FENCHEL DUALITY THEORY AND A PRIMAL-DUAL ALGORITHM ON RIEMANNIAN MANIFOLDS*

RONNY BERGMANN†, ROLAND HERZOG†, DANIEL TENBRINCK‡, AND JOSÉ VIDAL-NÚÑEZ†

Abstract. This paper introduces a new duality theory that generalizes the classical Fenchel conjugation to functions defined on Riemannian manifolds. We investigate its properties, e.g., the Fenchel–Young inequality and the characterization of the convex subdifferential using the analogue of the Fenchel–Moreau Theorem. These properties of the Fenchel conjugate are employed to derive a Riemannian primal-dual optimization algorithm, and to prove its convergence for the case of Hadamard manifolds under appropriate assumptions. Numerical results illustrate the performance of the algorithm, which competes with the recently derived Douglas–Rachford algorithm on manifolds of nonpositive curvature. Furthermore we show numerically that our novel algorithm even converges on manifolds of positive curvature.

Key words. convex analysis, Fenchel conjugate function, Riemannian manifold, Hadamard manifold, primal-dual algorithm, Chambolle–Pock algorithm, total variation

AMS subject classifications. 49N15, 49M29, 90C26, 49Q99,

1. Introduction.

Convex analysis plays an important role in optimization, and an elaborate theory on convex analysis and conjugate duality is available on locally convex vector spaces. Among the vast references on this topic, we mention Bauschke, Combettes, 2011 for convex analysis and monotone operator techniques, Ekeland, Temam, 1999 for convex analysis and the perturbation approach to duality, or Rockafellar, 1970 for an in-depth development of convex analysis on Euclidean spaces. Rockafellar, 1974 focuses on conjugate duality on Euclidean spaces, Zălinescu, 2002; Boţ, 2010 on conjugate duality on locally convex vector spaces, and Martínez-Legaz, 2005 on some particular applications of conjugate duality in economics.

We wish to emphasize in particular the role of convex analysis in the analysis and numerical solution of regularized ill-posed problems. Consider for instance the total variation functional, which was introduced for imaging applications in the famous ROF model (Rudin, Osher, Fatemi, 1992) and which is known for its ability to preserve sharp edges. We refer the reader to Chambolle, Caselles, et al., 2010 for further details about total variation for image analysis. Further applications and regularizers can be found in Chambolle, Lions, 1997; Strong, Chan, 2003; Chambolle, 2004;

---

*

Funding: This work was partially supported by DFG grants BE 5888/2–1 and HE 6077/10–1, the latter within the Priority Program SPP 1962, which is gratefully acknowledged.

†Technische Universität Chemnitz, Faculty of Mathematics, Workgroup Numerical Mathematics (Partial Differential Equations), D–09107 Chemnitz, Germany (RB: ronny.bergmann@mathematik.tu-chemnitz.de, https://www.tu-chemnitz.de/mathematik/part_dgl/people/bergmann; RH: roland.herzog@mathematik.tu-chemnitz.de, https://www.tu-chemnitz.de/mathematik/part_dgl/people/herzog); JVN: jose.vidal-nunez@mathematik.tu-chemnitz.de, https://www.tu-chemnitz.de/mathematik/part_dgl/people/vidal).

‡Friedrich-Alexander Universität Erlangen-Nürnberg, Department of Mathematics, Professorship of Applied Mathematics, D–91058 Erlangen, Germany, daniel.tenbrinck@fau.de, https://en.www.math.fau.de/applied-mathematics-prof-burger/members/dr-daniel-tenbrinck.

Chan, Esedoglu, et al., 2006; Wang et al., 2008. In addition, higher order differences or differentials can be taken into account, see for example Chan, Marquina, Mulet, 2000; Papafitsoros, Schönlieb, 2014 or most prominently the total generalized variation (TGV) Bredies, Kunisch, Pock, 2010. These models use the idea of the pre-dual formulation of the energy functional and Fenchel duality to derive efficient algorithms. Within the image processing community the resulting algorithms of primal-dual hybrid gradient type are often referred to as the Chambolle–Pock algorithm (Chambolle, Pock, 2011).

In recent years, optimization on Riemannian manifolds has gained a lot of interest. Starting in the 1980s and 90s with Udrişte, 1994, optimization on Riemannian manifolds and corresponding algorithms have been investigated. For a comprehensive textbook on optimization on matrix manifolds, see Absil, Mahony, Sepulchre, 2008. With the emergence of manifold-valued imaging, for example in InSAR imaging Bürgmann, Rosen, Fielding, 2000, data consisting of orientations for example in electron backscattered diffraction (EBSD) Adams, Wright, Kunze, 1993; Kunze et al., 1993 or for diffusion tensors in magnetic resonance imaging (DT-MRI), for example discussed in Pennec, Fillard, Ayache, 2006, the development of optimization techniques and/or algorithms on manifolds (especially for non-smooth functionals) has gained a lot of attention. Within these applications, the same tasks appear as for classical, Euclidean imaging, such as denoising, inpainting or segmentation. Both Lellmann et al., 2013 as well as Weinmann, Demaret, Storath, 2014 introduced the total variation as a prior in a variational model for manifold-valued images. While the first extends a lifting approach previously introduced for cyclic data in Strekalovskiy, Cremers, 2011 to Riemannian manifolds, the latter introduces a cyclic proximal point algorithm (CPPA) to compute a minimizer of the variational model. Such an algorithm was previously introduced by Bačák, 2014a on CAT(0) spaces based on the proximal point algorithm introduced by Ferreira, Oliveira, 2002 on Riemannian manifolds. Based on these models and algorithms, higher order models have been derived Bergmann, Laus, et al., 2014; Bačák et al., 2016; Bergmann, Fitschen, et al., 2018; Bredies, Holler, et al., 2018. Using a relaxation, the half-quadratic minimization Bergmann, Chan, et al., 2016, also known as iteratively reweighted least squares (IRLS) Grohs, Sprecher, 2016, has been generalised to manifold-valued image processing tasks and employs a quasi-Newton method. Finally, the parallel Douglas–Rachford algorithm (PDRA) was introduced on Hadamard manifolds Bergmann, Persch, Steidl, 2016 and its convergence proof is, to the best of our knowledge, limited to manifolds with constant nonpositive curvature. Numerically, the PDRA still performs well on arbitrary Hadamard manifolds. However, for the classical Euclidean case the Douglas–Rachford algorithm is equivalent to applying the alternating directions method of multipliers (ADMM) Gabay, Mercier, 1976 on the dual problem and hence is also equivalent to the algorithm of Chambolle, Pock, 2011.

In this paper we introduce a new notion of Fenchel duality for Riemannian manifolds, which allows us to derive a conjugate duality theory for convex optimization problems posed on such manifolds. Our theory allows new algorithmic approaches to be devised for optimization problems on manifolds. In the absence of a global concept of convexity on general Riemannian manifolds, our approach is local in nature. On so-called Hadamard manifolds, however, it is global.

The work closest to ours is Ahmadi Kakavandi, Amini, 2010, who introduce a Fenchel conjugacy-like concept on Hadamard metric spaces, using a quasilinearization map in terms of distances as the duality product. In contrast, our work makes use of intrinsic tools from differential geometry such as geodesics, tangent and cotangent vectors to establish a conjugation scheme which extends the theory from locally convex vector spaces to Riemannian manifolds. We investigate the application of the correspondence of a primal problem

$$\text{Minimize} \quad F(p) + G(\Lambda(p))$$

to a suitably defined dual and derive a primal-dual algorithm on Riemannian manifolds. In the absence of a concept of linear operators between manifolds we follow the approach of Valkonen, 2014 and state an exact and a linearized variant of our newly established Riemannian Chambolle–Pock algorithm (RCPA). We then study convergence of the latter on Hadamard manifolds.

As an example, we detail the algorithm for the anisotropic and isotropic total variation with squared distance data term, i.e., the variants of the ROF model on Riemannian manifolds. After illustrating the correspondence to the Euclidean (classical) Chambolle–Pock algorithm, we compare the numerical performance of the RCPA to the CPPA and the PDRA. While the latter has only been shown to converge on Hadamard manifolds of constant curvature, it performs quite well on Hadamard manifolds in general. On the other hand, the CPPA is known to converge possibly arbitrarily slowly; even in the Euclidean case. We illustrate that our linearized algorithm competes with the PDRA, and it even performs favorably on manifolds with non-negative curvature, like the sphere.

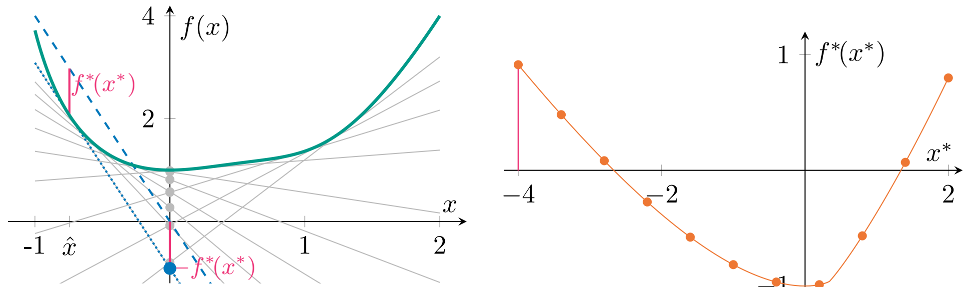
The remainder of the paper is organized as follows. In Section 2 we recall a number of classical results from convex analysis in $\mathbb{R}^d$. In an effort to make the paper self-contained, we also briefly state the required concepts from differential geometry. Section 3 is devoted to the development of a complete Fenchal conjugation scheme for functions defined on manifolds. To this end, we extend some classical results from convex analysis and locally convex vector spaces to manifolds, like the Fenchel–Moreau Theorem (also known as the Biconjugation Theorem) and useful characterizations of the subdifferential in terms of the conjugate function. In Section 4 we formulate the primal-dual hybrid gradient method (also referred to as the Riemannian Chambolle–Pock algorithm, RCPA) for general optimization problems on manifolds involving non-linear operators. We present an exact and a linearized formulation of this novel method and prove, under suitable assumptions, convergence for the linearized variant to a minimizer of a linearized problem on arbitrary Hadamard manifolds. As an application of our theory, Section 5 focuses on the analysis of several total variation models on manifolds. In Section 6 we carry out numerical experiments to illustrate the performance of our novel primal-dual algorithm. Finally, we give some conclusions and further remarks on future research in Section 7.

2. Preliminaries on Convex Analysis and Differential Geometry.

In this section we review some well known results from convex analysis in $\mathbb{R}^d$ as well as concepts of differential geometry. We also revisit the intersection of both topics, convex analysis on Riemannian manifolds, including its subdifferential calculus.

2.1. Convex Analysis.

In this subsection let $f\colon \mathbb{R}^d \to \overline{\mathbb{R}}$, where $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ denotes the extended real line. For standard definitions like *closedness, properness,*



(a) The function f (solid) and the linear function of $x^* = -4$ (dashed) and its shifted tangent (dotted).

(b) The Fenchel conjugate f^*.

Fig. 1: Illustration of the Fenchel conjugate as an interpretation by the tangents of slope x^*.

lower semicontinuity (lsc) and *convexity* of f we refer the reader, e.g., to the textbooks Rockafellar, 1970; Bauschke, Combettes, 2011. The Euclidean inner product is denoted by $\langle \cdot, \cdot \rangle$.

THEOREM 2.1 (Rockafellar, 1970, Thm. 12.1). *A closed, convex function $f\colon \mathbb{R}^d \to \overline{\mathbb{R}}$ is the pointwise supremum of the collection of all affine functions $h\colon \mathbb{R}^d \to \mathbb{R}$ such that $h \le f$.*

COROLLARY 2.2 (Rockafellar, 1970, Cor. 12.1.2). *Given a proper, convex function $f\colon \mathbb{R}^d \to \overline{\mathbb{R}}$, there exists some $b \in \mathbb{R}^d$ and $\beta \in \mathbb{R}$ such that $f(x) \ge \langle x, b \rangle - \beta$ for every $x \in \mathbb{R}^d$.*

DEFINITION 2.3. *The* Fenchel conjugate *of a function $f\colon \mathbb{R}^d \to \overline{\mathbb{R}}$ is defined as the function $f^*\colon \mathbb{R}^d \to \overline{\mathbb{R}}$ such that*

$$f^*(x^*) \coloneqq \sup_{x \in \mathbb{R}^d} \{\langle x^*, x \rangle - f(x)\}. \tag{2.1}$$

We recall some properties of the classical Fenchel conjugate function in the following lemma.

LEMMA 2.4 (Bauschke, Combettes, 2011, Chap. 13). *Let $f, g\colon \mathbb{R}^d \to \overline{\mathbb{R}}$ be proper functions, $\alpha \in \mathbb{R}$ and $\lambda > 0$. Then the following statements hold.*

i) *f^* is proper, convex and lsc.*
ii) *If $f(x) \le g(x)$ for all $x \in \mathbb{R}^d$, then $f^*(x^*) \ge g^*(x^*)$ for all $x^* \in \mathbb{R}^d$.*
iii) *If $g(x) = f(x) + \alpha$ for all $x \in \mathbb{R}^d$, then $g^*(x^*) = f^*(x^*) - \alpha$ for all $x^* \in \mathbb{R}^d$.*
iv) *If $g(x) = \lambda f(x)$ for all $x \in \mathbb{R}^d$, then $g^*(x^*) = \lambda f^*(x^*/\lambda)$ for all $x^* \in \mathbb{R}^d$.*
v) *If $g(x) = f(x + b)$ for all $x \in \mathbb{R}^d$, then $g^*(x^*) = f^*(x^*) - \langle x^*, b \rangle$*
vi) *The Fenchel–Young inequality holds, i.e., for all $x, x^* \in \mathbb{R}^d$ we have*

$$\langle x^*, x \rangle \le f(x) + f^*(x^*). \tag{2.2}$$

The Fenchel conjugate can be interpreted as a maximum seeking problem on the *epigraph* $\operatorname{epi} f \coloneqq \{(x, \alpha) \in \mathbb{R}^d \times \mathbb{R} \mid f(x) \le \alpha\}$. For the case $d = 1$ and some fixed x^*

the conjugate maximizes the (signed) distance $\langle x^*, x \rangle - f(x)$ of the line of slope x^* to f. For instance, let us focus on the case $x^* = -4$ highlighted in Figure 1a. For the linear functional $g_{x^*}(x) = \langle x^*, x \rangle$ (dashed), the maximal distance is attained at $\hat{x}$. We can find the same value by considering the shifted functional $h_{x^*}(x) = g_{x^*}(x) - f^*(x^*)$ (dotted line) and its negative value at the origin, i.e., $-h_{x^*}(0) = f^*(x^*)$. Furthermore h_{x^*} is actually tangent to f at the aforementioned maximizer $\hat{x}$. The function h_{x^*} also illustrates the shifting property from Lemma 2.4 v) and its linear offset $-\langle x^*, b \rangle$. The overall plot of the Fenchel conjugate f^* over an interval of values x^* is shown in Figure 1b.

We now recall some results related to the definition of the subdifferential of a proper, convex function.

Definition 2.5 (Rockafellar, 1970, Chap. 23). *Let $f\colon \mathbb{R}^d \to \overline{\mathbb{R}}$ be a proper, convex function. Its subdifferential is defined as*

$$\partial f(x) \coloneqq \{x^* \in \mathbb{R}^d \mid f(z) \geq f(x) + \langle x^*, z - x \rangle \text{ for all } z \in \mathbb{R}^d\}. \tag{2.3}$$

Theorem 2.6 (Rockafellar, 1970, Thm. 23.5). *Let $f\colon \mathbb{R}^d \to \overline{\mathbb{R}}$ be a proper, convex function and $x \in \mathbb{R}^d$. Then, $x^* \in \partial f(x)$ holds if and only if*

$$f(x) + f^*(x^*) = \langle x^*, x \rangle. \tag{2.4}$$

Corollary 2.7 (Rockafellar, 1970, Cor. 23.5.1). *Let $f\colon \mathbb{R}^d \to \overline{\mathbb{R}}$ be a closed, proper and convex function and $x^* \in \mathbb{R}^d$. Then $x \in \partial f^*(x^*)$ holds if and only if $x^* \in \partial f(x)$.*

The Fenchel biconjugate $f^{**}\colon \mathbb{R}^d \to \overline{\mathbb{R}}$ of a function $f\colon \mathbb{R}^d \to \overline{\mathbb{R}}$ is given by

$$f^{**}(x) = (f^*)^*(x) = \sup_{x^* \in \mathbb{R}^d} \{\langle x^*, x \rangle - f^*(x^*)\}. \tag{2.5}$$

Finally, we conclude this section with the following result known as the Fenchel–Moreau or Biconjugation Theorem.

Theorem 2.8 (Rockafellar, 1970, Thm. 12.2). *Given a proper function $f\colon \mathbb{R}^d \to \overline{\mathbb{R}}$, the equality $f^{**}(x) = f(x)$ holds if and only if f is closed and convex.*

Done till here.

2.2. Differential Geometry. This section is devoted to the collection of necessary concepts from differential geometry. For details concerning the subsequent definitions, the reader may with to consult do Carmo, 1992; Lee, 2003; Jost, 2017.

Suppose that $\mathcal{M}$ is a d-dimensional smooth manifold. The tangent space at $p \in \mathcal{M}$ is a vector space of dimension d and it is denoted by $\mathcal{T}_p\mathcal{M}$. Elements of $\mathcal{T}_p\mathcal{M}$, i.e., *tangent vectors*, will be denoted by X_p and Y_p etc. or simply X and Y when the base point is clear from the context. The disjoint union of all tangent spaces, i.e.,

$$\mathcal{T}\mathcal{M} \coloneqq \bigcup_{p \in \mathcal{M}} \mathcal{T}_p\mathcal{M}, \tag{2.6}$$

is called the *tangent bundle* of $\mathcal{M}$. It is a smooth manifold of dimension $2d$.

