Persistent holes in the Universe

Pranav, Pratyush

5

Discussions and Conclusions

Cosmology has entered the era of *big data* in the last decade. With the advent of new ground-based as well as space telescopes in commission or scheduled to be commision, the amount of data is in the order of petabytes and exabytes. On the one hand, the PLANCK mission and the PLANCK satellite are furnishing us with an unprecedented detailed measurement of the temperature ianisotropies in the Cosmic microwave background. On the other hand, ground based redshift surveys and telescopes like the Large Synoptic Survey Telescope (LSST) commisioned under the Dark Energy Survey (DES) project will trace billions or remote galaxies and provide multiple probes for the mysterious dark matter and dark energy. These surveys and telescopes are just a few to name in the plethora of missions planned to launch a coordinated attack on some of the recently discovered mysteries spewed at us. One of the most interesting among them is the realization that the Universe is expanding at an accelerating pace rather than slowing down. This has called for a renewed interest in the nature of dark energy and its properties. The recent influx of massive data in cosmology and related disciplines, calls for new methods of data analysis tailored to harness and extract the relevant information from the massive data in a systematic fashion.

Topological anaysis of cosmological datasets has a long history, and has been one of the principal pillars of investigation in the cosmological community for decades. This has mainly been achieved by describing and analyzing the features of the cosmic mass distrubution through Eluer characteristic, genus and Minkowsk functionals. There use also has been more heuristic in nature mainly aimed at discriminating between various models of cosmic mass distribution. While they have supplied a wealth of information on the nature of the topology and morphology of the cosmic mass distribution, there is a motivativation to introduce new topological methods for analysis, that tie into the salient features of the structure seen in the cosmos.

This thesis is motivated by recognizing the limitations of the existing methods in describing topology, to introduce new measures that are able to harness the topological information of the cosmic mass distribution in a greater detail. Topological data analysis, in particular persistence based analysis, of structural patterns has gained interest across various disciplines like medical imaging, cartography, agriculture etc., motivated by the similarity in the nature of the probelms approached. Morse theory and persistence based approach has also picked up interest in the cosmological community in the last few year, and has been applied to develop recipes for detecting and describing the structural patterns seen in the Cosmos. Noteworthy of metion among such methods are the SpineWeb formalism and the DiSperSe formalism (Aragón-Calvo et al. 2010; Sousbie 2011). These formalisms however have focussed solely within the scope of pattern recognition.

This thesis is a culmination of an interisciplinary collaboration beween the fields of cosmology, mathematics and computer science. On the front of cosmology, it has aimed to expand the scope of the analysis based on topological formalisms emanating from Morse theory, homology and persistence of the cosmic mass distribution. This has been through a comprehensive approach of integrating the quantitative topological analysis of models of cosmic mass distribution in terms of persistence and homology, with the more traditional approach of structure identification and detection. On the mathematical side, it has attempted to explore the general properties and characteristics of persistence homology and persistence diagrams. The properties of peristence diagrams are a topic of active research in the topo-mathematical community (Bubenik 2012). While the thesis has succeeded in establishing an empirical probabilistic view of persistence homology and diagrams through the introduction of the concepts of intensity and intensity maps, an analytical and theoretical framework has yet to be established. It is noteworthy to remember in this context is that a full analytical description of persistence may not be an easy challange, as has aptly been recognized in the wider mathematical community.

5.1 Future directions

5.1.1 Future directions

At the end of this thesis, there are a number of follow up investigations that come to mind. We list here a possible, but not exhaustive, set of directions that may be promising.

Statistics of persistence. We have established empirically that the ensemble average of persistence diagrams of stochastic processes are well-defined. An analytic probabilistic and statistical description of persistence will be a natural and important extension of the framework.

Persistence (Homology) of the large scale Universe. In this thesis, we presented a topological characterization of Gaussian random fields, with a view to understand the nature of fluctuations in the primordial Universe. Subsequently, we analyze the Voronoi models that mimic aspects of the matter distribution in the large scale Uni-

verse. A logical and important extension of this would be to investigate the topology of the genuine distributions of dark matter and halos in LCDM simulations.

Persistence characterization of the anisotropies in CMB. A persistence based hierarchical characterization of the anisotropies in the Cosmic Microwave Background, as provided by the latest measurements from the PLANCK satellite provides an interesting challenge Persistence has the potential to shed light into the hierarchical nature of primordial fluctuations. It may also be interesting to investigate the CMB with a view to isolate signatures of primordial non-Gaussianities. Using existing methodologies, the PLANCK team reports the absence of statistically significant non-Gaussian signals in the CMB maps. The hierarchical description of persistence has the potential to probe deeper in this direction.

Filament catalogues of the large scale Universe. Having tested the robustness of the filament finding software on test models and simulations, an important aspect to follow up is to produce filamentary catalogues from simulations as well as observation. These catalogs will have an option of querying for the properties of the associated galaxies they host, in view of investigating the formation and evolution of galaxies vis-a-vis the large scale environment they reside in.

Hierarchical characterization of Cosmic Voids. 2-dimensional persistence diagrams, that capture the formation and evolution of topological voids, can potentially provide a powerful tool for the characterization of the hierarchical aspects of cosmic voids. This is especially interesting, in view of the fact that topology, perhaps, is the only method that defines a void uniquely and unambiguously, without a choice of any free parameter. It has theoretically been suspected for long, and confirmed recently observationally that the cosmic voids form and evolve hierarchically. Studies have revealed cosmic voids are highly sensitive to dark energy and may contain information about modified gravity.

Characterization of the epoch of reionization. Due to its hierarchical nature, persistence is naturally tailored towards the study of reionization and the evolving network of ionization bubbles in the Epoch of reionization, marking the onset of formation of stars and galaxies. 0-dimensional persistence diagrams are optimally suited for capturing the evolving network of ionization bubbles, which are essentially isolated objects undergoing mergers and possibly more complex procedures. Additionally, studying the characteristics of the 1-dimensional diagrams will reveal details about the percolation properties of the distribution at that epoch. The 2-dimensional diagrams have a potential to reveal the evolving hierarchical network of voids in their infancy.

Appendices

A

Topology

A.1 Minkowski functionals

Suppose we have a solid body, $\mathbb{M}$, whose boundary is a smoothly embedded surface in $\mathbb{R}^3$. This surface may be a sphere or have holes, like the torus, and it may consist of one or several connected components, each with its own holes. Similarly, we do not require that $\mathbb{M}$ is connected. Write $\mathbb{M}^r$ for the set of points at distance r or less from $\mathbb{M}$. For small values of r, the boundary of $\mathbb{M}^r$ will be smoothly embedded in $\mathbb{R}^3$, but as r grows, it will develop singularities and self-intersections. Before this happens, the volume of $\mathbb{M}^r$ can be written as a degree-3 polynomial in r,

$$\mathrm{vol}\,\mathbb{M}^r \;=\; Q_0 + Q_1 r + Q_2 r^2 + Q_3 r^3\,. \tag{A.1}$$

The Q_i are known as the *Minkowski functionals* of $\mathbb{M}$, which are important concepts in integral geometry. For a d-dimensional manifold $\mathbb{M}$ there are $(d+1)$ Minkowski functionals. Minkowski functionals were first introduced as measures of the spatial cosmic mass distribution by Mecke et al. (1994) and have become an important measure of clustering of mass and galaxies (Schmalzing & Buchert 1997; Schmalzing et al. 1999; Sahni et al. 1998). In mathematics, they are closely related to concepts like the *Quermassintegrals*, *mixed volumes*, and *Killing-Lipschitz curvatures* $\mathcal{L}_i(\mathbb{M})$ in differential geometry. These names relate to different geometric interpretations of the Q_i, and in most cases involve a different ordering and normalization.

With respect to Killing-Lipschitz curvatures, we may observe that the expression above is the 3D version of *Weyl's tube formula* (Adler 1981). It is the general expression for the volume of the set of points that are at a distance $\leq r$ from an object $\mathbb{M}$ in d-dimensional space in terms of the Killing-Lipschitz curvatures (Adler & Taylor 2010;

Taylor & Adler 2009; Bobrowski & Borman 2012),

$$\text{vol}\,\mathbb{M}^r \;=\; \sum_0^d r^{d-i}\omega_{d-i}\mathcal{L}_i(\mathbb{M})\,. \tag{A.2}$$

In this expression ω_n is the volume of an n-dimensional unit sphere[1]. From this, we immediately find the identity between the Killing-Lipschitz curvatures $\mathcal{L}_{d-n}$ and the corresponding Minkowski functionals Q_n,

$$Q_n = \mathcal{L}_{d-n}/\omega_n\,. \tag{A.3}$$

In terms of their interpration in the three-dimensional context, following Equation A.1, we see that Q_0 is the volume of $\mathbb{M}$, Q_1 is the area of its boundary, Q_2 is the total mean curvature, and Q_3 is one third of the total Gaussian curvature of the boundary. These interpretations suggest that the Minkowski functionals are essentially geometric in nature, and they are, but there are strong connections to topological concepts as well. The key connection is established via the Euler characteristic, $\chi(S)$, of a surface S. We will discuss the latter notion shortly (see next subsection) but for now we just mention that the Euler characteristic – traditionally denoted as χ – is equal to 2 minus twice the number of holes. For example, the sphere has $\chi = 2$ and the torus has $\chi = 0$. If the boundary of $\mathbb{M}$ consists of k components with a total of h holes, then we have $\chi = 2(k-h)$. The connection between the topological characteristics of a manifold and its geometrical properties is stated by the famous Gauss-Bonnet theorem. For a connected surface S in $\mathbb{R}^3$, the Gauss-Bonnet theorem asserts that the total Gaussian curvature of a closed, is 2π times the Euler characteristic $\chi(S)$,

$$\chi(S) \;=\; \frac{1}{2\pi}\oint\left(\frac{1}{R_1R_2}\right)dS\,, \tag{A.4}$$

where R_1 and R_2 are the principal radii of curvature at each point of the surface. For the situation sketched above, a boundary of manifold $\mathbb{M}$ consisting of k components with a total of h holes, it tells that the total Gaussian curvature will be equal to $4\pi(k-h)$. Finally, $Q_3 = \frac{4\pi}{3}(k-h)$, i.e.

$$\begin{aligned} Q_3 \;&=\; \frac{2\pi}{3}\,\chi(S) \\ \mathcal{L}_0 \;&=\; \frac{1}{2}\,\chi(S)\,. \end{aligned} \tag{A.5}$$

For example, the Gaussian curvature of a sphere with radius r is $1/r^2$ at every point. Multiplying with the area, which is $4\pi r^2$, we get the total Gaussian curvature equal to 4π, which is independent of the radius. This agrees with $\chi = 4\pi(k-h)$ given above since $k-h=1$ in this case.

