arXiv:2603.02483v1 [stat.ML] 3 Mar 2026

James对称正定矩阵双锥域上的几何结构与偏差

Jacek Karwowski* Frank Nielsen*计算机科学系 索尼计算机科学实验室
牛津大学，英国 东京，日本

摘要

对称正定（SPD）矩阵数据集在众多科学领域中发挥着核心作用，包括信号处理、统计学、金融、计算机视觉、信息论和机器学习等。对称正定矩阵集合构成一个锥，可视为底层 SPD 流形的全局坐标图。对称正定锥流形上可以定义丰富的微分几何结构。该流形上应用最广泛的几何框架包括仿射不变黎曼结构和对偶信息几何对数行列式障碍结构，它们分别与不相似性度量（距离和散度）相关。本文介绍两种新结构：一种芬斯勒结构和一种对偶信息几何结构，二者均源自 James 对对称正定域的双锥重参数化。这些结构确保测地线在相应的坐标系中对应直线。闭双锥域将谱单纯形（由迹为单位的正半定对角矩阵构成的集合）包含为一个仿射子空间，并且我们证明，希尔伯特 VPM 距离是希尔伯特单纯形距离的推广；后者已在机器学习中得到广泛应用。最后，我们讨论这些芬斯勒／对偶 Hessian 结构的若干应用，并给出新型与传统不相似性度量之间的多种不等式。

关键词：对称正定锥；SPD双锥；黎曼结构；对偶平坦结构；芬斯勒结构；Hilbert/Birkhoff距离；Bregman散度；谱单纯形；障碍函数；效应（量子信息论中的正算子值测量）；Riccati方程（控制理论）。

1 引言

令 Mat(n) 表示实系数 $n \times n$ 矩阵构成的向量空间，令 Sym(n) $\subset$Mat(n) 表示对称矩阵构成的向量子空间。我们在 Mat(n) 上赋予内积$\langle M_1,\ M_2\rangle = \mathrm{tr}(M_1^\top M_2)$，该内积诱导出 Frobenius范数 $\|M\|_F = \sqrt{\langle M,\ M\rangle}$ 以及相应的距离 $d_F(M_1,\ M_2) = \|M_1 - M_2\|_F$，并考虑其度量拓扑。令 PD($n$) $\subset$ Sym(n) 表示对称正定（SPD）矩阵构成的子集（锥），其特征值均为正；令 ND(n) $= -$ PD(n) 表示对称负定锥。SPD 矩阵数据集广泛存在于众多科学领域，从统计学和概率论（高斯分布）到信号处理（卡尔曼滤波）、医学成像 [33] （扩散张量成像）、优化 [8],和计算机视觉 [40, 44]，仅举几例。

*贡献相同。

本文旨在比较应用中常用的对称正定锥上的两种不相似性：即 §1.1 中介绍的仿射不变黎曼度量距离（简称 AIRM 距离）和 §1.2 中详述的对数行列式散度 [10] ），并将它们与两种基于对称正定锥双锥表示 [18] 的新型不相似性进行比较：

- 希尔伯特对称正定双锥距离 [20],以及
- 双锥logdet散度。

我们首先在 §1.1 和 §1.2 中回顾 AIRM 结构与对称正定锥上的信息几何 logdet 黑塞结构之间的联系，然后分别在 §2 和 §3 中描述并研究新的 Finsler 结构和对偶双锥logdet Hessian结构的性质，它们分别构成希尔伯特差异度和对数行列式差异度的基础。接着，我们在 §4 中比较 AIRM/对数行列式差异度与新的希尔伯特/双锥 logdet 差异度。最后，我们在 §5 中总结研究结果并讨论本文的展望。

我们的符号汇总见附录 A 的表 2。

1.1 黎曼几何：仿射不变黎曼度量

从微分几何的角度来看，定义域 PD(n) 可解释为维数为 $^{n}{}_{(n+1)\ 2}$ 的对称正定锥流形 $\mathcal{C}$ 的全局坐标图，其在 $p \in \mathcal{C}$ 处的切平面 $T_p\mathcal{C}$ 可识别为~~Sym~~(n)。在 $\mathcal{C}$上众多可能的黎曼度量 [42] 中，仿射不变黎曼度量 [16](AIRM) 在应用中十分常用，因为其黎曼距离$d_{\mathrm{AIRM}}(X_1, X_2)$ 和测地线 $\gamma^{\mathrm{AIRM}}_{X_1\ X_2,}(t)$ 均有闭式表达：

$$\begin{aligned} \mathrm{d}_{\mathrm{AIRM}}(X_1, X_2) &= \sqrt{\sum_{i=1}^{n} \log^2 \lambda_i(X_2^{-1} X_1)}, \\ \gamma^{\mathrm{AIRM}}_{X_1,X_2}(t) &= X_1^{\frac{1}{2}} \left(X_1^{-\frac{1}{2}} X_2 X_1^{-\frac{1}{2}} \right)^t X_1^{\frac{1}{2}}, \end{aligned}$$

并满足对合同变换和矩阵求逆的不变性：

$$\begin{aligned} \mathrm{d}_{\mathrm{AIRM}}(X_1, X_2) &= \mathrm{d}_{\mathrm{AIRM}}(AX_1A^\top, AX_2A^\top), \forall A \in \mathrm{GL}(n) \\ \mathrm{d}_{\mathrm{AIRM}}(X_1, X_2) &= \mathrm{d}_{\mathrm{AIRM}}(X_1^{-1}, X_2^{1}), \end{aligned}$$

其中 GL(n) 表示 Mat(n) 的一般线性群。

此外，AIRM 度量 g_P^{AIRM}与中心高斯族 [18, 39]的 Fisher信息度量仅相差一个比例因子：

$$g_X^{\mathrm{AIRM}}(S_1, S_2) = \mathrm{tr}\left(X^{-1}S_1X^{-1}S_2\right), \forall X \in \mathrm{PD}(n), \forall S_1, S_2 \in \mathrm{Sym}(n).$$

因此，AIRM 也称为迹度量 [8] ，相应的黎曼距离最早是在更一般的 Siegel 上半空间 [38] 中计算的，该空间由虚部为正定的复矩阵构成。

1.2 对偶信息几何：Hessian 度量

在信息几何 [2, 37], 中，定义在配备无挠平坦联络 ∇ 的 n维仿射流形 $(\mathcal{C},\nabla)$ 上的任意势函数 ϕ 都会确定一个黎曼度量 $g=\nabla d\phi$，其中 d 表示外微分。坐标系 $\theta(\cdot)$ 在 $\mathcal{C}$上给出向量场 $\left\{\partial_1=\frac{\partial}{\partial\theta_1},\right.$ ....., $\left.\partial_n=\frac{\partial}{\partial\theta_n}\right\}$。若某坐标系满足该联络 ∇ 的克里斯托费尔符号$\Gamma_{ij}^k(\theta)$ 均为零，则称其为 ∇-仿射坐标系 [37]。对 θ作仿射变换 $\theta=A\theta+b$，其中 $A\in$ 属于 GL(n) 且 $b\in\mathbb{R}^n$，可得到另一个 ∇-仿射坐标系。度量 g 可用局部坐标 θ 表示为 $G(\theta)=[G_{ij}(\theta)]$，其中 $G_{ij}(\theta)=\frac{\partial^2\Psi(\theta)}{\partial\theta^i\partial\theta^j}$（简写为 $G_{ij}=\partial_i\partial_j\Psi$），且 $\phi(p)=\Psi(\theta(p))$（即 Ψ 是用 θ坐标系表示的势函数）。结构 (∇,g) 称为 $\mathcal{C}$上的 Hessian 结构 [37]，并由此得到另一个对偶 Hessian 结构 (∇^*,g)，其中 ∇^* 满足 $\frac{\nabla+\nabla^*}{2}=\nabla^g$，这里 ∇^g 为 Levi-Civita 度量联络。与度量 g 相配的对偶联络 ∇^* 由 Legendre对偶势函数 ϕ^*诱导，即 $g=\nabla^*d\phi^*$ 在局部坐标 η下表示为 $G^{ij}(\eta)=\frac{\partial^2\Psi^*(\eta)}{\partial\eta^i\partial\eta^j}$，其中 $\phi^*(p)=\Psi^*(\eta(p))$。设 $\{\partial^1=\frac{\partial}{\partial\eta_l},\ldots,\partial^n=\frac{\partial}{\partial\eta_n}\}$为由 ∇^*-仿射坐标系 η诱导的向量场。（一旦 ∇-仿射坐标系在仿射变换意义下选定，对偶的 ∇-仿射坐标系便唯一确定。）于是$G^{ij}(\eta)=\partial^i\partial^j\Psi^*(\eta)$，且对偶坐标系彼此正交：$G(\theta)G^(\eta)=I_{n,n}$，单位矩阵。

对偶势函数 ϕ 和 ϕ^* 组合成典范散度 $\mathcal{D}(p,q)=\phi(p)+\phi^*(q)-\sum_{i=1}^n\theta^i(p)\eta_i(q)$，它也可以等价地表示为对偶Bregman散度 [6]，分别对应对偶的 ∇-仿射和 ∇^*-仿射坐标系 $\theta(\cdot)$ 和 $\eta(\cdot)$。对称正定锥流形 $\mathcal{C}$ 上的这种对偶平坦信息几何在 [32]中得到了研究，其中 $\theta(p(X))=X$且 $\eta(p(X))=-X^{-1}$。$\mathcal{C}$ 上的对偶势函数为 $\phi(p)=-\log\det P(\theta(p))$ 和 $\phi^*(p)=-\log\det P(\eta)-n$，相应的对偶Bregman散度为 $B_\Psi(P_1:P_2)=\log\det(P_2P_1^{-1})+\mathrm{tr}(P_2^{-1}P_1)-n$，这是在 $\Psi(\theta)=-\log\det\theta$（通常称为 log-det 散度 [10]，有时也称为 Burg矩阵散度 [12] 或 Itakura-Saito矩阵散度 [11]）时得到的；以及 $B_{\Psi^*}(P_2^{-1}:P_1^{-1})=B_F(P_1:P_2)$，这是在 $\Psi^*(\eta)=-\log\det\eta-n$时得到的。Hessian势函数$\nabla^2\Psi(\theta)$ 和 $\nabla^2\Psi^*(\eta)$ 诱导出黎曼度量 g^Ψ，该度量与仿射不变黎曼度量的迹度量一致。尽管 logdet（ld）散度常用于信息几何，文献中也研究了 $\mathcal{C}$ 上的其他对偶势函数 [45, 30, 1]。

1.3 James 双锥

James 考虑了以下两个将正定矩阵(n)映射到 VPM 双锥的微分同胚映射：

$$\mathrm{VPM}^\circ(n)=\{X\in\mathrm{PD}(n):0\prec X\prec I\},$$

其中 I 表示单位矩阵：

$$\begin{aligned}V(X)&=X(I+X)^{-1},\quad V^{-1}(X)=X(I-X)^{-1}\\P(X)&=(I+X)^{-1},\quad P^{-1}(X)=X^{-1}-I.\end{aligned}$$

我们有特征值 $\lambda_i(V(X))\in(0,1)$ 和 $\lambda_i(P(X))\in(0,1)$，即 $V(X)\in\mathrm{VPM}^\circ(n)$ 和 $P(X)\in\mathrm{VPM}^\circ(n)$ 均成立。

命题 1 特征值 $\lambda_i(V(X))\in(0,1)$ 和 $\lambda_i(P(X))\in(0,1)$，即 $V(X)\in\mathrm{VPM}^\circ(n)$ 和 $P(X)\in\mathrm{VPM}^\circ(n)$均如此。

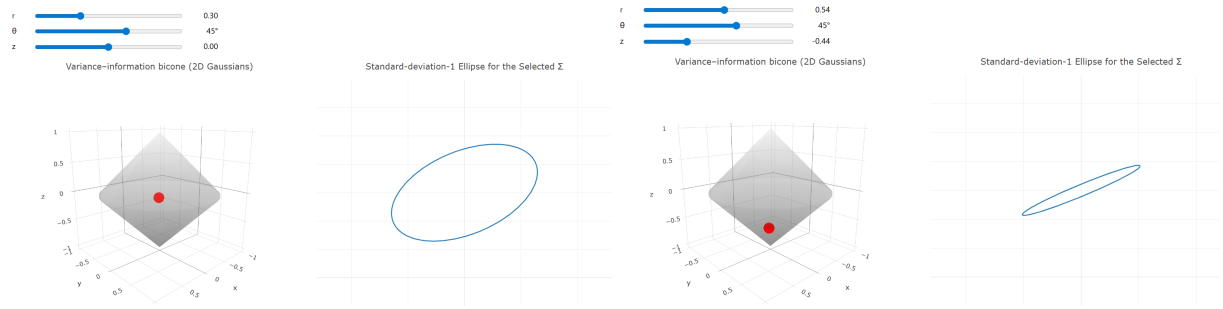


图 1：James 的三维双锥模型及其对应的双变异中心高斯分布截图。

证明：令 $X = U D U^\top$ 表示对称正定矩阵 X 的特征分解，其中 U 是 $O(n)$ 的正交矩阵，使得 $U U^\top = I$，且 $D = \mathrm{diag}((\lambda_i)_i)$ 是一个对角矩阵，其特征值为 $\lambda_i = \lambda_i(X) > 0$。由于我们可以将矩阵恒等式改写为 $I = U\,\mathrm{diag}((1)_i)\,U^\top$，对任意 $U \in O(n)$ 均成立，因此有 $I + X = U\,\mathrm{diag}((1+\lambda_i)_i)U^\top$，由此可得 $P(X) = (I+X)^{-1} = U^{-\top}\,\mathrm{diag}((\frac{1}{1+\lambda_i})_i)\,U^{-1}$。因此，我们可以将变换 $P(X)$ 改写为 $(I+X)^{-1} = V\,\mathrm{diag}((\frac{1}{1+\lambda_i})_i)\,V^\top$，其中 $V = R^{-\top} \in O(n)$ 是另一个正交矩阵。因此，$P(X) = (I+X)^{-1}$ 的特征值属于区间 $(0, 1)$，因为 $\lambda_i > 0$。现在，类似地，对于 $O \in O(n)$，有 $V(X) = O\,\mathrm{diag}((\frac{\lambda_i}{1+\lambda_i})_i)O^{-1}$。因此，其特征值落在区间 $(0, 1)$ 内。□

请注意，若 $J \in \mathrm{VPM}^\circ(n)$，则 $J^{-1} \notin \mathrm{VPM}^\circ(n)$。映射 $V(X)$ 可视为将协方差矩阵映射到 VPM 上，而映射 $P(X)$ 可解释为将精度矩阵映射到 VPM 上。由此可以考虑具有可能退化的协方差矩阵和/或精度矩阵的广义高斯分布 [18, 41]。

值得注意的是，VPM 定义域（及其闭包 $\overline{\mathrm{VPM}(n)} = \{X \in \text{正定矩阵}(n) : 0 \preceq X \preceq I\}$）在量子信息论 [28] 和效应代数 [14]中都有研究。事实上，定义域 $\mathrm{VPM}(n)$ 描述了正算子值测度（POVMs）中的矩阵效应 [23]。

VPM 定义域对于研究代数 Riccati 方程 [21, 19] 也很重要，其中映射V：正定矩阵$(n) \to \mathrm{VPM}^\circ(n)$ $X \mapsto V(X) = X(I+X)^{-1}$ 引入了一种归一化矩阵变换，确保特征值位于 $(0, 1)$ 内。

定义 1（James 映射） 令 ι：正定矩阵$(n) \to \mathrm{VPM}^\circ(n)$ 表示 $\iota(P) = P(I+P)^{-1}$，它是从维数为 n 的对称正定矩阵空间 正定矩阵 (n) 到 $\mathrm{VPM}^\circ(n)$ 的微分同胚。其逆映射为 $\iota^{-1}(X) = X(I-X)^{-1}$。

命题 2（ι 的微分） 对于任意 $P \in$ 正定矩阵(n) 和 $V \in T_P$ 正定矩阵$(n) \simeq \mathrm{Sym}(n)$，我们有：

$$d\iota_P(V) = (I+P)^{-1} V (I+P)^{-1}$$

引理1（ι 的逆公式） 设 $P \in$正定矩阵(n)。则：

1. $\iota(P) = (I + P^{-1})^{-1}$, *and*
2. $(I - \iota(P))^{-1} = I + P$.

