

PROBABILISTIC AND NUMERICAL VALIDATION OF HOMOLOGY COMPUTATIONS FOR NODAL DOMAINS

SARAH DAY, WILLIAM D. KALIES, KONSTANTIN MISCHAIKOW,
AND THOMAS WANNER

ABSTRACT. Homology has long been accepted as an important computable 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 accurate the resulting homology computations are. In this paper we present a probabilistic approach to quantifying the validity of homology computations for nodal domains of random Fourier series in one and two space dimensions, which furnishes explicit probabilistic a-priori bounds for the suitability of certain discretization sizes. In addition, we introduce a numerical method for verifying the homology computation using interval arithmetic.

1. INTRODUCTION

The practical need to extract low-dimensional nonlinear structures from high-dimensional data sets has led to the introduction of topological methods in statistical analysis, and currently there is a growing body of literature [22, 25, 10, 9, 6] that addresses the following problem: provide efficient algorithms for estimating topological properties of an unknown manifold X given a point-cloud data set that lies on or near X . It should be mentioned that for some of the proposed applications it is reasonable to assume that the point-cloud data set is obtained via a random sampling process.

To a large extent the emphasis of the above-mentioned papers is on the method of estimation. From now on we restrict our attention to homology where to the best of our knowledge the first result concerning the accuracy of such estimation is due to Niyogi, Smale, and Weinberger [21]. In this paper, 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 derive explicit bounds 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

Date: November 14, 2006.

1991 Mathematics Subject Classification. Primary 60G60, 55-04; Secondary 55N99, 60G15.

Key words and phrases. Homology, random Fourier series, nodal domain.

Sarah Day was partially supported by NSF grant DMS-9983660 at Cornell University and NSF grant DMS-0441170 at MSRI. .

William Kalies was partially supported by NSF grant DMS-0511208 and DOE grant 97888.

Konstantin Mischaikow was partially supported by NSF grants DMS-0511115 and DMS-0107396, by DARPA, and by DOE grant 97891.

Thomas Wanner was partially supported by NSF grant DMS-0406231 and DOE grant 97889.

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 $r < \tau$.

Consider now the problem of an evolutionary system which produces complicated spatial patterns, such as for example phase separation in materials, turbulent fluid flow, predator-prey populations in spatially explicit systems, etc. Again, these are typically high-dimensional systems for which there is considerable interest in understanding the temporal changes of the low-dimensional nodal domains, i.e., patterns, and homological methods appear to provide new techniques through which one can explore these problems [13, 14, 18]. Notice, however, that in this setting the topology of the level sets which define the nodal domains changes as a function of time, and thus, there is no fixed manifold X for which one is trying to compute the homology. Furthermore, as the topology changes the condition number $1/\tau$ becomes unbounded.

While the topological complexity of tracking patterns appears to be greater, the sampling issues can be assumed to be simpler. These patterns are typically obtained either experimentally using various imaging techniques, for example from a digital camera, or numerically by solving a deterministic or stochastic partial differential equations based on some discretization of the underlying evolution equation. This motivates the following assumption on the approximation of the nodal domains (see Figure 1).

Definition 1.1 (Cubical Approximation of a Nodal Domain). Let M be an arbitrary positive integer and let $[a, b]$ denote a compact interval in \mathbb{R} . Define the *equidistant M -discretization* of $[a, b]$ as the collection of the grid points

$$x_k = a + \frac{b-a}{M}, \quad \text{for } k = 0, \dots, M.$$

Let $u : G \rightarrow \mathbb{R}$ be a function defined on a compact rectangular domain $G \subset \mathbb{R}^d$. The *cubical approximations* Q_M^\pm of the nodal domains

$$N^\pm = \{x \in G : \pm u(t, x) \geq 0\}$$

are defined as the sets

$$Q_M^\pm = \bigcup \left\{ \prod_{\ell=1}^d [k_\ell, k_\ell + 1] : \pm u(x_{1,k_1}, \dots, x_{d,k_d}) \geq 0, \quad k \in \{0, \dots, M\}^d \right\},$$

where $k = (k_1, \dots, k_d)$, and $x_{j,0}, \dots, x_{j,M}$ denotes the equidistant M -discretization of the j -th component interval of G .

While [21] marks a significant first step towards quantifying the validity of homology computations, their framework was not designed to deal with the situation described above. For example, it seems difficult to obtain sharp estimates on the condition number τ of nodal domains of a smooth function u , which depends only on easily computable properties of u . In addition, we are assuming a fixed sample size, namely $(M+1)^d$. Since $1/\tau$ becomes unbounded, there are positive intervals of time during which the results of [21] provide no useful information.

