The difference of convex algorithm on Riemannian manifolds

O. P. Ferreira * E. M. Santos* J. C. O. Souza †

December 13, 2021

Abstract

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. Equivalence between the classical and a simplified Riemannian version of the DCA is established. 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. Some duality relations between the DC problem and its dual are also established. Applications to the problem of maximizing a convex function in a compact set and manifold-valued image denoising are presented in the DC context.

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

MSC: 90C30, 90C26, 49N14, 49Q99

1 Introduction

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

$$\min_{p \in \mathcal{M}} \phi(p) := g(p) - h(p),$$  \hspace{1cm} (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, and the functions $g, h : \mathcal{M} \to \mathbb{R} := \mathbb{R} \cup \{+\infty\}$ are convex, lower semi-continuous and proper functions (called DC components).

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 concerned with practical issues at this point, we emphasize that practical applications appear 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 [7, 11, 13, 40, 61], and problems in medical imaging modeled in this context are addressed in [21]. Problems of tracking, robotics and scene motion analysis are also posed in Riemannian manifolds, see [25, 43], machine learning [42], artificial intelligence [40], neural circuits [48], low-rank approximations of hyperbolic embeddings [27, 52].
procrustes problems \[51\], financial networks \[29\], complex networks \[30,39\], embeddings of data \[63\] and strain analysis \[56,64\]. We also mention that there are many papers on statistics in the Riemannian context, see for example \[12,24\].

As aforementioned, the number of works dealing with concepts and techniques of nonlinear programming and convex analysis in the Riemannian scenario have also increased, see \[2,55\]. 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 \[1,20,26,36,38,41,49,57,62\] and references therein. In this sense, the concept of conjugate of a convex function was recently presented in the Riemannian setting, which is an important tool in convex analysis and play an important role in the theory of duality on Riemannian manifolds, see \[8,9\]. In particular, this definition enables us to propose a Riemannian version of 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 \[1\]. It was proposed by Pham Dinh and Souad \[53\] in the Euclidean setting. The basic idea of DCA is to compute a subgradient of each (convex) DC component separately, i.e., at each iterate \(k\), DCA calculates \(y_k \in \partial h(x_k)\) using this trial point to compute \(x^{k+1} \in \partial g^*(y^k)\), where \(\partial g^*\) 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 \(x^{k+1} \in \arg\min\{g(x) - h_k(x) : x \in \mathbb{R}^n\} \) (here called alternative version of DCA). On the other hand, computing \(y^k \in \partial h(x_k)\) is equivalent to find a solution of the dual problem \(\min\{h^*(y) - g^*(y_k - 1 - \langle y - y_k - 1, x_k \rangle : y \in \mathbb{R}^n\}\). 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 \[4,17,32\]. To the best of our knowledge Souza and Oliveira \[50\] was the first work dealing with DC functions in Riemannian manifolds, more precisely, the authors proposed the proximal point algorithm for DC functions studying the convergence of the method in Hadamard manifolds. Recently, Almeida et al. \[3\] proposed a modified version of \[50\] in the same Riemannian setting in order to accelerate the convergence of the method considered in \[50\].

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 \[1\]. We also extend some relations between the DC problem \[1\] and its dual to the Riemannian setting. Applications to the problem of maximizing a convex function in a compact set and and manifold-valued image denoising are provided.

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 \[4\] is presented our formulation of the Riemannian DCA and its well definition. In Section \[5\] we study the convergence properties of the proposed method. In Section \[6\] we provide some applications to the problem of maximizing a convex function in a compact set and manifold-valued image denoising. The last section contains some conclusions.
2 Notions and basic results

In this section, we recall some concepts, notations, and basics results about Riemannian manifolds and optimization. For more details see, for example, [19, 44, 17, 55]. 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 plane of $\mathcal{M}$ at $p$. The corresponding norm associated to the Riemannian metric $\langle \cdot, \cdot \rangle$ is denoted by $\| \cdot \|$. For each, $p \in \mathcal{M}$, denotes by $T_p^*\mathcal{M}$ the dual of $T_p\mathcal{M}$. The Riemannian metric of $\mathcal{M}$ provides a linear bijective correspondence between $T_p\mathcal{M}$ and $T_p^*\mathcal{M}$ via the Riesz map and its inverse. Therefore, we also denotes the Riemannian metric and the corresponding norm in $T_p^*\mathcal{M}$ by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, respectively. Moreover, the tangent bundle and cotangent bundle of $\mathcal{M}$, will be denoted respectively by $T\mathcal{M}$ and $T^*\mathcal{M}$; for more details see [5].

We use $\ell(\gamma)$ to denote the length of a piecewise smooth curve $\gamma : [a, b] \to \mathcal{M}$. 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. Throughout the paper all manifolds are assumed to be Hadamard manifolds. For $\mathcal{M}$ a Hadamard manifold and $p \in \mathcal{M}$, the exponential map $\exp_p : T_p\mathcal{M} \to \mathcal{M}$ is a diffeomorphism and $\exp_p^{-1} : \mathcal{M} \to T_p\mathcal{M}$ denotes its inverse. In this case, $d(p, q) = \| \exp_p^{-1}q \|$ and the function $d^2_{pq} : \mathcal{M} \to \mathbb{R}$ defined by

$$d^2_{pq}(p) := d^2(q, p)$$

is $C^\infty$ and $\text{grad}d^2_{pq}(p) := -2\exp_p^{-1}q$. Now, we recall some concepts and basic properties about optimization in the Riemannian context. For that, given two points $p, q \in \mathcal{M}$, $\Gamma_{pq}$ denotes the set of all geodesic segments $\gamma : [0, 1] \to \mathcal{M}$ with $\gamma(0) = p$ and $\gamma(1) = q$. The domain of a function $f : \mathcal{M} \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ 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 $p, q \in \mathcal{M}$ and $\gamma \in \Gamma_{pq}$, the composition $f \circ \gamma : [0, 1] \to \mathbb{R}$ is convex (resp. strictly convex), i.e., $(f \circ \gamma)(t) \leq (1 - t)f(p) + tf(q)$ (resp. $(f \circ \gamma)(t) < (1 - t)f(p) + tf(q)$), for all $t \in [0, 1]$. A function $f : \mathcal{M} \to \overline{\mathbb{R}}$ is said to be $\sigma$-strongly convex for $\sigma > 0$ if, for any $p, q \in \mathcal{M}$ and $\gamma \in \Gamma_{pq}$, the composition $f \circ \gamma : [0, 1] \to \mathbb{R}$ is $\sigma$-strongly convex, i.e., $(f \circ \gamma)(t) \leq (1 - t)f(p) + tf(q) - \frac{\sigma}{2}t(1 - t)d^2(q, p)$, for all $t \in [0, 1]$.

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

$$\partial f(p) := \{ \xi \in T_p^*\mathcal{M} \mid f(q) \geq f(p) + \langle \xi, \exp_p^{-1}q \rangle, \forall q \in \mathcal{M} \}.$$ 

The next result can be found for example in [55].

**Theorem 2.2.** Let $\mathcal{M}$ be a Hadamard manifold and $f : \mathcal{M} \to \mathbb{R}$ be a function. Then,

(i) The function $f$ is convex if and only if $f(p) \geq f(q) + \langle v, \exp^{-1}q \rangle$, for all $p, q \in \mathcal{M}$ and all $v \in \partial f(q)$;

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

The proof of the next result can be found in [57, Proposition 2.5].

**Proposition 2.3.** Let $f : \mathcal{M} \to \overline{\mathbb{R}}$ be a convex and lower semicontinuous function and consider $(x^k)_{k \in \mathbb{N}} \subset \text{int dom}(f)$ such that $\lim_{k \to +\infty} x^k = \bar{x} \in \text{int dom}(f)$. If $(v^k)_{k \in \mathbb{N}}$ is a sequence such that $v^k \in \partial f(x^k)$ for every $k \in \mathbb{N}$, then $(v^k)_{k \in \mathbb{N}}$ is bounded and its cluster points belongs to $\partial f(\bar{x})$.

The following definition play an important role in the paper, see [13, p. 363].

**Definition 2.4.** Let $\mathcal{M}$ be a Hadamard manifold. A function $f : \mathcal{M} \to \overline{\mathbb{R}}$ is said to be lower semi-continuous, or lsc, at $p \in \mathcal{M}$ if $\lim \inf_{x \to p} f(x) = f(p)$. 

3
Definition 2.5. A function $f : \mathcal{M} \to \mathbb{R}$ is said to be 1-coercive if
\[
\lim_{d(\bar{p}, p) \to +\infty} f(p)/d(\bar{p}, p) = +\infty,
\]
for some at $\bar{p} \in \mathcal{M}$.

Proposition 2.6. Assume that $f : \mathcal{M} \to \mathbb{R}$ is lsc and 1-coercive. Then the global minimizer set of $f$ is nonempty.

