

スパースモデリングと情報統計力学



小 渕 智 之

京都大学大学院情報学研究科
obuchi@i.kyoto-u.ac.jp



樺 島 祥 介

東京大学大学院理学系研究科
kaba@phys.s.u-tokyo.ac.jp

「史上初、ブラックホールの撮影に成功」というニュースを覚えておられる読者も多いだろう。この成功の裏には、観測数の足りない推定問題という不良設定問題をいかにして解くかという数理科学・情報科学における技術の進展があった。ポイントは「適切な表現のもとでデータはスパースに表現できる」と仮定することで、不良設定問題を現実的に解ける問題に置き換えることであった。このような方法論を**スパースモデリング (SpM)** とよぶ。もちろんこの仮定の良し悪しは別途検証されなければならない。ブラックホールの例では、人工擬似観測データや較正天体データを用いて、SpMによる方法がきちんとした性能を上げることが慎重に検証された。

一方、個別問題における検証以外に、SpMによる方法論自体の限界を理論的に押さえておくことも重要である。すなわちデータの観測過程をモデル化、推定方式を定式化した上で、どういう条件なら真の信号が正しく復元できるかを理論的に問うのである。伝統的には統計学や情報理論で扱われる問題設定であるが、近年では統計力学を用いたアプローチも行われるようになってきている。**情報統計力学**とよばれる分野である。

情報統計力学における近年の成果の1つとして、平均場近似を用いた汎用的な推定アルゴリズムの導出法と、その挙動を解析するためのマクロなダイナミクスの理論が整備されたことが挙げられる。このアルゴリズムは、**Cavity法**という平均場近似の一種を、**確率伝搬法**というベイズ推定における近似的アルゴリズムの観点から見直すことで導出される。このアルゴリズムの特徴

は、計算量が非常に少ないこと、およびアルゴリズムを記述するパラメータ間の相関が熱力学極限で無視できるという点にある。後者のおかげで、系のマクロなダイナミクスが、それらパラメータの平均や二乗平均のみで記述できるという単純化が起こる。これにより、アルゴリズムによって到達可能な推定精度や収束までのスピードなどが議論できる。つまりアルゴリズムのある種の性能保証をすることが可能となる。面白いことに、このマクロなダイナミクスは系の大域的な平衡解析による結果と厳密に対応する。すなわち、このアルゴリズムによる推定精度限界(アルゴリズム限界)は、原理的に到達可能な限界(情報理論限界)と密接に関わっている、場合によっては厳密に一致する、ことが示される。

この平均場アルゴリズムとマクロダイナミクス解析を、SpMの問題に応用することができる。推定方式としてベイズ推定や ℓ_1 正則化付き線形回帰などが考えられるが、いずれの方式もこの方法論で系統的に解析することができる。特に興味深いのは、真の信号の復元に必要な観測数である。解析の結果、復元に必要な観測数はベイズ推定のほうが少なく済むこと、真の信号の非ゼロ要素の分布形状によってはベイズ推定のアルゴリズム限界が情報理論限界と一致することなど、SpMの理論性能を明らかにする上で有用な情報が明らかとなる。また、推定誤差は必要な観測数の前後でゼロから有限の値に立ち上がるが、これが物理的には相転移に対応し、 ℓ_1 線形回帰とベイズ推定では相転移の次数が異なることも同様に明らかとなる。

——用語解説——

スパースモデリング：

「データを適切な基底で表現した時、スパースに(ゼロ要素が多く)なるという性質を利用して、効率的な情報処理を可能にする技術。

情報統計力学：

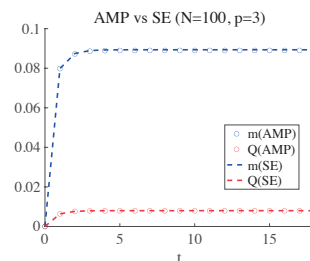
統計力学的手法を用いて情報科学の問題にアプローチする分野。スピングラス理論が根っこにあるが、そこから独自に進化した高度な平均場近似が強み。

Cavity法：

平均場近似であるベータ近似を一般化したもの。スピンを1つ取り除いた系と全スピンのある系を比較することで、磁化の自己無撞着方程式を導く。

確率伝搬法 (Belief Propagation, BP)：

ベイズ推定において、周辺分布を計算するための近似的アルゴリズム。変数間の依存構造が木構造を成しているときは正しい解を与える。信号処理や情報理論でも別名で類似アルゴリズムが使われていることがある。



平均場近似に基づくアルゴリズムによる磁化およびスピングラス秩序変数の時間発展。ミクロなアルゴリズムから計算(AMP)したものと、理論解析のためのマクロダイナミクスから計算(SE)したものとを比較している。両者はよく一致している。