

# 位相的データ解析による銀河分布の定量化と バリオン音響振動抽出

竹内 努<sup>1,2</sup>・河野 海<sup>1</sup>・クレ スチエータ<sup>1,3,4</sup>・西澤 淳<sup>5,6,7</sup>・  
村上 広耶<sup>1</sup>・馬 海霞<sup>1</sup>・本武 陽一<sup>8</sup>

(受付 2022 年 9 月 21 日 ; 改訂 2023 年 3 月 16 日 ; 採択 3 月 20 日)

## 要 旨

宇宙空間において、銀河は銀河団、銀河群、フィラメント(1次元的な銀河の連なり)、ヴォイド(銀河の存在しない領域)といった構造を形作り不均一に分布している。これを宇宙の大規模構造と呼ぶ。宇宙論では通常の物質をバリオンと呼び、我々が直接観測するのはこのバリオンが形作る大規模構造である。

大規模構造を形成する主要因は重力不安定だが、これとは別種の構造の源となる機構が存在する。これがバリオン-光子脱結合の時期に生じるバリオンの音波である。これはバリオン音響振動(BAO)として知られ、銀河の空間分布に刻印されている。本研究では、銀河分布のBAO信号を検証するために、位相的データ解析(topological data analysis: TDA)の方法で銀河の空間分布を解析した。TDAでは、点群データを「穴」の集合として解釈し、その幾何学的構造からパーシステントホモロジー群を構成することで、1つ1つの穴構造の特徴を抽出する手法である。従って、それぞれの穴の情報を保持するパーシステントホモロジー群と穴の位置情報などを組み合わせることで、TDAによって点群データから定義される穴集合のそれぞれの穴のサイズ、位置などの特徴が定量評価できる。

我々はまず、バリオンの物理あり、なしの場合のシミュレーションデータセットを解析し、TDAのパフォーマンスを調べた。この結果、TDAは、宇宙の大規模構造の中からBAO信号を実際に検出できることがわかった。バリオンありのシミュレーションはBAOからの有意な信号を示したのに対し、バリオンなしデータはこの信号が検出されなかった。この結果をもとに、次にスローンデジタルスカイサーベイデータリリース14(SDSS DR14)のeBOSS(extended Baryon Oscillation Spectroscopic Survey)から $z < 1.0$ のクエーサー(活動銀河核の一種)を抽出してサンプルを構成し、TDAを適用した。この解析で、我々は $r \sim 150$  [Mpc]のスケールで特徴的な穴(中空のシェル)を発見した。これは、銀河/クエーサー分布に刻印されたBAOの信

<sup>1</sup>名古屋大学 素粒子宇宙物理学専攻 : 〒464-8602 愛知県名古屋市千種区不老町

<sup>2</sup>統計数理研究所 客員 : 〒190-8562 東京都立川市緑町 10-3

<sup>3</sup>国立天文台 科学研究部 : 〒181-8588 東京都三鷹市大沢 2-21-1

<sup>4</sup>日本学術振興会特別研究員(PD)

<sup>5</sup>岐阜聖徳学園大学 DX推進センター : 〒501-6194 岐阜県岐阜市柳津高桑西 1-1

<sup>6</sup>名古屋大学 高等研究院 : 〒464-8601 愛知県名古屋市千種区不老町

<sup>7</sup>名古屋大学 素粒子宇宙起源研究所 : 〒464-8602 愛知県名古屋市千種区不老町

<sup>8</sup>統計数理研究所(現 一橋大学 ソーシャル・データサイエンス研究科) : 〒186-8601 東京都国立市中 2-1)

号に正確に対応している。この解析は2000個のクエーサーからなるサブサンプルで実行した。これはTDAが、たとえ非常に疎なサンプリングであっても中空の穴のような位相的構造を検出するのに有効であることを示している。

キーワード：位相的データ解析，パーシステントホモロジー，銀河分布，宇宙の大規模構造，バリオン音響振動。

## 1. はじめに

### 1.1 宇宙の物質分布の特徴づけと記述方法

各宇宙年齢における物質および銀河の分布には、宇宙の構造の形成と進化に関する基本的な情報が含まれている (e.g., Peebles, 1980; Efstathiou and Silk, 1983; Bernardeau et al., 2002)。まず強調しておかなくてはならない事実として、宇宙論が理論的に予言するのは物質分布あるいは銀河分布の統計的性質であり、時空間のどの位置にどの銀河が誕生するかなど具体的な天体についての予言力はない。よって、宇宙論の理論と観測の比較は統計的記述を通じて行うしかない。この理由から、物質分布のゆらぎの統計的特性を特徴づけるために、膨大な数の洗練された方法が提案されてきた (e.g., Peebles, 1980; Martinez and Saar, 2001; Bernardeau et al., 2002)。その中でも、 $n$ 点相関関数は、観測される銀河分布 (e.g., Peebles, 1980; Martinez and Saar, 2001; Bernardeau et al., 2002)の解析で最も一般的でよく研究されている方法である。宇宙空間のある位置 $\vec{x}$ における銀河の数密度 $n(\vec{x})$ を測定し、その統計平均 $\langle n \rangle$ と比較する。宇宙は我々の住んでいるもの1つしか観測できないが、仮想的に多数の宇宙の集合(アンサンブル: ensemble)を想定し、ある量 $Q$ の多数の宇宙での平均を考えることを統計平均あるいはアンサンブル平均と呼び、記号 $\langle Q \rangle$ で表す。これに対し、空間内で平均した値を体積平均とよび、記号 $\bar{Q}$ で表す。宇宙論では、この2種類の平均が一致するという仮定(宇宙論的エルゴード仮説)に基づいて議論を展開する。

この準備のもと、 $d^3x_1, d^3x_2$ という2つの微小体積を考える。双方の微小体積中に銀河が見つかる同時確率を $\mathcal{P}(\vec{x}_1, \vec{x}_2)d^3x_1d^3x_2$ とする。銀河分布が一様、即ち無相関(クラスタリングがない)場合、微小体積内に銀河が見つかる確率は

$$(1.1) \quad \mathcal{P}(\vec{x})d^3x = \langle n \rangle d^3x,$$

なので、体積 $d^3x_1$ 内の銀河と体積 $d^3x_2$ 内に銀河が見つかる同時確率は

$$(1.2) \quad \mathcal{P}(\vec{x}_1, \vec{x}_2)d^3x_1d^3x_2 = \langle n \rangle^2 d^3x_1d^3x_2$$

となる。銀河がクラスタリングしている場合、同時確率は

$$(1.3) \quad \mathcal{P}(\vec{x}_1, \vec{x}_2)d^3x_1d^3x_2 = \langle n(\vec{x}_1)n(\vec{x}_2) \rangle d^3x_1d^3x_2 \equiv \langle n \rangle^2 [1 + \xi(\vec{x}_1, \vec{x}_2)] d^3x_1d^3x_2$$

となり、一様な場合に比べて $\langle n \rangle^2 \xi(\vec{x}_1, \vec{x}_2)d^3x_1d^3x_2$ だけ高くなる。この $\xi(\vec{x}_1, \vec{x}_2)$ を銀河分布の2点相関関数と呼ぶ。同様に、一般に $n$ 点相関関数は $n$ 個の微小体積内に銀河を見出す同時確率を用いて定義される。

ここまでは銀河分布を点群として定式化した。銀河分布を議論する際には数密度の分布を平滑化し、密度場として扱うことも多い。点群と密度場の数学的扱いは異なるが、宇宙物理ではこの違いは重視しないで同様に扱うことが多く、本稿でもそれに倣う。宇宙論的定式化のため、銀河分布の密度場 $\rho(\vec{x})$ の密度ゆらぎ

$$(1.4) \quad \delta(\vec{x}) \equiv \frac{\rho(\vec{x}) - \bar{\rho}}{\bar{\rho}}$$

を考えるのが便利である。ここで  $\bar{\rho}(\bar{x})$  は平均密度を表す。密度  $\rho(\bar{x})$  は点群としての銀河の数密度、あるいは平滑化した銀河分布の質量密度のどちらを表現する場合もありうる。密度ゆらぎ  $\delta(\bar{x})$  を用いると、相関関数  $\xi(\bar{x}_1, \bar{x}_2)$  は

$$(1.5) \quad \xi(\bar{x}_1, \bar{x}_2) = \frac{\langle [\rho(\bar{x}_1) - \bar{\rho}] [\rho(\bar{x}_2) - \bar{\rho}] \rangle}{\bar{\rho}^2} = \langle \delta(\bar{x}_1) \delta(\bar{x}_2) \rangle$$

と表せる。なお、宇宙論では宇宙は一様等方であると仮定するので、対称性より

$$(1.6) \quad \xi(\bar{x}_1, \bar{x}_2) = \xi(|\bar{x}_1 - \bar{x}_2|) \equiv \xi(r)$$

として扱うことが一般的である。式(1.5)の形で表すと、相関関数が密度場の2次のキュミュラントになっていることが分かる。一般的に、ある空間上のベクトル  $\bar{x} = (x_1, \dots, x_n)^\top$  を添字とする確率変数  $X(\bar{x})$  を確率場 (random field) とよぶ。数学的には、ある確率場は無次元までのモーメントあるいはキュミュラントで一意に定まることから、キュミュラントと等価な  $n$  点相関関数によって確率場が決まる。これは確率場の特性関数のキュミュラント展開定理を用いて示せる (e.g., Daley and Vere-Jones, 2003)。即ち、相関関数を測定することによって原理的には宇宙の密度場の統計的性質を特定できる。

しかし実際には、高次の  $n$  点相関の推定は極めて難しく非現実的であるため、高次相関の情報を扱う別のアプローチが検討されている。銀河分布の3次元構造の発見 (e.g., de Lapparent et al., 1986) により、平滑化した密度場の幾何学 (トポロジー) を用いて大規模構造を定量化し、背後にある物理過程を特定しようというアイデアが提示された (e.g., Shandarin and Zeldovich, 1983; Gott et al., 1986, 1987, 1989; Boerner and Mo, 1989)。早くから用いられてきたのがジーナス (種数, genus:  $G$ ) によるトポロジーの特徴付けである (Gott et al., 1986)。銀河分布の密度ゆらぎ  $\delta(\bar{x})$  の標準偏差  $\sigma$  の  $\nu$  倍 ( $\nu$  は定数) を閾値とすれば、 $\nu\sigma$  に対する上側レベル集合が各  $\nu$  について定義できる。これをエクスカージョン集合 (ここでは  $E_\nu$  と表記) とよび、 $\rho(\bar{x}) = \nu\sigma$  となる等密度面がその表面となる。密度場を考えているので、 $E_\nu$  は3次元の集合である。銀河分布のジーナスは、平滑化した銀河密度場のオイラー標数の  $-1/2$ 、即ち

$$(1.7) \quad G = -\frac{1}{2} \int K dA,$$

( $K$ : 等密度面のガウス曲率,  $dA$ : 等密度面上の面積要素) と定義される。なお、ここでの定義は宇宙論で用いられるもので、数学における通常定義と1だけずれていることに注意されたい (松原, 1996)。エクスカージョン集合  $E_\nu$  の単位体積あたりのジーナスを  $\nu$  の関数として表した量  $G(\nu)$  をジーナスカーブと呼ぶ。ジーナスカーブを銀河分布の集約統計量として用い、宇宙論的研究を行う方法がジーナス統計である。このアプローチは、大規模構造の重力による成長を密度分布の初期値からの微小なずれ (摂動) として扱うことで、大規模構造進化を記述する理論体系、いわゆる構造形成の摂動論的アプローチに関連して開発されてきた (e.g., Matsubara, 1994, 1996, 2003; Matsubara and Suto, 1996; Matsubara et al., 2022; Matsubara and Yokoyama, 1996; Pogosyan et al., 2009; Gay et al., 2012)。摂動論を用いると、ジーナスカーブは大規模構造のスケールではモーメントの関数として解析的に書けるので (e.g., Matsubara, 1994, 1996; Matsubara and Suto, 1996; Matsubara and Yokoyama, 1996)、これによって位相幾何学的情報を定量的に表現できる。特にジーナスカーブはガウス場と他の確率場を区別できるため、大規模構造を解析するための一般的な道具として定着している。

ジーナス統計の方法を含む、より一般的な体系がいわゆる“ミンコフスキー汎関数 (Minkowski functionals)” である (e.g., Hikage et al., 2003; Matsubara, 2003; Sullivan et al., 2019)。ミンコフスキー汎関数の大規模構造への応用でも式(1.4)で定義した平滑化した銀河密度場の密度ゆ

