The Annals of Applied Probability
2010, Vol. 20, No. 3, 1068–1097
DOI: 10.1214/09-AAP652

TOPOLOGY-GUIDED SAMPLING OF NONHOMOGENEOUS RANDOM PROCESSES

BY KONSTANTIN MISCHAIKOW[1] AND THOMAS WANNER[2]

Rutgers University and George Mason University

Topological measurements are increasingly being accepted as an important tool for quantifying complex structures. In many applications, these structures can be expressed as nodal domains of real-valued functions and are obtained only through experimental observation or numerical simulations. In both cases, the data on which the topological measurements are based are derived via some form of finite sampling or discretization. In this paper, we present a probabilistic approach to quantifying the number of components of generalized nodal domains of nonhomogeneous random processes on the real line via finite discretizations, that is, we consider excursion sets of a random process relative to a nonconstant deterministic threshold function. Our results furnish explicit probabilistic a priori bounds for the suitability of certain discretization sizes and also provide information for the choice of location of the sampling points in order to minimize the error probability. We illustrate our results for a variety of random processes, demonstrate how they can be used to sample the classical nodal domains of deterministic functions perturbed by additive noise and discuss their relation to the density of zeros.

1. Introduction. The motivation for this work comes from our attempts to create novel metrics for quantifying, comparing and cataloging large sets of complicated varying geometric patterns. Random fields (for a general background, see [1, 4, 12, 19, 22], as well as the references therein) provide a framework in which to approach these problems and have, over the last few decades, emerged as an important tool for studying spatial phenomena which involve an element of randomness [1, 2, 24, 27, 29]. For the types of applications, we have in mind [13, 14, 21], we are often satisfied with a topological classification of sub- or superlevel sets of a scalar function. Algebraic topology, and in particular homology, can be used in a computationally efficient manner [18] to coarsely quantify these geometric properties. In past work [7, 23], we developed a probabilistic framework for assessing the correctness of homology computations for random fields via uniform discretizations. The approach considers the homology of nodal domains of random

Received April 2008; revised October 2009.

[1]Supported in part by NSF Grants DMS-05-11115 and DMS-01-07396, by DARPA and by the U.S. Department of Energy.

[2]Supported in part by NSF Grant DMS-04-06231 and the U.S. Department of Energy under Contract DE-FG02-05ER25712.

AMS 2000 subject classifications. 60G15, 60G17, 55N99.

Key words and phrases. Gaussian process, nodal domains, excursion set, components.

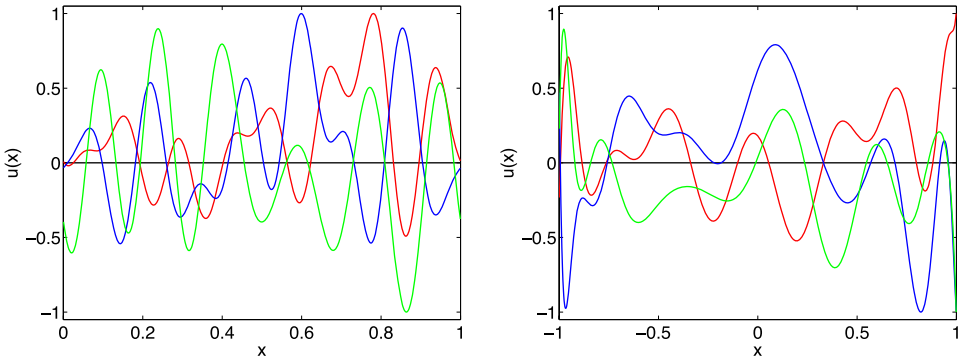

FIG. 1. *Sample functions from a random sum of the form* $u(x,\omega) = \sum_{k=0}^{N} g_k(\omega)\varphi_k(x)$ *where* $g_1, \ldots, g_N$ *are independent standard Gaussian random variables. In the left diagram, we consider random periodic functions, that is, the basis functions* φ_k *are given by* $\varphi_{2k}(x) = \cos(2\pi kx)$ *and* $\varphi_{2k-1}(x) = \sin(2\pi kx)$, *in the right diagram they are the Chebyshev polynomials* $\varphi_k(x) = \cos(k \arccos x)$. *In each case, we choose* $N = 16$.

fields which are given by classical Fourier series in one and two space dimensions, and it provides explicit and sharp error bounds as a function of the discretization size and averaged Sobolev norms of the random field. While we do not claim it is trivial—there are complicated combinatorial questions that need to be resolved—we believe that it is possible to extend the methods and hence the results of [23] to higher-dimensional domains.

The more serious restriction in [23] is the use of periodic random fields, which due to the fact that the associated spatial correlation function is homogeneous, simplifies many of the estimates. In general, however, one expects to encounter nonhomogeneous random fields. In such cases, it seems unreasonable to expect that uniform sampling provides the optimal choice. For example, in Figure 1, three sample functions each are shown for a random sum involving periodic basis functions and Chebyshev polynomials. As one would expect, the zeros of the random Chebyshev sum are more closely spaced at the boundary, and therefore small uniform discretization are most likely not optimal for determining the topology of the nodal domains.

With this as motivation, we allow for a more general sampling technique. We remark that because of the subtlety of some of the necessary estimates we restrict our attention in this paper to one-dimensional domains.

DEFINITION 1.1 (Nonuniform approximation of generalized nodal domains). Consider a compact interval $[a,b] \subset \mathbb{R}$, a threshold function $\mu : [a,b] \to \mathbb{R}$, and a function $u : [a,b] \to \mathbb{R}$. Then we define the *generalized nodal domains of* u by

$$N_\mu^{\pm} = \{x \in [a,b] : \pm(u(x) - \mu(x)) \geq 0\}, \tag{1}$$

which for the case of $\mu(x) \equiv 0$ reduces to the classical definition of a nodal domain in [5]. An *M-discretization of* $[a, b]$ is a collection of $M + 1$ grid points

$$a = x_0 < x_1 < \cdots < x_M = b,$$

and we define $x_{M+1} = x_M = b$ in the following. The *cubical approximations* $Q_M^{\pm}$ of the *generalized nodal domains* $N_\mu^{\pm}$ *of* u are defined as the sets

$$Q_{\mu,M}^{\pm} := \bigcup \{[x_k, x_{k+1}] : \pm\big((u - \mu)(x_k)\big) \geq 0, k = 0, \ldots, M\}.$$

Given a subset $X \subset [a, b]$, let $\beta_0(X)$ denote the number of components of X. Consider a random field $u : [a, b] \times \Omega \to \mathbb{R}$ over the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We are interested in optimally characterizing the topology, that is, determining the number of components, of the nodal domains $N_\mu^{\pm}$ in terms of the cubical approximations $Q_{\mu,M}^{\pm}$. In other words, our goal is to choose the M-discretization of $[a, b]$ in such a way as to optimize

$$\mathbb{P}\{\beta_0(N_\mu^{\pm}) = \beta_0(Q_{\mu,M}^{\pm})\}.$$

We provide two results addressing this question. The first characterizes the choice of the sampling points $a = x_0 < x_1 < \cdots < x_M = b$ under reasonably general abstract conditions. More precisely, consider the following assumptions:

(A1) For every $x \in [a, b]$, we have $\mathbb{P}\{u(x) = \mu(x)\} = 0$.
(A2) The random field is such that $\mathbb{P}\{u - \mu$ has a double zero in $[a, b]\} = 0$.
(A3) For $\sigma \in \{\pm 1\}$, $x \in [a, b]$ and $\delta > 0$ with $x + \delta \in [a, b]$ define

$$p_\sigma(x, \delta) = \mathbb{P}\Big\{\sigma u(x) \geq \sigma \mu(x), \sigma u\Big(x + \frac{\delta}{2}\Big) \leq \sigma \mu\Big(x + \frac{\delta}{2}\Big), \sigma u(x + \delta) \geq \sigma \mu(x + \delta)\Big\}.$$

Then there exists a continuously differentiable function $\mathcal{C}_0 : [a, b] \to \mathbb{R}^+$ as well as a constant $\mathcal{C}_1 > 0$ such that for all $x \in [a, b]$ with $x + \delta \in [a, b]$ we have

$$p_{+1}(x, \delta) + p_{-1}(x, \delta) \leq \mathcal{C}_0(x) \cdot \delta^3 + \mathcal{C}_1 \cdot \delta^4.$$

In Section 3, we prove the following result.

