Computational Homology Tutorial:
Accuracy of Homology Computations

Thomas Wanner
wanner@math.gmu.edu

Department of Mathematical Sciences
George Mason University
Fairfax, VA 22030
Phase field models are widely used to describe the dynamics of phase separation in materials science. The resulting microstructures and their patterns are generally a transient phenomenon and evolve with time.

(Figure from Mendoza, Alkemper, Voorhees (2003))
Homological Analysis of Microstructures

For a given phase variable $u$ and a threshold $\mu$ one can consider the Betti numbers $\beta_0$ and $\beta_1$ of the sets

$$N^+(t) = \{ x \in \Omega \mid u(t, x) \geq \mu \}$$

and

$$N^-(t) = \{ x \in \Omega \mid u(t, x) \leq \mu \}$$

Sample set $N^+(t)$ for $\mu = 0$, $\sigma = 0$, and $t = 0.0036$.
The set has $\beta_0 = 26$ components and $\beta_1 = 4$ loops.
Sample Betti number evolution for the Cahn-Hilliard model and the Cahn-Hilliard-Cook model.

Are these fluctuations numerical artifacts?
Homology via Discretization

Fundamental questions:

- From a mathematical point of view, the objects of interest — microstructures or patterns — are manifolds, which are often defined through level sets of differentiable functions.
- To make these objects amenable to a computational treatment, it is necessary to introduce some sort of finite discretization.
- Yet, how can one be sure that the computational results yield the correct homology of the underlying geometric object?

Is it enough to choose a sufficiently fine discretization? If so, can we determine the correct discretization size a-priori?
Example: Approximation of nodal domains by cubical sets

\[ N^\pm = \{ x \in [a, b] : \pm u(x) \geq 0 \} \]

\[ Q^\pm = \bigcup \{ [k, k+1] : \pm u(x_k) > 0 \} \]
**Errors Caused by Discretization Effects**

**Problem:**

Certain errors in homology computations which are caused by discretization effects persist even for finer discretizations — and are therefore more or less unavoidable.

Example from *Computational Homology* by Kaczynski, Mischaikow, and Mrozek (2003):
Is it possible to determine the likelihood of success or failure of performing a homology computation with a given discretization?

**Practical considerations:**

- For simulations such as the ones described earlier, the function values are known only on a fixed regular grid which is determined by the numerical method.
- The nodal domains are not given directly, only implicitly.
- On the other hand, there is a natural notion of randomness intrinsic to the problem:
  - Random ensemble of initial conditions,
  - Stochastic evolution equation.

Related recent work due to Niyogi, Smale, Weinberger (2004) considers the case of explicitly given manifolds.
Random Fourier Series

Typical situation:

- For evolution equations (deterministic or stochastic) with random ensembles of initial conditions, the solution at some point in time is given as a random Fourier series

\[ u(x, \omega) = \sum_{k=0}^{\infty} \alpha_k \cdot g_k(\omega) \cdot \varphi_k(x) \]

- The numbers \( \alpha_k \) are real constants, the random variables \( g_k \) are independent, and the functions \( \varphi_k : I \to \mathbb{R}, k \in \mathbb{N}_0 \), form a complete orthogonal set in the considered function space.

- One is interested in the homology of the nodal domains

\[ N^\pm = \{ x \in I : \pm u(x) \geq 0 \} \]
Random Fourier Series in 1D

- We consider linear evolution equations with periodic boundary conditions and Gaussian ensembles.

- The random variables $g_k$ are independent and normally distributed with mean 0 and variance 1.

- The orthogonal basis is given by $\varphi_0(x) = 1$ and

$$\varphi_{2k}(x) = \cos(kx) \ , \ \varphi_{2k-1}(x) = \sin(kx) \ , \ k \in \mathbb{N}$$

- The real numbers $\alpha_k$ are given by

$$\alpha_{2k} = \alpha_{2k-1} = a_k \ , \ k \in \mathbb{N}$$

- The series $u(x, \omega)$ is a homogeneous Gaussian random field with mean 0 and spatial covariance function

$$R(x, y) = r(x - y) = \sum_{k=0}^{\infty} a_k^2 \cdot \cos(k(x - y))$$
Homology via Discretization in 1D

Computing the homology of nodal domains:

- We are interested in the random nodal domains

\[ N^{±}(ω) = \{ x ∈ I = [0, 2π] : ±u(x, ω) ≥ 0 \} \]

- Consider the discretization of \( I \) of size \( M \) given by

\[ x_k = k \cdot \frac{2π}{M}, \quad k = 0, \ldots, M \]

- With this discretization we associate the random cubical complexes

\[ Q^{±}(ω) = \bigcup \{ [k, k + 1] : ±u(x_k, ω) > 0 \} \]
\[ \mathbb{P} \{ \omega : H_* (N^\pm(\omega)) = H_* (Q^\pm(\omega)) \} = ? \]
Resolving Nodal Domains in 1D

\[ u(\cdot, \omega) \]

\[ N^-(\omega) \]

\[ N^+(\omega) \]

Accuracy of Homology Computations – p. 13
Main Probabilistic Result in 1D

Consider the random Fourier series \( u \) as before and assume that

\[
\sum_{k=0}^{\infty} k^6 a_k^2 < \infty \quad \text{and} \quad a_{k_1} \neq 0, \ a_{k_2} \neq 0 \quad \text{for} \quad k_1 < k_2
\]

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

\[
1 - P \leq \frac{\pi^2}{6M^2} \cdot \frac{A_2 A_0 - A_1^2}{A_0^{3/2} A_1^{1/2}} + O \left( \frac{1}{M^4} \right)
\]

where

\[
A_\ell = \sum_{k=0}^{\infty} k^{2\ell} a_k^2 = \frac{1}{2\pi} \cdot \mathbb{E} \left| D_x^\ell u \right|_{L^2(0,2\pi)}^2
\]
Main Probabilistic Result in 1D

What does the result imply?