らぎを考える。この場合ミンコフスキー汎関数  $M_k$  は  $k = 0$  については

$$(1.8) \quad M_0 = \frac{1}{V} \int_{E_\nu} dV,$$

$k = 1, 2, 3$  については閾値  $\nu$  の等密度面上の曲率の積分によって定義されるので、原理的には銀河密度場はミンコフスキー汎関数によって完全に特定できる (Schmalzing and Buchert, 1997; Schmalzing and Gorski, 1998) (これ以降、本論文を通して  $k = 0, 1, 2, 3$  は次元の意味でのみ用いる)。実際、この方法は宇宙論的解析に広く応用されている (密度場の解析: e.g., Mecke et al., 1994; Kerscher et al., 1997; Beisbart et al., 2002; Hikage et al., 2003; Kerscher and Tikhonov, 2010; Wiegand et al., 2014; Wiegand and Eisenstein, 2017; 赤方偏移歪み: e.g., Appleby et al., 2018; 弱重力レンズ宇宙論: e.g., Spurio Mancini et al., 2018; Vicinanza et al., 2019; Parroni et al., 2020; 宇宙再電離: e.g., Yoshiura et al., 2017; Bag et al., 2018, 2019; Chen et al., 2019; 宇宙論/重力理論の検証: e.g., Ling et al., 2015; Junaid and Pogosyan, 2015; Shiraiishi et al., 2016; Fang et al., 2017; および宇宙マイクロ波背景放射 (CMB): e.g., Matsubara, 2010; Planck Collaboration et al., 2016; Santos et al., 2016; Chingangbam et al., 2017; Ganesan and Chingangbam, 2017; Joby et al., 2019)。

## 1.2 ミンコフスキー汎関数から位相的データ解析へ

このように、銀河密度場は密度場のトポロジーによって定量的評価ができる。上で述べたジーナス統計は、点群データの性質を特徴付ける。ジーナスのように、点群データ全体をその穴構造についての一つの特徴量で表現するのではなく、点群データに含まれる個別の穴構造についてのスケールのような特徴をそれぞれ保持できる表現がパーシステントホモロジー群である (TDA: e.g., Edelsbrunner et al., 2002; Zomorodian and Carlsson, 2004; Wasserman, 2018)。以下ではパーシステントホモロジー群に基づくデータ解析法である TDA についての簡易な説明を行う。

データが持つ一つ一つの穴構造を評価できる TDA を用いることで、宇宙論におけるデータ解析をより発展させることができる。本研究で対象となる点群データの穴構造は、銀河や任意の天体などで構成される  $N$  体系の各点に配置された半径  $r$  の  $n$  次元球の集合によって形成される。宇宙論分野黎明期のこのアプローチによる研究は、基本的にはランダムパーコレーションの応用であった (Gott et al., 1986)。すなわち、各銀河の周りに球を置いて半径  $r$  を増加させ、半径  $r$  の関数として考えている体積内の連結成分の幾何学的構造を調べる方法である。これは TDA で用いられる考え方そのものである。

萌芽的な意味での TDA の概念の誕生は 90 年代に遡るが、実質的な誕生は 2000 年代に入ってからである (see e.g., Edelsbrunner et al., 2002; Edelsbrunner and Harer, 2010)。TDA では、よくデータセットをパーシステントホモロジー群として表現することが行われる。ランダムパーコレーションと同様、パーシステントホモロジー群では点群データの各点の周りに半径  $r$  の (中空でない) 球を考え、さまざまな次元の穴構造の生成と消滅を扱う。この過程を図 1 で概略的に説明する。2 次元平面内に 3 つの点からなる集合を考える。半径  $r$  が小さい場合、これは点の周りに 3 つの円盤が置かれた構造である (図 1: 左)。次に  $r$  がある値より大きくなると、円盤が連結してサイクル構造が形成される (図 1: 中央)、これが穴の誕生とみなされる。さらに  $r$  が大きくなると、サイクル構造の内部が円盤で埋め尽くされて、穴は消滅する (図 1: 右)。ある穴が発生 (誕生) する  $r$  を  $x$  座標、消滅 (死亡) する  $r$  を  $y$  座標として、データに存在する全ての穴構造の生成と消滅をそれぞれプロットして構成されるのがパーシステンス図 (persistence diagram) と呼ばれるパーシステントホモロジー群の中心的な表現方法である。 $n$  次元空間中の

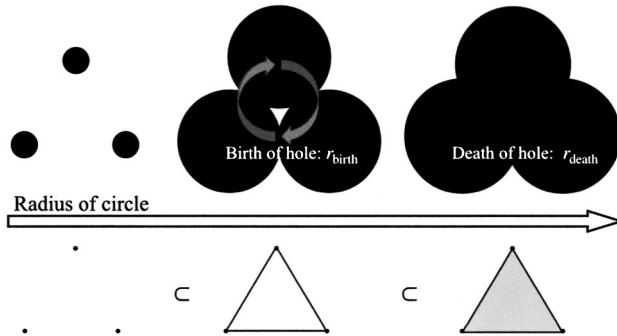


図 1. パーシステントホモロジー群の概念図 (Kono et al., 2020). 位相的情報は、離散的なデータ点からの半径  $r$  の  $n$  次元の球で構成された「穴」によって特徴づけられる。

点群データに対してパーシステンス図は、0次元の穴(連結成分)、1次元の穴(リング構造)、2次元の穴(シェル構造)…といった複数の次元の穴に対して個別に定義できる。より厳密な定義と方法論は2章で示す。

天体物理学や宇宙論における近代的な TDA の応用例はここ数年で徐々に増加している。Pranav et al. (2017, 2019a), Feldbrugge et al. (2019), Wilding et al. (2021), そして Biagetti et al. (2021) は一般の確率場への応用を念頭に置き、ガウスおよび非ガウス確率場の確率的ホモロジーに関する詳細な定式化を示した。この方法により、(Pranav et al., 2019b) は CMB に特異な信号を検出した。Cole and Shiu (2018) は、TDA を適用して CMB の非ガウス性を評価する方法を開発した。Xu et al. (2019) は、 $N$  体シミュレーション (Libeskind et al., 2018) から生成された大きな銀河データ集合に TDA を適用し、シミュレーションデータで 23 個のヴォイドを検出した。これは、TDA のヴォイド検出における有用性を示しており、従来の方法 (e.g., CLASSIC: Kitaura and Angulo, 2012) の結果ともよく一致している。さらに彼らは、MassiveNu シミュレーション (Liu et al., 2018) と比較して宇宙論的パラメータを制限し、パラメータ間の縮退を解く方法について検証した。Elbers and van de Weygaert (2019) と Giri and Mellema (2021) は TDA による宇宙再電離期の電離バブルの構造と進化、およびサイズ分布とトポロジーの定量的推定の有効性について議論した。

### 1.3 バリオン音響振動(BAO)

宇宙の構造形成は質量において支配的なダークマターが駆動する。しかし、電磁波で観測できる宇宙、即ち目に見える宇宙の姿を形作るのは電磁波と相互作用する通常の物質のなす構造である。宇宙論では通常の物質をバリオンとよぶ慣習があるので、本論文でも以降通常物質をバリオンと表現する。バリオンは、電磁相互作用(たとえば、輻射加熱/冷却、ガス圧、流体力学的過程)を介して複雑な進化をたどる。このような非自明な現象の典型例が、初期の宇宙のバリオン-輻射 2 流体で生成されたバリオン音響振動(BAO)である (Peebles and Yu, 1970; Sunyaev and Zeldovich, 1970)。ここでは、BAO と関連する観測量について簡単に説明する。

初期宇宙のバリオン-光子プラズマにおける点源による初期摂動を考える。プラズマでは、バリオンと光子がトムソン散乱(Thomson scattering)を解して結合し、単一の流体としてふるまう。光子は光速で動き、また非常に多数存在するため、バリオン-輻射流体は密度に対してとてつもなく大きな圧力を持つ。圧力は周囲と均一になろうとし、その結果外側に向かって伝播する球形の音波を生じる。この時期の音速は、次式によって評価される。

$$(1.9) \quad c_s = \frac{c}{\sqrt{3 \left(1 + \frac{3\rho_B}{4\rho_R}\right)}}.$$

ここで  $c$  は光速であり,  $\rho$  はエネルギー密度を, 下付き文字 B と R はそれぞれバリオンと輻射の成分を表す. 式(1.9)は, バリオンと輻射の脱結合前の音速が, 初期の光速の  $\sim 57\%$  に及ぶことを示している. 音響的地平線(sound horizon)の最終半径  $r_s$  は, 次の式で与えられる.

$$(1.10) \quad \begin{aligned} r_s &= \int_0^{t_{\text{dec}}} (1+z)c_s dt \\ &= \int_{z_{\text{dec}}}^{\infty} \frac{c_s}{H_0 \sqrt{\Omega_{R,0}(1+z)^4 + \Omega_{M,0}(1+z)^3 + \Omega_{\Lambda,0}}} dz. \end{aligned}$$

ここで  $z$  は宇宙論的赤方偏移,  $t$  は宇宙年齢,  $H_0$  はハッブルパラメータ,  $\Omega_{R,0}$  は輻射のエネルギー密度パラメータ,  $\Omega_{M,0}$  は物質のエネルギー密度パラメータ,  $\Omega_{\Lambda,0}$  は宇宙項の密度パラメータ, 下付き添字 dec は光子と物質の脱結合の時刻での値を, 0 は現在時刻での値を意味する. それぞれの詳細は付録 A.3 節を参照のこと. 我々が観測するのは, 大規模構造に刻印された, 原始揺らぎから生じた多くの音波の重ね合わせである. BAO は定義上バリオンに生じる現象だが, バリオンと暗黒物質の密度ゆらぎの構造は重力相互作用を通して作用し合うため, 最終状態では一致することが知られている (バリオンの追いつき (catch-up): e.g., Ma and Bertschinger, 1995; Eisenstein et al., 2007).

BAO のスケールは, 共動座標 (宇宙膨張による距離の拡大の影響を除いた座標系. 付録 A を参照) で一定である. このことから, BAO シグナルは基本的には銀河の 2 点相関関数で検出することができる. しかし, BAO スケールは銀河の典型的なスケールと比較して非常に大きいため, 2 点相関に現れる BAO シグナルはかなり弱い. したがって, BAO シグナルを測定するには, 広大な観測体積において高密度でサンプリングされた大規模銀河探査が必要とされる.

Sloan Digital Sky Survey (SDSS) は, 全天の 3 分の 1 をカバーする最も広い可視測光・分光探査である. SDSS により, 2 点相関関数で約 150 Mpc の BAO シグナルが検出された (Eisenstein et al., 2005). ここで単位 [pc] はパーセクを意味し,  $1 \text{ pc} = 3.08 \times 10^{18} \text{ cm}$  である. Mpc, Gpc はそれぞれメガパーセク, ギガパーセクと読み,  $10^6 \text{ pc}$ ,  $10^9 \text{ pc}$  を意味する. とはいえ, 伝統的方法による BAO 解析は SDSS のような大規模広域探査データでのみようやく可能になる.

よって, データへの要求がより少なくてすむ BAO シグナルの検出法および定量化の開拓は, 宇宙論モデルの制限において極めて重要な課題である. 本研究ではまず TDA を導入し, 銀河の空間分布データへの応用を紹介する. そして有意性検定の方法についても議論し, TDA が BAO スケールの評価に有用であることを示す. さらに, TDA の長所である逆解析についても議論する. これは, 銀河分布上に刻まれたバリオン音響振動に対応するトポロジー構造の視覚化を可能にする初めての方法である.

本論文は次のように構成されている. 2 章では, TDA の概念と方法, 逆解析を紹介する. 次に, 3 章ではバリオン物理ありとなしの場合のシミュレーションデータを用い, BAO 研究の道具としての TDA の有効性を検証する. 4 章では SDSS DR14 (Pâris et al., 2018) の extended Baryon Oscillation Spectroscopic Survey (eBOSS; Dawson et al., 2016) Large Scale Structure (LSS) catalog (Ata et al., 2018) に TDA を適用した結果を示す. 5 章では, 本論文で導入した方法論の結果と将来展望についてのまとめと議論を述べる. 本稿の物理的内容は Kono et al. (2020) の内容と議論を基礎にしている.