THEOREM 1.2 (Sampling based on local probabilities). *Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a continuous threshold function $\mu : [a, b] \to \mathbb{R}$, and a random field $u : [a, b] \times \Omega \to \mathbb{R}$ over $(\Omega, \mathcal{F}, \mathbb{P})$ such that for $\mathbb{P}$-almost all $\omega \in \Omega$ the function $u(\cdot, \omega) : [a, b] \to \mathbb{R}$ is continuous. Choose the sampling points $a = x_0 < \cdots < x_M = b$ such that*

$$\int_{x_{k-1}}^{x_k} \sqrt[3]{\mathcal{C}_0(x)}\, dx = \frac{1}{M} \cdot \int_a^b \sqrt[3]{\mathcal{C}_0(x)}\, dx \qquad \text{for all } k = 1, \ldots, M,$$

and consider the generalized nodal domains $N_\mu^\pm(\omega)$ *and their approximations* $Q_{\mu,M}^\pm(\omega)$ *as in Definition* 1.1. *If assumptions* (A1), (A2) *and* (A3) *hold, then*

$$\mathbb{P}\{\beta_0(N_\mu^\pm) = \beta_0(Q_{\mu,M}^\pm)\} \geq 1 - \frac{4}{3M^2} \cdot \left(\int_a^b \sqrt[3]{\mathcal{C}_0(x)}\, dx \right)^3 + O\left(\frac{1}{M^3}\right). \tag{2}$$

This theorem is a direct generalization of the corresponding result in ([23], Theorem 1.3). Numerical computations presented in Section 2 suggest that for certain nonhomogeneous random fields this estimate is sharp—and in fact an enormous improvement over the homogeneous result where $\mathcal{C}_0(x)$ is replaced by $\max_{x\in G} \mathcal{C}_0(x)$.

Of course in practice one is interested in applying Theorem 1.2 to specific random fields. This requires the verification of assumptions (A1), (A2) and (A3), preferably in terms of central random field characteristics.

DEFINITION 1.3. For a random field $u : [a, b] \times \Omega \to \mathbb{R}$ over a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we define its *spatial correlation function* $R : [a, b]^2 \to \mathbb{R}$ as

$$R(x, y) = \mathbb{E}\big(\big(u(x) - \mathbb{E}u(x)\big)\big(u(y) - \mathbb{E}u(y)\big)\big) \qquad \text{for all } x, y \in [a, b],$$

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

If the random field is sufficiently smooth, then the derivatives of the spatial correlation function,

$$R_{k,\ell}(x) = \frac{\partial^{k+\ell} R}{\partial x^k\, \partial y^\ell}(x, x), \tag{3}$$

have a natural interpretation in terms of spatial derivatives of the random field u. Since

$$R_{k,\ell}(x) = \mathbb{E}\big(\big(u^{(k)}(x) - \mathbb{E}u^{(k)}(x)\big)\big(u^{(\ell)}(y) - \mathbb{E}u^{(\ell)}(y)\big)\big),$$

the function $R_{k,k}$ contains averaged information on the square of the kth derivative of the random function u, more precisely, its variance.

To relate the spatial correlation function to the function $\mathcal{C}_0$ in Theorem 1.2, we specialize to Gaussian random fields. To be more precise, we make the following assumptions.

(G1) Consider a Gaussian random field $u : [a, b] \times \Omega \to \mathbb{R}$ over a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $u(\cdot, \omega) : [a, b] \to \mathbb{R}$ is twice continuously differentiable for $\mathbb{P}$-almost all $\omega \in \Omega$. Furthermore, assume that for every $x \in [a, b]$ the expected value of $u(x)$ satisfies

$$\mathbb{E}u(x) = 0.$$

(G2) The spatial correlation function R is three times continuously differentiable in a neighborhood of the diagonal $x = y$ and the matrix

$$\mathcal{R}(x) = \begin{pmatrix} R_{0,0}(x) & R_{1,0}(x) & R_{2,0}(x) \\ R_{1,0}(x) & R_{1,1}(x) & R_{2,1}(x) \\ R_{2,0}(x) & R_{2,1}(x) & R_{2,2}(x) \end{pmatrix} \tag{4}$$

is positive definite for all $x \in [a, b]$.

We make considerable use of $\mathcal{R}$, and thus introduce the following notation

$$\begin{aligned} \mathcal{R}^m_{33} &:= R_{0,0}R_{1,1} - R^2_{1,0}, \\ \mathcal{R}^m_{32} &:= R_{0,0}R_{2,1} - R_{1,0}R_{2,0}, \\ \mathcal{R}^m_{31} &:= R_{1,0}R_{2,1} - R_{1,1}R_{2,0}. \end{aligned} \tag{5}$$

These expressions are just the determinants of minors of $\mathcal{R}$. This allows us to state the following theorem.

THEOREM 1.4 (Sampling based on spatial correlation). *Consider a Gaussian random field $u : [a, b] \times \Omega \to \mathbb{R}$ satisfying* (G1) *and* (G2), *and a threshold function $\mu : [a, b] \to \mathbb{R}$ of class C^3. Choose the sampling points $a = x_0 < \cdots < x_M = b$ in such a way that*

$$\int_{x_{k-1}}^{x_k} \sqrt[3]{\mathcal{C}(x)}\, dx = \frac{1}{M} \cdot \int_a^b \sqrt[3]{\mathcal{C}(x)}\, dx \qquad \text{for all } k = 1, \ldots, M,$$

where

$$\mathcal{C}(x) = \frac{\det \mathcal{R}(x)}{48\pi \mathcal{R}^m_{3,3}(x)^{3/2}} \cdot \big(1 + \mathcal{A}(x)\big) \cdot e^{-\mathcal{B}(x)}, \tag{6}$$

given

$$\begin{aligned} \mathcal{A}(x) &= \frac{(\mathcal{R}^m_{3,1}(x)\mu(x) - \mathcal{R}^m_{3,2}(x)\mu'(x) + \mathcal{R}^m_{3,3}(x)\mu''(x))^2}{\mathcal{R}^m_{3,3}(x) \det \mathcal{R}(x)} \geq 0, \\ \mathcal{B}(x) &= \frac{(R_{1,0}(x)\mu(x) - R_{0,0}(x)\mu'(x))^2 + \mathcal{R}^m_{3,3}(x)\mu(x)^2}{2R_{0,0}(x)\mathcal{R}^m_{3,3}(x)} \geq 0. \end{aligned}$$

Let $Q^{\pm}_{\mu,M}(\omega)$ denote the cubical approximations of the random generalized nodal domains $N^{\pm}_{\mu}(\omega)$ of $u(\cdot, \omega)$. Then

$$\mathbb{P}\{\beta_0(N^{\pm}_{\mu}) = \beta_0(Q^{\pm}_{\mu,M})\} \geq 1 - \frac{1}{M^2} \cdot \left(\int_a^b \sqrt[3]{\mathcal{C}(x)}\, dx \right)^3 + O\left(\frac{1}{M^3}\right). \tag{7}$$

The proof of Theorem 1.4 is presented in Section 5. However, it depends on nontrivial results concerning the asymptotic behavior of sign-distribution probabilities of parameter-dependent Gaussian random variables. These results are developed in Section 4.

The number of nodal domains $\beta_0(N_\mu^\pm)$ is clearly dependent upon the zeros of $u - \mu$. Thus, it is reasonable to expect that there is some relationship between the function $\mathcal{C}$ derived in Theorem 1.4 and the density of the zeros of the random field u. The first step is to obtain a density function. For this, a weaker form of (G2) is sufficient.

(G3) Assume that the spatial correlation function R is two times continuously differentiable in a neighborhood of the diagonal $x = y$ and that $R(x, x) > 0$ for all $x \in [a, b]$.

Finding the density of the zeros of random fields has been studied in a variety of settings, see, for example, [2, 4, 6, 11, 12], as well as the references therein. The following theorem can be found in [6], (13.2.1), page 285.

THEOREM 1.5 (Density of zeros of a random field). *Consider a Gaussian random field $u : [a, b] \times \Omega \to \mathbb{R}$ satisfying* (G1) *and* (G3). *Then the density function for the number of zeros of u is given by*

$$\mathcal{D}(x) = \frac{\mathcal{R}^m_{3,3}(x)^{1/2}}{\pi \cdot R_{0,0}(x)}. \tag{8}$$

In other words, for every interval $I \subset [a, b]$ the expected number of zeros of u in I is given by $\int_I \mathcal{D}(x)\, dx$.

While Theorem 1.5 has been known for quite some time, its implications are surprising. As is demonstrated through examples in Section 2 there is no simple discernible relationship between the function $\mathcal{C}^{1/3}$ of Theorem 1.4 and the density function $\mathcal{D}$.

As is made clear at the beginning of this Introduction, our motivation is to develop optimal sampling methods for the analysis of complicated time-dependent patterns. Thus, before turning to the proofs of the above-mentioned results, we begin, in Section 2, with demonstrations of possible applications and implications of Theorem 1.4. In particular, we consider several random generalized Fourier series $u : [a, b] \times \Omega \to \mathbb{R}$ defined by

$$u(x, \omega) = \sum_{k=0}^{\infty} g_k(\omega) \cdot \varphi_k(x), \tag{9}$$

where $\varphi_k : [a, b] \to \mathbb{R}$, $k \in \mathbb{N}_0$, denotes a family of smooth functions and we assume that the Gaussian random variables $g_k : \Omega \to \mathbb{R}$, $k \in \mathbb{N}_0$, are defined over a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with mean 0.

