

# 双曲空間における乱択化 Helgason-Fourier 特徴法と 双曲ガウス過程潜在変数モデル

渡部航史<sup>†</sup> 持橋大地<sup>††</sup> 小川貴弘<sup>†††</sup> 長谷山美紀<sup>†††</sup>

<sup>†</sup> 北海道大学大学院 情報科学院 〒060-0814 北海道札幌市北区北 14 条西 9 丁目

<sup>††</sup> 統計数理研究所/国立国語研究所 〒190-8562 東京都立川市緑町 10-3

<sup>†††</sup> 北海道大学大学院 情報科学研究院 〒060-0814 北海道札幌市北区北 14 条西 9 丁目

E-mail: <sup>†</sup>{koshi,ogawa,mhaseyama}@lmd.ist.hokudai.jp, <sup>††</sup>daichi@ism.ac.jp

**あらまし** 階層構造を効果的に表現可能な双曲空間における機械学習は、従来の機械学習に代わるアプローチとして注目を集めている。双曲空間上のカーネル法も種々提案されており、少数サンプルからの画像分類等において高精度となることが報告されている一方、双曲空間のカーネル法において問題となる計算量の増加を抑えるアプローチは限られている。そこで、本研究では、代表的なスパースカーネル法である乱択化 Fourier 特徴法を双曲空間に拡張した乱択化 Helgason-Fourier 特徴法を提案する。提案手法では、ユークリッド空間における Fourier 変換を対称空間の調和解析に拡張した Helgason の結果 [1] を用いることで、双曲空間における乱択化特徴法を構成した。実験では、構築した乱択化 Helgason-Fourier 特徴法をグラフのノード分類と潜在変数モデルに適用することで、提案手法の有効性を検証した。

**キーワード** 双曲空間, カーネル法, 乱択化特徴法, ガウス過程

## Hyperbolic Kernel Method and Latent Variable Models with Hyperbolic Random Fourier Features

Koshi WATANABE<sup>†</sup>, Daichi MOCHIHASHI<sup>††</sup>, Takahiro OGAWA<sup>†††</sup>, and Miki HASEYAMA<sup>†††</sup>

<sup>†</sup> Graduate School of Information Science and Technology, Hokkaido University Kita 14, Nishi 9, Kita-ku, Sapporo, Hokkaido, 060-0814, Japan

<sup>††</sup> The Institute of Statistical Mathematics/National Institute for Japanese Language and Linguistics 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan

<sup>†††</sup> Graduate School of Information Science and Technology, Hokkaido University Kita 14, Nishi 9, Kita-ku, Sapporo, Hokkaido, 060-0814, Japan

E-mail: <sup>†</sup>{koshi,ogawa,mhaseyama}@lmd.ist.hokudai.jp, <sup>††</sup>daichi@ism.ac.jp

**Abstract** Hyperbolic machine learning is gaining attention for its ability to represent hierarchical structures effectively. While hyperbolic kernel methods show promise in tasks like few-shot classification, they often suffer from high computational costs. To address this, we propose Randomized Helgason-Fourier Features, extending the standard Randomized Fourier Features to hyperbolic space using Helgason's harmonic analysis of symmetric spaces [1]. We validated the proposed method through experiments on graph node classification and latent variable models, demonstrating its effectiveness in reducing computational complexity while maintaining performance.

**Key words** Hyperbolic space, kernel methods, random features, Gaussian process

### 1. ま え が き

リーマン多様体を用いた機械学習は、従来の機械学習に代わるアプローチとして注目を集めている。特に双曲空間はデータ

が持つ階層構造を効率的に表現でき、単語辞書に代表される階層的なデータの埋め込みにおいて、効率的な表現学習が可能であると報告されている。例えば、文献 [2] における WordNet の埋め込みでは、5次元双曲空間上で学習された埋め込みが、ユー

クリッド空間における 200 次元の埋め込みと同等の精度を達成できると報告されている。近年では、双曲空間におけるカーネル法 [3]~[5] が盛んに研究されており、ユークリッド空間における手法と比較して高精度であることから、さらなる進展が期待されている。

カーネル法において、サンプルサイズに三乗で増加する計算量の削減は中心的な課題の一つである。このボトルネックは  $N \times N$  の大きさをもつグラム行列の逆行列計算のために生じるものであり、これまで様々なスパース手法が提案されてきた。代表的なアプローチである誘導点法 [6]~[8] は、少数の代表点によって逆行列計算を近似する手法である。一方、乱択化 Fourier 特徴法 [9] は誘導点法と異なるスパース手法であり、カーネル関数を有限の基底関数により近似することでガウス過程を線形モデルに帰着させて計算量を削減し、データ量に依存しないモデルを実現する。

文献 [10] では、誘導点法が入力データに基づき基底関数を定める一方で、乱択化 Fourier 特徴法ではデータに依存しない基底関数を定めるような違いがあると報告されている。通常、少数の次元数しか持たない双曲埋め込みの場合、少数の代表点を数値的に不安定な双曲空間上で推定する必要がある誘導点法と比較して、関数自体を近似する乱択化 Fourier 特徴法はより適した方法であることが期待できる。しかしながら、双曲空間における乱択化 Fourier 特徴法の研究は数値的に不安定であることが知られるポアンカレボールモデルを対象とした手法に限られており [11]、また双曲埋め込みの分類や潜在変数モデルへの拡張はされていない。

そこで本研究では、双曲空間における乱択化 Fourier 特徴法として乱択化 Helgason-Fourier 特徴法を提案し、さらに分類とガウス過程に基づく潜在変数モデルへの適用を検討する。具体的には、Fourier 変換を一般のリーマン多様体に発展させた Helgason の結果を用いて、双曲空間上のカーネル関数を有限の基底関数により近似する方法を構築する。基底関数のパラメータはカーネル関数の周波数表現に従いサンプリングすることで、カーネル関数が構成するヒルベルト空間の有限近似として、基底関数を定める。また、潜在変数モデルへの拡張では、得られた基底関数を用いた MCMC による推定を行う。双曲空間において、ギブスサンプリングとメトロポリスヘイスティング法を組み合わせた推定法を構築し、対数尤度の差分で評価することで効率的な潜在変数の推定を実現する。

## 2. 前 提

### 2.1 双曲空間

はじめに、本手法で用いる双曲空間モデルのローレンツモデル [12] について定める。ローレンツモデルは、 $Q$  次元の (上側) 一葉双曲面  $\mathbb{H}^Q$  に計量テンソル  $\mathbf{G}_t = \text{diag}(-1, 1, \dots, 1) \in \mathbb{R}^{(Q+1) \times (Q+1)}$  を備え付けたリーマン多様体:  $\mathcal{L}^Q = (\mathbb{H}^Q, g_t)$  である。 $Q$  次元の一葉双曲面上の変数は、 $(Q+1)$  個の要素を持つベクトル  $\mathbf{x} = [x_t, x_1, x_2, \dots, x_Q]^\top \equiv [x_t, \mathbf{x}_s^\top]^\top$  で、同時に  $x_t = \sqrt{1 + \|\mathbf{x}_s\|_2^2}$  を満たす。以上を用いると、まずローレンツモデルにおける内積は次式のように定められる。

