The 33rd Annual Conference of the Japanese Society for Artificial Intelligence, 2019

Stiefel空間上の変分オートエンコーダ

Variational Auto-Encoder On Stiefel Space

三條 嵩明 *1 Takaaki Sanjoh　小宮山 純平 *2 Junpei Komiyama　豊田 正史 *2 Masashi Toyoda　喜連川 優 *2*3 Masaru Kitsuregawa

*1東京大学 大学院情報理工学系研究科
Graduate School of Information Science and Technology, The University of Tokyo

*2東京大学 生産技術研究所
Institute of Industrial Science, The University of Tokyo

*3国立情報学研究所
National Institute of Informatics

This paper presents a reformulation of Variational Auto-Encoder (VAE) framework on a non-Euclidean manifold, the Stiefel space $\mathcal{V}_{m,k}$. By assuming the latent space to be Stiefel manifold, we can use its intrinsic orthonormality to impose structure on the learned latent space representations. We derive an objective function and gradient descendant method for learning VAE using a probabilistic distribution on the Stiefel space.

1 はじめに

変分オートエンコーダ (Variational Auto-Encoder; VAE) は，教師無し生成モデルとして最も広く用いられる手法の一つであり，画像生成など様々な分野に応用されている [Kingma 14, Rezende 14]．VAE は自己符号化器に変分推論の手法を取りこんだモデルであり，何らかの事前分布を仮定して，それに KL divergence の意味で近くなるような正則化を行いながら確率分布を推定する．

この VAE の学習の際には，計算を簡単にするために，データがユークリッド空間 $\mathbb{R}^m$ 上に分布することを仮定し，事前，事後分布としてガウス分布を用いるのが一般的である．しかし，この仮定は $\mathbb{R}^m$ と同相でない標本空間上に分布するデータを学習する際には不適切である．

ユークリッド空間と同相でない標本空間の例として超球面が挙げられる．例えば，タンパク質構造の二面角や風向きといった方向データを扱う上では超球面の標本空間を用いた方が適切であることが古くから知られている [Mardia 75, Fisher 87]．また近年の自然言語処理や画像処理の分野においても，cos 類似度を重視したい等の理由により特徴量ベクトルのノルムによる正規化が行われるような場合は，方向データとして扱われる方が適切である．実際に，いくつかの機械学習タスクについて，ガウス分布の代わりに超球面上の確率分布である von Mises-Fisher 分布を事前分布として用いた VAE の方が安定して学習を行えることが報告されている．[Hasnat 17, Davidson 18, Xu 18].

この超球面を一般化した空間として，Stiefel 空間と呼ばれる空間を考えることができる．Stiefel 空間は空間上の一点が k 個の正規直交基底の組に対応するような空間であり，この空間上の統計に関する研究が近年進められてきた．映像など，それぞれのデータ点が潜在的に正規直交性を持つデータについては，Stiefel 空間上の統計を取り入れることで機械学習手法の性能が向上することが報告されている [Turaga 08].

VAE の学習に Stiefel 空間上の確率分布を用いることで，獲得される潜在表現に正規直交性を課すことができる．そのため，もし取り扱うデータが正規直交性を持つと分かっている場合，Stiefel 空間上の VAE を用いることでその本質的な構造を損なわずに確率的生成モデルを学習できると考えられる．

そこで本稿では，matrix Langevin 分布と呼ばれる，ガウス分布を Stiefel 空間上に制限した，Stiefel 空間上の代表的な確率分布を利用することで，Stiefel 空間を埋め込み空間として持つような VAE の学習手法を構成し，正規直交性を持つようなデータを扱うことを可能とする確率的生成モデルを提案する．

連絡先: {sanjoh, jkomiyama, toyoda, kitsure}@tkl.iis.u-tokyo.ac.jp

2 関連研究

VAE の拡張の多くは，取り扱うデータに適した事前分布と事後分布を潜在空間に課すことによって行われる．代表的な手法としては normalizing flow が挙げられる [Rezende 15]．この手法では，事後確率分布に変形を施すことで，学習可能な確率分布の表現力を高め，より複雑な事後分布を扱うことを可能にしている．しかしながら，この手法は依然としてユークリッド空間上のガウス分布を仮定している．