We conclude the paper with a general discussion of future work concerning natural generalizations to higher dimensions.

2. Sampling of specific random sums. To demonstrate the applicability and implications of Theorem 1.4, we consider in this section several random generalized Fourier series $u : [a, b] \times \Omega \to \mathbb{R}$ of the form in (9). As mentioned before, the functions $\varphi_k : [a, b] \to \mathbb{R}$, $k \in \mathbb{N}_0$, denote a family of smooth functions and we assume that the random variables $g_k : \Omega \to \mathbb{R}$, $k \in \mathbb{N}_0$, are Gaussian with vanishing mean, and defined over a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We would like to point out that these random variables do not need to be independent, and we define

$$\alpha_{k,m} = \mathbb{E}(g_k g_m) \qquad \text{for all } k, m \in \mathbb{N}_0.$$

Then one can easily show that

$$R_{k,\ell}(x) = \mathbb{E}\big(u^{(k)}(x)u^{(\ell)}(x)\big) = \sum_{i,j=0}^{\infty} \alpha_{i,j}\varphi_i^{(k)}(x)\varphi_j^{(\ell)}(x).$$

If in addition the random variables g_k are pairwise independent, then we have

$$R_{k,\ell}(x) = \sum_{j=0}^{\infty} \alpha_{j,j}\varphi_j^{(k)}(x)\varphi_j^{(\ell)}(x),$$

where $\alpha_{j,j} \geq 0$ for all $j \in \mathbb{N}_0$. One can show that this diagonalization can always be achieved for Gaussian random fields, provided the basis functions φ_k are chosen appropriately. For more details, we refer the reader to [2], Theorems 3.1.1 and 3.1.2, Lemma 3.1.4.

Within the above framework of random generalized Fourier series, we specifically consider several classes:

- *Random Chebyshev polynomials* $u : [-1, 1] \times \Omega \to \mathbb{R}$ of the form

$$u(x, \omega) = \sum_{k=0}^{N} g_k(\omega) \cdot \cos(k \arccos x) \qquad \text{with } \mathbb{E}(g_k g_\ell) = \delta_{k,\ell}. \tag{10}$$

- *Random cosine series* $u : [0, 1] \times \Omega \to \mathbb{R}$ of the form

$$u(x, \omega) = \sum_{k=0}^{N} g_k(\omega) \cdot \cos(k\pi x) \qquad \text{with } \mathbb{E}(g_k g_\ell) = \delta_{k,\ell}. \tag{11}$$

- *Random L-periodic functions* $u : \mathbb{R} \times \Omega \to \mathbb{R}$ of the form

$$u(x, \omega) = \sum_{k=0}^{\infty} a_k \cdot \left(g_{2k}(\omega) \cdot \cos \frac{2\pi kx}{L} + g_{2k-1}(\omega) \cdot \sin \frac{2\pi kx}{L} \right) \qquad \text{with } \mathbb{E}(g_k g_\ell) = \delta_{k,\ell}, \tag{12}$$

with real constants a_k.

- *Random polynomials* $u:[-3,3]\times\Omega\to\mathbb{R}$ *with Gaussian coefficients of binomial variance* of the form

$$u(x,\omega)=\sum_{k=0}^{N} g_k(\omega)\cdot x^k \qquad \text{with } \mathbb{E}(g_k g_\ell)=\delta_{k,\ell}\cdot\binom{N}{k}, \tag{13}$$

- *Random polynomials* $u:[-3,3]\times\Omega\to\mathbb{R}$ *with Gaussian coefficients of unit variance* of the form

$$u(x,\omega)=\sum_{k=0}^{N} g_k(\omega)\cdot x^k \qquad \text{with } \mathbb{E}(g_k g_\ell)=\delta_{k,\ell}. \tag{14}$$

As is indicated in Section 1, we assume that all the random coefficients are centered Gaussian random variables over a common probability space $(\Omega,\mathcal{F},\mathbb{P})$.

2.1. *The case of vanishing threshold function.* We begin our applications by thresholding sample random sums at their expected value, that is, we use the threshold function $\mu\equiv 0$. In this particular case, the function $\mathcal{C}(x)$ defined by (6) in Theorem 1.4 simplifies to

$$\mathcal{C}(x)=\frac{\det\mathcal{R}(x)}{48\pi\,\mathcal{R}^m_{3,3}(x)^{3/2}}, \tag{15}$$

since both $\mathcal{A}(x)$ and $\mathcal{B}(x)$ vanish.

For the case of random Chebyshev polynomials (10), the left diagram in Figure 2 shows three normalized sample functions

$$\frac{\mathcal{C}^{1/3}(x)}{\int_{-1}^{1}\mathcal{C}^{1/3}(x)\,dx}$$

for $N=3,5,10$. The right diagram shows the expected number of zeros of the random Chebyshev polynomials as a function of N (red curve), which grows proportional to N. Thus, in order to sample the random field sufficiently fine, we expect to use significantly more than $O(N)$ discretization points. The blue curve in the right diagram of Figure 2 shows the values of M for which the bound in (7) of Theorem 1.4 implies a correctness probability of 95%, and a least squares fit of this curve furnishes $M\sim N^{3/2}$. For comparison, the green curve in the same diagram shows the values of M for which the bound in our previous result ([23], Theorem 1.4) implies a correctness probability of 95%, provided we apply this theorem with $\mathcal{C}_0$ given as the $\max_{x\in[-1,1]}\mathcal{C}_0(x)$. Notice that in this case we have $M\sim N^3$. In other words, only the topology-guided sampling result of the current paper yields a reasonable growth for the number of sampling points. In fact, based on our results for periodic random fields in [23] and the numerical simulations in [7], we expect that $M\sim N^{3/2}$ is the optimal discretization size.

For the case of random cosine sums (11), that is, random trigonometric sums satisfying homogeneous Neumann boundary conditions, the analogue of the right

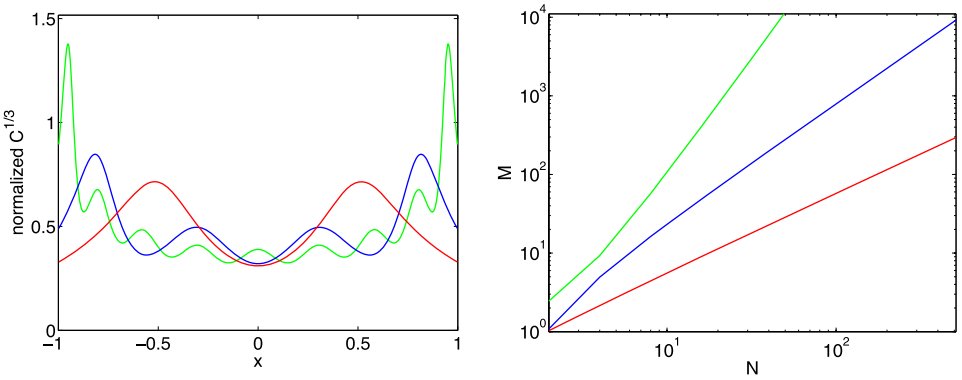

FIG. 2. *Topology-guided sampling of random Chebyshev polynomials* (10). *The left diagram shows the functions* $C^{1/3}$ *for* $N = 3, 5, 10$ (*red, blue and green, respectively—increasing values of* N *increase the number of extrema*); *for comparison reasons, each curve has been scaled in such a way that the area under the graph is one. The right diagram shows the expected number of zeros of the random Chebyshev polynomials as a function of* N (*bottom red curve*), *the value of* M *for which Theorem* 1.4 *gives a correctness probability of* 95% (*middle blue curve*), *and the value of* M *for which* [23] *gives a correctness probability of* 95% (*top green curve*) *with* $\mathcal{C}_0 = \max \mathcal{C}_0(x)$.

diagram in Figure 2 is depicted in the left diagram of Figure 3. Notice that for the random cosine sums the expected number of zeros is proportional to N, and the required number of sampling points has to be proportional to $N^{3/2}$ for both Theorem 1.4 and [23], Theorem 1.4. In other words, in this situation the gains from topology-guided sampling are no longer as large as in the context of Chebyshev

