

Total Variation Regularization for Functions with Values in a Manifold

Jan Lellmann

Department of Applied Mathematics
and Theoretical Physics,
University of Cambridge

Evgeny Strekalovskiy Sabrina Koetter Daniel Cremers

Department of Computer Science
Technical University of Munich

Abstract

While total variation is among the most popular regularizers for variational problems, its extension to functions with values in a manifold is an open problem. In this paper, we propose the first algorithm to solve such problems which applies to arbitrary Riemannian manifolds. The key idea is to reformulate the variational problem as a multi-label optimization problem with an infinite number of labels. This leads to a hard optimization problem which can be approximately solved using convex relaxation techniques. The framework can be easily adapted to different manifolds including spheres and three-dimensional rotations, and allows to obtain accurate solutions even with a relatively coarse discretization. With numerous examples we demonstrate that the proposed framework can be applied to variational models that incorporate chromaticity values, normal fields, or camera trajectories.

1. Introduction

1.1. Total Variation for Manifold-Valued Functions

For functions $u : \Omega \rightarrow \mathbb{R}^l$, $\Omega \subseteq \mathbb{R}^d$, the total variation

$$TV(u) = \sup_{p \in C_c^\infty(\Omega, \mathbb{R}^l \times \mathbb{R}^d), \|p\|_\infty \leq 1} \int_\Omega \langle u, \text{Div } p \rangle dx, \quad (1)$$

plays a central role in variational image processing because of its numerous favorable properties: it *preserves discontinuities*, as a sharp transition from 0 to 1 has the same cost as a smooth monotone transition. It is also *convex* and thus amenable to efficient and globally optimal solutions. Moreover, applied to the indicator function of a set it gives the perimeter of that set, which makes it well suited for *geometric optimization problems*.

In many applications of computer vision, however, the functions of interest take on values which do not lie in a Euclidean space such as \mathbb{R}^l , but rather on a *manifold* – see Figure 1. Smoothing a camera trajectory, for example, requires to define regularizers for functions $u : \Omega \rightarrow SE(3)$

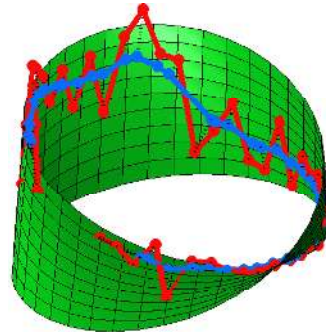


Figure 1. **Denoising of a function with values in a manifold.** We propose an algorithm for total variation (TV)-regularization of functions with values on arbitrary Riemannian manifolds. This example shows the TV-denoising (**blue**) of a one-dimensional function (**red**) $u : [0, 1] \rightarrow \mathcal{M}$ with values in the Moebius band \mathcal{M} . Characteristic of TV-denoising is the reduction of contrast, which corresponds to the shrinking of the curve. The key idea is to represent the manifold by a set of grid points (labels) and to solve a convex optimization problem that admits *sub-label accuracy*: in the above example, the blue points may lie between grid-points.

with values in the special Euclidean group $SE(3)$. Processing the normal field of a geometric structure, such as a surface or three-dimensional shape, leads to variational problems in terms of functions that assume values on the two-dimensional sphere S^2 .

Such problems with constraints on the *range* are considerably harder than problems in which the *domain* Ω of the function is a manifold. While the latter can usually be dealt with by a suitable modification of the differential operators, constraining the range of u to a manifold is generally a non-convex constraint, which makes optimization much harder.

Another major challenge in extending the concept of total variation to such manifold-valued data is that we need to assure algorithmically that discontinuities in the values are properly handled. In particular, jumps in the values should be measured with respect to the *geodesic* distance on that manifold. For general manifolds the estimation of the geodesic distance itself may also be a challenging computational problem.

1.2. Related Work

Giaquinta and Mucci [8, 9] studied the notion of total variation for functions $u : \Omega \rightarrow \mathcal{M}$ with values in general manifolds \mathcal{M} . They defined it as¹

$$TV_{\mathcal{M}}(u) = \int_{\Omega \setminus S_u} |\nabla u| dx + \int_{S_u} d_{\mathcal{M}}(u^-, u^+) d\mathcal{H}^{d-1}. \quad (2)$$

It provides a separation of the differentiable part and the jump part where the jump set S_u is penalized with the *geodesic distance* $d_{\mathcal{M}}(u^-, u^+)$ between the two values u^- and u^+ on either side of the jump. The authors theoretically study the functions with bounded total variation and prove several properties such as lower-semicontinuity and structure theorems. However, they do not provide an algorithm for implementing this regularizer in a variational setting.

In fact, their theoretical analysis is based on an embedding of the given manifold \mathcal{M} in a higher-dimensional Euclidean space which is infeasible to implement: Firstly, the higher-dimensional embedding space increases the computational complexity. Secondly, one cannot numerically constrain the values of the estimated solution to the generally non-convex manifold \mathcal{M} .

For the special case of the circle $\mathcal{M} = S^1$, Giaquinta et al. [7] prove the existence of minimizers for certain energies in the space of functions with bounded *total cyclic variation*, again using an embedding in the Euclidean plane. For the case $\mathcal{M} = S^1$, Cremers and Strekalovskiy [5] recently proposed an implementation of various models for cyclic data, including total variation, quadratic, Huber-TV and Mumford-Shah regularization. To this end, they unwrapped the values of u to the real axis \mathbb{R} and proposed an efficient algorithmic solution to account for the periodicity.

Unfortunately, this solution does not extend to general Riemannian manifolds, as unwrapping the manifold to a simple Euclidean one is typically not feasible. Moreover their representation is based on identifying u with the *sub-level set* $\{(x, t) | u(x) > t\}$, which is not available for higher-dimensional manifolds.