非ユークリッド空間上で確率モデルの学習を行う手法は，明示的に幾何学的構造を仮定するかどうかで大きく二分される．幾何学的構造を明示しない場合については，RSVGD と呼ばれる手法によって一般のリーマン多様体上で変分推論を行う手法が提案されている [Liu 18]．この手法では多数のパーティクルによって事後分布を近似するため，ノンパラメトリックに確率分布の近似を行うことができ，表現力が高い一方で，高次元での計算量は急峻に増大してしまう．

幾何学的構造を明示する場合については，超球面や双曲空間などの空間を標本空間に持つような確率モデルの学習手法が研究されている [Davidson 18, Ovinnikov 18]．それぞれ，規格化されたベクトルとして表現されるようなデータや階層構造を持つデータに対して，リンク予測などのタスクで低次元でも良い性能が達成されることが報告されている．

Stiefel 空間と呼ばれる，空間の一点が k 個の正規直交基底の順序付き集合に対応する空間上で機械学習手法を考えることで，正規直交性を持つデータに対しては良い性能が得られることが知られている．例えば，[Turaga 08] では状態空間モデルの観測方程式中の行列に正規直交性を課し，Stiefel 空間上の確率分布を用いて行動認識タスクを解くことで，低次元でも良い性能を達成できることを示している．

以上の研究ではいくつかのタスクについて，VAE に対する

非ユークリッド空間上の確率分布を考慮することの有効性と，Stiefel 空間上の機械学習の有効性とが示されてはいるものの，VAE に Stiefel 空間上の統計を取り入れた研究は知りうる限り存在しない．

3 変分オートエンコーダ

VAE では，生成モデルとして潜在状態からデータが生成されているようなモデルを考える．潜在状態を Z として，n 個の潜在状態 $\boldsymbol{Z}$ の分布を $P(\boldsymbol{Z})$ とする．また，n 個の潜在状態 $\boldsymbol{Z}$ それぞれから n 個の観測データ $\boldsymbol{x}$ が生成される確率を $p_\phi(\boldsymbol{x}|\boldsymbol{Z})$ として条件付確率で表す．VAE ではこの条件付確率は $\boldsymbol{Z}$ を入力とし，$\boldsymbol{x}$ を出力とする，パラメータ ϕ を持つニューラルネットとしてあらわされる．目的関数はデータ $\boldsymbol{x}$ が生成される対数尤度 (evidence 関数) $\log \int p_\phi(\boldsymbol{x}|\boldsymbol{Z})p(\boldsymbol{Z})d\boldsymbol{Z}$ であり，この関数を最大化するようにニューラルネットのパラメータ ϕ を学習する．しかし，$\boldsymbol{Z}$ について積分をして直接この関数を計算することは不可能であることが多く，代わりに以下のような evidence 関数の下限 (Evidence Lower BOund; ELBO) の最大化を行うことで，目的関数の最大化を行う．

$$\begin{aligned}&\log \int p_\phi(\boldsymbol{x}|\boldsymbol{Z})p(\boldsymbol{Z})d\boldsymbol{Z}\\&\geq \mathbb{E}_{q(\boldsymbol{Z})}[\log p_\phi(\boldsymbol{x}|\boldsymbol{Z})] - D_{KL}(q(\boldsymbol{Z})||p(\boldsymbol{Z})).\end{aligned} \tag{1}$$

ここで，$q(Z)$ は潜在状態 Z の事後分布 $p_\phi(Z|x)$ の近似事後分布であり，実際に $q(Z) = p_\phi(Z|x)$ となるときに上記の不等式の等号が成立し，evidence 関数が最大化される．しかし，通常は正確に $q(Z) = p_\phi(Z|x)$ を求めることはせず，現実的な計算量で最適化できるように $q(Z)$ の関数クラスを制限し，$q_\psi(Z|x;\theta)$ というようにパラメータ ψ を持つニューラルネットで $q(Z)$ のパラメタ θ を出力することによって近似的に $q(Z)$ の推論をできるようにモデルを組み，そのようなアーキテクチャの下で ELBO の最大化を行う．結局，最終的な目的関数は以下のようになる．

$$\mathcal{L}(\phi,\psi) = \underbrace{\mathbb{E}_{q_\psi(\boldsymbol{Z}|\boldsymbol{x};\theta)}[\log p_\phi(\boldsymbol{x}|\boldsymbol{Z})]}_{\text{reconstruction error}} - \underbrace{D_{KL}(q_\psi(\boldsymbol{Z}|\boldsymbol{x};\theta)||p(\boldsymbol{Z}))}_{\text{KL divergence}}. \tag{2}$$

