公開

研 究 終 了 報 告 書

「データドリブン計算代数幾何」

研究期間: 2020 年 11 月～2023 年 3 月
研 究 者: 計良 宥志

1. 研究のねらい

本研究のねらいは，代数幾何・計算代数（以下，計算代数幾何）で発展してきた非線形な数学をデータドリブンな計算代数幾何として刷新した新たな学問領域を立ち上げ，機械学習をはじめとした現代のデータドリブンの情報科学における「観測（データ）からの学習」というフレームワークに応用可能にすることにある．取り組む課題は大きく二つに分かれる．一つはデータ（ベクトルの集合）が与えられた時，そのデータが所属する代数多様体を発見することである．代数多様体の発見とは，具体的には，その代数多様体を記述する多項式系を計算することにあたり，消失イデアルの基底計算と呼ばれる問題として取り組まれている．既存の代数幾何や計算代数は多項式系の分析・処理を得意としているため，まずデータを特徴づける良い多項式系を得ることが重要である．その点で，消失イデアルの基底計算は大きな役割を担っている．もう一つの課題は，得られた近似的な多項式系を数値的な処理のみで解析することである．現在の大規模計算は数値計算によるものが中心であり，手法やライブラリ等もそれらについて発展している．そのため，従来の記号計算中心の代数処理から脱却し数値計算へ移行可能にすることは，提案フレームワークの柔軟性を飛躍的に高める．総じて，本研究の提案するデータドリブン計算代数幾何は，現代のデータドリブンな情報科学応用の動機から，計算代数幾何という強力な非線形数学を刷新するという独創的な挑戦であり，数理科学と情報科学の境界領域に対し基礎的な貢献を為すものである．これは，「観測（データ）からの学習」という現代の情報科学応用で最も本質的な問題に取り組むものであり，統計的手法と線形代数に支えられる現代のデータドリブンな情報科学から，統計的手法と非線形代数で展開される新たな情報科学への発展を目指す．

2. 研究成果

（1）概要

本研究計画を通じて，消失イデアルの基底計算に関わる手法と理論解析で大きな成果を得た．これらの成果は実応用上重要な近似基底計算に自然に拡張できる点でも重要である．主な成果としては次のとおりである．

（1）従来の係数正規化が実用上さまざまな問題を抱えることを発見し，それを解決するデータドリブン正規化の手法と理論的基盤を打ち立てた．

（2）近似基底計算の経験的な計算量と理論的計算量のギャップを説明する新たな解析を与え，近似基底計算の計算量がサンプル数に関して線形であることを明らかにし，その実用性を理論的に示した．

（3）一部のタスクで求められる識別的な多項式集合を，外部サンプルなしに計算するアルゴリズムを提案し，そこから得られる多項式集合に代数的解釈を与えた．

（1）に関しては，多項式の正規化が近似基底計算で非常に重要な処理である点，そして従来見過ごされていた係数正規化に問題がある点を新たな課題として発見したこと自体に価値が

ある. また実用上不都合な単項式順序に関する仮定を避けた場合でも多項式時間で実行できる初の正規化を提案した. これは係数正規化と異なりデータドリブンな正規化であり, 係数正規化では実現できないさまざまな性質を基底計算の出力に実現できる. また単項式順序の仮定を許容した場合でもこのデータドリブン正規化を用いる利点は多く存在する.
(2)に関しては, 従来の解析が厳密な基底計算に限定される問題を取り上げ, 近似的な基底計算に適用可能な新たな解析を考案した. これにより, 従来の悲観的な理論計算量を大きく改善し, 経験的な計算量に沿う理論解析を得た.
(3)に関しては, クラス分類等で求められる識別的な多項式集合を得るための従来手法が, その出力に対する代数的解釈を諦めていた問題に取り組んだ. 結果として代数的解釈を伴う初めての識別的多項式集合の計算アルゴリズムを得た. また従来手法は識別的多項式集合を得るために追加の(外部)サンプルを必要としたが, 提案手法では外部サンプルなしに識別的多項式を得られる点も実用上嬉しい.
総じて, 消失イデアルの近似基底計算において, 実用を意識した効率的・効果的なアルゴリズムの提案とその理論的基盤の構築が成果の中心である. 本問題はデータと代数的世界を結ぶ重要な問題であり, それに対しさまざまな手法と理論を得たことは大きな成果であると考える.

