様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科研費 KAKENHI

科学研究費助成事業　　研究成果報告書

平成 ３０ 年　 ６ 月　 ５ 日現在

機関番号：１２６０１
研究種目：基盤研究(C)（一般）
研究期間：2014～2017
課題番号：２６３３００３５

研究課題名（和文）多次元尺度構成法における効率的点配置構成法

研究課題名（英文）On efficient configuration in multidimensional scaling

研究代表者
倉田　博史（Kurata, Hiroshi）

東京大学・大学院情報学環・教授

研究者番号：５０２８４２３７

交付決定額（研究期間全体）：（直接経費）　　3,700,000円

研究成果の概要（和文）：　多次元尺度構成法における効率的な点配置の構成法について明らかにするため、ユークリッド２乗距離行列（Euclidean distance matrix、以下EDMと略）の性質について詳細に研究した。まず「セル行列（cell matrix）」という非常に単純な構造を持ちつつ応用が豊富なEDMのクラスに注目し、その性質を明らかにした。続いて、所与のEDMに対し、それに最も近いセル行列を与える式を導いた。また、EDMの逆行列が持つ情報を抽出するため、そのMoore-Penrose一般逆行列の表現を導出した。また、EDMの積を定義し、その性質を応用について議論した。

研究成果の概要（英文）：In this research, I discussed the problem of constructing efficient configuration in multidimensional scaling. In particular, I studied some properties of Euclidean distance matrices (EDMs) in detail. First I focused on a class of cell matrices, which is a subset of EDMs and has a very simple strucuture. Since the class has various potential applications to real data analysis, I studied its properties from geometric and linear algebraic points of view. Next I derived a general expression for the EDM closest to a given cell matrix and the Moore-Penrose inverse of an EDM. Both results give a theoretic basis for efficient configuration in multidimnsional scaling. I also defined a prodct on the set of EDMs and studied its properties.

研究分野：統計科学

キーワード：多次元尺度構成法　ユークリッド距離行列　一般逆行列　行列の順序

１．研究開始当初の背景

多次元尺度構成法（multidimensional scaling、以下 MDS とも表す）はデータの視覚的表現に優れている一方で、点配置の構成や解釈が分析者の主観や選択した近似方法に依存するという短所がある。この問題は、点配置の経年変化を調べる場合などのように、多数の点配置を比較しなければならないというような状況ではより大きなものとなり得る。

この視点に立って、倉田は本研究開始以前の一連の論文 (例えば、Kurata and Sakuma (2007)、Kurata and Tarazaga (2010, 1012)、Kurata and Matsuura (2013)、Tarazaga and Kurata (2013)などにおいて非類似度行列よりも純粋な概念であるユークリッド 2 乗距離行列（Euclidean squared distance matrix）に焦点を当て、その数理的構造を順序構造を中心に研究してきた。なお、ここにユークリッド 2 乗距離行列とは、非類似度行列であって、かつ適当な次元のユークリッド空間の点配置（の 2 乗距離）として表現可能なものを指す。この概念は、統計科学における L2 埋込問題、分子生物学における分子構造の記述を始め、認知科学、マーケティングサイエンス。計算化学など多くの分野に幅広く応用されており、Linear Algebra and its Applications 誌を中心に多くの研究者によって研究されている。上記の論文も 5 本のうち 4 本は同誌に掲載されており、ユークリッド距離行列の理論研究として一定の評価を得ている。

２．研究の目的

本研究ではこれらの数学的研究を発展させ、多次元尺度更生法において最も基本的かつ重要な問いである点配置の構成問題を扱う。具体的には、

（１） 点配置の順序付けと上界・下界の利用、
（２） 次元縮約により損失した情報を回復する方法
（３） 非類似度行列の逆行列の表現
（４） 上記の３つの問題のダイナミック MDS への応用

