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 \( N > 2 \) 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.

Key words. Positive definite matrices, Statistics, Optimization, Matrix means, Midranges, Thompson metric, Minimal geodesic, Affine-invariance

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

1. Introduction. The midrange of a collection of real numbers \( a_1, \ldots, a_N \) 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 non-robust 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 \([37, 12, 38]\). It is an important notion in the statistics of extreme events \([21]\).

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 \([15, 41]\). It has been noted in many works that using nonlinear geometries related to generalized spectral properties of positive definite matrices yield significantly improved performance \([2, 35]\). It is in this context that much attention has been paid to developing geometric statistical methods on the cone of positive definite matrices \([24, 1, 31, 5, 4, 36]\). A fundamental geometry that is associated to such spaces is the affine-invariant geometry \([4, 32]\), 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 \( \mathbb{R}^+ = \{x \in \mathbb{R} : x > 0\} \)

\(^*\)Submitted to the editors DATE.

Funding: This work benefited from funding by the European Research Council under the Advanced ERC Grant Agreement Switchlet n.670645. C. Mostajeran is supported by the Cambridge Philosophical Society.

†Department of Engineering, University of Cambridge, United Kingdom (csm54@cam.ac.uk).

\(^1\)This manuscript substantially expands upon ideas partially introduced by the authors in a paper accepted by the 4th Conference on Geometric Science of Information (GSI2019).
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

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

Note that (1.3) is the unique solution of the optimization problem

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

In the matrix setting, the natural geometric midrange problem on the cone of positive definite matrices takes the form

$$\min_{X \succeq 0} \max_i \| \log(Y^{-1/2}_i X Y^{-1/2}_i) \|_\infty,$$

where $\{Y_1, \ldots, Y_N\}$ are a collection of $N$ positive definite matrices of dimension $n$ and $\| \cdot \|_\infty$ denotes the spectral operator norm on the space of Hermitian matrices of dimension $n$ defined by $\|A\|_\infty = \max\{|\lambda_1(A)|, \ldots, |\lambda_n(A)|\}$. We will return to this formulation and justify its use when we consider the geometry of positive definite matrices in more detail. For now, note the similarity of (1.5) to the scalar version (1.3) of the affine-invariant midrange problem. An important point to be made about (1.5) is that it does not generally admit a unique solution. Indeed, even in the case of two matrices $Y_1 = A$ and $Y_2 = B$, the solution $X$ to the optimization problem (1.5) is generally not unique in contrast to the scalar problem. One particular analytic solution for the geometric midrange of two positive definite matrices $A$ and $B$ that will receive special consideration is $A \ast B$ defined by

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

where $\lambda_{\max}$ and $\lambda_{\min}$ 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 $\lambda_{\max}$ and $\lambda_{\min}$ also coincide with the maximum and minimum eigenvalues of $BA^{-1}$, respectively. We will consider the properties of the expression in (1.6) in some detail in section 2. For now, it is instructive to compare $A \ast B$ with the well-known geometric mean $A \# B$ of positive definite matrices $A$ and $B$ given by

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

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 [23, 8, 22]. These include optimization based approaches [6] as well as inductive sequential constructions [1, 7, 29, 30].

It is noteworthy that the formula (1.6) for a matrix midrange of $A$ and $B$ is considerably less expensive to compute than the the geometric mean (1.7), particularly for high dimensional matrices. This is because $A \ast B$ only relies on the computation of extremal generalized eigenvalues that can be computed efficiently using a variety
of techniques [19]. In contrast, the Cholesky-Schur algorithm for computing the geometric mean (1.7) of two matrices has a complexity of $O(n^3)$ [22]. 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.

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 (1.6) 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 mean and median. In subsection 2.2, we present a characterization of the midrange formula (1.6) based on an extremal ordering property defined using the Löwner order. In subsection 2.3, we review the differential geometry of $\mathcal{P}(n)$ 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 quasiconvex 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 $\mathcal{P}(n)$ 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 [39, 27] on $C$ is defined to be $d_T(x, y) = \log \max \{M(x/y; C), M(y/x; C)\}$, 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 $A, B \in \mathcal{P}(n)$, we have $M(A/B) = \lambda_{\max}(AB^{-1})$, so that

$$d_T(A, B) = \log \max \{\lambda_{\max}(AB^{-1}), \lambda_{\max}(BA^{-1})\}.$$  

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 $\Sigma \in \mathcal{P}(n)$, we find that the 2-point midrange problem (1.5) for data $A$ and $B$ takes the form

$$\min_{X \succeq 0} \max \{d_T(A, X), d_T(B, X)\}.$$  

A point $X$ is said to be a Thompson midpoint of the pair $(A, B)$ if $d_T(A, X) = d_T(B, X) = \frac{1}{2}d_T(A, B)$. As $(\mathcal{P}(n), d_T)$ forms a metric space [27], the minimizers of (2.3) coincide with the Thompson midpoints of $(A, B)$, which are generally non-unique. The geometry of the set of Thompson midpoints of a given pair of points in $A, B \in \mathcal{P}(n)$ is studied in detail in [28], 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 $A \ast B$ given by (1.6) due to its scalable computational properties.

Note that (2.3) is equivalent to $\min_{X \succeq 0} f(X)$, 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
\( A \ast B \) 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
\( c_1M + c_2I \) has eigenvalues \( c_1\lambda_i(M) + c_2 \).

Specifically, we find that for \( X = A \ast B \), we have
\[
XA^{-1} = \frac{1}{\sqrt{\lambda_{\min} + \lambda_{\max}}} \left( BA^{-1} + \sqrt{\lambda_{\min}\lambda_{\max}}I \right),
\]
\[
XB^{-1} = \frac{1}{\sqrt{\lambda_{\min} + \lambda_{\max}}} \left( I + \sqrt{\lambda_{\min}\lambda_{\max}}AB^{-1} \right),
\]

where \( \lambda_{\max} \) and \( \lambda_{\min} \) 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 \( f(A \ast B) \) simplifies to \( \frac{1}{2}d_T(A, B) \) as required.

