**１　研究目的、研究方法など**

様式Ｓ－１４ 研究計画調書（添付ファイル項目）

本研究計画調書は「小区分」の審査区分で審査されます。記述に当たっては、「科学研究費助成事業における審査及び評価に関する規程」（公募要領１８頁参照）を参考にすること。

本研究の目的と方法などについて、４頁以内で記述すること。

冒頭にその概要を簡潔にまとめて記述し、本文には、(1)本研究の学術的背景、研究課題の核心をなす学術的「問い」、(2)本研究の目的および学術的独自性と創造性、(3)本研究の着想に至った経緯や、関連する国内外の研究動向と本研究の位置づけ、(4)本研究で何をどのように、どこまで明らかにしようとするのか、(5)本研究の目的を達成するための準備状況、について具体的かつ明確に記述すること。

本研究を研究分担者とともに行う場合は、研究代表者、研究分担者の具体的な役割を記述すること。

**（概要）**

* データサイエンスにおいて基礎的な離散データ構造に、**順序**と**半順序**とがある。それらは**有向グラフ**に含まれる**。**
* 順序・半順序・有向グラフは、**箙**と呼ばれるある種の有向グラフとみなせる。この箙には**団代数**と呼ばれる**対称性**を持った**組み合わせ論的代数**が付随し、団代数の**変異**操作により、**変異同値な箙の集合**が定義される。
* 順序を箙としてみなしたとき、すべての順列は変異同値な箙の集合に含まれることが知られており、その変異同値な箙の集合には半順序やある特定の有向グラフが含まれる。
* 順列の集合を対象とするとき、ある順列から別の順列への**置換**は**互換**の積に分解できることから、互換は**単位的な操作**とみなすことができるが、団代数の枠組みでは、互換は**変異**と呼ばれる団代数上の単位的操作を複数回、繰り返すことで実現される。
* このことから、順列に対応する箙を含む箙の集合であって、半順序・有向グラフを含む変異同値な箙の集合に付随する団代数の操作は、互換の分解をその部分として含むなんらかの**「拡張された操作」**としての性質を持つと考えられる。
* **本研究で明らかにすること**
  + データ構造が団代数においては整数要素**交代行列**で表されることに着目し、交代行列の**線形代数操作**を「拡張された操作」とし、その操作の繰り返し(積)が行と列の互換・置換を実現する線形代数操作となる。この**線形代数操作の特徴**を明らかにする。
  + また、交代行列は**特殊直交群(回転群)**を**Lie群**としたときの**Lie代数**をなす性質を利用して、その代数的特徴についても検討する。これらの検討を通じて、順序を団代数によって拡張することによって得られる、順序・半順序・有向グラフからなる**新たな離散データ構造の存在と性質**を明らかにする。