Inspired by Ishikawa’s graph-theoretic solution to spatially discrete multi-label optimization [12], Chambolle et al. [3], Zach et al. [22], and Lellmann et al. [15, 17] proposed relaxations of the labeling problem on continuous domains that allow to find good – and often globally optimal – solutions using convex optimization, see also [20, 1]. While these approaches do consider a continuous domain, the set of feasible labels remains a finite discrete set.

To address this problem and devise an algorithm which admits an infinite number of labels representing a manifold, we build on the formulation proposed in [17]: for a finite

¹For simplicity, we assume that the Cantor part and vanishes constrain our analysis to the *special* functions of bounded variation $SBV(\Omega; \mathcal{M})$.

label set $\mathcal{J} = \{1, \dots, l\}$, a metric $d : \mathcal{J}^2 \rightarrow \mathbb{R}$, and unary costs $s : \Omega \rightarrow \mathbb{R}^l$, they solve

$$\begin{aligned} \min_{u' : \Omega \rightarrow \Delta_l} \sup_{p : \Omega \rightarrow \mathbb{R}^{l \times d}} \int_{\Omega} \langle u', s \rangle dx + \lambda \int_{\Omega} \langle u', \text{Div } p \rangle dx, \quad (3) \\ \text{s.t. } \|p^{i_1}(x) - p^{i_2}(x)\|_2 \leq d(i_1, i_2) \quad \forall i_1, i_2 \in \mathcal{J}, \forall x \in \Omega. \quad (4) \end{aligned}$$

where Δ_l is the l -dimensional unit simplex, and $\text{Div } p = (\text{div } p^1, \dots, \text{div } p^l)$. This constitutes a *convex relaxation* of the finite labeling problem similar to Linear Programming (LP) relaxation [19, 21]: each of the values in \mathcal{J} is associated with one of the unit vectors $e^1, \dots, e^l \in \mathbb{R}^l$, but intermediate values are allowed in order to obtain a convex problem. The regularizer in (3) can be seen as total variation for functions with values in the finite set \mathcal{J} , where the metric d controls the weighting of a jump from label i_1 to label i_2 . We note that recently extensions to the non-metric case have also been proposed [20].

A straightforward approach to apply this idea to manifolds is to choose a finite set of points $z^1, \dots, z^l \in \mathcal{M}$ and to set $d(i, j) = d_{\mathcal{M}}(z^i, z^j)$. In fact, with these definitions J implements $TV_{\mathcal{M}}$ if one restricts u to the values $\{z^1, \dots, z^l\}$ in (2).

This approach has two major drawbacks: Firstly, it generally requires a quadratic number of constraints, which severely restricts the resolution with which the manifold can be discretized. Secondly, it is in fact designed so that minimizers u' are likely to only assume values in $\{e^1, \dots, e^l\}$, effectively limiting the range of u to $\{z^1, \dots, z^l\}$.

While for finite labeling problems this is desirable, it sets a hard limit for the accuracy that can be expected when applying it to manifolds. This issue is inherent of the approach (3)–(4), as the latter provides no means to incorporate knowledge about the local structure of the manifold.

1.3. Contribution

In this paper, we present a framework and algorithmic solution for regularization of signals with values in an arbitrary Riemannian manifold. The key idea is to consider the problem as a multi-labeling problem as in (3), but to extend this approach to an *infinite* number of labels.

This allows to derive an improved formulation of the dual constraints (4) that specifically respects the local manifold structure, and at the same time requires only a linear number of constraints.

Together, these features ensure that the minimization problem is computationally feasible, and allow to obtain accurate solutions with values on the manifold that are not restricted to a finite set.

We validate our method on a variety of inverse problems, including the denoising of chromaticity values, the inpainting and denoising of normal fields and the denoising of camera trajectories.

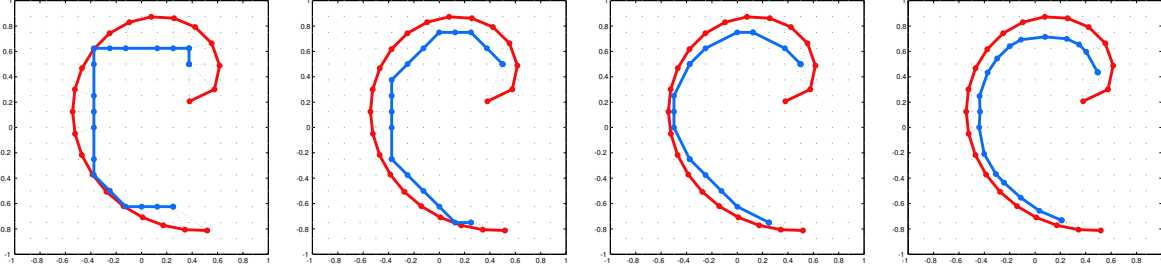


Figure 2. Rudin-Osher-Fatemi (ROF) denoising (blue) of a vector-valued signal $u : [0, 1] \rightarrow \mathbb{R}^2$ (red), visualized as a curve in \mathbb{R}^2 . **Left to right:** pairwise approach (6) restricted to 4-neighborhood; 8-neighborhood; full model (6) with quadratic number of pairwise terms; proposed gradient-based method (7). The grey points show the quantization of the range of u , i.e., $[-1, 1]^2$. Apart from a *directional bias* caused by restricting the number of pairwise interaction terms to a local neighborhood (**left, second from left**), the pairwise formulation exhibits a strong bias towards *grid points* (**second from right**). The proposed method has greatly reduced bias and allows the minimizer to assume values *between* grid points, which can be considered as a form of *sub-label accuracy* in the range of u (**right**). The slight shrinkage again amounts to the typical contrast loss associated with ROF models.