We now consider the merits of the midrange \( A \ast B \) as a measure of central tendency for \( \{A, B\} \). The following are a number of properties that are desirable for such a mapping \( \mu : \mathbb{P}(n) \times \mathbb{P}(n) \to \mathbb{P}(n) \):

1. Continuity: \( \mu \) 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.

Note that \( X^{*} \) in the third property denotes the conjugate transpose of \( X \). We will now prove that the map \( \mu(A, B) := A \ast B \) 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 \( \mu : \mathbb{P}(n) \times \mathbb{P}(n) \to \mathbb{P}(n) \) defined by \( \mu(A, B) = A \ast B \) satisfies properties 1-3.

*Proof.* 1. The continuity of \( \mu \) follows directly from the expression for \( A \ast B \) in (1.6), 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 \( \lambda_{\max}(BA^{-1}) = 1/\lambda_{\min}(AB^{-1}) \), so that
\[
(2.7) \quad B \ast A = \frac{1}{\sqrt{1/\lambda_{\min} + 1/\lambda_{\max}}} \left( A + \frac{1}{\sqrt{\lambda_{\min}\lambda_{\max}}}B \right)
\]
\[
(2.8) \quad = \frac{1}{\sqrt{\lambda_{\min}\lambda_{\max}}} \left( \frac{1}{\sqrt{\lambda_{\min}\lambda_{\max}}}B + A \right) = A \ast B.
\]
3. Affine-invariance follows immediately by noting that
\[
(2.9) \quad \lambda_i(CBC^{*}(CAC^{*})^{-1}) = \lambda_i(CBC^{*}(C^{*})^{-1}A^{-1}C^{-1}) = \lambda_i(BA^{-1}). \quad \square
\]

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 \text{ or } b \leq a, \text{ two Hermitian matrices } A \text{ and } B \text{ may fail to satisfy both } A \preceq B \text{ and } B \preceq A. \text{ 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 \text{ implies that } XAX^* \preceq XBX^* \text{. In particular, } A \preceq B \text{ if and only if } I \preceq A^{-1/2}BA^{-1/2}. \text{ Thus, by affine-invariance of } \mu, \text{ it suffices to prove point 5 in the case where } A = I \text{ since } A \leq \mu(A, B) \preceq B \text{ if and only if } I \preceq \mu(I, A^{-1/2}BA^{-1/2}) \preceq A^{-1/2}BA^{-1/2}. \text{ To establish the 4th property for } \mu(A, B) = A \ast B, \text{ we make use of Lemma 2.1. Let } \Sigma \in \mathbb{P}(n) \text{ be such that } I \preceq \Sigma \text{ and note that this is equivalent to } \lambda_i(\Sigma) \geq 1 \text{ for } i = 1, \ldots, n. \text{ Writing } \lambda_{\min} = \lambda_{\min}(\Sigma), \ \lambda_{\max} = \lambda_{\max}(\Sigma), \text{ and } \lambda_i(\Sigma) = 1 + \delta_i \text{ for } \delta_i \geq 0, \text{ we have by Lemma 2.1 that }$

\begin{align*}
(2.10) \quad \lambda_i(I \ast \Sigma) - 1 &= \lambda_i \left( \frac{1}{\sqrt{\lambda_{\min} + \lambda_{\max}}} \left( \Sigma + \sqrt{\lambda_{\min}\lambda_{\max}} I \right) \right) - 1 \\
(2.11) &= \lambda_i(\Sigma) + \sqrt{\lambda_{\min}\lambda_{\max}} - 1 \\
(2.12) &= \delta_i + \frac{\sqrt{\lambda_{\min}\lambda_{\max}} - \sqrt{\lambda_{\min}} - \sqrt{\lambda_{\max}}}{\sqrt{\lambda_{\min} + \lambda_{\max}}} \\
(2.13) &= \delta_i + \frac{(\sqrt{\lambda_{\min}} - 1)(\sqrt{\lambda_{\max}} - 1)}{\sqrt{\lambda_{\min} + \lambda_{\max}}} \geq 0,
\end{align*}

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

\begin{align*}
(2.14) \quad \lambda_i(\Sigma - I \ast \Sigma) &= \lambda_i \left( \left( \frac{\sqrt{\lambda_{\min} + \lambda_{\max}} - 1}{\sqrt{\lambda_{\min} + \lambda_{\max}}} \right) \Sigma - \frac{\sqrt{\lambda_{\min}\lambda_{\max}}}{\sqrt{\lambda_{\min} + \lambda_{\max}}} I \right) \\
(2.15) &= \left( \frac{\sqrt{\lambda_{\min} + \lambda_{\max}} - 1}{\sqrt{\lambda_{\min} + \lambda_{\max}}} \right) \lambda_i(\Sigma) - \frac{\sqrt{\lambda_{\min}\lambda_{\max}}}{\sqrt{\lambda_{\min} + \lambda_{\max}}} \\
(2.16) &= \frac{\lambda_{\min}(\sqrt{\lambda_{\min} + \lambda_{\max}} - 1) - \sqrt{\lambda_{\min}\lambda_{\max}}}{\sqrt{\lambda_{\min} + \lambda_{\max}}} + \left( \frac{\sqrt{\lambda_{\min} + \lambda_{\max}} - 1}{\sqrt{\lambda_{\min} + \lambda_{\max}}} \right) \epsilon_i \\
(2.17) &= \sqrt{\lambda_{\min}} \left( \sqrt{\lambda_{\min}} - 1 \right) + \left( \frac{\sqrt{\lambda_{\min} + \lambda_{\max}} - 1}{\sqrt{\lambda_{\min} + \lambda_{\max}}} \right) \epsilon_i \geq 0,
\end{align*}