The dual space of $\mathcal{T}_p\mathcal{M}$ is denoted by $\mathcal{T}_p^*\mathcal{M}$ and it is called the *cotangent space* to $\mathcal{M}$ at p. The disjoint union

$$\mathcal{T}^*\mathcal{M} \coloneqq \bigcup_{p\in\mathcal{M}} \mathcal{T}_p^*\mathcal{M} \tag{2.7}$$

is known as the *cotangent bundle*. Elements of $\mathcal{T}_p^*\mathcal{M}$ are called *cotangent vectors* to $\mathcal{M}$ at p and they will be denoted by ξ_p and η_p or simply ξ and η. The natural duality product between $X \in \mathcal{T}_p\mathcal{M}$ and $\xi \in \mathcal{T}_p^*\mathcal{M}$ is denoted by $\langle\, \xi, X\,\rangle = \xi(X) \in \mathbb{R}$.

We suppose that $\mathcal{M}$ is equipped with a Riemannian metric, i.e., a smoothly varying family of inner products on the tangent spaces $\mathcal{T}_p\mathcal{M}$. The metric at $p \in \mathcal{M}$ is denoted by $(\,\cdot,\cdot\,)_p\colon \mathcal{T}_p\mathcal{M} \times \mathcal{T}_p\mathcal{M} \to \mathbb{R}$. The induced norm on $\mathcal{T}_p\mathcal{M}$ is denoted by $\|\cdot\|_p$. The Riemannian metric furnishes a linear bijective correspondence between the tangent and cotangent spaces via the Riesz map and its inverse, the so-called *musical isomorphisms*; see Lee, 2003, Chap. 8. They are defined as

$$\flat\colon \mathcal{T}_p\mathcal{M} \ni X \mapsto X^\flat \in \mathcal{T}_p^*\mathcal{M} \tag{2.8}$$

satisfying

$$\langle\, X^\flat, Y\,\rangle = (\,X, Y\,)_p \quad \text{for all } Y \in \mathcal{T}_p\mathcal{M}, \tag{2.9}$$

and its inverse,

$$\sharp\colon \mathcal{T}_p^*\mathcal{M} \ni \xi \mapsto \xi^\sharp \in \mathcal{T}_p\mathcal{M} \tag{2.10}$$

satisfying

$$(\,\xi^\sharp, Y\,)_p = \langle\, \xi, Y\,\rangle \quad \text{for all } Y \in \mathcal{T}_p\mathcal{M}. \tag{2.11}$$

The $\sharp$-isomorphism further introduces an inner product and an associated norm on the cotangent space $\mathcal{T}_p^*\mathcal{M}$, which we will also denote by $(\,\cdot,\cdot\,)_p$ and $\|\cdot\|_p$, since it is clear which inner product or norm we refer to based on the respective arguments.

The tangent vector of a curve $c\colon I \to \mathcal{M}$ defined on some open interval I is denoted by $\dot{c}(t)$. A curve is said to be geodesic if the directional (covariant) derivative of its tangent in the direction of the tangent vanishes, i.e., if $\nabla_{\dot{c}(t)}\dot{c}(t) = 0$ holds for all $t \in I$. As a consequence, geodesic curves have constant speed.

We say that a geodesic connects p to q if $c(0) = p$ and $c(1) = q$ holds. Notice that a geodesic connecting p to q need not always exist, and it if exists, it need not be unique. If a geodesic connecting p to q exists, there also exists a shortest geodesic among them, which may in turn not be unique. If it is, we denote the unique minimal geodesic connecting p and q by $\gamma_{\widehat{p,q}}$.

Using minimal geodesics, one can introduce a notion of metric $d_{\mathcal{M}}$ on $\mathcal{M}$, which may attain the value $+\infty$ if p and q are not on the same connected component of $\mathcal{M}$. As usual, we denote by

$$\mathcal{B}_r(p) \coloneqq \{y \in \mathcal{M} \mid d_{\mathcal{M}}(p,q) < r\} \tag{2.12}$$

the open metric ball of radius $r > 0$ with center $p \in \mathcal{M}$.

We denote by $\gamma_{p,X}\colon (-\varepsilon, \varepsilon) \to \mathcal{M}$ a geodesic starting at p with $\dot\gamma_{p,X}(0) = X$ where $X \in \mathcal{T}_p\mathcal{M}$. We denote the subset of $\mathcal{T}_p\mathcal{M}$ for which these geodesics are well defined until $t = 1$ by $\mathcal{G}_p$. A Riemannian manifold $\mathcal{M}$ is said to be *complete* if $\mathcal{G}_p = \mathcal{T}_p\mathcal{M}$ holds for some, and equivalently for all $p \in \mathcal{M}$.

The *exponential map* is defined as the function $\exp_p\colon \mathcal{G}_p \to \mathcal{M}$ with $\exp_p X \coloneqq \gamma_{p,X}(1)$. Note that $\exp_p(tX) = \gamma_{p,X}(t)$ holds for every $t \in [0,1]$. We further introduce the set $\mathcal{G}'_p \subset \mathcal{T}_p\mathcal{M}$ as some open ball of radius $0 < r \leq \infty$ about the origin such that $\exp_p\colon \mathcal{G}'_p \to \exp_p(\mathcal{G}'_p)$ is a diffeomorphism. The *logarithmic map* is defined as the inverse function of the exponential map, i.e., $\log_p\colon \exp_p(\mathcal{G}'_p) \to \mathcal{G}'_p \subset \mathcal{T}_p\mathcal{M}$.

In the particular case where the sectional curvature of the manifold is nonpositive everywhere, all geodesics connecting distinct points are unique. If furthermore, the manifold is complete, the manifold is called *Hadamard manifold*, see Bačák, 2014b, Def. 1.2.3. Then the exponential and logarithmic maps are defined globally.

Given $p, q \in \mathcal{M}$ and $X \in \mathcal{T}_p\mathcal{M}$, we denote by $\mathcal{P}_{p\to q} X$ the so-called *parallel transport* of X along a unique minimal geodesic $\gamma_{\widehat{p,q}}$. Using the musical isomorphisms presented above, we can also parallelly transport cotangent vectors along geodesics according to

$$\mathcal{P}_{p\to q}\,\xi_p \coloneqq \bigl(\mathcal{P}_{p\to q}\,\xi_p^\sharp\bigr)^\flat. \tag{2.13}$$

Finally, by a Euclidean space we mean $\mathbb{R}^d$ (where $\mathcal{T}_p\mathbb{R} = \mathbb{R}$ holds), equipped with the Riemannian metric given by the Euclidean inner product. In this case, $\exp_p X = p + X$ and $\log_p q = q - p$ hold.

2.3. Convex Analysis on Riemannian Manifolds. Throughout this section, $\mathcal{M}$ is assumed to be a Riemannian manifold. In this subection we recall the basic concepts of convex analysis on manifolds. The central idea is to replace straight lines in the definition of convex sets in Euclidean vector spaces by geodesics.

Definition 2.9 (Sakai, 1996, Def. IV.5.1). *A subset $\mathcal{C} \subset \mathcal{M}$ of a Riemannian manifold $\mathcal{M}$ is said to be* strongly convex *if, for all $p, q \in \mathcal{C}$, minimal geodesics exist, are unique, and lie completely in $\mathcal{C}$.*

Definition 2.10. *Let $\mathcal{C} \subset \mathcal{M}$ and $p \in \mathcal{C}$. We introduce the tangent subset $\mathcal{L}_{\mathcal{C},p} \subset \mathcal{T}_p\mathcal{M}$ as*

$$\mathcal{L}_{\mathcal{C},p} \coloneqq \bigl\{X \in \mathcal{T}_p\mathcal{M} \mid \exp_p X \in \mathcal{C} \text{ and } \|X\|_p = d_{\mathcal{M}}\bigl(\exp_p X, p\bigr)\bigr\},$$

a localized variant of the pre-image of the exponential map.

Note that if $\mathcal{C}$ is strongly convex, the exponential and logarithmic map introduce bijections between $\mathcal{C}$ and $\mathcal{L}_{\mathcal{C},p}$ for any $p \in \mathcal{C}$.

The following definition states the important concept of convex functions on Riemannian manifolds.

Definition 2.11 (Sakai, 1996, Def. IV.5.9).

i) *A function $F\colon \mathcal{M} \to \overline{\mathbb{R}}$ is* proper *if* $\operatorname{dom} F \coloneqq \{x \in \mathcal{M} \mid F(x) < \infty\} \neq \emptyset$ *and $F(x) > -\infty$ holds for all $x \in \mathcal{M}$.*

ii) *Suppose that $\mathcal{C} \subset \mathcal{M}$ is strongly convex. A proper function $F\colon \mathcal{M} \to \overline{\mathbb{R}}$ is called convex on $\mathcal{C} \subset \mathcal{M}$ if for all $p, q \in \mathcal{C}$ the composition $F \circ \gamma_{\widehat{p,q}}$ is a convex function on $[0,1]$ in the classical sense. Similarly F is called strictly or strongly convex if $F \circ \gamma_{\widehat{p,q}}$ fulfills these properties.*

iii) *Suppose that $A \subset \mathcal{M}$. The* epigraph *of a function $F\colon A \to \overline{\mathbb{R}}$ is defined as*

$$\operatorname{epi} F \coloneqq \{(x, \alpha) \in A \times \mathbb{R} \mid F(x) \le \alpha\}. \tag{2.14}$$

iv) *Suppose that $A \subset \mathcal{M}$. A proper function $F\colon A \to \overline{\mathbb{R}}$ is called* lower semicontinuous (lsc) *if* $\operatorname{epi} F$ *is closed.*

A proper function $F\colon \mathcal{C} \to \overline{\mathbb{R}}$ on a strongly convex $\mathcal{C} \subset \mathcal{M}$ is convex if and only if $\operatorname{epi} F \cap (\mathcal{C} \times \mathbb{R})$ is a strongly convex set in the product Riemannian manifold $\mathcal{M} \times \mathbb{R}$. An equivalent way to express iv) is to require that the composition

$$F \circ \exp_m\colon \mathcal{L}_{\mathcal{C},m} \to \overline{\mathbb{R}}$$

is lsc for any $m \in \mathcal{C}$ in the classical sense.

We now recall the notion of the subdifferential of a convex function defined on a Riemannian manifold.

Definition 2.12 (Ferreira, Oliveira, 1998, Udrişte, 1994, Ch. 3, Def. 4.4). *Suppose that $\mathcal{C} \subset \mathcal{M}$ is strongly convex. The* subdifferential $\partial_{\mathcal{M}} F$ *on $\mathcal{C}$ at a point $p \in \mathcal{C}$ of a proper, convex function $F\colon \mathcal{C} \to \overline{\mathbb{R}}$ is given by*

$$\partial_{\mathcal{M}} F(p) \coloneqq \{\xi \in \mathcal{T}_p^* \mathcal{M} \mid F(q) \ge F(p) + \langle \xi, \log_p q \rangle \text{ for all } q \in \mathcal{C}\}. \tag{2.15}$$

In the above notation, the index $\mathcal{M}$ refers to the fact that it is the Riemannian subdifferential; the set $\mathcal{C}$ should always be clear from the context.

Definition 2.13. *Let $\mathcal{C} \subset \mathcal{M}$ be closed. The projection of $p \in \mathcal{M}$ onto $\mathcal{C}$ is defined as the set-valued function $P_{\mathcal{C}}(\cdot)\colon \mathcal{M} \to \mathcal{C}$ such that*

$$P_{\mathcal{C}}(p) \coloneqq \operatorname*{Arg\,min}_{q \in \mathcal{C}} d_{\mathcal{M}}(p, q). \tag{2.16}$$

Note that although the projection mapping is possibly set-valued (even if, additionally, $\mathcal{C}$ is strongly convex), we still have a generalization of the classical variational inequality which characterizes certain solutions of (2.16); see for instance Bauschke, Combettes, 2011, Thm. 3.14 for the Hilbert space case.

Proposition 2.14. *Suppose that $\mathcal{C} \subset \mathcal{M}$ is strongly convex and closed, and $m \in \mathcal{C}$. Then there exists an open ball $\mathcal{B}_r(m)$ such that for any $p \in \mathcal{B}_r(m)$, $m \in P_{\mathcal{C}}(p)$ holds if and only if*

$$(\log_m p, X)_m \le 0 \quad \text{for all } X \in \mathcal{L}_{\mathcal{C} \cap \mathcal{B}_r(m), m}. \tag{2.17}$$

Proof. The proof can be directly adapted from Li et al., 2009, Thm. 3.2, where it was shown even for weakly convex sets. We use the fact that $\log_m$ is well defined near m. ∎

As a corollary, we apply the result above by using the epigraph of an appropriate function as the set $\mathcal{C}$ to project onto. The following two corollaries correspond to Propositions 9.17 and 9.18 from Bauschke, Combettes, 2011, respectively.

COROLLARY 2.15. *Suppose that* $\mathcal{C} \subset \mathcal{M}$ *is strongly convex, and* $m \in \mathcal{C}$*. Moreover, suppose that* $F \colon \mathcal{C} \to \overline{\mathbb{R}}$ *is proper, convex and lsc. Then there exists an open ball* $\mathcal{B}_r(m)$ *such that the following holds. Whenever* $p \in \mathcal{C} \cap \mathcal{B}_r(m)$*, then* $(m, \beta) \in P_{\operatorname{epi} F}(p, \alpha)$ *if and only if* $\max\{\alpha, F(m)\} \le \beta$ *and*

$$(\log_m p, X)_m + (F(\exp_m X) - \beta)(\alpha - \beta) \le 0 \quad \text{for all } X \in \mathcal{L}_{\mathcal{C}\cap\mathcal{B}_r(m),m}. \tag{2.18}$$

COROLLARY 2.16. *Suppose that* $\mathcal{C} \subset \mathcal{M}$ *is strongly convex, and* $m \in \mathcal{C}$*. Moreover, suppose that* $F \colon \mathcal{C} \to \overline{\mathbb{R}}$ *is proper, convex and lsc. Then there exists an open ball* $\mathcal{B}_r(m)$ *such that the following holds. Whenever* $p \in \mathcal{C} \cap \mathcal{B}_r(m)$ *and* $\alpha < F(p) < \infty$*, then* $(m, \beta) \in P_{\operatorname{epi} F}(p, \alpha)$ *if and only if* $\alpha < F(m) = \beta$ *and*

$$(\log_m p, X)_m \le (F(m) - \alpha)(F(\exp_m X) - F(m)) \quad \text{for all } X \in \mathcal{L}_{\mathcal{C}\cap\mathcal{B}_r(m),m}. \tag{2.19}$$

Remark 2.17. Due to Li et al., 2009, Cor. 3.1, the open ball $\mathcal{B}_r(m)$ can be chosen in a way that $\mathcal{C} \subset \mathcal{B}_r(m)$. Hence (2.19) becomes

$$(\log_m p, X)_m \le (F(m) - \alpha)(F(\exp_m X) - F(m)) \quad \text{for all } X \in \mathcal{L}_{\mathcal{C},m}. \tag{2.20}$$

We further recall the definition of the proximal map, that was generalized to Hadamard manifolds in Ferreira, Oliveira, 2002.

DEFINITION 2.18. *Let* $\mathcal{M}$ *be a Riemannian manifold,* $F \colon \mathcal{M} \to \overline{\mathbb{R}}$ *be proper, and* $\lambda > 0$*. The* proximal map *of* F *is defined as*

$$\operatorname{prox}_{\lambda F}(p) \coloneqq \operatorname*{Arg\,min}_{q\in\mathcal{M}} \Big\{ \frac{1}{2} d^2_{\mathcal{M}}(p,q) + \lambda F(q) \Big\}. \tag{2.21}$$

Note that on Hadamard manifolds, the proximal map is single-valued for proper convex functions; see Bačák, 2014b, Chap. 2.2 or Ferreira, Oliveira, 2002, Lem. 4.2 for details. The following lemma is used later on to characterize the proximal map using the subdifferential on Hadamard manifolds.

LEMMA 2.19 (Ferreira, Oliveira, 2002, Lem. 4.2). *Let* $F \colon \mathcal{M} \to \overline{\mathbb{R}}$ *be a proper, convex function on the Hadamard manifold* $\mathcal{M}$*. Then the equality* $q = \operatorname{prox}_{\lambda F}(p)$ *is equivalent to*

$$\frac{1}{\lambda}(\log_q p)^\flat \in \partial_{\mathcal{M}} F(q). \tag{2.22}$$

Remark 2.20. When $F \colon \mathcal{C} \to \overline{\mathbb{R}}$ is defined on a strongly convex set $\mathcal{C} \subset \mathcal{M}$, the characterization Lemma 2.19 continues to hold. The main reason is that Proposition 2.14 replaces Ferreira, Oliveira, 2002, Cor. 3.1 in this setting, and that there exists a unique geodesic connecting two points in $\mathcal{C}$ due to its strong convexity.

3. Fenchel Conjugation Scheme on Manifolds. In this section we present a novel Fenchel conjugation scheme for extended real-valued functions defined on manifolds. We generalize ideas from Bertsekas, 1978, who defined local conjugation on manifolds embedded in $\mathbb{R}^d$ specified by nonlinear equality constraints.

Throughout this section, suppose that $\mathcal{M}$ is a Riemannian manifold and $\mathcal{C} \subset \mathcal{M}$ is strongly convex. The definition of the Fenchel conjugate of F is motivated by Theorem 2.1.

DEFINITION 3.1. *Suppose that $F \colon \mathcal{C} \to \overline{\mathbb{R}}$ and $m \in \mathcal{C}$. The m-*Fenchel conjugate *of F is defined as the function $F_m^* \colon \mathcal{T}_m^*\mathcal{M} \to \overline{\mathbb{R}}$ such that*

$$F_m^*(\xi_m) \coloneqq \sup_{X \in \mathcal{L}_{\mathcal{C},m}} \bigl\{ \langle \xi_m, X \rangle - F(\exp_m X) \bigr\}, \quad \xi_m \in \mathcal{T}_m^*\mathcal{M}. \tag{3.1}$$

Remark 3.2. Note that the Fenchel conjugate F_m^* depends on both the strongly convex set $\mathcal{C}$ and on the base point m. Observe as well that when $\mathcal{M}$ is a Hadamard manifold, it is possible to have $\mathcal{C} = \mathcal{M}$. In the particular case of the Euclidean space $\mathcal{M} = \mathcal{C} = \mathbb{R}^d$, Definition 3.1 becomes

$$F_m^*(\xi) = \sup_{X \in \mathbb{R}^d} \{ \langle \xi, X \rangle - F(m + X) \} = \sup_{Y \in \mathbb{R}^d} \{ \langle \xi, Y - m \rangle - F(Y) \} = F^*(\xi) - \langle \xi, m \rangle$$

for $\xi \in \mathbb{R}^d$. Hence, taking m to be the zero vector we recover the classical (Euclidean) conjugate F^* from Definition 2.3.