2. Proposed Model

In order to extend (3) to the case where \mathcal{J} consists of all points on a (connected, Riemannian) s -dimensional manifold \mathcal{M} , $s \geq 1$, we replace the unit simplex formulation $u' : \Omega \rightarrow \Delta_l$ by $u' : \Omega \rightarrow \mathcal{P}(\mathcal{M})$, where $\mathcal{P}(\mathcal{M})$ denotes the set of *probability measures* on \mathcal{M} . The distance d is naturally given by the geodesic distance $d_{\mathcal{M}} : \mathcal{M}^2 \rightarrow \mathbb{R}$ on \mathcal{M} .

The set $\mathcal{J} = \mathcal{M}$ then consists of an *infinite* number of points. We propose to use the following formulation:

$$\min_{u' : \Omega \rightarrow \mathcal{P}(\mathcal{M})} \sup_{p : \Omega \times \mathcal{M} \rightarrow \mathbb{R}^d} \int_{\Omega} \langle u', s \rangle dx + \lambda \int_{\Omega} \langle u', \text{Div } p \rangle dx \quad (5)$$

$$\text{s.t. } \|p(x, z_1) - p(x, z_2)\|_2 \leq d_{\mathcal{M}}(z_1, z_2), \quad \forall z_1, z_2 \in \mathcal{M}. \quad (6)$$

In order to model (2), ideally one would like to minimize (5), (6) over all u' that correspond to actual *functions* $u : \Omega \rightarrow \mathcal{M}$ in the sense that $u'(x) \in \mathcal{P}(\mathcal{M})$ is given by $u'(x) = \delta_{u(x)}$, i.e., all $u'(x)$ are Dirac measures. By permitting arbitrary probability measures instead, we obtain a convex and therefore more tractable problem, at the cost of allowing “non-binary” solutions u' where $u'(x)$ is not a point measure at one or multiple points. Finding rounding strategies for generating u from such u' without sacrificing too much in terms of optimality can still be a difficult task, we refer to [16] for recent results on the labeling problem.

The usefulness of formulation (5), (6) is still slightly limited, as solving the problem numerically requires to discretize the problem, i.e., choosing a finite set of labels \mathcal{J} , in which case (5) reduces to the finite labeling problem (3).

However, the Riemannian manifold structure allows to reformulate the dual constraints (6) in a more elaborate way. In particular, it allows to compute, for every point $x \in \Omega$ and every point $z \in \mathcal{M}$ the gradient $D_z p(x, \cdot) =$

$(D_z p_1(x, \cdot), \dots, D_z p_d(x, \cdot)) \in (T_z \mathcal{M})^d$, where $T_z \mathcal{M}$ is the tangent space to \mathcal{M} at the point z .

The key idea is to replace the *pairwise* constraints (6) by a local condition on the *gradient* in terms of the spectral norm $\|\cdot\|_{\sigma}$, $\|M\|_{\sigma} := \sup\{\|Mv\|_2 / \|v\|_2 \mid v \in T_z \mathcal{M}\}$:

$$\|D_z p(x, \cdot)\|_{\sigma} \leq 1, \quad \forall z \in \mathcal{M}, \forall x \in \Omega. \quad (7)$$

This amounts to replacing a Lipschitz condition with respect to the geodesic distance $d_{\mathcal{M}}$ by an equivalent constraint on the gradient. We provide a short proof of the equivalency in the appendix, as it is instructive to verify that in fact the spectral norm appears in the constraint, rather than a numerically more convenient – and more commonly used – matrix norm such as the Frobenius norm.

The gradient-based formulation (7) has two major benefits:

- It can be accurately discretized using $O(l)$ *local* terms in contrast to $O(l^2)$ terms required for the full pairwise formulation (6).
- It allows to incorporate knowledge about the *local manifold structure* in terms of the differential operator D_z .

While it is possible to reduce the computational cost for the pairwise constraints by restricting the interaction to local neighborhoods, the approximation of the geodesic distance $d_{\mathcal{M}}$ is more accurate when using the local terms obtained through the gradient-based formulation.

The effect can be seen in Fig. 2, where we compare the result of Rudin-Osher-Fatemi (ROF) denoising of a function $u : [0, 1] \rightarrow [-1, 1]^2$ using a discretization of the set $[-1, 1]^2$ with $l = 289$ equally spaced points and gradients evaluated at the center of each cell. For such a moderate number of points the full pairwise model (6) already requires 83521 constraints at each image point, and shows a

pronounced bias towards grid points. The proposed model has much less bias, while only requiring 289 constraints.

An interesting observation is that for $\mathcal{M} = \mathbb{R}^s$ our approach leads to a regularizer based on the nuclear norm, as the dual to $\|\cdot\|_\sigma$. Such norms have been considered in a slightly different context in [18] and more recently in [10].

3. Discretized Model and Optimization

In the following we outline a discretization for the problem (5), (7) that captures these favorable properties. We assume that the manifold \mathcal{M} is discretized using $l \in \mathbb{N}$ points $z^1, \dots, z^l \in \mathcal{M}$. The weights vector $b \in \mathbb{R}^l$ defines the integration over the manifold, i.e., if $f \in \mathbb{R}^l$ discretizes a function f' on the manifold via $f^k = f'(z^k)$ then $\langle b, f \rangle := \sum_{k=1}^l b^k f^k \approx \int_{\mathcal{M}} f(x) ds(x)$.

