4.6)*, where* K *is given by Corollary 4.1.*

Proof. Linearity of θ_i, $i = 1, \dots, M$ is inherited from the pointwise multiplicative definition in (4.7). Let $p \in V = H_0^{\operatorname{div}}(\Omega)^m$, then

$$\begin{aligned} \|\tilde{\theta}_i p\|_{L^2} &\le \|\tilde{\theta}_i\|_{L^\infty} \|p\|_{L^2}, \\ \|\operatorname{div}(\tilde{\theta}_i p)\|_{L^2} &= \|\nabla\tilde{\theta}_i p + \tilde{\theta}_i \operatorname{div} p\|_{L^2} \\ &\le \|\nabla\tilde{\theta}_i\|_{L^\infty}\|p\|_{L^2} + \|\tilde{\theta}_i\|_{L^\infty}\|\operatorname{div} p\|_{L^2}, \end{aligned}$$

and thus, since $\tilde{\theta}_i \in W^{1,\infty}(\Omega)$ and in particular $\nabla\tilde{\theta}_i \in L^\infty(\Omega)$, we have proven that $\theta_i : V \to V$ is indeed well-defined and bounded.

Due to the pointwise nature of (4.7) we see that for $p \in V$:

$$\Big(\sum_{i=1}^M \theta_i p\Big)(x) = \sum_{i=1}^M \tilde{\theta}_i(x) p(x) = p(x)$$

which shows $I = \sum_{i=1}^M \theta_i$.

We have $K \subseteq \sum_{i=1}^M \theta_i K$ per definition. To show the other inclusion, let $p^i \in \theta_i K$, $i = 1, \dots, M$. Then we see that for $x \in \Omega$ a.e.

$$\Big|\sum_{i=1}^M p^i(x)\Big|_F \le \sum_{i=1}^M |p^i(x)|_F \le \sum_{i=1}^M \tilde{\theta}_i(x)\lambda \le \lambda,$$

thus showing that $p^i \in K$. □

4.1.1 Related Work

The authors of [33] have provided parallel and sequential decomposition methods specifically for the case $\Lambda^* = \operatorname{div}$ and $B^{-1} = I$ while assuming exact local minimization. Their work serves as a motivational basis for our generalization.

Recently in [75] a general framework for analyzing additive Schwarz methods of convex optimization problems as gradient methods has been presented. For the special case of parallel decomposition their analysis covers ours, while we extend our results to sequential decomposition which [75] does not cover. We also consider a slightly different notion of approximate minimization (see Definition 4.4) for the local subproblems which does not seem to map to the approximate notion considered from [75] in an obvious way.

4.2 Algorithm

Let us first introduce our notion of approximate minimization.

Definition 4.4. *For $q \in V$, $\varrho \in (0,1]$ we call*

$$\underset{p\in K}{\overset{\varrho,q}{\arg\min}}\, D(p) := \Big\{ \tilde{p} \in K : D(q) - D(\tilde{p}) \geq \varrho(D(q) - D(\hat{p})), \\ \hat{p} \in \underset{p\in K}{\arg\min}\, D(p) \Big\} \tag{4.8}$$

the set of ϱ-approximate minimizers of D on K with respect to q.

The condition in (4.8) means that the improvement in functional value needs to be at least within a constant factor of the remaining difference in functional value towards a true minimizer. For $\varrho = 1$ and arbitrary $q \in V$ this reduces to the usual notion of minimizers.

We present the decomposition procedures, namely Algorithms 4.5 and 4.6, which are structurally similar. Apart from our introduced generalizations and notation changes these correspond to those presented in [33].

Algorithm 4.5 (Parallel decomposition)**.**

```
Initialize: $p^0 \in K$ and $\sigma \in (0, \frac{1}{M}]$, $\varrho \in (0,1]$
for $n = 0, 1, 2, \dots$ do
    for $i = 1, \dots, M$ do
        $\tilde{v}_i^n \in \arg\min^{\varrho,\theta_i p^n}_{v_i \in \theta_i K} D\big(p^n + (v_i - \theta_i p^n)\big)$
    end for
    $p^{n+1} = p^n + \sum_{i=1}^{M} \sigma(\tilde{v}_i^n - \theta_i p^n)$
end for
```

Algorithm 4.6 (Sequential decomposition)**.**

```
*Initialize:* $p^0 \in K$ *and* $\sigma \in (0,1]$, $\varrho \in (0,1]$
for $n = 0, 1, 2, \dots$ do
    $p_0^n = p^n$
    for $i = 1, \dots, M$ do
        $\tilde{v}_i^n \in \arg\min^{\varrho,\theta_i p^n}_{v_i \in \theta_i K} D\big(p_{i-1}^n + (v_i - \theta_i p^n)\big)$
        $p_i^n = p_{i-1}^n + \sigma(\tilde{v}_i^n - \theta_i p^n)$
    end for
    $p^{n+1} = p_M^n$
end for
```

To treat both algorithms in a similar way, we use the convention $p_{i-1}^n := p^n$ for Algorithm 4.5 independent of $i \in \{1, \dots, M\}$. Having defined $\tilde{v}_i^n \in \theta_i K$ for $i \in \{1, \dots, M\}$ we also set $\tilde{p}_i^n := p_{i-1}^n + (\tilde{v}_i^n - \theta_i p^n) \in K$.

We observe that in each step $n \in \mathbb{N}_0, i \in \{1, \dots, M\}$ of Algorithms 4.5 and 4.6 the subproblem

$$\inf_{v_i \in \theta_i K} \tfrac{1}{2} \|\Lambda^* v_i - f_i^n\|_{B^{-1}}^2 \tag{4.9}$$

with $f_i^n = f - \Lambda^*(p_{i-1}^n - \theta_i p^n)$ needs to be solved approximately.

For a locally acting operator B^{-1} and suitable θ_i these problems may be solved on $\operatorname{supp}(\theta_i) \subseteq \overline{\Omega}$. If B^{-1} on the other hand is global then in order to avoid having to solve the subproblems globally on Ω a surrogate technique will be introduced in Section 4.5.

Definition 4.7. *For $p, q \in V$ we introduce the notation*

$$\langle p, q \rangle_* := \langle \Lambda B^{-1} \Lambda^* p, q \rangle_V, \qquad \|p\|_* := \sqrt{\langle p, p \rangle_*}.$$

Note that we have $\|p\|_*^2 = \|\Lambda^* p\|_{B^{-1}}^2$ in particular and that $\langle \cdot, \cdot \rangle_*$ and $\| \cdot \|_*$ are not necessarily positive definite.

Lemma 4.8. *Let $D' : V \to V$ be the Fréchet derivative of D. For any $p, q, r \in V$ we have*

(i) $D(p) - D(q) = \langle D'(q), p - q \rangle_V + \frac{1}{2}\|p - q\|_*^2$,

(ii) $\langle D'(p) - D'(q), r\rangle_V = \langle p - q, r\rangle_*$.

Proof. (i) We expand the quadratic functional D at q to obtain

$$\begin{aligned}D(p) &= D(q) + \langle D'(q), p - q\rangle_V + \tfrac{1}{2}\langle \Lambda B^{-1}\Lambda^*(p-q), p-q\rangle_V \\ &= D(q) + \langle D'(q), p - q\rangle_V + \tfrac{1}{2}\|p-q\|_*^2.\end{aligned}$$

(ii) We see directly

$$\begin{aligned}&\langle D'(p) - D'(q), r\rangle_V \\ &\quad= \langle \Lambda B^{-1}(\Lambda^* p - T^* g) - \Lambda B^{-1}(\Lambda^* q - T^* g), r\rangle_V \\ &\quad= \langle \Lambda B^{-1}\Lambda^*(p-q), r\rangle_V = \langle p - q, r\rangle_*.\end{aligned}$$ □

We first note that Lemma 4.8 actually holds true for any quadratic functional. Further, while the equations in Lemma 4.8 hold true with equality, the decomposition in this chapter only requires the left hand side to be less or equal than the right hand side correspondingly. In particular, this suggests a generalization to strongly convex, but not necessarily quadratic D.

4.3 Convergence Analysis

Following [33] we first establish monotonicity for the energy of iterates.

Lemma 4.9. *The iterates $(p^n)_{n\in\mathbb{N}}$ of Algorithms 4.5 and 4.6 with corresponding constraints on σ satisfy*

$$D(p^n) - D(p^{n+1}) \geq \varrho\sigma \sum_{i=1}^{M} \big(D(p_i^n) - D(\hat{p}_i^n)\big) \geq 0$$

where $\hat{p}_i^n = p_{i-1}^n + (\hat{v}_i^n - \theta_i p^n)$, $\hat{v}_i^n \in \arg\min_{v_i\in\theta_i K} D(p_{i-1}^n + (v_i - \theta_i p^n))$ denotes any exact minimizer in the i-th substep of the corresponding algorithm. The non-negative sequence $(D(p^n))_{n\in\mathbb{N}}$ is in particular monotonically decreasing and thus convergent.

Proof. The update step for p^{n+1} in the parallel case of Algorithm 4.5 is given as

$$\begin{aligned}p^{n+1} &= p^n + \sigma \sum_{i=1}^{M} (\tilde{v}_i^n - \theta_i p^n) \\ &= (1 - \sigma M) p^n + \sigma \sum_{i=1}^{M} \big(p^n + (\tilde{v}_i^n - \theta_i p^n)\big).\end{aligned}$$

We denote $\tilde{p}_i^n = p_{i-1}^n + (\tilde{v}_i^n - \theta_i p^n) = p^n + (\tilde{v}_i^n - \theta_i p^n)$. Since $\sigma \in (0, \frac{1}{M}]$, convexity of D yields

$$D(p^{n+1}) \leq (1 - \sigma M) D(p^n) + \sigma \sum_{i=1}^{M} D\big(\tilde{p}_i^n\big).$$

We use this and the definition of $\tilde{v}_i^n$ to estimate

$$\begin{aligned}D(p^n) - D(p^{n+1}) &\geq \sigma M D(p^n) - \sigma \sum_{i=1}^{M} D\big(\tilde{p}_i^n\big) \\ &= \sigma \sum_{i=1}^{M} \Big(D(p^n) - D\big(\tilde{p}_i^n\big)\Big) \\ &\geq \varrho \sigma \sum_{i=1}^{M} \Big(D(p^n) - D\big(\hat{p}_i^n\big)\Big),\end{aligned}$$

where we denoted $\hat{p}_i^n = p_{i-1}^n + (\hat{v}_i^n - \theta_i p^n) = p^n + (\hat{v}_i^n - \theta_i p^n)$ in the last inequality.

For the sequential case of Algorithm 4.6 we have similarly

$$\begin{aligned}p_i^n &= p_{i-1}^n + \sigma(\tilde{v}_i^n - \theta_i p^n) \\ &= (1 - \sigma) p_{i-1}^n + \sigma(p_{i-1}^n - (\tilde{v}_i^n - \theta_i p^n)) \\ &= (1 - \sigma) p_{i-1}^n + \sigma \tilde{p}_i^n\end{aligned}$$

and thus $D(p_i^n) \leq (1 - \sigma) D(p_{i-1}^n) + \sigma D(\tilde{p}_i^n)$. Rewriting we see that

$$\begin{aligned}D(p_{i-1}^n) - D(p_i^n) &\geq \sigma(D(p_{i-1}^n) - D(\tilde{p}_i^n)) \\ &\geq \varrho \sigma (D(p_{i-1}^n) - D(\hat{p}_i^n)),\end{aligned} \tag{4.10}$$

where we again used the definition of $\tilde{v}_i^n$ in the second inequality. A telescope sum over $i = 1, \dots, M$ then yields

$$\begin{aligned} D(p^n) - D(p^{n+1}) &= \sum_{i=1}^{M} (D(p_{i-1}^n) - D(p_i^n)) \\ &\geq \varrho\sigma \sum_{i=1}^{M} (D(p_{i-1}^n) - D(\hat{p}_i^n)). \end{aligned} \qquad \square$$

In particular, Lemma 4.9 shows monotonicity of energies, i.e. $D(p^n) \geq D(p^{n+1})$. Because of the coercivity assumption (4.5), the set of iterates $\{p^n : n \in \mathbb{N}_0\} \subseteq V$ is therefore bounded. We thus denote for some fixed minimizer $\hat{p} \in K$ of D the finite radius

$$R_{\hat{p}} := \sup\{\|p - \hat{p}\|_V : p \in K, D(p) \leq D(p^0)\} < \infty. \tag{4.11}$$

In the following we employ ideas from alternating minimization [18] to achieve a convergence rate estimate. To that end we first collect two elementary results.

Lemma 4.10. *Let $c > 0$ and $(a_k)_{k \in \mathbb{N}_0} \subseteq \mathbb{R}^+$ be a sequence such that for all $k \in \mathbb{N}_0$:*

$$a_k - a_{k+1} \geq c a_k^2.$$

Then $\lim_{k\to\infty} a_k \to 0$ with rate

$$0 < a_k < \frac{1}{ck + \frac{1}{a_0}} < \frac{1}{ck}$$

for all $k \in \mathbb{N}$.

Proof. We proceed similar to [18]. Since the iterates a_k, $k \in \mathbb{N}_0$ are monotonically decreasing, we can write for $k \in \mathbb{N}_0$:

$$\frac{1}{a_{k+1}} - \frac{1}{a_k} = \frac{a_k - a_{k+1}}{a_{k+1} a_k} \geq \frac{c a_k}{a_{k+1}} > c$$

and use it to reduce the telescope sum for $k > 0$:

$$\frac{1}{a_k} = \sum_{j=0}^{k-1} \Big(\frac{1}{a_{j+1}} - \frac{1}{a_j} \Big) + \frac{1}{a_0} > ck + \frac{1}{a_0}.$$

Inverting the inequality yields the statement. □

Lemma 4.11. *Let $a, b > 0$, $c, x, y \geq 0$ such that for all $\mu \in (0, 1]$ the inequality*

$$y \leq a\mu + \frac{b}{\mu} x + c\sqrt{x}$$

holds. Then the following split inequality holds:

$$y \leq \begin{cases} (2b + c\frac{\sqrt{b}}{\sqrt{a}})x & \text{if } x > \frac{a}{b}, \\ (2\sqrt{ab} + c)\sqrt{x} & \text{if } x \leq \frac{a}{b}, \end{cases} \tag{4.12}$$

or equivalently

$$x \geq \begin{cases} (2b + c\frac{\sqrt{b}}{\sqrt{a}})^{-1} y & \text{if } y > 2a + c\frac{\sqrt{a}}{\sqrt{b}}, \\ (2\sqrt{ab} + c)^{-2} y^2 & \text{if } y \leq 2a + c\frac{\sqrt{a}}{\sqrt{b}}. \end{cases}$$

Proof. If $x > \frac{a}{b}$ we choose $\mu = 1$ to arrive at

$$y \leq a + bx + c\sqrt{x} < 2bx + c\sqrt{x} \leq (2b + c\frac{\sqrt{b}}{\sqrt{a}})x.$$

Otherwise we minimize the expression by choosing $\mu = \frac{\sqrt{b}}{\sqrt{a}}\sqrt{x}$ and get

$$y \leq a\frac{\sqrt{b}}{\sqrt{a}}\sqrt{x} + b\frac{\sqrt{a}}{\sqrt{b}}\sqrt{x} + c\sqrt{x} = (2\sqrt{ab} + c)\sqrt{x}.$$

Both statements together yield the estimate.

Noting that the right-hand side of estimate (4.12) is continuous and monotonic in x, the case distinction can equivalently be written in terms of y by splitting at $x = \frac{a}{b}$, $y = (2b + c\frac{\sqrt{b}}{\sqrt{a}})\frac{a}{b} = 2a + c\sqrt{a}b$. Seperately solving the inequalities for x thus yields the equivalent representation. □

Lemma 4.12. *We have*

$$\sum_{i=1}^{M} \|\theta_i p\|_*^2 \leq \|B^{-1}\| \|\Lambda\|^2 C_\theta^2 \|p\|_V^2$$

with $C_\theta^2 := \sum_{i=1}^{M} \|\theta_i\|^2$.

Proof. Application of the Cauchy-Schwarz inequality yields

$$\begin{aligned}\sum_{i=1}^{M} \|\theta_i p\|_*^2 &= \sum_{i=1}^{M} \langle \Lambda^* \theta_i p, B^{-1} \Lambda^* \theta_i p\rangle_V \\ &\leq \sum_{i=1}^{M} \|B^{-1}\| \|\Lambda\|^2 \|\theta_i\|^2 \|p\|_V^2 \\ &= \|B^{-1}\| \|\Lambda\|^2 \Big(\sum_{i=1}^{M} \|\theta_i\|^2 \Big) \|p\|_V^2 .\end{aligned}$$

□

Lemma 4.13. *We may estimate the step distance in terms of the corresponding energy change as follows:*

$$\tfrac{1}{2}\|p_{i-1}^n - p_i^n\|_*^2 \leq \tfrac{\sigma}{\varrho}\big(2 - \varrho + 2\sqrt{1-\varrho}\big)(D(p_{i-1}^n) - D(p_i^n)).$$

Proof. Let $\omega > 0$ to be chosen later and denote $\tilde{p}_i^n := p_{i-1}^n + (\tilde{v}_i^n - \theta_i p^n)$.

$$\begin{aligned}&\tfrac{1}{2\sigma^2}\|p_{i-1}^n - p_i^n\|_*^2 \\ &\quad = \tfrac{1}{2\sigma^2}\|\sigma(\tilde{v}_i^n - \theta_i p^n)\|_*^2 = \tfrac{1}{2}\|p_{i-1}^n - \tilde{p}_i^n\|_*^2 \\ &\quad \leq \tfrac{1}{2}\Big((1+\omega)\|p_{i-1}^n - \hat{p}_i^n\|_*^2 + (1+\omega^{-1})\|\tilde{p}_i^n - \hat{p}_i^n\|_*^2\Big) \\ &\quad \leq (1+\omega)(D(p_{i-1}^n) - D(\hat{p}_i^n)) + (1+\omega^{-1})(D(\tilde{p}_i^n) - D(\hat{p}_i^n)) \\ &\qquad\qquad \text{(Lemma 4.8 (i) and optimality)} \\ &\quad \leq \tfrac{1+\omega}{\varrho}(D(p_{i-1}^n) - D(\tilde{p}_i^n)) + \tfrac{(1+\omega^{-1})(1-\varrho)}{\varrho}(D(p_{i-1}^n) - D(\tilde{p}_i^n)) \\ &\qquad\qquad \text{(due to (4.8))} \\ &\quad = \tfrac{1}{\varrho}\big(1 + \omega + (1+\omega^{-1})(1-\varrho)\big)(D(p_{i-1}^n) - D(\tilde{p}_i^n)) \\ &\quad \leq \tfrac{1}{\sigma\varrho}\big(1 + \omega + (1+\omega^{-1})(1-\varrho)\big)(D(p_{i-1}^n) - D(p_i^n)). \\ &\qquad\qquad \text{(using (4.10))}\end{aligned}$$

Choosing $\omega := \sqrt{1-\varrho}$ to minimize the expression we arrive at

$$\tfrac{1}{2\sigma^2}\|p_{i-1}^n - p_i^n\|_*^2 \le \tfrac{1}{\sigma\varrho}\big(2-\varrho+2\sqrt{1-\varrho}\big)(D(p_{i-1}^n) - D(p_i^n)). \qquad \square$$

Proposition 4.14. *Let $(p^n)_{n\in\mathbb{N}_0}$ be the iterates from either one of Algorithms 4.5 and 4.6 and let $\hat{p} \in K$ denote a minimizer of D.*

Then $D(p^n) \to D(\hat{p})$ as $n \to \infty$ owing to

$$\begin{aligned} &D(p^n) - D(\hat{p}) \\ &\quad \le \begin{cases} \frac{2}{\varrho\sigma}\alpha\big(D(p^n) - D(p^{n+1})\big) & \text{if } D(p^n) - D(p^{n+1}) > \frac{1}{2}\sigma\varrho\Phi^2, \\ \sqrt{\frac{2}{\varrho\sigma}}\Phi\alpha\sqrt{D(p^n) - D(p^{n+1})} & \text{else,} \end{cases} \end{aligned}$$

where $\alpha := 1 + M\sigma\sqrt{2-\varrho+2\sqrt{1-\varrho}}$ for Algorithm 4.6 and $\alpha := 1$ for Algorithm 4.5, and $\Phi := \sqrt{\|B^{-1}\|}\|\Lambda\|C_\theta R_{\hat{p}}$.

Proof. Using convexity we expand

$$\begin{aligned} &D(p^n) - D(\hat{p}) \\ &\quad \le \langle D'(p^n), p^n - \hat{p}\rangle_V = \sum_{i=1}^M \langle D'(p^n), \theta_i(p^n - \hat{p})\rangle_V \\ &\quad = \sum_{i=1}^M \Big(\langle D'(p_{i-1}^n), \theta_i(p^n - \hat{p})\rangle_V \\ &\qquad\qquad + \sum_{j=1}^{i-1} \langle D'(p_{j-1}^n) - D'(p_j^n), \theta_i(p^n - \hat{p})\rangle_V\Big). \end{aligned} \tag{4.13}$$

Let $\Phi_n := (\sum_{i=1}^M \|\theta_i(p^n - \hat{p})\|_*^2)^{\frac{1}{2}}$, $\hat{v}_i^n \in \arg\min_{v_i \in \theta_i K} D(p_{i-1}^n + (v_i - \theta_i p^n))$ and $\hat{p}_i^n := p_{i-1}^n + (\hat{v}_i^n - \theta_i p^n)$. We now estimate the first summand in the expansion above:

$$\begin{aligned}
&\sum_{i=1}^{M}\langle D'(p_{i-1}^n), \theta_i(p^n-\hat{p})\rangle_V \\
&\quad = \tfrac{1}{\mu}\sum_{i=1}^{M}\langle D'(p_{i-1}^n), \mu\theta_i(p^n-\hat{p})\rangle_V \\
&\quad = \tfrac{\mu}{2}\sum_{i=1}^{M}\|\theta_i(p^n-\hat{p})\|_*^2 + \tfrac{1}{\mu}\sum_{i=1}^{M}\big(D(p_{i-1}^n) - D(p_{i-1}^n - \mu\theta_i(p^n-\hat{p}))\big) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(Lemma 4.8 (i))} \\
&\quad = \tfrac{\Phi_n^2\mu}{2} + \tfrac{1}{\mu}\sum_{i=1}^{M}\Big(D(p_{i-1}^n) - D\big(p_{i-1}^n + ((1-\mu)\theta_i p^n + \mu\theta_i\hat{p} - \theta_i p^n)\big)\Big) \\
&\quad \le \tfrac{\Phi_n^2\mu}{2} + \tfrac{1}{\mu}\sum_{i=1}^{M}(D(p_{i-1}^n) - D(\hat{p}_i^n)) \qquad\qquad \text{(optimality)} \\
&\quad \le \tfrac{\Phi_n^2\mu}{2} + \tfrac{1}{\mu\varrho\sigma}(D(p^n) - D(p^{n+1})), \qquad\qquad \text{(Lemma 4.9)}
\end{aligned}$$

where optimality was used by realizing that $(1-\mu)\theta_i p^n + \mu\theta_i\hat{p} \in \theta_i K$. For the second summand we see

$$\begin{aligned}
&\sum_{i=1}^{M}\sum_{j=1}^{i-1}\langle D'(p_{j-1}^n) - D'(p_j^n), \theta_i(p^n-\hat{p})\rangle_V \\
&\quad = \sum_{i=1}^{M}\sum_{j=1}^{i-1}\big\langle p_{j-1}^n - p_j^n, \theta_i(p^n-\hat{p})\big\rangle_* \qquad\qquad \text{(Lemma 4.8 (ii))} \\
&\quad \le \sum_{i=1}^{M}\sum_{j=1}^{i-1}\|p_{j-1}^n - p_j^n\|_*\|\theta_i(p^n-\hat{p})\|_* \\
&\quad \le M\Big(\sum_{j=1}^{M}\|p_{j-1}^n - p_j^n\|_*^2\Big)^{\frac{1}{2}}\Big(\sum_{i=1}^{M}\|\theta_i(p^n-\hat{p})\|_*^2\Big)^{\frac{1}{2}} \\
&\quad \le M\Phi_n\Big(\sum_{j=1}^{M}\|p_{j-1}^n - p_j^n\|_*^2\Big)^{\frac{1}{2}}.
\end{aligned}$$

Applying Lemma 4.13 completes the estimate of the second summand, yielding

$$\begin{aligned}\sum_{i=1}^{M}\sum_{j=1}^{i-1}&\langle D'(p_{i-1}^n) - D'(p_i^n), \theta_j(p^n - \hat{p})\rangle_V \\ &\le M\Phi_n\sqrt{\tfrac{2\sigma}{\varrho}(2-\varrho+2\sqrt{1-\varrho})}\big(D(p^n) - D(p^{n+1})\big)^{\frac{1}{2}}.\end{aligned}$$

Combining both estimates and roughly bounding $\Phi_n \le \Phi$ due to Lemma 4.12 we have

$$\begin{aligned}D(p^n) - D(\hat{p}) \le \tfrac{\Phi^2\mu}{2} + \tfrac{1}{\mu\varrho\sigma}\big(D(p^n) - D(p^{n+1})\big)& \\ + M\Phi\sqrt{\tfrac{2\sigma}{\varrho}(2-\varrho+2\sqrt{1-\varrho})}\big(D(p^n) - D(p^{n+1})\big)^{\frac{1}{2}}&.\end{aligned}$$

Invoking Lemma 4.11 with the constants $a = \frac{\Phi^2}{2}$, $b = \frac{1}{\varrho\sigma}$ and $c = M\Phi\sqrt{\frac{2\sigma}{\varrho}(2-\varrho+2\sqrt{1-\varrho})}$ yields the split bound with the following coefficients:

$$\begin{aligned}2b + c\sqrt{\tfrac{b}{a}} &= \tfrac{2}{\varrho\sigma} + M\Phi\sqrt{\tfrac{2\sigma}{\varrho}(2-\varrho+2\sqrt{1-\varrho})}\sqrt{\tfrac{2}{\sigma\varrho\Phi^2}} \\ &= \tfrac{2}{\varrho\sigma}(1 + M\sigma\sqrt{2-\varrho+2\sqrt{1-\varrho}}), \\ 2\sqrt{ab} + c &= 2\sqrt{\tfrac{\Phi^2}{2\varrho\sigma}} + M\Phi\sqrt{\tfrac{2\sigma}{\varrho}(2-\varrho+2\sqrt{1-\varrho})} \\ &= \sqrt{\tfrac{2}{\varrho\sigma}}\Phi(1 + M\sigma\sqrt{2-\varrho+2\sqrt{1-\varrho}}),\end{aligned}$$

which concludes the proof for Algorithm 4.6.

For Algorithm 4.5, examining the proof above, we notice that for the parallel version we have $p_i^n = p_{i-1}^n = p^n$ for $i = 1, \dots, M-1$ and thus the second summand in (4.13) vanishes completely. This allows us to invoke Lemma 4.11 with $c = 0$ and leads to the desired statement. □

Theorem 4.15. *Let $(p^n)_{n\in\mathbb{N}_0}$ be the iterates from either one of Algorithms 4.5 and 4.6 and let $\hat{p} \in K$ denote a minimizer of D. Algorithms 4.5 and 4.6 converge in the sense that $D(p^n) \to D(\hat{p})$. More* *specifically,*

$$D(p^n) - D(\hat{p}) \leq \begin{cases} (1 - \frac{\varrho\sigma}{2\alpha})^n \big(D(p^0) - D(\hat{p})\big) & \text{if } n \leq n_0 \\ \frac{2\Phi^2}{\varrho\sigma}\alpha^2(n - n_0 + 1)^{-1} & \text{if } n \geq n_0, \end{cases}$$

where $\alpha := 1 + M\sigma\sqrt{2 - \varrho + 2\sqrt{1-\varrho}}$ *for Algorithm 4.6 and* $\alpha := 1$ *for Algorithm 4.5,* $\Phi := \sqrt{\|B^{-1}\|\|\Lambda\|C_\theta R_{\hat{p}}}$ *and* $n_0 := \min\{n \in \mathbb{N}_0 : D(p^n) - D(\hat{p}) < \Phi^2\alpha\}$*.*

Proof. We first observe that since $(D(p^n))_{n\in\mathbb{N}_0}$ is monotonically decreasing, n_0 is well-defined and we have $D(p^n) - D(\hat{p}) \geq \Phi^2\alpha$ for all $n \in \mathbb{N}_0$, $n < n_0$ and likewise $D(p^n) - D(\hat{p}) < \Phi^2\alpha$ for all $n \in \mathbb{N}_0$, $n \geq n_0$.