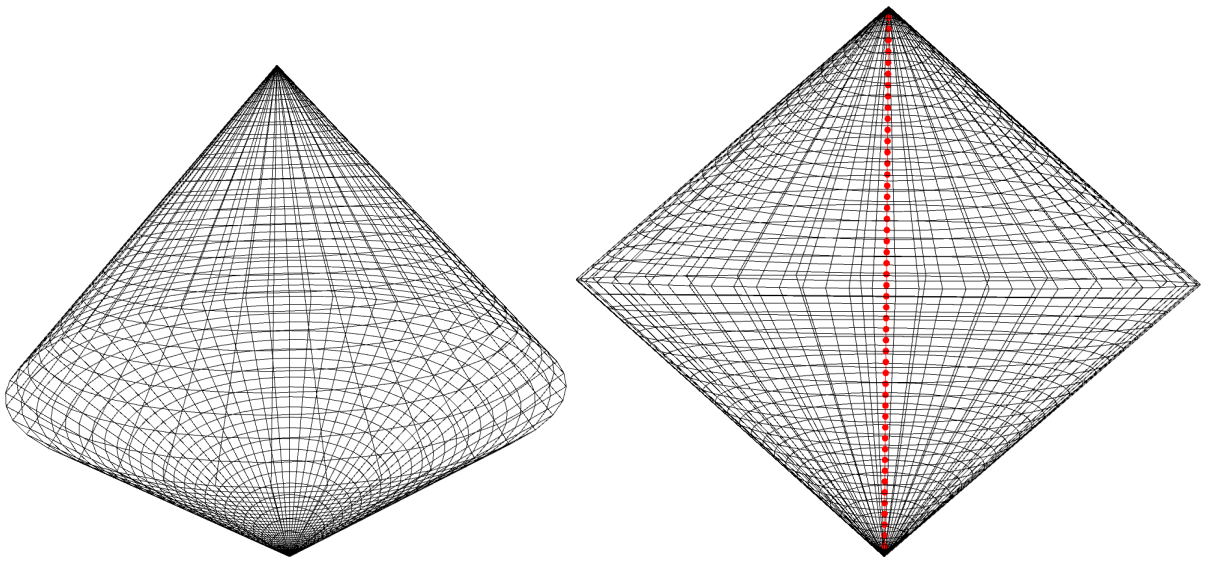


图 2：VPM°(2) 可视化为一个三维洛伦兹双锥。左侧以倾斜视角展示，以便更好地观察；其中还画出了连接 0 与 I（即 $\{0 \prec \alpha I \prec I$ ）的开预测地线： $\alpha \in (0,1)\}$。

证明： 对于第一部分，我们有：

$$(I+P)P^{-1} = P^{-1} + I = I + P^{-1} \implies P(I+P)^{-1} = (I+P^{-1})^{-1}$$

For 第二部分，根据任意 $X \in$VPM°(n) 的伍德伯里公式可得：

$$(I-X)^{-1} = I + X(I-X)^{-1} = I + \iota^{-1}(X)$$

因此：

$$(I-\iota(P))^{-1} = I + \iota^{-1}(\iota(P)) = I + P$$

□

当 $n = 2$ 时，VPM 可视化为一个三维洛伦兹双锥（图 2）。特别地，James [18] 采用了以下三维双锥的柱坐标参数化：

$$\begin{aligned} J = \begin{bmatrix} a & c \\ c & b \end{bmatrix} &\mapsto \left(r = \frac{\lambda_1 - \lambda_2}{1 + \lambda_1 + \lambda_2 + \lambda_1\lambda_2}, \theta = \arctan\frac{2c}{a-b}, z = \frac{\lambda_1\lambda_2 - 1}{1 + \lambda_1 + \lambda_2 + \lambda_1\lambda_2} \right) \\ &\mapsto (r\cos\theta, r\sin\theta, z). \end{aligned}$$

图 1 展示了交互式应用程序的一些屏幕截图。

J 的两个特征值 λ_1 和 λ_2 为

$$\lambda_1 = \frac{a+b-\sqrt{(a-b)^2+4c^2}}{2}, \quad \lambda_2 = \frac{a+b+\sqrt{(a-b)^2+4c^2}}{2}.$$

VPM°(n) 的闭包使我们能够考虑可能具有奇异协方差矩阵和/或精度矩阵的广义高斯分布 [41]。特别地，狄拉克分布被视为具有奇异协方差矩阵的广义高斯分布，而仿射子空间则被视为

具有奇异精度矩阵的广义高斯分布。处理广义高斯分布使我们能够用变量间关系所引起的非确定性来建模不确定性，这是 Willems 为建模开放随机系统所开创的框架 [43]。

有界凸集 VPM$^\circ(n)$ 上的对数交比希尔伯特距离已在 [20] 中研究。对称正定锥上的希尔伯特距离在信息论 [35]与统计学 [7] 等领域有众多应用。黎曼锥流形 $\mathcal{C}$ 具有许多潜在的全局坐标图：由对偶对数行列式势函数 $\Psi_{\mathrm{ld}}(X) = -\log\det(X)$ 与 Ψ^*_{ld} 诱导的信息几何给出两个全局坐标图 $\Theta =$ 正定矩阵(n) 与 $H = -$正定矩阵$(n) =$ ND(n)（负定锥），而定义在 SPD双锥 VPM$^\circ(n)$ 上的希尔伯特距离的微分几何诱导出芬斯勒结构 [36] 以及 $\mathcal{C}$ 上的距离。此外，通过考虑 VPM 上的双对数行列式函数 $\Psi_{\mathrm{bild}}(X) = -\log\det(X) - \log\det(I - X)$，我们得到对称矩阵空间 Sym$(n)$ 作为梯度空间。注意下列恒等式成立：$\log\det(X) = \mathrm{tr}\log(X)$。因此，计算 logdet势函数需要 $O(n^\omega)$ 次运算，其中 ω 表示 Mat(n) 中矩阵求逆、乘法与特征值分解复杂度的指数（$\omega \approx 2.37$）。

1.4 主要贡献与论文结构

我们的主要贡献概述如下：

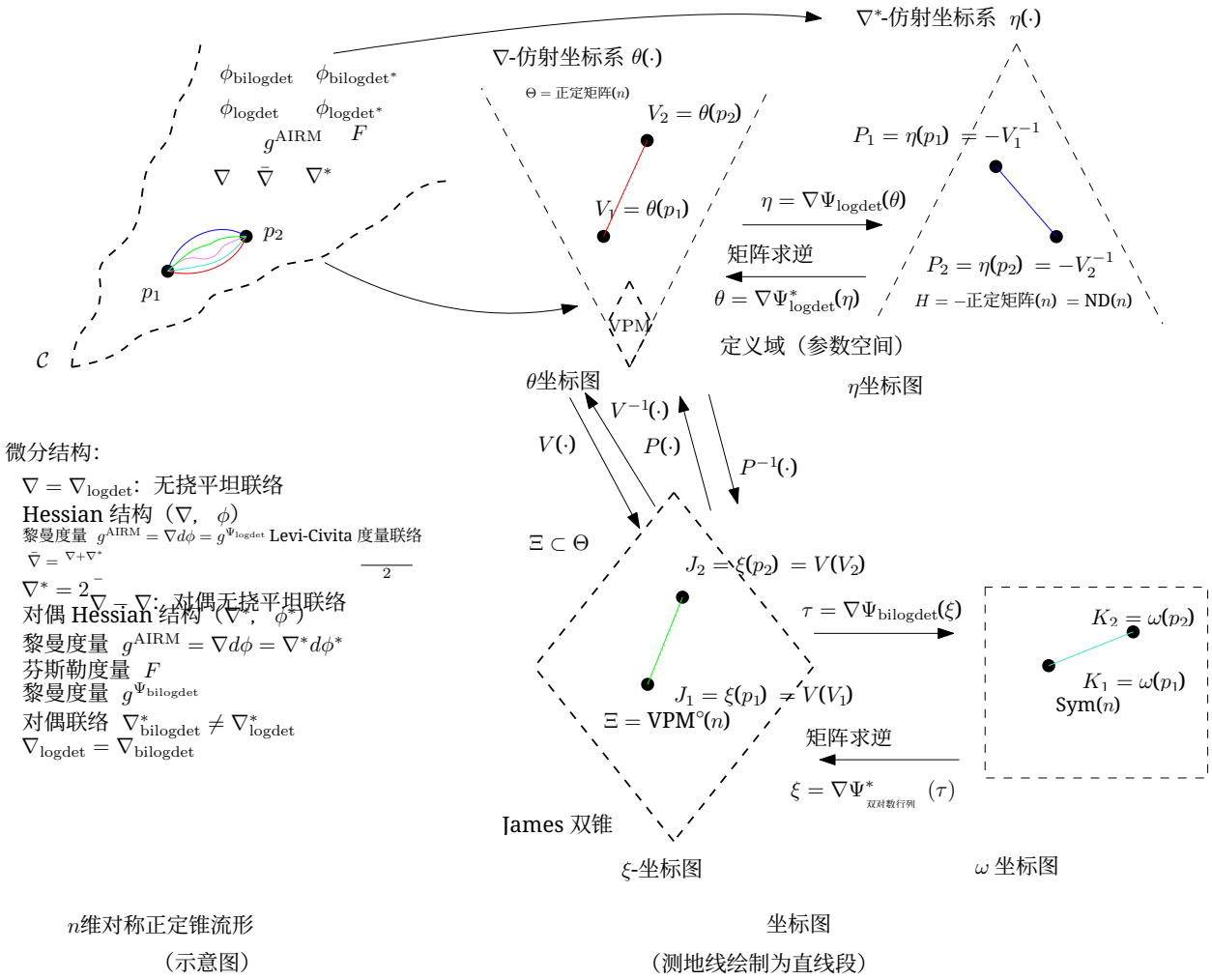
- 我们在定理 2 中证明，希尔伯特 VPM 距离推广了希尔伯特单纯形距离，
- 我们在定理 3 中给出了标准单纯形内希尔伯特测地线恒速参数化的闭式表达，
- 作为类比，我们在定理 4 中给出了 VPM 定义域内希尔伯特测地线恒速参数化的闭式表达，
- 我们在定理 5 中给出了 VPM 定义域内仿射不变黎曼距离与希尔伯特距离之间的紧下界，
- 类似地，我们在定理 6 中给出了希尔伯特距离相对于推前的仿射不变黎曼距离的一个紧上界，

本文结构安排如下：我们首先在第 2 节中描述由 VPM 距离的希尔伯特几何所诱导的芬斯勒结构（该几何最初在我们此前的工作 [20] 中提出）。随后，我们在第 3 节中介绍 VPM 上的双对数行列式障碍函数，并研究其诱导的信息几何。在第 4 节中，我们通过考察仿射不变黎曼度量在 VPM 定义域上的限制或其推前版本，将 VPM 希尔伯特距离与广泛使用的 AIRM 距离进行比较。最后，我们在第 5 节总结主要结果，并讨论其在量子信息论和控制理论中的潜在应用。附录中，§A 提供了符号表，§B 则给出了一些符号计算代码片段。

2 VPM 希尔伯特距离与芬斯勒结构

2.1 预备知识

定义 2（矩阵范数）对于 $V \in \mathrm{Sym}(n)$，其具有实特征值 $\lambda_1(V) \ldots, \lambda_n(V)$，定义：



微分结构：

$\nabla = \nabla_{\text{logdet}}$：无挠平坦联络
Hessian 结构（∇, ϕ）
黎曼度量 $g^{\text{AIRM}} = \nabla d\phi = g^{\Psi_{\text{logdet}}}$ Levi-Civita 度量联络
$\bar{\nabla} = \frac{\nabla+\nabla^*}{2}$
$\nabla^* = 2\bar{\nabla} - \nabla$：对偶无挠平坦联络
对偶 Hessian 结构（∇^*, ϕ^*）
黎曼度量 $g^{\text{AIRM}} = \nabla d\phi = \nabla^* d\phi^*$
芬斯勒度量 F
黎曼度量 $g^{\Psi_{\text{bilogdet}}}$
对偶联络 $\nabla^*_{\text{bilogdet}} \neq \nabla^*_{\text{logdet}}$
$\nabla_{\text{logdet}} = \nabla_{\text{bilogdet}}$

n维对称正定锥流形
（示意图）

坐标图
（测地线绘制为直线段）

图3：对偶SPD锥 $\Theta = \text{PD}(n)$与 $H = \text{ND}(n) = -\text{PD}(n)$经由logdet势函数，以及双锥 $\Xi = \text{VPM}^\circ(n)$及其对偶Sym$(n)$，均为对称正定锥流形 $\mathcal{C}$的全局坐标图。

*1. Frobenius*范数： $\|V\|_F = \sqrt{\sum_{i=1}^{n} \lambda_i^2(V)}$*2.* 算子范数（2-范数）： $\|V\|_{op} = \max_{1\le i\le n} |\lambda_i(V)|$

引理2（Frobenius-算子范数不等式）对于任意 $V \in \mathrm{Sym}(n)$，都有以下不等式：

$$\|V\|_{op} \le \|V\|_F \le \sqrt{n}\,\|V\|_{op}$$

引理 3（值域-l_2 不等式）对于任意 $x \in \mathbb{R}^n$，都有 $\max_{1\le k\le n} x_k - \min_{1\le k\le n} x_k \le \sqrt{2}\|x\|_2$。

证明：令 $i = \mathrm{argmax}_k\{x_k\}$ 和 $j = \arg\min_k\{x_k\}$。则：

$$\max_{1\le k\le n} x_k - \min_{1\le k\le n} x_k = x_i - x_j = \langle x, e_i - e_j\rangle \le \|x\|_2\|e_i - e_j\|_2 \le \sqrt{2}\|x\|_2.$$

其中该不等式由柯西-施瓦茨不等式得出。 □

希尔伯特几何已在各种有界凸区域 [26, 15] 上得到研究。文献 [24] 研究了对称正定域上的希尔伯特几何及其诱导的芬斯勒结构。

希尔伯特 文献 [20] 计算了 VPM 上两个参数 J_1 和 J_2 的距离：

定理 1（[20, 定理 14]）给定两个矩阵 $J_1, J_2 \in \mathrm{VPM}^\circ(n)$：

$$\mathrm{d_H}(J_1, J_2) = \log \frac{\max\left(\lambda_{\max}(J_2^{-1}J_1), \lambda_{\max}\big((I-J_2)^{-1}(I-J_1)\big)\right)}{\min\left(\lambda_{\min}(J_2^{-1}J_1), \lambda_{\min}\big((I-J_2)^{-1}(I-J_1)\big)\right)}$$

其中，$\lambda_{\min}(V)$ 和 $\lambda_{\max}(V)$ 分别表示对称矩阵V的最小和最大实特征值。

由于对于任意正数 $a, b > 0$，都有 $\max\{a,b\} \ge \frac{a+b}{2}$ 和 $\min\{a,b\} \le \frac{a+b}{2}$，因此得到 VPM 希尔伯特距离的以下下界：

