次世代銀河分光観測に向けたフィールドレベル解析の確立

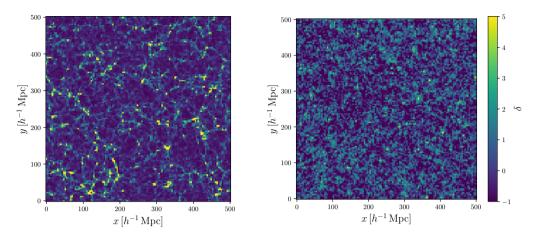
大里 健

千葉大学 先進科学センター

1. はじめに

現代の宇宙物理学は、電子や陽子といった通常の物質に加え、重力のみを通して相互作用する **ダークマター**と空間の加速膨張を引き起こすエネルギー源である**ダークエネルギー**が宇宙を構 成し、一般相対性理論によって記述される重力によって星や銀河といった構造が形成されると いう描像を確立した。一方で、ダークマターとダークエネルギーの物理的実体は未だ解明されて おらず、一般相対性理論とは異なる重力理論の可能性も排除できていないのが現状である。これ まで提唱されてきた多数の宇宙モデルの中から一つを決定する有効な手段は、現実の宇宙の観測 結果と理論モデルによる予言とを統計的に照合することである。この研究分野は**観測的宇宙論** と呼ばれ、観測技術の向上も相まって近年急速な発展を遂げた。ダークマターは宇宙の物質の大 部分を占めており、ダークエネルギーは宇宙膨張を加速させることで、重力による構造形成を妨 げる働きがあるため、その性質もまた宇宙の物質分布(**宇宙の大規模構造**)に刻み込まれている。 したがって、宇宙の大規模構造の観測は、重力の振る舞いや暗黒成分の物理的性質を解明する際 に重要な役割を果たす。観測的宇宙論における最も基本的な観測とは、銀河の発する輝線の分光 観測からその三次元位置を測定し、銀河の三次元地図を作成する銀河分光観測である。大規模構 造は宇宙の構造形成史を色濃く反映しているが、大部分がダークマターによって構成されている ため、光学望遠鏡では直接姿を捉えられない。一方で、銀河は密度の高い場所で形成されるため、 その空間分布は背景にある物質分布を反映している。したがって、銀河の三次元地図を元に、物 質分布の三次元構造を間接的に描き出すことが可能である[1]。

これまで銀河分光観測のデータ解析においては、銀河分布の情報を要約した統計量が広く用いられてきた。具体的な統計量として、二点相関関数やそのフーリエ変換に対応するパワースペクトルが挙げられる。統計量は理論的な取り扱いが容易である一方、統計量で表現できない情報が失われてしまうという問題がある。第1図に数値シミュレーションによって得られた現実的な宇宙モデルに基づいた宇宙の物質分布とそれと同じパワースペクトルを持つランダムな場(ランダムガウス場)を示す。シミュレーションで得られた物質分布は密度の濃淡がより明確であり、重力によって密度の濃いところにより物質が集積するという性質を反映している。これらの図は目で見ても識別できるほど異なる形態を示しているが、パワースペクトルは完全に同一であり、パワースペクトルの情報だけでは両者を区別できないことを意味している。高次の統計量を考えることで、より多くの情報を引き出すことが可能であるが、情報の損失を防ぐには無限次の統計量が必要になる上、高次統計量の解析的な理論予言は計算量の観点から既に現実的ではない。そこで、場の持つ情報を最大限利用するため、フィールドレベル解析という手法が提唱された[2]。これは初期宇宙の密度場から出発し、重力による非線形進化を解くことで、観測された時点における銀河分布と直接比較する手法である。したがって統計量に頼らず場の持つ全ての情報を解析に取り入れるため、観測で得られた情報を余すことなく用いることが可能である。