We now make use of Proposition 4.14. The equivalence in Lemma 4.11 then yields

$$D(p^{n-1}) - D(p^n) \geq \begin{cases} \frac{\varrho\sigma}{2\alpha}(D(p^{n-1}) - D(\hat{p})) & \text{if } n - 1 < n_0 \\ \frac{\varrho\sigma}{2\Phi^2\alpha^2}(D(p^{n-1}) - D(\hat{p}))^2 & \text{if } n - 1 \geq n_0. \end{cases}$$

In the former case we invert the inequality and add $D(p^{n-1}) - D(\hat{p})$ to arrive at

$$D(p^n) - D(\hat{p}) \leq (1 - \tfrac{\varrho\sigma}{2\alpha})(D(p^{n-1}) - D(\hat{p})),$$

which recursively yields the required statement for all $n \leq n_0$. In the latter case we may assume without loss of generality that $n_0 = 0$ since we can shift the sequence if necessary. Thus for all $n \in \mathbb{N}_0$:

$$D(p^n) - D(p^{n+1}) \geq \tfrac{\varrho\sigma}{2\Phi^2\alpha^2}(D(p^n) - D(\hat{p}))^2.$$

Invoking Lemma 4.10 with constant $c := \frac{\varrho\sigma}{2\Phi^2\alpha^2}$ we obtain

$$\begin{aligned} D(p^n) - D(\hat{p}) &\leq \frac{1}{cn + \frac{1}{D(p^0)-D(\hat{p})}} \leq \frac{1}{cn + \frac{1}{\Phi^2\alpha}} \\ &\leq \frac{1}{cn + \frac{\varrho\sigma}{2\Phi^2\alpha^2}} = \frac{1}{c(n+1)} \end{aligned}$$

since $0 \leq \sigma, \varrho \leq 1$ and $\alpha \geq 1$, thereby showing the second inequality. □

4.4 Comparison

We conclude that in special cases the results obtained here are either in agreement with or may improve upon other known estimates.

Gradient Method Framework [75]

In the special case of parallel decomposition, i.e. $\alpha = 1$, and exact local solutions, i.e. $\varrho = 1$, the framework of [75] is applicable to our model and their estimate [75, Algorithm 4.1] reproduces ours. We show this by specializing and transforming their estimate.

Using notation from [75] we employ [75, Algorithm 4.1] by setting $E(u) = F(u) + G(u) := D(u) + \chi_K$. The space decomposition is specified by the images of θ_k, $k = 1, \dots, M$, i.e. $V_k := \operatorname{im} \theta_k \subseteq V$ with $R_k^* : V_k \to V$ then being the inclusion map. We choose to use exact local solvers, i.e. $\varrho = 1$ in our notation, since it is not obvious to us how our notion of approximate minimization maps to theirs. In particular, d_k and G_k are chosen as in [75, (4.3)] and $\omega := \omega_0 := 1$. We now verify [75, Assumptions 4.1 to 4.3] in order to apply [75, Theorem 4.7]. [75, Assumption 4.1] is fulfilled due to Lemmas 4.8 and 4.12 with $C_{0,K} := C_\theta \|\Lambda\| \sqrt{\|B^{-1}\|}$ and $q := 2$. We fulfill [75, Assumption 4.2] by choosing $\tau_0 := \frac{1}{N}$ (their τ corresponds to our σ). [75, Assumption 4.3] is trivialized in the case of exact local solvers. Applying [75, Theorem 4.7] with $C_{q,\tau} = 2$ and $\kappa = \frac{1}{\tau} C_\theta^2 \|\Lambda\|^2 \|B^{-1}\|$ yields

$$\begin{aligned} D(p^1) - D(\hat{p}) &\leq (1 - \sigma(1 - \tfrac{1}{2}))(D(p^0) - D(\hat{p})) \\ &= (1 - \tfrac{\sigma}{2})(D(p^0) - D(\hat{p})) \end{aligned} \tag{4.14}$$

if $D(p^0) - D(\hat{p}) \geq \tau R_{\hat{p}}^2 \kappa = \Phi^2$ and

$$D(p^n) - D(\hat{p}) \leq \tfrac{C_{q,r} R_{\hat{p}}^2 \kappa}{(n+1)^{q-1}} = \tfrac{2\Phi^2}{\sigma}(n+1)^{-1} \tag{4.15}$$

otherwise. Applying estimate (4.14) recursively and shifting the sequence by n_0 for the estimate (4.15) finally yields the formulation

$$D(p^n) - D(\hat{p}) \leq \begin{cases} (1 - \frac{\sigma}{2})^n \big(D(p^0) - D(\hat{p})\big) & \text{if } n \leq n_0 \\ \frac{2\Phi^2}{\sigma}(n - n_0 + 1)^{-1} & \text{if } n \geq n_0, \end{cases}$$

which is in agreement with Theorem 4.15.

Decomposition of the Rudin-Osher-Fatemi Model [33]

In order to compare with the convergence rate in [33], we specialize our model to their setting by chosing $V = H_0^{\operatorname{div}}(\Omega)$, $\Lambda^* = \operatorname{div} : V \to L^2(\Omega)$, $T = S = I : L^2(\Omega) \to L^2(\Omega)$, $\alpha_2 = \beta = 1$ (thus $B = I$) and $\varrho = 1$. Next we introduce some notation from [33], namely $C_0, \delta > 0$ such that $\|\nabla\tilde{\theta}_i\|_{L^\infty} \leq \frac{C_0}{\delta}$ for $i = 1, \dots, M$, c.f. [33, (2.10)], $\zeta^0 := 2(D(p^0) - D(\hat{p}))$ (our D has an additional factor of $\frac{1}{2}$) and $N_0 := \max_{x\in\Omega} |\{i \in \{1, \dots, M\} : x \in \Omega_i\}|$. Then [33, Theorem 3.1] and [33, Theorem 3.6] provide the following estimate:

$$\tfrac{1}{2}\|u^n - \hat{u}\|^2 \leq D(p) - D(\hat{p}) \leq Cn^{-1} \tag{4.16}$$

where $u^n := -\operatorname{div} p^n + g$, $\hat{u} := -\operatorname{div}\hat{p} + g$ and

$$C := \tfrac{1}{2}\zeta^0\Big(\tfrac{2}{\sigma}(2M+1)^2 + 8\sqrt{2}C_0\lambda|\Omega|^{\frac{1}{2}}(\zeta^0)^{-\frac{1}{2}}\frac{M\sqrt{N_0}}{\delta\sqrt{\sigma}} + \sqrt{2} - 1\Big)^2.$$

Note that we used our notation for M and σ.

In order to compare favorably in this setting, we slightly refine the estimate $\Phi_n \leq \Phi$ from the proof of Proposition 4.14. First, we quantify an estimate from the proof of Lemma 4.3. For all $p \in V$ we have

$$\begin{aligned}
\sum_{i=1}^M \|\operatorname{div}\theta_i p\|_{L^2}^2 &\leq \sum_{i=1}^M \|\nabla\tilde{\theta}_i \cdot p + \tilde{\theta}_i \operatorname{div} p\|_{L^2}^2 \\
&\leq \sum_{i=1}^M \Big((1+\omega)\|\nabla\tilde{\theta}_i \cdot p\|_{L^2}^2 + (1+\omega^{-1})\|\tilde{\theta}_i \operatorname{div} p\|_{L^2}^2 \\
&= (1+\omega)\int_\Omega \sum_{i=1}^M |\nabla\tilde{\theta}_i \cdot p|^2 \,\mathrm{d}x + (1+\omega^{-1})\int_\Omega \sum_{i=1}^M |\tilde{\theta}_i \operatorname{div} p|^2 \,\mathrm{d}x \\
&\leq (1+\omega)\int_\Omega \Big(\sum_{i=1}^M |\nabla\tilde{\theta}_i|^2\Big)|p|^2 \,\mathrm{d}x \\
&\qquad + (1+\omega^{-1})\int_\Omega \Big(\sum_{i=1}^M |\tilde{\theta}_i|^2\Big)|\operatorname{div} p|^2 \,\mathrm{d}x
\end{aligned}$$

$$\begin{aligned}&\le (1+\omega)N_0\|\nabla\tilde{\theta}_i\|_{L^\infty}^2\|p\|_{L^2}^2+(1+\omega^{-1})\|\operatorname{div}p\|_{L^2}^2\\&\le (1+\omega)N_0\tfrac{C_0^2}{\delta^2}\|p\|_{L^2}^2+(1+\omega^{-1})\|\operatorname{div}p\|_{L^2}^2,\end{aligned}$$

for any $\omega > 0$. The pointwise box-constraints $|p| \le \lambda$ imply $\|p^n - \hat{p}\|_{L^2}^2 = \int_\Omega |p^n - \hat{p}|^2 \,\mathrm{d}x \le (2\lambda)^2|\Omega|$. Combining this allows us to estimate

$$\begin{aligned}\Phi_n^2 &= \sum_{i=1}^M \|\theta_i(p^n-\hat{p})\|_*^2 = \sum_{i=1}^M \|\operatorname{div}\theta_i(p^n-\hat{p})\|_{L^2}^2\\&\le (1+\omega)N_0\tfrac{C_0^2}{\delta^2}\|p^n-\hat{p}\|_{L^2}^2+(1+\omega^{-1})\|\operatorname{div}(p^n-\hat{p})\|_{L^2}^2\\&\le (1+\omega)\cdot 4\lambda^2|\Omega|N_0\tfrac{C_0^2}{\delta^2}+(1+\omega^{-1})\zeta^0\\&= \Big(2\lambda|\Omega|^{\frac12}N_0^{\frac12}\tfrac{C_0}{\delta}+(\zeta^0)^{\frac12}\Big)^2 =: \tilde{\Phi}^2\end{aligned}$$

by optimally choosing $\omega := (4\lambda^2|\Omega|N_0\frac{C_0^2}{\delta^2})^{-\frac12}(\zeta^0)^{\frac12}$. We therefore conclude that in this specific setting Theorem 4.15 holds true with Φ replaced by $\tilde{\Phi}$. Their and our estimate thus amount to

$$\begin{aligned}&\tfrac12\|u^n-\hat{u}\|^2 \le Cn^{-1},\\&\tfrac12\|u^n-\hat{u}\|^2 \le \tfrac{2\tilde{\Phi}^2}{\sigma}\alpha^2(n-n_0+1)^{-1},\end{aligned}$$

where for the lower estimate α and n_0 are defined as in Theorem 4.15 and $n \ge n_0$. Rewriting the involved constants,

$$\begin{aligned}C &= \bigg(\Big(\tfrac{\sqrt{2}}{2}\tfrac{(2M+1)^2}{\sigma}+\sqrt{2}-1\Big)\sqrt{\zeta^0}+8\tfrac{M}{\sqrt{\sigma}}\lambda|\Omega|^{\frac12}\sqrt{N_0}\tfrac{C_0}{\delta}\bigg)^2,\\ \tfrac{2\tilde{\Phi}^2\alpha^2}{\sigma} &\le \tfrac{2(1+\sigma M)^2}{\sigma}\Big(2\lambda|\Omega|^{\frac12}\sqrt{N_0}\tfrac{C_0}{\delta}+\sqrt{\zeta^0}\Big)^2\\&= \Big(\sqrt{2}\tfrac{1+\sigma M}{\sqrt{\sigma}}\sqrt{\zeta^0}+2\sqrt{2}\tfrac{1+\sigma M}{\sqrt{\sigma}}\lambda|\Omega|^{\frac12}\sqrt{N_0}\tfrac{C_0}{\delta}\Big)^2,\end{aligned}$$

we see that $\frac{2\tilde{\Phi}^2\alpha^2}{\sigma} \le C$ by comparing the relevant terms before $\sqrt{\zeta^0}$ and $\lambda|\Omega|^{\frac12}\sqrt{N_0}\frac{C_0}{\delta}$ under the square separately using $0 < \sigma \le 1$ and

$M \geq 1$:

$$\sqrt{2}\tfrac{1+\sigma M}{\sqrt{\sigma}} \leq \tfrac{\sqrt{2}}{\sqrt{\sigma}}(1+M) < \tfrac{\sqrt{2}}{\sigma}\tfrac{(2M+1)^2}{2} \leq \tfrac{\sqrt{2}}{2}\tfrac{(2M+1)^2}{\sigma} + \sqrt{2} - 1$$
$$2\sqrt{2}\tfrac{1+\sigma M}{\sqrt{\sigma}} \leq 3\tfrac{1+M}{\sqrt{\sigma}} < 4\tfrac{2M}{\sqrt{\sigma}} = 8\tfrac{M}{\sqrt{\sigma}}.$$

Consequently, Theorem 4.15 provides a strictly better estimate than [33, Theorems 3.1, 3.6] both for sufficiently large $n \in \mathbb{N}$ and for all $n \in \mathbb{N}$ whenever $n_0 = 0$ (i.e. the initial guess is close enough). While we expect Theorem 4.15 to prevail for $n_0 > 0$ as well, a complete comparison in that case seems to be more involved and remains to be done.

4.5 Surrogate Technique

A surrogate iteration substitutes minimization of one functional with minimization of different, simpler functionals at the cost of an additional iterative process. In particular one can substitute the minimization problem $\inf_{p\in K} \frac{1}{2}\|\Lambda^* p - f\|^2_{B^{-1}}$ by the iteration

$$\inf_{p^{n+1}\in K} \tfrac{1}{2}\|\Lambda^* p^{n+1} - f^n\|^2_W, \qquad f^n = \Lambda^* p^n - \tfrac{1}{\tau}B^{-1}(\Lambda^* p^n - f),$$

producing iterates $(p^n)_{n\in\mathbb{N}}$ for some initialization $p^0 \in V$ that converge to the same minimizer, provided $\tau \in (\|B^{-1}\|, \infty)$. Though its properties have been studied extensively in e.g. [71], we will analyze it as a nested subalgorithm of the domain decomposition scheme for approximate minimization following the notion from Definition 4.4. The main motivation for the surrogate technique in our case is to avoid having the local problems (4.9) depend directly on the potentially costly operator B^{-1}.

To that end we introduce an auxiliary functional D_i^s defined as

$$D_i^s(v_i, w_i) := D(p^n_{i-1} + (v_i - \theta_i p^n)) + \tfrac{1}{2}\|\Lambda^*(v_i - w_i)\|^2_{\tau I - B^{-1}}$$

with $\tau > \|B^{-1}\|$ for $v_i, w_i \in \theta_i K$ and $i = 1, \ldots, M$.

Algorithm 4.16 (Surrogate approximation)**.**

```
***Parameters:*** $N_{sur} \in \mathbb{N}$
***Input:*** $n \in \mathbb{N}_0$, $i \in \{1, \dots, M\}$, $p^n \in K$, $p^n_{i-1} \in K$
***Output:*** $\tilde{v}^n_i \in \theta_i K$
$v^{n,0}_i = \theta_i p^n_{i-1}$
for $\ell = 0, 1, \dots, N_{sur} - 1$ do
    $v^{n,\ell+1}_i \in \arg\min_{v_i \in \theta_i K} D^s_i(v_i, v^{n,\ell}_i)$
end for
$\tilde{v}^n_i = v^{n,N_i}_i$
```

We note that the subproblems in Algorithm 4.16 can be written as

$$\begin{aligned}\inf_{v_i\in\theta_i K} D^s_i(v_i, v^{n,\ell}_i) \iff & \inf_{v_i\in\theta_i K} \tfrac{1}{2}\|\Lambda^*(p^n_{i-1} + (v_i - \theta_i p^n)) - f\|^2_{B^{-1}} \\ & \qquad + \tfrac{1}{2}\|\Lambda^*(v_i - v^{n,\ell}_i)\|^2_{\tau I - B^{-1}} \\ \iff & \inf_{v_i\in\theta_i K} \tfrac{1}{2}\|\Lambda^* v_i - f^n_i\|^2_W,\end{aligned}$$

where $f^n_i = \Lambda^* v^{n,\ell}_i - \frac{1}{\tau}B^{-1}(\Lambda^*(p^n_{i-1} + (v^{n,\ell}_i - \theta_i p^n)) - f)$. The dependence on the operator B^{-1} has thereby been moved into the preparation of fixed data f^n_i for every subproblem, while the subproblem itself for fixed f^n_i is independent of B^{-1}.

Algorithm 4.16 produces approximations v^n_i to be used in Algorithms 4.5 and 4.6. Following ideas from [71, Proposition 2.2] it will soon become clear, that the surrogate approximation converges linearly and any fixed number of surrogate iterations N_{sur} is enough to receive the convergence rate from Theorem 4.15 for the resulting combined algorithm.

Lemma 4.17. *Using notation and assumptions from Algorithm 4.16 the functional $D^n_i : V_i \to \mathbb{R}$,*

$$D^n_i(v) := D(p^n_{i-1} + (v - \theta_i p^n)),$$

has quadratic growth in the sense that

$$D^n_i(v) - D^n_i(\hat{v}) \geq \tfrac{1}{2\|\tau I - B^{-1}\|\|B\|}\|\Lambda^*(v - \hat{v})\|^2_{\tau I - B^{-1}}$$

for any minimizer $\hat{v} \in \theta_i K$ of D^n_i.

Proof. Using Lemma 4.8 and optimality of $\hat{v} \in \theta_i K$ we see that

$$\begin{aligned} D_i^n(v) - D_i^n(\hat{v}) &= \langle D'(p_{i-1}^n + (\hat{v} - \theta_i p^n)), v - \hat{v}_i \rangle_V + \tfrac{1}{2}\|v - \hat{v}\|_*^2 \\ &\geq \tfrac{1}{2}\|\Lambda^*(v - \hat{v})\|_{B^{-1}}^2. \end{aligned}$$

Further noting that $\tau I - B^{-1}$ is positive definite, since $\tau > \|B^{-1}\|$,

$$\begin{aligned} \|\Lambda^*(v - \hat{v})\|_{\tau I - B^{-1}}^2 &\leq \|\tau I - B^{-1}\| \|\Lambda^*(v - \hat{v})\|_W^2 \\ &\leq \|\tau I - B^{-1}\| \|B\| \|\Lambda^*(v - \hat{v})\|_{B^{-1}}^2. \end{aligned}$$

Combining both inequalities yields the statement. □

Proposition 4.18. *Using the notation and assumptions from Algorithm 4.16 and Lemma 4.17 the surrogate iterates $v_i^{n,\ell}$ satisfy*

$$D_i^n(v_i^{n,\ell}) - D_i^n(v_i^{n,\ell+1}) \geq \eta(D_i^n(v_i^{n,\ell}) - D_i^n(\hat{v}_i^n))$$

for any minimizer $\hat{v}_i^n \in \theta_i K$ of D_i^n while $\eta \in (0,1)$ is given by

$$\eta = \begin{cases} \frac{1}{4\|\tau I - B^{-1}\|\|B\|} & \textit{if } \|\tau I - B^{-1}\|\|B\| \geq \frac{1}{2} \\ 1 - \|\tau I - B^{-1}\|\|B\| & \textit{else.} \end{cases}$$

Proof. Since $D_i^s(v, w) = D_i^n(v) + \frac{1}{2}\|\Lambda^*(w - v)\|_{\tau I - B^{-1}}^2$ we have

$$\begin{aligned} & D_i^n(v_i^{n,\ell+1}) + \tfrac{1}{2}\|\Lambda^*(v_i^{n,\ell} - v_i^{n,\ell+1})\|_{\tau I - B^{-1}}^2 \\ &\quad = D_i^s(v_i^{n,\ell+1}, v_i^{n,\ell}) \\ &\quad = \min_{v_i \in \theta_i K} D_i^n(v_i) + \tfrac{1}{2}\|\Lambda^*(v_i^{n,\ell} - v_i)\|_{\tau I - B^{-1}}^2 \\ &\quad \leq \min_{\mu \in [0,1]} D_i^n((1-\mu)v_i^{n,\ell} + \mu \hat{v}_i^n) + \tfrac{\mu^2}{2}\|\Lambda^*(v_i^{n,\ell} - \hat{v}_i^n)\|_{\tau I - B^{-1}}^2 \\ &\quad \leq \min_{\mu \in [0,1]} (1-\mu) D_i^n(v_i^{n,\ell}) + \mu D_i^n(\hat{v}_i^n) + \tfrac{\mu^2}{2}\|\Lambda^*(v_i^{n,\ell} - \hat{v}_i^n)\|_{\tau I - B^{-1}}^2, \end{aligned}$$

where we searched for the minimum along the line $v_i = (1-\mu)v_i^{n,\ell} + \mu \hat{v}_i^n \in \theta_i K$, $\mu \in [0,1]$ and used convexity afterwards. After reordering we use the quadratic growth property from Lemma 4.17 to see that

$$\begin{aligned}
&D_i^n(v_i^{n,\ell}) - D_i^n(v_i^{n,\ell+1}) - \tfrac{1}{2}\|\Lambda^*(v_i^{n,\ell} - v_i^{n,\ell+1})\|^2_{\tau I - B^{-1}} \\
&\quad\geq \max_{\mu\in[0,1]} \mu(D_i^n(v_i^{n,\ell}) - D_i^n(\hat{v}_i^n)) - \tfrac{\mu^2}{2}\|\Lambda^*(v_i^{n,\ell} - \hat{v}_i^n)\|^2_{\tau I - B^{-1}} \\
&\quad\geq \max_{\mu\in[0,1]} \big(\mu - \mu^2\|\tau I - B^{-1}\|\|B\|\big)(D_i^n(v_i^{n,\ell}) - D_i^n(\hat{v}_i^n)).
\end{aligned}$$

Discarding the last term on the left-hand side and evaluating the maximum optimally at $\mu = \min\{1, \frac{1}{2\|\tau I - B^{-1}\|\|B\|}\} \in (0,1]$ yields

$$D_i^n(v_i^{n,\ell}) - D_i^n(v_i^{n,\ell+1}) \geq \eta(D_i^n(v_i^{n,\ell}) - D_i^n(\hat{v}_i^n))$$

where $\eta \in (0,1)$ is given by

$$\eta = \begin{cases} \frac{1}{4\|\tau I - B^{-1}\|\|B\|} & \text{if } \|\tau I - B^{-1}\|\|B\| \geq \frac{1}{2} \\ 1 - \|\tau I - B^{-1}\|\|B\| & \text{else.} \end{cases} \qquad \square$$

Proposition 4.18 is sharp in the sense that for trivial $B^{-1} = I$ and minimizing $1 < \tau \to 1$, we recover the optimal factor $\eta \to 1$.

Lemma 4.19. *The surrogate iterates $(v_i^{n,\ell})_\ell$ from Algorithm 4.16 yield approximate solutions to the subproblems in the sense that*

$$D_i^n(v_i^{n,0}) - D_i^n(v_i^{n,\ell}) \geq \big(1 - (1-\eta)^\ell\big)(D_i^n(v_i^{n,0}) - D_i^n(\hat{v}_i^n))$$

for any minimizer $\hat{v}_i^n \in \theta_i K$ of D_i^n, $i \in \{1\ldots, M\}$, $n \in \mathbb{N}_0$ and $\eta \in (0,1)$ defined as in Proposition 4.18.

Proof. Elementary calculation using Proposition 4.18 yields a linear energy decrease

$$\begin{aligned}
&D_i^n(v_i^{n,\ell+1}) - D_i^n(\hat{v}_i^n) \\
&\quad= -(D_i^n(v_i^{n,\ell}) - D_i^n(v_i^{n,\ell+1})) + D_i^n(v_i^{n,\ell}) - D_i^n(\hat{v}_i^n) \\
&\quad\leq -\eta(D_i^n(v_i^{n,\ell}) - D_i^n(\hat{v}_i^n)) + D_i^n(v_i^{n,\ell}) - D_i^n(\hat{v}_i^n) \\
&\quad= (1-\eta)(D_i^n(v_i^{n,\ell}) - D_i^n(\hat{v}_i^n))
\end{aligned}$$

which we use to find

$$\begin{aligned}
&D_i^n(v_i^{n,0}) - D_i^n(v_i^{n,\ell}) \\
&\quad = D_i^n(v_i^{n,0}) - D_i^n(\hat{v}_i^n) - (D_i^n(v_i^{n,\ell}) - D_i^n(\hat{v}_i^n)) \\
&\quad \geq D_i^n(v_i^{n,0}) - D_i^n(\hat{v}_i^n) - (1-\eta)^\ell (D_i^n(v_i^{n,0}) - D_i^n(\hat{v}_i^n)) \\
&\quad \geq (1-(1-\eta)^\ell)(D_i^n(v_i^{n,0}) - D_i^n(\hat{v}_i^n)).
\end{aligned}$$

□

Finally, combining Theorem 4.15 with Lemma 4.19 then immediately yields the following corollary.

Corollary 4.20. *Algorithms 4.5 and 4.6 with subproblems solved using Algorithm 4.16 converge in the sense that $D(p^n) \to D(\hat{p})$. Furthermore*

$$D(p^n) - D(\hat{p}) \leq \begin{cases} (1 - \frac{\varrho\sigma}{2\alpha})^n \big(D(p^0) - D(\hat{p})\big) & \textit{if } n \leq n_0 \\ \frac{2\Phi^2}{\varrho\sigma}\alpha^2 (n - n_0 + 1)^{-1} & \textit{if } n \geq n_0, \end{cases}$$

where $\alpha := 1 + M\sigma\sqrt{2 - \varrho + 2\sqrt{1-\varrho}}$ for Algorithm 4.6 and $\alpha := 1$ for Algorithm 4.5, $\Phi := \sqrt{\|B^{-1}\|}\|\Lambda\| C_\theta R_{\hat{p}}$, $n_0 := \min\{n \in \mathbb{N}_0 : D(p^n) - D(\hat{p}) < \Phi^2\alpha\}$ and

$$\begin{aligned}
\varrho &= (1 - (1-\eta)^{N_{sur}}), \\
\eta &= \begin{cases} \frac{1}{4\|\tau I - B^{-1}\|\|B\|} & \textit{if } \|\tau I - B^{-1}\|\|B\| \geq \frac{1}{2} \\ 1 - \|\tau I - B^{-1}\|\|B\| & \textit{else.} \end{cases}
\end{aligned}$$

for any fixed number of inner surrogate iterations $N_{sur} \in \mathbb{N}$.

Concluding, in this section we were able to prove a convergence rate for the nested surrogate algorithm in Corollary 4.20 using Theorem 4.15 because we successfully managed to apply our notion of ϱ-approximate minimizers from Definition 4.4 to any fixed number of surrogate iterations with the help of Proposition 4.18 and Lemma 4.19.

5 Discretization and Algorithms

In the preceeding chapters we have been using a continuous setting. More specifically the topology of $\Omega \subseteq \mathbb{R}^d$ as a non-empty open set is inherently non-discrete, the function spaces used by the model functional (3.13) and its predual (3.12) have no finite basis (indeed they include polynomials of arbitrary degree) and the duality has been established by Theorem 2.28 in a general Hilbert space setting. Images in practice, however, are given in discrete form (e.g. as an array of brightness values) and algorithms may only act on discrete data, since available memory and computing resources are limited. In this chapter we apply the results from previous chapters to show their relevance and verify their claims numerically.

