DIFFERENTIAL GEOMETRY WITH EXTREME EIGENVALUES IN THE POSITIVE SEMIDEFINITE CONE[*]

CYRUS MOSTAJERAN[†], NATHAËL DA COSTA[†], GRAHAM VAN GOFFRIER[‡], AND RODOLPHE SEPULCHRE[§]

Abstract. Differential geometric approaches to the analysis and processing of data in the form of symmetric positive definite (SPD) matrices have had notable successful applications to numerous fields including computer vision, medical imaging, and machine learning. The dominant geometric paradigm for such applications has consisted of a few Riemannian geometries associated with spectral computations that are costly at high scale and in high dimensions. We present a route to a scalable geometric framework for the analysis and processing of SPD-valued data based on the efficient computation of extreme generalized eigenvalues through the Hilbert and Thompson geometries of the semidefinite cone. We explore a particular geodesic space structure based on Thompson geometry in detail and establish several properties associated with this structure. Furthermore, we define a novel iterative mean of SPD matrices based on this geometry and prove its existence and uniqueness for a given finite collection of points. Finally, we state and prove a number of desirable properties that are satisfied by this mean.

Key words. affine-invariance, convex cones, differential geometry, geodesics, geometric statistics, Hilbert metric, positive definite matrices, matrix means, Thompson metric

AMS subject classifications. 15B48, 53B50, 53C22, 53C80, 65F15

1. Introduction. Geometric data that lie in convex cones appear in a wide variety of applications. Of particular interest is the space of symmetric positive definite (SPD) matrices of a given dimension, which forms the interior of the convex cone of positive semidefinite matrices in the corresponding vector space of symmetric matrices. In medical imaging, SPD matrices model the covariance matrices of Brownian motion of water in Diffusion Tensor Imaging (DTI) [48]. In radar data processing, circular complex random processes with a null mean are characterized by Toeplitz Hermitian positive definite matrices [6]. In the context of brain-computer interfaces (BCI), where the objective is to enable users to interact with computers via brain activity alone (e.g. to enable communication for severely paralyzed users), the time-correlation of electroencephalogram (EEG) signals are encoded by SPD matrices [9]. SPD matrices appear as kernel matrices in machine learning [33]. SPD representations also find applications in process control, monitoring, and anomaly detection [23, 55, 64], object detection [62, 65], and the study of functional brain networks [28, 56].

Since SPD matrices do not form a Euclidean space, standard linear analysis techniques applied directly to such data are often inappropriate and result in poor performance. For instance, the regularization of DTI images using gradient descent algorithms that utilize the classical Euclidean (Frobenius) norm almost inevitably lead

[†]School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore.

[‡]Department of Physics and Astronomy, University College London, London, UK.

[§]Department of Engineering, University of Cambridge, UK & Department of Electiral Engineering, KU Leuven, Belgium.

to points in the image with negative eigenvalues. Even if we remain in the SPD cone, use of Euclidean (linear) geometry often results in other problems such as 'swelling' phenomena in interpolation in DTI [7, 48] or poor classification results in the context of BCI [9, 10, 22].

In order to cope with these problems, several Riemannian geometries on SPD matrices have been proposed and used effectively in a variety of applications in computer vision [30, 31, 38], medical data analysis [7, 48, 49], machine learning [21, 39, 66], and optimization [1, 16, 15, 40]. In particular, the affine-invariant Riemannian metric—so-called because it is invariant to affine transformations of the underlying spacial coordinates—has received considerable attention in recent years and applied successfully to problems such as EEG signal processing in BCI where it has been shown to be superior to classical techniques based on feature vector classification [9, 10, 22]. More recently, geometric deep learning architectures have been proposed to learn statistical representations of SPD-valued data that respect the underlying Riemannian geometry [18, 29, 32]. The affine-invariant Riemannian geometry has also been applied in the field of geometric statistics where it has been used to construct Riemannian Gaussian distributions, which are used as building blocks for learning models that describe the structure of statistical populations of SPD matrices [20, 50, 51, 52, 53, 61].

The affine-invariant Riemannian metric endows the space of SPD matrices of a given dimension with the structure of a Hadamard manifold with non-constant negative curvature [34]. Computing standard geometric objects such as distances, geodesics, Riemannian exponentials and logarithms in this geometry often amounts to the computation of the generalized eigenspectrum of a pair of SPD matrices, which typically means a significant increase in computational complexity, particularly for larger matrices.

An important point that has not received much attention in the literature on geometric optimization and statistics involving SPD-valued data is that there are natural non-Riemannian geometries that can be associated with SPD matrices based on the conic structure of the space. In particular, the Hilbert and Thompson metrics [8, 35, 42, 45, 60] on the cone of SPD matrices generate non-Euclidean geometries with a rich set of properties including distance and geodesic computations that rely only on extreme generalized eigenvalues [42, 63], which are efficiently computable using techniques such as Krylov subspace methods based on matrix-vector products [25, 27, 57, 58]. The full utilization of non-Euclidean geometries that are naturally suited to the SPD cone in the design of cost functions and optimization algorithms for problems involving SPD-valued data offers the potential for enhanced analytic insights and dramatic improvements in computational efficiency over existing costly Riemannian methods.

1.1. Hilbert and Thompson geometries. Let V be a finite-dimensional real vector space. A subset K of V is called a cone if it is convex, $\mu K \subseteq K$ for all $\mu \geq 0$, and $K \cap (-K) = \{0\}$. It is said to be a closed cone if it is a closed set in V with respect to the standard topology. A cone is said to be solid if it has non-empty interior. We say that a cone is almost Archimedean if the closure of its restriction to any two-dimensional subspace is also a cone. Examples of solid closed cones include the positive orthant $\mathbb{R}^n_+ = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : x_i \geq 0, \ 1 \leq i \leq n\}$ and the set of positive semidefinite matrices in the space of real $n \times n$ matrices.

A cone K in a vector space V induces a partial ordering on V given by $x \leq y$ if and only if $y - x \in K$. For each $x \in K \setminus \{0\}$, $y \in V$, define $M(y/x) := \inf\{\lambda \in \mathbb{R} : y \leq \lambda x\}$.

Hilbert's projective metric on K is defined to be

$$d_H(x, y) = \log(M(y/x)M(x/y)). \tag{1.1}$$

Hilbert's projective metric is a pseudo-metric on the cone since it can be shown that $d_H(x, y) = 0$ if and only if $x = \lambda y$ for some $\lambda > 0$. Indeed, d_H defines a metric on the space of rays of the cone [35]. A specific example of Hilbert geometry is n-dimensional hyperbolic space, which is isometric to the the Lorentz cone $\{(t, x_1, \ldots, x_n) \in \mathbb{R}^{n+1} : t^2 > x_1^2 + \ldots + x_n^2\}$ endowed with its Hilbert metric. However, Hilbert geometry only corresponds to a CAT(0) space if the cone is Lorentzian [17]. Thus, Hilbert geometry is certainly more general than hyperbolic geometry. Beyond geometry, Hilbert's projective metric finds important applications in analysis, where many naturally arising linear and nonlinear maps are either non-expansive or contractive with respect to it [13, 19, 35, 54].

Thompson's part metric on K is a closely related metric that is defined to be

$$d_T(x, y) = \log(\max\{M(y/x), M(x/y)\}). \tag{1.2}$$

Two points in K are said to be in the same part if the distance between them is finite in the Thompson metric. If K is almost Archimedean, then each part of K is a complete metric space with respect to the Thompson metric [60].

Turning our attention to the case of the positive semidefinite cone, we find that for strictly positive definite matrices $X, Y \succ 0$, $M(Y/X) = \lambda_{\max}(YX^{-1}) = 1/\lambda_{\min}(XY^{-1})$, where $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ denote the maximum and minimum eigenvalues of the matrix A, respectively. Note that $\lambda_{\max}(YX^{-1})$ is well-defined since YX^{-1} is a diagonalizable matrix with real and positive eigenvalues. It follows that the Hilbert and Thompson metrics take the form

$$d_H(X, Y) = \log\left(\frac{\lambda_{\max}(YX^{-1})}{\lambda_{\min}(YX^{-1})}\right) \tag{1.3}$$

and

$$d_T(X, Y) = \log\left(\max\{\lambda_{\max}(YX^{-1}), 1/\lambda_{\min}(YX^{-1})\}\right). \tag{1.4}$$

1.2. Paper organization and contributions. The main aim of this paper is to provide a connection between the differential geometry of SPD matrices—which has been the subject of significant research interest in recent years accompanied by notable successful applications—and numerical linear algebra, specifically iterative methods for computing extreme eigenvalues—a cornerstone of modern applied mathematics and computing. In this paper, the Hilbert and Thompson geometries of the semidefinite cone are used as a route to establish such a connection.

In section 2, we review affine-invariant metric geometry in the SPD cone and observe how the Thompson metric arises naturally as a member of a family of affine-invariant metrics generated by a collection of Finsler metrics. In section 3, we briefly note several results on the feasibility of isometrically embedding Hilbert and Thompson geometries in vector spaces. In section 4, we consider geodesics in Thompson geometry and a choose a particular geodesic with attractive computational properties as a distinguished geodesic whose properties we examine closely. In section 5, we consider the proximity of this distinguished Thompson geodesic to the affine-invariant Riemannian geodesic and establish general bounds for the distance between their associated midpoints. In section 6, we introduce a novel iterative mean of any finite

collection of SPD matrices as the limit of a sequence that is generated through constructions of Thompson geodesics (Algorithm 6.1) that can be efficiently computed in high dimensions using extreme generalized eigenvalues. We prove that this novel iterative mean of SPD matrices is well-defined by showing that any sequence generated by Algorithm 6.1 converges to a unique point that is independent of the choice of initialization (Theorem 6.14) and the ordering of the SPD matrices. Furthermore, we state and prove a number of desirable properties that are satisfied by this mean in Theorem 6.16.

2. Affine-invariant metric geometry. Let $\mathbb{S}_{++}^n$ denote the space of $n \times n$ real symmetric positive definite matrices. It is well-known that $\mathbb{S}_{++}^n$ admits a Riemannian distance function $d_2 : \mathbb{S}_{++}^n \times \mathbb{S}_{++}^n \to \mathbb{R}$

$$(2.1) \qquad d_2(X,Y) = \left(\sum_{i=1}^n \log^2 \lambda_i(YX^{-1}) \right)^{1/2},$$

where $\lambda_i(YX^{-1}) = \lambda_i(X^{-1/2}YX^{-1/2})$ denote the n real and positive eigenvalues of YX^{-1}. (2.1) endows $\mathbb{S}_{++}^n$ with the structure of a Riemannian symmetric space and a metric space of nonpositive curvature [53]. It can be viewed as a Riemannian extension of the logarithmic distance between positive scalars $d(x,y) = |\log(y/x)|$ to positive definite matrices [14, 36, 42] and possesses a number of remarkable symmetries that lie behind its utility in a variety of applications including brain-computer interfaces [9, 10, 22, 32], computer vision [30], medical imaging [7, 48], radar signal processing [6], statistical inference [50, 51], and machine learning [29, 66]. These symmetries include affine-invariance, i.e., invariance under congruence transformations: $d_2(X,Y) = d_2(AXA^T, AYA^T)$ for any invertible $A \in \mathrm{GL}(n,\mathbb{R})$, where A^T denotes the transpose of A [24, 41, 43, 44, 48, 59]. Another key symmetry satisfied by this metric is invariance under matrix inversion: $d_2(X,Y) = d_2(X^{-1}, Y^{-1})$.

While the Riemannian distance (2.1) has been the subject of significant research interest due to its symmetries and use in applications, it should be noted that it is only one member of a family of distance functions on $\mathbb{S}_{++}^n$ that enjoy the same properties. Indeed, the distances d_Φ on $\mathbb{S}_{++}^n$ defined as

$$(2.2) \qquad d_\Phi(X,Y) = \|\log X^{-1/2}YX^{-1/2}\|_\Phi,$$

where $\|\cdot\|_\Phi$ is an orthogonally invariant norm on the space of $n \times n$ symmetric matrices given by $\|Z\|_\Phi = \Phi(\lambda_1(Z), \cdots, \lambda_n(Z))$, $\lambda_i(Z)$ denote the eigenvalues of Z, and Φ is a symmetric gauge function on $\mathbb{R}^n$, are affine-invariant and inversion-invariant distances [12]. The symmetric gauge functions corresponding to the l_p-norms in $\mathbb{R}^n$ induce the Schatten p-norms $\|\cdot\|_\Phi$ for $1 \le p \le \infty$. If we take $\Phi(x_1, \cdots, x_n) = (\sum_i x_i^2)^{1/2}$, d_Φ yields the Riemannian distance function (2.1), whereas the choice of $\Phi(x_1, \cdots, x_n) = \max_i |x_i|$ yields the Thompson metric (1.4), which can equivalently be expressed as

$$(2.3)$$
$$d_\infty(X,Y) = \max_{1 \le i \le n} |\log \lambda_i(YX^{-1})| = \max\{\log \lambda_{\max}(YX^{-1}), \log \lambda_{\max}(XY^{-1})\}.$$

The form of the right-hand side of (2.3) is of computational significance since it only involves the computation of the largest generalized eigenvalues of the pairs (X,Y) and (Y,X). Thus, we see that the Thompson metric is both affine-invariant and inversion-invariant.

The space $\mathbb{S}_{++}^n$ is an open subset of the vector space of $n \times n$ real symmetric matrices and inherits a natural structure of a real differentiable manifold as a result.
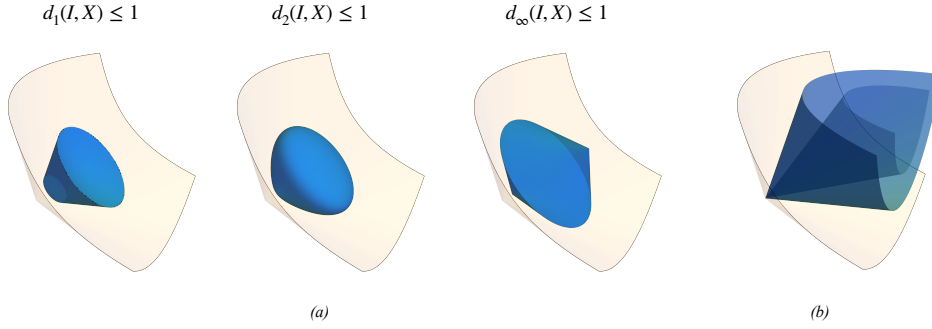
$d_1(I, X) \le 1$ $d_2(I, X) \le 1$ $d_\infty(I, X) \le 1$

(a)   (b)

FIG. 1. *(a) Unit balls $d_\Phi(I, X) \le 1$ in the affine-invariant geometries induced by the gauge functions Φ corresponding to the l_1-, l_2-, and l_∞-norms in $\mathbb{R}^2$ visualized as points in the interior of the closed convex cone $\{(a, b, c) \in \mathbb{R}^3 : a \ge 0, ac - b^2 \ge 0\}$, which we identify with the set of 2×2 SPD matrices. Note that d_∞ corresponds to the Thompson metric. (b) The sets $d_H(I, X) \le 1/2$ and $d_H(I, X) \le 1$ in Hilbert's projective metric applied to 2×2 SPD matrices visualized in $\mathbb{R}^3$.*

From a differential viewpoint, the distance functions d_Φ are induced by affine-invariant Finsler metrics on $\mathbb{S}^n_{++}$ given by the norm $\|d\Sigma\|_{\Sigma, \Phi} := \|\Sigma^{-1/2} d\Sigma \Sigma^{-1/2}\|_\Phi$ defined on the tangent space at $\Sigma \in \mathbb{S}^n_{++}$. In particular, the Thompson distance $d_T(X, Y)$ is induced by the norm

$$\|d\Sigma\|_\Sigma = \inf\{\alpha > 0 : -\alpha\Sigma \le d\Sigma \le \alpha\Sigma\} \tag{2.4}$$

and is recovered by minimizing the length

$$L[\gamma] = \int_0^1 \|\gamma'(t)\|_{\gamma(t)} dt \tag{2.5}$$

over all piecewise C^1 curves $\gamma : [a, b] \to \mathbb{S}^n_{++}$ with $\gamma(0) = X$ and $\gamma(1) = Y$ [46]. The Hilbert metric is recovered through a similar procedure by replacing the above norm with the semi-norm $\|d\Sigma\|_\Sigma = M(d\Sigma/\Sigma) - m(d\Sigma/\Sigma)$, where $M(d\Sigma/\Sigma) = \inf\{\lambda \in \mathbb{R} : d\Sigma \le \lambda\Sigma\}$ and $m(d\Sigma/\Sigma) = \sup\{\lambda \in \mathbb{R} : d\Sigma \ge \lambda\Sigma\}$ [47]. Various unit balls centered on the identity matrix in these affine-invariant geometries are depicted in Figure 1 in the case of 2×2 SPD matrices visualized as the interior of a convex cone $\{(a, b, c) \in \mathbb{R}^3 : a \ge 0, ac - b^2 \ge 0\}$.

