From the Institute of Mathematics and Image Computing
of the University of Lübeck
Director: Prof. Dr. Jan Modersitzki

Measure-Valued Variational Models with Applications in Image Processing

Dissertation
for Fulfillment of
Requirements
for the Doctoral Degree
of the University of Lübeck

from the Department of Computer Sciences

Submitted by
Thomas Goar Vogt
from Fulda

Lübeck, 2019

First referee: Prof. Dr. Jan Lellmann
Second referee: Prof. Dr. Dirk Lorenz
Date of oral examination: January 27, 2020
Approved for printing: January 27, 2020

Abstract

This thesis introduces the notion of measure-valued image processing models both as a mathematical framework for the study of images that take values in a space of probability measures, such as data from diffusion-weighted MRI, and as a global solution strategy for a broad class of image processing problems, such as image enhancement, inpainting and registration, that are often hard to solve globally optimally due to non-convexities. A novel total variation regularization energy is proposed for edge-preserving denoising and restoration of Q-ball data from diffusion-weighted MRI that naturally comes in the form of probability measures describing the direction of diffusivity of water in fibrous organic tissue. Furthermore, the global solution strategy that is known as *functional lifting* is adapted to a large class of manifold-valued imaging problems with an efficient discretization and implementation strategy based on finite elements. A generalization of functional lifting to energies that depend on second-order derivatives is proposed and its application to image registration problems is discussed. Finally, it is demonstrated that a large class of state-of-the-art functional lifting approaches including all of the previous models can be described through measure-valued image processing models whose mathematical description and analysis involves tools from functional analysis and the theory of optimal transport, especially dynamical models of optimal transportation.

Zusammenfassung

Im Zentrum dieser Arbeit stehen maßwertige Modelle für die Bildverarbeitung. Solche Modelle dienen einerseits als mathematischer Rahmen für die Handhabung von diffusionsgewichteten MRT-Datensätzen. Andererseits bilden sie die Grundlage einer globalen Lösungsstrategie für eine große Klasse von Bildverarbeitungsproblemen wie Bildverbesserung, -restauration oder -registrierung, die häufig aufgrund von Nichtkonvexität nur schwierig global optimal zu lösen sind. Sogenannte Q-Ball-Daten aus der diffusionsgewichteten Magnetresonanztomographie beschreiben die lokale Verteilung der Diffusionsrichtung von Wasser in faserigem organischem Gewebe mit Hilfe von Wahrscheinlichkeitsmaßen. In dieser Arbeit wird ein Regularisierer entwickelt und analysiert, der – analog zur in der Bildverarbeitung häufig verwendeten Totalvariationsnorm – für die strukturerhaltende Rauschentfernung von Q-Ball-Daten geeignet ist. Weitergehend wird die als *Funktionalanheben* (engl. *functional lifting*) bekannte globale Lösungsstrategie so angepasst, dass sie auf eine große Klasse von Bildverarbeitungsproblemen mit Werten in Mannigfaltigkeiten anwendbar ist. Die numerische Implementierung dieses Ansatzes basiert auf einer effizienten Diskretisierung mittels finiter Elemente. Im nächsten Schritt wird das Funktionalanheben auch für Probleme mit Ableitungen höherer Ordnung verallgemeinert, was vor allem in der Bildregistrierung eine interessante Anwendung findet. Den Abschluss der Arbeit bildet die Ausformulierung eines mathematischen Models, das alle zuvor genannten Strategien des Funktionalanhebens in einer gemeinsamen Sprache vereint, der maßwertige Funktionen und Konzepte aus den Bereichen der Funktionalanalysis und des dynamischen optimalen Transports zugrunde liegen.

Acknowledgments

This thesis evolved, first and foremost, under the supervision of Prof. Jan Lellmann. He was not only permanently available for my questions or for elaborate discussions and provided invaluable advice, but he also introduced me, coming from a different field, to the fascinating world of image processing and to the scientific community. I am deeply grateful for the time we spent together, for his efforts and encouragements.

I warmly thank my present and former co-workers and students at the Institute of Mathematics and Image Computing (MIC) and Fraunhofer MEVIS in Lübeck who contributed to a peaceful and friendly atmosphere of openness. Special thanks go to Hari, Kai, Daniel, Yury, Rosa, Caterina and Benjamin for innumerable both scientific and private conversations and for making office work such an enjoyable experience.

I also like to mention Thomas Möllenhoff, Bernhard Schmitzer, Emanuel Laude and Daniel Cremers from the Technical University in Munich, who contributed to my research work through fruitful discussions.

Finally, I express my gratitude to my family and friends, especially my parents, who unconditionally supported me at all times.

Notation and Abbreviations

Abbreviations

ADMM	Alternating Direction Method of Multipliers (optimization algorithm)
CB	Chromaticity Brightness (color model)
CSD	Constrained Spherical Deconvolution
CT	Computer Tomography
DR	Douglas-Rachford (optimization algorithm)
DTI	Diffusion Tensor Imaging
DW-MRI	Diffusion-weighted Magnetic Resonance Imaging
EBSD	Electron Backscatter Diffraction
EMD	Earth Mover's Distance
GPU	Graphics Processing Unit
HSV	Hue Saturation Value (color model)
InSAR	Interferometric Synthetic Aperture Radar
MRF	Markov Random Field
MRI	Magnetic Resonance Imaging
NP-hard	Problem that is at least as hard in complexity as the hardest problem in the class NP
OT	Optimal Transport
ODF	Orientation Distribution Function
QBI	Q-Ball Imaging
PDE	Partial Differential Equation
PDHG	Primal-Dual Hybrid Gradient (optimization algorithm)
PET	Positron-Emission Tomography
RGB	Red Green Blue (color model)
ROF	Rudin-Osher-Fatemi (L^2-L^1) model for image restoration
SSD	Sum of Squared Distances (data discrepancy term)
TGV	Total Generalized Variation (seminorm or regularizer)
TV	Total Variation (seminorm or regularizer)

Sets

$\mathcal{B}(X)$	Borel σ-algebra in X
$\mathrm{BV}(\Omega, \mathbb{R}^n)$	(Vectorial) functions of bounded variation on Ω
$C(Q)$	Continuous selections with respect to the set-valued map Q
$C(X)$	Same as $C(X, \mathbb{R})$
$C_b(X)$	Same as $C_b(X, \mathbb{R})$
$C_c(X)$	Same as $C_c(X, \mathbb{R})$
$C_0(X)$	Same as $C_0(X, \mathbb{R})$
$C(X, Y)$	Continuous maps from X to Y
$C_b(X, Y)$	Bounded continuous maps from X to Y
$C_c(X, Y)$	Continuous maps from X to Y with compact support
$C_0(X, Y)$	Closure of $C_c(X, Y)$ with respect to supremum norm
$C^k(\Omega, Y)$	k-times continuously differentiable functions from Ω to Y
$C^k_c(\Omega, Y)$	Compactly supported functions in $C^k(\Omega, Y)$
$\partial\Omega$	Set of boundary points of the set Ω
$\ker(A)$	Algebraic kernel of the linear operator A: vectors x with $Ax = 0$
$\mathrm{KR}(X)$	Kantorovich-Rubinstein space, completion of $\mathfrak{M}_0(X)$ with respect to $\|\cdot\|_{\mathrm{KR}}$
$L^p(\Omega)$	Lebesgue p-integrable functions on Ω
$L^p_\mu(\Omega)$	Functions that are p-integrable with respect to the measure μ
$L^\infty_w(\Omega, V)$	Weakly measurable and essentially bounded functions from Ω to V
$L^\infty_{w*}(\Omega, V^*)$	Weakly* measurable and essentially bounded functions from Ω to V^*
$\mathrm{Lip}(X)$	Real-valued Lipschitz functions on X
$\mathrm{Lip}(X, Y)$	Lipschitz functions from X to Y
$\mathrm{Lip}_0(X, Y)$	Lipschitz functions on X that vanish at a given point x_0
$\mathfrak{M}(X)$	Finite signed Radon measures on X
$\mathfrak{M}_0(X)$	Finite signed Radon measures on X with zero mean
$\mathfrak{M}(X, \mathbb{R}^n)$	Finite vectorial Radon measures on X
$\mathbb{N}$	Natural numbers (integers) $\{1, 2, 3, \dots\}$
$\mathbb{N}_0$	Natural numbers (integers) including zero $\{0, 1, 2, 3, \dots\}$
$N_p\mathcal{M}$	Normal space of submanifold $\mathcal{M}$ at p
$\mathcal{P}(X)$	Borel probability measures on a given topological space X
$\Pi(\mu, \nu)$	Measures in $\mathcal{P}(X \times X)$ with marginals μ and ν
$\mathbb{R}^d$	d-dimensional Euclidean space
$\mathbb{R}^{n,k}$	$n \times k$ matrices with real entries
$\mathbb{S}^1$	Circle as a compact manifold in $\mathbb{R}^2$

$\mathbb{S}^n$	n-dimensional sphere in $\mathbb{R}^{n+1}$ as a compact manifold
$SE(n)$	n-dimensional group of rigid motions
$SO(n)$	Special orthogonal group of $n \times n$ matrices
$T_p\mathcal{M}$	Tangent space of $\mathcal{M}$ at p

Operators and Functions

$\mathbb{1}_A$	Measure-theoretic indicator function of A; 1 on A and 0 otherwise
$\mathbb{1}_u$	Indicator function of the subgraph of the function u
$\operatorname{adj}_s \xi$	Vector of $s \times s$ minors of the matrix ξ
$\operatorname{argmin}_x F(x)$	Set of minimizers of F or the unique minimizer if it exists
$d_{\mathcal{M}}(p, q)$	Geodesic distance on given manifold $\mathcal{M}$
δ_x	Dirac unit point measure concentrated at x
δ_C	Convex indicator function of the set C; 0 on C and ∞ otherwise
$\partial\Omega$	Set of boundary points of the set Ω
$\partial_i^k u$	k-th partial derivative of u in direction i
Du	Distributional derivative (measure) of u
∇u	Gradient/Jacobian of u
$\nabla_x u$	Gradient/Jacobian of u with respect to the variable x
$\operatorname{div} u$	Divergence of the vector field u
$\operatorname{Div} u$	(Column-wise) divergence of the matrix-valued function u
Δu	(Channel-wise) Laplacian differential operator applied to u
epi_f	Epigraphical set of the function f
$\exp_p$	Riemannian exponential map at p
$\mathcal{H}^m(A)$	Mass of the set A with respect to the m-dimensional Hausdorff measure
$\mathcal{L}^n(A)$	Mass of the set A with respect to the n-dimensional Lebesgue measure
$\log_p$	Riemannian logarithmic (inverse exponential) map at p
$\pi_i \gamma$	i-th marginal Distribution of Γ
$\operatorname{proj}_C(x)$	Orthogonal projection of x onto the convex closed set C
$\operatorname{prox}_f(\bar{x})$	Proximal mapping of the convex function f
σ_C	Support function of the set C; convex conjugate of δ_C
$\operatorname{TV}(u)$	Total variation seminorm of u
$W_p(\mu, \nu)$	p-Wasserstein distance between given probability measures μ and ν

Expressions

$A \setminus B$	Set difference: set of elements that are in A, but not in B
$A \subset B$	Each element of A is contained in B, the statement is true in particular if $A = B$
$\mathring{\Omega}$	The (topological) interior of Ω
$n \wedge m$	Shorthand for $\min\{n, m\}$
$\binom{n}{k}$	Binomial coefficient, number of ways to choose k out of a set of n
$n!$	Factorial: $0! := 1$ and $(n+1)! := (n+1) \cdot n!$
$F \colon U \rightrightarrows \mathbb{R}^m$	A set-valued function: $F(x) \subset \mathbb{R}^m$ for each $x \in U$
$f^{-1}(A)$	Preimage of a set A under a function f
$f^*(y)$	Convex conjugate of the function f
$T_*\mu(A)$	Push-forward measure $\mu(T^{-1}(A))$ of μ under T
X^*	Topological dual space of a topological vector space Y
x_k	k-th component of the vector x
x^i	i-th element in the sequence (x^k)
$f^k \rightharpoonup g$	f^k converges weakly (in the weak topology) to g
$\mu^k \overset{*}{\rightharpoonup} \nu$	μ^k converges weakly* (in the weak* topology) to ν
μ^+, μ^-	Hahn decomposition of the signed measure μ
$\nu \ll \mu$	ν is absolutely continuous with respect to μ
$d\nu/d\mu$	Radon-Nikodym derivative of ν with respect to μ
$\lvert\mu\rvert$	Total variation measure of signed measure μ
$\lVert x\rVert$	Euclidean norm $\lVert x\rVert_2$ of vector x
$\lVert x\rVert_p$	l^p-norm of vector x
$\lVert\mu\rVert$	Total variation measure of vectorial measure μ
$\lVert x\rVert_Y$	Norm with respect to given normed space Y
$\lVert A\rVert_F$	Frobenius norm of matrix A
$\lVert A\rVert_{\sigma,\infty}$	Matrix operator (or spectral) norm
$\lVert A\rVert_{\sigma,1}$	Matrix nuclear norm, Schatten-1-norm
$\lVert f\rVert_\infty$	Supremum norm of the function f
$\lVert\mu\rVert_{\mathfrak{M}}$	Total variation norm of measure μ
$[x]_{\mathrm{Lip}}$	Lipschitz seminorm
$\langle x, y\rangle$	Dual pairing or scalar product between x and y
$\int_\Omega f(x)\,dx$	Integral of f with respect to the Lebesgue measure over the set Ω
$\int_\Omega f\,d\mu$	Integral of f with respect to the (signed) measure μ over the set Ω
$\int_\Omega f(x,y)\,\mu(dy)$	Integral of $f(x,\cdot)$ with respect to μ for x fixed

Contents

1 Introduction

In this thesis, we will be concerned with the mathematical modeling, approximation and numerical solution of certain classes of variational image processing problems

$$\inf_{u \in \mathcal{U}} F(u) \tag{1.1}$$

with $\mathcal{U}$ a space of *images* $u \colon \Omega \to \Gamma$, that are defined on an open and bounded *image domain* $\Omega \subset \mathbb{R}^d$. In particular, we will consider problems where the images u take values in a metric space Γ of *features*. The features might be discrete, such as classification labels, or Γ might be a scalar interval of gray values, a vectorial linear or nonlinear continuous range, such as Euclidean space in case of RGB color values or displacement vectors, or a Riemannian manifold in case of more abstract features, such as chromaticity.

However, a pivot between the different approaches considered in this work will be the case where u takes values in an infinite-dimensional space of *measures*. Namely, we will approximate the previous cases with a measure-valued setting, discuss benefits and difficulties of such an approximation, and show that the measure-valued setting is a mathematically sound framework for the processing of so-called diffusion-weighted magnetic resonance imaging or, more precisely, Q-ball data (Chap. 3).

In all cases, the starting point will be a model where the objective functional F is of integral form, depending pointwise on u and, possibly, on one or more of its derivatives:

$$F(u) := \int_\Omega f(x, u(x), \nabla u(x), \dots)\, dx. \tag{1.2}$$

The integrand f will be assumed to depend on the derivatives of u in a convex way, while the dependence on x and $u(x)$ is allowed to be non-convex. Computational minimization strategies for (1.2) typically start from a first guess that is iteratively improved by varying it according to the local energy landscape, e.g., based on the functional's gradient. Local in nature as they are, these strategies can get stuck in local minima and often fail to find the globally optimal minimizers. In recent years [Ish03; Zac+08; Poc+10; SGC11; Lel+13a; Bae+14], the convex approximation of problem (1.2) through so-called *functional lifting* has been investigated as an alternative solution strategy that allows to apply

efficient *convex* optimization methods and has been proved in some cases to yield *global* minimizers of the original energy (see Sect. 1.2.2). In this thesis, we apply the functional lifting idea to new classes of manifold-valued (Chap. 4) and second-order (Chap. 5) image processing problems with a new efficient discretization strategy. We demonstrate in Chap. 6 that a large class of these functional lifting approaches can be described through *measure-valued* image processing models and investigate their mathematical properties.

After a short motivational section where we illustrate our approach on the basis of explicit applications (Sect. 1.1), we introduce the basic notions and problems that this thesis will be concerned with and discuss related work (Sect. 1.2) before summarizing our contribution (Sect. 1.3).

1.1 Motivation

One specific application of our proposed model is the problem of *image registration*, where a given template image $I_t \colon \mathbb{R}^d \to \mathbb{R}$ is deformed to match a given reference image $I_r \colon \mathbb{R}^d \to \mathbb{R}$ in a bounded region of interest $\Omega \subset \mathbb{R}^d$ [Mod04; Gos05; FM08; Mod09]. As an example, consider the **variational** image registration model

$$\inf_{u \in \mathcal{U}} F(u), \quad F(u) := \frac{1}{2} \int_\Omega (I_t(x + u(x)) - I_r(x))^2 dx + \frac{\lambda}{2} \int_\Omega \|\Delta u(x)\|^2 dx, \tag{1.3}$$

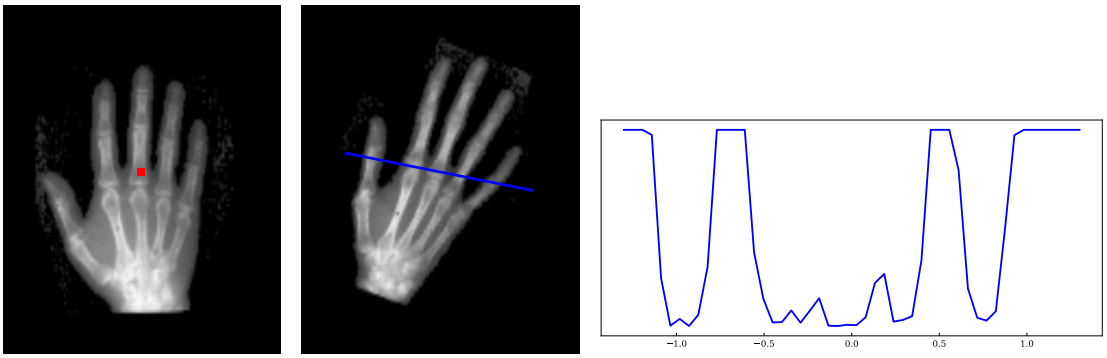
where $\mathcal{U}$ is a suitable space of *displacement vector fields* $u \colon \Omega \to \mathbb{R}^d$, $\lambda > 0$ is a regularization parameter and Δ is the Laplacian differential operator, applied componentwise:

$$[\Delta u(x)]_k := \sum_{i=1}^{d} \partial_i^2 u_k(x), \ k = 1, \dots, d. \tag{1.4}$$

The first integral in (1.3), the *data term*, describes the pointwise match of image intensities between I_r and the deformed template image I_t. The *regularization term*, the second integral in (1.3), measures the "smoothness" of the displacement u. The regularization parameter λ ensures that a solution to the minimization problem (1.3) is a suitable balance between data discrepancy and smoothness of deformation according to the specific application.

The variational image registration model (1.3) is usually solved numerically by employing local optimization algorithms such as gradient descent or quasi-Newton methods [Mod09]. However, the data term in (1.3) is highly non-convex in general (Fig. 1.1) so that local optimization methods often fail to find globally optimal solutions.

An alternative solution strategy for non-convex variational problems is the method of



Reference image I_r Template image I_t $\frac{1}{2}|I_r(x) - I_t(x+z)|^2$

Figure 1.1: In image registration, the task is to align a pair of images: Given $\Omega \subset \mathbb{R}^2$ open and bounded and two images $I_r, I_t \colon \mathbb{R}^2 \to \mathbb{R}$ **(left and center)**, the goal is to find a displacement $u(x) = z \in \mathbb{R}^2$ for each $x \in \Omega$ such that $x \mapsto I_r(x)$ and $x \mapsto I_t(x+u(x))$ are "similar" in the region Ω. In variational formulations of image registration, the similarity is expressed in a data discrepancy term, such as the sum of squared distances of intensity values **(right)**. For a fixed reference point $x \in \Omega$ (red square in the left image), the data term is typically a highly non-convex function of the displacement $z = u(x)$ (blue line in the center image). Our functional lifting framework approximates the non-convex variational formulation by a convex formulation in a space of measure-valued images. In contrast to previous approaches, our framework is applicable to models with second-order regularization, which is commonly used in image registration. Image data from [Mod09].

functional lifting, where the non-convex objective functional F in (1.3) is approximated by a convex functional $\mathcal{F}$ that is defined over a higher-dimensional set $\mathcal{V}$. The auxiliary minimization problem

$$\inf_{v \in \mathcal{V}} \mathcal{F}(v) \tag{1.5}$$

can be solved globally using efficient and robust convex optimization algorithms. The lifted functional $\mathcal{F}$ and the admissible set $\mathcal{V}$ are chosen in such a way that a minimizer v^* of the lifted problem (1.5) is a suitable approximation to a global minimizer u^* of the original problem (1.3).

In this work, we consider functional lifting strategies that replace the "hard" decision for a displacement vector $u(x) \in \mathbb{R}^d$ in each $x \in \Omega$ by a "soft" probability measure $v(x) \in \mathcal{P}(\mathbb{R}^d)$ over the set of all possible displacement vectors. The admissible set $\mathcal{V}$ in the lifted problem (1.5) is chosen to be a suitable space of **measure-valued** functions $v \colon \Omega \to \mathcal{P}(\mathbb{R}^d)$. Our proposed measure-valued functional lifting framework is applicable to **image processing** problems that depend on first- or second-order derivatives of the unknown

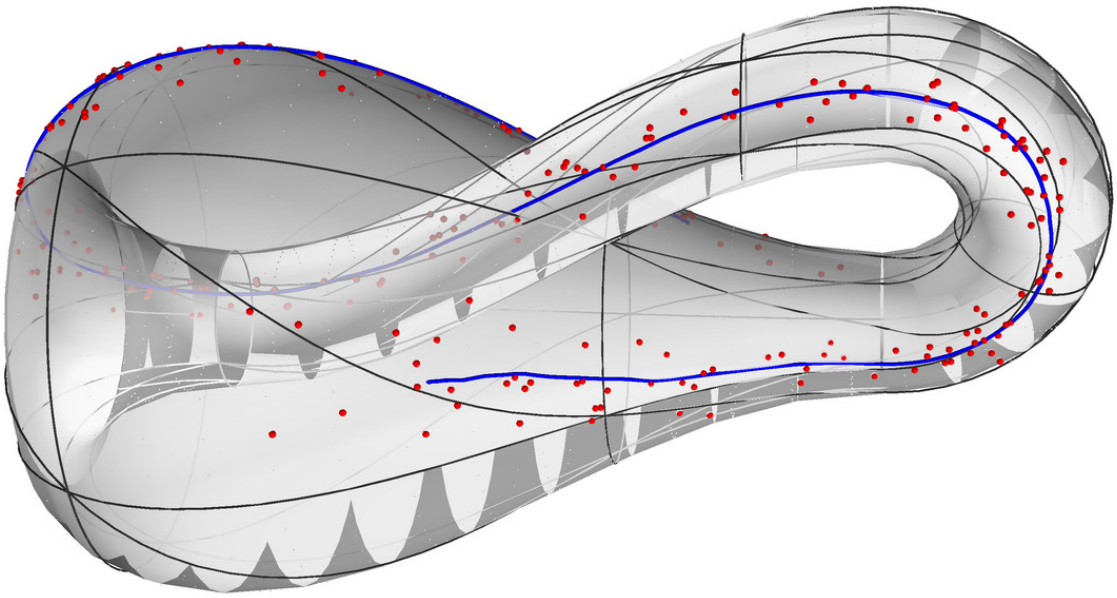


Figure 1.2: Images whose values underly symmetry or periodicity constraints can often be interpreted as taking values in a manifold. In this case, the admissible set fails to be convex in general so that variational models for denoising are non-convex. Local optimization algorithms on manifolds might get stuck in local minima. Instead, our functional lifting framework approximates the non-convex variational problem by a convex problem in a higher-dimensional space of measure-valued images to which efficient convex optimization algorithms can be applied. Here, a noisy one-dimensional signal or image on the Klein bottle **(red)** is denoised **(blue)** using our functional lifting framework.

function, as is the case in the image registration problem (1.3) (Chap. 5). Except for a few cases [LL18; SG19], previous work on functional lifting is only applicable to variational models that only depend on first-order derivatives of the unknown function u,

$$\inf_{u\in\mathcal{U}} F(u), \quad F(u) := \int_\Omega f(x, u(x), \nabla u(x))\,dx, \tag{1.6}$$

with $f\colon \Omega \times \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ convex in the last, but non-convex in the first two variables.. In many functional lifting models, the *range* of the admissible functions u is required to be scalar. Our proposed framework is compatible with first- and second-order models with vectorial- and, more generally, manifold-ranges, i.e., where $\mathcal{U}$ is a set of functions $u\colon \Omega \to \mathcal{M}$ for a smooth Riemannian manifold $\mathcal{M}$ (Fig. 1.2).

If we are concerned with *manifold-valued* images, the admissible set $\mathcal{U}$ of functions $u\colon \Omega \to \mathcal{M}$ in (1.6) is often non-convex. While local optimization algorithms for manifold-

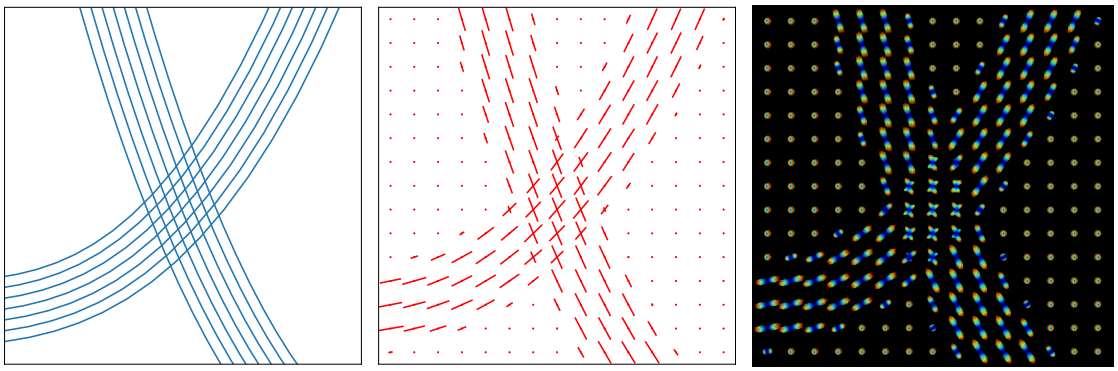


Figure 1.3: In diffusion-weighted MRI and, especially, Q-ball imaging, measure-valued images are used to describe mathematically the diffusivity of water in fibrous tissue. However, the signals are prone to noise and choosing a meaningful notion of similarity between neighboring signals (measures) is crucial in the process of restoration. In this work, optimal transport metrics are used to define regularity or smoothness of measure-valued images which can be used to denoise Q-ball images (Chap. 3). **Left:** 2-D fiber phantom of two crossing fiber bundles. **Center:** Peak directions on a 15×15 grid, derived from the phantom and used for the generation of synthetic DW-MRI data. **Right:** An ODF reconstruction of DW-MRI data represents fiber orientation using probability measures at each point, which allows to accurately recover fiber crossings in the center region.

valued problems are available [AMS09], they can get stuck in local minima. Our proposed functional lifting framework can be applied to a larger class of manifold-valued variational models than previous functional lifting frameworks [Lel+13b] and comes with a more efficient discretization. It is compatible with a large class of convex regularization terms and applicable to manifold-valued denoising and inpainting problems (Chap. 4).

Beyond the auxiliary role played by measure-valued functions in functional lifting frameworks, they are central to the processing of so-called *Q-ball images* in diffusion-weighted (DW) magnetic resonance imaging (MRI). In DW-MRI, the diffusivity of water in tissues that exhibit fibrous microstructures, such as axons in cerebral white matter, is measured in order to extract information about the fiber architecture in the living organism. For many medical applications, this is valuable information. Mathematically, these measurements are commonly represented as orientation distribution functions (ODFs), so-called Q-ball images (QBI): For each location x in the region of interest $\Omega \subset \mathbb{R}^3$, a probability measure $u(x) \in \mathcal{P}(\mathbb{S}^2)$ describes the distribution of diffusion directions over the two-dimensional sphere $\mathbb{S}^2$, the set of all possible diffusion directions (Fig. 1.3).

A difficulty in the processing of Q-ball images is the high level of noise that distorts

the signals during the acquisition process. For the denoising of image data, **variational models** are a common solution strategy: On the suitably chosen set $\mathcal{U}$ of admissible images $u\colon \Omega \to \mathcal{P}(\mathbb{S}^2)$, an energy functional $F\colon \mathcal{U} \to \mathbb{R}$,

$$F(u) := D(u, f) + \lambda R(u), \tag{1.7}$$

is defined. The data term $D(u, f)$ describes the closeness of the solution candidate u to the measured signal $f\colon \Omega \to \mathcal{P}(\mathbb{S}^2)$ while the regularization term $\lambda R(u)$ with $\lambda > 0$ measures the spatial regularity of u. As in variational models for image registration (1.3), a solution of the denoising problem is expected to be a minimizer of the minimization problem

$$\inf_{u \in \mathcal{U}} F(u). \tag{1.8}$$

As Q-ball images are measure-valued, defining spatial regularity is a non-trivial task. Our proposed approach starts by defining a measure of similarity between neighboring signals, which are probability measures on the metric space of directions $\mathbb{S}^2$. Distances between probability measures that take the metric structure of the underlying space into consideration are defined and investigated in the theory of optimal transport (see Sect. 1.2.3). We employ optimal transport distances to generalize the total variation (TV) seminorm, a popular regularization energy in image processing applications, to measure-valued functions. The proposed TV seminorm is then used as a measure of regularity for Q-ball images. When employed in the variational model (1.7) it preserves structural features such as edges while removing noise substantially (Chap. 3).

1.2 Related Work

The subject of this thesis is interdisciplinary in character and, naturally, we will encounter notions and results from numerous fields of research along the way:

- As illustrated in the previous motivational section, the main applications for our mathematical framework will be from the field of **mathematical signal and image processing**. We briefly introduce the reader to the field and give an overview of image denoising, -segmentation, and -registration as well as manifold-valued- and Q-ball images as far as these areas are relevant to this thesis (Sect. 1.2.1).

- The main contribution of this thesis is a novel measure-valued **functional lifting** framework (Chap. 6) that is applicable to a large class of scalar, vectorial, manifold-valued (Chap. 4), and first- and second-order (Chap. 5) problems. We introduce

the main concepts and discuss recent work on scalar, vectorial, sublabel-accurate, and higher-order models (Sect. 1.2.2).

- At the heart of our proposed models are methods from the mathematical theory of **optimal transport** that define a natural geometry on spaces of measures and therefore help in defining regularity of measure-valued functions. After a brief introduction to optimal transport, we discuss the main three aspects that are relevant to this work: the Kantorovich-Rubinstein formulation of optimal transport distances, the dynamical formulation of optimal transport due to Benamou and Brenier, and the more recent notion of (harmonic) functions with values in the Wasserstein space (Sect. 1.2.3).

- Since almost every image processing model discussed in this work comes in variational form, usually defined over an infinite-dimensional space of functions, notions from **variational calculus** are encountered in almost every chapter. Motivated by the direct method and lower semicontinuous relaxations of functionals, we discuss gradient Young measures, Cartesian currents, and the method of calibrations (Sect. 1.2.4), all of which play an important role in functional lifting approaches.

- We implemented most of our proposed models numerically in order to illustrate the applicability to real-world image processing problems. Since the models come in a convex variational form, we give a brief overview of recent algorithms from **convex finite-dimensional optimization**, especially non-smooth primal-dual splitting algorithms (Sect. 1.2.5).

1.2.1 Mathematical Signal and Image Processing

The main applications for the mathematical framework proposed in this work will be from the field of image processing. The abstract mathematical formulation of our framework on spaces of functions $u\colon \Omega \to \Gamma$, with $\Omega \subset \mathbb{R}^d$ open and bounded and Γ a metric space, allows to apply the proposed framework to almost arbitrary kinds of images, where we refer to an *image* as a physical signal that is acquired, modeled or interpreted in a *spatially coherent* way. This definition of an image comprises the common notion of an optical color or gray-value representation of a physical object or phenomenon, such as a photography or a painting. But it also includes more abstract signals, in particular biological, geological, ecological and meteorological features, such as acidity, density, radioactivity, height, temperature, humidity, plant covering, land use, geological reserves or seismic activity. These signals might be given on a bounded two-dimensional domain

$\Omega \subset \mathbb{R}^2$, such as a geographical area or a petri dish, or inside a three-dimensional volume $\Omega \subset \mathbb{R}^3$, such as the earth's atmosphere, oceanic covering and crust, or the human body.

Depending on the application, the mathematical representation of a single signal $u(x)$, $x \in \Omega$, is constrained to a specific set of values Γ: scalar or intensity values from a continuous interval $\Gamma := [a, b] \subset \mathbb{R}$, colors represented in a bounded three dimensional color space $\Gamma \subset \mathbb{R}^3$, classification labels from a finite or infinite discrete set $\Gamma \subset \mathbb{N}$, high-dimensional spectral multichannel information $\Gamma \subset \mathbb{R}^N$ or periodically, symmetrically or geometrically constrained values from a topological or geometric manifold $\Gamma = \mathcal{M}$.

For an overview of the highly active field, we point to introductions [VFG09; Sch+09; BL11; VG16; LH17; Sun17; BL18], comprehensive surveys, and handbooks [AK06; Pin14; Sch15; Jäh19] on mathematical image processing that appeared in recent years.

In the following paragraphs, we briefly introduce several image processing tasks and techniques that are relevant to the results presented in this work. An emphasis will be on variational (energy minimization) frameworks and we will point out the role of non-convexity and local minimizers in each case.

Image enhancement, denoising and restoration When a given image is distorted by noise, stains, or blur, lacks smoothness, sharpness, or contrast, the assumptions on the spatial coherence of the underlying signal can be employed to enhance or restore signals in damaged regions. Usually, restoration problems are ill-posed even if information about the location and type of distortion as well as about the underlying physical data acquisition process is available. In those cases, a common approach is to choose an image that is close to the acquired signal while being sufficiently regular according to a chosen *regularization* energy. We will discuss this approach in a bit more detail, as it is prototypical for many other image processing problems and, in particular, for the class of variational problems that our proposed framework applies to.

Given a distorted signal $f\colon \Omega \to \Gamma$, the restored image $u\colon \Omega \to \Gamma$ is chosen to satisfy an optimality criterion of the form

$$\inf_{u\colon \Omega \to \Gamma} F(u), \quad F(u) := D(u; f) + \lambda R(u), \tag{1.9}$$

where D describes the closeness to the measured signal and R is concerned with the regularity of u. This method, especially the delicate question of choosing the regularization parameter $\lambda > 0$, has its roots in the Tikhonov-Phillips regularization for ill-posed problems [Phi62; Tik63; TA77].

In the mathematical image processing community, the name Tikhonov regularizer is

frequently used for the particular choice of a squared image gradient,

$$R(u) := \frac{1}{2} \int_\Omega \|\nabla u(x)\|^2 \, dx. \tag{1.10}$$

Moreover, a generic choice for the data term D in case of gray-value images is the sum of squared differences (SSD)

$$D(u; f) := \frac{1}{2} \int_\Omega (u(x) - f(x))^2 \, dx. \tag{1.11}$$

Since images commonly contain signals from more than one physical object, they exhibit sharp edges between neighboring structures while being smooth away from these edges – they are only *piecewise smooth*. The Tikhonov regularizer heavily penalizes outliers in the gradient of u and, because of that, this restoration method will smooth out edges in the input signal f as an undesirable side-effect.

David Mumford and Jayant Shah [MS89] therefore explicitly added a dependence on the $(d-1)$-dimensional edge or boundary set B that is assumed to be compactly contained in the d-dimensional image domain Ω:

$$F(u, B) = \frac{1}{2} \int_{\Omega \setminus B} (u(x) - f(x))^2 \, dx + \frac{\lambda_1}{2} \int_{\Omega \setminus B} \|\nabla u(x)\|^2 \, dx + \lambda_2 \mathcal{H}^{d-1}(B). \tag{1.12}$$

As in the Tikhonov-Phillips regularization method, the first term of this functional encodes closeness to the measured signal, the other two terms are concerned with the regularity of the candidate u: large gradients and long edges are penalized proportionally to the chosen regularization parameters $\lambda_1, \lambda_2 > 0$. However, the dependence on a *set* B introduces a problem of parametrization as well as a problem of non-convexity.

The problem of parametrization in the Mumford-Shah functional is tackled by the Ambrosio-Tortorelli approximation [AT90], where the indicator function of B is approximated pointwise by a continuous function so that the resulting expressions Γ-converge [GF75; MM77] to the original functional. These approximate functionals are still non-convex, but local minimizers can be computed via gradient descent. Moreover, by the definition of Γ-convergence, every sequence of minimizers of the approximate functionals converges to a minimizer of the original functional. We refer the reader to the papers [Fus03; Foc16] for an overview of the Mumford-Shah problem.

On the other hand, nonsmooth convex approximations of the Mumford-Shah functional in the scalar [Poc+09b] and vectorial [SCC12] case are based on a sufficient condition for minimizers of the Mumford-Shah functional known as *calibration method* [ABD03]

(see Sect. 1.2.4). Even though the convex relaxation has been demonstrated to avoid local minima compared to the Ambrosio-Tortorelli method in practice, no explicit algorithm is known that provably produces minimizers of the Mumford-Shah functional from minimizers of the convex relaxation yet [Cha01; Poc+09b] – an issue that can also be observed in some of our proposed lifting approaches.

For the restoration of images with sharp edges, there are convex models that are based on the total variation seminorm [AFP00; Cha+10] for so-called functions of *bounded variation* $u\colon \Omega \to \mathbb{R}$,

$$\mathrm{TV}(u) := \sup \left\{ \int_\Omega \langle u, \operatorname{div} \phi \rangle \, dx : \phi \in C_c^1(\Omega, \mathbb{R}^d), \|\phi(x)\| \leq 1 \right\}. \tag{1.13}$$

Compared with the quadratic Tikhonov regularizer, the total variation seminorm grows linearly in ∇u for smooth functions $u\colon \Omega \to \mathbb{R}$:

$$\mathrm{TV}(u) = \int_\Omega \|\nabla u(x)\| \, dx, \tag{1.14}$$

and it is well-defined for piecewise smooth functions with jump discontinuities. The following three TV-based models are most relevant to this work: The *Rudin-Osher-Fatemi (ROF)* model [ROF92],

$$F(u) = \frac{1}{2} \int_\Omega (u(x) - f(x))^2 \, dx + \lambda \, \mathrm{TV}(u), \tag{1.15}$$

is known to preserve edges, but also to enforce piecewise constant solutions resulting in an undesired staircasing effect. The *Huber regularizer* [Wer+09] is based on an estimation procedure in robust statistics [Hub64] that – similar to Winsorizing, a data clipping procedure in statistics [HR09] – avoids being dominated by outliers while still being close to the mean-unbiased mean-square estimation procedure:

$$R(u) = \int_\Omega \phi_\alpha(\nabla u) \, dx, \tag{1.16}$$

$$\phi_\alpha(\xi) := \begin{cases} \frac{\|\xi\|^2}{2\alpha} & \text{if } \|\xi\| \leq \alpha, \\ \|\xi\| - \frac{\alpha}{2} & \text{if } \|\xi\| > \alpha, \end{cases} \tag{1.17}$$

The Huber regularizer uses a quadratic Tikhonov term for small gradients and a linear growth at infinity just as TV, which allows to preserve most edges as well as smooth transitions. Finally, in [BKP10], the *total generalized variation (TGV)* has been introduced as a regularizer that enforces piecewise smooth functions by making use of higher-order

derivatives.

Since many regularizers come in integral form and this work is concerned with integral energy minimization models, we will frequently come back to these regularizers and variants thereof.

Image segmentation, partitioning and multi-class labeling The grouping of shapes and features arranged in an image according to given criteria is called image segmentation, -partitioning or -labeling. While there are models, especially statistical in nature, that are able to apply prior semantic knowledge to this task [Guo+18], we will restrict to models that only use the measured signal values $f(x)$ together with assumptions on spatial coherence of regions to identify each point x as belonging to one of several objects or classes. Since this kind of image segmentation model is a prototypical application of functional lifting methods (Sect. 1.2.2) and, historically, many functional lifting techniques were developed in the context of image segmentation problems, we briefly give an overview of relevant image segmentation techniques in the following.

In the most generic case of images segmentation, a foreground region is to be separated from the background, a task known as 2-class labeling or binarization. Multi-Class labeling, on the other hand, corresponds to approximating the input image by a piecewise constant image. In fact, the Mumford-Shah functional (1.12) is known to produce piecewise constant images in the boundary case $\lambda_1 \to \infty$. The corresponding problem (1.12) with the second term removed and with a finite set of admissible values Γ, $u(x) \in \Gamma$, is called *Potts model* in the discrete setting, and its solution is known to be NP-hard in general [BVZ01].

For the 2-class labeling case, efficient algorithms are available: The Chan-Vese model [CV01] parametrizes the boundary curve in the two-phase piecewise constant Mumford-Shah energy as the zero level set of a Lipschitz function and applies a fast solution strategy for this kind of problem known as the level-set method [OS88]. Similarly, the Ambrosio-Tortorelli method [AT90] can be adapted to the piecewise constant Mumford-Shah energy. However, both methods are local in nature and might get stuck in local minima. In [CEN06], it was shown that the binarization problem can provably be solved globally by thresholding the solution of a convex denoising model.

There are several convex variational models for the solution of the multi-class labeling problem [Zac+08; Poc+09a; Lel+09; LBS09; BT09]. Contrary to the 2-class case, thresholding the solutions to take discrete values does not yield the same result as solving the same variational problems with values constrained to a discrete set. Instead, it was possible to prove bounds on the relative error introduced by the convex relaxation for

specific thresholding or *rounding* techniques [LLS13b]. Many of these convex relaxation approaches belong to the larger class of functional lifting approaches (see Sect. 1.2.2).

Intensity-based image registration In *image registration* or image matching, sometimes referred to as large-displacement optical flow, multiple images that represent the same scene from different sensors, times, depths or viewpoints are aligned by applying transformations of the image domain. In this work, we will restrict ourselves to the registration of pairs of images based on their pointwise *intensities* instead of salient geometric features, such as edges. As illustrated in the motivational Sect. 1.1, one application of our mathematical framework are image registration models in which a *template* or *moving* image $I_t \colon \mathbb{R}^d \to \mathbb{R}^k$ is deformed using a *deformation* $y \colon \Omega \to \mathbb{R}^d$ to match a *reference* image $I_r \colon \mathbb{R}^d \to \mathbb{R}^k$ in a bounded region of interest $\Omega \subset \mathbb{R}^d$ according to a pointwise distance measure

$$D(u) := \int_\Omega d(I_t(x + u(x)), I_r(x))\, dx, \tag{1.18}$$

where $u(x) := y(x) - x$ is the *displacement* vector field. The distance measure $d(\cdot, \cdot)$ is a nonnegative continuous function, such as the squared Euclidean distance $d(a, b) = \frac{1}{2}\|a - b\|^2$. For an overview of more general approaches to image registration [Mod04; Gos05; FM08; Mod09], including optimal control [You10; MTY15; Pol18], statistical [Yan+17] and feature-based [RV10] frameworks we refer to the broad literature on the subject.

Just as image restoration, image registration is an ill-posed problem that can only be solved in a meaningful way by adding further assumptions on the regularity or smoothness of the displacement u. Usually, this is done by applying a Tikhonov-Phillips regularization term as in (1.9). The choice of regularizer has to be well-adapted to the specific setting. When objects in the image undergo sliding or tearing movements, the deformation cannot be expected to be continuous everywhere and edge-enhancing regularizers, such as total (generalized) variation, will yield reasonable results. On the other hand, in situations where all parts of an image are sufficiently connected, as is often the case with organic tissue, the deformation can be assumed to be smooth. In these cases, first-order regularizers tend to be too restrictive since they heavily penalize affine and especially rigid deformations, such as rotations, which are often a mere artifact of a changed viewpoint. The second-order *curvature* regularizer [FM03]

$$R(u) := \frac{1}{2} \int_\Omega \|\Delta u(x)\|^2\, dx \tag{1.19}$$

is an alternative that is easy to implement efficiently. The fact that the curvature regu-

larizer does not only ignore affine deformations but all harmonic functions, i.e., solutions of $\Delta u = 0$, is unphysical in theory. In practice, however, this regularizer is known to perform reasonably well.

The resulting variational formulation of the image registration problem is usually tackled using quasi-Newton optimization methods. However, the highly non-convex nature of the data discrepancy term D (Fig. 1.1) can lead to situations where these local optimization methods get stuck in local minima, especially in case of large displacements or many fine structures, such as thin vessels or fibers. Multi-level models and pre-registration steps are often employed to mitigate this deficiency in practice. In this work, we try to find global solutions to image registration problems by refining the work in [LL18] on convex relaxation methods based on functional lifting (Chap. 5).

Manifold-valued images As noted above, the signals underlying an image might come with periodicity, symmetry or geometric constraints that favor the mathematical representation as functions with nonlinear range, such as a manifold $\mathcal{M}$. Such a choice of mathematical modeling is widespread in the processing of Interferometric Synthetic Aperture Radar (InSAR) [MF98], Electron Backscatter Diffraction (EBSD) [BHS11], Diffusion Tensor Imaging (DTI) [BML94], orientational/positional [Ros+12] data or of images with values in non-flat color spaces, such as hue-saturation-value (HSV) or chromaticity-brightness (CB) color spaces [CKS01].

They come with an inherent non-convexity (Fig. 1.4), as the space of images $u\colon \Omega \to \mathcal{M}$ is generally non-convex, with few exceptions, such as if $\mathcal{M}$ is a Euclidean space, or if $\mathcal{M}$ is a Hadamard manifold, if one allows for the more general notion of geodesic convexity [Bač14; Bač+16]. Efficient and robust convex numerical optimization algorithms therefore cannot be applied and global optimization is generally out of reach.

In order to address the inherent non-convexity, functional lifting (see Sect. 1.2.2) methods have been applied to the restoration of cyclic data [SC11; CS12] with $\mathcal{M} = \mathbb{S}^1$, which was later [Lel+13b] generalized for the case of total variation regularization to data with values in more general manifolds. In practice, the lifting methods are applicable to variational problems with values in manifolds of dimension $s \leq 3$. The theoretical framework applies to manifolds of arbitrary dimension, but the numerical costs increase exponentially. Based on the work in [Lel+13b], we propose a functional lifting framework for manifold-valued images that is applicable to a wider class of (convex) regularizers and present a refined discretization strategy based on finite element spaces on manifolds (Chap. 4).

There also exists a wide range of *local* smooth [AMS09] and nonsmooth [WDS14;

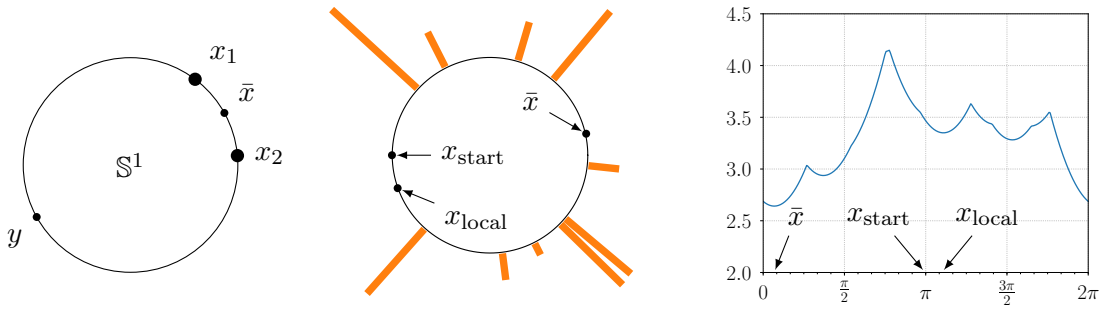


Figure 1.4: Variational problems where the feasible set is a non-Euclidean manifold are prone to local minima and non-uniqueness, which makes them generally much more difficult than their counterparts in $\mathbb{R}^n$. The inherent non-convexity of the feasible set is not only an issue of representation. The example shows the generalization of the weighted mean to manifolds: the Riemannian center of mass $\bar{x}$ of points x_i on a manifold – in this case, the unit circle $\mathbb{S}^1$ – is defined as the minimizer (if it exists and is unique) of the problem $\inf_{x \in \mathbb{S}^1} \sum_i \lambda_i d(x_i, x)^2$, where d is the geodesic (angular) distance and $\lambda_i > 0$ are given weights. **Left:** Given the two points x_1 and x_2, the energy for computing their "average" has a local minimum at y in addition to the global minimum at $\bar{x}$. Compare this to the corresponding problem in $\mathbb{R}^n$, which has a strictly convex energy with the unique and explicit solution $(x_1 + x_2)/2$. **Center and right:** When the number of points is increased and non-uniform weights are used (represented by the locations and heights of the orange bars), the energy structure becomes even less predictable. The objective function (right, parametrized by angle) exhibits a number of non-trivial local minimizers that are not easily explained by global symmetries. Again, the corresponding problem – computing a weighted mean – is trivial in $\mathbb{R}^n$.

WDS16; Bau+16; BPS16; Bre+18; Ber+18a] optimization methods on manifolds. These methods are generally applicable to manifolds of any dimension whose (inverse) exponential mapping can be evaluated in reasonable time and are quite efficient in finding a local minimum, but can get stuck in local extrema. We mention that, beyond variational models, there exist statistical [Fle12], discrete graph-based [BT18], wavelet-based [SW18], PDE-based [Che+04] and patch-based [LPS17] models for the processing and regularization of manifold-valued signals.

Diffusion-weighted MRI and Q-ball imaging In medical applications, the diffusivity of water in tissues that exhibit fibrous microstructures, such as muscle fibers or axons in cerebral white matter, contains valuable information about the fiber architecture in the living organism. *Diffusion-weighted* (DW) magnetic resonance imaging (MRI) is well-established as a way of measuring the main diffusion directions by consecutively applying six or more magnetic field gradients. However, this imaging technique is prone to noise. Consequently, DW-MRI data is a particularly interesting target for post-processing in terms of denoising and regularization [Del+07]. A widely used reconstruction scheme for DW-MRI data is *Q-ball imaging* (QBI) [Tuc04] where the quantity of interest is the marginal probability of diffusion in a given direction, the *orientation distribution function* (ODF) [ALS09]. Reconstructing orientation distributions rather than a single orientation at each point allows to recover directional information of structures, such as vessels or nerve fibers, that may overlap or have crossings.

Existing mathematical frameworks for QBI generally follow the standard literature on the physics of MRI [Cal91, p. 330] in assuming ODFs to be given by a *probability density function* in $L^1(\mathbb{S}^2)$, often with an explicit parametrization. While practical, the probability density-based approach raises some modeling questions, which lead to deeper mathematical issues. We rather propose to use the interpretation of Q-ball data as measure-valued images and apply optimal transport metrics (see Sect. 1.2.3) that properly take into account distances on the underlying set $\mathbb{S}^2$ (Chap. 3).

Markov random fields While, in this work, the mathematical definition of an image is as a function over a continuous domain whose regularity or smoothness defines its spatial coherence properties, we note that there are alternative mathematical frameworks that model images as signals given on discrete sample points. In those models, the spatial coherence is encoded in a graph structure defined on the sample points and in the specification of the statistical dependence of the values at different points in the graph. A very popular framework are Markov Random Fields (MRFs), where the value of a

vertex in the graph is assumed to be statistically independent from all vertices outside of its immediate neighborhood.

The functional lifting method (see Sect. 1.2.2) was originally introduced into the image processing community through corresponding insights in the context of MRF models [KT02; Ish03]. Furthermore, the discrete mathematics involved in the study of these models allows for a rigorous complexity analysis of common image processing tasks that continues to hold for many continuous image processing frameworks.

For a detailed introduction to MRF approaches, we refer to the literature on discrete and statistical methods of image processing [WJ08; BKR11; Fie11].

1.2.2 Functional Lifting

The main contribution of this thesis is a novel measure-valued *functional lifting* framework (Chap. 6) that is applicable to a large class of scalar, vectorial, manifold-valued (Chap. 4), and first- and second-order (Chap. 5) problems. In this section, we introduce the main concepts and discuss recent work on scalar, vectorial, sublabel-accurate, and higher-order models.

Multi-class labeling problems The historical starting point of functional lifting methods in the image processing community are *continuous multi-label-* or *multi-class labeling problems* [Poc+08; LBS09], the problem of finding a function $u\colon \Omega \to \Gamma$ that assigns a *label* $u(x) \in \Gamma$ from a *discrete* range Γ to each point x in a continuous domain $\Omega \subset \mathbb{R}^d$, while minimizing an energy functional $F(u)$. The name stems from the interpretation as the continuous counterpart to the fully discrete problem of assigning to each vertex of a graph one of finitely many labels $\gamma^1, \dots, \gamma^L$ while minimizing a given cost function [GPS89; CKR98; KT02; Ish03].

The prototypical application of multi-labeling techniques is multi-class image segmentation (see Sect. 1.2.1), where the task is to partition a given image into finitely many regions. In this case, the label set Γ is discrete and each label represents one of the regions so that $u^{-1}(\{\gamma\}) \subset \Omega$ is the region that is assigned label γ.

In the fully-discrete setting, one way of tackling first-order multi-label problems is to look for good linear programming relaxations [CKR98; KT02; Ish03]. These approaches were subsequently translated to continuous domains Ω for the two-class [CEN06], multi-class [Zac+08; Poc+08; LS11; Bae+14; BT15], and vectorial [GC10] case, resulting in non-linear, but convex, relaxations. By honoring the continuous nature of Ω, they reduce metrication errors and improve isotropy [SGC11; SNC12; GSC13; Str15], see [Lel+13a]

for a discussion and more references. Many of these relaxations fit into the concept of *functional lifting*.

General concept The general strategy of functional lifting methods, which we will also follow in this work, is to replace the energy minimization problem

$$\inf_{u\colon \Omega\to\Gamma} F(u), \tag{1.20}$$

by a problem

$$\inf_{v\colon \Omega\to X} \mathcal{F}(v), \tag{1.21}$$

where X is some "nice" convex set of larger dimension than Γ with the property that there is an embedding $i\colon \Gamma \hookrightarrow X$ and $F(u) \approx \mathcal{F}(i \circ u)$ in some sense whenever $u\colon \Omega \to \Gamma$.

In general, the lifted functional $\mathcal{F}$ is chosen in such a way that it exhibits favorable numerical or qualitative properties compared with the original functional F while being sufficiently close to the original functional so that minimizers of $\mathcal{F}$ can be expected to have some recoverable relationship with global minimizers of F. Usually, $\mathcal{F}$ is chosen to be convex when F is not, which will make the problem amenable for convex optimization algorithms and allows to find a global minimizer of the lifted problem.

Projection of lifted solutions While recent lifting strategies generally avoid local minimizers of the original problem, they are still an approximation and they are generally not guaranteed to find the global minimizers of the original problem. A central difficulty is that some simplifications have to be performed in the lifting process in order to make it computationally feasible, which may lose information about the original problem. As a result, global minimizers $v\colon \Omega \to X$ of the lifted problem need not be in the image of Γ under the embedding $i\colon \Gamma \hookrightarrow X$ and therefore are not directly associated with a function in the original space.

The process of projecting a solution back to the original space of functions $u\colon \Omega \to \Gamma$ is a difficult problem and, unless Γ is a continuous scalar range [Poc+10], the projection cannot be expected to be a minimizer of the original functional. These difficulties may be related to the fact that the original problems are NP-hard [CS12]. As in the discrete labeling setting [KT02], so-called *rounding strategies* have been investigated in the continuous case [LLS13a; Lel11] that come with an *a priori* bound for the relative gap between the minimum of the original functional and the value attained at the projected version of a minimizer to the lifted functional.

The issue of projecting minimizers of the lifted functional to minimizers of the original

problem is well-understood in the continuous scalar-valued case where a simple thresholding can be applied [Poc+10, Theorem 3.1]. Thresholding is also used for vectorial data with component-wise lifting as in [SCC14] and, in [Lel+13b; Möl+16; Lau+16], simple averaging produces useful results even though no theoretical proof is given addressing the accuracy in general. The considerations in [Fed74; BVZ01; Lav19a; VL19] indicate that there might be general theoretical obstructions to provably accurate projection methods for vector-valued models.

Continuous scalar range In addition to the case of a *discrete* range Γ, relaxation methods have been derived for *continuous* (non-discrete) ranges, most notably the scalar case $\Gamma \subseteq \mathbb{R}$ [ABD03; Poc+10]. They typically consider *first-order* energies that depend pointwise on u and ∇u only:

$$F(u) = \int_\Omega f(x, u(x), \nabla u(x))\, dx. \tag{1.22}$$

The integrand f is assumed to be convex in the third component and nonnegative. The equivalent problem class in the fully discrete setting consists of the energies with only unary (depending on one vertex's label) and pairwise (depending on two vertices' labels) terms.

For the problem (1.22), applying a strategy as in (1.20)–(1.21) comes with a substantial increase in dimensions that originally [Poc+09b] coined the term *functional lifting*. In [Poc+09b], the (non-convex) Mumford-Shah functional for edge-preserving image regularization and segmentation (see Sect. 1.2.1) is lifted to a space of functions $v\colon \Omega \times \Gamma \to [0,1]$, $\Gamma \subset \mathbb{R}$. The authors use the special "step function" lifting $X = \{v\colon \Gamma \to [0,1]\}$ and $i(z^*) = v$ with $v(z) = 1$ if $z \leq z^*$ and 0 otherwise, which is only available in the scalar case. The approach was motivated by the method of calibrations [ABD03] (see Sect. 1.2.4).

Vectorial range The application of the step function lifting to *vectorial* data $\Gamma \subset \mathbb{R}^s$, $s > 1$, is not straightforward, as the concept of subgraphs, which is central to the idea, does not translate easily to higher-dimensional ranges. There are functional lifting approaches for vectorial data with first-order regularization that consider the subgraphs of the components of u [GSC13; SCC14]. A variant without subgraphs [Lel+13b] applies to manifold-valued ranges Γ and total variation-based functionals. In the vectorial case, the integrand $f\colon \Omega \times \Gamma \times \mathbb{R}^{s,d} \to \mathbb{R}$ in (1.22) is usually assumed to be convex in the third component and nonnegative, but some lifting strategies [WC16; MC19], based on

the theory of currents [GMS98a; GMS98b] (see Sect. 1.2.4), are applicable to polyconvex integrands so that also minimal surface problems fit into this framework.

