Hilbert geometry of the symmetric positive-definite bicone

Jacek Karwowski* JACEK.KARWOWSKI@CS.OX.AC.UK
Department of Computer Science
University of Oxford, UK

Frank Nielsen* NIELSEN@CSL.SONY.CO.JP
Sony Computer Science Laboratories Inc.
Tokyo, Japan

Editors: List of editors' names

Abstract

The extended Gaussian family is the closure of the Gaussian family obtained by completing the Gaussian family with the counterpart elements induced by degenerate covariance or degenerate precision matrices, or a mix of both degeneracies. The parameter space of the extended Gaussian family forms a symmetric positive semi-definite matrix bicone, i.e. intersection of two partial symmetric positive semi-definite matrix cones. In this paper, we study the Hilbert geometry of such an open bounded convex symmetric positive-definite bicone. We report the closed-form formula for the corresponding Hilbert metric distance and study exhaustively its invariance properties. We also touch upon potential applications of this geometry for dealing with extended Gaussian distributions.

Keywords: Extended Gaussian distributions; Symmetric positive semi-definite cone; bicone; Hilbert geometry; projective geometry; invariance; open stochastic systems.

1. Introduction

In this paper, we consider the Hilbert geometry for the covariance-precision bicone domain. To first motivate our study, we introduce the bicone closed parameter space of the extended Gaussian family. We can parametrize the Gaussian family centered at the origin[1] using either its covariance matrix Σ, or its precision matrix $P = \Sigma^{-1}$. That is, the Gaussian density with respect to the Lebesgue measure can be written:

$$p_\Sigma(x) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}\langle x, \Sigma^{-1}x\rangle\right) = \frac{\sqrt{\det(P)}}{(2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2}\langle x, Px\rangle\right) = p_P(x)$$

where $\langle x, x'\rangle = \sum_{i=1}^n x_i x_i'$ denotes the Euclidean inner product, and $N_0(0,\Sigma) \sim p_\Sigma(x) = p_P(x)$ are two parameterizations.

The parameter space of the Gaussian family $\{N_0(\Sigma)\}$ then consists of symmetric positive-definite cone (SPD) $\mathrm{PD} = \{X \in \mathrm{Sym}(\mathbb{R}, n) : X \succ 0\}$, where $\succ$ denotes the Loewner partial ordering on the vector space $\mathrm{Sym}(\mathbb{R}, n)$ of $n \times n$ real symmetric matrices (written concisely as $\mathrm{Sym}(n)$ in the remainder): that is, $A \succ B$ if and only if $A - B \in \mathrm{PD}$. The topological closure of the PD cone is the positive semi-definite (PSD) cone: $\mathrm{PSD} = \{X \in \mathrm{Sym}(n) : X \succeq 0\}$ where $A \succeq B$ iff $A - B \in \mathrm{PSD}$. From the viewpoint of geometry, we consider the Gaussian

* Equal contribution.
1. We discuss the generalization to non-centered Gaussians in Section 6.

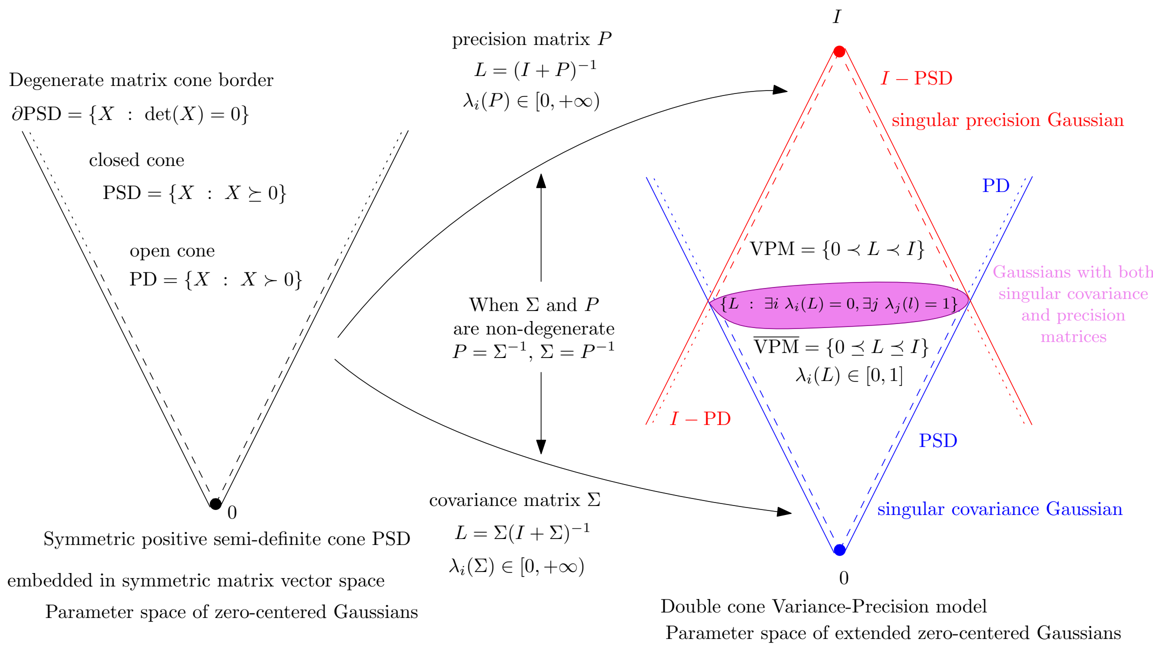


Figure 1: The parameter space of the extended Gaussian family forms a closed bicone.

family as a cone manifold equipped with a single global coordinate system (the covariance or precision coordinate systems which form a pair of dual coordinate systems $(\Sigma(\cdot), \mathrm{PD})$ and $(P(\cdot), \mathrm{PD})$ in the setting of information geometry (Ohara et al., 1996; Ohara, 2019)).

Gaussian distributions with degenerate covariance matrices Σ are commonly considered as Gaussian distributions defined on lower-dimensional affine subspaces $\mathrm{range}(\Sigma) = \{\Sigma x : x \in \mathbb{R}^n\}$ of $\mathbb{R}^n$. James (1973) considered Gaussian distributions with degenerate precision matrices, and more generally proposed to extend the Gaussian family by considering the following reparameterization of Gaussians:

$$
\begin{aligned}
L(\Sigma) &= \Sigma(I + \Sigma)^{-1}, & \Sigma(L) &= L(I - L)^{-1}, \\
L(P) &= (I + P)^{-1}, & P(L) &= L^{-1} - I.
\end{aligned}
$$

where $I = I_n$ denotes the $n \times n$ identity matrix. For full rank covariance/precision matrices, the eigenvalues $\lambda_i(L)$ fall inside $(0, 1)$ for $i \in \{1, \ldots, n\}$. James (1973) first considered the general case of extended Gaussian distributions by considering "Gaussian elements", i.e., the counterpart of Gaussians with degenerate covariance and/or precision matrices, lying on the boundary of the bicone[2](see Figure 1). The variance-precision model $\overline{\mathrm{VPM}}$ is the intersection of the two PSD cones and the variance-precision manifold (named the variance-information manifold in James (1973)) VPM is its interior: $\mathrm{VPM} = \overline{\mathrm{VPM}} \backslash \partial\overline{\mathrm{VPM}}$.

The pioneer approach of James was later studied in depth using the framework of category theory of Stein and Samuelson (2023), further highlighting interpretations of the different types of degenerate extended Gaussian distributions and their duality relationships. Gaussian distributions with degenerate precision matrices encode non-determinism in the

2. Terminology: a bicone is two cones joined at their bases; a double cone is two cones joined at their apex.

open stochastic framework of Willems (2012). Loosely speaking, those degenerate precision Gaussians can be interpreted as "uniform distributions on $\mathbb{R}^n$"; which, even if strictly speaking are ill-defined with respect to Lebesgue measure, are sometimes considered as *improper priors* in Bayesian data analysis (Gelman et al., 1995).

