Variational Methods for the Reconstruction of Signals with Discontinuities

Andreas Weinmann

Habilitationsschrift

Universität Osnabrück, Fakultät für Mathematik

Osnabrück, Januar 2018

List of Included Articles

[1] A. Weinmann, M. Storath, and L. Demaret. The L^1-Potts functional for robust jump-sparse reconstruction. *SIAM Journal on Numerical Analysis*, 53(1):644–673, 2015.

[2] M. Storath and A. Weinmann. Fast partitioning of vector-valued images. *SIAM Journal on Imaging Sciences*, 7(3):1826–1852, 2014.

[3] A. Weinmann and M. Storath. Iterative Potts and Blake-Zisserman minimization for the recovery of functions with discontinuities from indirect measurements. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 471(2176), 2015.

[4] M. Storath, A Weinmann, and L. Demaret. Jump-sparse and sparse recovery using Potts functionals. *IEEE Transactions on Signal Processing*, 62(14):3654–3666, 2014.

[5] A. Weinmann, L. Demaret, and M. Storath. Mumford–Shah and Potts regularization for manifold-valued data. *Journal of Mathematical Imaging and Vision*, 55(3):428–445, 2016.

[6] A. Weinmann, L. Demaret, and M. Storath. Total variation regularization for manifold-valued data. *SIAM Journal on Imaging Sciences*, 7(4):2226–2257, 2014.

[7] M. Baust, L. Demaret, M. Storath, N. Navab, and A. Weinmann. Total variation regularization of shape signals. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2075–2083, 2015.

[8] R. Bergmann, F. Laus, G. Steidl, and A. Weinmann. Second order differences of cyclic data and applications in variational denoising. *SIAM Journal on Imaging Sciences*, 7(4):2916–2953, 2014.

[9] M. Bačák, R. Bergmann, G. Steidl, and A. Weinmann. A second order nonsmooth variational model for restoring manifold-valued images. *SIAM Journal on Scientific Computing*, 38(1):A567–A597, 2016.

Contents

Chapter 1

Preface

This thesis is mainly located in the area of applied analysis. In particular, it deals with variational methods for direct and inverse problems involving functions with discontinuities. There are tight connections with numerical analysis and with iterative optimization schemes. Since we also consider manifold-valued data we base on concepts from differential geometry as well. Further, there is a (formal) link to maximum a posteriori estimators in Bayesian statistics. Application areas are signal and image processing, imaging, as well as data analysis.

Functions having discontinuities appear when modeling real-life problems in various branches of applied science, for instance, in connection with crack formation in mechanics or as earth mantle layers in geophysical exploration. In biology and medicine, on which the applications in this thesis are primarily focusing, discontinuities appear in different forms in microarray data [SNS$^+$01], tomography [RR07], single-molecule fluorescence resonance energy transfer [JBI$^+$08], as well as microscopy [ZLMY$^+$02]. Another very concrete example is the rotation of the bacterial flagella motor [SRL$^+$05]. The latter data represent angles and are thus not living in a linear space, but on the cricle. Circle valued data also appear in SAR imaging [MF98]. Further nonlinear data spaces appear when considering non-flat models for color image processing [CKS01]. In medical imaging, a prominent example with manifold-valued data is diffusion tensor imaging [BML94].

The discontinuities often represent highly significant information; for instance, they determine cell or tissue boundaries, or boundary layers between nerve fibers. Hence, it is important to capture this information from the measurements which are often corrupted by noise. Classical regularization techniques such as Tikhonov regularization do not respect the discontinuities of the underlying signals, and we lose significant information when applying them. Further, classical techniques typically do not apply to data living in nonlinear spaces.

Modern regularization approaches aim at preserving discontinuities. One direction of research aims at finding suitable sparse representation systems such as

wavelets [DeV98]. For multivariate functions, shearlets [LLKW05] and curvelets [CD04], collectively known as parabolic molecules [GK14], are suitable systems. Another approach to discontinuity preserving regularization using sparse expansions are wedgelets, introduced in [Don99].

We here primarily follow the approach via free-discontinuity problems. Free-discontinuity problems describe situations where the solution of interest is defined by a function and a lower-dimensional set consisting of the discontinuities of the function [DG91, FW10]. One of the historically first free-discontinuity problem is given by the Potts model [Pot52, GG84] or piecewise-constant Mumford-Shah model [MS85, MS89]. Here the task is to minimize an energy functional which consists of a data term and a regularizing or prior term. The data term penalizes the distance to the data, and the prior term penalizes the size of the discontinuity set, i.e., its Hausdorff measure. In the Potts model, the underlying target functions u are assumed to change only at the discontinuity set. The Mumford-Shah model [MS89] additionally penalizes the variation on the complement of the discontinuity set which promotes piecewise smooth minimizers. The target variables are functions of bounded variation where the corresponding signed measure has no Cantor part. From an algorithmic point of view, the considered free-discontinuity problems are challenging nonsmooth and nonconvex problems. A related variational regularization technique is total variation (TV) regularization. Here the regularizer penalizes the total variation of the candidate function. Whereas the TV problem is convex for linear space data, it is nonconvex for manifold-valued data in general.

This thesis provides several contributions to variational regularization based on free-discontinuity problems and to TV regularization. We consider vector space and manifold-valued data. The central focus is on the development of computationally feasible algorithms. We further provide theoretical foundations as well as application to concrete problems to show the potential of the derived methods. The articles comprising this thesis can be grouped as follows:

- Free-discontinuity methods for directly measured vector space data [1, 2];
- Free-discontinuity methods for indirectly measured vector space data [3, 4];
- Free-discontinuity methods for manifold-valued data [5];
- TV regularization for manifold-valued data [6, 7];
- Higher order TV type methods for manifold-valued data [8, 9].

The thesis consist of a summarizing chapter as well as the included articles.

Acknowledgments. I would like to thank the "Fast Algorithms for Biomedical Imaging" Group at Helmholtz Center Munich and, in particular, Stefan Kunis for providing an excellent working environment. I thank the members of the Institute of Computational Biology, respectively, the Institute of Biomathematics and Biometry, at Helmholtz Center Munich, and the members of M12 at Technical University Munich for the nice atmosphere. Further, I would like to thank all my coauthors for the fruitful cooperations. In particular, I would like to thank Martin Storath for an excellent collaboration over the years.

Furthermore I would like to thank Wolfgang Erb, Massimo Fornasier, Philipp Grohs, Rupert Lasser, Volkmar Liebscher, Tom März, Josef Obermaier, Ulrich Reif, Konrad Sandau, Ruben Seyfried, Fabian Theis, and Johannes Wallner. This list is by no means complete, and I apologize to all those which I forgot to mention.

I acknowledge support within the young investigator group VH-NG-526 of the Helmholtz Association, by the DFG Grant WE 5886/3-1 and by DFG Young Researchers' Network "Mathematics for Magnetic Particle Imaging".

Munich, August, 8, 2016 Andreas Weinmann

Chapter 2

Summary

We here summarize the content of the publications comprising this thesis. The chapter is organized as follows. We start out by briefly giving motivation for the considered problems and by reviewing the state of the art in Chapter 2.1. Here, we consider data living in linear space and in manifolds as well as the direct and indirect measurement case. In the rest of the chapter, we summarize the contributions of this thesis. The contributions concerning free-discontinuity problems for directly measured data in a linear space are the topic of Chapter 2.2. We deal with free-discontinuity problems in connection with inverse problems in Chapter 2.3. Then, we consider variational methods for manifold-valued data. The topic of Chapter 2.4 are free-discontinuity problems for manifold-valued data. In Chapter 2.5, we consider total variation minimization for manifold-valued data as well as higher order generalizations.

2.1 Motivation and State of the Art

Let us start to motivate the use of edge-preserving variational methods and why it is important to consider nonlinear data spaces in this context.

Edge-preserving variational methods try to restore the discontinuities of the signals or functions from the measured data. This is important because we encounter such functions with discontinuities almost everywhere in our environment and, in particular, in various branches of applied science. For example, they appear in connection with boundary layers in geophysics. In the biomedical context, on which we are concentrating here, functions with discontinuities appear in tomography [RR07], in microscopy [ZLMY$^+$02], and in connection with microarray data [SNS$^+$01, DMH$^+$03, HST$^+$04] to mention some examples. For further examples, we refer to the papers [LJ11a, LJ11b, FMS14] and the references therein.

A more specific, but prominent example of data representing functions with discontinuities is the time evolution of the rotation angle of the bacterial flagella motor [SRL$^+$05, SB08]. As angular data, these data live on the unit circle. Circle valued data also appear in SAR imaging [MF98] where they represent phase values. These examples show that frequently the data does not live in a linear, but in a nonlinear space. Further examples of nonlinear data spaces appear in connection with non-flat model spaces for color image processing such as HSV, HSI as well as chromaticity-based spaces; see, e.g., [CKS01, VO02]. Another data space important for application is the Euclidean motion group which appears for instance in the context of registration problems; see, e.g., [URDDS05, RBB$^+$12].

In medical imaging, a very prominent example with manifold-valued data is diffusion tensor imaging (DTI). To give the reader a more concrete impression, let us briefly discuss DTI in more detail here. DTI allows to quantify diffusional characteristics non-invasively [BML94, AP08, JBB09]. DTI is helpful in the context of neurodegenerative pathologies such as, e.g., schizophrenia [FMC$^+$00, KMW$^+$07], autism [ALL$^+$07] or Huntington's disease [RLB$^+$10]. In DTI, the data can be viewed as living in a Riemannian manifold, namely the manifold of positive (definite) matrices; see, e.g., [PFA06]. The underlying distance corresponds to the Fisher-Rao metric [Rao45] which is statistically motivated since the positive matrices (which are called diffusion tensors) represent covariance matrices. The diffusion tensors model the diffusivity of water molecules. The discontinuities represent the boundaries between different kind of tissue.

As the mentioned examples show, data representing functions with discontinuities appear frequently in applications, and these data may be manifold-valued. Typically, the discontinuities encode most significant information such as the boundaries between different nerve fiber bundles in DTI. Therefore, it is important to reliably extract this information from the data. Unfortunately, classical regularization techniques, such as Tikhonov regularization based on Sobolev seminorms, do not capture the discontinuities of the underlying signals; they treat the objectives as smooth functions, and we lose most significant information. Modern regularization approaches try to preserve discontinuities.

Regularization based on free-discontinuity problems as well as on total variation models and corresponding higher order generalizations are examples of such modern regularization approaches. Since these approaches are the topic of this thesis, we start out discussing them in detail first. Further approaches are discussed later on.

Free-discontinuity problems describe situations where the solution of interest is defined by a function and a lower-dimensional set describing the discontinuities of the function [DG91, FW10]. One of the earliest free-discontinuity problems is the *Potts model.* It has its roots in the work of R. B. Potts on statistical mechanics [Pot52] in 1952. It has been first used in imaging by Geman and Geman [GG84].

(a) Noisy data f. (b) Piecewise-constant u. (c) Discontinuity set C.

Figure 2.1: Segmentation of a noisy color image based on the Potts model using the method developed in [2].

The Potts problem is given by

$$\underset{u,C}{\text{minimize}}\ \gamma\text{length}(C)+E(u,f). \tag{2.1}$$

Here, we optimize w.r.t. the piecewise-constant function u where we denote its discontinuity set by $C \subset \Omega \subset \mathbb{R}^2$; Ω is the domain of the image u. In the described bivariate situation, the length is the one-dimensional Hausdorff measure which agrees with the intuitive length for smooth boundaries; in the univariate situation, length(C) corresponds to the number of jumps of u, where C is the jump set of u. In a general multivariate setup $\Omega \subset \mathbb{R}^L$, the penalty is given by the $(L-1)$-dimensional Hausdorff measure of C. The parameter $\gamma > 0$ controls the trade-off between the boundary length penalty and the data fidelity given by the functional E. The data fidelity term E is determined by the data acquisition modality. It measures how good u explains the acquired data f. In the simplest case, E is the squared L^2 distance $E(u,f) = \|u-f\|_2^2$ and a minimizer u is a piecewise constant regularization of f. Its discontinuity set C yields a partitioning of the image which makes image segmentation a typical application [GG84, BVZ01]; for an illustration, see Figure 2.1. Another important class of free-discontinuity problems is given by the *Mumford-Shah models.* They are of the form

$$\underset{u,C}{\text{minimize}}\ \gamma\text{length}(C)+\alpha\int_{\Omega\setminus C}|\nabla u(x)|^q\,dx+E(u,f). \tag{2.2}$$

Instead of assuming u to be piecewise constant, we now penalize its L^q-variation on the complement of the discontinuity set C. This promotes piecewise smooth minimizers. Here, the parameter $\alpha > 0$ controls the influence of the variation penalty. In particular, the Potts model corresponds to the degenerate case $\alpha \to \infty$. (This might explain why the Potts model is also known as piecewise-constant Mumford-Shah model.) We notice that the total variation model corresponds to $\gamma \to \infty$ and $q = 1$. As for the Potts model, the set C is interpreted as the discontinuity set of the image u. It can serve as a basis for segmentation as well. The

Mumford-Shah model has been studied in the seminal work of Mumford and Shah [MS89]; its discrete variant has been investigated by Blake and Zisserman [BZ87].

From an algorithmic point of view, the considered free-discontinuity problems are very challenging. The corresponding discretized multivariate versions are NP-hard; cf. [Vek99, AW10] for related statements. Due to their importance in data processing, various strategies have been proposed. The discussion of the state-of-the-art algorithms for directly measured data, indirectly measured data, and manifold-valued data are central topics of Chapter 2.1.1, Chapter 2.1.2, and Chapter 2.1.3, respectively. A central focus of this thesis is the design of new algorithmic strategies for these classes of problems.

Instead of penalizing the length of the discontinuity set C and the L^q-variation on its complement in (2.2), total variation (TV) regularization penalizes the total variation $TV(u)$ of the candidate u. For differentiable functions u, a version of the total variation can be defined by $TV(u) = \int_\Omega |\nabla u| d\lambda$. Then, the TV functional is given by

$$\underset{u}{\text{minimize}} \ \alpha \mathrm{TV}(u) + E(u, f). \tag{2.3}$$

As above, the parameter α regulates the trade-off between the data fidelity $E(u, f)$ and the regularity which is measured with respect to the total variation here. The natural underlying Banach space for the TV problem is the space BV of functions of bounded variation, i.e., the space of functions whose derivatives (in the sense of distributions) are regular finite signed vector-valued measures. The total variation defines a seminorm on BV. We note that there are functions with discontinuities/jumps which are contained in BV. In contrast to the Mumford-Shah penalty, the TV penalty measures the jump height as well. TV regularization was first introduced in the early 1990ies by Rudin, Osher and Fatemi [ROF92]. It has been used in many applications, e.g., [DBFZ$^+$06, ZPB07, AS12], and various theoretical properties have been shown, e.g., [CL97, PCXD99].

A lot of different algorithms for TV minimization have been proposed since the early 1990ies, e.g., [ROF92, Cha04, NNZC08, GO09, YZY10, CJK10, CP11a]. TV regularization for vector and matrix-valued data were, for instance, considered in [BC98, RWT$^+$14, LO14]. In contrast to linear space data, the investigation of manifold-valued data has started only recently. Related theoretical work is [GM06, GM07]; first algorithmic approaches are [SC11, LSKC13] and [6] which is part of the present thesis. A focus of this thesis are algorithms for total variation regularization for manifold-valued data as well as second order TV type generalizations for manifold-valued data. A more detailed discussion including the state of the art is given in Chapter 2.1.4.