Ideally, given a dynamical system and a fixed cubical approximation of the nodal domain we would like to be able to estimate the amount of time for which the homology computations of the nodal domains are correct. For now, we pose the following simpler goal. Consider a random field over the probability space $(\Omega, \mathcal{F}, \mathbb{P})$

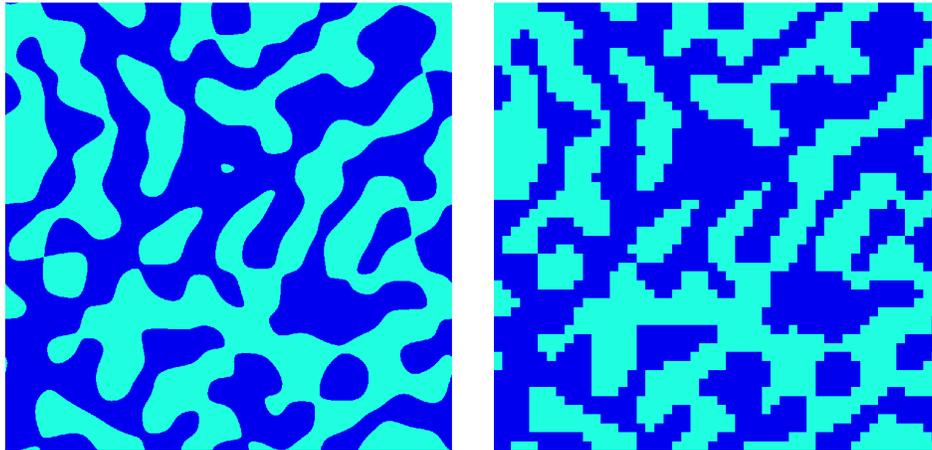


FIGURE 1. Nodal domains of a random trigonometric polynomial in two space dimensions and their cubical approximations with $M = 50$.

and the compact rectangular domain $G \subset \mathbb{R}^d$. Let M be an arbitrary positive integer. Approximate the random nodal domains $N^\pm(\omega)$ of the realization $u(\cdot, \omega) : G \rightarrow \mathbb{R}$ by cubical sets $Q_M^\pm(\omega)$ as in Definition 1.1. We are interested in the following problem:

(P) *Find sharp lower bounds on the probability*

$$\mathbb{P} \{ \omega \in \Omega : H_*(N^\pm(\omega)) \cong H_*(Q_M^\pm(\omega)) \}$$

as a function of M .

As is made clear in this announcement, we approach this problem from two distinct points of view. The first, which is discussed in Section 2, is to obtain probabilistic lower bounds. Since it is not a priori clear that these bounds are sharp, in Section 3 we describe a numerical procedure to rigorously determine when for a given sufficiently smooth $u : G \rightarrow \mathbb{R}$ we have $H_*(N^\pm) \cong H_*(Q_M^\pm)$.

We conclude this paper in Section 4 by applying the probabilistic and numerical techniques to two problems. The first is to determine the number of subdivisions, M , needed in order to guarantee with high probability that the homology of the nodal domain of a random trigonometric polynomial is computed correctly using a cubical approximation. The second is to study the patterns produced by the stochastic Cahn-Hilliard model for phase separation in binary alloys.

2. PROBABILISTIC ESTIMATES

Random fields and in particular random Fourier series have been studied extensively during the last few decades. See for example [1, 2, 12, 16, 19], as well as the references therein. With this in mind, we present an abstract probability estimate for general random fields. Then using this abstract result we consider, separately, the case of random Fourier series for one and two space dimensions.

2.1. An Abstract Result. We begin by presenting an abstract probability estimate for the correctness of homology computations in the one-dimensional case. Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $G = [a, b] \subset \mathbb{R}$ denote a compact interval. Moreover, let $u : G \times \Omega \rightarrow \mathbb{R}$ denote a random field over $(\Omega, \mathcal{F}, \mathbb{P})$ such that for \mathbb{P} -almost all $\omega \in \Omega$ the function $u(\cdot, \omega) : G \rightarrow \mathbb{R}$ is continuous. In addition, assume that the following hold:

- (A1) For every $x \in G$ we have $\mathbb{P}\{u(x) = 0\} = 0$.
- (A2) The random field is such that $\mathbb{P}\{u \text{ has a double zero in } G\} = 0$.
- (A3) For $x \in G$ and $\delta > 0$ with $x + \delta \in G$ and

$$p_\sigma(x, \delta) = \mathbb{P} \left\{ \sigma \cdot u(x) \geq 0, \quad \sigma \cdot u\left(x + \frac{\delta}{2}\right) \leq 0, \quad \sigma \cdot u(x + \delta) \geq 0 \right\}$$

there exists a constant $\mathcal{C}_0 > 0$ such that

$$p_\sigma(x, \delta) \leq \mathcal{C}_0 \cdot \delta^3 \text{ for all } \sigma \in \{\pm 1\} \text{ and } x \in G \text{ with } x + \delta \in G.$$

Of the above assumptions, the first two are usually satisfied for reasonable random fields. Assumption (A3), however, lies at the heart of our result, and establishing its validity generally requires some work. In particular, determining or estimating the constant \mathcal{C}_0 has to be done with significant care, since it has direct implications for the tightness of the resulting bounds.

Under the above assumptions, we can now formulate the following result for the one-dimensional setting.

Theorem 2.1. [20, Theorem 1.3] *Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $G = [a, b] \subset \mathbb{R}$ denote a compact interval. Moreover, let $u : G \times \Omega \rightarrow \mathbb{R}$ denote a random field over $(\Omega, \mathcal{F}, \mathbb{P})$ which satisfies all of the above assumptions. For each $\omega \in \Omega$, denote the nodal domains of $u(\cdot, \omega)$ by $N^\pm(\omega) \subset G$, and denote their cubical approximations as in Definition 1.1 by $Q_M^\pm(\omega)$. Then for every discretization size M , the probability that the homologies of $N^\pm(\omega)$ and $Q_M^\pm(\omega)$ coincide satisfies*

$$\mathbb{P} \{H_*(N^\pm) \cong H_*(Q_M^\pm)\} \geq 1 - \frac{8\mathcal{C}_0(b-a)^3}{3M^2},$$

where \mathcal{C}_0 is the constant introduced in (A3).