この目的関数を最大化するようにエンコーダネットワークのパラメタ ψ とデコーダネットワークのパラメタ ϕ の最適化を行う．この目的関数は，推定事後分布 $q_\psi(Z|x;\theta)$ が事前分布 $p(Z)$ から離れ過ぎないように KL divergence によって正則化を行いながら，データ x を生成する尤度を reconstruction error の項によって最大化しようとしている式として理解できる．この目的関数によって VAE の学習を行い，生成モデル $p_\phi(x|Z)p(Z)$ を得ることができる．事前分布 $p(Z)$ や近似事後分布 $q_\psi(Z|x;\theta)$ の関数クラスには正規分布が広く用いられているが，この事前分布分布をデータに適した分布にすることにより，VAE の性能が向上することが報告されている．この事前分布と近似事後分布に Stiefel 空間上の確率分布を適用することを試みる．

4 Stiefel 空間と Stiefel 空間上の確率分布

k 個の m 次元正規直交基底の順序付き集合を k-frame (枠) といい，空間の一点が k-frame 一つに対応するような空間を Stiefel 空間という．この節ではまず Stiefel 空間について説明をした後，Stiefel 空間上で定義される一様分布と，代表的な非一様分布である matrix Langevin distribution について説明する．

4.1 Stiefel 空間

Stiefel 空間 $\mathcal{V}_{m,k}$ は k 個の m 次元正規直交ベクトルの順序付き集合全体からなる空間であり，以下のように定義される．

$$\mathcal{V}_{m,k} = \left\{X \in \mathbb{R}^{m\times k} : X^T X = \boldsymbol{I}_k\right\}. \tag{3}$$

ただし，$\mathbb{R}^{m\times k}$ は $m\times k$ 実行列全体からなる空間であり，$\boldsymbol{I}_k$ は $k\times k$ の単位行列とする．$\mathcal{V}_{m,k}$ はコンパクトな $mk-k(k+1)/2$ 次元リーマン多様体であり，mk 次元ユークリッド空間 $\mathbb{R}^{mk}$ の部分多様体である．$k=1$ の場合に $\mathcal{V}_{m,k}$ は $(m-1)$ 超球面 $\mathbb{S}^{m-1}$ となり，$k=m$ の場合に $O(m)$ となる．ただし，$O(m)$ は m 次元直交群であり，$m\times m$ 実直交行列全体からなり，積が行列積として定義されるような群である．つまり，Stiefel 空間 $\mathcal{V}_{m,k}$ は m 次元正規直交縦ベクトルを k 個順に横に並べたものの全体からなる空間として考えることができて，更に特殊な場合として，$k=1$ の時に正規化された m 次元ベクトル全体からなる空間，$k=m$ の時に m 次元直交行列全体からなる空間として考えることができるような空間である．

4.2 Stiefel 空間上の一様分布

$X \in \mathcal{V}_{m,k}$ として，$\mathcal{V}_{m,k}$ 上の微分形式 $(X^T dX) = \bigwedge_{i=1}^{k}\bigwedge_{j=i+1}^{m} x_j^T dx_i$ は不変測度 (Haar 測度) を定める．これにより，$\mathcal{V}_{m,k}$ の表面積は以下のように計算できる．

$$Vol(\mathcal{V}_{m,k}) := \int_{\mathcal{V}_{m,k}} (X^T dX) = \frac{2^k(\sqrt{\pi})^{mk}}{\Gamma_k(m/2)}. \tag{4}$$

ただし，$\Gamma_m(a)$ は multivariate gamma function と呼ばれるものであり，以下のように定義される．

$$\begin{aligned}\Gamma_m(a) &:= \int_{S>0} \exp\left(\mathrm{tr}\left(-S\right)\right)|S|^{a-(m+1)/2}(dS)\\&= \pi^{m(m-1)/4}\prod_{i=1}^{m}\Gamma\left[a-\frac{1}{2}(i-1)\right],\\&\quad with \qquad a > \frac{1}{2}(m-1).\end{aligned}$$