3. Vector space embeddings. Consider the cone $K = \mathbb{R}^n_+$ in $\mathbb{R}^n$ and let $L : \text{int}(\mathbb{R}^n_+) \to \mathbb{R}^n$ denote the coordinatewise log function. Noting that $\log M(y/x) = \max_i(\log y_i - \log x_i)$ and substituting into (1.2), we find that

$$d_T(x, y) = \max\{\max_i(\log x_i - \log y_i), \max_i(\log y_i - \log x_i)\} = \|L(x) - L(y)\|_\infty, \tag{3.1}$$

for all $x, y \in \text{int}(\mathbb{R}^n_+)$. Thus, L is an isometry from $(\text{int}(\mathbb{R}^n_+), d_T)$ onto $(\mathbb{R}^n, \|\cdot\|_\infty)$. This result extends to polyhedral cones more generally, where a cone K in a vector space V is said to be polyhedral if it is the intersection of finitely many closed half-spaces. If P is a part of K, then there exists an isometric embedding from (P, d_T) to $(\mathbb{R}^m, \|\cdot\|_\infty)$ for some fixed m. Similarly, Hilbert geometry in a polyhedral cone can be related through an isometry to the geometry induced by the l_∞-norm in a vector space (see Lemma 2.2.2 and Proposition 2.2.3 of [35]).

However, such isometric embeddings in vector spaces do not generalize to Thompson and Hilbert geometries of non-polyhedral cones. Indeed a result by Nussbaum and Walsh in [47] shows that any finite collection of N points in a Thompson or Hilbert geometry in a part of an almost Archimedean cone K can be isometrically embedded in $\mathbb{R}_+^{N(N-1)}$ with respect to the Thompson or Hilbert metrics in K and $\mathbb{R}_+^{N(N-1)}$. The crucial point here is that the dimension of the target vector space is not fixed and depends on the number of points whose distances we seek to preserve. This result confirms that the Thompson and Hilbert geometries of non-polyhedral cones are quite non-trivial.

4. Geodesics. A geodesic path in a metric space (M, d) is a map $\gamma : I \to (M, d)$ such that $d(\gamma(s), \gamma(t)) = |s - t|$ for all $s, t \in I$, where $I \subseteq \mathbb{R}$ is a (possibly unbounded) interval. The image of a geodesic path is called a geodesic and a metric space is said to be a geodesic space if there exists a geodesic path joining any two points. Each of the metric spaces $(\mathbb{S}_{++}^n, d_\Phi)$ with d_Φ defined in (2.2) is a geodesic space. Indeed, the curve $\gamma : [0, 1] \to \mathbb{S}_{++}^n$ defined by

$$(4.1) \qquad \gamma(t) = X \#_t Y := X^{1/2}(X^{-1/2}YX^{-1/2})^t X^{1/2}$$

is a geodesic path from X to Y in each of these metric spaces and is unique provided that the geodesics in $\mathbb{R}^n$ induced by Φ are unique [12, 34]. Thus, uniqueness of geodesics in $(\mathbb{P}_n, d_\Phi)$ is inherited from $\mathbb{R}^n$ when Φ corresponds to the l_p-norms for $1 < p < \infty$, but not for $p = 1, \infty$.

In general, the Thompson metric does not admit unique geodesic paths between points. Indeed, a construction by Nussbaum in [46] describes a family of geodesics that generally consists of an infinite number of curves connecting a pair of points in a cone K. In particular, setting $\alpha := 1/M(x/y; K)$ and $\beta := M(y/x; K)$, the curve $\phi : [0, 1] \to K$ given by

$$(4.2) \qquad \phi(t; x, y) = x *_t y := \begin{cases} \left(\dfrac{\beta^t - \alpha^t}{\beta - \alpha}\right) y + \left(\dfrac{\beta\alpha^t - \alpha\beta^t}{\beta - \alpha}\right) x & \text{if } \alpha \neq \beta, \\ \alpha^t x & \text{if } \alpha = \beta, \end{cases}$$

is a geodesic path from x to y with respect to the Thompson metric. If we take K to be the cone of positive semidefinite matrices with interior int $K = \mathbb{S}_{++}^n$, then for a pair of points $X, Y \in \mathbb{S}_{++}^n$, we have $\beta = M(Y/X; K) = \lambda_{\max}(YX^{-1})$ and $\alpha = 1/M(X/Y; K) = \lambda_{\min}(YX^{-1})$. Therefore, $X *_t Y$ reduces to a linear combination of X and Y with coefficients that are nonlinear functions of the extreme generalized eigenvalues of (X, Y) and t.

PROPOSITION 4.1. *If $A \in \mathrm{GL}(n)$ and $X, Y \in \mathbb{S}_{++}^n$, then $(AXA^T) *_t (AYA^T) = A(X *_t Y)A^T$ for any $t \in \mathbb{R}$.*

Proof. The proof follows by noting that $(AYA^T)(AXA^T)^{-1} = AYX^{-1}A^{-1}$ and YX^{-1} have the same eigenvalues and using elementary algebra. $\qquad \square$

PROPOSITION 4.2. *If $X, Y \in \mathbb{S}_{++}^2$, then $X \#_t Y = X *_t Y$ for all $t \in [0, 1]$.*

Proof. By the density of dyadic rationals in the real line, it is sufficient to prove that $X \#_{1/2} Y = X *_{1/2} Y$ for arbitrary X and Y. Moreover, by affine-invariance and the uniqueness of the Riemannian geodesic, it is sufficient to prove that $I \#_{1/2} \Sigma = I *_{1/2} \Sigma$ for arbitrary $\Sigma \in \mathbb{S}_{++}^2$. This is equivalent to

$$\Sigma^{1/2} = \frac{1}{\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}}} \left(\Sigma + \sqrt{\lambda_{\max}\lambda_{\min}}\, I\right),$$

where λ_i denote the eigenvalues of Σ. However, this equality is seen to hold since $\Sigma^{1/2}$ is a 2×2 matrix with spectrum $\{\sqrt{\lambda_{\min}}, \sqrt{\lambda_{\max}}\}$ and characteristic equation $p(\lambda) = \lambda^2 - (\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}})\lambda + \sqrt{\lambda_{\max}\lambda_{\min}} = 0$, which is of course satisfied by $\Sigma^{1/2}$ by the Cayley-Hamilton theorem. $\square$

In general, of course, the geodesics $X \#_t Y$ and $X *_t Y$ do not agree in higher dimensions. Indeed, the two choices of geodesic agree in $\mathbb{S}^n_{++}$ if and only if the spectrum of YX^{-1} consists of at most two distinct eigenvalues [37]. It should be noted that even in $\mathbb{S}^2_{++}$ where the $\#_t$ and $*_t$ geodesics agree, the Thompson geodesic is still not unique. Indeed, it is shown in [37] that there exists a unique Thompson geodesic from X to Y in $\mathbb{S}^n_{++}$ if and only if the spectrum of YX^{-1} is contained in $\{\lambda, \lambda^{-1}\}$ for some fixed $\lambda > 0$. For example, the following construction describes another geodesic $X \diamond_t Y$ of $(\mathbb{S}^n_{++}, d_T)$ from X to Y when $\lambda_{\max}(YX^{-1}) \neq \lambda_{\min}(YX^{-1})$:

$$(4.3) \qquad X \diamond_t Y = \begin{cases} \dfrac{\lambda_{\max}^t - \lambda_{\max}^{-t}}{\lambda_{\max} - \lambda_{\max}^{-1}} Y + \dfrac{\lambda_{\max}^{1-t} - \lambda_{\max}^{t-1}}{\lambda_{\max} - \lambda_{\max}^{-1}} X, & \lambda_{\max}\lambda_{\min} \geq 1 \\[2em] \dfrac{\lambda_{\min}^t - \lambda_{\min}^{-t}}{\lambda_{\min} - \lambda_{\min}^{-1}} Y + \dfrac{\lambda_{\min}^{1-t} - \lambda_{\min}^{t-1}}{\lambda_{\min} - \lambda_{\min}^{-1}} X, & \lambda_{\max}\lambda_{\min} \leq 1, \end{cases}$$

where $\lambda_{\max}$ and $\lambda_{\min}$ refer to the corresponding eigenvalues of YX^{-1} [37, 46]. A depiction of these various geodesics for an example computed in the set $\mathbb{S}^2_{++}$ visualized as the interior of a cone in $\mathbb{R}^3$ is shown in Figure 2. We thus note that $*_t$ is special among the Thompson geodesics constructed by Nussbaum [46] in that it coincides with the Riemannian geodesic for 2×2 SPD matrices. The $*_t$ geodesic satisfies other desirable properties that do not generally hold for other Thompson geodesics such as joint homogeneity, which is also satisfied by the Riemannian geodesic in all dimensions.

PROPOSITION 4.3 (Joint homogeneity). *Let* $X_1, X_2 \in \mathbb{S}^n_{++}$. *If* μ_1 *and* μ_2 *are positive scalars, then*

$$(4.4) \qquad (\mu_1 X_1) *_t (\mu_2 X_2) = \mu_1^{1-t} \mu_2^t (X_1 *_t X_2)$$

for any $t \in \mathbb{R}$.

Proof. The result follows from the equality $\lambda_i \left((\mu_2 X_2 (\mu_1 X_1)^{-1} \right) = \frac{\mu_2}{\mu_1} \lambda_i (X_2 X_1^{-1})$ and substitution into the expression for $(\mu_1 X_1) *_t (\mu_2 X_2)$ arising from (4.2). $\square$

COROLLARY 4.4. *If* $X_1, X_2 \in \mathbb{S}^n_{++}$, *then* $(\mu_1 X_1) *_{\frac{1}{2}} (\mu_2 X_2) = \sqrt{\mu_1 \mu_2} (X_1 *_{\frac{1}{2}} X_2)$ *for any positive scalars* μ_1 *and* μ_2.

We will view the $*_t$ Thompson geodesic (4.2) as a distinguished geodesic of (K, d_T), which makes the resulting structure a geodesic space. For the remainder of this paper, by "Thompson geodesic" we refer specifically to the $*_t$ geodesic unless stated otherwise.

4.1. Metric inequalities in Hilbert and Thompson geometries. The following theorem from [47] establishes two important inequalities in the Thompson and Hilbert geometries of convex cones that provide insight into the curvature properties of these geometries. These inequalities can be viewed as describing how far the Thompson and Hilbert geometries are from being non-positively curved.

THEOREM 4.5 (Theorems 1.1 and 1.2 of [47]). *Let* K *be an almost Archimedean cone and* $u, x, y \in K$ *be in the same part of* K. *Suppose that* $0 < s < 1$ *and* $R > 0$,
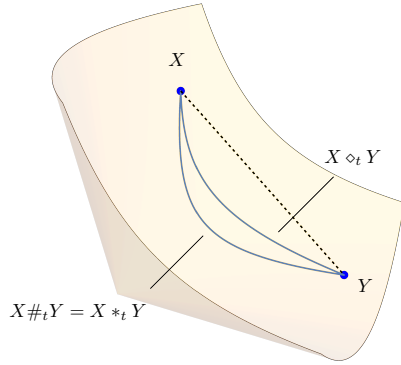
FIG. 2. *Two geodesics in $(\mathbb{S}^2_{++}, d_T)$ between a pair of matrices visualized as points in the interior of the closed convex cone $\{(a, b, c) \in \mathbb{R}^3 : a > 0, ac - b^2 > 0\}$. The dashed straight line between the endpoints does not represent a geodesic.*

and that $d_H(u, x) \leq R$ and $d_H(u, y) \leq R$. If the linear span of $\{u, x, y\}$ is 1- or 2-dimensional, then $d_T(u *_s x, u *_s y) \leq sd_T(x, y)$ and $d_H(u *_s x, u *_s y) \leq sd_H(x, y)$. In general,

$$d_T(u *_s x, u *_s y) \leq \left[\frac{2(1 - e^{-Rs})}{1 - e^{-R}} - s \right] d_T(x, y) \tag{4.5}$$

$$d_H(u *_s x, u *_s y) \leq \left(\frac{1 - e^{-Rs}}{1 - e^{-R}} \right) d_H(x, y). \tag{4.6}$$

A remarkable feature of Theorem 4.5 is that it ties the Hilbert and Thompson geometries of a convex cone together and suggests that one should consider both of these metrics in geometric analysis in convex cones rather than making a choice of one over the other. A consequence of Theorem 4.5 is that both the Hilbert and Thompson geometries are semihyperbolic in the sense of Alonso and Bridson [4].

