

# A Second Order Non-Smooth Variational Model for Restoring Manifold-Valued Images

Miroslav Bačák\*    Ronny Bergmann†    Gabriele Steidl†  
 Andreas Weinmann‡

October 26, 2016

## Abstract

We introduce a new non-smooth variational model for the restoration of manifold-valued data which includes second order differences in the regularization term. While such models were successfully applied for real-valued images, we introduce the second order difference and the corresponding variational models for manifold data, which up to now only existed for cyclic data. The approach requires a combination of techniques from numerical analysis, convex optimization and differential geometry. First, we establish a suitable definition of absolute second order differences for signals and images with values in a manifold. Employing this definition, we introduce a variational denoising model based on first and second order differences in the manifold setup. In order to minimize the corresponding functional, we develop an algorithm using an inexact cyclic proximal point algorithm. We propose an efficient strategy for the computation of the corresponding proximal mappings in symmetric spaces utilizing the machinery of Jacobi fields. For the  $n$ -sphere and the manifold of symmetric positive definite matrices, we demonstrate the performance of our algorithm in practice. We prove the convergence of the proposed exact and inexact variant of the cyclic proximal point algorithm in Hadamard spaces. These results which are of interest on its own include, e.g., the manifold of symmetric positive definite matrices.

**Keywords.** manifold-valued data, second order differences, TV-like methods on manifolds, non-smooth variational methods, Jacobi fields, Hadamard spaces, proximal mappings, DT-MRI

## 1. Introduction

In this paper, we introduce a non-smooth variational model for the restoration of manifold-valued images using first and second order differences. The model can be seen as a second order generalization of the Rudin-Osher-Fatemi (ROF) functional [60] for images taking

---

\*Max Planck Institute for Mathematics in the Sciences, Inselstr. 22, 04103 Leipzig, Germany, [bachak@mis.mpg.de](mailto:bachak@mis.mpg.de)

†Department of Mathematics, Technische Universität Kaiserslautern, Paul-Ehrlich-Str. 31, 67663 Kaiserslautern, Germany, [\{bergmann, steidl\}@mathematik.uni-kl.de](mailto:\{bergmann, steidl\}@mathematik.uni-kl.de).

‡Department of Mathematics, Technische Universität München and Fast Algorithms for Biomedical Imaging Group, Helmholtz-Zentrum München, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany, [andreas.weinmann@tum.de](mailto:andreas.weinmann@tum.de).

their values in a Riemannian manifold. For scalar-valued images, the ROF functional in its discrete, anisotropic penalized form is given by

$$\frac{1}{2}\|f - u\|_2^2 + \lambda\|\nabla u\|_1, \quad \lambda > 0,$$

where  $f \in \mathbb{R}^{N,M}$  is a given noisy image and the symbol  $\nabla$  is used to denote the discrete first order difference operator which usually contains the forward differences in vertical and horizontal directions. The frequently used ROF denoising model preserves important image structures as edges, but tends to produce staircasing: instead of reconstructing smooth areas as such, the reconstruction consists of constant plateaus with small jumps. An approach for avoiding this effect incorporates higher order differences, respectively derivatives, in a continuous setting. The pioneering work [15] couples the TV term with higher order terms by infimal convolution. Since then, various techniques with higher order differences/derivatives were proposed in the literature, among them [13, 17, 21, 22, 36, 43, 45, 46, 50, 61, 62, 63]. We further note that the second-order total generalized variation was extended for tensor fields in [72].

In various applications in image processing and computer vision the functions of interest take values in a Riemannian manifold. One example is diffusion tensor imaging where the data lives in the Riemannian manifold of positive definite matrices; see, e.g., [7, 14, 53, 65, 75, 78]. Other examples are color images based on non-flat color models [16, 40, 41, 73] where the data lives on spheres. Motion group and  $SO(3)$ -valued data play a role in tracking, robotics and (scene) motion analysis and were considered, e.g., in [28, 52, 55, 58, 70]. Because of the natural appearance of such nonlinear data spaces, processing manifold-valued data has gained a lot of interest in applied mathematics in recent years. As examples, we mention wavelet-type multiscale transforms [35, 54, 76], robust principal component pursuit on manifolds [37], and partial differential equations [19, 32, 69] for manifold-valued functions. Although statistics on Riemannian manifolds is not in the focus of this work, we want to mention that, in recent years, there are many papers on this topic.

In [30, 31], the notion of total variation of functions having their values on a manifold was investigated based on the theory of Cartesian currents. These papers extend the previous work [29] where circle-valued functions were considered. The first work which applies a TV approach of circle-valued data for image processing tasks is [66, 67]. An algorithm for TV regularized minimization problems on Riemannian manifolds was proposed in [44]. There, the problem is reformulated as a multilabel optimization problem which is approached using convex relaxation techniques. Another approach to TV minimization for manifold-valued data which employs cyclic and parallel proximal point algorithms and does not require labeling and relaxation techniques was given in [77]. In a recent approach [34] the restoration of manifold-valued images was done using a smoothed TV model and an iteratively reweighted least squares technique. A method which circumvents the direct work with manifold-valued data by embedding the matrix manifold in the appropriate Euclidean space and applying a back projection to the manifold was suggested in [59]. This can be also extended to higher order derivatives since the derivatives (or differences) are computed in the Euclidean space.

Our paper is the next step in a program already consisting of a considerable body of work of the authors: in [77], variational models using first order differences for general manifold-valued data were developed. In [9], variational models using first and second order differences for circle-valued data were introduced. Using a suitable definition of second order

differences on the circle  $\mathbb{S}^1$  the authors incorporate higher order differences into the energy functionals to improve the denoising results for circle-valued data. Furthermore, convergence for locally nearby data is shown. Our paper [11] extends this approach to product spaces of arbitrarily many circles and a vector space, and [10] to inpainting problems. Product spaces are important for example when dealing with nonlinear color spaces such as HSV.

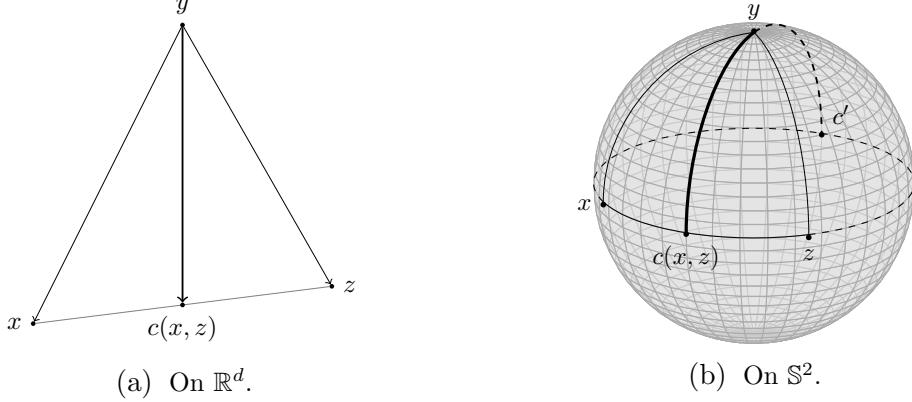
This paper continues our recent work considerably by generalizing the combined first and second order variational models to general symmetric Riemannian manifolds. Besides cyclic data this includes general  $n$ -spheres, hyperbolic spaces, symmetric positive definite matrices as well as compact Lie groups and Grassmannians. First we provide a novel definition of absolute second order differences for data with values in a manifold. The definition is geometric and particularly appealing since it avoids using the tangent bundle for its definition. As a result, it is computationally accessible by the machinery of Jacobi fields which, in particular, in symmetric spaces yields rather explicit descriptions – even in this generality. Employing this definition, we introduce a variational model for denoising based on first and second order differences in the Riemannian manifold setup. In order to minimize the corresponding functional, we follow [9, 11, 77] and use a cyclic proximal point algorithm (PPA). In contrast to the aforementioned references, in our general setup, no closed form expressions are available for some of the proximal mappings involved. Therefore, we use as approximate strategy, a subgradient descent to compute them. For this purpose, we derive an efficient scheme. We show the convergence of the proposed exact and inexact variant of the cyclic PPA in a Hadamard space. This extends a result from [4], where the exact cyclic PPA in Hadamard spaces was proved to converge under more restrictive assumptions. Note that the basic (batch) version of the PPA in Hadamard spaces was introduced in [3]. Another related result is due to S. Banert [6], who developed both exact and inexact PPA for a regularized sum of two functions on a product of Hadamard spaces. In the context of Hadamard manifolds, the convergence of an inexact proximal point method for multivalued vector fields was studied in [74].

In this paper we prove the convergence of the (inexact) cyclic PPA under the general assumptions required by our model which differs from the cited papers.

Our convergence statements apply in particular to the manifold of symmetric positive definite matrices. Finally, we demonstrate the performance of our algorithm in numerical experiments for denoising of images with values in spheres as well as in the space of symmetric positive definite matrices.

Our main application examples, namely  $n$ -spheres and manifolds of symmetric positive definite matrices are, with respect to the sectional curvature, two extreme instances of symmetric spaces. The spheres have positive constant curvature, whereas the symmetric positive definite matrices are non-positively curved. Their geometry is totally different, e.g., in the manifolds of symmetric positive definite matrices the triangles are slim and there are no cut locus which means that geodesics are always shortest paths. In  $n$ -spheres however, every geodesic meets a cut point and triangles are always fat, meaning that the sum of the interior angles is always bigger than  $\pi$ . In our setup, however, it turns out that the sign of the sectional curvature is not important, but the important thing is the structure provided by symmetric spaces.

The outline of the paper is as follows: We start by introducing our variational restoration model in Section 2. In Section 3 we show how the (sub)gradients of the second order difference operators can be computed. Interestingly, this can be done by solving appropriate



**Figure 1.** Illustration of the absolute second order difference on (a) the Euclidean space  $\mathbb{R}^n$  and (b) the sphere  $\mathbb{S}^2$ . In both cases the second order difference is the length of the line connecting  $y$  and  $c(x, z)$ . Nevertheless on  $\mathbb{S}^2$  there is a second minimizer  $c'$ , which is the mid point of the longer arc of the great circle defined by  $x$  and  $z$ .

Jacobi equations. We describe the computation for general symmetric spaces. Then we focus on  $n$ -spheres and the space of symmetric positive definite matrices. The (sub)gradients are needed within our inexact cyclic PPA which is proposed in Section 4. A convergence analysis of the exact and inexact cyclic PPA is given for Hadamard manifolds. In Section 5 we validate our model and illustrate the good performance of our algorithms by numerical examples. The appendix provides some useful formulas for the computations. Further, Appendix C gives a brief introduction into the concept of parallel transport on manifolds in order to make our results better accessible for non-experts in differential geometry.

## 2. Variational Model

Let  $\mathcal{M}$  be a complete  $n$ -dimensional Riemannian manifold with Riemannian metric  $\langle \cdot, \cdot \rangle_x: T_x \mathcal{M} \times T_x \mathcal{M} \rightarrow \mathbb{R}$ , induced norm  $\|\cdot\|_x$ , and geodesic distance  $d_{\mathcal{M}}: \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}_{\geq 0}$ . Let  $\nabla_{\mathcal{M}} F$  denote the Riemannian gradient of  $F: \mathcal{M} \rightarrow \mathbb{R}$  which is characterized for all  $\xi \in T_x \mathcal{M}$  by

$$\langle \nabla_{\mathcal{M}} F(x), \xi \rangle_x = D_x F[\xi], \quad (1)$$

where  $D_x F$  denotes the differential of  $F$  at  $x$ , see Appendix C.

Let  $\gamma_{x,\xi}(t)$ ,  $x \in \mathcal{M}$ ,  $\xi \in T_x \mathcal{M}$  be the unique geodesic starting from  $\gamma_{x,\xi}(0) = x$  with  $\dot{\gamma}_{x,\xi}(0) = \xi$ . Further let  $\tilde{\gamma}_{x,z}$  denote a unit speed geodesic connecting  $x, z \in \mathcal{M}$ . Then it fulfills  $\tilde{\gamma}_{x,z}(0) = x$ ,  $\tilde{\gamma}_{x,z}(\mathcal{L}) = z$ , where  $\mathcal{L} = \mathcal{L}(\tilde{\gamma}_{x,z})$  denotes the length of the geodesic. We further denote by  $\gamma_{x,z}$  a minimizing geodesic, i.e. a geodesic having minimal length  $\mathcal{L}(\tilde{\gamma}_{x,z}) = d_{\mathcal{M}}(x, z)$ . If it is clear from the context, we write geodesic instead of minimizing geodesic, but keep the notation of using  $\tilde{\gamma}_{x,z}$  when referring to all geodesics including the non-minimizing ones. We use the exponential map  $\exp_x: T_x \mathcal{M} \rightarrow \mathcal{M}$  given by  $\exp_x \xi = \gamma_{x,\xi}(1)$  and the inverse exponential map denoted by  $\log_x = \exp_x^{-1}: \mathcal{M} \rightarrow T_x \mathcal{M}$ . The core of our restoration model are absolute second order differences of points lying in a manifold. In the following we define such

differences in a sound way. The basic idea is based on rewriting the Euclidean norm of componentwise second order differences in  $\mathbb{R}^n$  as  $\|x - 2y + z\|_2 = 2\|\frac{1}{2}(x+z) - y\|_2$ , see Fig. 1(a). We define the set of midpoints between  $x, z \in \mathcal{M}$  as

$$\mathcal{C}_{x,z} := \{c \in \mathcal{M} : c = \tilde{\gamma}_{\widehat{x},z}(\frac{1}{2}\mathcal{L}(\tilde{\gamma}_{\widehat{x},z})) \text{ for any geodesic } \tilde{\gamma}_{\widehat{x},z}\}$$

and the *absolute second difference operator*  $d_2: \mathcal{M}^3 \rightarrow \mathbb{R}_{\geq 0}$  by

$$d_2(x, y, z) := \min_{c \in \mathcal{C}_{x,z}} d_{\mathcal{M}}(c, y), \quad x, y, z \in \mathcal{M}. \quad (2)$$

The definition is illustrated for  $\mathcal{M} = \mathbb{S}^2$  in Fig. 1(b). For the manifold  $\mathcal{M} = \mathbb{S}^1$ , definition (2) coincides, up to the factor  $\frac{1}{2}$ , with those of the absolute second order differences in [9]. Similarly we define the *second order mixed differences*  $d_{1,1}$  based on  $\|w - x + y - z\|_2 = 2\|\frac{1}{2}(w+y) - \frac{1}{2}(x+z)\|_2$  as

$$d_{1,1}(w, x, y, z) := \min_{c \in \mathcal{C}_{w,y}, \tilde{c} \in \mathcal{C}_{x,z}} d_{\mathcal{M}}(c, \tilde{c}), \quad x, y, z, w \in \mathcal{M}.$$

Let  $\mathcal{G} := \{1, \dots, N\} \times \{1, \dots, M\}$ . We want to denoise manifold-valued images  $f: \mathcal{G} \rightarrow \mathcal{M}$  by minimizing functionals of the form

$$\mathcal{E}(u) = \mathcal{E}(u, f) := F(u; f) + \alpha \operatorname{TV}_1(u) + \beta \operatorname{TV}_2(u) \quad (3)$$

where

$$\begin{aligned} F(u; f) &:= \frac{1}{2} \sum_{i,j=1}^{N,M} d_{\mathcal{M}}(f_{i,j}, u_{i,j})^2, \\ \alpha \operatorname{TV}_1(u) &:= \alpha_1 \sum_{i,j=1}^{N-1,M} d_{\mathcal{M}}(u_{i,j}, u_{i+1,j}) + \alpha_2 \sum_{i,j=1}^{N,M-1} d_{\mathcal{M}}(u_{i,j}, u_{i,j+1}), \\ \beta \operatorname{TV}_2(u) &:= \beta_1 \sum_{i=2,j=1}^{N-1,M} d_2(u_{i-1,j}, u_{i,j}, u_{i+1,j}) + \beta_2 \sum_{i=1,j=2}^{N,M-1} d_2(u_{i,j-1}, u_{i,j}, u_{i,j+1}) \\ &\quad + \beta_3 \sum_{i,j=1}^{N-1,M-1} d_{1,1}(u_{i,j}, u_{i,j+1}, u_{i+1,j}, u_{i+1,j+1}). \end{aligned}$$

For minimizing the functional we want to apply a cyclic PPA [4, 12]. This algorithm sequentially computes the proximal mappings of the summands involved in the functional. While the proximal mappings of the summands in data term  $F(u; f)$  and in the first regularization term  $\operatorname{TV}_1(u)$  are known analytically, see [27] and [77], respectively, the proximal mappings of  $d_2: \mathcal{M}^3 \rightarrow \mathbb{R}_{\geq 0}$  and  $d_{1,1}: \mathcal{M}^4 \rightarrow \mathbb{R}_{\geq 0}$  are only known analytically in the special case  $\mathcal{M} = \mathbb{S}^1$ , see [9]. In the following section we deal with the computation of the proximal mapping of  $d_2$ . The difference  $d_{1,1}$  can be treated in a similar way.

