

# Stochastic Homology of Gaussian vs. non-Gaussian Random Fields:

## Graphs towards Betti Numbers and Persistence Diagrams

**Job Feldbrugge<sup>a,b</sup> Matti van Engelen,<sup>c</sup> Rien van de Weygaert<sup>b</sup>  
Pratyush Pranav<sup>d,b,e</sup> Gert Vegter<sup>c</sup>**

<sup>a</sup>Perimeter Institute for Theoretical Physics,  
31 Caroline St. North, Waterloo, ON N2L 2Y5 Ontario, Canada

<sup>b</sup>Kapteyn Astronomical Institute, University of Groningen,  
P.O. Box 800, 9747 AV Groningen, The Netherlands

<sup>c</sup>Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence, University  
of Groningen, Nijenborgh 9, 9747 AG Groningen, the Netherlands

<sup>d</sup>Univ Lyon, ENS de Lyon, Univ Lyon, CNRS, Centre de Recherche Astrophysique de Lyon  
UMR5574, FV69007, Lyon, France

<sup>e</sup>Technion – Israel Institute of Technology,  
Haifa, 32000, Israel

E-mail: [jfeldbrugge@perimeterinstitute.ca](mailto:jfeldbrugge@perimeterinstitute.ca)

**Abstract.** The topology and geometry of random fields – in terms of the Euler characteristic and the Minkowski functionals – has received a lot of attention in the context of the Cosmic Microwave Background (CMB), as the detection of primordial non-Gaussianities would form a valuable clue on the physics of the early Universe. The virtue of both the Euler characteristic and the Minkowski functionals in general, lies in the fact that there exist closed form expressions for their expectation values for Gaussian random fields. However, the Euler characteristic and Minkowski functionals are summarizing characteristics of topology and geometry. Considerably more topological information is contained in the homology of the random field, as it completely describes the creation, merging and disappearance of topological features in superlevel set filtrations.

In the present study we extend the topological analysis of the superlevel set filtrations of two-dimensional Gaussian random fields by analysing the statistical properties of the Betti numbers – counting the number of connected components and loops – and the persistence diagrams – describing the creation and mergers of homological features. Using the link between homology and the critical points of a function – as illustrated by the Morse-Smale complex – we derive a one-parameter fitting formula for the expectation value of the Betti numbers and forward this formalism to the persistent diagrams. We, moreover, numerically demonstrate the sensitivity of the Betti numbers and persistence diagrams to the presence of non-Gaussianities.

---

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Random fields: definitions, probabilities &amp; singularities</b>	<b>6</b>
2.1	Gaussian random fields	6
2.1.1	Multiscale Field Structure & Fourier Components	8
2.2	Non-Gaussian Random Fields	8
2.3	Critical Points	9
2.3.1	Critical Point Densities	10
<b>3</b>	<b>Homology and Persistence: Topological Structure of Random fields</b>	<b>11</b>
3.1	Random Field Realizations and Manifolds	13
3.2	Filtrations and Superlevel Sets	13
3.3	Homology of Random Fields: Betti numbers	13
3.4	Betti, Euler, and Genus	15
3.5	Persistence Diagrams	17
<b>4</b>	<b>The Graph Formalism &amp; Betti Number Statistics</b>	<b>19</b>
4.1	Morse-Smale Complex	20
4.2	Morse-Smale Graph	23
4.3	The Incremental Algorithm: Graphic Extraction of Homological Information	24
4.4	Approximating the Probability $g$	25
4.5	Numerical Validation of $g$	27
<b>5</b>	<b>Statistics of Persistence diagrams of Random Fields</b>	<b>27</b>
<b>6</b>	<b>Non-Gaussian Random Fields: Persistence &amp; Betti numbers</b>	<b>30</b>
<b>7</b>	<b>Conclusion and Discussion</b>	<b>34</b>
<b>A</b>	<b>Concepts of Homology theory</b>	<b>41</b>
A.1	Simplices	41
A.2	Chains and the Boundary Operator	42
A.2.1	Chain	42
A.2.2	Boundary & Boundary Operator	42
A.3	Homology Groups and Betti Numbers	43
A.4	Example	44

---

## 1 Introduction

The Cosmic Microwave Background (CMB) radiation field is one of the principal cosmological probes, it is the source of revolutionary new insights into the physics of the early Universe and the hitherto unimaginable accurate measurement of cosmological quantities that ushered the era of precision cosmology (see [1–4]). The radiation field is a relic of an early dense phase

of our Universe in which the Universe was filled with a hot photon-electron plasma. The primordial radiation field consists of a near perfect uniform radiation field with a blackbody spectrum with a temperature of  $T_0 = 2.725\text{K}$  [5, 6]. Perhaps the most outstanding property of this radiation field is the tiny temperature fluctuations with an amplitude in the order of  $10^{-5}$ . The statistical properties of these fluctuations represent a true treasure trove of cosmological information, and form a direct reflection of processes that took place in the very early Universe. As evidenced by a range of CMB observations, the CMB temperature fluctuations and the corresponding mass density fluctuations, turn out to be close to a homogeneous and isotropic Gaussian random field [7–12]. The statistical properties of Gaussian random fields have been the subject of a range of fundamental studies [13–16]. Statistically, Gaussian fields are completely described in terms of their autocorrelation function or power spectrum. Cosmological properties such as the curvature of space, the energy content and the expansion rate determine the functional form of the power spectrum, so that one may infer constraints on these from the CMB measurements.

Recent analyses of the Planck satellite data have set stringent limits on the level of possible deviations from Gaussianity of the primordial density field [11, 12, 17]. The detection of non-Gaussian perturbations will have major repercussions for our understanding of the physical processes at the beginning of the Universe, in particular at its inflationary epoch [10, 18–21]. There has been a considerable effort into determining the level and nature of these primordial non-Gaussianities, but these have not led to any significant detection [12]. A considerable fraction of these studies is based on the measurement of the three-point and four-point correlation function, or of their Fourier-space equivalents of the bispectrum and trispectrum. They concentrate on a restricted set of deviations from Gaussianity, most notably the local, enfolded, equilateral and orthogonal templates for the bispectrum.

Most of the present searches for primordial non-Gaussianities have been focused at detecting deviations of an a priori known form. In the light of the failure to detect such signals, an arguably more profitable and objective approach would be to use a probe that is sensitive to generic non-Gaussian perturbations. We base ourselves on the idea that the geometry and topology of the primordial field is potentially one of the most direct and sensitive statistics, as it captures both the morphology, spatial distribution and connectivity between individual topological features in the primordial density field. It is these aspects that are very sensitive to the nature of the random field [see e.g. 22, 23]. In fact, preliminary indications for this might be the mildly unusual behaviour of the Euler characteristic pointed out by Eriksen et al. [24] and Park et al. [17].

Topology is the branch of mathematics that addresses the shapes, boundaries and connectivity of structural features in a field. Traditionally the Minkowski functionals, which include the genus and the Euler characteristic, have been used to study the topology, morphology and geometry of cosmological density fluctuations [15, 25–29]. For the study of Gaussian random fields, these topological and geometric measures have the virtue that there exist closed analytical expressions for their expectation values as a function of the field’s density level [14, 15, 22, 28, 30, 31]. While potential primordial non-Gaussianities might have expressed themselves via deviations from these statistical distribution functions [32, 33], no significant signals have been detected [see 17, 24, 34–40]. Nonetheless, while the Minkowski functionals have been instructive, the topological information contained in them is limited and convolved with geometric information. Moreover, they are not equipped to address the hierarchical aspects of the matter distribution directly. Recent advances in the field of algebraic and computational topology [see e.g. 41, 42] have opened the possibility to infer informa-

tion on the contribution of individual features on the overall structural patterns, in terms of persistent topology and Betti numbers.

In this study we therefore focus on the topological characterization of the cosmic mass (and temperature) distribution in terms of Betti numbers [43] and persistence diagrams [41, 42]. These are homology measures, concepts of algebraic and computational topology, describing in a quantitative manner how features in a manifold are connected through their boundaries [44]. It concerns a language to assess the multiscale nature of the topology of the megaparsec cosmic mass distribution. *Betti numbers* are topological invariants that formalize the topological information content of the cosmic mass distribution by counting features in terms of the number of  $p$ -dimensional holes [42, 45–48]. For two-dimensional manifolds, the zeroth Betti number counts the number of disjoint components, while the first Betti number counts the number of inequivalent loops enclosing troughs. One often does not differentiate between the loops and the troughs they enclose. It is important to appreciate that the homological measures (with the exception of the Euler characteristic) are *fundamentally non-local* as they depend on the correlation and connectivity of critical points in the random field. While homology and the Betti numbers do not fully quantify the topology of a manifold<sup>1</sup>, they extend the information beyond conventional cosmological studies of topology in terms of genus and Euler characteristics.

The profound significance of Betti numbers is underlined by their intimate relationship to the singularity structure and connectivity of the cosmic density field. Van de Weygaert and collaborators [49, 50] introduced the concept of homology and Betti numbers in a cosmological context, in a study of Betti number systematics in a range of weblike spatial galaxy distributions. This was followed up by recent studies that invoked homology in a cosmological context along more systematic and formalized lines [22, 50–52]. The Betti numbers provide a summary of information on the topology of the cosmic mass distribution. A more detailed, extensive and profound representation of related topological information is that of persistence [41, 42]

Persistence describes the formation and demise of individual topological features such as “islands”, and “troughs” in the primordial density field, as we see them connecting and merging with surrounding topological features at different density levels. Introduced by Edelsbrunner [41], persistence describes the density range over which topological features exist and hence reflects its multiscale character<sup>2</sup>. By focussing on the topological significance of features instead of their spatial scale, it provides a natural means of identifying structures over a range of scales without having to invoke artificial filters. Adding to its significance is the fact that it is intimately coupled to the singularity structure of the density field. This reflects the notion, stemming from Morse theory [53, 54], that the existence of and connectivity between topological features is determined by the location and nature of critical points – *i.e.*, the maxima, minima and saddle points – in the scalar field. In other words, persistence provides us with the mathematical language for assessing and analyzing the hierarchy of topological structure in a density field.

Recent years have seen a proliferation of scientific studies invoking persistent topology to characterize the complexity of a large diversity of systems and processes (see [48] for a

---

<sup>1</sup>One can change the topology without changing the homology of a manifold by removing a ‘hole’ in one component of the manifold and introducing a ‘hole’ in another connected component. The full topological information is captured in the homotopy. However, the homotopy is currently computationally out of reach.

<sup>2</sup>Note that persistence can be invoked for any scalar field.

recent review), ranging from brain research [55, 56], materials science [57] to astrophysics and cosmology. Sousbie et al. [51, 58], Shivashankar et al. [59], and Pranav et al. [52] invoke persistence with the purpose to characterize the spine of the cosmic web [60–64] and its connectivity structure. Recently, Kimura et al. [65] determined persistence diagrams for (small) volume-limited samples of the DR12 release of the SDSS galaxy redshift survey in an attempt to characterize the topology of the spatial galaxy distribution, while Xu et al. [66] used persistence to identify individual voids and filaments in heuristic models of the cosmic matter distribution [also see 59]. At a more fundamental level, Codis et al. [67] based their assessment of the connectivity of the nodes of the cosmic web on the persistent characterization of the cosmic web’s spine. The concept of persistence and Betti numbers also offer a natural means of following the evolving topology of the reionization bubble network [68]. In another astrophysical context, they were used to describe the topological structure of interstellar magnetic fields [69].

The main purpose of the present study is the formulation and inference of analytical descriptions and expressions of key topological characteristics, in particular Betti numbers and topological persistence of two-dimensional random fields on  $\mathbb{R}^2$ . For an early report on this study we refer to [70]. To this end, we specifically concentrate on Gaussian random fields, and on possible small non-Gaussian deviations. The accompanying extensive numerical analysis of Gaussian field homology is described in [22, 23]. While fully analytical expressions for the Euler characteristic and the Minkowski functionals of Gaussian random fields exist [14, 15, 22, 28, 30], to date no closed analytical expressions for topological quantities such as Betti numbers have been derived. All indications are that this may remain so [see e.g. 22, 71]. While there are some high level, asymptotic results about Betti numbers for high density thresholds of Gaussian excursion sets in the mathematical literature, these are a consequence of the simple structure of Gaussian fields at these levels. As a consequence, nearly without exception, the information on Betti numbers is obtained indirectly, through numerical and statistical evaluations [22, 29]. The Gaussian Kinematic Formula (GKF) implies that the Minkowski functionals and the Euler characteristic can be expressed as integrals of local functionals [see 16, 22]. This is not true for topological quantities, for which the localization is crucial. Note that the Euler characteristic is an exception, as it can be written as the oscillating sum of the critical point densities. From a geometric perspective, while the Euler characteristic is topological, it can with the Gauss-Bonnet theorem be expressed in terms of a local characterization of the curvature.

Motivated by the benefit of having an insightful and versatile analytical expression for Betti numbers and persistence diagrams, for Gaussian random fields, the present study follows an alternative route for inferring an accurate approximate formula. To this end, we follow a graph theoretical approach to Morse theory, developing path integral expressions via the connectivity of singularities – maxima, minima, and saddle points – in a Gaussian random field [also see 70]. While it is not trivial to convert the Morse formalism into concise formulae such as entailed in the Gaussian Kinematic Formula [16], in this study we demonstrate that the numerically evaluated approximate expressions for the two-dimensional Betti numbers turns out to be remarkably accurate.

Morse theory is the branch of mathematics that studies the singularity structure of a scalar field, *i.e.*, the positions and connectivity of minima, maxima and saddle points. Of fundamental importance is Morse theory’s principal topological tenet that there is a close relationship between the topology of the space and the critical points of any smooth function

on the topological space [42, 53, 54]. Following this realization, Morse theory describes the topology of the space by studying the critical points of a corresponding *Morse function*, *i.e.*, a smooth scalar function defined on the topological space with no degenerate critical points. Of particular interests are the observations that submanifolds defined as the regions where the Morse function is in excess of a particular functional threshold value, *i.e.*, superlevel sets, are topologically equivalent when the interval between the two defining threshold values does not contain any critical point. The important implication of this is that all changes in the topology of a space occur only at critical points.

The close relation between the topology of a manifold and the spatial distribution of singularities finds its expression in the *Morse-Smale* complex [72]. It is the segmentation of space defined by the spatial distribution of the singularities, consisting of regions which connect minima and maxima via the field’s integral lines. To formalize the connections between the singularities in the Morse-Smale complex, we introduce the concept of the *Morse Graph*. The graph is a diagrammatic representation of the connectivity structure of the Morse-Smale complex. We subsequently investigate the topological structure of the field using the *incremental algorithm* [73], which defines a filtration process to follow the addition and removal of topological features as new singularities in the Morse graph are included. In combination with the analytically known field distribution functions for the maxima, minima, and saddles in a Gaussian random field, we are led to the integral expressions for the *Betti numbers* as well as the *persistence diagrams* of Gaussian random fields. By investigating the asymptotic behaviour of the Betti numbers we are then able to infer an accurate fitting function. Note that these analytical expressions have been derived on the basis of a probabilistic calculation, and do not involve and resort to realizations of Gaussian random fields.

Having established accurate analytical expressions for the Betti numbers of Gaussian random fields, and the corresponding integral expressions for persistence diagrams, we subsequently investigate the sensitivity of persistent Betti numbers and persistence diagrams on the presence of (subtle) non-Gaussianities in two-dimensional random fields. In this study, we restrict our numerical study to the local template version of the primordial non-Gaussianities and explore the changes in Betti numbers and persistence diagrams as a function of the non-Gaussianity parameter  $f_{NL}$  [see 10]. The results of this analysis do underline the considerable potential for exploiting these instruments for a more profound topological data analysis of the cosmic microwave background – and other cosmological probes – for assessing the presence of primordial non-Gaussianities. A few earlier studies already indicated this potential [74, 75]. In fact, the recent study by Pranav et al. [76] reports on the detection of possibly anomalous topological signatures in the Planck CMB maps.

In section 2, we introduce basic concepts of (Gaussian) random fields, in particular focusing on the singularities in a random field and their statistical properties in a Gaussian random field. In section 3, we discuss and describe the key concepts of homology, Betti numbers and persistence diagrams. In section 4, we embark on the statistical properties of Betti numbers, following the description of the Morse-Smale complex for Gaussian fields, the introduction of Morse graphs, and the use of the incremental algorithm in conjunction with Gaussian field singularity statistics to derive integral expressions for Betti numbers. We conclude this section with the derivation of an accurate analytical fitting formula for Betti numbers. Extending this analysis to the larger information content of the multiscale topological properties of Gaussian fields represented by persistence diagrams is the topic of section 5. Finally, we turn towards the potential of Betti numbers and persistence diagrams

towards the detection of non-Gaussianities in section 6. The conclusions of this study are summarized and discussed in section 7.

## 2 Random fields: definitions, probabilities & singularities

Throughout this paper we describe the topology of random fields, whereby we largely follow the notation introduced by [15] [also see 77]. A typical example of a random field in the context of cosmology is the three-dimensional density perturbation field

$$\delta\rho(\mathbf{x}, t) = \frac{\rho(\mathbf{x}, t) - \rho_u(t)}{\rho_u(t)}, \quad (2.1)$$

with  $\rho(\mathbf{x}, t)$  the density at location  $\mathbf{x} \in \mathbb{R}^3$  at time  $t$ , and  $\rho_u(t)$  the universal cosmological density for that cosmic epoch. In the present study we consider the primordial temperature fluctuations  $\delta T(\mathbf{x})$  of the current cosmic microwave background (CMB) temperature field  $T(\mathbf{x})$  at sky position  $\mathbf{x} \in \mathbb{S}^2$  with respect to the current global CMB temperature  $T_0 = 2.726\text{K}$  [6],

$$\delta T(\mathbf{x}) = \frac{T(\mathbf{x}) - T_0}{T_0}. \quad (2.2)$$

The temperature field  $T$  is a two-dimensional imprint of the density field  $\rho$  at the time of last scattering when our Universe became neutral. As a consequence, the temperature perturbation  $\delta T$  is a realization of a random field on the 2-sphere  $\mathbb{S}^2$ . However, since our study of random fields is local, we for simplicity restrict the current analysis to random fields  $[f]$  on Euclidean space  $\mathbb{R}^2$ .

Important for the cosmological context of the random fields is the *statistical cosmological principle*, which states that the statistical properties of the cosmic density distribution in our Universe are uniform throughout space<sup>3</sup>. In other words, the distribution function and moments of a field are the same in each direction and at each location, with the direct implication of the ensemble averages being dependent only on the distance between points.

To high accuracy, the primordial density perturbation  $\delta\rho$ , and the CMB temperature fluctuation field  $\delta T$  – which is a reflection of former – are realizations of Gaussian random fields [2, 3], whose statistical character is fully specified by their first and second order moment. In this study, we largely focus on two-dimensional Gaussian random fields. However, the framework extends naturally to more general random fields which include non-Gaussianities. The detection or stringent bounds on primordial non-Gaussianities in the CMB are expected to lead to major improvements in our understanding of the early Universe [see 78–85].

### 2.1 Gaussian random fields

The random fields  $f(\mathbf{x})$  studied in this paper are assumed to be smooth and continuous<sup>4</sup>. We generalize our study to any field  $f(\mathbf{x})$  that is a linear functional  $\mathcal{L}[g]$  of a random field  $g(\mathbf{x})$ ,

$$f(\mathbf{x}) = \mathcal{L}[g; \mathbf{x}]. \quad (2.3)$$

---

<sup>3</sup>Also crucial for the cosmological reality is the *ergodic principle*. Based on this we are able to measure the value of ensemble averages by means of spatial averages: these will be equal to the expectations over an ensemble of Universes. Given the fact that the Universe is unique, and its density distribution the only realization we have of the underlying probability distribution, this is of key significance for the ability to test theoretical predictions for stochastic processes like the cosmic mass distribution with observational reality.

<sup>4</sup>In this section, the fields  $f(\mathbf{x})$  may either be the raw unfiltered field or, without loss of generality, a filtered field  $f_s(\mathbf{x})$ . A filtered field is a convolution with a filter kernel  $W(\mathbf{x}, \mathbf{y})$ ,  $f_s(\mathbf{x}) = \int d\mathbf{y} f(\mathbf{y})W(\mathbf{x}, \mathbf{y})$ .

