

THE DIFFERENCE OF CONVEX ALGORITHM ON HADAMARD MANIFOLDS

Ronny Bergmann* Orizon P. Ferreira† Elianderson M. Santos‡
João Carlos de O. Souza§

May 2nd, 2023

In this paper, we propose a Riemannian version of the difference of convex algorithm (DCA) to solve a minimization problem involving the difference of convex (DC) function. We establish the equivalence between the classical and simplified Riemannian versions of the DCA. We also prove that, under mild assumptions, the Riemannian version of the DCA is well-defined, and every cluster point of the sequence generated by the proposed method, if any, is a critical point of the objective DC function. Additionally, we establish some duality relations between the DC problem and its dual. To illustrate the effectiveness of the algorithm, we present some numerical experiments.

Keywords. DC programming, DCA, Fenchel conjugate function, Riemannian manifolds

AMS subject classifications (MSC2010). 90C30 90C26 49N14 49Q99

1 INTRODUCTION

In this paper, we consider a general non-convex and non-smooth constrained optimization problem involving a difference of convex functions (shortly, *DC problem*) as follows

$$\arg \min_{p \in \mathcal{M}} f(p), \quad \text{where } f(p) := g(p) - h(p), \quad (1)$$

where the constrained set \mathcal{M} is endowed with a structure of a *complete, simply connected Riemannian manifold of non-positive sectional curvature*, i.e., a *Hadamard manifold*, the functions $g, h: \mathcal{M} \rightarrow \overline{\mathbb{R}}$,

*Norwegian University of Science and Technology, Department of Mathematical Sciences, Trondheim, Norway (ronny.bergmann@ntnu.no, <https://www.ntnu.edu/employees/ronny.bergmann>, orcid: 0000-0001-8342-7218).

†IME/UFG, Avenida Esperança, s/n, Campus Samambaia, Goiânia, GO, 74690-900, Brazil (orizon@ufg.br).

‡Instituto Federal de Educação, Ciência e Tecnologia do Maranhão, Barra do Corda, MA, 65950-000, Brazil (elianderson.santos@ifma.edu.br, orcid: 0000-0003-0936-9640).

§Department of Mathematics, Federal University of Piauí, Teresina, PI, Brazil (joacos.mat@ufpi.edu.br, orcid: 0000-0003-4053-8211)

are convex, lower semi-continuous and proper functions (called DC components), and $\overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ is the extended real line.

Due to the increasing number of optimization problems arising from practical applications posed in a Riemannian setting, the interest in this topic has increased significantly over the years. Even though we are not currently concerned with practical issues at this point, we emphasize that practical applications arise whenever the natural structure of the data is modeled as an optimization problem on a Riemannian manifold. For example, several problems in image processing, computational vision and signal processing can be modeled as problems in this setting. Papers dealing with this subject include Bačák et al. 2016; Bergmann, Persch, and Steidl 2016; Bergmann and Weinmann 2016; Bredies et al. 2018; Weinmann, Demaret, and Storath 2014, 2016, and problems in medical imaging modeled in this context are addressed in Esposito et al. 2019. Problems of tracking, robotics and scene motion analysis are also posed on Riemannian manifolds, as seen in Freifeld and Black 2012; Park, Bobrow, and Ploen 1995. Machine learning Nickel and Kiela 2018, artificial intelligence Muscoloni et al. 2017, neural circuits Sharpee 2019, low-rank approximations of hyperbolic embeddings Jawanpuria, Meghwanshi, and Mishra 2019; Tabaghi and Dokmanić 2020, Procrustes problems Tabaghi and Dokmanic 2021, financial networks Keller-Ressel and Nargang 2021, complex networks Krioukov et al. 2010; Moshiri, Safaei, and Samei 2021, embeddings of data Wilson et al. 2014 and strain analysis Vollmer 2018; Yamaji 2008 are some of the other areas where optimization problems on Riemannian manifolds are prevalent. Additionally, we mention that there are many papers on statistics in the Riemannian context, as seen in Bhattacharya and Bhattacharya 2008; Fletcher 2013.

As previously mentioned, there has been a significant increase in the number of works focusing on concepts and techniques of nonlinear programming and convex analysis in the Riemannian setting, see Absil, Mahony, and Sepulchre 2008; Udriște 1994. In addition to the theoretical issues addressed, which have an interest of their own, the Riemannian machinery provides support to design efficient algorithms to solve optimization problem in this setting; papers on this subject include Absil, Baker, and Gallivan 2007; Edelman, Arias, and Smith 1999; Huang, Gallivan, and Absil 2015; Li, Mordukhovich, et al. 2011; Manton 2015; Miller and Malick 2005; Nesterov and Todd 2002; Smith 1994; Wen and Yin 2012; Wang et al. 2015 and references therein.

Recently, the concept of the conjugate of a convex function was introduced in the Riemannian context. This is an important tool in convex analysis and plays a significant role in the theory of duality on Riemannian manifolds, see Silva Louzeiro, Bergmann, and Herzog 2022; Bergmann, Herzog, et al. 2021. In particular, this definition enables us to propose a Riemannian version of the DCA.

DC problems cover a broad class of non-convex optimization problems and DCA was the first method introduced especially for the standard DC problem Equation (1). It was proposed by Tao and Souad 1986 in the Euclidean setting. The basic idea behind DCA is to compute a subgradient of each (convex) DC component separately, i.e., at each iterate k , DCA calculates $y^{(k)} \in \partial h(x^{(k)})$ and uses this trial point to compute $x^{(k+1)} \in \partial g^*(y^{(k)})$, where ∂g^* denotes the subdifferential of the conjugate function of g in the sense of convex analysis. Equivalently, DCA approximates the second DC component $h(x)$ by its affine minorization $h_k(x) = h(x^{(k)}) + \langle x - x^{(k)}, y^{(k)} \rangle$, with $y^{(k)} \in \partial h(x^{(k)})$, and minimizes the resulting convex function

$$x^{(k+1)} \in \arg \min_{x \in \mathbb{R}^n} g(x) - h_k(x),$$

which is called the alternative version of DCA therein. On the other hand, computing $y^{(k)} \in \partial h(x^{(k)})$ is equivalent to find a solution of the dual problem

$$\arg \min_{y \in \mathbb{R}^n} h^*(y) - g^*(y^{k-1}) - \langle y - y^{k-1}, x^{(k)} \rangle.$$

Therefore, DCA can also be viewed as an iterative primal-dual subgradient method.

DC optimization algorithms have been proved to be particularly successful for analyzing and solving a variety of highly structured and practical problems; see for instance de Oliveira 2020; Thi and Pham Dinh 2018; An and Tao 2005. To the best of our knowledge, the work in Souza and Oliveira 2015 was the first to deal with DC functions in Riemannian manifolds. More precisely, the authors proposed the proximal point algorithm for DC functions (DCPPA) and studied the convergence of the method in Hadamard manifolds. Recently, Almeida et al. 2020 proposed a modified version of the DCPPA in the same Riemannian setting in order to accelerate the convergence of the method considered in Souza and Oliveira 2015.

The aim of this paper is to propose, for the first time, a Riemannian version of the DCA. We obtain an equivalence between the classical and a simplified version of the Riemannian DCA. Therefore, under mild assumptions, we prove that the Riemannian DCA is well-defined, and every cluster point of the sequence generated by the proposed method, if any, is a critical point of the objective DC function in Equation (1). We also extend some relations between the DC problem Equation (1) and its dual to the Riemannian setting. To illustrate the effectiveness of DCA, we present some numerical experiments.

This paper is organized as follows. In Section 2 we present some notations and preliminary results that will be used throughout the paper. In Section 3 some relations between the DC problem and its dual are established on Hadamard manifolds. In Section 4 we present a formulation of the Riemannian DCA. In Section 5 we study the convergence properties of the proposed method. In Section 7 we provide some applications to the problem of maximizing a convex function in a compact set and manifold-valued image denoising. Finally, Section 8 presents some conclusions.

2 PRELIMINARIES

In this section, we recall some concepts, notations, and basics results about Riemannian manifolds and optimization. For more details see, for example, do Carmo 1992; Rapcsák 1997; Sakai 1996; Udriște 1994. Let us begin with concepts about Riemannian manifolds. We denote by \mathcal{M} a finite dimensional Riemannian manifold and by $T_p\mathcal{M}$ the *tangent space* of \mathcal{M} at p . The corresponding norm associated to the Riemannian metric $\langle \cdot, \cdot \rangle$ is denoted by $\|\cdot\|$. Moreover, the *tangent bundle* of \mathcal{M} , will be denoted by $T\mathcal{M}$. We use $\ell(\gamma)$ to denote the length of a piecewise smooth curve $\gamma: [a, b] \rightarrow \mathcal{M}$. The Riemannian distance between p and q in \mathcal{M} is denoted by $d(p, q)$, which induces the original topology on \mathcal{M} , namely, (\mathcal{M}, d) , which is a complete metric space. A complete, simply connected Riemannian manifold of non-positive sectional curvature is called a Hadamard manifold. *All Riemannian manifold considered in this paper will be Hadamard manifolds and will be denoted \mathcal{M} .* For a $p \in \mathcal{M}$, the exponential map $\exp_p: T_p\mathcal{M} \rightarrow \mathcal{M}$ is a diffeomorphism and $\exp_p^{-1}: \mathcal{M} \rightarrow T_p\mathcal{M}$ denotes its inverse. In

this case, $d(q, p) = \|\exp_p^{-1} q\|$ holds, the function $d_q^2: \mathcal{M} \rightarrow \mathbb{R}$ defined by $d_q^2(p) := d^2(q, p)$ is C^∞ and its gradient is given by $\text{grad } d_q^2(p) = -2 \exp_p^{-1} q$. Now, we recall some concepts and basic properties about optimization in the Riemannian context. For that, given two points $p, q \in \mathcal{M}$, denotes by γ_{pq} the geodesic segment joining p to q , i.e., $\gamma_{pq}: [0, 1] \rightarrow \mathcal{M}$ with $\gamma_{pq}(0) = p$ and $\gamma_{pq}(1) = q$. We denote by $\overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ the extended real line. The domain of a function $f: \mathcal{M} \rightarrow \overline{\mathbb{R}}$ is denoted by $\text{dom}(f) := \{p \in \mathcal{M} : f(p) < +\infty\}$. The function f is said to be *convex* (resp. *strictly convex*) if, for any $p, q \in \mathcal{M}$, the composition $f \circ \gamma_{pq}: [0, 1] \rightarrow \overline{\mathbb{R}}$ is convex (resp. strictly convex), i.e., $(f \circ \gamma_{pq})(t) \leq (1-t)f(p) + tf(q)$ (resp. $(f \circ \gamma_{pq})(t) < (1-t)f(p) + tf(q)$), for all $t \in [0, 1]$. A function $f: \mathcal{M} \rightarrow \overline{\mathbb{R}}$ is said to be σ -strongly convex for $\sigma > 0$ if, for any $p, q \in \mathcal{M}$, the composition $f \circ \gamma_{pq}: [0, 1] \rightarrow \overline{\mathbb{R}}$ is σ -strongly convex, i.e., $(f \circ \gamma_{pq})(t) \leq (1-t)f(p) + tf(q) - \frac{\sigma}{2}t(1-t)d^2(q, p)$, for all $t \in [0, 1]$.

Definition 2.1. The subdifferential of a proper, convex function $f: \mathcal{M} \rightarrow \overline{\mathbb{R}}$ at $p \in \text{dom}(f)$ is the set

$$\partial f(p) := \{X \in T_p \mathcal{M} : f(q) \geq f(p) + \langle X, \exp_p^{-1} q \rangle, \quad \text{for all } q \in \mathcal{M}\}.$$

The proof of the first item of the following theorem can be found in Udriște 1994, Theorem 4.10, p. 76, while the proof of the second one follows the same idea as the first one.

Theorem 2.2. Let $f: \mathcal{M} \rightarrow \overline{\mathbb{R}}$ be a function. Then,

- i) The function f is convex if and only if $f(p) \geq f(q) + \langle X, \exp_q^{-1} p \rangle$, for all $p, q \in \mathcal{M}$ and all $X \in \partial f(q)$.
- ii) The function f is σ -strongly convex if and only if $f(p) \geq f(q) + \langle X, \exp_q^{-1} p \rangle + \frac{\sigma}{2}d^2(p, q)$, for all $p, q \in \mathcal{M}$ and all $X \in \partial f(q)$.

The following definition play an import role in the paper, see Bourbaki 1995, p. 363.

Definition 2.3. A function $f: \mathcal{M} \rightarrow \overline{\mathbb{R}}$ is said to be lower semi-continuous (lsc), at $p \in \mathcal{M}$ if $\liminf_{q \rightarrow p} f(q) = f(p)$. If f is lower semi-continuous at all points along \mathcal{M} , we simply state that f is lower semi-continuous.

The proof of the following result is an immediate consequence of Wang et al. 2015, Proposition 2.5.

Proposition 2.4. Let $f: \mathcal{M} \rightarrow \overline{\mathbb{R}}$ be a convex and lower semi-continuous function. Consider the sequence $(p^{(k)})_{k \in \mathbb{N}} \subset \text{int dom}(f)$ such that $\lim_{k \rightarrow \infty} p^{(k)} = \bar{p} \in \text{int dom}(f)$. If $(X^{(k)})_{k \in \mathbb{N}}$ is a sequence such that $X^{(k)} \in \partial f(p^{(k)})$ for every $k \in \mathbb{N}$, then $(X^{(k)})_{k \in \mathbb{N}}$ is bounded and its cluster points belongs to $\partial f(\bar{p})$.

Definition 2.5. A function $f: \mathcal{M} \rightarrow \overline{\mathbb{R}}$ is said to be 1-coercive if there exists a point $\bar{p} \in \mathcal{M}$ such that

$$\lim_{d(\bar{p}, p) \rightarrow +\infty} \frac{f(p)}{d(\bar{p}, p)} = +\infty.$$

The *global minimizer set* of a function $f: \mathcal{M} \rightarrow \overline{\mathbb{R}}$ is defined by

$$\Omega^* := \{q \in \mathcal{M} : f(q) \leq f(p), \text{ for all } p \in \mathcal{M}\}.$$

Proposition 2.6. *Assume that $f: \mathcal{M} \rightarrow \overline{\mathbb{R}}$ is lsc and 1-coercive. Then the global minimizer set of f is non-empty.*

Proof. Take $\bar{p} \in \mathcal{M}$ such that $\lim_{d(\bar{p}, p) \rightarrow +\infty} (f(p)/d(\bar{p}, p)) = +\infty$. In particular, we conclude that $\lim_{d(\bar{p}, p) \rightarrow +\infty} f(p) = +\infty$. Thus, there exists $\bar{r} > 0$ such that $\bar{r} < d(\bar{p}, p)$ implies that $f(\bar{p}) \leq f(p)$. Consider the set $B[\bar{p}, \bar{r}] := \{p \in \mathcal{M} : d(p, \bar{p}) \leq \bar{r}\}$. Since \mathcal{M} is a Hadamard manifold, the Hopf-Rinow theorem ensures that $B[\bar{p}, \bar{r}]$ is compact. Thus, taking into account that f is lsc., by Bourbaki 1995, Theorem 3, p. 361 there exists $\hat{p} \in B[\bar{p}, \bar{r}]$ such that $f(\hat{p}) \leq f(p)$, for all $p \in B[\bar{p}, \bar{r}]$. Therefore, \bar{p} or \hat{p} is a global minimizer of f . \square

Lemma 2.7. *Let $g: \mathcal{M} \rightarrow \mathbb{R}$ be a σ -strongly convex function. Take $\bar{p} \in \mathcal{M}$ and $X \in T_{\bar{p}}\mathcal{M}$. Then, the function $f: \mathcal{M} \rightarrow \mathbb{R}$ defined by $f(p) = g(p) - \langle X, \exp_{\bar{p}}^{-1} p \rangle$ is 1-coercive. Consequently, the global minimizer set of f is non-empty.*

Proof. Since the function $g: \mathcal{M} \rightarrow \mathbb{R}$ is a σ -strongly convex, [Theorem 2.2 Item i\)](#) implies that

$$g(p) \geq g(\bar{p}) + \langle Y, \exp_{\bar{p}}^{-1} p \rangle + \frac{\sigma}{2} d^2(\bar{p}, p), \quad \text{for all } p \in \mathcal{M} \text{ and all } Y \in \partial g(\bar{p}).$$

Thus, considering that $f(p) = g(p) - \langle X, \exp_{\bar{p}}^{-1} p \rangle$ and using the last inequality we conclude

$$\frac{f(p)}{d(\bar{p}, p)} \geq \frac{g(\bar{p})}{d(\bar{p}, p)} + \left\langle Y, \frac{\exp_{\bar{p}}^{-1} p}{d(\bar{p}, p)} \right\rangle + \frac{\sigma}{2} d(\bar{p}, p) - \left\langle X, \frac{\exp_{\bar{p}}^{-1} p}{d(\bar{p}, p)} \right\rangle, \quad \text{for all } Y \in \partial g(\bar{p}).$$

Since $d(\bar{p}, p) = \|\exp_{\bar{p}}^{-1} p\|$, we obtain that the inner products in the last inequality are bounded. Hence, we have

$$\lim_{d(\bar{p}, p) \rightarrow +\infty} \frac{f(p)}{d(\bar{p}, p)} = +\infty.$$

Therefore, f is 1-coercive. The second part of the proposition is an immediate consequence of the first one combined with [Proposition 2.6](#). \square

The statement and proof of the next proposition can be found in Li, López, and Martín-Márquez 2009, Lemma 2.4, p. 666.

