

# Total Generalized Variation for Manifold-valued Data

K. Bredies\*      M. Holler\*      M. Storath<sup>†</sup>      A. Weinmann<sup>‡</sup>

## Abstract

In this paper we introduce the notion of second-order total generalized variation (TGV) regularization for manifold-valued data in a discrete setting. We provide an axiomatic approach to formalize reasonable generalizations of TGV to the manifold setting and present two possible concrete instances that fulfill the proposed axioms. We provide well-posedness results and present algorithms for a numerical realization of these generalizations to the manifold setup. Further, we provide experimental results for synthetic and real data to further underpin the proposed generalization numerically and show its potential for applications with manifold-valued data.

**Mathematical subject classification:** 94A08, 68U10, 90C90 53B99, 65K10.

**Keywords:** Total Generalized Variation, Manifold-valued Data, Denoising, Higher Order Regularization.

## 1 Introduction

In this work we introduce and explore a generalization of second-order total generalized variation (TGV) regularization for manifold-valued data. The TGV functional has originally been introduced in [29] for the vector space setting as generalization of the total variation (TV) functional [73], which is extensively used for regularization in image processing and beyond. The advantage of TV regularization compared to, e.g., classical  $H^1$  regularization approaches, is that jump discontinuities can be much better reconstructed. This can be seen in the function space setting since functions of bounded variation, as opposed to Sobolev functions, can have jump discontinuities. It is also reflected in numerical realizations where TV minimization allows to effectively preserve sharp interfaces. A disadvantage of TV regularization, however, is its tendency to produce piecewise constant results even in non-piecewise-constant regions, which is known as the staircasing effect. Employing a regularization with higher order derivatives, such as second-order TV regularization, overcomes this drawback, but again does not allow for jump discontinuities. As a result, a lot of recent research aims at finding suitable extensions of TV that overcome the staircasing effect, but still allow for jumps [31, 75, 21]. While infimal-convolution-type approaches can be seen as the first methods to achieve this, the introduction of the TGV functional (of arbitrary order  $k$ ) in 2010 finally provided a complete model for piecewise smooth

---

\*K. Bredies and M. Holler are with Institute of Mathematics and Scientific Computing of the University of Graz. The institute is a member of NAWI Graz ([www.nawigraz.at](http://www.nawigraz.at)). K. B. and M. H. are members of BioTechMed Graz ([www.biotechmed.at](http://www.biotechmed.at)); email: [kristian.bredies@uni-graz.at](mailto:kristian.bredies@uni-graz.at), [martin.holler@uni-graz.at](mailto:martin.holler@uni-graz.at).

<sup>†</sup>Martin Storath is with the Image Analysis and Learning Group, Heidelberg Collaboratory for Image Processing, Universität Heidelberg, Germany; email: [martin.storath@iwr.uni-heidelberg.de](mailto:martin.storath@iwr.uni-heidelberg.de).

<sup>‡</sup>Andreas Weinmann is with the Institute of Computational Biology, Helmholtz Zentrum München, and with the Department of Mathematics and Natural Sciences, Hochschule Darmstadt, Germany; email: [andreas.weinmann@h-da.de](mailto:andreas.weinmann@h-da.de)

data with jump discontinuities. This is achieved by an optimal balancing between first and higher order derivatives (up to the order  $k$ ), which is carried out as part of the evaluation of the functional. We refer to [29] for more motivation and to [25] for a detailed analysis of TGV in the context of inverse problems. In particular, second-order TGV, which balances between first and second-order derivatives, achieves piecewise linear – as opposed to piecewise constant – image reconstructions while it still allows for jumps. This renders second-order TGV a well suited model for piecewise smooth images and can be seen as the motivation of the use of second-order TGV regularization in a plethora of applications [57, 26, 27, 58, 65]. Up to now, TGV regularization was only available for vector space data and applications are hence limited to this situation.

In various problems of applied sciences, however, the data do not take values in a linear space but in a nonlinear space such as a smooth manifold. Examples of manifold-valued data are circle and sphere-valued data as appearing in interferometric SAR imaging [62], wind directions [74], orientations of flow fields [3, 79], and color image processing [33, 89, 56, 60]. Other examples are data taking values in the special orthogonal group  $SO(3)$  expressing vehicle headings, aircraft orientations or camera positions [87], Euclidean motion group-valued data [72] as well as shape-space data [63]. Another prominent manifold is the space of positive (definite) matrices endowed with the Fisher-Rao metric [70]. This space is the data space in DTI [69]. It is a Cartan-Hadamard manifold and as such it has particularly nice differential-geometric properties. DTI allows to quantify the diffusional characteristics of a specimen non-invasively [11, 54] which is helpful in the context of neurodegenerative pathologies such as schizophrenia [44] or autism [5]. Because of the natural appearance of these nonlinear data spaces quite a lot of recent work deals with them. Examples are wavelet-type multiscale transforms for manifold-valued data [87, 51, 91, 90, 92] as well as manifold-valued partial differential equations [85, 35, 48]. Work on statistics on Riemannian manifolds can be found in [66, 23, 43, 24, 68, 42]. Optimization problems for manifold-valued data are for example the topic of [2, 1], of [49] and of [52] with a view towards learning in manifolds. We also mention related work on optimization in Hadamard spaces [10, 9, 7] and on the Potts and Mumford-Shah models for manifold-valued data [94, 81].

TV functionals for manifold-valued data have been considered from the analytical side in [46, 47, 45]; in particular, the existence of minimizers of certain TV-type energies has been shown. A convex relaxation based algorithm for TV regularization for  $\mathbb{S}^1$ -valued data was considered in [82, 37]. Approaches for TV regularization for manifold-valued data are considered in [61] which proposes a reformulation as multi-label optimization problem and a convex relaxation, in [50] which proposes iteratively reweighted minimization, and in [93] which proposes cyclic and parallel proximal point algorithms. An exact solver for the TV problem for circle-valued signals has been proposed in [80]. Furthermore, [15] considers half-quadratic minimization approaches, which may be seen as an extension of [50], and [19] considers an extension of the Douglas-Rachford algorithm for manifold-valued data. Applications of TV regularization to shape spaces may be found in [13, 78]. TV regularization with a data term involving an imaging type operator has been considered in [14] where a forward-backward type algorithm is proposed to solve the corresponding inverse problem involving manifold-valued data.

As with vector space data, TV regularization for manifold-valued data has a tendency to produce piecewise constant results in regions where the data is smooth. As an alternative which prevents this, second-order TV type functionals for circle-valued data have been considered in [18, 20] and, for general manifolds, in [10]. However, similar to the vector space situation, regularization with second-order TV type functionals tends towards producing solutions which do not preserve the edges as desired. To address this drawback, the vector space situation guides us to considering TGV models and models based on infimal convolution to address this issue. The most difficult part in this respect is to define suitable notions in the manifold setup which have both reasonable analytic properties on the one hand and which are algorithmically

realizable on the other hand. Concerning infimal-convolution type functionals such as  $\text{TV}-\text{TV}^2$  infimal convolutions, a first effort towards a generalization to the manifold setting has been made in the recent conference proceeding [17] which has later been extended in an *arXiv* preprint [16]. The present manuscript [28], which was submitted to *arXiv* at the same time as [16], emerged independently. In this work, we focus on TGV and aim at providing a thorough study of reasonable generalizations of TGV for piecewise smooth, manifold-valued data by first investigating crucially required properties of generalizations of TGV to the manifold setting. Then we propose suitable extensions that fulfill these properties and which are, in addition, computationally feasible. In this respect, it is important to note that due to the cascadic nature of TGV (as opposed to infimal-convolution-type functionals), its generalization to manifolds requires a conceptually new approach. For this reason, both the techniques we propose for a generalization of TGV as well as our algorithmic setting are rather different from the one of [17], which uses a mid-point approach to generalize  $\text{TV}-\text{TV}^2$  infimal convolutions and gradient descent for numerical minimization.

## 1.1 Contributions

The contributions of the paper are as follows: *(i)* we lay the foundations for – and provide concrete realizations of – a suitable variational model for second-order total generalized variation for manifold-valued data, *(ii)* we provide algorithms for the proposed models, *(iii)* we show the potential of the proposed algorithms by applying them to synthetic and real data. Concerning *(i)*, we use an axiomatic approach. We first formalize reasonable fundamental properties of vector-valued TGV which should be conserved in the manifold setting. Then we propose two concrete realizations which we show to fulfill the axioms. The one is based on parallel transport and the other is motivated by the Schild’s approximation of parallel transport. We obtain well-posedness results for TGV-based denoising of manifold valued data for both variants. Concerning *(ii)* we provide the details for a numerical realization of variational regularization for general manifolds using either of the two proposed generalizations of TGV. We build on the well-established concept of cyclic proximal point algorithms (using the inexact variant). The challenging part and the main contribution consists in the computation of the proximal mappings of the TGV atoms involved. In particular, for the class of symmetric spaces, we obtain closed form expressions of the related derivatives in terms of Jacobi fields for the Schild variant; for the parallel transport variant, we obtain closed form expressions for the related derivatives for general symmetric spaces up to a derivative of a certain parallel transport term for which we can still analytically compute the derivative for the class of manifolds considered in the paper. Concerning *(iii)*, we provide a detailed numerical investigation of the proposed generalization and its fundamental properties. Furthermore, we provide experiments with real and synthetic data and a comparison to existing methods.

## 1.2 Outline of the paper

The paper is organized as follows. The topic of Section 2 is the derivation of suitable TGV models for manifold-valued data. We start out with a detailed discussion of fundamental properties expected by a reasonable TGV version for manifold-valued data. To this end, we first reconsider vector-space TGV in a form suitable for our purposes in Subsection 2.1. Then we derive an axiomatic extension of  $\text{TGV}_\alpha^2$  to the manifold setting where we – for a better understanding – first consider the univariate case in Subsection 2.2. Next we suggest two realizations which we call the Schild variant of TGV and the parallel transport variant of TGV and show that these realizations indeed fulfill all desired properties. Then, in Subsection 2.3, we extend the axiomatic

framework to the bivariate setting and provide bivariate versions of the Schild variant of TGV and the parallel transport variant of TGV, respectively, that we show to fulfill the required axioms. In Section 3 we then provide existence and lower semi-continuity results for the proposed variants of TGV from which the existence of minimizers for the TGV-regularized denoising problems of manifold-valued data follows. The topic of Section 4 is the algorithmic realization of the proposed models. We start out recalling the concept of a cyclic proximal point algorithm (CPPA). Then, in Subsection 4.1, we consider the implementation of the CPPA for manifold-valued TGV. We identify certain TGV atoms whose proximal mappings are challenging to compute. The necessary derivations needed for their computation are provided for the Schild variant of TGV in Subsection 4.2, and, for the parallel transport variant of TGV, in Subsection 4.3. The topic of Section 5 is the numerical evaluation of the proposed schemes. There, we first carry out a detailed numerical evaluation of the proposed model for synthetic data. We test extreme parameter choices and ensure consistency of the results of our numerical method with a reference implementation for vector spaces. Then we consider various application scenarios and compare to existing methods. Finally, we draw conclusions in Section 6.

### 1.3 Basic Notation and Concepts from Differential Geometry

Throughout this work,  $\mathcal{M} \subset \mathbb{R}^K$  will always denote a complete manifold with a Riemannian structure (with its canonical metric connection, its Levi-Civita connection). Let us explain these notions more precisely. We consider a manifold  $\mathcal{M}$  with a Riemannian metric which is a smoothly varying inner product  $\langle \cdot, \cdot \rangle_p$  in the tangent space of each point  $p$ . (We note that any paracompact manifold can be equipped with a Riemannian structure.) As usual, we will frequently omit the dependence on  $p$  in the notation in the following. According to the Hopf-Rinow theorem, the complete manifold  $\mathcal{M}$  is geodesically complete in the sense that any geodesic can be prolonged arbitrarily. Here a geodesic is a curve  $\gamma$  of zero acceleration, i.e.,  $\frac{D}{dt} \frac{d}{dt} \gamma = 0$ , where  $\frac{D}{dt}$  denotes the covariant derivative along the curve  $\gamma$  (see below for details on the covariant derivative). Geodesics are invariant under affine reparametrizations and usually identified with their image in  $\mathcal{M}$ . (We parametrize them on  $[0, 1]$  or, more generally, on  $[t_0, t_1]$  with  $t_0 < t_1$ , depending on the context.) We always denote by  $d : \mathcal{M} \times \mathcal{M} \rightarrow [0, \infty)$  the distance on  $\mathcal{M}$  which is induced by the Riemannian structure and note that, since  $\mathcal{M}$  is complete, for any  $a, b \in \mathcal{M}$  there is always a geodesic from  $a$  to  $b$  whose length equals  $d(a, b)$ . Such geodesics are called length minimizing geodesics between  $a$  and  $b$ . By  $T_a \mathcal{M}$  we denote the tangent space of  $\mathcal{M}$  at  $a \in \mathcal{M}$  and by  $T\mathcal{M}$  the tangent bundle of  $\mathcal{M}$ . Further, we need the notion of parallel transport. The parallel transport of a vector  $v \in T_a \mathcal{M}$  with  $a \in \mathcal{M}$  to a point  $b \in \mathcal{M}$  along a curve  $\gamma : [0, 1] \rightarrow \mathcal{M}$  such that  $\gamma(0) = a$ ,  $\gamma(1) = b$  is the vector  $V_1 = V(1) \in T_b \mathcal{M}$ , where  $V : [0, 1] \rightarrow T\mathcal{M}$  is given as the solution of the ODE  $\frac{D}{dt} V(t) = 0$  on  $[0, 1]$  with the initial condition  $V(0) = v$ , where the covariant derivative  $\frac{D}{dt}$  is taken along the curve  $\gamma$ . The covariant derivative  $\frac{D}{dt}$  is induced by the Levi-Civita connection  $\nabla_X Y$  of the Riemannian manifold  $\mathcal{M}$ ; a connection  $\nabla_X Y$  is a  $C^\infty$ -linear mapping w.r.t. the vector field  $X$  and a derivation w.r.t. the vector field  $Y$ . (Connections are needed since, in a general manifold  $\mathcal{M}$ , there is no a priori canonical way to define the directional derivative in direction  $X$  of a vector field  $Y$ .) For the covariant derivative  $\frac{D}{dt} Y$  along a curve  $\gamma$ , a direction  $X$  along the curve  $\gamma$  is given by  $\frac{d}{dt} \gamma$  and the vector field  $Y$  is given along the curve; hence, if  $\frac{d}{dt} \gamma \neq 0$ ,  $\frac{D}{dt} Y$  is just given by  $\nabla_{\frac{d}{dt} \gamma} Y$ . In a Riemannian manifold  $\mathcal{M}$ , there is a (uniquely determined) canonical metric connection, its Levi-Civita connection. The Levi-Civita connection is the only connection which is symmetric (i.e.,  $\nabla_X Y = \nabla_Y X$  for any vector fields  $X, Y$  which commute) and which is compatible with the metric (i.e., in terms of the induced covariant derivative which we need later on,  $\frac{d}{dt} \langle X, Y \rangle = \langle \frac{D}{dt} X, Y \rangle + \langle X, \frac{D}{dt} Y \rangle$  for any two vector fields  $X, Y$  along a curve  $\gamma$ , along which the covariant derivatives are taken.) For an account on



Riemannian geometry we refer to the books [76, 38] or to [59].

For later use, we fix some further notation. To this end, let  $a, b \in \mathcal{M}$ ,  $v \in T_a\mathcal{M}$ , and  $\gamma : [0, L] \rightarrow \mathcal{M}$  be a curve connecting  $a$  and  $b$  such that  $\gamma(t_0) = a$ ,  $\gamma(t_1) = b$  with  $t_0 < t_1$ ,  $t_0, t_1 \in [0, L]$ . We define

1.  $\exp_a(v) = \psi(1)$  where  $\psi : [0, 1] \rightarrow \mathcal{M}$  is the unique geodesic such that  $\psi(0) = a$ ,  $\frac{d}{dt}\psi(0) = v$ ,
2.  $\exp(v) = \exp_a(v)$  when it is clear from the context that  $v \in T_a\mathcal{M}$ ,
3.  $\log_a(b) = \{z \in T_a\mathcal{M} \mid \exp(z) = b \text{ and } [0, 1] \ni t \mapsto \exp(tz) \text{ is a length-minimizing geodesic}\}$ ,
4. in case  $\gamma$  is a geodesic and length-minimizing between  $a$  and  $b$  on  $[t_0, t_1]$ , we denote  $\log_a^\gamma(b) = (t_1 - t_0)\frac{d}{dt}\gamma(t_0)$ , i.e., the vector in  $\log_a(b)$  that is parallel to  $\frac{d}{dt}\gamma(t_0)$  and such that  $\exp_a(\log_a^\gamma(b)) = b$ ,
5.  $\text{pt}_{a,b}^\gamma(v) = z \in T_b\mathcal{M}$ , where  $z$  is the vector resulting from the parallel transport of  $v$  from  $a$  to  $b$  along the curve  $\psi$  that reparametrizes  $[t_0, t_1]$  to  $[0, 1]$  in an affine way, i.e.,  $\psi(t) = \gamma(t_0 + t(t_1 - t_0))$  such that  $\psi(0) = a$ ,  $\psi(1) = b$ ,
6.  $\text{pt}_{a,b}(v) = \{\text{pt}_{a,b}^\psi(v) \mid \psi \text{ is a length-minimizing geodesic connecting } a \text{ and } b\}$ ,
7.  $\text{pt}_b(v) = \text{pt}_{a,b}(v)$ ,  $\text{pt}_{a,b}^\gamma(v) = \text{pt}_b^\gamma(v)$  when it is clear from the context that  $v \in T_a\mathcal{M}$ ,
8. for  $t \in [0, 1]$ ,

$$[a, b]_t = \{\psi(t) \mid \psi : [0, 1] \rightarrow \mathcal{M} \text{ is a length-minimizing geodesic with } \psi(0) = a, \psi(1) = b\}.$$

We extend this definition for  $t \in \mathbb{R} \setminus [0, 1]$  by extending the corresponding geodesic.

We also note that throughout the paper we identify sets having only one element with the corresponding element.

We further need the concept of geodesic variations for the existence results in Section 3 and the derivation of the algorithms in Section 4. A variation of a curve  $\gamma : I \rightarrow \mathcal{M}$  defined on an interval  $I$  is a smooth mapping  $V : I \times J \rightarrow \mathcal{M}$ ,  $J$  an interval containing 0, such that  $V(s, 0) = \gamma(s)$  for all  $s \in I$ . A variation is called geodesic, if all curves  $s \mapsto V(s, t)$  are geodesics for any  $t$  in  $J$ .

## 2 Definition of TGV for manifold-valued data

The goal of this section is to define a discrete total generalized variation (TGV) functional of second order for manifold-valued data. To this end, we first state some fundamental properties of the TGV functional in infinite and finite dimensional vector spaces. The definition of a generalization of TGV to the manifold setting will then be driven by the goal of preserving these fundamental properties of vector-space TGV.

### 2.1 TGV on vector spaces

We first recall the definition of TGV on infinite-dimensional vector spaces via its minimum representation which is, according to results in [30, 88, 25, 26] covering the second- and general order case as well as the scalar and vectorial case, equivalent to the original definition as provided in [29]. Then we present a discretization which is slightly different from the one typically used.

**Definition 2.1** (Minimum representation of TGV in vector spaces). For  $\alpha_0, \alpha_1 \in (0, \infty)$ ,  $u \in L^1_{\text{loc}}(\Omega)^K$  with  $\Omega \subset \mathbb{R}^d$  a bounded Lipschitz domain, we define

$$\text{TGV}_\alpha^2(u) = \min_{w \in \text{BD}(\Omega, \mathbb{R}^d)^K} \alpha_1 \|Du - w\|_{\mathcal{M}} + \alpha_0 \|\mathcal{E}w\|_{\mathcal{M}}. \quad (1)$$

Here,  $\|\cdot\|_{\mathcal{M}}$  denotes the Radon norm in the space of Radon measures  $\mathcal{M}(\Omega, X)^K := (C_0(\Omega, X)^K)^*$  with  $X \in \{\mathbb{R}^d, S^{d \times d}\}$ ,  $S^{d \times d}$  the space of symmetric matrices and further  $\text{BD}(\Omega, \mathbb{R}^d)^K := \{w \in L^1(\Omega, \mathbb{R}^d)^K \mid \mathcal{E}w \in \mathcal{M}(\Omega, S^{d \times d})^K\}$ . The derivatives  $Du \in \mathcal{D}(\Omega, \mathbb{R}^d)^K$  and  $\mathcal{E}w \in \mathcal{D}(\Omega, S^{d \times d})^K$  are defined in the weak sense by

$$\begin{aligned} \langle Du, \varphi \rangle &= -\langle u, \text{div } \varphi \rangle, \quad \varphi \in C_c^\infty(\Omega, \mathbb{R}^d)^K, \\ \langle \mathcal{E}w, \varphi \rangle &= -\langle w, \text{div } \varphi \rangle, \quad \varphi \in C_c^\infty(\Omega, S^{d \times d})^K, \end{aligned}$$

with  $\text{div } \varphi = (\text{div } \varphi^1, \dots, \text{div } \varphi^K) \in C_c^\infty(\Omega)^K$  for  $\varphi = (\varphi^1, \dots, \varphi^K) \in C_c^\infty(\Omega, \mathbb{R}^d)^K$  and, for  $\varphi = (\varphi^1, \dots, \varphi^K) \in C_c^\infty(\Omega, S^{d \times d})^K$  with  $\varphi^i = (\varphi^i_1, \dots, \varphi^i_d) \in C_c^\infty(\Omega, S^{d \times d})$ , we denote  $\text{div } \varphi = (\text{div } \varphi^1, \dots, \text{div } \varphi^K) \in C_c^\infty(\Omega, \mathbb{R}^d)^K$  with  $\text{div } \varphi^i = (\text{div } \varphi^i_1, \dots, \text{div } \varphi^i_d) \in C_c^\infty(\Omega, \mathbb{R}^d)$ . See [25, 26] for details.

Note that  $\text{TGV}_\alpha^2(u)$  is finite if and only if  $u \in \text{BV}(\Omega)^K$  and, in this case, the minimum is actually obtained [25]. Hence the term  $\min$  in the above definition is justified.

For a discretization and generalization later on, it is convenient to list some of the main properties of second-order TGV in function space (see [25, 26]):

**(P1)** If the minimum in (1) is obtained at  $w = 0$ , then  $\text{TGV}_\alpha^2(u) = \alpha_1 \text{TV}(u)$ ,

**(P2)** If the minimum in (1) is obtained at  $w = Du$ , then  $\text{TGV}_\alpha^2(u) = \alpha_0 \text{TV}^2(u)$ ,

**(P3)**  $\text{TGV}_\alpha^2(u) = 0$  if and only if  $u$  is affine.

Here,  $\text{TV}^2$  denotes a second-order TV functional which can be defined as  $\text{TV}^2(u) = \|D^2u\|_{\mathcal{M}}$  where  $D^2u$  denotes the second-order distributional derivative of  $u$  and the above quantities are finite if and only if  $D^2u \in \mathcal{M}(\Omega, S^{d \times d})^K$ . We note that  $w = 0$  and  $w = Du$  can trivially be obtained when  $u$  is constant and affine, respectively, but  $w = 0$  is for instance also obtained under some symmetry conditions when  $u$  is piecewise constant or when the ratio  $\alpha_0/\alpha_1$  is large enough [29, 67].

Using the minimum representation above, a discretization of  $\text{TGV}_\alpha^2$  in vector spaces on two-dimensionals grids can be given as follows.

**Definition 2.2** (Discrete isotropic and anisotropic TGV in vector spaces). Set  $U = \mathbb{R}^K$ . For  $u = (u_{i,j})_{i,j} \in U^{N \times M}$  with  $u_{i,j} \in U$ , we define the discrete second-order TGV functional as

$$\text{TGV}_\alpha^2(u) = \min_{w \in U^{(N-1) \times M} \times U^{N \times (M-1)}} \alpha_1 \|Du - w\|_1 + \alpha_0 \|\mathcal{E}w\|_1 \quad (2)$$

where

$$Du := (\delta_{x+}u, \delta_{y+}u) \in U^{(N-1) \times M} \times U^{N \times (M-1)}$$

and

$$\mathcal{E}w := \mathcal{E}(w^1, w^2) := (\delta_{x-}w^1, \delta_{y-}w^2, \frac{\delta_{y-}w^1 + \delta_{x-}w^2}{2}) \in U^{(N-2) \times M} \times U^{N \times (M-2)} \times U^{(N-1) \times (M-1)}$$

with  $\delta_{x+}$ ,  $\delta_{y+}$  and  $\delta_{x-}$ ,  $\delta_{y-}$  being standard forward and backward differences, respectively. Introducing a parameter  $p \in [1, \infty)$ , the one-norms are given as

$$\|w\|_1 = \sum_{i,j} |(w^1_{i,j}, w^2_{i,j})|_p \text{ for } w \in U^{(N-1) \times M} \times U^{N \times (M-1)}$$

and

$$\|z\|_1 = \sum_{i,j} \left| \begin{pmatrix} z_{i,j}^1 & z_{i,j}^3 \\ z_{i,j}^3 & z_{i,j}^2 \end{pmatrix} \right|_p \text{ for } z \in U^{(N-2) \times M} \times U^{N \times (M-2)} \times U^{(N-1) \times (M-1)},$$

where we extend the signals by zero to have the same size. Here,  $|(w^1, w^2)|_p := (|w^1|^p + |w^2|^p)^{1/p}$  and  $\left| \begin{pmatrix} z^1 & z^3 \\ z^3 & z^2 \end{pmatrix} \right|_p := (|z^1|^p + |z^2|^p + 2|z^3|^p)^{1/p}$  with  $|\cdot|$  the Euclidean norm on  $U = \mathbb{R}^K$ .

Note that we incorporate a pointwise  $\ell^p$ -norm with  $p \in [1, \infty)$  in the definition of  $\text{TGV}_\alpha^2$ . In the function space setting of Definition 2.1, this corresponds to choosing the appropriate dual norm on  $X \in \{\mathbb{R}^d, S^{d \times d}\}$ . There, any choice of point-wise norm (also beyond  $\ell^p$ -norms) yields equivalent functionals and function spaces, with the Euclidean norm being used in the first paper on TGV [29] and the most popular choices being  $\ell^p$  norms with  $p \in \{1, 2\}$ . The focus of this paper is the case  $p = 1$  since this is more frequently used in the manifold context because of being more amenable to numerical realization. In the theory part we include the case of  $p \in [1, \infty)$  since it can be done without additional effort.

We remark that the above discretization of  $\text{TGV}_\alpha^2$  is slightly different from the standard one as provided, e.g., in [29]. The purpose of this is to achieve consistency of the zero set of both the continuous and discrete version as follows. Also, note that, notation-wise, we do not distinguish between the continuous and the discrete version of TGV (and of TV and  $\text{TV}^2$ ). In the following, we will always refer to discrete versions.

**Proposition 2.3** (Zero set of  $\text{TGV}_\alpha^2$  in vector spaces). *For  $u \in U^{N \times M}$  we have that  $\text{TGV}_\alpha^2(u) = 0$  if and only if  $u$  is affine, i.e., there exist  $a, b, c \in U$  such that  $u_{i,j} = ai + bj + c$ .*

*Proof.* We provide only the basic steps: Setting  $w = Du$  we get that

$$\mathcal{E}w = \mathcal{E}Du = \left( \delta_x - \delta_{x+u}, \delta_y - \delta_{y+u}, \frac{\delta_y - \delta_{x+u} + \delta_x - \delta_{y+u}}{2} \right).$$

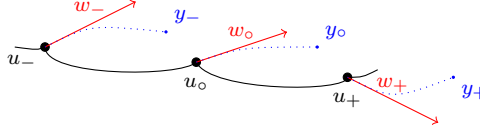
Now it is easy to see that if  $u$  is linear as above,  $\mathcal{E}Du = 0$  and hence the choice  $w = Du$  renders  $\text{TGV}_\alpha^2(u)$  to be zero. Conversely,  $\text{TGV}_\alpha^2(u) = 0$  implies  $w = Du$  and hence  $\mathcal{E}Du = 0$ . It is easy to see that  $\delta_x - \delta_{x+u} = \delta_y - \delta_{y+u} = 0$  implies that  $u$  is of the form  $u_{i,j} = rij + ai + bj + c$  with  $r, a, b, c \in U$ . But in this case, the last component of  $\mathcal{E}Du$  being zero implies that  $r = \delta_y - \delta_{x+u} = -\delta_x - \delta_{y+u} = -r$ , hence  $r = 0$ .  $\square$

The latter proposition shows that, also after discretization, the kernel of  $\text{TGV}_\alpha^2$  consists exactly of (discrete) affine functions. In fact, this is one of the fundamental properties of TGV (corresponding to **(P3)**) which should also be transferred to an appropriate generalization of TGV for manifolds. Regarding appropriate counterparts of **(P1)** and **(P2)**, we introduce the following definition.

**Definition 2.4.** *Using the notation of Definition 2.2, we define  $D^2u = \mathcal{E}Du$  and*

$$\text{TV}(u) = \|Du\|_1, \quad \text{and} \quad \text{TV}^2(u) = \|D^2u\|_1.$$

In summary, using the discretized version of TV,  $\text{TV}^2$  and  $\text{TGV}_\alpha^2$  and the discrete notion of affine functions as in Proposition 2.3, we can conclude that the properties **(P1)** to **(P3)** of TGV in continuous vector spaces also transfer to its discretization.



**Figure 1:** A three-point section of a signal  $u$  together with tangent vectors  $w$  represented by the endpoints  $y$ . The red vectors show the tangent vectors at the signal points, the blue dotted lines indicate the geodesics  $t \mapsto \exp_u(tw)$  and the  $y$  are the endpoints  $\exp_u(w)$ . The black lines indicate geodesic interpolations between the signal points and are for visualization purposes only.

## 2.2 Univariate $\text{TGV}_\alpha^2$ on manifolds

When moving from the vector space to the manifold setting, a main difference is that vectors representing derivatives can no longer be made independent of their location, but are attached to a base point on the manifold. In other words, in the Euclidean setting, all tangent spaces at all locations can be identified, which is not possible for manifolds in general. Accordingly, when aiming to extend TGV to the manifold setting, important questions are how to introduce the vector fields  $(w_{i,j})_{i,j}$  appearing in the minimum representation, how to define  $\mathcal{E}w$  for such elements, and, most importantly, how to define a suitable distance-type function between such tangent vectors sitting in different points.

In the following, we describe our main ideas to resolve these questions. In order to allow the reader to understand the underlying ideas more easily, we consider the univariate setting first. Let  $u = (u_i)_i$  be a finite sequence of points in a manifold  $\mathcal{M}$  with distance  $d : \mathcal{M} \times \mathcal{M} \rightarrow [0, \infty)$ . Our first goal is to suitably extend the notion of forward differences  $(\delta_{x+}u)_i = u_{i+1} - u_i$  and introduce auxiliary variables  $w_i$  which they can be compared to. To this end, the central idea is to identify tangent vectors  $w_i$  in the space  $T_{u_i}\mathcal{M}$  with point-tuples, i.e.,  $w_i \simeq [u_i, y_i]$  with  $y_i \in \mathcal{M}$ . In the vector-space case, this can be done via the correspondence  $w_i = y_i - u_i$ . For manifolds, the correspondence can be established via the exponential and the logarithmic map. That is, any  $w_i \in T_{u_i}\mathcal{M}$  can be assigned to a unique point-tuple  $[u_i, y_i]$  such that  $\exp_{u_i}(w_i) = y_i$ . Conversely, any point-tuple  $[u_i, y_i]$  can be assigned to (generally multiple) tangent vectors  $w_i$  such that  $w_i \in \log_{u_i}(y_i)$ . (Note that the ambiguities in logarithmic map result from non-uniqueness of distance-minimizing geodesics, which is a rather degenerate situation in the sense that it only occurs for a set of points with measure zero [53, 34].) Figure 1 visualizes the correspondence between tangent vectors  $w$  and point-tuples  $[u, y]$ .

Exploiting this correspondence, our approach is to work with a discrete “tangent bundle” that is defined as the set of point-tuples  $[u, y]$  with  $u, y \in \mathcal{M}$ , rather than with the continuous one. We refer to the elements in this discrete “tangent bundle” as tangent tuples. The identification of forward differences with a tangent tuple is then naturally given via  $(\delta_{x+}u)_i = [u_i, u_{i+1}]$ .

The discretized vector-space version of second-order TGV (for which there is a one-to-one correspondence of tangent vectors and point tuples) can then in the univariate case be rewritten as

$$\begin{aligned} \text{TGV}_\alpha^2(u) &= \min_{(w_i)_i} \sum_i \alpha_1 |(\delta_{x+}u)_i - w_i| + \alpha_0 |(\delta_{x-}w)_i| \\ &= \min_{(y_i)_i} \sum_i \alpha_1 D([u_i, u_{i+1}], [u_i, y_i]) + \alpha_0 D([u_i, y_i], [u_{i-1}, y_{i-1}]), \end{aligned}$$

where we define  $D([x, y], [u, v]) = |[x, y] - [u, v]|$  and  $[x, y] - [u, v] = (y - x) - (v - u)$ . Note that, in this context,  $(w_i)_i$ ,  $(y_i)_i$  always denote finite sequences of points with their length being the same as the one of  $\delta_{x+}u$  and, whenever out of bound indices are accessed in such a summation,

we assume the signals to be extended such that the evaluation of  $D$  yields zero cost.