Examples of such functionals are the value of the field itself at the point  $\mathbf{x}$ , the gradient of the field  $g(\mathbf{x})$ , its Hessian, and more generally a convolution of  $g(\mathbf{x})$  with a kernel function  $h(\mathbf{x})$ ,

$$\begin{aligned} \mathcal{L}[g; \mathbf{x}] &= g(\mathbf{x}); & \mathcal{L}_i[g; \mathbf{x}] &= \frac{\partial g}{\partial x_i}(\mathbf{x}); & \mathcal{L}_{ij}[g; \mathbf{x}] &= \frac{\partial^2 g}{\partial x_i \partial x_j}(\mathbf{x}); \\ \mathcal{L}_\zeta[g; \mathbf{x}] &= \int h(\mathbf{x} - \mathbf{x}_\zeta) g(\mathbf{x}_\zeta) d\mathbf{x}_\zeta. \end{aligned} \quad (2.4)$$

A random field  $f$  is defined by its  $N$ -point joint probability,

$$\mathcal{P}_{\mathbf{x}}(\mathbf{f}) d\mathbf{f} = \text{Prob}[f(\mathbf{x}_i) \in [f_i, f_i + df_i], i = 1, \dots, N], \quad (2.5)$$

the probability that a realization assumes a value in the interval  $[f_i, f_i + df_i]$ , at the locations  $\mathbf{x}_i$  for  $1 \leq i \leq N$ . Throughout this paper, we use the shorthand  $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$  for the  $N$  points  $\mathbf{x}_i$  and the vector  $\mathbf{f} = (f_1, f_2, \dots, f_N)$  for the corresponding field values.

A *Gaussian random field* is fully described by the first and second order moment, as can be seen from the probability distribution of a Gaussian random field with zero mean

$$\begin{aligned} \mathcal{P}_{\mathbf{x}}(\mathbf{f}) &= \frac{\exp[-\mathbf{f}\mathbf{M}^{-1}\mathbf{f}^T/2]}{[(2\pi)^N(\det \mathbf{M})]^{1/2}} \\ &= \frac{\exp\left[-\frac{1}{2} \sum_{i,j} f_i (\mathbf{M}^{-1})_{ij} f_j\right]}{[(2\pi)^N(\det \mathbf{M})]^{1/2}}, \end{aligned} \quad (2.6)$$

with  $\mathbf{M}^{-1}$  the inverse of the  $N \times N$  covariance matrix  $\mathbf{M}$ , *i.e.*,

$$M_{ij} = \langle f(\mathbf{x}_i) f(\mathbf{x}_j) \rangle = \xi(\mathbf{x}_i - \mathbf{x}_j), \quad (2.7)$$

with the angle bracket denoting the ensemble average and the autocorrelation function  $\xi(\mathbf{r})$  [see 14–16]. The covariance matrix  $\mathbf{M}$  is the  $N$ -dimensional generalization of the variance  $\sigma^2$  of the one-point Gaussian distribution,

$$\sigma^2 = \langle f(\mathbf{x}) f(\mathbf{x}) \rangle. \quad (2.8)$$

The prefactor  $[(2\pi)^N(\det \mathbf{M})]^{-1/2}$  in equation (2.6) normalizes the distribution, *i.e.*,

$$\int \mathcal{P}_{\mathbf{x}}(\mathbf{f}) d^N \mathbf{f} = 1. \quad (2.9)$$

Given the fact that cosmological random fields obey the statistical cosmological principle, *i.e.*, the fields are statistically isotropic and homogeneous, the correlation function  $\xi(\mathbf{r})$  only depends on the distance  $r = \|\mathbf{r}\|$  between points,

$$\xi(r) = \xi(\mathbf{r}). \quad (2.10)$$

With respect to the higher order correlation functions, it is straightforward to infer from the distribution (2.6) that they can be expressed in terms of the two-point correlation function.

According to Wick's theorem, the odd correlation functions vanish, while the even correlation function are given by the sum of the product of all possible pairings of points, *i.e.*,

$$\begin{aligned}\langle f(\mathbf{x}_1)f(\mathbf{x}_2)f(\mathbf{x}_3) \rangle &= 0, \\ \langle f(\mathbf{x}_1)f(\mathbf{x}_2)f(\mathbf{x}_3)f(\mathbf{x}_4) \rangle &= \xi(r_{12})\xi(r_{34}) + \xi(r_{13})\xi(r_{24}) + \xi(r_{14})\xi(r_{23}), \\ \langle f(\mathbf{x}_1)f(\mathbf{x}_2)f(\mathbf{x}_3)f(\mathbf{x}_4)f(\mathbf{x}_5) \rangle &= 0,\end{aligned}\tag{2.11}$$

with  $r_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|$ . Finally, it is important to note that when  $f(\mathbf{x})$  is Gaussian random field, so too are the linear functionals  $\mathcal{L}[f]$  of the field. *I.e.* the gradient, Hessian and higher order gradients of a Gaussian field are themselves Gaussian fields.

### 2.1.1 Multiscale Field Structure & Fourier Components

The *multiscale* structure of a Gaussian random field is most transparently described in terms of the Fourier components  $\hat{f}(\mathbf{k})$  of the field  $f(\mathbf{x})$ ,

$$f(\mathbf{x}) = \int \frac{d\mathbf{k}}{(2\pi)^3} \hat{f}(\mathbf{k}) e^{-i\mathbf{k}\cdot\mathbf{x}}.\tag{2.12}$$

In terms of the Fourier components of the field, the second order character of a Gaussian field is fully specified by the Fourier transform of the correlation function  $\xi(r)$ , known as the power spectrum  $P(k)$ ,

$$P(k) = \int \xi(r) e^{-ikr} \frac{dk}{2\pi},\tag{2.13}$$

which can be interpreted as the two point function of the Fourier modes

$$\langle \hat{f}(\mathbf{k})\hat{f}(\mathbf{k}') \rangle = (2\pi)^3 P(k) \delta_D(\mathbf{k} - \mathbf{k}'),\tag{2.14}$$

with the magnitude  $k = \|\mathbf{k}\|$  and the Dirac delta function  $\delta_D$  [see 77, 86]. Note that the correlation function in Fourier space is diagonal, *i.e.*,  $\langle \hat{f}(\mathbf{k})\hat{f}(\mathbf{k}') \rangle = 0$  for  $\mathbf{k} \neq \mathbf{k}'$ .

The power spectrum  $P(k)$  describes the prominence of the Fourier component at scale  $\sim 2\pi/k$  in the field  $f(\mathbf{x})$ , by means of the square of the average amplitude of that component. Cosmologically speaking, this is a factor of central significance: the power spectrum fully specifies the nature of the primordial density and velocity field, out of which all structure in the cosmos emerges, and is itself a direct manifestation of processes in the very early (inflationary) Universe.

## 2.2 Non-Gaussian Random Fields

A Gaussian random field is completely determined by the power spectrum or two-point correlation function. The higher order correlation functions can, with Wick's theorem, be expressed in term of the two-point function. For more general non-Gaussian random fields, the higher order correlation functions take a different form. The presence of these corrections are known as *non-Gaussianities*. The most straightforward one is that of a non-vanishing three-point function, and its Fourier transform, the bispectrum  $B(k_1, k_2, k_3)$ . The bispectrum is defined as

$$\langle \hat{f}(\mathbf{k}_1)\hat{f}(\mathbf{k}_2)\hat{f}(\mathbf{k}_3) \rangle = (2\pi)^2 B(k_1, k_2, k_3) \delta^2(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3),\tag{2.15}$$

and provides a first measure of non-Gaussianities in the density perturbation field. Evidently, a non-Gaussian field may also have deviating higher order correlations. An immediate example is the four-point correlation function and its Fourier transform, the trispectrum  $T(k_1, k_2, k_3, k_4)$ ,

$$\langle \hat{f}(\mathbf{k}_1)\hat{f}(\mathbf{k}_2)\hat{f}(\mathbf{k}_3)\hat{f}(\mathbf{k}_4) \rangle = (2\pi)^2 T(k_1, k_2, k_3, k_4)\delta^2(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4). \quad (2.16)$$

To account for the Gaussian contribution to the four-point function, it is common to subtract the Gaussian four-point function from the trispectrum, yielding the reduced trispectrum. In other words, a deviation of  $T(k_1, k_2, k_3, k_4)$  from the value predicted by Wick's theorem indicates the presence of non-Gaussianities.

### 2.3 Critical Points

The topological structure of a field  $f$  is intimately connected to the character and distribution of the *singular* or *stationary points* in the field, as is illustrated by the Morse-Smale complex (see section 4.1). We identify the *critical* or *stationary points* in a differentiable field  $f$  with the points for which the gradient of  $f$  vanishes, *i.e.*,

$$\nabla f = 0. \quad (2.17)$$

Near a critical point  $\mathbf{x}_0$ , the function  $f(\mathbf{x})$  can be approximated with the Taylor expansion

$$\begin{aligned} f(\mathbf{x}) &= f(\mathbf{x}_0) + (\mathbf{x} - \mathbf{x}_0)^T \nabla f(\mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^T \mathcal{H}(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + \mathcal{O}(\|\mathbf{x} - \mathbf{x}_0\|^3) \\ &= f(\mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^T \mathcal{H}(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + \mathcal{O}(\|\mathbf{x} - \mathbf{x}_0\|^3) \end{aligned} \quad (2.18)$$

with the Hessian matrix field  $\mathcal{H}$ , whose components  $\mathcal{H}_{ij}$  are the second order partial derivatives of the function  $f$ , *i.e.*,

$$\mathcal{H}_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}. \quad (2.19)$$

The function  $f$  is said to be a Morse function if all its critical points are non-degenerate, *i.e.*, the Hessian matrix  $\mathcal{H}$  is non-singular in the critical points of  $f$ . According to Morse lemma, a Morse function  $f$  is, in the neighbourhood of a critical point, approximately a quadratic form

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \frac{1}{2}(\lambda_1 y_1^2 + \lambda_2 y_2^2 + \dots + \lambda_d y_d^2) + \mathcal{O}(\|\mathbf{y}\|^3), \quad (2.20)$$

with the coordinates  $\mathbf{y} = \mathbf{x} - \mathbf{x}_0 = (y_1, y_2, \dots, y_d)$  defined by the (normalized) eigenvectors of the Hessian matrix  $\mathcal{H}$  and its non-zero eigenvalues  $(\lambda_1, \lambda_2, \dots, \lambda_d)$  [see 53, 54].

The character of a *singular point* is determined by the signature of the eigenvalues  $(\lambda_1, \lambda_2, \dots, \lambda_d)$ , *i.e.*, the number of positive and negative eigenvalues. The index  $i$  of a singularity is defined as the number of negative eigenvalues. When all eigenvalues are negative, the singularity is a (local) *maximum* (index  $i = d$ ). When all eigenvalues are positive, it is a (local) *minimum* (index  $i = 0$ ). The critical points for which the Hessian has a number of positive and negative eigenvalues are saddle points (index  $0 < i < d$ ). In two dimensions, critical points with index 0, 1 and 2 are minima, saddle points and maxima.

One of the implications of the Morse lemma is that non-degenerate singular points are isolated, *i.e.*, two critical points of a Morse function are separated by a finite distance.

### 2.3.1 Critical Point Densities

The density of critical points of a random field, at function value  $\nu = f/\sigma$ , can be evaluated with Rice's formula [13–16, 87]. For a two-dimensional random field, the density of minima  $\mathcal{N}_0$ , saddle points  $\mathcal{N}_1$  and maxima  $\mathcal{N}_2$  are given by the three-dimensional integral

$$\mathcal{N}_i(\nu) = \iiint P(f = \nu\sigma, f_1 = f_2 = 0, f_{11}, f_{12}, f_{22}) \begin{vmatrix} f_{11} & f_{12} \\ f_{12} & f_{22} \end{vmatrix} df_{11} df_{12} df_{22}, \quad (2.21)$$

with the notation

$$f_i = \frac{\partial f}{\partial x_i}, \quad f_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}, \quad (2.22)$$

for  $i, j = 1, 2$ . The integral for  $\mathcal{N}_0, \mathcal{N}_1$ , and  $\mathcal{N}_2$  is performed over the configurations  $(f_{11}, f_{12}, f_{22})$  corresponding to minima, saddle points and maxima. Explicitly, the integral is performed over configurations  $(f_{11}, f_{12}, f_{22})$  for which the eigenvalues  $\lambda_1, \lambda_2$  of the Hessian  $\mathcal{H}$ , *i.e.*,

$$\lambda_1 = \frac{1}{2} \left[ f_{11} + f_{22} + \sqrt{4f_{12}^2 + (f_{11} - f_{22})^2} \right], \quad (2.23)$$

$$\lambda_2 = \frac{1}{2} \left[ f_{11} + f_{22} - \sqrt{4f_{12}^2 + (f_{11} - f_{22})^2} \right], \quad (2.24)$$

are either both positive  $\lambda_1 \geq \lambda_2 > 0$  for minima, of mixed signature  $\lambda_1 > 0 > \lambda_2$  for saddle points, or both negative  $0 > \lambda_1 \geq \lambda_2$  for the maxima. For statistically isotropic random fields, the three-dimensional integral can be reduced to a two-dimensional integral, by expressing the three second order derivatives  $f_{11}, f_{12}, f_{22}$  in terms of two rotationally invariant variables. In terms of the two variables

$$J_1 = \lambda_1 + \lambda_2 = f_{11} + f_{22}, \quad (2.25)$$

$$J_2 = (\lambda_1 - \lambda_2)^2 = (f_{11} - f_{22})^2 + 4f_{12}^2, \quad (2.26)$$

as proposed by Pogosyan et al. [32], the number density  $\mathcal{N}_i(\nu)$  of critical points in a Gaussian field of index  $i$  takes the form

$$\mathcal{N}_i(\nu) = \iint \frac{|J_1^2 - J_2|}{8\pi^2 \sqrt{1 - \gamma^2}} \exp \left[ -\frac{1}{2} J_1^2 - J_2 - \frac{(\nu + \gamma J_1)^2 \sigma^2}{2(1 - \gamma^2)} \right] dJ_1 dJ_2. \quad (2.27)$$

In the expression above, we use the transformation of the measure

$$df_{11} df_{12} df_{22} = \pi |\lambda_1 - \lambda_2| d\lambda_1 d\lambda_2. \quad (2.28)$$

The integration domains for the minima, saddle points and maxima in the integral expression are given in table 1. The power spectrum, that characterizes the nature of the Gaussian random field, influences the number densities via the spectral parameter  $\gamma$  [also see 15],

$$\gamma = \frac{\sigma_1^2}{\sigma_2 \sigma_0} \quad (2.29)$$

critical point	index $i$	$J_1$	$J_2$
minima	0	$(0, \infty)$	$(0, J_1^2)$
saddle points	1	$(-\infty, \infty)$	$(J_1^2, \infty)$
maxima	2	$(-\infty, 0)$	$(0, J_1^2)$

**Table 1:** Field Hessian invariant variables  $J_1$  and  $J_2$ : parameter domains for the number density of minima, saddle points, and maxima  $\mathcal{N}_i$  in terms of the invariants  $J_1$  and  $J_2$  (for definitions see equations (2.25), (2.26) and (2.27)).

which is a combination of the spectral moments

$$\sigma_i^2 = \int_0^\infty P(k) k^{2i+1} \frac{dk}{2\pi}. \quad (2.30)$$

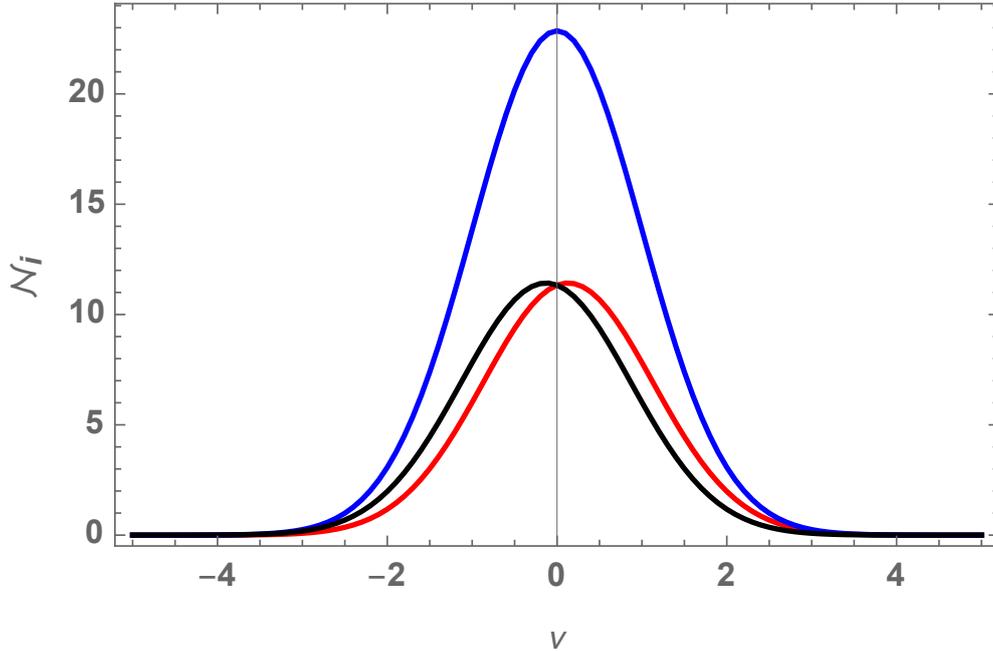
The critical point densities of Gaussian random fields, and the subsequent analysis, is fully determined by the spectral parameter  $\gamma$ .

In figure 1, we show the critical point densities  $\mathcal{N}_0(\nu)$ ,  $\mathcal{N}_1(\nu)$  and  $\mathcal{N}_2(\nu)$  for a two-dimensional Gaussian random field. Observe that for nearly the entire density range  $\nu$  the saddle points are approximately twice as abundant as the maxima and minima. The saddle point number density is symmetric and centers around the mean  $\nu = 0$ . The number density of maxima tends to be skewed towards positive field values, while the minima density is skewed towards negative field values. In Pranav et al. [22], we analyze the number density of critical points in three-dimensional Gaussian random fields. In that situation, there is hardly any overlap between the number density curves for minima and maxima. While in a two-dimensional random field both minima and maxima can be found for a large range of field values, in three-dimensional Gaussian fields they hardly co-exist. This is related to the separation of homologies as demonstrated in [22]. Also note from figure 1 that for Gaussian random fields, the number densities of maxima and minima are symmetric around  $\nu = 0$ , *i.e.*,  $\mathcal{N}_0(\nu) = \mathcal{N}_2(-\nu)$ . For non-Gaussian random fields, this symmetry is broken.

### 3 Homology and Persistence: Topological Structure of Random fields

The conventional analysis and characterization of random fields in terms of correlation functions inform us about their spatial coherence and clustering. In the case of Gaussian random fields we know that this is limited to the two-point correlation function, which fully encapsulates the statistical nature of the field.

However, when the primary interest is that of the features and patterns that define the spatial structure field, it is of greater benefit to study the topological characteristics of the



**Figure 1:** Critical point number densities  $\mathcal{N}_i$  as a function of the field value (obtained using equation (2.27)). The number density of maxima (red), saddle points (blue) and minima (black) in a Gaussian random field (with LCDM power spectrum), at function value  $\nu = f/\sigma$ .

field. Topology addresses the identity of topological features and their spatial connectivity. These include islands, tunnels and cavities or voids. Islands are the regions that have a field value in excess of a specific threshold, and one may study their connection, assess the number of cavities they encompass, how many tunnels percolate their interior, and a range of additional questions of interest.

This section presents the key concepts for a topological analysis and description of random fields that we concentrate on in the present study. These are homology, and specifically, the aspects of Betti numbers and persistence. For the inference of analytical expressions for homology measures, which is one of the key targets of the present study, we follow a path along a sequence of key topological concepts. Starting with the definition of a manifold, we are led to the concept of filtrations, in particular that of superlevel set filtrations. The central step in our analysis is the intimate relationship between the topology of a manifold and its singularity structure. The latter refers to the spatial distribution and connections between maxima, minima and saddle points. It is one of the central tenets of *Morse theory*, and leads us to the concept of Morse-Smale complex for codifying the connectivity between the manifold's singularities and its topological structure.

The final, decisive and unique, element of our formalism concerns the graphical translation of the Morse-Smale complex, producing a graphical representation of a random field realization that directly reflects its singularity and topological structure. Mathematically, it is known as a *graph* [88]. The graph representation facilitates our ability to infer integral relations for the Betti numbers, and even for the persistence diagram. While in principle we may follow this strategy in a three-dimensional context, in the present study we restrict ourselves to the situation of a two-dimensional random field and show that it yields an approximate

but accurate analytical expression for the two Betti numbers  $\beta_0$  and  $\beta_1$ .