1.1. Hilbert geometry and Birkhoff geometry

Hilbert geometry (Hilbert, 1895; Goldman, 2022) is a geometry defined on any open bounded convex domain C of a Hilbert space. It defines a metric distance d_H^C and ensures that the straight lines are geodesics, with uniqueness of geodesics depending on the smoothness of the boundary ∂C (Nielsen and Shao, 2017). Birkhoff geometry (Birkhoff, 1957; Lemmens and Nussbaum, 2014b) is a geometry defined on any convex pointed cone K. The cone defines a (partial) order which allows one to define a Birkhoff projective "distance" d_B on the K, satisfying $d_B(\lambda r, \lambda' r') = d_B(r, r')$ for any $\lambda, \lambda' > 0$, therefore descending to a well-defined distance on equivalence classes $[r]$ and $[r']$, where $r \sim \lambda r$ for any $\lambda > 0$. The Birkhoff distance is also called *Hilbert projective distance* (Nussbaum, 1988); they are related to each other via projectivization of K. Hilbert geometry for many different domains has been studied: e.g., the Hilbert geometry of the unit disk yields Klein model of hyperbolic geometry (Nielsen, 2020). The Hilbert geometry of simplices has been studied by de la Harpe (1993) (and more generally for polytopes by Lemmens and Walsh (2009)) with applications in machine learning in (Nielsen and Sun, 2018, 2023; Vanecek et al., 2024).

Birkhoff geometry of SPD matrices found many applications due to the fact that it exhibits a contraction property (Birkhoff, 1957; Chen et al., 2021) and only requires the extreme eigenvalues of matrices to compute the Birkhoff distance (Mostajeran et al., 2024). In comparison, the affine-invariant Riemannian metric distance (Thanwerdas and Pennec, 2023) (AIRM) requires the full spectral decomposition. The Hilbert geometry of the space of correlation matrices, whose domains are called elliptopes (Tropp, 2018), was studied in (Nielsen and Sun, 2018). Recently, Hilbert geometry gained interest in computational geometry due to its mathematical and computational tractability of their Voronoi diagrams (Gezalyan and Mount, 2023) and dual Delaunay triangulations (Gezalyan et al., 2024), or smallest enclosing balls (Nielsen and Sun, 2018; Banerjee et al., 2024), etc.

1.2. Contributions and paper outline

In this paper, we consider the Hilbert geometry for the covariance-precision bicone domain. The paper is organized as follows. In §2, we describe the Birkhoff projective geometry of regular open cones and the Hilbert geometry defined on open bounded convex domains. The Hilbert geometry of the open variance-precision manifold is studied in §3. We give an equivalent definition of the VPM of James (1973) (in Definition 10), which is proven open convex and bounded (Proposition 12). Automorphisms of the VPM bicones are reported in Proposition 13, and a closed-form formula for the Hilbert distance in the VPM is obtained in Theorem 14. In §4, we report two invariance properties of the Hilbert VPM distance: under transformations $X \mapsto I - X$ in Proposition 17 and under orthonormal matrix conjugation $X \mapsto U^T X U$ in Proposition 19. Section 5 proves that these two isometric transformations characterize all the VPM isometries (Theorem 39). Finally, we hint at some future directions, including applications, in §6. All proofs are given in Appendix A.

2. Birkhoff and Hilbert geometries

2.1. Cones and partial orders

Definition 1 (Cone order) *Given any vector space V, if $K \subseteq V$ satisfies the following:*

- *K is an open cone, i.e. for any $v \in K$ and any $c \in \mathbb{R}_{>0}$, we have $cv \in K$*
- *K is convex, i.e. for any $v, w \in K$ and any $c \in [0,1]$, we have $cv + (1-c)w \in K$*
- *K is pointed, i.e. $\overline{K} \cap (-\overline{K}) = \{0\}$*

then K defines a partial order $\prec_K$, or simply $\prec$, on V, by $v \prec w \iff w - v \in K$. We write $v \preceq w$ if $w - v \in \overline{K}$.

Definition 2 (Loewner order) *An example of cone order is the Loewner order $\preceq$ on the set of symmetric matrices $\mathrm{Sym}(n)$, arising from the convex open cone of symmetric positive definite matrices $\mathrm{PD}(n) \subset \mathrm{Sym}(n)$. That is, define a partial order on the space of symmetric matrices $A \prec B$ if and only if $0 \prec B - A \iff B - A \in \mathrm{PD}(n)$.*

Proposition 3 (Loewner order implies eigenvalues ordering) *Given two symmetric matrices A, B of $V = \mathrm{Sym}(n)$, if $A \preceq B$, then $\lambda_i(A) \leq \lambda_i(B)$, where $\lambda_i(X)$ denotes the i-th smallest eigenvalue of X (including multiplicities). The converse holds if the ordering is true element-wise in a common eigenbasis.*

2.2. Birkhoff projective distance

Definition 4 (Birkhoff projective distance) *Given a vector space V and an open convex pointed cone $K \subseteq V$ defining an order $\preceq$ on V, we define the Birkhoff distance d_B on K as follows:*

$$d_B(v, w) = \log \frac{M(v, w)}{m(v, w)}$$

where:

$$M(v, w) = \inf_{\lambda > 0} \{v \preceq \lambda w\} \qquad m(v, w) = \sup_{\mu > 0} \{\mu w \preceq v\}$$

Example 1 *Consider the symmetric positive-definite cone $\mathrm{PD}(n)$. Then the Birkhoff projective distance (Nielsen and Sun, 2018; Chen et al., 2021) is:*

$$d_B(P, Q) = \log \frac{\lambda_{\max}(PQ^{-1})}{\lambda_{\min}(PQ^{-1})}.$$

2.3. Hilbert metric distance

Definition 5 (Hilbert distance function) *Given an open convex set X in a normed vector space $(V, \|\cdot\|)$ with the boundary ∂X, we define the Hilbert distance function $d_H(x, y)$ on elements $x, y \in X$ by the following formula:*

$$d_H(x, y) = \log \frac{\|x - y'\|}{\|x' - x\|} \frac{\|x' - y\|}{\|y - y'\|}$$

where x', x, y, y' are points lying in this order on the line l_{xy}, such that $\{x', y'\} = \partial X \cap l_{xy}$.

Remark 6 *Hilbert distance in Definition 5 is independent of the norm on V, by the equivalence of one-dimensional norms to the absolute value $|\cdot|$ on l_{xy}.*

Example 2 *Let $X = (0,1)$ the open unit interval of $V = \mathbb{R}$. Then the Hilbert distance is:*

$$d_H(x,y) = \left|\log \frac{(1-x)y}{x(1-y)}\right|.$$

2.4. Relationship between Hilbert and Birkhoff distances

Definition 7 (Open pointed cone induced by a convex set) *Given a convex set $C \subseteq V$, we define the open pointed cone over C as $K(C) = \{(\lambda x, \lambda) : x \in C, \lambda \in \mathbb{R}_{>0}\}$.*

Definition 8 (Projective space over an open pointed cone) *For a pointed cone K in a vector space V, we define the projective space over K to be the space of half-lines in K, that is, $\mathbf{P}(K) = K/\sim$ where $x \sim y$ for $x, y \in K$ if and only if $x = \lambda y$ for some $\lambda \in \mathbb{R}_{>0}$.*

Proposition 9 (Birkhoff characterization of the Hilbert metric) *(Lemmens and Nussbaum, 2014a, Theorem 2.2) For a vector space V and an open bounded convex set $C \subseteq V$, the space $\mathbf{P}(K(C))$ equipped with the Birkhoff metric d_B is isometric to the space C equipped with the Hilbert metric d_H .*

3. Hilbert geometry of the Variance-Precision Model

Definition 10 (Variance-precision manifold $\mathrm{VPM}(n)$) *We define the variance-precision manifold of dimension n as the space $\mathrm{VPM}(n) := \{X \in \mathrm{Sym}(n) : 0 \prec X \prec I\}$.*

