

# 関数データに対する正則化部分空間クラスタリング

大阪大学 大学院基礎工学研究科, 理化学研究所 革新知能統合研究センター 寺田 吉彦  
岡山大学 大学院環境生命科学研究科, 理化学研究所 革新知能統合研究センター 山本 倫生

近年, 計測技術の発展により, データが連続的に観測されることが多くなっている. 例えば, 計量化学分野では, 近赤外線分光法 (NIR) により各対象に対して, 図 1 のような連続的に変化するスペクトラルデータ (Kalivas, 1997) が得られる. この様なデータは, 変数の並びに意味があり, 従来の変数解析を考えると適切ではない. これらのデータをある区間や領域上のランダムな関数の実現値として捉えたデータ解析は, 関数データ解析 (FDA) と呼ばれている.

本稿では, 連続的に変化するデータからその背後にある潜在的なクラスター (グループ) 構造を推定する問題を考える. 最も単純にクラスター構造を把握する方法としては, 関数主成分分析 (FPCA) による可視化が挙げられる. しかし, Yamamoto (2012) では, FPCA によって得られる部分空間が必ずしもクラスタリングに適していないことを指摘し, クラスタリングに適した部分空間を見つける方法を提案している. 一方で, 教師あり判別問題では, 関数データの実数空間  $\mathbb{R}$  への射影が有効であることが明らかになっている (Delaigle and Hall, 2012). したがって, 関数データの分類問題であるクラスタリング法においても, 部分空間への射影を考えた方法は有効であると考えられる. 寺田・山本 (2018) と Delaigle et al. (2019) では, データの背後のクラスター構造を反映した部分空間を適切に推定することができる方法を独立に提案した. 2つの提案法の目的関数は同じであるが, データを部分空間へ射影するための関数の推定方法が異なる. Delaigle et al. (2019) の方法は, 射影関数の推定に対して, Haar 基底などの固定した基底を用いた推定を考えており, 数値的に安定した結果が得られる. 一方で, 選択した基底が必ずしも推定に適しているとは限らないため, クラスタ構造の推定精度が低くなる場合がある. 寺田・山本 (2018) では, この問題を解決するために, adaptive な基底を用いて推定する方法を提案しており, Delaigle et al. (2019) の推定方法よりも良い精度を得ることができる. しかし, 推定アルゴリズムにおいて必ずしも目的関数の単調減少性が保証できず, 推定が不安定となる問題があった.

本発表では, 寺田・山本 (2018) の不安定性を解消するため, 連続的な正則化を用いた新しい部分空間クラスタリング法を提案する. この方法により, 数値的に安定した推定が行えるだけでなく, 多くの場合に寺田・山本 (2018) や Delaigle et al. (2019) で提案されている方法よりも良い精度でクラスター構造の推定が行える. また, 提案手法の理論的性質について議論し, 数値実験や実データ解析を通じて既存手法との性能を比較することで, 提案手法の有効性を示す.

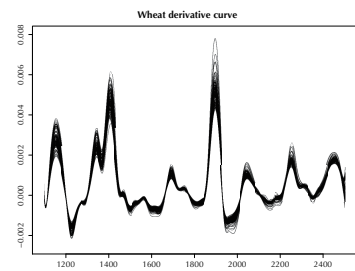


図 1: 小麦サンプルに対する NIR スペクトラルデータ

## 参考文献

- [1] Delaigle, A. and Hall, P. (2012). Achieving near perfect classification for functional data. *Journal of the Royal Statistical Society: Series B*, **74**, 267–286.
- [2] Delaigle, A., Hall, P. and Pham, T. (2019). Clustering functional data into groups using projections. *Journal of the Royal Statistical Society: Series B*, **81**, 271–304.
- [3] Kalivas, J. H. (1997). Two data sets of near infrared spectra. *Chemometrics and Intelligent Laboratory System*, **37**, 255–259.
- [4] Yamamoto, M. (2012). Clustering of functional data in a low-dimensional subspace. *Advances in Data Analysis and Classification*, **6**, 219–247.
- [5] 寺田吉彦・山本倫生 (2018). 関数データに対する部分空間クラスタリング法とその性質. 2018 年度統計関連学会連合大会.