Sublabel-accurate liftings While the above models come with a fully continuous description, a numerical implementation requires the discretization of Ω as well as the range Γ. This introduces two possible causes for errors: *metrication errors* and *label bias*.

Metrication errors are artifacts related to the graph or grid representation of the spatial image domain Ω, finite-difference operators, and the associated choice of metric. They manifest mostly in unwanted anisotropy, missing rotational invariance, or blocky diagonals. They constitute a common difficulty with all variational problems and lifting approaches [Klo+08; Lel+13b].

In contrast, *label bias* means that the discretization favors solutions that assume values at the chosen "labels" (discretization points) $Z^1, \dots, Z^L$ in the *range* Γ (see Fig. 1.5). This is very desirable for discrete Γ, but in the continuous case severely limits accuracy and forces a suitably fine discretization of the range.

In more recent so-called *sublabel-accurate* approaches for scalar and vectorial ranges Γ, more emphasis is put on the discretization [ZK12; Möl+16; Lau+16] to get rid of label bias in models with total variation regularization, which allows to greatly reduce the number of discretization points for the range Γ. In a recent publication [MC17], the gain in sublabel accuracy is explained to be caused by an implicit application of first-order finite elements on Γ as opposed to previous approaches that can be interpreted as using zero-order elements, which naturally introduces label-bias. Finally, there is a recent extension of the sublabel-accurate approaches to arbitrary convex regularizers using the theory of currents [MC19].

In Sect. 2.7, we propose a finite element-based discretization, that reduces label bias in the manifold-valued case (Chap. 4) compared to previous approaches [Lel+13b]. The proposed discretization agrees with the sublabel-accurate approaches [Möl+16; Lau+16] for Euclidean ranges.

Models with higher-order terms Only very recently, attempts at generalizing the continuous lifting strategies to models with higher-order regularization have been made – for regularizers that depend on the Laplacian [LL18; VL19] in case of vectorial ranges $\Gamma \subset \mathbb{R}^s$ and for the total generalized variation [RPB13; SG19] in case of a scalar range $\Gamma \subset \mathbb{R}$. However, in contrast to the first-order theory, the higher-order models, although empirically useful, are still considerably less mathematically validated. Furthermore, we

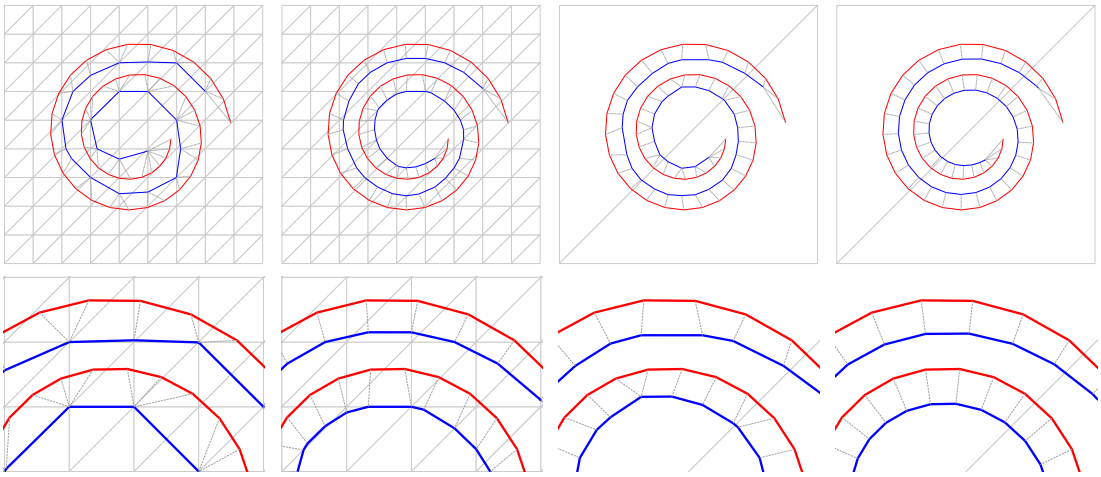


Figure 1.5: Rudin-Osher-Fatemi (ROF) L^2-TV denoising (blue) of a (Euclidean) vector-valued signal $u\colon [0,1] \to \mathbb{R}^2$ (red), visualized as a curve in $\mathbb{R}^2$. The problem is solved by the continuous multi-labeling framework with functional lifting described in Chap. 4. The discretization points (labels) in the range $\mathbb{R}^2$, which are necessary for the implementation of the lifted problem, are visualized by the gray grid. **Left:** The method in [Lel+13b] does not force the solution to assume values at the grid points (labels), but still shows significant bias towards edges of the grid (blue curve). **Second from left:** With the same number of labels, the method from [Lau+16] is able to reduce label bias by improving data term discretization. **Second from right:** Furthermore, the method from [Lau+16] allows to exploit the convexity of the data term to get decent results with as little as four grid points. **Right:** Further exploiting the quadratic form of the data term even produces the numerically exact reference solution, which in this case can be precisely computed using the unlifted formulation due to the convexity of the original problem. This shows that, for the Euclidean fully convex case, the sublabel-accurate lifting allows to recover the exact solution with careful discretization. We extend these sublabel-accurate techniques to the case of manifold-valued images in Chap. 4, e.g., for curves on the sphere $\mathbb{S}^2$ (Fig. 4.3).

mention that there are models where the image *domain* Ω is replaced by a shape (or manifold) [Del+09; Ber+17]. Our proposed functional lifting framework is applicable to first- (Chap. 4) and second-order (Chap. 5) variational problems and comes with a robust mathematical formulation (Chap. 6).

Measure-valued liftings In this work, we will be concerned with measure-valued frameworks for functional lifting, i.e., where the lifted space is chosen to be $X = \mathcal{P}(\Gamma)$, the space of Borel probability measures over Γ, with embedding $i\colon \Gamma \hookrightarrow \mathcal{P}(\Gamma)$, where $i(z) := \delta_z$ is the Dirac point measure with unit mass concentrated at $z \in \Gamma$. We consider a generalization of existing functional lifting approaches to a general class of first-order models with manifold ranges $\Gamma = \mathcal{M}$ (Chap. 4) and to second-order models with vectorial range (Chap. 5). Furthermore, we introduce a measure-valued framework motivated by the theory of dynamical optimal transport (Chap. 6) as an alternative to the popular framework based on the theory of currents, that exclusively applies to first-order problems, while our model applies to first- and second-order problems.

Models with submodular objective functions This thesis investigates variational problems (1.2) where the integrand is convex in the last and non-convex in the first two components. However, the concept of functional lifting has been applied to other problems, as well, in particular to (non-convex) *submodular* objective functions $F\colon \mathcal{X} \to \mathbb{R}$ that are defined on the cartesian product $\mathcal{X} = X^1 \times \cdots \times X^N$ of compact sets of real numbers $X^i \in \mathbb{R}$. The functional lifting approach [Bac19] for submodular functions uses multi-marginal optimal transport to define a convex relaxation $\mathcal{F}$ of F on the product space of measures $\mathcal{P}(X^1) \times \cdots \times \mathcal{P}(X^N)$. This framework has the property that minimizers of F may be obtained from minimizers of $\mathcal{F}$. It has applications in signal processing, especially in the denoising of sparse signals. For an overview of research on submodular problems, their convex relaxations and applications, we refer to [Bac19].

1.2.3 Optimal Transport

At the heart of our proposed models are methods from the mathematical theory of *optimal transport* that define a natural geometry on spaces of measures and therefore help in defining regularity of measure-valued functions. After a brief introduction to optimal transport, we discuss the main three aspects that are relevant to this work: the Kantorovich-Rubinstein formulation of optimal transport distances, the dynamical formulation of optimal transport due to Benamou and Brenier, and the more recent notion of (harmonic) functions with values in the Wasserstein space.

Historical background and basic notions The theory of optimal transport dates back to 1784, when the French mathematician Gaspard Monge published a mathematical description of earthwork [Mon84]: Monge described the problem of minimizing the amount of carriage needed to transport earth from a given area (*déblai*, excavation) to a given equal area (*remblai*, embankment). In modern mathematical terms, given a measure space X and two probability measures $\mu, \nu \in \mathcal{P}(X)$, we seek to find a *transport map* $T\colon X \to X$ that minimizes the total transport cost

$$\int_X c(x, T(x))\, \mu(dx), \tag{1.23}$$

under the constraint $T_*\mu = \nu$, where $T_*\mu(A) := \mu(T^{-1}(A))$, $A \subset X$ measurable, is the push-forward measure of μ under T, and $c(\cdot, \cdot)$ is a non-negative function of local transport cost, such as $c(x, y) = \|x - y\|^p$, $p \in [1, \infty)$, in the case $X \subset \mathbb{R}^d$. Under these assumptions, however, the problem is ill-posed in general: When the total mass of μ is concentrated in a single point while ν distributes the mass between two points, the constraint $T_*\mu = \nu$ clearly can never be satisfied.

When the Russian mathematician Leonid Kantorovich described a mathematical model for the leveling of land area and for the assignment of consumption locations to production locations in 1942 [Kan06b], he used an alternative problem description: Given a measure space X and two probability measures $\mu, \nu \in \mathcal{P}(X)$, we seek to find a *transport plan* or *coupling* $\gamma \in \mathcal{P}(X \times X)$ that minimizes the total transport cost

$$\int_{X \times X} c(x, y)\, \gamma(dx, dy), \tag{1.24}$$

subject to the constraints $\pi_1\gamma = \mu$ and $\pi_2\gamma = \nu$, where $\pi_1\gamma(A) := \gamma(A \times X)$, $\pi_2\gamma(A) := \gamma(X \times A)$, $A \subset X$ measurable, are the marginal distributions of γ. The existence of such a transport plan can be shown for every μ and ν under very weak assumptions on X and c [Vil09, Theorem 4.1]. Shortly after that, Kantorovich gave a sketch of how his theory can be applied to Monge's problem [Kan06a]. In fact, if the given measures μ and ν are absolutely continuous with respect to the Lebesgue measure and $c(x, y) = \|x - y\|_2^2$, the (in this case, unique) optimal transport plan according to Kantorovich is actually *deterministic*, i.e., supported on the graph of an optimal transport map T in the sense of Monge [Vil09, Theorem 9.4].

The minimal transport cost defines a statistical distance on $\mathcal{P}(X)$. Most notable, in case of a metric space (X, d), the *Kantorovich-Rubinstein* [KR57] or *Wasserstein* [Vas69]

distance between probability measures $\mu, \nu \in \mathcal{P}(X)$ is defined, for $1 \leq p < \infty$, as

$$W_p(\mu,\nu)^p := \inf\left\{ \int_{X\times X} d(x,y)^p \, \gamma(dx,dy) : \gamma \in \Pi(\mu,\nu) \right\}, \tag{1.25}$$

$$\Pi(\mu,\nu) := \{\gamma \in \mathcal{P}(X\times X) : \pi_1\gamma = \mu,\ \pi_2\gamma = \nu\}. \tag{1.26}$$

In fact, W_p defines a metric on $\mathcal{P}(X)$ that metrizes weak* convergence of measures [Vil09, Thm. 6.9]. Moreover, in the case $p = 1$, this distance is induced by a norm [KR57] and known under the name *earth mover's distance (EMD)* in the computer vision [RTG00] and statistical [LB01] community.

Dynamical optimal transport It was more than 50 years after Kantorovich's formulation that a dynamic (Eulerian) formulation of the optimal transport problem was introduced by Benamou and Brenier [BB00] for compact Euclidean sets $X \subset \mathbb{R}^s$. They proved that the Wasserstein distance between to measures $\mu_0, \mu_1 \in \mathcal{P}(X)$ is equal to the minimal Dirichlet energy of absolutely continuous curves in $\mathcal{P}(X)$ connecting the two measures:

$$W_p(\mu,\nu)^p = \inf\Bigg\{ \int_0^1 \int_X \|v_t\|^p \, d\mu_t \, dt : \partial_t\mu + \operatorname{div}(v\mu) = 0, \\ \mu(0) = \mu_0, \mu(1) = \mu_1 \Bigg\}, \tag{1.27}$$

where the continuity equation $\partial_t\mu + \operatorname{div}(v\mu) = 0$ is to be understood in the sense of distributions. A rigorous description of this setting, including a definition of absolutely continuous curves in metric spaces, is given in [AGS04; AGS08; San15] and in [Amb03] for $p = 1$. In [AGS08, Section 8.2], it is also established that absolutely continuous curves in the Wasserstein space can be decomposed as the superposition of absolutely continuous curves in X.

Recently, the above notion of regularity for curves in the space of probability measures has been used for the reconstruction of temporal positron-emission tomography (PET) imaging data, which allows to follow singular immune cells over time in-vivo [SSW19].

For the mathematical framework that we introduce in Chap. 6, a generalized version of dynamical optimal transport is a central theoretical motivation and background: the study of *mappings with values in Wasserstein spaces* that we describe in the following paragraph.

Mappings with values in Wasserstein spaces The minimizers of the dynamical formulation of Wasserstein distances are geodesics in the corresponding Wasserstein space. As a generalization of this, the concept of harmonic functions with values in the Wasserstein space was introduced in [Bre03]: A harmonic function on $\Omega \subset \mathbb{R}^d$ with values in $\mathcal{P}(X)$ is a pair of measures (μ, E) on $\Omega \times X$ valued in $\mathbb{R}$ and $\mathbb{R}^{d,n}$ that minimizes the energy

$$\int_{\Omega \times X} \|v\|_F^2 \, d\mu \tag{1.28}$$

subject to $\nabla_x \mu + \mathrm{Div}_z E = 0$ (in the sense of distributions) with $E := v\mu$, and prescribed boundary data on $\partial\Omega$. Whenever the boundary data is concentrated on the graph of a continuous function, the infimum of this Dirichlet energy is achieved for $\mu := \delta_u$ and $E := \nabla u \, \delta_u$ where u is the corresponding harmonic extension in the classical sense and $\delta_u(x) := \delta_{u(x)}$ is, for each $x \in \Omega$, the Dirac point mass concentrated at $u(x)$. As was recently investigated in [Lav19a; Lav19b], existence of harmonic extensions can be shown for general Lipschitz boundary data. However, contrary to the one-dimensional Benamou-Brenier setting, there is no superposition principle for generalized harmonic functions.

In this work, we are interested in the regularity of measure-valued images, similar to the case of generalized harmonic functions introduced above. Furthermore, our lifting approaches are based on concepts from variational calculus that are also at the heart of the Benamou-Brenier as well as the Kantorovich relaxation of the Monge transport problem.

Optimal transport in image processing In the introduction to mathematical image processing (Sect. 1.2.1), we identified images with *functions* on some domain on which a notion of spatial coherence is available. Just as we can identify every L^1-function as the density function of a (signed) *measure*, we can also model images as measures on some domain. In case of gray-value images, we can assume those measures to be probability measures after normalization. Then, the Monge problem of optimal transport can be interpreted as a model for image registration between two given images (measures) μ and ν, as noted by [HT01; Cha+09]. The constraint $T_*\mu = \nu$ then corresponds to an exact match between the reference ν and transformed template image $T_*\mu$ while the cost term $c(x, T(x))$ in (1.23) serves as a regularizer for the deformation T. The exact matching constraint is too restrictive in many applications of image registration where the alignment is expected to undo large deformations while preserving small and medium-sized anomalies. While there exist generalizations of optimal transport that

relax this constraint [Fer+14], it has not been applied to the particular problem of image registration so far.

On the other hand, the dynamical optimal transport viewpoint has been successfully applied to image metamorphosis, where the transition process is more interesting than the actual deformation function [Maa+15]. In the same spirit, dynamic inverse problems can be regularized using the notion of absolutely continuous curves in the Wasserstein space [BF19; Bre+19] and barycenters in the Wasserstein space of images have been applied to texture mixing [Rab+12].

Another popular application of Wasserstein distances is as a similarity measure in image processing. Variants of the Kantorovich-Rubinstein formulation have been applied in [BFS12; Lel+14; BL15] and, more recently, in [FLS16; TPG16; FLS17] to the problems of real-, RGB- and manifold-valued image denoising. For more applications, we refer to review articles on optimal transport in image processing [Pey+09] and statistics [PC19].

1.2.4 Variational Calculus

The *calculus of variations* is concerned with the minimization of real-valued mappings over (infinite-dimensional) function spaces, often given in integral form as introduced in (1.2). Central questions are existence, uniqueness and stability of minimizers (well-posedness) as well as necessary or sufficient criteria and qualitative properties of minimizers. As minimization is usually over an infinite-dimensional topological vector space of functions, theoretical results from functional analysis are a crucial ingredient. Through the optimality criterion known as Euler-Lagrange equation, there is a natural relationship with the theory of partial differential equations. For a general overview of and an introduction to the field, we recommend the numerous monographs on the topic [GH96a; GH96b; ABM06; Dac08; Str08; Cla13; Rin18]. An alternative and nowadays very influential point of view is taken by Rockafellar [RW04], where key concepts from traditional variational calculus like stationarity are generalized to the case of non-smooth objects and their subderivatives.

We will make use of both traditional and modern concepts from the calculus of variations in order to motivate, define and analyze our proposed functional lifting framework. Furthermore, the modern concept of functional lifting (see Sect. 1.2.2) is similar to methods that have originally been introduced in the theory of variational calculus, as we will see in the following paragraphs.

Direct method When tackling variational problems of the form (1.1), one of the first questions is whether the minimum is attained at all – whether minimizers *exist*. A

general method for constructing a proof of existence is the *direct method*: Given that the feasible set is non-empty, there always exists a sequence of functions (u^k) such that $F(u^k)$ converges to its infimum. A direct method proof now chooses a topology with respect to which F is sufficiently regular so that every minimizing sequence has a cluster point in that topology and so that this cluster point is actually a minimizer of F. The two regularity properties that F usually has to satisfy are:

1. F is *coercive* in the sense that the sublevel set $\{u : F(u) \leq s\}$ is precompact in the chosen topology for each $s \in \mathbb{R}$.

2. F is sequentially *lower semicontinuous* with respect to the chosen topology in the sense that $F(u^*) \leq \liminf_{k\to\infty} F(u^k)$ whenever u^k converges to u^* in that topology.

A setup of such a proof often starts by embedding the feasible set X into a larger topological space Y and defining a relaxed functional $\mathcal{F}$ that agrees with F on X while being lower semicontinuous and coercive in the topology of Y. The direct method then yields a minimizer of $\mathcal{F}$ in Y, and regularity results (necessary conditions) are applied to show that this minimizer is actually an element of X and, therefore, a minimizer of F.

The procedure of replacing a functional F by a functional $\mathcal{F}$ defined on a larger space and minimizing $\mathcal{F}$ in order to draw conclusions about minimizers of the original functional F is a central concept of functional lifting strategies (see Sect. 1.2.2).

Lower semicontinuous relaxations While the regularity results mentioned at the end of the last paragraph are not easily available in practice, there is a large theory about lower semicontinuous relaxations of functionals. In the case of function spaces with scalar range and integral functionals (1.2) with first-order terms, it is well-known that the lower semicontinuity with respect to the weak topology of Sobolev spaces is closely related to convexity of the integrand with respect to the gradient variable. In the case of a vectorial range, a similar connection exists for generalized notions of convexity (quasi- and polyconvexity) [Dac08].

Due to this connection to convexity, the functional lifting approach (see Sect. 1.2.4) that aims to approximate a non-convex functional by a convex one, has natural theoretical intersections with the study of lower semicontinuous relaxations.

Cartesian currents A useful viewpoint in the study of lower semicontinuous relaxations is to identify functions $u\colon \Omega \to \Gamma$ with their (oriented) graph in $\Omega \times \Gamma$. The feasible set of such graphs can then be embedded into a larger class of oriented objects in $\Omega \times \Gamma$ called *currents*, which are the dual concept to continuously differentiable vector fields or, more

precisely, differential forms in $\Omega \times \Gamma$. The theory of Cartesian currents [GMS98a] then studies those currents that arise as the weak limit of graph currents.

Every integral functional (1.2) of first order that depends on the Jacobian ∇u in a polyconvex way while being non-convex in u can be extended in a completely convex way to the space of currents [GMS98b]. This property has been used in functional lifting approaches with vectorial range [MC19], and many functional lifting approaches with scalar range, in particular those based on calibrations, fit into this framework after simple reformulations (see Sect. 6.4.2).

Furthermore, the language of currents has been used to parametrize spaces of shapes [Ben+19] and to solve correspondence or shape-matching problems [WC16].

Gradient Young measures Minimizing sequences of non-convex functionals often develop singularities and show highly oscillatory behavior. Even if the regularity results that are necessary for the last step in the direct method fail to hold, minimizers in the lifted space can give interesting insight into the problem setting – for example in the study of microstructures arising from solid-solid phase transitions in elastic crystals or alloys [Mül99].

More precisely, some of the qualitative properties of a (minimizing) sequence of functions $v^k \colon \Omega \to \mathbb{R}^N$ can be expressed in the form of a parametrized measure $\nu_x \in \mathfrak{M}(\mathbb{R}^N)$, $x \in \Omega$, or, equivalently, a measure-valued function $\nu \colon \Omega \to \mathfrak{M}(\mathbb{R}^N)$, $\nu(x) := \nu_x$: the so-called *Young measure* generated by that sequence that has, under appropriate assumptions, the property that $f(v^k)$ converges weakly to $\int_\Omega \int_{\mathbb{R}^N} f(A)\, \nu_x(dA)\, dx$ for every continuous function $f \colon \mathbb{R}^N \to \mathbb{R}$. Typically, minimizing sequences of many functionals are, by Sobolev embedding theorems, strongly convergent in some L^p-space so that non-trivial oscillatory behavior is only observed in the gradient sequence. That is why the typical application of Young measures is as *gradient Young measures* that are generated by a sequence of gradients $v^k = \nabla u^k$ of functions in some Sobolev space.

We mention this as another example where measure-valued functions play an important role in the calculus of variations. An overview of Young measures in the calculus of variations is given in [Ped97], where also the relationship with lower semicontinuous relaxations is pointed out. Another introduction to the field is contained in the recent book [Rin18].

The calibration method In the theory of minimal surfaces, a sufficient criterion for a submanifold to be a surface of minimal area with given boundary – a *minimal surface* – is that there exists a closed differential form on the enclosing manifold that satisfies

certain compatibility properties on that submanifold [HL82; Mor90]. Such a differential form is called a *calibration*. Inspired by this concept, a sufficient criterion for minimizers of the Mumford-Shah functional has been developed [ABD03]: It can be shown that a function $u\colon \Omega \to \mathbb{R}$ is a minimizer of the Mumford-Shah functional if there exists a divergence-free vector field on $\Omega \times \mathbb{R}$ that satisfies certain compatibility properties on the graph of u. Later, the duality between graphs of functions $u\colon \Omega \to \mathbb{R}$ and vector fields on $\Omega \times \mathbb{R}$ was used to study a duality theory for the Mumford-Shah functional and similar variational problems [Cha01; Poc+09b; Poc+10; BF18]. Even though, apart from [Mor02], these approaches have been applicable to problems with scalar range only, this so-called *calibration method* sparked interest in general frameworks, such as those presented in this work, for the convex relaxation of variational problems via functional lifting in the following years.

1.2.5 Convex Finite-Dimensional Optimization

In contrast to the calculus of variations, we will discuss solution methods for *finite-dimensional* energy minimization problems in this section. In image processing, and in our work, finite-dimensional optimization plays an important role when variational models are discretized and solved numerically. Since, in this work, we will always consider convex energies or convex approximations of convex energies that are nonsmooth in many cases, we introduce some key notions and issues from convex finite-dimensional optimization.

In *convex optimization*, minimization problems of the form

$$\inf_{x \in S} f(x) \tag{1.29}$$

are studied where the objective function $f\colon S \to \mathbb{R}$ as well as the feasible set $S \subset \mathbb{R}^N$ are known to be convex. The convexity does not only guarantee that stationary points are global minimizers, but it automatically comes with additional regularity of the objective function such as local Lipschitz-continuity and the existence of meaningful subderivatives at each $x \in S$. Furthermore, similar to the setting in linear programming, every convex minimization problem comes with dual concave maximization problems that are, under suitable assumptions, guaranteed to take the same extremal value. For an introduction and overview of general concepts in convex analysis and optimization, we refer to [Roc97; ET99]. Applications and numerical basics are covered in [BV04; Nes04].

Many convex optimization problems come with an objective function that is itself the sum of two convex functions, each of which is often better understood or easier

to handle than the sum. This is for example the case for (convex) image processing problems with a (convex) Tikhonov-Phillips regularization term as in (1.9). Similarly, the objective functions of the functional lifting approaches proposed in this work split up into a sum of two convex functions. Over the years, many splitting optimization algorithms have been investigated: To name some of the most prominent, we mention the proximal forward backward splitting algorithm, the Douglas-Rachford algorithm and the Alternating Direction Method of Multipliers (ADMM) [LM79].

Many of the popular convex optimization algorithms are primal-dual methods: they update, in each iteration, a primal and a dual variable. Those algorithms are often more stable or manage to overcome complexities and non-smoothness in the structure of the primal or dual objectives. By evaluating a primal and a dual objective function, they can give a guaranteed certificate of approximation quality through the concept of a (relative) primal-dual gap, as opposed to conventional stopping criteria based on first-order residuals, such as the gradient of the objective function [BV04, Chapter 5.5].

All numerical experiments in this work have been implemented based on the primal-dual splitting algorithm known as (modified) *Primal-Dual Hybrid Gradient* (PDHG) [ZC08; EZC10] or *Chambolle-Pock* [CP11] algorithm (see Sect. 2.6). It has been introduced in [Poc+09b] as an algorithm to solve a convex relaxation of the Mumford-Shah functional and was then generalized to more general convex models with objective functions of the form $F(x) = g(x) + f(Ax)$, where g and f are convex while A is linear. A convergence analysis [CP11] was able to show that its complexity is optimal in a class of convex problems and complexity automatically improves when either f or g (or both) are strictly convex or smooth. Recently, preconditioning [PC11], adaptive step sizes [GEB13] and a line-search procedure [MP18] have been investigated for a further increase in performance. Extensions to nonlinear operators A exist [Val14; CV17]. We use the PDHG algorithm for our very high-dimensional lifting models, as it is highly parallelizable on GPUs while still being comparably easy to implement and applicable to a wide range of nonsmooth objectives. The variant of the algorithm used for our implementations is documented in detail in Sect. 2.6.

1.3 Contributions and Outline

We specify our notation and some mathematical foundations that we will use throughout this work in Chap. 2. The main results are then presented in the following four chapters. Each chapter starts with a short introduction to the setting where related work is referenced. A demonstration of the theoretical novelties is contained in the main sections

and, if applicable, numerical experiments are discussed.

1. In Chap. 3, we give a functional analytically rigorous framework for the mathematical processing of Q-ball data from diffusion-weighted MRI as images with values in the Wasserstein space of measures. In particular, a new total variation regularizer is proposed for the denoising and restoration of measure-valued (Q-ball) images.

   Related Publications:

   - T. Vogt and J. Lellmann. "An Optimal Transport-Based Restoration Method for Q-Ball Imaging". In: *Proc SSVM 2017*. Ed. by F. Lauze, Y. Dong, and A. B. Dahl. Springer, 2017, pp. 271–282
   - T. Vogt and J. Lellmann. "Measure-Valued Variational Models with Applications to Diffusion-Weighted Imaging". In: *J Math Imaging Vis* 60.9 (2018), pp. 1482–1502

2. In Chap. 4, we generalize existing lifting methods for manifold-valued imaging problems to the general first-order variational model and give a rigorous mathematical framework in the continuous setting. We compare our proposed finite element-based discretization scheme with previous manifold-valued functional lifting approaches.

   Related Publications:

   - T. Vogt, E. Strekalovskiy, D. Cremers, and J. Lellmann. "Lifting methods for manifold-valued variational problems". In: *Variational Methods for Nonlinear Geometric Data and Applications*. Ed. by P. Grohs, M. Holler, and A. Weinmann. Springer, 2019, In press

3. In Chap. 5, the lifting strategy is adapted to be applicable to second-order variational models. Based on this, a global solution strategy for image registration problems with (Laplacian) curvature regularization is proposed and discussed.

   Related Publications:

   - T. Vogt and J. Lellmann. "Functional Liftings of Vectorial Variational Problems with Laplacian Regularization". In: *Proc SSVM 2019*. Ed. by J. Lellmann, M. Burger, and J. Modersitzki. Springer, 2019, pp. 559–571

4. In Chap. 6, a new mathematical description of existing functional lifting strategies based on notions from the theory of dynamical optimal transport is proposed. This has the potential to serve as a basis for future efficient second- and higher-order

lifting approaches. Connections with the functional lifting approaches introduced in the previous chapters are discussed and the proposed framework is compared with existing frameworks in the language of Cartesian currents and calibrations.

2 Notation and Mathematical Foundations

Throughout this work, the set of natural numbers $\mathbb{N} := \{1, 2, \dots\}$ starts at 1 and we write $\mathbb{N}_0 := \{0, 1, 2, \dots\}$. Whenever we write $\mathbb{R}^k$, we assume the usual Euclidean topology, metric, norm $\|\cdot\|$ and scalar product $\langle\cdot,\cdot\rangle$, unless explicitly stated otherwise. For the $n \times k$ matrices $\mathbb{R}^{n,k}$ we explicitly specify the norm used. Most importantly, $\|A\|_F^2 := \sum_{i=1}^n \sum_{j=1}^k a_{ij}^2$ is the *Frobenius norm* and $\|A\|_{\sigma,\infty} := \sup_{\|x\|=1} \|Ax\|$ is the (matrix) operator norm or *spectral norm*.

We use the terms *function*, *map* and *mapping* synonymously. We say that a function is *vector-valued* (or *vectorial*) and *scalar-valued* (or *real-valued*) if its values are vectors or scalars. We use *Banach space-valued* as a synonym for *taking values in a Banach space* even though we acknowledge the ambiguity carried by this expression. Analogously, *manifold-valued* means *taking values in a manifold* and *metric space-valued* means *taking values in a metric space*.

Sequences of functions, measures or vectors, (u^k), (μ^k), (x^k), etc., are indexed by a superscript. The subscript is used to refer to the components of a vectorial object: u_k, μ_k, x_k. We refer to a whole sequence of objects by adding parentheses: (u^k).

For topological spaces X and Y, we denote by $C(X, Y)$ the set of continuous maps $f\colon X \to Y$. For a Banach space Y, we denote by $C_b(X, Y)$ the Banach space of bounded continuous maps equipped with the supremum norm $\|f\|_\infty := \sup_{x\in X} \|f(x)\|_Y$. The corresponding subset of maps with compact support is denoted by $C_c(X, Y)$, and its closure in $C_b(X, Y)$ with respect to the supremum norm is denoted by $C_0(X, Y)$. As, in this work, many topological spaces of interest are compact, we note that, for compact topological spaces X, we have $C(X, Y) = C_b(X, Y) = C_c(X, Y) = C_0(X, Y)$. For real-valued functions, i.e., $Y = \mathbb{R}$, we write $C(X) = C(X, \mathbb{R})$, $C_b(X) = C_b(X, \mathbb{R})$, etc.

For metric spaces X and Y and a function $f \in C(X, Y)$, we denote by

$$[f]_{\mathrm{Lip}} := \sup \left\{ \frac{d_Y(f(x_1), f(x_2))}{d_X(x_1, x_2)} : x_1, x_2 \in X, x_1 \neq x_2 \right\} \tag{2.1}$$

the Lipschitz seminorm of f and we write $\mathrm{Lip}(X, Y)$ for the set of functions with finite Lipschitz seminorm. In the special case $Y = \mathbb{R}^n$, the set $\mathrm{Lip}(X, \mathbb{R}^n)$ is a Banach space when equipped with the norm $\|f\|_{\mathrm{Lip}} := \|f\|_\infty + [f]_{\mathrm{Lip}}$.

For $k \in \mathbb{N}$, Ω an open subset of $\mathbb{R}^d$ and Y a topological vector space, we denote by $C^k(\Omega, Y)$ the k-times continuously differentiable functions $f\colon \Omega \to Y$. The corresponding subsets of functions with compact support are denoted by $C_c^k(\Omega, Y)$. The k-th partial derivative of $u \in C^k(\Omega, Y)$ in the i-th direction is denoted $\partial_i^k u \in C(\Omega, Y)$, the gradient is the vectorial function $\nabla u \in C(\Omega, Y^d)$ with components $(\nabla u)_i := \partial_i u$. The Laplacian differential operator Δ is defined for $u \in C^2(\Omega, \mathbb{R}^n)$ as

$$\Delta u(x) := \sum_{i=1}^{d} \partial_i^2 u(x) \tag{2.2}$$

and the (column-wise) divergence $\mathrm{Div}\, u \in C(\Omega, \mathbb{R}^n)$ of $u \in C^1(\Omega, \mathbb{R}^{d,n})$ is

$$[\mathrm{Div}\, u(x)]_k := \sum_{i=1}^{d} \partial_i u_{ik}(x), \qquad \text{for } k = 1, \dots, n. \tag{2.3}$$

For $n = 1$, we usually write div instead of Div. If a function is defined on a Cartesian product $\Omega \times \Gamma$ for some $\Omega \subset \mathbb{R}^d$ and $\Gamma \subset \mathbb{R}^s$, the components associated to Ω and Γ are referred to using different variables x and z, and the associated differential operators are marked with a corresponding subscript: $\Delta_x u$, $\mathrm{Div}_z\, u$, $\nabla_x u$.

2.1 Measure Theory

Let X be a locally compact separable metric space. The *Borel σ-algebra* $\mathcal{B}(X)$ is the smallest σ-algebra in X that contains all open subsets of X. A (nonnegative) *measure* on X is a countably additive function $\mu\colon \mathcal{B}(X) \to [0, \infty]$ with $\mu(\emptyset) = 0$. A function $\mu\colon \mathcal{B}(X) \to \mathbb{R}^n$ is a *finite vectorial Radon measure* on X if there exists a finite nonnegative measure $\|\mu\|$ (the *total variation measure* of μ) and a measurable function $g\colon X \to \mathbb{R}^n$ with $\|g(x)\| = 1$ almost everywhere and

$$[\mu(A)]_i = \int_A g_i \, d\|\mu\| \tag{2.4}$$

for each $1 \le i \le n$. We write $\mu = g\|\mu\|$ and all integrals are in the sense of Lebesgue integration.

We denote by $\mathfrak{M}(X, \mathbb{R}^n)$ the vector space of finite vectorial Radon measures on X and

write $\mathfrak{M}(X) = \mathfrak{M}(X, \mathbb{R})$ for the finite *signed* Radon measures. For each $\mu \in \mathfrak{M}(X)$, there are nonnegative measures with disjoint support $\mu^+, \mu^- \in \mathfrak{M}(X)$, the *Hahn decomposition* of μ, such that $\mu = \mu^+ - \mu^-$. For the total variation measure $|\mu|$ of a signed measure μ it holds that $|\mu| = \mu^+ + \mu^-$. Equipped with the total variation norm $\|\mu\|_{\mathfrak{M}} := \|\mu\|(X)$, the set $\mathfrak{M}(X, \mathbb{R}^n)$ is a Banach space and isometrically isomorphic to the topological dual space of $C_0(X, \mathbb{R}^n)$ with dual pairing

$$\langle \mu, f \rangle := \int_X \langle f,\, d\mu \rangle := \int_X \langle f, g \rangle \, d\|\mu\|, \tag{2.5}$$

whenever $f \in C_0(X, \mathbb{R}^n)$ and $\mu = g\|\mu\| \in \mathfrak{M}(X, \mathbb{R}^n)$ [HS65, p. 364]. Accordingly, a sequence $(\mu^k) \subset \mathfrak{M}(X, \mathbb{R}^n)$ is *weakly* convergent* to μ^* if, for each $f \in C_0(X, \mathbb{R}^n)$,

$$\int_X \langle f,\, d\mu^k \rangle \to \int_X \langle f,\, d\mu^* \rangle \text{ as } k \to \infty. \tag{2.6}$$

For every pair of measures $\mu, \nu \in \mathfrak{M}(X, \mathbb{R}^n)$, by a variant of the Radon-Nikodym theorem [HS65, Thm. 19.42], there is a *Lebesgue decomposition* of ν with respect to μ, i.e., there exist measurable functions g and $g^\perp$ on X and a nonnegative measure $\|\mu\|^\perp \in \mathfrak{M}(X)$ with support disjoint from the support of $\|\mu\|$ such that $\nu = g\|\mu\| + g^\perp \|\mu\|^\perp$. If μ is a nonnegative measure, we say that ν is *absolutely continuous* with respect to μ (or $\nu \ll \mu$) if $\mu(A) = 0$ implies $\nu(A) = 0$ for each $A \in \mathcal{B}(X)$. In this case, $\|\mu\|^\perp = 0$ and we denote by $(d\nu/d\mu)(x) := g(x)$ the *Radon-Nikodym derivative* of ν with respect to μ. On the other hand, if $g = 0$ (or $\nu = g^\perp \|\mu\|^\perp$), then ν and μ are called *mutually singular*.

Integration with respect to a measure is denoted by $d\mu$ and $\mu(dx)$ if the variable cannot be omitted. Integration with respect to the Lebesgue measure is denoted by dx. For a measurable set $\Omega \subset \mathbb{R}^d$, the Lebesgue spaces are denoted by $L^p(\Omega, \mathbb{R}^n)$ for $p \in [1, \infty]$ and the Sobolev spaces are $W^{k,p}(\Omega, \mathbb{R}^n)$ for $k = 1, 2, \ldots$ and $p \in [1, \infty]$. The n-dimensional Lebesgue measure of a measurable set $A \subset \mathbb{R}^n$ is denoted $\mathcal{L}^n(A)$ and, for $s \in [0, \infty)$, the s-dimensional Hausdorff-measure is $\mathcal{H}^s(A)$.

2.2 Optimal Transport

For a Polish (metric, complete, and separable) space X, we briefly recall the definition of optimal transport metrics. We write $\mathcal{P}(X)$ for the set of Borel probability measures on X, i.e., the nonnegative Borel measures in $\mathfrak{M}(X)$ with unit total mass. The *Kantorovich-Rubinstein* [KR57] or *Wasserstein* [Vas69] distance between probabil-

ity measures $\mu, \nu \in \mathcal{P}(X)$ is defined, for $1 \leq p < \infty$, as

$$W_p(\mu, \mu') := \left(\inf_{\gamma \in \Pi(\mu,\mu')} \int_{X \times X} d(x,y)^p \, \gamma(dx, dy) \right)^{1/p}, \tag{2.7}$$

where

$$\Pi(\mu, \mu') := \left\{ \gamma \in \mathcal{P}(X \times X) \colon \ \pi_1 \gamma = \mu, \ \pi_2 \gamma = \mu' \right\}. \tag{2.8}$$

Here, $\pi_i \gamma \in \mathcal{P}(X)$ denotes the i-th marginal of the measure γ on the product space $X \times X$, i.e., $\pi_1 \gamma(A) := \gamma(A \times X)$ and $\pi_2 \gamma(B) := \gamma(X \times B)$ whenever $A, B \subset X$. In fact, W_p defines a metric on $\mathcal{P}(X)$ that metrizes the weak* convergence of measures (see Sect. 2.1) [Vil09, Thm. 6.9].

Fix some arbitrary $x_0 \in X$. Then, the seminorm $[\cdot]_{\mathrm{Lip}}$ is actually a *norm* on the set

$$\mathrm{Lip}_0(X, \mathbb{R}^d) := \{ p \in \mathrm{Lip}(X, \mathbb{R}^d) \colon p(x_0) = 0 \}. \tag{2.9}$$

The Kantorovich-Rubinstein duality [KR57] states that, for $p = 1$, the Wasserstein *metric* is actually induced by a *norm*, namely $W_1(\mu, \mu') = \|\mu - \mu'\|_{\mathrm{KR}}$, where

$$\|\nu\|_{\mathrm{KR}} := \sup \left\{ \int_X p \, d\nu : \ p \in \mathrm{Lip}_0(X), \ [p]_{\mathrm{Lip}} \leq 1 \right\}, \tag{2.10}$$

whenever $\nu \in \mathfrak{M}_0(X) := \{ \mu \in \mathfrak{M} \colon \ \int_X d\mu = 0 \}$. The completion $\mathrm{KR}(X)$ of $\mathfrak{M}_0(X)$ with respect to $\| \cdot \|_{\mathrm{KR}}$ is a predual space of $(\mathrm{Lip}_0(X), [\cdot]_{\mathrm{Lip}})$ [Wea99, Thm. 2.2.2 and Cor. 2.3.5].

In the Euclidean case $X \subset \mathbb{R}^n$, the Wasserstein distance between to measures $\mu_0, \mu_1 \in \mathcal{P}(X)$ can be expressed using the Benamou-Brenier [BB00] formulation:

$$W_p(\mu, \nu)^p = \inf \left\{ \int_0^1 \int_X \|v_t\|^p \, d\mu_t \, dt : \partial_t \mu + \mathrm{div}(v\mu) = 0, \mu(0) = \mu_0, \mu(1) = \mu_1 \right\}, \tag{2.11}$$

where we write $\mu_t := \mu(t)$, whenever $\mu \colon [0,1] \to \mathcal{P}(X)$ is a curve in the space of Borel probability measures. The continuity equation $\partial_t \mu + \mathrm{div}(v\mu) = 0$ is to be understood in the sense of distributions, i.e., for each $\phi \in C_c^1((0,1) \times X)$, we have

$$\int_0^1 \int_X \partial_t \phi(t,x) \, d\mu_t \, dt + \int_0^1 \int_X \langle \nabla_x \phi(t,x), v(t,x) \rangle \, d\mu_t \, dt = 0. \tag{2.12}$$

In other words, the Wasserstein distance is the minimal Dirichlet energy of absolutely continuous curves in $\mathcal{P}(X)$ connecting the two measures. A rigorous description of this

setting, including a definition of absolutely continuous curves in metric spaces, is given in [AGS04; AGS08; San15] and in [Amb03] for $p = 1$.

2.3 Functions of Bounded Variation

In image processing, images that are piecewise smooth with sharp discontinuities (edges) occur frequently. A popular function space for modeling this kind of data is the space of *functions of bounded variation*. For an open set $\Omega \subset \mathbb{R}^d$, we say that a function $u \in L^1(\Omega, \mathbb{R}^n)$ is of bounded variation, if its distributional Jacobian exists as a finite vectorial Radon measure $\mu \in \mathfrak{M}(\Omega, \mathbb{R}^{d,n})$ (see above). This means that, for each $\phi \in C_c^1(\Omega, \mathbb{R}^{d,n})$, we have

$$\int_\Omega \langle u, \operatorname{Div} \phi \rangle \, dx = \int_\Omega \langle \phi, d\mu \rangle. \tag{2.13}$$

For the set of these functions, we write $BV(\Omega, \mathbb{R}^n)$ and we introduce the *total variation seminorm* of $u \in BV(\Omega, \mathbb{R}^n)$ by

$$\mathrm{TV}(u) := \sup \left\{ \int_\Omega \langle u, \operatorname{Div} \phi \rangle \, dx : \phi \in C_c^1(\Omega, \mathbb{R}^{d,n}), \|\phi(x)\|_F \leq 1 \right\}. \tag{2.14}$$

Note that $\mathrm{TV}(u) = \|Du\|_{\mathfrak{M}}$ if we denote the distributional derivative of u by $Du \in \mathfrak{M}(\Omega, \mathbb{R}^{d,n})$. The vector space $BV(\Omega, \mathbb{R}^n)$ equipped with the norm $\|f\|_{\mathrm{BV}} := \|f\|_{L^1} + \mathrm{TV}(f)$ is a Banach space that is, remarkably, neither separable nor the completion of $C^1(\Omega, \mathbb{R}^n)$ with respect to $\|\cdot\|_{BV}$ [AFP00, Chapter 3]. In fact, $\|f\|_{\mathrm{BV}} = \|f\|_{W^{1,1}}$, whenever f is in the Sobolev space $W^{1,1}(\Omega, \mathbb{R}^n)$.

We illustrate the relationship with sharp discontinuities, as they occur in image processing, with the following property: If U is compactly contained in Ω with C^1-boundary ∂U and $u \colon \Omega \to \mathbb{R}^n$ is a *cartoon-like jump function* that takes only the values $u^+, u^- \in \mathbb{R}^n$,

$$u(x) := \begin{cases} u^+, & x \in U, \\ u^-, & x \in \Omega \setminus U, \end{cases} \tag{2.15}$$

then the distributional derivative Du is concentrated on the $(d-1)$-dimensional jump set ∂U with $\mathrm{TV}(u) = \mathcal{H}^{d-1}(\partial U) \cdot \|u^+ - u^-\|$. Hence, in this case, the total variation seminorm is the length or area of the jump set times the jump size. Similar properties hold for more general piecewise smooth functions and more general jump sets [AFP00, Section 3.9].

2.4 Manifolds

In this work, all manifolds are *smooth Riemannian manifolds* of a fixed finite dimension s. We often denote manifolds by the symbol $\mathcal{M}$. Usually, we can assume that $\mathcal{M}$ is smoothly embedded into $\mathbb{R}^N$ by the Whitney embedding theorem [Lee13, Thm. 6.15]. Furthermore, $\mathcal{M}$ is metrized by the geodesic distance $d_{\mathcal{M}}(\cdot,\cdot)$. The *exponential map* $\exp_p\colon T_p\mathcal{M} \to \mathcal{M}$ maps from the tangent space $T_p\mathcal{M}$ at a point $p \in \mathcal{M}$ to the point $\exp_p(v) \in \mathcal{M}$ that is reached by the geodesic curve with initial velocity $v \in T_p\mathcal{M}$ in unit time. The exponential map is invertible in a neighborhood $U \subset T_p\mathcal{M}$ of 0 and its inverse map is the *inverse exponential* or *logarithmic map* $\log_p\colon \exp_p(U) \to T_p\mathcal{M}$. Particular manifolds that will appear in this work are the compact s-dimensional spheres $\mathbb{S}^s$ in $\mathbb{R}^{s+1}$ and the compact Lie group of n-dimensional special orthogonal matrices $SO(n)$.

2.5 Convex Analysis

Convexity for functions on a topological vector space X that take values in the extended real line $\overline{\mathbb{R}} := [-\infty, \infty]$ is defined via convexity of their epigraphical sets. For a proper function $f\colon X \to \overline{\mathbb{R}}$, the *convex conjugate* or *Legendre-Fenchel transformation*

$$f^*(y) := \sup_{x \in X} \{\langle y, x\rangle + f(x)\} \tag{2.16}$$

is a proper, lower semicontinuous and convex function on the topological dual space X^*. For a reflexive space X, the biconjugate f^{**} is the lower semicontinuous *convex envelope* or *convex hull* of f. For a Hilbert space X, the *proximal mapping* of a proper, lower semicontinuous and convex function $f\colon X \to \overline{\mathbb{R}}$ is

$$\operatorname{prox}_f(\bar{x}) := \operatorname*{argmin}_x \left\{ \frac{1}{2}\|x - \bar{x}\|_X^2 + f(x) \right\} \in X \tag{2.17}$$

We give some examples of proximal mappings and convex conjugates that are used in this work:

- For a lower-semicontinuous and proper convex function f, the proximal mapping of f^* satisfies, for each $t > 0$,

$$\operatorname{prox}_{tf^*}(\bar{x}) = \bar{x} - t \operatorname{prox}_{t^{-1}f}(t^{-1}\bar{x}). \tag{2.18}$$

- For $C \subset X$ convex and closed, let $f(x) = \delta_C(x)$ be the *indicator function of* C which is 0 on C and ∞ otherwise. Then $f^*(y) = \sigma_C(y) := \sup_{y \in C}\langle y, x\rangle$ is the

support function of C and $\operatorname{prox}_f(\bar{x}) = \operatorname{proj}_C(\bar{x})$ is the *orthogonal projection* (with respect to the Hilbert space geometry of X) onto C.

- If $f(x) = \|x\|_X$ is a norm, then $f^*(y) = \delta_{B^*}(y)$ is the indicator function of the closed unit norm ball B^* in the dual norm $\|y\|_{X^*} := \sup_{\|x\|_X \leq 1} \langle y, x \rangle$.

- If $X = \mathbb{R}^n$, $p \in [1, \infty]$ and $f(x) = \|x\|_p$ is the vector l^p-norm, then $f^*(y) = \|y\|_q$ where $1/p + 1/q = 1$.

- For $X = \mathbb{R}^n$, $v \in X$ and $f(x) = \|x - v\|^2/2$, we have $f^*(y) = \|y\|^2 + \langle v, y \rangle$ and, for $t > 0$, $\operatorname{prox}_{tf}(\bar{x}) = (\bar{x} + v)/(1 + t)$.

2.6 Numerical Optimization

For the numerical experiments in this work, we rely on the primal-dual splitting algorithm known as (modified) *Primal-Dual Hybrid Gradient* (PDHG) [ZC08; EZC10] or *Chambolle-Pock* [CP11] algorithm:

$$\begin{aligned} y^{k+1} &= \operatorname{prox}_{\sigma^k f^*}(y^k + \sigma^k A \bar{x}^k), \\ x^{k+1} &= \operatorname{prox}_{\tau^k g}(x^k - \tau^k A^* y^{k+1}), \\ \bar{x}^{k+1} &= x^{k+1} + \theta(x^{k+1} - x^k). \end{aligned}$$

It is applicable to convex models with objective functions $F \colon \mathbb{R}^N \to \overline{\mathbb{R}}$ of the form $F(x) = g(x) + f(Ax)$, where g and f are convex while A is linear. Extensions to nonlinear operators A exist [Val14; CV17]. The algorithm converges to a saddle point (x^*, y^*) of the minimax problem

$$\min_x \max_y \quad g(x) + \langle Ax, y \rangle - f^*(y),$$

so that x^* is a minimizer of $F(x) = g(x) + f(Ax)$ and y^* maximizes the dual objective $G(y) = -g^*(A^*y) - f^*(y)$.

Throughout this work, we set $\theta = 1$ and choose the step sizes τ^k and σ^k according to the update scheme introduced in [GEB13; Gol+13; GLY15]. More precisely, we numerically estimate the operator norm $\|A\|_{\sigma,\infty}$ using power (von Mises) iterations applied to A^*A [MP29; Ste73] and initialize $\tau^0 = \sigma^0 = \sqrt{\|A\|_{\sigma,\infty}}$ and $\alpha^0 = 1/2$. In each iteration, we

compute the primal and dual residuals

$$R_p = \|(x^k - x^{k+1})/\tau^k - A^*(y^k - y^{k+1})\|,$$
$$R_d = \|(y^k - y^{k+1})/\sigma^k - \theta A(x^k - x^{k+1})\|,$$

and update the step sizes according to

$$\tau^{k+1} = \tau^k/(1-\alpha^k),$$
$$\sigma^{k+1} = \sigma^k \cdot (1-\alpha^k),$$
$$\alpha^{k+1} = \eta\alpha^k,$$

if $R_p/R_d > s\Delta$, and according to

$$\tau^{k+1} = \tau^k \cdot (1-\alpha^k),$$
$$\sigma^{k+1} = \sigma^k/(1-\alpha^k),$$
$$\alpha^{k+1} = \eta\alpha^k,$$

if $R_p/R_d < s/\Delta$. Here $\eta = 0.95$, $\Delta = 3/2$ and s is an appropriate scaling factor.

We stop the iteration as soon as the relative primal dual gap $(g(y^k) - f(x^k))/g(y^k)$ drops below a given tolerance [BV04, Chapter 5.5]. Alternatively, if the evaluation of the objectives f or g involves itself a minimization, we stop as soon as the residuals R_p and R_d drop below a given tolerance to avoid large numerical errors in the evaluation of f and g.

2.7 Discretization of Measure-Valued Functions

At the heart of all models discussed in this work are measure-valued functions. When it comes to the numerical implementation of these models, a reasonable discretization of measure-valued functions is crucial. Since we will be concerned with measures on subsets of Euclidean space or of Riemannian manifolds, we formulate our discretization scheme for Riemannian manifolds, which includes Euclidean space.

Furthermore, we formulate the discretization for submanifolds of $\mathbb{R}^N$, which is no restriction by the Whitney embedding theorem [Lee13, Thm. 6.15], and employ the notation from surface finite element methods [DE13]. For an s-dimensional submanifold $\mathcal{M}$ of $\mathbb{R}^N$ and $\Omega \subset \mathbb{R}^d$ open and bounded, differentiable functions $u\colon \Omega \to \mathcal{M}$ are regarded as a subset of differentiable functions with values in $\mathbb{R}^N$. For such functions, a Jacobian

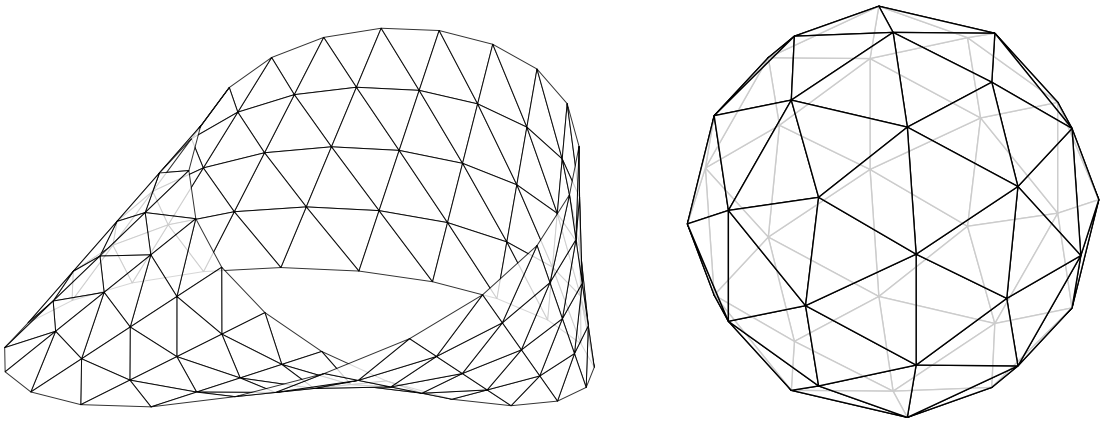


Figure 2.1: Triangulated approximations of the Moebius strip **(left)** and the two-dimensional sphere **(right)** as surfaces embedded into $\mathbb{R}^3$.

$\nabla u(x) \in \mathbb{R}^{d,N}$ in the Euclidean sense exists that can be identified with the push-forward of the tangent space $T_x\Omega$ to $T_{u(x)}\mathcal{M}$, i.e., for each $x \in \Omega$ and $\xi \in \mathbb{R}^d = T_x\Omega$, we have

$$(\nabla u(x))^\mathsf{T} \xi \in T_{u(x)}\mathcal{M} \subset T_{u(x)}\mathbb{R}^N. \tag{2.19}$$

On the other hand, for differentiable maps $p\colon \mathcal{M} \to \mathbb{R}^d$, there exists an extension of p to a neighborhood of $\mathcal{M} \subset \mathbb{R}^N$ that is constant in normal directions [BC92, p. 656], and we denote by $\nabla p(z) \in \mathbb{R}^{N,d}$ the Jacobian of this extension evaluated at $z \in \mathcal{M}$. Being constant in normal directions means that $(\nabla p(z))^\mathsf{T}\zeta = 0$ whenever $\zeta \in N_z\mathcal{M}$, the orthogonal complement of $T_z\mathcal{M}$ in $\mathbb{R}^N$. This property ensures that the value of $\nabla p(z)$ does not depend on the choice of extension.

The manifold $\mathcal{M} \subset \mathbb{R}^N$ is approximated by a triangulated topological manifold $\mathcal{M}_h \subset \mathbb{R}^N$ in the sense that there is some homeomorphism $\iota\colon \mathcal{M}_h \to \mathcal{M}$ (Fig. 2.1 and 2.2). The following discretization does not depend on the choice of homeomorphism ι. By $\mathcal{T}_h$, we denote the set of simplices that make up $\mathcal{M}_h$:

$$\bigcup_{T \in \mathcal{T}_h} T = \mathcal{M}_h. \tag{2.20}$$

For $T, \tilde{T} \in \mathcal{T}_h$, either $T \cap \tilde{T} = \emptyset$ or $T \cap \tilde{T}$ is an $(s-k)$-dimensional face for $k \in \{1, \dots, s\}$. Each simplex $T \in \mathcal{T}_h$ spans an s-dimensional linear subspace of $\mathbb{R}^N$ and there is an orthogonal basis representation $P_T \in \mathbb{R}^{s,N}$ of vectors in $\mathbb{R}^N$ to that subspace. Furthermore, we enumerate the vertices of the triangulation as $Z^1, \dots, Z^L \in \mathcal{M} \cap \mathcal{M}_h$.

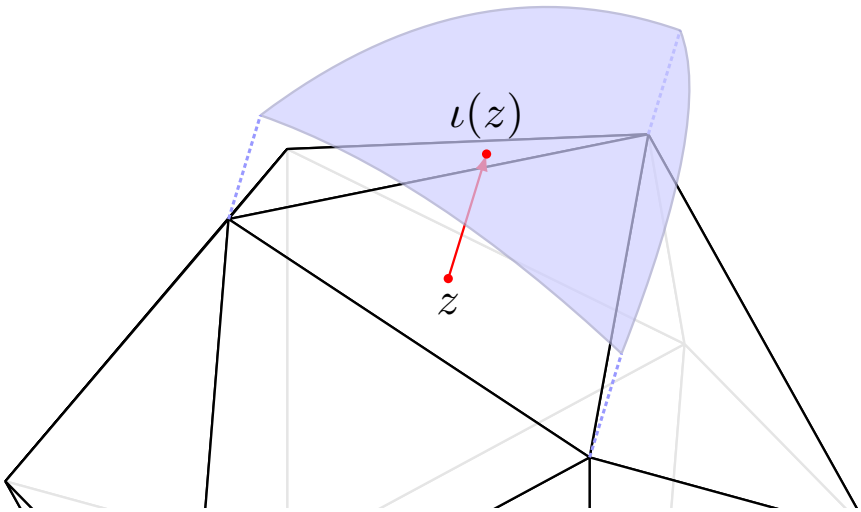


Figure 2.2: Each simplex T in a triangulation (black wireframe plot) is in homeomorphic correspondence to a piece $\iota(T)$ of the original manifold (blue) through the map $\iota\colon \mathcal{M}_h \to \mathcal{M}$.

