Riemannian optimization on unit sphere with *p*-norm and its applications

Hiroyuki Sato*

February 24, 2022

Abstract

This paper deals with Riemannian optimization on the unit sphere in terms of p-norm with general p > 1. As a Riemannian submanifold of the Euclidean space, the geometry of the sphere with p-norm is investigated, and several geometric tools used for Riemannian optimization, such as retractions and vector transports, are proposed and analyzed. Applications to Riemannian optimization on the sphere with nonnegative constraints and L_p -regularization-related optimization are also discussed. As practical examples, the former includes nonnegative principal component analysis and the latter is closely related to the Lasso regression and box-constrained problems. Numerical experiments verify that Riemannian optimization on the sphere with p-norm has substantial potential for such applications, and the proposed framework provides a theoretical basis for such optimization.

Keywords: *p*-norm, Sphere, Riemannian optimization, Nonnegative PCA, Lasso regression, Box-constrained optimization

1 Introduction

In the Euclidean space \mathbb{R}^n , the *p*-norm of a vector $a \in \mathbb{R}^n$ whose *i*th element is $a_i \in \mathbb{R}$ is defined by

$$||a||_{p} := \sqrt[p]{\sum_{i=1}^{n} |a_{i}|^{p}},\tag{1}$$

where $p \ge 1$ is a real value. When $p = \infty$, the ∞ -norm, or maximum norm, is defined by

$$|a||_{\infty} := \max\left\{|a_1|, |a_2|, \dots, |a_n|\right\}.$$
(2)

In optimization and related fields, discussions are usually based on the 2-norm. The 1-norm is also important in, e.g., Lasso regression for sparse estimation [7]. Furthermore, for $x \in \mathbb{R}^n$, the constraint $||x||_{\infty} \leq c$ for some $c \geq 0$ is equivalent to the box constraint $-c \leq x_i \leq c$ for all elements x_i of x.

Funding: This work was funded by JSPS KAKENHI Grant number JP20K14359.

^{*}Department of Applied Mathematics and Physics, Kyoto University, Kyoto, Japan

⁽hsato@i.kyoto-u.ac.jp).

For $p \ge 1$ or $p = \infty$, we define the unit sphere with p-norm in \mathbb{R}^n as

$$S_p^{n-1} := \{ x \in \mathbb{R}^n \mid ||x||_p = 1 \}.$$
(3)

A particularly important and well-studied example is the case of p = 2, which reduces to the standard (hyper)sphere $S_2^{n-1} = \{x \in \mathbb{R}^n \mid ||x||_2 = 1\}$ in the sense of the Euclidean norm. In terms of optimization, as we discuss in Section 7, the case of p = 2p' can be used to implicitly impose the nonnegativity constraints on $x \in S_{p'}^{n-1}$. A practical example of this is the case of p = 4 and p' = 2, which leads to a constrained optimization on the standard unit sphere S_2^2 with the constraint $x \ge 0$. Furthermore, the case of p = 1 is closely related to L_1 regularization in, e.g., Lasso [7], and the case of $p = \infty$ is closely related to the box constraint.

In this paper, we address the geometry of S_p^{n-1} with $p \in (0, \infty)$ and provide several mathematical tools required for Riemannian optimization, i.e., optimization on Riemannian manifolds, such as retractions and vector transports [1,15]. A natural and practical retraction is defined through normalization in terms of *p*-norm, and we provide mathematical support for the validity of this retraction. Furthermore, we discuss projective and orthographic retractions on S_p^{n-1} . Although it may be harder to use such retractions practically than the retraction based on normalization, their inverses are efficient and easy to implement. Thus, discussing them is meaningful. We also provide an explicit expression for the vector transport defined as the differentiated retraction associated with the retraction by normalization. Other contributions of this paper include applications of the sphere S_p^{n-1} to practical optimization problems related to, e.g., the nonnegative principal component analysis (PCA) and Lasso regression.

This paper is organized as follows. In Section 2, we introduce the notations used. We also review the differentiability and derivative of the *p*-norm, which are used throughout this paper. In Section 3, we prove that S_p^{n-1} is a Riemannian submanifold of \mathbb{R}^n and, as such, investigate its geometry. Section 4 provides a retraction on S_p^{n-1} based on normalization and its inverse. The respective formulas for the inverses of projective and orthographic retractions are also provided. In Section 5, we discuss a vector transport on S_p^{n-1} derived by differentiating a retraction. We also remark another vector transport based on the orthogonal projection. Section 6 is a reference to the geometric results in this paper. We present two types of applications of Riemannian optimization on S_p^{n-1} in Section 7. One is the application to Riemannian optimization problems on the sphere with the nonnegative constraint, which include nonnegative PCA as an important example. The other is the application to L_p regularization-related optimization problems, which include the Lasso regression and boxconstrained problems. Section 8 concludes the paper.

2 Preliminaries

In this section, we provide preliminaries for the discussion in the later sections.

2.1 Notation

Throughout the paper, we use the following notation. The vector space of *n*-dimensional real column vectors is denoted by \mathbb{R}^n . We use the notation \cdot^T to indicate transposition. The *n*-dimensional real vector whose *i*th element is $a_i \in \mathbb{R}$ is denoted by $(a_i) \in \mathbb{R}^n$, and we denote the *i*th element of $b \in \mathbb{R}^n$ by b_i or $(b)_i$. For $a = (a_i) \in \mathbb{R}^n$, we denote the element-wise power of $r \in \mathbb{R}$ by $a^r := (a_i^r) \in \mathbb{R}^n$ and the element-wise absolute value by $|a| := (|a_i|) \in \mathbb{R}^n$.

Furthermore, the binary relation \leq (resp. \geq) for vectors $a = (a_i), b = (b_i) \in \mathbb{R}^n$ means the element-wise relation \leq (resp. \geq), i.e., $a \leq b$ (resp. $a \geq b$) is equivalent to $a_i \leq b_i$ (resp. $a_i \geq b_i$) for i = 1, 2, ..., n. In particular, $a \geq 0$ means that all elements of a are nonnegative. We define the all-one vector as $\mathbf{1} := (1, 1, ..., 1)^T \in \mathbb{R}^n$. Then, the condition $||x||_p = 1$ is equivalent to $||x||_p^p = 1$ and rewritten as $\mathbf{1}^T |x|^p = 1$. The identity matrix of nth order is denoted by I. For $\mathbf{1} \in \mathbb{R}^n$ and $I \in \mathbb{R}^{n \times n}$, the size n is determined by context.

We denote the sign function by sgn, i.e.,

$$\operatorname{sgn}(w) := \begin{cases} 1 & \text{if } w > 0, \\ 0 & \text{if } w = 0, \\ -1 & \text{if } w < 0 \end{cases}$$
(4)

for $w \in \mathbb{R}$. Note that $\operatorname{sgn}(w)|w| = w$ always holds. We also use the same notation for the element-wise application of sgn, i.e., for $a = (a_i) \in \mathbb{R}^n$, we define $\operatorname{sgn}(a) := (\operatorname{sgn}(a_i)) \in \mathbb{R}^n$.

The operator \odot denotes the Hadamard product, which is the element-wise product, i.e., for $a = (a_i), b = (b_i) \in \mathbb{R}^n$, we define $a \odot b := (a_i b_i) \in \mathbb{R}^n$. We consider the Hadamard product only for vectors in this paper. It is clear that the commutative law $a \odot b = b \odot a$ holds. Furthermore, for $c = (c_i) \in \mathbb{R}^n$, we have $a^T(b \odot c) = (a \odot b)^T c$ because both sides are equal to $\sum_{i=1}^n a_i b_i c_i$. Using these facts, we can rewrite the condition $||x||_p^p = 1$ as $x^T(\operatorname{sgn}(x) \odot |x|^{p-1}) = 1$ because we have

$$x^{T}(\operatorname{sgn}(x) \odot |x|^{p-1}) = (\operatorname{sgn}(x) \odot x)^{T} |x|^{p-1} = |x|^{T} |x|^{p-1} = \mathbf{1}^{T} |x|^{p} = ||x||_{p}^{p}.$$
 (5)

Although \mathbb{R}^n can be equipped with the *p*-norm to be a normed vector space, no inner product is associated with the *p*-norm unless n = 1 or p = 2. Therefore, we equip \mathbb{R}^n with the standard inner product $\langle a, b \rangle := a^T b$ and the induced norm $||a|| := \sqrt{\langle a, a \rangle} = ||a||_2$, which coincides with the 2-norm, even when we discuss the sphere S_p^{n-1} for general *p*. As discussed in Section 3, we regard \mathbb{R}^n as a Riemannian manifold with the Riemannian metric induced by the standard inner product and consider S_p^{n-1} for $p \in (1, \infty)$ as a Riemannian submanifold of \mathbb{R}^n .

For a manifold \mathcal{M} , we denote the tangent space of \mathcal{M} at $x \in \mathcal{M}$ by $T_x\mathcal{M}$. Furthermore, when the manifold \mathcal{M} is a Riemannian manifold with a Riemannian metric $\langle \cdot, \cdot \rangle_{x}$ each tangent space $T_x\mathcal{M}$ is endowed with the inner product $\langle \cdot, \cdot \rangle_{x}$ via the Riemannian metric $\langle \cdot, \cdot \rangle_{x}$, and the Riemannian gradient grad f(x) of a C^1 function $f: \mathcal{M} \to \mathbb{R}$ at x is defined as the unique tangent vector at x satisfying $Df(x)[\xi] = \langle \operatorname{grad} f(x), \xi \rangle_{x}$ for all $\xi \in T_x\mathcal{M}$, where $Df(x): T_x\mathcal{M} \to T_{f(x)}\mathbb{R} \simeq \mathbb{R}$ is the derivative of f at $x \in \mathcal{M}$. For \mathbb{R}^n as a Riemannian manifold with the Riemannian metric $\langle \xi, \eta \rangle_x := \xi^T \eta$ for any $x \in \mathbb{R}^n$ and $\xi, \eta \in T_x\mathbb{R}^n \simeq \mathbb{R}^n$, the Riemannian gradient of a function $\overline{f}: \mathbb{R}^n \to \mathbb{R}$ coincides with the standard Euclidean gradient $\nabla \overline{f}$, i.e., $\nabla \overline{f}(x) := (\partial \overline{f}(x)/\partial x_i) \in T_x\mathbb{R}^n \simeq \mathbb{R}^n$ for $x \in \mathbb{R}^n$.

2.2 Derivatives of *p*-norm functions

Here, we investigate the derivative or Euclidean gradient of the *p*-norm-related functions in \mathbb{R}^n . First, although the *p*-norm is defined for any $p \in [1, \infty]$, it is of class C^1 only for $p \in (1, \infty)$. In the remainder of this section, we assume $p \in (1, \infty)$. Then, it is easy to verify that

$$\frac{d|w|^p}{dw} = p\operatorname{sgn}(w)|w|^{p-1} \tag{6}$$

for $w \in \mathbb{R}$. Regarding the *p*-norm of $x \in \mathbb{R}^n$, because $||x||_p^p = \mathbf{1}^T |x|^p$, its partial derivative with respect to the variable x_i for $i \in \{1, 2, ..., n\}$ is

$$\frac{\partial \|x\|_p^p}{\partial x_i} = \frac{\partial |x_i|^p}{\partial x_i} = p \operatorname{sgn}(x_i) |x_i|^{p-1}.$$
(7)

Therefore, the gradient of the function $x \to ||x||_p^p$ is equal to

$$\nabla(x \mapsto ||x||_p^p)(x) = (p \operatorname{sgn}(x_i) |x_i|^{p-1}) = p \operatorname{sgn}(x) \odot |x|^{p-1}.$$
(8)

In the subsequent sections, we exploit the fact that the conditions $||x||_p = 1$ and $||x||_p^p = 1$ both of which characterize the unit sphere S_p^{n-1} —are equivalent to each other. Furthermore, $||x||_p^p$ usually seems to be easier to handle than $||x||_p$. For example, the gradient of the *p*-norm function is computed as

$$\nabla(x \mapsto ||x||_{p})(x) = \nabla\left(x \mapsto \left(||x||_{p}^{p}\right)^{\frac{1}{p}}\right)(x)$$

$$= \frac{1}{p}(||x||_{p}^{p})^{\frac{1}{p}-1} \cdot p \operatorname{sgn}(x) \odot |x|^{p-1}$$

$$= \frac{\operatorname{sgn}(x) \odot |x|^{p-1}}{||x||_{p}^{p-1}}.$$
(9)

We prefer to use (8), which provides a simpler expression, rather than (9), unless (9) is essential in the discussion.