The recovery of functions with discontinuities from incomplete linear measurements can also be performed assuming as a regularizing prior its nearly-sparse expansion with respect to a prescribed dictionary of functions. It is, for instance,

well known that one-dimensional piecewise smooth functions exhibit sparse expansions with respect to wavelets [DeV98]. For multivariate functions, shearlets [LLKW05] and curvelets [CD04], collectively known as parabolic molecules [GK14], play the same role. These approaches complement the ones mentioned in the previous paragraphs as they employ the molecules to model the edges (in an a priori semidiscrete way) whereas the previous employ derivatives (in the generalized sense) as analysis operators. In the context of discontinuity preserving regularization, curvelets were employed for the reconstruction of frescos using iterative thresholding algorithms [FR08a, FR08b]. Another approach to discontinuity preserving regularization using sparse expansions are wedgelets, introduced in [Don99]. Here, the (weighted) number of wedges related to a quadtree decomposition together with the complexity of this decomposition is penalized.

2.1.1 Free-Discontinuity Problems for Directly Measured Data

In case of direct measurements, the data f are noisy versions of the underlying signal or image. Hence, a natural data fidelity is given by

$$E(u,f) = \|u-f\|_p^p = \int_\Omega |f(x)-u(x)|^p \, dx \tag{2.4}$$

with $p \geq 1$. Here, a Gaussian noise model corresponds to the squared L^2 distance $(p=2)$ which is the most classical setup and which has been extensively studied [MS89]. For Laplacian noise, or if there are severe outliers, the L^1 data fidelity term is typically a better choice. The symbol $\Omega \subset \mathbb{R}^L$ denotes the domain of the data f and f can be vector-valued meaning that $f(x)$ lives in $\mathbb{R}^K$; for instance, $K=3$ for color images, $K=10$ to $K=40$ for multispectral images, or $K=10$ to $K=200$ for feature images as they are used in texture segmentation.

In the classical direct setup of data terms of the form (2.4), many theoretical aspects of free-discontinuity problems are well-understood nowadays. A rich source of information in this respect is the book of Ambrosio et al. [AFP00]. To mention one relevant topic treated there, showing the existence of minimizers required sophisticated methods from geometric measure theory as approached first in [Amb89]. The Potts problem and the Mumford-Shah problem on discrete domains have been studied by Geman and Geman [GG84] and Blake and Zisserman [BZ87]. Later, Chambolle [Cha95] has shown that the discrete functionals of Blake and Zisserman converge (in sense of Γ-convergence) for increasing sampling density to an anisotropic variant of the Mumford-Shah functional. Γ-convergence to the classical Mumford-Shah functional was obtained using adaptive finite elements discretization [CDM99]. Boysen et al. [BLMW07, BKL$^+$09] have investigated the stability (in the statistical sense) of univariate Potts estimators with L^2 data fidelity. They have shown consistency which means that the

minimizers of a sequence of increasingly finer discretized noisy problems converge to a minimizer of the continuous domain noise-free Potts problem as the discretization gets finer and if the noise level does not increase too fast.

Existing computational approaches. Classical algorithmic approaches to free-discontinuity problems are simulated annealing [GG84], graduated non-convexity [BZ87], and approximations by elliptic functionals [AT90]. Current state-of-the-art methods are based on active contours, convex relaxations, and graph cuts. In the active contour approach, the discontinuity set C is parametrized (e.g, via snakes [KWT88] or, implicitly, via level sets [CV01, VC02].) One iteratively evolves the discontinuity set C and updates the target function u while C is fixed. The direct access to the discontinuity set allows to include shape priors [CRD07] or user interaction [DGTSU12]. On the flipside, they are computationally expensive on large domains. Dimensionality reduction strategies – such as the narrow band method – reduce the computational load, but come with a rather high sensitivity to the initialization of C. Furthermore, active contours require the a priori knowledge of the number of labels. In convex relaxation approaches, the original problem is replaced by a related convex problem which is then solved using convex optimization techniques [PCBC09, LKY$^+$09, SCC12]. They are less affected by certain metrization errors related to the discretization of the jump penalty, but their computational costs are very high for high-dimensional range spaces. The graph-cut based α-expansion algorithm of Boykov et al. [BVZ01] iteratively solves a series of binary segmentation problems. Each subproblem constitutes a max-flow/min-cut problem which is solved using graph cuts. Graph cuts are presently the benchmark in terms of trade-off between computational costs and quality. However, their computational costs grow exponentially in the dimension of the range space K.

For univariate data, the considered free-discontinuity problems are not NP-hard. They can be solved exactly using dynamic programming [MS89, Bla89, WL02, FKLW08]. Solvers for the univariate problems are interesting in their own right as univariate data appear for instance as time series. Furthermore, they are important as basic building block of the multivariate algorithms developed in the thesis. The state-of-the-art solver for the univariate Potts problem with L^2 data term has $O(n^2)$ time and $O(n)$ space complexity w.r.t. the length of the signal [FKLW08]; the previous state-of-the-art solver for the univariate Potts problem with L^1 data term has $O(n^2 \log n)$ time and $O(n^2)$ space complexity. The solvers proposed for the Mumford-Shah problem have cubic complexity [Bla89].

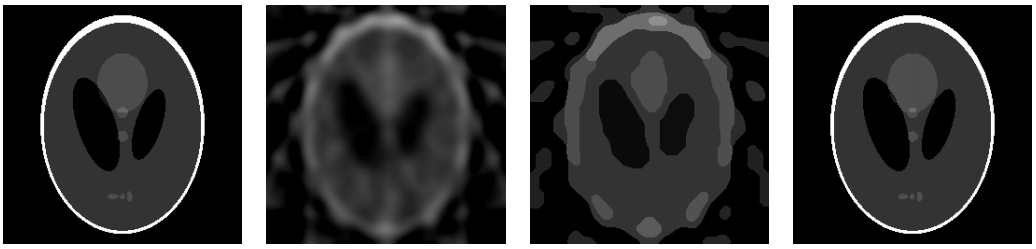
2.1.2 Free-Discontinuity Problems for Indirectly Measured Data

Often the function of interest u is only measured indirectly; that is, we have data f given as a noisy version of $H(u)$, where H is a (possibly nonlinear) measurement operator. Examples are medical imaging modalities such as computed tomography, photoacoustic tomography, and magnetic particle imaging. For segmentation in such setups, it is common to first reconstruct the image by a stabilized inversion of the operator H and then to segment the reconstructed image. In particular for incomplete or very noisy measurements, the results produced by this two-stage process are often unsatisfactory. Combined approaches can provide a significant improvement [RR07]; for a visualization, see Figure 2.2. Joint reconstruction and segmentation can be modeled by the data term

$$E(u,f) = \|H(u) - f\|_p^p = \int_{\Omega'} |H(u)(x) - f(x)|^p \, dx. \tag{2.5}$$

For such data terms, additional assumptions (which are often fulfilled in applications) are required to guarantee the existence of minimizers in a continuous setting; cf. [RS01, RR10, FW10, FMS13]. The regularizing properties in the sense of inverse problems were analyzed for the Potts problem in [RR10] and for the Mumford-Shah problem in [Ron08, JMP14]. In contrast to (2.4), the data term (2.5), in general even in the univariate setting, leads to NP-hard problems [AW10].

Existing computational approaches. In contrast to the direct measurement case, there is significantly less work dealing with algorithms for the indirect measurement variants. The earliest approaches are based on Ambrosio-Tortorelli approximations [RS01, BSK04]. Rondi and Santosa [RS01] use this approach for the Mumford-Shah and the Potts model in electrical impedance tomography; Bar et al. [BSK04] apply it for blind image deblurring. Ambrosio-Tortorelli approximations are easy to implement but computationally rather expensive. For linear inverse problems with L^2 data term, Fornasier and Ward [FW10] propose iterative thresholding algorithms and apply them to inpainting problems; see also [FMS13, AFS13]. Candès, Wakin, and Boyd [CWB08] use iteratively reweighted total variation minimization for piecewise constant recovery problems. Nikolova et al. [NNZC08, NNT10] derive algorithms of graduated non-convexity type. Kim et al. [KTCW02] propose a level-set active contour approach for deconvolution problems. Ramlau and Ring [RR07] use a level-set approach for the joint reconstruction and segmentation of x-ray tomographic images; related methods are applied to electron tomography [Kla11] and single photon emission computed tomography [KRR11]. As in the case of direct measurements, the main advantage



(a) Shepp-Logan phantom (256×256 pixel).

(b) FBP reconstruction from 7 angles, tuned w.r.t. PSNR.

(c) Segmentation of FBP result using graph cuts.

(d) Joint reconstruction and segmentation.

Figure 2.2: Reconstruction and segmentation from highly undersampled Radon data (7 projection angles). The joint reconstruction and segmentation algorithm of [SWFU15] (which bases on [2, 4]) outperforms the two-stage process (b),(c).

of level-set approaches lies in the access to the parametrization. The drawbacks are the computational costs on large domains.

In a noise free setup and under certain assumptions on the measurement operator (e.g., restricted isometry property) the minimizers of the Potts problem agree with the minimizers of the (convex) total variation model; this is a central result of compressed sensing [Don06, CRT06, NW13]. However, in general, and in particular for noisy data, the results of the Potts model and total variation minimization are often different; for a discussion see [Cha09].

2.1.3 Free-Discontinuity Problems for Manifold-Valued Data

When the data live in a (Riemannian) manifold, there is no norm available to devise a data term. Instead, we may use the distance d induced by the Riemannian metric to obtain

$$E(u,f) = \int_\Omega d^p(u(x), f(x))\, dx.$$

In the context of manifold-valued data, the symbol $|\nabla u(x)|$ in (2.2) can be understood in the sense of metric differentials [Kir94]. Examples of manifold-valued data are circle and sphere-valued data as appearing in SAR imaging [MF98] and color image processing [CKS01, VO02, KS02, LO14]. Other examples are data taking values in the special orthogonal group $SO(3)$ expressing vehicle headings, aircraft orientations or camera positions [URDDS05], Euclidean motion group-valued data [RBB$^+$12] as well as shape-space data [MM07]. As already mentioned, another prominent manifold is the space of positive (definite) matrices endowed with the Fisher-Rao metric [Rao45] which is the data space in DTI [PFA06]. This space is a Cartan-Hadamard manifold; it has particularly nice differential-geometric properties such as non-positive sectional curvature.

Related work dealing with the processing of data with values in a manifold are wavelet-type multiscale transforms [URDDS05, GW09, Wei10, WYW11, Wei12a, Wei12b] as well as manifold-valued partial differential equations [TD01, CTDF04, GHS15]. Work on statistics on Riemannian manifolds is [OC95, BP03, FLPJ04, BP05, Pen06, FJ07].

Existing computational approaches. In the context of DTI, Wang and Vemuri consider a Chan-Vese model for manifold-valued data which is a variant of the Potts model for the case of two segments. They further consider a piecewise smooth analogue [WV04, WV05] of their Chan-Vese model. Their method is based on a level-set active-contour approach. In order to reduce the computational load in their algorithms the authors resort to non-Riemannian distance measures in [WV04, WV05]. Recently, also DTI in the Riemannian setup in combination with the piecewise constant Chan-Vese model has been considered in [CSV12].

Tikhonov-type regularization using manifold-valued partial differential equations has been considered by Chefd'hotel et al. [CTDF04]; they numerically solve a corresponding evolution equation.

2.1.4 Total Variation Models for Manifold-valued Data and Higher-Order Generalizations

Total variation (TV) regularization was proposed by Rudin, Osher and Fatemi in the seminal paper [ROF92]; the corresponding minimization problem is frequently called the ROF model. A major advantage of TV regularization is that it better preserves sharp edges [GM01, SC03] in contrast to classical Tikhonov regularization. Especially because of this property TV minimization is used in a lot of applications. Examples are biomedical imaging [DBFZ$^+$06], geophysics [AS12] and computer vision [CYZ$^+$06, ZPB07].

There is a lot of work on theoretical properties of total variation regularization for scalar data. For example, results on existence and uniqueness of minimizers are shown in [CL97]. Connections to wavelet shrinkage are a topic in [PCXD99]; equivalences between diffusion techniques, thresholding strategies and TV minimization are given in [SWB$^+$04]. Many algorithms for TV minimization for scalar- and vector-valued data have been developed. Rudin, Osher, Fatemi [ROF92] consider ℓ^2 data terms. They use gradient descent on the Euler-Lagrange equations of the total variation functional. Other methods are based on Fenchel duals [Cha04], the alternating direction method of multipliers (ADMM) [YZY10], as well as split Bregman methods [GO09]. Total variation regularization for matrix-valued data is the topic of [RWT$^+$14] and [LO14]. Total variation problems with ℓ^1 data terms were for instance considered in [All92, Nik02, CE05]. Ap-

proaches using the ℓ^1-TV functional have edge preserving properties and are additionally more robust to outliers. For ℓ^1-TV regularization various algorithmic approaches have been proposed: for example, schemes based on smooth approximations are considered in [Nik04, NNZC08]; semi-smooth Newton approaches are proposed in [CJK10]; primal-dual methods are considered in [DHN09, CP11a].

In order to avoid staircasing, higher order generalizations of the TV functional are often employed. In particular, second order differences/derivatives are used. Pioneering work was done by Chambolle and Lions in [CL97] as well as by Bredies, Kunisch and Pock in [BKP10] which deals with total generalized variation. We further mention the papers [Sch98, CMM00, LLT03, LT06, HS06, CEP07, DWB09, SST11, LBU12] which include applications, discussions and references as well.

For TV related functionals for manifold valued data, a theoretical analysis is the topic of [GM06, GM07]. These two articles extend the paper [GMS93] on $\mathbb{S}^1$-valued functions where, in particular, the existence of minimizers of certain TV-type energies is shown.

Existing computational approaches. A convex relaxation based algorithm for TV regularization for $\mathbb{S}^1$-valued data was considered in [SC11, CS13]. An approach for TV regularization for general Riemannian manifolds was proposed in [LSKC13]. It is based on a reformulation as multilabel optimization problem with an infinite number of labels and a subsequent convex relaxation. This work emerged in parallel to the work [6] which uses a different approach.

Concerning higher order TV type functional, we are not aware of any algorithms developed previously neither for $\mathbb{S}^1$ nor for general manifold-valued data.

2.2 Free-discontinuity Problems for Directly Measured Data

We here summarize our results on free-discontinuity methods for directly measured vector space data which are contained in the papers [1, 2].

In the article [1], we study univariate L^1-Potts functionals. We here use the capital letter L^1 notation since we consider the space of (equivalence classes of) absolutely integrable functions w.r.t. the Lebesgue measure. Considering an L^1 data term instead of an L^2 data term is motivated by the observation that it produces better reconstructions when the data is corrupted by non-Gaussian noise. The main contributions of [1] are as follows. First, we develop a discretization framework for the continuous L^1-Potts functional. Secondly, we develop a fast algorithm to minimize the discretized L^1-Potts functionals. Thirdly, we show

that both continuous and discretized L^1-Potts functionals have blind deconvolution properties.

We start with our first main contribution. The continuous L^1-Potts functional is defined on $L^1[0,1]$ by

$$P_\gamma(u) = \gamma \cdot J(u) + \|u - f\|_1, \qquad \text{if } u \in \mathrm{PC}, \tag{2.6}$$

and by $P_\gamma(u) = \infty$ else. Here PC denotes the space of piecewise constant functions on the interval $[0,1]$. The regularity term $J(u) = \|\nabla u\|_0$ counts the number of jumps of u. (The abusive but appealing norm notation $\|\cdot\|_0$ is very popular in the sparsity community [Ela10].) A minimizer of the L^1-Potts functional may be seen as a *jump-sparse* approximation to the data f. We emphasize the non-uniqueness of minimizers. We consider (possibly non-equidistant) samplings $f_k = S_k f$ of continuous time data f obtained from f by a sampling operator S_k at some level k, $k \in \mathbb{N}$. One implementation of S_k is the integral sampling using the local averages

$$f^k(j) = S_k f(j) = \tfrac{1}{|I_j|} \int_{I_j} f \, \mathrm{d}\lambda, \tag{2.7}$$

where the intervals I_j live on level k and form a partition of the unit interval. If f is continuous, one can also consider point sampling

$$f^k(j) = S_k f(j) = f(x_j), \tag{2.8}$$

where x_j might be taken as the midpoint of the interval I_j. Then, the discrete Potts functionals read

$$P^k_\gamma(u) = \gamma \cdot J(u) + \|u - \textstyle\sum_j S_k f(j) \cdot \mathbf{1}_{I_j}\|_1, \qquad \text{if } u \in \mathrm{PC_k}, \tag{2.9}$$

and $P^k_\gamma(u) = \infty$ else. Here, $\mathrm{PC_k}$ are those functions which only jump at the interval boundaries of the I_j. Our first result consists of convergence statements of the discretizations to the continuous model as well as corresponding statements for the minimizers. It is stated as Theorem 2.4 in [1].