Hence, in order to extend TGV to the manifold-setting, we are left to appropriately measure the distance of two tangent tuples, i.e., generalize the expression  $D([x, y], [u, v]) = |[x, y] - [u, v]|$ . In case both tuples have the same base-point on the manifold, a simple and rather natural generalization is to measure their distance by the distance of their endpoints, that is, we set  $D([x, y], [x, v]) = d(y, v)$ . In the other case, i.e., for evaluating  $D([u_i, y_i], [u_{i-1}, y_{i-1}])$  for  $u_i \neq u_{i-1}$ , there is no such simple or even unique generalization since the two tangent tuples belong to different (tangential) spaces. A quite direct approach to overcome this is to first employ the concept of parallel transport on manifolds to shift both tangent tuples to the same location, and then to measure the distance of their endpoints. One possible generalization for TGV proposed in this paper builds on this idea. This is not the only possible way of measuring the distance of two tangent tuples. Alternatively, one can for instance build on a discrete approximation of parallel transport, i.e., on a variant of the Schild approximation, to measure distance of point tuples at different locations. This second approach is simpler to realize numerically, and it can be build on simpler differential-geometric tools, while the first one is more straight-forward. Both approaches have their motivation and, given this ambiguity in possible generalizations, we take a more systematic approach to the topic. That is, we first formulate fundamental properties that we require from reasonable generalizations of TGV to the manifold setting. Then we translate these properties to requirements on admissible distance-type functions for measuring the distance of two tangent tuples. Having established the later, we propose two possible concrete realizations that fulfill the required properties.

**Axiomatic extension.** Assuming  $D : \mathcal{M}^2 \times \mathcal{M}^2 \rightarrow [0, \infty)$  to be a distance-type function for tangent tuples, we define a generalization of second-order TGV for univariate manifold-valued data as

$$\text{M-TGV}_\alpha^2(u) = \min_{(y_i)_i} \sum_i \alpha_1 d(u_{i+1}, y_i) + \alpha_0 D([u_i, y_i], [u_{i-1}, y_{i-1}]). \quad (3)$$

In this context, we note that, while with TGV for discrete and continuous vector spaces it is known that the minimum in the above expression is attained, this is not clear a-priori for our generalization. In order not to lose focus, we shift the discussion on existence to Section 3, noting at this point that for all versions of  $\text{M-TGV}_\alpha^2$  proposed in this work, existence can be shown.

Accounting for the fact that the tangent tuples represent vectors in the tangent space, some basic identifications and requirements for the distance-type function  $D$  follow naturally. Zero elements in the discrete “tangent bundle” for instance correspond to tangent tuples of the form  $[u, u]$  with  $u \in \mathcal{M}$  and the distance of  $[x, x]$  and  $[y, y]$  should be zero. Also, the distance of two identical elements should also be zero, i.e.,  $D([x, y], [x, y]) = 0$ , and, in the vector space case, the distance function should reduce to the difference of the corresponding tangent vectors.

Our goal is now to further restrict possible choices of distance-type function in order to obtain a meaningful generalization of TGV. To this aim, we want to ensure that appropriate counterparts of the properties **(P1)** to **(P3)** are preserved for the resulting version of  $\text{M-TGV}_\alpha^2$  also in the manifold setting. In order to make this more precise, we first have to generalize the involved concepts. We start with the notion of “affine”.

Given that the geodesics in a manifold play the role of straight lines in vector space, a natural generalization for  $u = (u_i)_i$  a finite sequence in  $\mathcal{M}$  to be affine is to require that all points  $(u_i)_i$  are on a single geodesic at equal distance. A difficulty that arises in connection with this definition is that, in general, as opposed to the vector space case, a geodesic connecting two points is not necessarily unique. As a consequence, even though all points might be at the same distance when following a joint geodesic, the distance of each single pair of points in the manifold is not necessarily equal. To account for that, we require in addition that the geodesic connecting all points also realizes the shortest connection between all involved points on the geodesic locally.

**Definition 2.5.** Let  $u = (u_i)_{i=0}^n$  be a finite sequence of points on a manifold  $\mathcal{M}$ . We say that  $u$  is geodesic if there exists a geodesic  $\gamma : [0, L] \rightarrow \mathcal{M}$  parametrized with unit speed such that  $\gamma(iL/n) = u_i$  for  $i = 0, \dots, n$  and  $d(u_i, u_{i+1}) = L/n$  for  $i = 0, \dots, n-1$ , i.e.,  $\gamma|_{[iL/n, (i+1)L/n]}$  is a geodesic of minimal length connecting  $u_i$  and  $u_{i+1}$  for all  $i$ .

A second issue arising from non-uniqueness of geodesics is the fact that, even though every triplet of points in a signal  $(u_i)_i$  might be connected by a geodesic, we cannot conclude that all points are connected by a unique geodesic. Consequently, as a reasonable generalization of TGV will typically act local, in particular will be based on three point stencils, we cannot hope to obtain more than a local assertion for signals in the zero set of TGV. To account for that, we introduce the following notion.

**Definition 2.6.** Let  $u = (u_i)_{i=0}^n$  be a finite sequence of points on a manifold  $\mathcal{M}$ . We say that  $u$  is locally geodesic if for each  $i \in \{2, \dots, n-1\}$ , the sequence  $(u_j)_{j=i-1}^{i+1}$  is geodesic.

As one might expect, in the situation that subsequent points of a signal are sufficiently close such that they admit a unique connecting geodesic we obtain equivalence of the notions of geodesic and locally geodesic signals. In this respect, we again note that, in the general non-local situation, the existence of unique minimizing geodesics is true for any set of points outside a set of measure zero [53, 34].

**Lemma 2.7.** If a sequence of points  $u = (u_i)_i$  in  $\mathcal{M}$  is such that the length minimizing geodesic connecting each pair  $u_i, u_{i'}$  with  $|i - i'| \leq 1$  is unique, then  $u$  is locally geodesic if and only if it is geodesic.

*Proof.* If  $u$  is geodesic, it is obviously locally geodesic. Now assume that  $u$  is locally geodesic and that the length minimizing geodesic connecting each pair  $u_i, u_{i'}$  with  $|i - i'| \leq 1$  is unique. Then there exists a geodesic connecting  $u_0, u_1, u_2$  at equal distance, i.e., the distance between two subsequent points equals the length of the geodesic segment. We proceed recursively: Now assume that  $u_{i-2}, u_{i-1}, u_i$  are connected by a geodesic at equal distance. As  $u$  is locally geodesic also  $u_{i-1}, u_i, u_{i+1}$  are connected by a geodesic. Now by uniqueness, the two geodesics between  $u_i$  and  $u_{i-1}$  must coincide. Hence all points  $u_{i-2}$  till  $u_{i+1}$  are on the same geodesic at equal speed. Proceeding iteratively, the result follows.  $\square$

In order to investigate the counterparts of properties **(P1)** and **(P2)**, we need to define generalizations of the TV and  $\text{TV}^2$  functionals to the manifold setting. For TV, a natural generalization is to set, for  $u = (u_i)_i$  in  $\mathcal{M}$ ,

$$\text{TV}(u) = \sum_i d(u_{i+1}, u_i), \quad (4)$$

see also [93]. For  $\text{TV}^2$ , a generalization is not that immediate. In [10], second-order TV was essentially (and up to a constant) defined as  $\text{TV}^2(u) = 2 \sum_i \inf_{c \in [u_{i-1}, u_{i+1}]_{\frac{1}{2}}} d(c, u_i)$ , which, in the vector space setting reduces to

$$\text{TV}^2(u) = 2 \sum_i d\left(\frac{u_{i-1} + u_{i+1}}{2}, u_i\right) = \sum_i |u_{i-1} - 2u_i + u_{i+1}|. \quad (5)$$

However, as can be deduced by considering  $\text{TV}^2$  as special case of the different versions of TGV proposed below, this is not the only generalization which fulfills such a property. We will call a function an *admissible generalization* of  $\text{TV}^2$  whenever, in the vector space setting, it reduces to  $\text{TV}^2$  as in Equation (5) above.

Using these prerequisites, we now formulate our requirements on an appropriate generalization of TGV to the manifold setting of the form (3).

**(M-P1)** In the vector space setting,  $M\text{-TGV}_\alpha^2$  reduces to the univariate version of (2).

**(M-P2)** If the minimum in (3) is attained at  $(y_i)_i = (u_i)_i$ , i.e., the tangent tuples  $[u_i, y_i]$  all correspond to zero vectors, then  $M\text{-TGV}_\alpha^2(u) = \alpha_1 \text{TV}(u)$  with  $\text{TV}$  as in (4).

**(M-P3)** If the minimum in (3) is attained at  $(y_i)_i = (u_{i+1})_i$ , i.e., the tangent tuples  $[u_i, y_i]$  all correspond to  $(\delta_{x+}u)_i$ , then  $M\text{-TGV}_\alpha^2(u) = \alpha_0 \text{TV}^2(u)$  with  $\text{TV}^2$  an admissible generalization of  $\text{TV}^2$ .

**(M-P4)**  $M\text{-TGV}_\alpha^2(u) = 0$  if and only if  $u$  is locally geodesic according to Definition 2.6.

As the following Proposition shows, the Properties **(M-P1)** to **(M-P3)** already follow from basic requirements on the distance function that arise from the intuition that tangent tuples represent tangent vectors. The most restrictive property is **(M-P4)**, which requires an additional assumption on the distance function.

**Proposition 2.8.** *Assume that the function  $D : \mathcal{M}^2 \times \mathcal{M}^2 \rightarrow [0, \infty)$  is such that*

- $D([x, x], [u, u]) = 0$  for any  $x, u \in \mathcal{M}$ ,
- $D([x, y], [u, v]) = |(y - x) - (v - u)|$  in case  $\mathcal{M} = \mathbb{R}^K$ .

*Then, for  $M\text{-TGV}_\alpha^2$  as in (3), the properties **(M-P1)** to **(M-P3)** hold. If we further assume that for any geodesic three-point signal  $(v_j)_{j=i-1}^{i+1}$  it follows that  $D([v_i, v_{i+1}], [v_{i-1}, v_i]) = 0$ , then  $M\text{-TGV}_\alpha^2(u) = 0$  for any locally geodesic signal  $u = (u_i)_i$ . Conversely, assume that for any three-point signal  $(v_j)_{j=i-1}^{i+1}$ , such that the geodesic connecting each pair  $v_i, v_{i'}$  with  $|i - i'| \leq 1$  is unique,  $D([v_i, v_{i+1}], [v_{i-1}, v_i]) = 0$  implies that  $(v_j)_{j=i-1}^{i+1}$  is geodesic. Then, for any signal  $(u_i)_i$  where the geodesic connecting each pair  $u_i, u_{i'}$  with  $|i - i'| \leq 1$  is unique,  $M\text{-TGV}_\alpha^2(u) = 0$  implies that  $u$  is geodesic.*

*Proof.* In the vector space case we get by our assumptions and since  $d(x, u) = |x - u|$ , that

$$\begin{aligned} M\text{-TGV}_\alpha^2(u) &= \min_{(y_i)_i} \sum_i \alpha_1 |u_{i+1} - u_i - (y_i - u_i)| + \alpha_0 |(y_i - u_i) - (y_{i-1} - u_{i-1})| \\ &= \min_{(w_i)_i} \sum_i \alpha_1 |(\delta_{x+}u)_i - w_i| + \alpha_0 |(\delta_{x-}w)_i| \end{aligned} \quad (6)$$

which coincides with the univariate version of  $\text{TGV}$  as in (2), hence **(M-P1)** holds.

Now in case the minimum in (3) is achieved for  $(y_i)_i = (u_i)_i$  we get that

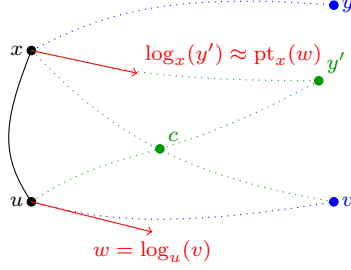
$$M\text{-TGV}_\alpha^2(u) = \sum_i \alpha_1 d(u_{i+1}, u_i) + \alpha_0 D([u_i, u_i], [u_{i-1}, u_{i-1}]) = \alpha_1 \text{TV}(u)$$

and **(M-P2)** holds. Similarly, if the minimum in (3) is achieved for  $(y_i)_i = (u_{i+1})_i$  we get that

$$M\text{-TGV}_\alpha^2(u) = \sum_i \alpha_0 D([u_i, u_{i+1}], [u_{i-1}, u_i]) \quad (7)$$

which is an admissible generalization of  $\text{TV}^2$  since, by assumption,  $D([u_i, u_{i+1}], [u_{i-1}, u_i]) = |u_{i+1} - 2u_i + u_{i-1}|$  in case  $\mathcal{M} = \mathbb{R}^K$ . Hence **(M-P3)** holds.

Now in case  $(u_i)_i$  is locally geodesic, we can choose  $(y_i)_i = (u_{i+1})_i$  to estimate  $M\text{-TGV}_\alpha^2$  from above with the right-hand side in (7) and obtain that  $M\text{-TGV}_\alpha^2(u) = 0$ . Conversely, in case  $M\text{-TGV}_\alpha^2(u) = 0$  we get that necessarily  $(y_i)_i = (u_{i+1})_i$  and  $M\text{-TGV}_\alpha^2$  reduces to the right hand side in (7). By the assumption on  $D$ , this implies that  $u$  is locally geodesic and by Lemma 2.7 the result follows.  $\square$



**Figure 2:** Approximate parallel transport of  $\log_u(v)$  to  $x$  via the Schild's ladder construction.

Guided by the aim of fulfilling the requirements of Proposition 2.8, in the following, we propose two possible choices for  $D(\cdot, \cdot)$  which result in two different concrete versions of  $\text{M-TGV}_\alpha^2$ . Both variants follow the general idea of first transporting different tangent tuples to the same location in  $\mathcal{M}$  and then measuring the distance there. The main difference between the two variants will be the way the transport is carried out. The first concept is motivated by the so called Schild's ladder approximation of parallel transport. Its implementation requires only basic differential geometric concepts and is therefore presented first. The second realization, which we call the parallel transport variant, requires more differential geometric concepts [76, 38] and is therefore presented afterwards. However, the Schild's ladder variant can be seen as an approximation of the parallel transport variant as explained below.

**Realization via Schild's-approximation.** Let  $[x, y]$  and  $[u, v]$  be two tuples in  $\mathcal{M}^2$  for which we want to define  $D([x, y], [u, v])$  and assume for the moment that distance-minimizing geodesics are unique. Motivated by the Schild's ladder approximation of parallel transport [39, 55], we consider the following construction (see Figure 2): take  $c$  to be the midpoint of the points  $v$  and  $x$ , i.e.,  $c = [x, v]_{\frac{1}{2}}$ ; set  $y' = [u, c]_2$ , i.e., reflect  $u$  at  $c$ . Then, we claim that the distance-type function  $D([x, y], [u, v]) = d(y, y')$  fulfills the requirements of Proposition 2.8.

The above construction may be motivated as follows. To compare the tangent vectors  $\log_x(y)$  and  $\log_u(v)$ , which correspond to the tangent tuples  $[x, y]$  and  $[u, v]$ , we could move them to the same point using parallel transport and then take the norm induced by the Riemannian metric in the corresponding tangent space. Using the above construction to obtain  $y'$  from  $x$  and  $[u, v]$ , the tangent vector  $\log_x(y')$  can be seen as an approximation of the parallel transport of  $\log_u(v)$  to  $x$  (in the limit  $x \rightarrow u$ ) [39, 55]. The difference of this vector to  $\log_x(y)$  is then approximated by  $d(y, y')$ . We notice that, as can be easily seen, this Schild's ladder approximation of parallel transport, i.e., the construction of  $y'$  and  $\log_x(y')$  as above, is exact in the vector space case.

Hence we propose to measure the deviation between the tangent tuples  $[x, y]$  and  $[u, v]$  by transporting  $[u, v]$  to  $x$  using the construction above to obtain the tangent tuple  $[x, y']$  and then to compare the resulting tangent tuples sitting in the same point  $x$  by measuring the distance of their endpoints  $y, y'$ . Since in general geodesics are not unique, we have to minimize over all possible constructions as above, i.e., consider all midpoints of all length-minimizing geodesics. This yields the following distance-type function

$$D_S([x, y], [u, v]) = \min_{y' \in \mathcal{M}} d(y', y) \text{ such that } y' \in [u, c]_2 \text{ with } c \in [x, v]_{\frac{1}{2}}. \quad (8)$$

It is immediate that  $D_S$  is positive and zero for identical elements. Furthermore if  $D_S([x, y], [u, v]) = 0$  then  $[x, y]$  can be interpreted to be equal to the approximate parallel transport of  $[u, v]$  to  $x$ . Indeed, in the vector space case,  $D_S([x, y], [u, v]) = 0$  if and only if  $v - u$  is parallel to  $y - x$  and has the same length and direction.



Plugging in  $D_S$  in (3) we define

$$S\text{-TGV}_\alpha^2(u) = \min_{(y_i)_i} \sum_i \alpha_1 d(u_{i+1}, y_i) + \alpha_0 D_S([u_i, y_i], [u_{i-1}, y_{i-1}]). \quad (9)$$

Regarding the properties **(M-P1)** to **(M-P4)**, we then get the following result.

**Theorem 2.9.** *The  $S\text{-TGV}_\alpha^2$  functional as in (9) satisfies **(M-P1)** and **(M-P3)**. If a finite sequence of points  $u = (u_i)_i$  is locally geodesic, then  $S\text{-TGV}_\alpha^2(u) = 0$ . Further, if  $S\text{-TGV}_\alpha^2(u) = 0$  and the geodesic connecting each pair  $u_i, u_{i'}$  with  $|i - i'| \leq 1$  is unique, then  $u$  is geodesic.*

*Proof.* We verify the assumptions of Proposition 2.8 for  $D = D_S$ . It can be easily seen that  $D_S([x, x], [u, u]) = D_S([x, y], [x, y]) = 0$  for  $x, u, y \in \mathcal{M}$ . Also, in the case  $\mathcal{M} = \mathbb{R}^K$ , the approximate parallel transport as in the definition of  $D_S$  above coincides with the parallel shift of vectors. Hence  $D_S([x, y], [u, v]) = D_S([0, y - x], [0, v - u]) = |(y - x) - (v - u)|$  and **(M-P1)** to **(M-P3)** holds. Now if  $v = (v_j)_{j=i-1}^{i+1}$  is locally geodesic, then  $v_i \in [v_i, v_i]_{\frac{1}{2}}$  and  $v_{i+1} \in [v_{i-1}, v_i]_2$ , hence  $D_S([v_i, v_{i+1}], [v_{i-1}, v_i]) = 0$ . Conversely,  $D_S([v_i, v_{i+1}], [v_{i-1}, v_i]) = 0$  implies that there exists  $y' \in [v_{i-1}, v_i]_2$  such that  $d(y', v_{i+1}) = 0$ . But this implies that  $v_{i+1} \in [v_{i-1}, v_i]_2$  and, by the assumption on unique geodesics, it follows that  $v$  is geodesic.  $\square$

**Remark 2.10.** *Assuming that, for each  $i$ ,  $u_i$  and  $y_i$  are sufficiently close such that they are connected by a unique length minimizing geodesic, we get that  $[u_i, [u_i, y_i]_{\frac{1}{2}}]_2 = y_i$  and hence  $D_S([u_i, u_{i+1}], [u_i, y_i]) = d(u_{i+1}, y_i)$ . Consequently, in this situation, an equivalent definition of  $S\text{-TGV}_\alpha^2$  can be given as*

$$S\text{-TGV}_\alpha^2(u) = \min_{(y_i)_i} \sum_i \alpha_1 D_S([u_i, u_{i+1}], [u_i, y_i]) + \alpha_0 D_S([u_i, y_i], [u_{i-1}, y_{i-1}])$$

*This definition regards the mapping  $(u_i)_i \mapsto ([u_i, u_{i+1}])_i$  as discrete gradient operator, mapping from  $\mathcal{M}$  to the discrete tangent space, and exclusively works in the discrete tangent space with  $D_S$  the canonical distance-type function.*

**Remark 2.11.** *We note that an alternative choice for  $D_S$  would have been to transport  $[u_i, y_i]$  to  $u_{i-1}$  rather than  $[u_{i-1}, y_{i-1}]$  to  $u_i$ . However, since the distance  $d(u_{i+1}, y_i)$  in  $S\text{-TGV}_\alpha^2$  as in (9) can be interpreted as evaluating the difference of the forward difference  $(\delta_{x+u})_i$  with  $[u_i, y_i]$  in the center point  $u_i$ , it seems natural to evaluate also the backward difference of the signal  $([u_i, y_i])_i$  at the center point.*

**Remark 2.12.** *We also note that, as opposed to the vector space setting, the distance function  $D_S$  is in general not symmetric, i.e.,  $D_S([x, y], [u, v]) \neq D_S([u, v], [x, y])$ . To obtain symmetry, alternative definitions of  $D_S$  could be given as  $D_S([x, y], [u, v]) = D_S([x, y], [u, v]) + D_S([u, v], [x, y])$  or*

$$\tilde{D}_S([x, y], [u, v]) = \min_{c_1, c_2} d(c_1, c_2) \quad \text{subject to } c_1 \in [x, v]_{\frac{1}{2}} \text{ and } c_2 \in [u, y]_{\frac{1}{2}}.$$

*For the sake of simplicity and in order to obtain the relation with parallel transport, however, we have defined  $D_S$  as in (8).*

**Realization via parallel transport.** The Schild's ladder construction defined above can be seen as a discrete approximation of parallel transport. Alternatively, we can use the identification of point-tuples  $[u, v], [x, y]$  with vectors in the tangent space at  $u$  and  $x$ , respectively, and use the parallel transport directly to transport the respective vectors to a common base point.

That is, assuming – for the moment – uniqueness of length-minimizing geodesics, we can identify each  $[u, v]$  with  $w \in T_u \mathcal{M}$  such that  $v = \log_u(w)$  and compare  $\text{pt}_x(w)$ , the vector

resulting from the parallel transport of  $w$  to  $T_x\mathcal{M}$ , to  $\log_x(y) \in T_x\mathcal{M}$ . This yields a distance-type function for two point-tuples  $[u, v]$ ,  $[x, y]$  as

$$D_{\text{pt}}([x, y], [u, v]) = \|\log_x(y) - \text{pt}_x(\log_u(v))\|_x \quad (10)$$

where  $\|\cdot\|_x$  denotes the norm in  $T_x\mathcal{M}$ . Comparing this to  $D_S$  we note that, besides using a different notion of transport, we now measure the distance with the norm in the tangent space rather than with the distance of endpoints. The reason for doing so is that, in this situation, approximating the norm in the tangent space via the distance of endpoints does not yield further simplification since the parallel transport forces us to work in the tangent space anyway.

Now in general, length minimizing geodesics are not necessarily unique. To deal with this issue, we have defined the log mapping and the parallel transport to be set-valued (see Section 1.3), i.e., for  $u, v \in \mathcal{M}$ , not necessarily close, and  $w \in T_v\mathcal{M}$  we define  $\text{pt}_u(w) \subset T_u\mathcal{M}$  to be the set of all vectors in  $T_u\mathcal{M}$  which can be obtained by parallel-transporting  $w$  to  $T_u\mathcal{M}$  along a length minimizing geodesic. Note that by isometry of the parallel transport, the length of all such vectors is the same but by holonomy their orientation might be different. In order to adapt the above definition of  $D_{\text{pt}}$  to this situation, we generalize the distance function  $D_{\text{pt}}$  for tangent tuples  $[x, y]$  and  $[u, v]$  as

$$D_{\text{pt}}([x, y], [u, v]) := \min_{\substack{z_1 \in \log_x(y) \\ z_2 \in T_x\mathcal{M}}} \|z_1 - z_2\|_x \text{ such that } z_2 \in \text{pt}_x(w) \text{ with } w \in \log_u(v). \quad (11)$$

Using this notation, a parallel-transport-based version of  $\text{M-TGV}_\alpha^2$ , denoted by  $\text{PT-TGV}_\alpha^2$ , can then be defined for the general situation as

$$\text{PT-TGV}_\alpha^2(u) = \min_{(y_i)_i} \sum_i \alpha_1 d(u_{i+1}, y_i) + \alpha_0 D_{\text{pt}}([u_i, y_i], [u_{i-1}, y_{i-1}]). \quad (12)$$

Similarly to the Schild's-version, we then get the following result.

**Theorem 2.13.** *The functional  $\text{PT-TGV}_\alpha^2$  satisfies **(M-P1)** and **(M-P3)**. If a finite sequence of points  $u = (u_i)_i$  is locally geodesic, then  $\text{PT-TGV}_\alpha^2(u) = 0$ . Further, if  $\text{PT-TGV}_\alpha^2(u) = 0$  and the geodesic connecting each pair  $u_i, u_{i'}$  with  $|i - i'| \leq 1$  is unique, then  $u$  is geodesic.*

*Proof.* The proof is similar to the one of Theorem 2.19 below considering the bivariate setting.

**Remark 2.14.** *Since the expression  $d(u_{i+1}, y_i)$  can be seen as approximation of  $\|\log_{u_i}(u_{i+1}) - w_i\|_{u_i}$ , where  $w_i \in \log_{u_i}(y_i)$ , an alternative definition of  $\text{PT-TGV}_\alpha^2$  (assuming uniqueness of geodesics for simplicity) only in terms of tangent vectors is given as*

$$\widetilde{\text{PT-TGV}}_\alpha^2(u) = \min_{w_i \in T_{u_i}\mathcal{M}} \sum_i \alpha_1 \|\log_{u_i}(u_{i+1}) - w_i\|_{u_i} + \alpha_0 \|w_i - \text{pt}_{u_i}(w_{i-1})\|_{u_i}. \quad (13)$$

*We believe, however, that the originally proposed version is preferable since the fact that the first term in  $\text{PT-TGV}_\alpha^2$  only involves the standard manifold distance of two points simplifies the numerical realization. That is, with the standard distance we are able to solve certain subproblems (proximal mappings) in the algorithm explicitly whereas the other version would require additional inner loops (see Section 4 for details).*

### 2.3 Bivariate $\text{TGV}_\alpha^2$ on manifolds

The goal of this section is to extend  $\text{M-TGV}_\alpha^2$  to the bivariate case. To this aim, we first observe that the bivariate version of  $\text{TGV}_\alpha^2$  for vector spaces as in (2) can be written as

$$\begin{aligned} \text{TGV}_\alpha^2(u) = & \min_{(w_{i,j}^1)_{i,j}, (w_{i,j}^2)_{i,j}} \sum_{i,j} \alpha_1 \left( |(\delta_x u)_{i,j} - w_{i,j}^1|^p + |(\delta_y u)_{i,j} - w_{i,j}^2|^p \right)^{1/p} \\ & + \alpha_0 \left( |w_{i-1,j}^1 - w_{i,j}^1|^p + |w_{i,j-1}^2 - w_{i,j}^2|^p + 2^{1-p} |(w_{i,j-1}^1 - w_{i,j}^1) + (w_{i-1,j}^2 - w_{i,j}^2)|^p \right)^{1/p}, \end{aligned} \quad (14)$$

where we set all norms  $|\cdot|^p$  to be zero whenever out of bound indices are involved.

The first four summands in the above definition of  $\text{TGV}_\alpha^2$  can all be transferred to the manifold setting using the previously introduced distance-type functions for tangent tuples. The additional difficulty of the bivariate situation arises from the fifth term, which first combines both differences and sums of tangent tuples and then measures the norm. Again possible generalizations are not unique and we employ an axiomatic approach to propose reasonable choices.

**Axiomatic extension.** Denote by  $D(\cdot, \cdot)$  one of the two previously introduced distance functions for tangent tuples and assume that  $D^{\text{sym}} : \mathcal{M}^2 \times \mathcal{M}^2 \times \mathcal{M}^2 \times \mathcal{M}^2 \rightarrow \mathbb{R}$  generalizes the fifth term in (14), which corresponds to the mixed derivatives. A bivariate version of  $\text{M-TGV}_\alpha^2$  can then be given as

$$\begin{aligned} \text{M-TGV}_\alpha^2(u) = & \min_{y_{i,j}^1, y_{i,j}^2} \alpha_1 \sum_{i,j} \left( d(u_{i+1,j}, y_{i,j}^1)^p + d(u_{i,j+1}, y_{i,j}^2)^p \right)^{1/p} \\ & + \alpha_0 \sum_{i,j} \left( D([u_{i,j}, y_{i,j}^1], [u_{i-1,j}, y_{i-1,j}^1])^p + D([u_{i,j}, y_{i,j}^2], [u_{i,j-1}, y_{i,j-1}^2])^p \right. \\ & \left. + 2^{1-p} D^{\text{sym}}([u_{i,j}, y_{i,j}^1], [u_{i,j}, y_{i,j}^2], [u_{i,j-1}, y_{i,j-1}^1], [u_{i-1,j}, y_{i-1,j}^2])^p \right)^{1/p}. \end{aligned} \quad (15)$$

The basis for this generalization is again the representation of tangent vectors with tangent tuples, only that now for each  $u_{i,j}$  we consider two tangent vectors  $w_{i,j}^1, w_{i,j}^2$  and corresponding points  $y_{i,j}^1, y_{i,j}^2$  in order to represent horizontal and vertical derivatives, see Figure 3. Remember that in this paper, we focus on the case  $p = 1$ , for which we derive an algorithmic realization later on. However, since it causes no additional effort in this section, we provide a formulation for all  $p \in [1, \infty)$  here.

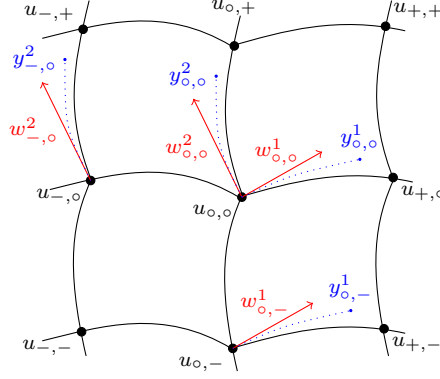
We now extend the requirements **(M-P1)** to **(M-P4)** to the bivariate case. The generalization of TV to bivariate manifold-valued data is quite straightforward [93]. For  $u = (u_{i,j})_{i,j}$  we define

$$\text{TV}(u) = \sum_{i,j} \left( d(u_{i+1,j}, u_{i,j})^p + d(u_{i,j+1}, u_{i,j})^p \right)^{1/p}. \quad (16)$$

A generalization of second-order TV is again less straightforward and we call any functional  $\text{TV}^2$  acting on  $u = (u_{i,j})_{i,j}$  an admissible generalization of  $\text{TV}^2$  if it reduces, in the vector space setting, to  $\text{TV}^2$  as given in Definition 2.4. Note that our version of  $\text{TV}^2$  differs from the definition of  $\text{TV}^2$  as given in [10] since we use a symmetrization of the mixed derivatives.

A generalization of affine functions for the bivariate manifold setting is given as follows.

**Definition 2.15.** Let  $u = (u_{i,j})_{i,j}$  be a finite sequence of points on a manifold  $\mathcal{M}$ . We say that  $u$  is (locally) geodesic if, for each  $(i_0, j_0)$ , the univariate signals  $(u_{i,j_0})_i$ ,  $(u_{i_0,j})_j$  and  $(u_{i_0+k,j_0-k})_k$  are (locally) geodesic.



**Figure 3:** A three-by-three point section of a bivariate signal  $u$  together with tangent vectors  $w$  represented by endpoints  $y$ . The blue, dotted lines indicate the geodesics  $t \mapsto \exp(tw)$  connecting the signal points  $u$  with the endpoints  $y$ . The black lines indicate a piecewise geodesic interpolation of the signal points and are for visualization purposes only.

Note that this indeed generalizes the notion of affine for vector space data. While the first two conditions ensure that  $u$  is of the form  $u_{i,j} = ai + bj + cj + d$ , the last condition ensures that  $a = 0$ , i.e., no mixed terms occur. In the vector space case the third condition is equivalent to requiring that  $(u_{i_0+k, j_0+k})_k$  is (locally) geodesic. In the manifold setting, this is not equivalent in general and hence our definition of geodesic is somewhat anisotropic. The reason for using the anti-diagonal  $((i_0 + k, j_0 - k))$  and not the diagonal  $((i_0 + k, j_0 + k))$  direction is that the former arises naturally from the usage of forward-backward differences, as will become clear in the discussion after Proposition 2.17. An alternative would also be to require both  $(u_{i_0+k, j_0+k})_k$  and  $(u_{i_0+k, j_0-k})_k$  to be geodesic. This, however, seems rather restrictive and for general manifolds such signals might even not exist. Hence we do not impose this additional restriction.

As a direct consequence of the corresponding result in the univariate setting, we obtain equivalence of the notion of locally geodesic and geodesic if neighboring points are sufficiently close.

**Lemma 2.16.** *Let  $u = (u_{i,j})_{i,j}$  be a finite sequence in  $\mathcal{M}$ . If the length minimizing geodesic connecting any two points  $u_{i,j}$  and  $u_{i',j'}$  with  $\max\{|i - i'|, |j - j'|\} \leq 1$  is unique, then  $u$  is locally geodesic if and only if it is geodesic.*

With these prerequisites, we now state our requirements for a reasonable generalization of TGV in the bivariate case.