Proposition 2.8. *Let $\bar{p} \in \mathcal{M}$ and $(p^{(k)})_{k \in \mathbb{N}} \subset \mathcal{M}$ be such that $\lim_{k \rightarrow +\infty} p^{(k)} = \bar{p}$. Then the following assertions hold:*

$$i) \text{ For any } p \in \mathcal{M}, \text{ we have } \lim_{k \rightarrow +\infty} \exp_{p^{(k)}}^{-1} p = \exp_{\bar{p}}^{-1} p \text{ and } \lim_{k \rightarrow +\infty} \exp_p^{-1} p^{(k)} = \exp_p^{-1} \bar{p}.$$

- ii) If $X^{(k)} \in T_{p^{(k)}}\mathcal{M}$ and $\lim_{k \rightarrow +\infty} X^{(k)} = \bar{X}$, then $\bar{X} \in T_{\bar{p}}\mathcal{M}$.
- iii) Given $X^{(k)} \in T_{p^{(k)}}\mathcal{M}$, $Y^{(k)} \in T_{p^{(k)}}\mathcal{M}$, $\bar{X} \in T_{\bar{p}}\mathcal{M}$, and $\bar{Y} \in T_{\bar{p}}\mathcal{M}$. If $\lim_{k \rightarrow +\infty} X^{(k)} = \bar{X}$ and $\lim_{k \rightarrow +\infty} Y^{(k)} = \bar{Y}$, then $\lim_{k \rightarrow +\infty} \langle X^{(k)}, Y^{(k)} \rangle = \langle \bar{X}, \bar{Y} \rangle$.

We end this section recalling several results from Fenchel duality on Hadamard manifolds, which play an important role in the following sections. It is worth emphasizing that we are limiting our study to finite-dimensional manifolds and that our emphasis is algorithmic. Consequently, we do not need to use cotangent space for our purposes. Due to this, we decided to exclusively employ tangent spaces in the following results of the paper Silva Louzeiro, Bergmann, and Herzog 2022. We begin by recalling the defining the conjugate of a proper function.

Definition 2.9. Let $f: \mathcal{M} \rightarrow \overline{\mathbb{R}}$ be a proper function. The Fenchel conjugate of f is the function $f^*: T\mathcal{M} \rightarrow \overline{\mathbb{R}}$ defined by

$$f^*(p, X) := \sup_{q \in \mathcal{M}} \{ \langle X, \exp_p^{-1} q \rangle - f(q) \}, \quad (p, X) \in T\mathcal{M}.$$

Theorem 2.10. Let $f: \mathcal{M} \rightarrow \overline{\mathbb{R}}$ be a proper function. Then, the Fenchel-Young inequality holds, i. e., for all $(p, X) \in T\mathcal{M}$ we have

$$f(q) + f^*(p, X) \geq \langle X, \exp_p^{-1} q \rangle, \quad \text{for all } q \in \mathcal{M}.$$

Theorem 2.11. Let $f: \mathcal{M} \rightarrow \overline{\mathbb{R}}$ be a proper lsc convex function and $p \in \mathcal{M}$. Then the function $f^*(p, \cdot): T_p\mathcal{M} \rightarrow \overline{\mathbb{R}}$ is convex and proper.

Definition 2.12. Let $p \in \mathcal{M}$ and suppose that $f^*(p, \cdot): T_p\mathcal{M} \rightarrow \overline{\mathbb{R}}$ is proper. The subdifferential of $f^*(p, \cdot)$ at $X \in T_p\mathcal{M}$, denoted by $\partial_2 f^*(p, X)$, is the set

$$\partial_2 f^*(p, X) = \{ Y \in T_p\mathcal{M} : f^*(p, Z) \geq f^*(p, X) + \langle Z - X, Y \rangle, \quad \text{for all } Z \in T_p\mathcal{M} \}.$$

Combining Silva Louzeiro, Bergmann, and Herzog 2022, Remark 3.3 with Bergmann, Herzog, et al. 2021, Corollary 3.16, we obtain the following result.

Theorem 2.13. Let $f: \mathcal{M} \rightarrow \overline{\mathbb{R}}$ be a proper function and $p \in \mathcal{M}$. Then, $Y \in \partial_2 f^*(p, X)$ if and only if $f(\exp_p Y) + f^*(p, X) = \langle X, Y \rangle$.

Remark 2.14 (Silva Louzeiro, Bergmann, and Herzog 2022, Remark 3.4). If $\mathcal{M} = \mathbb{R}^n$, then $f^*(p, X) = f^*(X) - \langle X, p \rangle$. Moreover, $\partial_2 f^*(p, X) := \partial f^*(X) + \{-p\}$.

Definition 2.15. The Fenchel biconjugate of a function $f: \mathcal{M} \rightarrow \overline{\mathbb{R}}$ is the function $f^{**}: \mathcal{M} \rightarrow \overline{\mathbb{R}}$ defined by

$$f^{**}(p) := \sup_{(q, X) \in T\mathcal{M}} \{ \langle X, \exp_p^{-1} q \rangle - f^*(q, X) \}, \quad \text{for all } p \in \mathcal{M}.$$

Theorem 2.16. Let $f: \mathcal{M} \rightarrow \overline{\mathbb{R}}$ be a proper lsc convex function. Then, $f^{**} = f$ holds.

Theorem 2.17. Let $f: \mathcal{M} \rightarrow \overline{\mathbb{R}}$ be a proper convex function. Then,

$$X \in \partial f(p) \text{ if and only if } f^*(p, X) = -f(p).$$

3 DUALITY IN DC OPTIMIZATION IN HADAMARD MANIFOLDS

In this section our aim is to state and study difference of convex optimization problem or DC problem, and its dual problem, called dual DC problem, in the Hadamard setting. The DC problem is defined as follows

$$\arg \min_{p \in \mathcal{M}} f(p), \quad \text{where } f(p) := g(p) - h(p), \quad (2)$$

and $g: \mathcal{M} \rightarrow \overline{\mathbb{R}}$ and $h: \mathcal{M} \rightarrow \overline{\mathbb{R}}$ are proper, lsc and convex functions. The DC problem is a non-convex and, in general, a non-smooth problem. In the following we further use the conventions

$$(+\infty) - (+\infty) = +\infty, \quad (+\infty) - \lambda = +\infty, \quad \text{and } \lambda - (+\infty) = -\infty, \quad \text{for all } \lambda \in \mathbb{R}. \quad (3)$$

Similarly to Euclidean context, see Tao and Souad 1988, the dual DC problem of the problem (2), is stated as follows

$$\arg \min_{(p, X) \in T\mathcal{M}} \varphi(p, X), \quad \text{where } \varphi(p, X) := h^*(p, X) - g^*(p, X). \quad (4)$$

Additional detail concerning the appropriateness of the previous definition will be provided later in Theorem 3.10. In the following remark, we will look at the details of the relationship between (4) and its Euclidean counterpart.

Remark 3.1. If $\mathcal{M} = \mathbb{R}^n$, then $T_p\mathcal{M} \simeq \mathbb{R}^n$ for all $p \in \mathcal{M}$. Consequently, $T\mathcal{M} \simeq \mathbb{R}^n$. Moreover, by using Remark 2.14, we obtain

$$h^*(p, X) - g^*(p, X) = h^*(X) - \langle X, p \rangle - (g^*(X) - \langle X, p \rangle) = h^*(X) - g^*(X), \quad \text{for all } X \in \mathbb{R}^n.$$

Therefore, if $\mathcal{M} = \mathbb{R}^n$ then problem (4) simplifies to

$$\arg \min_{X \in \mathbb{R}^n} h^*(X) - g^*(X). \quad (5)$$

In conclusion, for $\mathcal{M} = \mathbb{R}^n$, the dual (4) of the problem (2) merges into the dual stated in Tao and Souad 1988.

To proceed with the study of problems (2) and (4), for now on we will assume that:

- A1) $g: \mathcal{M} \rightarrow \overline{\mathbb{R}}$ and $h: \mathcal{M} \rightarrow \overline{\mathbb{R}}$ are σ -strongly convex and lsc functions, where $\sigma > 0$;
- A2) $f_{\inf} := \inf_{x \in \mathcal{M}} f(x) > -\infty$;

A3) $\text{dom}(g) \subseteq \text{int dom}(h)$;

A4) $\partial_2 g^*(p, X) \neq \emptyset$, for every $X \in \text{dom}(g^*(p, \cdot)) := \{X \in T_p \mathcal{M} : g^*(p, X) < +\infty\}$.

Next, we discuss the above assumptions. First, we show that (A1) is not restrictive.

Remark 3.2. Let $q \in \mathcal{M}$ and $\sigma > 0$. Consider the function $\mathcal{M} \ni p \mapsto \frac{\sigma}{2} d^2(q, p)$, which is σ -strongly convex, see Neto, Ferreira, and Pérez 2002, Corollary 3.1. If $\tilde{g}: \mathcal{M} \rightarrow \mathbb{R}$ and $\tilde{h}: \mathcal{M} \rightarrow \mathbb{R}$ are convex, then taking $g \in \mathcal{M}$ and setting $g(p) = \tilde{g}(p) + \frac{\sigma}{2} d^2(q, p)$ and $h(p) = \tilde{h}(p) + \frac{\sigma}{2} d^2(q, p)$ we obtain two σ -strongly convex functions g and h in \mathcal{M} . In addition, $f(p) = \tilde{g}(p) - \tilde{h}(p) = g(p) - h(p)$, for all $p \in \mathcal{M}$.

Remark 3.3. If assumption (A2) holds, then $\text{dom}(f) = \text{dom}(g) \subseteq \text{dom}(h)$. Indeed, if $\text{dom}(g) \not\subseteq \text{dom}(h)$, then there exists $p \in \text{dom}(g)$ such that $p \notin \text{dom}(h)$, and hence by (3), we have that $f(p) = g(p) - h(p) = g(p) - (+\infty) = -\infty$, which contradicts assumption (A2). Thus, $\text{dom}(g) \subseteq \text{dom}(h)$, which implies that $\text{dom}(g) \subseteq \text{dom}(f)$. On the other hand, assume by contradiction that $\text{dom}(f) \not\subseteq \text{dom}(g)$. Then, there exists $p \in \text{dom}(f)$ such that $g(p) = +\infty$. From (3) we obtain that $f(p) = g(p) - h(p) = (+\infty) - h(p) = +\infty$, which contradicts the fact that $p \in \text{dom}(f)$. Therefore, we conclude that $\text{dom}(f) = \text{dom}(g)$. Since under assumption (A2), we have $\text{dom}(f) = \text{dom}(g)$, which implies that $\text{dom}(g) \subseteq \text{dom}(h)$. Hence, assumption (A3) is only slightly more restrictive than assumption (A2). We also note that if $\text{dom}(h) = \mathcal{M}$, then assumption (A3) holds, and if $\text{dom}(g^*(p, \cdot)) = T_p \mathcal{M}$, then assumption (A4) holds. It is worth to note that assumption (A4) is used here to establish the relationship between problems (2) and (4).

A necessary condition for the point $p^* \in \mathcal{M}$ to be a local minimum of $f = g - h$ is that $0 \in \partial f(p^*) \subset \partial g(p^*) - \partial h(p^*)$. Hence, if $p^* \in \mathcal{M}$ is the solution of problem (2), then $\partial h(p^*) \subset \partial g(p^*)$. Consequently, $\partial g(p^*) \cap \partial h(p^*) \neq \emptyset$. In this sense, we define a *critical point* of problem (2).

Definition 3.4. A point $p^* \in \mathcal{M}$ is a *critical point* of f in (2) if $\partial g(p^*) \cap \partial h(p^*) \neq \emptyset$.

The next lemma establishes a necessary condition for a point $(\bar{p}, \bar{X}) \in T\mathcal{M}$ be a solution of problem (4).

Lemma 3.5. If (\bar{p}, \bar{X}) is a solution of problem (4), then $\partial_2 g^*(\bar{p}, \bar{X}) \subseteq \partial_2 h^*(\bar{p}, \bar{X})$ holds.

Proof. Let (\bar{p}, \bar{X}) be a solution of problem (4). Then, $h^*(p, Y) - g^*(p, Y) \geq h^*(\bar{p}, \bar{X}) - g^*(\bar{p}, \bar{X})$, for all $(p, Y) \in T\mathcal{M}$. Thus, we have

$$h^*(\bar{p}, Y) - h^*(\bar{p}, \bar{X}) \geq g^*(\bar{p}, Y) - g^*(\bar{p}, \bar{X}), \quad \text{for all } Y \in T_{\bar{p}} \mathcal{M}.$$

Take $Z \in \partial_2 g^*(\bar{p}, \bar{X})$. By Definition 2.12, we have $g^*(\bar{p}, Y) - g^*(\bar{p}, \bar{X}) \geq \langle Y - \bar{X}, Z \rangle$, for all $Y \in T_{\bar{p}} \mathcal{M}$, which combined with the last inequality yields

$$h^*(\bar{p}, Y) - h^*(\bar{p}, \bar{X}) \geq \langle Y - \bar{X}, Z \rangle, \quad \text{for all } Y \in T_{\bar{p}} \mathcal{M}.$$

This implies, by Definition 2.12, that $v \in \partial_2 h^*(\bar{p}, \bar{X})$, and the statement is proved. \square

Remark 3.6. If $\mathcal{M} = \mathbb{R}^n$, then by using [Remark 2.14](#) we have $\partial_2 g^*(\bar{p}, \bar{X}) = \partial g^*(\bar{X}) - \{\bar{p}\}$ and $\partial_2 h^*(\bar{p}, \bar{X}) = \partial h^*(\bar{X}) - \{\bar{p}\}$. Thus, from [Remark 3.1](#) and [Lemma 3.5](#) we conclude that if $\bar{X} \in \mathbb{R}^n$ is a solution of problem (5), then we have $\partial g^*(\bar{X}) \subseteq \partial h^*(\bar{X})$, which yields [Tao and Souad 1988, Theorem 2.1 \(2\)](#).

We have already defined the critical point for the primal problem in [Definition 3.4](#), so let us continue on dual problem. Please keep in mind that, it follows from [Lemma 3.5](#), that if (\bar{p}, \bar{X}) is a solution of the problem (4), then the set $\partial_2 h^*(\bar{p}, \bar{X}) \cap \partial_2 g^*(\bar{p}, \bar{X})$ is non-empty. Hence, we define the notion of critical point for the problem (4) as follows:

Definition 3.7. A point (\bar{p}, \bar{X}) is a critical point for problem (4) if $\partial_2 h^*(\bar{p}, \bar{X}) \cap \partial_2 g^*(\bar{p}, \bar{X}) \neq \emptyset$.