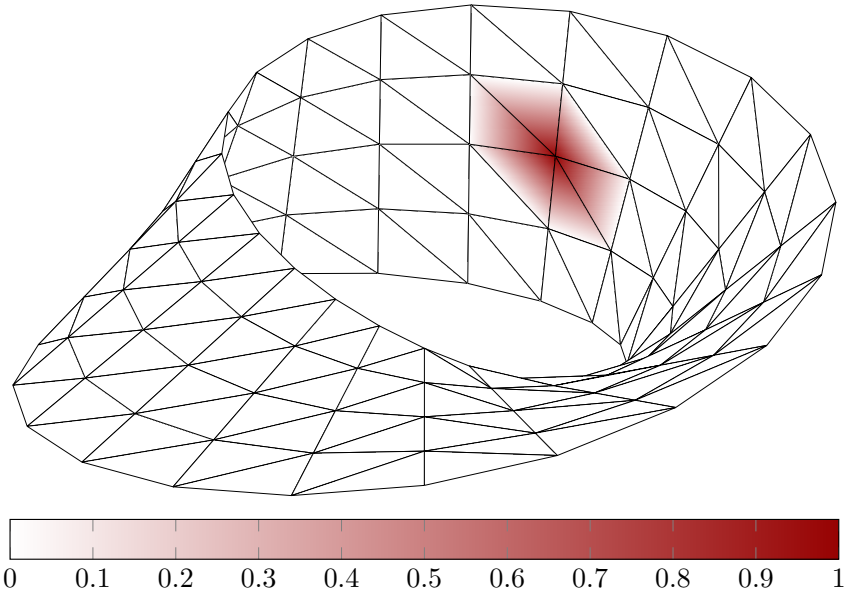


Figure 2.3: The first-order finite element space S_h is spanned by a nodal basis $\chi_1, \ldots, \chi_L$ which is uniquely determined by the property $\chi_k(Z^l) = 1$ if $k = l$ and $\chi_k(Z^l) = 0$ otherwise. The illustration shows a triangulation of the Moebius strip with a color plot of a nodal basis function.

For the numerics, we assume the first-order finite element space

$$S_h := \{\phi_h \in C(\mathcal{M}_h) : \phi_h|_T \text{ is affine linear for each } T \in \mathcal{T}_h\}. \tag{2.21}$$

The functions in S_h are piecewise differentiable on $\mathcal{M}_h$, and we define the surface gradient $\nabla_T \phi_h \in \mathbb{R}^N$ of $\phi_h \in S_h$ by the gradient of the linear affine extension of $\phi_h|_T$ to $\mathbb{R}^N$. However, most of the time, we will consider the projected surface gradient $P_T \nabla_T \phi_h \in \mathbb{R}^s$ where $P_T \in \mathbb{R}^{s,N}$ is the orthogonal basis representation of vectors in $\mathbb{R}^N$ in the subspace spanned by the simplex $T \in \mathcal{T}_h$. If L is the number of vertices in the triangulation of $\mathcal{M}_h$, then S_h is a linear space of dimension L with nodal basis $\chi_1, \dots, \chi_L$ which is uniquely determined by the property $\chi_k(Z^l) = 1$ if $k = l$ and $\chi_k(Z^l) = 0$ otherwise (Fig. 2.3).

The dual space of S_h, which we denote by $\mathfrak{M}_h(\mathcal{M}_h)$, is a space of signed measures. We identify $\mathfrak{M}_h(\mathcal{M}_h) = \mathbb{R}^L$ via dual pairing with the nodal basis $\chi_1, \dots, \chi_L$, i.e., to each $\mu_h \in \mathfrak{M}_h(\mathcal{M}_h)$ we associate the vector $(\langle \mu_h, \chi_1 \rangle, \dots, \langle \mu_h, \chi_L \rangle)$. We then replace the space $\mathcal{P}(\mathcal{M})$ of probability measures over $\mathcal{M}$ by the convex subset

$$\mathcal{P}_h(\mathcal{M}_h) = \left\{ \mu_h \in \mathfrak{M}_h(\mathcal{M}_h) : \mu_h \geq 0, \sum_{k=1}^{L} \langle \mu_h, \chi_k \rangle = 1 \right\}. \tag{2.22}$$

Finally, for the fully discrete setting, the d-dimensional image domain Ω, $d = 2, 3$ of a measure-valued function $u \colon \Omega \to \mathcal{P}(\mathcal{M})$ is replaced by a Cartesian rectangular grid using N points $X^1, \dots, X^N \in \Omega$. Differentiation in Ω is performed on a staggered grid with homogeneous Neumann boundary conditions such that the dual operator to the differential operator ∇_x is the negative divergence with vanishing boundary values. We choose homogeneous Neumann boundary conditions as an alternative to Dirichlet boundary conditions that require to fix values at the image boundary that are often not available, in particular for manifold-valued images.

3 Measure-Valued Image Processing with Applications to Q-Ball Imaging

In this chapter, we are concerned with variational problems in which the unknown function $u\colon \Omega \to \mathcal{P}(\mathbb{S}^2)$ maps from an open and bounded set $\Omega \subseteq \mathbb{R}^3$, the *image domain*, into the set of Borel *probability measures* $\mathcal{P}(\mathbb{S}^2)$ on the two-dimensional unit sphere $\mathbb{S}^2$ as the set of spatial directions or, more generally, on some metric space: each value $u_x := u(x) \in \mathcal{P}(\mathbb{S}^2)$ is a Borel probability measure on $\mathbb{S}^2$, and can be viewed as a distribution of directions in $\mathbb{R}^3$.

Such measures $\mu \in \mathcal{P}(\mathbb{S}^2)$, in particular when represented using density functions, are known as *orientation distribution functions* (ODFs). We will keep to the term due to its popularity in the literature, although we will be mostly concerned with *measures* instead of functions on $\mathbb{S}^2$. Accordingly, an *ODF-valued image* is a function $u\colon \Omega \to \mathcal{P}(\mathbb{S}^2)$. ODF-valued images appear in reconstruction schemes for diffusion-weighted magnetic resonance imaging (MRI), such as Q-ball imaging (QBI) [Tuc04] and constrained spherical deconvolution (CSD) [Tou+04].

3.1 Diffusion-Weighted MRI and Q-Ball Imaging

In diffusion-weighted (DW) magnetic resonance imaging (MRI), the diffusivity of water in biological tissues is measured non-invasively. In medical applications where tissues exhibit fibrous microstructures, such as muscle fibers or axons in cerebral white matter, the diffusivity contains valuable information about the fiber architecture [Mai+17; Sch+19]. For DW measurements, several full 3D MRI volumes are acquired with varying magnetic field gradients that are able to sense diffusion [Chi+15].

Under the assumption of anisotropic Gaussian diffusion, positive definite matrices (tensors) can be used to describe the diffusion in each voxel. This model, known as *diffusion tensor imaging* (DTI) [BML94], requires few measurements while giving a good estimate of the main diffusion direction in the case of well-aligned fiber directions. However, crossing and branching of fibers at a scale smaller than the voxel size, also called intra-voxel orientational heterogeneity (IVOH), often occurs in human cerebral white matter due to

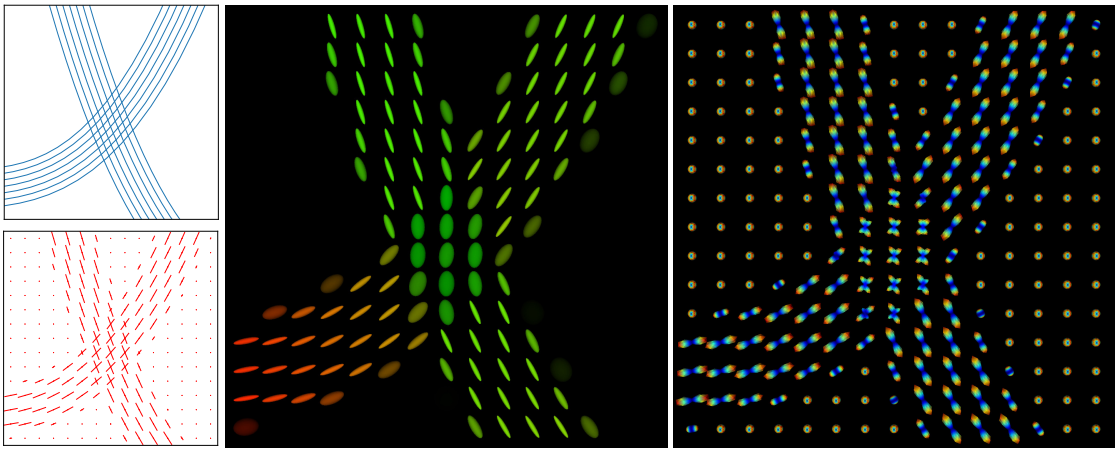


Figure 3.1: **Top left:** 2-D fiber phantom of two crossing fiber bundles. **Bottom left:** Peak directions on a 15×15 grid, derived from the phantom and used for the generation of synthetic HARDI data. **Center:** The diffusion tensor (DTI) reconstruction approximates diffusion directions in a parametric way using tensors, visualized as ellipsoids. **Right:** The QBI-CSA ODF reconstruction represents fiber orientation using probability measures at each point, which allows to accurately recover fiber crossings in the center region.

the relatively large (millimeter-scale) voxel size of DW-MRI data. Therefore, DTI data is insufficient for accurate fiber tract mapping in regions with complex fiber crossings (Fig. 3.1).

By increasing the number of applied magnetic field gradients, more accurate restoration of IVOH is possible with more refined approaches that are based on *high angular resolution diffusion imaging* (HARDI) [Tuc+02] measurements. Reconstruction schemes for HARDI data yield *orientation distribution functions (ODFs)* instead of tensors. In *Q-ball imaging* (QBI) [Tuc04], an ODF is interpreted to be the marginal probability of diffusion in a given direction [ALS09]. In contrast, ODFs in *constrained spherical deconvolution* (CSD) approaches [Tou+04], also denoted *fiber ODFs*, estimate the density of fibers per direction for each voxel of the volume.

In all of these approaches, ODFs are modeled as antipodally symmetric functions or measures on the sphere, which could be modeled just as well on the projective space, i.e., a sphere where antipodal points are identified. Novel approaches [Del+07; EOK11; RKK12; KÖU18] allow for asymmetric ODFs to account for intra-voxel geometry. Therefore, we keep to modeling ODFs on a sphere even though all results presented in this work could be easily adapted to models on the projective space.

3.1.1 Variational Models for Orientation Distributions

As a common denominator, in the above applications, reconstructing orientation distributions rather than a single orientation at each point allows to recover directional information of structures, such as vessels or nerve fibers, that may overlap or have crossings: For a given set of directions $A \subset \mathbb{S}^2$ and a measure-valued function $u\colon \Omega \to \mathcal{P}(\mathbb{S}^2)$, the integral $\int_A du_x$, $u_x := u(x)$, describes the fraction of fibers crossing the point $x \in \Omega$ that are oriented in any of the given directions $v \in A$.

However, modeling ODFs as probability measures in a non-parametric way is surprisingly difficult. We propose a new formulation of the classical total variation seminorm (TV) [AFP00; Cha+10] for nonparametric QBI that allows to formulate the variational restoration model

$$\inf_{u:\Omega\to\mathcal{P}(\mathbb{S}^2)} \int_\Omega \rho(x, u_x)\, dx + \lambda\, \mathrm{TV}_{W_1}(u), \tag{3.1}$$

with various pointwise data fidelity terms

$$\rho\colon \Omega \times \mathcal{P}(\mathbb{S}^2) \to [0, \infty). \tag{3.2}$$

This involves in particular a non-parametric concept of total variation for ODF-valued functions that is mathematically robust and computationally feasible: The idea is to build upon the TV-formulations developed in the context of functional lifting [Lel+13b]

$$\sup \left\{ \int_\Omega - \operatorname{div}_x p(x, z)\, u_x(dz)\, dx : \ p \in C_c^1(\Omega \times \mathbb{S}^2; \mathbb{R}^3),\ \|\nabla_z p\|_\infty \le 1 \right\}. \tag{3.3}$$

One distinguishing feature of this approach is that it is applicable to arbitrary Borel probability measures. In contrast, existing mathematical frameworks for QBI and CSD generally follow the standard literature on the physics of MRI [Cal91, p. 330] in assuming ODFs to be given by a *probability density function* in $L^1(\mathbb{S}^2)$, often with an explicit parametrization. As an example of one such approach, we point to the fiber continuity regularizer [RS13] which is defined for ODF-valued functions u where, for each $x \in \Omega$, the measure u_x can be represented by a probability density function $\phi(x, \cdot)$ on $\mathbb{S}^2$:

$$R_{\mathrm{FC}}(u) := \int_\Omega \int_{\mathbb{S}^2} (z \cdot \nabla_x \phi(x, z))^2\, dz\, dx. \tag{3.4}$$

A rigorous generalization of this functional to measure-valued functions for arbitrary Borel probability measures is not straightforward, as it requires a notion of spatial gradients, similar to $\nabla_x \phi$, for measures that are not absolutely continuous with respect to

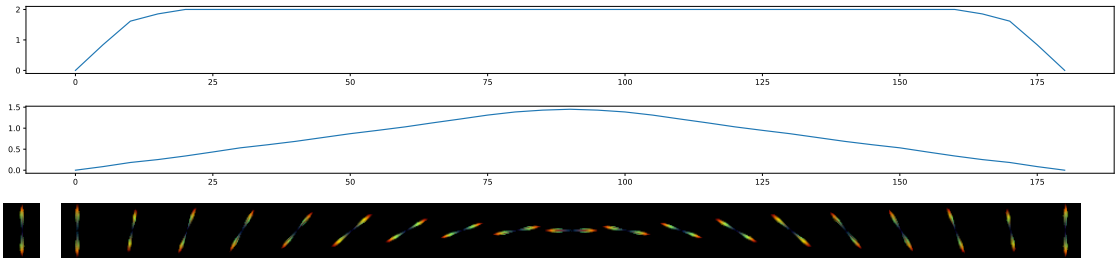


Figure 3.2: **Horizontal axis:** Angle of main diffusion direction relative to the reference diffusion profile in the bottom left corner. **Vertical axis:** Distances of the ODFs in the bottom row to the reference ODF in the bottom left corner (L^1-distances in the top row and W^1-distance in the second row). L^1-distances do not reflect the linear change in direction, whereas the W^1-distance exhibits an almost-linear profile. L^p-distances for other values of p, such as $p = 2$, show a behavior similar to L^1-distances.

the Lebesgue measure.

While practical, the probability density-based approach raises some modeling questions, which lead to deeper mathematical issues. In particular, comparing probability densities using the popular L^p-norm-based data fidelity terms – in particular the squared L^2-norm – does not incorporate the structure naturally carried by probability densities, such as nonnegativity and unit total mass, and ignores metric information about $\mathbb{S}^2$. To illustrate the last point, assume that two probability measures are given in terms of density functions $f, g \in L^p(\mathbb{S}^2)$ satisfying $\operatorname{supp}(f) \cap \operatorname{supp}(g) = \emptyset$, i.e., having disjoint support on $\mathbb{S}^2$. Then $\|f - g\|_{L^p} = \|f\|_{L^p} + \|g\|_{L^p}$, irrespective of the size and relative position of the supporting sets of f and g on $\mathbb{S}^2$.

One would prefer to use statistical metrics, such as optimal transport metrics (see Sect. 2.2), that properly take into account distances on the underlying set $\mathbb{S}^2$ (Fig. 3.2). However, replacing the L^p-norm with such a metric in density-based variational imaging formulations will generally lead to ill-posed minimization problems, as the minimum might not be attained in $L^p(\mathbb{S}^2)$, but possibly in $\mathcal{P}(\mathbb{S}^2)$ instead.

Therefore, it is interesting to investigate whether one can derive a mathematical basis for variational image processing with ODF-valued functions without making assumptions about the parametrization of ODFs nor assuming ODFs to be given by density functions.

3.1.2 Contribution

We derive a rigorous mathematical framework for a generalization of the total variation seminorm to Banach space-valued and, as a special case, ODF-valued functions on a

finite-dimensional Euclidean domain $\Omega \subset \mathbb{R}^d$. Well-definedness of the TV-seminorm and of variational problems of the form (3.1) is established by carefully considering measurability of the functions involved. Furthermore, a functional-analytic explanation for the dual structure that is inherent in (3.3) is given. Building on this framework, we show existence of minimizers and discuss properties of TV, such as rotational invariance and the behavior on cartoon-like jump functions.

We demonstrate that our framework can be numerically implemented as a primal-dual saddle point problem involving only convex functions. Applications to synthetic and real-world data sets show significant reduction of noise as well as qualitatively convincing results when combined with existing ODF-based imaging approaches, including Q-ball and CSD.

3.1.3 Related Work

The high angular resolution of HARDI results in a large amount of noise compared with DTI. Moreover, most QBI and CSD models reconstruct the ODFs in each voxel separately. Consequently, HARDI data is a particularly interesting target for post-processing in terms of denoising and regularization in the sense of contextual processing. Some techniques apply a total variation or diffusive regularization to the HARDI signal before ODF reconstruction [McG+09; KTV10; DF11; Bec+12] and others regularize in a post-processing step [Del+07; Dui+13; WDS16].

Variational Regularization of DW-MRI Data

A Mumford-Shah model for edge-preserving restoration of Q-ball data was introduced in [WDS16]. There, jumps were penalized using the Fisher-Rao metric, which depends on a parametrization of ODFs as discrete probability distribution functions on sampling points of the sphere. Furthermore, the Fisher-Rao metric does not take the metric structure of $\mathbb{S}^2$ into consideration and is not amenable to biological interpretations [Ncu11]. Our formulation avoids any parametrization-induced bias.

The recent TV-based regularization scheme [OCW14] for the reconstruction of Q-ball images from HARDI data relies on the underlying parametrization of ODFs by spherical harmonics basis functions. Similarly, DTI-based models, such as the second-order model for regularizing general manifold-valued data [Bač+16], make use of an explicit approximation using positive semidefinite matrices, which the proposed model avoids.

The application of spatial regularization to CSD reconstruction is known to signifi-

cantly enhance the results [Dad+14]. However, total variation [Can+15] and other regularizers [HR15] are based on a representation of ODFs by square-integrable probability density functions instead of the mathematically more general probability measures that we base our method on.

Regularization of DW-MRI by Linear Diffusion

In another approach, the orientational part of ODF-valued images is included in the image domain so that images are identified with functions $U\colon \mathbb{R}^3 \times \mathbb{S}^2 \to \mathbb{R}$ that allow for contextual processing via PDE-based models on the space of positions and orientation or, more precisely, on the group $SE(3)$ of 3D rigid motions. This technique comes from the theory of stochastic processes on the coupled space $\mathbb{R}^3 \times \mathbb{S}^2$. In this context, it has been applied to the problems of contour completion [MS13] and contour enhancement [DF11; Dui+13]. Its practical relevance in clinical applications has been demonstrated [Prč+15].

This approach has been used to enhance the quality of CSD as a prior in a variational formulation [RS13] or in a post-processing step [Por+15] that also includes additional angular regularization. Due to the linearity of the underlying linear PDE, convolution-based explicit solution formulas are available [DF11; PD17]. Implemented efficiently [Mee+16b; Mee+16a], they outperform our more computationally demanding model, which is not tied to the specific application of DW-MRI, but allows arbitrary metric spaces. Furthermore, nonlinear Perona and Malik extensions to this technique have been studied [Cre+13] that do not allow for explicit solutions.

As an important distinction, in these approaches, spatial location and orientation are coupled in the regularization. Since our model starts from the more general setting of measure-valued functions on an arbitrary metric space, rather than $\mathbb{S}^2$ only, it does not currently realize an equivalent coupling. An extension to anisotropic total variation for measure-valued functions might close this gap in the future.

In contrast to these diffusion-based methods, our approach is able to preserve edges by design, even though the coupling of positions and orientations is able to compensate for this shortcoming at least in part since edges in DW-MRI are, most of the time, oriented in parallel to the direction of diffusion. Furthermore, the diffusion-based methods are formulated for square-integrable density functions, excluding point masses. Our method avoids this limitation by operating on mathematically more general probability measures.

Other Related Theoretical Work

Variants of the Kantorovich-Rubinstein formulation of the Wasserstein distance that appears in our framework have been applied in [Lel+14; BL15] and, more recently, in [FLS16; FLS17] to the problems of real-, RGB- and manifold-valued image denoising.

Total variation regularization for functions on the space of positions and orientations was recently introduced in [CP19] based on [CMC17]. Similarly, the work and toolbox in [SR17] is concerned with the implementation of so-called *orientation fields* in 3D image processing.

A Dirichlet energy for measure-valued functions based on Wasserstein metrics was recently developed in the context of harmonic mappings in [Lav19a], which can be interpreted as a diffusive (L^2) version of our proposed (L^1) regularizer.

In the literature, Banach space-valued functions of bounded variation (BV) mostly appear as a special case of metric space-valued BV-functions as introduced in [Amb90]. Apart from that, the case of one-dimensional domains attracts some attention [DD11], and the case of Banach space-valued BV-functions defined on a metric space is studied in [Mir03]. In contrast to these approaches, we give a definition of Banach space-valued BV-functions that live on a *finite-dimensional Euclidean* domain. In analogy with the real-valued case, we formulate the TV seminorm by duality, inspired by the functional-analytic framework from the theory of lifting [II69] as used in the context of Young measures [Bal89].

Due to the functional-analytic approach, our model does not depend on the specific parametrization of the ODFs and can be combined with the QBI and CSD frameworks for ODF reconstruction from HARDI data, either in a post-processing step or during reconstruction. Combined with suitable data fidelity terms such as least-squares or Wasserstein distances, it allows for an efficient implementation using state-of-the-art primal-dual methods.

3.2 A Mathematical Framework for Measure-Valued Functions

Our work is motivated by the study of ODF-valued functions $u\colon \Omega \to \mathcal{P}(\mathbb{S}^2)$ for $\Omega \subset \mathbb{R}^3$ open and bounded. However, from an abstract viewpoint, the unit sphere $\mathbb{S}^2 \subset \mathbb{R}^3$ equipped with the metric induced by the Riemannian manifold structure [Lee13] – i.e., the distance between two points is the arc length of the great circle segment through the two points – is simply a particular example of a compact metric space.

As it turns out, most of the analysis only relies on this property. Therefore, in the following we generalize the setting of ODF-valued functions to the study of functions taking values in the space of Borel probability measures on an *arbitrary* compact metric space (instead of $\mathbb{S}^2$).

More precisely, throughout this section, let

1. $\Omega \subset \mathbb{R}^d$ be an open and bounded set, and let
2. (X, d) be a compact metric space, e.g., a compact Riemannian manifold equipped with the commonly-used metric induced by the geodesic distance, such as $X = \mathbb{S}^2$.

Not all of the statements below require boundedness of Ω and compactness of X. However, as we are ultimately interested in the case of $X = \mathbb{S}^2$ and rectangular image domains, we impose these restrictions. Apart from DW-MRI, one natural application of this generalized setting are two-dimensional ODFs where $d = 2$ and $X = \mathbb{S}^1$ which is similar to the setting introduced in [CP19] for the edge enhancement of color or grayscale images.

The goal of this section is a mathematically well-defined formulation of TV as given in (3.3) that exhibits all the properties that the classical total variation seminorm is known for: anisotropy (Prop. 3.5), preservation of edges and compatibility with piecewise-constant signals (Prop. 3.2). Furthermore, for variational problems as in (3.1), we give criteria for the existence of minimizers (Thm. 3.10) and discuss (non-)uniqueness (Prop. 3.11).

A well-defined formulation of TV as given in (3.3) requires a careful inspection of topological and functional analytic concepts from optimal transport and general measure theory. We present the theoretical background for a rigorous understanding of the notation and definitions underlying the notion of TV for Banach-space valued and measure-valued functions.

3.2.1 Functions of Bounded Variation

We first give a definition of the total variation seminorm TV for Banach space-valued functions, i.e., functions that take values in a Banach space, which a definition of TV for measure-valued functions will turn out to be a special case of.

Measurability of Banach Space-Valued Functions

Let $(V, \|\cdot\|_V)$ be a real Banach space with (topological) dual space V^*, i.e., V^* is the set of bounded linear operators from V to $\mathbb{R}$. The dual pairing is denoted by $\langle p, v\rangle := p(v)$ whenever $p \in V^*$ and $v \in V$.

We say that $u\colon \Omega \to V$ is *weakly measurable* [DU77, p. 41] if $x \mapsto \langle p, u(x)\rangle$ is measurable for each $p \in V^*$ and say that $u \in L^\infty_w(\Omega, V)$ if u is weakly measurable and essentially bounded in V, i.e.,

$$\|u\|_{\infty,V} := \operatorname{ess\,sup}_{x\in\Omega} \|u(x)\|_V < \infty. \tag{3.5}$$

Note that the essential supremum is well-defined even for non-measurable functions as long as the measure is complete. In our case, we assume the Lebesgue measure on Ω which is complete.

The following Lemma ensures that the integrand in (3.7) is measurable.

Lemma 3.1. *Assume that $u\colon \Omega \to V$ is weakly measurable and $p\colon \Omega \to V^*$ is weakly* continuous, i.e., for each $v \in V$, the map $x \mapsto \langle p(x), v\rangle$ is continuous. Then the map $x \mapsto \langle p(x), u(x)\rangle$ is measurable.*

Proof. Define $f\colon \Omega \times \Omega \to \mathbb{R}$ via

$$f(x, \xi) := \langle p(x), u(\xi)\rangle. \tag{3.6}$$

Then f is continuous in the first and measurable in the second variable. In the calculus of variations, functions with this property are called Carathéodory functions and have the property that $x \mapsto f(x, g(x))$ is measurable whenever $g\colon \Omega \to \Omega$ is measurable, which is proven by approximation of g as the pointwise limit of simple functions [Dac08, Prop. 3.7]. In our case we can simply set $g(x) := x$, which is measurable, and the assertion follows. □

Banach Space-Valued Functions of Bounded Variation

For weakly measurable functions $u\colon \Omega \to V$ with values in a Banach space V, that we will replace by a space of measures later on, we define,

$$\begin{aligned}\mathrm{TV}_V(u) := \sup \Bigg\{ \int_\Omega \langle -\operatorname{div} p(x), u(x)\rangle \, dx :\ & p \in C^1_c(\Omega, (V^*)^d), \\ & \forall x \in \Omega\colon \|p(x)\|_{(V^*)^d} \le 1 \Bigg\}.\end{aligned} \tag{3.7}$$

By V^*, we denote the (topological) dual space of V, i.e., V^* is the set of bounded linear operators from V to $\mathbb{R}$. The criterion $p \in C^1_c(\Omega, (V^*)^d)$ means that p is a compactly supported function on $\Omega \subset \mathbb{R}^d$ with values in the Banach space $(V^*)^d$ and the directional derivatives $\partial_i p\colon \Omega \to (V^*)^d$, $1 \le i \le d$, (in Euclidean coordinates) lie in $C_c(\Omega, (V^*)^d)$.

We write

$$\operatorname{div} p(x) := \sum_{i=1}^{d} \partial_i p_i(x). \tag{3.8}$$

Lemma 3.1 ensures that the integrals in (3.7) are well-defined. There is one subtlety about formulation (3.7) of the total variation: The choice of norm for the product space $(V^*)^d$ affects the properties of our total variation seminorm.

Properties of TV for Banach-Space Valued Functions

In this section, we show that the classical properties of the classical total variation seminorm continue to hold for definition (3.7) in the case of Banach space-valued functions.

Cartoon functions A reasonable demand is that the new formulation should behave similarly to the classical total variation on cartoon-like jump functions $u\colon \Omega \to V$,

$$u(x) := \begin{cases} u^+, & x \in U, \\ u^-, & x \in \Omega \setminus U, \end{cases} \tag{3.9}$$

for some fixed measurable set $U \subset \Omega$ with smooth boundary ∂U, and $u^+, u^- \in V$. The classical total variation assigns to such functions a penalty of

$$\mathcal{H}^{d-1}(\partial U) \cdot \|u^+ - u^-\|_V, \tag{3.10}$$

where the Hausdorff measure $\mathcal{H}^{d-1}(\partial U)$ describes the length or area of the jump set. The following proposition, which generalizes [VL17, Prop. 1], provides conditions on the norm $\|\cdot\|_{(V^*)^d}$ which guarantee this behavior.

Proposition 3.2. *Assume that U is compactly contained in Ω with C^1-boundary ∂U. Let $u^+, u^- \in V$ and let $u\colon \Omega \to V$ be defined as in* (3.9)*. If the norm $\|\cdot\|_{(V^*)^d}$ in* (3.7) *satisfies*

$$\left|\sum_{i=1}^{d} x_i \langle p_i, v\rangle\right| \le \|x\|_2 \|p\|_{(V^*)^d} \|v\|_V, \tag{3.11}$$

$$\|(x_1 q, \dots, x_d q)\|_{(V^*)^d} \le \|x\|_2 \|q\|_{V^*}, \tag{3.12}$$

whenever $q \in V^*$, $p \in (V^*)^d$, $v \in V$, and $x \in \mathbb{R}^d$, then

$$\mathrm{TV}_V(u) = \mathcal{H}^{d-1}(\partial U) \cdot \|u^+ - u^-\|_V. \tag{3.13}$$

Proof. Let $p\colon \Omega \to (V^*)^d$ satisfy the constraints in (3.7) and denote by ν the outer unit normal of ∂U. The set Ω is bounded, p and its derivatives are continuous and $u \in L^\infty_w(\Omega, V)$ since the range of u is finite and U, Ω are measurable. Therefore all of the following integrals converge absolutely. Due to linearity of the divergence,

$$\langle \operatorname{div} p(x), u^\pm \rangle = \operatorname{div}(\langle p(\cdot), u^\pm \rangle), \tag{3.14}$$

$$\langle p(x), u^\pm \rangle := (\langle p_1(x), u^\pm \rangle, \dots, \langle p_d(x), u^\pm \rangle) \in \mathbb{R}^d. \tag{3.15}$$

Using this property and applying Gauss' theorem, we compute

$$\begin{aligned} &\int_\Omega \langle -\operatorname{div} p(x), u(x) \rangle \, dx \\ &= -\int_{\Omega \setminus U} \operatorname{div}(\langle p(x), u^- \rangle) \, dx - \int_U \operatorname{div}(\langle p(x), u^+ \rangle) \, dx \\ &\overset{\text{Gauss}}{=} \int_{\partial U} \sum_{i=1}^d \langle \nu_i(x) p_i(x), u^+ - u^- \rangle \, \mathcal{H}^{d-1}(dx) \\ &\leq \mathcal{H}^{d-1}(\partial U) \cdot \|u^+ - u^-\|_V. \end{aligned} \tag{3.16}$$

For the last inequality, we used our first assumption on $\|\cdot\|_{(V^*)^d}$ together with the norm constraint for p in (3.7). Taking the supremum over p as in (3.7), we arrive at

$$\operatorname{TV}_V(u) \leq \mathcal{H}^{d-1}(\partial U) \cdot \|u^+ - u^-\|_V. \tag{3.17}$$

For the reverse inequality, let $\tilde{p} \in V^*$ be arbitrary with the property $\|\tilde{p}\|_{V^*} \leq 1$ and $\phi \in C^1_c(\Omega, \mathbb{R}^d)$ satisfying $\|\phi(x)\|_2 \leq 1$. Now, by (3.12), the function

$$p(x) := (\phi_1(x)\tilde{p}, \dots, \phi_d(x)\tilde{p}) \in (V^*)^d \tag{3.18}$$

has the properties required in (3.7). Hence,

$$\operatorname{TV}_V(u) \geq \int_\Omega \langle -\operatorname{div} p(x), u(x) \rangle \, dx \tag{3.19}$$

$$= -\int_\Omega \operatorname{div} \phi(x) \, dx \cdot \langle \tilde{p}, u^+ - u^- \rangle. \tag{3.20}$$

Taking the supremum over all $\phi \in C^1_c(\Omega, \mathbb{R}^d)$ satisfying $\|\phi(x)\|_2 \leq 1$, we obtain

$$\operatorname{TV}_V(u) \geq \operatorname{Per}(U, \Omega) \cdot \langle \tilde{p}, u^+ - u^- \rangle, \tag{3.21}$$

where $\mathrm{Per}(U,\Omega)$ is the perimeter of U in Ω. In the theory of *Caccioppoli sets* (or *sets of finite perimeter*), the perimeter is known to agree with $\mathcal{H}^{d-1}(\partial U)$ for sets with C^1-boundary [AFP00, p. 143].

Now, taking the supremum over all $\tilde{p} \in V^*$ with $\|\tilde{p}\|_{V^*} \leq 1$ and using the fact that the canonical embedding of a Banach space into its bidual is isometric, i.e.,

$$\|u\|_V = \sup_{\|p\|_{V^*}\leq 1} \langle p, u\rangle, \tag{3.22}$$

we arrive at the desired reverse inequality, which concludes the proof. □

The following proposition gives some examples for norms that satisfy or fail to satisfy the conditions (3.11) and (3.12) in Prop. 3.2.

Proposition 3.3. *The following norms for $p \in (V^*)^d$ satisfy* (3.11) *and* (3.12) *for every normed space V:*

1. *For $s = 2$:*
$$\|p\|_{(V^*)^d,s} := \left(\sum_{i=1}^{d} \|p_i\|_{V^*}^s\right)^{1/s}. \tag{3.23}$$

2. *Writing $p(v) := (\langle p_1, v\rangle, \ldots, \langle p_d, v\rangle) \in \mathbb{R}^d$, $v \in V$,*
$$\|p\|_{\mathcal{L}(V,\mathbb{R}^d)} := \sup_{\|v\|_V\leq 1} \|p(v)\|_2. \tag{3.24}$$

On the other hand, for each $1 \leq s < 2$ and $s > 2$, there is a normed space V such that at least one of the properties (3.11), (3.12) *is not satisfied by the corresponding product norm* (3.23).

Remark 3.4. In the finite-dimensional Euclidean case $V = \mathbb{R}^n$ with norm $\|\cdot\|_2$, we have $(V^*)^d = \mathbb{R}^{d,n}$, thus p is matrix-valued and $\|\cdot\|_{\mathcal{L}(V,\mathbb{R}^d)}$ agrees with the spectral norm $\|\cdot\|_{\sigma,\infty}$. The norm defined in (3.23) is the Frobenius norm $\|\cdot\|_F$ for $s = 2$.

Prop. 3.3. By Cauchy-Schwarz,

$$\left|\sum_{i=1}^d x_i\langle p_i, v\rangle\right| \leq \|x\|_2 \left(\sum_{i=1}^d |\langle p_i, v\rangle|^2\right)^{1/2} \tag{3.25}$$

$$\leq \|x\|_2 \left(\sum_{i=1}^d \|p_i\|_{V^*}^2 \|v\|_V^2\right)^{1/2} \tag{3.26}$$

$$\leq \|x\|_2 \|v\|_V \left(\sum_{i=1}^d \|p_i\|_{V^*}^2\right)^{1/2}, \tag{3.27}$$

whenever $p \in (V^*)^d$, $v \in V$, and $x \in \mathbb{R}^d$. Similarly, for each $q \in V^*$,

$$\left(\sum_{i=1}^d \|x_i q\|_{V^*}^2\right)^{1/2} = \|x\|_2 \|q\|_{V^*}. \tag{3.28}$$

Hence, for $s = 2$, the properties (3.11) and (3.12) are satisfied by the product norm (3.23).

For the operator norm (3.24), consider

$$\begin{aligned}\left|\sum_{i=1}^d x_i \langle p_i, v\rangle\right| &\le \|x\|_2 \left(\sum_{i=1}^d |\langle p_i, v\rangle|^2\right)^{1/2} && (3.29)\\ &= \|x\|_2 \|p(v)\|_2 && (3.30)\\ &\le \|x\|_2 \|p\|_{\mathcal{L}(V,\mathbb{R}^d)} \|v\|_V, && (3.31)\end{aligned}$$

which is property (3.11). On the other hand, (3.12) follows from

$$\begin{aligned}\|(x_1 q, \dots, x_d q)\|_{\mathcal{L}(V,\mathbb{R}^d)} &= \sup_{\|v\|_V \le 1} \left(\sum_{i=1}^d |x_i q(v)|^2\right)^{1/2} && (3.32)\\ &= \|x\|_2 \sup_{\|v\|_V \le 1} |q(v)| && (3.33)\\ &= \|x\|_2 \|q\|_{V^*}. && (3.34)\end{aligned}$$

Now, for $s > 2$, property (3.11) fails for $d = 2$, $V = V^* = \mathbb{R}$, $p = x = (1, 1)$ and $v = 1$ since

$$\left|\sum_{i=1}^d x_i \langle p_i, v\rangle\right| = 2 > 2^{1/2} \cdot 2^{1/s} = \|x\|_2 \|p\|_{(V^*)^d,s} \|v\|_V. \tag{3.35}$$

For $1 \le s < 2$, consider $d = 2$, $V^* = \mathbb{R}$, $q = 1$ and $x = (1, 1)$, then

$$\|(x_1 q, \dots, x_d q)\|_{(V^*)^d,s} = 2^{1/s} > 2^{1/2} = \|x\|_2 \|q\|_{V^*}, \tag{3.36}$$

which contradicts property (3.12). □

The classical total variation has the property that the distributional derivative Du of a function $u\colon \Omega \to \mathbb{R}^n$ of bounded variation is given by a vector measure, i.e., $Du \in \mathfrak{M}(\Omega, \mathbb{R}^n)$. The cartoon-like jump function property (3.13) then follows from general *decomposition* theorems for vector measures [AFP00, Section 3.9]. In case of Banach space-valued functions $u\colon \Omega \to V$, we generally expect the distributional derivative Du to be a vector measure with values in the bidual space V^{**} by the well-known representation

theorem [DU77, p. 182], and a decomposition theorem continues to hold for Banach space-valued measures [DU77, p. 31]. In this chapter, our main application is the case where V is a space of measures. In that case, V^{**} is not easily available and, therefore, we will not go into the details about decomposition theorems, but leave an in-depth analysis of Banach space-valued functions of bounded variation for future work.

Rotational invariance The cartoon-like jump function property (3.13) is inherently rotationally invariant: we have $\mathrm{TV}_V(u) = \mathrm{TV}_V(\tilde{u})$ whenever $\tilde{u}(x) := u(Rx)$ for some $R \in SO(d)$ and u as in (3.9), with the domain Ω rotated accordingly. The reason is that the jump size is the same everywhere along the edge ∂U. More generally, we have the following proposition:

Proposition 3.5. *Assume that $\|\cdot\|_{(V^*)^d}$ satisfies the rotational invariance property*

$$\|p\|_{(V^*)^d} = \|Rp\|_{(V^*)^d} \quad \forall p \in (V^*)^d, R \in SO(d), \tag{3.37}$$

where $Rp \in (V^*)^d$ is defined via

$$(Rp)_i = \sum_{j=1}^{d} R_{ij} p_j \in V^*. \tag{3.38}$$

Then TV_V is rotationally invariant, i.e., $\mathrm{TV}_V(u) = \mathrm{TV}_V(\tilde{u})$ whenever $u \in L^\infty_w(\Omega, V)$ and $\tilde{u}(x) := u(Rx)$ for some $R \in SO(d)$.

Prop. 3.5. Let $R \in SO(d)$ and define

$$R^\mathsf{T}\Omega := \{R^\mathsf{T} x : x \in \Omega\}, \tag{3.39}$$

$$\tilde{p}(y) := R^\mathsf{T} p(Ry). \tag{3.40}$$

In (3.7), the norm constraint on $p(x)$ is equivalent to the norm constraint on $\tilde{p}(y)$ by condition (3.37). Now, consider the integral transform

$$\int_\Omega \langle -\operatorname{div} p(x), u(x)\rangle \, dx = \int_{R^\mathsf{T}\Omega} \langle -\operatorname{div} p(Ry), \tilde{u}(y)\rangle \, dy \tag{3.41}$$

$$= \int_{R^\mathsf{T}\Omega} \langle -\operatorname{div} \tilde{p}(y), \tilde{u}(y)\rangle \, dy. \tag{3.42}$$

where, using $R^\mathsf{T} R = I$,

$$\operatorname{div} \tilde{p}(y) = \sum_{i=1}^{d} \partial_i \tilde{p}_i(y) = \sum_{i=1}^{d} \sum_{j=1}^{d} R_{ji} \partial_i \left[p_j(Ry) \right] \tag{3.43}$$

$$= \sum_{i=1}^{d} \sum_{j=1}^{d} \sum_{k=1}^{d} R_{ji} R_{ki} \partial_k p_j(Ry) \tag{3.44}$$

$$= \sum_{j=1}^{d} \sum_{k=1}^{d} \partial_k p_j(Ry) \sum_{i=1}^{d} R_{ji} R_{ki} \tag{3.45}$$

$$= \sum_{j=1}^{d} \partial_j p_j(Ry) = \operatorname{div} p(Ry), \tag{3.46}$$

which implies $\mathrm{TV}_V(u) = \mathrm{TV}_V(\tilde{u})$. □

For $V = (\mathbb{R}^n, \|\cdot\|_2)$, property (3.37) in Prop. 3.5 is satisfied by the Frobenius norm as well as the spectral norms on $(V^*)^d = \mathbb{R}^{d,n}$. In general, the following proposition holds:

Proposition 3.6. *For every normed space V, the rotational invariance property* (3.37) *is satisfied by the operator norm* (3.24)*. For each $s \in [1, \infty)$, there is a normed space V such that the rotational invariance property* (3.37) *does not hold for the product norm* (3.23)*.*

Proof. By definition of the operator norm and rotational invariance of the Euclidean norm $\|\cdot\|_2$,

$$\|Rp\|_{\mathcal{L}(V,\mathbb{R}^d)} = \sup_{\|v\|_V \leq 1} \|Rp(v)\|_2 \tag{3.47}$$

$$= \sup_{\|v\|_V \leq 1} \|p(v)\|_2 = \|p\|_{\mathcal{L}(V,\mathbb{R}^d)}. \tag{3.48}$$

For the product norms (3.23), without loss of generality, we consider the case $d = 2$, $V := (\mathbb{R}^2, \|\cdot\|_1)$, $p_1 = (1, 0)$, $p_2 = (0, 1)$ and

$$R := \begin{pmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{pmatrix} \in SO(2). \tag{3.49}$$

Then $V^* := (\mathbb{R}^2, \|\cdot\|_\infty)$ and

$$\|p\|_{(V^*)^d,s} = \left(\textstyle\sum_{i=1}^{2} \|p_i\|_\infty^s \right)^{1/s} = 2^{1/s}, \tag{3.50}$$

whereas

$$(Rp)_1 = (1/2, -\sqrt{3}/2),\ (Rp)_2 = (\sqrt{3}/2, 1/2), \tag{3.51}$$

$$\|Rp\|_{(V^*)^d,s} = \left(\sum_{i=1}^{2} (\sqrt{3}/2)^s\right)^{1/s} = 2^{1/s} \cdot \sqrt{3}/2 \neq 2^{1/s} = \|p\|_{(V^*)^d,s}, \tag{3.52}$$

for each $1 \leq s < \infty$. □

Measure-Valued Functions of Bounded Variation

Recall that $\mathfrak{M}(X)$ and $\mathcal{P}(X) \subset \mathfrak{M}(X)$ denote the sets of signed Radon measures and Borel probability measures supported on X, and $\mathfrak{M}(X)$ can be identified with the (topological) dual space of $C_0(X) = C(X)$ (see Sect. 2.1). Hence, $\mathcal{P}(X)$ is a bounded subset of a dual space.

Furthermore, by Sect. 2.2, after subtracting a point mass at x_0, the set $\mathcal{P}(X) - \delta_{x_0}$ is a subset of the Banach space $\mathrm{KR}(X)$, the predual (in the sense that its dual space can be identified with a "meaningful" function space) of $\mathrm{Lip}_0(X)$.

Consequently, the embeddings

$$\mathcal{P}(X) \hookrightarrow (\mathrm{KR}(X), \|\cdot\|_{\mathrm{KR}}), \tag{3.53}$$

$$\mathcal{P}(X) \hookrightarrow (\mathfrak{M}(X), \|\cdot\|_{\mathfrak{M}}) \tag{3.54}$$

define two different topologies on $\mathcal{P}(X)$. The first embedding space $(\mathfrak{M}(X), \|\cdot\|_{\mathfrak{M}})$ is isometrically isomorphic to the dual of $C(X)$ while the second embedding space $(\mathrm{KR}(X), \|\cdot\|_{\mathrm{KR}})$ is a metrization of the weak*-topology on the bounded subset $\mathcal{P}(X)$ of the dual space $\mathfrak{M}(X) = C(X)^*$. Importantly, while $(\mathcal{P}(X), \|\cdot\|_{\mathfrak{M}})$ is not separable unless X is discrete, $(\mathcal{P}(X), \|\cdot\|_{\mathrm{KR}})$ is in fact compact, in particular complete and separable [Vil09, Thm. 6.18] which is crucial in our result on the existence of minimizers (Thm. 3.10).

Now, for $u\colon \Omega \to \mathcal{P}(X)$, our proposed definition (3.7) of TV for Banach space-valued functions is

$$\begin{aligned}\mathrm{TV}_{\mathrm{KR}}(u) := \sup \Bigg\{ \int_\Omega \langle -\operatorname{div} p(x), u(x)\rangle \, dx :\ & p \in C_c^1(\Omega, [\mathrm{Lip}_0(X)]^d), \\ & \|p(x)\|_{[\mathrm{Lip}_0(X)]^d} \leq 1 \Bigg\}.\end{aligned} \tag{3.55}$$

It can be regarded as a variant of the TV-formulation (3.3) introduced in the context of functional lifting [Lel+13b].

We give some considerations about the choice of product norm $\|\cdot\|_{[\mathrm{Lip}_0(X)]^d}$: For $p \in [\mathrm{Lip}_0(X)]^d$, the most natural choice is

$$[p]_{\mathrm{Lip}(X,\mathbb{R}^d)} := \sup_{z \neq z'} \frac{\|p(z) - p(z')\|_2^2}{d(z, z')}, \tag{3.56}$$

which is automatically rotationally invariant. On the other hand, the product norm defined in (3.23) (with $s = 2$), namely $\sqrt{\sum_{i=1}^d [p_i]_{\mathrm{Lip}}^2}$, is not rotationally invariant for general metric spaces X. However, in the special case $X \subset (\mathbb{R}^n, \|\cdot\|_2)$ and $p \in C^1(X, \mathbb{R}^d)$, the norms (3.56) and (3.23) coincide with $\sup_{z \in X} \|\nabla p(z)\|_{\sigma,\infty}$ (spectral norm of the Jacobian) and $\sup_{z \in X} \|\nabla p(z)\|_F$ (Frobenius norm of the Jacobian) respectively, both satisfying rotational invariance.

3.2.2 $\mathrm{TV}_{\mathrm{KR}}$ as a Regularizer in Variational Problems

In this section, we will prove that, in the case of measure-valued functions $u\colon \Omega \to \mathcal{P}(X)$, the functional $\mathrm{TV}_{\mathrm{KR}}$ exhibits a regularizing property, i.e., it establishes existence of minimizers.

For $\lambda \in [0, \infty)$ and $\rho\colon \Omega \times \mathcal{P}(X) \to [0, \infty)$ fixed, we consider the functional

$$T_{\rho,\lambda}(u) := \int_\Omega \rho(x, u(x))\, dx + \lambda\, \mathrm{TV}_{\mathrm{KR}}(u). \tag{3.57}$$

for $u\colon \Omega \to \mathcal{P}(X)$. In order for $T_{\rho,\lambda}$ to be well-defined, the mapping $x \mapsto \rho(x, u(x))$ needs to be measurable. In the following Lemma, we show that this is the case under mild conditions on ρ.

Lemma 3.7. *Let $\rho\colon \Omega \times \mathcal{P}(X) \to [0, \infty)$ be a globally bounded function that is measurable in the first and convex in the second variable, i.e., $x \mapsto \rho(x, \mu)$ is measurable for each $\mu \in \mathcal{P}(X)$, and $\mu \mapsto \rho(x, \mu)$ is convex for each $x \in \Omega$. Then the map $x \to \rho(x, u(x))$ is measurable for every $u \in L^\infty_w(\Omega, \mathcal{P}(X))$.*

Remark 3.8. As will become clear from the proof, the convexity condition can be replaced by the assumption that ρ be continuous with respect to $(\mathcal{P}(X), W_1)$ in the second variable. However, in order to ensure weak* lower semi-continuity of $T_{\rho,\lambda}$, we will require convexity of ρ in the existence proof (Thm. 3.10) anyway. Therefore, for simplicity we also stick to the (stronger) convexity condition in Lemma 3.7.

Remark 3.9. One example of a function satisfying the assumptions in Lemma 3.7 is given by

$$\rho(x, \mu) := W_1(f(x), \mu),\ x \in \Omega,\ \mu \in \mathcal{P}(\mathbb{S}^2). \tag{3.58}$$

Indeed, boundedness follows from the boundedness of the Wasserstein metric in the case of an underlying bounded metric spaces (here $\mathbb{S}^2$). Convexity in the second argument follows from the fact that the Wasserstein metric is induced by a norm (2.10).

Proof of Lemma 3.7. The metric space $(\mathcal{P}(X), W_1)$ is compact, hence separable. By Pettis' measurability theorem [Bou04, Chapter VI, §1, No. 5, Prop. 12], weak and strong measurability coincide for separably-valued functions so that u is actually strongly measurable as a function with values in $(\mathcal{P}(X), W_1)$. Note, however, that this does not imply strong measurability with respect to the norm topology of $(\mathfrak{M}(X), \|\cdot\|_{\mathfrak{M}})$ in general!

As bounded convex functions are locally Lipschitz continuous [Cla13, Thm. 2.34], ρ is continuous in the second variable with respect to W_1. As in the proof of Lemma 3.1, we now note that ρ is a Carathéodory function, for which compositions with measurable functions such as $x \mapsto \rho(x, u(x))$ are known to be measurable. □

Then, minimizers of $T_{\rho,\lambda}$ exist in the following sense:

Theorem 3.10. *Let $\Omega \subset \mathbb{R}^d$ be open and bounded, let (X, d) be a compact metric space and assume that ρ satisfies the assumptions in Lemma 3.7. Then the variational problem*

$$\inf_{u \in L^\infty_w(\Omega, \mathcal{P}(X))} T_{\rho,\lambda}(u) \tag{3.59}$$

with $T_{\rho,\lambda}$ as in (3.57) admits a (not necessarily unique) solution.

Before we can continue with the proof of existence of minimizers to $T_{\rho,\lambda}$, we need to introduce the notion of *weak* measurability* because this will play a crucial role in the proof. Analogously with the notion of weak measurability and with $L^\infty_w(\Omega, \mathrm{KR}(X))$ introduced above, we say that a measure-valued function $u\colon \Omega \to \mathfrak{M}(X)$ is *weakly* measurable* [DU77, p. 41] if the mapping

$$x \mapsto \int_X f(z)\, u_x(dz) \tag{3.60}$$

is measurable for each $f \in C(X)$. $L^\infty_{w*}(\Omega, \mathfrak{M}(X))$ is defined accordingly as the space of weakly* measurable functions. For functions $u\colon \Omega \to \mathcal{P}(X)$ mapping into the space of probability measures, there is an immediate connection between weak* measurability and weak measurability: u is weakly measurable if the mapping

$$x \mapsto \int_X p\, du_x \tag{3.61}$$

is measurable whenever $p \in \mathrm{Lip}_0(X)$. However, since, by the Stone-Weierstrass theorem, the Lipschitz functions $\mathrm{Lip}(X)$ are dense in $(C(X), \|\cdot\|_\infty)$ [Car00, p. 198], both notions

of measurability coincide for probability measure-valued functions $u\colon \Omega \to \mathcal{P}(X)$, so that

$$L^\infty_w(\Omega, \mathcal{P}(X)) = L^\infty_{w*}(\Omega, \mathcal{P}(X)). \tag{3.62}$$

However, as this equivalence does not hold for the larger spaces $L^\infty_{w*}(\Omega, \mathfrak{M}(X))$ and $L^\infty_w(\Omega, \mathfrak{M}(X))$, it will be crucial to keep track of the difference between weak and weak* measurability in the existence proof.

Proof of Thm. 3.10. The proof is guided by the direct method from the calculus of variations. The first part is inspired by the proof of the Fundamental Theorem for Young measures as formulated and proven in [Bal89].

Let $u^k\colon \Omega \to \mathcal{P}(X)$, $k \in \mathbb{N}$, be a minimizing sequence for $T_{\rho,\lambda}$, i.e.,

$$T_{\rho,\lambda}(u^k) \to \inf_u T_{\rho,\lambda}(u) \quad \text{as} \quad k \to \infty. \tag{3.63}$$

As $\mathfrak{M}(X)$ is the dual space of $C(X)$, $L^\infty_{w*}(\Omega, \mathfrak{M}(X))$ with the norm defined in (3.5) is dual to the Banach space $L^1(\Omega, C(X))$ of Bochner integrable functions on Ω with values in $C(X)$ [II69, p. 93]. Now, $\mathcal{P}(X)$ as a subset of $\mathfrak{M}(X)$ is bounded so that our sequence u^k is bounded in $L^\infty_{w*}(\Omega, \mathfrak{M}(X))$. Here, we used again that $L^\infty_{w*}(\Omega, \mathcal{P}(X)) = L^\infty_w(\Omega, \mathcal{P}(X))$.

Note that we get boundedness of our minimizing sequence "for free", without any assumptions on the coercivity of $T_{\rho,\lambda}$! Hence we can apply the Banach-Alaoglu theorem [Cla13, Thm. 3.14], which states that there exist $u^\infty \in L^\infty_{w*}(\Omega, \mathfrak{M}(X))$ and a subsequence, also denoted by u^k, such that

$$u^k \overset{*}{\rightharpoonup} u^\infty \text{ in } L^\infty_{w*}(\Omega, \mathfrak{M}(X)). \tag{3.64}$$

This means by definition, for each $p \in L^1(\Omega, C(X))$,

$$\int_\Omega \langle u^k(x), p(x)\rangle \, dx \to \int_\Omega \langle u^\infty(x),\, p(x)\rangle \, dx \quad \text{as } k \to \infty. \tag{3.65}$$

We now show that $u^\infty(x) \in \mathcal{P}(X)$ almost everywhere, i.e., u^∞ is a nonnegative measure of unit mass: The convergence (3.65) holds in particular for the choice $p(x,s) := \phi(x)f(s)$, where $\phi \in L^1(\Omega)$ and $f \in C(X)$. For nonnegative functions ϕ and f, we have

$$\int_\Omega \phi(x)\langle u^k(x), f\rangle \, dx \geq 0 \tag{3.66}$$

for all k, which implies

$$\int_\Omega \phi(x)\langle u^\infty(x), f\rangle \, dx \geq 0. \tag{3.67}$$

Since this holds for all nonnegative ϕ and f, we deduce that $u^\infty(x)$ is a nonnegative measure for almost every $x \in \Omega$. The choice $f(s) \equiv 1$ in (3.65) shows that u^∞ has unit mass almost everywhere.

Therefore $u^\infty(x) \in \mathcal{P}(X)$ almost everywhere and we have shown that u^∞ lies in the feasible set $L^\infty_w(\Omega, \mathcal{P}(X))$. It remains to show that u^∞ is in fact a minimizer.

In order to do so, we prove weak* lower semi-continuity of $T_{\rho,\lambda}$. We consider the two integral terms in the definition (3.57) of $T_{\rho,\lambda}$ separately. For the $\mathrm{TV}_{\mathrm{KR}}$ term, for each $p \in C^1_c(\Omega, \mathrm{Lip}(X, \mathbb{R}^d))$, we have $\operatorname{div} p \in L^1(\Omega, C(X))$ so that

$$\lim_{k\to\infty} \int_\Omega \langle u^k(x), \operatorname{div} p(x)\rangle \, dx = \int_\Omega \langle u^\infty(x), \operatorname{div} p(x)\rangle \, dx. \tag{3.68}$$

Taking the supremum over all p with $[p(x)]_{[\mathrm{Lip}(X)]^d} \leq 1$ almost everywhere, we deduce lower semi-continuity of the regularizer:

$$\mathrm{TV}_{\mathrm{KR}}(u^\infty) \leq \liminf_{k\to\infty} \mathrm{TV}_{\mathrm{KR}}(u^k). \tag{3.69}$$

The data fidelity term $u \mapsto \int_\Omega \rho(x, u(x))\, dx$ is convex and bounded on the closed convex subset $L^\infty_w(\Omega, \mathcal{P}(X))$ of the space $L^\infty_{w*}(\Omega, \mathfrak{M}(X))$. It is also continuous, as convex and bounded functions on normed spaces are locally Lipschitz-continuous. This implies weak* lower semi-continuity on $L^\infty_w(\Omega, \mathcal{P}(X))$.

Therefore, the objective function $T_{\rho,\lambda}$ is weakly* lower semicontinuous, and we obtain

$$T_{\rho,\lambda}(u^\infty) \leq \liminf_{k\to\infty} T_{\rho,\lambda}(u^k) \tag{3.70}$$

for the minimizing sequence (u^k), which concludes the proof. □

Non-uniqueness of minimizers of $T_{\rho,\lambda}$ is clear for pathological choices such as $\rho \equiv 0$. However, there are non-trivial cases where uniqueness fails to hold:

Proposition 3.11. *Let $X = \{0, 1\}$ be the metric space consisting of two discrete points of distance 1 and define $\rho(x, \mu) := W_1(f(x), \mu)$ where*

$$f(x) := \begin{cases} \delta_1, & x \in \Omega \setminus U, \\ \delta_0, & x \in U, \end{cases} \tag{3.71}$$

for a non-empty subset $U \subset \Omega$ with C^1-boundary. In the definition (3.55) of $\mathrm{TV}_{\mathrm{KR}}$, assume the coupled norm on $[\mathrm{Lip}_0(X)]^d$ to be defined as in (3.56).

Then there is a one-to-one correspondence between feasible solutions u of problem (3.59)

and feasible solutions $\tilde{u}$ of the classical L^1-TV functional

$$\inf_{\tilde{u}\in L^1(\Omega,[0,1])} \tilde{T}_\lambda(u) := \|\mathbf{1}_U - \tilde{u}\|_{L^1} + \lambda \,\mathrm{TV}(\tilde{u}) \tag{3.72}$$

via the mapping

$$u(x) = \tilde{u}(x)\delta_0 + (1 - \tilde{u}(x))\delta_1. \tag{3.73}$$

Under this mapping, $\tilde{T}_\lambda(\tilde{u}) = T_{\rho,\lambda}(u)$ holds so that our proposed Banach space-valued problem (3.59) *and the classical L^1-TV-problem* (3.72) *are equivalent.*

Furthermore, there exists $\lambda > 0$ for which the minimizer of $T_{\rho,\lambda}$ is not unique.