Note that $h(x) := ||x||_p^p$ is not necessarily a C^{∞} function in \mathbb{R}^n . For example, consider the case p = 3 and n = 2, where $h(x) = |x_1|^3 + |x_2|^3$. Then, we have $\nabla h(x) = 3 \begin{pmatrix} |x_1|x_1 \\ |x_2|x_2 \end{pmatrix}$ and $\nabla^2 h(x) = 6 \begin{pmatrix} |x_1| & 0 \\ 0 & |x_2| \end{pmatrix}$. Hence, h is of class C^2 in \mathbb{R}^2 . However, since $\partial^2 h(x)/\partial x_1^2 = 6|x_1|$ (resp. $\partial^2 h(x)/\partial x_2^2 = 6|x_2|$) is not partially differentiable with respect to x_1 (resp. x_2) at any $(0, x_2)^T \in \mathbb{R}^2$ (resp. $(x_1, 0)^T$), h is not of class C^3 in \mathbb{R}^2 . This causes nonsmoothness of S_3^2 , which includes the points $(\pm 1, 0)^T$ and $(0, \pm 1)^T$, as a submanifold of \mathbb{R}^2 . In the next section, we will prove that S_p^{n-1} with $p \in (1, \infty)$ is still at least a C^1 submanifold of \mathbb{R}^n (Theorem 3.1).

3 Geometry of S_p^{n-1} and tools for Riemannian optimization

In this section, we discuss the geometry of the unit sphere with p-norm, i.e.,

$$S_p^{n-1} = \{ x \in \mathbb{R}^n \mid ||x||_p = 1 \},$$
(10)

where 1 . We use the following equivalent conditions interchangeably:

$$\|x\|_p = 1 \iff \|x\|_p^p = 1 \iff \mathbf{1}^T |x|^p = 1 \iff x^T (\operatorname{sgn}(x) \odot |x|^{p-1}) = 1.$$
(11)

As expected, many properties of the Euclidean sphere S_2^{n-1} analogically hold for S_p^{n-1} with any $p \in (1, \infty)$, especially even integer p, while some do not hold for S_1^{n-1} or S_{∞}^{n-1} .

3.1 S_n^{n-1} as a Riemannian submanifold of \mathbb{R}^n

First, we prove that S_p^{n-1} is an embedded submanifold of \mathbb{R}^n .

Theorem 3.1. For $p \in (1, \infty)$, the unit sphere S_p^{n-1} with p-norm is an (n-1)-dimensional C^r embedded submanifold of \mathbb{R}^n , where $r = \infty$ if p is an even integer, r = p - 1 if p is an odd integer, and $r = \lfloor p \rfloor$, which is the largest integer less than p, if p is not an integer.¹

Proof. We define $h: \mathbb{R}^n \to \mathbb{R}$ as $h(x) := ||x||_p^p$. We can observe that h is a C^r function in \mathbb{R}^n , where r is the integer in the statement of the theorem, as follows: If p is an even integer, $h(x) = \sum_{i=1}^n |x_i|^p = \sum_{i=1}^n x_i^p$ is clearly a C^∞ function. If p is an odd integer, $h(x) = \sum_{i=1}^n |x_i|^p$ is of class C^{p-1} because $\partial^{p-1}h(x)/\partial x_i^{p-1} = (p!)|x_i|$ is continuous for any $i \in \{1, 2, \ldots, n\}$. Similarly, if p is not an integer, h is of class $C^{|p|}$ because we have

$$\frac{\partial^{\lfloor p \rfloor} h}{\partial x_i^{\lfloor p \rfloor}}(x) = p(p-1)\cdots(p-\lfloor p \rfloor+1)\operatorname{sgn}(x)^{\lfloor p \rfloor}|x_i|^{p-\lfloor p \rfloor}$$
$$=\frac{\Gamma(p+1)}{\Gamma(p+1-\lfloor p \rfloor)}\operatorname{sgn}(x)^{\lfloor p \rfloor}|x_i|^{p-\lfloor p \rfloor},$$
(12)

which is continuous because $p - \lfloor p \rfloor > 0$ in this case, where $\Gamma(\cdot)$ is the gamma function. Therefore, h is of class C^r in every case.

Using the formula (8), the Jacobian matrix of h at $x \in \mathbb{R}^n - \{0\}$, which is defined as $(Jh)_x := (\partial h(x)/\partial x_i)^T \in \mathbb{R}^{1 \times n}$, is computed as

$$(Jh)_x = \nabla h(x)^T = p(\operatorname{sgn}(x) \odot |x|^{p-1})^T.$$
 (13)

For any $x \in \mathbb{R}^n$ satisfying h(x) = 1, we have $(Jh)_x \neq 0$ because such x is not 0. This implies that 1 is a regular value of h. Therefore, it follows from the regular level set theorem [19, Theorem 9.9] that $h^{-1}(\{1\}) = S_p^{n-1}$ is a C^r embedded submanifold of \mathbb{R}^n , whose dimension is $n - \dim \mathbb{R} = n - 1$. This completes the proof.

Remark 3.1. Note that the integer r in Theorem 3.1 is not less than 1 in every case. Therefore, S_p^{n-1} with $p \in (1, \infty)$ is always a C^1 submanifold of \mathbb{R}^n . In contrast, if p = 1 or $p = \infty$, the unit sphere S_p^{n-1} is not a C^1 embedded submanifold of \mathbb{R}^n because of their corners. Indeed, the above proof fails if p = 1 or $p = \infty$ because $x \mapsto ||x||_p$ is not a C^1 function in such cases.

In what follows, we assume $p \in (1, \infty)$ and define smoothness regarding S_p^{n-1} as C^r with $r \ge 1$ in Theorem 3.1. For example, we say that a function f on S_p^{n-1} is smooth if f is of class C^r .

We endow the sphere S_p^{n-1} with the Riemannian metric as

$$\langle \xi, \eta \rangle_x := \xi^T \eta, \qquad \xi, \ \eta \in T_x S_p^{n-1}, \quad x \in S_p^{n-1},$$
(14)

which is induced from the Riemannian metric (the standard inner product)

$$\langle a, b \rangle_x := a^T b, \qquad a, b \in T_x \mathbb{R}^n \simeq \mathbb{R}^n, \quad x \in \mathbb{R}^n$$

$$\tag{15}$$

in the ambient space \mathbb{R}^n . Thus, S_p^{n-1} is a Riemannian submanifold of \mathbb{R}^n .

¹The statement can be rewritten as follows: for any positive integer k, S_p^{n-1} is a C^{2k-1} submanifold of \mathbb{R}^n if 2k - 1 < n < 2k, C^{∞} submanifold if n = 2k, and C^{2k} submanifold if $2k < n \le 2k + 1$.

3.2 Tangent space, normal space, and orthogonal projection

Defining $h(x) := ||x||_p^p$, the tangent space $T_x S_p^{n-1}$ of $S_p^{n-1} = h^{-1}(\{1\})$ at x is equal to the kernel of the linear map $Dh(x) : \mathbb{R}^n \simeq T_x \mathbb{R}^n \to T_{h(x)} \mathbb{R} \simeq \mathbb{R}$, i.e., $(Dh(x))^{-1}(\{0\})$. Here, it follows from (13) that the derivative Dh(x) acts on $y \in \mathbb{R}^n$ as

$$Dh(x)[y] = (Jh)_x(y) = p(sgn(x) \odot |x|^{p-1})^T y.$$
(16)

Therefore, we have

$$T_x S_p^{n-1} = (\mathrm{D}h(x))^{-1}(\{0\}) = \{\xi \in \mathbb{R}^n \mid \xi^T(\mathrm{sgn}(x) \odot |x|^{p-1}) = 0\}.$$
 (17)

Since S_p^{n-1} is a Riemannian submanifold of \mathbb{R}^n , we can define the normal space $N_x S_p^{n-1}$ of S_p^{n-1} at a point x as the orthogonal complement of $T_x S_p^{n-1} \subset T_x \mathbb{R}^n \simeq \mathbb{R}^n$ in \mathbb{R}^n with respect to the Riemannian metric in \mathbb{R}^n , i.e., the standard inner product. From the expression (17), we can observe that $T_x S_p^{n-1}$ is a hyperplane orthogonal to the vector $\operatorname{sgn}(x) \odot |x|^{p-1} \in \mathbb{R}^n$. Hence, we have

$$N_x S_p^{n-1} := (T_x S_p^{n-1})^{\perp} = \{ \alpha \operatorname{sgn}(x) \odot |x|^{p-1} \mid \alpha \in \mathbb{R} \}.$$
(18)

For minimizing a smooth function $f: S_p^{n-1} \to \mathbb{R}$ on S_p^{n-1} , the Riemannian gradient of f is important. Here, the Riemannian gradient grad f(x) of f at $x \in S_p^{n-1}$ can be obtained by orthogonally projecting $\nabla \bar{f}(x) \in \mathbb{R}^n$ onto the tangent space $T_x S_p^{n-1}$ at x, where \bar{f} is a smooth extension of f to the ambient space \mathbb{R}^n and $\nabla \bar{f}(x) := (\partial \bar{f}(x)/\partial x_i) \in \mathbb{R}^n$ is the Euclidean gradient. That is, we have

$$\operatorname{grad} f(x) = P_x(\nabla f(x)),\tag{19}$$

where P_x is the orthogonal projection to the tangent space $T_x S_p^{n-1}$ at x. The projection $P_x \colon \mathbb{R}^n \to T_x S_p^{n-1}$ acts on any $d \in \mathbb{R}^n$ so that $d - P_x(d) \in N_x S_p^{n-1}$ holds. From (18), the normal vector $d - P_x(d)$ is written as $\alpha \operatorname{sgn}(x) \odot |x|^{p-1}$ for some $\alpha \in \mathbb{R}$. Thus, we obtain the decomposition of d as

$$d = P_x(d) + \alpha \operatorname{sgn}(x) \odot |x|^{p-1}.$$
(20)

By noting the expression (17) and multiplying (20) by $(\operatorname{sgn}(x) \odot |x|^{p-1})^T$ from the left, we obtain $\alpha = ((\operatorname{sgn}(x) \odot |x|^{p-1})^T d) / ||x|^{p-1}||_2^2$, where we used the relation

$$(\operatorname{sgn}(x) \odot |x|^{p-1})^T (\operatorname{sgn}(x) \odot |x|^{p-1}) = ((\operatorname{sgn}(x))^2)^T (|x|^{p-1})^2 = ||x|^{p-1}||_2^2 \neq 0.$$
(21)

Substituting the expression of α to (20), we obtain

$$P_{x}(d) = d - \frac{(\operatorname{sgn}(x) \odot |x|^{p-1})^{T} d}{\||x|^{p-1}\|_{2}^{2}} \operatorname{sgn}(x) \odot |x|^{p-1} = \left(I - \frac{(\operatorname{sgn}(x) \odot |x|^{p-1})(\operatorname{sgn}(x) \odot |x|^{p-1})^{T}}{\||x|^{p-1}\|_{2}^{2}}\right) d.$$
(22)

In other words, the linear map P_x is represented as the matrix

$$P_x = I - \frac{(\operatorname{sgn}(x) \odot |x|^{p-1})(\operatorname{sgn}(x) \odot |x|^{p-1})^T}{\||x|^{p-1}\|_2^2}.$$
(23)

4 Retractions and their inverses

In an iterative Riemannian optimization algorithm on a Riemannian manifold \mathcal{M} , to compute the next point from the current point $x \in \mathcal{M}$ and search direction $\eta \in T_x \mathcal{M}$, a retraction on \mathcal{M} is important [1,3,18]. A map $R: T\mathcal{M} \to \mathcal{M}$ is said to be a retraction on \mathcal{M} if the restriction $R_x := R|_{T_x\mathcal{M}}$ of R to $T_x\mathcal{M}$ for $x \in \mathcal{M}$ satisfies $R_x(0_x) = x$ and $DR_x(0_x) = \mathrm{id}_{T_x\mathcal{M}}$, where 0_x is the zero vector in $T_x\mathcal{M}$ and $\mathrm{id}_{T_x\mathcal{M}}$ is the identity map in $T_x\mathcal{M}$. Although retractions are usually discussed on C^{∞} manifolds, the manifold S_p^{n-1} is a C^r submanifold of \mathbb{R}^n , where r is in Theorem 3.1 and may not be ∞ . Therefore, we define a retraction on S_p^{n-1} as a C^r , which we say smooth, map on S_p^{n-1} satisfying the above properties.