の４つの問題を考察することを通して、より効率的な点配置の構成法を得ることを狙う。以下それぞれについて具体的に述べる。

３．研究の方法

点配置を構成するプロセスは、所与の非類似度行列 $\boldsymbol{S}$ に最も「近い」ユークリッド 2 乗距離行列 $\boldsymbol{D}$ を求め、$\boldsymbol{D}$ から点配置を計算するという手続きからなる。このプロセスは数学的には非類似度行列 $\boldsymbol{S}$ を「ユークリッド距離行列の全体からなる集である合」への射影（フロベニウス・ノルムに関する正射影）を求めることと言い換えることが出来る。ここに、「ユークリッド距離行列の全体からなる集合」は「対角要素が全て零であるような対称行列の全体からなる線形空間」の閉凸錐である。従って、最小点即ち最良の近似を与えるユークリッド 2 乗距離行列 $\boldsymbol{D}$ が一意に存在する。しかし、多くの応用研究で $\boldsymbol{D}$ が非類似度行列 $\boldsymbol{S}$ の持つ特徴を保っていないことがしばしばあることが指摘されている。この問題への対処として、正射影ではなく、非類似度行列 $\boldsymbol{S}$ の各要素 s_{ij} に適当な定数 c を加えることによって得られるユークリッド距離行列 $D=(s_{ij}+c)$ を求め、これに基づく点配置を用いるという手法 (additive constant technique)が広く利用されている。この手法は、数学的には、正射影の代わりに「全ての点が互いに等距離にあるようなユークリッド 2 乗距離行列がなす半直線」に沿った射影を用いることに対応する。しかし、いずれの射影を用いるにせよ、結果として得られたユークリッド 2 乗距離行列が $\boldsymbol{S}$ の特徴を保持している保証はなく、また保持しているか否かの確認も分析者の主観によってなされる。このことは多次元尺度構成法に基づく分析結果の客観性を減ずるものであり、改善を要する問題と思われる。

そこで本研究では、ユークリッド 2 乗距離行列の順序構造や固有値の散らばり、逆行列の情報などに注目することにより、これらの問題について新しい見方を提供しようとするものである。

４．研究成果

Kurata and Tarazaga (2015)では、所与のユークリッド 2 乗行列に対し、それにフロベニウス・ノルムの意味で最も近いセル行列を導出した。ここに、セル行列 $C=(c_{ij})$ とは、ある $n\times 1$ ベクトル $a=(a_1,\cdots,\ a_n)$ を用いて、その第 (i,j) 要素が $c_{ij}=a_i+a_j$（$i\neq j$），$c_{ii}=0(i=1,\cdots,n)$ と表現されるような非負行列であり、グラフの距離行列としては星グラフに対応するものである。セル行列の全体はユークリッド 2 乗距離行列の部分錐をなし、それが生成する部分空間は先行研究でも議論されている。本論文では、所与のユークリッド 2 乗距離行列に対し、それに最も近いセル行列を導出することでその構造をより簡潔に捉えることを狙うものである。特に、セル行列の固有値の散らばりはベクトル a の散らばりと密接な関係を持つため、点配置の形状をこれらから把握することができる可能性がある。本研究はそのための基礎を与えるものである。具体的には、２つのセル行列の固有値の間にマジョライゼーションの順序関係があるときには、対応する２つのベクトル a 同士にもマジョライゼーションの順序関係があることを明らかにした。これにより、セル行列の散らばり具合の順序付けが得られ、その情報はユークリッド 2 乗距離行列の順序付けにも応用可能と考えられる。これについては今後の課題である。