花冠 定理 1（下界）希尔伯特 *VPM* 距离具有如下下界：

$$\mathrm{d_H}(J_1, J_2) \ge \log \frac{\lambda_{\max}(J_2^{-1}J_1) + \lambda_{\max}\big((I-J_2)^{-1}(I-J_1)\big)}{\lambda_{\min}(J_2^{-1}J_1) + \lambda_{\min}\big((I-J_2)^{-1}(I-J_1)\big)}.$$

注意，矩阵 $J_2^{-1}J_1$ 的特征值可能具有重数：例如，考虑$J_1 = \alpha I$和 $H_2 = \beta I$，其中 $\alpha \in$取值于 (0，1)，$\beta \in$取值于 (0，1)。于是，$J_2^{-1}J_1 = \frac{\alpha}{\beta} I$仅有一个不同的特征值，而 $(I-J_2)^{-1}(I-J_1) = \frac{1-\alpha}{1-\beta} I$也仅有一个不同的特征值。在这种情况下，我们得到 $\mathrm{d_H}(J_1, J_2) = \log \max\left\{\frac{\alpha}{\beta}, \frac{1-\alpha}{1-\beta}\right\} \min\left\{\frac{\alpha}{\beta}, \frac{1-\alpha}{1-\beta}\right\}$。我们验证了，当且仅当 $\alpha = \beta$时，$\mathrm{d_H}(J_1, J_2) = 0$。

例 1 考虑以下两个矩阵

$$J_1 = \begin{bmatrix} \frac{7}{20} & -\frac{3\sqrt{3}}{20} \\ -\frac{3\sqrt{3}}{20} & \frac{13}{20} \end{bmatrix}, \quad J_2 = \begin{bmatrix} \frac{11}{20} & -\frac{\sqrt{3}}{20} \\ -\frac{\sqrt{3}}{20} & \frac{9}{20} \end{bmatrix}.$$

我们有 $\Lambda(J_1) = \{\frac{1}{5}, \frac{4}{5}\}$和 $\Lambda(J_2) = \{\frac{2}{5}, \frac{3}{5}\}$。利用附录 B 中报告的符号计算，我们得到

$$\mathrm{d_H}(J_1, J_2) = \log \frac{47 + \sqrt{673}}{47 - \sqrt{673}} \approx 1.242398973577776.$$

注意，谱单纯形 $\mathrm{Spect}_n = \{X \succeq 0 : \mathrm{tr}(X) = 1\} \subset \mathrm{VPM}$ 是闭 $\overline{\mathrm{VPM}}$ 的一个仿射子空间。谱单纯形是标准单纯形的半定对应物。令 $\mathrm{Spect}_n^\circ = \{X \in \text{正定矩阵}(n) : \mathrm{tr}(X) = 1\} \subset \mathrm{VPM}^\circ$ 为开正定谱单纯形。开谱单纯形对应于将开标准单纯形（概率单纯形） Δ_n°嵌入对角矩阵：$p \in \Delta_n^\circ \mapsto D_p := \text{对角矩阵}(p_1, \ldots, p_n) \in \mathrm{Spect}_n$。单纯形定义域 [26, 27]上的希尔伯特距离为

$$d_H^{\Delta_n}(p,q) = \log \frac{\max_i \frac{p_i}{q_i}}{\min_i \frac{p_i}{q_i}}.$$

这种希尔伯特单纯形距离在线性变换下具有收缩性质，因此适用于分析最优传输中的 Sinkhorn 型算法 [17, 9]。

我们有

$$\mathrm{d_H}(D_p, D_q) = \log \frac{\max\{\max_{i=1}^n \frac{p_i}{q_i}, \max_{i=1}^n \frac{1-p_i}{1-q_i}\}}{\min\{\min_{i=1}^n \frac{p_i}{q_i}, \min_{i=1}^n \frac{1-p_i}{1-q_i}\}}.$$

T定理 2 希尔伯特 *VPM* 距离推广了希尔伯特单纯形距离 *ce:*

$$\mathrm{d_H}(D_p, D_q) = d_H^{\Delta_n}(p,q), \quad \forall p, q \in \Delta_n^\circ.$$

证明： 定义 $a_i = \frac{p_i}{q_i}$和 $b_i = \frac{1-p_i}{1-q_i}$。由于 $0 < q_i < 1$，考虑 $b_i - 1 = \frac{1-p_i}{1-q_i} - 1 = \frac{q_i - p_i}{1-q_i}$。

由于 $1 - q_i > 0$， $b_i - 1$ 的符号与 $q_i - p_i$的符号相同，而后者与 $p_i - q_i$的符号相反。但 $a_i - 1 = \frac{p_i - q_i}{q_i}$。因此，我们有$(a_i - 1)(b_i - 1) \le 0$。所以，对于每个 i，数值 a_i 和 b_i 位于 1 的两侧。

令 $M = \text{为}_i\, a_i$ 的最大值， $m = \text{为}_i\, a_i$的最小值。若 $M > 1$，则对于相应的指标 k ，我们有$b_k < 1$，因此没有 b_i 能够超过 M。若相反地 $M \le 1$，则所有 $b_i \ge 1$，同样的对称性论证仍然适用。因此我们得到

$$\max\{\max_i a_i, \max_i b_i\} = M.$$

同理，我们有

$$\min\{\min_i a_i, \min_i b_i\} = m.$$

因此，我们得到

$$\frac{\max\{\max_i a_i, \max_i b_i\}}{\min\{\min_i a_i, \min_i b_i\}} = \frac{M}{m} = \frac{\max_i \frac{p_i}{q_i}}{\min_i \frac{p_i}{q_i}}.$$

由于 $D_p D_q^{-1} = \mathrm{diag}(\frac{p_1}{q_1}, \ldots, \frac{p_n}{q_n})$ 且 $(I - D_p)(I - D_q)^{-1} = \mathrm{diag}(\frac{1-p_1}{1-q_1}, \ldots, \frac{1-p_n}{1-q_n})$，因此 $\mathrm{d_H}(D_p, D_q) = d_H^{\Delta_n}(p,q)$。□

[13, 26] 中指出，尽管标准单纯形的希尔伯特几何中直线是测地线，但一般而言测地线并不唯一。因此，直线段是希尔伯特 VPM 几何中的测地线，但这些测地线并不唯一。希尔伯特几何中存在线段（预测）测地线这一事实，使得人们能够推广并实现各种算法，例如点集的最小包围球等 [29], 。连接 J_1 与 J_2 的预测地线为$\tilde{\gamma}_{J_1 J_2}(t) = (1-t)J_1 + tJ_2$。要得到测地线，我们需要将预测地线重新参数化为单位速度：

$$\gamma_{J_1 J_2}(s) = \tilde{\gamma}_{J_1 J_2}(s(t))$$

使得 $\mathrm{d_H}(\gamma_{J_1 J_2}(s)\ \gamma_{J_1 J_2}(s')) = |s - s'|\,\mathrm{d_H}(J_1, J_2) = |s(t) - s(t')|\,\mathrm{d_H}(J_1, J_2)$。与希尔伯特测地线不同，仿射不变黎曼度量的测地线是指数弧：事实上， $\gamma^{\mathrm{AIRM}}_{D_p D_q}(s) = \text{对角矩阵}\left(q_1^s p_1^{1-s}, \ldots, q_n^s p_n^{1-s}\right)$。

定理 3（希尔伯特单纯形测地线参数化）希尔伯特单纯形几何中， p 与 q 之间的等速测地线为：

$$\gamma_{pq}^{\Delta_n}(s) = (1 - t(s))p + t(s)q,$$

其中

$$t(s) = \frac{1 - \left(\frac{M}{m}\right)^{1-s}}{\left(\frac{M}{m}\right)^{1-s}\left(\frac{1}{M} - 1\right) - \left(\frac{1}{m} - 1\right))}$$

且 $M = \text{最大值}_{i=1}^{n} \frac{p_i}{q_i}$ 和 $m = \text{最小值}\ \frac{p_i}{q_i}$。

定理 4（希尔伯特 VPM 测地线参数化）在希尔伯特单纯形几何中， J_1 与 J_2 之间测地线的等速测地线为：

$$\gamma_{pq}^{\mathrm{VPM}^\circ}(s) = (1 - t(s))J_1 + t(s)J_2,$$

其中

$$t(s) = \frac{1 - \left(\frac{M}{m}\right)^{1-s}}{\left(\frac{M}{m}\right)^{1-s}\left(\frac{1}{M} - 1\right) - \left(\frac{1}{m} - 1\right))}$$

and

$$M = \max\left(\lambda_{\max}(J_2^{-1}J_1), \lambda_{\max}\big((I - J_2)^{-1}(I - J_1)\big)\right)$$

and

$$m = \min\left(\lambda_{\min}(J_2^{-1}J_1), \lambda_{\min}\big((I - J_2)^{-1}(I - J_1)\big)\right).$$

我们可以将缩放后的 VPM 定义为 $\mathrm{VPM}_\lambda^\circ(n) = \{X \in \mathrm{PD}(n) : \ 0 \prec X \prec \lambda I\}$，并将 VPM 锥 $K_{\mathrm{VPM}} = \{(\lambda, \mathrm{VPM}_\lambda^\circ(n)) \text{ 定义为：} \ \lambda > 0\}$。于是， K_{VPM} 上的比尔霍夫射影度量与子空间 $(1, \mathrm{VPM}_1^\circ)$ 上的 $\mathrm{d_H}$ 一致，见 [22]。进一步地，我们可以通过映射 $V_\lambda(X) = \lambda X(I + X)^{-1}$ 将对称正定矩阵嵌入 $\mathrm{VPM}_\lambda^\circ(n)$，从而保证特征值落在区间 $(0, \lambda)$ 内。

图 4 展示了关于 AIRM 结构（蓝色）和希尔伯特-芬斯勒结构（红色）的测地线中点，其中两个输入的 2×2 对称正定矩阵以黑色表示。

下文中的许多定理将用到以下从 $\mathrm{VPM}^\circ(n)$ 到乘积锥 $\mathrm{PD}(n) \times \mathrm{PD}(n)$ 的嵌入。

定义 3（帽嵌入）我们定义帽嵌入：

$$(\hat{}) : \mathrm{VPM}^\circ(n) \to \mathrm{VPM}^\circ(n) \times \mathrm{VPM}^\circ(n) \subset \mathrm{PD}(n) \times \mathrm{PD}(n) \qquad as \qquad \hat{X} = (X, I - X).$$

2.2 希尔伯特距离作为矩阵展宽

定义 4 对于对称矩阵 $S \in \mathrm{Sym}(n)$，我们将 S 的矩阵展宽记为 *：*

$$\mathbf{spread}(S) = \lambda_{\max}(S) - \lambda_{\min}(S)$$

为了考察**离散度**与希尔伯特距离之间的关系，我们需要借用某个量 M，它出现在希尔伯特距离的比尔科夫表示中。

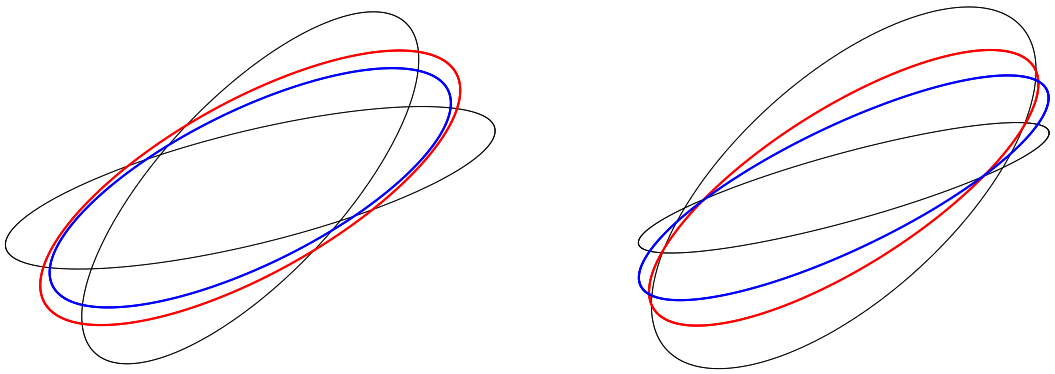


图 4：两个对称正定矩阵（黑色）的仿射不变黎曼度量（蓝色）和希尔伯特-芬斯勒（红色）测地线中点。

定义 5（上比尔霍夫比率 M）设 $\mathcal{V}$ 为一个向量空间，$K \subseteq \mathcal{V}$ 为诱导序 $\preceq$的开凸尖锥。我们对任意 $w \in K$、$v \in \mathcal{V}$定义 $M(\cdot,\cdot)$ 如下：

$$M_K(v,w) = \inf\{\lambda > 0 : \ v \preceq \lambda w\} \ \in [0,\infty].$$

备注 1（比尔霍夫射影距离）给定向量空间 $\mathcal{V}$ 以及定义 $\mathcal{V}$上序 $\preceq$ 的开凸尖锥 $K \subseteq \mathcal{V}$，我们可以在 K 上定义比尔霍夫距离$\mathrm{d_B}$：

$$\mathrm{d_B}(v,w) = \log M(v,w) + \log M(w,v)$$

命题 3（通过帽映射对$\mathrm{d_H}$ 进行对称化）对于任意 X，$Y \in \mathrm{VPM}^{\circ}(n)$，以下恒等式成立：

$$\begin{aligned}
\mathrm{d_H}(X,Y) &= \log M(\hat{X},\hat{Y}) + \log M(\hat{Y},\hat{X}) \\
M(\hat{X},\hat{Y}) &= \max\left\{\lambda_{\max}(Y^{-1}X),\, \lambda_{\max}\big((I-Y)^{-1}(I-X)\big)\right\}
\end{aligned}$$

证明：等式：

$$\mathrm{d_H}(X,Y) = \log M(\hat{X},\hat{Y}) + \log M(\hat{Y},\hat{X})$$

$\hat{}$ 以及 $M(X,Y)$ 的公式已在定理 [20, 14], 的证明中给出，不过其中使用了量 $m(X,Y) = \sup\{\lambda > 0 : \ \lambda Y \preceq X\}$。注意到$m(X,Y) = 1/M(Y,X)$，即可轻松将其转换为 M。□

命题 4 对于 $J_1, J_2 \in \mathrm{VPM}^{\circ}(n)$，希尔伯特距离$\mathrm{d_H}(J_1,J_2)$ 可表示为矩阵展宽：

$$\begin{aligned}
\mathrm{d_H}(J_1,J_2) &= \mathbf{spread}\left(\log\left(\hat{J_2}^{-1/2}\hat{J_1}\hat{J_2}^{-1/2}\right)\right) \\
&= \mathbf{spread}\left(\log\begin{bmatrix} J_2^{-1/2}J_1J_2^{-1/2} & 0 \\ 0 & (I-J_2)^{-1/2}(I-J_1)(I-J_2)^{-1/2}\end{bmatrix}\right)
\end{aligned}$$

证明： 由命题 3 可知：

$$\mathrm{d_H}(X,Y) = \log M(\hat{X},\hat{Y}) + \log M(\hat{Y},\hat{X})$$

我们构造一个大的分块对角矩阵：

$$P = \begin{bmatrix} Y^{-1/2}XY^{-1/2} & 0 \\ 0 & (I-Y)^{-1/2}(I-X)(I-Y)^{-1/2} \end{bmatrix} \in \mathrm{PD}(2n)$$

其特征值恰好是 $Y^{-1}X$ 和 $(I-Y)^{-1}(I-X)$ 的特征值的并集，因此：

$$\mathrm{d_H}(X,Y) = \log \frac{\lambda_{\max}(P)}{\lambda_{\min}(P)} = \log \lambda_{\max}(P) - \log \lambda_{\min}(P) = \lambda_{\max}(\log P) - \lambda_{\min}(\log P)$$