Remark 11 (Variance-precision model $\overline{\mathrm{VPM}(n)}$) *James (1973) defines the variance information manifold of dimension n as the space of real symmetric matrices $n \times n$, such that for all eigenvalues $\lambda_i, 1 \leq i \leq n$, we have $0 \leq \lambda_i \leq 1$. This corresponds to the closure $\overline{\mathrm{VPM}(n)} = \{X \in \mathrm{Sym}(n) : 0 \preceq X \preceq I\}$ in $\mathrm{Sym}(n)$ in our Definition 10, by the equivalence of Loewner order to eigenvalue ordering, as in Proposition 3.*

Proposition 12 $\mathrm{VPM}(n) \subset \mathrm{Sym}(n) \simeq \mathbb{R}^{n(n+1)/2}$ *is open in* $\mathrm{Sym}(n)$, *convex and bounded.*

Proposition 13 $\mathrm{VPM}(n)$ *is invariant under the mapping $X \mapsto I - X$ and conjugation with orthonormal matrices $X \mapsto U^T X U$ for $U \in O(n)$.*

Theorem 14 (Hilbert distance on $\mathrm{VPM}(n)$) *Given two matrices $A, B \in \mathrm{VPM}(n)$:*

$$d_H(A,B) = \log \frac{\max\left(\lambda_{\max}(B^{-1}A), \lambda_{\max}\big((I-B)^{-1}(I-A)\big)\right)}{\min\left(\lambda_{\min}(B^{-1}A), \lambda_{\min}\big((I-B)^{-1}(I-A)\big)\right)}$$

where $\lambda_{\min}$ and $\lambda_{\max}$ denote the minimal and maximal eigenvalues of respective matrices.

Remark 15 *Because of the logarithm of ratios, in Theorem 14, the Hilbert distance on* $\mathrm{VPM}(n)$ *can be equivalently written using matrices $A^{-1}B$ and $(I-A)^{-1}(I-B)$ in place of $B^{-1}A$ and $(I-B)^{-1}(I-A)$ respectively. The transformation $\mathrm{Mob}(A,B) = (I-A)^{-1}(I-B)$ is called a matrix Möbius transformation.*

4. Invariance properties of the Hilbert VPM distance

Let us first state the following two propositions:

Proposition 16 (ι map, James (1973)) *The variance-precision manifold is diffeomorphic to the set of symmetric positive-definite matrices of dimension n via the map $\iota(X) = X(I+X)^{-1}$ and its inverse $\iota^{-1}(A) = A(I-A)^{-1}$.*

Those mappings can be interpreted as matrix Möbius transformations. The formula of the Hilbert VPM distance from Theorem 14 immediately implies the following.

Proposition 17 (Identity-complement invariance) *The map $X \mapsto I - X$ is an isometry of $\mathrm{VPM}(n)$.*

Remark 18 *We observe that the map $X \mapsto X^{-1}$ is an isometry for the affine invariant Riemannian metric (AIRM) Pennec et al. (2006) (and for the log-Euclidean metric (Arsigny et al., 2006)). Thus, the invariance of the Hilbert metric on $\mathrm{VPM}(n)$ under map $X \mapsto I - X$ relates to the invariance of the standard Riemannian metric on $\mathrm{PD}(n)$ under matrix inverse.*

Proposition 19 (Conjugation invariance) *Conjugation under orthonormal matrix U, that is, a map $X \mapsto U^T X U$ is an isometry in the Hilbert metric on $\mathrm{VPM}(n)$.*

Remark 20 *Similarly, the conjugation with the orthonormal matrix $X \mapsto U^T X U$ is an isometry for the AIRM metric (and for the log-Euclidean metric). Thus, the invariance of the Hilbert metric on $\mathrm{VPM}(n)$ under conjugation with orthonormal matrix $X \mapsto U^T X U$ relates to the invariance of the standard Riemannian metric on $\mathrm{PD}(n)$ under conjugation.*

5. Characterization of VPM isometries

We now show that these two invariant transformations are the only ones by characterizing the isometries of the VPM when $n \geq 2$, which we assume everywhere below. The argument is based on the simplification from isometries of Hilbert metric to collineations of the underlying cone (for the Birkhoff characterization). We will denote the group of isometries of a metric space X by $\mathrm{Isom}(X)$.

Definition 21 *Given a real vector space V, three points $x, y, z \in V$ are called collinear if there exist real constants λ, λ' such that $(x - y) = \lambda(x - z) = \lambda'(y - z)$.*

Definition 22 (Collineation) *Given two vector spaces V, W and subsets $X \subseteq V$ and $Y \subseteq W$, an invertible transformation $f : X \to Y$ is called a collineation if for every three collinear points $x, y, z \in X$, their images $f(x), f(y), f(z)$ are also collinear. The group of collineations of W is denoted $\mathrm{Coll}(W)$.*

Proposition 23 (Goldman (2022)) *For an open convex bounded set $C \subseteq \mathbb{R}^n$ equipped with the Hilbert geometry d_H, we have $\mathrm{Coll}(C) \subseteq \mathrm{Isom}(C)$.*

The converse of Proposition 23 is not true in general. But we establish it for $\mathrm{VPM}(n)$ in particular, using the theory developed in Walsh (2017).

5.1. Isometries of VPM are collineations

Definition 24 (Lorentzian cone) *A cone $C \subseteq \mathbb{R}^n$ is called Lorentzian if it's of a form $C = \left\{(x_1, \dots x_n) : x_1 > 0,\ x_1^2 - \sum_{i>1} x_i^2 > 0\right\}$.*

Definition 25 (Dual cone) *For a cone C in a linear space V equipped with an inner product $\langle \cdot, \cdot \rangle$, its dual cone is defined to be $C^* = \{y \in V : \langle x, y \rangle > 0 \text{ for all } x \in C\}$.*

Definition 26 (Homogeneous cone) *A cone $C \subseteq V$ is said to be homogeneous, if for every two points $x, y \in C$ there exists a linear automorphism F of V, such that F restricts to an automorphism of C, and $F(x) = y$.*

Definition 27 (Symmetric cone) *An open convex pointed cone C is said to be symmetric if it is homogeneous, and equal to its dual $C = C^*$.*

Theorem 28 (Walsh (2017), Corollary 1.4) *Given a cone C and $D = \mathbf{P}(C)$ its projective space, we have that $\mathrm{Isom}(D)$ is a normal subgroup of $\mathrm{Coll}(D)$ if and only if C is symmetric and non-Lorentzian. Otherwise, $\mathrm{Isom}(D) = \mathrm{Coll}(D)$.*

Definition 29 (Cone C_n over $\mathrm{VPM}(n)$) *The cone over $\mathrm{VPM}(n)$ is given by:*

$$C_n = K(\mathrm{VPM}(n)) = \{(tX, t) : X \in \mathrm{VPM}(n),\ t \in \mathbb{R}_{>0}\} = \left\{(Y, t) : \frac{1}{t} Y \in \mathrm{VPM}(n),\ t \in \mathbb{R}_{>0}\right\}$$

for which $\mathrm{VPM}(n) = \mathbf{P}(C_n)$ is the projective space - in the projective coordinates we have $\mathrm{VPM}(n) \simeq [X : 1]$.

Proposition 30 *The cone C_n for $n > 1$ is not symmetric.*

Therefore, from Proposition 30 and Theorem 28 we derive the desired characterisation.

Corollary 31 *We have $\mathrm{Isom}(\mathrm{VPM}(n)) = \mathrm{Coll}(\mathrm{VPM}(n))$ for all $n \geq 2$.*

5.2. Classification of collineations of VPM

Let us classify the collineations of $\mathrm{VPM}(n)$. To do that, we use the fact that $\mathrm{VPM}(n)$ can be seen as the projective space of its cone $\mathbf{P}(C_n)$, as in in Definition 29.