The above result isolates the crucial conditions which are necessary for obtaining probabilistic estimates for homology validation in one space dimension. In principle, this result can be applied to any choice of random fields, as long as (A1), (A2), and (A3) can be verified. The third assumption is incorporated into the proof via dyadic subdivisions. Given the interval $G = [a, b]$ the *dyadic points* are

$$d_{n,k} := a + (b-a) \cdot k \cdot 2^{-n} \quad \text{for all } k = 0, \dots, 2^n - 1 \quad \text{and } n \in \mathbb{N}_0.$$

Observe that (A3) is related to the asymptotic probability of sign changes at three consecutive dyadic points which in turn is related to the choice of M , the number of subdivisions.

An analogous abstract result for the two-dimensional case can be obtained as well, which also requires (A1) and (A2). However, the assumption corresponding to (A3) is presented in terms of the asymptotic probabilities of various sign configurations on subsets of the nine dyadic points associated with a square with edge length size 2^{1-n} . Thus, we refer the reader to [20, Theorem 3.8] for a precise formulation.

Because of the applications we have in mind, we are interested in applying these results to explicit classes of random functions. Consider a series of the form

$$(2.1) \quad u(x, \omega) = \sum_{k=0}^{\infty} \alpha_k \cdot g_k(\omega) \cdot \varphi_k(x), \quad u : G \times \Omega \rightarrow \mathbb{R},$$

where $G \subset \mathbb{R}^d$ denotes a compact rectangular set, and φ_k , $k \in \mathbb{N}_0$, denotes a sequence of basis functions which are specified below. Assume that the g_k are independent, identically distributed real-valued random variables over a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The coefficients α_k denote deterministic real numbers.

2.2. The One-Dimensional Periodic Case. Random functions of the form (2.1) arise naturally in the context of partial differential equations. In this case, many of the characteristics of the underlying evolution equation are reflected in the choice of the basis functions φ_k , $k \in \mathbb{N}_0$. Though in principle Theorem 2.1 applies rather broadly, for the sake of simplicity we restrict ourselves to evolution equations subject to periodic boundary conditions. This leads to the following set of assumptions.

- (B1) Consider the interval $G = [0, 1]$, and assume that the basis functions in (2.1) are defined by

$$\varphi_{2k}(x) = \cos(2\pi kx) \quad \text{and} \quad \varphi_{2k-1}(x) = \sin(2\pi kx)$$

for arbitrary $k \in \mathbb{N}$ and $x \in G$ with $\varphi_0(x) = 1$ for $x \in G$.

- (B2) The random variables g_k in (2.1) are defined over a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and they are independent and normally distributed with mean 0 and variance 1.
- (B3) The constants α_k in (2.1) are given by $\alpha_0 = a_0$ and

$$\alpha_{2k} = \alpha_{2k-1} = a_k \quad \text{for} \quad k \in \mathbb{N}.$$

At least two of the constants a_k , $k \in \mathbb{N}_0$, are nonzero, and we have

$$\sum_{k=0}^{\infty} k^6 a_k^2 < \infty.$$

Observe that under the above assumptions, the constants a_k are directly related to smoothness properties of the random function u . More precisely, one can show that if

$$\sum_{k=0}^{\infty} k^{2p} a_k^2 < \infty \quad \text{for some} \quad p > 0,$$

then \mathbb{P} -almost all realizations $u(\cdot, \omega)$ are contained in the Hölder space $C^q[0, 1]$, for any real $0 < q < p$. See for example [16, Section 7.4]. Furthermore, one can easily show that the spatial covariance function of u is given by

$$R(x, y) = r(x - y) = \sum_{k=0}^{\infty} a_k^2 \cdot \cos(2\pi k(x - y)).$$

In other words, under the above assumptions the random function defined in (2.1) is a homogeneous random field in the sense of [1]. This simplifies some of the considerations.

Turning to the problem of determining the homology of the nodal domains of u using a spatial discretization of size M we have the following result.

Theorem 2.2. [20, Theorem 2.7] *Consider the random Fourier series u defined in (2.1), and assume that (B1), (B2), and (B3) are satisfied. Let M denote an arbitrary positive integer, and let $Q_M^\pm(\omega)$ denote the cubical approximations of the random nodal domains $N^\pm(\omega)$ of $u(\cdot, \omega)$ as in Definition 1.1.*

Then the probability that the homology of the random nodal domains $N^\pm(\omega)$ is computed correctly with the discretization of size M satisfies

$$\mathbb{P} \{ H_*(N^\pm) \cong H_*(Q_M^\pm) \} \geq 1 - \frac{\pi^2}{6M^2} \cdot \frac{A_0 A_2 - A_1^2}{A_0^{3/2} A_1^{1/2}} + O\left(\frac{1}{M^3}\right),$$

where

$$A_p = \sum_{k=0}^{\infty} k^{2p} a_k^2 = \frac{1}{(2\pi)^{2p}} \cdot \mathbb{E} \|D_x^p u\|_{L^2(0,1)}^2,$$