- (M-P1') In the vector space setting,  $\text{M-TGV}_\alpha^2$  reduces to vector-space TGV as in (2).
- (M-P2') If the minimum in (15) is attained at  $(y_{i,j}^1)_{i,j} = (y_{i,j}^2)_{i,j} = (u_{i,j})_{i,j}$ , i.e., the tangent tuples all correspond to zero vectors, then  $\text{M-TGV}_\alpha^2(u) = \alpha_1 \text{TV}(u)$  with  $\text{TV}$  as in (16).
- (M-P3') If the minimum in (15) is attained at  $(y_{i,j}^1)_{i,j} = (u_{i+1,j})_{i,j}$  and  $(y_{i,j}^2)_{i,j} = (u_{i,j+1})_{i,j}$ , i.e., the tangent tuples  $[u_{i,j}, y_{i,j}^1]$  and  $[u_{i,j}, y_{i,j}^2]$  all correspond to  $(\delta_x + u)_{i,j}$  and  $(\delta_y + u)_{i,j}$ , respectively, then  $\text{M-TGV}_\alpha^2(u) = \alpha_0 \text{TV}^2(u)$  with  $\text{TV}^2$  an admissible generalization of  $\text{TV}^2$ .
- (M-P4')  $\text{M-TGV}_\alpha^2(u) = 0$  if and only if  $u$  is locally geodesic according to Definition 2.15.

Those properties translate to requirements for the involved distance-type functions as follows.

**Proposition 2.17.** Assume that the function  $D : \mathcal{M}^2 \times \mathcal{M}^2 \rightarrow [0, \infty)$  satisfies the assumptions of Proposition 2.8 and assume that  $D^{\text{sym}}$  is such that  $D^{\text{sym}}([x, x], [x, x], [u, u], [u, u]) = 0$  for any  $x, u \in \mathcal{M}$  and, in case  $\mathcal{M} = \mathbb{R}^K$ ,

$$D^{\text{sym}}([u_{\circ, \circ}, y_{\circ, \circ}^1], [u_{\circ, \circ}, y_{\circ, \circ}^2], [u_{\circ, -}, y_{\circ, -}^1], [u_{-, \circ}, y_{-, \circ}^2]) \\ = |y_{\circ, \circ}^1 - u_{\circ, \circ} - (y_{\circ, -}^1 - u_{\circ, -}) + y_{\circ, \circ}^2 - u_{\circ, \circ} - (y_{-, \circ}^2 - u_{-, \circ})|.$$

Then, for  $M\text{-}TGV_\alpha^2$  as in (3), the properties **(M-P1')** to **(M-P3')** hold. If further

$$D^{\text{sym}}([v_{i,j}, v_{i+1,j}], [v_{i,j}, v_{i,j+1}], [v_{i,j-1}, v_{i+1,j-1}], [v_{i-1,j}, v_{i-1,j+1}]) = 0,$$

for any geodesic three-by-three signal  $v = (v_{i,j})_{i,j}$ , then  $M\text{-}TGV_\alpha^2(u) = 0$  for any locally geodesic  $u$ . Conversely, assume that for any three-by-three signal in  $v = (v_{i,j})_{i,j}$  where the geodesic connecting each pair  $v_{i,j}, v_{i',j'}$  are unique and where  $(v_{i,j})_i$  and  $(v_{i,j})_j$  are geodesic, it holds that

$$D^{\text{sym}}([v_{i,j}, v_{i+1,j}], [v_{i,j}, v_{i,j+1}], [v_{i,j-1}, v_{i+1,j-1}], [v_{i-1,j}, v_{i-1,j+1}]) = 0,$$

implies also  $(v_{i+k,j-k})_{k=-1}^1$  being geodesic. Then, for any  $u = (u_{i,j})_{i,j}$  such that the geodesic connecting each pair  $u_{i,j}$  and  $u_{i',j'}$  with  $\max\{|i - i'|, |j - j'|\} \leq 2$  is unique, we get that  $M\text{-}TGV_\alpha^2(u) = 0$  implies  $u$  being geodesic.

*Proof.* See Section A.1 in the Appendix. □

Similar to the univariate case, the most restrictive requirement for  $D^{\text{sym}}$  is that, in case of uniqueness of length-minimizing geodesics, for  $(u_{i,j})_{i,j}$  being locally geodesic in horizontal and vertical direction, it holds that

$$D^{\text{sym}}([u_{i,j}, u_{i+1,j}], [u_{i,j}, u_{i,j+1}], [u_{i,j-1}, u_{i+1,j-1}], [u_{i-1,j}, u_{i-1,j+1}]) = 0 \\ \iff (u_{i+k,j-k})_k \text{ is locally geodesic.} \quad (17)$$

Let us discuss, again temporarily assuming uniqueness of geodesics, a general strategy to design the function  $D^{\text{sym}}$  such that this property holds. We consider the situation around a point  $u_{\circ, \circ}$  and denote by  $w_{\circ, \circ}^1, w_{\circ, \circ}^2, w_{\circ, -}^1, w_{-, \circ}^2$  four tangent vectors corresponding to the four tangent tuples  $[u_{\circ, \circ}, y_{\circ, \circ}^1], [u_{\circ, \circ}, y_{\circ, \circ}^2], [u_{\circ, -}, y_{\circ, -}^1], [u_{-, \circ}, y_{-, \circ}^2]$  at which  $D^{\text{sym}}$  is evaluated, see Figure 3. The evaluation of  $D^{\text{sym}}$  at these points should, in the vector space case, correspond to one of the equivalent formulations

$$|w_{\circ, \circ}^1 - w_{\circ, -}^1 + w_{\circ, \circ}^2 - w_{-, \circ}^2| = |(w_{\circ, \circ}^1 + w_{\circ, \circ}^2) - (w_{\circ, -}^1 + w_{-, \circ}^2)| = |(w_{\circ, \circ}^1 - w_{\circ, -}^1) - (w_{-, \circ}^2 - w_{\circ, \circ}^2)|.$$

Disregarding for the moment the fact that the corresponding tangent tuples live on different locations, the difficulty here is how to define algebraic operations, i.e., sum or difference operations, on two tangent tuples. For the sum, there is no direct way of doing so, the difference operation, however, quite naturally transfers to tangent tuples as

$$[x, y_1] - [x, y_2] = (y_1 - x) - (y_2 - x) = y_1 - y_2 = [y_2, y_1].$$

Hence, in the situation that all tangent tuples live on the same location, we can define  $D^{\text{sym}}$  by first evaluating pair-wise differences of the four point tuples and then measuring the distance of the two resulting tuples by  $D(\cdot, \cdot)$ .

Now in the general situation, we first need to transport tangent tuples to the same location. As described in Section 2.2, we have means of doing this that build either on parallel transport

or its Schild's approximation. Taking this into account, we note that the evaluation of  $D^{\text{sym}}$  in the vector space case can be written purely in terms of distances of tangent-vector-differences in various ways, e.g.,

$$|w_{o,o}^1 - w_{o,-}^1 + w_{o,o}^2 - w_{-,o}^2| = |(w_{o,o}^1 - w_{o,-}^1) - (w_{-,o}^2 - w_{o,o}^2)| = |(w_{o,-}^1 - w_{o,o}^2) - (w_{o,o}^1 - w_{-,o}^2)|. \quad (18)$$

While any order of taking the differences is equivalent in the vector space case, in the manifold setting the order defines how the corresponding tangent tuples need to be transported to the same location. Looking again at Figure 3, the simplest idea seems to be to transport the tangent tuple  $[u_{o,-}, y_{o,-}^1]$  corresponding to  $w_{o,-}^1$  and the tangent tuple  $[u_{-,o}, y_{-,o}^2]$  corresponding to  $w_{-,o}^2$  to the point  $u_{o,o}$  and carry out all operations there.

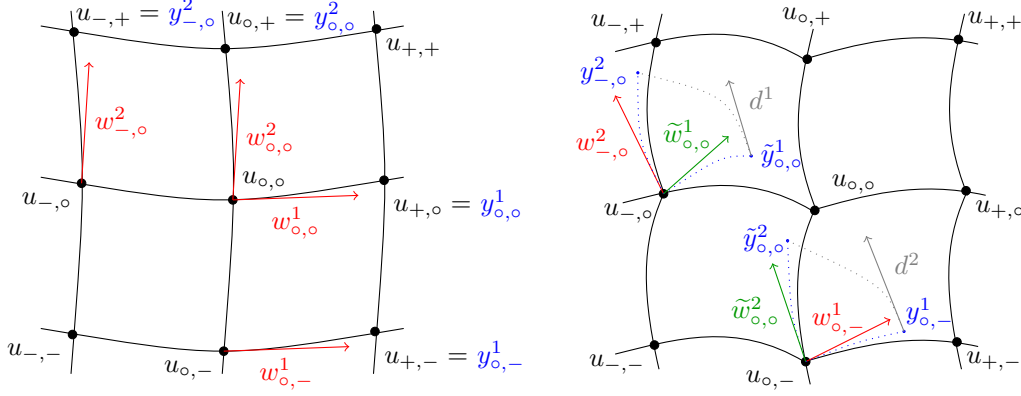
The drawback of this simple solution can be found when looking at the conditions of Proposition 2.17. There, the main difficulty is to ensure that

$$D^{\text{sym}}([u_{o,o}, u_{+,o}], [u_{o,o}, u_{o,+}], [u_{o,-}, u_{+,-}], [u_{-,o}, u_{-,+}]) = 0$$

allows to conclude that the signal is geodesic in the anti-diagonal direction. The situation that is relevant for this condition is when both  $(u_{i,j_0})_i$  and  $(u_{i_0,j})_j$  are locally geodesic for each  $(i_0, j_0)$  and the endpoints  $y$  coincide with the respective signal points, i.e.,  $y_{i,j}^1 = u_{i+1,j}$  and  $y_{i,j}^2 = u_{i,j+1}$  (see Figure 4, left). In this situation, in order to fulfill the above condition, we need to know something about the transported tangent tuples, e.g., if  $[u_{o,-}, y_{o,-}^1]$  and  $[u_{-,o}, y_{-,o}^2]$  corresponding to  $w_{o,-}^1$  and  $w_{-,o}^2$  are both transported to  $u_{o,o}$ , we would need to know for instance that the tangent tuple  $[u_{o,o}, x]$  resulting from the transport of  $[u_{o,-}, y_{o,-}^1]$  to  $u_{o,o}$  points to  $u_{+,o}$ , i.e.,  $x = u_{+,o}$ . Due to holonomy however, even if the transport is carried out along a geodesic, there is, to the best of our knowledge, no way of obtaining such a result. That is, the vector corresponding to the transported tangent tuple  $[u_{o,o}, x]$  might be arbitrarily rotated such that  $x$  is far away from  $u_{+,o}$ .

On the other hand, a well-known fact for manifolds is that the parallel transport of a vector that is tangential to a geodesic along the geodesic itself again results in a vector that is tangential to the geodesic. As we will see, this implies an equivalent assertion for both our variants for transporting tangent tuples. That is, the transport of a tangent tuple corresponding to a tangential vector of a geodesic along the geodesic itself again results in a tangent tuple that corresponds to a tangential vector of the geodesic. In the above-described particular situation (see again Figure 4, left), this means for example that, since the points  $u_{o,+}, u_{o,o}, u_{o,-}$  are on a geodesic and  $w_{o,o}^2 \simeq [u_{o,o}, u_{o,+}]$  corresponds to a tangent vector that is tangential to this geodesic, we know that the transport of  $[u_{o,o}, u_{o,+}]$  to  $u_{o,-}$  will result in the tangent tuple  $[u_{o,-}, u_{o,o}]$ . More generally, for the situation that both  $(u_{i,j_0})_i$  and  $(u_{i_0,j})_j$  are locally geodesic for each  $(i_0, j_0)$ , it means that we can always transport the tangent tuples corresponding to  $w_{o,o}^1$  and  $w_{o,-}^1$  in horizontal direction and the tangent tuples corresponding to the  $w_{o,o}^2$  and  $w_{-,o}^2$  in vertical direction and still know something about the transported tangent tuples.

In view of **(M-P4')**, a natural approach to define  $D^{\text{sym}}$  in the general situation is hence to restrict ourselves to these particular transport directions. Of course, we also want to define  $D^{\text{sym}}$  in an as-simple-as-possible way, meaning that we want to carry out as few transport operations as possible. A quick case study in Figure 3, shows that we have to transport at least three times and that, accounting for the facts that we need to generalize the expression (18) and that we can take differences but not sums of tangent tuples in the same location, there is only one possibility to achieve this. As highlighted in Figure 4 on the right for the general situation, our proposed approach is to transport the tangent tuple corresponding to  $w_{o,o}^2$  down to  $u_{o,-}$  giving  $[u_{o,-}, \tilde{y}_{o,o}^2]$ , the tangent tuple corresponding to  $w_{o,o}^1$  left to  $u_{-,o}$  giving  $[u_{-,o}, \tilde{y}_{o,o}^1]$  and take the differences, which results in the tangent tuples  $[y_{o,-}^1, \tilde{y}_{o,o}^2] \simeq d^2$  and  $[\tilde{y}_{o,o}^1, y_{-,o}^2] \simeq d^1$ . The distance of the



**Figure 4:** Two three-by-three point sections of a bivariate signal. Left: Signal points ( $u$ ), tangent vectors ( $w$ ) and endpoints ( $y$ ) in the particular situation that  $M\text{-TGV}_\alpha^2 = 0$ . Right: Signal with tangent vectors ( $\tilde{w}$ ) corresponding to transported tangent tuples, transported endpoints ( $\tilde{y}$ ) and ( $d$ ) corresponding to difference of tangent tuples. The dotted lines indicate geodesics that are determined by tangent vectors and the black lines show a piecewise geodesic interpolation of the signal points.

tangent tuples corresponding to  $d^1$  and  $d^2$  can then be measured with one of our previously defined distance functions, i.e., by evaluating  $D([\tilde{y}_{o,o}^1, y_{-,o}^2], [y_{o,-}^1, \tilde{y}_{o,o}^2])$ , which again requires one transport operation. This corresponds to the reformulation of

$$|w_{o,o}^1 - w_{o,-}^1 + w_{o,o}^2 - w_{-,o}^2| = |(w_{o,o}^2 - w_{o,-}^1) - (w_{-,o}^2 - w_{o,o}^1)|.$$

In the particular situation of Figure 4, left, due to the usage of the particular transport directions, the two tangent tuples resulting from the transport of the tangent tuples corresponding to  $w_{o,o}^1$  and  $w_{o,o}^2$  and taking the corresponding difference operations will be  $[u_{o,o}, u_{-,+}]$  and  $[u_{+,-}, u_{o,o}]$  (corresponding to  $d^1$  and  $d^2$ ), whose distance will then be measured by  $D([u_{o,o}, u_{+,-}], [u_{-,+}, u_{o,o}])$ . Using the assumptions as in Proposition 2.8,  $D([u_{o,o}, u_{+,-}], [u_{-,+}, u_{o,o}]) = 0$  then allows us to conclude that  $u_{-,+}, u_{o,o}, u_{+,-}$  are on a distance minimizing geodesic, hence the assumptions for **(M-P4')** will be fulfilled.

The above-described strategy is now used to extend both  $S\text{-TGV}_\alpha^2$  and  $PT\text{-TGV}_\alpha^2$  to the bivariate setting. As such, the only difference will be how to carry out the transport operations and which of the two functions  $D_S, D_{pt}$  is used to measure the distance of tangent tuples.

**Realization via Schild's-approximation.** We now realize the axiomatic setting for bivariate TGV for manifold-valued data of Section 2.3 using the proposed Schild's approximation. In particular, we use the Schild's approximation of parallel transport to obtain an instance of the distance-type function  $D^{\text{sym}}$ .

We define the following bivariate version  $S\text{-TGV}_\alpha^2$  of bivariate TGV for manifold-valued data using the Schild's approximation by

$$\begin{aligned} S\text{-TGV}_\alpha^2(u) = & \min_{y_{i,j}^1, y_{i,j}^2} \alpha_1 \sum_{i,j} \left( d(u_{i+1,j}, y_{i,j}^1)^p + d(u_{i,j+1}, y_{i,j}^2)^p \right)^{1/p} \\ & + \alpha_0 \sum_{i,j} \left( D_S([u_{i,j}, y_{i,j}^1], [u_{i-1,j}, y_{i-1,j}^1])^p + D_S([u_{i,j}, y_{i,j}^2], [u_{i,j-1}, y_{i,j-1}^2])^p \right. \\ & \left. + 2^{1-p} D_S^{\text{sym}}([u_{i,j}, y_{i,j}^1], [u_{i,j}, y_{i,j}^2], [u_{i,j-1}, y_{i,j-1}^1], [u_{i-1,j}, y_{i-1,j}^2])^p \right)^{1/p}. \end{aligned} \quad (19)$$

Here  $D_S$  is given as in Equation (8) and  $D_S^{\text{sym}}$  is defined by

$$\begin{aligned} D_S^{\text{sym}}([u_{o,o}, y_{o,o}^1], [\tilde{u}_{o,o}, y_{o,o}^2], [u_{o,-}, y_{o,-}^1], [u_{-,o}, y_{-,o}^2]) = \\ \min_{r^1, r^2} D_S([r^1, y_{-,o}^2], [y_{o,-}^1, r^2]) \quad \text{s.t. } r^1 \in [u_{o,o}, c^1]_2 \text{ with } c^1 \in [u_{-,o}, y_{o,o}^1]_{\frac{1}{2}} \\ \text{and } r^2 \in [\tilde{u}_{o,o}, c^2]_2 \text{ with } c^2 \in [u_{o,-}, y_{o,o}^2]_{\frac{1}{2}}. \end{aligned} \quad (20)$$

We note that, for the sake of a formally correct definition, we have introduced  $\tilde{u}_{o,o}$ , but in fact we will always choose  $\tilde{u}_{o,o} = u_{o,o}$  since  $D_S^{\text{sym}}$  effectively depends only on seven variables.

Also note that  $D_S^{\text{sym}}$  exactly carries out the construction of  $D^{\text{sym}}$  as described in Section 2.3, using the Schild's approximation as shown in Figure 2 for transporting tangent tuples and  $D_S$  to measure their difference. In the notation of Figure 4, right, we have that  $\tilde{y}_{o,o}^1 = r^1$  and  $\tilde{y}_{o,o}^2 = r^2$  correspond to the transported endpoints,  $[u_{-,o}, r^1] \simeq \tilde{w}_{o,o}^1$  and  $[u_{o,-}, r^2] \simeq \tilde{w}_{o,o}^2$  to the transported tangent tuples and  $[r^1, y_{-,o}^2] \simeq d^1$  and  $[y_{o,-}^1, r^2] \simeq d^2$  to the differences of tangent tuples whose distance is measured with  $D_S$ . Due to our careful design of  $S\text{-TGV}_\alpha^2$  we get the following result.

**Theorem 2.18.** *The  $S\text{-TGV}_\alpha^2$  functional as in (19) satisfies the properties (M-P1') to (M-P3'). If  $u = (u_{i,j})_{i,j}$  is locally geodesic, then  $S\text{-TGV}_\alpha^2(u) = 0$ . If  $u$  is such that the geodesic connecting each pair  $u_{i,j}$  and  $u_{i',j'}$  with  $\max\{|i-i'|, |j-j'|\} \leq 2$  is unique and  $S\text{-TGV}_\alpha^2(u) = 0$ , then  $u$  is geodesic.*

*Proof.* See Section A.1 in the Appendix.  $\square$

**Realization via parallel transport.** Using parallel transport in order to realize a symmetrized gradient as described in Section 2.3 yields a parallel transport version  $\text{PT-TGV}_\alpha^2$  of TGV for bivariate manifold-valued data case as follows. We let

$$\begin{aligned} \text{PT-TGV}_\alpha^2(u) = \min_{y_{i,j}^1, y_{i,j}^2} \alpha_1 \sum_{i,j} (d(u_{i+1,j}, y_{i,j}^1)^p + d(u_{i,j+1}, y_{i,j}^2)^p)^{1/p} \\ + \alpha_0 \sum_{i,j} \left( D_{\text{pt}}([u_{i,j}, y_{i,j}^1], [u_{i-1,j}, y_{i-1,j}^1])^p + D_{\text{pt}}([u_{i,j}, y_{i,j}^2], [u_{i,j-1}, y_{i,j-1}^2])^p \right. \\ \left. + 2^{1-p} D_{\text{pt}}^{\text{sym}}([u_{i,j}, y_{i,j}^1], [u_{i,j}, y_{i,j}^2], [u_{i,j-1}, y_{i,j-1}^1], [u_{i-1,j}, y_{i-1,j}^2])^p \right)^{1/p}. \end{aligned} \quad (21)$$

Here  $D_{\text{pt}}$  is given as in (11) and we define

$$\begin{aligned} D_{\text{pt}}^{\text{sym}}([u_{o,o}, y_{o,o}^1], [\tilde{u}_{o,o}, y_{o,o}^2], [u_{o,-}, y_{o,-}^1], [u_{-,o}, y_{-,o}^2]) = \min_{r^1, r^2} D_{\text{pt}}([r^1, y_{-,o}^2], [y_{o,-}^1, r^2]) \\ \text{s.t. } r^1 \in \exp(\text{pt}_{u_{-,o}}(w^1)) \text{ with } w^1 \in \log_{u_{o,o}}(y_{o,o}^1) \\ \text{and } r^2 \in \exp(\text{pt}_{u_{o,-}}(w^2)) \text{ with } w^2 \in \log_{\tilde{u}_{o,o}}(y_{o,o}^2). \end{aligned} \quad (22)$$

Note that, as for  $D_S^{\text{sym}}$ , for the sake of a formal correctness, we have introduced  $\tilde{u}_{o,o}$ , but in fact we will always choose  $\tilde{u}_{o,o} = u_{o,o}$  since also  $D_{\text{pt}}^{\text{sym}}$  effectively depends only on seven variables.

Also, we note again that  $D_{\text{pt}}^{\text{sym}}$  realizes exactly the construction of  $D^{\text{sym}}$  as discussed in Section 2.3 (see also Figure 4), only that now the parallel transport is used to transport tangent tuples and the distance of two tangent-tuple-differences is again measured with  $D_{\text{pt}}$ . Due to our careful construction, the following result follows easily.

**Theorem 2.19.** *The  $\text{PT-TGV}_\alpha^2$  functional as in (21) satisfies the properties (M-P1') to (M-P3'). If  $u = (u_{i,j})_{i,j}$  is locally geodesic, then  $\text{PT-TGV}_\alpha^2(u) = 0$ . If  $u$  is such that the geodesic connecting each pair  $u_{i,j}$  and  $u_{i',j'}$  with  $\max\{|i-i'|, |j-j'|\} \leq 2$  is unique and  $\text{PT-TGV}_\alpha^2(u) = 0$ , then  $u$  is geodesic.*



*Proof.* See Section A.1 in the Appendix.  $\square$

### 3 Existence results for TGV for manifold-valued data

This section provides existence results for the minimization problem appearing the definition of  $M\text{-TGV}_\alpha^2$  as well as for variational  $M\text{-TGV}_\alpha^2$ -regularized denoising. For the sake of brevity, we only consider the bivariate setting and note that the univariate counterpart follows as special case. We first state the main existence results in a general setting, that builds on lower semi-continuity of the involved distance-type function. The latter is then proven in subsequent lemmata.

**Theorem 3.1.** *Assume that  $D : \mathcal{M}^2 \times \mathcal{M}^2 \rightarrow [0, \infty)$  and  $D^{\text{sym}} : \mathcal{M}^2 \times \mathcal{M}^2 \times \mathcal{M}^2 \times \mathcal{M}^2 \rightarrow [0, \infty)$  are distance-type functions that are lower semi-continuous. Then, for any  $u = (u_{i,j})_{i,j}$  there exist  $(y_{i,j}^1)_{i,j}, (y_{i,j}^2)_{i,j}$  in  $\mathcal{M}$  that attain the minimum in the definition of  $M\text{-TGV}_\alpha^2$  as in Equation (15). Further,  $M\text{-TGV}_\alpha^2$  is lower semi-continuous and for any  $(f_{i,j})_{i,j}$  in  $\mathcal{M}$  there exists a solution to*

$$\min_u M\text{-TGV}_\alpha^2(u) + \sum_{i,j} d(u_{i,j}, f_{i,j})^2. \quad (23)$$

*Also, all previously introduced distance-type functions  $D_S, D_S^{\text{sym}}, D_{\text{pt}}, D_{\text{pt}}^{\text{sym}}$  are lower semi-continuous and a solution to (23) exists if  $M\text{-TGV}_\alpha^2$  is replaced by  $S\text{-TGV}_\alpha^2$  or  $PT\text{-TGV}_\alpha^2$ . In particular, all the models proposed in this paper have a minimizer.*

*Proof.* The existence results in the general setting follow by standard arguments and the Hopf-Rinow theorem and we only provide a sketch of proof. Regarding existence for the evaluation of  $M\text{-TGV}_\alpha^2$ , we note that all involved terms are positive and, for fixed  $u$ , any infimizing sequence  $((y^1)^n, (y^2)^n)_n$  is bounded due to the first two distance functions involved in  $M\text{-TGV}_\alpha^2$  as in Equation (15). Hence it admits a convergent subsequence and from lower semi-continuity of all the involved terms, it follows that the limit is a minimizer for the evaluation of  $M\text{-TGV}_\alpha^2$ .

Now take any sequence  $(u^n)_n$  converging to some  $u$  for which, without loss of generality, we assume that  $\liminf_n M\text{-TGV}_\alpha^2(u^n) = \lim_n M\text{-TGV}_\alpha^2(u^n)$ . We can pick elements  $((y^1)^n, (y^2)^n)$  for which the minimum in  $M\text{-TGV}_\alpha^2(u^n)$  is attained. Again by boundedness of  $(u^n)_n$  and the first two distance functions in the definition of  $M\text{-TGV}_\alpha^2$ , also  $((y^1)^n, (y^2)^n)_n$  is bounded and we can extract a subsequence  $((y^1)^{n_k}, (y^2)^{n_k})_k$  converging to some  $(\hat{y}^1, \hat{y}^2)$ . Defining  $E(\cdot)$  such that  $M\text{-TGV}_\alpha^2(u) = \inf_{y^1, y^2} E(u, y^1, y^2)$ , we obtain from lower semi-continuity of all involved terms that

$$M\text{-TGV}_\alpha^2(u) \leq E(u, \hat{y}^1, \hat{y}^2) \leq \liminf_k E(u^{n_k}, (y^1)^{n_k}, (y^2)^{n_k}) = \liminf_k M\text{-TGV}_\alpha^2(u^{n_k})$$

and since  $\liminf_k M\text{-TGV}_\alpha^2(u^{n_k}) = \liminf_n M\text{-TGV}_\alpha^2(u^n)$ ,  $M\text{-TGV}_\alpha^2$  is lower semi-continuous.

Given lower semi-continuity of  $M\text{-TGV}_\alpha^2$ , existence for (23) follows by similar arguments. Regarding the particular realizations  $S\text{-TGV}_\alpha^2$  and  $PT\text{-TGV}_\alpha^2$ , it suffices to show lower semi-continuity of the involved distance-type functions, which is done on the sequel.  $\square$

We are now left to show lower semi-continuity of the proposed distance-type functions  $D_S, D_{\text{pt}}, D_S^{\text{sym}}$  and  $D_{\text{pt}}^{\text{sym}}$ . To this aim, we exploit that they are all defined by minimizing a distance-type function subject to a constraint. Using this joint structure, we first provide a standard abstract lower semi-continuity result that covers this setting and reduces lower semi-continuity to a closedness condition on the constraint.

**Proposition 3.2.** Take  $K, N \in \mathbb{N}$  and functions  $G : \mathcal{M}^N \times \mathcal{M}^K \rightarrow [0, \infty)$  and  $C : \mathcal{M}^N \rightarrow \mathcal{P}(\mathcal{M}^K)$  such that  $G$  is lower semi-continuous,  $C(x) \neq \emptyset$  for all  $x \in \mathcal{M}^N$  and for any bounded sequence  $(x^n)_n$  with elements in  $\mathcal{M}^N$  and  $(y^n)_n$  such that  $y^n \in C(x^n)$  for all  $n$ , there exist subsequences  $(x^{n_k})_k$  and  $(y^{n_k})_k$  converging to some  $x \in \mathcal{M}^N$ ,  $y \in \mathcal{M}^K$  such that  $y \in C(x)$ . Define  $S : \mathcal{M}^N \rightarrow [0, \infty)$  as

$$S(x) = \inf_{y \in \mathcal{M}^K} G(x, y) \text{ such that } y \in C(x).$$

Then, for any  $x \in \mathcal{M}^N$  there exists  $y \in C(x)$  such that  $S(x) = G(x, y)$ . Further,  $S$  is lower semi-continuous.

*Proof.* See Section A.2 in the Appendix.  $\square$

In the application of this result to the proposed distance-type functions, the constraint  $y \in C(x)$  will correspond to constraining points to particular positions on distance minimizing geodesics. Consequently, in order to verify the closedness condition, the general result that the set of shortest geodesics connecting two points is closed w.r.t. perturbation of these points will be necessary and is provided as follows.

**Lemma 3.3.** Let  $\mathcal{M}$  be a (geodesically) complete Riemannian manifold. Let  $(p^n)_n$  and  $(q^n)_n$  be two sequences in  $\mathcal{M}$  converging to  $p$  and  $q$ , respectively. Let each  $\gamma^n : [0, 1] \rightarrow \mathcal{M}$  be a distance-minimizing geodesic such that  $\gamma^n(0) = p^n$ ,  $\gamma^n(1) = q^n$ . Then, there exists a distance-minimizing geodesic  $\gamma : [0, 1] \rightarrow \mathcal{M}$  such that  $\gamma(0) = p$ ,  $\gamma(1) = q$  and a subsequence  $(\gamma^{n_k})_k$  such that  $\gamma^{n_k} \rightarrow \gamma$  uniformly on  $[0, 1]$ . Furthermore, extending the geodesics to any interval  $[a, b]$  with  $[0, 1] \subset [a, b]$  we get that, up to subsequences,  $(\gamma^{n_k})_{n_k}$  also converges uniformly to  $\gamma$  on  $[a, b]$ .

*Proof.* See Section A.2 in the Appendix.  $\square$

We are now in the position of showing closedness of the constraints appearing in the definitions of  $D_S$  and  $D_S^{\text{sym}}$ .

**Lemma 3.4.** Let  $((x^n, y^n, u^n, v^n))_n$  be a bounded sequence in  $\mathcal{M}^4$  and take  $((c^n, \tilde{y}^n))_n$  in  $\mathcal{M}^2$  such that

$$c^n \in [x^n, v^n]_{\frac{1}{2}} \quad \text{and} \quad \tilde{y}^n \in [u^n, c^n]_2.$$

Then there exist subsequences  $((x^{n_k}, y^{n_k}, u^{n_k}, v^{n_k}))_k$  and  $((c^{n_k}, \tilde{y}^{n_k}))_k$  converging to  $(x, y, u, v)$  and  $(c, \tilde{y})$ , respectively, such that

$$c \in [x, v]_{\frac{1}{2}} \quad \text{and} \quad \tilde{y} \in [u, c]_2.$$

In particular,  $C : \mathcal{M}^4 \rightarrow \mathcal{P}(\mathcal{M}^2)$  defined as  $C(x, y, u, v) := \{(c', y') \mid c' \in [x, v]_{\frac{1}{2}}, y' \in [u, c']_2\}$  satisfies the assumption of Proposition 3.2.

*Proof.* Since each  $c^n$  is on a length-minimizing geodesic between  $x^n$  and  $v^n$ , we get that  $d(x^n, c^n) \leq d(x^n, v^n)$  and hence  $(c^n)_n$  is bounded. Also  $d(c^n, \tilde{y}^n) = d(c^n, u^n)$  and hence  $(\tilde{y}^n)_n$  is bounded. Consequently we can pick subsequences  $((x^{n_k}, y^{n_k}, u^{n_k}, v^{n_k}))_k$  and  $((c^{n_k}, \tilde{y}^{n_k}))_k$  converging to  $(x, y, u, v)$  and  $(c, \tilde{y})$ . Now with  $\gamma_c^{n_k} : [0, 1] \rightarrow \mathcal{M}$  being a shortest geodesic connecting  $x^{n_k}$  and  $v^{n_k}$  such that  $\gamma_c^{n_k}(1/2) = c^{n_k}$  we get by Lemma 3.3 that, up to a subsequence,  $\gamma_c^{n_k} \rightarrow \gamma_c$  uniformly, where  $\gamma_c$  is again a shortest geodesic connecting  $x$  and  $v$ . Consequently,  $c = \gamma_c(1/2) \in [x, v]_{\frac{1}{2}}$ .

Now pick  $\gamma_{\tilde{y}}^{n_k} : [0, 1] \rightarrow \mathcal{M}$  to be a length minimizing geodesic between  $\gamma_{\tilde{y}}^{n_k}(0) = u^{n_k}$  and  $\gamma_{\tilde{y}}^{n_k}(1) = c^{n_k}$  such that  $\gamma_{\tilde{y}}^{n_k}(2) = \tilde{y}^{n_k}$ . By Lemma 3.3 we get that, up to subsequences,  $\gamma_{\tilde{y}}^{n_k}$  converges uniformly to a geodesic  $\gamma_{\tilde{y}}$  on  $[0, 2]$  such that  $\gamma_{\tilde{y}}$  is length-minimizing between  $\gamma_{\tilde{y}}(0) = u$  and  $\gamma_{\tilde{y}}(1) = c$ . By uniform convergence on  $[0, 2]$  we get that  $y = \lim_k y^{n_k} = \lim_k \gamma_{\tilde{y}}^{n_k}(2) = \gamma_{\tilde{y}}(2)$  and the assertion follows.  $\square$

Combining this with the general assertion of Proposition 3.2, existence and lower semi-continuity results for  $D_S$  and  $D_S^{\text{sym}}$  follow as direct consequences.

**Lemma 3.5.** *The minimum in the definition of  $D_S$  as in Equation (8) is attained and  $D_S$  is lower semi-continuous. Also, the minimum in the definition of  $D_S^{\text{sym}}$  as in Equation (20) is attained and  $D_S^{\text{sym}}$  is lower semi-continuous.*

*Proof.* See Section A.2 in the Appendix.  $\square$

Using similar techniques, existence and lower semi-continuity results for the parallel transport variants can be established as follows.