**（本文）**

(1)本研究の学術的背景、研究課題の核心をなす学術的「問い」

本研究の背景には複数の概念がある。課題の核心をなす「問い」は、これらの概念を結びつける方法を見出すことである。

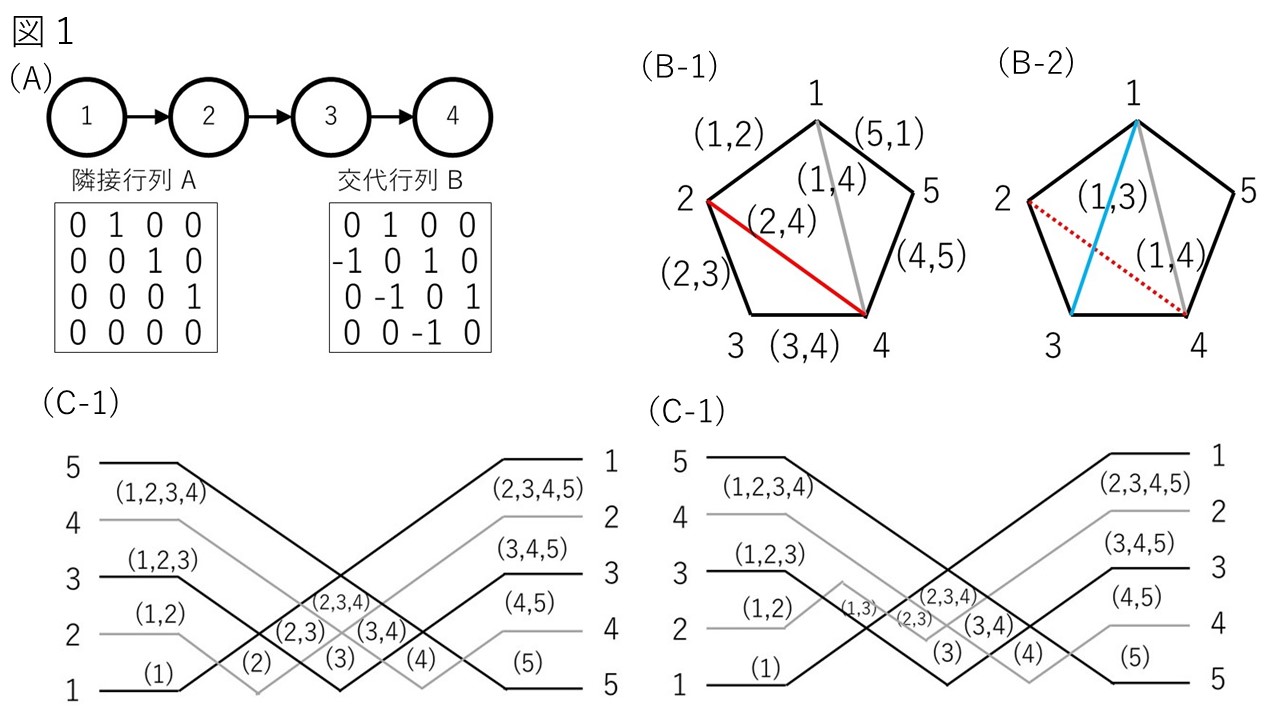
以下に、その背景となる概念を順番に述べ、最後にそれらを結びつける方向性について説明する。

1. 背景となる概念

概念リスト

* 1. 団代数と離散データ構造
  2. 箙・団代数と変異同値と置換
  3. 置換と箙変異
  4. 団代数の交代行列の変異
  5. 交代行列とLie代数・Lie群
  6. **団代数**と離散データ構造

団代数とは、Fomin とZelevinsky が2000 年頃に導入したある種の**可換代数**のクラスであり、数学の諸分野に遍在的に現れる一つの基礎的な**代数的・組み合わせ論的構造**のことである、と説明される。団代数では、**交代行列**と個の変数との組を「**種**」とし、その種が通りの変異によって変化することで、「種」による-**斉次グラフ**構造が定まる。交代行列は**箙**と呼ばれる**有向グラフ**に対応する。箙はループ(始点・終点を同じくする辺) を持たず、任意の２頂点間には、0本以上の有向辺を与えうるが、すべての辺の向きは同じであるという制約がある。また長さ2 のサイクル(2 頂点間を往復するような２辺)は存在しないと言う制約も持つ。箙の**隣接行列** に対して で定義される**交代行列**が「種」に含まれ、と個の変数とが、一定のルールで**変異**し、異なる「種（交代行列と変数のセット）」になる。図１(A)に、 の鎖状有向グラフであり(1,2,3,4)という順列に対応する箙とそれに対応する隣接行列と交代行列を示した。また、図1(B),(C)には、団代数が付随する対象の例を２つ示した。(B)は多角形の**三角化**(非交叉対角線のセットの取り方)であり、(C)は**組み紐**の配置とそれが作る領域への冪集合要素の割り当てである。(B-1)と(B-2)、(C-1)と(C-2)のそれぞれで、離散的なダイアグラムが変化しているが、それぞれのダイアグラムが「種」に対応しており、図示してはいないが、対応する箙が定義されており、ダイアグラムの変化が団代数の変異に相当する。団代数が組み合わせ論的構造であることを示す例である。