Important for our purpose is the observation that Minkowski functionals can be

[1] for the 3D situation, the relevant values of ω_k are $\omega_0 = 1$, $\omega_1 = 2$, $\omega_2 = \pi$ and $\omega_3 = \frac{4}{3}\pi$

expressed in terms of Euler integrals. The *Crofton intersection formula* of integral geometry (Crofton 1868) encapsulates a very profound statement on the length of curves, area of surfaces and a plethora of interesting geometric properties in terms of an integral over lower-dimensional intersecting hyperplanes. In a sense, it is the generalization of the famous Buffon's Needle problem (Ramaley 1969). The specific version of Crofton's formula pertaining to integrals over the Euler characteristic is known as *Hadwiger's Formula* (Hadwiger 1957; Adler & Taylor 2010). To evaluate the k-th Minkowski functional of a d-dimensional manifold $\mathbb{M}$, one has to consider the Euler characteristic of the intersection of k-dimensional hyperplanes S_k with $\mathbb{M}$, $\chi(S_k \cap M)$. The value of the Minkowski functional $Q_k(\mathbb{M})$ is equal to the integral of the Euler characteristic $\chi(S_k \cap M)$ over the space $\mathcal{E}_k^d$ of all conceivable hyperplanes S_k (Schmalzing & Buchert 1997),

$$Q_k(\mathbb{M}) \;=\; \frac{\omega_d}{\omega_{d-k}\omega_k} \int_{\mathcal{E}_k^d} \mathrm{d}\mu_k(S_k)\, \chi(S_k \cap M)\,, \tag{A.6}$$

with the normalization constants ω_j are the volumes of j-dimensional unit spheres.

A.2 Euler characteristic and genus

Suppose now that we have the boundary of $\mathbb{M}$ triangulated, using v vertices, e edges, and t triangles. Named after Leonhard Euler (Euler 1758), the *Euler characteristic* of the surface is the alternating sum of the number of simplices:

$$\chi = v - e + t\,. \tag{A.7}$$

It does not depend on the triangulation, only on the surface. For example, we can triangulate the sphere with 4 vertices, 6 edges, and 4 triangles, like the boundary of the tetrahedron, which gives $\chi = 4 - 6 + 4 = 2$. Alternatively, we may triangulate it with 6 vertices, 12 edges, and 8 triangles, like the boundary of the octahedron, which again gives $\chi = 6 - 12 + 8 = 2$.

As mentioned above, the Euler characteristic of a connected, closed surface with $h \geq 0$ holes is $\chi = 2 - 2h$. To make this more concrete, we formalize the number of holes of a closed, connected surface to its *genus*, denoted as $g = h$. It is defined as the maximum number of closed curves we can draw on the surface such that cutting along them leaves the surface in a single connected piece. For example, for a sphere we have $g = 0$, and for a torus we have $g = 1$. If we now drop the assumption that the surface is connected, we get the Euler characteristic and the genus by taking the sum over all components. Since $\chi_i = 2 - 2g_i$ for the i-th component, we have

$$\chi = \sum_{i=1}^{k} \chi_i = \sum_{i=1}^{k} (2 - 2g_i) = 2k - 2g\,. \tag{A.8}$$

We see that a minimum amount of topological information is needed to translate between Euler characteristic and genus. This is different from what the cosmologists

have traditionally called the genus, which is defined as $\tilde{g} = -\frac{1}{2}\chi$. Relating the two notions, we get $g = k + \tilde{g}$. We will abandon both in this paper, $\tilde{g}$ because it is redundant, and g because it is limited to surfaces. Indeed, the Euler characteristic can also be defined for a 3-dimensional body, taking the alternating sum of the simplices used in a triangulation, while the genus has no satisfactory generalization beyond 2-dimensional surfaces.

The Gauss-Bonnet theorem (eq. A.4) and Crofton's formula (eqn. A.6) underline the key position of the Euler characteristic at the core of the topological and geometric characterization of manifolds. The Euler characteristic establishes profound and perhaps even surprising links between seemingly widely different areas of mathematics (Adler & Taylor 2010). While in simplicial topology Euler's polyhdron formula states that it is the alternating sum of the number of k-dimensional simplices of a simplicial complex (eq. A.7), its role in *algebraic topology* as the alternating sum of Betti numbers is expressed by the Euler-Poincaré formula (see eq. A.9 in the next subsection). Even more intricate is the connection that it establishes between these topological aspects and the singularity structure of a field, which is the realm of *differential topology*. In particular interesting is the relation established by Morse theory of the Euler characteristic being equal to the alternating sum of the number of different field singularities, ie. of maxima, minima and saddle points. Finally, its significance in *integral geometry* is elucidated via Crofton's formula (eq A.6), which establishes the fact that Minkowski functionals are integrals over the Euler characteristic.

A.3 Homology and Betti numbers

While the Euler characteristic can distinguish between connected, closed surfaces in $\mathbb{R}^3$, it has no discriminative power if applied to 3-manifolds, which is the most direct generalization of surfaces to the next higher dimension. Indeed, Poincaré duality implies $\chi = 0$ for all 3-manifolds. Fortunately, we can write the Euler characteristic as an alternating sum of more descriptive topological invariants named after Enrico Betti (Betti 1871). To introduce them, we find it convenient to generalize the space $\mathbb{M}$ by dropping most limitations, such as that it be embedded or even embeddable in $\mathbb{R}^3$. Letting the intrinsic dimension of $\mathbb{M}$ be d, we get $d+1$ possibly non-zero *Betti numbers*, which traditionally are denotes as $\beta_0, \beta_1, \ldots, \beta_d$. The relationship to the Euler characteristic is give by the Euler-Poincaré Formula:

$$\chi \;=\; \beta_0 - \beta_1 + \beta_2 - \ldots (-1)^d \beta_d. \tag{A.9}$$

This relation holds in great generality, requiring only a triangulation of the space, and even this limitation can sometimes be lifted. In this paper, we only consider subspaces of the 3-torus: $\mathbb{M} \subseteq \mathbb{X}$. For this case, only β_0, β_1, β_2, and β_3 are possibly non-zero, and we have $\beta_3 \neq 0$ only if $\mathbb{M} = \mathbb{X}$, in which case $\beta_3 = 1$. The first three Betti numbers have intuitive interpretations: β_0 is the number of *components*, β_1 is the number of *loops*, and β_2 is the number of *shells* in $\mathbb{M}$. Often, it is convenient to consider the complement of $\mathbb{M}$, which shows $\beta_0 - 1$ *gaps* between the components, β_1 *tunnels* going through the loops, and β_2 *voids* enclosed by the shells.

A formal definition of the Betti numbers requires the algebraic notion of a ho-

mology group. While a serious discussion of this topic is beyond the scope of this paper, we provide a simplified exposition and refer to texts in the algebraic topology literature for details (see e.g. Munkres 1984).

For simplicity, we assume a triangulated space and we use the coefficients 0 and 1 and addition, modulo 2. A *p-chain* is a formal sum of the p-simplices in the triangulation, which we may interpret as a subset of all p-simplices, namely those with coefficients 1. The *sum* of two p-chains is again a p-chain. Interpreted as sets, the sum is the symmetric difference of the two sets. Note that each p-simplex has $p + 1$ (p-1)-simplices as faces. The *boundary* of the *p-chain* is then the sum of the boundaries of all p-simplices in the chain. Equivalently, it is the set of (p-1)-simplices that belong to an odd number of p-simplices in the chain. We call the p-chain a *p-cycle* if it is the boundary of a (p+1)-chain. Importantly, every p-boundary is a p-cycle. The reason is simply that the boundaries of the (p-1)-simplices in the boundary of a p-simplex contain all (p-2)-simplices twice, meaning the boundary of the boundary is necessarily empty. To get homology, we still need to form *classes*, which we do by not distinguishing between two p-cycles that together form the boundary of a $(p + 1)$-chain.

To get the group structure, we add p-cycles by taking their symmetric difference or, equivalently, by adding simplices modulo 2. Homology classes can now be added simply by adding representative p-cycles and taking the class that contains the sum. The collection of classes together with this group structure is the *p-th homology group*, which is traditionally denoted as H_p. Finally, the *p-th Betti number* is the rank of this group, and since we use modulo 2 arithmetic to add, this rank is the binary logarithmic of the order: $\beta_p = \log_2 |H_p|$. We note that modulo 2 arithmetic has multiplicative inverses and therefore forms what in algebra is called a *field*. For example, arithmetic with integers is not a field. Whenever we use a field to construct homology groups, we get vector spaces. In particular, the groups H_p defined above are vector spaces, and the β_p are their dimensions, as defined in standard linear algebra.

In our study, we forward Betti numbers for the characterization of the topological aspects of the cosmic mass distribution. In this context, we should also appreciate the significance of the observation that Minkowski functionals may be written as Euler integrals, expressed by Hadwiger's formula (eqn. A.6). Wintraecken (2012) recently demonstrated that Betti numbers, as opposed to Minkowski functionals, cannot be expressed in terms of integrals over the Euler characteristic. The importance of this finding for our purpose is that Betti numbers contain topological information which is different and complementary to that contained in Minkowski functionals and genus, and that an analysis of their characteristics will shed new light on the connectivity of the different morphological elements of the cosmic web.

