

# 天文学とインフォマティクス、特にスパースモデリングの応用 Astronomy and Informatics, Especially about the Application of Sparse Modeling

植村 誠  
Makoto Uemura

広島大学 宇宙科学センター  
Hiroshima Astrophysical Science Center, Hiroshima University

## 1. はじめに

天文学では物理の第一原理計算からデータを説明するフォワードモデリング的な研究が多いが、近年のサンプル数の増加、及び観測される特徴量の種類の増加に伴い、インフォマティクスの分野で発展してきた手法を応用したデータ駆動型の研究も増えている。本講演では、推定したい情報の「スパース性」を利用した「スパースモデリング」と呼ばれる手法のうち、特に LASSO の応用例を紹介する。

データに対してパラメータの線形和で表されるモデルがある時、LASSO では最小二乗項にパラメータの絶対値の総和（一次ノルム項）を付け加えた評価関数を用いる。この評価関数を最小化して得られる解は、モデルパラメータの多くがゼロに、すなわちスパースな解になる。パラメータの数は多いが、その多くはゼロであることが期待される問題で有用な手法である。

## 2. 変光星の周期解析

恒星のなかには周期的に明るさを変化させるものがあり、変光星と呼ばれる。変光星のなかには複数の変光周期が重なり合った光度変動が観測されることがあり、その周期から恒星内部の構造を探ることができる。したがって、光度の時系列データから、そこに含まれる周期（振動数）成分を知ることは重要である。通常、この解析にはフーリエ変換によるパワースペクトル推定が用いられる。ここで天文学特有の問題となるのはデータの非均一性である。

図 1 上の実線は人工衛星からとられたある変光星の光度変動である。人工衛星の軌道によっては、このように昼夜関係なく数日間途切れることのないデータを得ることができる。図 1 中はそのデータをフーリエ変換して得られたパワースペクトルである。一方、地上からの天体観測は昼夜や天候の影響を受けるため、光度の時系列データは図 1 上の赤点のような非等間隔なものとなる。そのため、フーリエ変換で

得られるパワースペクトルには図 1 下の灰色の線のように多数のエイリアスが現れる。これでは真の信号とエイリアスが区別できない。

そこでパワースペクトルの LASSO 推定が有力な手法となる。[1] フーリエ（逆）変換は  $\sin$  関数と  $\cos$  関数の畳み込みとして線形モデルで表現できる。ここでは  $\sin$ ,  $\cos$  の係数がモデルパラメータで、それを光度の時系列データから

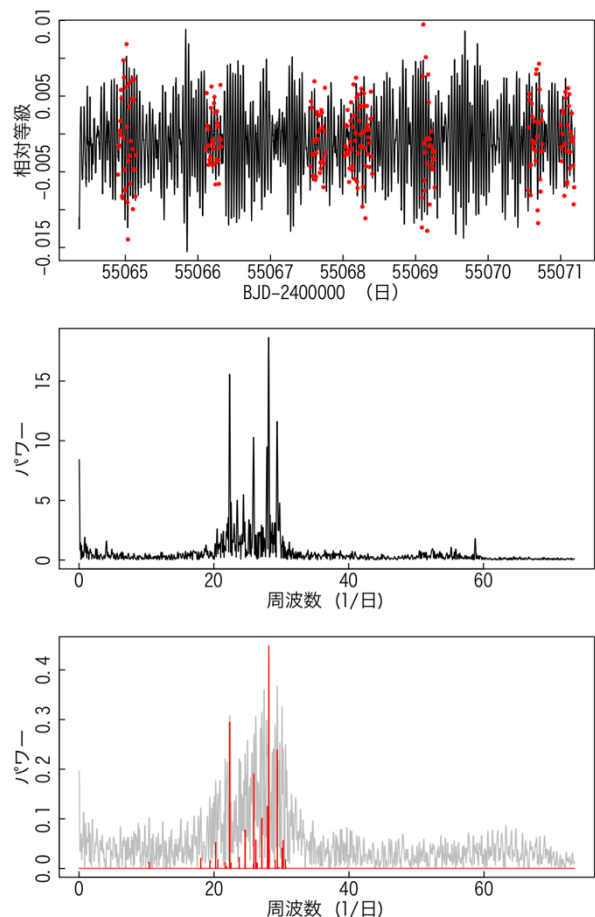


図 1 上：変光星の光度変動（実線）とそのサブサンプル（赤点）。中：上の実線から得られるパワースペクトル。下：上の赤点から得られるパワースペクトル（灰色線）と、LASSO 推定されたもの（赤線）