Theorem 1 *Let f be an integrable function and let P_γ be the corresponding continuous L^1-Potts functional. Then the discrete L^1-Potts functionals P^k_γ converge to P_γ as $k \to \infty$ in the sense of Γ-convergence. Each sequence u_k, where u_k is a minimizer of P^k_γ, has at least one accumulation point. Each such accumulation point u is a minimizer of P_γ, i.e., $P_\gamma(u) = \inf_{v \in L^1[0,1]} P_\gamma(v)$.*

We recall that Γ-convergence and convergence of minimizers in connection with the L^2-Potts functional were topics of [BLMW07, BKL$^+$09]. In these papers, Γ-convergence is shown for a modified L^2-Potts functional. This approach relies

on the Hilbert space structure of L^2, and therefore does not carry over to the L^1 context.

Our second main contribution in [1] is a fast algorithm for the L^1 Potts problem which has the same complexity as the state-of-the-art algorithm proposed for the L^2 Potts problem in [FKLW08]. The algorithm in [FKLW08] computes an exact minimizer of the L^2 Potts problem in $O(n^2)$ time and $O(n)$ space, where n denotes the length of the discrete data. In this paper, the authors also consider the discrete L^1-Potts model. Using a red-black tree approach the authors obtain an $O(n^2 \log n)$ time algorithm to minimize the discrete L^1 Potts functional which improves the time complexity of a naive implementation by a factor of n. However, this comes with $O(n^2)$ space consumption which means losing a factor of n. In [1], we introduce a suitable data structure which we call *indexed linked histogram* and combine it with the dynamic programming approach. Concerning runtimes we obtain the following statement

Theorem 2 *The algorithm proposed for the L^1-Potts problem in [1] computes an exact minimizer of the discrete L^1-Potts functional within $O(n^2)$ time and $O(n)$ space.*

This statement is formulated as Corollary 3.2 in [1] and deals with equidistant sampling. A corresponding statement for non-equidistant sampling is [1, Theorem 3.1]. Since it involves more technicalities we refer to [1] for details. It incorporates weights which arise naturally in applications where data are available only on a non-uniform grid [SSM+11]. Besides this asymptotic result, the actual runtime of the L^1 algorithm is, for data of larger size, only less than 20% slower than that of the L^2 version. We compare the L^1-Potts algorithm with the L^2-Potts algorithm of [FKLW08] and the L^1-TV algorithm of [CJK09] for various types of noise. The numerical experiments in [1] indicate that L^1-Potts minimization outperforms L^1-TV minimization whenever the underlying signal is jump-sparse. Furthermore, the L^1-Potts functional is more robust to non-Gaussian noise than the L^2-Potts functional; further, it yields comparable results for Gaussian noise. An efficient implementation of the proposed algorithm can be found at `http://www.pottslab.de`.

We next discuss our third main contribution in [1]. We observe that besides its high robustness to noise, the L^1-Potts functional has blind deconvolution properties which are neither shared by the L^2-Potts nor by the L^1-TV functional. In fact, we show that the continuous time L^1-Potts functional exactly recovers piecewise constant signals g from mildly blurred measurements $K * g$ *without* knowledge of the (narrowly supported) convolution kernel K. This result is stated as [1, Theorem 4.4]. For its formulation, we denote the minimal and maximal jump height of g by $h_{\min}$ and $h_{\max}$, respectively. Furthermore, $l_{\min}$ denotes the minimal interval length between two jumps of g.

Theorem 3 *Let g be a piecewise constant function on $[0,1]$ and $f = K * g$. If the support size κ of the convolution kernel K satisfies*

$$\kappa \leq \frac{h_{\min} l_{\min}}{2(8h_{\max} + h_{\min})} \tag{2.10}$$

then g is the unique minimizer of the Potts functional P_γ associated with f, i.e.,

$$P_\gamma(g) = \inf_{u \in \mathrm{PC}} P_\gamma(u) = \inf_{u \in \mathrm{PC}} \gamma J(u) + \|u - f\|_1, \tag{2.11}$$

for any Potts parameter γ satisfying

$$2\kappa h_{\max} \leq \gamma \leq \tfrac{1}{2} h_{\min} l_{\min} - (h_{\min} + 6h_{\max})\,\kappa. \tag{2.12}$$

This property is inherited by the discretizations of the L^1-Potts functional. A corresponding statement is formulated as [1, Theorem 4.5] where we refer to for details. The experiments conducted in [1] show that the blind deconvolution property approximately persists under noise. For the reconstruction of severely blurred signals (and known kernel), we consider the related deconvolution problem

$$\gamma \cdot \|\nabla u\|_0 + \|K * u - f\|_1 \to \min. \tag{2.13}$$

We derive a heuristic approach to solve (2.13) where our fast algorithm to solve the L^1 Potts problem for K being the identity is employed as basic building block of an iterative algorithm.

In [2], we develop fast solvers for the 2D Potts problem with higher-dimensional range spaces. We develop and extend the ideas to employ the dynamic programming solvers for univariate Potts problems as basic building blocks to approach more complex problems. We first show the potential gain of our approach by comparing our strategy with the current state-of-the-art. We see in Figure 2.3 that the current benchmark algorithm, the graph-cut based α-expansion [BVZ01, BK04, SZS+08] already breaks down at a moderate dimension. Our approach, in contrast, scales much better with the problem size in terms of computational costs, and it reaches lower functional values at the same time.

Let us explain our approach. For an image f defined on an $n \times m$ grid with values in $\mathbb{R}^s$, we consider a discrete domain version of (2.1) which reads

$$u^* = \operatorname{argmin}_{u \in \mathbb{R}^{m \times n \times s}} \Big\{ \gamma \sum_{i,j} \sum_{(a,b) \in \mathcal{N}} \omega_{a,b} \cdot [u_{i,j,:} \neq u_{i+a,j+b,:}] + \sum_{i,j,k} |u_{i,j,k} - f_{i,j,k}|^2 \Big\}. \tag{2.14}$$

Here $u_{i,j,:}$ is a vector sitting in the pixel (i,j) and the Iverson bracket $[\cdot]$ equals one whenever the expression in brackets is true, and it yields zero else. The neighborhood relation $\mathcal{N}$ together with the positive weights ω define a discrete boundary

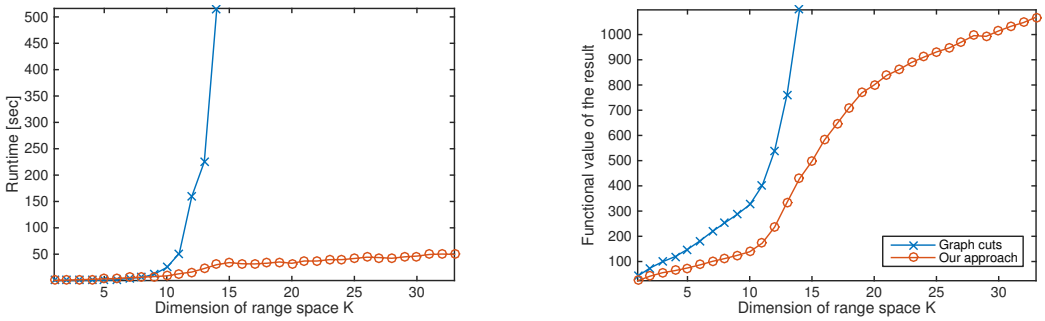


Figure 2.3: Comparison of our algorithm for the Potts problem with the graph-cut based benchmark algorithm for a multispectral image (255×355). The runtime of graph cuts grows exponentially in the dimension of the range space whereas our approach scales approximately linearly. At the same time, our method reaches lower functional values.

length of the corresponding partitions. The simplest case is given by the coordinate unit vectors as neighborhood relation and unit weights. It corresponds to the penalty $\|\nabla_1 u\|_0 + \|\nabla_2 u\|_0$. This penalty counts the non-zero elements of the directional difference operators ∇_1 and ∇_2 applied to u. Since this measures the boundary length of the partitions in the so called Manhattan metric (ℓ^1-norm), the results suffers from block artifacts. To deal with such effects, we consider larger neighborhoods and derive appropriate weights to obtain a more isotropic discretization.

To approach (2.14) we reformulate it as a suitable constrained optimization problem. In case of the Manhattan discretization, this problem reads

$$\gamma\|\nabla_1 u\|_0 + \gamma\|\nabla_2 v\|_0 + \tfrac{1}{2}\|u-f\|_2^2 + \tfrac{1}{2}\|v-f\|_2^2 \to \min, \text{ s.t. } u - v = 0,$$

where $u, v \in \mathbb{R}^{m\times n}$. The augmented Lagrangian of this consensus form is given by

$$L_\mu(u,v,\lambda) = \gamma\|\nabla_1 u\|_0 + \gamma\|\nabla_2 v\|_0 + \tfrac{1}{2}\|u-f\|_2^2 + \tfrac{1}{2}\|v-f\|_2^2 + \langle\lambda, u-v\rangle + \tfrac{\mu}{2}\|u-v\|_2^2. \quad (2.15)$$

The parameter $\mu > 0$ regulates how strongly the difference between u and v is penalized. The dual variable λ is an $(m \times n)$-dimensional matrix of Lagrange multipliers. We approach this problem using the alternating direction methods of multipliers (ADMM). In the corresponding iteration we first minimize $L_\mu(u,v,\lambda)$ with respect to u, then with respect to v. The third step is the update of the dual variable λ. As a result, we obtain computationally accessible subproblems. The crucial point is that these subproblems reduce to univariate Potts problems which can be solved fast and exactly using dynamic programming. We propose an acceleration strategy for the dynamic program which, in our experiments, resulted in a

speed up of the algorithm by a factor of four to five. The underlying theoretical basis for this speed-up is Theorem 2 in [2]. The corresponding algorithm is given as Algorithm 2 in [2]. We note that the proposed algorithm convergences; this result is stated as Theorem 1 in [2].

Theorem 4 *Algorithm 1 of [2] converges in the sense that* $(u^k, v^k) \to (u^*, v^*)$ *with* $u^* = v^*$.

Since the original problem is NP-hard, we cannot expect that the limit point is in general a minimizer of the cost function (2.14). In [2], we observe that, in practice, we get slightly lower functional values than the state-of-the art method graph cuts. The visual quality of our results is often slightly better than the result obtained by the graph cuts method.

The proposed algorithm has the advantage that it does not need any discretization in the codomain or, synonymously, range space of u. This contrasts most state of the art schemes which require a finite set of discrete labels. The main benefit of the proposed algorithm is its efficiency with respect to runtime and memory. Already for color images with a relatively coarse discretization of the color cube $[0,1]^3$, the computational costs of our approach are significantly lower than that of the graph cuts method as well as lower than the costs of convex relaxation methods. The advantage becomes even more prominent for higher dimensional codomains since our method grows linearly in the dimension of the codomain. This contrasts the exponential growth of graph cut based methods. Due to the linear scaling, we can process images taking values in a high-dimensional vector space in a reasonable time. Here a prominent example are multispectral images which may have more than 30 channels. The applications in [2] illustrate that our method is well suited for both image segmentation and the restoration of cartoon-like images.

We point out that the proposed method allowed us to process feature images of range space dimensions of a few hundreds as they appear in texture segmentation problems; see [SWU14].

2.3 Free-discontinuity Problems for Indirectly Measured Data

In [3], we propose a majorization-minimization approach for the discrete univariate Mumford-Shah and Potts problems. Discretizations of the Mumford-Shah functional are known under the name *Blake-Zisserman functionals*. The *weak string model* as it is called by Blake and Zisserman in [BZ87] is given in its uni-

variate version for inverse problems by

$$\text{minimize} \quad \gamma \sum_i \min(|u_i - u_{i-1}|^q, s^q) + \|Au - f\|_2^2. \tag{2.16}$$

Here the data f lives in a real-valued finite dimensional linear space. The parameter $\gamma > 0$ controls the trade-off between data fidelity and regularizing term. The parameter $s > 0$ determines the discontinuity in the sense that, if $|u_i - u_{i-1}| > s$, we say there is a jump between $i-1, i$. The penalty for a jump does not depend on its magnitude. If the distance is smaller than s, the penalty is the discrete ℓ^q-th variation of u. In the inverse setup considered in this section, the discrete Potts model reads

$$\text{minimize} \quad \gamma \|\nabla u\|_0 + \|Au - f\|_2^2. \tag{2.17}$$

Here, ∇ denotes the backward difference operator. Formally, the Potts problem (2.17) can be seen as the "Lagrange formulation" of the following constrained problem which we call *J-jump sparsity problem*. It reads

$$\begin{array}{ll} \text{minimize} & \|Au - f\|_2^2, \\ \text{subject to} & \|\nabla u\|_0 \leq J, \end{array} \tag{2.18}$$

where J is a nonnegative integer. To our knowledge, for general A, the J-jump sparsity problem (2.18) has only recently appeared in the literature [FHM14]; the authors obtain asymptotic statements in the context of inverse regression. The situation $A = I$, where I denotes the identity, is well studied; see [Bru65, AL89]. Related sparsity problems are the topic of [BD08].

We explain our majorization-minimization approach. We start with the Potts problem (2.17). We need the corresponding surrogate functional P_γ^{surr} which is defined by

$$P_\gamma^{\text{surr}}(u, v) = \gamma \|\nabla u\|_0 + \|Au - f\|_2^2 - \|Au - Av\|_2^2 + \|u - v\|_2^2. \tag{2.19}$$

By a straightforward calculation, we see that $P_\gamma^{\text{surr}}(u, v) = \gamma \|\nabla u\|_0 + \|u - v + A^*Av - A^*f\|_2^2 + C$, with some constant C which does not depend on u and which is therefore negligible. It follows that, for fixed v,

$$\operatorname{argmin}_u P_\gamma^{\text{surr}}(u, v) = \operatorname{argmin}_u \left\{ \gamma \|\nabla u\|_0 + \|u - v + A^*Av - A^*f\|_2^2 \right\}. \tag{2.20}$$

We now successively compute $u^{k+1} = \operatorname{argmin}_u P^{\text{surr}}(u, u^k)$. We obtain the following *iterative Potts minimization algorithm* given by the iteration

$$\begin{cases} d^{k+1} = (I - A^*A)u^k + A^*f, \\ u^{k+1} = \operatorname{argmin}_u \gamma \|\nabla u\|_0 + \|u - d^{k+1}\|_2^2, \end{cases} \tag{2.21}$$

where I denotes the identity matrix. The first step is a matrix-vector multiplication. The crucial point is that the second step consists of minimizing a Potts functional with $A = I$, which can be solved by dynamic programming in quadratic time. Proceeding similarly for the J-jump sparsity problem (2.18), we obtain the iteration

$$\begin{cases} d^{k+1} &= (I - A^*A)u^k + A^*f, \\ u^{k+1} &= \operatorname{argmin} \|u - d^{k+1}\|_2^2, \quad \text{s.t. } \|\nabla u\|_0 \leq J. \end{cases} \tag{2.22}$$

We call this iterative scheme *iterative J-jump sparsity algorithm*. We have to solve a J-jump sparsity problem for $A = I$ in (2.22). As the Potts problem, it can be solved by dynamic programming; the algorithm was reinvented several times in a different context; references are [Bru65, AL89]. Finally, we use the above principle to derive the *iterative Blake-Zissermann minimization algorithm*

$$\begin{cases} d^{k+1} = (I - A^*A)u^k + A^*f, \\ u^{k+1} = \operatorname{argmin}_u \gamma \sum_i \min(|u_i - u_{i-1}|^q, s^q) + \|u - d^{k+1}\|_2^2. \end{cases} \tag{2.23}$$

Again, the second line constitutes a Blake-Zisserman problem for $A = I$. For such Blake-Zisserman problems with $A = I$, we have proposed an exact solver of quadratic runtime in [HSW15].