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

Very much in the spirit of [21], our result furnishes an explicit bound on the likelihood of computing the correct homology. The bound depends on the discretization size M and global properties of the underlying function u . Unlike [21], however, the necessary information on u can easily be computed. In fact, it is given explicitly in terms of the coefficient sequence (a_k) , which in turn is related to smoothness properties of u . Observe, that the probability estimate involves the A_2 term and hence it is necessary that $u \in C^2$. In light of the importance played by the condition number $1/\tau$ in [21] and its relationship to the curvature of the manifold, it is reasonable to expect that this is a minimal requirement for the computability of homology.

The details of the proof of Theorem 2.2 can be found in [20]. However, we remark that it involves extending results of Dunnage [11] on the zeros of random trigonometric polynomials.

2.3. The Two-Dimensional Periodic Case. Consider random Fourier series on $G = [0, 1]^2$ of the form

$$\begin{aligned} u(x, \omega) = \sum_{k, \ell=0}^{\infty} a_{k, \ell} \cdot & (g_{k, \ell, 1}(\omega) \cos(2\pi k x_1) \cos(2\pi \ell x_2) \\ & + g_{k, \ell, 2}(\omega) \cos(2\pi k x_1) \sin(2\pi \ell x_2) \\ & + g_{k, \ell, 3}(\omega) \sin(2\pi k x_1) \cos(2\pi \ell x_2) \\ & + g_{k, \ell, 4}(\omega) \sin(2\pi k x_1) \sin(2\pi \ell x_2)) \end{aligned} \quad (2.2)$$

under the following assumptions.

- (C1) The random variables $g_{k, \ell, m}$ in (2.2) are defined over a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and they are independent and normally distributed with mean 0 and variance 1.
- (C2) There are positive integers $k_1, \ell_1 \in \mathbb{N}$ and nonnegative integers $k_2, \ell_2 \in \mathbb{N}_0$ with $k_1 \neq k_2$ 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, and in addition

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

As before, the summability condition in (C2) is related to smoothness properties of the random field u . In fact, assumption (C2) guarantees that the function $u(\cdot, \omega)$ has continuous partial derivatives up to order two, for \mathbb{P} -almost all $\omega \in \Omega$. In this setting, we obtain the following result.

Theorem 2.3. [20, Theorem 3.10] *Consider the random Fourier series u defined in (2.2), and assume that both (C1) and (C2) are satisfied. Let M denote an arbitrary positive integer, and let $Q_M^\pm(\omega)$ denote the cubical approximations of the random nodal domains $N^\pm(\omega)$ of $u(\cdot, \omega)$ as in Definition 1.1.*

Then the probability that the homology of the random nodal domains $N^\pm(\omega)$ is computed correctly with the discretization of size M satisfies

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

where

$$A_{p,q} = \sum_{k,\ell=0}^{\infty} k^{2p} \ell^{2q} a_{k,\ell}^2 = \frac{1}{(2\pi)^{2p+2q}} \cdot \mathbb{E} \|D_{x_1}^p D_{x_2}^q u\|_{L^2(G)}^2.$$

Establishing this result is considerably more involved than the one-dimensional case, details can be found in [20].

3. NUMERICAL HOMOLOGY VERIFICATION

The results of the previous section focus on the probability of correctly computing the homology of the nodal domain of a random function given a fixed level of discretization. The complementary question is: given a fixed function can one correctly compute the homology of the nodal domains? In particular, our motivation to address this issue stems in part from a need to understand the optimality of the probabilistic bounds given in Theorems 2.2 and 2.3. This led to the development of numerical techniques which rely on having explicit formulas (or at least appropriate bounds) for the function and its first and second derivatives, and when successful (we return to this point at the end of the section), produce the right homology information.

To control the computational cost, the basic approach is to adaptively divide the domain into boxes on which we verify that the sign structure on the vertices captures the topology of the nodal domain inside of the box. More explicitly, we use u , its derivatives, and interval arithmetic to prove that the double zero condition considered in the probabilistic study does not occur along appropriate rays in each box. Boxes in a grid are tested and are subdivided if the test does not rule out the possibility of a double zero along appropriate rays. The result is a non-uniform grid containing smaller boxes as required to resolve the nodal domain (see Figure 2).

We now describe the procedure for a two-dimensional, positive nodal domain N^+ , and an analogous approach works for the negative nodal domain N^- and the simpler setting of one-dimensional nodal domains. The first step is to subdivide the domain evenly in each coordinate direction to obtain a uniform grid. We next compute the sign of u on each of the vertices of the grid. It is important to note that all numerical computations will be performed using interval arithmetic to account for round-off errors. If the sign of u cannot be rigorously determined at a vertex, then to proceed further requires a modification of the grid.