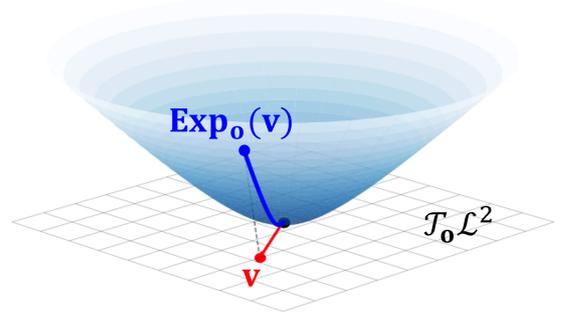


図 1: 2 次元ローレンツモデルの原点  $\mathbf{o} = [1, 0, 0]^\top$  の接空間と指数写像の例。

$$\langle \mathbf{x}, \mathbf{x}' \rangle_{\mathcal{L}^Q} = \mathbf{x}^\top \mathbf{G}_t \mathbf{x}' = -x_t x'_t + \mathbf{x}_s^\top \mathbf{x}'_s \quad (1)$$

以上から、 $\langle \mathbf{x}, \mathbf{x} \rangle_{\mathcal{L}^Q} = -(1 + \|\mathbf{x}_s\|_2^2) + \|\mathbf{x}_s\|_2^2 = -1$  を満たす。

続いて、多様体上の点に定まる接空間を定める。接空間とは、多様体上の点  $\mathbf{x}$  に直交するベクトルの集合:  $\mathcal{T}_x \mathcal{L}^Q = \{\mathbf{p} \in \mathbb{R}^{Q+1} : \langle \mathbf{p}, \mathbf{x} \rangle_{\mathcal{L}^Q} = 0\}$  である。指数写像  $\text{Exp}_x(\mathbf{v})$  とは、点  $\mathbf{x}$  を始点として、初速度ベクトル  $\mathbf{v} \in \mathcal{T}_x \mathcal{L}^Q$  で進む測地線  $\gamma(t)$  の  $t = 1$  における位置  $\gamma(1) \in \mathcal{L}^Q$  を表し、次式により与えられる。

$$\text{Exp}_x(\mathbf{v}) = \cosh(\|\mathbf{v}\|_{\mathcal{L}^Q}) \mathbf{x} + \sinh(\|\mathbf{v}\|_{\mathcal{L}^Q}) \frac{\mathbf{v}}{\|\mathbf{v}\|_{\mathcal{L}^Q}} \quad (2)$$

図 1 に原点  $\mathbf{o} = [1, 0, 0]^\top$  における接空間と接空間上のベクトルの指数写像の例を示す。指数写像は、その定義から、接空間上のベクトルを多様体上へ貼り付ける操作として解釈される。ローレンツモデルにおける測地線の距離は次式のように表される。

$$d_{\mathcal{L}^Q}(\mathbf{x}, \mathbf{x}') = \cosh^{-1}(-\langle \mathbf{x}, \mathbf{x}' \rangle_{\mathcal{L}^Q}) \quad (3)$$

ローレンツモデルの変数を可視化するには、 $\mathbf{x} \mapsto \frac{\mathbf{x}_s}{1+x_t}$  の変換によって得られるポアンカレ球上の変数として可視化する。

### 2.2 乱択化 Fourier 特徴法

続いて、ユークリッド空間における乱択化 Fourier 特徴法について説明する。まず、カーネル関数における Bochner の定理 [13] は定常カーネルにおける周波数領域の存在を保証し、それは次式によって表現される。

$$k(\mathbf{x}, \mathbf{x}') = \int_{\mathbb{R}^Q} p(\boldsymbol{\omega}) e^{j\boldsymbol{\omega}^\top (\mathbf{x} - \mathbf{x}')} d\boldsymbol{\omega} \quad (4)$$

乱択化 Fourier 特徴法は、式 (4) を  $H$  個の関数の和として近似する手法であり、次式のように表される。

$$\int_{\mathbb{R}^Q} p(\boldsymbol{\omega}) e^{j\boldsymbol{\omega}^\top (\mathbf{x} - \mathbf{x}')} d\boldsymbol{\omega} \approx \frac{1}{H} \sum_{h=1}^H \zeta_{\boldsymbol{\omega}_h}(\mathbf{x}) \zeta_{\boldsymbol{\omega}_h}^*(\mathbf{x}') \quad (5)$$

ただし、 $\zeta_{\boldsymbol{\omega}}(\mathbf{x}) = e^{j\boldsymbol{\omega}^\top \mathbf{x}} = \cos(\boldsymbol{\omega}^\top \mathbf{x}) + j \sin(\boldsymbol{\omega}^\top \mathbf{x})$ ,  $\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \dots, \boldsymbol{\omega}_H \sim p(\boldsymbol{\omega})$ ,  $p(\boldsymbol{\omega})$  は定常カーネルに一意に定まり、式 (5) の逆変換により算出可能である。ここで、それぞれの積は次式のように表される。

$$\zeta_{\boldsymbol{\omega}}(\mathbf{x}) \zeta_{\boldsymbol{\omega}}^*(\mathbf{x}')$$

$$= \cos \{ \boldsymbol{\omega}^\top (\mathbf{x} - \mathbf{x}') \} + j \sin \{ \boldsymbol{\omega}^\top (\mathbf{x} - \mathbf{x}') \} \quad (6)$$

式 (6) と正定値カーネルで虚部がゼロとなることを踏まえると、以下が成立する。

$$\begin{aligned} \frac{1}{H} \sum_{h=1}^H \zeta_{\boldsymbol{\omega}_h}(\mathbf{x}) \zeta_{\boldsymbol{\omega}_h}^*(\mathbf{x}') &= \boldsymbol{\psi}(\mathbf{x})^\top \boldsymbol{\psi}(\mathbf{x}') \\ \boldsymbol{\psi}(\mathbf{x}) &= \frac{1}{\sqrt{H}} [\cos(\boldsymbol{\omega}_h^\top \mathbf{x}), \sin(\boldsymbol{\omega}_h^\top \mathbf{x})]_{h=1, \dots, H}^\top \end{aligned} \quad (7)$$

このとき、グラム行列は  $\mathbf{\Psi} = [\boldsymbol{\psi}(\mathbf{x}_1), \boldsymbol{\psi}(\mathbf{x}_2), \dots, \boldsymbol{\psi}(\mathbf{x}_N)]^\top$  を用いて  $\mathbf{K} = \mathbf{\Psi} \mathbf{\Psi}^\top$  のように表される。つまり、乱択化特徴法ではカーネル法を  $H$  個の基底関数で表現された線形モデルに近似しており、 $O(N^3)$  から  $O(H^3)$  の時間計算量に削減することが可能となる。

