

GEOMETRIC MATRIX MIDRANGES*

CYRUS MOSTAJERAN[†], CHRISTIAN GRUSSLER^{†‡}, AND RODOLPHE SEPULCHRE[†]

Abstract. We define geometric matrix midranges for positive definite Hermitian matrices and study the midrange problem from a number of perspectives. Special attention is given to the midrange of two positive definite matrices before considering the extension of the problem to more than two matrices. We compare matrix midrange statistics with the scalar and vector midrange problem and note the special significance of the matrix problem from a computational standpoint. We also study various aspects of geometric matrix midrange statistics from the viewpoint of linear algebra, differential geometry, and convex optimization. A solution to the N -point problem is offered via convex optimization.

Key words. positive definite matrices, statistics, optimization, matrix means, midranges, Thompson metric, minimal geodesic, affine-invariance

AMS subject classifications. 15B48, 53C22, 90C26

DOI. 10.1137/19M1273475

1. Introduction. The midrange of a collection of real numbers a_1, \dots, a_N is defined as the arithmetic average of the extremal values. That is,

$$x = \frac{1}{2} \left(\min_i a_i + \max_i a_i \right).$$

This is the unique solution to the optimization problem

$$\min_{x \in \mathbb{R}} \max_i |x - a_i|.$$

In this paper, we are interested in midrange statistics in convex cones and in particular the cone of positive definite Hermitian matrices of a fixed dimension.¹ The midrange of scalar-valued data is sensitive to outliers and is therefore a nonrobust statistic. Despite this, it can be a useful measure in some contexts. For instance, the midrange is the maximally efficient estimator for the center of a uniform distribution. Thus, it can be an appropriate tool for data that is devoid of extreme outliers. It can also be useful in clustering algorithms that require the isolation of outlying clusters [15, 46, 48]. It is an important notion in the statistics of extreme events [25].

Data representations based on symmetric positive definite matrices are common in a variety of applications, from computer vision to machine learning. Often such matrices arise as covariance matrices that encode the correlations implicit in data and are thus highly structured [55]. Specific applications include brain-computer interface (BCI) systems [44, 54], radar data processing [3], and diffusion tensor imaging (DTI)

*Received by the editors July 9, 2019; accepted for publication (in revised form) by R. Vandebril June 8, 2020; published electronically September 16, 2020.

<https://doi.org/10.1137/19M1273475>

Funding: This work was supported by the European Research Council under the Advanced ERC grant agreement Switchlet 670645. The work of the first author was supported by Fitzwilliam College and the Cambridge Philosophical Society.

[†]Department of Engineering, University of Cambridge, Cambridge, CB2 1PZ, United Kingdom (csm54@cam.ac.uk, christian.grussler@berkeley.edu, r.sepulchre@eng.cam.ac.uk).

[‡]Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA 94720.

¹This paper further develops ideas partially introduced by the authors in [38].

[18]. It has been noted in many works that using nonlinear geometries related to generalized spectral properties of positive definite matrices yields significantly improved performance [43]. Indeed, Euclidean techniques for statistics and analysis on covariance matrices often result in poor accuracy and undesirable effects, such as swelling phenomena in DTI [5]. It is in this context that much attention has been paid to developing geometric statistical methods on the cone of positive definite matrices [2, 8, 9, 28, 37, 45]. A fundamental geometry that is associated to such spaces is the affine-invariant geometry [8, 39], whereby congruence transformations play the role of translations between matrices. The analogue of this geometry for scalars defined in the cone of positive real numbers $\mathbb{R}_+ = \{x \in \mathbb{R} : x > 0\}$ simply reduces to working with the logarithms of the data points and then mapping the result back to the positive cone \mathbb{R}_+ via the exponential map. Thus, we can define the affine-invariant midrange of N positive numbers $y_i > 0$ to be

$$(1.1) \quad x = \exp \left(\frac{1}{2} \left[\min_i \log y_i + \max_i \log y_i \right] \right) = \left(\min_i y_i \cdot \max_i y_i \right)^{1/2}.$$

Note that (1.1) is the unique solution of the optimization problem

$$\min_{x>0} \max_i |\log x - \log y_i| = \min_{x>0} \max_i \left| \log \frac{x}{y_i} \right|.$$

In the matrix setting, we define the geometric midrange problem on the cone of positive definite matrices as

$$(1.2) \quad \min_{X \succ 0} \max_i \|\log(Y_i^{-1/2} X Y_i^{-1/2})\|_\infty,$$

where $\{Y_1, \dots, Y_N\}$ are a collection of N positive definite matrices of dimension n , $\|\cdot\|_\infty$ denotes the spectral operator norm on the space of Hermitian matrices of dimension n defined by $\|A\|_\infty = \max\{|\lambda_1(A)|, \dots, |\lambda_n(A)|\}$, where $\lambda_j(A)$ denotes the n real eigenvalues of A and $X \succ 0$ indicates the positive definiteness of X . Note that (1.2) can be interpreted as the smallest enclosing ball problem for the collection of data $\{Y_i\}$ in Thompson geometry. In particular, (1.2) can be expressed as $\min_{X \succ 0} \max_i d_\infty(X, Y_i)$, where d_∞ denotes the Thompson metric (see section 2). The smallest enclosing ball problem of a finite set of points in Euclidean space was first posed by Sylvester in [49] and is a fundamental problem in computational geometry. The problem is also known as the minimum enclosing ball, the 1-center, or the minimax optimization problem and has been studied by several authors [6, 53]. It is an important problem that finds many applications in computer graphics and machine learning, including in collision detection, support vector clustering, and similarity search [40, 52]. On manifolds, the Riemannian smallest enclosing ball problem has been studied by Arnaudon and Nielsen in [4]. In particular, the authors consider

$$(1.3) \quad \min_{X \succ 0} \max_i \|\log(Y_i^{-1/2} X Y_i^{-1/2})\|_2,$$

which is the corresponding problem with respect to the standard affine-invariant Riemannian geometry of positive definite matrices (see subsection 2.3). Note that $\|\cdot\|_2$ in (1.3) denotes the standard Frobenius norm. Although this problem is clearly closely related to (1.2), there are fundamental differences between them. For instance, as proved by Afsari in [1], there exists a unique point that minimizes the cost function in (1.3). In contrast, even in the case of two matrices $Y_1 = A$ and $Y_2 = B$, the

solution to the optimization problem (1.2) is generally not unique. In subsection 2.1, we provide an interpretation of the Riemannian distance d_2 and Thompson distance d_∞ as members of a family of affine-invariant Finsler distances.

One particular analytic solution for the geometric midrange of two positive definite matrices A and B that will receive special consideration is $A * B$ defined by

$$(1.4) \quad A * B = \frac{1}{\sqrt{\lambda_{\min}} + \sqrt{\lambda_{\max}}} \left(B + \sqrt{\lambda_{\min} \lambda_{\max}} A \right),$$

where λ_{\max} and λ_{\min} denote the maximum and minimum generalized eigenvalues of the pencil (B, A) , which are determined by the equation $\det(B - \lambda A) = 0$. Note that λ_{\max} and λ_{\min} also coincide with the maximum and minimum eigenvalues of BA^{-1} , respectively. We will consider the properties of the expression in (1.4) in some detail in section 2. For now, it is instructive to compare $A * B$ with the well-known geometric mean $A \# B$ of positive definite matrices A and B given by

$$(1.5) \quad A \# B = A^{1/2} \left(A^{-1/2} B A^{-1/2} \right)^{1/2} A^{1/2}.$$

The matrix geometric mean has been studied in great detail by several authors and is used in a variety of applications. Much research has been devoted to extending the notion of a geometric mean from two matrices to an arbitrary number of matrices and finding efficient algorithms for computing such a mean [11, 26, 27]. These include optimization-based approaches [9] as well as inductive sequential constructions [2, 10, 34, 35].

It is noteworthy that the formula (1.4) for a matrix midrange of A and B is considerably less expensive to compute than the geometric mean (1.5), particularly for high dimensional matrices. This is because $A * B$ mainly relies on the evaluation of extremal generalized eigenvalues that can be computed efficiently using a variety of techniques such as Krylov subspace methods [22, 23, 36, 47]. In contrast, the Cholesky–Schur algorithm for computing the geometric mean (1.5) of two matrices has a complexity of $O(n^3)$ [26]. Thus, we already see an important difference between the scalar and matrix midrange problems: in the scalar case, the mean and midrange of *two* points are trivially the same, whereas a geometric midrange of two matrices may be much cheaper to compute than their geometric mean. The plots in Figures 1(a) and 1(b) illustrate how the computational cost of the midrange $A * B$ evolves with the matrix dimension as compared to the arithmetic and geometric means. The computations are based on a large number of randomly generated positive definite matrices of dimension $n = 5$ to $n = 10000$. Figure 1(c) provides a similar comparison for the cost of computing the Thompson distance versus the Euclidean and affine-invariant Riemannian distances. The logarithms in Figures 1(b) and 1(c) refer to the natural logarithm. All computations were performed on a 2017 Apple MacBook Pro laptop in MATLAB. The extremal generalized eigenvalues of (B, A) that appear in $A * B$ were computed using the `eigs` function in MATLAB with the default settings, which utilizes algorithms outlined in [31, 47].

1.1. Paper organization and contributions. The paper is organized as follows. In section 2, the midrange of two positive definite matrices is studied in detail from a variety of perspectives. We begin by proving a number of key properties of (1.4) that are expected of a measure of central tendency, including suitable order and monotonicity properties. In subsection 2.1, we present an interpretation of the geometric midrange within a unified optimization framework alongside the geometric

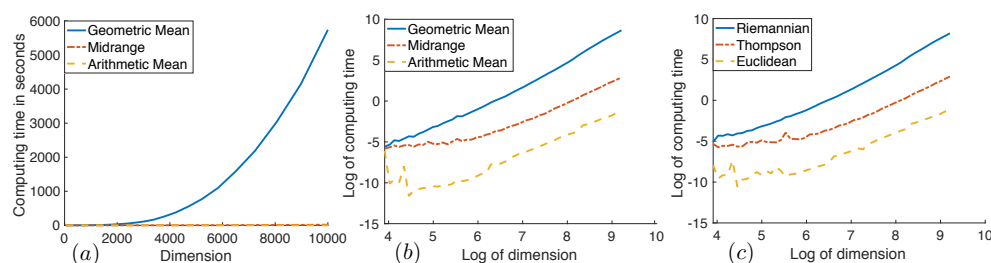


FIG. 1. (a),(b) Computing time for the geometric mean ($A\#B$), midrange ($A*B$), and arithmetic mean ($\frac{A+B}{2}$) of a pair of positive definite matrices. (c) Evolution of the cost of computing Riemannian, Thompson, and Euclidean distances with matrix dimension.

mean and median. In subsection 2.2, we present a characterization of the midrange formula (1.4) based on an extremal ordering property defined using the Löwner order. In subsection 2.3, we review the differential geometry of the manifold of $n \times n$ positive definite Hermitian matrices and consider midranges arising as midpoints of geodesics. In section 3, we define the geometric midrange problem for N positive definite matrices and study its properties in some detail. We offer a solution to the problem via convex optimization in subsection 3.1 before proving a number of optimality conditions and related results in subsection 3.2.