式 (4) より Stiefel 空間 $\mathcal{V}_{m,k}$ 上で一様な基準測度を以下のように定めることができる．

$$[dX] := \frac{(X^T dX)}{Vol(\mathcal{V}_{m,k})}. \tag{5}$$

4.3 Matrix Langevin 分布

Matrix Langevin ($\mathcal{ML}$) distribution は $\mathcal{V}_{m,k}$ 上の確率密分布として広く用いられており，[Downs 72] によって導入され，その後，[Khatri 77, Jupp 79] による初期研究により，古典的状況下でのパラメタの最尤推定量について調べられた．また，漸近性などその他の性質について [Chikuse 03] に良くまとめられている．$\mathcal{ML}$ distribution は $F \in \mathbb{R}^{m\times k}$ によってパラメタライズされ，基準測度 $[dX]$ に対する確率密度関数 $f_{\mathcal{ML}}(X;F)$ は以下によって与えられる．

$$f_{\mathcal{ML}}(X;F) = \frac{\exp\left(\mathrm{tr}\left(F^T X\right)\right)}{{}_0F_1\left(;\frac{1}{2}m;\frac{1}{4}F^T F\right)}. \tag{6}$$

ただし，${}_0F_1\left(;\frac{1}{2}m;\frac{1}{4}F^T F\right)$ は行列引数超幾何関数である [James 64, Khatri 77].

この分布は多変量正規分布を $X^TX = \boldsymbol{I}_k$ によって条件付けた条件付き確率分布として得られる．また，モーメント $\mathbb{E}[X]$ を定めた時に最大エントロピーを達成する分布として得ることもできる．$F = 0$ の時に一様分布と一致し，$k = 1$ の時に von Mises-Fisher 分布と一致する．

X の期待値はスコア関数の期待値が 0 になることを利用して以下のように求めることができる．

$$\begin{aligned}
&\mathbb{E}\left[\frac{\partial}{\partial F}\log\frac{\exp\left(\mathrm{tr}\left(F^TX\right)\right)}{{}_0F_1\left(;\frac{1}{2}m;\frac{1}{4}F^TF\right)}\right]\\
&=\mathbb{E}\left[\frac{\partial}{\partial F}\mathrm{tr}\left(F^TX\right)-\frac{\partial}{\partial F}\log{}_0F_1\left(;\frac{1}{2}m;\frac{1}{4}F^TF\right)\right]\\
&=\mathbb{E}\left[\frac{\partial}{\partial F}\mathrm{tr}\left(FX^T\right)-\frac{\partial}{\partial F}\log{}_0F_1\left(;\frac{1}{2}m;\frac{1}{4}F^TF\right)\right]\\
&=\mathbb{E}\left[X\right]-\frac{\partial}{\partial F}\log{}_0F_1\left(;\frac{1}{2}m;\frac{1}{4}F^TF\right)\\
&=0,\\
&\therefore\mathbb{E}[X]=\frac{\partial}{\partial F}\log{}_0F_1\left(;\frac{1}{2}m;\frac{1}{4}F^TF\right).
\end{aligned}\tag{7}$$

5 $\mathcal{ML}$ 分布を用いた VAE の学習

4 章の議論により，Stiefel 空間上の VAE の学習手法を構成できる．まず 5.1 節では目的関数の具体的な表式を求める．次に 5.2 節では目的関数の最適化に必要となる勾配計算について議論する．最後に 5.3 節では勾配計算でネックとなる行列引数超幾何関数 ${}_0F_1$ の偏微分の数値計算について説明する．

5.1 目的関数の導出

VAE の目的関数式 (2) について，事前分布 $p(Z)$ に $\mathcal{V}_{m,k}$ 上の無情報分布として一様分布 U を用い，近似事後分布 $q_\psi(Z|x;\theta)$ に $\mathcal{ML}$ 分布を用いることで，Stiefel 空間上の VAE を学習できることが期待される．

