

研究終了報告書

「レーブグラフの順序定式化のデータ解析に向けた数理」

研究期間：2019年10月～2023年3月

研究者：宇田 智紀

1. 研究のねらい

計算機が発達し多種多様なデータが溢れる現代科学・社会において、その有効な利活用のためにデータ解析が果たす役割は極めて大きなものとなっている。位相的データ解析は位相不変量に着目してデータの「形」を解析する学問分野であり、その中でも、パーシステントホモロジーやレーブグラフはある種の空間の形を捉える基本的な考察対象である。レーブ順序は、順序構造に着目したレーブグラフの新たな離散的定式化であって、既存手法になかった特徴(0次・1次パーシステントホモロジーからの構成、離散データからの非補間直接計算法等)を持つ。本研究は、その理論的な基礎を確立しつつデータ解析のために特に重要と思われる課題を扱い、位相的データ解析のさらなる発展を目指すものである。

2. 研究成果

(1) 概要

レーブ順序法は、流線トポロジー分類理論に基づくデータ解析(流線トポロジー解析)を実現するために2018年に提案された新しい手法および数学的定式化である。所与のデータを軸に沿って連続的に輪切りにしていく時のトポロジー変化の様子をグラフ構造で表現したものがレーブグラフであり、これを離散データの順序構造から復元するのがレーブ順序法である。本研究は、大きく分けて以下の三つの研究テーマに取り組むものである：

- A. 数理的基礎の整理と確立
- B. 発展的数理課題
- C. 計算機科学的課題

数理的基礎の確立は理論・応用のいずれの面でも極めて重要であり、研究テーマ A では「レーブ順序法に関連する圏論的知見の整理」「レーブ順序法の安定性定理の証明」等に取り組んだ。レーブグラフと同様に、レーブ順序法が入力データに連続的に依存することを証明した。離散的・半離散的・連続的なデータを画一的に取り扱うために半順序空間の圏の性質を整理し、その上で interleaving 距離を定義した。また、安定性を証明するために、半順序集合の基本亜群、一般化木の定義、半順序集合の分岐完備性、半順序集合上の道の強正規化定理といった新たな概念や定理を導入した。

研究テーマ B では、レーブ順序法のデータ解析的応用をより広範なものとするために「レーブ順序集合間の計算可能な距離構造の考察」に取り組んだ。Interleaving 距離は(計算機上での)計算が困難である。代替として、レーブグラフを測度距離空間とみなし Gromov-Wasserstein (GW) 距離を検討した。具体的には interleaving 距離との不等式評価を試みたが、両者の本質的な差のためにこの方向性では有意義な結果が得られないことが判明した。また、Python-OT に実装されている正則化 Gromov-Wasserstein 距離を用いてレーブグラフを比較する数値実験にも取り組んだ。

研究テーマ C では、「計算量評価と高速化」に取り組んだ。理論的な計算量評価はあまり改善が見込めず、実用上の組み合わせ爆発の問題も考慮し Python 実装のライブラリの一部を C++ 実装に置き換えることを試みた(ソフトウェア開発は外注)。現状のライブラリ実装のインタフェースなどが障壁となり単純な高速化には至らなかったが、高速化における懸念事項を洗い出すことができ、得られた知見を整理して現在企業との共同研究を進めている。

(2) 詳細

研究テーマ A「数理的基礎の整理と確立」

[安定性定理]

Silva らによるレーブグラフの interleaving 距離を拡張することでレーブ順序に適した interleaving 距離を定義し、これを用いてレーブ順序の安定性を証明した。レーブ順序の安定性とは、空間(あるいはグリッド) X 上のスカラー関数 f のレーブ順序 $R(X, f)$ について

$$d(R(X, f), R(X, g)) \leq \|f - g\|_{\infty}$$

が成り立つことである。

[半順序空間の圏と平滑化関手]

半順序空間の interleaving 距離を定義するために半順序空間の圏 **PoHaus** の性質を整理した。具体的には、圏 **PoHaus** が余完備であって、前順序位相空間の圏 **QoTop** への包含関手 **PoHaus** \hookrightarrow **QoTop** が左随伴 **QoTop** \rightarrow **PoHaus** (pospace reflector) を持つことを証明した。余完備性より **PoHaus** 上の平滑化関手 u_{ε} が定められる。

[Interleaving 距離]