A.4 Morse theory

In Morse theory, we consider a compact manifold $\mathbb{X}^\nu$,and a generic smooth function on this manifold. In the context of this paper, the manifold is the 3-torus and the function is a density distribution, $\varrho : \mathbb{X} \to \mathbb{R}$. Assuming ϱ is smooth, we can take derivatives, and we call a point $x \in \mathbb{X}$ *critical* if all partial derivatives vanish. Corre-

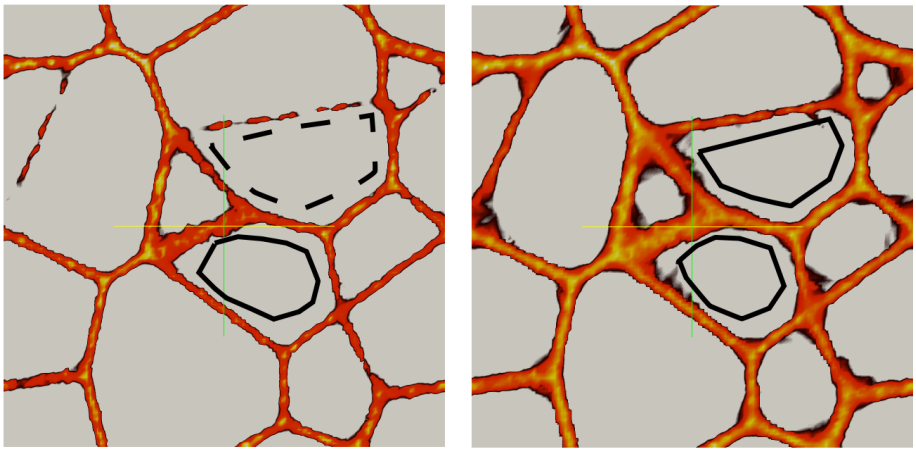

Figure A.1 Density rendering of the superlevel set of a 2-dimensional cross section of the voronoi wall models, illustrating chains and cycles. We focus our attention on the structures traced by the black lines. For high superlevel set values, in the left panel, the structure traced by the broken D-shape does not form a loop. The multiple broken segments are all *chains*. For lower superlevel set, the structures thicken and the individula segments merge together to form a loop, a 1-dimensional *cycle*.

spondingly, $\varrho(x)$ is a *critical value* of the function. All points of $\mathbb{X}$ that are not critical are *regular points*, and all values in $\mathbb{R}$ that are not the function value of critical points are *regular values*. Finally, we call ϱ *generic* if all critical points are non-degenerate in the sense that they have invertible Hessians. In this case, critical points are isolated from each other, and since $\mathbb{X}$ is compact, we have only finitely many critical points and therefore only finitely many critical values. The *index* of a non-degenerate critical point is the number of negative eigenvalues of the Hessian. Since $\mathbb{X}$ is 3-dimensional, we have 3-by-3 Hessians and therefore only four possibilities for the index. A *minimum* of ϱ has index 0, a *maximum* has index 3, and there are two types of *saddles*, with index 1 and 2.

The significance of the critical points and their indices becomes apparent when we look at the sequence of growing superlevel sets: $\mathbb{X}^\nu = \varrho^{-1}[\nu, \infty)$, for $0 \leq \nu < \infty$. If $\nu > \mu$ are regular values for which $[\mu, \nu]$ contains no critical value then $\mathbb{X}^\nu$ and $\mathbb{X}^\mu$ are topologically the same, the second obtained from the first by diffeomorphic thickening all around. If $[\mu, \nu]$ contains the critical value of exactly one critical point, x, then the difference between the two superlevel sets depends only on the index of x. If x has index 3 then $\mathbb{X}^\mu$ has one more component than $\mathbb{X}^\nu$, and that component is a topological ball. If x has index 2, then $\mathbb{X}^\mu$ can be obtained from $\mathbb{X}^\nu$ by attaching an arc at its two endpoints and thickening all around. This extra arc can have one of two effects on the homology of the superlevel set. If its endpoints belong to different components of $\mathbb{X}^\nu$, then $\mathbb{X}^\mu$ has one less component, while otherwise $\mathbb{X}^\mu$ has one

more loop. If x has index 1, then $\mathbb{X}^\mu$ can be obtained from $\mathbb{X}^\nu$ by attaching a disk, which has again one of two effects on the homology groups. Finally, if x has index 0 then $\mathbb{X}^\mu$ is obtained by attaching a ball. In all cases but one, this ball fills a void, the exception being the last ball that is attached when we pass the global minimum of ϱ. At this time, the superlevel set is completed to $\mathbb{X}^\mu = \mathbb{X}$.

A.5 Persistence homology

In Morse theory, we learned that each critical point either increases the rank of a homology group by one, or it decreases the rank of another group by one. Equivalently, it gives birth to a generator of one group or death to a generator of another group. Our goal is to pair up births with deaths such that we can talk about the subsequence in the filtration over which a homology class exists. This is precisely what persistent homology accomplishes. The hierarchical definition of topology that emerges due to taking the path of filtration is ideally suited to describe the topology of the mass distribution in the Universe, on account of it being hierarchical in nature as well.

To describe how this is achieved, we map each superlevel set in the filtration to the direct sum of its homology groups. With this construction, we capture the homology classes of all dimensions as once, and we simplify the notation by making it unnecessary to write the dimension in the subscript. Recall that between two consecutive critical values, the homology of the superlevel sets is constant. It therefore suffices to pick one regular value within each such interval. Writing $r_0 > r_1 > \ldots > r_n$ for these regular values and H_i for the direct sum of all homology groups of the superlevel set $\mathbb{X}^{r_i} = \varrho^{-1}[r_i, \infty)$,we get a sequence of homology groups:

$$H_0 \to H_1 \to \ldots \to H_n,$$

where $H_0 = 0$ and H_n is the homology group of $\mathbb{X}$. The arrows represent homomorphisms induced by the inclusion between superlevel sets. Assuming coefficients in a field, as before, we have a sequence of vector spaces with linear maps between them. These maps connect the groups by telling us where to find the cycles of a homology group within later homology groups. Sometimes, there are new cycles that cannot be found as images of incoming maps, and sometimes classes merge to form larger classes, which happens when we get chains that further wash out the difference between cycles.

We are now more specific about these connections. Letting γ be a class in H_i, we say γ is *born* at H_i and *dies entering* H_j if

- γ is not in the image of H_{i-1} in H_i;
- the image of γ is not in the image of H_{i-1} in H_{j-1}, but it is in the image of H_{i-1} in H_j.

Letting $r_{i-1} > \nu_i > r_i$ and $r_{j-1} > \nu_j > r_j$ be the critical values in the relevant intervals, we represent γ by (ν_i, ν_j), which we call a *birth-death pair*. Furthermore, we call $\operatorname{pers}(\gamma)$ the *persistence* of γ, but also of its birth-death pair.

To avoid any misunderstanding, we note that there is an entire coset of homology classes that are born and die together with γ, and all these classes are represented

by the same birth-death pair. Calling the image of H_i in H_{j-1} a *persistent homology group*, we note that its rank is equal to the number of birth-death pairs (ν_b, ν_d) that satisfy $\nu_b \geq \nu_i > \nu_j \geq \nu_d$. They represent the classes that are born at or before H_i and that die entering H_j or later.

B

Computation

The geometric and topological concepts outlined in Appendix A have all matured to a stage at which we have fast software to run on simulated and observed data. In this section, we describe the principles of these algorithms, and we provide sufficient information for the reader to understand the connection between the mathematics, the data, and the computed results.

The computational framework of our study involves three major components. One concerns the definition and calculation of the density field on which we apply the field's filtration. A directly related issue is the representation of the density field in the homology calculation, ie. whether we retain its representation by density estimates at the original sampling points or whether we evaluate it on the basis of a density image on a regular grid. The procedure for constructing the filtration is different in each case, and we detail them in Section B.1 and Section B.2. The algorithm used to implement the actual homology computation is the core of our study, and is same for both the filtration defined on particles or a filtration defined on image data. The third major aspect of our study concerns the representation of the results of the homology computation. The principal analytical tools of our study consist of *intensity maps* and *Betti numbers*, which form the visual representation and summary of the persistent homology of the analyzed data samples.

B.1 Density filtration from point samples

We use DTFE (Schaap & van de Weygaert 2001; van de Weygaert & Schaap 2009b; Cautun & van de Weygaert 2011) to construct a piecewise linear scalar-valued density field from a particle distribution. The DTFE formalism, whose details are outlined in appendix F.0.6, involves the computation of the Delaunay triangulation of

Model	# particles	# simplices	Tri. (s)	Pers. (s)
Poisson	500,000	14,532,164	10.15	6414.16
Cluster	262,144	7,491,308	81.48	12.58
Filament	262,144	7,346,712	77.76	402.36
Wall	262,144	7,345,520	5.26	555.46
Voronoi Kinematic Stage 3	262,144	7,409,364	5.93	125.33
Soneira-Peebles $\zeta = 9.0$	531,441	14,300,836	162.42	168.15

Table B.1 Parameters of computation for the various models described in this paper. All computations are performed on an Intel(R) Xeon(R) CPU @ 2.00GHz. Columns 1 & 2 present the models described in the later sections, and the number of particles used for the computation. Column 3 gives the total number of simplices of the Delaunay triangulation. Columns 4 & 5 give the time required to compute the triangulation and persistence respectively, in seconds.

the particles in $\mathbb{X}$, the determination of tessellation based density estimates and the subsequent piecewise linear interpolation of the density values at the Delaunay vertices, ie. the sample points, to the higher dimensional simplices, yielding a field $\varrho : \mathbb{X} \to \mathbb{R}$.

For the calculation of the Delaunay tessellation, we use software in the CGAL library. We use the 3-torus option of CGAL, which is the the periodic form of the original data set in a cubic box, which is accomplished by identifying and glueing opposite faces of the box.

Table B.1 presents the noteworthy parameters of computations for a single realization of the different models used in the results section of this paper. Naming the models in Column 1, we see the number of particles and simplices in the Delaunay triangulation in Columns 2 and 3 (also see Okabe et al. 2000; van de Weygaert 1994), and the number of seconds needed to compute the Delaunay triangulation and the persisten pairs in Columns 4 and 5. Apparently, the number of particles is not strongly correlated with the time it takes to construct the Delaunay triangulation. Indeed, the algorithm is also sensitive to other parameters – such as the number of simplices in the final triangulation or ever constructed and destroyed during the runtime of the algorithm – that depend on how the particles are distributed in space.