2. Midrange of two positive definite matrices. Let \mathbb{P}_n denote the set of $n \times n$ positive definite Hermitian matrices, which is the interior of the pointed, closed, and convex cone of positive semidefinite matrices of the same dimensions. A pointed, closed, and convex cone C in a vector space V induces a partial order on V given by $x \leq y$ if and only if $y - x \in C$. The Thompson metric [32, 50] on C is defined to be $d_\infty(x, y) = \log \max\{M(x/y; C), M(y/x; C)\}$, where $M(y/x; C) = \inf\{\lambda \in \mathbb{R} : y \leq \lambda x\}$ for $x \in C \setminus \{0\}$ and $y \in V$. For $A, B \in \mathbb{P}_n$, we have $M(A/B) = \lambda_{\max}(AB^{-1})$, so that

$$(2.1) \quad d_\infty(A, B) = \log \max\{\lambda_{\max}(AB^{-1}), \lambda_{\max}(BA^{-1})\}.$$

Noting that $\lambda_i(A^{-1/2}BA^{-1/2}) = \lambda_i(BA^{-1})$ and $\lambda_{\max}(\Sigma^{-1}) = 1/\lambda_{\min}(\Sigma)$ for any $\Sigma \in \mathbb{P}_n$, we find that the 2-point midrange problem (1.2) for data A and B takes the form

$$(2.2) \quad \min_{X \succ 0} \max\{d_\infty(A, X), d_\infty(B, X)\}.$$

A point X is said to be a Thompson midpoint of the pair (A, B) if $d_\infty(A, X) = d_\infty(B, X) = \frac{1}{2}d_\infty(A, B)$. As (\mathbb{P}_n, d_∞) forms a complete metric space [32], the minimizers of (2.2) coincide with the Thompson midpoints of (A, B) , which are generally nonunique. The geometry of the set of Thompson midpoints of a given pair of points $A, B \in \mathbb{P}_n$ is studied in detail in [33], where it is shown that the midpoint is unique if and only if the spectrum of BA^{-1} lies in a set $\{\lambda, \lambda^{-1}\}$ for some $\lambda > 0$. In this paper, we will pay special attention to the midrange $A*B$ given by (1.4) due to its scalable computational properties.

Note that (2.2) is equivalent to $\min_{X \succ 0} f(X)$, where $f(X)$ is given by

$$\max\{\log \lambda_{\max}(XA^{-1}), \log \lambda_{\max}(XB^{-1}), -\log \lambda_{\min}(XA^{-1}), -\log \lambda_{\min}(XB^{-1})\}.$$

Using this expression and the following elementary lemma, it is easy to verify that $A*B$ is indeed a Thompson metric midpoint of (A, B) .

LEMMA 2.1. *If $c_1, c_2 \in \mathbb{R}$ and M is an $n \times n$ matrix with eigenvalues $\lambda_i(M)$, then $c_1 M + c_2 I$ has eigenvalues $c_1 \lambda_i(M) + c_2$.*

Specifically, we find that for $X = A * B$, we have

$$\begin{aligned} XA^{-1} &= \frac{1}{\sqrt{\lambda_{\min}} + \sqrt{\lambda_{\max}}} \left(BA^{-1} + \sqrt{\lambda_{\min} \lambda_{\max}} I \right), \\ XB^{-1} &= \frac{1}{\sqrt{\lambda_{\min}} + \sqrt{\lambda_{\max}}} \left(I + \sqrt{\lambda_{\min} \lambda_{\max}} AB^{-1} \right), \end{aligned}$$

where λ_{\max} and λ_{\min} refer to the extremal eigenvalues of BA^{-1} . Using Lemma 2.1 and $\lambda_{\max}(BA^{-1}) = 1/\lambda_{\min}(AB^{-1})$, we find that $f(A * B)$ simplifies to $\frac{1}{2}d_{\infty}(A, B)$, as required.

We now consider the merits of the midrange $A * B$ as a measure of central tendency for $\{A, B\}$. The following are a number of properties that are desirable for such a mapping $\mu : \mathbb{P}_n \times \mathbb{P}_n \rightarrow \mathbb{P}_n$. We denote the conjugate transpose of X by X^* and the general linear group of $n \times n$ matrices by $GL(n)$.

1. Continuity: μ is a continuous map.
2. Symmetry: $\mu(A, B) = \mu(B, A)$ for all $A, B \in \mathbb{P}_n$.
3. Affine-invariance: $\mu(XAX^*, XBX^*) = X\mu(A, B)X^*$ for all $X \in GL(n)$.
4. Order property: $A \preceq B \implies A \preceq \mu(A, B) \preceq B$.
5. Monotonicity: $\mu(A, B)$ is monotone in its arguments.

We will now prove that $\mu(A, B) := A * B$ indeed satisfies properties 1–3 listed above before turning our attention to the order and monotonicity properties 4 and 5, which merit special consideration.

PROPOSITION 2.2. *The map $\mu(A, B) = A * B$ satisfies properties 1–3.*

Proof. 1. The continuity of μ follows directly from the expression for $A * B$ in (1.4), the invertibility of A , and the continuous dependence of eigenvalues on matrix entries, which itself follows from consideration of the roots of the characteristic polynomial of a matrix. 2. For symmetry, we note that $\lambda_{\min}(AB^{-1}) = 1/\lambda_{\max}(BA^{-1})$ and $\lambda_{\max}(AB^{-1}) = 1/\lambda_{\min}(BA^{-1})$, so that

$$\begin{aligned} B * A &= \frac{1}{\sqrt{1/\lambda_{\min}} + \sqrt{1/\lambda_{\max}}} \left(A + \frac{1}{\sqrt{\lambda_{\min} \lambda_{\max}}} B \right) \\ &= \frac{\sqrt{\lambda_{\min} \lambda_{\max}}}{\sqrt{\lambda_{\min}} + \sqrt{\lambda_{\max}}} \left(\frac{1}{\sqrt{\lambda_{\min} \lambda_{\max}}} B + A \right) = A * B. \end{aligned}$$

3. Affine-invariance follows immediately by noting that

$$\lambda_i(CBC^*(CAC^*)^{-1}) = \lambda_i(CBC^*(C^*)^{-1}A^{-1}C^{-1}) = \lambda_i(BA^{-1}). \quad \square$$

The order property is a generalization of the property of means of positive numbers, whereby a mean of a pair of points is expected to lie between the two points on the number line. For Hermitian matrices, a standard partial order \preceq exists according to which $A \preceq B$ if and only if $B - A$ is positive semidefinite. This partial order is known as the Löwner order, and the monotonicity in condition 5 is also with reference to this order. Unlike the case of real positive numbers $a, b > 0$, which always satisfy $a \leq b$ or $b \leq a$, two Hermitian matrices A and B may fail to satisfy both $A \preceq B$ and $B \preceq A$. It is well known that the Löwner order is affine-invariant in the sense that for all $A, B \in \mathbb{P}_n$, $X \in GL(n)$, $A \preceq B$ implies that $XAX^* \preceq XBX^*$. In particular, $A \preceq B$ if and only if $I \preceq A^{-1/2}BA^{-1/2}$. Thus, by affine-invariance of μ ,

it suffices to prove point 4 in the case where $A = I$ since $A \preceq \mu(A, B) \preceq B$ if and only if $I \preceq \mu(I, A^{-1/2}BA^{-1/2}) \preceq A^{-1/2}BA^{-1/2}$. To establish the fourth property for $\mu(A, B) = A * B$, we make use of Lemma 2.1. Let $\Sigma \in \mathbb{P}_n$ be such that $I \preceq \Sigma$, and note that this is equivalent to $\lambda_i(\Sigma) \geq 1$ for $i = 1, \dots, n$. Writing $\lambda_{\min} = \lambda_{\min}(\Sigma)$, $\lambda_{\max} = \lambda_{\max}(\Sigma)$, and $\lambda_i(\Sigma) = 1 + \delta_i$ for $\delta_i \geq 0$, we have by Lemma 2.1 that

$$\begin{aligned} \lambda_i(I * \Sigma) - 1 &= \lambda_i \left(\frac{1}{\sqrt{\lambda_{\min}} + \sqrt{\lambda_{\max}}} \left(\Sigma + \sqrt{\lambda_{\min}\lambda_{\max}} I \right) \right) - 1 \\ &= \frac{\lambda_i(\Sigma) + \sqrt{\lambda_{\min}\lambda_{\max}}}{\sqrt{\lambda_{\min}} + \sqrt{\lambda_{\max}}} - 1 \\ &= \frac{\delta_i + (\sqrt{\lambda_{\min}} - 1)(\sqrt{\lambda_{\max}} - 1)}{\sqrt{\lambda_{\min}} + \sqrt{\lambda_{\max}}} \geq 0, \end{aligned}$$

since $\lambda_i(\Sigma) \geq 1$ implies that $\sqrt{\lambda_i(\Sigma)} \geq 1$. Thus, we have shown that $I \preceq \Sigma$ implies $I \preceq I * \Sigma$. To prove the other inequality, let $\lambda_i(\Sigma) = \lambda_{\min}(\Sigma) + \epsilon_i$ for $\epsilon_i \geq 0$, and note that

$$\begin{aligned} \lambda_i(\Sigma - I * \Sigma) &= \lambda_i \left(\left(\frac{\sqrt{\lambda_{\min}} + \sqrt{\lambda_{\max}} - 1}{\sqrt{\lambda_{\min}} + \sqrt{\lambda_{\max}}} \right) \Sigma - \frac{\sqrt{\lambda_{\min}\lambda_{\max}}}{\sqrt{\lambda_{\min}} + \sqrt{\lambda_{\max}}} I \right) \\ &= \left(\frac{\sqrt{\lambda_{\min}} + \sqrt{\lambda_{\max}} - 1}{\sqrt{\lambda_{\min}} + \sqrt{\lambda_{\max}}} \right) \lambda_i(\Sigma) - \frac{\sqrt{\lambda_{\min}\lambda_{\max}}}{\sqrt{\lambda_{\min}} + \sqrt{\lambda_{\max}}} \\ &= \sqrt{\lambda_{\min}} (\sqrt{\lambda_{\min}} - 1) + \left(\frac{\sqrt{\lambda_{\min}} + \sqrt{\lambda_{\max}} - 1}{\sqrt{\lambda_{\min}} + \sqrt{\lambda_{\max}}} \right) \epsilon_i \geq 0, \end{aligned}$$

as $I \preceq \Sigma$ ensures that $\sqrt{\lambda_{\min}} \geq 1$. Therefore, we have also shown that $\Sigma - I * \Sigma \succeq 0$. That is,

$$(2.3) \quad I \preceq \Sigma \implies I \preceq I * \Sigma \preceq \Sigma$$

for all $\Sigma \in \mathbb{P}_n$. In particular, upon substituting $\Sigma = A^{-1/2}BA^{-1/2}$ in (2.3) and using the affine-invariance properties of both the Löwner order and the mean $\mu(A, B) = A * B$, we establish the following important property.

