Verified Homology Computations for Nodal Domains

Sarah Day[*]   William D. Kalies[†]   Thomas Wanner[‡]

March 11, 2009

Abstract

Homology has long been accepted as an important computational tool for quantifying complex structures. In many applications these structures arise as nodal domains of real-valued functions and are therefore amenable only to a numerical study based on suitable discretizations. Such an approach immediately raises the question of how accurately the resulting homology can be computed. In this paper we present an algorithm for correctly computing the homology of one- and two-dimensional *nodal domains*. The approach relies on constructing an appropriate cubical approximation for the nodal domain based on the behavior of the defining function at the vertices of a fixed grid. Betti numbers for these cubical sets are readily computable using [25, 26]. Here, we present a technique to verify that the cubical representation is homeomorphic to the nodal domain, and therefore preserves homology. To illustrate this approach we consider examples from three classes of nodal domains, including the time-dependent patterns generated by the Cahn-Hilliard model for spinodal decomposition. We use these results to examine the probability of correct homology computations given specific grid sizes as related to the analytic estimates presented in [14, 34, 35, 40].

1   Introduction

1.1   Topology of complicated evolving patterns

The formation of complex patterns is ubiquitous throughout the applied sciences. In many cases these patterns exhibit a clear geometric structure, such as periodicity or certain symmetries. Experience, however, tells us that this is not always the case. Most patterns observed in biological systems are considerably more complicated, and standard symmetry or periodicity arguments cannot easily be applied in their study. Also in the context of materials science one can observe irregular time-varying patterns such as, for example, complex microstructures generated in binary metal alloys through a process called spinodal decomposition [9, 10, 30, 31, 45].

Due to the irregularity of these patterns, we choose to make a quantitative study based on topological properties in order to focus on *coarse* characteristics. In previous work we have shown that the use of homology groups and the resulting Betti numbers can provide valuable information about the patterns. While a precise definition of homology groups is not necessary for the current paper, we briefly recall the notion of Betti numbers of a topological space, which measure connectivity in different dimensions. More precisely, the zero-dimensional Betti number β_0 counts the number of components of a set, and the one-dimensional Betti number β_1 counts the number of loops that cannot be shrunk to a point within the set and cannot be transformed into one another. In a three-dimensional complex, the two-dimensional Betti number β_2 counts the number of

---

[*]Department of Mathematics, College of William and Mary, P.O. Box 8795, Williamsburg, VA 23187, (`sday@math.wm.edu`).

[†]Department of Mathematical Sciences, Florida Atlantic University, 777 Glades RD, Boca Raton, FL 33431, (`wkalies@fau.edu`).

[‡]Department of Mathematical Sciences, George Mason University, Fairfax, VA 22030, (`wanner@math.gmu.edu`).

enclosed cavities. Thus, the Betti numbers provide quantitative information on basic topological information. For more a more detailed discussion, see [20, 21, 25, 37]. As an example, we refer the reader to the left image in Figure 2 below. In this image, the dark regions form a set which as zero-dimensional Betti number $\beta_0 = 11$ and one-dimensional Betti number $\beta_1 = 3$. If instead one considers the light colored region, one has $\beta_0 = 12$ and $\beta_1 = 2$.

From a mathematical point of view, the first step towards studying the phenomenon of pattern formation in physical systems centers around deriving accurate mathematical models that can be used to predict system behavior, either analytically or through numerical simulation. In the context of phase separation in materials, one of the central approaches is the derivation of so-called phase-field models [12], which describe the evolution of a phase variable $u(t, x)$ as a function of time t and a spatial variable x through (in many cases) parabolic partial differential equations. The resulting function u is a continuous real-valued function which describes the composition of the underlying material through its function values. For example, Cahn and Hilliard [9, 10] proposed a fourth-order parabolic partial differential equation to model the phenomenon of spinodal decomposition. In this model, a phase variable u records the composition of the alloy as follows: values of $u(t, x)$ close to $+1$ indicate that at time t and position x the material consists almost exclusively of the first material, values close to -1 correspond to the second material, and values in between represent mixtures of the two materials. With this convention, values of u close to zero correspond to an equal mixture of the two underlying materials. Since we are interested in the microstructures created through the phase separation process, we study the sets where u is of one sign, i.e., we want to describe the so-called *nodal domains*

$$ N^+(t) = \{x \,:\, u(t, x) \geq 0\} \qquad \text{and} \qquad N^-(t) = \{x \,:\, u(t, x) \leq 0\} \,. $$

For the classical Cahn-Hilliard model in two dimensions, these nodal domains typically have the form shown in the left image of Figure 1, where the dark region corresponds to $N^+(t)$ and the light region to $N^-(t)$. Similar microstructures can be observed in three space dimensions (see the right image of Figure 1).

In [21] it is demonstrated that by studying the evolution of the Betti numbers of the nodal domains $N^{\pm}(t)$ one can uncover quantitative microstructure differences between the classical Cahn-Hilliard model and its stochastic extension, the Cahn-Hilliard-Cook model [13, 29]. Both models are given by

$$ \frac{\partial u}{\partial t} = -\Delta \left(\epsilon^2 \Delta u - F'(u) \right) + \sigma \cdot \xi \quad \text{for} \quad x \in \Gamma \quad \text{and} \quad t \geq 0 \,, \tag{1} $$

subject to no-flux boundary conditions for both u and Δu. The domain Γ is bounded, and the nonlinearity F is a double-well potential, usually defined as $F(u) = (u^2 - 1)^2/4$. Moreover, $\epsilon > 0$ is a small parameter which models the interaction length, and ξ denotes a white noise process (for more details see for example [7]). The parameter $\sigma \geq 0$ is a measure for the intensity of the random fluctuations. For $\sigma = 0$ we obtain the deterministic Cahn-Hilliard model, whereas $\sigma > 0$ corresponds to the stochastic Cahn-Hilliard-Cook model.

In order to quantitatively describe the evolution of the pattern complexity in (1), the study in [21] considers ensembles of initial conditions which are random small-amplitude perturbations of a homogeneous initial state, generally $u(0, x) \equiv 0$, which corresponds to an equal mixture of the two alloy components. The evolution equation is then solved numerically for each initial condition up to some fixed time, and the dimensions of the homology groups of the nodal domains $N^{\pm}(t)$ of u are computed. This procedure furnishes averaged Betti number evolution curves, which can be viewed as characteristic descriptors of the complexity evolution of the microstructures. By varying the noise intensity σ and the size of the domain Γ, one can then study their effect on the resulting patterns. The main results of [21] can be summarized as follows.
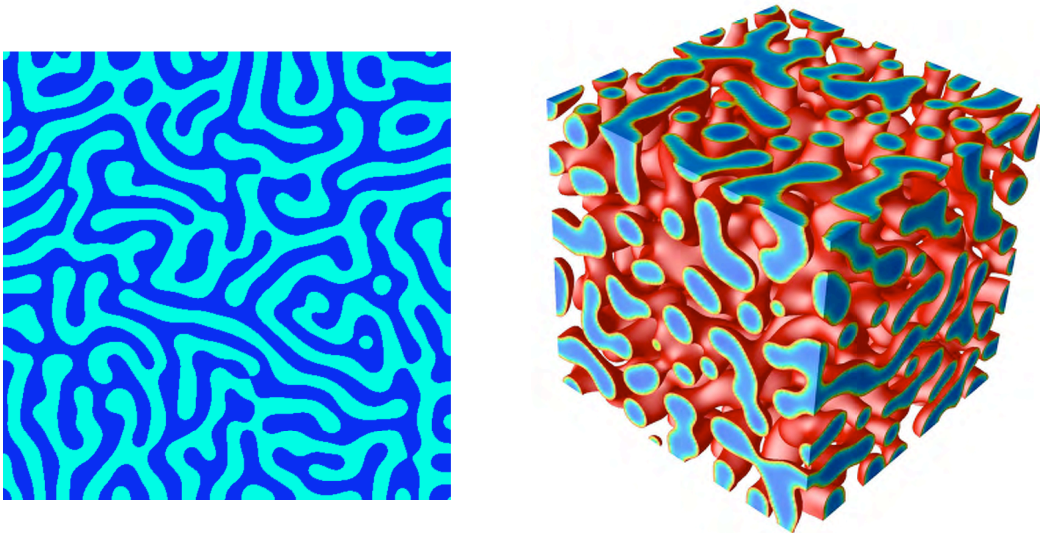
Figure 1: Complicated microstructures generated from simulations of the Cahn-Hilliard model.

- The deterministic Cahn-Hilliard model exhibits a surprising non-monotone behavior of the complexity evolution curves. This effect weakens with increasing noise intensity σ. If the noise intensity exceeds a certain threshold, the complexity evolution exhibits monotone decay, as would be expected from the experimental data in [24]. It is possible to heuristically explain the non-monotone behavior using the results of [43, 45].

- By combining the homology information of the complementary sets $N^{\pm}(t)$, one can distinguish between boundary effects and bulk behavior in the material. It turns out that the average number of components touching the boundary of Γ is given by the average of the Euler characteristic of the pattern, where the average is taken over the ensemble of initial conditions. The Euler characteristic is a weaker topological invariant than the set of Betti numbers and can be calculated as the alternating sum of the Betti numbers, which in two or three dimensions is given by $\beta_0 - \beta_1 + \beta_2$, see [37]. Bulk effects can only be described by the complete set of Betti numbers. Moreover, scaling the underlying domain has different effects on the topological quantities. While the averaged Euler characteristic scales with the length of the boundary, the averaged Betti numbers scale with the area.

These results demonstrate that the topological information contained in the Betti numbers of the microstructures can be used to distinguish different models, or to compare model behavior to experiments. Moreover, the provided connectivity information exceeds by far what can be obtained from studies using the Euler characteristic, which had been used extensively in the past [5, 11, 24, 32, 33].

The above discussion describes one particular instance in which the study of topological properties of nodal domains is important, but there are many others. For example, in the context of simulating problems involving moving fronts, level set methods have been used to describe the evolution of a front via the evolution of an underlying auxiliary function, which describes the front implicitly as a level set [41, 44]. One of the main reasons for the introduction of this concept was that it can easily deal which topological changes in the level set. In a probabilistic context, the topology of sub- or super-level sets can be used to determine the asymptotic behavior of excursion probabilities of random fields, see for example [1, 2]. Furthermore, homological techniques have been used successfully for quantitative studies involving spatio-temporal chaos [20], complicated flow patterns in Rayleigh-Benard convection [28], nonlocal and stochastic extensions of the classical phase-field model for non-isothermal phase separation [22, 23], and even in the study of residual

stress networks in polycrystals [18].

1.2 Validation of homology computations

In this paper, we address the topological study of nodal domains from a rigorous computational perspective. For this, let $f : \Gamma \to \mathbb{R}$ be a smooth function, with Γ a compact, rectangular domain in $\mathbb{R}$ or $\mathbb{R}^2$. By rescaling f, the domain can be taken to be the unit interval or square, that is, we have $\Gamma = [0, 1] \subset \mathbb{R}$ or $\Gamma = [0, 1] \times [0, 1] \subset \mathbb{R}^2$. We are interested in the topology of the nodal domains

$$N^+ = \{x \in \Gamma \, : \, f(x) \geq 0\} \qquad \text{and} \qquad N^- = \{x \in \Gamma \, : \, f(x) \leq 0\} \, ,$$

in particular, in computing the homology groups $H_*(N^\pm)$ of $N^\pm$.

To the best of our knowledge one of the first results concerning the accuracy of estimation of this topological information is due to Niyogi, Smale, and Weinberger [40]. Similar methods had been introduced earlier for the reconstruction of surfaces in [3, 4], and by combining these results with the ones of [15] it is possible to recover the main conclusions of [40]. In the paper by Niyogi et al., the authors propose a stochastic algorithm for computing the homology of a given manifold $X \subset \mathbb{R}^d$ by randomly sampling M points from the manifold, and explicit bounds are derived on the probability that their algorithm computes the correct homology. The probability bound depends on the number M and a condition number $1/\tau$. The latter parameter encodes both local curvature information of the manifold X, as well as global separation properties. More precisely, the inverse condition number τ is the largest number such that the open normal bundle about $X \subset \mathbb{R}^d$ of radius r is embedded in $\mathbb{R}^d$ for all $r < \tau$.

For nodal domains, it seems difficult to estimate τ based on computable properties of f. However, in this context there is a simpler method of sampling. Cover Γ with a uniform cubical grid, and then the numerical computation of the values of f on the vertices of the grid can be used to determine a set of points in N^+. This determination of roughly evenly spaced points in N^+ replaces the notion of random sampling in the Niyogi, Smale, and Weinberger approach. The computed collection of points is then used to construct a cubical approximation $\mathcal{N}_M^+$ of N^+ such as in Figure 2, and cubical homology software packages such as [26, 25] may then be applied to compute the Betti numbers of the cubical approximation. As can be seen in Figure 2, the homology of this cubical approximation need not be the same as that of N^+.

Given certain assumptions on the properties of the nodal domain (related to the assumptions in Niyogi, Smale, and Weinberger's approach), and a sufficiently fine grid, we expect this approach to yield the correct homology of the nodal domain. In a probabilistic sense this has been justified in [34], where rigorous lower bounds are derived for the probability of correctly computing the homology of nodal domains of random fields given a fixed uniform grid size. Since we will be addressing the sharpness of these bounds later in this paper, we briefly recall one of the main results in [34]. Consider a random Fourier series in two space dimensions of the form

$$\begin{aligned} f(x, \omega) \;\; = \;\; & \sum_{k,\ell=0}^{\infty} a_{k,\ell} \cdot \left(g_{k,\ell,1}(\omega) \cos(2\pi k x_1) \cos(2\pi \ell x_2) + g_{k,\ell,2}(\omega) \cos(2\pi k x_1) \sin(2\pi \ell x_2) \right. \\ & \left. + g_{k,\ell,3}(\omega) \sin(2\pi k x_1) \cos(2\pi \ell x_2) + g_{k,\ell,4}(\omega) \sin(2\pi k x_1) \sin(2\pi \ell x_2) \right) \, , \end{aligned} \qquad (2)$$

which defines a periodic random function on the square domain $\Gamma = [0, 1]^2$. In the definition of f, the functions $g_{k,\ell,m}$ denote Gaussian random variables over a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ which are independent and normally distributed with mean 0 and variance 1, and the $a_{k,\ell}$ are real constants. In addition, assume that there are positive $k_1, \ell_1 \in \mathbb{N}$ and nonnegative $k_2, \ell_2 \in \mathbb{N}_0$ which satisfy $k_1 \neq k_2$ and $\ell_1 \neq \ell_2$, as well as $k_1^2 + \ell_1^2 \neq k_2^2 + \ell_2^2$, such that both a_{k_1,ℓ_1} and a_{k_2,ℓ_2} are nonzero.
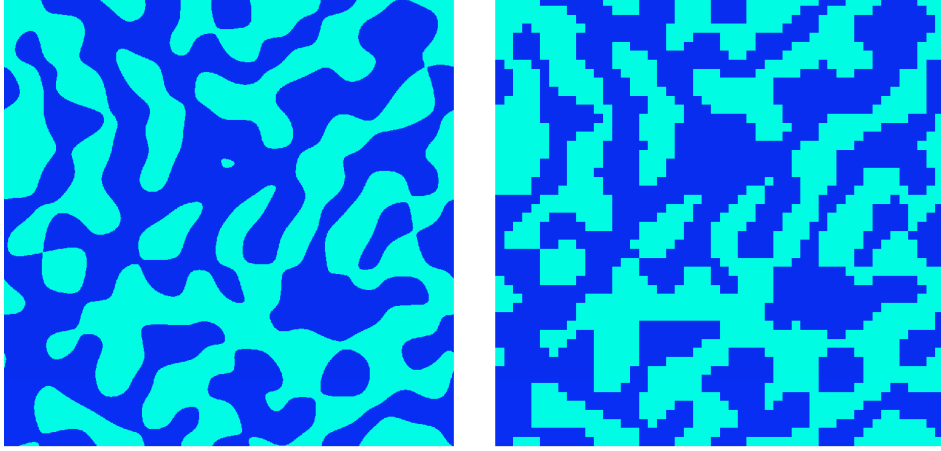
Figure 2: Nodal domains of a random trigonometric polynomial in two space dimensions and their cubical approximations with $M = 50$. Note that the topology of the cubical representation can possibly be different from the topology of the nodal domain. One quantitative measurement that implies a different topology is the set of Betti numbers, which for the dark region are $\beta_0 = 11$, $\beta_1 = 3$ for the nodal domain and $\beta_0 = 9$, $\beta_1 = 4$ for the cubical domain.