COROLLARY 4.6. $\mathbb{S}^n_{++}$ *is semihyperbolic when endowed with Hilbert's projective metric or Thompson's part metric.*

4.2. Determinants along geodesics. Here we consider the evolution of determinants along various interpolations of SPD matrices, including Euclidean, Riemannian, and Thompson geodesics. An application in which the evolution of determinants along interpolations is of practical significance is diffusion tensor imaging (DTI), which is an imaging modality used to produce non-invasive reconstructions of brain tissue connectivity. DTI is based on the assumption that the motion of water molecules in each voxel of the image is well approximated by Brownian motion. This Brownian motion is characterized by an SPD matrix, called the diffusion tensor [11]. Interpolating diffusion tensors is a basic operation in many data processing algorithms in DTI. It has been well documented that Euclidean interpolation and averaging of tensors generally results in a tensor swelling effect [7, 48] where the determinant of the Euclidean average of SPD matrices is often larger than the determinant of the original SPD matrices. This phenomenon is problematic as the determinant of the diffusion tensor is a measure of the dispersion of the local diffusion process, whose increase upon averaging is physically unrealistic. To overcome this difficulty, the affine-invariant and Log-Euclidean Riemannian geometries are often used in data processing tasks in
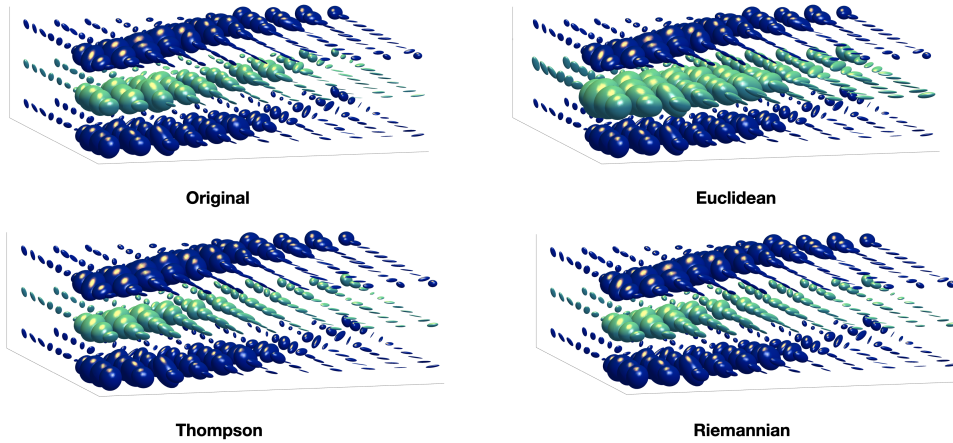
Original

Euclidean

Thompson

Riemannian

FIG. 3. *Reconstructions of the middle layer of an image consisting of three layers of ellipsoids representing diffusion tensors using Euclidean, Riemannian, and Thompson geodesic interpolations. The Euclidean interpolation exhibits an undesirable swelling effect. The Riemannian and Thompson interpolations generate similar results and are more faithful to the original data.*

DTI such as interpolation and regularization [7, 48]. In either of these Riemannian geometries, the swelling effect is eliminated and the determinants along the geodesics between two SPD matrices with the same determinant are constant. Indeed, for $X, Y \in \mathbb{S}^n_{++}$, we have

$$\det(X \#_t Y) = \det\left(X^{1/2}(X^{-1/2}YX^{-1/2})^t X^{1/2}\right) = (\det X)^{1-t}(\det Y)^t.$$

In particular,

(4.7) $$\det(X \#_{1/2} Y) = \sqrt{\det(XY)}.$$

Thus, if $\det X = \det Y$, then $\det(X \#_t Y) = \det X = \det Y$ for all $t \in \mathbb{R}$. More generally, we see that the logarithm of the determinant interpolates linearly along the affine-invariant Riemannian geodesics.

These observations raise the question of how determinants evolve along the $*_t$ Thompson geodesics. In Figure 3, we compare interpolations using three different geodesics with real data extracted from [3] in the context of DTI. Each image depicts three layers of ellipsoids corresponding to positive definite matrices. The middle layer of the original image consisting of real data was removed and reconstructed using Euclidean, Riemannian, and Thompson interpolations of data at voxels in the top and bottom layers. We observe that in the case of 3×3 matrices, the Riemannian and Thompson interpolations produce very similar reconstructions. The swelling effect in the case of the Euclidean interpolation is also clearly visible.

Further simulations suggest that in contrast to the Euclidean case, we seem to encounter a 'shrinkage phenomenon' when using the $*_t$ Thompson geodesics, whereby determinants along the geodesic between two SPD matrices with the same determinant are reduced in value. While this effect appears modest in low dimensions, it can become quite dramatic in higher dimensions as seen in Figure 4. In the following proposition, we provide mathematical explanations for these observations and pre-
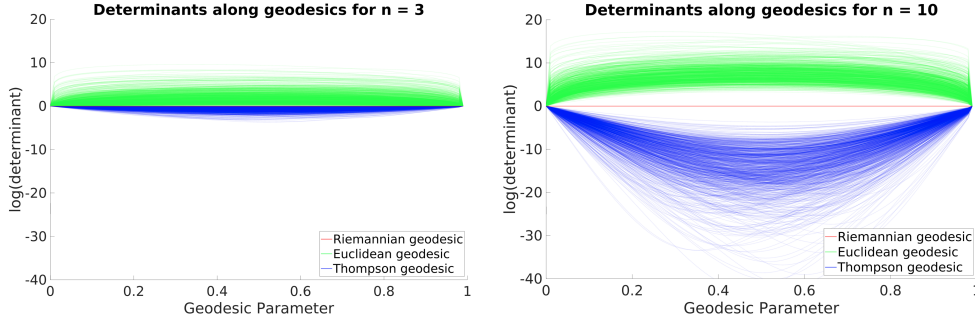
FIG. 4. *Plots of the logarithm of the determinants along Euclidean, Riemannian, and Thompson geodesics connecting 1,000 pairs of randomly generated $n \times n$ SPD matrices of unit determinant for $n = 3$ and $n = 10$. We observe that the determinants along the Riemannian geodesics are constant and equal to one as expected. The determinants along Euclidean interpolations exhibit swelling, whereas an opposite shrinkage effect is observed for interpolations along the Thompson geodesics. Both of these effects are magnified in higher dimensions.*

cisely characterize the circumstances under which the shrinkage phenomenon is not present along interpolations of a pair of SPD matrices.

PROPOSITION 4.7. *Let $X, Y \in \mathbb{S}_{++}^n$ and λ_i denote the eigenvalues of YX^{-1}. Then*

$$(4.8) \qquad \det\left(\frac{X+Y}{2}\right) = \sqrt{\det(XY)} \prod_{i=1}^{n}\left(\frac{\sqrt{\lambda_i} + \frac{1}{\sqrt{\lambda_i}}}{2}\right) \geq \sqrt{\det(XY)}$$

where equality holds if and only if $X = Y$. Furthermore,

$$\det\left(X *_{1/2} Y\right) = \sqrt{\det(XY)} \prod_{i=1}^{n}\left[\frac{1}{\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}}}\left(\sqrt{\lambda_i} + \frac{\sqrt{\lambda_{\max}\lambda_{\min}}}{\sqrt{\lambda_i}}\right)\right]$$

$$(4.9) \qquad\qquad\qquad \leq \sqrt{\det(XY)}$$

where equality holds if and only if $\lambda_i \in \{\lambda_{\min}, \lambda_{\max}\}$ for $i = 1, \cdots, n$, where $\lambda_{\max}$ and $\lambda_{\min}$ denote the maximum and minimum eigenvalues of YX^{-1}, respectively.

Proof. $X^{-1/2}YX^{-1/2}$ admits a diagonalization $X^{-1/2}YX^{-1/2} = QDQ^T$, where

Q is an orthogonal matrix and $D = \mathrm{diag}(\lambda_1, \cdots, \lambda_n)$. We have

$$\det\left(\frac{X+Y}{2}\right) = \det\left(X^{1/2}\frac{(I + X^{-1/2}YX^{-1/2})}{2}X^{1/2}\right)$$

$$= \det(X)\det\left(\frac{I + QDQ^T}{2}\right)$$

$$= \det(X)\det\left(\frac{I + D}{2}\right)$$

$$= \det(X)\det(D^{1/2})\det\left(\frac{D^{1/2} + D^{-1/2}}{2}\right)$$

$$= \sqrt{\det(XY)}\det\left(\frac{D^{1/2} + D^{-1/2}}{2}\right)$$

$$= \sqrt{\det(XY)}\prod_{i=1}^{n}\left(\frac{\sqrt{\lambda_i} + \frac{1}{\sqrt{\lambda_i}}}{2}\right),$$

where we have repeatedly used the fact that $\det(AB) = \det(A)\det(B)$ for square $n \times n$ matrices A and B. The inequality (4.8) now follows by noting the identity

$$\frac{1}{2}\left(\sqrt{\lambda_i} + \frac{1}{\sqrt{\lambda_i}}\right) \geq 1,$$

where equality holds if and only if $\lambda_i = 1$. Thus, each term in the product in (4.8) results in a scaling of the geometric mean $\sqrt{\det(XY)}$ by a factor greater than or equal to 1, which contributes to the swelling effect. For the swelling effect to completely disappear, we would need $\lambda_i = 1$ for each i, which is only possible in the trivial case when $X = Y$.

We now apply a similar analysis to $\det\left(X *_{1/2} Y\right)$ to find

$$\det\left(X *_{1/2} Y\right) = \det\left(\frac{1}{\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}}}(Y + \sqrt{\lambda_{\max}\lambda_{\min}}X)\right)$$

$$= \det(X)\det\left(\frac{1}{\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}}}(X^{-1/2}YX^{-1/2} + \sqrt{\lambda_{\max}\lambda_{\min}}I)\right)$$

$$= \det(X)\det\left(\frac{1}{\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}}}(QDQ^T + \sqrt{\lambda_{\max}\lambda_{\min}}I)\right)$$

$$= \det(X)\det\left(\frac{1}{\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}}}(D + \sqrt{\lambda_{\max}\lambda_{\min}}I)\right)$$

$$= \sqrt{\det(XY)}\det\left(\frac{1}{\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}}}(D^{1/2} + \sqrt{\lambda_{\max}\lambda_{\min}}D^{-1/2})\right)$$

$$= \sqrt{\det(XY)}\prod_{i=1}^{n}\left[\frac{1}{\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}}}\left(\sqrt{\lambda_i} + \frac{\sqrt{\lambda_{\max}\lambda_{\min}}}{\sqrt{\lambda_i}}\right)\right].$$

Now, the inequality (4.9) follows from

$$\frac{1}{\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}}}\left(\sqrt{\lambda_i} + \frac{\sqrt{\lambda_{\max}\lambda_{\min}}}{\sqrt{\lambda_i}}\right) - 1 = \frac{(\sqrt{\lambda_{\max}} - \sqrt{\lambda_i})(\sqrt{\lambda_{\min}} - \sqrt{\lambda_i})}{\sqrt{\lambda_i}(\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}})} \leq 0,$$

for each $i = 1, \cdots, n$, which holds since $\lambda_{\min} \leq \lambda_i \leq \lambda_{\max}$. Moreover, we see that equality holds if and only if $\lambda_i = \lambda_{\min}$ or $\lambda_i = \lambda_{\max}$ for all i. Thus, we see that each
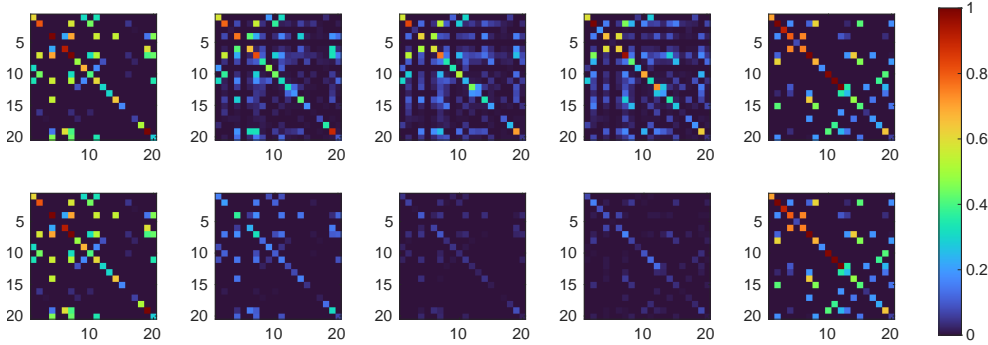
FIG. 5. *Points along the Riemannian (top row) and Thompson (bottom row) geodesic interpolations of a pair of* 20×20 *SPD matrices with 68 non-zero entries. The matrices represent equidistant points along the geodesics as measured by the corresponding metric. Each pixel is colored according to the value of the corresponding matrix element. We observe that in the Riemannian case, most of the matrix elements along the interpolation are non-zero.*

term in the product results in a scaling of the geometric mean $\sqrt{\det(XY)}$ by a factor less than or equal to 1, which contributes to the shrinkage effect. As the dimension increases, we expect this product to include more and more factors smaller than 1, which explains the intensification of this shrinkage effect with the dimension n. The shrinkage is fully eliminated if and only if each of the factors in the product is equal to 1, which occurs precisely if $\lambda_i \in \{\lambda_{\min}, \lambda_{\max}\}$ for all $i = 1, \cdots, n$. ☐

4.3. Sparsity preservation. Sparse matrices are matrices whose non-zero elements form a relatively small proportion of the matrix entries. They appear in many areas of applied mathematics and engineering including the numerical analysis of partial differential equations, network theory, and machine learning. They arise naturally in multi-agent systems that include relatively few pairwise interactions. From a computational perspective, sparsity is an important property due to the existence of specialized algorithms and data structures that enable the efficient storage and manipulation of large sparse matrices [26].

An interesting property of the $*_t$ Thompson geodesic is that it preserves sparsity. That is, if X and Y are sparse SPD matrices, then $X *_t Y$ is sparse for every $t \in \mathbb{R}$. This is simply a consequence of $X *_t Y$ being a linear combination of X and Y for any fixed t. In contrast, the Riemannian geodesic $X \#_t Y$, whose construction involves computing principal matrix square roots, matrix products, and matrix inverses, does not preserve sparsity. Thus, the use of Riemannian interpolation to process large sparse SPD matrices may be problematic. For instance, kernel matrices in machine learning are often built as sparse matrices to facilitate the analysis of large datasets. Applying the standard affine-invariant Riemannian geometry to process such SPD matrices will typically corrupt the sparse structure, potentially resulting in intractable computations. See Figure 5 for a visualization of Riemannian and $*_t$ Thompson geodesic interpolations of a pair of 20×20 SPD matrices with 68 non-zero entries.

5. The distance between the Riemannian and Thompson geodesics. We have seen that the affine-invariant Riemannian and $*_t$ Thompson geodesics coincide in $\mathbb{S}^2_{++}$. Moreover, simulations in $\mathbb{S}^3_{++}$ suggest that while the two geodesics no longer agree for 3×3 SPD matrices, they result in qualitatively similar interpolations. This

raises the question of how close these two geodesics may be in higher dimensions. To investigate this, we consider the affine-invariant Riemannian distance between the midpoints of the two geodesics connecting a pair of arbitrary SPD matrices X and Y. To simplify the notation, we denote the Riemannian and $*_t$ Thompson geodesic midpoints by $X \# Y$ and $X * Y$, respectively.

