arXiv:2601.01097v1 [stat.ML] 3 Jan 2026

非紧型对称空间上的神经网络

Xuan Son Nguyen, Shuo Yang, AymericHistaceETIS，UMR 8051，CY 塞尔吉巴黎大学，ENSEA，法国国家科学研究中心，法国{
xuan-son.nguyen,shuo.yang,aymeric.histace}@ensea.fr

摘要

近期研究已展示了神经网络在双曲空间和对称正定（SPD）流形上的良好性能。这些空间属于一类黎曼流形，称为非紧型对称空间。本文提出一种在此类空间上构建神经网络的新方法。我们的方法基于对所研究空间上点到超平面距离的统一表述。我们证明，在特定设置下，我们的方法可以还原点到超平面距离的一些现有表述。此外，我们推导出配备 G-不变黎曼度量的非紧型高秩对称空间中点到超平面距离的闭式表达。随后，我们利用推导出的距离，设计适用于所研究空间上神经网络的全连接（FC）层和注意力机制。我们在具有挑战性的图像分类、脑电图（EEG）信号分类、图像生成和自然语言推理基准上验证了该方法。

1 引言

非欧空间中的神经网络已成为强大的工具，可用于解决自然语言处理（Chami 等人，2019；Ganea 等人，2018b）、计算机视觉（Huang & Gool，2017；Nguyen 等人，2019b）和医学（Liu 等人，2019）等广泛领域中的问题。由于双曲空间能够以较低维度高保真地表示层次数据，现有文献中有大量研究聚焦于双曲神经网络（HNN）（Chami 等人，2021）。对称正定流形是另一类常见的非欧空间。本文将研究范围限定于具有流形值输出的神经网络。

超平面这一概念已被证明有助于构建双曲神经网络（Ganea 等人，2018b；Shimizu 等人，2021）以及用于双曲空间的分类算法（Fan 等，2023）。双曲空间中存在两类超平面，即庞加莱超平面（Ganea 等人，2018b；Shimizu 等人，2021），它们被定义为测地线的集合；以及等距曲线（Fan 等人，2023；Helgason，1984），它们被描述为与一族平行测地线正交的流形。近年来，一些方法（Chen 等人，2024a；Nguyen & Yang，2023；Nguyen 等人，2024）成功构造了庞加莱超平面的矩阵流形类比。然而，这些方法要么仅适用于与特定黎曼度量族相关的对称正定流形（Chen 等人，2024a），要么要求所研究空间具有丰富的代数结构，从而限制了其通用性。

本文提出了一种在非紧型对称空间上构建神经网络的新方法（Helgason，1979）。这些空间包括双曲空间和对称正定流形，通常被认为是数学中最基本且最优美的对象之一（Bridson & H¨afliger，2011；Helgason，1994）。我们的贡献总结如下：

- 我们提出一种在非紧型对称空间中构造点到超平面距离的新方法。与 Ganea 等人（2018b）、Nguyen & Yang（2023）以及 Chen 等人（2024a）不同，后者仅关注双曲空间（Ganea 等人，2018b）或对称正定流形（Nguyen & Yang，2023；Chen 等人，2024a）；我们的方法适用于所有这些空间，并为该距离提供了统一的表达式。

• 我们推导出在配备 G-不变黎曼度量的非紧型对称空间中点到超平面距离的表达式。

• 我们提出了适用于非紧型对称空间上神经网络的全连接层和注意力机制。据我们所知，在这一研究领域中，我们首次开发了此类构建模块。

• 我们在图像分类、脑电信号分类、图像生成和自然语言推理方面提供了实验结果，展示了我们方法的有效性。

2 相关工作

2.1 双曲空间

自 Ganea 等人（2018b）的开创性工作以来，双曲神经网络受到了越来越多的关注。该工作受 Lebanon 与 Lafferty（2004）的启发，提出了深度神经网络中若干构建模块的双曲类比。随后，Shimizu 等人（2021）引入了 Ganea 等人（2018b）中缺失的一些构建模块（例如全连接层和卷积层）。Ganea 等人（2018b）和 Shimizu 等人（2021）的工作主要依赖于 Poincar´e 超平面的构造。Fan 等人（2023）研究了双曲空间上的另一种超平面概念（等距曲线）。该方法利用 horosphere 投影推导点与等距曲线之间的距离，而 horosphere 投影最初用于双曲空间中的降维（Chami 等人，2021）。受图神经网络（GNNs）出色性能的启发（Veli˘ckovi´c 等人，2018），研究人员也对双曲空间中的图神经网络进行了研究（Chami 等人，2019；Gulcehre 等人，2018）。

2.2 矩阵流形

现有的大多数研究关注定义在对称正定流形、格拉斯曼流形和特殊正交群上的神经网络。SPDNet、LieNet 和 GrNet 是最早针对这些空间设计的网络之一（Huang & Gool，2017；Huang 等人，2017；2018）。在 Brooks 等人（2019）；Ju 与 Guan（2023）；Kobler 等人（2022）；Nguyen 等人（2019a；b；2020）；Nguyen（2021）；Pan 等人（2022）；Wang 等人（2021）的研究中，作者要么引入了黎曼批归一化层，要么改进了 Bimap 层（Huang & Gool，2017）。L´opez 等人（2021）；Nguyen（2022a；b）；Nguyen & Yang（2023）；Nguyen 等（2024）的研究利用对称正定流形和格拉斯曼流形丰富的代数结构，将欧几里得空间中的一些基本运算和概念推广到这些流形上。受 Nguyen & Yang（2023）的启发，Chen 等人（2024a）的研究在两类黎曼度量下，将多项逻辑回归（MLR）推广到了对称正定流形。

2.3 一般黎曼流形

此外，也有研究尝试为黎曼流形构建更通用的框架。Chakraborty 等人（2020）；Zhen 等人（2019）的研究主张使用加权 Fr´echet 均值，为黎曼流形上的神经网络构建若干层（例如卷积层和残差层）。在 Katsman 等人（2023）的研究中，作者对向量场进行参数化，以设计黎曼残差神经网络。我们的工作与此相关，因为可以使用我们推导出的距离来参数化这类向量场。此外，研究者还提出了李群上 SPD 批归一化层（Brooks 等人，2019）的扩展（Chen 等人，2024b）。

3 数学背景

3.1 双曲空间与对称正定流形

我们简要讨论机器学习应用中常见的两类对称空间的几何性质。

双曲空间 维度为 m 的双曲几何的庞加莱模型 $\mathbb{B}_m$ 定义为流形

$$\mathbb{B}_m = \{x \in \mathbb{R}^m \quad \|x\| < 1\} \qquad \langle u\ v\rangle_x =: \text{其配备的黎曼度量为} \ {}_{,(1-\|x\|^2)^2}\langle u, v\rangle \text{其中 } u, v \in \mathbb{R}^m\text{。两点 } x, y \in \mathbb{B}_m \text{ 之间的黎曼距离由下式给出}$$

[4]

by $d_{\mathbb{B}}(x,y) = \cosh^{-1}\left(1+2\frac{\|x-y\|^2}{(1-\|x\|^2)(1-\|y\|^2)}\right)$。附录 L.1 从对称空间的角度详细讨论了双曲空间。

对称正定流形 这里我们考虑 PEM（Chen 等人，2024c）（见附录 L.2），它比成熟的对数欧几里得框架（Arsigny 等人，2005）更为一般。令 Sym_m为 $m \times m$ 对称矩阵的空间。在 PEM 下，对称正定流形 Sym_m^+ 定义为 $\mathrm{Sym}_m^+ = \{x \in \mathrm{Sym}_m : u^T x u > 0$对所有 $u \in \mathbb{R}^m, u \neq \mathbf{0}\}$都成立，并配备度量 $\langle u,v\rangle_x^{\phi} = \langle D_x\phi(u)\ D_x\phi(v)\rangle$，其中 $\phi: \mathrm{Sym}_m^+ \to \mathrm{Sym}_m$是一个微分同胚，$D_x\phi: T_x\mathrm{Sym}_m^+ \to T_{\phi(x)}\mathrm{Sym}_m$是映射 ϕ 在点 x处的方向导数，T_xX是 X 在 $x \in X$处的切空间。对称正定流形 Sym_m^+ 中两点 $x,y \in$ 之间的黎曼距离为$d_{\mathbb{PEM}}(x,y) = \|\phi(x) - \phi(y)\|$。附录 L.3 从对称空间的角度详细讨论了对称正定流形。

黎曼流形上现有的点到超平面距离通常是针对上述某一类对称空间构建的，复合距离除外（Helgason，1984；1994）。然而，将复合距离用于我们的目的并非易事，因为它在高秩对称空间（例如对称正定流形）中是一种向量值距离。下文将建立一个统一框架，以解决现有研究的这一局限。

3.2 非紧型对称空间

本节简要回顾本文中用到的重要概念。更多内容请参阅 Ballmann（2012）；Bridson & H¨afliger（2011）；Helgason（1979）。

粗略地说，对称空间 X 是一种连通的黎曼流形，且在任意点处都具有反射对称性。也就是说，对于任意 $x \in X$，都存在 X 的一个局部等距变换 s_x，使得 $s_x(x) = x$，且其微分 $D_x s_x = -\mathrm{id}_{T_xX}$。每个（单连通）对称空间都是不可约对称空间的黎曼乘积。如果一个对称空间不能进一步分解为对称空间的黎曼乘积，则称其为不可约对称空间。非平坦的不可约对称空间有两种类型：紧型和非紧型。这两种类型可通过嘉当对偶互换。更多讨论请参阅附录 L.4。下文将只讨论非紧型对称空间。

形式上，设 G是一个中心有限的连通非紧半单李群，K 是 G的极大紧子群。那么，非紧型对称空间 X 由左陪集组成：

$$X := G/K := \{x = gK | g \in G\}.$$

G 在 $X = G/K$ 上的作用定义为 $g[x] = g[hK] = ghK$，其中 $x = hK \in X$、g，$h \in G$。令 o 为 X 中的原点 K，则映射 φ：$gK \mapsto g[o]$ 是从 G/K 到 X的微分同胚。

设 $G = KAN$ 为 G的伊瓦萨瓦分解（Helgason，1979；Sawyer，2016），并设 $\mathfrak{g}$ 和$\mathfrak{a}$ 分别为 G 和 A的李代数。对于 $\mathfrak{a}$上的任意线性形式 α，令 $\mathfrak{g}_\alpha := \{v \in \mathfrak{g} | \forall u \in \mathfrak{a},\ [u,v] = \alpha(u)v\}$。设 $\mathfrak{a}^*$为 $\mathfrak{a}$的对偶空间。则限制根集合定义为$\Sigma := \{\alpha \in \mathfrak{a}^* \setminus \{0\} | \mathfrak{g}_\alpha \neq \{0\}\}$。每个限制根的卷积核都是 $\mathfrak{a}$的一个超平面。$\mathfrak{a}$ 中的 Weyl 室是 $\mathfrak{a} \setminus \cup_{\alpha\in\Sigma} \ker(\alpha)$的一个连通分支。我们固定一个 Weyl 室 $\mathfrak{a}^+$，并用$\mathfrak{a}^+$ 表示其闭包。

几何边界 在非紧型对称空间中，边界（理想）点可以被视为欧几里得空间中方向概念的推广。直观地说，边界点表示点沿其可以趋向无穷远的方向（Chami 等人，2021年）。X 的边界点集合被称为 X的（几何）边界。例如，庞加莱 ́圆盘模型（二维双曲几何的一个模型）表示为$\mathbb{D} = \{(x_1,\ x_2):\ x_1^2 + x_2^2 < 1\}$（可以将该集合视为所有长度小于 1 的复数的集合，即 $\mathbb{D} = \{x \in \mathbb{C} :\ \|x\| < 1\}$）。$\mathbb{D}$ 的边界 $\partial\mathbb{D}$ 是单位圆$\partial\mathbb{D} = \{(x_1,\ x_2):\ x_1^2 + x_2^2 = 1\}$。

令 $d(.,.)$ 表示由黎曼度量诱导的距离。X 中的测地射线是一个映射 $\delta: [0,\infty) \to X$，满足 $d(\delta(t)\ \delta(t')) = |t-t'|, \forall t,t' \geq 0$。$X$ 中的测地线是一个映射 $\delta: \mathbb{R} \to X$，满足 $d(\delta(t)\ \delta(t')) = |t-t'|, \forall t,t' \in \mathbb{R}$。若两条测地射线 δ, δ' 满足以下条件，则称它们渐近

$d(\delta(t)\ \delta'(t))$ 在 t 中一致有界。这是 X 中所有测地射线集合上的一个等价关系。 X 的边界点集合 ∂X 是测地射线的等价类集合。测地射线 δ 的等价类记为 $\delta(\infty)$。

角度度量 对于 $x \in X$ 和 $\xi,\ \xi' \in \partial X$，分别存在唯一的测地射线 δ 和 δ'，它们从 x 出发，并分别属于类别 ξ 和 ξ'（Ballmann，2012年）。于是可以将 $\angle_x(\xi,\ \xi')$ 定义为 x 处 δ 与 δ' 之间的角（见附录 L.5）。角 $\angle(\xi,\ \xi')$ 定义为

$$\angle(\xi, \xi') = \sup_{x \in X} \angle_x(\xi, \xi').$$

函数 $(\xi,\ \xi') \mapsto \angle(\xi,\ \xi')$ 在 ∂X上定义了角度度量（Bridson & H¨afliger，2011年）.

布塞曼坐标 布塞曼函数（坐标）可以看作欧几里得空间中坐标概念的推广。在欧几里得空间中，给定一个点 x和一个表示方向的单位向量w，有

$$-\langle x, w\rangle = \lim_{t\to\infty}(d(x, tw) - d(0, tw)) = \lim_{t\to\infty}(d(x, tw) - t),$$

其中，$tw,\ t > 0$ 可以看作一条沿着 w 的方向向无穷远移动的射线，当 $t \to \infty$时。注意，内积 $\langle x,\ w\rangle$ 给出了 x 沿 w方向的坐标。这一观察可用于计算 X中的坐标。令 $\delta\colon\ [0, \infty) \to X$ 为一条（单位速率）测地射线，且 $\xi = \delta(\infty) \in \partial X$。于是，用测地射线 $\delta(t)$ 替换 tw，即可将点 $x \in X$ 沿 ξ 方向的布塞曼坐标定义为

$$B_\xi(x) = \lim_{t\to\infty}(d(x, \delta(t)) - t).$$

函数 $B_\xi\colon\ X \to \mathbb{R}$ 称为与测地射线 δ相关联的布塞曼函数。

等距曲线 等距曲线如同与一族平行直线正交的欧几里得超平面，也与一族平行测地线正交（Helgason，1984；1994）。因此，等距曲线可视为欧几里得超平面在对称空间中的类比。设 M 为 K中 A的中心化子，即，对所有 $a \in A\}$，$M := C_K(A) :$ $= \{k \in K | ka = ak$。等距曲线空间 Ξ 可识别为（Helgason，1994）

$$\Xi := G/MN := \{\eta = gMN | g \in G\}.$$

复合距离 复合距离的概念是欧几里得内积在对称空间中的类比（Helgason，1984；1994）。设 $\eta = gMN$ 为一条满足 $g \in G$的等距曲线，且 $g = kan$ 满足 $k \in K$、$a \in A$和 $n \in N$。则称 $\xi = kM \in \partial X$ 垂直于 η，而 log(a)是从原点 o到 η的复合距离。更一般地，若 $x = gK \in X$，且$\eta = hMN \in \Xi$满足 $g,\ h \in G$，则 $H(g^{-1}h)$是从 x到 η的复合距离，其中映射 $H\colon\ G \to \mathfrak{a}$由 $g_1 = k_1 \exp\ H(g_1)n_1$确定，且 $g_1 \in G$、$k_1 \in K$和 $n_1 \in N$。

4 方法

我们分别在第4.1节和第4.2节中定义超平面，并针对所考虑的空间提出点到超平面距离的一般形式。随后，我们在第4.3节中研究所提出的形式在双曲空间和对称正定流形中的应用。在第4.4节中，我们推导出配备 G-不变黎曼度量的空间中的距离。在第5节中，我们展示如何在所考虑的空间上为神经网络构建全连接层和注意力机制。

4.1 对称空间上的超平面

在欧几里得空间 $\mathbb{R}^m$ 中，超平面 $\mathcal{H}^E_{a,\ b}$ 定义为

$$\mathcal{H}^E_{a,b} = \{x \in \mathbb{R}^m : \langle x, a\rangle - b = 0\},$$

其中 $a \in \mathbb{R}^m \setminus \{\mathbf{0}\}$、$b \in \mathbb{R}$ 和 $\langle .,\ .\rangle$ 表示欧几里得内积。超平面 $\mathcal{H}^E_{a,\ b}$ 可改写为

$$\mathcal{H}^E_{a,b} = \{x \in \mathbb{R}^m : \langle p - x, a\rangle = 0\},$$

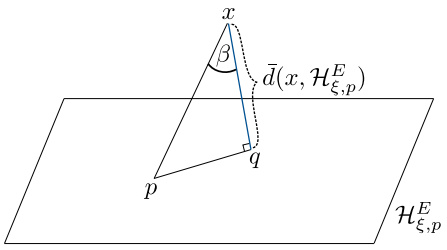


图1：点 x 与超平面 $\mathcal{H}^E_{\xi,\,p}$之间的距离。

其中 $p \in \mathbb{R}^m$ 和 $\langle p,\ a\rangle = b$。

为了将欧几里得超平面推广到矩阵流形，Nguyen & Yang（2023）的研究将参数 a视为所考虑流形 X上的一个点。随后，通过定义运算$-$ 和 $+$ 的矩阵流形类比，以及欧几里得内积的矩阵流形类比，将超平面$\mathcal{H}^E_{a,\,b}$的方程推广到矩阵流形情形。这里我们采用不同的方法，将 $\langle p - x,\ a\rangle$ 重写为布塞曼函数。设 ξ 为测地射线$\delta(t) = t\frac{a}{\|a\|}$的等价类，其中 $\|\cdot\|$ 为欧几里得范数。根据$\mathbb{R}^m$ 中的布塞曼函数表达式（见附录 L.6），我们有

$$\langle p - x, \frac{a}{\|a\|}\rangle = B_\xi(-p + x).$$

假设可以在 X 上定义适当的运算 $\ominus$ 和 $\oplus$，它们分别是运算 $-$ 和 $+$的对称空间类比。这引出了以下定义。

定义 4.1 (对称空间上的超平面）。对于 $p \in X$ 和 $\xi \in \partial X$，X上的超平面定义如下：

$$\mathcal{H}_{\xi,p} = \{x \in X : B_\xi(\ominus p \oplus x) = 0\},$$

其中，$\ominus$ 和 $\oplus$ 分别是 X上的逆运算和二元运算。

在对称空间中，horocycle 是一种与一族平行测地线正交的流形（Helgason，1994）。因此，horocycle 推广了欧几里得超平面的概念；欧几里得超平面与一族平行直线正交。在我们的方法中，超平面包含一个固定点$p \in X$ 以及满足以下条件的每个点 $x \in X$：线段 $\ominus p \oplus x$ 与一个固定方向ξ正交。形如 $\ominus p \oplus x$ 的线段可视为欧几里得直线在对称空间中的对应物。因此，这些超平面也以自然的方式推广了欧几里得超平面的概念。