Proof. First note that $\lim_{d(\bar{p}, p) \to +\infty} f(p)/d(\bar{p}, p) = +\infty$ implies that
\[
\lim_{d(\bar{p}, p) \to +\infty} f(p) = +\infty.
\]
Given $\bar{p} \in \mathcal{M}$, there exists $\bar{r} > 0$ such that $\bar{r} < d(\bar{p}, p)$ implies that $f(\bar{p}) \leq f(p)$. Consider $B[\bar{p}, \bar{r}] := \{p \in \mathcal{M} : d(p, \bar{p}) \leq \bar{r}\}$. Since $\mathcal{M}$ is a Hadamard manifold, the Hopf-Rinow’s theorem ensures that $B[\bar{p}, \bar{r}]$ is compact. Thus, taking into account that $f$ is lsc, by [13 Theorem 3, p. 361] there exists $\bar{p} \in B[\bar{p}, \bar{r}]$ such that $f(\bar{p}) \leq f(p)$, for all $p \in B[\bar{p}, \bar{r}]$. Therefore, the global minimizer of $f$ is $\bar{p}$ or $\hat{\bar{p}}$. □

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

Proof. Since the function $g : \mathcal{M} \to \mathbb{R}$ is a $\sigma$-strongly convex, Theorem 2.2 (i) implies that
\[
g(p) \geq g(\bar{p}) + \langle s, \exp_{\bar{p}}^{-1} p \rangle + \frac{\sigma}{2} d(\bar{p}, p), \quad \forall p \in \mathcal{M}, \forall s \in \partial g(\bar{p}).
\]
Thus, considering that $f(p) = g(p) - \langle w, \exp_{\bar{p}}^{-1} p \rangle$ and using the last inequality we conclude
\[
\frac{f(p)}{d(\bar{p}, p)} \geq \frac{g(\bar{p})}{d(\bar{p}, p)} + \langle s, \frac{\exp_{\bar{p}}^{-1} p}{d(\bar{p}, p)} \rangle + \frac{\sigma}{2} d(\bar{p}, p) - \langle w, \frac{\exp_{\bar{p}}^{-1} p}{d(\bar{p}, p)} \rangle, \quad \forall s \in \partial g(\bar{p}).
\]
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
\[
\lim_{d(\bar{p}, p) \to +\infty} \frac{f(p)}{d(\bar{p}, p)} = +\infty.
\]
Therefore, $f$ is 1-coercive. The second part of the proposition is an immediate consequence of the first one combined with Proposition 2.6. □

The statement and proof of the next proposition can be found in [35, Lemma 2.4, p. 666].

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

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

(ii) If $v^k \in T_{\bar{x}}\mathcal{M}$ and $\lim_{k \to +\infty} v^k = \bar{v}$, then $\bar{v} \in T_{\bar{x}}\mathcal{M}$.

(iii) Given $u^k, v^k \in T_{\bar{x}}\mathcal{M}$ and $\bar{u}, \bar{v} \in T_{\bar{x}}\mathcal{M}$, if $\lim_{k \to +\infty} u^k = \bar{u}$ and $\lim_{k \to +\infty} v^k = \bar{v}$, then $\lim_{k \to +\infty} \langle u^k, v^k \rangle = \langle \bar{u}, \bar{v} \rangle$. 

4
The dual DC problem of the problem (2), is stated as follows

\[ \text{Theorem 2.10. Let } f : \mathcal{M} \to \mathbb{R} \text{ be a proper function. Then, the Fenchel-Young inequality holds, i.e., for all } (p, \xi) \in \mathcal{T}^{*} \mathcal{M} \text{ we have} \]

\[ f(p) + f^{*}(p, \xi) \geq \langle \xi, \exp_{p}^{-1} q \rangle, \quad \forall q \in \mathcal{M}. \]

\[ \text{Theorem 2.11. Let } f : \mathcal{M} \to \mathbb{R} \text{ be a proper lsc convex function and } p \in \mathcal{M}. \text{ Then the function} \]

\[ f^{*}(p, \cdot) : \mathcal{T}_{p}^{*} \mathcal{M} \to \mathbb{R} \text{ is convex and proper.} \]

\[ \text{Definition 2.15. Let } f : \mathcal{M} \to \mathbb{R}. \text{ The Fenchel biconjugate of } f \text{ is the function } f^{**} : \mathcal{M} \to \mathbb{R} \text{ defined by} \]

\[ f^{**}(p) := \sup_{(q, \xi) \in \mathcal{T}^{*} \mathcal{M}} \{ \langle \xi, \exp_{p}^{-1} q \rangle - f^{*}(q, \xi) \}, \quad p \in \mathcal{M}. \]

\[ \text{Theorem 2.16. Let } f : \mathcal{M} \to \mathbb{R} \text{ be a proper lsc convex function. Then, } f^{**} = f \text{ holds.} \]

\[ \text{Theorem 2.17. Let } f : \mathcal{M} \to \mathbb{R} \text{ be a proper convex function. Then, } \eta \in \partial f(p) \text{ if and only if} \]

\[ f^{*}(p, \eta) = -f(p). \]

3 Duality in DC optimization in Hadamard manifolds

In this section our aim is to state and study difference of convex problem or DC problem, and its dual problem, called dual DC problem, in the Riemannian context. The DC problem is a non-convex and, in general, a non-smooth problem, and its statement is as follows

\[ \min_{x \in \mathcal{M}} \phi(x) := g(x) - h(x), \quad (2) \]

where \( g : \mathcal{M} \to \mathbb{R} \) and \( h : \mathcal{M} \to \mathbb{R} \) are proper, lsc and convex functions, with the conventions

\[ (+\infty) - (+\infty) = +\infty, \quad (+\infty) - \lambda = +\infty, \quad \text{and } \lambda - (+\infty) = -\infty, \quad \forall \lambda \in \mathbb{R}. \quad (3) \]

The dual DC problem of the problem (2), is stated as follows

\[ \min_{(p, \xi) \in \mathcal{T}^{*} \mathcal{M}} \varphi(p, \xi) := h^{*}(p, \xi) - g^{*}(p, \xi). \quad (4) \]
Lemma 3.5. If \( M = \mathbb{R}^n \), then \( T_pM \simeq \mathbb{R}^n \) for all \( p \in M \). Consequently, \( T^*M \simeq \mathbb{R}^n \). Moreover, by using Remark 2.14, we obtain

\[
h^*(p, \xi) - g^*(p, \xi) = h^*(\xi) - \langle \xi, p \rangle - (g^*(\xi) - \langle \xi, p \rangle) = h^*(\xi) - g^*(\xi), \quad \forall \xi \in \mathbb{R}^n.
\]

Therefore, if \( M = \mathbb{R}^n \) then problem (4) becomes the problem

\[
\min_{\xi \in \mathbb{R}^n} \{ h^*(\xi) - g^*(\xi) \}.
\]

In conclusion, for \( M = \mathbb{R}^n \), the dual (4) of the problem (2) merges into the dual stated in [54].

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

1. \( g : M \to \mathbb{R} \) and \( h : M \to \mathbb{R} \) are \( \sigma \)-strongly convex and lsc functions, where \( \sigma > 0 \);
2. \( \phi_{\text{inf}} := \inf \{ \phi(x) := g(x) - h(x) : x \in M \} > -\infty \);
3. \( \text{dom}(g) \subseteq \text{int dom}(h) \);
4. \( \partial g^*(p, \xi) \neq \emptyset \), for every \( \xi \in \text{dom}(g^*(p, \cdot)) := \{ \xi \in T_p^*M : g^*(p, \xi) < +\infty \} \).

Next, we discuss the above assumptions. First, we show that (1) is not restrictive.

Remark 3.2. Let \( q \in M \) and \( \sigma > 0 \). Consider the function \( M \ni p \mapsto \frac{\sigma}{2}d^2(q, p) \), which is \( \sigma \)-strongly convex, see [10, Corollary 3.1]. If \( g : M \to \mathbb{R} \) and \( h : M \to \mathbb{R} \) are convex, then taking \( q \in M \) and setting \( g(p) = g(p) + \frac{\sigma}{2}d^2(q, p) \) and \( h(p) = h(p) + \frac{\sigma}{2}d^2(q, p) \) we obtain two \( \sigma \)-strongly convex functions \( g \) and \( h \) in \( M \). In addition, \( \phi(p) = g(p) - h(p) = g(p) - h(p), \) for all \( p \in M \).

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

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

Definition 3.4. A point \( x^* \in M \) is critical of \( \phi \) in (2) if \( \partial g(x^*) \cap \partial h(x^*) \neq \emptyset \) holds.

Next lemma establishes a necessary condition for a point \((\bar{p}, \bar{\xi}) \in T^*M \) be a solution of problem (4).

Lemma 3.5. If \((\bar{p}, \bar{\xi}) \) is a solution of problem (4), then \( \partial_2 g^*(\bar{p}, \bar{\xi}) \subseteq \partial_2 h^*(\bar{p}, \bar{\xi}) \).
Proof. Let \((\bar{p}, \bar{\xi})\) be a solution of problem (4). Then, \(h^*(p, \eta) - g^*(p, \eta) \geq h^*(\bar{p}, \bar{\xi}) - g^*(\bar{p}, \bar{\xi})\), for all \((p, \eta) \in T^*\mathcal{M}\), or equivalently,

\[
h^*(p, \eta) - h^*(\bar{p}, \bar{\xi}) \geq g^*(p, \eta) - g^*(\bar{p}, \bar{\xi}), \quad \forall (p, \eta) \in T^*\mathcal{M}.
\]

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

\[
h^*(\bar{p}, \eta) - h^*(\bar{p}, \bar{\xi}) \geq \langle X, \eta - \bar{\xi} \rangle, \quad \forall \xi \in T_{\bar{p}}\mathcal{M}.
\]

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

Remark 3.6. If \(\mathcal{M} = \mathbb{R}^n\), then by using Remark 2.14 we have \(\partial_2 g^*(\bar{p}, \bar{\xi}) = \partial g^*(\bar{\xi}) - \{\bar{p}\}\) and \(\partial_2 h^*(\bar{p}, \bar{\xi}) = \partial h^*(\bar{\xi}) - \{\bar{p}\}\). Thus, from Remark 2.1 and Lemma 3.5 we conclude that if \(\bar{\xi} \in \mathbb{R}^n\) is a solution of problem (5), then we have \(\partial g^*(\bar{\xi}) \subseteq \partial h^*(\bar{\xi})\), which retrieves [54, Theorem 2.1-(2)].

It follows from Lemma 3.5 that if \((\bar{p}, \bar{\xi})\) is a solution of problem (4), then the set \(\partial_2 h^*(\bar{p}, \bar{\xi}) \cap \partial_2 g^*(\bar{p}, \bar{\xi})\) is nonempty. Hence, we define the notion of critical point for the problem (4) as follows.