本論文では *Planck* 衛星による CMB 観測で推定された宇宙論パラメータ  $h = H_0/100 \text{ [km s}^{-1} \text{ Mpc}^{-1}] = 0.6766$ ,  $\Omega_{\Lambda,0} = 0.6889$ ,  $\Omega_{M,0} = 0.3111$ ,  $\ln(10^{10} A_s) = 3.047$  を採

用する。この物理的意味については付録 A.3 節を参照していただきたい。

## 2. 位相的データ解析 (TDA)

TDA では、 $N$  次元ユークリッド空間  $\mathbb{R}^N$  中の有限個の点群  $P$  の位相幾何学的性質を検証する (この論文全体を通して、 $\mathbb{R}$ ,  $\mathbb{N}$  と  $\mathbb{N}_0$  は、それぞれ実数、自然数、および非負自然数の集合を表す)。点群  $P$  が有限集合であるため、伝統的なトポロジーの議論を単純に適用しても  $P$  の特徴に関する重要な情報を抽出することはできない。TDA は、スケール (e.g., Edelsbrunner et al., 2002; Boissonnat et al., 2018) を変えてもある程度保持される (persistent) 性質を調べることにより、点群データが持つ意味のある位相幾何学的特徴を抽出できる。このため、TDA で基本的な役割を果たすこのような幾何学的モデルを導入する。この解析の主要な道具がパーシステントホモロジー群である。

### 2.1 パーシステントホモロジー群

#### 2.1.1 点データの「形」

複雑な構造を持つ大規模データの分析では、データの性質を代表する少数の特徴量によってデータを縮約して表現することが重要である。このため、次元縮小 (dimensionality reduction) が中心的役割を果たす。例えば、以下のようなユークリッド空間に設定された点群データを考える。

$$(2.1) \quad \mathcal{P} \equiv \{x_i \in \mathbb{R}^N \mid i = 1, \dots, m\}.$$

例えばデータ  $\mathcal{P}$  に周期的な構造がある場合はフーリエ変換が非常にうまく機能し、データ点がクラスタリングしている場合は相関関数による解析が有効である (e.g., Peebles, 1980)。

しかし、このようなアプローチは、非周期的、あるいは明確なクラスター構造を持たない複雑なデータでは有効ではない。フーリエ変換や相関関数を用いる従来の解析方法では、このような点群の形状をうまく特徴付けることは難しい。代わりに、各データ点を中心とする球を置き、その球の結合関係としてデータ点の接続関係を導入することでデータの特徴付ける方法が有効である。例えば接続関係を穴構造の数によって特徴づけると、それはドーナツとコーヒーカップを同一視するような同相変換に対して不変なホモロジー群として、点群データを大胆に縮約して表現したことに対応する。一方これでは、穴の大きさのようなスケールの情報が失われてしまう問題がある。そこで前述の、各データ点に置いた球の半径を大きくしていく際の穴構造の生成・消滅を抽出することが有用となる。これによって、穴の形状やサイズといった情報を保持した点群データの縮約表現が得られる。これがホモロジー群を拡張した“パーシステントホモロジー群”である。

#### 2.1.2 点群データにおける「穴」の代数表現

前述の通り、点群データの分析ではランダムパーコレーションのように、各点の周りに球を置いて半径を増加させ、半径の関数として考えている体積内の連結成分の穴構造を調べるのが有用である。このような点群を元に構成された球の集合は、球という凸閉集合の集合と見なすことができる。凸閉集合の有限個の集まりで被覆された構造があった場合、その構造はトポロジカルな特徴を保存 (ホモトピー同値) しながら、代数操作が可能な単体複体による表現に変換できる。単体複体は図形の構造を表現したものであり、 $k$  次元の三角形に対応する概念である単体の組み合わせとして構成される。

一般には単体複体は位相空間の構造を持たないが、ここではまず具体的なイメージを持ちやすいユークリッド空間を導入した単体複体について説明する。ユークリッド空間内において  $k$

単体は以下で定義される.

**Definition 1.** (ユークリッド空間内での単体 (simplex))  $\mathbb{R}^n$  内の  $k+1$  個の点  $p_0, p_1, \dots, p_k$  で構成される  $k$  個のベクトル  $p_0\vec{p}_1, p_0\vec{p}_2, \dots, p_0\vec{p}_k$  が一次独立な関係にある場合, その  $k+1$  個の点によって  $k$  次単体  $|p_0, p_1, \dots, p_k|$  は以下のように定義される.

$$|p_0, p_1, \dots, p_k| := \left\{ \sum_{i=0}^k \lambda_i p_i \mid \lambda_i \geq 0 (i=0, 1, \dots, k), \sum_{i=0}^k \lambda_i = 1 \right\}$$

具体的には, 0 次元の単体は点, 1 次元の単体は三角形, 2 次元の単体は四面体に対応する (図 2). この単体をずれなく綺麗に貼り合わせたものが単体複体である. ずれなく貼り合わせるという操作に基づき, ユークリッド空間内における単体複体は以下のように定義される.

**Definition 2.** (ユークリッド空間内での単体複体 (simplicial complex))  $\mathbb{R}^n$  内の有限個の単体の集まり  $K$  が以下の条件を満たすとき,  $K$  を単体複体と呼ぶ.

- (1)  $K$  に属する単体  $\tau$  の面  $\sigma$  が  $K$  に含まれる.
- (2) 2 個の単体  $\tau, \sigma \in K$  の共通部分  $\tau \cap \sigma$  は空集合でない限り,  $\tau$  の面かつ  $\sigma$  の面である.

ここで単体の面とは, 直観的には 4 面体なら境界 (表面) の 4 枚の 3 角形という 1 次元低い単体の集合, 線分なら両端の 2 点という 0 次元単体の集合を指す. 図 3 にユークリッド空間内での単体複体の例を示す.

前述したように, 単体複体そのものは位相空間によらず, 以下のような組み合わせ論的構造として定義される.

**Definition 3.** (単体複体 (simplicial complex)) 有限集合  $V$  上の単体複体  $X$  とは, 次の条件

- (1)  $\{a\} \in X$  ( $\forall a \in V$ ),
- (2)  $\sigma \in X$  のとき,  $\tau \subset \sigma$  ならば  $\tau \in X$ .

を満たす  $V$  の部分集合の族である.

$X$  における単体の最大次元は複体  $X$  の次元と呼ばれ,  $\dim X$  と書く. また,  $X$  の  $k$  単体の集合を  $X_k$  と表記する.

次に, 点群データから単体複体を構成する手順を説明する. より具体的には, 本研究における「穴」の抽出では, 単体複体の一種である Čech 複体を構成する.

$\mathcal{P}$  上の Čech 複体は次のように定義される.

**Definition 4.** (Čech 複体 (Čech complex)) 点群  $\mathcal{P} \in \mathbb{R}^N$  を考える. それぞれの  $x_i$  の周りに半径  $r$  の球

$$(2.2) \quad B_r(x_i) = \{x \in \mathbb{R}^N; \|x - x_i\| \leq r\},$$

を置く. ここで  $\|x\|$  はユークリッドノルムである. Čech 複体  $\mathcal{C}(\mathcal{P}, r)$  とは,  $\mathcal{P}$  上の単体複体で, それを構成する単体が

$$(2.3) \quad \{x_{i_0}, \dots, x_{i_k}\} \in \mathcal{C}(\mathcal{P}, r) \Leftrightarrow \bigcap_{j=0}^k B_r(x_{i_j}) \neq \emptyset$$

で定義されるものである.

すなわち点群  $\mathcal{P}$  の周りに置いた半径  $r$  の球の関係 (重なり) から単体を構成する. 半径  $r$  を変



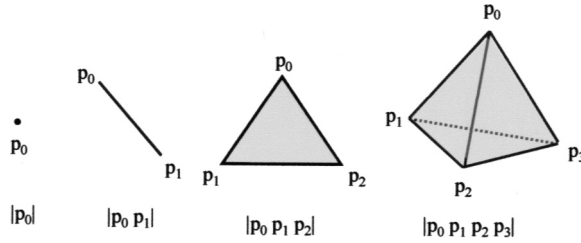


図 2. 単体 (simplex) の例. 図中の影をつけた部分は中身が詰まっていることを示している.

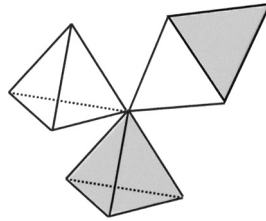


図 3. 単体複体 (simplicial complex) の例. 図 2 と同様に, 影をつけた部分は中空ではなく中身の詰まった図形であることを示す.

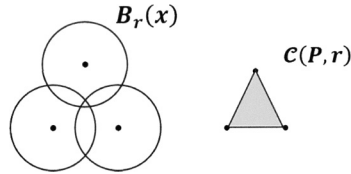


図 4. Čech 複体の例. 左パネル:  $x_i$  の周りの半径  $r$  の球の和集合  $B$ , 右パネル: 対応する Čech 複体  $C$ .

化させたときの Čech 複体の振舞を図 1 に示す. 図 1 からわかる通り, 球の和集合と対応する Čech 複体は連続変形によって互いに移りあえる. これは, トポロジーの基本的定理の一つである脈体定理 (Nerve Theorem) から導かれる (e.g., Kozlov, 2008). つまり, 点群の連結性をモデル化した半径  $r$  の球の和集合  $\bigcap_{j=0}^k B_r(x_{ij})$  と単体複体  $C(P, r)$  は, 連続変形によって互いに変形できることが保証される (ホモトピー同値). したがって, 単体複体には穴に関する情報が保存される. このようにして, 点群データから代数操作が可能な「穴」の表現が実現される.

Čech 複体はデータ  $\mathcal{P}$  と半径  $r$  によって決まる. 半径  $r$  が小さいと Čech 複体は元のデータ点群  $\mathcal{P}$  と同じ離散構造を持つ. 一方  $r$  を増大させるとデータ点は球によって連結し始め,  $r$  を十分大きく取ると最終的に全体が連結する (図 4). つまり前述のように,  $r$  はデータ点間の連結性を扱うための「解像度」とみなすことができる. 半径  $r_1 < r_2 < \dots < r_n$  に対応して Čech 複体の増大列  $C(\mathcal{P}, r_1) \subset C(\mathcal{P}, r_2) \subset \dots \subset C(\mathcal{P}, r_n)$  が得られる. 一般にこのような単体複体の増大列をフィルトレーションと呼ぶ.

実際の応用では, Čech 複体と共通の性質を持ちつつ計算機上での扱いに適した alpha 複体を使用されることが多い. Alpha 複体は, ヴォロノイ分割 (Voronoi tessellation) を用いた Čech 複体の改良版である. ここでは alpha 複体の厳密な導入はせず, 図 5 で概念のみ示すにとどめる. 以下の解析では alpha 複体を採用する.

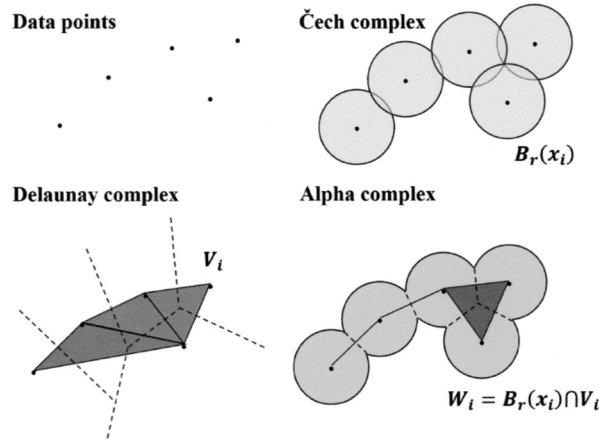


図5. 点データ集合の例(左上), その Čech 複体(右上), Delauney 複体(左下), alpha 複体(右下) (Kono et al., 2020).

### 2.1.3 ホモロジー

次に, 得られた単体複体から「穴」構造を代数的に算出する方法を説明する. 穴構造は, ホモロジー群として表現される. ホモロジー (homology) 群は, 穴に対応する幾何学的意味を持つ代数的な対象である. ホモロジー群の代数的計算方法は複数あるが, その中の1つについて, 以下に概説する.

単体複体  $X$  に含まれる全ての頂点の集合を  $V = \{1, \dots, m\}$  とする.  $X$  に含まれる全ての  $k$  単体の集合を  $X_k$  とする.  $X_k$  に含まれるある  $k$ -単体  $\sigma_k = |v_0, \dots, v_k|$  の頂点に順序を設定する. つまり, 頂点集合に順序  $v_0 < \dots < v_k$  を設定する. このようにして向きを指定した  $k$ -単体を  $\langle \sigma_k \rangle \equiv \langle v_0, \dots, v_k \rangle$  とする. この向きを指定した  $k$ -単体  $\langle \sigma_k \rangle$  を基底として整数  $\mathbb{Z}$  を係数とするベクトル空間を導入する.

$$(2.4) \quad C_k(X) \equiv \left\{ \sum_{\sigma_k \in X_k} a_{\sigma_k} \langle \sigma_k \rangle \mid a_{\sigma_k} \in \mathbb{Z} \right\},$$

ここで,  $k > \dim X$  あるいは  $k < 0$  ならば  $C_k(X) = 0$  である.

単体複体の穴を定義するため, サイクル構造を導入する. サイクル構造とは, 結合した単体を辿って1周回って元に戻ってくる構造である.  $k$ 次元の穴は,  $X_k$  に含まれる  $k$ 次元の単体を組み合わせてできるサイクル構造で,  $X_{k+1}$  に含まれる  $k+1$ 単体の境界となっていない構造のことを指す. 従って, サイクル構造を抽出するには,  $X_k$  から境界のない  $k$ 単体の集合を見つければよい. このように, 穴構造の抽出には単体の境界を算出することが必要となる. 向きが指定された  $k$ -単体  $\langle \sigma_k \rangle = \langle v_0, \dots, v_k \rangle$  の境界は, 以下で定義される境界作用素  $\partial_k$  を適用することで得られる.