4.2 对称空间上的点到超平面的距离

设 $\mathcal{H}^E_{\xi,\,p}$ 是 $\mathbb{R}^m$ 中的一个超平面。那么，点 $x \in \mathbb{R}^m$ 与 $\mathcal{H}^E_{\xi,\,p}$ 之间的距离 $\bar{d}(x,\ \mathcal{H}^E_{\xi,\,p})$ 可按如下方式计算（见图 1）：

$$\bar{d}(x, \mathcal{H}^E_{\xi,p}) = d(x,p)\cos(\beta), \tag{1}$$

其中，β 是线段 $[x,p]$ 和 $[x,q]$ 之间的夹角，q 是 x 在 $\mathcal{H}^E_{\xi,\,p}$上的投影。按照惯例，对于任意 $x \in \mathcal{H}^E_{\xi,\,p}$，$\bar{d}(x,\ \mathcal{H}^E_{\xi,\,p}) = 0$。注意，式（1）可改写为

$$\bar{d}(x, \mathcal{H}^E_{\xi,p}) = d(x,p)\cos\angle_x(\xi', \xi),$$

其中，ξ 和 ξ' 分别是测地射线 δ 和 δ' 的等价类；这两条测地射线从 x 出发，其像分别为线段 $[x,q]$ 和 $[x,p]$。设 $x = \delta(t)$，则

$$\bar{d}(x, \mathcal{H}^E_{\xi,p}) = d(x,p)\cos\angle_{\delta(t)}(\xi', \xi) = -d(x,p)\lim_{t\to+\infty}\frac{B_\xi(\delta'(t))}{t}.$$

[-]最后一个表达式（Kapovich 等人，2017年）非常引人注目，因为它将距离 $d(.,.)$ 与布塞曼函数联系起来。还要注意，

$$B_\xi(\delta'(t)) = -\langle\delta'(t), a\rangle = -\langle ta', a\rangle,$$

其中 $\delta(t) = ta, \delta'(t) = ta'$， a 是一个单位向量，且 $a' = \frac{p-x}{\|p-x\|}$。因此

$$\bar{d}(x, \mathcal{H}^E_{\xi,p}) = d(x,p)\langle a', a\rangle = d(x,p)\langle \frac{p-x}{\|p-x\|}, a\rangle = d(x,p)\frac{B_\xi(-p+x)}{\|-p+x\|}.$$

这促使我们给出以下定义。

定义 4.2。设 $\mathcal{H}_{\xi,p}$ 是定义 *4.1* 中给出的超平面，且 $\|\cdot\|_{\mathbb{S}}$ 是 X 上的范数。那么，点 $x \in X$ 与 $\mathcal{H}_{\xi,p}$ 之间的（有符号）距离 $d(x,\mathcal{H}_{\xi,p})$ 定义为

$$\bar{d}(x, \mathcal{H}_{\xi,p}) = d(x,p)\frac{B_\xi(\ominus p \oplus x)}{\|\ominus p \oplus x\|_{\mathbb{S}}}.$$

4.3 双曲空间和对称正定流形上的点到超平面距离 S

We 现在推导第 3.1 节所讨论的对称空间中的点到超平面距离 .

双曲空间 以下结果显然成立。

推论 4.3。设 $\ominus$ 和 $\oplus$ 分别是 $\mathbb{B}_m$ 中的 *M¨obius* 减法 $\ominus_M$ 和 *M¨obius* 加法 $\oplus_M$，且 $\|\cdot\|_{\mathbb{S}}$ 为欧几里得范数 $\|\cdot\|$（参见附录 *L.7.1*）。设 $p \in \mathbb{B}_m$、$\xi \in \partial\mathbb{B}_m$、 ，且 $\mathcal{H}_{\xi,p}$ 是定义 *4.1* 中给出的超平面。那么，点 $x \in \mathbb{B}_m$ 与 $\mathcal{H}_{\xi,p}$ 之间的距离 $d(x,\mathcal{H}_{\xi,p})$ 计算如下

$$\bar{d}(x, \mathcal{H}_{\xi,p}) = -\frac{d_{\mathbb{B}}(x,p)}{\|-p \oplus_M x\|}\log\frac{1-\|-p\oplus_M x\|^2}{\|-p\oplus_M x - \xi\|^2}.$$

对称正定流形命题 4.4 表明，Chen 等人（2024a）研究的点到超平面距离是我们提出的距离的一种特殊情况（命题 4.4 的证明见附录 M.1）。

命题 4.4。设 ϕ : 对称$^+_m \to$ 对称$_m$ 为一个微分同胚。设 $\oplus$ 和 $\ominus$为由下式定义的二元运算和逆运算

$$x \oplus y = \phi^{-1}(\phi(x) + \phi(y)),$$
$$\ominus x = \phi^{-1}(-\phi(x)),$$

其中 $x, y \in$ 对称^+_m。设 $\|\cdot\|_{\mathbb{S}}$ 为由内积 $\langle\cdot\rangle_{\mathbb{S}}$ 诱导的范数，定义如下：

$$\langle x, y\rangle_{\mathbb{S}} = \langle\phi(x), \phi(y)\rangle.$$

设 $\delta(t) = \phi^{-1}(ta)$是对称$^+_m$中的一条测地线，其中 $a \in$ 对称$_m$ 且 $\|a\| = 1$。设 $\xi = \delta(\infty), p \in$ 对称^+_m，并设 $\mathcal{H}_{\xi,p}$ 为定义 *4.1* 中给出的超平面。则点 $x \in$ 对称^+_m 与$\mathcal{H}_{\xi,p}$ 之间的距离 $\bar{d}(x,\mathcal{H}_{\xi,p})$计算如下：

$$\bar{d}(x, \mathcal{H}_{\xi,p}) = \langle a, \phi(p) - \phi(x)\rangle.$$

命题 4.4 的一个直接推论是，SPD 矩阵与 SPD 双曲超平面之间的距离（Nguyen & Yang, 2023）也是我们所提出的距离在对数欧几里得框架和对数 Cholesky 框架下的一个特例（例如，在对数欧几里得框架中，映射 ϕ 是矩阵对数）。

4.4 与 A G 不变度量相关的点到超平面距离

在前一节中，我们在 PEM 下计算了双曲空间和 SPD 流形中的点到超平面距离的闭式表达式。本节将推导配备 A G 不变黎曼度量的非紧型高秩对称空间 X 中的这一距离。这需要我们：(1) 在 X 上定义二元运算 $\oplus$ 和逆运算 $\ominus$；(2) 在 X 上定义范数 $\|\cdot\|_{\mathbb{S}}$；以及 (3) 计算布塞曼函数。

设 $x = gK, y = hK \in X$，其中 $g, h \in G$。

定义 4.5（**二元运算**）。二元运算 $\oplus$定义为

$$x \oplus y = ghK.$$

定义 4.6（逆运算）。逆运算 $\ominus$定义为

$$\ominus x = g^{-1}K.$$

上述定义的动机在于，空间 G/K 在运算 $\oplus$ 下构成一个群（单位元为 K，任意元素的逆元由逆运算给出）。为了计算范数 $\|\cdot\|_{\mathbb{S}}$，我们将定义一个内积 $\langle\cdot,\ \cdot\rangle_{\mathbb{S}}$，其构造基于以下自然的观点：

- 内积 $\langle\cdot,\ \cdot\rangle_{\mathbb{S}}$ 应与黎曼距离一致。
- 内积 $\langle\cdot,\ \cdot\rangle_{\mathbb{S}}$ 应在 K的作用下保持不变。Helgason（1994）以及 Nguyen & Yang（2023）提出的内积具有这一性质。

因此，我们考虑以下内积。

定义 4.7（对称空间上的内积）。设 $x = gK,\ y = hK \in X,\ g,\ h \in G$。则 $\langle\cdot,\ \cdot\rangle_{\mathbb{S}}$ 在 X 上的内积定义为

$$\langle x, y\rangle_{\mathbb{S}} = \langle \mu(g), \mu(h)\rangle,$$

其中，映射（*Cartan* 投影）$\mu:\ G \to \mathfrak{a}^+$ 由 $g \in \overline{=} k\exp(\mu(g))k'$ 确定，其中 $g \in G$和 $k,\ k' \in K$（这由 *Helgason (1979)* 的 *Cartan* 分解得出）G，其中 $\mu(\cdot)$ 是映射到闭 *Weyl* 室的连续、适当且满射的映射 $\mathfrak{a}^+$）。

命题 4.8 表明，上述性质对于所考虑的内积成立（命题 4.8 的证明见附录 M.2）。

命题 4.8。设 $x = gK$、$y = hK \in X$、g、$h \in G$，并设 $\langle\cdot\cdot\rangle_{\mathbb{S}}$ 为定义 *4.7* 中给出的内积。则

（*i*）有：

$$\|\ominus x \oplus y\|_{\mathbb{S}} = d(x, y),$$

其中，范数 $\|\cdot\|_{\mathbb{S}}$ 由内积 $\langle\cdot、\ \cdot\rangle_{\mathbb{S}}$诱导。

(ii) 对任意 $k \in K$，有：

$$\langle x, y\rangle_{\mathbb{S}} = \langle k[x], k[y]\rangle_{\mathbb{S}}.$$

最后，命题 4.9 给出了布塞曼函数的闭式表达式（命题 4.9 的证明见附录 M.3）。

命题 4.9。设 $\delta(t) = k\exp(ta)K$ 为一条测地射线，其中 $k \in K, a \in \mathfrak{a}, \|a\| = 1$，并令$\xi = \delta(\infty)$。则

$$B_\xi(x) = \langle a, H(g^{-1})\rangle,$$

其中 $x \in X$，且 $g \in G$由 $k^{-1}[x] = gK$给出。

作为命题 4.9 的推论，推论 4.10 给出了对称空间中一点到超平面的距离表达式（推论 4.10 的证明见附录 M.4）。

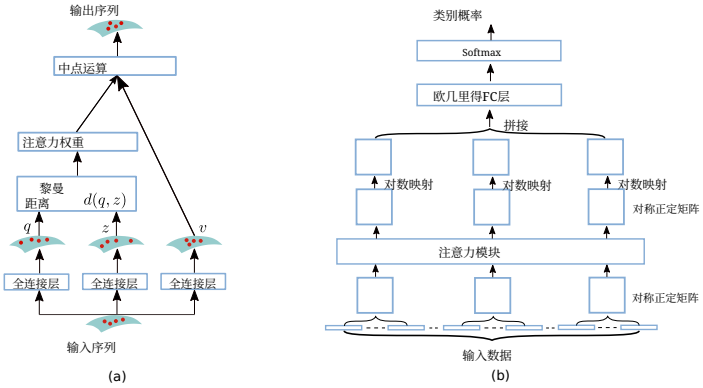
推论 4.10。设 $\delta(t) = k\exp(ta)K$ 为一条测地射线，其中 $k \in K, a \in \mathfrak{a}, \|a\| = 1$，并令$\xi = \delta(\infty)$。设 $p = hK \in X, h \in G$，并令 $\mathcal{H}_{\xi,p}$ 为定义 *4.1* 中给出的超平面。则$\bar{}$点 $x = gK \in X, g \in G$ 到 $\mathcal{H}_{\xi,p}$ 的距离 $d(x,\mathcal{H}_{\xi,p})$ 可计算为

$$\bar{d}(x, \mathcal{H}_{\xi,p}) = \langle a, H(g^{-1}hk)\rangle. \tag{2}$$

公式（2）中的距离与现有研究之间的联系将在附录 G 中讨论。

5 对称空间上的神经网络

在本节中，我们将构建 DNNs 中两个重要构建模块的对称空间类比，即全连接层和注意力机制。我们的出发点是上一节中给出的点到超平面距离的构造。



Fig图 2：我们提出的注意力块 (a) 以及用于 EEG分类的网络架构 (b).

5.1 全连接层

全连接层可以通过以下线性变换来描述：

$$y = ax - b, \tag{3}$$

其中 $a \in \mathbb{R}^{m \times m'}$、 $x \in \mathbb{R}^{m'}$ 和 y， $b \in \mathbb{R}^m$。式 (3) 可以改写为一个方程组，每个方程对应一个维度，即输出 y 的第 j 个维度 y_j， $j = 1 \ldots,$ m 表示为

$$y_j = \langle x, a_j \rangle - b_j,$$

其中 $a_j \in \mathbb{R}^{m'}$， $b_j \in \mathbb{R}$。令 $\tilde{\xi}_j \in \partial X$ ，并令 $\mathcal{H}_{\tilde{\xi}_j,\ K}$ 为包含原点（即 K）且与输出空间的第 j 个坐标轴正交的超平面。于是， y_j 可以解释为从输出 y 到超平面 $\mathcal{H}_{\tilde{\xi}_j,\ K}$ 的¯有符号距离 d（y， $\mathcal{H}_{\tilde{\xi}_j,\ K}$）。因此有

$$\bar{d}(y, \mathcal{H}_{\tilde{\xi}_j, K}) = \langle x, a_j \rangle - b_j.$$

根据定义 4.1，我们可以将表达式 $\langle x,\ a_j \rangle - b_j$ 写作 $B_{\xi_j}(\ominus p_j \oplus x)$，其中 $p_j \in X$ 且$\xi_j \in \partial X$。因此

$$\bar{d}(y, \mathcal{H}_{\tilde{\xi}_j, K}) = B_{\xi_j}(\ominus p_j \oplus x). \tag{4}$$

由于输出空间的坐标轴是正交归一的，因此很容易想到构造一组正交归一的˜边界点 $\{\xi_j\}_{j=1}^m$ ，使得输出 y 与输入 x 通过式 (4) 相关。若两个边界˜ ˜ ˜ ˜点 ξ_l和 ξ_j， l， $j = 1$、.....、 m， $l \neq j$满足 $\angle(\xi_l,\ \xi_j) = \frac{\pi}{2}$，则称它们正交归一。根据命题 5.1 可以确定这样一组边界点（其证明见附录 M.5）。

命题 5.1。 设 $\delta(t) = \exp(ta)K$ 和 $\delta'(t) = \exp(ta')K$ 为测地线，其中 a 和 a' 是 $\mathbb{R}^m$、 $a \neq a'$中的标准基向量。令 $\xi = \delta(\infty)$， $\xi' = \delta'(\infty)$。则 ξ 和 ξ' 正交归一。

现在我们来阐述所提出的全连接层（命题 5.2 的证明见附录 M.6） .