**Lemma 3.6.** *The minimum in the definition of  $D_{pt}$  as in Equation (11) is attained and  $D_{pt}$  is lower semi-continuous. Also, the minimum in the definition of  $D_{pt}^{\text{sym}}$  as in Equation (22) is attained and  $D_{pt}^{\text{sym}}$  is lower semi-continuous.*

*Proof.* See Section A.2 in the Appendix.  $\square$

## 4 Algorithmic approach to TGV for manifold-valued data

In order to algorithmically approach denoising problems using TGV regularization in the manifold setting, we employ the concept of cyclic proximal point algorithms (CPPAs). A reference for cyclic proximal point algorithms in vector spaces is [22]. In the context of Hadamard spaces, the concept of CPPAs was developed by [8], where it is used to compute means and medians. In the context of variational regularization methods for nonlinear, manifold-valued data, they were first used in [93]. More precisely, the reference [93] deals with TV regularization as well as classical smooth methods for manifold-valued data. The CPPA approach was later used for higher-order TV-type methods in [18] for circle-valued data and in [10] for manifold-valued data.

**Principle of a CPPA.** The idea of a CPPAs is to compose a functional  $f : \mathcal{M} \rightarrow \mathbb{R}$  into basic atoms  $f_i$  and then to compute the proximal mappings of the  $f_i$  in a cyclic, iterative way. More precisely, assume that

$$f = \sum_{i=1}^n f_i \quad (24)$$

and consider the proximal mappings [64, 41, 6]  $\text{prox}_{\lambda f_i} : \mathcal{M} \rightarrow \mathcal{M}$  given as

$$\text{prox}_{\lambda f_i} x = \underset{y}{\text{argmin}} f_i(y) + \frac{1}{2\lambda} d(x, y)^2. \quad (25)$$

One cycle of a CPPA then consists of applying each proximal mapping  $\text{prox}_{\lambda f_i}$  once in a prescribed order, e.g.,  $\text{prox}_{\lambda f_1}, \text{prox}_{\lambda f_2}, \text{prox}_{\lambda f_3}, \dots$ , or, generally,  $\text{prox}_{\lambda f_{\sigma(1)}}, \text{prox}_{\lambda f_{\sigma(2)}}, \text{prox}_{\lambda f_{\sigma(3)}}, \dots$ , where the symbol  $\sigma$  is employed to denote a permutation. The cyclic nature is reflected in the fact that the output of  $\text{prox}_{\lambda f_{\sigma(i)}}$  is used as input for  $\text{prox}_{\lambda f_{\sigma(i+1)}}$ . Since the  $i$ th update is immediately used for the  $(i+1)$ st step, it can be seen as a Gauss-Seidel-type scheme. A CPPA now consists of iterating these cycles, i.e., in the  $k$ th cycle, we have

$$x_{i+1}^{(k)} = \text{prox}_{\lambda_k f_{\sigma(i)}} x_i^{(k)}, \quad (26)$$

and  $x_1^{(k+1)}$  is obtained by applying  $\text{prox}_{\lambda_k f_{\sigma(n)}}$  to  $x_n^{(k)}$ . Here,  $n$  denotes the number of elements in the cycle. During the iteration, the parameter  $\lambda_k$  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  $(\lambda^k)_k$  is square-summable but not summable. Provided that this condition on the stepsize parameters holds, the cyclic proximal point algorithm can be shown to converge to the optimal solution of the underlying minimization problem, at least in the context of Hadamard manifolds and convex  $(f_i)_i$  [9, Theorem 3.1].

For general manifolds, the proximal mappings (25) are not globally defined, and the minimizers are not unique, at least for general possibly far apart points; cf. [41, 6]. This is a general issue in the context of manifolds that are – in a certain sense – a local concept involving objects that are often only locally well defined. In case of ambiguities, we hence consider the above objects as set-valued quantities. Furthermore, we cannot guarantee – and in fact do not expect – the second distance-type functions  $D$  in the definition of the  $\text{M-TGV}_\alpha^2$  functional to be convex. Hence convergence of the CPPA algorithm to a globally optimal solution cannot be ensured. It thus should be seen as an approximative strategy. Nevertheless, as will be seen in the numerical experiments section, we experience a good convergence behavior of the CPPA algorithm in practice. This was also observed in previous works, where the CPPA algorithm was employed to minimize second-order TV-type functionals [18, 10], which are also non-convex. In order to be precise, we further point out that we approximately compute the proximal mappings of the distance-type functions  $D$ ; we hence use an inexact proximal point algorithm.

We also point out that a parallel proximal point algorithm has been proposed in [93]. Here the proximal mappings of the  $f_i$  are computed in parallel and then averaged using intrinsic means. In order to reduce the costs for averaging, an approximative variant of the parallel proximal point algorithm has been proposed in [93] as well. Principally, the cyclic proximal point algorithm actually applied in this paper might be replaced by this parallel strategy; the derivations in the following provide all information necessary.

#### 4.1 Algorithms for manifold-valued TGV

We employ a cyclic proximal point algorithm for the minimization of a TGV-regularized variational approach in the manifold context. For simplicity, we consider the univariate setting first; all aspects discussed in the univariate situation are prototypic and basic for the bivariate situation. Starting with the general version of TGV for manifolds as in (3), we aim to solve the denoising problem

$$\min_{(u_i)_i, (y_i)_i} \frac{1}{2} \sum_i d(u_i, h_i)^2 + \sum_i \alpha_1 d(u_{i+1}, y_i) + \sum_i \alpha_0 D([u_i, y_i], [u_{i-1}, y_{i-1}]), \quad (27)$$

where  $(h_i)_i$ , being a finite sequence in  $\mathcal{M}$ , denotes the given data.

We decompose (27) into the atoms

$$\begin{aligned} g_i(u) &:= \frac{1}{2} d(u_i, h_i)^2; & g'_i(u, y) &:= \alpha_1 d(u_{i+1}, y_i); \\ g''_i(u, y) &:= \alpha_0 D([u_i, y_i], [u_{i-1}, y_{i-1}]). \end{aligned} \quad (28)$$

Now we use this decomposition into the atoms  $g_i, g'_i, g''_i$  for the decomposition (24) for the CPPA. We apply the iteration (26) to this decomposition. We remark that the data terms  $g_i$  are not coupled w.r.t. the index  $i$ . The same applies to the  $g'_i$ . This allows the parallel computation of the proximal mappings of the  $g_i$  for all  $i$  and, separately, of the  $g'_i$  for all  $i$ . The  $g''_i$  are actually coupled. However, grouping even and odd indices, we see that the  $g''_i$ ,  $i$  even, are not coupled and that the  $g''_i$ ,  $i$  odd, are not coupled. So we may compute their proximal mappings in parallel. Together, this results in a cycle of length four per iteration, one step for the  $g_i$ , one step for the  $g'_i$  and two steps (even, odd) for the  $g''_i$ .

In the following, the task is to compute the proximal mappings of the atoms  $g_i, g'_i, g''_i$  of (28). To this end, we will from now on always assume that the involved points are locally sufficiently

close such that there exist unique length minimizing geodesics connecting them. We note that this restriction is actually not severe: in the general (non-local) situation we have to remove the cut points – which are a set of measure zero – to end up with the corresponding setup.

We remark that, as we will see, an explicit computation of the proximal mapping is possible for  $g_i$  and  $g'_i$ , but not for  $g''_i$ . We believe that it is an important feature of the proposed definition of manifold-TGV via point-tuples that the proximal mappings of the first term  $g'_i$  is still explicit and hence only one part of the overall problem does not allow an explicit proximal mapping. Indeed, also existing generalizations of second-order TV to the manifold setting incorporate one part with non-explicit proximal mappings [10]. Hence, from the algorithmic viewpoint, the step from second-order TV to our proposed version of TGV does not introduce essential additional computational effort.

For the data atoms  $g_i$ , the proximal mappings  $\text{prox}_{\lambda g_i}$  are given by

$$(\text{prox}_{\lambda g_i})_j(u) = \begin{cases} [u_i, h_i]_t, & \text{if } i = j, \\ u_j, & \text{else,} \end{cases} \quad \text{where } t = \frac{\lambda}{1+\lambda}. \quad (29)$$

They have been derived in [41]. We recall that the symbol  $[\cdot, \cdot]_t$  denotes the point reached after time  $t$  on the (non unit speed) geodesic starting at the first argument reaching the second argument at time 1. Further, the proximal mappings of the distance terms  $g'_i$  have a closed form representation as well (see [93]), which is given by

$$\text{prox}_{\lambda g'_i}(u, y) = \begin{cases} [u_{i+1}, y_i]_t, & \text{at position } (i+1, 1), \\ [y_i, u_{i+1}]_t, & \text{at position } (i, 2), \\ u_j, & \text{at position } (j, 1), \ j \neq i+1, \\ y_j & \text{at position } (j, 2), \ j \neq i, \end{cases} \quad (30)$$

where  $t = \lambda \alpha_1 / d(u_{i+1}, y_i)$ , if  $\lambda \alpha_1 < \frac{1}{2}$ , and  $t = 1/2$  else. Note that for determining the position, we view  $u, y$  as column vectors of a matrix with two columns.

The next task is to compute the proximal mappings of the  $g''_i$ . Unfortunately, no closed form for the proximal mappings of the  $g''_i$  is available. Instead, we use a subgradient descent scheme to compute these proximal mappings, i.e., to solve the problem

$$\min_{u, y} \quad \frac{1}{2} \sum_j d(u_j, u'_j)^2 + \frac{1}{2} \sum_j d(y_j, y'_j)^2 + \lambda \alpha_0 D([u_i, y_i], [u_{i-1}, y_{i-1}]) \quad (31)$$

where the  $u'_j, y'_j$  are the input of the proximal mapping. A subgradient descent scheme has already been used to compute proximal mappings in [10]. Looking at (31), the optimal  $u, y$  fulfill  $u_j = u'_j, y_j = y'_j$  whenever  $j \notin \{i, i-1\}$ . Hence we may restrict ourselves to consider the four arguments  $u_i, y_i, u_{i-1}, y_{i-1}$ . Further the gradient of the mapping  $u_j \mapsto \frac{1}{2} \sum_j d(u_j, u'_j)^2$  is given as  $-\log_{u_j} u'_j$ . (For background information on (Riemannian) differential geometry, we refer to the books [76, 38].) So we have to compute the (sub)gradient of the mapping  $D([u_i, y_i], [u_{i-1}, y_{i-1}])$  as a function of  $u_i, y_i, u_{i-1}, y_{i-1}$ . To this end, we have to specify the setting to the two versions of  $D$  as proposed in Section 2; the respective derivations are topics of Section 4.2 and Section 4.3.

We next consider the bivariate denoising problem. We use M-TGV $^2_\alpha$  regularization with the M-TGV $^2_\alpha$  functional given in (15) where, as pointed out, we focus on  $p = 1$ . We obtain the

denoising problem

$$\begin{aligned}
\min_{(u_{i,j})_{i,j}, (y_{i,j}^1)_{i,j}, (y_{i,j}^2)_{i,j}} \quad & \frac{1}{2} \sum_{i,j} d(u_{ij}, h_{ij})^2 + \alpha_1 \sum_{i,j} d(u_{i+1,j}, y_{i,j}^1) + d(u_{i,j+1}, y_{i,j}^2) \\
& + \alpha_0 \sum_{i,j} D([u_{i,j}, y_{i,j}^1], [u_{i-1,j}, y_{i-1,j}^1]) + D([u_{i,j}, y_{i,j}^2], [u_{i,j-1}, y_{i,j-1}^2]) \\
& + \alpha_0 \sum_{i,j} D^{\text{sym}}([u_{i,j}, y_{i,j}^1], [u_{i,j}, y_{i,j}^2], [u_{i,j-1}, y_{i,j-1}^1], [u_{i-1,j}, y_{i-1,j}^2])
\end{aligned} \tag{32}$$

where the  $(h_{ij})_{i,j}$  denotes the bivariate data.

We decompose the objective functional in (32) into the atoms

$$\begin{aligned}
g_{ij}^{(1)}(u, y^1, y^2) &:= \frac{1}{2} d(u_i, h_{ij})^2; \\
g_{ij}^{(2)}(u, y^1, y^2) &:= \alpha_1 d(u_{i+1,j}, y_{i,j}^1); \quad g_{ij}^{(3)}(u, y^1, y^2) := \alpha_1 d(u_{i,j+1}, y_{i,j}^2); \\
g_{ij}^{(4)}(u, y^1, y^2) &:= \alpha_0 D([u_{i,j}, y_{i,j}^1], [u_{i-1,j}, y_{i-1,j}^1]); \\
g_{ij}^{(5)}(u, y^1, y^2) &:= \alpha_0 D([u_{i,j}, y_{i,j}^2], [u_{i,j-1}, y_{i,j-1}^2]); \\
g_{ij}^{(6)}(u, y^1, y^2) &:= \alpha_0 D^{\text{sym}}([u_{i,j}, y_{i,j}^1], [u_{i,j}, y_{i,j}^2], [u_{i,j-1}, y_{i,j-1}^1], [u_{i-1,j}, y_{i-1,j}^2]).
\end{aligned} \tag{33}$$

We use the decomposition w.r.t. the atoms  $g_{ij}^{(k)}$  as the decomposition (24) for the CPPA, and apply the iteration (26) w.r.t. the derived decomposition. The proximal mappings  $\text{prox}_{\lambda g_{ij}^{(1)}}$  are given by (29), and the proximal mappings  $\text{prox}_{\lambda g_{ij}^{(2)}}$ ,  $\text{prox}_{\lambda g_{ij}^{(3)}}$  are given by (30). The proximal mappings  $\text{prox}_{\lambda g_{ij}^{(4)}}$ ,  $\text{prox}_{\lambda g_{ij}^{(5)}}$  are computed as in the univariate case explained above. It remains to compute the proximal mappings of the atoms  $g_{ij}^{(6)}$ . As before, there is no explicit formula available and we use a subgradient descent for their computation as well. The computation of the corresponding derivatives are topics of Section 4.2 and Section 4.3.

## 4.2 Riemannian gradients for the Schild version

We here derive the derivatives needed to compute the proximal mappings of the mappings  $D$  and  $D^{\text{sym}}$  in Section 4.1 when specifying to the Schild's variant. We start out to compute the derivative of the mapping  $D_S$  given by

$$D_S(u_{i-1}, y_{i-1}, u_i, y_i) = D_S([u_i, y_i], [u_{i-1}, y_{i-1}]) = d([u_{i-1}, [u_i, y_{i-1}]_{\frac{1}{2}}]_2, y_i). \tag{34}$$

Recall that, for our computations, we assume that geodesics are unique, which is, as pointed out above, the case up to a set of measure zero.

We directly see that  $D_S$  is symmetric w.r.t. interchanging  $y_{i-1}, u_i$ . Hence we only have to compute the differential with respect to one of these variables  $y_{i-1}, u_i$ .

Furthermore, the gradient  $\nabla_{y_i} D_S$  w.r.t. the fourth argument  $y_i$  of  $D_S$  for  $y_i$  with  $y_i \neq [u_{i-1}, [u_i, y_{i-1}]_{\frac{1}{2}}]_2$  is just the well-known gradient of the Riemannian distance function which is known to be given by [38, 4]

$$\nabla_{y_i} D_S = -\log_{y_i} S(u_{i-1}, y_{i-1}, u_i) / |\log_{y_i} S(u_{i-1}, y_{i-1}, u_i)| \tag{35}$$

where  $|\cdot|$  is the norm in the tangent space associated with the point  $y_i$ , i.e., the square root of the Riemannian scalar product. Here,

$$S(u_{i-1}, y_{i-1}, u_i) = [u_{i-1}, [u_i, y_{i-1}]_{\frac{1}{2}}]_2 \tag{36}$$

denotes the result of applying the Schild's construction to the respective arguments.

In order to determine the gradients w.r.t. the other variables we have to apply the adjoint of the respective differentials to the gradient of the distance function w.r.t. the first argument. More precisely, we have to calculate

$$\nabla_{u_{i-1}} D_S = -T_1 \left( \log_{S(u_{i-1}, y_{i-1}, u_i)} y_i / \left| \log_{S(u_{i-1}, y_{i-1}, u_i)} y_i \right| \right), \quad (37)$$

$$\nabla_{y_{i-1}} D_S = -T_2 \left( \log_{S(u_{i-1}, y_{i-1}, u_i)} y_i / \left| \log_{S(u_{i-1}, y_{i-1}, u_i)} y_i \right| \right), \quad (38)$$

where  $T_1$  is the adjoint of the differential of the mapping  $u_{i-1} \mapsto [u_{i-1}, [u_i, y_{i-1}]_{\frac{1}{2}}]_2$ , and where  $T_2$  is the adjoint of the differential of the mapping  $y_{i-1} \mapsto [u_{i-1}, [u_i, y_{i-1}]_{\frac{1}{2}}]_2$ . The differential w.r.t.  $u_i$  is obtained by symmetry as pointed out above. In the following we derive these adjoint mappings in terms of Jacobi fields.

To this aim, we first explain the notion of a Jacobi field and then point out the connection with geodesic variations. In a Riemannian manifold  $\mathcal{M}$ , the Riemannian curvature tensor  $R$  is given by  $R(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X, Y]} Z$  where  $X, Y, Z$  are vector fields and  $\nabla$  denotes the Levi-Civita connection, and where  $[X, Y] = XY - YX$  denotes the Lie bracket of the vector fields  $X, Y$ . A Jacobi field  $Y$  along a geodesic  $\gamma$  is the solution of the differential equation  $\frac{D}{ds} \frac{D}{ds} Y + R(\frac{d}{ds} \gamma, Y) \frac{d}{ds} \gamma = 0$ . The space of Jacobi fields is a  $2N$ -dimensional linear space where  $N$  denotes the dimension of  $\mathcal{M}$ . The connection to geodesic variations is as follows: the “derivative” vector field  $Y(s) = \frac{d}{dt}|_{t=0} V(s, t)$  of a geodesic variation  $V$  is a Jacobi field, and conversely, any Jacobi field is obtained from a geodesic variation. Further details may be found in the books [34, 76, 38].

We also need the notion of (Riemannian) symmetric spaces for the formulation of the next lemma. A symmetric space is a Riemannian manifold for which its curvature tensor  $R$  is invariant under parallel transport,  $R(\text{pt}_{x,y} X, \text{pt}_{x,y} Y) \text{pt}_{x,y} Z = \text{pt}_{x,y} R(X, Y)Z$  where  $X, Y, Z$  are tangent vectors at  $x$  and  $\text{pt}_{x,y}$  denotes the parallel transport from  $x$  to  $y$  along a curve  $\gamma$  (here not reflected in the notation). Then, in particular, the covariant derivative of the curvature tensor  $R$  along any curve equals zero in a symmetric space. A reference for symmetric spaces is [34].

**Lemma 4.1.** *The mapping  $T_1$  of Equation (37) can be computed using Jacobi fields. In particular, we explicitly have for the class of symmetric spaces and points with  $D_S \neq 0$  that*

$$\nabla_{u_{i-1}} D_S = \text{pt}_{S(u_{i-1}, y_{i-1}, u_i), u_{i-1}} \left( \log_{S(u_{i-1}, y_{i-1}, u_i)} y_i / \left| \log_{S(u_{i-1}, y_{i-1}, u_i)} y_i \right| \right), \quad (39)$$

which means that the gradient of the distance function w.r.t. the second argument is reflected at  $[u_i, y_{i-1}]_{\frac{1}{2}}$ , or in other words, parallel transported from the Schild point  $[u_{i-1}, [u_i, y_{i-1}]_{\frac{1}{2}}]_2$  to  $u_{i-1}$  and multiplied by  $-1$ .

*Proof.* The proof is given in Section A.3. □

Regarding the fact that we consider points with  $D_S \neq 0$ , we remark that the points  $u_i, u_{i-1}, y_i, y_{i-1}$  such that  $D_S(u_i, u_{i-1}, y_i, y_{i-1}) = 0$  which corresponds to  $y_i = [u_{i-1}, [u_i, y_{i-1}]_{\frac{1}{2}}]_2$  form a set of measure zero. On this zero set, for instance, the four-tuple consisting of the four zero-tangent vectors sitting in  $u_i, u_{i-1}, [u_{i-1}, [u_i, y_{i-1}]_{\frac{1}{2}}]_2, y_{i-1}$  belong to the subgradient of  $D_S$ .

**Lemma 4.2.** *The gradient of the function  $D_S$  for points with  $D_S \neq 0$  given by (34) w.r.t. the variable  $y_{i-1}$  is given by*

$$\nabla_{y_{i-1}} D_S = -T_4 \left( T_3 \left( \log_{S(u_{i-1}, y_{i-1}, u_i)} y_i / \left| \log_{S(u_{i-1}, y_{i-1}, u_i)} y_i \right| \right) \right), \quad (40)$$

where  $T_3$  is the adjoint of the derivative of the mapping  $m \mapsto [u_{i-1}, m]_2$ , and where  $T_4$  is the adjoint of the derivative of the mapping  $y_{i-1} \mapsto [u_i, y_{i-1}]_{\frac{1}{2}}$ , that is,  $T_2$  as in (38) is given as  $T_2 = T_4 \circ T_3$ . Further  $T_3$  and  $T_4$  can be computed using Jacobi fields.

*Proof.* The proof is given in Section A.3.  $\square$

For the class of symmetric spaces, we make the mappings  $T_3, T_4$  more explicit.

**Lemma 4.3.** *Let  $\mathcal{M}$  be a symmetric space. Let  $\gamma : [0, 1] \rightarrow \mathcal{M}$  denote the geodesic connecting  $\gamma(0) = u_{i-1}$  and  $\gamma(1) = m = [u_i, y_{i-1}]_{\frac{1}{2}}$ . We consider an orthonormal basis  $v_n$  of eigenvectors of the self-adjoint Jacobi operator  $J \mapsto R(\frac{\gamma'(1)}{|\gamma'(1)|}, J) \frac{\gamma'(1)}{|\gamma'(1)|}$  with corresponding eigenvalues  $\lambda_n$ , and  $v_1$  tangent to  $\gamma$ . W.r.t. this basis, the mapping  $T_3$  is given by*

$$T_3 : \sum_n \alpha_n \text{pt}_{[u_i, y_{i-1}]_{\frac{1}{2}}, S(u_{i-1}, y_{i-1}, u_i)} v_n \mapsto \sum_n \alpha_n f(\lambda_n) v_n, \quad (41)$$

where, using the symbol  $d$  for the distance between  $u_{i-1}$  and  $[u_i, y_{i-1}]_{\frac{1}{2}}$ ,

$$f(\lambda_n) = \begin{cases} 2, & \text{if } \lambda_n = 0, \\ \sin(2\sqrt{\lambda_n}d)/\sin(\sqrt{\lambda_n}d), & \text{if } \lambda_n > 0, \quad d < \pi/\sqrt{\lambda_n}, \\ \sinh(2\sqrt{-\lambda_n}d)/\sinh(\sqrt{-\lambda_n}d), & \text{if } \lambda_n < 0. \end{cases} \quad (42)$$

Here, the  $\alpha_n$  are the coefficients of the corresponding basis representation.

Further, let  $\xi : [0, 1] \rightarrow \mathcal{M}$  be the geodesic connecting  $y_{i-1} = \xi(0)$  and  $u_i = \xi(1)$  and  $w_n$  be an orthonormal basis of eigenvectors of  $J \mapsto R(\frac{\xi'(0)}{|\xi'(0)|}, J) \frac{\xi'(0)}{|\xi'(0)|}$  with eigenvalues  $\mu_n$ , and  $w_1$  tangent to  $\xi$ . Then, w.r.t. this basis, the mapping  $T_4$  is given by

$$T_4 : \sum_n \beta_n \text{pt}_{y_{i-1}, [u_i, y_{i-1}]_{\frac{1}{2}}} w_n \mapsto \sum_n \beta_n g(\lambda_n) w_n, \quad (43)$$

where, using the symbol  $d'$  for the distance between  $y_{i-1}$  and  $u_i$ ,

$$g(\lambda_n) = \begin{cases} 1/2, & \text{if } \lambda_n = 0, \\ \sin(\frac{1}{2}\sqrt{\lambda_n}d')/\sin(\sqrt{\lambda_n}d'), & \text{if } \lambda_n > 0, \quad d < \pi/\sqrt{\lambda_n}, \\ \sinh(\frac{1}{2}\sqrt{-\lambda_n}d')/\sinh(\sqrt{-\lambda_n}d'), & \text{if } \lambda_n < 0. \end{cases} \quad (44)$$

Here, the  $\beta_n$  are the coefficients of the corresponding basis representation.

*Proof.* The proof is given in Section A.3.  $\square$

Next, we consider the Riemannian gradients for the Schild version  $D^{\text{sym}} = D_S^{\text{sym}}$  as given in (20) which are needed for the CPPA for the bivariate TGV functional. To this end, we note that  $D_S^{\text{sym}}$  can be expressed in terms of the function  $S$  of (36) by

$$D_S^{\text{sym}}([u_{i,j}, y_{i,j}^1], [u_{i,j}, y_{i,j}^2], [u_{i,j-1}, y_{i,j-1}^1], [u_{i-1,j}, y_{i-1,j}^2]) = d(S(y_{i,j-1}^1, r^2, r^1), y_{i-1,j}^2) \\ \text{s.t. } r^1 = S(u_{i,j}, y_{i,j}^1, u_{i-1,j}), r^2 = S(u_{i,j}, y_{i,j}^2, u_{i,j-1}).$$

Hence, the mapping  $D_S^{\text{sym}}$  (which depends on seven arguments) is an expression of the Riemannian distance function and three realizations of  $S$ . We have seen how to differentiate  $S$  in Lemma 4.1 and Lemma 4.2. We derive the gradient of  $D_S^{\text{sym}}$  by iterated application of these mappings and the concepts of these lemmata. We point out the symmetry of  $D_S^{\text{sym}}$  with respect



to  $u_{i-1,j}, y_{i,j}^1$ , and with respect to  $u_{i,j-1}, y_{i,j}^2$ . This reduces the task to actually considering five different arguments. The derivative of  $D_S^{\text{sym}}$  is provided in the following proposition. Its proof is a direct consequence of the previous results and the decomposition of  $D_S^{\text{sym}}$  into decompositions of the function  $S$  and the distance function  $d$  as above.

**Proposition 4.4.** *We consider constellations of points with  $D_S^{\text{sym}} \neq 0$ . Then, the gradient of  $D_S^{\text{sym}}$  w.r.t. the variable  $y_{i,j}^1$  is given by*

$$\nabla_{y_{i,j}^1} D_S^{\text{sym}} = -T_4 \circ T_3 \circ \tilde{T}_4 \circ \tilde{T}_3 \left( \log_{S(y_{i,j-1}^1, r^2, r^1)} y_{i-1,j}^2 / \left| \log_{S(y_{i,j-1}^1, r^2, r^1)} y_{i-1,j}^2 \right| \right), \quad (45)$$

with the adjoint operators  $T_3, T_4$  given as in Lemma 4.2 formed w.r.t. the points  $u_{i,j}, u_{i-1,j}, y_{i,j}^1$  and the adjoint operators  $\tilde{T}_3, \tilde{T}_4$  given as in Lemma 4.2 formed w.r.t. the points  $y_{i,j-1}^1, r^2, r^1$ . The gradient of  $D_S^{\text{sym}}$  w.r.t.  $u_{i-1,j}$  has the same form by symmetry.

Similarly, the gradient of  $D_S^{\text{sym}}$  w.r.t. the variable  $y_{i,j}^2$  is given by

$$\nabla_{y_{i,j}^2} D_S^{\text{sym}} = -T_4 \circ T_3 \circ \tilde{T}_4 \circ \tilde{T}_3 \left( \log_{S(y_{i,j-1}^1, r^2, r^1)} y_{i-1,j}^2 / \left| \log_{S(y_{i,j-1}^1, r^2, r^1)} y_{i-1,j}^2 \right| \right), \quad (46)$$

with the adjoint operators  $T_3, T_4$  as in Lemma 4.2 formed w.r.t. the points  $u_{i,j}, u_{i,j-1}, y_{i,j}^2$  and the adjoint operators  $\tilde{T}_3, \tilde{T}_4$  as in Lemma 4.2 formed w.r.t. the points  $y_{i,j-1}^1, r^2, r^1$ . The gradient of  $D_S^{\text{sym}}$  w.r.t.  $u_{i,j-1}$  has the same form by symmetry.

The gradient of  $D_S^{\text{sym}}$  w.r.t. the variable  $y_{i,j-1}^1$  is given by

$$\nabla_{y_{i,j-1}^1} D_S^{\text{sym}} = -T_1 \left( \log_{S(y_{i,j-1}^1, r^2, r^1)} y_{i-1,j}^2 / \left| \log_{S(y_{i,j-1}^1, r^2, r^1)} y_{i-1,j}^2 \right| \right), \quad (47)$$

with the adjoint operator  $T_1$  given as in Lemma 4.1 formed w.r.t. the points  $y_{i,j-1}^1, r^2, r^1$ . The gradient with respect to the variable  $y_{i-1,j}^2$  is simply given by

$$\nabla_{y_{i-1,j}^2} D_S^{\text{sym}} = -\log_{y_{i-1,j}^2} S(y_{i,j-1}^1, r^2, r^1) / \left| \log_{y_{i-1,j}^2} S(y_{i,j-1}^1, r^2, r^1) \right|. \quad (48)$$

Finally, the gradient of  $D_S^{\text{sym}}$  w.r.t. the variable  $u_{i,j}$  is given by

$$\nabla_{u_{i,j}} D_S^{\text{sym}} = -T_1 \circ T_4 \circ T_3 \left( \frac{\log_{S(y_{i,j-1}^1, r^2, r^1)} y_{i-1,j}^2}{\left| \log_{S(y_{i,j-1}^1, r^2, r^1)} y_{i-1,j}^2 \right|} \right) - \tilde{T}_1 \circ T_4 \circ T_3 \left( \frac{\log_{S(y_{i,j-1}^1, r^2, r^1)} y_{i-1,j}^2}{\left| \log_{S(y_{i,j-1}^1, r^2, r^1)} y_{i-1,j}^2 \right|} \right). \quad (49)$$

Here the adjoint operators  $T_3$  and  $T_4$  are given as in Lemma 4.2 formed w.r.t. the points  $y_{i,j-1}^1, r^2, r^1$ ,  $T_1$  is given as in Lemma 4.1 formed w.r.t. the points  $u_{i,j}, y_{i,j}^1, u_{i-1,j}$  and  $\tilde{T}_1$  is given as in Lemma 4.1 formed w.r.t. the points  $u_{i,j}, y_{i,j}^2, u_{i,j-1}$ .

### 4.3 Riemannian gradients for the parallel transport version

We here derive the derivatives needed to compute the proximal mappings in Section 4.1 when specifying to the parallel transport variant. We use the short-hand notation

$$F(u_i, u_{i-1}, y_i, y_{i-1}) = D_{\text{pt}}([u_i, y_i], [u_{i-1}, y_{i-1}]). \quad (50)$$

For the following computations recall that, by our assumption of uniqueness of length-minimizing geodesics, the log mapping, initially defined as set-valued, always maps to single points in the tangent space.

**Lemma 4.5.** *The function  $F$  given by (50) is symmetric with respect to interchanging  $(u_i, y_i)$  with  $(u_{i-1}, y_{i-1})$ . In particular, for points with  $F \neq 0$ , the gradient of  $F$  w.r.t. the third variable  $y_i$ , is given by  $\nabla_{y_i} F(u_i, u_{i-1}, y_i, y_{i-1}) = \nabla_{y_{i-1}} F(u_{i-1}, u_i, y_{i-1}, y_i)$ . Further, again for points with  $F \neq 0$ , the gradient of the function  $F$  w.r.t. the first component variable  $u_i$ , is given by  $\nabla_{u_i} F(u_i, u_{i-1}, y_i, y_{i-1}) = \nabla_{u_{i-1}} F(u_{i-1}, u_i, y_{i-1}, y_i)$ .*

*Proof.* The proof is given in Section A.3 by specifying  $F = F_0$  in Lemma A.5.  $\square$

Regarding the fact that we consider points with  $F \neq 0$  in the above statement, we remark that the points  $u_i, u_{i-1}, y_i, y_{i-1}$  such that  $F(u_i, u_{i-1}, y_i, y_{i-1}) = 0$  form a set of measure zero. Only on this zero set,  $F$  is not differentiable. Further, in this case, the four-tuple consisting of the four zero-tangent vectors sitting in  $u_i, u_{i-1}, y_i, y_{i-1}$  belong to the subgradient of  $F$ . We note that Lemma 4.5 tells us that, in order to compute the gradient of the second order type difference  $F$ , we only need to compute the respective gradient of  $F$  w.r.t.  $u_{i-1}$  and  $y_{i-1}$ . This is done in the following.

**Lemma 4.6.** *The gradient of the function  $F$  for points with  $F \neq 0$  w.r.t. the variable  $y_{i-1}$  is given by*

$$\nabla_{y_{i-1}} F = T \left( \left( \log_{u_{i-1}} y_{i-1} - \text{pt}_{u_i, u_{i-1}} \log_{u_i} y_i \right) / \left| \log_{u_{i-1}} y_{i-1} - \text{pt}_{u_i, u_{i-1}} \log_{u_i} y_i \right| \right), \quad (51)$$

where  $T = (d_{y_{i-1}} \log_{u_{i-1}})^*$  is the adjoint of the (Fréchet) derivative (denoted by the symbol  $d_{y_{i-1}}$ ) of the log mapping w.r.t. the variable  $y_{i-1}$ .  $T$  can be computed using Jacobi fields.

*Proof.* The proof is given in Section A.3.  $\square$

In Riemannian symmetric spaces the above mapping  $T$  can be made more explicit.