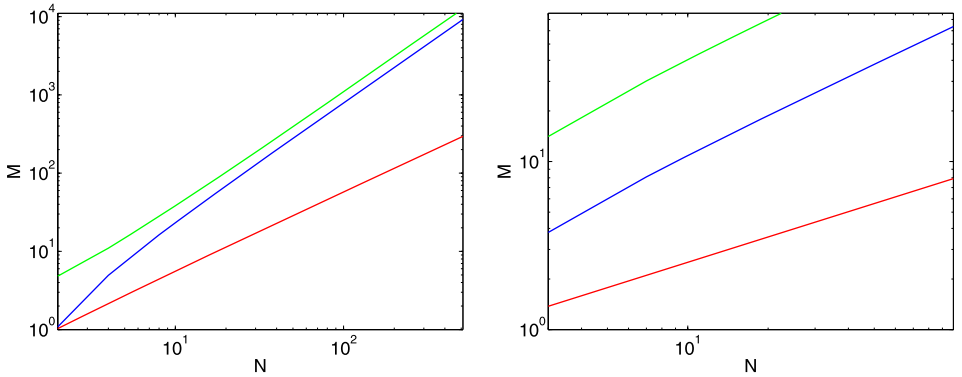

FIG. 3. *Topology-guided sampling of random trigonometric polynomials* (11) *satisfying Neumann boundary conditions* (*left diagram*) *and random algebraic polynomials* (13) *with binomial variances* (*right diagram*). *The curves show the expected numbers of zeros* (*bottom red curve*), *the discretization size required by Theorem* 1.4 *to achieve* 95% *correctness* (*middle blue curve*), *and the discretization size required by* [23] *for a correctness probability of* 95% (*top green curve*), *with* $\mathcal{C}_0 = \max \mathcal{C}_0(x)$.

polynomials. Also in this case, the curves for M are obtained in such a way that the right-hand side in (7) or the corresponding bound in [23] equals 95%

Similar behavior can be seen in the case of random polynomials (13) with Gaussian coefficients of binomial variance; see the right diagram of Figure 3. For the random algebraic polynomials (13), one can show that the expected number of zeros is proportional to $N^{1/2}$, and the required number of sampling points implied by (7) or [23] has to be proportional to $N^{3/4}$ for both results. In fact, the function $\mathcal{C}$ can be computed explicitly in this case. Due to (13), the spatial correlation function R is given by

$$R(x, y) = \sum_{k=0}^{N} \binom{N}{k} x^k y^k = (1 + xy)^N,$$

which after some elementary computations furnishes

$$\mathcal{C}(x) = \frac{N^{1/2}(N-1)}{24\pi(1+x^2)^3}. \tag{16}$$

As for the case of random polynomials with Gaussian coefficients of unit variance, a classical result due to Kac [16, 17] implies that the expected number of zeros is proportional to $\log N$. In this case, Theorem 1.4 implies that the required number of sampling points has to be proportional to $(\log N)^{3/2}$.

2.2. *The case of constant threshold function.* We now turn our attention to a constant threshold function $\mu(x) = \tau$, for some real number τ. In this case, the function $\mathcal{C}(x)$ in Theorem 1.4 simplifies to

$$\mathcal{C}(x) = \frac{\det \mathcal{R}(x)}{48\pi \mathcal{R}^m_{3,3}(x)^{3/2}} \cdot \mathcal{S}(x), \tag{17}$$

where

$$\mathcal{S}(x) = \left(1 + \frac{\mathcal{R}^m_{3,1}(x)^2\tau^2}{\mathcal{R}^m_{3,3}(x)\det \mathcal{R}(x)}\right) \cdot \exp\left(-\frac{R_{1,0}(x)^2 + \mathcal{R}^m_{3,3}(x)}{2R_{0,0}(x)\mathcal{R}^m_{3,3}(x)} \cdot \tau^2\right). \tag{18}$$

For large values of $|\tau|$, the scaling function $\mathcal{S}(x)$ will be close to zero, and it therefore effectively decreases the probability for mistakes in the homology computation. In fact, it decreases exponentially fast with respect to $|\tau|$. However, as is shown in Figure 4 for the random Chebyshev polynomials (10), for values of τ close to zero, there can be regions in which the probability for mistakes actually increases. This behavior is even more pronounced in the case of random algebraic polynomials (13) and (14), which is shown in Figure 5.

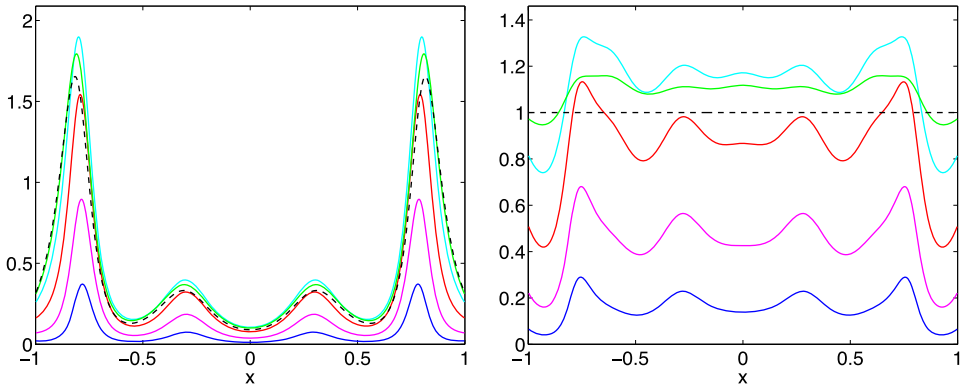

FIG. 4. *Effect of varying the threshold* τ *on the function* $\mathcal{C}(x)$ *in* (17) *for random Chebyshev polynomials* (10) *with* $N = 5$. *The left diagram shows the function* $\mathcal{C}(x)$ *for* $\tau = 0, 1, 2, 3, 4, 5$ (*black*, *green*, *cyan*, *red*, *magenta*, *blue*), *the right diagram shows only the function* $\mathcal{S}(x)$ *defined in* (18).

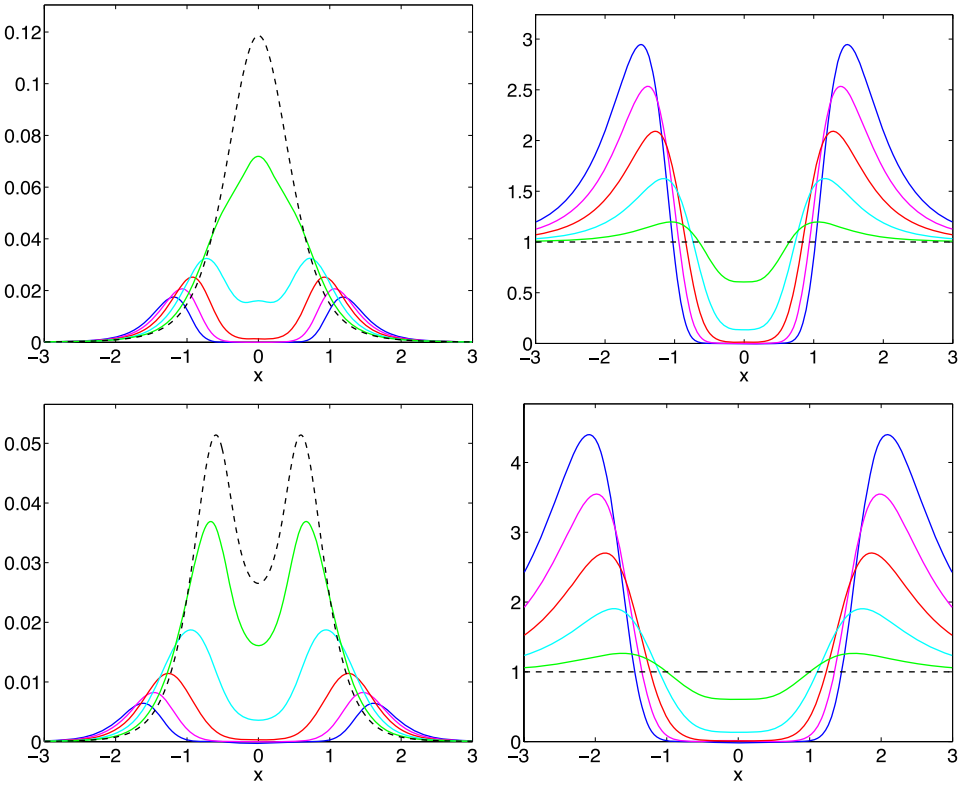

FIG. 5. *Effect of varying the threshold* τ *on the function* $\mathcal{C}(x)$ *in* (17) *for random algebraic polynomials* (13) (*top row*) *and* (14) (*bottom row*) *with* $N = 5$. *In each row, the left diagram shows* $\mathcal{C}(x)$ *for* $\tau = 0, 1, 2, 3, 4, 5$ (*black*, *green*, *cyan*, *red*, *magenta*, *blue*), *and the right diagram shows only the function* $\mathcal{S}(x)$ *defined in* (18).