### 3. Subgradients of Second Order Differences on $\mathcal{M}$

Since we work in a Riemannian manifold, it is necessary to impose an assumption that guarantees that the involved points do not take pathological (practically irrelevant)

constellations to make the following derivations meaningful. In particular, we assume in this section that there is exactly one shortest geodesic joining  $x, z$ , i.e.,  $x$  is not a cut point of  $z$ ; cf. [23]. We note that this is no severe restriction since such points form a set of measure zero. Moreover, we restrict our attention to the case where the minimizer in (2) is taken for the corresponding geodesic midpoint which we denote by  $c(x, z)$ .

We want to compute the proximal mapping of  $d_2: \mathcal{M}^3 \rightarrow \mathbb{R}_{\geq 0}$  by a (sub)gradient descent algorithm. This requires the computation of the (sub)gradient of  $d_2$  which is done in the following subsections. For  $c(x, z) \neq y$  the subgradient of  $d_2$  coincides with its gradient

$$\nabla_{\mathcal{M}^3} d_2 = (\nabla_{\mathcal{M}} d_2(\cdot, y, z), \nabla_{\mathcal{M}} d_2(x, \cdot, z), \nabla_{\mathcal{M}} d_2(x, y, \cdot))^T. \quad (4)$$

If  $c(x, z) = y$ , then  $d_2$  is not differentiable. However, we will characterize the subgradients in Remark 3.4. In particular, the zero vector is a subgradient which is used in our subgradient descent algorithm.

### 3.1. Gradients of the Components of Second Order Differences

We start with the computation of the second component of the gradient (4). In general we have for  $d_{\mathcal{M}}(\cdot, y): \mathcal{M} \rightarrow \mathbb{R}_{\geq 0}$ ,  $x \mapsto d_{\mathcal{M}}(x, y)$ , see [68], that

$$\nabla_{\mathcal{M}} d_{\mathcal{M}}^2(x, y) = -2 \log_x y, \quad \nabla_{\mathcal{M}} d_{\mathcal{M}}(x, y) = -\frac{\log_x y}{\|\log_x y\|_x}, \quad x \neq y. \quad (5)$$

**Lemma 3.1.** *The second component of  $\nabla_{\mathcal{M}^3} d_2$  in (4) is given for  $y \neq c(x, z)$  by*

$$\nabla_{\mathcal{M}} d_2(x, \cdot, z)(y) = \frac{\log_y c(x, z)}{\|\log_y c(x, z)\|_y}.$$

*Proof.* Applying (5) for  $d_2(x, \cdot, z) = d_{\mathcal{M}}(c(x, z), \cdot)$  we obtain the assertion.  $\square$

By the symmetry of  $c(x, z)$  both gradients  $\nabla_{\mathcal{M}} d_2(\cdot, y, z)$  and  $\nabla_{\mathcal{M}} d_2(x, y, \cdot)$  can be realized in the same way so that we can restrict our attention to the first one. For fixed  $z \in \mathcal{M}$  we will use the notation  $c(x)$  instead of  $c(x, z)$ . Let  $\gamma_{\widehat{x,z}} = \gamma_{x,v}$  denote the unit speed geodesic joining  $x$  and  $z$ , i.e.,

$$\gamma_{x,v}(0) = x, \quad \dot{\gamma}_{x,v}(0) = v = \frac{\log_x z}{\|\log_x z\|_x} \quad \text{and} \quad \gamma_{x,v}(T) = z, \quad T := d_{\mathcal{M}}(x, z).$$

We will further need the notation of parallel transport. For readers which are not familiar with this concept we give a brief introduction in Appendix C. We denote a parallel transported orthonormal frame along  $\gamma_{x,v}$  by

$$\{\Xi_1 = \Xi_1(t), \dots, \Xi_n = \Xi_n(t)\}. \quad (6)$$

For  $t = 0$  we use the special notation  $\{\xi_1, \dots, \xi_n\}$ .

**Lemma 3.2.** *The first component of  $\nabla_{\mathcal{M}^3} d_2$  in (4) is given for  $c(x, z) \neq y$  by*

$$\nabla_{\mathcal{M}} d_2(\cdot, y, z)(x) = \sum_{k=1}^n \left\langle \frac{\log_{c(x)} y}{\|\log_{c(x)} y\|_{c(x)}}, D_x c[\xi_k] \right\rangle_{c(x)} \xi_k. \quad (7)$$

*Proof.* For  $F: \mathcal{M} \rightarrow \mathbb{R}$  defined by  $F := d_2(\cdot, y, z)$  we are looking for the coefficients  $a_k = a_k(x)$  in

$$\nabla_{\mathcal{M}} F(x) = \sum_{k=1}^n a_k \xi_k. \quad (8)$$

For any tangential vector  $\eta := \sum_{k=1}^n \eta_k \xi_k \in T_x \mathcal{M}$  we have

$$\langle \nabla_{\mathcal{M}} F(x), \eta \rangle_x = D_x F[\eta] = \sum_{k=1}^n \eta_k a_k. \quad (9)$$

Since  $F = f \circ c$ , with  $f: \mathcal{M} \rightarrow \mathbb{R}$ ,  $x \mapsto d_{\mathcal{M}}(x, y)$  we obtain by the chain rule

$$D_x F[\eta] = (D_{c(x)} f \circ D_x c)[\eta].$$

Now the differential of  $c$  is determined by

$$D_x c[\eta] = \sum_{k=1}^n \eta_k D_x c[\xi_k] \in T_{c(x)} \mathcal{M}.$$

Then it follows by (1) and (5) that

$$\begin{aligned} D_x F[\eta] &= D_{c(x)} f[D_x c[\eta]] = \langle \nabla_{\mathcal{M}} f(c(x)), D_x c[\eta] \rangle_{c(x)} \\ &= \left\langle -\frac{\log_{c(x)} y}{\|\log_{c(x)} y\|_{c(x)}}, \sum_{k=1}^n \eta_k D_x c[\xi_k] \right\rangle_{c(x)}, \end{aligned}$$

and consequently

$$\langle \nabla_{\mathcal{M}} F(x), \eta \rangle_x = \sum_{k=1}^n \eta_k \left\langle -\frac{\log_{c(x)} y}{\|\log_{c(x)} y\|_{c(x)}}, D_x c[\xi_k] \right\rangle_{c(x)}.$$

By (9) and (8) we obtain the assertion (7).  $\square$

For the computation of  $D_x c[\xi_k]$  we can exploit Jacobi fields which are defined as follows: for  $s \in (-\varepsilon, \varepsilon)$ , let  $\sigma_{x, \xi_k}(s)$ ,  $k = 1, \dots, n$ , denote the unit speed geodesic with  $\sigma_{x, \xi_k}(0) = x$  and  $\sigma'_{x, \xi_k}(0) = \xi_k$ . Let  $\zeta_k(s)$  denote the tangential vector in  $\sigma_{x, \xi_k}(s)$  of the geodesic joining  $\sigma_{x, \xi_k}(s)$  and  $z$ . Then

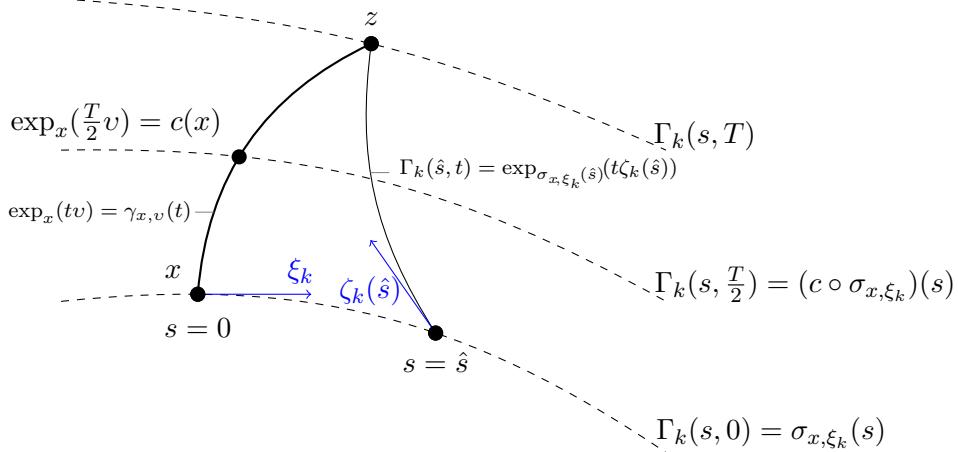
$$\Gamma_k(s, t) := \exp_{\sigma_{x, \xi_k}(s)}(t \zeta_k(s)), \quad s \in (-\varepsilon, \varepsilon), t \in [0, T],$$

with small  $\varepsilon > 0$  are variations of the geodesic  $\gamma_{x, v}$  and

$$J_k(t) := \frac{\partial}{\partial s} \Gamma_k(s, t) \Big|_{s=0}, \quad k = 1, \dots, n,$$

are the corresponding Jacobi field along  $\gamma_{x, v}$ . For an illustration see Fig. 2. Since  $\Gamma_k(s, 0) = \sigma_{x, \xi_k}(s)$  and  $\Gamma_k(s, T) = z$  for all  $s \in (-\varepsilon, \varepsilon)$  we have

$$J_k(0) = \xi_k, \quad J_k(T) = 0, \quad k = 1, \dots, n. \quad (10)$$



**Figure 2.** Illustration of the variations  $\Gamma_k(s, t)$  of the geodesic  $\gamma_{x,v}$  to define the Jacobi field  $J_k(t) = \frac{\partial}{\partial s} \Gamma_k(s, t)|_{s=0}$  along  $\gamma_{x,v}$  with respect to  $\xi_k$ .

Since  $\Gamma_k(s, \frac{T}{2}) = (c \circ \sigma_{x, \xi_k})(s)$  we conclude by definition (37) of the differential

$$D_x c[\xi_k] = \frac{d}{ds} (c \circ \sigma_{x, \xi_k})(s)|_{s=0} = \frac{d}{ds} \Gamma_k(s, \frac{T}{2})|_{s=0} = J_k(\frac{T}{2}), \quad k = 1, \dots, n.$$

Any Jacobi field  $J$  of a variation through  $\gamma_{x,v}$  fulfills a linear system of ordinary differential equations (ODE) [42, Theorem 10.2]

$$\frac{D^2}{dt^2} J + R(J, \dot{\gamma}_{x,v}) \dot{\gamma}_{x,v} = 0, \quad (11)$$

where  $R$  denotes the Riemannian curvature tensor defined by  $(\xi, \zeta, \eta) \mapsto R(\xi, \zeta)\eta := \nabla_\xi \nabla_\zeta \eta - \nabla_\zeta \nabla_\xi \eta - \nabla_{[\xi, \zeta]} \eta$ . Here  $[\cdot, \cdot]$  is the Lie bracket, see Appendix C. Our special Jacobi fields have to meet the boundary conditions (10). We summarize:

**Lemma 3.3.** *The vectors  $D_x c[\xi_k]$ ,  $k = 1, \dots, n$ , in (7) are given by*

$$D_x c[\xi_k] = J_k(\frac{T}{2}), \quad k = 1, \dots, n,$$

where  $J_k$  are the Jacobi fields given by

$$\frac{D^2}{dt^2} J_k + R(J_k, \dot{\gamma}_{x,v}) \dot{\gamma}_{x,v} = 0, \quad J_k(0) = \xi_k, \quad J_k(T) = 0.$$

Finally, we give a representation of the subgradients of  $d_2$  in the case  $c(x, z) = y$ .

**Remark 3.4.** *Let  $(x, y, z) \in \mathcal{M}^3$  with  $c(x, z) = y$  and*

$$\mathcal{V} := \left\{ \xi := (\xi_x, \xi_y, \xi_z) : \xi_x \in T_x \mathcal{M}, \xi_y = J_{\xi_x, \xi_z}(\frac{T}{2}), \xi_z \in T_z \mathcal{M} \right\},$$

where  $J_{\xi_x, \xi_z}$  is the Jacobi field along the unit speed geodesic  $\gamma_{x,z}$  determined by  $J(0) = \xi_x$  and  $J(T) = \xi_z$ . Then the subdifferential  $\partial d_2$  at  $(x, y, z) \in \mathcal{M}^3$  reads

$$\partial d_2(x, y, z) = \left\{ \alpha \eta : \eta \in \mathcal{V}^\perp, \alpha \in [l_\eta, u_\eta] \right\}, \quad (12)$$

where  $\mathcal{V}^\perp$  denotes the set of normalized vectors  $\eta := (\eta_x, \eta_y, \eta_z) \in T_x\mathcal{M} \times T_y\mathcal{M} \times T_z\mathcal{M}$  fulfilling

$$\langle \xi_x, \eta_x \rangle_x + \langle \xi_y, \eta_y \rangle_y + \langle \xi_z, \eta_z \rangle_z = 0$$

for all  $(\xi_x, \xi_y, \xi_z) \in \mathcal{V}$ , and the interval endpoints  $l_\eta, u_\eta$  are given by

$$\begin{aligned} l_\eta &= \lim_{\tau \uparrow 0} \frac{d_2(\exp_x(\tau\eta_x), \exp_y(\tau\eta_y), \exp_z(\tau\eta_z))}{\tau}, \\ u_\eta &= \lim_{\tau \downarrow 0} \frac{d_2(\exp_x(\tau\eta_x), \exp_y(\tau\eta_y), \exp_z(\tau\eta_z))}{\tau}. \end{aligned} \quad (13)$$

This can be seen as follows: For arbitrary tangent vectors  $\xi_x, \xi_z$  sitting in  $x, z$ , respectively, we consider the uniquely determined geodesic variation  $\Gamma(s, t)$  given by the side conditions  $\Gamma(0, t) = \gamma_{x,z}(t)$ ,  $\Gamma(s, 0) = \exp_x(s\xi_x)$  as well as  $\Gamma(s, T) = \exp_z(s\xi_z)$ . We note that  $c(\Gamma(s, 0), \Gamma(s, T)) = \Gamma(s, T/2)$  which implies for  $s \in (-\varepsilon, \varepsilon)$  that

$$d_2(\Gamma(s, 0), \Gamma(s, T/2), \Gamma(s, T)) = 0. \quad (14)$$

In view of the definition of a subgradient, see [26, 33], it is required, for a candidate  $\eta$ , that

$$d_2(\exp_x(h_x), \exp_y(h_y), \exp_z(h_z)) - d_2(x, y, z) \geq \langle h, \eta \rangle + o(h). \quad (15)$$

for any sufficiently small  $h := (h_x, h_y, h_z)$ . Setting  $h := (\xi_x, J_{\xi_x, \xi_z}(T/2), \xi_z)$ , equation (14) tells us that the left hand side above equals 0 up to  $o(h)$ , and thus, for a candidate  $\eta$ ,

$$\langle \xi_x, \eta_x \rangle_x + \langle J_{\xi_x, \xi_z}(\frac{T}{2}), \eta_y \rangle_y + \langle \xi_z, \eta_z \rangle_z = o(h)$$

for all  $\xi_x, \xi_y$  of small magnitude. Since these are actually linear equations for  $w = (\eta_x, \eta_y, \eta_z)$  in the tangent space, we get

$$\langle \xi_x, \eta_x \rangle_x + \langle J_{\xi_x, \xi_z}(\frac{T}{2}), \eta_y \rangle_y + \langle \xi_z, \eta_z \rangle_z = 0.$$

Then, if  $\eta \in \mathcal{V}^\perp$ , the nominators in (13) are nonzero and the limits exist. Finally, we conclude (12) from (15).

In the following subsection we recall how Jacobi fields can be computed for general symmetric spaces and have a look at two special examples, namely  $n$ -spheres and manifolds of symmetric positive definite matrices.

### 3.2. Jacobi Equation for Symmetric Spaces

Due to their rich structure symmetric spaces have been the object of differential geometric studies for a long time, and we refer to [24, 25] or the books [8, 18] for more information.