Finally, suppose that

$$\sum_{k,\ell=0}^{\infty} \left(k^6 + \ell^6 \right) a_{k,\ell}^2 < \infty .$$

These assumptions guarantee that the random function f includes sufficient randomness and that the sample realizations are almost surely at least twice continuously differentiable.

Now consider the random nodal domains

$$N^+(\omega) = \{ x \in \Gamma \; : \; f(x,\omega) \geq 0 \} \qquad \text{and} \qquad N^-(\omega) = \{ x \in \Gamma \; : \; f(x,\omega) \leq 0 \} ,$$

as well as cubical approximations $\mathcal{N}_M^{\pm}(\omega)$ constructed from the evaluation of $f(\cdot,\omega)$ at the $(M+1)^2$ discretization points $x_{k,\ell} = (k/M, \ell/M)$ for $k, \ell = 0, \ldots, M$. Then [34, Theorem 3.10] gives a lower bound on the probability that the nodal domains $N^{\pm}(\omega)$ have the same homology as their cubical approximations $\mathcal{N}_M^{\pm}(\omega)$. More precisely, it is shown that

$$\mathbb{P}\left\{ H_*(N^{\pm}) \cong H_*(\mathcal{N}_M^{\pm}) \right\} \; \geq \; 1 - \frac{1067\pi^2}{18M^2} \cdot \frac{(A_{2,0} + A_{1,1} + A_{0,2})^2}{A_{0,0}^{1/2} A_{0,1}^{1/2} A_{1,0}^{1/2} A_{1,1}^{1/2}} + O\left(\frac{1}{M^3} \right) , \tag{3}$$

where $A_{p,q}$ is defined by

$$A_{p,q} \;\; = \;\; \sum_{k,\ell=0}^{\infty} k^{2p} \ell^{2q} a_{k,\ell}^2 . \tag{4}$$

Notice that these values are related to averaged $L^2(\Gamma)$-norms of the random function f and its derivatives, since we have

$$\mathbb{E} \left\| D_{x_1}^p D_{x_2}^q f \right\|_{L^2(\Gamma)}^2 \;\; = \;\; (2\pi)^{2p+2q} \cdot A_{p,q} ,$$

where $\mathbb{E}$ denotes the expected value of a random variable over $(\Omega, \mathcal{F}, \mathbb{P})$.

The above result provides fundamental insight into the relationship between the discretization size and averaged "curvature" information of the random function f. More specifically, specifying
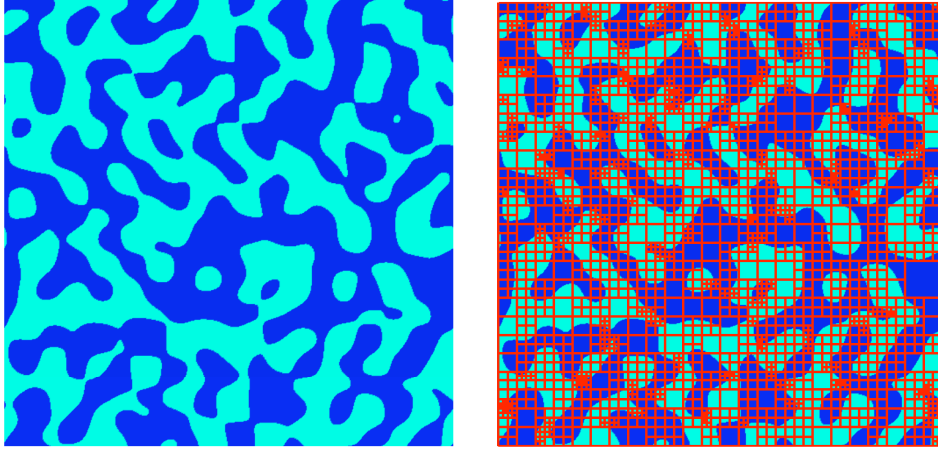
Figure 3: Nodal domains of a bivariate random trigonometric polynomial of degree $K = 10$ and boxes on which the topology of N^+ can be determined from the corner function values.

a confidence probability for the correctness of the homology computation determines a sufficient discretization size a-priori. However, this result does leave a number of questions unanswered. First, it does not address the tightness of the lower bound. More importantly, the theory presented in [34] relies on the use of Gaussian fields which is not appropriate in a number of interesting applications, most notably for patterns created by nonlinear stochastic partial differential equations.

1.3 Structure of the paper and main results

Partially motivated by the questions just described, in the present paper, we develop a *verified* numerical approach to computing the homology of $N^\pm$. In particular, by including bounds on first and second derivatives of f and interval arithmetic to account for round-off error in the numerical computations, we may check that the structure of the nodal domain has been computed accurately within each grid element. For efficiency, we construct an adaptive grid which is finer where greater resolution is required to accurately represent and verify $N^\pm$, as shown in Figure 3. Our verified numerical approach is described in detail in Sections 2 and 3. After introducing necessary mathematical tools and results on interval arithmetic, the complete adaptive algorithm is described in the remainder of Section 2. Its validity is then established in Section 3.

In Section 4, we use the verified homology procedure on three specific families of nodal domains: double-well potentials, where we study connections to the condition number in [40], random trigonometric polynomials, where we explore the sharpness of the estimates in [34], and solutions to the Cahn-Hilliard equation. The latter case is of particular importance due to the above-mentioned restrictions of the results in [34]. First, we study the effect of the inherently non-Gaussian distribution of the random fields produced by the Cahn-Hilliard-Cook evolution. In addition, we note that for functions where the boundary of the nodal domain is not smooth, the homology need not be correctly computed for any grid size. When considering the time evolution of the patterns generated by the Cahn-Hilliard-Cook model, one therefore has to expect that near times where the topology of the pattern is changing, the homology is not computed correctly for computationally tractable grid sizes. For this reason, we also investigate the amount of time in which the homology is computed correctly in the time series of Betti numbers. The main findings of our studies in Section 4 can be summarized as follows.

- For simple geometric features generated through sublevel sets of a double-well potential, the main performance parameters of our verified numerical algorithm scale logarithmically in the

feature size or feature curvature. This is in stark contrast to the polynomial scaling observed in the probabilistic algorithm of [40].

- For two-dimensional doubly-periodic random polynomials we show that the theoretical estimates of [34] on the discretization size which is necessary to determine the nodal domain homology with high probability are sharp in general. Moreover, we demonstrate that while these theoretical estimates cannot readily take advantage of any significant spatial correlations, our adaptive algorithm does. In other words, the complexity of our algorithm is related to the actual topological complexity of the nodal domains, and not on worst-case behavior.

- In the context of temporally evolving complicated patterns which are observed during phase separation we draw two important conclusions. For the case of one-dimensional base domains, we show that the theoretical predictions for the discretization size given in [34] are sharp as long as the random field describing the solution snapshot is close to Gaussian, i.e., during the linear initial phase separation regime. As soon as nonlinearity effects set in, the probabilistic estimates are overestimations. In addition, we show that as a function of time our adaptive algorithm validates the nodal domains almost always. In fact, even though one would expect algorithm failure whenever the topology of the nodal domains changes, for realistic time discretizations it does not fail, even in the two-dimensional case.

2   An algorithm for verified homology computations

In this section, we present an algorithm to construct a cubical domain which has the same homology as the nodal domain N^+. Indeed, as shown in Section 3, the resulting cubical complex is homeomorphic to N^+. The idea behind this construction is to combine an adaptive, binary subdivision approach for constructing a cubical decomposition of the domain with a verification step to check the behavior of the function on each cube. The procedure is described for a two-dimensional nodal domain, and a similar approach works in the simpler one-dimensional setting.

Before describing our procedure, we must address some computational issues. We use the symbols $\{\oplus, \ominus, \otimes, \oslash\}$ to denote the standard interval arithmetic operations as defined in [27, 36, 38, 39]. If one of the operands is a real number a, then it should be considered as a degenerate interval $[a] := [a, a]$. If F is a continuous real-valued function and B is a compact, connected subset of its domain, then $F(B)$ is an interval. Numerically, interval arithmetic operations and interval function evaluation are performed using outward rounding so that the interval computed using floating-point numbers is guaranteed to contain the true interval result. In this case, we emphasize that an interval I is a computationally rounded interval by the notation $\widetilde{I}$. There are a number of software libraries that perform these calculations, and we use CAPD [19] for the computations shown in Section 4.

Now, the first step in the algorithm to compute a cubical approximation to a nodal domain is to subdivide $\Gamma = [0, 1] \times [0, 1]$ into a uniform grid. The amount of this initial subdivision can be chosen arbitrarily, including no initial subdivision. We compute the sign of f on each of the vertices of the grid. It is again important to note that all numerical computations are performed using interval arithmetic to account for round-off errors. For each vertex v we compute a small interval $\widetilde{f(v)}$ containing $f(v)$. If $\widetilde{f(v)}$ does not contain zero, then we have verified the sign of $f(v)$. For a random function in the classes of functions we typically consider, the probability that f has a zero at a vertex is zero. However, it is possible, due to round-off error or the wrapping effect, that $0 \in \widetilde{f(v)}$, and hence the sign of $f(v)$ cannot be verified. For simplicity, we declare that the algorithm fails if we cannot verify the sign at a vertex.

For each box B in the grid, we now define a *verification step* whereby we determine the topological structure of $N^+ \cap B$. The verification step will depend on the sign configuration on the
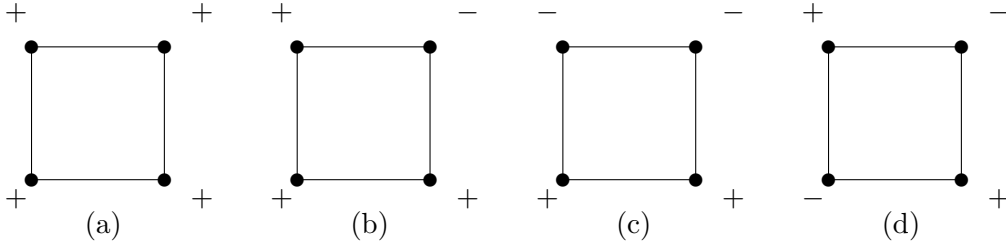
Figure 4: Possible sign structures (up to rotation and negation) on the vertices of a grid element.

vertices of B, which falls into one of the four configurations (up to orientation and negation) shown in Figure 4. If the verification step on B fails, we subdivide B in each coordinate direction and perform the verification step on each of the smaller boxes contained in B. This procedure continues until all boxes in the grid have passed the verification step, or until the grid is refined beyond a preset resolution.

The verification step involves checking computable conditions that guarantee that the topology of the nodal domain in B is captured by the sign structure on the vertices of B (see Figure 5). In particular, we attempt to establish that f is either bounded away from zero on B in case (a) in Figure 4 or is monotonic along appropriate rays through B in cases (b) and (c) in Figure 4. The required bounds on $f(B)$, $f_x(B)$, and $f_y(B)$ are directly computable using interval arithmetic. However, because the outward rounding employed in this approach may produce very large bounds, we employ a form of the Mean Value Theorem to obtain tighter bounds. The motivating idea behind this approach is given in the following proposition.

Proposition 2.1 *Suppose that the function $g : [a,b] \to \mathbb{R}$ is continuous on the closed interval $[a,b] \subset \mathbb{R}$ and is continuously differentiable on (a,b). Then*

$$g([a,b]) \subset g\left(\frac{a+b}{2}\right) \oplus \frac{b-a}{2} \otimes g'([a,b]) \otimes [-1,1] \tag{5}$$

Proof. By the Mean Value Theorem, for $x \in [a,b]$, there exists $c \in [a,b]$ such that

$$
\begin{aligned}
g(x) &= g\left(\frac{a+b}{2}\right) + g'(c)\left(x - \frac{a+b}{2}\right) \\
&\in g\left(\frac{a+b}{2}\right) \oplus \frac{b-a}{2} \otimes g'([a,b]) \otimes [-1,1] \quad \text{for all } x \in [a,b].
\end{aligned}
$$

■

Corollary 2.2 *Let $B = [a_1, b_1] \times [a_2, b_2]$ and $f : B \to \mathbb{R}$ be a C^1 function. If*

$$0 \notin f\left(\frac{a_1+b_1}{2}, \frac{a_2+b_2}{2}\right) \oplus \frac{b_1-a_1}{2} \otimes f_x(B) \otimes [-1,1] \oplus \frac{b_2-a_2}{2} \otimes f_y(B) \otimes [-1,1], \tag{6}$$

then f is bounded away from 0 on B.

Proof. For $x \in B$ let $\gamma : [-1,1] \to \mathbb{R}^2$ be a linear parametrization of the intersection of B with the line through the points x and $(\frac{1}{2}(a_1 + b_1), \frac{1}{2}(a_2 + b_2))$ with $\gamma(0) = (\frac{1}{2}(a_1 + b_1), \frac{1}{2}(a_2 + b_2))$. We now

apply Lemma 2.1 to the function $g : [-1, 1] \to \mathbb{R}$ defined by $g = f \circ \gamma$ to obtain

$$
\begin{aligned}
f(x) &= g(\gamma^{-1}(x)) \\
&\in g(0) + g'([-1, 1]) \otimes [-1, 1] \\
&\subset f\left(\frac{a_1 + b_1}{2}, \frac{a_2 + b_2}{2}\right) \oplus (f_x(B), f_y(B)) \odot \left(\gamma([-1, 1]) \ominus \left(\frac{a_1 + b_1}{2}, \frac{a_2 + b_2}{2}\right)\right) \\
&\subset f\left(\frac{a_1 + b_1}{2}, \frac{a_2 + b_2}{2}\right) \oplus \frac{b_1 - a_1}{2} \otimes f_x(B) \otimes [-1, 1] \oplus \frac{b_2 - a_2}{2} \otimes f_y(B) \otimes [-1, 1].
\end{aligned}
$$

Note that $\odot$ denotes the dot product using interval arithmetic. $\blacksquare$

In practice, the bound computed from the right-hand side of (6) is often much smaller than the bound given by evaluating $f(B)$ directly using interval arithmetic. A similar approach may be adopted for studying monotonicity along rays in B. We again apply Proposition 2.1 to obtain the necessary bounds.