Remark 3.8. If $\mathcal{M} = \mathbb{R}^n$, then by using [Remark 2.14](#) we have $\partial_2 g^*(\bar{p}, \bar{X}) = \partial g^*(\bar{X}) - \{\bar{p}\}$ and $\partial_2 h^*(\bar{p}, \bar{X}) = \partial h^*(\bar{X}) - \{\bar{p}\}$. Thus, if (\bar{p}, \bar{X}) is a critical point of problem (4), then there exist $Z \in \partial_2 h^*(\bar{p}, \bar{X}) \cap \partial_2 g^*(\bar{p}, \bar{X})$. Hence, $Z + \bar{p} \in \partial g^*(\bar{X}) \cap \partial h^*(\bar{X}) \neq \emptyset$. Therefore, $\bar{X} \in \mathbb{R}^n$ is a critical point of problem (5).

To proceed with our analysis we need the next lemma. For a proof of it see [Bartle and Sherbert 2000, p. 46](#).

Lemma 3.9. Let X and Y be non-empty sets and $f: X \times Y \rightarrow \mathbb{R}$ a function. Then, it holds

$$\inf_{(x,y) \in X \times Y} f(x,y) = \inf_{x \in X} \inf_{y \in Y} f(x,y) = \inf_{y \in Y} \inf_{x \in X} f(x,y).$$

The next theorem presents the relation between the optimum values of problems (2) and (4).

Theorem 3.10. Let $g: \mathcal{M} \rightarrow \overline{\mathbb{R}}$ and $h: \mathcal{M} \rightarrow \overline{\mathbb{R}}$ be proper, lsc and convex functions. Then, there holds

$$\inf_{(q,X) \in T\mathcal{M}} \left\{ h^*(q, X) - g^*(q, X) \right\} = \inf_{p \in \mathcal{M}} \{g(p) - h(p)\}.$$

Proof. Since h is convex, [Theorem 2.16](#) implies that $h^{**} = h$. Thus, using [Definition 2.15](#) we have

$$\begin{aligned} \inf_{p \in \mathcal{M}} \{g(p) - h(p)\} &= \inf \{g(p) - h^{**}(p) : p \in \mathcal{M}\} \\ &= \inf \left\{ g(p) - \sup_{(q,X) \in T\mathcal{M}} \left\{ \langle X, \exp_q^{-1} p \rangle - h^*(q, X) \right\} : p \in \mathcal{M} \right\}. \end{aligned}$$

Since $\sup_{(q,X) \in T\mathcal{M}} \{ \langle X, \exp_q^{-1} p \rangle - h^*(q, X) \} = - \inf_{(q,X) \in T\mathcal{M}} \{ h^*(q, X) - \langle X, \exp_q^{-1} p \rangle \}$, the last equality is equivalent to

$$\inf_{p \in \mathcal{M}} \{g(p) - h(p)\} = \inf_{p \in \mathcal{M}} \inf_{(q,X) \in T\mathcal{M}} \left\{ g(p) + h^*(q, X) - \langle X, \exp_q^{-1} p \rangle \right\},$$

which, using [Lemma 3.9](#), can still be expressed equivalently as

$$\inf_{p \in \mathcal{M}} \{g(p) - h(p)\} = \inf_{(q, X) \in T\mathcal{M}} \inf_{p \in \mathcal{M}} \{g(p) + h^*(q, X) - \langle X, \exp_q^{-1} p \rangle\}.$$

Due to $\inf_{p \in \mathcal{M}} \{g(p) + h^*(q, X) - \langle X, \exp_q^{-1} p \rangle\} = h^*(q, X) - \sup_{p \in \mathcal{M}} \langle X, \exp_q^{-1} p \rangle - \{g(p)\}$, the final equality is as follows

$$\inf_{p \in \mathcal{M}} \{g(p) - h(p)\} = \inf_{(q, X) \in T\mathcal{M}} \left\{ h^*(q, X) - \sup_{p \in \mathcal{M}} \{ \langle X, \exp_q^{-1} p \rangle - g(p) \} \right\},$$

which, by using [Definition 2.9](#), yields the desired equality and the proof is concluded. \square

Theorem 3.11. *The following statements hold:*

- i) *If $\bar{p} \in \mathcal{M}$ is a solution of problem (2), then $(\bar{p}, \bar{Y}) \in T\mathcal{M}$ is a solution of the problem (4), for all $\bar{Y} \in \partial h(\bar{p}) \cap \partial g(\bar{p})$.*
- ii) *If $(\bar{p}, \bar{Y}) \in T\mathcal{M}$ is a solution of problem (4), for some $\bar{Y} \in \partial h(\bar{p}) \cap \partial g(\bar{p})$, then $\bar{p} \in \mathcal{M}$ is a solution of problem (2).*

Proof. To prove [Item i](#)), assume that $\bar{p} \in \mathcal{M}$ is a solution of problem (2). Thus, we have $\partial h(\bar{p}) \cap \partial g(\bar{p}) \neq \emptyset$. Let $\bar{Y} \in \partial h(\bar{p}) \cap \partial g(\bar{p})$. Since g and h are convex, by [Theorem 2.17](#) we have $-g^*(\bar{p}, \bar{Y}) = g(\bar{p})$ and $h^*(\bar{p}, \bar{Y}) = -h(\bar{p})$, which implies that $h^*(\bar{p}, \bar{Y}) - g^*(\bar{p}, \bar{Y}) = g(\bar{p}) - h(\bar{p})$. Using again that $\bar{p} \in \mathcal{M}$ is a solution of problem (2), the last equality together with [Theorem 3.10](#) ensure that (\bar{p}, \bar{Y}) is a solution of problem (4), and hence, the [Item i](#)) is proved. We proceed to prove [Item ii](#)). To this end, we assume that (\bar{p}, \bar{Y}) is a solution of problem (4) with $\bar{Y} \in \partial h(\bar{p}) \cap \partial g(\bar{p})$. Since g and h are convex and $\bar{Y} \in \partial h(\bar{p}) \cap \partial g(\bar{p})$, it follows from [Theorem 2.17](#) that $-g^*(\bar{p}, \bar{Y}) = g(\bar{p})$ and $h^*(\bar{p}, \bar{Y}) = -h(\bar{p})$, which implies

$$g(\bar{p}) - h(\bar{p}) = h^*(\bar{p}, \bar{Y}) - g^*(\bar{p}, \bar{Y}) = \inf_{(p, X) \in T\mathcal{M}} \{h^*(p, X) - g^*(p, X)\}. \quad (6)$$

On the other hand, [Theorem 3.10](#) implies that

$$\inf_{(p, X) \in T\mathcal{M}} \{h^*(p, X) - g^*(p, X)\} = \inf_{q \in \mathcal{M}} \{g(q) - h(q)\} \leq g(\bar{p}) - h(\bar{p})$$

Combining the last inequality with (6) yields $g(\bar{p}) - h(\bar{p}) = \inf_{q \in \mathcal{M}} \{g(q) - h(q)\}$. Hence, $\bar{p} \in \mathcal{M}$ is a solution of problem (2). \square

4 DCA ON HADAMARD MANIFOLDS

The aim of this section is present an extension of the DCA to Hadamard manifolds. To this end, we first propose an extension of the classical DCA, which is based on the Fenchel conjugate introduced in [Definition 2.9](#). As the DCA is dependent on the Fenchel conjugate of the first component of the objective function, which is in general difficult to compute, we provide a much simpler version of

DCA on Hadamard manifolds based on a first-order approximation of the second component. We also show the well-definition of these algorithms and their equivalence in the Riemannian setting, such as in the linear setting. The DCA based on Fenchel conjugate is stated in [Algorithm 1](#), and the second version in [Algorithm 2](#).

Algorithm 1 The DC Algorithm on Hadamard Manifolds (DCA1)

- 1: Choose an initial point $p^{(0)} \in \text{dom}(g)$. Set $k = 0$.
- 2: Take $X^{(k)} \in \partial h(p^{(k)})$, and compute

$$\begin{aligned} Y^{(k)} &\in \partial_2 g^*(p^{(k)}, X^{(k)}), \\ p^{(k+1)} &:= \exp_{p^{(k)}} Y^{(k)}. \end{aligned} \tag{7}$$

- 3: If $p^{(k+1)} = p^{(k)}$, then STOP and return $p^{(k)}$. Otherwise, go to Step 4.
 - 4: Set $k \leftarrow k + 1$ and go to Step 2.
-

As mentioned before, [Algorithm 1](#) relies on the computation of the Fenchel conjugate, which can be difficult to compute in practice. However, this algorithm is conceptually useful and can be shown to be is equivalent to more practical and computable algorithm that does not rely on the Fenchel conjugate. The following two results will be used to demonstrate the well-definedness of [Algorithm 1](#).

Lemma 4.1. *If $p \in \text{dom}(h)$ and $Y \in \partial h(p)$, then $\text{dom}(h^*(p, \cdot)) \subseteq \text{dom}(g^*(p, \cdot))$ and*

$$Y \in \text{dom}(g^*(p, \cdot)) = \{X \in T_p \mathcal{M} : g^*(p, X) < +\infty\}.$$

In particular, $\partial_2 g^(p, Y) \neq \emptyset$.*

Proof. Assume that $p \in \text{dom}(h)$ and take $Y \in \partial h(p)$. Thus, by using [Theorem 2.17](#) we obtain

$$h^*(p, Y) = -h(p) < +\infty. \tag{8}$$

From [Theorem 3.10](#) and assumption (A2) we have that

$$h^*(p, Y) - g^*(p, Y) \geq \inf_{(q, X) \in T\mathcal{M}} \{h^*(q, X) - g^*(q, X)\} = \inf_{q \in \mathcal{M}} \{g(q) - h(q)\} > -\infty. \tag{9}$$

To prove the first statement, assume by contradiction that $\text{dom}(h^*(p, \cdot)) \not\subseteq \text{dom}(g^*(p, \cdot))$. Thus, there exists $\bar{Y} \in T_p \mathcal{M}$ such that $h^*(p, \bar{Y}) < +\infty$ and $g^*(p, \bar{Y}) = +\infty$. By using (3), we have $h^*(p, \bar{Y}) - g^*(p, \bar{Y}) = h^*(p, \bar{Y}) - (+\infty) = -\infty$, which contradicts the equality in (9) and the first statement is proved. Since $\text{dom}(h^*(p, \cdot)) \subseteq \text{dom}(g^*(p, \cdot))$, it follows from (8) that $g^*(p, Y) < +\infty$. Thus, $Y \in \text{dom}(g^*(p, \cdot))$ and by assumption (A4) we conclude that $\partial_2 g^*(p, Y) \neq \emptyset$. \square

Proposition 4.2. *Algorithm 1 is well defined.*

Proof. Assume $p^{(k)} \in \text{dom}(g)$. From [Remark 3.3](#), we have that $\text{dom}(f) = \text{dom}(g) \subseteq \text{dom}(h)$, and hence $p^{(k)} \in \text{dom}(h)$. By assumption (A3), we have that $\partial h(p^{(k)}) \neq \emptyset$. Let $X^{(k)} \in \partial h(p^{(k)})$.

Since h is convex, [Theorem 2.17](#) implies that $h^*(p^{(k)}, X^{(k)}) = -h(p^{(k)}) < +\infty$. By the first part of [Lemma 4.1](#), we have that $g^*(p^{(k)}, X^{(k)}) < +\infty$ and $\partial_2 g^*(p^{(k)}, X^{(k)}) \neq \emptyset$. Let $Y^{(k)} \in \partial_2 g^*(p^{(k)}, X^{(k)})$. Since \mathcal{M} is Hadamard, the point $p^{(k+1)} = \exp_{p^{(k)}} Y^{(k)}$ is well defined and belongs to \mathcal{M} . Moreover, applying [Theorem 2.13](#) with $f = g$, $p = p^{(k)}$, $X = X^{(k)}$ and $Y = Y^{(k)}$ we have $g(p^{(k+1)}) + g^*(p^{(k)}, X^{(k)}) = \langle X^{(k)}, Y^{(k)} \rangle$ or equivalently $g(p^{(k+1)}) = \langle X^{(k)}, Y^{(k)} \rangle - g^*(p^{(k)}, X^{(k)}) < +\infty$, which implies that $p^{(k+1)} \in \text{dom}(g) = \text{dom}(f) \subseteq \text{dom}(h)$. Therefore, [Algorithm 1](#) is well defined. \square

In the following, we present a second version of the DCA that is equivalent to [Algorithm 1](#), which is described in [Algorithm 2](#).

Algorithm 2 The DC Algorithm on Hadamard Manifolds (DCA2)

- 1: Choose an initial point $p^{(0)} \in \text{dom}(g)$. Set $k = 0$.
- 2: Take $X^{(k)} \in \partial h(p^{(k)})$, and the next iterated $p^{(k+1)}$ is define as following

$$p^{(k+1)} \in \arg \min_{p \in \mathcal{M}} \left(g(p) - \langle X^{(k)}, \exp_{p^{(k)}}^{-1} p \rangle \right). \quad (10)$$

- 3: If $p^{(k+1)} = p^{(k)}$, then STOP and return $p^{(k)}$. Otherwise, go to Step 4.
 - 4: Set $k \leftarrow k + 1$ and go to Step 2.
-

It should be noted that the stopping criterion in step 3 of [Algorithm 2](#) allows it to generate an infinite sequence. Therefore, in practice, to implement [Algorithm 2](#), an appropriate stopping criterion will be required, which will be addressed further in the implementation section. Let us now analyze [Algorithm 2](#). First of all, note that due to the point $p^{(k+1)}$ be a solution of (10), we have

$$g(p) - \langle X^{(k)}, \exp_{p^{(k)}}^{-1} p \rangle \geq g(p^{(k+1)}) - \langle X^{(k)}, \exp_{p^{(k)}}^{-1} p^{(k+1)} \rangle, \quad \text{for all } p \in \mathcal{M}. \quad (11)$$

This inequality will now have an important role in the paper.

Proposition 4.3. *Algorithm 2 is well defined.*

Proof. Assume that $p^{(k)} \in \text{dom}(g)$. From [Remark 3.3](#), we have that $\text{dom}(g) = \text{dom}(f) \subseteq \text{dom}(h)$, which implies that $p^{(k)} \in \text{dom}(h)$. Thus, by Assumption (A3), we have that $\partial h(p^{(k)}) \neq \emptyset$. Let $X^{(k)} \in \partial h(p^{(k)})$. From [Lemma 2.7](#), we have that $g_k: \mathcal{M} \rightarrow \overline{\mathbb{R}}$ given by $g_k(p) := g(p) - \langle X^{(k)}, \exp_{p^{(k)}}^{-1} p \rangle$ is 1-coercive. Consequently, its minimizer set is non-empty and is contained in $\text{dom}(g)$. Therefore, there exists $p^{(k+1)} \in \text{dom}(g) = \text{dom}(f)$ such that $p^{(k+1)} \in \arg \min_{p \in \mathcal{M}} (g(p) - \langle X^{(k)}, \exp_{p^{(k)}}^{-1} p \rangle)$, which implies that [Algorithm 2](#) is well defined. \square

Remark 4.4. *If $\mathcal{M} = \mathbb{R}^n$, then by [Remark 2.14](#), we have $\partial_2 g^*(p^{(k)}, X^{(k)}) = \partial g^*(X^{(k)}) - \{p^{(k)}\}$ and consequently $Y^{(k)} + p^{(k)} = \exp_{p^{(k)}} Y^{(k)} = p^{(k+1)} \in \partial g^*(X^{(k)}) = \partial_2 g^*(p^{(k)}, X^{(k)}) + \{p^{(k)}\}$, i.e., $p^{(k+1)} \in \partial g^*(X^{(k)})$ and $X^{(k)} \in \partial h(p^{(k)})$. Therefore, [Algorithm 1](#) coincides with the classical*

formulation of the DCA; see Tao and Souad 1988; An and Tao 2005. Moreover, if $\mathcal{M} = \mathbb{R}^n$, then (11) becomes

$$g(p) - \langle X^{(k)}, p - p^{(k)} \rangle \geq g(p^{(k+1)}) - \langle X^{(k)}, p^{(k+1)} - p^{(k)} \rangle, \quad \text{for all } p \in \mathbb{R}^n,$$

which is equivalent to $p^{(k+1)} = \arg \min_{p \in \mathbb{R}^n} \{g(p) - \langle X^{(k)}, p - p^{(k)} \rangle\}$. As a conclusion, Algorithm 2 yields an alternative version of the classical DCA

In the next result, we show that Algorithm 1 is equivalent to Algorithm 2 in the Riemannian setting, similar to the linear setting.