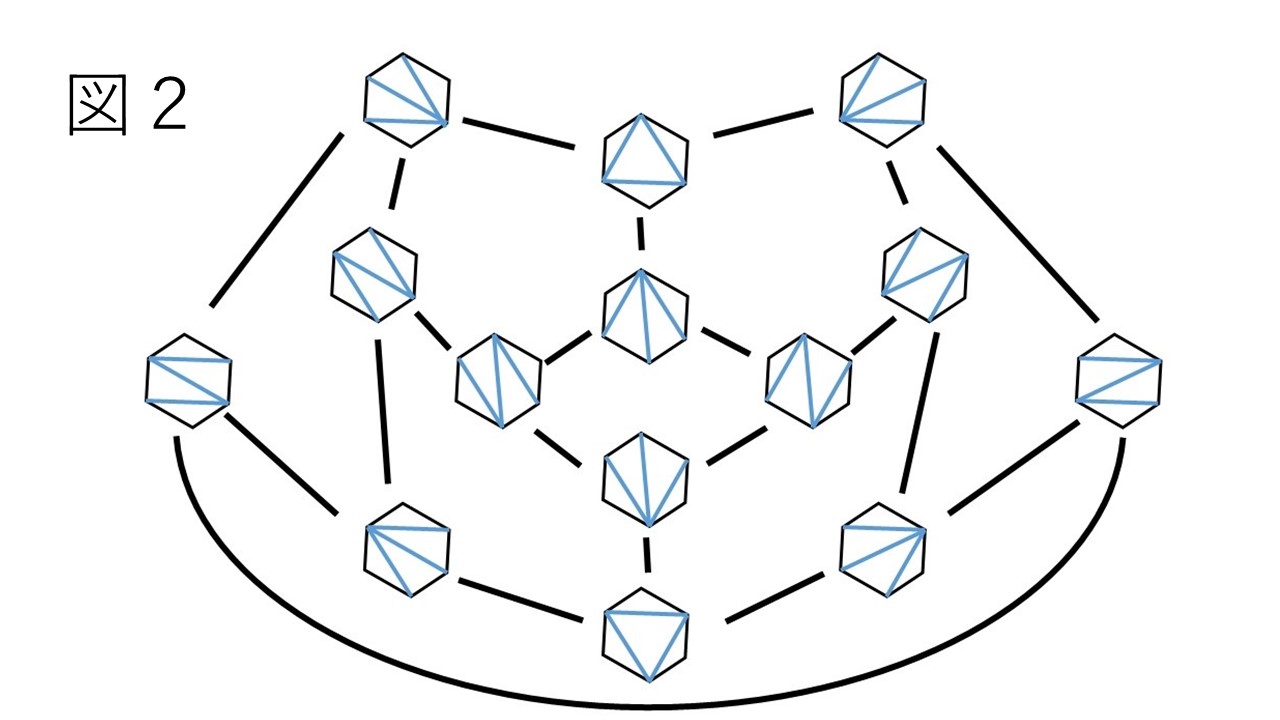


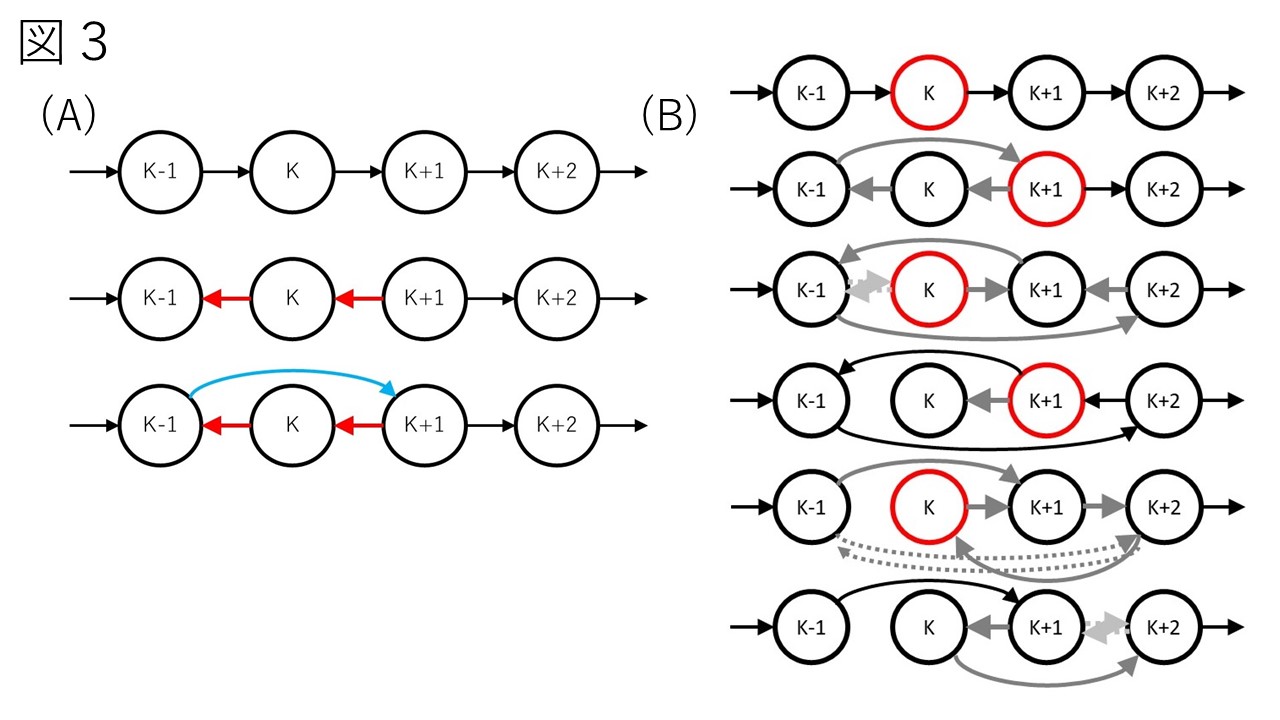
1. 箙・団代数と変異同値と置換

図1(B)(C)にダイアグラムのバリエーション(B-1),(B-2)、ならびに(C-1),(C-2)はそれぞれ団代数の「種」に相当し、相互に変異によって行き合える関係になっている。この種を集め、相互交換関係で結びつけると、図2に示すような**斉次グラフ**になる。このように変異によって行き合える「種」の間柄を変異同値と呼ぶ。団代数は変異によってパターンが変化するが、付随する交代行列の行番・列番の順列を入れ替えた「種」が**変異同値**セット内に現れることが知られている。