命题 5.2. 令 $\delta_j(t) = k_j \exp(ta_j)K, j = 1, \ldots, m$ 为测地射线，其中 $k_j \in K, a_j \in \mathfrak{a}, \|a_j\| = 1$。令 $v_j(x) = B_{\xi_j}(\ominus p_j \oplus x)$ $j = 1, \ldots, m$，其中 $\xi_j = \delta_j(\infty)$， $p_j \in X$，且 $x \in X$ 是全连接层的输入。那么，全连接层的输出 y 可表示为

$$y = n \exp([-v_1(x) \ldots - v_m(x)])K,$$

其中 $n \in N$。

在我们的方法中，全连接层执行的变换被设计为式（3）中线性变换在对称空间中的对应形式，这使我们的方法有别于现有方法（Huang & Gool，2017；Huang 等人，2018；Chakraborty 等人，2020年；Wang，2021年；Sonoda 等人，2022年）。有关我们的方法与这些方法的比较，请参阅附录 H。

表1： $\mathbb{B}_m$上点到超平面距离的不同表述。

g-距离	h-距离	b-距离
$\sinh^{-1}\left(\frac{2\lvert\langle -p\oplus_M x, a\rangle\rvert}{\lVert -p\oplus_M x\rVert^2}\right)\lVert a\rVert$	$\frac{1}{a}\left\lvert a\text{对数}\frac{1-\lVert x\rVert^2}{\lVert x-\xi\rVert^2}-b\right\rvert$	$\frac{d_{\mathbb{B}}(x,p)}{1-\lVert -p\oplus_M x\rVert^2}\lVert -p\oplus_M x\rVert\text{对数}\frac{1-\lVert -p\oplus_M x\rVert^2}{\lVert -p\oplus_M x-\xi\rVert^2}$
$x,\ p\in\mathbb{B}_m,\ a\in T_p\mathbb{B}_m\setminus\{\mathbf{0}\}$	$x\in\mathbb{B}_m, a,b\in\mathbb{R}, a>0, \xi\in\partial\mathbb{B}_m$	$x,\ p\in\mathbb{B}_m,\ \xi\in\partial\mathbb{B}_m$
（Ganea 等人，2018b）	（Fan 等，2023）	本研究

表 2：混合 ResNet-18 模型在图像分类任务上的准确率（%）。

方法	CIFAR-10	CIFAR-100
混合庞加莱（Guo 等，2022）	95.04±0.13	77.19±0.50
庞加莱 ResNet（van Spengler 等，2023）	94.51±0.15	76.60±0.32
欧几里得-庞加莱-H（Fan 等，2023）	81.72±7.84	44.35±2.93
欧几里得-庞加莱-G（Ganea 等人，2018b）	95.14±0.11	**77.78**±0.09
欧几里得-庞加莱-B（我们的方法）	**95.23**±0.08	**77.78**±0.15

5.2 注意力机制

We 采用与 Shimizu 等人（2021）类似的方法。缩放点积注意力（Vaswan i 等人., 2017）表述为

$$\mathrm{att}(q,z,v)=\mathrm{softmax}\left(\frac{qz^T}{\sqrt{m_z}}\right)v, \tag{5}$$

其中， $q,z\in\mathbb{R}^{l\times m_z}$ 和 $v\in\mathbb{R}^{l\times m_v}$ 分别是查询、键和值， l 是序列长度， m_z 和 m_v 分别是查询（键）和值的隐藏维度，而 functionsoftmax(.)通过对该矩阵的每一行应用 s oftmax 函数，生成一个与输入矩阵大小相同的矩阵。

矩阵乘积 qz^T 对应一个注意力函数，用于确定所有查询-键对之间的相似度。function softmax(.) 与 $v=[v_1^T;\ldots;v_l^T]$ 的乘积得到值 $v_j, j=1,\ldots,l$ 的加权均值，因此可视为一种中点运算。在自注意力中，查询、键和值是对同一输入序列进行不同线性投影所得。因此，式（5）可改写为

$$\mathrm{att}(f_{lin}^q, f_{lin}^z, f_{lin}^v, (x_j)_{j=1}^l)=f_{mid}(\{f_{lin}^v(x_j),\pi_{j'j}\}_{j=1}^l)$$

对所有 $j'=1,\ldots,l$ 均成立，其中 $(x_j)_{j=1}^l$ 是输入序列， $f_{lin}^q(.)$、 $f_{lin}^z(.)$、 $f_{lin}^v(.)$ 是将输入点分别投影到查询、键和值的线性函数，$(\pi_{j'j})_{j=1}^l$ =softmax $\left(\left(f_{att}(f_{lin}^q(x_{j'})\ f_{lin}^z(x_j))\right)_{j=1}^l\right)$, $f_{att}(.,.)$ 是注意力函数，而 $f_{mid}(.)$ 是中点运算。我们使用所提出的全连接层（见图 2 (a)）在$f_{lin}^q(.)$、 $f_{lin}^z(.)$ 和 $f_{lin}^v(.)$ 中进行线性投影。注意力函数（Gulcehre 等人，2018；Shimizu 等人，2021）定义为

$$f_{att}(f_{lin}^q(x_{j'}), f_{lin}^z(x_j))=-c_1 d(f_{lin}^q(x_{j'}), f_{lin}^z(x_j))-c_2,$$

其中 c_1， $c_2\in\mathbb{R}$， $c_1>0$ 是可学习参数。我们采用加权弗雷歇均值（wFM）进行中点运算。

6 实验

本节报告我们在图像分类和脑电信号分类任务上的实验评估。实验详情请参阅附录 B；关于图像生成和自然语言推理的实验评估，请参阅附录 C 和 D。

Tab表 3：我们提出的网络与先进方法在脑电信号分类中的准确率（%） .

方法	BCIC-IV-2a	MAMEM-SSVEP-II	BCI-NER
EEG-TCNet（Ingolfsson 等人，2020年）	67.09±4.6	55.45±7.6	77.05±2.4
MBEEGSE（Altuwaijri 等人，2022年）	64.58±6.0	56.45±7.2	75.46±2.3
MAtt（Pan 等，2022年）	74.71±5.0	65.50±8.2	76.01±2.2
Graph-CSPNet（Ju 与 Guan，2023）	71.95±13.3	-	-
AttSymSpd-LE（我们的方法）	**78.24** ± 5.4	**70.96** ± 8.6	**78.02** ± 2.3
AttSymSpd-GI（我们的方法）	78.08 ± 4.8	67.24 ± 7.4	75.88 ± 2.2

6.1 双曲空间

我们遵循 Bdeir 等人（2024）的做法，设计了一种混合架构[1]，由 ResNet-18（He 等人，2016）和 Poincar´é MLR（Ganea 等人，2018b）组成。ResNet-18 的输出会先映射到 Poincar´é 球，再输入 Poincar´é MLR。我们在 Poincar´é MLR 中采用了所提出的点到超平面距离，以及 Ganea 等人（2018b）和 Fan 等人（2023）提出的距离（见表 1）。实验使用 CIFAR-10 和 CIFAR-100 数据集（Krizhevsky，2009）进行。表 2 展示了由此得到的三个网络的结果，以及取自 Bdeir 等人（2024）的 Hybrid Poincar´é（Guo 等，2022）和 Poincar´é ResNet（van Spengler 等，2023）的结果。Hybrid Poincar´é 与 Euclidean-Poincar´é-G 的唯一区别在于其 Poincar´é MLR，该模型采用了 Shimizu 等人（2021）的重参数化方法。我们的网络在两个数据集上都取得了最高的平均准确率。尤其是，在 CIFAR-10 数据集上，它优于 Bdeir 等人（2024）的所有双曲神经网络模型，包括全双曲模型。请注意，g-距离是点到 Poincar´é 超平面的最近距离，而 b-距离则被设计为对称空间中点到欧几里得超平面最近距离的对应形式。不过，h-距离是通过 horosphere 投影得到的，旨在保留主成分分析中的一项重要性质，即点之间的距离在沿正交方向平移时保持不变。因此，h-距离与 g-距离和 b-距离的性质并不相同。这或许可以解释为何 Euclidean-Poincar´é-H 的表现逊于其他模型。在我们关于图像生成和自然语言推理的实验中，也能观察到 h-距离表现较差（见附录 C 和 D）。此外，这些实验表明：（1）对于图像生成，b-距离在所有情况下的平均表现都优于 g-距离；（2）对于自然语言推理，在大多数情况下，前者的平均表现优于后者。这表明，我们提出的距离有望提升现有的双曲神经网络。

6.2 对称正定流形

我们在三个脑电图信号分类数据集上验证了所提出的对称正定神经网络构建模块（见第 5 节）：BCIC-IV-2a（Brunner等人，2008年）、MAMEM-SSVEP-II（Nikolopoulos，2021年）和 BCI-NER（Perrin等人，2012年）。我们测试了图 2 (b) 所示网络架构的两个变体。第一个网络 AttSymSpd-LE 的注意力块中使用的全连接层基于对数欧几里得度量（见第 4.3 节），而第二个网络 AttSymSpd-GI 的全连接层则基于 G 不变度量（见第 4.4 节和第 5.1 节）。
表 3 展示了我们的网络与一些先进方法的结果。这些方法大多依据两项标准选取（Pan 等，2022）：（1）代码可用且完整；（2）评估扎实（例如跨会话评估），且不需要额外的辅助流程。如结果所示，AttSymSpd-LE 在所有数据集上表现最佳。在 BCIC-IV-2a 数据集上，AttSymSpd-GI 的表现与 AttSymSpd-LE 不相上下。尽管 AttSymSpd-GI 在 MAMEM- 稳态视觉诱发电位-II 和 BCI-NER 数据集上的表现不及 AttSymSpd-LE，但前者的优势是参数数量远少于后者。例如，在 BCIC-IV-2a 数据集上，AttSymSpd-GI 和 AttSymSpd-LE 分别使用 0.007 MB 和 0.034 MB 的可学习参数（另见附录B.2.3）。

[1]https://github.com/nguyenxuanson10/symspaces-ic.

致谢

我们感谢匿名审稿人提出的建设性意见和反馈。参考文献 Ghadir Ali Altuwaijri、Ghulam Muhammad、Hamdi Altaheri 和 Mansour Alsulaiman。《一种用于脑电图运动想象信号分类的多分支卷积神经网络与挤压-激励注意力模块》。*Diagnostics*，12(4):995，2022年。

Vincent Arsigny、Pierre Fillard、Xavier Pennec 和 Nicholas Ayache。基于对数欧几里得度量的张量快速简便计算。技术报告 RR-5584，INRIA，2005。

Werner Ballmann。 非正曲率空间讲义。Birkh¨auser，2012年。

Francesca Bartolucci、Filippo De Mari 和 Matteo Monti。对称空间上的双曲环状拉东变换的酉化。*CoRR*，abs/2108.04338，2021年。网址：http://arxiv.org/ abs/2108.04338。

Ahmad Bdeir、Kristian Schwethelm 和 Niels Landwehr。用于计算机视觉的全双曲卷积神经网络。发表于*ICLR*，2024。Cl´ement Bonet、Benoit Mal´ezieux、Alain Rakotomamonjy、Lucas Drumetz、Thomas Moreau、Matthieu Kowalski 和 Nicolas Courty。用于脑磁/脑电信号的对称正定矩阵切片 Wasserstein。发表于*ICML*，第 2777–2805 页，2023 年。Silv`ere Bonnabel。黎曼流形上的随机梯度下降。*IEEETransactions onAutomatic Control*，58(9):2217–2229，2013。Samuel R. Bowman、Gabor Angeli、Christopher Potts 和 Christopher D. Manning。用于自然语言推理学习的大型标注语料库。发表于*EMNLP*，第 632–642 页，2015。M.R. Bridson 和 A. H¨afliger。非正曲率度量空间。Springer Berlin Heidelberg，2011。

Michael M. Bronstein、Joan Bruna、Yann LeCun、Arthur Szlam 和 Pierre Vandergheynst。《几何深度学习：超越欧几里得数据》。*IEEE* 信号处理杂志，34(4)：18–42，2017。

Daniel A. Brooks、Olivier Schwander、Fr´ed´eric Barbaresco、Jean-Yves Schneider 和 Matthieu Cord。面向 SPD 神经网络的黎曼批归一化。载于*NeurIPS*，第 15463–15474 页，2019 年。Clemens Brunner、Robert Leeb、Gernot M¨uller-Putz、Alois Schl¨ogl 和 Gert Pfurtscheller。BCI Competition 2008-Graz 数据集 A。知识发现研究所（脑机接口实验室），格拉茨技术大学，16:1–6，2008 年。Rudrasis Chakraborty、Jose Bouza、Jonathan H. Manton 和 Baba C. Vemuri。ManifoldNet：一种用于流形值数据的深度神经网络及其应用。*TPAMI*，44(2):799–810，2020 年。Ines Chami、Rex Ying、Christopher R´e 和 Jure Leskovec。双曲图卷积神经网络。*CoRR*，abs/1910.12933，2019 年。网址：https://arxiv.org/abs/1910.12933。Ines Chami、Albert Gu、Dat Nguyen 和 Christopher R´e。HoroPCA：通过 horosphere 投影实现双曲降维。载于*ICML*，第 1419–1429 页，2021 年。Ziheng Chen、Yue Song、Gaowen Liu、Ramana Rao Kompella、Xiaojun Wu 和 Nicu Sebe。用于 SPD 神经网络的黎曼多项逻辑回归。*CoRR*，abs/2305.11288，2024a。网址：http://arxiv.org/abs/2305.11288。Ziheng Chen、Yue Song、Yunmei Liu 和 Nicu Sebe。一种基于李群的黎曼批归一化方法。*CoRR*，abs/2403.11261，2024b。网址：http://arxiv.org/abs/2403. 11261。

Ziheng Chen、Yue Song、Tianyang Xu、Zhiwu Huang、Xiao-Jun Wu 和 Nicu Sebe。用于 SPD 矩阵学习的自适应对数欧几里得度量。*IEEE*图像处理汇刊，33:5194–5205，2024c。Taco S. Cohen、Mario Geiger、Jonas Koehler 和 Max Welling。球面卷积神经网络。*CoRR*，abs/1801.10130，2018 年。网址：https://arxiv.org/abs/1801.10130。Terrance DeVries 和 Graham W. Taylor。利用 Cutout 改进卷积神经网络的正则化。*CoRR*，abs/1708.04552，2017 年。网址：https://arxiv.org/abs/1708. 04552。Xiran Fan、Chun-Hao Yang 和 Baba C. Vemuri。双曲空间中用于大间隔分类的超球面决策边界。*CoRR*，abs/2302.06807，2023。网址：http: //arxiv.org/abs/2302.06807。Octavian Ganea、Gary Becigneul 和 Thomas Hofmann。用于学习层次嵌入的双曲蕴含锥。载于*ICML*，第 1646–1655 页，2018a。Octavian-Eugen Ganea、Gary B´ecigneul 和 Thomas Hofmann。双曲神经网络。载于*NeurIPS*，第 5350–5360 页，2018b。Caglar Gulcehre、Misha Denil、Mateusz Malinowski、Ali Razavi、Razvan Pascanu、Karl Moritz Hermann、Peter Battaglia、Victor Bapst、David Raposo、Adam Santoro 和 Nando de Freitas。双曲注意力网络。*CoRR*，abs/1805.09786，2018 年。网址：https://arxiv.org/ abs/1805.09786。Yunhui Guo、Xudong Wang、Yubei Chen 和 Stella X Yu。截断双曲分类器即超双曲分类器。载于*CVPR*，第 11–20 页，2022 年。Kaiming He、Xiangyu Zhang、Shaoqing Ren 和 Jian Sun。用于图像识别的深度残差学习。载于*CVPR*，第 770–778 页，2016 年。S. Helgason。微分几何、李群与对称空间。ISSN。Elsevier Science，1979 年。S. Helgason。群与几何分析：积分几何、不变微分算子与球函数。群与几何分析丛书第 1 卷。Academic Press，1984 年。S. Helgason。对称空间上的几何分析。数学调查与专著。美国数学会，1994 年。Zhiwu Huang 和 Luc Van Gool。用于 SPD 矩阵学习的黎曼网络。载于*AAAI*，第 2036–2042 页，2017 年。Zhiwu Huang、Chengde Wan、Thomas Probst 和 Luc Van Gool。用于基于骨架的动作识别的李群上的深度学习。载于*CVPR*，第 6099–6108 页，2017 年。Zhiwu Huang、Jiqing Wu 和 Luc Van Gool。在格拉斯曼流形上构建深度网络。载于*AAAI*，第 3279–3286 页，2018 年。Thorir Mar Ingolfsson、Michael Hersche、Xiaying Wang、Nobuaki Kobayashi、Lukas Cavigelli 和 Luca Benini。EEG-TCNet：一种用于嵌入式运动想象脑机接口的高精度时序卷积网络。载于*2020*年*IEEE*系统、人类与控制论国际会议，第 2958–2965 页，2020 年。Catalin Ionescu、Orestis Vantzos 和 Cristian Sminchisescu。用于结构化层深度网络的矩阵反向传播。载于*ICCV*，第 2965–2973 页，2015 年。Ce Ju 和 Cuntai Guan。用于运动想象分类的 SPD 流形上的图神经网络：时频分析视角。*IEEE*神经网络与学习系统汇刊，第 1–15 页，2023 年。

Michael Kapovich、Bernhard Leeb 和 Joan Porti。《Anosov 子群：动力学与几何刻画》。*CoRR*，abs/1703.01647，2017。网址：https://arxiv.org/abs/1703. 01647。Hermann Karcher。《黎曼质心与磨光函数平滑》。《纯粹与应用数学通讯》，30(5):509–541，1977。Fanny Kassel。《余秩一约化齐性空间上的适当作用》。*CoRR*，abs/0807.3980，2009。网址：http://arxiv.org/abs/0807.3980。Isay Katsman、Eric Ming Chen、Sidhanth Holalkere、Anna Asch、Aaron Lou、Ser-Nam Lim 和 Christopher De Sa。《黎曼残差神经网络》。*CoRR*，abs/2310.10013，2023。网址：http://arxiv.org/abs/2310.10013。Diederik P. Kingma 和 Jimmy Ba。Adam：一种随机优化方法。载于 *ICLR*，2015。Thomas N. Kipf 和 Max Welling。基于图卷积网络的半监督分类。*CoRR*，abs/1609.02907，2017。网址：https://arxiv.org/abs/1609.02907。Reinmar J. Kobler、Jun ichiro Hirayama、Qibin Zhao 和 Motoaki Kawanabe。用于破解脑电图可解释无监督域适应难题的 SPD 领域特定批归一化。载于*NeurIPS*，第 6219–6235 页，2022。Max Kochurov、Rasul Karimov 和 Serge Kozlukov。Geoopt：PyTorch 中的黎曼优化。*CoRR*，abs/2005.02819，2020。网址：https://arxiv.org/abs/2005.02819。Alex Krizhevsky。从微型图像中学习多层特征。第 32–33 页，2009。网址：https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf。Vernon J Lawhern、Amelia J Solon、Nicholas R Waytowich、Stephen M Gordon、Chou P Hung 和 Brent J Lance。EEGNet：一种用于脑机接口的紧凑型卷积神经网络。《神经工程学杂志》，15(5):056013，2018。Guy Lebanon 和 John Lafferty。多项式流形上的超平面间隔分类器。载于*ICML*，第 66 页，2004。Mario Lezcano-Casado。流形上基于梯度的优化的平凡化。载于*NeurIPS*，第 9154–9164 页，2019。Zhenhua Lin。通过 Cholesky 分解研究对称正定矩阵的黎曼几何。《*SIAM* 矩阵分析与应用杂志》，40(4):1353–1370，2019。Qi Liu、Maximilian Nickel 和 Douwe Kiela。双曲图神经网络。载于*NeurIPS*，第 8228–8239 页，2019。Federico L´opez、Beatrice Pozzetti、Steve Trettel、Michael Strube 和 Anna Wienhard。对称正定矩阵空间上的向量值距离与陀螺演算。载于 *NeurIPS*，第 18350–18366 页，2021。Aaron Lou、Isay Katsman、Qingxuan Jiang、Serge Belongie、Ser-Nam Lim 和 Christopher De Sa。通过 Fr´echet 均值求导。载于 *ICML*，第 6393–6403 页，2020。Ravikiran Mane、Effie Chew、Karen Chua、Kai Keng Ang、Neethu Robinson、A. P. Vinod、Seong-Whan Lee 和 Cuntai Guan。FBCNet：一种用于脑机接口的多视图卷积神经网络。*CoRR*，abs/2104.01233，2021。网址：https://arxiv.org/abs/2104. 01233。Yazeed K. Musallam、Nasser I. AlFassam、Ghulam Muhammad、Syed Umar Amin、Mansour Al-sulaiman、Wadood Abdul、Hamdi Altaheri、Mohamed A. Bencherif 和 Mohammed Algabri。利用时序卷积网络融合进行基于脑电图的运动想象分类。《生物医学信号处理与控制》，69:102826，2021。Yoshihiro Nagano、Shoichiro Yamaguchi、Yasuhiro Fujita 和 Masanori Koyama。用于基于梯度学习的双曲空间上的包裹正态分布。载于*ICML*，第 4693–4702 页，2019。