**Lemma 4.7.** *Let  $\mathcal{M}$  be a symmetric space. Consider the geodesic  $\gamma : [0, 1] \rightarrow \mathcal{M}$  connecting  $\gamma(0) = u_{i-1}$  and  $\gamma(1) = y_{i-1}$ , and let  $d$  denote the distance between  $u_{i-1}$  and  $y_{i-1}$ . Let  $(v_n)_n$  be an orthonormal basis of eigenvectors of the self-adjoint Jacobi operator  $J \mapsto R(\frac{\gamma'(0)}{|\gamma'(0)|}, J) \frac{\gamma'(0)}{|\gamma'(0)|}$  with  $v_1$  tangent to  $\gamma$ , and denote, for each  $n$ , the eigenvalue associated with  $v_n$  by  $\lambda_n$ . W.r.t. this basis, the operator  $T$  of Lemma 4.6 can be represented by*

$$T : \sum_n \alpha_n v_n \mapsto \sum_n \alpha_n f(\lambda_n) \text{pt}_{u_{i-1}, y_{i-1}} v_n, \quad (52)$$

where the  $\alpha_n$  are the coefficients of the corresponding basis representation and the function  $f$ , depending on the sign of  $\lambda_n$  is given by

$$f(\lambda_n) = \begin{cases} 1, & \text{if } \lambda_n = 0, \\ \sqrt{\lambda_n} d / \sin(\sqrt{\lambda_n} d), & \text{if } \lambda_n > 0, \\ \sqrt{-\lambda_n} d / \sinh(\sqrt{-\lambda_n} d), & \text{if } \lambda_n < 0. \end{cases} \quad d < \pi / \sqrt{\lambda_n}, \quad (53)$$

*Proof.* The proof is given in Section A.3.  $\square$

Finally, we consider the gradient of  $F$  w.r.t. the variable  $u_{i-1}$ .

**Lemma 4.8.** *The gradient of the function  $F$  for points with  $F \neq 0$  w.r.t. the variable  $u_{i-1}$  is given by*

$$\nabla_{u_{i-1}} F_1 = \sum \alpha_n v_n, \quad (54)$$

where the  $(v_n)_n$  form an orthonormal basis of the tangent space at  $u_{i-1}$ , and the coefficients  $\alpha_n$  are given by

$$\alpha_n = \frac{d}{dt}|_{t=0} |L_t^n - B_t^n| = \left\langle \frac{L(u_{i-1}) - B(u_{i-1})}{|L(u_{i-1}) - B(u_{i-1})|}, \frac{D}{dt}|_{t=0} L_t^n - \frac{D}{dt}|_{t=0} B_t^n \right\rangle. \quad (55)$$

Here  $L_t^n, B_t^n$  denote the vector fields  $L, B$  (defined by  $L : u_{i-1} \mapsto \log_{u_{i-1}} y_{i-1}$  and  $B : u_{i-1} \mapsto \text{pt}_{u_i, u_{i-1}} z$ , where  $z = \log_{u_i} y_i$ ), along the (specific) geodesic  $t \mapsto \exp_{u_{i-1}} t v_n$ ,  $t \in [0, 1]$ , determined by  $v_n$ , i.e.,

$$L_t^n = \log_{\exp_{u_{i-1}} t v_n} y_{i-1} \quad \text{and} \quad B_t^n = \text{pt}_{u_i, \exp_{u_{i-1}} t v_n} z, \quad (56)$$

with  $z = \log_{u_i} y_i$ .

*Proof.* The proof is given in Section A.3.  $\square$

The precise computation of  $\frac{D}{dt}|_{t=0} L_t^n$  in symmetric spaces is topic of Lemma 4.9. Further, the computation of  $\frac{D}{dt}|_{t=0} B_t^n$  is carried out for the manifolds explicitly considered in this paper: this is done for the sphere in Lemma 4.10, and for the space of positive matrices in Lemma 4.11. Lemma 4.11 and its proof may serve as a prototypic guide for deriving similar expressions for other symmetric spaces such as the rotation groups or the Grassmannians for instance. We note that the approach is by no means restricted to the two considered classes of spaces and might serve as a guide for other manifolds; we only did not derive a more explicit representation on the general level of symmetric spaces. We further note that numerical differentiation of the particular term is a second option as well.

**Lemma 4.9.** *Assume that the manifold  $\mathcal{M}$  is a symmetric space. Let  $(v_n)_n$  be an orthonormal basis of eigenvectors of the self-adjoint Jacobi operator  $J \mapsto R(\frac{\gamma'(0)}{|\gamma'(0)|}, J) \frac{\gamma'(0)}{|\gamma'(0)|}$  where the (constant speed) geodesic  $\gamma : [0, 1] \rightarrow \mathcal{M}$  connects  $u_{i-1} = \gamma(0)$  and  $y_{i-1} = \gamma(1)$ , and where  $R$  denotes the Riemannian curvature tensor. For each  $n$ , we denote by  $\lambda_n$  the eigenvalue associated with  $v_n$ .*

*The covariate derivatives  $\frac{D}{dt}|_{t=0} L_t^n$ , of the vector fields  $L_t^n = \log_{\exp_{u_{i-1}} t v_n} y_{i-1}$  at  $t = 0$  can be computed jointly for all  $n$  using Jacobi fields as follows:*

$$\frac{D}{dt}|_{t=0} L_t^n = \begin{cases} -v_n, & \text{if } \lambda_n = 0, \\ -d\sqrt{\lambda_n} \frac{\cos(\sqrt{\lambda_n} d)}{\sin(\sqrt{\lambda_n} d)} v_n, & \text{if } \lambda_n > 0, \quad d < \pi/\sqrt{\lambda_n}, \\ -d\sqrt{-\lambda_n} \frac{\cosh(\sqrt{-\lambda_n} d)}{\sinh(\sqrt{-\lambda_n} d)} v_n, & \text{if } \lambda_n < 0. \end{cases} \quad (57)$$

Here  $d = d(u_{i-1}, y_{i-1})$  denotes the length of the geodesic connecting  $u_{i-1}, y_{i-1}$ . (If the term  $\sqrt{\lambda_n} d = 0$  in the denominators of the second line in (57), then  $u_{i-1} = y_{i-1}$ , and the formula is still valid since we are facing a removable singularity then.)

*Proof.* The proof is given in Section A.3.  $\square$

**Lemma 4.10.** *Consider the unit sphere  $S^2$  embedded into euclidean space  $\mathbb{R}^3$ . For  $u_i, u_{i-1}$  with  $u_i \neq u_{i-1}$ , the differential  $\frac{D}{dt}|_{t=0} B_t^n$  is given by*

$$\frac{D}{dt}|_{t=0} B_t^n = \frac{D}{dt}|_{t=0} \text{pt}_{u_i, \exp_{u_{i-1}} t v_n} z = \begin{cases} 0 & \text{for } v_n \parallel \log_{u_{i-1}} u_i, \\ L_\omega \text{pt}_{u_i, u_{i-1}} z, & \text{for } v_n \perp \log_{u_{i-1}} u_i, \quad |v_n| = 1, \end{cases} \quad (58)$$

and  $v_n$  to the left of  $\log_{u_{i-1}} u_i$  (otherwise multiplied by  $-1$  accounting for the change of orientation). Here the skew-symmetric matrix  $L_\omega = \begin{pmatrix} 0 & \omega \\ -\omega & 0 \end{pmatrix}$  is taken w.r.t. the basis  $\{\log_{u_{i-1}} u_i, (\log_{u_{i-1}} u_i)^\perp\}$  of the tangent space  $T_{u_{i-1}}$ , and  $\omega$  is given by  $\omega = \frac{1}{\sin d} - \frac{1}{\tan d}$ , where  $d = d(u_i, u_{i-1})$ . For general  $v_n$ ,

$$\frac{D}{dt}|_{t=0} B_t^n = \frac{D}{dt}|_{t=0} \text{pt}_{u_i, \exp_{u_{i-1}} t v_n} z = \langle v_n, w \rangle L_\omega \text{pt}_{u_i, u_{i-1}} z, \quad (59)$$

where  $w$  is the vector determined by  $w \perp \log_{u_{i-1}} u_i$ ,  $|w| = 1$ , and  $w$  is to the left of  $\log_{u_{i-1}} u_i$ . In other words, we have to multiply the second line of (58) with the signed length of the projection of  $v_n$  to the normalized vector  $(\log_{u_{i-1}} u_i)^\perp$ .

If  $u_i = u_{i-1}$ , then the differential  $\frac{D}{dt}|_{t=0} B_t^n = 0$  (which is consistent with letting  $d \rightarrow 0$  in the above formulae.)

*Proof.* The proof is given in Section A.3.  $\square$

**Lemma 4.11.** *Let  $\mathcal{M}$  be the space of symmetric positive definite matrices. Then, the covariate derivative  $\frac{D}{dt}|_{t=0} B_t^n$  (which is a tangent vector sitting in  $u_{i-1}$ ) is given by the following sum of matrices*

$$\frac{D}{dt}|_{t=0} B_t^n = (T - \frac{1}{2}S) + (T - \frac{1}{2}S)^\top, \quad (60)$$

where  $(T - \frac{1}{2}S)^\top$  denotes the transpose of the matrix  $T - \frac{1}{2}S$ . The matrix  $S$  is determined in terms of elementary matrix operations of the data (by (107) in the proof of the statement). The matrix  $T$  is determined in terms of elementary matrix operations and the solution of a Sylvester equation (by (109) in the proof of the statement with the Sylvester equation given by (112) there).

*Proof.* The proof is given in Section A.3.  $\square$

Summing up, we have computed the derivatives of all building blocks necessary to compute the derivative of  $F(u_i, u_{i-1}, y_i, y_{i-1}) = D_{\text{pt}}([u_i, y_i], [u_{i-1}, y_{i-1}])$  for the non-degenerate case  $F \neq 0$ .

**Remark 4.12.** *For the bivariate version of the parallel transport based  $M\text{-}TGV_\alpha^2$  realization of Section 2.3, we can use the analogue of the decomposition (33) with  $D_S, D_S^{\text{sym}}$  replaced by the corresponding parallel transport versions (10) and (22). Then we can use this analogue of the decomposition (33) and apply the CPPA iteration (26) to this decomposition. The proximal mappings  $\text{prox}_{\lambda g_{ij}^{(1)}}$  are given by (29),  $\text{prox}_{\lambda g_{ij}^{(2)}}, \text{prox}_{\lambda g_{ij}^{(3)}}$  are given by (30) as for the Schild variant above. The proximal mappings  $\text{prox}_{\lambda g_{ij}^{(4)}}, \text{prox}_{\lambda g_{ij}^{(5)}}$  are computed as in the univariate case considered in Section 4.3. In order to compute the proximal mappings of the atoms  $g_{ij}^{(6)}$  using a subgradient descent, it is necessary to differentiate the mapping  $D_{\text{pt}}^{\text{sym}}$  of (22) (which is the analogue of  $D_S^{\text{sym}}$  for the Schild case) w.r.t. its seven arguments. As in the Schild case, it is possible to decompose the mapping into simpler functions which we have already considered in Section 4.3. We do not carry out the rather space consuming derivation here.*

## 5 Numerical results

This section provides numerical results for synthetic and real signals and images. We first describe the experimental setup.

We carry out experiments for  $S^1$ ,  $S^2$  and  $\text{Pos}_3$  (the manifold of symmetric positive definite  $3 \times 3$  matrices equipped with the Fisher-Rao metric) valued data.  $S^1$  data is visualized by the phase angle, and color-coded as hue value in the HSV color space when displaying image data.

We visualize  $S^2$  data either by a scatter plot on the sphere as in Figure 6, or by a color coding as in Figure 9. Data on the  $\text{Pos}_3$  manifold is visualized by the isosurfaces of the corresponding quadratic forms. More precisely, the ellipse visualizing the point  $f_p$  at voxel  $p$  are the points  $x$  fulfilling  $(x - p)^\top f_p (x - p) = c$ , for some  $c > 0$ .

To quantitatively measure the quality of a reconstruction, we use the manifold variant of the *signal-to-noise ratio improvement*

$$\Delta\text{SNR} = 10 \log_{10} \left( \frac{\sum_{ij} d(g_{ij}, f_{ij})^2}{\sum_{ij} d(g_{ij}, u_{ij})^2} \right) \text{ dB},$$

see [86, 93]. Here  $f$  is the noisy data,  $g$  the ground truth, and  $u$  a regularized restoration. A higher  $\Delta\text{SNR}$  value means better reconstruction quality.

For adjusting the model parameters  $\alpha_0, \alpha_1$  of  $\text{M-TGV}_\alpha^2$ , it is convenient to parametrize them by

$$\alpha_0 = r \frac{(1-s)}{\min(s, 1-s)}, \quad \text{and} \quad \alpha_1 = r \frac{s}{\min(s, 1-s)}, \quad (61)$$

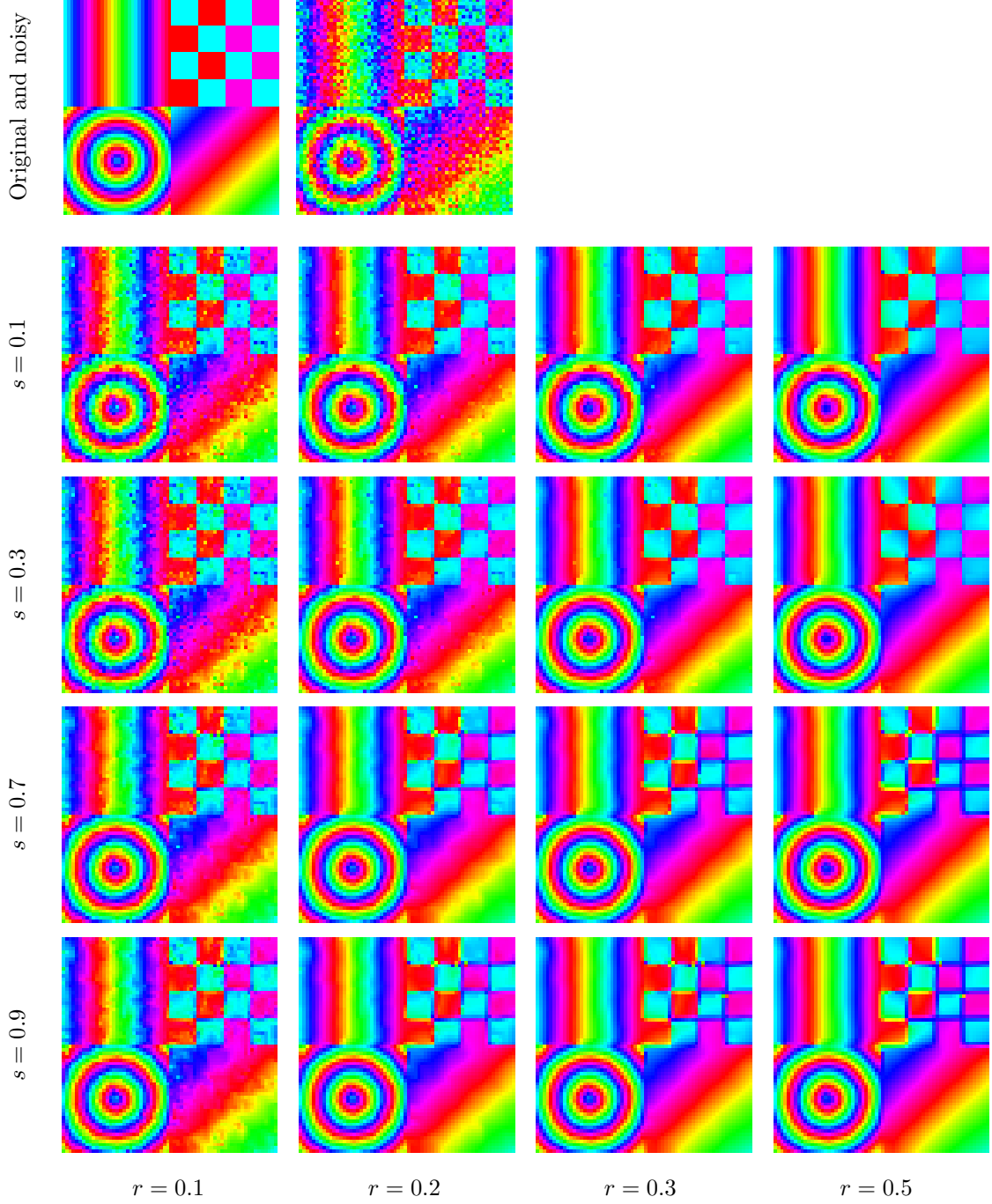
so that  $r \in (0, \infty)$  controls the overall regularization strength and  $s \in (0, 1)$  the balance of the two TGV penalties. For  $s \rightarrow 0$  we get  $\alpha_0 \rightarrow \infty$  and  $\alpha_1 = r$ , so that TGV minimization approximates the minimization of TV modulo a linear term which can be added at no cost. For  $s \rightarrow 1$  we have  $\alpha_0 = r$  and  $\alpha_1 \rightarrow \infty$  which corresponds to pure second-order TV regularization. One may think of  $r$  as the parameter mainly depending on the noise level, and of  $s$  as the parameter mainly depending on the geometry of the underlying image. Figure 5 illustrates the influence on  $r$  and  $s$  for a synthetic  $S^1$ -valued image corrupted by von Mises noise with concentration parameter  $\kappa = 5$ . There and in most of the following experiments, we observed satisfactory results for the fixed value  $s = 0.3$ . Based on these observations, we suggest to use  $s = 0.3$  as a starting point, to decrease or increase it if the image is dominated by edges or smooth parts, respectively.

We have implemented the presented methods in Matlab 2016b. We use the toolbox MVIRT [10] for the basic manifold operations such as log, exp and parallel transport, and for parts of the visualization<sup>1</sup>. We used 100000 iterations for all experiments with univariate data and 1000 iterations for the image data. The cooling sequence  $(\lambda^k)_{k \in \mathbb{N}}$  used as stepsize in the gradient descent for computing the non-explicit proximal mappings was chosen as  $\lambda^k = \lambda^0 k^{-\tau}$  with  $\tau = 0.55$ . For the univariate spherical data we observed faster convergence when using a stagewise cooling, i.e., letting the sequence fixed to  $\lambda_0$  for 500 iterations in the first stage, use the moderate cooling  $\tau = 0.35$  in the second stage until iteration 1000 and then the cooling  $\tau = 0.55$  afterwards.

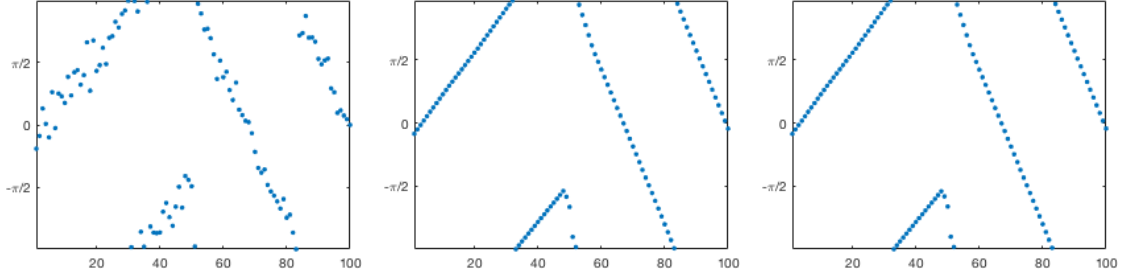
## 5.1 Numerical evaluation of the proposed models

**Comparison between Schild variant and parallel transport variant.** First we compare the two proposed realizations of manifold TGV: the parallel transport variant and the Schild variant. Figure 6 shows the results with both variants for some typical univariate signals for the  $S^1$ , the  $S^2$  and the  $\text{Pos}_3$  manifold. (The  $S^1$ -valued signal was corrupted by von Mises noise with concentration parameter  $\kappa = 10$ . The  $S^2$ -valued signal was taken from [10] and corrupted by applying the exponential mapping of Gaussian distributed tangent vectors as in [10] with  $\sigma = 0.1$ . The  $\text{Pos}_3$ -valued signal was corrupted by applying the exponential mapping of Gaussian noise distributed on the tangent vectors with  $\sigma = 0.2$ . The  $r$ -parameters were chosen as  $r = 1$  for the spherical signals and as  $r = 0.2$  for the  $\text{Pos}_3$  signal. In all cases  $s = 0.3$ .) The results for the spherical data appear very similar. For the  $\text{Pos}_3$  manifolds, there are slight visible differences at

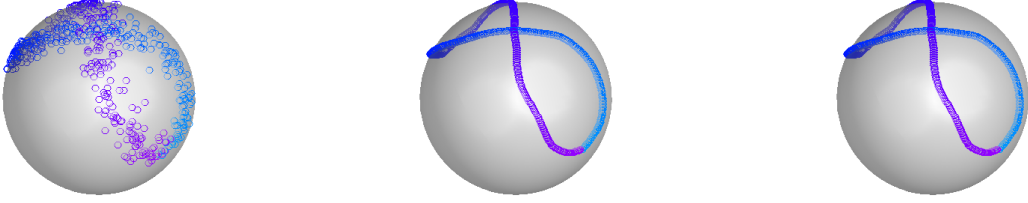
<sup>1</sup>Implementation available at <https://github.com/kellertuer/MVIRT>



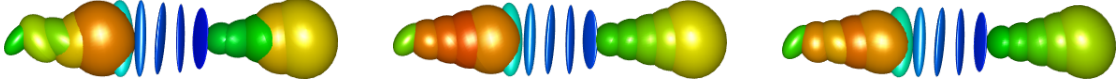
**Figure 5:** Effect of the model parameters of S-TGV using the parametrization by  $r > 0$  and  $s \in (0, 1)$  according to (61) for an  $S^1$ -valued image. A higher value of  $r$  results in a stronger smoothing. For small values of  $s$ , the edges are well-preserved but some staircasing effects appear. For high values of  $s$ , the linear trends are recovered but the edges are smoothed out. When using an intermediate value such as  $s = 0.3$ , we observe a combination of positive effects of TV and  $TV^2$  regularization: rather sharp edges and good recovery of linear trends.



(a)  $S^1$ -valued signal, visualized by phase angle.



(b)  $S^2$ -valued signal, visualized as scatter plot on the sphere.

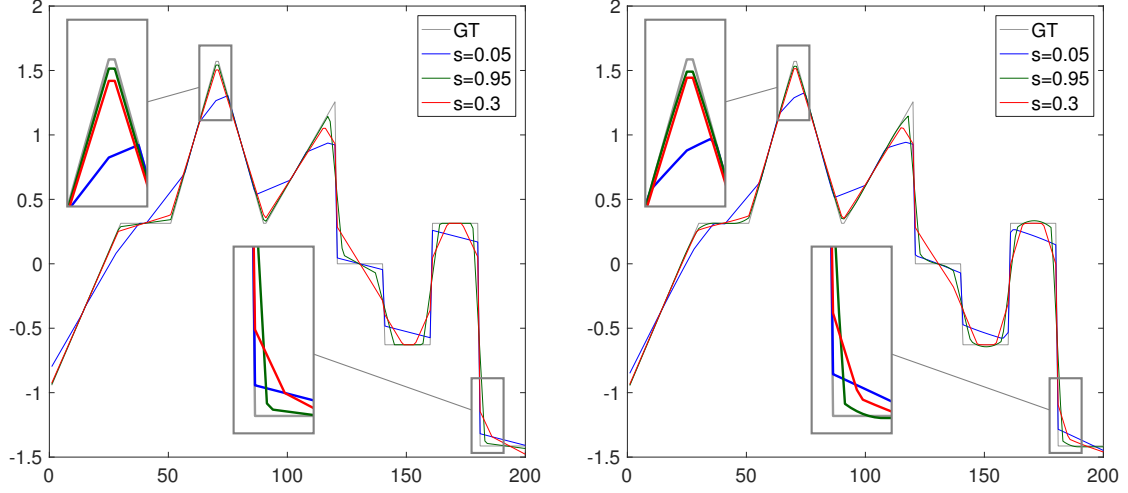


(c)  $\text{Pos}_3$ -valued signal, visualized as ellipsoids.

**Figure 6:** Comparison of parallel transport and Schild variant of  $\text{M-TGV}_\alpha^2$  for various data spaces. The subfigures show the noisy data (*left*), the result of the Schild variant S-TGV (*center*), and the result of the parallel transport variant PT-TGV (*right*).

some points, e.g. near the discontinuity of the second to fourth position of the signal and at the last position. In summary, we observe a qualitatively similar result in the sense that the geodesic parts are well reconstructed and that the edges are preserved. In the following, we focus on the Schild variant.

**Comparison of manifold TGV with vector-space TGV.** In order to validate both our generalizations of TGV to the manifold setting as well as numerical feasibility of our optimization algorithms, we carry out a comparison to the vector-space case. This is done for  $S^1$ , the unit sphere in  $\mathbb{R}^2$ . By generating a signal with values that are strictly contained in one hemisphere, we can unroll the signal and compare to vector-space TGV-denoising in a way that the same results can be expected. We carried out this comparison for synthetic data without noise and different values of the balancing parameter  $s$ . We tested the setting of  $s = 0.3$ , which comprises a good balance between first and second order terms, as well as the rather extreme settings of  $s = 0.05$  and  $s = 0.95$ . The regularization parameter  $r$  was fixed to  $r = 1$ . In order to approximate the ground-truth solution of second-order TGV denoising in the vector space setting, we implemented the Chambolle-Pock algorithm [32] for this situation. To ensure a close approximation of the



**Figure 7:** Comparison of the approximate vector-space ground truth and the result obtained with the manifold TGV implementation for univariate  $S^1$  data and different ratios of  $s$ . *Left:* Approximate ground truth obtained with reference vector space implementation. *Right:* Result with TGV manifold code.

ground truth for all parameter settings, we carried out a dedicate stepsize tuning to accelerate convergence of the algorithm, computed  $2 \times 10^5$  iterations and ensured optimality by measuring the duality gap.

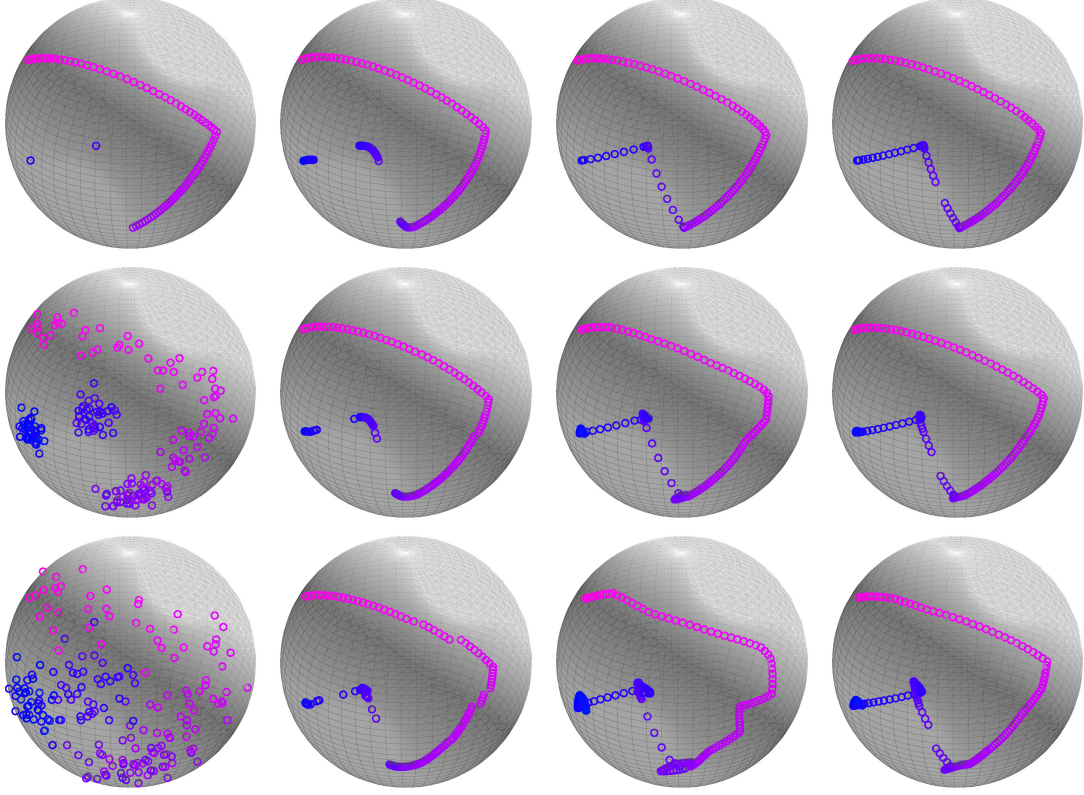
The result of this evaluation can be found in Figure 7. It can be observed that the qualitative properties of the numerical solution obtained with the manifold-TGV code are similar to the ones of the approximate ground-truth of the vector space setting, confirming overall feasibility of our model and implementation.

For the case  $s = 0.05$ , one can see in particular in the right part of the bottom plot (starting with the first plateau after 180 on the x-axis), that the solution is piecewise constant up to a linear part, which is approximately the same for all four plateaus. This is what one would expect for the extreme case  $s \rightarrow 0$ , as in this case, TGV minimization mimics TV minimization up to a globally linear term. On the other hand, for the case  $s = 0.95$ , one can see that the solution is piecewise linear with no jumps. This can be expected from second-order TV minimization, which coincides with second order TGV minimization for  $s \rightarrow 1$ . In particular, the piecewise linear part on the left is approximated well, while the plateaus on the right are not well captured. The case  $s = 0.3$  provides a good compromise here: The linear parts on the left are still well captured, but the solution still admits jumps on the plateau parts, as can be seen in particular outermost right jump after point.

We also note that there are still some differences of the solution obtained with the manifold code and the Chambolle-Pock result interpreted as approximate ground truth of the vector space case, in particular for the cases  $s = 0.05$  and  $s = 0.95$ . We believe that this is mostly an issue of the algorithmic realization rather than the model itself caused by the numerical solutions obtained for the more extreme cases  $s = 0.05$  and  $s = 0.95$ .

**Basic situations.** The aim of this experiment is to investigate the performance of the proposed manifold-TGV model for different basic situations and noise levels. The experiment is carried out on a univariate signal on the two-sphere  $S^2$ , that comprises jumps in the signal as well as its





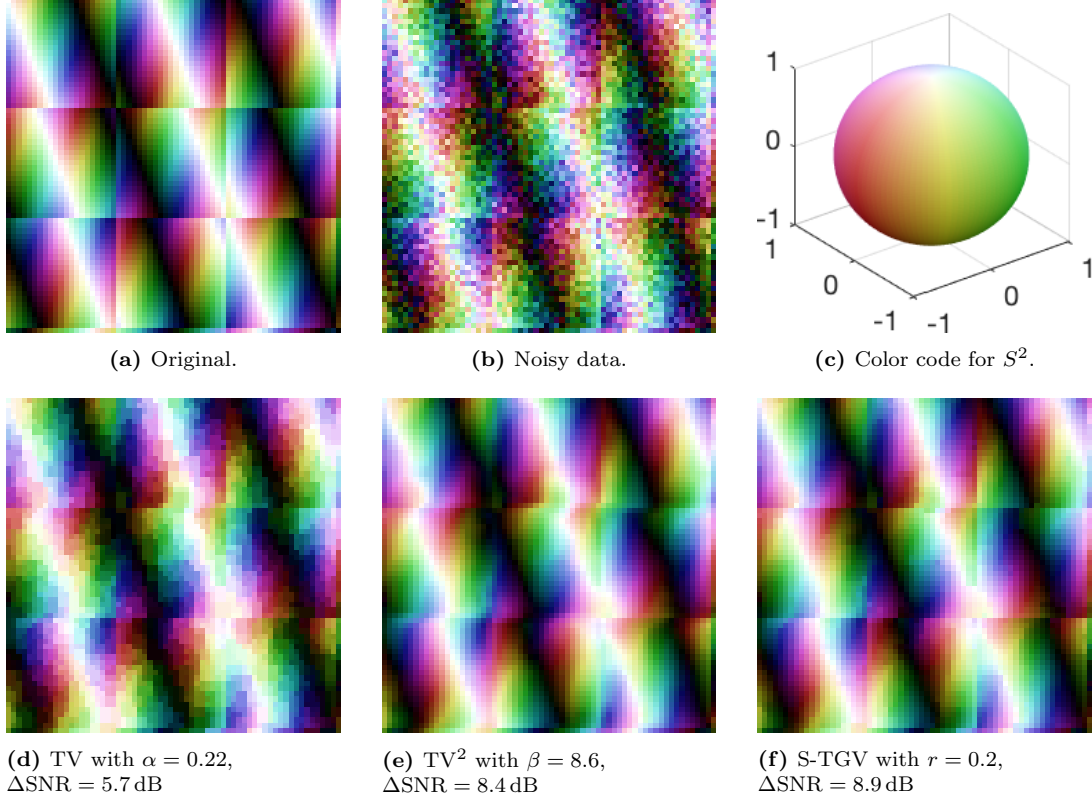
**Figure 8:** Results for univariate  $S^2$  data. *Top:* Ground truth data and reconstruction. *Middle:* Data with intermediate noise and reconstruction. *Bottom:* Data with strong noise and reconstruction. *From left to right:* Data, S-TGV $^2_\alpha$  regularized reconstruction for  $s = 0.05$ ,  $s = 0.95$ ,  $s = 0.3$ , respectively. The color gradation indicates the ordering of the signal.

derivative and is composed of piecewise constant and geodesic signals.