### 3.1 Random Field Realizations and Manifolds

The ultimate objective of our study, the Gaussian temperature fluctuation field of the CMB, concerns a realization of a random field on a 2-sphere  $\mathbb{S}^2$ . However, because of the local nature of our analysis, we may restrict our analysis to random fields on the Euclidian plane  $\mathbb{R}^2$ . Hence, we consider a function  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  that is a realization of a two-dimensional real random field over  $\mathbb{R}^2$ . In principle, our analysis can be generalized to arbitrary higher dimensional random fields.

The realization  $f(\mathbf{r})$  of the random field may be considered in terms of a manifold  $M$ . Roughly speaking, a manifold is a subset of  $\mathbb{R}^N$  which, for some  $n \leq N$ , locally looks like  $\mathbb{R}^n$ . For the field realization  $f(\mathbf{r})$ , the corresponding manifold  $M$  is defined as

$$M = \{(\mathbf{r}, f(\mathbf{r})) | \mathbf{r} \in \mathbb{R}^2\} \subset \mathbb{R}^3. \quad (3.1)$$

### 3.2 Filtrations and Superlevel Sets

On the manifold  $M$  defined by the random field  $f(\mathbf{r})$ , we may construct a superlevel set. To this end, we apply a threshold  $\nu$  to the function values. The corresponding superlevel set filtration  $M(\nu)$  of the manifold  $M$  is

$$M(\nu) = \{(\mathbf{r}, f(\mathbf{r})) \in M | f(\mathbf{r})/\sigma \geq \nu\}. \quad (3.2)$$

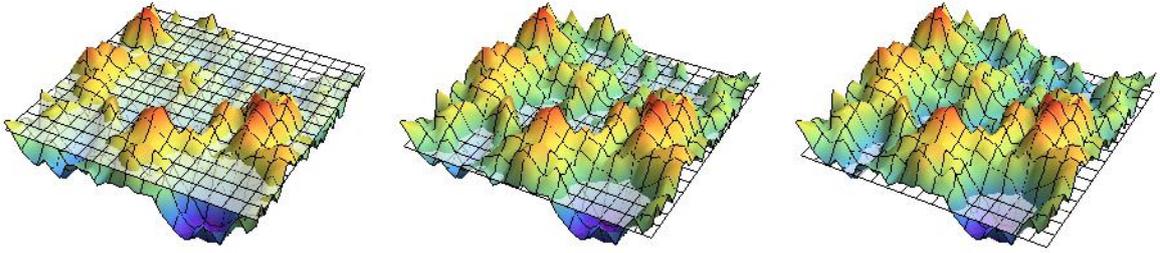
A key property of superlevel set filtrations is that

$$M(\nu_1) \subseteq M(\nu_2) \quad (3.3)$$

for all  $\nu_1, \nu_2$  with  $\nu_1 \geq \nu_2$ . This can also be visually appreciated from the example of superlevel filtrations of a Gaussian random field realization at a sequence of three different threshold levels  $\nu$  shown in figure 2. In our analysis,  $\nu$  always runs from  $\infty$  to  $-\infty$ , incorporating more and more points of  $M$  in  $M(\nu)$  as we descend along  $\nu$ . By definition therefore,  $M(\infty) = \emptyset$  and  $M(-\infty) = M$ .

### 3.3 Homology of Random Fields: Betti numbers

Homology is the topological formalism for specifying in a quantitative and unambiguous way the connectivity of manifolds through the connectivity of their boundaries [42, 44, 89]. A branch of algebraic topology, it uses the mathematical formalism of homology groups to represent holes in a space in terms of the connectivity with surrounding structures. We refer to appendix A for a formal treatment and presentation of the basic concepts of homology [also see 42, 90, 91, for complete mathematical treatments]. An extensive outline of homology, within the context of their application to the large-scale cosmological mass distribution, can also be found in [50] and [52]. Our recent numerical study of the Betti numbers, Euler characteristic and Minkowski functionals of three-dimensional Gaussian fields [22] can be seen as a companion paper to the present analytical and theoretical study.



**Figure 2:** Superlevel set filtrations. Superlevel set filtrations  $M(\nu)$  of a realization of a Gaussian random field for three different field thresholds  $\nu = f/\sigma$ : the regions with field value above the threshold  $f/\sigma > \nu$  belong to the superlevel set at  $\nu$  (see equation (3.2)).

Formally, Betti numbers are a concept from algebraic topology enabling a (partial) characterization of topological spaces. The Betti numbers characterize topology in terms of the homology of a manifold  $M$ , concentrating on the equivalence classes of homological chains and cycles on the manifold  $M$  (or its simplicial decomposition). This is intimately related to a study of the boundary  $\partial M$  of the manifold in the setting of superlevel set filtrations of functions on  $\mathbb{R}^D$  studied in this paper. The homology of a  $D$ -dimensional manifold is characterized by  $D + 1$  Betti numbers  $\beta_i, i = 0, \dots, D$ . In a sense, the  $i^{\text{th}}$  Betti number may be considered as the number of  $i$ -dimensional independent holes in a manifold  $M$ . For example, in three-dimensional space the zeroth Betti number  $\beta_0$  counts the number of separate components in  $M$ , while the first Betti number  $\beta_1$  counts the number of tunnels and the second Betti number  $\beta_2$  the number of cavities, in the context of superlevel sets on  $\mathbb{R}^3$ .

A formal and rigorous definition of the Betti numbers is given in appendix A. It follows simplicial homology theory, and concentrates on a characterization of a manifold in terms of the number of independent  $p$ -dimensional homological boundaries that it contains, each enclosing  $p$ -dimensional holes. Formally, within the context of algebraic topology, the boundary structure of the manifold is assessed in terms of  $p$ -cycles (see app. A). The collection of independent  $p$ -dimensional homological cycles, up to topological equivalence, is the  $p$ -th *homology group*  $H_p(M)$ <sup>5</sup>. The rank of the homology group  $H_p(M)$  is the number of all linearly independent topological cycles, and is denoted by the *Betti number*  $\beta_p$  [43]. A  $D$ -dimensional manifold  $M$  has one homology group  $H_p(M)$  for each of  $D + 1$  ranks  $0 \leq p \leq D$ , and the set of homology groups characterizes the boundary structure and homology of the superlevel set.

We may also observe that as the number of  $i$ -dimensional holes is an integer this implies Betti numbers  $\beta_i$  intrinsically to be an integer, *i.e.*,  $\beta_i : [-\infty, \infty] \rightarrow \mathbb{Z}$ . For random fields over  $\mathbb{R}^2$ , a realization can have infinitely many  $i$ -dimensional holes. It is more sensible therefore to define normalized Betti numbers  $\tilde{\beta}_i$ , the number of  $i$ -dimensional holes per unit volume/area. Hence,  $\tilde{\beta}_i$  is a real number, *i.e.*,  $\tilde{\beta}_i : [-\infty, \infty] \rightarrow \mathbb{R}$ . For practical applications, such as the numerical evaluation of Betti numbers in experimental or simulation datasets, de facto we nearly always deal with normalized Betti numbers. For reasons of convenience, we therefore always retain the notation  $\beta_i$ , even when it concerns a normalized Betti number  $\tilde{\beta}_i$ .

It is straightforward to extend the concept of the Betti numbers for a manifold to that of the superlevel set  $M(\nu)$ . The  $i^{\text{th}}$  Betti number  $\beta_i(\nu)$  of the superlevel set  $M(\nu)$  is the number of  $i$ -dimensional holes in  $M(\nu)$  as a function of filtration level  $\nu$ . Dependent on the nature of

<sup>5</sup>Correctly defined, the  $p$ -th homology group is the  $p$ -th cycle group modulo the  $p$ -th boundary group.

the manifold, we may also identify symmetries with respect to its topological structure. The statistical symmetry of Gaussian random fields between underdense and overdense regions implies an equal number of overdense regions above level  $\nu$  versus the number of underdense troughs below level  $-\nu$ . This means that in two-dimensional space the Betti numbers obey the statistical symmetry relation

$$\beta_0(\nu) = \beta_1(-\nu). \quad (3.4)$$

### 3.4 Betti, Euler, and Genus

The homology and connectivity of a manifold are globally quantified by the Betti numbers, which represent a (partial) characterization of topological spaces<sup>6</sup>. In this sense, they extend the principal topological characterization known in cosmology in terms of *genus* or *Euler characteristics*. Numerous cosmological studies have assessed and discussed the *genus* of the isodensity surfaces defined by the Megaparsec galaxy distribution [26, 92, 93].

The genus  $G$  specifies the number of handles defining an orientable surface, *i.e.*, the maximum number of incisions one can make without partitioning the manifold into separate components. The genus has a direct and simple relation to the Euler characteristic  $\chi$  of an isodensity surface.

Originally stemming from the description of polyhedral surfaces and three-dimensional simplicial complexes, the Euler characteristic is a topological invariant that plays a key role in homology. For a polyhedral surface with  $V$  vertices,  $E$  edges and  $F$  faces, the Euler characteristic is

$$\chi = V - E + F. \quad (3.5)$$

This definition of the Euler characteristic extends to manifold, as any manifold has a polyhedral decomposition. Note that this definition is well-defined, as the Euler characteristic does not depend on the chosen decomposition. While in principle one may define the Euler characteristic in any dimension  $D$ , its significance may be best appreciated in the context of surfaces in 3-dimensional space. To this end, we consider a 2-manifold in three-dimensional space, the surface  $\mathcal{S} = \partial M$  of a 3-manifold  $M$ . It follows from equation (3.5) that, when the surface  $\mathcal{S}$  consisting of  $c$  discrete components, the genus of the surface is given by

$$G = c - \frac{1}{2} \chi(\mathcal{S}), \quad (3.6)$$

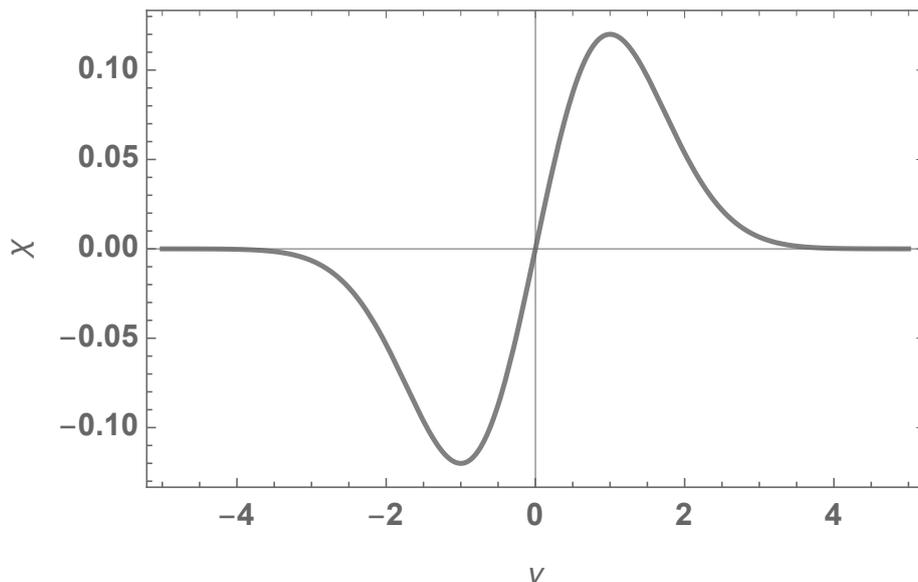
where the Euler characteristic  $\chi(\mathcal{S})$  is, by the Gauss-Bonnet theorem, expressed as the integrated intrinsic curvature of the surface

$$\chi(\mathcal{S}) = \frac{1}{2\pi} \oint \left( \frac{1}{R_1 R_2} \right) dS. \quad (3.7)$$

Here  $R_1$  and  $R_2$  are the principal radii of curvature at each point of the surface, and the integrand  $1/R_1 R_2$  is known as the Gaussian curvature. The integral of the curvature is invariant under continuous deformation of the surface  $\mathcal{S}$ . This is perhaps one of the most

---

<sup>6</sup>Strictly speaking this is only true when ignoring the torsion-free part of a homology group



**Figure 3:** The Euler characteristic  $\chi(\nu)$  for the superlevel set filtration of a Gaussian random field on a  $\mathbb{R}^2$  at field threshold  $\nu = f/\sigma$ . The well-known analytical expression for  $\xi(\nu)$  is given in equation (3.10).

surprising results in mathematics, establishing an intricate and profound relation between differential geometry and topology.

The profound central significance of the Euler characteristic is not only reflected in its intrinsic and deep relationship with the topology of simplicial complexes and the geometric properties of manifolds, and hence with simplicial topology and differential geometry. Similar intricate relations exist between the Euler characteristic and the homology of a manifold, and also with the singularity structure of manifolds. This is reflected in the relation between the Euler characteristic  $\chi(M)$  of a manifold and its Betti numbers  $\beta_i(M)$ .

The *Euler-Poincaré formula* [42] establishes that for any arbitrary  $D$ -dimensional manifold  $M$ , its Euler characteristic <sup>7</sup>  $\chi(M)$  is the alternating sum of the corresponding Betti numbers  $\beta_p(M)$  of the manifold,

$$\chi = \sum_{p=0}^D (-1)^p \beta_p. \quad (3.8)$$

In addition, according to Morse theory (see sect. 4) there is an intimate link between the topology of a field and its singularity structure. It expresses itself, amongst others, in the relation between the number density of critical points and the Euler characteristic  $\chi(\nu)$ . Akin to the relations above, the Euler characteristic  $\chi(\nu)$  of the manifold filtration at threshold level  $\nu$  appears to be equal to the alternating sum of all critical points with a field value in excess of  $\nu$ ,

$$\chi(\nu) = \int_{\nu}^{\infty} (\mathcal{N}_0(\nu') - \mathcal{N}_1(\nu') + \mathcal{N}_2(\nu')) d\nu'. \quad (3.9)$$

---

<sup>7</sup>For the  $D$ -dimensional definition of the Euler characteristic see [42].

As calculated by [15], for a Gaussian field this implies the following analytical expression for the Euler characteristic at level  $\nu$ ,

$$\chi(\nu) = \left( \frac{(2\pi)^{3/2}}{2} \frac{\sigma_1^2}{\sigma_0^2} \right) \nu e^{-\nu^2/2}. \quad (3.10)$$

For a visual impression, this expression for  $\chi(\nu)$  is shown in figure 3. With respect to eqn. 3.10, it is of importance to realize that the analytical expression implicitly assumes the manifold  $M$  to be a smooth, closed, manifold in 2-D Euclidian space  $\mathbb{R}^2$ . In nearly all practical – observational – circumstances this is not true. Any irregularities, convoluted boundaries, systematic biases, and incompleteness issues will lead to a more complex expression and a non-symmetric Euler characteristic. The correct expression in those circumstances may be inferred from the *Gaussian kinematic formula* or GKF [see 16, 22, 94]. The number density of minima, saddles, and maxima, and the Euler characteristic are also somewhat modified for a Gaussian random field on the two-sphere,  $\mathbb{S}^2$ , for which we refer to [95].

### 3.5 Persistence Diagrams

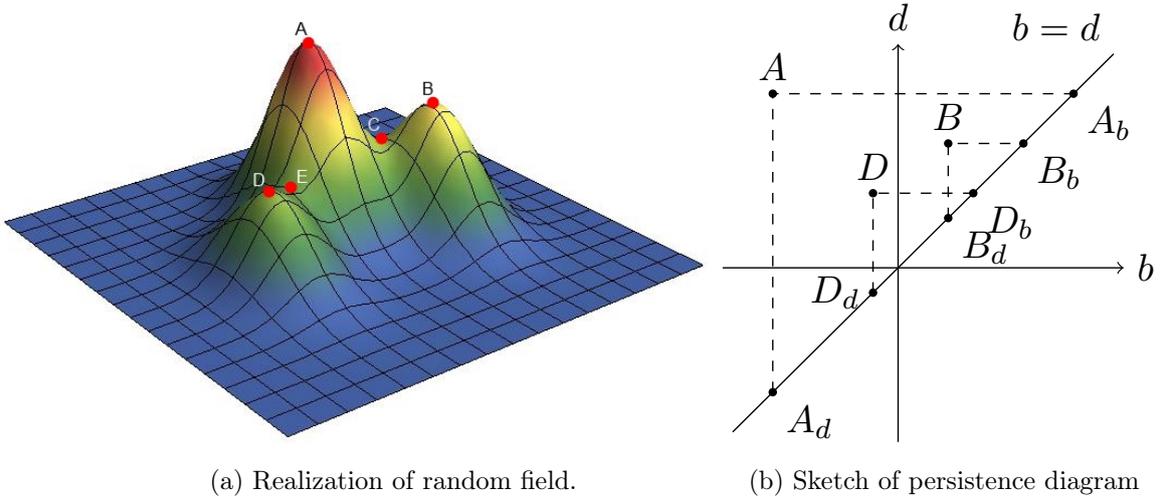
While Betti numbers provide a global characterization of the topology of a manifold, a richer description emerges when we assess the detailed changes in the topological structure of a manifold as the field threshold changes.

Complementary to the Betti numbers of a manifold, the structure of the features in the superlevel sets  $M(\nu)$  for different threshold values  $\nu$  forms a second signature of a random field. As mentioned above,  $M(\infty) = \emptyset$ . As  $\nu$  decreases, connected features enter the superlevel set filtration. If a connected component of  $M$  enters the superlevel set filtration at a threshold value  $\nu = a$ , we define the (0-dimensional) feature to be born at  $a$ . If the superlevel set  $M(\nu)$  consists of at least two connected (0-dimensional) components, lowering  $\nu$  can lead to the merger of two disconnected components. We define the youngest component to die at  $b$ , when two features merge at some threshold value  $\nu = b$ . For example, let two features be born at  $a_1 = 7$  and  $a_2 = 5$  respectively. If these features merge at  $b = 0$  the second component dies because it is 2 younger. Similar behaviour can be observed for 1-dimensional holes.

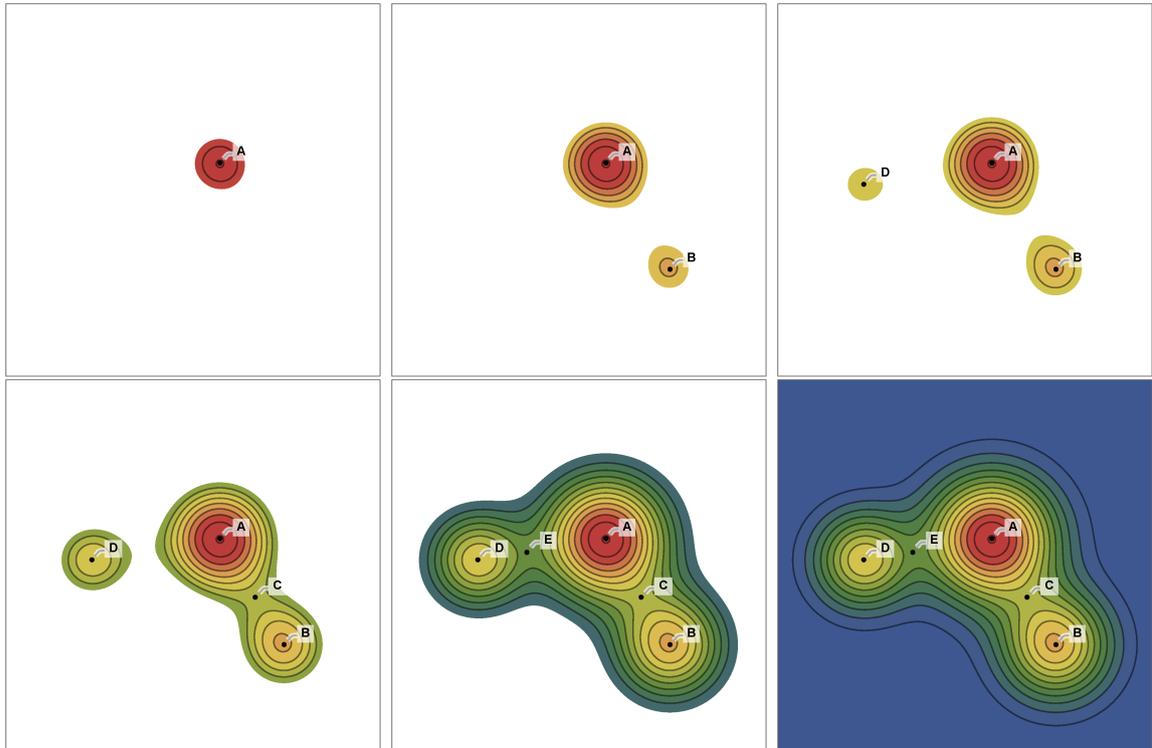
This process is illustrated in figure 4a. Initially, at threshold  $\nu = \infty$ , the superlevel set is the empty set. As the threshold decreases, the maximum  $A$  is the first topological feature to appear in the superlevel set. As the threshold  $\nu$  is lowered further, first  $B$  appears.  $B$  will then merge with  $A$  as the saddle point  $C$  appears. At this point, the topological feature  $B$  dies. We say that the topological feature  $A$  is more persistent than  $B$ . Analogously, as the threshold decreases further, the maximum  $D$  is born and subsequently killed by  $A$  as the saddle point  $E$  appears in the superlevel set filtration. The resulting persistence diagram sorts the topological features in the superlevel sets (or extrema) by their persistence.