Proof. Let $u \in L^\infty_w(\Omega, \mathcal{P}(X))$. With the given choice of X, there exists a measurable function $\tilde{u}\colon \Omega \to [0,1]$ such that

$$u(x) = \tilde{u}(x)\delta_0 + (1 - \tilde{u}(x))\delta_1. \tag{3.74}$$

The measurability of $\tilde{u}$ is equivalent to the weak measurability of u by definition:

$$\begin{aligned} \langle p, u(x)\rangle &= \tilde{u}(x)\cdot p_0 + (1-\tilde{u}(x))\cdot p_1 && (3.75)\\ &= \tilde{u}(x)\cdot(p_0 - p_1) + p_1. && (3.76) \end{aligned}$$

The constraint

$$p \in C^1_c(\Omega, [\mathrm{Lip}_0(X)]^d),\ [p(x)]_{\mathrm{Lip}(X,\mathbb{R}^d)} \le 1 \tag{3.77}$$

from the definition of $\mathrm{TV}_{\mathrm{KR}}$ in (3.55) translates to

$$p_0, p_1 \in C_c(\Omega, \mathbb{R}^d),\ \|p_0(x) - p_1(x)\|_2 \le 1. \tag{3.78}$$

Furthermore,

$$\begin{aligned} \langle -\operatorname{div} p(x), u(x)\rangle &= -\operatorname{div} p_0(x)\cdot\tilde{u}(x) - \operatorname{div} p_1(x)\cdot(1-\tilde{u}(x)) && (3.79)\\ &= -\operatorname{div}(p_0 - p_1)(x)\cdot\tilde{u}(x) - \operatorname{div} p_1(x). && (3.80) \end{aligned}$$

By the compact support of p_1, the last term vanishes when integrated over Ω. Conse-

quently,

$$\mathrm{TV}_{\mathrm{KR}}(u) = \sup\Bigg\{ \int_\Omega -\operatorname{div}(p_0 - p_1)(x) \cdot \tilde{u}(x)\, dx : p_0, p_1 \in C_c(\Omega, \mathbb{R}^d), \tag{3.81}$$

$$\|(p_0 - p_1)(x)\|_2 \leq 1 \Bigg\} \tag{3.82}$$

$$= \sup\left\{ \int_\Omega -\operatorname{div} p(x) \cdot \tilde{u}(x)\, dx : p \in C_c(\Omega, \mathbb{R}^d),\ \|p(x)\|_2 \leq 1 \right\} \tag{3.83}$$

$$= \mathrm{TV}(\tilde{u}). \tag{3.84}$$

and therefore

$$T_{\rho,\lambda}(u) = \int_{\Omega\setminus U} \tilde{u}(x)\, dx + \int_U (1 - \tilde{u}(x))\, dx + \lambda\, \mathrm{TV}(\tilde{u}) \tag{3.85}$$

$$= \int_\Omega |\mathbf{1}_U(x) - \tilde{u}(x)|\, dx + \lambda\, \mathrm{TV}(\tilde{u}) \tag{3.86}$$

$$= \|\mathbf{1}_U - \tilde{u}\|_{L^1} + \lambda\, \mathrm{TV}(\tilde{u}). \tag{3.87}$$

Thus we have shown that the functional $T_{\rho,\lambda}$ is equivalent to the classical L^1-TV functional with the indicator function $\mathbf{1}_U$ as input data and evaluated at $\tilde{u}$ which is known to have non-unique minimizers for a certain choice of λ [CE05]. □

3.2.3 Application to ODF-Valued Images

For ODF-valued images, we consider the special case $X = \mathbb{S}^2$ equipped with the metric induced by the standard Riemannian manifold structure on $\mathbb{S}^2$, and $\Omega \subset \mathbb{R}^3$. Let $f \in L^\infty_w(\Omega, \mathcal{P}(\mathbb{S}^2))$ be an ODF-valued image and denote by W_1 the Wasserstein metric from the theory of optimal transport (see Sect. 2.2). Then the function

$$\rho(x, \mu) := W_1(f(x), \mu),\ x \in \Omega,\ \mu \in \mathcal{P}(\mathbb{S}^2), \tag{3.88}$$

satisfies the assumptions in Lemma 3.7 and hence Thm. 3.10. For denoising of an ODF-valued function f in a post-processing step after ODF reconstruction, similar to [VL17] we propose to solve the variational minimization problem

$$\inf_{u:\Omega\to\mathcal{P}(\mathbb{S}^2)} \int_\Omega W_1(f(x), u(x))\, dx + \lambda\, \mathrm{TV}_{\mathrm{KR}}(u), \tag{3.89}$$

using the definition of $\mathrm{TV}_{\mathrm{KR}}(u)$ in (3.55).

The following statement shows that this in fact penalizes jumps in u by the Wasserstein

distance as desired, correctly taking the metric structure of $\mathbb{S}^2$ into account.

Corollary 3.12. *Assume that U is compactly contained in Ω with C^1-boundary ∂U. Let the function $u\colon \Omega \to \mathcal{P}(\mathbb{S}^2)$ be defined as in (3.9) for some $u^+, u^- \in \mathcal{P}(\mathbb{S}^2)$. Choosing the norm (3.56) (or (3.23) with $s = 2$) on the product space $\mathrm{Lip}(\mathbb{S}^2)^d$, we have*

$$\mathrm{TV}_{\mathrm{KR}}(u) = \mathcal{H}^{d-1}(\partial U) \cdot W_1(u^+, u^-). \tag{3.90}$$

Proof. In the functional-analytic framework established above, this statement follows as a simple corollary to Prop. 3.2. □

3.3 Numerical Scheme

The framework presented in Sect. 3.2 applies to arbitrary compact metric spaces X. However, for an efficient implementation of the Lipschitz constraint in (3.55), we will assume an s-dimensional manifold $X = \mathcal{M}$ and follow the discretization scheme from Sect. 2.7. This includes the case of ODF-valued images ($X = \mathcal{M} = \mathbb{S}^2$, $s = 2$), but neglects the reasonable parametrization of $\mathbb{S}^2$ using spherical harmonics in the case of DW-MRI. Moreover, note that the following discretization does not apply to arbitrary metric spaces X. Finally, we will write out the problem in a saddle point form that is amenable to standard primal-dual algorithms (as discussed in Sect. 2.6).

3.3.1 Implementation of the Lipschitz Constraint

Functions $p \in C_c^1(\Omega, \mathrm{Lip}(X, \mathbb{R}^d))$ as they appear in our proposed formulation of $\mathrm{TV}_{\mathrm{KR}}$ in (3.7) are identified with functions on $\Omega \times \mathcal{M}$ and discretized via $p^i_{kt} := p_t(X^i, Z^k)$ in $\mathbb{R}^{N,L,d}$ where $X^1, \dots, X^N$ are the discretization points for Ω and $Z^1, \dots, Z^L$ are the nodal basis points of the finite element space S_h on the triangulated manifold $\mathcal{M}_h$ (see Sect. 2.7).

The Lipschitz constraint in the definition (2.10) of W_1 and in the definition (3.55) of $\mathrm{TV}_{\mathrm{KR}}$ is implemented as a norm constraint on the projected surface gradients

$$g^{ij} := P_{T^j} \nabla_{T^j} p^i \in \mathbb{R}^{d,s}, \tag{3.91}$$

where $T^1, \dots, T^M$ is an enumeration of the simplices that make up $\mathcal{M}_h$. Alternatively,

$$g^{ij} = B^j P^j p^i, \text{ for each } j \in \{1, \dots, M\}, \tag{3.92}$$

for sparse indexing matrices P^j. Following our discussion about the choice of norm, the (Lipschitz) norm constraint $\|g^{ij}\| \leq 1$ can be implemented using the Frobenius norm or the spectral norm, both being rotationally invariant and both acting as desired on cartoon-like jump functions (cf. Prop. 3.2).

3.3.2 Discretized W_1-TV Model

Based on the above discretization, we can formulate saddle point forms for (3.89) that allow to apply a primal-dual first-order method (as discussed in Sect. 2.6). In the following, the measure-valued input or reference image is given by $f \in \mathbb{R}^{L,N}$ and the dimensions of the primal and dual variables are

$$u \in \mathbb{R}^{L,N}, \qquad p \in \mathbb{R}^{L,N,d}, \qquad g \in \mathbb{R}^{N,M,s,d}, \tag{3.93}$$

$$p_0 \in \mathbb{R}^{L,N}, \qquad g_0 \in \mathbb{R}^{N,M,s}, \tag{3.94}$$

where p is the dual variable in $\mathrm{TV}_{\mathrm{KR}}$ and p_0 is the dual variable in W_1.

Using a W_1 data term, the saddle point form of the overall problem reads

$$\min_u \max_p \quad W_1(u, f) + \langle \nabla_x u, p \rangle \tag{3.95}$$

$$\text{s.t.} \quad u^i \geq 0,\ \textstyle\sum_{k=1}^{L} u_k^i = 1,\ \forall i, \tag{3.96}$$

$$\|B^j P^j p^i\| \leq \lambda\ \forall i, j, \tag{3.97}$$

or, applying the Kantorovich-Rubinstein duality (2.10) to the data term,

$$\min_u \max_{p,g,p_0,g_0} \quad \langle u - f, p_0 \rangle + \langle \nabla_x u, p \rangle \tag{3.98}$$

$$\text{s.t.} \quad u^i \geq 0,\ \textstyle\sum_{k=1}^{L} u_k^i = 1\ \forall i, \tag{3.99}$$

$$g^{ij} = B^j P^j p^i,\ \|g^{ij}\| \leq \lambda\ \forall i, j, \tag{3.100}$$

$$g_0^{ij} = B^j P^j p_0^i,\ \|g_0^{ij}\| \leq 1\ \forall i, j. \tag{3.101}$$

3.3.3 Discretized L^2-TV Model

For comparison, we also implemented the Rudin-Osher-Fatemi (ROF) model

$$\inf_{u:\Omega\to\mathcal{P}(\mathbb{S}^2)} \int_\Omega \int_{\mathbb{S}^2} (f_x(z) - u_x(z))^2 \, dz \, dx + \lambda \, \mathrm{TV}(u) \tag{3.102}$$

using $\mathrm{TV} = \mathrm{TV}_{\mathrm{KR}}$. The quadratic data term can be implemented using the saddle point form

$$\min_{u} \max_{p,g} \quad \langle u - f, u - f \rangle + \langle \nabla_x u, p \rangle \tag{3.103}$$

$$\text{s.t.} \quad u^i \geq 0,\ \textstyle\sum_{k=1}^{L} u_k^i = 1, \tag{3.104}$$

$$g_t^{ij} = B^j P^j p_t^i,\ \|g^{ij}\| \leq \lambda\ \forall i, j, t. \tag{3.105}$$

From a functional-analytic viewpoint, this approach requires to assume that u_x can be represented by an L^2 density, suffers from well-posedness issues, and ignores the metric structure on $\mathbb{S}^2$ as mentioned in the introduction (see Fig. 3.2). Nevertheless we include it for comparison, as the L^2 norm is a common choice and the discretized model is a straightforward modification of the W_1-TV model.

3.3.4 Implementation Using a Primal-Dual Algorithm

Based on the saddle point forms (3.98) and (3.103), we applied the primal-dual first-order method with adaptive step sizes as described in Sect. 2.6. We also evaluated the diagonal preconditioning [PC11]. However, we found that while it led to rapid convergence in some cases, the method frequently became unacceptably slow before reaching the desired accuracy. The adaptive step size strategy exhibited a more robust overall convergence.

The equality constraints in (3.98) and (3.103) were included into the objective function by introducing suitable Lagrange multipliers. As far as the norm constraint on g_0 is concerned, the spectral and Frobenius norms agree, since the gradient of p_0 is one-dimensional. For the norm constraint on the Jacobian g of p, we found the spectral and Frobenius norm to give visually indistinguishable results.

Furthermore, since $\mathcal{M} = \mathbb{S}^2$ and therefore $s = 2$ in the ODF-valued case, explicit formulas for the orthogonal projections on the spectral norm balls that appear in the proximal steps are available [GSC12]. The experiments below were calculated using spectral norm constraints, as in our experience this choice led to slightly faster convergence.

3.4 Results

We implemented our model in Python 3 (see `https://github.com/room-10/opymize`) using the libraries NumPy and PyCUDA. The examples were computed on an Intel Xeon X5670 2.93GHz with 24 GB of main memory and an NVIDIA GeForce GTX 480 graphics card with 1,5 GB of dedicated video memory. For each step in the primal-dual algorithm,

a set of kernels was launched on the GPU, while the primal-dual gap was computed and termination criteria were tested every 5 000 iterations on the CPU.

For the following experiments, we applied our models presented in Sects. 3.3.2 (W_1-TV) and 3.3.3 (L_2-TV) to ODF-valued images reconstructed from HARDI data using the reconstruction methods that are provided by the Dipy project [Gar+14]:

- For voxel-wise QBI reconstruction within constant solid angle (CSA-ODF) [ALS09], we used `CsaOdfModel` from `dipy.reconst.shm` with spherical harmonics functions up to order 6.
- We used `ConstrainedSphericalDeconvModel` as provided with the submodule `dipy.reconst.csdeconv` for voxel-wise CSD reconstruction [TCC07].

The response function that is needed for CSD reconstruction was determined using the recursive calibration method [Tax+14] as implemented in `recursive_response`, which is also part of `dipy.reconst.csdeconv`. We generated the ODF plots using VTK-based `sphere_funcs` from `dipy.viz.fvtk`.

It is equally possibly to use other methods for Q-ball reconstruction for the preprocessing step, or even integrate the proposed TV-regularizer directly into the reconstruction process. Furthermore, our method is compatible with different numerical representations of ODFs, including sphere discretization [Goh+09], spherical harmonics [ALS09], spherical wavelets [Kez+08], ridgelets [MR10] or similar basis functions [KK11; Ahr+13], as it does not make any assumptions on regularity or symmetry of the ODFs. We leave a comprehensive benchmark to future work, as the main goal of this work is to investigate the mathematical foundations.

3.4.1 Synthetic Data

L^2-TV vs. W_1-TV In order to justify the more involved approach based on optimal transport distances, we demonstrate the different behaviors of the L^2-TV model compared to the W_1-TV model with the help of one-dimensional synthetic images (Figs. 3.3 and 3.4) generated using the multi-tensor simulation method from `dipy.sims.voxel`, which is based on [ST65] and [Des08, p. 42]; see also [VL17].

By choosing very high regularization parameters λ in the experiment illustrated in Fig. 3.3, we force the models to produce constant results. The L^2-based data term prefers a blurred mixture of diffusion directions, essentially averaging the probability measures. The W_1 data term tends to concentrate the mass close to the median of the diffusion directions on the unit sphere, properly taking into account the metric structure of $\mathbb{S}^2$.

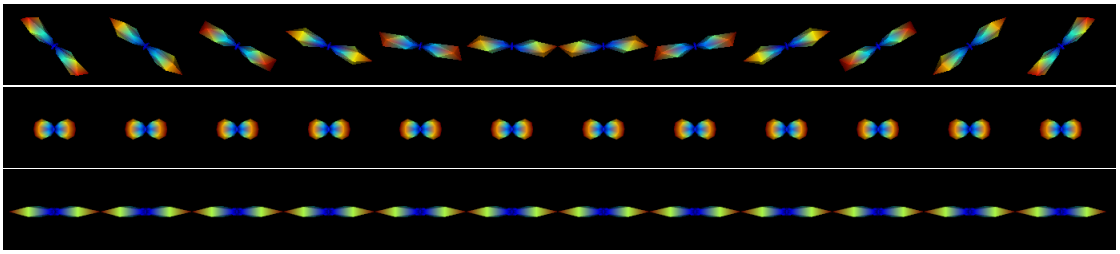


Figure 3.3: **Top:** 1D image of synthetic unimodal ODFs where the angle of the main diffusion direction varies linearly from left to right. This is used as input image for the center and bottom row. **Center:** Solution of L^2-TV model with $\lambda = 5$. **Bottom:** Solution of W_1-TV model with $\lambda = 10$. In both cases, the regularization parameter λ was chosen sufficiently large to enforce a constant result. The quadratic data term mixes all diffusion directions into one blurred ODF, whereas the Wasserstein data term produces a tight ODF that is concentrated close to the median diffusion direction.

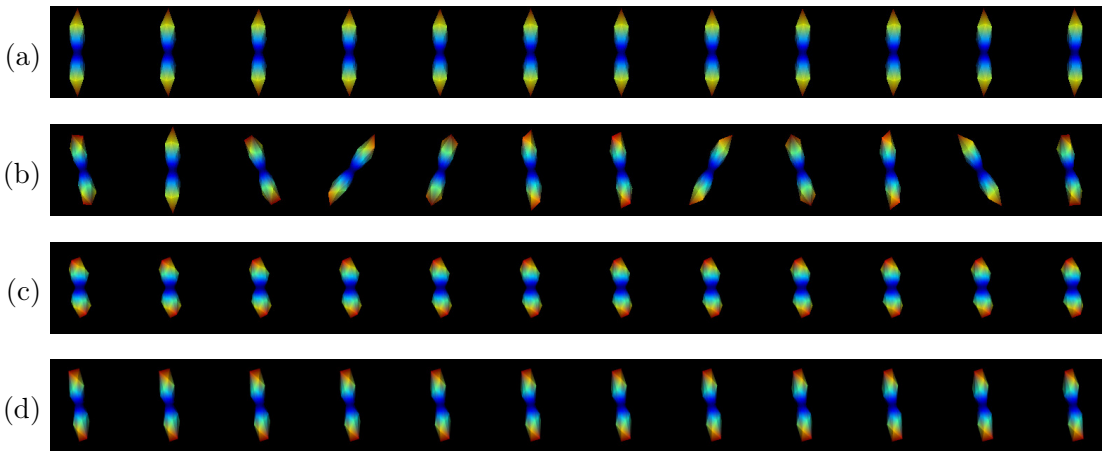


Figure 3.4: Synthetic 1D Q-ball image of unimodal ODFs. (a) ground truth with constant modes and (b) noisy image with distorted diffusion directions. The noisy image was denoised using (c) a quadratic L^2-data term ($\lambda = 0.85$) and (d) the Wasserstein-1 data term ($\lambda = 2.5$). The directional noise causes blurring in the L^2 case (c), whereas the W_1 data term keeps the mass concentrated (d): The entropy of the original data is 2.267 compared to 2.336 (L^2) and 2.279 (W_1).

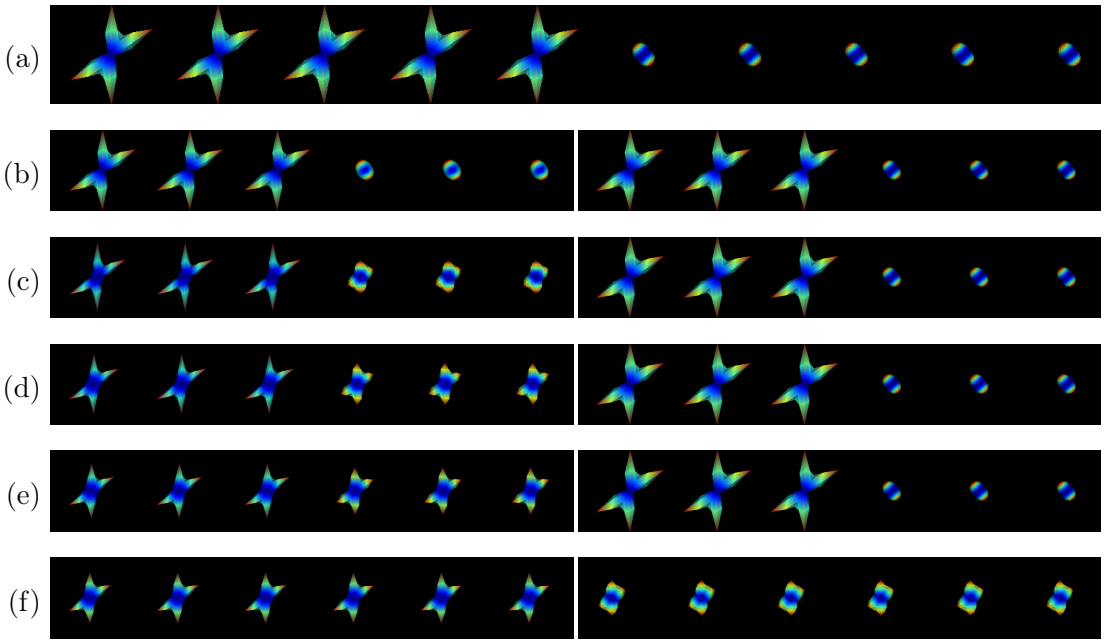


Figure 3.5: Synthetic 1D Q-ball image of bimodal and almost uniform ODFs: The (a) original data was denoised (b)–(f) using an L^2 data term **(left)** and a W_1 data term **(right)** for increasing values of λ (on the left-hand side $\lambda = 0.05, 0.55, 1.05, 1.55, 2.05$ and on the right-hand side $\lambda = 0.05, 1.35, 2.65, 3.95, 5.25$). Both models preserve the edge. However, as is known from classical ROF models, the L^2 data term produces a gradual transition – i. e., contrast loss – towards the constant image, while the W_1 data term exhibits an abrupt phase transition.

A similar experiment (Fig. 3.4) demonstrates that the behavior of the W_1 model is preferable if the main diffusion directions of the unimodal ODFs underlie random distortion (noise). The ground truth consists of 12 identical unimodal ODFs, while the main diffusion directions have been randomly distorted in the input image following a Gaussian distribution on the angle with 20° standard deviation.

Scale-space behavior The next two examples (Figs. 3.5 and 3.6) demonstrate the scale space behavior of our variational models.

The first experiment is inspired by [WDS16], where a similar 1D image is used for demonstration of edge preservation properties. Six voxels of the ground truth are chosen to be bimodal ODFs with the two main diffusion directions separated by 55°. The remaining six voxels are almost uniform ODFs (unimodal ODFs with almost uniform distribution of eigenvalues corresponding to the main diffusion directions). As both

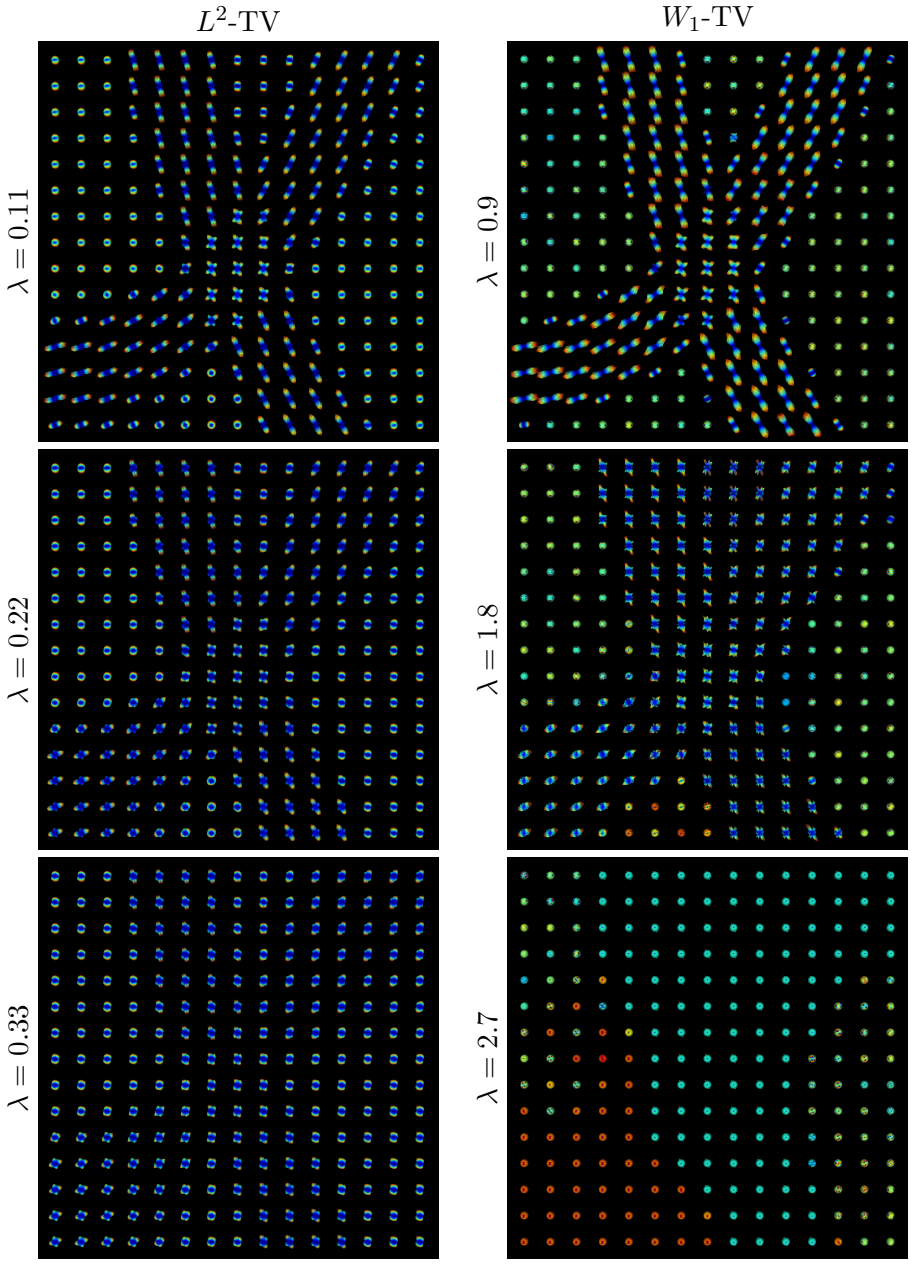


Figure 3.6: Numerical solutions of the proposed variational models (see Sects. 3.3.2 and 3.3.3) applied to the phantom (Fig. 3.1) for increasing values of the regularization parameter λ. **Left column:** Solutions of L^2-TV model for $\lambda = 0.11, 0.22, 0.33$. **Right column:** Solutions of W_1-TV model for $\lambda = 0.9, 1.8, 2.7$. As is known from classical ROF models, the L^2 data term produces a gradual transition/loss of contrast towards the constant image, while the W_1 data term stabilizes contrast as is especially visible along the edges.

models use the proposed TV regularizer, the edge is preserved and a piecewise constant image is produced in both cases. However, just as classical ROF models tend to reduce jump sizes across edges and lose contrast, the result produced using the L^2 model exhibits bimodal ODFs on both sides of the jump and the tightness of the original bimodals gets lost.

Secondly, we implemented a 2-D phantom of two crossing fiber bundles as depicted in Fig. 3.1, inspired by [OCW14]. From this phantom we computed the peak directions of fiber orientations on a 15×15 grid. This was used to generate synthetic HARDI data simulating a DW-MRI measurement with 162 gradients and a b-value of 3 000, again using the multi-tensor simulation framework from `dipy.sims.voxel`.

We then applied our models to the CSA-ODF reconstruction of this data set for increasing values of the regularization parameter λ in order to demonstrate the scale-space behaviors of the different data terms (Fig. 3.6).

As in the previous case, edges are preserved, and the L^2-TV model results in the background and foreground regions becoming gradually more similar as regularization strength increases. The W_1-TV model preserves the unimodal ODFs in the background regions and demonstrates a behavior more akin to robust L^1-TV models [DAG09], with structures disappearing abruptly rather than gradually depending on their scale.

Denoising We applied our model to the CSA-ODF reconstruction of a slice (NumPy coordinates `[12:27,22,21:36]`) from the synthetic HARDI data set with added noise at SNR = 10, provided in the ISBI 2013 HARDI reconstruction challenge. We evaluated the angular precision of the estimated fiber compartments using the script provided on the challenge homepage [Dad+14, `compute_local_metrics.py`]. The script computes the mean μ and standard deviation σ of the angular error between the estimated fiber directions inside the voxels and the ground truth as also provided on the challenge page (Fig. 3.7).

The noisy input image exhibits a mean angular error of $\mu = 34.52$ degrees ($\sigma = 19.00$). The reconstructions using W_1-TV ($\mu = 17.73$, $\sigma = 17.25$) and L^2-TV ($\mu = 17.82$, $\sigma = 18.79$) clearly improve the angular error and give visually convincing results: The noise is effectively reduced and a clear trace of fibers becomes visible (Fig. 3.8). In these experiments, the regularizing parameter λ was chosen optimally in order to minimize the mean angular error to the ground truth.

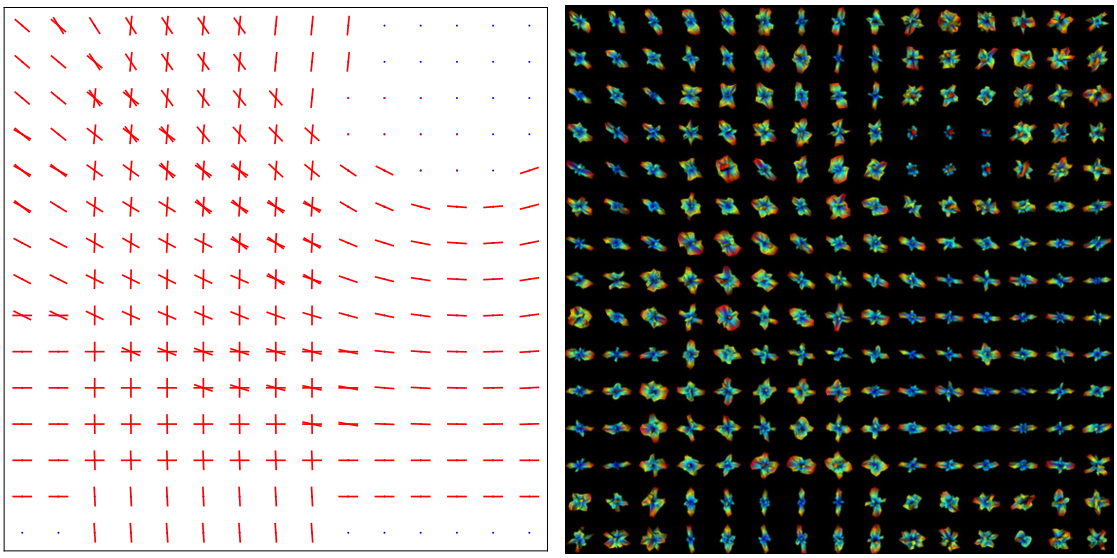


Figure 3.7: Slice of size 15×15 from the data provided for the ISBI 2013 HARDI reconstruction challenge [Dad+14]. **Left:** Peak directions of the ground truth. **Right:** Q-ball image reconstructed from the noisy (SNR = 10) synthetic HARDI data, without spatial regularization. The low SNR makes it difficult to visually recognize the fiber directions.

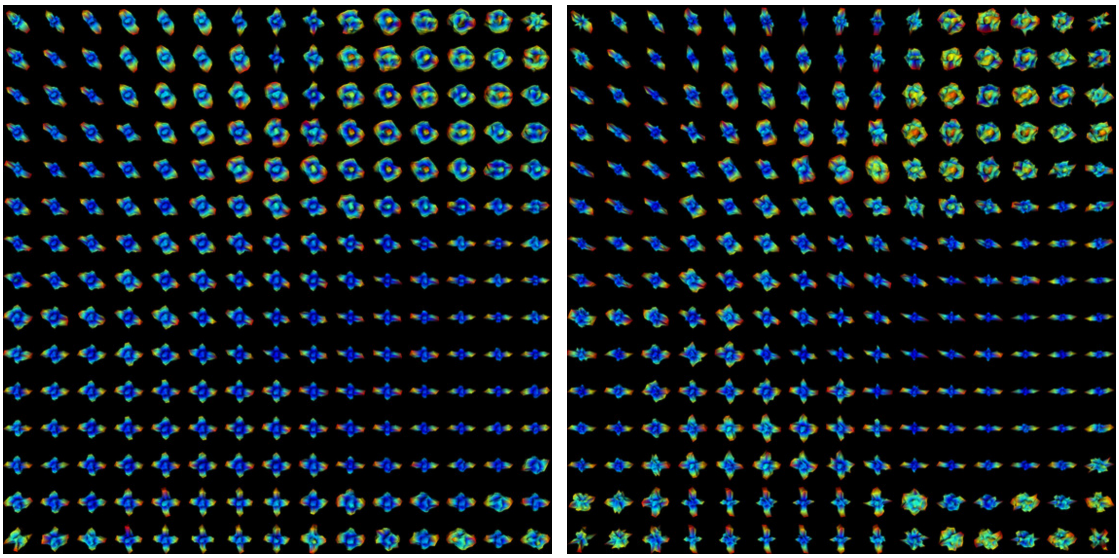


Figure 3.8: Restored Q-ball images reconstructed from the noisy input data in Fig. 3.7. **Left:** Result of the L^2-TV model ($\lambda = 0.3$). **Right:** Result of the W_1-TV model ($\lambda = 1.1$). The noise is reduced substantially so that fiber traces are clearly visible in both cases. The W_1-TV model generates less diffuse distributions.

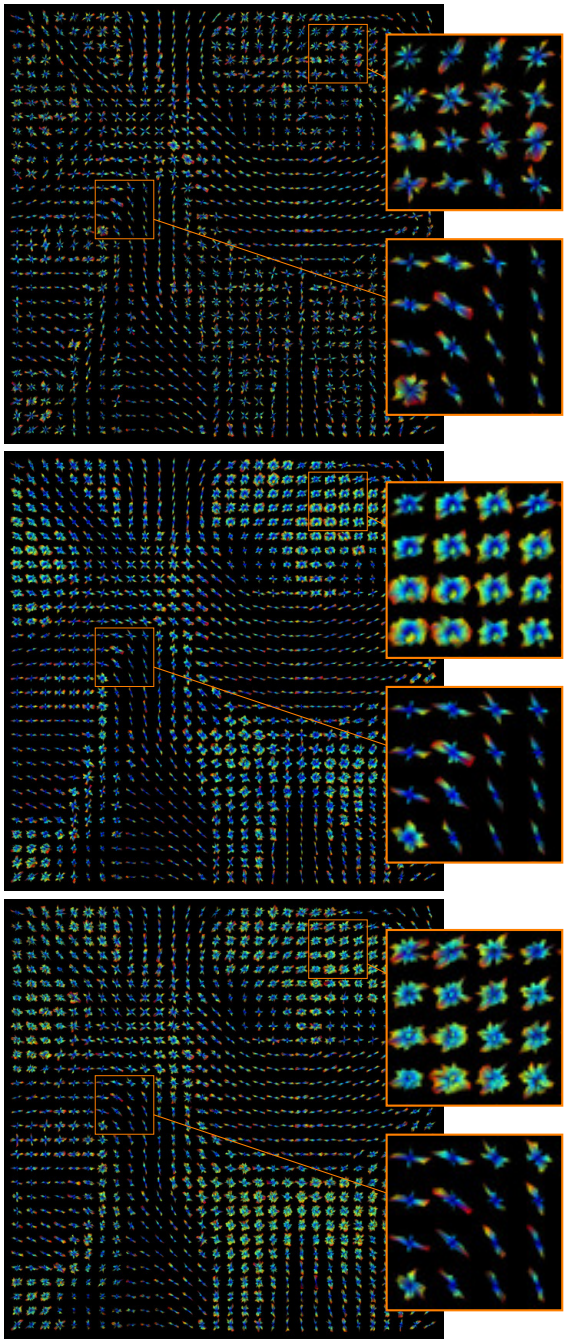


Figure 3.9: ODF image of the corpus callosum, reconstructed with CSD from HARDI data of the human brain [Rok+]. **Top:** Noisy input. **Middle:** Restored using L^2-TV model ($\lambda = 0.6$). **Bottom:** Restored using W_1-TV model ($\lambda = 1.1$). The results do not show much difference: Both models enhance contrast between regions of isotropic and anisotropic diffusion while the anisotropy of ODFs is preserved.

3.4.2 Human Brain HARDI Data

One slice (NumPy coordinates `[20:50, 55:85, 38]`) of HARDI data from the human brain data set [Rok+] was used to demonstrate the applicability of our method to real-world problems and to images reconstructed using CSD (Fig. 3.9). Run times of the W_1-TV and L^2-TV model are approximately 35 minutes (10^5 iterations) and 20 minutes ($6 \cdot 10^4$ iterations).

As a stopping criterion, we require the primal-dual gap to fall below 10^{-5}, which corresponds to a deviation from the global minimum of less than 0.001%, and is a rather challenging precision for the first-order methods used. The regularization parameter λ was manually chosen based on visual inspection.

Overall, contrast between regions of isotropic and anisotropic diffusion is enhanced. In regions where a clear diffusion direction is already visible before spatial regularization, W_1-TV tends to conserve this information better than L^2-TV.

3.5 Conclusion and Outlook

Our mathematical framework for ODF- and, more general, measure-valued images allows to perform total variation-based regularization of measure-valued data without assuming a specific parametrization of ODFs, while correctly taking the metric on $\mathbb{S}^2$ into account. The proposed model penalizes jumps in cartoon-like images proportional to the jump size measured on the underlying normed space, in our case the Kantorovich-Rubinstein space, which is built on the Wasserstein-1-metric. Moreover, the full variational problem was shown to have a solution and can be implemented using off-the-shelf numerical methods.

With the first-order primal-dual algorithm chosen in this work, solving the underlying optimization problem for DW-MRI regularization is computationally demanding due to the high dimensionality of the problem. However, numerical performance was not a priority in this work and can be improved. For example, optimal transport norms are known to be efficiently computable using Sinkhorn's algorithm [Cut13; Bra+17].

A particularly interesting direction for future research concerns extending the approach to simultaneous reconstruction and regularization, with an additional (non-)linear operator in the data fidelity term [ALS09]. For example, one could consider an integrand of the form $\rho(x, u(x)) := d(S(x), Au(x))$ for some measurements S on a metric space (H, d) and a forward operator A mapping an ODF $u(x) \in \mathcal{P}(\mathbb{S}^2)$ to H.

Furthermore, modifications of our total variation seminorm that take into account the coupling of positions and orientations according to the physical interpretation of ODFs in DW-MRI could close the gap to state-of-the-art approaches such as [DF11; PD17].

The model does not require symmetry of the ODFs, and therefore could be adapted to novel asymmetric ODF approaches [Del+07; EOK11; RKK12; KÖU18]. Finally, it is easily extendable to images with values in the probability space over a different manifold, or even a metric space, as they appear for example in statistical models of computer vision [SJJ07], in histogram propagation [Sol+14] and in recent lifting approaches [Möl+16; Lau+16; Åst+17] for combinatorial and non-convex optimization problems.

4 Convex Lifting of Manifold-Valued Variational Problems

Image processing problems where $\Omega \subset \mathbb{R}^d$ is open and bounded and u takes values in an s-dimensional *manifold* $\mathcal{M}$,

$$\inf_{u:\Omega\to\mathcal{M}} F(u) \tag{4.1}$$

are wide-spread in image processing and especially in the processing of manifold-valued images, such as InSAR [MF98], EBSD [BHS11], DTI [BML94], orientational/positional [Ros+12] data or images with values in non-flat color spaces, such as hue-saturation-value (HSV) or chromaticity-brightness (CB) color spaces [CV01].

They come with an inherent non-convexity, as the space of images $u\colon \Omega \to \mathcal{M}$ is generally non-convex, with few exceptions, such as if $\mathcal{M}$ is a Euclidean space, or if $\mathcal{M}$ is a Hadamard manifold, if one allows for the more general notion of geodesic convexity [Bač14; Bač+16]. Except for these special cases, efficient and robust convex numerical optimization algorithms therefore cannot be applied and global optimization is generally out of reach.

The inherent non-convexity of the feasible set is not only an issue of representation. Even for seemingly simple problems, such as the problem of computing the Riemannian center of mass for a number of points on the unit circle, it can affect the energy in surprisingly intricate ways, creating multiple local minimizers and non-uniqueness (Fig. 4.1). The equivalent operation in Euclidean space, computing the weighted mean, is a simple convex (even linear) operation, with a unique, explicit solution.

The problem of non-convexity is not unique to our setting, but rather ubiquitous in a much broader context of image and signal processing (see introductory Sect. 1.2.1). When applied to such non-convex problems, local optimization strategies often get stuck in local minimizers. In *convex relaxation* approaches, an energy functional is approximated by a convex one whose global optimum can be found numerically and whose minimizers lie within a small neighborhood around the actual solution of the problem. A popular convex relaxation technique that applies to a wide range of problems from image and signal processing is *functional lifting* (see introductory Sect. 1.2.2). With this

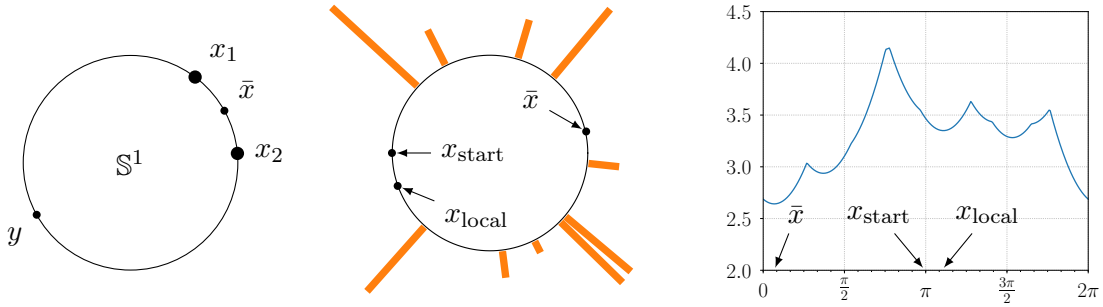


Figure 4.1: Variational problems where the feasible set is a non-Euclidean manifold are prone to local minima and non-uniqueness, which makes them generally much more difficult than their counterparts in $\mathbb{R}^n$. The example shows the generalization of the (weighted) mean to manifolds: the Riemannian center of mass $\bar{x}$ of points x_i on a manifold – in this case, the unit circle $\mathbb{S}^1$ – is defined as the minimizer (if it exists and is unique) of the problem $\inf_{x\in\mathbb{S}^1} \sum_i \lambda_i d(x_i, x)^2$, where d is the geodesic (angular) distance and $\lambda_i > 0$ are given weights. **Left:** Given the two points x_1 and x_2, the energy for computing their "average" has a local minimum at y in addition to the global minimum at $\bar{x}$. Compare this to the corresponding problem in $\mathbb{R}^n$, which has a strictly convex energy with the unique and explicit solution $(x_1 + x_2)/2$. **Center and right:** When the number of points is increased and non-uniform weights are used (represented by the locations and heights of the orange bars), the energy structure becomes even less predictable. The objective function (right, parametrized by angle) exhibits a number of non-trivial local minimizers that are not easily explained by global symmetries. Again, the corresponding problem – computing a weighted mean – is trivial in $\mathbb{R}^n$. Starting from $x_{\text{start}} = \pi$, our functional lifting implementation finds the global minimizer $\bar{x}$, while gradient descent (a local method) gets stuck in the local minimizer x_{local}. Empirically, this behavior can be observed for every other choice of points and weights, but there is no theoretical result in this direction.

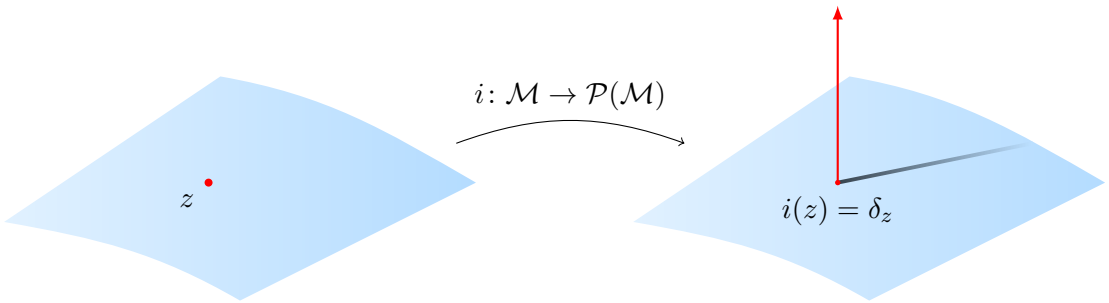


Figure 4.2: A manifold $\mathcal{M}$ is embedded into the space $\mathcal{P}(\mathcal{M})$ of probability measures via the identification of a point $z \in \mathcal{M}$ with the Dirac point measure δ_z concentrated at z. This "lifts" the problem into a higher-dimensional *linear* space, which is much more amenable to global optimization methods.

technique, the feasible set is embedded into a higher-dimensional space where efficient convex approximations of the energy functional are easier available.

In this chapter, we will explore the generalization of functional lifting methods to manifold-valued problems. In contrast to prior work, we will explain existing results in the updated finite element-based framework described in Sect. 2.7. Moreover, we propose extensions to handle general regularizers other than the total variation on manifolds, and to apply the "sublabel-accurate" methods to manifold-valued problems.

4.1 Related Work

Manifold-Valued Functional Lifting

A general introduction to the concept of functional lifting in Euclidean spaces is given in the introductory Sect. 1.2.2. In this chapter, we will consider the more general case of functions $u\colon \Omega \to \Gamma$ with $\Gamma = \mathcal{M}$ having a manifold structure. We will also restrict ourselves to first-order models

$$\inf_{u\colon \Omega \to \Gamma} F(u), \quad F(u) := \int_\Omega f(x, u(x), \nabla u(x))\, dx. \tag{4.2}$$

The first step towards applying lifting methods to problems where Γ has a manifold structure was an application to the restoration of cyclic data [SC11; CS12] with $\Gamma = \mathbb{S}^1$, which was later [Lel+13b] generalized for the case of total variation regularization to data with values in more general manifolds. In [Lel+13b], the functional lifting approach

was applied to a first-order model with total variation regularizer,

$$F(u) = \int_\Omega \rho(x, u(x))dx + \lambda\,\mathrm{TV}(u), \tag{4.3}$$

for $u\colon \Omega \to \mathcal{M}$, where $\Gamma = \mathcal{M}$ is an s-dimensional manifold and $\rho\colon \Omega \times \mathcal{M} \to \mathbb{R}$ is a pointwise data discrepancy. The lifted space is chosen to be $X = \mathcal{P}(\mathcal{M})$, the space of Borel probability measures over $\mathcal{M}$, with embedding $i\colon \mathcal{M} \hookrightarrow \mathcal{P}(\mathcal{M})$, where $i(z) := \delta_z$ is the Dirac point measure with unit mass concentrated at $z \in \mathcal{M}$ (see Fig. 4.2). The lifted functional is

$$\mathcal{F}(v) = \int_\Omega \langle \rho(x, \cdot), v(x)\rangle\, dx + \lambda\widetilde{\mathrm{TV}}(v), \tag{4.4}$$

where $\langle g, \mu\rangle := \int_{\mathcal{M}} g\, d\mu$ for $g \in C(\mathcal{M})$ and $\mu \in \mathcal{P}(\mathcal{M})$. Furthermore,

$$\widetilde{\mathrm{TV}}(v) := \sup\left\{ \int_\Omega \langle \mathrm{Div}_x\, p(x, \cdot), v(x)\rangle\, dx : p\colon \Omega \times \mathcal{M} \to \mathbb{R}, \|\nabla_z p\|_\infty \le 1 \right\}. \tag{4.5}$$

As pointed out in Sect. 3.2, the Lipschitz constraint $\|\nabla_z p\|_\infty \le 1$, where

$$\|\nabla_z p\|_\infty := \sup\left\{ \|\nabla_z p(x, z)\|_{\sigma,\infty} : (x, z) \in \Omega \times \mathcal{M} \right\}, \tag{4.6}$$

can be explained by a functional-analytic perspective on this lifting strategy: The lifted total variation functional is the vectorial total variation semi-norm for functions over Ω with values in the Kantorovich-Rubinstein Banach space of measures $\mathrm{KR}(\mathcal{M})$ (see Sect. 3.2). However, this interpretation does not generalize easily to other regularizers. We will instead loosely base our model for general convex regularizers on the functional lifting strategies presented in [MC17] for scalar and in [MC19] for vectorial problems. In Sect. 6.4.3, we demonstrate that our proposed functional lifting strategy for manifold-valued problems is a special case of the novel, more general functional lifting framework that we propose and discuss in Chap. 6.

Motivated by recent advances in so-called sublabel-accurate liftings (see introductory Sect. 1.2.2), we propose to extend the methods from [Lel+13b] for manifold-valued images to arbitrary convex regularizers, making use of the finite element discretization introduced in Sect. 2.7. This reduces label bias and thus the amount of labels necessary in the discretization (Fig. 4.3).

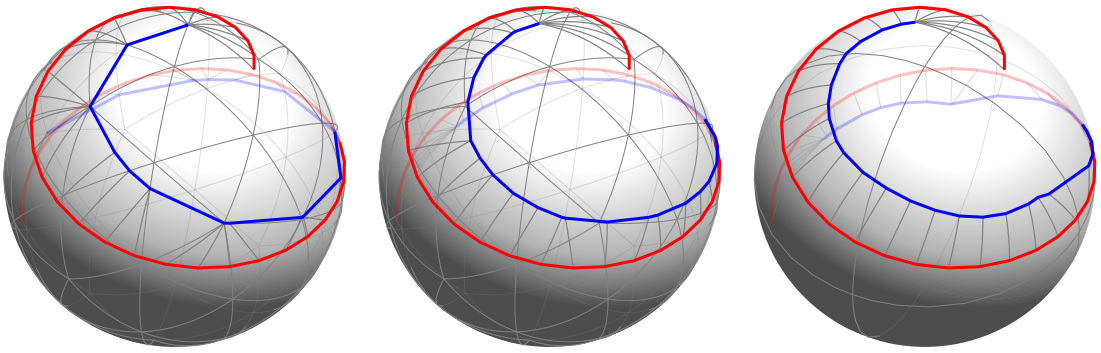


Figure 4.3: Total Variation denoising (blue) of a signal $u\colon [0,1] \to \mathbb{S}^2$ with values in $\mathbb{S}^2$ (red), visualized as curves on the two-dimensional sphere embedded into $\mathbb{R}^3$. The problem is solved by the continuous multi-labeling framework with functional lifting described in this chapter. The discretization points (labels), which are necessary for the implementation of the lifted problem, are visualized by the gray grid. **Left:** The method in [Lel+13b] does not force the solution to take values at the grid points, but still shows significant grid bias. **Center:** With the same number of labels, our proposed method, motivated by [Lau+16], reduces label bias by improving data term discretization. **Right:** Furthermore, our method can get excellent results with as little as 6 grid points (right). Note that the typical contrast reduction that occurs in the classical Euclidean ROF can also be observed in the manifold-valued case in the form of a shrinkage towards the Fréchet mean in the "center" of the helical curve.

Further Related Work

The methods proposed in this work are applicable to variational problems with values in manifolds of dimension $s \leq 3$. The theoretical framework applies to manifolds of arbitrary dimension, but the numerical costs increase exponentially with the dimension s and become prohibitively large in practice for dimensions 4 and larger.

An alternative is to use *local* optimization methods on manifolds. A reference for the smooth case is [AMS09]. For non-smooth energies, methods such as the cyclic proximal point, Douglas-Rachford, ADMM and (sub-)gradient descent algorithm have been applied to first- and second-order TV and TGV as well as Mumford-Shah and Potts regularization approaches in [WDS14; WDS16; Bau+16; BPS16; BL18; Ber+18a]. These methods are generally applicable to manifolds of any dimension whose (inverse) exponential mapping can be evaluated in reasonable time. They are quite efficient in finding a local minimum, but can get stuck in local extrema. Furthermore, the use of total variation regularization in these frameworks is currently limited to anisotropic formulations; instead, quadratic

regularization is isotropic [WDS14; Ber+18b]. An overview of applications, variational models and local optimization methods is given in [Ber+18b].

Furthermore, we mention that, beyond variational models, there exist discrete graph-based [BT18], statistical [Fle12], wavelet-based [SW18], PDE-based [Che+04] and patch-based [LPS17] models for the processing and regularization of manifold-valued signals.

4.2 Calculus of Variations on Submanifolds of $\mathbb{R}^N$

We formulate our model for submanifolds of $\mathbb{R}^N$ as discussed at the beginning of Sect. 2.7.

4.2.1 General First-Order Variational Problems

In this section, we generalize the total variation-based approach in [Lel+13b] to less restrictive first-order variational problems by applying the ideas from functional lifting of vectorial problems [MC19] to manifold-valued problems. Most derivations will be formal; we leave a rigorous choice of function spaces as well as an analysis of well-posedness for future work. We note that theoretical work is available for the scalar-valued case in [ABD03; Poc+10; BF18] and for the vectorial and for selected manifold-valued cases in [GMS98a; GMS98b].

We consider variational models on functions $u\colon \Omega \to \mathcal{M}$,

$$F(u) := \int_\Omega f(x, u(x), Du(x))\,dx, \tag{4.7}$$

for which the integrand $f\colon \Omega \times \mathcal{M} \times \mathbb{R}^{N,d} \to \mathbb{R}$ is convex in the last component. Note that the dependence of f on the full Jacobian of u avoids the coordinate-free tangent bundle push-forward $T\Omega \to T\mathcal{M}$, thus facilitating discretization in a coordinate system later on.

Formally, the lifting strategy for vectorial problems [MC19] can be generalized to this setting by replacing the range Γ with $\mathcal{M}$. As the lifted space, we consider the space of probability measures on the Borel σ-Algebra over $\mathcal{M}$, $X = \mathcal{P}(\mathcal{M})$, with embedding $i\colon \mathcal{M} \to \mathcal{P}(\mathcal{M})$, where $i(z) = \delta_z$ is the Dirac point mass concentrated at $z \in \mathcal{M}$. Furthermore, we write $\Sigma := \Omega \times \mathcal{M}$ and, for $(x, z) = y \in \Sigma$, we define the coordinate projections $\pi_1 y := x$ and $\pi_2 y := z$. Then, for $v\colon \Omega \to \mathcal{P}(\mathcal{M})$, we define the lifted functional

$$\mathcal{F}(v) := \sup\left\{ \int_\Omega \langle -\operatorname{Div}_x p(x,\cdot) + q(x,\cdot), v(x)\rangle\,dx : (\nabla_z p, q) \in \mathcal{K} \right\}, \tag{4.8}$$

where $\langle g, \mu \rangle := \int_{\mathcal{M}} g \, d\mu$ is the dual pairing between $g \in C(\mathcal{M})$ and $\mu \in \mathcal{P}(\mathcal{M})$ and

$$\mathcal{K} := \left\{ (P, q) \in C(\Sigma; \mathbb{R}^{N,d} \times \mathbb{R}) : f^*(\pi_1 y, \pi_2 y, P(y)) + q(y) \leq 0 \, \forall y \in \Sigma \right\}, \tag{4.9}$$

where $f^*(x, z, \zeta) := \sup_\xi \langle \zeta, \xi \rangle - f(x, z, \xi)$ is the convex conjugate of f with respect to the last variable.

In the following, the integrand $f \colon \Omega \times \mathcal{M} \times \mathbb{R}^{N,d} \to \mathbb{R}$ is assumed to decompose as

$$f(x, z, \xi) = \rho(x, z) + \eta(P_z \xi) \tag{4.10}$$

into a pointwise data term $\rho \colon \Omega \times \mathcal{M} \to \mathbb{R}$ and a convex regularizer $\eta \colon \mathbb{R}^{s,d} \to \mathbb{R}$ that only depends on an s-dimensional representation of vectors in $T_z\mathcal{M}$ given by a surjective linear map $P_z \in \mathbb{R}^{s,N}$ with $\ker(P_z) = N_z\mathcal{M}$, the orthogonal complement of $T_z\mathcal{M}$ in $\mathbb{R}^N$.

This very general integrand covers most first-order models in the literature on manifold-valued imaging problems. It applies in particular to isotropic and anisotropic regularizers that depend on (matrix) norms of $Du(x)$, such as the Frobenius or spectral/operator norm where P_z is taken to be an arbitrary orthogonal basis transformation. Since $z \mapsto P_z$ is not required to be continuous, it can also be applied to non-orientable manifolds, such as the Moebius strip or the Klein bottle, where no continuous orthogonal basis representation of the tangent bundle $T\mathcal{M}$ exists.

Regularizers of this particular form depend on the manifold through the choice of P_z only. This is important because, in the course of discretization (see Sect. 2.7), $\mathcal{M}$ is approximated by a discrete (simplicial) manifold $\mathcal{M}_h$ and the tangent spaces $T_z\mathcal{M}$ are replaced by the linear spaces spanned by the simplicial faces of $\mathcal{M}_h$.

4.2.2 Finite Element Formulation

We translate the finite element approach for functional lifting [MC17] to the manifold-valued setting by employing the notation introduced in Sect. 2.7. The energy functional is translated to the discretized setting by redefining the integrand f on $\mathcal{M}_h$ for each $x \in \Omega$, $z \in \mathcal{M}_h$ and $\xi \in \mathbb{R}^{N,d}$ as

$$f(x, z, \xi) := \rho(x, \iota(z)) + \eta(P_T \xi), \tag{4.11}$$

where $P_T \in \mathbb{R}^{s,N}$ is the orthogonal basis representation of vectors in $\mathbb{R}^N$ in the subspace spanned by the simplex $T \in \mathcal{T}_h$ that satisfies $z \in T$. The epigraphical constraints in $\mathcal{K}$

translate to

$$\forall x \in \Omega \forall z \in \mathcal{M}_h \colon \quad \eta^*(P_T \nabla_z p(x,z)) - \rho(x, \iota(z)) + q(x,z) \leq 0, \tag{4.12}$$

for functions $p \in S_h^d$ and $q \in S_h$. The constraints can be efficiently implemented on each $T \in \mathcal{T}_h$ where $\nabla_z p$ is constant and $q(x,z) = \langle q_{T,1}(x), z \rangle + q_{T,2}(x)$ is linear affine in z:

$$\eta^*(P_T \nabla_T p(x)) + \langle q_{T,1}(x), z \rangle - \rho(x, \iota(z)) \leq -q_{T,2}(x), \tag{4.13}$$

for each $x \in \Omega$, $T \in \mathcal{T}_h$ and $z \in T$. Following the approach in [MC17], we define

$$\rho_T^*(x,z) := \sup_{z' \in T} \left\{ \langle z, z' \rangle - \rho(x, \iota(z')) \right\}, \tag{4.14}$$

and introduce auxiliary variables a_T, b_T to split the epigraphical constraint (4.13) into two epigraphical and one linear constraint for $x \in \Omega$ and $T \in \mathcal{T}_h$:

$$\eta^*(P_T \nabla_T p(x)) \leq a_T(x), \tag{4.15}$$

$$\rho_T^*(q_{T,1}(x)) \leq b_T(x), \tag{4.16}$$

$$a_T(x) + b_T(x) = -q_{T,2}(x). \tag{4.17}$$

The resulting optimization problem is described by the following saddle point form over functions $v \in L^1(\Omega, \mathcal{P}_h(\mathcal{M}_h))$, $p \in C^1(\Omega, S_h^{d+1})$ and $q \in C(\Omega, S_h)$:

$$\inf_v \sup_{p,q} \int_\Omega \langle -\operatorname{Div}_x p(x,\cdot) + q(x,\cdot), v(x) \rangle \, dx \tag{4.18}$$

$$\text{subject to} \quad \eta^*(P_T \nabla_T p(x)) \leq a_T(x), \tag{4.19}$$

$$\rho_T^*(q_{T,1}(x)) \leq b_T(x), \tag{4.20}$$

$$a_T(x) + b_T(x) + q_{T,2}(x) = 0. \tag{4.21}$$

Finally, for the fully discrete setting, the domain Ω is replaced by a Cartesian rectangular grid with finite differences operator ∇_x and Neumann boundary conditions.