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

\begin{equation}
(2.18) \quad I \preceq \Sigma \implies I \preceq I \ast \Sigma \preceq \Sigma,
\end{equation}

\text{for all } \Sigma \in \mathbb{P}(n). \text{ In particular, upon substituting } \Sigma = A^{-1/2}BA^{-1/2} \text{ in (2.18) and using the affine-invariance properties of both the Löwner order and the mean } \mu(A, B) = A \ast B, \text{ we establish the following important property.}$

**Proposition 2.3.** For \( A, B \in \mathbb{P}(n), \ A \preceq B \) implies that \( A \preceq A \ast B \preceq B. \**

We now consider the 5th and final desirable property of \( \mu : \mathbb{P}(n) \times \mathbb{P}(n) \to \mathbb{P}(n), \) which is monotonicity of \( \mu \) in its arguments. First recall that a map \( F : \mathbb{P}(n) \to \mathbb{P}(n) \) 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 \( \Sigma. \) That is, monotonicity is established by showing that

\begin{equation}
(2.19) \quad \Sigma_1 \preceq \Sigma_2 \implies I \ast \Sigma_1 \preceq I \ast \Sigma_2.
\end{equation}
However, it turns out that $F(\Sigma) := I \ast \Sigma$ is not monotone with respect to $\Sigma$ as we demonstrate below. Nonetheless, $F$ is seen to enjoy certain weaker monotonicity properties. Considering the eigenvalues of $I \ast \Sigma$, we find that

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

where $\lambda_{\min}$ and $\lambda_{\max}$ refer to the smallest and largest eigenvalues of $\Sigma$.

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

**Proof.** Considering the cases $i = 1$ and $i = n$, we find that (2.20) yields

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

both of which are seen to be monotone functions of $\Sigma$.

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

**Theorem 2.5.** The midrange $\mu(A, B) = A \ast B$ defined in (1.6) yields a Thompson metric midpoint of $A, B \in P(n)$ 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 $\Sigma \in P(n)$.

We also note that $A \ast B$ 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 $A, B \in P(n)$, we have

$$\lambda_i\left( (aA) \ast (bB) \right) = \frac{\lambda_i((bB)(aA)^{-1})}{\lambda_i(BA^{-1})} = \frac{b}{a} \lambda_i(bA^{-1}) \lambda_i(\Sigma).$$

**Proof.** The result follows upon substituting $\lambda_i((bB)(aA)^{-1}) = \frac{b}{a} \lambda_i(bA^{-1})$ into the formula (1.6).
several measures of aggregation of data to the space of positive definite matrices 

\[ \text{distance function that coincides with the Thompson metric } [39] \text{ on the cone complex matrices is said to be unitarily invariant if } \]

\[ \| (2.29) \]

choice of symmetric gauge \( \Phi \). With \( \Phi(\cdot) \), the family of affine-invariant metric distances \( d \) of two matrices. Indeed, it does not generally hold for means arising as \( d_\infty \)-midpoints either. For instance, [28] identifies \( d \) as another \( d_\infty \)-midpoint of \( A \) and \( B \). Clearly \( A \circ B \) does not scale geometrically in the sense of (2.24).

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 \( \Phi \) 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 \( d_\Phi \) on \( \mathbb{P}(n) \) defined as

\[ (2.26) \]

\[ d_\Phi(A, B) = \| \log A^{-1/2}BA^{-1/2} \|_\Phi, \]

where \( \| \cdot \|_\Phi \) is any unitarily invariant norm on the space of Hermitian matrices of dimension \( n \) defined by \( \|X\|_\Phi := \Phi(\lambda_1(X), \ldots, \lambda_n(X)) \), with \( \lambda_{\min}(X) = \lambda_\Lambda(X) \leq \ldots \leq \lambda_1(X) = \lambda_{\max}(X) \) denoting the \( n \) real eigenvalues of \( X \) and \( \Phi \) a symmetric gauge norm on \( \mathbb{R}^n \) [3]. 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 \( \Phi(x_1, \ldots, x_n) = (\sum_i x_i^2)^{1/2} \), \( d_2 := d_\Phi \) 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 of this geometry is sometimes denoted by

\[ (2.27) \]

\[ ds = \left[ \text{tr} \left( \Sigma^{-1}d\Sigma \right)^2 \right]^{1/2}. \]

The unique Riemannian geodesic from \( A \) to \( B \) is given by the curve \( \gamma_G : [0, 1] \to \mathbb{P}(n) \) defined by

\[ (2.28) \]

\[ \gamma_G(t) = A^{1/2} \left( A^{-1/2}BA^{-1/2} \right)^t A^{1/2}. \]

This curve is significant as a minimal geodesic for any of the affine-invariant metrics \( d_\Phi \) [3]. The midpoint of \( \gamma_G \) is the matrix geometric mean \( A \# B \) (1.7), 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 \( \Phi \). With \( \Phi(x_1, \ldots, x_n) = \max_i |x_i| \), \( d_\infty := d_\Phi \) yields the distance function that coincides with the Thompson metric [39] on the cone \( \mathbb{P}(n) \)

\[ (2.29) \]

