Mathematical Analysis Team
Shin-ichi Ohta
数理解析チーム　太田 慎一

RIKEN
Center for Advanced Intelligence Project

滑らかとは限らない部分多様体に対する Laplacian Eigenmapsの収束とそのレート

数理解析チーム　基礎科学特別研究員　相野眞行

次元削減

高次元空間のデータ
$x_1, x_2, \ldots, x_n \in \mathbb{R}^D$

⇩ 変換

低次元表現
$y_1, y_2, \ldots, y_n \in \mathbb{R}^d$

線形次元削減

主成分分析（PCA）

非線形次元削減

多様体学習

Laplacian Eigenmaps

Laplacian Eigenmaps

Laplacianの固有値を用いた次元削減
閉部分多様体$\mathrm{M} \subset \mathbb{R}^D$のLaplacianの固有関数
$f_1, f_2, \ldots, f_d$を用いて

$$y_i = (f_1(x_i), f_2(x_i), \ldots, f_d(x_i)) \in \mathbb{R}^d$$

と定めたい。

実際は有限個のサンプル$x_1, x_2, \ldots, x_n \in \mathbb{R}^D$から
$f_1, f_2, \ldots, f_d$に対応するものを計算

アルゴリズム

パラメータ$\varepsilon \in (0, \infty)$を選択。
行列$K, D, L \in \mathbb{R}^{n \times n}$を次のように定義：

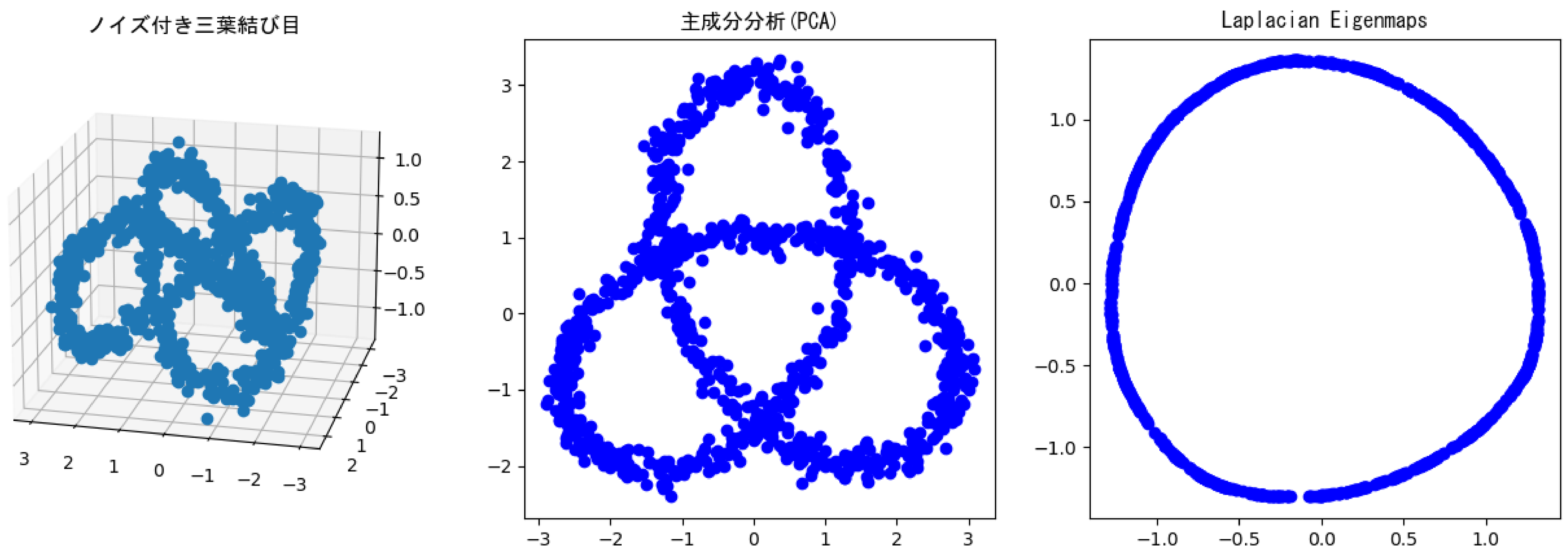
$$K_{ij} = \begin{cases} 1 & \|x_i - x_j\|_{\mathbb{R}^D} < \varepsilon \\ 0 & \text{その他の場合,} \end{cases}$$

$$D_{ij} = \delta_{ij} \sum_{l=1}^{n} K_{il}, \quad L = D - K$$

固有値問題$Lv = \lambda Dv$を解き固有値$\lambda = 0$を除いて小さい順の固有値に対応する固有ベクトルd個$v^1, v^2, \ldots, v^d \in \mathbb{R}^n$を取る。

$$y_i = (v_i^1, v_i^2, \ldots, v_i^d) \in \mathbb{R}^d$$

と定める。ここでv_i^jはv^jの第i成分。



Laplacian Eigenmapsはデータのユークリッド空間への入り方ではなく内在的な性質を見ようとする。

サンプリング条件

閉部分多様体$\mathrm{M} \subset \mathbb{R}^D$上の確率測度$\mu = \rho \mathrm{Vol}_g$に従ってランダムに (i.i.d.に)サンプル
$x_1, x_2, \ldots, x_n \in \mathrm{M}$が得られているとする。
ρはLipschitz定数L_ρのLipschitz関数で

$$\frac{1}{\alpha} \leqq \rho \leqq \alpha \ (\alpha > 1)$$

更にReachと呼ばれる量の下からの評価Reach $\geqq R$のもとで次の重み付きLaplacianにスペクトル収束する。

[Belkin, Niyogi, NIPS, 2006], [Trillos-Gerlach-Hein-Slepčev, FOCM, 2020], [Calder-Trillos, arXiv:1910.13476v2]

$$\Delta_\rho = \Delta - 2\langle \nabla \log \rho, \nabla \cdot \rangle$$

ただし$\Delta = -\mathrm{tr}\,\mathrm{Hess}$.
しかしReach $\geqq$ Rを満たす部分多様体列では、
滑らかとは限らない部分多様体は近似できない。
[Breuning, J. Geom. Anal, 2015]を使って示される。
➡より弱い第二基本形式のL^1評価$\|II\|_{L^1} \leqq S$を仮定。
ただしReach $\geqq R$は第二基本形式、断面曲率、単射半径の評価 $\|II\|_{L^\infty} \leqq 1/R$, $|\mathrm{Sect}| \leqq 1/R^2$, inj$\geqq \pi R$を導く。

主結果 (arXiv:2110.08138)

サンプリング条件が成り立ち、ある$S, K, i_0, L > 0$について$\|II\|_{L^1} \leqq S$, $|\mathrm{Sect}| \leqq K$, inj$\geqq i_0$および$d_M(x, y) \leqq L\|x - y\|_{\mathbb{R}^D}$ $(x, y \in \mathrm{M})$とする。
パラメータ$\varepsilon \in (0, \infty)$をMの次元とサンプル数$n$に対して適切に選びつつ$n \to \infty$とすると、
固有値問題$Lv = \lambda Dv$はΔ_ρにスペクトル収束。

上記の仮定を満たす滑らかな部分多様体列によりある種の微分不可能な部分多様体は近似される。
➡ そのような部分多様体についてもスペクトル収束が成立。