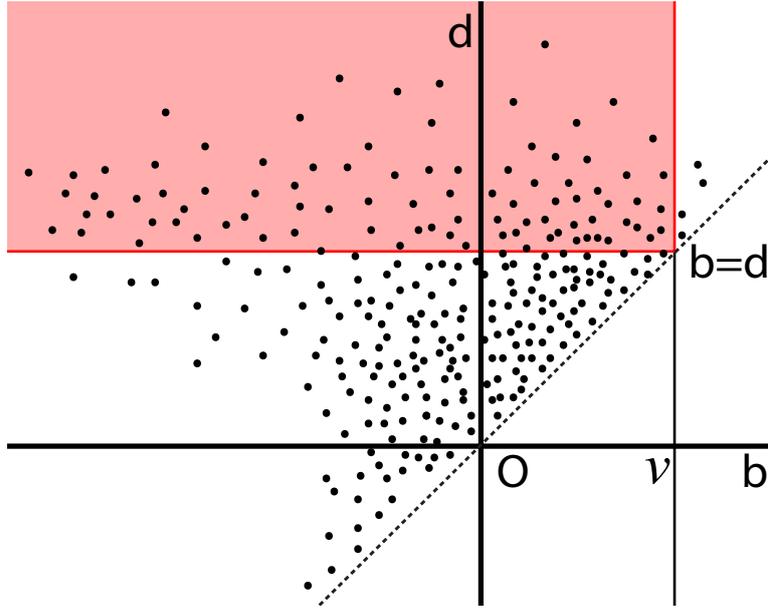
As  $\nu$  runs from  $\infty$  to  $-\infty$  all  $i$ -dimensional topological features<sup>8</sup> are born at some threshold  $a$  and die at some lower threshold value  $b$ . Hence each  $i$ -dimensional feature can be labeled by its birth and death  $(a, b)$ . This pair is called an  $i$ -dimensional event. The collection of all  $i$ -dimensional events of a realization of a random field form the  $i$ -dimensional persistence diagram [42]. The shape of the point cloud, of which the persistence diagram consists, forms a second signature of a random field (see figure 6), which has been the subject

<sup>8</sup>A feature is a specified geometric object, while a topological feature is the set of all features which are topologically identical.



**Figure 4:** Persistence and Field singularity structure: Panel (a) illustrates a realization of a random field. The critical points are denoted by  $A, B, C, D$  and  $E$ . The corresponding superlevel set filtrations are plotted in the six lower panels. Panel (b) illustrates the persistence of the superlevel set filtrations of the realization shown in the left figure. The persistence diagram consists of the points  $A, B$  and  $D$ , corresponding to the birth-death pair or the corresponding critical point in the realization.  $A_b, B_b$  and  $D_b$  are the times of birth corresponding to these critical points, while  $A_d, B_d$  and  $D_d$  are the corresponding times of death. Panels bottom: Persistence and Field singularity structure. A top view of the superlevel set filtration  $M(\nu)$  for decreasing threshold  $\nu$ .





**Figure 6:** Persistence diagram  $\Pi_i$ , schematic. The diagram plots the birth ( $b$ ) versus death ( $d$ ) of a topological feature. The points indicate the birth-death ( $b,d$ ) values for the topological features of a manifold  $M$ . For a specific superlevel set filtration,  $M(\nu)$ , the red region indicates the birth-death events that mark its topological structure. The number of events in this region yields the corresponding Betti number of  $M(\nu)$ .

of several studies including [96, 97]. A random field over  $\mathbb{R}^2$  generically consists of infinitely many connected components and holes. We will, therefore, normalize to the area.

The Betti numbers and persistence diagrams are intimately related. The Betti numbers can be computed from the persistence diagrams (see fig. 6), from which we may immediately infer that they represent a compression of topological information. The Betti numbers at threshold  $\nu$  are the number of (living) connected components or holes in the superlevel set  $M(\nu)$ . These components were born before or at  $\nu$  and die at a lower threshold. The Betti number  $\beta_i(\nu)$  can therefore be obtained by summing over all events  $(a,b)$  with  $a \geq \nu$  and  $b \leq \nu$ . In figure 6 we see the region over which one should sum to obtain the corresponding Betti number at a specific threshold  $\nu$ .

#### 4 The Graph Formalism & Betti Number Statistics

One of the central tenets of Morse theory is the fact that the full topological structure of a manifold  $M(\nu)$  is contained in the identity, spatial distribution, and connectivity of the critical points of the manifold. Because there are a finite number of critical points in a volume, the implication is that the full topological structure can be condensed into a discrete representation of the manifold that properly reflects the spatial singularity distribution. Accordingly, one of the principal concepts in Morse theory is the *Morse-Smale complex* [42, 53, 54]. It transcribes the topological structure of a manifold into a spatial tessellation of cells centering around

the minima and maxima, in which the edges that define the cell’s perimeter encapsulate the connectivity structure of the manifold.

The Morse-Smale complex allows the prospect of developing a formalism for inferring integral expressions for the expectation values of a manifold’s Betti numbers  $\beta_i$ , and for additional and more intricate homology measures such as the corresponding persistence diagrams. Here we introduce the *Graph Formalism* and describe the steps that lead us towards the formulation of accurate analytical expressions for Betti numbers. At the core of the formalism is the step in which the Morse-Smale complex is condensed into a corresponding simplicial complex, whose edges – 1-faces – represent the connections between the various field singularities.

Assessing the evolving network of connecting simplicial edges as we proceed through the manifold filtrations  $M(\nu)$  as a function of threshold  $\nu$ , we witness the unfolding of a network tree. In mathematics, these are known as graphs [88]. They establish a profound link between the manifold’s topology and graph theory. As the filtration systematically descends along the graph, and edge by edge is added to the simplicial complex, we witness the formation, absorption, and demise of simplicial elements in the complex. On the basis of this *incremental algorithm* [see 42, 73], we are enabled to follow the emergence and disappearance of the manifold’s topological features represented by the corresponding simplicial elements.

With the addition of the crucial aspect of the known probability of critical points – at density level  $\nu$  – in a (Gaussian) random field, we may subsequently arrive at a probabilistic evaluation of the formation and destruction of topological features and assess the expected number of topological features at a field level  $\nu$ . In the following subsections we describe how our *Graph Formalism* yields integral expressions for the Betti numbers  $\beta_0$  and  $\beta_1$  in a two-dimensional setting.

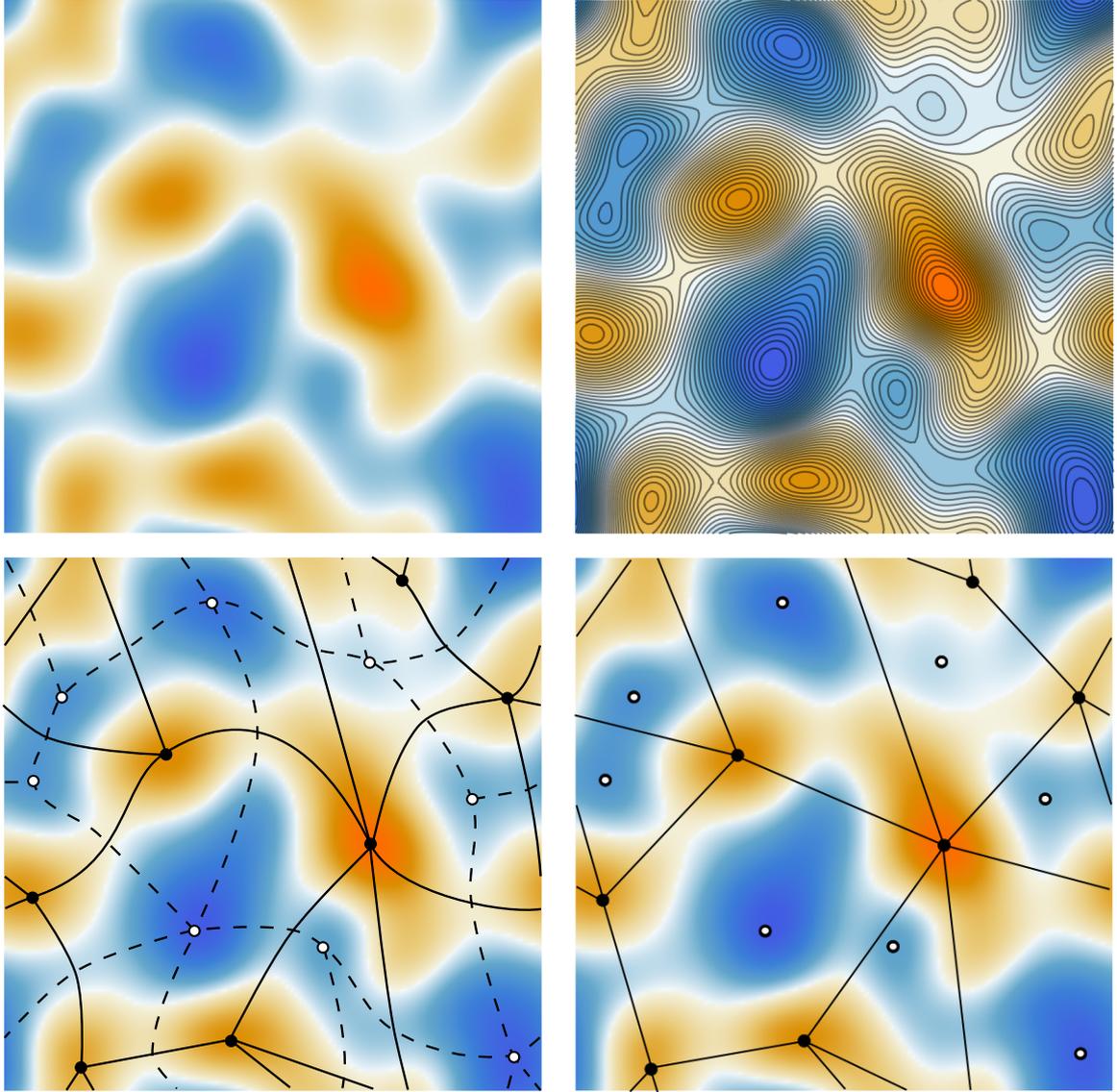
The integral expressions for the expectation value of the Betti numbers involve the (a priori unknown) probability function  $g(\nu)$  for a topological transition at the level  $\nu$ . For the Gaussian case, we infer a one-parameter fitting function for  $g(\nu)$ , leading us to a closed analytical expression for the expectation values of the Betti numbers.

#### 4.1 Morse-Smale Complex

Morse theory is the study of the topology of the level sets of a manifold via the critical points of a smooth scalar function that are defined on it (see sect. 2.3). A major result of Morse theory is the observation that the global shape and topology of a manifold is defined only by the character, spatial distribution and connectivity of its critical points [42, 54]. In a cosmological context, this has been invoked in several studies seeking to delineate the spatial patterns in the cosmic matter distribution [52, 98–100].

A key concept in Morse theory is the Morse-Smale complex [see *e.g.* 53, 72, 101]. The Morse-Smale complex is a partition of space defined by the critical points, and their connections via integral lines, that completely encapsulates the topology of (filtration) manifolds  $M(\nu)$ . It consists of a spatial tessellation of stable and unstable manifolds which forms a complete representation of the topological structure of the manifold.

The topological partition and connections between the critical points are established via the *integral lines*, which are the paths in the manifold  $M$  whose tangent vectors agree with the gradient of the function  $f$  at each point. In other words, the integral lines trace the gradient



**Figure 7:** Gaussian random field: Singularity structure, Morse-Smale complex and Singularity Graph. The upper left plot is a realization of a Gaussian random field on the unit torus. The upper left figure is an iso-contour plot of the realization in which we can distinguish the minima, saddle points and maxima. The lower left plot is the Morse-Smale complex of the random field. The lower right figure is an impression of the derived tessellation, the Morse-Smale graph, used in our analysis.

flow of the field  $f$ . Formally, the integral line  $\gamma_x : \mathbb{R} \rightarrow M$  of a smooth function  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ ,<sup>9</sup> with corresponding manifold  $M$ , is defined as the path satisfying

<sup>9</sup>Formally, such a function  $f$  should be a *Morse function*, *i.e.*, a function whose critical points are non-degenerate (*i.e.* its Hessian is non-singular for all critical points) and no two critical points have the same function value. Gaussian random fields always obey these conditions.

$$\frac{d}{ds}\gamma_x(s) = \nabla f(\gamma_x(s)), \quad \forall s \in \mathbb{R}. \quad (4.1)$$

with boundary condition  $\gamma_x(0) = x \in M$ . The *origin* and *destination* of an integral line  $\gamma$  are defined as

$$\text{org } \gamma_x = \lim_{s \rightarrow -\infty} \gamma_x(s), \quad (4.2)$$

$$\text{dest } \gamma_x = \lim_{s \rightarrow \infty} \gamma_x(s). \quad (4.3)$$

Since integral lines always follow the steepest slope of the function  $f$ , they generically originate and end up at critical points of  $f$ .

Assembling the integral lines with the critical point  $e$  as common origin demarcates the region of the manifold  $M$  that is known as *unstable* or *ascending* manifold. Likewise, one may identify the region that clusters all integral lines for which a critical point  $f$  is the common destination. This delineates the *stable* or *descending* manifold of the critical point  $f$ ,

$$\text{Stable manifold:} \quad S(e) = \{x \in M \mid \text{dest } \gamma_x = e\} \cup \{e\}, \quad (4.4)$$

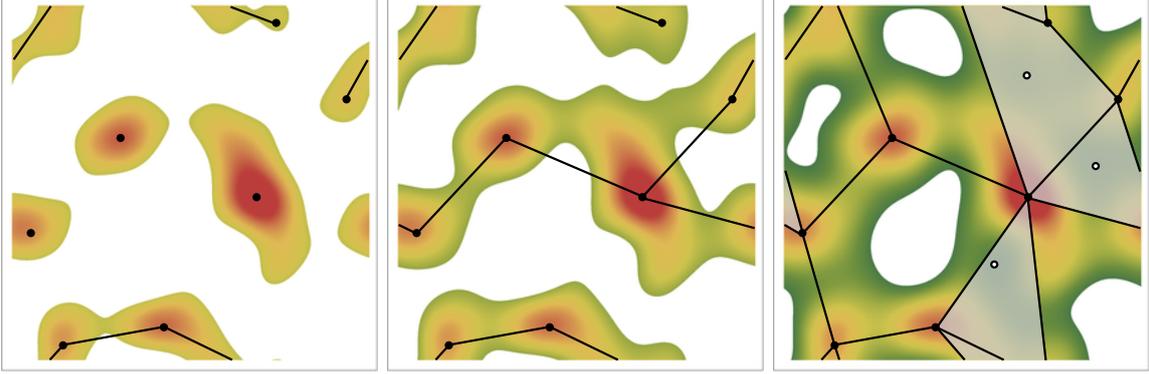
$$\text{Unstable manifold:} \quad U(e) = \{x \in M \mid \text{org } \gamma_x = e\} \cup \{e\}, \quad (4.5)$$

for all integral lines  $\gamma_x$ . The decomposition of the manifold  $M$  into its stable manifolds is called the Morse complex of the function  $f$ . It is a tessellation of  $M$  into Morse domains. Likewise, the decomposition of  $M$  into its unstable manifolds is the Morse function of  $-f$ .

One may obtain a visual impression of the process from figure 7. It is straightforward to identify the maxima (red peaks) and minima (blue troughs) in the 2-D Gaussian field realization shown in the top-left frame. In addition to the presence and location of the peaks and troughs, the contour map in the top righthand frame also reveals nicely the location of the saddle points. In the bottom left-hand frame of figure 7 we find the Morse complex of stable manifolds delineated by the dashed black lines originating at the minima in the field. It corresponds to a cellular partition of the manifold into stable Morse domains. The related Morse complex of unstable manifolds is indicated by the solid lines emanating from the maxima in the field. Furthermore, this defines a spatial tessellation, consisting of unstable Morse domains.

The intersections between the stable and unstable Morse domains yields a further decomposition of the manifold  $M$  into cells. These are the *Morse-Smale* cells. The Morse-Smale cells are the connected components of the set  $U(e_1) \cap S(e_2)$ , for all critical points  $e_1, e_2 \in M$ . Each of the Morse-Smale cells is the domain sharing a common origin ( $e_1$ ) and a common destination ( $e_2$ ). The *Morse-Smale complex* is the partition of the manifold  $M$  into the Morse-Smale cells<sup>10</sup>. It is straightforward to infer the identity of the Morse-Smale cells from the bottom lefthand panel in figure 7: the intersection between the stable and unstable domains results in Morse-Smale cells that have a maximum, minimum and two saddle points as a vertex. For a more elaborate analysis of Morse-Smale complexes see [102].

<sup>10</sup>Note that, in general, compactness of the manifold  $M$  is required for the origin and destination of the Morse-Smale complex to exist. Due to the large number of critical points in realizations of Gaussian random fields, the origin and destination generically exist for a Gaussian random field  $f$ .



**Figure 8:** Morse-Smale Graph. The 3 frames show the Morse-Smale graph of a manifold for three different filtrations, starting from high to low field value  $\nu$ . Left-hand panel: at high field values, the graph contains a few isolated vertices (marking the location of maxima), which in two cases are connected by edges (marking ridges via saddle points). Central panel: at a medium field value, we notice that all vertices have been connected via edges. This marks the formation of larger islands in the manifold, containing several peaks connected by ridges via saddle points. Righthand panel: several edges have connected into a loop/cycle, surrounding a cavity. Three loops even developed into a face, marking a trough with a density higher than the threshold.

## 4.2 Morse-Smale Graph

To condense the information contained in the Morse-Smale complex into one that is straightforward to process for a homology analysis, we translate the Morse-Smale complex of a manifold  $M$  into a simplicial complex. In a sense, it is a straightforward and natural step, following the circumstance that the Morse-Smale complex is a spatial tessellation of the manifold, which allows it to simplify into a graph representation that still retains the needed topological information. We call the resulting simplicial complex the *Morse-Smale Graph*. By construction, the resulting simplicial complex of the *Morse-Smale Graph* preserves the homological information of the superlevel set filtration.

The conversion involves the transformation of maxima into vertices, of saddle points into edges connecting the associated vertices and of the minima into faces. In the case the manifold for which we seek to construct the simplicial *Morse-Smale Graph* is the superlevel set filtration  $M(\nu)$ , the graph only includes the simplices – *i.e.* the vertices, edges and faces – for which the function  $f$  at the corresponding critical point  $\mathbf{x}_0$  assumes a value larger or equal than  $\nu$ , *i.e.*  $f(\mathbf{x}_0) \geq \nu$  (see figure 8). Dependent on the level, features around maxima (islands) can get connected. In this case, two vertices are connected by an edge. If the features are disconnected, they are reduced to vertices in the graph. In other words, two maxima get connected by an edge if and only if they are connected by a saddle point in the Morse-Smale complex. Likewise, troughs in the manifold  $M(\nu)$  are reduced to loops or cycles when the corresponding minimum is not yet included in the superlevel set (see appendix A).

It should be noted that not all Morse-Smale complexes can be reduced to a unique associated simplicial complex. There exist non-degenerate cases in which parts of the Morse-Smale complex cannot be converted according to the above rules. However, since these cases are non-generic for the random fields of interest and since we are analysing the statistical properties of homological features, we do not study these cases here.