PROPOSITION 5.1. *Let $X, Y \in \mathbb{S}_{++}^n$ and denote the eigenvalues of $Y X^{-1}$ by λ_i. Then,*

$$(5.1) \quad d_2(X \# Y, X * Y) = \left[\sum_{i=1}^n \log^2 \left(\frac{1}{\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}}} \left(\sqrt{\lambda_i} + \frac{\sqrt{\lambda_{\max} \lambda_{\min}}}{\sqrt{\lambda_i}} \right) \right) \right]^{1/2},$$

where $\lambda_{\min} = \min_i \lambda_i$, and $\lambda_{\max} = \max_i \lambda_i$. Furthermore,

$$(5.2) \qquad 0 \le d_2(X \# Y, X * Y) \le \sqrt{n-2} \log \left(\cosh \frac{d_\infty(X, Y)}{2} \right),$$

where $d_2(X \# Y, X * Y) = 0$ if and only if $\lambda_i \in \{\lambda_{\min}, \lambda_{\max}\}$ for $i = 1, \cdots, n$, and the upper bound in (5.2) is attained if and only if $\lambda_{\min} = 1/\lambda_{\max}$ and $\lambda_i = 1$ for the remaining $n - 2$ eigenvalues of $Y X^{-1}$.

Proof. By affine-invariance of d_2, we have

$$d_2(X \# Y, X * Y) = d_2 \left(X^{1/2} (X^{-1/2} Y X^{-1/2})^{1/2} X^{1/2}, \frac{Y + \sqrt{\lambda_{\max} \lambda_{\min}} X}{\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}}} \right)$$

$$= d_2 \left((X^{-1/2} Y X^{-1/2})^{1/2}, \frac{X^{-1/2} Y X^{-1/2} + \sqrt{\lambda_{\max} \lambda_{\min}} I}{\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}}} \right)$$

$$(5.3) \qquad = \left\| \log \left(\frac{1}{\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}}} \left(\Sigma^{1/2} + \sqrt{\lambda_{\max} \lambda_{\min}} \Sigma^{-1/2} \right) \right) \right\|_F,$$

where $\Sigma = X^{-1/2} Y X^{-1/2}$ and $\| \cdot \|_F$ denotes the Frobenius norm. By noting that Σ admits a diagonalization $\Sigma = Q D Q^T$, where Q is an orthogonal matrix and $D = \text{diag}(\lambda_1, \cdots, \lambda_n)$, and considering the eigenvalues of the principal matrix logarithm in (5.3), we arrive at (5.1).

It follows from (5.1) that $d_2(X \# Y, X * Y) = 0$ if and only if

$$\frac{1}{\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}}} \left(\sqrt{\lambda_i} + \frac{\sqrt{\lambda_{\max} \lambda_{\min}}}{\sqrt{\lambda_i}} \right) = 1$$

for $i = 1, \cdots, n$, which holds if and only if $(\sqrt{\lambda_i} - \sqrt{\lambda_{\max}})(\sqrt{\lambda_i} - \sqrt{\lambda_{\min}}) = 0$. That is, $d_2(X \# Y, X * Y) = 0$ if and only if $\lambda_i \in \{\lambda_{\min}, \lambda_{\max}\}$ for $i = 1, \cdots, n$.

Now denote by r the Thompson distance $d_\infty(X, Y) = \max_i |\log \lambda_i|$ between X and Y. It follows that $\lambda_{\max} = e^r$ or $\lambda_{\min} = e^{-r}$. Suppose that $\lambda_{\max} = e^r$ and note that (5.1) can be expressed as

$$(5.4) \qquad d_2(X \# Y, X * Y) = \left(\sum_{i=1}^n \rho(\lambda_i) \right)^{1/2},$$

where

$$(5.5) \qquad \rho(x) = \log^2 \left(\frac{1}{\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}}} \left(\sqrt{x} + \frac{\sqrt{\lambda_{\max} \lambda_{\min}}}{\sqrt{x}} \right) \right),$$

for $x \in [\lambda_{\min}, \lambda_{\max}]$. To establish the upper bound in (5.2), we seek to maximize (5.1) subject to the constraint that $e^{-r} \leq \lambda_{\min} \leq \lambda_i \leq \lambda_{\max} = e^r$. For a given $\lambda_{\min}$, this is achieved by maximizing $\rho(\lambda_i)$ for every choice λ_i not corresponding to the extreme eigenvalues. ρ is a non-negative continuous function on $[\lambda_{\min}, \lambda_{\max}]$ that is smooth on $(\lambda_{\min}, \lambda_{\max})$ and satisfies $\rho'(x) = 0$ if and only if $x = \sqrt{\lambda_{\min}\lambda_{\max}}$. Since $\rho(\lambda_{\min}) = \rho(\lambda_{\max}) = 0$, it follows that (5.4) is maximized for a given $\lambda_{\min}$ and $\lambda_{\max}$ if and only if all remaining eigenvalues satisfy $\lambda_i = \sqrt{\lambda_{\min}\lambda_{\max}}$. For such a distribution of eigenvalues, (5.4) reduces to

$$(5.6) \qquad d_2(X \# Y, X * Y) = \left((n-2) \log^2 \left(\frac{2 e^{r/4} \lambda_{\min}^{1/4}}{e^{r/2} + \lambda_{\min}^{1/2}} \right) \right)^{1/2},$$

which in turn is maximized as a function of $\lambda_{\min}$ subject to the constraint $e^{-r} \leq \lambda_{\min} \leq \lambda_{\max} = e^r$ when $\lambda_{\min} = e^{-r}$, at which point

$$(5.7) \quad d_2(X \# Y, X * Y) = \left((n-2) \log^2 \left(\frac{2}{e^{r/2} + e^{-r/2}} \right) \right)^{1/2} = \sqrt{n-2} \, \log \left(\cosh \frac{r}{2} \right).$$

Repeating the analysis by starting with the assumption that $\lambda_{\min} = e^{-r}$ instead of $\lambda_{\max} = e^r$ would yield the same conclusion. That is, the upper bound in (5.2) holds and is attained if and only if $\lambda_{\min} = e^{-r}$, $\lambda_{\max} = e^r$, and $\lambda_i = \sqrt{\lambda_{\min}\lambda_{\max}} = 1$ for the remaining $n-2$ eigenvalues of YX^{-1}, where $r = d_\infty(X, Y)$. □

While we see that the bounds in (5.2) are attainable, it is generally unlikely that two SPD matrices X and Y will have the generalized eigenvalues required to attain these bounds. Indeed, the required distributions of generalized eigenvalues become increasingly unlikely as the dimension n grows. To develop a more practical sense of the typical values of the distance $d_2(X \# Y, X * Y)$, we can compute this distance for a large number of randomly generated SPD matrices. As (5.1) provides an expression for $d_2(X \# Y, X * Y)$ in terms of the eigenvalues of YX^{-1}, it allows us to efficiently simulate a large number of matrix distance computations by sampling vectors of generalized eigenvalues instead of generating a large number of pairs of SPD matrices and computing the corresponding generalized eigenvalues. Thus, we consider the 'normalized' distance function

$$(5.8) \quad f(\boldsymbol{\lambda}) := \frac{\sqrt{\sum_i \log^2 \left(\frac{1}{\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}}} \left(\sqrt{\lambda_i} + \frac{\sqrt{\lambda_{\max}\lambda_{\min}}}{\sqrt{\lambda_i}} \right) \right)}}{\max_i |\log \lambda_i|} = \frac{d_2(X \# Y, X * Y)}{d_\infty(X, Y)}$$

where f is a positive-valued function of $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_n)$. The results in Figure 6 were produced by generating 100,000 points $\boldsymbol{\lambda}$ for values of $r = \max_i |\log \lambda_i|$ from 0 to 100 and computing the corresponding $f(\boldsymbol{\lambda})$ in the cases $n = 4$ and $n = 20$. The solid blue curve above each plot indicates the corresponding upper bound as a function of r. We note that in either case most points concentrate away from the bounds and increasingly so in higher dimensions. By (5.2), we have

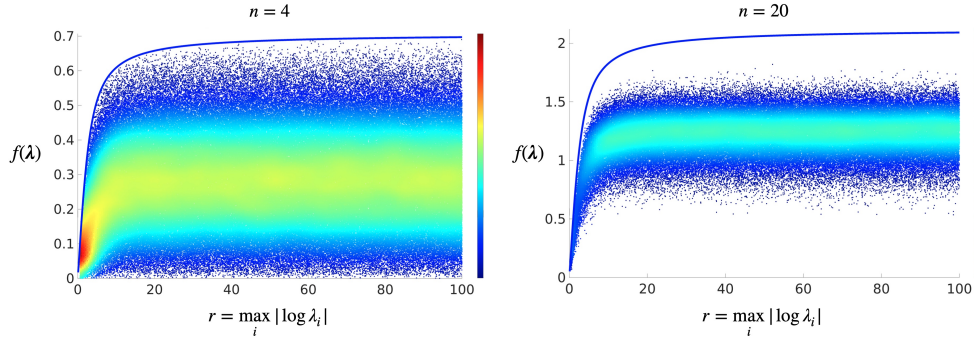$$(5.9) \qquad\qquad 0 \leq f(\boldsymbol{\lambda}) \leq \sqrt{n-2} \, \frac{\log \left(\cosh \frac{r}{2} \right)}{r}.$$

FIG. 6. *Plots of $f(\boldsymbol{\lambda})$ (5.8) against $r = \max_i |\log \lambda_i|$ for 100,000 samples $\boldsymbol{\lambda} \in \mathbb{R}_+^n$ for $n = 4$ and $n = 20$. The solid blue curves represent the upper bounds of $f(\boldsymbol{\lambda})$ given in (5.9). The color scheme is used to indicate the density of plot points $(r, f(\boldsymbol{\lambda}))$.*

6. An iterative mean of SPD matrices based on Thompson geometry. A crucial step in developing a scalable computational framework for performing analysis and statistics on SPD-valued data using extreme generalized eigenvalues is to provide a suitable definition for the mean of a collection of k SPD matrices whose computation can be based primarily on finding a sequence of extreme generalized eigenvalues. In this section, we introduce such a notion for any finite collection of SPD matrices through an iterative algorithm based on Thompson geodesics and prove that it yields a well-defined and unique point in each case. Furthermore, we highlight and prove a number of desirable properties that are satisfied by this novel iterative mean in subsection 6.4.

Specifically, given any finite ordered set $\mathcal{P} = (Y_1, \cdots, Y_k) \subset \mathbb{S}_{++}^n$, we generate a sequence of SPD matrices $(X_i)_{i \geq 1}$ from an arbitrary initialization $X_1 \in \mathbb{S}_{++}^n$ according to Algorithm 6.1. We will then prove that any sequence generated by this algorithm converges to a point X^* that is independent of the choice of initialization X_1 and the ordering of the Y_j, and thus can be viewed as a mean of the set of points $\{Y_j\}$.

Algorithm 6.1 Generate iterative sequence of SPD matrices $(X_i)_{i \geq 1}$ from an initial point X_1 and the finite ordered set $\mathcal{P} = (Y_1, \cdots, Y_k) \subset \mathbb{S}_{++}^n$

1: **for** $i \geq 1$ **do**
2:     Set $j \equiv i \mod k$ for $1 \leq j \leq k$
3:     Define $X_{i+1} = X_i *_{\frac{1}{i+1}} Y_j$
4: **end for**
5: **return** $(X_1, X_2, X_3, \cdots)$.

6.1. Mathematical preliminaries. We begin by presenting a number of technical lemmas that are used in the proof of our main theorem. First note that the Thompson geodesic (4.2) in $\mathbb{S}_{++}^n$ can be written as

$$(6.1) \qquad\qquad X *_t Y = \varphi_{\alpha\beta}(t)Y + \psi_{\alpha\beta}(t)X,$$

where

$$\varphi_{\alpha\beta}(t) = \begin{cases} \frac{\beta^t - \alpha^t}{\beta - \alpha} & \text{if } \beta > \alpha \\ t\alpha^{t-1} & \text{if } \beta = \alpha \end{cases} \qquad \psi_{\alpha\beta}(t) = \begin{cases} \frac{\beta\alpha^t - \alpha\beta^t}{\beta - \alpha} & \text{if } \beta > \alpha \\ (1-t)\alpha^t & \text{if } \beta = \alpha \end{cases}$$

(6.2)
$$\frac{d\varphi_{\alpha\beta}}{dt}(0) = \begin{cases} \frac{\log\beta - \log\alpha}{\beta - \alpha} & \text{if } \beta > \alpha \\ \frac{1}{\alpha} & \text{if } \beta = \alpha \end{cases} \qquad \frac{d\psi_{\alpha\beta}}{dt}(0) = \begin{cases} \frac{\beta\log\alpha - \alpha\log\beta}{\beta - \alpha} & \text{if } \beta > \alpha \\ \log\alpha - 1 & \text{if } \beta = \alpha \end{cases}.$$

for $\alpha = \lambda_{\min}(YX^{-1})$ and $\beta = \lambda_{\max}(YX^{-1})$.

LEMMA 6.1 (Geodesic consistency). *For $X, Y \in \mathbb{S}_{++}^n$ and $s, t \in [0,1]$*

(6.3)
$$X *_t Y = Y *_{1-t} X,$$

(6.4)
$$X *_s (X *_t Y) = X *_{st} Y$$

and

(6.5)
$$(X *_s Y) *_t Y = X *_{s+t-st} Y.$$

Proof. (6.3) follows from the observation that for $0 < \alpha \le \beta$,

$$\varphi_{\alpha\beta}(t) = \psi_{\beta^{-1}\alpha^{-1}}(1-t).$$

So writing $\alpha = \lambda_{\min}(YX^{-1})$, $\beta = \lambda_{\max}(YX^{-1})$ we have by (6.1)

$$X *_t Y = \varphi_{\alpha\beta}(t)Y + \psi_{\alpha\beta}(t)X = \varphi_{\beta^{-1}\alpha^{-1}}(1-t)X + \psi_{\beta^{-1}\alpha^{-1}}(1-t)Y = Y *_{1-t} X.$$

For (6.4), suppose $\beta > \alpha$. Then

$$\begin{aligned} \lambda_{\max}((X *_t Y)X^{-1}) &= \lambda_{\max}((\varphi_{\alpha\beta}(t)Y + \psi_{\alpha\beta}(t)X)X^{-1}) \\ &= \varphi_{\alpha\beta}(t)\lambda_{\max}(YX^{-1}) + \psi_{\alpha\beta}(t) \\ &= \frac{\beta^t - \alpha^t}{\beta - \alpha}\beta + \frac{\beta\alpha^t - \alpha\beta^t}{\beta - \alpha} \\ &= \beta^t. \end{aligned}$$

Similarly,

$$\lambda_{\min}((X *_t Y)X^{-1}) = \alpha^t.$$

These also hold when $\beta = \alpha$, and the proof is easier. Now use (6.1) substituting the variables appropriately, or alternatively use [46, Equation 1.25], to get (6.4). For (6.5),

$$(X *_s Y) *_t Y = Y *_{1-t} (Y *_{1-s} X) = Y *_{(1-t)(1-s)} X = X *_{s+t-st} Y$$

by (6.3) and (6.4). $\qquad\square$

For $i \in \mathbb{N}$ and $p \in \mathbb{Z}_{\ge 0}$ define the maps $S_i : \mathbb{S}_{++}^n \to \mathbb{S}_{++}^n$ and $T_p : \mathbb{S}_{++}^n \to \mathbb{S}_{++}^n$ by

$$S_i : X \mapsto X *_{\frac{1}{i+1}} Y_j,$$

where $0 \leq j \leq k$ is such that $j \equiv i \mod k$, and

$$T_p : X \mapsto \left(\ldots (X *_{\frac{1}{pk+2}} Y_1) *_{\frac{1}{pk+3}} \ldots \right) *_{\frac{1}{(p+1)k+1}} Y_k.$$

So if we pick an initialization $X_1 \in \mathbb{S}^n_{++}$ for the algorithm, we have

$$X_{i+1} = S_i(X_i) \quad \text{and} \quad X_{(p+1)k+1} = T_p(X_{pk+1}).$$

We will later need the following observation.