The image domain $\Omega \subseteq \mathbb{R}^d$ is discretized using $n \in \mathbb{N}$ points $x^1, \dots, x^n \in \Omega$. The function $u' : \Omega \rightarrow \mathcal{P}(\mathcal{M})$ is represented as a vector $u \in (P_{\mathcal{M}})^n$ with $u^i = u'(x^i) = (u_1^i, \dots, u_l^i) \in P_{\mathcal{M}}$ for $1 \leq i \leq n$, where $P_{\mathcal{M}}$ is the discretized space of probability measures on the manifold \mathcal{M} :

$$P_{\mathcal{M}} := \{y \in \mathbb{R}^l \mid y \geq 0, \langle y, b \rangle = 1\}. \quad (8)$$

The local costs $\langle u', s \rangle$ are represented as the weighted inner product $\langle u, s \rangle_b$ with the cost vector $s \in \mathbb{R}^{n \times l}$, where

$$\langle u, s \rangle_b := \sum_{i=1}^n \sum_{k=1}^l u_k^i s_k^i b^k. \quad (9)$$

The linear differential operator $D : \mathbb{R}^{n \times l} \rightarrow \mathbb{R}^{n \times l \times d}$ computes the gradients in Ω using a staggered-grid scheme, and can be implemented as a sparse matrix.

For computation of the gradients on the s -dimensional manifold \mathcal{M} we allow to specify $m \in \mathbb{N}$ evaluation points $y^1, \dots, y^m \in \mathcal{M}$ at which the gradient constraint (7) should be enforced; these could for example be the cell centers if $\mathcal{M} = \mathbb{R}^s$ and the z^k form a regular mesh. For every such point y^j , a neighborhood $\mathcal{N}_j \subseteq \{1, \dots, l\}$ of $r \in \mathbb{N}$ points, i.e., $\#\mathcal{N}_j = r$, is given, whose values will be used to compute the gradient at y^j .

3.1. Gradient Discretization on the Manifold

In every point $y^j \in \mathcal{M}$ we compute, for all points $z^k \in \mathcal{M}, k \in \mathcal{N}_j$ in the neighborhood, the inverse exponential map $v^{j,k} := \exp_{y^j}^{-1}(z^k) \in T_{y^j} \mathcal{M}$. We see that using this definition, $p(x, y^j) + \langle v^{j,k}, D_{y^j} p(x, \cdot) \rangle$ provides an estimate of $p(x, z^k)$ through the Taylor expansion.

In order to approximate the gradient $g^{i,j} := D_{y^j} p(x^i, \cdot)$ at an image domain point $x^i \in \Omega, 1 \leq i \leq n$, and an evaluation point $y^j \in \mathcal{M}, 1 \leq j \leq m$, in a generic fashion, we define it as the vector in the tangent space that best explains the values of p through these estimates in an ℓ_2 -sense:

$$g_t^{i,j} := \arg \min_{g \in \mathbb{R}^s} \min_{c \in \mathbb{R}} \sum_{k \in \mathcal{N}_j} (c + \langle v^{j,k}, g \rangle - p_t(x^i, z^k))^2, \quad (10)$$

where $t = 1, \dots, d$. The extra variable c in the minimization receives the estimate for the value $p_t(x^i, y^j)$, which is unknown as p is discretized only on the points z^k .

By choosing a suitable parametrization of the tangent space at each point y^j , problem (10) can be written in terms of matrices $M^j \in \mathbb{R}^{r \times s}$ (where s is the dimension of the manifold and therefore of the tangent space) and a sparse indexing matrix $P^j \in \mathbb{R}^{r \times l}$ (both independent of i) as

$$\min_{g \in \mathbb{R}^s, c \in \mathbb{R}} \|ce + M^j g - P^j p_t\|_2^2, \quad (11)$$

where $e = (1, \dots, 1) \in \mathbb{R}^r$ is the constant 1-vector. The optimality condition of this joint minimization problem is

$$\begin{pmatrix} M^{j\top} \\ e^\top \end{pmatrix} \begin{pmatrix} M^j & e \end{pmatrix} \begin{pmatrix} g \\ c \end{pmatrix} = \begin{pmatrix} M^{j\top} \\ e^\top \end{pmatrix} P^j p_t. \quad (12)$$

Solving the system explicitly for c and substituting the result back into (12), we obtain the optimality condition for g ,

$$M^{j\top} E M^j g = M^{j\top} E P^j p_t, \quad E := (I - r^{-1} e e^\top). \quad (13)$$

We can therefore represent everything required to compute the gradient according to (10) in the compact form

$$A^j g_t^{i,j} = B^j P^j p_t^i, \quad (14)$$

with the two matrices $A^j := M^{j\top} E M^j \in \mathbb{R}^{s \times s}$ and $B^j := M^{j\top} E \in \mathbb{R}^{s \times r}$, together with the neighborhood-defining sparse matrices P^j , and $p_t^i := (p_t(x^i, z^k))_{1 \leq k \leq l} \in \mathbb{R}^l$.

3.2. Discretized Model

With the above remarks, and identifying matrices with their vector representation where necessary, the fully assembled discretized problem takes the following form:

$$\min_{u \in \mathbb{R}^{n \times l}} \max_{p \in \mathbb{R}^{n \times l \times d}, g \in \mathbb{R}^{n \times m \times s \times d}} \langle u, s \rangle_b + \langle Du, p \rangle_b \quad (15)$$

$$\text{s.t. } u^i \geq 0, \langle u^i, b \rangle = 1 \quad \forall i, \quad (16)$$

$$A^j g_t^{i,j} = B^j P^j p_t^i \quad \forall i, j, t, \quad (17)$$

$$\|g^{i,j}\|_\sigma \leq \lambda \quad \forall i, j, \quad (18)$$

where $i = 1, \dots, n, j = 1, \dots, m$ and $t = 1, \dots, d$. The scaling by b can be removed by setting $b = e$, in which case the problem remains the same except for an element-wise scaling of u^i by b^i , and the constraints for u turn into the unit simplex constraints $u^i \in \Delta_l$. Note that each of the $g^{i,j}$ is a matrix in $\mathbb{R}^{s \times d}$, where s is the dimension of the manifold and d is the dimension of the image domain Ω .