Furthermore, the inverse of a retraction can be used in, e.g., the Riemannian conjugate gradient method [22]. In the following, we discuss three types of retractions on S_p^{n-1} and their respective inverses.

4.1 Retraction by normalization and its inverse

Intuitively, for any $x \in S_p^{n-1}$ and $\eta \in T_x S_p^{n-1}$, $x + \eta \in \mathbb{R}^n$ appears to be outside S_p^{n-1} unless $\eta = 0$. This is actually true from the following proposition. However, its proof for general p > 1 is not as easy as in the case of p = 2.

Proposition 4.1. Assume that $p \in (1, \infty)$. For any $x \in S_p^{n-1}$ and $\eta \in T_x S_p^{n-1}$, if $\eta \neq 0$, then $||x + \eta||_p > 1$ holds.

Proof. Note that the function $h(y) := ||y||_p^p$ is not of class C^2 in the entire \mathbb{R}^n when 1 . Therefore, we avoid using the Hessian matrix in the following discussion to address the general case.

We first show that h is a strictly convex function in \mathbb{R}^n . For $y, z \in \mathbb{R}^n$ with $y \neq z$ and $\alpha \in (0,1)$, Minkowski's inequality (the triangle inequality for the p-norm) as well as convexity and monotonicity of the function $w \mapsto w^p$ on $\mathbb{R}_+ := \{w \in \mathbb{R} \mid w \ge 0\}$ yield that

$$\|\alpha y + (1-\alpha)z\|_p^p \le (\alpha \|y\|_p + (1-\alpha)\|z\|_p)^p \le \alpha \|y\|_p^p + (1-\alpha)\|z\|_p^p.$$
(24)

We now assume that both equalities in (24) simultaneously hold. Then, the first equality implies that y = cz for some $c \ge 0$ or z = 0 from Minkowski's inequality theory for $p \in$ $(1,\infty)$. Furthermore, from the second equality and the strict convexity of $w \mapsto w^p$ on \mathbb{R}_+ , we have $\|y\|_p = \|z\|_p$. If y = cz with $c \ge 0$, then $\|y\|_p = \|z\|_p$ implies c = 1 or $\|y\|_p = \|z\|_p = 0$. Otherwise, we have z = 0; and $\|y\|_p = \|z\|_p$ then means y = z = 0. In any case, we have y = z, which contradicts the assumption that $y \ne z$. Therefore, both equalities in (24) do not hold at the same time, meaning

$$h(\alpha y + (1 - \alpha)z) = \|\alpha y + (1 - \alpha)z\|_p^p < \alpha \|y\|_p^p + (1 - \alpha)\|z\|_p^p = \alpha h(y) + (1 - \alpha)h(z).$$
(25)

This proves that h is strictly convex.

By using the strict convexity of h, we can show that $\phi(t) := h(x + t\eta) = ||x + t\eta||_p^p$ is a strictly convex function on \mathbb{R} for $x \in S_p^{n-1}$ and $\eta \in T_x S_p^{n-1}$ with $\eta \neq 0$. Indeed, for any $s, t \in \mathbb{R}$ with $s \neq t$ and $\alpha \in (0, 1)$, it follows from the strict convexity of h and the fact $x + s\eta \neq x + t\eta$ that

$$\phi(\alpha s + (1 - \alpha)t) = h(x + (\alpha s + (1 - \alpha)t)\eta)$$

= $h(\alpha(x + s\eta) + (1 - \alpha)(x + t\eta))$
< $\alpha h(x + s\eta) + (1 - \alpha)h(x + t\eta)$
= $\alpha \phi(s) + (1 - \alpha)\phi(t).$ (26)

Subsequently, we show that t = 0 is the unique minimizer of ϕ . Since ϕ is strictly convex, it suffices to prove that $\phi'(0) = 0$, which is shown as

$$\phi'(0) = \nabla h(x)^T \eta = p(\operatorname{sgn}(x) \odot |x|^{p-1})^T \eta = 0$$
(27)

from (8) and (17).

In conclusion, we obtain $||x + t\eta||_p^p = \phi(t) > \phi(0) = ||x||_p^p = 1$ for all $t \neq 0$, where the case of t = 1 implies that the desired inequality $||x + \eta||_p > 1$ holds.

Considering Proposition 4.1, we propose a retraction R on S_p^{n-1} as

$$R_x(\eta) := \frac{x+\eta}{\|x+\eta\|_p}, \qquad \eta \in T_x S_p^{n-1}, \quad x \in S_p^{n-1}.$$
(28)

This is simply the normalization (with respect to the *p*-norm) of $x + \eta$, which is not on S_p^{n-1} when $\eta \neq 0$. Note that the denominator in (28) is ensured to be nonzero from Proposition 4.1.

Proposition 4.2. Assume that $p \in (1, \infty)$. The map R defined by (28) is a retraction on S_p^{n-1} .

Proof. It is clear that $||R_x(\eta)||_p = 1$ and $R_x(0_x) = x$ hold for any $x \in S_p^{n-1}$ and $\eta \in T_x S_p^{n-1}$. To prove that $DR_x(0_x) = \operatorname{id}_{T_x S_p^{n-1}}$ holds, we use (9), i.e., the fact that the gradient of $x \mapsto ||x||_p$ is written as $||x||_p^{1-p} \operatorname{sgn}(x) \odot |x|^{p-1}$. Then, we can compute $DR_x(0_x)[\eta]$ for $\eta \in T_x S_p^{n-1}$ as

$$DR_{x}(0_{x})[\eta] = \frac{d}{dt}R_{x}(t\eta)\Big|_{t=0}$$

$$= \frac{\eta \|x + t\eta\|_{p} - (x + t\eta)(\|x + t\eta\|_{p}^{1-p}\operatorname{sgn}(x + t\eta) \odot |x + t\eta|^{p-1})^{T}\eta}{\|x + t\eta\|_{p}^{2}}\Big|_{t=0}$$

$$= \eta - x(\operatorname{sgn}(x) \odot |x|^{p-1})^{T}\eta = \eta, \qquad (29)$$

where we used $||x||_p = 1$ and $(sgn(x) \odot |x|^{p-1})^T \eta = 0$ from (17).

To derive the inverse of R, we fix $x, y \in S_p^{n-1}$ and assume that $\eta \in T_x S_p^{n-1}$ satisfies $R_x(\eta) = y$. Then, η should satisfy $x + \eta = \alpha y$ for some $\alpha > 0$. Multiplying the equality by $(\operatorname{sgn}(x) \odot |x|^{p-1})^T$ from the left and noting that $x \in S_p^{n-1}$ and $\eta \in T_x S_p^{n-1}$, we obtain $\alpha = 1/((\operatorname{sgn}(x) \odot |x|^{p-1})^T y)$. Therefore, η should satisfy

$$\eta = \alpha y - x = \frac{y}{(\operatorname{sgn}(x) \odot |x|^{p-1})^T y} - x.$$
(30)

However, this is necessary but not sufficient for $R_x(\eta) = y$. In fact, for certain $x, y \in S_p^{n-1}$, there may not exist η such that $R_x(\eta) = y$. The following proposition elaborates on this issue.

Proposition 4.3. Assume that $p \in (1, \infty)$. For any $x \in S_p^{n-1}$, the inverse of R_x defined in (28) is given by

$$R_x^{-1}(y) = \frac{y}{(\operatorname{sgn}(x) \odot |x|^{p-1})^T y} - x, \qquad y \in D_x$$
(31)

where the domain D_x of R_x^{-1} is $D_x = \{y \in S_p^{n-1} \mid (\text{sgn}(x) \odot |x|^{p-1})^T y > 0\}.$

Proof. For y satisfying $(\operatorname{sgn}(x) \odot |x|^{p-1})^T y = 0$, the right-hand side of (31) is not defined. We assume that $(\operatorname{sgn}(x) \odot |x|^{p-1})^T y \neq 0$ and denote the right-hand side of (31) by $\eta_{x,y}$, which is in $T_x S_p^{n-1}$ because $(\operatorname{sgn}(x) \odot |x|^{p-1})^T \eta_{x,y} = 1 - 1 = 0$. Then, we have

$$R_{x}(\eta_{x,y}) = \frac{x + \eta_{x,y}}{\|x + \eta_{x,y}\|_{p}} = \frac{y}{\operatorname{sgn}((\operatorname{sgn}(x) \odot |x|^{p-1})^{T}y)} \\ = \begin{cases} y & \text{if } (\operatorname{sgn}(x) \odot |x|^{p-1})^{T}y > 0 \\ -y & \text{if } (\operatorname{sgn}(x) \odot |x|^{p-1})^{T}y < 0 \end{cases}$$
(32)

Furthermore, if $\eta \in T_x S_p^{n-1}$ satisfies $R_x(\eta) = y$, then η should be equal to $\eta_{x,y}$, as discussed in (30). Therefore, $R_x(\eta) = y$ holds if and only if $(\operatorname{sgn}(x) \odot |x|^{p-1})^T y > 0$ and $\eta = \eta_{x,y}$. This completes the proof.

4.2 Inverse of projective retraction

Another natural retraction is the projective retraction [2]. The projective retraction R^{proj} on S_p^{n-1} is given by

$$R_x^{\text{proj}}(\eta) = \underset{y \in S_p^{n-1}}{\operatorname{min}} \| (x+\eta) - y \|_2, \qquad \eta \in T_x S_p^{n-1}, \quad x \in S_p^{n-1}.$$
(33)

Remark 4.1. Note that the projection onto any closed convex set in \mathbb{R}^n regarding the 2-norm is unique [5, Section 8.1]. Therefore, because the unit ball $B_p^n := \{x \in \mathbb{R}^n \mid ||x||_p \leq 1\}$ with *p*norm is obviously a closed convex set in \mathbb{R}^n , vector $y \in B_p^n$ that minimizes the distance $||(x+\eta)-y||_2$ uniquely exists for a given $x \in S_p^{n-1}$ and $\eta \in T_x S_p^{n-1}$. Since $x+\eta$ is outside B_p^{n-1} unless $\eta = 0$ from Proposition 4.1, the right-hand side in (33) is equal to the uniquely existing projection of $x + \eta$ onto B_p^n (clearly, we have $R_x^{\text{proj}}(\eta) = x$ when $\eta = 0$).

The vector $R_x^{\text{proj}}(\eta)$ satisfies $(x + \eta) - R_x^{\text{proj}}(\eta) \in N_x S_p^{n-1}$, which is implied by [2] or is a direct consequence of the Lagrange multiplier method. Therefore, there exists $\alpha \in \mathbb{R}$ such that

$$R_x^{\text{proj}}(\eta) = x + \eta - \alpha \operatorname{sgn}(R_x^{\text{proj}}(\eta)) \odot |R_x^{\text{proj}}(\eta)|^{p-1},$$
(34)

where α is determined such that $R_x^{\text{proj}}(\eta) \in S_p^{n-1}$ holds, i.e.,

$$||x + \eta - \alpha \operatorname{sgn}(R_x^{\operatorname{proj}}(\eta)) \odot |R_x^{\operatorname{proj}}(\eta)|^{p-1}||_p = 1.$$
(35)

However, it may be difficult to explicitly express $R_x^{\text{proj}}(\eta)$ by solving (34) and (35).

Remark 4.2. When p = 2, Eq. (34) is reduced to $R_x^{\text{proj}}(\eta) = x + \eta - \alpha R_x^{\text{proj}}(\eta)$, i.e., we have $(\alpha + 1)R_x^{\text{proj}}(\eta) = (x + \eta)$. Then, $||R_x^{\text{proj}}(\eta)||_2 = 1$ implies $|\alpha + 1| = ||x + \eta||_2$. Hence, we obtain $(x + \eta)/(\alpha + 1) = \pm (x + \eta)/||x + \eta||_2$, among which $(x + \eta)/||x + \eta||_2$ is closer to $x + \eta$. In summary, when p = 2, we have $R_x^{\text{proj}}(\eta) = (x + \eta)/||x + \eta||_2$, which is equal to the retraction by normalization in Section 4.1.