**Definition 5.** (単体の境界作用素) 境界作用素  $\partial_k$  は

$$(2.5) \quad \partial_k \langle v_0, \dots, v_k \rangle \equiv \sum_{i=0}^k (-1)^i \langle v_0, \dots, \check{v}_i, \dots, v_k \rangle,$$

$$(2.6) \quad \partial_k \sum_{\sigma_k \in X_k} a_{\sigma_k} \langle \sigma_k \rangle = \sum_{\sigma_k \in X_k} a_{\sigma_k} \partial_k \langle \sigma_k \rangle$$

と定義される(ここで  $\tilde{v}_i$  は  $i$  番目の成分を取り除いたことを表す).

前述したように, 単体複体  $X$  に含まれる  $k$  次元の穴は,  $X$  に含まれる境界のない  $k$ -単体の集まりで,  $X$  に含まれる  $k+1$ -単体の集まりの境界になっていないもののことをいう. サイクルに対応する  $X$  に含まれる境界のない  $k$ -単体複体  $c$  の集合は,

$$(2.7) \quad \text{Ker } \partial_k := \{c \in C_k(X) \mid \partial_k(c) = 0\}$$

与えられる. また,  $X$  に含まれる  $k+1$ -単体の境界の集合は,

$$(2.8) \quad \text{Im } \partial_{k+1} := \{c \in C_k(X) \mid c = \partial_{k+1}(c'), c' \in C_{k+1}(X)\}$$

与えられる. 従って,  $k$  次元のサイクルで  $k+1$  次元の単体複体の境界でないものは, これらの商  $\text{Ker } \partial_k / \text{Im } \partial_{k+1}$  をとることで与えられる.

このように, 単体複体  $X$  に存在する  $k$  次元の穴構造に対応する数学的意味を持つ,  $k$  次ホモロジー群は以下のように定義される.

**Definition 6.** (ホモロジー (homology)) 商ベクトル空間

$$(2.9) \quad H_k(X) = \text{Ker } \partial_k / \text{Im } \partial_{k+1}$$

は単体複体  $X$  の  $k$  次ホモロジー群である. なお  $k=0$  のホモロジー群  $H_0$  と Hubble パラメーター  $H_0$  とを混同しないように注意されたい.

このように, データ  $\mathcal{D}$  に内在する穴の情報は, その Čech 複体  $\mathcal{C}(\mathcal{D}, r)$  のホモロジー群を計算することで抽出できる. しかし, この方法では穴のスケール情報を抽出することはできず, またある半径  $r_2$  で穴が出現し, 半径  $r_3$  で消滅するといった穴のパーシステンス (persistence) についての検証をすることもできない. こういった情報を抽出するためには, パーシステントホモロジー群を導入する必要がある.

### 2.1.4 パーシステントホモロジー群

穴の生成消滅の情報を持つパーシステントホモロジー群も, ホモロジー群と同様に代数的演算によって求められる. そのためにまず, 半径  $r$  を増大させ, 新たな単体に加わった半径を記録する. また, ちょうど  $t$  番目 ( $0 \leq t \leq n$ ) の単体が追加された半径  $r_t$  を用いた場合の単体複体を  $X^t := \mathcal{C}(\mathcal{D}, r_t)$  として記録する. このようにして, 単体の発生回数  $n$  の数の半径  $r_t$  と単体複体  $X^t$  のセットが得られる. 次に, この  $r_t$  と  $X^t$  を用いて, 半径  $r$  の増大によって生成・消滅する穴構造を代数的に算出することを考える.  $X^t \subset X^{t+1}$  であることから, 穴の消滅は必ずその穴を境界とする高次の単体複体の誕生により生じる. 従って, 図形構造の変化を捉える半径としては, 単体の発生時刻を保持すれば十分である. そこで, 単体が発生した時刻に, その単体を要素とする基底空間を設定する. 例えば,  $r_t$  で  $k$  単体  $\langle \sigma_k \rangle$  が誕生したとすると, それは以下のように表現される.

$$(2.10) \quad e_{\langle \sigma_k \rangle} = \begin{pmatrix} 0, & 0, & \dots, & \langle \sigma_k \rangle, & \dots, & 0, & 0 \\ 1 & 2 & & t & & n-1 & n \end{pmatrix}$$

単体複体  $X^t$  は, このような基底空間で表現された単体の和として代数的に表現することができる.

ホモロジー群の算出と同様に, 上で与えた基底空間で表現された単体複体からサイクルや境界の発生・消滅を算出することで, 穴構造の発生・消滅を抽出することを考える. その際に重要な点は, 境界に対応する単体・単体複体の発生とサイクルのような図形構造に対応する単体



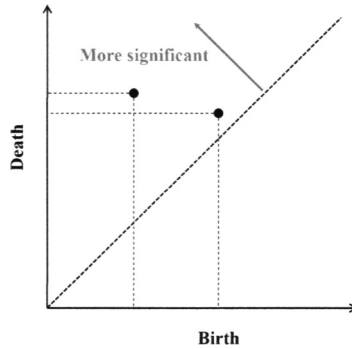


図 6. パーシステンス図の模式図. 対角線近くに位置するデータ点は位相的にはノイズと見なされる一方, 対角線から遠い点はデータ点群の持つ有意でロバストな構造を表す.

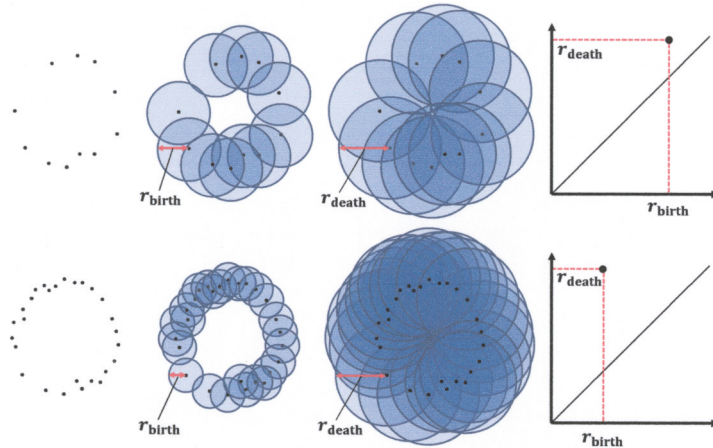


図 7. 点データの密度がパーシステンス図上の構造の生成(birth)および消滅(death)に与える影響を表した概念図 (Kono et al., 2020). 左から右の順に, 点データ, データ上に半径  $r_{\text{birth}}$  の球(ball)を描いた図, 半径  $r_{\text{death}}$  の球を描いた図, および対応する  $H_1$  のパーシステンス図を示す. 上段の 4 つの図が低密度データ点の場合, 下段が高密度の場合である.

点是对角線よりも上に存在する. 定義により, 対角線の近くにある点の「寿命」は短く, 誕生後すぐに死滅することが示されている. 対照的に, 対角線から遠く離れた点は長く生存できる穴(生成元)を表す. 図 6 はこれを端的に示している. 本研究は銀河分布の空間構造を議論するため, 以下では誕生と死滅に対応する半径を  $b$  と  $d$  ではなく  $r_{\text{birth}}$  および  $r_{\text{death}}$  と表記する. 直感的には,  $r_{\text{birth}}$  は  $H_k$  を構成する点の平均間隔を表し,  $r_{\text{death}}$  は構造のサイズを表している.

パーシステンス図の重要な特性の 1 つについて言及する. リング状の構造を持つ点群データを考える. 図 7 ではこのような点群データにおける 2 つの場合のパーシステンス図が示されている. 上のパネルはデータが低密度の場合のパーシステントホモロジー群と対応するパーシステンス図, 下のパネルは高密度の場合の同じ図を示している. 低密度のデータは, 実際には図 7 の高密度データの部分集合となっている. つまり, 2 つのデータはリング構造に関して基本的に同じ情報を共有している. ただし, 高密度データの  $r_{\text{birth}}$  and  $r_{\text{death}}$  の間隔は低密度

データの間隔よりも小さい。これは、2つのデータが同じ構造を持つ場合、データの密度が高い(サンプリングレートが高い)ほど検出された特徴が有意に長い寿命を持つ可能性が高くなることを示唆する。図 7 では低密度データの  $r_{\text{birth}}$  が大きく  $r_{\text{death}}$  に近いことが見て取れる。これは  $r_{\text{birth}}$  が上記のデータ点間の平均距離の指標であることの直接的な帰結である。対照的に、 $r_{\text{death}}$  は注目する構造のサイズの指標であるためほとんど変化しない。この事実は、次章以降で示すシミュレーションおよび観測データの解析で検出された特徴の有意性についての議論と関連する。

### 2.1.6 ボトルネック距離

様々な応用において、異なる 2つのパーシステンス図  $D_1$  と  $D_2$  の差異を比較する方法が必要になる。パーシステンス図はベクトルとならないため、この比較には注意が必要である。これを可能にする方法の 1つがボトルネック距離(bottleneck distance)である (e.g., Edelsbrunner and Harer, 2010; Kerber et al., 2017; Cohen-Steiner et al., 2007; Chazal et al., 2008)。

**Definition 9.** (ボトルネック距離(bottleneck distance)) ボトルネック距離は

$$(2.15) \quad d_b(D_1, D_2) \equiv \inf_m \sup_{x_1 \in |D_1|} \|x_1 - m(x_1)\|_\infty$$

と定義される。ここで  $\|x\|_\infty$  はチェビシェフ距離(Chebyshev distance)  $L^\infty \equiv \max\{|x_1|, |x_2|\}$  で、 $|D|$  はパーシステンス図  $D$  に対角線  $(x, x)$  上を重なりを許して埋め尽くす無限個の生成元の集合  $\Delta$  を加えた集合  $|D| \equiv D \cup \Delta$ ,  $m$  は  $|D_1|$  と  $|D_2|$  の間の全単射である。対角線の多重集合  $\Delta$  は、対応する相手のない生成元の対応相手として用意された、寿命が 0 の生成元の集合である。これによって、2つのパーシステンス図の生成元の個数が異なるような場合でも直接比較が可能となる。

ボトルネック距離の概念を図 8 に示す。2つのパーシステンス図が完全に同一である場合、ボトルネック距離は最小値となる。ここでパーシステンス図の距離の安定性を定義する。

**Definition 10.** (パーシステンス図の距離の安定性) ハウスドルフ距離  $D_H$  を

$$(2.16) \quad D_H = \max \left\{ \sup_{x \in X} (x, Y), \sup_{y \in Y} (y, X) \right\}$$

とする。パーシステンス図の距離  $d$  が安定とは、ある定数  $L$  が存在し、任意の点集合  $X = \{x_i\}$ ,  $Y = \{y_j\}$  に対し

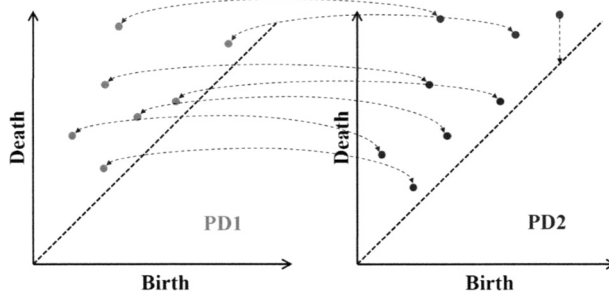


図 8. 2つのパーシステンス図間のボトルネック距離を計算する方法の概念。他方のパーシステンス図に対応する点がない点は、対角線に存在するとみなして計算する。

$$(2.17) \quad d(D_1, D_2) \leq LD_H(X, Y)$$

が成り立つことをいう (Lipschitz 連続性)。

ポトルネック距離はこの意味での安定性が保証されており (Cohen-Steiner et al., 2007), パーシステンス図の検証のために用いることができる。

## 2.2 パーシステンス図の逆解析

2点相関関数による従来の BAO 解析では, データにおける宇宙論的な構造の位置や形状を特定することはできない. これは, この種の解析が相関の統計的な特徴のみを表現するため, データの持つ平均的な特徴しか抽出できないからである. TDA では, 得られたパーシステンス図の各生成元と対応する図形構造を特定することができる. このような逆解析は, 解析結果の物理的議論を進める上で有用である. 半径  $r$  についてのフィルトレーションを通じて, 対応する単体複体  $\sigma(r)$  の位相的情報は保持される. この場合, 逆解析の目的は特異的に長寿命な穴 (生成元)  $(r_{\text{birth},i}, r_{\text{death},i})$  (つまりパーシステンス図に現れた有意な点に対応するデータ点群の特徴) に対応する単体複体  $(\sigma(r_{\text{birth},i}), \sigma(r_{\text{death},i}))$  を見つけることである. 例えば, 穴が消滅する際の穴を構成する単体複体やそのときの半径  $r_{\text{death},i}$  から, その穴の特徴的な構造や位置, スケールといった情報が推定できる. 本研究では R TDA パッケージを用いて逆解析を行った. R TDA では, 逆解析だけでなく, 異なるパーシステンス図の間の距離, パーシステンス図にある穴構造の存在についての検定などを実施することができる. R TDA はまた C++ で書かれている位相的データ解析アルゴリズム GUDHI (The GUDHI Project, 2015), Dionysus (Morozov, 2007), PHAT (Bauer et al., 2012) のインターフェースも与える.