To solve the problem numerically, we introduce suitably-sized multipliers w and q for the equality constraints, with $w_t^{i,j} \in \mathbb{R}^s$ and $q^i \in \mathbb{R}$, and obtain

$$\min_{u, w, p, g, q} \langle u, s \rangle + \langle Du, p \rangle + \sum_{i,j,t} \langle w_t^{i,j}, A^j g_t^{i,j} - B^j P^j p_t^i \rangle + \sum_i \langle q^i, e^\top u^i - 1 \rangle \quad (19)$$

$$\text{s.t. } u^i \geq 0, \quad \|g^{i,j}\|_\sigma \leq \lambda. \quad (20)$$

The problem can then be solved using a standard primal-dual approach such as [4]. In order to deal with the non-trivial spectral norm constraints we use an approach based on [10], which allows to explicitly compute projections on the constraint set for image domains Ω with dimension 1 and 2. Another possible way is to replace the spectral norm by the Frobenius norm, in which case (19)–(20) becomes a standard second-order cone program (SOCP).

Let us summarize why in our opinion the proposed scheme is well suited as a generic model for optimization problems with values on manifolds:

- The solver implementation is fully independent of the actual structure of the manifold. Implementing a new manifold only requires to supply a set of points z^k , $k = 1, \dots, l$, on the manifold, the weight vector b that defines the integration, and the matrices A^j and B^j , $j = 1, \dots, m$, that define the local manifold structure.
- All constraint sets have a simple separable structure and can be easily projected upon. This makes the use of iterative, inexact projections obsolete when using first-order solvers such as [4], and eliminates related convergence and infeasibility issues.
- By choosing the y^j on the midpoints between pairs of points in the mesh defined by the z^k and choosing suitable A^j , B^j , and P^j , the discretized model also covers the *pairwise* formulation (6), as well as variants with a reduced number of interaction terms.

The latter point allows to easily compare the performance to that of the pairwise formulation. In our opinion, the most prominent feature is the modularity: in fact, all examples in the experimental section were computed using the same, unmodified solver implementation.

3.3. Means on Manifolds

As (5) is essentially a convex relaxation approach, it is possible – and actually desirable – to obtain probability measures u^i with several non-zero components. This raises the question of how to map such probability measures into single points on \mathcal{M} .

To simplify notation, for a fixed point $x^i \in \Omega$ we write u for $u^i = u'(x^i)$, and restrict ourselves to the case $b = (1, \dots, 1)$. Consequently u is constrained to the unit simplex, $u \in \Delta_l$, and we denote $u = (u_1, \dots, u_l)$.

The classical way to interpret non-binary u is to associate it with the corresponding *convex combination* z' of the z^k ,

$$z' = \sum_{k=1}^l u_k z^k. \quad (21)$$

This approach fails for manifolds other than convex subsets of \mathbb{R}^s with the usual Euclidean distance, as otherwise z

may not be a point in \mathcal{M} . Therefore we replace (21) by its *Fréchet mean* (also known as *Karcher mean*) with respect to the geodesic distance,

$$z' = \arg \min_{z' \in \mathcal{M}} \sum_{k=1}^l u_k d_{\mathcal{M}}(z', z^k)^2. \quad (22)$$

For $\mathcal{M} = \mathbb{R}^s$ and $d_{\mathcal{M}}(z', z) = \|z - z'\|_2$, this coincides with (21). Note that there may be cases when (22) does not have a unique minimizer, such as when $\mathcal{M} = \mathcal{S}^1$ and u assumes the value $1/2$ on two opposing points on the circle. While such border cases can be exercised by constructing symmetric noise-free academic problems, we found that in all our experiments on real-world data such a case never occurred.

As (22) cannot directly be solved as a convex problem, we use an iterative method [14]. The idea is to iteratively consider an approximated version of (22) within the tangent space of the current iterate, and with $d_{\mathcal{M}}$ replaced by the usual Euclidean norm. Each approximated problem can then be solved in closed form (21). More precisely, starting from an initial estimate z for z' , we iterate

$$z'^k \leftarrow \exp_z^{-1}(z^k) \in T_z \mathcal{M}, \quad (23)$$

$$\bar{z}' \leftarrow \sum_{k=1}^l u_k z'^k \in T_z \mathcal{M}, \quad (24)$$

$$z \leftarrow \exp_z(\bar{z}'). \quad (25)$$

In (23) all points z^k are mapped to the tangent space $T_z \mathcal{M}$ by the inverse exponential map, the mean \bar{z}' is then computed *in the tangent space* in (24), and finally projected back onto the manifold to update z in (25).

We observed that this process is generally very fast. On the unit sphere \mathcal{S}^2 it consistently required 5 – 10 iterations to converge to almost machine precision, and 200 iterations were sufficient to obtain reliable results even in cases where u' has a large number of non-zero components. As the Fréchet mean only needs to be computed once after the solver has finished, efficiency is also non-critical.

4. Experimental Results

4.1. Chromaticity-Brightness Denoising

Kang and March, and Bao et al. [13, 2] suggested a denoising model for RGB-valued color images $I : \Omega \rightarrow \mathbb{R}^3$ that separates denoising of the brightness $|I| : \Omega \rightarrow \mathbb{R}$ from denoising of the chromaticity $C = I/|I| : \Omega \rightarrow \mathcal{S}^2$. The former encodes mainly geometric information, while the latter consists of vectors in the two-dimensional unit sphere and captures the color information.