A Riemannian manifold  $\mathcal{M}$  is called *locally symmetric* if the geodesic reflection  $s_x$  at each point  $x \in \mathcal{M}$  given by mapping  $\gamma(t) \mapsto \gamma(-t)$  for all geodesics  $\gamma$  through  $x = \gamma(0)$  is a local isometry, i.e., an isometry at least locally near  $x$ . If this property holds globally,  $\mathcal{M}$  is called a (*Riemannian globally*) *symmetric space*. More formally,  $\mathcal{M}$  is a symmetric space if for any  $x \in \mathcal{M}$  and all  $\xi \in T_x\mathcal{S}$  there is an isometry  $s_x$

on  $\mathcal{M}$  such that  $s_x(x) = x$  and  $D_x s_x[\xi] = -\xi$ . A Riemannian manifold  $\mathcal{M}$  is locally symmetric if and only if there exists a symmetric space which is locally isometric to  $\mathcal{M}$ . As a consequence of the Cartan–Ambrose–Hicks theorem [18, Theorem 1.36], every simply connected, complete, locally symmetric space is symmetric. Symmetric spaces are precisely the homogeneous spaces with a symmetry  $s_x$  at some point  $x \in \mathcal{M}$ . Beyond  $n$ -spheres and the manifold of symmetric positive definite matrices, hyperbolic spaces, Grassmannians as well as compact Lie groups are examples of symmetric spaces. The crucial property we need is that a Riemannian manifolds is locally symmetric if and only if the covariant derivative of the Riemannian curvature tensor  $R$  along curves is zero, i.e.,

$$\nabla R = 0. \quad (16)$$

**Proposition 3.5.** *Let  $\mathcal{M}$  be a symmetric space. Let  $\gamma: [0, T] \rightarrow \mathcal{M}$  be a unit speed geodesic and  $\{\Theta_1 = \Theta_1(t), \dots, \Theta_n = \Theta_n(t)\}$  a parallel transported orthonormal frame along  $\gamma$ . Let  $J(t) = \sum_{i=1}^n a_i(t) \Theta_i(t)$  be a Jacobi field of a variation through  $\gamma$ . Set  $a := (a_1, \dots, a_n)^T$ . Then the following relations hold true:*

i) *The Jacobi equation (11) can be written as*

$$a''(t) + G a(t) = 0, \quad (17)$$

*with the constant coefficient matrix  $G := (\langle R(\Theta_i, \dot{\gamma})\dot{\gamma}, \Theta_j \rangle_\gamma)_{i,j=1}^n$ .*

ii) *Let  $\{\theta_1, \dots, \theta_n\}$  be chosen as the initial orthonormal basis which diagonalizes the operator*

$$\Theta \mapsto R(\Theta, \dot{\gamma})\dot{\gamma} \quad (18)$$

*at  $t = 0$  with corresponding eigenvalues  $\kappa_i$ ,  $i = 1, \dots, n$ , and let  $\{\Theta_1, \dots, \Theta_n\}$  be the corresponding parallel transported frame along  $\gamma$ . Then the matrix  $G$  becomes diagonal and (17) decomposes into the  $n$  ordinary linear differential equations*

$$a_i''(t) + \kappa_i a_i(t) = 0, \quad i = 1, \dots, n.$$

iii) *The Jacobi fields*

$$J_k(t) := \begin{cases} \sinh(\sqrt{-\kappa_k} t) \Theta_k(t), & \text{if } \kappa_k < 0, \\ \sin(\sqrt{\kappa_k} t) \Theta_k(t), & \text{if } \kappa_k > 0, \\ t \Theta_k(t), & \text{if } \kappa_k = 0, \end{cases}$$

$k = 1, \dots, n$  form a basis of the  $n$  dimensional linear space of Jacobi fields of a variation through  $\gamma$  fulfilling the initial condition  $J(0) = 0$ .

Part i) of Proposition 3.5 is also stated as Property A in Rauch's paper [56] in the particularly nice form "The curvature of a 2-section propagated parallel along a geodesic is constant." For convenience we add the proof.

*Proof.* i) Using the frame representation of  $J$ , (38) and the linearity of  $R$  in the first argument, the Jacobi equation (11) becomes

$$0 = \sum_{i=1}^n a_i''(t)\Theta_i(t) + a_i(t)R(\Theta_i, \dot{\gamma})\dot{\gamma},$$

and by taking inner products with  $\Theta_j$  further

$$0 = a_j''(t) + \langle R(\Theta_i, \dot{\gamma})\dot{\gamma}, \Theta_j \rangle a_i(t), \quad j = 1, \dots, n.$$

Now (16) implies for  $R(\Theta_i, \dot{\gamma})\dot{\gamma} = \sum_{k=1}^n r_{ik}(t)\Theta_k(t)$

$$0 = \nabla_{\dot{\gamma}} R = \sum_{k=1}^n r'_{ik}(t)\Theta_k(t)$$

which is, by the linear independence of the  $\Theta_k$ ,  $k = 1, \dots, n$ , only possible if all  $r_{ik}$  are constants and we get  $G = (r_{ij})_{i,j=1}^n$ .

Parts ii) and iii) follow directly from i). For iii) we also refer to [18, p. 77].  $\square$

With respect to our special Jacobi fields in Lemma 3.3 we obtain the following corollary.

**Corollary 3.6.** *The Jacobi fields  $J_k$ ,  $k = 1, \dots, n$  of a variation through  $\gamma_{x,z} = \gamma_{x,v}$  with boundary conditions  $J_k(0) = \xi_k$  and  $J_k(T) = 0$  fulfill*

$$J_k\left(\frac{T}{2}\right) = \begin{cases} \frac{\sinh(\sqrt{-\kappa_k}\frac{T}{2})}{\sinh(\sqrt{-\kappa_k}T)}\Xi_k\left(\frac{T}{2}\right), & \text{if } \kappa_k < 0, \\ \frac{\sin(\sqrt{\kappa_k}\frac{T}{2})}{\sin(\sqrt{\kappa_k}T)}\Xi_k\left(\frac{T}{2}\right), & \text{if } \kappa_k > 0, \\ \frac{1}{2}\Xi_k\left(\frac{T}{2}\right), & \text{if } \kappa_k = 0. \end{cases} \quad (19)$$

*Proof.* The Jacobi fields  $\bar{J}_k(t) := \alpha_k J_k(T-t)$  of a variation through  $\gamma_{z,x} := \gamma_{x,v}(T-t)$  satisfy  $J_k(0) = 0$  and by Proposition 3.5 iii) they are given by

$$\bar{J}_k(t) := \begin{cases} \sinh(\sqrt{-\kappa_k}t)\Xi_k(T-t), & \text{if } \kappa_k < 0, \\ \sin(\sqrt{\kappa_k}t)\Xi_k(T-t), & \text{if } \kappa_k > 0, \\ t\Xi_k(T-t), & \text{if } \kappa_k = 0. \end{cases}$$

In particular we have

$$\bar{J}_k(T) := \begin{cases} \sinh(\sqrt{-\kappa_k}T)\xi_k, & \text{if } \kappa_k < 0, \\ \sin(\sqrt{\kappa_k}T)\xi_k, & \text{if } \kappa_k > 0, \\ T\xi_k, & \text{if } \kappa_k = 0. \end{cases}$$

Now  $\alpha_k\xi_k = \alpha_k J_k(0) = \bar{J}_k(T)$  determines  $\alpha_k$  as

$$\alpha_k = \begin{cases} \sinh(\sqrt{-\kappa_k}T), & \text{if } \kappa_k < 0, \\ \sin(\sqrt{\kappa_k}T), & \text{if } \kappa_k > 0, \\ T, & \text{if } \kappa_k = 0 \end{cases}$$

and  $J_k(t) = \frac{1}{\alpha_k} \bar{J}_k(T-t)$ . We notice that the denominators of the appearing fractions are nonzero since  $x$  and  $z$  were assumed to be non-conjugate points. Finally, we get (19).  $\square$

Let us apply our findings for the  $n$ -sphere and the manifold of symmetric positive definite matrices.

**The Sphere  $\mathbb{S}^n$ .** We consider the  $n$ -sphere  $\mathbb{S}^n$ . Then, the Riemannian metric is just the Euclidean distance  $\langle \cdot, \cdot \rangle_x = \langle \cdot, \cdot \rangle$  in  $\mathbb{R}^{n+1}$ . For the definitions of the geodesic distance, the exponential map and parallel transport see Appendix A. Let  $\gamma := \gamma_{x,v}$ . We choose  $\xi_1 := v = \dot{\gamma}(0)$  and complete this to an orthonormal basis  $\{\xi_1, \dots, \xi_n\}$  of  $T_x \mathbb{S}^n$  with corresponding parallel frame  $\{\Xi_1, \dots, \Xi_n\}$  along  $\gamma$ . Then diagonalizing the operator (18) is especially simple. Since  $\mathbb{S}^n$  has constant curvature  $C = 1$ , the Riemannian curvature tensor fulfills [42, Lemma 8.10]  $R(\Theta, \Xi)\Upsilon = \langle \Xi, \Upsilon \rangle \Theta - \langle \Theta, \Upsilon \rangle \Xi$ . Consequently,

$$R(\Theta, \dot{\gamma})\dot{\gamma} = \langle \dot{\gamma}, \dot{\gamma} \rangle \Theta - \langle \Theta, \dot{\gamma} \rangle \dot{\gamma} = \Theta - \langle \Theta, \dot{\gamma} \rangle \dot{\gamma}$$

so that at  $t = 0$  the vector  $\xi_1$  is an eigenvector with eigenvalue  $\kappa_1 = 0$  and  $\xi_i$ ,  $i = 2, \dots, n$ , are eigenvectors with eigenvalues  $\kappa_i = 1$ . Consequently, we obtain by Lemma 3.3 and Corollary 3.6 the following corollary.

**Corollary 3.7.** *For the sphere  $\mathbb{S}^n$  and the above choice of the orthonormal frame system, the following relations hold true:*

$$D_x c[\xi_1] = \frac{1}{2} \Xi_1\left(\frac{T}{2}\right), \quad D_x c[\xi_k] = \frac{\sin \frac{T}{2}}{\sin T} \Xi_k\left(\frac{T}{2}\right), \quad k = 2, \dots, n.$$

**Symmetric Positive Definite Matrices.** Let  $\text{Sym}(r)$  denote the space of symmetric  $r \times r$  matrices with (Frobenius) inner product and norm

$$\langle A, B \rangle := \sum_{i,j=1}^r a_{ij} b_{ij}, \quad \|A\| := \left( \sum_{i,j=1}^r a_{ij}^2 \right)^{\frac{1}{2}}. \quad (20)$$

Let  $\mathcal{P}(r)$  be the manifold of symmetric positive definite  $r \times r$  matrices. It has the dimension  $\dim \mathcal{P}(r) = n = \frac{r(r+1)}{2}$ . The tangent space of  $\mathcal{P}(r)$  at  $x \in \mathcal{P}(r)$  is given by  $T_x \mathcal{P}(r) = \{x\} \times \text{Sym}(r) = \{x^{\frac{1}{2}} \eta x^{\frac{1}{2}} : \eta \in \text{Sym}(r)\}$ , in particular  $T_I \mathcal{P}(r) = \text{Sym}(r)$ , where  $I$  denotes the  $r \times r$  identity matrix. The Riemannian metric on  $T_x \mathcal{P}$  reads

$$\langle \eta_1, \eta_2 \rangle_x := \text{tr}(\eta_1 x^{-1} \eta_2 x^{-1}) = \langle x^{-\frac{1}{2}} \eta_1 x^{-\frac{1}{2}}, x^{-\frac{1}{2}} \eta_2 x^{-\frac{1}{2}} \rangle,$$

where  $\langle \cdot, \cdot \rangle$  denotes the matrix inner product (20). For the definitions of the geodesic distance, exponential map, parallel transport see the Appendix B.

Let  $\gamma := \gamma_{x,v}$  and let the matrix  $v \in T_x \mathcal{P}$  have the eigenvalues  $\lambda_1, \dots, \lambda_r$  with a corresponding orthonormal basis of eigenvectors  $v_1, \dots, v_r$  in  $\mathbb{R}^r$ , i.e.,

$$v = \sum_{i=1}^r \lambda_i v_i v_i^T.$$

We will use a more appropriate index system for the frame (6), namely

$$\mathcal{I} := \{(i, j) : i = 1, \dots, r; j = i, \dots, r\}.$$

Then the matrices

$$\xi_{ij} := \begin{cases} \frac{1}{2}(v_i v_j^T + v_j v_i^T), & (i, j) \in \mathcal{I} \quad \text{if } i = j, \\ \frac{1}{\sqrt{2}}(v_i v_j^T + v_j v_i^T), & (i, j) \in \mathcal{I} \quad \text{if } i \neq j, \end{cases} \quad (21)$$

form an orthonormal basis of  $T_x \mathcal{P}(r)$ .

In other words, we will deal with the parallel transported frame  $\Xi_{ij}$ ,  $(i, j) \in \mathcal{I}$ , of (21) instead of  $\Xi_k$ ,  $k = 1, \dots, n$ . To diagonalize the operator (18) at  $t = 0$  we use that the Riemannian curvature tensor for  $\mathcal{P}(r)$  has the form

$$R(\Theta, \Xi)\Upsilon = -\frac{1}{4}x^{\frac{1}{2}} \left[ [x^{-\frac{1}{2}}\Theta x^{-\frac{1}{2}}, x^{-\frac{1}{2}}\Xi x^{-\frac{1}{2}}], x^{-\frac{1}{2}}\Upsilon x^{-\frac{1}{2}} \right] x^{\frac{1}{2}}$$

with the Lie bracket  $[A, B] = AB - BA$  of matrices. Then

$$R(\Theta, \dot{\gamma})\dot{\gamma} = -\frac{1}{4}x^{\frac{1}{2}} \left[ [x^{-\frac{1}{2}}\Theta x^{-\frac{1}{2}}, x^{-\frac{1}{2}}\dot{\gamma} x^{-\frac{1}{2}}], x^{-\frac{1}{2}}\dot{\gamma} x^{-\frac{1}{2}} \right] x^{\frac{1}{2}}$$

and for  $t = 0$  with  $\theta = \Theta(0)$  the right-hand side becomes

$$\begin{aligned} T(\theta) &= -\frac{1}{4}x^{\frac{1}{2}} \left[ [x^{-\frac{1}{2}}\theta x^{-\frac{1}{2}}, x^{-\frac{1}{2}}vx^{-\frac{1}{2}}], x^{-\frac{1}{2}}vx^{-\frac{1}{2}} \right] x^{\frac{1}{2}} \\ &= -\frac{1}{4}x^{\frac{1}{2}}(wb^2 - 2bw + b^2w)x^{\frac{1}{2}}, \end{aligned}$$

where  $b := x^{-\frac{1}{2}}vx^{-\frac{1}{2}}$  and  $w := x^{-\frac{1}{2}}\theta x^{-\frac{1}{2}}$ . Expanding  $\theta = \sum_{(i,j) \in \mathcal{I}} \mu_{ij}\xi_{ij}$  into the orthonormal basis of  $T_x \mathcal{P}$  and substituting this into  $T(\theta)$  gives after a straightforward computation

$$T(\theta) = \sum_{(i,j) \in \mathcal{I}} \mu_{ij}(\lambda_i - \lambda_j)^2 \xi_{ij}.$$

Thus  $\{\xi_{ij} : (i, j) \in \mathcal{I}\}$  is an orthonormal basis of eigenvectors of  $T$  with corresponding eigenvalues

$$\kappa_{ij} = -\frac{1}{4}(\lambda_i - \lambda_j)^2, \quad (i, j) \in \mathcal{I}.$$

Let  $\mathcal{I}_1 := \{(i, j) \in \mathcal{I} : \lambda_i = \lambda_j\}$  and  $\mathcal{I}_2 := \{(i, j) \in \mathcal{I} : \lambda_i \neq \lambda_j\}$ . Then, by Lemma 3.3 and Corollary 3.6, we get the following corollary.

**Corollary 3.8.** *For the manifold of symmetric positive definite matrices  $\mathcal{P}(r)$  it holds*

$$D_x c[\xi_{ij}] = \begin{cases} \frac{1}{2}\Xi_{ij}(\frac{T}{2}) & \text{if } (i, j) \in \mathcal{I}_1, \\ \frac{\sinh(\frac{T}{4}|\lambda_i - \lambda_j|)}{\sinh(\frac{T}{2}|\lambda_i - \lambda_j|)} \Xi_{ij}(\frac{T}{2}) & \text{if } (i, j) \in \mathcal{I}_2. \end{cases}$$

## 4. Inexact Cyclic Proximal Point Algorithm

In order to minimize the functional in (3), we follow the approach in [9] and employ a cyclic proximal point algorithm (cyclic PPA).

For a proper, closed, convex function  $\phi: \mathbb{R}^m \rightarrow (-\infty, +\infty]$  and  $\lambda > 0$  the *proximal mapping*  $\text{prox}_{\lambda\phi}: \mathbb{R}^m \rightarrow \mathbb{R}^m$  at  $x \in \mathbb{R}^m$  is defined by

$$\text{prox}_{\lambda\phi}(x) := \arg \min_{y \in \mathbb{R}^m} \left\{ \frac{1}{2\lambda} \|x - y\|_2^2 + \phi(y) \right\},$$

see [49]. The above minimizer exists and is uniquely determined. Many algorithms which were recently used in variational image processing reduce to the iterative computation of values of proximal mappings. An overview of applications of proximal mappings is given in [51].