2.3. *The case of varying threshold function.* We now consider the case of a general threshold function under the following assumptions. Suppose a deterministic function $\mu(x)$ is perturbed by a centered Gaussian random field $u(x,\omega)$, and that we are interested in determining the classical nodal domains of the sum

$$v(x,\omega) = \mu(x) + u(x,\omega).$$

Sampling $v(x,\omega)$ at the threshold zero is obviously equivalent to sampling $u(x,\omega)$ at the threshold $-\mu(x)$. Thus, we can use Theorem 1.4 to find the optimal location of the sampling points using the function $\mathcal{C}(x)$ defined in (6).

In order to demonstrate the effects of the varying threshold function $-\mu(x)$ more clearly, we now assume that the perturbing random field u is homogeneous, that is, we assume that u is a random L-periodic function of the form (12). Furthermore, we assume that the real scaling factors a_k in (12) satisfy

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

and that at least two of the a_k do not vanish. It was shown in [23] that in this case the spatial correlation function R is given by

$$R(x,y) = \mathbb{E}u(x)u(y) = \sum_{k=0}^{\infty} a_k^2 \cdot \cos\frac{2\pi k(x-y)}{L}.$$

From this, one can readily see that the matrix function $\mathcal{R}(x)$ defined in (4) is constant and given by

$$\mathcal{R}(x) = \begin{pmatrix} A_0 & 0 & -\dfrac{4\pi^2 A_1}{L^2} \\ 0 & \dfrac{4\pi^2 A_1}{L^2} & 0 \\ -\dfrac{4\pi^2 A_1}{L^2} & 0 & \dfrac{16\pi^4 A_2}{L^4} \end{pmatrix},$$

where

$$A_\ell = \sum_{k=0}^{\infty} k^{2\ell} a_k^2.$$

Thus, the function $\mathcal{C}(x)$ in (6) is now given as

$$\mathcal{C}(x) = \frac{\pi^2}{6L^3} \cdot \frac{A_0A_2 - A_1^2}{A_0^{3/2}A_1^{1/2}} \cdot \mathcal{S}(x), \tag{19}$$

where

$$\mathcal{S}(x) = \left(1 + \frac{(A_1\mu(x) + A_0\mu''(x)\cdot(L^2/(4\pi^2)))^2}{A_0(A_0A_2 - A_1^2)}\right) \times \exp\left(-\frac{A_1\mu(x)^2 + A_0\mu'(x)^2\cdot(L^2/(4\pi^2))}{2A_0A_1}\right).$$

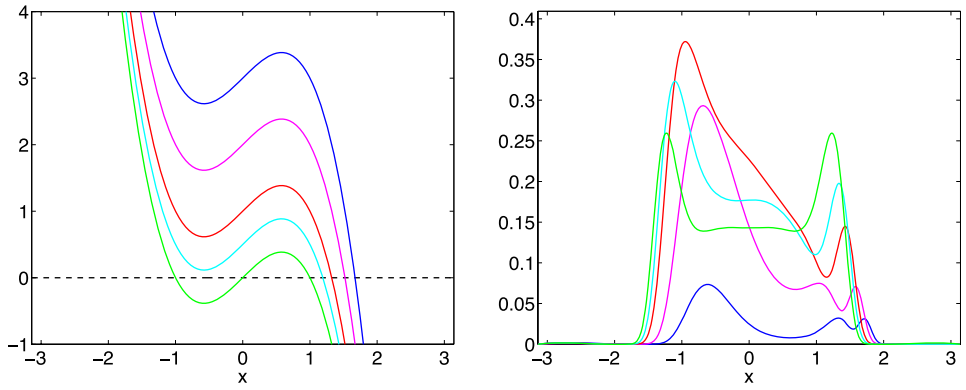

FIG. 6. *Sampling of deterministic functions $\mu(x)$ perturbed by homogeneous random noise. The left image shows the functions $\mu(x) = x - x^3 + \tau$ for $\tau = 0, 0.5, 1, 2, 3$ (green, cyan, red, magenta, blue), the right images shows the corresponding functions $\mathcal{C}(x)$ defined in* (19).

Notice that the exponential factor is bounded above by $\exp(-\mu(x)^2/(2A_0))$, that is, large function values of $\mu(x)$ lead to small failure probabilities.

We close this subsection by visualizing the function $\mathcal{C}(x)$ defined in (19) for the deterministic function $\mu(x) = x - x^3 + \tau$ and τ-values between 0 and 3. The specific functions $\mu(x)$ are shown in the left image of Figure 6. In the right image, the corresponding functions $\mathcal{C}(x)$ are shown, where u is defined as in (12) with $a_k = 0$ for $k = 0$ and $k > N$, as well as $a_k = N^{-1/2}$ for $k = 1, \ldots, N$. This implies that the variance of $u(x)$ equals 1. In Figure 6, we use $N = 5$.

2.4. *Comparison with density-guided sampling.* In order to illustrate the differences between the density of zeros $\mathcal{D}$ derived in Theorem 1.5 and the function $\mathcal{C}^{1/3}$ from Theorem 1.4, we return to our examples from the last section. For each of these examples, Figure 7 depicts both

$$\frac{\mathcal{C}^{1/3}(x)}{\int \mathcal{C}^{1/3}(x)\,dx} \quad \text{and} \quad \frac{\mathcal{D}(x)}{\int \mathcal{D}(x)\,dx}$$

for the case $N = 5$. It is evident from these graphs that in most cases, the homology-based sampling density is different from the actual density of zeros. In fact, in many cases it behaves anticyclic to $\mathcal{D}$ in the sense that the local extrema of $\mathcal{C}^{1/3}$ alternate with the local extrema of $\mathcal{D}$.

There are, however, exceptions, as the case of the random algebraic polynomial (13) demonstrates. In this case, it follows from Theorem 1.5 that

$$\mathcal{D}(x) = \frac{N^{1/2}}{\pi(1 + x^2)},$$

and together with (16) this shows that the normalized $\mathcal{C}^{1/3}$- and $\mathcal{D}$-functions coincide.

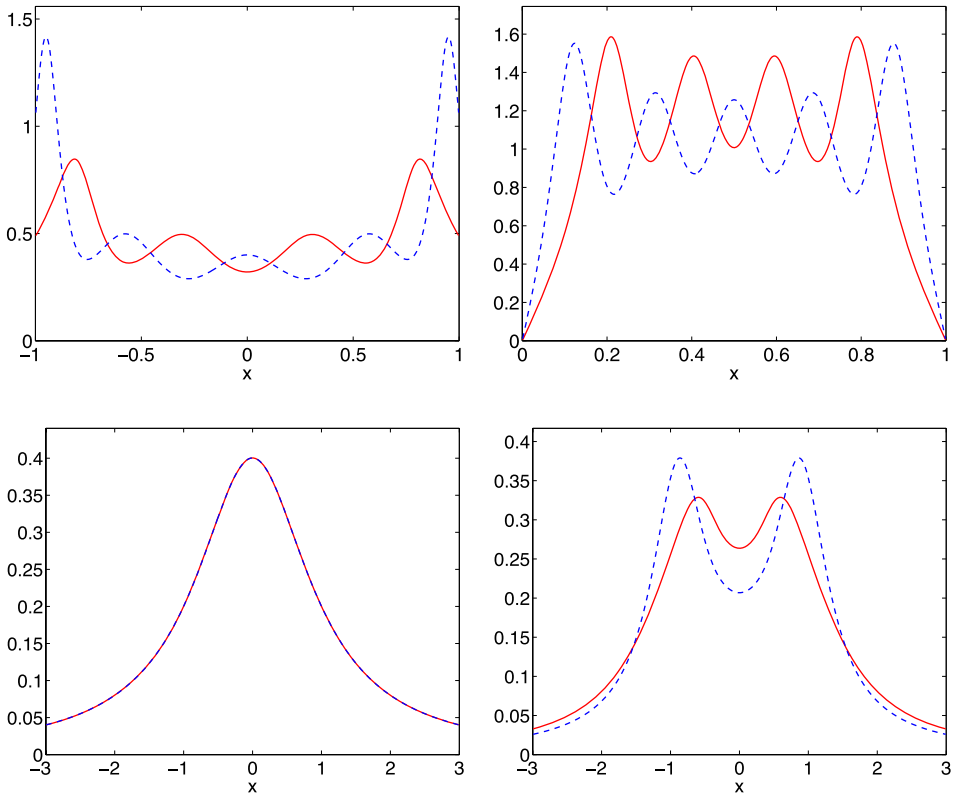

FIG. 7. *A comparison of the function* $\mathcal{C}^{1/3}$ *and the density function* $\mathcal{D}$ *for random Chebyshev polynomials* (10) (*top left diagram*), *random trigonometric polynomials* (11) (*top right*), *random algebraic polynomials* (13) (*bottom left*), *and random algebraic polynomials* (14) (*bottom right*). *In all cases, the areas under the graphs have been normalized to one, and we chose* $N = 5$.