EXAMPLE 3.3. *Let $\mathcal{M}$ be a Hadamard manifold, $m \in \mathcal{M}$ and $F \colon \mathcal{M} \to \mathbb{R}$ defined as $F(p) = \frac{1}{2} d^2(p, m)$. Due to the fact that*

$$F(p) = \frac{1}{2} d^2(p, m) = \|\log_m p\|_m^2,$$

we obtain from Definition 3.1 the following representation of the m-conjugate of F:

$$\begin{aligned} F_m^*(\xi_m) &= \sup_{X \in \mathcal{T}_m\mathcal{M}} \Bigl\{ \langle \xi_m, X \rangle - \frac{1}{2} \|\log_m \exp_m X\|_m^2 \Bigr\} \\ &= \sup_{X \in \mathcal{T}_m\mathcal{M}} \Bigl\{ \langle \xi_m, X \rangle - \frac{1}{2} \|X\|_m^2 \Bigr\} \\ &= \frac{1}{2} \|\xi_m\|_m^2. \end{aligned}$$

Notice that the conjugate w.r.t. base points other than m does not have a similarly simple expression. In the Euclidean setting with $\mathcal{M} = \mathbb{R}^d$ and $F(p) = \frac{1}{2}\|p - m\|^2$, it is well known that

$$F_0^*(\xi) = F^*(\xi) = \frac{1}{2}\|\xi + m\|^2 - \frac{1}{2}\|m\|^2$$

holds and thus, by Remark 3.2,

$$F_m^*(\xi) = F^*(\xi) - \langle \xi, m \rangle = \frac{1}{2}\|\xi\|^2$$

in accordance with the expression obtained above.

We now establish a result regarding the properness of the m-conjugate function.

LEMMA 3.4. *Suppose that $F \colon \mathcal{C} \to \overline{\mathbb{R}}$ and $m \in \mathcal{C}$. If F_m^* is proper, then F is also proper.*

Proof. Since F_m^* is proper by assumption we can suppose without loss of generality that $\xi_m \in \operatorname{dom} F_m^*$. Hence, applying Definition 3.1 we get

$$\sup_{X \in \mathcal{L}_{\mathcal{C},m}} \{\langle \xi_m, X \rangle - F(\exp_m X)\} < +\infty,$$

so there must exist at least one $\bar{X} \in \mathcal{L}_{\mathcal{C},m}$ such that $F(\exp_m \bar{X}) \in \mathbb{R}$. This shows that $F \not\equiv +\infty$. On the other hand, let $p \in \mathcal{C}$ and take $X \coloneqq \log_m p$. If $F(p)$ were equal to $-\infty$, then $F_m^*(\xi_m) = +\infty$ for any $\xi_m \in \mathcal{T}_m^*\mathcal{M}$, which would contradict the properness of F_m^*. ∎

Definition 3.5. *Suppose that* $F\colon \mathcal{C} \to \overline{\mathbb{R}}$ *and* $m, m' \in \mathcal{C}$. *Then the* (mm')-Fenchel biconjugate function $F_{mm'}^{**}\colon \mathcal{C} \to \mathbb{R}$ *is defined as*

$$F_{mm'}^{**}(p) = \sup_{\xi_{m'} \in \mathcal{T}_{m'}^*\mathcal{M}} \bigl\{\langle \xi_{m'}, \log_{m'} p \rangle - F_m^*(\mathcal{P}_{m' \to m}\, \xi_{m'})\bigr\}, \quad p \in \mathcal{C}. \tag{3.2}$$

Lemma 3.6. *Suppose that* $F\colon \mathcal{C} \to \overline{\mathbb{R}}$ *and* $m \in \mathcal{C}$. *Then* $F_{mm}^{**}(p) \le F(p)$ *holds for all* $p \in \mathcal{C}$.

Proof. Applying (3.2), we have

$$\begin{aligned} F_{mm}^{**}(p) &= \sup_{\xi_m \in \mathcal{T}_m^*\mathcal{M}} \{\langle \xi_m, \log_m p \rangle - F_m^*(\xi_m)\} \\ &= \sup_{\xi_m \in \mathcal{T}_m^*\mathcal{M}} \{\langle \xi_m, \log_m p \rangle - \sup_{X \in \mathcal{L}_{\mathcal{C},m}} \{\langle \xi_m, X \rangle - F(\exp_m X)\}\} \\ &= \sup_{\xi_m \in \mathcal{T}_m^*\mathcal{M}} \{\langle \xi_m, \log_m p \rangle + \inf_{X \in \mathcal{L}_{\mathcal{C},m}} \{-\langle \xi_m, X \rangle + F(\exp_m X)\}\} \\ &\le \sup_{\xi_m \in \mathcal{T}_m^*\mathcal{M}} \{\langle \xi_m, \log_m p \rangle - \langle \xi_m, \log_m p \rangle + F\bigl(\exp_m(\log_m p)\bigr)\} \\ &= F(p). \end{aligned}$$ ∎

The following lemma collects the properties i)–iv) from Lemma 2.4 that Definition 3.1 inherits from the classical setting in $\mathbb{R}^d$.

Lemma 3.7. *Let* $F, G\colon \mathcal{C} \to \overline{\mathbb{R}}$ *be two proper functions and suppose that* $m \in \mathcal{C}$, $\alpha \in \mathbb{R}$ *and* $\lambda > 0$. *Then the following statements hold.*

i) *The* m*-conjugate* F_m^* *is proper, convex and lsc.*
ii) *If* $F(p) \le G(p)$ *for all* $p \in \mathcal{C}$, *then* $F_m^*(\xi_m) \ge G_m^*(\xi_m)$ *for all* $\xi_m \in \mathcal{T}_m^*\mathcal{M}$.
iii) *If* $G(p) = F(p) + \alpha$ *for all* $p \in \mathcal{C}$, *then* $G_m^*(\xi_m) = F_m^*(\xi_m) - \alpha$ *for all* $\xi_m \in \mathcal{T}_m^*\mathcal{M}$.
iv) *If* $G(p) = \lambda\, F(p)$ *for all* $p \in \mathcal{C}$, *then* $G_m^*(\xi_m) = \lambda\, F_m^*\bigl(\frac{\xi_m}{\lambda}\bigr)$ *for all* $\xi_m \in \mathcal{T}_m^*\mathcal{M}$.
v) *It holds* $F_{mmm}^{***}(\xi_m) = F_m^*(\xi_m)$ *for all* $\xi_m \in \mathcal{T}_m^*\mathcal{M}$.

Proof. Let us start with ii) since i) is a direct application of Bauschke, Combettes, 2011, Prop. 13.11. Observe that F_m^* is defined on a vector space. If $F(p) \le G(p)$ for all $p \in \mathcal{C}$, then it also holds $F(\exp_m X) \le G(\exp_m X)$ for every $X \in \mathcal{L}_{\mathcal{C},m}$. Then we have for any $\xi_m \in \mathcal{T}_m^*\mathcal{M}$ that

$$\begin{aligned} F_m^*(\xi_m) &= \sup_{X \in \mathcal{L}_{\mathcal{C},m}} \{\langle \xi_m, X \rangle - F(\exp_m X)\} \\ &\ge \sup_{X \in \mathcal{L}_{\mathcal{C},m}} \{\langle \xi_m, X \rangle - G(\exp_m X)\} = G_m^*(\xi_m). \end{aligned}$$

Similarly, we prove iii): let us suppose that $G(p) = F(p) + \alpha$ for all $p \in \mathcal{C}$. Then $G(\exp_m X) = F(\exp_m X) + \alpha$ for every $X \in \mathcal{L}_{\mathcal{C},m}$. Hence, for any $\xi_m \in \mathcal{T}_m^*\mathcal{M}$ we obtain

$$\begin{aligned} G_m^*(\xi_m) &= \sup_{X \in \mathcal{L}_{\mathcal{C},m}} \{\langle \xi_m, X \rangle - G(\exp_m X)\} \\ &= \sup_{X \in \mathcal{L}_{\mathcal{C},m}} \{\langle \xi_m, X \rangle - (F(\exp_m X) + \alpha)\} \\ &= \sup_{X \in \mathcal{L}_{\mathcal{C},m}} \{\langle \xi_m, X \rangle - F(\exp_m X)\} - \alpha = F_m^*(\xi_m) - \alpha. \end{aligned}$$

Let us now prove iv) and suppose that $\lambda > 0$ and $G(\exp_m X) = \lambda\, F(\exp_m X)$ for all $X \in \mathcal{L}_{\mathcal{C},m}$. Then we have for any $\xi_m \in \mathcal{T}_m^*\mathcal{M}$ that

$$\begin{aligned} G_m^*(\xi_m) &= \sup_{X \in \mathcal{L}_{\mathcal{C},m}} \{\langle \xi_m, X \rangle - G(\exp_m X)\} \\ &= \sup_{X \in \mathcal{L}_{\mathcal{C},m}} \{\langle \xi_m, X \rangle - \lambda\, F(\exp_m X)\} \\ &= \lambda \sup_{X \in \mathcal{L}_{\mathcal{C},m}} \Big\{\langle \frac{\xi_m}{\lambda}, X \rangle - F(\exp_m X)\Big\} = \lambda\, F_m^*\big(\tfrac{\xi_m}{\lambda}\big). \end{aligned}$$

Finally, we prove v). Applying Lemma 3.6 to F, we get $F_{mm}^{**} \le F$, and due to ii) we get $F_{mmm}^{***} \ge F_m^*$. The converse inequality is ensured by a direct application of Bauschke, Combettes, 2011, Prop. 13.14 to F_m^*. ∎

Remark 3.8. Observe that an analogue of v) in Lemma 2.4 cannot be expected for $F\colon \mathcal{M} \to \mathbb{R}$ due to the lack of a concept of linearity on manifolds.

PROPOSITION 3.9 (Fenchel–Young inequality). *Suppose that $F\colon \mathcal{C} \to \overline{\mathbb{R}}$ and $m \in \mathcal{C}$. Then*

$$F(p) + F_m^*(\xi_m) \ge \langle \xi_m, \log_m p \rangle \tag{3.3}$$

holds for all $p \in \mathcal{C}$ and $\xi_m \in \mathcal{T}_m^*\mathcal{M}$.

Proof. Suppose that $\xi_m \in \mathcal{T}_m^*\mathcal{M}$, $p \in \mathcal{C}$ and set $X \coloneqq \log_m p$. From Definition 3.1 we obtain

$$F_m^*(\xi_m) \ge \langle \xi_m, \log_m p \rangle - F\big(\exp_m(\log_m p)\big),$$

which is equivalent to (3.3). ∎

We continue introducing the manifold counterpart of the Fenchel–Moreau Theorem; compare Theorem 2.8.

THEOREM 3.10. *Let $F\colon \mathcal{C} \to \overline{\mathbb{R}}$ be a proper, convex function and $m \in \mathcal{C}$. Then the following statements hold.*

i) *$F(p) = F_{mm}^{**}(p)$ for all $p \in \mathcal{C}$ if and only if F is lsc on $\mathcal{C}$.*
ii) *F_{mm}^{**} is proper if F is.*

Proof. Suppose that $\mathcal{B}_r(m) \supset \mathcal{C}$ is as in Remark 2.17. This implies, in particular, that $\log_m p$ is well-defined for all $p \in \mathcal{C}$. We will prove the result along the lines of Bauschke, Combettes, 2011, Thm. 13.32.

In order to show i), let us first suppose that $F(p) = F^{**}_{mm}(p)$ for all $p \in \mathcal{C}$. Due to Lemma 3.7 i), F is proper, convex and lsc since it is the m-Fenchel conjugate of F^*_m.

Let us prove the converse. To this end, recall that $\mathcal{C} \subset \mathcal{B}_r(m)$. Let F be a proper, convex and lsc function on $\mathcal{C}$. Fix any $p \in \mathcal{C}$ and $\alpha < F(p)$, and set $(m, \beta) \in P_{\operatorname{epi} F}(p, \alpha)$. From Corollary 2.15 we get $\alpha \le F(m) = \beta$ and

$$\langle \xi_m, X \rangle \le (\beta - \alpha)(F(\exp_m X) - \beta) \quad \text{for all } X \in \mathcal{L}_{\mathcal{C},m}, \tag{3.4}$$

with $\xi_m = (\log_m p)^\flat$.

If $\beta > \alpha$, taking the supremum over $X \in \mathcal{L}_{\mathcal{C},m}$ and choosing $\xi'_m = \frac{1}{\beta-\alpha}\xi_m \in (\mathcal{L}_{\mathcal{B}_r(m),m})^\flat$, we get

$$F^*_m(\xi'_m) \le -\beta + (\, \eta_m, \eta_m \,)_m, \qquad \text{for all } \eta_m \in (\mathcal{L}_{\mathcal{B}_r(m),m})^\flat. \tag{3.5}$$

In particular, it also holds for $\eta_m = \xi_m$ and since $\beta > \alpha$,

$$F^*_m(\xi'_m) \le -\beta + \left(\, \tfrac{1}{\beta-\alpha}\xi_m, \xi_m \,\right)_m. \tag{3.6}$$

Now, we use the isometry of the musical isomorphisms (2.8) and (2.10) to obtain

$$(\, \tfrac{1}{\beta-\alpha}\xi_m, \xi_m \,)_m = \left(\left(\tfrac{1}{\beta-\alpha}\xi_m\right)^\sharp, \xi_m^\sharp \right)_m = \langle\, \xi'_m, \xi_m^\sharp \,\rangle = \langle\, \xi'_m, \log_m p \,\rangle.$$

Then, (3.6) becomes

$$F^*_m(\xi'_m) \le -\beta + \langle\, \xi'_m, \log_m p \,\rangle.$$

Rearranging terms, we get

$$\beta \le \langle\, \xi'_m, \log_m p \,\rangle - F^*_m(\xi'_m) \le F^{**}_{mm}(p).$$

Hence, if $\beta > \alpha$, we obtain

$$F^{**}_{mm}(p) \ge \beta > \alpha. \tag{3.7}$$

Let us show that $F^{**}_{mm}(p) = F(p)$ holds. Since $\operatorname{dom} F \cap \mathcal{C} \neq \emptyset$, we first look at $p \in \operatorname{dom} F \cap \mathcal{C}$. Since $(m, \beta) \in P_{\operatorname{epi} F}(p, \alpha)$, $\beta = F(m) > \alpha$, and due to Lemma 3.6 and (3.7) it follows that $F(p) \ge F^{**}_{mm}(p) > \alpha$. Since $\alpha < F(p)$ and the fact that α can be chosen arbitrarily, we finally get $F^{**}_{mm}(p) = F(p)$ on $\operatorname{dom} F \cap \mathcal{C}$. Let us now suppose that $p \notin \operatorname{dom} F$ but $p \in \mathcal{C}$. If $\beta > \alpha$, by (3.7) we get $F^{**}_{mm} > \alpha$ and since α can be chosen arbitrarily, $F^{**}_{mm}(p) = F(p) = +\infty$.

Otherwise, if $\beta = \alpha$ and due to the fact that $(p, \alpha) \notin \operatorname{epi} F$ but $(m, \beta) \in \operatorname{epi} F$, we have $d_{\mathcal{M}}(p, m) > 0$ so the tangent vector $\log_m p \in \mathcal{L}_{\mathcal{C},m}$ has positive length, i.e., $\lVert \log_m p \rVert_m > 0$. Now, fix $\xi_m \in \operatorname{dom} F^*_m$ with $\xi_m \in \mathcal{L}_{\mathcal{B}_r(m),m}$. Observe that if $\xi_m \in \mathcal{L}_{\mathcal{B}_r(m),m'}$ with $m \neq m'$, the biconjugate would be $F^{**}_{mm'}$ and we want to recover F^{**}_{mm}. Then, by Definition 3.1 and using Proposition 2.14 we obtain

$$\begin{gathered} \langle\, \xi_m, X \,\rangle - F(\exp_m X) \le F^*_m(\xi_m), \\ \langle\, \xi_m, X \,\rangle \le 0 \end{gathered} \tag{3.8}$$

for all $X \in \mathcal{L}_{\mathcal{C},m}$ for which we recall that $\xi_m \in \mathcal{L}_{\mathcal{B}_r(m),m}$ such that $\xi_m^\sharp = \log_m p$. Taking $\mu > 0$ and combining the two inequalities from (3.8) we get

$$\langle \xi_m, X \rangle + \langle \mu\, \xi_m, X \rangle - F(\exp_m X) \leq F_m^*(\xi_m), \tag{3.9}$$

for every (fixed) $\xi_m \in \operatorname{dom} F_m^*$ with $\xi_m \in \mathcal{L}_{\mathcal{B}_r(m),m}$. We analyze (3.9) and we get for all $X \in \mathcal{L}_{\mathcal{C},m}$ that

$$\begin{aligned}\langle \xi_m\,(1+\mu), X \rangle - F(\exp_m X)& \\ &\leq F_m^*(\xi_m) + \langle (1+\mu)\,\xi_m, \log_m p \rangle - \langle (1+\mu)\,\xi_m,\, \log_m p\rangle\,.\end{aligned}$$

Taking the supremum over $X \in \mathcal{L}_{\mathcal{C},m}$ yields

$$F_m^*((1+\mu)\,\xi_m) \leq F_m^*(\xi_m) + \langle (1+\mu)\,\xi_m, \log_m p \rangle - \langle (1+\mu)\,\xi_m, \log_m p \rangle,$$

and rearranging the above inequality, we obtain

$$\langle (1+\mu)\,\xi_m, \log_m p \rangle - F_m^*((1+\mu)\,\xi_m) \geq \langle (1+\mu)\,\xi_m, \log_m p \rangle - F_m^*(\xi_m),$$

which implies that

$$F_{mm}^{**}(p) \geq \langle (1+\mu)\,\xi_m, \log_m p \rangle - F_m^*(\xi_m).$$

Analyzing the summand $\langle (1+\mu)\,\xi_m, \log_m p \rangle$, we get

$$\begin{aligned}\langle (1+\mu)\,\xi_m, \log_m p \rangle &= (1+\mu)\langle \xi_m, \log_m p \rangle = (1+\mu)(\, \xi_m^\sharp, \log_m p\,)_m \\ &= (1+\mu)(\log_m p, \log_m p\,)_m = (1+\mu)\,\|\log_m p\|_m > 0.\end{aligned}$$

Thus, the positivity of $\langle (1+\mu)\,\xi_m, \log_m p \rangle$ implies that $F_{mm}^{**}(p) = F(p) = \infty$ also in the last case where $p \notin \operatorname{dom} F$ but $p \in \mathcal{C}$.