Proximal mappings were generalized for functions on Riemannian manifolds in [27], replacing the squared Euclidean norm by the squared geodesic distances. For  $\phi: \mathcal{M}^m \rightarrow (-\infty, +\infty]$  and  $\lambda > 0$  let

$$\text{prox}_{\lambda\phi}(x) := \arg \min_{y \in \mathcal{M}^m} \left\{ \frac{1}{2\lambda} \sum_{j=1}^m d_{\mathcal{M}}(x_j, y_j)^2 + \phi(y) \right\}. \quad (22)$$

For proper, closed, convex functions  $\phi$  on Hadamard manifolds the minimizer exists and is uniquely determined. More generally, one can define proximal mappings in certain metric spaces. In particular, such a definition was given independently in [38] and [47] for Hadamard spaces, which was later on used for the PPA [3] and cyclic PPA [4].

### 4.1. Algorithm

We split the functional in (3) into the summands

$$\mathcal{E} = \sum_{l=1}^{15} \mathcal{E}_l, \quad (23)$$

where  $\mathcal{E}_1(u) := F(u; f)$  and

$$\begin{aligned} \alpha \text{TV}_1(u) &= \sum_{\nu_1=0}^1 \alpha_1 \sum_{i,j=1}^{\lfloor \frac{N-1}{2} \rfloor, M} d_{\mathcal{M}}(u_{2i-1+\nu_1, j}, u_{2i+\nu_1, j}) \\ &\quad + \sum_{\nu_2=0}^1 \alpha_2 \sum_{i,j=1}^{N, \lfloor \frac{M-1}{2} \rfloor} d_{\mathcal{M}}(u_{i, 2j-1+\nu_2}, u_{i, 2j+\nu_2}) \\ &=: \sum_{\nu_1=0}^1 \mathcal{E}_{2+\nu_1}(u) + \sum_{\nu_2=0}^1 \mathcal{E}_{4+\nu_2}(u) \end{aligned}$$

and

$$\begin{aligned}
& \beta \operatorname{TV}_2(u) \\
&= \sum_{\nu_1=0}^2 \beta_1 \sum_{i,j=1}^{\left[\frac{N-1}{3}\right], M} d_2(u_{3i-2+\nu_1,j}, u_{3i-1+\nu_1,j}, u_{3i+\nu_1}) \\
&\quad + \sum_{\nu_2=0}^2 \beta_2 \sum_{i,j=1}^{N, \left[\frac{M-1}{3}\right]} d_2(u_{i,3j-2+\nu_2}, u_{i,3j-1+\nu_2}, u_{i,3j+\nu_2}) \\
&\quad + \sum_{\nu_3, \nu_4=0}^1 \beta_3 \sum_{i,j=1}^{\left[\frac{N-1}{2}\right], \left[\frac{M-1}{2}\right]} d_{1,1}(u_{2i-1+\nu_3, 2j-1+\nu_4}, u_{2i+\nu_3, 2j-1+\nu_4}, u_{2i-1+\nu_3, 2j+\nu_4}, u_{2i+\nu_3, 2j+\nu_4}) \\
&=: \sum_{\nu_1=0}^2 \mathcal{E}_{6+\nu_1}(u) + \sum_{\nu_1=0}^2 \mathcal{E}_{9+\nu_1}(u) + \sum_{\nu_3, \nu_4=0}^1 \mathcal{E}_{12+\nu_3+2\nu_4}(u).
\end{aligned}$$

Then the exact cyclic PPA computes starting with  $u^{(0)} = f$  until a convergence criterion is reached the values

$$u^{(k+1)} := \operatorname{prox}_{\lambda_k \mathcal{E}_{15}} \circ \operatorname{prox}_{\lambda_k \mathcal{E}_{14}} \circ \dots \circ \operatorname{prox}_{\lambda_k \mathcal{E}_1}(u^{(k)}) \quad (24)$$

where the parameters  $\lambda_k > 0$  in the  $k$ -th cycle have to fulfill

$$\sum_{k=0}^{\infty} \lambda_k = \infty, \quad \text{and} \quad \sum_{k=0}^{\infty} \lambda_k^2 < \infty. \quad (25)$$

By construction, the functional  $\mathcal{E}_l$ ,  $l \in \{1, \dots, 15\}$  in (24), contains every entry of  $u$  at most once. Hence the involved proximal mappings of  $\operatorname{prox}_{\lambda \mathcal{E}_l}$  consists of can be evaluated by computing all involved proximal mappings, one for every summand, in parallel, i.e. for

(D0)  $d_{\mathcal{M}}^2(u_{ij}, f_{ij})$  of the data fidelity term,

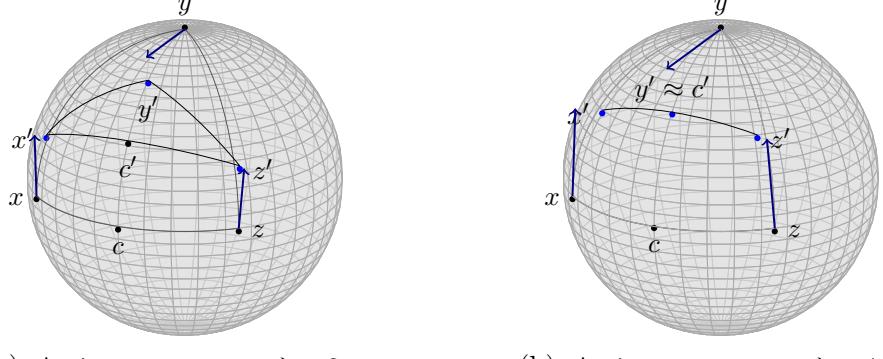
(D1)  $\alpha_1 d_{\mathcal{M}}(u_{2i-1+\nu_1,j}, u_{2i-\nu_1,j})$ ,  $\alpha_2 d_{\mathcal{M}}(u_{i,2j-1+\nu_2}, u_{i,2j-\nu_2})$  of the first order differences,

(D2)  $\beta_1 d_2(u_{3i-2+\nu_1,j}, u_{3i-1+\nu_1,j}, u_{3i+\nu_1})$ ,  $\beta_2 d_2(u_{i,3j-2+\nu_2}, u_{i,3j-1+\nu_2}, u_{i,3j+\nu_2})$  of the second order differences, and  $\beta_3 d_{1,1}(u_{2i-1+\nu_3, 2j-1+\nu_4}, u_{2i+\nu_3, 2j-1+\nu_4}, u_{2i-1+\nu_3, 2j+\nu_4}, u_{2i+\nu_3, 2j+\nu_4})$  of the second order mixed differences.

Taking these as the functions  $\phi$  which are of interest in (22) we can reduce our attention to  $m = 1, 2$  and  $m = 3, 4$ , respectively. Analytical expressions for the minimizers defining the proximal mappings, for the data fidelity terms (D0) are given in [27], and for the first order differences (D1) in [77]. For the second order difference in (D2) such expressions are only available for the manifold  $\mathcal{M} = \mathbb{S}^1$ , see [9].

In order to derive an approximate solution of

$$\operatorname{prox}_{\lambda d_2}(g_1, g_2, g_3) = \arg \min_{x \in \mathcal{M}^3} \left\{ \frac{1}{2} \sum_{j=1}^3 d_{\mathcal{M}}(x_j, g_j)^2 + \lambda d_2(x_1, x_2, x_3) \right\} =: \arg \min_{x \in \mathcal{M}^3} \psi(x)$$



(a) An inexact  $\text{prox}_{\lambda d_2}$ ,  $\lambda = 2\pi$ .

(b) An inexact  $\text{prox}_{\lambda d_2}$ ,  $\lambda = 4\pi$ .

**Figure 3.** Illustration of the inexact proximal mapping  $(x', y', z') = \text{prox}_{\lambda d_2}(x, y, z)$ . For all points the negative gradients are shown in dark blue.

---

**Algorithm 1** Subgradient Method for  $\text{prox}_{\lambda d_2}$

---

**Input** data  $g = (g_1, g_2, g_3) \in \mathcal{M}^3$ , a sequence  $\tau = \{\tau_k\}_k \in \ell_2 \setminus \ell_1$ .

**function** SUBGRADIENTPROXD2( $g, \tau$ )  
 Initialize  $x^{(0)} = g$ ,  $x^* = x^{(0)}$ ,  $k = 1$ .  
**repeat**  

$$x^{(k)} \leftarrow \exp_{x^{(k-1)}} \left( -\tau_k \nabla_{\mathcal{M}^3} \psi(x^{(k-1)}) \right)$$
**if**  $\psi(x^*; g) > \psi(x^{(k)}; g)$  **then**  $x^* \leftarrow x^{(k)}$   
 $k \leftarrow k + 1$   
**until** a convergence criterion is reached  
**return**  $x^*$

---

we employ the (sub)gradient descent method to  $\psi$ . For gradient descent methods on manifolds including convergence results we refer to [1, 71]. The subgradient method is one of the classical algorithms for nondifferentiable optimization which was extended for manifolds, e.g., in [26, 33]. In [26] convergence results for Hadamard manifolds were established. A subgradient method on manifolds is given in Algorithm 1. In particular, again restricting to the second order differences, we have to compute the gradient of  $\psi$ :

$$\nabla_{\mathcal{M}^3} \psi(x) = - \begin{pmatrix} \log_{x_1} g_1 \\ \log_{x_2} g_2 \\ \log_{x_3} g_3 \end{pmatrix} + \lambda \nabla_{\mathcal{M}^3} d_2(x_1, x_2, x_2), \quad c(x_1, x_3) \neq x_2.$$

The computation of  $\nabla_{\mathcal{M}^3} d_2$  was the topic of Section 3. A result of Algorithm 1 for the points already used in Fig. 1(b), the Fig. 3 illustrates the proximal mapping for two different values of  $\lambda$ .

In summary this means that we perform an inexact cyclic PPA as in Algorithm 2. We will prove the convergence of such an algorithm in the following subsection for Hadamard spaces.

---

**Algorithm 2** Inexact Cyclic PPA for minimizing (3)

---

**Input** data  $f \in \mathcal{M}^{N \times M}$ ,  $\alpha \in \mathbb{R}_{\geq 0}^2$ ,  $\beta \in \mathbb{R}_{\geq 0}^3$ , a sequence  $\lambda = \{\lambda_k\}_k$ ,  $\lambda_k > 0$ , fulfilling (25), and a sequence of positive reals  $\epsilon = \{\epsilon_k\}_k$  with  $\sum \epsilon_k < \infty$

**function** CPPA( $\alpha, \beta, \lambda, f$ )

Initialize  $u^{(0)} = f$ ,  $k = 0$

Initialize the cycle length as  $L = 15$  (or  $L = 6$  for the case  $M = 1$  ).

**repeat**

**for**  $l \leftarrow 1$  **to**  $L$  **do**

$u^{(k+\frac{l}{L})} \leftarrow \text{prox}_{\lambda_k \varphi_l}(u^{(k+\frac{l-1}{L})})$ ,

where the proximal operators are given analytically for (D0)

and (D1) as in [27, 77] and approximately for (D2) via Algorithm 1

and the error is bounded by  $\epsilon_k$

$k \leftarrow k + 1$

**until** a convergence criterion is reached

**return**  $u^{(k)}$

---

## 4.2. Convergence Analysis

We now present the convergence analysis of the above algorithms in the setting of Hadamard spaces, which include, for instance, the manifold of symmetric positive definite matrices. Recall that a complete metric space  $(X, d)$  is called Hadamard if every two points  $x, y$  are connected by a geodesic and the following condition holds true

$$d(x, v)^2 + d(y, w)^2 \leq d(x, w)^2 + d(y, v)^2 + 2d(x, y)d(v, w), \quad (26)$$

for any  $x, y, v, w \in X$ . Inequality (26) implies that Hadamard spaces have nonpositive curvature [2, 57] and Hadamard spaces are thus a natural generalization of complete simply connected Riemannian manifolds of nonpositive sectional curvature. For more details, the reader is referred to [5, 39].

In this subsection, let  $(\mathcal{H}, d)$  be a locally compact Hadamard space. We consider

$$\varphi = \sum_{l=1}^L \varphi_l, \quad (27)$$

where  $\varphi_l: \mathcal{H} \rightarrow \mathbb{R}$  are convex continuous functions and assume that  $\varphi$  attains a (global) minimum.

For Hadamard spaces  $\mathcal{H} := \mathcal{M}^N$ ,  $N = N \cdot M$ , the functional  $\varphi = \mathcal{E}$  in (23) fits into this setting with  $L = 15$ . Alternatively we may take the single differences in (D0)-(D2) as summands  $\varphi_l$ . Our aim is to show the convergence of the (inexact) cyclic PPA. To this end, recall that, given a metric space  $(X, d)$ , a mapping  $T: X \rightarrow X$  is nonexpansive if  $d(Tx, Ty) \leq d(x, y)$ . In the proof of Theorem 4.3, we shall need the following well known lemmas. Lemma 4.1 is a consequence of the strong convexity of a regularized convex function and expresses how much the function's value decreases after applying a single PPA step. Lemma 4.2 is a refinement of the fact that a bounded monotone sequence has a limit.

**Lemma 4.1** ([5, Lemma 2.2.23]). *If  $h: \mathcal{H} \rightarrow (-\infty, +\infty]$  is a convex lower semi-continuous function, then, for every  $x, y \in \mathcal{H}$ , we have*

$$h(\text{prox}_{\lambda h}(x)) - h(y) \leq \frac{1}{2\lambda} d(x, y)^2 - \frac{1}{2\lambda} d(\text{prox}_{\lambda h}(x), y)^2.$$

**Lemma 4.2.** *Let  $\{a_k\}_{k \in \mathbb{N}}$ ,  $\{b_k\}_{k \in \mathbb{N}}$ ,  $\{c_k\}_{k \in \mathbb{N}}$  and  $\{\eta_k\}_{k \in \mathbb{N}}$  be sequences of nonnegative real numbers. For each  $k \in \mathbb{N}$  assume*

$$a_{k+1} \leq (1 + \eta_k) a_k - b_k + c_k,$$

along with

$$\sum_{k=1}^{\infty} c_k < \infty \quad \text{and} \quad \sum_{k=1}^{\infty} \eta_k < \infty.$$

Then the sequence  $\{a_k\}_{k \in \mathbb{N}}$  converges and  $\sum_{k=1}^{\infty} b_k < \infty$ .

Let us start with the exact cyclic PPA. The following theorem generalizes [4, Theorem 3.4] in a way that is required for proving convergence for our setting. The point  $p$  in Theorem 4.3 is a reference point chosen arbitrarily. In linear spaces it is natural to take the origin. Condition (28) then determines how fast the functions  $\varphi_l$  can change their values across the space.

**Theorem 4.3** (Cyclic PPA). *Let  $(\mathcal{H}, d)$  be a locally compact Hadamard space and let  $\varphi$  in (27) have a global minimizer. Assume that there exist  $p \in \mathcal{H}$  and  $C > 0$  such that for each  $l = 1, \dots, L$  and all  $x, y \in \mathcal{H}$  we have*

$$\varphi_l(x) - \varphi_l(y) \leq Cd(x, y)(1 + d(x, p)). \quad (28)$$

Then the sequence  $\{x^{(k)}\}_{k \in \mathbb{N}}$  defined by the cyclic PPA

$$x^{(k+1)} := \text{prox}_{\lambda_k \varphi_L} \circ \text{prox}_{\lambda_{k-1} \varphi_{L-1}} \circ \dots \circ \text{prox}_{\lambda_1 \varphi_1}(x^{(k)})$$

with  $\{\lambda_k\}_{k \in \mathbb{N}}$  as in (25) converges for every starting point  $x^{(0)}$  to a minimizer of  $\varphi$ .