### 3. 乱択化 Helgason-Fourier 特徴法

本研究では、 $Q$  次元ローレンツモデル上の正定値カーネルの乱択化特徴法について検討する。始めに、Helgason による Fourier 変換の対称空間への一般化である Helgason-Fourier 変換を導入し、それを用いた乱択化特徴法について説明する。

#### 3.1 球変換

対称空間上の調和解析では、等方カーネル  $k(\mathbf{x}, \mathbf{x}') = k(d_{\mathcal{L}^Q}(\mathbf{x}, \mathbf{x}')) = k(r)$  に対し定められる、以下の球変換 [1](p.39) を適用することが有効である。

$$\tilde{k}(\lambda) = \int_0^\infty k(r) \phi_\lambda(r) \sinh^{Q-1}(r) dr \quad (8)$$

ただし、 $r = d_{\mathcal{L}^Q}(\mathbf{x}, \mathbf{x}')$ 、 $\lambda \in \mathbb{R}_+$  は周波数、 $\phi_\lambda(r)$  は球関数であり、ガウスの超幾何関数  ${}_2F_1$  を用いて  $\phi_\lambda(r) = {}_2F_1(\rho + i\lambda, \rho - i\lambda; \frac{Q}{2}; -\sinh^2(\frac{r}{2}))$ 、 $\rho = \frac{Q-1}{2}$  と表される<sup>(注1)</sup>。特に  $Q = 3$  のときには簡単に表すことができ、

$$\phi_\lambda(r) = \frac{\sin(\lambda r)}{\lambda \sinh(r)} \quad (9)$$

で与えられる。

続いて、球関数の別の表現により、ユークリッド空間における平面波との関連について検討する。具体的には、Harish-Chandra の積分公式 [1] (Chapter IV Theorem 4.3) を用いると、球関数は次式のように表すことが可能である。

$$\phi_\lambda(r_0) = \int_{\partial \mathcal{L}^Q} e^{(i\lambda + \rho)B(\mathbf{x}, \boldsymbol{\xi})} d\boldsymbol{\xi} \quad (10)$$

ただし、 $r_0 = d_{\mathcal{L}^Q}(\mathbf{x}, \mathbf{o})$  は  $\mathbf{x}$  と原点との距離、 $\boldsymbol{\xi} = [1, \boldsymbol{\xi}^\top]^\top \in \partial \mathcal{L}^Q$ 、 $\boldsymbol{\xi}' \in \mathcal{S}^Q$  であり、 $B(\mathbf{x}, \boldsymbol{\xi})$  は次式の Bussemann 関数である。

$$B(\mathbf{x}, \boldsymbol{\xi}) = \log(x_t - \mathbf{x}_s^\top \boldsymbol{\xi}') \quad (11)$$

ここで以下の不等式から、対数の真数は必ず正の値となる。

$$\begin{aligned} x_t - \mathbf{x}_s^\top \boldsymbol{\xi}' &= \sqrt{1 + \|\mathbf{x}_s\|_2^2} - \mathbf{x}_s^\top \boldsymbol{\xi}' \\ &\geq \sqrt{1 + \|\mathbf{x}_s\|_2^2} - \|\mathbf{x}_s\|_2 \|\boldsymbol{\xi}'\|_2 \\ &= \sqrt{1 + \|\mathbf{x}_s\|_2^2} - \|\mathbf{x}_s\|_2 > 0 \end{aligned}$$

ここで、2 行目の不等式では、コーシー=シュワルツの不等式を適用した。式から、球関数は振動成分  $e^{i\lambda B(\mathbf{x}, \boldsymbol{\xi})}$  と波の位相を決定する係数  $e^{\rho B(\mathbf{x}, \boldsymbol{\xi})}$  を重ね合わせた積分とみなせる。このことから、球関数は距離に依存する振動成分の重みづけ和により表される関数であり、周波数  $\lambda$  を持つ動径方向の固有関数とみなすことができる。

さらに、Planchrel の定理 [1] (Chapter IV Theorem 7.5) は、次式の球変換の逆変換の存在を保証する。

$$k(r) = \int_0^\infty \phi_\lambda(r) \frac{\tilde{k}(\lambda)}{|c(\lambda)|^2} d\lambda \quad (12)$$

ただし、 $c(\lambda)$  は Harish-Chandra の  $c$  関数 [1] (Chapter IV Theorem 6.14) で、

$$|c(\lambda)|^2 = \frac{1}{\lambda \tanh(\pi\lambda)} \quad (Q = 2) \quad (13)$$

$$|c(\lambda)|^2 = \frac{1}{\lambda^2} \quad (Q = 3) \quad (14)$$

と表すことができる。 $|c(\lambda)|^{-2}$  は、双曲空間における体積の指数的な増加を補正する項であり、こうして双曲空間のカーネル関数の周波数表現を得ることが可能となる。

#### 3.2 カーネルの周波数表現

続いて、カーネル関数の周波数表現  $\tilde{k}(\lambda)$  を定める。例えば、周波数領域で滑らかに減衰していくカーネルを定義する際には、以下の熱カーネルが有効である。

$$\tilde{k}_t(\lambda) \propto e^{-t\lambda^2} \quad (15)$$

$t$  は温度ハイパーパラメータである。この熱カーネルの指数的な減衰を抑えることを目的として、以下の Matérn カーネルも広く用いられる。

$$\tilde{k}_{\nu, \kappa}(\lambda) \propto \frac{1}{(\lambda^2 + \rho^2 + \frac{2\nu}{\kappa^2})^{\nu + \frac{Q}{2}}} \quad (16)$$

ここで、 $\nu, \kappa$  はハイパーパラメータであり、有限の  $\nu$  に対しては多項式的な減衰をすることがわかる。

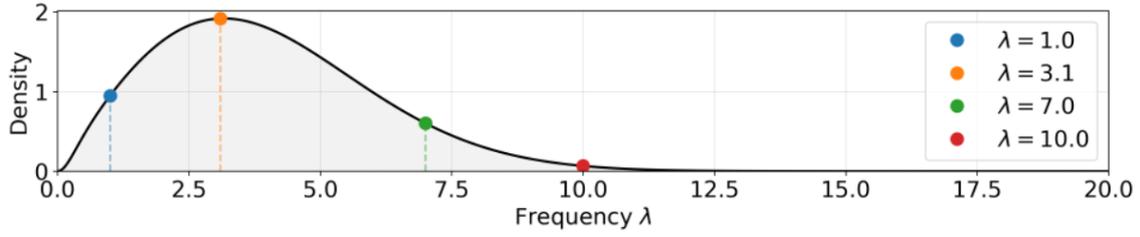
#### 3.3 乱択化による近似

最後に、球変換を用いた乱択化に基づく近似を行う。乱択化 Fourier 特徴法を式 (12) に適用するためには、球関数を変数  $\mathbf{x}, \mathbf{x}'$  それぞれの項に分解する必要がある。そのために、以下の球関数の表現が有効である ([1], Chapter IV Lemma 4.4)。