Proposition 4.5. *If $p^{(k)} \in \text{dom}(g)$, $X^{(k)} \in \partial h(p^{(k)})$ and $Y^{(k)} \in \partial_2 g^*(p^{(k)}, X^{(k)})$, then $p^{(k+1)} = \exp_{p^{(k)}} Y^{(k)}$ if and only if $p^{(k+1)} \in \arg \min_{p \in \mathcal{M}} (g(p) - \langle X^{(k)}, \exp_{p^{(k)}}^{-1} p \rangle)$. Consequently, Algorithm 1 is equivalent to Algorithm 2.*

Proof. Let $p^{(k)} \in \text{dom}(g)$, $X^{(k)} \in \partial h(p^{(k)})$, $Y^{(k)} \in \partial_2 g^*(p^{(k)}, X^{(k)})$, and $p^{(k+1)} = \exp_{p^{(k)}} Y^{(k)}$ be given by Algorithm 1. By applying Theorem 2.13 with $f = g$, $p = p^{(k)}$, $Y = \exp_{p^{(k)}}^{-1} p^{(k+1)}$, and $X = X^{(k)}$, we have $g(p^{(k+1)}) + g^*(p^{(k)}, X^{(k)}) = \langle X^{(k)}, \exp_{p^{(k)}}^{-1} p^{(k+1)} \rangle$, which by using Definition 2.9 is equivalent to

$$g(p^{(k+1)}) - \langle X^{(k)}, \exp_{p^{(k)}}^{-1} p^{(k+1)} \rangle = -g^*(p^{(k)}, X^{(k)}) = -\sup_{q \in \mathcal{M}} (\langle X^{(k)}, \exp_{p^{(k)}}^{-1} q \rangle - g(q)),$$

or equivalently,

$$g(p^{(k+1)}) - \langle X^{(k)}, \exp_{p^{(k)}}^{-1} p^{(k+1)} \rangle = \inf_{q \in \mathcal{M}} (g(q) - \langle X^{(k)}, \exp_{p^{(k)}}^{-1} q \rangle).$$

This is also equivalent to $p^{(k+1)} \in \arg \min_{p \in \mathcal{M}} (g(p) - \langle X^{(k)}, \exp_{p^{(k)}}^{-1} p \rangle)$. Therefore, Algorithm 1 is equivalent to Algorithm 2. \square

5 CONVERGENCE ANALYSIS OF DCA

The aim of this section is to study the convergence properties of DCA. It is worth mentioning that the results in this section can be proved using either of the formulations of DCA in Algorithm 1 and 2, as they are equivalent according to Proposition 4.5. For simplicity, we present the results only using Algorithm 2, but the proofs of the results for Algorithm 1 are quite similar. We begin by showing a descent property of the algorithm.

Proposition 5.1. *Let $(p^{(k)})_{k \in \mathbb{N}}$ be generated by Algorithm 2. Then, the following inequality holds*

$$f(p^{(k+1)}) \leq f(p^{(k)}) - \frac{\sigma}{2} d^2(p^{(k)}, p^{(k+1)}). \quad (12)$$

Moreover, if $p^{(k+1)} = p^{(k)}$, then $p^{(k)}$ is a critical point of f .

Proof. By using inequality in (11) with $p = p^{(k)}$ we have $g(p^{(k)}) - g(p^{(k+1)}) \geq \langle -X^{(k)}, \exp_{p^{(k)}}^{-1} p^{(k+1)} \rangle$. On the other hand, since h is σ -strongly convex and $X^{(k)} \in \partial h(p^{(k)})$, we obtain that

$$h(p^{(k+1)}) - h(p^{(k)}) \geq \langle X^{(k)}, \exp_{p^{(k)}}^{-1} p^{(k+1)} \rangle + \frac{\sigma}{2} d^2(p^{(k+1)}, p^{(k)}).$$

Hence, using that $f = g - h$ together with two previous inequalities we obtain (12). To prove the last statement, we assume that $p^{(k+1)} = p^{(k)}$. Thus, (11) implies that $g(p) \geq g(p^{(k)}) + \langle X^{(k)}, \exp_{p^{(k)}}^{-1} p \rangle$, for all $p \in \mathcal{M}$, which shows that $X^{(k)} \in \partial g(p^{(k)})$. Hence, taking into account that $X^{(k)} \in \partial h(p^{(k)})$, we conclude that $X^{(k)} \in \partial g(p^{(k)}) \cap \partial h(p^{(k)}) \neq \emptyset$. Therefore, it follows from Definition 3.4 that $p^{(k)}$ is a critical point of f in problem (2). \square

Proposition 5.2. *Let $(p^{(k)})_{k \in \mathbb{N}}$ be generated by Algorithm 2. Then,*

$$\sum_{k=0}^{+\infty} d^2(p^{(k)}, p^{(k+1)}) < +\infty.$$

In particular, $\lim_{k \rightarrow +\infty} d(p^{(k)}, p^{(k+1)}) = 0$.

Proof. It follows from (12) that $0 \leq (\sigma/2)d^2(p^{(k)}, p^{(k+1)}) \leq f(p^{(k)}) - f(p^{(k+1)})$, for all $k \in \mathbb{N}$. Thus,

$$\sum_{k=0}^T d^2(p^{(k)}, p^{(k+1)}) \leq \frac{2}{\sigma} \sum_{k=0}^T (f(p^{(k)}) - f(p^{(k+1)})) \leq \frac{2}{\sigma} (f(p^{(0)}) - f_{\inf}),$$

for each $T \in \mathbb{N}$, where $f_{\inf} > -\infty$ is given by assumption (A2). Taking the limit in the last inequality, as T goes to $+\infty$, we obtain the first statement. The second statement is an immediate consequence of the first one. \square

Theorem 5.3. *Let $(p^{(k)})_{k \in \mathbb{N}}$ and $(X^{(k)})_{k \in \mathbb{N}}$ be generated by Algorithm 2. If \bar{p} is a cluster point of $(p^{(k)})_{k \in \mathbb{N}}$, then $\bar{p} \in \text{dom}(g)$ and there exists a cluster point \bar{X} of $(X^{(k)})_{k \in \mathbb{N}}$ such that $\bar{X} \in \partial g(\bar{p}) \cap \partial h(\bar{p})$. Consequently, every cluster point of $(p^{(k)})_{k \in \mathbb{N}}$, if any, is a critical point of f .*

Proof. Let $\bar{p} \in \mathcal{M}$ be a cluster point of $(p^{(k)})_{k \in \mathbb{N}}$. Without loss of generality we can assume that $\lim_{k \rightarrow +\infty} p^{(k)} = \bar{p}$. It follows from Proposition 5.1 together with assumption (A2) that $(f(p^{(k)}))_{k \in \mathbb{N}}$ is non-increasing and converges. Moreover, due to $f(p^{(0)}) \geq f(p^{(k)}) = g(p^{(k)}) - h(p^{(k)})$ and g be lsc, we have

$$f(p^{(0)}) \geq \liminf_{k \rightarrow +\infty} g(p^{(k)}) - \limsup_{k \rightarrow +\infty} h(p^{(k)}) \geq g(\bar{p}) - \limsup_{k \rightarrow +\infty} h(p^{(k)}).$$

Thus, using the convention (3) we conclude that $\bar{p} \in \text{dom}(g)$. Hence, using assumption (A3), we conclude that $\bar{p} \in \text{int dom}(h)$. We know that $X^{(k)} \in \partial h(p^{(k)})$, for all $k \in \mathbb{N}$. Thus, by Proposition 2.4, we can also conclude that $\lim_{k \rightarrow +\infty} X^{(k)} = \bar{X} \in \partial h(\bar{p})$. Due to the point $p^{(k+1)}$ being a solution

of (10), it satisfies (11). Thus, taking the inferior limit in (11), as k goes to $+\infty$, and using the fact that $\lim_{k \rightarrow +\infty} p^{(k)} = \bar{p}$, g is lsc together with Proposition 2.8, Item iii) and Proposition 5.2, we obtain

$$g(p) \geq \liminf_{k \rightarrow +\infty} \left(g(p^{(k+1)}) + \langle X^{(k)}, \exp_{p^{(k)}}^{-1} p \rangle - \langle X^{(k)}, \exp_{p^{(k)}}^{-1} p^{(k+1)} \rangle \right) \geq g(\bar{p}) + \langle \bar{X}, \exp_{\bar{p}}^{-1} p \rangle,$$

for each $p \in \mathcal{M}$, which implies that $g(p) \geq g(\bar{p}) + \langle \bar{X}, \exp_{\bar{p}}^{-1} p \rangle$, for all $p \in \mathcal{M}$. Hence, $\bar{X} \in \partial g(\bar{p})$. Therefore, $\bar{X} \in \partial g(\bar{p}) \cap \partial h(\bar{p})$, and hence \bar{p} is a critical point of f in problem (2). \square

Proposition 5.4. *Let $(p^{(k)})_{k \in \mathbb{N}}$ be generated by Algorithm 2. Then, for all $N \in \mathbb{N}$, there holds*

$$\min_{k=0,1,\dots,N} d(p^{(k)}, p^{(k+1)}) \leq \left(\frac{2(f(p^0) - f_{\inf})}{(N+1)\sigma} \right)^{1/2}.$$

Proof. It follows from (12) that $d^2(p^{(k)}, p^{(k+1)}) \leq (2/\sigma)(f(p^{(k)}) - f(p^{(k+1)}))$, for all $k \in \mathbb{N}$. Thus,

$$(N+1) \min_{k=0,1,\dots,N} \left(d^2(p^{(k)}, p^{(k+1)}) \right) \leq \sum_{k=0}^N \frac{2}{\sigma} \left(f(p^{(k)}) - f(p^{(k+1)}) \right) \leq \frac{2}{\sigma} \left(f(p^0) - f_{\inf} \right),$$

where $f_{\inf} > -\infty$ is given by assumption (A2). Therefore, the desired inequality directly follows. \square

The last result of this section establishes a primal-dual asymptotic convergence of the sequences generated by the DCA. This result extends the known result from the Euclidean case, cf. Tao and Souad 1988, Theorem 3, to Hadamard manifolds. Due to the nature of the problem, we will use the formulation of the DCA given in Algorithm 1.

Theorem 5.5. *Let $(p^{(k)})_{k \in \mathbb{N}}$ and $(X^{(k)})_{k \in \mathbb{N}}$ be the sequences generated by Algorithm 1. Then, the following statements hold:*

- i) $g(p^{(k+1)}) - h(p^{(k+1)}) \leq h^*(p^{(k)}, X^{(k)}) - g^*(p^{(k)}, X^{(k)}) \leq g(p^{(k)}) - h(p^{(k)})$, for all $k = 0, 1, \dots$
- ii) $\lim_{k \rightarrow +\infty} (g(p^{(k)}) - h(p^{(k)})) = \lim_{k \rightarrow +\infty} (h^*(p^{(k)}, X^{(k)}) - g^*(p^{(k)}, X^{(k)})) = \bar{f} \geq f_{\inf}$.
- iii) *If the sequence $(p^{(k)})_{k \in \mathbb{N}}$ is bounded and \bar{p} is a cluster point of $(p^{(k)})_{k \in \mathbb{N}}$, then $\bar{p} \in \text{dom}(g)$ and there exists a cluster point \bar{X} of $(X^{(k)})_{k \in \mathbb{N}}$ such that*

$$\partial g(\bar{p}) \cap \partial h(\bar{p}) \neq \emptyset, \tag{13a}$$

$$\lim_{k \rightarrow +\infty} (h(p^{(k)}) + h^*(p^{(k)}, X^{(k)})) = h(\bar{p}) + h^*(\bar{p}, \bar{X}) = 0, \tag{13b}$$

$$\lim_{k \rightarrow +\infty} (g(p^{(k)}) + g^*(p^{(k)}, X^{(k)})) = g(\bar{p}) + g^*(\bar{p}, \bar{X}) = 0. \tag{13c}$$

$$\partial_2 h^*(\bar{p}, \bar{X}) \cap \partial_2 g^*(\bar{p}, \bar{X}) \neq \emptyset, \tag{13d}$$

$$g(\bar{p}) - h(\bar{p}) = h^*(\bar{p}, \bar{X}) - g^*(\bar{p}, \bar{X}) = \bar{f}, \tag{13e}$$

Proof.

- i) By applying [Theorem 2.10](#) with $q = p^{(k+1)}$, $p = p^{(k)}$, $X = X^{(k)}$, and $f = h$, we obtain that $h(p^{(k+1)}) + h^*(p^{(k)}, X^{(k)}) \geq \langle X^{(k)}, \exp_{p^{(k)}}^{-1} p^{(k+1)} \rangle$. Since (7) implies that $\exp_{p^{(k)}}^{-1} p^{(k+1)} \in \partial_2 g^*(p^{(k)}, X^{(k)})$, we can apply [Theorem 2.13](#) with $f = g$, $p = p^{(k)}$, $Y = \exp_{p^{(k)}}^{-1} p^{(k+1)}$, and $X = X^{(k)}$ to obtain $\langle X^{(k)}, \exp_{p^{(k)}}^{-1} p^{(k+1)} \rangle = g(p^{(k+1)}) + g^*(p^{(k)}, X^{(k)})$. Hence, we have $h(p^{(k+1)}) + h^*(p^{(k)}, X^{(k)}) \geq g(p^{(k+1)}) + g^*(p^{(k)}, X^{(k)})$, which is equivalent to the first inequality of [Item i](#). To prove the second one, we first note that since $X^{(k)} \in \partial h(p^{(k)})$ and h is convex, by using [Theorem 2.17](#), we have $h^*(p^{(k)}, X^{(k)}) + h(p^{(k)}) = 0$. Thus, applying [Theorem 2.10](#) with $q = p = p^{(k)}$, $X = X^{(k)}$ and $f = g$, we have $0 \leq g^*(p^{(k)}, X^{(k)}) + g(p^{(k)})$, which combined with the last equality yields the second inequality of [Item i](#).
- ii) First we recall that $f = g - h$ satisfies assumption (A2). Thus, [Item i](#) implies that $(f(p^{(k)}))_{k \in \mathbb{N}}$ is non-increasing and convergent. Hence $\lim_{k \rightarrow +\infty} (g(p^{(k)}) - h(p^{(k)})) =: \bar{f} \in \mathbb{R}$. Moreover, by using again [Item i](#)), we also have

$$\lim_{k \rightarrow +\infty} (h^*(p^{(k)}, X^{(k)}) - g^*(p^{(k)}, X^{(k)})) =: \bar{f} \in \mathbb{R}.$$

Finally, the inequality in [Item ii](#)) follows from assumption (A2).

- iii) To prove the first part, we assume that $(p^{(k)})_{k \in \mathbb{N}}$ is bounded and \bar{p} a cluster point of $(p^{(k)})_{k \in \mathbb{N}}$. By using [Theorem 5.3](#), we conclude that $\bar{p} \in \text{dom}(g)$ and that there exists a cluster point \bar{X} of $(X^{(k)})_{k \in \mathbb{N}}$, such that $\bar{X} \in \partial g(\bar{p}) \cap \partial h(\bar{p})$. Therefore, (13a) is proved. Before proceeding with the proof we note that due to $\bar{p} \in \text{dom}(g)$, assumption (A3) implies that $\bar{p} \in \text{dom}(h)$. To prove (13b) note that since $X^{(k)} \in \partial h(p^{(k)})$, for all $k \in \mathbb{N}$, and h is convex, from [Theorem 2.17](#), we have $h(p^{(k)}) + h^*(p^{(k)}, X^{(k)}) = 0$, for all $k \in \mathbb{N}$. Consequently, $\lim_{k \rightarrow +\infty} (h(p^{(k)}) + h^*(p^{(k)}, X^{(k)})) = 0$. Since $\bar{X} \in \partial h(\bar{p})$, using again [Theorem 2.17](#), we have $h(\bar{p}) + h^*(\bar{p}, \bar{X}) = 0$ and (13b) follows directly.