4.2.3 Connection to [Lel+13b]

In [Lel+13b], the special case of total variation regularization is approached using a similar functional lifting without the finite elements interpretation. More precisely, the regularizing term is chosen to be $\eta(\xi) = \lambda \|\xi\|_{\sigma,1}$ for $\xi \in \mathbb{R}^{s,d}$, where $\|\cdot\|_{\sigma,1}$ is the matrix nuclear norm, also known as Schatten-1-norm, which is given by the sum of singular

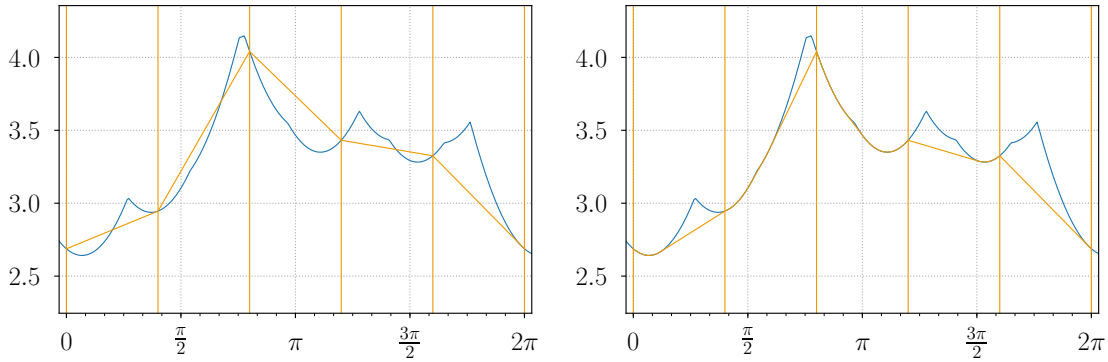


Figure 4.4: Data term discretization for the lifting approach applied to the Riemannian center of mass problem introduced in Fig. 4.1. For each $x \in \Omega$, the data term $z \mapsto \rho(x, z)$ (blue graph) is approximated (orange graphs) between the label points Z^k (orange vertical lines). **Left:** In the lifting approach [Lel+13b] for manifold-valued problems, the data term is interpolated linearly between the labels. **Right:** Based on ideas from recent scalar and vectorial lifting approaches [Möl+16; Lau+16], we interpolate piecewise convex between the labels.

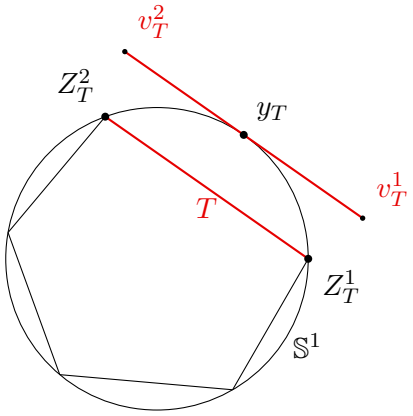


Figure 4.5: Mapping a simplex T, spanned by $Z_T^1, \ldots, Z_T^{s+1}$, to the tangent space at its center-of-mass y_T using the logarithmic map. The proportions of the simplex spanned by the mapped points $v_T^1, \ldots, v_T^{s+1}$ may differ from the proportions of the original simplex for curved manifolds. The illustration shows the case of a circle $\mathbb{S}^1 \subset \mathbb{R}^2$, where the deformation reduces to a multiplication by a scalar α_T, the ratio between the geodesic (angular) and Euclidean distance between Z_T^1 and Z_T^2. The gradient $\nabla_T p$ of a finite element $p \in S_h$ can be modified according to this change in proportion in order to compensate for some of the geometric (curvature) information lost in the discretization.

values of a matrix. It is the dual to the matrix operator or spectral norm $\|\cdot\|_{\sigma,\infty}$. If we substitute this choice of η into the discretization given above, the epigraphical constraint (4.12) translates to the two constraints

$$\|P_T \nabla_T p(x)\|_{\sigma,\infty} \leq \lambda \text{ and } q(x,z) \leq \rho(x,\iota(z)). \tag{4.22}$$

The first one is a Lipschitz constraint just as in the model from [Lel+13b], but two differences remain:

1. Sublabel-accuracy. In [Lel+13b], the lifted and discretized form of the data term reads

$$\int_\Omega \sum_{k=1}^{L} \rho(x, Z^k) v(x)^k \, dx. \tag{4.23}$$

This agrees with our setting if $z \mapsto \rho(x,\iota(z))$ is affine linear on each simplex $T \in \mathcal{T}_h$, as then $q(x,z) = \rho(x,\iota(z))$ maximizes the objective function for each p and v. Hence, the model in [Lel+13b] does not take into account any information about ρ below the resolution of the triangulation. We improve this by implementing the epigraph constraints $\rho_T^*(q_{T,1}(x)) \leq b_T(x)$ as in [Lau+16] using a convex approximation of ρ_T (see Fig. 4.4). The approximation is implemented numerically with piecewise affine linear functions in a "sublabel-accurate" way, i.e., at a resolution below the resolution of the triangulation.

2. Gradient discretization on the manifold. A very specific discretization of the gradients $\nabla_T p(x)$ is used in [Lel+13b]: To each simplex in the triangulation a mid-point $y_T \in \mathcal{M}$ is associated. The vertices $Z_T^1, ..., Z_T^{s+1}$ of the simplex are projected to the tangent space at y_T as $v_T^k := \log_{y_T} Z_T^k$. The gradient is then computed as the vector g in the tangent space $T_{y_T}\mathcal{M}$ describing the affine linear map on $T_{y_T}\mathcal{M}$ that takes values $p(Z_T^k)$ at the points v_T^k, $k = 1, \ldots, s+1$.

This procedure aims to outweigh some of the error introduced by the simplicial discretization and amounts to a different choice of P_T than in our proposed model. We did not observe any significant positive or negative effects from using either discretization; the difference between the minimizers is very small in practice.

For problems with a one-dimensional range, the two approaches differ only in a constant factor: Denote by $P_T \in \mathbb{R}^{s,N}$ the orthogonal basis representation of vectors in $\mathbb{R}^N$ in the subspace spanned by the simplex $T \in \mathcal{T}_h$ and denote by $\tilde{P}_T \in \mathbb{R}^{s,N}$ the alternative approach from [Lel+13b]. Now, consider a triangulation $\mathcal{T}_h$ of the circle $\mathbb{S}^1 \subset \mathbb{R}^2$ and a one-dimensional simplex $T \in \mathcal{T}_h$. A finite element $p \in S_h$ that takes values $p_1, p_2 \in \mathbb{R}$ at

the vertices $Z_T^1, Z_T^2 \in \mathbb{R}^2$ that span T has the gradient

$$\nabla_T p = (p_1 - p_2) \frac{Z_T^1 - Z_T^2}{\|Z_T^1 - Z_T^2\|_2^2} \in \mathbb{R}^2 \tag{4.24}$$

and $P_T, \tilde{P}_T \in \mathbb{R}^{1,2}$ are given by

$$P_T := \frac{(Z_T^1 - Z_T^2)^\top}{\|Z_T^1 - Z_T^2\|_2}, \qquad \tilde{P}_T := \frac{(Z_T^1 - Z_T^2)^\top}{d_{\mathbb{S}^1}(Z_T^1, Z_T^2)}. \tag{4.25}$$

Hence $P_T = \alpha_T \tilde{P}_T$ for $\alpha_T = d_{\mathbb{S}^1}(Z_T^1, Z_T^2)/\|Z_T^1 - Z_T^2\|_2$ the ratio between geodesic (angular) and Euclidean distance between the vertices. If the vertices are equally spaced on $\mathbb{S}^1$, this is a constant factor independent of T that typically scales the discretized regularizer by a small constant factor. On higher-dimensional manifolds, more general linear transformations $P_T = A_T \tilde{P}_T$ come into play. For very irregular triangulations and coarse discretization, this may affect the minimizer; however, in our experiments the observed differences were negligible.

4.2.4 Full Discretization and Numerical Implementation

A prime advantage of the lifting method when applied to manifold-valued problems is that it is formulated over linear or convex sets of functions, even though the original spaces of manifold-valued functions are non-convex in general. This allows to apply established solution strategies for the non-manifold case, which rely on non-smooth convex optimization: After discretization, the convex-concave saddle point form allows for a solution using the primal-dual hybrid gradient method described in Sect. 2.6. In this optimization framework, the epigraph constraints are realized by projections onto the epigraphs in each iteration step. For the regularizers to be discussed in this paper (TV, quadratic and Huber), we refer to the instructions given in [Poc+10]. For the data term ρ, we follow the approach in [Lau+16]: For each $x \in \Omega$, The data term $z \mapsto \rho(x, \iota(z))$ is sampled on a subgrid of $\mathcal{M}_h$ and approximated by a piecewise affine linear function. The quickhull algorithm [BDH96] can then be used to get the convex hull of this approximation. Projections onto the epigraph of ρ_T^* are then projections onto convex polyhedra, which amounts to solving many low-dimensional quadratic programs; see [Lau+16] for more details.

Following [Lel+13b], the numerical solution $u\colon \Omega \to \mathcal{P}_h(\mathcal{M}_h)$, taking values in the lifted space $\mathcal{P}_h(\mathcal{M}_h)$, is projected back to a function $u\colon \Omega \to \mathcal{M}$, taking values in the original space $\mathcal{M}$, by mapping, for each $x \in \Omega$ separately, a probability measure $u(x) =$

$(\lambda_1, \dots, \lambda_L) = \mu_h \in \mathcal{P}_h(\mathcal{M}_h)$ to the following Riemannian center of mass on the original manifold $\mathcal{M}$:

$$\mu_h = (\lambda_1, \dots, \lambda_L) \mapsto \operatorname*{argmin}_{z \in \mathcal{M}} \sum_{k=1}^{L} \lambda_k d_{\mathcal{M}}(z, Z^k)^2. \tag{4.26}$$

For $\mathcal{M} = \mathbb{R}^s$, this coincides with the usual weighted mean $\bar{z} = \sum_{k=1}^{L} \lambda_k Z_k$. However, on manifolds this minimization is known to be a non-convex problem with non-unique solutions (compare Fig. 4.1). Still, in practice the iterative method described in [Kar77] yields reasonable results for all real-world data considered in this work: Starting from a point $z_0 := Z^k$ with maximum weight λ_k, we proceed for $i \geq 0$ by projecting the Z^k, $k = 1, \dots, L$, to the tangent space at z_i using the inverse exponential map, taking the linear weighted mean v_i there and defining z_{i+1} as the projection of v_i to $\mathcal{M}$ via the exponential map:

$$V_i^k := \log_{z_i}(Z^k) \in T_{z_i}\mathcal{M}, \ k = 1, \dots, L, \tag{4.27}$$

$$v_i := \sum_{k=1}^{L} \lambda_k V_i^k \in T_{z_i}\mathcal{M}, \tag{4.28}$$

$$z_{i+1} := \exp_{z_i}(v_i). \tag{4.29}$$

The method converges rapidly in practice. It has to be applied only once for each $x \in \Omega$ after solving the lifted problem so that efficiency is non-critical.

4.3 Numerical Results

We apply our model to problems with quadratic data term $\rho(x, z) := d_{\mathcal{M}}^2(I(x), z)$ and Huber, total variation (TV) and quadratic regularization with parameter $\lambda > 0$:

$$\eta_{\mathrm{TV}}(\xi) := \lambda \|\xi\|_2, \tag{4.30}$$

$$\eta_{\mathrm{Huber}}(\xi) := \lambda \phi_\alpha(\xi), \tag{4.31}$$

$$\eta_{\mathrm{quad}}(\xi) := \frac{\lambda}{2} \|\xi\|_2^2, \tag{4.32}$$

where the Huber function ϕ_α for $\alpha > 0$ is defined by

$$\phi_\alpha(\xi) := \begin{cases} \frac{\|\xi\|_2^2}{2\alpha} & \text{if } \|\xi\|_2 \leq \alpha, \\ \|\xi\|_2 - \frac{\alpha}{2} & \text{if } \|\xi\|_2 > \alpha. \end{cases} \tag{4.33}$$

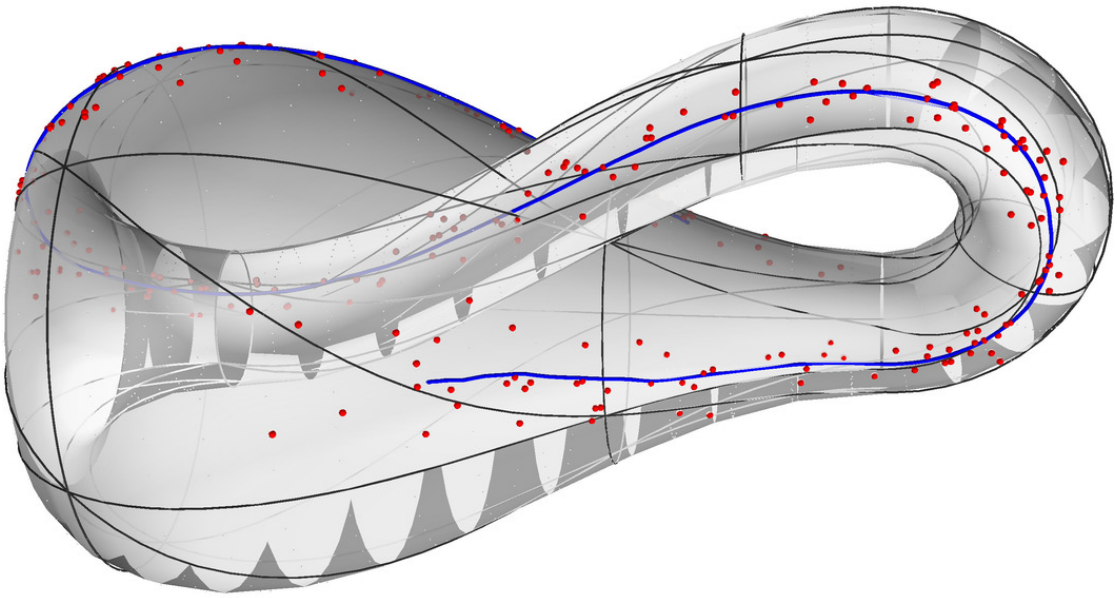


Figure 4.6: Quadratic denoising (blue) of a one-dimensional signal (red) $u\colon [0,1] \to \mathcal{M}$ with values on the two-dimensional Klein surface (commonly referred to as Klein bottle) $\mathcal{M} \subset \mathbb{R}^3$. The black wireframe lines on the surface represent the triangulation used by the discretization of our functional lifting approach. The numerical implementation recovers the denoised signal at a resolution far below the resolution of the manifold's discretization. The lifting approach does not require the manifold to be orientable.

Note that previous lifting approaches for manifold-valued data were restricted to total variation regularization η_{TV}.

The methods were implemented in Python 3 with NumPy and PyCUDA, running on an Intel Core i7 4.00 GHz with an NVIDIA GeForce GTX 1080 Ti 12 GB and 16 GB RAM. The iteration was stopped as soon as the relative gap between primal and dual objective fell below 10^{-5}. Approximate runtimes ranged between 5 and 45 minutes. The code is available from `https://github.com/room-10/mfd-lifting`.

4.3.1 One-Dimensional Denoising on a Klein Bottle

Our model can be applied to both orientable and non-orientable manifolds. Figure 4.6 shows an application of our method to quadratic denoising of a synthetic one-dimensional signal $u\colon [0,1] \to \mathcal{M}$ on the two-dimensional Klein surface embedded in $\mathbb{R}^3$, a non-orientable closed surface that cannot be embedded into $\mathbb{R}^3$ without self-intersections. Our numerical implementation uses a triangulation with a very low count of 5×5 vertices

and 50 triangles. The resolution of the signal (250 one-dimensional data points) is far below the resolution of the triangulation. Nevertheless, our approach is able to restore a smooth curve.

4.3.2 Three-Dimensional Manifolds: $SO(3)$

Signals with rotational range $u\colon \Omega \to SO(3)$ occur in the description of crystal symmetries in EBSD (Electron Backscatter Diffraction Data) and in motion tracking. The rotation group $SO(3)$ is a three-dimensional manifold that can be identified with the three-dimensional unit-sphere $\mathbb{S}^3$ up to identification of antipodal points via the quaternion representation of 3D rotations. A triangulation of $\mathbb{S}^3$ is given by the vertices and simplicial faces of the hexacosichoron (600-cell), a regular polytope in $\mathbb{R}^4$ akin to the icosahedron in $\mathbb{R}^3$. As in [Lel+13b], we eliminate opposite points in the hexacosichoron and obtain a discretization of $SO(3)$ with 60 vertices and 300 tetrahedral faces.

Motivated by Bézier surface interpolation [Abs+16], we applied quadratic regularization to a synthetic inpainting (interpolation) problem with added noise (Fig. 4.7). In our variational formulation, we chose $\rho(x, z) = 0$ for x in the inpainting area and $\rho(x, z) = \delta_{\{z=I(x)\}}$ (a hard constraint to the input signal $I\colon \Omega \to SO(3)$) for x in the known area.

Using the proposed sublabel-accurate handling of data terms, we obtain visually appealing results with only 60 vertices, in contrast to [Lel+13b], where the discretization is refined to 720 vertices (Fig. 4.7).

4.3.3 Normals Fields from Digital Elevation Data

In digital elevation models (DEM), elevation information for earth science studies and mapping applications often includes surface normals which can be used to produce a shaded coloring of elevation maps. Normal fields $u\colon \Omega \to \mathbb{S}^2$ are defined on a rectangular image domain $\Omega \subset \mathbb{R}^2$; variational processing of the normal fields is therefore a manifold-valued problem on the two-dimensional sphere $\mathbb{S}^2 \subset \mathbb{R}^3$.

Denoising using variational regularizers from manifold-valued image processing before computing the shading considerably improves visual quality (Figures 4.8 and 4.9). For our framework, the sphere was discretized using 12 vertices and 20 triangles, chosen to form a regular icosahedron. The lifting approach in [Lel+13b] required 162 vertices – and solving a proportionally larger optimization problem – in order to produce comparable results on the same dataset.

We applied our approach with TV, Huber and quadratic regularization. Interestingly,

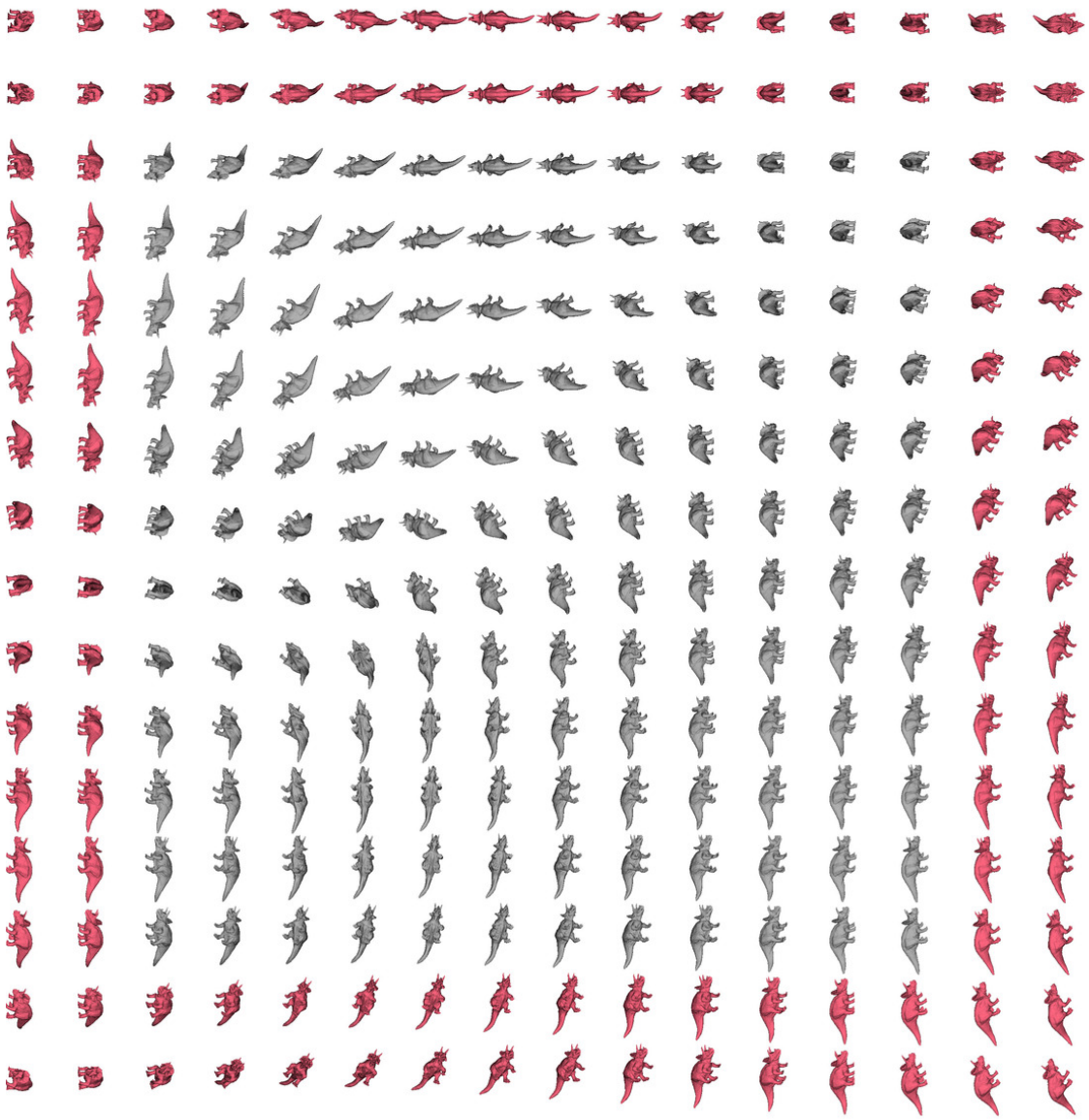


Figure 4.7: Diffusive inpainting of a two-dimensional signal of (e.g., camera) orientations, elements of the three-dimensional special orthogonal group of rotations $SO(3)$, a manifold of dimension $s = 3$. The masked input signal (red) is inpainted (gray) using our model with quadratic regularization. The interpolation into the central area is smooth. Shape: *Triceratops* by BillyOceansBlues (CC-BY-NC-SA, `https://www.thingiverse.com/thing:3313805`).

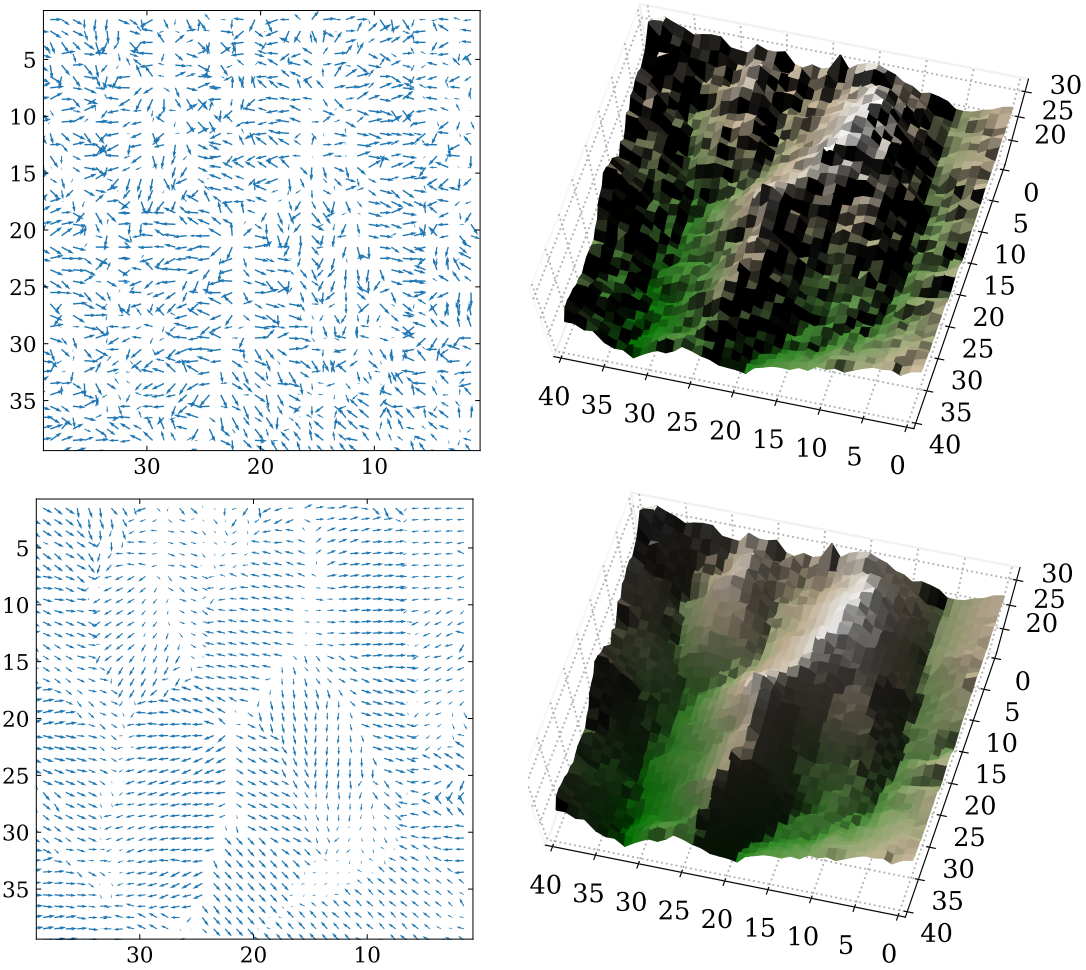


Figure 4.8: Denoising of $\mathbb{S}^2$-valued surface normals on the digital elevation model (DEM) dataset from [Ges+09] **(noisy input on top)**. The total variation model ($\lambda = 0.4$) enforces flat hillsides in the denoised image **(bottom)**. See Fig. 4.9 for results from Huber and quadratic regularization.

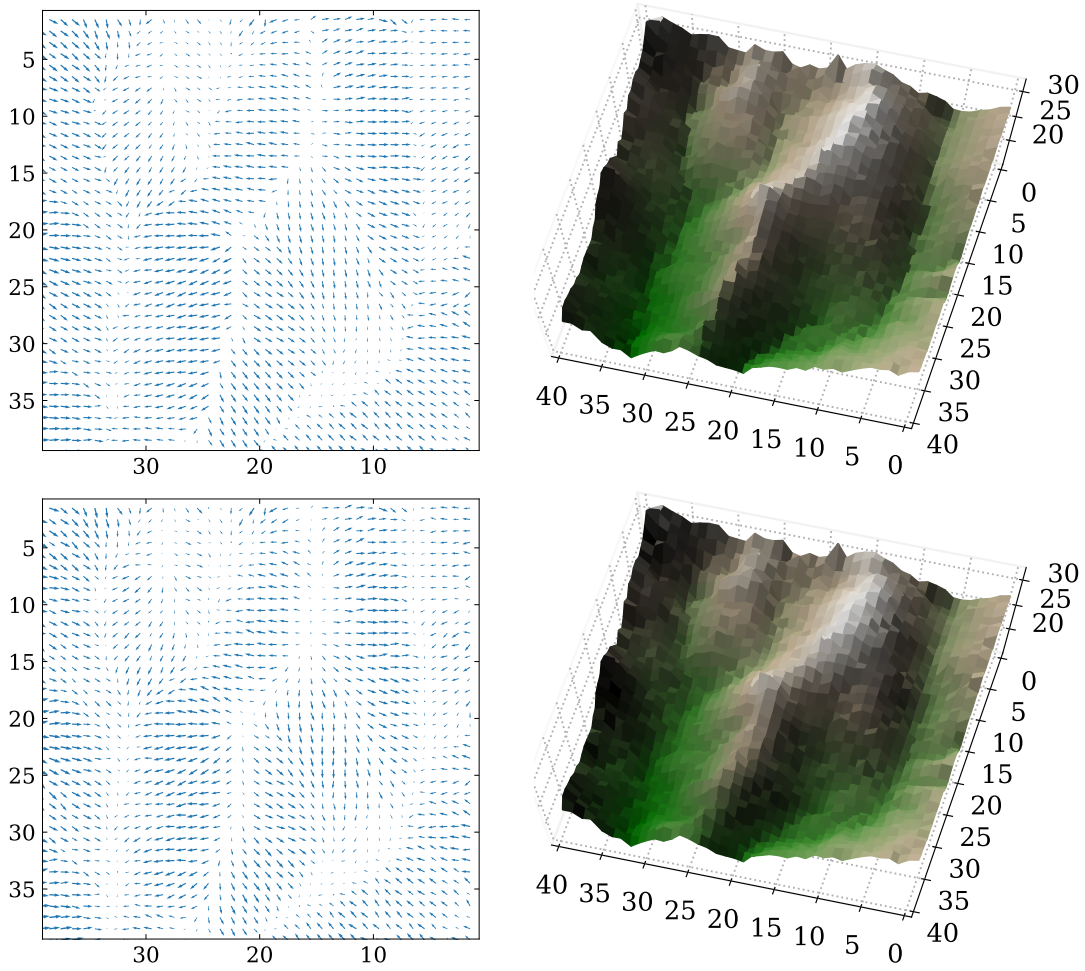


Figure 4.9: Denoising of $\mathbb{S}^2$-valued surface normals on the digital elevation model (DEM) dataset from [Ges+09] (noisy input is in Fig. 4.8). Mountain ridges are sharp while hillsides remain smooth with Huber (**top**, $\alpha = 0.1$, $\lambda = 0.75$). In the quadratically ($\lambda = 3.0$) denoised image **(bottom)** all contours are smoothed out.

many of the qualitative properties known from RGB and grayscale image processing appear to transfer to the manifold-valued case: TV enforces piecewise constant areas, such as flat hillsides, but preserves edges, such as mountain ridges (Fig. 4.8). Quadratic regularization gives overall very smooth results, but tends to lose edge information. With Huber regularization, edges (mountain ridges) remain sharp while hillsides are smooth, and flattening is avoided (Fig. 4.9).

4.3.4 Denoising of High-Resolution InSAR Data

While the resolution of the DEM dataset is quite limited (40×40 data points), an application to high resolution (432×426 data points) Interferometric Synthetic Aperture Radar (InSAR) denoising shows that our model is also applicable in a more demanding scenario (Fig. 4.10).

In InSAR imaging, information about terrain is obtained from satellite or aircraft by measuring the phase difference between the outgoing signal and the incoming reflected signal. This allows a very high relative precision, but no immediate absolute measurements, as all distances are only recovered modulo the wavelength. After normalization to $[0, 2\pi)$, the phase data is correctly viewed as lying on the one-dimensional unit sphere $\mathbb{S}^1$. Therefore, handling the data before any phase unwrapping is performed requires a manifold-valued framework.

Again, denoising with TV, Huber and quadratic regularizations demonstrates properties comparable to those known from scalar-valued image processing while all regularization approaches reduce noise substantially (Fig. 4.10).

4.4 Conclusion and Outlook

We provided an overview and framework for functional lifting techniques for the variational regularization of functions with values in arbitrary Riemannian manifolds. The framework is motivated from continuous multi-label relaxations, but generalizes these from the context of scalar and vectorial ranges to geometrically more challenging manifold ranges.

Using this approach, it is possible to solve variational problems for manifold-valued images that consist of a possibly non-convex data term and an arbitrary, smooth or non-smooth, convex first-order regularizer, such as quadratic, total variation or Huber. A refined discretization based on manifold finite element methods achieves sublabel-accurate results, which allows to use coarser discretization of the range and reduces computational effort compared to previous lifting approaches on manifolds.

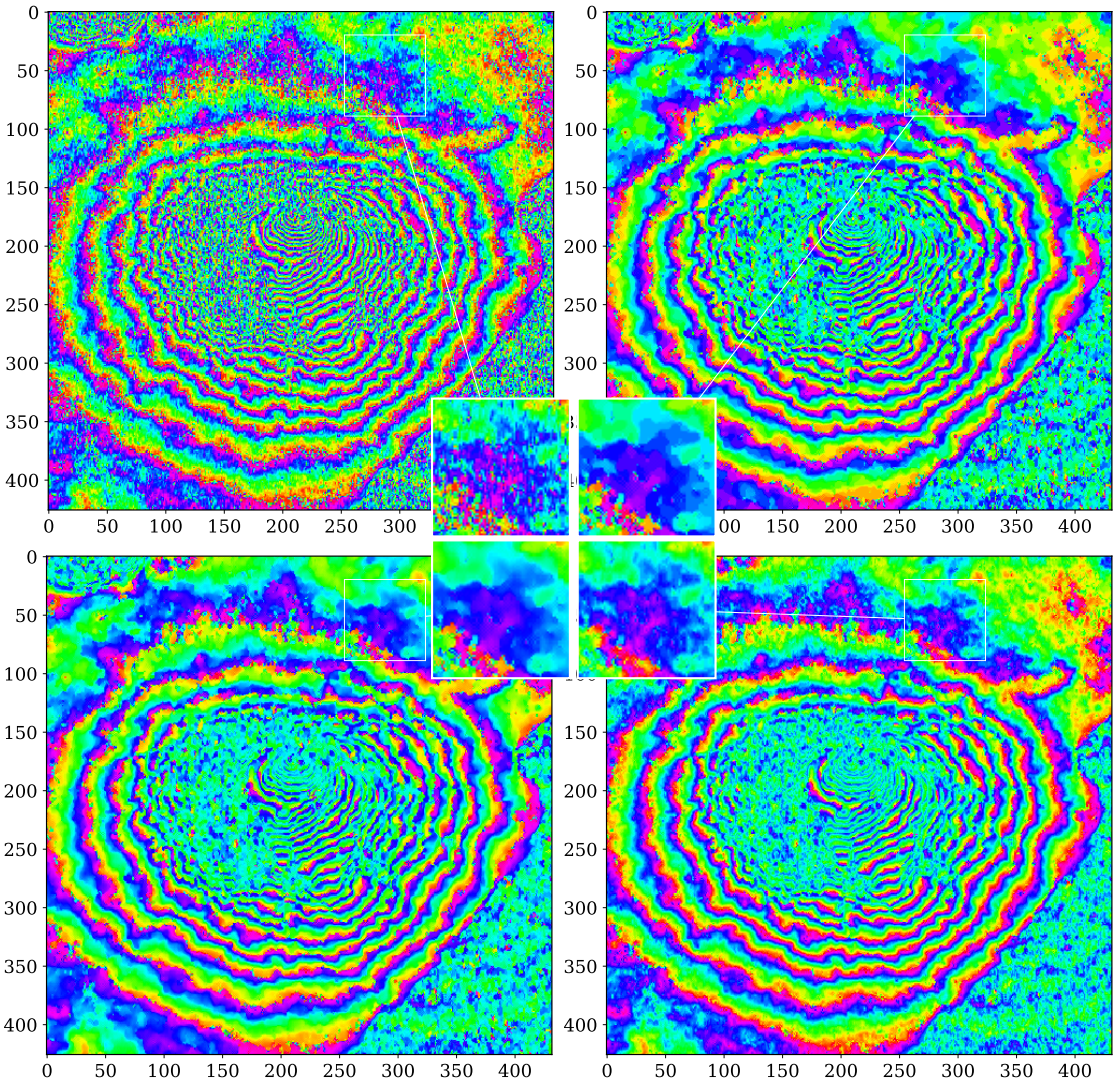


Figure 4.10: Denoising of $\mathbb{S}^1$-valued InSAR measurements from Mt. Vesuvius, dataset from [RPF97]: Noisy input **(top left)**, total variation ($\lambda = 0.6$) denoised image **(top right)**, Huber ($\alpha = 0.1$, $\lambda = 0.75$) denoised image **(bottom left)**, quadratically ($\lambda = 1.0$) denoised image **(bottom right)**. All regularization strategies successfully remove most of the noise. The total variation regularizer enforces clear contours, but exhibits staircasing effects. The staircasing is removed with Huber while contours are still quite distinct. Quadratic smoothing preserves some of the finer structures, but produces an overall more blurry and less contoured result.

A primary limitation of functional lifting methods, which equally applies to manifold-valued models, is dimensionality: The numerical cost increases exponentially with the dimensionality of the manifold due to the required discretization of the range. Addressing this issue appears possible, but will require a significantly improved discretization strategy.

5 Lifting of Functionals with Higher-Order Regularization

Let $\Omega \subset \mathbb{R}^d$ and $\Gamma \subset \mathbb{R}^s$ both be bounded sets. In the following, we consider the variational problem of minimizing the functional

$$F(u) = \int_\Omega f(x, u(x), \Delta u(x))dx, \tag{5.1}$$

that acts on vector-valued functions $u \in C^2(\Omega, \Gamma)$, whose Laplacian Δu is understood component-wise. Convexity of the integrand $f\colon \Omega \times \Gamma \times \mathbb{R}^s \to \mathbb{R}$ is only assumed in the last entry so that $u \mapsto F(u)$ is generally *non-convex*.

Variational problems of this form occur in a wide variety of image processing tasks, including image reconstruction, restoration, and interpolation. Commonly, the integrand is split into data term and regularizer:

$$f(x, z, p) = \rho(x, z) + \eta(p). \tag{5.2}$$

As an example, in *image registration* (sometimes referred to as large-displacement optical flow), the data term $\rho(x, z) = d(R(x), T(x + z))$ encodes the pointwise distance of a reference image $R\colon \mathbb{R}^d \to \mathbb{R}^k$ to a deformed template image $T\colon \mathbb{R}^d \to \mathbb{R}^k$ according to a given distance measure $d(\cdot, \cdot)$, such as the squared Euclidean distance $d(a, b) = \frac{1}{2}\|a-b\|_2^2$. While often a suitable convex regularizer η can be found, the highly non-convex nature of ρ renders the search for global minimizers of (5.1) a difficult problem.

Instead of directly minimizing F using gradient descent or other local solvers, we will aim to *replace* it by a convex functional $\mathcal{F}$ that acts on a higher-dimensional (*lifted*) function space (see introductory Sect. 1.2.2). If the lifting is chosen in such a way that we can construct global minimizers of F from global minimizers of $\mathcal{F}$, we can find a global solution of the original problem by applying convex solvers to $\mathcal{F}$. While we cannot claim this property for our choice of lifting, we believe that the mathematical motivation and some of the experimental results show that this approach can be a good basis for future work on global solutions of variational models with higher-order regularization.

5.1 Related Work and Overview

The lifted functional $\mathcal{F}$ proposed in this work is motivated by previous lifting approaches for *first-order* variational problems of the form

$$\inf_u \int_\Omega f(x, u(x), \nabla u(x))dx, \tag{5.3}$$

that are defined on functions $u\colon \Omega \to \Gamma$ with $\Omega \subset \mathbb{R}^d$ and $\Gamma \subset \mathbb{R}^s$. For a general overview of functional lifting, see the introductory Sect. 1.2.2 and Sect. 4.1 for manifold ranges.

Not much is known about functional lifting for models with higher-order regularizers, such as the Laplacian-based curvature regularizer in image registration [FM03]. Recently, a functional lifting approach has been successfully applied to second-order regularized image registration problems [LL18], but the approach was limited to a single regularizer, namely the integral over the 1-norm of the Laplacian (*absolute Laplacian regularization*). It is stated for scalar problems and then applied componentwise for vectorial problems. Similarly, the continuous lifting strategy for models with total generalized variation regularization [RPB13; SG19] is limited to a scalar range $\Gamma \subset \mathbb{R}$.

Contribution In Sect. 5.2, we propose a functional lifting approach in the fully continuous vector-valued setting for functionals that depend in a convex way on Δu. We show that the lifted functional satisfies $\mathcal{F}(\delta_u) \leq F(u)$, where δ_u is the lifted version of a function u and briefly discuss the question of whether the inequality is actually an equality. For the case of absolute Laplacian regularization, we show that our model is a generalization of [LL18]. Section 5.3 clarifies how convex saddle point solvers can be applied to our discretized model and Sect. 5.4 presents experimental results. We discuss the problem of projection and demonstrate that the model can be applied to image registration problems.

5.2 Functional Lifting with Vectorial Second-Order Terms

We propose the following lifted substitute for F:

$$\mathcal{F}(v) := \sup_{(p,q)\in X} \int_\Omega \int_\Gamma (\Delta_x p(x,z) + q(x,z))\, v_x(dz)dx, \tag{5.4}$$

acting on functions $v\colon \Omega \to \mathcal{P}(\Gamma)$ with values in the space $\mathcal{P}(\Gamma)$ of Borel probability measures on Γ. This means that, for each $x \in \Omega$ and every measurable set $U \subset \Gamma$, the

expression $v_x(U) \in \mathbb{R}$ can be interpreted as the "confidence" of an assumed underlying function on Ω to take a value inside of U at point x. A function $u\colon \Omega \to \Gamma$ can be *lifted* to a function $v\colon \Omega \to \mathcal{P}(\Gamma)$ by defining $v_x := \delta_{u(x)}$, the Dirac mass at $u(x) \in \Gamma$, for each $x \in \Omega$.

We propose the following set of test functions in the definition of $\mathcal{F}$:

$$\begin{aligned} X = \{(p,q) : p \in C_c^2(\Omega \times \Gamma), q \in L^1(\Omega \times \Gamma), & \quad (5.5)\\ z \mapsto p(x,z) \text{ concave} & \quad (5.6)\\ \text{and } q(x,z) + f^*(x,z,\nabla_z p(x,z)) \leq 0 & \quad (5.7)\\ \text{for every } (x,z) \in \Omega \times \Gamma\}, & \quad (5.8) \end{aligned}$$

where $f^*(x,z,q) := \sup_{p\in\mathbb{R}^s} \langle q,p\rangle - f(x,z,p)$ is the convex conjugate of f with respect to the last argument.

A thorough analysis of $\mathcal{F}$ requires a careful choice of function spaces in the definition of X as well as a precise definition of the properties of the integrand f and the admissible functions $v\colon \Omega \to \mathcal{P}(\Gamma)$. In Chap. 6, we will show that a primal formulation of $\mathcal{F}$ – as opposed to the dual formulation given here – is easier to handle in a certain sense. Here, we present a proof that the lifted functional $\mathcal{F}$ bounds the original functional F from below.

Proposition 5.1. *Let $f\colon \Omega \times \Gamma \times \mathbb{R}^s \to \mathbb{R}$ be measurable in the first two, and convex in the third entry, and let $u \in C^2(\Omega,\Gamma)$ be given. Then, for $v\colon \Omega \to \mathcal{P}(\Gamma)$ defined by $v_x := \delta_{u(x)}$, it holds that*

$$F(u) \geq \mathcal{F}(v). \tag{5.9}$$

Proof. Let p,q be a pair of functions satisfying the properties from the definition of X. By the chain rule, we compute

$$\begin{aligned} \Delta_x p(x,u(x)) = \Delta\left[p(x,u(x))\right] - \sum_{i=1}^{d} \langle \partial_i u(x), D_z^2 p(x,u(x)) \partial_i u(x)\rangle & \quad (5.10)\\ - 2\langle \nabla_x \nabla_z p(x,u(x)), \nabla u(x)\rangle - \langle \nabla_z p(x,u(x)), \Delta u(x)\rangle. & \end{aligned}$$

Furthermore, the divergence theorem ensures

$$\begin{aligned} -\int_\Omega \langle \nabla_x \nabla_z p(x,u(x)), \nabla u(x)\rangle dx = \int_\Omega \langle \nabla_z p(x,u(x)), \Delta u(x)\rangle dx & \quad (5.11)\\ + \int_\Omega \sum_{i=1}^{d} \langle \partial_i u(x), D_z^2 p(x,u(x)) \partial_i u(x)\rangle dx, & \end{aligned}$$

as well as $\int_\Omega \Delta\left[p(x,u(x))\right]dx = 0$ by the compact support of p. As $p \in C_c^2(\Omega\times\Gamma)$, concavity of $z \mapsto p(x,z)$ implies a negative semi-definite Hessian $D_z^2 p(x,z)$ so that, together with (5.10)–(5.11),

$$\int_\Omega \Delta_x p(x,u(x))\,dx \le \int_\Omega \langle \nabla_z p(x,u(x)), \Delta u(x)\rangle\,dx. \tag{5.12}$$

We conclude

$$\begin{aligned}
\mathcal{F}(v) &= \int_\Omega \int_\Gamma (\Delta_x p(x,z) + q(x,z))\,dv_x(z)dx && (5.13)\\
&= \int_\Omega \Delta_x p(x,u(x)) + q(x,u(x))\,dx && (5.14)\\
&\overset{(5.7)}{\le} \int_\Omega \Delta_x p(x,u(x)) - f^*(x,u(x),\nabla_z p(x,u(x)))\,dx && (5.15)\\
&\overset{(5.12)}{\le} \int_\Omega \langle \nabla_z p(x,u(x)), \Delta u(x)\rangle - f^*(x,u(x),\nabla_z p(x,u(x)))\,dx && (5.16)\\
&\le \int_\Omega f(x,u(x),\Delta u(x))\,dx, && (5.17)
\end{aligned}$$

where we used the definition of f^* in the last inequality. □

By a standard result from convex analysis [Roc97, Thm. 23.5], $\langle p,g\rangle - f^*(x,z,g) = f(x,z,p)$ whenever $g \in \partial_p f(x,z,p)$, the subdifferential of f with respect to p. Hence, for equality to hold in (5.9), we would need to find a function $p \in C_c^2(\Omega\times\Gamma)$ with

$$\nabla_z p(x,u(x)) \in \partial_p f(x,u(x),\Delta u(x)) \tag{5.18}$$

and associated $q(x,z) := -f^*(x,z,\Delta u(x))$, such that $(p,q) \in X$ or (p,q) can be approximated by functions from X.

If the integrand can be decomposed into $f(x,z,p) = \rho(x,z) + \eta(p)$ as in (5.2), with $\eta \in C^1(\mathbb{R}^s)$ and u sufficiently smooth, the optimal pair (p,q) in the sense of (5.18) is

$$\begin{aligned}
p(x,z) &:= \langle z, \nabla\eta(\Delta u(x))\rangle, && (5.19)\\
q(x,z) &:= \rho(x,z) - \eta^*(\nabla\eta(\Delta u(x))). && (5.20)
\end{aligned}$$

For nonsmooth u, these candidates can be approximated by compactly supported functions from the admissible set X using suitable cut-off functions on $\Omega\times\Gamma$.

We will see in Chap. 6 that considering F as a function of the pair $(u,\Delta u)$ instead of u leads to a functional lifting framework that avoids the problems discussed here.

Connection to the Discretization-First Approach [LL18]

In [LL18], data term ρ and regularizer η are lifted independently from each other for the case $\eta = \|\cdot\|_1$. Following the continuous multilabeling approaches in [CCP12; Möl+16; Lau+16], the setting is fully discretized in $\Omega \times \Gamma$ in a first step. Then the lifted data term and regularizer are defined to be the convex hull of a constraint function, which enforces the lifted terms to agree on the Dirac measures δ_u with the original functional applied to the corresponding function u. The data term is taken from [Lau+16], while the main contribution concerns the regularizer that now depends on the Laplacian of u.

In this section, we show that our fully continuous lifting is a generalization of the result from [LL18] after discretization.

Discretization In order to formulate the discretization-first lifting approach given in [LL18], we have to clarify the used discretization. Note that this discretization is loosely based on the one given in Sect. 2.7.

For the image domain $\Omega \subset \mathbb{R}^d$, discretized using points $X^1, \dots, X^N \in \Omega$ on a rectangular grid, we employ a finite-differences scheme: We assume that, on each grid point X^{i_0}, the discrete Laplacian of $u \in \mathbb{R}^{N,s}$, $u^i \approx u(X^i) \in \mathbb{R}^s$, is defined using the values of u on $m+1$ grid points $X^{i_0}, \dots, X^{i_m}$ such that

$$(\Delta u)^{i_0} = \textstyle\sum_{l=1}^{m} (u^{i_l} - u^{i_0}) \in \mathbb{R}^s. \tag{5.21}$$

For example, in the case $d = 2$, the popular five-point stencil means $m = 4$ and the X^{i_l} are the neighboring points of X^{i_0} in the rectangular grid. More precisely,

$$\textstyle\sum_{l=1}^{4} (u^{i_l} - u^{i_0}) = [u^{i_1} - 2u^{i_0} + u^{i_2}] + [u^{i_3} - 2u^{i_0} + u^{i_4}]. \tag{5.22}$$

The range $\Gamma \subset \mathbb{R}^s$ is triangulated into simplices $\Delta_1, \dots, \Delta_M$ with altogether L vertices (or *labels*) $Z^1, \dots, Z^L \in \Gamma$. We write $T := (Z^1 | \dots | Z^L)^\mathsf{T} \in \mathbb{R}^{L,s}$, and define the sparse indexing matrices $P^j \in \mathbb{R}^{s+1,L}$ in such a way that the rows of $T_j := P^j T \in \mathbb{R}^{s+1,s}$ are the labels associated to Δ_j.

There exist piecewise linear finite elements $\Phi_k \colon \Gamma \to \mathbb{R}$, $k = 1, \dots, L$ satisfying $\Phi_k(t_l) = 1$ if $k = l$, and $\Phi_k(t_l) = 0$ otherwise. In particular, the Φ_k form a partition of unity for Γ, i.e., $\sum_k \Phi_k(z) = 1$ for each $z \in \Gamma$. For a function $p \colon \Gamma \to \mathbb{R}$ in the function space spanned by the Φ_k, with a slight abuse of notation, we write $p = (p_1, \dots, p_L)$, where $p_k = p(Z^k)$ so that $p(z) = \sum_k p_k \Phi_k(z)$.

Functional lifting of the discretized absolute Laplacian Along the lines of classical continuous multilabeling approaches, we lift the absolute Laplacian regularizer using the convex hull of the constraint function $\phi : \mathbb{R}^L \to \mathbb{R} \cup \{+\infty\}$,

$$\phi(p) := \begin{cases} \mu \left\| \sum_{l=1}^{m} (T_{j_l}\alpha^l - T_{j_0}\alpha^0) \right\|, & \text{if } p = \mu \sum_{l=1}^{m} (P^{j_l}\alpha^l - P^{j_l}\alpha^0), \\ +\infty, & \text{otherwise,} \end{cases} \tag{5.23}$$

where $\mu \geq 0$, $\alpha^l \in \Delta_{s+1}^U$ (for Δ_{s+1}^U the unit simplex) and $1 \leq j_l \leq M$ for each $l = 0, \ldots, m$. The parameter $\mu \geq 0$ is enforcing positive homogeneity of ϕ which ensures that the convex conjugate ϕ^* of ϕ is given by the characteristic function $\delta_{\mathcal{K}}$ of a set $\mathcal{K} \subset \mathbb{R}^L$. Namely,

$$\mathcal{K} = \bigcap\nolimits_{1 \leq j_l \leq M} \{ f \in \mathbb{R}^L : \sum_{l=1}^{m} (f(t^l) - f(t^0)) \leq \left\| \sum_{l=1}^{m} (t^l - t^0) \right\|, \tag{5.24}$$

$$\text{for every } \alpha^l \in \Delta_{s+1}^U, l = 0, 1, \ldots, m \}, \tag{5.25}$$

where $t^l := T_{j_l}\alpha^l$ and $f(t^l)$ is the evaluation of the piecewise linear function f defined by the coefficients $(f_1, \ldots, f_L)$. The formulation of $\mathcal{K}$ comes with infinitely many constraints so far.

We now show two propositions which give a meaning to this set of constraints for arbitrary dimensions s of the labeling space and an arbitrary choice of norm in the definition of $\eta = \|\cdot\|$. They extend the component-wise (anisotropic) absolute Laplacian result in [LL18] to the vector-valued case.

Proposition 5.2. *The set $\mathcal{K}$ can be written as*

$$\mathcal{K} = \left\{ f \in \mathbb{R}^L : f \colon \Gamma \to \mathbb{R} \text{ is concave and 1-Lipschitz continuous} \right\}.$$

Proof. If the piecewise linear function induced by $f \in \mathbb{R}^L$ is concave and 1-Lipschitz continuous, then

$$\frac{1}{m} \sum_{l=1}^{m} (f(t^l) - f(t^0)) = \left(\frac{1}{m} \sum_{l=1}^{m} f(t^l) \right) - f(t^0) \tag{5.26}$$

$$\leq f\left(\frac{1}{m} \sum_{l=1}^{m} t^l \right) - f(t^0) \tag{5.27}$$

$$\leq \left\| \left(\frac{1}{m} \sum_{l=1}^{m} t^l \right) - t^0 \right\| \tag{5.28}$$

$$= \frac{1}{m} \left\| \sum_{l=1}^{m} (t^l - t^0) \right\|. \tag{5.29}$$

Hence, $f \in \mathcal{K}$. On the other hand, if $f \in \mathcal{K}$, then we recover Lipschitz continuity by choosing $t^l = t^1$, for each l in (5.24). For concavity, we first prove mid-point concavity. That is, for every $t^1, t^2 \in \Gamma$, we have

$$\frac{f(t^1)+f(t^2)}{2} \leq f\left(\frac{t^1+t^2}{2}\right) \tag{5.30}$$

or, equivalently, $[f(t^1) - f(t^0)] + [f(t^2) - f(t^0)] \leq 0$, where $t^0 = \frac{1}{2}(t^1 + t^2)$. This follows from (5.24) by choosing $t^0 = \frac{1}{2}(t^1 + t^2)$ and $t^l = t^0$ for $l > 2$. With this choice, the right-hand side of the inequality in (5.24) vanishes and the left-hand side reduces to the desired statement. Now, f is continuous by definition and, for these functions, mid-point concavity is equivalent to concavity [RV73, pp. 220–221]. □

The following theorem is an extension of [LL18, Thm. 1] to the vector-valued case and is crucial for numerical performance, as it shows that the constraints in Prop. 5.2 can be reduced to a finite number:

Proposition 5.3. *The set $\mathcal{K}$ can be expressed using not more than $|\mathcal{E}|$ (nonlinear) constraints, where $\mathcal{E}$ is the set of faces (or edges in the 2D-case) in the triangulation.*

Proof. Usually, Lipschitz continuity of a piecewise linear function requires one constraint on each of the simplices in the triangulation, and thus as many constraints as there are gradients. However, together with concavity, it suffices to enforce a gradient constraint on each of the boundary simplices, of which there are fewer than the number of outer faces in the triangulation. This can be seen by considering the one-dimensional case where Lipschitz constraints on the two outermost pieces of a concave function enforce Lipschitz continuity on the whole domain. Concavity of a function $f\colon \Gamma \to \mathbb{R}$ expressed in the basis (Φ_k) is equivalent to its gradient being monotonously decreasing across the common boundary between each pair of adjacent simplices (see [CL01, Lemma 3] and [Dud77, Thm. 3.1]). Together, we need one gradient constraint for each inner, and at most one for each outer face in the triangulation. □

5.3 Numerical Implementation

For the numerical experiments, we restrict ourselves to the special case of integrands $f(x, z, p) = \rho(x, z) + \eta(p)$ as motivated in the previous section.

Full discretization We base our discretization on the considerations in the previous section. For a function $p\colon \Gamma \to \mathbb{R}$ in the function space spanned by the Φ_k, we note that

$$p(z) = \sum_{k=1}^{L} p_k \Phi_k(z) = \langle A^j z - b^j, P^j p \rangle \text{ whenever } z \in \Delta_j, \tag{5.31}$$

where A^j and b^j are such that $\alpha = A^j z - b^j \in \Delta_{s+1}^U$ contains the barycentric coordinates of z with respect to Δ_j. More precisely, for $\bar{T}^j := (P^j T| - e)^{-1} \in \mathbb{R}^{s+1,s+1}$ with $e = (1, \dots, 1) \in \mathbb{R}^{s+1}$, we set

$$A^j := \bar{T}^j\texttt{(1:s,:)} \in \mathbb{R}^{s,s+1}, \tag{5.32}$$

$$b^j := \bar{T}^j\texttt{(s+1,:)} \in \mathbb{R}^{s+1}. \tag{5.33}$$

The functions $v\colon \Omega \to \mathcal{P}(\Gamma)$ are discretized as

$$v^{ik} := \int_\Gamma \Phi_k(z) dv_{X^i}(z), \tag{5.34}$$

hence $v \in \mathbb{R}^{N,L}$. Furthermore, whenever $v_x = \delta_{u(x)}$, the discretization v^i contains the barycentric coordinates of $u(X^i)$ relative to Δ_j. In the context of first-order models, this property is described as sublabel-accuracy in [Lau+16; MC17].

Dual admissibility constraints The admissible set X of dual variables is realized by discretizing the conditions (5.6) and (5.7).

Concavity (5.6) of a function $p\colon \Gamma \to \mathbb{R}$ expressed in the basis (Φ_k) is equivalent to its gradient being monotonously decreasing across the common boundary between each pair of neighboring simplices. This amounts to

$$\langle g^{j_2} - g^{j_1}, n_{j_1,j_2} \rangle \leq 0, \tag{5.35}$$

where g^{j_1}, g^{j_2} are the (piecewise constant) gradients $\nabla p(z)$ on two neighboring simplices $\Delta_{j_1}, \Delta_{j_2}$, and $n_{j_1,j_2} \in \mathbb{R}^s$ is the normal of their common boundary pointing from Δ_{j_1} to Δ_{j_2}.