Finally, part ii) of the theorem is a straightforward application of Lemma 3.6. ∎

We now address the manifold counterpart of Theorem 2.6.

THEOREM 3.11. *Let $F\colon \mathcal{C} \to \overline{\mathbb{R}}$ be a proper, convex function and $m, p \in \mathcal{C}$. Then, $\xi_p \in \partial_{\mathcal{M}} F(p)$ holds if and only if*

$$F(p) + F_m^*(\mathcal{P}_{p\to m}\,\xi_p) = \langle \mathcal{P}_{p\to m}\,\xi_p, \log_m p \rangle. \tag{3.10}$$

Proof. From (3.10), we get

$$\begin{aligned}F(p) + F_m^*(\mathcal{P}_{p\to m}\,\xi_p) &= \langle \mathcal{P}_{p\to m}\,\xi_p, \log_m p \rangle = \langle \mathcal{P}_{m\to p}\,\mathcal{P}_{p\to m}\,\xi_p, \mathcal{P}_{m\to p}\log_m p \rangle \\ &= \langle \xi_p, -\log_p m \rangle = -\langle \xi_p, \log_p m \rangle\end{aligned}$$

and hence

$$F(p) + \langle \xi_p, \log_p m \rangle = -F_m^*(\mathcal{P}_{p\to m}\,\xi_p). \tag{3.11}$$

On the other hand, for every $m \in \mathcal{M}$,

$$\begin{aligned}-F(m) &= \langle \mathcal{P}_{p\to m}\,\xi_p, 0_m \rangle - F(\exp_m 0_m) \\ &\leq \sup_{X \in \mathcal{L}_{\mathcal{C},m}} \{\langle \mathcal{P}_{p\to m}\,\xi_p, X \rangle - F(\exp_m X)\} = F_m^*(\mathcal{P}_{p\to m}\,\xi_p),\end{aligned}$$

and hence

$$F(m) \geq -F_m^*(\mathcal{P}_{p\to m}\,\xi_p). \tag{3.12}$$

Combining (3.11) and (3.12), we get that

$$F(p) + \langle\, \xi_p, \log_p m \,\rangle = -F_m^*(\mathcal{P}_{p\to m}\,\xi_p) \leq F(m), \tag{3.13}$$

holds true for all $m \in \mathcal{C}$, or equivalently, $\xi_p \in \partial_{\mathcal{M}} F(p)$ according to Definition 2.12. ∎

Given $F\colon \mathcal{C} \to \overline{\mathbb{R}}$ and $m \in \mathcal{C}$, we can state the subdifferential from Definition 2.12 for the conjugate function $F_m^*\colon \mathcal{T}_m^*\mathcal{M} \to \mathbb{R}$,

$$\begin{aligned}\partial_{\mathcal{M}} F_m^*(\xi_m) &\\ \coloneqq \{X \in \mathcal{T}_m\mathcal{M} \mid F_m^*(\eta_m) &\geq F_m^*(\xi_m) + \langle\, X, \eta_m - \xi_m \,\rangle \text{ for all } \eta_m \in \mathcal{T}_m^*\mathcal{M}\}.\end{aligned} \tag{3.14}$$

Before providing the manifold counterpart of Corollary 2.7, let us show how Theorem 3.11 looks like for F_m^* instead of F.

Corollary 3.12. *Let $F\colon \mathcal{C} \to \overline{\mathbb{R}}$ be a proper, convex and lsc function and m, $p \in \mathcal{C}$. Then for all $\xi_m \in \mathcal{T}_m^*\mathcal{M}$ it holds*

$$\log_m p \in \partial F_m^*(\xi_m) \Leftrightarrow F_m^*(\xi_m) + F(p) = \langle\, \xi_m, \log_m p \,\rangle. \tag{3.15}$$

Proof. Similarly as in the proof of Theorem 3.11, for every $\eta_m \in \mathcal{T}_m^*\mathcal{M}$, it holds

$$\begin{aligned}\log_m p \in \partial_{\mathcal{M}} F_m^*(\xi_m) &\Leftrightarrow F_m^*(\eta_m) \geq F_m^*(\xi_m) + \langle\, \log_m p, \eta_m - \xi_m \,\rangle\\ &\Leftrightarrow -F_m^*(\xi_m) + \langle\, \log_m p, \xi_m \,\rangle \geq \langle\, \log_m p, \eta_m \,\rangle - F_m^*(\eta_m).\end{aligned}$$

Taking the supremum over η_m, we get

$$\begin{aligned}\log_m p \in \partial_{\mathcal{M}} F_m^*(\xi_m) &\Leftrightarrow -F_m^*(\xi_m) + \langle\, \log_m p, \xi_m \,\rangle \geq F_{mm}^{**}(p)\\ &\Leftrightarrow \langle\, \log_m p, \xi_m \,\rangle \geq F_m^*(\xi_m) + F_{mm}^{**}(p).\end{aligned}$$

Hence, applying Proposition 3.9 and using that F is proper, convex, and lsc, Theorem 3.10 implies that $F(p) = F_{mm}^{**}(p)$ holds and thus

$$\langle\, \log_m p, \xi_m \,\rangle \geq F_m^*(\xi_m) + F(p) \geq \langle\, \log_m p, \xi_m \,\rangle,$$

and we finally get the right-hand-side of (3.15).

The converse follows in a straightforward way since, in particular, we have

$$\langle\, \log_m p, \xi_m \,\rangle \geq F_m^*(\xi_m) + F_{mm}^{**}(p),$$

for all $\xi_m \in \mathcal{T}_m^*\mathcal{M}$, which completes the proof. ∎

To conclude this section, we state the following result which generalizes Corollary 2.7 and shows the symmetric relation between the conjugate function and the subdifferential when the involved function is proper, convex and lsc.

Corollary 3.13. *Let $F\colon \mathcal{C} \to \overline{\mathbb{R}}$ be a proper, convex and lsc function and $m, p \in \mathcal{C}$. Then*

$$\xi_p \in \partial_{\mathcal{M}} F(p) \Leftrightarrow \log_m p \in \partial F_m^*(\mathcal{P}_{p\to m}\,\xi_p). \tag{3.16}$$

Proof. The proof is a straightforward combination of Theorem 3.11 and taking as a particular cotangent vector $\xi_m = \mathcal{P}_{p\to m}\,\xi_p$ in Corollary 3.12. ∎

4. Optimization on Manifolds. In this section we derive a primal-dual optimization algorithm to solve minimization problems on manifolds of the form

$$\text{Minimize} \quad F(p) + G(\Lambda(p)), \quad p \in \mathcal{C}. \tag{4.1}$$

Here $\mathcal{C} \subset \mathcal{M}$ and $\mathcal{D} \subset \mathcal{N}$ are strongly convex sets, $F\colon \mathcal{C} \to \overline{\mathbb{R}}$ and $G\colon \mathcal{D} \to \overline{\mathbb{R}}$ are proper functions, $\Lambda\colon \mathcal{M} \to \mathcal{N}$ is a general differentiable map such that $\Lambda(\mathcal{C}) \subset \mathcal{D}$. We emphasize that convexity of F or G is not necessarily assumed.

Our algorithm requires a choice of a pair of base points $m \in \mathcal{C}$ and $n \in \mathcal{D}$. The role of m is to serve as a possible linearization point for Λ, while n is the base point of the Fenchel conjugate for G. More generally, the points can be allowed to change during the iterations. We emphasize this possibility by writing $m^{(k)}$ and $n^{(k)}$ when appropriate.

In the particular case that the functions involved in the primal problem (4.1) are proper, convex, and lsc the following saddle-point formulation is equivalent to (4.1),

$$\text{Minimize} \quad \max_{\xi_n \in \mathcal{T}_n^* \mathcal{N}} \langle \log_n \Lambda(p), \xi_n \rangle + F(p) - G_n^*(\xi_n), \quad p \in \mathcal{C}. \tag{4.2}$$

The proof of equivalence uses Theorem 3.10.

The solution of problem (4.2) by primal-dual optimization algorithms is challenging due to the lack of a vector space structure, which implies in particular the absence of a concept of linearity of Λ. This is also the reason why we cannot derive a dual problem associated to (4.1) following the same reasoning as in vector spaces.

Therefore we concentrate on the saddle-point problem. One can show that if $(\widehat{p}, \widehat{\xi_n}) \in \mathcal{C} \times \mathcal{T}_n^* \mathcal{N}$ solves (4.2), then it satisfies the first-order optimality conditions

$$\begin{aligned} -(D\Lambda(\widehat{p}))^*[\mathcal{P}_{n \to \Lambda(\widehat{p})}\, \xi_n] &\in \partial_{\mathcal{M}} F(p), \\ \log_n \Lambda(p) &\in \partial G_n^*(\xi_n). \end{aligned} \tag{4.3}$$

Following the line of arguments in Valkonen, 2014, Sect. 2.1, it is impossible to directly convert (4.5) into a primal-dual algorithm, which would require knowledge of the solution. Instead, we propose to replace $\widehat{p}$ by m, which suggests to consider the system

$$\begin{aligned} \mathcal{P}_{m \to p} \left(-(D\Lambda(m))^*[\mathcal{P}_{n \to \Lambda(m)}\, \xi_n] \right) &\in \partial_{\mathcal{M}} F(p), \\ \log_n \Lambda(p) &\in \partial G_n^*(\xi_n), \end{aligned} \tag{4.4}$$

for the unknowns (p, ξ_n).

Remark 4.1. In the specific case that $\mathcal{X} = \mathcal{M}$ and $\mathcal{Y} = \mathcal{N}$ are Hilbert spaces, $F\colon \mathcal{X} \to \mathbb{R}$ is C^1, $\Lambda\colon \mathcal{X} \to \mathcal{Y}$ is a linear operator, $m = n = 0$, and either $(D\Lambda(m))^*$ has empty null space or $\operatorname{dom} G = \mathcal{Y}$, we observe (similar to Valkonen, 2014) that the conditions (4.4) simplify to

$$\begin{aligned} -\Lambda^* \xi &\in \partial F(p), \\ \Lambda p &\in \partial G^*(\xi), \end{aligned} \tag{4.5}$$

where $p \in \mathcal{X}$ and $\xi \in \mathcal{T}_n^* \mathcal{N} = \mathcal{Y}^*$.

Algorithm 4.1 Exact (primal relaxed) Riemannian Chambolle–Pock for (4.2)

Input: $m \in \mathcal{C}$, $n \in \mathcal{D}$, $p^{(0)} \in \mathcal{C}$, and $\xi_n^{(0)} \in \mathcal{T}_n^*\mathcal{N}$, and parameters $\sigma_0, \tau_0, \theta_0, \gamma$

$k \leftarrow 0, \quad \bar{p}^{(0)} \leftarrow p^{(0)}$

while not converged **do**

$\quad \xi_n^{(k+1)} \leftarrow \operatorname{prox}_{\tau_k G_n^*}\Big(\xi_n^{(k)} + \tau_k \big(\log_n \Lambda(\bar{p}^{(k)})\big)^\flat \Big),$

$\quad p^{(k+1)} \leftarrow \operatorname{prox}_{\sigma_k F}\bigg(\exp_{p^{(k)}}\Big(\mathcal{P}_{m \to p^{(k)}} \big(-\sigma_k (D\Lambda(m))^*[\mathcal{P}_{n \to \Lambda(m)}\, \xi_n^{(k+1)}]\big)^\sharp \Big) \bigg),$

$\quad \theta_k = (1 + 2\gamma\sigma_k)^{-\frac{1}{2}}, \quad \sigma_{k+1} \leftarrow \sigma_k \theta_k, \quad \tau_{k+1} \leftarrow \tau_k / \theta_k$

$\quad \bar{p}^{(k+1)} \leftarrow \exp_{p^{(k+1)}}\big(-\theta_k \log_{p^{(k+1)}} p^{(k)}\big)$

$\quad k \leftarrow k+1$

end while

Output: $p^{(k)}$

4.1. Exact Riemannian Chambolle–Pock. In this subsection we develop the *exact* Riemannian Chambolle–Pock algorithm summarized in Algorithm 4.1. The name "exact", introduced by Valkonen, 2014, refers to the fact that the operator Λ in the dual step is used in its exact form and only the primal step employs a linearization in order to obtain the adjoint $(D\Lambda(m))^*$. Indeed, our Algorithm 4.1 can be seen to generalize Valkonen, 2014, Alg. 2.1.

Let us motivate the formulation of Algorithm 4.1. We start from the second inclusion in (4.4) and obtain, for any $\tau > 0$, the equivalent condition

$$\xi_n + \tau(\log_n \Lambda(p))^\flat \in \xi_n + \big(\tau \partial G_n^*(\xi_n)\big)^\flat = (I + (\tau\partial G_n^*)^\flat)(\xi_n). \tag{4.6}$$

Similarly we obtain that the first inclusion in (4.4) is equivalent to

$$-\frac{1}{\sigma}\big(\sigma \mathcal{P}_{m \to p}\,(D\Lambda(m))^*[\mathcal{P}_{n \to \Lambda(m)}\,\xi_n]\big) \in \partial_{\mathcal{M}} F(p) \tag{4.7}$$

for any $\sigma > 0$. Lemma 2.19 now suggests the following alternating algorithmic scheme

$$\begin{aligned} \xi_n^{(k+1)} &= \operatorname{prox}_{\tau G_n^*}\big(\,\widetilde{\xi}_n^{(k)}\,\big), \\ p^{(k+1)} &= \operatorname{prox}_{\sigma F}\big(\,\widetilde{p}^{(k)}\,\big), \end{aligned}$$

where

$$\widetilde{\xi}_n^{(k)} \coloneqq \xi_n^{(k)} + \tau\Big(\log_n \Lambda(\bar{p}^{(k)})\Big)^\flat, \tag{4.8}$$

$$\widetilde{p}^{(k)} \coloneqq \exp_{p^{(k)}}\Big(\mathcal{P}_{m \to p^{(k)}} - \big(\sigma(D\Lambda(m))^*[\mathcal{P}_{n \to \Lambda(m)}\,\xi_n^{(k+1)}]\big)^\sharp\Big), \tag{4.9}$$

and

$$\bar{p}^{(k+1)} = \exp_{p^{(k+1)}}(-\theta \log_{p^{(k+1)}} p^{(k)}), \tag{4.10}$$

i.e., we perform an over-relaxation of the primal variable. This basic form of an algorithm can be combined with an acceleration by step size selection as described in Chambolle, Pock, 2011, Sec. 5. This is already reflected in Algorithm 4.1.

4.2. Linearized Riemannian Chambolle–Pock. The main obstacle in deriving a complete duality theory for problem (4.2) is the lack of a concept of linearity of operators Λ between manifolds. In the previous section, we chose to linearize Λ in the primal update step only, in order to have an adjoint. By contrast, we now replace it by its first order approximation

$$\Lambda(p) \approx \exp_{\Lambda(m)} D\Lambda(m)[\log_m p] \tag{4.11}$$

throughout. Here $D\Lambda(m)\colon \mathcal{T}_m\mathcal{M} \to \mathcal{T}_{\Lambda(m)}\mathcal{N}$ denotes the derivative of Λ at m. Since $D\Lambda\colon \mathcal{T}\mathcal{M} \to \mathcal{T}\mathcal{N}$ is a linear operator between tangent bundles, we can then utilize the adjoint operator $D\Lambda(m)^*\colon \mathcal{T}^*_{\Lambda(m)}\mathcal{N} \to \mathcal{T}^*_m\mathcal{M}$. We further point out that we can work algorithmically with cotangent vectors $\xi_n \in \mathcal{T}^*_n\mathcal{N}$ with a fixed base point n since, at least locally, we can obtain a cotangent vector $\xi_{\Lambda(m)} \in \mathcal{T}^*_{\Lambda(m)}\mathcal{N}$ from it by parallel transport using $\xi_{\Lambda(m)} = \mathcal{P}_{n\to\Lambda(m)}\,\xi_n$. The duality pairing reads

$$\langle\, D\Lambda(m)[\log_m p], \mathcal{P}_{n\to\Lambda(m)}\,\xi_n \,\rangle = \langle\, \log_m p, (D\Lambda(m))^*[\mathcal{P}_{n\to\Lambda(m)}\,\xi_n] \,\rangle \tag{4.12}$$

for every $p \in \mathcal{C}$ and $\xi_n \in \mathcal{T}^*_n\mathcal{N}$.

We plug in (4.11) into (4.1), which yields the *linearized* primal problem

$$\text{Minimize} \quad F(p) + G(\exp_{\Lambda(m)} D\Lambda(m)[\log_m p]), \quad p \in \mathcal{C}. \tag{4.13}$$

For simplicity, we assume $\Lambda(m) = n$ in this subsection. Hence, the associated *linearized* saddle-point problem reads

$$\text{Minimize} \quad \max_{\xi_n \in \mathcal{T}^*_n\mathcal{N}} \langle\, D\Lambda(m)[\log_m p], \xi_n \,\rangle + F(p) - G^*_n(\xi_n), \quad p \in \mathcal{C}. \tag{4.14}$$

In contrast to (4.1), we are able to also derive a Fenchel dual problem associated with (4.13).