Theorem 32 (Fundamental theorem of projective geometry) *Any bijective collineation $\hat{f} : \mathbb{RP}^n \to \mathbb{RP}^n$ is induced by some linear isomorphism $f : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$, such that $\hat{f} = [f]$, where $[f][x] = [fx]$.*

Definition 33 ($\mathrm{PGL}(n)$) *By $\mathrm{PGL}(n)$ we denote the projective general linear group of dimension n, that is, the group $\mathrm{GL}(\mathbb{R}, n)/\mathrm{GL}(\mathbb{R}, 1)$. That is, $F \sim G$ iff $F = aG$ for some real constant $a \neq 0$.*

Proposition 34 (Shiffman (1995), Lemma 4) *Given an open set $U \subseteq \mathbb{RP}^n$ for $n \geq 2$, and a continuous injective map $f : U \to \mathbb{RP}^n$, such that for all projective lines $L \subseteq \mathbb{RP}^n$, the set $f(L \cap U)$ is collinear, then there exist a collineation $\hat{f} : \mathbb{RP}^n \to \mathbb{RP}^n$ such that $\hat{f}|_U = f$.*

Combining Lemma 34 and Theorem 32 instead of classifying collineations of $\mathrm{VPM}(n)$, we can instead classify the linear cone isomorphisms, as we do in the next section.

5.2.1. Classification of the projectivizations of linear cone isomorphism

Definition 35 (Linear cone isomorphism) *For a vector space V and a cone $C \subseteq V$, we call a map $L : V \to V$ a linear cone automorphism for C, if L is a linear isomorphism and $L(C) = C$.*

Using the two lemmas below, we prove our major result following.

Lemma 36 (Gowda et al. (2013), Example 1) *Linear cone isomorphism $\mathcal{A} : \mathrm{Sym}(n) \to \mathrm{Sym}(n)$ for $\mathrm{PD}(n)$ exactly correspond to conjugation with some $U \in \mathrm{GL}(\mathbb{R}, n)$, that is $\mathcal{A}(X) = U^T XU$.*

Lemma 37 *Given a vector space V, a pointed closed cone K such that $0 \in K$, and a smooth curve $\eta(t) \in K$, for $t \in (-\epsilon, \epsilon)$ and $\epsilon > 0$, if $\eta(0) = 0$, then necessarily $\eta'(0) = 0$.*

Lemma 38 *Given an affine isomorphism $L : \mathrm{Sym}(n) \to \mathrm{Sym}(n)$ where $L(\mathrm{VPM}(n)) = \mathrm{VPM}(n)$, we either have:*

$$\begin{aligned} L(0) &= 0 \\ L(I) &= I \end{aligned} \qquad \textit{or} \qquad \begin{aligned} L(0) &= I \\ L(I) &= 0 \end{aligned}$$

Theorem 39 (Classification of VPM isometries) *The group of isometries of $\mathrm{VPM}(n)$ for $n > 1$ is generated by conjugation by orthonormal matrices $X \mapsto U^T XU$ for $U \in O(n)$ and inversion $X \mapsto I - X$.*

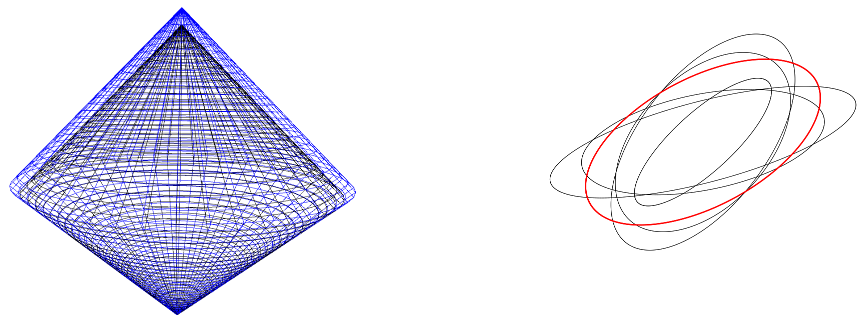
6. Discussion with some perspectives

We motivated this work by showing that the parameter space of the extended Gaussian family is a closed bicone called the variance-precision model. We then considered the Hilbert geometry for the corresponding interior of this parameter space, reported the closed-form formula for the Hilbert distance, and fully characterized its isometries. Below, we outline some extensions which we delegate to future work, along the full stratification of $\overline{\mathrm{VPM}}(n)$.

Comparison with AIRM (see Table 1) Coincidentally, the well-known affine invariant Riemannian metric (AIRM) distance (Pennec et al., 2006) obtained by taking the Riemannian trace metric in PD with length element $\mathrm{d}s_Q^2(\mathrm{d}Q) = \mathrm{tr}\left((Q^{-1}\mathrm{d}Q)\right)$ at $Q \in \mathrm{PD}$ was calculated in the same paper of James (1973). The AIRM distance between $Q_1, Q_2 \in \mathrm{PD}(n)$, computed via $\rho(Q_1, Q_2) = \sqrt{\sum_{i=1}^n \log^2 \lambda_i(Q_1 Q_2^{-1})}$, requires the full set of eigenvalues. Hilbert VPM distance requires only four extreme eigenvalues, which might make it of interest in applications like in Diffusion Tensor Imaging (Arsigny et al., 2006).

Table 1: Comparison of the AIRM vs Hilbert VPM distances. By $\mathrm{Mob}(Q_1, Q_2)$ we denote the Möbius transformation $\mathrm{Mob}(Q_1, Q_2) = (I - Q_1)^{-1}(I - Q_2)$.

	AIRM distance	Hilbert VPM distance
Eigenvalues:	$\{\lambda_i(Q_1 Q_2^{-1})\}_{1 \le i \le n}$	$\lambda_1(Q_1 Q_2^{-1}), \lambda_n(Q_1 Q_2^{-1})$ $\lambda_1(\mathrm{Mob}(Q_1, Q_2)), \lambda_n(\mathrm{Mob}(Q_1, Q_2))$
Invariance under a map:	$X \mapsto X^{-1}$	$X \mapsto I - X$
Invariance under congruence:	$\mathrm{GL}(n)$	$O(n)$



(a) (b)

Figure 2: Left: boundary of the 3D bicone with $\overline{\mathrm{VPM}}_\epsilon(2)$ (in blue) encapsulating $\overline{\mathrm{VPM}}(2)$ (in black). Right: fine approximation of the smallest enclosing ball (SEB, in red) w.r.t. the Hilbert distance on VPM(2) of a set of 2D SPD matrices (in black).

Hilbert distance on the extended Gaussians So far, we have only considered the Hilbert distance on the open set VPM, that is, full-rank covariance/precision matrices. However, we observe that we could enlarge $\overline{\mathrm{VPM}}$ by defining $\overline{\mathrm{VPM}}_\epsilon = \{-\epsilon\, I \preceq X \preceq (1+\epsilon)\, I\}$ for $\epsilon \geq 0$ (see Figure 2(a)). Let $d_{H,\epsilon}$ denote the Hilbert distance induced by $\overline{\mathrm{VPM}}_\epsilon$. Hilbert distances defined on nested VPM_ϵ domains satisfy the following monotonicity property (Beardon, 1999; Nielsen, 2020): for all $\epsilon > 0$ and all $P, Q \in \mathrm{VPM}(n)$, we have $d_H(P,Q) \geq d_{H,\epsilon}(P,Q)$. Even if $\forall P, Q \in \partial\overline{\mathrm{VPM}}$, $d_H(P,Q) = +\infty$, we have $d_{H,\epsilon}(P,Q) < +\infty$, resulting in a practically useful bounded distance between degenerate covariance/precision matrices.

Non-centered Gaussians The paper concerns the centered Gaussians, but we may embed (Calvo and Oller, 2002) the full Gaussian family $\{N(\mu,\Sigma) : \mu \in \mathbb{R}^n, \Sigma \in \mathrm{PD}(n)\}$ by:

$$(\mu,\Sigma) \mapsto \Sigma_\mu^+ = \begin{bmatrix} \Sigma + \mu\mu^T & \mu \\ \mu^T & 1 \end{bmatrix} \in \mathrm{PD}(n+1) \hookrightarrow \mathrm{VPM}(n+1)$$

Smallest enclosing ball Having straight lines being geodesics allows one to implement easily computational geometric primitives. For example, we may extend the Badoiu and Clarkson iterative geodesic-cut algorithm for approximating the smallest enclosing ball (Bâdoiu and Clarkson, 2003; Arnaudon and Nielsen, 2013) (SEB). Figure 2(b) shows a fine approximation of the Hilbert VPM SEB which we implemented as a proof-of-concept.