In a second step, we compute the DTFE density value for each vertex, u, of the Delaunay triangulation. The DTFE density value at the vertices is the inverse of the volume of its *star*. The star consists of all simplices that contain u as a vertex (see Figure B.1 for an illustration) , and we assign one over this volume as the density value to u. Finally, we use piece-wise linear interpolation to define $\varrho : \mathbb{X} \to \mathbb{R}$.

Evidently, we should ask ourselves in how far the results of the homology analysis are dependent on the density estimator used. In Section 2.9 we present a com-

parison between the results of our homology analysis obtained using the DTFE density estimate and that using the spatially adaptive SPH density estimates (see Appendix F.0.7), demonstrating that the results based on DTFE and on SPH are largely consistent.

B.2 Density filtration of fields on a regular grid

A first step towards computing persistence of a field ρ sampled on a regular cubical grid is the construction of a triangulation on the sample voxels. The components of a triangulation - vertices, edges, faces and tetrahedra - define a simplicial complex whose topological characteristics are equivalent to that of the sampled field. It is not possible to construct a unique triangulation K from a regular cubical grid of sample voxels. This is because a cubical grid suffers from degeneracies caused by corners common to eight and not four voxels, and the edges shared by four and not three faces.

Bendich et al. (2010) provide an algorithm for constructing triangulations of images represented on regular voxel grids. The algorithm solves the degeneracy arising from the regular grid by slightly perturbing the grid cells leading to a deformed grid where the corners are shared by four voxels, and the edges by three voxels. This transformation defines the elements of the dual triangulation uniquely – the vertices of this triangulation are defined by voxel centers, the edges are defined by the centers of the voxels which share a common face, the triangles by the centers of the voxels which share a common edge, and the tetrahedra by the centers of the voxels which share a common corner.

In a second step of the algorithm, the field values at the vertices in the triangulation are used to interpolate the values on the higher dimensional simplices, much akin to that used in the DTFE formalism developed by Schaap & van de Weygaert (2000b) (also see (Bernardeau & van de Weygaert 1996; van de Weygaert & Schaap 2009b)). This results in a continuous linearly extrapolated simplicial field - ie. a field defined on the edges, faces and tetrahedra of the resulting simplicial complex - that preserves the topology of the original density field (see (Pranav et al. 2013)). Of crucial importance is the fact that the choice of interpolation - linear, or constant - has no effect on topology. In this paper we we use a piece wise constant interpolation: $\rho(\sigma) = max[\rho(\tau)|\tau \subset \text{vertex of} \sigma]$.

B.3 Field sampling

To further process the density field towards its topological analysis, we may follow a range of field sampling strategies. The most suggestive option is to take the raw DTFE field, including all details of the discretely sampled field. In the context of our study, we mostly follow this *Raw DTFE sampling* strategy. The particular nature of the discretely sampled density field involves a complication. Because the number density of the sample points represents a measure of the value of the density field itself, the DTFE density field has a much higher spatial resolution in high density regions than in low density regions. This might be a source of a strong bias in the re-

trieved topological information, given that most of this will focus on the topological structure of the high-density regions.

To alleviate a density bias in the topological analysis, one may invoke a range of strategies. One option is to sample the density field on a regular grid. In other words, to create an image of the DTFE density field reconstruction. Details of the image construction are described in appendix F.0.6. It has the advantage of representing a uniformly sampled density field, with a uniform spatial resolution dictated by the voxel size of the image. The homology analysis of such a gridbased image involves a few extra complications, the details of which are most extensively discussed in the follow-up study analyzing the homology of Gaussian random fields (see Chapter 3). In appendix 2.9 we have compared the results of the homology analysis involving this *DTFE image sampling* strategy with those obtained following the *Raw DTFE sampling* strategy.

Another strategy to moderate the bias towards high-density regions is to use the singularity structure of the piecewise linear density field, and use the persistence of singularity pairs to remove insignificant topological features. This natural feature-based smoothing of the density field has been described extensively by Edelsbrunner et al. (2003), and has been applied in studies of cosmic structures in Chapter 4.

B.4 Topology through critical points and filtrations.

As mentioned in the paragraph on Morse theory, the superlevel set does not change topology as long as ν does not pass a critical value of the function, and this is also true for piecewise linear functions, except that we need to adjust the concept of critical point. Here we do the obvious, looking at how ϱ varies in the *link*, of a vertex. The link consists of all faces of simplices in the star, that do not themselves belong to the star (Edelsbrunner & Harer 2010, Chapter VI). Indeed, the topology can change only when ν passes the value of a vertex, so it suffices to consider only one (regular) value between any two contiguous vertex values. To describe this, we let n be the number of vertices in the triangulation, and we assume $\nu_i = \varrho(u_i) < \nu_{i+1} = \varrho(u_{i+1})$ for $1 \leq i < n$.[1] We thus consider superlevel sets at the regular values in the sequence

$$r_0 > \nu_1 > r_1 > \nu_2 > \ldots > \nu_n > r_n.$$

Constructing these superlevel sets and computing their homology individually would be impractical for the data-sets we study in this paper. Fortunately, there are short-cuts we can take that speed up the computations while having no effect on the computed results. The first short-cut is based on the observation that $\mathbb{X}^\nu$ has the same homotopy type as the subcomplex K^ν of the triangulation K of $\mathbb{X}$ that consists of all vertices with $\varrho(u_i) \geq \nu$ and all simplices connecting them. There is a convenient alternative description of K^ν. Define the *upper star* of a vertex u as the collection of simplices in the star for which u is the vertex with smallest density value (see fig. B.1 for the upper star of a regular vertex, a 1-saddle, a 2-saddle an a maximum). Then K^ν

[1] It is unlikely that the estimated density values at two vertices are the same, and if they are, we can pretend they are different, eg. by simulating a tiny perturbation that agrees with the ordering of the vertices by index; see eg. (Edelsbrunner 2001, Section I.4).

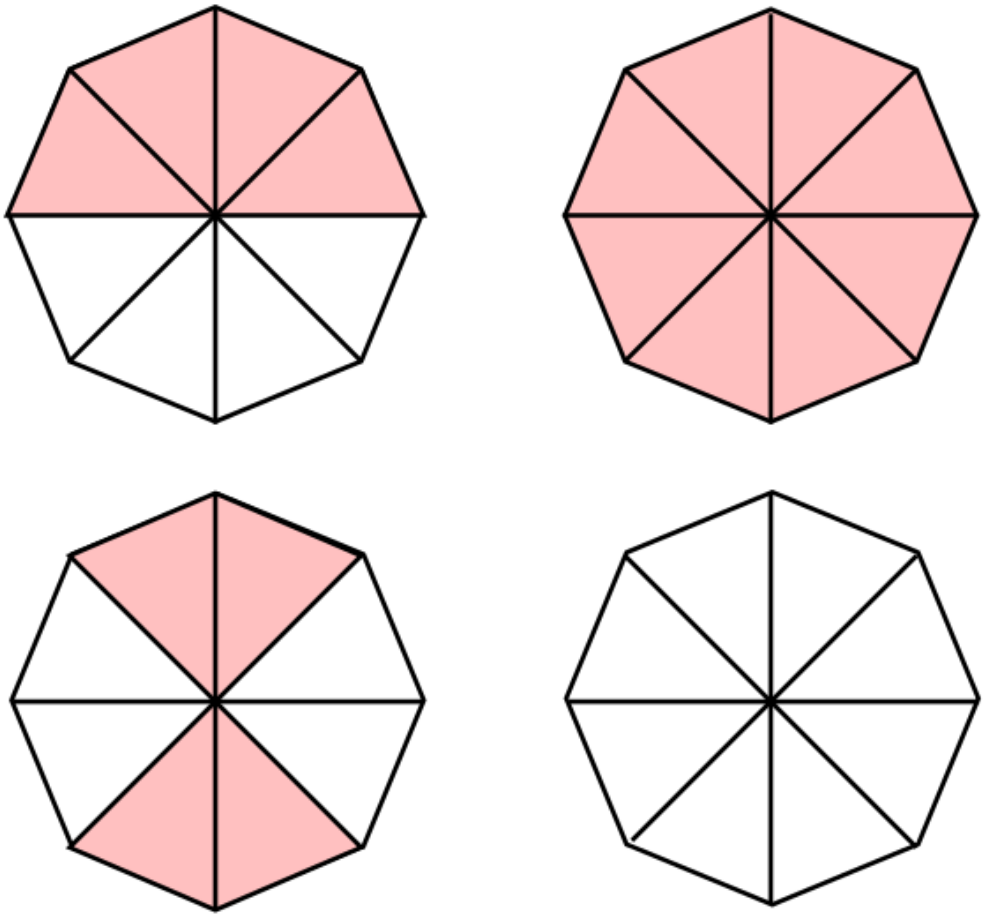

Figure B.1 Figure illustrating the upper star of a regular vertex, minimum, saddle and maximum, respectively in top-left, top-right, bottom-left, bottom-right panels. The star of a vertex is consists of all the simplices incident to it. The shaded simplices in pink have a function value higher than the vertex.

is the union of the upper stars of all vertices with $\varrho(u_i) \geq \nu$. This description is computational convenient because it tells us that $K^{r_{i+1}}$ can be obtained from K^{r_i} simply by adding the simplices in the upper star of u_{i+1}. We say the superlevel sets can be computed *incrementally*, and we will be careful to follow this paradigm in every step of our computational pipeline. This incremental construction of the superlevel sets is equivalent to constructing the upper-star filtration, which is an essential pre-cursor to computing persistence homology which we describe next.

B.5 Persistence homology

Next, we sketch the algorithm that computes the persistent homology of the sequence of superlevel sets. We begin with a linear ordering of the simplices in K that contains all K^v as prefixes. To describe it, let $u_i = \sigma_{j_i}, \sigma_{j_i+1}, \ldots, \sigma_{j_{i+1}-1}$ be the simplices in the upper star of u_i, sorted in increasing order of dimension. Setting $j_1 = 1$ and $m = j_{n+1} - 1$, this linear ordering of the simplices is $\sigma_1, \sigma_2, \ldots, \sigma_m$. It has the property that each simplex is preceded by its faces, which implies that every prefix, $K_j = \{\sigma_1, \sigma_2, \ldots, \sigma_j\}$, is a simplicial complex. We require this property so that every step of our incremental algorithm is well defined. It should be clear that $K^{v_i} = K_j$ for $j = j_{i+1} - 1$.