THEOREM 4.2. *The dual problem of* (4.13) *is given by*

$$\textit{Maximize} \quad -F^\star_m(-(D\Lambda(m))^*[\xi_n]) - G^\star_n(\xi_n), \quad \xi_n \in \mathcal{T}^*_n\mathcal{N}. \tag{4.15}$$

Weak duality holds, i.e.,

$$\min_{p\in\mathcal{C}} F(p) + G(\exp_{\Lambda(m)} D\Lambda(m)[\log_m p]) \geq \max_{\xi_n \in \mathcal{T}^*_n\mathcal{N}} -F^\star_m(-(D\Lambda(m))^*[\xi_n]) - G^\star_n(\xi_n). \tag{4.16}$$

Moreover, the optimality conditions for the primal-dual pair (4.13), (4.15) *as well as the saddle-point formulation* (4.14) *are*

$$\begin{aligned} \mathcal{P}_{m\to p}\left(-(D\Lambda(m))^*[\xi_n]\right) &\in \partial_{\mathcal{M}} F(p),\\ D\Lambda(m)[\log_m p] &\in \partial G^*_n(\xi_n). \end{aligned} \tag{4.17}$$

Proof. The proof of (4.15) and (4.16) follows from the application of Zălinescu, 2002, eq. (2.80) and Definition 3.1 in (4.14). From the linearized primal-dual pair (4.13)–(4.15), the development of (4.17) derives from classical duality theory on general vector spaces, see for instance Ekeland, Temam, 1999, Chap. 3.4, and Corollary 3.13. ∎

Algorithm 4.2 Linearized (dual relaxed) Riemannian Chambolle–Pock for (4.14)

Input: $m^{(k)} \in \mathcal{C}$, $n^{(k)} \in \mathcal{D}$, $p^{(0)} \in \mathcal{C}$, $\xi_n^{(0)} \in \mathcal{T}^*_{n^{(0)}}\mathcal{N}$, and parameters $\sigma_0, \tau_0, \theta_0, \gamma$

1: $k \leftarrow 0, \quad \bar{p}^{(0)} \leftarrow p^{(0)}$

2: **while** not converged **do**

3: $\quad p^{(k+1)} \leftarrow \operatorname{prox}_{\sigma_k F}\Bigl(\exp_{p^{(k)}}\Bigl(\mathcal{P}_{m^{(k)}\to p^{(k)}}\bigl(-\sigma_k (D\Lambda(m^{(k)}))^*\bigl[\mathcal{P}_{n^{(k)}\to\Lambda(m^{(k)})}\,\bar{\xi}^{(k)}_{n^{(k)}}\bigr]\bigr)^\sharp\Bigr)\Bigr)$

4: $\quad \xi^{(k+1)}_{n^{(k)}} \leftarrow \operatorname{prox}_{\tau_k G^*_{n^{(k)}}}\Bigl(\xi^{(k)}_{n^{(k)}} + \tau_k\bigl(\mathcal{P}_{\Lambda(m^{(k)})\to n^{(k)}}\, D\Lambda(m^{(k)})[\log_{m^{(k)}} p^{(k+1)}]\bigr)^\flat\Bigr)$

5: $\quad \theta_k = (1+2\gamma\sigma_k)^{-\frac{1}{2}}, \quad \sigma_{k+1} \leftarrow \sigma_k\theta_k, \quad \tau_{k+1} \leftarrow \tau_k/\theta_k$

6: $\quad \bar{\xi}^{(k+1)}_{n^{(k+1)}} \leftarrow \mathcal{P}_{n^{(k)}\to n^{(k+1)}}\bigl(\xi^{(k+1)}_{n^{(k)}} + \theta(\xi^{(k+1)}_{n^{(k)}} - \xi^{(k)}_{n^{(k)}})\bigr)$

7: $\quad \xi^{(k+1)}_{n^{(k+1)}} \leftarrow \mathcal{P}_{n^{(k)}\to n^{(k+1)}}\,\xi^{(k+1)}_{n^{(k)}}$

8: $\quad k \leftarrow k+1$

9: **end while**

Output: $p^{(k)}$

Notice that under the particular scenario stated in Remark 4.1, the first order optimality conditions (4.17) boil down to (4.5). Motivated by the statement of the linearized primal-dual pair (4.13)–(4.15), a further development of duality theory and an investigation of the linearization error is left for future research.

Both the exact and the linearized variants of our Riemannian Chambolle–Pock algorithm (RCPA) can be stated to over-relax either the primal variable as in Algorithm 4.1 or the dual variable as in Algorithm 4.2. In total this yields four possibilities — exact vs. linearized, and primal vs. dual over-relaxation. This generalizes the analogous cases discussed in Valkonen, 2014 for the Hilbert space setting. In each of the four cases it is possible to allow changes in the base points, and $n^{(k)}$ may be equal or different from $\Lambda(m^{(k)})$. Letting $m^{(k)}$ depend on k changes the linearization point of the operator, while allowing $n^{(k)}$ to change introduces different $n^{(k)}$-Fenchel conjugates $G^*_{n^{(k)}}$, and also incurs a parallel transport on the dual variable. These possibilities are reflected in the statement of Algorithm 4.2.

Reasonable choices for the base points include, e.g., to set both $m^{(k)} = m$ and $n^{(k)} = \Lambda(m)$, for $k \geq 0$ and some $m \in \mathcal{M}$. This choice eliminates the parallel transport in the dual update step as well as the most inner parallel transport of the primal update step. Another choice is, to fix just n and set $m^{(k)} = p^{(k)}$, which eliminates the parallel transport in the primal update step. It further eliminates both parallel transports of the dual variable in steps 6 and 7 of Algorithm 4.2.

4.3. Relation to the Chambolle–Pock Algorithm on Hilbert Spaces. In this section we confirm that both Algorithm 4.1 and Algorithm 4.2 boil down to the classical Chambolle–Pock method in Hilbert spaces. To this end, suppose in this section that $\mathcal{M} = \mathcal{X}$ and $\mathcal{N} = \mathcal{Y}$ are finite-dimensional Hilbert spaces with inner products $(\cdot,\cdot)_\mathcal{X}$ and $(\cdot,\cdot)_\mathcal{Y}$, respectively, and that $\Lambda\colon \mathcal{X} \to \mathcal{Y}$ is a linear operator. In Hilbert spaces, geodesics are straight lines. Moreover, $\mathcal{X}$ and $\mathcal{Y}$ can be identified with their tangent spaces at arbitrary points, the exponential map equals addition, and the logarithmic map equals subtraction. In addition, all parallel transports are identity maps.

We are now showing that Algorithm 4.1 reduces to the classical Chambolle–Pock method when $n = 0 \in \mathcal{Y}$ is chosen. The same holds true for Algorithm 4.2 since Λ is already linear. Notice that the iterates $p^{(k)}$ belong to $\mathcal{X}$ while the iterates $\xi^{(k)}$ belong to $\mathcal{Y}^*$. We can drop the fixed base point $n = 0$ from their notation. Also notice that G_0^* agrees with the classical Fenchel conjugate and will be denoted by $G^*\colon \mathcal{Y} \to \overline{\mathbb{R}}$.

We only need to consider steps 3, 4 and 6 in Algorithm 4.1. The dual update step becomes

$$\xi^{(k+1)} \leftarrow \operatorname{prox}_{\tau_k G^*}\Big(\xi^{(k)} + \tau_k\big(\Lambda\bar{p}^{(k)}\big)^\flat\Big).$$

Here $\flat\colon \mathcal{Y} \to \mathcal{Y}^*$ denotes the Riesz isomorphism for the space $\mathcal{Y}$. Next we address the primal update step, which reads

$$p^{(k+1)} \leftarrow \operatorname{prox}_{\sigma_k F}\Big(p^{(k)} - \sigma_k\big(\Lambda^*\xi^{(k+1)}\big)^\sharp\Big).$$

Here $\sharp\colon \mathcal{X}^* \to \mathcal{X}$ denotes the inverse Riesz isomorphism for the space $\mathcal{X}$. Finally, the (primal) extrapolation step becomes

$$\bar{p}^{(k+1)} \leftarrow p^{(k+1)} - \theta_k\left(p^{(k)} - p^{(k+1)}\right) = p^{(k+1)} + \theta_k\left(p^{(k+1)} - p^{(k)}\right).$$

The steps above agree with Chambolle, Pock, 2011, Alg. 1 (with the roles of F and G reversed).

4.4. Convergence of the Linearized Chambolle–Pock Algorithm. In the following we adapt the proof of Chambolle, Pock, 2011 to solve the linearized saddle-point problem (4.14). We restrict the discussion to the case where $\mathcal{M}$ and $\mathcal{N}$ are Hadamard manifolds and $\mathcal{C} = \mathcal{M}$ and $\mathcal{D} = \mathcal{N}$. Moreover, we fix m and $n \coloneqq \Lambda(m)$ during the iteration and set the acceleration parameter γ to zero and $\theta_k \equiv 1$ in Algorithm 4.2.

Before presenting the main result of this section and motivated by the condition introduced after Valkonen, 2014, eq. (2.4), we introduce the following constant

$$L \coloneqq \|D\Lambda(m)\|_n, \tag{4.18}$$

i.e., the operator norm of $D\Lambda(m)\colon \mathcal{T}_m\mathcal{M} \to \mathcal{T}_n\mathcal{N}$.

THEOREM 4.3. *Let $\mathcal{M}$ and $\mathcal{N}$ be two Hadamard manifolds and $F\colon \mathcal{M} \to \overline{\mathbb{R}}$, $G\colon \mathcal{N} \to \overline{\mathbb{R}}$ be proper, convex, lsc, and $\Lambda\colon \mathcal{M} \to \mathcal{N}$. Fix $m \in \mathcal{M}$ and $n \coloneqq \Lambda(m) \in \mathcal{N}$. Suppose that the linearized saddle-point problem (4.14) has a saddle-point $(\widehat{p}, \widehat{\xi}_n)$. Choose σ, τ such that $\sigma\tau L^2 < 1$, with L defined in (4.18), and let the iterates $(\xi_n^{(k)}, p^{(k)}, \bar{\xi}_n^{(k)})$ be given by Algorithm 4.2. Suppose that there exists $K \in \mathbb{N}$ such that for all $k \geq K$, the following holds:*

$$C(k) \coloneqq \frac{1}{\sigma} d_{\mathcal{M}}^2(p^{(k)}, \widetilde{p}^{(k)}) + \langle\, \bar{\xi}_n^{(k)}, D\Lambda(m)[\zeta_k]\,\rangle \geq 0, \tag{4.19}$$

where $\widetilde{p}^{(k)}$ is defined in (4.9),

$$\zeta_k \coloneqq \mathcal{P}_{p^{(k)}\to m}\left(\log_{p^{(k)}} p^{(k+1)} - \mathcal{P}_{\widetilde{p}^{(k)}\to p^{(k)}} \log_{\widetilde{p}^{(k)}} \widehat{p}\right) - \log_m p^{(k+1)} + \log_m \widehat{p},$$

with $\bar{\xi}_n^{(k)} = 2\xi_n^{(k)} - \xi_n^{(k-1)}$. Then the following statements are true.

i) *The sequence* $(p^{(k)}, \xi_n^{(k)})$ *remains bounded, i.e.,*

$$\frac{1}{2\tau}\|\widehat{\xi}_n - \xi_n^{(k)}\|_n^2 + \frac{1}{2\sigma}d_{\mathcal{M}}^2(p^{(k)}, \widehat{p}) \le \frac{1}{2\tau}\|\widehat{\xi}_n - \xi_n^{(0)}\|_n^2 + \frac{1}{2\sigma}d_{\mathcal{M}}^2(p^{(0)}, \widehat{p}). \tag{4.20}$$

ii) *There exists a saddle-point* (p^*, ξ_n^*) *such that* $p^{(k)} \to p^*$ *and* $\xi_n^{(k)} \to \xi_n^*$.

Remark 4.4. A main difference of Theorem 4.3 to the Hilbert space case is the condition on $C(k)$. Restricting this theorem to the setting of subsection 4.3, the parallel transport and the logarithmic map simplify to the identity and subtraction, respectively. Then

$$\begin{aligned}\zeta_k &= p^{(k+1)} - p^{(k)} - \widehat{p} + \widetilde{p}^{(k)} - p^{(k+1)} + m + \widehat{p} - m = \widetilde{p}^{(k)} - p^{(k)}\\ &= -\sigma(D\Lambda(m))^*[\bar{\xi}_n^{(k)}]^\sharp\end{aligned}$$

and hence $C(k)$ simplifies to

$$C(k) = \sigma\|(D\Lambda(m))^*[\bar{\xi}_n^{(k)}]\|_{\mathcal{Y}^*}^2 - \sigma\big\langle\bar{\xi}_n^{(k)},\, D\Lambda(m)\big[((D\Lambda(m))^*[\bar{\xi}_n^{(k)}])^\sharp\big]\big\rangle = 0$$

for any $\bar{\xi}_n^{(k)}$, so condition (4.19) is satisfied for all $k \in \mathbb{N}$.

Proof of Theorem 4.3. Recall that we assume $\Lambda(m) = n$ in the following. Following along the lines of Chambolle, Pock, 2011, Thm. 1, we first write the iterations from Algorithm 4.2 in a general form

$$\begin{aligned}p^{(k+1)} &= \operatorname{prox}_{\sigma F}(\,\widetilde{p}^{(k)}\,), \quad \widetilde{p}^{(k)} := \exp_{p^{(k)}}\Big(\mathcal{P}_{m\to p^{(k)}}\left(-\sigma(D\Lambda(m))^*[\bar{\xi}_n]\right)^\sharp\Big)\\ \xi_n^{(k+1)} &= \operatorname{prox}_{\tau G_n^*}(\,\widetilde{\xi}_n^{(k)}\,), \quad \widetilde{\xi}_n^{(k)} := \xi_n^{(k)} + \tau(D\Lambda(m)[\log_m \bar{p}])^\flat,\end{aligned} \tag{4.21}$$

for some $\bar{\xi}_n$ and $\bar{p}$ to be specified later on. Applying Lemma 2.19, we get

$$\begin{aligned}\frac{1}{\sigma}\big(\log_{p^{(k+1)}}\widetilde{p}^{(k)}\big)^\flat &\in \partial_{\mathcal{M}}F(p^{(k+1)}),\\ \left(\frac{\xi_n^{(k)} - \xi_n^{(k+1)}}{\tau}\right)^\sharp + D\Lambda(m)[\log_m \bar{p}] &\in \partial G_n^*(\xi_n^{(k+1)}).\end{aligned} \tag{4.22}$$

Due to Definition 2.5 and Definition 2.12, we obtain that for every $\xi_n \in \mathcal{T}_n^*\mathcal{N}$ and $p \in \mathcal{M}$ it holds

$$\begin{aligned}F(p) &\ge F(p^{(k+1)}) + \frac{1}{\sigma}(\,\log_{p^{(k+1)}}\widetilde{p}^{(k)}, \log_{p^{(k+1)}} p\,)_{p^{(k+1)}},\\ G_n^*(\xi_n) &\ge G_n^*(\xi_n^{(k+1)}) + \Big(\,\frac{\xi_n^{(k)} - \xi_n^{(k+1)}}{\tau}, \xi_n - \xi_n^{(k+1)}\,\Big)_n\\ &\quad + \langle\,\xi_n - \xi_n^{(k+1)}, D\Lambda(m)[\log_m \bar{p}]\,\rangle.\end{aligned} \tag{4.23}$$

Now we consider the geodesic triangle $\Delta = (\widetilde{p}^{(k)}, p^{(k+1)}, p)$. Applying the law of cosines in Hadamard manifolds (Ferreira, Oliveira, 2002, Thm. 2.2), we obtain

$$\begin{aligned}\frac{1}{\sigma}(\,\log_{p^{(k+1)}}\widetilde{p}^{(k)}, \log_{p^{(k+1)}} p\,)_{p^{(k+1)}} \ge \frac{1}{2\sigma}d_{\mathcal{M}}^2(\widetilde{p}^{(k)}, p^{(k+1)}) + \frac{1}{2\sigma}d_{\mathcal{M}}^2(p, p^{(k+1)})\\ - \frac{1}{2\sigma}d_{\mathcal{M}}^2(\widetilde{p}^{(k)}, p).\end{aligned}$$

Rearranging the law of cosines for the triangle $\Delta = (p^{(k)}, \widetilde{p}^{(k)}, p)$ yields

$$-\frac{1}{2\sigma} d_{\mathcal{M}}^2(\widetilde{p}^{(k)}, p) \geq \frac{1}{2\sigma} d_{\mathcal{M}}^2(\widetilde{p}^{(k)}, p^{(k)}) - \frac{1}{2\sigma} d_{\mathcal{M}}^2(p^{(k)}, p) - \frac{1}{\sigma} (\log_{\widetilde{p}^{(k)}} p^{(k)}, \log_{\widetilde{p}^{(k)}} p)_{\widetilde{p}^{(k)}}.$$

We rephrase the last term as

$$\begin{aligned} -\frac{1}{\sigma} (\log_{\widetilde{p}^{(k)}} p^{(k)}, \log_{\widetilde{p}^{(k)}} p)_{\widetilde{p}^{(k)}} &= -\frac{1}{\sigma} (\mathcal{P}_{\widetilde{p}^{(k)} \to p^{(k)}} \log_{\widetilde{p}^{(k)}} p^{(k)}, \mathcal{P}_{\widetilde{p}^{(k)} \to p^{(k)}} \log_{\widetilde{p}^{(k)}} p)_{p^{(k)}} \\ &= -\frac{1}{\sigma} (-\log_{p^{(k)}} \widetilde{p}^{(k)}, \mathcal{P}_{\widetilde{p}^{(k)} \to p^{(k)}} \log_{\widetilde{p}^{(k)}} p)_{p^{(k)}} \\ &= -((D\Lambda(m))^*[\bar{\xi}_n]^\sharp, \mathcal{P}_{p^{(k)} \to m} \mathcal{P}_{\widetilde{p}^{(k)} \to p^{(k)}} \log_{\widetilde{p}^{(k)}} p)_m \\ &= -\langle \xi_n, D\Lambda(m)[\mathcal{P}_{p^{(k)} \to m} \mathcal{P}_{\widetilde{p}^{(k)} \to p^{(k)}} \log_{\widetilde{p}^{(k)}} p] \rangle. \end{aligned}$$