To prove (13c) we first note that

$$\begin{aligned} g(p^{(k)}) + g^*(p^{(k)}, X^{(k)}) &= g(p^{(k)}) - h(p^{(k)}) - (h^*(p^{(k)}, X^{(k)}) \\ &\quad - g^*(p^{(k)}, X^{(k)})) + h(p^{(k)}) + h^*(p^{(k)}, X^{(k)}). \end{aligned}$$

Thus, using [Item ii](#)) together with (13b), we have $\lim_{k \rightarrow +\infty} (g(p^{(k)}) + g^*(p^{(k)}, X^{(k)})) = 0$. Since $\bar{X} \in \partial g(\bar{p})$, using again [Theorem 2.17](#), we have $g(\bar{p}) + g^*(\bar{p}, \bar{X}) = 0$, which combined with the last equality yields (13c).

We proceed to prove (13d). For that, we assume without loss of generality that $\lim_{k \rightarrow +\infty} p^{(k)} = \bar{p}$. Now, by applying [Theorem 2.10](#) with $f = h$, $p = \bar{p}$, $q = p^{(k)}$, we obtain

$$h(p^{(k)}) + h^*(\bar{p}, Y) \geq \langle Y, \exp_{\bar{p}}^{-1} p^{(k)} \rangle, \quad \text{for all } Y \in T_{\bar{p}}\mathcal{M}, \quad \text{and all } k \in \mathbb{N}.$$

Thus, by using [Definition 2.3](#), $\lim_{k \rightarrow +\infty} p^{(k)} = \bar{p}$, [Proposition 2.8](#), [Item i](#)) and [Item iii](#)), and that h is lsc, we have $h(\bar{p}) + h^*(\bar{p}, Y) = \liminf_{k \rightarrow +\infty} h(p^{(k)}) + h^*(\bar{p}, Y) \geq 0$. Thus, the second equality in (13b) implies that

$$h^*(\bar{p}, Y) \geq h^*(\bar{p}, X), \quad \text{for all } Y \in T_{\bar{p}}\mathcal{M}.$$

Hence, $0 \in \partial_2 h^*(\bar{p}, X)$. Similarly, by using (13c), we can also show that $0 \in \partial_2 g^*(\bar{p}, X)$. Therefore, $0 \in \partial_2 h^*(\bar{p}, X) \cap \partial_2 g^*(\bar{p}, X)$, which proves (13d).

Finally, we prove (13e). Combining the second equality in (13b) and (13c), we obtain the first equality in (13e). To prove the second inequality, we first note that $\bar{p} \in \text{dom}(g) \subset \text{int dom}(h)$.

Since h is convex, it is continuous in $\text{int dom}(h)$, which implies that $\lim_{k \rightarrow +\infty} h(p^{(k)}) = h(\bar{p})$. Thus, using [Item ii](#)), we conclude that

$$\lim_{k \rightarrow +\infty} g(p^{(k)}) = \lim_{k \rightarrow +\infty} (g(p^{(k)}) - h(p^{(k)})) + \lim_{k \rightarrow +\infty} h(p^{(k)}) = \bar{f} + h(\bar{p}).$$

Hence, using [Definition 2.3](#), we have $\lim_{k \rightarrow +\infty} g(p^{(k)}) = \liminf_{k \rightarrow +\infty} g(p^{(k)}) = g(\bar{p})$. Therefore, we obtain that $g(\bar{p}) - h(\bar{p}) = \bar{f}$, which concludes the proof. \square

6 EXAMPLES

In this section we consider examples of DC functions on the Hadamard manifold of symmetric positive definite matrices. These examples can also be seen as constraint problems on the Euclidean space of square matrices, but they are not DC problems thereon. Only by imposing the manifold structure on the constrained set, namely the symmetric positive definite matrices set, both components of the problem become convex.

Formerly, we consider the symmetric positive definite (SPD) matrices cone \mathbb{P}_{++}^n . Following [Rothaus 1960](#), see also [Nesterov and Todd 2002](#), Section 6.3, we introduce the Hadamard manifold,

$$\mathcal{M} := (\mathbb{P}_{++}^n, \langle \cdot, \cdot \rangle) \tag{14}$$

endowed with the Riemannian metric given by

$$\langle X, Y \rangle_p := \text{tr}(Xp^{-1}Yp^{-1}), \tag{15}$$

for $p \in \mathcal{M}$ and $X, Y \in T_p\mathcal{M}$, where $\text{tr}(p)$ denotes the trace of the matrix $p \in \mathbb{P}_{++}^n$, $T_p\mathcal{M} \approx \mathbb{P}^n$ is the tangent space of \mathcal{M} at p and \mathbb{P}^n denotes the set of symmetric matrices of order $n \times n$. Further details about the Hadamard manifold \mathcal{M} can be found, for example, in [Lang 1999](#), Theorem 1.2. p. 325. The *exponential map and its inverse* at a point $p \in \mathcal{M}$ are given, respectively, by

$$\exp_p X := p^{1/2} e^{p^{-1/2} X p^{-1/2}} p^{1/2}, \quad X \in T_p\mathcal{M}, p \in \mathcal{M} \tag{16}$$

$$\exp_p^{-1} q := p^{1/2} \log(p^{-1/2} q p^{-1/2}) p^{1/2}, \quad p, q \in \mathcal{M}. \tag{17}$$

The *dimension* of the manifold is given by $\dim_{\mathbb{P}_{++}^n} = \frac{n(n+1)}{2}$.

The *gradient* of a differentiable function $f: \mathbb{P}_{++}^n \rightarrow \mathbb{R}$ is given by

$$\text{grad } f(p) = p f'(p) p. \tag{18}$$

where $f'(p)$ is the Euclidean gradient of f at p . If f is twice differentiable, then the *hessian* of f is given by

$$\text{Hess } f(p)X = p f''(p) X p + \frac{1}{2} [X f'(p) p + p f'(p) X], \tag{19}$$

where $X \in T_p\mathcal{M}$ and $f''(p)$ is the Euclidean hessian of f at p .

In general, subproblem (10) in Algorithm 2 is not convex; nevertheless, in some special cases, as illustrated by the following examples, it actually is convex. To begin, recall that the gradient and hessian of a function $p \mapsto \varphi(\det(p))$, where $\varphi: \mathbb{R}_{++} \rightarrow \mathbb{R}$ is twice differentiable, is given by

$$\text{grad } \varphi(\det(p)) = (\varphi'(\det(p)) \det(p))p, \quad (20)$$

$$\text{Hess } \varphi(\det(p))v = (\varphi''(\det(p))(\det(p))^2 + \varphi'(\det(p)) \det(p))\text{tr}(p^{-1}v)p, \quad (21)$$

where $v \in T_p\mathcal{M}$, φ' and φ'' are the first and second derivative of φ , respectively.

Example 6.1. Consider the following optimization problem

$$\arg \min_{p \in \mathcal{M}} f(p), \quad \text{where } f(p) := \varphi_1(\det(p)) - \varphi_2(\det(p)), \quad (22)$$

where the function $\varphi_i: \mathbb{R}_{++} \rightarrow \mathbb{R}$ are twice differentiable satisfying $\varphi_i''(t)t^2 + \varphi_i'(t)t \geq 0$, for all $t \in \mathbb{R}_{++}$ and $i = 1, 2$. Indeed, by using (18) and (19), we can show that (22) is a DC problem with components

$$g(p) = \varphi_1(\det(p)), \quad h(p) = \varphi_2(\det(p)). \quad (23)$$

This follows from (21) and $\varphi_i''(t)t^2 + \varphi_i'(t)t \geq 0$, for all $t \in \mathbb{R}_{++}$ that $\langle \text{Hess } \varphi_i(\det(p))X, X \rangle \geq 0$, for all $X \in T_p\mathcal{M}$ and $i = 1, 2$, which implies that g and h are convex.

By using (20) we conclude that critical points of f are matrices $\bar{p} \in \mathbb{P}_{++}^n$ such that

$$\varphi_1'(\det(\bar{p})) = \varphi_2'(\det(\bar{p})). \quad (24)$$

Now, considering that h is a differentiable function, consider the subproblem associated to the problem (22)

$$\arg \min_{p \in \mathcal{M}} \psi(p), \quad \text{where } \psi(p) := \varphi_1(\det(p)) - \langle \text{grad}(\varphi_2(\det(q))), \exp_q^{-1} p \rangle \quad (25)$$

It is worth noting that if we use Algorithm 2 to solve problem (22), the subproblem (10) to be addressed has the form (25). In general, subproblem (10) is not convex; nevertheless, we will show now that (25) is a convex problem. In fact, by using second equality in (20) it follows from (25) that

$$\psi(p) = \varphi_1(\det(p)) - (\varphi_2'(\det(q)) \det(q)) \langle q, \exp_q^{-1} p \rangle. \quad (26)$$

On the other hand, by using the exponential in (17) and the metric in (15) we obtain that

$$\langle q, \exp_q^{-1} p \rangle = \text{tr}(\log(q^{-1/2} p q^{-1/2})).$$

Since $\text{tr} \log Z = \log \det Z$, for any matrix Z , the last equality becomes

$$\langle q, \exp_q^{-1} p \rangle = \log \det(p) - \log \det(q). \quad (27)$$

Combining (26) with (27), the function ψ in subproblem (25) is rewritten equivalently as

$$\psi(p) = \varphi_1(\det(p)) - (\varphi_2'(\det(q)) \det(q)) (\log \det(p) - \log \det(q)). \quad (28)$$

Since the matrix $q \in \mathcal{M}$ is fixed and the function $g(p) = \varphi_1(\det(p))$ is convex, proving that ψ is convex is sufficient to prove that the function $\Upsilon(p) = -\log \det(p)$ is convex. Applying (21) with $\varphi = \log$ we conclude that $\text{Hess } \Upsilon(p) = 0$, for all p , which implies that Υ is convex. In conclusion, the objective function f in problem (22) is not convex in general, while the function ψ in the associated subproblem (25) is. Let us conclude by presenting some functions $\varphi: \mathbb{R}_{++} \rightarrow \mathbb{R}$ satisfying the condition $\varphi''(t)t^2 + \varphi'(t)t \geq 0$, for all $t \in \mathbb{R}_{++}$:

- i) $\varphi_1(t) = a_1(\log(t))^{2(b+1)}$ and $\varphi_2(t) = a_2(\log(t))^{2b}$ with $a_1, a_2 \in \mathbb{R}_{++}$ and $b \geq 1$.
- ii) $\varphi_1(t) = \bar{a} \log(t^b + c_1) - \hat{a} \log(t)$ and $\varphi_2(t) = \log(t + c_2)$ with $\bar{a}, \hat{a}, b, c_1, c_2 \in \mathbb{R}_{++}$. Note that, if $ab > d + 1$, then $\varphi_1 - \varphi_2$ has a critical point.
- iii) $\varphi_1(t) = a_1 t^{b_1+2}$ and $\varphi_2(t) = a_2 t^{b_2+2}$ with $a_1, a_2, b_1, b_2 \in \mathbb{R}_+$.

Finally, it is worth noting that these functions g and h in (23) associated with these problems are in general not Euclidean convex functions. Consequently, (22) is not a Euclidean DC problem. As we just derived, they are DC in the Hadamard manifold (14).

Let us examine at another set of examples that are not DC Euclidean problems but are DC in the Hadamard manifold (14) described above.

Example 6.2. Consider the following optimization problem

$$\arg \min_{p \in \mathcal{M}} f(p), \quad \text{where } f(p) := \varphi_1(\text{tr}(p)) - \varphi_2(\det(p)), \quad (29)$$

where the function $\varphi_i: \mathbb{R}_{++} \rightarrow \mathbb{R}$ are twice differentiable satisfying the following conditions

$$\varphi_1'(t) \geq 0, \quad \varphi_1''(t) \geq 0, \quad \varphi_2''(t)t^2 + \varphi_2'(t)t \geq 0 \quad \forall t \in \mathbb{R}_{++}. \quad (30)$$

In general, the objective function f in the problem (29) is not convex in either the Euclidean context nor the Hadamard manifold (14). However, by using (18) and (19), we prove that the components in (29), denoted by

$$g(p) = \varphi_1(\text{tr}(p)), \quad \text{and} \quad h(p) = \varphi_2(\det(p)), \quad (31)$$

which, in general, are not convex Euclidean, are convex functions on the Hadamard manifold (14) since conditions in (30) hold. Therefore, (29) is a DC optimization problem. In addition, by using (18), we can show that the gradients of g and h are given by

$$\text{grad } g(p) = \varphi_1'(\text{tr}(p))p^2, \quad \text{grad } h(p) = (\varphi_2'(\det(p)) \det(p))p, \quad (32)$$

respectively. By using (32) we conclude that critical points of f are matrices $\bar{p} \in \mathbb{P}_{++}^n$ such that

$$\varphi_1'(\text{tr}(\bar{p}))\bar{p} = (\varphi_2'(\det(\bar{p})) \det(\bar{p}))I. \quad (33)$$

Using the same arguments as in Example 6.1, we can show that the subproblem associated with problem (29) is given by

$$\arg \min_{p \in \mathcal{M}} \psi(p), \quad \text{where } \psi(p) = \varphi_1(\text{tr}(p)) - (\varphi_2'(\det(q)) \det(q))(\log \det(p) - \log \det(q)), \quad (34)$$

for a fixed $q \in \mathcal{M}$, and the objective function ψ is convex. Finally, let us present some functions satisfying the condition (30).

- i) $\varphi_1(t) = a_1 t^{b_1}$ and $\varphi_2(t) = a_2 t^{b_2}$ with $a_1 \geq 1$ and $a_2, b_1, b_2 > 0$ such that $a_1 b_1 n^{b_1-1} = a_2 b_2$.
- ii) $\varphi_1(t) = a e^{bt}$ and $\varphi_2(t) = \frac{1}{2} a b e^{nb t^2}$, with $a, b > 0$.

7 NUMERICS

In this section, we present several numerical examples. On the one hand, we compare the algorithm to two existing algorithms and, on the other hand, illustrate in a third example how optimization problems can be reformulated into DC problems to use this structure as an advantage in numerical computations. For all numerical examples, the [Algorithm 2](#) is implemented in Julia 1.8.5 Bezanson et al. 2017 within the package `Manopt.jl` Bergmann 2022 version 0.4.12, using a trust region solver to solve the optimization problem in (10) within every step, including a generic implementation of the corresponding cost and gradient. This way, the algorithm is easy-to-use, while when a more efficient computation for either cost and gradient of the sub problem or even a closed form solution is available, they can benefit to speed up the computation, when provided. Together with `Manifolds.jl` Axen et al. 2021 this algorithm can be used on arbitrary manifolds. All times refer to running the experiments on an Apple MacBook Pro M1 (2021), 16 GB Ram, Mac OS Ventura 13.0.1.

7.1 A COMPARISON TO THE DIFFERENCE OF CONVEX PROXIMAL POINT ALGORITHM

We first consider the problem

$$\arg \min_{p \in \mathcal{M}} (\log(\det(p)))^4 - (\log \det(p))^2.$$

on \mathcal{P}_{++}^n . Here we have $f(p) = g(p) - h(p)$ where $g(p) = \varphi_1(\det(p))$, $\varphi_1(t) = (\log t)^4$, and $h(p) = \varphi_2(\det(p))$, $\varphi_2(t) = (\log t)^2$, which fits [Example 6.1, Item i](#)). The critical points of this problem are the matrices $p^* \in \mathcal{P}_{++}^n$ such that $\det(p^*) = e^{1/\sqrt{2}}$. We have $f(p^*) = -\frac{1}{4}$, for each critical point p^* .