したがって、変異同値な「種」のペアには付番が**置換**された関係のものが含まれていることがわかる。

1. 置換と箙変異

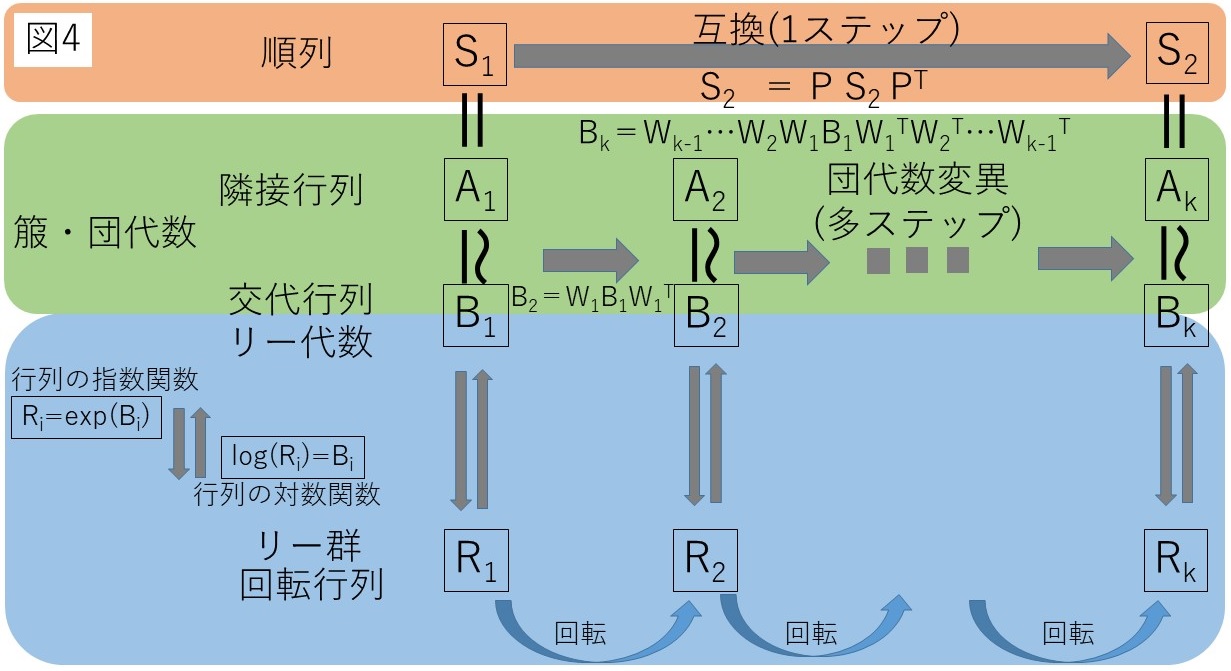
図3には、鎖状の箙における変異の様子を示している。図3(A)は、鎖状箙のある頂点における１回の変異による有向辺の変化を示している。図3(B)は、隣接2頂点を交互に変異頂点として、計5回の変異を繰り返すことによって、2頂点を入れ替えて、新たな鎖状の箙を生成する過程を示している。任意の置換は**互換**の積で表せることと併せて考えると。鎖状の箙に変異を施すことによって、任意の順列のペアに相当する箙のペアの間には、変異の列が定まることがわかる。