We insert the estimates above into the first inequality in (4.23) to obtain

$$\begin{aligned} F(p) \geq F(p^{(k+1)}) &+ \frac{1}{2\sigma} d_{\mathcal{M}}^2(\widetilde{p}^{(k)}, p^{(k+1)}) + \frac{1}{2\sigma} d_{\mathcal{M}}^2(p^{(k+1)}, p) + \frac{1}{2\sigma} d_{\mathcal{M}}^2(\widetilde{p}^{(k)}, p^{(k)}) \\ &- \frac{1}{2\sigma} d_{\mathcal{M}}^2(p^{(k)}, p) - \langle \bar{\xi}_n, D\Lambda(m)[\mathcal{P}_{p^{(k)} \to m} \mathcal{P}_{\widetilde{p}^{(k)} \to p^{(k)}} \log_{\widetilde{p}^{(k)}} p] \rangle. \end{aligned}$$

Considering now the geodesic triangle $\Delta = (\widetilde{p}^{(k)}, p^{(k)}, p^{(k+1)})$ we get

$$\begin{aligned} \frac{1}{2\sigma} d_{\mathcal{M}}^2(p^{(k+1)}, \widetilde{p}^{(k)}) \geq \frac{1}{2\sigma} d_{\mathcal{M}}^2(p^{(k)}, p^{(k+1)}) &+ \frac{1}{2\sigma} d_{\mathcal{M}}^2(p^{(k)}, \widetilde{p}^{(k)}) \\ &- \frac{1}{\sigma} (\log_{p^{(k)}} \widetilde{p}^{(k)}, \log_{p^{(k)}} p^{(k+1)})_{p^{(k)}}, \end{aligned}$$

and, noticing that

$$-\frac{1}{\sigma} (\log_{p^{(k)}} \widetilde{p}^{(k)}, \log_{p^{(k)}} p^{(k+1)})_{p^{(k)}} = \langle \bar{\xi}_n, D\Lambda(m)[\mathcal{P}_{p^{(k)} \to m} \log_{p^{(k)}} p^{(k+1)}] \rangle,$$

we write

$$\begin{aligned} F(p) \geq F(p^{(k+1)}) &+ \frac{1}{2\sigma} d_{\mathcal{M}}^2(p^{(k+1)}, p) - \frac{1}{2\sigma} d_{\mathcal{M}}^2(p^{(k)}, p) + \frac{1}{2\sigma} d_{\mathcal{M}}^2(p^{(k)}, p^{(k+1)}) \\ &+ \frac{1}{\sigma} d_{\mathcal{M}}^2(p^{(k)}, \widetilde{p}^{(k)}) \\ &+ \langle \bar{\xi}_n, D\Lambda(m)[\mathcal{P}_{p^{(k)} \to m} \log_{p^{(k)}} p^{(k+1)} - \mathcal{P}_{p^{(k)} \to m} \mathcal{P}_{\widetilde{p}^{(k)} \to p^{(k)}} \log_{\widetilde{p}^{(k)}} p] \rangle. \end{aligned}$$

Adding this inequality with the second inequality from (4.23), we get

$$
\begin{aligned}
&\frac{1}{2\tau}\|\xi_n - \xi_n^{(k)}\|_n^2 + \frac{1}{2\sigma} d_{\mathcal{M}}^2(p^{(k)}, p)\\
&\qquad \geq \langle\, D\Lambda(m)[\log_m p^{(k+1)}], \xi_n \,\rangle + F(p^{(k+1)}) - G_n^*(\xi_n)\\
&\qquad\quad - [\langle\, D\Lambda(m)[\log_m p], \xi^{(k+1)} \,\rangle + F(p) - G_n^*(\xi_n^{(k+1)})]\\
&\qquad\quad + \frac{1}{2\tau}\|\xi_n - \xi_n^{(k+1)}\|_n^2 + \frac{1}{2\tau}\|\xi_n^{(k)} - \xi_n^{(k+1)}\|_n^2\\
&\qquad\quad + \frac{1}{2\sigma} d_{\mathcal{M}}^2(p^{(k+1)}, p) + \frac{1}{2\sigma} d_{\mathcal{M}}^2(p^{(k)}, p^{(k+1)})\\
&\qquad\quad + \frac{1}{\sigma} d_{\mathcal{M}}^2(p^{(k)}, \widetilde{p}^{(k)}) \qquad \text{(4.24a)}\\
&\qquad\quad + \langle\, \bar{\xi}_n, D\Lambda(m)[\mathcal{P}_{p^{(k)}\to m} \log_{p^{(k)}} p^{(k+1)} - \mathcal{P}_{p^{(k)}\to m}\, \mathcal{P}_{\widetilde{p}^{(k)}\to p^{(k)}} \log_{\widetilde{p}^{(k)}} p] \,\rangle \qquad \text{(4.24b)}\\
&\qquad\quad + \langle\, \xi_n^{(k+1)} - \xi_n, D\Lambda(m)[\log_m p^{(k+1)} - \log_m \bar{p}] \,\rangle \qquad \text{(4.24c)}\\
&\qquad\quad - \langle\, \xi_n^{(k+1)} - \bar{\xi}_n, D\Lambda(m)[\log_m p^{(k+1)} - \log_m p] \,\rangle \qquad \text{(4.24d)}\\
&\qquad\quad - \langle\, \bar{\xi}_n, D\Lambda(m)[\log_m p^{(k+1)} - \log_m p] \,\rangle. \qquad \text{(4.24e)}
\end{aligned}
$$

Choosing now $\bar{p} = p^{(k+1)}$ and $\bar{\xi}_n = 2\xi_n^{(k)} - \xi_n^{(k-1)}$ (4.24c) vanishes. We continue with (4.24d) and estimate it according to

$$
\begin{aligned}
&-\langle\, \xi_n^{(k+1)} - \bar{\xi}_n, D\Lambda(m)[\log_m p^{(k+1)} - \log_m p] \,\rangle\\
&\qquad = -\langle\, \xi_n^{(k+1)} - \xi_n^{(k)} - (\xi_n^{(k)} - \xi_n^{(k-1)}), D\Lambda(m)[\log_m p^{(k+1)} - \log_m p] \,\rangle\\
&\qquad = -\langle\, \xi_n^{(k+1)} - \xi_n^{(k)}, D\Lambda(m)[\log_m p^{(k+1)} - \log_m p] \,\rangle\\
&\qquad\quad + \langle\, \xi_n^{(k)} - \xi_n^{(k-1)}, D\Lambda(m)[\log_m p^{(k)} - \log_m p] \,\rangle\\
&\qquad\quad - \langle\, \xi_n^{(k-1)} - \xi_n^{(k)}, D\Lambda(m)[\log_m p^{(k+1)} - \log_m p^{(k)}] \,\rangle\\
&\qquad \geq -\langle\, \xi_n^{(k+1)} - \xi_n^{(k)}, D\Lambda(m)[\log_m p^{(k+1)} - \log_m p] \,\rangle\\
&\qquad\quad + \langle\, \xi_n^{(k)} - \xi_n^{(k-1)}, D\Lambda(m)[\log_m p^{(k)} - \log_m p] \,\rangle\\
&\qquad\quad - L\|\xi_n^{(k)} - \xi^{(k-1)}\|_n \|\log_m p^{(k+1)} - \log_m p^{(k)}\|_m.
\end{aligned}
$$

Using that $2ab \leq \alpha a^2 + b^2/\alpha$ holds for every $a, b \geq 0$ and $\alpha > 0$, and choosing in particular $\alpha = \frac{\sqrt{\tau}}{\sqrt{\sigma}}$, we get

$$
\begin{aligned}
&-\langle\, \xi_n^{(k+1)} - \bar{\xi}_n, D\Lambda(m)[\log_m p^{(k+1)} - \log_m p] \,\rangle\\
&\qquad \geq -\langle\, \xi_n^{(k+1)} - \xi_n^{(k)}, D\Lambda(m)[\log_m p^{(k+1)} - \log_m p] \,\rangle\\
&\qquad\quad + \langle\, \xi_n^{(k)} - \xi_n^{(k-1)}, D\Lambda(m)[\log_m p^{(k)} - \log_m p] \,\rangle\\
&\qquad\quad - \frac{L\sqrt{\tau}}{2\sqrt{\sigma}} d_{\mathcal{M}}^2(p^{(k+1)}, p^{(k)}) - \frac{L\sqrt{\sigma}}{2\sqrt{\tau}}\|\xi_n^{(k-1)} - \xi_n^{(k)}\|_n^2, \qquad \text{(4.25)}
\end{aligned}
$$

where L is the constant defined in (4.18). We notice now that (4.24a), (4.24b) and (4.24e) correspond to $C(k+1)$, hence we simplify the inequality of interest by writing $C(k+1)$ instead and also noticing that the first two lines on the right hand side of (4.25) are the primal dual gap, denoted in the following by $\mathrm{PDG}(k+1)$.

With these modifications to (4.24a)–(4.24e), we get

$$
\begin{aligned}
&\frac{1}{2\tau}\|\xi_n - \xi_n^{(k)}\|_n^2 + \frac{1}{2\sigma}d_{\mathcal{M}}^2(p^{(k)}, p) \\
&\qquad \geq \mathrm{PDG}(k+1) + C(k+1) \\
&\qquad\quad + \Big(\frac{1}{2\sigma} - \frac{L\sqrt{\tau}}{2\sqrt{\sigma}}\Big)d_{\mathcal{M}}^2(p^{(k)}, p^{(k+1)}) + \frac{1}{2\sigma}d_{\mathcal{M}}^2(p^{(k+1)}, p) \\
&\qquad\quad + \frac{1}{2\tau}\|\xi_n - \xi_n^{(k+1)}\|_n^2 + \frac{1}{2\tau}\|\xi_n^{(k)} - \xi_n^{(k+1)}\|_n^2 - \frac{L\sqrt{\sigma}}{2\sqrt{\tau}}\|\xi_n^{(k-1)} - \xi_n^{(k)}\|_n^2 \\
&\qquad\quad - \langle\, \xi_n^{(k+1)} - \xi_n^{(k)}, D\Lambda(m)[\log_m p^{(k+1)} - \log_m p]\,\rangle \\
&\qquad\quad + \langle\, \xi_n^{(k)} - \xi_n^{(k-1)}, D\Lambda(m)[\log_m p^{(k)} - \log_m p]\,\rangle.
\end{aligned}
\tag{4.26}
$$

We continue to sum (4.26) from 0 to $N-1$, where we set $\xi_n^{(-1)} \coloneqq \xi_n^{(0)}$ in coherence with the initial choice $\bar{\xi}_n^{(0)} = \xi_n^{(0)}$. For every p and ξ_n, we get

$$
\begin{aligned}
&\frac{1}{2\tau}\|\xi_n - \xi_n^{(0)}\|_n^2 + \frac{1}{2\sigma}d_{\mathcal{M}}^2(p^{(0)}, p) \\
&\quad \geq \sum_{k=0}^{N-1}\mathrm{PDG}(k+1) + \sum_{k=0}^{N-1}C(k+1) + \frac{1}{2\tau}\|\xi_n - \xi_n^{(N)}\|_n^2 + \frac{1}{2\sigma}d_{\mathcal{M}}^2(p^{(N)}, p) \\
&\quad + \Big(\frac{1}{2\sigma} - \frac{L\sqrt{\tau}}{2\sqrt{\sigma}}\Big)\sum_{k=1}^{N} d_{\mathcal{M}}^2(p^{(k)}, p^{(k-1)}) + \Big(\frac{1}{2\tau} - \frac{L\sqrt{\sigma}}{2\sqrt{\tau}}\Big)\sum_{k=1}^{N-1}\|\xi_n^{(k)} - \xi_n^{(k-1)}\|_n^2 \\
&\quad + \frac{1}{2\tau}\|\xi_n^{(N-1)} - \xi_n^{(N)}\|_n^2 - \langle\, \xi_n^{(N)} - \xi_n^{(N-1)}, D\Lambda(m)[\log_m p^{(N)} - \log_m p]\,\rangle.
\end{aligned}
\tag{4.27}
$$

We further develop the last term in (4.27) and get

$$
\begin{aligned}
-\langle\, \xi_n^{(N)} - \xi_n^{(N-1)}, &D\Lambda(m)[\log_m p^{(N)} - \log_m p]\,\rangle \\
&\geq -L\|\xi_n^{(N)} - \xi_n^{(N-1)}\|_n d_{\mathcal{M}}(p^{(N)}, p) \\
&\geq -\frac{L\alpha}{2}\|\xi_n^{(N)} - \xi_n^{(N-1)}\|_n^2 - \frac{L}{2\alpha}d_{\mathcal{M}}^2(p^{(N)}, p).
\end{aligned}
$$

Choosing $\alpha = 1/(\tau L)$, it follows

$$
\begin{aligned}
-\langle\, \xi_n^{(N)} - \xi_n^{(N-1)}, &D\Lambda(m)[\log_m p^{(N)} - \log_m p]\,\rangle \\
&\geq -\frac{1}{2\tau}\|\xi_n^{(N)} - \xi_n^{(N-1)}\|_n^2 - \frac{\tau L^2}{2}d_{\mathcal{M}}^2(p^{(N)}, p).
\end{aligned}
$$

Hence (4.27) becomes

$$
\begin{aligned}
&\frac{1}{2\tau}\|\xi_n - \xi_n^{(0)}\|_n^2 + \frac{1}{2\sigma}d_{\mathcal{M}}^2(p^{(0)}, p) \\
&\qquad \geq \sum_{k=0}^{N-1}\mathrm{PDG}(k+1) + \sum_{k=0}^{N-1}C(k+1) \\
&\qquad\quad + \frac{1}{2\tau}\|\xi_n - \xi_n^{(N)}\|_n^2 + \Big(\frac{1}{2\tau} - \frac{L\sqrt{\sigma}}{2\sqrt{\tau}}\Big)\sum_{k=1}^{N-1}\|\xi_n^{(k)} - \xi_n^{(k-1)}\|_n^2 \\
&\qquad\quad + \Big(\frac{1}{2\sigma} - \frac{\tau L^2}{2}\Big)d_{\mathcal{M}}^2(p^{(N)}, p) + \Big(\frac{1}{2\sigma} - \frac{L\sqrt{\tau}}{2\sqrt{\sigma}}\Big)\sum_{k=1}^{N} d_{\mathcal{M}}^2(p^{(k)}, p^{(k-1)}).
\end{aligned}
\tag{4.28}
$$

Taking in particular $(p, \xi_n) = (\widehat{p}, \widehat{\xi_n})$, the combination of the feasibility of the saddle-point $(\widehat{p}, \widehat{\xi_n})$ together with (4.19) and the inequality $1 > \sigma\tau L^2$ implies that the sequence $\{p^{(k)}, \xi_n^{(k)}\}$ is bounded, which is the statement i).

Part ii) follows completely analogously to the steps of Chambolle, Pock, 2011, Thm. 1(c) adapted to (4.26). ∎

5. ROF Models on Manifolds. A starting point of the work of Chambolle, Pock, 2011 is the ROF ℓ^2-TV denoising model Rudin, Osher, Fatemi, 1992, which was generalized to manifolds in Lellmann et al., 2013 for the so-called isotropic and anisotropic cases. This class of ℓ^2-TV models can be formulated in the discrete setting as follows: let $F = (f_{i,j})_{i,j} \in \mathcal{M}^{d_1 \times d_2}$, $d_1, d_2 \in \mathbb{N}$ be a manifold-valued image, i.e., each pixel $f_{i,j}$ takes values in a manifold $\mathcal{M}$. Then the manifold-valued ℓ^2-TV energy functional reads

$$\mathcal{E}_q(P) \coloneqq \frac{1}{2\alpha} \sum_{i,j=1}^{d_1,d_2} d^2_{\mathcal{M}}(f_{i,j}, p_{i,j}) + \|\nabla P\|_{g,q,1}, \quad P = (p_{i,j})_{i,j} \in \mathcal{M}^{d_1 \times d_2}, \tag{5.1}$$

where $q \in \{1, 2\}$. Moreover, $\nabla \colon \mathcal{M}^{d_1 \times d_2} \to \mathcal{T}\mathcal{M}^{d_1 \times d_2 \times 2}$ denotes the generalization of the finite difference operator, which is defined as

$$(\nabla P)_{i,j,k} = \begin{cases} 0 \in \mathcal{T}_{p_{i,j}}\mathcal{M} & \text{if } i = d_1 \text{ and } k = 1, \\ 0 \in \mathcal{T}_{p_{i,j}}\mathcal{M} & \text{if } j = d_2 \text{ and } k = 2, \\ \log_{p_{i,j}} p_{i+1,j} & \text{if } i < d_1 \text{ and } k = 1, \\ \log_{p_{i,j}} p_{i,j+1} & \text{if } j < d_2 \text{ and } k = 2. \end{cases} \tag{5.2}$$

The corresponding norm is then given by

$$\|\nabla P\|_{g,q,1} = \sum_{i,j=1}^{d_1,d_2} \bigl(\|(\nabla P)_{i,j,1}\|_g^q + \|(\nabla P)_{i,j,2}\|_g^q\bigr)^{\frac{1}{q}}. \tag{5.3}$$

For simplicity of notation we do not explicitly state the base point in the Riemannian metric but denote the norm on $\mathcal{T}\mathcal{M}$ by $\|\cdot\|_g$. Depending on the value of $q \in \{1, 2\}$, we call the energy functional (5.1) *isotropic* when $q = 2$ and *anisotropic* for $q = 1$. Note that previous algorithms like CPPA Weinmann, Demaret, Storath, 2014 or Douglas–Rachford (DR) Bergmann, Persch, Steidl, 2016 are only able to tackle the anisotropic case $q = 1$ due to a missing closed form of the prox for the isotropic TV summands. A relaxed version of the isotropic case can be computed using the half-quadratic minimization Bergmann, Chan, et al., 2016. Looking at the optimality conditions of the isotropic or anisotropic energy functional, the authors in Bergmann, Tenbrinck, 2018 derived and solved the corresponding q-Laplace equation. This can be generalized even to the cases $q > 0$, $q \notin \{1, 2\}$.