We briefly comment on related algorithms for the sparsity problem. Fornasier and Ward [FW10] rewrite the Blake-Zisserman problem as a problem with separable penalty. They derive generalized iterative thresholding algorithms for the rewritten problem. Related algorithms are iterative soft thresholding for ℓ^1 penalized problems, analyzed by Daubechies, Defrise, and De Mol in the seminal paper [DDDM04], and the iterative hard thresholding algorithms for ℓ^0 penalizations, analyzed by Blumensath and Davies in [BD08, BD09]. In contrast to the approaches in [FW10] and [BD08, BD09], which lead to thresholding algorithms, our approach leads to non-separable problems in the backward step.

Facing the challenge due to the non-separability of the backward step, we provide a convergence analysis. In particular, we obtain convergence statements towards local minimizers. We start our analysis by showing that the Potts problem (2.17) and the J-jump sparsity problem (2.18) are not equivalent; this result is stated as Theorem 3.1 in [3].

Theorem 5 *The problems* (2.17) *and* (2.18) *are not equivalent. More precisely, if* u^* *is a solution of the Potts problem* (2.17) *with* $\gamma > 0$ *then it is also a solution of the constrained problem* (2.18) *with parameter* $k = \|\nabla u^*\|_0$. *On the other hand, a minimizer of* (2.18) *need not necessarily be a minimizer of* (2.17) *– not even for some parameter.*

In the following theorems on convergence, the measurement matrix A is always assumed to fulfill $\|A\| < 1$ with respect to the operator norm on ℓ^2. We note that

this can always be achieved by rescaling which, however, results in a smaller step size in the respective algorithm. We first obtain the following convergence results for iterative Potts minimization; it is stated as Theorem 3.2 in [3].

Theorem 6 *We let* $\|A\| < 1$. *Then the iterative Potts minimization algorithm* (2.21) *converges to a local minimizer of the inverse Potts functional* (2.17) *for any starting point. The convergence rate is linear. Furthermore, we have the following relation between local minimizers* $\mathscr{L}$, *global minimizers* $\mathscr{G}$ *and the fixed points* $\mathrm{Fix}(\mathbb{I})$ *of the iteration* (2.21),

$$\mathscr{G} \subset \mathrm{Fix}(\mathbb{I}) \subset \mathscr{L}. \tag{2.24}$$

We next consider the constrained J-jump sparsity problem (2.18). The formulation of the corresponding convergence result, Theorem 3.3 in [3], is somewhat more involved which is due to additional assumptions. These additional assumptions are usually fulfilled when there is noise in the data. We recall that u is J-jump sparse if it has at most J jumps.

Theorem 7 *We let* $\|A\| < 1$ *and assume that the unrestricted problem of minimizing* $\|Au - f\|^2$ *with respect to* $u \in \mathbb{R}^L$ *has at most one J-jump sparse solution. (This is for example the case if A is injective.)*

If the unrestricted problem has no J-jump sparse solution, then the iterative J-jump sparsity algorithm (2.22) *converges towards a local minimizer of* (2.18).

Otherwise, the iterative J-jump sparsity algorithm (2.22) *produces iterates* u^k *that either converge to a local minimizer of* (2.18) *with (exactly) J jumps or they have a cluster point which is a global minimizer of* (2.18) *with (strictly) less than J jumps. If, in this situation, f is in the range of A and A is injective, then the iterates converge to a local minimizer (which is a global minimizer when it has (strictly) less than J jumps.)*

Regarding the analysis of the iterative Blake-Zissermann minimization algorithm, we obtain the following convergence result which is stated as Theorem 3.4 in [3].

Theorem 8 *For* $\|A\| < 1$, *the iterative Blake-Zisserman minimization algorithm* (2.23) *converges to a local minimizer of the Blake-Zisserman functional* (2.16) *for any starting point. Furthermore, the relation* (2.24) *holds true in the context of the Blake-Zisserman functionals* $B_{\gamma,s}$ *of* (2.16) *as well.*

In [3], we show the applicability of our approach in several signal recovery experiments. We consider deconvolution problems with full as well as with partial data. Finally, we apply our methods to real data. We use it for estimating the steps in the rotation of the bacterial flagellar motor [SRL$^+$05].

In [4], we consider the univariate discrete Potts problem with indirect measurements for the more general case $p \in [1,\infty)$ which, in particular, includes $p = 1$. It is given by

$$P_\gamma(u) = \gamma \cdot \|\nabla u\|_0 + \|Au - b\|_p^p \to \min. \tag{2.25}$$

We recall that ∇ denotes the backward difference operator, i.e., $\nabla u_i = u_i - u_{i-1}$. We first clarify the existence of minimizers.

Theorem 9 *The discrete inverse Potts problem* (2.25) *has a minimizer.*

The corresponding result is stated as Theorem 1 in [4]. We recall that the corresponding continuous models need not have a minimizer without additional assumptions.

Then, we propose an algorithm for (2.25) based on the alternating directions method of multipliers (ADMM). For its derivation, we rewrite (2.25) as the bivariate constrained optimization problem

$$\begin{aligned} &\text{minimize} && \gamma\|\nabla u\|_0 + \|Av - b\|_p^p \\ &\text{subject to} && u - v = 0. \end{aligned} \tag{2.26}$$

We incorporate the constraint $u - v$ into the target functional to obtain the augmented Lagrangian of (2.26) given by $L_\mu(u,v,\lambda) = \gamma\|\nabla u\|_0 + \langle \lambda, u - v\rangle + \frac{\mu}{2}\|u - v\|_2^2 + \|Av - b\|_p^p$ Here, the parameter $\mu > 0$ regulates the coupling of u and v. The dual variable λ is an n-dimensional vector of Lagrange multipliers. Some computation yields the problem

$$L_\mu(u,v,\lambda) = \gamma\|\nabla u\|_0 - \tfrac{\mu}{2}\|\tfrac{\lambda}{\mu}\|_2^2 + \tfrac{\mu}{2}\|u - v + \tfrac{\lambda}{\mu}\|_2^2 + \|Av - b\|_p^p \to \min. \tag{2.27}$$

In order to minimize (2.27) we use the ADMM, see e.g. [BPC+11], i.e., we iteratively minimize (2.27) w.r.t. one variable keeping the others fixed. The resulting ADMM algorithm for the Potts problem with indirect measurements (2.17) reads

$$\begin{cases} u^{k+1} \in \operatorname{argmin}_u \gamma\|\nabla u\|_0 + \frac{\mu_k}{2}\|u - (v^k - \frac{\lambda^k}{\mu_k})\|_2^2, \\ v^{k+1} = \operatorname{argmin}_v \|Av - b\|_p^p + \frac{\mu_k}{2}\|v - (u^{k+1} + \frac{\lambda^k}{\mu_k})\|_2^2, \\ \lambda^{k+1} = \lambda^k + \mu_k(u^{k+1} - v^{k+1}). \end{cases} \tag{2.28}$$

Here, we employ a coupling sequence μ_k such that $\sum_k \mu_k^{-1/2} < \infty$. Both subproblems appearing in the first and the second line of (2.28) are computationally tractable. The first subproblem is the minimization of a classical Potts problem for A being the identity matrix which is accessible via dynamic programming as already explained above. For $p = 2$, the second subproblem is the minimization of a

classical Tikhonov-type problem which results in a linear problem. In particular, for convolution operators, the resulting linear problem can be solved efficiently using Fourier transform techniques. For $p = 1$, we can employ a fast semismooth Newton method [CJK10].

In [4], we further clarify the relation between the Potts problem and the sparse recovery problem

$$S_\gamma(x) = \gamma \cdot \|x\|_0 + \|Ax - b\|_p^p \to \min. \tag{2.29}$$

We obtain the following statements which are formulated as Theorems 4 and 5 in [4].

Theorem 10 *Let $x^* \in \mathbb{R}^{n+1}$ be a minimizer of the inverse Potts functional associated with the matrix $B = A\nabla$, i.e.,*

$$x^* \in \operatorname{argmin}_{x \in \mathbb{R}^{n+1}} \gamma \|\nabla x\|_0 + \|Bx - b\|_p^p. \tag{2.30}$$

Then $u^* = \nabla x^*$ minimizes the sparsity problem (2.29) *related to the matrix A and data b.*

We use the relation given by Theorem 10 to rewrite sparsity problems as Potts problems.

Theorem 11 *For the inverse Potts problem* (2.25) *associated with the matrix A and data b we consider the sparsity problem associated with the matrix $B = A'\nabla^+$ and data b', where A', ∇^+ and b' are transformed matrices and vectors, repectively, given as in [4, Theorem 5]. Let u^* be a minimizer of the sparsity problem with respect to B, b',i. e.,*

$$u^* \in \operatorname{argmin}_{u \in \mathbb{R}^{n-1}} \gamma \|u\|_0 + \|Bu - b'\|_2^2. \tag{2.31}$$

Then $x^* = \nabla^+ u^* + \mu(\nabla^+ u^*)e$ (with μ given by [4, Formula (28)]) is a solution of the inverse Potts problem (2.25) *associated with A, b.*

We note that Theorem 11 only holds for $p = 2$; it is not clear how to get a similar result for $p \neq 2$. A related statement for Blake-Zissermann problems is given in [FW10].

In [4], we apply the proposed algorithm to reconstruct jump-sparse signals from indirect measurements. In particular we consider blurred data as well as Fourier data. We consider the case of incomplete measurements and data corrupted with Gaussian noise, Laplacian noise, or impulsive noise. Our algorithm recovers jump sparse signals almost perfectly from a reasonable level of noise. In average, it yields higher reconstruction quality than TV minimization. We further apply the developed method to sparse recovery problems. We consider blurred

data under different types of noise. We show the potential of our method by comparing it with orthogonal matching pursuit [Tem03, Tro04, Pet06], basis pursuit denoising [YZ11], iterative hard thresholding [BD09] and iteratively reweighted ℓ^1 minimization [CWB08], which are the state-of-the-art approaches to sparse recovery. In order to the guarantee the reproducibility of the derived results, an implementation of the proposed algorithms is available at `www.pottslab.de`.

As concluding remark, we point out that, we have proposed an ADMM based algorithm for the bivariate Potts problem in [SWFU15]. Furthermore, we have developed an related algorithmic framework for the inverse first order Mumford-Shah model in [HSW15].

2.4 Free-Discontinuity Problems for Data with Values in a Manifold

In this section, we consider manifold-valued data. We summarize the paper [5], where we have derived algorithms for Mumford-Shah and Potts regularization for manifold-valued data. We note that, for manifold-valued data, an additional challenge arises from the fact that the elementary features of vector spaces such as addition and scalar multiplication are not available. Besides the approach to the Chan-Vese model for the particular instance of DTI data [WV04, WV05] (cf. Chapter 2.1.3), our recently proposed algorithms in [5] seem to be the first algorithmic approaches to the Mumford-Shah problem and the Potts problem for manifold-valued data. We point out that our algorithms are applicable for any Riemannian manifold whose exponential mapping and its inverse can be evaluated within reasonable time.

We start with the univariate problems. These are not only important in their own right; variants of the derived solvers are also used as a basic building block for the proposed algorithm for the multivariate problems. In the univariate case, the discretization of the Mumford-Shah functional and the Potts functional is rather straightforward. The (equidistantly sampled) discrete Mumford-Shah functional reads

$$B_{\alpha,\gamma}(x) = \frac{1}{p}\sum_{i=1}^{n} d(x_i, f_i)^p + \frac{\alpha}{q}\sum_{i \notin \mathscr{J}(x)} d(x_i, x_{i+1})^q + \gamma |\mathscr{J}(x)|, \tag{2.32}$$

where d is the distance with respect to the Riemannian metric in the manifold M, $f \in M^n$ is the data, and $\mathscr{J}$ is the jump set of x. The jump set is given by $\mathscr{J}(x) = \{i : 1 \leq i < n \text{ and } d(x_i, x_{i+1}) > s\}$ where the jump height s is related to the parameter γ via $\gamma = \alpha s^q/q$. Using a truncated power function we may rewrite

(2.32) in the Blake-Zisserman type form

$$B_{\alpha,s}(x) = \frac{1}{p}\sum_{i=1}^{n} d(x_i, f_i)^p + \frac{\alpha}{q}\sum_{i=1}^{n-1} \min(s^q, d(x_i, x_{i+1})^q), \tag{2.33}$$

where s is the argument the power function $t \mapsto t^q$ is truncated at.

The discrete univariate Potts functional for manifold-valued data reads

$$P_\gamma(x) = \frac{1}{p}\sum_{i=1}^{n} d(x_i, f_i)^p + \gamma |\mathscr{J}(x)|, \tag{2.34}$$

where d is the distance in the manifold and i belongs to the jump set of x if $x_i \neq x_{i+1}$.

We first show in [5] that the problems (2.32) and (2.34) have a minimizer. We note that certain variants of the Mumford-Shah and Potts functional do not have a minimizer without additional assumptions; see for instance [FMS13]. The following result is stated as Theorem 1 in [5].

Theorem 12 *In a complete Riemannian manifold the discrete Mumford-Shah functional* (2.32) *and the discrete Potts functional* (2.34) *have a minimizer.*

We note that the data spaces in applications are typically complete Riemannian manifolds.

For univariate Mumford-Shah and Potts problems, we derive solvers based on a combination of dynamic programming techniques developed in [MS89, Cha95, WL02, FKLW08] and proximal point based methods for manifold-valued data developed by the authors in [6]. In fact, the motivation for [6] (which studies the TV problem for manifold-valued data) was to find a solver for the univariate Mumford problem for manifold-valued data.

We obtain that our algorithms compute global minimizers for data living in Cartan-Hadamard manifolds which includes many symmetric spaces. Prominent examples are the spaces of positive matrices (which are the data space in diffusion tensor imaging) and the hyperbolic spaces. Cartan-Hadamard manifolds are simply-connected complete Riemannian manifolds of nonpositive sectional curvature. For details we refer to [dC92] or to [BGS85]. In particular, in these manifolds, geodesics always exist and are unique shortest paths. The following result is formulated as Theorem 2 in [5].

Theorem 13 *In a Cartan-Hadamard manifold, Algorithm 1 in [5] computes a global minimizer for the univariate Mumford-Shah problem* (2.32) *and the univariate Potts problem* (2.34).

This result generalizes to the more general class of (locally compact) Hadamard spaces. These are certain metric spaces generalizing the concept of Cartan-Hadamard manifolds [Stu03]. Examples of Hadamard spaces which are not Cartan-Hadamard manifolds are given in [Stu03]. We point out that, for general complete Riemannian manifolds, we obtain an analogous result under the additional assumption of restricting the search space to candidates whose jump sets correspond to admissible partitions as explained in [5].