Algorithm 2 MATRIX REDUCTION

```
1: R = Δ
2: for j = 1 to m do
3:    while there exists j_0 < j with low(j_0) = low(j) do
4:       add colum j_0 to column j
5:    end while
6: end for
```

B.5.1 Boundary matrix and its reduction

The persistence algorithm is easiest to describe as a matrix reduction algorithm, with the input matrix being the ordered boundary matrix of K.[2] Specifically, this is the m-by-m matrix Δ whose rows and columns correspond to the simplices in the mentioned linear ordering. Specifically, the j-th column records the boundary of σ_j, namely $\Delta_{i,j} = 1$, if σ_i is a face of σ_j and the dimension of σ_i is one less than that of σ_j, and $\Delta_{i,j} = 0$, otherwise. Symmetrically, the i-th row records the star of σ_i. The persistence algorithm transforms Δ into *reduced form*, in which every row contains the lowest non-zero entry of at most one column. Making sure that we do not permute rows, and we add columns strictly from left to right, the lowest non-zero entries in the reduced matrix correspond to the birth-death pairs of the density field – precisely the information we are after. To describe the transformation, we write $\text{low}(j) = i$ if i is the maximum row index of a non-zero entry in column j, and we set $\text{low}(j) = 0$ if the entire column is 0. Algorithm 2 presents the algorithm for such a reduction. Section 2.3.3 illustrates these concepts and steps through an example.

The search for the fastest algorithm to reduce an ordered boundary matrix is an interesting question of active research in the field of computational topology. Most known algorithms use row and column operations, like in Gaussian elimination, which takes time proportional to m^3 in the worst case. A fortunate but largely not understood phenomenon is the empirical observation that some of these algorithms

[2] We hasten to mention that storing this matrix explicitly is too costly for our purposes. Instead, we use the triangulation as a sparse matrix representation, and we implement all steps of the matrix reduction algorithm accordingly. However, for the purpose of explaining the algorithm, we maintain the illusion of an explicit representation of the matrix.

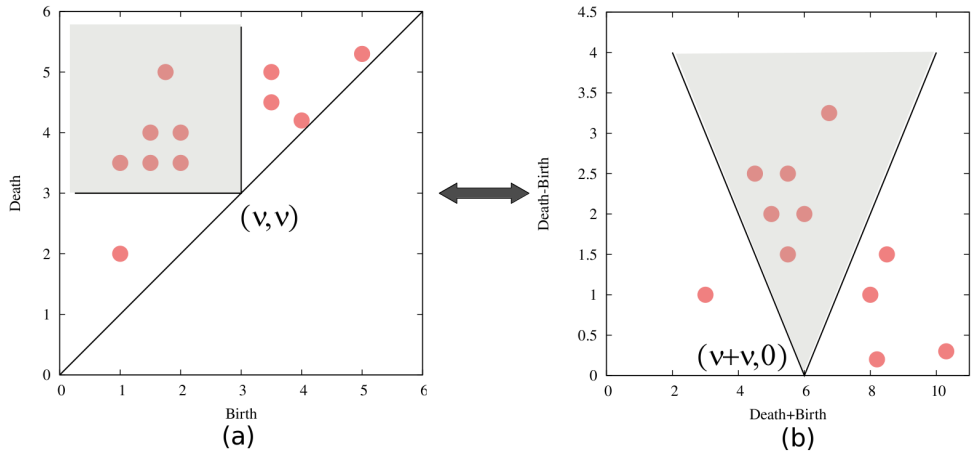

Figure B.2 Figure illustrating the transition from the *birth-death* to the *mean age-persistence* plane. If the coordinates of a point in panel (a) are (b,d), the coordinates in panel (b) are (d+b,d-b). The Betti numbers can be read off from the persistence diagrams. The contribution to the Betti numbers for a level set ν comes from all the persistent dots that are born before ν and die after ν – in other words, the shaded region in panel (a) anchored at (ν, ν). The shaded region transforms in panel (b) to a V-shaped region anchored at $(\nu + \nu, 0)$. The arms of the V have slope -1 and 1 respectively.

are significantly faster than cubic time for most practical input data. This is lucky but also necessary since we could otherwise not compute the results we present in this paper. The time to compute the persistence pairs for different models is displayed in Column 5 of Table B.1.

B.6 Persistence Diagrams

Given the reduced boundary matrix, we generate the birth-death pairs of ϱ from the lowest non-zero entries in the columns. Specifically, for every non-zero $i' = \mathrm{low}(j')$, the addition of $\sigma_{i'}$ gives birth to a homology class that dies when we add $\sigma_{j'}$. If $\sigma_{i'}$ is in the upper star of u_i, and $\sigma_{j'}$ is in the upper star of u_j, then we get (ν_i, ν_j) as the corresponding birth-death pair. It is quite possible that $i = j$, namely if both simplices belong to the same upper star, in which case we talk of a *still-birth*. We draw this birth-death pair as the point (ν_i, ν_j) in the *birth-death* plane. Alternatively, we can also draw them as $(\nu_i + \nu_j, \nu_j - \nu_i)$ in the plane. This amounts to a scaling by a factor of $\sqrt{2}$ and a rotation of coordinates by 45 degrees clock-wise. This is our preferred representation of the persistence diagrams throughout this paper. An illustration of the transformation is depicted in Figure B.2. Drawing all points representing p-dimensional homology classes gives the *p-th persistence diagram* of ϱ, which

we denote as $\mathrm{Dgm}_p(\varrho)$. Recall that the second coordinate is the persistence, and because a still-birth has zero persistence, it is drawn right on the horizontal axis. The persistence is a measure of significance of the feature represented by a birth-death point, and still-births are artifacts of the representation of ϱ and have indeed no significance. The first coordinate is the sum of birth- and death-values, and we refer to half that coordinate as the *mean age*. It gives information about the range of density values the corresponding feature is visible.[3]

Persistence diagrams contain more information than the Betti numbers. Indeed, we can read the p-th Betti number of the superlevel set for ν as the number of points of $\mathrm{Dgm}_p(\varrho)$. The contribution to the Betti numbers for a level set ν comes from all the dots in the persistence diagram corresponding to cycles that are born before ν and die after ν – in other words, the shaded region in panel (a) anchored at (ν, ν) in Figure B.2. The shaded region transforms appropriately in panel (b) to a V-shaped region anchored at $(\nu, 0)$ on the horizontal axis. The arms of the V have slope -2 and 2 respectively. Another useful property is the stability of the diagram under small perturbations of the input. Specifically, the diagram of a density function, ϱ', that differs from ϱ by at most ε at every point of the space, has bottleneck distance at most ε from $\mathrm{Dgm}_p(\varrho)$; see (Cohen-Steiner et al. 2007). This implies that every points of $\mathrm{Dgm}_p(\varrho')$ is at distance at most ε from a point in $\mathrm{Dgm}_p(\varrho)$ or from the horizontal axis.

[3] Almost every homology class that is ever born will also die at finite time, but there are eight exceptions, namely the classes that describe the 3-torus itself. They are not relevant for the study in this paper, and we do not draw them in the diagrams.

C

Stochastic Random Fields

The cosmic density perturbation field is a realization of a stochastic random field. A *random field*, f, on a spatial volume assigns a value, $f(\mathbf{x})$, to each location, $\mathbf{x}$, of that volume. The fields of interest are smooth and continuous [1]. The stochastic properties of a random field are defined by its *N-point joint probabilities*, where N can be any arbitrary positive integer. To denote them, we write $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N)$ for a vector of N points and $\mathbf{f} = (f_1, f_2, \ldots, f_N)$ for a vector of N field values. The joint probability is

$$\text{Prob}[f(\mathbf{x}_1) = f_1, \ldots, f(\mathbf{x}_N) = f_N] \quad = \quad \mathcal{P}_{\mathbf{X}}(\mathbf{f})\,\mathrm{d}\mathbf{f}\,, \tag{C.1}$$

which is the probability that the field f at the locations $\mathbf{x}_i$ has values in the range f_i to $f_i + \mathrm{d}f_i$, for each $1 \leq i \leq N$.

In cosmological circumstances, we use the *statistical cosmological principle*, which states that statistical properties of e.g. the cosmic density distribution in the Universe are uniform throughout the Universe. It means that the distribution functions and moments of fields are the same in each direction and at each location. The latter implies that ensemble averages depend only on one parameter, namely the distance between the points.

Important for the cosmological reality is the validity of the *ergodic principle*. The Universe is unique, and its density distribution is the only realization we have of the underlying probability distribution. The ergodic principle allows us to measure the value of ensemble averages on the basis of spatial averages. These will be equal to the expectations over an ensemble of Universes, something which is of key significance for the ability to test theoretical predictions for stochastic processes like the cosmic mass distribution with observational reality.

[1] 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 $W(\mathbf{x}, \mathbf{y})$, $f_s(\mathbf{x}) = \int \mathrm{d}\mathbf{y} f(\mathbf{y}) W(\mathbf{x}, \mathbf{y})$.

C.1 Gaussian Random Fields

The primordial density field is, to high accuracy, a Gaussian random field. A Gaussian random field is an example of a stochastic process. Given a parameter space T, a stochastic process f over T is a collection of random variables

$$\{f(t) : t \in T\}\,. \tag{C.2}$$

f is a (N, d) random field if T is a set of dimension N and $f(t)$ are vector-valued with dimension d. In the scope of this paper, we are interested in scalar random fields defined in $\mathbb{R}^3$. In this case, $N = 3$ and $d = 0$. Finally, f is a *gaussian* random field if $f(t)$ are gaussian distributed.