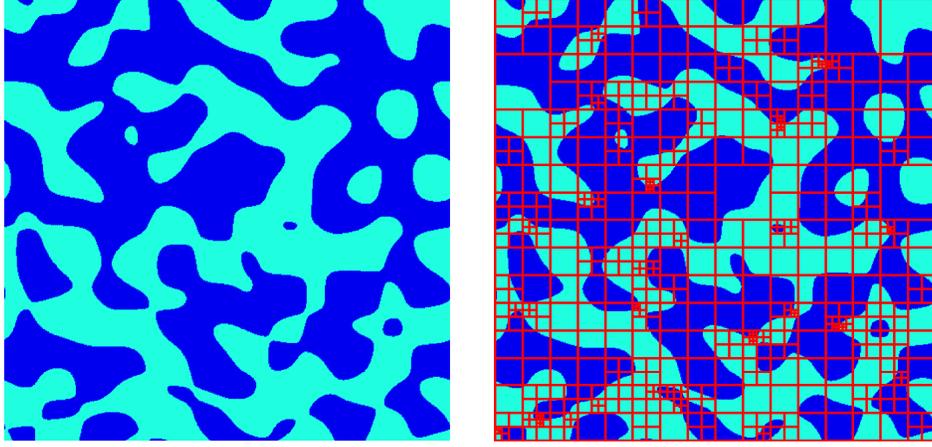


FIGURE 2. Nodal domains of a bivariate random trigonometric polynomial with $N = 7$ and boxes on which the topology can be determined from the corner function values.

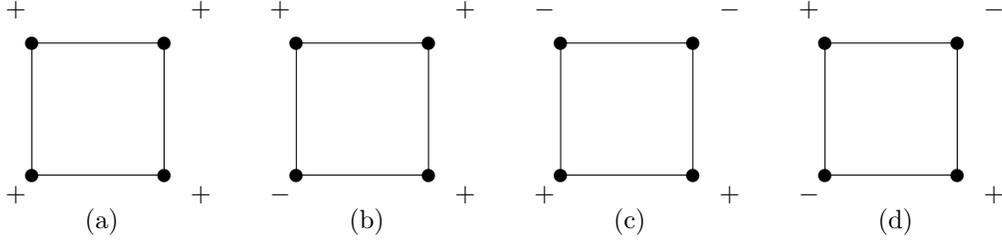


FIGURE 3. Possible sign structure up to rotation and sign inversion on the vertices of a grid element B .

For each possible vertex sign configuration on a grid element B , we now define a *verification step* whereby we try to determine the topological structure of $N^+ \cap B$. The verification step will depend on the sign configuration on the vertices of B which falls into one of the four categories shown up to rotation and sign inversion in Figure 3. 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.

Case (a): In this case, we check that $u(x) > 0$ for all $x \in B$. In other words, $N^+ \cap B = B$. This check is implemented by using interval arithmetic to evaluate an outer approximation $\widetilde{u(B)}$ of $u(B)$. In practice we compute

$$\widetilde{u(B)} := u(c) + Du(B) \cdot (B - c)^t$$

where $c = (c_x, c_y)$ is the center point of $B = (B_x, B_y)$, with B_x and B_y the intervals of B in the x and y directions respectively, $Du(B) = (\partial_x u(B_x, B_y), \partial_y u(B_x, B_y))$,

and all computations are performed using interval arithmetic to also bound round-off error. If all points in this outer approximation are strictly positive, then we have verified that $u(x) > 0$ for all $x \in B$.

Cases (b),(c): To rule out the possibility of a double zero occurring along vertical segments in Case (c) and horizontal and vertical segments in Case (b), we try to show that u is monotone in appropriate directions. For case (c), we use interval arithmetic to verify that $u > 0$ on the bottom edge of B , $u < 0$ on the top edge of B , and $0 \notin u_y(B)$ so that u is monotone in the vertical direction. For Case (b), we analogously show that u is monotone in both the vertical and the horizontal direction.

Case (d): 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 will be subdivided.

If this subdivision and verification procedure terminates successfully, meaning that we obtain a grid where each grid element passes the corresponding verification test, then we have obtained a resolution sufficient for approximating N^+ . We define the cubical representation \mathcal{N}^+ for the nodal domain N^+ on a uniform grid whose diameter matches the smallest elements in the adaptive grid used for verification. This grid is also augmented with an extra set of boxes on the left side and bottom of the domain to account for a shift in the cubical representation introduced by the following choice. For the cubical representation, we make the choice that the grid element B is in \mathcal{N}^+ if and only if the vertex sign on the upper right hand corner of B , as determined by outer approximation based on interval arithmetic, is positive. The homology of the cubical set \mathcal{N}^+ may now be computed using existing software [15, 17].

The proof of the following theorem is presented in [8].

Theorem 3.1. *Let \mathcal{N}^+ be the cubical representation of N^+ produced by the adaptive verification technique described above. Then $H_*(\mathcal{N}^+) = H_*(N^+)$.*

It can be shown, [20, Theorem 1.2], that if the random field $u : G \times \Omega \rightarrow \mathbb{R}$ over the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is \mathbb{P} -almost surely twice differentiable and satisfies assumptions (A1) and (A2), then with probability 1 the homology of a nodal domain can be computed given a sufficiently fine grid. In practice, however, failure can occur. For example, the boxes in the adaptive grid may reach a preset minimal size so that the algorithm is terminated before the result is obtained. The adaptive verification technique outlined above is used in [8] to study the probabilistic estimates in Theorem 2.3 in greater depth.