1. 団代数の交代行列の変異

箙という有向グラフは、その隣接行列に対して、という行列を定めるが、これは交代行列になっている。団代数の置換行列と呼ぶ。この置換行列は変異によって変化して異なる交代行列となるが、に依存する行列が存在して、 と表せることが知られている。したがって、図3(B)には5回の変異によって互換が起きる例を示しているが、それに相当する置換行列の変異は と表せる。他方、行列の行と列とに同じ置換を施すとき、置換行列を用いてと表せる。したがって、ととはに対して同じ作用を持つことがわかる。その様子を一般化して図4に示している。



1. 交代行列とLie代数・Lie群

交代行列が「**特殊直交群(回転群)**を**Lie群**とする**Lie代数**」をなすことは、**群の表現論**が教えることである。箙の変異同値集合は交代行列の部分集合になっていることは上述の通りであるから、順列を包含する変異同値な箙の集合には、Lie代数・Lie群(特殊直交群)上での考察が可能となる。その対応関係を図4に示した。

1. 学術的な「問い」：(ア)-(オ)を結びつける方法を見出すこと

　上述の通り、複数の要素間に相互関係が認められることはそれぞれ断片的に示されているが、それらのすべてを包括的に結びつけることが、本申請研究の学術的な問いである。そして、その結びつけ作業を通じて、順列を含む団代数が示す箙部分集合のデータ構造としての特性と応用可能性を探りたい。この問いに答えるにあたって、次に挙げる複数の方向性で研究を展開する予定である。

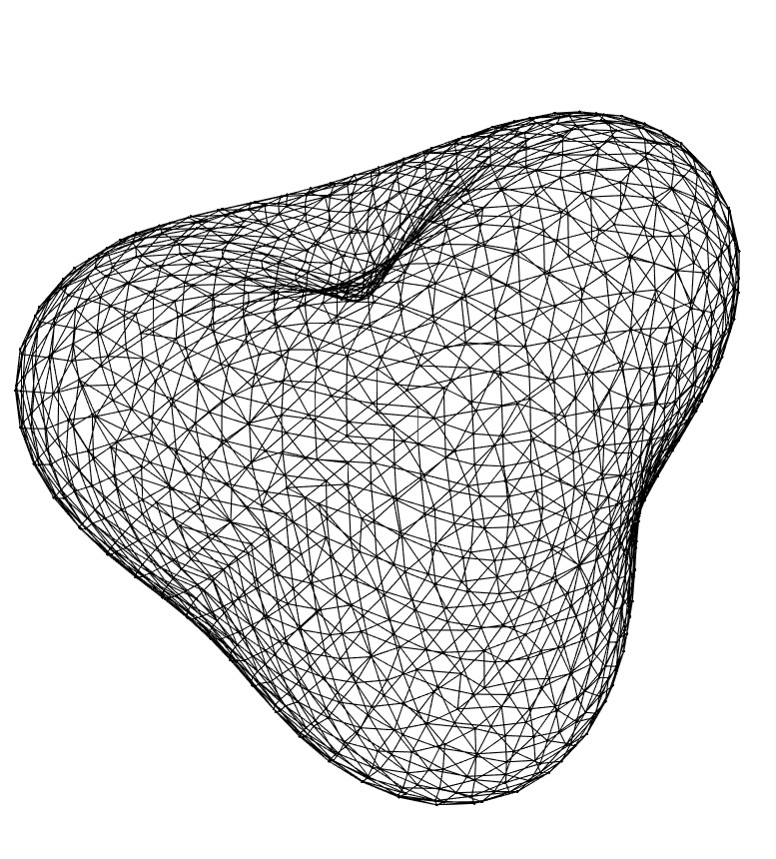
第一に、本申請研究では、**離散性の要素(**「順序・半順序」「グラフ」「整数行列(箙の隣接行列と置換行列の要素は整数である)」)と**連続性の要素(**「特殊直交群のリー群・リー代数」)とを扱うが、この相異なる性質を結びつける関係性が何かに着目している。