Correlation matrices We may also consider the VPM subdomain of correlation matrices, potentially degenerate. For the symmetric positive-definite cone $\mathrm{PD}(n)$, the subdomain of correlation matrices is called an elliptope, an open bounded convex domain whose simplicial facial structure was studied in Tropp (2018) and corresponding Hilbert geometry investigated in Nielsen and Sun (2018).

Supplementary material We provide supplementary material, code, and visualizations at `https://franknielsen.github.io/ExtendedGaussianGeometry/` and maintain an extended version of the paper on arXiv:2508.14369.

References

Marc Arnaudon and Frank Nielsen. On approximating the Riemannian 1-center. *Computational Geometry*, 46(1):93–104, 2013.

Vincent Arsigny, Pierre Fillard, Xavier Pennec, and Nicholas Ayache. Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 56(2): 411–421, 2006.

Mihai Bâdoiu and Kenneth L Clarkson. Smaller core-sets for balls. In *Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 801–802, 2003.

Hridhaan Banerjee, Carmen Isabel Day, Megan Hunleth, Sarah Hwang, Auguste H Gezalyan, Olya Golovatskaia, Nithin Parepally, Lucy Wang, and David M Mount. On the Heine-Borel property and minimum enclosing balls. *arXiv preprint arXiv:2412.17138*, 2024.

Alan F. Beardon. The Klein, Hilbert and Poincaré metrics of a domain. *Journal of computational and applied mathematics*, 105(1-2):155–162, 1999.

Garrett Birkhoff. Extensions of Jentzsch's theorem. *Transactions of the American Mathematical Society*, 85(1):219–227, 1957.

Miquel Calvo and Josep M Oller. A distance between elliptical distributions based in an embedding into the Siegel group. *Journal of Computational and Applied Mathematics*, 145(2):319–334, 2002.

Yongxin Chen, Tryphon T Georgiou, and Michele Pavon. Stochastic control liaisons: Richard Sinkhorn meets Gaspard Monge on a Schrödinger bridge. *Siam Review*, 63(2): 249–313, 2021.

Pierre de la Harpe. On Hilbert's Metric for Simplices. In Graham A. Niblo and Martin A. Roller, editors, *Geometric Group Theory*, volume 1 of *London Mathematical Society Lecture Note Series*, pages 97–119. Cambridge University Press, Cambridge, 1993. ISBN 978-0-521-43529-1. doi: 10.1017/CBO9780511661860.009.

Andrew Gelman, John B Carlin, Hal S Stern, and Donald B Rubin. *Bayesian data analysis*. Chapman and Hall/CRC, 1995.

Auguste H. Gezalyan and David M. Mount. Voronoi Diagrams in the Hilbert Metric. In Erin W. Chambers and Joachim Gudmundsson, editors, *39th International Symposium on Computational Geometry, SoCG 2023, June 12-15, 2023, Dallas, Texas, USA*, volume 258 of *LIPIcs*, pages 35:1–35:16. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2023. doi: 10.4230/LIPICS.SOCG.2023.35. URL https://doi.org/10.4230/LIPIcs.SoCG.2023.35.

Auguste H Gezalyan, Soo H Kim, Carlos Lopez, Daniel Skora, Zofia Stefankovic, and David M Mount. Delaunay Triangulations in the Hilbert Metric. In *19th Scandinavian Symposium on Algorithm Theory*, 2024.

William M Goldman. *Geometric structures on manifolds*, volume 227. American Mathematical Society, 2022.

M. Seetharama Gowda, Roman Sznajder, and Jiyuan Tao. The automorphism group of a completely positive cone and its Lie algebra. *Linear Algebra and its Applications*, 438 (10):3862–3871, May 2013. ISSN 0024-3795. doi: 10.1016/j.laa.2011.10.006.

David Hilbert. Über die gerade linie als kürzeste verbindung zweier punkte: Aus einem an herrn f. klein gerichteten briefe. *Mathematische Annalen*, 46(1):91–96, 1895.

Alan Treleven James. The variance information manifold and the functions on it. In *Multivariate Analysis–III*, pages 157–169. Elsevier, 1973.

Bas Lemmens and Roger Nussbaum. Birkhoff's version of Hilbert's metric and its applications in analysis. *Handbook of Hilbert Geometry*, pages 275–303, 2014a.

Bas Lemmens and Roger D. Nussbaum. Birkhoff's version of Hilbert's metric and its applications in analysis. In Athanase Papadopoulos and Marc Troyanov, editors, *Handbook of Hilbert geometry*, pages 275–303. EMS press, 2014b.

Bas Lemmens and Cormac Walsh. Isometries of polyhedral Hilbert geometries. Technical Report arXiv:0904.3306, April 2009.

Cyrus Mostajeran, Nathaël Da Costa, Graham Van Goffrier, and Rodolphe Sepulchre. Differential geometry with extreme eigenvalues in the positive semidefinite cone. *SIAM Journal on Matrix Analysis and Applications*, 45(2):1089–1113, 2024.

Frank Nielsen. The Siegel–Klein disk: Hilbert geometry of the Siegel disk domain. *Entropy*, 22(9):1019, 2020.

Frank Nielsen and Laetitia Shao. On balls in a Hilbert polygonal geometry. In *33rd International Symposium on Computational Geometry (SoCG)*, pages 67–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2017.

Frank Nielsen and Ke Sun. Clustering in Hilbert's projective geometry: The case studies of the probability simplex and the elliptope of correlation matrices. In *Geometric Structures of Information*, pages 297–331. Springer, 2018.

Frank Nielsen and Ke Sun. Non-linear embeddings in Hilbert simplex geometry. In *Topological, Algebraic and Geometric Learning Workshops 2023*, pages 254–266. PMLR, 2023.

Roger D Nussbaum. *Hilbert's projective metric and iterated nonlinear maps*, volume 1. American Mathematical Soc., 1988.

Atsumi Ohara. Doubly autoparallel structure on positive definite matrices and its applications. In *International Conference on Geometric Science of Information (GSI)*, pages 251–260. Springer, 2019.

Atsumi Ohara, Nobuhide Suda, and Shun-ichi Amari. Dualistic differential geometry of positive definite matrices and its applications to related problems. *Linear Algebra and Its Applications*, 247:31–53, 1996.

Xavier Pennec, Pierre Fillard, and Nicholas Ayache. A Riemannian framework for tensor computing. *International Journal of computer vision*, 66(1):41–66, 2006.

Bernard Shiffman. Synthetic projective geometry and Poincaré's theorem on automorphisms of the ball. 1995. doi: 10.5169/SEALS-61825.

Dario Stein and Richard Samuelson. A Category for Unifying Gaussian Probability and Nondeterminism. In *10th Conference on Algebra and Coalgebra in Computer Science (CALCO 2023)*, pages 13–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2023.

Viktor Stein. Loewner order in terms of eigenvalues. Mathematics Stack Exchange. URL https://math.stackexchange.com/q/4606814. URL:https://math.stackexchange.com/q/4606814 (version: 2022-12-28).

Yann Thanwerdas and Xavier Pennec. $O(n)$-invariant Riemannian metrics on SPD matrices. *Linear Algebra and its Applications*, 661:163–201, 2023.

Joel A Tropp. Simplicial faces of the set of correlation matrices. *Discrete & Computational Geometry*, 60(2):512–529, 2018.

Julian Vanecek, Auguste H Gezalyan, and David M Mount. Support Vector Machines in the Hilbert Geometry. In *31st Annual Fall Workshop onComputational Geometry*, 2024. URL https://www.cs.tufts.edu/research/geometry/FWCG24/program2.html.

Cormac Walsh. Gauge-reversing maps on cones, and Hilbert and Thompson isometries. *Geometry & Topology*, 22(1):55–104, October 2017. ISSN 1364-0380, 1465-3060. doi: 10.2140/gt.2018.22.55.