目的関数式 (2) の第一項の reconstruction error については以下のように書くことができる．

$$\mathbb{E}_{q_\psi(Z|x;\theta)}[\log p_\phi(x|Z)]=\mathbb{E}_{f_{\mathcal{ML}}\left(Z;F_\psi(x)\right)}[\log p_\phi(x|Z)].\tag{8}$$

ただし，$F_\psi(x)$ はパラメータ ψ を持つエンコーダニューラルネットに対し，あるデータ x を入力した時の出力とする．また，$p_\phi(x|Z)$ はパラメータ ϕ を持つデコーダニューラルネットに対し，$\mathcal{ML}$ 分布からサンプリングした潜在変数 Z を入力して，エンコーダへの入力 x がどの程度復元されそうかを表す尤度である．

次に，目的関数式 (2) 中の第二項の KL divergence を求める．パラメータ F，G を持つ 2 つの $\mathcal{ML}$ 分布 f，g 間の KL divergence は以下の形となる．

$$\begin{aligned}
&D_{KL}(f_{\mathcal{ML}}(Z;F)||g_{\mathcal{ML}}(Z;G))\\
&=\mathrm{tr}\left((F-G)^T\frac{\partial}{\partial F}\log{}_0F_1\left(;\frac{1}{2}m;\frac{1}{4}F^TF\right)\right)\\
&\quad+\log\frac{{}_0F_1\left(;\frac{1}{2}m;\frac{1}{4}G^TG\right)}{{}_0F_1\left(;\frac{1}{2}m;\frac{1}{4}F^TF\right)}.
\end{aligned}$$

これは以下のように導出できる．

$$\begin{aligned}
&D_{KL}(f_{\mathcal{ML}}(Z;F)||g_{\mathcal{ML}}(Z;G))\\
&=\int_{\mathcal{V}_{m,k}}f(Z;F)\log\frac{f(Z;F)}{g(Z;G)}[dZ]\\
&=\int_{\mathcal{V}_{m,k}}f(Z;F)\left(\mathrm{tr}\left(F^TZ\right)-\mathrm{tr}\left(G^TZ\right)\right)[dZ]\\
&\quad+\int_{\mathcal{V}_{m,k}}f(Z;F)\log\frac{{}_0F_1\left(;\frac{1}{2}m;\frac{1}{4}G^TG\right)}{{}_0F_1\left(;\frac{1}{2}m;\frac{1}{4}F^TF\right)}[dZ]\\
&=\mathbb{E}_{f(Z;F)}\left[\mathrm{tr}\left((F-G)^TZ\right)\right]+\log\frac{{}_0F_1\left(;\frac{1}{2}m;\frac{1}{4}G^TG\right)}{{}_0F_1\left(;\frac{1}{2}m;\frac{1}{4}F^TF\right)}\\
&=\mathrm{tr}\left((F-G)^T\mathbb{E}_{f(Z;F)}[Z]\right)+\log\frac{{}_0F_1\left(;\frac{1}{2}m;\frac{1}{4}G^TG\right)}{{}_0F_1\left(;\frac{1}{2}m;\frac{1}{4}F^TF\right)}\\
&=\mathrm{tr}\left((F-G)^T\frac{\partial}{\partial F}\log{}_0F_1\left(;\frac{1}{2}m;\frac{1}{4}F^TF\right)\right)\\
&\quad+\log\frac{{}_0F_1\left(;\frac{1}{2}m;\frac{1}{4}G^TG\right)}{{}_0F_1\left(;\frac{1}{2}m;\frac{1}{4}F^TF\right)}.
\end{aligned}$$

ここで，$G = 0$ とすれば分布 g が一様分布となるため，$\mathcal{ML}$ 分布と一様分布 U の間の KL divergence を求めることができる．これにより，式 (2) の第二項の KL divergence の項は以下の形となる．

$$\begin{aligned}
&D_{KL}(f_{\mathcal{ML}}(Z;F)||U)\\
&=\mathrm{tr}\left(F^T\frac{\partial}{\partial F}\log{}_0F_1\left(;\frac{1}{2}m;\frac{1}{4}F^TF\right)\right)-\log{}_0F_1\left(;\frac{1}{2}m;\frac{1}{4}F^TF\right).
\end{aligned}\tag{9}$$