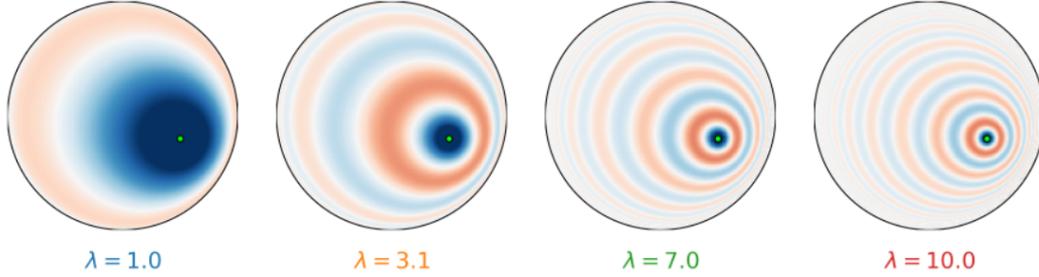
$$\phi_\lambda(r) = \int_{\mathcal{S}^Q} e^{(\rho + j\lambda)B(\mathbf{x}, \boldsymbol{\xi})} e^{(\rho - j\lambda)B(\mathbf{x}', \boldsymbol{\xi})} d\boldsymbol{\xi} \quad (17)$$

ここで、式 (17) は、上述の原点を用いた球関数の表現 (10) を 2 点間の場合に拡張した表現である。これは、前の議論と同様に、球関数が無限遠点  $\partial \mathcal{L}^Q = \mathcal{S}^Q$  から到来する平面波の積分により分解できることを表している。この結果を用いると、球関

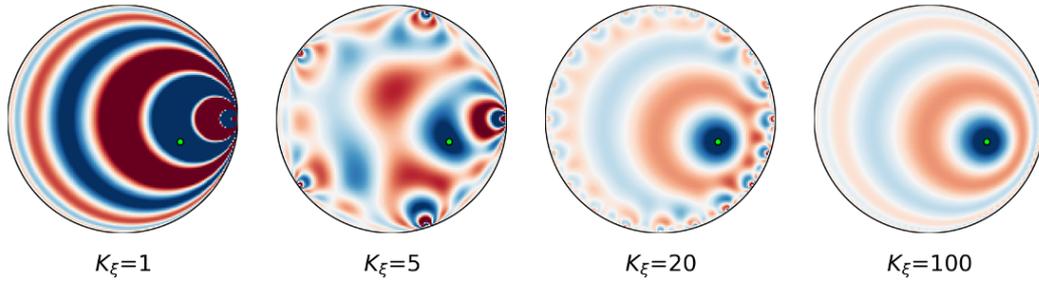
(注1) : ガウスの超幾何関数  ${}_2F_1(x, a; b; c)$  は、常微分方程式  $x(x-1)y'' + (c - (a+b+1)x)y' - aby = 0$  の解である。



(a)  $\frac{\tilde{k}(\lambda)}{|c(\lambda)|^2}$  の分布.



(b) 上図の各周波数に対応する球関数のポアンカレ球モデルを用いた可視化.



(c) 有限の平面波による球関数の近似のポアンカレ球モデルを用いた可視化.

図 2: 乱択化 Helgason-Fourier 特徴法による近似の概要図.

数は次の様に展開される.

$$\begin{aligned} \phi_\lambda(r) &= \int_{S^Q} e^{(\rho+j\lambda)B(\mathbf{x}, \boldsymbol{\xi})} e^{(\rho-j\lambda)B(\mathbf{x}', \boldsymbol{\xi})} d\boldsymbol{\xi} \\ &= \int_{S^Q} \psi_{\mathcal{L}Q}(\mathbf{x}, \boldsymbol{\xi})^\top \psi_{\mathcal{L}Q}(\mathbf{x}', \boldsymbol{\xi}) d\boldsymbol{\xi} \end{aligned} \quad (18)$$

ただし,  $\psi_{\mathcal{L}Q}(\mathbf{x}, \boldsymbol{\xi}) = [e^{\rho B(\mathbf{x}, \boldsymbol{\xi})} \cos \lambda B(\mathbf{x}, \boldsymbol{\xi}), e^{\rho B(\mathbf{x}, \boldsymbol{\xi})} \sin \lambda B(\mathbf{x}, \boldsymbol{\xi})]^\top$  であり, 虚数部を無視することで分解した. これにより, 球関数を変数ごとに分解した表現が得られる.

以上を用いて, 最終的な逆変換公式によるカーネルの表現は次のように表される.

$$k(r) = \int_0^\infty \left\{ \int_{S^Q} \psi_{\mathcal{L}Q}(\mathbf{x}, \boldsymbol{\xi})^\top \psi_{\mathcal{L}Q}(\mathbf{x}', \boldsymbol{\xi}) d\boldsymbol{\xi} \right\} \frac{\tilde{k}(\lambda)}{|c(\lambda)|^2} d\lambda$$

(双曲空間におけるカーネルの Helgason-Fourier 表現)

(19)

これは,  $\boldsymbol{\xi}$  は境界方向に一樣にサンプリングすることで近似可能であり,  $\lambda$  は  $p(\lambda) \propto \tilde{k}(\lambda)|c(\lambda)|^{-2}$  の分布からサンプリングすることで近似可能である.

以上の近似の概要を図 2 に示した. 双曲空間上の任意の等方カーネルは, 図 2 (a) に示すような周波数領域を持つ. 図 2 (b) に示すように, 各周波数領域に対応する球関数は各点からの距離  $r$  に依存する関数であり, 任意の等方カーネルはこれらの

固有関数の重ね合わせにより表現される. また, 球関数は, 図 2 (c) に示すように, 無限遠方より到来する平面波の重ね合わせで表現される.

提案手法では, カーネルの周波数領域の密度関数に従って周波数  $\lambda$  をサンプリングする. その後, 各周波数に対応する球関数をサンプリングされた平面波の重ね合わせにより表現することで, 等方カーネルの内積表現を得ることができる. このような手順により, 双曲空間上の乱択化特徴法が実現可能である.

続いて, サンプリングの手順について説明する. 平面波の到来方向  $\boldsymbol{\xi}$  は, 対象とする空間で等間隔になるようサンプリングを行い, 周波数  $\lambda$  は密度関数に従ってサンプリングする. ここで, 多くの密度関数において正規化定数は自明でないため, 棄却サンプリングに基づいて周波数をサンプリングする<sup>(注2)</sup>. 基底関数の計算時には, 次式のように得られた周波数と到来方向の直積で表される全ての組を用いて計算を行う.

$$\begin{aligned} \psi_{\mathcal{L}Q}(\mathbf{x}_n) &= \frac{1}{\sqrt{2K_\lambda K_\xi}} \left[ e^{\rho B_{nj}} \cos(\lambda_i B_{nj}), e^{\rho B_{nj}} \sin(\lambda_i B_{nj}) \right]_{\substack{i=1, \dots, K_\lambda \\ j=1, \dots, K_\xi}}^\top \end{aligned}$$

(双曲空間における乱択化 Helgason-Fourier 表現)

(20)

(注2):  $Q \geq 4$  では, 標準正規分布に従う変数を正規化してサンプリングする

ただし,  $B_{nj} = B(\mathbf{x}_n, \xi_j)$  として簡略化した. そのため, 基底関数の次元は, 周波数の個数  $K_\lambda$  と平面波の個数  $K_\xi$  の積の2倍となることに注意されたい.