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

The invariant Finsler metrics (2.26) 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
\(a_1, \ldots, a_N\) can be defined as
\[
\text{argmin}_{x \in \mathbb{R}} \left( \sum_i (x - a_i)^2 \right)^{1/2} = \text{argmin}_{x \in \mathbb{R}} \|x1 - a\|_2, \tag{2.30}
\]
\[
\text{argmin}_{x \in \mathbb{R}} \sum_i |x - a_i| = \text{argmin}_{x \in \mathbb{R}} \|x1 - a\|_1, \tag{2.31}
\]
\[
\text{argmax}_{i} |x - a_i| = \text{argmin}_{x \in \mathbb{R}} \|x1 - a\|_\infty, \tag{2.32}
\]
respectively, where \(1 = (1, \ldots, 1) \in \mathbb{R}^N\) and \(a = (a_1, \ldots, a_N)\). By analogy, one can extend these notions to geometric averages for a collection of data \(Y_1, \ldots, Y_N \in \mathbb{P}(n)\) arising as
\[
\text{argmin}_{X \succeq 0} \Phi_N(\mathcal{d}_\Phi(X, Y_i)), \tag{2.33}
\]
where \(\Phi_n\) denotes the gauge norm on the space of \(n \times n\) Hermitian matrices and \(\Phi_N\) denotes the corresponding gauge function acting on the \(N\) distances \(\mathcal{d}_\Phi(X, Y_i)\). If \(\Phi\) corresponds to the \(l_2\) vector norm, (2.33) yields the geometric mean \(\mathcal{G}_2\) of \(Y_1, \ldots, Y_N\), also known as the Karcher mean [31, 6, 4]:
\[
\mathcal{G}_2(Y_1, \ldots, Y_N) = \text{argmin}_{X \succeq 0} \sum_{i=1}^N \mathcal{d}_2(X, Y_i)^2. \tag{2.34}
\]
If \(N = 2\), the unique solution \(\mathcal{G}_2(A, B)\) of (2.34) coincides with the geometric mean \(A \# B\). One can also define geometric medians of \(Y_1, \ldots, Y_N\) to be solutions to (2.33) for the choice of \(\Phi(x_1, \ldots, x_n) = \sum_i |x_i|\):
\[
\mathcal{G}_1(Y_1, \ldots, Y_N) = \text{argmin}_{X \succeq 0} \sum_{i=1}^N \mathcal{d}_1(X, Y_i). \tag{2.35}
\]
Note that the \(d_1\) distance of \(X \in \mathbb{P}(n)\) to the identity \(I\) takes the form
\[
d_1(X, I) = \|\log X\|_1 = \sum_{i=1}^N |\log \lambda_i(X)| = \text{tr} \left( (\log X \log X)^{1/2} \right). \tag{2.36}
\]
It is interesting to compare (2.36) to the function \(F(X) = \log \det(X)\), which plays an important role in convex optimization [10]. In particular, we have
\[
F(X) = \log \det(X) = \text{tr}(\log X) = \sum_{i=1}^N \log \lambda_i(X). \tag{2.37}
\]
If \(X \succeq I\), then \(\log X \succeq 0\) and hence \(d_1(X, I) = \text{tr}(\log X) = \log \det X\).

If \(\Phi\) corresponds to the \(l_\infty\)-norm, then (2.33) yields the geometric midrange problem (1.5). In the \(N = 2\) case, we have already seen that \(A \ast B\) is a solution to the corresponding midrange optimization problem
\[
\text{argmin}_{X \succeq 0} \max\{d_\infty(X, A), d_\infty(X, B)\}. \tag{2.38}
\]
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 $A \ast B$ of $A, B \in \mathbb{P}(n)$ that does not rely on any additional structures on $\mathbb{P}(n)$ except for the standard L"owner 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,

\[
A \ast B = \max_{X \in \text{span}\{A,B\}} \left\{ X = X^* : \begin{pmatrix} A & X \\ X & B \end{pmatrix} \succeq 0 \right\} 
\]

\[
= \max_{a,b \in \mathbb{R}} \left\{ aA + bB : \begin{pmatrix} A & aA + bB \\ aA + bB & B \end{pmatrix} \succeq 0 \right\}.
\]

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

\[
\begin{pmatrix} A & X \\ X & B \end{pmatrix} \sim \begin{pmatrix} I & -XB^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} A & X \\ X & B \end{pmatrix} \begin{pmatrix} I & -B^{-1}X \\ 0 & I \end{pmatrix}
\]

\[
= \begin{pmatrix} A - XB^{-1}X & 0 \\ 0 & B \end{pmatrix}.
\]

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

\[
B^{-1/2}AB^{-1/2} \succeq B^{-1/2}XB^{-1}X B^{-1/2} = (B^{-1/2}XB^{-1/2})^2.
\]

Using the monotonicity of the square root function $\Sigma \mapsto \Sigma^{1/2}$ on $\mathbb{P}(n)$ and the affine-invariance of the L"owner order, (2.43) holds if and only if

\[
X \preceq B^{1/2}(B^{-1/2}AB^{-1/2})^{1/2} B^{1/2}.
\]

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 $B$, (2.44) becomes

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

\[
\iff a\Sigma + bI \preceq \Sigma^{1/2},
\]

where $\Sigma = B^{-1/2}AB^{-1/2}$. 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 $D = \text{diag}(\lambda_1(\Sigma), \ldots, \lambda_n(\Sigma))$. Therefore, we have $aD + bI \preceq D^{1/2}$, which is equivalent to

\[
a\lambda_i(\Sigma) + b \leq \sqrt{\lambda_i(\Sigma)},
\]

for $i = 1, \ldots, n$. If we require that equality hold in (2.48) for $i = 1$ and $i = n$, so that

\[
\begin{cases}
  a\lambda_{\min}(\Sigma) + b = \sqrt{\lambda_{\min}(\Sigma)} \\
  a\lambda_{\max}(\Sigma) + b = \sqrt{\lambda_{\max}(\Sigma)}
\end{cases}
\]
we find that
\begin{equation}
(2.50) \quad a = \frac{\sqrt{\lambda_{\text{max}}(\Sigma)} - \sqrt{\lambda_{\text{min}}(\Sigma)}}{\sqrt{\lambda_{\text{max}}(\Sigma)} - \sqrt{\lambda_{\text{min}}(\Sigma)}} = \frac{1}{\sqrt{\lambda_{\text{max}}(\Sigma)} + \sqrt{\lambda_{\text{min}}(\Sigma)}}
\end{equation}
\begin{equation}
(2.51) \quad b = \frac{\sqrt{\lambda_{\text{min}}(\Sigma)} \lambda_{\text{max}}(\Sigma)}{\sqrt{\lambda_{\text{max}}(\Sigma)} + \sqrt{\lambda_{\text{min}}(\Sigma)}}.
\end{equation}