3. Sampling based on local probabilities. The goal of this section is the proof of Theorem 1.2, which is a generalization of [23], Theorem 1.3. Thus, we begin by recalling some basic definitions and results.

As is indicated in Section 1, given a continuous function $u : [a, b] \to \mathbb{R}$ and a continuous threshold $\mu : [a, b] \to \mathbb{R}$ we are interested in determining the number of components of the generalized nodal domain $N_\mu^\pm$ in terms of a cubical approximation $Q_{\mu,M}^\pm$ obtained via sampling at $M + 1$ points as described in Definition 1.1. For suitably chosen discretization points, and under appropriate regularity and nondegeneracy conditions on u one can then expect that the number of components of $Q_{\mu,M}^\pm$ and $N_\mu^\pm$ agree. One only has to be able to verify that the function u has at most one zero (counting multiplicity) in each of the intervals $[x_{k-1}, x_k]$, for $k = 1, \ldots, M$. This is accomplished using the following framework which goes back to Dunnage [10].

DEFINITION 3.1. A continuous function $u : [a, b] \to \mathbb{R}$ has a *double crossover on the interval* $[\alpha, \beta] \subset [a, b]$, if

$$\sigma \cdot u(\alpha) \geq 0, \qquad \sigma \cdot u\left(\frac{\alpha+\beta}{2}\right) \leq 0 \quad \text{and} \quad \sigma \cdot u(\beta) \geq 0 \tag{20}$$

for one choice of the sign $\sigma \in \{\pm 1\}$.

DEFINITION 3.2. Let $u : [a, b] \to \mathbb{R}$ be a continuous function.

- The *dyadic points* in the interval $[\alpha, \beta]$ are defined as

$$d_{n,k} = \alpha + (\beta - \alpha) \cdot \frac{k}{2^n} \qquad \text{for all } k = 0, \ldots, 2^n \text{ and } n \in \mathbb{N}_0.$$

  The *dyadic subintervals* of $[\alpha, \beta]$ are the intervals $[d_{n,k}, d_{n,k+1}]$ for all $k = 0, \ldots, 2^n - 1$ and $n \in \mathbb{N}_0$.
- The interval $[\alpha, \beta] \subset [a, b]$ is *admissible* for u, if the function u does not have a double crossover on any of the dyadic subintervals of $[\alpha, \beta]$.

It was shown in [23] that the concept of admissibility implies the suitability of our nodal domain approximations. More precisely, the following is a slight rewording of [23], Proposition 2.5.

PROPOSITION 3.3 (Validation criterion). *Let $u : [a, b] \to \mathbb{R}$ be a continuous function and let $\mu : [a, b] \to \mathbb{R}$ be a continuous threshold function. Let $N_\mu^\pm$ denote the generalized nodal domains of u, and let $Q_{\mu,M}^\pm$ denote their cubical approximations as in Definition 1.1. Furthermore, assume that the following hold*:

(a) *The function $u - \mu$ is nonzero at all grid points x_k, for $k = 0, \ldots, M$.*
(b) *The function $u - \mu$ has no double zero in (a, b), that is, if $x \in (a, b)$ is a zero of u, then $u - \mu$ attains both positive and negative function values in every neighborhood of x.*
(c) *For every $k = 1, \ldots, M$, the interval $[x_{k-1}, x_k]$ between consecutive discretization points is admissible for $u - \mu$ in the sense of Definition 3.2.*

Then we have

$$\beta_0(N_\mu^\pm) = \beta_0(Q_{\mu,M}^\pm).$$

The following lemma provides bounds on the probability for admissibility of a given interval.

LEMMA 3.4. *Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a continuous threshold function $\mu : [a, b] \to \mathbb{R}$, and a random field $u : [a, b] \times \Omega \to \mathbb{R}$ over $(\Omega, \mathcal{F}, \mathbb{P})$ such*

that $u(\cdot,\omega)$ *is continuous for* $\mathbb{P}$*-almost all* $\omega\in\Omega$. *In addition, assume that* (A1), (A2) *and* (A3) *hold. If* $[x,x+\delta]\subset[a,b]$, *then*

$$\begin{aligned}&\mathbb{P}([x,x+\delta]\ \textit{is not admissible for}\ u-\mu)\\&\quad\le\frac{4\mathcal{C}_0(x)}{3}\cdot\delta^3+\left(\frac{4L}{3}+\frac{8\mathcal{C}_1}{7}\right)\cdot\delta^4,\end{aligned}\tag{21}$$

where $L=\max\{|\mathcal{C}_0'(y)|:y\in[a,b]\}$.

PROOF. If the interval $I=[x,x+\delta]$ is not admissible, then the function $u-\mu$ has a double crossover on one of its dyadic subintervals. If we now denote the dyadic points in I by $d_{n,k}$ as in Definition 3.2, then together with (A3) one obtains the estimate

$$\begin{aligned}\mathbb{P}\{I\text{ is not admissible}\}&\le\sum_{n=0}^{\infty}\sum_{k=0}^{2^n-1}\left(p_{+1}(d_{n,k},\delta/2^n)+p_{-1}(d_{n,k},\delta/2^n)\right)\\&\le\sum_{n=0}^{\infty}\sum_{k=0}^{2^n-1}\left(\mathcal{C}_0(d_{n,k})\cdot\left(\frac{\delta}{2^n}\right)^3+\mathcal{C}_1\cdot\left(\frac{\delta}{2^n}\right)^4\right).\end{aligned}$$

Since $\mathcal{C}_0$ is continuously differentiable, we can define $L=\max\{|\mathcal{C}_0'(y)|:y\in[a,b]\}$, and the definition of the dyadic points implies

$$\mathcal{C}_0(d_{n,k})\le\mathcal{C}_0(x)+L\cdot(d_{n,k}-x)\le\mathcal{C}_0(x)+L\delta.$$

This finally furnishes

$$\begin{aligned}\mathbb{P}\{I\text{ is not admissible}\}&\le\sum_{n=0}^{\infty}\sum_{k=0}^{2^n-1}\left((\mathcal{C}_0(x)+L\delta)\cdot\left(\frac{\delta}{2^n}\right)^3+\mathcal{C}_1\cdot\left(\frac{\delta}{2^n}\right)^4\right)\\&=\frac{4\mathcal{C}_0(x)}{3}\cdot\delta^3+\left(\frac{4L}{3}+\frac{8\mathcal{C}_1}{7}\right)\cdot\delta^4.\end{aligned}$$

□

Combining Proposition 3.3, Lemma 3.4, and restricting to the leading order term in (21), one obtains

$$\mathbb{P}\{\beta_0(N_\mu^{\pm})=\beta_0(Q_{\mu,M}^{\pm})\}\ge1-\frac{4}{3}\cdot\sum_{k=1}^{M}\mathcal{C}_0(x_{k-1})\cdot(x_k-x_{k-1})^3.\tag{22}$$

Clearly, the resulting bound depends on the location of the sampling points, which suggests maximizing the bound to optimize the location.

We first provide a heuristic argument for this optimal location, and present the precise result afterwards. One can show that for arbitrary nonnegative numbers $\delta_1,\dots,\delta_M\ge0$ the inequality

$$\sum_{k=1}^{M}\delta_k^3\ge\frac{1}{M^2}\cdot\left(\sum_{k=1}^{M}\delta_k\right)^3$$

holds, with equality if and only if $\delta_1 = \delta_2 = \cdots = \delta_M$. Applying this inequality to the sum in the right-hand side of (22), implies

$$\sum_{k=1}^{M} \mathcal{C}_0(x_{k-1}) \cdot (x_k - x_{k-1})^3 \geq \frac{1}{M^2} \cdot \left(\sum_{k=1}^{M} \sqrt[3]{\mathcal{C}_0(x_{k-1})} \cdot (x_k - x_{k-1}) \right)^3 \tag{23}$$

with equality if and only if

$$\sqrt[3]{\mathcal{C}_0(x_{k-1})} \cdot (x_k - x_{k-1}) = \sqrt[3]{\mathcal{C}_0(x_{\ell-1})} \cdot (x_\ell - x_{\ell-1}) \quad \text{for all } k, \ell = 1, \ldots, M. \tag{24}$$

For large M, the sum on the right-hand side of (23) converges to the integral of $\mathcal{C}_0^{1/3}$ over $[a, b]$. The motivation for Theorem 1.2 is now clear: Condition (24) suggests that for $M \to \infty$, the optimal estimate can be achieved by choosing the sampling points in an equi-$\mathcal{C}_0^{1/3}$-area fashion, since the term $\mathcal{C}_0(x_{k-1})^{1/3}(x_k - x_{k-1})$ approximates the intergral of $\mathcal{C}_0^{1/3}$ over $[x_{k-1}, x_k]$. This heuristic forms the basis for the following proof of our first main result.