□

2.3 Finsler 范数与芬斯勒度量

为了给出与希尔伯特距离对应的 Finsler 结构，我们先提出非对称范数的概念。

定义 6（非对称范数）向量空间 V 上的函数 $f:\ \mathcal{V} \to \mathbb{R}$ 称为非对称范数，若其满足：

- 三角不等式：对于任意 $x, y \in \mathcal{V}$, $f(x+y) \le f(x) + f(y)$,
- 齐次性：对于任意 $x \in \mathcal{V}$和 $\lambda \in \mathbb{R}_+$， $f\left(\lambda x\right) = \lambda f\left(x\right)$，
- 正定性： $f\left(x\right) \ge 0$对所有 $x \in V$成立，并且当且仅当 $x = 0$时， $f\left(x\right) = 0$。

定义 7（Finsler 结构）设 M 为一个光滑流形， $T\mathcal{M}$ 为其切丛。在 $\mathcal{M}$ 上的 *Finsler* 结构是一个连续函数 $F:\ T\mathcal{M} \to [0,\infty)$，使得对于每个 $x \in \mathcal{M}$, $||\cdot||_F:\ T_x\mathcal{M} \to \mathbb{R}$ 是一个非对称范数，我们称其为 M上的 *Finsler* 范数。我们在 M 中的分段光滑曲线上定义长度泛函如下：

$$L_F(\gamma) = \int_a^b ||\dot{\gamma}(t)||_x^F \ dt \qquad for \qquad \gamma : [a,b] \to M$$

以及 x与 $y \in M$ 之间的相应 *Finsler* 距离如下：

$$d_F(x,y) = \inf_{\gamma} L_F(\gamma) \qquad where \qquad \gamma(a) = x, \gamma(b) = y.$$

为展示 Finsler 结构如何由希尔伯特距离产生，我们需要定义退出时间，它们与比尔霍夫比率 $M(\cdot,\ \cdot)$ 有关。

引理 4（退出时间 $t^{\pm}$）设 $K \subseteq \mathcal{V}$ 为一个开凸尖锥。对于 $w \in K$ 和 $v \in \mathcal{V}$，定义前向和后向锥退出时间：

$$t_K^+(w,v) = \sup\{t > 0:\ w + tv \in K\} \in (0,\infty] \qquad t_K^-(w,v) = t_K^+(w,-v)$$

那么，按照约定 $\frac{1}{\infty} = 0$，我们有：

$$\frac{1}{t_K^+(w,v)} = M_K(-v,w) \qquad \frac{1}{t_K^-(w,v)} = M_K(v,w)$$

证明： 对于 $t>0$，我们有：

$$w+tu\in K \quad\Longleftrightarrow\quad \frac{1}{t}w+u\in K \quad\Longleftrightarrow\quad -u\preceq\frac{1}{t}w.$$

因此，集合 $\{t>0:\ \ w+tu\in K\}$ 经由 $\lambda=\frac{1}{t}$ 对应于集合 $\{\lambda>0:\ \ -u\preceq\lambda w\}$，而对 t 取上确界等价于对 λ 取下确界。由此得到 $t_K^+(w,u)=1/M_K(-u,w)$，而 t_K^- 的公式则通过替换 $u\mapsto -u$ 得出。□

命题 5（由希尔伯特距离给出的芬斯勒范数） 设 $C\subset\mathbb{R}^n$ 为配备希尔伯特距离 $\mathrm{d_H}$ 的开有界凸集。则 C 带有一个芬斯勒结构，其在 $x\in C$ 处对于 $v\in T_xC\simeq\mathbb{R}^n$ 的范数由下式给出：

$$\|v\|_x^H=\frac{1}{t^+(x,v)}+\frac{1}{t^-(x,v)}$$

证明： 固定 $x\in C$ 与 $v\neq 0$。由于 C 是开的、凸的且有界的，直线 $\ell=\{x+sv:\ \ s\in\mathbb{R}\}$ 与 ∂C 恰好交于两点：

$$a=x-t^-(x,v)v\qquad b=x+t^+(x,v)v$$

其中 $t^{\pm}(x,v)\in(0,\infty)$。对于 $s\in(-t^-(x,v)\ t^+(x,v))$，令 $y_s=x+sv\in C$。

根据直线上的希尔伯特距离的交比公式 ℓ（在坐标 s中）：

$$\mathrm{d_H}(x,y_s)=\log\frac{\|a-y_s\|\,\|b-x\|}{\|a-x\|\,\|b-y_s\|}=\log\left(\frac{(t^-(x,v)+s)\,t^+(x,v)}{t^-(x,v)\,(t^+(x,v)-s)}\right).$$

因此：

$$\frac{d}{ds}\ \mathrm{d_H}(x,y_s)=\frac{1}{t^-(x,v)+s}+\frac{1}{t^+(x,v)-s},$$

并在 $s=0$ 处求值，得到：

$$\left.\frac{d}{ds}\right|_{s=0}\mathrm{d_H}(x,x+sv)=\frac{1}{t^-(x,v)}+\frac{1}{t^+(x,v)}.$$

这就在每个 T_xC上定义了一个芬斯勒范数。 □

现在，我们希望显式计算由希尔伯特距离给出的芬斯勒范数，这可以通过帽映射转到乘积锥正定矩阵$^2(n)$来实现。我们首先需要了解比尔霍夫比率在乘积锥正定矩阵上的表现。

引理 5（乘积锥上的比尔霍夫比率） 令 $K\subset\mathcal{V}$ 为诱导序 $\preceq$的开凸尖锥。则对于乘积锥 $K\times K\subset\mathcal{V}\times\mathcal{V}$及其乘积序 $\preceq^2$：

$$(v,v')\preceq^2(w,w')\iff v\preceq w\ \ and\ v'\preceq w'$$

我们得到：

$$M((v,v'),(w,w'))=\max(M(v,w),M(v',w'))$$

证明： 展开乘积锥上 M 的定义，我们得到：

$$M((v,v'),(w,w')) = \inf\{\lambda > 0 : (v,v') \preceq \lambda(w,w')\}$$

这意味着：

$$(v,v') \preceq^2 \lambda(w,w') \iff v \preceq \lambda w \text{ and } v' \preceq \lambda w'$$

因此，最小可行的 λ 是两个分量 M中的最大值。 □

命题 6（由希尔伯特距离导出的芬斯勒范数）设 $\mathrm{PD}^2(n) = \mathrm{PD}(n) \times \mathrm{PD}(n)$ 为乘积锥，配备乘积洛温纳序。对于 $X \in \mathrm{VPM}^\circ(n)$ 和 $V \in \mathrm{Sym}(n)$，令 $\hat{X} = (X, I-X) \in \mathrm{PD}^2(n)$ 且 $\hat{V} = (V, -V) \in \mathcal{V}$。那么，由 d_{H} 在 X 处诱导的芬斯勒范数满足：

$$\|V\|_X^H = \max\Big(\max(0, -\lambda_{\min}(A)),\ \max(0, \lambda_{\max}(B))\Big) + \max\Big(\max(0, \lambda_{\max}(A)),\ \max(0, -\lambda_{\min}(B))\Big)$$

其中：

$$A = X^{-1/2} V X^{-1/2} \qquad B = (I-X)^{-1/2} V (I-X)^{-1/2}.$$

证明： 我们证明：

$$\begin{aligned}
\|V\|_X^H &= \frac{1}{t^+(X,V)} + \frac{1}{t^-(X,V)} \\
&= M_{\mathrm{PD}^2(n)}(\hat{V}, \hat{X}) + M_{\mathrm{PD}^2(n)}(-\hat{V}, \hat{X}) \\
&= \max\big(M_{\mathrm{PD}}(V, X),\ M_{\mathrm{PD}}(-V, I-X)\big) + \max\big(M_{\mathrm{PD}}(-V, X),\ M_{\mathrm{PD}}(V, I-X)\big) \\
&= \max\Big(\max(0, -\lambda_{\min}(A)),\ \max(0, \lambda_{\max}(B))\Big) + \max\Big(\max(0, \lambda_{\max}(A)),\ \max(0, -\lambda_{\min}(B))\Big)
\end{aligned}$$

第一个恒等式由命题 5 得出。根据定义：

$$t^+(X,V) = \sup\{t > 0 :\ X + tV \in \mathrm{VPM}^\circ(n)\}.$$

利用帽嵌入，$X + tV \in \mathrm{VPM}^\circ(n)$ 等价于 $\hat{X} + t\hat{V} \in$ 正定矩阵$^2(n)$。因此：

$$t^+(X,V) = t^+_{\mathrm{PD}^2(n)}(\hat{X}, \hat{V}) \qquad t^-(X,V) = t^-_{\mathrm{PD}^2(n)}(\hat{X}, \hat{V})$$

第二个恒等式由引理 4 得出。第三个恒等式则是因为，根据引理 5，我们有：

$$M_{\mathrm{PD}^2(n)}(\hat{V}, \hat{X}) = \max\big(M_{\mathrm{PD}}(V, X),\ M_{\mathrm{PD}}(-V, I-X)\big)$$

因为在乘积锥上，$M_{\mathrm{PD}^2(n)}(\cdot,\ \cdot)$ 是各分量 M 的最大值。接下来只需计算 $M_{\mathrm{PD}}(P,\ Q)$，其中 $Q \in$ 属于正定矩阵(n)，$P \in$ 属于 $\mathrm{Sym}(n)$。根据定义：

$$M_{\mathrm{PD}}(P,Q) = \inf\{\lambda > 0 :\ P \preceq \lambda Q\} = \inf\{\lambda > 0 :\ \lambda Q - P \succeq 0\}.$$

对 $Q^{-1/2}$ 进行合同变换，等价地可得 $\lambda I - Q^{-1/2} P Q^{-1/2} \succeq 0$，即 $\lambda \geq \lambda_{\max}(Q^{-1/2} P Q^{-1/2})$。对 $\lambda > 0$ 取下确界，得到：

$$M_{\mathrm{PD}}(P,Q) = \max\big(0, \lambda_{\max}(Q^{-1/2} P Q^{-1/2})\big).$$

将此应用于 $(P,Q) = (-V, X)$ 与 $(P,Q) = (V, I-X)$，可得 $1/t^+(X,V)$ 的表达式，对 $1/t^-(X,V)$ 亦然。 □

注 2 对于 $n \geq 2$，考虑按如下方式对对称正定矩阵 X 进行归一化：$\tilde{X} = \frac{1}{\mathrm{tr}(X)}X$。于是 $\tilde{X}$ 我们有 $\mathrm{tr}(X) = 1$，并且由于迹是 X 的正特征值之和，且 X 是 $\tilde{}$ 对称正定的，其所有特征值均落在区间 $(0,1)$ 内。因此 X 属于 *VPM*。由此我们可以定义射影 *VPM* 距离：

$$\tilde{d}_H(X_1, X_2) = d_H(\tilde{X}_1, \tilde{X}_2).$$

芬斯勒结构使得可以计算均值和中位数 [3]。

3 双对数 VPM 上的 det 散度及其 Hessian 结构

本节在 VPM 上定义一种对偶平坦结构，该结构由一个严格凸且可微的函数诱导，我们称之为双对数行列式函数。

定义 8 在集合 $\mathrm{VPM}^{\circ}(n)$上，我们定义双对数行列式函数 $\Psi : \mathrm{VPM}^{\circ}(n) \to [0, \infty]$ 如下：

$$\Psi_{\mathrm{bild}}(X) = -\log\det X - \log\det(I - X) = \Psi_1(X) + \Psi_2(X)$$

为简化记号，以下我们简记 Ψ 为 Ψ_{bild}。

命题 7 函数 $\Psi(X)$ 是严格凸函数。

证明： 由 [5], 可知，函数 $\Psi_1(X) = -\log\det X$ 是凸函数。函数$\Psi_2(X) = -\log\det(I - X)$也是严格凸函数，因为它就是 $\Psi_1 \circ (I - \cdot)$；后者是仿射函数，而与仿射函数复合会保持凸性。最后，两个严格凸函数之和仍然是严格凸函数。□

推论 2 Ψ 的 *Hessian* 在$\mathrm{VPM}^{\circ}(n)$上给出一个黎曼度量，其形式为 $g_X(V, W) = \nabla^2_X \Psi(V, W)$。

证明： 我们将 $g_X(V, V)$ 写成非负数的平方和：

$$g_X(V, V) = ||A||^2_F + ||B||^2_F > 0$$

对于 $A = X^{-1/2}VX^{-1/2}$ 和 $B = (I - X)^{-1/2}V(I - X)^{-1/2}$。 □

注 3 我们有 $\Psi(X) \to \infty$，因为 $X \to \partial\overline{\mathrm{VPM}(n)}$，因此势函数 Ψ 适配于$\mathrm{VPM}(n)$的边界。

命题 8 内部集合 $\mathrm{VPM}^{\circ}(n)$可赋予黎曼流形结构，其黎曼度量写为：

$$g_X\Psi(V, W) = g_X(V, W) = \mathrm{tr}(X^{-1}VX^{-1}W) + \mathrm{tr}((I - X)^{-1}V(I - X)^{-1}W)$$

证明： VPM$^\circ(n)$ 的内部是 Sym(n) 中的开集，因此继承了环境光滑流形结构。我们有：

$$d\Psi_1(X)[V] = d(-\log\det X)[V] = -\mathrm{tr}(X^{-1}V)$$
$$d\Psi_2(X)[V] = d(-\log\det(I-X))[V] = \mathrm{tr}((I-X)^{-1}V)$$

再来看：

$$\nabla^2\Psi_1(X)[V,W] = \mathrm{tr}(X^{-1}VX^{-1}W)$$
$$\nabla^2\Psi_2(X)[V,W] = \mathrm{tr}((I-X)^{-1}V(I-X)^{-1}W)$$

因此，障碍度量为：

$$g_X(V,W) = \nabla^2_X\Psi(V,W) = \mathrm{tr}(X^{-1}VX^{-1}W) + \mathrm{tr}((I-X)^{-1}V(I-X)^{-1}W)$$

□

注 4 对于 $X, Y \in \mathrm{VPM}^\circ(n)$ 和 $V \in T_X\,\mathrm{VPM}^\circ(n)$，我们将 $\mathrm{VPM}^\circ(n)$ 上相应的 Ψ 度量、其中分段光滑曲线的 Ψ长度，以及 $\mathrm{VPM}^\circ(n)$ 中两点之间的 Ψ距离记为：

$$||V||^\Psi_X = \sqrt{g^\Psi_X(V,V)} \qquad L_\Psi(\gamma) = \int ||\dot\gamma(t)||_{\gamma(t)}dt \qquad \mathrm{d}_\Psi(X,Y) = \inf_\gamma L_\Psi(\gamma)$$

由双对数行列式生成元诱导的梯度映射为

$$J \mapsto \nabla\Psi_{\mathrm{bild}}(J) = -J^{-1} + (I-J)^{-1}. \tag{1}$$

命题 9 对于任意 $X \in \mathrm{VPM}^\circ(n)$，对 $U \in O(n)$进行正交共轭 $(X \mapsto UXU^T)$ 以及 $X \mapsto I - X$ 反演都是对数行列式障碍度量的等距变换。