### 4.3 The Incremental Algorithm: Graphic Extraction of Homological Information

In the following, we outline the procedure for the extraction of homological information, by means of an incremental filtration algorithm. The incremental algorithm is a procedure to compute the Betti numbers of a simplicial complex by tracking the evolution of the Betti numbers while incrementally introducing simplices [see 42, 73].

Given a realization of a random field  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ , we consider the change in homology of the superlevel set filtration  $M(\nu)$  as a function of  $\nu$ . Starting from infinite density  $\nu = \infty$ , we follow the evolving Morse-Smale graph as a function of descending density threshold  $\nu$ . Given that – at each step – the graph represents the full topology of the manifold filtration  $M(\nu)$ . The gradual unfolding of the Morse-Smale graphs enables us to follow the evolution of its topology in terms of the emergence, merging and annihilation of the corresponding topological features. In the two-dimensional situation, this concerns the appearance of an “island” around a maximum of the field, their merging into larger island features, and the appearance of holes as such islands connect into features surrounding an empty interior.

Morse theory teaches us that the topology can only change by inserting or removing a critical point. In other words, the persistent homology of the Morse-Smale complex changes if and only if a critical point is introduced to the filtration. The maximum introduces a new connected feature, while the minimum fills up a hole. In the two-dimensional case, the saddle point – depending on the position of the saddle point in the Morse-Smale complex – either connects two connected components or introduces a 1-dimensional hole to the structure. The Morse-Smale complex of a realization  $f$  thus completely determines the persistent homology.

At threshold  $\nu = \infty$  the superlevel set  $M(\infty)$ , and corresponding Morse-Smale complex, is empty. The corresponding Betti numbers  $\beta_0(\infty), \beta_1(\infty), \beta_2(\infty)$  vanish. As  $\nu$  decreases, the elements of the simplicial complex of  $M$  are incrementally added to the simplicial complex of  $M(\nu)$ . In this process, the zeroth Betti number  $\beta_0$  is increased by one for every vertex introduced to the simplicial complex. Similarly,  $\beta_1$  is decreased by one for every face introduced to the simplicial complex. The introduction of an edge – corresponding to a saddle point – either connects two previously disconnected parts, decreasing  $\beta_0$  by one or introduces a cycle to the simplicial complex, increasing  $\beta_1$  by one. The incremental algorithm is summarized in table 2. The second Betti number  $\beta_2$  is included, as the addition of the last face of the graph formally increases  $\beta_2$  from 0 to 1.

Based on the steps of the incremental procedure outlined in table 2, we may subsequently infer the following integral relations for  $\beta_0$ ,  $\beta_1$  and  $\beta_2$ :

$$\beta_0(\nu) = \int_{\nu}^{\infty} [\mathcal{N}_2(\mu) - (1 - g(\mu))\mathcal{N}_1(\mu)]d\mu, \quad (4.6)$$

$$\beta_1(\nu) = \int_{\nu}^{\infty} [g(\mu)\mathcal{N}_1(\mu) - \mathcal{N}_0(\mu)]d\mu, \quad (4.7)$$

$$\beta_2(\nu) = 0, \quad (4.8)$$

with  $\mathcal{N}_0, \mathcal{N}_1, \mathcal{N}_2 : \mathbb{R} \rightarrow \mathbb{R}$  the density of minima, saddle points and maxima at threshold  $\nu$ . In these integral expressions, the function  $g(\nu)$  is the probability that the introduction of a saddle point at threshold  $\nu$  creates a cycle increasing  $\beta_1$ ,

$$g(\nu) = P[\text{At } \nu, \beta_1 \mapsto \beta_1 + 1 \mid \text{At } \nu \text{ a saddle point is added to the simplicial complex}] , \quad (4.9)$$

with  $g : [-\infty, \infty] \rightarrow [0, 1]$ . Inversely, the function  $1 - g(\nu)$  denotes the probability that a saddle point connects two disconnected features and decreases  $\beta_0$  by one, *i.e.*,

$$1 - g(\nu) = P[\text{At } \nu, \beta_0 \mapsto \beta_0 - 1 \mid \text{At } \nu \text{ a saddle point is added to the simplicial complex}] . \quad (4.10)$$

Following the above, the problem of computing the Betti numbers  $\beta_0, \beta_1$  of a random field has been reduced to the determination of the probability  $g$ . Note that this formulation applies to arbitrary two-dimensional random fields, *i.e.*, it does not presume the Gaussianity condition.

#### 4.4 Approximating the Probability $g$

In order to deduce an accurate approximation for the probability function  $g(\nu)$ , we investigate its asymptotic behaviour in the limits  $\nu \rightarrow \pm\infty$ .

In the limit  $\nu = \infty$ , the superlevel set filtration  $M(\infty)$  is empty and the Betti numbers vanish. As the threshold  $\nu$  is decreased, we first see the addition of isolated islands. These are 0-dimensional vertices in the corresponding simplicial complex. These vertices dominate over the edges and faces, as at high function values, as maxima are more numerous than saddle points and minima (see figure 1). At high  $\nu$ , saddle points are therefore more likely to connect disjoint vertices than to form cycles. This translates into a simplicial complex that consists of many isolated vertices and relatively few clusters.

As we proceed to medium levels of  $\nu$ , we encounter the reverse situation (see figure 1). The introduction of a saddle point is likely to form a cycle since the simplicial complex consists of many clusters and relatively few connected components. As a consequence, as we decrease  $\nu$  starting from  $\nu = \infty$ , the Betti number  $\beta_0$  is first to increase while the Betti number  $\beta_1$  remains approximately constant. At low  $\nu$ , we reach a phase in which  $\beta_1$  decreases whereas  $\beta_0$  is approximately constant. At  $\nu = -\infty$ , the superlevel set filtration  $M(\nu) = M$  and the Betti numbers are determined by the homological structure of  $M$ . Differentiating equations (4.6) and (4.7) in these two regimes results in the asymptotic behaviour

$$g(\nu) \sim g_+(\nu) \equiv \frac{\mathcal{N}_0(\nu)}{\mathcal{N}_1(\nu)} \quad \text{as } \nu \rightarrow \infty , \quad (4.11)$$

$$g(\nu) \sim g_-(\nu) \equiv 1 - \frac{\mathcal{N}_2(\nu)}{\mathcal{N}_1(\nu)} \quad \text{as } \nu \rightarrow -\infty . \quad (4.12)$$

These asymptomatics apply to general two-dimensional random fields.

Subsequently restricting ourselves to a Gaussian random field, it is straightforward to infer that the derivatives of the Betti numbers at  $\nu = 0$  with respect to the threshold are related by

$$\left. \frac{d\beta_0}{d\nu} \right|_{\nu=0} = - \left. \frac{d\beta_1}{d\nu} \right|_{\nu=0} . \quad (4.13)$$

Critical point	index	$\beta_0$	$\beta_1$	$\beta_2$	addition of
Minimum	0		↓	↑	face
Saddle point	1	↓	↑		edge
Maximum	2	↑			vertex

**Table 2:** Betti number trend as a function of superlevel filtration threshold: decrease (↓)/increase (↑) with decreasing threshold in the incremental algorithm.

From the integral expressions (4.6) and (4.7), we then easily infer a relation between the number densities  $\mathcal{N}_0(0)$ ,  $\mathcal{N}_1(0)$ ,  $\mathcal{N}_2(0)$  of maxima, saddles and minima at  $\nu = 0$ ,

$$-\mathcal{N}_2(0) + [1 - g(0)]\mathcal{N}_1(0) = g(0)\mathcal{N}_1(0) - \mathcal{N}_0(0). \quad (4.14)$$

In addition, the statistical reflection invariance of Gaussian random fields,  $f \mapsto -f$ , implies a statistical symmetry between the corresponding Betti numbers,  $\beta_0(\nu) = \beta_1(-\nu)$ , and number densities,  $\mathcal{N}_0(0) = \mathcal{N}_2(0)$ . From this we obtain the additional constraint

$$g(0) = \frac{1}{2}. \quad (4.15)$$

Following the asymptotic relations (4.11), (4.12) and constraint (4.15), we are therefore led to propose the following fitting formula for  $g(\nu)$ ,

$$g(\nu) \approx \frac{g_+(\nu) e^{\alpha\nu} + g_-(\nu) e^{-\alpha\nu}}{e^{\alpha\nu} + e^{-\alpha\nu}}, \quad (4.16)$$

which includes the fitting parameter  $\alpha$ . Note that the proposed fitting function has the asymptotic limits,

$$\begin{aligned} \lim_{\nu \rightarrow -\infty} g(\nu) &= 1, \\ \lim_{\nu \rightarrow \infty} g(\nu) &= 0, \end{aligned} \quad (4.17)$$

as expected. For the fitting function  $g$  and the two asymptotics for a Gaussian random field near  $\nu = 0$  we refer to figure 9. The two asymptotics  $\mathcal{N}_0(\nu)/\mathcal{N}_1(\nu)$  and  $1 - \mathcal{N}_2(\nu)/\mathcal{N}_1(\nu)$  are very close to the constraint  $g(0) = 1/2$  at  $\nu = 0$ . The fitting function  $g$  is a small quantitative correction and quickly approaches the asymptotes for non-zero  $\nu$ . On the other hand, the correction is significant for the behaviour of the Betti numbers  $\beta_i(\nu)$  near  $\nu = 0$ . For a Gaussian random field, the parameter  $\alpha$  should be expressible in terms of the power spectrum  $P(k)$ . This relation will be the subject of an upcoming study.

Given the fitting function for  $g$ , equations (4.6) and (4.7) form a one-parameter fitting function for the Betti curves  $\beta_0(\nu)$  and  $\beta_1(\nu)$ . In figure 10, we compare the Betti numbers

measured for a realization of a Gaussian random field with the best fitting function  $g(\nu)$ . Equation (4.16) turns out to be a good approximation of the probability that a saddle point leads to a cycle. Also, equations (4.6) and (4.7) capture all the features of the Betti curves. We have attempted several alternative interpolation schemes for equation (4.16). However, the Betti numbers predicted by equations (4.6) and (4.7) turn out to be very sensitive to the behaviour of  $g$  near  $\nu$ . Many weightings can quickly be ruled out, while the exponential weighting turns out to lead to the most accurate results.

#### 4.5 Numerical Validation of $g$

In figure 10, numerically determined Betti numbers and the analytically predicted Betti numbers of a two-dimensional Gaussian random field are shown. The analytic approximation deviates at  $\nu = 0$  and at larger  $|\nu|$ . Nonetheless, the approximation appears to capture the qualitative nature of the Betti numbers. For this plot, we used a fitting parameter  $\alpha = 2.71$ . Empirically, the Betti numbers turn out to be very sensitive to the choice of the procedure for gluing the two branches of  $g(\nu)$  and to the value of the fitting parameter. In turn, this implies that equation 4.16 must be a good approximation of  $g$  and  $\alpha$  should be approximately constant.

### 5 Statistics of Persistence diagrams of Random Fields

In the previous section, we inferred an (approximate) analytical expression for the expected persistent Betti numbers of a Gaussian random field, in terms of the distribution of critical points. Here we forward integral expressions for the persistence diagrams of two-dimensional Gaussian random fields, following the same graph formalism.

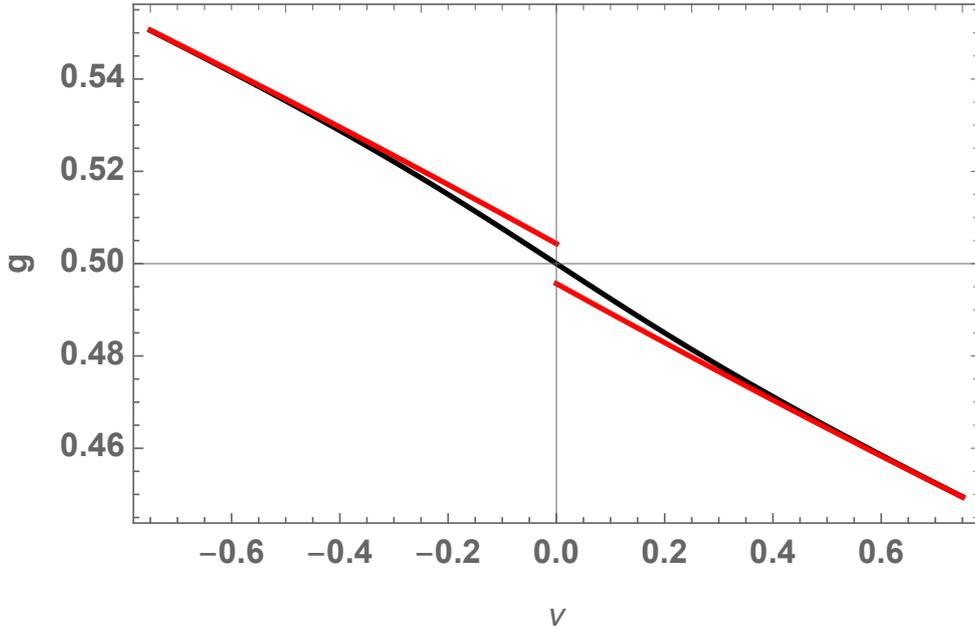
The  $i$ -dimensional persistence diagram  $\Pi_i$  of the superlevel set filtration  $M(\nu)$  of a function  $f$  consists of the set of events  $(b, d)$  corresponding to the birth at  $f = \sigma b$  and death at  $f = \sigma d$  of an  $i$ -dimensional topological feature. The persistence diagram of a random field is a continuous function  $\Pi_i : \mathbb{R}^2 \rightarrow \mathbb{R}$  defined as the expectation value

$$\Pi_i(b, d) = \text{E}[\text{number of events } (b, d) \text{ per unit area/volume in a realization}]. \quad (5.1)$$

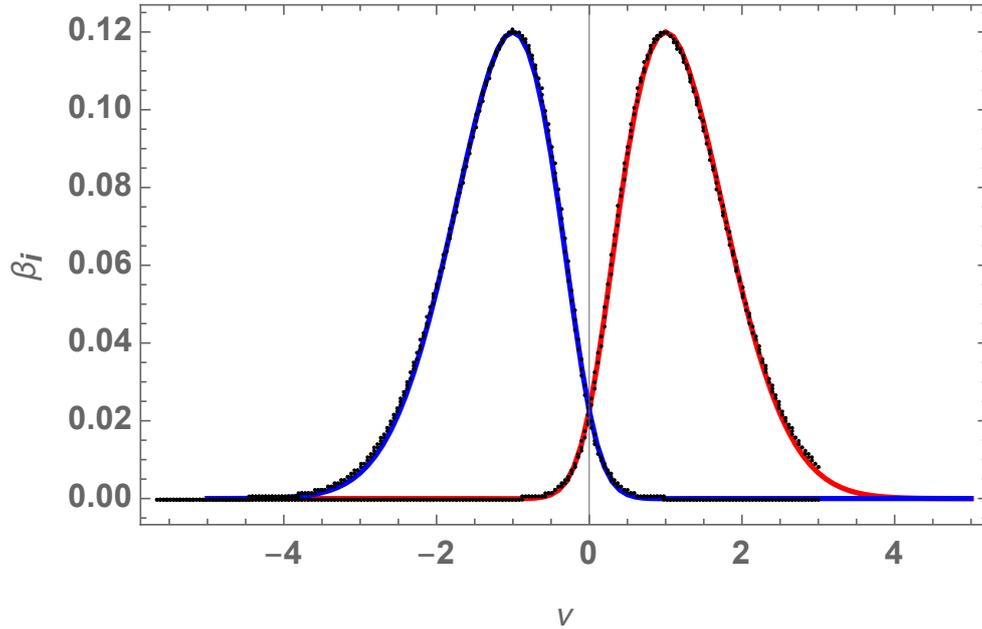
Using the graph formalism presented in section 4.2, the persistence diagram  $\Pi_i$  is directly related to the distribution of and connections between critical points. Following the growth of the Morse graph as we proceed through the field filtration, singularities get absorbed into the graph and establish new graphic connections. As they do so, topological features are born or vanish. The levels at which a feature emerges and at which it disappears defines a (birth,death) point in the persistence diagram.

The zero-dimensional persistence diagram  $\Pi_0$  consists of the (birth, death) pairs that represent the birth of an island and the subsequent merging of this feature with a neighbouring island. The birth of an island corresponds to the appearance of a local maximum at function value  $f = \sigma b$  in the Morse graph. The merger with a neighbouring island occurs upon the introduction of a saddle point at function value  $f = \sigma d$ . The saddle point joins two disconnected components. Evidently, it is the birth *c.q.* death of an island feature that contributes to the evolution of the zeroth persistent Betti number  $\beta_0(\nu)$ .

The (birth, death) pairs in the one-dimensional persistence diagram  $\Pi_1$  represents the formation of a loop and the subsequent filling up of the loop. The creation of the loop corresponds to the introduction of a saddle point at function value  $f = \sigma b$ , while the loop fills up



**Figure 9:** Probability distribution  $g$ : comparison between function  $g(\nu)$  (black) and the 2 fitting relations (red) for  $g_-(\nu)$  for  $\nu < 0$  and  $g_+(\nu)$  for  $\nu > 0$ . For their expressions see eqns. (4.11), (4.12) and (4.16).



**Figure 10:** Betti numbers, fitting formula *v.s.* numerical Betti curve. Lines: analytical relation for the Betti curves  $\beta_0(\nu)$  (red) and  $\beta_1(\nu)$  (blue) for a two-dimensional Gaussian random field (with a LCDM power spectrum), as specified by expressions (4.6) and (4.7). Dots: the corresponding numerically computed Betti numbers for a realization of the same LCDM Gaussian random field.

and disappears due to the introduction of a minimum at function value  $f = \sigma d$ . In the case of  $\Pi_1$ , a saddle point connects two features which were already connected. It is self-evident that this phenomenon relates to the evolution of the first persistence Betti number  $\beta_1(\nu)$ .

Following the observations above, we may formulate the corresponding expressions for the expected persistence diagrams  $\Pi_i$  of a two-dimensional random field in terms of the statistics of the critical points. As a function of the birth and death field levels  $(b, d)$ , these are the integral expressions

$$\Pi_0(b, d) = \int_0^\infty G_0(b, d, r) p_0(b, d, r) dr, \quad (5.2)$$

$$\Pi_1(b, d) = \int_0^\infty G_1(b, d, r) p_1(b, d, r) dr, \quad (5.3)$$

in which  $p_0(a, b, r)$  is the number density of (maximum, saddle point)-pairs whose mutual distance is  $r = \|\mathbf{x}_m - \mathbf{x}_s\|$ , where the maximum is located at  $\mathbf{r}_m$ , at function value  $f(\mathbf{r}_m) = \sigma b$ , and the saddle point at position  $\mathbf{r}_s$ , at function value  $f(\mathbf{r}_s) = \sigma d$ . Likewise,  $p_1(b, d, r)$  is the number density of (saddle point, minimum)-pairs with  $r$  the distance between the saddle point, at function value  $b$ , and the minimum at function value  $d$ .

