

氏名（本籍） 佐々木 亮 平（宮城県）  
学位の種類 博士（工学）  
学位記番号 甲第1042号  
学位授与の日付 2020年3月17日  
学位授与の要件 学位規則第4条第1項該当  
学位論文題目 局所線形空間の低次元化に基づく行列補完法  
と各種応用に関する研究

論文審査委員 （主査）教授 古川 利博  
教授 八嶋 弘幸 教授 谷口 行信  
教授 赤倉 貴子 准教授 池辺 淑子  
教授 相川 直幸  
法政大学 情報科学部 デジタルメディア学科 小西 克巳

## 論文内容の要旨

本論文では行列補完問題を扱う。行列補完問題とは、観測要素と未観測要素が混在する行列に対して未観測要素を推定する問題である。観測データから未観測データを推定する一般の問題において、推定変数を行列という構造として定義できるのであれば行列補完問題を設計できるため、その単純さを理由に応用の汎用性に富む。例として、推薦システムにおける協調フィルタリング、無線通信におけるチャンネル推定、マルチメディア信号処理においては画像・音声修復などが挙げられる。

近年では同問題に対して、対象の行列が低ランクであるとの仮定に基づいた行列の低ランク近似による行列補完手法が盛んに研究されている。同手法は、対象とする行列の各列ベクトル（または行ベクトル）が低次元線形空間に属すると仮定し、その線形空間を構成する基底と行列の部分欠損値を同時に推定する手法である。しかしながら、現在までに提案されている手法のほとんどは、推定対象の行列の列空間が低次元線形空間であるという仮定に基づいており、同仮定が成り立たない場合、すなわち、推定対象の行列が高ランクとなる場合には推定精度が劣化しやすい。同仮定が成り立たない場合の重要な例としては線型部分空間の和集合(Union of Linear Subspaces : ULS)と可微分多様体(Differentiable Manifold : DM)が挙げられる。ULSの場合、一つ一つの線型部分空間が低次元であっても、その和集合をすべて含む線形空間の次元は最大で各線型部分空間の次元の和となり、推定対象の行列は高ランク行列となる。また、機械学習におけるデータ分析の一つとして、実用として用いられるデータの多くは高次元線形空間上の低次元多様体に属するという多

様体仮説に基づく様々な分析手法が提案されており，同応用における行列補完問題に関しても推定対象の行列は高ランクとなる．このため，ULS と DM のいずれの場合にも，推定対象の行列が高ランク行列となりやすく，既存の行列の低ランク近似に基づく手法は適切でない．

本論文は上述の行列補完問題における欠損推定精度の向上を目的とする．推定対象の行列の各列ベクトルが ULS に属する場合と DM に属する場合のそれぞれについて最適な行列補完アルゴリズムを提案し，各種実データを用いた数値実験により有効性を示す．

第 2 章では，行列補完問題，ならびに，既存手法である低ランク行列近似による行列補完手法について述べる．既存手法は大きく分けて 2 つあり，1 つは行列ランク最小化，もう 1 つは行列因子分解による行列補完手法である．本論文ではそれぞれにおける具体的な解法を含めて解説する．

第 3 章では，ULS 上の部分欠損推定手法について述べる．ULS 上の代表的なデータ分析手法として，ULS を構成する各低次元線型部分空間を分割する Subspace Clustering (SC) と呼ばれる手法が提案されている．同手法は与えられた行列の各列ベクトルが ULS に属するという仮定の下，どの列がどの線型部分空間に属するかを推定する手法である．このとき，SC は与えられた行列を複数の低ランク部分行列に分割する問題を解く手法であるとみなせる．この考えに基づき，行列ランク最小化を用いた部分行列分割手法が提案されている．同手法は複数の部分行列のランク総和を最小にする問題を解くことで達成される．そこで本論文では，同最小化問題に基づき，部分行列分割と欠損推定を交互に行うアルゴリズムを提案する．提案アルゴリズムの有効性は，音声のクリッピングノイズ除去応用に関する数値実験の定量評価により示される．