Definition 2.3 Let $B = [a_1, b_1] \times [a_2, b_2]$. We will say that $f : B \to \mathbb{R}$ is **monotone in the x-direction** on B if $0 \notin f_x([a_1, b_1], y)$ for each $y \in [a_2, b_2]$. Similarly, f is **monotone in the y-direction** on B if $0 \notin f_y(x, [a_2, b_2])$ for each $x \in [a_1, b_1]$.

Corollary 2.4 Let $B = [a_1, b_1] \times [a_2, b_2]$ and $f : B \to \mathbb{R}$ be a C^2 function. If

$$
0 \notin f_x\left(\frac{a_1 + b_1}{2}, [a_2, b_2]\right) \oplus \frac{b_1 - a_1}{2} \otimes f_{xx}(B) \otimes [-1, 1] \tag{7}
$$

then f is monotone in the x-direction on B. Similarly, if

$$
0 \notin f_y\left([a_1, b_1], \frac{a_2 + b_2}{2}\right) \oplus \frac{b_2 - a_2}{2} \otimes f_{yy}(B) \otimes [-1, 1] \tag{8}
$$

then f is monotone in the y-direction on B.

Proof. For $(x^*, y^*) \in B$, let $g = f_x(\cdot, y^*) : [a_1, b_1] \to \mathbb{R}$. Then by Proposition 2.1,

$$
\begin{aligned}
f_x(x^*, y^*) &= g(x^*) \\
&\in g\left(\frac{a_1 + b_1}{2}\right) \oplus \frac{b_1 - a_1}{2} \otimes g'([a_1, b_1]) \otimes [-1, 1] \\
&\subset f_x\left(\frac{a_1 + b_1}{2}, y^*\right) \oplus \frac{b_1 - a_1}{2} \otimes f_{xx}([a_1, b_1], y^*) \otimes [-1, 1] \\
&\subset f_x\left(\frac{a_1 + b_1}{2}, [a_2, b_2]\right) \oplus \frac{b_1 - a_1}{2} \otimes f_{xx}(B) \otimes [-1, 1].
\end{aligned}
$$

Therefore, if (7) holds, then $f_x(z) \neq 0$ for all $z \in B$ and f is monotone in the x-direction on B. The second part of the corollary follows by a similar argument. $\blacksquare$

While it is also possible to study monotonicity along rays in directions other than the coordinate directions, we find the tests listed in Corollary 2.4 to be both efficient for coding purposes and sufficient for our studies. On one further technical point, we found that in computing bounds on $f(B)$, the process of subdividing B into D_{box}^2 uniform boxes $B_1, \ldots, B_{D_{box}^2}$, where D_{box} denotes a suitable integer, then performing interval arithmetic to obtain bounds on each $\widetilde{f(B_i)}$, $i = 1, \ldots, D_{box}^2$, and finally setting $\widetilde{f(B)}$ to be the smallest bounding interval of $\cup_{i=1}^{D_{box}^2} \widetilde{f(B_i)}$, can produce considerably
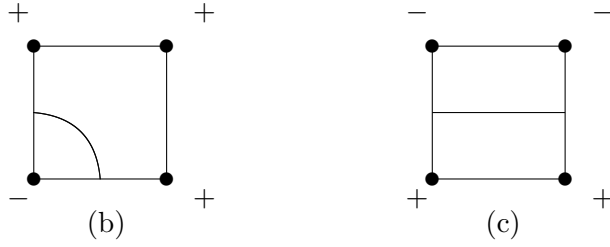
Figure 5: Possible structure of $\partial(N^+) \cap B$ in B up to rotation and negation.

tighter bounds. Due to the additional computational effort, the integer D_{box} has to be chosen appropriately. Through initial experiments, we found that choosing $D_{box} = 4$ significantly increases the tightness of the bounds without adding too much additional computational expense. Nevertheless, for more complicated patterns, such as for example random trigonometric polynomials of higher degree, the value of D_{box} has to be increased. For the simulations in this paper we use values of D_{box} between 4 and 12.

We now describe how to apply the tools offered by Corollaries 2.2 and 2.4 to the study of the sign configurations shown in Figure 4.

Case (a) *in Figure* 4: In this case, we try to verify the hypothesis of Corollary 2.2 in order to show that $N^+ \cap B = B$ or $N^+ \cap B = \emptyset$.

Cases (b) *and* (c) *in Figure* 4: Since the sign structure on the vertices for each of these cases indicates that there must be a sign change for f in the interior of B, we now check that this sign change occurs in the simplest possible way. In other words, we check that $\partial(N^+) \cap B$ looks topologically like the pictures listed in Figure 5 by verifying the appropriate hypotheses in Corollaries 2.2 and 2.4. More specifically, for Case (c) with the orientation depicted in Figure 5, we first use Corollary 2.2 with input boxes $[a_1, b_1] \times [a_2, a_2]$ and $[a_1, b_1] \times [b_2, b_2]$ to check that f is bounded away from 0 on the top and bottom edges of B and then use Corollary 2.4 to test that f is monotone in the y-direction on B. In Case (b) we check that f is monotone in both the x and y-directions on B. Note that we are checking a condition that is stronger than necessary in Case (b). This stronger condition can be verified in the work we present here and has the added benefit that it simplifies the coded algorithm.

Case (d) *in Figure* 4: The sign structure on the vertices in this case indicates that more resolution is required to approximate $N^+ \cap B$. We consider a box of this sign structure to automatically fail the verification step, and therefore it is subdivided.

Suppose this subdivision and verification procedure terminates successfully. Then we obtain a possibly nonuniform cubical decomposition of Γ as in Figure 3. In order to construct a cubical approximation of N^+, we consider the uniform grid with grid size equal to the minimal cube size in the nonuniform decomposition, $1/M$, and augment the region Γ to $\Gamma_M = [-1/(2M), 1 + 1/(2M)]$ or $[-1/(2M), 1 + 1/(2M)]^2$. Let $\mathcal{G}_M$ denote the uniform grid of cubes of size $1/M$ on Γ_M. Now we perform one last verification test to determine the sign of f at the center of each cube in $\mathcal{G}_M$. (Note that these centers are exactly the vertices of cubes in our subdivision and verification procedure.) If this test is passed, we define the cubical approximation $\mathcal{N}^+$ of N^+ by the condition that the grid element $B \in \mathcal{G}_M$ is in $\mathcal{N}^+$ if and only if the vertex sign at the center of B is positive. As previously mentioned, the homology of the cubical set $\mathcal{N}^+$ may now be computed using [26]. We prove in the next section that if this procedure is successful, then in fact one has $H_*(\mathcal{N}^+) \approx H_*(N^+)$.

The above-described procedure works well in many applications. However, particularly for time-evolving patterns, it is possible that due to grid alignment issues of the nodal lines the minimal cube size $1/M$ in the nonuniform decomposition can become extremely small, which in turn leads

10

to large cubical approximations $\mathcal{N}^+$ and long computational times for their homology. We have therefore extended the above algorithm by a procedure which allows for the verification step to be performed recursively. More precisely, the recursive implementation of our algorithm uses the following modified verification step for a given box B:

- For each of the four edges of the box B, verify the sign behavior on this edge via interval arithmetic. In other words, if adjacent corners have the same sign, the function has to be of one sign throughout the adjoining edge, if the corners have opposite signs, the respective partial derivative of f has to be nonzero on the edge. For finding enclosures of function values on the edges, we subdivide the edge into D_{edge} subintervals, similarly to the earlier discussion of D_{box}. For the computations in this paper we always use $D_{edge} = 4 \cdot D_{box}$.

- If the appropriate sign and/or monotonicity conditions can be verified for the box given the sign configuration on its vertices as described above, then the box also verifies in the recursive version of the algorithm.

- If the box cannot be verified using the above procedure, subdivide the box into four subboxes and verify these recursively. If all of them can be verified, then mark the original box as verified.

By allowing this recursive verification procedure, the minimal cube size $1/M$ in the nonuniform decomposition increases significantly, in most cases by several orders of magnitude. Moreover, if a box B passes the recursive verification step, one can still guarantee that the location of the nodal line within B can be deduced from the signs of the function values at the corners. However, the nodal line no longer has to be monotone within B as in the non-recursive algorithm, which may result in the cubical approximation $\mathcal{N}^+$ as described above furnishing the wrong homology. To see this, note that a snaking nodal line might introduce artificial loops for certain discretization sizes M.

In order to avoid this problem, we have implemented the construction of a cubical approximation $\tilde{\mathcal{N}}^+$ which is based solely on the sign configurations at the vertices of the recursively verified cubes and avoids evaluating additional function values inside the cubes. Consider again the collection $\mathcal{G}_M$ of $(M + 1)^2$ cubes covering Γ_M as described previously. Then the cubical set $\tilde{\mathcal{N}}^+$ is constructed as follows. Let B denote a recursively verified cube in the final adaptive grid, say with side length k/M for some integer k. If the signs at the four vertices of B are all positive, then we add all $(k+1)^2$ cubes of $\mathcal{G}_M$ which overlap with B to $\tilde{\mathcal{N}}^+$. If on the other hand B contains negative vertices, then we only add the cubes of $\mathcal{G}_M$ which overlap with the edges of B with positive vertices. In other words, if B has only one such edge we add $k+1$ cubes, and if B has exactly two such edges we add $2k + 1$ cubes. If all of the vertices of B have negative signs, no cubes from $\mathcal{G}_M$ are added to $\tilde{\mathcal{N}}^+$. Notice that if some cube in $\mathcal{G}_M$ overlaps with two adjacent recursively verified cubes B_1 and B_2, then it might only be included in $\tilde{\mathcal{N}}^+$ when considering, say, the verified cube B_1. This situation arises if B_2 is a larger recursively verified cube, and the smaller cube B_1 intersects B_2 on an edge which has a sign change for B_2, whereas the intersecting edge $B_1 \cap B_2 \subset B_1$ is all positive. In rare situations, this observation can result in a recursively verified cube B for which only $4k - 1$ cubes overlapping with the edges of B are added, while exactly one corner cube is not. (For this to happen the nodal line has to pass within distance $1/M$ from the only negative corner of B.) In this case, one also has to add the $(k - 2)^2$ cubes from $\mathcal{G}_M$ which lie completely in the interior of B to $\tilde{\mathcal{N}}^+$, in order to avoid the creation of an artificial loop. In our implementation this is achieved by a second sweep through the recursively verified cubes with exactly one negative corner.

We illustrate the construction of $\tilde{\mathcal{N}}^+$ for the positive nodal domain N^+ of a random trigonometric polynomial with $K = 8$, shown in dark blue in the left image of Figure 6. The center image shows the final recursively verified adaptive grid, and the resulting cubical approximation $\tilde{\mathcal{N}}^+$ of N^+ is shown on the right in dark blue. Notice the appearance of the little "spines" which are due to
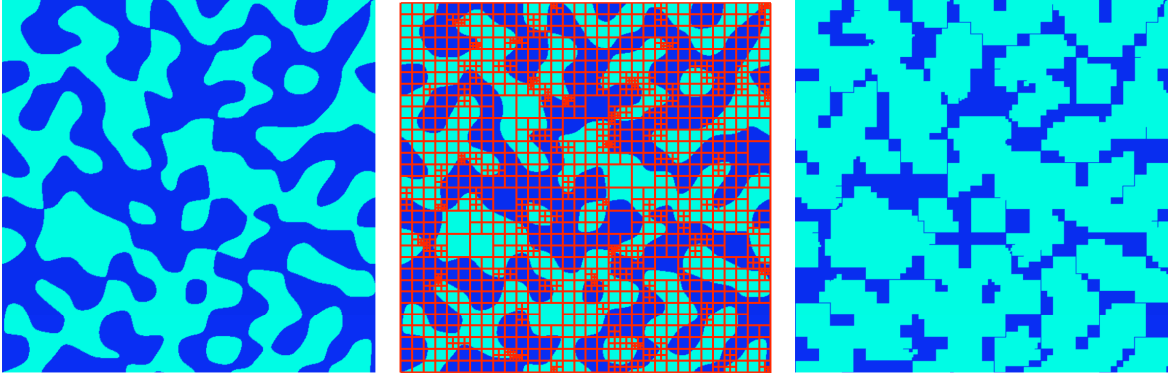
Figure 6: Nodal domains of a bivariate random trigonometric polynomial of degree $K = 8$, its recursively verified adaptive grid, and the cubical approximation $\tilde{\mathcal{N}}^+$ of N^+.

the above-mentioned fact that cubes from $\mathcal{G}_M$ which intersect an edge of the verified adaptive grid might only be added when processing one of the adjacent verified cubes.

3   Proof of correctness for verified homology

In this section, we address the fundamental concern of whether the computed topological information is accurate. Here we prove that using the construction outlined in Section 2, we either obtain the correct Betti numbers or the algorithm returns a failure. For now, we return to the discussion of the original version of the algorithm described in Section 2, i.e., the version without recursion. Recall that we define a cubical approximation $\mathcal{N}^+$ of the nodal domain N^+ based on the following construction.

We first compute a nonuniform cubical decomposition of Γ in which every element passes the appropriate verification test. Setting $1/M$ to be the finest (smallest) cube size in this grid, we next construct an extended uniform grid $\mathcal{K}_M$ of size $1/M$ on Γ by adding cubes of the same size $1/M$ along the left and bottom edges of Γ. The function signs on the vertices of $\mathcal{K}_M \cap \Gamma$ are then determined rigorously. Finally, we set

$$\mathcal{N}^+ := \bigcup \{B \in \mathcal{K}_M \mid \text{the upper right hand vertex of } B \text{ has sign } +\}. \tag{9}$$

(Notice that $\mathcal{K}_M$ is just a translation of $\mathcal{G}_M$ from the last section.) An adaptive grid allows us to more efficiently approximate the topology of N^+ (see for example Figure 3). However, the input to the computational homology software should be a cubical complex on a grid that is homeomorphic to an integer lattice. This necessitates the refinement to $\mathcal{K}_M$ and the definition of $\mathcal{N}^+$ on $\mathcal{K}_M$. However, some of the cubes in $\mathcal{K}_M$ may not pass the appropriate verification. Figure 7 shows an example where a larger cube satisfies the verification criteria, but the verification step necessarily fails on a cube in the refined grid. For simplicity of presentation, in the proof below we assume that each cube in the uniform grid $\mathcal{K}_M$ satisfies the verification criteria. In this case we prove that the cubical approximation $\mathcal{N}^+$ is homeomorphic to the nodal domain N^+, and, therefore, that the homology computed for $\mathcal{N}^+$ is also the homology of N^+. One can use the techniques from the proof below to show that if each cube in an adaptive grid is verified and M is chosen so that $\mathcal{K}_M$ is a uniform refinement of this grid, then $\mathcal{N}^+$ is again homeomorphic to N^+.

The construction of a homeomorphism from $N^+ \cap \Gamma$ to a cubical set begins by defining a map h on $N^+ \cap B$ for each $B \in \mathcal{G}_M$ and $B \subset \Gamma$. For a fixed B we consider the origin of $\mathbb{R}^2$ to be the lower left vertex of B, and the coordinates (x, y) are the horizontal and vertical distances from this vertex. Let $\Pi : \mathbb{R} \times [0, \pi/2] \to [0, \infty)^2$ denote the polar coordinate map $\Pi(r, \theta) = (r\cos(\theta), r\sin(\theta))$.
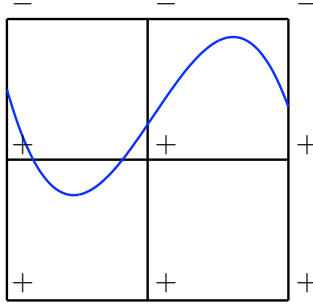
Figure 7: The large square has been verified so that f is monotone in the y-direction and the nodal line (blue) is the graph of a function of x, but the lower left subcube does not pass the verification test since f does not have constant sign there.