4. SPECIFIC APPLICATIONS

In this section, we use the probabilistic and numerical techniques described above to study the homology of nodal sets in two specific contexts, random trigonometric polynomials and the linearized stochastic Cahn-Hilliard model.

4.1. Random Trigonometric Polynomials. Observe that Theorem 2.2 applies to random trigonometric polynomials

$$(4.1) \quad u(x, \omega) = \sum_{k=1}^N (g_{2k}(\omega) \cdot \cos(2k\pi x) + g_{2k-1}(\omega) \cdot \sin(2k\pi x))$$

if the coefficients in (B3) are chosen to be $a_k = 1$ for $1 \leq k \leq N$, $a_0 = 0$, and $a_k = 0$ for $k > N$, where $N \geq 3$. In particular,

$$\frac{A_0 A_2 - A_1^2}{A_0^{3/2} A_1^{1/2}} = \frac{\sqrt{6}}{180} (N-1)(8N+11) \sqrt{(N+1)(2N+1)} \sim \frac{4\sqrt{3}}{45} \cdot N^3$$

which suggests that, in order for the homology computation to be accurate with high confidence, we have to choose the discretization size M in such a way that

$$(4.2) \quad M \sim N^{3/2} \quad \text{for} \quad N \rightarrow \infty .$$

A cautionary remark is that because of the relationship between cubical approximations and assumption (A3), a priori Theorem 2.2 only provides an asymptotic bound. With this in mind we turn to the question of measuring the validity of the result for “small” values of N and determining how sharp the result is.

The asymptotic behavior of the number of zeros of random trigonometric polynomials has been studied extensively. Specifically, one can show that as $N \rightarrow \infty$, most trigonometric polynomials of the form (4.1) have on the order of $2N/3^{1/2}$ real zeros in the interval $[0, 1]$. See for example [2, 12]. Thus, in order for our homology computation to be correct, we need to at least make sure that the discretization size M satisfies $1/M = O(1/N)$. Due to the almost certain occurrence of zeros which are more closely spaced, it seems implausible that this constraint would be optimal.

To obtain better bounds we turn to the numerical techniques described in Section 3. Figure 4 contains the results for various values of N between 5 and 1000 of the computations of four different quantities:

- (1) the expected number of zeros (red, lower-most curve),
- (2) the inverse of the expected minimal distance between two consecutive zeros of the random polynomial (blue curve, second from below),
- (3) the prediction of Theorem 2.2 which would ensure a correctness probability of at least 95% is given by the magenta curve (second from the top), i.e.,

$$M^2 = \frac{8\pi^2}{9\sqrt{3}} \cdot N^3 ,$$

- (4) the integer M for which 95% of the generated random polynomials have a minimal distance between consecutive zeros which is larger than $1/M$ (green, upper-most curve).

The fact that the curve (3) lies exactly between the curves (2) and (4) and shows the same asymptotic growth, leads us to conjecture that (4.2) provides an appropriate ratio between the number of modes and the discretization size.

The same analysis can be applied to the two-dimensional case. Again, let $N \geq 3$. If the coefficients of (2.2) are chosen to be $a_{k,\ell} = 1$ for $1 \leq k, \ell \leq N$, and $a_{k,\ell} = 0$ otherwise, then u is a bivariate random trigonometric polynomial and

$$\frac{(A_{2,0} + A_{1,1} + A_{0,2})^2}{A_{0,0}^{1/2} A_{0,1}^{1/2} A_{1,0}^{1/2} A_{1,1}^{1/2}} = \frac{1}{900} \cdot (46N^2 + 51N - 7)^2 \sim \frac{529}{225} \cdot N^4 .$$

This implies that in order for the homology computation to be accurate with high confidence, we have to choose the discretization size M in such a way that

$$M \sim N^2 \quad \text{for} \quad N \rightarrow \infty ,$$

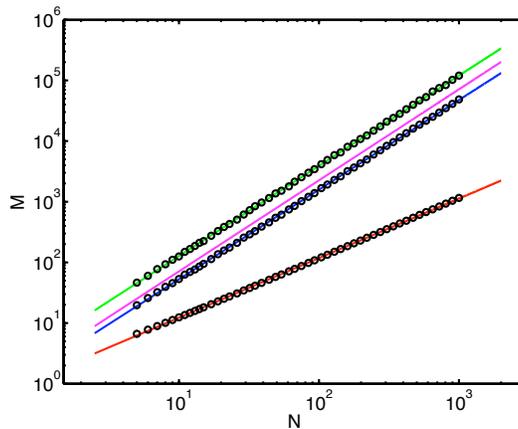


FIGURE 4. Numerical results for random trigonometric polynomials. Curves (1), (2), (3), (4) from bottom to top, respectively.

in contrast to the one-dimensional case. A numerical study of the accuracy of this ratio is presented in [8].