PROPOSITION 2.3. *For $A, B \in \mathbb{P}_n$, $A \preceq B$ implies that $A \preceq A * B \preceq B$.*

We now consider the fifth and final desirable property of $\mu : \mathbb{P}_n \times \mathbb{P}_n \rightarrow \mathbb{P}_n$, which is monotonicity of μ in its arguments. First recall that a map $F : \mathbb{P}_n \rightarrow \mathbb{P}_n$ is said to be monotone if $\Sigma_1 \preceq \Sigma_2$ implies that $F(\Sigma_1) \preceq F(\Sigma_2)$. By symmetry and affine-invariance, it is sufficient to consider monotonicity of $\mu(I, \Sigma)$ with respect to Σ . That is, monotonicity is established by showing that

$$\Sigma_1 \preceq \Sigma_2 \implies I * \Sigma_1 \preceq I * \Sigma_2.$$

However, it turns out that $F(\Sigma) := I * \Sigma$ is not monotone with respect to Σ , as we demonstrate below. Nonetheless, F is seen to enjoy certain weaker monotonicity properties. Considering the eigenvalues of $I * \Sigma$, we find that

$$(2.4) \quad \lambda_i(I * \Sigma) = \frac{\lambda_i(\Sigma) + \sqrt{\lambda_{\min}\lambda_{\max}}}{\sqrt{\lambda_{\min}} + \sqrt{\lambda_{\max}}},$$

where λ_{\min} and λ_{\max} refer to the smallest and largest eigenvalues of Σ .

PROPOSITION 2.4. *The maximum and minimum eigenvalues of $F(\Sigma) = I * \Sigma$ are monotone with respect to Σ .*

Proof. Considering the cases $i = 1$ and $i = n$, we find that (2.4) yields

$$\lambda_{\min}(I * \Sigma) = \sqrt{\lambda_{\min}(\Sigma)} \quad \text{and} \quad \lambda_{\max}(I * \Sigma) = \sqrt{\lambda_{\max}(\Sigma)},$$

both of which are seen to be monotone functions of Σ . \square

It is in the sense of the above that $\mu(A, B) = A * B$ inherits a weak monotonicity property. The monotonic dependence of the extremal eigenvalues of $I * \Sigma$ on Σ ensures that if $\Sigma_1 \preceq \Sigma_2$, then we can at least rule out the possibility that $I * \Sigma_1 \succ I * \Sigma_2$, where $\succ 0$ here denotes positive definiteness. To prove that monotonicity is generally not satisfied in the full sense, consider a diagonal matrix $\Sigma = \text{diag}(a, b, x) \in \mathbb{P}(3)$, where $\lambda_{\min}(\Sigma) = a < b \leq x = \lambda_{\max}(\Sigma)$ and x is thought of as a variable. We have $I * \Sigma = \text{diag}(\sqrt{a}, f(x), \sqrt{x})$, where

$$\lambda_2(I * \Sigma) = f(x) := \frac{b + \sqrt{ax}}{\sqrt{a} + \sqrt{x}}.$$

Taking the derivative of f with respect to x , we find that

$$f'(x) = \frac{a - b}{2\sqrt{x}(\sqrt{a} + \sqrt{x})^2} < 0 \quad \text{for all } x \geq b,$$

which shows that the second eigenvalue of $I * \Sigma$ decreases as x increases. Thus, we see that $I * \Sigma$ cannot depend monotonically on Σ in this example.

As a summary, we collect the main results so far in the following theorem.

THEOREM 2.5. *The midrange $\mu(A, B) = A * B$ defined in (1.4) yields a Thompson metric midpoint of $A, B \in \mathbb{P}_n$ that is continuous, symmetric, and affine-invariant. Moreover, if $A \preceq B$, then $A \preceq \mu(A, B) \preceq B$, and the extremal eigenvalues of $\mu(I, \Sigma)$ depend monotonically on $\Sigma \in \mathbb{P}_n$.*

We also note that $A * B$ satisfies a key scaling property, which suggests that it may be a plausible candidate for a computationally scalable substitute for the standard geometric mean $A \# B$ of two positive definite matrices.

PROPOSITION 2.6. *For any real scalars $a, b > 0$ and matrices $A, B \in \mathbb{P}_n$, we have*

$$(2.5) \quad (aA) * (bB) = \sqrt{ab}(A * B).$$

Proof. The result follows upon substituting $\lambda_i((bB)(aA)^{-1}) = \frac{b}{a}\lambda_i(BA^{-1})$ into the formula (1.4). \square

Remark 2.7. The scaling in (2.5) of course does not generally hold for a mean of two matrices. Indeed, it does not generally hold for means arising as d_∞ -midpoints either. For instance, [33] identifies

$$(2.6) \quad A \diamond B = \begin{cases} \frac{\sqrt{\lambda_{\max}}}{1 + \lambda_{\max}}(A + B) & \text{if } \lambda_{\min}\lambda_{\max} \geq 1, \\ \frac{\sqrt{\lambda_{\min}}}{1 + \lambda_{\min}}(A + B) & \text{if } \lambda_{\min}\lambda_{\max} \leq 1 \end{cases}$$

as another d_∞ -midpoint of A and B . Clearly $A \diamond B$ does not scale geometrically in the sense of (2.5).

2.1. An optimization-based formulation. A norm $\|\cdot\|$ on the space of $n \times n$ complex matrices is said to be unitarily invariant if $\|UXV\| = \|X\|$ for all $n \times n$ matrices X and unitary matrices U, V . A norm Φ on \mathbb{R}^n is called a symmetric gauge norm if it is invariant under permutations and sign changes of coordinates. Consider the family of affine-invariant metric distances d_Φ on \mathbb{P}_n defined as

$$(2.7) \quad d_\Phi(A, B) = \|\log A^{-1/2} B A^{-1/2}\|_\Phi,$$

where $\|\cdot\|_\Phi$ is any unitarily invariant norm on the space of Hermitian matrices of dimension n defined by $\|X\|_\Phi := \Phi(\lambda_1(X), \dots, \lambda_n(X))$, with $\lambda_{\min}(X) = \lambda_n(X) \leq \dots \leq \lambda_1(X) = \lambda_{\max}(X)$ denoting the n real eigenvalues of X and Φ a symmetric gauge norm on \mathbb{R}^n [7]. The norms $\|\cdot\|_\Phi$ induced by the l_p -norms on \mathbb{R}^n for $1 \leq p \leq \infty$ are called the Schatten p -norms. For the choice of $\Phi(x_1, \dots, x_n) = (\sum_i x_i^2)^{1/2}$, $d_2 := d_\Phi$ corresponds to the metric distance generated by the standard affine-invariant Riemannian metric on \mathbb{P}_n given by $\langle X, Y \rangle_\Sigma = \text{tr}(\Sigma^{-1} X \Sigma^{-1} Y)$ for $\Sigma \in \mathbb{P}_n$ and Hermitian matrices $X, Y \in T_\Sigma \mathbb{P}_n$. The length element ds of this geometry satisfies $ds^2 = \text{tr}(\Sigma^{-1} d\Sigma)^2$. The unique (up to parametrization) Riemannian geodesic from A to B is given by the curve $\gamma_G : [0, 1] \rightarrow \mathbb{P}_n$ defined by

$$(2.8) \quad \gamma_G(t) = A^{1/2} \left(A^{-1/2} B A^{-1/2} \right)^t A^{1/2}.$$

This curve is significant as a minimal geodesic for *any* of the affine-invariant metrics d_Φ [7]. The midpoint of γ_G is the matrix geometric mean $A \# B$ (1.5), which is a metric midpoint in the sense that $d_\Phi(A, A \# B) = d_\Phi(A \# B, B) = \frac{1}{2} d_\Phi(A, B)$ for any choice of symmetric gauge Φ . With $\Phi(x_1, \dots, x_n) = \max_i |x_i|$, $d_\Phi = d_\infty$ yields the distance function that coincides with the Thompson metric [50] on the cone \mathbb{P}_n ,

$$d_\infty(A, B) = \|\log A^{-1/2} B A^{-1/2}\|_\infty = \max\{\log \lambda_{\max}(B A^{-1}), \log \lambda_{\max}(A B^{-1})\}.$$

Therefore, we see that the geometric mean $A \# B$ is also a geometric midrange of A and B .

The invariant Finsler metrics (2.7) provide a route to geometrically generalize several measures of aggregation of data to the space of positive definite matrices \mathbb{P}_n . Specifically, the mean, median, and midrange of a collection of real numbers a_1, \dots, a_N can be defined as

$$\begin{aligned} \operatorname{argmin}_{x \in \mathbb{R}} \left(\sum_i (x - a_i)^2 \right)^{1/2} &= \operatorname{argmin}_{x \in \mathbb{R}} \|x \mathbf{1} - \mathbf{a}\|_2, \\ \operatorname{argmin}_{x \in \mathbb{R}} \sum_i |x - a_i| &= \operatorname{argmin}_{x \in \mathbb{R}} \|x \mathbf{1} - \mathbf{a}\|_1, \\ \operatorname{argmin}_{x \in \mathbb{R}} \max_i |x - a_i| &= \operatorname{argmin}_{x \in \mathbb{R}} \|x \mathbf{1} - \mathbf{a}\|_\infty, \end{aligned}$$

respectively, where $\mathbf{1} = (1, \dots, 1) \in \mathbb{R}^N$ and $\mathbf{a} = (a_1, \dots, a_N)$. By analogy, one can extend these notions to geometric averages for a collection of data $Y_1, \dots, Y_N \in \mathbb{P}_n$ arising as

$$(2.9) \quad \operatorname{argmin}_{X \succ 0} \Phi_N(d_{\Phi_n}(X, Y_i)),$$

where Φ_n denotes the gauge norm on the space of $n \times n$ Hermitian matrices and Φ_N denotes the corresponding gauge function acting on the N distances $d_{\Phi_n}(X, Y_i)$. If Φ

corresponds to the l_2 vector norm, (2.9) yields the *geometric mean* \mathcal{G}_2 of Y_1, \dots, Y_N , also known as the Karcher mean [8, 9, 37]:

$$(2.10) \quad \mathcal{G}_2(Y_1, \dots, Y_N) = \operatorname{argmin}_{X \succ 0} \sum_{i=1}^N d_2(X, Y_i)^2.$$

If $N = 2$, the unique solution $\mathcal{G}_2(A, B)$ of (2.10) coincides with the geometric mean $A \# B$. One can also define *geometric medians* of Y_1, \dots, Y_N to be solutions to (2.9) for the choice of $\Phi(x_1, \dots, x_n) = \sum_i |x_i|$:

$$(2.11) \quad \mathcal{G}_1(Y_1, \dots, Y_N) = \operatorname{argmin}_{X \succ 0} \sum_{i=1}^N d_1(X, Y_i).$$

Note that the d_1 distance of $X \in \mathbb{P}_n$ to the identity I takes the form

$$(2.12) \quad d_1(X, I) = \|\log X\|_1 = \sum_{i=1}^N |\log \lambda_i(X)| = \operatorname{tr} \left((\log X \log X)^{1/2} \right).$$

It is interesting to compare (2.12) to the function $F(X) = \log \det(X)$, which plays an important role in convex optimization [13]. In particular, we have

$$F(X) = \log \det(X) = \operatorname{tr}(\log X) = \sum_{i=1}^N \log \lambda_i(X).$$

If $X \succeq I$, then $\log X \succeq 0$ and hence $d_1(X, I) = \operatorname{tr}(\log X) = \log \det X$.

If Φ corresponds to the l_∞ -norm, then (2.9) yields the geometric midrange problem (1.2). In the $N = 2$ case, we have already seen that $A * B$ is a solution to the corresponding midrange optimization problem $\min_{X \succ 0} \max\{d_\infty(X, A), d_\infty(X, B)\}$. The N -point problem is studied in more detail in section 3.