We consider two standard discretization approaches: finite differences in Section 5.1 and finite elements in Section 5.2 and discuss how to apply them in our setting. For finite elements we derive in Sections 5.2.2 and 5.2.3 two distinct a-posteriori error estimates for use in adaptive algorithms. In Section 5.3 two classic optimization algorithms from [28, 31] are reviewed and reformulated in a general Hilbert space setting, before we derive and evaluate a practical semi-smooth Newton method for our model in Section 5.4. Section 5.6 then concludes with selected numerical examples of the decomposition method from Chapter 4.

New contributions include the discussion of finite element interpolation methods in the context of image processing, the proposal of a pixel-adapted L^2-projection, derivation of two a-posteriori estimates and a semi-smooth Newton method for our generalized model, proposal of an adaptive warping scheme for optical flow, as well as numerical verification of the theoretical convergence results from Chapter 4.

Preliminary work in deriving the finite element residual a-posteriori error estimate has been carried out by Martin Alkämper. The source code to reproduce all numerical examples of this chapter has been made publicly available at [50, 53]. Results of this chapter paired with their theoretical underpinnings from Chapters 3 and 4 are in preparation to be published separately [55, 56].

5.1 Finite Differences

The method of finite differences is concerned with functions defined on a finite, discrete subset $\Omega_h \subseteq \Omega$ and provides derivative operators for those functions, which approximate the corresponding derivatives in the continuous space. Classically, Ω_h is taken as a lattice of equidistantly spaced points, similar to the arrangement of pixels on a screen or in a computer image. Functions can then be represented as arrays of point evaluations, which makes this discretization method a straight-forward fit for image processing.

Consider a rectangular grid Ω_h spanning from $a = (a_1, \dots, a_d) \in h\mathbb{Z}^d$ to $b = (b_1, \dots, b_d) \in h\mathbb{Z}^d$ with grid gap $h > 0$ defined by

$$\Omega_h := \Omega_{h,[a,b]} := \big\{x = (x_1, \dots, x_d) \in h\mathbb{Z}^d : a \leq x \leq b\big\}.$$

Digital images given by an array $A \in [0,1]^{n_1 \times \cdots \times n_d}$, $n = (n_1, \dots, n_d) \in \mathbb{N}^d$ of intensity values between 0 (black) and 1 (white) are then mapped to a discrete function $u : \Omega_{1,[1,n]} \to \mathbb{R}$ by defining $u(x) := A_x$, $x \in \prod_{i=1}^d [1, n_i]$.

Definition 5.1 (Finite Difference Operators). *For $u_h : \Omega_h \to \mathbb{R}^m$ and $p_h = (p_{h,1}, \dots, p_{h,d}) : \Omega_h \to \mathbb{R}^{d \times m}$ let forward differences $\partial^+_{h,k} : \Omega_h \to \mathbb{R}^m$ and backward differences $\partial^-_{h,k} : \Omega_h \to \mathbb{R}^m$ be given by*

$$h\partial^+_{h,k} u_h(x) := \begin{cases} 0 & \text{if } x_k = b_k, \\ u_h(x + he^k) - u_h(x) & \text{else}, \end{cases},$$

$$h\partial^-_{h,k} u_h(x) := \begin{cases} u_h(x) & \text{if } x_k = a_k, \\ -u_h(x - he^k) & \text{if } x_k = b_k, \\ u_h(x) - u_h(x - he^k) & \text{else}, \end{cases}$$

where $e^k \in \mathbb{N}^d$ denotes the k-th unit vector, $k = 1, \ldots, d$. The discrete gradient $\nabla_h u_h : \Omega_h \to \mathbb{R}^{d \times m}$ and discrete divergence $\operatorname{div}_h p_h : \Omega_h \to \mathbb{R}^m$ are then defined as

$$\nabla_h u_h := (\partial^+_{h,k} u_h)_{k=1}^d, \qquad \operatorname{div}_h p_h := \sum_{k=1}^d \partial^-_{h,k} p_{h,k}.$$

Remark 5.2. *The discrete operators ∇_h and $-\operatorname{div}_h$ in Definition 5.1 are adjoint, i.e. $\langle \nabla_h u_h, p_h \rangle_h = \langle u_h, -\operatorname{div}_h p_h \rangle_h$ where $\langle \cdot\,, \cdot \rangle_h$ denotes the L^2 discrete inner product on functions $\Omega_h \to \mathbb{R}^m$.*

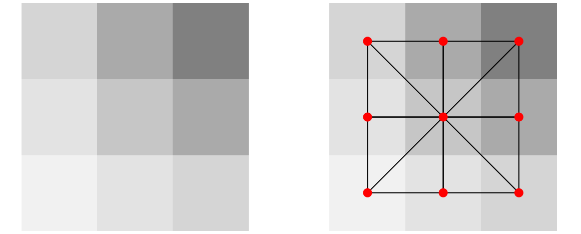
Finite difference discretization following Definition 5.1 often implicitly assumes a continuous model using smooth functions $u \in C^1(\Omega)^m$, $p \in C^1(\Omega)^{d \times m}$. In that model point evaluations are well-defined and approximation results for the derivative operators in Definition 5.1 hold as $h \to 0$. In image processing $h = 1$ is often used implicitly in order to avoid additional interpolation of image data. For the total variation functional from Definition 2.29 this means: while approximation guarantees for $h \to 0$ are important, discretizing in a way as to preserve geometric properties of the total variation as faithfully as possible for fixed $h = 1$ is generally of more practical value.

Let $u_h : \Omega_h \to \mathbb{R}^m$ be a discrete image. Discretizing the constraint on the predual variable in a pointwise sense leads to a primal total variation discretization well-known as *isotropic total variation*:

$$\mathrm{TV}_F(u_h) := \sum_{x \in \Omega_h} |\nabla_h u_h(x)|_F.$$

Though TV_F has been shown to adhere to some of the geometrical properties in Proposition 2.33, it is also, despite its name, not quite isotropic (here: invariant under interpolated spatial rotation) [35]. This topic has sparked renewed theoretical interest and new isotropic discretizations are being proposed [1, 24, 30, 32, 35].

While we will not use TV_F directly, the pointwise finite difference discretization of the predual problem (4.2) corresponds to using TV_F in the primal problem.



(a) 3x3 image given by square pixels

(b) simplicial grid with nodes in pixel centers

Figure 5.1: image aligned simplicial grid construction

5.2 Finite Elements

Central to the idea of finite element discretization is the idea to find a solution within a finite dimensional subspace of the original space. This discrete space usually consists of cellwise polynomial functions defined on a mesh of cells, e.g. simplices. In contrast to finite differences the discretization may be adaptively refined to accomodate for certain local error indicators, thereby reducing the number of degrees of freedom needed to represent a solution up to a certain accuracy. For more details on adaptive finite element schemes in general, we refer the reader to the survey [74].

For a two-dimensional computer image given by an array $A \in [0,1]^{n_1 \times n_2}$ define the domain $\Omega := [1, n_1] \times [1, n_2]$. If not otherwise noted, Ω is triangulated using simplices with nodes at integer coordinates $(x_1, x_2) \in \mathbb{Z}^2$, $1 \leq x_1 \leq n_1$, $1 \leq x_2 \leq n_2$ corresponding to pixel centers as depicted in Figure 5.1.

Let $\mathcal{T}$ denote the set of cells and Γ the set of oriented facets (i.e. edges for $d = 2$) of the simplicial triangulation. For any cell $K \in \mathcal{T}$ let $P_k(K)$ be the space of polynomial functions on K with total degree $k \in \mathbb{N}$. We choose finite subspaces $V_h \subseteq H^1(\Omega)^m \subseteq V$, $W_h^* \subseteq L^2(\Omega) \times L^2(\Omega)^{d \times m} \subseteq$

W^*, $Z_h \subseteq L^2(\Omega)$ as follows:

$$
\begin{aligned}
V_h &:= \{u \in C(\Omega)^m : u|_K \in P_1(K)^m, K \in \mathcal{T}\},\\
W_h^* &:= \{(p_1, \boldsymbol{p}_2) \in C(\Omega) \times L^2(\Omega)^{d\times m} :\\
&\qquad\qquad p_1|_K \in P_1(K), \boldsymbol{p}_2|_K \in P_0(K)^{d\times m}, K \in \mathcal{T}\},\\
Z_h &:= \{g \in C(\Omega) : g|_K \in P_1(K), K \in \mathcal{T}\},
\end{aligned}
\tag{5.1}
$$

i.e. piecewise linear continuous elements for u, g, p_1 and piecewise constant discontinuous elements for p_2.

There are different options to approach the finite element discretization of (3.13). This is due to the fact, that dualization and discretization do not necessarily commute. Indeed, the simple pointwise representations deduced for the dual problem in Proposition 3.9 do not necessarily hold true for subspaces of V. For that reason a modified primal discrete energy is introduced in [49], which allows for a manageable dual representation with direct constraints on the degrees of freedom. Here, we explore a suitable discretization of the continuous optimality conditions (3.16) instead. Namely, in the discrete finite element setting we will search for solutions $p = (p_1, \boldsymbol{p}_2) \in W_h^*$, $u \in V_h$ which satisfy

$$
\begin{aligned}
0 &= \Lambda^* \boldsymbol{p} - \alpha_2 T^* g + Bu,\\
0 &= p_1 \max\{\gamma_1, |Tu - g|\} - \alpha_1(Tu - g), \qquad |p_1| \leq \alpha_1,\\
0 &= \boldsymbol{p}_2 \max\{\gamma_2, |\nabla u|_F\} - \lambda \nabla u, \qquad\quad |\boldsymbol{p}_2|_F \leq \lambda,
\end{aligned}
\tag{5.2}
$$

where the last two equations are enforced on vertices only. This is due to the fact, that the expression $|Tu - g|$ is not necessarily cellwise linear, even though $Tu - g$ is.

For refinement we bisect triangles using the newest-vertex strategy [74], i.e. the bisection edge is chosen to be opposite of the vertex which was inserted last. In an adaptive refinement setting we mark cells for refinement using the greedy Dörfler marking strategy [74] with $\theta_{\text{mark}} = 0.5$, i.e. given error indicators in descending order $(\eta_{K_n})_{1\leq n\leq|\mathcal{T}|}$

for triangles $K_n \in \mathcal{T}$ we refine the first $n_{\text{mark}} \in \mathbb{N}$ triangles that satisfy

$$\sum_{n=1}^{n_{\text{mark}}} \eta_{K_n} \geq \theta_{\text{mark}} \sum_{n=1}^{|\mathcal{T}|} \eta_{K_n}.$$

We start out with an L^2-norm estimate of the gradient operator in our finite element setting, which will be handy for limiting stepsizes in the algorithm to come.

Lemma 5.3. *Let $d = 2$. For every cell $K \in \mathcal{T}$ and every $u \in V_h$ we have the upper bound*

$$\|\nabla u\|_{L^2(K)} \leq \tfrac{6\sqrt{2}}{\varrho_K} \|u\|_{L^2(K)}.$$

Proof. Let $F : \hat{K} \to K$, $\hat{x} \mapsto Ax + b$ be the affine transformation bijectively mapping the reference cell $\hat{K}$ to K and set $\hat{u} := u \circ F$ to be u transformed onto $\hat{K}$. As in the proof of [80, Proposition 3.38], since K contains a ball with diameter ϱ_K and $\hat{K}$ is contained in a ball with diameter $h_{\hat{K}}$, we have

$$\frac{\|\nabla u\|_{L^2(K)}}{\|u\|_{L^2(K)}} = \frac{\|A^{-t}\nabla \hat{u}\|_{L^2(\hat{K})}}{\|\hat{u}\|_{L^2(\hat{K})}} \leq \frac{h_{\hat{K}}}{\varrho_K} \frac{\|\nabla \hat{u}\|_{L^2(\hat{K})}}{\|\hat{u}\|_{L^2(\hat{K})}} = \frac{\sqrt{2}}{\varrho_k} \frac{\|\nabla \hat{u}\|_{L^2(\hat{K})}}{\|\hat{u}\|_{L^2(\hat{K})}}$$

and it remains to bound $\frac{\|\nabla \hat{u}\|_{L^2(\hat{K})}}{\|\hat{u}\|_{L^2(\hat{K})}}$. Representing $\hat{u}$ in local coordinates: $\hat{u}(x, y) = ax + by + c(1 - x - y)$, $a, b, c \in \mathbb{R}$ we explicitly calculate

$$\begin{aligned} \|\nabla \hat{u}\|^2_{L^2(\hat{K})} &= \int_0^1 \int_0^{1-x} |\nabla \hat{u}(x, y)|^2 \,\mathrm{d}y \,\mathrm{d}x \\ &= \tfrac{1}{2}(a^2 + b^2 + 2c^2 - 2ac - 2bc), \\ \|\hat{u}\|^2_{L^2(\hat{K})} &= \int_0^1 \int_0^{1-x} |\hat{u}(x, y)|^2 \,\mathrm{d}y \,\mathrm{d}x \\ &= \tfrac{1}{12}(a^2 + b^2 + c^2 + ab + ac + bc). \end{aligned}$$

Using $0 \leq (a+b+c)^2 = a^2+b^2+c^2+2ab+2ac+2bc$ and $0 \leq (\sqrt{2}x + \frac{c}{\sqrt{2}})^2 = 2x^2 + \frac{c^2}{2} + 2xc$, $x \in \{a, b\}$ we bound

$$\begin{aligned} a^2+b^2+2c^2-2ac-2bc &\leq 4a^2+4b^2+5c^2+6ab+4ac+4bc \\ &\leq 6a^2+6b^2+6c^2+6ab+6ac+6bc \end{aligned}$$

and infer $\|\nabla\hat{u}\|^2_{L^2(\hat{K})} \leq 6 \cdot \frac{12}{2}\|\hat{u}\|^2_{L^2(\hat{K})} = 36\|\hat{u}\|^2_{L^2(\hat{K})}$. Combining this with the transformation above, we get

$$\frac{\|\nabla u\|_{L^2(K)}}{\|u\|_{L^2(K)}} \leq \frac{\sqrt{2}}{\varrho_k}\frac{\|\nabla\hat{u}\|_{L^2(\hat{K})}}{\|\hat{u}\|_{L^2(\hat{K})}} \leq \frac{6\sqrt{2}}{\varrho_k}. \qquad \square$$

5.2.1 On Image Interpolation Methods

For aligned grids as in Figure 5.1 image data $A \in [0,1]^{n_1 \times n_2}$ is interpolated to a grid function $g_h \in Z_h$ by nodal sampling. This way, when resampling at pixel centers one receives the original image. For non-aligned grids , i.e. when pixel centers do not correspond to vertices of the grid, a mapping of the image A to some grid function $g_h \in Z_h$ has to be computed. Chosing nodal sampling (e.g. of bilinearly interpolated image data) for this map can result in aliasing and is a well known undesired effect in image processing [21]. The Shannon-Whittaker sampling theorem provides ample conditions for correct equidistant sampling of bandwith-limited functions [21, 78], namely to remove high-frequency contributions (e.g. by a linear smoothing filter) before sampling on a coarser grid. It is however unclear to us how to apply this to general unstructured sampling points.

We evaluate three projection schemes and then explore one alternative. Let $g \in C(\Omega)$ be the piecewise bilinear interpolation of the nodal image data A. For methods involving quadratures the quadrature over a cell $E \in \mathcal{T}$ is computed using a simple averaging quadrature on a Lagrange lattice of degree $\lceil \operatorname{diam}(E) \rceil$, i.e. we properly scale the number of quadrature points depending on the size of the cell (and thus the number of pixels it covers) to avoid aliasing effects.

The projection schemes are defined as follows:

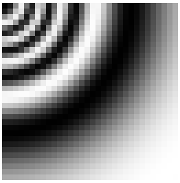


Figure 5.2: Discrete input image for mesh interpolation comparison, 32×32 pixels

(i) `nodal`: Nodal interpolation, i.e. $g_h \in Z_h$ such that $g_h(x) := g(x)$ at every mesh vertex x.

(ii) `l2_lagrange`: Standard L^2-projection, i.e. $g_h \in Z_h$ such that $\int_\Omega g_h \cdot \varphi \, \mathrm{d}x = \int_\Omega g \cdot \varphi \, \mathrm{d}x$ for all $\varphi \in Z_h$.

(iii) `qi_lagrange`: The general L^1-stable quasi-interpolation operator as proposed in [45], i.e. the continuous $g_h \in Z_h$ is given by setting its nodal degrees of freedom at each vertex to the arithmetic mean of the corresponding local degrees of freedom of a discontinuous interpolant $\widetilde{g}_h \in \widetilde{Z}_h := \{f \in L^\infty(\Omega) : f|_K \in P_1(K), K \in \mathcal{T}\}$, which is defined on each cell $K \in \mathcal{T}$ by its local nodal degrees of freedom $\sigma_{K,i} = \frac{1}{|K|} \int_K g(x) \cdot \varrho_{K,i}(x) \, \mathrm{d}x$, $i = 1, \dots, 3$ where the test function $\varrho_{K,i}$ in our case is given in barycentric coordinates λ_K by $\varrho_{K,i}(\lambda_K) = 12\lambda_{K,i} - 3$. The reader may refer to [45] for details on the general construction.

(iv) `l2_pixel`: Our proposed method, i.e. minimizing the sum of pointwise squared errors over all pixel coordinates amounting to: $\inf_{g_h \in Z_h} \sum_{x \in \Omega \cap \mathbb{Z}^2} \frac{1}{2} \|g_h(x) - g(x)\|^2$.

We note, that `l2_pixel` may be interpreted as an L^2-projection with a cell-dependent averaging quadrature rule, adapted to the orginal image pixel locations.

We evaluate these schemes by interpolating the discrete image given in Figure 5.2 onto V_h for regular meshes of different sizes. In Figure 5.3 we see the interpolated results as cellwise linear functions.

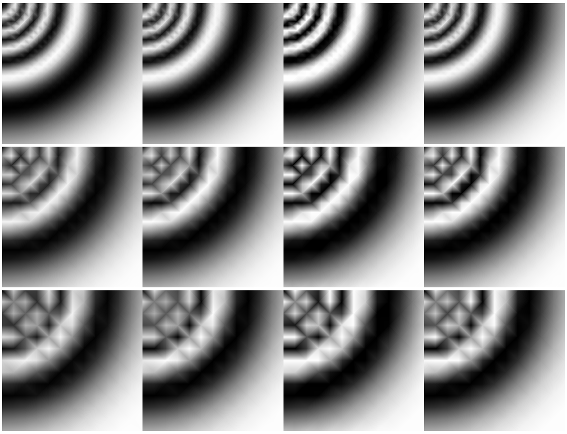


Figure 5.3: Interpolated finite element functions for varying mesh sizes and interpolation methods. Methods from left to right: `nodal`, `l2_lagrange`, `qi_lagrange`, `l2_pixel`. Number of mesh vertices in each dimension from top to bottom: 32, 16, 13.

Table 5.1: PSNR values (higher is better) of interpolated mesh functions from Figure 5.3, sampled at image coordinates.

mesh_size	nodal	l2_lagrange	qi_lagrange	l2_pixel
32	∞	55.826	22.606	327.597
16	21.141	22.020	21.483	23.106
13	18.861	19.650	19.053	19.981

Table 5.2: SSIM values (higher is better) of interpolated mesh functions from Figure 5.3, sampled at image coordinates.

mesh_size	nodal	l2_lagrange	qi_lagrange	l2_pixel
32	1.00000	0.99998	0.97846	1.00000
16	0.93868	0.95133	0.94757	0.96693
13	0.89307	0.90888	0.90053	0.92080

Note that e.g. for mesh size 16 the results from `l2_pixel` appear to be less blurry than `nodal` or `l2_lagrange`, while avoiding an excessive sharpening effect observed for `qi_lagrange` in the upper left of the image. Next, we sample the interpolated functions on the original image grid Ω_h and quantify the error to the original image. In particular, Tables 5.1 and 5.2 list the peak signal-to-noise ratio PSNR, given for two scalar images $u, v : \Omega_h \to [0, 1]$ by $\mathrm{PSNR}(u, v) := -10 \log_{10} \big(\frac{1}{|\Omega_h|} \sum_{x \in \Omega_h} (u(x) - v(x))^2 \big)$, and the structural similarity index SSIM as defined in [81]. Perhaps unsurprisingly, due to construction, `l2_pixel` produces results closest to the original image for both metrics. We note that, theoretically, `l2_pixel` in the finest mesh setting should provide exact results (i.e. infinite PSNR) and the discrepancy is due to numerical imprecision.

It should be noted, that while `l2_pixel` seems to provide visually superior results, it does not necessarily preserve other quantities of the image, such as the total mass, which may be relevant for the image processing task in question. Further, it is not necessarily well-defined for meshes finer than the original image and in that case a regularization needs to be applied.

We conclude that, since digital images are usually given as an array of brightness values and the output of image processing algorithms is expected to be so too, working in unstructured finite element spaces comes with the drawback of information loss due to mesh interpolation. Apart from increased complexity, this may be considered a problem for applying unstructured adaptive finite element methods to image processing tasks in practice.

5.2.2 Primal-Dual A-Posteriori Error Estimator

In the following we use the same approach as in [12, 14] to derive a-posteriori error estimates from the primal dual energy gap. We slightly adjust the arguments to account for the Huber regularization and potentially non-local operators T and B.

Lemma 5.4. *Let H be a real Hilbert space with inner product $\langle \cdot, \cdot \rangle_H$ and associated norm $\| \cdot \|_H$. Then for every $v, w_h, w \in H$ one has*

$$\begin{aligned} \|v - w\|_H^2 - \|v - w_h\|_H^2 &= \langle 2v - w_h - w, w_h - w \rangle_H \\ &\leq (\|v - w\|_H + \|v - w_h\|_H) \|w - w_h\|_H. \end{aligned}$$

Proof. For the equality one has

$$\begin{aligned} &\|v - w\|_H^2 - \|v - w_h\|_H^2 \\ &\quad = \langle v - w, v - w \rangle_H - \langle v - w_h, v - w + w - w_h \rangle_H \\ &\quad = \langle w_h - w, v - w \rangle_H - \langle v - w_h, w - w_h \rangle_H \\ &\quad = \langle v - w, w_h - w \rangle_H + \langle v - w_h, w_h - w \rangle_H \\ &\quad = \langle 2v - w - w_h, w_h - w \rangle_H. \end{aligned}$$

For the inequality one has

$$\begin{aligned} \langle 2v - w_h - w, w_h - w \rangle_H &= \langle v - w_h, w_h - w \rangle_H + \langle v - w, w_h - w \rangle_H \\ &\leq (\|v - w\|_H + \|v - w_h\|_H) \|w - w_h\|_H. \qquad \square \end{aligned}$$

Note that we may apply Lemma 5.4 in particular to the weighted scalar product $\langle \cdot, B^{-1} \cdot \rangle$ with associated norm $\| \cdot \|_{B^{-1}}$.

Let $g_h \in Z_h$ be the L^2-projection of g onto Z_h. Recalling (3.12) and (3.13) we define discretized functionals of E^* and E by $E_h^* : W_h^* \to \overline{\mathbb{R}}$ and $E_h : V_h \to \overline{\mathbb{R}}$ using the discretized data g_h as

$$\begin{aligned} E_h^*(\boldsymbol{p}_h) &:= \tfrac{1}{2} \|T^* p_{h,1} + \nabla^* \boldsymbol{p}_{h,2} - \alpha_2 T^* g_h\|_{B^{-1}}^2 - \tfrac{\alpha_2}{2} \|g_h\|_{L^2}^2 + \langle g_h, p_{h,1} \rangle \\ &\qquad + \chi_{|p_{h,1}| \leq \alpha_1} + \chi_{|\boldsymbol{p}_{h,2}|_F \leq \lambda} + \tfrac{\gamma_1}{2\alpha_1} \|p_{h,1}\|_{L^2}^2 + \tfrac{\gamma_2}{2\lambda} \|\boldsymbol{p}_{h,2}\|_{L^2}^2, \\ E_h(u_h) &:= \tfrac{\alpha_2}{2} \|T u_h - g_h\|_{L^2}^2 + \tfrac{\beta}{2} \|S u_h\|_{L^2}^2 \\ &\qquad + \alpha_1 \int_\Omega \varphi_{\gamma_1}(|T u_h - g_h|) \, \mathrm{d}x + \lambda \int_\Omega \varphi_{\gamma_2}(|\nabla u_h|_F) \, \mathrm{d}x. \end{aligned}$$

Theorem 5.5. *Let $u \in V$ be the solution to (3.13). Further, let $u_h \in V_h$ be the discrete minimizer of E_h. Then one has for any $v_h \in V_h$, $\boldsymbol{q}_h \in W_h^*$ the a-posteriori error estimate*

$$\tfrac{1}{2}\|u_h - u\|_B^2 \leq \eta_h^2(v_h, \boldsymbol{q}_h) + c\|g_h - g\|_{L^2}$$

where the estimator $\eta_h : V_h \times W_h^* \to \overline{\mathbb{R}}$ is given as

$$\eta_h^2(v_h, \boldsymbol{q}_h) := E_h(v_h) + E_h^*(\boldsymbol{q}_h) \tag{5.3}$$

and $c > 0$ is a constant depending only on the model parameters, the domain size $|\Omega|$ and data g.

Proof. Let $\boldsymbol{p} \in W^*$ be a solution to (3.12) and $\boldsymbol{p}_h \in W_h^*$ be a discrete minimizer of E_h^*.

Let $v_h \in V_h$ and $\boldsymbol{q}_h \in W_h^*$ be arbitrary. Then we have due to Lemma 3.7, strong duality from Theorem 2.28, optimality of p in E^* and optimality of u_h and $\boldsymbol{p}_h$ in E_h and E_h^* respectively:

$$\begin{aligned}
\tfrac{1}{2}\|u_h - u\|_B^2 &\leq E(u_h) - E(u) \\
&= E(u_h) + E^*(p) \\
&\leq E(u_h) + E^*(\boldsymbol{p}_h) \\
&= \eta_h^2(u_h, \boldsymbol{p}_h) + E(u_h) - E_h(u_h) + E^*(\boldsymbol{p}_h) - E_h^*(\boldsymbol{p}_h) \\
&\leq \eta_h^2(v_h, \boldsymbol{q}_h) + E(u_h) - E_h(u_h) + E^*(\boldsymbol{p}_h) - E_h^*(\boldsymbol{p}_h).
\end{aligned}$$

It remains to bound the data approximation errors $E(u_h) - E_h(u_h)$ and $E^*(\boldsymbol{p}_h) - E_h^*(\boldsymbol{p}_h)$.