LEMMA 6.2. *If $c > 0$, $i \in \mathbb{N}$ and $p \in \mathbb{Z}_{\geq 0}$, then*

$$S_i(cX) = c^{\frac{i}{i+1}} S_i(X) \quad \text{and} \quad T_p(cX) = c^{\frac{pk+1}{(p+1)k+1}} T_p(X).$$

Proof. Observe that for $0 < \alpha \leq \beta$ and $c > 0$,

$$\varphi_{(\alpha/c)(\beta/c)}(t) = c^{1-t} \varphi_{\alpha\beta}(t) \quad \text{and} \quad \psi_{(\alpha/c)(\beta/c)}(t) = c^{-t} \psi_{\alpha\beta}(t)$$

and use the expression (6.1). □

For $1 \leq j \leq k$ and $X \in \mathbb{S}^n_{++}(n)$, let $\varphi^j_X = \varphi_{\alpha\beta}$ and $\psi^j_X = \psi_{\alpha\beta}$ where $\alpha = \lambda_{\min}(Y_j X^{-1})$ and $\beta = \lambda_{\max}(Y_j X^{-1})$. We will also write $m^j_X = \frac{d\varphi^j_X}{dt}(0)$ and $o^j_X = \frac{d\psi^j_X}{dt}(0)$. Note that by (6.2) we see that m^j_X is always positive, while o^j_X may be positive, negative or zero.

From now on we will write $\|\cdot\|$ for the Euclidean (i.e. Frobenius) norm on matrices.

LEMMA 6.3. *For $1 \leq j \leq k$, the maps from $\mathbb{S}^n_{++}$ to $\mathbb{R}$ given by*

$$X \mapsto \varphi^j_X, \quad X \mapsto \psi^j_X, \quad X \mapsto \frac{d^2\varphi^j_X}{dt^2}, \quad X \mapsto \frac{d^2\psi^j_X}{dt^2}$$

are continuous. Moreover, if $\mathcal{K} \subset \mathbb{S}^n_{++}(n)$ is a compact set, the maps

$$X \mapsto m^j_X, \quad X \mapsto o^j_X$$

from $\mathcal{K}$ to $\mathbb{R}$ are Lipschitz with respect to the metric induced by $\|\cdot\|$ on $\mathcal{K}$, and in particular they are continuous.

Proof. We will show that $X \mapsto m^j_X$ is locally Lipschitz. The proof for $X \to o^j_X$ is analogous and the continuity of the other maps can also be shown in a similar fashion.

$X \mapsto m^j_X$ can be expressed as the composition of the five maps

$$\mathcal{K} \xrightarrow{\text{I}} \mathcal{K} \times \mathcal{K} \xrightarrow{\text{II}} \mathcal{K} \times \mathcal{K} \xrightarrow{\text{III}} \mathbb{R}_{>0} \times \mathbb{R}_{>0} \xrightarrow{\text{IV}} \mathbb{R}_{>0} \times \mathbb{R}_{>0} \xrightarrow{\text{V}} \mathbb{R}_{>0}$$

$$X \xmapsto{\text{I}} (X, X^{-1}) \xmapsto{\text{II}} (Y_j X^{-1}, XY_j^{-1})$$

$$\xmapsto{\text{III}} (\lambda_{\max}(Y_j X^{-1}), \lambda_{\max}(XY_j^{-1})) \xmapsto{\text{IV}} (\lambda_{\max}(Y_j X^{-1}), \lambda_{\min}(Y_j X^{-1})) \xmapsto{\text{V}} m^j_X.$$

Let us consider whether each of these five maps is Lipschitz.

    I: Inversion is a smooth operation on the invertible matrices and $\mathcal{K}$ is a compact set of invertible matrices, so I is Lipschitz.

    II: Matrix multiplication is a smooth operation on matrices, so II is Lipschitz.

    III: This map is Lipschitz (and in fact non-expansive) with respect to the matrix norm $\|\cdot\|$.

IV: This map is given by inversion of the second coordinate. This is not Lipschitz. However we could restrict the domain to a compact subset of $\mathbb{R}_{>0} \times \mathbb{R}_{>0}$, since the image of the continuous map (map I) $\circ$ (map II) $\circ$ (map III) is a compact set. Then IV is Lipschitz on this domain.

V: Explicitly, this map takes the form

$$(\alpha, \beta) \mapsto \begin{cases} \frac{\log \beta - \log \alpha}{\beta - \alpha} & \text{if } \beta \neq \alpha \\ \frac{1}{\alpha} & \text{if } \beta = \alpha. \end{cases}$$

We do not need V to be Lipschitz, we only need it to be locally Lipschitz, and then restrict the domain to a compact set like we did for IV. For this we show that its partial derivatives exist and are continuous. For $\beta \neq \alpha$,

$$\frac{\partial}{\partial \beta}(\text{map V})(\alpha, \beta) = \frac{1 - \alpha/\beta - \log \beta + \log \alpha}{(\beta - \alpha)^2}.$$

As $(\alpha, \beta) \to (\gamma, \gamma)$, the above tends to $-\frac{1}{2\gamma^2}$. This can be shown by letting $\alpha = \gamma + ta$ and $\beta = \gamma + tb$ for some b and a, letting $t \to 0$ and applying l'Hôpital's rule twice. Moreover we have

$$\frac{\partial}{\partial \beta}(\text{map V})(\gamma, \gamma) = -\frac{1}{2\gamma^2}.$$

So the $\frac{\partial}{\partial \beta}$ derivatives exist and are continuous. The argument for the $\frac{\partial}{\partial \alpha}$ derivatives is analogous. This shows V is C^1 and hence locally Lipschitz.

Now m_X^j is Lipschitz in $X \in \mathcal{K}$ since it is a composition of Lipschitz maps.   □

Pick an initialization $X_1 \in \mathbb{S}_{++}^n$ and let $\mathcal{C} = \text{conv}\{X_1, Y_1, \ldots, Y_k\}$, where conv is used to denote the Euclidean convex hull. If $X \in \mathbb{S}_{++}^n$, $1 \leq j \leq k$ and $t \in [0, 1]$,

$$(6.6) \qquad X *_t Y_j = \left(\varphi_X^j(t) + \psi_X^j(t)\right)\left(\frac{\varphi_X^j(t)}{\varphi_X^j(t) + \psi_X^j(t)}Y_j + \frac{\psi_X^j(t)}{\varphi_X^j(t) + \psi_X^j(t)}X\right).$$

Write $\mathbb{R}_{>0} \cdot \mathcal{C} = \{cX : X \in \mathcal{C}, c > 0\}$. Then (6.6) tells us that for all $i \in \mathbb{N}$, S_i maps $\mathcal{C}$ to $\mathbb{R}_{>0} \cdot \mathcal{C}$. So by Lemma 6.2 it maps $\mathbb{R}_{>0} \cdot \mathcal{C}$ to itself. In particular $X_i \in \mathbb{R}_{>0} \cdot \mathcal{C}$ for all $i \in \mathbb{N}$.

LEMMA 6.4. *There is* $X^* \in \mathbb{R}_{>0} \cdot \mathcal{C}$ *such that*

$$(6.7) \qquad m_{X^*}^k Y_k + \cdots + m_{X^*}^1 Y_1 + \sum_{j=1}^{k} o_{X^*}^j X^* = 0.$$

Proof. Consider the map $F : \mathcal{C} \to \mathcal{C}$ defined by

$$F : X \mapsto \frac{m_X^k Y_k + \cdots + m_X^1 Y_1}{\sum_{j=1}^{k} m_X^j}.$$

This is a continuous map (by continuity of the m_X^j in X, Lemma 6.3) from the convex compact set $\mathcal{C}$ to itself, so it has a fixed point X^{**} by Brouwer's fixed point theorem [2, Theorem 4.10]. Thus we have

$$X^{**} = \frac{m_{X^{**}}^k Y_k + \cdots + m_{X^{**}}^1 Y_1}{\sum_{j=1}^{k} m_{X^{**}}^j}.$$

Now try $X^* = cX^{**}$ for $c > 0$ in (6.7). Using the relations $m^j_{cX^{**}} = cm^j_{X^{**}}$ and $o^j_{cX^{**}} = o^j_{X^{**}} - \log c$ and solving for c we get a solution

$$(6.8) \qquad X^* = \exp\left(\frac{\sum_{j=1}^{k} m^j_{X^{**}} + \sum_{j=1}^{k} o^j_{X^{**}}}{k}\right) X^{**}. \qquad \square$$

From now on X^* will denote a point satisfying the conditions of Lemma 6.4.

Remark 6.5. Eventually we will show that $X_i \to X^*$ as $i \to \infty$ (and thus that X^* is uniquely defined). This can be understood intuitively: writing out (6.7) in a more explicit notation we have

$$(6.9) \qquad \frac{d\varphi^k_{X^*}}{dt}(0)Y_k + \cdots + \frac{d\varphi^1_{X^*}}{dt}(0)Y_1 + \left(\frac{d\psi^k_{X^*}}{dt}(0) + \cdots + \frac{d\psi^1_{X^*}}{dt}(0)\right)X^* = 0.$$

(6.9) is, loosely speaking, the infinitesimal version (taking $p \to \infty$) of the equation

$$T_p(X) = X,$$

which characterises the fixed point(s) of T_p.

We are now in a position to prove the following crucial lemma.

LEMMA 6.6. *Let $\mathcal{K} \subset \mathbb{S}^n_{++}$ be a compact set with $X^* \in \mathcal{K}$. Then there exist $O, K > 0$ such that for all $X \in \mathcal{K}$ and $p \in \mathbb{Z}_{\geq 0}$,*

$$(6.10) \qquad \|T_p(X) - X\| \leq \frac{K}{p}\|X - X^*\| + \frac{O}{p^2}.$$

In particular,

$$(6.11) \qquad \|T_p(X^*) - X^*\| \leq \frac{O}{p^2}.$$

Proof. For $1 \leq j \leq k$, we define recursively

$$\mathcal{K}_j = \{X *_t Y_j : X \in \mathcal{K}_{j-1}, t \in [0,1]\}$$

where $\mathcal{K}_0 = \mathcal{K}$. The $\mathcal{K}_j$ are continuous images of compact sets since the φ^j_X and ψ^j_X are continuous in X (Lemma 6.3). Then for $i \in \mathbb{N}$ and $1 \leq j \leq k$ such that $j \equiv i$ mod k, if $X \in \mathcal{K}_{j-1}$,

$$S_i(X) = \varphi^j_X\left(\frac{1}{i+1}\right)Y_j + \psi^j_X\left(\frac{1}{i+1}\right)X$$

so taking a Taylor expansion to first order we get

$$(6.12) \qquad S_i(X) = \frac{m^j_X}{i+1}Y_j + \left(1 + \frac{o^j_X}{i+1}\right)X + R_j(X,i).$$

where $\|R_j(X,i)\| \leq \frac{M_j}{i^2}$ for some $M_j > 0$ independent of $X \in \mathcal{K}_{j-1}$. This bound on R_j is possible because $\frac{d^2\varphi^j_X}{dt^2}$ and $\frac{d^2\psi^j_X}{dt^2}$ are uniformly bounded for $X \in \mathcal{K}_{j-1}$ and $t \in [0, \frac{1}{i+1}] \subset [0,1]$ by continuity on these compact sets (Lemma 6.3).

Now for $X \in \mathcal{K}$ and $p \in \mathbb{N}$

$$T_p(X) =$$

$$\frac{m^k_{S_{(p+1)k}(\ldots S_{pk+1}(X)\ldots)}}{(p+1)k+1} Y_k + \left(1 + \frac{o^k_{S_{(p+1)k}(\ldots S_{pk+1}(X)\ldots)}}{(p+1)k+1}\right) \frac{m^{k-1}_{S_{(p+1)k-1}(\ldots S_{pk+1}(X)\ldots)}}{(p+1)k} Y_{k-1}$$

$$+ \cdots + \left(1 + \frac{o^k_{S_{(p+1)k}(\ldots S_{pk+1}(X)\ldots)}}{(p+1)k+1}\right) \cdots \left(1 + \frac{o^1_X}{pk+2}\right) X + R(X,p)$$

$$= \frac{m^k_{S_{(p+1)k}(\ldots S_{pk+1}(X)\ldots)}}{pk} Y_k + \cdots + \frac{m^1_X}{pk} Y_1$$

$$+ \left(1 + \frac{o^k_{S_{(p+1)k}(\ldots S_{pk+1}(X)\ldots)} + \cdots + o^1_X}{pk}\right) X + S(X,p)$$

$$= \frac{m^k_X}{pk} Y_k + \cdots + \frac{m^1_X}{pk} Y_1 + \left(1 + \frac{\sum_{j=1}^{k} o^j_X}{pk}\right) X + T(X,p)$$

$$= \frac{m^k_X - m^k_{X^*}}{pk} Y_k + \cdots + \frac{m^1_X - m^1_{X^*}}{pk} Y_1 + \frac{\sum_{j=1}^{k}(o^j_X - o^j_{X^*})}{pk} X$$

$$+ \frac{\sum_{j=1}^{k} o^j_{X^*}}{pk}(X - X^*) + X + T(X,p)$$

$$= X + U(X,p)$$

where $R(X,p) \le \frac{M}{p^2}$, $S(X,p) \le \frac{N}{p^2}$, $T(X,p) \le \frac{O}{p^2}$ and $U(X,p) \le \frac{K}{p}\|X - X^*\| + \frac{O}{p^2}$ for some $L, M, N, O, K > 0$ independent of $X \in \mathcal{C}$. The bound on R comes from the expansion (6.12) applied k times and using the fact that $m^j_{X'}$ and $o^j_{X'}$ are continuous in X' (Lemma 6.3), so bounded on the compact set $\mathcal{K}_k$. This last observation also gives us the bound on S. The bound on T uses the fact that $\|S_j \circ \cdots \circ S_1(X) - X\|$ vanishes to order $\frac{1}{p}$ for $X \in \mathcal{K}$ since the $m^j_{X'}$ and $o^j_{X'}$ are bounded for $X' \in \mathcal{K}_k$. Then we use the fact that $m^j_{X'}$ and $o^j_{X'}$ are Lipschitz in $X' \in \mathcal{K}_k$ (Lemma 6.3). Finally, the bound on U uses the fact that $m^j_{X'}$ and $o^j_{X'}$ are Lipschitz in $X' \in \mathcal{K}$ (Lemma 6.3) and bounded on that set. This proves the lemma. □

Remark 6.7. Using similar estimates as in the proof of Lemma 6.6, we can show that there are $O', K' > 0$ such that for $X \in \mathcal{K}$ and $p \in \mathbb{Z}_{\ge 0}$,

$$(6.13) \qquad \|T_p(X) - X^*\| \le K'\|X - X^*\| + \frac{O'}{p^2}.$$

However it is not clear whether $K' < 1$. If not, (6.13) is not good enough to show that the point X^* is attractive under our dynamics, so we will need to use more machinery involving the Hilbert projective metric.