The results of the experiment for different noise levels (no noise, intermediate noise, strong noise), a fixed value of  $r = 1$  and different values of  $s$ , namely  $s \in \{0.05, 0.95, 0.3\}$ , can be found in Figure 8. As can be seen there, the manifold-TGV functional regularizes the data quite well and is able to achieve good results, even in the case with relatively strong noise where it is hard to see any structure in the data. As in the previous experiment, the choice  $s = 0.05$  promotes piecewise constant solutions, which is naturally best for regions where the signal is piecewise constant, but leads to “staircasing” in smooth regions, as can be particularly seen in the geodesic parts of the strong-noise case. The choice  $s = 0.95$  approximates geodesics well, but does not allow for jumps and also produces some oscillations in the smooth parts of the case with strong noise. Again the choice  $s = 0.3$  is a good compromise. Even though it does not reconstruct jumps as well as the TV-like version, it still allows for jumps and reconstructs the geodesic parts rather accurately.

## 5.2 Results for synthetic data

Next we investigate the denoising performance of M-TGV $^2_\alpha$  on synthetic images. We compare the results of the proposed algorithm with the results of manifold TV regularization [93], and



**Figure 9:** Comparison of S-TGV with first and second-order total variation on an  $S^2$ -valued image from [10]. The spherical values are visualized according to the scheme (c) which means that the north pole is white, the south pole is black and the equator gets hue values according to its longitude.

with the result of second-order manifold TV, denoted by  $TV^2$  [10]. The model parameter of first-order TV is denoted by  $\alpha$  and that of  $TV^2$  by  $\beta$ . For all algorithms, 1000 iterations were used.

First we look at  $S^2$ -valued images. As in [10], the noisy data  $f$  were created from the original image  $g$  by  $f_{ij} = \exp_{g_{ij}} \eta_{ij}$  where  $\eta_{ij}$  is a tangent vector at  $g_{ij}$  and both its components are Gaussian distributed with standard deviation  $\sigma = \frac{4}{45}\pi$ . For comparison with first order TV we scanned the parameter range  $\alpha = 0.1, 0.2, \dots, 1$  and the special parameters  $\alpha = 0.22$  and  $\alpha = 3.5 \cdot 10^{-2}$  given in [10] and the corresponding implementation. Similarly, for  $TV^2$  we scanned the parameter range  $\beta = 1, 5, 10, 15, \dots, 30$ , and the special parameters  $\beta = 8.6$  and  $\beta = 29.5$  suggested in [10]. As no beneficial effect of combining first and second-order TV was observed in [10], we used pure TV and  $TV^2$  regularization. For the proposed method, we fixed  $s = 0.3$  and report the best result among the six parameters  $r = 0.05, 0.1, 0.15, \dots, 0.3$ . The results in Figure 9 show that the second-order methods give a significantly smoother result than first order TV and that they do not suffer from staircasing effects. On the flipside,  $TV^2$  regularization results in an undesired smoothing effect of the edges between the tiles, best seen at the bottom left tile. The proposed TGV regularization preserves these edges which results in an improved reconstruction quality.

Next we look at  $Pos_3$ -valued images. Such images appear naturally in diffusion tensor imaging

(DTI), so we briefly describe the setup. DTI is a medical imaging modality based on diffusion weighted magnetic resonance images (DWI) which for example allows to reconstruct fiber tract orientations [11]. A DWI captures the diffusivity of water molecules with respect to a direction  $v \in \mathbb{R}^3$ . The relation between a diffusion tensor  $f(p)$  and the DWIs  $D_v(p)$  at the voxel  $p$  is given by the Stejskal-Tanner equation

$$D_v(p) = A_0 e^{-b v^\top f(p) v} \quad (62)$$

with constants  $b, A_0 > 0$ . A standard noise model for the DWIs is the Rician model which assumes a complex-valued Gaussian noise in the original frequency domain measurements [12]. This means that assuming the model (62) the actual measurement in direction  $v$  at pixel  $p$  is given by  $D'_v(p) = ((X + D_v(p))^2 + Y^2)^{1/2}$ , with the Gaussian random variables  $X, Y \sim N(0, \sigma^2)$ . Building on this model, we impose the noise as follows. From the synthetic tensor image, we compute DWIs according to (62) with respect to 21 different directions. Then we impose Rician noise to all derived DWIs, and we estimate the corresponding diffusion tensor image  $f$  using a least squares approach on (62). Due to the noise, some of the fitted tensors might not be positive definite and thus have no meaningful interpretation. To handle such invalid tensor we exclude them from the data fidelity term. On the other hand, we keep the corresponding pixels in the regularizing term so that we achieve a reasonable inpainting. In Figure 10, we illustrate the effects of the regularization for a synthetic  $\text{Pos}_3$ -valued image. For comparison with (combined) first and second-order TV, we scanned all combinations of the parameters  $\alpha = 0, 0.01, 0.035, 0.1, 0.3125, 0.375$ , and  $\beta = 0.02, 0.05, 0.125, 0.625$ . As before, this comprises the parameters used in [10] and the corresponding implementation for that manifold. For the proposed method, we report the best result among the possible combinations of the parameters  $r = 0.06, 0.08, 0.09, 0.1, 0.12$  and  $s = 0.3, 0.35, 0.4, 0.45$ . In the example we observe that combined  $\text{TV}/\text{TV}^2$  regularization yields a similar reconstruction quality as pure first order TV, whereas S-TGV gives a significantly higher reconstruction quality.

### 5.3 Results for real data

We illustrate the effects of TGV regularization on real manifold-valued signals and images.

First we look at smoothing a time series of wind directions. The natural data space for a signal of orientations is the unit circle  $S^1$ . The present data shows wind directions recorded every hour in 2016 by the weather station SAUF1, St. Augustine, FL.; see Figure 11<sup>2</sup>. We observe that the proposed method smooths the orientations while respecting the phase jump from  $-\pi$  to  $\pi$  and preserving linear trends in the data.

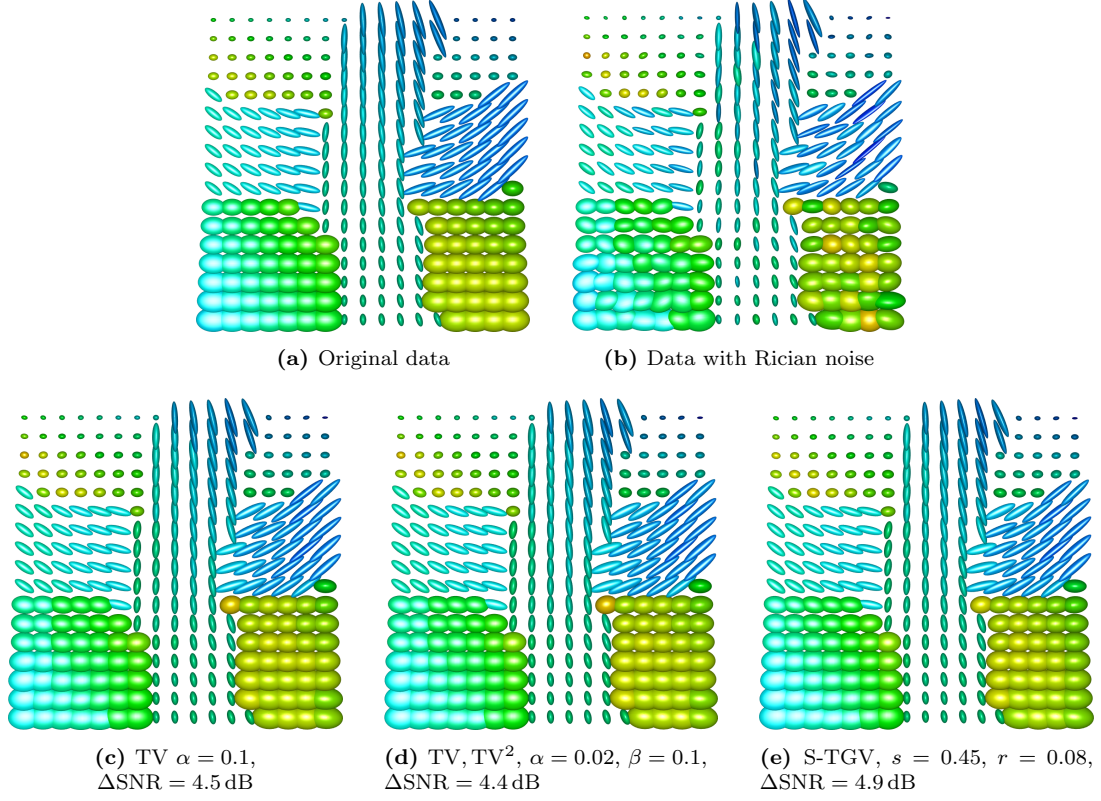
Next we look at smoothing of interferometric synthetic aperture radar (InSAR) images. Synthetic aperture radar (SAR) is a radar technique for sensing the earth's surface from measurements taken by aircrafts or satellites. InSAR images consist of the phase difference between two SAR images, recording a region of interest either from two different angles of view or at two different points in time. Important applications of InSAR are the creation of accurate digital elevation models and the detection of terrain changes; cf. [62, 71]. As InSAR data consists of phase values that are defined modulo  $2\pi$ , their natural data space is the unit circle. In Figure 12, we illustrate the effect of S-TGV to an InSAR image taken from [83]<sup>3</sup>. We observe a clear smoothing effect while sudden phase changes are preserved.

At last we consider real DTI data which was taken from the Camino project [36]<sup>4</sup>. In Figure 13, we see an axial slice of a human brain (slice 20 of the Camino data). We also display a zoom to the corpus callosum which connects the right and the left hemisphere. The

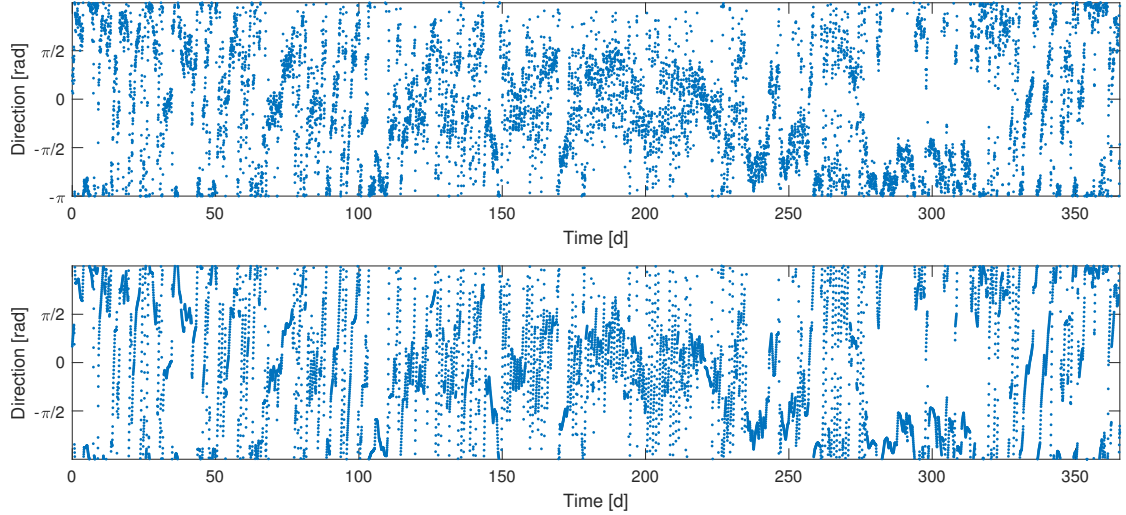
<sup>2</sup>Data available at [http://www.ndbc.noaa.gov/historical\\_data.shtml](http://www.ndbc.noaa.gov/historical_data.shtml).

<sup>3</sup>Data available at <https://earth.esa.int/workshops/ers97/papers/thiel/index-2.html>.

<sup>4</sup>Data available at <http://camino.cs.ucl.ac.uk/>

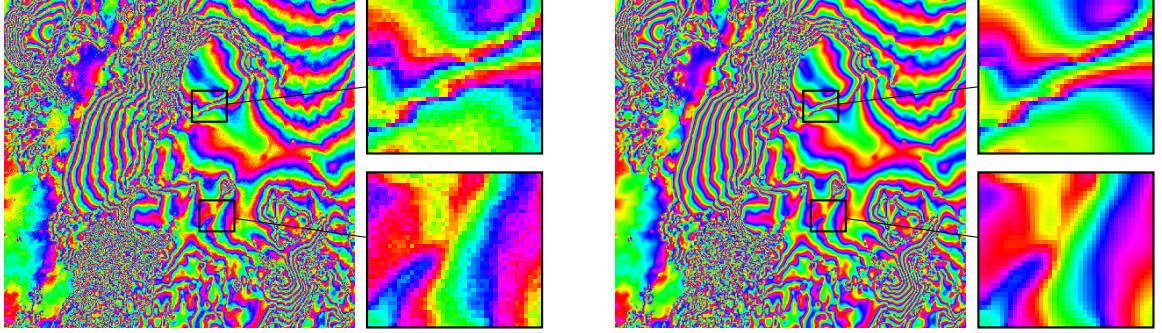


**Figure 10:** Denoising results for a synthetic  $\text{Pos}_3$ -valued image corrupted by Rician noise. Combined TV /  $\text{TV}^2$ -regularization yields a similar reconstruction quality as first order TV, whereas S-TGV gives a significantly higher reconstruction quality.



**Figure 11:** *Top:* Hourly wind directions at weather station SAUF1 (St. Augustine, FL) in the year 2016. *Bottom:* S-TGV ( $s = 0.3$ ,  $r = 1$ ) smooths the signal while not over-smoothing jump discontinuities and properly dealing with the phase jump from  $-\pi$  to  $\pi$  and preserving linear trends in the data.





**Figure 12:** *Left:* InSAR image from [83]. *Right:* Result of S-TGV with  $r = 0.2$ ,  $s = 0.3$ . The image is smoothed and sharp edges are preserved.

original tensors were computed from the diffusion weighted images by a least squares method based on the Stejskal-Tanner equation. We observe that the proposed method smooths the tensors in the corpus callosum while it preserves the sharp edges to the adjacent tissue. Also observe the inpainting of the invalid tensors.

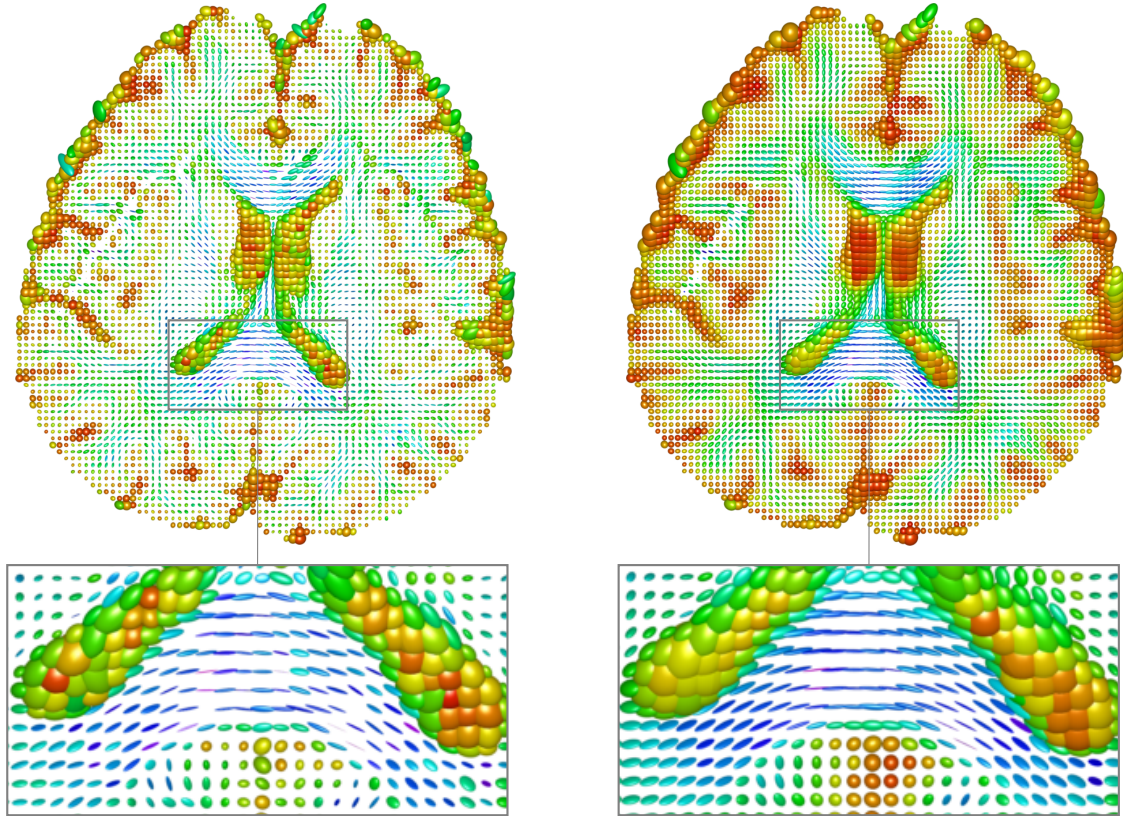
## 6 Conclusion

In this work, we have introduced and explored a notion of second-order total generalized variation (TGV) regularization for manifold-valued data. First, we have derived a variational model for total generalized variation for manifold-valued data. For this purpose, we have used an axiomatic approach. We have first formalized reasonable fundamental properties of vector-valued TGV that should be conserved in the manifold setting. Then we have proposed two realizations which we have shown to fulfill the required axioms. The realization based on parallel transport is rather natural – although not straight forward – from the point of view of differential geometry. The realization based on the Schild’s ladder may be seen as an approximation of the parallel transport variant. It requires less differential geometric concepts and it is easier to realize numerically while yielding comparable numerical results as shown in the experimental section. Existence results for  $M\text{-TGV}_\alpha^2$ -based denoising have been obtained for the proposed variants. Next, we have derived an algorithm for the proposed model. To this end we built on the well-established concept of a cyclic proximal algorithm. As main contribution, we have performed the challenging task to derive all quantities necessary to compute the proximal mappings of the involved atoms. Finally, we have conducted a numerical study of the proposed scheme. The experiments revealed that  $M\text{-TGV}$  regularization reliably removes noise while it preserves edges and smooth trends. A quantitative comparison on images with groundtruth indicates that TGV regularization improves upon first and second-order TV for manifold-valued data.

An interesting topic of future research are improved numerics such as earlier stopping criteria.

## Acknowledgment

Kristian Bredies and Martin Holler acknowledge support by the Austrian Science Fund (FWF) (Grant P 29192). Martin Storath acknowledges support by the German Research Foundation DFG (Grant STO1126/2-1). Andreas Weinmann acknowledges support by the German Research Foundation DFG (Grants WE5886/4-1, WE5886/3-1).



**Figure 13:** *Left:* Diffusion tensor image of a human brain (axial slice). The magnified image shows the corpus callosum. Note the missing (invalid) tensors at several voxels. *Right:* Result of S-TGV regularization with  $r = 0.1$ ,  $s = 0.3$ . The streamlines are smoothed whereas sharp transitions are preserved. Invalid voxels are reasonably inpainted.

## References

- [1] P.-A. Absil, R. Mahony, and R. Sepulchre. Riemannian geometry of Grassmann manifolds with a view on algorithmic computation. *Acta Applicandae Mathematicae*, 80(2):199–220, 2004.
- [2] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.
- [3] Y. Adato, T. Zickler, and O. Ben-Shahar. A polar representation of motion and implications for optical flow. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1145–1152, 2011.
- [4] B. Afsari. Riemannian  $L^p$  center of mass: existence, uniqueness, and convexity. *Proceedings of the American Mathematical Society*, 139(2):655–673, 2011.
- [5] 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.
- [6] D. Azagra and J. Ferrera. Proximal calculus on Riemannian manifolds. *Mediterranean Journal of Mathematics*, 2:437–450, 2005.
- [7] M. Bačák. The proximal point algorithm in metric spaces. *Israel Journal of Mathematics*, 194(2):689–701, 2013.

- [8] M. Bačák. Computing medians and means in Hadamard spaces. *SIAM Journal on Optimization*, 24(3):1542–1566, 2014.
- [9] M. Bačák. *Convex analysis and optimization in Hadamard spaces*. de Gruyter, 2014.
- [10] M. Bačák, R. Bergmann, G. Steidl, and A. Weinmann. A second order non-smooth variational model for restoring manifold-valued images. *SIAM Journal on Scientific Computing*, 38(1):A567–A597, 2016.
- [11] P. Basser, J. Mattiello, and D. LeBihan. MR diffusion tensor spectroscopy and imaging. *Biophysical Journal*, 66(1):259–267, 1994.
- [12] S. Basu, T. Fletcher, and R. Whitaker. Rician noise removal in diffusion tensor MRI. In *Medical Image Computing and Computer-Assisted Intervention 2006*, pages 117–125. Springer, 2006.
- [13] 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.
- [14] M. Baust, A. Weinmann, M. Wiecek, T. Lasser, M. Storath, and N. Navab. Combined tensor fitting and TV regularization in diffusion tensor imaging based on a Riemannian manifold approach. *IEEE Transactions on Medical Imaging*, 35(8):1972–1989, 2016.
- [15] R. Bergmann, R. H. Chan, R. Hielscher, J. Persch, and G. Steidl. Restoration of manifold-valued images by half-quadratic minimization. *Inverse Problems and Imaging*, 10:281–304, 2016.
- [16] R. Bergmann, J. Fitschen, J. Persch, and G. Steidl. Priors with coupled first and second order differences for manifold-valued image processing. *arXiv preprint arXiv:1709.01343*, 2017.
- [17] R. Bergmann, J. H. Fitschen, J. Persch, and G. Steidl. Infimal convolution type coupling of first and second order differences on manifold-valued images. *Scale Space and Variational Methods in Computer Vision 2017*, pages 447–459, 2017.
- [18] 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.
- [19] R. Bergmann, J. Persch, and G. Steidl. A parallel Douglas-Rachford algorithm for restoring images with values in symmetric hadamard manifolds. *SIAM Journal on Imaging Sciences*, 9:901–937, 2016.
- [20] R. Bergmann and A. Weinmann. A second-order TV-type approach for inpainting and denoising higher dimensional combined cyclic and vector space data. *Journal of Mathematical Imaging and Vision*, 55(3):401–427, 2016.
- [21] M. Bergounioux. Mathematical analysis of a inf-convolution model for image processing. *J. Optim. Theory Appl.*, pages 1–21, 2015.
- [22] D. Bertsekas. Incremental proximal methods for large scale convex optimization. *Mathematical Programming*, 129:163–195, 2011.
- [23] 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.
- [24] 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.
- [25] K. Bredies and M. Holler. Regularization of linear inverse problems with total generalized variation. *Journal of Inverse and Ill-posed Problems*, 22(6):871–913, 2014.
- [26] K. Bredies and M. Holler. A TGV-based framework for variational image decompression, zooming and reconstruction. Part I: Analytics. *SIAM J. Imaging Sci.*, 8(4):2814–2850, 2015.
- [27] K. Bredies and M. Holler. A TGV-based framework for variational image decompression, zooming and reconstruction. Part II: Numerics. *SIAM J. Imaging Sci.*, 8(4):2851–2886, 2015.
- [28] K. Bredies, M. Holler, M. Storath, and A. Weinmann. Total generalized variation for manifold-valued data. *arXiv preprint arXiv:1709.01616*, 2017.

- [29] K. Bredies, K. Kunisch, and T. Pock. Total generalized variation. *SIAM Journal on Imaging Sciences*, 3(3):492–526, 2010.
- [30] K. Bredies and T. Valkonen. Inverse problems with second-order total generalized variation constraints. In *Proceedings of SampTA 2011 - 9th International Conference on Sampling Theory and Applications*, Singapore, 2011.
- [31] A. Chambolle and P.-L. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76(2):167–188, 1997.
- [32] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40:120–145, 2011.
- [33] 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.
- [34] J. Cheeger and D. Ebin. *Comparison theorems in Riemannian geometry*, volume 9. North-Holland, 1975.
- [35] C. Ched’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.
- [36] P. Cook, Y. Bai, S. Nedjati-Gilani, K. Seunarine, M. Hall, G. Parker, and D. Alexander. Camino: Open-source diffusion-MRI reconstruction and processing. In *14th Scientific Meeting of the International Society for Magnetic Resonance in Medicine*, page 2759, 2006.
- [37] D. Cremers and E. Strelakovsky. Total cyclic variation and generalizations. *Journal of Mathematical Imaging and Vision*, 47(3):258–277, 2013.
- [38] M. do Carmo. *Riemannian geometry*. Birkhauser, 1992.
- [39] J. Ehlers, F. A. E. Pirani, and A. Schild. The geometry of free fall and light propagation. In O’Raifeartaigh, editor, *General Relativity*. Oxford University Press, 1972.
- [40] J. Eschenburg. Lecture notes on symmetric spaces. *preprint*, 1997.
- [41] O. Ferreira and P. Oliveira. Proximal point algorithm on Riemannian manifolds. *Optimization*, 51:257–270, 2002.
- [42] P. Fletcher and S. Joshi. Riemannian geometry for the statistical analysis of diffusion tensor data. *Signal Processing*, 87(2):250–262, 2007.
- [43] 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.
- [44] 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.
- [45] 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.
- [46] 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.
- [47] 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.
- [48] 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.
- [49] P. Grohs and S. Hosseini.  $\varepsilon$ -subgradient algorithms for locally Lipschitz functions on Riemannian manifolds. *Advances in Computational Mathematics*, 42(2):333–360, 2016.
- [50] P. Grohs and M. Sprecher. Total variation regularization on riemannian manifolds by iteratively reweighted minimization. *Information and Inference*, 5(4):353–378, 2016.



- [51] P. Grohs and J. Wallner. Interpolatory wavelets for manifold-valued data. *Applied and Computational Harmonic Analysis*, 27(3):325–333, 2009.
- [52] S. Hawe, M. Seibert, and M. Kleinsteuber. Separable dictionary learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 438–445, 2013.
- [53] J. Itoh and M. Tanaka. The dimension of a cut locus on a smooth Riemannian manifold. *Tohoku Mathematical Journal, Second Series*, 50(4):571–575, 1998.
- [54] H. Johansen-Berg and T. Behrens. *Diffusion MRI: From quantitative measurement to in-vivo neuroanatomy*. Academic Press, London, 2009.
- [55] A. Kheyfets, W. A. Miller, and G. A. Newton. Schild’s ladder parallel transport procedure for an arbitrary connection. *Internat. J. Theoret. Phys.*, 39(12):2891–2898, 2000.
- [56] 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.
- [57] F. Knoll, K. Bredies, T. Pock, and R. Stollberger. Second order total generalized variation (TGV) for MRI. *Magn. Reson. Med.*, 65(2):480–491, 2011.
- [58] F. Knoll, M. Holler, T. Koesters, R. Otazo, K. Bredies, and D. Sodickson. Joint MR-PET reconstruction using a multi-channel image regularizer. *IEEE Trans. Med. Imag.*, 2016. In press.
- [59] S. Kobayashi and K. Nomizu. *Foundations of differential geometry*, volume 2. Wiley New York, 1969.
- [60] R. Lai and S. Osher. A splitting method for orthogonality constrained problems. *Journal of Scientific Computing*, 58(2):431–449, 2014.
- [61] 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.
- [62] 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.
- [63] 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.
- [64] 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.
- [65] S. Niu, Y. Gao, Z. Bian, J. Huang, W. Chen, G. Yu, Z. Liang, and J. Ma. Sparse-view x-ray ct reconstruction via total generalized variation regularization. *Physics in medicine and biology*, 59(12):2997, 2014.
- [66] J. Oller and J. Corcuera. Intrinsic analysis of statistical estimation. *Annals of Statistics*, 23(5):1562–1581, 1995.
- [67] K. Papafitsoros and T. Valkonen. Asymptotic behaviour of total generalised variation. In *5th International Conference on Scale Space and Variational Methods in Computer Vision.*, pages 702–714. Springer, 2015.
- [68] X. Pennec. Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127–154, 2006.
- [69] X. Pennec, P. Fillard, and N. Ayache. A Riemannian framework for tensor computing. *International Journal of Computer Vision*, 66(1):41–66, 2006.
- [70] C. Rao. Information and accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, 37(3):81–91, 1945.
- [71] F. Rocca, C. Prati, and A. Ferretti. An overview of SAR interferometry. In *Proceedings of the 3rd ERS Symposium on Space at the Service of our Environment, Florence*, 1997.

- [72] 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.
- [73] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.
- [74] H. Schultz. A circular median filter approach for resolving directional ambiguities in wind fields retrieved from spaceborne scatterometer data. *Journal of Geophysical Research: Oceans*, 95(C4):5291–5303, 1990.
- [75] S. Setzer, G. Steidl, and T. Teuber. Infimal convolution regularizations with discrete l1-type functionals. *Comm. Math. Sci.*, 9(3):797–872, 2011.
- [76] M. Spivak. *Differential Geometry*, volume 1-5. Publish or Perish, Berkeley, 1975.
- [77] S. Sra and R. Hosseini. Conic geometric optimization on the manifold of positive definite matrices. *SIAM Journal on Optimization*, 25(1):713–739, 2015.
- [78] A. Stefanoiu, A. Weinmann, M. Storath, N. Navab, and M. Baust. Joint segmentation and shape regularization with a generalized forward-backward algorithm. *IEEE Transactions on Image Processing*, 25(7):3384–3394, 2016.
- [79] M. Storath and A. Weinmann. Fast median filtering for phase or orientation data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, to appear.
- [80] 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.
- [81] M. Storath, A. Weinmann, and M. Unser. Jump-penalized least absolute values estimation of scalar or circle-valued signals. *Information and Inference*, 2017.
- [82] E. Strelakovsky 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.
- [83] K. Thiel, X. Wu, and P. Hartl. ERS-tandem-interferometric observation of volcanic activities in Iceland. *ESA SP*, pages 475–480, 1997.
- [84] I. Todhunter. *Spherical trigonometry, for the use of colleges and schools: with numerous examples*. Macmillan, 1863.
- [85] 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.
- [86] M. Unser and P. Tafti. *An introduction to sparse stochastic processes*. Cambridge University Press, 2014.
- [87] 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.
- [88] T. Valkonen, K. Bredies, and F. Knoll. Total generalized variation in diffusion tensor imaging. *SIAM Journal on Imaging Sciences*, 6(1):487–525, 2013.
- [89] 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.
- [90] 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.
- [91] A. Weinmann. Nonlinear subdivision schemes on irregular meshes. *Constructive Approximation*, 31(3):395–415, 2010.
- [92] A. Weinmann. Interpolatory multiscale representation for functions between manifolds. *SIAM Journal on Mathematical Analysis*, 44:162–191, 2012.

- [93] A. Weinmann, L. Demaret, and M. Storath. Total variation regularization for manifold-valued data. *SIAM Journal on Imaging Sciences*, 7(4):2226–2257, 2014.
- [94] 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.

## A Appendix

### A.1 Proofs for Section 2

*Proof of Proposition 2.17.* Given that  $D$  satisfies the assumptions of Proposition 2.8 and the particular form of  $D^{\text{sym}}$  in the vector-space case, it is immediate that  $\text{M-TGV}_\alpha^2$  reduces to vector-space TGV, hence **(M-P1’)** holds.

Now in case the minimum in (15) is attained at  $y_{i,j}^1 = y_{i,j}^2 = u_{i,j}$  we get, since

$$D^{\text{sym}}([u_{i,j}, u_{i,j}], [u_{i,j}, u_{i,j}], [u_{i,j-1}, u_{i,j-1}], [u_{i-1,j}, u_{i-1,j}]) = 0$$

for all  $i, j$ , that  $\text{M-TGV}_\alpha^2(u) = \alpha_1 \sum_{i,j} \left( d(u_{i+1,j}, u_{i,j})^p + d(u_{i,j+1}, u_{i,j})^p \right)^{1/p} = \alpha_1 \text{TV}(u)$ .

On the other hand, in case the minimum is attained at  $y_{i,j}^1 = u_{i+1,j}$  and  $y_{i,j}^2 = u_{i,j+1}$  we get, in the vector space situation, that

$$\begin{aligned} \text{M-TGV}_\alpha^2(u) &= \alpha_0 \sum_{i,j} \left( |u_{i+1,j} - 2u_{i,j} + u_{i-1,j}|^p + |u_{i,j+1} - 2u_{i,j} + u_{i,j-1}|^p \right. \\ &\quad \left. + 2^{1-p} |(u_{i+1,j} - u_{i,j}) - (u_{i+1,j-1} - u_{i,j-1}) + (u_{i,j+1} - u_{i,j}) - (u_{i-1,j+1} - u_{i-1,j})|^p \right)^{1/p} \end{aligned}$$

which coincides with  $\alpha_0 \text{TV}^2$  and  $\text{TV}^2$  as in Definition 2.4.

Now suppose that  $(u_{i,j})_{i,j}$  is locally geodesic. Choosing  $y_{i,j}^1 = u_{i+1,j}$  and  $y_{i,j}^2 = u_{i,j+1}$  for all  $i, j$  we get, as in the proof of Theorem 2.9, that  $\text{M-TGV}_\alpha^2$  is bounded above by

$$\alpha_0 \sum_{i,j} D^{\text{sym}}([u_{i,j}, u_{i+1,j}], [u_{i,j}, u_{i,j+1}], [u_{i,j-1}, u_{i+1,j-1}], [u_{i-1,j}, u_{i-1,j+1}]),$$

which is zero by assumption.