Galileo Namata、Ben London、Lise Getoor 和 Bert Huang。用于集体分类的查询驱动主动调查。载于图挖掘与学习研讨会，2012。Xuan Son Nguyen。GeomNet：一种基于 SPD 矩阵空间和 Cholesky 空间的黎曼几何、用于基于三维骨架的交互识别的神经网络。载于*ICCV*，第 13379–13389 页，2021。Xuan Son Nguyen。一种用于对称半正定矩阵学习的陀螺向量空间方法。载于*ECCV*，第 52–68 页，2022a。Xuan Son Nguyen。若干矩阵流形的陀螺结构。载于*NeurIPS*，第 26618–26630 页，2022b。Xuan Son Nguyen 和 Shuo Yang。构建矩阵流形上的神经网络：一种陀螺向量空间方法。载于*ICML*，第 26031–26062 页，2023。Xuan Son Nguyen、Luc Brun、Olivier Lezoray 和 S´ebastien Bougleux。通过神经网络学习 SPD 矩阵，实现基于骨架的手势识别。*CoRR*，abs/1905.07917，2019a。Xuan Son Nguyen、Luc Brun、Olivier L´ezoray 和 S´ebastien Bougleux。一种基于 SPD 流形学习、用于基于骨架的手势识别的神经网络。载于*CVPR*，第 12036–12045 页，2019b。Xuan Son Nguyen、Luc Brun、Olivier L´ezoray 和 S´ebastien Bougleux。学习用于基于骨架的手势识别的高阶循环统计量。载于*ICPR*，第 975–982 页，2020。Xuan Son Nguyen、Shuo Yang 和 Aymeric Histace。矩阵流形神经网络++。载于*ICLR*，2024。Spiros Nikolopoulos。MAMEM EEG SSVEP数据集II（256个通道、11名受试者、同时呈现5种频率）。2021。Yue-Ting Pan、Jing-Lun Chou 和 Chun-Shu Wei。MAtt：用于脑电解码的流形注意力网络。*CoRR*，abs/2210.01986，2022。网址：https://arxiv.org/abs/2210.01986。Margaux Perrin、Emmanuel Maby、S´ebastien Daligault、Olivier Bertrand 和 J´er´emie Mattout。基于P300的拼写中在线错误校正的客观与主观评估。《*Adv. Hum. Comput. Interact.*》，2012:578295:1–578295:13，2012。P. Sawyer。利用 QR 分解计算非紧型经典李群的伊瓦萨瓦分解。《*Linear Algebra and its Applications*》，493:573–579，2016。Robin Tibor Schirrmeister、Jost Tobias Springenberg、Lukas Dominique Josef Fiederer、Martin Glasstetter、Katharina Eggensperger、Michael Tangermann、Frank Hutter、Wolfram Burgard 和 Tonio Ball。利用卷积神经网络进行脑电解码与可视化的深度学习。人脑映射，38(11):5391–5420，2017。Prithviraj Sen、Galileo Mark Namata、Mustafa Bilgic、Lise Getoor、Brian Gallagher 和 Tina Eliassi-Rad。网络数据中的集体分类。《人工智能杂志》，29(3):93–106，2008。Ryohei Shimizu、Yusuke Mukuta 和 Tatsuya Harada。双曲神经网络++。*CoRR*，abs/2006.08210，2021。网址：https://arxiv.org/abs/2006.08210。Sho Sonoda、Isao Ishikawa 和 Masahiro Ikeda。非紧型对称空间上的全连接网络及基于 Helgason-傅里叶分析的脊波变换。载于*ICML*，第 20405–20422 页，2022。Max van Spengler、Erwin Berkhout 和 Pascal Mettes。Poincar´e ResNet。载于*ICCV*，第 5396–5405 页，2023。Ashish Vaswani、Noam Shazeer、Niki Parmar、Jakob Uszkoreit、Llion Jones、Aidan N Gomez、Lukasz Kaiser 和 Illia Polosukhin。注意力机制就是你所需要的一切。载于*NIPS*，第 5998–6008 页，2017。

Petar Veli˘ckovi´c、Guillem Cucurull、Arantxa Casanova、Adriana Romero、Pietro Li`o 和 Yoshua Bengio。《图注意力网络》。*CoRR*，abs/1710.10903，2018年。网址：https://arxiv. org/abs/1710.10903。

Ming-Xi Wang。《双曲空间上的拉普拉斯特征空间、等距曲线与神经元模型》，2021年。

Rui Wang、Xiao-Jun Wu 和 Josef Kittler。《SymNet：一种用于图像集分类的简单对称正定流形深度学习方法》。*IEEE*神经网络与学习系统汇刊，第 1–15 页，2021年。

Chun-Shu Wei、Toshiaki Koike-Akino 和 Ye Wang。《用于运动想象脑电图分类的空间分量卷积网络（SCCNet）》。载于*2019*年第九届*IEEE/EMBS*国际神经工程会议 *(NER)*，第 328–331 页，2019年。

Maurice Weiler、Patrick Forr´e、Erik Verlinde 和 Max Welling。《坐标无关卷积网络——黎曼流形上的等距与规范等变卷积》。*CoRR*，abs/2106.06020，2021年。网址：https://arxiv.org/abs/2106.06020。

Tao Yu 和 Christopher De Sa。《用于双曲空间学习的随机拉普拉斯特征》 In*ICLR*，2023。

Xingjian Zhen、Rudrasis Chakraborty、Nicholas Vogt、Barbara B. Bendlin 和 Vikas Singh。《用于序列流形值数据的膨胀卷积神经网络》。发表于*ICCV*，第10620–10630页，2019年。

A 实现细节

本节将详细介绍我们对全连接层和注意力模块的实现，以及相关的反向传播过程。

A.1 全连接层

输入： $x \in X$.

可训练参数： $k_j \in K, a_j \in \mathfrak{a}, \|a_j\| = 1, p_j \in X, j = 1, \ldots, m, n \in N$.

输出： $y \in X$.

对于对称正定流形，我们注意到：

- $K = O_m$ （ $m \times m$ 正交矩阵群）。
- A 是由对角线元素为正的 $m \times m$ 对角矩阵组成的子群。
- N 是由对角线元素为 1 的 $m \times m$ 上三角矩阵组成的子群。

因此，可以相应地推导出输入、输出和可训练参数的维度（$a_j \in \mathbb{R}^m$、 $j = 1$、.....、 m）。

全连接层执行的计算如下。

步骤 *1*： 计算 $\ominus p_j \oplus x = h_j^{-1} g K$，其中 $p_j = h_j K$、 $x = gK$、 h_j、 $g \in G$、 $j = 1$、.....、 m。

步骤 *2*： 计算 $g_j \in G$、 $j = 1$、.....、 m，使得 $k_j^{-1}[\ominus p_j \oplus x] = g_j K$。

步骤 *3*： 根据 g_j^{-1} 的伊瓦萨瓦分解计算 $H(g_j^{-1})$ $j = 1$，.....， m，即我们需要确定映射 H： $G \to \mathfrak{a}$，使得 $g_j^{-1} = K \exp(H(g_j^{-1}))N$。

步骤 *4*： 计算 $B_{\xi_j}(\ominus p_j \oplus x)$ $j = 1$，.....， m，即

$$B_{\xi_j}(\ominus p_j \oplus x) = \langle a_j, H(g_j^{-1}) \rangle.$$

步骤 *5*： 计算全连接层的输出，即

$$y = n \exp\left(\begin{bmatrix} -B_{\xi_1}(\ominus p_1 \oplus x) & \cdot & \cdot \\ \cdot & -B_{\xi_2}(\ominus p_2 \oplus x) & \cdot \\ \cdot & \cdot & -B_{\xi_m}(\ominus p_m \oplus x) \end{bmatrix}\right) K$$

反向传播 下面我们描述对称正定流形情况下全连接层的反向传播过程。

步骤 *1*： 设 $p_j = u_j s_j u_j^T$和 $x = u_x s_x u_x^T$分别是 p_j和 x的特征分解，其中 u_j, u_x是正交矩阵， s_j, s_x是对角矩阵。于是

$$\ominus p_j \oplus x = s_j^{-\frac{1}{2}} u_j^{-1} u_x s_x^{\frac{1}{2}} K.$$

在 TensorFlow 和 PyTorch 框架中，特征分解是可微运算。关于奇异值分解和特征分解的梯度计算，详见 Ionescu 等人（2015）的讨论。

Step 2: Compute $g_j = k_j^{-1} s_j^{-\frac{1}{2}} u_j^{-1} u_x s_x^{\frac{1}{2}}$.

步骤 *3*： 设 $g \in G$ 和 $g = kan$ 满足 $k \in K$、 $a \in A$和 $n \in N$。于是

$$g^T g = n^T a^T k^T kan = n^T a^2 n = n^T a (n^T a)^T.$$

设 $g^T g = cc^T$是 $g^T g$的 Cholesky 分解（c是下三角矩阵），并设 s是包含 c主对角线元素的对角矩阵。于是 $n = (cs^{-1})^T$且 $a = s$。映射 $H: G \to \mathfrak{a}$ 可表示为 $H(g) = \log(a)$。我们看到，映射 H 可通过 Cholesky 分解确定，而该分解在 TensorFlow 和 PyTorch 框架中是可微运算。

步骤 *4*： 反向传播按正常方式进行，计算内积 $\langle a_j, H(g_j^{-1}) \rangle$。

步骤 *5*： 全连接层的输出计算如下

$$y = n \exp\left(2 \begin{bmatrix} -B_{\xi_1}(\ominus p_1 \oplus x) & \cdot & \cdot \\ \cdot & -B_{\xi_2}(\ominus p_2 \oplus x) & \cdot \\ \cdot & \cdot & -B_{\xi_m}(\ominus p_m \oplus x) \end{bmatrix}\right) n^T,$$

其中仅涉及矩阵乘法。

A.2 注意力模块

注意力模块的流程如图 2(a) 所示。

输入： 点序列 $x_j \in X$， $j = 1, ..., l$。

可训练参数：其中包括三个全连接层的可训练参数，以及 c_1， $c_2 \in \mathbb{R}$（由注意力函数使用）。

输出：点序列 $y_j \in X$， $j = 1, ..., l$。

对于对称正定矩阵流形，输入、输出和可训练参数的维度由全连接层中的维度推得（见上文）。

该模块执行的计算如下。

步骤 *1*：将三个全连接层应用于输入点序列 x_j，并获得三个点序列（q_j）$_{j=1}^{l}$（查询）、（z_j）$_{j=1}^{l}$（键）和（v_j）$_{j=1}^{l}$（值）：

$$(q_j)_{j=1}^{l} = f_{lin}^{q}((x_j)_{j=1}^{l}), \quad (z_j)_{j=1}^{l} = f_{lin}^{z}((x_j)_{j=1}^{l}), \quad (v_j)_{j=1}^{l} = f_{lin}^{v}((x_j)_{j=1}^{l}),$$

whe其中 f_{lin}^{q}（。）、 f_{lin}^{z}（。）和 f_{lin}^{v}（。）是三个全连接层执行的线性变换。

步骤 *2*： 根据黎曼距离函数计算查询与键之间的相似度 $\bar{\pi}_{j'j}$， j'， $j = 1, ..., l$

$$\bar{\pi}_{j'j} = -c_1 d(q_{j'}, z_j) - c_2.$$

步骤 *3*： 计算注意力权重 $\pi_{j'j}$， j'， $j=1$，……， l

$$(\pi_{j'j})_{j=1}^{l}=\operatorname{softmax}\left((\bar{\pi}_{j'j})_{j=1}^{l}\right).$$

步骤 *4*： 执行中点运算，以获得输出点序列 $y_{j'}$、 $j'=1$、……、 l，其形式为

$$y_{j'}=f_{mid}\left(\{v_j,\pi_{j'j}\}_{j=1}^{l}\right),$$

其中，$\pi_{j'j}$ 是与 v_j相关联的注意力权重。

反向传播 我们描述对称正定流形情况下注意力模块的反向传播过程。这里我们只关注“黎曼距离”和“中点运算”这两个模块。对于 AttSymSpd-GI，“黎曼距离”模块中使用的黎曼距离为

$$d(x,y)=\|\log(y^{-\frac{1}{2}}xy^{-\frac{1}{2}})\|,$$

其中 x、 $y\in\mathrm{Sym}_m^+$。项 $y^{-\frac{1}{2}}$的计算方式为

$$y^{-\frac{1}{2}}=us^{-\frac{1}{2}}u^T,$$

其中， $y=usu^T$ 是 y的特征值分解， u 是正交矩阵， s 是对角矩阵。log(。)函数也通过对其参数进行特征值分解获得。

“中点运算”模块在对数欧几里得框架下执行加权弗雷歇均值操作。设 $\{x_j,w_j\}_{j=1}^{L}$ 为一组点 $x_j\in\mathrm{Sym}_m^+$，其对应的权重为 $w_j\in\mathbb{R}$，其中 $w_j>0$ 且$\sum_{j=1}^{L}w_j=1$。这些点的加权弗雷歇均值为

$$\mathrm{wFM}(\{x_j,w_j\}_{j=1}^{L})=\exp\left(\sum_{j=1}^{L}w_j\log(x_j)\right).$$

exp(.)函数通过特征值分解计算为 exp(y) $=u\exp(s)u^T$，其中 u 是正交矩阵， s 是对角矩阵。因此，wFM计算中涉及的所有函数都基于特征分解，反向传播过程如上文所述。

B 实验详情

B.1 图像分类

B.1.1 数据集

CIFAR-10 和 CIFAR-100（Krizhevsky，2009）CIFAR-10 和 CIFAR-100 数据集分别包含来自 10 个和 100 个不同类别的 60K32×32 张彩色图像。我们使用 PyTorch 中实现的数据集划分，其中包含 50K 张训练图像和 10K 张测试图像。

B.1.2 实验设置

网络架构 Euclidean-Poincar´e-G、Euclidean-Poincar´e-B 和 Euclidean-Poincar´e-H 具有相同的网络架构，由 ResNet-18 和 Poincar´e MLR 组成。这里我们仅介绍 Poincar´e MLR。设 L 为类别数，则 MLR 计算每个输出类别的概率，如下所示

$$\mathrm{prop}(y=l|x)=\frac{\exp(a_l^Tx-b_l)}{\sum_{j=1}^{L}\exp(a_j^Tx-b_j)}\propto\exp(a_l^Tx-b_l),\tag{6}$$

其中 $x\in\mathbb{R}^m$ 是输入， $b_j\in\mathbb{R}$、 $a_j\in\mathbb{R}^m$， $j=1$，……， L 是模型参数。可以将（Lebanon 与 Lafferty，2004）的式（6）表示为

$$\mathrm{prop}(y=l|x)\propto\exp(\mathrm{sign}(a_l^Tx-b_l)\|a_l\|\bar{d}(x,\mathcal{H}_{a_l,b_l}^{E})),\tag{7}$$

表4：图像分类实验每个训练轮次的计算时间（秒）（在 Quadro RTX 8000 GPU 上测量）。

方法	欧几里得-Poincar´e-H	欧几里得-Poincar´e-G	欧几里得-Poincar´e-B
CIFAR-10	52	57	59
CIFAR-100	90	114	118

其中 $\bar{d}\left(x,\ \mathcal{H}^{E}_{a_l,\ b_l}\right)$ 是 x 与超平面 $\mathcal{H}^{E}_{a_l,\ b_l}$ 之间的距离（见第4.1节）。在庞加莱´MLR（Ganea 等人，2018b）中，式（7）写作

$$\operatorname{prop}(y=l|x) \propto \exp\left(\frac{2}{(1-\|p_l\|^2)}\|a_l\|\sinh^{-1}\left(\frac{2|\langle -p_l \oplus_M x, a_l\rangle|}{(1-\|-p_l\oplus_M x\|^2)\|a_l\|}\right)\right), \tag{8}$$

其中 $x, p_l \in \mathbb{B}_m,\quad a_l \in T_{p_l}\mathbb{B}_m \setminus \{\mathbf{0}\}, l = 1, \ldots, L$。

Euclidean-Poincar´e-G 使用公式 (8) 计算每个输出类别的概率。Euclidean-Poincar´e-B 和 Euclidean-Poincar´e-H 分别通过将公式 (8) 中的点到超平面的距离替换为我们提出的距离和 Fan 等人（2023）提出的距离构建而成。

超参数 我们严格遵循 DeVries和Taylor（2017）以及 Bdeir 等人（2024）的设置。训练时采用随机镜像和裁剪。批量大小和训练轮数分别设为 128 和 200。学习率和权重衰减分别设为 $1e-1$ 和 $5e-4$。对于自适应学习率调度，训练轮次设为 60、120 和 160，伽马因子设为 0。2。

优化与评估 所有模型均使用 PyTorch 实现。我们使用 Geoopt 库（Kochurov等，2020）进行黎曼优化。使用黎曼随机梯度下降训练网络。每个模型的结果均为 5 次运行的平均值。所有实验均使用 Quadro RTX 8000 GPU。

B.1.3 更多结果

表 4 展示了用于图像分类实验的 Euclidean-Poincar´e-H、Euclidean-Poincar´e-G 和 Euclidean-Poincar´e-B 的计算时间。可以观察到，与其他方法相比，Euclidean-Poincar´e-B 的计算成本较高，而 Euclidean-Poincar´e-H 是这三种方法中最快的。