According to Proposition 3.10 the Huber function satisfies $|\varphi'_{\gamma_1}(x)| \leq 1$ for all $x \in \mathbb{R}$, so φ is in particular Lipschitz-continuous with Lipschitz-constant 1 and we may use the inverse triangle inequality to obtain

$$\begin{aligned}
\varphi_{\gamma_1}(|Tu_h - g|) - \varphi_{\gamma_1}(|Tu_h - g_h|) &\leq \big||Tu_h - g| - |Tu_h - g_h|\big| \\
&\leq |g - g_h|.
\end{aligned}$$

We further note in advance that since g_h is the projection of g onto Z_h we have due to Lemma 2.8 $\|g_h\| \leq \|g\|$ and

$$\begin{aligned}\|g_h - g\|_{L^2} &\leq \|g\|_{L^2},\\ \|Tu_h - g\|_{L^2} &\leq \|Tu_h - g_h\|_{L^2} + \|g_h - g\|_{L^2}.\end{aligned}$$

Using that and Lemma 5.4 we estimate:

$$\begin{aligned}
&E(u_h) - E_h(u_h)\\
&\quad= \alpha_1 \int_\Omega \varphi_{\gamma_1}(|Tu_h - g|) - \varphi_{\gamma_1}(|Tu_h - g_h|)\,\mathrm{d}x\\
&\qquad\quad + \tfrac{\alpha_2}{2}\big(\|Tu_h - g\|_{L^2}^2 - \|Tu_h - g_h\|_{L^2}^2\big)\\
&\quad\leq \alpha_1 \int_\Omega |g - g_h|\,\mathrm{d}x + \tfrac{\alpha_2}{2}\big(\|Tu_h - g\|_{L^2} + \|Tu_h - g_h\|_{L^2}\big)\|g - g_h\|_{L^2}\\
&\quad\leq \alpha_1 \int_\Omega |g - g_h|\,\mathrm{d}x + \tfrac{\alpha_2}{2}\big(2\|Tu_h - g_h\|_{L^2} + \|g - g_h\|_{L^2}\big)\|g - g_h\|_{L^2}\\
&\quad\leq \alpha_1 |\Omega|^{\frac{1}{2}} \|g - g_h\|_{L^2} + \tfrac{\alpha_2}{2}\big(2E_h(0)^{\frac{1}{2}} + \|g\|_{L^2}\big)\|g - g_h\|_{L^2}\\
&\quad= c_E \|g - g_h\|_{L^2}
\end{aligned}$$

for the constant

$$c_E := \alpha_1 |\Omega|^{\frac{1}{2}} + \alpha_2 E_h(0)^{\frac{1}{2}} + \tfrac{\alpha_2}{2}\|g\|_{L^2} > 0$$

where $E_h(0)$ may be bounded independently of g_h:

$$\begin{aligned}
E_h(0) &= \alpha_1 \int_\Omega \varphi_{\gamma_1}(|g_h|) + \tfrac{\alpha_2}{2}\|g_h\|_{L^2}^2\\
&\leq \alpha_1 \|g_h\|_{L^1} + \tfrac{\alpha_2}{2}\|g\|_{L^2}^2\\
&\leq 2\alpha_1 |\Omega|^{\frac{1}{2}} \|g\|_{L^2} + \tfrac{\alpha_2}{2}\|g\|_{L^2}^2.
\end{aligned}$$

Then for the predual data approximation error one has

$$
\begin{aligned}
&E^*(\boldsymbol{p}_h) - E_h^*(\boldsymbol{p}_h) \\
&\quad = \tfrac{1}{2}\|\Lambda^*\boldsymbol{p}_h - \alpha_2 T^* g\|_{B^{-1}}^2 - \tfrac{1}{2}\|\Lambda^*\boldsymbol{p}_h - \alpha_2 T^* g_h\|_{B^{-1}}^2 \\
&\qquad\quad + \tfrac{\alpha_2}{2}(\|g_h\|_{L^2}^2 - \|g\|_{L^2}^2) + \langle g - g_h, p_{h,1}\rangle \\
&\quad \le \tfrac{1}{2}\Big(\|\Lambda^*\boldsymbol{p}_h - \alpha_2 T^* g\|_{B^{-1}} + \|\Lambda^*\boldsymbol{p}_h - \alpha_2 T^* g_h\|_{B^{-1}}\Big)\|g - g_h\|_{B^{-1}} \\
&\qquad\quad + \tfrac{\alpha_2}{2}\langle g_h - g, g_h + g\rangle + \|p_{h,1}\|_{L^2}\|g - g_h\|_{L^2} \\
&\quad \le \tfrac{1}{2}\Big(2\|\Lambda^*\boldsymbol{p}_h - \alpha_2 T^* g\|_{B^{-1}} + \|\alpha_2 T^* g - \alpha_2 T^* g_h\|_{B^{-1}}\Big)\|g - g_h\|_{B^{-1}} \\
&\qquad\quad + \tfrac{\alpha_2}{2}\|g_h - g\|_{L^2}\|g_h + g\|_{L^2} + |\Omega|^{\frac{1}{2}}\|p_{h,1}\|_{L^\infty}\|g - g_h\|_{L^2} \\
&\quad \le \tfrac{1}{2}\Big(2E_h^*(0)^{\frac{1}{2}} + \alpha_2\|B^{-1}\|^{\frac{1}{2}}\|T\|\|g - g_h\|_{L^2}\Big)\|B^{-1}\|^{\frac{1}{2}}\|g - g_h\|_{L^2} \\
&\qquad\quad + \alpha_2\|g\|_{L^2}\|g_h - g\|_{L^2} + |\Omega|^{\frac{1}{2}}\alpha_1\|g - g_h\|_{L^2} \\
&\quad = c_{E^*}\|g - g_h\|_{L^2}
\end{aligned}
$$

for the constant $c_{E^*} := \|B^{-1}\|^{\frac{1}{2}}E_h^*(0)^{\frac{1}{2}} + \frac{\alpha_2}{2}\|B^{-1}\|\|T\|\|g\|_{L^2} + \alpha_2\|g\|_{L^2} + |\Omega|^{\frac{1}{2}}\alpha_1 > 0$, where $E_h^*(0)$ may be bounded independent of g_h:

$$
\begin{aligned}
E_h^*(0) &= \tfrac{1}{2}\|\alpha_2 T^* g_h\|_{B^{-1}}^2 - \tfrac{\alpha_2}{2}\|g_h\|_{L^2}^2 \\
&\le \tfrac{\alpha_2^2}{2}\|T\|^2\|B^{-1}\|\|g_h\|_{L^2}^2 - \tfrac{\alpha_2}{2}\|g_h\|_{L^2}^2 \\
&\le \max\{0, \tfrac{\alpha_2}{2}(\alpha_2\|T\|^2\|B^{-1}\| - 1)\}\|g\|_{L^2}.
\end{aligned}
$$

We finish the proof by combining both results:

$$
\tfrac{1}{2}\|u_h - u\|_B^2 \le \eta_h^2(v_h, \boldsymbol{q}_h) + (c_E + c_{E^*})\|g_h - g\|_{L^2}
$$

with $c := c_E + c_{E^*}$ independent of g_h. □

Remark 5.6. *The estimator* (5.3) *has the equivalent representation*

$$\begin{aligned}\eta_h^2(v_h, \boldsymbol{q}_h) = \alpha_1 \int_\Omega \varphi_{\gamma_1}(|Tv_h - g_h|)\,\mathrm{d}x - \langle Tv_h - g_h, q_{h,1}\rangle + \tfrac{\gamma_1}{2\alpha_1}\|q_{h,1}\|_{L^2}^2 \\ + \lambda \int_\Omega \varphi_{\gamma_2}(|\nabla v_h|_F)\,\mathrm{d}x - \langle \nabla v_h, \boldsymbol{q}_{h,2}\rangle + \tfrac{\gamma_2}{2\lambda}\|\boldsymbol{q}_{h,2}\|_{L^2}^2 \\ + \tfrac{1}{2}\left\|B^{-1}(\Lambda^*\boldsymbol{q}_h - \alpha_2 T^* g_h) + v_h\right\|_B^2\end{aligned}$$

for all $v_h \in V_h$, $\boldsymbol{q}_h \in W_h^*$ with $|q_{h,1}| \le \alpha_1$, $|\boldsymbol{q}_{h,2}|_F \le \lambda$.

In the limits $\alpha_1 \to 0$ or $\lambda \to 0$ the terms $\frac{\gamma_1}{2\alpha_1}\|q_{h,1}\|_{L^2}^2$ and $\frac{\gamma_2}{2\lambda}\|\boldsymbol{q}_{h,2}\|_{L^2}^2$ vanish respectively due to the constraints on $q_{h,1}, q_{h,2}$. Thus the above estimate is to be interpreted for $\alpha_1 = 0$ or $\lambda = 0$ by omitting the corresponding terms.

Proof. We break up the terms of

$$\begin{aligned}\eta_h^2(v_h, \boldsymbol{q}_h) &= E_h(v_h) + E_h^*(\boldsymbol{q}_h) \\ &= \alpha_1 \int_\Omega \varphi_1(|Tv_h - g_h|)\,\mathrm{d}x + \lambda \int_\Omega \varphi_{\gamma_2}(|\nabla v_h|_F)\,\mathrm{d}x \\ &\qquad + \tfrac{\alpha_2}{2}\|Tv_h - g_h\|_{L^2}^2 + \tfrac{\beta}{2}\|Sv_h\|_{L^2}^2 \\ &\qquad + \tfrac{1}{2}\|\Lambda^*\boldsymbol{q}_h - \alpha_2 T^* g_h\|_{B^{-1}}^2 - \tfrac{\alpha_2}{2}\|g_h\|_{L^2}^2 + \langle g_h, q_{h,1}\rangle \\ &\qquad + \tfrac{\gamma_1}{2\alpha_1}\|q_{h,1}\|_{L^2}^2 + \tfrac{\gamma_2}{2\lambda}\|\boldsymbol{q}_{h,2}\|_{L^2}^2.\end{aligned}$$

First we expand

$$\begin{aligned}\alpha_1 &\int_\Omega \varphi_1(|Tv_h - g_h|)\,\mathrm{d}x + \tfrac{\gamma_1}{2\alpha_1}\|q_{h,1}\|_{L^2}^2 \\ &= \alpha_1 \int_\Omega \varphi_1(|Tv_h - g_h|)\,\mathrm{d}x - \langle Tv_h - g_h, q_{h,1}\rangle \\ &\qquad + \tfrac{\gamma_1}{2\alpha_1}\|q_{h,1}\|_{L^2}^2 + \langle Tv_h - g_h, q_{h,1}\rangle \\ \lambda &\int_\Omega \varphi_{\gamma_2}(|\nabla v_h|_F)\,\mathrm{d}x + \tfrac{\gamma_2}{2\lambda}\|\boldsymbol{q}_{h,2}\|_{L^2}^2 \\ &= \lambda \int_\Omega \varphi_{\gamma_2}(|\nabla v_h|_F)\,\mathrm{d}x - \langle \nabla v_h, q_{h,2}\rangle \\ &\qquad + \tfrac{\gamma_2}{2\lambda}\|\boldsymbol{q}_{h,2}\|_{L^2}^2 + \langle \nabla v_h, \boldsymbol{q}_{h,2}\rangle\end{aligned}$$

which takes care of the first two summands. The remaining terms add up to

$$\begin{aligned}
&\tfrac{1}{2}\|\Lambda^*\boldsymbol{q}_h - \alpha_2 T^* g_h\|_{B^{-1}}^2 + \tfrac{\alpha_2}{2}\|Tv_h - g_h\|_{L^2}^2 - \tfrac{\alpha_2}{2}\|g_h\|_{L^2}^2 + \tfrac{\beta}{2}\|Sv_h\|_{L^2}^2 \\
&\qquad + \langle g_h, q_{h,1}\rangle + \langle Tv_h - g_h, q_{h,1}\rangle + \langle \nabla v_h, \boldsymbol{q}_{h,2}\rangle \\
&\quad = \tfrac{1}{2}\|B^{-1}(\Lambda^*\boldsymbol{q}_h - \alpha_2 T^* g_h)\|_B^2 + \tfrac{\alpha_2}{2}\|Tv_h\|_{L^2}^2 - \alpha_2\langle Tv_h, g_h\rangle \\
&\qquad + \tfrac{\beta}{2}\|Sv_h\|_{L^2}^2 + \langle v_h, T^* q_{h,1}\rangle + \langle v_h, \nabla^*\boldsymbol{q}_{h,2}\rangle \\
&\quad = \tfrac{1}{2}\|B^{-1}(\Lambda^*\boldsymbol{q}_h - \alpha_2 T^* g_h)\|_B^2 + \tfrac{1}{2}\|v_h\|_B^2 \\
&\qquad + \langle v_h, BB^{-1}(\Lambda^*\boldsymbol{q}_h - \alpha_2 T^* g_h)\rangle \\
&\quad = \tfrac{1}{2}\|B^{-1}(\Lambda^*\boldsymbol{q}_h - \alpha_2 T^* g_h) + v_h\|_B^2
\end{aligned}$$

which constitutes the third summand. □

Remark 5.6 motivates a corresponding cell-wise error indicator as follows.

Definition 5.7. *Let $u_h \in V_h$, $\boldsymbol{p}_h \in W_h^*$ be feasible discrete approximations to the solution of* (3.16)*. The primal reconstruction from the dual variable is given weakly by*

$$w := -B^{-1}(\Lambda^*\boldsymbol{p}_h - \alpha_2 T^* g_h).$$

For every cell $K \in \mathcal{T}$ we then define the cell indicator as

$$\begin{aligned}
&\eta_{K,1}^2(u_h, p_h) \\
&\quad := \int_K \alpha_1\varphi_{\gamma_1}(|Tu_h - g_h|) - (Tu_h - g_h)\cdot p_{h,1} + \tfrac{\gamma_1}{2\alpha_1}|p_{h,1}|^2 \,\mathrm{d}x \\
&\qquad + \int_K \lambda\varphi_{\gamma_2}(|\nabla u_h|_F) - \nabla u_h \cdot \boldsymbol{p}_{h,2} + \tfrac{\gamma_2}{2\lambda}|\boldsymbol{p}_{h,2}|_F^2 \,\mathrm{d}x \\
&\qquad + \tfrac{1}{2}\int_K |w - u_h|^2 \,\mathrm{d}x.
\end{aligned}$$

5.2.3 Residual A-Posteriori Error Estimator

In [61] an a-posteriori error estimate for a smooth functional, composed of an L^2-data term and a total variation term, was derived using residual-based methods for variational inequalities. We try to use a similar technique for our general setting to establish a suitable guiding criterion for adaptive spatial discretization.

Our presentation avoids the variational inequality setting in [61] by making use of strong convexity. Additionally, we consider both choices of Settings (S.i) and (S.ii), which requires different interpolation estimates depending on the space V and the corresponding coercivity of a_B stated in Assumption (A1).

Since the derivation requires local (cell-wise) adjoints T^*, we will make the following locality assumption on the operator T.

(A2) For almost every $x \in \Omega$ there exists $f_x : \mathbb{R}^m \to \mathbb{R}$ such that:

$$(Tu)(x) = f_x(u(x)). \tag{5.4}$$

Note that while Assumption (A2) excludes global operators such as Fourier transforms. Implementing global operators in an unstructured finite element setting efficiently, however, is considered impractical at best because of the non-local access pattern and will therefore not be pursued here.

We fix $V := H^1(\Omega)^m$ and recall $E : V \to \overline{\mathbb{R}}$ from (3.13) in the split form $E(u) = F(u) + G(\Lambda u)$, where $\Lambda = (T, \nabla) : V \to W$ as in Chapter 3 and $F : V \to \overline{\mathbb{R}}$, $G : W \to \overline{\mathbb{R}}$ are given by

$$\begin{aligned} F(u) &:= \tfrac{\alpha_2}{2}\|Tu - g\|_{L^2}^2 + \tfrac{\beta}{2}\|Su\|_{L^2}^2, \\ G(\Lambda u) &:= \alpha_1 \int_\Omega \varphi_{\gamma_1}(|Tu - g|)\,\mathrm{d}x + \lambda \int_\Omega \varphi_{\gamma_2}(|\nabla u|_F)\,\mathrm{d}x. \end{aligned}$$

The optimality conditions in this setting, see Theorem 2.28, are given by

$$\begin{aligned} -\Lambda^* p &\in \partial F(u), \\ p &\in \partial G(\Lambda u), \end{aligned}$$

where we accounted for the sign change in the dual variable p to be consistent with the formulation in (3.16). We assume that the discrete solution pair $u_h \in V_h$, $p_h \in W_h^*$ satisfies the corresponding discrete optimality conditions

$$\begin{aligned} -\Lambda^* p_h &\in \partial F_h(u_h), \\ p_h &\in \partial G_h(\Lambda u_h), \end{aligned} \tag{5.5}$$

where for clarity we explicitly denoted with $F_h : V_h \to \overline{\mathbb{R}}$, $G_h : W_h \to \overline{\mathbb{R}}$ the restrictions of F, G on the respective discrete spaces, since their subdifferentials differ from those of F and G respectively. One may obtain this system analagously to Chapter 3 by application of the Fenchel duality from Theorem 2.28 on the corresponding discrete energy functional. Indeed, Theorem 3.5 in general allows for V and W to be appropriate subspaces. This way existence of a discrete solution pair $u_h \in V_h$, $p_h \in W_h^*$ is ensured. Since F is Fréchet-differentiable we see that

$$\begin{aligned} \langle \partial F(u), v\rangle_{V^*,V} &= \alpha_2\langle Tu - g, Tv\rangle + \beta\langle Su, Sv\rangle \\ &= a_B(u,v) - l(v) \end{aligned}$$

with $l(v) := \alpha_2\langle Tv, g\rangle$. Analogously $\langle \partial F_h(u_h), v_h\rangle_{V_h^*,V_h} = a_B(u_h, v_h) - l(v_h)$ and from (5.5) we infer for all $v_h \in V_h$:

$$-\langle p_h, \Lambda v_h\rangle = a_B(u_h, v_h) - l(v_h). \tag{5.6}$$

For ∂G on the other hand we have the following non-obvious observation.

Lemma 5.8. *Let $p_h = (p_{h,1}, p_{h,2}) \in W_h^*$ satisfy* (5.5)*. If $\alpha_1 = 0$ then*

$$p_h \in \partial G(\Lambda u_h),$$

where $p_h \in W_h^* \subseteq W^*$ is identified with an element of the dual space $(W^*)^* = W$.

Proof. Let $q = (q_1, q_2) \in \partial G(\Lambda u_h)$. Due to (3.16) and (5.5) we deduce $q_1 = p_{h,1} = 0$ since $\alpha_1 = 0$. From (3.16) we infer for q_2, and analogously for $p_{h,2}$, due to (5.5) that pointwise, whenever $\nabla u_h \neq 0$, we have

$$q_2 = \frac{\lambda \nabla u_h}{\max\{\gamma_2, |\nabla u_h|_F\}} = p_{h,2}.$$

Therefore q_2 is piecewise constant whenever $\nabla u_h \neq 0$. Further both q_2 and $p_{h,2}$ are bounded pointwise by $|q_2|_F \leq \lambda$ and $|p_{h,2}|_F \leq \lambda$. On cells where $\nabla u_h = 0$ holds, q_2 may assume any value bounded by $|q_2|_F \leq \lambda$. Since this holds true for $p_{h,2}$ in particular, one may choose $q := (0, p_{h,2}) \in \partial G(\Lambda u_h)$. □

We note that Lemma 5.8 does not necessarily hold true for $\alpha_1 > 0$. Indeed $\partial G(\Lambda u_h) = \alpha_1 \frac{Tu_h - g}{\max\{\gamma_1, |Tu_h - g|\}}$ does not need to be piecewise linear, even if u_h is. We will continue the derivation nevertheless, noting that theoretical justification is lacking for $\alpha_1 > 0$.

With these observations we are ready to estimate the error $\|u_h - u\|_B$. Denote $e_h := u_h - u \in V$ and bound using Lemma 3.7:

$$\begin{aligned}\tfrac{1}{2}\|u_h - u\|_B^2 &\leq E(u_h) - E(u)\\ &= F(u_h) - F(u) + G(\Lambda u_h) - G(\Lambda u)\\ &\leq \langle \partial F(u_h), e_h\rangle + \langle \partial G(\Lambda u_h), \Lambda e_h\rangle\\ &= a_B(u_h, e_h) - l(e_h) + \langle \partial G(\Lambda u_h), \Lambda e_h\rangle.\end{aligned}$$

In particular due to Lemma 5.8 and discrete optimality (5.6) we have for any bounded linear operator $I_h : V \to V_h$:

$$\begin{aligned}\tfrac{1}{2}\|u_h - u\|_B^2 &\leq a_B(u_h, e_h) - l(e_h) + \langle p_h, \Lambda e_h\rangle\\ &= a_B(u_h, e_h - I_h e_h) - l(e_h - I_h e_h)\\ &\qquad + \langle p_h, \Lambda(e_h - I_h e_h)\rangle \qquad (5.7)\\ &= \langle \alpha_2(Tu_h - g) + p_{h,1}, T(e_h - I_h e_h)\rangle\\ &\qquad + \langle \beta S u_h, S(e_h - I_h e_h)\rangle + \langle p_{h,2}, \nabla(e_h - I_h e_h)\rangle,\end{aligned}$$

Now choose $I_h : V \to V_h$ to be a quasi-interpolation operator which satisfies the interpolation estimates [80, Proposition 1.3] [3, Theorem 1.7]:

$$\|v - I_h v\|_{L^2(K)} \leq c_1 h_K \|\nabla v\|_{L^2(\omega_K)}, \qquad (5.8)$$

$$\|v - I_h v\|_{L^2(F)} \leq c_2 h_F^{\frac{1}{2}} \|\nabla v\|_{L^2(\omega_F)}, \qquad (5.9)$$

$$\|v - I_h v\|_{L^2(F)} \leq c_3 h_F^{-\frac{1}{2}} \|v\|_{L^2(\omega_F)}, \qquad (5.10)$$

where ω_K, ω_F denote the union of triangles which share a common vertex with the cell K or the facet F respectively and h_K, h_F denote the diameter of K and F respectively. Depending on S, we will now use the interpolation estimates to further bound the error $\frac{1}{2}\|u_h - u\|_B^2$.

Setting (S.ii): the Case $\mathrm{S} = \nabla$

Using Setting (S.ii) we have $V = H^1(\Omega)^m$ and will make use of the fact that a_B is coercive on V, see Assumption (A1), to derive an estimate on the error $\|e_h\|_{H^1(\Omega)^m}$.

Let $F \in \Gamma$ be an oriented inner facet with adjacent cells $K_1, K_2 \in \mathcal{T}$ and $\varphi \in L^2(K_1 \cup K_2)^{d\times m}$ with $\varphi|_{K_i} \in C(K_i)^{d\times m}$, $i \in \{1,2\}$ a square integrable function which allows continuous representations $\varphi|_{K_i}$ on each cell K_i individually. We then define the jump term $[\varphi]_F \in C(F)^{d\times m}$ by $[\varphi]_F(x) := \varphi|_{K_1}(x) - \varphi|_{K_2}(x)$ and omit the index as in $[\varphi]$ when the facet in question is clear. For outer facets only the term corresponding to the existing adjacent cell is considered. Note that $[\varphi]_F$ is in general dependent on the orientation of F, while e.g. $[n^T\varphi]_F$ for the oriented facet-normal n is not.

We will now bound the error in the norm induced by the bilinear form $a_B(\cdot,\cdot)$ starting from (5.7).

$$
\begin{aligned}
&\tfrac{1}{2}\|u_h - u\|_B^2 \\
&\quad \le \big\langle \alpha_2(Tu_h - g) + p_{h,1}, T(e_h - I_h e_h)\big\rangle \\
&\qquad\quad + \big\langle \beta\nabla u_h + p_{h,2}, \nabla(e_h - I_h e_h)\big\rangle \\
&\quad = \sum_{K\in\mathcal{T}} \Bigg(\Big\langle \alpha_2 T^*(Tu_h - g) + T^* p_{h,1} - \beta\Delta u_h - \operatorname{div} p_{h,2}, \\
&\qquad\qquad\qquad e_h - I_h e_h \Big\rangle_{L^2(K)^m} \\
&\qquad\qquad + \Big\langle n^T(\beta\nabla u_h + p_{h,2}), e_h - I_h e_h \Big\rangle_{L^2(\partial K)^m} \Bigg),
\end{aligned}
$$

where we used locality of T from Assumption (A2).

Denoting

$$\begin{aligned}
&\xi := \alpha_2 T^*(Tu_h - g) + T^* p_{h,1} - \beta \Delta u_h - \operatorname{div} p_{h,2} \in L^2(\Omega)^m \\
&\zeta := \beta \nabla u_h + p_{h,2} \in L^2(\Omega)^{d\times m}
\end{aligned}$$

we obtain using interpolation estimates (5.8) and (5.9):

$$\begin{aligned}
&\tfrac{1}{2}\|u_h - u\|_B^2 \\
&\quad\le \sum_{K\in\mathcal{T}} \Big(\langle \xi, e_h - Ie_h\rangle_{L^2(K)^m} + \langle n^T\zeta, e_h - Ie_h\rangle_{L^2(\partial K)^m} \Big) \\
&\quad= \sum_{K\in\mathcal{T}} \langle \xi, e_h - Ie_h\rangle_{L^2(K)^m} + \sum_{F\in\Gamma} \langle [n^T\zeta], e_h - Ie_h\rangle_{L^2(F)^m} \\
&\quad\le \max\{c_1, c_2\} \Big(\sum_{K\in\mathcal{T}} \|\nabla e_h\|_{L^2(\omega_K)^{d\times m}} h_K \|\xi\|_{L^2(K)^m} \\
&\qquad\qquad\qquad + \sum_{F\in\Gamma} \|\nabla e_h\|_{L^2(\omega_F)^{d\times m}} h_F^{\frac{1}{2}} \|[n^T\zeta]\|_{L^2(F)^m} \Big) \\
&\quad\le \max\{c_1, c_2\} \Big(\sum_{K\in\mathcal{T}} \|\nabla e_h\|^2_{L^2(\omega_K)^{d\times m}} + \sum_{F\in\Gamma} \|\nabla e_h\|^2_{L^2(\omega_F)^{d\times m}} \Big)^{\frac{1}{2}} \\
&\qquad\qquad\qquad \Big(\sum_{K\in\mathcal{T}} h_K^2 \|\xi\|^2_{L^2(K)^m} + \sum_{F\in\Gamma} h_F \|[n^T\zeta]\|^2_{L^2(F)^m} \Big)^{\frac{1}{2}}.
\end{aligned}$$

Realizing that $\omega_F \subseteq \omega_K$ for some, to F adjacent cell K we may bound

$$\sum_{F\in\Gamma} \|\nabla e_h\|^2_{L^2(\omega_F)^{d\times m}} \le \tfrac{3}{2} \sum_{K\in\mathcal{T}} \|\nabla e_h\|^2_{L^2(\omega_K)^{d\times m}} \le \tfrac{3}{2} c_{\mathcal{T}}^2 \|\nabla e_h\|^2_{L^2(\Omega)^{d\times m}},$$

where $c_{\mathcal{T}}$ denotes the shape regularity constant of the mesh independent of the mesh size. We thus arrive at

$$
\begin{aligned}
\tfrac{1}{2}\|u_h - u\|_B^2 \le \max\{c_1, c_2\}\Big(\tfrac{5}{2}\sum_{K\in\mathcal{T}} \|\nabla e_h\|^2_{L^2(\omega_K)^{d\times m}}\Big)^{\frac{1}{2}}& \\
\cdot\Big(\sum_{K\in\mathcal{T}} h_K^2\|\xi\|^2_{L^2(K)^m} + \sum_{F\in\Gamma} h_F\|[n^T\zeta]\|^2_{L^2(F)^m}\Big)^{\frac{1}{2}}& \\
\le \tfrac{\sqrt{10}}{2}\max\{c_1, c_2\}c_{\mathcal{T}}\|\nabla e_h\|_{L^2(\Omega)^{d\times m}}& \\
\cdot\Big(\sum_{K\in\mathcal{T}} h_K^2\|\xi\|^2_{L^2(K)^m} + \sum_{F\in\Gamma} h_F\|[n^T\zeta]\|^2_{L^2(F)^m}\Big)^{\frac{1}{2}}& \\
\le \tfrac{\sqrt{10}}{2}\max\{c_1, c_2\}c_{\mathcal{T}}\|e_h\|_{H^1(\Omega)^m}& \\
\cdot\Big(\sum_{K\in\mathcal{T}} h_K^2\|\xi\|^2_{L^2(K)^m} + \sum_{F\in\Gamma} h_F\|[n^T\zeta]\|^2_{L^2(F)^m}\Big)^{\frac{1}{2}}&.
\end{aligned}
$$

Together with coercivity $\|v\|_B^2 = a_B(v, v) \ge c_B\|v\|_V^2 = c_B\|v\|^2_{H^1(\Omega)^m}$ from Assumption (A1) we arrive at an a-posteriori error bound

$$
\begin{aligned}
&\|u_h - u\|^2_{H^1(\Omega)^m} \\
&\quad\le C\Big(\sum_{K\in\mathcal{T}} h_K^2\|\alpha_2 T^*(Tu_h - g) + T^*p_{h,1} - \beta\Delta u_h - \operatorname{div} p_{h,2}\|^2_{L^2(K)^m} \\
&\qquad\qquad + \sum_{F\in\Gamma} h_F\|[n^T(\beta\nabla u_h + p_{h,2})]\|^2_{L^2(F)^m}\Big)
\end{aligned}
$$

with constant $C := \frac{10}{c_B}\max\{c_1^2, c_2^2\}c_{\mathcal{T}}^2$.

