

スパースモデリングによるトモグラフィーと変数選択：天文学データへの応用

植村誠 (広島大学 宇宙科学センター)

スパースモデリングとは推定した情報の「スパース性」を利用することで、より適切な推定を行う手法である。この最も基本的なものとして Least Absolute Shrinkage and Selection Operator (LASSO) と呼ばれる手法は、推定したいベクトルの要素の1次ノルムを正則化項として用いる制約付き最小二乗法である。[1] 本講演では天文学におけるデータ解析について、LASSO を応用した例を紹介する。

明るさが時間とともに変化する恒星は「変光星」と呼ばれる。変光星のなかには周期的に明るさを変えるものがあり、その明るさの時系列データから周期を推定することは基本的な解析となる。この解析で一般的に用いられるのはフーリエ変換である。時系列データをフーリエ変換することでパワースペクトルが推定でき、観測データに含まれる周期がわかる。しかし、地上からの天体観測では天候や季節の影響でデータに欠損が生じることが多い。その結果、推定されたパワースペクトルには多数のエイリアスが生じる(図1、灰色線)。このとき、もし天体が周期的に変動していることを仮定して良い状況であれば、パワースペクトルが「スパース」であることを活かし、LASSO でパワースペクトルを推定することでエイリアスを軽減することができる(図1、赤線)。[2] 欠損のない完全な時系列データを得るためには人工衛星を用いた観測などコストが高い研究が必要となるが、データに欠損があっても LASSO を用いることで高コストなデータを用いたものと同様の結果を得ることができる。

パワースペクトル推定と同様の応用例として、天体構造のトモグラフィー的な再構成の問題が挙げられる。電波干渉計と呼ばれる観測手法では天球面上の電波輝度分布を2次元フーリエ変換した複素ビジビリティと呼ばれる量が測定され、そこから輝度分布を再構成する。電波源がコンパクトな領域に限定され、多くの場所で輝度がゼロであることが期待される場合、やはり LASSO を用いて推定することで限られたデータでもより良い再構成結果が得られる(図2)。[3]

LASSO を用いた別の応用としてデータ駆動型の変数選択が挙げられる。Ia 型超新星と呼ばれる天体は爆発極大時の光度がわかっているため、それを利用して遠方銀河までの距離の推定に用いられる。しかし、その光度は実際には星間物質による光の吸収や超新星が暗くなる速度などによって補正する必要がある。極大光度を決める変数は多数の候補があるため、従来は重要と予想される変数のみを考えることも多かった。しかし、LASSO と交差検証法を用いることでデータから客観的に最も良い変数の組み合わせを選択し、必要な変数の数を推定することができる。[4] このようなデータ駆動型の変数選択はデータに潜む法則性の抽出が可能で、大規模・大次元のデータから新たな知見を得るためにも重要である。

[1] Tibshirani, R. 1996, Journal of the Royal Statistical Society: Series B (Methodological), 58, 267, [2] Kato, T. & Uemura, M., 2012, PASJ, 64, 122, [3] Honma, M., et al., 2014, PASJ, 66, 95, [4] Uemura, M., et al., 2015, PASJ, 67, 55

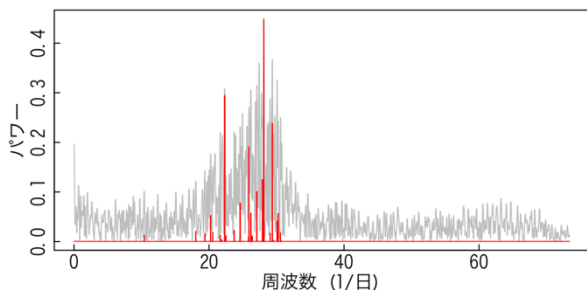


図1：恒星の明るさの時系列データから推定されたパワースペクトル。灰色線は欠損の多い実データを模したもので、多数のエイリアスが見られる。赤線は同じデータを用いて LASSO 推定されたもの。エイリアスが軽減されている。

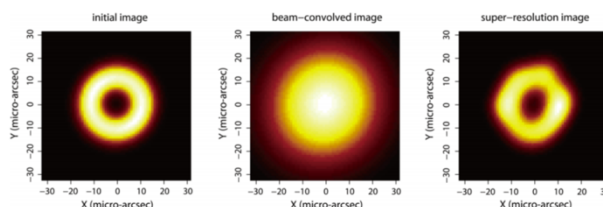


図2：左：仮定する電波マップ。中央：電波干渉計データから従来法によって再構成された結果。右：LASSO 推定で再構成された結果。[3]