Definition 3.7. A point \((\bar{p}, \bar{\xi})\) is critical for problem (4) if

\[
\partial_2 h^*(\bar{p}, \bar{\xi}) \cap \partial_2 g^*(\bar{p}, \bar{\xi}) \neq \emptyset.
\]

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

To proceed with our analysis we need the following lemma, for a proof of it see [6, p. 46].

Lemma 3.9. Let \(X\) and \(Y\) be nonempty sets and \(f : X \times Y \rightarrow \mathbb{R}\). There holds

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

Next theorem presents the relation between the optimum values of Problems 2 and 4.

Theorem 3.10. There holds

\[
\inf_{(q, \xi) \in T^*\mathcal{M}} \{h^*(q, \xi) - g^*(q, \xi)\} = \inf_{p \in \mathcal{M}} \{g(p) - h(p)\}.
\]

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

\[
\inf_{p \in \mathcal{M}} \{g(p) - h(p)\} = \inf \{g(p) - h^{**}(p) : p \in \mathcal{M}\}
\]

\[
= \inf \left\{ g(p) - \sup_{(q, \xi) \in T^*\mathcal{M}} \{\langle \xi, \exp_q^{-1} p \rangle - h^*(q, \xi)\} : p \in \mathcal{M} \right\}.
\]

Some algebraic manipulations in the right hand side of the last equality show that

\[
\inf_{p \in \mathcal{M}} \{g(p) - h(p)\} = \inf_{p \in \mathcal{M}} \inf_{(q, \xi) \in T^*\mathcal{M}} \{g(p) + h^*(q, \xi) - \langle \xi, \exp_q^{-1} p \rangle\},
\]
which by Lemma 3.9 is equivalent to

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

Since

$$\inf_{p \in \mathcal{M}} \{ g(p) + h^*(q,\xi) - \langle \xi, \exp^{-1}_q p \rangle \} = h^*(q,\xi) + \sup_{p \in \mathcal{M}} \langle \xi, \exp^{-1}_q p \rangle - \{ g(p) \},$$

we have

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

which, by using Definition 2.9, gives the desired equality and the proof is concluded. □ □

**Theorem 3.11.** The following statements hold:

(i) If \( \bar{p} \in \mathcal{M} \) is a solution of problem (2), then \( (\bar{p}, \eta) \in T^*\mathcal{M} \) is a solution of problem (4), for all \( \eta \in \partial h(\bar{p}) \cap \partial g(\bar{p}) \).

(ii) If \( (\bar{p}, \eta) \in T^*\mathcal{M} \) is a solution of problem (4), for some \( \eta \in \partial h(\bar{p}) \cap \partial g(\bar{p}) \), 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 \( \eta \in \partial h(\bar{p}) \cap \partial g(\bar{p}) \). Since \( g \) and \( h \) are convex, by Theorem 2.17 we have

$$-g^*(\bar{p}, \eta) = g(\bar{p}) \text{ and } h^*(\bar{p}, \eta) = -h(\bar{p}),$$

which implies that \( h^*(\bar{p}, \eta) - g^*(\bar{p}, \eta) = g(\bar{p}) - h(\bar{p}) \). Using again that \( \bar{p} \in \mathcal{M} \) is a solution of problem (2), the last equality together with Theorem 3.10 ensure that \( (\bar{p}, \eta) \) 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 \( (\bar{p}, \eta) \) is a solution of problem (4) with \( \eta \in \partial h(\bar{p}) \cap \partial g(\bar{p}) \). Since \( g \) and \( h \) are convex and \( \eta \in \partial h(\bar{p}) \cap \partial g(\bar{p}) \), it follows from Theorem 2.17 that

$$-g^*(\bar{p}, \eta) = g(\bar{p}) \text{ and } h^*(\bar{p}, \eta) = -h(\bar{p}),$$

which implies

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

On the other hand, Theorem 3.10 implies that

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

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). □ □

### 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. In the sequel, we introduce a second version of the DCA on Hadamard manifolds based on a first-order approximation of the second component. We also show the well definition of those 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.

Next two results will be used to show the well definition of Algorithm 1.
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 \( \xi^k \in \partial h(p^k) \), and compute
\[
v^k \in \partial 2g^*(p^k, \xi^k), \quad p^{k+1} := \exp_{p^k} v^k.
\] (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.

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
\[
y \in \text{dom}(g^*(p, \cdot)) = \{ \xi \in T_p^* \mathcal{M} : g^*(p, \xi) < +\infty \}.
\]
In particular, \( \partial 2g^*(p, y) \neq \emptyset \).

Proof. Assume that \( p \in \text{dom}(h) \) and take \( y \in \partial h(p) \). Thus, by using Theorem 2.17 we obtain
\[
h^*(p, y) = -h(p) < +\infty.
\] (8)
From Theorem 3.10 and assumption 2 we have that
\[
h^*(p, y) - g^*(p, y) \geq \inf_{(q, \xi) \in T^* \mathcal{M}} \{ h^*(q, \xi) - g^*(q, \xi) \} = \inf_{x \in \mathcal{M}} \{ g(x) - h(x) \} > -\infty.
\] (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 \( \eta \in T^* \mathcal{M} \) such that \( h^*(p, \eta) < +\infty \) and \( g^*(p, \eta) = +\infty \). By using (8) we have \( h^*(p, \eta) - g^*(p, \eta) = h^*(p, \eta) - (+\infty) = -\infty \), 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 3 we conclude that \( \partial 2g^*(p, y) \neq \emptyset \).

Proposition 4.2. Algorithm 1 is well defined.

Proof. Assume \( p^k \in \text{dom}(g) \). From Remark 3.3 we have that \( \text{dom}(\phi) = \text{dom}(g) \subseteq \text{dom}(h) \), and hence \( p^k \in \text{dom}(h) \). By assumption 3 we have that \( \partial h(p^k) \neq \emptyset \). Let \( \xi^k \in \partial h(p^k) \). Since \( h \) is convex, Theorem 2.17 implies that \( h^*(p^k, \xi^k) = -h(p^k) < +\infty \). By the first part of Lemma 4.1 we have that \( g^*(p^k, \xi^k) < +\infty \) and \( \partial 2g^*(p^k, \xi^k) \neq \emptyset \). Let \( v^k \in \partial 2g^*(p^k, \xi^k) \). Since \( \mathcal{M} \) is Hadamard, the point \( p^{k+1} = \exp_{p^k} v^k \) is well defined and belongs to \( \mathcal{M} \). Moreover, applying Theorem 2.13 with \( f = g, p = p^k, v = s^k \) and \( \xi = y^k \) we have \( g(p^{k+1}) + g^*(p^k, \xi^k) = \langle \xi^k, v^k \rangle \) or equivalently \( g(p^{k+1}) - \langle \xi^k, v^k \rangle < +\infty \), which implies that \( p^{k+1} \in \text{dom}(g) = \text{dom}(\phi) \subseteq \text{dom}(h) \). Therefore, Algorithm 1 is well defined.

In the following, we present a second form of the DCA, which is equivalent to Algorithm 1. First of all, note that due to the point \( p^{k+1} \) be a solution of (10) we have
\[
g(p) - \langle \xi^k, \exp_{p^k}^{-1} p \rangle \geq g(p^{k+1}) - \langle \xi^k, \exp_{p^k}^{-1} p^{k+1} \rangle, \quad \forall p \in \mathcal{M}.
\] (11)
This inequality plays an important role throughout the paper. Next, we show the well definition of Algorithm 2.

Proposition 4.3. Algorithm 2 is well defined.
Algorithm 2: The DC Algorithm on Hadamard Manifolds (DCA)

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

\[
    p^{k+1} \in \text{argmin}_{p \in \mathcal{M}} \left( g(p) - \langle \xi^k, \exp_{p^k}^{-1} p \rangle \right).
\]

(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.

**Proof.** Assume that \( p^k \in \text{dom}(g) \). From Remark 3.3 we have that \( \text{dom}(g) = \text{dom}(\phi) \subseteq \text{dom}(h) \), which implies that \( p^k \in \text{dom}(h) \). Thus, by Assumption 2.1 we have that \( \partial h(p^k) \neq \emptyset \). Let \( \xi^k \in \partial h(p^k) \). From Lemma 2.7, we have that \( g_k : \mathcal{M} \rightarrow \mathbb{R} \) given by \( g_k(p) = g(p) - \langle \xi^k, \exp_{p^k}^{-1} p \rangle \) is 1-coercive. Consequently, its minimizer set is non-empty and is contained in \( \text{dom}(g) \). Therefore, there exists \( p^{k+1} \in \text{dom}(g) = \text{dom}(\phi) \) such that \( p^{k+1} \in \text{argmin}_{p \in \mathcal{M}} (g(p) - \langle \xi^k, \exp_{p^k}^{-1} p \rangle) \), which implies that Algorithm 2 is well defined.

**Remark 4.4.** If \( \mathcal{M} = \mathbb{R}^n \), then by Remark 2.14 we have that \( \partial^2 g^*(p^k, \xi^k) = \partial g^*(\xi^k) - \{ p^k \} \) and consequently \( v^k + p^k = \exp_{p^k} v^k = p^{k+1} \in \partial g^*(\xi^k) = \partial g^*(p^k, \xi^k) + \{ p^k \} \), i.e., \( p^{k+1} \in \partial g^*(\xi^k) \) and \( \xi^k \in \partial h(p^k) \). Therefore, Algorithm 2 merges into the classical formulation of the DCA; see [5,54]. Moreover, if \( \mathcal{M} = \mathbb{R}^n \), then (11) becomes

\[
    g(p) - \langle \xi^k, p - p^k \rangle \geq g(p^{k+1}) - \langle \xi^k, p^{k+1} - p^k \rangle, \quad \forall p \in \mathbb{R}^n,
\]

which is equivalent to \( p^{k+1} = \text{argmin}_{x \in \mathbb{R}^n} \{ g(p) - \langle \xi^k, p - p^k \rangle \} \). Hence, Algorithm 2 retrieves alternative version of the DCA.