6.2. Hilbert projective convergence. Here we establish convergence of any sequence generated by Algorithm 6.1 in Hilbert's projective geometry. Recall that Hilbert's projective metric d_H takes the form (1.3) in $\mathbb{S}^n_{++}$ and satisfies $d_H(cX, c'X) = d_H(X, X')$ for any $X, X' \in \mathbb{S}^n_{++}$ and $c, c' > 0$. Moreover, d_H is a metric in the usual sense on the projective space (space of rays) $\mathbb{S}^n_{++}/\mathbb{R}_{>0}$ [35, Proposition 2.1.1]. To proceed further, we need to be able to translate our estimates in the Euclidean norm to the Hilbert projective metric. This is achieved by the following lemma.

LEMMA 6.8. *Let $\mathcal{K} \subset \mathbb{S}^n_{++}$ be a compact set. Then there is $C > 0$ such that for $X, X' \in \mathcal{K}$*

$$d_H(X, X') \le C\|X - X'\|.$$

Proof. For $X, X' \in \mathcal{K}$,

$$d_H(X, X') = \log\left(\frac{\lambda_{\max}(X'X^{-1})}{\lambda_{\min}(X'X^{-1})}\right).$$

So d_H is the composition of the five maps

$$\mathcal{K} \times \mathcal{K} \xrightarrow{\mathrm{I}} \mathcal{K} \times \mathcal{K} \times \mathcal{K} \times \mathcal{K} \xrightarrow{\mathrm{II}} \mathcal{K} \times \mathcal{K} \xrightarrow{\mathrm{III}} \mathbb{R}_{>0} \times \mathbb{R}_{>0} \xrightarrow{\mathrm{IV}} \mathbb{R}_{>0} \times \mathbb{R}_{>0} \xrightarrow{\mathrm{V}} \mathbb{R}_{\ge 0}$$

$$(X, X') \xmapsto{\mathrm{I}} (X, X^{-1}, X', X'^{-1}) \xmapsto{\mathrm{II}} (X'X^{-1}, XX'^{-1})$$

$$\xmapsto{\mathrm{III}} (\lambda_{\max}(X'X^{-1}), \lambda_{\max}(XX'^{-1})) \xmapsto{\mathrm{IV}} (\lambda_{\max}(X'X^{-1}), \lambda_{\min}(X'X^{-1}))$$

$$\xmapsto{\mathrm{V}} \log\left(\frac{\lambda_{\max}(X'X^{-1})}{\lambda_{\min}(X'X^{-1})}\right).$$

Then we can show d_H is Lipschitz analogously to the proof of Lemma 6.3. $\qquad\square$

Let $\mathcal{K} \subset \mathbb{S}^n_{++}$ be a compact set. By Theorem 4.5 ([47, Theorem 1.2]), we have for $X, X', Y \in \mathbb{R}_{>0} \cdot \mathcal{K}$ and $t \in [0, 1]$,

$$(6.14) \qquad d_H(X *_t Y, X' *_t Y) \le \gamma_{1-t}(R) d_H(X, X')$$

where

$$\gamma_{1-t}(R) = \frac{1 - e^{-R(1-t)}}{1 - e^{-R}}$$

and

$$R = \operatorname{diam}_{d_H}(\mathbb{R}_{>0} \cdot \mathcal{K}) = \operatorname{diam}_{d_H}(\mathcal{K}) = \sup\{d_H(X, X') : X, X' \in \mathcal{K}\} < \infty$$

since $d_H(cX, c'X') = d_H(X, X')$ for all $c, c' > 0$ and $\mathcal{K}$ is compact. We immediately get the following lemma.

LEMMA 6.9 (Hilbert contractivity). *Let $\mathcal{K} \subset \mathbb{S}^n_{++}$ be a compact set. If $X, X' \in \mathbb{R}_{>0} \cdot \mathcal{K}$, $i \in \mathbb{N}$ and $p \in \mathbb{Z}_{\ge 0}$*

$$d_H(S_i(X), S_i(X')) \le \gamma_{\frac{i}{i+1}}(R) d_H(X, X').$$

and so

$$d_H(T_p(X), T_p(X')) \le \gamma_{\frac{(p+1)k}{(p+1)k+1}}(R) \cdots \gamma_{\frac{pk+1}{pk+2}}(R) d_H(X, X')$$

where $R = \operatorname{diam}_{d_H}(\mathcal{K})$.

Note that taking the tangent line at $-R$ of the function $x \to e^x$ and using that this function is convex we get $e^x \ge e^{-R} + (x + R)e^{-R}$. So

$$(6.15) \qquad \gamma_{1-t}(R) = \frac{1 - e^{-R(1-t)}}{1 - e^{-R}} \le 1 - \frac{Re^{-R}}{1 - e^{-R}}t.$$

Now the following observation will turn out to be useful:

$$(6.16) \qquad \prod_{i=1}^{\infty} \gamma_{\frac{i}{i+1}}(R) = 0.$$

This holds because

$$0 \leq \prod_{i=1}^{\infty} \gamma_{\frac{i}{i+1}}(R) \leq \prod_{i=1}^{\infty} \left(1 - \frac{Re^{-R}}{1-e^{-R}} \frac{1}{i+1}\right) = 0$$

where the second inequality holds by (6.15) and the last identity holds by [5, Corollary 2.2.3] and using the divergence of the harmonic series.

Remark 6.10. (6.16) combined with Lemma 6.9 tells us that, given any two initializations for the algorithm, the resulting sequences will come arbitrarily close together in the Hilbert projective metric. However, this is not enough to show convergence in this projective metric. The key to showing this will be to also use Lemma 6.6, with the help of Lemma 6.8.

PROPOSITION 6.11 (Hilbert convergence). *Let $(X_i)_{i\geq 1}$ be any sequence generated by Algorithm 6.1 and X^* denote a point satisfying the conditions of Lemma 6.4. Then, we have*

$$d_H(X_{pk+1}, X^*) \to 0 \ as \ p \to \infty.$$

Proof. For $p \in \mathbb{Z}_{\geq 0}$ we have

(6.17)
$$d_H(X_{(p+1)k+1}, X^*) \leq d_H(X_{(p+1)k+1}, T_p(X^*)) + d_H(T_p(X^*), X^*)$$
$$\leq \gamma_{\frac{(p+1)k}{(p+1)k+1}}(R) \cdots \cdots \gamma_{\frac{pk+1}{pk+2}}(R) d_H(X_{pk+1}, X^*) + \frac{D}{p^2}$$

for some $D > 0$, where we used Lemma 6.9 for the first term and Lemma 6.8 followed by (6.11) from Lemma 6.6 for the second term. Lemma 6.9 was applied with $\mathcal{K} = \mathcal{C}$ and Lemma 6.8 was applied with

$$\mathcal{K} = \{(\dots (X^* *_{t_1} Y_1) *_{t_2} \dots) *_{t_k} Y_k : t_1, \dots, t_k \in [0,1]\}$$

which is compact by continuity of the φ_X^j and ψ_X^j in X (Lemma 6.3).

Now applying (6.17) recursively we get

$$d_H(X_{pk+1}, X^*) \leq \left(\prod_{i=k+1}^{pk} \gamma_{\frac{i}{i+1}}(R)\right) d_H(X_{k+1}, X^*) + \sum_{q=1}^{p-1} \left(\prod_{i=qk+1}^{(p-1)k} \gamma_{\frac{i}{i+1}}(R)\right) \frac{D}{q^2}$$

$$\leq \left(\prod_{i=k+1}^{pk} \gamma_{\frac{i}{i+1}}(R)\right) d_H(X_{k+1}, X^*) + \sum_{q=1}^{\infty} \left(\prod_{i=qk+1}^{(p-1)k} \gamma_{\frac{i}{i+1}}(R)\right) \frac{D}{q^2}$$

$$\to 0 \ as \ p \to \infty$$

using (6.16) and that $\sum_{q=1}^{\infty} \frac{D}{q^2} < \infty$, where we interpret $\prod_{i=l}^{m} a_i$ as 1 when $m < l$. $\square$

6.3. Convergence. Let

$$\mathcal{S} = \{X \in \mathbb{S}_{++}^n : \|X\| = \|X^*\|\}.$$

So $X^* \in \mathcal{S}$. Moreover, if $X \in \mathbb{S}_{++}^n$ there is a unique $c > 0$ such that $c^{-1}X \in \mathcal{S}$. So we have the natural identification

$$\mathcal{S} \times \mathbb{R}_{>0} \xrightarrow{\cong} \mathbb{S}_{++}^n, \quad (\hat{X}, c) \mapsto c\hat{X}.$$

Write $(\hat{X}_i, c_i)_{i \geq 1} \subset \mathcal{S} \times \mathbb{R}_{>0}$ for the sequence corresponding to $(X_i)_{i \geq 1} \subset \mathbb{S}^n_{++}$. By Proposition 6.11, $(X_{pk+1})_{p \geq 0}$ tends to X^* in the Hilbert projective metric, hence so does $(\hat{X}_{pk+1})_{p \geq 0}$. The Hilbert projective metric is a metric in the proper sense on the set $\mathcal{S}$ [35, Proposition 2.1.1]. Moreover, the topology it generates is the Euclidean topology [46, Proposition 1.1]. Hence $(\hat{X}_{pk+1})_{p \geq 0}$ actually converges to X^* in the Euclidean topology, and so in the Euclidean norm. Now note that $X_{pk+1} = c_{pk+1}\hat{X}_{pk+1}$ for all $p \in \mathbb{Z}_{\geq 0}$. So we need to show $(c_{pk+1})_{p \geq 0}$ converges.

We will slightly abuse the notation and view T_p as a map from $\mathcal{S} \times \mathbb{R}_{>0}$ to itself. With this in mind, for $\hat{X} \in \mathcal{S}$ write $b_{\hat{X},p}$ for the positive number corresponding to the second coordinate of $T_p(\hat{X}, 1)$. We will need to analyse these.

LEMMA 6.12. *Let $\mathcal{K} \subset \mathcal{S}$ be a compact set with $X^* \in \mathcal{K}$. Then there is $I, L > 0$ such that for all $\hat{X} \in \mathcal{K}$ and $p \in \mathbb{Z}_{\geq 0}$*

$$\|b_{\hat{X},p} - 1\| \leq \frac{L}{p}\|\hat{X} - X^*\| + \frac{I}{p^2}.$$

Proof. This is an exercise in Euclidean geometry using Lemma 6.6. Let $\hat{X} \in \mathcal{K}$ and $p \in \mathbb{Z}_{\geq 0}$. Write $T_p(\hat{X}, 1) = (\hat{X}', b_{\hat{X},p})$. Then let x, y and z be the lengths of the segments $\hat{X}'$ to $b_{\hat{X},p}\hat{X}'$, $b_{\hat{X},p}\hat{X}'$ to $\hat{X}$ and $\hat{X}$ to $\hat{X}'$ respectively. Furthermore, let h be the distance from $\hat{X}$ to the line through 0 and $\hat{X}'$, and α be the angle between this line and the line through 0 and $\hat{X}$ (see Figure 7). Then

$$(6.18) \qquad y \leq \frac{K}{p}\|\hat{X} - X^*\| + \frac{O}{p^2}$$

for some $O, K > 0$ independent of $\hat{X} \in \mathcal{K}$ by (6.10) from Lemma 6.6. So

$$(6.19) \qquad \alpha = \arcsin\frac{h}{\|\hat{X}\|} \leq \arcsin\frac{y}{\|\hat{X}\|} \to 0 \text{ as } p \to \infty$$

uniformly for $\hat{X} \in \mathcal{S}$. Now

$$(6.20) \qquad z = \frac{h}{\cos(\alpha/2)} \leq \frac{y}{\cos(\alpha/2)}.$$

So

$$x \leq y + z \leq \left(1 + \frac{1}{\cos(\alpha/2)}\right)y \leq \frac{K'}{p}\|\hat{X} - X^*\| + \frac{O'}{p^2}$$

by (6.20), (6.19) and (6.18) for some $O', K' > 0$ independent of $\hat{X} \in \mathcal{K}$. Dividing by $\|\hat{X}'\| = \|X^*\|$, we get the lemma. $\square$
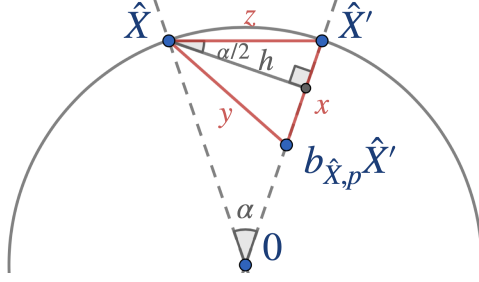
FIG. 7. *Sketch of the geometry involved in the proof of Lemma 6.12. Informally, we are trying to show x is small. We know from Lemma 6.6 that y is small, and we deduce that z is small and thus that x must be small.*

Now we have all the necessary ingredients to prove convergence of $(c_{pk+1})_{p \geq 0}$.

PROPOSITION 6.13 (Radial convergence). *Let $(X_i)_{i \geq 1}$ be any sequence generated by Algorithm 6.1 and $(c_i)_{i \geq 1}$ be the corresponding sequence in $\mathbb{R}_{>0}$ defined at the beginning of subsection 6.3. Then, we have*

$$c_{pk+1} \to 1 \ as \ p \to \infty.$$

Proof. Lemma 6.2 says that if $\hat{X}, \hat{X}' \in \mathcal{S}$ and $a, a' > 0$ are such that $T_p(\hat{X}, a) = (\hat{X}', a')$, then for $c > 0$,

$$(6.21) \qquad T_p(\hat{X}, ca) = (\hat{X}', c^{\frac{pk+1}{(p+1)k+1}} a').$$

Then by (6.21)

$$(6.22) \qquad c_{(p+1)k+1} = c_{pk+1}^{\frac{pk+1}{(p+1)k+1}} b_{\hat{X}_{pk+1}, p}.$$

Now since $\hat{X}_{pk+1} \to X^*$ as $p \to \infty$, Lemma 6.12 applied to $\mathcal{K} = (\mathbb{R}_{>0} \cdot \mathcal{C}) \cap \mathcal{S}$ implies that for $\epsilon > 0$ there is $r \in \mathbb{Z}_{\geq 0}$ such that

$$(6.23) \qquad |b_{\hat{X}_{qk+1}, q} - 1| \leq \frac{\epsilon}{q}$$

for all $q \geq r$. Also note that for $x \in \mathbb{R}$, $(1 + x/q)^q \to e^x$ as $q \to \infty$. Thus, possibly by increasing r, we have

$$(6.24) \qquad \left(1 + \frac{\epsilon}{q}\right)^q \leq \exp(2\epsilon)$$

and

$$(6.25) \qquad \left(1 - \frac{\epsilon}{q}\right)^q \geq \exp(-2\epsilon)$$

for all $q \geq r$. Now applying (6.22) recursively we get for $p > r$

$$c_{pk+1} = c_{rk+1}^{\frac{rk+1}{pk+1}} \prod_{q=r}^{p-1} b_{\hat{X}_{qk+1}, q}^{\frac{(q+1)k+1}{pk+1}}$$

so using (6.23) followed by (6.24)

$$c_{pk+1} \leq c_{rk+1}^{\frac{rk+1}{pk+1}} \prod_{q=r}^{p-1} \left(1 + \frac{\epsilon}{q}\right)^{\frac{(q+1)k+1}{pk+1}}$$

(6.26)
$$\leq c_{rk+1}^{\frac{rk+1}{pk+1}} \prod_{q=r}^{p-1} \exp(2\epsilon)^{\frac{k+(k+1)/q}{pk+1}}$$

$$= c_{rk+1}^{\frac{rk+1}{pk+1}} \exp(2\epsilon)^{\sum_{q=r}^{p-1} \frac{k+(k+1)/r}{pk+1}}$$

$$\leq c_{rk+1}^{\frac{rk+1}{pk+1}} \exp(2\epsilon)^{1+2/r} \to \exp(2\epsilon)^{1+2/r} \text{ as } p \to \infty.$$

We can similarly use (6.23) followed by (6.25) to get

(6.27)
$$c_{pk+1} \geq c_{rk+1}^{\frac{rk+1}{pk+1}} \exp(-2\epsilon)^{1+2/r} \to \exp(-2\epsilon)^{1+2/r} \text{ as } p \to \infty.$$

Since $\epsilon > 0$ is arbitrary and $r \in \mathbb{Z}_{\geq 0}$ is arbitrarily large, we deduce from (6.26) and (6.27) that $c_{pk+1} \to 1$ as $p \to \infty$. □

Finally, we are in a position to state and prove our main theorem.