For this choice of $a$ and $b$, we have $aA + bB = A \ast B$. Moreover, (2.48) is satisfied for each $i = 1, \ldots, n$ since
\begin{equation}
\sqrt{\lambda_{i}(\Sigma)} - a\lambda_{i}(\Sigma) - b = \sqrt{\lambda_{\text{min}}(\Sigma)} \left( \frac{\sqrt{\lambda_{\text{max}}(\Sigma)} \lambda_{i}(\Sigma) - \lambda_{\text{max}}(\Sigma)}{\sqrt{\lambda_{\text{max}}(\Sigma)} + \sqrt{\lambda_{\text{min}}(\Sigma)}} \right) \geq 0.
\end{equation}

Imposing equality in (2.48) 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 (2.48). Therefore, our choice of $a$ and $b$ is indeed optimal.

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
\begin{equation}
(2.54) \quad de^{X}(Z) = \left. \frac{d}{dt} \right|_{t=0} e^{X+tz}.
\end{equation}

The following exponential metric increasing property is established by Bhatia in [3]. See [25] for an earlier version of the theorem.

Theorem 2.9. For any symmetric gauge norm $\Phi$ and Hermitian matrices $X$ and $Z$, we have
\begin{equation}
(2.55) \quad \|Z\|_{\Phi} \leq \|e^{-X/2}de^{X}(Z)e^{-X/2}\|_{\Phi},
\end{equation}
where $\|\cdot\|_{\Phi}$ denotes the unitarily invariant norm induced by $\Phi$.

This theorem has several important consequences, which we will briefly review. First note that the distance functions $d_{\Phi}$ defined in (2.26) are induced by the affine-invariant Finsler structures on $\mathbb{P}(n)$ given by
\begin{equation}
(2.56) \quad \|d\Sigma\|_{\Sigma,\Phi} := \|\Sigma^{-1/2}d\Sigma\Sigma^{-1/2}\|_{\Phi},
\end{equation}
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 $\gamma$ in $\mathbb{P}(n)$. We can express any such curve as the image of a curve $\Gamma$ in $\mathbb{H}(n)$ under the exponential map. In particular, any smooth curve $\gamma : [0,1] \rightarrow \mathbb{P}(n)$ from $I$ to $\Sigma$ 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 (2.56) is

\begin{align}
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}d\Gamma'(t)(\Gamma'(t))^{-1/2}\|_{\Phi} dt \\
&\geq \int_0^1 \|\Gamma'(t)\|_{\Phi} dt.
\end{align}

The last integral is simply the length of the curve \(\Gamma\) 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 \(\Sigma\) is defined as

\begin{equation}
\inf_{\gamma} L_{\Phi}[\gamma],
\end{equation}

where the infimum is taken over all smooth curves \(\gamma\) from \(I\) to \(\Sigma\). Therefore, we see that

\begin{equation}
d_{\Phi}(I, \Sigma) = \inf_{\gamma} L_{\Phi}[\gamma],
\end{equation}

and that this distance is attained by the curve \(\gamma(t) = e^{t \log \Sigma}\), which is a geodesic from \(I\) to \(\Sigma\). See Figure 2.1.

Since congruence transformations are isometries of \((\mathbb{P}(n), d_{\Phi})\), it follows that the curve

\begin{equation}
\gamma^{A,B}_{0}(t) = A^{1/2} \exp(t \log(A^{-1/2}BA^{-1/2}))A^{1/2},
\end{equation}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure2.1}
\caption{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 \(\mathcal{M}\). \((\mathbb{P}(n), d_{\Phi})\) is a manifold that satisfies such an exponential metric increasing property.}
\end{figure}
is a geodesic from \( A \) to \( B \). Note that this is precisely in agreement with (2.28). This geodesic is unique provided that the geodesics in \( \mathbb{R}^n \) induced by \( \Phi \) are unique.

In particular, uniqueness of geodesics in \((\mathbb{P}(n), d_\Phi)\) is inherited from \( \mathbb{R}^n \) when \( \Phi \) 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 \((\mathbb{P}(n), d_\Phi)\) is a space of non-positive curvature for any choice of \( \Phi \). See [11, 3] for further details. A closely related result is the geodesic convexity [26] of \( d_\Phi \), which follows from the inequality

\[
(2.64) \quad d_\Phi(e^{tX}, e^{tZ}) \leq t d_\Phi(e^X, e^Z),
\]

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

\[
(2.65) \quad t \mapsto d_\Phi(\tilde{\gamma}^{A_1, A_2}(t), \tilde{\gamma}^{B_1, B_2}(t))
\]

is convex for any symmetric gauge norm \( \Phi \) [3].