第1図:数値シミュレーションで得られた宇宙の物質分布(左)と それと同一のパワースペクトルを持つランダムガウス場(右)。

明るい場所は密度が高く、暗い場所は密度が低い。実際の宇宙の物質分布は三次元場であるが、ここでは薄い二次元スライスを示している。

2. フィールドレベル解析

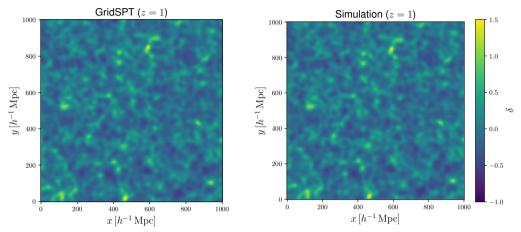
次にフィールドレベル解析の詳細を解説する。フィールドレベル解析で推定する宇宙モデルは一般的に**宇宙論パラメータ**というパラメータ(最大でも 10 個程度)で記述される。宇宙論パラメータの例として、ダークマターの総量や初期宇宙で生成された密度揺らぎの振幅などが挙げられる。統計量を用いた解析では主にこれら宇宙論パラメータのみを推定するが、フィールドレベル解析では同時に宇宙の初期密度場そのものについても推定する。観測データが与えられたときのこれらパラメータの確率分布は以下のように与えられる。

$$\log P(\boldsymbol{p}_{\text{cosmo}}, \delta_{\text{in}}(\boldsymbol{x}) | \delta_{\text{g,obs}}(\boldsymbol{x})) = -\frac{1}{2} \sum_{\boldsymbol{x}} \frac{[\delta_{\text{g,obs}}(\boldsymbol{x}) - \delta_{\text{g,fwd}}(\boldsymbol{x}; \boldsymbol{p}_{\text{cosmo}}, \delta_{\text{in}}(\boldsymbol{x}))]^2}{\sigma^2} + \text{const}$$

ここで p_{cosmo} は宇宙論パラメータセット、 $\delta_{\text{in}}(x)$ は初期密度場、 $\delta_{\text{g,obs}}(x)$ は観測された銀河個数密度場であり、xは三次元座標である。フィールドレベル解析で重要な要素の一つが右辺に現れる非線形密度場 $\delta_{\text{g,fwd}}(x;p_{\text{cosmo}},\delta_{\text{in}}(x))$ であり、これは重力による初期密度場の時間発展を考慮し、観測された時点で理論予言した銀河個数密度場を表している。この時間発展を解く理論モデルはフォワードモデルと呼ばれ、フィールドレベル解析の計算コストの大部分を占める。本研究ではフォワードモデルの具体的なモデルとしてGridSPT [3, 4]を採用する。このGridSPT モデルは既存の手法と比較して、微細な構造の重力進化を精度良くかつ高速に解くことが可能である(第2図)。右辺の意味するところは観測された場($\delta_{\text{g,obs}}(x)$)と予言された場($\delta_{\text{g,fwd}}(x;p_{\text{cosmo}},\delta_{\text{in}}(x))$)が近い値を持つモデルがより尤もらしいということである。 σ^2 は場の分散であり、ここでは簡単のため場所によらず一定とする。

フィールドレベル解析は場の持つ情報を余すことなく考慮するため、従来の統計量を元にした解析と比べ宇宙論パラメータに対する制限は改善することが見込まれる。一方で、フィールドレベル解析はフォワードモデルによる密度場の時間進化を計算する必要があり、統計量と比較するとより多くの計算量が必要となる。さらに、統計解析において初期密度場を推定するが、初期密度場は三次元グリッドとして定義されており、例えば観測した体積を空間一次元あたり100分割

したとすると、合計で 10^6 =100 万個のパラメータを推定する必要がある。このような多次元のパラメータ空間において、現実的な時間の範囲内で確率分布をサンプリングするには工夫が必要となる。そこで、本研究では Graphic Processing Unit (GPU) を用いたフォワードモデルの高速化と Hamiltonian Monte Carlo 法 (HMC 法) による効率的なサンプリング手法の実装に着手した。



第2図: GridSPT(左)と数値シミュレーション(右)で予言した宇宙の物質分布。 数値シミュレーションは物質分布の重力進化を非常に精度良く解くことが可能な一方、計算コストが非常に高い (1日程度)。一方でGridSPTでは数秒である程度のスケールまでシミュレーションと遜色ない時間発展した物質場を予言することが出来る。