Although the above discussion implies that the projective retraction on S_p^{n-1} for general $p \in (1, \infty)$ may not provide as successful a result as the retraction by normalization, the inverse of R_x^{proj} can be discussed more practically. For given $x, y \in S_p^{n-1}$, if $\eta \in T_x S_p^{n-1}$ satisfies $R_x^{\text{proj}}(\eta) = y$, then $x + \eta - y \in N_y S_p^{n-1}$ should hold. Therefore, there exists $\alpha_{x,y} \in \mathbb{R}$ such that $x + \eta - y = \alpha_{x,y} \operatorname{sgn}(y) \odot |y|^{p-1}$. From $x \in S_p^{n-1}$ and $\eta \in T_x S_p^{n-1}$, we can obtain an explicit expression for $\alpha_{x,y}$ as in Proposition 4.4. Furthermore, it seems that $\alpha_{x,y}$ should be nonnegative by analogy with the discussion of the case p = 2 in Remark 4.2. We discuss these rigorously in the proof of the proposition using the Karush–Kuhn–Tucker (KKT) conditions.

Proposition 4.4. Assume that $p \in (1, \infty)$. For any $x \in S_p^{n-1}$, the inverse of R_x^{proj} in (33) is given by

$$(R_x^{\text{proj}})^{-1}(y) = y - x + \alpha_{x,y} \operatorname{sgn}(y) \odot |y|^{p-1} = \left(I - \frac{(\operatorname{sgn}(y) \odot |y|^{p-1})(\operatorname{sgn}(x) \odot |x|^{p-1})^T}{(\operatorname{sgn}(y) \odot |y|^{p-1})^T (\operatorname{sgn}(x) \odot |x|^{p-1})}\right)(y - x), \qquad y \in D_x,$$
(36)

where

$$\alpha_{x,y} := \frac{1 - (\operatorname{sgn}(x) \odot |x|^{p-1})^T y}{(\operatorname{sgn}(x) \odot |x|^{p-1})^T (\operatorname{sgn}(y) \odot |y|^{p-1})}$$
(37)

and the domain of $(R_x^{\text{proj}})^{-1}$ is

$$D_x = \{ y \in S_p^{n-1} \mid (\operatorname{sgn}(x) \odot |x|^{p-1})^T (\operatorname{sgn}(y) \odot |y|^{p-1}) \neq 0, \ \alpha_{x,y} \ge 0 \}.$$
(38)

Proof. The second equality in (36) directly follows from $(\operatorname{sgn}(x) \odot |x|^{p-1})^T x = 1$. We define $\eta_{x,y} := y - x + \alpha_{x,y} \operatorname{sgn}(y) \odot |y|^{p-1}$ with $\alpha_{x,y}$ in (37). Then, what we need to prove is that $R_x^{\operatorname{proj}}(\eta) = y$ holds for $\eta \in T_x S_p^{n-1}$ if and only if y belongs to the right-hand side of (38) and $\eta = \eta_{x,y}$.

To see this in light of (33) and Remark 4.1, we must verify that z = y is the optimal solution to the following optimization problem with a fixed $\eta \in T_x S_p^{n-1}$ if and only if $\alpha_{x,y}$ in (37) is well-defined and nonnegative and $\eta = \eta_{x,y}$:

minimize
$$\|(x+\eta) - z\|_2^2$$

subject to $\|z\|_p^p \le 1, \ z \in \mathbb{R}^n,$

where the decision variable vector is z. This is a convex optimization problem because both $z \mapsto ||(x + \eta) - z||_2^2$ and $z \mapsto ||z||_p^p - 1$ are convex. Furthermore, the problem satisfies Slater's condition [5, Section 5.2.3], i.e., it is strictly feasible (e.g., with z = 0). Therefore, the condition that z = y is optimal for the optimization problem is equivalent to saying that there exists $\lambda \in \mathbb{R}$ such that z = y and λ satisfy the KKT conditions for the problem, which are written as

$$2(y - (x + \eta)) + \lambda p \operatorname{sgn}(y) \odot |y|^{p-1} = 0,$$
(39)

$$\|y\|_p^p \le 1,\tag{40}$$

$$\lambda \ge 0,\tag{41}$$

$$\lambda(\|y\|_p^p - 1) = 0. \tag{42}$$

Since $||y||_p = 1$, they are equivalent to

$$\eta = y - x + \frac{p}{2}\lambda\operatorname{sgn}(y) \odot |y|^{p-1},\tag{43}$$

$$\lambda \ge 0. \tag{44}$$

Noting that $x \in S_p^{n-1}$ and $\eta \in T_x S_p^{n-1}$, we multiply (43) by $(\operatorname{sgn}(x) \odot |x|^{p-1})^T$ from the left to obtain

$$2(1 - (\operatorname{sgn}(x) \odot |x|^{p-1})^T y) = p\lambda(\operatorname{sgn}(x) \odot |x|^{p-1})^T(\operatorname{sgn}(y) \odot |y|^{p-1}).$$
(45)

If $(\operatorname{sgn}(x) \odot |x|^{p-1})^T (\operatorname{sgn}(y) \odot |y|^{p-1}) = 0$ holds, $1 - (\operatorname{sgn}(x) \odot |x|^{p-1})^T y = 0$ should hold, and λ can be any value. However, it then follows from (43) that y = 0, contradicting $y \in S_p^{n-1}$. Hence, we have $(\operatorname{sgn}(x) \odot |x|^{p-1})^T (\operatorname{sgn}(y) \odot |y|^{p-1}) \neq 0$, and λ is written as

$$\lambda = \frac{2}{p} \frac{1 - (\operatorname{sgn}(x) \odot |x|^{p-1})^T y}{(\operatorname{sgn}(x) \odot |x|^{p-1})^T (\operatorname{sgn}(y) \odot |y|^{p-1})} = \frac{2}{p} \alpha_{x,y}.$$
(46)

Therefore, there exists $\lambda \in \mathbb{R}$ such that z = y and λ satisfy the KKT conditions (43) and (44) if and only if $\eta = y - x + \alpha_{x,y} \operatorname{sgn}(y) \odot |y|^{p-1} = \eta_{x,y}$ and $\alpha_{x,y}$ is well-defined and nonnegative. This completes the proof.

4.3 Inverse of orthographic retraction

Other possibilities of retractions on S_p^{n-1} include the orthographic retraction. See [2] for a discussion of orthographic retractions on general Riemannian submanifolds.

For $x \in S_p^{n-1}$ and $\eta \in T_x S_p^{n-1}$, the orthographic retraction R^{orth} is defined to satisfy $R_x^{\text{orth}}(\eta) = x + \eta + \zeta \in S_p^{n-1}$ for some $\zeta \in N_x S_p^{n-1}$ with the smallest norm among all normal vectors in $\{\xi \in N_x S_p^{n-1} \mid x + \eta + \xi \in S_p^{n-1}\}$. Since we can express $\zeta \in N_x S_p^{n-1}$ as $\zeta = -\alpha \operatorname{sgn}(x) \odot |x|^{p-1}$ for some $\alpha \in \mathbb{R}$, the relation $||R_x^{\text{orth}}(\eta)||_p = 1$ yields the equation on α as

$$\|x + \eta - \alpha \operatorname{sgn}(x) \odot |x|^{p-1}\|_p^p = 1.$$
(47)

Remark 4.3. When p = 2, Eq. (47) is reduced to $(1 - \alpha)^2 + \eta^T \eta = 1$, the smaller solution (with smaller absolute value) of which is given by $\alpha = 1 - \sqrt{1 - \eta^T \eta}$ if $\|\eta\|_2 \le 1$. This gives the expression $R_x^{\text{orth}}(\eta) = \sqrt{1 - \eta^T \eta} x + \eta$, which is a well-known result.

For general $p \in (1, \infty)$, we have

$$R_x^{\text{orth}}(\eta) = x + \eta - \alpha \operatorname{sgn}(x) \odot |x|^{p-1}, \qquad \eta \in T_x S_p^{n-1}, \quad x \in S_p^{n-1},$$
 (48)

where η should be a tangent vector such that Eq. (47) has a solution and α is the one with the smallest absolute value of the solutions. Unfortunately, as in the projective retraction in Section 4.2, it may be difficult to explicitly express such α for general p.

However, the discussion on this retraction is still important because its inverse can be practically computed. Subsequently, we assume that $\eta \in T_x S_p^{n-1}$ satisfies $R_x^{\text{orth}}(\eta) = y$ for given $x, y \in S_p^{n-1}$. Then, there exists $\alpha_{x,y} \in \mathbb{R}$ such that $x + \eta - \alpha_{x,y} \operatorname{sgn}(x) \odot |x|^{p-1} = y$. Since $\eta \in T_x S_p^{n-1}$, multiplying both sides by $(\operatorname{sgn}(x) \odot |x|^{p-1})^T$ from the left yields

$$\alpha_{x,y} = \frac{1 - (\operatorname{sgn}(x) \odot |x|^{p-1})^T y}{(\operatorname{sgn}(x) \odot |x|^{p-1})^T (\operatorname{sgn}(x) \odot |x|^{p-1})} = \frac{1 - (\operatorname{sgn}(x) \odot |x|^{p-1})^T y}{\||x|^{p-1}\|_2^2}.$$
(49)

Note that the denominator is nonzero because of $x \neq 0$. This observation, together with the discussion on when (49) is sufficient for $R_x^{\text{orth}}(\eta) = y$, leads to the following proposition.

Proposition 4.5. Assume that $p \in (1, \infty)$. For any $x \in S_p^{n-1}$, the inverse of the retraction R_x^{orth} is given by

$$(R_x^{\text{orth}})^{-1}(y) = y - x + \alpha_{x,y} \operatorname{sgn}(x) \odot |x|^{p-1}, \qquad y \in D_x,$$
 (50)

where

$$\alpha_{x,y} := \frac{1 - (\operatorname{sgn}(x) \odot |x|^{p-1})^T y}{\||x|^{p-1}\|_2^2}$$
(51)

and the domain of $(R_x^{\text{orth}})^{-1}$ is

$$D_x = \{ y \in S_p^{n-1} \mid \alpha_{x,y} \text{ is the solution to } (47) \text{ with the smallest absolute value} \}.$$
(52)

Proof. Let $\eta_{x,y} := y - x + \alpha_{x,y} \operatorname{sgn}(x) \odot |x|^{p-1}$ be the right-hand side of (50) with $\alpha_{x,y}$ in (51). From the above discussion on (49), for a given $x \in S_p^{n-1}$, if $\eta \in T_x S_p^{n-1}$ satisfies $R_x^{\operatorname{orth}}(\eta) = y$, then y should belong to the right-hand side of (52) and $\eta = \eta_{x,y}$ should hold.

To prove the converse, we show that $R_x^{\text{orth}}(\eta) = y$ holds if y belongs to the right-hand side of (52) and $\eta = \eta_{x,y}$ holds. Assume that y and η be such vectors, i.e., $\alpha = \alpha_{x,y}$ is the solution to (47) with the smallest absolute value and $\eta = \eta_{x,y}$. Then, from the definition of the orthographic retraction and the expression of $\eta_{x,y}$, we have

$$R_x^{\text{orth}}(\eta) = R_x^{\text{orth}}(\eta_{x,y}) = x + \eta_{x,y} - \alpha_{x,y} \operatorname{sgn}(x) \odot |x|^{p-1} = y.$$
(53)

This completes the proof.

4.4 Discussion on exponential retraction

On a general Riemannian manifold, another important retraction is the exponential retraction R := Exp, where Exp is the exponential map. However, it may be difficult to use practically. Here, we discuss this issue. In the following discussion, we assume that $p \ge 2$, which ensures that S_p^{n-1} is a C^2 submanifold of \mathbb{R}^n from Theorem 3.1.