分析対象である 3次元点群データから検出された 3次元の穴 (シェル) 構造の逆解析のために, R パッケージ SCHU (<https://github.com/xinxuyale/SCHU>; Xu et al., 2019) を用いた. SCHU は Dionysus での位相的データ解析に最適化されている. これにより, パーシステンス図を元に検出された生成元に対応する特徴的な図形構造の位置と形状を特定することが可能になる.

逆解析法の研究は急速に発展しているため, 他のさまざまな洗練された方法が提案されている (たとえば Obayashi et al., 2018 を参照されたい). 様々な方法の比較は本研究の範囲外であり, 今後の研究で行う予定である.

## 3. 宇宙論的シミュレーションを用いた解析法の性能評価

### 3.1 データ

本研究で開発した方法の性能を検証するため, まず一連の暗黒物質のみの宇宙論的  $N$  体シミュレーションを実行し, データを生成した. この計算では公開  $N$  体シミュレーションコード Gadget-2 (Springel, 2005) を使い, 初期条件は Crocce et al. (2006) の方法を用いて生成した. 初期条件は赤方偏移  $z = 20$  に設定し, シミュレーションボックスは通常の BAO スケール  $\sim 150$  Mpc よりも十分に大きい一辺 2 Gpc に設定した. BAO 検出のためには質量分解能は重要でないため, シミュレーションボックスには  $256^3$  個の粒子を配置した. これは, 粒子あたりの質量  $6.4 \times 10^{13} M_\odot$  ( $M_\odot$  は太陽質量) に対応する. このシミュレーションには冷たい暗黒物質 (cold dark matter: CDM) 粒子しか含まれていないが, 本研究で注目する時代は光子-バリオン脱結合のかなり後であり, また見る必要があるのは重力だけなので, BAO の特徴は初期条件にのみ刻印しておけば一般的に十分である. 初期ゆらぎのパワースペクトルは, 線型ボルツマン方程式を数値的に解くコード CLASS (Lesgourgues, 2011) を用いて計算している. これを元に  $z = 0$  でのシミュレーションのスナップショットを取り, パワースペクトル解析を行うこ

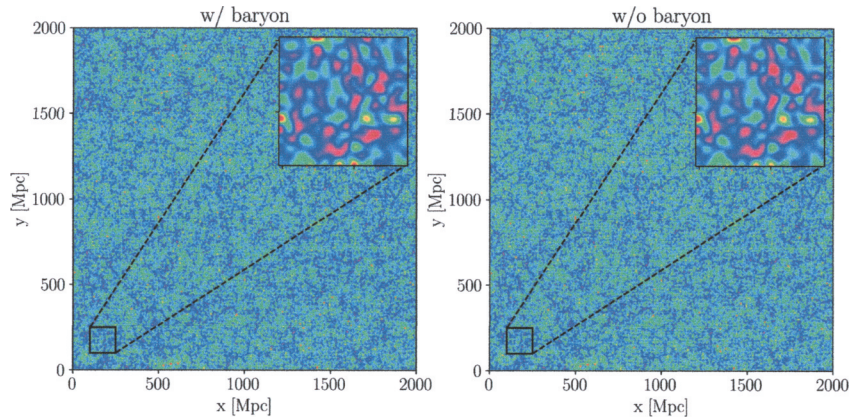


図 9. 赤方偏移  $z = 0$  でのバリオンあり(右)およびなし(左)の場合のシミュレーションデータ。インサートは 100–250 Mpc のボックス領域の拡大図を表す。密度揺らぎは  $z$  軸方向に 100 Mpc の厚さの領域を射影しており、見やすくするために  $\text{arcsinh}$  をとった値を示している。

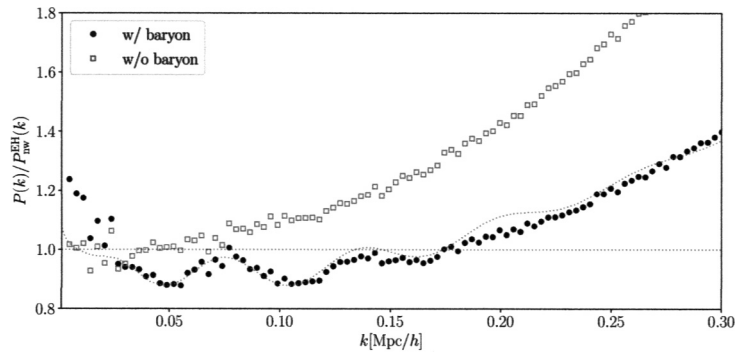


図 10. シミュレーションによる  $z = 0$  でのパワースペクトル (Kono et al., 2020). Eisenstein and Hu (1998) のバリオン振動のない線型パワースペクトルで規格化されている。黒丸はバリオン振動ありのシミュレーションデータ、点線は同じ宇宙論的パラメーターのセットを用い、ハローフィットモデル (Takahashi et al., 2012) によって計算した解析解である。比較のため、バリオン振動なしのシミュレーションデータを四角で示す。バリオンありと比べ、 $k < 0.2 \text{ Mpc}/h$  において振動がみられない。

とによってデータに BAO が確かに存在することが確認できる。

コントロールサンプルとして、BAO なしで初期パワースペクトルを置き換えたまったく同じシミュレーションセットを生成した。これは初期パワースペクトルを計算する際、物質の総質量(CDM+ バリオン)を変更せずにバリオンの量を極限まで減らすことで実現している。すなわち、通常モデルでは  $\Omega_B = 0.049$  であるが、コントロールサンプルは  $\Omega_B = 0.002204$  とした。以降便宜上両者を BAO あり/なし、やバリオンあり/なしと表現し区別することにする。図 9 はバリオンがある場合とない場合のシミュレーションのスナップショットを示している。どちらのシミュレーションセットにも同じランダムシードを使用しているため、目視によって違いを確認することは非常に難しい。一方で、図 10 は両者のシミュレーションセットのバ



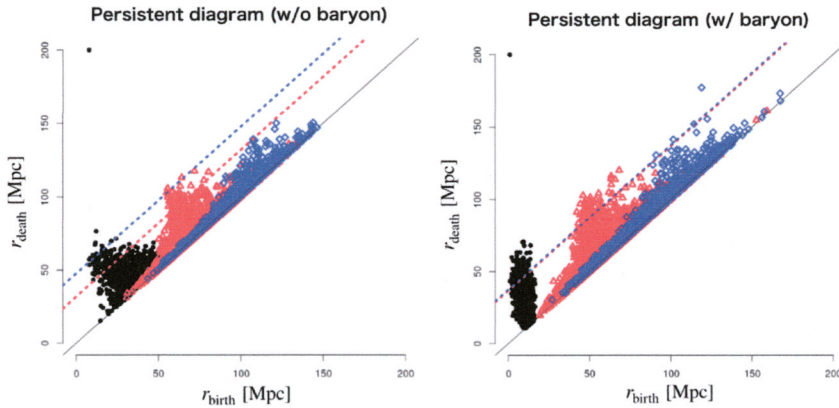


図 11. シミュレーションデータのパーシステンス図. 黒い点, 赤い三角形, 青いひし形はそれぞれ  $H_0$ ,  $H_1$ ,  $H_2$  を表す. 黒い対角線は  $r_{\text{birth}} = r_{\text{death}}$  である. 左パネルはバリオンなし, 右パネルはバリオンありシミュレーションの結果を示している.

ワースペクトルを示している. それぞれのワースペクトルは見やすさのために, BAO の振動的な特徴を平滑化したワースペクトル (Eisenstein and Hu, 1998) によって規格化されている. 予想されるとおり, 得られたコントロールサンプルのワースペクトルは BAO の信号を示さない. 本論文では, 計算機システム上の制約のため  $256^3$  粒子の親サンプルから 2000 個の粒子をランダムに抽出したサブサンプルを用いた. 今後はコードを並列化すると同時に計算機をアップグレードすることで, はるかに大きなデータセットが処理可能になる予定である.

### 3.2 シミュレーションデータの解析結果

シミュレーションデータのパーシステンス図を検証しよう. 検出された信号の統計的有意性を評価するには, 信頼区間を計算する方法が必要になる. しかし, パーシステンス図の信頼区間推定は自明な議論ではなく, いくつかの方法が提案されている (e.g., Fasy et al., 2014). そのひとつとして, ここではボトルネック距離を使用する方法を採用する. 本研究ではまず, ブートストラップリサンプリングでサンプルのセットを生成する. ここから対を抽出し, その 2 つのパーシステンス図の間の距離をボトルネック距離で測ることによって信頼区間を推定する.

シミュレーションで得られたパーシステンス図を図 11 に示す. 左のパネルはバリオンなし, 右のパネルはバリオンありの結果を表す. 黒い点, 赤い三角形, 青いひし形はそれぞれ 0 次 ( $H_0$ ), 1 次 ( $H_1$ ), 2 次 ( $H_2$ ) のパーシステントホモロジー群に対応する. 赤と青の点線は, それぞれ  $H_1$  と  $H_2$  の 90-% 信頼区間である. これらパーシステンス図の信頼区間はリサンプリング回数  $N_{\text{boot}} = 30$  のブートストラップリサンプリングによる方法で計算したものである. この研究では, 対応する信頼区間の上にある点を統計的に有意であるとみなす.

図 11 の右パネルで示されているように, バリオンありデータから得られた  $H_2$  には 4 個の有意な生成元, すなわち 3 次元の穴 (青いひし形) が検出された. 一方バリオンなしデータでは対応するスケールの有意な穴は検出されなかった.

バリオンありのサンプルで検出された有意な穴の  $r_{\text{birth}}$  と  $r_{\text{death}}$  を表 1 に示す. 検出された中空の穴の平均半径は  $r_{\text{death}}$  は  $150.16 \pm 8.46$  [Mpc] であった. 検出個数は 4 つだが, 取得された  $\bar{r}_{\text{death}}$  は, BAO で期待される半径と一致する. 図 12 に検出された生成元に対応するデータ点群の構造の位置と形状を示す. この 3 次元構造は, 中空の穴となっている.

表 1. シミュレーションデータで検出された  $p < 0.2$  の穴の  $r_{\text{birth}}$  と  $r_{\text{death}}$ .

No.	$r_{\text{birth}}$ [Mpc]	$r_{\text{death}}$ [Mpc]
1	116.34	130.65
2	127.42	141.80
3	140.17	151.70
4	144.60	176.47

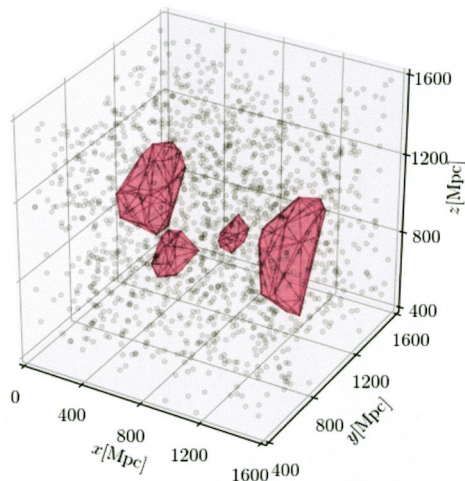


図 12. 逆解析の例. シミュレーションデータにおいて最も優位性の高い 4 個の  $H_2$  に対応する実空間での構造を示す. 網掛け部分は  $H_2$  に対応する構造のすべての構成粒子の凸包であり, 中空の構造となっている.

一方  $H_1$  に属する生成元(赤い三角形)については, 双方のデータで重要な特徴が得られた. この生成元は 2 次元のループ構造に対応する. バリオンありのデータでは  $p < 0.2$  の 17 個の有意な  $H_1$  が検出された. これら有意な生成元の平均  $r_{\text{death}}$  は  $99.00 \pm 2.26$  [Mpc] である. バリオンなしデータの  $H_1$  では 34 個の生成元が検出され,  $\bar{r}_{\text{death}} = 100.49 \pm 3.24$  [Mpc] となった.  $H_1$  で検出された生成元に対応する図形構造の特徴的スケールの一致は, バリオンの存在によって影響されない 2 次元のループのような構造があることを示唆する. 大規模構造のフィラメントはこのような構造の候補の可能性がある. これとは対照的に,  $\bar{r}_{\text{birth}}$  はバリオンの有無で違いがあり, バリオンありで  $46.13 \pm 2.24$  [Mpc], バリオンなしで  $\bar{r}_{\text{birth}} = 62.34 \pm 2.84$  [Mpc] となった. この解析から, ループを構成する銀河の間の距離はバリオンの存在によって減少することが分かる.