第 4 章では，DM 上の部分欠損推定手法について述べる．多様体仮説に基づく，多様体学習と呼ばれるデータ分析の 1 つとして，局所線形埋め込み(Locally Linear Embedding : LLE)と呼ばれる手法が提案されている．同手法は与えられた行列の各列ベクトルに関してより小さな次元での埋め込み表現を求めることを目的とし，高次元空間の基底に関する結合係数を求める手法である．同手法は各列ベクトルがその近傍ベクトル(ユークリッド距離の小さいものが少数選択される)の線形結合で近似されるという仮定の下，近似誤差二乗和の最小化により各列ベクトルに関する結合係数を求める手法である．本論文では LLE の考えに基づき，各列ベクトルとその近傍ベクトルで為す部分行列が低ランク近似できると仮定する．その後，各列ベクトルごとに生成されるすべての部分行列に関して，その部分行列に低ランク近似される行列のランク総和を最小化する問題を設計する．同問題の解法として，各列ベクトルごとに近傍ベクトルを推定し，生成される部分行列について行列ランク最小化を逐次的に行うアルゴリズムを提案する．提案アルゴリズムの有効性は，画像インペイント応用に関する数値実験の定量評価により示される．

第 5 章では，DM 上の欠損推定問題のうち，対象とする行列サイズが大きい場合に有効な手法について述べる．同問題の代表例として推薦システムにおける協調フィルタリングが挙げられる．推薦システムとは Amazon や Netflix など用いられる，各ユーザに対してユーザの好むアイテムを推薦するシステムを示し，この方法論である協調フィルタリング

は一般に、ユーザ・アイテムごとの嗜好評価値を蓄え、疎な行列の補完問題として定式化し、各ユーザの未知の嗜好評価値を行列補完により推定することで達成される。同行列補完問題においても低ランク行列近似に基づくアルゴリズムがいくつか提案されている。推薦システムの実応用上において、推定対象の行列は一般に規模の大きな行列となる。このとき、行列ランク最小化に基づく低ランク行列近似アルゴリズムのほとんどは行列の特異値分解を繰り返し用いる手法であり、大規模な行列に対しては計算時間が大幅に増大することとなる。そのため、協調フィルタリングにおいては行列因子分解に基づく低ランク行列近似による行列補完アルゴリズムが多く提案されている。行列因子分解とは、行列  $X$  を  $X = P^T Q$  として  $P$  と  $Q$  の2つの行列に分解するモデル化の手法を示す。行列因子分解に基づく行列補完は、推定対象の行列を  $P^T Q$  とし、観測要素との誤差二乗和を最小化する問題へと定式化され、 $P$  と  $Q$  に関する交互最小二乗法により求めた  $P$  と  $Q$  を用いて未知の要素を推定する手法である。同手法は予め最大行列ランク( $P$  と  $Q$  の行数)を与えなければならないが、 $P$  と  $Q$  の要素数は与えた最大行列ランクに依存するため、予め低い最大行列ランクを与えた場合には行列ランク最小化と比べて大幅に計算時間を削減できる。しかしながら、同行列因子分解に基づく行列補完アルゴリズムのほとんどは低ランク行列を仮定しているが、実データである行列が低ランク近似できる保証はない。そこで本論文では上述した多様体仮説に基づき、埋め込み空間が低次元ながらも高次元空間上の点を表現できる、多様体学習に基づいた行列補完手法により推定精度を向上させることを考える。提案手法は  $P$  と  $Q$  の各列ベクトルが低次元 DM に属するという制約条件の下で行列因子分解を行う手法である。提案アルゴリズムの有効性は、協調フィルタリング問題の代表例である MovieLens データセットを用いた数値実験の定量評価により示される。