In geometric midrange statistics we are interested in the distance \( d_\infty \), 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 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 \( \phi : [0, 1] \to C \) given by

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

is always a minimal geodesic from \( x \) to \( y \) with respect to the Thompson metric. The curve \( \phi \) defines a projective straight line in the cone [33]. 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 (2.66) takes the form

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

where \( \lambda_{\max} \) and \( \lambda_{\min} \) denote the largest and smallest eigenvalues of \( BA^{-1} \), respectively. Taking the midpoint \( t = 1/2 \) of this geodesic, we recover the \( d_\infty \)-midpoint \( A*B \) (1.6) of \( A \) and \( B \). Thus, we have arrived at another interpretation of \( A*B \) as the midpoint of a suitable geodesic in \((\mathbb{P}(n), d_\infty)\). 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 (1.6). It is shown in [28] that \( \phi = \phi(t; A, B) \) is the unique \( d_\infty \) 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 \( d_\infty \) minimal geodesics from \( A \) to \( B \), and that the set of \( d_\infty \)-midpoints of \( A \) and \( B \) is compact and convex in both Riemannian and Euclidean senses [28].
3. The $N$-point geometric midrange problem. Given a collection of $N$ points $Y_1, \ldots, Y_N$ in $\mathbb{P}(n)$, the midrange problem can be formulated as the following optimization problem

\begin{equation}
\min_{X \geq 0} \max_i d_\infty(X, Y_i).
\end{equation}

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 not smooth.

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

\begin{equation}
l = \frac{1}{2} \operatorname{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.
\end{equation}

**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 $i, j = 1, \ldots, N$,

\begin{equation}
d_\infty(Y_i, Y_j) \leq d_\infty(Y_i, X^*) + d_\infty(X^*, Y_j) \leq t^* + t^* = 2t^*.
\end{equation}

Taking the maximum of the left-hand side over $i, j$, we arrive at $l = \frac{1}{2} \operatorname{diam}_\infty(\{Y_i\}) \leq t^*$. 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 $t^* \leq u = \min_i \max_j d_\infty(Y_i, Y_j)$. \qed

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 still yield an upper bound on the optimum cost $t^*$. Thus we have $t^* \leq u = \min_i \max_j d_\infty(Y_i, Y_j)$.

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

\begin{equation}
\min_{x > 0} \max_i \|\log x - \log y_i\|_\infty := \min_{x > 0} \max_i \|\log x^a - \log y_i^a\|,
\end{equation}

where $x > 0$ means that $x = (x^a)$ satisfies $x^a > 0$ for $a = 1, \ldots, n$ and $y_i$ are a collection of $N$ given points in $\mathbb{R}_+^n$. As in the matrix case, the optimum cost $t^* = \min_{x > 0} f(x)$ has a lower bound

\begin{equation}
l = \frac{1}{2} \max_{i,j} \|\log y_i - \log y_j\|_\infty.
\end{equation}

**Proposition 3.2.** The lower bound (3.5) is attained by $x = (x^a) \in \mathbb{R}_+^n$ defined by

\begin{equation}
x^a = \left(\min_i y_i^a \cdot \max_i y_i^a\right)^{1/2}.
\end{equation}

**Proof.** Note that

\begin{equation}
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 \max_{i,j} \frac{y_i^a}{y_j^a}\right|.
\end{equation}
With \( x \) as defined in (3.6), we have
\[
f(x) = \max_{k,a} \left[ \frac{1}{2} \log \left( \min_i y_i^a \cdot \max_i y_i^a \right) - \log y_0 \right] = \max_{k,a} \left[ \frac{1}{2} \log \left( \frac{\min_i y_i^a \cdot \max_i y_i^a}{y_k^0} \right) \right] = \frac{1}{2} \log \left( \frac{\min_i y_i^a \cdot \max_i y_i^a}{\min_j y_j^a} \right) \]
(3.8)  \[
= \max_{a} \frac{1}{2} \log \left( \frac{\min_i y_i^a \cdot \max_i y_i^a}{\min_i y_i^a} \right) \]
which has the equivalent epigraph formulation
\[
\left\{ \begin{array}{l}
\min_{X \succeq 0} t \\
-t \leq \log(\lambda_{\min}(Y_i^{-1/2}XY_i^{-1/2})) \\
-t \leq \log(\lambda_{\min}(Y_j^{-1/2}XY_j^{-1/2})) \end{array} \right. \quad \text{for all } i, j.
\]
This can be rewritten as the quasiconvex problem
\[
\left\{ \begin{array}{l}
\min_{X \succeq 0, t \in \mathbb{R}} t \\
e^{-t} Y_i \preceq X \preceq e^t Y_i.
\end{array} \right.
\]
(3.11)

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 [20]. Given a \( t \) that is greater than or equal to the optimum value \( t^* = \min_X \max_i d_\infty(X, Y_i) \), we can solve \( (3.11) \) using the bisection method [10] 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 \( (3.11) \) in one step by taking \( t = l \) and solving the feasibility condition once. Unfortunately, and rather remarkably, numerical examples show that unlike the scalar and vector case, 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 \( P(2) \) for
\[
Y_1 = \begin{pmatrix} 0.95 & -0.6 \\ -0.6 & 1.1 \end{pmatrix}, \quad Y_2 = \begin{pmatrix} 1.0 & 0.5 \\ 0.5 & 2.1 \end{pmatrix}, \quad Y_3 = \begin{pmatrix} 2.5 & -0.2 \\ -0.2 & 1.2 \end{pmatrix}.
\]
(3.12)

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 \( (3.11) \) via the bisection method yields the midrange
\[
X^* = \begin{pmatrix} 1.3154 & -0.5321 \\ -0.5321 & 1.6217 \end{pmatrix}
\]
(3.13)
with minimum cost \( t^* = 0.7901 > 0.7880 = l \). Indeed, we have \( t^* = \lambda_\infty(X^*, Y_1) = \lambda_\infty(X^*, Y_2) = \lambda_\infty(X^*, Y_3) \).