4.2. 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 which is given by

$$(4.3) \quad \frac{\partial u}{\partial t} = -\Delta (\varepsilon^2 \Delta u + u - u^3) + \sigma \cdot \xi \quad \text{in } G \subset \mathbb{R}^d,$$

where G denotes a square domain, ξ is the generalized derivative of a suitable Q -Wiener process, and ε and σ are small positive parameters. This stochastic partial differential equation has been proposed as a model for phase separation in metallic alloys and produces complicated patterns, see for example [3, 4, 5, 7, 26] and the references therein. As we mentioned in the introduction, computational homology can be used to quantify these complicated structures [14], 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 (4.3) originating at a random field, then for any time $t > 0$ the solution $u(t, \cdot)$ is a random field over G . 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 model

$$(4.4) \quad \frac{\partial u}{\partial t} = -\Delta (\varepsilon^2 \Delta u + u) + \sigma \cdot \xi \quad \text{in } G \subset \mathbb{R}^d,$$

provided the initial condition satisfies the assumptions of our theorems.

For the sake of brevity, we demonstrate the applicability of our results on homology validation only for the deterministic special case $\sigma = 0$ and the one-dimensional base domain $G = (0, 1)$ under periodic boundary conditions, the general case can be found in [8]. Furthermore, suppose that the initial condition $u(0, \cdot)$ is a random periodic field as in (4.1) for an appropriate choice of $N = N_\varepsilon$. Then for every $t > 0$

the solution of (4.4) with $\sigma = 0$ is given by

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

where $\lambda_k = 4\pi^2 k^2 \cdot (1 - 4\pi^2 k^2 \varepsilon^2)$ denotes the k -th eigenvalue of the linearized Cahn-Hilliard operator. Choosing $N_\varepsilon = \lfloor r/(2\pi\varepsilon) \rfloor$, for some fixed $r > 1$ guarantees that for every small $\varepsilon > 0$ the initial condition contains all unstable modes, i.e., all modes which are responsible for the formation of the complicated patterns. Again, see [8] for details.

The explicit representation of the solution u of (4.4) shows that our results for random Fourier series are readily applicable. Consider

$$S_{\ell, \varepsilon}(\tau) = \sum_{k=1}^{N_\varepsilon} (2\pi\varepsilon)^{2\ell+1} k^{2\ell} e^{2\varepsilon^2 \lambda_k \tau} \xrightarrow{\varepsilon \rightarrow 0} \int_0^\tau s^{2\ell} e^{2\tau s^2(1-s^2)} ds ,$$

then using the notation of Theorem 2.2 we have

$$\frac{A_0 A_2 - A_1^2}{A_0^{3/2} A_1^{1/2}} = \frac{I_\varepsilon(t/\varepsilon^2)}{8\pi^3 \varepsilon^3} , \quad \text{where} \quad I_\varepsilon(\tau) = \frac{S_{0, \varepsilon}(\tau) S_{2, \varepsilon}(\tau) - S_{1, \varepsilon}(\tau)^2}{S_{0, \varepsilon}(\tau)^{3/2} S_{1, \varepsilon}(\tau)^{1/2}} .$$

Thus, the probability estimate takes the form

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

Notice that for every fixed $\tau > 0$ the value $I_\varepsilon(\tau)$ converges as $\varepsilon \rightarrow 0$. The above estimate implies that in order to compute the homology of the nodal domains correctly with high probability, we have to choose $M \sim \varepsilon^{-3/2}$. This is in accordance with the fact that the observed patterns exhibit a typical thickness which is proportional to ε as $\varepsilon \rightarrow 0$.

In order to demonstrate the accuracy of our probabilistic predictions also in the partial differential equations case, we employ the method of validated computations from [8]. For $\varepsilon = 0.005$ 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 $I_\varepsilon(t/\varepsilon^2)/(48\pi\varepsilon^3 M^3)$ for large values of M , and this is confirmed in the left graph of Figure 5. In fact, the asymptotic behavior predicted by Theorem 2.2 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 [12] 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

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

The graph of $\mathbb{E}Z(t)$ is qualitatively similar to the $M = 25$ curve in Figure 5: After an initial decrease to a minimal value of 41.18 at $t/\varepsilon^2 \approx 2.4$, the graph increases again and limits to 44.94 as $t \rightarrow \infty$. Thus, the probability that the interval $[0, 1/M]$ for $M = 25$ contains more than two zeros is fairly large. In fact, the computations

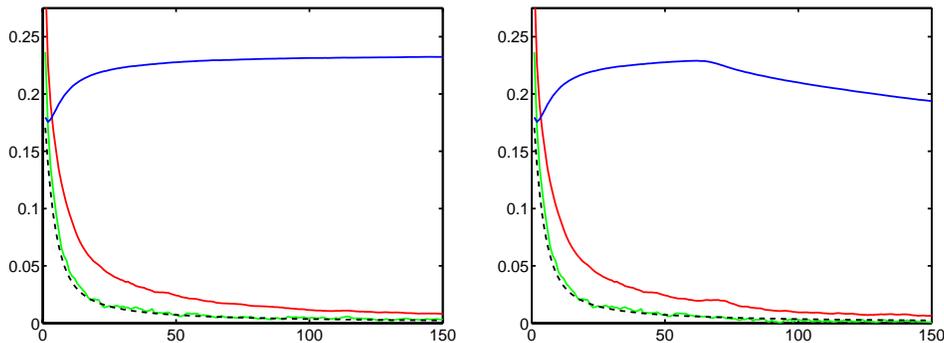


FIGURE 5. Validated numerical results for the probability of a false homology computation in the linearized Cahn-Hilliard model (left) and the nonlinear Cahn-Hilliard model (right). From top to bottom the solid lines correspond to $M = 25, 50, 100$, the dashed line shows the function $I_\varepsilon(\tau)$ from the probabilistic estimate. All curves have been scaled by the factor $48\pi\varepsilon^3 M^3$.

of Figure 5 show that for $t/\varepsilon^2 \approx 2.4$ this probability is 59.6%, while for large t it stabilizes at 79.0%.