Setting (S.i): the case $\mathrm{S} = \mathrm{I}$

Using Setting (S.i) we set $V = L^2(\Omega)^m$ and will make use of the fact that a_B is coercive on V, see Assumption (A1) , to derive an estimate on the error $\|e_h\|_{L^2(\Omega)^m}$.

Similarly to above we obtain from (5.7) with $S = I$:

$$\begin{aligned}
\tfrac{1}{2}a_B(e_h, e_h) &\leq \langle \alpha_2 T^*(Tu_h - g) + T^* p_{h,1} + \beta u_h, e_h - I_h e_h \rangle \\
&\qquad\qquad\qquad + \langle p_{h,2}, \nabla(e_h - I_h e_h)\rangle \\
&\leq \sum_{K \in \mathcal{T}} \Bigg(\Big\langle \alpha_2 T^*(Tu_h - g) + T^* p_{h,1} + \beta u_h - \operatorname{div} p_{h,2}, \\
&\qquad\qquad e_h - I_h e_h \Big\rangle_{L^2(K)^m} \\
&\qquad\qquad + \Big\langle n^T p_{h,2}, e_h - I_h e_h \Big\rangle_{L^2(\partial K)^m} \Bigg).
\end{aligned}$$

Denoting $\xi := \alpha_2 T^*(Tu_h - g) + T^* p_{h,1} + \beta u_h - \operatorname{div} p_{h,2}$ and $\zeta := n^T p_{h,2}$ for brevity one continues to derive using (5.10):

$$\begin{aligned}
&\tfrac{1}{2}a_B(e_h, e_h) \\
&\quad \leq \sum_{K \in \mathcal{T}} \|\xi\|_{L^2(K)^m} \|e_h - I_h e_h\|_{L^2(K)^m} \\
&\qquad + \sum_{F \in \Gamma} \|[\zeta]\|_{L^2(F)^m} \|e_h - I_h e_h\|_{L^2(F)^m} \\
&\quad \leq (1 + \tfrac{3}{2}c_3)^{\frac{1}{2}} \Big(\sum_{K \in \mathcal{T}} \|e_h - I_h e_h\|^2_{L^2(\omega_K)^m} \Big)^{\frac{1}{2}} \\
&\qquad \cdot \Big(\sum_{K \in \mathcal{T}} \|\xi\|^2_{L^2(K)^m} + \sum_{F \in \Gamma} h_F^{-1} \|[\zeta]\|^2_{L^2(F)^m} \Big)^{\frac{1}{2}} \\
&\quad \leq (1 + \tfrac{3}{2}c_3)^{\frac{1}{2}} c_{\mathcal{T}}^2 \|e_h\|_{L^2(\Omega)^m} \\
&\qquad \Big(\sum_{K \in \mathcal{T}} \|\xi\|^2_{L^2(K)^m} + \sum_{F \in \Gamma} h_F^{-1} \|[\zeta]\|^2_{L^2(F)^m} \Big)^{\frac{1}{2}}
\end{aligned}$$

Together with L^2-coercivity of $a_B(\,\cdot\,,\,\cdot\,)$ from Assumption (A1) we arrive at

$$\begin{aligned}&\|u_h - u\|^2_{L^2(\Omega)^m} \\ &\quad\le C\biggl(\sum_{K\in\mathcal{T}} \|\alpha_2 T^*(Tu_h - g) + T^* p_{h,1} + \beta u_h - \operatorname{div} p_{h,2}\|^2_{L^2(K)^m} \\ &\qquad\qquad + \sum_{F\in\Gamma} h_F^{-1} \|[n^T p_{h,2}]\|^2_{L^2(F)^m}\biggr)\end{aligned}$$

for some constant $C > 0$.

Error Indicators

For a cell $K \in \mathcal{T}$ we define our local error indicators as follows

$$\eta^2_{2,K} := \tilde{\eta}^2_{2,K} + \sum_{\substack{F\in\Gamma\\ F\cap K\neq\emptyset}} \tilde{\eta}^2_{2,F}, \tag{5.11}$$

where $\tilde{\eta}_{2,K}$ and $\tilde{\eta}_{2,F}$ are chosen as follows. For Setting (S.ii) and $S = \nabla$ we set

$$\eta^2_{2,K} := h_K^2 \bigl\|\alpha_2 T^*(Tu_h - g) + T^* p_{h,1} - \beta \Delta u_h - \operatorname{div} \boldsymbol{p}_{h,2}\bigr\|^2_{L^2(K)^m}, \tag{5.12}$$

$$\eta^2_{2,F} := h_F \bigl\|[n^T(\beta \nabla u_h + \boldsymbol{p}_{h,2})]\bigr\|^2_{L^2(F)^m}, \tag{5.13}$$

where for piecewise linear u_h and piecewise constant $\boldsymbol{p}_{h,2}$ the terms Δu_h and $\operatorname{div} \boldsymbol{p}_{h,2}$ in (5.12) vanish. For Setting (S.i) and $S = I$ we set

$$\eta^2_{2,K} := \bigl\|\alpha_2 T^*(Tu_h - g) + T^* p_{h,1} + \beta u_h - \operatorname{div} \boldsymbol{p}_{h,2}\bigr\|^2_{L^2(K)^m}, \tag{5.14}$$

$$\eta^2_{2,F} := h_F^{-1} \bigl\|[n^T \boldsymbol{p}_{h,2}]\bigr\|^2_{L^2(F)^m}. \tag{5.15}$$

Note that the indicators are computable locally on each cell again because of Assumption (A2). For Setting (S.i) with $S = I$ the facet indicator $\eta_{2,F}$ scales inversely with the diameter and is therefore not very useful in the context of adaptive refinement. This is due to a_B only being L^2-coercive in that case, which limits our choice of interpolation estimates and it is unclear to us whether this result can be improved. Showing efficiency of these estimators may be considered in future work and is expected to work out in a similar way as in [61].

5.3 Classic Algorithms

We now review two classical first-order optimization algorithms from [28, 31] in the general Hilbert space setting and apply them to our model (4.1). In particular, this presentation makes the algorithms applicable for both the finite difference setting as well as the finite element one.

5.3.1 Semi-Implicit Dual Algorithm

The problem (4.1) may be solved using its (pre-)dual formulation (4.2). This is especially relevant in our domain decomposition setting of Chapter 4 since we have to solve subproblems (4.9) of the same general form and it is unclear how to decompose the primal variable.

One specific such algorithm is the semi-implicit Lagrange multiplier method due to Chambolle [28] which solves (4.2) for the special case $B = I$. While [28] uses finite differences, we present the algorithm in a Hilbert space setting and for general B.

As in Equation (4.2) let $K := \{\boldsymbol{p} \in W^* : |\boldsymbol{p}|_F \leq \lambda\}$ denote the set of feasible dual variables. Similar to [28] there exists a Lagrange multiplier $\mu \in L^2(\Omega)$ corresponding to the constraint in K, c.f. [63, Theorem 1.6], such that

$$0 = \Lambda B^{-1}(\Lambda^* p - T^* g) + \mu p \tag{5.16}$$

with $\mu \geq 0$, $|p|_F \leq 1$ and $\frac{\mu}{2}(|p|_F^2 - \lambda^2) = 0$. Here μp is to be understood as pointwise multiplication.

Let for now $\lambda > 0$. Observing that in a pointwise sense $\mu = 0$ implies $\xi := \Lambda B^{-1}(\Lambda^* p - T^* g) = 0$ and $\mu > 0$ implies $|p|_F = \lambda$ almost everywhere, we deduce from the above condition that in either case

$\mu = \frac{|\xi|}{\lambda}$. Thus (5.16) becomes

$$0 = \xi + \tfrac{|\xi|}{\lambda} p.$$

The semi-implicit iterative method then uses for some starting value $p^0 \in K$ and stepsize $\tau > 0$ iterates $(p^n)_{n\in\mathbb{N}_0} \subseteq W^*$ satisfying

$$p^{n+1} = p^n - \tau(\xi^n + \tfrac{|\xi^n|}{\lambda} p^{n+1}), \tag{5.17}$$

where $\xi^n := \Lambda B^{-1}(\Lambda^* p^n - T^* g)$, $n \in \mathbb{N}_0$. Solving (5.17) for p^{n+1} then yields the following algorithm.

Algorithm 5.9 (Semi-implicit dual multiplier method [28]).

```
***Initialize:*** $p^0 := 0 \in K$ *and* $\tau \in (0, \frac{1}{\|\Lambda B^{-1}\Lambda^*\|})$
for $n = 0, 1, 2, \dots$ do
    $\xi^n = \Lambda B^{-1}(\Lambda^* p^n - T^* g)$
    $p^{n+1} = \lambda \dfrac{p^n - \tau\xi^n}{\lambda + \tau|\xi^n|}$
end for
```

Before we prove convergence of Algorithm 5.9 in the upcoming Theorem 5.11, let us first make some comments on Algorithm 5.9.

Remark 5.10. *Regarding Algorithm 5.9, take note of the following:*

- *In the trivial case $\lambda = 0$ one sets $p^{n+1} = 0$.*

- *If B^{-1} is a local operator, the computation of ξ^n and p^{n+1} are both local. They can therefore be merged together and carried out in parallel over the whole domain.*

- *One may solve the domain decomposition subproblems (4.9) by replacing K with $\theta_i K$ and λ with the pointwise function $\theta_i \lambda$.*

- *A more explicit bound for the maximum stepsize τ to still analytically ensure convergence is given by*

$$\|\Lambda B^{-1}\Lambda^*\| \leq \max\{\|T\|^2, \|\nabla\|^2\}\|B^{-1}\|.$$

In the finite element setting we have due to Lemma 5.3:

$$\|\nabla\|^2 \leq \tfrac{72}{\varrho_{\min}^2},$$

where $\varrho_{\min} = \min_{k \in \mathcal{T}} \varrho_K$ *denotes the diameter of the largest ball that fits into every cell. For finite differences on the other hand we have [28]*

$$\|\nabla\|^2 \leq 8.$$

Theorem 5.11 (c.f. [28, Theorem 3.1])**.** *Let* $(p^n)_{n \in \mathbb{N}_0} \subseteq W^*$ *be the iterates of Algorithm 5.9 and let* $\hat{p} \in K$ *denote a minimizer of* (4.2)*. Then the algorithm converges in the sense that* $D(p^n) \to D(\hat{p})$ *if* W^* *is finite dimensional.*

Proof. We follow along the lines of [28, Theorem 3.1]. Notice that $|p^0|_F \leq \lambda$ and thus inductively

$$|p^{n+1}|_F \leq \lambda \tfrac{|p^n|_F + \tau |\xi^n|_F}{\lambda + \tau |\xi^n|_F} \leq \lambda,$$

i.e. $p^n \in K$ for all $n \in \mathbb{N}$. Let $F : W^* \to W^*$ denote the iteration function of Algorithm 5.9, such that $p^{n+1} = F(p^n)$, $n \in \mathbb{N}_0$. Any fixed point of F or equivalently of (5.17) satisfies the stationary point condition (5.16) per construction and, since D is convex, will be a minimizer of (4.2).

Denote $\eta^n := \frac{1}{\tau}(p^n - p^{n+1})$ and bound the energy difference

$$\begin{aligned}
&D(p^n) - D(p^{n+1}) \\
&\quad = -\tfrac{1}{2}\|p^n - p^{n+1}\|_*^2 + \langle D'(p^n), p^n - p^{n+1}\rangle \qquad \text{(Lemma 4.8 (i))} \\
&\quad = -\tfrac{\tau^2}{2}\|\eta^n\|_*^2 + \tau\langle \xi^n, \eta^n\rangle \\
&\quad = \tfrac{\tau}{2}(\|\eta^n\|^2 - \tau\|\eta^n\|_*^2) + \tau\langle \xi^n - \tfrac{1}{2}\eta^n, \eta^n\rangle \\
&\quad = \tfrac{\tau}{2}(\|\eta^n\|^2 - \tau\|\eta^n\|_*^2) + \tfrac{\tau}{2}\langle \xi^n - \tfrac{|\xi^n|_F}{\lambda}p^{n+1}, \xi^n + \tfrac{|\xi^n|_F}{\lambda}p^{n+1}\rangle \\
&\qquad \text{(applying (5.17))} \\
&\quad = \tfrac{\tau}{2}(\|\eta^n\|^2 - \tau\langle \Lambda B \Lambda^* \eta^n, \eta^n\rangle) + \tfrac{\tau}{2}(\|\xi^n\|^2 - \|\tfrac{|\xi^n|_F}{\lambda}p^{n+1}\|^2) \\
&\quad \geq \tfrac{\tau}{2}\big(1 - \tau\big\|\Lambda B^{-1}\Lambda^*\big\|\big)\|\eta^n\|^2 + \tfrac{\tau}{2}\|\xi^n\|^2\big(1 - \big\|\tfrac{|p^{n+1}|_F^2}{\lambda^2}\big\|_{L^\infty}\big) \\
&\quad \geq \tfrac{1}{2\tau}(1 - \tau\|\Lambda B^{-1}\Lambda^*\|)\|p^n - p^{n+1}\|^2.
\end{aligned}$$

We see that as long as $\tau < \|\Lambda B^{-1}\Lambda^*\|^{-1}$, the sequence $(D(p^n))_{n\in\mathbb{N}_0}$ is non-increasing and thus, since it is non-negative, also convergent. The feasible set $K \subseteq W^*$ is closed and bounded, see Lemma 2.5, and compact since W^* is finite dimensional. Consequently there exists a convergent subsequence $(q^n)_{n\in\mathbb{N}} \subseteq (p^n)_{n\in\mathbb{N}} \subseteq K$, $q^n \to q \in K$ and with continuity of F we get $F(q^n) \to F(q)$. Using the estimate above and the convergence of energies we see that for some $c > 0$ we have $c\|q^n - F(q^n)\|^2 \leq D(q^n) - D(F(q^n)) \to 0$ and therefore the limit needs to be a fixed point, $q = F(q)$, and thus a minimizer of (4.2). □

We note that Theorem 5.11 guarantees the convergence to a minimal dual energy, which allows us to reconstruct the optimal primal solution due to Proposition 4.2. It does, however, not guarantee convergence of the dual iterates $(p^n)_{n\in\mathbb{N}}$ themselves.

5.3.2 Semi-Implicit Primal-Dual Algorithm

In [31] the authors derive an accelerated semi-implicit primal-dual algorithm to minimize a general functional of the form

$$\inf_{u\in X} \mathcal{F}(u) + \mathcal{G}(\Lambda u), \tag{5.18}$$

where X, Y are Hilbert spaces, $\Lambda : X \to Y$ is a linear bounded operator, $\mathcal{F} : X \to \overline{\mathbb{R}}$, $\mathcal{G} : Y \to \overline{\mathbb{R}}$ are proper, convex, lower semi-continuous functionals with $\mathcal{F}(u_0) + \mathcal{G}(\Lambda u_0) < \infty$ and $\mathcal{G}$ continuous at Λu_0 for some $u_0 \in X$. Note that this is a special case of the Fenchel duality setting from Theorem 2.28.

Let $\hat{\mu} \geq 0$ denote a constant such that for all $w^* \in \partial\mathcal{F}(w)$:

$$\mathcal{F}(v) - \mathcal{F}(w) \geq \langle w^*, v - w\rangle + \tfrac{\hat{\mu}}{2}\|v - w\|^2. \tag{5.19}$$

Since $\mathcal{F}$ is convex, the choice $\hat{\mu} = 0$ will always satisfy (5.19) while for strongly convex $\mathcal{F}$ one can find $\hat{\mu} > 0$. Then [31] states the following algorithm.

Algorithm 5.12 (Semi-implicit primal-dual algorithm [31, ALG2]).

```
Parameters: $\tau_0 > 0, \sigma_0 := \frac{1}{\tau_0 L^2} > 0$, $\theta \in [0,1]$, $0 \le \mu \le \hat{\mu}$, $\|\Lambda\| \le L < \infty$
Initialization: $u^0 \in X$, $\overline{u}^0 = u^0$, $(p_1^0, p_2^0) \in Y$
for $n = 0, 1, 2, \dots$ do
    $p^{n+1} = (I + \sigma_n \partial G^*)^{-1}(p^n + \sigma_n \Lambda \overline{u}^n)$
    $u^{n+1} = (I + \tau_n \partial F)^{-1}(u^n - \tau_n \Lambda^* p^{n+1})$
    $\theta_n = (1 + 2\mu\tau_n)^{-\frac{1}{2}}$
    $\tau_{n+1} = \theta_n \tau_n$
    $\sigma_{n+1} = \theta_n^{-1} \sigma_n$
    $\overline{u}^{n+1} = u^{n+1} + \theta_n (u^{n+1} - u^n)$
end for
```

We note that a practical bound for $\|\Lambda\|$, in order to select L as small as possible may be obtained through Lemma 5.3. The non-accelerated variant of Algorithm 5.12, i.e. [31, ALG2], may be obtained by leaving σ, τ, θ constant throughout the algorithm.

Theorem 5.13 ([31, Theorem 2]). *In the setting of Algorithm 5.12 if $\mu > 0$ then for any $\varepsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that for any $n \in \mathbb{N}$, $n \ge n_0$ one has the following a-priori error estimate:*

$$\begin{aligned} &\|\hat{u} - u^n\|^2 + \tfrac{1+\varepsilon}{n^2} \tfrac{1 - L^2 \sigma_0 \tau_0}{\sigma_0 \tau_0} \|\hat{p} - p^n\|^2 \\ &\quad \le \tfrac{1+\varepsilon}{n^2} \big(\tfrac{1}{\mu^2 \tau_0^2} \|\hat{u} - u^0\|^2 + \tfrac{L^2}{\mu^2} \|\hat{p} - p^0\|^2 \big). \end{aligned}$$

Proof. This is exactly [31, Theorem 2] but applied to [31, eq. (42)] and thereby relaxing $L \ge \|\Lambda\|$. Also the following notational identifications were made: $F := \mathcal{G}$, $G := \mathcal{F}$, $K := \Lambda$, $\gamma := \mu$. □

We apply Algorithm 5.12 to our setting from Chapter 3, using

$$\begin{aligned} \mathcal{F}(u) &= \tfrac{\alpha_2}{2} \|Tu - g\|_{L^2}^2 + \tfrac{\beta}{2} \|Su\|_{L^2}^2, \\ \mathcal{G}^*(p) &= \langle g, p_1 \rangle + \chi_{|p_1| \le \alpha_1} + \chi_{|\boldsymbol{p}_2|_F \le \lambda} + \tfrac{\gamma_1}{2\alpha_1} \|p_1\|_{L^2}^2 + \tfrac{\gamma_2}{2\lambda} \|\boldsymbol{p}_2\|_{L^2}^2, \end{aligned}$$

thus yielding the updates

$$
\begin{aligned}
p_1^{n+1} &= \operatorname{proj}_{|\cdot|\le\alpha_1}\Big((1+\tfrac{\gamma_1\sigma_n}{\alpha_1})^{-1}(p_1^n+\sigma_n(T\overline{u}^n-g))\Big)\\
p_2^{n+1} &= \operatorname{proj}_{|\cdot|_F\le\lambda}\Big((1+\tfrac{\gamma_2\sigma_n}{\lambda})^{-1}(p_2^n+\sigma_n\nabla\overline{u}^n)\Big)\\
u^{n+1} &= (I+\tau_n B)^{-1}\Big(u^n-\tau_n(T^*p_1^n+\nabla^*p_2^n-\alpha_2T^*g)\Big)
\end{aligned}
$$

Note that if $\alpha_1 = 0$ or $\lambda = 0$ one implicitly has $p_1 = 0$ or $p_2 = 0$ respectively.

Algorithm 5.14 (Semi-implicit primal-dual algorithm [31, ALG2]).

```
***Parameters:*** $\tau_0 > 0, \sigma_0 := \frac{1}{\tau_0 L^2} > 0,\ \theta \in [0,1],\ 0 \le \mu \le \|B\|, \|\Lambda\| \le L < \infty$
***Initialization:*** $u^0 \in V,\ \overline{u}^0 = u^0,\ (p_1^0, p_2^0) \in W^*$
for $n = 0, 1, 2, \ldots$ do
    $p_1^{n+1} = \operatorname{proj}_{|\cdot|\le\alpha_1}\Big((1+\tfrac{\gamma_1\sigma_n}{\alpha_1})^{-1}(p_1^n+\sigma_n(T\overline{u}^n-g))\Big)$
    $p_2^{n+1} = \operatorname{proj}_{|\cdot|_F\le\lambda}\Big((1+\tfrac{\gamma_2\sigma_n}{\lambda})^{-1}(p_2^n+\sigma_n\nabla\overline{u}^n)\Big)$
    $u^{n+1} = (I+\tau_n B)^{-1}\Big(u^n-\tau_n(T^*p_1^n+\nabla^*p_2^n-\alpha_2T^*g)\Big)$
    $\theta_n = (1+2\mu\tau_n)^{-\frac{1}{2}}$
    $\tau_{n+1} = \theta_n\tau_n$
    $\sigma_{n+1} = \theta_n^{-1}\sigma_n$
    $\overline{u}^{n+1} = u^{n+1} + \theta_n(u^{n+1}-u^n)$
end for
```

Again the non-accelerated variant of Algorithm 5.14 may be obtained by leaving σ, τ, θ constant throughout the algorithm.

Note that with the exception of u^{n+1} in Algorithm 5.14 all steps can be performed locally as a simple update, whereas for u^{n+1} in general the solution of the variational equality

$$
\begin{aligned}
\langle u^{n+1}, v\rangle + \tau_n\big(\alpha_2\langle Tu^{n+1}, Tv\rangle + \beta\langle Su^{n+1}, Sv\rangle\big)&\\
= \langle u^n, v\rangle - \tau_n\big(\langle p_1^n - \alpha_2 g, Tv\rangle + \langle p_2^n, \nabla v\rangle\big)&
\end{aligned}
\tag{5.20}
$$

for all $v \in H^1(\Omega)^m$ is required.

In practice Algorithm 5.12 exhibits sublinear convergence [31] in agreement with Theorem 5.13 and therefore a high number of iterations for a sufficient approximation is to be expected. Computing a costly solution to (5.20) in each step is therefore undesirable in practice. In the case $T = I$, $S = I$, $\lambda = 0$, however, (5.20) simplifies to

$$u^{n+1} = (1 + \tau_n(\alpha_2 + \beta))^{-1}(u^n - \tau_n(p_1^n - \alpha_2 g)),$$

which can indeed be realized as a simple local update.

5.4 Primal-Dual Semi-Smooth Newton Algorithm

In this section we derive a primal-dual semi-smooth Newton method, cf. [57], in order to find an approximate solution of (3.13). Note that such Newton methods have been already used for the L^2-TV model [41, 61, 62, 67] and L^1-TV model [40, 65] in image reconstruction, i.e., $m = 1$. We extend the approach of semi-smooth Newton methods to a vector-valued setting and to the L^1-L^2-TV model.

Due to the dualization in vector-valued spaces results for the scalar case derived in [58] and [60] will be adjusted to our setting.

5.4.1 Derivation

In general (3.14) has a solution $\hat{u} \in V$ which can be approximated using continuous piecewise linear finite elements [12, Chapter 10.2]. Since all such discrete functions $u_h \in V_h$ are elements of $H^1(\Omega)^m$, we derive the semi-smooth Newton system using the spaces $H^1(\Omega)^m$ and $L^2(\Omega) \times L^2(\Omega)^{d\times m}$ for the primal and predual variable respectively. This simplification is sufficient for our discrete setting in any case, and sufficient for the continuous setting as long as $V = H^1(\Omega)^m$.

Let us denote for convenience

$$m_1 := m_1(u) := \max\{\gamma_1, |Tu-g|\}, \quad \chi_1 := \begin{cases} 1 & \text{if } |Tu-g| \geq \gamma_1 \\ 0 & \text{else} \end{cases},$$

$$m_2 := m_2(u) := \max\{\gamma_2, |\nabla u|_F\}, \qquad \chi_2 := \begin{cases} 1 & \text{if } |\nabla u|_F \geq \gamma_2 \\ 0 & \text{else} \end{cases}.$$

The Newton system of (3.16) in the unknowns $d_u \in H^1(\Omega)$, $d_{p_1} \in L^2(\Omega)$, $d_{\boldsymbol{p}_2} \in L^2(\Omega)^{d\times m}$ then reads

$$\begin{aligned} &\alpha_2 T^*Td_u + \beta S^*Sd_u + T^*d_{p_1} + \nabla^* d_{\boldsymbol{p}_2} \\ &\qquad = -\Big(\nabla^*(\boldsymbol{p}_2) + T^*p_1 + \alpha_2 T^*(Tu-g) + \beta S^*Su\Big), \end{aligned} \tag{5.21}$$

$$\begin{aligned} &\chi_1 \frac{(Tu-g)\cdot Td_u}{|Tu-g|} p_1 - \alpha_1 Td_u + m_1 d_{p_1} \\ &\qquad = -\Big(m_1 p_1 - \alpha_1(Tu-g)\Big), \end{aligned} \tag{5.22}$$

$$\begin{aligned} &\chi_2 \frac{\nabla u \cdot \nabla d_u}{|\nabla u|} \boldsymbol{p}_2 - \lambda \nabla d_u + m_2 d_{\boldsymbol{p}_2} \\ &\qquad = -\Big(m_2 \boldsymbol{p}_2 - \lambda \nabla u\Big), \end{aligned} \tag{5.23}$$

where $u \in H^1(\Omega)^m$, $p_1 \in L^2(\Omega)$, $\boldsymbol{p}_2 \in L^2(\Omega)^{d\times m}$ represent the variables from the previous Newton step.

Rearranging (5.22) and (5.23) for d_{p_1} and $d_{\boldsymbol{p}_2}$ yields

$$d_{p_1} = -p_1 + \frac{\alpha_1}{m_1}(T(u+d_u)-g) - \chi_1 \frac{(Tu-g)\cdot Td_u}{|Tu-g|^2} p_1, \tag{5.24}$$

$$d_{\boldsymbol{p}_2} = -\boldsymbol{p}_2 + \frac{\lambda}{m_2}\nabla(u+d_u) - \chi_2 \frac{\nabla u \cdot \nabla d_u}{|\nabla u|^2} \boldsymbol{p}_2. \tag{5.25}$$

Plugging these two equations into (5.21) leads to

$$\begin{aligned} 0 = T^*\Big(&\frac{\alpha_1}{m_1}\big(T(u+d_u)-g\big) - \chi_1 \frac{(Tu-g)\cdot Td_u}{|Tu-g|^2} p_1\Big) \\ &+ \nabla^*\Big(\frac{\lambda}{m_2}\nabla(u+d_u) - \chi_2 \frac{\nabla u \cdot \nabla d_u}{|\nabla u|^2}\boldsymbol{p}_2\Big) \\ &+ \alpha_2 T^*\big(T(u+d_u)-g\big) + \beta S^*S(u+d_u), \end{aligned}$$

which is to be understood in a weak sense.

Recall $a_B : H^1(\Omega)^m \times H^1(\Omega)^m \to \mathbb{R}$ from (3.6) and define $a_1, a_2 : H^1(\Omega)^m \times H^1(\Omega)^m \to \mathbb{R}$ and $l : H^1(\Omega)^m \to \mathbb{R}$ as follows

$$\begin{aligned} a_B(d_u, \varphi) &= \alpha_2 \langle T d_u, T\varphi\rangle + \beta \langle S d_u, S\varphi\rangle \\ a_1(d_u, \varphi) &:= \Big\langle \tfrac{\alpha_1}{m_1} T d_u - \tfrac{\chi_1}{m_1^2}(Tu - g)(T d_u) p_1, T\varphi \Big\rangle, \\ a_2(d_u, \varphi) &:= \Big\langle \tfrac{\lambda}{m_2} \nabla d_u - \tfrac{\chi_2}{m_2^2}(\nabla u \cdot \nabla d_u) \boldsymbol{p}_2, \nabla\varphi \Big\rangle \\ l(\varphi) &:= -a_B(u, \varphi) - \langle \tfrac{\lambda}{m_2} \nabla u, \nabla\varphi\rangle \\ &\qquad - \langle \tfrac{\alpha_1}{m_1}(Tu - g), T\varphi\rangle + \langle \alpha_2 g, T\varphi\rangle. \end{aligned}$$

We then have the following result.