そして,  $H_0$  に属する生成元(黒い点)についてはバリオンの有無でパーシステンス図上の分布に明らかな違いが見られた  $H_0$  は連結成分を現し, 銀河団やフィラメントなどの 1 次元的な連続構造に対応する. バリオンなしのデータでは  $r_{\text{birth}}$  は対角線に沿って幅広く分布するのに対し, バリオンありでは  $r_{\text{birth}}$  は低い値に集中している. バリオンありの場合のような  $r_{\text{birth}}$  が局在したような分布は, 連結構造がある特定の半径で誕生したことを意味する. このことは, 銀河団やフィラメント構造を構成する銀河の近某銀河間距離が均等であり, 従って特定の  $r_{\text{birth}}$  で一気に結合して連結構造を形成することを意味する. 一方で,  $r_{\text{death}}$  の分布はバリオンあり・なしに関わらず広く分布していた. このことは, 銀河団やフィラメント間の距離はバリ

オンの影響を受けないことを示唆する．このように  $H_0$  のパーシステンス図の特徴は，バリオンの存在が銀河団やフィラメント構造内の近傍銀河間距離に何らかの制約を与える効果を持つことを示唆する．更なる解析を行うことで， $H_0$  と  $H_1$  のパーシステンス図上の特徴を物理的に解釈できることが期待される．

#### 4. Sloan Digital Sky Survey Data Release 14 による解析

##### 4.1 SDSS 銀河探査データ

本研究では，Sloan Digital Sky Survey (SDSS) Data Release 14 (DR14) の extended Baryon Oscillation Spectroscopic Survey (eBOSS) (Dawson et al., 2016) 大規模構造カタログ (Ata et al., 2018) のクエーサーデータを用いた．これは SDSS-IV (Blanton et al., 2017) の一部である．観測されたバンド(波長帯)は  $u, g, r, i$ , および  $z$  である．天球上の探査領域は  $2044 \text{ deg}^2$ ，赤方偏移は  $0.8 < z < 2.2$  の範囲である．有効面積は北銀河半球(NGC)で  $1288 \text{ deg}^2$ ，南銀河半球で  $995 \text{ deg}^2$  となっている．クエーサーサンプルは Myers et al. (2015) によって選択されたデータを使用した．このサンプルにおける BAO の存在は Ata et al. (2018) によって検証され，すでに確認されている．

サンプル全体には 147,000 個のクエーサーが含まれているが，本研究ではシミュレーションデータと同様 2000 個の銀河をランダム抽出して用いた．

##### 4.2 SDSS データ解析結果

SDSS DR14 データのパーシステンス図を図 13 に示す．図 13 より，SDSS DR14 データでは  $H_2$  に属する 4 つの有意な生成元すなわち「シェル」が検出された．平均の  $r_{\text{death}}$  は  $146.6 \pm 2.0 \text{ [Mpc]}$  である(表 2)．有意な生成元の数 は 4 個だが， $\bar{r}_{\text{birth}}$  は BAO から予想される半径の期待値と一致している． $H_2$  の逆解析の結果を図 14 に示す．

この解析では， $H_1$  に属する 19 個の生成元が有意な穴(ループ)として検出されている．これらの生成元の平均  $r_{\text{death}}$  は  $101.82 \pm 3.54 \text{ [Mpc]}$  となり，シミュレーションデータで得られた  $\bar{r}_{\text{birth}}$  と一致している．また観測データの  $r_{\text{birth}}$  は  $57.92 \pm 3.05 \text{ [Mpc]}$  となった．そして  $H_0$  ホモロジーの分布は，興味深いことにバリオンなしシミュレーションの分布と類似している．これら低次元ホモロジーの解釈には今後のより詳細な研究が待たれる．

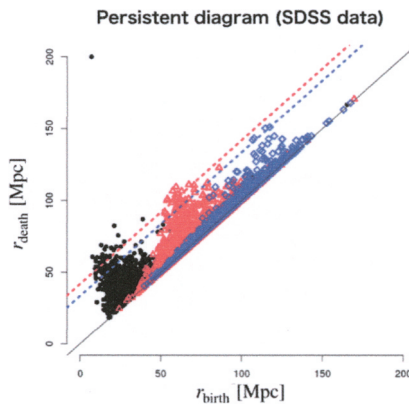
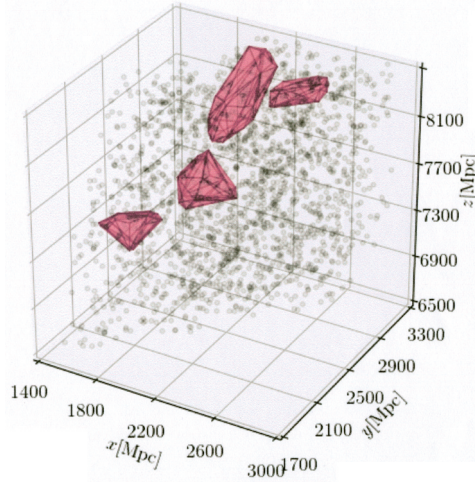


図 13. SDSS データのパーシステンス図．シンボルは図 11 と同様である．

表 2. 観測データ (SDSS DR14) で検出された  $p < 0.2$  の  $r_{\text{birth}}$  および  $r_{\text{death}}$ .

No.	$r_{\text{birth}}$ [Mpc]	$r_{\text{death}}$ [Mpc]
1	107.00	141.98
2	107.83	143.30
3	114.86	150.20
4	117.56	150.98

図 14. 実際の銀河探索 (SDSS) データで検出された  $H_2$  に属する 4 つの有意な生成元に対応する実空間構造. 網掛け部分の意味は図 12 と同様.

## 5. まとめと議論

本研究では、銀河分布の BAO 信号検出における位相的データ解析の有用性を検証した。本研究の位相的データ解析で対象としたパーシステントホモロジー群は、様々なスケールに対応した単体複体の集合列によって定義されたフィルトレーションを構成する。この性質はデータに内在する穴のスケールと形状を分析するのに特に有用である。

まず  $N = 256^3$  の  $N$  体シミュレーションを用い、パーシステントホモロジー群に対するバリオンの影響を検証した。バリオンの物理を入れた場合とない場合のシミュレーションで、パーシステンス図は大幅に異なることが示された。バリオンありのデータからは 2 次のパーシステンス図から 4 つの有意な生成元 (穴構造) が検出されたが、バリオンなしのデータからは検出されなかった。また  $H_1$  と  $H_0$  に属する生成元のパーシステンス図上での分布にも明らかな違いが見られた。検出された 1 次元の穴の  $\bar{r}_{\text{death}}$  は同程度の値を取るが、 $\bar{r}_{\text{birth}}$  は大きく異なる。同様に 0 次元の穴の  $\bar{r}_{\text{birth}}$  は、バリオンの有無で大きく異なる傾向を示した。

次に、実際の観測データとして SDSS のクエーサー探索から抽出した高々 2000 サブサンプルで BAO 信号を検出することに成功した。得られた  $\bar{r}_{\text{death}}$  は  $H_2$  に対して  $146.6 \pm 2.0$  [Mpc] であり、BAO スケールと一致する。このことは、銀河分布の大規模構造検出においてパーシステントホモロジー群の計算コストが相関関数よりもはるかに安価であることを意味し、BAO の宇宙論的分析の新たな可能性を与える。

本研究の方法による BAO 検出は、伝統的な 2 点相関関数による解析と好対照をなす。2 点

相関関数の推定には、高密度にサンプリングされた銀河データが必要とされるが、位相的データ解析は～2000個の銀河のスパースサンプリングされたデータでBAOのような弱い信号を検出できる。これは、位相的データ解析を活用することで、スパースにサンプリングされた銀河探査も宇宙論的研究に使用できることを意味する。このことは次世代の宇宙論的銀河調査のデザインと戦略に新しい方向性を与えると期待される。

さらに、従来の解析と比較した位相的データ解析の著しい利点は、逆解析によりBAO構造を可視化できることである。パーシステンス図の逆分析が持つ機能は、従来の解析とは一線を画す明確な特長であり、強調に値する。たとえば本研究の直接的発展として、BAO信号に寄与する実空間での銀河分布構造を簡単に特定できるため、伝統的方法に比べさまざまな宇宙論的議論がずっと簡単に可能になる。ただし、最も単純な位相的データ解析はノイズに対して頑健でないという弱点があり、今後のより詳細な応用ではこの点を克服する必要がある。ロバストな位相的データ解析の方法は現在も研究されており（たとえば Vishwanath et al., 2022）。将来の研究ではその成果を取り入れられると期待できる。

## 謝 辞

本稿について大変貴重な改善点の指摘および助言を頂いた査読者のお2人、および担当編集委員の島谷健一郎氏には心からの感謝を申し上げる。池田思朗、福水健次、栗木哲の各氏からは本研究に関する貴重な議論および有益な示唆を頂いた。ここに深く感謝する。本研究はJSPS科学研究費助成事業(19H05076, 20H01932, 21H01128, 21J23611)、新学術領域研究「加速宇宙」(15H05890 および 16H01096)、住友財団2018年度基礎科学研究助成金(180923)、および統計数学研究所の共同利用「データサイエンス的方法による銀河進化研究の新展開」の補助を受けて行った。

## 付 録

### A. 宇宙論の基礎

#### A.1 Friedmann-Lemaître-Robertson-Walker 計量とスケール因子

一般相対性理論は、時空がその幾何学的構造によって特徴付けられることを示し、微分幾何学が使用するべき基本的な枠組みであることを示した。微分幾何において、多様体上の局所的性質を記述する基本的な概念が線素(line element)あるいは計量(metric)である。Friedmann-Lemaître-Robertson-Walker (FLRW) 計量は、均一で等方的な時空の(局所的)幾何構造を表すために提唱された計量で、宇宙論のモデルとして広く用いられている。

$$(A.1) \quad ds^2 = g_{\mu\nu} dx_\mu dx_\nu = -c^2 dt^2 + a^2(t) \left[ \frac{dr^2}{1 - Kr^2} + r^2 (d\theta^2 + \sin^2 \theta d\phi^2) \right]$$

ここで  $g_{\mu\nu}$  が計量テンソル、 $t$  は時刻、 $r, \theta, \phi$  は極座標の動径距離および角度方向の座標を表す。また  $K$  はガウス曲率、 $a(t)$  はスケール因子である。スケール因子は宇宙膨張による空間のスケールの変化を表現しており、宇宙論の慣習上  $a_0 \equiv a(t_0) = 1$  ( $t_0$ : 現在の宇宙年齢)と規格化して用いられる。ここでは  $K$  は(長さ) $^{-2}$ の次元を持つように定義を採用する。この場合  $r$  は長さの次元を持ち、 $a(t)$  は無次元になる。

スケール因子を導入すると、宇宙の進化による天体間の距離の変化から宇宙膨張による「自明な」距離の変化を除いた正味の変化分を表現することができる。即ち、距離をあらわすベクトルを  $\vec{r}$  とすると

$$(A.2) \quad \vec{r} = a(t)\vec{x}$$

と書ける。この  $\vec{x}$  は宇宙膨張の影響を除いた距離の指標であり、共動座標 (comoving coordinate) と呼ばれる。

## A.2 赤方偏移と宇宙論パラメータ

過去のある時刻  $t = t_{\text{em}}$  で共動座標  $(r, \phi, \theta) = (r_e, 0, 0)$  で放射された光が、時刻  $t = t_0$  において原点  $(r, \phi, \theta) = (0, 0, 0)$  に位置する観測者に到達する状況を考える。相対性理論では、光はヌル測地線 (null geodesic) に沿って移動する。ヌル測地線とは  $ds^2 = 0$  を満たす経路のことで、最小作用の原理で実現する軌道と考えてよい。式 (A.1) を用いれば、距離と時間を

$$(A.3) \quad \frac{cdt}{a(t)} = \frac{dr}{\sqrt{1 - Kr^2}}$$

のように関係付けられる。宇宙膨張の光への影響を議論するため、変数  $r$  を

$$(A.4) \quad \chi \equiv \int_0^{r_{\text{em}}} \frac{dr}{\sqrt{1 - Kr^2}}$$

によって座標距離  $\chi$  に、時間  $t$  を

$$(A.5) \quad \eta \equiv \int^t \frac{dt'}{a(t')}$$

によって共形時間  $\eta$  に変換すると、光の満たすべき方程式は

$$(A.6) \quad c(\eta_{\text{obs}} - \eta_{\text{em}}) = \chi$$

となる。ここで  $\eta_{\text{obs}}$  は光が観測者に到達した時刻を表す。

時刻  $\eta = \eta_0$  および  $\eta = \eta_0 + \delta\eta_0$  で放出された光が、それぞれ  $\eta_1$  および  $\eta_1 + \delta\eta_1$  で観測者に到達したとする。式 (A.6) の右辺は  $\eta$  によらないことから

$$(A.7) \quad \delta\eta_0 = \delta\eta_1$$

すなわち

$$(A.8) \quad \frac{\delta t_0}{a(t_0)} = \frac{\delta t_1}{a(t_1)}$$

となる。時間間隔  $\delta t_0$  および  $\delta t_1$  において、光の位相が保存することから、

$$(A.9) \quad a(t_0)\nu_0 = a(t_1)\nu_1 \iff \frac{\lambda_0}{a(t_0)} = \frac{\lambda_1}{a(t_1)}$$