For a *Gaussian random field*, the joint probabilities for $N = 1$ and $N = 2$ determine all others (see Adler 1981; Adler & Taylor 2010; Bardeen et al. 1986). Specifically, the probability density functions take the simple form

$$\mathcal{P}_{\mathbf{X}}(\mathbf{f}) \;=\; C \cdot \exp\left[-\mathbf{f}\mathsf{M}^{-1}\mathbf{f}^T/2\right]\,, \tag{C.3}$$

where $C = 1/[(2\pi)^N(\det \mathsf{M})]^{1/2}$ normalizes the expression, making sure that the integral of $\mathcal{P}_{\mathbf{X}}(\mathbf{f})$, over all $\mathbf{f} \in \mathbb{R}^N$, is equal to 1. Here, we assume that each 1-point distribution is Gaussian with zero mean. The matrix M^{-1} is the inverse of the $N \times N$ covariance matrix with entries

$$M_{ij} \;=\; \langle f(\mathbf{x}_i) f(\mathbf{x}_j)\rangle\,, \tag{C.4}$$

in which the angle bracket denotes the ensemble average of the product, over the 2-point probability density function. In effect, M is the generalization of the variance of a 1-point normal distribution, and we indeed have $\mathsf{M} = [\sigma_0^2]$ for the case $N = 1$.

Equation (C.3) shows that a Gaussian random is fully specified by the autocorrelation function, $\xi(r)$, which expresses the correlation between the density values at two points separated by a distance $r = |\mathbf{r}|$,

$$\xi(r) \;=\; \xi(|\mathbf{r}|) \;\equiv\; \langle f(\mathbf{x}) f(\mathbf{x}+\mathbf{r})\rangle\,. \tag{C.5}$$

In other words, the entries in the matrix are the values of the *autocorrelation function* for the distance between the points: $M_{ij} = \xi(r_{ij})$, with $r_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|$. For a multi-scale Gaussian density field, its structure can be characterized in a more transparent fashion in terms of the Fourier transform of the correlation function, the power spectrum $P(k)$. It specifies the mean square of the fluctuations of the Fourier components $\hat{f}(\mathbf{k})$ of the field $f(\mathbf{x})$,

$$f(\mathbf{x}) = \int \frac{\mathrm{d}\mathbf{k}}{(2\pi)^3}\, \hat{f}(\mathbf{k})\, e^{-i\mathbf{k}\cdot\mathbf{x}}\,, \tag{C.6}$$

via the relation

$$\langle \hat{f}(\mathbf{k})\hat{f}(\mathbf{k}')\rangle \;=\; (2\pi)^{3/2}\, P(k)\, \delta_D(\mathbf{k}-\mathbf{k}')\,, \tag{C.7}$$

where $\delta_D(\mathbf{k})$ is the Dirac delta function and the

C.1.1 Number distribution of peaks in Gaussian random fields

Bardeen et al. (1986) derive the marginal number distribution of peaks for Gaussian random fields, as a function of the dimensionless density threshold ν

$$\mathbb{N}_{pk}(\nu)\, d\nu = \frac{1}{(2\pi)^2 R_\star^3} e^{-\nu^2} G(\gamma, \gamma\nu). \tag{C.8}$$

Here, the function $G(\gamma, \gamma\nu)$ is a fitting function given by

$$G(\gamma, w) = \frac{w^3 - 3\gamma^2 w + \left[B(\gamma)w^2 + C_1(\gamma)\right] \exp\left[-A(\gamma)w^2\right]}{1 + C_2(\gamma) \exp\left[-C_3(\gamma)w\right]}. \tag{C.9}$$

The various coefficients are given by

$$\begin{aligned} A &= \frac{5/2}{9 - 5\gamma^2} \\ B &= \frac{432}{(10\pi^2)(9 - 5\gamma^2)^{5/2}} \\ C1 &= 1.84 + 1.13(1 - \gamma^2)^{5.72} \\ C2 &= 8.91 + 1.27 \exp(6.51\gamma^2) \\ C3 &= 2.58 \exp(1.05\gamma^2). \end{aligned} \tag{C.10}$$

For the power law power spectrum, the parameters γ and $R_\star$ are related to the various moments of the power spectrum, value of the spectral index n and the co-moving filtering radius R_f as

$$\begin{aligned} \gamma &= \frac{\langle k^2 \rangle}{\langle k^4 \rangle^{1/2}} \\ &= \frac{\sigma_1^2}{\sigma_2 \sigma_0} \\ &= \frac{n+3}{n+5} \\ R_\star &= \sqrt{3}\frac{\sigma_1}{\sigma_2} \\ &= \left(\frac{6}{n+5}\right)^{1/2} R_f. \end{aligned} \tag{C.11}$$

C.2 Gaussian random fields: Minkowski functionals, Euler characteristic, and genus

The goal of this sub-section is to present a brief account of the known topological characteristics of Gaussian random fields in terms of the existing decriptors in use, namely Minkowski Functionals, Euler Characteristic and genus.

C.2.1 Euler characteristic and genus

The Gauss-Bonnet Theorem relates the total Gaussian curvature κ of a connected 2-dimensional manifold surface to either their genus g or the Euler characteristic χ, suggesting that the knowledge of either κ, g or χ is sufficient to compute the others:

$$C = 4\pi(1-g) = 2\pi\chi. \tag{C.12}$$

Gott et al. (1986) derive a closed form analytical expression for the expected value of κ to arrive at expected analytic equation of g in case of Gaussian random fields

$$g(\nu) \;=\; -\frac{1}{8\pi^2}\left(\frac{\langle k^2\rangle}{3}\right)^{3/2}(1-\nu^2)e^{-\nu^2/2}, \tag{C.13}$$

where $\nu = \delta/\sigma$. Here, δ is the over- or under-density at a spatial location and σ is the rms of the density fluctuation field.
Often in cosmological literature (Doroshkevich 1970; Bardeen et al. 1986), the distinction between Euler characteristic and genus is blurred. Indeed, the expression of genus used in cosmology (Gott et al. 1986; Hamilton et al. 1986) is the same as the standard expression of Euler characteristic (Adler 1981).

C.2.2 Minkowski Functionals

There are $(d+1)$ Minkowski functionals defined for a d-dimensional manifold(Mecke et al. 1994; Schmalzing & Buchert 1997; Schmalzing et al. 1999; Sahni et al. 1998). Predominantly geometric in nature, the first four Minkowski functionals ($d \leq 3$) are respectively the volume, surface area, integrated mean curvature and the Gaussian curvature. For Gaussian random fields, the first four Minkowski functionals have known analytical epressions as a function of the superlevel sets of rms density threshold ν (Tomita 1993; Schmalzing & Buchert 1997)

$$\begin{aligned}
\mathrm{Q}_0(\nu) &= \frac{1}{2} - \frac{1}{2}\Phi\left(\frac{1}{\sqrt{2}}\nu\right), \\
\mathrm{Q}_1(\nu) &= \frac{2}{3}\frac{\lambda}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\nu^2\right), \\
\mathrm{Q}_2(\nu) &= \frac{2}{3}\frac{\lambda^2}{\sqrt{2\pi}}\nu\exp\left(-\frac{1}{2}\nu^2\right), \\
\mathrm{Q}_3(\nu) &= \frac{\lambda^3}{\sqrt{2\pi}}(\nu^2-1)\exp\left(-\frac{1}{2}\nu^2\right).
\end{aligned} \tag{C.14}$$

where $\lambda = \sqrt{|\xi''(0)|/[2\pi\xi(0)]}$, and the function $\Phi(x) = \int_0^x \mathrm{d}te^{(-t^2)}$ is the standard error function.

The functional $Qspace_3(\nu)$ in (C.14) and the equation for genus in (C.13) differ only by a factor. This is not surprising, as the fourth Minkowski functional can also be formulated as an integral of the Gaussian curvature, which is related to Euler characteristic and genus via the Gauss-Bonneť Theorem. In fact, Crofton's intersection formula (Crofton 1868) goes further and establishes that all the Minkowski functionals can be formulated as appropriate integrals of Euler characteristic (Pranav et al. 2013). Evidently, Euler characteristic is the more fundamental quantity, and is the bridge between geometry and topology (Adler & Taylor 2009).

D

Voronoi Clustering Models

Voronoi clustering models are heuristic models for cellular spatial patterns which use the geometric (and convex) structure of the Voronoi tessellation (Voronoi 1908; Okabe et al. 2000) to emulate the cosmic matter distribution van de Weygaert & Icke (1989); van de Weygaert (1991, 2002). They offer flexible templates for cellular patterns and are easy to tune towards a specific spatial cellular morphology. This makes them very suitable for studying clustering properties of nontrivial geometric spatial patterns. Unless otherwise specified, the seeds of the tessellation usually involve a set of Poisson distributed points.

The Voronoi models use Voronoi tessellations for defining the structural frame around which matter assembles as cosmic structure emerges and grows. Particles are distributed within this skeleton by assigning them to one of the four distinct structural components of a Voronoi tessellation. The interior of Voronoi cells is identified with *void regions*, the Voronoi cell faces with *walls*, the edges with *filaments* and the vertices with *cluster nodes*. What is usually described as a flattened "supercluster" consists of an assembly of various connecting walls in the Voronoi foam, while elongated "superclusters" of "filaments" usually include a few coupled edges. Vertices are the most outstanding structural elements, corresponding to the very dense compact nodes within the cosmic web where one finds the rich clusters of galaxies.

Among a variety of possible Voronoi clustering realizations, two distinct yet complementary classes of models are the most frequently used ones, the structurally rigid *Voronoi Element Models* and the evolving *Voronoi evolution models*. Both the Voronoi Element Models and the Voronoi Evolution Models are obtained by projecting an initially random distribution of N sample points/galaxies onto the walls, edges or vertices of the Voronoi tessellation defined by M nuclei.

D.0.3 Voronoi element models

"Voronoi element models" are fully heuristic models. They are user-specified spatial galaxy distributions within the *cells* (field), *walls*, *edges* and *vertices* of a Voronoi tessellation. *Pure Voronoi element Models* place their model galaxies exclusively in either walls, edges or vertices.

The practical implementation of the Voronoi model consists of an initial random distribution of N particles in a box of volume V with periodic boundary conditions (ie. a 3-torus). The initial spatial distribution of these N galaxies within the sample volume V is purely random, their initial locations $\mathbf{x}_{n0}$ $(n = 1, \ldots, N)$ defined by a homogeneous Poisson process. A random distribution of M nuclei defines the geometric structure of the Voronoi tessellation.