*Proof.* For  $l = 1, \dots, L$  we set

$$x^{(k+\frac{l}{L})} := \text{prox}_{\lambda_k \varphi_l}(x^{(k+\frac{l-1}{L})}).$$

1. First we prove that for any fixed  $q \in \mathcal{H}$  and all  $k \in \mathbb{N}_0$  there exists a constant  $C_q > 0$  such that

$$d(x^{(k+1)}, q)^2 \leq (1 + C_q \lambda_k^2) d(x^{(k)}, q)^2 - 2\lambda_k (\varphi(x^{(k)}) - \varphi(q)) + C_q \lambda_k^2. \quad (29)$$

For any fixed  $q \in \mathcal{H}$  we obtain by (28) and the triangle inequality

$$\varphi_l(x) - \varphi_l(y) \leq C_q d(x, y)(1 + d(x, q)), \quad C_q := 1 + d(q, p). \quad (30)$$

Applying Lemma 4.1 with  $h := \varphi_l$ ,  $x := x^{(k+\frac{l-1}{L})}$  and  $y := q$  we conclude

$$d(x^{(k+\frac{l}{L})}, q)^2 \leq d(x^{(k+\frac{l-1}{L})}, q)^2 - 2\lambda_k (\varphi_l(x^{(k+\frac{l}{L})}) - \varphi_l(q))$$

for  $l = 1, \dots, L$ . Summation yields

$$\begin{aligned} d(x^{(k+1)}, q)^2 &\leq d(x^{(k)}, q)^2 - 2\lambda_k \sum_{l=1}^L (\varphi_l(x^{(k+\frac{l}{L})}) - \varphi_l(q)) \\ &= d(x^{(k)}, q)^2 - 2\lambda_k (\varphi(x^{(k)}) - \varphi(q)) + 2\lambda_k \sum_{l=1}^L (\varphi_l(x^{(k)}) - \varphi_l(x^{(k+\frac{l}{L})})), \end{aligned} \quad (31)$$

where we used (27). The growth condition in (30) gives

$$\varphi_l(x^{(k)}) - \varphi_l(x^{(k+\frac{l}{L})}) \leq C_q d(x^{(k)}, x^{(k+\frac{l}{L})}) (1 + d(x^{(k)}, q)). \quad (32)$$

By the definition of the proximal mapping we have

$$\varphi_l(x^{(k+\frac{l}{L})}) + \frac{1}{2\lambda_k} d(x^{(k+\frac{l-1}{L})}, x^{(k+\frac{l}{L})})^2 \leq \varphi_l(x^{(k+\frac{l-1}{L})})$$

and by (30) further

$$\begin{aligned} d(x^{(k+\frac{l-1}{L})}, x^{(k+\frac{l}{L})}) &\leq 2\lambda_k \frac{\varphi_l(x^{(k+\frac{l-1}{L})}) - \varphi_l(x^{(k+\frac{l}{L})})}{d(x^{(k+\frac{l-1}{L})}, x^{(k+\frac{l}{L})})} \\ &\leq 2\lambda_k C_q (1 + d(x^{(k+\frac{l-1}{L})}, q)). \end{aligned} \quad (33)$$

for every  $l = 1, \dots, L$ . For  $l = 1$  this becomes

$$d(x^{(k)}, x^{(k+\frac{1}{L})}) \leq 2\lambda_k C_q (1 + d(x^{(k)}, q)), \quad (34)$$

and for  $l = 2$  using (34) and the triangle inequality

$$\begin{aligned} d(x^{(k+\frac{1}{L})}, x^{(k+\frac{2}{L})}) &\leq 2\lambda_k C_q (1 + d(x^{(k+\frac{1}{L})}, q)) \\ &\leq 2\lambda_k C_q (1 + 2\lambda_k C_q) (1 + d(x^{(k)}, q)). \end{aligned}$$

By (25) we can assume that  $\lambda_k < 1$ . Then replacing  $2C_q (1 + 2C_q)$  by a new constant which we call  $C_q$  again, we get

$$d(x^{(k+\frac{1}{L})}, x^{(k+\frac{2}{L})}) \leq \lambda_k C_q (1 + d(x^{(k)}, q)).$$

This argument can be applied recursively for  $l = 3, \dots, L$ . In the rest of the proof we will use  $C_q$  as a generic constant independent of  $\lambda_k$ . Using

$$d(x^{(k)}, x^{(k+\frac{l}{L})}) \leq d(x^{(k)}, x^{(k+\frac{1}{L})}) + \dots + d(x^{(k+\frac{l-1}{L})}, x^{(k+\frac{l}{L})})$$

we obtain

$$d(x^{(k)}, x^{(k+\frac{l}{L})}) \leq \lambda_k C_q (1 + d(x^{(k)}, q)),$$

for  $l = 1, \dots, L$ . Consequently we get by (32) that

$$\varphi_l(x^{(k)}) - \varphi_l(x^{(k+\frac{l}{L})}) \leq \lambda_k C_q \left(1 + d(x^{(k)}, q)^2\right),$$

for  $l = 1, \dots, L$ . Plugging this inequality into (31) yields

$$d(x^{(k+1)}, q)^2 \leq d(x^{(k)}, q)^2 - 2\lambda_k (\varphi(x^{(k)}) - \varphi(q)) + C_q \left(1 + d(x^{(k)}, q)^2\right)$$

which finishes the proof of (29).

2. Assume now that  $q \in \mathcal{H}$  is a minimizer of  $\varphi$  and apply Lemma 4.2 with  $a_k := d(x^{(k)}, q)^2$ ,  $b_k := 2\lambda_k (\varphi(x^{(k)}) - \varphi(q))$ ,  $c_k := C_q \lambda_k^2$  and  $\eta_k := C_q \lambda_k^2$  to conclude that the sequence  $\{d(x^{(k)}, q)\}_{k \in \mathbb{N}_0}$  converges and

$$\sum_{k=0}^{\infty} \lambda_k (\varphi(x^{(k)}) - \varphi(q)) < \infty. \quad (35)$$

In particular, the sequence  $\{x^{(k)}\}_{k \in \mathbb{N}}$  is bounded. From (35) and (25) we immediately obtain  $\min \varphi = \liminf_{k \rightarrow \infty} \varphi(x^{(k)})$ , and thus there exists a cluster point  $z \in \mathcal{H}$  of  $\{x^{(k)}\}_{k \in \mathbb{N}}$  which is a minimizer of  $\varphi$ . Now convergence of  $\{d(x^{(k)}, z)\}_{k \in \mathbb{N}_0}$  implies that  $\{x^{(k)}\}_{k \in \mathbb{N}}$  converges to  $z$  as  $k \rightarrow \infty$ . (By (33) we see moreover that  $\{x^{(k+\frac{l}{L})}\}_{k \in \mathbb{N}}$ ,  $l = 1, \dots, L$  converges to the same point.)  $\square$

Next we consider the inexact cyclic PPA which iteratively generates the points  $x^{(k+\frac{l}{L})}$ ,  $l = 1, \dots, L$ ,  $k \in \mathbb{N}_0$ , fulfilling

$$d(x^{(k+\frac{l}{L})}, \text{prox}_{\lambda_k \varphi_l}(x^{(k+\frac{l-1}{L})})) < \frac{\varepsilon_k}{L}, \quad (36)$$

where  $\{\varepsilon_k\}_{k \in \mathbb{N}_0}$  is a given sequence of positive reals.

**Theorem 4.4** (Inexact Cyclic PPA). *Let  $(\mathcal{H}, d)$  be a locally compact Hadamard space and let  $\varphi$  be given by (27). Assume that for every starting point, the sequence generated by the exact cyclic PPA converges to a minimizer of  $\varphi$ . Let  $\{x^{(k)}\}_{k \in \mathbb{N}}$  be the sequence generated by the inexact cyclic PPA in (36), where  $\sum_{k=0}^{\infty} \varepsilon_k < \infty$ . Then the sequence  $\{x^{(k)}\}_{k \in \mathbb{N}}$  converges to a minimizer of  $\varphi$ .*

We note that the assumptions for Theorem 4.4 are fulfilled if the assumptions of Theorem 4.3 are given.

*Proof.* For  $k, m \in \mathbb{N}_0$ , set

$$y_{m,k} := \begin{cases} x^{(k)} & \text{if } k \leq m, \\ T_{k-1}(x^{(k-1)}) & \text{if } k > m, \end{cases}$$

where

$$T_k := \text{prox}_{\lambda_k \varphi_L} \circ \dots \circ \text{prox}_{\lambda_k \varphi_1}.$$

Hence, for a fixed  $m \in \mathbb{N}_0$ , the sequence  $\{y_{m,k}\}_k$  is obtained by inexact computations until the  $m$ -th step and by exact computations from the step  $(m+1)$  on. In particular,

the sequence  $\{y_{0,k}\}_k$  is the exact cyclic PPA sequence and  $\{y_{k,k}\}_k$  the inexact cyclic PPA sequence. By assumption we know that, for a given  $m \in \mathbb{N}_0$ , the sequence  $\{y_{m,k}\}_k$  converges to minimizer  $y_m$  of  $\varphi$ . Next we observe that the fact  $\sum_{k=0}^{\infty} \varepsilon_k < \infty$  implies that the set  $\{y_{m,k} : k, m \in \mathbb{N}_0\}$  is bounded: Indeed, by our assumptions the sequence  $\{y_{0,k}\}_k$  converges and therefore lies in a bounded set  $D_0 \subset \mathcal{H}$ . By (36) and since the proximal mapping is nonexpansive, see [5, Theorem 2.2.22], we obtain

$$\begin{aligned} d(x^{(\frac{1}{L})}, \text{prox}_{\lambda_0 \varphi_1}(x^{(0)})) &\leq \frac{\varepsilon_0}{L}, \\ d\left(x^{(\frac{2}{L})}, \text{prox}_{\lambda_0 \varphi_2}(\text{prox}_{\lambda_0 \varphi_1}(x^{(0)}))\right) &\leq d(x^{(\frac{2}{L})}, \text{prox}_{\lambda_0 \varphi_2}(x^{(\frac{1}{L})})) \\ &\quad + d\left(\text{prox}_{\lambda_0 \varphi_2}(x^{(\frac{1}{L})}), \text{prox}_{\lambda_0 \varphi_2}(\text{prox}_{\lambda_0 \varphi_1}(x^{(0)}))\right) \\ &\leq \frac{\varepsilon_0}{L} + d(x^{(\frac{1}{L})}, \text{prox}_{\lambda_0 \varphi_1}(x^{(0)})) \leq \frac{2\varepsilon_0}{L} \end{aligned}$$

and using the argument recursively

$$d(x^{(1)}, T_0(x^{(0)})) \leq \varepsilon_0.$$

Hence  $\{y_{1,k}\}_k$  lies in a bounded set  $D_1 := \{x \in \mathcal{H} : d(x, D_0) < \varepsilon_0\}$ . The same argument yields that the sequence  $\{y_{m,k}\}_k$  with  $m \geq 1$  lies in a bounded set

$$D_m := \left\{ x \in \mathcal{H} : d(x, D_0) < \sum_{j=0}^{m-1} \varepsilon_j \right\}.$$

Finally the set  $\{y_{m,k} : k, m \in \mathbb{N}_0\}$  is contained in  $\{x \in \mathcal{H} : d(x, D_0) < \sum_{j=0}^{\infty} \varepsilon_j\}$ .

Consequently, also the sequence  $\{y_m\}_m$  is bounded and has at least one cluster point  $z$  which is also a minimizer of  $\varphi$ . Since the proximal mappings are nonexpansive, we have  $d(y_m, y_{m+1}) < \varepsilon_m$ . Using again the fact that  $\sum_{j=0}^{\infty} \varepsilon_j < \infty$ , we obtain that the sequence  $\{y_m\}_m$  cannot have two different cluster points and therefore  $\lim_{m \rightarrow \infty} y_m = z$ .

Next we will show that the sequence  $\{x^{(k)}\}_k$  converges to  $z$  as  $k \rightarrow \infty$ . To this end, choose  $\delta > 0$  and find  $m_1 \in \mathbb{N}$  such that  $\sum_{j=m_1}^{\infty} \varepsilon_j < \frac{\delta}{3}$  and  $d(y_{m_1}, z) < \frac{\delta}{3}$ . Next find  $m_2 > m_1$  such that whenever  $m > m_2$  we have

$$d(y_{m_1, m}, y_{m_1}) < \frac{\delta}{3}.$$

Since the proximal mappings is nonexpansive, we get

$$d(y_{m_1, m}, y_{m, m}) < \sum_{j=m_1}^{m-1} \varepsilon_j < \sum_{j=m_1}^{\infty} \varepsilon_j < \frac{\delta}{3}.$$

Finally, the triangle inequality gives

$$d(z, y_{m, m}) < d(z, y_{m_1}) + d(y_{m_1}, y_{m_1, m}) + d(y_{m_1, m}, y_{m, m}) < \frac{\delta}{3} + \frac{\delta}{3} + \frac{\delta}{3} = \delta$$

and the proof is complete. (For each  $l = 1, \dots, L$  the sequence  $\{x^{(k+\frac{l}{L})}\}_k$  has the same limit as  $\{x^{(k)}\}_k$ .)  $\square$

**Remark 4.5.** Note that the condition  $\sum_{k=0}^{\infty} \varepsilon_k < \infty$  is necessary. Indeed, let  $C := \{(x, 0) \in \mathbb{R}^2 : x \in \mathbb{R}\}$  and let  $\varphi := d(\cdot, C)$ . Then one can easily see that an inexact PPA sequence with errors  $\varepsilon_k$  satisfying  $\sum_{k=0}^{\infty} \varepsilon_k = \infty$  does not converge.

**Remark 4.6.** While the theory of convergence for the inexact proximal point algorithm, especially the convergence Theorem 4.4 is valid, we noticed that some functions in our splitting from Section 4.1 do not fulfill the assumptions of the theorem. For the convergence of Algorithm 2 claimed in Corollary 4.6 of the former arXiv version, all involved functions  $\varphi_l$  have to be geodesically convex. Unfortunately the second order differences  $d_2$  are not jointly convex in their three arguments as the following discussion shows. Let  $\mathcal{M}$  be a finite dimensional Hadamard manifold.

- i) Let  $x(t) := \gamma_{x_1, x_2}(t)$  and  $z(t) := \gamma_{z_1, z_2}(t)$  be two geodesics connecting  $x_1, x_2$  and  $z_1, z_2$  respectively, and  $c(t) := \gamma_{x(t), z(t)}\left(\frac{1}{2}\right)$  the midpoint function. In particular we have  $c(0) = \gamma_{x_1, z_1}\left(\frac{1}{2}\right)$  and  $c(1) = \gamma_{x_2, z_2}\left(\frac{1}{2}\right)$ . In general this midpoint function does not coincide with the geodesic  $\gamma_c := \gamma_{c(0), c(1)}$ . Take for example the Poincaré disc  $\mathcal{M} := \mathbb{D}$  and

$$x_1 := \begin{pmatrix} \sin \frac{\pi}{3} \tanh 1 \\ \cos \frac{\pi}{3} \tanh 1 \end{pmatrix}, \quad x_2 := \begin{pmatrix} -\sin \frac{\pi}{3} \tanh 1 \\ \cos \frac{\pi}{3} \tanh 1 \end{pmatrix},$$

$$z_1 := \begin{pmatrix} \sin \frac{\pi}{4} \tanh \frac{1}{2} \\ \cos \frac{\pi}{4} \tanh \frac{1}{2} \end{pmatrix}, \quad z_2 := \begin{pmatrix} -\sin \frac{\pi}{4} \tanh \frac{1}{2} \\ \cos \frac{\pi}{4} \tanh \frac{1}{2} \end{pmatrix}.$$

The computed the mid point curve  $c$  and the geodesic  $\gamma_c$  are depicted Figure 4.

- ii) The second order difference  $f(x, y, z) := d_2(x, y, z)$  is in general not (jointly) convex. To this end, we use the example in i) and consider the second order difference function along  $x(t), y(t) := \gamma_c(t)$  and  $z(t)$ ,  $t \in [0, 1]$ . Since the mid point curve is not the geodesic, there exists a point  $t_0 \in [0, 1]$  such that

$$d_{\mathcal{M}}(c(t_0), \gamma_c(t_0)) > 0.$$

Then we get

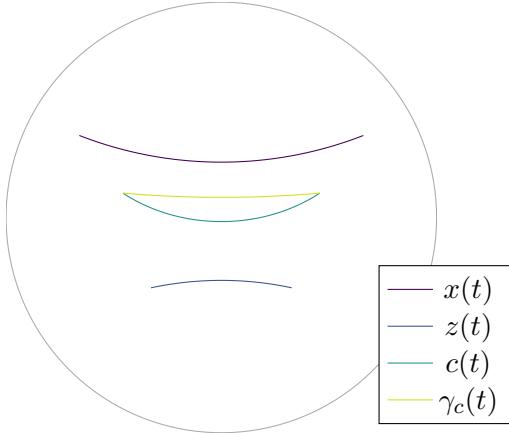
$$f(x(t_0), y(t_0), z(t_0)) = d_{\mathcal{M}}(c(t_0), y(t_0)) = d_{\mathcal{M}}(c(t_0), \gamma_c(t_0)) > 0$$

and

$$(1 - t_0)f(x(0), y(0), z(0)) + t_0f(x(1), y(1), z(1)) = (1 - t_0)d_{\mathcal{M}}(c(x(0), z(0)), \gamma_c(0)) + t_0d_{\mathcal{M}}(c(x(1), z(1)), \gamma_c(1)) = 0$$

so that  $f$  is not convex.

**Remark 4.7** (Random PPA). Instead of considering the cyclic PPA in Theorem 4.4, one can study an inexact version of the random PPA, generalizing hence [4, Theorem 3.7]. This would rely on the supermartingale convergence theorem and yield the almost sure convergence of the inexact PPA sequence. We however choose to focus on the cyclic variant and develop its inexact version, because it is appropriate for our applications.



**Figure 4.** The geodesics  $x(t)$ ,  $z(t)$  (violet) on  $\mathbb{D}$ , its mid point curve  $c$  (cyan) and the geodesic  $\gamma_c$  (light green) from the above example i) from Remark 4.6. The curves  $c$  and  $\gamma_c$  do not coincide.