#### 4. 潜在変数モデルへの拡張

本章では, 双曲空間上の乱択化 Helgason-Fourier 特徴法の潜在変数モデルへの拡張について説明する. 本研究では, ガウス過程潜在変数モデルに基づき目的関数を定め, MCMC 法によって潜在変数の推定を行う,

##### 4.1 ガウス過程潜在変数モデル

観測変数を  $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N]^\top \in \mathbb{R}^{N \times D}$  とする. ガウス過程潜在変数モデルでは, 次式で与えられる対数尤度を最大化することで潜在変数を推定する.

$$-\frac{D}{2} \log |\mathbf{K} + \beta^{-1} \mathbf{I}| - \frac{1}{2} \text{tr}((\mathbf{K} + \beta^{-1} \mathbf{I})^{-1} \mathbf{Y} \mathbf{Y}^\top) \quad (21)$$

ただし,  $\beta$  はノイズの精度パラメータであり, 定数部分は無視して表記した. この対数尤度は, ガウス過程により入力となる潜在変数  $\mathbf{X}$  から観測変数が生成されると仮定し, ガウス過程事前分布を周辺化することで得られる. 式に含まれる逆行列計算には  $O(N^3)$  の計算量が要求され, 計算上のボトルネックとなることに注意されたい.

##### 4.2 乱択特徴に基づく対数尤度の近似

そこで, 乱択特徴  $\Psi = [\psi_{\mathcal{L}Q}(\mathbf{x}_1), \psi_{\mathcal{L}Q}(\mathbf{x}_2), \dots, \psi_{\mathcal{L}Q}(\mathbf{x}_N)]^\top \in \mathbb{R}^{N \times 2K_\lambda K_\xi}$  を用いることで, 対数尤度の計算を近似する.  $\mathbf{K} = \Psi \Psi^\top$  が成立することから, 行列式および逆行列には次式の関係式が成立する.

$$|\Psi \Psi^\top + \beta^{-1} \mathbf{I}| = \beta^{-N} |\mathbf{I} + \beta \Psi^\top \Psi| \quad (22)$$

$$(\Psi \Psi^\top + \beta^{-1} \mathbf{I})^{-1} = \beta \mathbf{I} - \beta^2 \Psi (\mathbf{I} + \beta \Psi^\top \Psi)^{-1} \Psi^\top \quad (23)$$

式 (22) および式 (23) 右辺の計算は  $O((2K_\lambda K_\xi)^3)$  で計算可能であり,  $2K_\lambda K_\xi < N$  となるように基底および周波数を選択することで, 計算量の大幅な削減が可能となる.

##### 4.3 潜在変数の更新

続いて, MCMC に基づく潜在変数の更新規則について説明する. 提案手法ではギブスサンプリングに基づき, 各サンプル  $\mathbf{x}_n \in \mathcal{L}^Q$  毎に潜在変数の更新を行う. また, メトロポリスヘイスティング法に基づき各サンプルを更新し, 尤度の差分により棄却と受け入れを評価する.

###### 4.3.1 潜在変数の提案分布からのサンプリング

提案手法では, 双曲空間上のランダムウォークに基づいて潜在変数の提案分布を定める. まず, 標準正規分布に従う乱数  $\mathbf{v} \in \mathbb{R}^{Q+1}$  を生成する. 続いて, 次式に従って得られた乱数を, 潜在変数  $\mathbf{x}_n$  の接空間まで移動する.

$$\mathbf{v}_{\mathcal{T}_x \mathcal{L}^Q} = \mathbf{v} - \langle \mathbf{v}, \mathbf{x} \rangle_{\mathcal{L}^Q} \mathbf{x} \quad (24)$$

ここで, ベクトル  $\mathbf{v}_{\mathcal{T}_x \mathcal{L}^Q}$  は次式を満たす.

$$\langle \mathbf{v}_{\mathcal{T}_x \mathcal{L}^Q}, \mathbf{x} \rangle_{\mathcal{L}^Q} = \langle \mathbf{v}, \mathbf{x} \rangle_{\mathcal{L}^Q} - \langle \mathbf{v}, \mathbf{x} \rangle_{\mathcal{L}^Q} \cdot \langle \mathbf{x}, \mathbf{x} \rangle_{\mathcal{L}^Q} = 0 \quad (25)$$

ここで,  $\langle \mathbf{x}, \mathbf{x} \rangle_{\mathcal{L}^Q} = 1$  を用いた. 以上から,  $\mathbf{v}_{\mathcal{T}_x \mathcal{L}^Q}$  は  $\mathbf{x}$  の接

空間上のベクトルであることが分かる. その後, 次式に従い指数写像に基づいて接ベクトルを多様体上に写すことで, ランダムウォークに従うサンプルを生成する.

$$\hat{\mathbf{x}} = \text{Exp}_{\mathbf{x}}(\mathbf{v}_{\mathcal{T}_x \mathcal{L}^Q}) \quad (26)$$

以上の操作に基づく提案分布からのサンプリングと詳細釣り合い条件に基づき, 棄却と受け入れの評価を行う.

###### 4.3.2 対数尤度の差分評価

得られた新しい状態の尤度を計算するためには, 同様に  $O((2K_\lambda K_\xi)^3)$  が必要であると考えられるが, 差分を評価することで大幅に計算量を削減できる. 以降では,  $n$  番目の潜在変数を更新する場合を考え,  $\mathbf{x}_n$  と提案分布からサンプリングされた  $\hat{\mathbf{x}}_n$  の尤度差の算出方法を説明する.