PROOF OF THEOREM 1.2. Let $\delta_{\max} := \max_{k=1,\ldots,M} |x_k - x_{k-1}|$, and define the positive number $m := \min_{x\in[a,b]} \mathcal{C}_0(x)^{1/3} > 0$. Furthermore, let $L := \max_{x\in[a,b]} |d\mathcal{C}_0^{1/3}/dx|$. Then the mean value theorem readily furnishes

$$\left| \sqrt[3]{\mathcal{C}_0(x_{k-1})} \cdot (x_k - x_{k-1}) - \int_{x_{k-1}}^{x_k} \sqrt[3]{\mathcal{C}_0(x)}\, dx \right| \leq L(x_k - x_{k-1})^2 \tag{25}$$

for all $k = 1, \ldots, M$. Due to the choice of the sampling points we further have

$$m \cdot (x_k - x_{k-1}) \leq \int_{x_{k-1}}^{x_k} \sqrt[3]{\mathcal{C}_0(x)}\, dx = \frac{1}{M} \cdot \underbrace{\int_a^b \sqrt[3]{\mathcal{C}_0(x)}\, dx}_{=:K}, \tag{26}$$

which in turn implies

$$0 < x_k - x_{k-1} \leq \delta_{\max} \leq \frac{K}{m \cdot M} \qquad \text{for all } k = 1, \ldots, M. \tag{27}$$

Applying Lemma 3.4 to every subinterval formed by adjacent sampling points, we now obtain together with (25), (26) and (27) the estimate

$$\begin{aligned} 1 - \mathbb{P}\{\beta_0(N_\mu^\pm) = \beta_0(Q_{\mu,M}^\pm)\} &\leq \frac{4}{3} \sum_{k=1}^{M} \mathcal{C}_0(x_{k-1}) \cdot (x_k - x_{k-1})^3 + \mathcal{C}_2 \sum_{k=1}^{M} (x_k - x_{k-1})^4 \\ &\leq \frac{4}{3} \cdot \sum_{k=1}^{M} \left(\frac{K}{M} + L(x_k - x_{k-1})^2 \right)^3 + \frac{\mathcal{C}_2 K^4}{m^4 M^3} \end{aligned}$$

$$\leq \frac{4}{3} \cdot \sum_{k=1}^{M} \left(\frac{K}{M} + \frac{LK^2}{m^2 M^2} \right)^3 + \frac{\mathcal{C}_2 K^4}{m^4 M^3}$$

$$= \frac{4K^3}{3M^2} + O\left(\frac{1}{M^3} \right)$$

for some constant $\mathcal{C}_2 \geq 0$. This is exactly (2). $\square$

4. Asymptotics of sign-change probabilities. Theorem 1.4 can be viewed as a special case of Theorem 1.2. The content lies in the fact that under the assumption of a Gaussian random field, the function $\mathcal{C}_0$ can be explicitly computed. However, this requires a quantitative understanding of the asymptotic behavior of sign-distribution probabilities of parameter-dependent Gaussian random variables, which is the focus of this section.

More precisely, let $T(\delta) = (T_1(\delta), \ldots, T_n(\delta))^t \in \mathbb{R}^n$ denote a one-parameter family of $\mathbb{R}^n$-valued random Gaussian variables over a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, indexed by $\delta > 0$, and choose a sign sequence $(s_1, \ldots, s_n) \in \{\pm 1\}^n$. Furthermore, let $\tau(\delta) \in \mathbb{R}^3$ denote an arbitrary threshold vector. We are interested in the precise asymptotic behavior as $\delta \to 0$ of the probability

$$P(\delta) = \mathbb{P}\big\{ s_j \big(T_j(\delta) - \tau_j(\delta) \big) \geq 0 \text{ for all } j = 1, \ldots, n \big\}. \tag{28}$$

The following result is an extension of ([23], Proposition 4.1) which dealt only with the special case $\tau \equiv 0$.

PROPOSITION 4.1. *Let $(s_1, \ldots, s_n) \in \{\pm 1\}^n$ denote a fixed sign sequence, and consider one-parameter families of a threshold vector $\tau(\delta) \in \mathbb{R}^3$ and an $\mathbb{R}^n$-valued random Gaussian variable $T(\delta)$ over a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, for $\delta > 0$. Assume that the following hold*:

(i) *For each $\delta > 0$, assume that the Gaussian random variable $T(\delta)$ has mean $0 \in \mathbb{R}^n$ and a positive definite covariance matrix $C(\delta) \in \mathbb{R}^{n \times n}$, whose positive eigenvalues are given by $0 < \lambda_1(\delta) \leq \cdots \leq \lambda_n(\delta)$. The corresponding orthonormalized eigenvectors are denoted by $v_1(\delta), \ldots, v_n(\delta)$.*

(ii) *There exists a vector $\bar{v}_1 = (\bar{v}_{11}, \ldots, \bar{v}_{1n})^t \in \mathbb{R}^n$ such that $v_1(\delta) \to \bar{v}_1$ as $\delta \to 0$, and $s_j \cdot \bar{v}_{1j} > 0$ for all $j = 1, \ldots, n$.*

(iii) *The quotient $\lambda_1(\delta)/\lambda_k(\delta)$ converges to 0 as $\delta \to 0$, for all $k = 2, \ldots, n$.*

(iv) *There exists a vector $\alpha = (\alpha_1, \ldots, \alpha_n)^t \in \mathbb{R}^n$ such that*

$$\lim_{\delta \to 0} \frac{\tau(\delta) \cdot v_k(\delta)}{\lambda_k(\delta)^{1/2}} = \alpha_k \qquad \text{for all } k = 1, \ldots, n. \tag{29}$$

Furthermore, for α as above define

$$S_\alpha = \frac{2}{2^{n/2} \cdot \Gamma(n/2)} \cdot e^{-\sum_{k=2}^{n} \alpha_k^2/2} \cdot \int_{\alpha_1}^{\infty} (s - \alpha_1)^{n-1} e^{-s^2/2} \, ds. \tag{30}$$

Then the probability $P(\delta)$ *defined in* (28) *satisfies*

$$\lim_{\delta\to 0} P(\delta)\cdot\sqrt{\frac{\det C(\delta)}{\lambda_1(\delta)^n}} = \frac{\Gamma(n/2)\cdot S_\alpha}{2\cdot\pi^{n/2}\cdot(n-1)!}\cdot\left|\prod_{j=1}^{n}\bar{v}_{1j}\right|^{-1}. \tag{31}$$

For specific values of n, the integral in (29) can be simplified further. For our one-dimensional application, we need the case $n=3$, which is the subject of the following remark.

REMARK 4.2. *Recall that* $\Gamma(1/2)=\pi^{1/2}$, $\Gamma(1)=1$, *and* $\Gamma(t+1)=t\Gamma(t)$ *for* $t>0$. *Furthermore, notice that* $S_\alpha = 1$ *for* $\alpha = 0\in\mathbb{R}^n$. *In addition, for* $n=3$ *one can readily verify that*

$$\begin{aligned} S_\alpha &= \frac{2^{1/2}}{\pi^{1/2}}\cdot e^{-(\alpha_2^2+\alpha_3^2)/2} \\ &\quad\times\left(-\alpha_1 e^{-\alpha_1^2/2}+(1+\alpha_1^2)\cdot\int_{\alpha_1}^{\infty} e^{-s^2/2}\,ds\right). \end{aligned} \tag{32}$$

PROOF. Define the diagonal matrix $S=(s_i\delta_{ij})_{i,j=1,\dots,n}$, where δ_{ij} denotes the Kronecker delta, and let $Z_+=\{z\in\mathbb{R}^n : z_j\geq 0 \text{ for } j=1,\dots,n\}$. Finally, let

$$D(\delta)=\lambda_1(\delta)\cdot SC(\delta)^{-1}S$$

and

$$d(\delta)=\frac{1}{\lambda_1(\delta)^{1/2}}\cdot S\tau(\delta).$$

Using the density of the Gaussian distribution of $T(\delta)$ according to ([3], Theorem 30.4), which exists since $C(\delta)$ is positive definite, in combination with a simple rescaling and shifting of the coordinate system, the probability in (28) can be rewritten as

$$\begin{aligned} P(\delta) &= \frac{(2\pi)^{-n/2}}{\sqrt{\det C(\delta)}}\cdot\int_{S\tau(\delta)+Z_+} e^{-z^t SC(\delta)^{-1}Sz/2}\,dz \\ &= \sqrt{\frac{\lambda_1(\delta)^n}{2^n\pi^n\det C(\delta)}}\cdot\int_{Z_+} e^{-(z+d(\delta))^t D(\delta)(z+d(\delta))/2}\,dz. \end{aligned}$$