第 6 章では、本論文で述べた研究成果に関する総括と今後の課題、展望について簡単に触れる。

## 論文審査の結果の要旨

本論文では、学長からの審査の付託を受けて、標記 ~~6~~<sup>7</sup> 名の審査委員で構成する審査委員会を組織し、提出された学位論文について審査を行った。

審査委員会では、学位申請者より学位論文の内容、あるいは前回審査における指摘事項に対する対応結果について説明し、その後、質疑応答を実施すると、博士論文として満たすべき条件や必要な修正点を確認する、という形式で進めた。

本論文は、行列の一部の観測要素から残りの要素を推定する行列補完問題を扱っている。行列補完問題は、観測データから未観測データを予測する一般の問題のうち、推定変数を行列として定義可能であれば行列補完問題を定式化できるため、その単純さを理由に応用の汎用性に富む。行列補完問題に対して、推定変数行列の各列ベクトルが低次元線形空間上に存在すると仮定し、行列の低ランク近似による行列補完手法が現在まで

盛んに研究されている。同手法は、主推定対象の行列が低ランク近似可能な場合には高い推定精度を持つ一方で、各列ベクトルが属する集合を包含する線形空間の次元が高い場合には推定精度が大幅に劣化する。そのような集合のうち重要なものとして、低次元線型部分空間の和集合(Union of Subspaces : UoS)と低次元可微分多様体(Differentiable Manifold : DM)が挙げられ、近年マルチメディア応用などにおいて盛んに研究されている。そこで本研究では、各列ベクトルが UoS や DM に属する行列に対し、推定変数行列の部分行列を低ランク行列に近似することで欠損推定を行う手法を提案している。

以下、第 2 章では、行列補完問題、ならびに、従来手法における数理を整理する。既存手法は大きく分けて、行列ランク最小化と行列因子分解に基づく手法があり、本論文ではそれぞれの利点と欠点を説明する。

第 3 章では、各列ベクトルが UoS に属する場合の行列補完について述べる。本章では、UoS を構成する複数の部分空間を推定することと、それら部分空間に属する列ベクトルで構成する部分行列に対して低ランク行列に近似することを交互に行うことで欠損要素を推定する手法を提案している。これは与えられた行列の各列ベクトルがどの部分空間に属するかを推定する手法である Subspace Clustering に基づいており、部分行列のランク総和最小化問題として定式化される。提案手法は音声のクリッピング除去に応用され、従来の低ランク行列近似に基づく手法と比べ高い修復精度を持つことを確認している。

第 4 章では、各列ベクトルが DM に属する場合の行列補完について述べる。本章では、多様体学習の一つである Locally Linear Embedding に基づき、多様体上に属する各列ベクトルに対して局所近傍ベクトルを推定することと、その近傍ベクトルを列ベクトルに持つ部分行列を低ランク行列に近似することを交互に行うことで欠損要素を推定する手法を提案している。提案手法は上述の手法と同様に、推定変数行列のすべての列ベクトル毎に部分行列を生成し、それら部分行列のランク総和最小化問題として定式化される。提案手法は画像の欠損修復に応用され、従来の低ランク行列近似に基づく手法と比べ高い修復精度を持つことを確認している。

第 5 章では、第 4 章で扱う DM 上の行列補完問題について、行列因子分解に基づく行列補完を提案する。行列因子分解とは推定変数行列  $X$  を  $X = P^T Q$  として  $P$  と  $Q$  の 2 つの行列を低ランク行列となるように推定することで変数行列  $X$  を推定する手法である。このとき、 $P$  と  $Q$  の行数が小さい程、行列推定に要する計算時間は短くなり、特に、大規模行列の補完に対して有効な手法である。この行列因子分解に基づき、提案手法では上述の DM 上の欠損推定手法を  $P$  と  $Q$  それぞれに適用することで、変数行列  $X$  の各列ベクトルが多様体に属している場合に高い推定精度で行列補完を行うことが可能となる手法である。

以上により、本論文は、博士（工学）の学位論文として十分に価値あるものと認められる。