THEOREM 6.14 (Convergence). *Let* $(X_i)_{i \geq 1}$ *denote any sequence generated by* Algorithm 6.1. *We have*

$$X_i \to X^* \text{ as } i \to \infty,$$

where X^* *is independent of the choice of initialization* X_1. *Moreover,* X^* *is the unique solution in* $\mathbb{S}_{++}^n$ *to the equation*

(6.28)
$$m_{X^*}^k Y_k + \cdots + m_{X^*}^1 Y_1 + \sum_{j=1}^{k} o_{X^*}^j X^* = 0,$$

where $m_X^j = \frac{d\varphi_X^j}{dt}(0)$, $o_X^j = \frac{d\psi_X^j}{dt}(0)$, $\varphi_X^j = \varphi_{\alpha\beta}$, $\psi_X^j = \psi_{\alpha\beta}$, $\alpha = \lambda_{\min}(Y_j X^{-1})$, *and* $\beta = \lambda_{\max}(Y_j X^{-1})$.

Proof. We have shown with projective convergence (Proposition 6.11) and radial convergence (Proposition 6.13) that

$$X_{pk+1} = c_{pk+1}\hat{X}_{pk+1} \to X^* \text{ as } p \to \infty.$$

With the same proof we can show that for $1 \leq j \leq k$, $(X_{pk+j})_{p \geq 0}$ converges, and it does so to the same point X^*. So

$$X_i \to X^* \text{ as } i \to \infty.$$

To be precise, we have shown $X_i \to X^*$ for any $X^* \in \mathbb{R}_{>0} \cdot \mathcal{C}$ satisfying (6.28). The proof can easily be extended to any $X^* \in \mathbb{S}_{++}^n$. Thus, by uniqueness of limits, the solution to (6.28) must be unique. Moreover we see that X^* is independent of the choice of initialization X_1 by the symmetries of (6.28), or alternatively by construction of X^* in the proof of Lemma 6.4. □

Remark 6.15. Theorem 6.14 holds more generally than just for the cone $\mathbb{S}_{++}^n$: given a closed convex cone in a finite dimensional real vector space V, take $\mathcal{P} =$

$(y_1, \ldots, y_k)$ in its interior. The cone is then almost Archimedean so we can still apply [47, Theorem 2] in Section 6.2, and it is also normal ([35, Lemma 1.2.5]) so we can also apply [46, Proposition 1.1] in subsection 6.3. The only other parts of the proof that need generalizing are the proofs of Lemma 6.3 and Lemma 6.8. For this, we only need the additional assumption that, in the interior of this cone, the map $(x, y) \mapsto M(y/x)$ is locally Lipschitz with respect to a norm on V. This is for example the case for the cone $\mathbb{R}^n_+ = \{(x_i)^n_{i=1} : x_i \geq 0, \ 1 \leq i \leq n\}$ where $M(y/x) = \max\{y_i/x_i\}^n_{i=1}$.

6.4. Properties of the limit. Since X^* only depends on the Y_j and not on X_1 we can write $M(Y_1, \ldots, Y_k) = X^*$ and study how M varies in terms of its arguments. We will see that M satisfies a number of nice properties, and thus can be viewed as a mean of the Y_j. Some of these properties can be proved in several different ways. However, in a lot of cases, the characterisation of M as being the unique solution to (6.28) grants us with some elegant proofs.

THEOREM 6.16 (Properties). *Let $Y_1, \ldots, Y_k \in \mathbb{S}^n_{++}$.*
  1. *For any $0 \leq l \leq k$,*

$$M(\underbrace{Y_1, \ldots, Y_1}_{k-l}, \underbrace{Y_2, \ldots, Y_2}_{l}) = Y_1 *_{\frac{l}{k}} Y_2,$$

  and in particular $M(Y_1, Y_2) = Y_1 *_{\frac{1}{2}} Y_2$.
  2. *Permutation invariance:*

$$M(Y_{\sigma(1)}, \ldots Y_{\sigma(k)}) = M(Y_1, \ldots, Y_k)$$

  for any permutation of k elements σ.
  3. *Affine-equivariance:*

$$M(AY_1A^T, \ldots, AY_kA^T) = AM(Y_1, \ldots, Y_k)A^T$$

  for any invertible matrix A.
  4. *Joint homogeneity:*

$$M(c_1 Y_1, \ldots, c_k Y_k) = (c_1 \cdots c_k)^{1/k} M(Y_1, \ldots, Y_k)$$

  for any $c_1, \ldots, c_k > 0$.
  5. *The map $M : (\mathbb{S}^n_{++})^k \to \mathbb{S}^n_{++}$ is continuous.*

Proof. 1. In this case it is perhaps easiest not to prove this property with (6.28) but instead to use Algorithm 6.1 and Lemma 6.1. Taking $X_1 = Y_1 *_{\frac{l}{k}} Y_2$ in the algorithm we have, using Lemma 6.1 recursively,

$$X_{pk+j} = \begin{cases} Y_1 *_{\frac{l}{k} \frac{pk+1}{pk+j}} Y_2 & \text{if } 1 \leq j \leq k - l \\ Y_1 *_{1 - (1 - \frac{l}{k}) \frac{(p+1)k+1}{pk+j}} Y_2 & \text{if } k - l + 1 \leq j \leq k \end{cases}$$

for all $p \in \mathbb{Z}_{\geq 0}$. In particular $X_{pk+1} = Y_1 *_{\frac{l}{k}} Y_2$ for all $p \in \mathbb{Z}_{\geq 0}$, and thus $X_i \to Y_1 *_{\frac{l}{k}} Y_2$ as $i \to \infty$.

2. Non-trivial from Algorithm 6.1, but follows immediately from the symmetries of (6.28).

3. Follows from Algorithm 6.1 and Proposition 4.1. We give an alternative proof:

writing $\tilde{X}^* = M(AY_1A^T, \ldots, AY_kA^T)$ and $X^* = M(Y_1, \ldots, Y_k)$, we have from (6.28)

$$m^k_{\tilde{X}^*}AY_kA^T + \cdots + m^1_{\tilde{X}^*}AY_1A^T + \sum_{j=1}^{k} o^j_{\tilde{X}^*}\tilde{X}^* = 0.$$

Here we used that, for all $1 \leq j \leq k$, $AY_jA^T(AXA^T)^{-1} = AY_jX^{-1}A^{-1}$ and Y_jX^{-1} have the same (extreme) eigenvalues, and thus the coefficients m^j_X and o^j_X remain the same as in (6.28). So we see that $\tilde{X}^* = AX^*A^T$ satisfies this equation.

4. Again, this is non-trivial from Algorithm 6.1. Writing $\tilde{X}^* = M(c_1Y_1, \ldots, c_kY_k)$ and $X^* = M(Y_1, \ldots, Y_k)$, we have from (6.28)

$$\tilde{m}^k_{\tilde{X}^*}c_kY_k + \cdots + \tilde{m}^1_{\tilde{X}^*}c_1Y_1 + \sum_{j=1}^{k} \tilde{o}^j_{\tilde{X}^*}\tilde{X}^* = 0$$

where $\tilde{m}^j_X = c_j^{-1}m^j_X$ and $\tilde{o}^j_X = o^j_X + \log c_j$ for all $1 \leq j \leq k$ and all X. So

$$m^k_{\tilde{X}^*}Y_k + \cdots + m^1_{\tilde{X}^*}Y_1 + \sum_{j=1}^{k} (o^j_{\tilde{X}^*} + \log c_j)\tilde{X}^* = 0.$$

Now it suffices to check that $\tilde{X}^* = (c_1 \cdots c_k)^{1/k}X^*$ satisfies this equation, using the relations $m^j_{cX} = cm^j_X$ and $o^j_{cX} = o^j_X - \log c$ for $c > 0$.

5. Consider the map $E : (\mathbb{S}^n_{++})^{k+1} \to \mathbb{S}^n_{++}$ given by

$$E : (Y_1, \ldots, Y_k, X) \mapsto m^k_X Y_k + \cdots + m^1_X Y_1 + \sum_{j=1}^{k} o^j_X X.$$

We have by (6.28)

$$E(Y_1, \ldots, Y_k, M(Y_1, \ldots, Y_k)) = 0$$

for all $Y_1, \ldots, Y_k \in \mathbb{S}^n_{++}$. One may try to apply the implicit function theorem to show that M is continuous. However E is not in general differentiable, only continuous (Lemma 6.3), so the implicit function theorem in its classical form cannot be applied. Instead, take $(Y_{1,i}, \ldots, Y_{k,i})_{i \geq 1} \subset (\mathbb{S}^n_{++})^{k+1}$ a convergent sequence, converging to $(Y_1, \ldots, Y_k)$, say. Then writing $\tilde{\mathcal{C}}$ for the closure of the convex hull of the bounded set $\{(Y_{1,i}, \ldots, Y_{k,i}) : i \in \mathbb{N}\}$, $\tilde{\mathcal{C}}$ is closed and bounded so compact. By the proof of Lemma 6.4, $(M(Y_{1,i}, \ldots, Y_{k,i}))_{i \geq 1} \subset \mathcal{K}$ where

$$\mathcal{K} = \left\{ \exp\left(\frac{\sum_{j=1}^{k} m^j_X + \sum_{j=1}^{k} o^j_X}{k} \right) X : X \in \tilde{\mathcal{C}} \right\},$$

which is compact (Lemma 6.3). Thus, there is a convergent subsequence

$$(M(Y_{1,i_l}, \ldots, Y_{k,i_l}))_{l \geq 1}$$

converging to M^*, say. Then by continuity of E we have $E(Y_1, \ldots, Y_k, M^*) = 0$. But by uniqueness of the solution to (6.28), $M(Y_1, \ldots, Y_k) = M^*$. If $(M(Y_{1,i}, \ldots, Y_{k,i}))_{i \geq 1}$ did not converge to M^*, then it would have another convergent subsequence converging to a $\tilde{M}^* \neq M^*$. But then $E(Y_1, \ldots, Y_k, \tilde{M}^*) = 0$, contradicting the uniqueness of the solution to (6.28). So $M(Y_{1,i}, \ldots, Y_{k,i}) \to M^* = M(Y_1, \ldots, Y_k)$ as $i \to \infty$. So M is continuous. $\square$

7. Conclusions. The Hilbert and Thompson metrics in the positive semidefinite cone provide a route to non-Euclidean geometries based on extreme generalized eigenvalue computations. We have seen that by focusing on a particular choice of geodesic of the Thompson metric with attractive computational properties, we can view $(\mathbb{S}^n_{++}, d_T)$ as a semihyperbolic geodesic space. We have noted several interesting properties of this distinguished Thompson geodesic, including the preservation of sparsity. Significantly, we have defined an iterative mean of any finite collection of SPD matrices based on the computation of a sequence of extreme generalized eigenvalues. Furthermore, we have proved that this new mean exists and is unique for any given finite collection of SPD matrices. Finally, we have established several important properties that are satisfied by this mean, including permutation invariance, affine-equivariance, and joint homogeneity. We hope that these contributions will provide a foundation for a computationally scalable geometric statistical framework for the processing of large SPD-valued data.

REFERENCES

[1] P.-A. ABSIL, R. MAHONY, AND R. SEPULCHRE, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, 2009, https://doi.org/doi:10.1515/9781400830244, https://doi.org/10.1515/9781400830244.

[2] R. P. AGARWAL, M. MEEHAN, AND D. O'REGAN, *Fixed Point Theory and Applications*, Cambridge Tracts in Mathematics, Cambridge University Press, Cambridge, 2001, https://doi.org/10.1017/CBO9780511543005.

[3] D. ALEXANDER, K. SEUNARINE, S. NEDJATI-GILANI, M. HALL, G. PARKER, Y. BAI, AND P. COOK, *Camino: Open-source diffusion-MRI reconstruction and processing*, in Proceedings of the 14th Scientific Meeting of the International Society for Magnetic Resonance in Medicine (ISMRM), Seattle, WA, USA, 2006.

[4] J. M. ALONSO AND M. R. BRIDSON, *Semihyperbolic Groups*, Proceedings of the London Mathematical Society, s3-70 (1995), pp. 56–114, https://doi.org/10.1112/plms/s3-70.1.56.

[5] G. ARFKEN, *5 - Infinite Series*, in Mathematical Methods for Physicists (Third Edition), G. Arfken, ed., Academic Press, Jan. 1985, pp. 277–351, https://doi.org/10.1016/B978-0-12-059820-5.50013-6.

[6] M. ARNAUDON, F. BARBARESCO, AND L. YANG, *Riemannian medians and means with applications to radar signal processing*, IEEE Journal of Selected Topics in Signal Processing, 7 (2013), pp. 595–604, https://doi.org/10.1109/JSTSP.2013.2261798.

[7] V. ARSIGNY, P. FILLARD, X. PENNEC, AND N. AYACHE, *Log-Euclidean metrics for fast and simple calculus on diffusion tensors*, Magnetic Resonance in Medicine, 56 (2006), pp. 411–421, https://doi.org/10.1002/mrm.20965.

[8] G. BAGGIO, A. FERRANTE, AND R. SEPULCHRE, *Conal distances between rational spectral densities*, IEEE Transactions on Automatic Control, 64 (2019), pp. 1848–1857, https://doi.org/10.1109/TAC.2018.2855114.

[9] A. BARACHANT, S. BONNET, M. CONGEDO, AND C. JUTTEN, *Multiclass brain–computer interface classification by Riemannian geometry*, IEEE Transactions on Biomedical Engineering, 59 (2012), pp. 920–928, https://doi.org/10.1109/TBME.2011.2172210.