B.2 脑电信号分类

B.2.1 数据集

BCIC-IV-2a 该数据集包含从 9 名受试者采集的脑电图数据。基于提示的脑机接口范式包含 4 种不同的运动想象任务，分别是想象左手运动（类别 1）、右手运动（类别 2）、双脚运动（类别 3）和舌头运动（类别 4）。每名受试者在不同日期进行了两次实验。每次实验包含 6 个由短暂休息隔开的运行。每个运行包含 48 次试验（4 种可能类别各 12 次），因此每次实验共计 288 次试验。信号由 22 个 Ag/AgCl传感器记录（电极间距为 3.5 厘米），采样率为 250 Hz。信号经过 0.5 Hz 至 100 Hz 的带通滤波。

MAMEM-SSVEP-II 该数据集包含 11 名受试者执行基于稳态视觉诱发电位的实验方案时采集的 256 通道脑电图数据。视觉刺激采用五种不同频率（6.66、7.50、8.57、10.00 和 12.00 Hz），并使用 EGI 300测地线脑电系统（GES 300）采集信号；该系统配备 256 通道 HydroCel测地线传感器网（HCGSN），采样率为 250 Hz。

BCI-NER 该数据集包含从26名受试者采集的脑电图数据。脑电电极的放置遵循扩展10–20系统。每名受试者均进行了五次实验（前四次实验各60个试次，

最后一次实验100个试次），单个脑电图试次的时长为1.25秒。信号由56个无源Ag/AgCl传感器（兼容VSM-CTF的系统）采集，采样频率为600 Hz。我们在实验中使用了Kaggle竞赛早期阶段发布的16名受试者的数据[2]。

B.2.2 实验设置

网络架构 受 Huang & Gool（2017）的启发，我们对输入数据应用若干卷积层以提取特征。提取出的特征序列被划分为若干不重叠子序列，每个子序列都构成一个 SPD 矩阵（Huang & Gool，2017）。这些步骤生成一个 SPD 矩阵序列，并将其输入注意力块（见第 5.2 节）。注意力块的每个输出 SPD 矩阵都通过对数映射投影到单位矩阵处的切空间（Huang & Gool，2017）。所得矩阵随后被转换为向量，再拼接起来，形成用于分类的最终特征。网络架构如图 2（b）所示。

对于 AttSymSpd-LE，我们使用命题 4.4 中推导的距离，并结合 $\phi(\text{。}) = \log(\text{。})$，在注意力模块中构建全连接层。如第 4.3 节所述，我们对超平面的定义以及推导出的距离，分别与对称正定双曲超平面的定义以及 SPD 矩阵与对称正定双曲超平面之间的距离相吻合。因此，我们可以采用 Nguyen 等（2024）的方法。设 $x \in \text{Sym}^+_{m'}$ 为全连接层的输入，并设 $v_{(l,j)}(x) = \langle \ominus_{le} p_{(l,j)} \oplus_{le} x,\ a_{(l,j)} \rangle^{le},\ p_{(l,j)},\ a_{(l,j)} \in \text{Sym}^+_{m'},\ l \le j,\ l,\ j = 1, \ldots,\ m$（运算 $\oplus_{le}$，$\ominus_{le}$ 以及 SPD内积 $\langle ., .\rangle^{le}$的定义见附录L.7.2）。那么，全连接层的输出 y 计算如下

$$y = \exp\left([z_{(l,j)}]_{l,j=1}^{m}\right),$$

其中$z_{(l,j)}$ 由下式给出

$$z_{(l,j)} = \begin{cases} v_{(l,j)}(x), & \text{if } l = j \\ \frac{1}{\sqrt{2}} v_{(l,j)}(x), & \text{if } l < j \\ \frac{1}{\sqrt{2}} v_{(j,l)}(x), & \text{if } l > j \end{cases}$$

对于 AttSymSpd-GI，注意力模块中全连接层的输出 y 计算如下

$$y = \exp([-v_1(x) \ldots - v_m(x)])K,$$

其中 $v_j(x) = B_{\xi_j = \delta_j(\infty)}(\ominus p_j \oplus x)\ p_j \in \text{Sym}^+_{m'}, \delta_j(t) = \exp(t a_j)K, j = 1, \ldots, m$。对于参数 $p_j = g_j K$，我们将其建模在对称矩阵空间上，并应用指数映射以获得 SPD矩阵（L´opez 等，2021年）。

映射 $H: G \to \mathfrak{a}$ 按如下方式计算。设 $g \in G$ 和 $g = kan$ 满足 $k \in K$、$a \in A$，且$n \in N$。那么 $g^T g = n^T a^T k^T kan = n^T a^2 n$，这表明可以通过对 $g^T g$进行 LDL分解来确定 $a = \exp(H(g))$ 和 n。

为了计算中点运算的加权弗雷歇均值，我们采用对数欧几里得框架。设 $\{x_j, w_j\}_{j=1}^{L}$ 是一组点 $x_j \in \text{Sym}^+_m$，其对应的权重为 $w_j \in \mathbb{R}$，其中 $w_j > 0$ 且 $\sum_{j=1}^{L} w_j = 1$。那么这些点的加权弗雷歇均值为

$$\text{wFM}(\{x_j, w_j\}_{j=1}^{L}) = \exp\left(\sum_{j=1}^{L} w_j \log(x_j)\right).$$

超参数 为注意力块创建 SPD 矩阵序列时，在 BCIC-IV-2a、MAMEM-SSVEP-II 和 BCI-NER 数据集上，不重叠子序列的数量分别设为 4、6 和 4。卷积层数设为 2。第一和第二个卷积层的输出通道数分别设为 20 和 15。注意力块中全连接层输出 SPD 矩阵的大小在 BCIC-IV-2a、MAMEM-SSVEP-II 和 BCI-NER 数据集上分别设为 6×6、4×4 和 4×4。BCIC-IV-2a、MAMEM-SSVEP-II 和 BCI-NER 数据集的训练轮数分别设为 400、100 和 100。BCIC-IV-2a、MAMEM-SSVEP-II 和 BCI-NER 数据集的批次大小分别设为 128、64 和 64（Pan 等，2022）。学习率和权重衰减分别设为 $1e-3$ 和 $1e-1$。

[2]https://www.kaggle.com/c/inria-bci-challenge.

Ta表 5：AttSymSpd-GI 和 AttSymSpd-LE 的参数数量（MB）比较 .

数据集	BCIC-IV-2a	MAMEM-SSVEP-II	BCI-NER
AttSymSpd-LE	0.034	0.024	0.034
AttSymSpd-GI	0.007	0.013	0.022

T表6：所提出的全连接层和注意力模块在 BCIC-IV-2a 数据集上的有效性。

方法	CovNet	AttSymSpd-GI-Bimap	AttSymSpd-GI
	73.04±6.34	75.82±5.1	78.08±4.8

优化与评估 所有模型均使用 TensorFlow 实现。我们使用交叉熵损失和 Adam（Kingma 和 Ba，2015）训练网络。评估方案基于 Mane 等人（2021）、Pan 等人（2022）和 Wei 等人（2019）。对于 BCIC-IV-2a 数据集，将某个受试者第 1 次实验的数据用作训练集，其中的 1/8 用作验证集。同一受试者第 2 次实验的数据用作测试集。对于 MAMEM-SSVEP-II（BCI-NER）数据集，将某个受试者的前 4 次实验用作训练集，其中的 1/4 用作验证集。同一受试者的第 5 次实验用作测试集。在所有实验中，使用验证集上损失最低的模型进行测试。BCIC-IV-2a 和 MAMEM-SSVEP-II 数据集的结果根据每位受试者 10 次运行所得的准确率计算，而 BCI-NER 数据集的结果则基于 AUC 分数。每个模型的结果均取 10 次运行的平均值。所有实验均使用 Quadro RTX 8000 GPU。

B.2.3 更多结果

表 5 报告了 AttSymSpd-GI 和 AttSymSpd-LE 的可学习参数数量。结果清楚地表明，AttSymSpd-GI 使用的参数远少于 AttSymSpd-LE。值得注意的是，这两个网络在 BCIC-IV-2a 数据集上的准确率相近，但 AttSymSpd-GI 比 AttSymSpd-LE 少约 $5\times$ 个参数。

我们还研究了所提出的全连接层和注意力模块对脑电信号分类的有效性。为此，我们评估了 AttSymSpd-GI 在以下两种情况下的性能：

- 从网络中移除注意力模块。由此得到的网络称为 Cov-Net。这使我们能够验证注意力模块的作用。
- 注意力模块中的全连接层被 Bimap 层（Huang & Gool，2017）取代。Bimap 层也称为全连接卷积类层，可谓 SPD 神经网络中最常用的全连接层类比结构。由此得到的网络称为 AttSymSpd-GI-Bimap。

表 6 报告了我们的实验结果。可以看出，这两个构建模块都很有效。特别是，使用注意力模块使平均准确率提高了 5% 以上；在平均准确率方面，基于我们提出的全连接层的网络比基于 Bimap 层的网络高出 2% 以上。

最后，表 7 展示了表 3 中的所有结果，以及一些先进方法的其他结果。可以看出，AttSymSpd-LE 在所有数据集上的表现均优于先进方法，而 AttSymSpd-GI 在 BCIC-IV-2a 和 MAMEM-SSVEP-II 数据集上的表现优于这些方法。

C 图像生成

在本节中，我们使用 CIFAR-10 和 CIFAR-100 数据集进行图像生成实验。我们基于 HCNN Lorentz（Bdeir 等人，2024）设计了一种新的双曲变分自编码器（VAE），其中将洛伦兹模型中的双曲包裹正态分布替换为庞加莱球中的双曲包裹正态分布，并将洛伦兹 MLR 替换为庞加莱 MLR。与我们的图像分类实验一样，我们在庞加莱 MLR 中使用三种不同的点到超平面距离。

表 7：我们的网络和先进方法在脑电图信号分类中的准确率 n.

方法	BCIC-IV-2a	MAMEM-SSVEP-II	BCI-NER
ShallowNet（Schirrmeister 等，2017）	61.84±6.39	56.93±6.97	71.86±2.64
EEGNet（Lawhern 等，2018）	57.43±6.25	53.72±7.23	74.28±2.47
SCCNet（Wei 等人，2019年）	71.95±5.05	62.11±7.70	70.93±2.31
EEG-TCNet（Ingolfsson 等人，2020年）	67.09±4.6	55.45±7.6	77.05±2.4
TCNet-Fusion（Musallam等人，2021年）	56.52±3.0	45.00±6.4	70.46±2.9
FBCNet（Mane 等人，2021年）	71.45±4.4	53.09±5.6	60.47±3.0
MBEEGSE（Altuwaijri 等人，2022年）	64.58±6.0	56.45±7.2	75.46±2.3
MAtt（Pan 等，2022年）	74.71±5.0	65.50±8.2	76.01±2.2
Graph-CSPNet（Ju 与 Guan，2023）	71.95±13.3	-	-
AttSymSpd-LE（我们的方法）	**78.24** ± 5.4	**70.96** ± 8.6	**78.02** ± 2.3
AttSymSpd-GI（我们的方法）	78.08 ± 4.8	67.24 ± 7.4	75.88 ± 2.2

表 8：用于图像生成的网络架构。PROJ$_{\mathbb{R}^m \to \mathbb{L}_m}$ 层将 $\mathbb{R}^m$中的数据映射到 $\mathbb{L}_m$。H-PROJ $_{\mathbb{L}_m \to \mathbb{B}_m}$ 和 H-PROJ$_{\mathbb{B}_m \to \mathbb{L}_m}$ 层在 $\mathbb{L}_m$ 和 $\mathbb{B}_m$ 之间映射数据（见正文）。CONV 和 CONVTR 层分别是卷积层和转置卷积层的洛伦兹类比层。BN 和 RELU 层分别是批归一化层和 ReLU 层的洛伦兹类比层。FC-MEAN、FC-VAR 和 FC 层分别是欧几里得FC层的洛伦兹类比层。SAMPLE 层根据网络的潜在分布在 $\mathbb{B}_m$ 中生成随机样本。MLR 层是庞加莱 MLR。卷积层和转置卷积层的卷积核大小分别为 3×3 和 4×4。 s和 p 分别表示步幅和零填充。

层	CIFAR-10 / CIFAR-100
编码器:	
$\to$ PROJ$_{\mathbb{R}^m \to \mathbb{L}_m}$	$8 \times 8 \times 3$
$\to$ 卷积 $_{65,\ s2,\ p1}$ $\to$ 批归一化 $\to$ ReLU	$4 \times 4 \times 65$
$\to$ 展平	1025
$\to$ 全连接均值$_{129}$	129
$\to$ 全连接方差$_{129}$ $\to$ Softplus	129
解码器:	
$\to$ H-PROJ $_{\mathbb{L}_m \to \mathbb{B}_m}$	128
$\to$ 采样 ($\mathbb{B}_m$)	128
$\to$ H-PROJ$_{\mathbb{B}_m \to \mathbb{L}_m}$	129
$\to$ FC$_{257}$ $\to$批归一化 $\to$ ReLU	257
$\to$ 重塑	$2 \times 2 \times 65$
$\to$ 转置卷积$_{33,\ s2,\ p1}$ $\to$ 批归一化 $\to$ ReLU	$4 \times 4 \times 33$
$\to$转置卷积$_{17,s2,p1}$ $\to$批归一化 $\to$ReLU	$8 \times 8 \times 17$
$\to$ 卷积 $_{65,\ s1,\ p1}$	$8 \times 8 \times 65$
$\to$ H-PROJ$_{\mathbb{L}_m \to \mathbb{B}_m}$	$8 \times 8 \times 64$
$\to$ MLR ($\mathbb{B}_m$)	$8 \times 8 \times 3$

C.1 洛伦兹模型

$\mathbb{L}_m$ 维 m双曲几何的洛伦兹模型定义为流形 $\mathbb{L}_m = \{x = [x_0, \ldots, x_m]^T \in \mathbb{R}^{m+1}, x_0 > 0 : -x_0^2 + \sum_{i=1}^m x_i^2 = -1\}$配备黎曼度量 $\langle ., . \rangle_x = \text{diag}(-1, \ldots, 1)$。两点 $x = [x_0, \ldots, x_m]^T$, $y = [y_0, \ldots, y_m]^T \in \mathbb{L}_m$之间的黎曼距离由 $d_{\mathbb{L}}(x, y) = \cosh^{-1}\left(x_0 y_0 - \sum_{i=1}^m x_i y_i\right)$ 给出。

C.2 网络架构

网络架构如表 8 所示。H-PROJ$_{\mathbb{L}_m \to \mathbb{B}_m}$ 层通过如下微分同胚将洛伦兹模型映射到庞加莱球：

$$\tau(x_0, x_1, \ldots, x_m) = \frac{(x_1, \ldots, x_m)}{x_0 + 1}.$$

H-PROJ$_{\mathbb{B}_m \to \mathbb{L}_m}$ 层通过如下微分同胚将庞加莱球映射到洛伦兹模型：

$$\tau^{-1}(x_1, \ldots, x_m) = \frac{(1 + \|x\|^2, 2x_1, \ldots, 2x_m)}{1 - \|x\|^2}.$$

下面简要介绍其他层。详情请参阅 Bdeir 等人（2024）。

洛伦兹全连接层

$$y = \mathrm{LFC}(x) = \begin{bmatrix} \sqrt{\|\rho(wx + b)\|^2 + 1} \\ \rho(wx + b) \end{bmatrix},$$

其中，x，y 分别表示该层的输入和输出，$w \in \mathbb{R}^{m' \times (m+1)}$，$b \in \mathbb{R}^{m'}$ 和 ρ分别表示偏置和激活函数。

洛伦兹卷积层 给定一幅图像，将每个图像像素的特征映射到洛伦兹模型。因此，图像可以视为由 m维双曲特征向量构成的有序集合。随后执行洛伦兹卷积：

$$y_{h,w} = \mathrm{LFC}(\mathrm{HCat}(\{x_{h'+s\tilde{h}, w'+s\tilde{w}}\}_{\tilde{h},\tilde{w}=1}^{\tilde{H},\tilde{W}})),$$

其中，$\{x_{h'+s\tilde{h},w'+s\tilde{w}}\}^{H,\ W}_{\tilde{h},\ \tilde{w}=1}$ 表示卷积核感受野内的特征，HCat(。)表示双曲向量的拼接，(h′，w′)表示起始位置，s 为步幅参数。

洛伦兹转置卷积层 转置卷积层通过交换卷积层的前向传播和反向传播来实现。在洛伦兹模型中，这是通过在特征之间进行原点填充实现的。

洛伦兹批归一化 给定一个包含 m个特征的批次 $\mathcal{B}$ x_i，传统的批归一化算法可描述为

$$\mathrm{BN}(x_i) = u \odot \frac{x_i - \mathrm{mean}(\mathcal{B})}{\sqrt{\mathrm{var}(\mathcal{B}) + \epsilon}} + v,$$

其中，mean($\mathcal{B}$) $= \frac{1}{m}\sum_{i=1}^{m} x_i$，var($\mathcal{B}$) $= \frac{1}{m}\sum_{i=1}^{m}(x_i - \mathrm{mean}(\mathcal{B}))^2$，$u$ 和 v 是用于重新缩放和重新居中特征的参数。

对于洛伦兹批归一化层，使用洛伦兹质心和平行移动操作进行重新居中，并使用 Fr´echet 方差以及原点切空间中的直测地线进行重新缩放。

洛伦兹 ReLU

$$y = \begin{bmatrix} \sqrt{\|\,\mathrm{ReLU}([x_1, \ldots, x_m])\|^2 + 1} \\ \mathrm{ReLU}([x_1, \ldots, x_m]) \end{bmatrix},$$

其中，$x = [x_0, \ldots, x_m]$ 和 y 分别是该层的输入和输出。

包裹正态分布 SAMPLE 层采用 Nagano 等人（2019）的方法，在 $\mathbb{B}_m$ 上生成随机样本。给定一个由双曲均值向量 $h \in \mathbb{B}_m$ 和欧几里得方差矩阵 $\Sigma \in \mathbb{R}^{m \times m}$ 参数化的正态分布，该层执行以下运算：