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

**Proposition 4.5.** If \( p^k \in \text{dom}(g) \), \( \xi^k \in \partial h(p^k) \) and \( v^k \in \partial^2 g^*(p^k, \xi^k) \), then \( p^{k+1} = \exp_{p^k} v^k \) if and only if, \( p^{k+1} \in \text{argmin}_{p \in \mathcal{M}} (g(p) - \langle \xi^k, \exp_{p^k}^{-1} p \rangle) \). Consequently, Algorithm 1 is equivalent to Algorithm 2.

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

\[
    g(p^{k+1}) - \langle \xi^k, \exp_{p^k}^{-1} p^{k+1} \rangle = -g^*(p^k, \xi^k) = -\sup_{q \in \mathcal{M}} \left( \langle \xi^k, \exp_{p^k}^{-1} q \rangle - g(q) \right),
\]

or equivalently, \( g(p^{k+1}) - \langle \xi^k, \exp_{p^k}^{-1} p^{k+1} \rangle = \inf_{q \in \mathcal{M}} (g(q) - \langle \xi^k, \exp_{p^k}^{-1} q \rangle) \). This is also equivalent to \( p^{k+1} \in \text{argmin}_{p \in \mathcal{M}} (g(p) - \langle \xi^k, \exp_{p^k}^{-1} p \rangle) \). Therefore, Algorithm 1 is equivalent to Algorithm 2.

**5 Convergence analysis of DCA**

The aim of this section is to study the convergence properties of DCA. It is worth to mention that the results in this section can be proved using both formulations of DCA in Algorithm 1.
and 2 due to the equivalence between them; see Proposition 4.5. For simplicity, we present the results only using Algorithm 2 but the proofs of the results for Algorithm 1 is 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

\[
\phi(p^{k+1}) \leq \phi(p^k) - \frac{\sigma}{2} d^2(p^k, p^{k+1}).
\]

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

**Proof.** By using inequality in (11) with \(p = p^k\) we have \(g(p^k) - g(p^{k+1}) \geq \langle -\xi^k, \exp^{-1}_p p^{k+1} \rangle\). On the other hand, due to \(h\) be \(\sigma\)-strongly convex and \(\xi^k \in \partial h(p^k)\), we obtain that

\[
h(p^{k+1}) - h(p^k) \geq \langle \xi^k, \exp^{-1}_p p^{k+1} \rangle + \frac{\sigma}{2} d^2(p^{k+1}, p^k).
\]

Hence, using that \(\phi = g - h\) 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 \(g(p) \geq g(p^k) + \langle \xi^k, \exp^{-1}_p p \rangle\), for all \(p \in \mathcal{M}\), which shows that \(y^k \in \partial g(p^k)\). Hence, taking into account that \(\xi^k \in \partial h(p^k)\), we conclude that \(\xi^k \in \partial g(p^k) \cap \partial h(p^k) = \emptyset\). Therefore, it follows from Definition 3.4 that \(p^k\) is a critical point of \(\phi\) 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, \(\lim_{k \to +\infty} d(p^k, p^{k+1}) = 0\).

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

\[
\sum_{k=0}^{T} d^2(p^k, p^{k+1}) \leq \frac{2}{\sigma} \left( \phi(p^0) - \phi(p^{k+1}) \right) \leq \frac{2}{\sigma} \left( \phi(p^0) - \phi_{\text{inf}} \right),
\]

for each \(T \in \mathbb{N}\), where \(\phi_{\text{inf}} > -\infty\) is given by assumption 2. Taking the limit in the last inequality, as \(T\) goes to +\(\infty\), we obtain that 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 \((\xi^k)_{k \in \mathbb{N}}\) 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 \(\xi\) a cluster point \((\xi^k)_{k \in \mathbb{N}}\) such that \(\xi \in \partial g(\bar{p}) \cap \partial h(\bar{p})\). Consequently, every cluster point of \((p^k)_{k \in \mathbb{N}}\), if any, is a critical point of \(\phi\).

**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 \(\lim_{k \to +\infty} p^k = \bar{p}\). It follows from Proposition 5.1 together with assumption 2 that \((\phi(p^k))_{k \in \mathbb{N}}\) is non-increasing and converges. Moreover, due to \(\phi(p^0) \geq \phi(p^k) = g(p^k) - h(p^k)\) and \(g\) be lsc, we have

\[
\phi(p^0) \geq \liminf_{k \to +\infty} g(p^k) - \limsup_{k \to +\infty} h(p^k) = g(\bar{p}) - \limsup_{k \to +\infty} h(p^k).
\]
Thus, using the convention \(\overline{p}\) we conclude that \(\overline{p} \in \text{dom}(g)\). Hence, using assumption \(\overline{p}\) we conclude that \(\overline{p} \in \text{int dom}(h)\). We know that \(\xi^k \in \partial h(p^k)\), for all \(k \in \mathbb{N}\). Thus, by Proposition \(\overline{p}\) we can also assume that \(\lim_{k \to +\infty} \xi^k = \overline{\xi} \in \partial h(\overline{p})\). Due to the point \(p^{k+1}\) be a solution of \((10)\), it satisfies \((11)\). Thus, taking the inferior limit in \((11)\), as \(k \to +\infty\), and using the fact that \(\lim_{k \to +\infty} p^k = \overline{p}\), \(g\) is lsc toghether with Proposition \(\overline{p}\) iii) and Proposition \(\overline{p}\), we obtain

\[
g(p) \geq \liminf_{k \to +\infty} \left( g(p^{k+1}) + \langle \xi^k, \exp_{\overline{p}}^{-1} p \rangle - \langle \xi^k, \exp_{\overline{p}}^{-1} p^{k+1} \rangle \right) \geq g(\overline{p}) + \langle \overline{\xi}, \exp_{\overline{p}}^{-1} p \rangle,
\]

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

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

\[
\min_{k=0,1,\ldots,N} d(p^k, p^{k+1}) \leq \left( \frac{2(\phi(p^0) - \phi_{\text{inf}})}{(N + 1)\sigma} \right)^{1/2}.
\]

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

\[
(N + 1) \min_{k=0,1,\ldots,N} d^2(p^k, p^{k+1}) \leq \sum_{k=0}^N \frac{2}{\sigma} (\phi(p^k) - \phi(p^{k+1})) \leq \frac{2}{\sigma} (\phi(p^0) - \phi_{\text{inf}}),
\]

where \(\phi_{\text{inf}} > -\infty\) is given by assumption \(\overline{p}\). Therefore, the desired inequality directly follows.

The last result of this section establishes a primal-dual asymptotic convergence of the sequences generated by the DCA. This result extends to Hadamard manifolds setting its counter-part for DCA on \(\mathbb{R}^n\), which appears in \(\overline{p}\) Theorem 3). Due to the nature of the problem, we will use the formulation of the DCA given in Algorithm \(\overline{p}\).

**Theorem 5.5.** Let \((p^k)_{k \in \mathbb{N}}\) and \((\xi^k)_{k \in \mathbb{N}}\) be the sequences generated by Algorithm \(\overline{p}\). The following statements hold:

(i) \(g(p^{k+1}) - h(p^{k+1}) \leq h^*(p^k, \xi^k) - g^*(p^k, \xi^k) \leq g(p^k) - h(p^k)\), for all \(k = 0, 1, \ldots\)

(ii) \(\lim_{k \to +\infty} (g(p^k) - h(p^k)) = \lim_{k \to +\infty} (h^*(p^k, \xi^k) - g^*(p^k, \xi^k)) = \overline{\phi} \geq \phi_{\text{inf}}\).

(iii) If the sequence \((p^k)_{k \in \mathbb{N}}\) is bounded and \(\overline{p}\) is a cluster point of \((p^k)_{k \in \mathbb{N}}\), then \(\overline{p} \in \text{dom}(g)\) and there exists a cluster point \(\xi\) of \((\xi^k)_{k \in \mathbb{N}}\) such that

\[
\partial g(\overline{p}) \cap \partial h(\overline{p}) \neq \emptyset,
\]

\[
\lim_{k \to +\infty} (h(p^k) + h^*(p^k, \xi^k)) = h(\overline{p}) + h^*(\overline{p}, \overline{\xi}) = 0,
\]

\[
\lim_{k \to +\infty} (g(p^k) + g^*(p^k, \xi^k)) = g(\overline{p}) + g^*(\overline{p}, \overline{\xi}) = 0.
\]

\[
\partial_2 h^*(\overline{p}, \overline{\xi}) \cap \partial_2 g^*(\overline{p}, \overline{\xi}) \neq \emptyset,
\]

\[
g(\overline{p}) - h(\overline{p}) = h^*(\overline{p}, \overline{\xi}) - g^*(\overline{p}, \overline{\xi}) = \overline{\phi},
\]
Proof. (i): By applying Theorem 2.10 with \( q = p^{k+1}, p = p^k, \xi = \xi^k \) and \( f = h \), we obtain that
\[
h(p^{k+1}) + h^*(p^k, \xi^k) \geq \langle \xi^k, \exp^{-1}_p p^{k+1} \rangle.
\]
Since \( \exp^{-1}_p p^{k+1} \in \partial g^*(p^k, \xi^k) \), we can apply Theorem 2.13 with \( f = g, p = p^k, v = \exp^{-1}_p p^{k+1} and \xi = \xi^k \) to obtain \( \langle \xi^k, \exp^{-1}_p p^{k+1} \rangle = g(p^{k+1}) + g^*(p^k, \xi^k). \) Hence, we have \( h(p^{k+1}) + h^*(p^k, \xi^k) \geq g(p^{k+1}) + g^*(p^k, \xi^k), \) which is equivalent to the first inequality of item (i). To prove the second one, we first note that since \( \xi^k \in \partial h(p^k) \) and \( h \) is convex, by using Theorem 2.17 we have \( h^*(p^k, \xi^k) + h(p^k) = 0. \) Thus, applying Theorem 2.10 with \( q = p^k, \xi = \xi^k \) and \( f = g \), we have \( 0 \leq g^*(p^k, \xi^k) + g(p^k), \) which combined with the last equality yields the second inequality of item (i).