3. 統計解析の高速化

まず、GridSPT フォワードモデルについて解説する。宇宙の物質分布は大スケールにおいて流体描像が有効であり、以下の Vlasov-Poisson 方程式系でその進化がよく記述される。

$$\begin{split} \frac{\partial \delta}{\delta t} + \frac{1}{a} \nabla \cdot [(1+\delta) \boldsymbol{v}] &= 0 \\ \frac{\partial \boldsymbol{v}}{\delta t} + H \boldsymbol{v} + \frac{1}{a} (\boldsymbol{v} \cdot \nabla) \cdot \boldsymbol{v} &= -\frac{1}{a} \nabla \psi \\ \frac{1}{a^2} \nabla^2 \psi &= 4\pi G \rho_{\rm m} \delta \end{split}$$

ここでtは時間、 δ は密度場の平均からのずれ 1 、vは速度場、aは宇宙の大きさを表すスケールファクター、 $H=\dot{a}/a$ は宇宙の膨張率を表すハッブルパラメータである。第三式は重力の Poisson 方程式であり、 ψ は重力ポテンシャル、Gは万有引力定数、 ρ mは宇宙の平均物質密度である。初期の宇宙では密度揺らぎは小さく($\delta \ll 1$)、重力による非線形成長が進んでいない場合、これらの式を摂動展開し解析的な表式を得ることができる。結果のみ示すと、非線形な密度場 $\delta(x,t)$ と速度場の発散 $\theta(x,t) \equiv \nabla \cdot v(x,t)$ の展開式は以下のようになる。

$$\delta(\boldsymbol{x},t) = \sum_{n} \delta_{n}(\boldsymbol{x},t), \ \theta(\boldsymbol{x},t) = \sum_{n} \theta_{n}(\boldsymbol{x},t),$$

¹ 正確な定義は $\delta(x) = (\rho(x) - \rho_{\rm m})/\rho_{\rm m}$ であり、 $\rho(x)$ は密度場である。今後、密度を表す物理量として $\delta(x)$ を主に用いるため、こちらを密度場とも呼称する。

$$\begin{pmatrix} \delta_n(\boldsymbol{x},t) \\ \theta_n(\boldsymbol{x},t) \end{pmatrix} = \frac{2}{(2n+3)(n-1)} \begin{pmatrix} n+\frac{1}{2} & 1 \\ \frac{3}{2} & n \end{pmatrix} \sum_{m=1}^{n-1} \begin{pmatrix} (\nabla \delta_m) \cdot \boldsymbol{u}_{n-m} + \delta_m \theta_{n-m} \\ \frac{1}{2} \nabla^2 (\boldsymbol{u}_m \cdot \boldsymbol{u}_{n-m}) \end{pmatrix}$$

ここで $\nabla \cdot u_m \equiv \theta_m$ である。最低次の項は初期密度場で与えられる。

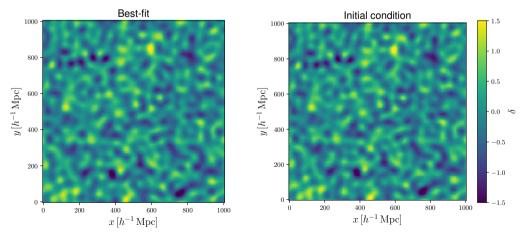
$$\begin{pmatrix} \delta_1(\boldsymbol{x},t) \\ \theta_1(\boldsymbol{x},t) \end{pmatrix} = D_+(t) \begin{pmatrix} \delta_{\rm in}(\boldsymbol{x}) \\ \delta_{\rm in}(\boldsymbol{x}) \end{pmatrix}$$

ここで $D_+(t)$ は線形成長率であり、宇宙論パラメータが与えられれば解析的に計算できる。したがって非線形な密度場 $\delta(x,t)$ は無限級数として表現されるが、現実の計算ではある次数で打ち切る必要がある。以後、5次の項まで計算する。5次までの寄与で数値シミュレーションをパーセントレベルで精度良く再現できることが分かっている。

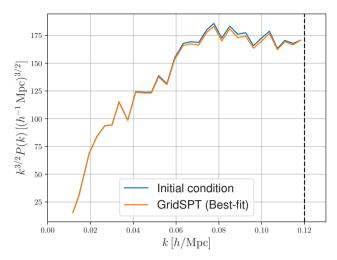
GridSPTの計算において、最も計算コストがかかる部分は多数回の三次元フーリエ変換である。上式において、微分演算子が関わる項は全てフーリエ空間で計算する必要がある。このフーリエ変換にかかる計算量を削減するため、本研究では GPU を用いて計算を行なった。具体的にはPython ライブラリである JAX [5]を用いて GridSPT を実装した。Wisteria-A 環境 GPU (NVIDIA A100) 1 基で実行した結果、5 次までの GridSPT による非線形密度場の計算にかかる時間は 5 秒であった。一方、CPU (8 threads)を用いた場合 5 分程度時間を要しており、30 倍以上の高速化を実現した。JAX を用いる利点はもう一つあり、自動微分が実装されている点である。自動微分により確率分布のパラメータに対する勾配を自動的に計算し、HMC 法を利用することが可能になった。パラメータの次元が高次元となるフィールドレベル解析においては、HMC 法のような効率的な手法は不可欠である。