The inequality (5.7) is discretized using (5.31) similar to the one-dimensional setting presented in [MC17]. We denote the dependence of p and q on $X^i \in \Omega$ by a superscript i as in q^i and p^i. Then, for each $j = 1, \dots, M$, we require

$$\sup_{z \in \Delta_j} \langle A^j z - b^j, P^j q^i \rangle - \rho(X^i, z) + \eta^*(g^{ij}) \leq 0 \tag{5.36}$$

which, for $\rho_j := \rho + \delta_{\Delta_j}$, can be formulated equivalently as

$$\rho_j^*(X^i, (A^j)^\mathsf{T} P^j q^i) + \eta^*(g^{ij}) \leq \langle b^j, P^j q \rangle. \tag{5.37}$$

The fully discretized problem can be expressed in convex-concave saddle point form

$$\begin{aligned} \min_u \max_{p,q,g} \quad & \textstyle\sum_i \langle u^i, (\Delta p)^i + q^i \rangle \\ \text{s.t.} \quad & \textstyle\sum_k u^{ki} = 1,\ u^i \geq 0, \\ & g^{ij} = B^j P^j p^i, \\ & \rho_j^*(X^i, (A^j)^\mathsf{T} P^j q^i) + \eta^*(g^{ij}) \leq \langle b^j, P^j q^i \rangle, \\ & \langle g^{j_2} - g^{j_1}, n_{j_1,j_2} \rangle \leq 0, \end{aligned}$$

to which we apply the primal-dual splitting algorithm with adaptive step sizes described in Sect. 2.6. The epigraph projections for ρ_j^* and η are implemented along the lines of [Möl+16] and [Poc+10].

5.4 Numerical Results

We implemented the proposed model in Python 3 (see `https://github.com/room-10/opymize`) with NumPy and PyCUDA. The examples were computed on an Intel Core i7 4.00 GHz with 16 GB of memory and an NVIDIA GeForce GTX 1080 Ti with 12 GB of dedicated video memory. The iteration was stopped when the Euclidean norms of the primal and dual residuals [GEB13] fell below $10^{-6} \cdot \sqrt{n}$ where n is the respective number of variables. We use this stopping criterion instead of the primal-dual gap (see introductory Sect. 2.6), as the evaluation of the primal objective function involves itself a minimization to avoid large numerical errors in the evaluation of the primal objective function.

Image registration We show that the proposed model can be applied to two-dimensional image registration problems (Figures 5.1 and 5.2). We used the sum of squared distances (SSD) data term $\rho(x, z) := \frac{1}{2}\|R(x) - T(x + z)\|_2^2$ and squared Laplacian (curvature) regularization $\eta(p) := \frac{1}{2}\|\cdot\|^2$. The image values $T(x + z)$ were calculated using bilinear interpolation with Neumann boundary conditions. After minimizing the lifted functional, we projected the solution by taking averages over Γ in each image pixel.

In the first experiment (Fig. 5.1), the reference image R was synthesized by numerically rotating the template T by 40 degrees. The grid plot of the computed deformation as well

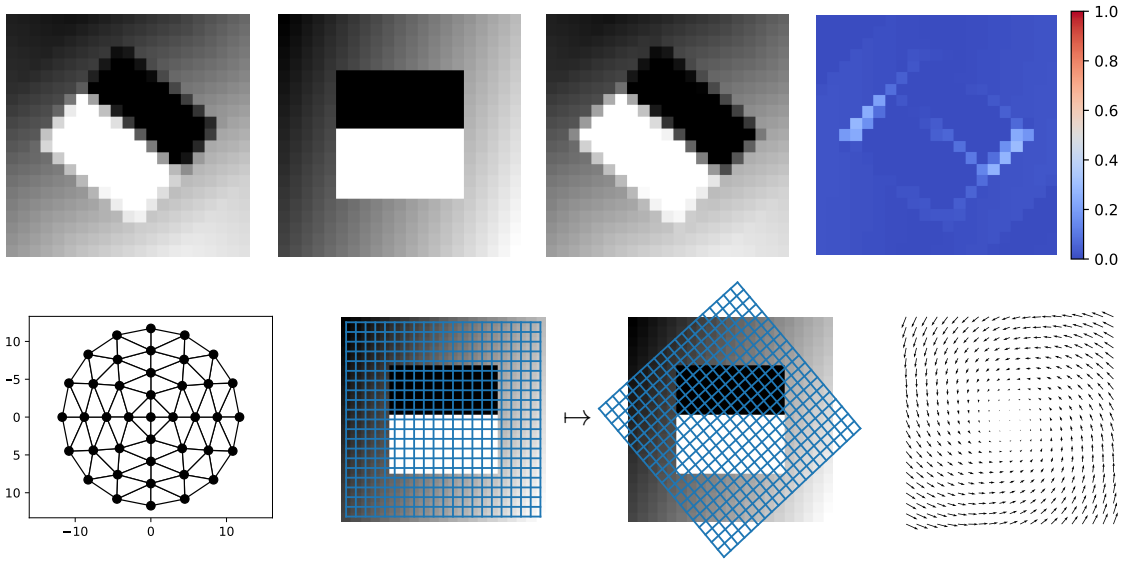


Figure 5.1: Application of the proposed higher-order lifting to image registration with SSD data term and squared Laplacian regularization. The method accurately finds a deformation **(bottom row, middle and right)** that maps the template image **(top row, second from left)** to the reference image **(top row, left)**, as also visible from the difference image **(top row, right)**. The result **(top row, second from right)** is almost pixel-accurate, although the range Γ of possible deformation vectors at each point is discretized using only 25 points **(second row, left)**.

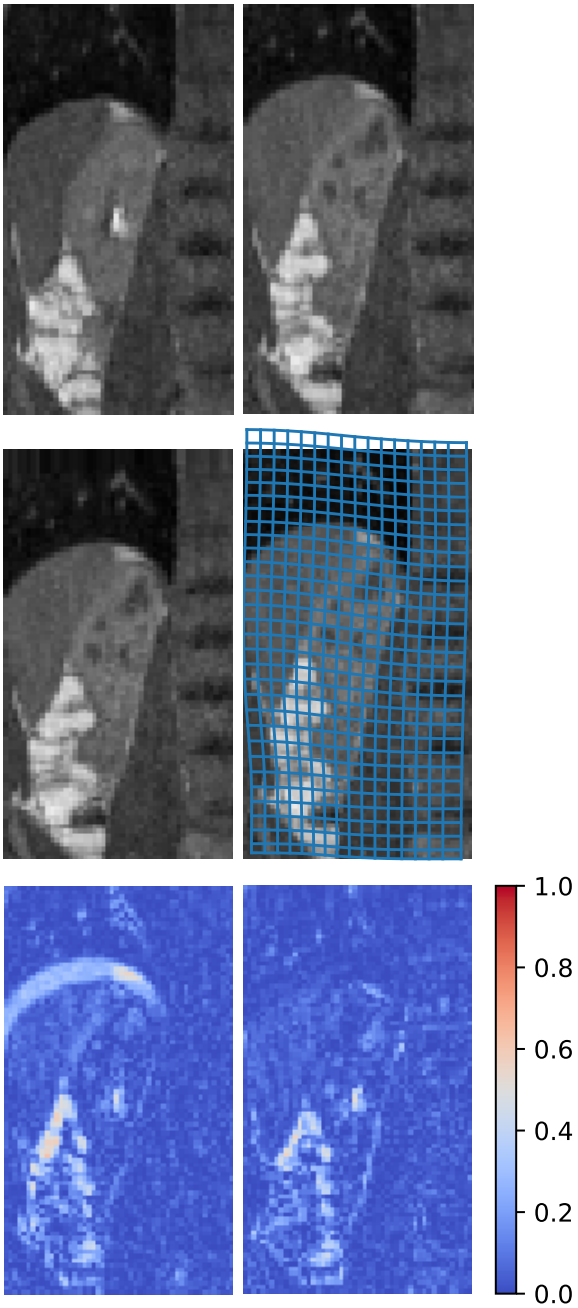


Figure 5.2: DCE-MRI data of a human kidney; data courtesy of Jarle Rørvik, Haukeland University Hospital Bergen, Norway; taken from [BWM18]. The deformation **(middle row)** mapping the template **(top left)** to the reference **(top right)** image, computed using our proposed model, is able to significantly reduce the misfit in the left half while fixing the spinal cord at the right edge as can be observed in the difference images from before **(bottom left)** and after **(bottom right)** registration.

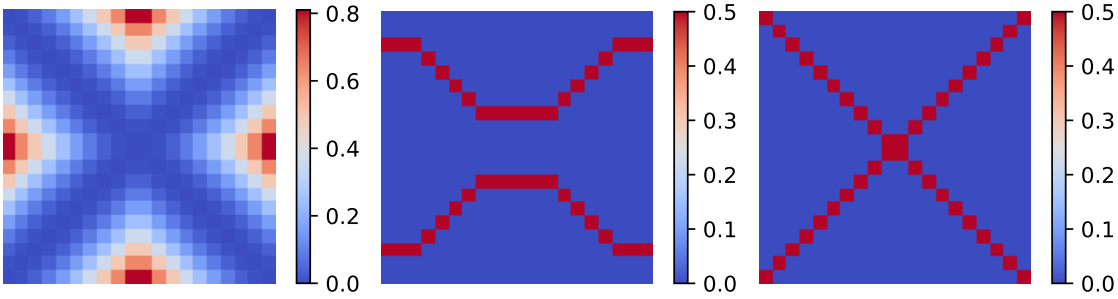


Figure 5.3: Minimizers of the lifted functional for a non-convex data term. **Left:** The data term $\rho(x, z) = (|x| - |z|)^2$. **Middle:** With classical first-order total variation-regularized lifting, the result is a composition of two solutions, which can be easily discriminated using thresholding. **Right:** For the new second-order squared-Laplacian regularized lifting, this simple approach fails to separate the two possible (straight line) solutions.

as the deformed template are visually very close to the rigid ground-truth deformation (a rotation by 40 degrees). Note that the method obtains almost pixel-accurate results although the range Γ of the deformation is discretized on a disk around the origin, triangulated using only 25 vertices, which is far less than the image resolution.

The second experiment (Fig. 5.2) consists of two coronal slices from a DCE-MRI dataset of a human kidney (data courtesy of Jarle Rørvik, Haukeland University Hospital Bergen, Norway; taken from [BWM18]). The deformation computed using our proposed model is able to significantly reduce the misfit in liver and kidney in the left half while accurately fixing the spinal cord at the right edge.

Projecting the lifted solution In the scalar-valued case with first-order regularization, the minimizers of the calibration-based lifting can be projected to minimizers of the original problem [Poc+10, Theorem 3.1]. In our notation, the thresholding technique used there corresponds to mapping v to

$$u(x) := \inf\{t : v_x((-\infty, t] \cap \Gamma) > s\}, \tag{5.38}$$

which is (provably) a global minimizer of the original problem for each $s \in [0, 1)$.

To investigate whether a similar property can hold in our higher-order case, we applied our model with Laplacian regularization $\eta(p) = \frac{1}{2}\|p\|^2$ as well as the calibration method approach with total variation regularization to the data term $\rho(x, z) = (|x| - |z|)^2$ with one-dimensional domain $\Omega = [-1, 1]$ and scalar data $\Gamma = [-1, 1]$ using 20 regularly-spaced discretization points (Fig. 5.3).

The result from the first-order approach is easily interpretable as a composition of two solutions to the original problem, each of which can be obtained by thresholding (5.38). In contrast, thresholding applied to the result from the second-order approach yields the two hat functions $u_1(x) = |x|$ and $u_2(x) = -|x|$, neither of which minimizes the original functional. Instead, the solution turns out to be of the form $v = \frac{1}{2}\delta_{u_3} + \frac{1}{2}\delta_{u_4}$, where u_3 and u_4 are in fact global minimizers of the original problem: namely, the straight lines $u_3(x) = x$ and $u_4(x) = -x$.

5.5 Conclusion and Outlook

In this chapter, we presented a novel fully continuous functional lifting approach for non-convex variational problems that involve Laplacian second-order terms and vectorial data, with the aim to ultimately provide sufficient optimality conditions and find global solutions despite the non-convexity. Experiments indicate that the method can produce accurate solutions for the non-convex image registration problem. We argued that more involved projection strategies than in the classical calibration approach will be needed for obtaining a good (approximate) solution of the original problem from a solution of the lifted problem. Another interesting direction for future work is the generalization to functionals that involve arbitrary second- or higher-order terms.

6 Measure-valued Liftings Motivated from Dynamical Optimal Transport

In this chapter, we propose a new mathematical framework for the description of a large class of first- and second-order functional lifting strategies for non-convex vectorial variational problems of the form

$$\inf_{u\in\mathcal{U}} F(u), \quad F(u) := \int_\Omega f(x, u(x), Lu(x))\,dx, \tag{6.1}$$

for $\Omega \subset \mathbb{R}^d$ open and bounded, $\Gamma \subset \mathbb{R}^s$ compact and L a linear differential operator of first or second order, i.e., $L = \nabla$ or $L = \nabla^2$ (furthermore, we explicitly discuss the case $L = \Delta$). According to the regularity imposed by L, we denote by $\mathcal{U}$ a suitable set of functions $u\colon \Omega \to \Gamma$ on which Lu is well-defined in a weak or strong sense, e.g., $\mathcal{U} = W^{1,1}(\Omega, \Gamma)$ or $\mathcal{U} = C^2(\Omega, \Gamma)$. The integrand f is expected to be convex in the third argument Lu and bounded from below, but f may be non-convex in u so that the whole problem is *non-convex* in general. Our proposed framework uses notions from dynamical optimal transport as introduced by Benamou and Brenier to define a *convex* functional $\mathcal{F}$ on the space of measures $\mathfrak{M}(\Omega \times \Gamma)$ with the property that

$$\mathcal{F}(\delta_u) = F(u) = \int_\Omega f(x, u(x), Lu(x))\,dx, \tag{6.2}$$

whenever $u\colon \Omega \to \Gamma$ is smooth enough and δ_u is the measure on $\Omega\times\Gamma$ that is concentrated on the graph of u (this measure is defined rigorously in (6.51) below). Minimizing $\mathcal{F}$ on the set $\{\delta_u : u \text{ smooth enough}\}$ is equivalent to the original problem (6.1). Our proposed lifting strategy then consists of minimizing the convex functional $\mathcal{F}$ over the *convex* set $\{\mu\colon \Omega \to \mathcal{P}(\Gamma)\}$ of measure-valued functions on Ω – a convex relaxation of the original problem in a higher-dimensional space.

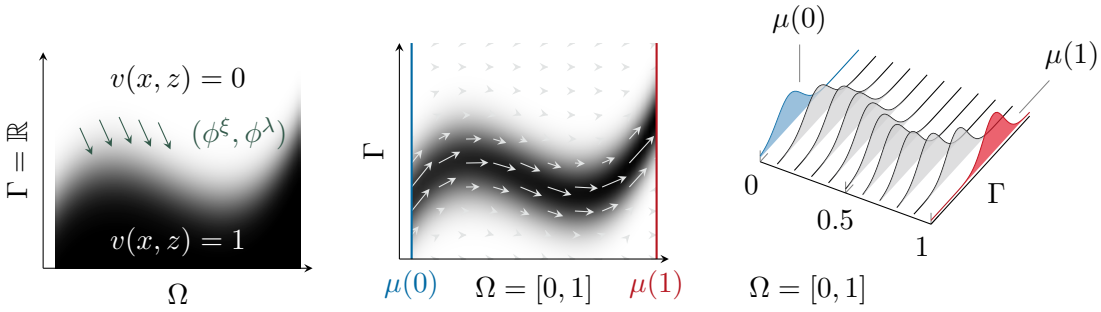


Figure 6.1: The theory of dynamical optimal transport exhibits conceptual similarities with state-of-the-art lifting strategies. **Left:** The calibration method-based lifting is defined on the space of functions $\mathrm{BV}(\Omega\times\mathbb{R},[0,1])$ that converge to 0 and 1 at ∞ and $-\infty$, respectively. A lifted functional is defined via dual pairings $\langle\phi, Dv\rangle$ with suitable vector fields $\phi=(\phi^\xi,\phi^\lambda)$ on the Cartesian product $\Omega\times\mathbb{R}$. **Center and Right:** In the Benamou-Brenier approach to dynamical optimal transport, the regularity of curves in the space of probability measures is measured by defining tangential vector fields on the Cartesian product $[0,1]\times\Gamma$. The derivative measures Dv from the calibration-based model can be interpreted to correspond to the tangential vector fields in dynamical optimal transport.

6.1 Related Work and Overview

Scalar-valued problems The functional lifting theory (see introductory Sect. 1.2.2) for scalar problems with first-order regularization, i.e., for (6.1) with $\Gamma=\mathbb{R}$ and $L=\nabla$, is well-understood when based on the theory of the calibration method as introduced in [ABD03; Poc+10]. In this case, the lifted functional $\mathcal{F}$ is defined on the space of functions $v\colon\Omega\times\mathbb{R}\to\mathbb{R}$ and, for each $u\in W^{1,1}(\Omega)$, satisfies $\mathcal{F}(\mathbb{1}_u)=F(u)$ where

$$\mathbb{1}_u(x,z):=\begin{cases}1 & \text{if } u(x)>z,\\ 0 & \text{otherwise.}\end{cases}\tag{6.3}$$

The convexified set of admissible functions is then defined as (a rigorous definition of this set is beyond the scope of this work)

$$\left\{v\in\mathrm{BV}(\Omega\times\mathbb{R},[0,1]):\lim_{z\to\infty}v(x,z)=0,\lim_{z\to-\infty}v(x,z)=1\right\}.\tag{6.4}$$

This framework has the property that, whenever v^* is a minimizer of $\mathcal{F}$, there is a minimizer u^* of F with $\mathcal{F}(v^*)=\min_u F(u)=F(u^*)$ that can be explicitly derived

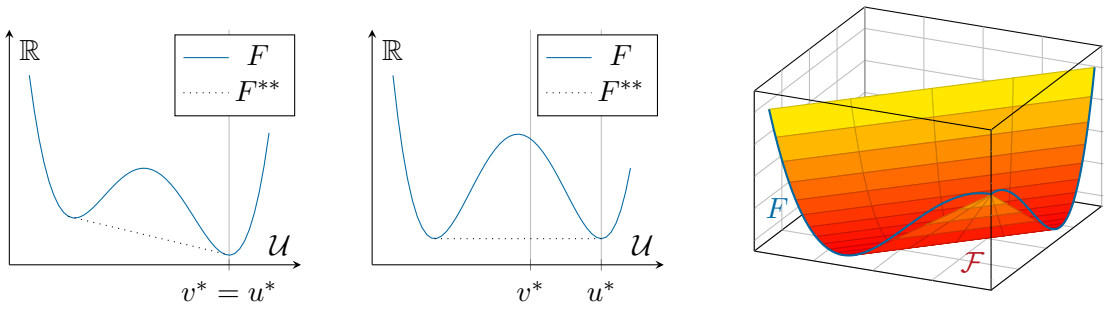


Figure 6.2: The convex hull F^{**} of a functional $F\colon \mathcal{U} \to \mathbb{R}$ defined on a Hilbert space $\mathcal{U}$ has the property that every minimizer $v^* \in \mathcal{U}$ of F^{**} satisfies $F^{**}(v^*) = \inf_{u\in\mathcal{U}} F(u)$. **Left:** Since F^{**} is convex, global minimizers can be computed efficiently. In some cases, it might hold that $F^{**}(v^*) = F(v^*)$. **Center:** In general, $F^{**}(v^*) < F(v^*)$ and it is a non-trivial task to derive any $u^* \in \mathcal{U}$ from v^* such that $F(u^*) = F^{**}(v^*)$. **Right:** Figuratively speaking, a "folding" of the objective functional along an additional axis is applied in the functional lifting theory for scalar variational problems with first-order regularization. Formally repeating the folding infinitely many times defines a convex functional $\mathcal{F}$ with the following property: Whenever v^* is a minimizer of $\mathcal{F}$, an explicit procedure is known how to compute a minimizer u^* of the original functional F. This is achieved at the price of a significant increase of dimension (*lifting*) in the domain of $\mathcal{F}$ compared to the domain of F, but it allows to apply efficient convex optimization algorithms and provably yields globally optimal solutions to the original non-convex minimization problem.

from v^* without any further optimization step. To demonstrate the power of this last property, we compare it with properties of the convex hull F^{**} of a functional $F\colon \mathcal{U} \to \mathbb{R}$ in case of a Hilbert space $\mathcal{U}$ (see Fig. 6.2): In fact, every minimizer $v^* \in \mathcal{U}$ of F^{**} satisfies $F^{**}(v^*) = \min_{u\in\mathcal{U}} F(u)$. But $F^{**}(v^*) < F(v^*)$ in general, and finding any $u^* \in \mathcal{U}$ such that $F(u^*) = F^{**}(v^*)$ is a non-trivial task in practice. On the other hand, F^{**} is defined on the same domain as F while $\mathcal{F}$ is defined on a space of much higher dimension, increasing numerical cost substantially. Our proposed measure-valued lifting framework is equivalent with the calibration-based lifting approach in the case of scalar first-order problems (see Sect. 6.4.1). However, for vectorial and higher-order problems, our approach does not come with a guarantee that minimizers of the lifted functional are somehow related to minimizers of the original functional, a trade-off that is shared with any other framework for the lifting of vectorial problems that has been investigated so far.

Vector-valued problems The concept of indicator functions of subgraphs $\mathbb{1}_u$ does not easily extend to vectorial data. The calibration criterion [ABD03] that is at the heart of the calibration-based lifting can be generalized to vectorial ranges [Mor02], but this generalization has not been translated into a corresponding functional lifting approach so far. Only very recently, a first attempt at a theory in the fully continuous vectorial setting using currents has been investigated in [MC19]. Based on this, general variational problems with manifold range $\Gamma = \mathcal{M}$ have been addressed [Vog+19] (see Chap. 4). Those approaches can be shown to be equivalent to our approach in the first-order case (Sect. 6.4.2). However, in contrast to our framework, neither of them easily extends to higher-order models.

Higher-order models Existing extensions of the functional lifting approach to higher-order models are restricted to a scalar range and to a single regularizer, such as the total generalized variation [RPB13; SG19] and the 1-norm of the Laplacian (*absolute Laplacian regularization*) [LL18]. On the other hand, the lifting approach [VL19] for more general Laplacian regularization terms applied to problems with vectorial range lacks a mathematically rigorous formulation (see also Chap. 5).

Dynamical optimal transport At the same time, the dynamical formulation of optimal transportation as introduced by Benamou and Brenier [Bre03] (see introductory Sect. 1.2.3) exhibits some striking conceptual similarities with some of these lifting theories (see Fig. 6.1). In fact, the Benamou-Brenier approach has been generalized to other

energy functionals in the language of one-dimensional currents [DGG06; Gra07; Gra09] and it inspired the concept of harmonic functions with values in the Wasserstein space as introduced in [Bre03]: A harmonic function on $\Omega \subset \mathbb{R}^d$ with values in $\mathcal{P}(\Gamma)$ is a pair of measures (μ, E) on $\Omega \times \Gamma$ valued in $\mathbb{R}$ and $\mathbb{R}^{d,n}$ that minimizes the energy

$$\int_{\Omega\times\Gamma} \|v\|_F^2 \, d\mu \tag{6.5}$$

subject to the continuity equation $\nabla_x \mu + \mathrm{Div}_z E = 0$ (in the sense of distributions, see Defn. 6.17 for a rigorous definition) with $E := v\mu$ for $v\colon \Omega \times \Gamma \to \mathbb{R}^{d,n}$ measurable and prescribed boundary data on $\partial\Omega$. We propose to apply this scheme to more general energy functionals and to more general continuity equations to lift problems such as (6.1) in the language of measure-valued functions:

Definition 6.1. For a measure-valued function $\mu\colon \Omega \to \mathcal{P}(\Gamma)$, we write $\mu_x := \mu(x)$ and we call μ *weakly measurable* if $x \mapsto \int_\Gamma \phi \, d\mu_x$ is measurable on Ω for each $\phi \in C_0(\Gamma)$. We denote by $L^\infty_w(\Omega, \mathcal{P}(\Gamma))$ the weakly measurable functions on Ω with values in $\mathcal{P}(\Gamma)$.

Our approach is easier accessible for readers that are not familiar with the theory of currents and the notation of differential geometry and exterior algebra. For simplicity, we restrict ourselves to the case of convex integrands rather than the general polyconvex setting, which avoids the need for introducing currents and exterior algebra.

Contribution In Sect. 6.2, we define a lifting $\mathcal{B}_f$ for *augmented* functionals

$$F(p, u) := \int_\Omega f(x, u(x), p(x)) \, dx \tag{6.6}$$

with the property that $\mathcal{B}_f(p\delta_u, \delta_u) = F(u, p)$ (Cor. 6.13). In Sect. 6.3 we translate the relationship $p = Lu$ to equivalent constraints in the lifted setting in the form of weak continuity equations for the cases $L = \nabla$ (Sect. 6.3.1), $L = \nabla^2$ (Sect. 6.3.2) and $L = \Delta$ (Sect. 6.3.3). In each case, we formulate the full lifting model and discuss some of its properties. Finally, in Sect. 6.4, we clarify its relationship to existing lifting and measure-valued frameworks, in particular the ones based on the calibration method (Sect. 6.4.1) and on the theory of currents (Sect. 6.4.2).

6.2 Integral Representation of Lifted Functional

For this section, let $U \subset \mathbb{R}^n$ be nonempty and let $f\colon U \times \mathbb{R}^m \to \overline{\mathbb{R}}$ be convex in the second argument. By $\overline{\mathbb{R}} := [-\infty, \infty]$ we denote the extended real line. Convexity of functions

with values in $\overline{\mathbb{R}}$ is defined via convexity of their epigraphical set and $g\colon \mathbb{R}^m \to \overline{\mathbb{R}}$ is said to be lower semicontinuous at $t^* \in \mathbb{R}^m$ if, for each $M < g(t^*)$, there exists $\delta > 0$ such that, for each $t \in B_\delta(t^*)$ we have $M \leq g(t)$. For the final functional lifting framework, we will set $U := \Omega \times \Gamma \subset \mathbb{R}^d \times \mathbb{R}^s$, $n := d + s$ and $m := ds$. We propose to use as a lifted version of (6.6) the *generalized Benamou-Brenier functional*,

$$\mathcal{B}_f(\nu) := \sup_{\phi \in \mathcal{K}_f} \langle \nu, \phi \rangle, \tag{6.7}$$

where $\nu \in \mathfrak{M}(U, \mathbb{R}^{m+1})$, by $\langle \nu, \phi \rangle$ we denote the dual pairing between $\nu \in \mathfrak{M}(U, \mathbb{R}^{m+1})$ and $\phi \in C_0(U, \mathbb{R}^{m+1})$ (see Sect. 2.1), and we define

$$\mathcal{K}_f := \{\phi = (\phi^\xi, \phi^\lambda) \in C_0(U, \mathbb{R}^m \times \mathbb{R}) : \phi^\lambda(t) + f^*(t, \phi^\xi(t)) \leq 0 \ \forall t \in U\}. \tag{6.8}$$

Here $f^*(t, \zeta) := \sup_{\xi \in \mathbb{R}^m} \langle \xi, \zeta \rangle - f(t, \xi)$ is the convex conjugate of f with respect to the second variable. The name *generalized Benamou-Brenier functional* and the symbol $\mathcal{B}_f$ are motivated by [San15, Section 5.3.1], where the name *Benamou-Brenier functional* is used for the special case

$$f(t, \xi) = \frac{1}{q}\|\xi\|^q, \quad q \in (1, \infty). \tag{6.9}$$

We start by stating the following basic properties of $\mathcal{B}_f$:

Proposition 6.2. *Let $f\colon U \times \mathbb{R}^m \to \overline{\mathbb{R}}$ be nonnegative, lower semicontinuous in both variables and convex in the second variable. Then the functional $\mathcal{B}_f$ is nonnegative, convex and lower semicontinuous on $\mathfrak{M}(U, \mathbb{R}^{m+1})$.*

Proof. $\mathcal{B}_f$ is defined as a pointwise supremum over continuous linear functions, since $\mathfrak{M}(U, \mathbb{R}^{m+1})$ is the dual space of $C_0(U, \mathbb{R}^{m+1})$. This implies lower-semicontinuity and convexity. Nonnegativity follows from nonnegativity of f since this implies that $\phi = (0, 0)$ is in $\mathcal{K}_f$. □

The goal of this section is to derive a primal formulation (an integral representation) for the functional $\mathcal{B}_f$ that will clarify the relation between $\mathcal{B}_f$ and (6.6). This includes a careful choice of suitable regularity conditions on the integrand f and its domain U.

6.2.1 Finite-Dimensional Analogon

We first note that $\mathcal{K}_f$ is a convex subset of $C_0(U, \mathbb{R}^{m+1})$, whose dual space is $\mathfrak{M}(U, \mathbb{R}^{m+1})$ (see Sect. 2.1). From this perspective, $\mathcal{B}_f$ is actually the support function of a convex

set in an infinite-dimensional setting. Furthermore, the specific form of $\mathcal{K}_f$ is a pointwise variant of a convex set whose support function is well-understood in standard finite-dimensional convex analysis and known in the literature under the name *perspective function*:

Proposition 6.3 ([Roc97, Corollary 13.5.1])**.** *Let $f\colon \mathbb{R}^m \to \overline{\mathbb{R}}$ be a proper convex function that is lower semicontinuous. Then the support function of*

$$K_f := \{(\xi, \lambda) : \lambda + f^*(\xi) \leq 0\} \subset \mathbb{R}^{m+1} \tag{6.10}$$

is the perspective function $h\colon \mathbb{R}^{m+1} \to \overline{\mathbb{R}}$ *of f defined by*

$$h(\xi, \lambda) := \begin{cases} \lambda f(\xi/\lambda) & \text{if } \lambda > 0, \\ f^\infty(\xi) & \text{if } \lambda = 0, \\ +\infty & \text{if } \lambda < 0, \end{cases} \tag{6.11}$$

Here, $f^\infty(\xi) := \lim_{s\to\infty} f(s\xi)/s$ is the recession function *of f.*

Remark 6.4. The function h is convex and positively 1-homogeneous in (ξ, λ). Many examples of perspective functions with their corresponding properties and applications are discussed in [Com18].

We are going to show that a similar statement holds in our pointwise infinite-dimensional setting. This requires a careful analysis of measurability and semicontinuity of all functions involved that we will discuss in the following subsections.

6.2.2 Main Result: Statement of Integral Representation

We define the function $h\colon U \times \mathbb{R}^{m+1} \to \overline{\mathbb{R}}$ by

$$h(t, \xi, \lambda) := \begin{cases} \lambda f(t, \xi/\lambda) & \text{if } \lambda > 0, \\ f^\infty(t, \xi) & \text{if } \lambda = 0, \\ +\infty & \text{if } \lambda < 0. \end{cases} \tag{6.12}$$

Moreover, we say that $U \subset \mathbb{R}^n$ is locally compact if, for each $t \in U$, there are $V \subset \mathbb{R}^n$ open and $C \subset U$ compact with $t \in (V \cap U) \subset C$ (or, equivalently, each $t \in U$ has a compact neighborhood in U).

Then, our main result in this section is that the explicit formula for the support function of K_f in the finite-dimensional setting carries over to the case of the infinite-dimensional

set $\mathcal{K}_f$ in the following sense:

Theorem 6.5. *Let $U \subset \mathbb{R}^n$ be locally compact and suppose that $f\colon U \times \mathbb{R}^m \to \overline{\mathbb{R}}$ is lower semicontinuous in both variables, convex in the second variable, and such that $t \mapsto f(t,0)$ is locally bounded. Then, the generalized Benamou-Brenier functional $\mathcal{B}_f$ as defined in* (6.7) *satisfies, for each $\nu \in \mathfrak{M}(U, \mathbb{R}^{m+1})$,*

$$\sup_{\phi \in \mathcal{K}_f} \langle \phi, \nu \rangle = \mathcal{B}_f(\nu) = \int_U h(t, (d\nu/d\|\nu\|)(t))\, \|\nu\|(dt). \tag{6.13}$$

Here, $\|\nu\|$ is the total variation measure of ν and $(d\nu/d\|\nu\|)\colon U \to \mathbb{R}^{m+1}$ denotes the corresponding vectorial density of the vector measure ν (its Radon-Nikodym derivative with respect to $\|\nu\|$, see Sect. 2.1).

Remark 6.6. For now, we simply assumed local compactness of U, but later this will follow from $U = \Omega \times \Gamma$ with Ω open and bounded and Γ compact (see also [Wil70, Theorem 18.6]). In many applications in image processing, f splits up as $f(t,\xi) = \rho(t) + \eta(\xi)$ with ρ bounded and $\eta(0) = 0$, which will imply the local boundedness assumption.

For the integral in (6.13) to be well-defined, h needs to satisfy appropriate measurability properties. In fact, h inherits lower semicontinuity from f if f is sufficiently regular, as noted in the proof of [Dal79, Theorem 3.1]:

Lemma 6.7. *Suppose that $U \subset \mathbb{R}^n$ is nonempty and $f\colon U \times \mathbb{R}^m \to \overline{\mathbb{R}}$ is lower semicontinuous in both variables and convex in the second variable. If $t \mapsto f(t,0)$ is locally bounded, then h as defined in* (6.12) *is lower semicontinuous (jointly in all variables).*

Remark 6.8. The following two examples show that, while it cannot be skipped without replacement, the criterion of local boundedness at $0 \in \mathbb{R}^m$ is not a necessary criterion for lower semicontinuity of h.

1. The function $f\colon [0,1] \times \mathbb{R} \to \mathbb{R}$,

$$f(t,\xi) := \begin{cases} 1/t, & \text{if } t > 0, \\ 0, & \text{otherwise,} \end{cases} \tag{6.14}$$

   is lower semicontinuous in both variables, but locally unbounded at $t^* = 0$ for each

$\xi \in \mathbb{R}$. Still, the corresponding perspective function

$$h(t,\xi,\lambda) := \begin{cases} \lambda/t, & \text{if } t > 0, \lambda \geq 0 \\ 0, & \text{if } t = 0, \lambda \geq 0, \\ +\infty, & \text{otherwise,} \end{cases} \tag{6.15}$$

is obviously lower semicontinuous in all variables.

2. The function $f\colon [0,1] \times \mathbb{R} \to \mathbb{R}$,

$$f(t,\xi) := \begin{cases} |\xi - 1/t| & \text{if } t > 0, \\ |\xi| & \text{otherwise,} \end{cases} \tag{6.16}$$

is lower semicontinuous in both variables and locally unbounded at $t^* = 0$ (for each $\xi \in \mathbb{R}$). The corresponding perspective function

$$h(t,\xi,\lambda) := \begin{cases} |\xi - \lambda/t| & \text{if } t > 0, \\ |\xi| & \text{otherwise,} \end{cases} \tag{6.17}$$

fails to be lower semicontinuous at $(t,\xi,\lambda) = (0,1,0)$ as can be seen from

$$h(1/k, 1, 1/k) = 0 < 1 = h(0,1,0). \tag{6.18}$$

Proof of Lemma 6.7. Since h is obviously lower semicontinuous at all points (t,ξ,λ) with $\lambda \neq 0$, it suffices to consider lower semicontinuity at $(t^*,\xi^*,0) \in U \times \mathbb{R}^m \times \mathbb{R}$. We show that, for $M < h(t^*,\xi^*,0)$, there exists $\delta > 0$ such that, for each $t \in B_\delta(t^*)$, $\xi \in B_\delta(\xi^*)$ and $\lambda \in B_\delta(0)$,

$$M \leq h(t,\xi,\lambda). \tag{6.19}$$

At first, let $\delta_1 > 0$ and $f_{\max} > 0$ be such that $|f(t,0)| \leq f_{\max}$ for each $t \in B_{\delta_1}(t^*)$. Let $\varepsilon > 0$ be such that $M + 2\varepsilon < h(t^*,\xi^*,0)$. From the definition of f^∞, there is $\delta_2 < \min\{\delta_1, \varepsilon/f_{\max}\}$ such that

$$M + 2\varepsilon \leq h(t^*,\xi^*,\delta_2). \tag{6.20}$$

On the other hand, lower semicontinuity at (t^*,ξ^*,δ_2) implies that there is $\delta \in (0,\delta_2)$

such that, for each $t \in B_\delta(t^*)$ and $\xi \in B_\delta(\xi^*)$,

$$M + \varepsilon \leq h(t, \xi, \delta_2). \tag{6.21}$$

By convexity of f, we can compute, for each $\lambda \in (0, \delta)$,

$$\begin{aligned} h(t, \xi, \delta_2) &= \delta_2 f(t, \xi/\delta_2) && (6.22)\\ &\leq \lambda f(t, \xi/\lambda) + (\delta_2 - \lambda) f(t, 0) && (6.23)\\ &\leq h(t, \xi, \lambda) + \delta_2 f(t, 0) && (6.24)\\ &\leq h(t, \xi, \lambda) + \varepsilon. && (6.25) \end{aligned}$$

Together we deduce, for each $t \in B_\delta(t^*)$, $\xi \in B_\delta(\xi^*)$ and $\lambda \in (0, \delta)$,

$$M \leq h(t, \xi, \lambda). \tag{6.26}$$

Note that the inequality trivially holds for $\lambda < 0$. □

6.2.3 Proof of the Main Result

The main result Thm. 6.5 is a consequence of a well-known statement of which variants have been known since the 1960s [Roc68; Roc71] at least. To state the general result, let us first introduce some notation: For sets X and Y, we write $Q\colon X \rightrightarrows Y$ if Q is a set-valued function, i.e., for each $x \in X$, the value $Q(x)$ is a subset of Y. The preimage $Q^{-1}(V)$ of $V \subset Y$ under Q is defined via

$$Q^{-1}(V) := \{x \in X : Q(x) \cap V \neq \emptyset\}. \tag{6.27}$$

For topological spaces X and Y we call a set-valued function $Q\colon X \rightrightarrows Y$ *inner semi-continuous* if $Q^{-1}(V) \subset X$ is open for each open set $V \subset Y$ (Fig. 6.3). Note that inner semicontinuity does not "correspond" to the usual concept of lower or upper semicontinuity in case of single-valued functions. Furthermore, in case of a normed vector space Y, we define the associated set of continuous selections

$$C(Q) := \{\phi \in C_0(X, Y) : \phi(x) \in Q(x), \forall x \in X\}. \tag{6.28}$$

For a topological vector space Y and a convex set $V \subset Y$, we write $\sigma_V := \delta_V^*$ for the support function of V, the convex conjugate of the convex indicator function δ_V of V (see Sect. 2.5). Finally, a second-countable locally compact Hausdorff space is a topological

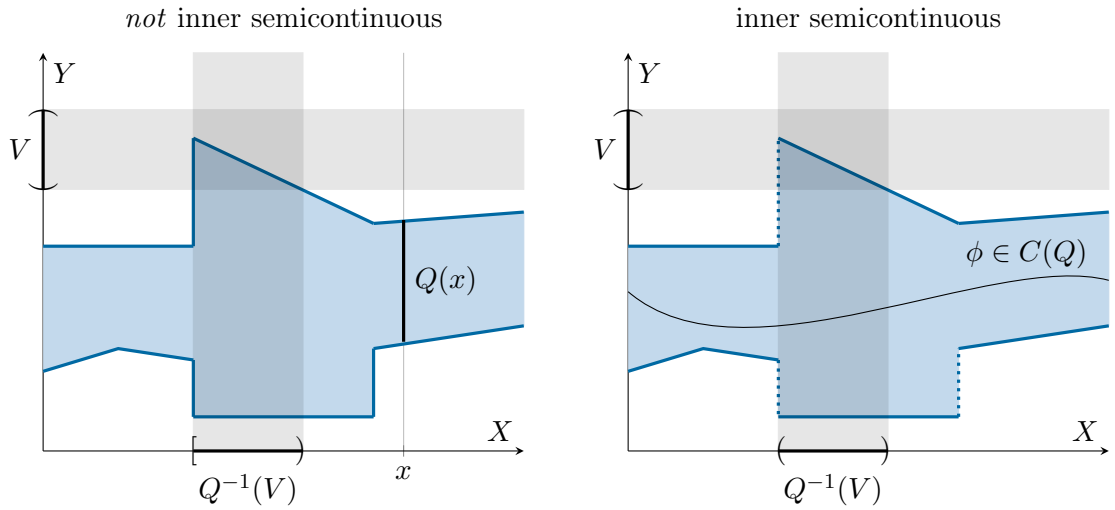


Figure 6.3: A set-valued function $Q\colon X \rightrightarrows Y$ is a function that assigns a *set* $Q(x)$ to each point $x \in X$. We call Q inner semicontinuous if the preimage $Q^{-1}(V)$ is open for each open set $V \subset Y$. **Left:** If the graph of a set-valued function contains vertical boundaries, the function fails to be inner semicontinuous. **Right:** Michael's selection theorem (Thm. 6.10) says that, under suitable assumptions on X, Y and Q, inner semicontinuity of Q is a sufficient condition for the existence of a *continuous selection* for Q, i.e., a continuous function $\phi\colon X \to Y$ with $\phi(x) \in Q(x)$ for each $x \in X$.

space where points can be separated by disjoint neighborhoods, where every point has a compact neighborhood and where there exists a countable family $\mathcal{U}$ of open sets such that every open set can be written as the union of sets from $\mathcal{U}$. Every separable metric space (including every subset of $\mathbb{R}^n$) is second-countable and Hausdorff [Wil70, Theorem 16.11] and we will only be interested in subsets of $\mathbb{R}^n$ in this chapter. For subsets $U \subset \mathbb{R}^n$, local compactness is satisfied, for example, if U is open or closed. An example of a set that fails to be locally compact is

$$\big\{(t_1, t_2) \in \mathbb{R}^2 : t_2 > 0\big\} \cup \{(0,0)\}, \tag{6.29}$$

which fails to have a compact neighborhood at $t = (0,0)$.

With this notation at hand, we will demonstrate in this subsection that Thm. 6.5 is a special case of the following

Theorem 6.9 ([Per18, Theorem 1])**.** *Assume X is a second-countable locally compact Hausdorff space and let $Q\colon X \rightrightarrows \mathbb{R}^m$ be such that $C(Q) \neq \emptyset$ and such that, for each $x \in X$, the set $Q(x)$ is closed and convex. Then the following two statements are equivalent*

1. *The support function of $C(Q)$ on $\mathfrak{M}(X, \mathbb{R}^m)$ is given by the formula*

$$\sup_{\phi \in C(Q)} \langle \nu, \phi \rangle = \sigma_{C(Q)}(\nu) = \int_X \sigma_{Q(x)}((d\nu/d\theta)(x))\, \theta(dx), \tag{6.30}$$

   whenever $\nu \in \mathfrak{M}(X, \mathbb{R}^m)$ and θ is a nonnegative measure in $\mathfrak{M}(X)$ with respect to which ν is absolutely continuous.

2. *Q is inner semicontinuous.*

In order to prove Thm. 6.9, we aim to apply this general result to the case where $X = U$ with U as in the statement of Thm. 6.5 and $Q\colon U \rightrightarrows \mathbb{R}^{m+1}$ is defined via

$$Q(t) := \{(\xi, \lambda) \in \mathbb{R}^m \times \mathbb{R} : \lambda + f^*(t, \xi) \leq 0\}. \tag{6.31}$$

First, U is a subset of $\mathbb{R}^n$, hence a separable metric space and, thus, second-countable and Hausdorff. Note that, by the definition of the convex conjugate,

$$Q(t) = \{(\xi, \lambda) \in \mathbb{R}^{m+1} : \lambda + \langle \xi, \zeta \rangle \leq f(t, \zeta)\ \forall \zeta \in \mathbb{R}^m\}. \tag{6.32}$$

By this alternative representation, we immediately see that Q is non-empty, closed and convex-valued. It remains to show that $C(Q) \neq \emptyset$ and that Q is inner semicontinuous.

Existence of a continuous selection for Q By Michael's selection theorem [Mic56], $C(Q) \neq \emptyset$ is a consequence of inner semicontinuity (which will be shown in the following paragraph):

Theorem 6.10 ([Mic56, Theorem 3.2"])**.** *Assume that X is a separable metric space, and let $Q\colon X \rightrightarrows \mathbb{R}^m$ be an inner semicontinuous set-valued mapping that has nonempty, convex and closed values. Then $C(Q) \neq \emptyset$.*

Inner semicontinuity of Q The inner semicontinuity of a mapping similar to Q is proved in [GMS98b, Sect. 1.1.3]. We present a proof of inner semicontinuity based on arguments from the proof of [BV88, Theorem 8] that requires less restrictive assumptions on f than the proof given in [GMS98b, Sect. 1.1.3]. In addition, we show that the lower semicontinuity of f is also a necessary condition for Q to be inner semicontinuous.

Proposition 6.11. *Suppose that $U \subset \mathbb{R}^n$ is nonempty, $f\colon U \times \mathbb{R}^m \to \mathbb{R}$ is convex in the second variable and that $t \mapsto f(t,0)$ is locally bounded. Let the map $Q\colon U \rightrightarrows \mathbb{R}^{m+1}$ be defined as in* (6.31)*. Then, the map Q is inner semicontinuous if and only if f is lower semicontinuous jointly in both variables.*

Proof. Suppose that f is lower semicontinuous, and, for the sake of contradiction, suppose that Q fails to be inner semicontinuous. That means, there exists an open set $V \subset \mathbb{R}^{m+1}$ such that $Q^{-1}(V)$ is not open. Choose $t^* \in Q^{-1}(V)$ such that no open neighborhood of t^* is contained in $Q^{-1}(V)$. We can assume $V = B_\varepsilon(\xi,\lambda)$ for some $\varepsilon > 0$ and $(\xi,\lambda) \in Q(t^*)$. Then there is a sequence $t^k \to t^*$ such that $B_\varepsilon(\xi,\lambda) \cap Q(t^k) = \emptyset$. After subtracting the linear and lower semicontinuous function $\zeta \mapsto \langle \xi, \zeta\rangle + \lambda$ from f, we can assume that $(\xi,\lambda) = 0 \in Q(t^*)$.

The sequence (t^k) is constructed in such a way that, for each k, the convex sets $Q(t^k)$ and $B_\varepsilon(0)$ are disjoint. By the hyperplane separation theorem, there exist (ξ^k,λ^k) such that

$$h(t^k,\xi^k,\lambda^k) = \sup_{(\xi,\lambda)\in Q(t^k)} \langle (\xi^k,\lambda^k),(\xi,\lambda)\rangle \leq -1 \leq \inf_{(\xi,\lambda)\in B_\varepsilon(0)} \langle (\xi^k,\lambda^k),(\xi,\lambda)\rangle. \tag{6.33}$$

The right hand side is

$$\inf_{(\xi,\lambda)\in B_\varepsilon(0)} \langle (\xi^k,\lambda^k),(\xi,\lambda)\rangle = -\varepsilon \sup_{(\xi,\lambda)\in B_1(0)} \langle (\xi^k,\lambda^k),(\xi,\lambda)\rangle = -\varepsilon \|(\xi^k,\lambda^k)\|, \tag{6.34}$$

so that $\|(\xi^k,\lambda^k)\| \leq 1/\varepsilon$ for each k. Hence, $\xi^k \to \xi^*$, $\lambda^k \to \lambda^*$ for a subsequence. By

Lemma 6.7, we know that h as in (6.12) is lower semicontinuous so that

$$h(t^*, \xi^*, \lambda^*) \leq \liminf_{k\to\infty} h(t^k, \xi^k, \lambda^k) \leq -1. \tag{6.35}$$

But, by $0 \in Q(t^*)$ we have $f(t^*, \cdot) \geq 0$ and therefore $h(t^*, \xi^*, \lambda^*) \geq 0$, a contradiction.

Now, suppose that Q is inner semicontinuous, fix $t^* \in U$, $\zeta^* \in \mathbb{R}^m$ and $M < f(t^*, \zeta^*)$. We want to show that there exists $\delta > 0$ such that, for each $t \in B_\delta(t^*)$ and $\zeta \in B_\delta(\zeta^*)$, we have

$$f(t, \zeta) \geq M. \tag{6.36}$$

Let $\varepsilon > 0$ be such that $M + 2\varepsilon < f(t^*, \zeta^*)$. The point $(\zeta^*, M + 2\varepsilon)$ has positive distance from the graph of $\zeta \mapsto f(t^*, \zeta)$ and hence, by convexity of f in the second variable, an affine linear function separates $\zeta \mapsto f(t^*, \zeta)$ from $M + 2\varepsilon$. This means that there exists $(\xi^*, \lambda^*) \in Q(t^*)$ such that $\langle \xi^*, \zeta^* \rangle + \lambda^* \geq M + 2\varepsilon$. By inner semicontinuity of Q, there exists $\tilde{\delta} > 0$, such that for each $t \in U$ with $\|t - t^*\| < \tilde{\delta}$, there exist $(\xi(t), \lambda(t)) \in \mathbb{R}^{m+1}$ with

$$\|\xi(t) - \xi^*\| + |\lambda(t) - \lambda^*| < \varepsilon / \max\{1, \|\zeta^*\|\} \tag{6.37}$$

and such that, for each $\zeta \in \mathbb{R}^m$,

$$f(t, \zeta) \geq \langle \xi(t), \zeta \rangle + \lambda(t). \tag{6.38}$$

Expanding the right hand side of (6.38) and applying the definition of λ^*, we get

$$\begin{aligned} f(t, \zeta) &\geq \langle \xi(t), \zeta \rangle + \lambda(t) && (6.39)\\ &= \langle \xi(t), \zeta - \zeta^* \rangle + \langle \xi(t) - \xi^*, \zeta^* \rangle + (\lambda(t) - \lambda^*) + \langle \xi^*, \zeta^* \rangle + \lambda^* && (6.40)\\ &\geq \langle \xi(t), \zeta - \zeta^* \rangle + \langle \xi(t) - \xi^*, \zeta^* \rangle + (\lambda(t) - \lambda^*) + M + 2\varepsilon. && (6.41) \end{aligned}$$

Using the Cauchy-Schwarz inequality, we further bound from below

$$f(t, \zeta) \geq M + 2\varepsilon + \langle \xi(t), \zeta - \zeta^* \rangle - \|\xi(t) - \xi^*\| \|\zeta^*\| - |\lambda(t) - \lambda^*|. \tag{6.42}$$

Applying (6.37) to this yields

$$f(t, \zeta) \geq M + \varepsilon + \langle \xi(t), \zeta - \zeta^* \rangle. \tag{6.43}$$

We again use (6.37) to bound $\|\xi(t)\|$ from below:

$$f(t,\zeta) \geq M + \varepsilon + \langle \xi(t), \zeta - \zeta^* \rangle \tag{6.44}$$
$$\geq M + \varepsilon - \|\xi(t)\| \|\zeta - \zeta^*\| \tag{6.45}$$
$$\geq M + \varepsilon - (\|\xi^*\| + \varepsilon)\|\zeta - \zeta^*\|. \tag{6.46}$$

Finally, we define $\delta := \min\{\tilde{\delta}, \varepsilon/(\|\xi^*\| + \varepsilon)\}$ so that $(\|\xi^*\| + \varepsilon)\|\zeta - \zeta^*\| \leq \varepsilon$ whenever $\zeta \in B_\delta(\zeta^*)$ and conclude

$$f(t,\zeta) \geq M + \varepsilon + \langle \xi(t), \zeta - \zeta^* \rangle \geq M \tag{6.47}$$

for every $t \in B_\delta(t^*)$ and $\zeta \in B_\delta(\zeta^*)$ which proofs lower semicontinuity of f at (t^*, ζ^*). □

Proposition 6.11 shows inner semicontinuity of Q under the assumptions given in Thm. 6.5 which implies that Thm. 6.5 is a direct consequence of Thm. 6.9.

6.2.4 Connection to the Unlifted Augmented Functional

To show the connection between the integral representation of $\mathcal{B}_f$ as in (6.13) and the original (augmented) functional (6.6), we will split vectorial measures in $\mathfrak{M}(U, \mathbb{R}^{m+1})$ into an $\mathbb{R}^m$-valued and a real-valued part (E, μ). For, brevity, we will then write $\mathcal{B}_f(E, \mu)$ instead of $\mathcal{B}_f((E, \mu))$. We first note that the integral representation established in Thm. 6.5 can be split up in the following way:

Proposition 6.12. *Suppose that the assumptions in Thm. 6.5 hold. Let $(E, \mu) \in \mathfrak{M}(U, \mathbb{R}^m \times \mathbb{R})$ and $E = v|\mu| + v^\perp|\mu|^\perp$ with $|\mu|^\perp \geq 0$ the part of E that is singular with respect to μ (the* Lebesgue decomposition *of E with respect to μ, see Sect. 2.1). Then*

$$\mathcal{B}_f(E,\mu) = \begin{cases} \int_U f(t, v(t))\,|\mu|(dt) + \int_U f^\infty(t, v^\perp(t))\,|\mu|^\perp(dt), & \text{if } \mu \geq 0, \\ +\infty, & \text{otherwise,} \end{cases} \tag{6.48}$$

where $f^\infty(t,\xi) := \lim_{s\to\infty} f(t, s\xi)/s$ is the recession function of $f(t,\cdot)$.

In particular, if μ is nonnegative and E is absolutely continuous with respect to μ, then we have, for $v \in L^1_\mu(U, \mathbb{R}^m)$ with $E = v \cdot \mu$,

$$\mathcal{B}_f(E,\mu) = \int_U f(t, v(t))\,\mu(dt). \tag{6.49}$$

Proof. Let $\theta := \|E\| + |\mu|$. By Thm. 6.5, we have

$$\mathcal{B}_f(E,\mu) = \int_U h(t,(dE/d\theta)(t),(d\mu/d\theta)(t))\,\theta(dt), \tag{6.50}$$

If there is a set of positive θ-measure on which μ is negative, then h assumes the value $+\infty$ on this set and hence $\mathcal{B}_f(E,\mu) = \infty$. Otherwise, θ can be split up into a singular and absolutely continuous part with respect to $|\mu|$ by the Radon-Nikodym theorem (see Sect. 2.1) and the assertion follows from the positive one-homogeneity of $h(t,\cdot,\cdot)$. □

Denote by δ_u the measure in $U = \Omega \times \Gamma$ that is concentrated on the graph of $u\colon \Omega \to \Gamma$ in the sense that, for each $\phi \in C_0(U)$,

$$\int_U \phi\, d\delta_u := \int_\Omega \phi(x,u(x))\,dx. \tag{6.51}$$

We establish conditions for equivalence of the original (augmented) functional (6.6) and the lifted functional:

Corollary 6.13. *Let $\Gamma \subset \mathbb{R}^s$ compact, $\Omega \subset \mathbb{R}^d$ open and $U = \Omega \times \Gamma$. Assume that $p\colon \Omega \to \mathbb{R}^m$ and $u\colon \Omega \to \mathbb{R}$ are measurable functions such that*

$$F(p,u) := \int_\Omega f(x,u(x),p(x))\,dx \tag{6.52}$$

is finite. Furthermore, define the vectorial measure $(E,\mu) \in \mathfrak{M}(U,\mathbb{R}^m \times \mathbb{R})$ by $\mu = \delta_u$ and $E = p \cdot \delta_u$ or, more precisely,

$$\int_U \langle \phi, dE\rangle := \int_\Omega \langle \phi(x,u(x)),p(x)\rangle\,dx, \tag{6.53}$$

whenever $\phi \in C_0(U,\mathbb{R}^m)$. Then

$$\mathcal{B}_f(E,\mu) = \int_\Omega f(x,u(x),p(x))\,dx = F(p,u) \tag{6.54}$$

with $\mathcal{B}_f(\nu) = \sup_{\phi\in\mathcal{K}_f}\langle \nu,\phi\rangle$ as in (6.7).

From Cor. 6.13, we infer that the problem

$$\begin{aligned} &\text{minimize } F(p,u)\\ &\text{subject to } p = Lu \end{aligned} \tag{6.55}$$

is equivalent to the lifted problem

$$\begin{aligned} &\text{minimize } \mathcal{B}_f(E,\mu) \\ &\text{subject to } E = p\mu,\ \mu = \delta_u,\ p = Lu. \end{aligned} \tag{6.56}$$

In both cases, E, μ, p and u are restricted to suitable spaces of measures and functions, according to the regularity imposed by L. In contrast to F, the functional $\mathcal{B}_f$ is convex, but the admissible set of measures (E,μ) in (6.56) is non-convex. As a convex relaxation of this set, we propose to consider measure-valued functions $\mu\colon \Omega \to \mathcal{P}(\Gamma)$ and $E\colon \Omega \to \mathfrak{M}(\Gamma,\mathbb{R}^m)$. Section 6.3 will be concerned with the lifting of $p = Lu$ to this relaxed setting. First, we prove an important property of $\mathcal{B}_f$ in the following subsection.

6.2.5 Absolute Continuity Enforced by the Lifted Functional

We defined the generalized Benamou-Brenier functional (6.7) for an *arbitrary* vectorial measure in $\mathfrak{M}(U,\mathbb{R}^{m+1})$, but we showed in Cor. 6.13 that it acts in a special meaningful way on split measures $(E,\mu) \in \mathfrak{M}(U,\mathbb{R}^m \times \mathbb{R})$ with $\mu = \delta_u$ and $E = p\cdot\delta_u$. In fact, we can show that, under stronger assumptions on the integrand f, the lifted functional forces E to be absolutely continuous with respect to μ whenever $\mathcal{B}_f(E,\mu) < \infty$.

Assumption 6.14. The integrand $f\colon U \times \mathbb{R}^m \to \overline{\mathbb{R}}$ is convex in the second argument and its convex conjugate f^* is continuous and satisfies

$$\sup_{t\in U, \|\xi\|<r} |f^*(t,\xi)| < \infty \qquad \forall r > 0. \tag{6.57}$$

Proposition 6.15. *Assume that U is locally compact and f satisfies Assumption 6.14. Let $(E,\mu) \in \mathfrak{M}(U,\mathbb{R}^m \times \mathbb{R})$ satisfy $\mathcal{B}_f(E,\mu) < \infty$. Then E is absolutely continuous with respect to μ.*

In order to prove Prop. 6.15, we show that the set $\mathcal{K}_f$ of continuous selections in the definition of $\mathcal{B}_f$ can be replaced by essentially bounded measurable selections:

Lemma 6.16. *Assume that U and f^* satisfy the assumptions from Prop. 6.15. Let $\nu \in \mathfrak{M}(U,\mathbb{R}^{m+1})$ and define $\theta := \|\nu\|$ as well as*

$$\mathcal{K}_f^\theta := \{(\phi^\xi,\phi^\lambda) \in L^\infty_\theta(U,\mathbb{R}^m\times\mathbb{R}) :\ \phi^\lambda(t) + f^*(t,\phi^\xi(t)) \le 0 \text{ for a.e. } t \in U\}. \tag{6.58}$$

Then,

$$\mathcal{B}_f(\nu) = \sup_{\phi\in\mathcal{K}_f^\theta} \int_U \langle \phi, d\nu\rangle. \tag{6.59}$$

Proof. Clearly, $\mathcal{K}_f \subset \mathcal{K}_f^\theta$, so that we only need to show that

$$\mathcal{B}_f(\nu) \geq \int_U \langle \phi, d\nu \rangle \tag{6.60}$$

hold for all $\phi \in \mathcal{K}_f^\theta$. Let $\phi = (\phi^\xi, \phi^\lambda) \in \mathcal{K}_f^\theta$ and $\varepsilon > 0$. We now construct $\tilde{\phi} = (\tilde{\phi}^\xi, \tilde{\phi}^\lambda) \in \mathcal{K}_f$ such that

$$\left| \int_U \langle \phi - \tilde{\phi}, d\theta \rangle \right| \leq \varepsilon. \tag{6.61}$$

Since ε was arbitrary and $\mathcal{B}_f$ is the supremum over $\mathcal{K}_f$, this will prove the assertion.