The minimization of (5.1) fits into the setting of the model problem (4.1). Indeed, $\mathcal{M}$ is replaced by $\mathcal{M}^{d_1 \times d_2}$, $\mathcal{N} = \mathcal{T}\mathcal{M}^{d_1 \times d_2 \times 2}$, F is given by the first sum in (5.1), $\Lambda = \nabla$ and $G_q = \|\cdot\|_{g,q,1}$. We apply Algorithm 4.2 to solve the linearized saddle-point problem (4.14). This procedure will yield an approximate minimizer of (5.1). To this end we require both the Fenchel conjugate and the proximal map of G. Its Fenchel dual can be stated using the dual norms, i.e., $\|\cdot\|_{g,q^*,\infty}$ similar to Thm. 2 of Duran et al., 2016, where $q^* \in \mathbb{R}$ is the dual exponent of q. Let

$$B_{q^*} \coloneqq \{ X \mid \|X\|_{g,q^*,\infty} \leq 1 \}$$

denote the 1-norm ball of the dual norm and

$$\iota_B(x) \coloneqq \begin{cases} 0 & \text{if } x \in B, \\ \infty & \text{else,} \end{cases}$$

the indicator function of the set B. Then the two cases of main interest, namely the Fenchel dual functions for the cases $q = 1$ and $q = 2$, read

$$G_2^*(\Xi) = \iota_{B_2}(\Xi), \quad \text{and} \quad G_\infty^*(\Xi) = \iota_{B_\infty}(\Xi).$$

The corresponding proximal maps read

$$\operatorname{prox}_{\tau G_2^*}(\,\Xi\,) = \Big(\big(\max\{1, \big\|\|\Xi_{i,j,:}\|_g\big\|_2\}\big)^{-1} \Xi_{i,j,k} \Big)_{i,j,k}$$

and

$$\operatorname{prox}_{\tau G_\infty^*}(\,\Xi\,) = \Big(\big(\max\{1, \|\Xi_{i,j,k}\|_g\}\big)^{-1} \Xi_{i,j,k} \Big)_{i,j,k}.$$

Finally, to derive the adjoint of $D\Lambda(m)$, let $P \in \mathcal{M}^{d_1 \times d_2}$ and $X \in \mathcal{T}_P\mathcal{M}^{d_1 \times d_2}$. Applying the chain rule it is not difficult to prove that

$$(D\nabla(P)[X])_{i,j,k} = D_1 \log_{p_{i,j}}(p_{i,j+e_k})[X_{i,j}] + D_2 \log_{p_{i,j}}(p_{i,j+e_k})[X_{i,j+e_k}] \tag{5.4}$$

with the obvious modifications at the boundary. In the above formula e_k represents either the vector $(0,1)$ or $(1,0)$ used to reach either the neighbor to the right ($k = 1$) or below ($k = 2$). The symbols D_1 and D_2 represent the differentiation of the logarithmic map w.r.t. the base point and its argument, respectively. We notice that $D_1 \log_{\cdot}(p_{i,j+e_k})$ and $D_2 \log_{p_{i,j}}(\,\cdot\,)$ can be computed by a straightforward application of Jacobi fields; see for example Bergmann, Fitschen, et al., 2018, Lem. 4.1 ii) and iii).

With $(D\nabla)(\,\cdot\,)[\,\cdot\,]\colon \mathcal{T}\mathcal{M}^{d_1 \times d_2} \to \mathcal{T}\mathcal{N}$ given by Jacobi fields its adjoint can be computed using the so-called adjoint Jacobi fields, see e.g. Bergmann, Gousenbourger, 2018, Sect. 4.2. Defining $N_{i,j}$ to be the set of neighbors of the pixel $p_{i,j}$, for every $X \in \mathcal{T}_P\mathcal{M}^{d_1 \times d_2}$ and $\eta \in \mathcal{T}^*_{\nabla P}\mathcal{N}$ we have

$$\begin{aligned}
&\langle\, D\nabla(P)[X], \eta \,\rangle \\
&\quad= \sum_{i,j,k} \langle\, (D\nabla(P)[X])_{i,j,k}, \eta_{i,j,k} \,\rangle \\
&\quad= \sum_{i,j} \sum_{k} \langle\, D_1 \log_{p_{i,j}} p_{i,j+e_k}[X_{i,j}], \eta_{i,j,k} \,\rangle + \sum_{k} \langle\, D_2 \log_{p_{i,j}} p_{i,j+e_k}[X_{i,j+e_k}], \eta_{i,j,k} \,\rangle \\
&\quad= \sum_{i,j} \sum_{k} \langle\, X_{i,j}, D_1^* \log_{p_{i,j}} p_{i,j+e_k}[\eta_{i,j,k}] \,\rangle + \sum_{k} \langle\, X_{i,j+e_k}, D_2^* \log_{p_{i,j}} p_{i,j+e_k}[\eta_{i,j,k}] \,\rangle \\
&\quad= \sum_{i,j} \Big\langle X_{i,j}, \sum_{k} D_1^* \log_{p_{i,j}} p_{i,j+e_k}[\eta_{i,j,k}] + \sum_{(i',j') \in N_{i,j}} D_2^* \log_{p_{i'j'}} p_{i,j}[\eta_{i'j'k}] \Big\rangle \\
&\quad= \sum_{i,j} \langle\, X_{i,j}, (D^*\nabla(P)[\eta])_{i,j} \,\rangle,
\end{aligned}$$

which leads to the component-wise entries in the linearized adjoint

$$(D^*\nabla(P)[\eta])_{i,j} = \sum_k D_1^* \log_{p_{i,j}} p_{i,j+e_k}[\eta_{i,j,k}] + \sum_{(i',j')\in N_{i,j}} D_2^* \log_{p_{i'j'}} p_{i,j}[\eta_{i'j'k}]. \tag{5.5}$$

As before, we recall that $D_1^* \log_{\cdot}(p_{i,j+e_k})$ and $D_2^* \log_{p_{i,j}}(\,\cdot\,)$ also follow from Bergmann, Fitschen, et al., 2018, Sect. 4.

6. Numerical Experiments. The numerical experiments are implemented in the toolbox MANOPT.JL[1] (Bergmann, 2019) in Julia[2]. They were run on a MacBook Pro, 2.5 Ghz Intel Core i7, 16 GB RAM, with Julia 1.1. All our examples are based on the linearized saddle-point formulation (4.14) for ℓ^2-TV.

6.1. A Signal with Known Minimizer. The first example uses signal data $\mathcal{M}^{d_1}$ instead of an image, where the data space is $\mathcal{M} = \mathbb{S}^2$. This gives us the opportunity to consider the same problem also on the embedding manifold $(\mathbb{R}^3)^{d_1}$ in order to illustrate the difference between the manifold-valued and Euclidean settings. We construct the data $(f_i)_i$ such that the unique minimizer of (5.1) is known in closed form. Therefore a second purpose of this problem is to compare the numerical solution obtained by Algorithm 4.2, i.e., an approximate saddle-point of the *linearized* problem (4.14), to the solution of the original saddle-point problem (4.2). Third, we wish to explore how the value $C(k)$ from (4.19) behaves numerically.

The piecewise constant signal is given by

$$f \in \mathcal{M}^{30} \quad f_i = \begin{cases} p_1 & \text{if } i \le 15, \\ p_2 & \text{if } i > 15, \end{cases}$$

for two values $p_1, p_2 \in \mathcal{M}$ specified below.

Notice that since $d_2 = 1$, the isotropic and anisotropic models (5.1) coincide. The exact minimizer $\widehat{p}$ of (5.1) is piecewise constant with the same structure as the data f. Its values are $\widehat{p}_1 = \gamma_{\overset{\frown}{p_1,p_2}}(\delta)$ and $\widehat{p}_2 = \gamma_{\overset{\frown}{p_2,p_1}}(\delta)$ where $\delta = \min\{d_{\mathcal{M}}(p_1,p_2)\frac{\alpha}{15}, \frac{1}{2}\}$. Notice that the notion of geodesics are different for both manifolds under consideration, and thus $\widehat{p}_{\mathbb{R}^3}$ is different from $\widehat{p}_{\mathbb{S}^2}$.

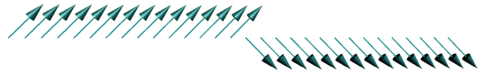
In the following we use $\alpha = 5$ and $p_1 = \frac{1}{\sqrt{2}}(1,1,0)^{\mathrm{T}}, p_2 = \frac{1}{\sqrt{2}}(1,-1,0)^{\mathrm{T}}$. The data f is shown in Figure 2a.

We ran the linearized Riemannian Chambolle–Pock Algorithm 4.2 with relaxation parameter $\theta = 1$ on the dual variable as well as $\sigma = \tau = \frac{1}{2}$, and $\gamma = 0$, i.e., without acceleration, as well as $p^{(0)} = f$ and $\xi_n^{(0)}$ as the zero vector. The stopping criterion is set to 500 iterations and as m we use the mean of the data, which is just $m = \gamma_{\overset{\frown}{p_1,p_2}}(\frac{1}{2})$. For the Euclidean case $\mathcal{M} = \mathbb{R}^3$, we obtain a shifted version of the original Chambolle–Pock algorithm, since $m \neq 0$.

While the algorithm on $\mathcal{M} = \mathbb{S}^2$ takes about 0.85 seconds, the Euclidean algorithm takes about 0.44 seconds, which is most likely due to the exponential, logarithmic map as well as the parallel transport on $\mathbb{S}^2$, which involve sines and cosines. The

[1] Available at http://www.manoptjl.org, following the same philosophy as the Matlab version available at https://manopt.org, see also Boumal et al., 2014.

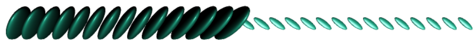
[2] https://julialang.org



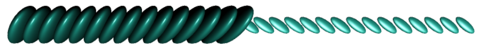
(a) Signal f of unit vectors.

(b) Minimizer with values in $\mathcal{M} = \mathbb{S}^2$.

(c) Minimizer with values in $\mathcal{M} = \mathbb{R}^3$.



(d) Signal of $\mathcal{P}(3)$ matrices.



(e) Minimizer on $\mathcal{M} = \mathcal{P}(3)$.

Fig. 2: Computing the minimizer of the manifold-valued ℓ^2-TV model for a signal of unit vectors shown in (a) with respect to both manifolds $\mathbb{R}^3$ and $\mathbb{S}^2$ with $\alpha = 5$: (b) on $(\mathbb{S}^2)^{30}$ and (c) on $(\mathbb{R}^3)^{30}$. The known effect, loss of contrast is different for both cases, since on $\mathbb{S}^2$ the vector remain of unit length. The same effect can be seen for a signal of spd matrices, i.e., $\mathcal{P}(3)$; see (d) and (e).

results obtained by the Euclidean algorithm is $2.18 \cdot 10^{-12}$ away in Euclidean norm from the analytical minimizer $\widehat{p}_{\mathbb{R}^3}$. Notice that the convergence of the Euclidean algorithm is covered by the theory in Chambolle, Pock, 2011. Moreover, notice that in this setting, Λ is a linear map between vector spaces. During the iterations, we confirmed that the value of $C(k)$ is numerically zero ($\pm 5.5511 \cdot 10^{-17}$), as expected from Remark 4.4.

Although Algorithm 4.2 on $\mathcal{M} = \mathbb{S}^2$ is based on the *linearized* saddle-point problem (4.14) instead of (4.2), we observed that it converges to the exact minimizer $\widehat{p}_{\mathbb{S}^2}$ of (5.1). Therefore it is meaningful to plug in $\widehat{p}_{\mathbb{S}^2}$ into the formula (4.19) to evaluate $C(k)$ numerically. The numerical values observed throughout the 500 iterations are in the interval $[-4.0 \cdot 10^{-13}, 4.0 \cdot 10^{-9}]$. We interpret this as confirmation that $C(k)$ is non-negative in this case. However, even with this observation the convergence of Algorithm 4.2 is not covered by Theorem 4.3 since $\mathbb{S}^2$ is not a Hadamard manifold. Quite to the contrary, it has constant positive sectional curvature.

The results are shown in Figure 2c and Figure 2b, respectively. They illustrate the well known loss of contrast and reduction of jump heights. This leads to shorter vectors in $\widehat{p}_{\mathbb{R}^3}$, while, of course, their unit length is preserved in $\widehat{p}_{\mathbb{S}^2}$.

We also constructed a similar signal on $\mathcal{M} = \mathcal{P}(3)$, the manifold of symmetric positive definite (SPD) matrices with affine metric; see Pennec, Fillard, Ayache, 2006. This is a Hadamard manifold with non-constant curvature. Let $I \in \mathbb{R}^{3\times 3}$ denote the unit matrix and

$$p_1 = \exp_I\Big(\frac{2}{\lVert X\rVert_I} X\Big),\ p_2 = \exp_I\Big(-\frac{2}{\lVert X\rVert_I} X\Big),\ \text{with } X = \frac{1}{2}\begin{pmatrix} 1 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 6 \end{pmatrix} \in \mathcal{T}_I\mathcal{P}(3).$$

In this case, the run time is 5.94 seconds, which is due to matrix exponentials and logarithms as well as singular value decompositions that need to be computed. Here,

$C(k)$ turns out to be numerically zero ($\pm 8 \cdot 10^{-15}$) and the distance to the analytical minimizer $\widehat{p}_{\mathcal{P}(3)}$ is $1.08 \cdot 10^{-12}$. The original data f and the result $\widehat{p}_{\mathcal{P}(3)}$ (again with a loss of contrast as expected) are shown in Figure 2d and Figure 2e, respectively.

6.2. A Comparison of Algorithms. As a second example we compare our Algorithm 4.2 to the cyclic proximal point algorithm (CPPA) Bačák, 2014a, which was first applied to ℓ^2-TV problems in Weinmann, Demaret, Storath, 2014. It is known to be a robust but generally slow method. We also compare the proposed method with the parallel Douglas–Rachford algorithm (PDRA), which was introduced in Bergmann, Persch, Steidl, 2016.

As an example, we use the anisotropic ℓ^2-TV model (5.1) on images of size 32×32 with values in the manifold of 3×3 SPD matrices $\mathcal{P}(3)$ as in the previous subsection. The original data is shown in Figure 3a. No exact solution is known for this example. We use a value of $\alpha = 6$. To generate a reference solution we allowed the CPPA with step size $\lambda_k = \frac{4}{k}$ to run for 4000 iterations. This required 1235.18 seconds and it yields a value of the objective function (5.1) of approximately 38.7370, see the bottom gray line in Figure 3c. The result is shown in Figure 3b.

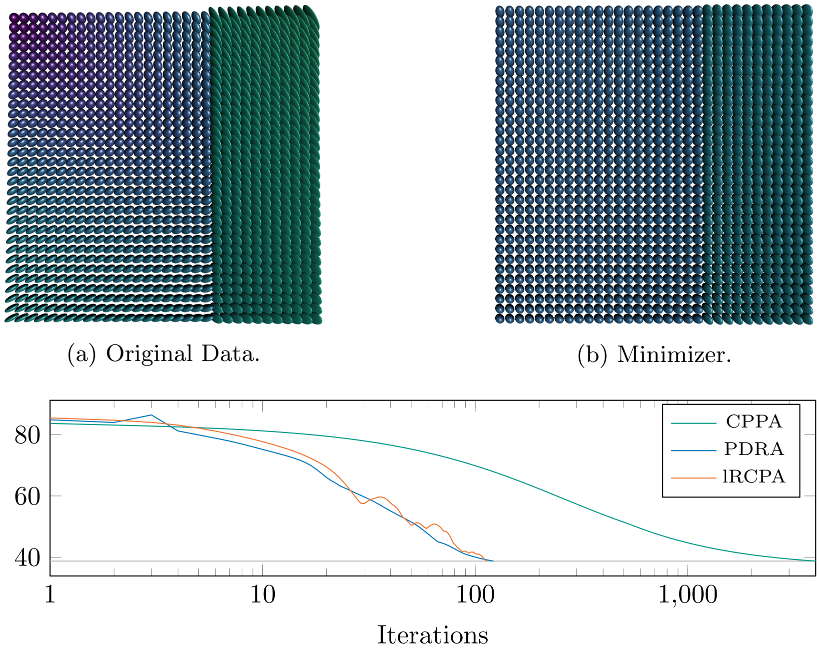
We compare CPPA to PDRA as well as to our Algorithm 4.2 using the value of the cost function and the run time as criteria. The PDRA was run with parameters $\eta = 0.58$, $\lambda = 0.93$, which where used by Bergmann, Persch, Steidl, 2016 for a similar example. It took 379.7 seconds to perform 122 iterations in order to reach the same value of the cost function as obtained by CPPA. The main bottleneck is the approximate evaluation of the involved mean, which has to be computed in every iteration. Here we performed 20 gradient descent steps for this purpose.

For Algorithm 4.2 we set $\sigma = \tau = 0.4$ and $\gamma = 0.2$. We choose the base point $m \in \mathcal{P}(3)^{32\times 32}$ to be the constant image of unit matrices so that $n = \Lambda(m)$ consists of zero matrices. We initialize the algorithm with $p^{(0)} = f$ and $\xi_n^{(0)}$ as the zero vector. Our algorithm stops after 113 iterations, which take 96.20 seconds, when the value of (5.1) was below the value obtained by the CPPA. While the CPPA requires about half a second per iteration, our method requires a little less than a second per iteration, but it also requires only a fraction of the iteration count of CPPA. The behavior of the cost function is shown in Figure 3c, where the iterates are in log scale, since the "tail" of CPPA is quite long.

6.3. Dependence on the Point of Linearization. We mentioned previously that Algorithm 4.2 depends on the base points m and n and it cannot, in general, be expected to converge to a saddle point of (4.2) since it is based on the *linearized* saddle-point problem (4.14). In this experiment we illustrate the dependence of the limit of sequence of primal iterates on the base point m.

As data f we use the S2Whirl image designed by Johannes Persch in Laus et al., 2017, adapted to Manopt.jl, see Figure 4a. We set $\alpha = 1.5$ in the manifold-valued anisotropic TV (5.1). We employed Algorithm 4.2 with $\sigma = \tau = 0.35$ and $\gamma = 0.2$ and ran it for 300 iterations. The initial iterate is $p^{(0)} = f$ and $\xi_n^{(0)}$ as the zero vector.