Jan C Willems. Open stochastic systems. *IEEE Transactions on Automatic Control*, 58(2): 406–421, 2012.

Appendix A. Deferred Proofs

Proof [Proof of Proposition 3] For the forward direction, following Stein, assume that $A \preceq B$, that is, $0 \preceq B - A$. Then from the min-max theorem, we have:

$$\lambda_i(A) = \min_{U \subset \mathbb{R}^n; \dim(U)=i} \max_{x \in U} \frac{v^T A v}{\|v\|_2^2} \leq \min_{U \subset \mathbb{R}^n; \dim(U)=i} \max_{x \in U} \frac{v^T B v}{\|v\|_2^2} = \lambda_i(B)$$

In the other direction, we assume $A = Q^T \operatorname{diag}(\alpha_1, \dots, \alpha_n)\, Q$ and $B = Q^T \operatorname{diag}(\beta_1, \dots, \beta_n)\, Q$ in the common eigenbasis given by Q. If $\alpha_i \leq \beta_i$, then we have $(B - A)q_i = (\beta_i - \alpha_i)$, proving the converse. ∎

Proof [Proof of Proposition 12] $\mathrm{PD}(n)$ is convex and open in $\mathrm{Sym}(n)$, and $\mathrm{VPM}(n) = \mathrm{PD}(n) \cap (I - \mathrm{PD}(n))$, i.e., an intersection of two open convex sets is open and convex. Moreover, all norms in a finite-dimensional vector space are equivalent, which means we only have to show boundedness in one norm. This means that $0 \preceq A \preceq I$ implies that:

$$0 \preceq A^2 \preceq A \preceq I$$

and therefore:

$$\|A\|_F^2 = \mathrm{tr}(A^T A) \leq \mathrm{tr}(A) \leq n.$$

■

Proof [Proof of Proposition 13] We use the characterization of the Loewner order in terms of eigenvalues, as in Proposition 3. For the first part, observe that for each eigenvalue λ of X, the inequality $0 < \lambda < 1$ implies $0 < 1 - \lambda < 1$. For the second part, spectrum is invariant under conjugation with $O(n)$.

■

Proof [Proof of Theorem 14]

VPM(n) is an open bounded convex set (Proposition 12), therefore VPM(n) can be equipped with the Hilbert distance. To prove the theorem, we employ Birkhoff's characterization of the Hilbert metric, using the Loewner order $\preceq$ (see Definition 2).

We consider the affine map $\hat{(\cdot)} : \mathrm{VPM}(n) \to \mathrm{PD}(n) \times \mathrm{PD}(n)$ given by $\hat{A} = (A, I - A)$. The image of this map is a convex set:

$$C = \{(X, Y) \in \mathrm{PD}(n) \times \mathrm{PD}(n) : X + Y = I\}$$

On the set C, we build the cone

$$K(C) = \{(\lambda X, \lambda Y, \lambda) : (X, Y) \in C, \lambda \in \mathbb{R}_{>0}\} \subset \mathrm{Sym}(n) \times \mathrm{Sym}(n) \times \mathbb{R}_{>0}$$

By Proposition 9, we have the equality:

$$d_H(A, B) = d_H(\hat{A}, \hat{B}) = d_B((\hat{A}, 1), (\hat{B}, 1)) = \log \frac{M\left((\hat{A}, 1), (\hat{B}, 1)\right)}{m\left((\hat{A}, 1), (\hat{B}, 1)\right)}$$

where d_H is the Hilbert metric on VPM(n) equal to the Hilbert metric on its affine image C, and d_B is the Birkhoff metric on $\mathbf{P}(K(C))$, which we obtain here by considering the projective slice $[\hat{X} : 1]$. We first compute the numerator M:

$$\begin{aligned} M\left((\hat{A}, 1), (\hat{B}, 1)\right) &= \inf\left\{\lambda > 0 : (\hat{A}, 1) \preceq_{K(C)} \lambda(\hat{B}, 1)\right\}, \\ &= \inf\left\{\lambda > 0 : \lambda(\hat{B}, 1) - (\hat{A}, 1) \in K(C)\right\}, \\ &= \inf\left\{\lambda > 0 : \exists_{Z \in \mathrm{VPM}(n)} \exists_{\theta \in \mathbb{R}_{>0}} \text{ s.t. } \lambda(\hat{B}, 1) - (\hat{A}, 1) = (\theta \hat{Z}, \theta)\right\}. \end{aligned}$$

The last equality can only happen if $\theta = \lambda - 1$, and therefore $\lambda \hat{B} - \hat{A} = (\lambda - 1)\hat{Z}$, meaning:

$$\lambda B - A = (\lambda - 1) Z, \qquad \lambda(I - B) - (I - A) = (\lambda - 1)(I - Z),$$

where the second equality is equivalent to the first one. Since $0 \prec Z \prec I$, we arrive at two inequalities:

$$A \prec \lambda B, \qquad (I - A) \prec \lambda(I - B)$$

Performing an analogous computation for $m\left((\hat{A},1),(\hat{B},1)\right)$, we get:

$$M(\hat{A},\hat{B}) = \inf_{\lambda>0}\{A \preceq \lambda B \text{ and } (I-A) \preceq \lambda(I-B)\},$$
$$m(\hat{A},\hat{B}) = \sup_{\mu>0}\{\mu B \preceq A \text{ and } \mu(I-B) \preceq I-A\}.$$

Still focusing on $M(\hat{A},\hat{B})$, if $A \preceq \lambda B$, then $A^{-1/2}AA^{-1/2} \preceq \lambda A^{-1/2}BA^{-1/2}$. The converse in Proposition 3 proves that this holds if and only $\frac{1}{\lambda} \leq \lambda_{\min}(A^{-1}B) = \lambda_{\min}(A^{-1/2}BA^{-1/2})$, and therefore:

$$\lambda \geq \frac{1}{\lambda_{\min}(A^{-1}B)} = \lambda_{\max}(B^{-1}A)$$

Similarly, applied to $(I-A) \preceq \lambda(I-B)$, we get $\lambda \geq \lambda_{\max}((I-B)^{-1}(I-A))$. Taking the infimum over λ, we obtain:

$$M(\hat{A},\hat{B}) = \max\left(\lambda_{\max}(B^{-1}A), \lambda_{\max}((I-B)^{-1}(I-A))\right).$$

The case of $m(\hat{A},\hat{B})$ follows in the same way. If $\mu I \leq B^{-1}A$ and $\mu I \leq (I-B)^{-1}(I-A)$, then:

$$\mu \leq \lambda_{\min}(B^{-1}A), \qquad \mu \leq \lambda_{\min}((I-B)^{-1}(I-A)),$$

and so taking the infimum:

$$m(\hat{A},\hat{B}) = \min\left(\lambda_{\min}(B^{-1}A), \lambda_{\min}((I-B)^{-1}(I-A))\right).$$

Substituting it in the original formula, we get the claimed formula:

$$d_H(A,B) = \log\frac{\max(\lambda_{\max},\mu_{\max})}{\min(\lambda_{\min},\mu_{\min})},$$

for $\lambda_{\min,\max}$ the minimal and maximal eigenvalues of $B^{-1}A$, and $\mu_{\min,\max}$ the minimal and maximal eigenvalues of $(I-B)^{-1}(I-A)$. ■

Proof [Proof of Remark 18] We compute:

$$\begin{aligned}(\iota^{-1}\circ(I-\cdot)\circ\iota)(X) &= \iota^{-1}(I-\iota(X)) \\ &= \iota^{-1}(I-X(I+X)^{-1}) \\ &= (I-X(I+X)^{-1})(I-I+X(I+X)^{-1})^{-1} \\ &= (I-X(I+X)^{-1})(I+X)X^{-1} \\ &= ((I+X)X^{-1}-I) \\ &= X^{-1}(I+X-X) = X^{-1}\end{aligned}$$

■

Proof [Proof of Proposition 19]