The number densities  $p_1$  and  $p_0$  are natural extensions of the number density of the field singularities,  $\mathcal{N}_0$ ,  $\mathcal{N}_1$  and  $\mathcal{N}_2$ , whereby they incorporate the spatial correlations between the critical points. Using Rice's formula [13–16, 87] we can express  $p_0$  and  $p_1$  as the four-dimensional integral

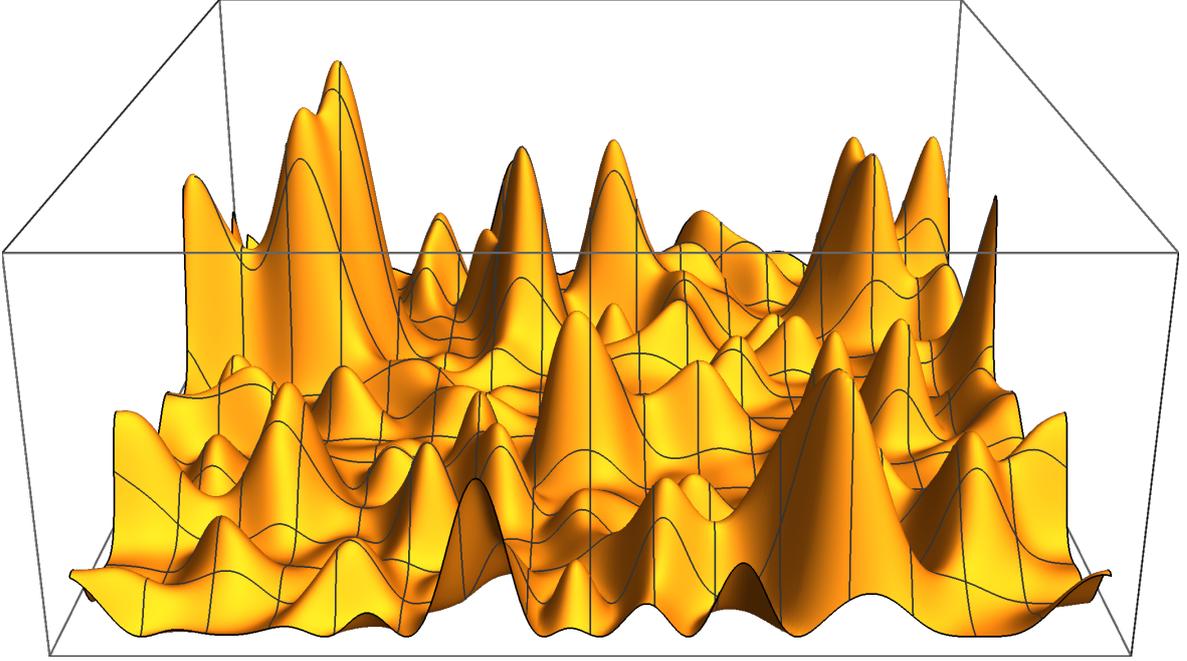
$$p_i(b, d, r) = \iiint\!\!\!\int P(b, f_k(0) = 0, J_k(0), d, f_k(r) = 0, J_k(r)) \times (|J_1(0)^2 - J_2(0)| + |J_1(r)^2 - J_2(r)|) dJ_1(0) dJ_2(0) dJ_1(r) dJ_2(r), \quad (5.4)$$

where the integral for  $i = 0$  and  $i = 1$  ranges over  $(J_1(0), J_2(0), J_1(r), J_2(r)) \in (-\infty, 0) \times (0, J_1(0)^2) \times (-\infty, \infty) \times (J_1(r)^2, \infty)$  and  $(J_1(0), J_2(0), J_1(r), J_2(r)) \in (-\infty, \infty) \times (J_1(0)^2, \infty) \times (0, \infty) \times (0, J_1(r)^2)$ .

In the intergral expression for persistence  $\Pi_0(b, d)$ , the function  $G_0(b, d, r)$  is the probability that the introduction of the saddle point with function value  $f = \sigma d$  at location  $\mathbf{r}_s$  induces the death of the island corresponding to the maximum with function value  $f = \sigma b$  at location  $\mathbf{r}_m$ . Likewise, the function  $G_1$  is the analogous probability that the saddle point in the (saddle point, minimum)-pair induces the introduction of a loop at function value  $f = \sigma b$  which is killed by the minimum at function value  $f = \sigma d$ . The two functions  $G_0$  and  $G_1$  are natural generalizations of the probability function  $g$  (equation (4.9)). The relation can be made explicit by integrating over the birth  $b$  and death  $d$  variables

$$g(\nu) = \int_0^\infty \int_{-\infty}^\nu \int_\nu^\infty G_0(b, d, r) db d d dr. \quad (5.5)$$

The integral over the separation  $r$  traces over the different configurations which lead to the same birth-death event  $(b, d)$  in the persistence diagram.



**Figure 11:** A highly local non-Gaussian random field of the local template with non-Gaussianity parameter  $f_{NL} = 1$ . From this field one can clearly observe an asymmetry in the statistics of maxima and minima.

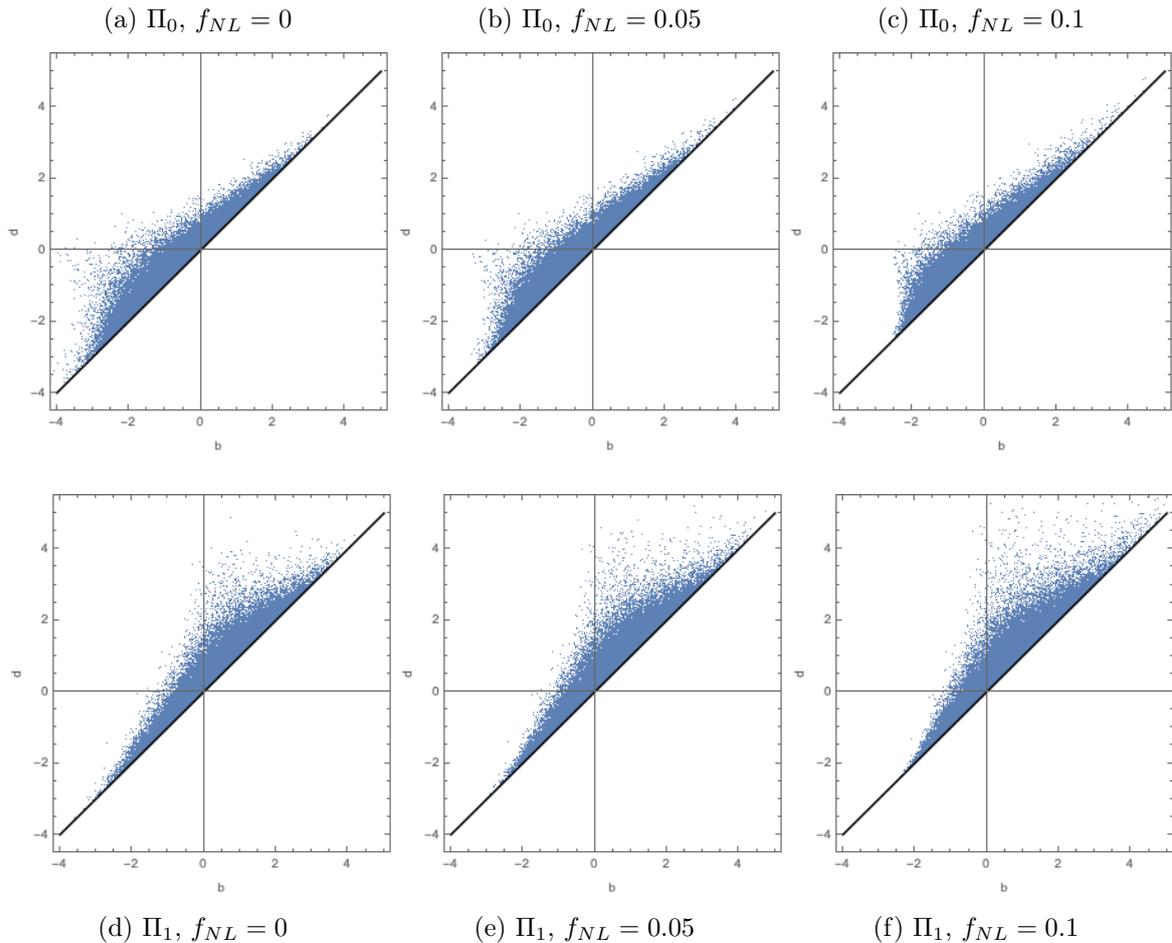
The statistical symmetry of Gaussian random fields with respect to reflection through the mean  $\mu = 0$  ensures that the persistence diagrams  $\Pi_0, \Pi_1$  are statistically symmetric under reflections through the line  $b = -d$ , *i.e.*,  $\Pi_0(b, d) = \Pi_1(d, b)$ . In addition, from the Morse-Smale complex we deduce that the probability  $G_0$  is related to the probability that in a realization of a random field  $f$ , the points  $A = (0, 0)$  and  $B = (r, 0)$ , with  $A$  a maximum and  $B$  a minimum of  $f$  are connected by an integral line  $\gamma$  between  $A$  and  $B$  defined by

$$\gamma(t) \in M \text{ and } \frac{d\gamma(t)}{dt} = \nabla f(\gamma(t)) \quad \forall t \in \mathbb{R}. \quad (5.6)$$

For Gaussian random fields, we expect this probability to be expressible in terms of a Gaussian path integral [103]. Such an expression would, besides predicting the persistence diagrams  $\Pi_i$ , yield an expression for the fitting parameter  $\alpha$  in terms of the power spectrum. Obtaining an approximation for  $Q_0$  and  $g$  using path integral methods is the subject of an upcoming paper [104].

## 6 Non-Gaussian Random Fields: Persistence & Betti numbers

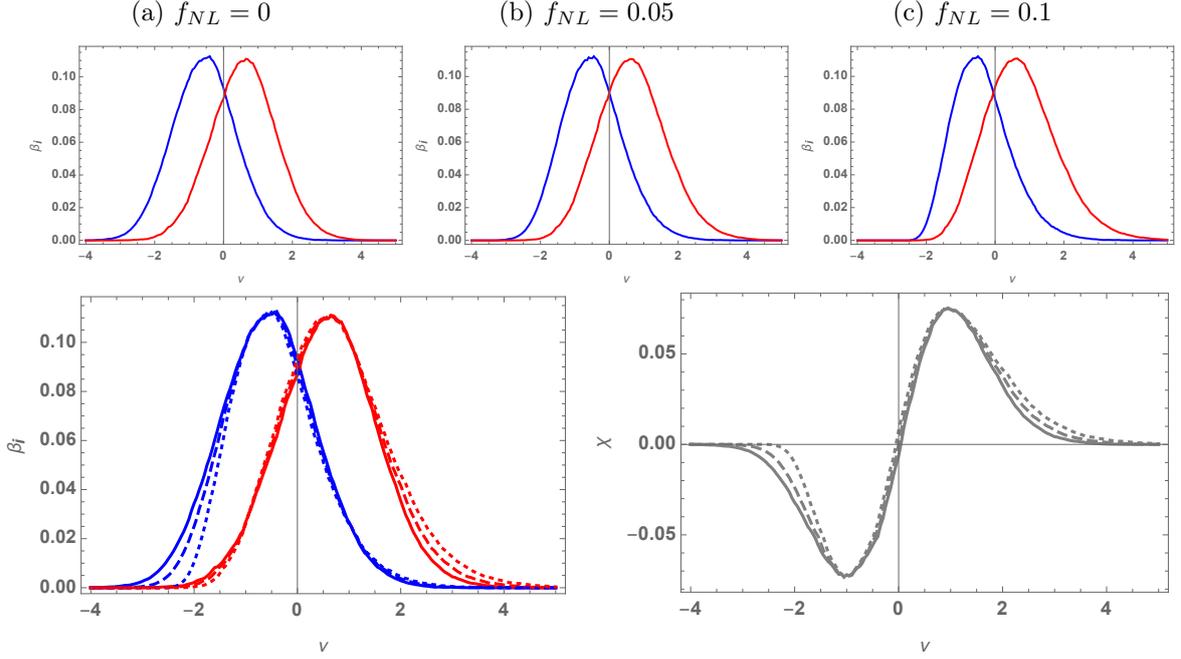
In the previous section, we analysed the persistent Betti numbers of the superlevel set filtration  $M(\nu)$  of a random field  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ , and derived the corresponding fitting formulas for the Gaussian case. The analysis can be extended naturally to random fields on more general two-dimensional manifolds. When applied to random fields on the two-dimensional sphere  $\mathbb{S}^2$ , the Gaussian persistent Betti numbers can be compared to the persistent homology of the CMB temperature perturbations  $\delta T$  to look for deviations from Gaussianity [12, 76, 105]. In this section, we numerically investigate the sensitivity of the persistent Betti numbers



**Figure 12:** Persistence Diagrams non-Gaussian random fields. The figure shows the 0- and 1-dimensional persistence diagram of a Gaussian and two non-Gaussian random fields. The non-Gaussian fields are of a local nature, characterized by parameter  $f_{NL}$ . Left: a Gaussian random field, with LCDM power spectrum (see text). By definition, it has non-Gaussianity parameter  $f_{NL} = 0$ . Centre: a random field with local non-Gaussianity parameter  $f_{NL} = 0.05$ . Right: a random field with local non-Gaussianity parameter  $f_{NL} = 0.1$ . The persistence diagrams clearly reveal a significant change in shape as a function of  $f_{NL}$ .

and persistence diagrams on non-Gaussianities in random fields on  $\mathbb{R}^2$ . An investigation of the behaviour of persistent homology of random fields on the sphere  $\mathbb{S}^2$  is the subject of an upcoming paper [106].

Within the context of primordial non-Gaussianities it is common to distinguish four different templates: local, enfolded, equilateral, and orthogonal. These cover a large range of physical models of the early Universe. For an elaborate review of these templates see [78, 79, 84, 85, 107]. In this study, we restrict our numerical study to the local template. A realization of a random field with local non-Gaussianities  $f_{NG} : \mathbb{R}^2 \rightarrow \mathbb{R}$  can be generated from a realization of a Gaussian random field  $f_G : \mathbb{R}^2 \rightarrow \mathbb{R}$  with zero mean  $\mu = 0$  and unit variance  $\sigma^2 = 1$ ,



**Figure 13:** Betti Numbers & Euler characteristic of non-Gaussian fields. For the same set of Gaussian and non-Gaussian random field realizations as in fig. 12 above, the panels show the Betti number curves  $\beta_0(\nu)$  (red) and  $\beta_1(\nu)$  (blue) and the corresponding Euler characteristic  $\chi(\nu)$  (black), as a function of field threshold level  $\nu = f/\sigma$ . The non-Gaussian fields are of a local nature, characterized by the parameter  $f_{NL}$ . Top row: the Betti numbers  $\beta_0$  (red) and  $\beta_1$  (blue) for the individual random field realizations. Top left: Gaussian random field, with LCDM power spectrum (see text), and  $f_{NL} = 0$  (by definition). Top centre: a random field with local non-Gaussianity parameter  $f_{NL} = 0.05$ . Top right: a random field with local non-Gaussianity parameter  $f_{NL} = 0.1$ . Note the characteristics shift of the Betti curves, manifesting itself as a systematic increase of skewness and (reduced) kurtosis (see table 3). Bottom row left: comparison of Betti curves  $\beta_0$  (red) and  $\beta_1$  (blue) of the 3 Gaussian and non-Gaussian random field realizations. Bottom row right: comparison between their Euler characteristic curves  $\chi(\nu)$ . Solid: Gaussian random field (LCDM power spectrum); dashed: local non-Gaussian field with  $f_{NL} = 0.05$ ; dotted: local non-Gaussian field with  $f_{NL} = 0.1$ .

using the non-linear transformation

$$f_{NG} = f_G + f_{NL}f_G^2, \quad (6.1)$$

with the non-Gaussianity parameter  $f_{NL} \in \mathbb{R}$  [see 10]. In order to compare the Betti numbers and persistence diagrams for different non-Gaussianity parameters  $f_{NL}$ , we normalize  $f_{NG}$  as follows,

$$f_{NG} \mapsto \frac{f_{NG} - \langle f_{NG} \rangle}{\langle f_{NG}^2 \rangle^{1/2}}. \quad (6.2)$$

This ensures that  $f_{NG}$  is centered around zero, *i.e.*,  $\langle f_{NG} \rangle = 0$ , and has unit variance, *i.e.*,  $\langle f_{NG}^2 \rangle = 1$ . For an illustration of a representative realization of a non-Gaussian field with local

$f_{NL}$	$\mu$ mean		$\sigma$ st. dev.		$\gamma_1$ skewness		$\gamma'_2$ kurtosis		$\gamma_2 = \gamma'_2 - 3$ exces kurtosis	
	$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$	$\beta_0$	$\beta_1$
0	0.64	-0.55	0.95	0.92	0.041	0.085	3.03	3.01	0.027	0.015
0.05	0.67	-0.48	0.99	0.88	0.25	0.32	3.12	3.18	0.12	0.18
0.1	0.76	-0.38	1.05	0.85	0.46	0.64	3.25	3.62	0.25	0.62

**Table 3:** Betti curve moments: the mean, standard deviation, skewness and kurtosis of the curves  $\beta_0(\nu)$  (the first entry) and  $\beta_1(\nu)$  (the second entry) of a (LCDM) Gaussian random field, and two corresponding (local) non-Gaussian fields. The level of non-Gaussianity is specified by the parameter  $f_{NL}$  (see expression (6.1)).

non-Gaussianities with non-Gaussian parameter  $f_{NL} = 1$  see figure 11. Note the asymmetry between the peaks and dips in the field.

In figure 12 we plot the persistence diagrams  $\Pi_0$  and  $\Pi_1$  as a function of parameter  $f_{NL}$ . The original random field  $f_G$  was constructed with the cold dark matter power spectrum described by [15]. In the Gaussian case, *i.e.*,  $f_{NL} = 0$ , the zero- and one-dimensional persistent diagrams are statistically symmetric, *i.e.*,  $\Pi_0(a, b) = \Pi_1(b, a)$ . As  $f_{NL}$  is increased, we see that this symmetry is broken. The zero-dimensional persistent diagram  $\Pi_0$  consists of many points in the upper right quadrant which are absent in the lower left quadrant of one-dimensional persistent diagram  $\Pi_1$ . The presence of a small local non-Gaussian component leads to an observable effect in the persistence diagram of the field. However, note that the scatter plots in particular emphasize the change in the outliers.

A more quantitative effect can be witnessed in the persistent Betti numbers (see figure 13) corresponding to the persistent diagrams presented in figure 12. For a Gaussian random field, *i.e.*,  $f_{NL} = 0$ , the Betti curves  $\beta_0(\nu)$  and  $\beta_1(\nu)$  are bell-shaped curves. Also, the frames show that statistically the two Betti numbers are reflection symmetric, *i.e.*,  $\beta_0(\nu) = \beta_1(-\nu)$ . In the non-Gaussian case, both symmetries are broken. The individual Betti curves  $\beta_0(\nu)$ ,  $\beta_1(\nu)$  get skewed and the symmetry between the two curves is broken. Quantitatively, this effect can be seen in the behaviour of the mean, standard deviation, skewness and kurtosis,

$$\begin{aligned} \mu &= E[\nu]; & \sigma^2 &= E[(\nu - \mu)^2]; \\ \gamma_1 &= E\left[\left(\frac{\nu - \mu}{\sigma}\right)^3\right]; & \gamma'_2 &= E\left[\left(\frac{\nu - \mu}{\sigma}\right)^4\right]; & \gamma_2 &= \gamma'_2 - 3. \end{aligned}$$

The inferred values of these statistical moments are presented in table 1. We observe that the mean  $\mu$  of both  $\beta_0$  and  $\beta_1$  increase for larger  $f_{NL}$ . Interestingly, we find a decrease of the standard deviation  $\sigma$  of the Betti curve  $\beta_0(\nu)$  for increasing  $f_{NL}$ , while the standard deviation for the Betti curve  $\beta_1(\nu)$  decreases. The skewness  $\gamma_1$  and excess kurtosis  $\gamma_2$  also

evolve non-trivially as a function of the non-Gaussianity parameter  $f_{NL}$ . A similar effect is present in the Euler characteristic  $\chi(\nu)$  (see figure 13). Note that the Betti curves  $\beta_i(\nu)$  completely determine the Euler characteristic  $\chi(\nu)$ , *i.e.*,

$$\chi(\nu) = \beta_0(\nu) - \beta_1(\nu) + \beta_2(\nu). \quad (6.3)$$

Note that the second Betti number,  $\beta_2$ , vanishes. The Euler characteristic, and more generally the Minkowski functionals, of the CMB has been extensively studied in [17, 24, 34–40]. We believe that an analysis in terms of the two Betti curves and the persistence diagrams will have more statistical power to detect non-Gaussian deviations. For an initial investigation see [76].

In summary, we see that the persistent homology of a random field is highly sensitive to the presence of non-Gaussianities. By studying the Betti curves of the cosmic microwave background radiation field, we can attempt to measure primordial non-Gaussianities. For a more detailed study of the persistent homology of non-Gaussianities on the two-dimensional sphere  $S^2$ , we refer to the upcoming paper [106].

## 7 Conclusion and Discussion

The topology of superlevel set filtrations of Gaussian random fields has over the last decades received a lot of attention in terms of the study of the Euler characteristic and the Minkowski functionals [17, 24, 34–40]. A major virtue of these functionals is that in the case of Gaussian random fields there are analytical expressions for their expectation value [see e.g. 22]. Deviations from the Gaussian nature of the primordial temperature perturbations in the Cosmic Microwave Background are then expected to reveal themselves as differences with the Gaussian expectation values.

Potential deviations in the cosmic microwave background radiation field, resulting from primordial non-Gaussianities, can provide us with valuable clues about the physics of the early Universe [10, 18–21, 24]. However, the Euler characteristic and Minkowski functionals are summarizing characteristics of the topology. Considerably more topological information is contained in the homology of manifolds, in particular in their persistent homology. Homology groups for superlevel set filtrations contain the complete information of the creation, merging and disappearance of topological features.