2.2. Extremal ordering property. Here we describe a characterization of the midrange $A * B$ of $A, B \in \mathbb{P}_n$ that does not rely on any additional structures on \mathbb{P}_n except for the standard Löwner partial order \succeq . It is remarkable that such a characterization that is independent of any metric or differential geometric structure on \mathbb{P}_n exists.

THEOREM 2.8. *Let $A, B \in \mathbb{P}_n$. Then*

$$\begin{aligned} A * B &= \max_{X \in \operatorname{span}\{A, B\}} \left\{ X = X^* : \begin{pmatrix} A & X \\ X & B \end{pmatrix} \succeq 0 \right\} \\ &= \max_{a, b \in \mathbb{R}} \left\{ aA + bB : \begin{pmatrix} A & aA + bB \\ aA + bB & B \end{pmatrix} \succeq 0 \right\}. \end{aligned}$$

Proof. For any $X = X^*$, we have the congruence relation

$$\begin{aligned} \begin{pmatrix} A & X \\ X & B \end{pmatrix} &\sim \begin{pmatrix} I & -XB^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} A & X \\ X & B \end{pmatrix} \begin{pmatrix} I & 0 \\ -B^{-1}X & I \end{pmatrix} \\ &= \begin{pmatrix} A - XB^{-1}X & 0 \\ 0 & B \end{pmatrix}. \end{aligned}$$

This matrix is clearly positive semidefinite if and only if $A \succeq XB^{-1}X$, which is equivalent to

$$(2.13) \quad B^{-1/2}AB^{-1/2} \succeq B^{-1/2}XB^{-1}XB^{-1/2} = (B^{-1/2}XB^{-1/2})^2.$$

Using the monotonicity of the square root function $\Sigma \mapsto \Sigma^{1/2}$ on \mathbb{P}_n and the affine-invariance of the Löwner order, (2.13) holds if and only if

$$(2.14) \quad X \preceq B^{1/2}(B^{-1/2}AB^{-1/2})^{1/2}B^{1/2}.$$

The expression on the right-hand side is of course the geometric mean $A\#B$, and thus we have

$$A\#B = \max_{X \succeq 0} \left\{ X = X^* : \begin{pmatrix} A & X \\ X & B \end{pmatrix} \succeq 0 \right\}.$$

If we restrict X to be in the real span of A and B , (2.14) becomes

$$aA + bB \preceq B^{1/2}(B^{-1/2}AB^{-1/2})^{1/2}B^{1/2},$$

which is equivalent to $a\Sigma + bI \preceq \Sigma^{1/2}$ for $\Sigma = B^{-1/2}AB^{-1/2}$. By diagonalizing $\Sigma^{1/2}$ we obtain a unitary matrix V such that $\Sigma = V^*DV$ and $\Sigma^{1/2} = V^*D^{1/2}V$, where $D = \text{diag}(\lambda_1(\Sigma), \dots, \lambda_n(\Sigma))$. Therefore, we have $aD + bI \preceq D^{1/2}$, which is equivalent to

$$(2.15) \quad a\lambda_i(\Sigma) + b \leq \sqrt{\lambda_i(\Sigma)}$$

for $i = 1, \dots, n$. If we require that equality hold in (2.15) for $i = 1$ and $i = n$, so that

$$a\lambda_{\min}(\Sigma) + b = \sqrt{\lambda_{\min}(\Sigma)} \quad \text{and} \quad a\lambda_{\max}(\Sigma) + b = \sqrt{\lambda_{\max}(\Sigma)},$$

we find that

$$\begin{aligned} a &= \frac{\sqrt{\lambda_{\max}(\Sigma)} - \sqrt{\lambda_{\min}(\Sigma)}}{\lambda_{\max}(\Sigma) - \lambda_{\min}(\Sigma)} = \frac{1}{\sqrt{\lambda_{\max}(\Sigma)} + \sqrt{\lambda_{\min}(\Sigma)}}, \\ b &= \frac{\sqrt{\lambda_{\min}(\Sigma)}\lambda_{\max}(\Sigma)}{\sqrt{\lambda_{\max}(\Sigma)} + \sqrt{\lambda_{\min}(\Sigma)}}. \end{aligned}$$

For this choice of a and b , we have $aA + bB = A\#B$. Moreover, (2.15) is satisfied for each $i = 1, \dots, n$ since

$$\begin{aligned} \sqrt{\lambda_i(\Sigma)} - a\lambda_i(\Sigma) - b &= \frac{\sqrt{\lambda_{\min}\lambda_i} + \sqrt{\lambda_{\max}\lambda_i} - \lambda_i - \sqrt{\lambda_{\min}\lambda_{\max}}}{\sqrt{\lambda_{\min}} + \sqrt{\lambda_{\max}}} \\ &= \frac{(\sqrt{\lambda_{\max}} - \sqrt{\lambda_i})(\sqrt{\lambda_i} - \sqrt{\lambda_{\min}})}{\sqrt{\lambda_{\max}} + \sqrt{\lambda_{\min}}} \geq 0. \end{aligned}$$

Imposing equality in (2.15) for any pair of indices other than $i = 1$ and $i = n$ would yield coefficients a and b that result in the violation of some of the other inequalities in (2.15). Therefore, our choice of a and b is indeed optimal. \square

2.3. Differential geometric viewpoint. The set \mathbb{P}_n is a smooth manifold whose tangent space $T_\Sigma \mathbb{P}_n$ at any point $\Sigma \in \mathbb{P}_n$ can be identified with the set of $n \times n$ Hermitian matrices \mathbb{H}_n . The matrix exponential map $X \mapsto e^X$ maps \mathbb{H}_n bijectively onto \mathbb{P}_n . Its differential $de^X : \mathbb{H}_n \rightarrow \mathbb{H}_n$ at X is the linear map given by $de^X(Z) = \frac{d}{dt}(e^{X+tZ})|_{t=0}$. The following exponential metric increasing property was established by Bhatia in [7]. See [29] for an earlier version of the theorem.

THEOREM 2.9. *For any symmetric gauge norm Φ and Hermitian matrices X and Z , we have*

$$\|Z\|_\Phi \leq \|e^{-X/2} de^X(Z) e^{-X/2}\|_\Phi,$$

where $\|\cdot\|_\Phi$ denotes the unitarily invariant norm induced by Φ .

This theorem has several important consequences, which we will briefly review. First note that the distance functions d_Φ defined in (2.7) are induced by the affine-invariant Finsler structures on \mathbb{P}_n given by

$$(2.16) \quad \|d\Sigma\|_{\Sigma, \Phi} := \|\Sigma^{-1/2} d\Sigma \Sigma^{-1/2}\|_\Phi$$

for $\Sigma \in \mathbb{P}_n$ and $d\Sigma \in T_\Sigma \mathbb{P}_n$. For our purposes, we can think of a Finsler structure on \mathbb{P}_n as a smoothly varying norm on the tangent bundle of \mathbb{P}_n . Such a structure can be used to calculate the length of any smooth curve γ in \mathbb{P}_n . We can express any such curve as the image of a curve Γ in \mathbb{H}_n under the exponential map. In particular, any smooth curve $\gamma : [0, 1] \rightarrow \mathbb{P}_n$ from I to Σ can be expressed as $\gamma(t) = e^{\Gamma(t)}$, where $\Gamma(0) = 0$ and $\Gamma(1) = \log(\Sigma) \in \mathbb{H}_n$. The length of this curve with respect to the Finsler structure (2.16) is

$$\begin{aligned} L_\Phi[\gamma] &= \int_0^1 \|\gamma'(t)\|_{\gamma(t), \Phi} dt = \int_0^1 \|\gamma(t)^{-1/2} \gamma'(t) \gamma(t)^{-1/2}\|_\Phi dt \\ &= \int_0^1 \|\gamma(t)^{-1/2} de^{\Gamma(t)}(\Gamma'(t)) \gamma(t)^{-1/2}\|_\Phi dt \\ &\geq \int_0^1 \|\Gamma'(t)\|_\Phi dt. \end{aligned}$$

The last integral is simply the length of the curve Γ in \mathbb{H}_n and the least value it can take is $\|\log \Sigma\|_\Phi$, which is attained by the straight line segment from 0 to $\log \Sigma$ in \mathbb{H}_n . The distance between I and Σ is defined as $d_\Phi(I, \Sigma) = \inf_\gamma L_\Phi[\gamma]$, where the infimum is taken over all smooth curves γ from I to Σ . Therefore, we see that

$$d_\Phi(I, \Sigma) = \|\log \Sigma\|_\Phi,$$

and note that this distance is attained by the curve $\gamma(t) = e^{t \log \Sigma}$, which is a geodesic from I to Σ . See Figure 2.

Since congruence transformations are isometries of (\mathbb{P}_n, d_Φ) , it follows that the curve $\gamma_G^{A,B}(t) = A^{1/2} \exp(t \log(A^{-1/2} B A^{-1/2})) A^{1/2}$ is a geodesic from A to B . Note that this is precisely in agreement with (2.8). This geodesic is unique provided that the geodesics in \mathbb{R}^n induced by Φ are unique. In particular, uniqueness of geodesics in (\mathbb{P}_n, d_Φ) is inherited from \mathbb{R}^n when Φ corresponds to the l_p -norms for $1 < p < \infty$, but not for $p = 1, \infty$.

The exponential metric increasing property can also be used to show that the metric space (\mathbb{P}_n, d_Φ) is a space of nonpositive curvature for any choice of Φ . See [7, 14] for further details. A closely related result is the geodesic convexity [30] of

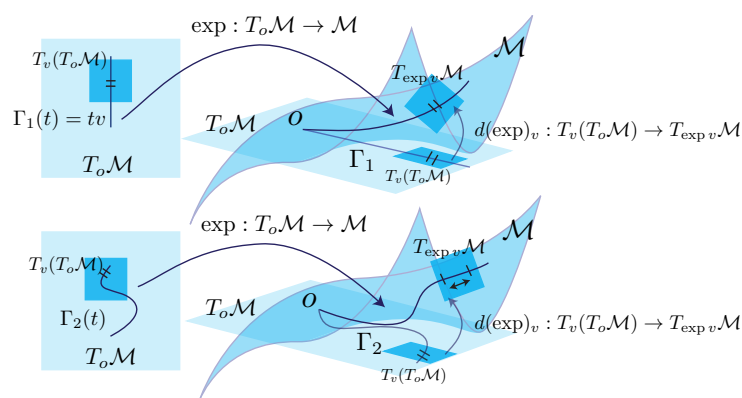


FIG. 2. The figure depicts a manifold \mathcal{M} whose exponential map preserves the length of rays through the origin (top), but generally increases the length of curves (bottom) as measured with respect to some Finsler structure on \mathcal{M} . (\mathbb{P}_n, d_Φ) is a manifold that satisfies such an exponential metric increasing property.

d_Φ , which follows from the inequality $d_\Phi(e^{tX}, e^{tZ}) \leq t d_\Phi(e^X, e^Z)$ for all $X, Z \in \mathbb{H}_n$ and $0 \leq t \leq 1$. More generally, the geodesic convexity theorem states that for all $A_1, A_2, B_1, B_2 \in \mathbb{P}_n$, the real function

$$(2.17) \quad t \mapsto d_\Phi \left(\gamma_{\mathcal{G}}^{A_1, A_2}(t), \gamma_{\mathcal{G}}^{B_1, B_2}(t) \right)$$

is convex for any symmetric gauge norm Φ [7].