We compare two different base points m. The first base point is the constant image whose value is the mean of all data pixels. The second base point is the constant image whose value is $p = (1, 0, 0)^{\mathrm{T}}$ ("west"). The final iterates are shown



(a) Original Data.

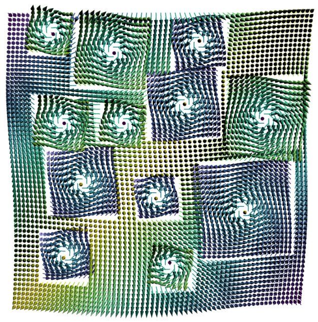
(b) Minimizer.

(c) Cost function.

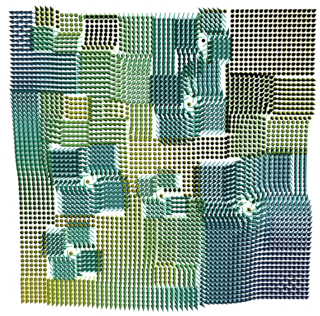
Fig. 3: Development of the three algorithms Cyclic Proximal Point (CPPA), parallel Douglas–Rachford (PDRA) as well as the linearized Riemannian Chambolle–Pock Algorithm 4.2 (lRCPA) starting all from the original data in (a) reaching the final value (image) in (b) is shown in (c), where the iterations on the x-axis are in log-scale.

in Figures 4b and 4c, respectively. The development of the cost function during the iterations is given in Figure 4d. Both runs yield piecewise constant solutions, but since their linearizations of Λ are using different base points and hence yield different linearized models. The resulting values of the cost function (5.1) differs, but both show a similar convergence behavior.

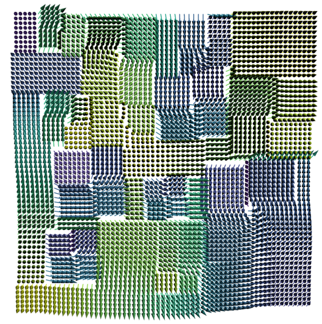
7. Conclusions. This paper introduces a novel concept of Fenchel duality for manifolds. We investigate properties of this novel duality concept and study corresponding primal-dual formulations of non-smooth optimization problems on manifolds. This leads to a novel primal-dual algorithm on manifolds, which comes in two variants, termed the exact and linearized Riemannian Chambolle–Pock algorithm. The convergence proof for the linearized version is given on arbitrary Hadamard manifolds under a suitable assumption. This accompanies the previous convergence proof for a comparable method, namely the Douglas–Rachford algorithm, where the proof is restricted to Hadamard manifolds of constant curvature. Numerical results illustrate not only that the linearized Riemannian Chambolle–Pock algorithm performs as well as state-of-the-art methods on Hadamard manifolds, but it also performs similarly well on manifolds with positive sectional curvature. Note that here it also has to deal with the absence of a global convexity concept of the functional.



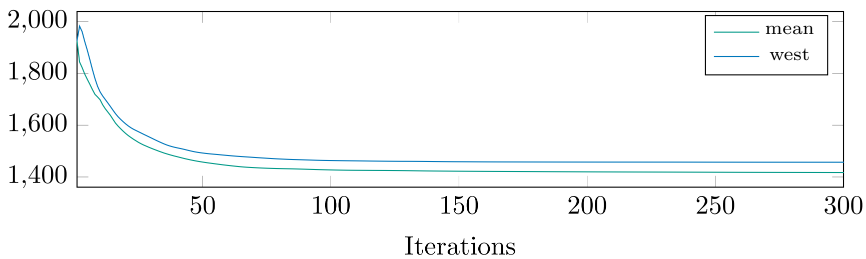
(a) The original S2Whirl data.



(b) The result with base m mean.



(c) The result with base m west.



(d) cost function

Fig. 4: The S2Whirl example illustrates that for manifolds with positive curvature, the algorithm still converges quite fast, but due to the nonconvexity of the distance, the effect of the linearization influences the result.

A more thorough investigation as well as a convergence proof for the exact variant are topics for future research. Another point of future research is an investigation of the choice of the base points $m \in \mathcal{M}$ and $n \in \mathcal{N}$ on the convergence, especially when the base points vary during the iterations.

Starting from the proper statement of the primal and dual problem for the linearization approach of Subsection 4.2, further aspects are open to investigation, for instance, regularity conditions ensuring strong duality. Well-known closedness-type conditions are then available, opening in this way a new line of rich research topics for optimization on manifolds.

Another point of potential future research is the measurement of the linearization error introduced by the model from Subsection 4.2. The analysis of the discrepancy term, as well as its behavior in the convergence of the linearized algorithm Algorithm 4.2, are closely related to the choice of the base points during the iteration, and should be considered in future research.

Furthermore, our novel concept of duality permits a definition of infimal convolution and thus offers a direct possibility to introduce the total generalized variation. In what way these novel priors correspond to existing ones, is another issue of ongoing research. Furthermore, the investigation of both a convergence rate as well as properties on manifolds with non-negative curvature are also open.

Acknowledgement. RB would like to thank Fjedor Gaede and Leon Bungert for fruitful discussions concerning the Chambolle–Pock algorithm in $\mathbb{R}^n$, especially concerning the choice of parameters as well as DT for hospitality in Münster and Erlangen. The authors would further like to thank Tuomo Valkonen for discussions on Hadamard manifolds and a three-point inequality remark, as well as Nicolas Boumal, Sebastian Neumayer and Gabriele Steidl for discussions and suggestions on preliminary versions of this manuscript. RB would like to acknowledge funding by the DFG project BE 5888/2. DT would like to acknowledge support within the EU grant No. 777826, the NoMADs project. RH and JVN would like to acknowledge the Priority Program SPP 1962 (*Non-smooth and Complementarity-based Distributed Parameter Systems: Simulation and Hierarchical Optimization*), which supported this work through the DFG grant HE 6077/10–1.

References.

Absil, P.-A.; Mahony, R.; Sepulchre, R. (2008). *Optimization Algorithms on Matrix Manifolds*. Princeton University Press. DOI: 10.1515/9781400830244.

Adams, B. L.; Wright, S. I.; Kunze, K. (1993). “Orientation imaging: The emergence of a new microscopy”. *Journal Metallurgical and Materials Transactions A* 24, pp. 819–831. DOI: 10.1007/BF02656503.

Ahmadi Kakavandi, B.; Amini, M. (2010). “Duality and subdifferential for convex functions on complete metric spaces”. *Nonlinear Analysis: Theory, Methods & Applications* 73.10, pp. 3450–3455. DOI: 10.1016/j.na.2010.07.033.

Bačák, M. (2014a). “Computing medians and means in Hadamard spaces”. *SIAM Journal on Optimization* 24.3, pp. 1542–1566. DOI: 10.1137/140953393.

Bačák, M.; Bergmann, R.; Steidl, G.; Weinmann, A. (2016). “A Second Order Non-Smooth Variational Model for Restoring Manifold-Valued Images”. *SIAM Journal on Scientific Computing* 38.1, A567–A597. DOI: 10.1137/15M101988X.

Bačák, M. (2014b). *Convex analysis and optimization in Hadamard spaces*. Vol. 22. De Gruyter Series in Nonlinear Analysis and Applications. De Gruyter, Berlin. DOI: 10.1515/9783110361629.

Bauschke, H. H.; Combettes, P. L. (2011). *Convex analysis and monotone operator theory in Hilbert spaces*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. With a foreword by Hédy Attouch. Springer, New York. DOI: 10.1007/978-1-4419-9467-7.

Bergmann, R.; Chan, R. H.; Hielscher, R.; Persch, J.; Steidl, G. (2016). “Restoration of Manifold-Valued Images by Half-Quadratic Minimization”. *Inverse Problems in Imaging* 10 (2), pp. 281–304. DOI: 10.3934/ipi.2016001.

Bergmann, R.; Laus, F.; Steidl, G.; Weinmann, A. (2014). “Second order differences of cyclic data and applications in variational denoising”. *SIAM Journal on Imaging Sciences* 7.4, pp. 2916–2953. DOI: 10.1137/140969993.

Bergmann, R.; Persch, J.; Steidl, G. (2016). “A parallel Douglas Rachford algorithm for minimizing ROF-like functionals on images with values in symmetric Hadamard manifolds”. *SIAM Journal on Imaging Sciences* 9.4, pp. 901–937. DOI: 10.1137/15M1052858.

Bergmann, R. (2019). *manopt.jl. Optimization on Manifolds in Julia.*

Bergmann, R.; Fitschen, J. H.; Persch, J.; Steidl, G. (2018). "Priors with Coupled First and Second Order Differences for Manifold-Valued Image Processing". *Journal of Mathematical Imaging and Vision* 60.9, pp. 1459–1481. DOI: 10.1007/s10851-018-0840-y.

Bergmann, R.; Gousenbourger, P.-Y. (2018). "A variational model for data fitting on manifolds by minimizing the acceleration of a Bézier curve". *Frontiers in Applied Mathematics and Statistics*. DOI: 10.3389/fams.2018.00059. arXiv: 1807.10090.

Bergmann, R.; Tenbrinck, D. (2018). "A graph framework for manifold-valued data". *SIAM Journal on Imaging Sciences* 11.1, pp. 325–360. DOI: 10.1137/17M1118567.

Bertsekas, D. P. (1978). "Local Convex Conjugacy and Fenchel Duality". *Preprints of Seventh World Congress of IFAC* 2, pp. 1079–1084.

Boţ, R. I. (2010). *Conjugate duality in convex optimization*. Vol. 637. Lecture Notes in Economics and Mathematical Systems. Berlin: Springer-Verlag. DOI: 10.1007/978-3-642-04900-2.

Boumal, N.; Mishra, B.; Absil, P.-A.; Sepulchre, R. (2014). "Manopt, a Matlab Toolbox for Optimization on Manifolds". *Journal of Machine Learning Research* 15, pp. 1455–1459.

Bredies, K.; Holler, M.; Storath, M.; Weinmann, A. (2018). "Total Generalized Variation for Manifold-Valued Data". *SIAM Journal on Imaging Sciences* 11.3, pp. 1785–1848. DOI: 10.1137/17M1147597.

Bredies, K.; Kunisch, K.; Pock, T. (2010). "Total generalized variation". *SIAM Journal on Imaging Sciences* 3.3, pp. 492–526. DOI: 10.1137/090769521.

Bürgmann, R.; Rosen, P. A.; Fielding, E. J. (2000). "Synthetic Aperture Radar Interferometry to Measure Earth′s Surface Topography and Its Deformation". *Annual Reviews Earth and Planetary Science* 28.1, pp. 169–209.

Chambolle, A. (2004). "An algorithm for total variation minimization and applications". *Journal of Mathematical Imaging and Vision* 20.1-2. Special issue on mathematics and image analysis, pp. 89–97. DOI: 10.1023/B:JMIV.0000011325.36760.1e.

Chambolle, A.; Caselles, V.; Cremers, D.; Novaga, M.; Pock, T. (2010). "An introduction to total variation for image analysis". *Theoretical foundations and numerical methods for sparse recovery*. Vol. 9. Radon Series on Computational and Applied Mathematics. Walter de Gruyter, Berlin, pp. 263–340. DOI: 10.1515/9783110226157.263.

Chambolle, A.; Lions, P.-L. (1997). "Image recovery via total variation minimization and related problems". *Numerische Mathematik* 76.2, pp. 167–188. DOI: 10.1007/s002110050258.

Chambolle, A.; Pock, T. (2011). "A first-order primal-dual algorithm for convex problems with applications to imaging". *Journal of Mathematical Imaging and Vision* 40.1, pp. 120–145. DOI: 10.1007/s10851-010-0251-1.

Chan, T.; Esedoglu, S.; Park, F.; Yip, A. (2006). "Total variation image restoration: overview and recent developments". *Handbook of mathematical models in computer vision*. Springer, New York, pp. 17–31. DOI: 10.1007/0-387-28831-7_2.

Chan, T.; Marquina, A.; Mulet, P. (2000). "High-order total variation-based image restoration". *SIAM Journal on Scientific Computing* 22.2, pp. 503–516. DOI: 10.1137/S1064827598344169.

Do Carmo, M. P. (1992). *Riemannian Geometry*. Mathematics: Theory & Applications. Birkhäuser Boston, Inc., Boston, MA.

Duran, J.; Moeller, M.; Sbert, C.; Cremers, D. (2016). "Collaborative total variation: a general framework for vectorial TV models". *SIAM Journal on Imaging Sciences* 9.1, pp. 116–151. DOI: 10.1137/15M102873X.

Ekeland, I.; Temam, R. (1999). *Convex Analysis and Variational Problems*. Vol. 28. Classics in Applied Mathematics. Philadelphia: SIAM.

Ferreira, O. P.; Oliveira, P. R. (1998). "Subgradient Algorithm on Riemannian Manifolds". *Journal of Optimization Theory and Applications* 97.1, pp. 93–104. DOI: 10.1023/A:1022675100677.

Ferreira, O. P.; Oliveira, P. R. (2002). "Proximal point algorithm on Riemannian manifolds". *Optimization. A Journal of Mathematical Programming and Operations Research* 51.2, pp. 257–270. DOI: 10.1080/02331930290019413.

Gabay, D.; Mercier, B. (1976). "A dual algorithm for the solution of nonlinear variational problems via finite element approximations". *Computer and Mathematics with Applications* 2, pp. 17–40. DOI: 10.1016/0898-1221(76)90003-1.

Grohs, P.; Sprecher, M. (2016). "Total variation regularization on Riemannian manifolds by iteratively reweighted minimization". *Information and Inference: A Journal of the IMA* 5.4, pp. 353–378. DOI: 10.1093/imaiai/iaw011.

Jost, J. (2017). *Riemannian geometry and geometric analysis*. 7th ed. Universitext. Springer, Cham. DOI: 10.1007/978-3-319-61860-9.

Kunze, K.; Wright, S. I.; Adams, B. L.; Dingley, D. J. (1993). "Advances in Automatic EBSP Single Orientation Measurements". *Textures and Microstructures* 20, pp. 41–54. DOI: 10.1155/TSM.20.41.

Laus, F.; Nikolova, M.; Persch, J.; Steidl, G. (2017). "A Nonlocal Denoising Algorithm for Manifold-Valued Images Using Second Order Statistics". *SIAM Journal on Imaging Sciences* 10.1, pp. 416–448. DOI: 10.1137/16M1087114.

Lee, J. M. (2003). *Introduction to smooth manifolds*. Vol. 218. Graduate Texts in Mathematics. Springer-Verlag, New York. DOI: 10.1007/978-0-387-21752-9.

Lellmann, J.; Strekalovskiy, E.; Koetter, S.; Cremers, D. (2013). "Total variation regularization for functions with values in a manifold". *IEEE ICCV 2013*, pp. 2944–2951. DOI: 10.1109/ICCV.2013.366.

Li, S.-L.; Li, C.; Liou, Y.-C.; Yao, J.-C. (2009). "Existence of solutions for variational inequalities on Riemannian manifolds". *Nonlinear Analysis* 71.11, pp. 5695–5706. DOI: 10.1016/j.na.2009.04.048.

Martínez-Legaz, J. E. (2005). "Generalized convex duality and its economic applications". *Handbook of generalized convexity and generalized monotonicity*. Vol. 76. Nonconvex Optimization and its Applications. Springer, New York, pp. 237–292. DOI: 10.1007/0-387-23393-8_6.

Papafitsoros, K.; Schönlieb, C. B. (2014). "A combined first and second order variational approach for image reconstruction". *Journal of Mathematical Imaging and Vision* 48.2, pp. 308–338. DOI: 10.1007/s10851-013-0445-4.

Pennec, X.; Fillard, P.; Ayache, N. (2006). "A Riemannian framework for tensor computing". *International Journal of Computer Vision* 66, pp. 41–66. DOI: 10.1007/s11263-005-3222-z.

Rockafellar, R. T. (1970). *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J.

Rockafellar, R. T. (1974). *Conjugate duality and optimization*. Lectures given at the Johns Hopkins University, Baltimore, Md., June, 1973, Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics, No. 16. Society for Industrial and Applied Mathematics.

Rudin, L. I.; Osher, S.; Fatemi, E. (1992). "Nonlinear total variation based noise removal algorithms". *Physica D* 60.1–4, pp. 259–268. DOI: 10.1016/0167-2789(92)90242-F.

Sakai, T. (1996). *Riemannian geometry.* Vol. 149. Translations of Mathematical Monographs. Translated from the 1992 Japanese original by the author. American Mathematical Society, Providence, RI.

Strekalovskiy, E.; Cremers, D. (2011). "Total variation for cyclic structures: Convex relaxation and efficient minimization". *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1905–1911. DOI: 10.1109/CVPR.2011.5995573.

Strong, D.; Chan, T. (2003). "Edge-preserving and scale-dependent properties of total variation regularization". *Inverse Problems. An International Journal on the Theory and Practice of Inverse Problems, Inverse Methods and Computerized Inversion of Data* 19.6. Special section on imaging, S165–S187. DOI: 10.1088/0266-5611/19/6/059.

Udrişte, C. (1994). *Convex functions and optimization methods on Riemannian manifolds.* Vol. 297. Mathematics and its Applications. Kluwer Academic Publishers Group, Dordrecht. DOI: 10.1007/978-94-015-8390-9.

Valkonen, T. (2014). "A primal–dual hybrid gradient method for nonlinear operators with applications to MRI". *Inverse Problems* 30.5, p. 055012. DOI: 10.1088/0266-5611/30/5/055012.

Wang, Y.; Yang, J.; Yin, W.; Zhang, Y. (2008). "A new alternating minimization algorithm for total variation image reconstruction". *SIAM Journal on Imaging Sciences* 1.3, pp. 248–272. DOI: 10.1137/080724265.

Weinmann, A.; Demaret, L.; Storath, M. (2014). "Total variation regularization for manifold-valued data". *SIAM Journal on Imaging Sciences* 7.4, pp. 2226–2257. DOI: 10.1137/130951075.

Zălinescu, C. (2002). *Convex analysis in general vector spaces.* World Scientific Publishing Co., Inc., River Edge, NJ. DOI: 10.1142/9789812777096.