The exponential map Exp is defined as

$$\operatorname{Exp}_{x}(\eta) := \gamma_{x,\eta}(1), \qquad \eta \in T_{x}S_{p}^{n-1}, \quad x \in S_{p}^{n-1}, \tag{54}$$

where $\gamma_{x,\eta}$ is the geodesic on S_p^{n-1} emanating from x in the direction of η . The geodesic satisfies the geodesic equation, which is derived from the condition $\ddot{\gamma}_{x,\eta}(t) \in N_{\gamma_{x,\eta}(t)}S_p^{n-1}$. For simplicity, we denote $\gamma_{x,\eta}(t)$ by x(t). Then, $x(t) \in S_p^{n-1}$ implies $\mathbf{1}^T |x(t)|^p = 1$. Differentiating both sides, we obtain $(\operatorname{sgn}(x(t)) \odot |x(t)|^{p-1})^T \dot{x}(t) = 0$. We further differentiate both sides to get

$$(p-1)(|x(t)|^{p-2} \odot \dot{x}(t))^T \dot{x}(t) + (\operatorname{sgn}(x(t)) \odot |x(t)|^{p-1})^T \ddot{x}(t) = 0.$$
(55)

From $\ddot{x}(t) \in N_{x(t)}S_p^{n-1}$, there exists $\alpha(t) \in \mathbb{R}$ such that $\ddot{x}(t) = \alpha(t)\operatorname{sgn}(x(t)) \odot |x(t)|^{p-1}$. Substituting this into (55), we obtain $\alpha(t) = -(p-1)((|x(t)|^{p-2})^T \dot{x}(t)^2)/||x(t)|^{p-1}||_2^2$. Therefore, x(t) satisfies the geodesic equation

$$\ddot{x}(t) + \frac{(p-1)(|x(t)|^{p-2})^T \dot{x}(t)^2}{\||x(t)|^{p-1}\|_2^2} \operatorname{sgn}(x(t)) \odot |x(t)|^{p-1} = 0.$$
(56)

Solving this equation for the case $p \neq 2$ may be difficult. Thus, this will be dealt in a future work.

Remark 4.4. When p = 2, Eq. (56) is reduced to $\ddot{x}(t) + (\dot{x}(t)^T \dot{x}(t))x(t) = 0$, whose solution is $x(t) = x \cos(\|\eta\|_2 t) + (\eta/\|\eta\|_2) \sin(\|\eta\|_2 t)$, where x(0) = x and $\dot{x}(0) = \eta$, as shown in [1, Example 5.4.1].

5 Vector transports

In addition to a retraction, a vector transport is also an important geometric tool in Riemannian optimization methods, e.g., Riemannian conjugate gradient methods [1, 12–14, 16], Riemannian quasi-Newton methods [8, 9], and Riemannian stochastic optimization methods [17, 21]. Let \mathcal{M} be a Riemannian manifold and $T\mathcal{M} \oplus T\mathcal{M} := \{(\eta, \xi) \mid \eta, \xi \in T_x\mathcal{M}, x \in \mathcal{M}\}$ be the Whitney sum. A map $\mathcal{T}: T\mathcal{M} \oplus T\mathcal{M} \to \mathcal{M}$ is called a vector transport on \mathcal{M} if there exists a retraction R on \mathcal{M} and the following conditions are satisfied for any $x \in \mathcal{M}$: (i) $\mathcal{T}_{\eta}(\xi) \in T_{R_x(\eta)}\mathcal{M}$ for any $\eta, \xi \in T_x\mathcal{M}$; (ii) $\mathcal{T}_{0_x} = \mathrm{id}_{T_x\mathcal{M}}$; (iii) \mathcal{T}_{η} is a linear transformation in $T_x\mathcal{M}$ for any $\eta \in T_x\mathcal{M}$.

5.1 Differentiated retraction

An important vector transport is the differentiated retraction \mathcal{T}^R [1, Section 8.1.2] associated with a retraction R on S_p^{n-1} defined by

$$\mathcal{T}^R_\eta(\xi) := \mathrm{D}R_x(\eta)[\xi], \qquad \eta, \, \xi \in T_x S^{n-1}_p, \quad x \in S^{n-1}_p.$$
(57)

The differentiated retraction appears in the Riemannian (strong) Wolfe conditions and is thus used for line search in various algorithms.

Here, we derive the expression of \mathcal{T}^R with the retraction R defined in (28). Noting (9), an analogous computation to (29) gives

$$\mathcal{T}_{\eta}^{R}(\xi) = \mathrm{D}R_{x}(\eta)[\xi] = \frac{d}{dt}R_{x}(\eta + t\xi)\Big|_{t=0}$$

$$= \frac{\xi \|x + \eta\|_{p} - (x + \eta)\left(\|x + \eta\|_{p}^{1-p}\operatorname{sgn}(x + \eta) \odot |x + \eta|^{p-1}\right)^{T}\xi}{\|x + \eta\|_{p}^{2}}$$

$$= \frac{\xi}{\|x + \eta\|_{p}} - \frac{\left(\operatorname{sgn}(x + \eta) \odot |x + \eta|^{p-1}\right)^{T}\xi}{\|x + \eta\|_{p}^{p+1}}(x + \eta).$$
(58)

5.2 Vector transport based on orthogonal projection

Since S_p^{n-1} is a Riemannian submanifold of \mathbb{R}^n , another vector transport \mathcal{T}^P on S_p^{n-1} is defined by the orthogonal projection [1, Section 8.1.3] as

$$\mathcal{T}^{P}_{\eta}(\xi) := P_{R_{x}(\eta)}(\xi), \qquad \eta, \, \xi \in T_{x} S_{p}^{n-1}, \quad x \in S_{p}^{n-1}, \tag{59}$$

where the orthogonal projection P is provided by (23). Specifically, if we use the retraction (28), we have

$$\mathcal{T}_{\eta}^{P}(\xi) = \left(I - \frac{(\operatorname{sgn}(R_{x}(\eta)) \odot |R_{x}(\eta)|^{p-1})(\operatorname{sgn}(R_{x}(\eta)) \odot |R_{x}(\eta)|^{p-1})^{T}}{\||R_{x}(\eta)|^{p-1}\|_{2}^{2}}\right)\xi$$
(60)

$$=\xi - \frac{(\operatorname{sgn}(x+\eta) \odot |x+\eta|^{p-1})^T \xi}{\||x+\eta|^{p-1}\|_2^2} \operatorname{sgn}(x+\eta) \odot |x+\eta|^{p-1}.$$
 (61)

6 Summary of theoretical results

We investigated the geometry of S_p^{n-1} and proposed several retractions and their inverses and vector transports. These results are summarized in Table 1.

Table 1: Summary of theoretical results. The sphere S_p^{n-1} is defined for $p \in [1, \infty]$. However, the above results are for the case of $p \in (1, \infty)$, where S_p^{n-1} is a C^1 submanifold of \mathbb{R}^n . In addition, we assume $x, y \in S_p^{n-1}$ and $\xi, \eta \in T_x S_p^{n-1}$.

Sphere with p -norm	$S_p^{n-1} = \{ x \in \mathbb{R}^n \mid x _p = 1 \}.$
Riemannian metric on S_p^{n-1}	$\langle \xi, \eta \rangle_x = \xi^T \eta.$
Induced norm in $T_x S_p^{n-1}$	$\ \xi\ _x = \ \xi\ _2 = \sqrt{\xi^T \xi}.$
Tangent space at x	$T_x S_p^{n-1} = \{ \xi \in \mathbb{R}^n \mid \xi^T (\operatorname{sgn}(x) \odot x ^{p-1}) = 0 \}.$
Normal space at x	$N_x S_p^{n-1} = \{ \alpha \operatorname{sgn}(x) \odot x ^{p-1} \mid \alpha \in \mathbb{R} \}.$
Orthogonal projection onto $T_x S_p^{n-1}$	$P_x = I - \frac{(\operatorname{sgn}(x) \odot x ^{p-1})(\operatorname{sgn}(x) \odot x ^{p-1})^T}{\ x ^{p-1}\ _2^2}.$
Retraction by normalization	$R_x(\eta) = \frac{x+\eta}{\ x+\eta\ _p}.$
Inverse of R_x	$R_x^{-1}(y) = \frac{y}{(\operatorname{sgn}(x) \odot x ^{p-1})^T y} - x,$
	where y satisfies $(\operatorname{sgn}(x) \odot x ^{p-1})^T y > 0.$
Inverse of projective retraction	$(R_x^{\text{proj}})^{-1}(y) = y - x + \alpha \operatorname{sgn}(y) \odot y ^{p-1},$
	where $\alpha = \frac{1 - (\operatorname{sgn}(x) \odot x ^{p-1})^T y}{(\operatorname{sgn}(x) \odot x ^{p-1})^T (\operatorname{sgn}(y) \odot y ^{p-1})}$
	and y is such that $\alpha \ge 0$.
Inverse of orthographic retraction	$(R_x^{\text{orth}})^{-1}(y) = y - x + \alpha \operatorname{sgn}(x) \odot x ^{p-1}$
	where $\alpha = \frac{1 - (\operatorname{sgn}(x) \odot x ^{p-1})^T y}{\ x ^{p-1}\ _2^2}$
	and y is such that α is the solution to
	$\ x+\eta-\alpha\operatorname{sgn}(x)\odot x ^{p-1}\ _p^p=1$
	with the smallest absolute value.
Differentiated retraction of R	$\mathcal{T}_{\eta}^{R}(\xi) = \mathrm{D}R_{x}(\eta)[\xi]$
	$= \frac{\xi}{\ z\ _p} - \frac{(\operatorname{sgn}(z) \odot z ^{p-1})^T \xi}{\ z\ _p^{p+1}} z,$
	where $z = x + \eta$.
Vector transport by projection	$\mathcal{T}_{\eta}^{P}(\xi) = P_{R_{x}(\eta)}(\xi)$ = $\xi - \frac{(\operatorname{sgn}(z) \odot z ^{p-1})^{T} \xi}{\ z ^{p-1}\ _{2}^{2}} \operatorname{sgn}(z) \odot z ^{p-1},$
	where $z = x + \eta$.

7 Applications

In this section, we discuss two types of applications of S_p^{n-1} for optimization.

7.1 Nonnegative constraints on spheres

In nonlinear optimization, we can introduce squared slack variables to handle nonnegative constraints [6]. Specifically, the constraint $v \ge 0$ for $v \in \mathbb{R}^n$ is equivalent to $v = x^2$ with $x \in \mathbb{R}^n$. This idea can be used to address optimization problems on the sphere whose decision variable vector is constrained to be nonnegative.

7.1.1 Unconstrained and constrained optimization problems on spheres with different norms

For $p' \ge 1$ and p = 2p', S_p^{n-1} can be used to handle the variable on $S_{p'}^{n-1}$ with the nonnegative constraint. To see this, we consider the following problem:

minimize
$$g(v)$$

subject to $v \ge 0, v \in S_{p'}^{n-1},$

where $g: S_{p'}^{n-1} \to \mathbb{R}$ is the objective function. This is a constrained Riemannian optimization problem on $S_{p'}^{n-1}$ with the constraint $v \ge 0$. Defining $v := x^2 \ge 0$ with $x = (x_i) \in \mathbb{R}^n$, we can observe that the conditions $v \in S_{p'}^{n-1}$ and $v \ge 0$ are equivalent to $||x^2||_{p'} = 1$. Regarding the left-hand side, we have the relation $||x^2||_{p'}^p = \sum_{i=1}^n |x_i|^{2p'} = \sum_{i=1}^n |x_i|^{2p'} = ||x||_{2p'}^p = ||x||_p^p$. Therefore, $||x^2||_{p'} = 1$ is equivalent to $||x||_p = 1$, i.e., $x \in S_p^{n-1}$. Hence, the aforementioned optimization problem is equivalent to the following problem:

minimize
$$f(x) := g(x^2)$$

subject to $x \in S_p^{n-1}$,

which is an unconstrained Riemannian optimization problem on S_p^{n-1} .

7.1.2 Application to nonnegative PCA

As a particular case of p' = 2 and p = 4, we can deduce from the above discussion that solving an optimization problem on S_2^{n-1} with the nonnegative constraint on the decision variable vector is equivalent to solving the corresponding optimization problem on S_4^{n-1} without constraint. An important example within this framework is the nonnegative PCA [20].

In [10], the nonnegative PCA is formulated as follows:

minimize
$$-v^T A v$$

subject to $v \ge 0, v \in S_2^{n-1}$, (62)

where A corresponds to the variance–covariance matrix of the data to be analyzed. We assume that A is an $n \times n$ symmetric positive definite matrix. The above problem is equivalent to the following unconstrained problem on the sphere S_4^{n-1} with 4-norm:

minimize
$$f(x) := -(x^2)^T A(x^2)$$

subject to $x \in S_4^{n-1}$. (63)

For an optimal solution x_* to the latter problem, $v_* := x_*^2$ is an optimal solution to the former problem.