が得られる。ここで  $\nu$  は光の振動数で、波長とは  $\lambda\nu = c$  の関係がある。

膨張する宇宙では  $a(t_0) > a(t_1)$  なので、観測される時刻での波長は放射のときよりも長くなる、即ち  $\lambda_0 > \lambda_1$  となる。この波長の伸びが宇宙膨張による赤方偏移と呼ばれている。赤方偏移  $z$  は

$$(A.10) \quad z \equiv \frac{\lambda_0 - \lambda_1}{\lambda_1} = \frac{a(t_0)}{a(t_1)} - 1$$

と定義される。

ここで、上記のようにスケール因子  $a(t)$  を  $a(t_0) = 1$  となるよう規格化する。このとき  $a(t)$  と  $z$  との関係は

$$(A.11) \quad z = \frac{1}{a} - 1,$$

あるいは

$$(A.12) \quad a(t) = \frac{1}{1+z}$$

となる。

### A.3 フリードマン方程式と宇宙論パラメータ

ここまでの準備, およびアインシュタイン方程式 (Einstein's equation) を用いれば, 膨張する宇宙を記述するための「運動方程式」を導くことができる。アインシュタイン方程式とは, 時空の歪みとエネルギー密度を関連付けた一般相対性理論の基本方程式であり,

$$(A.13) \quad R^{\mu\nu} - \frac{1}{2}Rg^{\mu\nu} + \Lambda g^{\mu\nu} = \frac{8\pi G}{c^4}T^{\mu\nu}$$

と書かれる。ここで  $R^{\mu\nu}$  はリッチテンソル (Ricci tensor),  $R \equiv R^\alpha_\alpha$  はそのトレースでリッチスカラー (Ricci scalar),  $G$  はニュートンの重力定数,  $\Lambda$  は宇宙定数である。詳細は相対性理論の文献に譲るが, リッチテンソルは計量  $g_{\mu\nu}$  とその微分量の関数であり, 時空の歪みを表現する 2 階の共変テンソルである。計量が FLRW メトリックの場合, 時空の運動方程式は

$$(A.14) \quad \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho}{3} - \frac{c^2 K}{a^2} + \frac{c^2 \Lambda}{3}$$

$$(A.15) \quad \frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{c^2 \Lambda}{3},$$

となる。これはフリードマン方程式 (Friedmann equations) と呼ばれている。ここで  $\rho$  は宇宙の密度,  $p$  は圧力を現す。式 (A.14) はアインシュタイン方程式 (式 (A.13)) の 00 成分, 式 (A.15) はトレースから求められる。

さらに, 宇宙論パラメータと呼ばれる以下の量を定義すると, 宇宙膨張の物理量依存性がより具体的に記述できる。

- ハッブルパラメータ:

$$(A.16) \quad H(t) \equiv \frac{\dot{a}(t)}{a(t)}.$$

これは時間依存する量であるが, 観測的には現在値  $H_0$  が用いられることが多い。また無次元化した  $h = H_0/100$  もよく用いる。

- 密度パラメータ:

$$(A.17) \quad \Omega_M(t) \equiv \frac{\rho(t)}{\rho_c(t)} \equiv \frac{8\pi G\rho(t)}{3H^2(t)},$$

ここで  $\rho_c$  は臨界密度パラメータと呼ばれる量で, 現在の宇宙では  $\rho_c = 1.88 \times 10^{-29} h^2$  [ $\text{g cm}^{-3}$ ] と測定されている。

- 無次元宇宙項パラメータ:

$$(A.18) \quad \Omega_\Lambda(t) \equiv \frac{c^2 \Lambda}{3H^2(t)}.$$

- 曲率パラメータ:



$$(A.19) \quad \Omega_K(t) \equiv -\frac{c^2 K}{a(t)^2 H^2(t)}.$$

式(A.16)–(A.19)を用いると、式(A.14)は

$$(A.20) \quad \Omega_M(t) + \Omega_\Lambda(t) + \Omega_K(t) = 1$$

と表せる。これは一般の宇宙年齢  $t$  で成り立つので、現在の宇宙年齢  $t_0$  では

$$(A.21) \quad \Omega_{M0} + \Omega_{\Lambda0} + \Omega_{K0} = 1$$

となる。

導入で述べているように、本論文では *Planck* 衛星による CMB 観測で推定された宇宙論パラメータ  $h = H_0/100$  [ $\text{km s}^{-1} \text{Mpc}^{-1}$ ] = 0.6766,  $\Omega_{\Lambda,0} = 0.6889$ ,  $\Omega_{M,0} = 0.3111$ ,  $\ln(10^{10} A_s) = 3.047$  を採用している。曲率パラメータ  $\Omega_{K0} = 0$  である。これらの値は、現在の宇宙は空間的に平坦で、宇宙定数(または暗黒エネルギー)により加速膨張している宇宙モデルを強く示唆する。パラメータ  $A_s$  は宇宙初期のホライズンスケールの揺らぎの振幅を決めるもので、シミュレーションデータを生成する際にこの値を採用している。

## 参 考 文 献

- Appleby, S., Chingangbam, P., Park, C., Yogendran, K. P. and Joby, P. K. (2018). Minkowski tensors in three dimensions: Probing the anisotropy generated by redshift space distortion, *Astrophysical Journal*, **863**(2), 200.
- Ata, M., et al. (2018). The clustering of the SDSS-IV extended Baryon Oscillation Spectroscopic Survey DR14 quasar sample: First measurement of baryon acoustic oscillations between redshift 0.8 and 2.2, *Monthly Notices of the Royal Astronomical Society*, **473**(4), 4773–4794.
- Bag, S., Mondal, R., Sarkar, P., Bharadwaj, S. and Sahni, V. (2018). The shape and size distribution of H II regions near the percolation transition, *Monthly Notices of the Royal Astronomical Society*, **477**(2), 1984–1992.
- Bag, S., Mondal, R., Sarkar, P., Bharadwaj, S., Choudhury, T. R. and Sahni, V. (2019). Studying the morphology of H I isodensity surfaces during reionization using Shapefinders and percolation analysis, *Monthly Notices of the Royal Astronomical Society*, **485**(2), 2235–2251.
- Bauer, U., Kerber, M. and Reininghaus, J. (2012). PHAT, a software library for persistent homology, <https://bitbucket.org/phat-code/phat/src/master/>.
- Beisbart, C., Dahlke, R., Mecke, K. and Wagner, H. (2002). Vector- and tensor-valued descriptors for spatial patterns, *Morphology of Condensed Matter*, Lecture Notes in Physics, Volume 600, 238–260, Springer, New York.
- Bernardeau, F., Colombi, S., Gaztañaga, E. and Scoccimarro, R. (2002). Large-scale structure of the universe and cosmological perturbation theory, *Physics Reports*, **367**(1-3), 1–248.
- Biagetti, M., Cole, A. and Shiu, G. (2021). The persistence of large scale structures I: Primordial non-Gaussianity, *Journal of Cosmology and Astroparticle Physics*, **2021**(4), 061.
- Blanton, M. R., et al. (2017). Sloan Digital Sky Survey IV: Mapping the Milky Way, nearby galaxies, and the distant universe, *Astronomical Journal*, **154**(1), 28.
- Boerner, G. and Mo, H. (1989). A percolation analysis of cluster superclustering, *Astronomy and Astrophysics*, **224**(1-2), 1–9.
- Boissonnat, J.-D., Chazal, F. and Yvinec, M. (2018). *Geometric and Topological Inference*, Cambridge Texts in Applied Mathematics, Cambridge University Press, Cambridge.
- Chazal, F., Cohen-Steiner, D., Glisse, M., Guibas, L. J. and Oudot, S. (2008). Proximity of persistence modules and their diagrams, Research Report, RR-6568, INRIA, 29.



- Chen, Z., Xu, Y., Wang, Y. and Chen, X. (2019). Stages of reionization as revealed by the Minkowski functionals, *Astrophysical Journal*, **885**(1), 23.
- Chingangbam, P., Ganesan, V., Yogendran, K. P. and Park, C. (2017). On Minkowski functionals of CMB polarization, *Physics Letters B*, **771**, 67–73.
- Cohen-Steiner, D., Edelsbrunner, H. and Harer, J. (2007). Stability of persistence diagrams, *Discrete & Computational Geometry*, **37**, 103–120.
- Cole, A. and Shiu, G. (2018). Persistent homology and non-Gaussianity, *Journal of Cosmology and Astroparticle Physics*, **2018**(3), 025.
- Croce, M., Pueblas, S. and Scoccimarro, R. (2006). Transients from initial conditions in cosmological simulations, *Monthly Notices of the Royal Astronomical Society*, **373**(1), 369–381.
- Daley, D. J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes. Vol. I, Probability and Its Applications*, 2nd ed., Springer-Verlag, Berlin.
- Dawson, K. S., et al. (2016). The SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Overview and early data, *Astronomical Journal*, **151**(2), 44.
- de Lapparent, V., Geller, M. J. and Huchra, J. P. (1986). A slice of the universe, *Astrophysical Journal*, **302**, L1–L4.
- Edelsbrunner, H. and Harer, J. (2010). *Computational Topology — An Introduction*, American Mathematical Society, Providence, Rhode Island, <http://www.ams.org/bookstore-getitem/item=MBK-69>.
- Edelsbrunner, H., Letscher, D. and Zomorodian, A. (2002). Topological persistence and simplification, *Discrete & Computational Geometry*, **28**(4), 511–533.
- Efstathiou, G. and Silk, J. (1983). The formation of galaxies, *Fundamentals of Cosmic Physics*, **9**, 1–138.
- Eisenstein, D. J. and Hu, W. (1998). Baryonic features in the matter transfer function, *Astrophysical Journal*, **496**(2), 605–614.
- Eisenstein, D. J., et al. (2005). Detection of the baryon acoustic peak in the large-scale correlation function of SDSS luminous red galaxies, *Astrophysical Journal*, **633**(2), 560–574.
- Eisenstein, D. J., Seo, H.-J., Sirko, E. and Spergel, D. N. (2007). Improving cosmological distance measurements by reconstruction of the baryon acoustic peak, *Astrophysical Journal*, **664**(2), 675–679.
- Elbers, W. and van de Weygaert, R. (2019). Persistent topology of the reionization bubble network — I. Formalism and phenomenology, *Monthly Notices of the Royal Astronomical Society*, **486**(2), 1523–1538.
- Fang, W., Li, B. and Zhao, G.-B. (2017). New probe of departures from general relativity using Minkowski functionals, *Physical Review Letters*, **118**(18), 181301.
- Fasy, B. T., Lecci, F., Rinaldo, A., Wasserman, L., Balakrishnan, S. and Singh, A. (2014). Confidence sets for persistence diagrams, *The Annals of Statistics*, **42**(6), 2301–2339.
- Feldbrugge, J., van Engelen, M., van de Weygaert, R., Pranav, P. and Vegter, G. (2019). Stochastic homology of Gaussian vs. non-Gaussian random fields: Graphs towards Betti numbers and persistence diagrams, *Journal of Cosmology and Astroparticle Physics*, **2019**(9), 052.
- Ganesan, V. and Chingangbam, P. (2017). Tensor Minkowski functionals: First application to the CMB, *Journal of Cosmology and Astroparticle Physics*, **2017**(6), 023.
- Gay, C., Pichon, C. and Pogosyan, D. (2012). Non-Gaussian statistics of critical sets in 2D and 3D: Peaks, voids, saddles, genus, and skeleton, *Physical Review D*, **85**(2), 023011.
- Giri, S. K. and Mellema, G. (2021). Measuring the topology of reionization with Betti numbers, *Monthly Notices of the Royal Astronomical Society*, **505**(2), 1863–1877.
- Gott, I., Gott, J. R., III, Melott, A. L. and Dickinson, M. (1986). The sponge-like topology of large-scale structure in the universe, *Astrophysical Journal*, **306**, 341–357.
- Gott, I., Gott, J. R., III, Weinberg, D. H. and Melott, A. L. (1987). A quantitative approach to the topology of large-scale structure, *Astrophysical Journal*, **319**, 1–8.