证明： 首先，根据 [20, 命题 4], 可知，VPM$^\circ(n)$ 在正交共轭和反演下保持不变。共轭映射为：

$$c_U(X) = UXU^T$$

关于 X 是线性的，因此其微分就等于其本身，$d(c_U)(V) = UVU^T$。现在，我们来计算黎曼度量 $g^1_X(V, W) = \nabla^2_X\Psi_1(V, W)$ 的第一项：

$$\begin{aligned} g^1_{c_U(X)}(d(c_U)V, d(c_U)W) &= g^1_{UXU^T}(UVU^T, UWU^T) \\ &= \mathrm{tr}((UXU^T)^{-1}(UVU^T)(UXU^T)^{-1}(UWU^T)) \\ &= \mathrm{tr}(UX^{-1}U^TUVU^TUX^{-1}U^TUWU^T) \\ &= \mathrm{tr}(UX^{-1}VX^{-1}WU^T) \\ &= \mathrm{tr}(X^{-1}VX^{-1}W) = g_X(V,W) \end{aligned}$$

其中最后一行来自迹的循环不变性。类似地，我们计算第二部分$g^2_X(V,W) = \nabla^2_X\Psi_2(V,W)$。

对于第二类等距变换，仿射反演映射 $X \mapsto i(X) = I - X$的微分为：

$$di_X V = -V$$

因此，我们分别计算障碍度量的两部分：

$$\begin{aligned} g_{i(X)}(d(i_X)V, diW) &= g^1_{i(X)}(d(i_X)V, d(i_X)W) + g^2_{i(X)}(d(i_X)V, d(i_X)W) \\ &= g^1_{(I-X)}(-V,-W) + g^2_{(I-X)}(-V,-W) \\ &= \mathrm{tr}((I-X)^{-1}(-V)(I-X)^{-1}(-W)) + \mathrm{tr}((I-I+X)^{-1}(-V)(I-I+X)^{-1}(-W)) \\ &= g^2_X(V,W) + g^1_X(V,W) = g_X(V,W) \end{aligned}$$

□

势函数 $\Psi_1 = \Psi_{\mathrm{ld}}$、$\Psi_2$ 和 Ψ_{bild} 分别诱导出以下 Bregman 对数行列式、补空间对数行列式和双对数行列式散度：

$$\begin{aligned} B_{\Psi_{\mathrm{ld}}}(V_1 : V_2) &= \mathrm{tr}(V_1 V_2^{-1}) - \log\det(V_1 V_2^{-1}) - n, \\ B_{\Psi_2}(J_1 : J_2) &= \mathrm{tr}((I-J_1)(I-J_2)^{-1}) - \log\det((I-J_1)(I-J_2)^{-1}) - n = B_{\Psi_{\mathrm{ld}}}(I-J_1 : I-J_2), \\ B_{\Psi_{\mathrm{bild}}}(J_1 : J_2) &= B_{\Psi_{\mathrm{ld}}}(J_1 : J_2) + B_{\Psi_{\mathrm{ld}}}(I-J_1 : I-J_2) \end{aligned}$$

注意，由于 $\log\det(V_1 V_2^{-1}) = \mathrm{tr}(\log(V_1 V_2^{-1}))$，对数行列式散度 $B_{\Psi_{\mathrm{ld}}}(V_1 : V_2)$ 可改写为：

$$B_{\Psi_{\mathrm{ld}}}(V_1 : V_2) = \mathrm{tr}(V_1 V_2^{-1} - \log(V_1 V_2^{-1})) - n = \sum_{\lambda_i \in \Lambda(V_1 V_2^{-1})} \lambda_i - \log\lambda_i - 1,$$

其中，$\Lambda(X) = \{\lambda_1(X) \ldots, \lambda_n(X)\}$ 是 X 的特征值谱。也就是说，对数行列式散度是一种 Bregman 谱散度（同时也是 AIRM 距离）。

一般而言，由 Hessian 度量 $g(\theta) = \nabla^2 F(\theta)$ 诱导的黎曼距离 ρ_F 有如下上界：$\sqrt{S_F}$，其中 S_F 是对称化 Bregman 散度 [25]：$\rho_F(\theta_1, \theta_2) \leq \sqrt{(\theta_2 - \theta_1)^\top (\nabla F(\theta_2) - \nabla F(\theta_1))}$.

定义 9（自协调函数）开凸域 $C \subseteq \mathbb{R}$ 上的凸函数 $f : C \to \mathbb{R}$ 称为自协调函数，如果：

$$|f'''(x)| \leq 2\,(f''(x))^{3/2}.$$

开凸域 $C \subseteq \mathbb{R}^n$ 上的凸函数 $f : C \to \mathbb{R}$ 称为自协调函数 *[34]*，如果对任意直线 $\ell \in \mathbb{R}^n$，其1维限制 $f|_{\ell \cap C}$ 都是自协调的。

命题 10（[5, 第 9] 章）我们有：

- 自协调函数在与仿射函数进行预复合以及 *er*
  此外，
- 对数行列式函数 $f(X) = -\log\det X$ 是自协调的。

推论 3 对数障碍函数 Ψ 在 $\mathrm{VPM}^\circ(n) \subseteq \mathrm{Sym}(n)$ 上是自协调的。

证明：这直接由命题 10 得出，同时注意到定义域 VPM°(n) 是凸的，如 [20,命题 3] 中所证明。□

4 仿射不变黎曼度量/对数行列式差异度与希尔伯特/双对数行列式 VPM 差异度之间的不等式

4.1 用仿射不变黎曼度量进行界定

定义 10（仿射不变黎曼距离（AIRM）在 正定矩阵(n)） 对于两个矩阵 P、$Q \in$ 正定矩阵(n)，我们将它们的仿射不变黎曼距离（*AIRM*）定义为：

$$\mathrm{d}_{\mathrm{AIRM}}(P,Q) = \sqrt{\sum_{i=1}^{n} \log^2 \lambda_i(Q^{-1}P)}.$$

我们寻求用 $\mathrm{d}_{\mathrm{AIRM}}$ 给出 d_{H} 的界。需要考虑几种情形。一种做法是直接将仿射不变黎曼度量从 PD(n) 限制到 VPM°(n)；另一种是通过 James 映射 ι 将其推前（等价地，通过 ι^{-1} 将希尔伯特度量从 VPM°(n) 拉回到 PD(n)）。

定义 11 在集合 $\mathrm{VPM}^{\circ}(n)$ 上，我们分别定义限制的仿射不变黎曼距离和推前仿射不变黎曼距离如下：

$$\mathrm{d}^{||}_{AIRM}(X,Y) = \mathrm{d}_{\mathrm{AIRM}}(X,Y), \tag{2}$$
$$\mathrm{d}^{\rightarrow}_{AIRM}(X,Y) = \mathrm{d}^{\rightarrow}_{AIRM}(\iota^{-1}(X),\iota^{-1}(Y)) = \mathrm{d}_{\mathrm{AIRM}}(X(I-X)^{-1},Y(I-Y)^{-1}). \tag{3}$$

4.2 用限制的 AIRM 距离给出的界 $\mathrm{d}^{||}$ AIRM

命题 11（仿射不变黎曼度量下无上界$\mathrm{d}^{||}$ **仿射不变黎曼度量）**不存在常数 $\kappa > 0$，使得对于所有 X, $Y \in \mathrm{VPM}^{\circ}(n)$，我们都能对$\mathrm{d}_{\mathrm{H}}$ 给出如下上界：

$$\mathrm{d}_{\mathrm{H}}(X,Y) \leq \kappa \cdot \mathrm{d}^{||}_{AIRM}(X,Y)$$

证明：我们来证明存在一个矩阵序列 X_t, Y_t，其中 $t = 3, 4, \ldots$，使得以下各式同时成立：

$$\lim_{t\to\infty} \mathrm{d}^{||}_{\mathrm{AIRM}}(X_t,Y_t) = 0, \qquad \lim_{t\to\infty} \mathrm{d}_{\mathrm{H}}(X_t,Y_t) > 0$$

这立即推出了结论。对于 $t \geq 3$，定义两个矩阵：

$$X_t = \mathrm{diag}\left(1-\frac{2}{t}, \frac{1}{2}, \ldots \frac{1}{2}\right) \qquad Y_t = \mathrm{diag}\left(1-\frac{1}{t}, \frac{1}{2}, \ldots \frac{1}{2}\right)$$

两个矩阵都属于 VPM°(n)，且在同一基下均为对角矩阵，因此我们可以直接计算 $Y_t^{-1}X_t$ 的特征值 λ：

$$\lambda_1 = \frac{1-\frac{2}{t}}{1-\frac{1}{t}}, \qquad \lambda_{2\leq i\leq n} = 1$$

注意，对所有 $t > 2$，都有 $\lambda_1 \leq 1$，这一点稍后会很重要。目前，我们计算：

$$\begin{aligned}\lim_{t\to\infty} \mathrm{d}^{||}_{\mathrm{AIRM}}(X_t, Y_t) &= \lim_{t\to\infty} \sqrt{\sum_{i=1}^{n} \log^2 \lambda_i(Y_t^{-1} X_t)} \\ &\leq \lim_{t\to\infty} \sqrt{\log^2 \left(\frac{1-\frac{2}{t}}{1-\frac{1}{t}}\right)} \\ &= \lim_{t\to\infty} \left|\log\left(1 + \frac{1}{1-t}\right)\right| \\ &= \lim_{t\to\infty} \left|\frac{1}{1-t} + o\left(\frac{1}{t^2}\right)\right| = 0\end{aligned}$$

其中最后一行是 $\log(1+x) = x + O(x^2)$ 的泰勒展开。另一方面，我们通过观察矩阵来计算希尔伯特距离：

$$(I - Y_t)^{-1}(I - X_t) = \mathrm{diag}\,(t, 2, \ldots 2) \cdot \mathrm{diag}\left(\frac{2}{t}, \frac{1}{2}, \ldots \frac{1}{2}\right) = \mathrm{diag}\,(2, 1 \ldots 1)$$

由此计算：

$$\begin{aligned}\lim_{t\to\infty} \mathrm{d_H}(X_t, Y_t) &= \lim_{t\to\infty} \log \frac{\max\left(\lambda_{\max}(Y_t^{-1}X_t), \lambda_{\max}\big((I-Y_t)^{-1}(I-X_t)\big)\right)}{\min\left(\lambda_{\min}(Y_t^{-1}X_t), \lambda_{\min}\big((I-Y_t)^{-1}(I-X_t)\big)\right)} \\ &= \log \frac{\max(1,2)}{\min\left(\lim_{t\to\infty} \frac{1-2/t}{1-1/t}, 1\right)} \\ &= \log 2 > 0\end{aligned}$$

□

定理 5（关于 $\mathrm{d_H}$ 的紧下界，其中 $\mathrm{d}^{||}$ 仿射不变黎曼度量）对于任意 $X,\ Y \in \mathrm{VPM}^{\circ}(n)$ 都有一个紧界：

$$\frac{1}{\sqrt{n}}\, \mathrm{d}^{||}_{AIRM}(X, Y) \leq \mathrm{d_H}(X, Y)$$

证明： 我们首先证明该界是紧的。取某个 $\varepsilon > 0,\ c > 1$，使得 $0 < \varepsilon < \frac{1}{c}$，并定义：

$$X_\varepsilon = \varepsilon I \qquad Y_\varepsilon = \varepsilon c I$$

我们计算仿射不变黎曼距离：

$$\mathrm{d}^{||}_{\mathrm{AIRM}}(X_\varepsilon, Y_\varepsilon) = \sqrt{\sum_{i=1}^{n} \log^2 \lambda_i(Y_\varepsilon^{-1} X_\varepsilon)} = \sqrt{n} \log c$$

对于希尔伯特距离，我们有：

$$\lambda_{\min}(Y_\varepsilon^{-1}X_\varepsilon) = \lambda_{\max}(Y_\varepsilon^{-1}X_\varepsilon) = \frac{1}{c} \qquad \lambda_{\min}((I-Y_\varepsilon)^{-1}(I-X_\varepsilon)) = \lambda_{\max}((I-Y_\varepsilon)^{-1}(I-X_\varepsilon)) = \frac{1-\varepsilon}{1-\varepsilon c}$$

因此，我们代入：

$$\begin{aligned}\lim_{\varepsilon\to 0}\mathrm{d_H}(X_\varepsilon,Y_\varepsilon)&=\lim_{\varepsilon\to 0}\log\frac{\max\left(\lambda_{\max}(Y_\varepsilon^{-1}X_\varepsilon),\lambda_{\max}\big((I-Y_\varepsilon)^{-1}(I-X_\varepsilon)\big)\right)}{\min\left(\lambda_{\min}(Y_\varepsilon^{-1}X_\varepsilon),\lambda_{\min}\big((I-Y_\varepsilon)^{-1}(I-X_\varepsilon)\big)\right)}\\&=\log\frac{\max\left(\frac{1}{c},\lim_{\varepsilon\to 0}\frac{1-\varepsilon}{1-\varepsilon c}\right)}{\min\left(\frac{1}{c},\lim_{\varepsilon\to 0}\frac{1-\varepsilon}{1-\varepsilon c}\right)}=\log\frac{1}{\frac{1}{c}}=\log c\end{aligned}$$

由此可得所需结果：

$$\lim_{\varepsilon\to 0}\mathrm{d}^{||}_{\mathrm{AIRM}}(X_\varepsilon,Y_\varepsilon)=\sqrt{n}\log c=\lim_{\varepsilon\to 0}\sqrt{n}\,\mathrm{d_H}(X_\varepsilon,Y\varepsilon)$$

现在证明该界对所有 X， $Y\in\mathsf{VPM}^{\circ}(n)$ 都成立。对于仿射不变黎曼距离，由引理2可得上界：

$$\mathrm{d}^{||}_{\mathrm{AIRM}}(X,Y)=\sqrt{\sum_{i=1}^{n}\log^2\lambda_i(Y^{-1}X)}\le\sqrt{n}\max_i|\log\lambda_i(Y^{-1}X)|$$

对于希尔伯特距离，根据命题 3 中的刻画，我们也有：

$$\mathrm{d_H}(X,Y)=\log M(\hat{X},\hat{Y})+\log M(\hat{Y},\hat{X})$$

并计算得：

$$\log M(\hat{X},\hat{Y})=\log\max\left(\lambda_{\max}(Y^{-1}X),\lambda_{\max}((I-Y)^{-1}(I-X))\right)\ge\log\lambda_{\max}(Y^{-1}X)$$

类似地：

$$\log M(\hat{Y},\hat{X})\ge\log\lambda_{\max}(X^{-1}Y)=-\log\frac{1}{\lambda_{\max}(X^{-1}Y)}=-\log\lambda_{\min}(Y^{-1}X)$$

因此，综合以上结果可得：

$$\begin{aligned}\mathrm{d_H}(X,Y)&\ge\log M(\hat{X},\hat{Y})+\log M(\hat{Y},\hat{X})\\&\ge\max\left(\log\lambda_{\max}(Y^{-1}X),-\log\lambda_{\min}(Y^{-1}X)\right)\\&=\max_i|\log\lambda_i(Y^{-1}X)|\\&\ge\frac{1}{\sqrt{n}}\,\mathrm{d}^{||}_{\mathrm{AIRM}}(X,Y)\end{aligned}$$

从而得到所需的不等式。 □

4.3 前推AIRM距离的界 $\mathrm{d}^{\rightarrow}_{\mathbf{AIRM}}$

现在，对于 $\mathsf{d}^{\rightarrow}_{\mathrm{AIRM}}$，情况正好相反——我们有 d_{H} 的上界，但没有下界。首先注意到，在维数为 1 时，希尔伯特距离与推前的 $\mathsf{d}^{\rightarrow}_{\mathrm{AIRM}}$ 距离相同。

注 5 我们有 $\mathsf{d}_{\overrightarrow{AIRM}}(X,\ Y) = \mathsf{d}_{\mathrm{H}}(X,\ Y)$ 对于 $X,\ Y \in \mathrm{VPM}^{\circ}(1) = (0,\ 1)$。