平滑化を用いて、高さ付き順序集合を高さ方向の誤差 ε 程度を許してトポロジーを(柔らかく)比較する ε -同型を定義できる。Interleaving 距離はこの誤差 ε の下限である。

[一般化木と順序集合の基本群]

位相幾何学的な木と数学基礎論における木の両方の良い特徴付けを併せ持つ、基本群による木の特徴付けを得た。ホモトピー同値に相当する同値関係によって順序集合上の道の同一視をして道が為す代数を導入し、これを順序集合の基本(亜)群と呼ぶことにする。連結順序集合 X が自明な基本群を持つとき X を木と呼ぶ。

[鎖道の強正規化定理]

順序集合が良い分岐構造を持つとき、鎖道の強正規化定理が成り立つ。大雑把に言えば、強正規化定理は鎖道から最短な鎖道へのホモトピー変形を陽に記述できる、強力な道具である。これを用いると一般化木の諸性質の証明が容易となる。

[当初の想定とは異なる成果について]

当初、安定性定理の証明は既存のレーブグラフに対するものと同様に容易に証明可能と考えていた。しかし、離散的なデータを取り扱えるレーブ順序固有の特徴のために、木の新たな特徴付け等の良い性質の確認を要した。順序集合の基本(亜)群や強正規化定理はその過程で副次的に得られたものではあるが、それら自体も数学的に独立した興味深い成果である。

研究テーマ B「発展的数理課題」

確率測度距離空間の間には Gromov-Wasserstein (GW) 距離が定まる。無向連結グラフ $G = (V, E)$ の頂点集合 V に総和が 1 となるように適当な重み μ_V を備えれば確率測度距離

空間とみなせる。したがって、GW 距離でグラフ同士を比較できる。

[Interleaving 距離と GW 距離の不等式評価]

GW 距離は測度距離空間としてグラフを比較するのに対して、interleaving 距離はグラフに備わる高さと分岐の構造を比較するという本質的な差異がある。この本質的な差異のために、GW 距離と interleaving 距離の間に有益な不等式評価を確立するのは困難であることが分かった。

[数値実験の結果]

ライブラリ PythonOT には GW 距離やその変種がいくつか既に実装されており、このうち正則化 GW 距離を用いて簡単な実験結果を行った。その結果を以下に示す。

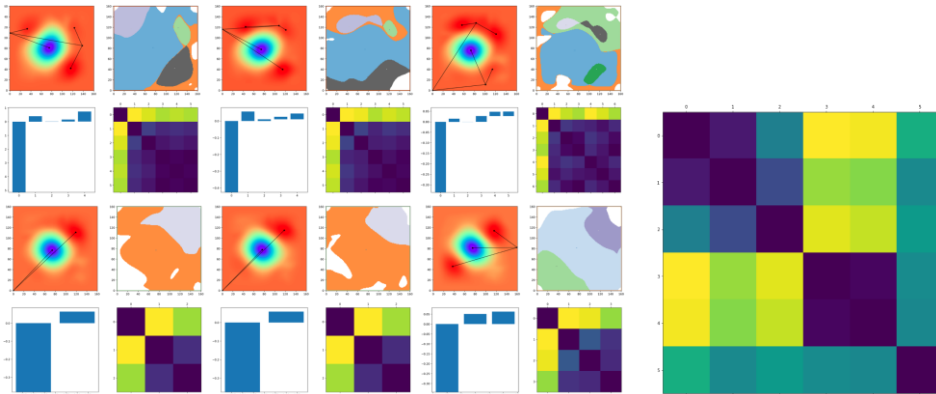


図 1. (左) 六つの入力データの計算結果を 2 行 × 3 列 × (2 行 × 2 列) に並べたもの。
(右) レーブグラフ同士の正則化 GW 距離を計算して得た距離行列のプロット。

図 1 の各 2 × 2 セルの内容はそれぞれ次の通りである。左上は入力データのスカラー関数(高さ関数) f のカラーマップの上にレーブ順序法で得たレーブグラフを重ね描きしたものである。右上はレーブ順序のパーティション図で、スカラー場の等高線のトポロジー関係を可視化している。左下はバーコード図で、レーブグラフの各辺 $e = (v, v')$ に対して $(f(v), f(v'), d_f(v, v'))$ をバーで図示している。右下はレーブグラフの距離行列 $d_G(v, v')$ で、暗色は距離が 0 に近いことを、明色は距離が大きいことを表している。これらのレーブグラフを一様重みによる確率測度距離空間 $G_i = (R(X, f_i), d_{R(X, f_i)}, n_i^{-1})$ とみなして正則化 GW_1 距離で比べたものが右の図である。 G_3, G_4 が似通っており他のグラフとは似ていないといった特徴が捉えられている。