We compare the DCA with the Difference of Convex Proximal Point Algorithm (DCPPA) as introduced in Souza and Oliveira 2015. The algorithm is also available in `Manopt.jl`, implemented in the same generic manner, as the DCA explained above. This means that the proximal map can be considered as a subproblem to solve. When only g and its gradient $\text{grad } g$ are provided, the subproblem is generated in a generic manner, that is, a default implementation of the minimization problem that corresponds to step 3 of the DCPPA-Algorithm from Souza and Oliveira 2015 is generated. This is also the scenario we use for our example. For the case that $\text{prox}_{\lambda g}(p)$ is available, e.g. in closed form, it can be provided to the algorithm for speed-up.

For both DCA and DCPPA, the generation of the generic subproblem is the default in `Manopt.jl` as soon as the gradient $\text{grad } g$ of g is provided. The function calls look like

```
difference_of_convex_algorithm(M, f, g, grad_h, p0; grad_g=grad_g)
difference_of_convex_proximal_point(M, grad_h, p0; g=g, grad_g=grad_g)
```

By default, further an approximation the Hessian of both sub problems by a Riemannian variant of forward difference from the gradient is used. This enables the use of the `trust_regions`¹ algorithm to solve the sub-problem. To make both algorithms comparable we

¹see manoptjl.org/stable/solvers/trust_regions/ for details

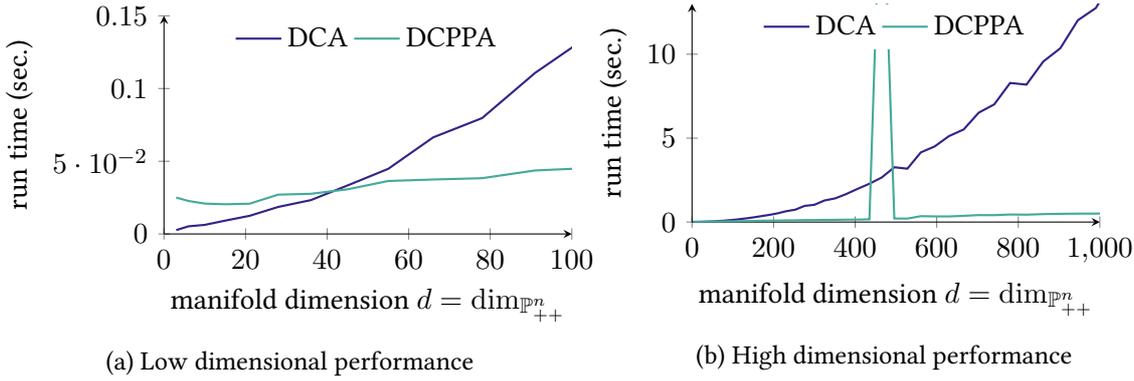


Figure 1: A comparison of the run times of DCA and DCPA for different manifold dimensions.

- for both sub solvers we *stop* when the gradient norm (of the subproblem’s gradient) is below 10^{-10} or after 5000 iterations if the gradient does not get small.
- for both algorithms DCA and DCPA when the gradient norm (of f) is below 10^{-10} . We also have a fall back to stop after 100 iterations if the gradient norm is not hit.
- the proximal parameter in the DCPA to a constant of $\lambda = \frac{1}{2n}$
- for both algorithms we set $p^{(0)} = \log(n)I_n$ as the initial point, where I_n denotes the $n \times n$ identity matrix

For the matrix size n of \mathbb{P}_{++}^n we set $n = 2, 3, \dots, 80$ to compare the algorithm for different manifold sizes, which yields manifolds of dimension $d = \frac{n(n+1)}{2}$.

In [Figure 1](#) we compare the different run times for both the DCA and DCPA. These were obtained using the `@benchmark` macro from `BenchmarkTools.jl` [Chen and Revels 2016](#),

Up to a dimension of approximately $d = 40$ (or 8×8 spd. matrices) the DCA is faster. This includes the important case of 3×3 spd. matrices, that is one representation of diffusion tensors, where the DCA takes only $5.2434 \cdot 10^{-3}$ seconds while the DCPA takes $2.2672 \cdot 10^{-2}$ seconds. For higher-dimensional problems, cf. [Figure 1b](#), the DCPA seems to only increase very slowly, where $d = 465$, or 30×30 spd. matrices, seems to be an outlier, where DCPA takes over 22 seconds, while otherwise it stays around about half a second, even for the last case shown, i.e. $d = 99$ (or 44×44 spd. matrices).

Comparing the number of iterations, we observe that after the first 5 experiments, so starting from a dimension of 21 (6×6 spd. matrices), the number of iterations stabilizes around 25 iterations for the DCA and 38 for the DCPA.

We compare different developments of the cost function in [Figure 2](#). Since for all dimensions we know that $f(p^*) = -\frac{1}{4}$ we plot $|f(p^{(k)}) - f(p^*)|$ over the iterations for the manifold dimensions $d = 15, 55, 210, 820, 3240$, that is the $n \times n$ matrices for $n = 5, 10, 20, 40, 80$. The initial value $p^{(0)}$ was chosen as above, which yields that the value $f(p^{(1)})$ is always below 10^3 in our experiments. All these different dimensions show the same slope in the decrease of the cost function $f(p)$ for both the DCA as well as the DCPA. While for DCPA the cost seems to be below 10^{-16} close to the minimum for a few iterations already, before the stopping criterion of a gradient norm $\|\text{grad } f(p^{(k)})\|_{p^{(k)}} < 10^{-10}$

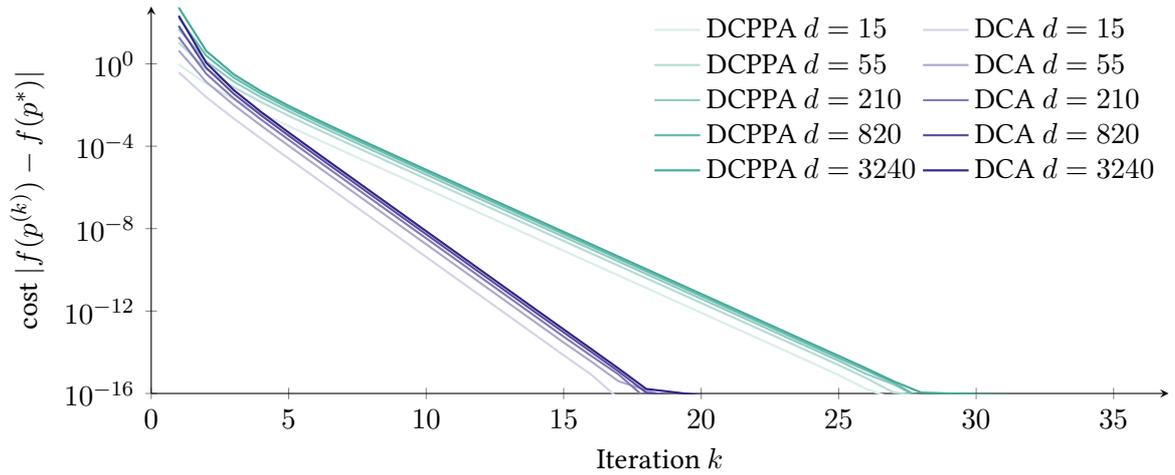


Figure 2: A comparison of how close the cost function is to the actual minimum for different sizes of problems and both algorithms.

is reached.

The development of the cost function illustrates, that DCA converges faster than DCPPA, such that the choice of the sub solver seems to be crucial for the run time, which for these experiments we configures equally to compare the algorithms and not sub solvers.

7.2 THE ROSENBRACK PROBLEM

The Rosenbrock problem consists of

$$\arg \min_{x \in \mathbb{R}^2} a(x_1^2 - x_2)^2 + (x_1 - b)^2, \tag{35}$$

where $a, b > 0$ are positive numbers, classically $b = 1$ and $a \gg b$, see Rosenbrock 1960. Note that the function is non-convex on \mathbb{R}^2 . The minimizer x^* is given by $x^* = (b, b^2)^T$, and also the (Euclidean) Gradient can be directly stated as

$$\nabla f(x) = \begin{pmatrix} 4a(x_1^2 - x_2) \\ -2a(x_1^2 - x_2) \end{pmatrix} + \begin{pmatrix} 2(x_1 - b) \\ 0 \end{pmatrix} \tag{36}$$

We introduce a new metric for $\mathcal{M} = \mathbb{R}^2$: For any $p \in \mathbb{R}^2$ we define

$$G_p := \begin{pmatrix} 1 + 4p_1^2 & -2p_1 \\ -2p_1 & 1 \end{pmatrix}, \text{ which has the inverse matrix } G_p^{-1} = \begin{pmatrix} 1 & 2p_1 \\ 2p_1 & 1 + 4p_1^2 \end{pmatrix}.$$

We define the inner product on $T_p\mathcal{M} = \mathbb{R}^2$ as

$$\langle X, Y \rangle_p = X^T G_p Y$$

In the following we refer to \mathbb{R}^2 with the default Euclidean metric further as just \mathbb{R}^2 and to the same space with this new metric as \mathcal{M} .

The exponential and logarithmic map are given as

$$\exp_p(X) = \begin{pmatrix} p_1 + X_1 \\ p_2 + X_2 + X_1^2 \end{pmatrix}, \quad \exp_p^{-1}(q) = \begin{pmatrix} q_1 - p_1 \\ q_2 - p_2 - (q_1 - p_1)^2 \end{pmatrix}.$$

Given some function $h: \mathcal{M} \rightarrow \mathbb{R}$, its Riemannian gradient $\text{grad } h: \mathcal{M} \rightarrow T\mathcal{M}$ can be computed from the Euclidean one by

$$\text{grad } h(p) = G_p^{-1} \nabla h(p).$$

Denoting the two components of the Euclidean gradient by $\nabla h(p) = (h'_1(p), h'_2(p))^T$ we can derive that given two points $p, q \in \mathcal{M}$ we have

$$\begin{aligned} \left\langle \text{grad } h(q), \exp_q^{-1}(p) \right\rangle_q &= (\exp_q^{-1}(p))^T \nabla h(q) \\ &= (p_1 - q_1)h'_1(q) + (p_2 - q_2 - (p_1 - q_1)^2)h'_2(q) \end{aligned} \quad (37)$$

For the difference of convex algorithm we split the cost function from the Rosenbrock problem (35) as $f(x) = g(x) - h(x)$ with

$$g(x) = a(x_1^2 - x_2)^2 + 2(x_1 - b)^2 \quad \text{and} \quad h(x) = (x_1 - b)^2.$$

Using the isometry $\psi: \mathbb{R}^2 \rightarrow \mathcal{M}, \mathbf{z} \mapsto (z_1, z_1^2 - z_2)$ we get

$$(h \circ \psi)(x) = h(x_1, x_1^2 - x_2) = (x_1 - b)^2$$

and hence h is (geodesically) convex on \mathcal{M} .

The corresponding Euclidean gradients of g and h are

$$\nabla g(p) = \begin{pmatrix} 4a(x_1^2 - x_2) \\ -2a(x_1^2 - x_2) \end{pmatrix} + \begin{pmatrix} 4(x_1 - b) \\ 0 \end{pmatrix} \quad \text{and} \quad \nabla h(p) = \begin{pmatrix} 2(p_1 - b) \\ 0 \end{pmatrix},$$

So especially the second component $h'_2(p) = 0$.

Considering the sub-problem Equation (10) from Algorithm 2, we obtain together with (37) for some fixed q that

$$\begin{aligned} \varphi(p) &= g(p) - \left\langle \text{grad } h(q), \exp_q^{-1}(p) \right\rangle_q \\ &= a(p_1^2 - p_2)^2 + 2(p_1 - b)^2 - 2(q_1 - b)(p_1 - q_1) \\ &= a(p_1^2 - p_2)^2 + 2(p_1 - b)^2 - 2(q_1 - b)p_1 + 2(q_1 - b)q_1, \end{aligned}$$

where the last term is constant with respect to p and hence irrelevant when determining a minimizer. The Euclidean Gradient reads

$$\nabla \varphi(p) = \begin{pmatrix} 4ap_1(p_1^2 - p_2) + 4(p_1 - b) - 2(q_1 - b) \\ -2a(p_1^2 - p_2) \end{pmatrix}$$

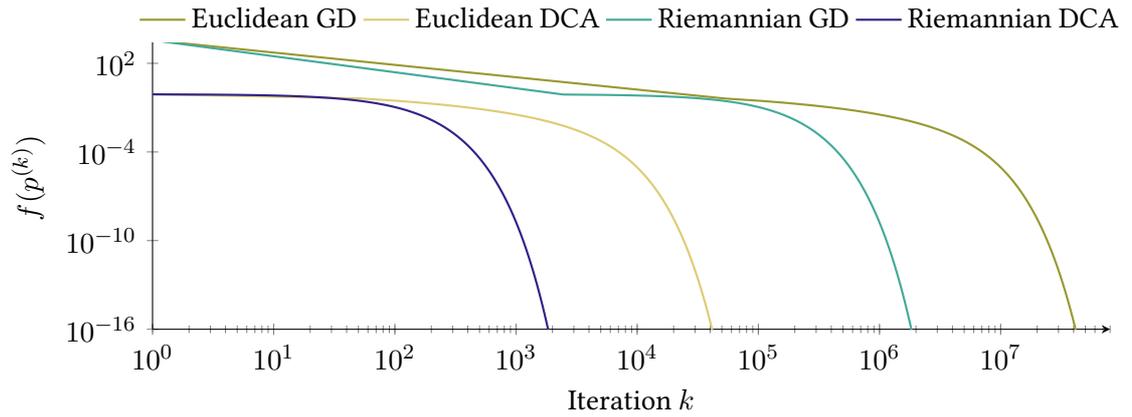


Figure 3: A comparison of the cost function during the iterations of the four experiments performed on the Rosenbrock example.

and the Riemannian gradient is similar to before $\text{grad } \varphi(p) = G_p^{-1} \nabla \varphi(p)$. This allows to use for example the Riemannian gradient descent from `Manopt.jl` to be used as a sub-solver for Equation (10) within Algorithm 2.

Since also for the Rosenbrock function the Riemannian gradient can be easily computed in the same manner from (36), we can now compare three different first order methods:

- i) The Euclidean gradient descent algorithm on \mathbb{R}^2 ,
- ii) The Euclidean Difference of Convex Algorithm on \mathbb{R}^2
- iii) The Riemannian gradient descent algorithm on \mathcal{M} ,
- iv) The Riemannian Difference of Convex Algorithm on \mathcal{M} , using Riemannian gradient descent as a sub-solver

All algorithms use `ArmiJoLinesearch(M)` when determining the step size in gradient descent, and all stop either after 10 million steps, or when the change between two successive iterates drops below $1e-16$. The sub solver in the DCA is set to stop when the gradient is at $1e-16$ in norm or at 1000 iterations.

We set $b = 1$ and $a = 2 \cdot 10^5$. All algorithms start in $p^{(0)} = \frac{1}{10}(1, 2)^T$. The initial cost is $f(p^{(0)}) \approx 7220.81$. The runtime and number of iterations is depicted in Table 1 and the development of the cost function during the iterations in Figure 3.

For the cost $f(p^{(k)})$ during the iterations, we can observe that both gradient algorithms as well as both difference of convex algorithms perform similar in shape, both groups even have similar gain in their first step.

Still, even for the Euclidean case, the gradient descent with Armijo step size requires several orders of magnitude more iterations than the Euclidean difference of convex algorithm. The Riemannian gradient descent outperforms the Euclidean one both in number of iterations as well as overall runtime. Since a single iteration in the difference of convex algorithm requires to solve a sub optimization problem, and we even employ a gradient descent per iteration, even the Euclidean DCA is slower than

Algorithm	Runtime	# Iterations
Euclidean GD	305.567 sec.	53 073 227
Euclidean DCA	58.268 sec.	50 588
Riemannian GD	18.894 sec.	2 454 017
Riemannian DCA	7.704 sec.	2 459

Table 1: Summary of the runtime and number of iterations of the four experiments performed on the Rosenbrock example.

the Riemannian gradient descent, while the DCA already requires about a factor of 50 less iterations. Similarly, the Riemannian DCA requires a factor of 1 000 less iterations than the Riemannian gradient descent, but since a single iteration is more costly, it is only about a factor of 2 faster.