(2)詳細

研究課題 A 「消失イデアルの近似基底計算とデータドリブン正規化」

消失イデアルの近似基底計算では, ベクトル集合が与えられた時, その集合上でゼロ値をとる多項式(消失多項式)の計算を行う. 意味のないゼロ多項式や係数が小さいだけの多項式を避けるため, 多項式の正規化が必要になる. 最も標準的なのが係数正規化であるが, 本研究はそれに変わる新たな正規化として, データを用いたデータドリブン正規化(勾配正規化・勾配重み付き正規化)を提案し, 係数正規化では実現できない以下の有用な理論的性質を示した.

(1) 単項式順序(実用上仮定したくない)を仮定しない場合, 係数正規化では正規化の指数オーダの計算量が必要となる. それに対してデータドリブン正規化は多項式時間で実現できるため, 非常に効率的である.
(2) 単項式順序を仮定しない場合の基底計算において, 勾配正規化を用いると冗長な多項式が出力に含まれることを避けられる. 従来の理論解析ではサンプル数の 2-3 乗程度の数の多項式が出力されるが, 本研究ではそれが 1 乗以下に抑えられることを示した.
(3) 実応用上重要となる近似基底計算において, 与えられるベクトル集合のスケーリング(計算の安定性などのための前処理)に対して, 係数正規化による基底計算は失敗しうる(失敗させるようなスケーリングが常に存在することを明らかにした). それに対し, データドリブン正規化は, どのようなスケーリングに対しても一貫した結果を出力できる.

（1），（2）はともに，効率の面で非常に重要である．（1）は理論的な側面で重要な処理を現実的な計算コストで初めて実現した点，また（2）も従来の手法は（少なくとも理論的には）非常に冗長な多項式集合を出力していたことを発見し，提案手法がそれを現実的な計算コストで解消する点に価値がある．また（3）は，データのスケーリングという極めて単純な前処理が近似基底計算に大きな影響を与えうることを初めて発見し，それを解消した．データドリブンに正規化を行うというのは，そもそも「データ」を処理しない従来の計算機代数・代数幾何の分野で考え得ないものである．そのため，データドリブン正規化は理論・実応用の両面で重要となる新たな視点と研究の発展の方向性を与えられたと考える．

研究課題 B「近似基底計算のスケール性」
消失イデアルの基底計算の実用面の課題の一つとして，そのスケール性の悪さが挙げられる．そのため，現在の主流である大規模・高次元データの扱いが難しい．具体的には，（個々の手法により差はあるが，）理論的にはサンプル数の 3–4 乗程度の計算量となる．そのため，一見大規模データへの適用が難しいように思われる．しかし，経験的な計算量はかなり小さく，およそサンプル数に対して線形に変化する．この差は，理論においては厳密な基底計算の場合を考えており，実用される近似基底計算に対する良い解析を与えられていないためである．本研究では，近似基底計算の設定に沿った新たな解析を行い，この経験と理論のギャップを解消した．具体的には，近似基底計算（多項式を低次から高次へ順次計算する）が停止する次数と近似の度合いの関係を明らかにする不等式を導いた．この不等式は停止次数がサンプル数に関して線形な量で上から抑えられることを示している．本成果により，近似基底計算の実用性を理論面から支えることができた．残る課題として，課題 A のデータドリブン正規化と組み合わせた場合と高次元データを扱う場合とでそれぞれ同様のことが実現できるかという点が挙げられる．

研究課題 C「識別的な多項式集合の獲得」
消失イデアルの基底計算が出力する多項式集合は，入力されたサンプル集合の消失イデアルの基底という形でサンプル集合を特徴づける．例えばクラス分類等で適用する場合，各クラスのサンプル集合それぞれから多項式集合を計算し，それらを総合して各サンプルの特徴ベクトルを構成することになる．このフレームワークでは，各クラスの多項式集合の計算において異なるクラス間のサンプルが影響しないため，識別的な多項式集合が得られない．これを解決するために，いくつかの既存手法は複数のクラスのサンプルを基底計算に同時に用いることを提案している．しかし，この場合に得られる多項式集合が，代数的にどのような意味を持つのか（その多項式集合が生成するイデアルが何であるのか）についての解釈を持たない．代数的な理論基盤は消失イデアルの基底計算の特徴であるため，本研究では，代数的な解釈を持ちつつ識別な多項式集合を出すアルゴリズムの構築を行なった．その際，機械学習分野で近年発展している自己教師あり学習のフレームワークを取り込むことで，他クラスのデータがない・定義されないような設定でも識別的な多項式集合を得ることを可能にした．これにより，例えば分布外検知という他クラスのデータが与えられない問題設定に適用可能となった．