Kurata and Bapat (2015)ではユークリッド 2 乗距離行列の Moore-Penrose 一般逆行列

の一つの表現を導いた。ユークリッド2乗距離行列は spherical と nonspherical に分類される。ユークリッド 2 乗距離行列が spherical であるとは、その点配置がある超球面上に存在することである。Spherical でないユークリッド 2 乗距離行列は nonspherical と呼ばれる。ユークリッド 2 乗距離行列が spherical であるための必要十分条件はこの分野の中心的テーマの一つであり、長く議論されてきた問題である。ユークリッド2乗距離行列の一般逆行列もこれに関連しており、例えば、spherical な場合の点配置の乗る超球面の半径は一般逆行列で表現される。本研究では Moore-Penrose 一般逆行列を、ユークリッド距離行列の Young-Householder-Gower 変換のみで表現するという結果を導いた。Young-Householder-Gower 変換はユークリッド2乗距離行列と1対1に対応する非負値定符号行列であり、このランク分解から点配置が作られる。従って、本研究の結果には、点配置に関する情報と一般逆行列の構造との関係がより明確になるという長所がある。

Kurata and Bapat (2016)は上記の Kurata and Bapat (2015)の拡張であり、2015 年の論文で導出した Moore-Penrose 一般逆行列の表現を hollow 行列すなわち対角要素が全て零であるような対称行列に対して拡張したものである。これにより、ユークリッド2乗距離行列だけでなく、(必ずしも点配置を持たない) 一般の非類似度行列の一般逆行列も扱うことが出来ることとなった。

よく知られる通り、一般逆行列は行列の集合上に順序（半順序）を誘導する。最も基礎的なものは「マイナス半順序」であり、これは $\boldsymbol{A}$ と $\boldsymbol{B}$ を n×m 複素行列とすると、$\boldsymbol{A} \leq^{-} \boldsymbol{B}$ なる不等号を「ある一般逆行列 $\boldsymbol{A}^{-}$ に対して、$\boldsymbol{A}^{\wedge}(\boldsymbol{B}-\boldsymbol{A})=\boldsymbol{0}$ かつ $(\boldsymbol{B}-\boldsymbol{A})\boldsymbol{A}^{-}=0$ が成り立つこと」と定義するものである。この考え方に基づいて、Moore-Penrose 一般逆行列が定める「スター半順序」、group inverse （定まった和訳なし）が定める「シャープ半順序」、core inverse（定まった和訳なし）が定める「core 半順序」などが定義され、広く研究されている。

Kurata (2018)では core inverse と core 半順序に注目し、幾つかの数学的事実を導いた。具体的には、core inverse の理論的研究として広く引用される Baksalary and Trenkler (2010, Linear Multilinear Algebra) と Malik (2013, Applied Mathematics and Computation)によって導出された幾つかの公式に注目し、それらのそれぞれが成り立つ行列の集合として最大のものを導出した。

これらは多次元尺度構成法の実際の分析手順に影響を与えるものではないが、点配置を与えるランク分解の行列に順序構造を求める際に応用可能と考えられる。

Bapat and Kurata (2018, 投稿中)は２つの（次元が異っていてもよい）ユークリッド2乗距離行列 $\boldsymbol{A}$、$\boldsymbol{B}$ に対し、そのデカルト積 $\boldsymbol{A} \times \boldsymbol{B}$ を（論文では異なる記号を用いているが本稿では×で表す）を $\boldsymbol{J} \otimes \boldsymbol{A} + \boldsymbol{B} \otimes \boldsymbol{J}$ と定義し(ここに⊗はクロネッカー積、$\boldsymbol{J}$ は全ての要素が１であるような正方行列)、その性質とグラフ論への応用を論じた。この結果は現時点では数学の範囲にとどまっているが、$\boldsymbol{u}$ を適当な次元の縦ベクトルとし、$\boldsymbol{J}$ という行列を $\boldsymbol{uu}'$ に置き換える（$\boldsymbol{u}'$ は $\boldsymbol{u}$ の転置を表す）などの操作でより広いユークリッド2乗距離行列のクラスを得ることが出来、実際の分析にも応用することが出来る。この点については今後論文の形で発表する予定である。