7.3 CONSTRAINED MAXIMIZATION OF THE FRÉCHET VARIANCE

Let \mathcal{M} be the manifold (14) of symmetric positive definite matrices \mathbb{P}_{++}^n , $n \in \mathbb{N}$ with the affine invariant metric $\langle \cdot, \cdot \rangle$, $\{q_1, \dots, q_m\} \subset \mathcal{M}$ be a data set of distinct points, i. e. $q_i \neq q_j$ for $i \neq j$, and μ_1, \dots, μ_m be non-negative weights with $\sum_{j=1}^m \mu_j = 1$. Let $h: \mathcal{M} \rightarrow \mathbb{R}$ be the function defined by

$$h(p) := \sum_{j=1}^m \mu_j d^2(p, q_j), \quad \text{where } d^2(p, q_i) := \text{tr}(\log^2(p^{-1/2} q_i p^{-1/2})). \quad (38)$$

Recall that

$$d(p, q) := \|\log(p^{-1/2} q p^{-1/2})\|_F = \sqrt{\text{tr} \log^2(p^{-1/2} q p^{-1/2})}$$

is the Riemannian distance between p and q on \mathcal{M} . When every one of the weights μ_1, \dots, μ_m are equal, this function h is known as the *Fréchet variance* of the set $\{q_1, \dots, q_m\}$, see Horev, Yger, and Sugiyama 2017. In this example we want to consider the constrained Fréchet variance maximization problem, which is stated as

$$\arg \max_{p \in \mathcal{C}} h(p) \quad (39)$$

where the constrained convex set is given by

$$\mathcal{C} := \{p \in \mathcal{M} \mid \bar{q}_- \preceq p \preceq \bar{q}_+\}, \quad (40)$$

where $\bar{q}_-, \bar{q}_+ \in \mathcal{M}$ with $\bar{q}_- \prec \bar{q}_+$. Here, $p \prec q$ ($p \preceq q$) denotes the (non-strict) partial ordering on the spd-matrices, i. e. that $q - p$ is positive (semi-)definite or for short $q - p \prec 0$ ($\preceq 0$). We point out that Lim 2012, Lemma 2 (iii) implies that the set \mathcal{C} is convex.

The problem (39) can be equivalently stated as a Difference of Convex problem or a non-convex minimization problem. The second formulation can be algorithmically solved by a Frank-Wolfe algorithm Weber and Sra 2022.

Maximizing the Fréchet variance as a DC problem. We define the *indicator function* of the set \mathcal{C} as

$$\iota_{\mathcal{C}}(p) = \begin{cases} 0 & \text{if } p \in \mathcal{C} \\ \infty & \text{else.} \end{cases}$$

Using $g = \iota_{\mathcal{C}}$ and the fact that

$$\arg \max_{p \in \mathcal{C}} h(p) = \arg \min_{p \in \mathcal{C}} -h(p)$$

to rephrase problem (39) to

$$-\arg \min_{p \in \mathcal{M}} f(p), \quad \text{where } f(p) := g(p) - h(p). \quad (41)$$

We obtain indeed for $p \in \mathcal{C}$ that $f(p) = -h(p)$ and hence at a minimizer of f we obtain a maximizer of h . This hence yields a DC problem as studied in the previous sections.

By using (18) and (38), the gradient $\text{grad } h(p)$ is given by

$$\text{grad } h(p) = -2 \sum_{j=1}^m \mu_j p^{1/2} \log(p^{-1/2} q_j p^{-1/2}) p^{1/2} \quad (42)$$

$$= 2 \sum_{j=1}^m \mu_j p^{1/2} \log(p^{1/2} q_j^{-1} p^{1/2}) p^{1/2}. \quad (43)$$

In this case, due to $g(p) = 0$, for $p \in \mathcal{C}$, the subproblem (10) for $X^{(k)} = \text{grad } h(p^{(k)})$ is given by

$$p^{(k+1)} \in \arg \min_{p \in \mathcal{C}} \langle -\text{grad } h(p^{(k)}), \exp_{p^{(k)}}^{-1} p \rangle. \quad (44)$$

On the other hand, it follows from (15) and (17) that

$$\langle -\text{grad } h(p^{(k)}), \exp_{p^{(k)}}^{-1} p \rangle \quad (45)$$

$$= \langle -\text{grad } h(p^{(k)}), (p^{(k)})^{1/2} \log((p^{(k)})^{-1/2} p (p^{(k)})^{-1/2}) (p^{(k)})^{1/2} \rangle \quad (46)$$

$$= \text{tr} \left(-(p^{(k)})^{-1/2} \text{grad } h(p^{(k)}) (p^{(k)})^{-1/2} \log((p^{(k)})^{-1/2} p (p^{(k)})^{-1/2}) \right). \quad (47)$$

Therefore, the problem (44) becomes

$$p^{(k+1)} \in \arg \min_{p \in \mathcal{C}} \text{tr} (s^{(k)} \log((p^{(k)})^{-1/2} p (p^{(k)})^{-1/2})), \quad (48)$$

where, by using (42), the matrix $s^{(k)}$ is given by

$$s^{(k)} = -(p^{(k)})^{-1/2} \text{grad } h(p^{(k)}) (p^{(k)})^{-1/2} = 2 \sum_{j=1}^m \mu_j \log((p^{(k)})^{-1/2} q_j (p^{(k)})^{-1/2}). \quad (49)$$

or, by using (43), the matrix $s^{(k)}$ is given equivalently by

$$s^{(k)} := -(p^{(k)})^{-1/2} \text{grad } h(p^{(k)}) (p^{(k)})^{-1/2} = -2 \sum_{j=1}^m \mu_j \log((p^{(k)})^{1/2} q_j^{-1} (p^{(k)})^{1/2}). \quad (50)$$

In order to deal with the subproblem (48) we consider the following theorem, which gives a closed formula for it, see Weber and Sra 2022, Theorem 4.

Theorem 7.1. Let $L, U \in \mathbb{P}_{++}^n$ such that $L \prec U$. Let S be a Hermitian $(n \times n)$ matrix and $X \in \mathbb{P}_{++}^n$ be arbitrary. Then, the solution to the optimization problem

$$\min_{L \preceq Z \preceq U} \text{tr}(S \log(XZX)),$$

is given by $Z = X^{-1}Q \left(P^\top [-\text{sgn}(D)]_+ P + \hat{L} \right) Q^\top X^{-1}$, where $S = QDQ^\top$ is a diagonalization of S , $\hat{U} - \hat{L} = P^\top P$ with $\hat{L} = Q^\top X L X Q$ and $\hat{U} = Q^\top X U X Q$, where $[-\text{sgn}(D)]_+$ is the diagonal matrix

$$\text{diag}([- \text{sgn}(d_{11})]_+, \dots, [- \text{sgn}(d_{nn})]_+)$$

and $D = (d_{ij})$.

Remark 7.2. The solution to (48) can be obtained [Theorem 7.1](#) setting $L = \bar{q}_-$, $U = \bar{q}_+$, $S = s^{(k)}$, $X = (p^{(k)})^{-\frac{1}{2}}$ and $Z = p$. Note that given $p^{(k)}$ both X and X^{-1} can be easily computed using the eigen decomposition and modifying the diagonal matrix.

To minimize a constrained, non-convex function $f_{\text{FW}}: \mathcal{X} \rightarrow \mathbb{R}$, $\mathcal{X} \subset \mathcal{M}$, [Weber and Sra 2022](#) propose the Riemannian Frank-Wolfe algorithm as summarized in [Algorithm 3](#).

Algorithm 3 The Riemannian Frank-Wolfe Algorithm, cf. [Weber and Sra 2022](#), [Algorithm 2](#).

- 1: Choose an initial point $p^{(0)} \in \mathcal{X}$. Set $k = 0$.
 - 2: **while** convergence criterion is not met **do**
 - 3: $q^{(k)} \leftarrow \underset{q \in \mathcal{X}}{\text{argmin}} \langle \text{grad } f_{\text{FW}}(p^{(k)}), \exp_{p^{(k)}}^{-1} q \rangle$
 - 4: $s_k \leftarrow \frac{2}{2+k}$
 - 5: $p^{(k+1)} = \gamma_{p^{(k)}q^{(k)}}(s_k)$
 - 6: $k \leftarrow k + 1$
 - 7: **end while**
-

In our example we have $\mathcal{X} = \mathcal{C}$ and $f_{\text{FW}} = -h$, i. e. we obtain a concave constrained problem. Since $f_{\text{FW}} = -h$, we obtain the same subproblem in Step 3 of the Frank-Wolfe Algorithm as stated in (44). Thus, we have two algorithms for solving the problem (39) or equivalently (41), namely [Algorithm 3](#) and [Algorithm 2](#). Both possess the same subproblem in this case. They treat the result of the subproblem differently, though. While [Algorithm 2](#) uses the subproblem solution directly for the next iteration, [Algorithm 3](#) uses the solution as an end point of a geodesic segment starting from the previous iterate. This geodesic segment is then evaluated at a certain interims point. This also means that [Algorithm 3](#) has to start in a *feasible point* $p^{(0)} \in \mathcal{C}$, while for [Algorithm 2](#) this is not necessary. In our numerical example we consider $\mathcal{M} = \mathbb{P}_{++}^{20}$, that is, the set of 20×20 symmetric positive definite matrices with the affine-invariant metric $\langle \cdot, \cdot \rangle$. This is a Riemannian manifold of dimension $d = 210$. We further generate $m = 100$ random spd matrices q_i with corresponding random weights w_i as the data set for the Fréchet variance. We set

$$q_- := \left(\sum_{i=1}^m w_i q_i^{-1} \right)^{-1}, \quad q_+ := \sum_{i=1}^m w_i q_i, \quad \text{and} \quad p^{(0)} := \frac{1}{2}(q_- + q_+).$$

A numerical implementation of [Theorem 7.1](#) is used as a closed-form solver of the subproblem. Numerically we observe, that these results might suffer from imprecisions, which means they might not

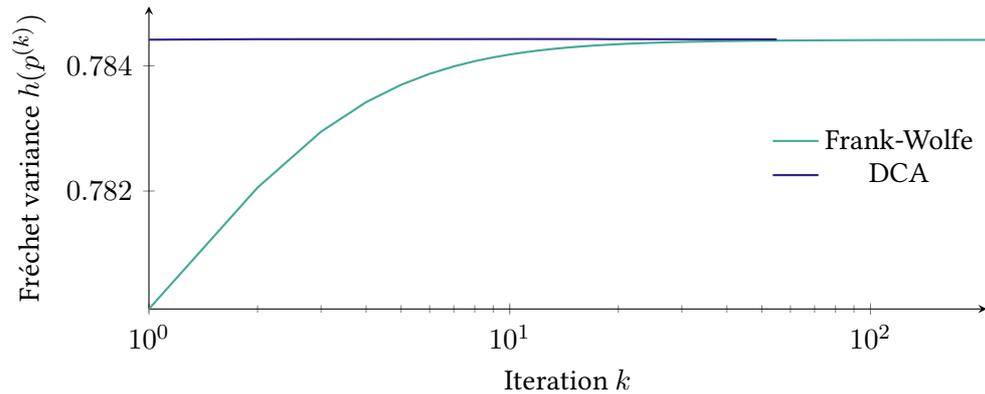


Figure 4: A comparison of the Fréchet variance $h(p)$ during the iterations of Algorithm 2 (indigo) and Algorithm 3 (teal).

meet the constraint, but only by around $2 \cdot 10^{-13}$. Since we use these points as iterates in Algorithm 2, only for this algorithm we add a “safeguard” and perform a small line search for the first matrix closest matrix to the sub-solver’s result q^* on the geodesic to the last iterate $p^{(k)}$ fulfilling the constraint.

Then we stop Algorithm 2 if the change $d(p^{(k)}, p^{(k+1)}) < 10^{-14}$ or if the gradient change between these two iterates (computes using parallel transport) is below 10^{-9} . The algorithm stops after 55 iterations due to a small gradient change.

For Frank-Wolfe a suitable stopping criterion is challenging. Note that the gradient $\text{grad } f_{\text{fw}}$ does not tend to 0 of the minimizer is on the boundary. Even after 100 000 iterations, Frank-Wolfe still has not reached either of the stopping criteria, both changes are still of order 10^{-4} . While the Difference of Convex Algorithm reaches it’s minimum (its maximum in h) after 11 iterations and then increases slightly, probably due to the closed-form solution not being precise, Frank-Wolfe reaches neither of these two values – after 11 or 55 in these 100 000 iterations.

Finally, comparing the time per iteration, both algorithms comparable. With the numerical safeguard for this specific problem, 1000 iterations of DCA take 16.01 seconds, Frank-Wolfe 8.13 seconds and DCA without the safeguard 7.475 seconds. That is, in runtime per iteration, using the same sub solver, both perform similarly, while DCA seems to have a vanishing gradient change.

8 CONCLUSION

In this paper, we investigated the extension of the Difference of Convex Algorithm (DCA) to the Riemannian case, enabling us to solve DC problems on Riemannian manifolds. We investigate its relation to Duality on manifolds and state a convergence result on Hadamard manifolds.

Numerically, the new algorithm outperforms the existing Difference of Convex Proximal Point algorithm (DCPPA) in terms of the number of iterations. However, for large-dimensional manifolds, the DCPPA is faster. Additionally, for a specific class of constrained maximization problems, the DCA is well-suited and outperforms the Riemannian Frank-Wolfe algorithm, especially because a suitable stopping criterion can be used but also in how close it gets to the actual minimizer. Finally, rephrasing

Euclidean problems into DC problems with a suitable metric is another field where using the DCA seems very beneficial.

Extending the numerical algorithms also to employing duality is a future research topic, where the iteration time and convergence speed might increase.

REFERENCES

- Absil, P.-A., C. G. Baker, and K. A. Gallivan (2007). “Trust-region methods on Riemannian manifolds”. In: *Foundations of Computational Mathematics. The Journal of the Society for the Foundations of Computational Mathematics* 7.3, pp. 303–330. DOI: [10.1007/s10208-005-0179-9](https://doi.org/10.1007/s10208-005-0179-9).
- Absil, P.-A., R. Mahony, and R. Sepulchre (2008). *Optimization Algorithms on Matrix Manifolds*. Princeton University Press. DOI: [10.1515/9781400830244](https://doi.org/10.1515/9781400830244).
- Almeida, Y. T., J. X. da Cruz Neto, P. R. Oliveira, and J. C. de Oliveira Souza (2020). “A modified proximal point method for DC functions on Hadamard manifolds”. In: *Computational Optimization and Applications* 76.3, pp. 649–673. DOI: [10.1007/s10589-020-00173-3](https://doi.org/10.1007/s10589-020-00173-3).
- An, L. T. H. and P. D. Tao (2005). “The DC (Difference of Convex Functions) Programming and DCA Revisited with DC Models of Real World Nonconvex Optimization Problems”. In: *Annals of Operations Research* 133.1-4, pp. 23–46. DOI: [10.1007/s10479-004-5022-1](https://doi.org/10.1007/s10479-004-5022-1).
- Axen, S. D., M. Baran, R. Bergmann, and K. Rzecki (2021). *Manifolds.jl: An Extensible Julia Framework for Data Analysis on Manifolds*. arXiv: [2106.08777](https://arxiv.org/abs/2106.08777).
- Bačák, M., R. Bergmann, G. Steidl, and A. Weinmann (2016). “A second order non-smooth variational model for restoring manifold-valued images”. In: *SIAM Journal on Scientific Computing* 38.1, A567–A597. DOI: [10.1137/15M101988X](https://doi.org/10.1137/15M101988X). arXiv: [1506.02409](https://arxiv.org/abs/1506.02409).
- Bartle, R. G. and D. R. Sherbert (2000). *Introduction to real analysis*. 3rd ed. Vol. 2. Wiley New York.
- Bergmann, R. (2022). “Manopt.jl: Optimization on Manifolds in Julia”. In: *Journal of Open Source Software* 7.70, p. 3866. DOI: [10.21105/joss.03866](https://doi.org/10.21105/joss.03866).
- Bergmann, R., R. Herzog, M. Silva Louzeiro, D. Tenbrinck, and J. Vidal-Núñez (2021). “Fenchel duality theory and a primal-dual algorithm on Riemannian manifolds”. In: *Foundations of Computational Mathematics* 21.6, pp. 1465–1504. DOI: [10.1007/s10208-020-09486-5](https://doi.org/10.1007/s10208-020-09486-5). arXiv: [1908.02022](https://arxiv.org/abs/1908.02022).
- Bergmann, R., J. Persch, and G. Steidl (2016). “A parallel Douglas Rachford algorithm for minimizing ROF-like functionals on images with values in symmetric Hadamard manifolds”. In: *SIAM Journal on Imaging Sciences* 9.4, pp. 901–937. DOI: [10.1137/15M1052858](https://doi.org/10.1137/15M1052858). arXiv: [1512.02814](https://arxiv.org/abs/1512.02814).
- Bergmann, R. and A. Weinmann (2016). “A second-order TV-type approach for inpainting and denoising higher dimensional combined cyclic and vector space data”. In: *Journal of Mathematical Imaging and Vision* 55.3, pp. 401–427. DOI: [10.1007/s10851-015-0627-3](https://doi.org/10.1007/s10851-015-0627-3). arXiv: [1501.02684](https://arxiv.org/abs/1501.02684).
- Bezanson, J., A. Edelman, S. Karpinski, and V. Shah (2017). “Julia: A Fresh Approach to Numerical Computing”. In: *SIAM Review* 59.1, pp. 65–98. DOI: [10.1137/141000671](https://doi.org/10.1137/141000671). URL: <https://epubs.siam.org/doi/10.1137/141000671> (visited on 01/18/2019).
- Bhattacharya, A. and R. Bhattacharya (2008). “Statistics on Riemannian manifolds: asymptotic distribution and curvature”. In: *Proceedings of the American Mathematical Society* 136.08, pp. 2959–2967. DOI: [10.1090/s0002-9939-08-09445-8](https://doi.org/10.1090/s0002-9939-08-09445-8).
- Bourbaki, N. (1995). *General Topology*. Springer Berlin Heidelberg. DOI: [10.1007/978-3-642-61701-0](https://doi.org/10.1007/978-3-642-61701-0).