This separation of geometry and color was found to be better suited for preserving details than simply treating I as an \mathbb{R}^3 vector. While [13, 2] replace the constraint $C \in \mathcal{S}^2$



Figure 3. Denoising of a color image $I : \Omega \rightarrow \mathbb{R}^3$ by separately denoising its intensity $|I| \in \mathbb{R}$ by the usual ROF model, and chromaticity $I/|I| \in \mathcal{S}^2$ by the ROF model for values on \mathcal{S}^2 using the proposed method. **Left to right, top to bottom:** input image; added noise; denoised chromaticity; combined denoising result.



Figure 4. Artistic effects using separate smoothing of brightness and chromaticity. **Left to right:** smoothed chromaticity; smoothed brightness and chromaticity; smoothed brightness.

by a penalizing term $(1 - |C|)^2$ in the energy, our framework allows to naturally include it into the optimization (Fig. 3). Smoothing the brightness and chromaticity components differently gives rise to interesting visual effects (Fig. 4).

4.2. Optical Flow

Another application of our framework is the computation of dense optical flow $u : \Omega \rightarrow \mathbb{R}^2$ between two images $I_1, I_2 : \Omega \rightarrow \mathbb{R}^3$, where one seeks to minimize $|I_1(x) - I_2(x + u(x))|$ while keeping the velocity map u regular. Although the range of u is discretized using a regular grid, the sub-label accuracy of our approach allows u to assume values between grid points, giving a more natural, smooth result compared to treating the problem as a labeling problem (Fig. 6).

4.3. Processing Normals Fields for Visualization

Normal fields often occur in computer graphics and computer vision as direction fields or surface normals. Variational processing of such data requires optimization over the two-dimensional unit sphere \mathcal{S}^2 .

Figure 5 shows an application of our method to ROF denoising of the normal field for terrain data obtained from [6] in order to compute a shaded model. Denoising the normal field considerably improves the visual quality, preserving sharp transitions in the normal field such as along the mountain ridges due to the total variation-based regularizer.

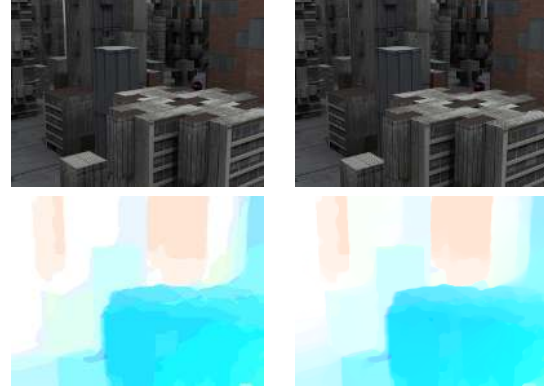


Figure 6. Optical flow computation. The optical flow vectors are represented as velocity vectors in \mathbb{R}^2 (cf. Fig. 2) **Top:** input image pair. **Bottom:** color-coded velocity vector map. Solving the problem as a labeling problem on a finite grid leads to an artificial piecewise constant solution (**bottom left**). Our sub-label accurate model generates a smoother, more realistic velocity map, as the vectors are not constrained to the grid (**bottom right**).

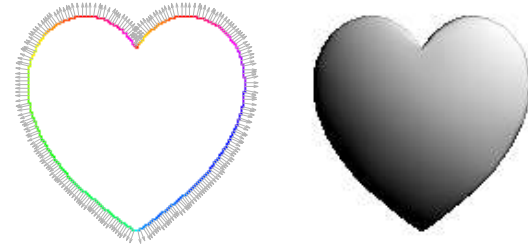


Figure 7. Optimization with values in the unit sphere \mathcal{S}^2 applied to the estimation of a surface normal field based on a sketched contour line. **Left to right:** Input contours with outer normal field; normal field estimated using a total variation regularizer with values on the unit sphere (x - and y -components of the normals color-coded as hue); shaded object based on the reconstructed normals.

The sphere was discretized using $l = 162$ points obtained by subdividing the edges of an icosahedron twice, and the gradients evaluated at the centers of all (triangular) faces.

The same approach can be used to inpaint normal fields based on given contour lines with outer normals by setting the data term to zero outside of the contours. By applying a shading model, pseudo three-dimensional views can be synthesized from a sketched two-dimensional contour (Fig. 7, we also refer to [11]).

4.4. Denoising of Rotation Data

Variational processing of rotation data is a highly non-trivial task, as the rotation space $SO(3)$ is difficult to parametrize in a way that is amenable to optimization. The proposed approach can be applied to this setting by representing the rotations as unit quaternions. These can be viewed as points on the three-dimensional unit sphere \mathcal{S}^3 , provided that antipodal points are identified, with the geodesic distance $d(a, b) = \arccos(|\langle a, b \rangle|)$.

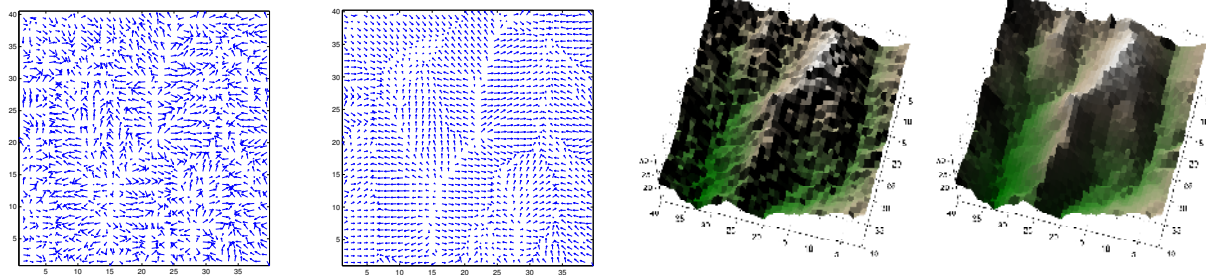


Figure 5. Optimization on the two-dimensional unit sphere applied to denoising of normals in S^2 for three-dimensional visualization. **Left to right:** Input normal field (two-dimensional top-projection) computed from noisy height map; result of Rudin-Osher-Fatemi denoising in S^2 ; height map with shading computed from unprocessed noisy normal field; height map with shading computed from denoised normal field. Similar to the scalar-valued case, the ROF model minimizer preserves discontinuities, e.g., along the ridges of the mountains.