The above result suggests that the \( N \)-point matrix midrange problem is more challenging than the vector case in fundamental ways. While the bisection method applied to the quasiconvex problem (3.11) offers a solution, we expect that 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 [40, 34], 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) [13, 9, 16, 14], 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 [17, 18]. 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 (3.11) 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 3.5.** If \( Y_1, \ldots, Y_N \) are such that \( Y_1 \preceq Y_i \preceq Y_N \) for all \( i = 1, \ldots, N \), then the geometric midrange of \( \{Y_i\} \) is given by the set of \( \lambda_\infty \)-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 (3.11) 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 \( \lambda_\infty \)-midpoint of this pair. Furthermore, the lower bound \( l = \frac{1}{2} \text{diam}_\infty(\{Y_i\}) = \frac{1}{2} \lambda_\infty(Y_1, Y_N) \) is trivially attained.

**Remark 3.6.** 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 case. 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.

**Definition 3.7.** (Active matrices) Let \( N \geq 2 \) and \((X^*, t^*)\) be a solution to (3.11). Then \( Y_j \) is called active if at least one of the following hold:

\[
(3.15) \quad - \log(\lambda_{\min}(Y_j^{-\frac{1}{2}}X^*Y_j^{-\frac{1}{2}})) = t^* \quad \text{or} \quad \log(\lambda_{\max}(Y_j^{-\frac{1}{2}}X^*Y_j^{-\frac{1}{2}})) = t^*.
\]
Proposition 3.8. Let $N \geq 2$ and $(X^*, t^*)$ be a solution to (3.11). Then there exist distinct $i^*, j^* \in \{1, \ldots, N\}$ such that

\begin{equation}
\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^* \tag{3.16}
\end{equation}

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

\begin{equation}
\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}})) \geq 0. \tag{3.17}
\end{equation}

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{align*}
|\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 \tag{3.18} \\
\log(\lambda_{\max}(Y_i^{-\frac{1}{2}}X^*Y_i^{-\frac{1}{2}})) \leq \log(\lambda_{\max}(Y_j^{-\frac{1}{2}}X^*Y_j^{-\frac{1}{2}})) = t^* & \text{ for all } i.
\end{align*}

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

\begin{align*}
|\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 \tag{3.19} \\
\log(\lambda_{\max}(Y_i^{-\frac{1}{2}}\tilde{X}^*Y_i^{-\frac{1}{2}})) \leq \log(\lambda_{\max}(Y_j^{-\frac{1}{2}}\tilde{X}^*Y_j^{-\frac{1}{2}})) & \leq t^* \text{ for all } i,
\end{align*}

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.

Proposition 3.9. Recall the $N = 2$ geometric midrange problem for $Y_1$ and $Y_2$ in $P(n)$. Set $\alpha := \lambda_{\max}(Y_1^{-\frac{1}{2}}Y_2Y_1^{-\frac{1}{2}})$ and $\beta := \lambda_{\min}(Y_1^{-\frac{1}{2}}Y_2Y_1^{-\frac{1}{2}})$. If $\alpha \neq \beta$, then

\begin{equation}
X^* = \frac{\sqrt{\alpha \beta}Y_1 + Y_2}{\sqrt{\alpha} + \sqrt{\beta}} = Y_1 + Y_2 \tag{3.20}
\end{equation}

is the only midrange of $\{Y_1, Y_2\}$ in $\{k_1Y_1 + k_2Y_2 : k_1, k_2 \geq 0\}$ for which the following is satisfied:

\begin{align*}
\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}})). \tag{3.21}
\end{align*}

The optimal cost to (3.11) 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{align*}
\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 \tag{3.22} \\
\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 \tag{3.23} \\
\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 \tag{3.24} \\
\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. \tag{3.25}
\end{align*}
Substituting these expressions into (3.21), we find that

\[ k_1 + k_2 \alpha = \frac{1}{k_1 + k_2} \alpha \]

and hence

\[ k_1 \beta + k_2 = \frac{1}{k_1 + k_2} \beta \]

which implies that \( k_1^2 = k_2^2 \alpha \beta \) since \( \alpha = \beta \). Substituting \( k_1 = \sqrt{\alpha \beta} \) into (3.26) gives

\[ k_2 \beta + 2k_2^2 \sqrt{\alpha \beta} + k_2^2 \alpha = k_2^2 (\sqrt{\alpha} + \sqrt{\beta})^2 = 1 \]

and hence (3.20). 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 3.8.

**Remark 3.10.** Note that if \( \alpha = \beta \) in the statement of the previous proposition, then \( Y_1^{-1/2} Y_2 Y_1^{-1/2} = \alpha I \), which is equivalent to \( Y_2 = \alpha Y_1 \). The midrange \( Y_1 \ast Y_2 = \sqrt{\alpha Y_1} \) can then be obtained as a conic combination of \( Y_1 \) and \( Y_2 = \alpha Y_1 \) in a non-unique 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 (3.11) when \( N \geq 2 \).

**Lemma 3.11.** Let \( N \geq 2 \) and \((X^*, t^*)\) be a solution to (3.11). Then the following are equivalent:

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

Analogously, we have the equivalences:

i. \( Y_j \) is active with \( -\log (\lambda_{\min}(Y_j^{-1/2} X^* Y_j^{-1/2})) = t^* \)
ii. \( \lambda_{\max}(X^* - e^{-t^*} Y_j) = 0 \)
iii. \( \exists \varepsilon > 0 : X^* \succeq e^{t^*} Y_j - \varepsilon I \)

**Lemma 3.12.** Let \( D = \text{diag}(d_1, \ldots, d_n) \) with \( d_1 \leq \cdots \leq d_n \) 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-Fisher, \( e_n = (0, 0, \ldots, 1) \) is an eigenvector of \( X \) with eigenvalue \( d_n \) and thus \( X \) has the required form.