(ii): First we recall that \( \phi = g - h \) satisfies assumption \( 2. \) Thus, item (i) implies that \( (\phi(p^k))_{k \in \mathbb{N}} \) is non-increasing and convergent. Hence \( \lim_{k \to +\infty} (g(p^k) - h(p^k)) =: \bar{\phi} \in \mathbb{R}. \) Moreover, by using again item (i), we also have
\[
\lim_{k \to +\infty} (h^*(p^k, \xi^k) - g^*(p^k, \xi^k)) =: \bar{\phi} \in \mathbb{R}.
\]
Finally, the inequality in item (ii) follows from assumption \( 2. \)

(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 there exists a cluster point \( \xi \) of \( (\xi^k)_{k \in \mathbb{N}} \) such that \( \xi \in \partial g(\bar{p}) \cap \partial h(\bar{p}). \) Therefore, (13a) is proved. Before proceeding with the proof we note that due to \( \bar{p} \in \text{dom}(g) \), assumption \( 3. \) implies that \( \bar{p} \in \text{dom}(h). \)

To prove (13b) note that since \( \xi^k \in \partial h(p^k) \), for all \( k \in \mathbb{N}, \) and \( h \) is convex, from Theorem 2.17 we have \( h(p^k) + h^*(p^k, \xi^k) = 0, \) for all \( k \in \mathbb{N}. \) Consequently, \( \lim_{k \to +\infty} (h(p^k) + h^*(p^k, \xi^k)) = 0. \) Since \( \xi \in \partial h(\bar{p}), \) using again Theorem 2.17, we have \( h(\bar{p}) + h^*(\bar{p}, \xi) = 0 \) and (13b) directly follows.

To prove (13c) we first note that
\[
g(p^k) + g^*(p^k, \xi^k) = g(p^k) - h(p^k) - (h^*(p^k, \xi^k) - g^*(p^k, \xi^k)) + h(p^k) + h^*(p^k, \xi^k).
\]
Thus, using item (ii) together with (13b), we have \( \lim_{k \to +\infty} (g(p^k) + g^*(p^k, \xi^k)) = 0. \) Since \( \xi \in \partial g(\bar{p}), \) using again Theorem 2.17 we have \( g(\bar{p}) + g^*(\bar{p}, \xi) = 0, \) which combined with the last equality yields (13c).

We proceed to prove (13d). For that, we assume without loss of generality that \( \lim_{k \to +\infty} p^k = \bar{p}. \) Now, by applying Theorem 2.10 with \( f = h, p = \bar{p}, q = p^k \) and \( \xi = \eta, \) we obtain
\[
h(p^k) + h^*(\bar{p}, \eta) \geq \langle \eta, \exp^{-1}_p p^k \rangle, \quad \forall \eta \in \mathcal{T}_{\bar{p}}^* \mathcal{M}, \quad \forall k \in \mathbb{N}.
\]
Thus, by using Definition 2.4 \( \lim_{k \to +\infty} p^k = \bar{p}, \) Proposition 2.8 (i) and (iii) and that \( h \) is lsc, we have \( h(\bar{p}) + h^*(\bar{p}, \eta) = \lim_{k \to +\infty} h(p^k) + h^*(\bar{p}, \eta) \geq 0. \) Thus, the second equality in (13b) implies that
\[
h^*(\bar{p}, \eta) \geq h^*(\bar{p}, \xi), \quad \forall \eta \in \mathcal{T}_{\bar{p}}^* \mathcal{M}.
\]
Hence, \( 0 \in \partial h^*(\bar{p}, \xi). \) Similarly, by using (13d), we can also show that \( 0 \in \partial g^*(\bar{p}, \xi). \) Therefore, \( 0 \in \partial h^*(\bar{p}, \xi) \cap \partial g^*(\bar{p}, \xi), \) 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 \( \lim_{k \to +\infty} h(p^k) = h(\bar{p}). \) Thus, using item (ii), we conclude that
\[
\lim_{k \to +\infty} g(p^k) = \lim_{k \to +\infty} (g(p^k) - h(p^k)) + \lim_{k \to +\infty} h(p^k) = \bar{\phi} + h(\bar{p}).
\]
Hence, using Definition 2.4, we have \( \lim_{k \to +\infty} g(p^k) = \lim_{k \to +\infty} \inf_{k \to +\infty} g(p^k) = g(\bar{p}). \) Therefore, we obtain that \( g(\bar{p}) - h(\bar{p}) = \bar{\phi}, \) which concludes the proof. \( \square \)
6 Applications

In this section we present some applications of the DCA to solve a constrained convex maximization problem and manifold-valued image denoising on Hadamard manifolds. In the remain of this paper, \([t]_+\) denotes the positive part of \(t\), i.e., the function defined by \([t]_+ := \max\{0, t\}\), and \(\text{sgn}(t)\) denotes the signal of \(t\), i.e., the function defined by \(\text{sgn}(t) := -1, \text{if } t < 0, \text{sgn}(t) := 0, \text{if } t = 0, \text{and } \text{sgn}(t) := 1, \text{if } t > 0\).

6.1 Constrained convex maximization

In the following, we present a general formulation of a constrained convex maximization (CCM) problem, which is a theoretical model for several practical applications. Let \(M\) be a Hadamard manifold, \(C \subset M\) be a compact and convex set and \(h : M \to \mathbb{R}\) be a convex function. The constrained convex maximization problem is stated as follows:

\[
\max_{p \in C} h(p) \quad (14)
\]

Next, we show that problem (14) is equivalently stated as a DC problem. In fact, define \(g : M \to \overline{\mathbb{R}}\) the indicate function of the set \(C\) as \(g(p) := 0, \text{for } p \in C\), and \(g(p) := +\infty, \text{otherwise}\). Thus, the problem (14) is equivalent to

\[
- \min_{p \in M} \phi(p) := g(p) - h(p),
\]

which is the DC problem studied in the previous sections. For proceed, we first note that for a given \(\xi \in \partial h(q)\), we have

\[
\min_{p \in M} \left(g(p) - \langle \xi, \exp^{-1}_q p \rangle \right) = \min_{p \in C} - \langle \xi, \exp^{-1}_q p \rangle.
\]

Hence, the DC algorithm for solving CCM problem is formally stated in Algorithm 3:

**Algorithm 3:** The DC Algorithm to CCM

1. Choose an initial point \(p^0 \in C\). Set \(k = 0\).
2. Take \(\xi^k \in \partial h(p^k)\), and the next iterated \(p^{k+1}\) is defined as following

\[
p^{k+1} \in \arg\min_{p \in C} \left(-\xi^k, \exp^{-1}_q p \right).
\]

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

It is worth to note that the subproblem (15) is closely related to the Frank-Wolfe subproblem studied in [58, Algorithm 2, p. 20], see also [59]. We recall that subproblem (15) is linear in the Euclidean space, but not in a general Riemannian manifolds. This can constitute a practical difficult for the implementation of the method. However, as we will show in next section, it can be solved in some special manifolds.

6.1.1 Convex maximization on the set of symmetric positive definite matrices

In this section, we consider the symmetric positive definite (SPD) matrices cone \(\mathbb{P}_+^{n+}\). Following Rothaus [15], see also [11 Section 6.3], let \(\mathcal{M} := (\mathbb{P}_+^{n+}, \langle \cdot, \cdot \rangle)\) be the Riemannian manifold endowed with the metric given by \(\langle U, V \rangle := \text{tr}(VX^{-1}UX^{-1})\), for \(X \in \mathcal{M}\) and \(U, V \in T_X\mathcal{M}\), where \(\text{tr}(X)\)...
denotes the trace of the matrix $X$, $T_XM \approx \mathbb{P}^n$ is the tangent plane of $M$ at $X$ and $\mathbb{P}^n$ denotes the set of symmetric matrices of order $n \times n$. We recall that $M$ is a Hadamard manifold, see for example \cite{31} Theorem 1.2. p. 325. Consider the problem \eqref{14} with
$$M = \mathbb{P}^n_{++}, \quad C := \{X \in M \mid L \preceq X \preceq U\},$$
where $L, U \in M$, $L \prec U$. In this case, if $X_k$ is the current iterate of Algorithm\cite{33} then subproblem \eqref{15} is equivalent to
$$\min_{L \preceq Z \preceq U} \langle -W_k, \exp^{-1}_{X_k}(X) \rangle,$$
where $W_k \in \partial h(X_k)$. We first recall that $\exp^{-1}_{X_k}(X) = X_k^{1/2} \log(X_k^{-1/2}XX_k^{-1/2})X_k^{1/2}$. Therefore, problem \eqref{16} becomes
$$\min_{L \preceq Z \preceq U} \text{tr}\left(S_k \log(\hat{X}_kZ\hat{X}_k)\right),$$
where $S_k := -X_k^{-1/2}W_kX_k^{-1/2}$ and $\hat{X}_k := X_k^{-1/2}$. The explicit solution of the problem \eqref{17} is given by the following theorem, which can be found in \cite{58} Theorem 2.4.1.

**Theorem 6.1.** 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 $Z = X^{-1}Q \left(P^T[-\text{sgn}(D)]_+P + \hat{L}\right)Q^TX^{-1}$, where $S = QDQ^T$ is a diagonalization of $S$, $\hat{U} - \hat{L} = P^TP$ with $\hat{L} = Q^TXLXQ$ and $\hat{U} = Q^TXUXQ$, where $[-\text{sgn}(D)]_+$ is the diagonal matrix
$$\text{diag}([-\text{sgn}(d_{11})]_+, \ldots, [-\text{sgn}(d_{nn})]_+)$$
and $D = (d_{ij})$.