Case (a): B has the sign structure shown in Figure 4(a).

In this case, the verification step assures that $f(x) > 0$ for all $x \in B$. We define the homeomorphism $h : B \to B$ on this cube to be the identity map $h(x) = x$ as shown in Figure 8(a).

Case (b): B has the sign structure shown in Figure 4(b).

The verification step establishes that f is monotone on all horizontal and vertical lines in B. Therefore, there exists a continuous function $r^* : [0, \pi/2] \to (0, \infty)$ such that $(r^*(\theta), \theta)$ is the unique zero of $f(\Pi(\cdot, \theta))$ in B and $N^+ \cap B = \{\Pi(r, \theta) \mid 0 \le r \le r^*(\theta)\}$. Consider the map $h : N^+ \cap B \to B$ defined by $h(0,0) = (0,0)$ and for $(x, y) \ne 0$

$$h(\Pi^{-1}(x, y)) = \Pi\left(\frac{r\rho(\theta)}{r^*(\theta)}, \theta\right) \quad \text{where} \quad \rho(\theta) = \begin{cases} \sec(\theta) & \text{for } 0 \le \theta \le \tan^{-1}(\frac{1}{2}) \\ \frac{1}{2}\csc(\theta) & \text{for } \tan^{-1}(\frac{1}{2}) \le \theta \le \frac{\pi}{4} \\ \frac{1}{2}\sec(\theta) & \text{for } \frac{\pi}{4} \le \theta \le \tan^{-1}(2) \\ \csc(\theta) & \text{for } \tan^{-1}(2) \le \theta \le \frac{\pi}{2} \end{cases}.$$

Then the image $h(N^+ \cap B) = \{\Pi(r, \theta) \mid 0 \le \theta \le \pi/2, 0 \le r \le \rho(\theta)\}$ is the set B with the upper right quarter square removed as shown in Figure 8(b). By definition, h is one-to-one and hence a homeomorphism onto its image.

Case (c): B has the sign structure shown in Figure 4(c).

The verification step assures that f is monotone on all vertical lines in B and that $f(x, 0) > 0$ and $f(x, 1) < 0$. Therefore, there exists a continuous function $y^* : [0, 1] \to (0, 1)$ such that $(x, y^*(x))$ is the unique zero of $f(x, \cdot)$ in B and $N^+ \cap B = \{(x, y) \mid 0 \le y \le y^*(x)\}$. Define $h : N^+ \cap B \to B$ by

$$h(x, y) = \left(x, \frac{y}{2y^*(x)}\right).$$

Then the image $h(N^+ \cap B) = [0, 1] \times [0, \frac{1}{2}]$ is the lower half of B as shown in Figure 8(c). By definition, h is one-to-one and hence a homeomorphism onto its image.

Case (d): B has the negation of sign structure shown in Figure 4(b) rotated by 180°.

The verification step establishes that f is monotone on all horizontal and vertical lines in B. Therefore, there exists a continuous function $r^* : [0, \pi/2] \to (0, \infty)$ such that $(r^*(\theta), \theta)$ is the unique zero of $f(\Pi(\cdot, \theta))$ in B and $N^+ \cap B = \{(\Pi(r, \theta) \mid 0 \le r \le r^*(\theta)\}$. Consider the map $h : N^+ \cap B \to B$ defined by $h(0,0) = (0,0)$ and for $(x, y) \ne 0$

$$h(\Pi^{-1}(x, y)) = \left(\frac{r\rho(\theta)}{r^*(\theta)}, \theta\right) \quad \text{where} \quad \rho(\theta) = \begin{cases} \frac{1}{2}\sec(\theta) & \text{for } 0 \le \theta \le \frac{\pi}{4} \\ \frac{1}{2}\csc(\theta) & \text{for } \frac{\pi}{4} \le \theta \le \frac{\pi}{2} \end{cases}.$$
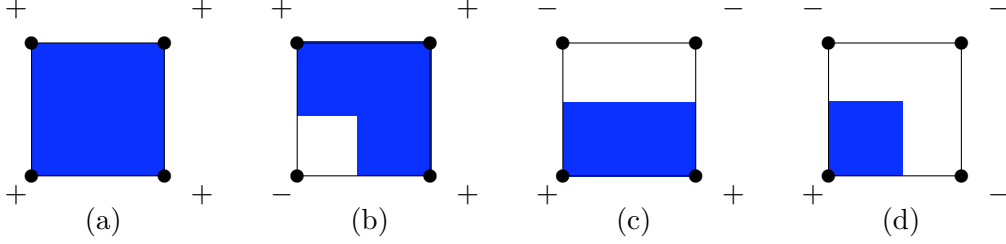
13

Figure 8: Possible cases for $\mathcal{N}_l^+ \cap B$ (shaded region), for $B \in \mathcal{G}$ with the specified sign structure.

Then the image $h(N^+ \cap B) = \{\Pi(r, \theta) \mid 0 \leq \theta \leq \pi/2, 0 \leq r \leq \rho(\theta)\}$ is the lower left quarter square in B as shown in Figure 8(d). By definition, h is one-to-one and hence a homeomorphism onto its image.

Note that in the case that the sign structure of B is all negative signs, the verification step establishes that $N^+ \cap B = \emptyset$. Moreover, the definitions of h on the internal edges of N^+ agree (after translation of the origin in each box) so that h is well-defined and one-to-one on all of N^+. We now consider the image of $h : N^+ \to \Gamma$. This image is related to $\mathcal{N}^+$ in a natural way. Let $\mathcal{N}_l^+ := \mathcal{N}^+ + (l, l)$ be the shift of $\mathcal{N}^+$ by the vector (l, l) (see Figure 8). Then $h(N^+) = \mathcal{N}_l^+ \cap \Gamma$ and $h : N^+ \to \mathcal{N}_l^+ \cap \Gamma$ is a homeomorphism. Furthermore, it is not difficult to check that $\mathcal{N}_l^+ \cap \Gamma$ is homeomorphic to $\mathcal{N}_l^+$. Since $\mathcal{N}_l^+$ is just a shift of $\mathcal{N}^+$, we now have that $\mathcal{N}^+$ is homeomorphic to N^+.

We close this section with a comment on the recursive version of the algorithm as presented at the end of Section 2. It is immediately clear that the above proof technique must be modified slightly because the monotonicity of the nodal line within each verified box in the final adaptive grid can no longer be guaranteed. Nevertheless, the result still applies to the recursive version of the algorithm if we construct approximating cubical complexes as mentioned at the end of Section 2 and as shown in Figure 6. While the resulting proof is a bit more involved technically, the fundamental arguments do not change. For the sake of clarity and brevity we have therefore included only the non-recursive case.

4   Results and comparison to probabilistic estimates

4.1   Two-dimensional level sets of a double-well potential

In this section we begin assessing the performance of our original, non-recursive verification algorithm by applying it to a simple special case. In particular, we compare our algorithm to the probabilistic result by Niyogi, Smale, and Weinberger [40] which was already mentioned briefly in the introduction. More precisely, we study the performance of the algorithm in relation to the condition number $1/\tau$ introduced in [40]. The inverse τ of the condition number is defined for a compact embedded manifold $X \subset \mathbb{R}^d$ and is the largest number such that the open normal bundle about $X \subset \mathbb{R}^d$ of radius r is embedded in $\mathbb{R}^d$ for all $r < \tau$. The results in [40] then provide a probabilistic algorithm that allows one to compute the homology of X by randomly choosing N sampling points $x_k \in X$, $k = 1, \ldots, N$, from the manifold, and then computing the homology of the union $\cup_{k=1}^N B_\epsilon(x_k)$ of ϵ-balls centered at these points. The number N depends on the inverse condition number τ, the volume of the manifold, and on the specified correctness probability for the homology computation. In addition, the radius ϵ has to be chosen sufficiently small— in particular smaller than τ. The sampling size N is the main measure for the complexity of the probabilistic algorithm in [40].
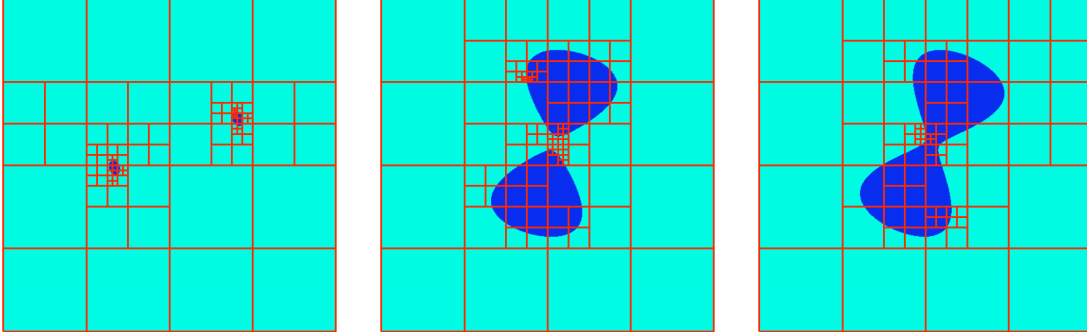
14

Figure 9: Sample images of the double-well nodal domains together with the final grid produced by the homology verification algorithm. From left to right the images correspond to C-values slightly larger than $c_0 = -1/4$, slightly less than $c_0 = 0$, and slightly larger than $c_0 = 0$, respectively. In each case we have $\gamma = |C - c_0| = 0.00625$.

At first glance, the results of [40] cannot be applied directly to the situation of nodal domains considered in the present paper, since our manifolds are full-dimensional manifolds with boundary. Thus, we concentrate on a simple special case by considering the nodal domains of the function

$$H_C(x, y) = \tfrac{1}{2}x^2 - \tfrac{1}{4}x^4 - \tfrac{1}{2}y^2 + C . \tag{10}$$

The positive nodal domain is empty for $C < -1/4$. When $-1/4 < C < 0$, the nodal domain consists of two topological disks which merge together at a single point at the origin as $C \to 0^-$ and contract to two points as $C \to -1/4^+$. For $C > 0$, the nodal domain is connected, but has a concave neck which pinches to single point at the origin as $C \to 0^+$. At both pinching events as $C \to 0^{\pm}$, the curvature of the nodal line becomes infinite. One can easily check that in this simple setting, the results in [40] still apply if one slightly modifies the definition of τ above. For this, let τ denote the largest number such that the open outward normal bundle on the boundary of X with radius r is still embedded in $\mathbb{R}^d$ for all $r < \tau$.

Using this modified definition of the manifold parameter τ, one can easily compute that for the explicit example of nodal domains of H_C we have

$$\tau(C) = \begin{cases} \sqrt{1 - \sqrt{1 + 4C}} = \dfrac{2\sqrt{|C|}}{\sqrt{1 + \sqrt{1 + 4C}}} & \text{for } -\tfrac{1}{4} \le C < 0 \\[2ex] \sqrt{2C} & \text{for } C > 0 . \end{cases}$$

For C-values close to $-1/4$ the condition number $1/\tau$ is actually close to one, even though the topology of the nodal sets changes as C crosses the threshold $-1/4$. For C-values close to 0 the condition number $1/\tau$ becomes unbounded.

To gain more insight into the performance of our adaptive verification algorithm, we apply it to the special case of the double-well potential defined above, but in a slightly modified version. Rather than considering $H_C : \mathbb{R}^2 \to \mathbb{R}$ as defined in (10), we consider the nodal sets of the composition of a scaled version of H_C and a rotation of the plane around the point

$$r_c = \left(\frac{3\sqrt{3}}{10}, \frac{2\sqrt{2}}{5} \right) \approx (0.5196, 0.5657)$$

with random angles $\theta \in [0, 2\pi)$. In other words, we consider scaled and randomly rotated versions of the nodal sets. In this way, it is possible to determine typical scalings of the central performance
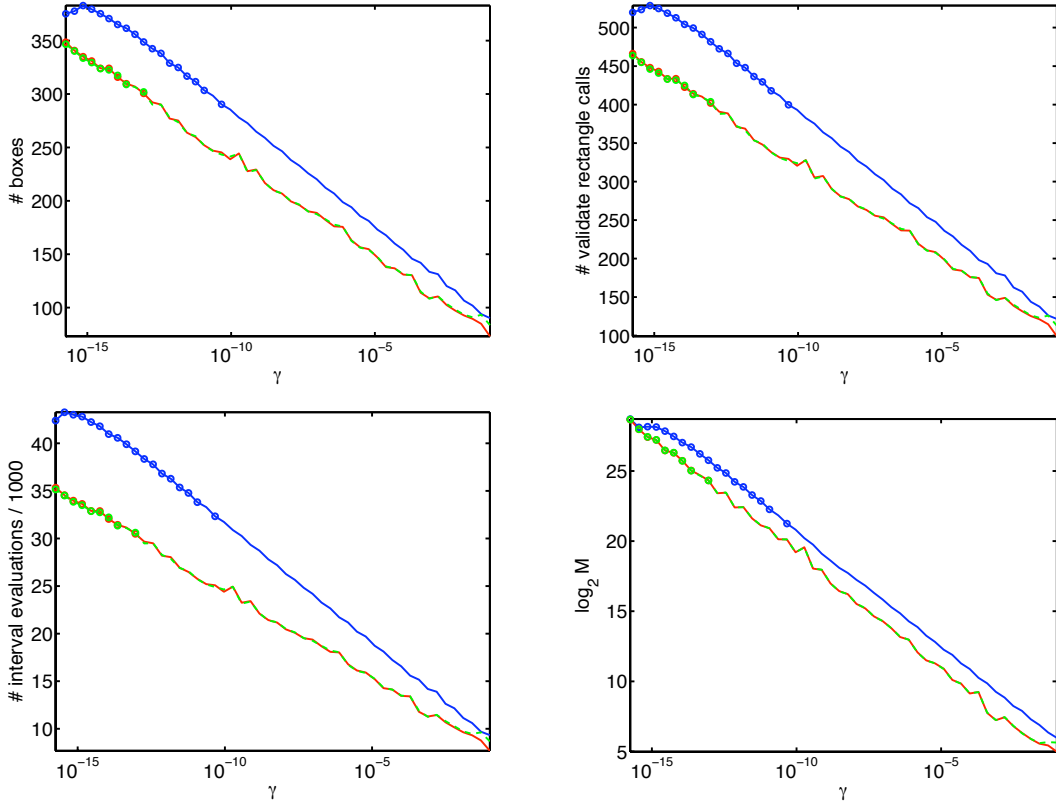
Figure 10: Dependence of averaged key performance parameters of our verification algorithm on the absolute value $\gamma = |C - c_0|$. From top left to bottom right the images show the dependence of the total number of boxes in the final adaptive grid, the number of calls to the central rectangle verification function, the number of interval computations, and the logarithm of the verification size. The solid blue, dashed green, and solid red curves correspond to values $C = -0.25^+$, $C = 0^-$, and $C = 0^+$, respectively. Circled data points are lower estimates, see the text for more details.

parameters of our algorithms, which are not affected by grid alignment issues. More precisely, we consider the θ-dependent potentials

$$H_{C,\theta}(x,y) = H_C \left(5 R_\theta^{-1} \left((x,y) - r_c\right)^t\right) , \qquad \text{with} \quad R_\theta = \left[\begin{array}{cc} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{array} \right] .$$

Some typical images of the resulting nodal domains are shown in Figure 9, together with the grids obtained from our verification algorithm.