**Proposition 3.13.** Let \((X^*, t^*)\) be a solution to (3.11) and

\[ t^* = \frac{1}{2} \log \left( \lambda_{\max}(Y_1^{-1/2} Y_2 Y_1^{-1/2}) \right) \]
Then, $Y_1$ and $Y_2$ are active with
\begin{equation}
\log \left( \lambda_{\max} \left( Y_1^{\frac{1}{2}} X^{\star} Y_1^{-\frac{1}{2}} \right) \right) = t^* = - \log \left( \lambda_{\min} \left( Y_2^{\frac{1}{2}} X^{\star} Y_2^{-\frac{1}{2}} \right) \right).
\end{equation}
Further, if $Y_2 V = Y_1 V D$ is a generalized eigenvalue decomposition such that $V^{\star} Y_1 V = I$ and $D = \text{diag}(d_1, \ldots, d_n)$ with $d_1 \leq \cdots \leq d_n = \lambda_{\max} \left( Y_1^{\frac{1}{2}} Y_2 Y_1^{-\frac{1}{2}} \right)$, then
\begin{equation}
V^{\star} X^{\star} V = \begin{pmatrix} X_{11} & 0 \\ 0 & \sqrt{d_n} \end{pmatrix}.
\end{equation}

Proof. Let $X_V := V^{\star} X^{\star} V$. We first show that $X_V$ has the claimed structure. To this end, note that by (3.11)
\begin{equation}
e^{-t^*} I \preceq X_V \preceq e^{t^*} I
\end{equation}
\begin{equation}
e^{-t^*} D \preceq X_V \preceq e^{t^*} D
\end{equation}
which implies that
\begin{equation}
e^{-t^*} D \preceq X_V \preceq e^{t^*} I.
\end{equation}
with $e^{t^*} = \sqrt{d_n}$. Therefore, Lemma 3.12 implies the required structure for $X_V$. Then by Lemma 3.11 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 $(X^{\star}, t^*)$. Then by Lemma 3.12 we can conclude the remaining claim.

Proposition 3.14. If there are only two active matrices at an optimum $(X^{\star}, t^*)$ of (3.11), 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 $(X^{\star}, t^*)$ and assume that the lower bound $l$ is not attained so that $l < t^*$. Denote the geodesic (2.28) from $X^{\star}$ to the $d_\infty$-midpoint set of the active pair of matrices, it does not imply that any $d_\infty$-midpoint of $Y_1, Y_2$ will be a solution. In particular, $Y_1 \ast Y_2$ may not be a solution even if the only active matrices at optimum are $Y_1$ and $Y_2$ since $Y_1 \ast Y_2$ may fail to satisfy one or more of the constraints in (3.11). This is in contrast to the scenario in Proposition 3.5, 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, an algorithmic solution to the $N$-point problem is offered via quasiconvex 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 work on statistics of extremes [21] and clustering algorithms such as $K$-midranges [37, 12] 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 offering a possible angle of attack.

REFERENCES

[1] T. Ando, C.-K. Li, and R. Mathias, Geometric means, Linear Algebra and its Applications, 385 (2004), pp. 305 – 334, https://doi.org/https://doi.org/10.1016/j.laa.2003.11.019, http://www.sciencedirect.com/science/article/pii/S0024379503008693. Special Issue in honor of Peter Lancaster.

[2] 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, https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.20965, https://arxiv.org/abs/https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrm.20965.

[3] R. Bhatia, On the exponential metric increasing property, Linear Algebra and its Applications, 375 (2003), pp. 211 – 220, https://doi.org/https://doi.org/10.1016/S0024-3795(03)00647-5, http://www.sciencedirect.com/science/article/pii/S0024379503006475.

[4] R. Bhatia, Positive Definite Matrices, Princeton University Press, 2007.

[5] R. Bhatia and J. Holbrook, Riemannian geometry and matrix geometric means, Linear Algebra and its Applications, 413 (2006), pp. 594 – 618, https://doi.org/https://doi.org/10.1016/j.laa.2005.08.025, http://www.sciencedirect.com/science/article/pii/S0024379505004350. Special Issue on the 11th Conference of the International Linear Algebra Society, Coimbra, 2004.

[6] R. Bhatia and J. Holbrook, Riemannian geometry and matrix geometric means, Linear Algebra and its Applications, 413 (2006), pp. 594 – 618, https://doi.org/https://doi.org/10.1016/j.laa.2005.08.025, http://www.sciencedirect.com/science/article/pii/S0024379505004350. Special Issue on the 11th Conference of the International Linear Algebra Society, Coimbra, 2004.

[7] D. Bini, B. Meini, and F. Poloni, An effective matrix geometric mean satisfying the Ando-Li-Mathias properties, Math. Comput., 79 (2010), pp. 437–452.

[8] D. A. Bini and B. Iannazzo, Computing the Karcher mean of symmetric positive definite matrices, Linear Algebra and its Applications, 438 (2013), pp. 1700 – 1710, https://doi.org/https://doi.org/10.1016/j.laa.2011.08.052, http://www.sciencedirect.com/science/article/pii/S0024379511006616. 16th ILAS Conference Proceedings, Pisa 2010.

[9] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Foundations and Trends in Machine Learning, 3 (2011), pp. 1–122.

[10] S. Boyd and L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004, https://doi.org/10.1017/CBO9780511804441.

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

[12] 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., Berlin, Heidelberg, 1998, Springer Berlin Heidelberg, pp. 3–14.

[13] P. L. Combettes and J.-C. Pesquet, Proximal Splitting Methods in Signal Processing, Springer New York, 2011, pp. 185–212.
[38] S. M. Stigler, The seven pillars of statistical wisdom, Harvard University Press, 2016.
[39] 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.
[40] 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.
[41] H. Zhu, H. Zhang, J. G. Ibrahim, and B. S. Peterson, Statistical analysis of diffusion tensors in diffusion-weighted magnetic resonance imaging data, Journal of the American Statistical Association, 102 (2007), pp. 1085–1102, https://doi.org/10.1198/016214507000000581, https://doi.org/10.1198/016214507000000581, https://arxiv.org/abs/https://doi.org/10.1198/016214507000000581.