- The result provides explicit probability estimates for the correctness of the homology computation.
- The probability estimate depends on the discretization size and on central parameters of the random field which relate to its smoothness properties and to its derivatives up to second order.
- The result provides a-priori information on choosing a suitable discretization size.

How sharp is this estimate?
Application: Finite Trigonometric Sums

Any random trigonometric polynomial of the form

\[ u(x, \omega) = \sum_{k=1}^{N} a_k \cdot (g_{2k}(\omega) \cdot \cos(kx) + g_{2k-1}(\omega) \cdot \sin(kx)) \]

has at most \(2N\) zeros. In this situation our result furnishes:

The probability \(P\) that the homology of the random nodal domains \(N^{\pm}(\omega)\) is computed correctly with the discretization of size \(M\) satisfies

\[
1 - P \leq \frac{2\sqrt{3}\pi^2}{135} \cdot \frac{N^3}{M^2} + O\left(\frac{1}{M^4}\right)
\]

In order to compute the homology correctly with high confidence we need to choose \(M \sim N^{3/2}\).
Numerical results confirm the optimality!

Shown are the expected number of zeros, the expected minimal distance between two consecutive zeros, the value of $M$ for which 95% of the functions had minimal distance at least $2\pi/M$, and the value of the discretization size $M$ for which the probability estimate yields $P = 95\%$. For each $N$ we considered 15,000 random trigonometric sums.
Application: Linear Cahn-Hilliard Model

The solution of the linearized Cahn-Hilliard equation originating at a Gaussian random field is given by

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

where $\lambda_k = k^2(1 - \varepsilon^2 k^2)$. In this situation one obtains the probability estimate

$$1 - P \leq \frac{\pi^2}{6\varepsilon^3 M^2} \cdot f(t\varepsilon^{-2}) + O\left(\frac{1}{M^4}\right)$$

for some $\varepsilon$-independent, decreasing function $f$ with $f(1) \approx 1/5$.

In order to compute the homology correctly with high confidence we need to choose $M \sim \varepsilon^{-3/2}$.
Towards a Two-dimensional Result

Can this result be generalized to two-dimensional domains?
Consider a random Fourier series on $\Omega = [0, 2\pi]^2$ of the form

$$u(x, \omega) = \sum_{k, \ell=0}^{\infty} a_{k, \ell} \cdot (g_{k, \ell, 1}(\omega) \cos(kx_1) \cos(\ell x_2) +$$

$$+ g_{k, \ell, 2}(\omega) \cos(kx_1) \sin(\ell x_2)$$

$$+ g_{k, \ell, 3}(\omega) \sin(kx_1) \cos(\ell x_2)$$

$$+ g_{k, \ell, 4}(\omega) \sin(kx_1) \sin(\ell x_2))$$

The random variables $g_{k, \ell, m}$ are independent and normally distributed with mean 0 and variance 1. There are integers $k_1, \ell_1 \in \mathbb{N}$ and $k_2, \ell_2 \in \mathbb{N}_0$ with $k_1 \neq k_2$ and $\ell_1 \neq \ell_2$ such that both $a_{k_1, \ell_1} \neq 0$ and $a_{k_2, \ell_2} \neq 0$, and in addition

$$\sum_{k, \ell=0}^{\infty} (k^6 + \ell^6) a_{k, \ell}^2 < \infty$$
Preliminary Probabilistic Result in 2D

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

$$1 - P \leq \frac{3\pi^2}{4M} \cdot \left( \frac{A_{0,2}A_{0,0} - A_{0,1}^2}{A_{0,0}^{3/2}A_{0,1}^{1/2}} + \frac{A_{2,0}A_{0,0} - A_{1,0}^2}{A_{0,0}^{3/2}A_{1,0}^{1/2}} \right)$$

$$+ \frac{32\pi^2}{9M^2} \cdot \frac{A_{1,1}^{3/2}}{A_{0,0}^{1/2}A_{0,1}^{1/2}A_{1,0}^{1/2}} + O \left( \frac{1}{M^3} \right),$$

where

$$A_{p,q} = \sum_{k,\ell=0}^{\infty} k^2p \ell^2q a_{k,\ell}^2 = \frac{1}{4\pi^2} \cdot \mathbb{E} \left\| D_{x_1}^p D_{x_2}^q u \right\|_{L^2(0,2\pi)}^2$$

This result is suboptimal and cannot be generalized to higher dimensions! But there is room for improvement...
Collaborators

- Sarah Day (Cornell University)
- Marcio Gameiro (Georgia Institute of Technology)
- Bill Kalies (Florida Atlantic University)
- Konstantin Mischaikow (Georgia Institute of Technology)
- Todd Moeller (Los Alamos National Laboratory)