## 5. Numerical Examples

Algorithm 2 was implemented in MATLAB and C++ with the Eigen library\* for both the sphere  $\mathbb{S}^2$  and the manifold of symmetric positive definite matrices  $\mathcal{P}(3)$  employing the subgradient method from Algorithm 1. In the latter algorithm we choose 0 from the subdifferential whenever it is multi-valued. Furthermore a suitable choice for the sequences in Algorithms 2 and 1 is  $\lambda := \{\frac{\lambda_0}{k}\}_k$ ,  $\lambda_0 = \frac{\pi}{2}$ , and  $\tau := \{\frac{\tau_0}{j}\}_j$ ,  $\tau_0 = \lambda_k$ , respectively. The parameters in our model (3) were chosen as  $\alpha := \alpha_1 = \alpha_2$  and  $\beta := \beta_1 = \beta_2 = \beta_3$  with an example depending grid search for an optimal choice. The experiments were conducted on a MacBook Pro running Mac OS X 10.10.3, Core i5, 2.6 GHz with 8 GB RAM using MATLAB 2015a, Eigen 3.2.4 and the clang-602.0.49 compiler. For all experiments we set the convergence criterion to 1000 iterations for one-dimensional signals and to 400 iterations for images. This yields the same number of proximal mapping applied to each point, because we have  $L = 15$  for the two-dimensional case and  $L = 6$  in one dimension. To measure quality, we look at the mean error

$$E(x, y) = \frac{1}{|\mathcal{G}|} \sum_{i \in \mathcal{G}} d_{\mathcal{M}}(x_i, y_i)$$

for two signals or images of manifold-valued data  $\{x_i\}_{i \in \mathcal{G}}, \{y_i\}_{i \in \mathcal{G}}$  defined on an index set  $\mathcal{G}$ .

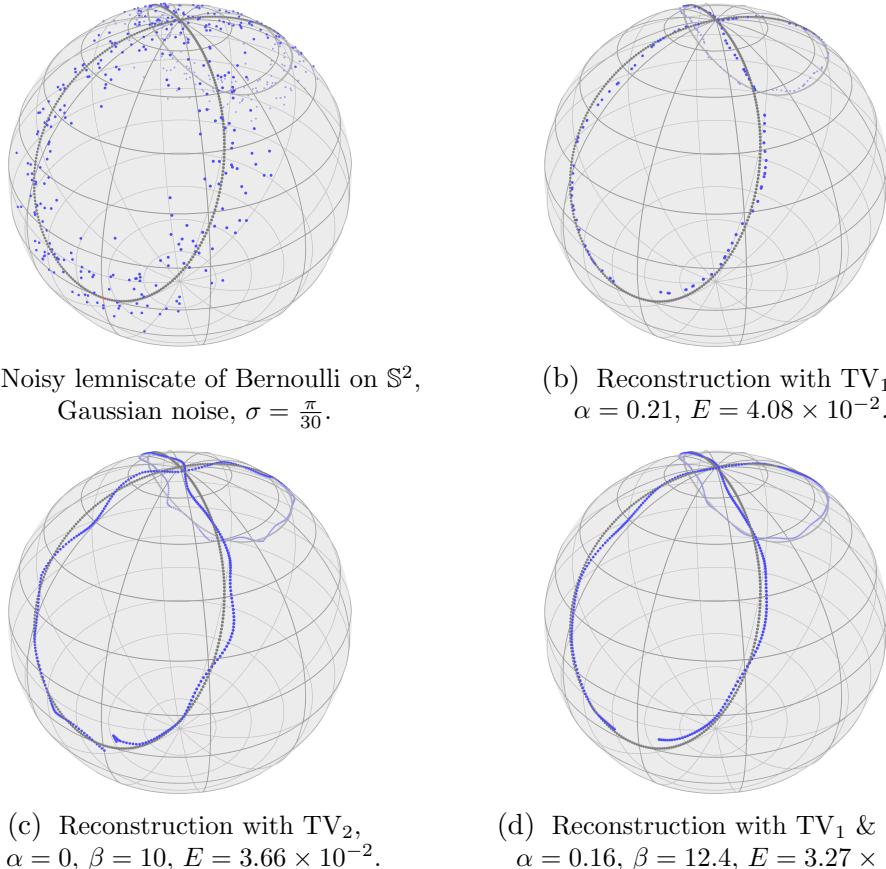
### 5.1. $\mathbb{S}^2$ -valued Data

**Sphere-Valued Signal.** As first example we take a curve on the sphere  $\mathbb{S}^2$ . For any  $a > 0$  the lemniscate of Bernoulli is defined as

$$\gamma(t) := \frac{a\sqrt{2}}{\sin^2(t) + 1} (\cos(t), \cos(t)\sin(t))^T, \quad t \in [0, 2\pi].$$

---

\*available at <http://eigen.tuxfamily.org>



**Figure 5.** Denoising an obstructed lemniscate of Bernoulli on the sphere  $\mathbb{S}^2$ . Combining first and second order differences yields the minimal value with respect to  $E(f_o, u_r)$ .

To obtain a curve on the sphere, we take an arbitrary point  $p \in \mathbb{S}^2$  and define the spherical lemniscate curve by

$$\gamma_S(t) = \log_p(\gamma(t))$$

Setting  $a = \frac{\pi}{2\sqrt{2}}$ , both extremal points of the lemniscate are antipodal, cf. the dotted gray line in Fig. 5(a). We sample the spherical lemniscate curve for  $p = (0, 0, 1)^T$  at  $t_i := \frac{2\pi i}{511}$ ,  $i = 0, \dots, 511$ , to obtain a signal  $(f_{o,i})_{i=0}^{511} \in (\mathbb{S}^2)^{512}$ . Note that the first and last point are identical. We colored them in red in Fig. 5(a), where the curve starts counterclockwise, i.e., to the right. This signal is affected by an additive Gaussian noise by setting  $f_i := \exp_{f_{o,i}} \eta_i$  with  $\eta$  having standard deviation of  $\sigma = \frac{\pi}{30}$  independently in both components. We obtain, e.g., the blue signal  $f = (f_i)_{i=0}^{511}$  in Fig. 5(a). We compare the TV regularization which was presented in [77] with our approach by measuring the mean error  $E(f_o, u_r)$  of the result  $u_r \in (\mathbb{S}^2)^{511}$  to the original data  $f_o$ , which is always shown in gray.

The TV regularized result shown in Fig. 5(b) suffers from the well known staircasing effect, i.e., the signal is piecewise constant which yields groups of points having the same value and the signal to look sparser. The parameter was optimized with respect to  $E$  by a parameter search on  $\frac{1}{100}\mathbb{N}$  for  $\alpha$  and  $\frac{1}{10}\mathbb{N}$  for  $\beta$ . When just using second order differences, i.e. setting  $\alpha = 0$ , we obtain a better value for the quality measure, namely for  $\beta = 10$  we obtain  $E = 3.66 \times 10^{-2}$ , see Fig. 5(c). Combining the first and second order differences yields the best result with respect to  $E$ , i.e.  $E = 3.27 \times 10^{-2}$  for  $\alpha = 0.16$  and  $\beta = 12.4$ .

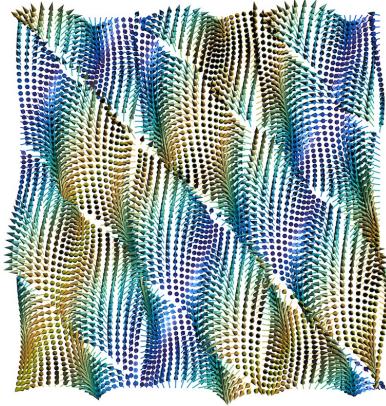
**Two-Dimensional Sphere-Valued Data Example.** We define an  $\mathbb{S}^2$ -valued vector-field by

$$G(t, s) = R_{t+s} S_{t-s} e_3, \quad t \in [0, 5\pi], s \in [0, 2\pi],$$

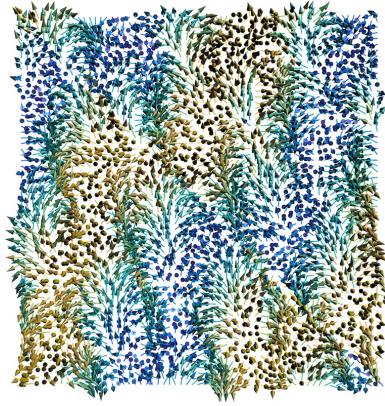
where  $R_\theta := \begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix}, S_\theta := \begin{pmatrix} \cos \theta & 0 & -\sin \theta \\ 0 & 1 & 0 \\ \sin \theta & 0 & \cos \theta \end{pmatrix}.$

We sample both dimensions with  $n = 64$  points, and obtain a discrete vector field  $f_o \in (\mathbb{S}^2)^{64 \times 64}$  which is illustrated in Fig. 6(a) the following way: on an equispaced grid the point on  $\mathbb{S}^2$  is drawn as an arrow, where the color emphasizes the elevation using the colormap `parula` from MATLAB, cf. Fig. 6(e). Similar to the sphere-valued signal, this vector field is affected by Gaussian noise imposed on the tangential plane at each point having a standard deviation of  $\sigma = \frac{5}{45}\pi$ . The resulting noisy data  $f$  is shown in Fig. 6(b).

We again perform a parameter grid search on  $\frac{1}{200}\mathbb{N}$  to find good reconstructions of the noisy data, first for the denoising with first order difference terms (TV). For  $\alpha = 3.5 \times 10^{-2}$  we obtain the vector field shown in Fig. 6(c) having  $E = 0.1879$ . Introducing the complete functional from (3), we obtain setting  $\beta = 8.6$  and  $\alpha = 0$  an error of just  $E = 0.1394$ , see Fig. 6(d). Indeed, just using a second order difference term yields the best result here. Still, both methods cannot reconstruct the jumps along the diagonal lines from the original signal, because they vanish in noise. Only the main diagonal jump can roughly been recognized in both cases.



(a) Original unit vector field.



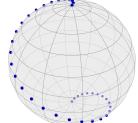
(b) Noisy unit vector field.



(c) Reconstruction with TV<sub>1</sub>.

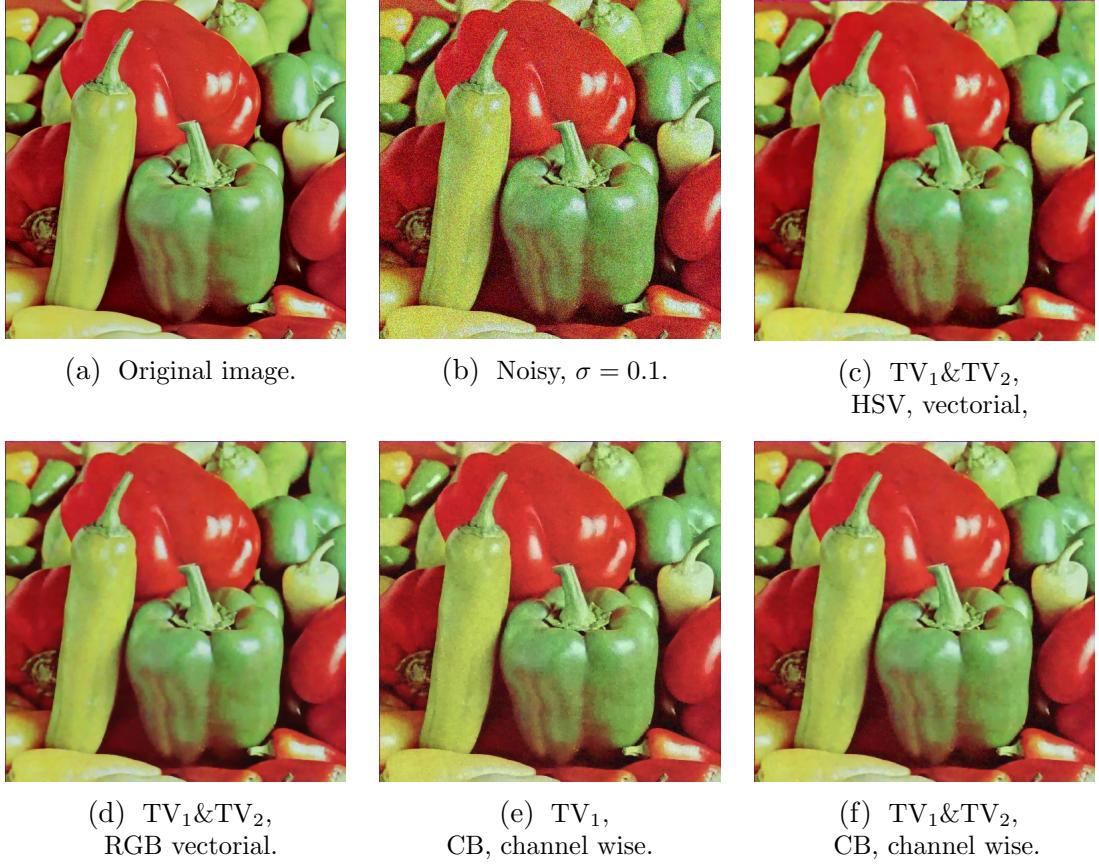


(d) Reconstruction with TV<sub>2</sub>.



(e) Colormap illustration: a signal on  $\mathbb{S}^2$  (left) and drawn using arrows and the colormap `parula` for elevation (right).

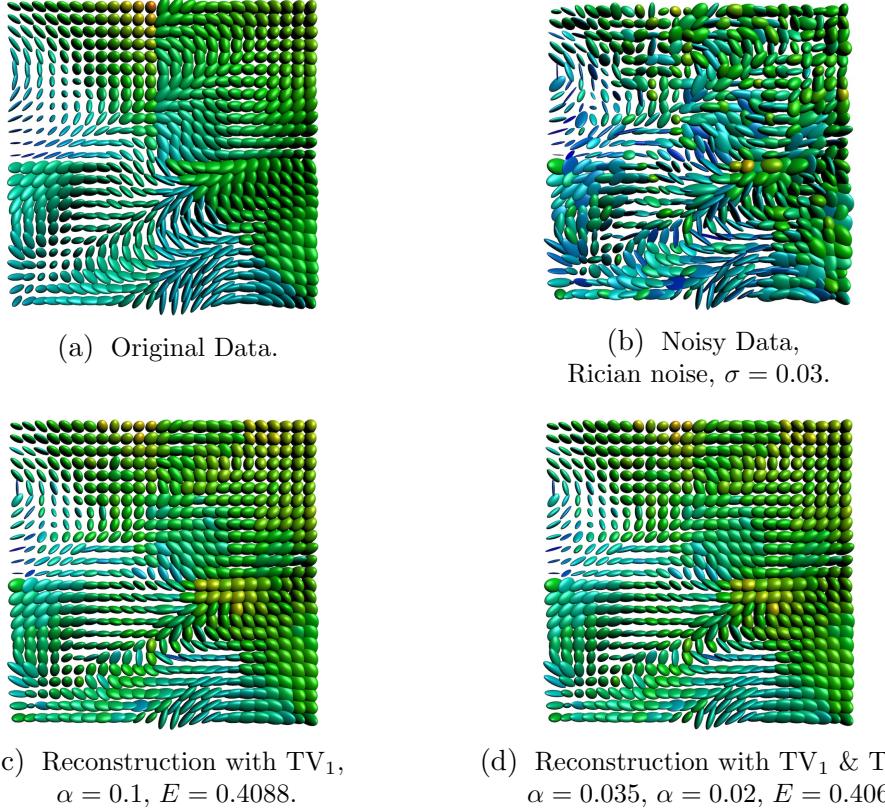
**Figure 6.** Denoising results of an (a)  $\mathbb{S}^2$ -valued vector field, which is obstructed by (b) Gaussian noise on  $T_f \mathbb{S}^2$ ,  $\sigma = \frac{4}{45}\pi$ . A reconstruction using (c) TV<sub>1</sub> approach,  $\alpha = 3.5 \times 10^{-2}$ , yields  $E = 0.1879$  while (d) the reconstruction with TV<sub>2</sub>,  $\alpha = 0$ ,  $\beta = 8.6$ , yields an error of just  $E = 0.1394$ .



**Figure 7.** Denoising the “Peppers” image using approaches in various color spaces: (c) On HSV with an vectorial approach using  $\alpha = 0.0625, \beta = 0.125$  which yields a PSNR of 28.155, (d) on RGB with  $\alpha = 0.05, \beta = 0.025$  yields a PSNR of 31.241, and two approaches channel wise on CB, where (e) a TV approach with  $\alpha = 0.05$  results in a PSNR of 28.969 and (f) a  $\text{TV}_1\&\text{TV}_2$  approach,  $\alpha = 0.024, \beta = 0.022$ , yields a PSNR of 29.7692.