5.2 勾配降下法による学習

目的関数の最適化には勾配降下法を用いる．そのため，式 (8)，式 (9) の勾配を求める必要がある．ここで，$\mathcal{ML}$ 分布からのサンプリング手法として，提案分布を Stiefel 空間上の一様分布 U とした棄却サンプリング法を用いるとする．この時，式 (8) の勾配推定には VAE で広く用いられる期待値の勾配推定法である reparameterization trick[Kingma 14] を用いることができない．そのため，代替として score function estimator を用いて期待値の勾配推定を行う．score function estimator による reconstruction error の勾配推定は以下のようになる．

$$\begin{aligned}
&\frac{\partial}{\partial F}\mathbb{E}_{f_{\mathcal{ML}}\left(Z;F_\psi(x)\right)}[\log p_\phi(x|Z)]\\
&=\mathbb{E}_{f_{\mathcal{ML}}\left(Z;F_\psi(x)\right)}\left[\log p_\phi(x|Z)\cdot\frac{\partial}{\partial F}\log f_{\mathcal{ML}}\left(Z;F_\psi(x)\right)\right].
\end{aligned}\tag{10}$$

式 (9) の勾配の計算については，$\frac{\partial}{\partial F}\log{}_0F_1$ や $\frac{\partial^2}{\partial F^2}\log{}_0F_1$ の項の計算を除けば通常の自動微分によって計算することができる．

5.3 ${}_0F_1$ の偏微分計算

式 (9) の勾配や式 (10) を求める際に，$\frac{\partial}{\partial F}\log{}_0F_1$ や $\frac{\partial^2}{\partial F^2}\log{}_0F_1$ の項の計算が問題となる．この項に含まれる ${}_0F_1$ は特殊関数となっており，偏微分の式を陽に求めることができない．そのため数値微分によって計算する．

[Khatri 77] によれば，${}_0F_1$ は以下のように単純化できる．

$${}_0F_1\left(;\frac{1}{2}m;\frac{1}{4}F^TF\right)={}_0F_1\left(;\frac{1}{2}m;\frac{1}{4}\Lambda^2\right).$$

ただし，Λ は $F = M\Lambda V^T$ として F を特異値分解した際の特異値対角行列 Λ である．

ここで，$D := \Lambda^2/4$，$D = \mathrm{diag}(d_1, \cdots, d_k)$ と書くとする．${}_0F_1\left(;\frac{1}{2}m; D\right)$ には低次元において効率的な近似的数値計算手法が存在する [Koev 06]．そのため，${}_0F_1\left(;\frac{1}{2}m; D\right)$ の $(d_1, \cdots, d_k)$ による二階までの偏微分は，以下のように刻み幅 h として中央差分による数値微分を用いて求めることができる．

$$\begin{aligned}&\frac{\partial}{\partial d_i}{}_0F_1\left(;\frac{1}{2}m; D\right)\\&= \frac{1}{2h}\left({}_0F_1\left(;\frac{1}{2}m; \mathrm{diag}(d_1, \cdots, d_i + h, \cdots, d_k)\right)\right.\\&\qquad\left.-{}_0F_1\left(;\frac{1}{2}m; \mathrm{diag}(d_1, \cdots, d_i - h, \cdots, d_k)\right)\right). \quad (11)\end{aligned}$$

$$\begin{aligned}&\frac{\partial^2}{\partial d_i \partial d_j}{}_0F_1\left(;\frac{1}{2}m; D\right)\\&= \frac{1}{4h^2}\left({}_0F_1\left(;\frac{1}{2}m; \mathrm{diag}(d_1, \cdots, d_i + h, \cdots, d_j + h, \cdots, d_k)\right)\right.\\&\qquad- {}_0F_1\left(;\frac{1}{2}m; \mathrm{diag}(d_1, \cdots, d_i + h, \cdots, d_j - h, \cdots, d_k)\right)\\&\qquad- {}_0F_1\left(;\frac{1}{2}m; \mathrm{diag}(d_1, \cdots, d_i - h, \cdots, d_j + h, \cdots, d_k)\right)\\&\qquad\left.+{}_0F_1\left(;\frac{1}{2}m; \mathrm{diag}(d_1, \cdots, d_i - h, \cdots, d_j - h, \cdots, d_k)\right)\right). \quad (12)\end{aligned}$$