Our simulations concentrate on the C-values at which topological changes occur, i.e., we consider the cases $C \approx -1/4$ and $C \approx 0$. At these threshold values our algorithm necessarily has to fail, and we study the performance of the algorithm for C-values close to these critical values, i.e., we consider

$$C = c_0 + c_s \gamma \quad \text{with} \quad \gamma = \frac{2^{-k}}{10} , \quad k = 0, \ldots, 49 , \quad \text{and} \quad (c_0, c_s) \in \{(-1/4, 1), (0, -1), (0, 1)\} .$$

These three possible choices for (c_0, c_s) are abbreviated by $C = -1/4^+$, $C = 0^-$, and $C = 0^+$, respectively. For each of these values, we choose 5000 angles θ from a uniform distribution on $[0, 2\pi)$ and apply the verification algorithm to the unit square $[0,1]^2$. If the algorithm verifies the topology of the nodal domains, four key performance parameters are recorded. These are the total number of boxes in the final adaptive grid, the number of calls to the central rectangle verification function, the

16

	$C = -0.25^+$	$C = 0^-$	$C = 0^+$
# boxes	$-9.410\ln\gamma + 67.4$	$-7.899\ln\gamma + 57.9$	$-7.939\ln\gamma + 57.2$
# verify rect. calls	$-13.125\ln\gamma + 88.7$	$-10.534\ln\gamma + 78.9$	$-10.587\ln\gamma + 77.9$
# interval eval./1000	$-1.089\ln\gamma + 6.5$	$-0.794\ln\gamma + 6.2$	$-0.798\ln\gamma + 6.1$
$\log_2 M$	$-0.719\ln\gamma + 4.2$	$-0.723\ln\gamma + 2.5$	$-0.722\ln\gamma + 2.5$

Table 1: Numerical least-squares fits for the data in Figure 10.

number of interval computations, and the logarithm of the verification size M. Finally, we average these parameters over all verified runs. The results of these simulations are contained in Figure 10, where the dependence of the averaged key parameters on the absolute value $\gamma = |C - c_0|$ is shown. From top left to bottom right the images show the dependence of the total number of boxes in the final adaptive grid, the number of calls to the central rectangle verification function, the number of interval computations, and the logarithm of the verification size. The solid blue, dashed green, and solid red curves correspond to values $C = -0.25^+$, $C = 0^-$, and $C = 0^+$, respectively. These results indicate that there exist affine relations between the parameters and the logarithm of γ, and least-squares fits can be found in Table 1.

The almost perfect scaling of these performance parameters is somewhat surprising, especially since the smallest value of γ is given by $1.77636 \cdot 10^{-16}$, and in fact the scaling seems to break down for γ-values close to machine precision, at least in the case $C = -1/4^+$. In order to understand this effect, notice that in our simulations we simply discard the runs which can not be verified using our algorithm, which of course biases the results. If this happens, the actual data points are only lower bounds. Data points for which not all of the 5000 runs could be verified using our algorithm are indicated as circles in Figure 10. However, it turns out that in practice, all of the 5000 simulations were verified for all γ-values larger than $5 \cdot 10^{-11}$ in the case $C = -1/4^+$, and for γ larger than 10^{-13} in the cases $C = 0^\pm$. In addition, the verification percentage remains well above 99% as long as γ is larger than $2 \cdot 10^{-13}$ for $C = -1/4^+$, and for γ larger than $5 \cdot 10^{-16}$ in the cases $C = 0^\pm$. As such, we do believe that the scalings shown in Figure 10 are correct, with the exception of γ-values close to machine precision for $C = -1/4^+$.

In order to relate our simulations to the probabilistic algorithm in [40], one has to distinguish between the cases $C = -1/4^+$ and $C = 0^\pm$. In the former case, the manifold parameter τ is close to 1, i.e., the algorithm in [40] is "well-conditioned." In fact, one can easily see that for any value $C \geq -1/4$ which is sufficiently close to $-1/4$ one only has to sample a few points from the nodal set N^+ to achieve a high correctness probability. Of course, this method does assume a priori knowledge of the location of the nodal set, which in general is not available. On the other hand, our algorithm finds the components of the nodal sets by using an adaptive grid whose size is logarithmic in the size of the actual nodal domains.

Finally, consider the case $C = 0^\pm$. In this case, one can readily see that the sample size N for the probabilistic algorithm in [40] has to be polynomial in $1/\tau$. This large growth rate is a consequence of the fact that the radius ϵ has to be smaller than the manifold parameter τ, yet one has to sample enough points to obtain a suitable cover of the whole of X and its volume converges to a positive number for $C \to 0^\pm$. On the other hand, our method still only requires an adaptive grid whose size is logarithmic in the condition number $1/\tau$, in addition to making no a priori assumption about knowledge of the nodal domains.

4.2   Random trigonometric polynomials

In previous theoretical work, Mischaikow and Wanner [34] study the question of determining the homology of nodal domains of random fields using uniform discretizations. In particular, for the

case of homogeneous Gaussian random fields, i.e., spatially periodic random Gaussian functions, they could establish lower bounds on the probability that the homology of the nodal domain is isomorphic to the homology of the cubical approximation obtained from a uniform discretization of size M. In this section, we employ our verified computational approach to assess the tightness of these theoretical bounds in two space dimensions.

The sharpness of these estimates was established in the one-dimensional case using non-rigorous computations in [14]. These one-dimensional simulations could be performed using regular numerical methods, since one can easily choose the spacing of the sampling points close enough to actually resolve all the nodal domains, and the computation of the number of components of these nodal domains is straightforward. In two dimensions, employing large discretization sizes is problematic, since they result in very long computation times for determining the homology. For this reason, we now use the verified homology algorithm described in the previous sections. More precisely, from now on, we employ the recursive version of the algorithm as described at the end of Section 2.

To investigate the sharpness of the rigorous probabilistic estimates derived in Mischaikow and Wanner [34] in two-dimensions, we concentrate on random periodic Gaussian fields as described in the introduction, see the discussion centered around equation (2). Rather than considering general random Fourier series, we consider a special class of random trigonometric polynomials on $\Gamma = [0,1]^2$ of the form

$$
\begin{aligned}
f(x,\omega) \;=\; \sum_{k,\ell=0}^{K} \alpha_k \alpha_\ell \cdot \big(& g_{k,\ell,1}(\omega)\cos(2\pi kx_1)\cos(2\pi\ell x_2) + g_{k,\ell,2}(\omega)\cos(2\pi kx_1)\sin(2\pi\ell x_2) \\
& + g_{k,\ell,3}(\omega)\sin(2\pi kx_1)\cos(2\pi\ell x_2) + g_{k,\ell,4}(\omega)\sin(2\pi kx_1)\sin(2\pi\ell x_2) \big)
\end{aligned} \tag{11}
$$

with $K \geq 3$, where $g_{k,\ell,m}$ are random variables defined over a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ which are independent and normally distributed with mean 0 and variance 1. As outlined in the introduction, these random trigonometric polynomials are covered by the theory on random Fourier series of the form (2) in [34, Theorem 3.10] if one defines $a_{k,\ell} = \alpha_k \alpha_\ell$. If we further assume that at least two of the numbers α_k are nonzero, then (3) shows that the probability of a correct homology computation for the random nodal domains $N^\pm$ via uniform cubical approximations $\mathcal{N}_M^\pm$ satisfies

$$
1 - \mathbb{P}\left\{ H_*(N^\pm) \cong H_*(\mathcal{N}_M^\pm) \right\} \;\leq\; \frac{1067\pi^2}{18M^2} \cdot \frac{\left(2A_2 A_0 + A_1^2\right)^2}{A_0^2 A_1^2} + O\left(\frac{1}{M^3}\right) , \tag{12}
$$

where

$$
A_p \;=\; \sum_{k=0}^{K} k^{2p}\alpha_k^2 .
$$

Notice that due to the particular choice $a_{k,\ell} = \alpha_k \alpha_\ell$ these constants replace the doubly indexed constants in (4). We begin by considering the case of standard random trigonometric polynomials with standard normal coefficients, i.e., we consider the choice

$$
\alpha_k = 1 \quad \text{for} \quad 1 \leq k \leq K , \quad \text{and} \quad \alpha_k = 0 \quad \text{otherwise.} \tag{13}
$$

In this case, the above formula (12) reduces to

$$
\begin{aligned}
1 - \mathbb{P}\{ H_*(N^\pm) \cong H_*(\mathcal{N}_M^\pm)\} \;&\leq\; \frac{1067\pi^2}{18M^2} \cdot \frac{1}{900} \cdot \left(46K^2 + 51K - 7\right)^2 + O\left(\frac{1}{M^3}\right) \\
&\sim\; \frac{564443\pi^2}{4050} \cdot \frac{K^4}{M^2} ,
\end{aligned}
$$

which suggests that, in order for the homology computation to be accurate with high confidence, we have to choose the discretization size M proportional to K^2 for $K \to \infty$. At first glance this result
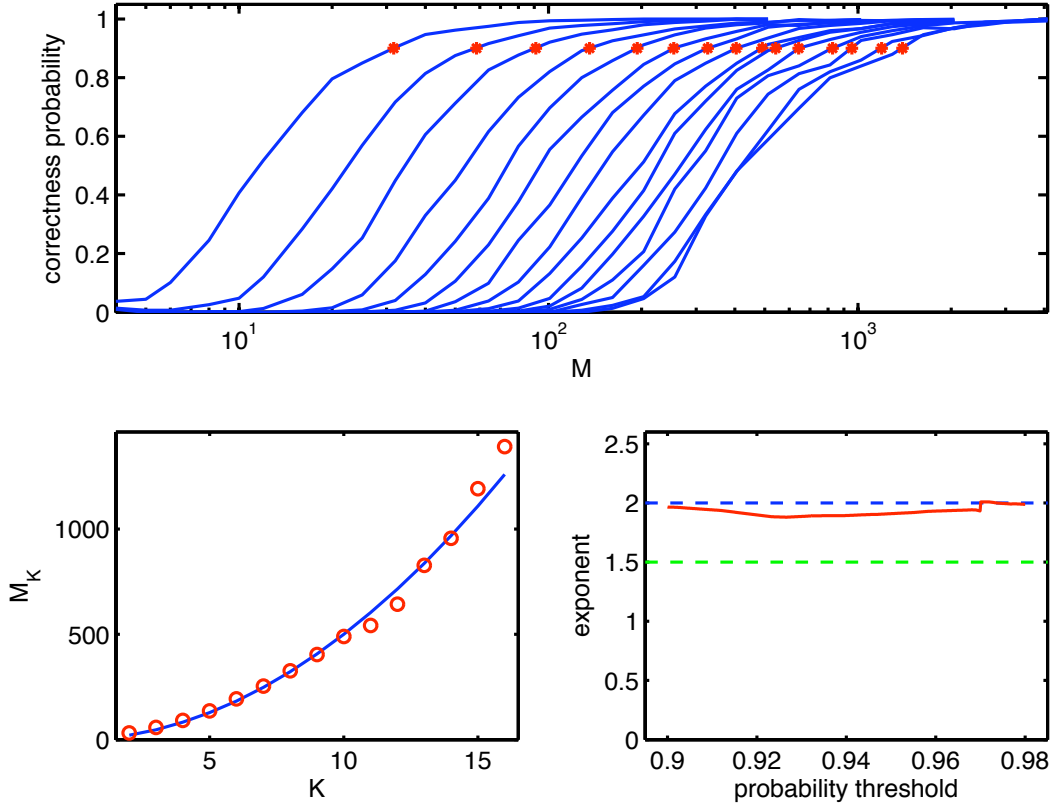
Figure 11: Numerical results for two-dimensional random trigonometric polynomials (11) satisfying (13). The top plot is the probability of computing the correct homology as a function of the discretization size M for degree $K = 2, \ldots, 16$ (from left to right). The discretization sizes M_K which give a 90% probability are marked red. The lower left figure contains these values as a function of K together with a fitted curve. The lower right figure shows the exponents α_p in the fitted relations $M_K = C_p \cdot K^{\alpha_p}$ as a function of the threshold probability p. The dashed blue line shows the predicted value of 2, the dashed green line is for the one-dimensional predicted exponent $3/2$.

does seem surprising, and it is not clear why the two-dimensional situation requires considerably finer discretizations.

In order to test this probabilistic prediction, we applied our recursive algorithm to trigonometric polynomials (11) satisfying (13) for $K = 2, \ldots, 16$. For each value of K we performed between 300 and 1000 runs of our verified homology computations (300 runs for $N = 13, \ldots, 16$, and 1000 runs for the remaining values). Within each of these runs, we then use the correct homology information to assess the correctness of regular homology computations based on sampling from equidistant grids with sizes between $M = 4$ and $M = 4096$. In this way, one obtains the probability of a correct homology computation as a function of M, for each value of K. The resulting correctness probability curves are shown in top image of Figure 11, where the curves from left to right correspond to increasing values $K = 2, \ldots, 16$. If one now specifies a desired correctness probability level $p \in (0, 1)$, one can determine for each K the value M_K which gives correctness probability p. Based on the probabilistic result in [34] one would then expect that

$$M_K = C_p \cdot K^{\alpha_p} \qquad \text{with} \qquad \alpha_p \approx 2 \quad \text{as} \quad p \to 1 \, .$$

For example, for the special case $p = 0.9$, the red dots in the top image of Figure 11 indicate the values M_K, and their behavior as a function of K is shown in the lower left image. A least-squares
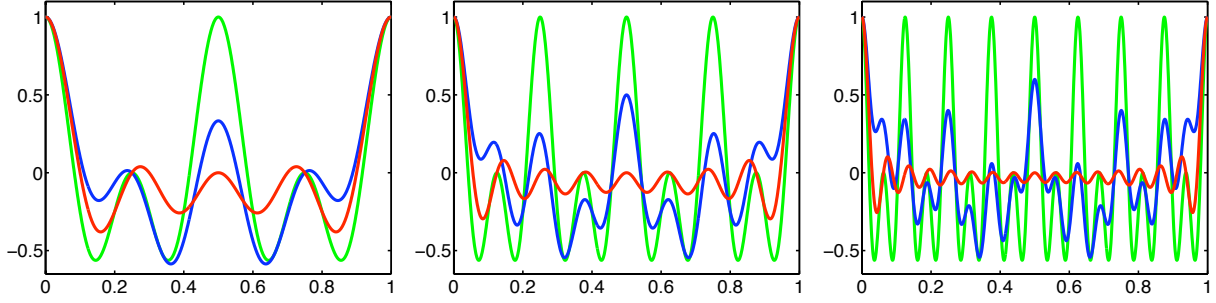
Figure 12: Normalized one-dimensional marginals $\rho(z)/\rho(0)$ of the spatial correlation functions for three classes of random fields of the form (11). From left to right the images correspond to $K = 4, 8, 16$, within each diagram the red, blue, and green curves correspond to trigonometric polynomials (11) satisfying (13), (14), and (15), respectively.

fit of the data gives the blue curve in the lower left image with values $C_p = 5.4202$ and $\alpha_p = 1.9652$. Since we are interested in results for large K, the fit is computed only from the values $K = 4, \dots, 16$. The dependence of the exponent α_p on the threshold probability p is shown in the lower right image of Figure 11, which indicates that $\alpha_p \approx 2$ for p close to 1.