The present study focusses on the theoretical and analytical aspects of the homology of Gaussian random fields. The insights, understanding and analytical expressions derived have the objective of providing a theoretical foundation for the numerical and data analysis studies within the context of our project for developing and applying advanced concepts of topological data analysis [see e.g. 42, 48] to the description and quantification of the topology and of the large scale Universe. These range from the topology of the primordial mass distribution up to the connectivity aspects of the complex spatial pattern of the cosmic web on Megaparsec scales [22, 29, 50, 52, 76, 108].

In this paper, we demonstrate that Betti numbers and persistence diagrams are a powerful tool for the more complete characterisation of the topology of the primordial mass distribution [22, 50, 70]. The study involves an analytical investigation of the topology of (Gaussian) random fields based on Morse theory. Our argument is based on Morse theory’s principal topological thesis, the fact that the topology of a manifold is fully specified by the spatial distribution and connectivity of the discrete set of its singularities (maxima, minima and saddle points). We formalize the connectivity structure of the field in terms of the Morse

graph, which condenses the connectivity structure of the Morse-Smale complex into a graph network. In other words, it introduces the concept of a graph to represent the connectivity of a manifold as specified by the Morse-Smale complex, the segmentation of space defined by the spatial distribution of the singularities and which consists of regions connecting minima and maxima via the field’s integral lines. Having compressed the topological information into the Morse graph, we proceed by a filtration process in terms of the *incremental algorithm* for adding or removing topological features as new singularities in the Morse graph get included. In combination with the analytically known field distribution functions for the maxima, minima, and saddles in a Gaussian random field, we are led to the integral expressions for the Betti numbers and persistence diagrams of Gaussian random fields. By investigating the asymptotic behaviour of the Betti numbers we are able to infer an accurate fitting function. Note that these analytical expressions have been derived on the basis of a probabilistic calculation, and do not involve and resort to realizations of Gaussian random fields.

Deviations from the obtained fitting formulas most likely indicate the presence of non-Gaussianities. We numerically demonstrate that both the Betti numbers and the persistence diagrams are sensitive probes for the search for primordial non-Gaussianities in the CMB [also see 70, 74–76]. We evaluate both the Betti numbers and the persistent diagrams for a range of non-Gaussian fields of the local template and demonstrate the existence of a systematic deformation of the distribution. We expect different non-Gaussianity templates to affect the distribution of the Betti numbers and persistence diagrams in distinct ways. For a systematic analysis of the persistent homology of non-Gaussian fields, we refer to the upcoming paper [106].

## Acknowledgment

We are particularly indebted to Robert Adler for his encouragement and illuminating discussions. We wish to thank Leon Doddema, whose help was essential for solving key computer and data issues. We are also grateful to Bernard Jones, Elisabetta Pallante, and Mathijs Wintraecken for useful and insightful discussions, in particular at the early stages of this project. Job Feldbrugge thanks Neil Turok for enabling the continuation of research on the presented study. Research at Perimeter Institute is supported by the Government of Canada through Industry Canada and by the Province of Ontario through the Ministry of Economic Development, Job Creation and Trade. Rien van de Weygaert acknowledges support by the John Templeton Foundation grant nr. FP5136-O. He is also very grateful for the hospitality of the Perimeter Institute, Waterloo, Canada, which allowed us to finish the present study. Pratyush Pranav acknowledges the support of the ERC advanced grand ARTHUS (no:740021).

## References

- [1] G. F. Smoot, C. L. Bennett, A. Kogut, E. L. Wright, J. Aymon, N. W. Boggess et al., *Structure in the COBE Differential Microwave Radiometer First-Year Maps*, *Astrophysical Journal, Letters* **396** (1992) L1.
- [2] G. Hinshaw, D. Larson, E. Komatsu, D. N. Spergel, C. L. Bennett, J. Dunkley et al., *Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results*, *Astrophysical Journal, Supplement* **208** (2013) 19 [1212.5226].
- [3] Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi et al., *Planck 2018 results. VI. Cosmological parameters*, *arXiv e-prints* (2018) [1807.06209].

- [4] B. J. T. Jones, *Precision Cosmology: The First Half Million Years*. Cambridge University Press, June, 2017, [10.1017/CBO9781139027809](https://doi.org/10.1017/CBO9781139027809).
- [5] D. J. Fixsen, E. S. Cheng, D. A. Cottingham, R. E. Eplee, Jr., R. B. Isaacman, J. C. Mather et al., *Cosmic microwave background dipole spectrum measured by the COBE FIRAS instrument*, *Astrophys. J.* **420** (1994) 445.
- [6] D. J. Fixsen, *The Temperature of the Cosmic Microwave Background*, *Astrophys. J.* **707** (2009) 916 [[0911.1955](https://arxiv.org/abs/0911.1955)].
- [7] G. F. Smoot, C. L. Bennett, A. Kogut and et. al., *Structure in the COBE differential microwave radiometer first-year maps*, *Astrophysical Journal, Letters* **396** (1992) L1.
- [8] C. L. Bennett, M. Halpern, G. Hinshaw and et. al., *First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Preliminary Maps and Basic Results*, *Astrophys. J. Suppl.* **148** (2003) 1 [[arXiv:astro-ph/0302207](https://arxiv.org/abs/astro-ph/0302207)].
- [9] D. N. Spergel, R. Bean, O. Doré and et. al., *Three-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Implications for Cosmology*, *Astrophys. J. Suppl.* **170** (2007) 377 [[arXiv:astro-ph/0603449](https://arxiv.org/abs/astro-ph/0603449)].
- [10] WMAP COLLABORATION collaboration, *Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation*, *Astrophys.J.Suppl.* **192** (2011) 18 [[1001.4538](https://arxiv.org/abs/1001.4538)].
- [11] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, F. Arroja, M. Ashdown et al., *Planck 2015 results. XVII. Constraints on primordial non-Gaussianity*, *Astron. Astrophys.* **594** (2016) A17 [[1502.01592](https://arxiv.org/abs/1502.01592)].
- [12] Planck Collaboration, Y. Akrami, F. Arroja, M. Ashdown, J. Aumont, C. Baccigalupi et al., *Planck 2018 results. IX. Constraints on primordial non-Gaussianity*, *arXiv e-prints* (2019) [arXiv:1905.05697](https://arxiv.org/abs/1905.05697) [[1905.05697](https://arxiv.org/abs/1905.05697)].
- [13] M. Longuet-Higgins, *Mathematical analysis of random noise*, *Philosophical transactions of the royal society london A* **249** (1957) 321.
- [14] R. J. Adler, *The Geometry of Random Fields*. Chichester: Wiley, 1981.
- [15] J. M. Bardeen, J. R. Bond, N. Kaiser and A. S. Szalay, *The statistics of peaks of Gaussian random fields*, *Astrophys. J.* **304** (1986) 15.
- [16] R. Adler and J. Taylor, *Random Fields and Geometry*, Springer Monographs in Mathematics. Springer New York, 2009.
- [17] C.-G. Park, *Non-Gaussian signatures in the temperature fluctuation observed by the Wilkinson Microwave Anisotropy Probe*, *Mon. Not. R. Astron. Soc.* **349** (2004) 313 [[astro-ph/0307469](https://arxiv.org/abs/astro-ph/0307469)].
- [18] A. H. Guth and S.-Y. Pi, *Fluctuations in the new inflationary universe*, *Physical Review Letters* **49** (1982) 1110.
- [19] D. Baumann, *TASI Lectures on Inflation*, *arXiv:0907.5424* (2009) [[0907.5424](https://arxiv.org/abs/0907.5424)].
- [20] X. Chen, *Primordial Non-Gaussianities from Inflation Models*, *Advances in Astronomy* **2010** (2010) 638979 [[1002.1416](https://arxiv.org/abs/1002.1416)].
- [21] N. Bartolo, S. Matarrese and A. Riotto, *Signatures of primordial non-Gaussianity in the large-scale structure of the universe*, *Journal of Cosmology and Astroparticle Physics* **10** (2005) 10 [[astro-ph/0501614](https://arxiv.org/abs/astro-ph/0501614)].
- [22] P. Pranav, R. van de Weygaert, G. Vegter, B. J. T. Jones, R. J. Adler, J. Feldbrugge et al., *Topology and geometry of Gaussian random fields I: on Betti numbers, Euler characteristic, and Minkowski functionals*, *Mon. Not. R. Astron. Soc.* **485** (2019) 4167 [[1812.07310](https://arxiv.org/abs/1812.07310)].

- [23] P. Pranav, R. van de Weygaert, G. Vegter, B. J. T. Jones, R. J. Adler, J. Feldbrugge et al., *Topology and Geometry of Gaussian random fields II: Persistence Characteristics*, *Mon. Not. R. Astron. Soc.* (2019, in prep.) .
- [24] H. K. Eriksen, D. I. Novikov, P. B. Lilje, A. J. Banday and K. M. Górski, *Testing for Non-Gaussianity in the Wilkinson Microwave Anisotropy Probe Data: Minkowski Functionals and the Length of the Skeleton*, *Astrophys. J.* **612** (2004) 64 [[astro-ph/0401276](#)].
- [25] A. G. Doroshkevich, *Spatial structure of perturbations and origin of galactic rotation in fluctuation theory*, *Astrophysics* **6** (1970) 320.
- [26] J. R. Gott, III, A. L. Melott and M. Dickinson, *The sponge-like topology of large-scale structure in the universe*, *Astrophys. J.* **306** (1986) 341.
- [27] K. R. Mecke, T. Buchert and H. Wagner, *Robust morphological measures for large-scale structure in the universe*, *Astron. Astrophys.* **288** (1994) 697 [[arXiv:astro-ph/9312028](#)].
- [28] J. Schmalzing and T. Buchert, *Beyond Genus Statistics: A Unifying Approach to the Morphology of Cosmic Structure*, *Astrophysical Journal, Letters* **482** (1997) L1+ [[arXiv:astro-ph/9702130](#)].
- [29] C. Park, P. Pranav, P. Chingangbam, R. van de Weygaert, B. Jones, G. Vegter et al., *Betti Numbers of Gaussian Fields*, *Journal of Korean Astronomical Society* **46** (2013) 125 [[1307.2384](#)].
- [30] H. Tomita, *Statistics and geometry of random interface systems*, in *Formation, Dynamics, and Statistics of Patterns*, K. Kawasaki and M. Suzuki, eds., vol. 1, p. 113, 1993, [DOI](#).
- [31] S. Appleby, P. Chingangbam, C. Park, S. E. Hong, J. Kim and V. Ganesan, *Minkowski Tensors in Two Dimensions: Probing the Morphology and Isotropy of the Matter and Galaxy Density Fields*, *Astrophys. J.* **858** (2018) 87 [[1712.07466](#)].
- [32] D. Pogosyan, C. Pichon and C. Gay, *Non-Gaussian extrema counts for CMB maps*, *Physical Review D* **84** (2011) 083510 [[1107.1863](#)].
- [33] S. Codis, C. Pichon, D. Pogosyan, F. Bernardeau and T. Matsubara, *Non-Gaussian Minkowski functionals and extrema counts in redshift space*, *Mon. Not. R. Astron. Soc.* **435** (2013) 531 [[1305.7402](#)].
- [34] D. Munshi, P. Coles and A. Heavens, *Probing CMB Secondary Anisotropies through Minkowski Functionals*, *arXiv e-prints* (2012) arXiv:1207.6217 [[1207.6217](#)].
- [35] H. I. Modest, C. Räth, A. J. Banday, G. Rossmannith, R. Sütterlin, S. Basak et al., *Scale-dependent non-Gaussianities in the CMB data identified with Minkowski functionals and scaling indices*, *Mon. Not. R. Astron. Soc.* **428** (2013) 551 [[1209.5106](#)].
- [36] T. Buchert, M. J. France and F. Steiner, *Model-independent analyses of non-Gaussianity in Planck CMB maps using Minkowski functionals*, *Classical and Quantum Gravity* **34** (2017) 094002 [[1701.03347](#)].
- [37] A. Ducout, *Planck results - Non Gaussianity constrained with Minkowski Functionals*, in *SF2A-2013: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics*, L. Cambresy, F. Martins, E. Nuss and A. Palacios, eds., pp. 337–342, Nov, 2013.
- [38] W. Zhao, *Probing the CMB cold spot through local Minkowski functionals*, *Research in Astronomy and Astrophysics* **14** (2014) 625 [[1209.4021](#)].
- [39] C. P. Novaes, A. Bernui, G. A. Marques and I. S. Ferreira, *Local analyses of Planck maps with Minkowski functionals*, *Mon. Not. R. Astron. Soc.* **461** (2016) 1363 [[1606.04075](#)].
- [40] V. Ganesan and P. Chingangbam, *Tensor Minkowski Functionals: first application to the CMB*, *Journal of Cosmology and Astro-Particle Physics* **2017** (2017) 023 [[1608.07452](#)].

- [41] Edelsbrunner, Letscher and Zomorodian, *Topological persistence and simplification*, *Discrete & Computational Geometry* **28** (2002) 511.
- [42] H. Edelsbrunner and J. Harer, *Computational Topology, An introduction*. AMS Press, 2009.
- [43] E. Betti, *Sopra gli spazi di un numero qualunque di dimensioni*, *Ann. Mat. Pura Appl.* **2** (1871) .
- [44] J. R. Munkres, *Topological results in combinatorics*, *Michigan Math. J.* **31** (1984) 113.
- [45] H. Edelsbrunner and E. Mücke, *Three-dimensional alpha shapes*, *arXiv Mathematics e-prints* (1994) math/9410208 [[math/9410208](#)].
- [46] A. Zomorodian and G. Carlsson, *Computing persistent homology*, *Discrete & Computational Geometry* **33** (2005) 249.
- [47] V. Robins, *Betti number signatures of homogeneous Poisson point processes*, *Physical Review E* **74** (2006) 061107 [[math-ph/0609061](#)].
- [48] L. Wasserman, *Topological Data Analysis*, *Annual Review of Statistics and Its Application* **5** (2018) 501.
- [49] R. van de Weygaert, E. Platen, G. Vegter, B. Eldering and N. Kruithof, *Alpha shape topology of the cosmic web*, in *Proceedings of the 2010 International Symposium on Voronoi Diagrams in Science and Engineering*, M. A. Mostafavi, ed., pp. 224–234, IEEE Computer Society, 2010, [1306.3640](#).
- [50] R. van de Weygaert, G. Vegter, H. Edelsbrunner, B. J. T. Jones, P. Pranav, C. Park et al., *Alpha, betti and the megaparsec universe: On the topology of the cosmic web*, in *Transactions on Computational Science XIV*, M. L. Gavrilova, C. K. Tan and M. A. Mostafavi, eds., (Berlin, Heidelberg), pp. 60–101, Springer-Verlag, 2011, [1306.3640](#).
- [51] T. Sousbie, *The persistent cosmic web and its filamentary structure - I. Theory and implementation*, *Mon. Not. R. Astron. Soc.* **414** (2011) 350 [[1009.4015](#)].
- [52] P. Pranav, H. Edelsbrunner, R. van de Weygaert, G. Vegter, M. Kerber, B. J. T. Jones et al., *The topology of the cosmic web in terms of persistent Betti numbers*, *Mon. Not. R. Astron. Soc.* **465** (2017) 4281 [[1608.04519](#)].
- [53] M. Morse, *Relations between the critical points of a real function on  $n$  independent variables*, *Transactions of the American Mathematical Society* **27** (1925) 345.
- [54] J. Milnor, *Morse theory*, Based on lecture notes by M. Spivak and R. Wells. Annals of Mathematics Studies, No. 51. Princeton University Press, Princeton, N.J., 1963.
- [55] G. Petri, P. Expert, F. Turkheimer, R. Carhart-Harris, D. Nutt, P. J. Hellyer et al., *Homological scaffolds of brain functional networks*, *Journal of The Royal Society Interface* **11** (2014) 20140873.
- [56] M. W. Reimann, M. Nolte, M. Scolamiero, K. Turner, R. Perin, G. Chindemi et al., *Cliques of neurons bound into cavities provide a missing link between structure and function*, *Frontiers in Computational Neuroscience* **11** (2017) 48.
- [57] Y. Hiraoka, T. Nakamura, A. Hirata, E. G. Escolar, K. Matsue and Y. Nishiura, *Hierarchical structures of amorphous solids characterized by persistent homology*, *Proceedings of the National Academy of Sciences* **113** (2016) 7035.
- [58] T. Sousbie, C. Pichon and H. Kawahara, *The persistent cosmic web and its filamentary structure - II. Illustrations*, *Mon. Not. R. Astron. Soc.* **414** (2011) 384 [[1009.4014](#)].
- [59] N. Shivashankar, P. Pranav, V. Natarajan, R. v. d. Weygaert, E. G. P. Bos and S. Rieder, *Felix: A topology based framework for visual exploration of cosmic filaments*, *IEEE Transactions on Visualization and Computer Graphics* **22** (2016) 1745.

- [60] J. R. Bond, L. Kofman and D. Pogosyan, *How filaments of galaxies are woven into the cosmic web*, *Nature* **380** (1996) 603 [[arXiv:astro-ph/9512141](#)].
- [61] R. van de Weygaert and J. R. Bond, *Clusters and the Theory of the Cosmic Web*, in *A Pan-Chromatic View of Clusters of Galaxies and the Large-Scale Structure*, M. Plionis, O. López-Cruz and D. Hughes, eds., vol. 740 of *Lecture Notes in Physics*, Berlin Springer Verlag, p. 335, 2008, [DOI](#).
- [62] M. A. Aragón-Calvo, R. van de Weygaert and B. J. T. Jones, *Multiscale phenomenology of the cosmic web*, *Mon. Not. R. Astron. Soc.* **408** (2010) 2163 [[1007.0742](#)].
- [63] M. Cautun, R. van de Weygaert, B. J. T. Jones and C. S. Frenk, *Evolution of the cosmic web*, *Mon. Not. R. Astron. Soc.* **441** (2014) 2923 [[1401.7866](#)].
- [64] N. I. Libeskind, R. van de Weygaert, M. Cautun, B. Falck, E. Tempel, T. Abel et al., *Tracing the cosmic web*, *Mon. Not. R. Astron. Soc.* **473** (2018) 1195 [[1705.03021](#)].
- [65] Y. Kimura and K. Imai, *Quantification of LSS using the persistent homology in the SDSS fields*, *Advances in Space Research* **60** (2017) 722.
- [66] X. Xu, J. Cisewski-Kehe, S. B. Green and D. Nagai, *Finding cosmic voids and filament loops using topological data analysis*, *Astronomy and Computing* **27** (2019) 34 [[1811.08450](#)].
- [67] S. Codis, D. Pogosyan and C. Pichon, *On the connectivity of the cosmic web: theory and implications for cosmology and galaxy formation*, *Mon. Not. R. Astron. Soc.* **479** (2018) 973 [[1803.11477](#)].
- [68] W. Elbers and R. van de Weygaert, *Persistent topology of the reionization bubble network - I. Formalism and phenomenology*, *Mon. Not. R. Astron. Soc.* **486** (2019) 1523 [[1812.00462](#)].
- [69] I. Makarenko, A. Shukurov, R. Henderson, L. F. S. Rodrigues, P. Bushby and A. Fletcher, *Topological signatures of interstellar magnetic fields - I. Betti numbers and persistence diagrams*, *ArXiv e-prints* (2017) [[1708.04558](#)].
- [70] J. Feldbrugge and M. van Engelen, *Analysis of Betti numbers and persistence diagrams of two-dimensional Gaussian random fields*. BSc. thesis, Univ. Groningen, 2012.
- [71] M. Wintraecken and G. Vegter, *A geometrical take on invariants of low-dimensional manifolds found by integration*, *Topology and its applications* **160** (2013) 2175.
- [72] A. Gyulassy, *Combinatorial Construction of Morse-Smale Complexes*, Ph.D. thesis, University of California Davis, 2008.
- [73] C. J. A. Delfinado and H. Edelsbrunner, *An incremental algorithm for betti numbers of simplicial complexes on the 3-sphere*, *Computer Aided Geometric Design* **12** (1995) 771.
- [74] P. Chingangbam, C. Park, K. P. Yogendran and R. van de Weygaert, *Hot and Cold Spot Counts as Probes of Non-Gaussianity in the Cosmic Microwave Background*, *Astrophys. J.* **755** (2012) 122 [[1206.0436](#)].
- [75] A. Cole and G. Shiu, *Persistent homology and non-Gaussianity*, *Journal of Cosmology and Astroparticle Physics* **3** (2018) 025 [[1712.08159](#)].
- [76] P. Pranav, R. J. Adler, T. Buchert, H. Edelsbrunner, B. J. T. Jones, A. Schwartzman et al., *Unexpected topology of the temperature fluctuations in the cosmic microwave background*, *Astron. Astrophys.* **627** (2019) A163 [[1812.07678](#)].
- [77] R. van de Weygaert and E. Bertschinger, *Peak and gravity constraints in Gaussian primordial density fields: An application of the Hoffman-Ribak method*, *Mon. Not. R. Astron. Soc.* **281** (1996) 84 [[astro-ph/9507024](#)].
- [78] X. Chen, M.-x. Huang, S. Kachru and G. Shiu, *Observational signatures and non-Gaussianities of general single-field inflation*, *Journal of Cosmology and Astro-Particle Physics* **2007** (2007) 002 [[hep-th/0605045](#)].