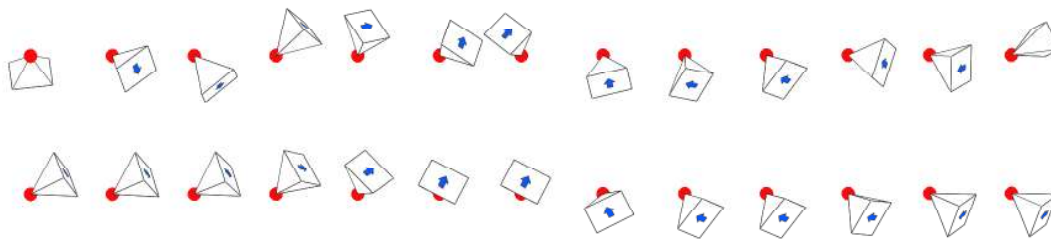


Figure 8. Optimization over the space $SO(3)$ of three-dimensional rotations applied to denoising of camera orientations. **Top:** input orientations. **Bottom:** denoised orientations. The rotations are discretized using a quaternion representation, and optimization is subsequently carried out over the three-dimensional manifold of unit quaternions.

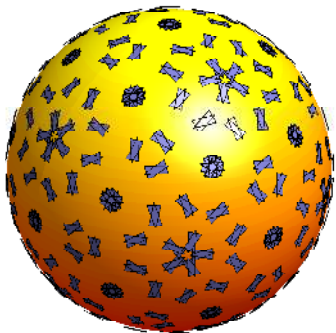


Figure 9. Discretization of the set of unit quaternions using 720 points as used in Fig. 8. Each quaternion is visualized as the rotated triangle that is obtained when applying the quaternion rotation to a fixed template triangle.

We generate the points z^k from the vertices of the hexacosichoron, which is a regular polytope in \mathbb{R}^4 akin to the icosahedron in \mathbb{R}^3 , by subdividing the faces and eliminating opposite points. The gradients are evaluated on the barycenters of the faces. This yields a set of 720 regularly spaced quaternions representing rotations in $SO(3)$ (Fig. 9).

Using the same numerical solver as for the previous experiments, this allows to apply Rudin-Osher-Fatemi denoising to the task of smoothing the rotation component of a camera trajectory (Fig. 8).

5. Conclusion

We proposed a framework for TV-regularization of functions with values in an arbitrary Riemannian manifold. To this end, we formulated the TV-regularization as a multi-label optimization problem with an infinite number of labels, and used a specialized discretization that allows to better incorporate knowledge about the manifold structure.

Using this approach, it becomes possible to solve variational problems for manifold-valued functions that consist of a possibly non-convex data term and a total variation regularizer. Suitable constraints on the dual variables ensure that the TV-regularizer correctly penalizes jumps according to the geodesic distance on the manifold.

We experimentally validated the proposed method on a variety of inverse problems, including the denoising of chromaticity values and camera trajectories, optical flow computation, as well as inpainting and denoising of normal fields. The proposed approach allows to obtain more accurate solutions compared to treating the problem as a finite labeling problem, while reducing the computational effort.

Many properties of minimizers of total variation-regularized models in manifolds are still not fully understood. The present work allows to gain numerical insight into this question, and will hopefully lead to a better theoretical understanding of their behavior. A generalization to other regularizers remains a subject of future work.

Acknowledgments. This publication is based on work supported by Award No. KUK-I1-007-43, made by King Abdullah University of Science and Technology (KAUST), EPSRC first grant No. EP/J009539/1, Royal Society International Exchange Award No. IE110314, Leverhulme Early Career Fellowship ECF-2013-436, and ERC Starting Grant Convex Vision.

Appendix

Proof of (6) \Leftrightarrow (7): Denote $p = p(x, \cdot)$ and assume that the gradient constraint (7) holds. For any two points $z_1, z_2 \in \mathcal{M}$ and any C^1 curve $c : [a, b] \rightarrow \mathcal{M}$ connecting the points, we have $c(a) = z_1$ and $c(b) = z_2$. Then

$$p(c(b)) - p(c(a)) = \int_a^b (p \circ c)'(s) ds = \int_a^b D_{c(s)} p \cdot c'(s) ds$$

by the definition of the gradient $D_{c(s)} p$, and consequently

$$\|p(c(b)) - p(c(a))\|_2 \leq \int_a^b \|D_{c(s)} p \cdot c'(s)\|_2 ds. \quad (26)$$

As the gradient constraint (7) holds by assumption and $c'(s) \in T_{c(s)} \mathcal{M}$, we can further bound

$$\|p(c(b)) - p(c(a))\|_2 \leq \int_a^b \|c'(s)\|_2 ds. \quad (27)$$

This estimate clearly requires to use the spectral norm in (7) in order to not introduce any unnecessary constants. The right-hand side is the length $L(c)$ of c , and we obtain

$$\|p(z_1) - p(z_2)\|_2 \leq L(c) \quad (28)$$

for all curves connecting z_1 and z_2 , which implies (6).