表 9：双曲变分自编码器的重建和生成 FID（越低越好）。

方法	CIFAR-10		CIFAR-100	
	重建 FID	生成 FID	重建 FID	生成 FID
Lorentz-Poincar´e-H（Fan 等，2023）	125.53±5.94	69.11±1.67	110.36±11.50	62.32±6.34
Lorentz-Poincar´e-G（Ganea 等人，2018b）	39.68±1.45	49.91±2.06	42.82±2.48	60.24±4.01
洛伦兹-庞加莱-B（我们的方法）	**38.32**±2.11	**48.45**±1.31	**42.05**±2.58	**59.76**±1.81

1. 从正态分布 $\mathcal{N}(0,\ \Sigma)$ 中采样一个欧几里得向量 $\tilde{v}$。

2. 计算 $v =$ $\tilde{v}$。

3. 将 v 从原点的切空间 0 平行移动到双曲均值的切空间 h，以获得切向量 $u \in T_h\mathbb{B}_m$，如下所示：

$$u = \mathcal{T}_{\mathbf{0}\to h}(v) = (1 - \|h\|^2)v.$$

4. 通过应用指数映射将 u 映射 $\mathbb{B}_m$，如下所示：

$$z = \exp_h(u) = h \oplus_M \left(\tanh\left(\frac{\|u\|}{1-\|h\|^2}\right)\frac{u}{\|u\|}\right),$$

其中，$\oplus_M$ 是 M¨obius 加法（见附录 L.7.1），z 是 $\mathbb{B}_m$中的最终样本。

C.3 实验设置

超参数 我们采用 Bdeir 等人（2024）的超参数。洛伦兹模型和庞加莱球的曲率均设为 1。学习率和权重衰减分别设为 $5e-4$和 0。批量大小和训练轮数均设为 100。KL 损失权重设为 0。024。

优化与评估所有模型均使用 PyTorch 实现。我们使用 Geoopt 库（Kochurov等，2020）进行黎曼优化。使用 RiemannianAdam 训练网络。我们使用重建 FID 和生成 FID 评估网络。重建 FID 通过比较测试图像与重建的验证图像计算得出。训练集中固定随机抽取的 10K 张图像用作验证集（Bdeir 等人，2024）。生成 FID 通过从潜在分布中生成随机图像并与测试集进行比较计算得出。每种模型的结果均为 5 次运行的平均值。所有实验均使用 Quadro RTX 8000 GPU。

C.4 结果

结果见表 9。我们的方法在所有情况下的平均重建 FID 和平均生成 FID 方面均表现最佳。我们还可以观察到，在这些实验中，h-距离的表现显著逊于 g-距离和 b-距离。

D 自然语言推理

在本节中，我们通过在文本蕴含和噪声前缀检测任务上进行与 Ganea 等人（2018b）相同的实验，将我们在庞加莱球中构造点到超平面距离的方法与 Ganea 等人（2018b）和 Fan 等人（2023）的方法进行比较。对于第一个任务，需要预测一个句子能否从另一个句子推断出来。第二个任务则是判断一个句子是否为另一个句子的噪声前缀。我们分别在 SNLI（Bowman等人，2015）和 PREFIX 数据集（Ganea 等人，2018b）上进行实验，以完成第一个和第二个任务。我们的实现[3] 基于 HypGRU 的开源实现[4]

[3]https://github.com/sohata24/nli。[4] https://github.com/dalab/hyperbolic_神经网络。

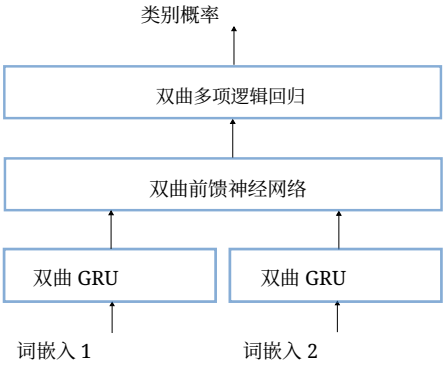


图3：用于自然语言推理的网络架构。

（Ganea 等人，2018b），该实现使用庞加莱 MLR 作为分类层。各个竞争网络的区别仅在于庞加莱 MLR 中点到超平面距离的计算方式（见表 1）。

D.1 数据集

SNLI（Bowman等人，2015）该数据集包含57万对训练句对、1万对验证句对和1万对测试句对。与 Ganea 等人（2018b）的做法类似，将“矛盾”和“中立”类别合并为一个负句对类别，而“蕴含”类别则构成正句对。

PREFIX（Ganea 等人，2018b）PREFIX-10%、PREFIX-30% 和 PREFIX-50% 是合成数据集，每个数据集包含50万对训练句对、1万对验证句对和1万对测试句对。每个数据集按如下方式构建：随机选取一个长度不超过20的第一句，并从中随机选取一个前缀；通过随机替换该前缀中 Z%（Z 为10、30或50）的词语，生成第二个正句；再随机生成一个长度相同的第二个负句。词汇表大小为100。

D.2 实验设置

网络架构 我们采用图 3 所示的全双曲 GRU（Ganea 等人，2018b）架构。该网络由一个双曲 GRU、一个双曲前馈神经网络（前馈神经网络）以及 Poincar´e MLR 组成（见附录 B.1.2）。双曲 GRU 的更新方程如下：

$$
\begin{aligned}
r_t &= \sigma\big(\log_{\mathbf{0}}(W^r \otimes_M h_{t-1} \oplus_M U^r \otimes_M x_t \oplus_M b^r)\big),\\
z_t &= \sigma\big(\log_{\mathbf{0}}(W^z \otimes_M h_{t-1} \oplus_M U^z \otimes_M x_t \oplus_M b^z)\big),\\
\tilde{h}_t &= \psi^{\otimes}((W\,\mathrm{diag}(r_t)) \otimes_M h_{t-1} \oplus_M U \otimes_M x_t \oplus_M b),\\
h_t &= h_{t-1} \oplus_M \mathrm{diag}(z_t) \otimes_M (-h_{t-1} \oplus_M \tilde{h}_t).
\end{aligned}
$$

其中，$x_t \in \mathbb{B}_{m_1}$ 是帧 t 的输入，$h_{t-1}, h_t \in \mathbb{B}_{m_2}$ 分别是帧 $t-1$ 和 t 的隐藏状态，$W^r, W^z, W \in \mathbb{R}^{m_2 \times m_2}$, $b^r, b^z, b \in \mathbb{B}_{m_2}$, $U^r, U^z, U \in \mathbb{R}^{m_1 \times m_2}$ 是模型参数，$\oplus_M$ 是 Möbius 加法（见附录 L.7.1），$\otimes_M$ 是 Möbius 矩阵-向量乘法（见附录 L.7.1），σ, ψ 是激活函数，$\log_{\mathbf{0}}(.)$ 是 $\mathbf{0}$ 处的对数映射（见附录 L.7.1），函数 $\psi^{\otimes}(.)$ 定义为

$$
\psi^{\otimes}(x) = \exp_{\mathbf{0}}(\psi(\log_{\mathbf{0}}(x))),
$$

其中，$x \in \mathbb{B}_m$ 且 $\exp_{\mathbf{0}}(.)$ 是在 $\mathbf{0}$ 处的指数映射（见附录 L.7.1）。

设 $x_1, x_2 \in \mathbb{B}_{m_2}$ 分别为对应第一句和第二句的双曲 GRU 输出。则双曲前馈神经网络的输出 z 计算如下：

$$
z = \tau(W_1 \otimes_M x_1 \oplus_M W_2 \otimes_M x_2 \oplus_M b_1 \oplus_M d_{\mathbb{B}}(x_1, x_2) \otimes_M b_2),
$$

表10：竞争网络在自然语言推理任务中的准确率。

方法	SNLI	PREFIX-10%	PREFIX-30%	PREFIX-50%
HypGRU（Ganea 等人，2018b）	80.89±0.17	96.75±0.40	87.59±0.46	**76.45**±0.61
HypGRU-H（Fan 等，2023）	80.66±0.46	92.20±8.33	83.29±6.34	74.67±3.12
HypGRU-B（我们的方法）	**81.01**±0.35	**97.03**±0.11	**87.69**±0.04	76.25±0.07

其中，$W_1, W_2 \in \mathbb{R}^{m_2 \times m_3}$, $b_1, b_2 \in \mathbb{B}_{m_3}$ 为模型参数，$\otimes_M$ 为 M¨obius 标量乘法（参见附录 L.7.1），且 τ 为激活函数。

超参数 词嵌入和隐藏状态嵌入的维度，以及双曲前馈神经网络的输出通道数均设为 5。训练轮数和批量大小分别设为 30 和 64。词嵌入和双曲权重的学习率分别设为 $1e-1$ 和 $1e-2$。双曲 GRU 和双曲前馈神经网络中的所有激活函数均设为恒等函数。

优化与评估 我们使用交叉熵作为损失函数。欧几里得参数和双曲参数分别使用 Adam（学习率设为 $1e-3$）和黎曼随机梯度下降（RSGD）（Bonnabel，2013；Ganea 等人，2018a）进行优化。每个模型的结果均取 5 次运行的平均值。所有实验均使用 Quadro RTX 8000 GPU。

D.3 结果

竞争网络的结果见表 10。在所有数据集上，我们的网络在平均准确率和标准差方面均优于 HypGRU-H。结果还表明，我们提出的距离有望提升现有双曲神经网络在所考虑任务上的性能。

E 节点分类

本节中，我们通过进行节点分类实验，将我们在 Poincar´e 球中构造点到超平面距离的方法，与 Ganea 等人（2018b）和 Fan 等人（2023）的方法进行比较。

E.1 数据集

Disease（Chami 等人，2019年）这是通过模拟 SIR 疾病传播模型创建的数据集的传导式变体，其中节点标签表示该节点是否受到感染，节点特征表示其对该疾病的易感性。

Airport 数据集（Chami 等人，2019年）这是一个来自 OpenFlights.org 的航班网络数据集，其中节点代表机场，边代表航空公司航线，节点标签则表示机场所属国家的人口。

Pubmed Namata 等人（2012年）这是一个用于描述引文网络的标准基准数据集，其中节点代表医学领域的科学论文，边表示论文之间的引用关系，节点标签则表示学术（子）领域。

Cora Sen 等人（2008年）这是一个引文网络，其中节点代表机器学习领域的科学论文，边表示论文之间的引用关系，节点标签则表示学术（子）领域。该数据集中每篇出版物都由一个取值为 0/1 的词向量表示，用于指示相应的词是否出现在词典中。

四个数据集的统计信息汇总于表11。按照 Chami 等人（2019）的做法，Disease数据集采用 30/10/60 的划分比例，Airport数据集采用 70/15/15 的划分比例；Pubmed 和 Cora 数据集则采用标准划分（Kipf & Welling，2017），每个类别使用 20 个训练样本。

数据集	#节点	#边	#类别	#特征
Disease 数据集	1044	1043	2	1000
Airport 数据集	3188	18631	4	4
Pubmed	19717	44338	3	500
Cora	2708	5429	7	1433

表1 1：节点分类实验的数据集描述。

数据集 双曲性 δ	Disease 数据集 $\delta = 0$	Airport 数据集 $\delta = 1$	Pubmed $\delta = 3.5$	Cora $\delta = 11$
HGCN-H	85.67 ± 2.58	69.82 ± 2.08	75.26 ± 1.82	77.09 ± 2.02
HGCN-G	88.98 ± 1.96	84.78 ± 1.48	**76.02 ± 1.09**	77.47 ± 1.15
HGCN-B（我们的方法）	**89.05 ± 0.78**	**85.04 ± 0.97**	75.89 ± 0.78	**77.90 ± 1.00**

表12：HGCN 模型的节点分类结果。HGCN-H、HGCN-G 和 HGCN-B 分别基于 h-距离、g-距离和 b-距离构建。较低的双曲度值 δ 表示双曲性更强。

E.2 实验设置

网络架构 我们采用 Chami 等人（2019）中的 HGCN架构[5]，该架构由一个图卷积网络（GCN）和一个作为分类最终层的 MLR层组成。GCN 和 MLR层均构建于庞加莱球上。

超参数 我们严格遵循 Chami 等人（2019）中的实验设置。训练轮数设为 5000。我们根据验证集性能进行早停，耐心值为 100 个训练轮次。所有实验的学习率均设为 $1e-2$。Diease 和 Airport 数据集的权重衰减设为 0。Pubmed 和 Cora 数据集的权重衰减分别设为$1e-4$和 $1e-3$。维度数设为 16。GCN 中的层数设为 3。

优化与评估 该网络使用 PyTorch 实现，并采用交叉熵损失和 Adam 优化器进行训练。在最终测试集上，对 10 次随机参数初始化的结果取平均。

E.3 结果

表12展示了我们的实验结果。可以观察到，在大多数情况下，b-距离的表现最佳，这与我们在自然语言推理实验中得到的结果相似。还可以看到，在双曲性较弱的数据集（双曲度值较高）上，即 Pubmed 和 Cora 数据集，h-距离的表现几乎可以媲美 g-距离和 h-距离。然而，在双曲性较强的数据集上，即 Disease 数据集和 Airport 数据集，h-距离的表现明显不如 g-距离和 b-距离。对于这些数据集，b-距离表现最佳，证明了它在层次结构更强的数据集上的有效性。此外，在所有情况下，b-距离的标准差都最低，这表明它在所考虑的应用中能够提供稳定的结果。

F 复杂度分析

我们分析基于对称正定流形上 G-不变度量的两种网络构建模块的复杂度。设 d_{in} 和 d_{out} 分别为全连接层输入和输出矩阵的维度。设 l 为注意力模块的输入矩阵数量。

[5]https://github.com/nguyenxuanson10/symspaces-nc。

• 全连接层：其内存复杂度为 $O(d_{in}^2 d_{out})$，时间复杂度为 $O(d_{in}^3 d_{out})$。
• 注意力模块：其内存复杂度为 $O(3d_{in}^2 d_{out})$。全连接层的时间复杂度为 $O(3l d_{in}^3 d_{out})$。“黎曼距离”和“中点运算”模块的时间复杂度均为 $O(l^2 d_{out}^3)$。

G 我们推导的点到超平面距离与已有工作的联系

当 $\delta(t) = \exp(ta)K$ 时，公式（2）给出的距离与复合距离（见第3.2节）有直接联系。也就是说，前者是通过泛函 $a \in \mathfrak{a}$ 作用于后者得到的，而后者是一个向量值距离。这类似于如何构造定义在 X 上的函数的 Helgason-傅里叶变换。具体而言，对于定义在 X 上的任意函数 f，其 Helgason-傅里叶变换定义为

$$\tilde{f}(\lambda, \xi) = \int_X f(x) \exp((-i\lambda + \varrho) A(x, \xi)) dx,$$

其中 λ，$\varrho \in \mathfrak{a}^*$、$\xi \in \partial X$ 和 $A(x, \xi) \in \mathfrak{a}$ 表示从原点 o 到经过点 $x \in X$ 且法向量为 ξ 的 horocycle 的复合距离。这里，指数项 $(-i\lambda + \varrho)\langle x, \xi \rangle_{\mathbb{H}}$ 被视为泛函 $-i\lambda + \varrho \in \mathfrak{a}^*$ 作用于向量值距离 $A(x, \xi)$ 的结果（Helgason，1994；Sonoda 等人，2022年）。

公式（2）给出的距离也与双曲空间上的随机特征密切相关（Yu & Sa，2023）。这些特征由映射 $\mathrm{HyLa}(.) : \mathbb{B}_m \to \mathbb{R}$ 生成，其定义如下：

$$\mathrm{HyLa}_{\lambda,b,\xi}(x) = \exp\left(\frac{m-1}{2} A(x, \xi)\right) \cos(\lambda A(x, \xi) + b),$$

其中 $x \in \mathbb{B}_m$、$\xi \in \partial \mathbb{B}_m$ 和 $\lambda, b \in \mathbb{R}$。从适当的分布中生成 m' 个元组 (λ, b, ξ) 的随机样本，可得到特征映射 $\omega : \mathbb{B}_m \to \mathbb{R}^{m'}$，该映射近似于双曲空间 $\mathbb{B}_m$上的等距不变核。在高秩对称空间的情形下，项 $\lambda A(x, \xi)$ 将被替换[6] 为一个向量与一个向量值距离之间的欧几里得内积，从而得到与公式（2）中距离类似的表达式。

H 我们的全连接层与现有层的比较

在 Huang & Gool（2017）和 Huang 等人（2018）中，作者引入了 Bimap 和 FRMap 层，并将它们称为全连接卷积类层。对于这两种层，输出矩阵的每个元素都是输入矩阵各元素的线性组合，而我们的全连接层并非如此。FRMap 层与我们的全连接层在输出方面也有所不同，因为前者不会在所考虑的流形上生成点。

Chakraborty 等（2020）采用加权弗雷歇均值（wFM）来构建黎曼卷积层。这些层无法通过简单地将全连接层视为卷积核大小覆盖全部输入的卷积层特例，轻易扩展为欧几里得全连接层的自然推广。这是因为由此得到的全连接层会将所考虑流形上的一组点作为输入，因此与式（3）中的线性变换没有明显联系。

用于双曲空间（Wang，2021）和对称空间（Sonoda 等人，2022）上神经网络的全连接层包含激活函数，而我们的全连接层不使用这些激活函数。此外，Sonoda 等人（2022）和 Wang（2021）的全连接层不会输出所考虑空间上的点。

Nguyen 等（2024）和 Shimizu 等人（2021）的全连接层并非基于布塞曼函数构建。尽管 Nguyen 等（2024）提出的方法面向矩阵流形，其中一些流形（例如格拉斯曼流形）不在我们的方法适用范围内，但前者依赖某些几何量的可微形式（例如对数映射和平行移动），而后者无需这些形式。

另一类研究（Cohen 等，2018；Weiler 等，2021）发展了黎曼神经网络，它们是函数 $f\colon X \to \mathbb{R}^m$，其中 m是输出通道的数量。这些研究的背景与我们的不同，因为我们的目标是构建函数 $f\colon X \to X$形式的神经网络。

[6]还需要进行其他调整，但这超出了本文的讨论范围。

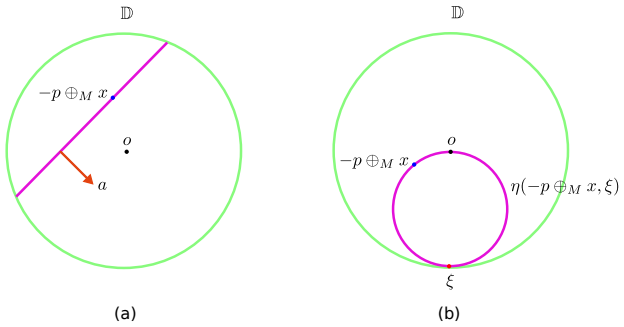


图4：Poincaré圆盘模型中 Poincaré 超平面与我们提出的超平面的比较。

.