- [79] J. Martin, C. Ringeval and V. Vennin, *Encyclopædia Inflationaris*, *Physics of the Dark Universe* **5** (2014) 75 [[1303.3787](#)].
- [80] D. Babich, P. Creminelli and M. Zaldarriaga, *The shape of non-Gaussianities*, *Journal of Cosmology and Astro-Particle Physics* **2004** (2004) 009 [[astro-ph/0405356](#)].
- [81] D. Seery and J. E. Lidsey, *Primordial non-Gaussianities in single-field inflation*, *Journal of Cosmology and Astro-Particle Physics* **2005** (2005) 003 [[astro-ph/0503692](#)].
- [82] X. Chen, *Primordial Non-Gaussianities from Inflation Models*, *Advances in Astronomy* **2010** (2010) 638979 [[1002.1416](#)].
- [83] X. Chen and Y. Wang, *Quasi-single field inflation and non-Gaussianities*, *Journal of Cosmology and Astro-Particle Physics* **2010** (2010) 027 [[0911.3380](#)].
- [84] D. Seery and J. E. Lidsey, *Primordial non-Gaussianities from multiple-field inflation*, *Journal of Cosmology and Astro-Particle Physics* **2005** (2005) 011 [[astro-ph/0506056](#)].
- [85] X. Chen, R. Easther and E. A. Lim, *Large non-Gaussianities in single-field inflation*, *Journal of Cosmology and Astro-Particle Physics* **2007** (2007) 023 [[astro-ph/0611645](#)].
- [86] E. Bertschinger, *Large-Scale Structures and Motions: Linear Theory and Statistics*, in *New Insights into the Universe*, V. J. Martinez, M. Portilla and D. Saez, eds., vol. 408 of *Lecture Notes in Physics*, Berlin Springer Verlag, p. 65, 1992, [DOI](#).
- [87] S. Rice, *Mathematical analysis of random noise*, *Bell system technology journal* **23** (1944) 282.
- [88] W. Tutte, W. Tutte and C. Nash-Williams, *Graph Theory*, Cambridge Mathematical Library. Cambridge University Press, 2001.
- [89] J. Vick, *Homology Theory: An Introduction to Algebraic Topology*, Graduate texts in mathematics. Academic Press, 1994.
- [90] G. Rote and G. Vegter, *Computational topology: An introduction*, in *Effective Computational Geometry for Curves and Surfaces*, J.-D. Boissonnat and M. Teillaud, eds., (Berlin, Heidelberg), pp. 277–312, Springer, 2006.
- [91] V. Robins, *Algebraic topology*, in *Mathematical Tools for Physicists*, *Encyclopedia of Applied Physics*, M. Grinfeld, ed., (UK), pp. 211–239, Wiley - VCH Verlag GmbH, 2015, [1304.7846](#).
- [92] A. J. S. Hamilton, J. R. Gott, III and D. Weinberg, *The topology of the large-scale structure of the universe*, *Astrophys. J.* **309** (1986) 1.
- [93] F. Hoyle, M. S. Vogeley, J. R. Gott, III, M. Blanton, M. Tegmark, D. H. Weinberg et al., *Two-dimensional Topology of the Sloan Digital Sky Survey*, *Astrophys. J.* **580** (2002) 663 [[astro-ph/0206146](#)].
- [94] R. J. Adler, J. E. Taylor and K. J. Worsley, *Applications of random fields and geometry: Foundations and case studies*, 2010.
- [95] D. Pogosyan, S. Codis and C. Pichon, *Non Gaussian Minkowski functionals and extrema counts for CMB maps*, in *The Zeldovich Universe: Genesis and Growth of the Cosmic Web*, R. van de Weygaert, S. Shandarin, E. Saar and J. Einasto, eds., vol. 308 of *IAU Symposium*, pp. 61–66, Oct., 2016, [1607.02268](#), [DOI](#).
- [96] R. J. Adler, O. Bobrowski, M. S. Borman, E. Subag and S. Weinberger, *Persistent Homology for Random Fields and Complexes*, *arXiv e-prints* (2010) arXiv:1003.1001 [[1003.1001](#)].
- [97] R. J. Adler, O. Bobrowski and S. Weinberger, *Crackle: The Persistent Homology of Noise*, *arXiv e-prints* (2013) arXiv:1301.1466 [[1301.1466](#)].
- [98] S. Colombi, D. Pogosyan and T. Souradeep, *Tree Structure of a Percolating Universe*, *Physical Review Letters* **85** (2000) 5515 [[astro-ph/0011293](#)].

- [99] T. Sousbie, C. Pichon, S. Colombi, D. Novikov and D. Pogosyan, *The 3D skeleton: tracing the filamentary structure of the Universe*, *Mon. Not. R. Astron. Soc.* **383** (2008) 1655 [[0707.3123](#)].
- [100] M. A. Aragón-Calvo, E. Platen, R. van de Weygaert and A. S. Szalay, *The Spine of the Cosmic Web*, *Astrophys. J.* **723** (2010) 364 [[0809.5104](#)].
- [101] A. Gyulassy, V. Natarajan, V. Pascucci and B. Hamann, *Efficient computation of Morse-Smale complexes for three-dimensional scalar fields*, *IEEE Trans Vis Comput Graph.* **13** (2007) 1440.
- [102] H. Edelsbrunner, J. Harer and A. Zomorodian, *Hierarchical morse complexes for piecewise linear 2-manifolds*, in *Proceedings of the Seventeenth Annual Symposium on Computational Geometry*, SCG '01, (New York, NY, USA), pp. 70–79, ACM, 2001, [DOI](#).
- [103] R. Feynman and A. Hibbs, *Quantum mechanics and path integrals*, International series in pure and applied physics. McGraw-Hill, 1965.
- [104] J. Feldbrugge and R. van de Weygaert *Journal of Cosmology and Astro-Particle Physics* (2020, in prep.) .
- [105] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, F. Arroja, M. Ashdown et al., *Planck 2015 results. XVII. Constraints on primordial non-Gaussianity*, *Astron. Astrophys.* **594** (2016) A17 [[1502.01592](#)].
- [106] J. Feldbrugge and R. van de Weygaert *Journal of Cosmology and Astro-Particle Physics* (2020, in prep.) .
- [107] A. P. S. Yadav and B. D. Wandelt, *Primordial Non-Gaussianity in the Cosmic Microwave Background*, *Advances in Astronomy* **2010** (2010) 565248 [[1006.0275](#)].
- [108] K. Nevenzeel, G. Wilding, R. van de Weygaert, G. Vegter, P. Pranav, B. J. T. Jones et al., *Persistent homology of the cosmic web. I. Method and application to  $\Lambda$ CDM simulations*, *Mon. Not. R. Astron. Soc.* (2019, to be subm.) .
- [109] G. Vegter, *Chapt. Computational Topology*, pp. 517 – 536. CRC Press, Inc., Boca Raton, FL, USA, 1997.
- [110] A. Okabe, *Spatial tessellations: concepts and applications of Voronoi diagrams*, Wiley series in probability and statistics: Applied probability and statistics. Wiley, 2000.

## A Concepts of Homology theory

This appendix presents a formal introduction to homology theory and Betti numbers. A more extensive treatment and discussion can be found in several books and reviews [[42](#), [90](#), [91](#), [109](#)]. Appendices [A.1](#), [A.2](#), [A.3](#) present the formal definitions of simplicial complexes, chains and cycles and persistence diagrams. Appendix [A.4](#) proceeds with a case study of a simple example of how to compute Betti numbers from first principle.

### A.1 Simplicies

Networks and graphs are well-known concepts in mathematics. They describe the connections and relations between points or objects. In geometry they appear as tessellations, tilings of space composed of points, lines, surfaces and cells that in combination represent a geometric object [see e.g. [110](#)]. They are known as a *Simplicial Complex*. A representative example of a simplicial complex in two-dimensional Euclidian space is shown in figure [14](#). The points, lines, and surfaces are known as vertices, edges, and faces or, in a more formal fashion, as

0-, 1- and 2-simplices. We write a simplex by listing its corresponding vertices (points), *i.e.*, the line segment between point  $A$  and  $B$  is represented by the edge  $AB$ . The vertices  $A$ ,  $B$ , and  $C$  can span a face  $ABC$ . In the present study we only consider simplicial complexes embedded in 2-dimensional Euclidean space. However, simplicial complexes can be defined in any  $D$ -dimensional space.

Remark that the homology of a manifold can be captured by these simplices using a polyhedral decomposition of the manifold. The homology of a manifold is well-defined as it is independent of the chosen decomposition. In the subsequent sections we always assume a single simplicial decomposition of a manifold.

## A.2 Chains and the Boundary Operator

A simplicial complex is composed of simplices: vertices, edges, and faces. Using a simplicial decomposition of the manifold  $M$ , we can construct the corresponding simplicial complex. Topology is interested in the global properties of the complex. We here we define a composite object known as a *chain*. On these chains we define the *boundary operator* which maps chains to their boundaries.

### A.2.1 Chain

A  $p$ -chain is a collection of  $p$ -simplices of the complex. Formally a  $p$ -chain is a finite sum

$$c = \sum_{\sigma} c_{\sigma} \sigma,$$

where  $\sigma$  are  $p$ -simplices and  $c_{\sigma} \in \mathbb{Z}/2\mathbb{Z}$ <sup>11</sup>. The simplex  $\sigma$  is included in the chain  $c$  if and only if  $c_{\sigma} = 1$ . The collection of all  $p$ -chains, corresponding to a simplicial decomposition of the manifold  $M$ , forms the chain group  $C_p(M)$ .

### A.2.2 Boundary & Boundary Operator

The "boundary" of a  $p$ -simplex consists of a number of  $(p - 1)$ -simplices. Given a  $p$ -simplex

$$\sigma = v_{i_0} v_{i_1} \cdots v_{i_p} \tag{A.1}$$

with vertices  $v_i$ , we define the boundary of the simplex as

$$\partial_p(\sigma) = \sum_{j=0}^p v_{i_0} \cdots v_{i_{j-1}} v_{i_{j+1}} \cdots v_{i_p},$$

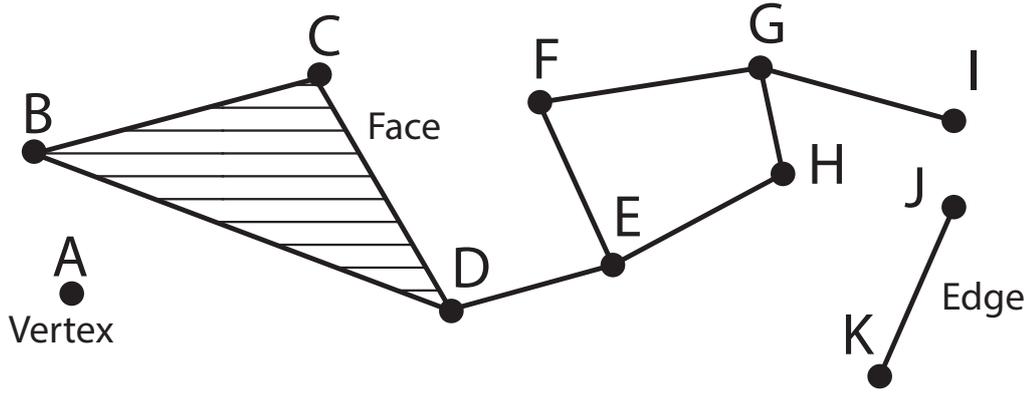
where the vertex  $v_{i_j}$  has been dropped from the simplex. The zero-dimensional boundary operator  $\partial_0$  is defined to be the zero-function, *i.e.*,  $\partial_0(v_i) = 0$  for all vertices  $v_i$ . The boundary operator can be extended to chains assuming linearity, *i.e.*,

$$\partial_p \left( \sum_{\sigma} c_{\sigma} \sigma \right) = \sum_{\sigma} c_{\sigma} \partial_p(\sigma), \tag{A.2}$$

making the boundary operator  $\partial_p$  a map from the  $p$ -dimensional chain group  $C_p$  to the  $(p - 1)$ -dimensional chain group  $C_{p-1}$ .

---

<sup>11</sup>The group  $\mathbb{Z}/2\mathbb{Z}$  can be seen as  $\{0, 1\}$  with the property that  $1 + 0 = 1$  and  $1 + 1 = 0$ .



**Figure 14:** Simplicial Complex. An example of a simplicial complex, or tessellation, in two-dimensional Euclidean space. The letters mark the points known as vertices. The lines represent the edges and the shaded region represents a face.

The boundary operator takes a chain and returns the boundary of that chain. For illustrative purposes consider examples extracted from the simplicial complex in figure 14. In these examples, we apply the boundary operator on several chains in the simplicial complex. The boundary of the edge  $JK$  consists of the boundary vertices  $J$  and  $K$ , *i.e.*, boundaries, *i.e.*

$$\partial_1(JK) = J + K. \quad (\text{A.3})$$

The boundary operator returns an empty chain when operating on a loop or more formally a *cycle*. Consider for example the boundary of the cycle  $EF + FG + GH + EH$ , *i.e.*,

$$\partial_1(EF + FG + GH + EH) = E + F + F + G + G + H + E + H \quad (\text{A.4})$$

$$= 0. \quad (\text{A.5})$$

By considering the face  $BCD$  we observe that the boundary of a boundary vanishes *i.e.*,

$$\partial_1(\partial_2(BCD)) = \partial_1(CD + BD + BC) \quad (\text{A.6})$$

$$= C + D + B + D + B + C \quad (\text{A.7})$$

$$= 0. \quad (\text{A.8})$$

The fact that the square of the boundary operator vanishes, *i.e.*,

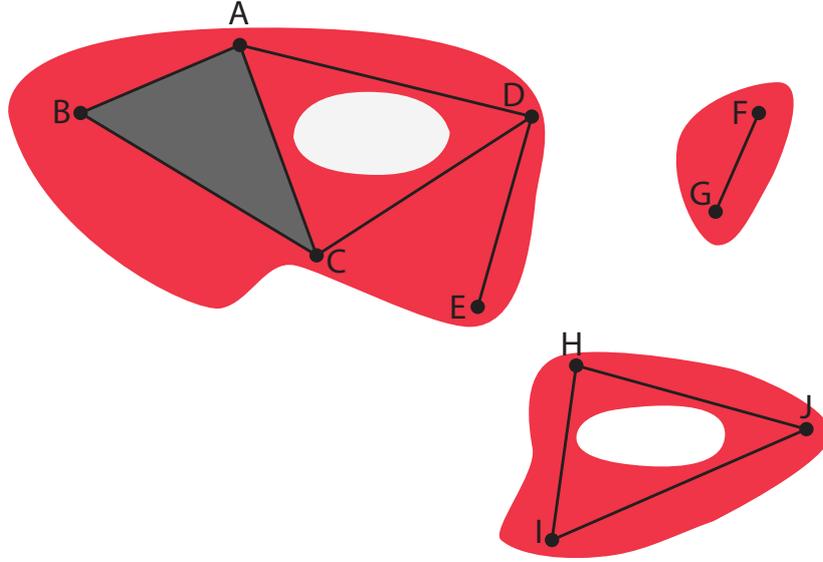
$$(\partial_{p-1} \circ \partial_p)(c) = 0. \quad (\text{A.9})$$

for a general  $p$ -chain  $c$ , can be proven from the definition of the boundary operator. In the next section we use this exactness property of chains to define the homology group.

### A.3 Homology Groups and Betti Numbers

Using the boundary operator acting on the chain groups corresponding to a simplicial decomposition of the manifold  $M$ , we construct an exact sequence

$$\dots \xrightarrow{\partial_{p+2}} C_{p+1}(M) \xrightarrow{\partial_{p+1}} C_p(M) \xrightarrow{\partial_p} C_{p-1}(M) \xrightarrow{\partial_{p-1}} \dots$$



**Figure 15:** Chain groups simplicial complex. A manifold is shown in red, with a specified collection of chains from the chain group.

The sequence is exact since  $\partial_{p-1} \circ \partial_p = 0$ . Using the boundary operator we construct two subsets of the chain group  $C_p(M)$ . We define the  $p$ -cycles  $Z_p(M)$  and the  $p$ -boundaries  $B_p(M)$  as the image and the kernel of the boundary operator, *i.e.*,

$$\begin{aligned} Z_p(M) &= \text{Ker } \partial_p, \\ B_p(M) &= \text{Im } \partial_{p+1}. \end{aligned}$$

The exactness ensures that  $B_p(M) \subseteq C_p(M)$  is a linear subgroup of  $Z_p(M) \subseteq C_p(M)$ .

The  $p^{\text{th}}$  homology group of  $M$  is defined as the quotient group

$$H_p(M) = Z_p(M) / B_p(M).$$

The group  $H_p(M)$  should be seen as a partition of  $Z_p(M)$  in sets (equivalence classes) that differ by an element in  $B_p(M)$ .

The  $p^{\text{th}}$  Betti number of  $M$  is defined as

$$\beta_p = \text{rank } H_p(M),$$

which is the number of equivalence classes in  $H_p(M)$ . Note that for a compact manifold, the Betti numbers are finite.

#### A.4 Example

On the basis of figure 15, we provide a visual illustration of the significance of the Betti numbers of a manifold  $M$  by following the method outlined above. Instead of working with a simplicial decomposition of the manifold, we consider the homology of the homologically equivalent simplicial representation denoted by the vertices, edges, and faces.

The chain groups of the simplicial complex are given by the formal sums of the vertices, edges and faces present in the simplicial complex:

$$\begin{aligned}\mathcal{C}_0 &= \{A, B, C, D, E, F, G, H, I, J\} \\ \mathcal{C}_1 &= \{AB, AC, AD, BC, CD, DE, FG, HI, HJ, IJ\} \\ \mathcal{C}_2 &= \{ABC\},\end{aligned}$$

where we denote a  $p$ -simplex by its corresponding vertices. For these chain groups the boundary groups are generated by the chains:

$$\begin{aligned}\mathcal{B}_0 &= \{A + B, A + C, A + D, B + C, C + D, D + E, F + G, H + I, H + J, I + J\} \\ \mathcal{B}_1 &= \{AB, AC, BC\} \\ \mathcal{B}_2 &= \{0\},\end{aligned}$$

since  $\partial_1(AB) = A + B$ ,  $\partial_1(AC) = A + C$  etc. The cycle groups are generated by the chains:

$$\begin{aligned}\mathcal{Z}_0 &= \mathcal{C}_0 \\ \mathcal{Z}_1 &= \{AB + AC + BC, AC + AD + CD, HI + HJ + IJ\} \\ \mathcal{Z}_2 &= \{0\},\end{aligned}$$

since  $\partial_1(AB + AC + BC) = A + B + A + C + B + C = 0$  etc. We evaluate homology groups by partitioning  $\mathcal{Z}_i$  into subsets whose elements differ by an element generated by  $\mathcal{B}_i$  for  $i = 0, 1, 2$ . The homology groups are given by

$$\begin{aligned}H_0 &= \{A, F, H\} \\ H_1 &= \{AC + AD + CD, HI + HJ + IJ\} \\ H_2 &= \{0\},\end{aligned}$$

where we represent the equivalence classes by a single element of the class. As the Betti numbers are the number of equivalence classes in the homology groups, we find the Betti numbers

$$\beta_0 = 3, \beta_1 = 2 \text{ and } \beta_2 = 0.$$

This is what one would expect, as the manifold  $M$  has three connected components and two holes.