- Gott, I., Gott, J. R., III, et al. (1989). The topology of large-scale structure. III. Analysis of observations, *Astrophysical Journal*, **340**, 625–646.
- Hikage, C., et al. (2003). Minkowski functionals of SDSS galaxies I: Analysis of excursion sets, *Publications of the Astronomical Society of Japan*, **55**, 911–931.
- Joby, P. K., Chingangbam, P., Ghosh, T., Ganesan, V. and Ravikumar, C. D. (2019). Search for anomalous alignments of structures in Planck data using Minkowski tensors, *Journal of Cosmology and Astroparticle Physics*, **2019**(1), 009.
- Junaid, M. and Pogosyan, D. (2015). Geometrical measures of non-Gaussianity generated from single field inflationary models, *Physical Review D*, **92**(4), 043505.
- Kerber, M., Morozov, D. and Nigmatov, A. (2017). Geometry helps to compare persistence diagrams, *Journal of Experimental Algorithmics*, **22**, 1.4.
- Kerscher, M. and Tikhonov, A. (2010). Morphology of the local volume, *Astronomy and Astrophysics*, **509**, A57.
- Kerscher, M., et al. (1997). Minkowski functionals of Abell/ACO clusters, *Monthly Notices of the Royal Astronomical Society*, **284**(1), 73–84.
- Kitaura, F.-S. and Angulo, R. E. (2012). Linearization with cosmological perturbation theory, *Monthly Notices of the Royal Astronomical Society*, **425**(4), 2443–2454.
- Kono, K. T., Takeuchi, T. T., Cooray, S., Nishizawa, A. J. and Murakami, K. (2020). A study on the baryon acoustic oscillation with topological data analysis, arXiv:2006.02905.
- Kozlov, D. (2008). *Combinatorial Algebraic Topology*, Springer, Berlin.
- Lesgourgues, J. (2011). The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview, arXiv:1104.2932
- Libeskind, N. I., et al. (2018). Tracing the cosmic web, *Monthly Notices of the Royal Astronomical Society*, **473**(1), 1195–1217.
- Ling, C., Wang, Q., Li, R., Li, B., Wang, J. and Gao, L. (2015). Distinguishing general relativity and  $f(R)$  gravity with the gravitational lensing Minkowski functionals, *Physical Review D*, **92**(6), 064024.
- Liu, J., Bird, S., Zorrilla Matilla, J. M., Hill, J. C., Haiman, Z., Madhavacheril, M. S., Petri, A. and Spergel, D. N. (2018). MassiveNuS: Cosmological massive neutrino simulations, *Journal of Cosmology and Astroparticle Physics*, **2018**(3), 049.
- Ma, C.-P. and Bertschinger, E. (1995). Cosmological perturbation theory in the synchronous and conformal Newtonian gauges, *Astrophysical Journal*, **455**, 7–52.
- Martinez, V. and Saar, E. (2001). *Statistics of the Galaxy Distribution*, CRC Press, London.
- Matsubara, T. (1994). Analytic expression of the genus in a weakly non-Gaussian field induced by gravity, *Astrophysical Journal*, **434**, L43–L46.
- Matsubara, T. (1996). Statistics of isodensity contours in redshift space, *Astrophysical Journal*, **457**, 13–17.
- 松原 隆彦 (1996). ジーナース統計を用いた宇宙の大構造の定量化, 日本物理學會誌, **51**(4), 266–273.
- Matsubara, T. (2003). Statistics of smoothed cosmic fields in perturbation theory. I. Formulation and useful formulae in second-order perturbation theory, *Astrophysical Journal*, **584**(1), 1–33.
- Matsubara, T. (2010). Analytic Minkowski functionals of the cosmic microwave background: Second-order non-Gaussianity with bispectrum and trispectrum, *Physical Review D*, **81**(8), 083505.
- Matsubara, T. and Suto, Y. (1996). Nonlinear evolution of genus in a primordial random Gaussian density field, *Astrophysical Journal*, **460**, 51–58.
- Matsubara, T. and Yokoyama, J. (1996). Genus statistics of the large-scale structure with non-Gaussian density fields, *Astrophysical Journal*, **463**, 409–419.
- Matsubara, T., Hikage, C. and Kuriki, S. (2022). Minkowski functionals and the nonlinear perturbation theory in the large-scale structure: Second-order effects, *Physical Review D*, **105**(2), 023527.
- Mecke, K. R., Buchert, T. and Wagner, H. (1994). Robust morphological measures for large-scale

- structure in the universe, *Astronomy and Astrophysics*, **288**, 697–704.
- Morozov, D. (2007). Dionysus, a C++ library for computing persistent homology, <https://www.mrzv.org/software/dionysus/>.
- Myers, A. D., et al. (2015). The SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Quasar target selection, *Astrophysical Journal Supplement Series*, **221**(2), 27.
- Obayashi, I., Hiraoka, Y. and Kimura, M. (2018). Persistence diagrams with linear machine learning models, *Journal of Applied and Computational Topology*, **1**, 421–449.
- Pâris, I., et al. (2018). The Sloan digital sky survey quasar catalog: Fourteenth data release, *Astronomy and Astrophysics*, **613**, A51.
- Parroni, C., Cardone, V. F., Maoli, R. and Scaramella, R. (2020). Going deep with Minkowski functionals of convergence maps, *Astronomy and Astrophysics*, **633**, A71.
- Peebles, P. J. E. (1980). *The Large-scale Structure of the Universe*, Princeton University Press, New Jersey.
- Peebles, P. J. E. and Yu, J. T. (1970). Primeval adiabatic perturbation in an expanding universe, *Astrophysical Journal*, **162**, 815–836.
- Planck Collaboration, Ade, P. A. R., et al. (2016). Planck 2015 results. XVII. Constraints on primordial non-Gaussianity, *Astronomy and Astrophysics*, **594**, A17.
- Pogosyan, D., Gay, C. and Pichon, C. (2009). Invariant joint distribution of a stationary random field and its derivatives: Euler characteristic and critical point counts in 2 and 3D, *Physical Review D*, **80**(8), 081301.
- Pranav, P., et al. (2017). The topology of the cosmic web in terms of persistent Betti numbers, *Monthly Notices of the Royal Astronomical Society*, **465**(4), 4281–4310.
- Pranav, P., et al. (2019a). Topology and geometry of Gaussian random fields I: on Betti numbers, Euler characteristic, and Minkowski functionals, *Monthly Notices of the Royal Astronomical Society*, **485**(3), 4167–4208.
- Pranav, P., et al. (2019b). Unexpected topology of the temperature fluctuations in the cosmic microwave background, *Astronomy and Astrophysics*, **627**, A163.
- Santos, L., Wang, K. and Zhao, W. (2016). Probing the statistical properties of CMB B-mode polarization through Minkowski functionals, *Journal of Cosmology and Astroparticle Physics*, **2016**(7), 029.
- Schmalzing, J. and Buchert, T. (1997). Beyond genus statistics: A unifying approach to the morphology of cosmic structure, *Astrophysical Journal*, **482**(1), L1–L4.
- Schmalzing, J. and Gorski, K. M. (1998). Minkowski functionals used in the morphological analysis of cosmic microwave background anisotropy maps, *Monthly Notices of the Royal Astronomical Society*, **297**(2), 355–365.
- Shandarin, S. F. and Zeldovich, I. B. (1983). Topology of the large-scale structure of the universe, *Comments on Astrophysics*, **10**, 33–45.
- Shiraishi, M., Hikage, C., Namba, R., Namikawa, T. and Hazumi, M. (2016). Testing statistics of the CMB B-mode polarization toward unambiguously establishing quantum fluctuation of the vacuum, *Physical Review D*, **94**(4), 043506.
- Springel, V. (2005). The cosmological simulation code GADGET-2, *Monthly Notices of the Royal Astronomical Society*, **364**(4), 1105–1134.
- Spurio Mancini, A., et al. (2018). 3D cosmic shear: Numerical challenges, 3D lensing random fields generation, and Minkowski functionals for cosmological inference, *Physical Review D*, **98**(10), 103507.
- Sullivan, J. M., Wiegand, A. and Eisenstein, D. J. (2019). The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Evolution of higher-order correlations demonstrated with Minkowski functionals, *Monthly Notices of the Royal Astronomical Society*, **485**(2), 1708–1719.

- Sunyaev, R. A. and Zeldovich, Y. B. (1970). Small-scale fluctuations of relic radiation, *Astrophysics and Space Science*, **7**(1), 3–19.
- Takahashi, R., Sato, M., Nishimichi, T., Taruya, A. and Oguri, M. (2012). Revising the halofit model for the nonlinear matter power spectrum, *Astrophysical Journal*, **761**(2), 152.
- The GUDHI Project (2015). GUDHI User and Reference Manual, <http://gudhi.gforge.inria.fr/doc/latest/>.
- Vicinanza, M., Cardone, V. F., Maoli, R., Scaramella, R., Er, X. and Tereno, I. (2019). Minkowski functionals of convergence maps and the lensing figure of merit, *Physical Review D*, **99**(4), 043534.
- Vishwanath, S., Sriperumbudur, B. K., Fukumizu, K. and Kuriki, S. (2022). Robust topological inference in the presence of outliers, arXiv:2206.01795.
- Wasserman, L. (2018). Topological data analysis, *Annual Review of Statistics and Its Application*, **5**(1), 501–532.
- Wiegand, A. and Eisenstein, D. J. (2017). The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Higher order correlations revealed by germ-grain Minkowski functionals, *Monthly Notices of the Royal Astronomical Society*, **467**(3), 3361–3378.
- Wiegand, A., Buchert, T. and Ostermann, M. (2014). Direct Minkowski functional analysis of large redshift surveys: A new high-speed code tested on the luminous red galaxy Sloan Digital Sky Survey-DR7 catalogue, *Monthly Notices of the Royal Astronomical Society*, **443**(1), 241–259.
- Wilding, G., et al. (2021). Persistent homology of the cosmic web — I. Hierarchical topology in  $\Lambda$ CDM cosmologies, *Monthly Notices of the Royal Astronomical Society*, **507**(2), 2968–2990.
- Xu, X., Cisewski-Kehe, J., Green, S. B. and Nagai, D. (2019). Finding cosmic voids and filament loops using topological data analysis, *Astronomy and Computing*, **27**, 34–57.
- Yoshiura, S., Shimabukuro, H., Takahashi, K. and Matsubara, T. (2017). Studying topological structure of 21-cm line fluctuations with 3D Minkowski functionals before reionization, *Monthly Notices of the Royal Astronomical Society*, **465**(1), 394–402.
- Zomorodian, A. and Carlsson, G. (2004). Computing persistent homology, *Proceedings of the Twentieth Annual Symposium on Computational Geometry*, SCG'04, 347–356, Association for Computing Machinery, New York.

## Quantification of Galaxy Distribution with Topological Data Analysis and Detection of the Baryon Acoustic Oscillation

Tsutomu T. Takeuchi<sup>1,2</sup>, Kai T. Kono<sup>1</sup>, Suchetha Cooray<sup>1,3,4</sup>, Atsushi J. Nishizawa<sup>5,6,7</sup>, Koya Murakami<sup>1</sup>, Hai-Xia Ma<sup>1</sup> and Yoh-Ichi Mototake<sup>8</sup>

<sup>1</sup>Division of Particle and Astrophysical Science, Nagoya University

<sup>2</sup>The Institute of Statistical Mathematics

<sup>3</sup>Division of Science, National Astronomical Observatory of Japan

<sup>4</sup>Research Fellow of the Japan Society of the Promotion of Science (PD)

<sup>5</sup>DX Center, Gifu Shotoku Gakuen University

<sup>6</sup>Institute for Advanced Research, Nagoya University

<sup>7</sup>Kobayashi Maskawa Institute, Nagoya University

<sup>8</sup>The Institute of Statistical Mathematics; Now at Graduate School of Data Science, Hitotsubashi University

Galaxies are distributed inhomogeneously in space, with clusters, groups, filaments and voids. This is called the large-scale structure in the Universe. In astronomy, it is customary to refer to the normal matter as baryons. What we directly observe is the large-scale structure made of baryons. The large-scale structure has formed mainly through the gravitational instability, but there is another origin of the structure exists, which is the acoustic oscillation of baryons at the baryon-photon decoupling. This is imprinted on the spatial distribution of galaxies in the Universe, known as the baryon acoustic oscillation (BAO). In this work, we analyzed the spatial distribution of galaxies with a method from the topological data analysis (TDA), in order to examine the BAO signal in the galaxy distribution. The TDA provides a method to treat various types of “holes” in point set data, by constructing the persistent homology group from the geometric structure of data points and handling the topological information of the dataset. We can obtain the information on the size, position, and statistical significance of the holes in the data. A particularly strong point of the persistent homology is that it can classify the holes by their spatial dimension. We first analyzed the simulation datasets with and without the baryon physics to examine the performance of the PH method. We found that the persistent homology is indeed able to detect the BAO signal among the large-scale structures in the Universe: simulation data with baryon physics present a prominent signal from the BAO, while data without baryon physics does not show this signal. Then, we applied the persistent homology to a quasar (a kind of active galaxies) sample at  $z < 1.0$  from extended Baryon Oscillation Spectroscopic Survey in Sloan Digital Sky Survey Data Release 14 (SDSS DR14). We discovered a characteristic hole (a hollow shell) at a scale  $r \sim 150$  [Mpc]. This corresponds exactly to the BAO signature imprinted in the galaxy/quasar distribution. We performed this analysis on a small subsample of 2000 quasars. This clearly demonstrates that the PH analysis is very efficient in finding this type of topological structures even if the sampling is very sparse.

---

Key words: Topological data analysis (TDA), persistent homology, galaxy distribution, the large-scale structure in the Universe, baryon acoustic oscillation.