We can further show that any critical point of f in Problem (63) satisfies the first-order optimality conditions for Problem (62). First, we show that the KKT conditions are first-order necessary conditions for Problem (62) and investigate the conditions.

Proposition 7.1. Let v_* be an optimal solution to Problem (62) with an $n \times n$ symmetric positive definite matrix A. Then, v_* satisfies

$$v_* \ge 0, \quad v_*^T v_* = 1, \quad (I - v_* v_*^T) A v_* \le 0.$$
 (64)

Specifically, the ith element $(Av_*)_i$ of Av_* satisfies $(Av_*)_i = 0$ if $(v_*)_i > 0$, and $(Av_*)_i \leq 0$ if $(v_*)_i = 0$. In particular, if $v_* > 0$, then $Av_* = (v_*^T Av_*)v_*$ holds, i.e., $v_*^T Av_*$ and v_* are an eigenvalue and associated eigenvector of A, respectively.

Proof. Problem (62) is equivalent to the following Euclidean optimization problem:

minimize
$$-v^T A v$$

subject to $v \ge 0, v^T v = 1, v \in \mathbb{R}^n$. (65)

Throughout this proof, let $\mathcal{A}(v_*) := \{i_1, i_2, \ldots, i_m\} \subset \{1, 2, \ldots, n\}$ be the set of indices for the active inequality constraints at v_* among the *n* constraints $v_1 \ge 0, v_2 \ge 0, \ldots, v_n \ge 0$, and $\overline{\mathcal{A}}(v_*) := \{1, 2, \ldots, n\} - \mathcal{A}(v_*)$ be the complement of $\mathcal{A}(v_*)$ in $\{1, 2, \ldots, n\}$, i.e.,

$$(v_*)_{i_1} = (v_*)_{i_2} = \dots = (v_*)_{i_m} = 0, \tag{66}$$

and $(v_*)_i \neq 0$ for all $i \in \overline{\mathcal{A}}(v_*)$. Then, letting $e_i \in \mathbb{R}^n$ be the vector whose *i*th element is 1 and the others are 0, the gradients of the *m* functions defining the active inequality constraints are $e_{i_1}, e_{i_2}, \ldots, e_{i_m}$. Since $v_*^T v_* = 1$, v_* is not 0; and $\overline{\mathcal{A}}(v_*)$ is not empty, i.e, there exists $i_0 \neq i_1, i_2, \ldots, i_m$ such that $(v_*)_{i_0} \neq 0$. Hence, the gradient of the equality constraint function $v^T v - 1$ at v_* , which is $2v_*$, and $e_{i_1}, e_{i_2}, \ldots, e_{i_m}$ are linearly independent. This means that the linear independent constraint qualification (LICQ) [11] holds at v_* , and the KKT conditions for (65) are necessary optimality conditions.

Writing the KKT conditions explicitly, there exist $\lambda \in \mathbb{R}^n$ and $\mu \in \mathbb{R}$ such that

$$-2Av_* - \lambda + 2\mu v_* = 0, (67)$$

$$v_* \ge 0, \tag{68}$$

$$v_*^T v_* = 1, (69)$$

$$\lambda \ge 0,\tag{70}$$

$$\lambda \odot v_* = 0. \tag{71}$$

Under (68) and (70), Eq. (71) is equivalent to the condition $\lambda^T v_* = 0$. Using this and (69), and multiplying (67) by v_*^T from the left, we obtain $\mu = v_*^T A v_*$. Therefore, (67) yields that $\lambda = 2((v_*^T A v_*)I - A)v_* \ge 0$, which implies $(I - v_*v_*^T)Av_* \le 0$. Thus, the conditions (64) are verified to hold.

Here, $I - v_* v_*^T$ is the orthogonal projection matrix to the orthogonal complement of the span of $v_* \geq 0$ with respect to the 2-norm. Therefore, from (66), the intersection of the image $\operatorname{Im}(I - v_* v_*^T)$ of $I - v_* v_*^T$ and the nonpositive orthant $\mathbb{R}^n_- := \{x \in \mathbb{R}^n \mid x \leq 0\}$ is

$$\operatorname{Im}(I - v_* v_*^T) \cap \mathbb{R}^n_- = \{ x = (x_i) \in \mathbb{R}^n_- \mid x_i = 0, \ i \in \mathcal{A}(v_*) \}.$$
(72)

It follows from (64) and (72) that the *i*th element of $(I - v_* v_*^T) A v_*$ is

$$((I - v_* v_*^T) A v_*)_i \begin{cases} = 0 & \text{if } i \in \overline{\mathcal{A}}(v_*), \\ \leq 0 & \text{if } i \in \mathcal{A}(v_*). \end{cases}$$
(73)

Rewriting this, we obtain the relations $(Av_*)_i = (v_*^T A v_*)(v_*)_i$ if $i \in \overline{\mathcal{A}}(v_*)$, i.e., $(v_*)_i > 0$, and $(Av_*)_i \leq (v_*^T A v_*)(v_*)_i = 0$ if $i \in \mathcal{A}(v_*)$, i.e., $(v_*)_i = 0$.

In particular, if $v_* > 0$, then $\mathcal{A}(v_*) = \emptyset$, and $(Av_*)_i = (v_*^T A v_*)(v_*)_i$ for all $i \in \{1, 2, \ldots, n\}$, which is equivalent to $Av_* = (v_*^T A v_*)v_*$. This completes the proof.

Remark 7.1. Conversely, it can be readily checked that if $v_* \in \mathbb{R}^n$ satisfies the conditions (64), then v_* , $\lambda = 2((v_*^T A v_*)I - A)v_*$, and $\mu = v_*^T A v_*$ satisfy the KKT conditions (67)–(71). In summary, there exist λ and μ such that the KKT conditions (67)–(71) are satisfied if and only if v_* satisfies (64).

Remark 7.2. From the last statement of Proposition 7.1, if there does not exist an eigenvector v associated with the largest eigenvalue of A such that v > 0, then at least one inequality constraint is active at an optimal solution v_* to Problem (62), i.e., v_* contains at least one zero element.

If v_* is an optimal solution to Problem (62), x_* satisfying $v_* = x_*^2$ is an optimal solution to Problem (63). Therefore, such x_* satisfies grad $f(x_*) = 0$ on S_4^{n-1} . More generally, as the following proposition claims, if v_* satisfies the first-order necessary conditions (64) for Problem (62), then x_* that satisfies $v_* = x_*^2$ is a critical point of f in (63).

Proposition 7.2. Consider Problem (63) with an $n \times n$ symmetric positive definite matrix A. The gradient of the objective function f on S_4^{n-1} satisfies

grad
$$f(x) = -4\left((Ax^2) \odot x - \frac{(x^4)^T Ax^2}{\|x^3\|_2^2}x^3\right)$$
 (74)

for any $x \in S_4^{n-1}$. Furthermore, if $v_* \in S_2^{n-1}$ satisfies (64) and $x_* \in S_4^{n-1}$ satisfies $x_* = v_*^2$, then grad $f(x_*) = 0$.

Proof. We first derive Eq. (74) for grad f. Let $\overline{f}(x) := -(x^2)^T A(x^2)$ in \mathbb{R}^n , which is a smooth extension of f to \mathbb{R}^n . For any $d \in \mathbb{R}^n$, the directional derivative of \overline{f} at x in the direction of d is computed as

$$D\bar{f}(x)[d] = -4(x^2)^T A(x \odot d) = -4(Ax^2)^T (x \odot d) = -4((Ax^2) \odot x)^T d.$$
(75)

Hence, we obtain $\nabla \bar{f}(x) = -4(Ax^2) \odot x$. The Riemannian gradient grad f is then obtained by using the orthogonal projection (23) as

$$\operatorname{grad} f(x) = P_x(\nabla \bar{f}(x)) \tag{76}$$

$$= -4\left(I - \frac{(\operatorname{sgn}(x) \odot |x|^3)(\operatorname{sgn}(x) \odot |x|^3)^T}{\|x^3\|_2^2}\right)((Ax^2) \odot x)$$
(77)

$$= -4\left((Ax^2) \odot x - \frac{(x^4)^T Ax^2}{\|x^3\|_2^2} x^3\right),\tag{78}$$

where we used $sgn(x) \odot |x|^3 = x \odot |x|^2 = x^3$. Thus, (74) is proved.

Subsequently, we assume that v_* satisfies (64) and x_* satisfies $v_* = x_*^2$. As in the proof of Proposition 7.1, let $\mathcal{A}(v_*) = \{i_1, i_2, \ldots, i_m\} \subset \{1, 2, \ldots, n\}$ be the set of indices such that $(v_*)_{i_1} = (v_*)_{i_2} = \cdots = (v_*)_{i_m} = 0$ holds and $\overline{\mathcal{A}}(v_*) := \{1, 2, \ldots, n\} - \mathcal{A}(v_*)$. Defining $\mu := v_*^T A v_*$, it follows from (73) that

$$(v_*^2)^T A v_* = \sum_{i \in \bar{\mathcal{A}}(v_*)} (v_*)_i^2 (A v_*)_i = \sum_{i \in \bar{\mathcal{A}}(v_*)} (v_*)_i^2 \mu(v_*)_i = \mu \sum_{i \in \bar{\mathcal{A}}(v_*)} (v_*)_i^3 = \mu \|v_*\|_3^3.$$
(79)

Here, from (73), we have $((I - v_* v_*^T) A v_*) \odot v_* = 0$, which, together with $v_* \neq 0$ and (79), yields $(Av_*) \odot v_* = (v_* v_*^T A v_*) \odot v_* = \mu v_*^2 = ((v_*^2)^T A v_* / \|v_*\|_3^3) v_*^2$. Substituting $v_* = x_*^2$, this is written as

$$(Ax_*^2) \odot x_*^2 = \frac{(x_*^4)^T A(x_*^2)}{\|x_*^2\|_3^3} x_*^4 = \frac{(x_*^4)^T A(x_*^2)}{\|x_*^3\|_2^2} x_*^4.$$
(80)

In general, for any $a, b, c \in \mathbb{R}$, $ac^2 = bc^4$ is equivalent to $ac = bc^3$. Therefore, (80) is reduced to

$$(Ax_*^2) \odot x_* = \frac{(x_*^4)^T A(x_*^2)}{\|x_*^3\|_2^2} x_*^3, \tag{81}$$

which shows that grad $f(x_*) = 0$ in light of (74). This completes the proof.

7.1.3 Numerical experiments for nonnegative PCA

Here, we demonstrate numerical experiments for the nonnegative PCA. To solve the constrained Problem (62) on S_2^{n-1} , we solve the unconstrained Problem (63) on S_4^{n-1} to obtain $x_n^{\text{proposed}} \in S_4^{n-1}$. Then, we obtain $v_n^{\text{proposed}} := (x_n^{\text{proposed}})^2$ as a solution to the original Problem (62) based on the proposed framework. For comparison, we also solve the constrained Euclidean optimization Problem (65), which is equivalent to Problem (62), using MATLAB's fmincon function, to obtain v_n^{fmincon} .

We consider the two cases of n = 10 and n = 1000. For each n, we constructed an $n \times n$ symmetric positive definite matrix A with randomly generated elements. Implementing the orthogonal projection (23) and retraction (28) based on Manopt [4], we applied the Riemannian conjugate gradient method on S_4^{n-1} to Problem (63) with n = 10 and n = 1000. The initial point x_0 for solving Problem (63) was also randomly constructed, and we used $v_0 := x_0^2$ as the initial point for solving Problem (65) by fmincon.

For n = 10, each elements of the two solutions v_{10}^{proposed} and v_{10}^{fmincon} are the same to the third decimal place, as $(0.000, 0.604, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.116, 0.788)^T$. Furthermore, they are sparse. This is consistent with the discussion in Remark 7.2.

Subsequently, for n = 1000, we have $||v_{1000}^{\text{proposed}} - v_{1000}^{\text{fmincon}}||_2 = 1.307$. Furthermore, the values of the function $g(v) := -v^T A v$, which should be minimized in Problems (62) and (65), are $g(v_{1000}^{\text{proposed}}) = -2.963 \times 10^3 < -1.805 \times 10^3 = g(v_{1000}^{\text{fmincon}})$. In addition, $v_{1000}^{\text{proposed}}$ is sparse because 479 of 1000 elements of $v_{1000}^{\text{proposed}}$ are less than 10^{-6} , while no element of $v_{1000}^{\text{fmincon}}$ is less than 10^{-6} . Therefore, the proposed framework yielded a much better solution in this case.