We next consider the multivariate situation. Here, the straight forward finite difference discretization with respect to the coordinate directions is known to produce undesired block artifacts in the reconstruction [Cha99]. The results improve significantly when including further finite difference directions such as the diagonal directions [Cha99, SWFU15]. We use the notation $d^p(x,y)$ for the pth power of the p-distance of two manifold-valued images x,y, i.e., $d^p(x,y) = \sum_{i,j} d^p(x_{ij},y_{ij})$. We further define the penalty function

$$\Psi_a(x) = \sum_{i,j} \psi(x_{(i,j)+a}, x_{ij})$$

with respect to some finite difference vector $a \in \mathbb{Z}^2 \setminus \{0\}$ using the potential function ψ which we instantiate in the Mumford-Shah case by $\psi(w,z) = \frac{1}{q}\min(s^q, d(w,z)^q)$. and in the Potts case by $\psi(w,z) = 1$ if $w \neq z$, and 0 else, for $w,z \in M$. We employ discretizations of the form

$$\min_{x \in M^{m\times n}} \frac{1}{p} d^p(x,f) + \alpha \sum_{s=1}^{R} \omega_s \Psi_{a_s}(x), \tag{2.35}$$

where the finite difference vectors $a_s \in \mathbb{Z}^2 \setminus \{0\}$ belong to a neighborhood system $\mathcal{N}$. For example, $\mathcal{N} = \{(1,0);(0,1);(1,1);(1,-1)\}$. The values $\omega_1,\ldots,\omega_R$ are non-negative weights. We obtain the existence of minimizers of the discrete functional (2.35). The corresponding result is stated as Theorem 4 in [5].

Theorem 14 *Let M be a complete Riemannian manifold. Then the discrete Mumford-Shah and Potts problems* (2.35) *both have a minimizer.*

For the discrete Mumford-Shah and Potts problems for manifold-valued images (where the problems become NP-hard), we propose a novel splitting approach. Starting from (2.35) we use a penalty method to split the problems into computationally tractable subproblems: we first rewrite (2.35) as the constrained problem

$$\min_{x_1,\ldots,x_R} \sum_{s=1}^{R} \frac{1}{pR} d^p(x_s,f) + \alpha\omega_s \Psi_{a_s}(x_s) \tag{2.36}$$

subject to $x_s = x_{s+1}$ for all $1 \leq s \leq R$, where we use the convention $x_{R+1} = x_1$. We include the constraints into the target functional using a penalty method (see for instance [Ber76]) and get the problem

$$\min_{x_1,\ldots,x_R} \sum_{s=1}^{R} \omega_s pR\alpha\Psi_{a_s}(x_s) + d^p(x_s, f) + \mu_k d^p(x_s, x_{s+1}).$$

We use an increasing coupling sequence $(\mu_k)_k$ which fulfills the summability condition $\sum_k \mu_k^{-1/p} < \infty$. Optimization with respect to all variables simultaneously is still not tractable, but our specific splitting allows us to minimize the functional with respect to the variables $x_1, \ldots, x_R$ separately. These subproblems are closely related to univariate Mumford-Shah and Potts problems. They can be solved using the methods we developed for the corresponding univariate problems. The proposed method does not require any a priori restrictions on the edge set; further, we do not have to discretize the data space. The presented algorithm is subsumed in [5, Equ. 18]. We eventually show convergence. The corresponding statement is Theorem 5 in [5].

Theorem 15 *For Cartan-Hadamard manifold-valued images, the proposed algorithm [5, Equ. 18] converges for both the Mumford-Shah and the Potts problem.*

In [5], we apply our method to diffusion tensor imaging (DTI) as well as Q-ball imaging. For DTI, we first consider several synthetic examples corrupted by Rician noise and show our algorithms potential for edge-preserving denoising. In certain cases, the edge set produced by our method can directly serve as a segmentation. We illustrate this for the corpus callosum of real human brain data. In DTI, oriented diffusivity along fiber structures is reflected by the anisotropy of the corresponding tensors; typically, there is one large eigenvalue and the corresponding eigenvector yields the orientation of the fiber. Potential problems arise in areas where two or more fiber bundles are crossing because the tensors are not able to represent multiple directions. In order to overcome this, the Q-ball imaging (QBI) approach [Tuc04, HMH$^+$06, DAFD07] uses higher angular information to allow for multiple directional peaks at each voxel; it was, for instance, applied to diffusion tractography [BJBJ$^+$07]. The Q-ball imaging data can be modeled by a probability density on the $3D$-unit sphere which is called orientation distribution function (ODF). The corresponding space of ODFs can in turn be endowed with a Riemannian manifold structure [GLTV09]. We show the capability of our methods by applying it to Q-ball imaging as well.

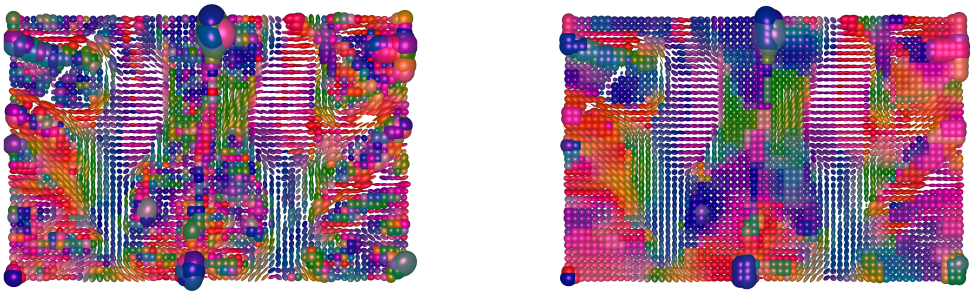


Figure 2.4: *Left:* Diffusion tensor image of a real human brain. The ellipsoids represent the diffusivity at a voxel. *Right:* Total variation denoising using the cyclic proximal point algorithm developed in [6]; figure taken from [6].

2.5 TV and Higher-Order Models for Manifold-Valued Data

In this section, we summarize our results for TV and related higher-order models for manifold-valued data. We start with [6], where we propose algorithms for TV minimization for (Riemannian) manifold-valued data; for an illustration in DTI we refer to Figure 2.4. The algorithms are parallelizable, fast and do not require any a priori discretization. We propose a cyclic proximal point algorithm as well as a parallel proximal point algorithm to minimize discrete TV functionals with ℓ^p-type data terms in the case of manifold-valued data. The algorithms are based on iterative geodesic averaging which makes them applicable to a large class of data manifolds.

We explain our approach. Let us consider the problem of bivariate discrete ℓ^p-TV^q minimization

$$\frac{1}{p}\sum_{i,j} d^p(x_{ij}, f_{ij}) + \alpha\frac{1}{q}\sum_{i,j} d^q(x_{ij}, x_{i+1,j}) + \alpha\frac{1}{q}\sum_{i,j} d^q(x_{ij}, x_{i,j+1}) \to \min. \quad (2.37)$$

The data f_{ij} and the targets x_{ij} to minimize take their values in a Riemannian manifold M. Setting $q = 1$ in (2.37), we get the discrete (anisotropic) TV functional with ℓ^p data term. In particular, if $p = 1$, we are in the ℓ^1-TV setting. The case $q = 2$ corresponds to the classical Tikhonov regularization term in the linear space case. Huber data and regularizing terms can be incorporated as well; for details, we refer to [6]. We rewrite (2.37) as the sum

$$F(x) + \alpha\sum_{i,j} G_{ij}(x) + \alpha\sum_{i,j} H_{ij}(x) \to \min; \quad (2.38)$$

here, $F : M \times \ldots \times M \to \mathbb{R}$ is the data term,

$$F(x) = \frac{1}{p} \sum_{i,j=1}^{n,m} d^p(x_{ij}, f_{ij}); \tag{2.39}$$

and the functions $G_{ij}, H_{ij} : M \times \ldots \times M \to \mathbb{R}$ are given by

$$G_{ij}(x) = \frac{1}{q} d^q(x_{ij}, x_{i,j+1}), \quad H_{ij}(x) = \frac{1}{q} d^q(x_{ij}, x_{i+1,j}). \tag{2.40}$$

For each summand in (2.38), we consider its proximal mapping [Mor62, FO02, AF05]. The proximal mappings of the G_{ij} are defined by the minimization problem

$$\operatorname{prox}_{\lambda G_{ij}} x = \operatorname{argmin}_{y \in M^{n \times m}} \left(\lambda G_{ij}(y) + \frac{1}{2} d^2(x, y) \right), \tag{2.41}$$

where the parameter $\lambda > 0$ and the distance d on the product manifold $M^{n \times m}$ is given by $d^2(x,y) = \sum_{i,j=1}^{n,m} d(x_{ij}, y_{ij})^2$. The proximal mappings of F and the H_{ij} are defined analogously. The crucial point is that, using the splitting (2.38), the proximal mappings of all appearing summands can be explicitly computed as geodesic averages. More precisely, solving the minimization problem of (2.41) reduces to computing points on shortest geodesics joining given points. The same is true for the analogous problems for F and the H_{ij}. We get

$$\begin{aligned} (\operatorname{prox}_{\lambda G_{ij}} x)_{ij} &= [x_{ij}, x_{i,j+1}]_t, \\ (\operatorname{prox}_{\lambda G_{ij}} x)_{i,j+1} &= [x_{i,j+1}, x_{ij}]_t, \end{aligned} \tag{2.42}$$

where the symbol $[\cdot,\cdot]_t$ denotes the point reached after time t on the unit speed geodesic starting at the first argument in direction of the second argument. In the TV case $(q = 1)$, we get

$$t = \begin{cases} \lambda, & \text{if } \lambda < \frac{1}{2} d(x_{ij}, x_{i,j+1}), \\ d(x_{ij}, x_{i,j+1})/2, & \text{else.} \end{cases} \tag{2.43}$$

For $q = 2$, which corresponds to quadratic variation, we have

$$t = \frac{\lambda}{1 + 2\lambda} d(x_{ij}, x_{i,j+1}). \tag{2.44}$$

The proximal mappings of the H_{ij} are obtained analogously. We state the following result which is part of [6, Proposition 1].

Proposition 16 *If there is a single shortest geodesic joining the data items $x_{i,j}$ and $x_{i,j+1}$ (which is always the case for nearby points and almost everywhere globally), then the proximal mapping of G_{ij} is well defined as a single-valued mapping. It is given by* (2.42),(2.43),(2.44). *Analogous statements hold for the H_{ij}.*

We note that [6, Proposition 1] also considers the case of non-unique geodesics; we refer to [6] for details. The proximal mapping of F corresponds to the proximal mapping of the distance function in M and is given by [FO02]

$$(\text{prox}_{\lambda F})_{ij}(x) = [x_{ij}, f_{ij}]_t, \tag{2.45}$$

where, for the ℓ^2 data term, $t = \frac{\lambda}{1+\lambda} d(x_{ij}, f_{ij})$. For the ℓ^1 data term, $t = \lambda$ if $\lambda < d(x_{ij}, f_{ij})$ and $d(x_{ij}, f_{ij})$ else. In order to make our algorithms work on a concrete manifold, the only operations we need are those needed for calculating geodesics. The spaces which frequently appear as data spaces are matrix groups or related symmetric spaces. So usually, there are explicit formulae available for this task.

Our algorithms are iterative schemes. In each iteration, we apply the above proximal mappings of the functionals decomposing the TV functional. The first algorithm is a *cyclic proximal point algorithm*. We consider the problem in the form $F(x) + \alpha \sum_{i,j} G_{ij}(x) + \alpha \sum_{i,j} H_{ij}(x)$ given by (2.38). We first apply the proximal mapping of F which is given as pointwise geodesic averages of data f_{ij} and the argument of the functional x_{ij}; see (2.45). Then we successively apply the proximal mappings of all the G_{ij}. As a last step, the analogous operations are executed for the H_{ij}. Iteration of all these steps yields the algorithm which is stated as Algorithm 1 in [6]. During the iteration, the parameter λ_r of the proximal mappings is successively decreased. In this way, the penalty for deviation from the previous iterate is successively increased. It is chosen in a way such that the sequence λ_r is square-summable but not summable. This is moderate enough not to prevent convergence towards a minimizer; cf. Theorem 17. The second algorithm is a *parallel proximal point algorithm*. Here the proximal mappings are calculated for the same initial point and averaged afterwards using the Riemannian center of mass. Since computing mean values on a manifold is a relatively expensive iterative procedure, we also consider a variant which only does approximative averaging (but yields comparable results). We call this variant *fast parallel proximal point algorithm*. The corresponding parallel proximal point algorithms are stated in Algorithm 2 in [6].

The proposed algorithms belong to the class of proximal splitting methods for manifold valued data. References for related proximal splitting methods for scalar data are for instance [CP11b, BPC$^+$11, Ber11]. In [Bač14], cyclic proximal point methods are applied for the computation of means and medians in Hadamard spaces.

As an application, we consider denoising images which take their values in a manifold. We apply our algorithms to diffusion tensor images, interferometric SAR images as well as sphere and cylinder valued images. Further we consider SO(3)-valued data.

We obtain the convergence of Algorithm 1 and Algorithm 2 of [6] to a global minimizer for Cartan-Hadamard manifolds. The proofs in [6] work in the more general setup of Hadamard spaces without additional effort. For details on Hadamard spaces we refer to [Stu03] and the references therein or to the book [BH99]. The next result is stated as Theorem 2 in [6].

Theorem 17 *For data in a (locally compact) Hadamard space Algorithm 1 of [6] converges towards a minimizer of the ℓ^p-TV^q functional. The statement remains true when using Huber data and regularizing terms.*

The following statement is Theorem 3 in [6].

Theorem 18 *The parallel proximal algorithm for ℓ^p-TV^q minimization (Algorithm 2 of [6]) and its approximate variant converge towards a minimizer in every (locally compact) Hadamard space. The statement remains true when using Huber regularization and Huber data terms.*

In the CVPR paper [7] we deal with TV regularization in shape spaces. The considered manifolds are particularly high-dimensional when dealing with shape spaces [Ken84, MM07, Fle13]. We consider *shape signals*, i.e., collections of shapes which appear in a spatial or temporal context. An obvious example is object tracking in video sequences where all shapes have a natural temporal ordering. Another example is organ segmentation from tomographic imaging modalities such as computed tomography or magnetic resonance imaging, where a three-dimensional organ can be segmented by obtaining its two-dimensional outlines from all containing slices.

We use active contours to obtain the shape of the object under consideration within each slice or frame. The idea of active-contour-based tracking is already considered in the seminal work [KWT88]. Further reference is [CRD07, CMP$^+$07, SYM08, SYMS09, SMSY09]. Then we use our approach for TV regularization developed in [6] to regularize the shape signals. Even if active-contour-based segmentation methods are augmented by sophisticated shape metrics, the regularity of the obtained segmentation is not comparable to the one obtained by our approach. We observe in [7] that our method nicely complements Sobolev-type active contours by further regularizing their results.

There are various models for shape spaces; see for instance [MM06, MM07, YMSM07, SKJJ11, BBM14, BBMM14]. Here, we consider the Kendall shape space [Ken84] as well as a related oriented variant which turned out to perform better in our experiments. We start by recalling the concept of the Kendall shape space. We define

$$V_{n-1} = \{z \in \mathbb{C}^n : \sum_{i=1}^{n} z_i = 0\} \subset \mathbb{C}^n, \tag{2.46}$$

which can itself be identified with $\mathbb{C}^{n-1}$. We note that $z \in V_{n-1}$ can be scaled by a factor $s > 0$ and rotated by an angle $\theta \in [0, 2\pi)$ by multiplying all complex components z_i with the complex number $w = s\exp(i\theta) = s\cos(\theta) + is\sin(\theta)$. Consequently, all $z \in V_{n-1}$ which are equivalent w.r.t. rotation, and scaling lie on the complex line $L_z = \{w \cdot z : w \in \mathbb{C}\backslash\{0\}\}$. Then a shape L_z, or loosely speaking z, is an equivalence class in V_{n-1} representing a polygon described by points in $\mathbb{C}^n$ up to rigid transformations and scalings. The set of these shapes can thus be identified with the complex projective space $\mathbb{C}\mathrm{P}^{n-2}$ or, more intuitively, the complex unit sphere $\mathbb{S}_{\mathbb{C}}^{n-2}$ (with antipodal points identified). As a consequence, the exponential mapping and the inverse exponential mapping are given by the respective mappings of $\mathbb{S}_{\mathbb{C}}^{n-2}$, i.e.,

$$\exp_z(v) = \cos(\phi) \cdot z + \frac{\|z\| \sin(\phi)}{\phi} \cdot v, \phi = \|v\| \tag{2.47}$$

and

$$\log_z(y) = \phi \cdot \frac{y - \Pi_z(y)}{\|y - \Pi_z(y)\|}, \phi = \arccos\left(\frac{|\langle z, y\rangle|}{\|z\| \, \|y\|}\right), \tag{2.48}$$

where $\Pi_z(y) = z \cdot \langle z, y\rangle \,/\, \|z\|^2$ denotes the projection of y onto z. Here, $\langle \cdot, \cdot \rangle$ denotes the complex scalar product, i.e., $\langle z, y\rangle = \sum_{i=1}^{n} z_i \overline{y_i}$. and $\| \cdot \|$ the induced norm.