We define

$$M_f = \sup_{t \in U, \|\xi\| < \|\phi\|_\infty} |f^*(t, \xi)| \tag{6.62}$$

and $M := \max\{\|\phi\|_\infty, M_f\}$. By Lusin's theorem [HS65, Theorem 11.36], there is a closed set $A \subset U$ with $\theta(U \setminus A) \leq \varepsilon/(2M)$ and continuous functions $(\hat{\phi}^\xi, \hat{\phi}^\lambda) \in C_c(U, \mathbb{R}^m \times \mathbb{R})$ that agree with (ϕ^ξ, ϕ^λ) on A and satisfy $\|\hat{\phi}^\xi\|_\infty \leq \|\phi^\xi\|_\infty$ and $\|\hat{\phi}^\lambda\|_\infty \leq \|\phi^\lambda\|_\infty$. Now, define

$$\tilde{\phi}^\lambda(t) := \min\{\hat{\phi}^\lambda(t), -f^*(t, \hat{\phi}^\xi(t))\} \tag{6.63}$$

$$\tilde{\phi}^\xi(t) := \hat{\phi}^\xi(t). \tag{6.64}$$

Then, $\|\tilde{\phi}\|_\infty \leq M$ and the scalar component $\tilde{\phi}^\lambda$ is continuous with $|\phi^\lambda(t) - \tilde{\phi}^\lambda(t)|$ vanishing on A, so that $(\tilde{\phi}^\xi, \tilde{\phi}^\lambda) \in \mathcal{K}_f$. Hence,

$$\left| \int_U \langle \phi - \tilde{\phi}, d\theta \rangle \right| = \left| \int_{U \setminus A} \langle \phi - \tilde{\phi}, d\theta \rangle \right| \tag{6.65}$$

$$\leq \theta(U \setminus A) \|\phi - \tilde{\phi}\|_\infty \tag{6.66}$$

$$\leq \theta(U \setminus A)(\|\phi\|_\infty + \|\tilde{\phi}\|_\infty) \tag{6.67}$$

$$\leq 2M\theta(U \setminus A) \leq \varepsilon, \tag{6.68}$$

where we used $\theta(U \setminus A) \leq \varepsilon/(2M)$ in the last step. □

Proof of Prop. 6.15. Let (E, μ) as above and assume that there is a measurable set A with $\mu(A) = 0$ and $E(A) \neq 0$. For $k \in \mathbb{N}$, define $v := E(A) \in \mathbb{R}^m$ and $\psi_k^\xi(x) = kv\mathbb{1}_A(x)$ as well as $\psi_k^\lambda(x) = -f^*(x, \psi_k^\xi(x))\mathbb{1}_A(x)$. Then, the pair (ψ^ξ, ψ^λ) is in $\mathcal{K}_f^\theta$ from Lemma 6.16

and hence

$$\mathcal{B}_f(E,\mu) = \sup_{(\phi^\xi,\phi^\lambda)\in\mathcal{K}_f^\theta} \int_U \langle \phi^\xi, dE\rangle + \int_U \phi^\lambda \, d\mu \geq \int_U \langle \psi^\xi, dE\rangle + \int_U \psi^\lambda \, d\mu. \tag{6.69}$$

Inserting the definition of (ψ^ξ, ψ^λ) on the right hand side yields

$$\mathcal{B}_f(E,\mu) \geq \langle kv, E(A)\rangle + 0 = k\|v\|^2. \tag{6.70}$$

Since k was arbitrary, this is only possible for $\mathcal{B}_f(E,\mu) = \infty$. □

The boundedness assumption Assumption 6.14 on f^* cannot be omitted. As an example, consider the norm function $f(t,\xi) = \|\xi\|$ with $f^*(t,\zeta) = \delta_B(\zeta)$ where $B = \{\xi \in \mathbb{R}^m : \|\xi\| \leq 1\}$. In this case,

$$\mathcal{B}_f(E,\mu) = \sup_{\|\phi^\xi\|_\infty \leq 1} \int_U \langle \phi^\xi, dE\rangle = \|E\|(U), \tag{6.71}$$

which is finite, independent of μ. This means that some of the statements of this chapter cannot be applied to problems (6.1) with linear growth in Lu. However, it is still possible that many of the main results continue to hold for the case of linear growth. Note that a similar conceptual difference between energies with linear and superlinear growth is observed in the theory of dynamical optimal transport [Amb03].

6.3 Constraining the Augmented and Lifted Functional

At the heart of our proposed model is the idea to lift the coupling $p = Lu$ of the variables u and p in the augmented functional (6.6) to the space of measures by suitable continuity equations. In the following subsections we will motivate and formulate the continuity equations for the cases $L = \nabla$, $L = \nabla^2$ and $L = \Delta$. Throughout this section, we will apply the previous section to the case $U = \Omega \times \Gamma$, where $\Omega \subset \mathbb{R}^d$ is open and bounded and $\Gamma \subset \mathbb{R}^s$ is compact.

We denote by $C_c^k(\mathring{\Omega} \times \Gamma, \mathbb{R}^m)$ the space of k-times continuously differentiable functions with compact support in the first component, i.e., for $\phi \in C_c^k(\mathring{\Omega} \times \Gamma, \mathbb{R}^m)$ there is a compact set $V \subset \mathring{\Omega}$, such that $\phi(x,z) = 0$ whenever $x \in \Omega \setminus V$ and $z \in \Gamma$. Here, $\mathring{\Omega}$ is the interior of Ω. For the notion of differentiability on the compact set Γ it is assumed that there is an open neighborhood $V \supset \Gamma$ and a function $\psi \in C_c^k(\mathring{\Omega} \times V, \mathbb{R}^m)$ such that $\phi(x,z) = \psi(x,z)$ for $x \in \Omega$ and $z \in \Gamma$. This is the usual notion of differentiability in

those cases where Γ is a smooth submanifold of $\mathbb{R}^s$ with or without boundary [Hes41, Theorem 1].

As defined in (6.51) in the previous section, we denote by δ_u the measure in $U = \Omega \times \Gamma$ that is concentrated on the graph of $u\colon \Omega \to \Gamma$.

6.3.1 The First-Order Continuity Equation

In this section, we cover the case $L = \nabla$ and apply the previous section with $\mathbb{R}^m = \mathbb{R}^{d,s}$. Our idea is to lift the constraint $p = \nabla u$ to a weak differential equation in (E, μ) that is satisfied if (E, μ) is concentrated on the graph of u with E taking the values of the Jacobian ∇u. An equation with this property is the first-order continuity equation

$$\nabla_x \mu + \operatorname{div}_z E = 0 \tag{6.72}$$

in the following weak sense:

Definition 6.17. We say that the pair $(E, \mu) \in \mathfrak{M}(\Omega \times \Gamma, \mathbb{R}^{d,s} \times \mathbb{R})$ *solves the first-order continuity equation* if, for every $\phi \in C_c^1(\mathring{\Omega} \times \Gamma, \mathbb{R}^d)$, we have

$$\int_{\Omega \times \Gamma} \operatorname{div}_x \phi \, d\mu + \int_{\Omega \times \Gamma} \langle \nabla_z \phi, dE \rangle = 0. \tag{6.73}$$

We now show that, for differentiable u, this agrees with $p = \nabla u$ if (E, μ) is concentrated on the graph of u with E taking the values of the Jacobian ∇u:

Proposition 6.18. *Let $u \in C^1(\Omega, \Gamma)$ and define $(E, \mu) \in \mathfrak{M}(\Omega \times \Gamma, \mathbb{R}^{d,s} \times \mathbb{R})$ by $\mu := \delta_u$ and $E = \nabla u \cdot \mu$, i.e.,*

$$\int_{\Omega \times \Gamma} \langle \phi, dE \rangle := \int_\Omega \langle \phi(x, u(x)), \nabla u(x) \rangle \, dx, \tag{6.74}$$

whenever $\phi \in C_0(\Omega \times \Gamma, \mathbb{R}^{d,s})$. Then (E, μ) solves the continuity equation.

Proof. Let $\phi \in C_c^1(\mathring{\Omega} \times \Gamma, \mathbb{R}^d)$ be arbitrary. By chain rule and by the regularity assumption on u, we have

$$\operatorname{div}_x [\phi(x, u(x))] = \operatorname{div}_x \phi(x, u(x)) + \langle \nabla_z \phi(x, u(x)), \nabla u(x) \rangle. \tag{6.75}$$

Then

$$\int_{\Omega\times\Gamma} \operatorname{div}_x \phi \, d\mu = \int_\Omega \operatorname{div}_x \phi(x, u(x)) \, dx \tag{6.76}$$

$$= \int_\Omega \operatorname{div}_x \left[\phi(x, u(x))\right] \, dx - \int_\Omega \langle \nabla_z \phi(x, u(x)), \nabla u(x) \rangle \, dx. \tag{6.77}$$

The first term vanishes by the compact support of ϕ and, by the definition of E, the second term equals $\int_{\Omega\times\Gamma} \langle \nabla_z \phi, dE \rangle$, which implies (6.73). □

Given a function u and letting $\mu := \delta_u$, the solution to the continuity equation is unique in the following sense:

Proposition 6.19. *Let $u \in C^1(\Omega, \Gamma)$, $\mu := \delta_u \in \mathfrak{M}(\Omega \times \Gamma)$, and suppose that $E \in \mathfrak{M}(\Omega \times \Gamma, \mathbb{R}^{d,s})$ is absolutely continuous with respect to μ. If (E, μ) solves the first-order continuity equation (Defn. 6.17), then $E = \nabla u \cdot \mu$ as given in Prop. 6.18 holds μ-almost everywhere on the support of μ.*

Proof. Let E be given as in Prop. 6.18 and let $\tilde{E} \in \mathfrak{M}(\Omega \times \Gamma, \mathbb{R}^{d,s})$ be a vector measure that is absolutely continuous with respect to μ and such that $(\tilde{E}, \mu)$ solves the continuity equation, say $\tilde{E} = w \cdot \mu$. Since both (E, μ) and $(\tilde{E}, \mu)$ solve the continuity equation, we have

$$\int_{\Omega\times\Gamma} \langle \nabla_z \phi, w - \nabla u \rangle \, d\mu = 0, \tag{6.78}$$

for every $\phi \in C_c^1(\mathring{\Omega} \times \Gamma, \mathbb{R}^d)$. Now, let $\psi \in C_c(\Omega, \mathbb{R}^{d,s})$. Then $\phi(x, z) := \psi(x)z$ is in $C_c^1(\mathring{\Omega} \times \Gamma, \mathbb{R}^d)$ with $\nabla_z \phi = \psi$. Therefore, (6.78) becomes

$$\int_\Omega \langle \psi(x), w(x, u(x)) - \nabla u(x) \rangle \, dx = 0, \tag{6.79}$$

by the definition of μ. Since ψ was arbitrary, we conclude that $w = \nabla u$ almost everywhere on the support of μ. □

Remark 6.20. This uniqueness result fails to hold if μ is not concentrated on a graph or without the assumption of absolute continuity of E with respect to μ, since the continuity equation is invariant in E under the addition of divergence-free vector fields on Γ: Let $v \in C^1(\Gamma, \mathbb{R}^{d,s})$ be a divergence-free vector field, e.g., $v(z_1, z_2) = (z_2, z_1)$ for $d = 1$, $\Gamma = [0, 1]^2$. Whenever (E, μ) solves the first-order continuity equation, another solution is given by $(\tilde{E}, \mu)$ where $\tilde{E}$ is defined by

$$\int_{\Omega\times\Gamma} \langle \phi, d\tilde{E} \rangle = \int_{\Omega\times\Gamma} \langle \phi, dE \rangle + \int_{\Omega\times\Gamma} \langle \phi(x, z), v(z) \rangle \, dz \, dx \tag{6.80}$$

for every $\phi \in C_0(\Omega \times \Gamma, \mathbb{R}^{d,s})$. To see this, note that, for every $\phi \in C_c^1(\mathring{\Omega} \times \Gamma, \mathbb{R}^d)$, we have

$$\int_\Omega \int_\Gamma \langle \nabla_z \phi(x,z), v(z) \rangle \, dz \, dx = \int_\Omega \int_\Gamma \langle \phi(x,z), -\operatorname{div}_z v(z) \rangle \, dz \, dx = 0, \tag{6.81}$$

and therefore $\int_\Omega \int_\Gamma \langle \nabla_z \phi, dE \rangle = \int_\Omega \int_\Gamma \langle \nabla_z \phi, d\tilde{E} \rangle$. Hence, both (E, μ) and $(\tilde{E}, \mu)$ solve the first-order continuity equation (Defn. 6.17). However, if μ is concentrated on a graph, as is the case in Prop. 6.19, not both E and $\tilde{E}$ are absolutely continuous with respect to μ, since $E - \tilde{E}$ is not concentrated on a graph.

Finally, for each $\mu \in \mathfrak{M}(\Omega \times \Gamma)$, we define the lifted version of our original functional (6.1) (without the augmentation introduced in (6.6)) and with $L = \nabla$ as

$$\mathcal{F}(\mu) := \inf \{ \mathcal{B}_f(E, \mu) : (E, \mu) \text{ solves (6.73)} \} \tag{6.82}$$

with suitable choice of spaces for E and μ, see Defn. 6.17. In practice, we are interested in minimizers of $\mathcal{F}$ which amounts to minimizing the convex and lower semicontinuous functional $\mathcal{B}_f$ subject to the linear constraints (6.73). That's why we omit a study of lower semicontinuity and convexity of $\mathcal{F}$. Furthermore, we leave a study of the set of vector measures E that satisfy the first-order continuity equation (6.73) for given μ for future work. The central result for this subsection is the following connection between the lifted functional $\mathcal{F}$ and the original functional F under the stronger assumptions on the integrand established in Prop. 6.15:

Proposition 6.21. *Let Ω open, Γ compact and suppose that f satisfies Assumption 6.14. If $u \in C^1(\Omega, \Gamma)$ satisfies $F(u) < \infty$ for F as in (6.1) with $L = \nabla$, then*

$$\mathcal{F}(\delta_u) = F(u) := \int_\Omega f(x, u(x), \nabla u(x)) \, dx. \tag{6.83}$$

Proof. By Prop. 6.18, a solution to the continuity equation is given by $\mu := \delta_u$ and $E = \nabla u \cdot \mu$, hence $\mathcal{F}(\delta_u) \leq F(u)$. On the other hand, whenever $\mathcal{B}_f(E, \mu)$ is finite in the infimum (6.82), the measure E is automatically absolutely continuous with respect to μ by Prop. 6.15. This implies that $E = \nabla u \cdot \mu$ is, in fact, the *unique* solution to the continuity equation by Prop. 6.19 so that $\mathcal{F}(\delta_u) = F(u)$. □

We conclude this subsection with a formal derivation of a dual formulation for $\mathcal{F}$ that will help in understanding the relationship to similar functionals in the literature on measure-valued models (see Sect. 6.4). By introducing a Lagrange multiplier $p \in$

$C_c^1(\mathring{\Omega} \times \Gamma, \mathbb{R}^d)$ for the first-order continuity equation and by the dual formulation (6.7) of $\mathcal{B}_f$, we deduce

$$\mathcal{F}(\mu) = \inf_E \sup_{\phi \in \mathcal{K}_f, p} \left[\int_{\Omega \times \Gamma} (\operatorname{div}_x p + \phi^\lambda)\, d\mu + \int_{\Omega \times \Gamma} \langle \nabla_z p + \phi^\xi, dE \rangle \right], \tag{6.84}$$

where $E \in \mathfrak{M}(\Omega \times \Gamma, \mathbb{R}^{d,s})$ and $p \in C_c^1(\mathring{\Omega} \times \Gamma, \mathbb{R}^d)$. Formally swapping infimum and supremum yields $\phi^\xi = -\nabla_z p$ almost everywhere and, renaming $q := \phi^\lambda$, we obtain the dual formulation:

$$\begin{aligned} \mathcal{F}_D(\mu) &= \sup_{\phi \in \mathcal{K}_f, p} \left[\int_{\Omega \times \Gamma} (\operatorname{div}_x p + \phi^\lambda)\, d\mu + \inf_E \int_{\Omega \times \Gamma} \langle \nabla_z p + \phi^\xi, dE \rangle \right] \\ &= \sup \left\{ \int_{\Omega \times \Gamma} (\operatorname{div}_x p + q)\, d\mu : p \in C_c^1(\mathring{\Omega} \times \Gamma, \mathbb{R}^d) \text{ and } (-\nabla_z p, q) \in \mathcal{K}_f \right\}. \end{aligned} \tag{6.85}$$

The equation $\mathcal{F}_D(\mu) = \mathcal{F}(\mu)$ holds only in a formal sense and is hard to justify in general. The dual formulation establishes a formal connection to the total variation regularization of measure-valued images introduced in Chap. 3, see Sect. 6.4.5) below. The relationship of our first-order model with the calibration method-based and currents-based liftings from the literature will be discussed in Sects. 6.4.1 and 6.4.2.

6.3.2 The Second-Order Continuity Equation

This section covers the case $L = \nabla^2$. For a function $u \colon \Omega \to \mathbb{R}^s$, the Hessian $\nabla^2 u(x)$ is a third-order tensor in $\mathbb{R}^{d,d,s}$, which means that, for each $k = 1, \dots, s$, the component-wise Hessian $\nabla^2 u_k(x)$ is a $d \times d$ matrix. Therefore, we apply the lifting introduced in the previous section to the case $\mathbb{R}^m = \mathbb{R}^{d,d,s}$ by considering an equation that is satisfied if (E, μ) is concentrated on the graph of u with E taking the values of the Hessian $\nabla^2 u$. An equation with this property is the second-order continuity equation

$$-\nabla_x^2 u - \operatorname{div}_z E + \operatorname{div}_z^2 H = 0, \tag{6.86}$$

for an additional auxiliary variable H, a measure whose values are fourth-order tensors in $\mathbb{R}^{d,d,s,s}$. The equation is to be understood in the following weak sense:

Definition 6.22. We say that the triple $(E, \mu, H) \in \mathfrak{M}(\Omega \times \Gamma, \mathbb{R}^{d,d,s} \times \mathbb{R} \times \mathbb{R}^{d,d,s,s})$ *solves the second-order continuity equation* if, for every $\phi \in C_c^2(\mathring{\Omega} \times \Gamma, \mathbb{R}^{d,d})$, we have

$$-\int_{\Omega \times \Gamma} \operatorname{div}_x^2 \phi\, d\mu + \int_{\Omega \times \Gamma} \langle \nabla_z \phi, dE \rangle + \int_{\Omega \times \Gamma} \langle \nabla_z^2 \phi, dH \rangle = 0. \tag{6.87}$$

Analogously to the previous subsection, this agrees with $p = \nabla^2 u$ if (E, μ, H) is concentrated on the graph of u with E taking as values the Hessian $\nabla^2 u$ and H playing the role of the tensor product $\nabla u \otimes \nabla u$ as follows:

Proposition 6.23. *Let $u \in C^2(\Omega, \Gamma)$ and define $\mu \in \mathfrak{M}(\Omega \times \Gamma)$, $E \in \mathfrak{M}(\Omega \times \Gamma, \mathbb{R}^{d,d,s})$ and $H \in \mathfrak{M}(\Omega \times \Gamma, \mathbb{R}^{d,d,s,s})$ by*

$$\mu := \delta_u, \quad E = \nabla^2 u \cdot \mu, \quad H = (\nabla u \otimes \nabla u) \cdot \mu, \tag{6.88}$$

where the latter is to be understood in the sense that

$$\int_{\Omega \times \Gamma} \langle \Phi, dH \rangle := \sum_{i,j=1}^{d} \int_{\Omega} \langle \Phi_{ij}(x, u(x)) \partial_i u(x), \partial_j u(x) \rangle \, dx, \tag{6.89}$$

whenever $\Phi \in C_0(\Omega \times \Gamma, \mathbb{R}^{d,d,s,s})$. Then (E, μ, H) solves the second-order continuity equation.

Proof. By the regularity assumption on u, the chain rule applies:

$$\begin{aligned}
\operatorname{div}_x^2 [\phi(x, u(x))] &= \operatorname{div}_x \left[\operatorname{div}_x \phi(x, u(x)) + \sum_{k=1}^{d} \langle \partial_{z_k} \phi(x, u(x)), \nabla u_k(x) \rangle \right] && (6.90)\\
&= \operatorname{div}_x^2 \phi(x, u(x)) + \langle \nabla_z \phi(x, u(x)), \nabla^2 u(x) \rangle && (6.91)\\
&\quad + \langle \nabla_z \operatorname{div}_x (\phi + \phi^{\mathsf{T}})(x, u(x)), \nabla u(x) \rangle && (6.92)\\
&\quad + \sum_{i,j=1}^{d} \langle \nabla_z^2 \phi_{ij}(x, u(x)) \partial_i u(x), \partial_j u(x) \rangle && (6.93)\\
&= \Delta_x \phi(x, u(x)) + \langle \nabla_z \phi(x, u(x)), \nabla^2 u(x) \rangle && (6.94)\\
&\quad + \langle \operatorname{div}_x \left[\nabla_z (\phi + \phi^{\mathsf{T}})(x, u(x)) \right], \nabla u(x) \rangle && (6.95)\\
&\quad - \sum_{i,j=1}^{d} \langle \nabla_z^2 \phi_{ij}(x, u(x)) \partial_i u(x), \partial_j u(x) \rangle, && (6.96)
\end{aligned}$$

for every $\phi \in C^2(\Omega \times \Gamma, \mathbb{R}^{d,d})$. Now consider (6.87) and let $\phi \in C_c^2(\mathring{\Omega} \times \Gamma, \mathbb{R}^{d,d})$. Then, after applying Gauss' theorem once,

$$\begin{aligned}
\int_{\Omega} \operatorname{div}_x^2 \phi(x, u(x)) \, dx &= \int_{\Omega} \operatorname{div}_x^2 [\phi(x, u(x))] \, dx + \int_{\Omega} \langle \nabla_z \phi(x, u(x)), \nabla^2 u(x) \rangle \, dx && (6.97)\\
&\quad + \sum_{i,j=1}^{d} \int_{\Omega} \langle \nabla_z^2 \phi_{ij}(x, u(x)) \partial_i u(x), \partial_j u(x) \rangle \, dx. && (6.98)
\end{aligned}$$

The first term vanishes by the compact support of ϕ and, by the definitions of E and H, the second and third terms equal $\int_{\Omega\times\Gamma}\langle\nabla_z\phi(x,z), dE\rangle$ and $\int_{\Omega\times\Gamma}\langle\nabla_z^2\phi(x,z), dH\rangle$. □

Uniqueness of solutions along the graph of a C^2 function holds in the following sense and a remark analogous to Rem. 6.20 applies:

Proposition 6.24. *Let $u \in C^2(\Omega,\Gamma)$ and $\mu = \delta_u \in \mathfrak{M}(\Omega\times\Gamma)$ and suppose that the vectorial measure $(E,H) \in \mathfrak{M}(\Omega\times\Gamma, \mathbb{R}^{d,d,s}\times\mathbb{R}^{d,d,s,s})$ is such that (E,μ,H) solves the second-order continuity equation.*

1. *If E is absolutely continuous with respect to μ, then it is as given in Prop. 6.23 almost everywhere on the support of μ.*

2. *If, in addition to the previous assumptions, H is symmetric in the last two components and absolutely continuous with respect to μ, then H is as given in Prop. 6.23 almost everywhere on the support of μ.*

Proof. Let E and H be given as in Prop. 6.23 and let $(\tilde{E},\tilde{H}) \in \mathfrak{M}(\Omega\times\Gamma, \mathbb{R}^{d,d,s}\times\mathbb{R}^{d,d,s,s})$ be a vector measure such that $(\tilde{E},\mu,\tilde{H})$ solves the second-order continuity equation. Suppose that $\tilde{E} = w\cdot\mu$. Then, by the weak formulation of the second-order continuity equation,

$$\int_{\Omega\times\Gamma}\langle\nabla_z\phi, w-\nabla^2 u\rangle\, d\mu + \int_{\Omega\times\Gamma}\langle\nabla_z^2\phi, d(\tilde{H}-H)\rangle = 0, \tag{6.99}$$

for each $\phi \in C_c^2(\mathring{\Omega}\times\Gamma, \mathbb{R}^{d,d})$. Let $\psi \in C_c^2(\Omega, \mathbb{R}^{d,d,s})$. Then $\phi_{ij}(x,z) := (\langle\psi_{ij}(x), z\rangle)_{i,j}$ is in $C_c^2(\mathring{\Omega}\times\Gamma, \mathbb{R}^{d,d})$ with $\nabla_z\phi = \psi$ and $\nabla_z^2\phi = 0$, so that

$$\int_\Omega \langle\psi(x), w(x,u(x)) - \nabla^2 u(x)\rangle\, dx = 0, \tag{6.100}$$

by the definition of μ. Since ψ was arbitrary, we infer $w = \nabla^2 u$ almost everywhere on the support of μ. This shows the first claim.

For the second claim, assume $\tilde{H} = A\cdot\mu$ and define $\phi_{ij}(x,z) := \frac{1}{2}(\langle z, \psi_{ij}(x)z\rangle)_{i,j}$ for some $\psi \in C_c^2(\Omega, \mathbb{R}^{d,d,s,s})$ with $\psi_{ij}^{\mathsf{T}} = \psi_{ij}$. Then $\phi \in C_c^2(\mathring{\Omega}\times\Gamma, \mathbb{R}^{d,d})$ and

$$\int_{\Omega\times\Gamma}\langle\psi, d(\tilde{H}-H)\rangle = \int_\Omega \langle\psi(x,u(x)), A(x,u(x)) - (\nabla u(x)\otimes\nabla u(x))\rangle\, dx = 0. \tag{6.101}$$

Since ψ was arbitrary (but symmetric) and A is symmetric by assumption, we get $A = (\nabla u\otimes\nabla u)$ almost everywhere on the support of μ. □

Finally, for each $\mu \in \mathfrak{M}(\Omega \times \Gamma)$, we define the lifted version of our original functional (6.1) without augmentation and with $L = \nabla^2$ as

$$\mathcal{F}(\mu) := \inf\{\mathcal{B}_f(E, \mu) : (E, \mu, H) \text{ solves (6.87) for } H \text{ sym.}\}. \tag{6.102}$$

with suitable choice of spaces for E, μ and H, see Defn. 6.22. Note that H remains an auxiliary variable that is not included in the evaluation of $\mathcal{B}_f$. In practice, we are interested in minimizers of $\mathcal{F}$ which amounts to minimizing the convex and lower semi-continuous functional $\mathcal{B}_f$ subject to the linear constraints (6.87). That's why we omit a study of lower semicontinuity and convexity of $\mathcal{F}$. Furthermore, we leave a study of the set of vector measures E and H tha satisfy the first-order continuity equation (6.87) for a given measure $\mu \in \mathfrak{M}(\Omega \times \Gamma)$ for future work. As in the first-order case (Prop. 6.21), the central result for this subsection is the connection between the lifted functional $\mathcal{F}$ and the original functional F, which is a consequence o Prop. 6.23 and Prop. 6.15:

Proposition 6.25. *Let $u \in C^2(\Omega, \Gamma)$ with $F(u) < \infty$ and suppose that f satisfies Assumption 6.14. Then*

$$\mathcal{F}(\delta_u) = F(u) := \int_\Omega f(x, u(x), \nabla^2 u(x))\, dx. \tag{6.103}$$

Stating the dual formulation of $\mathcal{F}$ is more involved than in the first-order case as the auxiliary variable H has hidden structure that translates into constraints on the dual variables. We restrict ourselves to the case where f depends on the Laplacian because there it will help in establishing the connection to the second-order lifting framework proposed in Chap. 5.

6.3.3 The Laplacian Continuity Equation

As a special case of the previous subsection, we explicitly discuss the case where $L = \Delta$ or, equivalently, where the integrand f only depends on the trace of each $\nabla^2 u_k$ for $k = 1, \dots, s$. In this setting, it suffices to consider the simplified second-order continuity equation

$$-\Delta_x \mu - \operatorname{div}_z E + \operatorname{div}_z^2 H = 0. \tag{6.104}$$

The difference to (6.104) is in the differential operator Δ_x (instead of ∇^2) in the first term and in the order of the tensor-valued measures E and H. The equation is to be understood in the following weak sense:

Definition 6.26. We say that the triple $(E, \mu, H) \in \mathfrak{M}(\Omega \times \Gamma, \mathbb{R}^s \times \mathbb{R} \times \mathbb{R}^{s,s})$ *solves the*

Laplacian continuity equation if, for every $\phi \in C_c^2(\mathring{\Omega} \times \Gamma)$, we have

$$-\int_{\Omega\times\Gamma} \Delta_x \phi \, d\mu + \int_{\Omega\times\Gamma} \langle \nabla_z \phi, dE\rangle + \int_{\Omega\times\Gamma} \langle \nabla_z^2 \phi, dH\rangle = 0. \tag{6.105}$$

All of the following statements are proved analogously to the corresponding statements in the previous subsection for $L = \nabla^2$:

Proposition 6.27. *Let $u \in C^2(\Omega, \Gamma)$ and define $\mu \in \mathfrak{M}(\Omega \times \Gamma)$, $E \in \mathfrak{M}(\Omega \times \Gamma, \mathbb{R}^s)$ and $H \in \mathfrak{M}(\Omega \times \Gamma, \mathbb{R}^{s,s})$ by $\mu := \delta_u$, $E = \Delta u \cdot \mu$ and $H = \sum_i (\partial_i u \otimes \partial_i u) \cdot \mu$, i.e.,*

$$\int_{\Omega\times\Gamma} \langle \Phi, dH\rangle := \sum_{i=1}^{d} \int_\Omega \langle \Phi(x, u(x))\partial_i u(x), \partial_i u(x)\rangle \, dx, \tag{6.106}$$

whenever $\Phi \in C_0(\Omega \times \Gamma, \mathbb{R}^{s,s})$. Then (E, μ, H) solves the Laplacian continuity equation.

Proposition 6.28. *Let $u \in C^2(\Omega, \Gamma)$ and $\mu = \delta_u \in \mathfrak{M}(\Omega \times \Gamma)$ and suppose that the vectorial measure $(E, H) \in \mathfrak{M}(\Omega\times\Gamma, \mathbb{R}^s \times \mathbb{R}^{s,s})$ is such that (E, μ, H) solves the Laplacian continuity equation.*

1. *If E is absolutely continuous with respect to μ, then it is as given in Prop. 6.27 almost everywhere on the support of μ.*

2. *If, in addition to the previous assumptions, H is symmetric and absolutely continuous with respect to μ, then H is as given in Prop. 6.27 almost everywhere on the support of μ.*

Finally, for each $\mu \in \mathfrak{M}(\Omega \times \Gamma)$, we define the lifted version of our original functional (6.1) without augmentation and with $L = \Delta$ as

$$\mathcal{F}(\mu) := \inf\{\mathcal{B}_f(E, \mu) : (E, \mu, H) \text{ solves (6.105) for } H \text{ sym. pos. semidef.}\}. \tag{6.107}$$

Note that we added the assumption of positive semidefiniteness on H because this is satisfied if $\mu = \delta_u$. As in the previous cases, we establish the following connection with the original functional (6.1):

Proposition 6.29. *Let $u \in C^2(\Omega, \Gamma)$ with $F(u) < \infty$ and suppose that f satisfies Assumption 6.14. Then*

$$\mathcal{F}(\delta_u) = F(u) := \int_\Omega f(x, u(x), \Delta u(x)) \, dx. \tag{6.108}$$

We conclude this subsection with a formal derivation of a dual formulation for $\mathcal{F}$ that will help in understanding the relationship to the literature on measure-valued models. By the method of Lagrange multipliers applied to the weak Laplacian continuity equation and by the dual formulation (6.7) of $\mathcal{B}_f$, we have

$$\mathcal{F}(\mu) = \inf_{E,H} \sup_{\phi \in \mathcal{K}_f, p} \left[\int_{\Omega\times\Gamma} (-\Delta_x p + \phi^\lambda)\, d\mu + \int_{\Omega\times\Gamma} \langle \nabla_z p + \phi^\xi, dE\rangle + \int_{\Omega\times\Gamma} \langle \nabla_z^2 p, dH\rangle \right], \tag{6.109}$$

where $(E,H) \in \mathfrak{M}(\Omega\times\Gamma, \mathbb{R}^s \times \mathbb{R}^{s,s})$ with H symmetric and positive semidefinite and $p \in C_c^2(\mathring{\Omega}\times\Gamma, \mathbb{R}^d)$. Formally swapping infimum and supremum enforces $\phi^\xi = -\nabla_z p$ as well as positive semidefiniteness of $\nabla_z^2 p$ almost everywhere and, setting $q := \phi^\lambda$, we obtain the dual formulation:

$$\begin{aligned}\mathcal{F}_D(\mu) = \sup \Bigg\{ \int_{\Omega\times\Gamma} (-\Delta_x p + q)\, d\mu : p \in C_c^2(\mathring{\Omega}\times\Gamma, \mathbb{R}^d),\ \nabla_z^2 p \text{ pos. semidef.} \\ \text{and } (-\nabla_z p, q) \in \mathcal{K}_f \Bigg\}.\end{aligned} \tag{6.110}$$

Again, as in the first-order case, the equation $\mathcal{F}_D(\mu) = \mathcal{F}(\mu)$ holds only in a formal sense and is hard to justify in general in our non-compact infinite-dimensional setting.

6.4 Connection to Previous Functional Lifting Models

From Props. 6.21, 6.25 and 6.29, we deduce that problem (6.1) (depending on L) is equivalent to the lifted problem

$$\begin{aligned}&\text{minimize } \mathcal{F}(\mu) \\ &\text{subject to } \mu = \delta_u,\end{aligned} \tag{6.111}$$

where u is assumed to be as smooth as is required by L. In contrast to the minimization of F, this is a convex problem, but the admissible set of measures $\mu = \delta_u$ in (6.111) is non-convex. As a convex relaxation of this set, we propose to consider weakly measurable functions $\mu \in L_w^\infty(\Omega, \mathcal{P}(\Gamma))$ in the sense of Defn. 6.1.

In the following subsections, we will demonstrate that our lifting approach is a generalization of a large class of functional lifting approaches from the literature.

6.4.1 Calibration Method-Based Functional Lifting

The functional lifting approach presented in [Poc+10] is based on the so-called *calibration method* It only applies to the first-order (i.e., $L = \nabla$ in (6.1)) scalar-valued case $\Gamma = \mathbb{R}$ and comes in a dual formulation of the form

$$\mathcal{F}_{\text{calib}}(v) = \sup_{\phi \in \mathcal{K}_{\text{calib}}} \int_{\Omega \times \mathbb{R}} \langle \phi, Dv \rangle \tag{6.112}$$

for each $v \in \mathrm{BV}(\Omega \times \mathbb{R})$, $\Omega \subset \mathbb{R}^d$ and

$$\mathcal{K}_{\text{calib}} := \{\phi = (\phi^\xi, \phi^\lambda) \in C_0(\Omega \times \mathbb{R}, \mathbb{R}^{d+1}) : \tag{6.113}$$

$$f^*(x, z, \phi^\xi(x, z)) \leq \phi^\lambda(x, z) \ \forall (x, z) \in \Omega \times \mathbb{R}\}. \tag{6.114}$$

After setting $U := \Omega \times \mathbb{R}$, the set $\mathcal{K}_{\text{calib}}$ agrees with our definition of $\mathcal{K}_f$ as given in (6.8) up to a change of sign in ϕ^λ.

As noted in Sect. 6.1, the central concept of the calibration-based lifting is to minimize $\mathcal{F}_{\text{calib}}$ over the set

$$\{\mathbb{1}_u : u \in W^{1,1}(\Omega)\}, \qquad \mathbb{1}_u(x, z) := \begin{cases} 1 & \text{if } u(x) > z, \\ 0 & \text{otherwise.} \end{cases} \tag{6.115}$$

The resulting constraint minimization problem is equivalent to minimizing the original functional (6.1) over the space $W^{1,1}(\Omega)$. The lifted problem is then defined on the convexified admissible set

$$\left\{ v \in \mathrm{BV}(\Omega \times \mathbb{R}, [0, 1]) : \lim_{z \to \infty} v(x, z) = 0, \lim_{z \to -\infty} v(x, z) = 1 \right\}. \tag{6.116}$$

We observe that the functional $\mathcal{F}_{\text{calib}}$ is actually not a functional in v, but in Dv. As v is a function of bounded variation on $\Omega \times \mathbb{R}$, Dv is a vector-valued measure on $U := \Omega \times \mathbb{R}$ and we can easily state the relationship with our functional (6.7) as

$$\mathcal{F}_{\text{calib}}(v) = \mathcal{B}_f(D^- v), \tag{6.117}$$

where $D^- v$ is Dv with the last component multiplied by (-1) to address the change of sign in $\mathcal{K}_{\text{calib}}$ compared to $\mathcal{K}_f$. Furthermore, we obtain the following central result from [Poc+10] as a corollary of our Thm. 6.5.

Corollary 6.30 ([Poc+10, Theorem 3.2])**.** *Suppose that $\Omega \subset \mathbb{R}^d$ is open and f satisfies*

the assumptions in Thm. 6.5. Then, for each $u \in W^{1,1}(\Omega)$*,*

$$\int_\Omega f(x, u(x), \nabla u(x))\,dx = \sup_{\phi \in \mathcal{K}_{calib}} \int_{\Omega\times\mathbb{R}} \langle \phi, D\mathbb{1}_u \rangle. \tag{6.118}$$

Proof. We set $v := \mathbb{1}_u$. Then $D^- v \in \mathfrak{M}(\Omega \times \mathbb{R}, \mathbb{R}^{d+1})$. Moreover, the set $U := \Omega \times \mathbb{R}$ is locally compact by [Wil70, Theorem 18.6] so that the assumptions in Thm. 6.5 are satisfied. As discussed above, the right-hand side in (6.118) is $\mathcal{B}_f(D^- v)$ so that we infer from Thm. 6.5

$$\sup_{\phi \in \mathcal{K}_{\text{calib}}} \int_{\Omega\times\mathbb{R}} \langle \phi, Dv \rangle = \int_{\Omega\times\mathbb{R}} h(x, z, (dD^- v/d\|D^- v\|)(x, z))\,\|D^- v\|(dx, dz). \tag{6.119}$$

On the other hand, since $u \in W^{1,1}(\Omega)$, we know that Dv is concentrated on the graph of u with vectorial density $(\nabla u, -1)$. From the definition of h, we get

$$h(x, z, (dD^- v/d\|D^- v\|)(x, z)) = h(x, z, (\nabla u(x), 1)) = f(x, z, \nabla u(x)). \tag{6.120}$$

Integrating this over the graph $(x, z) = (x, u(x))$ of u we conclude

$$\sup_{\phi \in \mathcal{K}_{\text{calib}}} \int_{\Omega\times\mathbb{R}} \langle \phi, Dv \rangle = \int_{\Omega\times\mathbb{R}} h(x, z, (\nabla u(x), 1))\,\|D^- v\|(dx, dz) \tag{6.121}$$

$$= \int_\Omega f(x, u(x), \nabla u(x))\,dx, \tag{6.122}$$

which is the claimed statement (6.118). □

While the lifted functional in the calibration-based lifting is equivalent to the functional in our proposed lifting (as pointed out above), the convexified admissible set (6.4) as used in the calibration-based lifting is slightly smaller than the set $L^\infty_w(\Omega, \mathcal{P}(\mathbb{R}))$ that we use for our lifting. For every function v in (6.116), its derivative measure Dv is in $L^\infty_w(\Omega, \mathcal{P}(\mathbb{R}))$ by the established slicing theory for functions of bounded variation [AFP00, Lemma 3.106]. However, for the converse inclusion (i.e., for each $\mu \in L^\infty_w(\Omega, \mathcal{P}(\mathbb{R}))$ there exists a function v in (6.4) such that $Dv = \mu$), more regularity in Ω is needed instead of just weak measurability. Intuitively, additional regularity of μ in Ω is imposed by the continuity equation constraint $\nabla_x \mu + \operatorname{div}_z E = 0$. We leave a more extensive discussion for future work.

Concluding, the calibration method-based framework and our framework are mathematically equivalent in some respects and very similar or supposedly equivalent in others. In practice, with the discretizations from [Poc+10] and [MC17], we found the calibration

method-based model to be a reformulation of our proposed lifting framework. However, our proposed framework has the advantage that it generalizes to vectorial and higher-order problems while the calibration method-based framework uses a notation that is restricted to scalar and first-order problems.

6.4.2 Currents-Based Functional Lifting

The lifting strategy [WC16; MC19] for vectorial first-order problems (i.e., $L = \nabla$ in (6.1)) with so-called *polyconvex* integrands is based on the theory of currents [GMS98a; GMS98b]. We will demonstrate that the functional lifting approach based on currents is equivalent to our proposed framework if the integrand is convex (which implies polyconvexity). Moreover, as noted in [MC19], the finite element-based description of functional lifting introduced in [MC17] is a reformulation of the currents-based framework for convex integrands and, hence, the same considerations apply to it. For a full discussion of the concept of currents and the associated multilinear algebra we refer to [MC19]. We will use the following definition of polyconvexity:

Definition 6.31 ([Dac08, Definition 5.1 (iii)])**.** For $s, d \in \mathbb{N}$ we write $s \wedge d := \min\{s, d\}$ and

$$\tau(s,d) := \sum_{n=1}^{s\wedge d} \sigma(n), \quad \text{where} \quad \sigma(n) := \binom{d}{n}\binom{s}{n} = \frac{d!s!}{(n!)^2(d-n)!(s-n)!}. \tag{6.123}$$

A function $f\colon \mathbb{R}^{d,s} \to \overline{\mathbb{R}}$ is said to be *polyconvex* if there exists $g\colon \mathbb{R}^{\tau(s,d)} \to \overline{\mathbb{R}}$ convex such that $f(\xi) = g(M(\xi))$ where $M\colon \mathbb{R}^{d,s} \to \mathbb{R}^{\tau(s,d)}$ maps a matrix onto all of its subdeterminants (minors), i.e.,

$$M(\xi) := (\xi, \operatorname{adj}_2 \xi, \dots, \operatorname{adj}_{s\wedge d} \xi). \tag{6.124}$$

Here, we denote by $\operatorname{adj}_n \xi$ the vector of all $n \times n$ minors of the matrix $\xi \in \mathbb{R}^{d,s}$ for $2 \leq n \leq s \wedge d := \min\{s, d\}$. The function g is called a *polyconvex extension* of f.

Every convex function f is polyconvex with a polyconvex extension given by the function that ignores all s-minors for $s > 1$ and agrees with f otherwise (the *trivial* polyconvex extension). However, note that the polyconvex extension of a polyconvex function f fails to be unique in general, even if f happens to be convex. An example for this case is the Dirichlet integrand

$$f(\xi) := \frac{1}{2}\|\xi\|_F^2 = \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{s}\xi_{ij}^2. \tag{6.125}$$

as discussed in [GMS98b, Section 1.2.1] and [GMS98b, Section 1.2.4, Prop. 9].

For a *current of finite mass and compact support* $T \in \mathfrak{M}(\Omega \times \Gamma, \mathbb{R}^{\tau(d,s)+1})$ (a vectorial measure with values in $\mathbb{R}^{\tau(d,s)+1}$ in our nomenclature), the lifted functional in [MC19, Eq. (30)] is given by

$$\mathcal{F}_{\mathrm{curr}}(T) := \sup_{\omega \in \mathcal{K}_{\mathrm{curr}}} \langle T, \omega \rangle, \tag{6.126}$$

where

$$\mathcal{K}_{\mathrm{curr}} := \left\{ \omega \in C_0(\Omega \times \Gamma, \mathbb{R}^{\tau(d,s)+1}) : \omega_0(x,z) + g^*(x,z,\omega_{\bar{0}}(x,z)) \leq 0 \right\}. \tag{6.127}$$

Here, $g(x,z,\cdot)$ is the chosen polyconvex extension of the integrand with respect to the last component $f(x,z,\cdot)$, and $\omega_0 \in \mathbb{R}$ denotes the last component of $\omega \in \mathbb{R}^{\tau(d,s)+1}$ while $\omega_{\bar{0}} \in \mathbb{R}^{\tau(d,s)}$ is the vector consisting of the first $\tau(d,s)$ components of $\omega = (\omega_{\bar{0}}, \omega_0)$. As a lifting strategy for problems with Neumann boundary conditions, the lifted functional $\mathcal{F}_{\mathrm{curr}}$ is minimized in [MC19, Eq. (34)] over the following set of admissible currents

$$\mathcal{C} := \left\{ T \in \mathfrak{M}(\Omega \times \Gamma, \mathbb{R}^{\tau(d,s)+1}) : \pi_{1\#} T = [\![\Omega]\!], \operatorname{supp} \partial T \subset (\partial\Omega) \times \Gamma \right\}. \tag{6.128}$$

The marginal constraint $\pi_{1\#} T = [\![\Omega]\!]$ translates in our nomenclature to the requirement that T is a measure-valued function $T\colon \Omega \to \mathfrak{M}(\Gamma, \mathbb{R}^{\tau(d,s)+1})$ with the property that the last component is a probability measure on Γ for almost every $x \in \Omega$. And the boundary constraint $\operatorname{supp} \partial T \subset (\partial\Omega) \times \Gamma$ is a way of writing that T is a weak solution to the partial differential equation $\partial T = 0$ when tested with functions $\omega \in C_c^1(\mathring{\Omega} \times \Gamma, \mathbb{R}^{\tau(d,s)+1})$ with compact support in $\mathring{\Omega} \times \Gamma$. The first-order differential operator ∂ is the *boundary operator* from differential geometry that is defined via the exterior derivative of differential k-forms [MC19, Eq. (20)] and cannot be easily written down without introducing a lot of notation from exterior algebra, which we prefer to avoid in this work.

The functional lifting framework in [MC19] depends on the choice of polyconvex extension g for the integrand f through the set $\mathcal{K}_{\mathrm{curr}}$ in (6.127). Without going into the details about the definitions of ∂T and $[\![\Omega]\!]$, we now demonstrate that, in fact, the functional lifting framework in [MC19] exactly agrees with our proposed framework for first-order functionals for the case of a convex integrand with its trivial polyconvex extension. If f is convex and g is the trivial polyconvex extension of f, then $g^*(x,z,\omega_{\bar{0}})$ takes the value $+\infty$ whenever $\omega_{\bar{0}}$ does not vanish in all but the first ds components. This means that all but the first ds components of the functions $\omega_{\bar{0}}$ in $\mathcal{K}_{\mathrm{curr}}$ are enforced to be 0 and it means that, in this case, the functional $\mathcal{F}_{\mathrm{curr}}$ in (6.126) does only depend on the first ds and the last component of $T \in \mathfrak{M}(\Omega \times \Gamma, \mathbb{R}^{\tau(d,s)+1})$ and actually satisfies $\mathcal{F}_{\mathrm{curr}}(T) = \mathcal{B}_f(E,\mu)$,

where $E \in \mathfrak{M}(\Omega \times \Gamma, \mathbb{R}^{d,s})$ denotes the first ds and $\mu \in \mathfrak{M}(\Omega \times \Gamma)$ the last component of T. On the other hand, if $T \in \mathfrak{M}(\Omega \times \Gamma, \mathbb{R}^{\tau(d,s)+1})$ vanishes in all but the first ds and the last component, the partial differential equation $\partial T = 0$ reduces to our first-order continuity equation $\nabla_x \mu + \operatorname{div}_z E = 0$ as in Defn. 6.17.

In summary, the framework in [MC19] is applicable to a larger class of integrands (polyconvex integrands) than our framework, but our framework incorporates second-order variational models, whereas [MC19] is limited to first-order terms.

6.4.3 First-Order Functional Lifting for Manifold-Valued Problems

A functional lifting strategy for manifold-valued imaging problems with convex first-order regularization terms is described in Chap. 4. The variational models considered there are equivalent to the problem (6.1) discussed in this chapter with the choice $L := \nabla$ and Γ a submanifold of $\mathbb{R}^s$. The functional lifting strategy for manifold-valued imaging problems consists of minimizing, for $v \colon \Omega \to \mathcal{P}(\mathcal{M})$, the lifted functional (see (4.8) and (4.9))

$$\mathcal{F}_{\mathrm{mfd}}(v) := \sup \left\{ \int_\Omega \langle -\operatorname{div}_x p(x,\cdot) + q(x,\cdot), v(x) \rangle \, dx : (\nabla_z p, q) \in \mathcal{K}_f \right\}, \tag{6.129}$$

where $\mathcal{K}_f$ is exactly the constraint set (6.8) used in the current chapter. Because the focus of Chap. 4 is on a heuristic generalization of previous lifting strategies and a new discretization strategy, the function space for p is not specified in (6.129). Apart from this, and up to a change of sign in the dual variable p, the lifting strategy in Chap. 4 is equivalent to the dual version (6.85) of our proposed functional in the first-order case:

$$\mathcal{F}_D(\mu) = \sup \left\{ \int_{\Omega \times \Gamma} (\operatorname{div}_x p + q) \, d\mu : p \in C_c^1(\mathring{\Omega} \times \Gamma, \mathbb{R}^d) \text{ and } (-\nabla_z p, q) \in \mathcal{K}_f \right\}. \tag{6.130}$$

The interpretation of the manifold-valued lifting strategy as a special case of our framework allows to generalize the ideas in Chap. 4 to second-order models. However, a discretization of second-order derivatives on manifolds is more involved and we leave this for future work.

6.4.4 Second-Order Functional Lifting with Laplacian

The functional lifting framework introduced in Chap. 5 for the lifting of variational problems with Laplacian regularization is equivalent to the dual formulation (6.110) of our proposed framework. More precisely, in Chap. 5, the original problem (6.1) is

approximated for $L = \Delta$ with the minimization of the lifted functional (5.5), given by

$$\mathcal{F}_{\mathrm{lapl}}(\mu) := \sup_{(p,q)\in\mathcal{K}_{\mathrm{lapl}}} \int_{\Omega\times\Gamma} (\Delta_x p + q)\, d\mu, \tag{6.131}$$

where, as in (5.5),

$$\begin{aligned}\mathcal{K}_{\mathrm{lapl}} = \{(p,q) : p \in C_c^2(\Omega\times\Gamma), q \in L^1(\Omega\times\Gamma),\\ z \mapsto p(x,z) \text{ concave and } (\nabla_z p, q) \in \mathcal{K}_f\},\end{aligned} \tag{6.132}$$

On the other hand, the dual formulation (6.110) of our proposed framework with Laplacian second-order terms is

$$\begin{aligned}\mathcal{F}_D(\mu) = \sup\Bigg\{ \int_{\Omega\times\Gamma} (-\Delta_x p + q)\, d\mu : p \in C_c^2(\mathring{\Omega}\times\Gamma, \mathbb{R}^d),\ \nabla_z^2 p \text{ pos. semidef.}\\ \text{and } (-\nabla_z p, q) \in \mathcal{K}_f \Bigg\}.\end{aligned} \tag{6.133}$$

The definiteness constraint on the Hessian of the dual variable p in (6.133) corresponds to the concavity constraint in (6.132). Because the focus of Chap. 5 is on a heuristic generalization of previous lifting strategies to second-order models, the choice of function spaces for p and q are not discussed. Apart from this, and up to a change of sign in the dual variable p, the lifting strategy in Chap. 5 is equivalent to the dual version (6.133) of our proposed functional.

While the connection $\mathcal{F}_{\mathrm{lapl}}(\delta_u) \leq F(u)$ between the lifted and the original functional could only be established as an inequality in Prop. 5.1, we proved in Prop. 6.29 that this connection holds as an equality for our proposed primal formulation $\mathcal{F}$. Note that Prop. 6.29 does not easily imply the equality $\mathcal{F}_{\mathrm{lapl}}(\delta_u) = F(u)$ since we did not show strong duality ($\mathcal{F} = \mathcal{F}_D$).

The framework proposed in the current chapter clarifies some mathematical details of the functional lifting model for problems with Laplacian regularization introduced in Chap. 5. Our framework allows to generalize the ideas in Chap. 5 to arbitrary second-order regularizers. However, a discretization of the full Hessian second-order derivatives in the finite element framework used in this work (see Sect. 2.7) is more involved and we leave this for future work.

6.4.5 Total Variation Regularization of ODFs

Finally, we mention that our proposed measure-valued lifting framework bears a striking similarity to the framework for total variation regularization of measure-valued (more precisely: orientation distribution functions from diffusion-weighted MRI) images introduced in Chap. 3. In (3.55), the following total variation functional for measure-valued functions $\mu\colon \Omega \to \mathcal{P}(\Gamma)$ is introduced:

$$\begin{aligned}\mathrm{TV}_{\mathrm{KR}}(\mu) := \sup \Bigg\{ \int_\Omega \langle -\operatorname{div} p(x), \mu(x) \rangle \, dx : \; & p \in C_c^1(\Omega, [\mathrm{Lip}_0(\Gamma)]^d), \\ & \|p(x)\|_{[\mathrm{Lip}_0(\Gamma)]^d} \leq 1 \Bigg\}.\end{aligned} \tag{6.134}$$

On the other hand, consider the dual formulation (6.85) of our proposed lifted functional for the integrand $f(x, z, \xi) := \|\xi\|_F$. Noting that $f^*(x, z, \xi) = \delta_{\|\xi\|_F \leq 1}$, the dual constraint set $\mathcal{K}_f$ simplifies a lot so that we can omit the dual variable q in (6.85) and obtain:

$$\mathcal{F}_D(\mu) = \sup \left\{ \int_{\Omega \times \Gamma} \operatorname{div}_x p \, d\mu : p \in C_c^1(\mathring{\Omega} \times \Gamma, \mathbb{R}^d) \text{ and } \|\nabla_z p\|_\infty \leq 1 \right\}. \tag{6.135}$$

The constraint $\|\nabla_z p\|_\infty \leq 1$ in (6.135) corresponds to the Lipschitz seminorm constraint in (6.134). Hence, the functional $\mathcal{F}_D$ differs from $\mathrm{TV}_{\mathrm{KR}}$ only in the choice of the dual function space $C_c^1(\Omega, [\mathrm{Lip}_0(\Gamma)]^d)$ instead of $C_c^1(\mathring{\Omega} \times \Gamma, \mathbb{R}^d)$. We leave a rigorous analysis of this correspondence for future work.

Summing up, our proposed framework not only defines a high-dimensional convex approximation of a non-convex problem. It also defines a meaningful notion of regularity for measure-valued functions. This is no surprise given its connection to functionals from dynamical optimal transport where the regularity of curves with values in spaces of measures is the object of study. Beyond the total variation functional used in Chap. 3, our framework can serve as a blueprint for new regularizers in the restoration of orientation distribution functions and similar measure-valued imaging problems.

6.5 Conclusion and Outlook

We proposed a functional lifting framework that is based on measure-valued functions and motivated by the Benamou-Brenier theory of dynamical optimal transport. The lifted functional is a convex energy on the space of vectorial measures and provably agrees with the original non-convex functional on the lifted versions of images for both first-order and second-order variational models. We provided a primal and dual formulation

for an augmented version of the lifted functional and showed that the two formulations agree. Under suitable assumptions on the integrand, we could prove that the lifted functional enforces absolute continuity of the vectorial measures with respect to their last component. This property allowed to view one part of the vectorial measure as defining the support and the other part as defining the derivative part. For first- and second-order derivatives, we proposed weak continuity equations that enforce this relationship in vectorial measures.

The proposed framework allows to view a large class of recent functional lifting strategies under a unified mathematical theory of measure-valued functions. This includes calibration-based and currents-based lifting approaches, but also recent second-order (Laplacian) lifting strategies. In the second-order case, our model improves the understanding of the relationship between the lifted and the original functional compared to previous considerations in the literature.

While the well-known properties in the scalar first-order case translate from the calibration-based lifting method, the following properties remain an open question in the vectorial or higher-order case:

- Does the minimal value of the lifted functional $\mathcal{F}$ agree with the minimum of the original functional F?
- Is every minimizer of $\mathcal{F}$ a superposition of minimizers of F in a suitable sense?
- Given a minimizer of $\mathcal{F}$, is there an explicit or efficient way that provably computes a minimizer of the original functional F?

In any case, our proposed framework manages to unite notationally diverse functional lifting approaches in a single consistent mathematical setting with the help of notions from dynamical optimal transport and functional analysis, and its modular structure will hopefully inspire interesting further developments in the future.

7 Discussion

Summary With this thesis, we hope to have established a broad and unifying perspective on recent functional lifting frameworks that will inspire more applications, improved implementations and a better understanding of the limitations and theoretical properties in the future. While our perspective was driven by applications throughout Chaps. 3–5, we aimed at bringing together the different approaches in a final theoretical superstructure for measure-valued models in Chap. 6.

Motivated by the mathematical perspective on Q-ball data as images taking values in a space of probability measures, our proposed total variation functional provides a generic way of measuring the spatial regularity of diffusion-weighted MRI signals. We discussed some of its geometric properties, such as isotropy, in theory and demonstrated its scale-space behavior in practice. While numerical performance is still limited by the non-smooth high-dimensional structure compared to state-of-the-art methods, applications to phantom and real-world datasets indicate that the proposed regularization strategy is able to remove noise while preserving structural information, such as edges (Chap. 3).

We proposed a measure-valued functional lifting approach for the convex optimization of a large class of image processing problems where the data satisfies geometric, periodical or symmetry constraints, so-called manifold-valued images. It allows to reliably handle arbitrary pointwise data discrepancy terms and many different convex regularization terms by employing an efficient finite element method on the underlying manifold (Chap. 4).