7.2 L_p -regularization-related optimization

In certain applications, L_p regularization is a frequently used technique, which considers an objective function as the weighted sum of the original objective function and the *p*-norm of the decision variable vector. In particular, L_1 regularization is used in Lasso for sparse estimation [7].

7.2.1 Relationship between regularized, constrained, and manifold optimization problems

We consider the following regularized optimization problem with $p \in [1, \infty]$:

minimize
$$L(w) + \lambda ||w||_p$$

subject to $w \in \mathbb{R}^n$, (82)

where $L: \mathbb{R}^n \to \mathbb{R}$ is a convex function, and $\lambda \ge 0$ is a predefined constant called a regularization parameter.

Intuitively, L_p regularization is closely related to considering the constraint that the *p*-norm of the decision variable vector is not larger than a predefined nonnegative constant. Specifically, the corresponding constrained optimization problem is written as follows:

minimize
$$L(w)$$

subject to $||w||_p \le C, w \in \mathbb{R}^n$, (83)

where $C \ge 0$ is a constant. For example, while Lasso regression is performed by solving the former unconstrained Problem (82), the latter Problem (83) is sometimes used to explain why Lasso tends to find a sparse solution [7]. This intuition is justified even for general $p \in [1, \infty]$ through the following proposition:

Proposition 7.3. Assume that $p \in [1, \infty]$, and let $L: \mathbb{R}^n \to \mathbb{R}$ be a convex function. If w_* is an optimal solution to Problem (82) with a predefined constant $\lambda \ge 0$, then there exists $C \ge 0$ such that w_* is an optimal solution to Problem (83) with C. Conversely, if w_* is an optimal solution to Problem (83) with a predefined constant $C \ge 0$, then there exists $\lambda \ge 0$ such that w_* is an optimal solution to Problem (82).

Proof. First, we fix $\lambda \geq 0$ and let w_* be an optimal solution to Problem (82). Then, for any $w \in \mathbb{R}^n$, we have

$$L(w_{*}) + \lambda \|w_{*}\|_{p} \le L(w) + \lambda \|w\|_{p}.$$
(84)

We show that w_* is an optimal solution to Problem (83) with $C := ||w_*||_p$. For any feasible solution $w \in \mathbb{R}^n$ to Problem (83), we have $||w||_p \leq C = ||w_*||_p$. Combining this and (84), we have $L(w_*) + \lambda ||w_*||_p \leq L(w) + \lambda ||w_*||_p$, which means $L(w_*) \leq L(w)$. Furthermore, w_* is clearly a feasible solution to Problem (83) since $||w_*||_p = C$. Therefore, w_* is an optimal solution to Problem (83).

Conversely, we fix $C \ge 0$ and let w_* be an optimal solution to Problem (83). Here, we additionally consider the Lagrange dual problem of (83):

maximize
$$\inf_{w \in \mathbb{R}^n} (L(w) + \mu(||w||_p - C))$$

subject to $\mu \ge 0, \ \mu \in \mathbb{R}.$ (85)

If C > 0, then Slater's condition for Problem (83), which is that there exists $w \in \mathbb{R}^n$ with $||w||_p < C$, clearly holds with w = 0. If C = 0, then the constraint $||w||_p \leq C$ in Problem (83) is rewritten as the equality constraint w = 0, and Slater's condition (which in this case is that a feasible solution exists) holds by taking w = 0. In each case, Slater's condition for Problem (83) holds. Furthermore, Problem (83) is a convex optimization problem. Therefore, it follows from Slater's theorem [5, Section 5.2.3] that strong duality holds. Hence, the optimal value $L(w_*)$ of Problem (83) and the optimal value of the dual problem (85) coincide. Letting $\mu_* \ge 0$ be an optimal solution to (85), we have

$$L(w_*) = \inf_{w \in \mathbb{R}^n} (L(w) + \mu_*(||w||_p - C)).$$
(86)

Since $||w_*||_p \leq C$ and $\mu_* \geq 0$, we obtain $L(w_*) \leq L(w_*) + \mu_*(||w_*||_p - C) \leq L(w_*)$. Thus, $L(w_*) = L(w_*) + \mu_*(||w_*||_p - C)$ holds, and $w = w_*$ attains the minimum value of $L(w) + \mu_*(||w||_p - C)$ over all $w \in \mathbb{R}^n$. Since μ_*C is a constant, $w = w_*$ also attains the minimum value of $L(w) + \mu_*||w||_p$ over \mathbb{R}^n . This implies that w_* is an optimal solution to Problem (82) with $\lambda = \mu_*$, thereby completing the proof.

Remark 7.3. Although we focus on the *p*-norm here, Proposition 7.3 can be straightforwardly generalized to the case with a general norm in \mathbb{R}^n . Indeed, in the proof of Proposition 7.3, we do not exploit any specific property of the *p*-norm but properties of a general norm.

From Proposition 7.3, we can observe the importance of Problem (83) in dealing with Problem (82). Furthermore, Problem (83) is closely related to the following problem:

minimize
$$L(w)$$

subject to $||w||_p = C, w \in \mathbb{R}^n$, (87)

where C is the constant in Problem (83). Indeed, if a minimum point w_* of L over the entire \mathbb{R}^n lies in the ball $\{w \in \mathbb{R}^n \mid ||w||_p \leq C\}$, then w_* is also an optimal solution to (83). Therefore, a practically more important case we focus on is when all minimum points of L over \mathbb{R}^n are outside the ball. In this case, we can show that there exists an optimal solution to Problem (83) that is on the sphere $\{w \in \mathbb{R}^n \mid ||w||_p = C\}$ as the following proposition:

Proposition 7.4. Assume $p \in [1, \infty]$, let $L: \mathbb{R}^n \to \mathbb{R}$ be convex, and consider Problem (83) with a constant $C \ge 0$. Assume that any minimum point y_* of L over \mathbb{R}^n satisfies $||y_*||_p > C$. Then, there exists an optimal solution w_* to Problem (83) that satisfies $||w_*||_p = C$.

Proof. Let y_* and z_* be a minimum point of L over \mathbb{R}^n and optimal solution to Problem (83), respectively. If $||z_*||_p = C$, then we can take z_* as w_* in the statement of the proposition.

In the remainder of the proof, we assume $||z_*|| < C$. From the assumption, we have $||z_*||_p < C < ||y_*||_p$ and $L(y_*) \le L(z_*)$. Since L is convex, for $\alpha \in [0, 1]$, we have

$$L(\alpha y_* + (1 - \alpha)z_*) \le \alpha L(y_*) + (1 - \alpha)L(z_*) \le \alpha L(z_*) + (1 - \alpha)L(z_*) = L(z_*).$$
(88)

Note that the function $\varphi(\alpha) := \|\alpha y_* + (1-\alpha)z_*\|_p$ is continuous with respect to α , where φ satisfies $\varphi(0) = \|z_*\|_p < C$ and $\varphi(1) = \|y_*\|_p > C$. Therefore, from the intermediate value theorem, there exists $\alpha_* \in (0,1)$ such that $\varphi(\alpha_*) = C$. With this α_* , defining $w_* := \alpha_* y_* + (1-\alpha_*)z_*$, we have $\|w_*\|_p = \varphi(\alpha_*) = C$, implying that w_* is feasible for Problem (83). Since z_* is optimal for (83), we have $L(z_*) \leq L(w_*)$. On the contrary, Eq. (88) yields that $L(w_*) \leq L(z_*)$. Thus, we obtain $L(w_*) = L(z_*)$, which means that w_* is an optimal solution to Problem (83) with $\|w_*\|_p = C$. This completes the proof.

From this proposition, if no minimum point of L over \mathbb{R}^n lies in the ball $\{w \in \mathbb{R}^n \mid ||w||_p \leq C\}$, then any optimal solution to Problem (87) is also an optimal solution to Problem (83), i.e., it is sufficient to solve Problem (87) for obtaining an optimal solution to Problem (83). Furthermore, upon scaling $w \mapsto w/C$ and $L \mapsto L \circ CI$ and writing w/C and $L \circ CI$ newly as x and f, respectively, i.e., f(x) := L(Cx) = L(w), Problem (87) essentially becomes equivalent to the following problem on the unit sphere S_p^{n-1} with *p*-norm:

minimize
$$f(x)$$

subject to $x \in S_p^{n-1}$. (89)

The important cases p = 1 and $p = \infty$ do not lie within the scope of the discussion in the previous sections. Therefore, we approximate S_1^{n-1} and S_{∞}^{n-1} by S_p^{n-1} with $p = 1 + \varepsilon$, where $\varepsilon > 0$ is sufficiently small, and S_p^{n-1} with sufficiently large p, respectively.

7.2.2 Numerical experiment for Lasso regression

Here, we consider the Lasso regression with simple artificial data. The data size is set as m = 100 and the number of variables as n = 13. We construct a data matrix $X \in \mathbb{R}^{m \times n}$ with randomly generated elements, set $w_* = (-5, -4, -3, -2, -1, 1, 2, 3, 4, 5, 0, 0, 0)^T \in \mathbb{R}^n$, and compute $y = Xw_* + \epsilon$, where each element of $\epsilon \in \mathbb{R}^n$ is randomly generated from a uniform distribution on the interval [-1, 1]. This means that, among n = 13 variables, the first 10 are essential and the last 3 have no effect in the data y. We now estimate the coefficient parameter vector w_* without any information on it, i.e., by using only the observed data Xand y. An appropriate sparse estimation should yield a coefficient parameter vector whose last 3 elements are close to 0.

With the data X and y, we consider Problem (87) with $L(w) := ||Xw - y||_2^2$ and a constant C > 0, i.e., with the equality constraint $||w||_p = C$. Note that we exclude the case when C = 0 since it yields the trivial solution w = 0. Solving this problem is equivalent to minimizing $f(x) := L(Cx) = ||CXx - y||_2^2$ with respect to $x \in S_p^{n-1}$, i.e., solving Problem (89), and multiplying the resultant solution x_* by C to obtain the solution $w_* = Cx_*$ to Problem (87).

The case of p = 1 corresponds to the Lasso regression. However, we can handle S_p^{n-1} with p > 1 using the Riemannian optimization techniques developed in the previous sections. Therefore, we adopt $p = 1.000001 = 1 + 10^{-6}$ and expect that solving the problem on S_p^{n-1} yields a sparse solution. Implementing the projection (23) and retraction (28) based on Manopt, we applied the Riemannian conjugate gradient method for Problem (89) on S_p^{n-1} .

In Table 2, $w^{\text{nonreg}} := (X^T X)^{-1} X^T y$ is the solution to the nonregularized optimization problem of minimizing L, i.e., Problem (82) with $\lambda = 0$. As expected, this is not sparse. Then, we applied the Riemannian conjugate gradient method in the proposed framework with several C and obtained the solution w_C^{proposed} to Problem (87) for each C. The results for C = 1, 5, 10, 20, 22, 25, 30, 50, 100 are shown in the table. For small C such as C = 1, 5, 10, the resultant solutions are sparse but do not provide a good estimation because the 5th and 6th entries are almost zero and the 11th and 13th are nonzero. On the contrary, large C does not contribute to sparse estimation at all. Although finding the best value of C is difficult, we observe that the case of C = 22 yields an appropriate solution in this experiment, which is a sparse solution with appropriate values.