The right graph in Figure 5 contains analogous numerical results for the nonlinear Cahn-Hilliard equation (4.3), again in the deterministic situation with $\sigma = 0$. Notice that now the curve for $M = 25$ exhibits a marked decay starting at around $t/\varepsilon^2 \approx 70$, and the remaining curves show similar, although not as pronounced, behavior. On the other hand, for times $t \leq 70\varepsilon^2$ 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 [3, 4, 23, 24, 26]. 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 [14].

REFERENCES

- [1] R. J. Adler. *The Geometry of Random Fields*. John Wiley & Sons Ltd., Chichester, 1981.
- [2] A. T. Bharucha-Reid and M. Sambandham. *Random Polynomials*. Academic Press, Orlando, 1986.
- [3] 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.
- [4] D. Blömker, S. Maier-Paape, and T. Wanner. Second phase spinodal decomposition for the Cahn-Hilliard-Cook equation. *Transactions of the American Mathematical Society*, to appear.
- [5] 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.
- [6] G. Carlsson and V. de Silva. Topological approximation by small simplicial complexes. *Preprint*, 2003.

- [7] H. Cook. Brownian motion in spinodal decomposition. *Acta Metallurgica*, 18:297–306, 1970.
- [8] S. Day, W. D. Kalies, and T. Wanner. Homology computations of nodal domains: Accuracy estimates and validation. *In preparation*, 2006.
- [9] V. de Silva and G. Carlsson. Topological estimation using witness complexes. In M. Alexa and S. Rusinkiewicz, editors, *Eurographics Symposium on Point-Based Graphics*. The Eurographics Association, 2004.
- [10] D. L. Donoho and C. Grimes. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. *Proceedings of the National Academy of Science*, 100(10):5591–5596, 2003.
- [11] J. E. A. Dunnage. The number of real zeros of a random trigonometric polynomial. *Proceedings of the London Mathematical Society*, 16:53–84, 1966.
- [12] K. Farahmand. *Topics in Random Polynomials*, volume 393 of *Pitman Research Notes in Mathematics*. Longman, Harlow, 1998.
- [13] M. Gameiro, K. Mischaikow, and W. Kalies. Topological characterization of spatial-temporal chaos. *Physical Review E*, 70(3):035203, 4, 2004.
- [14] 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.
- [15] T. Kaczynski, K. Mischaikow, and M. Mrozek. *Computational homology*, volume 157 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2004.
- [16] J.-P. Kahane. *Some Random Series of Functions*. Cambridge University Press, Cambridge – London – New York, second edition, 1985.
- [17] W. Kalies, M. Mrozek, and P. Pilarczyk. Computational homology project. <http://www.math.gatech.edu/~chomp/>, 2006.
- [18] K. Krishan, M. Gameiro, K. Mischaikow, and M. F. Schatz. Homological characterization of spiral defect chaos in Rayleigh-Benard convection. *Preprint*, 2005.
- [19] M. B. Marcus and G. Pisier. *Random Fourier Series with Applications to Harmonic Analysis*, volume 101 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, 1981.
- [20] K. Mischaikow and T. Wanner. Probabilistic validation of homology computations for nodal domains. *Submitted for publication*, 2006.
- [21] P. Niyogi, S. Smale, and S. Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete and Computational Geometry*, 2006. To appear.
- [22] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290:2323–2326, 2000.
- [23] E. Sander and T. Wanner. Monte Carlo simulations for spinodal decomposition. *Journal of Statistical Physics*, 95(5–6):925–948, 1999.
- [24] E. Sander and T. Wanner. Unexpectedly linear behavior for the Cahn-Hilliard equation. *SIAM Journal on Applied Mathematics*, 60(6):2182–2202, 2000.
- [25] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–2323, 2000.
- [26] T. Wanner. Maximum norms of random sums and transient pattern formation. *Transactions of the American Mathematical Society*, 356(6):2251–2279, 2004.

COLLEGE OF WILLIAM AND MARY, DEPARTMENT OF MATHEMATICS, P.O. BOX 8795, WILLIAMSBURG, VA 23187

E-mail address: `sday@math.wm.edu`

DEPARTMENT OF MATHEMATICAL SCIENCES, FLORIDA ATLANTIC UNIVERSITY, 777 GLADES ROAD, BOCA RATON, FL 33431

E-mail address: `wkalies@fau.edu`

DEPARTMENT OF MATHEMATICS, RUTGERS UNIVERSITY, 110 FRELINGHUSEN RD, PISCATAWAY, NJ 08854-8019, USA

E-mail address: `mischaik@math.rutgers.edu`

DEPARTMENT OF MATHEMATICAL SCIENCES, GEORGE MASON UNIVERSITY, 4400 UNIVERSITY DRIVE, MS 3F2, FAIRFAX, VA 22030

E-mail address: `wanner@math.gmu.edu`