First, we note that eigenvalues are invariant under conjugation, and conjugation with an orthonormal matrix $U \in O(n)$ is a congruence, which preserves symmetry, therefore conjugation $X \mapsto U^T XU = U^{-1}XU$ is an automorphism of $\mathrm{VPM}(n)$. Given two matrices $X, Y \in \mathrm{VPM}(n)$ and an orthonormal matrix U, we write $A = U^{-1}XU$ and $B = U^{-1}YU$ and calculate:

$$d_H(A,B) = \log \frac{\max(\lambda_{\max}(A^{-1}B), \lambda_{\max}((I-A)^{-1}(I-B)))}{\min(\lambda_{\min}(A^{-1}B), \lambda_{\min}((I-A)^{-1}(I-B)))}$$

But then:

$$A^{-1}B = (U^{-1}XU)^{-1}(U^{-1}YU) = (U^{-1}X^{-1}U)(U^{-1}YU) = U^{-1}(X^{-1}Y)U$$

and again, the eigenvalues are invariant under conjugation, so $\lambda_{\min,\max}\left(A^{-1}B\right) = \lambda_{\min,\max}\left(X^{1-}Y\right)$. We also observe that:

$$\begin{aligned}(I-A)^{-1}(I-B) &= (I-Q^{-1}XQ)^{-1}(I-Q^{-1}YQ) \\ &= (Q^{-1}(I-X)Q)^{-1}(Q^{-1}(I-Y)Q) \\ &= Q^{-1}(I-X)^{-1}(I-Y)Q\end{aligned}$$

and therefore the same argument applies here. Thus, the Hilbert distance is preserved. ■

Proof [Proof of Remark 20]
We compute:

$$\begin{aligned}U^T\iota(X)U &= U^TX(I+X)^{-1}U \\ &= (U^TXU)\left(U^T(I+X)^{-1}U\right) \\ &= (U^TXU)\left((U^T(I+X)U)^{-1}\right) \\ &= (U^TXU)\left((I+U^TXU)^{-1}\right) \\ &= \iota(U^TXU)\end{aligned}$$

and similarly for $U^T\iota^{-1}(X)U = \iota^{-1}(U^TXU)$. Thus:

$$(\iota^{-1} \circ (U^T \cdot U) \circ \iota)(X) = U^TXU$$

■

Proof [Proof of Proposition 30]
We form the dual cone:

$$C_n^* = \{(Y,s) \in \mathrm{Sym}(n) \times \mathbb{R} : \langle (X,t),(Y,s)\rangle > 0 \text{ for all } (X,t) \in C\}$$

where the inner product is induced on the product space as:

$$\langle (X,t),(Y,s)\rangle = \langle X,Y\rangle + ts = \mathrm{tr}(XY) + ts$$

We denote $Z = \frac{1}{t}X$, where $Z \in \mathrm{VPM}(n)$ and $t > 0$ are now arbitrary, and write:

$$t(\mathrm{tr}(ZY) + s) > 0$$

which is equivalent to $\mathrm{tr}(ZY)+s > 0$. Because $\mathrm{VPM}(n)$ is invariant under conjugation with orthonormal matrices as Proposition 19 showed, we can diagonalize $Y = Q^T \mathrm{diag}(\lambda_1, \dots, \lambda_n)Q$ and compute the infimum:

$$\begin{aligned}
\inf_{Z\in\mathrm{VPM}(n)} \mathrm{tr}(ZY) &= \inf_{Z\in\mathrm{VPM}(n)} \sum_{i=1}^{n} \lambda_i q_i^T Z q_i \\
&\geq \inf_{Z\in\mathrm{VPM}(n)} \sum_{\lambda_i<0} \lambda_i q_i^T Z q_i \\
&\geq \sum_{\lambda_i<0} \lambda_i = -\mathrm{tr}(Y_-)
\end{aligned}$$

and the infimum is achieved by letting $Z \to Q^T \mathrm{diag}(\mathbb{1}_{\lambda_1<0}, \dots, \mathbb{1}_{\lambda_n<0})Q$ (which is attained in the closure of $\mathrm{VPM}(n)$ in $\mathrm{Sym}(n)$). This means that the sufficient and necessary condition defining the dual cone is:

$$C_n^* = \{(Y, s) \in \mathrm{Sym}(n) \times \mathbb{R} : s > \mathrm{tr}(Y_-)\}$$

which is clearly different from $\mathrm{VPM}(n)$, since it does not even depend on the positive eigenvalues of Y. ■

Proof [Proof of Lemma 37]

For any $t \in \mathbb{R}_{>0}$, we have $\frac{\eta(t)}{t} \in K$, because K is a cone. Since K is closed, for $t > 0$, we also have

$$\lim_{t\to 0^+} \frac{\eta(t)}{t} = \lim_{t\to 0^+} \frac{\eta(t) - \eta(0)}{t} = \eta'(0) \in K$$

By a similar argument, this time taking $t < 0$, we have:

$$\lim_{t\to 0^-} \frac{\eta(t)}{-t} = -\eta'(0) \in K$$

But K is pointed, so $K \cap (-K) = \{0\}$, and therefore $\eta'(0) = 0$. ■

Proof [Proof of Lemma 38] A general affine isomorphism $L : \mathrm{Sym}(n) \to \mathrm{Sym}(n)$ has a form $LX = \mathcal{A}X + B$ for some linear isomorphism $\mathcal{A} : \mathrm{Sym}(n) \to \mathrm{Sym}(n)$ and $B \in \mathrm{Sym}(n)$. Since L preserves $\mathrm{VPM}(n)$ and is smooth, it also maps the closure $\overline{\mathrm{VPM}(n)}$, and thus the $\partial\,\mathrm{VPM}(n) = \overline{\mathrm{VPM}(n)} \setminus \mathrm{VPM}(n)$ to itself $L(\partial\mathrm{VPM}(n)) = \partial\mathrm{VPM}(n)$.

Let us assume that $L(X) = 0$ for some $X \in \overline{\mathrm{VPM}(n)}$ and $X \notin \{0, I\}$. First, we consider a case where $X = \alpha I$ for some $0 < \alpha < 1$. Take some $0 < \beta < \min(\alpha, 1 - \alpha)$, therefore $\alpha I \pm \beta I \in \mathrm{VPM}(n)$, and compute:

$$L(\alpha I \pm \beta I) = \mathcal{A}(\alpha I \pm \beta I) + B = (\mathcal{A}(\alpha I) + B) \pm \mathcal{A}\beta I = L(\alpha I) \pm \mathcal{A}\beta I = \pm\mathcal{A}\beta I \in \mathrm{VPM}(n)$$

But having both $\mathcal{A}I \prec 0$ and $0 \prec \mathcal{A}I$ is impossible.

Next, assume there exist at least two distinct eigenvalues $\lambda_1 < \lambda_2$ of X, with corresponding unit eigenvectors v_1, v_2. Let U_θ for $\theta \in (0, 2\pi]$ be a rotation matrix which rotates the subspace $V = \operatorname{span}(v_1, v_2)$ by angle θ, that is:

$$U_\theta = \exp(\theta Z) \qquad Z = v_1 v_2^T - v_2 v_1^T$$

Since $U_\theta \in O(n)$, both $\mathrm{Sym}(n)$ and $\mathrm{PSD}(n)$ are invariant under conjugation with U_θ.

We thus have a smooth curve $\eta : S^1 \to \mathrm{Sym}(n)$ given by $\eta(\theta) = L(U_\theta^T X U_\theta)$ lying inside $\overline{\mathrm{VPM}(n)} \subset \mathrm{PSD}(n)$, such that for $\theta = 0$, we have $\eta(0) = 0$. This means that $\eta'(0) = 0$, by Lemma 37. By standard computation, we have:

$$\frac{d}{d\theta} U_\theta^T X U_\theta = U_\theta^T (XZ - ZX) U_\theta$$

and therefore:

$$\frac{d}{d\theta}\eta(\theta) = \mathcal{A}\left(U_\theta^T (XZ - ZX) U_\theta\right)$$

Thus, $\mathcal{A}(XZ - ZX) = 0$, meaning $XZ - ZX = 0$. Unfolding:

$$X(v_1v_2^T - v_2v_1^T) - (v_1v_2^T - v_2v_1^T)X = \lambda_1 v_1 v_2^T - \lambda_2 v_2 v_1^T - \lambda_2 v_1 v_2^T + \lambda_1 v_2 v_1^T = (\lambda_1 - \lambda_2)(v_1 v_2^T + v_2 v_1^T)$$

But if $v_1 v_2^T + v_2 v_1^T$ is a non-zero symmetric matrix, implying $\lambda_1 = \lambda_2$, a contradiction.