The above results indicate that asymptotically, the results in [34] do indeed provide the correct scaling for the discretization size M as a function of the degree K of the random trigonometric polynomial (11). There are, however, situations in which these results are an overestimation. To illustrate this, we now consider two more classes of random trigonometric polynomials (11). For the first class, we assume that

$$\alpha_{2^\ell} = 1 \quad \text{for} \quad 1 \le 2^\ell \le K , \quad \alpha_K = 1 , \quad \text{and} \quad \alpha_k = 0 \quad \text{otherwise,} \tag{14}$$

while for the second class we assume

$$\alpha_{2^\ell} = 1 \quad \text{for} \quad \ell = \left\lfloor \log_2 \frac{2K - 1}{4} \right\rfloor , \quad \alpha_K = 1 , \quad \text{and} \quad \alpha_k = 0 \quad \text{otherwise,} \tag{15}$$

i.e., only the K-th and the 2^ℓ-th coefficients are non-zero, where 2^ℓ denotes the largest power of 2 strictly less than K. One can easily show that for both of these choices of the coefficients α_k the estimate (12) remains qualitatively unchanged, i.e., we still have to choose the discretization size M proportional to K^2 for $K \to \infty$. However, if one repeats the above simulations for the two new classes of trigonometric polynomials, different scalings are obtained. For the case (14) one obtains a growth rate $M_K \sim K^{1.73}$, whereas for the case (15) one observes linear growth $M_K \sim K^{1.00}$. Thus, while the probabilistic estimates of [34] clearly are suboptimal in both of these cases, our adaptive numerical method takes advantage of the reduced necessary discretization size.

The discrepancy between the uniform grid size suggested by the probabilistic estimates and the grid size required for our numerical approach can be explained with the help of the spatial correlation function of the random trigonometric polynomials (11). Due to the special form of these random fields, their spatial correlation function is explicitly given by

$$R(x, y) = \mathbb{E}f(x)f(y) = \sum_{k,\ell=0}^{K} \alpha_k^2 \alpha_\ell^2 \cdot \cos\left(2\pi k \left(x_1 - y_1\right)\right) \cdot \cos\left(2\pi \ell \left(x_2 - y_2\right)\right) ,$$

i.e., we have

$$R(x, y) = \rho\left(x_1 - y_1\right) \cdot \rho\left(x_2 - y_2\right) , \quad \text{where} \quad \rho(z) = \sum_{k=0}^{K} \alpha_k^2 \cdot \cos\left(2\pi k z\right) . \tag{16}$$
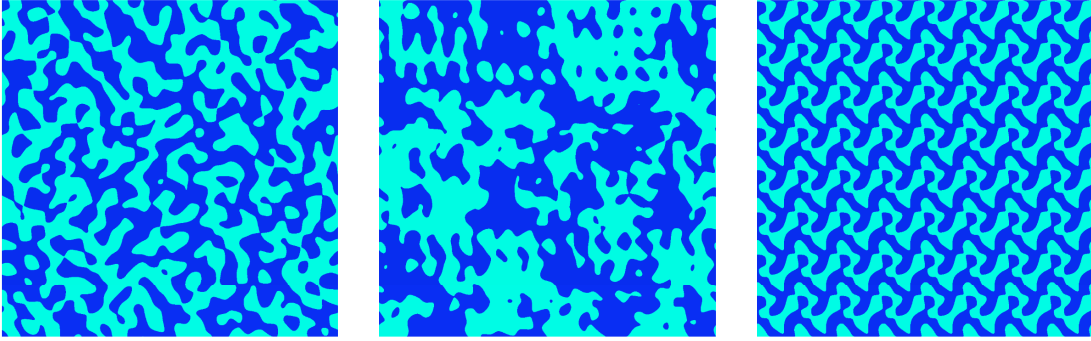
20

Figure 13: Nodal domains of random trigonometric polynomials (11) for $K = 16$. From left to right the images correspond to trigonometric polynomials satisfying (13), (14), and (15), respectively.

The spatial correlation function is the central tool for assessing whether the function values $f(x)$ and $f(y)$ of the random field f at points x in y behave independently or not. To see this, note that

$$\mathbb{E}\,|f(x) - f(y)|^2 = R(x,x) - 2R(x,y) + R(y,y) = 2\rho(0)^2 \cdot \left(1 - \frac{\rho(x_1 - y_1)}{\rho(0)} \cdot \frac{\rho(x_2 - y_2)}{\rho(0)}\right),$$

as well as

$$-1 \leq \frac{\rho(z)}{\rho(0)} \leq 1 \qquad \text{for all} \qquad z \in \mathbb{R}\,.$$

These formulas show that if x and y are chosen in such a way that $\rho(x_1 - y_1) = \rho(x_2 - y_2) = \rho(0)$, then we have $f(x) = f(y)$ almost surely. Furthermore, even if the value of $\rho(x_1 - y_1) \cdot \rho(x_2 - y_2)$ is only sufficiently close to $\rho(0)^2$, the above formulas still imply that $f(x) \approx f(y)$ with high probability. Similarly, one can show that if $\rho(x_1 - y_1) \cdot \rho(x_2 - y_2)$ is close to $-\rho(0)^2$, then we have $f(x) \approx -f(y)$ with high probability. Finally, one can show that if $\rho(x_1 - y_1) \cdot \rho(x_2 - y_2)$ is close to 0, the behavior of the function values $f(x)$ and $f(y)$ is basically independent.

The effects of spatial correlations on the geometry of nodal domains can be seen in Figures 12 and 13. The first of these figures contains sample plots of the normalized one-dimensional marginals $\rho(z)/\rho(0)$ of the spatial correlation functions for the three classes of random fields considered above. These plots indicate that for random trigonometric polynomials satisfying (13) the function values $f(x)$ and $f(y)$ are basically uncorrelated, unless of course x is close to y. A sample resulting nodal domain patterns is shown in the left-most image of Figure 13. However, for random trigonometric polynomials satisfying (15), the marginal $\rho(z)/\rho(0)$ attains the value 1 periodically throughout the domain, which leads to periodic nodal domain patterns, see the right-most image in Figure 13. In contrast, random trigonometric polynomials satisfying (14) exhibit a normalized marginal which does attain values far away from zero on significant parts of the domain, and this effect seems to increase with increasing K. (It can be shown that if z is a real number with finite dyadic representation, then $\rho(z)/\rho(0) \to 1$ for $K \to \infty$.) In fact, as one can see from the center image in Figure 13, the resulting nodal domains are less "random" than the left-most image.

How can these observations be used to explain the numerical results from above? For this one has to take a closer look at the proof techniques used in [34], which are fundamentally local in nature. By introducing a suitable notion of admissibility for square subdomains of Γ, it is first shown that if all the basic cubes determined by the discretization grid points are admissible, then the homology of the nodal domains $N^\pm$ is isomorphic to the homology of the cubical approximations $\mathcal{N}_M^\pm$. The probabilistic part of the main results in [34] then consists in deriving upper bounds on the probability that a square of small side length is not admissible. These local probability estimates are indeed sharp, as follows a posteriori from the first simulation presented in this section.

However, in order to estimate the probability of a correct homology computation for the complete domain Γ, the results in [34] simply add all these local probability estimates. This approach can only be expected to lead to sharp estimates if the admissibility of different subsquares are independent events, i.e., if the spatial correlation function is close to zero for $x \neq y$. Clearly, this is only the case if the coefficients α_k satisfy (13).

4.3 The stochastic Cahn-Hilliard model

One of the main motivations for our results is the study of deterministic or stochastic evolution equations. As an example, consider the Cahn-Hilliard-Cook model (1). This stochastic partial differential equation has been proposed as a model for phase separation in metallic alloys and produces complicated patterns, see for example [6, 8, 10, 13, 45] and the references therein. As we mentioned in the introduction, computational homology can be used to quantify these complicated structures [21], and the question of choosing the correct discretization size M for the homology computations is of utmost importance. Notice that if we are interested in the evolution of (1) originating at a random field, then for any time $t > 0$ the solution $u(t, \cdot)$ is a random field over Γ. In general, however, the coefficients in the Fourier expansion of this random field will be neither Gaussian nor independent. An important special case where these properties are realized is the linearized Cahn-Hilliard-Cook model

$$\frac{\partial u}{\partial t} = -\Delta \left(\epsilon^2 \Delta u - F''(\bar{u})u \right) + \sigma \cdot \xi \quad \text{in} \quad \Gamma \subset \mathbb{R}^d , \tag{17}$$

provided the random initial condition satisfies the assumptions of Theorem 2.7 in [34]. In (17), the function $\bar{u}$ denotes a given spatially homogeneous equilibrium solution of the deterministic Cahn-Hilliard model.

In the remainder of this section, we will apply our rigorous computational techniques both in a one- and a two-dimensional setting. Thereby, we are not only able to assess the sharpness of the probabilistic results in [34] for case of the linearized Cahn-Hilliard model (17), but we also study the effect of non-Gaussianity in the nonlinear case. In addition, we address the question of homology accuracy as a function of time, since it is well-known that whenever the nodal lines of a function exhibit singularities, homology computations via discretizations will introduce errors — regardless of the discretization size. In all of our studies below, we consider the Cahn-Hilliard-Cook model (1) or its linearization (17) for the classical choice of double-well potential $F(u) = (u^2 - 1)^2/4$, as well as $\bar{u} = 0$, i.e., we have $-F'(u) = u - u^3$ in (1) and $-F''(\bar{u}) = 1$ in (17).

4.3.1 The one-dimensional case

We begin our study of the Cahn-Hilliard model by considering the deterministic one-dimensional setting, i.e., we consider $\Gamma = [0, 1]$ and $\sigma = 0$ in (1) and (17). Specifically, our goal is to assess the sharpness of the results in [34] for the linearized equation, and to determine how well they apply in the nonlinear, and therefore non-Gaussian, situation. All of our simulations use a one-dimensional version of the rigorous techniques described in Sections 2 and 3, as well as periodic boundary conditions. For the linearized equation in this setting, we assume further that the initial condition $u(0, \cdot)$ is a random periodic Gaussian field of degree $K = K_\epsilon$. Then for every $t > 0$ the solution of (17) with $\sigma = 0$ is explicitly given by

$$u(t, x, \omega) = \sum_{k=1}^{K_\epsilon} e^{\lambda_k t} \cdot \left(g_{2k}(\omega) \cdot \cos(2k\pi x) + g_{2k-1}(\omega) \cdot \sin(2k\pi x) \right) ,$$

where $\lambda_k = 4\pi^2 k^2 \cdot (1 - 4\pi^2 k^2 \epsilon^2)$ denotes the k-th eigenvalue of the linearized Cahn-Hilliard operator. Choosing $K_\epsilon = \lfloor r/(2\pi\epsilon) \rfloor$, for some fixed $r > 1$ guarantees that for every small $\epsilon > 0$ the initial
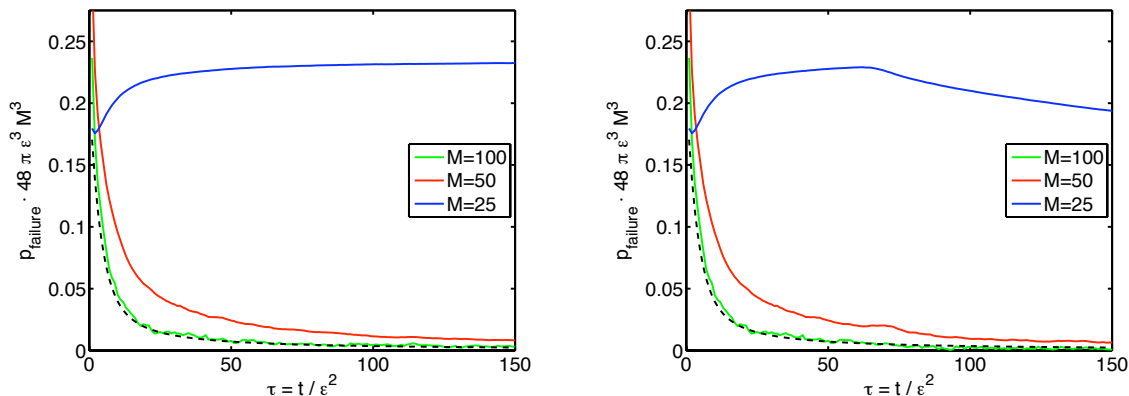
Figure 14: Verified numerical results for the probability p_{failure} of a false homology computation in the linearized Cahn-Hilliard model (left) and the nonlinear Cahn-Hilliard model (right). From top to bottom the solid lines correspond to $M = 25, 50, 100$, and the dashed line shows the function $I_\epsilon(\tau)$ from the probabilistic estimate. All curves have been scaled by the factor $48\pi\epsilon^3 M^3$.

condition contains all unstable modes, i.e., all modes which are responsible for the formation of the complicated patterns.

The explicit representation of the solution u of (17) shows that the results for random Fourier series in [34, Theorem 2.7] are readily applicable. In fact, if we denote the nodal domains of $u(t, \cdot)$ by $N^\pm(t)$ and their cubical approximations by $\mathcal{N}_M^\pm(t)$, then the probability for a correct homology computation is bounded by

$$\mathbb{P}\left\{H_*(N^\pm(t)) = H_*(\mathcal{N}_M^\pm(t))\right\} \quad \geq \quad 1 - \frac{1}{48\pi\epsilon^3 M^2} \cdot I_\epsilon\left(t/\epsilon^2\right) + O\left(\frac{1}{M^3}\right) .$$

In this estimate, the function $I_\epsilon(\tau)$ is defined by

$$I_\epsilon(\tau) = \frac{S_{0,\epsilon}(\tau)S_{2,\epsilon}(\tau) - S_{1,\epsilon}(\tau)^2}{S_{0,\epsilon}(\tau)^{3/2}S_{1,\epsilon}(\tau)^{1/2}} ,$$

where

$$S_{\ell,\epsilon}(\tau) = \sum_{k=1}^{K_\epsilon} (2\pi\epsilon)^{2\ell+1} k^{2\ell} e^{2\epsilon^2 \lambda_k \tau} \quad \xrightarrow{\epsilon \to 0} \quad \int_0^r s^{2\ell} e^{2\tau s^2(1-s^2)} \, ds .$$

Notice that the last limit shows that for every fixed $\tau > 0$ the value $I_\epsilon(\tau)$ converges as $\epsilon \to 0$, say to a limit function $I_0(\tau)$.

The above estimate implies that in order to compute the homology of the nodal domains correctly with high probability, we have to choose $M \sim \epsilon^{-3/2}$. This is in accordance with the fact that the observed patterns exhibit a typical thickness which is proportional to ϵ as $\epsilon \to 0$. In addition, the form of the above probability estimate automatically yields the correct time scaling for the phenomenon of spinodal decomposition: for small $\epsilon > 0$, one expects the time frame to be proportional to ϵ^2, which is of course reflected by the fact that the t-dependent prefactor in the probability estimate is basically given by $I_0(t/\epsilon^2)$.