**Application to Image Denoising.** Next we deal with denoising in different color spaces. Therefore we take the image “Peppers”<sup>†</sup>, cf. Fig. 7(a). This image is distorted with Gaussian noise on each of the red, green and blue (RGB) channels with  $\sigma = 0.1$ , cf. Fig. 7(b). Besides the RGB space, we consider the Hue-Value-Saturation (HSV) color space consisting of a  $\mathbb{S}^1$ -valued hue component  $H$  and two real valued components  $S, V$ . For the latter one, there are many methods, e.g., vector valued TV. For both, the authors presented a vector-valued first and second order TV-type approach in [10, 11]. We compare the vectorial approaches to the Chromaticity-Brightness (CB), where we apply a second order TV on the real-valued brightness and the  $\mathbb{S}^2$ -valued chromaticity separately. To be precise, the obtained chromaticity values are in the positive octant of  $\mathbb{S}^2$ . Again we search for the best value —here with respect to PSNR— of the denoising models at hand on a grid of  $\frac{1}{500}\mathbb{N}$  for the available parameters. For the component based approach of CB, both components are treated with the same parameters.

<sup>†</sup>Taken from the USC-SIPI Image Database, available online at <http://sipi.usc.edu/database/database.php?volume=misc&image=15>



**Figure 8.** Denoising an artificial image of SPD-valued data.

While for this example already the TV-based approach on the separate channels C and B outperforms the HSV vectorial approach, both the TV and the combined first and second order approach on CB are outperformed by the vectorial RGB approach. The reason for that is, that both channels of brightness and chromaticity in the latter model are not coupled. It would be interesting to couple the channels in the CB color model in a future work.

## 5.2. $\mathcal{P}(3)$ -valued Images

**An Artificial Matrix-Valued Image.** We construct an artificial image of  $\mathcal{P}(3)$ -valued pixels by sampling

$$G(s, t) := A(s, t) \operatorname{diag} \begin{pmatrix} 1 + \delta_{x+y, 1} \\ 1 + s + t + \frac{3}{2} \delta_{s, \frac{1}{2}} \\ 4 - s - t + \frac{3}{2} \delta_{t, \frac{1}{2}} \end{pmatrix} A(s, t)^T, \quad s, t \in [0, 1],$$

where  $A(s, t) = R_{x_2, x_3}(\pi s) R_{x_1, x_2}(|2\pi s - \pi|) R_{x_1, x_2}(|\pi(t - s - \lfloor t - s \rfloor) - \pi|)$ ,

$R_{x_i, x_j}(t)$  rotation in the  $x_i, x_j$ -plane and  $\delta_{a,b} = \begin{cases} 1 & \text{if } a > b \\ 0 & \text{else.} \end{cases}$

Despite the outer rotations the diagonal, i.e., the eigenvalues introduce three jumps along both center vertical and horizontal lines and along the diagonal, see Fig. 8(a), where this function is sampled to obtain an  $25 \times 25$  matrix valued image  $f = (f_{i,j})_{i,j=1}^{25} \in$

$\mathcal{P}(3)^{25,25}$ . We visualize any symmetric positive definite matrix  $f_{i,j}$  by drawing a shifted ellipsoid given by the surface niveau  $\{x \in \mathbb{R}^3 : (x^T - c(i,j))f_{i,j}(x - c(i,j)^T) = 1\}$  for some grid scaling parameter  $c > 0$ . As coloring we use the anisotropy index relative to the Riemannian distance [48] normalized onto  $[0, 1]$ , which is also known as the geodesic anisotropy index. Together with the hue color map from MATLAB both the unit matrix yielding a sphere and the case where one eigenvalue dominates by far get colored in red.

**Application to DT-MRI.** Finally we explore the capabilities of applying the denoising technique to real world data. The Camino project<sup>‡</sup>[20] provides a dataset of a Diffusion Tensor Magnetic Resonance Image (DT-MRI) of the human head, which is freely available.<sup>§</sup> From the complete dataset of  $\tilde{f} = (\tilde{f}_{i,j,k}) \in \mathcal{P}(3)^{112 \times 112 \times 50}$  we take the traversal plane  $k = 28$ , see Fig. 9(a). By combining a first and second order model for denoising, the noisy parts are reduced, while constant parts as well as basic features are kept, see Fig. 9(b). To see more detail, we focus on the subset  $(i, j) \in \{28, \dots, 87\} \times \{24, \dots, 73\}$ , which is shown in Figs. 9(c) and (d), respectively.

**Acknowledgement.** This research was partly conducted when AW visited the University Kaiserslautern and when MB and RB were visiting the Helmholtz-Zentrum München. We would like to thank A. Trouvé for valuable discussions. AW is supported by the Helmholtz Association within the young investigator group VH-NG-526. AW also acknowledges the support by the DFG scientific network “Mathematical Methods in Magnetic Particle Imaging”. GS acknowledges the financial support by DFG Grant STE571/11-1.

## A. The Sphere $\mathbb{S}^2$

We use the parametrization

$$x(\theta, \varphi) = (\cos \varphi \cos \theta, \sin \varphi \cos \theta, \sin \theta)^T, \quad \theta \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right), \varphi \in [0, 2\pi)$$

with north pole  $z_0 := (0, 0, 1)^T = x(\frac{\pi}{2}, \varphi_0)$  and south pole  $(0, 0, -1)^T = x(-\frac{\pi}{2}, \varphi_0)$ . Then we have for the tangent spaces

$$T_x(\mathbb{S}^d) = T_{x(\theta, \varphi)}(\mathbb{S}^2) := \{v \in \mathbb{R}^{d+1} : v^T x = 0\} = \text{span}\{e_1(\theta, \varphi), e_2(\theta, \varphi)\}$$

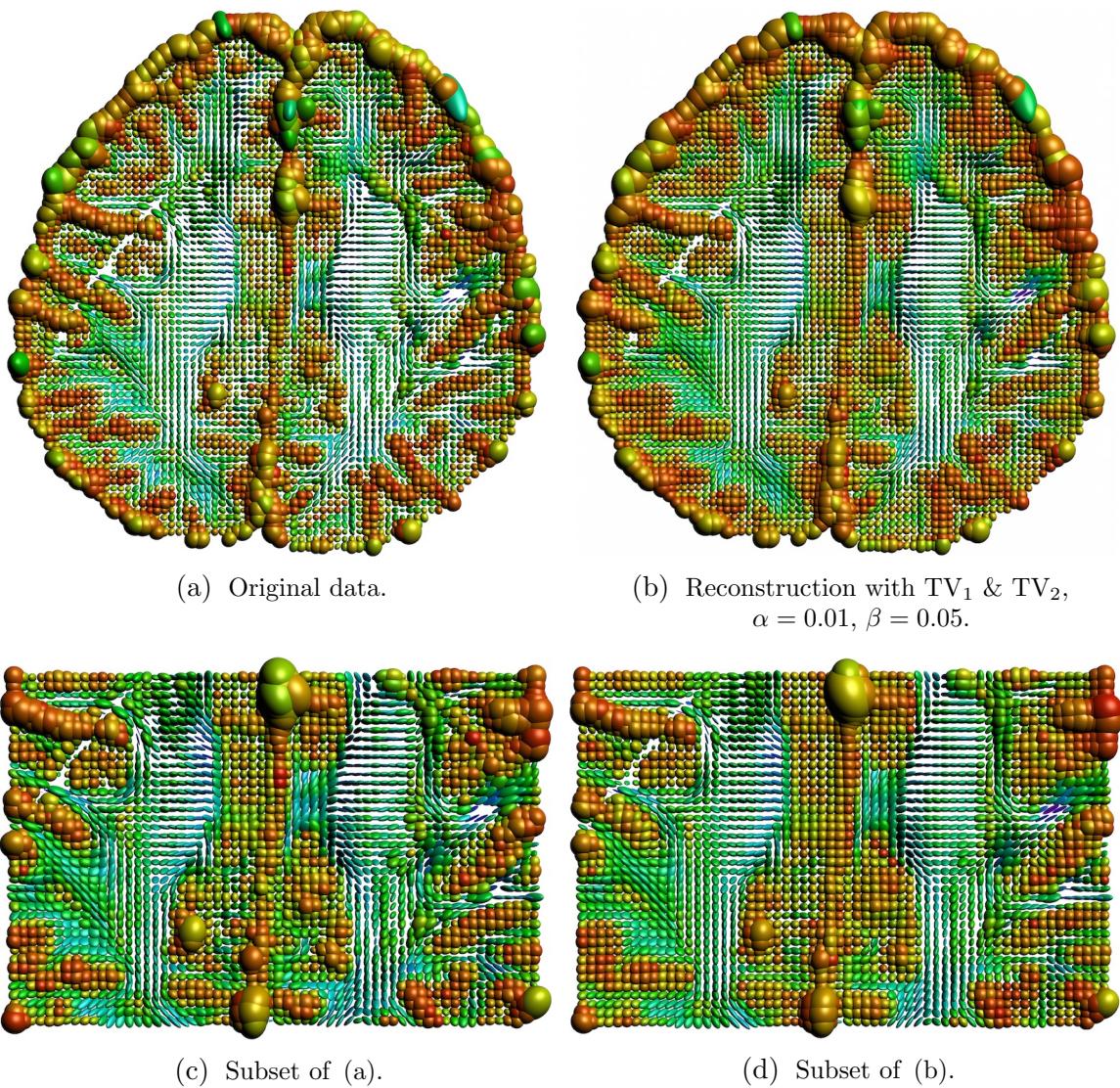
with the normed orthogonal vectors  $e_1(\theta, \varphi) = \frac{\partial x}{\partial \theta} = (\cos \varphi \sin \theta, -\sin \varphi \sin \theta, \cos \theta)^T$  and  $e_2(\theta, \varphi) = \frac{1}{\cos \theta} \frac{\partial x}{\partial \varphi} = (-\sin \varphi, \cos \varphi, 0)^T$ . The geodesic distance is given by  $d_{\mathbb{S}^2}(x_1, x_2) = \arccos(x_1^T x_2)$ , the unit speed geodesic by  $\gamma_{x, \eta}(t) = \cos(t)x + \sin(t)\frac{\eta}{\|\eta\|_2}$ , and the exponential map by

$$\exp_x(t\eta) = \cos(t\|\eta\|_2)x + \sin(t\|\eta\|_2)\frac{\eta}{\|\eta\|_2}.$$

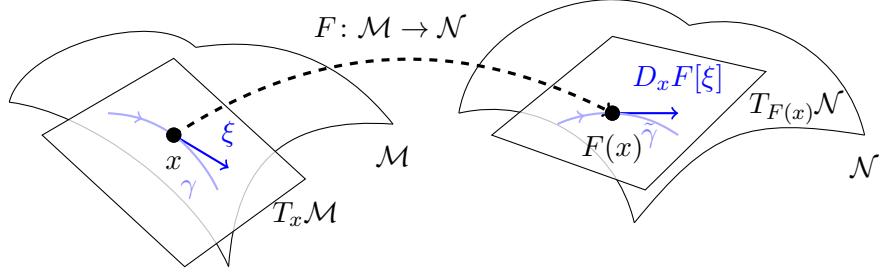
---

<sup>‡</sup>see <http://cmic.cs.ucl.ac.uk/camino>

<sup>§</sup>follow the tutorial at <http://cmic.cs.ucl.ac.uk/camino//index.php?n=Tutorials.DTI>



**Figure 9.** The Camino DT-MRI data of slice 28: (a) the original data, (b) the  $\text{TV}_1\&\text{TV}_2$ -regularized model keeps the main features but smoothes the data. For both data a subset is shown in (c) and (d), respectively.



**Figure 10.** Illustration of the differential of  $F: \mathcal{M} \rightarrow \mathcal{N}$ . Here  $\tilde{\gamma} = F \circ \gamma$ .

Finally, a unit speed geodesic through  $x$  (with  $\theta \geq 0$ ) and the north pole  $z_0$  reads

$$\gamma_{x,e_1}(t) = \cos(t)x + \sin(t)e_1, \quad \gamma_{x,e_1}(T) = z_0, \quad c(x) = \gamma_{x,e_1}\left(\frac{T}{2}\right), \quad T = \frac{\pi}{2} - \theta_x$$

and the orthogonal frame along this geodesic as  $E_1(t) = e_1(\theta(t), \varphi)$ ,  $E_2(t) = e_2(\theta(t), \varphi)$ ,  $\theta(t) := \theta_x + \frac{t}{T} \left(\frac{\pi}{2} - \theta_x\right)$ .

## B. The Manifold $\mathcal{P}(r)$ of Symmetric Positive Definite Matrices

We provide definitions for the manifold of positive definite symmetric  $r \times r$  matrices  $\mathcal{P}(r)$  which are required in our computations, see [64]. By Exp and Log we denote the matrix exponential and logarithm defined by  $\text{Exp } x := \sum_{k=0}^{\infty} \frac{1}{k!} x^k$  and  $\text{Log } x := \sum_{k=1}^{\infty} \frac{1}{k} (I - x)^k$ ,  $\rho(I - x) < 1$ . The geodesic distance is given by

$$d_{\mathcal{P}}(x_1, x_2) := \|\text{Log}(x_1^{-\frac{1}{2}} x_2 x_1^{-\frac{1}{2}})\|.$$

Further, we have the exponential map  $\exp_x(t\eta) := x^{\frac{1}{2}} \text{Exp}(tx^{-\frac{1}{2}} \eta x^{-\frac{1}{2}}) x^{\frac{1}{2}}$ . The unit speed geodesic linking  $x$  and  $z$  for  $t = T = d_{\mathcal{P}}(x, z)$  is

$$\gamma_{x,z}(t) = \exp_x(tv) := x^{\frac{1}{2}} \text{Exp}\left(\frac{t}{T} \text{Log}(x^{-\frac{1}{2}} zx^{-\frac{1}{2}})\right) x^{\frac{1}{2}},$$

where  $v = x^{\frac{1}{2}} \text{Log}(x^{-\frac{1}{2}} zx^{-\frac{1}{2}}) x^{\frac{1}{2}} / \|\text{Log}(x^{-\frac{1}{2}} zx^{-\frac{1}{2}})\|$ . In particular we obtain

$$c(x, z) = \gamma_{x,z}\left(\frac{1}{2}\right) = x^{\frac{1}{2}} (x^{-\frac{1}{2}} zx^{-\frac{1}{2}})^{\frac{1}{2}} x^{\frac{1}{2}}$$

which is known as the geometric mean of  $x$  and  $z$ . The parallel transport of  $\eta \in T_x \mathcal{P}$  along the geodesic  $\gamma_{x,\xi}(t) = \exp_x(t\xi)$  is given by  $P_t(\eta) = \exp_x\left(\frac{t}{2}\xi\right) x^{-1} \eta x^{-1} \exp_x\left(\frac{t}{2}\xi\right)$ .

## C. Basics on Parallel Transport

In this section we review some concepts from differential geometry which were used in the paper. For more details we refer to [42]. Let  $\mathcal{C}^\infty(\mathcal{M})$  denote the set of smooth

real-valued functions on a manifold  $\mathcal{M}$  and  $\mathcal{C}^\infty(x)$  the functions defined on some open neighborhood of  $x \in \mathcal{M}$  which are smooth at  $x$ . Further, let  $\mathcal{C}^\infty(\mathcal{M}, \mathcal{N})$  denote the smooth maps from  $\mathcal{M}$  to a manifold  $\mathcal{N}$ . For computational purposes we introduce tangent vectors by their curve realizations. A tangent vector  $\xi$  to a manifold  $\mathcal{M}$  at  $x \in \mathcal{M}$  is a mapping from  $\mathcal{C}^\infty(x)$  to  $\mathbb{R}$  such that there exists a curve  $\gamma: \mathbb{R} \rightarrow \mathcal{M}$  with  $\gamma(0) = x$  satisfying

$$\xi f = \dot{\gamma}(0)f := \frac{d}{dt}f(\gamma(t))|_{t=0}.$$

The set of all tangent vectors at  $x \in \mathcal{M}$  forms the tangent space  $T_x \mathcal{M}$ . Further,  $T\mathcal{M} := \cup_x T_x \mathcal{M}$  is the tangent bundle of  $\mathcal{M}$ . Given  $F \in \mathcal{C}^\infty(\mathcal{M}, \mathcal{N})$  and a curve  $\gamma: (-\varepsilon, \varepsilon) \rightarrow \mathcal{M}$  with  $\gamma(0) = x$  and  $\dot{\gamma}(0) = \xi$ , then

$$D_x F: T_x \mathcal{M} \rightarrow T_{F(x)} \mathcal{N}, \quad \xi \mapsto D_x F[\xi] := (F \circ \gamma)'(0) \quad (37)$$

is a linear map between vector spaces, called differential or derivative of  $F$  at  $x$ . This is illustrated in Fig. 10.