On the other hand, assume that (6) holds. For any point $z \in \mathcal{M}$ and tangent vector $v \in T_z \mathcal{M}$ we can find a C^1 curve $c : [-\varepsilon, \varepsilon] \rightarrow \mathcal{M}$ with $c(0) = z$ and $c'(0) = v$. Then

$$\begin{aligned} \|D_z(s) p \cdot v\|_2 &= \|D_z(s) p \cdot c'(0)\|_2 = \|(p \circ c)'(0)\|_2 \\ &= \left\| \lim_{\delta \rightarrow 0} (p(c(\delta)) - p(c(0))) / \delta \right\|_2 \quad (29) \\ &= \lim_{\delta \rightarrow 0} \|p(c(\delta)) - p(c(0))\|_2 / \delta \quad (30) \end{aligned}$$

Using the assumption (6), we obtain

$$\|D_z(s) p \cdot v\|_2 \leq \liminf_{\delta \rightarrow 0} \|d_{\mathcal{M}}(c(0), c(\delta))\|_2 / \delta \quad (31)$$

$$= \liminf_{\delta \rightarrow 0} \left\| \int_0^\delta c'(s) ds \right\|_2 / \delta = \|c'(0)\|_2 = \|v\|_2, \quad (32)$$

using the continuity of c' . Thus $\|D_z(s) p \cdot v\|_2 / \|v\|_2 \leq 1$ for all v in $T_z \mathcal{M}$, which implies (7).

References

- [1] E. Bae, J. Yuan, X.-C. Tai, and Y. Boykov. A fast continuous max-flow approach to non-convex multilabeling problems. CAM Report 10-62, UCLA, June 2012. 2
- [2] Z. Bao, Y. Liu, Y. Peng, and G. Zhang. Variational color image segmentation via chromaticity-brightness decomposition. In *Adv. Multim. Mod.*, pages 295–302, 2010. 5
- [3] A. Chambolle, D. Cremers, and T. Pock. A convex approach to minimal partitions. *J. Imaging Sci.*, 5(4):1113–1158, 2012. 2
- [4] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.*, 2010. published online. 5
- [5] D. Cremers and E. Strelakovsky. Total cyclic variation and generalizations. *Journal of Mathematical Imaging and Vision*, Nov. 2012. 2
- [6] D. Gesch, G. Evans, J. Mauck, J. Hutchinson, and W. J. Carswell Jr. The national map – elevation. U.S. Geological Survey Fact Sheet 3053, 2009. 6
- [7] M. Giaquinta, G. Modica, and J. Souček. Variational problems for maps of bounded variation with values in S^1 . *Calc. Var. Part. Diff.*, 1(1):87–121, 1993. 2
- [8] M. Giaquinta and D. Mucci. The BV-energy of maps into a manifold: relaxation and density results. *Ann. Scuola Norm. Sup. Pisa Cl. Sci.*, 5(4):483–548, 2006. 2
- [9] M. Giaquinta and D. Mucci. Maps of bounded variation with values into a manifold: Total variation and relaxed energy. *Pure Appl. Math. Q.*, 3(2):513–538, 2007. 2
- [10] B. Goldluecke, E. Strelakovsky, and D. Cremers. The natural vectorial total variation which arises from geometric measure theory. *SIAM J. Imag. Sci.*, 5(2):537–563, 2012. 4, 5
- [11] J. Hahn, J. Qiu, E. Sugisaki, L. Jia, X.-C. Tai, and H. Seah. Stroke-based surface reconstruction. CAM Report 12-18, UCLA, 2012. 6
- [12] H. Ishikawa. Exact optimization for Markov random fields with convex priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10):1333–1336, Oct. 2003. 2
- [13] K. H. Kang and R. March. Variational models for image colorization via chromaticity and brightness decomposition. *IEEE Trans. Imag. Proc.*, 16(9):2251–2261, 2007. 5
- [14] H. Karcher. Riemannian center of mass and mollifier smoothing. *Pure Appl. Math.*, 30(5):509–541, 1977. 5
- [15] J. Lellmann, F. Becker, and C. Schnörr. Convex optimization for multi-class image labeling with a novel family of total variation based regularizers. In *Int. Conf. Comp. Vis.*, pages 646–653, 2009. 2
- [16] J. Lellmann, F. Lenzen, and C. Schnörr. Optimality bounds for a variational relaxation of the image partitioning problem. *J. Math. Imaging Vis.*, 47(3):239–257, 2013. 3
- [17] J. Lellmann and C. Schnörr. Continuous multiclass labeling approaches and algorithms. *J. Imaging Sci.*, 4(4):1049–1096, 2011. 2
- [18] G. Sapiro and D. L. Ringach. Anisotropic diffusion of multivalued images with applications to color filtering. *IEEE Trans. Image Proc.*, 5(11):1582–1586, 1996. 4
- [19] M. I. Schlesinger. Sintaksicheskiy analiz dvumernykh zritelnykh signalov v usloviyakh pomekh (Syntactic analysis of two-dimensional visual signals in noisy conditions). *Kibernetika, in Russian*, 4:113–130, 1976. 2
- [20] E. Strelakovsky, C. Nieuwenhuis, and D. Cremers. Nonmetric priors for continuous multilabel optimization. In *Europ. Conf. Comp. Vis.*, volume 7578 of *Springer LNCS*, pages 208–221, 2012. 2
- [21] T. Werner. A linear programming approach to maxsum problem: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(7):1165–1179, 2007. 2
- [22] C. Zach, D. Gallup, J.-M. Frahm, and M. Niethammer. Fast global labeling for real-time stereo using multiple plane sweeps. In *Workshop on Vision, Modeling and Visualization*, October 2008. 2