In geometric midrange statistics we are interested in the distance d_∞ , which coincides with the Thompson metric on the cone of positive definite Hermitian matrices. It is known that the Thompson metric does not admit unique minimal geodesics. Indeed, a remarkable construction by Nussbaum in [41] describes a family of geodesics that generally consists of an infinite number of curves connecting a pair of points in a cone C . In particular, setting $\alpha := 1/M(x/y; C)$ and $\beta := M(y/x; C)$, the curve $\phi: [0, 1] \rightarrow C$ given by

$$(2.18) \quad \phi(t; x, y) := \begin{cases} \left(\frac{\beta^t - \alpha^t}{\beta - \alpha} \right) y + \left(\frac{\beta\alpha^t - \alpha\beta^t}{\beta - \alpha} \right) x & \text{if } \alpha \neq \beta, \\ \alpha^t x & \text{if } \alpha = \beta \end{cases}$$

is always a minimal geodesic from x to y with respect to the Thompson metric. The curve ϕ defines a projective straight line in the cone. If we take C to be the cone of positive semidefinite matrices with interior $\text{int } C = \mathbb{P}_n$, then for a pair of points $A, B \in \mathbb{P}_n$, we have $\beta = M(B/A; C) = \lambda_{\max}(BA^{-1})$ and $\alpha = 1/M(A/B; C) = \lambda_{\min}(BA^{-1})$. Thus, the minimal geodesic described by (2.18) takes the form

$$(2.19) \quad \phi(t) := \begin{cases} \left(\frac{\lambda_{\max}^t - \lambda_{\min}^t}{\lambda_{\max} - \lambda_{\min}} \right) B + \left(\frac{\lambda_{\max}\lambda_{\min}^t - \lambda_{\min}\lambda_{\max}^t}{\lambda_{\max} - \lambda_{\min}} \right) A & \text{if } \lambda_{\min} \neq \lambda_{\max}, \\ \lambda_{\min}^t A & \text{if } \lambda_{\min} = \lambda_{\max}, \end{cases}$$

where λ_{\max} and λ_{\min} denote the largest and smallest eigenvalues of BA^{-1} , respectively. Taking the midpoint $t = 1/2$ of this geodesic, we recover the d_∞ -midpoint

$A * B$ (1.4) of A and B . Thus, we have arrived at another interpretation of $A * B$ as the midpoint of a suitable geodesic in (\mathbb{P}_n, d_∞) . The result follows from elementary algebraic simplification upon setting $A * B = \phi(1/2; A, B)$ in the case $\lambda_{\min} \neq \lambda_{\max}$. If $\lambda_{\min} = \lambda_{\max}$, then $\phi(1/2; A, B) = \sqrt{\lambda_{\min}} A$ also agrees with the formula in (1.4). It is shown in [33] that $\phi = \phi(t; A, B)$ is the unique d_∞ geodesic connecting A to B if and only if the spectrum of BA^{-1} consists of at most two distinct eigenvalues, one of which is the reciprocal of the other. Moreover, it is shown that otherwise there are infinitely many d_∞ minimal geodesics from A to B , and that the set of d_∞ -midpoints of A and B is compact and convex in both Riemannian and Euclidean senses [33].

3. The N -point geometric midrange problem. Given a collection of N points Y_1, \dots, Y_N in \mathbb{P}_n , the midrange problem can be formulated as the following optimization problem:

$$(3.1) \quad \min_{X \succ 0} \max_i d_\infty(X, Y_i).$$

We call a solution X^* to the above problem a midrange of $\{Y_i\}$. Note that the cost function $f(X) := \max_i d_\infty(X, Y_i)$ is continuous but not smooth. That is, (3.1) is a nonsmooth continuous optimization problem on a smooth Finsler manifold.

PROPOSITION 3.1. *The optimum cost $t^* = \min_{X \succ 0} \max_i d_\infty(X, Y_i)$ of (3.1) satisfies $l \leq t^* \leq u$, where the lower and upper bounds are given by*

$$(3.2) \quad l = \frac{1}{2} \text{diam}_\infty(\{Y_i\}) := \frac{1}{2} \max_{i,j} d_\infty(Y_i, Y_j), \quad u = \min_i \max_j d_\infty(Y_i, Y_j) \leq 2l.$$

Proof. Let X^* denote a midrange of $\{Y_i\}$ so that $t^* = \max_i d_\infty(X^*, Y_i)$. By the triangle inequality, we have for any $i, j = 1, \dots, N$,

$$d_\infty(Y_i, Y_j) \leq d_\infty(Y_i, X^*) + d_\infty(X^*, Y_j) \leq t^* + t^* = 2t^*.$$

Taking the maximum of the left-hand side over i, j , we arrive at $l = \frac{1}{2} \text{diam}_\infty(\{Y_i\}) \leq t^*$. For the upper bound, note that taking $X = Y_i$ for each i , we obtain a cost $f(Y_i) = \max_j d_\infty(Y_i, Y_j)$. The minimum value of these N cost evaluations will clearly still yield an upper bound on the optimum cost t^* . Thus we have $t^* \leq u = \min_i \max_j d_\infty(Y_i, Y_j)$. \square

Note that it is possible to have a collection of points $\{Y_i\}$ for which either l or u is attained. For instance, l is clearly attained by the d_∞ -midpoint when $\{Y_i\}$ consists of a pair of points. Similarly, u is attained if we have 3 points Y_1, Y_2, Y_3 , where Y_3 happens to be a d_∞ -midpoint of Y_1 and Y_2 . In general, the upper bound is attained if the midrange coincides with one of the data points.

It is instructive to consider the N -point affine-invariant midrange of vectors in the positive orthant. In the vector case, the midrange problem in \mathbb{R}_+^n takes the form

$$(3.3) \quad \min_{\mathbf{x} \succ 0} \max_i \|\log \mathbf{x} - \log \mathbf{y}_i\|_\infty := \min_{\mathbf{x} \succ 0} \max_i \max_a |\log x^a - \log y_i^a|,$$

where $\mathbf{x} \succ 0$ means that $\mathbf{x} = (x^a)$ satisfies $x^a > 0$ for $a = 1, \dots, n$ and \mathbf{y}_i are a collection of N given points in \mathbb{R}_+^n . As in the matrix case, the optimum cost $t^* = \min_{\mathbf{x} \succ 0} f(\mathbf{x}) = \min_{\mathbf{x} \succ 0} \max_i \|\log \mathbf{x} - \log \mathbf{y}_i\|_\infty$ has a lower bound

$$(3.4) \quad l = \frac{1}{2} \max_{i,j} \|\log \mathbf{y}_i - \log \mathbf{y}_j\|_\infty.$$

PROPOSITION 3.2. *The lower bound (3.4) is attained by $\mathbf{x}^* = (x^a) \in \mathbb{R}_+^n$ defined by $x^a = (\min_i y_i^a \cdot \max_i y_i^a)^{1/2}$.*

Proof. Note that

$$l = \frac{1}{2} \max_{i,j} \max_a \left| \log \frac{y_i^a}{y_j^a} \right| = \frac{1}{2} \max_a \left| \log \frac{\max_i y_i^a}{\min_j y_j^a} \right|.$$

With \mathbf{x}^* as defined in Proposition 3.2, we have

$$\begin{aligned} f(\mathbf{x}^*) &= \max_{k,a} \left| \frac{1}{2} \log \left(\min_i y_i^a \cdot \max_i y_i^a \right) - \log y_k^a \right| = \max_{k,a} \left| \frac{1}{2} \log \left(\frac{\min_i y_i^a \cdot \max_i y_i^a}{y_k^a \cdot y_k^a} \right) \right| \\ &= \max_a \left| \frac{1}{2} \log \left(\frac{\min_i y_i^a \cdot \max_i y_i^a}{(\min_i y_i^a)^2} \right) \right| = \max_a \left| \frac{1}{2} \log \left(\frac{\max_i y_i^a}{\min_j y_j^a} \right) \right| = l. \quad \square \end{aligned}$$

Remark 3.3. The midrange problem does not generally have a unique solution in the vector case, as can be readily seen through simple examples. For instance, the problem in \mathbb{R}_+^2 with $N = 2$ and $\mathbf{y}_1 = (a, 1)$, $\mathbf{y}_2 = (1/a, 1)$ for some $a > 1$ has the solution $\mathbf{x} = (1, s)$ for any s satisfying $1/a < s < a$.

3.1. Geometric midranges via convex optimization. The geometric matrix midrange problem (3.1) can be written as

$$\min_{X \succ 0} \max_i \{ |\log(\lambda_{\min}(Y_i^{-1/2} X Y_i^{-1/2}))|, |\log(\lambda_{\max}(Y_i^{-1/2} X Y_i^{-1/2}))| \},$$

which has the equivalent epigraph formulation

$$\begin{cases} \min_{X \succ 0, t \in \mathbb{R}} t \\ -t \leq \log(\lambda_{\max}(Y_i^{-\frac{1}{2}} X Y_i^{-\frac{1}{2}})) \leq t \text{ for all } i, \\ -t \leq \log(\lambda_{\min}(Y_j^{-\frac{1}{2}} X Y_j^{-\frac{1}{2}})) \leq t \text{ for all } j. \end{cases}$$

This can be rewritten as the *quasiconvex* problem

$$(3.5) \quad \begin{cases} \min_{X \succ 0, t \in \mathbb{R}} t \\ e^{-t} Y_i \preceq X \preceq e^t Y_i. \end{cases}$$

While this problem is not convex due to the presence of the log function, the feasibility condition $e^{-t} Y_i \preceq X \preceq e^t Y_i$ is convex for *fixed* t and can be solved using standard convex optimization packages such as CVX [24]. Given a t that is greater than or equal to the optimum value $t^* = \min_{X \succ 0} \max_i d_\infty(X, Y_i)$, we can solve (3.5) using the bisection method [13] by successively solving the feasibility problem as we effectively decrease t . In the bisection method it is desirable to have a good estimate for the initial t as the successive reductions in t can be quite slow. In particular, if the lower bound $l = \frac{1}{2} \text{diam}_\infty(\{Y_i\})$ is attained as in the vector case, then we can solve (3.5) in one step by taking $t = l$ and solving the feasibility condition once. However, rather remarkably, numerical examples show that, unlike the scalar and vector cases, the lower bound l is not always attained in the geometric matrix midrange problem.