推定する問題と考える。光度変動はせいぜい複数の周波数信号の重ね合わせだと仮定して良いので、モデルパラメータはスパースだと期待できる。図1上の赤点からパワースペクトルをLASSO推定した結果が図1下の赤線のパワースペクトルである。エイリアスが軽減され、パワーの強い信号は少ないデータからでも正しく検出されている。

人工衛星による観測データは貴重だが、この手法を用いれば地上観測データからでも同様の研究が可能であり、可能な研究対象が広がる。

### 3. 天体の構造再構成

前節では周波数という1次元空間における「構造」をデータから再構成したが、2次元空間における「構造」の再構成にもスパース性を活かすことができる。

電波干渉計は複数の電波アンテナで検出される天体からの電波の信号を干渉させることで、巨大な電波アンテナとして高い角度分解能を実現する観測手法である。電波干渉計の観測量は天球面上の電波源マップを2次元フーリエ変換した複素数になる。電波源が非常に狭い領域にしか存在せず、かつその領域の構造を高い分解能で知りたい場合は、前節のLASSO推定の枠組みが応用できる。

図2は銀河系中心核の巨大ブラックホール(BH)を電波干渉計で観測した時のシミュレーションである。[2] 左が予想されるBH周辺からの電波源マップで、中央の低輝度な部分がBHに対応する。このような電波源マップを地球規模の電波干渉計で観測し、従来法によって画像を再構成すると図2中のようになり、BHの影は検出できない。一方、同じデータでもLASSO推定すると図2右のようになり、正しくBHの影が認識できる。現在、地球規模の電波干渉計によるBHの影を検出する計画が進行中で、スパースモデリングによるBHの直接撮像が実現するかもしれない。

### 4. Ia型超新星の極大等級の変数選択

前の二例は本来なら解けない問題を解くために、推定したい情報のスパース性を利用している。一方で、積極的にパラメータベクトルをスパースにすることで、重要なパラメータ(変数)をデータから選択することもできる。

Ia型超新星はサンプルに依らず極大時の等級がほぼ等しいため、見かけの等級から超新星、および超新星が付随する銀河までの距離を推

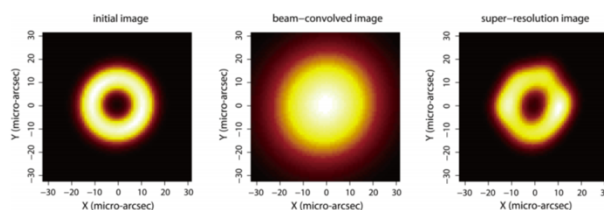


図2：左：仮定する電波マップ。中央：電波干渉計データから従来法によって再構成された結果。右：LASSO推定で再構成された結果。[2]

定することができる。超新星を使った銀河の距離推定は宇宙の加速膨張の発見にも利用されている。

しかし、実際の極大等級には有意なばらつきがあり、距離の指標として使用するには他の観測量を用いて補正する必要がある。従来は超新星の「色」と「減光速度」が極大等級と関連し、良い補正パラメータとなることが知られているが、さらに精度の高い補正をするため、超新星のスペクトルなど他の観測量を加えた新たなモデルも検討されている。しかし、スペクトルには多くの情報が含まれており、どの変数を用いるのが最適なモデルか選択が難しい。

この問題は超新星の極大等級を目的変数とし、他の観測量を説明変数とする線形モデルで表される。スペクトルの情報まで加えると、説明変数候補の数は天体サンプル数より多くなってしまふ。そこで、極大等級を決める要因はそれほど多くないという仮定に基づき、変数ベクトルをスパースにすることで、有力な変数をデータから選択する手法として、LASSOを用いることができる。我々は78個の超新星サンプルに対してスペクトルデータから抽出する276個の説明変数候補を用いてこのような解析を行い、極大等級の説明変数として従来モデルである「色」「減光速度」が最も良いモデルで、それに別の観測量を加えたモデルは極大等級の予測精度を向上しないことを示した。[3]

利用できるデータの種類・数が増えていく中で、このようなデータ駆動で重要変数を選択し、選択された変数から天体の物理を探る手法は今後も重要となるだろう。

### 参考文献

- [1] Kato, T. & Uemura, M., 2012, PASJ, 64, 122
- [2] Honma, M., Akiyama, K., Uemura, M., & Ikeda, S., 2014, PASJ, 66, 95
- [3] Uemura, M., Kawabata, K. S., Ikeda, S., & Maeda, K., 2015, PASJ, 67, 55