Theorem 5.15. *Let $H \subseteq H^1(\Omega)^m$ be a subspace such that there exists $c_S > 0$ with $\|\nabla u\|_{L^2} \le c_S \|Su\|_{L^2}$ for all $u \in H$. If $p_1 \in L^2(\Omega)$, $\boldsymbol{p}_2 \in L^2(\Omega)^{d\times m}$ such that $|p_1| \le \alpha_1$, $|\boldsymbol{p}_2|_F \le \lambda$ holds, then the problem*

$$a(d_u, \varphi) := a_1(d_u, \varphi) + a_2(d_u, \varphi) + a_B(d_u, \varphi) = l(\varphi), \quad \forall \varphi \in H \quad (5.26)$$

admits a unique solution $d_u \in H$.

Proof. We verify the prerequisites for the Lax-Milgram Theorem 2.10, i.e. boundedness of a and l, as well as coercivity of a.

We verify boundedness of l

$$\begin{aligned} |l(\varphi)| &\le \|B\| \|u\|_{L^2} \|\varphi\|_{L^2} + \lambda |\Omega| \|\nabla\varphi\|_{L^2} \\ &\qquad + \alpha_1 |\Omega| \|T\varphi\|_{L^2} + \alpha_2 \|g\|_{L^2} \|T\varphi\|_{L^2} \\ &\le c \|\varphi\|_{H^1(\Omega)^m} \end{aligned}$$

for some constant $c > 0$, since T and Ω are bounded and $u, g \in L^2(\Omega)$.

Boundedness of a_1, a_2 follows from

$$\begin{aligned}
|a_1(v,w)| &\le \Big(\|\tfrac{\alpha_1}{m_1}Tv\|_{L^2} + \|\tfrac{\chi_1}{m_1^2}(Tu-g)(Tv)p_1\|\Big)\|Tw\|_{L^2} \\
&\le \tfrac{2\alpha_1}{\gamma_1}\|T\|^2\|v\|_{L^2}\|w\|_{L^2}, \\
|a_2(v,w)| &\le \Big(\|\tfrac{\lambda}{m_2}\nabla v\|_{L^2} + \|\tfrac{\chi_1}{m_2^2}(\nabla u\cdot\nabla v)\boldsymbol{p}_2\|\Big)\|\nabla w\|_{L^2} \\
&\le \tfrac{2\lambda}{\gamma_2}\|\nabla v\|_{L^2}\|\nabla w\|_{L^2}.
\end{aligned}$$

This implies together with boundedness of the bilinear form a_B that $|a(v,w)| \le c\|v\|_{H^1(\Omega)}\|w\|_{H^1(\Omega)}$ for some constant $c > 0$.

Since coercivity of $a_B(v,v)$ follows from Assumption (A1), it is sufficient to show that a_1 and a_2 are positive semi-definite. Using the vectorization operator $\mathrm{vec} : \mathbb{R}^{d\times m} \to \mathbb{R}^{dm} : X \mapsto (X_{(k-1 \bmod d)+1, \lfloor \frac{k-1}{d}\rfloor+1})_{k=1}^{k=dm}$ applied in a pointwise sense for convenience we see that

$$\begin{aligned}
a_1(v,w) &= \Big\langle \big(\tfrac{\alpha_1}{m_1} - \chi_1\tfrac{p_1(Tu-g)}{m_1^2}\big)Tv, Tw\Big\rangle =: \langle A_1 Tv, Tw\rangle, \\
a_2(v,w) &= \Big\langle \big(\tfrac{\lambda}{m_2}I_{dm\times dm} - \chi_2\tfrac{\mathrm{vec}(\boldsymbol{p}_2)\,\mathrm{vec}(\nabla u)^T}{m_2^2}\big)\,\mathrm{vec}(\nabla v), \mathrm{vec}(\nabla w)\Big\rangle \\
&=: \langle A_2\nabla v, \nabla w\rangle,
\end{aligned}$$

where $I_{dm\times dm} \in \mathbb{R}^{dm\times dm}$ denotes the unit matrix. It thus remains to show pointwise positive semi-definiteness for $A_1 : \Omega \to \mathbb{R}$ and $A_2 : \Omega \to \mathbb{R}^{dm\times dm}$. We see this by evaluating for $x \in \mathbb{R}^{dm}$:

$$\begin{aligned}
A_1 &\ge \tfrac{\alpha_1}{m_1} - \chi_1\tfrac{|p_1|}{m_1}\tfrac{|Tu-g|}{m_1} \ge \tfrac{\alpha_1}{m_1} - \chi_1\tfrac{\alpha_1}{m_1} \ge 0, \\
x^T A_2 x &\ge \tfrac{\lambda}{m_2}|x|^2 - \chi_2\tfrac{|\mathrm{vec}(\boldsymbol{p}_2)|}{m_2}\tfrac{|\mathrm{vec}(\nabla u)|}{m_2}|x|^2 \ge \big(\tfrac{\lambda}{m_2} - \chi_2\tfrac{\lambda}{m_2}\big)|x|^2 \ge 0.
\end{aligned}$$

This concludes the coercivity of the sum $a = a_1 + a_2 + a_B$ and applying the Lax-Milgram theorem yields the required result. □

Corollary 5.16. *Assume $|p_1| \leq \alpha_1$, $|\boldsymbol{p}_2|_F \leq \lambda$ holds. Then the discrete problem of finding $d_u \in V_h$ such that $a(d_u, \varphi) = l(\varphi)$ for all $\varphi \in V_h$ admits a unique solution.*

Proof. If $S = \nabla$, the statement follows immediately from Theorem 5.15 using $c_S = 1$. Let $S = I$, then the finite element inverse inequality (see e.g. [34, Theorem 3.2.6] or [3, Theorem 1.3]) yields

$$\|\nabla u\|_{L^2} \leq ch^{-1}\|u\|_{L^2} = ch^{-1}\|Su\|_{L^2},$$

where h is the smallest cell diameter and c is a constant independent of h. Then Theorem 5.15 with $c_S = ch^{-1}$ again yields the required result. □

Theorem 5.15 and Corollary 5.16 prove the solvability of the semi-smooth Newton step and thus ensure that the following semi-smooth Newton algorithm is well-defined in a general Hilbert space setting.

Let $H \subseteq H^1(\Omega)^m$ satisfy the requirements from Theorem 5.15. The semi-smooth Newton scheme for (3.13) is given by the following algorithm.

Algorithm 5.17 (Semi-smooth Newton)**.**

```
Parameters: model parameters $\alpha_1, \alpha_2, \lambda, \beta$, regularization parameters $\gamma_1, \gamma_2 > 0$
Input: data $g \in L^2(\Omega)$, initial guesses $u_0 \in H^1(\Omega)^m$, $\boldsymbol{p}_0 \in L^2(\Omega) \times L^2(\Omega)^{d\times m}$
Output: sequence $(u_k, \boldsymbol{p}_k)$ approximating the solution to (3.16)
for $k = 1, 2, \dots$ do
    solve $a(d_u, \varphi) = l(\varphi)$, $\varphi \in H$ from Theorem 5.15
    assign $d_{p_1}, d_{\boldsymbol{p}_2}$ according to (5.24) and (5.25)
    $u_k = u_{k-1} + d_u$
    $\boldsymbol{p}_k = \boldsymbol{p}_{k-1} + (d_{p_1}, d_{\boldsymbol{p}_2})$
end for
```

If not otherwise specified, we use for Algorithm 5.17 the Cauchy stopping criterion

$$\tfrac{1}{|\Omega|}\left(\|u_n - u_{n-1}\|^2 + \|p_{1,n} - p_{1,n-1}\|^2 + \|\boldsymbol{p}_{2,n} - \boldsymbol{p}_{2,n-1}\|^2\right) < \varepsilon_{\text{newton}},$$

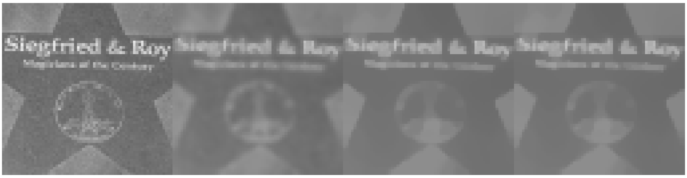


Figure 5.4: from left to right: 64x64 pixel input image g and respective denoised outputs for semi-implicit, semi-implicit accelerated and semi-smooth Newton algorithms

for some specified constant $\varepsilon_{\text{newton}}$.

5.4.2 Numerical Behaviour

Convergence Rate

To numerically observe the asymptotic convergence properties of Algorithm 5.17, a small image g as depicted in Figure 5.4 has been chosen along with the denoising setting $T = I$, $S = I$, $\alpha_1 = 0$, $\alpha_2 = 30$, $\lambda = 1$, $\beta = 0$ and $\gamma_1 = 1 \cdot 10^{-2}, \gamma_2 = 1 \cdot 10^{-3}$. We iterate until $|\Omega|^{-\frac{1}{2}} \|u^k - u^{k-1}\|_{L^2} < 1 \cdot 10^{-10}$ or $k \geq 10{,}000$. The energy $\hat{E} := 112.47$ was obtained as the minimal energy over all iterations and algorithms and assumed by Algorithm 5.17.

From the step lengths and energies in Figure 5.5 one can see the sublinear convergence of the semi-implicit method and its accelerated variant. The semi-smooth Newton method displays superlinear convergence, reaches the desired tolerance after only a few iterations and assumes the minimal energy $\hat{E}$ in the last step which is excluded in the logarithmic plot.

5.5 Applications

In the following we aim show that our model together with Algorithm 5.17 can indeed be applied in practice to solve the image processing tasks introduced in Section 3.1.1.

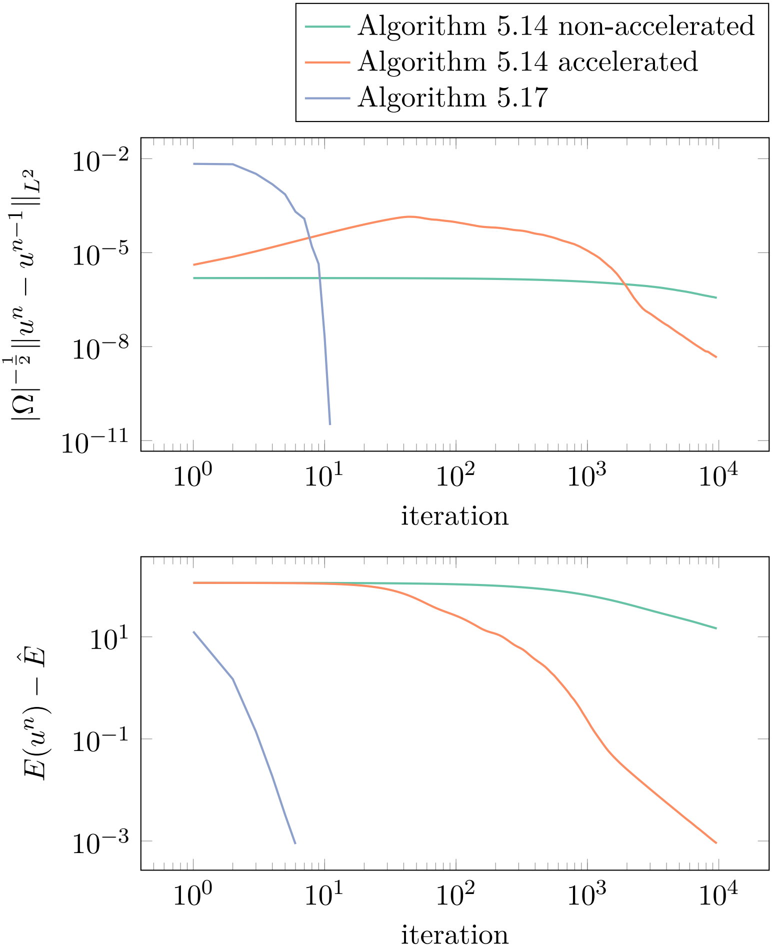


Figure 5.5: comparison of steps and energy

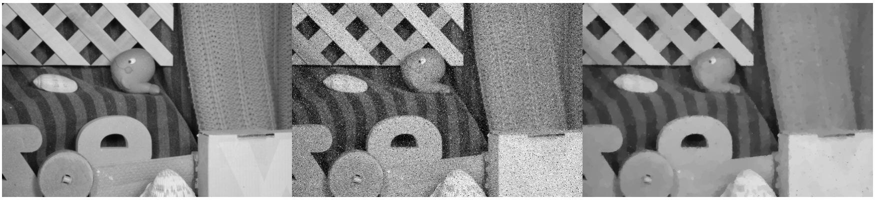


Figure 5.6: from left to right: original image $\tilde{g}$, noisy input image g, denoised output image u

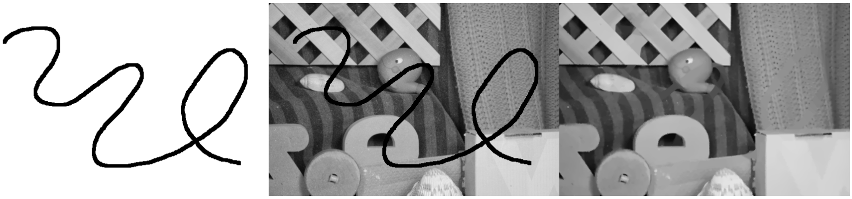


Figure 5.7: from left to right: inpainting mask, masked input image, inpainted output image

5.5.1 Denoising

From an original image $\tilde{g}$ we generate an artificial noisy input $g := \varphi(\tilde{g} + \eta)$, where η denotes zero mean additive Gaussian noise with variance 0.1 and $\varphi(x) \in \{0, 1, x\}$ with probability $\frac{p}{2}, \frac{p}{2}, 1-p$ respectively and $p = 2 \cdot 10^{-2}$.

We denoise the image g in Figure 5.6 by setting $T = I$ as described in Section 3.1.1, $S = I$ and using manually chosen parameters $\alpha_1 = 0.2$, $\alpha_2 = 8$, $\lambda = 1$, $\beta = 0$, $\gamma_1 = 1 \cdot 10^{-4}$, $\gamma_2 = 1 \cdot 10^{-4}$ to obtain visually pleasing results.

The result visible in Figure 5.6 matches the expected behaviour of total variation denoising, i.e. coherent noisy regions are flattened out, while hard edges are mostly preserved.

5.5.2 Inpainting

For inpainting we aim to use the setting from Section 3.1.1 by use of the masking operator from (3.2) for T. The inpainting domain D is specified by an inpainting mask as shown in Figure 5.7.

Special care has to be taken, however, in our finite element setting. Image interpolation may leak corrupt data from within the inpainting area and the inpainting mask needs to be extended to cover this area. In particular, global interpolation methods, such as L^2-projection in the case of cellwise linear continuous elements, should be avoided and for other interpolation methods, the inpainting mask needs to be extended to cover the area of influence. We do this by choosing the interpolation method `qi_lagrange` for g, interpolating the image mask using the same method and defining the operator T pointwise to be zero whenever the image mask is not one.

The parameters are manually chosen by visual preference as follows: $\alpha_1 = 0$, $\alpha_2 = 50$, $\lambda = 1$, $\beta = 1 \cdot 10^{-5}$, $\gamma_1 = 1 \cdot 10^{-4}$, $\gamma_2 = 1 \cdot 10^{-4}$. The result can be seen in Figure 5.7, where the lost information within the masked region has been filled in.

5.5.3 Optical Flow

The application to motion estimation was discussed in Section 3.1.1. We now additionally introduce a combined warping and adaptation algorithm.

While the linearized optical flow equation (3.4) has localized the global condition (3.3) it comes at the cost of misrepresenting large displacements. One may alleviate this problem by repositioning the linearization point as in Algorithm 5.18.

Algorithm 5.18 (Warping technique for optical flow).

```
Input: images $f_0, f_1$, initial guess $u_0$
Output: motion fields $(u_k)$
for $k = 1, 2, \dots$ do
    $f_{w,k-1}(x) = f_1(x + u_{k-1}(x))$, $x \in \Omega$
    $0 = \nabla f_{w,k-1} \cdot u_k - \nabla f_{w,k-1} \cdot u_{k-1} + f_w - f_0$ for $u_k$
end for
```

A few remarks on Algorithm 5.18 are in order.

- The images f_0 and f_1 are assumed to be smooth, since otherwise a measure-theoretic definition of $\nabla f_{w,k}$ is necessary and the existence of solutions seems unclear in that case [9].

- Stopping after the first iteration corresponds to solving the linearized optical flow equation $0 = \nabla f_1 \cdot u + f_1 - f_0$.

- The algorithm has resemblence to an underdetermined Newton algorithm. Indeed, when solving for the minimum norm solution in each step the algorithm relates to the *normal flow algorithm* [69] in a pointwise sense.

- The warping technique may be combined with a coarse-to-fine scheme, where u_k is solved on increasingly finer scales, resolving large displacements on an early coarse scale and filling in detail later.

To approximately solve for u_k in Algorithm 5.18, we may use our model (3.13) by choosing $Tu := \nabla f_{w,k-1} \cdot u$, $g := \nabla f_{w,k-1} \cdot u_{k-1} - (f_{w,k-1} - f_0)$.

The images f_0, f_1, $f_{w,k}$, except for the warping step detailed below, use the discrete space Z_h, whereas for g we instead use a cellwise linear discontinuous space to capture the discontinuous component $\nabla f_{w,k-1}$.

Algorithm 5.19 (Optical flow algorithm with adaptive warping).

```
Parameters: warping threshold $\varepsilon_{warp}$ and parameters for Algorithm 5.17
Input: images $f_0, f_1$, initial guess $u_0$
Output: motion fields $(u_k)$
$f_{w,0}(x) = f_1(x + u_0(x))$
for $k = 1, 2, \dots$ do
    find approximate discrete solution $u_k$ to (3.13) using Algorithm 5.17
    $f_{w,k}(x) = f_1(x + u_k(x))$
    if $\frac{\|f_{w,k}-f_0\|_{L^2} - \|f_{w,k-1}-f_0\|_{L^2}}{\|f_{w,k-1}-f_0\|_{L^2}} > -\varepsilon_{warp}$ then
        refine mesh and reproject image data
    end if
end for
```

In Algorithm 5.19 we combine the warping technique from Algorithm 5.18 with adaptive refinement, starting from a coarse grid. Loosely speaking we solve the linearized optical flow equation for u_k and warp the input data by the computed flow field until we no longer improve on the data difference $f_{w,k} - f_0$, which indicates displacement. In that case, the mesh is refined using the indicators from (5.11) and the process repeats, now including more detailed image data.

The warping step $f_{w,k}(x) = f_1(x + u_k(x))$ is carried out at original image resolution by evaluating f_1 using bicubic interpolation and in a second step projected onto the current finite element space in order to capture more detailed displacement information.

We note, that this approach to adaptivity allows us to start off with a coarse mesh and refine cells only if deemed necessary by the error indicator. In that respect it is different from the only other adaptive finite element methods for optical flow we are aware of, see [15, 16], where the mesh is initialized at fine image resolution first and iteratively coarsened only after a costly computation of the flow field and a suitable metric for adaptivity on this fine mesh has been established.

Experiments

For all benchmarks the same manually chosen model parameters were used. We use $\alpha_1 = 10$, $\alpha_2 = 0$, $\lambda = 1$ to obtain visually pleasing results, cf. superiority of L1-TV in [39], $\beta = 1 \cdot 10^{-5}$, $\gamma_1 = 1 \cdot 10^{-4}$, $\gamma_2 = 1 \cdot 10^{-4}$ to balance between speed and quality of the reconstruction and $u_0 := 0$. For the interior method $\varepsilon_{\text{newton}} = 1 \cdot 10^{-3}$ was chosen. The mesh is initialized to $\frac{1}{6}$ of the image resolution, rounded down to integer values, $\varepsilon_{\text{warp}} = 5 \cdot 10^{-2}$ and a constant number of 6 total refinements are carried out before stopping the algorithm.

In Figure 5.8 we evaluate Algorithm 5.19 visually against the middlebury optical flow benchmark [11]. The color-coded images representing optical flow fields are normalized by the maximum motion of the ground truth flow data and black areas of the ground truth data represent unknown flow information, e.g. due to occlusion. A good resemblance of the computed optical flow to the ground truth and the effect of total variation regularization, i.e. sharp edges separating homogeneous regions, can be seen clearly. Large displacements are resolved quite well, e.g. the fast moving triangle-shaped object at the bottom of the RubberWhale benchmark, thanks to the warping algorithm employed. Using the same example, smaller slow-moving parts adjacent to the larger moving objects tend to get somewhat distorted however. It is unclear how much visual improvement more careful or adaptive parameter selection may give and further study remains to be done. Exemplary, the adapted mesh for the Dimetrodon benchmark can be seen in Figure 5.9, where refinement seems to take place largely around image edges.

5.6 Decomposition

We aim to implement the decomposition from Chapter 4 using the finite difference setting from Section 5.1. Let for $d \in \mathbb{N}$, $h = 1$, $a, b \in \mathbb{Z}^d$ the discrete domain be given by $\Omega_h := \Omega_{h,[a,b]} \subseteq \mathbb{Z}^d$ as in Section 5.1.

For a given discrete overlap $r \in \mathbb{N}$ and a desired number of domains $M \in \mathbb{N}$ we first define a discrete covering of Ω_h in dimension $d = 1$.

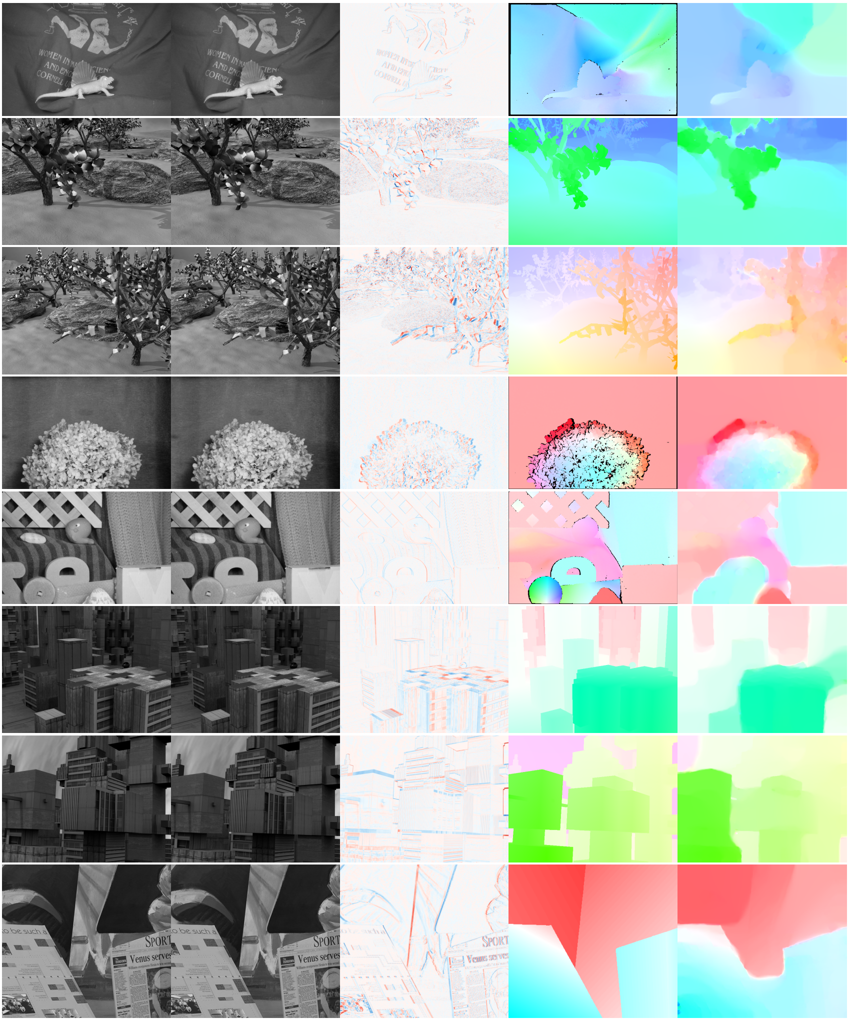


Figure 5.8: Middlebury Optical Flow Benchmark: columns from left to right: f_0, f_1, image difference $f_1 - f_0$, ground truth optical flow and computed optical flow u using the adaptive warping algorithm. Benchmarks from top to bottom: Dimetrodon, Grove2, Grove3, Hydrangea, RubberWhale, Urban2, Urban3, Venus.

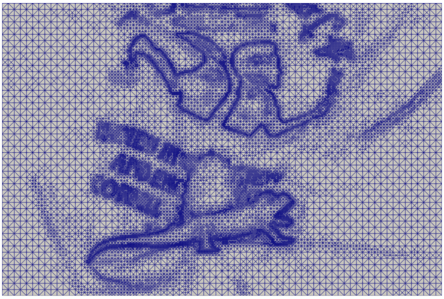


Figure 5.9: Exemplary adapted mesh for adaptive optical flow on the Dimetrodon Middlebury optical flow benchmark.

Let $s := |b - a|$ be the diameter of Ω, i.e. its length. Define M approximately equal integer sublengths given recursively by

$$a_i := \left\lfloor \frac{s + (M-1)r - \sum_{j=1}^{i-1}(a_j - r)}{M - (i-1)} \right\rfloor, \qquad i = 1, \dots, M.$$

These give rise to the subdomains

$$\Omega_i := \{b_i, b_i + 1, \dots, b_i + a_i\}, \qquad b_i := \sum_{j=1}^{i-1}(a_j - r), \qquad i = 1, \dots, M$$

of diameter a_i and the partition functions $\theta_i : \Omega \to [0,1]$ by

$$\theta_i(x) := \min\left\{1, \tfrac{1}{r}\operatorname{dist}(x, [0,s] \setminus [b_i, b_i + a_i])\right\},$$

where dist is the (Euclidean) distance function. The above construction in one dimension yields M discrete subdomains Ω_i and a corresponding partition of unity θ_i for a discrete domain Ω of any size provided M and r are chosen such that $a_i \geq 2r$.

Higher dimensions $d > 1$ are realized through a standard tensor-product formulation based on the above construction, yielding $M = \prod_{k=1}^{d} M_k$ subdomains with overlaps $r = (r_1, \dots, r_d)$.

In all our decomposition examples we use Algorithm 5.9 as a subproblem solver if not specified otherwise and choose its stepsize $\tau = \frac{1}{8\|B^{-1}\|}$ in accordance with Remark 5.10.

Convergence

We numerically verify the theoretical sublinear convergence properties of Algorithm 4.5 and Algorithm 4.6 due to Theorem 4.15 for different applications. In each case small data of size 48×32 was used together with $M = 2 \cdot 2 = 4$ domains and an overlap of $r = 5$ to make a high number of iterations timely feasible. For denoising, a maximum of 10^5 iterations was chosen, while for inpainting and optical flow 10^6 iterations were made. For denoising we use artificial additive Gaussian noise ($\sigma = 0.1$) on the ground truth image and model parameters $\lambda = 0.1$, $\beta = 0.0$.

We observe in Figure 5.10 similar behaviour as in [28], i.e. the sequential decomposition has a slight edge on the global algorithm due to domain-overlap, while the energy curve of the parallel averaging algorithm displays a characteristic bulge in the beginning.