I 仿射不变度量与 PEM 度量的优势和局限性

我们使用 G-不变度量的动机在于，这类度量在所考虑的空间上始终存在（例如，由基灵型诱导的度量（Helgason，1979））。因此，第4.4节提出的框架适用于所有这些空间。由于与G-不变度量相关的布塞曼函数由 G 的极大阿贝尔 A 上的布塞曼函数决定（参见命题 4.9 的证明；回顾一下，$G = KAN$），而后者的维度远小于 X，因此，与 PEM度量相比，G-不变度量的一大优势是，它能让全连接层更好地扩展到高维输入和输出矩阵。不过，出于同样的原因，与 G-不变度量相关的布塞曼函数无法充分捕捉 G的结构，这可能导致性能不佳。

采用 PEM 度量的动机在于，与对数欧几里得度量一样，它能将对称正定流形变为平坦空间（即其截面曲率处处为零），从而大幅简化计算与分析（例如，在对数域中，黎曼计算可转化为欧几里得计算（Arsigny 等人，2005））。PEM 度量也比对数欧几里得度量更为一般。但其缺点是无法实现完全的仿射不变性（Arsigny 等人，2005；Lin，2019）。这可能会破坏几何深度学习架构中发挥关键作用的几何稳定性原理（Bronstein等人，2017）。

J b-距离与 g-距离的比较

b-距离与 g-距离之间的差异，最适合通过 Poincar´e 圆盘模型来解释 $\mathbb{D}$（参见附录 L.1）。Poincar´e 超平面（Ganea 等人，2018b）定义为

$$\mathcal{H}_{a,p} = \{x \in \mathbb{D} : \langle \log_p(x), a \rangle_p = 0\},$$

其中 $p \in \mathbb{D}$、$a \in T_p\mathbb{D} \setminus \{\mathbf{0}\}$ 和 $\langle ., . \rangle_p$ 是 Poincar´e 圆盘模型的黎曼度量。已有研究（Ganea 等人，2018b）表明，超平面$\mathcal{H}_{a,\ p}$ 也可以描述为

$$\mathcal{H}_{a,p} = \{x \in \mathbb{D} : \langle -p \oplus_M x, a \rangle = 0\},$$

如图 4(a) 中的线段（紫色）所示。

由于我们使用 M¨obius 加法 $\oplus_M$ 来定义二元运算 $\oplus$，我们的方法中的超平面由以下条件刻画

$$\mathcal{H}_{\xi,p} = \{x \in \mathbb{D} : B_\xi(-p \oplus_M x) = 0\},$$

其中 $p \in \mathbb{D}$ 和 $\xi \in \partial\mathbb{D}$。可以将 $B_\xi(-p \oplus_M x)$ 理解为原点 o 与 horocycle $\eta(-p \oplus_M x,\ \xi)$ 之间的有符号距离；该 horocycle 是经过 $-p \oplus_M x$ 且法向量为 ξ 的唯一 horocycle（Helgason，1979）。这意味着 o 必定位于 horocycle $\eta(-p \oplus_M x,\ \xi)$ 上。因此，点 $-p \oplus_M x$ 必定位于经过原点 o 且法向量为 ξ 的（唯一）horocycle（紫色；见图 4(b)）上。可以看出，我们的超平面与 Poincaré 超平面的刻画方式不同，从而导致点到超平面距离的公式也不同。

K 我们方法的局限性

在我们的工作中，我们针对所有非紧型高秩对称空间推导了点与超平面之间的距离，并且还为这些空间上的神经网络设计了所提出的全连接层。然而，注意力模块依赖于加权弗雷歇均值的计算，而加权弗雷歇均值在一般情况下没有闭式解。为解决这一问题，可以考虑使用 Karcher 算法（Karcher，1977）；事实证明，该算法在 SPD 神经网络的批归一化层实现中非常有效（Brooks 等人，2019）。由于其迭代性质，该算法在实践中可能会带来较高的计算开销。为缓解这一挑战，Chakraborty 等（2020）提出了一种高效的加权弗雷歇均值估计器，值得进一步研究。还可以考虑使用 Lou 等人（2020）提出的方法。特别是，其中依赖指数映射重参数化（Lezcano-Casado，2019）的方法可以为我们的问题提供一种有效的解决方案。

要获得点到超平面距离的闭式表达，必须推导出所考虑的空间及其黎曼度量上的布塞曼函数的闭式表达，这并非总是易事。
我们的全连接层具有 $O(d_{in}^2 d_{out})$ 的内存复杂度和 $O(d_{in}^3 d_{out})$ 的时间复杂度。尽管我们的方法显著降低了 Nguyen 等（2024）提出的全连接层的复杂度，但对于高维全连接层输入和输出矩阵而言，扩展性仍然不佳。我们的注意力块由这些全连接层构成，因此也存在同样的问题。

最后，由于我们的方法是针对非紧型对称空间开发的，因此无法应用于紧型对称空间，例如在许多机器学习应用中也会遇到的格拉斯曼流形。

L 定义与基本事实

L.1 双曲空间作为非紧型对称空间

这里我们从对称空间的角度描述二维双曲空间的庞加莱圆盘模型。记 $\mathbb{D} = \{x \in \mathbb{C} \text{ 为 } \mathbb{C} \text{ 中的开单位圆盘 } \|x\| < 1\}$，其上赋予黎曼度量

$$\langle u, v\rangle_x = \frac{\langle u, v\rangle}{(1-\|x\|^2)^2},$$

其中 $u,\ v \in T_x\mathbb{D}$ 是 $x \in \mathbb{D}$ 处的切向量。令 G 为如下定义的群:

$$G = SU(1,1) := \left\{ \begin{bmatrix} a & b \\ \bar{b} & \bar{a} \end{bmatrix} : a, b \in \mathbb{C}, \|a\|^2 - \|b\|^2 = 1 \right\}.$$

令 SO_m 为行列式为 1 的正交矩阵群。于是 $\mathbb{D}$ 可被视为 as

$$\mathbb{D} \simeq SU(1,1)/SO_2.$$

伊瓦萨瓦分解 $G = KAN$ 中的子群 K、A 和 N，以及 K 中 A 的中心化子 M，分别为

$$\begin{aligned}
K &= \left\{ \begin{bmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{bmatrix} : \theta \in [0, 2\pi) \right\}, \\
A &= \left\{ \begin{bmatrix} \cosh t & \sinh t \\ \sinh t & \cosh t \end{bmatrix} : t \in \mathbb{R} \right\}, \\
N &= \left\{ \begin{bmatrix} 1+is & -is \\ is & 1-is \end{bmatrix} : s \in \mathbb{R} \right\}, \\
M &= \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \right\}.
\end{aligned}$$

群 G 通过由下式定义的莫比乌斯作用，以等距变换的方式作用于 $\mathbb{D}$

$$g[x] := \frac{ax+b}{\bar{b}x+\bar{a}}.$$

此映射是保角的，并将圆和直线映射为圆和直线。

L.2 拉回欧几里得度量

在 PEM 下，黎曼运算定义如下:

$$\exp_x(u) = \phi^{-1}(\phi(x) + D_x\phi(u)),$$
$$\log_x(y) = D_{\phi(x)}\phi^{-1}(\phi(y) - \phi(x)),$$
$$\mathcal{T}_{x\to y}(u) = D_{\phi(y)}\phi^{-1} \circ D_x\phi(u),$$

其中，$\exp_x(.)$、$\log_x(.)$ 和 $\mathcal{T}_{x\to y}(.)$分别是指数映射、对数映射和平行传输。

L.3 非紧型对称空间上的对称正定流形

对称正定流形 Sym_m^+ 是维度为 $m(m+1)/2$ 的可微流形。该流形上点 $x \in \text{Sym}_m^+$ 处的切空间$T_x\text{Sym}_m^+$ 与 Sym_m 同构。黎曼度量定义如下:

$$\langle u, v\rangle_x = \text{Tr}(x^{-1}ux^{-1}v),$$

其中 $u, v \in T_x\,\text{Sym}_m^+$。该度量具有 G-不变性。

设 GL_m 为 $m \times m$ 可逆矩阵群， O_m 为 $m \times m$ 正交矩阵群。则 Sym_m^+ 可表示为

$$\text{Sym}_m^+ \simeq GL_m/O_m.$$

设 $G = GL_m$。那么，伊瓦萨瓦分解 $G = KAN$ 中的子群 K、 A 和 N 定义如下:

- $K = O_m$。
- A 是由对角元素为正的 $m \times m$ 对角矩阵构成的子群。
- N 是由对角元素均为 1 的 $m \times m$ 上三角矩阵构成的子群。

任意 $g \in G$ 都可以唯一地写成 $g = kan$，其中三元组 $(k,\ a,\ n) \in K \times A \times N$，且将 $(k,\ a,\ n)$ 映射到 kan 的映射$K \times A \times N \to G$是一个微分同胚。 A在 K 中的中心化子 M为 $M := C_K(A) := \{k \in K | ka = ak$ 对所有 $a \in A\}$均成立，它是对角元素为 ±1的对角矩阵集合。 G 对 Sym_m^+ 的（传递）作用定义为 $g[x] := gxg^T$，对任意 $g \in G$ 和$x \in \text{Sym}_m^+$均成立。

L.4 非紧型对称空间

紧型对称空间与非紧型对称空间通过嘉当对偶相互转换。它们中的每一类还可进一步分为两个类别。

紧型对称空间具有非负截面曲率。紧型对称空间分为以下两个类别:

- 由对合定义的紧李群的齐性空间（类别 1）。
- 具有双不变度量的紧李群（类别 2）。

非紧型对称空间具有非正截面曲率，并且与欧几里得空间微分同胚。非紧型对称空间分为以下两个类别:

- 非紧致非复李群的齐性空间，取其极大紧子群（第 3 类，与第 1 类对偶）。
- 复李群关于实形式的齐性空间（第 4 类，与第 2 类对偶）。

非紧型对称空间与具有平凡中心且无紧因子的半单李群之间存在对应关系：对于任意具有平凡中心且无

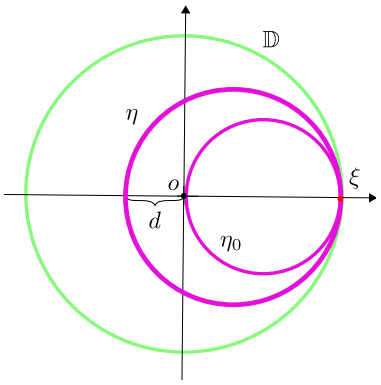


图 5：Poincar´e 圆盘模型 $\mathbb{D}$的边界 $\partial\mathbb{D}$ 由绿色圆圈表示。边界点 ξ 同时垂直于 horocycle η 和基本 horocycle η_0。原点 o到 horocycle ξ 的距离为 $-d$。

紧因子的李群，若取其极大紧子群 K，则商空间 $X = G/K$ 配备 G-不变黎曼度量后构成一个非紧型对称空间。
非紧型对称空间的一个重要不变量是其秩。非紧型对称空间 X 的秩是其中 X 平坦子空间的最大维度（ X 中的平坦子空间是与欧几里得空间等距的子空间）。非紧型对称空间的秩可以为一，也可以更高。

- 非紧型秩一对称空间：实、复和四元数双曲空间，以及 Cayley 数上的双曲平面。
- 非紧型高秩对称空间：所有秩大于一的非紧型对称空间。典型例子包括对称正定流形。

非紧型秩一对称空间与高秩对称空间的许多几何性质各不相同。因此，要将为双曲空间设计的方法推广到高秩对称空间的情形，可能颇具挑战。例如，在 Poincar´e 模型中，由测地线投影得出的点到超平面的距离（Chami 等人，2021年）可以用闭式计算（Ganea 等人，2018b）。然而，对于对称正定流形，情况并非如此（特定黎曼度量下的情形除外，例如对数欧几里得度量）（Nguyen & Yang，2023年）。
In t接下来，我们给出几个例子，以说明第3.2节回顾的一些概念。

几何边界 下面给出一些非紧型对称空间的几何边界示例：

(1) 对于第L.1节所述二维双曲空间的庞加莱圆盘模型，其几何边界 $\partial\mathbb{D}$ 是单位圆 $\partial\mathbb{D} = \{x \in \mathbb{C} : \|x\| = 1\}$（见图 5）。
(2) 考虑第3.1节讨论的 $\mathbb{B}_m$ 维双曲几何的庞加莱模型 $\mathbb{B}_m$。在 $\mathbb{B}_m$ 中，两条测地射线 δ 和 δ' 渐近，当且仅当 $\delta(t)$ 和 $\delta'(t)$ 在 $t \to \infty$ 趋于无穷时收敛到 $\mathbb{R}^m$ 的同一个点。因此，$\mathbb{B}_m$ 的几何边界 $\partial\mathbb{B}_m$ 可自然地与以 $\mathbf{0} \in \mathbb{R}^m$ 为中心、欧几里得半径为1的球面相识别。

（3）对于 Sym_m^+，其几何边界 $\partial\mathrm{Sym}_m^+$ 由左陪集组成

$$\partial\,\mathrm{Sym}_m^+ := K/M := \{\xi = kM | k \in K\}.$$

(4) 如果 X 是一個完備的 m 維非正截面曲率黎曼流形，那麼∂X 與 $\mathbb{S}^{m-1}$ 同胚，即與 ($m-1$) 維球面同胚。這是因為，對於給定的基點 $x \in X$，存在一個同胚，將每個在 x 處與 X 相切的單位向量 u，對應到從 x 出發、速度向量為 u 的測地射線 δ 所屬的等價類。

等距曲線 $\mathbb{D}$ 和 Sym_m^+ 的 horocycle 可描述如下。

(1) $\mathbb{D}$ 中的测地线是与边界 $\partial\mathbb{D}$垂直的圆弧。所有在同一点 $\partial\mathbb{D}$ 处垂直的圆弧都可以看作平行直线。因此，horocycle 的自然定义是与 $\partial\mathbb{D}$ 相切的圆（见图 5）。

(2) 令 I_m 为 $m \times m$ 单位矩阵。那么，任意 horocycle η 都可以写成 $\eta = kaN[I_m]$，其中 $k \in K$ 和 $a \in A$为某些值。特别地，它是这样一组矩阵：在关于 $\mathbb{R}^m$-基$\{ke_i\}_{i=1,\dots,m}$的 UDU 分解中，其对角矩阵为 a^2；其中$\{e_i\}_{i=1,\dots,m}$ 是 $\mathbb{R}^m$ 的标准基（Bartolucci 等人，2021年）。

复合距离 对于 Sym_m^+，从原点 o 到经过点 $x \in \mathrm{Sym}_m^+$ 且法向量为 ξ 的 horocycle 的复合距离 $A(x, \xi)$ 由 Sonoda 等人（2022年）给出：

$$A(x = gK, \xi = kM) = \frac{1}{2}\log\gamma(k^T[x]),$$

其中映射 $\gamma : \mathrm{Sym}_m^+ \to A$ 由 $y = v\gamma(y)v^T$ 确定，其中 $y \in \mathrm{Sym}_m^+, v \in N$。

L.5 角

定义 *L.1.* 对于 X 中的三个点 (p, q, r)，其在 E^2 中的比较三角形是欧几里得平面上的一个三角形，其顶点为 $\bar{p}, \bar{q}, \bar{r}$，满足 $d(p, q) = d(\bar{p}, \bar{q})$ $d(q, r) = d(\bar{q}, \bar{r})$，且 $d(p, r) = d(\bar{p}, \bar{r})$。这样的三角形在等距意义下唯一，记作 $\Delta(p, q, r)$。$\Delta(p, q, r)$在 p 处的内角称为 p处 q与 r 之间的比较角，记作 $\angle_p(q, r)$。只要 q 和 r 均不同于 p，比较角便有良好定义。

定义 *L.2.* 设 $\delta : [0, a] \to X$ 和 $\delta' : [0, a'] \to X$ 是两条测地路径，且 $\delta(0) = \delta'(0)$。给定 $t \in (0, a]$ 和 $t' \in (0, \underline{a}']$，考虑比较三角形 $\Delta(\delta(0)\ \delta(t)\ \delta'(t'))$，以及比较角 $\angle_{\delta(0)}(\delta(t)\ \delta'(t'))$。测地路径 δ 和 δ' 之间的（亚历山德罗夫）角或上角是数值 $\angle_{\delta,\delta'} \in [0, \pi]$，定义如下：

$$\angle_{\delta,\delta'} := \lim_{t,t'\to 0} \sup \overline{\angle}_{\delta(0)}(\delta(t), \delta'(t')) = \lim_{\varepsilon\to 0} \sup_{0<t,t'<\varepsilon} \overline{\angle}_{\delta(0)}(\delta(t), \delta'(t')).$$

L.6 布塞曼函数

欧几里得空间 (Bridson & H¨afliger, 2011) 设 $\delta(t) = tu$ 是 $\mathbb{R}^m$ 中的一条射线，其中 u 是单位向量。那么布塞曼函数 $B_{\xi=\delta(\infty)}(.)$ 定义为

$$B_\xi(x) = -\langle x, u\rangle.$$

对称正定流形在对数欧几里得框架下（Bonet 等人，2023）令 $\delta(t) = \exp(ta)$ 为 Sym_m^+ 中的一条测地线，其中 $a \in \mathrm{Sym}_m$ 且 $\|a\| = 1$。则布塞曼函数 $B_{\xi=\delta(\infty)}(。)$为

$$B_\xi(x) = -\operatorname{Tr}(a\log(x)).$$

L.7 运算

L.7.1 庞加莱´模型与 M ¨obius 陀螺向量空间

在 m 维双曲几何的庞加莱´模型 $\mathbb{B}_m$ 中，对数映射 $\log_{\mathbf{0}}(。)$ 和指数映射 $\exp_{\mathbf{0}}(。)$分别为

$$\log_{\mathbf{0}}(x) = \tanh^{-1}(\|x\|)\frac{x}{\|x\|}, \quad \exp_{\mathbf{0}}(v) = \tanh(\|v\|)\frac{v}{\|v\|},$$