为了证明在维数 $n > 1$ 下不存在界，我们需要给出二维空间中某些正定矩阵的特征值和迹的若干闭式表达式，因为这将构成我们的反例（随后我们会将其推广到维数 $n > 2$。这些表达式略显技术性。

引理 6（单位行列式矩阵的特征值） 设 $P, Q \in$ 正定矩阵(2)，满足 行列式 $P =$行列式 Q。则， $Q^{-1}P$ 的特征值为：

$$\lambda_{\pm} = \exp(\pm\delta) \qquad \delta = \operatorname{arccosh}\left(\frac{1}{2}\operatorname{tr}(Q^{-1}P)\right)$$

证明： 记 $A = Q^{-1}P$。由于 $\det P = \det Q$，根据行列式是幺半群映射这一事实，我们有 $\det A = 1$。因此，我们知道 A 的（无序）特征值必为 λ、$\frac{1}{\lambda}$，其中 $\lambda \geq 1$。定义 $\delta = \log \lambda > 0$。我们利用 $\operatorname{tr}(A) = \lambda + \frac{1}{\lambda} = 2\frac{e^{\delta}+e^{-\delta}}{2} = 2\cosh(\delta)$ 这一事实，并注意到 cosh 在 $[0, \infty)$ 上严格递增，因此逆是唯一的。□

引理 7（共轭乘法公式） 令 $D =$ 对角矩阵(a, b) 为正定对角矩阵，并令 $U = \begin{bmatrix} \text{余弦 } \theta_t & -\text{正弦 } \theta_t \\ \text{正弦 } \theta_t & \text{余弦 } \theta_t \end{bmatrix}$ 为旋转角为 θ 的旋转矩阵。则有以下公式：

$$\frac{1}{2}\operatorname{tr}(U^T D^{-1} U D) = \cos^2\theta + \frac{\sin^2\theta}{2}\left(\frac{a}{b} + \frac{b}{a}\right)$$

证明：

$$\begin{aligned}
\frac{1}{2}\operatorname{tr}(U^T D^{-1} U D) &= \frac{1}{2}\operatorname{tr}\left(\begin{bmatrix}\cos\theta & -\sin\theta \\ \sin\theta & \cos\theta\end{bmatrix}\begin{bmatrix}\frac{1}{a} & 0 \\ 0 & \frac{1}{b}\end{bmatrix}\begin{bmatrix}\cos\theta & \sin\theta \\ -\sin\theta & \cos\theta\end{bmatrix}\begin{bmatrix}a & 0 \\ 0 & b\end{bmatrix}\right) \\
&= \frac{1}{2}\operatorname{tr}\left(\begin{bmatrix}\cos^2\theta\frac{1}{a} + \sin^2\theta\frac{1}{b} & \cos\theta\sin\theta\left(\frac{1}{a} - \frac{1}{b}\right) \\ \cos\theta\sin\theta\left(\frac{1}{a} - \frac{1}{b}\right) & \sin^2\theta\frac{1}{a} + \cos^2\theta\frac{1}{b}\end{bmatrix}\begin{bmatrix}a & 0 \\ 0 & b\end{bmatrix}\right) \\
&= \cos^2\theta + \frac{\sin^2\theta}{2}\left(\frac{a}{b} + \frac{b}{a}\right)
\end{aligned}$$

□

引理 8 令 $X, Y \in \mathrm{VPM}^{\circ}(2)$ 满足 $\operatorname{tr}(X) = \operatorname{tr}(Y) = 1$ 且 $\det(X) = \det(Y)$。则 ***both*** $Y^{-1}X$ ***and*** $(I - Y)^{-1}(I - X)$ 具有相同的谱。

证明： 根据 Cayle y-Hamilton 定理，每个 2×2矩阵 M都满足其特征多项式：

$$M^2 - \operatorname{tr}(M)M + (\det M)I = 0$$

我们将此应用于 X 和 Y，并考虑到 $\operatorname{tr}(X) = \operatorname{tr}(Y) = 1$:

$$X^2 - X + (\det X)I = 0 \qquad Y^2 - Y + (\det Y)I = 0$$

因此，重新排列得：

$$I - X = (\det X)X^{-1} \qquad I - Y = (\det Y)Y^{-1}$$

因此，我们有：

$$(I-Y)^{-1}(I-X)=\frac{1}{\det Y}Y(\det X)X^{-1}=YX^{-1}$$

因为它们的行列式被假设为相等。但现在，由于我们有 $\det YX^{-1}=1$，且我们处于 Sym(2) 中，这个矩阵有两个特征值 $\{\lambda,1/\lambda\}$。因此，其逆矩阵 $Y^{-1}X$ 也具有相同的特征值 $\{1/\lambda,\lambda\}$。□

命题12（不存在关于$\mathrm{d}^{\rightarrow}_{\mathbf{AIRM}}$的下界）不存在常数 $\kappa>0$，使得对于所有 $X,Y\in \mathrm{VPM}^{\circ}(n)$，都能得到$\mathrm{d}_{\mathrm{H}}$ 的如下下界：

$$\kappa\cdot \mathrm{d}^{\rightarrow}_{AIRM}(X,Y)\leq \mathrm{d}_{\mathrm{H}}(X,Y)$$

证明：和之前一样，我们给出一个元素序列 $P_t,Q_t\in$ PD(n)，使得：

$$\lim_{t\to\infty}\mathrm{d}_{\mathrm{AIRM}}(P_t,Q_t)=\lim_{t\to\infty}\mathrm{d}^{\rightarrow}_{\mathrm{AIRM}}(\iota(P_t),\iota(Q_t))>0\qquad \lim_{t\to\infty}\mathrm{d}_{\mathrm{H}}(\iota(P_t),\iota(Q_t))=0$$

如前所述，我们先给出维数为 $n=2$ 时的序列，并在最后将其推广到任意 n。定义以下矩阵：

$$P_t=\begin{bmatrix}t & 0\\ 0 & \frac{1}{t}\end{bmatrix}\qquad Q_t=U_tP_tU_t^T\qquad U_t=\begin{bmatrix}\cos\theta_t & -\sin\theta_t\\ \sin\theta_t & \cos\theta_t\end{bmatrix}\qquad \theta_t=\frac{1}{t}$$

由于 P_t 和 Q_t 仅相差一个旋转共轭，它们具有相同的行列式。因此，根据引理 6，我们知道 $Q_t^{-1}X$ 的特征值具有如下形式：

$$\lambda_{\pm}=\exp\left(\pm\operatorname{arccosh}\frac{1}{2}\mathrm{tr}(Q_t^{-1}X)\right)$$

利用引理 7 中的公式，并代入对角矩阵 $P_t=$ 对角矩阵(t, $1/t$)，得到：

$$\begin{aligned}\frac{1}{2}\mathrm{tr}(Q_t^{-1}X)&=\frac{1}{2}\mathrm{tr}(U_tX_t^{-1}U_t^TP_t)\\ &=\cos^2\theta_t+\frac{\sin^2\theta_t}{2}\left(t^2+\frac{1}{t^2}\right)\\ &=(1-\sin^2\theta_t)+\frac{\sin^2\theta_t}{2}\left(t^2+\frac{1}{t^2}\right)\\ &=1+\frac{\sin^2\frac{1}{t}}{2}\left(t^2+\frac{1}{t^2}-1\right)\end{aligned}$$

根据泰勒展开，我们知道 $\sin^2\frac{1}{t}=\frac{1}{t^2}+O\left(\frac{1}{t^4}\right)$，得到：

$$\lim_{t\to\infty}1+\frac{\sin^2\theta_t}{2}\left(t^2+\frac{1}{t^2}-1\right)=\lim_{t\to\infty}1+\frac{1}{2}\left(\frac{1}{t^2}+o\left(\frac{1}{t^4}\right)\right)\left(t^2+\frac{1}{t^2}-1\right)=\frac{3}{2}$$

因此，将其代入仿射不变黎曼度量的公式可得：

$$\mathrm{d}^{||}_{\mathrm{AIRM}}(P_t,Q_t)=\sqrt{\log^2\lambda_++\log^2(1/\lambda_+)}=\sqrt{2}\left|\operatorname{arccosh}\left(\frac{3}{2}\right)\right|>0$$

我们将通过 James 映射推送得到的矩阵记为 P_t, Q_t：

$$X_t = \iota(P_t) = P_t(I + P_t)^{-1} \qquad Y_t = \iota(Q_t) = Q_t(I + Q_t)^{-1}$$

根据定义，$\mathsf{d}^{\rightarrow}_{\mathrm{AIRM}}(X_t,\ Y_t) = \mathsf{d}^{||}$ 仿射不变黎曼度量$(P_t,\ Q_t)$。

另一方面，现在我们明确计算 $X_t,\ Y_t$ 的希尔伯特距离。首先：

$$X_t = \begin{bmatrix} t & 0 \\ 0 & \frac{1}{t} \end{bmatrix} \begin{bmatrix} \frac{1}{1+t} & 0 \\ 0 & \frac{t}{t+1} \end{bmatrix} = \begin{bmatrix} \frac{t}{1+t} & 0 \\ 0 & \frac{1}{t+1} \end{bmatrix}$$

通过这一直接计算，我们得到 $\mathrm{tr}(X_t) = 1$。根据 [20, 备注 20], ，我们知道旋转可以与 James 映射互换，得到 $Y_t = U_t X_t U_t^T$；反演也同样如此，得到 $I - Y_t = I - U_t X_t U_t^T = U_t(I - X_t)U_t^T$。由于它们相差一个旋转共轭，因此具有相同的行列式和迹，据此我们应用引理 8，得到 $(I - Y_t)^{-1}(I - X_t)$ 和 $\lambda_{\pm}(Y_t^{-1}X_t$ 的谱相同。根据引理 6，它们的特征值为：

$$\lambda_{\pm}((I - Y_t)^{-1}(I - X_t)) = \lambda_{\pm}(Y_t^{-1}X_t) = e^{\pm\delta_t} \qquad \delta_t = \operatorname{arccosh}\left(\frac{1}{2}\mathrm{tr}(Y_t^{-1}X_t)\right) > 0$$

现在来看 VPM$^\circ$ 距离公式：

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

由于 $(I - Y_t)^{-1}(I - X_t)$ 与 $Y_t^{-1}X_t$ 的谱相同，min,max 中的因子也相同，因此可以将比值简化为：

$$\mathrm{d_H}(X_t, Y_t) = \log \frac{e^{\delta_t}}{e^{-\delta_t}} = 2\delta_t$$

最后一步是证明 $\lim_{t\to\infty} \delta_t = 0$：

$$\begin{aligned} \lim_{t\to\infty} \operatorname{arccosh}\left(1 + \frac{\sin^2\theta_t}{2}\left(t + \frac{1}{t} - 1\right)\right) &= \operatorname{arccosh}\left(1 + \lim_{t\to\infty}\frac{1}{2}\left(\frac{1}{t^2} + O\left(\frac{1}{t^4}\right)\right)\left(t + \frac{1}{t} - 1\right)\right) \\ &= \operatorname{arccosh}(1) = 0 \end{aligned}$$

最后，为了 将论证推广到一般的 $n > 2$，我们定义分块对角矩阵：

$$A_t^n = X_t \oplus \frac{1}{2}I_{n-2} \qquad B_t^n = Y_t \oplus \frac{1}{2}I_{n-2}$$

于是 $A_t^n, B_t^n \in \mathrm{VPM}^\circ(n)$。对于 $\mathsf{d}^{\rightarrow}_{\mathrm{AIRM}}$，这些额外维数不会对求和产生贡献（因为取对数后它们相互抵消，结果为零），因此 $\lim_{t\to\infty} \mathsf{d}^{\rightarrow}_{\mathrm{AIRM}}(A_t^n, B_t^n) = \lim_{t\to\infty} \mathsf{d}^{\rightarrow}_{\mathrm{AIRM}}(X_t, Y_t) > 0$。对于希尔伯特距离，我们有：

$$(B_t^n)^{-1}A_t^n = (Y_t^{-1}X_t) \oplus I_{n-2} \qquad (I - B_t^n)^{-1}(I - A_t^n) = (I - Y_t)^{-1}(I - X_t) \oplus I_{n-2}$$

由于 $\delta_t > 0$，我们有 $e^{\delta_t} \geq 1 \geq e^{-\delta_t}$，因此最小值和最大值仍在与之前相同的元素处取得，这也使 $\mathrm{d_H}$ 保持不变。□

另一方面，证明依赖于几个映射的复合，而我们将分别证明这些映射具有带有明确常数的利普希茨性质。具体而言，对于定义 3 中的帽嵌入映射，我们将证明两次：一次是在 VPM$^\circ$ 赋予希尔伯特距离时，另一次是在赋予对数障碍距离时。

引理 9（帽嵌入是 $\sqrt{2}$-余利普希茨）如果在定义域 $\mathrm{VPM}^{\circ}(n)$ 上取希尔伯特距离，并在陪域 $\mathrm{PD}(n) \times \mathrm{PD}$ 上取乘积 ***AIRM*** 距离，则帽嵌入 $(\hat{}) : \mathrm{VPM}^{\circ}(n) \to \mathrm{PD}(n) \times \mathrm{PD}(n)$ 是$\sqrt{2}$-余利普希茨的。

证明：对于任意 X，$Y \in \mathrm{VPM}^{\circ}(n)$，我们需要证明：

$$\mathrm{d_H}(X,Y) \leq \sqrt{2}\sqrt{\left(\mathrm{d}^{||}_{\mathrm{AIRM}}(X,Y)\right)^2 + \left(\mathrm{d}^{||}_{\mathrm{AIRM}}(I-X,I-Y)\right)^2}$$

根据命题 4，我们知道：

$$\mathrm{d_H}(X,Y) = \mathbf{spread}(\log P)$$

对于矩阵：

$$P = \begin{bmatrix} Y^{-1}X & 0 \\ 0 & (I-Y)^{-1}(I-X) \end{bmatrix} \in \mathrm{PD}(2n)$$

因此，我们可以将范围-l_2 不等式（引理 3）应用于 $\log\ P$ 的特征值向量。由于 $\log\ P$ 是对称矩阵，且 $||\log\ P||_F^2$ 是其特征值的平方和，因此我们得到：

$$\begin{aligned} \mathrm{d_H}^2(X,Y) \leq 2||\log P||_F^2 &= 2(||\log Y^{-1}X||_F^2 + ||\log(I-Y)^{-1}(I-X)||_F^2) \\ &= 2\left(\left(\mathrm{d}^{||}_{\mathrm{AIRM}}(X,Y)\right)^2 + \left(\mathrm{d}^{||}_{\mathrm{AIRM}}(I-X,I-Y)\right)^2\right) \end{aligned}$$

□

引理 10（无穷小不等式蕴含 κ-利普希茨）设 F：$\mathcal{M} \to \mathcal{N}$ 是芬斯勒流形 $(\mathcal{M}, g)$ 与 $(\mathcal{N}, h)$ 之间的光滑映射，其诱导距离分别为 d_g 和 d_h，并假设存在常数 $\kappa > 0$，使得对于任意 $p \in \mathcal{M}$ 和任意 $v \in T_p\mathcal{M}$，都有以下逐点不等式：

$$||dF_p(v)||_{F(p)} \leq \kappa||v||_p$$

那么 F 是 κ-利普希茨的，也就是说，对于所有 $x, y \in \mathcal{M}$，我们有：

$$d_h(F(x),F(y)) \leq \kappa d_g(x,y)$$

证明：令 γ：$[0,1] \to \mathcal{M}$ 为一条分段 C^1 曲线。根据链式法则，$(F \circ \gamma)'(t) = dF_{\gamma(t)}(\gamma'(t))$，因此根据假设：

$$||(F \circ \gamma)'(t)||_{F(\gamma(t))} \leq \kappa||\gamma'(t)||_{\gamma(t)}.$$