PROPOSITION 3.4. *The lower bound $l = \frac{1}{2} \text{diam}_\infty(\{Y_i\})$ is not necessarily attained in (3.1).*

Proof. Consider the $N = 3$ geometric midrange problem in \mathbb{P}_2 for

$$Y_1 = \begin{pmatrix} 0.95 & -0.6 \\ -0.6 & 1.1 \end{pmatrix}, \quad Y_2 = \begin{pmatrix} 1.0 & 0.5 \\ 0.5 & 2.1 \end{pmatrix}, \quad Y_3 = \begin{pmatrix} 2.5 & -0.2 \\ -0.2 & 1.2 \end{pmatrix}.$$

The lower bound l is computed to be $\frac{1}{2} \text{diam}_\infty(\{Y_1, Y_2, Y_3\}) = 0.7880$. On the other hand, solving the quasiconvex optimization problem (3.5) via the bisection method yields the midrange

$$X^* = \begin{pmatrix} 1.3154 & -0.5321 \\ -0.5321 & 1.6217 \end{pmatrix}$$

with minimum cost $t^* = 0.7901 > 0.7880 = l$. Indeed, we have $t^* = d_\infty(X^*, Y_1) = d_\infty(X^*, Y_2) = d_\infty(X^*, Y_3)$. \square

DEFINITION 3.5 (active matrices). *Let $N \geq 2$, and let (X^*, t^*) be a solution to (3.5). Then Y_j is called active if $d_\infty(X^*, Y_j) = t^*$. In particular, at least one of the following must hold:*

$$-\log(\lambda_{\min}(Y_j^{-\frac{1}{2}} X^* Y_j^{-\frac{1}{2}})) = t^* \quad \text{or} \quad \log(\lambda_{\max}(Y_j^{-\frac{1}{2}} X^* Y_j^{-\frac{1}{2}})) = t^*.$$

Proposition 3.4 suggests that the N -point matrix midrange problem is richer than the vector case in fundamental ways. While the bisection method applied to the quasiconvex problem (3.5) offers a solution, it can be quite slow due to the need for multiple bisection steps and the requirement to compute a reasonable upper bound estimate of the optimum cost for initialization. However, it is possible to recast (3.5) as a convex optimization problem and thereby obtain a dramatic improvement in efficiency by introducing new variables. Specifically, by introducing $\xi = e^t$ and $\tau = e^{-t}$, and adding the extra convex constraint that $1/\xi - \tau \leq 0$, we find that (3.5) can be reformulated as the convex optimization problem

$$(3.6) \quad \begin{cases} \min_{X \succeq 0, \xi \in \mathbb{R}, \tau \in \mathbb{R}} & \xi \\ \tau Y_i \preceq X \preceq \xi Y_i, \\ 1/\xi - \tau \leq 0, \end{cases}$$

which can generally be solved much more efficiently than (3.5) using standard convex optimization techniques and software packages.

Example 3.6. As an example, we use (3.6) to compute the geometric midrange of $N = 1000$ real symmetric positive definite 2×2 matrices. The data was generated as $Y_j = \Sigma + A_j^T A_j$ for $j = 1, \dots, N$, where the A_j are 2×2 matrices with normally distributed entries and $\Sigma \succ 0$ is a fixed matrix. The data matrices can be represented as points in a cone in \mathbb{R}^3 via the bijection

$$\begin{pmatrix} a & b \\ b & c \end{pmatrix} \mapsto \left(\sqrt{2}b, \frac{1}{\sqrt{2}}(a-c), \frac{1}{\sqrt{2}}(a+c) \right)$$

as described in [39]. Figure 3 shows a visualization of the results of the computation in \mathbb{R}^3 from two perspectives. The surrounding open cone represents the boundary of the set of real symmetric positive definite matrices, and the cloud of points in gray are the data points Y_i . At optimum, we find that there are four active points (highlighted in red), two of which nearly coincide in the figure. The geometric midrange is highlighted in blue and is the center of the Thompson sphere of radius t^* that defines the smallest enclosing Thompson ball of the data. Note how the active points lie on

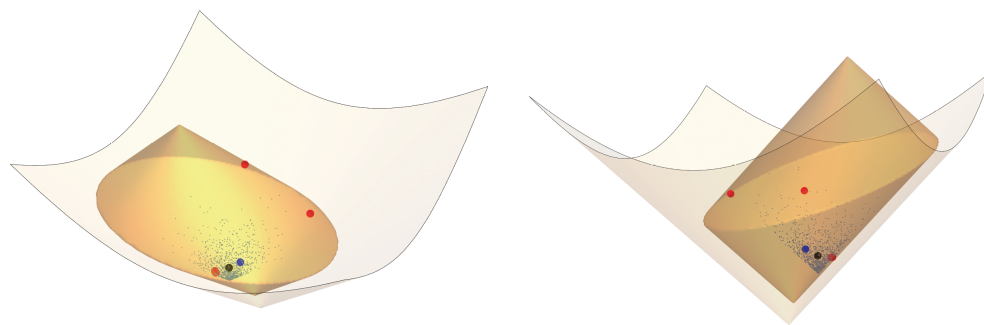


FIG. 3. The geometric midrange (blue point), Karcher mean (black point), active data matrices (red points), and smallest enclosing Thompson ball for an example with $N = 1000$ matrices depicted in \mathbb{R}^3 from two perspectives.

this sphere. It is interesting that the Thompson ball is the intersection of two cones in this representation. For the sake of comparison, the geometric mean of the data is also included as a solid black point.

Remark 3.7. The preceding analysis provides an interesting example of a non-convex optimization problem that admits a reformulation as a convex optimization problem in the Euclidean sense through a nonlinear change of coordinates.

While the convex formulation (3.6) offers a dramatic improvement to the bisection algorithm applied to the quasiconvex formulation of the problem, we expect that yet more efficient solutions to the problem can be found. In particular, conventional SDP solvers are based on interior point methods with fast convergence but high cost per iteration [42, 51], which makes them less suitable for matrices of larger size. Alternatively, one may consider so-called proximal splitting methods such as alternating projections, Douglas–Rachford, or the alternating direction method of multipliers (ADMM) [12, 16, 17, 19] applied to (3.5), which have cheap cost per iteration. Unfortunately, these methods tend to have poor convergence properties when the optimal solution is an intersection point of the boundaries of two convex sets with a small intersection angle [20, 21]. Indeed, our numerical experiments indicate that the rates of convergence of such methods degrade as t gets close to the true minimum. This can be expected as each Y_i in (3.5) defines a bounding box for X through the inequality constraint and X cannot be an interior point to all of them. Ideally, an efficient algorithm for solving this problem would principally rely on the computation of dominant generalized eigenpairs as in the $N = 2$ case for which very efficient algorithms exist. In the next subsection, we will consider the optimality conditions for the geometric midrange problem in more detail. Before doing so, we note the following special case for which the N -point midrange problem reduces to the 2-point problem as in the scalar case.

PROPOSITION 3.8. *If Y_1, \dots, Y_N are such that $Y_1 \preceq Y_i \preceq Y_N$ for all $i = 1, \dots, N$, then the geometric midrange of $\{Y_i\}$ is given by the set of d_∞ -midpoints of Y_1 and Y_N .*

Proof. The ordering $Y_1 \preceq Y_i \preceq Y_N$ means that the intersection of the feasibility constraints $e^{-t}Y_i \preceq X \preceq e^tY_i$ in the epigraph formulation (3.5) is simply

$$e^{-t}Y_N \preceq X \preceq e^tY_1.$$

Thus, the optimization problem is unchanged following the elimination of all Y_i for $i \neq 1, N$. Hence, the problem is equivalent to the midrange problem for $\{Y_1, Y_N\}$ and is solved by any d_∞ -midpoint of this pair. Furthermore, the lower bound $l = \frac{1}{2} \text{diam}_\infty(\{Y_i\}) = \frac{1}{2} d_\infty(Y_1, Y_N)$ is trivially attained. \square

Remark 3.9. Note that in the above we do not assume an order relation between Y_i and Y_j for $i, j \neq 1, N$. The value of this result lies in the insight that it provides in how and why the matrix N -point midrange problem diverges from the scalar and vector cases. Fundamentally, no order relation need exist between a pair of matrices, whereas in the scalar case such an ordering is always possible, and similarly an unambiguous ordering is possible at the level of coordinates for vectors.

3.2. Necessary optimality conditions. Finally, we prove a number of results on the optimality conditions of the geometric matrix midrange problem and the connection between the attainment of the lower bound $l = \frac{1}{2} \text{diam}_\infty(\{Y_i\})$ and the number of active matrices at optimum.

PROPOSITION 3.10. *Let $N \geq 2$, and let (X^*, t^*) be a solution to (3.5). Then there exist distinct $i^*, j^* \in \{1, \dots, N\}$ such that*

$$\log(\lambda_{\max}(Y_{i^*}^{-\frac{1}{2}} X^* Y_{i^*}^{-\frac{1}{2}})) = -\log(\lambda_{\min}(Y_{j^*}^{-\frac{1}{2}} X^* Y_{j^*}^{-\frac{1}{2}})) = t^*.$$

Proof. By the definition of (X^*, t^*) , there exists at least one index i^* or j^* such that $|\log(\lambda_{\max}(Y_{i^*}^{-\frac{1}{2}} X^* Y_{i^*}^{-\frac{1}{2}}))| = t^*$ or $|\log(\lambda_{\min}(Y_{j^*}^{-\frac{1}{2}} X^* Y_{j^*}^{-\frac{1}{2}}))| = t^*$. In particular, for such i^* and j^* it must hold that

$$\log(\lambda_{\max}(Y_{i^*}^{-\frac{1}{2}} X^* Y_{i^*}^{-\frac{1}{2}})), \quad -\log(\lambda_{\min}(Y_{j^*}^{-\frac{1}{2}} X^* Y_{j^*}^{-\frac{1}{2}})) \geq 0.$$

Next we will show that if there exists only one i^* or j^* , then (X^*, t^*) would not be a solution. To this end, assume that no index such as j^* exists, so that

$$\begin{aligned} |\log(\lambda_{\min}(Y_j^{-\frac{1}{2}} X^* Y_j^{-\frac{1}{2}}))| &< \log(\lambda_{\max}(Y_{i^*}^{-\frac{1}{2}} X^* Y_{i^*}^{-\frac{1}{2}})) = t^* \text{ for all } j, \\ \log(\lambda_{\max}(Y_i^{-\frac{1}{2}} X^* Y_i^{-\frac{1}{2}})) &\leq \log(\lambda_{\max}(Y_{i^*}^{-\frac{1}{2}} X^* Y_{i^*}^{-\frac{1}{2}})) = t^* \text{ for all } i. \end{aligned}$$

Then for sufficiently large $0 < k < 1$ and $\tilde{X}^* := kX^*$, it holds that

$$\begin{aligned} |\log(\lambda_{\min}(Y_j^{-\frac{1}{2}} \tilde{X}^* Y_j^{-\frac{1}{2}}))| &< \log(\lambda_{\max}(Y_{i^*}^{-\frac{1}{2}} \tilde{X}^* Y_{i^*}^{-\frac{1}{2}})) < t^* \text{ for all } j, \\ \log(\lambda_{\max}(Y_i^{-\frac{1}{2}} \tilde{X}^* Y_i^{-\frac{1}{2}})) &\leq \log(\lambda_{\max}(Y_{i^*}^{-\frac{1}{2}} \tilde{X}^* Y_{i^*}^{-\frac{1}{2}})) < t^* \text{ for all } i, \end{aligned}$$

which would mean that \tilde{X}^* is a feasible solution of smaller cost than t^* . Analogously, it follows that there always exists an index i^* with the required property. \square

PROPOSITION 3.11. *Recall the $N = 2$ geometric midrange problem for Y_1 and Y_2 in \mathbb{P}_n . Set $\alpha := \lambda_{\max}(Y_1^{-\frac{1}{2}} Y_2 Y_1^{-\frac{1}{2}})$ and $\beta := \lambda_{\min}(Y_1^{-\frac{1}{2}} Y_2 Y_1^{-\frac{1}{2}})$. If $\alpha \neq \beta$, then*

$$(3.7) \quad X^* = \frac{\sqrt{\alpha\beta}Y_1 + Y_2}{\sqrt{\alpha} + \sqrt{\beta}} = Y_1 * Y_2$$

is the only midrange of $\{Y_1, Y_2\}$ in $\{k_1 Y_1 + k_2 Y_2 : k_1, k_2 \geq 0\}$ for which the following is satisfied:

$$(3.8) \quad \begin{cases} \log(\lambda_{\max}(Y_1^{-\frac{1}{2}} X^* Y_1^{-\frac{1}{2}})) = -\log(\lambda_{\min}(Y_2^{-\frac{1}{2}} X^* Y_2^{-\frac{1}{2}})), \\ \log(\lambda_{\max}(Y_2^{-\frac{1}{2}} X^* Y_2^{-\frac{1}{2}})) = -\log(\lambda_{\min}(Y_1^{-\frac{1}{2}} X^* Y_1^{-\frac{1}{2}})). \end{cases}$$

The optimal cost to (3.5) is given by $t^* = \frac{1}{2} \max\{|\log(\alpha)|, |\log(\beta)|\}$. Furthermore, if $Y_2V = Y_1VD$ is a generalized eigenvalue decomposition such that $V^*Y_1V = I$ and D is diagonal, then V^*X^*V is diagonal.