Let  $\mathcal{X}(\mathcal{M})$  denote the linear space of smooth vector fields on  $\mathcal{M}$ , i.e., of smooth mappings from  $\mathcal{M}$  to  $T\mathcal{M}$ . On every Riemannian manifold there is the Riemannian or Levi-Civita connection

$$\nabla: T\mathcal{M} \times T\mathcal{M} \rightarrow T\mathcal{M}$$

which is uniquely determined by the following properties:

- i)  $\nabla_{f\Xi+g\Theta} X = f\nabla_\Xi X + g\nabla_\Theta X$  for all  $f, g \in \mathcal{C}^\infty(\mathcal{M})$  and all  $\Xi, \Theta, X \in \mathcal{X}(\mathcal{M})$ ,
- ii)  $\nabla_\Xi(aX + bY) = a\nabla_\Xi X + b\nabla_\Xi Y$  for all  $a, b \in \mathbb{R}$  and  $\Xi, X, Y \in \mathcal{X}(\mathcal{M})$ ,
- iii)  $\nabla_\Xi(fX) = \Xi(f)X + f\nabla_\Xi X$  for all  $f \in \mathcal{C}^\infty(\mathcal{M})$  and all  $\Xi, X \in \mathcal{X}(\mathcal{M})$ ,
- iv)  $\nabla$  is compatible with the Riemannian metric  $\langle \cdot, \cdot \rangle$ , i.e.,  $\Xi \langle X, Y \rangle = \langle \nabla_\Xi X, Y \rangle + \langle X, \nabla_\Xi Y \rangle$ ,
- v)  $\nabla$  is symmetric (torsion-free):  $[\Xi, \Theta] = \nabla_\Xi \Theta - \nabla_\Theta \Xi$ , where  $[\cdot, \cdot]$  denotes the Lie bracket.

For some real interval  $I$ , a map  $\Xi: I \rightarrow T\mathcal{M}$  is called a vector field along a curve  $\gamma: I \rightarrow \mathcal{M}$  if  $\Xi(t) \in T_{\gamma(t)} \mathcal{M}$  for all  $t \in I$ . Let  $\mathcal{X}(\gamma)$  denote the smooth vector fields along  $\gamma$ . The Riemannian connection determines for each curve  $\gamma: I \rightarrow \mathcal{M}$  a unique operator  $\frac{D}{dt}: \mathcal{X}(\gamma) \rightarrow \mathcal{X}(\gamma)$  with the properties:

- T1)  $\frac{D}{dt}(a\Xi + b\Theta) = a\frac{D}{dt}\Xi + b\frac{D}{dt}\Theta$  for all  $a, b \in \mathbb{R}$ ,
- T2)  $\frac{D}{dt}(f\Xi) = \dot{f}\Xi + f\frac{D}{dt}\Xi$  for all  $f \in \mathcal{C}^\infty(I)$ ,
- T3) If  $\Xi$  is extendible (to a neighborhood of the image of  $\gamma$ ), then for any extension  $\tilde{\Xi}$  it holds  $\frac{D}{dt}\Xi(t) = \nabla_{\dot{\gamma}(t)}\tilde{\Xi}$ .

Then  $\frac{D}{dt}\Xi$  is called covariant derivative of  $\Xi$  along  $\gamma$ . A vector field  $\Xi$  along a curve  $\gamma$  is said to be parallel along  $\gamma$  if

$$\frac{D}{dt}\Xi = 0.$$

The fundamental fact about parallel vector fields is that any tangent vector at any point on a curve can be uniquely extended to a parallel vector field along the entire curve. In this sense we can extend an (orthonormal) basis  $\{\xi_1, \dots, \xi_n\}$  of  $T_x\mathcal{M}$  parallel along a curve  $\gamma$  and call this a parallel transported (orthonormal) frame  $\{\Xi_1, \dots, \Xi_n\}$  along  $\gamma$ . Then any vector field  $\Xi \in \mathcal{X}(\gamma)$  can be written as  $\Xi(t) = \sum_{j=1}^n a_j(t)\Xi_j(t)$  and we obtain by T1) and T2) that

$$\frac{D}{dt}\Xi = \frac{D}{dt}\left(\sum_{j=1}^n a_j\Xi_j\right) = \sum_{j=1}^n \frac{D}{dt}(a_j\Xi_j) = \sum_{j=1}^n \dot{a}_j(t)\Xi_j. \quad (38)$$

## References

- [1] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton and Oxford, Princeton University Press, 2008.
- [2] A. D. Aleksandrov. A theorem on triangles in a metric space and some of its applications. In *Trudy Mat. Inst. Steklov.*, v 38, pages 5–23. Izdat. Akad. Nauk SSSR, Moscow, 1951.
- [3] M. Bačák. The proximal point algorithm in metric spaces. *Israel Journal of Mathematics*, 194(2):689–701, 2013.
- [4] M. Bačák. Computing medians and means in Hadamard spaces. *SIAM Journal on Optimization*, 24(3):1542–1566, 2014.
- [5] M. Bačák. *Convex analysis and optimization in Hadamard spaces*, volume 22 of *De Gruyter Series in Nonlinear Analysis and Applications*. De Gruyter, Berlin, 2014.
- [6] S. Banert. Backward–backward splitting in Hadamard spaces. 414(2):656–665, 2014.
- [7] P. Basser, J. Mattiello, and D. LeBihan. MR diffusion tensor spectroscopy and imaging. *Biophysical Journal*, 66:259–267, 1994.
- [8] M. Berger. *A Panoramic View of Riemannian Geometry*. Springer Science & Business Media, 2003.
- [9] 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.
- [10] R. Bergmann and A. Weinmann. Inpainting of cyclic data using first and second order differences. In *EMCVPR2015*, Lecture Notes in Computer Science, pages 155–168, Berlin, 2015. Springer.
- [11] R. Bergmann and A. Weinmann. A second order TV-type approach for inpainting and denoising higher dimensional combined cyclic and vector space data. *ArXiv Preprint*, 1501.02684, 2015.
- [12] D. P. Bertsekas. Incremental proximal methods for large scale convex optimization. *Mathematical Programming*, 129(2, Ser. B):163–195, 2011.
- [13] K. Bredies, K. Kunisch, and T. Pock. Total generalized variation. *SIAM Journal on Imaging Sciences*, 3(3):1–42, 2009.
- [14] B. Burgeth, M. Welk, C. Feddern, and J. Weickert. Morphological operations on matrix-valued images. In *Computer Vision - ECCV 2004*, Lecture Notes in Computer Science, 3024, pages 155–167, Berlin, 2004. Springer.
- [15] A. Chambolle and P.-L. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76(2):167–188, 1997.

- [16] T. F. 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.
- [17] T. F. Chan, A. Marquina, and P. Mulet. High-order total variation-based image restoration. *SIAM Journal on Scientific Computation*, 22(2):503–516, 2000.
- [18] J. Cheeger and D. Ebin. *Comparison Theorems in Riemannian Geometry*, volume 365. American Mathematical Society, 1975.
- [19] C. Chefd’Hotel, D. Tschumperlé, R. Deriche, and O. Faugeras. Regularizing flows for constrained matrix-valued images. *Journal of Mathematical Imaging and Vision*, 20:147–162, 2004.
- [20] P. A. Cook, Y. Bai, S. Nedjati-Gilani, K. K. Seunarine, M. G. Hall, G. J. Parker, and D. C. Alexander. Camino: Open-source diffusion-mri reconstruction and processing. In *Proc. Intl. Soc. Mag. Reson. Med. 14*, page 2759, Seattle, WA, USA, 2006.
- [21] S. Didas, G. Steidl, and S. Setzer. Combined  $\ell_2$  data and gradient fitting in conjunction with  $\ell_1$  regularization. *Advances in Computational Mathematics*, 30(1):79–99, 2009.
- [22] S. Didas, J. Weickert, and B. Burgeth. Properties of higher order nonlinear diffusion filtering. *Journal of Mathematical Imaging and Vision*, 35:208–226, 2009.
- [23] M. P. do Carmo. *Riemannian Geometry*. Birkhäuser, 1992.
- [24] J. H. Eschenburg. Lecture notes on symmetric spaces. *Preprint*, 1997.
- [25] J. H. Eschenburg. Symmetric spaces, topology, and linear algebra. *Preprint*, 2014.
- [26] O. P. Ferreira and P. R. Oliveira. Subgradient algorithm on Riemannian manifolds. *Journal of Optimization Theory and Applications*, 97(1):93–104, 1998.
- [27] O. P. Ferreira and P. R. Oliveira. Proximal point algorithm on Riemannian manifolds. *Optimization*, 51(2):257–270, 2002.
- [28] O. Freifeld and M. J. Black. Lie bodies: A manifold representation of 3D human shape. In *ECCV 2012*, pages 1–14. Springer, 2012.
- [29] M. Giaquinta, G. Modica, and J. Souček. Variational problems for maps of bounded variation with values in  $S^1$ . *Calculus of Variation*, 1(1):87–121, 1993.
- [30] M. Giaquinta and D. Mucci. The BV-energy of maps into a manifold: relaxation and density results. *Ann. Sc. Norm. Super. Pisa Cl. Sci.*, 5(4):483–548, 2006.
- [31] M. Giaquinta and D. Mucci. Maps of bounded variation with values into a manifold: total variation and relaxed energy. *Pure Applied Mathematics Quarterly*, 3(2):513–538, 2007.
- [32] P. Grohs, H. Hardering, and O. Sander. Optimal a priori discretization error bounds for geodesic finite elements. *Foundations of Computational Mathematics*, 2015. to appear.
- [33] P. Grohs and S. Hosseini.  $\varepsilon$ -subgradient algorithms for locally Lipschitz functions on Riemannian manifolds. *SAM Report 2013-49, ETH Zürich*, 2013.
- [34] P. Grohs and M. Sprecher. Total variation regularization by iteratively reweighted least squares on Hadamard spaces and the sphere. *Preprint 2014-39, ETH Zürich*, 2014.
- [35] P. Grohs and J. Wallner. Interpolatory wavelets for manifold-valued data. *Applied and Computational Harmonic Analysis*, 27(3):325–333, 2009.
- [36] W. Hinterberger and O. Scherzer. Variational methods on the space of functions of bounded Hessian for convexification and denoising. *Computing*, 76(1):109–133, 2006.

- [37] M. Hintermüller and T. Wu. Robust principal component pursuit via inexact alternating minimization on matrix manifolds. *Journal of Mathematical Imaging and Vision*, 51(3):361–377, 2014.
- [38] J. Jost. Convex functionals and generalized harmonic maps into spaces of nonpositive curvature. *Commentarii Mathematici Helvetici*, 70(4):659–673, 1995.
- [39] J. Jost. *Nonpositive curvature: geometric and analytic aspects*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 1997.
- [40] R. Kimmel and N. Sochen. Orientation diffusion or how to comb a porcupine. *Journal of Visual Communication and Image Representation*, 13:238–248, 2002.
- [41] R. Lai and S. Osher. A splitting method for orthogonality constrained problems. *Journal of Scientific Computing*, 58(2):431–449, 2014.
- [42] J. M. Lee. *Riemannian Manifolds. An Introduction to Curvature*. Springer-Verlag, New York-Berlin, New York-Berlin-Heidelberg, 1997.
- [43] S. Lefkimiatis, A. Bourquard, and M. Unser. Hessian-based norm regularization for image restoration with biomedical applications. *IEEE Transactions on Image Processing*, 21(3):983–995, 2012.
- [44] J. Lellmann, E. Strekalovskiy, S. Koetter, and D. Cremers. Total variation regularization for functions with values in a manifold. In *IEEE ICCV 2013*, pages 2944–2951, 2013.
- [45] M. Lysaker, A. Lundervold, and X.-C. Tai. Noise removal using fourth-order partial differential equations with applications to medical magnetic resonance images in space and time. *IEEE Transactions on Image Processing*, 12(12):1579–1590, 2003.
- [46] M. Lysaker and X.-C. Tai. Iterative image restoration combining total variation minimization and a second-order functional. *International Journal of Computer Vision*, 66(1):5–18, 2006.
- [47] U. F. Mayer. Gradient flows on nonpositively curved metric spaces and harmonic maps. *Communications in Analysis and Geometry*, 6(2):199–253, 1998.
- [48] M. Moakher and P. G. Batchelor. Symmetric positive-definite matrices: From geometry to applications and visualization. In *Visualization and Processing of Tensor Fields*, pages 285–298. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.
- [49] J. J. Moreau. Fonctions convexes duales et points proximaux dans un espace hilbertien. *C. R. Acad. Sci. Paris Ser. A Math.*, 255:2897–2899, 1962.
- [50] K. Papafitsoros and C. B. Schönlieb. A combined first and second order variational approach for image reconstruction. *Journal of Mathematical Imaging and Vision*, 2(48):308–338, 2014.
- [51] N. Parikh and S. Boyd. Proximity algorithms. *Foundations and Trends in Optimization*, 1(3):123–231, 2013.
- [52] F. C. Park, J. E. Bobrow, and S. R. Ploen. A Lie group formulation of robot dynamics. *The International Journal of Robotics Research*, 14(6):609–618, 1995.
- [53] X. Pennec, P. Fillard, and N. Ayache. A Riemannian framework for tensor computing. *International Journal of Computer Vision*, 66:41–66, 2006.
- [54] I. U. Rahman, I. Drori, V. C. Stodden, and D. L. Donoho. Multiscale representations for manifold-valued data. *SIAM Journal on Multiscale Modeling and Simulation*, 4(4):1201–1232, 2005.
- [55] M. Raptis and S. Soatto. Tracklet descriptors for action modeling and video analysis. In *ECCV 2010*, pages 577–590. Springer, 2010.

- [56] H. E. Rauch. The global study of geodesics in symmetric and nearly symmetric Riemannian manifolds. *Commentarii Mathematici Helvetici*, 35:111–125, 1961.
- [57] J. G. Rešetnjak. Non-expansive maps in a space of curvature no greater than  $K$ . *Akademija Nauk SSSR. Sibirskoe Otdelenie. Sibirskiĭ Matematičeskiĭ Žurnal*, 9:918–927, 1968.
- [58] 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.
- [59] G. Rosman, X.-C. Tai, R. Kimmel, and A. M. Bruckstein. Augmented-Lagrangian regularization of manifold-valued maps. *Methods and Applications of Analysis*, 21(1):105–122, 2014.
- [60] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60(1):259–268, 1992.
- [61] O. Scherzer. Denoising with higher order derivatives of bounded variation and an application to parameter estimation. *Computing*, 60:1–27, 1998.
- [62] S. Setzer and G. Steidl. Variational methods with higher order derivatives in image processing. In *Approximation XII: San Antonio 2007*, pages 360–385, 2008.
- [63] S. Setzer, G. Steidl, and T. Teuber. Infimal convolution regularizations with discrete l1-type functionals. *Communications in Mathematical Sciences*, 9(3):797–872, 2011.
- [64] S. Sra and R. Hosseini. Conic geometric optimization on the manifold of positive definite matrices. *ArXiv Preprint 1320.1039v3*, 2014.
- [65] G. Steidl, S. Setzer, B. Popilka, and B. Burgeth. Restoration of matrix fields by second order cone programming. *Computing*, 81:161–178, 2007.
- [66] E. Strekalovskiy and D. Cremers. Total variation for cyclic structures: Convex relaxation and efficient minimization. In *IEEE CVPR 2011*, pages 1905–1911. IEEE, 2011.
- [67] E. Strekalovskiy and D. Cremers. Total cyclic variation and generalizations. *Journal of Mathematical Imaging and Vision*, 47(3):258–277, 2013.
- [68] R. Tron, B. Afsari, and R. Vidal. On the convergence of gradient descent for finding the Riemannian center of mass. *SIAM Journal on Control and Optimization*, 51(3):2230–2260, 2013.
- [69] D. Tschumperlé and R. Deriche. Diffusion tensor regularization with constraints preservation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1948–1953, 2001.
- [70] O. Tuzel, F. Porikli, and P. Meer. Learning on Lie groups for invariant detection and tracking. In *CVPR 2008*, pages 1–8. IEEE, 2008.
- [71] C. Udriște. *Convex functions and optimization methods on Riemannian manifolds*, volume 297 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1994.
- [72] T. Valkonen, K. Bredies, and F. Knoll. Total generalized variation in diffusion tensor imaging. *SIAM Journal on Imaging Sciences*, 6(1):487–525, 2013.
- [73] 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.
- [74] J. Wang, C. Li, G. Lopez, and J.-C. Yao. Convergence analysis of inexact proximal point algorithms on Hadamard manifolds. *Journal of Global Optimization*, 61(3):553–573, 2014.
- [75] J. Weickert, C. Feddern, M. Welk, B. Burgeth, and T. Brox. PDEs for tensor image processing. In *Visualization and Processing of Tensor Fields*, pages 399–414, Berlin, 2006. Springer.

- [76] A. Weinmann. Interpolatory multiscale representation for functions between manifolds. *SIAM Journal on Mathematical Analysis*, 44(1):162–191, 2012.
- [77] A. Weinmann, L. Demaret, and M. Storath. Total variation regularization for manifold-valued data. *SIAM Journal on Imaging Sciences*, 7(4):2226–2257, 2014.
- [78] M. Welk, C. Feddern, B. Burgeth, and J. Weickert. Median filtering of tensor-valued images. In *Pattern Recognition*, Lecture Notes in Computer Science, 2781, pages 17–24, Berlin, 2003. Springer.