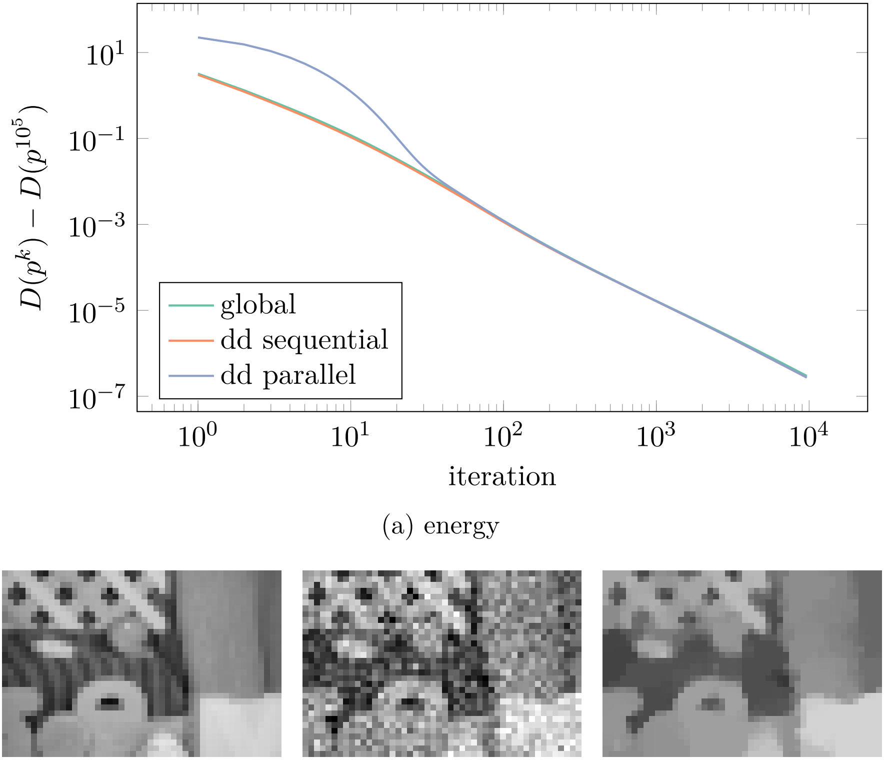
For inpainting we use model parameters $\lambda = 0.05$, $\beta = 0.001$, while for optical flow estimation we use model parameters $\lambda = 0.002$, $\beta = 0.001$. In Figures 5.11 and 5.12 the difference between the sequential and parallel algorithm is less visible for both inpainting and optical flow estimation. In all cases the domain decomposition algorithms converge at a sublinear rate comparable to the respective global algorithm.

Surrogate

For local operators B we compare (i) nesting the surrogate iteration (Algorithm 4.16) within domain decomposition and (ii) nesting domain decomposition within a global surrogate iteration. Note that for $B = I$, $\tau \to 1$ and a single surrogate iteration both of these are identical.

We use the optical flow problem with frames of original size 584×388 and model parameters $\beta = 0.001$, $\lambda = 0.01$. The number of subdomains is $M = 4 \cdot 4 = 16$ with overlap $r = 50$.

Both nestings perform similarly as can be seen in Figure 5.13, while nesting the surrogate iteration within the domain decomposition has a slight edge. This can be attributed to additional evaluations of B in regions of overlap.



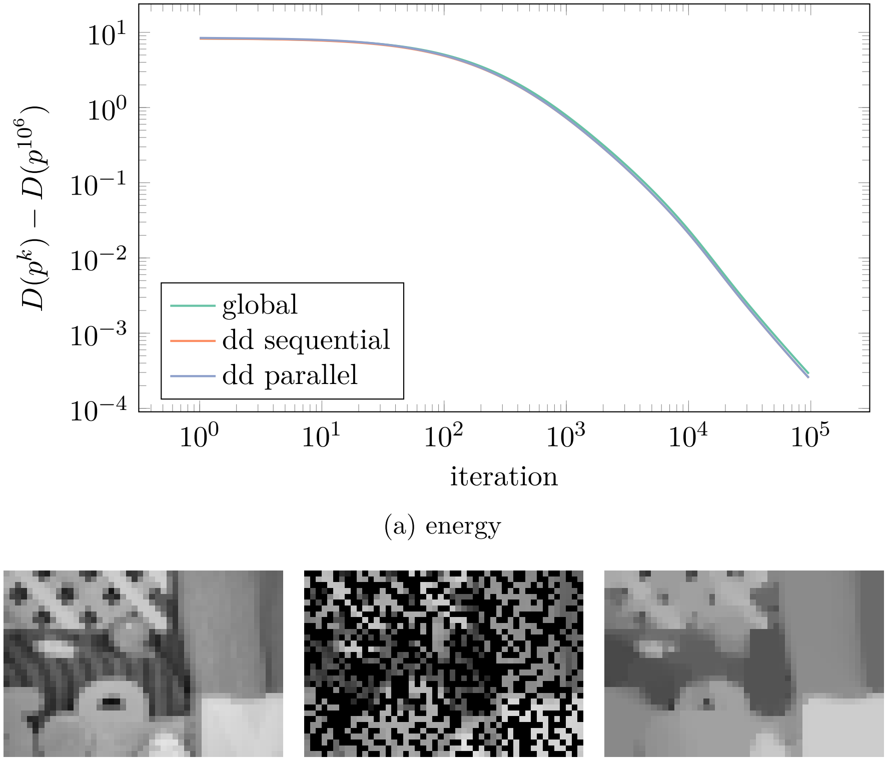
(a) energy

(b) ground truth image

(c) noisy input image

(d) denoised output image

Figure 5.10: denoising: convergence of energy and results



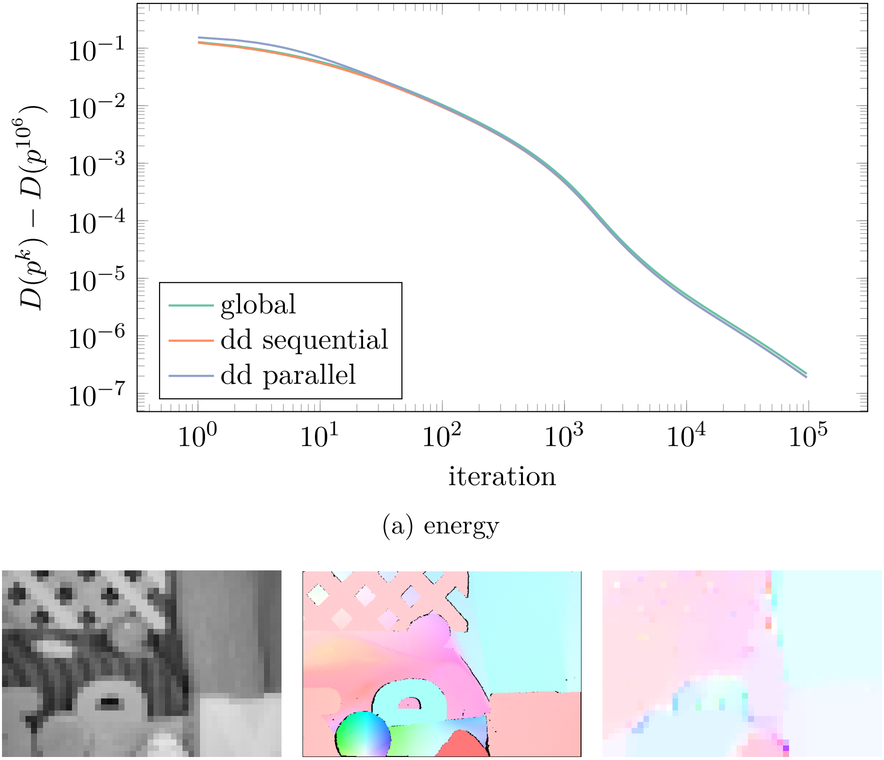
(a) energy

(b) ground truth image

(c) corrupted input image: half of all pixels black

(d) inpainted output image

Figure 5.11: inpainting: convergence of energy and results



(a) energy

(b) first image f_0 of image sequence

(c) optical flow ground truth from [11] (original resolution)

(d) optical flow computed

Figure 5.12: optical flow: convergence of energy and results

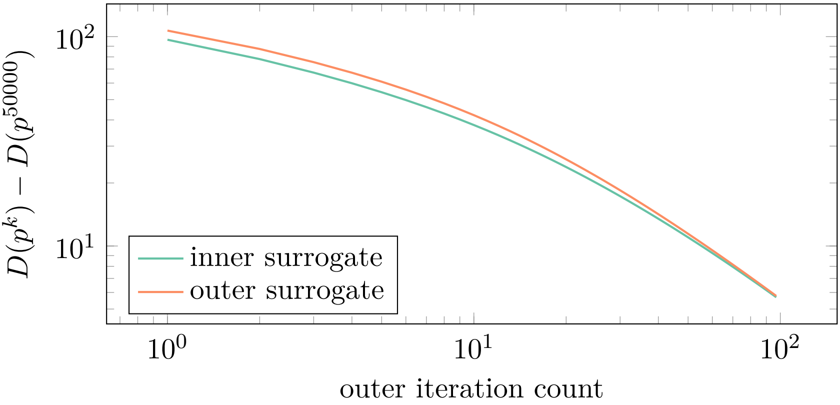


Figure 5.13: Comparison of outer and inner surrogate, 50 inner iterations per domain decomposition iteration, one single inner iteration per surrogate iteration

Wavelet Transformation

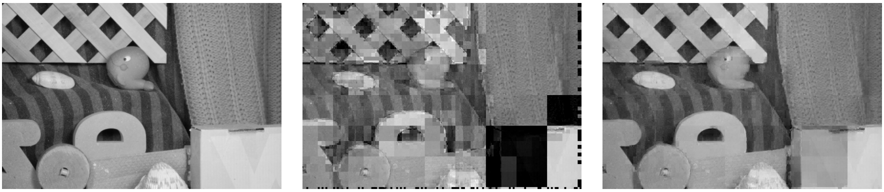
To demonstrate feasibility of our method even for global operators, we aim to apply it to the reconstruction of wavelet coefficients.

Let $u : \Omega_s \to \mathbb{R}$ and $k \in \mathbb{N}_0^d$ be such that $2k \leq s \leq 2k+1$. Define the d-dimensional n-th level discrete Haar wavelet transform $T^n : \mathbb{R}^{\Omega_s} \to \mathbb{R}^{\Omega_s}$ through $T^0 := I$ and for $n \geq 1$ recursively through

$$(T^n u)(\alpha \cdot k + x) := \begin{cases} (T^{n-1} T_0 u|_{\Omega_{2k}})(x) & \text{if } \alpha = 0,\ k \geq 1, \\ (T_\alpha u|_{\Omega_{2k}})(x) & \text{if } 0 \neq \alpha \leq 1,\ k \geq 1, \\ u(\alpha \cdot k + x) & \text{else,} \end{cases}$$

for all $\alpha \cdot k + x \in \Omega_s$ where $\alpha, x \in \mathbb{N}_0^d$, $x < k$ and the transformation on the orthant indicated by $\alpha \in \{0,1\}^d$ is given by $T_\alpha : \mathbb{R}^{\Omega_{2k}} \to \mathbb{R}^{\Omega_k}$ with

$$(T_\alpha u)(x) := 2^{-\frac{d}{2}} \sum_{\substack{\beta \in \mathbb{N}_0^d \\ \beta \leq 1}} (-1)^{|\alpha \cdot \beta|} u(2x + \beta)$$



(a) original image (b) corrupted image (c) reconstruction

Figure 5.14: wavelet inpainting

for all $x \in \mathbb{N}_0^d$, $x < k$.

Since $T_\alpha : \mathbb{R}^{\Omega_{2k}} \to \mathbb{R}^{\Omega_k}$ halves the size and for $s \leq 1$ we have $T^n = I$ for any $n \in \mathbb{N}$ the operator T^n becomes idempotent for large enough n and we thus conveniently denote by $T^\infty := \lim_{n\to\infty} T^n$ the full wavelet transform.

Let further $R : \mathbb{R}^{\Omega_s} \to \mathbb{R}^{\Omega_s}$ be an operator that sets a fixed set $I \subseteq \Omega_s$, $|I| = \frac{1}{2}|\Omega_s|$ of coefficients to zero. On may simulate corruption of wavelet coefficients by setting $T := R \circ T^\infty$ and $g = Tg_0$ for some original image $g_0 : \mathbb{R}^{\Omega_s} \to [0, 1]$.

We use $M = 4 \cdot 4 = 8$ domains and an overlap of $r = 5$. In Figure 5.14 the result of an reconstruction for $\lambda = 0.02$ and $\beta = 0.001$ can be seen after 10^2 outer iterations and 10^3 inner iterations each.

Parallel scaling

While our decomposition methods Algorithms 4.5 and 4.6 allow us to dissect the original problem into smaller ones, a computational benefit can only be achieved when those smaller problems are solved in parallel. We aim to show the feasibility of a parallel implementation for these decomposition methods and measure its parallel scaling behaviour.

Note that the subproblems of the parallel method from Algorithm 4.5 are trivially parallelizable, while Algorithm 4.6 may be parallelized using an appropriate coloring of the subdomains similar to [33] by calculating the solution on disjoint subdomains of the same color in parallel and subdomains of different colors in sequence. We test our

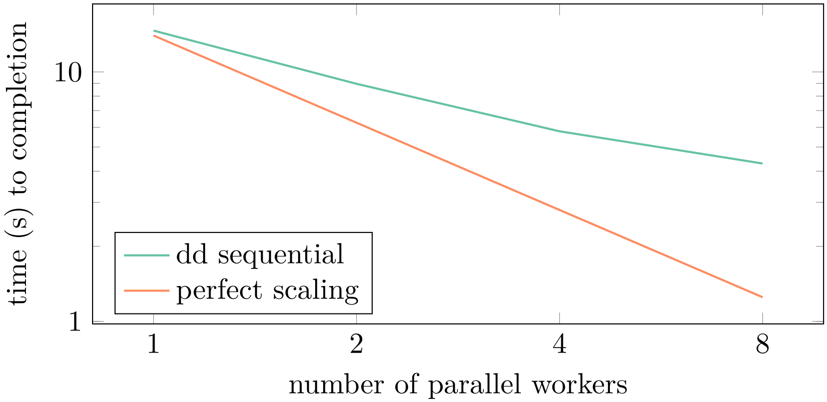


Figure 5.15: parallel scaling

parallel implementation on an Intel(R) Core(TM) i7-5820K CPU @ 3.30GHz (6 cores, 12 processing units) processor for scaling efficiency. Data and parameters are chosen as in the surrogate comparison above while the algorithms are terminated after reaching an energy of 130.0.

In Figure 5.15 one can see that the parallel implementation shows desirable almost linear scaling (note the logarithmic axes) though with a bad factor which we attribute to the data preparation and communication steps that are carried out on a single worker and apparently do not scale well in this implementation.

Conclusion

We have confirmed the convergence of Algorithms 4.5 and 4.6 numerically and shown applicability of the decomposition to a wider range of image processing tasks, namely inpainting and optical flow estimation. When considering the total number iterations of the inner subalgorithm, both decomposition methods did not differ substantially in terms of convergence speed from the global one, i.e. applying the subalgorithm directly to the global problem, which suggests a minimal overhead of the decomposition method. An expected runtime improvement by parallel execution compared to sequential execution of the same decomposition algorithm could be verified. Using domain decomposition for a memory-constrained computing environment is expected to be possible but requires a careful implementation and therefore remains to be shown.

6 Implementation

In Chapter 5 all algorithms were presented in a formal mathematical notation, disregarding any implementation details. While a formal presentation may be sufficient to fully describe the algorithm in question and analyze it in a theoretical setting, the implementation details influence various aspects of practical numerical research: development time, performance characteristics, independent reproducibility, reuse and maintainability.

In this chapter we will shortly outline the implementation which enabled us to produce the results of Chapter 5 and highlight various deliberate design choices. We used the Julia programming language [17] for our implementation and note that all Julia source code developed for this thesis has been made publicly available with permissive license through [50, 51, 52, 53, 54].

6.1 Optical Flow Utilities

While for digital images standardized image file formats are common, there seems to be no standardized file format available for storing the vector-valued optical flow field. The Middlebury optical flow benchmark [11] contains an informal description of an uncompressed binary format to store flow field data as `flo`-files [77]. Since then the `flo` file format has been used in various benchmarks as a submission format or to provide ground-truth optical flow data [11, 23].

Since the Julia package ecosystem did not provide a way to read or write flow field data, the package `OpticalFlowUtils` [51] was developed and published under the MIT ”Expat” License. The implementation is Julia-native and thus does not derive from the original C++-implementation provided in [11] which has unclear licensing.

This package has since been accepted to the default Julia package registry and registered as the default handler for reading and writing `flo` files by the `FileIO` framework [36]. It loads flow data as an array of type `Array{Union{Missing, Float32}, 3}`, where the last two ranks extend over the image grid and the first rank over the two vector field dimensions. The singleton value `missing` indicates that the flow data is unavailable, e.g. due to occlusion. For visualization a function `colorflow` converts the flow array into a color-coded image as used for visualization in e.g. Figure 5.8 and similar to [11]. A minimal code example for loading, plotting and saving of optical flow data is shown below.

```julia
using OpticalFlowUtils, FileIO, Plots

x = load("input.flo")
plot(colorflow(x))
save("output.flo", x)
```

6.2 Kernel Operations

Numerical algorithms often perform operations on arrays of data, evaluating the same function at every index by passing it data from within a small window (sometimes called ‘stencil’) of array data around that index. We call the function evaluated this way *kernel function* and the resulting operation *kernel operation*. Best known examples include finite difference operators, linear and non-linear image filters, morphological image operations as well as convolutional neural nets.

Implementation usually involves writing a loop ranging over all indices and performing the kernel operation within the loop body. Special care has to be taken to account for index ranges and boundary conditions depending on the window size. When slice indexing over multiple dimensions is available (e.g. in array programming languages) one may replace the explicit loop with broadcasted operations over shifted array slices. Both methods involve tedious index-notation and manual handling of boundary conditions.

The package StaticKernels [54] has been developed to ease the process of writing custom kernel operations in Julia. It has since been registered and accepted into the default Julia package registry. The custom kernel operation is created by defining the kernel function operating on the window of array data using relative indexing (with 0 being the current position). Applying the two-dimensional Laplace finite difference operator to an array of random floating point values may be performed as follows:

```julia
using StaticKernels
a = rand(100, 100)

k = @kernel w ->
    w[0,-1] + w[-1,0] - 4*w[0,0] + w[1,0] + w[0,1]
b = map(k, a) # size(b) == (98, 98)
```

The map function, traditionally accepting a function to apply pointwise to every entry of an array, has been extended to accept a kernel object k consisting of the kernel function and a window range and performs here the application of the Laplace finite difference kernel operation. Note that the resulting array size has diminished since we did not specify any boundary handling.

The boundary may be incorporated by tagging the array to be operated on with an extension specification. In the following example the anisotropic total variation of a grey-scaled image is computed.

```julia
k = @kernel w ->
    abs(w[1,0] - w[0,0]) + abs(w[0,1] - w[0,0])
sum(k, extend(a, StaticKernels.ExtensionReplicate()))
```

Note again that sum has been extended to accept a kernel object in place of a pointwise function.

For more complicated examples when the window size cannot be inferred from the kernel function, explicit creation of the kernel object is still possible. The following example illustrates this for the kernel operation of Conway's game of life.

```
a = rand(Bool, 1000, 1000)
k = Kernel{(-1:1,-1:1)}(@inline w ->
    count(w) - w[0,0] == 3 || count(w) == 3 && w[0,0])
a .= map(k, extend(a,
    StaticKernels.ExtensionConstant(false)))
```

We see in Table 6.1 for linear image filter operations that the kernel operation is carried out without heap memory allocations and performs competitively with packages applicable to the same specific task. Note that for comparison reasons the benchmark does not incorporate boundary conditions. `StaticKernels` and the (auto-vectorized) code produced manages to keep up with the package `LoopVectorization` [44] which transforms loop bodies under various additional assumptions to vectorized LLVM assembly code using hand-picked optimized rules. The package `LoopVectorization` could in principle be used as a backend for transforming the loop generated by the `StaticKernels` package in order to harness these optimizations, which become more relevant for larger window sizes.

6.3 Stateful Parallelism

For executing CPU-based computing tasks in parallel the Julia language natively supports shared-memory multi-threading and distributed process-parallelism. The former (provided by the standard library package `Threads`) schedules a number of computing tasks to run in parallel on multiple CPU-threads or CPU-cores sharing access to the same working memory. While the overhead of launching these tasks is small, care has to be taken to avoid concurrent writes to the same memory location. Further multi-threading does not scale well beyond a single computing node due to the shared-memory requirement. Dis-

	100 × 100		1000 × 1000	
package name	time, µs	memory	time, ms	memory
StaticKernels (v0.6.1)	15.292	0 B	2.867	0 B
LoopVectorization (v0.12.89)	12.396	0 B	2.657	0 B
NNlib (v0.7.29)	231.295	678.55 KiB	57.642	68.39 MiB
ImageFiltering (v0.7.0)	146.377	81.11 KiB	20.015	7.63 MiB

Table 6.1: Results for benchmark code [54] on an AMD A10-7860K Radeon R7 CPU using Julia v1.6.3, applying a 3 × 3 linear filter to an array of the specified size.

tributed process-parallelism (provided by the standard library package `Distributed`) schedules computing tasks to execute on separate Julia processes through an asynchronous communication protocol. These Julia processes run independently and may execute either locally in order to make use of multi-core CPUs or remotely for use in a computing cluster. While distributed parallelism does not suffer from race conditions and can scale well beyond a single computing node, an efficient algorithm design to minimize the communication overhead between the computing nodes becomes crucial.

Stateless computation can be scheduled with relative ease using e.g. the method `Distributed.pmap`, which transfers all inputs to the worker nodes, executes the computation and collects the outputs again. The transfer of input and output data thus introduces a latency overhead for each parallel computation. If the remote computation can be done incrementally it makes sense to let the remote computing tasks be stateful, i.e. run computation on demand for an incremental input while having an internal persistent program state for each remote task.

The package `Outsource` [52] was developed to provide a minimal wrapper atop the `Distributed` interface to ease management of stateful remote tasks, including spawning and communicating with them using a simple bidirectional channel-like interface built upon `Distributed.RemoteChannel`.

The following minimal example launches one remote worker to incrementally calculate terms of the Fibonacci sequence upon request. The bidirectional channel `wc` (`c` from the perpective of the remote worker) is interfaced through `put!` and `take!` to convey the request order and to receive a Fibonacci number as a result. From a remote perspective this interface appears reversed.

```julia
using Distributed; addprocs(1)
@everywhere using Outsource

# spawn worker
wc = outsource() do c
    state = (0, 1)
    while isopen(c)
        n = take!(c)
        # compute next n Fibonacci numbers
        for _ in 1:n
            state = (state[2], state[1] + state[2])
        end
        put!(c, state[1])
    end
end

# issue stateful work and retrieve result
put!(wc, 10)
take!(wc) # = 55
put!(wc, 10)
take!(wc) # = 6765
```

While the example above may appear trivial, more complex scenarios may spawn multiple workers with different states by making use of function closures. These workers can then execute remotely in parallel while being orchestrated on the main thread using one bidirectional channel per worker.

6.4 Domain Decomposition

The domain decomposition methods featured in Section 5.6 were implemented as the Julia package `DualTVDD` [50]. Apart from all numerical experiments of Section 5.6, the package contains general code to subdivide domains and apply subalgorithms according to the decomposition Algorithms 4.5 and 4.6, the general surrogate Algorithm 4.16 as well as an implementation of Algorithm 5.9. Notably the implementation supports arbitrary dimensions.

Finite difference operations make use of `StaticKernels` from Section 6.2 and parallel execution of subdomain algorithms is accomplished using `Outsource` from Section 6.3.

6.4.1 Algorithm Interface

An algorithm interface has been designed to allow for efficient execution and arbitrary nesting of algorithms.

```
abstract type Problem end
abstract type Algorithm{P <: Problem} end
abstract type State end
```

The abstract subtype hierarchy of `Problem` specifies the problem model interface (problem, available data, available oracles, solution form) and any concrete type instance fully specifies all problem parameters.

Any abstract subtype of `Algorithm` represents a fully specified iterative process of usually infinite or a-priori fixed but unknown finite length, producing iterates that approximate the solution to the dependent problem in some way. The hierarchy of abstract subtypes of `Algorithm` is based on the algorithm interface (e.g. accepted parameters) and the specific numerical scheme. A concrete type `Algorithm` represents an implementation of its supertype algorithm and its instance is a complete specification of the algorithm with all its inputs and should guarantee to produce a deterministic sequence of iterates (randomized algorithms should use a seeded pseudorandom number generator). An concrete algorithm type needs to implement at least

(i) `state = init(::Algorithm)`: allocates and initializes the algorithm state.

(ii) `state = step!(::State)`: performs one iteration of the algorithm by updating `state`. This method should never allocate dynamic memory.

(iii) `solution = fetch(::State)`: fetches the solution in a format specified by the problem. In some cases this can be a non-trivial operation if the algorithm uses a different internal representation.

Depending on its supertype interface additional methods may be implemented.

An instance of `State` represents the algorithm workspace. It contains sufficient information to act as a checkpoint for continuing the algorithm at that point.

Using the method dispatch capabilities of Julia this interface design allows to e.g. apply the same algorithm to different problems satisfying the same problem interface as well as solving the same problem by different algorithms, provided they adhere to a common problem interface. Algorithms may also be wrapped to produce e.g. logging variants without modifying the actual algorithm code.

6.5 Finite Elements

The numerical results from Section 5.4 use a finite element implementation that has been developed from scratch in Julia. The package `SemiSmoothNewton` [53] contains scripts to reproduce the numerical results and a module which hosts the finite element framework and utility functions. The finite element framework supports the following main features:

- unstructured conforming two-dimensional triangle grid with local bisection refinement and prolongation of grid functions,
- elements of type DP0/DP1 (piecewise constant/linear discontinuous), P1 (piecewise linear continuous)
- statically sized array-valued grid functions of arbitrary rank and pointwise expressions using native Julia syntax
- interpolation and stable projections from pointwise expressions to grid functions
- sampling of array-valued grid functions to array-valued images

While existing Julia finite element packages (notably [10, 26, 46]) could have been used and suitably extended to support all of these features, a minimal reimplementation was expected to be simpler to achieve and manipulate.

As a minimal example, suppose $\Omega = (0,1)^2 \subseteq \mathbb{R}^2$, $f(x) = 5e^{-5\|x-p\|^2}$, $p = (\frac{1}{4}, \frac{1}{2})$. The screened Poisson equation with natural boundary conditions in the unknown u is given by:

$$\begin{aligned} 0 &= -\Delta u + \lambda^2 u - f && \text{in } \Omega, \\ 0 &= \partial_\nu u && \text{on } \partial\Omega. \end{aligned}$$

The corresponding weak formulation amounts to finding $u \in H^1(\Omega)$ with $a(u,\varphi) = l(\varphi)$ for all $\varphi \in H^1(\Omega)$, where $a : H^1(\Omega) \times H^1(\Omega) \to \mathbb{R}$, $l : H^1(\Omega) \to \mathbb{R}$ are given by

$$\begin{aligned} a(u,\varphi) &:= \int_\Omega \nabla u \cdot \nabla\varphi + \lambda^2 u \cdot \varphi \,\mathrm{d}x, \\ l(\varphi) &:= \int_\Omega f \cdot \varphi \,\mathrm{d}x. \end{aligned}$$

The following code snippet computes a discrete solution on a 32 × 32 grid of piecewise linear continuous finite elements.

```julia
using LinearAlgebra: dot, norm
using SemiSmoothNewton
using SemiSmoothNewton:
    vtk_mesh, vtk_append!, vtk_save

# simple 2d unit square triangulation
mesh = init_grid(32, 32)

# scalar piecewise linear continuous elements
space = FeSpace(mesh, P1(), (1,))
u = FeFunction(space, name = "u")

# inhomogenity
const p = (0.25, 0.5)
f(x) = 5 * exp(-5 * norm(x .- p)^2)
# parameter
const lambda = 19.7

# weak formulation
a(x, u, d_u, phi, d_phi) =
    dot(d_u, d_phi) + lambda^2 * dot(u, phi)
l(x, phi, d_phi) =
    dot(f(x), phi)

# solve u
A, b = assemble(space, a, l)
u.data .= A \ b

# vtk output
vtk = vtk_mesh("output.vtu", mesh)
vtk_append!(vtk, u)
vtk_save(vtk)
```

Notably the weak formulation is provided by defining functions `a` and

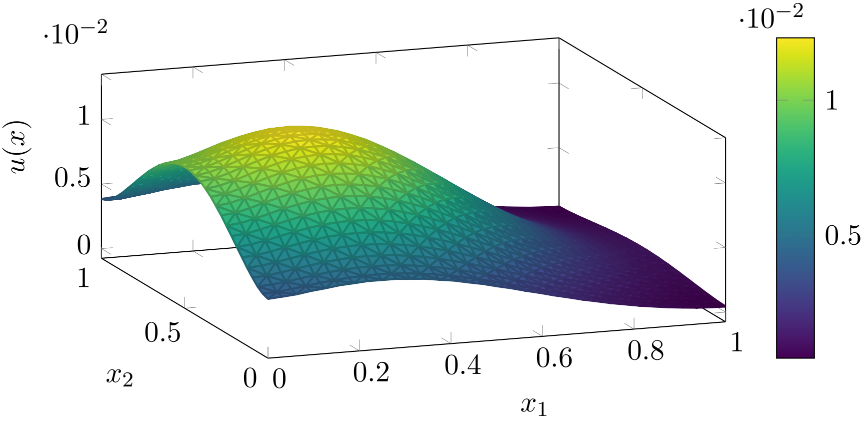


Figure 6.1: Computed solution u on corresponding mesh.