まず, 現在の状態の基底関数を用いて, サンプリングされた新しい基底関数は次式のように表される.

$$\begin{aligned} \hat{\Psi} &= \Psi + \mathbf{e}_n (\hat{\psi}_n - \psi_n)^\top \\ &\triangleq \Psi + \mathbf{e}_n \mathbf{d}_n^\top \end{aligned} \quad (27)$$

ただし,  $\mathbf{e}_n$  は  $n$  番目の要素が 1 で他が 0 のベクトルを表す. 式 (27) の表現を用いると, 新しい状態のグラム行列は次式により与えられる.

$$\begin{aligned} \hat{\Psi} \hat{\Psi}^\top + \beta^{-1} \mathbf{I} \\ = \mathbf{K} + \mathbf{e}_n \mathbf{d}_n^\top \Psi^\top + \Psi \mathbf{d}_n \mathbf{e}_n^\top + \mathbf{e}_n \mathbf{d}_n^\top \mathbf{d}_n \mathbf{e}_n^\top \end{aligned} \quad (28)$$

ただし,  $\mathbf{K} = \Psi \Psi^\top + \beta^{-1} \mathbf{I}$  である. ここで,  $\mathbf{u}_n = \Psi \mathbf{d}_n$ ,  $\gamma_n = \mathbf{d}_n^\top \mathbf{d}_n$  とおくと, 次のように変形できる.

$$\begin{aligned} \mathbf{K} + \mathbf{e}_n \mathbf{d}_n^\top \Psi^\top + \Psi \mathbf{d}_n \mathbf{e}_n^\top + \mathbf{e}_n \mathbf{d}_n^\top \mathbf{d}_n \mathbf{e}_n^\top \\ = \mathbf{K} + \mathbf{e}_n \mathbf{u}_n^\top + \mathbf{u}_n \mathbf{e}_n^\top + \gamma_n \mathbf{e}_n \mathbf{e}_n^\top \end{aligned} \quad (29)$$

さらに,  $\mathbf{U}_n = [\mathbf{u}_n + \frac{\gamma_n}{2} \mathbf{e}_n, \mathbf{e}_n] \in \mathbb{R}^{N \times 2}$ ,  $\mathbf{V}_n = [\mathbf{e}_n, \mathbf{u}_n + \frac{\gamma_n}{2} \mathbf{e}_n] \in \mathbb{R}^{N \times 2}$  を用いて, 次のように変形する.

$$\begin{aligned} \mathbf{e}_n \mathbf{u}_n^\top + \mathbf{u}_n \mathbf{e}_n^\top + \gamma_n \mathbf{e}_n \mathbf{e}_n^\top \\ = \mathbf{U}_n \mathbf{V}_n^\top \end{aligned} \quad (30)$$

式 (30) を用いると, 新しい状態に対応する基底の行列式と逆行列は次式により算出される.

$$\begin{aligned} \log |\hat{\Psi} \hat{\Psi}^\top + \beta^{-1} \mathbf{I}| \\ = \log |\mathbf{K}| + \log |\mathbf{I} + \mathbf{V}_n^\top \mathbf{K}^{-1} \mathbf{U}_n| \end{aligned} \quad (31)$$

$$\begin{aligned} (\hat{\Psi} \hat{\Psi}^\top + \beta^{-1} \mathbf{I})^{-1} \\ = \mathbf{K}^{-1} - \mathbf{K}^{-1} \mathbf{U}_n (\mathbf{I} + \mathbf{V}_n^\top \mathbf{K}^{-1} \mathbf{U}_n)^{-1} \mathbf{V}_n^\top \mathbf{K}^{-1} \end{aligned} \quad (32)$$

以上から, 対数尤度の差分  $\Delta$  は次式で計算できる.

$$\begin{aligned} \Delta = \log |\mathbf{I} + \mathbf{V}_n^\top \mathbf{K}^{-1} \mathbf{U}_n| \\ - \text{tr}(\mathbf{K}^{-1} \mathbf{U}_n (\mathbf{I} + \mathbf{V}_n^\top \mathbf{K}^{-1} \mathbf{U}_n)^{-1} \mathbf{V}_n^\top \mathbf{K}^{-1}) \end{aligned} \quad (33)$$

式 (33) の対数尤度の差分は, MCMC において更新前の状態の行列  $\mathbf{K} = \Psi \Psi^\top + \beta^{-1} \mathbf{I}$  の逆行列と行列式を計算して保

表 1: グラフの双曲埋め込みの分類結果. 全ての結果は 10 回行った平均の精度であり, 太字は最も高い数値, 下線は二番目に高い数値を表す.

Domain	Ontology						Network				
	WordNet Noun		DOID		MeSH		Cora	CS	Comp.	Photo	WikiCS
Dataset	クラス	深さ	クラス	深さ	クラス	深さ	クラス	クラス	クラス	クラス	クラス
ノード数	82,114	82,110	11,999	11,987	28,824	28,824	2,708	18,333	13,471	7,535	11,367
カテゴリ数	26	17	8	9	16	13	7	15	10	8	10
Linear	0.658	0.275	0.830	0.385	0.622	0.338	0.605	0.739	0.691	0.755	0.580
RFF	0.120	0.204	0.411	0.231	0.267	0.227	<b>0.825</b>	0.277	0.248	0.169	0.161
HyLa	<u>0.871</u>	0.448	0.854	0.427	0.849	0.390	0.795	<b>0.831</b>	0.805	<u>0.849</u>	<u>0.680</u>
RHFF (Matérn)	<b>0.878</b>	<u>0.491</u>	<b>0.917</b>	<b>0.541</b>	<b>0.859</b>	<u>0.429</u>	<u>0.803</u>	<u>0.809</u>	<b>0.808</b>	<b>0.855</b>	<b>0.691</b>
RHFF (Heat)	0.867	<b>0.492</b>	0.889	<u>0.540</u>	0.825	<b>0.431</b>	0.802	0.807	0.765	0.836	0.654

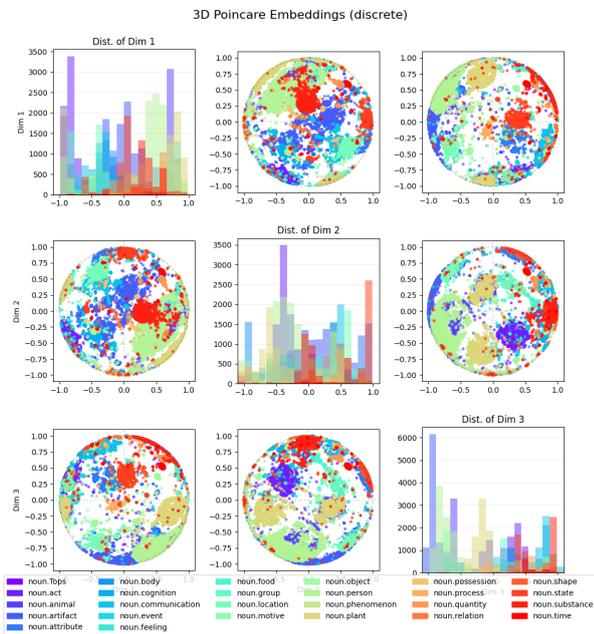


図 3: WordNet Noun における 3 次元のローレンツ埋め込みの例. ローレンツモデルで算出された埋め込みをポアンカレボール上に表示した.

持しておくことで  $O(1)$  で計算可能であり, 全てのサンプルは  $O(N)$  で更新可能である. しかしながら, 実装においては, 差分に応じた更新を繰り返すことで数値誤差が積み重なることによる精度劣化が想定される. そこで提案手法では, 全ての点の更新が終了した後一度だけ, 逆行列と行列式を計算し直す処理を行った.

## 5. 実験

以降では, 提案手法の有効性を検証する実験を行う. 始めに, 提案手法を用いた線形分類の精度を比較することで, 乱択化 Helgason-Fourier 特徴の有効性を検証する. 続いて, 潜在変数

モデルに適用した結果を比較し, 提案手法のさらなる有効性の検証を行う.