对 t 积分，得到 $L_h(F \circ \gamma) \leq \kappa L_g(\gamma)$。对所有连接 x 与 y 的分段 C^1 曲线 γ 取下确界，得到 $d_h(F(x)\ F(y)) \leq \kappa d_g(x,y)$。□

引理 11（帽嵌入是等距嵌入）考虑配备对数障碍度量 d_Ψ 的$\mathrm{VPM}^{\circ}(n)$，以及配备乘积 ***AIRM*** 度量的正定矩阵(n) ×正定矩阵(n)，帽嵌入是等距浸入。

证明： 将乘积 AIRM 黎曼度量记为 $g^{\times} = \mathrm{g}^{\mathrm{AIRM}} \times \mathrm{g}^{\mathrm{AIRM}}$。映射 $\hat{}$ (^) 是仿射的，因此 $d\hat{X}(V) = (V, -V)$。因此：

$$\begin{aligned} g^{\times}_{\hat{X}}\big(d\hat{X}(V), d\hat{X}(W)\big) &= \mathrm{g}^{\mathrm{AIRM}}{}_X(V, W) + \mathrm{g}^{\mathrm{AIRM}}{}_{I-X}(-V, -W) \\ &= \mathrm{tr}(X^{-1}VX^{-1}W) + \mathrm{tr}((I-X)^{-1}V(I-X)^{-1}W) \end{aligned}$$

这正是 $g^{\Psi}_X(V, W)$。根据引理 10，随即可得 1-Lipschitz 性质。 □

推论 4（帽嵌入是 1-Lipschitz 的） 考虑赋予对数障碍度量 d_{Ψ} 的 $\mathrm{VPM}^{\circ}(n)$，以及赋予乘积 *AIRM* 度量的 $\mathrm{PD}(n) \times \mathrm{PD}(n)$，帽嵌入是 1-*Lipschitz* 的。

引理 12（坐标中的 AIRM 计算） 设 $P = \mathrm{diag}(p_1, \ldots p_n) \in \mathrm{PD}(n)$ 且 $V = (v_{ij}) \in \mathrm{Sym}(n)$。则：

$$\mathrm{tr}(P^{-1}VP^{-1}V) = \sum_{i,j} \frac{v_{ij}^2}{p_i p_j}$$

证明： 我们将矩阵 $P^{-1}VP^{-1}$ 写在对角 P 基下：

$$(P^{-1}VP^{-1})_{ij} = \sum_{k,l} (P^{-1})_{ik} V_{kl} (P^{-1})_{lj} = p_i^{-1}\, v_{ij}\, p_j^{-1} = \frac{v_{ij}}{p_i p_j}.$$

现在计算迹：

$$\mathrm{tr}(P^{-1}VP^{-1}V) = \sum_i ((P^{-1}VP^{-1})V)_{ii} = \sum_{i,j} (P^{-1}VP^{-1})_{ij} V_{ji} = \sum_{i,j} \frac{v_{ij}}{p_i p_j} v_{ji} = \sum_{i,j} \frac{v_{ij}^2}{p_i p_j}$$

□

引理 13（James 映射是映入对数障碍几何的 1-Lipschitz 映射） *James* 映射 ι：$\mathrm{PD}(n) \to \mathrm{VPM}^{\circ}(n)$ 是 1-*Lipschitz* 的，其中我们考虑带有 $\mathrm{d}_{\mathrm{AIRM}}$ 距离的 $\mathrm{PD}(n)$，以及带有对数障碍距离 d_{Ψ} 的 $\mathrm{VPM}^{\circ}(n)$。

Proof: 我们固定 $P \in$正定矩阵(n)与 $V \in$ 对称矩阵(n)。由命题 2，我们有：

$$d\iota_P(V) = AVA \qquad A = (I+P)^{-1}$$

记 $S = d\iota_P(V)$。因此，我们需要证明如下逐点估计：

$$||d\iota_P(V)||^{\Psi}_{\iota(P)} = ||S||^{\Psi}_{\iota(P)} \leq \|V\|^{\mathrm{AIRM}}_P$$

然后由引理 10 ,这将意味着 ply 即 ι 是1-利 p普希茨，从 $(\mathrm{PD}(n), \mathrm{d}^{\text{仿射不变黎曼度量}})$ to $(\mathrm{VPM}^{\circ}(n), d_{\Psi})$.

不等式两边在正交共轭下保持不变 $P \mapsto U^T PU$，且 $V \mapsto U^T VU$：仿射不变黎曼度量和 Ψ 在共轭下均保持不变（命题 9），$\iota(P)$ 是等变的（[20,备注 20]），而 S 也是等变的：

$$U^T SU = (U^T AU)(U^T VU)(U^T AU) = (I + U^T PU)^{-1}(U^T VU)(I + U^T PU)^{-1}$$

因此，不失一般性， $P=$对角矩阵(p_1，.....， p_n)，其中 $p_i>0$。接着令 $X=\iota(P)$，并且：

$$A=\operatorname{diag}(a_1,\ldots,a_n),\quad a_i=\frac{1}{1+p_i},\qquad X=\operatorname{diag}(x_1,\ldots,x_n),\quad x_i=\frac{p_i}{1+p_i}=1-a_i.$$

其中我们还注意到 $0<a_i<1$。在此基下写出 $V=(v_{ij})$，并将引理 12 应用于点对 $(P,\ V)$，可得：

$$\left(\|V\|_P^{\mathrm{AIRM}}\right)^2=\operatorname{tr}(P^{-1}VP^{-1}V)=\sum_{i,j}v_{ij}^2/p_ip_j=\sum_{i,j}b_{ij}$$

其中我们定义了 $b_{ij}=v_{ij}^2/p_ip_j$。此外，$S=AVA$ 的元素为 $s_{ij}=a_ia_jv_{ij}$，因此再次将引理 12 应用于点对 $(X,\ S)$，得到：

$$\operatorname{tr}(X^{-1}SX^{-1}S)=\sum_{i,j}\frac{s_{ij}^2}{x_ix_j}=\sum_{i,j}\frac{a_i^2a_j^2}{(p_ia_i)(p_ja_j)}\,v_{ij}^2=\sum_{i,j}a_ia_j\,b_{ij}$$

类似地

$$\operatorname{tr}((I-X)^{-1}S(I-X)^{-1}S)=\operatorname{tr}(A^{-1}SA^{-1}S)=\sum_{i,j}\frac{s_{ij}^2}{a_ia_j}=\sum_{i,j}(p_ia_i)(p_ja_j)\,b_{ij}=\sum_{i,j}x_ix_j\,b_{ij}$$

因此，利用柯西-施瓦茨不等式以及 $\sqrt{a_i^2+(1-a_i)^2}=\sqrt{1-2a_i(1-a_i)}<1$ 这一事实：

$$\begin{aligned}\left(||S||_X^{\Psi}\right)^2&=\sum_{i,j}(a_ia_j+x_ix_j)b_{ij}\\&=\sum_{i,j}(a_ia_j+(1-a_i)(1-a_j)b_{ij}\\&=\sum_{i,j}\langle(a_i,1-a_i),(a_j,1-a_j)\rangle b_{ij}\\&\le\sum_{i,j}||(a_i,1-a_i)||_2||(a_j,1-a_j)||_2b_{ij}\\&\le\sum_{i,j}\sqrt{a_i^2+(1-a_i)^2}\sqrt{a_j^2+(1-a_j)^2}b_{ij}\\&\le\sum_{i,j}b_{ij}=\left(\|V\|_P^{\mathrm{AIRM}}\right)^2\end{aligned}$$

□

推论 *5* 设 *PD(n)* 配备 $d_{AIRM}^{||}$ 距离，且 *PD(n)* $\times$*PD(n)* 配备相应的乘积距离。那么，由 $F(P)=(\iota(P),(I+P)^{-1})$ 给出的映射 $F:PD(n)\to PD(n)\times PD(n)$ 是 *1-Lipschitz*。

证明： 由引理 13 和推论 4 的组合可得。 □

定理 6（$\mathrm{d_H}$ **与** $\mathrm{d}^{\rightarrow}_{\mathbf{AIRM}}$**的紧上界**）对于任意 $X,Y\in\mathrm{VPM}^{\circ}(n)$ 我们有一个紧界

$$\mathrm{d_H}(X,Y)\le\sqrt{2}\,\mathrm{d}^{\rightarrow}_{AIRM}(X,Y)$$

证明：记作：

$$P = \iota^{-1}(X) = X(I-X)^{-1} \qquad Q = \iota^{-1}(Y) = Y(I-Y)^{-1}$$

首先与引理 9 中的不等式复合，然后再与推论 5 中的不等式复合，可得：

$$\begin{aligned}\frac{1}{\sqrt{2}}\,\mathrm{d_H}(X,Y) &\le \sqrt{\left(\mathrm{d}^{||}_{\mathrm{AIRM}}(X,Y)\right)^2 + \left(\mathrm{d}^{||}_{\mathrm{AIRM}}(I-X,I-Y)\right)^2} \\ &= \sqrt{\left(\mathrm{d}^{||}_{\mathrm{AIRM}}(\iota(P),\iota(Q))\right)^2 + \left(\mathrm{d}^{||}_{\mathrm{AIRM}}(I-\iota(P),I-\iota(Q))\right)^2} \\ &\le \mathrm{d}_{\mathrm{AIRM}}(P,Q) = \mathrm{d}_{\mathrm{AIRM}}(\iota^{-1}(X),\iota^{-1}(Y)) = \mathrm{d}^{\rightarrow}_{\mathrm{AIRM}}(X,Y)\end{aligned}$$

□

4.4 利用对数障碍度量进行界定

对于对数障碍度量，我们有下界和上界。

定理 7（关于 $\|\cdot\|^{\Psi}$ 的紧上界）对于所有 $X \in \mathrm{VPM}^{\circ}(n)$ 以及所有 $V \in T_X\,\mathrm{VPM}^{\circ}(n) \simeq \mathrm{Sym}(n)$，我们有：

$$\|V\|^H_X \le \sqrt{2}\,\|V\|^{\Psi}_X$$

证明： 根据命题 8，障碍度量范数为：

$$\left(||V||^{\Psi}_X\right)^2 = \mathrm{tr}(X^{-1}VX^{-1}V) + \mathrm{tr}((X-I)^{-1}V(X-I)^{-1}V) = ||A||^2_F + ||B||^2_F$$

其中：

$$A = X^{-1/2}VX^{-1/2} \qquad B = (I-X)^{-1/2}V(I-X)^{-1/2}$$

另一方面，根据命题 5，我们将切空间 $T_X\,\mathrm{VPM}^{\circ}(n)$ 上的希尔伯特度量写为：

$$||V||^H_X = \frac{1}{t^+(X,V)} + \frac{1}{t^-(X,V)}$$

其中：

$$\begin{aligned}\frac{1}{t^+(X,V)} &= \max\left(\max(0,-\lambda_{\min}(A)),\max(0,\lambda_{\max}(B))\right) \\ \frac{1}{t^-(X,V)} &= \max\left(\max(0,\lambda_{\max}(A)),\max(0,-\lambda_{\min}(B))\right)\end{aligned}$$

这一点已在命题 6 中证明。现在，令：

$$u = \max(0,-\lambda_{\min}(A)) \qquad p = \max(0,\lambda_{\max}(A)) \qquad v = \max(0,\lambda_{\max}(B)) \qquad q = \max(0,-\lambda_{\min}(B))$$

那么，$||V||^H_X = \max(u,v) + \max(p,q)$。注意，我们始终有：

$$||A||^2_F = \sum_i \lambda_i(A)^2 \ge \lambda_{\max}(A)^2 + \lambda_{\min}(A)^2 \ge p^2 + u^2 \qquad ||B||^2_F \ge v^2 + q^2$$

因此，我们比较：

$$\begin{aligned}||V||_X^H &= \max(u,v) + \max(p,q) \\ &\le \sqrt{u^2+v^2} + \sqrt{p^2+q^2} \\ &\le \sqrt{2}\sqrt{u^2+v^2+p^2+q^2} \\ &\le \sqrt{2}\sqrt{||A||_F^2 + ||B||_F^2} = \sqrt{2}||V||_X^\Psi\end{aligned}$$

其中我们在第一个不等式中使用了 $\max(a,b) \le \sqrt{a^2+b^2}$。

为证明该界是紧的，我们取 $X = \frac{1}{2}I_n$ 和 $V = e_1 e_1^T$。由此得到：

$$A = B = 2e_1 e_1^T$$

因此 $||V||_X^\Psi = 2(||A||_X^\Psi) = 2\sqrt{2}$。另一方面，$A$ 和 B 仅有一个非零特征值 $\lambda_{\max} = 2$，故 $||V||_X^H = 2+2 = 4$。因此，该比值为 $\sqrt{2}$ □

定理 8（带有 $\|\cdot\|^\Psi$ 的紧下界） 对所有 $X \in \mathrm{VPM}^\circ(n)$ 以及所有 $V \in T_X \mathrm{VPM}^\circ(n) \simeq \mathrm{Sym}(n)$，我们有：

$$||V||_X^\Psi \le \sqrt{n}||V||_X^H$$

证明： 与上文定理 7 的证明类似，我们有：

$$A = X^{-1/2} V X^{-1/2} \qquad B = (I-X)^{-1/2} V (I-X)^{-1/2}$$

因此：

$$\left(||V||_X^\Psi\right)^2 = \mathrm{tr}(X^{-1}VX^{-1}V) + \mathrm{tr}((X-I)^{-1}V(X-I)^{-1}V) = ||A||_F^2 + ||B||_F^2$$

和之前一样，我们进行如下操作：

$$u = \max(0, -\lambda_{\min}(A)) \qquad p = \max(0, \lambda_{\max}(A)) \qquad v = \max(0, \lambda_{\max}(B)) \qquad q = \max(0, -\lambda_{\min}(B))$$

其中 $||V||_X^H = \max(u,v) + \max(p,q)$。现在考虑 A 的非负特征值数量和负特征值数量，分别记为 r^+ 和 r^-。我们注意到，对于 B，这些数量是相同的；根据西尔维斯特惯性定律，这是因为 A 和 B 与同一个矩阵 V 合同。因此，我们有：

$$\|A\|_F^2 \le r^- u^2 + r^+ p^2 \qquad \|B\|_F^2 \le r^+ v^2 + r^- q^2$$

同样地，对于希尔伯特范数 $||V||_X^H = \max(u,v) + \max(p,q)$，我们有以下界：

$$\|V\|_X^H \ge u + p \qquad \|V\|_X^H \ge v + q$$

由此得到最终不等式：

$$\begin{aligned}\left(||V||_X^\Psi\right)^2 &= ||A||_F^2 + ||B||_F^2 \\ &\le r^- u^2 + r^+ p^2 + r^+ v^2 + r^- q^2 \\ &\le r^-(u^2+q^2) + r^+(v^2+p^2) \\ &\le r^-(u+q)^2 + r^+(v+p)^2 \\ &\le r^- \left(||V||_X^H\right)^2 + r^+ \left(||V||_X^H\right)^2 \\ &\le (r^- + r^+)\left(||V||_X^H\right)^2 = n\left(||V||_X^H\right)^2\end{aligned}$$

距离	上界常数	下界常数
$\mathrm{d}^{\parallel}_{\mathrm{AIRM}}$	✗	$\frac{1}{\sqrt{n}}$
$\mathrm{d}^{\rightarrow}_{\mathrm{AIRM}}$	$\sqrt{2}$	✗
d_{Ψ}	$\sqrt{2}$	$\frac{1}{\sqrt{n}}$