Proof. Using the linear ansatz $X^* = k_1Y_1 + k_2Y_2$, we obtain

$$\begin{aligned}\lambda_{\max}(Y_1^{-\frac{1}{2}}X^*Y_1^{-\frac{1}{2}}) &= k_1 + k_2\lambda_{\max}(Y_1^{-\frac{1}{2}}Y_2Y_1^{-\frac{1}{2}}) = k_1 + k_2\alpha, \\ \lambda_{\max}(Y_2^{-\frac{1}{2}}X^*Y_2^{-\frac{1}{2}}) &= k_1\lambda_{\max}(Y_2^{-\frac{1}{2}}Y_1Y_2^{-\frac{1}{2}}) + k_2 = \frac{k_1}{\beta} + k_2, \\ \lambda_{\min}(Y_1^{-\frac{1}{2}}X^*Y_1^{-\frac{1}{2}}) &= k_1 + k_2\lambda_{\min}(Y_1^{-\frac{1}{2}}Y_2Y_1^{-\frac{1}{2}}) = k_1 + k_2\beta, \\ \lambda_{\min}(Y_2^{-\frac{1}{2}}X^*Y_2^{-\frac{1}{2}}) &= k_1\lambda_{\min}(Y_2^{-\frac{1}{2}}Y_1Y_2^{-\frac{1}{2}}) + k_2 = \frac{k_1}{\alpha} + k_2.\end{aligned}$$

Substituting these expressions into (3.8), we find that

$$(3.9) \quad k_1 + k_2\alpha = \frac{1}{\frac{k_1}{\alpha} + k_2} \quad \text{and} \quad \frac{k_1}{\beta} + k_2 = \frac{1}{k_1 + k_2\beta}.$$

Hence

$$(k_1 + k_2\alpha) \left(\frac{k_1}{\alpha} + k_2 \right) = \left(\frac{k_1}{\beta} + k_2 \right) (k_1 + k_2\beta),$$

which is equivalent to $(k_2^2\alpha\beta - k_1^2)(\alpha - \beta) = 0$ and implies that $k_1^2 = k_2^2\alpha\beta$ since $\alpha \neq \beta$. Substituting $k_1 = \sqrt{\alpha\beta}k_2$ into the first equation of (3.9) gives $k_2^2\beta + 2k_2^2\sqrt{\alpha\beta} + k_2^2\alpha = k_2^2(\sqrt{\alpha} + \sqrt{\beta})^2 = 1$ and thus (3.7). That the cost is given by t^* is trivial and optimality of X^* follows by the attainment of the lower bound. Finally, V^*X^*V is diagonal by Proposition 3.10. \square

Remark 3.12. Note that if $\alpha = \beta$ in the statement of the previous proposition, then $Y_1^{-1/2}Y_2Y_1^{-1/2} = \alpha I$, which is equivalent to $Y_2 = \alpha Y_1$. The midrange $Y_1 \star Y_2 = \sqrt{\alpha}Y_1$ can then be obtained as a conic combination of Y_1 and $Y_2 = \alpha Y_1$ in a nonunique way.

In the remainder of this section, we explore the significance of the number of active points at optimum for the attainment of the lower bound of (3.5) when $N \geq 2$.

LEMMA 3.13. Let $N \geq 2$, and let (X^*, t^*) be a solution to (3.5). Then the following are equivalent:

1. Y_j is active with $\log(\lambda_{\max}(Y_j^{-\frac{1}{2}}X^*Y_j^{-\frac{1}{2}})) = t^*$.
2. $\lambda_{\min}(e^{t^*}Y_j - X^*) = 0$.
3. $\nexists \varepsilon > 0 : X^* \preceq e^{t^*}Y_j - \varepsilon I$.

Analogously, we have the following equivalences:

1. Y_j is active with $-\log(\lambda_{\min}(Y_j^{-\frac{1}{2}}X^*Y_j^{-\frac{1}{2}})) = t^*$.
2. $\lambda_{\max}(X^* - e^{-t^*}Y_j) = 0$.
3. $\nexists \varepsilon > 0 : X^* \succeq e^{-t^*}Y_j + \varepsilon I$.

LEMMA 3.14. Let $D = \text{diag}(d_1, \dots, d_n)$ with $d_1 \leq \dots \leq d_n$ and $D \preceq X \preceq d_n I$. Then

$$X = \begin{pmatrix} X_{11} & 0 \\ 0 & d_n \end{pmatrix},$$

and thus $d_n = \lambda_{\max}(X) \geq \lambda_{\max}(X_{11})$.

Proof. From the inequality it follows that $X_{nn} = d_n$ and $\lambda_{\max}(X) \leq d_n$. Thus, by Courant–Fischer, $e_n = (0, 0, \dots, 1)$ is an eigenvector of X with eigenvalue d_n , and thus X has the required form. \square

PROPOSITION 3.15. *Let (X^*, t^*) be a solution to (3.5) and*

$$t^* = \frac{1}{2} \log \left(\lambda_{\max} \left(Y_1^{-\frac{1}{2}} Y_2 Y_1^{-\frac{1}{2}} \right) \right).$$

Then Y_1 and Y_2 are active with

$$\log \left(\lambda_{\max} \left(Y_1^{-\frac{1}{2}} X^* Y_1^{-\frac{1}{2}} \right) \right) = t^* = -\log \left(\lambda_{\min} \left(Y_2^{-\frac{1}{2}} X^* Y_2^{-\frac{1}{2}} \right) \right).$$

Further, if $Y_2 V = Y_1 V D$ is a generalized eigenvalue decomposition such that $V^ Y_1 V = I$ and $D = \text{diag}(d_1, \dots, d_n)$ with $d_1 \leq \dots \leq d_n = \lambda_{\max} \left(Y_1^{-\frac{1}{2}} Y_2 Y_1^{-\frac{1}{2}} \right)$, then*

$$V^* X^* V = \begin{pmatrix} X_{11} & 0 \\ 0 & \sqrt{d_n} \end{pmatrix}.$$

Proof. Let $X_V := V^* X^* V$. We first show that X_V has the claimed structure. To this end, note that by (3.5)

$$e^{-t^*} I \preceq X_V \preceq e^{t^*} I \quad \text{and} \quad e^{-t^*} D \preceq X_V \preceq e^{t^*} D,$$

which implies that $e^{-t^*} D \preceq X_V \preceq e^{t^*} I$ with $e^{t^*} = \sqrt{d_n}$. Therefore, Lemma 3.14 implies the required structure for X_V . Then by Lemma 3.13 it follows that $e^{t^*} I$ and $e^{-t^*} D$ are active matrices for (X_V, t^*) , and thus Y_1 and Y_2 are active matrices for (X^*, t^*) . Then by Lemma 3.14 we can conclude the remaining claim. \square

PROPOSITION 3.16. *If there are only two active matrices at an optimum (X^*, t^*) of (3.5), then the lower bound $l = \frac{1}{2} \text{diam}_{\infty}(\{Y_i\})$ is attained.*

Proof. Suppose that Y_1 and Y_2 are the only two active matrices at (X^*, t^*) and assume that the lower bound l is not attained so that $l < t^*$. Denote the geodesic (2.8) from X^* to the d_{∞} -midpoint $Y_1 * Y_2$ of Y_1 and Y_2 by $\gamma_{\mathcal{G}}(s) = \gamma_{\mathcal{G}}(s, X^*, Y_1 * Y_2)$. By the geodesic convexity of $(\mathbb{P}_n, d_{\infty})$, the function $s \mapsto d_{\infty}(Y_j, \gamma_{\mathcal{G}}(s, X^*, Y_1 * Y_2))$ is convex for $j = 1, 2$. Thus, we have

$$\begin{aligned} d_{\infty}(Y_j, \gamma_{\mathcal{G}}(s, X^*, Y_1 * Y_2)) &\leq (1-s)d_{\infty}(Y_j, X^*) + s d_{\infty}(Y_j, Y_1 * Y_2) \\ &= (1-s)t^* + \frac{s}{2} d_{\infty}(Y_1, Y_2) \\ &\leq (1-s)t^* + sl \\ &< (1-s)t^* + st^* = t^* \end{aligned}$$

for any $s > 0$ and $j = 1, 2$. As all matrices other than Y_1 and Y_2 are inactive at $s = 0$, we can achieve a local reduction in the cost function by moving a sufficiently small $s > 0$ along the geodesic from X^* to $Y_1 * Y_2$, which would contradict the optimality of (X^*, t^*) . Thus, we have $t^* = l$. \square

Remark 3.17. Note that although Proposition 3.16 implies that the optimum X^* will lie in the d_{∞} -midpoint set of the active pair of matrices, it does not imply that any d_{∞} -midpoint of Y_1, Y_2 will be a solution. In particular, $Y_1 * Y_2$ may not be a solution even if the only active matrices at optimum are Y_1 and Y_2 since $Y_1 * Y_2$ may fail to satisfy one or more of the constraints in (3.5). This is in contrast to the scenario in Proposition 3.8, where any midrange of Y_1 and Y_N will be a solution.

4. Conclusion. We have introduced a theory of geometric midrange statistics for positive definite Hermitian matrices within an optimization framework. We have also established a number of key results including bounds on the optimization problem as well as necessary conditions for optimality. Furthermore, a solution to the N -point problem is offered via convex optimization. Special consideration has been given to the 2-point midrange problem, which was studied in detail from a number of complementary perspectives. The existence of solutions to the 2-point problem that can be computed using only extremal generalized eigenvalues has significant implications for computational scalability of matrix midrange statistics. We expect this work to offer a solid foundation for future research in statistics based on Thompson geometry and related topics such as K -midranges [15, 46] for matrix-valued data. The development of a fast algorithm for the computation of a midrange of N matrices would be an important step in this direction, with weighted inductive schemes and stochastic algorithms offering a promising angle of attack.

Acknowledgment. We are most grateful to Yurii Nesterov for suggesting the change of variables that facilitated the conversion of the quasiconvex formulation of the N -point geometric matrix midrange problem to a convex optimization problem.