4. 結果

ここでは実装したフィールドレベル解析の結果について述べる。まず、実際の観測データに適用する前に、数値シミュレーションのデータをフィールドレベル解析の観測データとして入力する。数値シミュレーションを実行する際に仮定された宇宙モデルがフィールドレベル解析によって正しく復元されるかを確かめることで、解析手法の妥当性を検証する。まず、確率分布のサンプリングを行う前に、最良推定値、すなわち事後確率を最大化するパラメータが十分正解の値に近いか検証を行った。



第3図:推定された宇宙の初期密度分布(左)と実際のシミュレーションで 初期条件として用いた初期密度分布(右)。



第4図: 推定された初期密度場(橙)と実際のシミュレーションで初期条件として用いた 初期密度分布(青)のパワースペクトル。

黒い点線は解析で用いた最大の波数を示す。フォワードモデルにおいてあまりに大きい波数(小スケール)は摂動展開の近似が破れ精度が悪くなるため、一般に大きいスケールに限って解析に用いる。

第3図に推定された宇宙の初期密度分布、第4図に初期密度分布のパワースペクトルの比較を示す。視覚的にも正しく初期密度場が見積もられており、パワースペクトルについてもシミュレーションで用いた初期密度場から計算した値と矛盾がないことが示された。現状、最良推定値のみではなく、HMC 法を用いた確率分布のサンプリングについて計算を行っている段階である。

5. まとめ

宇宙物理学の最大の謎であるダークマター・ダークエネルギーの正体に迫るべく、大型の銀河分光観測が計画されている。その際、観測された銀河分布から如何に情報を引き出すか、ということが課題となっているが、フィールドレベル解析は原理的に場の持つ情報を余すことなく引き出せるという可能性を秘めている。本研究では、フィールドレベル解析で課題となっていた計算量の問題と高次元パラメータ空間での確率分布のサンプリングという問題に対して、GPUと自動微分を利用した HMC 法を用いて解決を試みた。実装した解析手法を数値シミュレーションで得られた現実的な宇宙の密度分布を基に初期密度場を推定した結果、正しい結果を復元できることを確認した。今後、HMC 法を用いた確率分布サンプリングを行い、既存の統計量と比較してフィールドレベル解析が宇宙モデルに対するより強い制限を与えられるかどうか検証を行う。また、今回用いたシミュレーションは宇宙の物質分布を直接用いているが、本来観測されるのは銀河の分布である。物質分布を元に銀河分布を予言する手法である銀河バイアス展開法[6]という方法が提唱されており、今後はこの手法を実装することで現実的な設定におけるフィールドレベル解析の実装、そしてその妥当性を検証する。

謝辞

本研究は東京大学情報基盤センター2022 年度若手・女性利用者推薦課題として、Wisteria/BDEC-01 Aquarius の利用支援を受けました。また科学研究費若手研究 22K14036「次世

代銀河分光観測に向けたフィールドレベル宇宙論解析の確立」(代表:大里 健)として支援頂きました。これらの援助に感謝いたします。

参考文献

- [1] Aubourg E. et al., Physical Review D, 92, 123516 (2015)
- [2] Jasche J., Wandelt B. D., Monthly Notices of the Royal Astronomical Society, 432, p. 894-913 (2013)
- [3] Taruya A., Nishimichi T., Jeong D., Physical Review D, 98, 103532 (2018)
- [4] Taruya A., Nishimichi T., Jeong D., Physical Review D, 105, 103507 (2022)
- [5] Bradbury J. et al., JAX: composable transformations of Python+NumPy programs, http://github.com/google/jax
- [6] Desjacques V., Jeong D., Schmidt F., Physics Reports, 733, p. 1-193 (2018)