In order to demonstrate the accuracy of these probabilistic predictions we performed the following computations. For $\epsilon = 0.005$ and three values of M, we computed the actual probability that the first discretization interval $[0, 1/M]$ contains more than one zero, i.e., that one cannot determine the correct topology of the nodal domains from the function values at 0 and $1/M$. According to the above discussion, this probability should asymptotically be given by $I_\epsilon(t/\epsilon^2)/(48\pi\epsilon^3 M^3)$ for large

23

values of M, and this is confirmed in the left graph of Figure 14. In fact, the asymptotic behavior predicted by [34, Theorem 2.7] is realized almost exactly for the discretization size $M = 100$, and at least qualitatively for $M = 50$. The significantly different behavior of the curve for $M = 25$ can be explained as follows. Using the results in [16, 17] one can easily show that the expected value $\mathbb{E}Z(t)$ of the number of zeros $Z(t, \omega)$ of the function $u(t, \cdot, \omega)$ is given by

$$\mathbb{E}Z(t) = 2 \cdot \left(\sum_{k=1}^{K_\epsilon} k^2 e^{2\lambda_k t} \right)^{1/2} \cdot \left(\sum_{k=1}^{K_\epsilon} e^{2\lambda_k t} \right)^{-1/2}.$$

The graph of $\mathbb{E}Z(t)$ is qualitatively similar to the $M = 25$ curve in Figure 14; after an initial decrease to a minimal value of 41.18 at $t/\epsilon^2 \approx 2.4$, the graph increases again and limits to 44.94 as $t \to \infty$. Thus, the probability that the interval $[0, 1/M]$ for $M = 25$ contains more than two zeros is fairly large. In fact, the computations show that for $t/\epsilon^2 \approx 2.4$ this probability is 59.6%, while for large t it stabilizes at 79.0%.

The right graph in Figure 14 contains analogous numerical results for the nonlinear Cahn-Hilliard equation (1), again in the deterministic situation with $\sigma = 0$. Notice that now the curve for $M = 25$ exhibits a marked decay starting at around $t/\epsilon^2 \approx 70$, and the remaining curves show similar, although not as pronounced, behavior. On the other hand, for times $t \le 70\epsilon^2$ the curves in both graphs are indistinguishable, despite the fact that they were obtained from a linear and a nonlinear model, respectively. Recent theoretical work has shown that in fact during the initial phase separation regime of the Cahn-Hilliard equation the effects of the nonlinearity are suppressed for an unexpectedly long time [6, 8, 42, 43, 45]. These results have established rigorous lower bounds on the duration of the linear regime. In contrast, our results provide an upper bound on the onset of nonlinear behavior in the Cahn-Hilliard equation, and complement our findings in [21]. We would like to point out, however, that even during the early stages of phase separation, the Gaussianity of the functions $u(t, \cdot)$ is lost, and it is remarkable that the results of [34] still correctly predict the homology correctness probability.

4.3.2 The two-dimensional case

In this final section of the paper we turn our attention to the two-dimensional nonlinear Cahn-Hilliard-Cook model (1) on the square domain $\Gamma = [0, 1]^2$. Our goals are two-fold. On the one hand, we want to assess the efficiency of our rigorous computational algorithm and describe the variations in key parameters as one follows the evolution of a sample solution to (1). On the other hand, we would like to obtain probabilistic information based on random ensembles of initial conditions which sheds light on the correctness of homology computations as a function of the used grid size and of time. These latter results are motivated by [21].

Throughout this section, we consider (1) on the square domain $\Gamma = [0, 1]^2$ subject to homogeneous Neumann boundary conditions and with $F(u) = (u^2 - 1)^2/4$. The interaction length is chosen as $\epsilon = 0.025$, and we simulate solutions originating at random perturbations of the unstable homogeneous state $\bar{u} \equiv 0$. These random perturbations have vanishing total mass and maximum norm equal to 10^{-4}. The solutions are computed on a time interval $[0, t_e]$ that is chosen as in [21], which covers both the spinodal decomposition process and the beginning of the coarsening regime. For more background information we refer the reader to [6, 7, 8, 9, 10, 13, 29, 30, 31, 42, 43, 45].

In order to assess the effects of noise, we consider both the deterministic equation with $\sigma = 0$ and the stochastic version with $\sigma = 0.025$. In the latter case, the noise process $\xi = \dot{W}$ is given as the generalized derivative of a Q-Wiener process W which is cut-off noise on the standard cosine basis functions used in the spectral method for simulating (1). In other words, the noise process is white in time and colored in space.
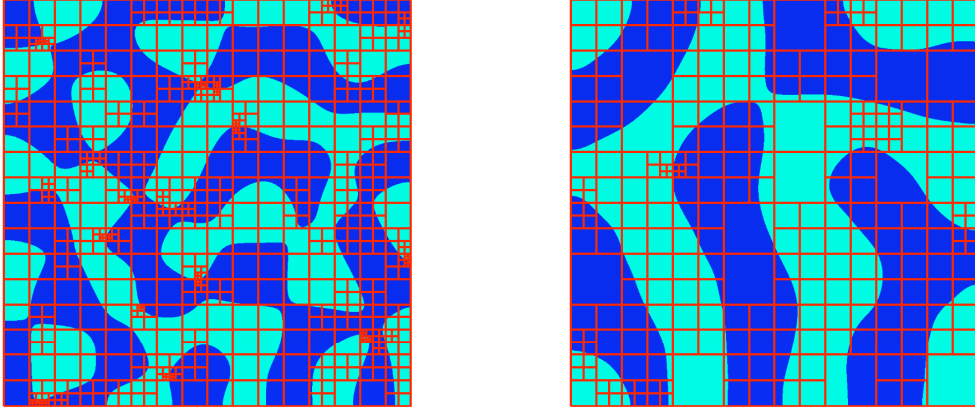
Figure 15: Sample nodal domains of a solution to the Cahn-Hilliard equation (1) with $\epsilon = 0.025$ and $\sigma = 0$. The left image shows the pattern at time $t = 0.1t_e$, the right image at time $t = 0.9t_e$. Both images also contain the adaptive grid produced by our recursive verification algorithm.

As a first test of the recursive verification algorithm described in Sections 2 and 3 we consider one solution path each for the deterministic Cahn-Hilliard equation and the stochastic Cahn-Hilliard-Cook model, with parameters as described above. As was mentioned earlier, verifying the nodal domains of evolving patterns can fail. Every time the nodal lines of the phase function $u(t, \cdot)$ exhibit singularities, no homology computation using a finite size grid will be able to determine the correct homology. See for example [25, Figure 8.1]. In the context of evolving patterns such singularities in the nodal lines occur every time the topology of the nodal sets changes. While one would expect that these times form a set of measure zero, any verification with a maximal resolution should fail on an open set of times. Thus, we do expect that at least on some noticeable portion of the underlying time interval our algorithm will fail.

In order to determine the actual size of these failure intervals, we run the verification code for the solution snapshots at times $t = k \cdot t_e/1000$, where $k = 1, \ldots, 1000$. The somewhat surprising results of these simulations are shown in Figure 16. It turns out that in all of the 1000 solution snapshots the algorithm was able to verify the nodal domains, i.e., the set of times where verification fails is extremely small. Furthermore, key parameters of the algorithm change in qualitative agreement with the underlying topology of the nodal domains. This can be seen in the diagrams of Figure 16, which depict the evolution of the total number of boxes in the final adaptive grid (top left), the number of calls to the central rectangle verification function (top right), and the number of interval computations (bottom left). In these images, the blue curves correspond to the deterministic model, the red curves to the stochastic one. It is evident that all of these parameters vary qualitatively in the same way. For comparison, the lower right image shows the evolution of the ℓ^1-norm of the Betti number vector $(\beta_0^+(t), \beta_0^-(t), \beta_1^+(t), \beta_1^-(t))$, where $\beta_k^\pm(t)$ denotes the k-th Betti number of the nodal domain $N^\pm(t)$. While the agreement is not precise, the evolution of the algorithm parameters qualitatively follows the topology evolution. Notice in particular that while the deterministic evolution shows the nonmonotone behavior described in [21], the stochastic evolution decays more or less monotonically.

Another quantity that is of importance for our algorithm is the verification size M, which is defined as the inverse of the side length of the smallest rectangle in the adaptive verifying grid. The verification size changes much more irregularly than the parameters shown in Figure 16, which can be explained by taking another look at Figure 15. In both the patterns and the resulting grids shown there, one can readily see regions with extremely fine adaptive grid, while the nodal domain structure is nowhere near a topology change. This is an artifact of our dyadic subdivision
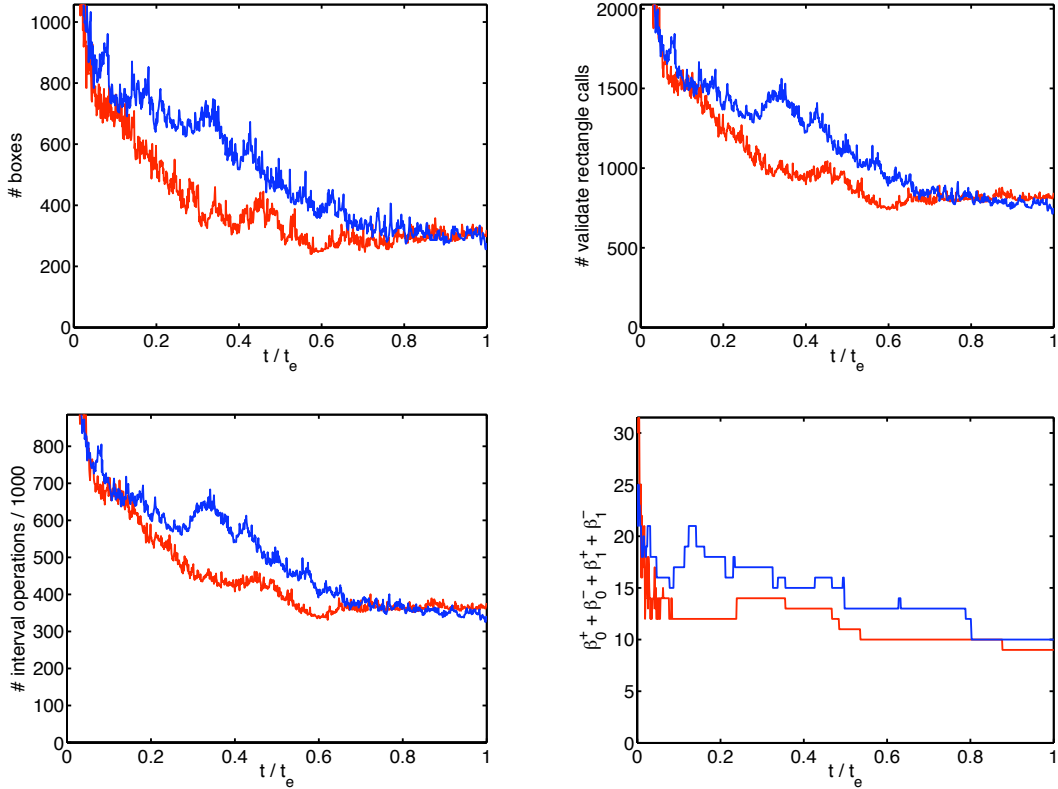
Figure 16: Evolution of key parameters of our verification algorithm as a function of time, both for a sample path of the deterministic (blue curves) and of the stochastic (red curves) Cahn-Hilliard model. From top left to bottom right the images show the evolution of the total number of boxes in the final adaptive grid, the number of calls to the central rectangle verification function, the number of interval computations, as well as the ℓ^1-norm of the Betti number vector $(\beta_0^+, \beta_0^-, \beta_1^+, \beta_1^-)$.

method described in Section 2. All of the line segments determining the final grid are aligned with the dyadic points in the base domain Γ. Thus, if the nodal lines get extremely close to points with dyadic components with a large denominator, then the algorithm has to refine significantly in that region. One easy possibility to address this issue is to not only start the binary subdivision algorithm from the square base domain Γ, but also from subdivided versions, where the domain is first subdivided into M_0^2 squares of equal size and the algorithm is then run on each of these smaller squares.

In order to describe "typical" behavior, we studied ensembles of 500 random initial conditions for each of these models. We begin by considering verification sizes $M(M_0)$ with initial discretization sizes $M_0 = 1, 3, 5, 7$. Assume we only allow for a maximal verification size $M_{\max}$, based for example on computational restrictions. How likely is it then that one cannot verify the nodal domain geometry at a given point in time with a verification size at most $M_{\max}$? We computed these failure probabilities from the ensembles of solution paths described above, in two scenarios. The results are shown in Figure 17, both for maximal verification size $M_{\max} = 500$ (top row) and for $M_{\max} = 1000$ (bottom row). For initial discretization sizes $M_0 = 1, 3, 5, 7$ we tried verification of the M_0^2 subsquares using the dyadic subdivision algorithm, i.e., we rejected the verification if the resulting verification size $M(M_0)$ satisfied $M(M_0) > M_{\max}$, where we use the convention $M(M_0) = \infty$ if our algorithm did not verify at all. These curves are shown in blue, green, cyan, and magenta for $M_0 = 1, 3, 5, 7$, respectively. Finally, we combined all of these four verification steps by letting M denote the smallest of these verification sizes, and rejecting the verification
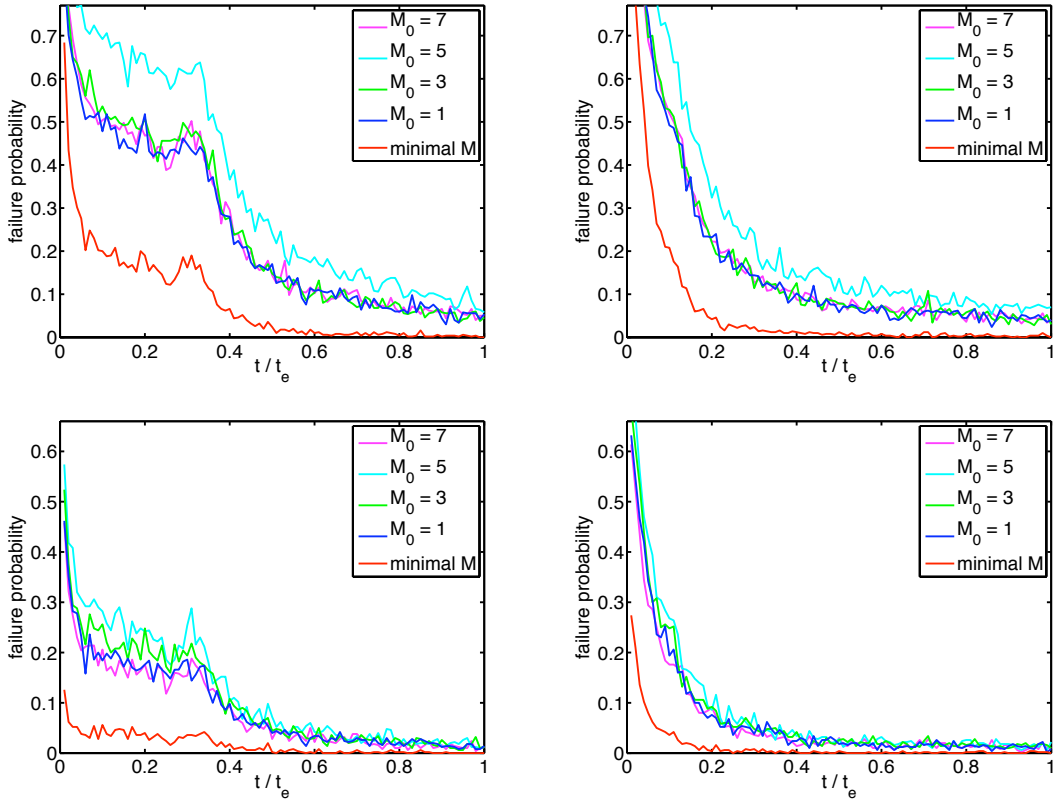
26

Figure 17: Temporal evolution of the probability of verification failure with verification size at most $M_{\max} = 500$ (top row) and $M_{\max} = 1000$ (bottom row). The left column shows the curves for the deterministic model, the right column corresponds to the stochastic one. In each case, the top four curves are obtained using initial subdivisions of size M_0^2 with $M_0 = 1, 3, 5, 7$, as described in the text. The red curve is obtained by using the smallest of these verification sizes.

if $M > M_{\max}$.