Now conversely, assume that  $\text{M-TGV}_\alpha^2(u) = 0$  and any two points in  $u_{i,j}, u_{i',j'}$  with  $\max\{|i - i'|, |j - j'|\} \leq 2$  are connected by unique geodesics. As in the proof of Proposition 2.8 this implies that both  $(u_{i,j})_i$  and  $(u_{i,j})_j$  are locally geodesic as univariate signals. Also we get that

$$0 = \sum_{i,j} D^{\text{sym}}([u_{i,j}, u_{i+1,j}], [u_{i,j}, u_{i,j+1}], [u_{i,j-1}, u_{i+1,j-1}], [u_{i-1,j}, u_{i-1,j+1}]),$$

which implies by assumption, that  $(u_{i+k,j-k})_k$  is locally geodesic. Hence, by definition,  $(u_{i,j})_{i,j}$  is locally geodesic as bivariate signal. Finally, by Lemma 2.16,  $u$  is geodesic.  $\square$

*Proof of Theorem 2.18.* It suffices to verify the assumptions of Proposition 2.17. It is easy to see that, since  $D_S([x, x], [u, u]) = 0$ , also  $D_S^{\text{sym}}([x, x], [x, x], [u, u], [u, u]) = 0$ . Further, in the vector space setting, we get that  $c^1 = \frac{u_{-,o} + y_{o,o}^1}{2}$  and  $c^2 = \frac{u_{o,-} + y_{o,o}^2}{2}$  as well as  $r^1 = u_{-,o} + y_{o,o}^1 - u_{o,o}$  and  $r^2 = u_{o,-} + y_{o,o}^2 - u_{o,o}$ . Consequently,

$$\begin{aligned} D_S([r^1, y_{-,o}^2], [y_{o,-}^1, r^2]) &= |y_{-,o}^2 - (u_{o,-} + y_{o,o}^1 - u_{o,o}) - (u_{o,-} + y_{o,o}^2 - u_{o,o} - y_{o,-}^1)| \\ &= |y_{o,o}^1 - u_{o,o} - (y_{o,-}^1 - u_{o,-}) + y_{o,o}^2 - u_{o,o} - (y_{-,o}^2 - u_{-,o})| \end{aligned}$$

and, using Proposition 2.17, **(M-P1')** to **(M-P3')** follow.

Now suppose that  $(u_{i,j})_{i,j}$  is a geodesic three-by-three signal. Hence, with the notation as in the definition of  $D_S^{\text{sym}}$ ,  $c^1 = u_{i,j} \in [u_{i-1,j}, u_{i+1,j}]_{\frac{1}{2}}$  and we can choose  $r^1 = u_{i,j} \in [u_{i,j}, u_{i,j}]_2$ . Similar,  $c^2 = u_{i,j} \in [u_{i,j-1}, u_{i,j+1}]_{\frac{1}{2}}$  and we can again choose  $r^2 = u_{i,j} \in [u_{i,j}, u_{i,j}]_2$ . Consequently,

$$D_S^{\text{sym}}([u_{i,j}, u_{i+1,j}], [u_{i,j}, u_{i,j+1}], [u_{i,j-1}, u_{i+1,j-1}], [u_{i-1,j}, u_{i-1,j+1}]) \leq D_S([u_{i,j}, u_{i-1,j+1}], [u_{i+1,j-1}, u_{i,j}]).$$

But, as shown in the proof of Theorem 2.9, the right-hand-side vanishes since  $(u_{i-k,j+k})_k$  is geodesic as univariate signal.

Now conversely, assume that  $u = (u_{i,j})_{i,j}$  is a three-by-three signal such that both  $(u_{i,j})_i$  and  $(u_{i,j})_j$  are locally geodesic and that

$$0 = D_S^{\text{sym}}([u_{i,j}, u_{i+1,j}], [u_{i,j}, u_{i,j+1}], [u_{i,j-1}, u_{i+1,j-1}], [u_{i-1,j}, u_{i-1,j+1}]).$$

By the assumption on uniqueness of geodesics we get, again with the notation as in the definition of  $D_S^{\text{sym}}$ , that  $c^1 = u_{i,j}$  and hence  $r^1 = u_{i,j}$  and also that  $c^2 = u_{i,j}$  and hence  $r^2 = u_{i,j}$ . Consequently,  $D_S^{\text{sym}}(\dots)$  reduces to  $D_S([u_{i,j}, u_{i-1,j+1}], [u_{i+1,j-1}, u_{i,j}])$ . But the latter being zero implies that  $u_{i-1,j+1} \in [u_{i+1,j-1}, u_{i,j}]_2$  and, consequently,  $u$  is geodesic.

Hence also the remaining assumptions of Proposition 2.17 are satisfied and the assertion follows.  $\square$

The following well-known fact on the parallel transport in manifolds will be required in the proof of Theorem 2.19. For the sake of completeness, we provide a short proof.

**Lemma A.1.** *Let  $u_-$ ,  $u_o$  and  $u_+$  be three points in a manifold such that there exists a pairwise distance minimizing geodesic  $\gamma : [0, 1] \rightarrow \mathcal{M}$  such that  $\gamma(0) = u_-$ ,  $\gamma(1/2) = u_o$ ,  $\gamma(1) = u_+$ . Then  $\log_{u_-}^\gamma(u_o) = \text{pt}_{u_-}^\gamma(\log_{u_o}^\gamma(u_+))$ .*

*Proof.* For  $t \in [0, 1]$ , set  $w(t) = (1/2) \frac{d}{dt} \gamma(\frac{t}{2})$ . Then  $w$  satisfies  $\frac{D}{dt} w(t) = 0$  on  $[0, 1]$ ,  $w(0) = \log_{u_-}^\gamma(u_o)$ . By definition of the parallel transport,  $\text{pt}_{u_o}^\gamma(\log_{u_-}^\gamma(u_o)) = w(1) = (1/2) \frac{d}{dt} \gamma(1/2) = \log_{u_o}^\gamma(u_+)$ . Applying  $\text{pt}_{u_-}^\gamma$  on both sides, the result follows.

*Proof of Theorem 2.19.* It suffices to verify the assumptions of Proposition 2.17. Since the parallel transport is isometric, reduces to the identity if starting- and endpoint coincide, and since  $|\log_x(x)|_x = 0$  for all  $x \in \mathcal{M}$ , it is easy to see that  $D_{\text{pt}}^{\text{sym}}([x, x], [x, x], [u, u], [u, u]) = 0$ . In the vector space setting, with the notation as in the definition of  $D_{\text{pt}}^{\text{sym}}$ , we get that  $w^1 = y_{o,o}^1 - u_{o,o}$  and  $r^1 = u_{-,o} + y_{o,o}^1 - u_{o,o}$  as well as  $w^2 = y_{o,o}^2 - u_{o,o}$  and  $r^2 = u_{o,-} + y_{o,o}^2 - u_{o,o}$ . Consequently, from the properties of  $D_{\text{pt}}$  it follows that

$$\begin{aligned} D_{\text{pt}}^{\text{sym}}([u_{o,o}, y_{o,o}^1], [u_{o,o}, y_{o,o}^2], [u_{o,-}, y_{o,-}^1], [u_{-,o}, y_{-,o}^2]) &= \\ &= D_{\text{pt}}([u_{-,o} + y_{o,o}^1 - u_{o,o}, y_{o,o}^2], [y_{o,-}^1, u_{o,-} + y_{o,o}^2 - u_{o,o}]) \\ &= |y_{o,-}^2 - (u_{-,o} + y_{o,o}^1 - u_{o,o}) - (u_{o,-} + y_{o,o}^2 - u_{o,o} - y_{o,-}^1)| \\ &= |y_{o,o}^1 - u_{o,o} - (y_{o,-}^1 - u_{o,-}) + y_{o,o}^2 - u_{o,o} - (y_{o,-}^2 - u_{-,o})| \end{aligned}$$

and from Proposition 2.17 it follows that **(M-P1')** to **(M-P3')** holds.

Now let  $u = (u_{i,j})_{i,j}$  be a three-by-three geodesic signal and choose  $y_{i,j}^1 = u_{i+1,j}$  and  $y_{i,j}^2 = u_{i,j+1}$  for each  $(i, j)$ . Denoting by  $\gamma^1$  and  $\gamma^2$  distance minimizing geodesics such that  $\gamma^1(0) = u_{i-1,j}$ ,  $\gamma^1(1/2) = u_{i,j}$ ,  $\gamma^1(1) = u_{i+1,j}$  and  $\gamma^2(0) = u_{i,j-1}$ ,  $\gamma^2(1/2) = u_{i,j}$ ,  $\gamma^2(1) =$

$u_{i,j-1}$ , respectively, Lemma A.1 implies that  $\log_{u_{i-1,j}}^{\gamma^1}(u_{i,j}) = \text{pt}_{u_{i-1,j}}(\log_{u_{i,j}}^{\gamma^1}(u_{i+1,j}))$  as well as  $\log_{u_{i,j-1}}^{\gamma^2}(u_{i,j}) = \text{pt}_{u_{i,j-1}}(\log_{u_{i,j}}^{\gamma^2}(u_{i,j+1}))$ . Hence we can choose  $r^1 = r^2 = u_{i,j}$  and

$$\begin{aligned} D_{\text{pt}}^{\text{sym}}([u_{i,j}, u_{i+1,j}], [u_{i,j}, u_{i,j+1}], [u_{i,j-1}, u_{i+1,j-1}], [u_{i-1,j}, u_{i-1,j+1}]) \\ \leq D_{\text{pt}}([u_{i,j}, u_{i-1,j+1}], [u_{i+1,j-1}, u_{i,j}]) = 0, \end{aligned}$$

where the last term is zero since  $(u_{i-k,j+k})_k$  is locally geodesic as univariate signal.

Now conversely, assume that  $u = (u_{i,j})_{i,j}$  is a three-by-three signal such that the geodesics connecting each pair of points are unique and such that  $(u_{i,j})_i$  and  $(u_{i,j})_j$  are locally geodesic. By Lemma A.1 and uniqueness of geodesics we get  $\log_{u_{i-1,j}}(u_{i,j}) = \text{pt}_{u_{i-1,j}}(\log_{u_{i,j}}(u_{i+1,j}))$  as well as  $\log_{u_{i,j-1}}(u_{i,j}) = \text{pt}_{u_{i,j-1}}(\log_{u_{i,j}}(u_{i,j+1}))$ . Again by uniqueness of geodesics, with the notation as in the definition of  $D_{\text{pt}}^{\text{sym}}$ , we conclude that  $r^1 = r^2 = u_{i,j}$  hence,

$$\begin{aligned} 0 = D_{\text{pt}}^{\text{sym}}([u_{i,j}, u_{i+1,j}], [u_{i,j}, u_{i,j+1}], [u_{i,j-1}, u_{i+1,j-1}], [u_{i-1,j}, u_{i-1,j+1}]) \\ = D_{\text{pt}}([u_{i,j}, u_{i-1,j+1}], [u_{i+1,j-1}, u_{i,j}]). \end{aligned}$$

From the properties of  $D_{\text{pt}}$  it hence follows that the points  $u_{i+1,j-1}$ ,  $u_{i,j}$ ,  $u_{i-1,j+1}$  are on a joint, length minimizing geodesic at equal distance. Hence  $u$  is geodesic. This ensures that all assumptions of Proposition 2.17 are fulfilled and hence the assertion follows.  $\square$

## A.2 Proofs for Section 3

*Proof of Proposition 3.2.* Take  $(y^n)_n$  such that  $y^n \in C(x)$  and  $S(x) = \lim_n G(x, y^n)$ . By assumption  $(y^n)_n$  admits a subsequence  $(y^{n_k})_k$  converging to some  $y \in C(x)$ . By lower semi-continuity of  $G$  it follows that  $S(x) = G(x, y)$ .

For lower semi-continuity of  $S$  now take  $(x^n)_n$  in  $\mathcal{M}^N$  converging to some  $x \in \mathcal{M}^N$  for which, without loss of generality, we assume that  $\liminf_n S(x^n) = \lim_n S(x^n)$ . Pick  $y^n \in C(x^n)$  such that  $S(x^n) = G(x^n, y^n)$ . By assumption and uniqueness of limits we can obtain a subsequence  $(x^{n_k})_k$  and  $(y^{n_k})_k$  converging to  $x$  and  $y$ , respectively, such that  $y \in C(x)$ . We conclude that

$$S(x) \leq G(x, y) \leq \liminf_k G(x^{n_k}, y^{n_k}) = \liminf_k S(x^{n_k}) = \liminf_n S(x^n)$$

and the assertion follows.  $\square$

*Proof of Lemma 3.3.* The proof relies on the following fact: for any bounded subset  $\mathcal{N}$  of  $\mathcal{M}$ , there is a constant  $D$  such that, if the length of any geodesic  $\psi : [r, s] \rightarrow \mathcal{M}$  with  $\psi([r, s]) \subset \mathcal{N}$  is smaller than  $D$ , then  $\psi$  is a unique distance minimizing geodesic between  $\psi(r)$  and  $\psi(s)$ , and the Jacobi fields have no zero along this geodesic  $\psi$ ; see for instance [34, Chapter 1.6/1.7]. In addition, there is a constant  $L$  which is uniform for all such geodesics such that, with  $f : [r, s] \times [0, 1] \rightarrow \mathcal{M}$  a geodesic variation of  $\psi$  with  $f(\cdot, 0) = \psi$ , we have

$$d(\psi(t), f(t, \tau)) \leq L \max\{d(\psi(r), f(r, \tau)), d(\psi(s), f(s, \tau))\}$$

for all  $\tau \in [0, 1]$  and all  $t \in [r, s]$ .

In order to obtain uniform convergence of  $(\gamma^n)_n$ , our approach is now to use compactness arguments and subdivide the curves into small segments where these assertions hold.

At first we define the bounded set  $\mathcal{N}$ . To this aim, define  $B = \{(p^n, q^n) \mid n \in \mathbb{N}\} \cup \{(p, q)\}$ ,  $K = \sup\{d(p', q') \mid (p', q') \in B\} < \infty$  and  $\mathcal{N}$  to be the union of all images of shortest geodesics connecting two points  $p'$  and  $q'$  with  $(p', q') \in B$ . Then, for any  $z \in \mathcal{N}$  there is a shortest

geodesic  $\psi : [r, s] \rightarrow \mathcal{M}$  such that  $(\psi(r), \psi(s)) \in B$  and  $z \in \psi([r, s])$ , and we get that  $d(p, z) \leq d(p, \psi(r)) + d(\psi(r), z) \leq d(p, \psi(r)) + d(\psi(r), \psi(s)) \leq \sup_{n \in \mathbb{N}} d(p, p^n) + K$ , hence  $\mathcal{N}$  is bounded.

Further, by construction,  $\gamma^n([0, 1]) \subset \mathcal{N}$  for any  $n$ . With this choice of  $\mathcal{N}$ , we now choose the constant  $D$  as above and we choose  $l \in \mathbb{N}$  large enough such that  $l \geq 2 \frac{d(p, q)}{D}$  and subdivide the interval  $[0, 1]$  into the  $l+1$  points  $0, \frac{1}{l}, \frac{2}{l}, \dots, 1$ . Then, by compactness, we can find a subsequence  $(\gamma^{n_k})_k$  such that  $\gamma^{n_k}(t)$  converges for all  $t \in [0, \frac{1}{l}, \dots, 1]$ . We set  $z^t := \lim_{k \rightarrow \infty} \gamma^{n_k}(t)$ . Then

$$d(z^t, z^{t+1/l}) = \lim_{k \rightarrow \infty} d(\gamma^{n_k}(t), \gamma^{n_k}(t + 1/l)) \leq \frac{1}{l} \lim_{k \rightarrow \infty} d(p^{n_k}, q^{n_k}) = \frac{d(p, q)}{l} < D.$$

Hence, for each  $l$ , there is a unique shortest geodesic connecting  $z^t$  and  $z^{t+1/l}$  and we define the curve  $\gamma : [0, 1] \rightarrow \mathcal{M}$  as the concatenation of these geodesics, parametrized proportionally. Then  $\gamma(0) = p$ ,  $\gamma(1) = q$  and the length of  $\gamma$  is given as

$$\sum_{t \in \{0, \frac{1}{l}, \dots, 1 - \frac{1}{l}\}} d(z^t, z^{t+1/l}) \leq \frac{d(p, q)}{l} l = d(p, q).$$

Hence the length of  $\gamma$  is less or equal to the distance of its two endpoints which implies that  $\gamma$  is a shortest geodesic connecting  $p$  and  $q$ ; see for instance [38]. Defining  $\psi^t = \gamma|_{[z^t, z^{t+1/l}]}$  we get that the image of  $\psi^t$  is in  $\mathcal{N}$  and its length is less than  $D$ . Defining the geodesic variation  $f^{t,k} : [t, t + 1/l] \times [0, 1]$  as

$$f^{t,k}(\eta, \tau) = [[z^t, \gamma^{n_k}(t)]_\tau, [z^{t+1/l}, \gamma^{n_k}(t + \frac{1}{l})]_\tau]_\theta,$$

with  $\theta$  chosen such that  $(1 - \theta)t + \theta(t + \frac{1}{l}) = \eta$ , we get, for sufficiently large  $k$  (such that the brackets are single-valued), that

$$d(\gamma(\eta), \gamma^{n_k}(\eta)) = d(\psi^t(\eta), f^{t,k}(\eta, 1)) \leq L \max\{d(z^t, \gamma^{n_k}(t)), d(z^{t+1/l}, \gamma^{n_k}(t + \frac{1}{l}))\}.$$

Hence  $\gamma^{n_k} \rightarrow \gamma$  uniformly on  $[0, 1]$ .

Now consider an arbitrary interval  $[a, b] \supset [0, 1]$ . First we uniquely extend the geodesics  $(\gamma^{n_k})_k$ ,  $\gamma$  to this interval. Then,  $\tilde{\mathcal{N}} := \{\gamma^{n_k}([a, b]) \mid k \in \mathbb{N}\} \cup \gamma([a, b])$  is a bounded set and we can again pick a constant  $\tilde{D}$  such that any geodesic in  $\tilde{\mathcal{N}}$  with length smaller than  $\tilde{D}$  is a unique length minimizing geodesic between its start- and endpoint. Now since geodesics have constant speed, we note that, for any  $a', b'$  such that  $[a', b'] \subset [a, b]$ , the length of  $\gamma^{n_k}|_{[a', b']}$  equals  $|b' - a'|$  times the length of  $\gamma^{n_k}|_{[0, 1]}$  which in turn is equal to  $|b' - a'|d(p^{n_k}, q^{n_k})$ . But the latter is uniformly bounded by convergence of  $(p^{n_k})_k$ ,  $(q^{n_k})_k$ , and hence we can pick a uniform  $\epsilon > 0$  such that the length of  $\gamma^{n_k}|_{[a', b']}$  (and of  $\gamma|_{[a', b]}$ ) is less than  $\tilde{D}$  whenever  $|b' - a'| \leq 2\epsilon$ .

Our approach is now to show that, whenever  $\gamma^{n_k}$  converges uniformly to  $\gamma$  on an interval  $I$  with  $[0, 1] \subset I \subset [a, b]$ , uniform convergence (up to subsequences) still holds if we extend the interval by  $\epsilon$  on both sides (up to the boundary of  $[a, b]$ ). By induction, the claimed assertion then follows. We show this extension result by considering the extension from  $[0, 1]$  to  $[0, 1 + \epsilon]$ , the general case follows similarly and by symmetry.

For this purpose, we observe that for each  $k$ ,  $\gamma^{n_k}|_{[1-\epsilon, 1+\epsilon]}$  is a unique length minimizing geodesic between  $\gamma^{n_k}(1 - \epsilon)$  and  $\gamma^{n_k}(1 + \epsilon)$ , both of which, up to subsequences, converge to some limit points  $q_{-\epsilon}$  and  $q_\epsilon$ , respectively. Hence, employing the first result of this lemma, we obtain that, again up to subsequences,  $\gamma^{n_k}|_{[1-\epsilon, 1+\epsilon]}$  converges uniformly to some  $\psi : [1 - \epsilon, 1 + \epsilon] \rightarrow \mathcal{M}$  with  $\psi(1 - \epsilon) = q_{-\epsilon}$ ,  $\psi(1 + \epsilon) = q_\epsilon$  being a length minimizing geodesics between those points. But by uniform convergence of  $\gamma^{n_k}$  on  $[0, 1]$  and uniqueness of geodesics, we get that  $\psi|_{[1-\epsilon, 1]} = \gamma|_{[1-\epsilon, 1]}$ . Hence they coincide also on  $[1 - \epsilon, 1 + \epsilon]$  and the assertion follows.  $\square$

*Proof of Lemma 3.5.* For the assertion on  $D_S$ , apply Proposition 3.2 with  $C : \mathcal{M}^4 \rightarrow \mathcal{P}(\mathcal{M}^2)$ ,  $C(x, y, u, v) := \{(c', y') \mid c' \in [x, v]_{\frac{1}{2}}, y' \in [u, c']_2\}$  and  $G((x, y, u, v), (c, \tilde{y})) = d(\tilde{y}, y)$ . For the assertion on  $D_S^{\text{sym}}$ , apply Proposition 3.2 with  $C : \mathcal{M}^8 \rightarrow \mathcal{P}(\mathcal{M}^4)$ ,  $C(u^1, v^1, u^2, v^2, x^1, y^1, x^2, y^2) := \{(\tilde{c}^1, \tilde{r}^1, \tilde{c}^2, \tilde{r}^2) \mid (\tilde{c}^1, \tilde{r}^1, \tilde{c}^2, \tilde{r}^2) \in [x^2, v^1]_{\frac{1}{2}} \times [u^1, \tilde{c}^1]_2 \times [x^1, v^2]_{\frac{1}{2}} \times [u^2, \tilde{c}^2]_2\}$  and  $G : \mathcal{M}^8 \times \mathcal{M}^4 \rightarrow \mathbb{R}$  according to  $G((u^1, v^1, u^2, v^2, x^1, y^1, x^2, y^2), (c^1, r^1, c^2, r^2)) := D_S([r^1, y^2], [y^1, r^2])$ .  $\square$

Our next goal is to provide the proof of Lemma 3.6, stating existence and lower semi-continuity results for the parallel-transport-based distance-type functions. Regarding  $D_{\text{pt}}$ , we note that a proof based on a statement similar to Proposition 3.2 (as done for the Schild's variant) seems possible. However, since this would require a generalization of Proposition 3.2 that involves metrics on the tangent bundle we chose to work out the proof for  $D_{\text{pt}}$  directly in order to avoid introducing additional technicalities and notation.

**Lemma A.2.** *The parallel-transport-based distance-type functional  $D_{\text{pt}}$  is lower semi-continuous. In particular, the minimum in (11) is attained for any  $x, y, u, v \in \mathcal{M}$ .*

*Proof.* We consider sequences  $x^n \rightarrow x$ ,  $y^n \rightarrow y$ ,  $u^n \rightarrow u$ , and  $v^n \rightarrow v$ , and show that

$$D_{\text{pt}}([x, y], [u, v]) \leq \liminf_n D_{\text{pt}}([x^n, y^n], [u^n, v^n]). \quad (63)$$

Here, by (11)

$$D_{\text{pt}}([x, y], [u, v]) = \inf_{z, w, \gamma} |z - \text{pt}_{\gamma(1), \gamma(0)}^\gamma w| = \inf_{z, w, \gamma} |\text{pt}_{\gamma(0), \gamma(1)}^\gamma z - w|, \quad (64)$$

where  $z \in \log_x(y)$ ,  $w \in \log_u(v)$ , and  $\gamma : [0, 1] \rightarrow \mathcal{M}$  is a shortest geodesic connecting  $\gamma(0) = x$  and  $\gamma(1) = u$ . We note that the second equation in (64) is true since the parallel transport is an isometry in a Riemannian manifold. We start out choosing subsequences  $x^{n_k} \rightarrow x$ ,  $y^{n_k} \rightarrow y$ ,  $u^{n_k} \rightarrow u$ ,  $v^{n_k} \rightarrow v$ , such that

$$\lim_{k \rightarrow \infty} D_{\text{pt}}([x^{n_k}, y^{n_k}], [u^{n_k}, v^{n_k}]) = \liminf_n D_{\text{pt}}([x^n, y^n], [u^n, v^n]). \quad (65)$$

Then, for each  $k \in \mathbb{N}$ , we choose tangent vectors  $z^{n_k}, w^{n_k}$ , and shortest geodesics  $\gamma^{n_k}$  such that

$$\left| \text{pt}_{\gamma^{n_k}(0), \gamma^{n_k}(1)}^{\gamma^{n_k}} z^{n_k} - w^{n_k} \right| \leq D_{\text{pt}}([x^{n_k}, y^{n_k}], [u^{n_k}, v^{n_k}]) + \frac{1}{n_k}. \quad (66)$$

Here,  $z^{n_k}$  is sitting in the tangent space at the point  $x^{n_k}$ ,  $w^{n_k}$  is sitting in the tangent space at the point  $u^{n_k}$ , and  $\gamma^{n_k}$  is one of the (potentially non-unique) shortest geodesics connecting the points  $\gamma^{n_k}(0) = x^{n_k}$  and  $\gamma^{n_k}(1) = u^{n_k}$ . The parallel transport is understood along  $\gamma^{n_k}$ . By Lemma 3.3 there is a subsequence  $\gamma^{n_l}$  of  $\gamma^{n_k}$  (for convenience, we suppress iterated indices and write  $\gamma^{n_l}$  instead of  $\gamma^{n_{k_l}}$  in the following) such that  $\gamma^{n_l} \rightarrow \gamma$  uniformly on  $[0, 1]$ , for some shortest geodesic  $\gamma : [0, 1] \rightarrow \mathcal{M}$  connecting  $\gamma(0) = x$  and  $\gamma(1) = u$ .

Since  $x^{n_l}$  converges to  $x$ , the geodesic connecting  $x^{n_l}$  and  $x$  is unique for sufficiently large  $l$ . The same is true for  $u^{n_l}$  and  $u$ . So we may identify the tangent vectors  $z^{n_l}$  sitting at  $x^{n_l}$ , and the tangent vectors  $w^{n_l}$  sitting at  $u^{n_l}$ , with their parallel transported versions

$$\bar{z}^{n_l} = \text{pt}_{x_{n_l}, x} z_{n_l}, \quad \bar{w}^{n_l} = \text{pt}_{u_{n_l}, u} w_{n_l}, \quad (67)$$

along the corresponding unique geodesic. Note that the  $\bar{z}^{n_l}$  are sitting in the common point  $x$ , and that the  $\bar{w}^{n_l}$  are sitting in the common point  $u$ . The sequences  $\bar{z}^{n_l}$  is bounded since parallel transport is an isometry and  $z^{n_l} \in \log_{x^{n_l}}(y^{n_l})$ , where both  $x^{n_l}$  and  $y^{n_l}$  converge. The analogous

statements hold for  $\bar{w}^{n_l}$  with the same argument. So, by going to subsequences  $\bar{z}^{n_r}$  of  $\bar{z}^{n_l}$ , and  $\bar{w}^{n_r}$  of  $\bar{w}^{n_l}$ , we get that

$$\bar{z}^{n_r} \rightarrow z, \quad \bar{w}^{n_r} \rightarrow w, \quad (68)$$

for a tangent vector  $z$  sitting in  $x$  and a tangent vector  $w$  sitting in  $u$ . We have that the limit  $z \in \log_x(y)$ , and that  $w \in \log_u(v)$ , since the exponential map from  $T\mathcal{M} \rightarrow \mathcal{M}$  is differentiable in a Riemannian manifold which implies  $\exp_x z = \lim_{r \rightarrow \infty} \exp_{x^{n_r}} \bar{z}^{n_r} = \lim_{r \rightarrow \infty} y^{n_r} = y$ , and  $\exp_u w = v$ .

We are now prepared to estimate  $|\text{pt}_{\gamma(0), \gamma(1)}^\gamma z - w|$  which allows us to bound  $D_{\text{pt}}([x, y], [u, v])$  from above; see (64). Since the parallel transport is an isometry, we have

$$|\text{pt}_{\gamma(0), \gamma(1)}^\gamma z - w| = |\text{pt}_{u, u^{n_r}} \text{pt}_{\gamma(0), \gamma(1)}^\gamma z - \text{pt}_{u, u^{n_r}} w|. \quad (69)$$

We further have that

$$\varepsilon_r := |\text{pt}_{u, u^{n_r}} \text{pt}_{\gamma(0), \gamma(1)}^\gamma z - \text{pt}_{\gamma^{n_r}(0), \gamma^{n_r}(1)}^{\gamma^{n_r}} \text{pt}_{x, x^{n_r}} \bar{z}^{n_r}| \rightarrow 0 \quad \text{as } r \rightarrow \infty \quad (70)$$

which is a consequence of the parallel transport along nearby geodesics being differentially dependent on the variation of the geodesics. Using (70) together with (69), we have

$$\begin{aligned} |\text{pt}_{\gamma(0), \gamma(1)}^\gamma z - w| &\leq |\text{pt}_{\gamma^{n_r}(0), \gamma^{n_r}(1)}^{\gamma^{n_r}} \text{pt}_{x, x^{n_r}} \bar{z}^{n_r} - \text{pt}_{u, u^{n_r}} w| + \varepsilon_r \\ &\leq |\text{pt}_{\gamma^{n_r}(0), \gamma^{n_r}(1)}^{\gamma^{n_r}} \text{pt}_{x, x^{n_r}} \bar{z}^{n_r} - \text{pt}_{u, u^{n_r}} \bar{w}^{n_r}| + |\bar{z}^{n_r} - z| + |\bar{w}^{n_r} - w| + \varepsilon_r. \end{aligned} \quad (71)$$

The second inequality is a consequence of the triangle inequality together with the fact that parallel transport is an isometry. We take the limit w.r.t.  $r$  on the right hand side of (71): by (68), we have that  $\bar{z}^{n_r} \rightarrow z$ , and that  $\bar{w}^{n_r} \rightarrow w$  and by (70) that  $\varepsilon_r \rightarrow 0$  which implies

$$\begin{aligned} |\text{pt}_{\gamma(0), \gamma(1)}^\gamma z - w| &\leq \lim_{r \rightarrow \infty} |\text{pt}_{\gamma^{n_r}(0), \gamma^{n_r}(1)}^{\gamma^{n_r}} \text{pt}_{x, x^{n_r}} \bar{z}^{n_r} - \text{pt}_{u, u^{n_r}} \bar{w}^{n_r}| \\ &\leq \lim_{r \rightarrow \infty} \left( D_{\text{pt}}([x^{n_r}, y^{n_r}], [u^{n_r}, v^{n_r}]) + \frac{1}{n_r} \right) = \liminf_n D_{\text{pt}}([x^n, y^n], [u^n, v^n]). \end{aligned} \quad (72)$$

The second inequality in (72) is a consequence of our choice of  $z^{n_k}, w^{n_k}$ , and  $\gamma^{n_k}$  in (66) and the definition of  $\bar{z}^{n_l}, \bar{w}^{n_l}$  as in (67), and the equality in the last line follows by our choice of subsequences in (65). Then passing to the infimum according to (64) shows (63) which shows the first assertion.

To see that the infimum in (11) is attained for any  $x, y, u, v \in \mathcal{M}$ , we choose the constant sequences  $x^n := x$ ,  $y^n := y$ ,  $u^n := u$ , and  $v^n := v$ , and apply the previous result to these sequences. This shows the second assertion and completes the proof.  $\square$

Having shown Lemma A.2 we can now employ Proposition 3.2 to show existence and lower semi-continuity results for  $D_{\text{pt}}^{\text{sym}}$ . As a preparation we need the following lemma.

**Lemma A.3.** *Let  $u^n \rightarrow u$ ,  $\tilde{u}^n \rightarrow \tilde{u}$ , and  $y^n \rightarrow y$  in the complete manifold  $\mathcal{M}$ . We consider a sequence  $(r^n)_n$  with*

$$r^n \in \exp(\text{pt}_{\tilde{u}^n}(w^n)) \quad \text{with} \quad w^n \in \log_{u^n} y^n. \quad (73)$$

*Then there is a subsequence  $(r^{n_k})_k$  which converges and a limit  $r = \lim_{k \rightarrow \infty} r^{n_k}$  such that  $r$  fulfills*

$$r \in \exp(\text{pt}_{\tilde{u}}(w)) \quad \text{for some} \quad w \in \log_u y. \quad (74)$$



*Proof.* The present proof essentially employs the techniques already used in the proof of Lemma A.2. For this reason we keep the following arguments rather short. The sequence  $\bar{w}^n := \text{pt}_u(w^n)$  is bounded since parallel transport is an isometry. So, by going to a subsequence  $\bar{w}^{n_l}$  of  $\bar{w}^n$ , we get that  $\bar{w}^{n_l} \rightarrow w$  for a tangent vector  $w$  sitting in  $u$ . We have that the limit  $w \in \log_u(y)$ , since the exponential map  $T\mathcal{M} \rightarrow \mathcal{M}$  is differentiable which implies  $\exp_u w = \lim_{l \rightarrow \infty} \exp_{u^{n_l}} w^{n_l} = \lim_{l \rightarrow \infty} y^{n_l} = y$ . For each  $l$  we choose a distance-minimizing geodesic  $\gamma^{n_l}$  connecting  $\gamma^{n_l}(0) = u^{n_l}$  with  $\gamma^{n_l}(1) = \tilde{u}^{n_l}$ . Then we use Lemma 3.3 to choose a subsequence of geodesics  $\gamma^{n_k}$  of  $\gamma^{n_l}$  (suppressing iterated subindices) which uniformly converges to a geodesic  $\gamma : [0, 1] \rightarrow \mathcal{M}$  connecting  $\gamma(0) = u$  with  $\gamma(1) = \tilde{u}$ . Then, since parallel transport along nearby geodesics is differentially dependent on the variation of the geodesics and by an argument similar to the one used for the convergence in Equation (70),  $\text{pt}_{\tilde{u}^{n_k}, \tilde{u}} \text{pt}_{u^{n_k}, \tilde{u}^{n_k}} w^{n_k} \rightarrow \text{pt}_{u, \tilde{u}}^\gamma w$  as  $k \rightarrow \infty$ . Then, by the continuity of the exponential map,  $r = \lim_{k \rightarrow \infty} r^{n_k} = \lim_{k \rightarrow \infty} \exp(\text{pt}_{u^{n_k}, \tilde{u}^{n_k}}^\gamma(w^{n_k})) = \exp(\text{pt}_{u, \tilde{u}}^\gamma(w))$ , i.e.,  $r^{n_k}$  converges and (74) holds true which was the assertion to show.  $\square$

Finally, the proof of Lemma 3.6 follows as consequence.