表1：希尔伯特距离下界/上界汇总。常数均相对于 d_{H} 给出。所有界都是紧的。

为了证明它是紧的，我们令 $X = \varepsilon I$ 对某个 $\varepsilon > 0$，且 $V = I$。则：

$$A = \frac{1}{\varepsilon} I \qquad B = \frac{1}{1-\varepsilon} I$$

and:

$$||V||_X^H = \max(u, v) + \max(p, q) = \frac{1}{\varepsilon} + \frac{1}{1-\varepsilon} = \frac{1}{\varepsilon(1-\varepsilon)}$$

$$||V||_X^{\Psi} = \sqrt{n\frac{1}{\varepsilon^2} + n\frac{1}{(1-\varepsilon)^2}} = \sqrt{n}\frac{1}{\varepsilon(1-\varepsilon)}\sqrt{((1-\varepsilon)^2 + \varepsilon^2)}$$

取极限时：

$$\lim_{\varepsilon \to 0} \frac{||V||_X^{\Psi}}{||V||_X^H} = \lim_{\varepsilon \to 0} \sqrt{n}((1-\varepsilon)^2 + \varepsilon^2) = \sqrt{n}$$

□

推论 6（关于 d_{Ψ} 的紧下界和上界）对于 X，$Y \in \mathrm{VPM}^{\circ}(n)$，有：

$$\frac{1}{\sqrt{2}}\,\mathrm{d}_{\mathrm{H}}(X, Y) \leq \mathrm{d}_{\Psi}(X, Y) \leq \sqrt{n}\,\mathrm{d}_{\mathrm{H}}(X, Y)$$

证明：直接由将引理 10 应用于定理 7 和定理 8 得出。 □

5 总结与讨论

在本文第一部分，我们通过考察 James 双锥域 [18]，介绍了对称正定锥流形上的两种新微分结构：即由双锥上的希尔伯特距离诱导的芬斯勒结构，以及由双对数行列式函数诱导的对偶平坦结构 [2]。双对数行列式函数是锥对数行列式函数在 James 双锥域上的自然推广。特别地，我们证明了希尔伯特 VPM 距离推广了希尔伯特单纯形距离 [27]，因为开谱单纯形是双锥域的子集。此外，在希尔伯特距离所对应的芬斯勒几何中，开谱单纯形是一个全测地子流形。我们给出了标准单纯形和 VPM 域中希尔伯特测地线的显式恒速参数化。在第二部分，我们将仿射不变黎曼度量（AIRM，也称迹度量）所诱导的距离和 Bregman 对数行列式散度，与新定义的 VPM 希尔伯特距离和 VPM Bregman 双对数行列式散度进行了比较。

在对称正定锥流形的 James 的 VPM 域中定义并研究不相似性，有望应用于许多领域。希尔伯特 VPM 距离仅取决于极端特征值

可将其解释为最坏方向畸变度量，因此在本质上有别于仿射不变几何或对数欧几里得几何。例如，VPM 希尔伯特距离或可用于鲁棒控制、李雅普诺夫理论或黎卡提方程，其中映射 $P(X)$ 将特征值归一化到 $(0,\ 1)$ 内，见 [21, 19]。另一个自然的应用领域是量子信息论 [28]，其中效应矩阵属于闭 VPM 域 [23]。希尔伯特 VPM 距离在 $J \to 0$ 或 $J \to I$ 时趋于无穷，这一性质对于采用障碍几何的优化至关重要。该性质源于对补运算 $J \to I - J$ 的不变性。因此，新颖的 VPM 障碍双对数行列式函数及其底层的对偶信息几何，或可为优化理论 [31]提供助益。

对称正定锥的方差-精度模型/流形网页为 https://franknielsen. github.io/VPM/index.html

A 符号说明

表 2：符号说明

符号	定义域
$U \in O(n)$	正交矩阵
$(\cdot)^T$	矩阵转置
$X,\ Y,\ Z \in \mathrm{VPM}^{\circ}(n)$	方差-精度流形的元素
$A, B, V, W \in \mathrm{Sym}(n)$	对称矩阵（切空间的元素）
$P,\ Q \in$ 正定矩阵(n)	正定矩阵
$X = (X,\ I - X)$	配对嵌入
$\lambda_{\min}(V)\ \lambda_{\max}(V)\ \lambda_i(V)$	对称矩阵的最小/最大/i 大特征值 V
$\mathcal{V}, \mathcal{W}$	有限维向量空间
$u, v, x, y, z \in \mathbb{R}^n$	实向量
$s, t, \lambda, \mu, \kappa, \theta \in \mathbb{R}$	实数
$i, j, k, l \in \mathbb{N}$	自然数
$C \subseteq \mathbb{R}^n$	开有界凸子集 $\mathbb{R}^n$
$\mathrm{d}_{\mathrm{H}},\ \|\|\cdot\|\|^H$	希尔伯特距离，对应的芬斯勒范数
$\mathrm{d}_{\mathrm{AIRM}}, \|\cdot\|^{\mathrm{AIRM}}, \mathrm{g}^{\mathrm{AIRM}}$	AIRM 距离、对应的黎曼范数和黎曼度量
$\mathrm{d}^{\|\|}_{\mathrm{AIRM}}$	AIRM限制距离
$\mathrm{d}^{\rightarrow}_{\mathrm{AIRM}}$	AIRM拉回距离
$\mathcal{M}, \mathcal{N}$	光滑流形
$T\mathcal{M}, T_p\mathcal{M}$	切丛，$p \in \mathcal{M}$处的切空间
K	向量空间中的开凸尖锥
$\gamma(t)\colon\ [0,1] \to \mathcal{M}$	$\mathcal{M}$中的分段光滑曲线
$L_g(\gamma)$	曲线 γ 在黎曼度量 g下的长度泛函
$D =$ 对角矩阵$(d_1,\ \ldots d_n)$	元素为 $d_1,\ \ldots d_n$的对角矩阵

B 希尔伯特 VPM 距离的符号计算

我们使用 Maxima[1] 符号计算包 [?] 来计算希尔伯特 VPM 距离并检验各种性质。下面是一段代码片段

/* 计算 2x2 对称正定矩阵的希尔伯特 VPM 距离 */ kill(all); I : matrix([1, 0], [0, 1]); /* 检查矩阵是否为对称正定矩阵 */ isSPD(M) := block([vals],vals : eigenvalues(M)[1],if every(lambda([x],is(x > 0)), vals) then return(true) else return(false))$

/* 生成随机 2x2 对称正定矩阵 */ randomSPD() :=block([B, A],B : matrix([random(1.0)+1, random(1.0)+1],[random(1.0)+1, random(1.0)+1]), A : transpose(B) . B, return(A))$

希尔伯特 VPM 距离(J1,J2):= block([T1, T2, lambda1 , lambda2], T1: invert(J2).J1 , T2: invert(I-J2).(I-J1) , lambda1: sort(eigenvalues(T1)[1]), lambda2: sort(eigenvalues(T2)[1]), return(log(max(lambda1[2],lambda2[2])/min(lambda1[1],lambda2[1]))))$

J1: 矩阵([7/20, -3*sqrt(3)/20],[- 3*sqrt(3)/20, 13/20])$

J2: 矩阵([11/20, -sqrt(3)/20],[- sqrt(3)/20, 9/20])$

[1] https://maxima.sourceforge.io/

希尔伯特 VPM 距离(J1,J2); 浮点数(%);

iota(X):=X.invert(I+X);

V1：randomSPD();J1：iota(V1); isSPD(J1); V2：
randomSPD();J2：iota(V2); isSPD(J2); dH12：
float(希尔伯特 VPM 距离(J1,J2));

alpha:random(1.0); J12:alpha*J1+(1-alpha)*J2; /* 希尔伯特预测
地线 */ dH1mid12: float(HilbertVPMDistance(J1,J12));
dHmid122: float(HilbertVPMDistance(J2,J12)); /* 检查希尔伯特
VPM 线段是否为预测地线 */ dH12-(dH1mid12+
dHmid122);float(%); /* 零 */

参考文献

[1] Shun-ichi Amari. 正测度与正定矩阵的信息几何：可分解对偶平坦结构。*Entropy*, 16(4):2131–2145, 2014.[2] Shun-ichi Amari. 信息几何及其应用。应用数学科学。Springer Japan, 2016.[3] Marc Arnaudon and Frank Nielsen. 芬斯勒几何中的中位数与均值。*LMS* 计算与数学杂志, 15:23–37, 2012.[4] Rajendra Bhatia. 正定矩阵。载于正定矩阵。普林斯顿大学出版社，2009.[5] Stephen Boyd and Lieven Vandenberghe. 凸优化。剑桥大学出版社，2004.[6] Lev M. Bregman. 求取凸集公共点的松弛法及其在求解凸规划问题中的应用。苏联计算数学与数学物理, 7(3):200–217, 1967.[7] Yongxin Chen, Tryphon T Georgiou, and Michele Pavon. 随机控制的联系：Richard Sinkhorn 与 Gaspard Monge 在 Schr¨odinger 桥上相遇。*SIAM* 评论, 63(2):249–313, 2021.[8] Anoop Cherian and Suvrit Sra. 正定矩阵的黎曼字典学习与稀疏编码。*IEEE* 神经网络与学习系统汇刊, 28(12):2859– 2871, 2016.[9] Lenaic Chizat, Gabriel Peyr´e, Bernhard Schmitzer, and Fran¸cois-Xavier Vialard. 非平衡最优传输问题的缩放算法。计算数学, 87(314):2563– 2609, 2018.

[10] Andrzej Cichocki, Sergio Cruces, and Shun-ichi Amari. 对数行列式散度再探：阿尔法-贝塔与伽马对数行列式散度。*Entropy*, 17(5):2988–3034, 2015.[11] Koby Crammer, Mehryar Mohri, and Fernando Pereira. 高斯间隔机。载于人工智能与统计, 第105–112页。PMLR, 2009.[12] Jason Davis and Inderjit Dhillon. 多元高斯分布的微分熵聚类。神经信息处理系统进展, 19, 2006.[13] Pierre de la Harpe. 关于单纯形上的希尔伯特度量。载于Graham A. Niblo and Martin A. Roller（编），几何群论，第1卷，伦敦数学学会讲义系列，第97–119页。剑桥大学出版社，剑桥，1993.[14]Gy¨orgy P´al Geh´er and Peter ˇSemrl. 希尔伯特空间效应代数中的共存性及其对称变换的刻画。《数学物理通讯》, 379(3):1077–1112, 2020.[15] Tryphon T Georgiou and Michele Pavon. 经典与量子 Schr¨odinger 系统的正压缩映射。《数学物理学杂志》, 56(3), 2015.[16] Mehrtash T Harandi, Mathieu Salzmann, and Richard Hartley. 从流形到流形：面向正定矩阵的几何感知降维。载于欧洲计算机视觉会议，第17–32页。施普林格，2014.[17] Tim Holliday, Peter Glynn, and Andrea Goldsmith. 香农遇上李雅普诺夫：信息论与动力系统之间的联系。载于第*44*届 *IEEE* 决策与控制会议论文集，第1756–1763页。电气电子工程师学会，2005.[18] Alan Treleven James. 方差信息流形及其上的函数。载于多元分析——*III*，第157–169页。爱思唯尔，1973.[19] Yassir Jedra and Alexandre Proutiere. 线性二次控制中的最小期望遗憾。载于国际人工智能与统计会议，第10234–10321页。PMLR, 2022.[20] Jacek Karwowski and Frank Nielsen. 对称正定双锥的希尔伯特几何。载于*NeurIPS 2025* 神经表征中的对称性与几何研讨会，2025.[21] Peter Lancaster and Leiba Rodman. 代数黎卡提方程。克拉伦登出版社，1995.[22] Bas Lemmens and Roger D Nussbaum. 伯克霍夫版本的希尔伯特度量及其在分析学中的应用。载于《希尔伯特几何手册》，第275–303页。欧洲数学学会出版，2014.[23] Oscar Montiel Ross. 基于效应的粒、代数性质与参考架构：量子粒计算基础。*arXiv* 电子预印本，第arXiv–2511页，2025.[24] Cyrus Mostajeran, Natha¨el Da Costa, Graham Van Goffrier, and Rodolphe Sepulchre. 半正定锥中极端特征值的微分几何。《*SIAM* 矩阵分析与应用杂志》, 45(2):1089–1113, 2024.

[25] Frank Nielsen。多元高斯流形上的 Fisher-Rao 与回拉希尔伯特锥距离及其在混合分布简化与混合分布量化中的应用。载于拓扑、代数与几何学习研讨会 *2023*，第 488–504 页。PMLR，2023。[26] Frank Nielsen 与 Ke Sun。希尔伯特射影几何中的聚类：概率单纯形与相关矩阵椭圆多胞形的案例研究。载于 信息的几何结构，第 297–331 页。施普林格，2018。[27] Frank Nielsen 与 Ke Sun。希尔伯特单纯形几何中的非线性嵌入。载于 拓扑、代数与几何学习研讨会 *2023*，第 254–266 页。PMLR，2023。[28] Michael A Nielsen 与 Isaac L Chuang。量子计算与量子信息。剑桥大学出版社，2010。[29] Richard Nock 与 Frank Nielsen。拟合最小包围 Bregman球。载于欧洲机器学习会议，第 649–656 页。施普林格，2005。[30] Atsumi Ohara 与 Shinto Eguchi。由 V- 势函数变形的正定矩阵几何及其子流形结构。载于 信息几何理论，第 31–55 页。施普林格，2014。[31] Atsumi Ohara、Hideyuki Ishi 与 Takashi Tsuchiya。双重自平行结构与曲率积分。信息几何，7（增刊 1）：555–586，2024。[32] Atsumi Ohara、Nobuhide Suda 与 Shun-ichi Amari。正定矩阵的对偶微分几何及其在相关问题中的应用。线性代数及其应用，247：31–53，1996。[33] Xavier Pennec、Stefan Sommer 与 Tom Fletcher。医学图像分析中的黎曼几何统计。学术出版社，2019。[34] Roman A Polyak。现代优化中的勒让德变换。优化及其在控制与数据科学中的应用，第 437–507 页，2016。[35] David Reeb、Michael J Kastoryano 与 Michael M Wolf。量子信息论中的希尔伯特射影度量。《数学物理学杂志》，52（8），2011。[36] Zhongmin Shen。芬斯勒几何讲义。世界科学出版社，2001。[37] Hirohiko Shima。黑塞结构的几何。世界科学出版社，2007。[38] Carl Ludwig Siegel。辛几何。美国数学杂志，65（1）：1–86，1943。[39] Lene Theil Skovgaard。多元正态模型的黎曼几何。斯堪的纳维亚统计学杂志，第 211–223 页，1984。[40] Suvrit Sra。正定矩阵与 S-散度。美国数学会会刊，144（7）：2787–2797，2016。

[41] Dario Stein 与 Richard Samuelson。统一高斯概率与非确定性的范畴。载于第十届计算机科学代数与余代数会议（*CALCO 2023*），第 13–1 页。Schloss Dagstuhl–Leibniz-Zentrum f¨ur Informatik，2023。[42] Yann Thanwerdas。协方差矩阵与相关矩阵上的黎曼几何与分层几何。博士论文，Universit´e Cˆote d'Azur，2022。[43] Jan C Willems。开放随机系统。*IEEE* 自动控制汇刊，58（2）：406–421，2012。[44] Or Yair、Mirela Ben-Chen 与 Ronen Talmon。用于领域自适应的对称正定矩阵锥流形上的平行移动。*IEEE* 信号处理汇刊，67（7）：1797–1811，2019。[45]Shintaro Yoshizawa 与 Kunio Tanabe。与高斯分布上的 Kullback-Leibler 信息相关联的对偶微分几何及其二参数变形。*SUT* 数学杂志，35（1）：113–137，1999。