The initially randomly distributed model particles are projected onto the relevant Voronoi wall, Voronoi edge or Voronoi vertex or retained within the interior of the Voronoi cell in which they are located, according to a process which is the asymptotic limit of the prescription outlined in equation D.2 (see van de Weygaert et al. 2011). The model walls, filaments and nodes are not infinitely thin, but have a Gaussian density profile with a user-specified width. Following their projection onto wall, filament or vertex, the particles are randomly displaced according to this profile.

The versatility of the Voronoi element model also allows combinations in which field (cell), wall, filament and vertex distributions are superimposed. These complete composite particle distributions, *Mixed Voronoi element Models*, include particles located in four distinct structural components. The characteristics of the patterns and spatial distribution in the composite Voronoi Element models can be varied and tuned according to the fractions of galaxies in in Voronoi walls, in Voronoi edges, in Voronoi vertices and in the field. These fractions are free parameters to be specified by the user.

D.0.4 Voronoi evolution models

The second class of Voronoi models is that of the *Voronoi Evolution models*. They attempt to provide weblike galaxy distributions that reflect the outcome of realistic cosmic structure formation scenarios. They are based upon the notion that voids play a key organizational role in the development of structure and makes the Universe resemble a soapsud of expanding bubbles Icke (1984). While the galaxies move away from the void centres, and stream out of the voids towards the sheets, filaments and clusters in the Voronoi network the fraction of galaxies in the voids (cell interior), the sheets (cell walls), filaments (wall edges) and clusters (vertices) is continuously changing and evolving. The details of the model realization depends on the time evolution specified by the particular Voronoi Evolution Model.

Within the class of Voronoi Evolution Models the most representative and most frequently used are the *Voronoi kinematic models*. They form the idealized and asymptotic description of the outcome of hierarchical gravitational structure formation process, with single-sized voids forming around depressions in the primordial density field. This is translated into a scheme for the displacement of initially randomly distributed galaxies within the Voronoi skeleton. Within a void, the mean distance between galaxies increases uniformly in the course of time. When a galaxy tries to

enter an adjacent cell, the velocity component perpendicular to the cell wall disappears. Thereafter, the galaxy continues to move within the wall, until it tries to enter the next cell; it then loses its velocity component towards that cell, so that the galaxy continues along a filament. Finally, it comes to rest in a node, as soon as it tries to enter a fourth neighbouring void.

As for the Voronoi element models, the Voronoi kinematic models are based on the displacement of a sample of N particles, or model galaxies, with an initial random distribution. In practice, these are distributed in a box of volume V with periodic boundary conditions (ie. a 3-torus). The initial spatial distribution of these N galaxies within the sample volume V is purely random, their initial locations $\mathbf{x}_{n0}$ $(n = 1, \ldots, N)$ defined by a homogeneous Poisson process. A set of M nuclei or *expansion centres* within the volume V corresponds to the cell centres, or "expansion centres" driving the evolving matter distribution. The nuclei have locations $\mathbf{y}_m$ $(m = 1, \ldots, M)$. The first step of the formalism is to determine for each galaxy n the Voronoi cell $\mathcal{V}_\alpha$ in which it is initially located.

The path $\mathbf{x}_n$ along which a galaxy moves within the Voronoi skeleton depends on how far it has moved away from its initial location $\mathbf{x}_{n0}$. For a specific galaxy n this path may consist of the following sequence,

- cell displacement $\mathbf{s}_\alpha$ radially directed away from the expansion centre j_α
- wall displacement $\mathbf{s}_{\alpha\beta}$ within the Voronoi wall $\Sigma_{\alpha\beta}$, $\Sigma_{\alpha\beta}$ defined by the nucleus j_α and its natural neighbour j_β
- edge displacement $\mathbf{s}_{\alpha\beta\gamma}$ along the Voronoi edge $\Lambda_{\alpha\beta\gamma}$, $\Lambda_{\alpha\beta\gamma}$ defined by j_α and its natural neighbours j_β and j_γ .

This path is encapsulated in the equation,

$$\begin{aligned} \mathbf{x}_n &= \mathbf{y}_\alpha + \mathbf{s}_{n\alpha} + \mathbf{s}_{n\alpha\beta} + \mathbf{s}_{n\alpha\beta\gamma} \\ &= \mathbf{y}_\alpha + s_{n\alpha}\hat{\mathbf{e}}_{n\alpha} + s_{n\alpha\beta}\hat{\mathbf{e}}_{n\alpha\beta} + s_{n\alpha\beta\gamma}\hat{\mathbf{e}}_{n\alpha\beta\gamma}\,, \end{aligned} \tag{D.1}$$

where the unity vectors $\hat{\mathbf{e}}_{n\alpha}$, $\hat{\mathbf{e}}_{n\alpha\beta}$ and $\hat{\mathbf{e}}_{n\alpha\beta\gamma}$ specify the direction of galaxy's path within a Voronoi cell, a Voronoi wall or Voronoi edge.

While the galaxy's location in the Voronoi element models (see section D.0.3) is restricted to one of the components of the Voronoi skeleton, the Voronoi kinematic model is characterized by an evolving global "void" expansion factor $R(t)$ which dictates whether a galaxy has entered a wall, proceeded towards a filament or has arrived within a vertex/cluster node. At any one cosmic epoch each galaxy n is displaced according to $R(t)$. At first, while still within the cell's interior, the galaxy proceeds radially from its expansion centre j_α, moving along the radial cell path emanating from its nucleus,

$$\begin{aligned} s_{n\alpha}(t) &= \mathbf{x}_n(t) - \mathbf{y}_\alpha \\ &= R(t)\,|\mathbf{x}_{n0} - \mathbf{y}_\alpha|\,\hat{\mathbf{e}}_{n\alpha}\,. \end{aligned} \tag{D.2}$$

During this stage the wall and edge path factors $s_{n\alpha\beta} = s_{n\alpha\beta\gamma} = 0$, while the mean distance between the particles increases uniformly with expansion factor $R(t)$. Once the galaxy tries to enter an adjacent cell j_β and reaches a Voronoi wall, i.e. when

$$R(t)\,|\mathbf{x}_{n0} - \mathbf{y}_\alpha| \;>\; v_n \tag{D.3}$$

the gravity of the wall, aided and abetted by dissipational processes, will slow down its motion. Subsequently, the velocity component perpendicular to the cell wall disappears and the galaxy continues to move within the wall along the direction of $\hat{\mathbf{e}}_{n\alpha\beta}$. This continues until it reaches the corresponding Voronoi edge, after which it moves along the direction $\hat{\mathbf{e}}_{n\alpha\beta\gamma}$. Finally, it will reach the Voronoi vertex towards its move, which represents the cluster node where it will culminate its cosmic journey. As in the case of the Voronoi element models, the walls, filaments and nodes in the Voronoi kinematic models have a Gaussian density profile, with a user-specified width. To effect this profile, following their arrival in wall, filament or vertex node, the particles are randomly displaced according to the specified profile. For further details of the complete formalism for generating these spatial distributions can be found in van de Weygaert et al. (2011).

The resulting evolutionary progression within the Voronoi kinematic scheme is that of an almost featureless random distribution, via a wall-like and filamentary morphology towards a distribution in which matter ultimately aggregates into conspicuous compact cluster-like clumps. Figure 2.16 provides an impression of the evolutionary progression.

E

Soneira-Peebles model.

The Soneira-Peebles model is an analytic self-similar spatial point distribution which was defined for the purpose of modelling the galaxy distribution, such that its statistical properties would be tuned to reality (Soneira & Peebles 1978). An important property of the Soneira-Peebles model is that it is one of the few nonlinear models of the galaxy distribution whose statistical properties can be fully and analytically evaluated. This concerns its power-law two-point correlation function, correlation dimension and its Hausdorff dimension. Here we shortly specify the main characteristics of the Soneira-Peebles model, for an extensive description of the fractal-like properties of the Soneira-Peebles model we refer to (Martinez 1990) and (van de Weygaert & Schaap 2009b).

The Soneira-Peebles model is specified by three parameters (for an illustrated descriptions see van de Weygaert & Schaap (2009b)). The starting point of the model is a level-0 sphere of radius R. At each level-m a number of ψ subspheres are placed randomly within their parent level-m sphere: the level-$(m+1)$ spheres have a radius R/ζ where $\zeta > 1$, the size ratio between parent sphere and subsphere. This process is repeated for L successive levels, yielding ψ^L level-L spheres of radius R/ζ^L. At the center of each of these spheres a point is placed, producing a point sample of ζ^L points. While this produces a pure *singular* Soneira-Peebles model, usually a set of these is superimposed to produce a somewhat more realistically looking model of the galaxy distribution, an *extended* Soneira-Peebles model.

E.0.5 Self-similarity

The Soneira-Peebles model involves a hierarchy of structures of varying densities and characteristic scales, with the higher level spheres corresponding to high density structures of small scale and the lower level spheres corresponding to low density

structures of larger scale. As each sphere is constructed in the same way, the resulting point distribution is self-similar and forms a bound fractal. The fractal geometry of a point set is often characterized by the fractal dimension D, which is defined as

$$D = \lim_{r \to 0} \frac{\log N(r)}{\log(1/r)} \tag{E.1}$$

Here $N(r)$ is the number of non-empty cells in a partition of constant cell size r. If the Soneira-Peebles model would contain an infinite amount of levels, the resulting point distribution would have fractal dimension

$$D = (\log \psi)/(\log \zeta)\,. \tag{E.2}$$

One important manifestation of the self-similarity of the defined Soneira-Peebles distribution is reflected in the power-law two-point correlation function. For M dimensions it is given by

$$\begin{aligned} \xi(r) &\sim r^{-\gamma}\,, \\ \gamma &= M - \left(\frac{\log \psi}{\log \zeta}\right) \quad \text{for} \quad \frac{R}{\zeta^{L-1}} < r < R\,. \end{aligned} \tag{E.3}$$

The parameters ψ and ζ may be adjusted such that they yield the desired value for the correlation slope γ.