第二に、団代数による変異同値集合には、**有限**要素のそれと**無限**要素のそれとがあることが知られているが、それを特殊直交群であるLie群とそのLie代数とに対応付けることを目指すが、その際、特殊直交群が閉じた多様体であることから、**有限要素集合の閉空間配置**と**無限要素集合の閉空間配置**としての側面が研究の対象になると考えている。

第三に、団代数には、**全正値性**と密接な関係があり、その例に、**全正値行列**(すべての小行列式が正である行列)や**Teichmuller空間**(空間上のすべての点の座標が正値で表される空間であり**位相幾何対象**を納めることができる空間)がある。**データサイエンス**では**距離**・**確率**が重要な要素となるが、これらも正値性を有している。この正値性が要素間の相互関係の解明によって何らかのデータサイエンス上の意義を持つ可能性を探求したい。

最後の方向性は、**計算機演算**的側面に関することになる。互換・置換は**スワップ**処理と言える。**一時変数**を用いて情報を退避させる方法が一般的になっているが、**XOR-swap**のように一時変数を用いない方法も知られている。団代数変異を経由する置換がSwap処理という文脈でどのような意味を持ちうるかについても方向性の１つとして検討している。

(2)本研究の目的および学術的独自性と創造性



団代数は比較的新しいがホット。

整数行列という意味で、色々なところにつながっていることは確か。

探索空間としても、その対称性などから、探索アルゴリズムが開発されれば素敵。

但し、団代数とリー代数との関係については余りペイパーは見当たらない。

また、特定の離散データ構造の取り出しという側面も進んでいない。

(3)本研究の着想に至った経緯や、関連する国内外の研究動向と本研究の位置づけ

３次元形状の三角メッシュ的取扱い。形オブジェクト。

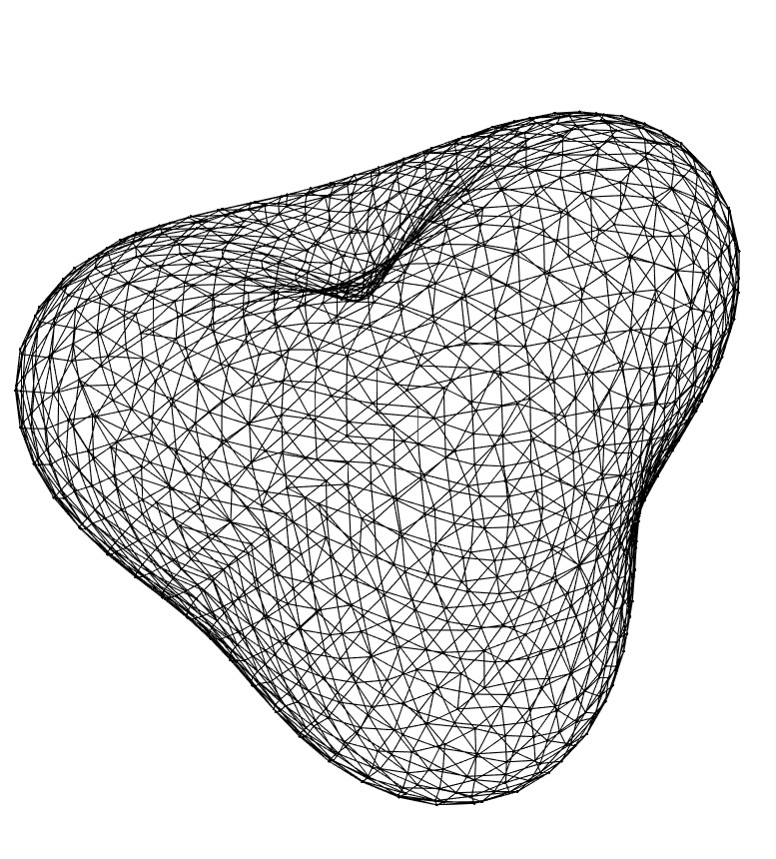
その表現として、箙・団代数・結び目があった。

(4)本研究で何をどのように、どこまで明らかにしようとするのか

(5)本研究の目的を達成するための準備状況、について具体的かつ明確に記述すること。

このくらいのディスカッションはして、このくらいの総説は書けている。

代数アプリケーションもあり、小規模では動くことは確認できている。



**２　応募者の研究遂行能力及び研究環境**

応募者（研究代表者、研究分担者）の研究計画の実行可能性を示すため、(1)これまでの研究活動、(2)研究環境（研究遂行に必要な研究施設・設備・研究資料等を含む）について２頁以内で記述すること。