また、本研究の周辺部分の研究として、Matsuura, Kurata and Tarpey (2015)は多次元分布のプリンシパル・ポイントの推定量を提案するものであり、それを通して多次元分布を少数の点で近似する方法についての数学的基礎を提供するものである。Kurata and Matsuura (2016)は見かけ上無関係な回帰（seemingly unrelated regression, SUR）モデルにおいて相関係数行列が既知であるときの回帰係数の最良共変推定量を導出している。Matsuura and Kurata (submitted) は同じモデルの下で誤差項の共分散行列の最良共変推定量を導出している。Tsukuda and Kurata (to appear)は誤差項の分散共分散行列が単位行列の定数倍とは限らない場合の一般線形回帰モデルにおいて、２つの一般化リッジ推定量が恒等的に等しくなるための必要十分条件を誤差項の分散共分散行列の構造に対する条件として導出した。

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 9 件）

[1] Shun Matsuura and <u>Hiroshi Kurata</u>, Covariance estimator for seemingly unrelated regression model with known correlation coefficient matrix (submitted for publication)

[2] Ravindra B. Bapat and <u>Hiroshi Kurata</u>, On Cartesian product of Euclidean distance matrices (submitted for publication)

[3] Koji Tsukuda and <u>Hiroshi Kurata</u>, Covariance structure associated with an equality between general ridge estinmators, *Statistical Papers* (to appear), 査読有

[4] <u>Hiroshi Kurata</u>, Some theorems on the core inverse of matrices and the core partial ordering, *Applied Mathematics and*

Computation, 316 (2018), 43-51.査読有, DOI:https://doi.org/10.1016/j.amc.2017.07.082

[5] Hiroshi Kurata and Shun Matsuura, Best equivariant estimator of regression coefficients in a seemingly unrelated regression model with known correlation matrix, *Annals of the Institute of Statistical Mathematics*, 68 (2016), no. 4, 705–723. 査読有, DOI 10.1007/s10463-015-0512-2

[6] Hiroshi Kurata and Ravindra B. Bapat, Moore-Penrose inverses of a hollow symmetric matrix and a predistance matrix, *Special Matrices*, 4 (2016), 270-282. 査読有, DOI 10.1515/spma-2016-0028

[7] Shun Matsuura, Hiroshi Kurata and Thaddeus Tarpey, Optimal estimators of principal points for minimizing expected mean squared distance, *Journal of Statistical Planning and Inference*, 167 (2015), 102-122. 査読有, DOI:http://dx.doi.org/10.1016/j.jspi.2015.05.005

[8] Hiroshi Kurata and Pablo Tarazaga, The cell matrix closest to a given Euclidean distance matrix, *Linear Algebra and its Applications*, 485 (2015), 194-207.査読有, http://dx.doi.org/10.1016/j.laa.2015.07.030

[9] Hiroshi Kurata and Ravindra B. Bapat, Moore-Penrose inverse of a Euclidean distance matrix, *Linear Algebra and its Applications*, 472 (2015), 106-117.査読有, DOI:http://dx.doi.org/10.1016/j.laa.2015.01.032

〔学会発表〕（計 3 件）

① 佃康司・倉田博史、二つの一般リッジ推定量を等しくする共分散構造、日本数学会、2017 年
② 松浦峻・倉田博史、SUR モデルにおける相関係数が既知のときの分散共分散行列の最良共変推定について、統計関連連合大会、2017 年
③ Hiroshi Kurata, Some theorems on the core inverse of matrices and the core partial ordering, International Linear Algebra Society, 2017 年

〔図書〕（計１件）

倉田博史 KADOKAWA『大学 4 年間の統計学が 10 時間でざっと学べる』2017 年 221 頁

〔産業財産権〕

○出願状況（計０件）

○取得状況（計０件）

〔その他〕

ホームページ等

http://www.geocities.jp/h_kurata_statistics/

６．研究組織

(1)研究代表者

倉田　博史（KURATA, Hiroshi）

東京大学・大学院情報学環・教授

研究者番号：50284237

(2)研究分担者　なし

研究者番号：

(3)連携研究者　なし

研究者番号：

(4)研究協力者　なし