研究テーマ C「計算機科学的課題」

レーブ順序法のアルゴリズムにおける理論上のボトルネックはパーシステントホモロジーに相当する合併木の計算の部分である。そこで、ライブラリの主要部を pybind11 を用いて C++ 実装で置き換えることで高速化を試みた(開発は外注)。結果、ライブラリは速度改善はできなかった。

その他の研究

中野直人氏、池祐一氏らと画像処理に関する共同研究を行い、簡単なセグメンテーション実験を行い TDA の新たな応用の一可能性を検討した。これは、レーブ順序法が持つパーティションの性質を応用してグレイスケール画像をピクセル値による“等高線(の離散的類似)”で分割し特徴的な部分を抽出するものである。レーブ順序法の安定性を検証するためノイズを加えた MNIST 画像で実験を行い、適切な設定下では画像データのトポロジカルな構造を抽出できた(図 2)。共同研究成果は JSIAM 研究部会連合発表会にて発表の上 JSIAM Letters 誌に投稿中である。

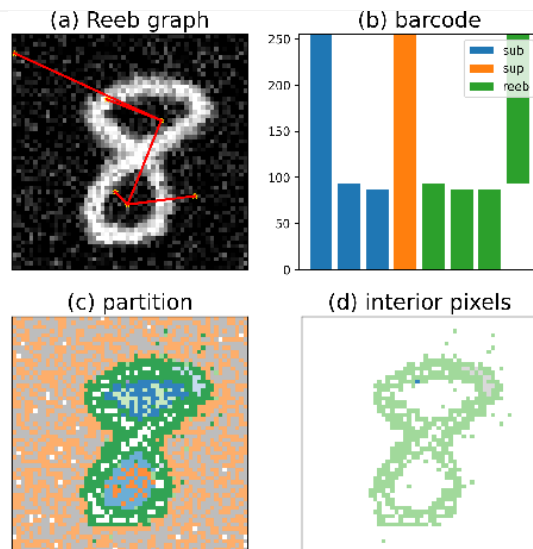


図 2. レーブ順序法による領域分割と特徴検出の実験

また、坂上貴之氏らとは流線トポロジー

解析の応用に関する共同研究を行い、気象学および海洋学への応用成果の論文がそれぞれ 2021 年、2022 年に出版された。流線トポロジー解析にはレーブ順序法が使われており、データ解析におけるレーブ順序法のポテンシャルを示すことができた。

3. 今後の展開

安定性定理はデータ解析への応用では非常に重要である一方で、その不等式評価自体は数学的には依然として粗いものである。よって、レーブ順序法の信頼性をより高めるためには、理論面では GW 距離など様々な距離による安定性定理の確立、応用面ではノイズを含むデータに対する統計実験など、一歩さらに踏み込んだ研究の展開が必要であると考えている。また、レーブ順序法は新しい手法であるため、現時点ではまだその数学的な詳細や応用上の価値が多くの研究者に理解されているとは言い難く、この現状を打破するためにはレーブ順序法の有用性を分かりやすく示す理論・応用研究を積み重ねていくべきであろうと考えている。引き続き数学、TDA、データ解析など様々な方面の研究集会での研究発表を積極的に行いつつ、アウトリーチ活動も進めていきたい。こうした地道な研究活動の積み重ねには、短くとも(程度にもよるが)5~10 年はかかると思われる。

また、安定性定理の証明の過程で副次的に得られた、木の一般化・順序の基本群・強正規化定理などの新たな数学的知見に関しても、今後レーブ順序とはまた独立した応用や理論の一般化を模索していきたいと考えている。これに関しても、現段階でレーブ順序法以外での応用や一般化の目処が立っている訳ではないため、方針の模索だけでも数年はかかるであろう。

GW 距離の理論解析は難しいが、一方で PythonOT ライブラリの普及により計算実験は比較的容易になっているため、この方向性での応用研究は短いタイムスパンで継続していきたい。

4. 自己評価

研究テーマ A に関しては、安定性定理の証明とそのための理論整備に想定よりもかなり時間を要してしまったものの、結果として副次的な理論的成果(半順序空間の圏の性質やリフレクター、