[10] A. BARACHANT, S. BONNET, M. CONGEDO, AND C. JUTTEN, *Classification of covariance matrices using a Riemannian-based kernel for BCI applications*, Neurocomputing, 112 (2013), pp. 172–178, https://doi.org/10.1016/j.neucom.2012.12.039. Advances in artificial neural networks, machine learning, and computational intelligence.

[11] P. J. BASSER, J. MATTIELLO, AND D. LEBIHAN, *MR diffusion tensor spectroscopy and imaging*, Biophysical journal, 66 (1994), pp. 259–267.

[12] R. BHATIA, *On the exponential metric increasing property*, Linear Algebra and its Applications, 375 (2003), pp. 211 – 220, https://doi.org/10.1016/S0024-3795(03)00647-5.

[13] G. BIRKHOFF, *Extensions of Jentzsch's theorem*, Transactions of the American Mathematical Society, 85 (1957), pp. 219–227, http://www.jstor.org/stable/1992971 (accessed 2023-03-15).

[14] S. BONNABEL AND R. SEPULCHRE, *Riemannian metric and geometric mean for positive semidefinite matrices of fixed rank*, SIAM Journal on Matrix Analysis and Applications, 31 (2010), pp. 1055–1070, https://doi.org/10.1137/080731347.

[15] N. Boumal, *An Introduction to Optimization on Smooth Manifolds*, Cambridge University Press, 2023, https://doi.org/10.1017/9781009166164.

[16] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre, *Manopt, a MATLAB toolbox for optimization on manifolds*, Journal of Machine Learning Research, 15 (2014), pp. 1455–1459, http://jmlr.org/papers/v15/boumal14a.html.

[17] M. R. Bridson and A. Haefliger, *Metric spaces of non-positive curvature*, vol. 319, Springer Science & Business Media, 2013.

[18] D. Brooks, O. Schwander, F. Barbaresco, J.-Y. Schneider, and M. Cord, *Riemannian batch normalization for SPD neural networks*, in Advances in Neural Information Processing Systems, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, and R. Garnett, eds., vol. 32, Curran Associates, Inc., 2019.

[19] P. J. Bushell, *Hilbert's metric and positive contraction mappings in a Banach space*, Archive for Rational Mechanics and Analysis, 52 (1973), pp. 330–338, https://doi.org/10.1007/BF00247467.

[20] B. Chen, C. Mostajeran, and S. Said, *Geometric learning of hidden Markov models via a method of moments algorithm*, Physical Sciences Forum, 5 (2022), https://doi.org/10.3390/psf2022005010.

[21] A. Cherian and S. Sra, *Riemannian dictionary learning and sparse coding for positive definite matrices*, IEEE Transactions on Neural Networks and Learning Systems, 28 (2017), pp. 2859–2871, https://doi.org/10.1109/TNNLS.2016.2601307.

[22] M. Congedo, A. Barachant, and R. Bhatia, *Riemannian geometry for EEG-based brain-computer interfaces; a primer and a review*, Brain-Computer Interfaces, 4 (2017), pp. 155–174, https://doi.org/10.1080/2326263X.2017.1297192.

[23] B. Feng, M. Fu, H. Ma, Y. Xia, and B. Wang, *Kalman filter with recursive covariance estimation—sequentially estimating process noise covariance*, IEEE Transactions on Industrial Electronics, 61 (2014), pp. 6253–6263, https://doi.org/10.1109/TIE.2014.2301756.

[24] P. T. Fletcher and S. Joshi, *Riemannian geometry for the statistical analysis of diffusion tensor data*, Signal Processing, 87 (2007), pp. 250–262, https://doi.org/10.1016/j.sigpro.2005.12.018. Tensor Signal Processing.

[25] R. Ge, C. Jin, P. Netrapalli, A. Sidford, et al., *Efficient algorithms for large-scale generalized eigenvector computation and canonical correlation analysis*, in International Conference on Machine Learning, 2016, pp. 2741–2750.

[26] J. R. Gilbert, C. Moler, and R. Schreiber, *Sparse matrices in MATLAB: Design and implementation*, SIAM Journal on Matrix Analysis and Applications, 13 (1992), pp. 333–356, https://doi.org/10.1137/0613024.

[27] G. H. Golub and H. A. van der Vorst, *Eigenvalue computation in the 20th century*, Journal of Computational and Applied Mathematics, 123 (2000), pp. 35 – 65, https://doi.org/10.1016/S0377-0427(00)00413-1. Numerical Analysis 2000. Vol. III: Linear Algebra.

[28] J. Goñi, M. P. van den Heuvel, A. Avena-Koenigsberger, N. V. de Mendizabal, R. F. Betzel, A. Griffa, P. Hagmann, B. Corominas-Murtra, J.-P. Thiran, and O. Sporns, *Resting-brain functional connectivity predicted by analytic measures of network communication*, Proceedings of the National Academy of Sciences, 111 (2014), pp. 833–838, https://doi.org/10.1073/pnas.1315529111.

[29] Z. Huang and L. Van Gool, *A Riemannian network for SPD matrix learning*, in Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17), AAAI Press, 2017-07, pp. 2036 – 2042. 31st AAAI Conference On Artificial Intelligence (AAAI-17); Conference Location: San Francisco, CA, USA; Conference Date: February 4-9, 2017.

[30] Z. Huang, R. Wang, X. Li, W. Liu, S. Shan, L. Van Gool, and X. Chen, *Geometry-aware similarity learning on SPD manifolds for visual recognition*, IEEE Transactions on Circuits and Systems for Video Technology, 28 (2018), pp. 2513–2523, https://doi.org/10.1109/TCSVT.2017.2729660.

[31] S. Jayasumana, R. Hartley, M. Salzmann, H. Li, and M. Harandi, *Kernel methods on Riemannian manifolds with Gaussian RBF kernels*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 37 (2015), pp. 2464–2477, https://doi.org/10.1109/TPAMI.2015.2414422.

[32] C. Ju and C. Guan, *Tensor-CSPNet: A novel geometric deep learning framework for motor imagery classification*, IEEE Transactions on Neural Networks and Learning Systems, (2022), pp. 1–15, https://doi.org/10.1109/TNNLS.2022.3172108.

[33] G. R. Lanckriet, N. Cristianini, P. Bartlett, L. E. Ghaoui, and M. I. Jordan, *Learning the kernel matrix with semidefinite programming*, Journal of Machine learning research, 5 (2004), pp. 27–72.

[34] S. Lang, *Fundamentals of Differential Geometry*, Springer-Verlag New York, 01 1999.

[35] B. LEMMENS AND R. NUSSBAUM, *Nonlinear Perron-Frobenius Theory*, Cambridge Tracts in Mathematics, Cambridge University Press, 2012, https://doi.org/10.1017/CBO9781139026079.

[36] L.-H. LIM, R. SEPULCHRE, AND K. YE, *Geometric distance between positive definite matrices of different dimensions*, IEEE Transactions on Information Theory, 65 (2019), pp. 5401–5405, https://doi.org/10.1109/TIT.2019.2913874.

[37] Y. LIM, *Geometry of midpoint sets for Thompson's metric*, Linear Algebra and its Applications, 439 (2013), pp. 211 – 227, https://doi.org/10.1016/j.laa.2013.03.012.

[38] H. Q. MINH AND V. MURINO, *Covariances in Computer Vision and Machine Learning*, Morgan & Claypool Publishers, 2017.

[39] N. MIOLANE, N. GUIGUI, A. L. BRIGANT, J. MATHE, B. HOU, Y. THANWERDAS, S. HEYDER, O. PELTRE, N. KOEP, H. ZAATITI, H. HAJRI, Y. CABANES, T. GERALD, P. CHAUCHAT, C. SHEWMAKE, D. BROOKS, B. KAINZ, C. DONNAT, S. HOLMES, AND X. PENNEC, *Geomstats: A Python package for Riemannian geometry in machine learning*, Journal of Machine Learning Research, 21 (2020), pp. 1–9, http://jmlr.org/papers/v21/19-027.html.

[40] B. MISHRA, G. MEYER, F. BACH, AND R. SEPULCHRE, *Low-rank optimization with trace norm penalty*, SIAM Journal on Optimization, 23 (2013), pp. 2124–2149, https://doi.org/10.1137/110859646.

[41] C. MOSTAJERAN, C. GRUSSLER, AND R. SEPULCHRE, *Affine-invariant midrange statistics*, in Geometric Science of Information, F. Nielsen and F. Barbaresco, eds., Cham, 2019, Springer International Publishing, pp. 494–501.

[42] C. MOSTAJERAN, C. GRUSSLER, AND R. SEPULCHRE, *Geometric matrix midranges*, SIAM Journal on Matrix Analysis and Applications, 41 (2020), pp. 1347–1368, https://doi.org/10.1137/19M1273475, https://doi.org/10.1137/19M1273475.

[43] C. MOSTAJERAN AND R. SEPULCHRE, *Affine-invariant orders on the set of positive-definite matrices*, in Geometric Science of Information, F. Nielsen and F. Barbaresco, eds., Cham, 2017, Springer International Publishing, pp. 613–620.

[44] C. MOSTAJERAN AND R. SEPULCHRE, *Ordering positive definite matrices*, Information Geometry, 1 (2018), pp. 287–313, https://doi.org/10.1007/s41884-018-0003-7.

[45] F. NIELSEN, *The Siegel–Klein disk: Hilbert geometry of the Siegel disk domain*, Entropy, 22 (2020), https://doi.org/10.3390/e22091019.

[46] R. D. NUSSBAUM, *Finsler structures for the part metric and Hilbert's projective metric and applications to ordinary differential equations*, Differential Integral Equations, 7 (1994), pp. 1649–1707, https://projecteuclid.org:443/euclid.die/1369329537.

[47] R. D. NUSSBAUM AND C. WALSH, *A metric inequality for the Thompson and Hilbert geometries.*, JIPAM. Journal of Inequalities in Pure & Applied Mathematics [electronic only], 5 (2004), pp. Paper No. 54, 14 p., electronic only–Paper No. 54, 14 p., electronic only, http://eudml.org/doc/124572.

[48] X. PENNEC, P. FILLARD, AND N. AYACHE, *A Riemannian framework for tensor computing*, International Journal of Computer Vision, 66 (2006), pp. 41–66, https://doi.org/10.1007/s11263-005-3222-z.

[49] X. PENNEC, S. SOMMER, AND T. FLETCHER, *Riemannian geometric statistics in medical image analysis*, Academic Press, 2020.

[50] S. SAID, L. BOMBRUN, Y. BERTHOUMIEU, AND J. H. MANTON, *Riemannian gaussian distributions on the space of symmetric positive definite matrices*, IEEE Transactions on Information Theory, 63 (2017), pp. 2153–2170, https://doi.org/10.1109/TIT.2017.2653803.

[51] S. SAID, H. HAJRI, L. BOMBRUN, AND B. C. VEMURI, *Gaussian distributions on Riemannian symmetric spaces: Statistical learning with structured covariance matrices*, IEEE Transactions on Information Theory, 64 (2018), pp. 752–772, https://doi.org/10.1109/TIT.2017.2713829.

[52] S. SAID, S. HEUVELINE, AND C. MOSTAJERAN, *Riemannian statistics meets random matrix theory: Toward learning from high-dimensional covariance matrices*, IEEE Transactions on Information Theory, 69 (2023), pp. 472–481, https://doi.org/10.1109/TIT.2022.3199479.

[53] S. SAID, C. MOSTAJERAN, AND S. HEUVELINE, *Chapter 10 - Gaussian distributions on Riemannian symmetric spaces of nonpositive curvature*, in Geometry and Statistics, F. Nielsen, A. S. Srinivasa Rao, and C. Rao, eds., vol. 46 of Handbook of Statistics, Elsevier, 2022, pp. 357–400, https://doi.org/10.1016/bs.host.2022.03.004.

[54] R. SEPULCHRE, A. SARLETTE, AND P. ROUCHON, *Consensus in non-commutative spaces*, in 49th IEEE Conference on Decision and Control (CDC), 2010, pp. 6596–6601, https://doi.org/10.1109/CDC.2010.5717072.

[55] A. SMITH, B. LAUBACH, I. CASTILLO, AND V. M. ZAVALA, *Data analysis using Riemannian geometry and applications to chemical engineering*, Computers & Chemical Engineering,

168 (2022), p. 108023, https://doi.org/10.1016/j.compchemeng.2022.108023.

[56] O. Sporns, *Network analysis, complexity, and brain function*, Complexity, 8 (2002), pp. 56–60, https://doi.org/https://doi.org/10.1002/cplx.10047.

[57] G. W. Stewart, *A Krylov–Schur algorithm for large eigenproblems*, SIAM Journal on Matrix Analysis and Applications, 23 (2002), pp. 601–614, https://doi.org/10.1137/S0895479800371529.

[58] J. Sun and A. Zhou, *Finite element methods for eigenvalue problems*, CRC Press, 2016, https://doi.org/10.1201/9781315372419.

[59] Y. Thanwerdas and X. Pennec, *Is affine-invariance well defined on SPD matrices? a principled continuum of metrics*, in Geometric Science of Information, F. Nielsen and F. Barbaresco, eds., Cham, 2019, Springer International Publishing, pp. 502–510.

[60] A. C. Thompson, *On certain contraction mappings in a partially ordered vector space*, Proceedings of the American Mathematical Society, 14 (1963), pp. 438–443, http://www.jstor.org/stable/2033816.

[61] Q. Tupker, S. Said, and C. Mostajeran, *Online learning of Riemannian hidden Markov models in homogeneous Hadamard spaces*, in Geometric Science of Information, F. Nielsen and F. Barbaresco, eds., Cham, 2021, Springer International Publishing, pp. 37–44.

[62] O. Tuzel, F. Porikli, and P. Meer, *Region covariance: A fast descriptor for detection and classification*, in Computer Vision – ECCV 2006, A. Leonardis, H. Bischof, and A. Pinz, eds., Berlin, Heidelberg, 2006, Springer Berlin Heidelberg, pp. 589–600.

[63] G. W. Van Goffrier, C. Mostajeran, and R. Sepulchre, *Inductive geometric matrix midranges*, IFAC-PapersOnLine, 54 (2021), pp. 584–589, https://doi.org/10.1016/j.ifacol.2021.06.120. 24th International Symposium on Mathematical Theory of Networks and Systems MTNS 2020.

[64] B. M. Wise and N. B. Gallagher, *The process chemometrics approach to process monitoring and fault detection*, Journal of Process Control, 6 (1996), pp. 329–348, https://doi.org/10.1016/0959-1524(96)00009-1.

[65] Y. Xu, Z. Wu, J. Li, A. Plaza, and Z. Wei, *Anomaly detection in hyperspectral images based on low-rank and sparse representation*, IEEE Transactions on Geoscience and Remote Sensing, 54 (2016), pp. 1990–2000, https://doi.org/10.1109/TGRS.2015.2493201.

[66] P. Zadeh, R. Hosseini, and S. Sra, *Geometric mean metric learning*, in Proceedings of The 33rd International Conference on Machine Learning, M. F. Balcan and K. Q. Weinberger, eds., vol. 48 of Proceedings of Machine Learning Research, New York, New York, USA, 20–22 Jun 2016, PMLR, pp. 2464–2471.