For comparison, we also applied MATLAB's **lasso** function, which successively increases the value of λ and solves Problem (82) for each λ . For small λ 's, the corresponding solutions are dense, whereas the solution is 0 for a sufficiently large λ . We focus on the λ 's and corresponding solutions $w_{\lambda}^{\text{Lasso}} \in \mathbb{R}^n$ such that only the last 3 elements of $w_{\lambda}^{\text{Lasso}}$ are 0. The **lasso** function yielded several λ 's satisfying this condition. Among them, $w_{0.029}^{\text{Lasso}}$ and $w_{0.746}^{\text{Lasso}}$ correspond to the smallest and largest values of λ , respectively. We can observe that w_{22}^{proposed} and $w_{0.746}^{\text{Lasso}}$ are close to each other.

abic 2. 103	uits obtai	ncu upon	solving u	IC Lasso-I	ciaica opi		problem	5. The i				field of ca	ch solutio
	1	2	3	4	5	6	7	8	9	10	11	12	13
w^{nonreg}	-5.055	-3.904	-3.022	-2.039	-1.036	0.967	1.972	3.028	4.036	5.060	-0.008	-0.032	0.052
$w_1^{\rm proposed}$	-0.167	-0.081	-0.150	-0.095	0.000	0.000	0.137	0.000	0.048	0.257	0.005	-0.000	-0.059
w_5^{proposed}	-0.874	-0.631	-0.727	-0.329	-0.000	0.003	0.685	0.044	0.370	1.334	0.004	-0.000	-0.000
$w_{10}^{\rm proposed}$	-1.683	-1.335	-1.359	-0.709	0.000	0.002	1.051	0.537	0.984	2.268	0.000	0.000	-0.071
$w_{20}^{\mathrm{proposed}}$	-3.357	-2.790	-2.452	-1.330	-0.048	0.255	1.564	1.765	2.437	3.907	0.076	-0.000	-0.018
$w_{22}^{\mathrm{proposed}}$	-3.779	-3.234	-2.537	-1.202	-0.119	0.294	1.819	1.914	2.829	4.272	0.000	0.000	-0.000
$w_{25}^{\mathrm{proposed}}$	-4.193	-3.422	-2.791	-1.599	-0.510	0.587	1.787	2.391	3.203	4.504	0.008	0.004	0.000
$w_{30}^{\mathrm{proposed}}$	-5.027	-3.895	-3.014	-2.016	-1.021	0.952	1.967	3.010	4.012	5.042	0.000	-0.013	0.030
$w_{50}^{\mathrm{proposed}}$	-8.040	-5.762	-4.521	-3.657	-2.872	-0.979	-0.608	5.474	6.346	7.002	-2.290	0.979	-1.471
$w_{100}^{\rm proposed}$	-15.03	0.127	-10.70	-9.352	-5.748	-7.317	-7.698	11.94	0.913	12.70	-7.959	4.992	-5.526
$w_{0.029}^{\text{Lasso}}$	-4.989	-3.881	-3.023	-1.986	-0.993	0.939	1.959	2.976	3.981	5.037	0	0	0
$w_{0.746}^{\mathrm{Lasso}}$	-3.727	-3.212	-2.712	-1.192	-0.002	0.224	1.831	1.890	2.791	4.314	0	0	0

Table 2: Results obtained upon solving the Lasso-related optimization problems. The *i*th row shows the *i*th element of each solution.

7.2.3 Numerical experiment for box-constrained problem

Here, we consider the following box-constrained optimization problem:

minimize
$$L(w)$$

subject to $l \le w \le u, \ w \in \mathbb{R}^n$, (90)

where $l = (l_i), u = (u_i) \in \mathbb{R}^n$ are given constant vectors with l < u.² The constraint $l \leq w \leq u$ means the box constraint $l_i \leq w_i \leq u_i$ for $i = 1, 2, \ldots, n$. Defining a := (u-l)/2 > 0 and b := (l+u)/2, this constraint is rewritten as $-a \leq w - b \leq a$, which is equivalent to $-\mathbf{1} \leq D^{-1}(w-b) \leq \mathbf{1}$, i.e., $\|D^{-1}(w-b)\|_{\infty} \leq 1$, with D being the $n \times n$ diagonal matrix with diagonal elements $a_1, a_2, \ldots, a_n > 0$. Therefore, with the transformation $x := D^{-1}(w-b) \in S_{\infty}^{n-1}$ and $f(x) := L(a \odot x + b) = L(Dx + b) = L(w)$, solving Problem (90) is essentially equivalent to minimizing f in the unit ball $B_{\infty}^n = \{x \in \mathbb{R}^n \mid \|x\|_{\infty} \leq 1\}$. Consider a practical case where no minimum point of f over the entire \mathbb{R}^n is in the ball B_{∞}^n . Then, as discussed in Section 7.2.1, we only have to solve Problem (89) on the sphere S_p^{n-1} with $p = \infty$. However, since $p = \infty$ was excluded from the discussion in the previous sections, we instead need to consider a sufficiently large finite value p when solving the problem numerically.

We performed a numerical experiment for the following problem with n = 10:

minimize
$$L(w) := \frac{1}{2}w^T A w + c^T w$$

subject to $l \le w \le u, \ w \in \mathbb{R}^n$, (91)

where the elements of the $n \times n$ symmetric positive definite matrix A and vector $c \in \mathbb{R}^n$ are randomly generated. We set $l = (-1, -2, \ldots, -10)^T$ and $u = (1, 2, \ldots, 10)^T$. Note that $\nabla L(w) = Aw + c$ and the minimum point of L over the entire \mathbb{R}^n is $-A^{-1}c$. We checked that $w^{\text{unconst}} := -A^{-1}c$ is not feasible for Problem (91) in this case. Therefore, as discussed above, if we minimize $f(x) := L(a \odot x + b)$ with a := (u - l)/2 and b := (l + u)/2 on the sphere S_{∞}^{n-1} to obtain x_* , then $w_* := a \odot x_* + b$ is an optimal solution to Problem (91). We approximated S_{∞}^{n-1} by S_p^{n-1} with p = 5, 10, 50, 100, 500, 1000, 5000, 10000, 50000, and solved Problem (89) by the Riemannian conjugate gradient method based on Manopt. We denote the resultant approximate solution to the original Problem (91) by w_p^{proposed} for each p and the solution to Problem (91) obtained using MATLAB's fmincon function by w^{fmincon} . The results are shown in Table 3. As expected, the larger the value of p, the more accurate is the obtained solution.

²If $l_i = u_i$ for some *i*, then the constant l_i is the only value that the corresponding w_i can take. By eliminating such a constant variable in advance if necessary, we can assume l < u without loss of generality.

	1	2	3	4	5	6	7	8	9	10	$\ w - w^{\text{fmincon}}\ _2$
w^{unconst}	-3.335	4.331	-1.575	-0.383	-1.127	5.731	-3.268	-0.024	2.072	-1.885	5.971
$w_5^{\rm proposed}$	-0.725	1.912	-0.629	-0.187	-0.693	1.697	-0.863	-0.011	0.500	-0.570	0.4488
$w_{10}^{\rm proposed}$	-0.855	1.954	-0.644	-0.243	-0.728	1.797	-0.931	0.008	0.576	-0.536	0.2357
$w_{50}^{\mathrm{proposed}}$	-0.969	1.991	-0.657	-0.294	-0.761	1.882	-0.989	0.026	0.643	-0.503	4.952×10^{-2}
$w_{100}^{\mathrm{proposed}}$	-0.985	1.995	-0.658	-0.301	-0.765	1.893	-0.997	0.028	0.652	-0.499	2.493×10^{-2}
$w_{500}^{\mathrm{proposed}}$	-0.997	1.999	-0.659	-0.306	-0.769	1.902	-1.003	0.030	0.659	-0.495	5.012×10^{-3}
$w_{1000}^{\mathrm{proposed}}$	-0.998	2.000	-0.660	-0.307	-0.769	1.903	-1.004	0.030	0.660	-0.494	2.508×10^{-3}
$w_{5000}^{\mathrm{proposed}}$	-1.000	2.000	-0.660	-0.308	-0.770	1.904	-1.004	0.030	0.660	-0.494	5.014×10^{-4}
$w_{10000}^{\mathrm{proposed}}$	-1.000	2.000	-0.660	-0.308	-0.770	1.904	-1.004	0.030	0.660	-0.494	2.508×10^{-4}
$w_{50000}^{\mathrm{proposed}}$	-1.000	2.000	-0.660	-0.308	-0.770	1.904	-1.004	0.031	0.660	-0.494	5.030×10^{-5}
$w^{\rm fmincon}$	-1.000	2.000	-0.660	-0.308	-0.770	1.904	-1.004	0.031	0.660	-0.494	0

Table 3: Results of solving box-constrained-optimization-related problems. The *i*th row shows the *i*th element of each solution, and the rightmost column shows the distance between the resultant vectors and w^{fmincon} .

8 Concluding remarks

In this paper, we investigated the geometry of the unit sphere defined via the *p*-norm as $S_p^{n-1} := \{x \in \mathbb{R}^n \mid ||x||_p = 1\}$ with $p \in [1, \infty]$, especially $p \in (1, \infty)$, in detail. In particular, we derived formulas for retractions, their inverses, and vector transports, which can be used in Riemannian optimization algorithms. The results are summarized in Table 1 of Section 6.

Furthermore, we discussed two types of applications of optimization on S_p^{n-1} . The first was for optimization problems on the sphere with the nonnegative constraint, which include the nonnegative PCA problem. The second was for L_p -regularization-related optimization problems, which are closely related to the Lasso regression and box-constrained problems. To this end, we provided mathematical support for the applications and performed numerical experiments to verify the validity of the theory.

The applications addressed in this paper are examples of the proposed theory, and the corresponding numerical experiments are preliminary ones. Therefore, developing more efficient algorithms by combining the present theory and existing Riemannian optimization theory than state-of-the-art algorithms for specific problems, e.g., the nonnegative PCA and Lasso problems, are left for future work.

References

- P.-A. Absil, R. Mahony, and R. Sepulchre. Optimization Algorithms on Matrix Manifolds. Princeton University Press, Princeton, NJ, 2008.
- [2] P.-A. Absil and J. Malick. Projection-like retractions on matrix manifolds. SIAM J. Optim., 22(1):135–158, 2012.
- [3] R. L. Adler, J.-P. Dedieu, J. Y. Margulies, M. Martens, and M. Shub. Newton's method on Riemannian manifolds and a geometric model for the human spine. *IMA J. Numer. Anal.*, 22(3):359–390, 2002.
- [4] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre. Manopt, a Matlab toolbox for optimization on manifolds. J. Mach. Learn Res., 15(1):1455–1459, 2014.
- [5] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
- [6] E. H. Fukuda and M. Fukushima. A note on the squared slack variables technique for nonlinear optimization. J. Oper. Res. Soc. Jpn., 60(3):262–270, 2017.
- [7] T. Hastie, R. Tibshirani, and M. Wainwright. Statistical Learning with Sparsity: The Lasso and Generalizations. CRC Press, 2015.
- [8] W. Huang, P.-A. Absil, and K. A. Gallivan. A Riemannian BFGS method without differentiated retraction for nonconvex optimization problems. SIAM J. Optim., 28(1):470–495, 2018.
- [9] 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.
- [10] C. Liu and N. Boumal. Simple algorithms for optimization on Riemannian manifolds with constraints. Appl. Math. Optim., 82(3):949–981, 2020.

- [11] J. Nocedal and S. Wright. Numerical Optimization, 2nd edn. Springer, 2006.
- [12] W. Ring and B. Wirth. Optimization methods on Riemannian manifolds and their application to shape space. SIAM J. Optim., 22(2):596–627, 2012.
- [13] H. Sakai and H. Iiduka. Sufficient descent Riemannian conjugate gradient methods. J. Optim. Theory Appl., 190(1):130–150, 2021.
- [14] H. Sato. A Dai-Yuan-type Riemannian conjugate gradient method with the weak Wolfe conditions. *Comput. Optim. Appl.*, 64(1):101–118, 2016.
- [15] H. Sato. Riemannian Optimization and Its Applications. Springer Nature, 2021.
- [16] H. Sato and T. Iwai. A new, globally convergent Riemannian conjugate gradient method. Optimization, 64(4):1011–1031, 2015.
- [17] H. Sato, H. Kasai, and M. Bamdev. Riemannian stochastic variance reduced gradient algorithm with retraction and vector transport. SIAM J. Optim., 29(2):1444–1472, 2019.
- [18] M. Shub. Some remarks on dynamical systems and numerical analysis. In Dynamical Systems and Partial Differential Equations: Proceedings of VII ELAM, pages 69–92, 1986.
- [19] L. W. Tu. An Introduction to Manifolds. Springer New York, 2010.
- [20] R. Zass and A. Shashua. Nonnegative sparse PCA. In Adv. Neural Inf. Process. Syst., volume 19, pages 1561–1568, 2007.
- [21] P. Zhou, X.-T. Yuan, S. Yan, and J. Feng. Faster first-order methods for stochastic nonconvex optimization on Riemannian manifolds. *IEEE Trans. Pattern Anal. Mach. Intell.*, 43(2):459–472, 2019.
- [22] X. Zhu and H. Sato. Riemannian conjugate gradient methods with inverse retraction. Comput. Optim. Appl., 77(3):779–810, 2020.