式 (11)，式 (12) の結果と自動微分を組み合わせることにより，最終的に目的関数式 (2) 中の reconstruction error(式 (8)) と KL divergence(式 (9)) の勾配計算を行うことができる．

6 おわりに

本稿では，matrix Langevin 分布を用いて Stiefel 空間を埋め込み空間として持つような VAE の学習手法を提案した．

今後の課題としては，(1) synthetic なデータで Stiefel 空間上に分布するデータについて，その構造をこの手法によって捉えることができるのかについての検証，(2) 実データを用いた検証，が挙げられる．

謝辞

本研究は JSPS 科研費 16H02905, 17K12736 の助成を受けたものです．

参考文献

[Chikuse 03] Chikuse, Y.: *Statistics on Special Manifolds*, Vol. 174, Springer Science & Business Media (2003)

[Davidson 18] Davidson, T. R., Falorsi, L., De Cao, N., Kipf, T., and Tomczak, J. M.: Hyperspherical Variational Auto-Encoders, in *34th Conference on Uncertainty in Artificial Intelligence* (2018)

[Downs 72] Downs, T. D.: Orientation Statistics, *Biometrika*, Vol. 59, No. 3, pp. 665–676 (1972)

[Fisher 87] Fisher, N. I., Lewis, T., and Embleton, B. J. J.: Statistical analysis of spherical data, *Cambridge: University Press, 1987* (1987)

[Hasnat 17] Hasnat, M. A., Bohné, J., Milgram, J., Gentric, S., and Chen, L.: von Mises-Fisher Mixture Model-based Deep learning: Application to Face Verification (2017)

[James 64] James, A. T.: Distributions of Matrix Variates and Latent Roots Derived from Normal Samples, *The Annals of Mathematical Statistics*, Vol. 35, No. 2, pp. 475–501 (1964)

[Jupp 79] Jupp, P. and Mardia, K.: Maximum likelihood estimators for the matrix von Mises-Fisher and Bingham distributions, *The Annals of Statistics*, Vol. 7, No. 3, pp. 599–606 (1979)

[Khatri 77] Khatri, C. G. and Mardia, K. V.: The Von Mises-Fisher Matrix Distribution in Orientation Statistics, *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 39, No. 1, pp. 95–106 (1977)

[Kingma 14] Kingma, D. P. and Welling, M.: Auto-Encoding Variational Bayes, *Proceedings of the 2nd International Conference on Learning Representations (ICLR)* (2014)

[Koev 06] Koev, P. and Edelman, A.: THE EFFICIENT EVALUATION OF THE HYPERGEOMETRIC FUNCTION OF A MATRIX ARGUMENT, *Mathematics of Computation*, Vol. 75, No. 254, pp. 883–846 (2006)

[Liu 18] Liu, C. and Zhu, J.: Riemannian Stein Variational Gradient Descent for Bayesian Inference, in *AAAI* (2018)

[Mardia 75] Mardia, K. V.: Statistics of Directional Data, Technical Report 3 (1975)

[Ovinnikov 18] Ovinnikov, I.: Poincaré Wasserstein Autoencoder, in *Third workshop on Bayesian Deep Learning (NeurIPS 2018)* (2018)

[Rezende 14] Rezende, D. J., Mohamed, S., and Wierstra, D.: Stochastic Backpropagation and Approximate Inference in Deep Generative Models, in *ICML*, pp. 1278–1286 (2014)

[Rezende 15] Rezende, D. J. and Mohamed, S.: Variational Inference with Normalizing Flows, in *ICML* (2015)

[Turaga 08] Turaga, P., Veeraraghavan, A., and Chellappa, R.: Statistical analysis on Stiefel and Grassmann Manifolds with applications in Computer Vision Pavan Turaga , Ashok Veeraraghavan and Rama Chellappa Center for Automation Research University of Maryland , College Park, *Cvpr* (2008)

[Xu 18] Xu, J. and Durrett, G.: Spherical Latent Spaces for Stable Variational Autoencoders, in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 4503–4513 (2018)