Thus, we can only have $L(0) = 0$ or $L(I) = 0$. Let us compute:

$$L(0) + L(I) = \mathcal{A}0 + B + \mathcal{A}I + B = \mathcal{A}X + B + \mathcal{A}(I - X) + B = L(X) + L(I - X)$$

Since L is onto $\mathrm{VPM}(n)$ and a diffeomorphism, it is also onto $\overline{\mathrm{VPM}(n)}$, so for each $Y = L(X)$, we also have $L(I - X) = C - L(X) \in \overline{\mathrm{VPM}(n)}$, for $C := L(0) + L(1)$. Thus, substituting $Y = 0$, we get $C \in \overline{\mathrm{VPM}(n)}$, i.e. $C \preceq I$, and substituting $Y = I$, we get $C - I \in \overline{\mathrm{VPM}(n)}$, i.e. $I \preceq C$, forcing $I = C$. Thus, $L(0) + L(I) = I$, and if $L(0) = 0$, then $L(I) = I$. Otherwise, $L(0) = I$ and $L(I) = 0$, proving the thesis. ∎

Proof [Proof of Theorem 39]

We will prove that given a linear cone automorphism $L : \mathrm{Sym}(n) \times \mathbb{R} \to \mathrm{Sym}(n) \times \mathbb{R}$ for C_n, there exists $P \in O(n)$ and $\epsilon \in \{0, 1\}$ such that:

$$[L(X, t)] = [P^T((1 - \epsilon)X + \epsilon(I - X))P : 1]$$

from which the theorem follows immediately.

A linear cone automorphism $L : \mathrm{Sym}(n) \times \mathbb{R} \to \mathrm{Sym}(n) \times \mathbb{R}$ is, in general, of a form:

$$L(X, t) = (\mathcal{A}X + tB, \langle C, X\rangle + st) = (\mathcal{A}X + tB, \operatorname{tr}(CX) + st)$$

for some linear $\mathcal{A} : \mathrm{Sym}(n) \to \mathrm{Sym}(n)$ and $B, C \in \mathrm{Sym}(n)$ and $s \in \mathbb{R}$. Because L maps the origin $(0, 0)$ to itself, and maps straight lines to straight lines, it has to map rays through the origin to rays through the origin. Thus, it is an isomorphism on each projective slice. That is, for each fixed $t_0 \in \mathbb{R}_{>0}$, we get an affine isomorphism $\pi \circ L(\cdot, t_0) \circ \iota_{t_0} : \mathrm{Sym}(n) \to$

$\mathrm{Sym}(n)$, where $\pi(X,t) = \frac{1}{t}X$ is the projectivization and $\iota_{t_0}(X) = (X, t_0)$ is an inclusion. After applying Lemma 38, we get that either $L(I,t) = (I,t')$ and $L(0,t) = L(0,t'')$ or $L(I,t) = (0,t')$ and $L(0,t) = (I,t'')$ for some $t', t'' \in \mathbb{R}_{>0}$. By possibly precomposing with a map $(X,t) \mapsto (It - X, t)$, we can ensure that $L(0,\mathbb{R}_{>0}) = (0,\mathbb{R}_{>0})$ and $L(I,\mathbb{R}_{>0}) = (I,\mathbb{R}_{>0})$. Setting $X = 0$, we have:

$$L(0,t) = (\mathcal{A}0 + tB, \mathrm{tr}(C0) + st) = (tB, st)$$

which means $B = 0$. Therefore, we can write simpler formula for the inverse:

$$L^{-1}(X,t) = \left(\mathcal{A}^{-1}X, -\frac{\mathrm{tr}(C\mathcal{A}^{-1}X)}{s} + \frac{t}{s}\right)$$

Next, substituting $X = I$, and remembering that $L(I,\mathbb{R}_{>0}) = (I,\mathbb{R}_{>0})$, we have:

$$L(I,t) = (\mathcal{A}I, \mathrm{tr}(CI) + st) = \left(I, -\mathrm{tr}(C) + \frac{t}{s}\right)$$
$$L^{-1}(I,t) = \left(\mathcal{A}^{-1}I, -\frac{\mathrm{tr}(C\mathcal{A}^{-1}I)}{s} + \frac{t}{s}\right) = \left(I, -\frac{\mathrm{tr}(C)}{s} + \frac{t}{s}\right)$$

From the fact that $\mathrm{VPM}(n)$ is mapped into itself, and the fact that L is a linear isomorphism and thus a diffeomorphism, we get that the closure $\overline{\mathrm{VPM}(n)}$, and thus the boundary $\mathrm{VPM}(n)\backslash\overline{\mathrm{VPM}(n)}$, is preserved by both L and L^{-1}. Therefore:

$$I \preceq (\mathrm{tr}(C) + s)I \implies 1 - s \leq \mathrm{tr}(C)$$
$$I \preceq \left(-\frac{\mathrm{tr}(C)}{s} + \frac{1}{s}\right)I \implies 1 - s \geq \mathrm{tr}(C)$$

where we also used the fact that $s \geq 0$, from the fact that L preserves the positivity of rays $(0,\mathbb{R}_{>0})$. Therefore, we conclude that $1 - s = \mathrm{tr}(C)$.

Now, we note that since we're interested only in the equivalence class $[L]$, we can assume w.l.o.g. that $s = 1$:

$$[L] = [\mathcal{A}X, \mathrm{tr}(CX) + st] = \left[\frac{\mathcal{A}}{s}X, \mathrm{tr}\left(\frac{C}{s}X\right) + t\right]$$

and this combined with the previous point, sets $\mathrm{tr}(C) = 0$. Finally, assume that C has at least one negative eigenvalue λ, with the unit eigenvector v. Set $X = vv^T$, and compute:

$$\mathrm{tr}(CX) = \mathrm{tr}(Cvv^T) = \mathrm{tr}(\lambda vv^T) = \mathrm{tr}(\lambda vv^T I) = \lambda\mathrm{tr}(v^T v) = \lambda||v||_2 = \lambda < 0$$

Now, pick some $0 < t_0 < -\lambda$, and compute:

$$L(t_0X, t_0) = (t_0\mathcal{A}X, \lambda - t_0)$$

making the second component negative, which is a contradiction with the positivity of rays. This means that all eigenvalues λ_i^C of C must be non-negative, and thus:

$$0 = \mathrm{tr}(C) = \sum_i \lambda_i^C \implies \lambda_i^C = 0$$

So far, the above argument proved that:

$$[L(X,t)] = [\mathcal{A}X, t]$$

Now, using Lemma 36 and the fact that

$$L(\mathrm{PD}(n), \mathbb{R}_{>0}) = (\mathrm{PD}(n), \mathbb{R}_{>0})$$

which holds because every element of $Y \in \mathrm{PD}(n)$ can be scaled to a point in $\mathrm{VPM}(n)$ by a non-negative factor $\frac{1}{\lambda_{\max}(X)+1}$, we have $[L(X,t)] = [P^T XP, t]$. This proves the statement of the theorem, since we have been possibly precomposing L with $I - X$, and:

$$P^T(I-X)P = P^T P - P^T XP = I - P^T XP$$

which used the fact that $\mathcal{A}I = I$, i.e. that P must be orthonormal. In other words, the group of isometries of $\mathrm{VPM}(n)$ for $n > 1$ is generated by conjugation by orthonormal matrix $X \mapsto U^T XU$ and inversion $X \mapsto I - X$. ■