Based on the considerations above, we derive a shape representation which is not rotationally invariant. We term this representation *oriented Kendall shapes*. At first, we normalize $x = (x_1, \ldots, x_n)$, with all $x_i \in \mathbb{R}^2$, to obtain invariance w.r.t. translation. This is done by restriction to the real subspace

$$V_{2n-2} = \{x \in \mathbb{R}^{2n} : \sum_{i=1}^{n} (x_i^{(1)}, x_i^{(2)}) = 0\} \subset \mathbb{R}^{2n}. \tag{2.49}$$

Next, we notice that an element $x \in V_{2n-2}$ is scaled by multiplying all real components x_i with a real number $s \neq 0$. Consequently, in order to require scaling invariance only, we consider the equivalence classes $L_x = \{s \cdot x : s \in \mathbb{R}\backslash\{0\}\}$. Then a shape L_x, or loosely x, is an equivalence class in V_{2n-2} representing a polygon described by points in $\mathbb{R}^{2n}$ up to translation and scaling. Thus, the corresponding shape can be identified with the real projective space $\mathbb{R}\mathrm{P}^{2n-3}$ or, more intuitively, with the real unit sphere $\mathbb{S}_{\mathbb{R}}^{2n-3}$ with antipodal points identified. As a consequence, the exponential mapping and the inverse exponential mapping are given by the respective mappings of $\mathbb{S}_{\mathbb{R}}^{n-3}$, i.e., formulas (2.47) and (2.48) but this time with the real-valued scalar product $\langle x, y\rangle$ as well as its induced norm. Hence, all mappings can be implemented very efficiently using basic linear algebra subroutines.

In [7], we apply the proposed scheme to synthetic shape signals as well as to shape signals obtained from real imaging data. A natural real data application of our method is the processing of shape signals obtained from video segmentation

algorithms, in particular, in case of low resolution and low quality video data. Thus, we applied the recently proposed video segmentation algorithm of [PF13] to the "monkey" sequence of the SegTrack database[1] of [TFR10]. The results are shown in [7, Figure 4]. It can be observed that the segmentation boundaries are significantly regularized without deviating too much from the original segmentation. Furthermore, we observe that our method is robust w.r.t. perturbations of the initial shape data.; see [7, Table 2]. As further application, we consider geometry processing applications. A typical scenario appears in the context of slice-wise segmentation of organs. As example, we consider the segmentation of the abdominal part of the aorta from computed tomography angiography [7, Figure 5].

To conclude our discussion on TV regularization, we point out that TV regularization for $\mathbb{S}^1$-valued data is the topic of our paper [SWU16]. There, we develop a non-iterative scheme particularly tailored for TV minimization of $\mathbb{S}^1$-valued data.

Based on [6], we develop an algorithm for higher order TV type functionals for $\mathbb{S}^1$-valued data in [8]. We extend the TV functional considered in (2.38), here denoted by TV_1, by the second order terms TV_2, i.e., we consider the problem

$$J(x) = F(x) + \alpha\,\mathrm{TV}_1(x) + \beta\,\mathrm{TV}_2(x) \to \min, \tag{2.50}$$

where $\alpha\,\mathrm{TV}_1(x) = \alpha\sum_{i,j} G_{ij}(x) + \alpha\sum_{i,j} H_{ij}(x)$ is as in (2.38) and

$$\begin{aligned}\beta\,\mathrm{TV}_2(x) =& \beta_1 \sum_{i,j} d_2(x_{i-1,j}, x_{i,j}, x_{i+1,j}) + \beta_2 \sum_{i,j} d_2(x_{i,j-1}, x_{i,j}, x_{i,j+1}) \\ &+ \beta_3 \sum_{i,j} d_{1,1}(x_{i,j}, x_{i+1,j}, x_{i,j+1}, x_{i+1,j+1}).\end{aligned} \tag{2.51}$$

Here, $d_2, d_{1,1}$ are second order differences for $\mathbb{S}^1$-valued data defined as follows. Recall that, for real data, the absolute value of the diagonal second order difference $d_{1,1}$ w.r.t. the stencil $(-1,1)\otimes(-1,1)$ is given by $d_{1,1}(x_{i,j}, x_{i+1,j}, x_{i,j+1}, x_{i+1,j+1}) = |x_{i,j} - x_{i+1,j} - x_{i,j+1} + x_{i+1,j+1}|$. For circle valued data, we identify $\mathbb{S}^1$ with $[-\pi, \pi)$ and consider all possible liftings of the items $x_{k,l}$, $k \in \{i, i+1\}$, $l \in \{j, j+1\}$, i.e., $x_{k,l} + \mathbb{Z}2\pi$; for each such 4-tuple we take the absolute value of the diagonal second order difference in the sense of real valued data, and define $d_{1,1}(x_{i,j}, x_{i+1,j}, x_{i,j+1}, x_{i+1,j+1})$ as the minimum of these values. For the second order difference d_2 we proceed analogously basing on the real-valued definition $d_2(x_{i-1,j}, x_{i,j}, x_{i+1,j}) = |x_{i-1,j} - 2x_{i,j} + x_{i+1,j}|$. We adapt the cyclic proximal point scheme of [6] for the first order total variation problem by including the proximal mappings of the TV_2 atoms, i.e., we apply the proximal mapping of each summand in (2.51) within the cyclic scheme. To compute the proximal mappings of these atoms

[1] cpl.cc.gatech.edu/projects/SegTrack/

$d_2, d_{1,1}$, we need the minimizers w.r.t. x of the functional

$$\mathscr{E}(x,f,w) = \frac{1}{2}\sum_{j=1}^{r} d(f_j,x_j)^2 + \lambda d(x,w), \tag{2.52}$$

on $[-\pi,\pi)^r$, for $\lambda > 0$. Here, $w \in \{b_1, b_2, b_{1,1}\}$ where b_1, b_2 represents the second order difference stencils corresponding to $(1,-2,1)$ and $b_{1,1}$ is the mixed bivariate second order stencil $(-1,1)\otimes(-1,1)$. The symbol $d(x,w)$ is the second order difference w.r.t. the weight vector w given by the corresponding stencils; for instance, $d(x,(1,-2,1))$, $r=3$, corresponds to $d_2(x_{i-1,j}, x_{i,j}, x_{i+1,j})$ above. We get the following result which is stated as [8, Theorem 3.5].

Theorem 19 *For $w \in \{b_1, b_2, b_{1,1}\}$ set $s = \operatorname{sign}(\langle f, w\rangle)_{2\pi}$. Let $f \in [-\pi,\pi)^r$, where r is adapted to the respective length of w. If $|(\langle f,w\rangle)_{2\pi}| < \pi$, then the unique minimizer of $\mathscr{E}(x,f,w)$ is given by*

$$\hat{x} = (f - s m w)_{2\pi}, \qquad m = \min\left\{\lambda, \frac{|(\langle f,w\rangle)_{2\pi}|}{\|w\|_2^2}\right\}. \tag{2.53}$$

If $|(\langle f,w\rangle)_{2\pi}| = \pi$, then $\mathscr{E}(x,f,w)$ has the two minimizers

$$\hat{x} = (f \mp s m w)_{2\pi}, \qquad m = \min\left\{\lambda, \frac{\pi}{\|w\|_2^2}\right\}. \tag{2.54}$$

The proposed scheme is subsumed in Algorithm 1,(21),(22) in [8].

In [8], we show its convergence for $\mathbb{S}^1$-valued images under certain conditions. Our first condition is that the data $f \in (\mathbb{S}^1)^{N\times M}$ is dense enough. This means that the distance between neighboring pixels

$$d_\infty(f) := \max_{(i,j)} \max_{(k,l)\in\mathcal{N}_{i,j}} d(f_{i,j}, f_{k,l}) \tag{2.55}$$

is sufficiently small. Here $N_{i,j}$ is the four-neighborhood of (i,j). Similar conditions also appear in the convergence analysis of nonlinear subdivision schemes for manifold-valued data in [Wei10, WYW11]. We further require that the regularization parameters α, β in (2.50) are sufficiently small. We note that, for large parameters α, β, the solutions become almost constant and the model looses its interpretation which is an inherent problem due to the cyclic structure of the data. Finally, the parameter sequence $\{\lambda_k\}_k$ has to be square summable, but not summable, with a small ℓ^2 norm. The latter can be achieved by rescaling. The following statement is formulated as Theorem 4.10 in [8].

Theorem 20 *We consider an image f defined on an $N \times M$ grid with values in $\mathbb{S}^1$ with $d_\infty(f) < \frac{\pi}{8}$. Let $\{\lambda_k\}_k$ be a sequence of positive numbers which is square summable but not summable and*

$$\sqrt{\varepsilon^2 + 2\|\lambda\|_2^2 L^2 c(c+1)} + 2\|\lambda\|_\infty cL < \frac{\pi}{16}, \tag{2.56}$$

for some $\varepsilon > 0$, where $c = 15$ and $L = 4$. Further, assume that the parameters α, β of the functional J in (2.50) satisfy $(1,1)\,\mathrm{TV}_1(f) + (1,1,1)\,\mathrm{TV}_2(f) \leq \frac{\varepsilon^2}{m}$, where $m = \max(\alpha, \beta)$. Then the sequence $\{x^{(k)}\}_k$ generated by [8, Algorithm 1] (implemented via [8, Eq.(21), Eq.(22)]) converges to a minimizer of J.

Our convergence analysis is based on an unwrapping procedure combined with a convergence result in [Bač14]. In [8], we apply the proposed algorithm to interferometric SAR data and to electroencephalography data [MHCS$^+$15].

In [9], we consider the second order TV type problem (2.50) for more general manifolds. We first find a suitable geometric definition of the absolute values of second order differences $d_2, d_{1,1}$ in a Riemannian manifold. We let

$$d_2(x,y,z) = 2\,\mathrm{dist}(\mathrm{mid}(x,z),y), \quad d_{1,1}(x,y,z,v) = 2\,\mathrm{dist}(\mathrm{mid}(x,z),\mathrm{mid}(y,v)), \tag{2.57}$$

where the symbol $\mathrm{dist}(\cdot,\cdot)$ denotes the distance induced by the Riemannian metric and the symbol $\mathrm{mid}(\cdot,\cdot)$ denotes the midpoint between the corresponding arguments. If the midpoint is non-unique, we take the minimum w.r.t. all midpoints in (2.57). We note that this geometric definition is particularly appealing since it avoids using the tangent bundle. As a result, it is computationally accessible by the machinery of Jacobi fields which, in particular, in symmetric spaces yields rather explicit descriptions. We recall that the symmetric spaces form a class of manifolds including many spaces interesting in applications such as the spheres, the spaces of symmetric positive definite matrices, rotation groups, and motion groups, to mention only a few.

For the minimization of (2.50), we employ the cyclic proximal point strategy of [6] explained above. In addition to the proximal mappings for the TV terms discussed in [6], we have to compute the proximal mappings of the second order differences $d_2, d_{1,1}$. In contrast to [8] where we derived closed formulae, we here employ an approximate strategy for the proximal mappings of $d_2, d_{1,1}$. Since the proximal mappings of $d_2, d_{1,1}$ are not evaluated exactly, we obtain an inexact variant of the cyclic proximal point algorithm. We show the convergence of the proposed inexact variant of the cyclic proximal point algorithm in a Hadamard space. This result is formulated as Theorem 4.4 in [9].

Theorem 21 *Let H be a locally compact Hadamard space and let φ be a real-valued function given by $\varphi = \sum_{l=1}^{L} \varphi_l$, where all φ_l are convex continuous func-*

tions, and assume that φ attains a (global) minimum. Assume that for every starting point, the sequence generated by the exact cyclic PPA converges to a minimizer of φ. Let $\{x^{(k)}\}_{k\in\mathbb{N}}$ be the sequence generated by the inexact cyclic PPA, where $\sum_{k=0}^{\infty}\varepsilon_k < \infty$ and ε_k is the perturbation in the kth step. Then the sequence $\{x^{(k)}\}_{k\in\mathbb{N}}$ converges to a minimizer of φ.

We explain the approximate strategy we employ to compute the proximal mappings of $d_2, d_{1,1}$. In order to minimize the functional defining the proximal mapping of $d_2, d_{1,1}$, we use a (sub)gradient descent scheme. We proceed with explaining the procedure for d_2; the approach for $d_{1,1}$ is analogous. We notice that for $\operatorname{mid}(x,z) \neq y$ the subgradient of d_2 coincides with its gradient

$$\nabla_{M^3} d_2 = \left(\nabla_M d_2(\cdot,y,z), \nabla_M d_2(x,\cdot,z), \nabla_M d_2(x,y,\cdot)\right)^{\mathrm{T}}. \tag{2.58}$$

In the special case $\operatorname{mid}(x,z) = y$, d_2 is not differentiable. For a discussing of the corresponding subgradients, we refer to [9]. It turns out that the zero vector is a subgradient which allows us to use it in our subgradient descent algorithm. The gradient of the second component of (2.58) is directly obtained from the gradient of the Riemannian distance function. For computing the gradient of the first and third component of (2.58), we use the machinery of Jacobi fields: in order to differentiate the mapping $x \mapsto \operatorname{mid}(x,z)$, we consider the corresponding Jacobi differential equation with boundary conditions. We use the properties of symmetric spaces, which allow a rather explicit analytic solution of the corresponding initial value problem. Then, in symmetric spaces, relating the initial and the boundary value problem can be done explicitly avoiding inversion. These analytic ingredients avoid tedious numerical computations and so guarantee computational feasibility.

In [9], we illustrate the performance of the algorithm by applying it to sphere valued data and to data taking values in the manifold of symmetric positive definite matrices.

Bibliography

[AF05] D. Azagra and J. Ferrera. Proximal calculus on Riemannian manifolds. *Mediterranean Journal of Mathematics*, 2:437–450, 2005.

[AFP00] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of bounded variation and free discontinuity problems*. Clarendon Press Oxford, 2000.

[AFS13] M. Artina, M. Fornasier, and F. Solombrino. Linearly constrained nonsmooth and nonconvex minimization. *SIAM Journal on Optimization*, 23(3):1904–1937, 2013.

[AL89] I. Auger and C. Lawrence. Algorithms for the optimal identification of segment neighborhoods. *Bulletin of Mathematical Biology*, 51(1):39–54, 1989.

[All92] S. Alliney. Digital filters as absolute norm regularizers. *IEEE Transactions on Signal Processing*, 40:1548–1562, 1992.

[ALL+07] A. Alexander, J. Lee, M. Lazar, R. Boudos, M. DuBray, T. Oakes, J. Miller, J. Lu, E.-K. Jeong, W. McMahon, et al. Diffusion tensor imaging of the corpus callosum in autism. *Neuroimage*, 34(1):61–73, 2007.