**6.1.2 Convex maximization on the positive orthant**

In this section, we present an example in the positive orthant. For that, we denote the set of $n \times n$ matrices with real entries by $\mathbb{R}^{n \times n}$, the $n$-dimensional Euclidean space by $\mathbb{R}^n = \mathbb{R}^{n \times 1}$, the positive orthant by
$$\mathbb{R}^n_+ = \{x = (x_1, \ldots, x_n)^T \in \mathbb{R}^{n \times 1} : x_i \geq 0, \ i = 1, \ldots, n\},$$
and its interior by $\mathbb{R}^n_{++}$. Set $\text{diag}(y) := \text{diag}(y_1, \ldots, y_n) \in \mathbb{R}^{n \times n}$ the diagonal matrix with $(i, i)$-th entry equal to $y_i$, $i = 1, \ldots, n$. Let $M := (\mathbb{R}^n_{++}, G)$ be the Hadamard manifold obtained by endowing $\mathbb{R}^n_{++}$ with the new Riemannian metric $(u, v) := u^T G(x) v$, for all $x \in \mathbb{R}^n_+$ and $u, v \in T_x \mathbb{R}^n_+ \equiv \mathbb{R}^n$, where
$$G(x) := \text{diag}(x_1^{-2}, \ldots, x_n^{-2}) \in \mathbb{R}^{n \times n}.$$ 
(18)
We recall that $M := (\mathbb{R}^n_{++}, G)$ is a Hadamard manifold and the inverse of the exponential is
$$\exp_{x_k}^{-1}(x) = \left(x_k \log(x_1/x_1^k), \ldots, x_n \log(x_n/x_n^k)\right),$$
for details see for example [41] Section 6.1, [22] Section 4.1 and [23] Example 4.4]. Consider the problem (14) with $\mathcal{M} := (\mathbb{R}^n_{+}, G)$, $\mathcal{C} := \{ x \in \mathcal{M} : \ a \preceq x \preceq b \}$, where $a \preceq x \preceq b$ means that $a_i \leq x_i \leq b_i$, for all $i = 1, \ldots, n$ with $a, b \in \mathbb{R}^n_{+}$. In this case, if $x^k$ is the current iterate of Algorithm 3, then subproblem (15) is given by

$$\min_{a \preceq x \preceq b} \ f(x) := -\sum_{i=1}^{n} \frac{\varsigma^k}{x_i} (\log x_i - \log x_i^k) = \sum_{i=1}^{n} \frac{\varsigma^k}{x_i} (\log x_i^k - \log x_i),$$

which attains the minimum on $\mathcal{C}$ at a vertex, i.e.,

$$x^{k+1} := \left( [-\text{sgn}(-\xi^k_1)]_+ (b_1 - a_1) + a_1, \ldots, [-\text{sgn}(-\xi^k_n)]_+ (b_n - a_n) + a_n \right).$$

Therefore, in this case, a solution of (14) is a vertex of the box $\mathcal{C}$.

6.1.3 Convex maximization on Euclidean space with a new metric

The aim of this section is to present a particular instance of problem (14). For that, endow $\mathbb{R}^{2n}$ with the new Riemannian metric $\langle u, v \rangle := u^T G(x)v$, where $u, v \in \mathbb{R}^{2n}$ and $G(x)$ is the $2n \times 2n$ block diagonal matrix $G(x) = \text{diag}(G_1(x), \ldots, G_n(x))$, where the blocks are given by

$$G_i(x) := \begin{pmatrix} 1 + 4x_{2i-1}^2 & -2x_{2i-1} \\ -2x_{2i-1} & 1 \end{pmatrix}, \quad i = 1, \ldots, n,$$

and $x = (x_1, \ldots, x_{2n}) \in \mathbb{R}^{2n}$. Thus, we obtain a Riemannian manifold $\mathcal{M} := (\mathbb{R}^{2n}, G)$. Given $x \in \mathcal{M}$, the exponential map $\text{exp}_x : T_x \mathcal{M} \to \mathcal{M}$ is given by

$$\text{exp}_x(v) = (x_1 + v_1, v_1, x_2 + v_2, v_2, \ldots, x_{2n-1} + v_{2n-1}, v_{2n-1} + x_{2n} + v_{2n}, v_{2n}),$$

where $x := (x_1, \ldots, x_{2n})$ and $v := (v_1, \ldots, v_{2n})$. Thus, the inverse of $\text{exp}_x$ at $x^k$ is given by

$$\text{exp}_{x}^{-1}(x^k) = (x_1^k - x_1, x_2^k - x_2 - (x_2^k - x_1)^2, \ldots, x_1^k - 2x_{2i-1} - x_2 - (x_{2i-1} - x_2)^2),$$

Consider the problem (14) with $h : \mathcal{M} \to \mathbb{R}$ is differentiable and

$$\mathcal{M} := (\mathbb{R}^{2n}, G), \quad \mathcal{C} := \{ x \in \mathbb{R}^{2n} : u_i \leq x_i \leq v_i, i = 1, \ldots, 2n \},$$

and $u, v \in \mathbb{R}^{2n}$. The gradient of $h$ at a point $x \in \mathcal{M}$ is given by $\text{grad} h(x) = G(x)^{-1} h'(x)$, where $h'(x)$ is the Euclidean gradient of $h$ at $x$. After some calculations we have $\langle -\text{grad} h(x^k), \text{exp}_{x}^{-1}(x^k) \rangle = -\text{exp}_{x}^{-1}(x^k)^T h'(x^k)$, or equivalently,

$$\langle -\text{grad} h(x^k), \text{exp}_{x}^{-1}(x^k) \rangle = -\sum_{i=1}^{n} (x_{2i-1}^k - x_{2i-1}) h'_{2i-1}(x^k) + (x_{2i}^k - x_{2i} - (x_{2i}^k - x_{2i-1})^2) h'_{2i}(x^k).$$

Therefore, if $x^k$ is the current iterate of Algorithm 3 then subproblem (15) is given by

$$x^{k+1} \in \arg\min_{p \in \mathcal{C}} -\sum_{i=1}^{n} (x_{2i-1}^k - x_{2i-1}) h'_{2i-1}(x^k) + (x_{2i}^k - x_{2i} - (x_{2i}^k - x_{2i-1})^2) h'_{2i}(x^k),$$

(19)
where the objective function is a quadratic function. To simplify the notations, we define
\[
\begin{align*}
a_j := \begin{cases} 
0, & \text{if } j \text{ even}, \\
2h'_{j+1}(x^k), & \text{if } j \text{ odd}
\end{cases}, \\
b_j := \begin{cases} 
h'_j(x^k), & \text{if } j \text{ even}, \\
-2h'_{j+1}(x^k)x_j^k + h'_j(x^k), & \text{if } j \text{ odd},
\end{cases}, \\
c_j := -x_j^k h'_j(x^k) + (((x_j^k)^2 - x_j^{k+1}) h'_{j+1}(x^k), & j = 1, \ldots, 2n.
\end{align*}
\]
By using definitions of \(a_j, b_j\) and \(c_j\), (19) becomes the following quadratic problem
\[
x^{k+1} \in \arg\min_{p \in C} \sum_{j=1}^{2n} a_j x_j^2 + b_j x_j + c_j.
\]
(20)

Next, we present the solution of problem (20), which can be found in [28, Corollary 2.2, p.476].

**Proposition 6.2.** Let \(\bar{x} = (\bar{x}_1, \ldots, \bar{x}_{2n}) \in C\). Consider \(A = \text{diag}(a_1, \ldots, a_{2n})\),
\[
\hat{a}_j = \min\{0, a_j\}, \quad \hat{x}_j := \begin{cases} 
-1, & \text{if } \bar{x}_j = u_j, \\
1, & \text{if } \bar{x}_j = v_j, \\
b_j + (A\bar{x}_j)_j, & \text{if } u_j < \bar{x}_j < v_j,
\end{cases}
\]
for \(j = 1, \ldots, 2n\). Then, the point \(\bar{x}\) is a global minimizer of problem (20) if and only if,
\[
\bar{x}_j(b_j + a_j \bar{x}_j) - \frac{1}{2} \hat{a}_j(v_j - u_j) \leq 0,
\]
for each \(j = 1, \ldots, 2n\).

### 6.2 Manifold-valued image denoising

In this section, we show how the manifold-valued image denoising can be modeled as a DC problem. Although we are not concerned with practical issues at this time, we emphasize that practical applications appear whenever the natural structure of the data is modeled as an optimization problem on a Riemannian manifold. As aforementioned, data taking values in a manifold appear naturally in various signal and image processing applications, and hence, processing manifold-valued data has gained a lot of interest in recent years. Recently, there has also been made progress in extending TV regularization to arbitrary Riemannian manifolds; see [9] and references therein. Lellmann et al. [34] presented a first framework and an algorithmic solution for TV regularization for arbitrary Riemannian manifolds.