The results in Figure 17 show that in general, significant improvements can be achieved if one runs the algorithm for a variety of different values of M_0, and then uses the smallest possible verification size. The computational overhead for this procedure is small. It can also be seen that in the Cahn-Hilliard models, significant failure probabilities only occur at the beginning of the phase separation process. Especially during the later coarsening regime the failure probabilities are basically zero. Finally, the results of Figure 17 demonstrate a surprising effect of noise. While in the very beginning of the evolution the added noise increases the failure probability, during the subsequent stages of spinodal decomposition, i.e., for times roughly between $0.2t_e$ and $0.4t_e$, the noise actually leads to a decrease in these probabilities. This is due to the fact that noise accelerates the phase separation and that the complexity of our algorithm is mostly dependent on the pattern complexity. For more details see [6, 7, 8].

So far, we have only addressed the size necessary for verifying the geometry of the nodal domains using interval arithmetic. But in many cases one would expect that computing the homology using a smaller grid size, say using the $(M_h + 1)^2$ equidistant sampling points $(k/M_h, \ell/M_h)$, with $k, \ell = 0, \ldots, M_h$, in the square Γ should already provide the correct Betti numbers. This can certainly be seen from the sample images in Figure 15. To study this question in more detail, we use again the ensemble simulations from above, both for the deterministic and the stochastic Cahn-Hilliard model. Recall that for these simulations we randomly choose initial conditions u_0 close to the homogeneous state $\bar{u} \equiv 0$. Then if $u(t, x; u_0)$ denotes the solution of (1), we consider
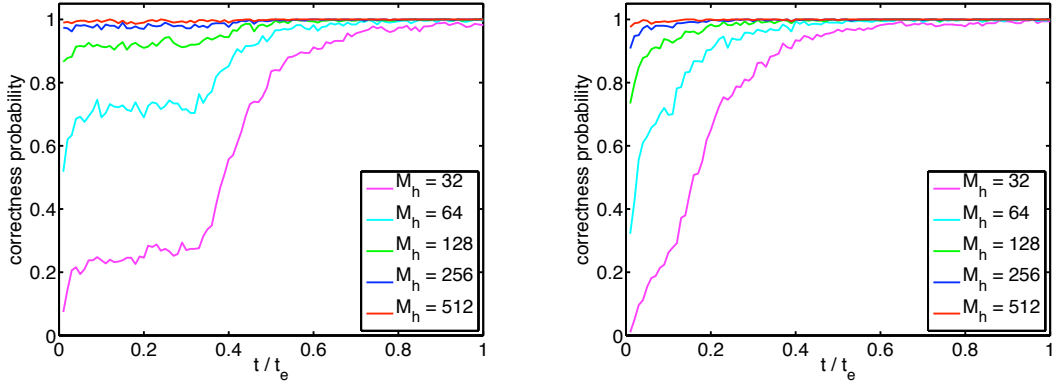
27

Figure 18: Temporal evolution of the probability of the event $\mathcal{E}(t, M_h, M_{\max})$ defined in (18), for the choice $M_{\max} = 50000$. The left image shows the curves for the deterministic Cahn-Hilliard model, the right image for the stochastic Cahn-Hilliard-Cook equation. In both cases, the magenta, cyan, green, blue, and red curves correspond to $M_h = 32, 64, 128, 256, 512$, respectively.

the event

$$\mathcal{E}(t, M_h, M_{\max}) \;\; = \;\; \{u_0 \; : \;\; \text{verification with size } M \leq M_{\max} \text{ is possible, and the Betti}$$
$$\text{numbers are correctly computed via the } M_h\text{-grid}\} \;, \qquad (18)$$

which consists of all initial conditions for which at time t the verification algorithm terminates for $u(t, \cdot; u_0)$ with verification size $M \leq M_{\max}$, and for which the Betti numbers of the nodal domains $N^\pm(t; u_0)$ of $u(t, \cdot; u_0)$ computed via the above sampling by $(M_h + 1)^2$ points agree with the verified Betti numbers. For our implementation, we use the minimal verification size M among the $M(M_0)$ with $M_0 = 1, 3, 5, 7$ as before. The probabilities of these events computed from ensembles of size 500 and $M_{\max} = 50000$ are shown in Figure 18. The left image shows the curves for the deterministic Cahn-Hilliard model, the right image for the stochastic Cahn-Hilliard-Cook equation. In both cases, the magenta, cyan, green, blue, and red curves correspond to $M_h = 32, 64, 128, 256, 512$, respectively. These computations show that already for $M_h = 512$, the likelihood of obtaining the correct homology information is close to one for all times t, both for the deterministic and the stochastic case. We would also like to mention that for the above choice of $M_{\max} = 50000$, all of the solution snapshots were able to be verified with $M \leq M_{\max}$.

We close this section by briefly commenting on the actual averaged Betti number evolution curves for the Cahn-Hilliard model (1) with $\epsilon = 0.025$. This ϵ-value is significantly larger than the ones used in [21], and was chosen specifically to speed up our simulations. Nevertheless, it was pointed out in [21] that the averaged Betti number evolution curves scale with ϵ^2, since changing ϵ can be thought of as rescaling the underlying domain Γ. Thus, our choice of ϵ corresponds to a very small domain, where the Betti numbers usually stay in the single digits. It was conjectured in [21] that upon averaging, even on such small domains one can still distinguish between the nonmonotone evolution of the averaged Betti number curves for the deterministic Cahn-Hilliard model, and the monotone decay resulting from the addition of noise. The curves shown in Figure 19 confirm this.
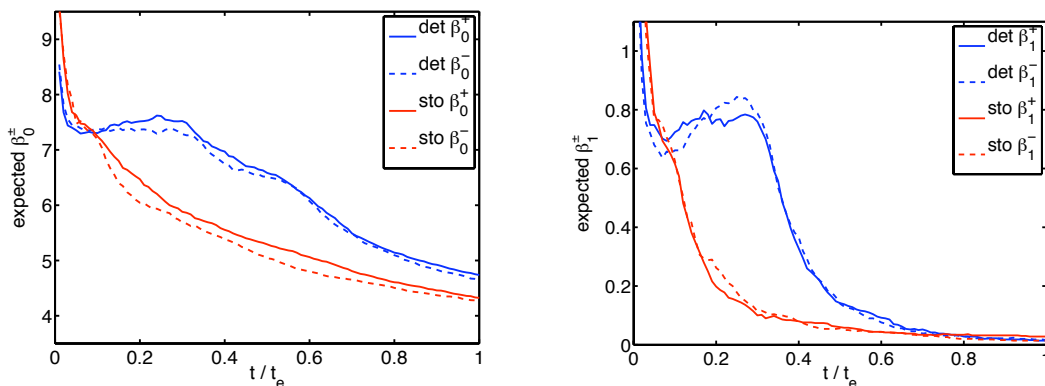
Acknowledgements

Figure 19: Averaged Betti number evolutions for the deterministic Cahn-Hilliard model with $\sigma = 0$ (blue curves) and the stochastic Cahn-Hilliard-Cook with $\sigma = 0.025$ equation (red curves), in each case for $\epsilon = 0.025$. The left image shows the curves for $\beta_0^{\pm}$, the right image is for $\beta_1^{\pm}$; in either image the solid curve represents β_k^+, while the dashed curve shows β_k^-.

References

[1] R. J. Adler. *The Geometry of Random Fields*. John Wiley & Sons Ltd., Chichester, 1981.

[2] R. J. Adler and J. E. Taylor. *Random Fields and Geometry*. Springer-Verlag, New York, 2007.

[3] N. Amenta and M. Bern. Surface reconstruction by Voronoi filtering. *Discrete & Computational Geometry*, 22(4):481–504, 1999.

[4] N. Amenta, S. Choi, T. K. Dey, and N. Leekha. A simple algorithm for homeomorphic surface reconstruction. *International Journal of Computational Geometry & Applications*, 12(1-2):125–141, 2002.

[5] C. H. Arns, M. A. Knackstedt, W. V. Pinczewski, and K. R. Mecke. Euler-Poincare characteristics of classes of disordered media. *Physical Review E*, 63(3):031112, 2001.

[6] D. Blömker, S. Maier-Paape, and T. Wanner. Spinodal decomposition for the Cahn-Hilliard-Cook equation. *Communications in Mathematical Physics*, 223(3):553–582, 2001.

[7] D. Blömker, S. Maier-Paape, and T. Wanner. Phase separation in stochastic Cahn-Hilliard models. In A. Miranville, editor, *Mathematical Methods and Models in Phase Transitions*, pages 1–41. Nova Science Publishers, New York, 2005.

[8] D. Blömker, S. Maier-Paape, and T. Wanner. Second phase spinodal decomposition for the Cahn-Hilliard-Cook equation. *Transactions of the American Mathematical Society*, 360(1):449–489, 2008.

[9] J. W. Cahn. Free energy of a nonuniform system. II. Thermodynamic basis. *Journal of Chemical Physics*, 30:1121–1124, 1959.

[10] J. W. Cahn and J. E. Hilliard. Free energy of a nonuniform system I. Interfacial free energy. *Journal of Chemical Physics*, 28:258–267, 1958.

[11] A. Cerezo, M. G. Hetherington, J. M. Hyde, and M. K. Miller. A topological approach to materials characterisation. *Scripta Metallurgica et Materialia*, 25:1435–1440, 1991.

[12] L.-Q. Chen. Phase-field models of microstructure evolution. *Annual Review of Materials Research*, 32:113–140, 2002.

[13] H. Cook. Brownian motion in spinodal decomposition. *Acta Metallurgica*, 18:297–306, 1970.

[14] S. Day, W. D. Kalies, K. Mischaikow, and T. Wanner. Probabilistic and numerical validation of homology computations for nodal domains. *Electron. Res. Announc. Amer. Math. Soc.*, 13:60–73 (electronic), 2007.

[15] T. K. Dey, H. Edelsbrunner, and S. Guha. Computational topology. In *Advances in discrete and computational geometry (South Hadley, MA, 1996)*, volume 223 of *Contemp. Math.*, pages 109–143. Amer. Math. Soc., Providence, RI, 1999.

[16] A. Edelman and E. Kostlan. How many zeros of a random polynomial are real? *Bulletin of the American Mathematical Society*, 32(1):1–37, 1995.

[17] K. Farahmand. *Topics in Random Polynomials*, volume 393 of *Pitman Research Notes in Mathematics*. Longman, Harlow, 1998.

[18] E. R. Fuller, D. M. Saylor, and T. Wanner. Homological characterization of microstructure response fields in polycrystals. *In preparation*, 2008.

[19] Z. Galias, T. Kapela, M. Mrozek, P. Pilarczyk, D. Wilczak, and P. Zgliczynski. Computer Asisted Proofs in Dynamics. `http://capd.wsb-nlu.edu.pl/`, 2006.

[20] M. Gameiro, K. Mischaikow, and W. Kalies. Topological characterization of spatial-temporal chaos. *Physical Review E*, 70(3):035203, 4, 2004.

[21] M. Gameiro, K. Mischaikow, and T. Wanner. Evolution of pattern complexity in the Cahn-Hilliard theory of phase separation. *Acta Materialia*, 53(3):693–704, 2005.

[22] T. Hartley. *An Analysis of Phase Separation Processes for Stochastic and Nonlocal Extensions of the Classical Phase-Field Model*. PhD thesis, George Mason University, 2008.

[23] T. Hartley and T. Wanner. A semi-implicit spectral method for stochastic nonlocal phase-field models. *In preparation*, 2008.

[24] J. M. Hyde, M. K. Miller, M. G. Hetherington, A. Cerezo, G. D. W. Smith, and C. M. Elliott. Spinodal decomposition in Fe-Cr alloys: Experimental study at the atomic level and comparison with computer models — III. Development of morphology. *Acta Metallurgica et Materialia*, 43:3415–3426, 1995.

[25] T. Kaczynski, K. Mischaikow, and M. Mrozek. *Computational Homology*, volume 157 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2004.

[26] W. Kalies, M. Mrozek, and P. Pilarczyk. Computational Homology Project. `http://chomp.rutgers.edu/`, 2006.

[27] R. B. Kearfott and V. Kreinovich. Applications of interval computations: an introduction. In *Applications of interval computations (El Paso, TX, 1995)*, volume 3 of *Appl. Optim.*, pages 1–22. Kluwer Acad. Publ., Dordrecht, 1996.

[28] K. Krishan, M. Gameiro, K. Mischaikow, M. Schatz, H. Kurtuldu, and S. Madruga. Homology and symmetry breaking in Rayleigh-Benard convection: Experiments and simulations. *Physics of Fluids*, 19:117105, 2007.

[29] J. S. Langer. Theory of spinodal decomposition in alloys. *Annals of Physics*, 65:53–86, 1971.

[30] S. Maier-Paape and T. Wanner. Spinodal decomposition for the Cahn-Hilliard equation in higher dimensions. Part I: Probability and wavelength estimate. *Communications in Mathematical Physics*, 195(2):435–464, 1998.

[31] S. Maier-Paape and T. Wanner. Spinodal decomposition for the Cahn-Hilliard equation in higher dimensions: Nonlinear dynamics. *Archive for Rational Mechanics and Analysis*, 151(3):187–219, 2000.

[32] K. R. Mecke and V. Sofonea. Morphology of spinodal decomposition. *Physical Review E*, 56(4):R3761–R3764, 1997.

[33] R. Mendoza, J. Alkemper, and P. W. Voorhees. The topology of coarsened microstructures. In M. Rappaz, C. Beckermann, and R. Trivedi, editors, *Solidification Processes and Microstructures: A Symposium in Honor of Wilfried Kurz*, pages 123–129. TMS, 2004.

[34] K. Mischaikow and T. Wanner. Probabilistic validation of homology computations for nodal domains. *Annals of Applied Probability*, 17(3):980–1018, 2007.

[35] K. Mischaikow and T. Wanner. Topology-guided sampling of nonhomogeneous random fields. *Submitted for publication*, 2008.

[36] R. E. Moore. *Methods and Applications of Interval Analysis*, volume 2 of *SIAM Studies in Applied Mathematics*. Society for Industrial and Applied Mathematics, Philadelphia, 1979.

[37] J. R. Munkres. *Elements of Algebraic Topology*. Addison-Wesley, Menlo Park, 1984.

[38] A. Neumaier. *Interval Methods for Systems of Equations*, volume 37 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1990.

[39] A. Neumaier. *Introduction to Numerical Analysis*. Cambridge University Press, Cambridge, 2001.

[40] P. Niyogi, S. Smale, and S. Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete and Computational Geometry*, 39:419–441, 2008.

[41] S. Osher and R. Fedkiw. *Level Set Methods and Dynamic Implicit Surfaces*, volume 153 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2003.

[42] E. Sander and T. Wanner. Monte Carlo simulations for spinodal decomposition. *Journal of Statistical Physics*, 95(5–6):925–948, 1999.

[43] E. Sander and T. Wanner. Unexpectedly linear behavior for the Cahn-Hilliard equation. *SIAM Journal on Applied Mathematics*, 60(6):2182–2202, 2000.

[44] J. A. Sethian. *Level Set Methods and Fast Marching Methods*. Cambridge University Press, New York, 1999.

[45] T. Wanner. Maximum norms of random sums and transient pattern formation. *Transactions of the American Mathematical Society*, 356(6):2251–2279, 2004.