[Amb89] L. Ambrosio. Variational problems in SBV and image segmentation. *Acta Applicandae Mathematicae*, 17(1):1–40, 1989.

[AP08] Y. Assaf and O. Pasternak. Diffusion tensor imaging (DTI)-based white matter mapping in brain research: a review. *Journal of Molecular Neuroscience*, 34(1):51–61, 2008.

[AS12] A. Anagaw and M. Sacchi. Edge-preserving seismic imaging using the total variation method. *Journal of Geophysics and Engineering*, 9:138, 2012.

[AT90] L. Ambrosio and V. Tortorelli. Approximation of functional depending on jumps by elliptic functional via Γ-convergence. *Communications on Pure and Applied Mathematics*, 43(8):999–1036, 1990.

[AW10] B. Alexeev and R. Ward. On the complexity of Mumford-Shah-type regularization, viewed as a relaxed sparsity constraint. *IEEE Transactions on Image Processing*, 19(10):2787–2789, 2010.

[Bač14] M. Bačák. Computing medians and means in Hadamard spaces. *SIAM Journal on Optimization*, 24(3):1542–1566, 2014.

[BBM14] M. Bauer, M. Bruveris, and P. Michor. Overview of the geometries of shape spaces and diffeomorphism groups. *Journal of Mathematical Imaging and Vision*, 50(1-2):60–97, 2014.

[BBMM14] M. Bauer, M. Bruveris, S. Marsland, and P. Michor. Constructing reparameterization invariant metrics on spaces of plane curves. *Differential Geometry and its Applications*, 34:139 – 165, 2014.

[BC98] P. Blomgren and T. Chan. Color TV: total variation methods for restoration of vector-valued images. *IEEE Transactions on Image Processing*, 7(3):304–309, 1998.

[BD08] T. Blumensath and M. Davies. Iterative thresholding for sparse approximations. *Journal of Fourier Analysis and Applications*, 14(5-6):629–654, 2008.

[BD09] T. Blumensath and M. Davies. Iterative hard thresholding for compressed sensing. *Applied and Computational Harmonic Analysis*, 27(3):265–274, 2009.

[Ber76] D. Bertsekas. Multiplier methods: a survey. *Automatica*, 12(2):133–145, 1976.

[Ber11] D. Bertsekas. Incremental proximal methods for large scale convex optimization. *Mathematical Programming*, 129:163–195, 2011.

[BGS85] W. Ballmann, M. Gromov, and V. Schroeder. *Manifolds of nonpositive curvature*. Birkhäuser, Boston, 1985.

[BH99] M. Bridson and A. Haefliger. *Metric spaces of non-positive curvature*. Springer, Berlin, 1999.

[BJBJ+07] T. Behrens, H. Johansen-Berg, S. Jbabdi, M. Rushworth, and M. Woolrich. Probabilistic diffusion tractography with multiple fibre orientations: what can we gain? *Neuroimage*, 34(1):144–155, 2007.

[BK04] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1124–1137, 2004.

[BKL+09] L. Boysen, A. Kempe, V. Liebscher, A. Munk, and O. Wittich. Consistencies and rates of convergence of jump-penalized least squares estimators. *Annals of Statistics*, 37(1):157–183, 2009.

[BKP10] K. Bredies, K. Kunisch, and T. Pock. Total generalized variation. *SIAM Journal on Imaging Sciences*, 3(3):492–526, 2010.

[Bla89] A. Blake. Comparison of the efficiency of deterministic and stochastic algorithms for visual reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(1):2–12, 1989.

[BLMW07] L. Boysen, V. Liebscher, A. Munk, and O. Wittich. Scale space consistency of piecewise constant least squares estimators: another look at the regressogram. *IMS Lecture Notes Monograph Series*, 55:65–84, 2007.

[BML94] P. Basser, J. Mattiello, and D. LeBihan. MR diffusion tensor spectroscopy and imaging. *Biophysical Journal*, 66(1):259–267, 1994.

[BP03] R. Bhattacharya and V. Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds I. *Annals of Statistics*, 31(1):1–29, 2003.

[BP05] R. Bhattacharya and V. Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds II. *Annals of Statistics*, 33(3):1225–1259, 2005.

[BPC+11] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.

[Bru65] J. Bruce. Optimum quantization. Technical report, MIT Research Laboratory of Electronics, 1965.

[BSK04] L. Bar, N. Sochen, and N. Kiryati. Variational pairing of image segmentation and blind restoration. In *European Conference on Computer Vision (ECCV)*, pages 166–177. Springer, 2004.

[BVZ01] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.

[BZ87] A. Blake and A. Zisserman. *Visual reconstruction*. MIT Press Cambridge, 1987.

[CD04] E. Candès and D. Donoho. New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities. *Communications on Pure and Applied Mathematics*, 57(2):219–266, 2004.

[CDM99] A. Chambolle and G. Dal Maso. Discrete approximation of the Mumford-Shah functional in dimension two. *ESAIM: Mathematical Modelling and Numerical Analysis*, 33(4):651–672, 1999.

[CE05] T. Chan and S. Esedoglu. Aspects of total variation regularized L^1 function approximation. *SIAM Journal on Applied Mathematics*, 65:1817–1837, 2005.

[CEP07] T. Chan, S. Esedoglu, and F. Park. Image decomposition combining staircase reduction and texture extraction. *Journal of Visual Communication and Image Representation*, 18(6):464–486, 2007.

[Cha95] A. Chambolle. Image segmentation by variational methods: Mumford and Shah functional and the discrete approximations. *SIAM Journal on Applied Mathematics*, 55(3):827–863, 1995.

[Cha99] A. Chambolle. Finite-differences discretizations of the Mumford-Shah functional. *ESAIM: Mathematical Modelling and Numerical Analysis*, 33(2):261–288, 1999.

[Cha04] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1):89–97, 2004.

[Cha09] R. Chartrand. Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 262–265, 2009.

[CJK09] C. Clason, B. Jin, and K. Kunisch. A duality-based splitting method for ℓ^1-TV image restoration with automatic regularization parameter choice. *SIAM Journal on Scientific Computing*, pages 1484–1505, 2009.

[CJK10] C. Clason, B. Jin, and K. Kunisch. A semismooth Newton method for L^1 data fitting with automatic choice of regularization parameters and noise calibration. *SIAM Journal on Imaging Sciences*, 3(2):199–231, 2010.

[CKS01] T. Chan, S. Kang, and J. Shen. Total variation denoising and enhancement of color images based on the CB and HSV color models. *Journal of Visual Communication and Image Representation*, 12:422–435, 2001.

[CL97] A. Chambolle and P.-L. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76:167–188, 1997.

[CMM00] T. Chan, A. Marquina, and P. Mulet. High-order total variation-based image restoration. *SIAM Journal on Scientific Computing*, 22(2):503–516, 2000.

[CMP+07] G. Charpiat, P. Maurel, J. Pons, R. Keriven, and O. Faugeras. Generalized gradients: priors on minimization flows. *International Journal on Computer Vision*, 73(3):325–344, 2007.

[CP11a] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

[CP11b] P. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In *Fixed-point algorithms for inverse problems in science and engineering*, pages 185–212. Springer, 2011.

[CRD07] D. Cremers, M. Rousson, and R. Deriche. A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape. *International Journal of Computer Vision*, 72(2):195–215, 2007.

[CRT06] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006.

[CS13] D. Cremers and E. Strekalovskiy. Total cyclic variation and generalizations. *Journal of Mathematical Imaging and Vision*, 47(3):258–277, 2013.

[CSV12] G. Cheng, H. Salehian, and B. Vemuri. Efficient recursive algorithms for computing the mean diffusion tensor and applications to DTI segmentation. In *European Conference on Computer Vision (ECCV)*, pages 390–401. Springer, 2012.

[CTDF04] C. Chefd'Hotel, D. Tschumperlé, R. Deriche, and O. Faugeras. Regularizing flows for constrained matrix-valued images. *Journal of Mathematical Imaging and Vision*, 20(1-2):147–162, 2004.

[CV01] T. Chan and L. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, 2001.

[CWB08] E. Candès, M. Wakin, and S. Boyd. Enhancing sparsity by reweighted ℓ^1 minimization. *Journal of Fourier Analysis and Applications*, 14(5):877–905, 2008.

[CYZ+06] T. Chen, W. Yin, X. Zhou, D. Comaniciu, and T. Huang. Total variation models for variable lighting face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28:1519–1524, 2006.

[DAFD07] M. Descoteaux, E. Angelino, S. Fitzgibbons, and R. Deriche. Regularized, fast, and robust analytical Q-ball imaging. *Magnetic Resonance in Medicine*, 58(3):497–510, 2007.

[DBFZ+06] N. Dey, L. Blanc-Feraud, C. Zimmer, P. Roux, Z. Kam, J.-C. Olivo-Marin, and J. Zerubia. Richardson-Lucy algorithm with total variation regularization for 3D confocal microscope deconvolution. *Microscopy Research and Technique*, 69:260–266, 2006.

[dC92] M. do Carmo. *Riemannian Geometry*. Birkhäuser, Boston, 1992.

[DDDM04] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on pure and applied mathematics*, 57(11):1413–1457, 2004.

[DeV98] R. DeVore. Nonlinear approximation. *Acta numerica*, 7:51–150, 1998.

[DG91] E. De Giorgi. Free discontinuity problems in calculus of variations. *Frontiers in pure and applied Mathematics, a collection of papers dedicated to J.L. Lions on the occasion of his 60th birthday, R. Dautray ed., North Holland*, 1991.

[DGTSU12] R. Delgado-Gonzalo, P. Thevenaz, C. Seelamantula, and M. Unser. Snakes with an ellipse-reproducing property. *IEEE Transactions on Image Processing*, 21(3):1258–1271, 2012.

[DHN09] Y. Dong, M. Hintermüller, and M. Neri. An efficient primal-dual method for l^1-*tv* image restoration. *SIAM Journal on Imaging Sciences*, 2:1168–1189, 2009.

[DMH+03] A. Drobyshev, C. Machka, M. Horsch, M. Seltmann, V. Liebscher, M. Hrabe de Angelis, and J. Beckers. Specificity assessment from fractionation experiments (SAFE): a novel method to evaluate microarray probe specificity based on hybridisation stringencies. *Nucleic Acids Research*, 31(2):1–10, 2003.

[Don99] D. Donoho. Wedgelets: Nearly minimax estimation of edges. *Annals of Statistics*, 27(3):859–897, 1999.

[Don06] D. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[DWB09] S. Didas, J. Weickert, and B. Burgeth. Properties of higher order nonlinear diffusion filtering. *Journal of Mathematical Imaging and Vision*, 35:208–226, 2009.

[Ela10] M. Elad. *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer Verlag, 2010.

[FHM14] S. Frick, T. Hohage, and A. Munk. Asymptotic laws for change point estimation in inverse regression. *Statistica Sinica*, 24:555–575, 2014.

[FJ07] P. Fletcher and S. Joshi. Riemannian geometry for the statistical analysis of diffusion tensor data. *Signal Processing*, 87(2):250–262, 2007.

[FKLW08] F. Friedrich, A. Kempe, V. Liebscher, and G. Winkler. Complexity penalized M-estimation: fast computation. *Journal of Computational and Graphical Statistics*, 17(1):201–224, 2008.

[Fle13] P. Fletcher. Geodesic regression and the theory of least squares on Riemannian manifolds. *International Journal of Computer Vision*, 105(2):171–185, 2013.

[FLPJ04] P. Fletcher, C. Lu, S. Pizer, and S. Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions on Medical Imaging*, 23(8):995–1005, 2004.

[FMC+00] J. Foong, M. Maier, C. Clark, G. Barker, D. Miller, and M. Ron. Neuropathological abnormalities of the corpus callosum in schizophrenia: a diffusion tensor imaging study. *Journal of Neurology, Neurosurgery & Psychiatry*, 68(2):242–244, 2000.

[FMS13] M. Fornasier, R. March, and F. Solombrino. Existence of minimizers of the Mumford-Shah functional with singular operators and unbounded data. *Annali di Matematica Pura ed Applicata*, 192(3):361–391, 2013.

[FMS14] K. Frick, A. Munk, and H. Sieling. Multiscale change point inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(3):495–580, 2014.

[FO02] O. Ferreira and P. Oliveira. Proximal point algorithm on Riemannian manifolds. *Optimization*, 51:257–270, 2002.

[FR08a] M. Fornasier and H. Rauhut. Iterative thresholding algorithms. *Applied and Computational Harmonic Analysis*, 25(2):187–208, 2008.

[FR08b] M. Fornasier and H. Rauhut. Recovery algorithms for vector-valued data with joint sparsity constraints. *SIAM Journal on Numerical Analysis*, 46(2):577–613, 2008.

[FW10] M. Fornasier and R. Ward. Iterative thresholding meets free-discontinuity problems. *Foundations of Computational Mathematics*, 10(5):527–567, 2010.

[GG84] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984.

[GHS15] P. Grohs, H. Hardering, and O. Sander. Optimal a priori discretization error bounds for geodesic finite elements. *Foundations of Computational Mathematics*, 15(6):1357–1411, 2015.

[GK14] P. Grohs and G. Kutyniok. Parabolic molecules. *Foundations of Computational Mathematics*, 14(2):299–337, 2014.

[GLTV09] A. Goh, C. Lenglet, P. Thompson, and R. Vidal. A nonparametric Riemannian framework for processing high angular resolution diffusion images (HARDI). In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2496–2503, 2009.

[GM01] Y. Gousseau and J.-M. Morel. Are natural images of bounded variation? *SIAM Journal on Mathematical Analysis*, 33:634–648, 2001.

[GM06] M. Giaquinta and D. Mucci. The BV-energy of maps into a manifold: relaxation and density results. *Annali della Scuola Normale Superiore di Pisa-Classe di Scienze*, 5(4):483–548, 2006.

[GM07] M. Giaquinta and D. Mucci. Maps of bounded variation with values into a manifold: total variation and relaxed energy. *Pure and Applied Mathematics Quarterly*, 3(2):513–538, 2007.

[GMS93] M. Giaquinta, G. Modica, and J. Souček. Variational problems for maps of bounded variation with values in S^1. *Calculus of Variations and Partial Differential Equations*, 1(1):87–121, 1993.

[GO09] T. Goldstein and S. Osher. The split bregman method for l1-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009.

[GW09] P. Grohs and J. Wallner. Interpolatory wavelets for manifold-valued data. *Applied and Computational Harmonic Analysis*, 27(3):325–333, 2009.

[HMH+06] C. Hess, P. Mukherjee, E. Han, D. Xu, and D. Vigneron. Q-ball reconstruction of multimodal fiber orientations using the spherical harmonic basis. *Magnetic Resonance in Medicine*, 56(1):104–117, 2006.

[HS06] W. Hinterberger and O. Scherzer. Variational methods on the space of functions of bounded Hessian for convexification and denoising. *Computing*, 76(1):109–133, 2006.

[HST+04] P. Hupé, N. Stransky, J. Thiery, F. Radvanyi, and E. Barillot. Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. *Bioinformatics*, 20(18):3413–3422, 2004.

[HSW15] K. Hohm, M. Storath, and A. Weinmann. An algorithmic framework for Mumford–Shah regularization of inverse problems in imaging. *Inverse Problems*, 31(11):115011, 2015.

[JBB09] H. Johansen-Berg and T. Behrens. *Diffusion MRI: From quantitative measurement to in-vivo neuroanatomy*. Academic Press, London, 2009.

[JBI+08] C. Joo, H. Balci, Y. Ishitsuka, C. Buranachai, and T. Ha. Advances in single-molecule fluorescence methods for molecular biology. *Annual Review of Biochemistry*, 77:51–76, 2008.

[JMP14] M. Jiang, P. Maaß, and T. Page. Regularizing properties of the Mumford-Shah functional for imaging applications. *Inverse Problems*, 30(3):035007, 2014.