In order to apply the functional lifting idea to classical variational formulations of image registration, we generalized it to energy functionals that take higher-order derivatives into account. The numerical cost of the proposed generalization is comparable to state-of-the-art first-order functional lifting methods. While we demonstrated its limitations in a synthetic example with symmetries, the proposed method performs well in real-world image registration tasks (Chap. 5).

We presented a concept that unifies a large class of recent functional lifting strategies using notions from dynamical optimal transport. At the heart of this concept stands the lifting of an augmented functional that can be formulated for variational models

of arbitrary order. By combining it with suitable continuity equation constraints, we demonstrated how it can be used for the functional lifting of non-convex first- and second-order variational problems. We hope that the framework's modular structure will inspire higher-order and other generalizations in the future (Chap. 6).

Future directions of research Our proposed perspective on functional lifting strategies in the language of measure-valued image processing opens up several directions for interesting further developments:

- **Histogram propagation.** When the distribution (or histogram) of traffic density, temperature or wind direction over a fixed period of time is measured for given geographical locations, e.g., by weather stations, a natural problem is the estimation of distributions in the areas away from the sample locations. This problem, sometimes called *histogram propagation*, is a classical image inpainting or interpolation problem and has been demonstrated [Sol+14] to benefit from optimal transport metrics. In the future, our proposed model for the restoration of Q-ball images can be adapted to the problem of histogram propagation.

- **Efficient optimization and implementation.** Even though the convex structure of the lifted problems allows to use state-of-the-art efficient and highly parallelizable convex optimization algorithms, the high-dimensionality and non-smoothness of the lifted setting limits the numerical performance and practical applicability, in particular in real-time applications. On the other hand, our proposed implementations are not optimized for efficiency and there are many starting points for improvement. For instance:
  - Our proposed finite element-based discretization can be improved by making use of more advanced [Her+19; Ben+19] or adaptive [PR00; BJ08; HR14] finite element methods.
  - The connection to (dynamical) optimal transport introduced in Chap. 6 suggests the adaptation of efficient optimization strategy from the context of numerical optimal transport, such as entropy [Cut13; Bra+17; Cla+19] or quadratic [LMM19] regularization.

- **Projection or interpretation of lifted solutions.** Minimizers of the convexified and lifted functional are measure-valued functions in general, that require some post-processing in order to be interpreted in the context of the original non-convex problem. In the scalar-valued first-order case, each solution of the lifted problem

is provably related to at least one global minimizer of the original functional. For higher-order or vectorial problems, it remains an open problem whether each minimizer of the lifted functional contains usable information about at least one solution of the original problem. Until a definite answer to this question is available, there is plenty of room for efficient rounding or projection strategies that might produce approximate solutions to the original problem from solutions in the lifted setting.

- **Generalizations to more differential operators.** The modular structure of our proposed model allows to decouple the lifting of the objective functional from the lifting of the features involved. While we only considered generic first- and second-order derivatives as features so far, models with higher-order, non-local or non-linear differential operators in their convex component and couplings of first- and second-order terms, such as with the total generalized variation [BKP10], or roto-translational features [Dui+19; CP19] are conceivable.

- **Non-local data terms.** Our proposed generalization of the functional lifting approach to models with second-order regularization is an important first step in making the functional lifting strategy applicable to modern image registration problems. However, popular image registration frameworks use non-pointwise features, such as Normalized Gradient Fields (NGF) in the highly non-convex data discrepancy term. This requires the study of models with first- and second-order derivatives where the dependency on the first-order term is allowed to be non-convex.

The numerous and diverse directions for further investigation and wider applications presented above once again demonstrate the appeal and modularity of our proposed functional lifting framework. Bearing in mind the framework's compelling properties that were discussed in this thesis, we think that it is justified to face the additional challenges that appear when using measure-valued variational models.

Bibliography

[ABD03] G. Alberti, G. Bouchitté, and G. Dal Maso. "The calibration method for the Mumford-Shah functional and free-discontinuity problems". In: *Calc Var Partial Differ Equ* 16.3 (2003), pp. 299–333.

[ABM06] H. Attouch, G. Buttazzo, and G. Michaille. *Variational analysis in Sobolev and BV spaces. Applications to PDEs and optimization.* Vol. 6. SIAM, 2006.

[Abs+16] P.-A. Absil, P.-Y. Gousenbourger, P. Striewski, and B. Wirth. "Differentiable Piecewise-Bézier Surfaces on Riemannian Manifolds". In: *SIAM J Imaging Sci* 9 (2016), pp. 1788–1828.

[AFP00] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of bounded variation and free discontinuity problems.* Clarendon Press, 2000.

[AGS04] L. Ambrosio, N. Gigli, and G. Savaré. "Gradient flows with metric and differentiable structures, and applications to the Wasserstein space". In: *Atti Accad Naz Lincei Rend Cl Sci Fis Mat Natur* 15.3–4 (2004), pp. 327–343.

[AGS08] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures. 2nd ed.* Birkhäuser, 2008.

[Ahr+13] C. Ahrens, J. Nealy, F. Pérez, and S. van der Walt. "Sparse reproducing kernels for modeling fiber crossings in diffusion weighted imaging". In: *Proc ISBI 2013.* 2013, pp. 688–691.

[AK06] G. Aubert and P. Kornprobst. *Mathematical problems in image processing: partial differential equations and the calculus of variations.* Springer, 2006.

[ALS09] I. Aganj, C. Lenglet, and G. Sapiro. "ODF Reconstruction in Q-Ball Imaging with Solid Angle Consideration". In: *Proc ISBI 2009.* 2009, pp. 1398–1401.

[Amb03] L. Ambrosio. "Lecture Notes on Optimal Transport Problems". In: *Mathematical Aspects of Evolving Interfaces.* Ed. by L. Ambrosio, K. Deckelnick, G. Dziuk, M. Mimura, V. A. Solonnikov, and H. M. Soner. Springer, 2003, pp. 1–52.

[Amb90] L. Ambrosio. "Metric space valued functions of bounded variation". In: *Ann Sc Norm Super Pisa, Cl Sci, IV Ser* 17.3 (1990), pp. 439–478.

[AMS09] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.

[Åst+17] F. Åström, S. Petra, B. Schmitzer, and C. Schnörr. "Image Labeling by Assignment". In: *J Math Imaging Vis* 58.2 (2017), pp. 211–238.

[AT90] L. Ambrosio and V. M. Tortorelli. "Approximation of functionals depending on jumps by elliptic functionals via Γ-convergence". In: *Commun Pure Appl Math* 43.8 (1990), pp. 999–1036.

[Bač+16] M. Bačák, R. Bergmann, G. Steidl, and A. Weinmann. "A Second Order Nonsmooth Variational Model for Restoring Manifold-Valued Images". In: *SIAM J Sci Comput* 38.1 (2016), A567–A597.

[Bač14] M. Bačák. *Convex Analysis and Optimization in Hadamard Spaces*. De Gruyter, 2014.

[Bac19] F. Bach. "Submodular functions: from discrete to continuous domains". In: *Math Program Series A* 175.1–2 (2019), pp. 419–459.

[Bae+14] E. Bae, J. Yuan, X.-C. Tai, and Y. Boykov. "A Fast Continuous Max-Flow Approach to Non-convex Multi-labeling Problems". In: *Efficient Algorithms for Global Optimization Methods in Computer Vision*. Ed. by A. Bruhn, T. Pock, and X.-C. Tai. Springer, 2014, pp. 134–154.

[Bal89] J. M. Ball. "A version of the fundamental theorem for young measures". In: *PDEs and Continuum Models of Phase Transitions*. Ed. by M. Rascle, D. Serre, and M. Slemrod. Springer, 1989, pp. 207–215.

[Bau+16] M. Baust, A. Weinmann, M. Wieczorek, T. Lasser, M. Storath, and N. Navab. "Combined Tensor Fitting and TV Regularization in Diffusion Tensor Imaging Based on a Riemannian Manifold Approach". In: *IEEE Trans Med Imaging* 35 (2016), pp. 1972–1989.

[BB00] J.-D. Benamou and Y. Brenier. "A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem". In: *Numer Math* 84.3 (2000), pp. 375–393.

[BC92] D. Bleecker and G. Csordas. *Basic partial differential equations*. Van Nostrand Reinhold, 1992.

[BDH96] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa. "The quickhull algorithm for convex hulls". In: *ACM Trans Math Softw* 22.4 (1996), pp. 469–483.

[Bec+12] S. Becker, K. Tabelow, H. Voss, A. Anwander, R. Heidemann, and J. Polzehl. "Position-orientation adaptive smoothing of diffusion weighted magnetic resonance data (POAS)". In: *Med Image Anal* 16.6 (2012), pp. 1142–1155.

[Ben+19] J. Benn, S. Marsland, R. I. McLachlan, K. Modin, and O. Verdier. "Currents and Finite Elements as Tools for Shape Space". In: *J Math Imaging Vis* 61.8 (2019), pp. 1197–1220.

[Ber+17] F. Bernard, F. R. Schmidt, J. Thunberg, and D. Cremers. "A Combinatorial Solution to Non-Rigid 3D Shape-to-Image Matching". In: *Proc ICCV 2017*. 2017, pp. 1436–1445.

[Ber+18a] R. Bergmann, J. H. Fitschen, J. Persch, and G. Steidl. "Priors with Coupled First and Second Order Differences for Manifold-Valued Image Processing". In: *J Math Imaging Vis* 60 (2018), pp. 1459–1481.

[Ber+18b] R. Bergmann, F. Laus, J. Persch, and G. Steidl. *Recent Advances in Denoising of Manifold-Valued Images*. Tech. rep. arXiv:1812.08540. arXiv, 2018.

[BF18] G. Bouchitté and I. Fragalà. "A Duality Theory for Non-convex Problems in the Calculus of Variations". In: *Arch Rational Mech Anal* 229.1 (2018), pp. 361–415.

[BF19] K. Bredies and S. Fanzon. *An optimal transport approach for solving dynamic Inverse Probl in spaces of measures*. Tech. rep. arXiv:1901.10162. arXiv, 2019.

[BFS12] M. Burger, M. Franek, and C.-B. Schönlieb. "Regularized regression and density estimation based on optimal transport". In: *Appl Math Res Express* 2012.2 (2012), pp. 209–253.

[BHS11] F. Bachmann, R. Hielscher, and H. Schaeben. "Grain detection from 2d and 3d EBSD data - Specification of the MTEX algorithm". In: *Ultramicroscopy* 111.12 (2011), pp. 1720–1733.

[BJ08] W. Bangerth and A. Joshi. "Adaptive finite element methods for the solution of inverse problems in optical tomography". In: *Inverse Probl* 24.3 (2008), p. 034011.

[BKP10] K. Bredies, K. Kunisch, and T. Pock. "Total generalized variation". In: *SIAM J Imaging Sci* 3.3 (2010), pp. 492–526.

[BKR11] A. Blake, P. Kohli, and C. Rother, eds. *Markov random fields for vision and image processing*. MIT Press, 2011.

[BL11] K. Bredies and D. Lorenz. *Mathematische Bildverarbeitung. Einführung in Grundlagen und moderne Theorie*. Vieweg+Teubner, 2011.

[BL15] C. Brauer and D. Lorenz. "Cartoon-Texture-Noise Decomposition with Transport Norms". In: *Scale Space and Variational Methods in Computer Vision*. Ed. by J.-F. Aujol, M. Nikolova, and N. Papadakis. Springer, 2015, pp. 142–153.

[BL18] K. Bredies and D. Lorenz. *Mathematical image processing. Translated from the German*. Birkhäuser, 2018.

[BML94] P. J. Basser, J. Mattiello, and D. LeBihan. "MR diffusion tensor spectroscopy and imaging". In: *Biophys J* 66.1 (1994), pp. 259–267.

[Bou04] N. Bourbaki. *Integration*. Springer, 2004.

[BPS16] R. Bergmann, J. Persch, and G. Steidl. "A Parallel Douglas-Rachford Algorithm for Minimizing ROF-like Functionals on Images with Values in Symmetric Hadamard Manifolds". In: *SIAM J Imaging Sci* 9 (2016), pp. 901–937.

[Bra+17] C. Brauer, C. Clason, D. Lorenz, and B. Wirth. *A Sinkhorn-Newton method for entropic optimal transport*. Tech. rep. arXiv:1710.06635. arXiv, 2017.

[Bre+18] K. Bredies, M. Holler, M. Storath, and A. Weinmann. "Total Generalized Variation for Manifold-Valued Data". In: *SIAM J Imaging Sci* 11 (2018), pp. 1785–1848.

[Bre+19] K. Bredies, M. Carioni, S. Fanzon, and F. Romero. *On the extremal points of the ball of the Benamou-Brenier energy*. Tech. rep. arXiv:1907.11589. arXiv, 2019.

[Bre03] Y. Brenier. "Extended Monge-Kantorovich Theory". In: *Optimal Transportation and Applications*. Ed. by L. Ambrosio, L. A. Caffarelli, Y. Brenier, G. Buttazzo, C. Villani, and S. Salsa. Springer, 2003, pp. 91–121.

[BT09] E. Bae and X.-C. Tai. "Graph Cut Optimization for the Piecewise Constant Level Set Method Applied to Multiphase Image Segmentation". In: *Scale Space and Variational Methods in Computer Vision*. Ed. by X.-C. Tai, K. Mørken, M. Lysaker, and K.-A. Lie. Springer, 2009, pp. 1–13.

[BT15] E. Bae and X.-C. Tai. "Efficient Global Minimization Methods for Image Segmentation Models with Four Regions". In: *J Math Imaging Vis* 51.1 (2015), pp. 71–97.

[BT18] R. Bergmann and D. Tenbrinck. "A Graph Framework for Manifold-Valued Data". In: *SIAM J Imaging Sci* 11 (2018), pp. 325–360.

[BV04] S. P. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.

[BV88] G. Bouchitté and M. Valadier. "Integral representation of convex functionals on a space of measures". In: *J Funct Anal* 80.2 (1988), pp. 398–420.

[BVZ01] Y. Boykov, O. Veksler, and R. Zabih. "Fast approximate energy minimization via graph cuts". In: *IEEE Trans Pattern Anal Mach Intell* 23.11 (2001), pp. 1222–1239.

[BWM18] K. Brehmer, B. Wacker, and J. Modersitzki. "A Novel Similarity Measure for Image Sequences". In: *Biomedical Image Registration*. Ed. by S. Klein, M. Staring, S. Durrleman, and S. Sommer. Springer, 2018, pp. 47–56.

[Cal91] P. T. Callaghan. *Principles of Nuclear Magnetic Resonance Microscopy*. Clarendon Press, 1991.

[Can+15] E. J. Canales-Rodríguez et al. "Spherical Deconvolution of Multichannel Diffusion MRI Data with Non-Gaussian Noise Models and Spatial Regularization". In: *PLOS ONE* 10.10 (2015). Ed. by A. Leemans, Article ID e0138910.

[Car00] N. L. Carothers. *Real Analysis*. Cambridge University Press, 2000.

[CCP12] A. Chambolle, D. Cremers, and T. Pock. "A convex approach to minimal partitions". In: *SIAM J Imaging Sci* 5.4 (2012), pp. 1113–1158.

[CE05] T. F. Chan and S. Esedoglu. "Aspects of Total variation regularized L^1 function approximation". In: *SIAM J Appl Math* 65.5 (2005), pp. 1817–1837.

[CEN06] T. F. Chan, S. Esedoglu, and M. Nikolova. "Algorithms for Finding Global Minimizers of Image Segmentation and Denoising Models". In: *SIAM J Appl Math* 66.5 (2006), pp. 1632–1648.

[Cha+09] R. Chartrand, B. Wohlberg, K. R. Vixie, and E. M. Bollt. "A Gradient Descent Solution to the Monge-Kantorovich Problem". In: *Appl Math Sci* 3.22 (2009), pp. 1071–1080.

[Cha+10] A. Chambolle, V. Caselles, D. Cremers, M. Novaga, and T. Pock. "An introduction to total variation for image analysis". In: *Theoretical foundations and numerical methods for sparse recovery* 9 (2010), pp. 263–340.

[Cha01] A. Chambolle. "Convex Representation for Lower Semicontinuous Envelopes of Functionals in L^1". In: *J Convex Anal* 8.1 (2001), pp. 149–170.

[Che+04] C. Chefd'Hotel, D. Tschumperlé, R. Deriche, and O. D. Faugeras. "Regularizing Flows for Constrained Matrix-Valued Images". In: *J Math Imaging Vis* 20 (2004), pp. 147–162.

[Chi+15] G. S. Chilla, C. H. Tan, C. Xu, and C. L. Poh. "Diffusion weighted magnetic resonance imaging and its recent trend—a survey". In: *Quant Imaging Med Surg* 5.3 (2015), pp. 407–422.

[CKR98] G. Călinescu, H. Karloff, and Y. Rabani. "An Improved Approximation Algorithm for Multiway Cut". In: *Proc STOC 1998*. 1998, pp. 48–52.

[CKS01] T. F. Chan, S. H. Kang, and J. Shen. "Total Variation Denoising and Enhancement of Color Images Based on the CB and HSV Color Models". In: *J Vis Commun Image Represent* 12 (2001), pp. 422–435.

[CL01] P. Choné and H. V. J. Le Meur. "Non-Convergence Result for Conformal Approximation of Variational Problems Subject to a Convexity Constraint". In: *Numer Func Anal Opt* 22.5–6 (2001), pp. 529–547.

[Cla+19] C. Clason, D. A. Lorenz, H. Mahler, and B. Wirth. *Entropic regularization of continuous optimal transport problems*. Tech. rep. arXiv:1906.01333. arXiv, 2019.

[Cla13] F. Clarke. *Functional Analysis, Calculus of Variations and Optimal Control*. Springer, 2013.

[CMC17] D. Chen, J.-M. Mirebeau, and L. D. Cohen. "Global Minimum for a Finsler Elastica Minimal Path Approach". In: *Int J Comput Vis* 122.3 (2017), pp. 458–483.

[Com18] P. L. Combettes. "Perspective Functions: Properties, Constructions, and Examples". In: *Set-Valued and Variational Analysis* 26.2 (2018), pp. 247–264.

[CP11] A. Chambolle and T. Pock. "A first-order primal-dual algorithm for convex problems with applications to imaging". In: *J Math Imaging Vis* 40.1 (2011), pp. 120–145.

[CP19] A. Chambolle and T. Pock. "Total roto-translational variation." In: *Numer Math* 142.3 (2019), pp. 611–666.

[Cre+13] E. Creusen, R. Duits, A. Vilanova, and L. Florack. "Numerical Schemes for Linear and Non-Linear Enhancement of DW-MRI". In: *Numer Math Theor Meth Appl* 6.1 (2013), pp. 138–168.

[CS12] D. Cremers and E. Strekalovskiy. "Total Cyclic Variation and Generalizations". In: *J Math Imaging Vis* 47 (2012), pp. 258–277.

[Cut13] M. Cuturi. "Sinkhorn Distances: Lightspeed Computation of Optimal Transport". In: *Proc NIPS 2013*. Ed. by C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger. Curran Associates, Inc., 2013, pp. 2292–2300.

[CV01] T. F. Chan and L. A. Vese. "Active contours without edges". In: *IEEE Trans Image Process* 10.2 (2001), pp. 266–277.

[CV17] C. Clason and T. Valkonen. "Primal-Dual Extragradient Methods for Nonlinear Nonsmooth PDE-Constrained Optimization". In: *SIAM J Optim* 27.3 (2017), pp. 1314–1339.

[Dac08] B. Dacorogna. *Direct methods in the calculus of variations*. Springer, 2008.

[Dad+14] A. Daducci et al. "Quantitative Comparison of Reconstruction Methods for Intra-Voxel Fiber Recovery From Diffusion MRI". In: *IEEE Trans Med Imaging* 33.2 (2014), pp. 384–399.

[DAG09] V. Duval, J.-F. Aujol, and Y. Gousseau. "The TVL1 Model: A Geometric Point of View". In: *Multiscale Model Simul* 8.1 (2009), pp. 154–189.

[Dal79] G. Dal Maso. "Integral representation on BV(Ω) of Γ-limits of variational integrals". In: *Manuscripta Math* 30.4 (1979), pp. 387–416.

[DD11] M. Duchon and C. Debiève. "Functions with bounded variation in locally convex space". In: *Tatra Mt Math Publ* 49.1 (2011), pp. 89–98.

[DE13] G. Dziuk and C. M. Elliott. "Finite element methods for surface PDEs". In: *Acta Numerica* 22 (2013), pp. 289–396.

[Del+07] S. Delputte, H. Dierckx, E. Fieremans, Y. D'Asseler, R. Achten, and I. Lemahieu. "Postprocessing of Brain White Matter Fiber Orientation Distribution Functions". In: *Proc ISBI 2007*. 2007, pp. 784–787.

[Del+09] A. Delaunoy, K. Fundana, E. Prados, and A. Heyden. "Convex multi-region segmentation on manifolds". In: *Proc ICCV 2009*. 2009, pp. 662–669.

[Des08] M. Descoteaux. "High Angular Resolution Diffusion MRI: from Local Estimation to Segmentation and Tractography". PhD Thesis. University of Nice-Sophia Antipolis, 2008.

[DF11] R. Duits and E. Franken. "Left-Invariant Diffusions on the Space of Positions and Orientations and their Application to Crossing-Preserving Smoothing of HARDI images". In: *Int J Comput Vis* 92.3 (2011), pp. 231–264.

[DGG06] L. De Pascale, M. S. Gelli, and L. Granieri. "Minimal measures, one-dimensional currents and the Monge–Kantorovich problem". In: *Calc Var* 27.1 (2006), pp. 1–23.

[DU77] J. Diestel and J. J. j. Uhl. *Vector measures*. AMS, 1977.

[Dud77] R. M. Dudley. "On second derivates of convex functions". In: *Mathematica Scandinavica* 41 (1977), pp. 159–174.

[Dui+13] R. Duits, T. Dela Haije, E. Creusen, and A. Ghosh. "Morphological and Linear Scale Spaces for Fiber Enhancement in DW-MRI". In: *J Math Imaging Vis* 46.3 (2013), pp. 326–368.

[Dui+19] R. Duits, E. St-Onge, J. Portegies, and B. Smets. "Total Variation and Mean Curvature PDEs on the Space of Positions and Orientations". In: *Scale Space and Variational Methods in Computer Vision*. Ed. by J. Lellmann, M. Burger, and J. Modersitzki. Springer, 2019, pp. 211–223.

[EOK11] H. H. Ehricke, K. M. Otto, and U. Klose. "Regularization of bending and crossing white matter fibers in MRI Q-ball fields". In: *Magn Reson Imaging* 29.7 (2011), pp. 916–926.

[ET99] I. Ekeland and R. Témam. *Convex analysis and variational problems*. SIAM, 1999.

[EZC10] E. Esser, X. Zhang, and T. Chan. "A General Framework for a Class of First Order Primal-Dual Algorithms for Convex Optimization in Imaging Science". In: *SIAM J Imaging Sci* 3.4 (2010), pp. 1015–1046.

[Fed74] H. Federer. "Real flat chains, cochains and variational problems". In: *Indiana U Math J* 24 (1974), pp. 351–407.

[Fer+14] S. Ferradans, N. Papadakis, G. Peyré, and J.-F. Aujol. "Regularized discrete optimal transport". In: *SIAM J Imaging Sci* 7.3 (2014), pp. 1853–1882.

[Fie11] P. Fieguth. *Statistical image processing and multidimensional modeling*. Springer, 2011.

[Fle12] P. T. Fletcher. "Geodesic Regression and the Theory of Least Squares on Riemannian Manifolds". In: *Int J Comput Vis* 105 (2012), pp. 171–185.

[FLS16] J. H. Fitschen, F. Laus, and G. Steidl. "Transport between RGB Images Motivated by Dynamic Optimal Transport". In: *J Math Imaging Vis* 56.3 (2016), pp. 409–429.

[FLS17] J. H. Fitschen, F. Laus, and B. Schmitzer. "Optimal Transport for Manifold-Valued Images". In: *Proc SSVM 2017*. 2017, pp. 460–472.

[FM03] B. Fischer and J. Modersitzki. "Curvature Based Image Registration". In: *J Math Imaging Vis* 18.1 (2003), pp. 81–85.

[FM08] B. Fischer and J. Modersitzki. "Ill-posed medicine—an introduction to image registration". In: *Inverse Probl* 24.3 (2008), Article ID 034008.

[Foc16] M. Focardi. "Fine regularity results for Mumford-Shah minimizers: porosity, higher integrability and the Mumford-Shah conjecture." In: *Free discontinuity problems. Lecture notes given at the school, Pisa, Italy, July 2014*. Pisa: Edizioni della Normale, 2016, pp. 1–68.

[Fus03] N. Fusco. "An Overview of the Mumford-Shah Problem". In: *Milan J Math* 71.1 (2003), pp. 95–119.

[Gar+14] E. Garyfallidis, M. Brett, B. Amirbekian, A. Rokem, S. Van Der Walt, M. Descoteaux, I. Nimmo-Smith, and D. Contributors. "Dipy, a library for the analysis of diffusion MRI data". In: *Front Neuroinform* 8.8 (2014), pp. 1–17.

[GC10] B. Goldlücke and D. Cremers. "Convex Relaxation for Multilabel Problems with Product Label Spaces". In: *Proc ECCV 2010*. 2010, pp. 225–238.

[GEB13] T. Goldstein, E. Esser, and R. Baraniuk. "Adaptive primal dual optimization for image processing and learning". In: *Proc NIPS Workshop Optim Mach Learn 2013*. 2013.

[Ges+09] D. Gesch, G. Evans, J. Mauck, J. Hutchinson, W. J. Carswell Jr, et al. "The national map - Elevation". In: *US Geol Surv Fact Sheet* 3053.4 (2009).

[GF75] E. D. Giorgi and T. Franzoni. "Su un tipo di convergenza variazionale". In: *Atti Accad Naz Lincei Rend Cl Sci Fis Mat Natur* 58 (1975), pp. 842–850.

[GH96a] M. Giaquinta and S. Hildebrandt. *Calculus of variations 1. The Lagrangian formalism*. Springer, 1996.

[GH96b] M. Giaquinta and S. Hildebrandt. *Calculus of variations 2. The Hamiltonian formalism*. Springer, 1996.

[GLY15] T. Goldstein, M. Li, and X. Yuan. "Adaptive Primal-Dual Splitting Methods for Statistical Learning and Image Processing". In: *Proc NIPS 2015*. 2015, pp. 2089–2097.

[GMS98a] M. Giaquinta, G. Modica, and J. Souček. *Cartesian currents in the calculus of variations I. Cartesian currents*. Springer, 1998.

[GMS98b] M. Giaquinta, G. Modica, and J. Souček. *Cartesian currents in the calculus of variations II. Variational integrals*. Springer, 1998.

[Goh+09] A. Goh, C. Lenglet, P. M. Thompson, and R. Vidal. "Estimating Orientation Distribution Functions with Probability Density Constraints and Spatial Regularity". In: *Proc MICCAI 2009*. Ed. by G.-Z. Yang, D. Hawkes, D. Rueckert, A. Noble, and C. Taylor. Springer, 2009, pp. 877–885.

[Gol+13] T. Goldstein, M. Li, X. Yuan, E. Esser, and R. Baraniuk. *Adaptive Primal-Dual Hybrid Gradient Methods for Saddle-Point Problems*. Tech. rep. arXiv: 1305.0546. arXiv, 2013.

[Gos05] A. Goshtasby. *2-D and 3-D image registration for medical, remote sensing, and industrial applications*. John Wiley & Sons, 2005.

[GPS89] D. M. Greig, B. T. Porteous, and A. H. Seheult. "Exact Maximum A Posteriori Estimation for Binary Images". In: *J R Stat Soc Series B Stat Methodol* 51.2 (1989), pp. 271–279.

[Gra07] L. Granieri. "On action minimizing measures for the Monge-Kantorovich problem". In: *NoDEA-Nonlinear Diff* 14.1–2 (2007), pp. 125–152.

[Gra09] L. Granieri. "Remarks on the stability of mass minimizing currents in the Monge-Kantorovich problem". In: *Indag Math* 20.4 (2009), pp. 527–536.

[GSC12] B. Goldluecke, E. Strekalovskiy, and D. Cremers. "The natural vectorial total variation which arises from geometric measure theory". In: *SIAM J Imaging Sci* 5.2 (2012), pp. 537–563.

[GSC13] B. Goldluecke, E. Strekalovskiy, and D. Cremers. "Tight convex relaxations for vector-valued labeling". In: *SIAM J Imaging Sci* 6.3 (2013), pp. 1626–1664.

[Guo+18] Y. Guo, Y. Liu, T. Georgiou, and M. S. Lew. "A review of semantic segmentation using deep neural networks". In: *Int J Multimed Inf Retr* 7.2 (2018), pp. 87–93.

[Her+19] M. Herrmann, R. Herzog, S. Schmidt, J. Vidal-Núñez, and G. Wachsmuth. "Discrete Total Variation with Finite Elements and Applications to Imaging". In: *J Math Imaging Vis* 61.4 (2019), pp. 411–431.

[Hes41] M. R. Hestenes. "Extension of the range of a differentiable function". In: *Duke Math J* 8 (1941), pp. 183–192.

[HL82] R. Harvey and H. B. Lawson. "Calibrated geometries". In: *Acta Mathematica* 148 (1982), pp. 47–157.

[HR09] P. J. Huber and E. M. Ronchetti. *Robust statistics. 2nd revised ed.* John Wiley & Sons, 2009.

[HR14] M. Hintermüller and M. Rincon-Camacho. "An adaptive finite element method in L^2-TV-based image denoising". In: *Inverse Probl Imaging* 8.3 (2014), p. 685.

[HR15] T. Hohage and C. Rügge. "A coherence enhancing penalty for Diffusion MRI: regularizing property and discrete approximation". In: *SIAM J Imaging Sci* 8.3 (2015), pp. 1874–1893.

[HS65] E. Hewitt and K. Stromberg. *Real and Abstract Analysis*. Springer, 1965.

[HT01] S. Haker and A. Tannenbaum. "Optimal mass transport and image registration". In: *Proc IEEE Workshop Var Level Set Meth Comput Vis*. 2001, pp. 29–36.

[Hub64] P. J. Huber. "Robust estimation of a location parameter". In: *Ann Math Stat* 35 (1964), pp. 73–101.

[II69] A. Ionescu Tulcea and C. Ionescu Tulcea. *Topics in the theory of lifting*. Springer, 1969.

[Ish03] H. Ishikawa. "Exact optimization for markov random fields with convex priors". In: *IEEE Trans Pattern Anal Mach Intell* 25.10 (2003), pp. 1333–1336.

[Jäh19] B. Jähne. *Digitale Bildverarbeitung und Bildgewinnung*. Springer Vieweg, 2019.

[Kan06a] L. V. Kantorovich. "On a problem of Monge". In: *J Math Sci* 133.4 (2006). English Translation of Uspekhi Mat Nauk, 3, No. 2, 225–226 (1948), pp. 1383–1383.

[Kan06b] L. V. Kantorovitch. "On the translocation of masses". In: *Comptes Rendus (Doklady) de l'Académie des Sciences de l'URSS, Nouvelle Série* 37 (2006). English Translation of Dokl Akad Nauk SSSR, 37, No. 7–8, 227–229 (1942), pp. 199–201.

[Kar77] H. Karcher. "Riemannian center of mass and mollifier smoothing". In: *Commun Pure Appl Math* 30 (1977), pp. 509–541.

[Kez+08] I. Kezele, M. Descoteaux, C. Poupon, P. Abrial, F. Poupon, and J.-F. Mangin. "Multiresolution decomposition of HARDI and ODF profiles using spherical wavelets". In: *Proc MICCAI Workshop Comput Diffus MRI 2008*. 2008, pp. 225–234.

[KK11] E. Kaden and F. Kruggel. "A Reproducing Kernel Hilbert Space Approach for Q-Ball Imaging". In: *IEEE Trans Med Imaging* 30.11 (2011), pp. 1877–1886.

[Klo+08] M. Klodt, T. Schoenemann, K. Kolev, M. Schikora, and D. Cremers. "An Experimental Comparison of Discrete and Continuous Shape Optimization Methods". In: *Proc ECCV 2008*. 2008, pp. 332–345.

[KÖU18] S. C. Karayumak, E. Özarslan, and G. Unal. "Asymmetric Orientation Distribution Functions (AODFs) revealing intravoxel geometry in diffusion MRI". In: *Magn Reson Imaging* 49 (2018), pp. 145–158.

[KR57] L. V. Kantorovich and G. S. Rubinshteĭn. "On a functional space and certain extremum problems". In: *Dokl Akad Nauk SSSR* 115 (1957), pp. 1058–1061.

[KT02] J. Kleinberg and E. Tardos. "Approximation algorithms for classification problems with pairwise relationships: Metric labeling and Markov random fields". In: *J ACM* 49.5 (2002), pp. 616–639.

[KTV10] Y. Kim, P. Thompson, and L. Vese. "HARDI data denoising using vectorial total variation and logarithmic barrier". In: *Inverse Probl Imaging* 4.2 (2010), pp. 273–310.

[Lau+16] E. Laude, T. Möllenhoff, M. Moeller, J. Lellmann, and D. Cremers. "Sublabel-accurate convex relaxation of vectorial multilabel energies". In: *Proc ECCV 2016*. Springer. 2016, pp. 614–627.

[Lav19a] H. Lavenant. "Harmonic mappings valued in the Wasserstein space". In: *J Funct Anal* 277.3 (2019), pp. 688–785.

[Lav19b] H. Lavenant. "Optimal curves and mappings valued in the Wasserstein space". PhD Thesis. Université Paris-Saclay, 2019.

[LB01] E. Levina and P. Bickel. "The Earth Mover's distance is the Mallows distance: some insights from statistics". In: *Proc ICCV 2001*. Vol. 2. IEEE, 2001, pp. 251–256.

[LBS09] J. Lellmann, F. Becker, and C. Schnorr. "Convex optimization for multi-class image labeling with a novel family of total variation based regularizers". In: *Proc ICCV 2009*. IEEE, 2009, pp. 646–653.

[Lee13] J. M. Lee. *Introduction to smooth manifolds. 2nd revised ed.* Springer, 2013.

[Lel+09] J. Lellmann, J. Kappes, J. Yuan, F. Becker, and C. Schnörr. "Convex Multi-class Image Labeling by Simplex-Constrained Total Variation". In: *Proc SSVM 2009*. Ed. by X.-C. Tai, K. Mørken, M. Lysaker, and K.-A. Lie. Springer, 2009, pp. 150–162.

[Lel+13a] J. Lellmann, B. Lellmann, F. Widmann, and C. Schnörr. "Discrete and Continuous Models for Partitioning Problems". In: *Int J Comput Vis* 104.3 (2013), pp. 241–269.

[Lel+13b] J. Lellmann, E. Strekalovskiy, S. Koetter, and D. Cremers. "Total variation regularization for functions with values in a manifold". In: *Proc ICCV 2013*. IEEE, 2013, pp. 2944–2951.

[Lel+14] J. Lellmann, D. A. Lorenz, C. Schönlieb, and T. Valkonen. "Imaging with Kantorovich-Rubinstein Discrepancy". In: *SIAM J Imaging Sci* 7.4 (2014), pp. 2833–2859.

[Lel11] J. Lellmann. "Nonsmooth convex variational approaches to image analysis". PhD Thesis. Ruprecht-Karls-Universität Heidelberg, 2011.

[LH17] J. L. Lancaster and B. Hasegawa. *Fundamental mathematics and physics of medical imaging*. CRC Press, 2017.

[LL18] B. Loewenhauser and J. Lellmann. "Functional Lifting for Variational Problems with Higher-Order Regularization". In: *Imaging, Vision and Learning Based on Optimization and PDEs*. Ed. by X.-C. Tai, E. Bae, and M. Lysaker. Springer, 2018, pp. 101–120.

[LLS13a] J. Lellmann, F. Lenzen, and C. Schnörr. "Optimality Bounds for a Variational Relaxation of the Image Partitioning Problem". In: *J Math Imaging Vis* 47.3 (2013), pp. 239–257.

[LLS13b] J. Lellmann, F. Lenzen, and C. Schnörr. "Optimality Bounds for a Variational Relaxation of the Image Partitioning Problem". In: *J Math Imaging Vis* 47.3 (2013), pp. 239–257.

[LM79] P.-L. Lions and B. Mercier. "Splitting algorithms for the sum of two nonlinear operators". In: *SIAM J Numer Anal* 16 (1979), pp. 964–979.

[LMM19] D. A. Lorenz, P. Manns, and C. Meyer. *Quadratically regularized optimal transport*. Tech. rep. arXiv:1903.01112. arXiv, 2019.

[LPS17] F. Laus, J. Persch, and G. Steidl. "A Nonlocal Denoising Algorithm for Manifold-Valued Images Using Second Order Statistics". In: *SIAM J Imaging Sci* 10 (2017), pp. 416–448.

[LS11] J. Lellmann and C. Schnörr. "Continuous Multiclass Labeling Approaches and Algorithms". In: *SIAM J Imaging Sci* 4.4 (2011), pp. 1049–1096.

[Maa+15] J. Maas, M. Rumpf, C. Schönlieb, and S. Simon. "A generalized model for optimal transport of images including dissipation and density modulation". In: *ESAIM Math Model Numer Anal* 49.6 (2015), pp. 1745–1769.

[Mai+17] K. H. Maier-Hein et al. "The challenge of mapping the human connectome based on diffusion tractography". In: *Nature Communications* 8.1 (2017), p. 1349.

[MC17] T. Möllenhoff and D. Cremers. "Sublabel-Accurate Discretization of Nonconvex Free-Discontinuity Problems". In: *Proc ICCV 2017*. IEEE, 2017, pp. 1183–1191.

[MC19] T. Möllenhoff and D. Cremers. "Lifting Vectorial Variational Problems: A Natural Formulation based on Geometric Measure Theory and Discrete Exterior Calculus". In: *Proc CVPR 2019*. 2019, pp. 11117–11126.

[McG+09] T. McGraw, B. Vemuri, E. özarslan, Y. Chen, and T. Mareci. "Variational denoising of diffusion weighted MRI". In: *Inverse Probl Imaging* 3.4 (2009), pp. 625–648.

[Mee+16a] S. Meesters, G. Sanguinetti, E. Garyfallidis, J. Portegies, and R. Duits. *Fast implementations of contextual PDE's for HARDI data processing in DIPY*. Tech. rep. Poster session presented at 24th ISMRM Annual Meeting and Exhibition, May 7–13, 2016, Singapore, 2016.

[Mee+16b] S. Meesters, G. Sanguinetti, E. Garyfallidis, J. Portegies, P. Ossenblok, and R. Duits. *Cleaning output of tractography via fiber to bundle coherence, a new open source implementation*. Tech. rep. Poster session presented at Organization for Human Brain Mapping Annual Meeting, 26–30 June 2016, Geneve, Switzerland, 2016.

[MF98] D. Massonnet and K. L. Feigl. "Radar interferometry and its application to changes in the Earth's surface". In: *Rev Geophys* 36.4 (1998), pp. 441–500.

[Mic56] E. Michael. "Continuous Selections. I". In: *Ann Math* 63.2 (1956), pp. 361–382.

[Mir03] M. Miranda. "Functions of bounded variation on "good" metric spaces". In: *J Math Pures Appl* 82.8 (2003), pp. 975–1004.

[MM77] L. Modica and S. Mortola. "Un esempio di Γ-convergenza". In: *Boll Unione Mat Ital. Ser V. B* 14 (1977), pp. 285–299.

[Mod04] J. Modersitzki. *Numerical methods for image registration*. Oxford University Press, 2004.

[Mod09] J. Modersitzki. *FAIR. Flexible algorithms for image registration*. Vol. 6. SIAM, 2009.

[Möl+16] T. Möllenhoff, E. Laude, M. Moeller, J. Lellmann, and D. Cremers. "Sublabel-accurate relaxation of nonconvex energies". In: *Proc CVPR 2016*. 2016, pp. 3948–3956.

[Mon84] G. Monge. "Mémoire sur la théorie des déblais et des remblais". In: *Histoire de l'Académie Royale des Sciences Avec les Mémoires de Mathématique & de Physique pour la même Année. Année 1781*. Ed. by Académie des sciences. L'Imprimerie Royale, 1784, pp. 666–704.

[Mor02] M. G. Mora. "The Calibration Method for Free-Discontinuity Problems on Vector-Valued Maps". In: *J Convex Anal* 9.1 (2002), pp. 1–29.

[Mor90] F. Morgan. "Calibrations and new singularities in area-minimizing surfaces: A survey". In: *Variational methods. Proc Conf Prog Nonlinear Differ Equ Appl 4, Paris/Fr 1988*. 1990, pp. 329–342.

[MP18] Y. Malitsky and T. Pock. "A First-Order Primal-Dual Algorithm with Linesearch". In: *SIAM J Optim* 28.1 (2018), pp. 411–432.

[MP29] R. von Mises and H. Pollaczek-Geiringer. "Praktische Verfahren der Gleichungsauflösung. I, II". In: *Z f angew Math* 9 (1929), pp. 58–77, 152–164.

[MR10] O. V. Michailovich and Y. Rathi. "On approximation of orientation distributions by means of spherical ridgelets". In: *IEEE Trans Image Process* 19.2 (2010), pp. 461–477.

[MS13] P. MomayyezSiahkal and K. Siddiqi. "3D Stochastic Completion Fields for Mapping Connectivity in Diffusion MRI". In: *IEEE Trans Pattern Anal Mach Intell* 35.4 (2013), pp. 983–995.

[MS89] D. Mumford and J. Shah. "Optimal approximations by piecewise smooth functions and associated variational problems". In: *Commun Pure Appl Math* 42.5 (1989), pp. 577–685.

[MTY15] M. I. Miller, A. Trouvé, and L. Younes. "Hamiltonian Systems and Optimal Control in Computational Anatomy: 100 Years Since D'Arcy Thompson". In: *Annu Rev Biomed Eng* 17.1 (2015), pp. 447–509.

[Mül99] S. Müller. "Variational models for microstructure and phase transitions". In: *Calculus of Variations and Geometric Evolution Problems*. Ed. by S. Hildebrandt and M. Struwe. Springer, 1999, pp. 85–210.

[Ncu11] S. Ncube. "A Novel Riemannian Metric For Analyzing Spherical Functions With Applications To Hardi Data". PhD Thesis. Florida State University, 2011.

[Nes04] Y. Nesterov. *Introductory lectures on convex optimization. A basic course*. Kluwer Academic Publishers, 2004.

[OCW14] Y. Ouyang, Y. Chen, and Y. Wu. "Vectorial total variation regularisation of orientation distribution functions in diffusion weighted MRI". In: *Int J Bioinform Res Appl* 10.1 (2014), pp. 110–127.

[OS88] S. Osher and J. A. Sethian. "Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations". In: *J Comput Phys* 79.1 (1988), pp. 12–49.

[PC11] T. Pock and A. Chambolle. "Diagonal preconditioning for first order primal-dual algorithms in convex optimization". In: *Proc ICCV 2011*. IEEE, 2011, pp. 1762–1769.

[PC19] G. Peyré and M. Cuturi. "Computational Optimal Transport". In: *Found Trends Mach Learn* 11.5–6 (2019), pp. 355–206.

[PD17] J. M. Portegies and R. Duits. "New exact and numerical solutions of the (convection-)diffusion kernels on SE(3)". In: *Differ Geom Appl* 53 (2017), pp. 182–219.

[Ped97] P. Pedregal. *Parametrized measures and variational principles*. Birkhäuser, 1997.

[Per18] A.-P. Perkkiö. "Conjugates of integral functionals on continuous functions". In: *J Math Anal Appl* 459.1 (2018), pp. 124–134.

[Pey+09] G. Peyré, M. Péchaud, R. Keriven, and L. D. Cohen. "Geodesic methods in computer vision and graphics". In: *Found Trends Comp Graphics and Vision* 5.3–4 (2009), pp. 197–397.

[Phi62] D. L. Phillips. "A Technique for the Numerical Solution of Certain Integral Equations of the First Kind". In: *J ACM* 9.1 (1962), pp. 84–97.

[Pin14] J.-C. Pinoli. *Mathematical foundations of image processing and analysis. volume 1 and 2*. John Wiley & Sons, 2014.

[Poc+08] T. Pock, T. Schoenemann, G. Graber, H. Bischof, and D. Cremers. "A Convex Formulation of Continuous Multi-label Problems". In: *Proc ECCV 2008*. 2008, pp. 792–805.

[Poc+09a] T. Pock, A. Chambolle, D. Cremers, and H. Bischof. "A Convex Relaxation Approach for Computing Minimal Partitions". In: *Proc CVPR 2009*. 2009, pp. 810–817.

[Poc+09b] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. "An algorithm for minimizing the Mumford-Shah functional". In: *Proc ICCV 2009*. IEEE, 2009, pp. 1133–1140.

[Poc+10] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. "Global Solutions of Variational Models with Convex Regularization". In: *SIAM J Imaging Sci* 3.4 (2010), pp. 1122–1145.

[Pol18] T. Polzin. "Large Deformation Diffeomorphic Metric Mappings - Theory, Numerics, and Applications". PhD Thesis. University of Lübeck, 2018.

[Por+15] J. M. Portegies, R. H. J. Fick, G. R. Sanguinetti, S. P. L. Meesters, G. Girard, and R. Duits. "Improving Fiber Alignment in HARDI by Combining Contextual PDE Flow with Constrained Spherical Deconvolution". In: *PLOS ONE* 10.10 (2015). Ed. by A. Leemans, Article ID e0138122.

[PR00] T. Preußer and M. Rumpf. "An Adaptive Finite Element Method for Large Scale Image Processing". In: *J Vis Commun Image Represent* 11.2 (2000), pp. 183–195.

[Prč+15] V. Prčkovska, M. Andorrà, P. Villoslada, E. Martinez-Heras, R. Duits, D. Fortin, P. Rodrigues, and M. Descoteaux. "Contextual Diffusion Image Post-processing Aids Clinical Applications". In: *Visualization and Processing of Higher Order Descriptors for Multi-Valued Data*. Springer, 2015, pp. 353–377.

[Rab+12] J. Rabin, G. Peyré, J. Delon, and M. Bernot. "Wasserstein Barycenter and Its Application to Texture Mixing". In: *Proc SSVM 2011*. Ed. by A. M. Bruckstein, B. M. ter Haar Romeny, A. M. Bronstein, and M. M. Bronstein. Springer, 2012, pp. 435–446.

[Rin18] F. Rindler. *Calculus of variations*. Springer, 2018.

[RKK12] M. Reisert, E. Kellner, and V. Kiselev. "About the geometry of asymmetric fiber orientation distributions". In: *IEEE Trans Med Imaging* 31.6 (2012), pp. 1240–1249.

[Roc68] R. T. Rockafellar. "Integrals which are convex functionals". In: *Pac J Math* 24 (1968), pp. 525–539.

[Roc71] R. T. Rockafellar. "Integrals which are convex functionals. II". In: *Pac J Math* 39 (1971), pp. 439–469.

[Roc97] R. T. Rockafellar. *Convex analysis*. Princeton University Press, 1997.

[ROF92] L. I. Rudin, S. Osher, and E. Fatemi. "Nonlinear total variation based noise removal algorithms". In: *Physica D* 60.1 (1992), pp. 259–268.

[Rok+] A. Rokem, J. Yeatman, F. Pestilli, and B. Wandell. *High angular resolution diffusion MRI*. Stanford Digital Repository 2013. Available at: `http://purl.stanford.edu/yx282xq2090`. Accessed: 2017–09–20.

[Ros+12] G. Rosman, M. M. Bronstein, A. M. Bronstein, A. Wolf, and R. Kimmel. "Group-Valued Regularization Framework for Motion Segmentation of Dynamic Non-rigid Shapes". In: *Proc SSVM 2011*. Ed. by A. M. Bruckstein, B. M. ter Haar Romeny, A. M. Bronstein, and M. M. Bronstein. Springer, 2012, pp. 725–736.

[RPB13] R. Ranftl, T. Pock, and H. Bischof. "Minimizing TGV-Based Variational Models with Non-convex Data Terms". In: *Proc SSVM 2013*. Ed. by A. Kuijper, K. Bredies, T. Pock, and H. Bischof. Springer, 2013, pp. 282–293.

[RPF97] F. Rocca, C. Prati, and A. Ferretti. "An overview of SAR interferometry". In: *Proc 3rd ERS Symp Spac Serv Env*. 1997.

[RS13] M. Reisert and H. Skibbe. "Fiber Continuity Based Spherical Deconvolution in Spherical Harmonic Domain". In: *Proc MICCAI 2013*. Ed. by K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab. Springer, 2013, pp. 493–500.

[RTG00] Y. Rubner, C. Tomasi, and L. J. Guibas. "The Earth Mover's Distance as a Metric for Image Retrieval". In: *Int J Comput Vis* 40.2 (2000), pp. 99–121.

[RV10] K. Rohr and M. A. Viergever. *Landmark-Based Image Analysis*. Springer, 2010.

[RV73] A. W. Roberts and D. E. Varberg. *Convex functions*. Academic Press, 1973.

[RW04] R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*. Springer, 2004.

[San15] F. Santambrogio. *Optimal transport for applied mathematicians. Calculus of variations, PDEs, and modeling*. Springer, 2015.

[SC11] E. Strekalovskiy and D. Cremers. "Total variation for cyclic structures: Convex relaxation and efficient minimization". In: *Proc CVPR 2011*. 2011, pp. 1905–1911.

[SCC12] E. Strekalovskiy, A. Chambolle, and D. Cremers. "A convex representation for the vectorial Mumford-Shah functional". In: *Proc CVPR 2012*. IEEE. 2012, pp. 1712–1719.

[SCC14] E. Strekalovskiy, A. Chambolle, and D. Cremers. "Convex relaxation of vectorial problems with coupled regularization". In: *SIAM J Imaging Sci* 7.1 (2014), pp. 294–336.

[Sch+09] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational methods in imaging*. Springer, 2009.

[Sch+19] K. G. Schilling et al. "Limits to anatomical accuracy of diffusion tractography using modern approaches". In: *NeuroImage* 185 (2019), pp. 1–11.

[Sch15] O. Scherzer, ed. *Handbook of mathematical methods in imaging. In 3 volumes. 2nd edition*. Springer, 2015.

[SG19] M. Strecke and B. Goldluecke. "Sublabel-Accurate Convex Relaxation with Total Generalized Variation Regularization". In: *Proc GCPR 2018*. Ed. by T. Brox, A. Bruhn, and M. Fritz. Springer, 2019, pp. 263–277.

[SGC11] E. Strekalovskiy, B. Goldlücke, and D. Cremers. "Tight convex relaxations for vector-valued labeling problems". In: *Proc ICCV 2011* (2011), pp. 2328–2335.

[SJJ07] A. Srivastava, I. H. Jermyn, and S. H. Joshi. "Riemannian analysis of probability density functions with applications in vision". In: *Proc CVPR 2007*. 2007, pp. 1–8.

[SNC12] E. Strekalovskiy, C. Nieuwenhuis, and D. Cremers. "Nonmetric Priors for Continuous Multilabel Optimization". In: *Proc ECCV 2012*. Ed. by A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid. Springer, 2012, pp. 208–221.

[Sol+14] J. Solomon, R. Rustamov, L. Guibas, and A. Butscher. "Wasserstein propagation for semi-supervised learning". In: *Proc ICML 2014*. 2014, pp. 306–314.

[SR17] H. Skibbe and M. Reisert. "Spherical Tensor Algebra: A Toolkit for 3D Image Processing". In: *J Math Imaging Vis* 58.3 (2017), pp. 349–381.

[SSW19] B. Schmitzer, K. P. Schäfers, and B. Wirth. *Dynamic Cell Imaging in PET with Optimal Transport Regularization*. Tech. rep. arXiv:1902.07521. arXiv, 2019.

[ST65] E. Stejskal and J. Tanner. "Spin diffusion measurements: spin echos in the presence of a time-dependent field gradient". In: *J Chem Phys* 42 (1965), pp. 288–292.

[Ste73] G. W. Stewart. *Introduction to matrix computations*. Academic Press, 1973.

[Str08] M. Struwe. *Variational methods: applications to nonlinear partial differential equations and Hamiltonian systems*. Springer, 2008.

[Str15] E. Strekalovskiy. "Convex Relaxation of Variational Models with Applications in Image Analysis". PhD Thesis. Technische Universität München, 2015.

[Sun17] D. Sundararajan. *Digital image processing. A signal processing and algorithmic approach*. Springer, 2017.

[SW18] M. Storath and A. Weinmann. *Wavelet Sparse Regularization for Manifold-Valued Data*. Tech. rep. arXiv:1808.00505. arXiv, 2018.

[TA77] A. N. Tikhonov and V. Y. Arsenin. *Solutions of ill-posed problems. Translation editor Fritz John*. John Wiley & Sons, 1977.

[Tax+14] C. M. W. Tax, B. Jeurissen, S. B. Vos, M. A. Viergever, and A. Leemans. "Recursive calibration of the fiber response function for spherical deconvolution of diffusion MRI data". In: *NeuroImage* 86 (2014), pp. 67–80.

[TCC07] J. D. Tournier, F. Calamante, and A. Connelly. "Robust determination of the fibre orientation distribution in diffusion MRI: Non-negativity constrained super-resolved spherical deconvolution". In: *NeuroImage* 35.4 (2007), pp. 1459–1472.

[Tik63] A. N. Tikhonov. "Solution of incorrectly formulated problems and the regularization method". In: *Sov Math Dokl* 5 (1963). English Translation of Dokl Akad Nauk SSSR 151, 501–504 (1963), pp. 1035–1038.

[Tou+04] J. D. Tournier, F. Calamante, D. Gadian, and A. Connelly. "Direct estimation of the fibre orientation density function from diffusion-weighted MRI data using spherical deconvolution". In: *NeuroImage* 23.3 (2004), pp. 1176–1185.

[TPG16] G. Tartavel, G. Peyré, and Y. Gousseau. "Wasserstein loss for image synthesis and restoration". In: *SIAM J Imaging Sci* 9.4 (2016), pp. 1726–1755.

[Tuc+02] D. S. Tuch, T. G. Reese, M. R. Wiegell, N. Makris, J. W. Belliveau, and V. J. Wedeen. "High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity". In: *Magn Reson Med* 48.4 (2002), pp. 577–582.

[Tuc04] D. S. Tuch. "Q-ball imaging". In: *Magn Reson Med* 52.6 (2004), pp. 1358–1372.

[Val14] T. Valkonen. "A primal–dual hybrid gradient method for nonlinear operators with applications to MRI". In: *Inverse Prob* 30.5 (2014), p. 055012.

[Vas69] L. N. Vaserstein. "Markovian processes on countable space product describing large systems of automata". In: *Probl Peredaci Inform* 5.3 (1969), pp. 64–72.

[VFG09] L. Velho, A. C. Frery, and J. d. M. Gomes. *Image processing for computer graphics and vision. 2nd ed.* Springer, 2009.

[VG16] L. A. Vese and C. L. Guyader. *Variational methods in image processing.* CRC Press, 2016.

[Vil09] C. Villani. *Optimal transport. Old and new.* Springer, 2009.

[VL17] T. Vogt and J. Lellmann. "An Optimal Transport-Based Restoration Method for Q-Ball Imaging". In: *Proc SSVM 2017.* Ed. by F. Lauze, Y. Dong, and A. B. Dahl. Springer, 2017, pp. 271–282.

[VL18] T. Vogt and J. Lellmann. "Measure-Valued Variational Models with Applications to Diffusion-Weighted Imaging". In: *J Math Imaging Vis* 60.9 (2018), pp. 1482–1502.

[VL19] T. Vogt and J. Lellmann. "Functional Liftings of Vectorial Variational Problems with Laplacian Regularization". In: *Proc SSVM 2019*. Ed. by J. Lellmann, M. Burger, and J. Modersitzki. Springer, 2019, pp. 559–571.

[Vog+19] T. Vogt, E. Strekalovskiy, D. Cremers, and J. Lellmann. "Lifting methods for manifold-valued variational problems". In: *Variational Methods for Nonlinear Geometric Data and Applications*. Ed. by P. Grohs, M. Holler, and A. Weinmann. Springer, 2019, In press.

[WC16] T. Windheuser and D. Cremers. "A Convex Solution to Spatially-Regularized Correspondence Problems". In: *Proc ECCV 2016*. Ed. by B. Leibe, J. Matas, N. Sebe, and M. Welling. Springer, 2016, pp. 853–868.

[WDS14] A. Weinmann, L. Demaret, and M. Storath. "Total variation regularization for manifold-valued data". In: *SIAM J Imaging Sci* 7.4 (2014), pp. 2226–2257.

[WDS16] A. Weinmann, L. Demaret, and M. Storath. "Mumford–Shah and Potts Regularization for Manifold-Valued Data". In: *J Math Imaging Vis* 55.3 (2016), pp. 428–445.

[Wea99] N. Weaver. *Lipschitz algebras*. World Scientific, 1999.

[Wer+09] M. Werlberger, W. Trobin, T. Pock, A. Wedel, D. Cremers, and H. Bischof. "Anisotropic Huber-L^1 Optical Flow". In: *Proc BMVC 2009*. British Machine Vision Association, 2009, pp. 108.1–108.11.

[Wil70] S. Willard. *General topology*. Addison-Wesley, 1970.

[WJ08] M. J. Wainwright and M. I. Jordan. "Graphical Models, Exponential Families, and Variational Inference". In: *Found Trends Mach Learn* 1.1–2 (2008), pp. 1–305.

[Yan+17] X. Yang, R. Kwitt, M. Styner, and M. Niethammer. "Quicksilver: Fast predictive image registration – A deep learning approach". In: *NeuroImage* 158 (2017), pp. 378–396.

[You10] L. Younes. *Shapes and diffeomorphisms*. Springer, 2010.

[Zac+08] C. Zach, D. Gallup, J.-M. Frahm, and M. Niethammer. "Fast Global Labeling for Real-Time Stereo Using Multiple Plane Sweeps". In: *Proc Workshop VMV 2008*. Ed. by O. Deussen, D. Keim, and D. Saupe. AKA, 2008, pp. 243–252.

[ZC08] M. Zhu and T. Chan. *An Efficient Primal-Dual Hybrid Gradient Algorithm For Total Variation Image Restoration*. Tech. rep. 08–34. UCLA CAM, 2008.

[ZK12] C. Zach and P. Kohli. "A Convex Discrete-Continuous Approach for Markov Random Fields". In: *Proc ECCV 2012*. Ed. by A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid. Springer, 2012, pp. 386–399.