F

Density Estimators

In the cosmological reality, a density field is discretely sampled by galaxies, or by particles in computer simulations. This discrete distribution is subsequently translated into a continuous density field, based on the key assumption that the discrete point distribution represents a fair discrete sampling of the underlying density field. There is a large array of techniques available for obtaining density estimates from a discrete particle distribution. Key references on the corresponding problems and solutions include those by Ripley (1981). An extensive and systematic survey of available mathematical methods in an astronomical context can be found in a set of publications by Lombardi & Schneider (2001); Lombardi (2002); Lombardi & Schneider (2003).

In this appendix we focus on two formalisms. This study is mainly based on the use of the Delaunay Tessellation Field Estimator (DTFE). In addition, we have compared the obtained topological results with that of a spatially adaptive kernel estimator used in Smooth Particle Hydrodynamics.

F.0.6 DTFE, the Delaunay Tessellation Field Estimator

DTFE - the Delaunay Tessellation Field Estimator - exploits three properties of Voronoi and Delaunay tessellations. The first is the sensitivity of the tessellation cell size to the local point density. The DTFE method uses this fact to define a local estimate of the density, on the basis of the inverse of the volume of the tessellation cells. Equally important is their sensitivity to the local shape of the point distribution, which allows them to trace anisotropic features. Finally, it uses the adaptive and minimum triangulation properties of Delaunay tessellations to use them as adaptive spatial interpolation intervals for irregular point distributions.

In the first step of the DTFE procedure, a local density estimate ϱ_i is computed at each point i of the point sample. The value is the (normalized) inverse of the volume $V(\mathcal{W}_i)$ of the corresponding *contiguous Voronoi cell* or *star*, $\mathcal{W}_i$. The star consists of all Delaunay simplices that contain $\vec{x}_i$ as a vertex. In D dimensions the DTFE density estimate at each sample point i is given by:

$$\varrho_i \;=\; \frac{(1+D)}{V(\mathcal{W}_i)}\,. \tag{F.1}$$

DTFE subsequently interpolates the values of these density field estimates $f_i = \varrho_i$ over the volume of the sample. This produces a piecewise linear interpolation of the field, in which the Delaunay tessellation is used as interpolation grid. The field gradient ∇f is defined to be constant over the volume of each Delaunay tetrahedron, with its value directly and uniquely determined from the $(1+D)$ field values f_j at the sample points constituting the vertices of a Delaunay simplex.

Given the location $\vec{r}_0$, $\vec{r}_1$, $\vec{r}_2$ and $\vec{r}_3$, of the four points forming the Delaunay tetrahedra's vertices, and the field value estimates at these locations, f_0, f_1, f_2 and f_3, the gradient $\widehat{\nabla} f$ follows from the inversion

$$\widehat{\nabla} f \;=\; \begin{pmatrix} \dfrac{\partial f}{\partial x} \\[2ex] \dfrac{\partial f}{\partial y} \\[2ex] \dfrac{\partial f}{\partial z} \end{pmatrix} \;=\; \mathbf{A}^{-1} \begin{pmatrix} \Delta f_1 \\ \Delta f_2 \\ \Delta f_3 \end{pmatrix} \;; \qquad \mathbf{A} \;=\; \begin{pmatrix} \Delta x_1 & \Delta y_1 & \Delta z_1 \\ \Delta x_2 & \Delta y_2 & \Delta z_2 \\ \Delta x_3 & \Delta y_3 & \Delta z_3 \end{pmatrix} \tag{F.2}$$

where the coordinate distance between the tetrahedral vertices is represented by $\Delta x_n = x_n - x_0$, $\Delta y_n = y_n - y_0$ and $\Delta z_n = z_n - z_0$, and the corresponding differences between the vertex field values is given by $\Delta f_n \equiv f_n - f_0\,(n = 1, 2, 3)$.

Once the value of ∇_f for each Delaunay tetrahedron has been determined, it is straightforward to determine the DTFE field value $\widehat{f}(\vec{x})$ for any location $\vec{x}$ by means of straightforward linear interpolation within the Delaunay tetrahedron m in which $\vec{x}$ is located (eqn. F.2),

$$\widehat{f}(\vec{x}) \;=\; \widehat{f}(\vec{x}_i) \;+\; \widehat{\nabla f}\big|_m \cdot (\vec{x} - \vec{x}_i)\,. \tag{F.3}$$

In many practical situations we need a DTFE reconstructed field sampled on a grid. The created DTFE image may be then be used for further analysis. To this end, the DTFE density field value is sampled at each gridpoint of the grid. Formally, this should involve the DTFE density field value averaged over the corresponding gridcell. That value could be determined geometrically, although a more practical implementation involves the averaging over the interpolated field values (eqn. F.3) over a number or randomly placed points in a gridcell (e.g. Cautun et al. 2013). In a number of practical circumstances, a reasonable shortcut is to limit the field value calculation to that at the grid location. This offers a reasonable approximation for

gridcells which are smaller or comparable to that of intersecting Delaunay cells, on the condition that the field gradient within the cell(s) is not too large.

F.0.7 SPH kernel density estimates

In the previous appendix, we have described in some detail the technical aspects of DTFE. DTFE produces volume-weighted density estimates that adapt themselves to the local number density and geometry/shape of the particle distribution. The majority of conventionally applied techniques produce mass-weighted estimates. They usually involve a suitably weighted sum over discretely sampled field values, involving kernel weight functions $W(\mathbf{x}, \mathbf{y})$. One may directly appreciate that such convolutions produce mass-weighted estimates by converting the discrete sum into an integral over Dirac delta functions (see van de Weygaert & Schaap 2009b),

$$\begin{aligned} \hat{f}(\mathbf{x}) &= \frac{\sum_{i=1}^{N} \tilde{f}_i \, W(\mathbf{x}-\mathbf{x}_i)}{\sum_{i=1}^{N} W(\mathbf{x}-\mathbf{x}_i)} \\ &= \frac{\int d\mathbf{y}\, f(\mathbf{y})\, W(\mathbf{x}-\mathbf{y}) \, \sum_{i=1}^{N} \delta_D(\mathbf{y}-\mathbf{x}_i)}{\int d\mathbf{y}\, W(\mathbf{x}-\mathbf{y}) \, \sum_{i=1}^{N} \delta_D(\mathbf{y}-\mathbf{x}_i)} \\ &= \frac{\int d\mathbf{y}\, f(\mathbf{y})\, \rho(\mathbf{y})\, W(\mathbf{x}-\mathbf{y})}{\int d\mathbf{y}\, \rho(\mathbf{y})\, W(\mathbf{x}-\mathbf{y})} . \end{aligned} \tag{F.4}$$

Both rigid grid-based convolution schemes as well as a convolution formalism involving a spatially adaptive technique produce a mass-weighted average.

One particularly interesting example of the latter are the adaptive density functions used in Smooth Particle Hydrodynamics (SPH) codes (Hernquist & Katz 1989; Monaghan 2005; Springel 2008, see e.g.). The basic feature of the SPH procedure for density estimation consist of the convolution of the discrete particle distribution with a user-specified kernel function W that adapts itself to the local point density. For a sample of N particles, with masses m_j and locations $\mathbf{r}_j$, the density estimate ρ_{SPH} at the location $\mathbf{r}_i$ of particle i is given by

$$\rho(\mathbf{r}_i) \, = \, \sum_{j=1}^{N} m_j \, W(\mathbf{r}_i - \mathbf{r}_j, h_i) \, , \tag{F.5}$$

in which the kernel resolution is determined through the smoothing scale h_i. In general, the scale h_i is set by the local particle density. In many applications, the smoothing length h_i is chosen such that the sum involves a specific number of M nearest neighbours, usually in the order of $M \approx 40$.

The functional dependence of the kernel W is nearly always spherically summetric, so that it is the function of $|\mathbf{r}_i - \mathbf{r}_j|$ only. One possibility would be the use of Gaussian kernels W whose width h would be set by the local number density. More elaborate schemes, such as described by Monaghan (2005), use spline kernels. A

more advanced SPH scheme, introduced by Hernquist & Katz (1989), uses a symmetrized form of Eq. F.5. In general, the variants of SPH density estimates produce comparable results.

G

Skew-normal distribution

The skew-normal distribution with the *skewness parameter* α is given by(O'Hagan & Leonard (1976); Azzalini (1985))

$$f(x) = 2\phi(\nu)\Phi(\alpha\nu). \tag{G.1}$$

Here,

$$\phi(\nu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\nu^2}{2}}. \tag{G.2}$$

is the standard normal distribution. The function

$$\begin{aligned} \Phi(\nu) &= \int_{-\infty}^{x} \phi(t)\, dt \\ &= \frac{1}{2}\left[1 + \text{erf}\left(\frac{\nu}{\sqrt{2}}\right)\right], \end{aligned} \tag{G.3}$$

is the cumulative distribution function, in which *erf(x)* is the error function. Note that one recovers the familiar normal distribution when the skewness parameter $\alpha = 0$. he absolute value of skewness increases as the absolute value of α increases. Note that the skewness parameter α is different than the skewness, i.e. the third moment of the distribution

$$\gamma_1 = \mu^3/\sigma^3. \tag{G.4}$$

By definition, a curve has positive skewness if it has a more prominent tail for increasing values of ν, and a negative skewness when its balance is shifted towards decreasing values of ν.

To account for the location of the peak and the width of the curve, one usually makes the transformation

$$\mathrm{x} \to \mathrm{x} - \xi/\omega, \tag{G.5}$$

where ξ and ω are the location and scale parameters respectively.
The probability distribution function with location ξ, scale ω and skewness parameter α becomes

$$f(x) = \frac{2}{\omega}\phi\left(\frac{\nu-\xi}{\omega}\right)\Phi\left(\alpha\left(\frac{\nu-\xi}{\omega}\right)\right). \tag{G.6}$$

Introducing the amplitude parameter A_0, this takes the form

$$f(\nu) = \frac{A_0}{\omega\pi}\mathrm{e}^{-\frac{(\nu-\xi)^2}{2\omega^2}}\int_{-\infty}^{\alpha\left(\frac{\nu-\xi}{\omega}\right)}\mathrm{e}^{-\frac{t^2}{2}}\,\mathrm{d}t. \tag{G.7}$$