### 5.1 学習済み双曲埋め込みの分類

本節では, 学習された双曲埋め込みを入力とする分類の有効性を検証する.

#### a) データセット

まず, オントロジーを有するデータセットとして, WordNet における名詞サブツリー (Noun) [2], [14], 疾患に関するオントロジーである DOID [15], 医学用語のシソーラスである MeSH [16] を用いた. また, 一般的なネットワーク構造を有するデータセットとして, 引用ネットワークである Cora [17] およびコンピュータサイエンス分野における共著者ネットワークである Coauthor CS (CS) [18], Amazon における共購入ネットワークである Amazon Computer および Amazon Photo [18], Wikipedia のコンピュータサイエンスの記事から構成される Web グラフである WikiCS [19] を用いた.

#### b) 実験設定

本実験では, 文献 [20] の手法により各データセットにおける 3 次元の双曲埋め込みを算出し, グラフにおけるカテゴリや深さを分類する実験を行った. 図 3 に, WordNet Noun において推定された 3 次元の双曲埋め込みを示す. 乱択化特徴法は, カーネル法を有限の基底関数により近似する手法である. そこで, 算出された双曲埋め込みを基底関数により変換し, 変換された特徴量を多クラスロジスティック回帰により分類することで検証を行った. また, 全てのデータセットにおいて, 各クラスに少なくとも 10 件のデータが含まれるようにフィルタリングを行なった. オントロジーの分類では, カテゴリとノード埋め込みの階層的な深さの分類の二つの実験を行っており, フィルタリングによりそれぞれのデータ数が異なっていることに注意されたい.

比較手法として, 埋め込みを直接分類する手法 (Linear), 通常の乱択化 Fourier 特徴法 (RFF) [9], 双曲空間を対象として

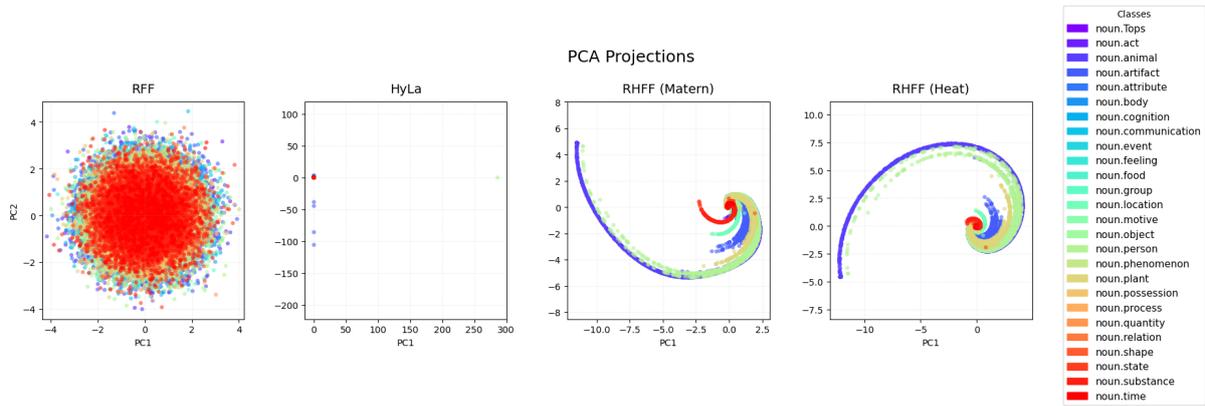
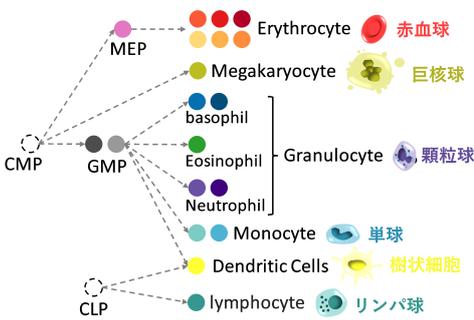
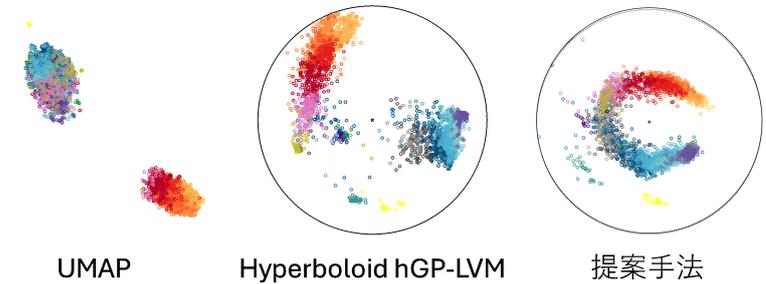


図 4: WordNet Noun の埋め込みにおいて, RFF, HyLa および提案手法で計算された基底関数の主成分の可視化. 埋め込みは右に示すクラス分けに従い色付けした.



(a) 埋め込みの色付けに対応するカラーコード.



(b) 各手法における次元削減の結果. UMAP は二次元平面上で可視化しており, Bayesian GP-LVM および提案手法ではポアンカレモデル上で可視化を行った.

提案された従来の乱択化特徴法 (HyLa) [11] を用いた. HyLa はポアンカレ球モデルで定義された手法であるため, 入力となるローレンツモデル上の変数を一度ポアンカレモデル上の変数に変換してから分類を行った. 提案手法では, 式 (16) で表される Matérn カーネルおよび式 (15) で表される熱カーネルを用いた. 乱択化手法におけるパラメータは, 1,000 次元の基底関数を構成するようにサンプリングされ, 提案手法である乱択化 Helgason-Fourier 特徴法 (RHFF) では  $K_\lambda = 10$ ,  $K_\xi = 50$  と設定した.

表 1 に実験結果を示す. 全ての結果は 10 回行った場合の平均の精度を表す. 全ての結果から, 提案手法である乱択化 Helgason-Fourier 特徴法 (RHFF) が高い精度を達成していることが確認される. 特に, 提案手法は Ontology の場合に一貫して優れた精度を達成しており, 階層構造に適した基底関数を構成可能であることが示唆された.

図 4 に WordNet Noun における基底関数の主成分の可視化を示す. 乱択化 Fourier 特徴法ではデータのクラス毎に分離する特性を捉えられておらず, 従来手法の HyLa では, 基底関数が離散的な表現となっており, データセットが有する連続的なノード同士の関係性を失っていることが確認された. 一方, 提案手法である乱択化 Helgason-Fourier 特徴法は, 螺旋構造により各クラスを連続的な表現を保ったまま, クラス同士を分離していることが確認できた. 以上から, 量的および定性的に提案手法の有効性を確認した.