その他の成果.
本研究計画を通じて，国内外のさまざまな研究者との繋がりを得，共同研究を行うことができた. 具体的には，池祐一氏（東京大学，本領域 1 期生）や，舛谷亮祐氏（東京都立大）との協力で課題 C の成果を得た. また Elias Wirth 氏 (Zuse Institute Berlin), Sebastian Pokutta 氏 (Zuse Institute Berlin)との共同研究では課題 B の成果を得た. さらにまだ成果としてまとめられてはいないが，Achim Kehrein 氏（Rhine-Waal 大）との共同研究も進んでいる.

3. 今後の展開

本研究成果は，主に与えられたデータに代数的手法で特徴づけを与えるためのアルゴリズムと理論的基盤を整備するものであった. 今後はこの理論的基盤をさらに拡張しつつ，具体的な問題に特化し実用性の高いものを目指すことを考えている. 具体的には，データドリブン正規化とスケール性の両立や高次元データへの対応を理論的に議論しつつ，多項式が強力なツールとなる設定やデータの模索を行う. 理論・アルゴリズム面が発展すれば，さまざまな設定で強力になる一方で，設定やデータが固定されれば，より特化した手法開発や議論を行うことができる.
理論・アルゴリズムの整備と適用課題の模索に 3-4 年程度を考えている. その後は，従来の統計中心のデータ処理・学習に対し，代数という新たなアプローチ群を適用可能にし，データドリブン科学の広がりに貢献したい.

4. 自己評価

本研究計画を通じて，当初計画していた二つの研究課題のうち，消失イデアルの基底計算に関わる部分で実用・理論の両面でさまざまな成果を得られた. 特に，従来の手法では見落とされていた点，実現できなかった点に取り組み実現できた点が良かった. この成果は，計算機代数と機械学習等の応用分野の双方に通じている自身の強みを活かした独自性のあるものだと考える.
もう一つの研究課題である，近似多項式系処理の数値的処理に関してはあまり取り組むことができなかった. 一つ目の課題に関して，新たな共同研究を通じて広がりを得たこと，また議論と試行錯誤の中で課題自体の再設定の必要性を感じたことが主な理由である.
総じて，得られた研究成果は分野横断的な視点で得られた独自性のあるものであり，新たな研究の展開を可能にする点で波及効果があると考える. またこれらの成果は，研究総括・領域アドバイザー・また同領域の若手研究者，そして海外研究者との議論の中で新たに広がりを得たものであり，研究者ネットワークと相互触発の効果であると考える.

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：1件

1. Hiroshi Kera, "Border Basis Computation with Gradient-Weighted Normalization," ISSAC '22: Proceedings of the 2022 International Symposium on Symbolic and Algebraic Computation, pp. 225—234, 2022.

   消失イデアルの近似基底計算において，正規化は非常に重要な役割を持つ. 本研究では，従来の係数正規化に変わる新たな正規化，勾配重み付き正規化を提案し，その有効

性を示した. この正規化は, 従来見過ごされていた係数正規化の問題点を解消し, また実用上有用な性質を実現する. また同等の計算量で実現できる点でも優れる.
2. Elias Wirth, Hiroshi Kera, Sebastian Pokutta, "Approximate Vanishing Ideal Computations at Scale," ICLR' 23, 2023.
消失イデアルの近似基底計算の計算量に関する新たな解析を与えた. 従来の解析は厳密な基底計算の計算量のみを与えていた. しかし近似基底計算の場合, 実験的にはサンプル数に対して計算コストがあまり増大せず, この理論との乖離があった. 本研究では, 近似基底計算の場合の計算量を明らかにし, この乖離を埋めた.

(2)特許出願

研究期間全出願件数:0件

(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)

a. Ryosuke Masuya, Yuichi Ike, Hiroshi Kera, "Vanishing Component Analysis with Contrastive Normalization"(査読中)
b. Hiroshi Kera and Yoshihiko Hasegawa, "Monomial-agnostic computation of vanishing ideals"(査読中)
c. Elias Wirth and Hiroshi Kera, "Approximate Vanishing Ideal Computations," POEMA Final Workshop, 2022(口頭発表)