REFERENCES

- [1] B. AFSARI, *Riemannian L^p center of mass: Existence, uniqueness, and convexity*, Proc. Amer. Math. Soc., 139 (2011), pp. 655–673.
- [2] T. ANDO, C.-K. LI, AND R. MATHIAS, *Geometric means*, Linear Algebra Appl., 385 (2004), pp. 305–334, <https://doi.org/10.1016/j.laa.2003.11.019>.
- [3] M. ARNAUDON, F. BARBARESCO, AND L. YANG, *Riemannian medians and means with applications to radar signal processing*, IEEE J. Selected Topics Signal Process., 7 (2013), pp. 595–604, <https://doi.org/10.1109/JSTSP.2013.2261798>.
- [4] M. ARNAUDON AND F. NIELSEN, *On approximating the Riemannian 1-center*, Comput. Geom., 46 (2013), pp. 93–104, <https://doi.org/10.1016/j.comgeo.2012.04.007>.
- [5] V. ARSIGNY, P. FILLARD, X. PENNEC, AND N. AYACHE, *Log-Euclidean metrics for fast and simple calculus on diffusion tensors*, Magnetic Resonance Med., 56 (2006), pp. 411–421, <https://doi.org/10.1002/mrm.20965>.
- [6] M. BADOIU AND K. L. CLARKSON, *Optimal core-sets for balls*, Comput. Geom., 40 (2008), pp. 14–22, <https://doi.org/10.1016/j.comgeo.2007.04.002>.
- [7] R. BHATIA, *On the exponential metric increasing property*, Linear Algebra Appl., 375 (2003), pp. 211–220, [https://doi.org/10.1016/S0024-3795\(03\)00647-5](https://doi.org/10.1016/S0024-3795(03)00647-5).
- [8] R. BHATIA, *Positive Definite Matrices*, Princeton University Press, Princeton, NJ, 2007.
- [9] R. BHATIA AND J. HOLBROOK, *Riemannian geometry and matrix geometric means*, Linear Algebra Appl., 413 (2006), pp. 594–618, <https://doi.org/10.1016/j.laa.2005.08.025>.
- [10] D. BINI, B. MEINI, AND F. POLONI, *An effective matrix geometric mean satisfying the Ando-Li-Mathias properties*, Math. Comp., 79 (2010), pp. 437–452.
- [11] D. A. BINI AND B. IANNAZZO, *Computing the Karcher mean of symmetric positive definite matrices*, Linear Algebra Appl., 438 (2013), pp. 1700–1710, <https://doi.org/10.1016/j.laa.2011.08.052>.
- [12] S. BOYD, N. PARIKH, E. CHU, B. PELEATO, AND J. ECKSTEIN, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Mach. Learn., 3 (2011), pp. 1–122.
- [13] S. BOYD AND L. VANDENBERGHE, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004, <https://doi.org/10.1017/CBO9780511804441>.
- [14] M. R. BRIDSON AND A. HAEFLIGER, *Metric Spaces of Non-positive Curvature*, Grundlehren Math. Wiss. 319, Springer Science & Business Media, New York, 1999.
- [15] J. D. CARROLL AND A. CHATURVEDI, *K-midranges clustering*, in Advances in Data Science and Classification, A. Rizzi, M. Vichi, and H.-H. Bock, eds., Springer, Berlin, Heidelberg, 1998, pp. 3–14.
- [16] P. L. COMBETTES AND J.-C. PESQUET, *Proximal Splitting Methods in Signal Processing*, Springer, New York, 2011, pp. 185–212.
- [17] J. DOUGLAS AND H. H. RACHFORD, *On the numerical solution of heat conduction problems in*

- two and three space variables, Trans. Amer. Math. Soc., 82 (1956), pp. 421–439.
- [18] I. L. DRYDEN, A. KOLOYDENKO, AND D. ZHOU, *Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging*, Ann. Appl. Statist., 3 (2009), pp. 1102–1123, <http://www.jstor.org/stable/30242879>.
 - [19] J. ECKSTEIN AND D. P. BERTSEKAS, *On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Programming, 55 (1992), pp. 293–318.
 - [20] M. FÄLT AND P. GISELSSON, *Line search for generalized alternating projections*, in 2017 American Control Conference (ACC), IEEE, 2017, pp. 4637–4642, <https://doi.org/10.23919/ACC.2017.7963671>.
 - [21] M. FÄLT AND P. GISELSSON, *Optimal convergence rates for generalized alternating projections*, in 2017 IEEE 56th Annual Conference on Decision and Control (CDC), 2017, pp. 2268–2274, <https://doi.org/10.1109/CDC.2017.8263980>.
 - [22] R. GE, C. JIN, S. KAKADE, P. NETRAPALLI, AND A. SIDFORD, *Efficient algorithms for large-scale generalized eigenvector computation and canonical correlation analysis*, in International Conference on Machine Learning, International Machine Learning Society, 2016, pp. 2741–2750.
 - [23] G. H. GOLUB AND H. A. VAN DER VORST, *Eigenvalue computation in the 20th century*, J. Comput. Appl. Math., 123 (2000), pp. 35–65, [https://doi.org/10.1016/S0377-0427\(00\)00413-1](https://doi.org/10.1016/S0377-0427(00)00413-1).
 - [24] M. GRANT AND S. BOYD, *CVX: Matlab Software for Disciplined Convex Programming*, Version 2.1, <http://cvxr.com/cvx>, 2014.
 - [25] E. GUMBEL, *Statistics of Extremes*, Columbia University Press, New York, 1967.
 - [26] B. IANNAZZO, *The geometric mean of two matrices from a computational viewpoint*, Numer. Linear Algebra Appl., 23 (2011), pp. 208–229, <https://doi.org/10.1002/nla.2022>.
 - [27] B. JEURIS, R. VANDEBRIL, AND B. VANDEREYCKEN, *A survey and comparison of contemporary algorithms for computing the matrix geometric mean*, Electron. Trans. Numer. Anal., 39 (2012), pp. 379–402.
 - [28] F. KUBO AND T. ANDO, *Means of positive linear operators*, Math. Ann., 246 (1980), pp. 205–224.
 - [29] S. LANG, *Fundamentals of Differential Geometry*, Springer-Verlag, New York, 1999.
 - [30] J. LAWSON AND Y. LIM, *Metric convexity of symmetric cones*, Osaka J. Math., 44 (2007), pp. 795–816, <https://projecteuclid.org:443/euclid.ojm/1199719405>.
 - [31] R. B. LEHOUCQ, D. C. SORENSEN, AND C. YANG, *ARPACK Users' Guide: Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*, Software Environ. Tools 6, SIAM, Philadelphia, 1998, <https://doi.org/10.1137/1.9780898719628>.
 - [32] B. LEMMENS AND R. NUSSBAUM, *Nonlinear Perron-Frobenius Theory*, Cambridge Tracts in Math. 189, Cambridge University Press, Cambridge, UK, 2012, <https://doi.org/10.1017/CBO9781139026079>.
 - [33] Y. LIM, *Geometry of midpoint sets for Thompson's metric*, Linear Algebra Appl., 439 (2013), pp. 211–227, <https://doi.org/10.1016/j.laa.2013.03.012>.
 - [34] Y. LIM AND M. PALFIA, *Weighted inductive means*, Linear Algebra Appl., 453 (2014), pp. 59–83, <https://doi.org/10.1016/j.laa.2014.04.002>.
 - [35] E. MASSART, J. M. HENDRICKX, AND P.-A. ABSIL, *Matrix geometric means based on shuffled inductive sequences*, Linear Algebra Appl., 542 (2018), pp. 334–359, <https://doi.org/10.1016/j.laa.2017.05.036>.
 - [36] B. MISHRA AND R. SEPULCHRE, *Riemannian preconditioning*, SIAM J. Optim., 26 (2016), pp. 635–660, <https://doi.org/10.1137/140970860>.
 - [37] M. MOAKHER, *A differential geometric approach to the geometric mean of symmetric positive-definite matrices*, SIAM J. Matrix Anal. Appl., 26 (2005), pp. 735–747, <https://doi.org/10.1137/S0895479803436937>.
 - [38] C. MOSTAJERAN, C. GRUSSLER, AND R. SEPULCHRE, *Affine-invariant midrange statistics*, in Geometric Science of Information, F. Nielsen and F. Barbaresco, eds., Springer, Cham, 2019, pp. 494–501.
 - [39] C. MOSTAJERAN AND R. SEPULCHRE, *Ordering positive definite matrices*, Inf. Geom., 1 (2018), pp. 287–313, <https://doi.org/10.1007/s41884-018-0003-7>.
 - [40] F. NIELSEN AND R. NOCK, *Approximating smallest enclosing balls with applications to machine learning*, Internat. J. Comput. Geom. Appl., 19 (2009), pp. 389–414.
 - [41] 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>.
 - [42] D. PEAUCELLE, D. HENRION, Y. LABIT, AND K. TAITZ, *User's Guide for SEDUMI INTERFACE 1.04*, LAAS-CNRS, Toulouse, France, 2002.

- [43] X. PENNEC, P. FILLARD, AND N. AYACHE, *A Riemannian framework for tensor computing*, Internat. J. Comput. Vis., 66 (2006), pp. 41–66, <https://doi.org/10.1007/s11263-005-3222-z>.
- [44] R. P. N. RAO, *Brain-Computer Interfacing: An Introduction*, Cambridge University Press, Cambridge, UK, 2013, <https://doi.org/10.1017/CBO9781139032803>.
- [45] S. SRA, *A new metric on the manifold of kernel matrices with application to matrix geometric means*, in Advances in Neural Information Processing Systems, Curran Associates, Red Hook, NY, 2012, pp. 144–152.
- [46] D. STEINLEY, *K-means clustering: A half-century synthesis*, British J. Math. Statist. Psych., 59 (2006), pp. 1–34, <https://doi.org/10.1348/000711005X48266>.
- [47] G. W. STEWART, *A Krylov–Schur algorithm for large eigenproblems*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 601–614, <https://doi.org/10.1137/S0895479800371529>.
- [48] S. M. STIGLER, *The Seven Pillars of Statistical Wisdom*, Harvard University Press, Cambridge, MA, 2016.
- [49] J. J. SYLVESTER, *A question in the geometry of situation*, Quart. J. Pure Appl. Math., 1 (1857), pp. 79–80.
- [50] A. C. THOMPSON, *On certain contraction mappings in a partially ordered vector space*, Proc. Amer. Math. Soc., 14 (1963), pp. 438–443, <http://www.jstor.org/stable/2033816>.
- [51] K. C. TOH, R. H. TUTUNCU, AND M. J. TODD, *On the implementation of SDPT3 (version 3.1): A MATLAB software package for semidefinite-quadratic-linear programming*, in IEEE International Conference on Robotics and Automation, 2004, pp. 290–296.
- [52] I. W. TSANG, A. KOCSOR, AND J. T. KWOK, *Simpler core vector machines with enclosing balls*, in Proceedings of the 24th International Conference on Machine Learning, ICML 2007, ACM, New York, 2007, pp. 911–918, <https://doi.org/10.1145/1273496.1273611>.
- [53] E. WELZL, *Smallest enclosing disks (balls and ellipsoids)*, in New Results and New Trends in Computer Science, Springer, Berlin, 1991, pp. 359–370.
- [54] P. ZANINI, M. CONGEDO, C. JUTTEN, S. SAID, AND Y. BERTHOUMIEU, *Transfer learning: A Riemannian geometry framework with applications to brain-computer interfaces*, IEEE Trans. Biomed. Engrg., 65 (2018), pp. 1107–1116.
- [55] H. ZHU, H. ZHANG, J. G. IBRAHIM, AND B. S. PETERSON, *Statistical analysis of diffusion tensors in diffusion-weighted magnetic resonance imaging data*, J. Amer. Statist. Assoc., 102 (2007), pp. 1085–1102, <https://doi.org/10.1198/016214507000000581>.