Noise is an unavoidable component of digital image acquisition and noise removal is a fundamental task in digital image processing. In 1992, Rudin, Osher, and Fatemi [40] proposed an approach for problem of real-valued image denoising and restoration which is known in the literature as ROF model. Their approach is based on minimizing the Total Variation (TV) functional \(TV(u) = \int_{\Omega} |\nabla u|\) of an image \(u : \Omega \to \mathbb{R}\) subject to the constraints \(\int_{\Omega} u = \int_{\Omega} a\) and \(\int_{\Omega} (u - a)^2 = \sigma^2\), where \(a\) denotes the noisy image and \(\sigma\) is a given constant. Chambolle and Lions [15] showed that this problem has a unique solution and is equivalent to minimizing the TV-functional
\[
J(u) = \frac{1}{2} ||u - a||_2^2 + \lambda TV(u),
\]
where \(\lambda > 0\) depends on \(\sigma\). A discrete version of the TV-functional can be done by using a finite difference discretization of \(\nabla u\). For 2D-images, there are in the literature the anisotropic version
\[
TV_{\text{aniso}}(u) = \sum_{i,j} |u_{i+1,j} - u_{i,j}| + |u_{i,j+1} - u_{i,j}|
\]
and the isotropic version
\[ TV_{iso}(u) = \sum_{i,j} \sqrt{(u_{i+1,j} - u_{i,j})^2 + (u_{i,j+1} - u_{i,j})^2}. \]

We are interested in restoration of images characterized by sparse gradients, a feature which, in general, characterizes piecewise constant images. The denoised image \( u \) is computed as the solution of the minimization problem
\[
\min_{u \in \mathbb{R}^n} \left\{ J(u) = \frac{\mu}{2} \| u - a \|^2 + \sum_{i=1}^{n} \varphi(||(\nabla u)_i||) \right\}, \tag{21}
\]
where where \( \mu > 0 \) is a fidelity parameter, \( \varphi : \mathbb{R} \to \mathbb{R} \) is the regularization (or penalty) function and \( (\nabla u)_i \in \mathbb{R}^2 \) denotes the discretization of the gradient image \( u \) at pixel \( i \), that is, \( (\nabla u)_i \) represents finite difference approximations of first-order horizontal and vertical partial derivatives. The purpose of the regularization term in (21) is to encode a prior on the image gradients magnitude, and the function \( \varphi \) is intended so as to strongly promote sparsity and consequently to better fit gradient distributions of real images. A very popular choice for \( \varphi \) is the Tikhonov regularization, which involves a smooth and convex regularization term, i.e., \( l_2 \) norm, \( \varphi(t) = t^2 \). However, it is well known that Tikhonov regularization does not well preserve edges in the restoration process. Non-smooth and convex regularization terms, for instance, TV regularization, i.e., \( l_1 \) norm, \( \varphi(t) = |t| \) better allow for discontinuities in the restored image, making them an acceptable penalty function for images. We denote by
\[ TV(u) = \sum_{i=1}^{n} ||\nabla u_i||. \]

Convex formulations benefit from convex optimization theory which leads to robust algorithms with guaranteed convergence. On the other hand, non-smooth and non-convex regularization has remarkable advantages over convex regularization for restoring images, in particular to restore high-quality piecewise constant images with neat edges. However, it may lead to challenging computation since it requires non-convex and non-smooth minimization. We denote by
\[ TV(\varphi)(u) = \sum_{i=1}^{n} \varphi(||\nabla u_i||), \]
for a given regularization function \( \varphi \).

In [9], the authors formulated the ROF model on manifolds using the \( l^2 \)-TV model in the discrete setting as follows: let \( F = (f_{i,j})_{i,j} \in \mathcal{M}^{d_1 \times d_2}, d_1, d_2 \in \mathbb{N} \) be a manifold-valued image, i.e., each pixel \( f_{i,j} \) takes values on a manifold \( \mathcal{M} \). Then, the manifold \( l^2 \)-TV energy functional reads as follows:
\[
\mathcal{E}_q(P) = \frac{1}{2\alpha} \sum_{i,j=1}^{d_1,d_2} d_{\mathcal{M}}^q(f_{i,j}, p_{i,j}) + \|\nabla P\|_{q,1}, \quad P = (p_{i,j})_{i,j} \in \mathcal{M}^{d_1 \times d_2},
\]
where \( q \in \{1, 2\} \) depending on the value of \( q \) the energy functional is called isotropic (\( q = 2 \)) and anisotropic (\( q = 1 \)). Moreover, \( \nabla : \mathcal{M}^{d_1 \times d_2} \to \mathcal{T}\mathcal{M}^{d_1 \times d_2} \) denotes the generalization of the one-sided finite difference operator which is defined as
\[
\|\nabla P\|_{q,1} = \left( \sum_{i,j} d_{\mathcal{M}}^q(x_{i,j}, x_{i+1,j}) + \sum_{i,j} d_{\mathcal{M}}^q(x_{i,j}, x_{i,j+1}) \right)^{\frac{1}{q}}.
\]
To fit into our setting we will consider the case where $\mathcal{M}$ is a Hadamard manifold. Now, denote by $TV(P) = ||\nabla P||_{1,1}$. Our goal is to prove that under mild assumptions, the following (anisotropic) $l^p$-TV energy functional

$$E_{\varphi}^p(P) = \frac{1}{p\alpha} \sum_{i,j} d^p_{\mathcal{M}}(f_{i,j}, p_{i,j}) + \varphi(||\nabla P||_{1,1})$$

(22)

can be written as a difference of two convex functions. To this end, assume that $\varphi : \mathbb{R}_+ \to \mathbb{R}$ is a concave and non-decreasing function and denote by $TV_{\varphi}(P) = \varphi(||\nabla P||_{1,1})$. The next results extend to the Riemannian setting the ones considered in [18]. In the Euclidean context, DC theory has shown a nice model to deal with image denoising; see for instance [18, 33].

**Lemma 6.3.** Let $f : \mathcal{M} \to \mathbb{R}_+$ be a convex function. If $\varphi : \mathbb{R}_+ \to \mathbb{R}$ is a concave and non-decreasing function such that $\varphi'_+(0) < \infty$, then $\tau f(x) - \varphi(f(x))$ is convex for all $\tau \geq \varphi'_+(0)$.

**Proof.** Since $\varphi$ is concave, its right derivative exists and $\varphi'_+(t) \geq 0$, for all $t \geq 0$. Furthermore, $\varphi'_+(t_1) \geq \varphi'_+(t_2)$, for all $0 \leq t_1 \leq t_2$. Define $\psi : \mathbb{R}_+ \to \mathbb{R}$ as

$$\psi(t) = \tau t - \varphi(t),$$

for $\tau \geq \varphi'_+(0) \geq 0$ fixed. Clearly, $\psi$ is convex and hence its right derivative exists and

$$\psi'_+(t) = \tau - \varphi'_+(t) \geq \varphi'_+(0) - \varphi'_+(t) \geq 0, \quad \forall t \geq 0.$$

Thus, $\psi$ is non-decreasing on $\mathbb{R}_+$. Therefore, from [55, Theorem 3.2], we have that $\psi(f(x)) = \tau f(x) - \varphi(f(x))$ is convex. \hfill \Box

**Proposition 6.4.** Let $\varphi : \mathbb{R}_+ \to \mathbb{R}$ be as in Lemma 6.3. If $\tau \geq \varphi'_+(0)$, then $\tau TV(\cdot) - TV_{\varphi}(\cdot)$ is a convex function.

**Proof.** The proposition directly follows from Lemma 6.3 taking into account that $TV(\cdot)$ is convex (sum of convex functions) and the definition of $TV_{\varphi}(\cdot)$. \hfill \Box

Now, we are able to consider a DC decomposition of the $l^p$-TV energy functional (22). Indeed,

$$E_{\varphi}^p(P) = \frac{1}{p\alpha} \sum_{i,j} d^p_{\mathcal{M}}(f_{i,j}, p_{i,j}) + \varphi(||\nabla P||_{1,1})$$

$$= \left[ \frac{1}{p\alpha} \sum_{i,j} d^p_{\mathcal{M}}(f_{i,j}, p_{i,j}) + TV(P) \right] - [\tau TV(P) - TV_{\varphi}(P)]$$

$$= f_1(P) - f_2(P),$$

where $f_1$ is a convex function since it is sum of convex functions and $f_2$ is a convex function due to Proposition 6.4.

### 7 Conclusions

This paper have proposed and studied convergence properties of a Riemannian version of the DCA to solve a DC problem. We show that this method can be extended to the Riemannian context. From this point of view, our paper is part of an effort to theoretically explaining the applicability of DC theory and DC algorithms in a nonlinear context. We expect that the results of this paper become a first step towards a better understanding of the convergence properties of DCA and a starting point for numerical implementation in this new setting. We foresee further progress in this topic in the nearby future.
References

[1] P.-A. Absil, C. G. Baker, and K. A. Gallivan. Trust-region methods on Riemannian manifolds. *Found. Comput. Math.*, 7(3):303–330, 2007.

[2] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, Princeton, NJ, 2008. With a foreword by Paul Van Dooren.

[3] Y. T. Almeida, J. X. da Cruz Neto, P. R. Oliveira, and J. C. de Oliveira Souza. A modified proximal point method for dc functions on hadamard manifolds. *Computational Optimization and Applications*, pages 1–25, 2020.

[4] L. T. H. An and P. D. Tao. The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems. *Ann. Oper. Res.*, 133:23–46, 2005.

[5] L. T. H. An, P. D. Tao, and L. D. Muu. Numerical solution for optimization over the efficient set by d.c. optimization algorithms. *Operations Research Letters*, 1996.

[6] R. G. Bartle and D. R. Sherbert. *Introduction to real analysis*, volume 2. Wiley New York, 2000.

[7] M. Bačák, R. Bergmann, G. Steidl, and A. Weinmann. A second order nonsmooth variational model for restoring manifold-valued images. *SIAM J. Sci. Comput.*, 38(1):A567–A597, 2016.

[8] R. Bergmann, R. Herzog, and M. S. Louzeiro. Fenchel duality and a separation theorem on hadamard manifolds. *arXiv preprint arXiv:2102.11155*, 2021.

[9] R. Bergmann, R. Herzog, M. Silva Louzeiro, D. Tenbrinck, and J. Vidal-Núñez. Fenchel Duality Theory and a Primal-Dual Algorithm on Riemannian Manifolds. *Found. Comput. Math.*, 21(6):1465–1504, 2021.

[10] R. Bergmann, J. Persch, and G. Steidl. A parallel Douglas-Rachford algorithm for minimizing ROF-like functionals on images with values in symmetric Hadamard manifolds. *SIAM J. Imaging Sci.*, 9(3):901–937, 2016.