順序集合の基本群, 強正規化定理等)が複数得られた点で非常に良かったと考えている。このため, レーブ順序法の数理的諸性質の理解を深めるという当初の目的も十分に達成されたと考える。また, 圏論や半順序集合の基礎的な理論という一見応用になかなか結びつきそうにない数学分野をデータ解析に役立てるといふ, ある種標語的にかつ学際的な貢献が出来たと考えており, 今後こうした側面を押し出してより積極的に社会へのアウトリーチ活動を進めていきたい。

研究テーマ B では, 数学的にはポジティブな結果こそ得られなかったものの, 今後研究をデータ解析方面へと展開する上で有意義な知見を蓄積できた点は良かったと考えている。レーブ順序法のデータ解析的研究のポテンシャルを広げるという目的は一部達成しているものの, 研究期間中に学術的成果として残せなかった点は残念である。

研究テーマ C に関しても, 研究費による高速化ライブラリ開発の外注は適正ではあったと考えているが, 結局高速化がうまく行かなかった点は極めて残念である。ソフトウェア開発における要件定義の難しさを再認識することができた。一方, 現在このテーマ研究を通して得た知見を部分的に活かす形で企業との共同開発・研究を開始しており, 研究成果から直接的な形ではないものの今後の科学技術発展の貢献に少なからず寄与しているとも考える。

5. 主な研究成果リスト

(1) 代表的な論文(原著論文)発表

研究期間累積件数: 2 件

1. Takashi Sakajo, Shun Ohishi, Tomoki Uda. Identification of Kuroshio meanderings south of Japan via a topological data analysis for sea surface height. *Journal of Oceanography*. 2022, 78, pp. 495–513, [DOI:10.1007/s10872-022-00656-3](https://doi.org/10.1007/s10872-022-00656-3)

海面高度観測再解析データに対する流線トポロジー解析に基づき黒潮大蛇行期間を同定するアルゴリズムを提唱した。アルゴリズムはレーブグラフと部分円順序根付き木表現を用いて東海沖渦流を時系列追跡する。黒潮南端座標の時系列は気象庁による既存結果と高い相関を示した。1993~2020 年の間に五つの蛇行エピソードが検出され, それぞれの蛇行終了間際には大きな渦が二つの小さな渦に分裂する事象が観測された。

2. Tomoki Uda, Takashi Sakajo, Masaru Inatsu, Kazuki Koga. Identification of Atmospheric Blocking with Morphological Type by Topological Flow Data Analysis. *Journal of the Meteorological Society of Japan*. Ser. II. 2021, 99, pp. 1169–1183, [DOI:10.2151/jmsj.2021-057](https://doi.org/10.2151/jmsj.2021-057)

流線トポロジー解析に基づき気圧のブロッキング現象期間を同定するアルゴリズムを提案した。アルゴリズムはレーブグラフと部分円順序根付き木表現を用いてブロッキング候補領域を時系列追跡する。従来手法と比べて, 気象学的先見知識に基づくパラメータが少なく, また結果もパラメータ選択に鋭敏ではない。さらに, Ω 型ブロッキングと双極型ブロッキングの半自動的な形態判定が可能である。

(2) 特許出願

研究期間全出願件数: 0 件(特許公開前のもも含む)



(3) その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

[主要な学会発表]

- 宇田智紀. 半順序集合の基本群と一般化木. 応用数学合同研究集会(解析系), 2022/12.
- 宇田智紀. 流線トポロジー解析の手法と応用. 日本学術会議第 11 回計算力学シンポジウム, 2021/12.
- Tomoki Uda. On Interleaving Distance between Reeb Trees as \mathbf{R} -Pospaces. Asia Pacific Seminar on Applied Topology and Geometry, 2021/11.
- Tomoki Uda. Stability of Reeb Trees and Application to Noisy Images. Topological Data Analysis and Machine Learning. 2021/07.
- 宇田 智紀, 中野 直人, 池 祐一. レーブ順序法の画像輪郭抽出への応用. 第 17 回研究部会連合発表会. 2021/03.

[受賞]

- 宇田 智紀, 横山 知郎, 坂上 貴之: パーシステントホモロジーとレーブグラフを用いた 2 次元ハミルトンベクトル場の流線位相構造の自動抽出アルゴリズム, 日本応用数理学会論文誌論文賞, 日本応用数理学会, 2022/09.