*Proof of Lemma 3.6.* For  $D_{\text{pt}}$ , this is the assertion of Lemma A.2. For  $D_{\text{pt}}^{\text{sym}}$ , we apply Proposition 3.2 with  $C : \mathcal{M}^8 \rightarrow \mathcal{P}(\mathcal{M}^2)$ ,  $C(u^1, v^1, u^2, v^2, x^1, y^1, x^2, y^2) := \{\exp(\text{pt}_{x^2} w) \mid w \in \log_{u^1}(v^1)\} \times \{\exp(\text{pt}_{x^1} w) \mid w \in \log_{u^2}(v^2)\}$ , and  $G : \mathcal{M}^8 \times \mathcal{M}^2 \rightarrow \mathbb{R}$  according to  $G((u^1, v^1, u^2, v^2, x^1, y^1, x^2, y^2), (r^1, r^2)) := D_{\text{pt}}([r^1, y^2], [y^1, r^2])$ . The Lemmata A.2 and A.3 ensure that  $C$  and  $G$  satisfy the assumption of Proposition 3.2.  $\square$

### A.3 Proofs for Section 4

*Proof of Lemma 4.1.* Let us consider the Schild's ladder mapping  $u_{i-1} \mapsto S(u_{i-1}, y_{i-1}, u_i) = [u_{i-1}, [u_i, y_{i-1}]_{\frac{1}{2}}]_{\frac{1}{2}}$  as a function of  $u_{i-1}$ . Since the points  $y_{i-1}, u_i$  are fixed, so is their midpoint  $m = [u_i, y_{i-1}]_{\frac{1}{2}}$ . Now  $S(u_{i-1}, y_{i-1}, u_i)$  is obtained by evaluating the geodesic  $\gamma : [0, 1] \rightarrow \mathcal{M}$  connecting  $u_{i-1} = \gamma(0)$  and  $m = \gamma(1)$  at time 2, thus considering  $\gamma(2)$ . Hence, the differential  $T_1$  of  $S$  w.r.t.  $u_{i-1}$  is related to the Jacobi fields along  $\gamma$  as follows: consider those Jacobi fields  $\mathcal{J}$  along  $\gamma$  for which  $J(1) = 0$  which means that they belong to geodesic variations leaving  $m$  fixed. Then the adjoint of the mapping

$$J(0) \mapsto J(2), \quad J \in \mathcal{J} \quad (75)$$

equals  $T_1$ .

If the manifold is a Riemannian symmetric space, then the mapping  $u_{i-1} \mapsto S(u_{i-1}, y_{i-1}, u_i) = [u_{i-1}, [u_i, y_{i-1}]_{\frac{1}{2}}]_{\frac{1}{2}}$  is a geodesic reflection, see, e.g., [40]. We consider an orthonormal basis  $(v_n)_n$  of eigenvectors of the self-adjoint Jacobi operator  $J \mapsto R(\frac{\gamma'(1)}{|\gamma'(1)|}, J) \frac{\gamma'(1)}{|\gamma'(1)|}$  with corresponding eigenvalues associated  $\lambda_n$ , and  $v_1$  tangent to  $\gamma'(1)$ . Then the mapping  $J(0) \mapsto J(2)$ ,  $J \in \mathcal{J}$ , in (75) can be written as

$$\sum_n \alpha_n f(\lambda_n) \text{pt}_{m, u_{i-1}} v_n \mapsto - \sum_n \alpha_n f(\lambda_n) \text{pt}_{m, S(u_{i-1}, y_{i-1}, u_i)} v_n, \quad (76)$$

where the  $\alpha_n$  are the coefficients of the corresponding basis representation and the function  $f$ , depending on the sign of  $\lambda_n$ , is given by  $f(\lambda_n) = d$ , if  $\lambda_n = 0$ , by  $f(\lambda_n) = (\sqrt{\lambda_n})^{-1} \sin(\sqrt{\lambda_n} d)$ , if  $\lambda_n > 0$  and by  $f(\lambda_n) = (\sqrt{-\lambda_n})^{-1} \sinh(\sqrt{-\lambda_n} d)$ , if  $\lambda_n < 0$ , where  $d$  is the distance between  $m$  and  $u_{i-1}$  (which equals the distance between  $m$  and  $S(u_{i-1}, y_{i-1}, u_i)$ ). This results from the fact that  $f(\lambda_n) = x(0)$ , the evaluation at 0 of the solution  $x(t)$  of the scalar initial value problem  $x'' = -\lambda_n d^2 x$ ,  $x(1) = 0, x'(1) = -d$ . We observe that (76) states that the mapping  $J(0) \mapsto J(2)$ ,  $J \in \mathcal{J}$ , equals the identity multiplied with  $-1$  (up to parallel transport). (We note

that this can also be concluded from the derivations in [40] near Theorem 1.) Hence the adjoint of the mapping  $J(0) \mapsto J(2)$  coincides with the parallel transport multiplied by  $-1$  as in (39), in particular, it is an isometry, and the proof is complete.  $\square$

*Proof of Lemma 4.2.* We observe that the concatenation of the mappings  $m \mapsto [u_{i-1}, m]_2$  and  $y_{i-1} \mapsto [u_i, y_{i-1}]_{\frac{1}{2}}$  equals the mapping  $y_{i-1} \mapsto [u_{i-1}, [u_i, y_{i-1}]_{\frac{1}{2}}]_2$ . Hence, we conclude from the discussion leading to (38) in connection with the chain rule the validity of (40).

It remains to express  $T_3, T_4$  in terms of Jacobi fields. Concerning  $T_3$  we consider the geodesic  $\gamma$  connecting  $u_{i-1} = \gamma(0)$  and  $m = \gamma(1)$ .  $T_3$  is related to the Jacobi fields along  $\gamma$  as follows: consider those Jacobi fields  $\mathcal{J}$  along  $\gamma$  for which  $J(0) = 0$  which means that they belong to geodesic variations leaving  $u_{i-1}$  fixed. Then, the adjoint of the mapping

$$J(1) \mapsto J(2), \quad J \in \mathcal{J} \quad (77)$$

equals  $T_3$ . Concerning  $T_4$ , we consider the geodesic  $\xi$  connecting  $y_{i-1} = \xi(0)$  and  $u_i = \xi(1)$ .  $T_4$  is related to the Jacobi fields  $\mathcal{J}'$  along  $\xi$  for which  $J(1) = 0$  by the mapping

$$J(0) \mapsto J(\tfrac{1}{2}), \quad J \in \mathcal{J}'. \quad (78)$$

$\square$

*Proof of Lemma 4.3.* We consider the assertion of Lemma 4.3 for the mapping  $T_3$ . By the proof of Lemma 4.2, we have to determine the adjoint of the mapping given by (77) more explicitly. Since the covariant derivative of the curvature tensor equals zero in a symmetric space, the differential equation for the Jacobi fields  $J$  in terms of the scalar coefficient  $x$  of the vector field  $\text{pt}_{m, [m, u_{i-1}]_t} v_n$  obtained by parallel transport of the eigenvector  $v_n$  along  $\gamma$  reads  $x'' = -\lambda_n d^2 x$ ,  $x(0) = 0$ . Solving this scalar ODE shows (41). Then,  $f(\lambda_n)$  corresponds to the value  $x(2)$  of the solution of the scalar ODE of the previous line with additional boundary condition  $x'(1) = 1$ . Solving this scalar ODE yields (42). We have already derived the corresponding statement for  $T_4$  in [10].  $\square$

Our next goal is to show Lemma 4.5. To this end, we introduce the mapping  $F_t$  which is a slight extension of  $D_{\text{pt}}$  for different parameters as follows. For  $t \in [0, 1]$ , we consider the mapping  $F_t : \mathcal{M} \times \mathcal{M} \times \mathcal{M} \times \mathcal{M} \rightarrow [0, \infty)$ , given by

$$F_t(u_i, u_{i-1}, y_i, y_{i-1}) = |\text{pt}_{0,t} \log_{u_i} y_i - \text{pt}_{1,t} \log_{u_{i-1}} y_{i-1}|, \quad (79)$$

where, for the definition of  $F_t$  and in the following Lemma, we use the shorthand notation  $\text{pt}_{s,t}$  to denote the parallel transport from  $[u_i, u_{i-1}]_s$  to  $[u_i, u_{i-1}]_t$  and the norm on the right hand side is the one introduced by the Riemannian scalar product in the point  $[u_i, u_{i-1}]_t$ . Note that  $F_0(u_i, u_{i-1}, y_i, y_{i-1}) = D_{\text{pt}}([u_i, y_i], [u_{i-1}, y_{i-1}])$ , so in order to obtain the derivative of  $D_{\text{pt}}$  it suffices to differentiate  $F_0$  w.r.t. its four arguments. The following lemma shows that we can also consider  $F_t$  instead.

**Lemma A.4.** *The function  $F_t$  given by (79) is independent of  $t$ .*

*Proof.* The Riemannian scalar product is invariant under parallel transport, i.e., for any tangent vectors  $v, v'$  sitting in the arbitrary point  $x$ , and the parallel transport along any curve  $\gamma$  with any points  $y, z$  sitting on that curve, we have  $\langle \text{pt}_{x,y} v, \text{pt}_{x,y} v' \rangle_y = \langle \text{pt}_{x,z} v, \text{pt}_{x,z} v' \rangle_z$ . Hence, for all  $s, t \in [0, 1]$ ,

$$\begin{aligned} |\text{pt}_{0,s} \log_{u_i} y_i - \text{pt}_{1,s} \log_{u_{i-1}} y_{i-1}|^2 &= |\text{pt}_{t,s} \text{pt}_{0,t} \log_{u_i} y_i - \text{pt}_{t,s} \text{pt}_{1,t} \log_{u_{i-1}} y_{i-1}|^2 \\ &= |\text{pt}_{0,t} \log_{u_i} y_i - \text{pt}_{1,t} \log_{u_{i-1}} y_{i-1}|^2. \end{aligned} \quad (80)$$

Since the first expression in (80) equals the square of  $F_s$ , and the last expression equals the square of  $F_t$ , this shows that  $F_s = F_t$  for all  $s, t \in [0, 1]$ .  $\square$

As pointed out below Lemma 4.5, Lemma 4.5 is obtained by specifying  $F = F_0$  in the following lemma.

**Lemma A.5.** *The function  $F_t$  given by (79) is symmetric with respect to interchanging  $(u_i, y_i)$  with  $(u_{i-1}, y_{i-1})$  i.e.,  $F_t(u_i, u_{i-1}, y_i, y_{i-1}) = F_t(u_{i-1}, u_i, y_{i-1}, y_i)$ . In particular, for points with  $F_t \neq 0$ , the gradient of the function  $F_t$  w.r.t. the third variable  $y_i$ , is given by  $\nabla_{y_i} F_t(u_i, u_{i-1}, y_i, y_{i-1}) = \nabla_{y_{i-1}} F_t(u_{i-1}, u_i, y_{i-1}, y_i)$ , where  $\nabla_{y_{i-1}} F_t$  denotes the derivative of  $F_t$  w.r.t. the fourth argument. Further, again for points with  $F_t \neq 0$ , the gradient of the function  $F_t$  w.r.t. the first component variable  $u_i$ , is given by  $\nabla_{u_i} F_t(u_i, u_{i-1}, y_i, y_{i-1}) = \nabla_{u_{i-1}} F_t(u_{i-1}, u_i, y_{i-1}, y_i)$ , where  $\nabla_{u_{i-1}} F_t$  denotes the derivative of  $F_t$  w.r.t. the second argument.*

*Proof.* By the definition of  $F_t$  in (79), we have  $F_t(u_i, u_{i-1}, y_i, y_{i-1}) = F_{1-t}(u_{i-1}, u_i, y_{i-1}, y_i)$ . By Lemma A.4,  $F_t$  is independent of  $t$ . Together, this implies the identity  $F_t(u_i, u_{i-1}, y_i, y_{i-1}) = F_t(u_{i-1}, u_i, y_{i-1}, y_i)$  which is the first assertion of the lemma. The following two assertions on the gradients are then an immediate consequence of this symmetry property.  $\square$

*Proof of Lemma 4.6.* We note that  $F$  agrees with the function  $F_0$  with  $F_t$  as in Equation (79) by definition and, consequently, also with  $F_1$  by Lemma A.4. For fixed  $u_i$ ,  $u_{i-1}$  and  $y_i$ , we decompose the mapping  $y_{i-1} \mapsto F_1(u_i, u_{i-1}, y_i, y_{i-1})$  into the mappings  $G, H$ , i.e.,  $F_1 = H \circ G$ , where

$$G(y_{i-1}) = \log_{u_{i-1}} y_{i-1} \quad (81)$$

locally maps the manifold  $\mathcal{M}$  to the tangent space at  $u_{i-1}$ , and

$$H(w) = |w - z|, \quad \text{where } z = \text{pt}_{u_i, u_{i-1}} \log_{u_i} y_i, \quad (82)$$

maps from the tangent space at  $u_{i-1}$  to the positive real numbers. The differential of  $H$  (as a map defined on the tangent bundle) is given by  $d_\xi H(\eta) = \langle \frac{\xi - z}{|\xi - z|}, \eta \rangle$ , for  $\xi \neq z$ . Here  $\xi$  is a point in the tangent space at  $u_{i-1}$ , and  $\eta$  is the direction in the tangent space with respect to which the differentiation is performed. Hence, the gradient of  $H$  at  $\xi$  equals  $\frac{\xi - z}{|\xi - z|}$ . So the gradient of  $H$  at the point  $\xi = \log_{u_{i-1}} y_{i-1}$  equals

$$\nabla H(\log_{u_{i-1}} y_{i-1}) = \left( \log_{u_{i-1}} y_{i-1} - z \right) / \left| \log_{u_{i-1}} y_{i-1} - z \right|. \quad (83)$$

In order to get the gradient of  $F_1$ , we have to multiply  $\nabla H(\log_{u_{i-1}} y_{i-1})$  with the adjoint of the differential of  $G$ . The mapping  $G$  is related to the Jacobi fields along the geodesic  $\gamma : [0, 1] \rightarrow \mathcal{M}$  connecting the points  $\gamma(0) = u_{i-1}$  and  $\gamma(1) = y_{i-1}$  as follows. Consider the collection  $\mathcal{J}$  of Jacobi fields  $J$  along  $\gamma$  for which  $J(0) = 0$ , which means that they belong to geodesic variations leaving  $u_{i-1}$  fixed. Then the mapping

$$J(1) \mapsto J'(0), \quad J \in \mathcal{J}. \quad (84)$$

equals the differential of  $G$ .  $\square$

*Proof of Lemma 4.7.* We use the basis  $\{v_n\}$  to express the operator  $T$  of Lemma 4.6 (which is given as the adjoint of the derivative of  $G$  defined by (81) evaluated at  $y_{i-1}$ ). We use the expression (84) for the derivative of  $G$ . Since the covariant derivative of the curvature tensor equals zero in a symmetric space, the differential equation for the Jacobi fields  $J$  in terms of the

scalar coefficient  $x$  of the vector field  $\text{pt}_{u_{i-1}, [u_{i-1}, y_{i-1}]_t} v_n$  obtained by parallel transport of the eigenvector  $v_n$  along  $\gamma$  reads

$$x'' = -\lambda_n d^2 x, \quad x(0) = 0.$$

This shows (52). Further,  $f(\lambda_n)$  corresponds to the value  $x'(0)$  of the solution of the scalar ODE of the previous line with additional boundary condition  $x(1) = 1$ . Solving this scalar ODE yields (53).  $\square$

*Proof of Lemma 4.8.* We again note that  $F$  agrees with the function  $F_1$  introduced in Equation (79). So our task is to determine the gradient of the function  $F_1$  for points with  $F_1 \neq 0$  given by (79) w.r.t. the variable  $u_{i-1}$ . We analyze the structure of  $F_1$  as a function of  $u_{i-1}$ . To this end, we consider the two vector fields  $L : u_{i-1} \mapsto \log_{u_{i-1}} y_{i-1}$  and  $B : u_{i-1} \mapsto \text{pt}_{u_i, u_{i-1}} z$ , where  $z = \log_{u_i} y_i$  introduced above. Remember that  $\text{pt}_{u_i, u_{i-1}}$  denotes the parallel transport from the point  $u_i$  to the (varying) point  $u_{i-1}$  along a shortest geodesic connecting these points. We note that the parallel transport here depends on the varying end point  $u_{i-1}$ . Further note that  $z = \log_{u_i} y_i$  does not depend on  $u_{i-1}$  and is therefore fixed. Using this notation we may write (79) as

$$F_1(u_i, \cdot, y_i, y_{i-1}) : u_{i-1} \mapsto |L(u_{i-1}) - B(u_{i-1})|. \quad (85)$$

In order to differentiate (85) w.r.t.  $u_{i-1}$  we need some more preparation. Recall that the Levi-Civita connection is metric. Hence, for any two vector fields  $V_t, W_t$  along a geodesic  $\gamma$ ,

$$\frac{d}{dt} \langle V_t, W_t \rangle = \langle \frac{D}{dt} V_t, W_t \rangle + \langle V_t, \frac{D}{dt} W_t \rangle. \quad (86)$$

Recall that we use the symbol  $\frac{D}{dt}$  to denote the covariant derivative of the corresponding vector field along the curve  $\gamma$ . Thus, for any two vector fields  $V_t, W_t$  with  $V_t \neq W_t$ , we have

$$\begin{aligned} \frac{d}{dt} |V_t - W_t| &= \frac{1}{2|V_t - W_t|} \cdot 2 \langle V_t - W_t, \frac{D}{dt} V - \frac{D}{dt} W \rangle \\ &= \langle \frac{V_t - W_t}{|V_t - W_t|}, \frac{D}{dt} V - \frac{D}{dt} W \rangle. \end{aligned} \quad (87)$$

Since we have chosen the  $v_n$  to be an orthogonal basis of the tangent space at  $u_{i-1}$ , the coordinate representation of the gradient in this basis is given as the directional derivatives w.r.t. the basis vectors. The curves  $t \mapsto \exp_{u_{i-1}} t v_n$  precisely yield these tangent vectors. This explains (54), (56), as well as the first identity in (55). The second identity in (55) is a consequence of (87).  $\square$

*Proof of Lemma 4.9.* We notice that the proof of this lemma uses well-known facts on the connection of Jacobi fields and the exponential map (see for instance the books [76, 38]) which is the reason why we kept it rather short. We consider the Jacobi field  $J^n$  associated with the following geodesic variation

$$f^n(s, t) = \exp_{c^n(t)} \left( s \log_{c^n(t)} y_{i-1} \right), \quad \text{where } c^n(t) = \exp_{u_{i-1}} t v_n. \quad (88)$$

Then the desired covariant derivative is connected with the Jacobi field  $J^n$  associated with the geodesic variation  $f^n$  by

$$\frac{D}{dt} \Big|_{t=0} L_t^n = \frac{D}{ds} \Big|_{s=0} J^n(s), \quad \text{where } J^n(s) = \frac{d}{dt} \Big|_{t=0} f^n(s, t). \quad (89)$$

This identity can be seen as follows. By the definition of  $f^n$ , we have  $L_t^n = \log_{c(t)} y_{i-1} = \frac{d}{ds}|_{s=0} f^n(s, t)$ . Hence,

$$\begin{aligned} \frac{D}{dt}|_{t=0} L_t^n &= \frac{D}{dt}|_{t=0} \frac{d}{ds}|_{s=0} f^n(s, t) \\ &= \frac{D}{ds}|_{s=0} \frac{d}{dt}|_{t=0} f^n(s, t) = \frac{D}{ds}|_{s=0} J^n(s), \end{aligned} \quad (90)$$

which shows (89). We further notice that

$$J^n(0) = v_n, \quad J^n(1) = 0. \quad (91)$$

The first part of (91) is a consequence of

$$J^n(0) = \frac{d}{dt}|_{t=0} f^n(0, t) = \frac{d}{dt}|_{t=0} c^n(t) = v_n. \quad (92)$$

The second equality of (91) is a consequence of the mapping  $t \mapsto f^n(1, t) = y_{i-1}$ , being constant.

We notice that (91) determines  $J^n$  uniquely which, in turn, yields the desired derivative of  $L^n$  via (88) as being equal to  $(J^n)'(0) := \frac{D}{ds}|_{s=0} J^n$ . So it only remains to determine this uniquely determined Jacobi field  $J^n$ .

Since  $\mathcal{M}$  is a symmetric space, and thus the curvature tensor is parallel, and since  $v_n$  is an eigenvector of the Jacobi operator, we end up with the problem of determining  $x'(0)$  where  $x$  is the solution of the scalar boundary value problem

$$x''(s) + d^2 \lambda_n x(s) = 0, \quad x(0) = 1, \quad x(1) = 0, \quad (93)$$

where  $d = d(y_{i-1}, u_{i-1}) = |\gamma'(t)|$  for all  $t \in [0, 1]$ . Here,  $\lambda_n$  is the corresponding eigenvalue of the Jacobi operator. Depending on the sign of  $\lambda_n$ , the solution of (93) is given as follows. If  $\lambda_n = 0$ ,  $x(s) = 1 - s$ , and so  $x'(0) = -1$ . If  $\lambda_n > 0$ , the general solution of the ODE is  $x(s) = \alpha \cos(d\sqrt{\lambda_n}s) + \beta \sin(d\sqrt{\lambda_n}s)$  with real parameters  $\alpha, \beta$ . Then  $x(0) = 1$  implies  $\alpha = 1$  which, in turn, yields using  $0 = x(1) = \cos(d\sqrt{\lambda_n}) + \beta \sin(d\sqrt{\lambda_n})$  that  $\beta = -\cos(d\sqrt{\lambda_n})/\sin(d\sqrt{\lambda_n})$ . Hence,  $x'(0) = -\cos(d\sqrt{\lambda_n})/\sin(d\sqrt{\lambda_n}) \cdot d\sqrt{\lambda_n}$ . If  $\lambda_n < 0$ , an analogous calculation replacing the trigonometric polynomials by their hyperbolic analogues yields that  $x'(0) = -\cosh(d\sqrt{-\lambda_n})/\sinh(d\sqrt{-\lambda_n}) \cdot d\sqrt{-\lambda_n}$ . This shows (57) and completes the proof.  $\square$

*Proof of Lemma 4.10.* In order to covariantly differentiate the mapping

$$t \mapsto B_t^n = \text{pt}_{u_i, \exp_{u_{i-1}} t v_n} z, \quad (94)$$

we differentiate the mapping

$$t \mapsto P_t^n z := \text{pt}_{\exp_{u_{i-1}} t v_n, u_{i-1}} \text{pt}_{u_i, \exp_{u_{i-1}} t v_n} z \quad (95)$$

in the tangent space  $T_{u_{i-1}} \mathcal{M}$  of  $u_{i-1}$ . This follows from the relation between parallel transport and the covariant derivative, see for instance [76].

If  $v_n$  is parallel to  $\log_{u_{i-1}} u_i$ , then the mapping in (95) is constant, and therefore, its derivative is 0 which shows the first statement in (58).

We show the second part of (58). We may assume that  $v_n$  is orthogonal to  $\exp_{u_{i-1}}^{-1} u_i$ . We have to differentiate the mapping in (95), which means calculating  $\lim_{t \rightarrow 0} \frac{1}{t} (P_t^n - P_0^n)$ . Since the parallel transport is an isometry, the differential of (95) is an infinitesimal rotation (up to the identity) applied to  $z$ . We start out to calculate  $P_t^n - P_0^n$  in the basis of  $T_{u_{i-1}} \mathcal{M}$ . We note that  $P_t^n - P_0^n$  corresponds to the holonomy along the (spherical) triangle  $\Delta$  connecting the points  $u_i$ ,  $\exp_{u_{i-1}} t v_n$  and  $u_{i-1}$ . We observe that the rotation angle  $\alpha_t$  of the rotation  $P_t^n - P_0^n$  equals the

spherical excess  $A_t + B_t + C_t - \pi$  of the triangle  $\Delta_t$ , where  $A_t$  is the angle at  $u_i$ ,  $C_t$  is the angle at  $u_{i-1}$  and  $B_t$  is the angle at  $\exp_{u_{i-1}} tv_n$  of the triangle  $\Delta_t$ . Hence,

$$\lim_{t \rightarrow 0} \frac{1}{t} (P_t^n - P_0^n) = \begin{pmatrix} 0 & \lim_{t \rightarrow 0} \frac{\alpha_t}{t} \\ -\lim_{t \rightarrow 0} \frac{\alpha_t}{t} & 0 \end{pmatrix} \text{pt}_{u_i, u_{i-1}}, \quad \text{where } \alpha_t = A_t + B_t + C_t - \pi. \quad (96)$$

Here, the first identity is a consequence of the chain rule combined with  $\alpha_0 = 0$ . Since  $\frac{\sin t}{t} = 1 + o(1)$ , and since  $C_t = \pi/2$  by the orthogonality of  $v_n$  and  $\log_{u_{i-1}} u_i$ , we get

$$\omega = \lim_{t \rightarrow 0} \frac{\alpha_t}{t} = \lim_{t \rightarrow 0} \frac{A_t + B_t + C_t - \pi}{\sin t} = \lim_{t \rightarrow 0} \left( \frac{A_t}{\sin t} + \frac{B_t - \pi/2}{\sin t} \right). \quad (97)$$

We recall that  $d = d(u_i, u_{i-1})$  and use the following identities for spherical triangles with an angle of  $\pi/2$  (cf. [84])

$$A_t = \arctan\left(\frac{\tan t}{\sin d}\right), \quad B_t = \arctan\left(\frac{\tan d}{\sin t}\right). \quad (98)$$

Using the Taylor expansion of the arctan function w.r.t. 0 we obtain that

$$\lim_{t \rightarrow 0} \frac{A_t}{\sin t} = \lim_{t \rightarrow 0} \frac{\frac{\tan t}{\sin d} + O\left(\left(\frac{\tan t}{\sin d}\right)^3\right)}{\sin t} = \lim_{t \rightarrow 0} \frac{1}{\cos t \sin d} = \frac{1}{\sin d}. \quad (99)$$

Further, by L'Hospital's rule,

$$\begin{aligned} \lim_{t \rightarrow 0} \frac{B_t - \pi/2}{\sin t} &= \lim_{t \rightarrow 0} \frac{\arctan\left(\frac{\tan d}{\sin t}\right) - \pi/2}{\sin t} = \lim_{t \rightarrow 0} \frac{-\frac{\tan d}{\sin^2 t} \cdot \cos t}{1 + \left(\frac{\tan d}{\sin t}\right)^2} \cdot \frac{1}{\cos t} \\ &= \lim_{t \rightarrow 0} \left(-\frac{1}{\tan d}\right) \cdot \left(1 - \frac{1}{1 + \left(\frac{\tan d}{\sin t}\right)^2}\right) = -\frac{1}{\tan d}. \end{aligned} \quad (100)$$

Now, we combine (99) with (100) and conclude, using (97), that  $\omega = \frac{1}{\sin d} - \frac{1}{\tan d}$ . Together with (96), this shows (58). To see (59), we notice that the connection is linear w.r.t. the direction of differentiation. Therefore, (59) is a consequence of (58) and the expression  $\langle v_n, w \rangle$  equals the coefficient of the corresponding linear combination.

If  $u_i = u_{i-1}$ , the mapping in (95) is constant; hence differentiation of this mapping yields zero which implies that the differential  $\frac{D}{dt}|_{t=0} B_t^n = 0$ . This shows the last assertion and completes the proof.  $\square$

*Proof of Lemma 4.11.* Since the space of positive matrices is a Riemannian symmetric space representable as quotient of matrix Lie groups there are explicit formulae for the objects of Riemannian geometry such as the exp mapping and the parallel transport available. The corresponding formulae may be found in, e.g., [77]. Our first task is to explicitly express the mapping  $B_t^n$  in the space of positive matrices which form a symmetric space. We use the notation  $\gamma_t : [0, 1] \rightarrow \mathcal{M}$  to denote the geodesic starting in  $u_{i-1}$  with direction  $v$ , i.e.,

$$\gamma_t := \exp_{u_{i-1}} tv = u_{i-1}^{\frac{1}{2}} \text{Exp} \left( u_{i-1}^{-\frac{1}{2}} tv u_{i-1}^{-\frac{1}{2}} \right) u_{i-1}^{\frac{1}{2}}. \quad (101)$$

Here, Exp denotes the ordinary matrix exponential. Then,  $B_t^n$  may be expressed in the space of positive matrices by

$$B_t^n = \text{pt}_{u_i, \gamma_t} z = u_i^{\frac{1}{2}} \bar{\gamma}_t^{\frac{1}{2}} \bar{z} \bar{\gamma}_t^{\frac{1}{2}} u_i^{\frac{1}{2}} \quad (102)$$

where

$$\bar{\gamma}_t = u_i^{-\frac{1}{2}} \gamma_t u_i^{-\frac{1}{2}}, \quad \text{and} \quad \bar{z} = u_i^{-\frac{1}{2}} z u_i^{-\frac{1}{2}}. \quad (103)$$

(We refer for instance to [77] for the corresponding formulae for the parallel transport.) The covariant derivative in the space of positive matrices may be expressed in terms of the ordinary derivative of a curve in the vector space of matrices plus some “correction terms” (as for instance explained in [77]). In our situation, we have

$$\frac{D}{dt}|_{t=0} B_t^n = \frac{d}{dt}|_{t=0} B_t^n - \frac{1}{2} \left( v u_{i-1}^{-1} \text{pt}_{u_i, u_{i-1}} z + \text{pt}_{u_i, u_{i-1}} z u_{i-1}^{-1} v \right). \quad (104)$$

We denote the terms in brackets in (104) by

$$S + S^\top, \quad S := v u_{i-1}^{-1} \text{pt}_{u_i, u_{i-1}} z. \quad (105)$$

We further have that, by (102),

$$\text{pt}_{u_i, u_{i-1}} z = B_0^n = u_i^{\frac{1}{2}} \bar{u}_{i-1}^{\frac{1}{2}} \bar{z} \bar{u}_{i-1}^{\frac{1}{2}} u_i^{\frac{1}{2}}, \quad \text{where} \quad \bar{u}_{i-1} = u_i^{-\frac{1}{2}} u_{i-1} u_i^{-\frac{1}{2}}, \quad (106)$$

and so we may explicitly express  $S$  in terms of matrix multiplications by

$$S := v u_{i-1}^{-1} u_i^{\frac{1}{2}} \bar{u}_{i-1}^{\frac{1}{2}} \bar{z} \bar{u}_{i-1}^{\frac{1}{2}} u_i^{\frac{1}{2}}. \quad (107)$$

In view of (105) and (104), we have to compute  $\frac{d}{dt}|_{t=0} B_t^n$  in order to derive an explicit representation of  $\frac{D}{dt}|_{t=0} B_t^n$  in terms of matrices. Differentiating (102), we have that

$$\frac{d}{dt}|_{t=0} B_t^n = u_i^{\frac{1}{2}} \left( \frac{d}{dt}|_{t=0} \bar{\gamma}_t^{\frac{1}{2}} \right) \bar{z} \bar{u}_{i-1}^{\frac{1}{2}} u_i^{\frac{1}{2}} + u_i^{\frac{1}{2}} \bar{u}_{i-1}^{\frac{1}{2}} \bar{z} \left( \frac{d}{dt}|_{t=0} \bar{\gamma}_t^{\frac{1}{2}} \right) u_i^{\frac{1}{2}}. \quad (108)$$

Introducing the notation

$$X := \frac{d}{dt}|_{t=0} \bar{\gamma}_t^{\frac{1}{2}}, \quad T := u_i^{\frac{1}{2}} X \bar{z} \bar{u}_{i-1}^{\frac{1}{2}} u_i^{\frac{1}{2}}, \quad (109)$$

the derivative  $\frac{d}{dt}|_{t=0} B_t^n$  in (108) may be rewritten as

$$\frac{d}{dt}|_{t=0} B_t^n = T + T^\top. \quad (110)$$

We express  $X$  more explicitly now. To that end, let  $f(A) = A^{1/2}$  be the matrix square root function (which is unambiguously defined for positive matrices). We have, by the inverse function theorem, that  $df_{A^2}(Z) = Y$ , where  $AY + YA = Z$ , meaning that, at the point  $A$ , the directional derivative of  $f$  in direction  $Z$  is given by the solution  $Y$  of the right-hand Sylvester equation. Hence, in order to get  $\frac{d}{dt}|_{t=0} \bar{\gamma}_t^{\frac{1}{2}}$ , we have to solve the following Sylvester equation for  $X$ ,

$$\bar{\gamma}_0^{\frac{1}{2}} X + X \bar{\gamma}_0^{\frac{1}{2}} = \bar{v} \quad \text{where} \quad \bar{v} := u_i^{-\frac{1}{2}} v u_i^{-\frac{1}{2}}. \quad (111)$$

Summing up,

$$X = \frac{d}{dt}|_{t=0} \bar{\gamma}_t^{\frac{1}{2}} \quad \text{is the solution of} \quad \bar{u}_{i-1}^{\frac{1}{2}} X + X \bar{u}_{i-1}^{\frac{1}{2}} = \bar{v}. \quad (112)$$

Plugging (112) and (109) together with (107) into (104) shows that  $\frac{D}{dt}|_{t=0} B_t^n = (S - \frac{1}{2}T) + (S - \frac{1}{2}T)^\top$  which completes the proof.  $\square$