- Bredies, K., M. Holler, M. Storath, and A. Weinmann (2018). “Total generalized variation for manifold-valued data”. In: *SIAM Journal on Imaging Sciences* 11.3, pp. 1785–1848. DOI: [10.1137/17M1147597](https://doi.org/10.1137/17M1147597). arXiv: [1709.01616](https://arxiv.org/abs/1709.01616).
- Chen, J. and J. Revels (2016). “Robust benchmarking in noisy environments”. In: *arXiv e-prints*, arXiv:1608.04295. arXiv: [1608.04295](https://arxiv.org/abs/1608.04295) [cs.PF].
- De Oliveira, W. (2020). “The ABC of DC Programming”. In: *Set-Valued and Variational Analysis* 28.4, pp. 679–706. DOI: [10.1007/s11228-020-00566-w](https://doi.org/10.1007/s11228-020-00566-w).
- Do Carmo, M. P. (1992). *Riemannian geometry*. Mathematics: Theory & Applications. Birkhäuser Boston, Inc., Boston, MA, pp. xiv+300. DOI: [10.1007/978-1-4757-2201-7](https://doi.org/10.1007/978-1-4757-2201-7).
- Edelman, A., T. A. Arias, and S. T. Smith (1999). “The geometry of algorithms with orthogonality constraints”. In: *SIAM Journal on Matrix Analysis and Applications* 20.2, pp. 303–353. DOI: [10.1137/S0895479895290](https://doi.org/10.1137/S0895479895290).
- Esposito, M., C. Hennersperger, R. Gobl, L. Demaret, M. Storath, N. Navab, M. Baust, and A. Weinmann (2019). “Total Variation Regularization of Pose Signals with an Application to 3D Freehand Ultrasound”. In: *J. Math. Imaging Vision*. DOI: [10.1109/TMI.2019.2898480](https://doi.org/10.1109/TMI.2019.2898480).
- Fletcher, P. T. (2013). “Geodesic regression and the theory of least squares on Riemannian manifolds”. In: *International Journal of Computer Vision* 105.2, pp. 171–185. DOI: [10.1007/s11263-012-0591-y](https://doi.org/10.1007/s11263-012-0591-y).
- Freifeld, O. and M. J. Black (2012). *Lie Bodies: A Manifold Representation of 3D Human Shape*. in Proceedings of ECCV 2012, Springer Berlin Heidelberg, pp. 1–14. DOI: [10.1007/978-3-642-33718-5_1](https://doi.org/10.1007/978-3-642-33718-5_1).
- Horev, I., F. Yger, and M. Sugiyama (2017). “Geometry-aware principal component analysis for symmetric positive definite matrices”. In: *Mach. Learn.* 106.4, pp. 493–522. DOI: [10.1007/s10994-016-5605-5](https://doi.org/10.1007/s10994-016-5605-5). URL: <https://doi.org/10.1007/s10994-016-5605-5>.
- Huang, W., K. A. Gallivan, and P.-A. Absil (2015). “A Broyden class of quasi-Newton methods for Riemannian optimization”. In: *SIAM Journal on Optimization* 25.3, pp. 1660–1685. DOI: [10.1137/140955483](https://doi.org/10.1137/140955483).
- Jawanpuria, P., M. Meghwanshi, and B. Mishra (2019). “Low-rank approximations of hyperbolic embeddings”. In: *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE. DOI: [10.1109/cdc40024.2019.902](https://doi.org/10.1109/cdc40024.2019.902).
- Keller-Ressel, M. and S. Nargang (2021). “The hyperbolic geometry of financial networks”. In: *Scientific Reports* 11.1. DOI: [10.1038/s41598-021-83328-4](https://doi.org/10.1038/s41598-021-83328-4).
- Krioukov, D., F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguñá (2010). “Hyperbolic geometry of complex networks”. In: *Physical Review E* 82.3. DOI: [10.1103/physreve.82.036106](https://doi.org/10.1103/physreve.82.036106).
- Lang, S. (1999). *Fundamentals of differential geometry*. Vol. 191. Graduate Texts in Mathematics. Springer-Verlag, New York, pp. xviii+535. DOI: [10.1007/978-1-4612-0541-8](https://doi.org/10.1007/978-1-4612-0541-8).
- Li, C., G. López, and V. Martín-Márquez (2009). “Monotone vector fields and the proximal point algorithm on Hadamard manifolds”. In: *Journal of the London Mathematical Society* 79.3, pp. 663–683. DOI: [10.1112/jlms/jdn087](https://doi.org/10.1112/jlms/jdn087).
- Li, C., B. S. Mordukhovich, J. Wang, and J.-C. Yao (2011). “Weak sharp minima on Riemannian manifolds”. In: *SIAM Journal on Optimization* 21.4, pp. 1523–1560. DOI: [10.1137/09075367X](https://doi.org/10.1137/09075367X).
- Lim, Y. (2012). “Factorizations and geometric means of positive definite matrices”. In: *Linear Algebra Appl.* 437.9, pp. 2159–2172. DOI: [10.1016/j.laa.2012.05.039](https://doi.org/10.1016/j.laa.2012.05.039).
- Manton, J. H. (2015). “A framework for generalising the Newton method and other iterative methods from Euclidean space to manifolds”. In: *Numerische Mathematik* 129.1, pp. 91–125. DOI: [10.1007/s00211-014-063](https://doi.org/10.1007/s00211-014-063).
- Miller, S. A. and J. Malick (2005). “Newton methods for nonsmooth convex minimization: connections among U-Lagrangian, Riemannian Newton and SQP methods”. In: *Mathematical Programming* 104.2-3, Ser. B, pp. 609–633. DOI: [10.1007/s10107-005-0631-2](https://doi.org/10.1007/s10107-005-0631-2).
- Moshiri, M., F. Safaei, and Z. Samei (2021). “A novel recovery strategy based on link prediction and hyperbolic geometry of complex networks”. In: *Journal of Complex Networks* 9.4. Ed. by M. D. Domenico. DOI: [10.1093/comnet/cnab007](https://doi.org/10.1093/comnet/cnab007).

- Muscoloni, A., J. M. Thomas, S. Ciucci, G. Bianconi, and C. V. Cannistraci (2017). “Machine learning meets complex networks via coalescent embedding in the hyperbolic space”. In: *Nature Communications* 1. DOI: [10.1038/s41467-017-01825-5](https://doi.org/10.1038/s41467-017-01825-5).
- Nesterov, Y. E. and M. J. Todd (2002). “On the Riemannian geometry defined by self-concordant barriers and interior-point methods”. In: *Foundations of Computational Mathematics. The Journal of the Society for the Foundations of Computational Mathematics* 2.4, pp. 333–361. DOI: [10.1007/s102080010032](https://doi.org/10.1007/s102080010032).
- Neto, J. X. D. C., O. P. Ferreira, and L. R. L. Pérez (2002). In: *Acta Mathematica Hungarica* 94.4, pp. 307–320. DOI: [10.1023/a:1015643612729](https://doi.org/10.1023/a:1015643612729).
- Nickel, M. and D. Kiela (2018). “Learning Continuous Hierarchies in the Lorentz Model of Hyperbolic Geometry”. In: *Proceedings of the 35th International Conference on Machine Learning*. Ed. by J. Dy and A. Krause. Vol. 80. Proceedings of Machine Learning Research. PMLR, pp. 3779–3788. URL: <https://proceedings.mlr.press/v80/nickel18a.html>.
- Park, F., J. Bobrow, and S. Ploen (1995). “A Lie Group Formulation of Robot Dynamics”. In: *The International Journal of Robotics Research* 14.6, pp. 609–618. DOI: [10.1177/027836499501400606](https://doi.org/10.1177/027836499501400606).
- Rapcsák, T. (1997). *Smooth nonlinear optimization in \mathbf{R}^n* . Vol. 19. Nonconvex Optimization and its Applications. Kluwer Academic Publishers, Dordrecht, pp. xiv+374. DOI: [10.1007/978-1-4615-6357-0](https://doi.org/10.1007/978-1-4615-6357-0).
- Rosenbrock, H. H. (1960). “An automatic method for finding the greatest or least value of a function”. In: *Comput. J.* 3, pp. 175–184. DOI: [10.1093/comjnl/3.3.175](https://doi.org/10.1093/comjnl/3.3.175). URL: <https://doi.org/10.1093/comjnl/3.3.175>.
- Rothaus, O. S. (1960). “Domains of positivity”. In: *Abh. Math. Sem. Univ. Hamburg* 24, pp. 189–235.
- Sakai, T. (1996). *Riemannian geometry*. Vol. 149. Translations of Mathematical Monographs. Translated from the 1992 Japanese original by the author. Providence, RI: American Mathematical Society, pp. xiv+358.
- Sharpee, T. O. (2019). “An argument for hyperbolic geometry in neural circuits”. In: *Current Opinion in Neurobiology* 58, pp. 101–104. DOI: [10.1016/j.conb.2019.07.008](https://doi.org/10.1016/j.conb.2019.07.008).
- Silva Louzeiro, M., R. Bergmann, and R. Herzog (2022). “Fenchel duality and a separation theorem on Hadamard manifolds”. In: *SIAM Journal on Optimization* 32.2, pp. 854–873. DOI: [10.1137/21m1400699](https://doi.org/10.1137/21m1400699). arXiv: [2102.11155](https://arxiv.org/abs/2102.11155).
- Smith, S. T. (1994). “Optimization techniques on Riemannian manifolds”. In: *Hamiltonian and gradient flows, algorithms and control*. Vol. 3. Fields Inst. Commun. Amer. Math. Soc., Providence, RI, pp. 113–136.
- Souza, J. C. O. and P. R. Oliveira (2015). “A proximal point algorithm for DC functions on Hadamard manifolds”. In: *Journal of Global Optimization* 63.4, pp. 797–810. DOI: [10.1007/s10898-015-0282-7](https://doi.org/10.1007/s10898-015-0282-7).
- Tabaghi, P. and I. Dokmanic (2021). “On Procrustes Analysis in Hyperbolic Space”. In: *IEEE Signal Processing Letters* 28, pp. 1120–1124. DOI: [10.1109/lsp.2021.3081379](https://doi.org/10.1109/lsp.2021.3081379).
- Tabaghi, P. and I. Dokmanić (2020). “Hyperbolic Distance Matrices”. In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM. DOI: [10.1145/3394486.3403224](https://doi.org/10.1145/3394486.3403224).
- Tao, P. D. and E. B. Souad (1986). “Algorithms for solving a class of nonconvex optimization problems. Methods of subgradients”. In: *FERMAT days 85: mathematics for optimization (Toulouse, 1985)*. Vol. 129. North-Holland Math. Stud. North-Holland, Amsterdam, pp. 249–271. DOI: [10.1016/S0304-0208\(08\)72402-2](https://doi.org/10.1016/S0304-0208(08)72402-2).
- Tao, P. D. and E. B. Souad (1988). “Duality in D.C. (Difference of Convex functions) Optimization. Subgradient Methods”. In: *Trends in Mathematical Optimization*. Birkhäuser Basel, pp. 277–293. DOI: [10.1007/978-3-0348-9297-1_18](https://doi.org/10.1007/978-3-0348-9297-1_18).
- Thi, H. A. L. and T. Pham Dinh (2018). “DC programming and DCA: thirty years of developments”. In: *Mathematical Programming* 169.1, pp. 5–68. DOI: [10.1007/s10107-018-1235-y](https://doi.org/10.1007/s10107-018-1235-y).
- Udriște, C. (1994). *Convex functions and optimization methods on Riemannian manifolds*. Vol. 297. Mathematics and its Applications. Kluwer Academic Publishers Group, Dordrecht, pp. xviii+348. DOI: [10.1007/978-94-015-8390-9](https://doi.org/10.1007/978-94-015-8390-9).

- Vollmer, F. W. (2018). “Automatic contouring of geologic fabric and finite strain data on the unit hyperboloid”. In: *Computers & Geosciences* 115, pp. 134–142. DOI: [10.1016/j.cageo.2018.03.006](https://doi.org/10.1016/j.cageo.2018.03.006).
- Wang, X., C. Li, J. Wang, and J.-C. Yao (2015). “Linear convergence of subgradient algorithm for convex feasibility on Riemannian manifolds”. In: *SIAM Journal on Optimization* 25.4, pp. 2334–2358. DOI: [10.1137/14099961X](https://doi.org/10.1137/14099961X).
- Weber, M. and S. Sra (2022). “Riemannian Optimization via Frank-Wolfe Methods”. In: *Mathematical Programming*. DOI: [10.1007/s10107-022-01840-5](https://doi.org/10.1007/s10107-022-01840-5).
- Weinmann, A., L. Demaret, and M. Storath (2014). “Total variation regularization for manifold-valued data”. In: *SIAM Journal on Imaging Sciences* 7.4, pp. 2226–2257. DOI: [10.1137/130951075](https://doi.org/10.1137/130951075).
- Weinmann, A., L. Demaret, and M. Storath (2016). “Mumford-Shah and Potts regularization for manifold-valued data”. In: *Journal of Mathematical Imaging and Vision* 55.3, pp. 428–445. DOI: [10.1007/s10851-015-0628-0](https://doi.org/10.1007/s10851-015-0628-0).
- Wen, Z. and W. Yin (2012). “A feasible method for optimization with orthogonality constraints”. In: *Mathematical Programming* 142.1-2, pp. 397–434. DOI: [10.1007/s10107-012-0584-1](https://doi.org/10.1007/s10107-012-0584-1).
- Wilson, R. C., E. R. Hancock, E. Pekalska, and R. P. W. Duin (2014). “Spherical and Hyperbolic Embeddings of Data”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36.11, pp. 2255–2269. DOI: [10.1109/tpami.2014.2316836](https://doi.org/10.1109/tpami.2014.2316836).
- Yamaji, A. (2008). “Theories of strain analysis from shape fabrics: A perspective using hyperbolic geometry”. In: *Journal of Structural Geology* 30.12, pp. 1451–1465. DOI: [10.1016/j.jsg.2008.07.011](https://doi.org/10.1016/j.jsg.2008.07.011).