其中 $x \in \mathbb{B}_m \setminus \{\mathbf{0}\}$ 且 $v \in T_{\mathbf{0}}\mathbb{B}_m \setminus \{\mathbf{0}\}$。

M¨obius 加法 $\oplus_M$ 定义为

$$x \oplus_M y = \frac{(1 + 2\langle x, y\rangle + \|y\|^2)x + (1 - \|x\|^2)y}{1 + 2\langle x, y\rangle + \|x\|^2\|y\|^2},$$

其中 $x, y \in \mathbb{B}_m$。M¨obius 减法 $\ominus_M$ 定义为

$$\ominus_M x = -x.$$

M¨obius 标量乘法 $\otimes_M$ 定义为

$$r \otimes_M x = \tanh(r \tanh^{-1}(\|x\|)) \frac{x}{\|x\|},$$

其中 $r \in \mathbb{R}$ 且 $x \in \mathbb{B}_m \setminus \{\mathbf{0}\}$。

M¨obius 矩阵-向量乘法 $\otimes_M$ 定义为

$$M \otimes_M x = \tanh\left(\frac{\|xM\|}{\|x\|} \tanh^{-1}(\|x\|)\right) \frac{xM}{\|xM\|},$$

其中 $x \in \mathbb{B}_m$、 $M \in \mathbb{R}^{m \times m'}$，且当 $xM = \mathbf{0}$时 $M \otimes_M x = \mathbf{0}$。请注意，我们对 M¨obius 标量乘法和 M¨obius 矩阵-向量乘法使用了与 Ganea 等人（2018b）相同的记号$\otimes_M$。

L.7.2 陀螺向量空间

对于 $x, y \in \text{Sym}_m^+$，二元运算 $\oplus_{le}$ 和逆运算 $\ominus_{le}$ 定义如下（Nguyen，2022a;b）

$$x \oplus_{le} y = \exp(\log(x) + \log(y)), \quad \ominus_{le} x = x^{-1},$$

其中 exp(.) 表示矩阵指数[7]。

SPD内积定义为

$$\langle x, y \rangle^{le} = \langle \log(x), \log(y) \rangle.$$

M 数学证明

M.1 命题 4.4 的证明

证明。 我们首先将 Chen 等人（2024a）（引理3.5）中的一个结果改写为以下命题的形式。

命题 M.1。 设 $\phi : \text{Sym}_m^+ \to \text{Sym}_m$ 为一个微分同胚， $p \in \text{Sym}_m^+$, $a' \in T_p \text{Sym}_m^+ \setminus \{\mathbf{0}\}$，并设 $\mathcal{H}^{pb}_{a',p}$ 为如下定义的超平面

$$\mathcal{H}^{pb}_{a',p} = \{x \in \text{Sym}_m^+ : \langle \text{Log}_p(x), a' \rangle_p^\phi = 0\},$$

其中， $\langle ,. \rangle_p^\phi$ 是定义对称正定流形时给出的点 p 处的 *PEM*。于是，点 $x \in \text{Sym}_m^+$ 与超平面 $\mathcal{H}^{pb}_{a',p}$ 之间的距离$d^{pb}(x, \mathcal{H}^{pb}_{a',p})$ 为

$$d^{pb}(x, \mathcal{H}^{pb}_{a',p}) = \frac{|\langle \phi(x) - \phi(p), D_p\phi(a') \rangle|}{\|a'\|_p^\phi},$$

其中， $\|.\|_p^\phi$是由黎曼内积 $\langle ,. \rangle_p^\phi$诱导的范数。

根据三角不等式,

$$d(\delta(0), \delta(t)) - d(x, \delta(0)) \leq d(x, \delta(t)) \leq d(\delta(0), \delta(t)) + d(x, \delta(0)),$$

由此可得

$$t - d(x, \delta(0)) \leq d(x, \delta(t)) \leq t + d(x, \delta(0)).$$

Thus

$$1 - \frac{d(x, \delta(0))}{2t} \leq \frac{d(x, \delta(t)) + t}{2t} \leq 1 + \frac{d(x, \delta(0))}{2t},$$

[7]至于对数函数.，指数函数.的含义应可从上下文中看出。

由此得到 $\lim_{t\to\infty} \frac{d(x,\delta(t))+t}{2t} = 1$)。因此

$$\begin{aligned}
B_{\xi=\delta(\infty)}(x) &= \lim_{t\to\infty} d(x,\delta(t)) - t \\
&= \lim_{t\to\infty} \big(d(x,\delta(t)) - t\big)\frac{d(x,\delta(t))+t}{2t} \\
&= \lim_{t\to\infty} \frac{1}{2t}\big(d(x,\delta(t))^2 - t^2\big) \\
&= \lim_{t\to\infty} \frac{1}{2t}\big(\|\phi(x)-\phi(\delta(t))\|^2 - t^2\big) \\
&= \lim_{t\to\infty} \frac{1}{2t}\big(\|\phi(x)\|^2 + \|\phi(\delta(t))\|^2 - 2\langle\phi(\delta(t)),\phi(x)\rangle - t^2\big) \\
&= \lim_{t\to\infty} \frac{1}{2t}\big(\|\phi(x)\|^2 + \|ta\|^2 - 2\langle ta,\phi(x)\rangle - t^2\big) \\
&= \lim_{t\to\infty} \frac{1}{2t}\big(\|\phi(x)\|^2 - 2t\langle a,\phi(x)\rangle\big) \\
&= \lim_{t\to\infty} \frac{1}{2t}\big(\|\phi(x)\|^2\big) - \langle a,\phi(x)\rangle \\
&= -\langle a,\phi(x)\rangle.
\end{aligned}$$

根据二元运算 $\oplus$和逆运算 $\ominus$的定义，我们有

$$\ominus p \oplus x = \phi^{-1}(\phi(x) - \phi(p)).$$

因此

$$\| \ominus p \oplus x\|_{\mathbb{S}} = \|\phi(\ominus p \oplus x)\| = \|\phi(x) - \phi(p)\|.$$

于是得到

$$\begin{aligned}
\bar{d}(x,\mathcal{H}_{\xi,p}) &= d(x,p).\frac{B_\xi(\ominus p\oplus x)}{\|\ominus p\oplus x\|_{\mathbb{S}}} \\
&= -d(x,p).\frac{\langle a,\phi(\ominus p\oplus x)\rangle}{\|\phi(x)-\phi(p)\|} \\
&= -\langle a,\phi(\phi^{-1}(\phi(x)-\phi(p)))\rangle \\
&= -\langle a,\phi(x)-\phi(p)\rangle.
\end{aligned} \tag{9}$$

由命题 M.1,

$$d^{pb}(x,\mathcal{H}^{pb}_{a',p}) = \frac{|\langle\phi(x)-\phi(p),D_p\phi(a')\rangle|}{\|a'\|^\phi_p} = \left|\langle\phi(x)-\phi(p),\frac{D_p\phi(a')}{\|a'\|^\phi_p}\rangle\right|,$$

根据拉回度量的性质,

$$\langle a_1,a_2\rangle^\phi_p = \langle D_p\phi(a_1),D_p\phi(a_2)\rangle,$$

其中 $a_1,\ a_2 \in T_p\,\mathrm{Sym}^+_m$。我们推得

$$\left\|\frac{D_p\phi(a')}{\|a'\|^\phi_p}\right\| = 1.$$

It c可以看出，无符号距离 $|\bar{d}(x,\ \mathcal{H}_{\xi,\ p})|$ 与 $d^{pb}(x,\ \mathcal{H}^{pb}_{a',\ p})$ 的形式完全相同 .

□

M.2 命题 4.8 的证明

证明。 为证明 (i)，我们需要使用 Kassel (2009) 中的一个结果。

引理 M.2。令 $\rho:\ X \to \mathfrak{a}^+$ 为将 $x = g[o] \in X$ 映射到 $\mu(g)$ 的映射，其中 $g \in G$。对于所有x、$x' \in X$,

$$\|\rho(x) - \rho(x')\| \leq d(x, x').$$

此外，若 $x, x' \in \exp(\mathfrak{a}^+)[o]$，则 $d(x, x') = \|\rho(x) - \rho(x')\|$。

令 $x = gK, y = hK$，其中 $g, h \in G$。由于距离 $d(.,.)$ 是 G-不变的，我们有

$$\begin{aligned} d(x, y) &= d(g^{-1}[x], g^{-1}[y]) \\ &= d(o, g^{-1}h[o]). \end{aligned}$$

令 $g^{-1}h = kak'$，其中 $a \in \exp(\overline{\mathfrak{a}^+})$, $k, k' \in K$。则

$$\begin{aligned} d(o, g^{-1}h[o]) &= d(k^{-1}[o], k^{-1}kak'[o]) \\ &= d(o, a[o]) \end{aligned}$$

由引理 M.2,

$$\begin{aligned} d(o, a[o]) &= \|\rho(o) - \rho(a[o])\| \\ &= \|\mu(a)\| \\ &= \|\mu(g^{-1}h)\|. \end{aligned}$$

注意到

$$\begin{aligned} \| \ominus x \oplus y\|_{\mathbb{S}} &= \|g^{-1}hK\|_{\mathbb{S}} \\ &= \sqrt{\langle g^{-1}hK, g^{-1}hK\rangle_{\mathbb{S}}} \\ &= \sqrt{\langle \mu(g^{-1}h), \mu(g^{-1}h)\rangle} \\ &= \|\mu(g^{-1}h)\|. \end{aligned}$$

因此

$$\| \ominus x \oplus y\|_{\mathbb{S}} = d(x, y).$$

为证明 (ii)，注意对任意 $k \in K$，我们有 $k[x] = kgK = kk_1a_1n_1K = k_2a_1n_1K$，其中$g = k_1a_1n_1$、$k_2 = kk_1$、$k_1 \in K$、$a_1 \in A$、$n_1 \in N$。因此，$\mu(kg) = \mu(g)$。同理可得$\mu(kh) = \mu(h)$。因此

$$\begin{aligned} \langle k[x], k[y]\rangle_{\mathbb{S}} &= \langle \mu(kg), \mu(kh)\rangle \\ &= \langle \mu(g), \mu(h)\rangle \\ &= \langle x, y\rangle_{\mathbb{S}}. \end{aligned}$$

□

M.3 命题 4.9 的证明

证明。我们首先将 Bridson 与 H¨afliger（2011）的一项结果（引理 10.26）改写为以下引理。

引理 M.3。设 X 是一个非紧型对称空间，G 是一个通过等距变换作用于 X 上的群。假设 $h \in G$ 保持一条测地线 $\delta:\ \mathbb{R} \to X$ 不变，并且$h[\delta(t)] = \delta(t + c)$，其中 $c > 0$。设 $x_0 = \delta(0)$，并令 $N \subset G$ 为满足当 $j \to \infty$ 时 $h^{-j}gh^j[x_0] \to x_0$ 的元素 $g \in G$ 的集合。则 N 固定 $\delta(\infty) \in \partial X$，并保持与 δ 相关联的布塞曼函数不变。

令 $\delta(t) = k\exp(ta)K, h = \exp(a) \in G$，其中 $a \in \mathfrak{a}, \|a\| = 1, k \in K$。设 $\delta'(t) = k^{-1}\delta(t)$。则

$$\begin{aligned} h[\delta'(t)] &= \exp(a + ta)K \\ &= k^{-1}\delta(t + 1) \\ &= \delta'(t + 1). \end{aligned}$$

由于距离 $d(\cdot,\cdot)$ 是 G-不变的，对于任意 $x \in X$，我们有

$$d(x, \delta(t)) = d(k^{-1}[x], \delta'(t)).$$

令 $g \in G$ 满足 $k^{-1}[x] = gK$，并令 $g = n_1 \exp A(g) k_1$，其中 $n_1 \in N$、$k_1 \in K$[8]。对任意 $n \in N$、$h^{-j} n h^j[o] \to o$，当 $j \to \infty$ 时。由引理 M.3（$c = 1$），可推出 $B_{\xi'=\delta'(\infty)}(k^{-1}[x]) = B_{\xi'=\delta'(\infty)}(n_1^{-1}k^{-1}[x])$。因此

$$d(k^{-1}[x], \delta'(t))^2 = d(n_1^{-1}k^{-1}[x], \delta'(t))^2.$$

由此得到以下等式链

$$\begin{aligned} d(k^{-1}[x], \delta'(t))^2 &= d(n_1 \exp A(g) K, \delta'(t))^2 \\ &= d(\exp A(g) K, \delta'(t))^2 \\ &= \langle A(g) - ta, A(g) - ta \rangle \\ &= \langle A(g), A(g) \rangle - 2t\langle a, A(g) \rangle + t^2. \end{aligned} \tag{10}$$

根据三角不等式,

$$d(\delta'(0), \delta'(t)) - d(k^{-1}[x], \delta'(0)) \le d(k^{-1}[x], \delta'(t)) \le d(\delta'(0), \delta'(t)) + d(k^{-1}[x], \delta'(0)),$$

由此可得

$$t - d(k^{-1}[x], \delta'(0)) \le d(k^{-1}[x], \delta'(t)) \le t + d(k^{-1}[x], \delta'(0)).$$

Thus

$$1 - \frac{d(k^{-1}[x], \delta'(0))}{2t} \le \frac{d(k^{-1}[x], \delta'(t)) + t}{2t} \le 1 + \frac{d(k^{-1}[x], \delta'(0))}{2t},$$

这导致 $\lim_{t\to\infty} \frac{d(k^{-1}[x], \delta'(t)) + t}{2t} = 1$。因此

$$\begin{aligned} B_{\xi=\delta(\infty)}(x) &= \lim_{t\to\infty} d(x, \delta(t)) - t \\ &= \lim_{t\to\infty} d(k^{-1}[x], \delta'(t)) - t \\ &= \lim_{t\to\infty} \big(d(k^{-1}[x], \delta'(t)) - t\big) \frac{d(k^{-1}[x], \delta'(t)) + t}{2t} \\ &= \lim_{t\to\infty} \frac{1}{2t} \big(d(k^{-1}[x], \delta'(t))^2 - t^2\big). \end{aligned}$$

利用公式（10），可得

$$\begin{aligned} B_\xi(x) &= \lim_{t\to\infty} \frac{1}{2t} \big(\langle A(g), A(g) \rangle - 2t\langle a, A(g) \rangle\big) \\ &= -\langle a, A(g) \rangle \\ &= \langle a, H(g^{-1}) \rangle, \end{aligned}$$

从而得出命题 4.9。

□

M.4 推论 4.10 的证明

证明。我们有

$$\begin{aligned} \bar{d}(x, \mathcal{H}_{\xi,p}) &= d(x, p). \frac{B_\xi(\ominus p \oplus x)}{\| \ominus p \oplus x \|_{\mathbb{S}}} \\ &= d(x, p). \frac{B_\xi(h^{-1} g K)}{\| \ominus p \oplus x \|_{\mathbb{S}}}. \end{aligned}$$

[8]我们使用与 Helgason（1994）中相同的记号， A表示从原点 o 到 horocycle 的复合距离。

注意， $k^{-1}[h^{-1}gK] = k^{-1}h^{-1}gK$。根据命题 4.8 和 4.9,

$$d(x,p).\frac{B_\xi(h^{-1}gK)}{\| \ominus p \oplus x\|_{\mathbb{S}}} = \langle a, H(g^{-1}hk)\rangle,$$

由此得证推论 4.10。

□

M.5 命题 5.1 的证明

证明。 我们首先将 Bridson 和 H¨afliger（2011）的一项结果（命题 9.8）改写为以下命题的形式。

命题 M.4。 设 X 是一个以 x_0 为基点的非紧型对称空间。设 $\xi, \xi' \in \partial X$，并设 δ, δ' 为测地射线，满足 $\delta(0) = \delta'(0) = x_0$, $\delta(\infty) = \xi$ 且 $\delta'(\infty) = \xi'$。那么

$$2\sin(\angle(\xi,\xi')/2) = \lim_{t\to\infty}\frac{1}{t}d(\delta(t),\delta'(t)).$$

对于任意 $t \in [0,\infty)$，有

$$\begin{aligned} d(\delta(t),\delta'(t)) &= d(\exp(ta)K, \exp(ta')K) \\ &= \|t(a-a')\| \\ &= \sqrt{2}t. \end{aligned}$$

由命题 M.4,

$$\begin{aligned} 2\sin(\angle(\xi,\xi')/2) &= \lim_{t\to\infty}\frac{1}{t}d(\delta(t),\delta'(t)) \\ &= \sqrt{2}. \end{aligned}$$

因此我们推得 $\angle(\xi, \xi') = \frac{\pi}{2}$。– □

M.6 命题 5.2 的证明

证明。设 $(e_j)_{j=1,\ldots,m}$ 是 $\mathbb{R}^m$ 的标准基，并设 $\xi_j = \delta_j(\infty)$，其中 $\delta_j(t) = \exp(te_j)K$，$j = 1, \ldots, m$ 是测地射线。那么对于 $y = gK \in X$ 和任意 $j \in \{1,\ldots,m\}$,

$$v_j(x) = \bar{d}(y, \mathcal{H}_{\tilde{\xi}_j,K}) = \langle e_j, H(g^{-1})\rangle = H(g^{-1})[j],$$

其中 $H(g^{-1})[j]$表示 $H(g^{-1})$的第 j个维度。因此

$$H(g^{-1}) = [v_1(x)\ldots v_m(x)]^T.$$

注意到 $g = n\exp(-H(g^{-1}))k$，其中 $n \in N$ 且 $k \in K$。因此

$$g = n\exp([-v_1(x)\ldots - v_m(x)])k,$$

由此可得

$$y = n\exp([-v_1(x)\ldots - v_m(x)])K.$$

□