「(1)これまでの研究活動」の記述には、研究活動を中断していた期間がある場合にはその説明などを含めてもよい。

1. これまでの研究活動（単名研究代表者　山田亮）

●　研究歴概略

1. 2000年度 – 2021年度

理化学研究所 研究員、京都大学医学部附属ゲノム医学センター 准教授・教授、東京大学医科学研究所ヒトゲノム解析センター 准教授、京都大学数理解析研究所 特任教授として**ゲノム解析・オミックス解析・量的生物学解析**の解析実務と手法開発を実施した。

1. 2015年度－2021年度

上記に加え、科学技術振興機構 CREST [１細胞]統合１細胞解析のための革新的技術基盤 「動く１細胞の「意思」を読み取るin vivo 網羅的動態・発現解析法の開発(石井優代表)」の主たる共同研究者として、その**数理モデル・統計解析手法**を担当した。また、並行して、[ビッグデータ基盤] ビッグデータ統合利活用のための次世代基盤技術の創出・体系化 「**離散構造統計学**の創出と癌科学への展開(津田宏冶代表)」の主たる共同研究者として、離散構造統計学の統計学的解釈とゲノム疫学分野への応用を担当した。

●　本申請課題と関係が深い研究活動

1. 離散データとしての**DNA配列の多様性**に関する手法開発研究

DNA配列情報は4種類の塩基からなる離散データ構造であり、その変異は**木構造**・**Directed Acyclic Graph構造**を取る。同データ構造を用いた統計学的検定は、**分割表検定**にて行われることが通例であり、大規模な多**重検定補正**と**組合せ検定**とが統計学的に特徴的な課題となる。この領域の成果の主なものは、末尾参照文献のが該当する[4,8-10]。

1. 確率過程を扱った研究

医療における個人の意思決定過程を集団の**Multi-armed bandit問題**の枠組みで検討し、集団の状態を**格子上の酔歩**として扱った研究は、実社会課題を離散データ構造の課題に翻訳した点で本申請研究に通じるものである[7]。また、２次元空間を運動する点集合のグローバルマッチング**最適化問題**を扱った研究[6]、**３次元**顕微鏡画像の時系列解析に**ガウシアン過程**を用いた研究も行った[2]。

1. 幾何・位相幾何・情報幾何

３次元顕微鏡画像から**データ駆動式**に**形特徴量**を抽出する研究は、上述の1細胞CRESTプロジェクトのテーマであったが、３次元立体画像を三角メッシュ化して離散対象とし、数学的に安定な**曲率フロー**・**共形変換**と球面調和関数分解を利用した解析的**直交基底分解**、**モース理論**に基づく測地距離を用いた分解、**グラフ・ラプラシアン**を用いた再帰的**グラフカット**による**位相的特徴量**抽出などを実施した[2]。

幾何的なアプローチの一環として、**情報幾何**にも取り組んだ。分布状のデータを**ノンパラメトリック**に**指数型分布族**的に扱うことで情報幾何の知見をデータマイニングに活用する試みも行った[3]。

1. 俯瞰的考察

ゲノム・オミックス・量的生物学分野の状況を、検定を中心とする統計学とデータマイニングを主眼とする機械学習との両面に渡って、現状を俯瞰し、今後の方向性について考察した[5]。また、本申請課題の中心となる団代数とそのデータサイエンスとの関りについても、総説にまとめている[1]。

1. 研究環境

京都大学医学部附属ゲノム医学センターは３分野体制で、大規模ゲノムオミックスデータと地域コホートを持ち、その解析環境として、小規模なインハウス計算機環境＋京都大学スパコン＋東京大学計算機環境(外部向けサービス)＋アマゾンウェブサービスがあり、その計算機リソースの利活用はセンター雇用のスタッフにより快適に提供されている。

情報学関係では、２つのＣＲＥＳＴを通じた複数の研究者と定期的に情報交換をしており、また、京都大学学内の情報学研究科・理学研究科数学科の教員らとの交流があり、フランクなディスカッションを行う環境がある。