`l` which are then passed to the `assemble` routine. Both functions represent the pointwise integrand when considering a, l as being functionals defined by a domain integral, e.g. $a(u, \varphi) = \int_\Omega \tilde{a}(x, u, \nabla u, \varphi, \nabla \varphi) \,\mathrm{d}x$. They are defined as standard native functions and operate on stack-allocated fixed-size operands.

The solution u generated by the code above can be seen in Figure 6.1, where the two-dimensional grid has been visualized as a surface plot.

7 Outlook

At last, we use this chapter to describe some related topics, which could potentially provide interesting experiments or future research. This is by no means an exhaustive list and mostly reflects the mindset of the author himself.

7.1 Finite Element Discretization of L^1-Type Functionals

In Chapter 5 we solved the discretized optimality condition (5.2). This discrete solution does not necessarily agree with minimizing the continuous primal functional over a discrete subspace. While this work concentrated on the former discretization, the latter approach may be interesting to investigate.

Choosing e.g. simplicial, cellwise linear finite elements, the involved L^1-term can be evaluated in the local Lagrange basis as follows. In one dimension for $u : \mathbb{R} \to \mathbb{R}$, $u(x) = u_0(1-x) + u_1 x$, $u_0, u_1 \in \mathbb{R}$ we have

$$\int_0^1 |u(x)| \,\mathrm{d}x = \frac{|u_0|u_0 - |u_1|u_1}{2(u_0 - u_1)}.$$

While in two dimensions for $u : \mathbb{R}^2 \to \mathbb{R}$, $u(x) = u_0(1 - x_1 - x_2) + u_1 x_1 + u_2 x_2$, $u_0, u_1, u_2 \in \mathbb{R}$ we have

$$\int_0^1 \int_0^{1-x_1} |u(x)| \,\mathrm{d}x_2 \,\mathrm{d}x_1 \\ = \frac{|u_0|u_0^2(u_1 - u_2) - |u_1|u_1^2(u_0 - u_2) + |u_2|u_2^2(u_0 - u_1)}{6(u_0 - u_1)(u_0 - u_2)(u_1 - u_2)}.$$

These expressions are still convex with respect to the basis coefficients (it is however unclear whether this is necessarily true for a higher order Lagrange basis). Using these expressions, one may now apply the Fenchel duality on the discrete finite element spaces.

While being more involved, by discretizing this way, a solution to the discrete primal-dual system will also be a minimizer of the continuous primal functional on the discrete subspace. Being able to guarantee that the error of discrete and continuous solution is orthogonal to the chosen subspace is expected to help in deriving approximation guarantees and a posteriori error estimates. Notably, convergence results towards a continous solution as the mesh size tends to zero, as seen e.g. in [5], would no longer be necessary, provided approximation properties of the involved discrete spaces hold.

7.2 Software

7.2.1 Sparse Jacobians for Kernel Operations

The kernel operations detailed in Section 6.2 may be subject to automatic differentiation. Using e.g. the package `ReverseDiff`, jacobians for kernel operations may readily be computed as follows:

```
using StaticKernels
using ReverseDiff: jacobian
k = @kernel w -> w[-1] - 2*w[0] + w[1]
A = jacobian(x -> map(k, x), rand(5))
# 3×5 Matrix{Float64}:
# 1.0 -2.0 1.0 0.0 0.0
# 0.0 1.0 -2.0 1.0 0.0
# 0.0 0.0 1.0 -2.0 1.0
```

Since the kernel operations are generally sparse operations, it is judicious to save time by not computing the dense jacobian. There exists e.g. the package `SparseDiffTools`, which facilitates creating sparse jacobian operators by detecting the sparsity pattern of an arbitrary operation, finding a suitable graph-coloring and evaluating a jacobian product in only few reverse automatic differentiation passes. But for our kernel operations in particular, the sparsity pattern is well known and could be leveraged to implement a sparse jacobian operator.

7.2.2 Broadcasted Kernel Fusion

Kernels from Section 6.2 are not yet subject to composition, i.e. combining multiple kernels by simple operations to produce a new kernel. For example, we would expect for the following two finite difference kernels

```
k1 = @kernel w -> w[1] - w[0]
k2 = @kernel w -> w[2] - 2*w[1] + w[0]
```

the code `k1 ∘ k1` to produce a kernel equivalent to `k2`.

While it is straightforward to apply `k1` twice, this would loop over the input array twice. It is therefore desirable to fuse both operations into one loop. That way, the compiler may optimize the combined operation further than possible if both operations were separated.

7.2.3 Differentiable Finite Element Toolbox

Differentiable programming is the paradigm to implement algorithms in a way which allows them to be subjected to automatic differentiation. This allows e.g. for sensitivity analysis of the algorithm or its integration into other gradient-based minimization pipelines, such as deep learning training loops.

While e.g. adjoint models of partial differential equations for sensitivity analysis are not a recent invention, we are not aware of a fully differentiable finite element toolbox which allows differentiation with regard to all its primitives. Besides model input and parameters, this should include differentiation with regard to e.g. mesh vertex coordinates, quadrature points and weights, potentially allowing for model-adaptive moving meshes and quadratures.

A major roadblock for more practical usage of differentiable programming techniques in the finite element community seems to be a lack of appropriate seamless frameworks.

References

[1] Rémy Abergel and Lionel Moisan. ‘The Shannon total variation’. In: *Journal of Mathematical Imaging and Vision* 59.2 (Oct. 2017), pages 341–370. DOI: 10.1007/s10851-017-0733-5.

[2] Robert A. Adams and John J. F. Fournier. *Sobolev Spaces.* 2nd edition. Volume 140. Pure and Applied Mathematics. Academic Press, June 2003, pages 1–305. ISBN: 978-0-12-044143-3.

[3] Mark Ainsworth and J. Tinsley Oden. *A Posteriori Error Estimation in Finite Element Analysis.* Wiley-Interscience, 2000. DOI: 10.1002/9781118032824.

[4] Charalambos D. Aliprantis and Kim C. Border. *Infinite Dimensional Analysis. A Hitchhiker’s Guide.* 3rd edition. Springer, 2006. DOI: 10.1007/3-540-29587-9.

[5] Martin Alkämper and Andreas Langer. ‘Using DUNE-ACFem for non-smooth minimization of bounded variation functions’. In: *Archive of Numerical Software* 5.1 (2017), pages 3–19. DOI: 10.11588/ans.2017.1.27475.

[6] Stefano Alliney. ‘A property of the minimum vectors of a regularizing functional defined by means of the absolute norm’. In: *IEEE Transactions on Signal Processing* 45.4 (Apr. 1997), pages 913–917. DOI: 10.1109/78.564179.

[7] Luigi Ambrosio, Nicola Fusco and Diego Pallara. *Functions of Bounded Variation and Free Discontinuity Problems.* Oxford Mathematical Monographs. New York: Oxford University Press, May 2000. ISBN: 0-19-850245-1.

[8] Hedy Attouch, Giuseppe Buttazzo and Gérard Michaille. *Variational Analysis in Sobolev and BV Spaces. Applications to PDEs and optimization.* 2nd edition. MOS-SIAM Series on Optimization. Mathematical Optimization Society and Society for Industrial and Applied Mathematics, 2014. DOI: 10.1137/1.9781611973488.

[9] Gilles Aubert and Pierre Kornprobst. *Mathematical Problems in Image Processing. Partial differential equations and the calculus of variations.* 2nd edition. Volume 147. Applied Mathematical Sciences. New York: Springer, 2006. DOI: 10.1007/978-0-387-44588-5.

[10] Santiago Badia and Francesc Verdugo. 'Gridap: an extensible finite element toolbox in Julia'. In: *Journal of Open Source Software* 5.52 (2020), page 2520. DOI: 10.21105/joss.02520.

[11] Simon Baker, Daniel Scharstein, James P. Lewis, Stefan Roth, Michael J. Black and Richard Szeliski. 'A database and evaluation methodology for optical flow'. In: *International Journal of Computer Vision* 92.1 (2011), pages 1–31. DOI: 10.1007/s11263-010-0390-2.

[12] Sören Bartels. *Numerical Methods for Nonlinear Partial Differential Equations.* Volume 47. Springer Series in Computational Mathematics. Springer, 2015. DOI: 10.1007/978-3-319-13797-1.

[13] Sören Bartels. 'Total variation minimization with finite elements: convergence and iterative solution'. In: *SIAM Journal on Numerical Analysis* 50.3 (2012), pages 1162–1180. DOI: 10.1137/11083277X.

[14] Sören Bartels and Marijo Milicevic. 'Primal-dual gap estimators for a posteriori error analysis of nonsmooth minimization problems'. In: *ESAIM: Mathematical Modelling and Numerical Analysis* 54.5 (2020), pages 1635–1660. DOI: 10.1051/m2an/2019074.

[15] Zakaria Belhachmi and Frédéric Hecht. 'An adaptive approach for the segmentation and the TV-filtering in the optic flow estimation'. In: *Journal of Mathematical Imaging and Vision* 54.3 (2016), pages 358–377. DOI: `10.1007/s10851-015-0608-6`.

[16] Zakaria Belhachmi and Frédéric Hecht. 'Control of the effects of regularization on variational optic flow computations'. In: *Journal of Mathematical Imaging and Vision* 40.1 (2011), pages 1–19. DOI: `10.1007/s10851-010-0239-x`.

[17] Jeff Bezanson, Alan Edelman, Stefan Karpinski and Viral B. Shah. 'Julia: a fresh approach to numerical computing'. In: *SIAM Review* 59.1 (2017), pages 65–98. DOI: `10.1137/141000671`.

[18] Jakub W. Both. 'On the rate of convergence of alternating minimization for non-smooth non-strongly convex optimization in Banach spaces'. In: *Optimization Letters* 16.2 (2022), pages 729–743. DOI: `10.1007/s11590-021-01753-w`.

[19] Andrea Braides. *Γ-convergence for Beginners*. Volume 22. Oxford Lecture Series in Mathematics and its Applications. Oxford University Press, 2002. DOI: `10.1093/acprof:oso/9780198507840.001.0001`.

[20] Kristian Bredies, Karl Kunisch and Thomas Pock. 'Total generalized variation'. In: *SIAM Journal on Imaging Sciences* 3.3 (2010), pages 492–526. DOI: `10.1137/090769521`.

[21] Kristian Bredies and Dirk Lorenz. *Mathematical Image Processing*. Edited by John J. Benedetto. 1st edition. Applied and Numerical Harmonic Analysis. Birkhäuser, 2018. DOI: `10.1007/978-3-030-01458-2`.

[22] Martin Burger, Konstantinos Papafitsoros, Evangelos Papoutsellis and Carola-Bibiane Schönlieb. 'Infimal convolution regularisation functionals of BV and L^p spaces'. In: *Journal of Mathematical Imaging and Vision* 55 (2016), pages 343–369. DOI: `10.1007/s10851-015-0624-6`.

[23] Daniel J. Butler, Jonas Wulff, Garret B. Stanley and Michael. J. Black. 'A naturalistic open source movie for optical flow evaluation'. In: *Computer Vision – ECCV 2012*. Edited by Andrew Fitzgibbon, Svetlana Lazebnik, Pietro Perona, Yoichi Sato and Cordelia Schmid. Volume 7577. Lecture Notes in Computer Science. Springer, Oct. 2012, pages 611–625. DOI: 10.1007/978-3-642-33783-3_44.

[24] Corentin Caillaud and Antonin Chambolle. 'Error estimates for finite differences approximations of the total variation'. Preprint. Apr. 2020. URL: https://hal.archives-ouvertes.fr/hal-02539136.

[25] Luca Calatroni, Juan Carlos De Los Reyes and Carola-Bibiane Schönlieb. 'Infimal convolution of data discrepancies for mixed noise removal'. In: *SIAM Journal on Imaging Sciences* 10.3 (2017), pages 1196–1233. DOI: 10.1137/16M1101684.

[26] Kristoffer Carlsson, Fredrik Ekre and Contributors. *Ferrite.jl: Finite Element Toolbox for Julia*. Version 0.3.0. Mar. 2021. URL: https://github.com/Ferrite-FEM/Ferrite.jl.

[27] Kevin W. Cassel. *Variational Methods with Applications in Science and Engineering*. Cambridge University Press, 2013. DOI: 10.1017/CBO9781139136860.

[28] Antonin Chambolle. 'An algorithm for total variation minimization and applications'. In: *Journal of Mathematical Imaging and Vision* 20.1-2 (2004), pages 89–97. ISSN: 0924-9907. DOI: 10.1023/B:JMIV.0000011325.36760.1e.

[29] Antonin Chambolle, Vicent Caselles, Daniel Cremers, Matteo Novaga and Thomas Pock. 'An introduction to total variation for image analysis'. In: *Theoretical Foundations and Numerical Methods for Sparse Recovery*. Volume 9. Radon Series on Computational and Applied Mathematics. De Gruyter, 2010, pages 263–340. DOI: 10.1515/9783110226157.263.

[30] Antonin Chambolle, Stacey E. Levine and Bradley J. Lucier. ‘An upwind finite-difference method for total variation–based image smoothing’. In: *SIAM Journal on Imaging Sciences* 4.1 (2011), pages 277–299. DOI: 10.1137/090752754.

[31] Antonin Chambolle and Thomas Pock. ‘A first-order primal-dual algorithm for convex problems with applications to imaging’. In: *Journal of Mathematical Imaging and Vision* 40.1 (2011), pages 120–145. DOI: 10.1007/s10851-010-0251-1.

[32] Antonin Chambolle and Thomas Pock. ‘Learning consistent discretizations of the total variation’. In: *SIAM Journal on Imaging Sciences* 14.2 (2021), pages 778–813. DOI: 10.1137/20M1377199.

[33] Huibin Chang, Xue-Cheng Tai, Li-Lian Wang and Danping Yang. ‘Convergence rate of overlapping domain decomposition methods for the Rudin–Osher–Fatemi model based on a dual formulation’. In: *SIAM Journal on Imaging Sciences* 8.1 (2015), pages 564–591. DOI: 10.1137/140965016.

[34] Philippe G. Ciarlet. *The Finite Element Method for Elliptic Problems.* Classics in Applied Mathematics. SIAM, 2002. DOI: 10.1137/1.9780898719208.

[35] Laurent Condat. ‘Discrete total variation: new definition and minimization’. In: *SIAM Journal on Imaging Sciences* 10.3 (2017), pages 1258–1290. DOI: 10.1137/16M1075247.

[36] Simon Danisch, Tim Holy and Contributors. *FileIO: A Common Framework for Detecting File Formats and Dispatching to Appropriate Readers/Writers.* Version 1.11.1. Sept. 2021. URL: https://github.com/JuliaIO/FileIO.jl.

[37] Robert Dautray and Jacques-Louis Lions. *Mathematical Analysis and Numerical Methods for Science and Technology. Spectral Theory and Applications.* Volume 3. Mathematical Analysis and Numerical Methods for Science and Technology. Springer, 2000. ISBN: 978-3-540-66099-6.

[38] Mauricio Delbracio, Damien Kelly, Michael S. Brown and Peyman Milanfar. 'Mobile computational photography: a tour'. In: *Annual Review of Vision Science* 7 (Sept. 2021), pages 571–604. DOI: 10.1146/annurev-vision-093019-115521.

[39] Hendrik Meinert Dirks. 'Variational methods for joint motion estimation and image reconstruction'. PhD thesis. University of Münster, 2015. URL: https://nbn-resolving.de/urn:nbn:de:hbz:6-59219499925.

[40] Yiqiu Dong, Michael Hintermüller and Marrick Neri. 'An efficient primal-dual method for L^1TV image restoration'. In: *SIAM Journal on Imaging Sciences* 2.4 (2009), pages 1168–1189. DOI: 10.1137/090758490.

[41] Yiqiu Dong, Michael Hintermüller and M. Monserrat Rincon-Camacho. 'Automated regularization parameter selection in multi-scale total variation models for image restoration'. In: *Journal of Mathematical Imaging and Vision* 40.1 (2011), pages 82–104. DOI: 10.1007/s10851-010-0248-9.

[42] Manfred Einsiedler and Thomas Ward. *Functional Analysis, Spectral Theory, and Applications.* Edited by Sheldon Axler and Kenneth Ribet. Volume 276. Graduate Texts in Mathematics. Springer, 2017. DOI: 10.1007/978-3-319-58540-6.

[43] Ivar Ekeland and Roger Témam. *Convex Analysis and Variational Problems.* English. Volume 28. Classics in Applied Mathematics. SIAM, 1999. DOI: 10.1137/1.9781611971088.

[44] Chris Elrod. *LoopVectorization: Macro(s) for vectorizing loops.* Version v0.12.89. 2021. URL: https://github.com/JuliaSIMD/LoopVectorization.jl.

[45] Alexandre Ern and Jean-Luc Guermond. 'Finite element quasi-interpolation and best approximation'. In: *ESAIM: Mathematical Modelling and Numerical Analysis* 51.4 (2017), pages 1367–1385. DOI: 10.1051/m2an/2016066.

[46] Tero Frondelius and Jukka Aho. 'JuliaFEM - open source solver for both industrial and academia usage'. In: *Rakenteiden Mekaniikka* 50.3 (2017), pages 229–233. DOI: 10.23998/rm.64224.

[47] Bastian Goldluecke, Evegeny Strekalovskiy and Daniel Cremers. 'The natural vectorial total variation which arises from geometric measure theory'. In: *SIAM Journal on Imaging Sciences* 5.2 (2012), pages 537–563. DOI: 10.1137/110823766.

[48] Zheng Gong, Zuowei Shen and Kim-Chuan Toh. 'Image restoration with mixed or unknown noises'. In: *Multiscale Modeling & Simulation* 12.2 (2014), pages 458–487. DOI: 10.1137/130904533.

[49] Marc Herrmann, Roland Herzog, Stephan Schmidt, José Vidal-Núñez and Gerd Wachsmuth. 'Discrete total variation with finite elements and applications to imaging'. In: *Journal of Mathematical Imaging and Vision* 61.4 (2019), pages 411–431. DOI: 10.1007/s10851-018-0852-7.

[50] Stephan Hilb. *DualTVDD: Dual total variation decomposition algorithm and related tools*. Version 0.1. 2021. URL: https://gitlab.mathematik.uni-stuttgart.de/stephan.hilb/DualTVDD.jl.

[51] Stephan Hilb. *OpticalFlowUtils: Basic operations for handling optical flow vector fields*. 2021. URL: https://github.com/stev47/OpticalFlowUtils.jl.

[52] Stephan Hilb. *Outsource: Simple and explicit asychronous handling of stateful worker tasks*. 2021. URL: https://github.com/stev47/Outsource.jl.

[53] Stephan Hilb. *SemiSmoothNewton: A tiny finite element framework, primal-dual algorithms and numerical examples*. Version 0.1. 2021. URL: https://gitlab.mathematik.uni-stuttgart.de/stephan.hilb/SemiSmoothNewton.jl.

[54] Stephan Hilb. *StaticKernels: Julia-native non-allocating kernel operations on arrays.* 2021. URL: `https://github.com/stev47/StaticKernels.jl`.

[55] Stephan Hilb and Andreas Langer. 'A general decomposition method for a convex problem related to total variation minimization'. In preparation. 2022.

[56] Stephan Hilb, Andreas Langer and Martin Alkämper. 'A primal-dual finite element method for scalar and vectorial total variation minimization'. In: *Journal of Scientific Computing* 96.1 (2023), page 24. ISSN: 1573-7691. DOI: `10.1007/s10915-023-02209-2`.

[57] Michael Hintermüller, Kazufumi Ito and Karl Kunisch. 'The primal-dual active set strategy as a semismooth Newton method'. In: *SIAM Journal on Optimization* 13.3 (2003). DOI: `10.1137/S1052623401383558`.

[58] Michael Hintermüller and Karl Kunisch. 'Total bounded variation regularization as a bilaterally constrained optimization problem'. In: *SIAM Journal on Applied Mathematics* 64.4 (2004), pages 1311–1333. DOI: `10.1137/S0036139903422784`.

[59] Michael Hintermüller and Andreas Langer. 'Subspace correction methods for a class of nonsmooth and nonadditive convex variational problems with mixed L^1/L^2 data-fidelity in image processing'. In: *SIAM Journal on Imaging Sciences* 6.4 (2013), pages 2134–2173. DOI: `10.1137/120894130`.

[60] Michael Hintermüller and Carlos N. Rautenberg. 'On the density of classes of closed convex sets with pointwise constraints in Sobolev spaces'. In: *Journal of Mathematical Analysis and Applications* 426.1 (2015), pages 585–593. DOI: `10.1016/j.jmaa.2015.01.060`.

[61] Michael Hintermüller and Monserrat Rincon-Camacho. 'An adaptive finite element method in L^2-TV-based image denoising'. In: *Inverse Problems and Imaging* 8.3 (2014), pages 685–711. DOI: `10.3934/ipi.2014.8.685`.

[62] Michael Hintermüller and Georg Stadler. 'An infeasible primal-dual algorithm for total bounded variation-based inf-convolution-type image restoration'. In: *SIAM Journal on Scientific Computing* 28.1 (2006), pages 1–23. DOI: 10.1137/040613263.

[63] Kazufumi Ito and Karl Kunisch. *Lagrange Multiplier Approach to Variational Problems and Applications*. Advances in Design and Control. SIAM, 2008. DOI: 10.1137/1.9780898718614.

[64] Licheng Jiao and Jin Zhao. 'A survey on the new generation of deep learning in image processing'. In: *IEEE Access* 7 (2019), pages 172231–172263. DOI: 10.1109/ACCESS.2019.2956508.

[65] Andreas Langer. 'Automated parameter selection for total variation minimization in image restoration'. In: *Journal of Mathematical Imaging and Vision* 57.2 (Feb. 2017), pages 239–268. DOI: 10.1007/s10851-016-0676-2.

[66] Andreas Langer. 'Automated parameter selection in the L^1-L^2-TV model for removing Gaussian plus impulse noise'. In: *Inverse Problems* 33.7 (June 2017), page 074002. DOI: 10.1088/1361-6420/33/7/074002.

[67] Andreas Langer. 'Investigating the influence of box-constraints on the solution of a total variation model via an efficient primal-dual method'. In: *Journal of Imaging* 4.1 (2018), page 12. DOI: 10.3390/jimaging4010012.

[68] Andreas Langer. 'Locally adaptive total variation for removing mixed Gaussian–impulse noise'. In: *International Journal of Computer Mathematics* 96.2 (2019), pages 298–316. DOI: 10.1080/00207160.2018.1438603.

[69] Yuri Levin and Adi Ben-Israel. 'A Newton method for systems of m equations in n variables'. In: *Nonlinear Analysis: Theory, Methods & Applications* 47.3 (Aug. 2001), pages 1961–1972. DOI: 10.1016/S0362-546X(01)00325-X.

[70] Ryan Wen Liu, Lin Shi, Simon C. H. Yu and Defeng Wang. ‘Box-constrained second-order total generalized variation minimization with a combined $L^{1,2}$ data-fidelity term for image reconstruction’. In: *Journal of Electronic Imaging* 24.3 (2015), page 033026. DOI: 10.1117/1.JEI.24.3.033026.

[71] Julien Mairal. ‘Optimization with first-order surrogate functions’. In: *Proceedings of the 30th International Conference on Machine Learning*. Edited by Sanjoy Dasgupta and David McAllester. Volume 28. Proceedings of Machine Learning Research 3. 2013, pages 783–791. URL: https://proceedings.mlr.press/v28/mairal13.html.

[72] Mila Nikolova. ‘A variational approach to remove outliers and impulse noise’. In: *Journal of Mathematical Imaging and Vision* 20.1-2 (2004), pages 99–120. DOI: 10.1023/B:JMIV.0000011326.88682.e5.

[73] Mila Nikolova. ‘Minimizers of cost-functions involving nonsmooth data-fidelity terms. application to the processing of outliers’. In: *SIAM Journal on Numerical Analysis* 40.3 (2002), pages 965–994. DOI: 10.1137/S0036142901389165.

[74] Ricardo H. Nochetto, Kunibert G. Siebert and Andreas Veeser. ‘Theory of adaptive finite element methods: an introduction’. In: *Multiscale, nonlinear and adaptive approximation*. Edited by Ronald DeVore and Angela Kunoth. Springer, 2009, pages 409–542. DOI: 10.1007/978-3-642-03413-8_12.

[75] Jongho Park. ‘Additive Schwarz methods for convex optimization as gradient methods’. In: *SIAM Journal on Numerical Analysis* 58.3 (2020), pages 1495–1530. DOI: 10.1137/19M1300583.

[76] Thomas Pock, Daniel Cremers, Horst Bischof and Antonin Chambolle. ‘Global solutions of variational models with convex regularization’. In: *SIAM Journal on Imaging Sciences* 3.4 (2010), pages 1122–1145. DOI: 10.1137/090757617.

[77] Daniel Scharstein. *".flo" file format used for optical flow evaluation.* 2007. URL: https://vision.middlebury.edu/flow/code/flow-code/README.txt. Defined in Simon Baker, Daniel Scharstein, James P. Lewis, Stefan Roth, Michael J. Black and Richard Szeliski. 'A database and evaluation methodology for optical flow'. In: *International Journal of Computer Vision* 92.1 (2011), pages 1–31. DOI: 10.1007/s11263-010-0390-2.

[78] Claude E. Shannon. 'A mathematical theory of communication'. In: *The Bell System Technical Journal* 27.4 (1948), pages 623–656. DOI: 10.1002/j.1538-7305.1948.tb00917.x.

[79] Andrea Toselli and Olof Widlund. *Domain Decomposition Methods: Algorithms and Theory.* Volume 34. Springer Series in Computational Mathematics. Springer, 2005. DOI: 10.1007/b137868.

[80] Rüdiger Verfürth. *A Posteriori Error Estimation Techniques for Finite Element Methods.* Numerical Mathematics and Scientific Computation. Oxford University Press, 2013. ISBN: 978-0-19-967942-3.

[81] Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh and Eero P Simoncelli. 'Image quality assessment: from error visibility to structural similarity'. In: *IEEE Transactions on Image Processing* 13.4 (2004), pages 600–612.

[82] Christopher Zach, Thomas Pock and Horst Bischof. 'A duality based approach for realtime TV-L^1 optical flow'. In: *DAGM: Joint Pattern Recognition Symposium.* Volume 4713. Lecture Notes in Computer Science. Springer. Sept. 2007, pages 214–223. DOI: 10.1007/978-3-540-74936-3_22.