[11] R. Bergmann and A. Weinmann. A second-order TV-type approach for inpainting and denoising higher dimensional combined cyclic and vector space data. *J. Math. Imaging Vision*, 55(3):401–427, 2016.

[12] A. Bhattacharya and R. Bhattacharya. Statistics on Riemannian manifolds: asymptotic distribution and curvature. *Proc. Amer. Math. Soc.*, 136(8):2959–2967, 2008.

[13] N. Bourbaki. *General Topology: Chapters 1–4*. Springer-Verlag Berlin Heidelberg, 1995.

[14] K. Bredies, M. Holler, M. Storath, and A. Weinmann. Total generalized variation for manifold-valued data. *SIAM J. Imaging Sci.*, 11(3):1785–1848, 2018.

[15] A. Chambolle and P.-L. Lions. Image recovery via total variation minimization and related problems. *Numer. Math.*, 76(2):167–188, 1997.

[16] J. X. da Cruz Neto, O. P. Ferreira, and L. R. Lucambio Pérez. Contributions to the study of monotone vector fields. *Acta Math. Hungar.*, 94(4):307–320, 2002.
[17] W. de Oliveira. The ABC of DC programming. *Set-Valued Var. Anal.*, 28(4):679–706, 2020.

[18] W. de Oliveira and M. P. Tcheou. An inertial algorithm for DC programming. *Set-Valued Var. Anal.*, 27(4):895–919, 2019.

[19] M. P. do Carmo. *Riemannian geometry*. Mathematics: Theory & Applications. Birkhäuser Boston, Inc., Boston, MA, 1992. Translated from the second Portuguese edition by Francis Flaherty.

[20] A. Edelman, T. A. Arias, and S. T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.*, 20(2):303–353, 1999.

[21] M. Esposito, C. Hennersperger, R. Gobl, L. Demaret, M. Storath, N. Navab, M. Baust, and A. Weinmann. Total variation regularization of pose signals with an application to 3d freehand ultrasound. *J. Math. Imaging Vision*, 2019.

[22] O. P. Ferreira, M. S. Louzeiro, and L. F. Prudente. Gradient method for optimization on Riemannian manifolds with lower bounded curvature. *SIAM J. Optim.*, 29(4):2517–2541, 2019.

[23] O. P. Ferreira, M. S. Louzeiro, and L. F. Prudente. Iteration-complexity and asymptotic analysis of steepest descent method for multiobjective optimization on Riemannian manifolds. *J. Optim. Theory Appl.*, 184(2):507–533, 2020.

[24] P. T. Fletcher. Geodesic regression and the theory of least squares on Riemannian manifolds. *Int. J. Comput. Vis.*, 105(2):171–185, 2013.

[25] O. Freifeld and M. J. Black. *Lie Bodies: A Manifold Representation of 3D Human Shape*. in Proceedings of ECCV 2012, Springer Berlin Heidelberg, 2012.

[26] W. Huang, K. A. Gallivan, and P.-A. Absil. A Broyden class of quasi-Newton methods for Riemannian optimization. *SIAM J. Optim.*, 25(3):1660–1685, 2015.

[27] P. Jawanpuria, M. Meghwanshi, and B. Mishra. Low-rank approximations of hyperbolic embeddings. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 7159–7164. IEEE, 2019.

[28] V. Jeyakumar, A. M. Rubinov, and Z. Y. Wu. Sufficient global optimality conditions for non-convex quadratic minimization problems with box constraints. *J. Global Optim.*, 36(3):471–481, 2006.

[29] M. Keller-Ressel and S. Nargang. The hyperbolic geometry of financial networks. *Scientific reports*, 11(1):1–12, 2021.

[30] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguñá. Hyperbolic geometry of complex networks. *Phys. Rev. E (3)*, 82(3):036106, 18, 2010.

[31] S. Lang. *Fundamentals of differential geometry*, volume 191 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1999.

[32] H. A. Le Thi and T. P. Dinh. Dc programming and dca: thirty years of developments. *Mathematical Programming*, 169(1):5–68, 2018.
[33] H. A. Le Thi and T. Pham Dinh. Difference of convex functions algorithms (DCA) for image restoration via a Markov random field model. Optim. Eng., 18(4):873–906, 2017.

[34] J. Lellmann, E. Strekalovskiy, S. Koetter, and D. Cremers. Total variation regularization for functions with values in a manifold. In Proceedings of the IEEE International Conference on Computer Vision, pages 2944–2951. 2013.

[35] C. Li, G. López, and V. Martín-Márquez. Monotone vector fields and the proximal point algorithm on hadamard manifolds. Journal of the London Mathematical Society, 79(3):663–683, 2009.

[36] C. Li, B. S. Mordukhovich, J. Wang, and J.-C. Yao. Weak sharp minima on Riemannian manifolds. SIAM J. Optim., 21(4):1523–1560, 2011.

[37] J. H. Manton. A framework for generalising the Newton method and other iterative methods from Euclidean space to manifolds. Numer. Math., 129(1):91–125, 2015.

[38] S. A. Miller and J. Malick. Newton methods for nonsmooth convex minimization: connections among U-Lagrangian, Riemannian Newton and SQP methods. Math. Program., 104(2-3, Ser. B):609–633, 2005.

[39] M. Moshiri, F. Safaei, and Z. Samei. A novel recovery strategy based on link prediction and hyperbolic geometry of complex networks. J. Complex Netw., 9(4):Paper No. cnab007, 2021.

[40] A. Muscoloni, J. M. Thomas, S. Ciucci, G. Bianconi, and C. V. Cannistraci. Machine learning meets complex networks via coalescent embedding in the hyperbolic space. Nature communications, 8(1):1–19, 2017.

[41] Y. E. Nesterov and M. J. Todd. On the Riemannian geometry defined by self-concordant barriers and interior-point methods. Found. Comput. Math., 2(4):333–361, 2002.

[42] M. Nickel and D. Kiela. Learning continuous hierarchies in the lorentz model of hyperbolic geometry. In International Conference on Machine Learning, pages 3779–3788. PMLR, 2018.

[43] F. C. Park, J. E. Bobrow, and S. R. Ploen. A lie group formulation of robot dynamics. Int. J. Rob. Res., 14(6):609–618, 1995.

[44] T. Rapcsák. Smooth nonlinear optimization in $\mathbb{R}^n$, volume 19 of Nonconvex Optimization and its Applications. Kluwer Academic Publishers, Dordrecht, 1997.

[45] O. S. Rothaus. Domains of positivity. Abh. Math. Sem. Univ. Hamburg, 24:189–235, 1960.

[46] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. volume 60, pages 259–268. 1992. Experimental mathematics: computational issues in nonlinear science (Los Alamos, NM, 1991).

[47] T. Sakai. Riemannian geometry, volume 149 of Translations of Mathematical Monographs. American Mathematical Society, Providence, RI, 1996. Translated from the 1992 Japanese original by the author.

[48] T. O. Sharpee. An argument for hyperbolic geometry in neural circuits. Current opinion in neurobiology, 58:101–104, 2019.
[49] S. T. Smith. Optimization techniques on Riemannian manifolds. In Hamiltonian and gradient flows, algorithms and control, volume 3 of Fields Inst. Commun., pages 113–136. Amer. Math. Soc., Providence, RI, 1994.

[50] J. Souza and P. Oliveira. A proximal point algorithm for dc functions on hadamard manifolds. Journal of Global Optimization, 63(4):797–810, 2015.

[51] P. Tabaghi and I. Dokmanić. On procrustes analysis in hyperbolic space. IEEE Signal Processing Letters, 28:1120–1124, 2021.

[52] P. Tabaghi and I. Dokmanić. Hyperbolic distance matrices. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pages 1728–1738, 2020.

[53] P. D. Tao and E. B. Souad. Algorithms for solving a class of nonconvex optimization problems. Methods of subgradients. In FERMAT days 85: mathematics for optimization (Toulouse, 1985), volume 129 of North-Holland Math. Stud., pages 249–271. North-Holland, Amsterdam, 1986.

[54] P. D. Tao and E. B. Souad. Duality in dc (difference of convex functions) optimization. Subgradient methods. Trends in mathematical optimization, pages 277–293, 1988.

[55] C. Udriște. Convex functions and optimization methods on Riemannian manifolds, volume 297 of Mathematics and its Applications. Kluwer Academic Publishers Group, Dordrecht, 1994.

[56] F. W. Vollmer. Automatic contouring of geologic fabric and finite strain data on the unit hyperboloid. Computers & Geosciences, 115:134–142, 2018.

[57] X. Wang, C. Li, J. Wang, and J.-C. Yao. Linear convergence of subgradient algorithm for convex feasibility on Riemannian manifolds. SIAM J. Optim., 25(4):2334–2358, 2015.

[58] M. Weber. On Geometric Optimization, Learning and Control. ProQuest LLC, Ann Arbor, MI, 2021. Thesis (Ph.D.)–Princeton University.

[59] M. Weber and S. Sra. Projection-free nonconvex stochastic optimization on Riemannian manifolds. IMA Journal of Numerical Analysis, 09 2021. drab066.

[60] A. Weinmann, L. Demaret, and M. Storath. Total variation regularization for manifold-valued data. SIAM J. Imaging Sci., 7(4):2226–2257, 2014.

[61] A. Weinmann, L. Demaret, and M. Storath. Mumford-Shah and Potts regularization for manifold-valued data. J. Math. Imaging Vision, 55(3):428–445, 2016.

[62] Z. Wen and W. Yin. A feasible method for optimization with orthogonality constraints. Math. Program., 142(1-2, Ser. A):397–434, 2013.

[63] R. C. Wilson, E. R. Hancock, E. Pekalska, and R. P. Duin. Spherical and hyperbolic embeddings of data. IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(11):2255–2269, 2014.

[64] A. Yamaji. Theories of strain analysis from shape fabrics: A perspective using hyperbolic geometry. Journal of Structural Geology, 30(12):1451–1465, 2008.