●　参照文献(発表年降順)

[1] 山田亮,木下瞬,和田みのり,三村和史,杉山磨人 三角化曲面に付随する団代数に対応する置換の変異の可能性について人工知能36(4) 461-469 (2021)

[2] Yusri Dwi Heryanto, Chen, CY., Uchida, Y., Mimura, K., Ishii, M., Yamada, R., Integrated Analysis of Cell Shape and Movement in Moving Frame. BiolOpen (2021)doi: <https://doi.org/10.1242/bio.058512>

[3] Okada, D., Yamada, R. Decomposition of a set of distributions in extended exponential family form for distinguishing multiple oligo-dimensional marker expression profiles of single-cell populations and visualizing their dynamics. (2020) PLoS ONE, 15 (4), art. no. e0231250, DOI: 10.1371/journal.pone.0231250

[4] Basak, T., Nagashima, K., Kajimoto, S., Kawaguchi, T., Tabara, Y., Matsuda, F., Yamada, R. A Geometry-Based Multiple Testing Correction for Contingency Tables by Truncated Normal Distribution (2020) Statistics in Biosciences, 12 (1), pp. 63-77. DOI: 10.1007/s12561-020-09271-6

[5] Yamada, R., Okada, D., Wang, J., Basak, T., Koyama, S. Interpretation of omics data analyses(2020) Journal of Human Genetics, . DOI: 10.1038/s10038-020-0763-5

[6] Okada, D., Nakamura, N., Wada, T., Iwasaki, A., Yamada, R. Extension of sinkhorn method: Optimal movement estimation of agents moving at constant velocity. (2019) Transactions of the Japanese Society for Artificial Intelligence, 34 (5), art. no. D-J13\_1-7 . DOI: 10.1527/tjsai.D-J13

[7] Wang, J., Yamada, R. In silico study of medical decision-making for rare diseases: Heterogeneity of decision-makers in a population improves overall benefit. (2018) PeerJ, 2018 (9), art. no. e5677, . DOI: 10.7717/peerj.5677

[8] Terada, A., Yamada, R., Tsuda, K., Sese, J. LAMPLINK: Detection of statistically significant SNP combinations from GWAS data. (2016) Bioinformatics, 32 (22), pp. 3513-3515. DOI: 10.1093/bioinformatics/btw418

[9] Narahara, M., Tamaki, K., Yamada, R. Application of permanents of square matrices for DNA identification in multiple-fatality cases. (2013) BMC Genetics, 14, art. no. 72, . DOI: 10.1186/1471-2156-14-72

[10] Yamada, R., Okada, Y. An optimal dose-effect mode trend test for SNP genotype tables. (2009) Genetic Epidemiology, 33 (2), pp. 114-127. DOI: 10.1002/gepi.20362

**３**　**人権の保護及び法令等の遵守への対応（公募要領４頁参照）**

本研究を遂行するに当たって、相手方の同意・協力を必要とする研究、個人情報の取扱いの配慮を必要とする研究、生命倫理・安全対策に対する取組を必要とする研究など指針・法令等（国際共同研究を行う国・地域の指針・法令等を含む）に基づく手続が必要な研究が含まれている場合、講じる対策と措置を、１頁以内で記述すること。

個人情報を伴うアンケート調査・インタビュー調査・行動調査（個人履歴・映像を含む）、提供を受けた試料の使用、ヒト遺伝子解析研究、遺伝子組換え実験、動物実験など、研究機関内外の倫理委員会等における承認手続が必要となる調査・研究・実験などが対象となります。

該当しない場合には、その旨記述すること。

該当しない。

**４　研究計画最終年度前年度応募を行う場合の記述事項（該当者は必ず記述すること（公募要領２６頁参照））**

本研究の研究代表者が行っている、令和４(2022)年度が最終年度に当たる継続研究課題の当初研究計画、その研究によって得られた新たな知見等の研究成果を記述するとともに、当該研究の進展を踏まえ、本研究を前年度応募する理由（研究の展開状況、経費の必要性等）を１頁以内で記述すること。

該当しない場合は記述欄を削除することなく、空欄のまま提出すること。

|  |  |  |  |
| --- | --- | --- | --- |
| **研究種目名** | **課題番号** | **研究課題名** | **研究期間** |
|  |  |  | 平成　年度～令和４年度 |

**当初研究計画及び研究成果**

**前年度応募する理由**