## 5.2 単一細胞データの可視化

### 5.2.1 実験設定

本節では, 提案手法を潜在変数モデルに拡張した場合の有効性の検証を行う. 本実験では, 文献 [21] で提供される骨髓系前駆細胞の遺伝子発現データを用いた. 本データセットには, 前駆細胞から赤血球やリンパ球といった血液細胞へ分化する細胞群を含んでおり, 細胞の分化に伴う階層構造を有している. 図 5a に本データセットに含まれる細胞群と対応する色付けを示す. 本データセットには,  $N = 2,730$  の細胞群が含まれており, また文献 [22] の手順に従い事前処理された  $D = 1,000$  個の特徴量を有している.

比較手法として, 広く用いられる次元削減手法である UMAP [23] と双曲空間の潜在変数モデルである Bayesian hGP-LVM [24] を採用した. Bayesian hGP-LVM では, 誘導点法に基づく近似を適用し, 勾配降下法に基づく潜在変数の推定を実現している. 提案手法では,  $K_\lambda = 30$ ,  $K_\xi = 40$  と設定し, また潜在変数を無作為の変数で初期化した後, MCMC に基づく潜在変数の推定を行った.

図 5b に可視化結果を示す. 始めに二つの離散的なクラスが生じている UMAP の結果と比較すると, 提案手法および Bayesian hGP-LVM は連続的な細胞の分化の過程を保持していることが確認された. また, Bayesian hGP-LVM では, 赤血球への分化の過程と顆粒球, 単球への分化の過程がクラスター化しており, 提案手法のように連続的な表現ではないことが確

認された。以上から、提案手法の有効性が定性的に確認された。

## 6. ま と め

本文では、双曲空間で効率的なカーネル法を実現するための乱択化 Helgason-Fourier 特徴法を提案した。双曲埋め込みや潜在変数の可視化の実験により、提案手法の有効性が確認された。今後は、複数の次元の双曲埋め込みを用いた手法の検証、および可視化の定量的な検証を行う予定である。

## 文 献

- [1] S. Helgason, *Groups and geometric analysis: integral geometry, invariant differential operators, and spherical functions*, vol.83, American Mathematical Society, 2022.
- [2] M. Nickel and D. Kiela, “Poincaré embeddings for learning hierarchical representations,” *Advances in Neural Information Processing Systems*, vol.30, pp.1–10, 2017.
- [3] H. Cho, B. DeMeo, J. Peng, and B. Berger, “Large-margin classification in hyperbolic space,” *Proceedings of International Conference on Artificial Intelligence and StatisticsPMLR*, pp.1832–1840 2019.
- [4] P. Fang, M. Harandi, and L. Petersson, “Kernel methods in hyperbolic spaces,” *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp.10665–10674, 2021.
- [5] X. Fan, C.-H. Yang, and B. Vemuri, “Horospherical decision boundaries for large margin classification in hyperbolic space,” *Advances in Neural Information Processing Systems*, vol.36, pp.11194–11204, 2023.
- [6] E. Snelson and Z. Ghahramani, “Sparse Gaussian processes using pseudo-inputs,” *Advances in Neural Information Processing Systems*, vol.18, pp.1–8, 2005.
- [7] M. Titsias, “Variational learning of inducing variables in sparse Gaussian processes,” *Proceedings of the International Conference on Artificial Intelligence and StatisticsPMLR*, pp.567–574 2009.
- [8] S. Kumar, M. Mohri, and A. Talwalkar, “Sampling methods for the nyström method,” *The Journal of Machine Learning Research*, vol.13, no.1, pp.981–1006, 2012.
- [9] A. Rahimi and B. Recht, “Random features for large-scale kernel machines,” *Advances in neural information processing systems*, vol.20, pp.1–8, 2007.
- [10] T. Yang, Y.-F. Li, M. Mahdavi, R. Jin, and Z.-H. Zhou, “Nyström method vs random fourier features: A theoretical and empirical comparison,” *Advances in Neural Information Processing Systems*, vol.25, pp.1–9, 2012.
- [11] T. Yu and C. De Sa, “Random laplacian features for learning with hyperbolic space,” *International Conference on Learning Representations*, pp.1–23, 2023.
- [12] J.W. Cannon, W.J. Floyd, R. Kenyon, W.R. Parry, et al., “Hyperbolic geometry,” *Flavors of geometry*, vol.31, no.59-115, p.2, 1997.
- [13] C.K. Williams and C.E. Rasmussen, *Gaussian processes for machine learning*, vol.2, MIT press Cambridge, MA, 2006.
- [14] G.A. Miller, “WordNet: a lexical database for english,” *Communications of the ACM*, vol.38, no.11, pp.39–41, 1995.
- [15] L.M. Schriml, C. Arze, S. Nadendla, Y.-W.W. Chang, M. Mazaitis, V. Felix, G. Feng, and W.A. Kibbe, “Disease ontology: a backbone for disease semantic integration,” *Nucleic acids research*, vol.40, no.D1, pp.D940–D946, 2012.
- [16] C.E. Lipscomb, “Medical subject headings (MeSH),” *Bulletin of the Medical Library Association*, vol.88, no.3, p.265, 2000.
- [17] P. Sen, G. Namata, M. Bilgic, L. Getoor, B. Galligher, and T. Eliassi-Rad, “Collective classification in network data,” *AI magazine*, vol.29, no.3, pp.93–93, 2008.
- [18] O. Shchur, M. Mumme, A. Bojchevski, and S. Günnemann, “Pitfalls of graph neural network evaluation,” *arXiv preprint arXiv:1811.05868*, pp.1–11, 2018.
- [19] P. Mernyei and C. Cangea, “Wiki-CS: A wikipedia-based benchmark for graph neural networks,” *arXiv preprint arXiv:2007.02901*, pp.1–7, 2020.
- [20] M. Nickel and D. Kiela, “Learning continuous hierarchies in the lorentz model of hyperbolic geometry,” *Proceedings of the International Conference on Machine LearningPMLR*, pp.3779–3788 2018.
- [21] F. Paul, Ya’ara Arkin, A. Giladi, D.A. Jaitin, E. Kenigsberg, H. Keren-Shaul, D. Winter, D. Lara-Astiaso, M. Gury, A. Weiner 他, “Transcriptional heterogeneity and lineage commitment in myeloid progenitors,” *Cell*, vol.163, no.7, pp.1663–1677, 2015.
- [22] G.X. Zheng, J.M. Terry, P. Belgrader, P. Ryvkin, Z.W. Bent, R. Wilson, S.B. Ziraldo, T.D. Wheeler, G.P. McDermott, J. Zhu, et al., “Massively parallel digital transcriptional profiling of single cells,” *Nature Communications*, vol.8, no.1, p.14049, 2017.
- [23] L. McInnes, J. Healy, and J. Melville, “UMAP: Uniform manifold approximation and projection for dimension reduction,” *arXiv preprint arXiv:1802.03426*, pp.1–51, 2018.
- [24] K. Watanabe, K. Maeda, T. Ogawa, and M. Haseyama, “Hyperboloid GPLVM for discovering continuous hierarchies via nonparametric estimation,” *Proceedings of the International Conference on Artificial Intelligence and StatisticsPMLR*, pp.1783–1791 2025.