[Ken84] D. Kendall. Shape manifolds, procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society*, 16(2):81–121, 1984.

[Kir94] B. Kirchheim. Rectifiable metric spaces: local structure and regularity of the Hausdorff measure. *Proceedings of the American Mathematical Society*, 121(1):113–123, 1994.

[Kla11] E. Klann. A Mumford-Shah-like method for limited data tomography with an application to electron tomography. *SIAM Journal on Imaging Sciences*, 4(4):1029–1048, 2011.

[KMW+07] M. Kubicki, R. McCarley, C.-F. Westin, H.-J. Park, S. Maier, R. Kikinis, F. Jolesz, and M. Shenton. A review of diffusion tensor imaging studies in schizophrenia. *Journal of Psychiatric Research*, 41(1):15–30, 2007.

[KRR11] E. Klann, R. Ramlau, and W. Ring. A Mumford-Shah level-set approach for the inversion and segmentation of SPECT/CT data. *Inverse Problems and Imaging*, 5(1):137–166, 2011.

[KS02] R. Kimmel and N. Sochen. Orientation diffusion or how to comb a porcupine. *Journal of Visual Communication and Image Representation*, 13(1):238–248, 2002.

[KTCW02] J. Kim, A. Tsai, M. Cetin, and A. Willsky. A curve evolution-based variational approach to simultaneous image restoration and segmentation. In *IEEE International Conference on Image Processing (ICIP)*, pages I109–I112, 2002.

[KWT88] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.

[LBU12] S. Lefkimmiatis, A. Bourquard, and M. Unser. Hessian-based norm regularization for image restoration with biomedical applications. *IEEE Transactions on Image Processing*, 21(3):983–995, 2012.

[LJ11a] M. Little and N. Jones. Generalized methods and solvers for noise removal from piecewise constant signals. I. Background theory. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 467(2135):3088–3114, 2011.

[LJ11b] M. Little and N. Jones. Generalized methods and solvers for noise removal from piecewise constant signals. II. New methods. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 467(2135):3115–3140, 2011.

[LKY+09] J. Lellmann, J. Kappes, J. Yuan, F. Becker, and C. Schnörr. Convex multiclass image labeling by simplex-constrained total variation. *Scale Space and Variational Methods in Computer Vision*, pages 150–162, 2009.

[LLKW05] D. Labate, W. Lim, G. Kutyniok, and G. Weiss. Sparse multidimensional representation using shearlets. In *Optics & Photonics 2005*, pages 59140U–59140U. International Society for Optics and Photonics, 2005.

[LLT03] M. Lysaker, A. Lundervold, and X.-C. Tai. Noise removal using fourth-order partial differential equations with applications to medical magnetic resonance images in space and time. *IEEE Transactions on Image Processing*, 12(12):1579–1590, 2003.

[LO14] R. Lai and S. Osher. A splitting method for orthogonality constrained problems. *Journal of Scientific Computing*, 58(2):431–449, 2014.

[LSKC13] J. Lellmann, E. Strekalovskiy, S. Koetter, and D. Cremers. Total variation regularization for functions with values in a manifold. In *International Conference on Computer Vision (ICCV)*, pages 2944–2951, 2013.

[LT06] M. Lysaker and X.-C. Tai. Iterative image restoration combining total variation minimization and a second-order functional. *International Journal of Computer Vision*, 66(1):5–18, 2006.

[MF98] D. Massonnet and K. Feigl. Radar interferometry and its application to changes in the earth's surface. *Reviews of Geophysics*, 36(4):441–500, 1998.

[MHCS+15] Z. Mortezapouraghdam, L. Haab, F. Corona-Strauss, G. Steidl, and D. Strauss. Assessment of long-term habituation correlates in event-related potentials using a von Mises model. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 23(3):363–373, 2015.

[MM06] P. Michor and D. Mumford. Riemannian geometries on the space of plane curves. *Journal of the European Mathematical Society*, 8:1–48, 2006.

[MM07] P. Michor and D. Mumford. An overview of the Riemannian metrics on spaces of curves using the Hamiltonian approach. *Applied and Computational Harmonic Analysis*, 23(1):74 – 113, 2007.

[Mor62] J.-J. Moreau. Fonctions convexes duales et points proximaux dans un espace hilbertien. *Comptes Rendus de l'Académie des Sciences. Series A Mathematics.*, 255:2897–2899, 1962.

[MS85] D. Mumford and J. Shah. Boundary detection by minimizing functionals. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 137–154, 1985.

[MS89] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42(5):577–685, 1989.

[Nik02] M. Nikolova. Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers. *SIAM Journal on Numerical Analysis*, 40:965–994, 2002.

[Nik04] M. Nikolova. A variational approach to remove outliers and impulse noise. *Journal of Mathematical Imaging and Vision*, 20:99–120, 2004.

[NNT10] M. Nikolova, M. Ng, and C.-P. Tam. Fast nonconvex nonsmooth minimization methods for image restoration and reconstruction. *IEEE Transactions on Image Processing*, 19(12):3073–3088, 2010.

[NNZC08] M. Nikolova, M. Ng, S. Zhang, and W. Ching. Efficient reconstruction of piecewise constant images using nonsmooth nonconvex minimization. *SIAM Journal on Imaging Sciences*, 1(1):2–25, 2008.

[NW13] D. Needell and R. Ward. Stable image reconstruction using total variation minimization. *SIAM Journal on Imaging Sciences*, 6(2):1035–1058, 2013.

[OC95] J. Oller and J. Corcuera. Intrinsic analysis of statistical estimation. *Annals of Statistics*, 23(5):1562–1581, 1995.

[PCBC09] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. An algorithm for minimizing the Mumford-Shah functional. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1133–1140, 2009.

[PCXD99] P. Petrushev, A. Cohen, H. Xu, and R. DeVore. Nonlinear approximation and the space $BV(\mathbb{R}^2)$. *American Journal of Mathematics*, 121:587–628, 1999.

[Pen06] X. Pennec. Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127–154, 2006.

[Pet06] A. Petukhov. Fast implementation of orthogonal greedy algorithm for tight wavelet frames. *Signal processing*, 86(3):471–479, 2006.

[PF13] A. Papazoglou and V. Ferrari. Fast object segmentation in unconstrained video. In *International Conference on Computer Vision (ICCV)*, pages 1777–1784, 2013.

[PFA06] X. Pennec, P. Fillard, and N. Ayache. A Riemannian framework for tensor computing. *International Journal of Computer Vision*, 66(1):41–66, 2006.

[Pot52] R. Potts. Some generalized order-disorder transformations. *Mathematical Proceedings of the Cambridge Philosophical Society*, 48(1):106–109, 1952.

[Rao45] C. Rao. Information and accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, 37(3):81–91, 1945.

[RBB+12] G. Rosman, M. Bronstein, A. Bronstein, A. Wolf, and R. Kimmel. Group-valued regularization framework for motion segmentation of dynamic non-rigid shapes. In *Scale Space and Variational Methods in Computer Vision*, pages 725–736. Springer, 2012.

[RLB+10] H. Rosas, S. Lee, A. Bender, A. Zaleta, M. Vangel, P. Yu, B. Fischl, V. Pappu, C. Onorato, J.-H. Cha, et al. Altered white matter microstructure in the corpus callosum in Huntington's disease: implications for cortical "disconnection". *Neuroimage*, 49(4):2995–3004, 2010.

[ROF92] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.

[Ron08] L. Rondi. On the regularization of the inverse conductivity problem with discontinuous conductivities. *Inverse Problems and Imaging*, 2(3):397–409, 2008.

[RR07] R. Ramlau and W. Ring. A Mumford-Shah level-set approach for the inversion and segmentation of X-ray tomography data. *Journal of Computational Physics*, 221(2):539–557, 2007.

[RR10] R. Ramlau and W. Ring. Regularization of ill-posed Mumford–Shah models with perimeter penalization. *Inverse Problems*, 26(11):115001, 2010.

[RS01] L. Rondi and F. Santosa. Enhanced electrical impedance tomography via the Mumford–Shah functional. *ESAIM: Control, Optimisation and Calculus of Variations*, 6:517–538, 2001.

[RWT+14] G. Rosman, Y. Wang, X.-C. Tai, R. Kimmel, and A. Bruckstein. Fast regularization of matrix-valued images. In *Efficient Algorithms for Global Optimization Methods in Computer Vision*, pages 19–43. Springer, 2014.

[SB08] Y. Sowa and R. Berry. Bacterial flagellar motor. *Quarterly Reviews of Biophysics*, 41(2):103–132, 2008.

[SC03] D. Strong and T. Chan. Edge-preserving and scale-dependent properties of total variation regularization. *Inverse Problems*, 19:S165, 2003.

[SC11] E. Strekalovskiy and D. Cremers. Total variation for cyclic structures: Convex relaxation and efficient minimization. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1905–1911, 2011.

[SCC12] E. Strekalovskiy, A. Chambolle, and D. Cremers. A convex representation for the vectorial Mumford-Shah functional. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1712–1719, 2012.

[Sch98] O. Scherzer. Denoising with higher order derivatives of bounded variation and an application to parameter estimation. *Computing*, 60:1–27, 1998.

[SKJJ11] A. Srivastava, E. Klassen, S. Joshi, and I. Jermyn. Shape analysis of elastic curves in Euclidean spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(7):1415–1428, 2011.

[SMSY09] G. Sundaramoorthi, A. Mennucci, S. Soatto, and A. Yezzi. Tracking deforming objects by filtering and prediction in the space of curves. In *IEEE Conference on Decision and Control*, 2009.

[SNS+01] A. Snijders, N. Nowak, R. Segraves, et al. Assembly of microarrays for genome-wide measurement of DNA copy number by CGH. *Nature Genetics*, 29:263–264, 2001.

[SRL+05] Y. Sowa, A. Rowe, M. Leake, T. Yakushi, M. Homma, A. Ishijima, and R. Berry. Direct observation of steps in rotation of the bacterial flagellar motor. *Nature*, 437(7060):916–919, 2005.

[SSM+11] R. Schönmeyer, G. Schmidt, S. Meding, A. Walch, and G. Binnig. Automated co-analysis of MALDI- and H&E-images of retinal tissue for an improved spatial MALDI resolution. In *53rd Symposium of the Society for Histochemistry*, pages 54–55, 2011.

[SST11] S. Setzer, G. Steidl, and T. Teuber. Infimal convolution regularizations with discrete l1-type functionals. *Communications in Mathematical Sciences*, 9(3):797–872, 2011.

[Stu03] K.-T. Sturm. Probability measures on metric spaces of nonpositive curvature. In *Heat kernels and analysis on manifolds, graphs, and metric spaces*, pages 357–390. American Mathematical Society, Providence, 2003.

[SWB+04] G. Steidl, J. Weickert, T. Brox, P. Mrázek, and M. Welk. On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and SIDEs. *SIAM Journal on Numerical Analysis*, 42:686–713, 2004.

[SWFU15] M. Storath, A. Weinmann, J. Frikel, and M. Unser. Joint image reconstruction and segmentation using the Potts model. *Inverse Problems*, 31(2):025003, 2015.

[SWU14] M. Storath, A. Weinmann, and M. Unser. Unsupervised texture segmentation using monogenic curvelets and the Potts model. In *IEEE International Conference on Image Processing (ICIP)*, pages 4348–4352, 2014.

[SWU16] M. Storath, A. Weinmann, and M. Unser. Exact algorithms for L^1-TV regularization of real-valued or circle-valued signals. *SIAM Journal on Scientific Computing*, 38(1):A614–A630, 2016.

[SYM08] G. Sundaramoorthi, A. Yezzi, and A. Mennucci. Coarse-to-fine segmentation and tracking using Sobolev active contours. *IEEE Transactions on Pattern Analysis and Machine Intelligence,*, 30(5):851 –864, 2008.

[SYMS09] G. Sundaramoorthi, A. Yezzi, A. Mennucci, and G. Sapiro. New possibilities with Sobolev active contours. *International Journal of Computer Vision*, 84(2):113–129, 2009.

[SZS+08] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother. A comparative study of energy minimization methods for Markov random fields with smoothness-based priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(6):1068–1080, 2008.

[TD01] D. Tschumperlé and R. Deriche. Diffusion tensor regularization with constraints preservation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages I948–I953, 2001.

[Tem03] V. Temlyakov. Nonlinear methods of approximation. *Foundations of Computational Mathematics*, 3(1):33–107, 2003.

[TFR10] D. Tsai, M. Flagg, and J. Rehg. Motion coherent tracking with multi-label MRF optimization. In *British Machine Vision Conference*, 2010.

[Tro04] J. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Transactions on Information Theory*, 50(10):2231–2242, 2004.

[Tuc04] D. Tuch. Q-ball imaging. *Magnetic Resonance in Medicine*, 52(6):1358–1372, 2004.

[URDDS05] I. Ur Rahman, V. Drori, I.and Stodden, D. Donoho, and P. Schröder. Multiscale representations for manifold-valued data. *Multiscale Modeling & Simulation*, 4(4):1201–1232, 2005.

[VC02] L. Vese and T. Chan. A multiphase level set framework for image segmentation using the Mumford and Shah model. *International Journal of Computer Vision*, 50(3):271–293, 2002.

[Vek99] O. Veksler. *Efficient graph-based energy minimization methods in computer vision*. PhD thesis, Cornell University, 1999.

[VO02] L. Vese and S. Osher. Numerical methods for p-harmonic flows and applications to image processing. *SIAM Journal on Numerical Analysis*, 40:2085–2104, 2002.

[Wei10] A. Weinmann. Nonlinear subdivision schemes on irregular meshes. *Constructive Approximation*, 31(3):395–415, 2010.

[Wei12a] A. Weinmann. Interpolatory multiscale representation for functions between manifolds. *SIAM Journal on Mathematical Analysis*, 44:162–191, 2012.

[Wei12b] A. Weinmann. Subdivision schemes with general dilation in the geometric and nonlinear setting. *Journal of Approximation Theory*, 164(1):105–137, 2012.

[WL02] G. Winkler and V. Liebscher. Smoothers for discontinuous signals. *Journal of Nonparametric Statistics*, 14(1-2):203–222, 2002.

[WV04] Z. Wang and B. Vemuri. An affine invariant tensor dissimilarity measure and its applications to tensor-valued image segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages I228–I233, 2004.

[WV05] Z. Wang and B. Vemuri. DTI segmentation using an information theoretic tensor dissimilarity measure. *IEEE Transactions on Medical Imaging*, 24(10):1267–1277, 2005.

[WYW11] J. Wallner, E. Yazdani, and A. Weinmann. Convergence and smoothness analysis of subdivision rules in Riemannian and symmetric spaces. *Advances in Computational Mathematics*, 34(2):201–218, 2011.

[YMSM07] L. Younes, P. Michor, J. Shah, and D. Mumford. A metric on shape spaces with explicit geodesics. *Technical report*, 2007.

[YZ11] J. Yang and Y. Zhang. Alternating direction algorithms for ℓ_1-problems in compressive sensing. *SIAM Journal on Scientific Computing*, 33(1):250–278, 2011.

[YZY10] J. Yang, Y. Zhang, and W. Yin. A fast alternating direction method for TV L1-L2 signal reconstruction from partial Fourier data. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):288–297, 2010.

[ZLMY+02] C. Zimmer, E. Labruyere, V. Meas-Yedid, N. Guillen, and J.-C. Olivo-Marin. Segmentation and tracking of migrating cells in videomicroscopy with parametric active contours: a tool for cell-based drug testing. *IEEE Transactions on Medical Imaging*, 21(10):1212–1221, 2002.

[ZPB07] C. Zach, T. Pock, and H. Bischof. A duality based approach for realtime TV-L^1 optical flow. In *Pattern Recognition*, pages 214–223. Springer, 2007.