

中西 義典

東京大学大学院総合文化研究科

助教

再標本化による情報計測のためのデータ駆動診断法開発

§ 1. 研究成果の概要

情報計測が対象とする最先端計測では、高次元のデータが得られるようになり、目的の情報を得るためにどの変数に着目するのがよいかという変数選択の問題は重要である。たとえば、コヒーレント軟 X 線散乱によるスキルミオン格子のイメージングでは、計測データとして得られる回折パターンに基づいて、スキルミオンが存在する場所を選択する必要がある。二次元実空間上に少数のスキルミオンが点在しているとき、スパースモデリングが有効に働くと考えられる。スパースモデリングを行う代表的なアルゴリズムの一つが Least Absolute Shrinkage and Selection Operator (LASSO) である。LASSO は L1 ノルムを正則化項としてもつ最小二乗法のことである。正則化項の係数である正則化パラメータの値を変えると推定結果が変わるため、適切な値をデータ駆動的に客観的に決める必要がある。従来、正則化パラメータの値は交差検証に基づいて決めることが多い。しかしながら、推定値と真の値との差(推定誤差)と、交差検証誤差とは必ずしも一致しない。交差検証誤差を最小化することにより求めた値を補正して、そもそもの目的である推定誤差を最小化するような正則化パラメータ値を決定する必要がある。本研究では統計力学で発展したレプリカ法により、推定誤差と交差検証誤差とを解析した。推定誤差を最小化する値と交差検証誤差を最小化する値との比を調べると、スパース極限において計測ノイズの大きさによらず一定値をとることが分かった。これは、交差検証誤差を最小化することによって求めた正則化パラメータ値を定数倍するのみという効率的な手法で補正することができることを示す。数値実験によりその適用条件等を明らかにしつつ、研究成果をまとめて論文を投稿中である。