According to our assumptions, the eigenvalues $\mu_1(\delta),\dots,\mu_n(\delta)$ of the matrix $D(\delta)$ are given by

$$\mu_1(\delta)=1 \quad\text{and}\quad \mu_k(\delta)=\frac{\lambda_1(\delta)}{\lambda_k(\delta)} \qquad \text{for } k=2,\dots,n,$$

with corresponding orthonormalized eigenvectors $w_k(\delta) = Sv_k(\delta)$, for $k = 1, \ldots, n$. Now let $B(\delta)$ denote the orthogonal matrix with columns $w_1(\delta), \ldots, w_n(\delta)$ and introduce the change of variables $z = B(\delta)\zeta$. Moreover, let

$$Z(\zeta_1, \delta) = \left\{ (\zeta_2, \ldots, \zeta_n) : \sum_{k=1}^{n} \zeta_k w_k(\delta) \in Z_+ \right\} \subset \mathbb{R}^{n-1}$$

define real numbers $\eta_1(\delta), \ldots, \eta_n(\delta)$ by

$$\eta_k(\delta) = S\tau(\delta) \cdot w_k(\delta) = \tau(\delta) \cdot v_k(\delta) \qquad \text{for } k = 1, \ldots, n,$$

and let

$$I(\zeta_1, \delta) = \int_{Z(\zeta_1, \delta)} \exp\left(-\sum_{k=1}^{n} \frac{\mu_k(\delta)}{2} \left(\zeta_k + \frac{\eta_k(\delta)}{\lambda_1(\delta)^{1/2}} \right)^2 \right) d(\zeta_2, \ldots, \zeta_n).$$

Due to (ii) and the definition of the signs s_k, the eigenvector $w_1(\delta)$ has strictly positive components for all sufficiently small $\delta > 0$, and therefore the identity

$$\big(z + d(\delta)\big)^t D(\delta) \big(z + d(\delta)\big) = \sum_{k=1}^{n} \mu_k(\delta) \left(\zeta_k + \frac{\eta_k(\delta)}{\lambda_1(\delta)^{1/2}} \right)^2$$

implies

$$\begin{aligned} &\int_{Z_+} e^{-(z+d(\delta))^t D(\delta)(z+d(\delta))/2}\, dz \\ &\quad = \int_{B(\delta)^{-1} Z_+} \exp\left(-\sum_{k=1}^{n} \frac{\mu_k(\delta)}{2} \left(\zeta_k + \frac{\eta_k(\delta)}{\lambda_1(\delta)^{1/2}} \right)^2 \right) d\zeta \\ &\quad = \int_0^\infty I(\zeta_1, \delta)\, d\zeta_1. \end{aligned} \tag{33}$$

From the definition of $I(\zeta_1, \delta)$, one can easily deduce

$$I(\zeta_1, \delta) = \zeta_1^{n-1} \cdot \int_{Z(1,\delta)} \exp\left(-\sum_{k=1}^{n} \frac{\mu_k(\delta)}{2} \left(\zeta_1 \xi_k + \frac{\eta_k(\delta)}{\lambda_1(\delta)^{1/2}} \right)^2 \right) d(\xi_2, \ldots, \xi_n),$$

where we define $\xi_1 = 1$. This representation furnishes for all $\zeta_1 > 0$ and $\delta > 0$ the estimate

$$I(\zeta_1, \delta) \le \zeta_1^{n-1} \cdot \mathrm{vol}_{n-1}(Z(1, \delta)) \cdot e^{-(\zeta_1 + \eta_1(\delta)\lambda_1(\delta)^{-1/2})^2/2}. \tag{34}$$

Again according to (ii), the $(n-1)$-dimensional volume of the simplex $Z(1, \delta)$ converges to the $(n-1)$-dimensional volume of the simplex

$$\widetilde{Z} = \{ z \in Z_+ : (z - S\bar{v}_1, S\bar{v}_1) = 0 \} \subset \mathbb{R}^n,$$

which can be computed as

$$\mathrm{vol}_{n-1}(\widetilde{Z}) = \frac{1}{(n-1)!} \cdot \left| \prod_{j=1}^{n} \bar{v}_{1j} \right|^{-1}.$$

Now let $\zeta_1 > 0$ be arbitrary, but fixed. Notice that since we did not make any assumptions about the asymptotic behavior of the eigenvectors $w_2(\delta), \ldots, w_n(\delta)$ for $\delta \to 0$, the sets $Z(1,\delta)$ do not have to converge. Yet, (ii) yields the existence of a compact subset $K \subset \mathbb{R}^{n-1}$ such that $Z(1,\delta) \subset K$ for all sufficiently small $\delta > 0$. Furthermore, we have

$$\begin{aligned}\sum_{k=1}^{n} \mu_k(\delta)\left(\zeta_1\xi_k + \frac{\eta_k(\delta)}{\lambda_1(\delta)^{1/2}}\right)^2 &= \zeta_1^2 + \frac{2\zeta_1\eta_1(\delta)}{\lambda_1(\delta)^{1/2}} + \frac{\eta_1(\delta)^2}{\lambda_1(\delta)} + \sum_{k=2}^{n} \frac{\zeta_1^2\xi_k^2\lambda_1(\delta)}{\lambda_k(\delta)} \\ &\quad + 2\sum_{k=2}^{n} \frac{\zeta_1\xi_k\eta_k(\delta)\lambda_1(\delta)^{1/2}}{\lambda_k(\delta)} + \sum_{k=2}^{n} \frac{\eta_k(\delta)^2}{\lambda_k(\delta)} \\ &\to \zeta_1^2 + 2\zeta_1\alpha_1 + \alpha_1^2 + \sum_{k=2}^{n} \alpha_k^2\end{aligned}$$

as $\delta \to 0$. Due to (iii) and (iv), this convergence is uniform on K. Therefore, we have

$$\lim_{\delta\to 0} I(\zeta_1,\delta) = \zeta_1^{n-1} \cdot \mathrm{vol}_{n-1}(\widetilde{Z}) \cdot e^{-(\zeta_1+\alpha_1)^2/2} \cdot e^{-(\alpha_2^2+\cdots+\alpha_n^2)/2}$$

for all $\zeta_1 > 0$.

Due to (34) and $\mathrm{vol}_{n-1}(Z(1,\delta)) \to \mathrm{vol}_{n-1}(\widetilde{Z})$, we can now apply the dominated convergence theorem to pass to the limit $\delta \to 0$ in (33), and this furnishes

$$\begin{aligned}&\lim_{\delta\to 0} \int_{Z_+} e^{-(z+d(\delta))^t D(\delta)(z+d(\delta))/2}\, dz \\ &\quad = \mathrm{vol}_{n-1}(\widetilde{Z}) \cdot e^{-(\alpha_2^2+\cdots+\alpha_n^2)/2} \cdot \int_0^\infty \zeta_1^{n-1} e^{-(\zeta_1+\alpha_1)^2/2}\, d\zeta_1 \\ &\quad = \mathrm{vol}_{n-1}(\widetilde{Z}) \cdot e^{-(\alpha_2^2+\cdots+\alpha_n^2)/2} \cdot \int_{\alpha_1}^\infty (s-\alpha_1)^{n-1} e^{-s^2/2}\, ds.\end{aligned}$$

□

We close this section with a corollary to Proposition 4.1. In our applications of the above result, we are not only interested in the asymptotic behavior of $P(\delta)$ as defined in (28), that is, for the fixed sign sequence $(s_1, \ldots, s_n)$, but also in the corresponding probability for the negative sign sequence $(-s_1, \ldots, -s_n)$.

More precisely, if $T(\delta) = (T_1(\delta), \ldots, T_n(\delta))^t \in \mathbb{R}^n$ denotes again a one-parameter family of $\mathbb{R}^n$-valued random Gaussian variables over a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, indexed by $\delta > 0$, and if we choose both a sign sequence $(s_1, \ldots,$

$s_n) \in \{\pm 1\}^n$ and a one-parameter family $\tau(\delta) \in \mathbb{R}^n$ of threshold vectors, then we are interested in the asymptotic behavior as $\delta \to 0$ of the probability

$$P^{\pm}(\delta) = \mathbb{P}\{s_j(T_j(\delta) - \tau_j(\delta)) \geq 0 \text{ for all } j = 1, \ldots, n\} + \mathbb{P}\{s_j(T_j(\delta) - \tau_j(\delta)) \leq 0 \text{ for all } j = 1, \ldots, n\}. \tag{35}$$

This is the subject of the following corollary.

COROLLARY 4.3. *Let $(s_1, \ldots, s_n) \in \{\pm 1\}^n$ denote a fixed sign sequence, let $\tau(\delta) \in \mathbb{R}^n$ denote a threshold vector, and consider a one-parameter family $T(\delta)$, $\delta > 0$, of $\mathbb{R}^n$-valued random Gaussian variables over a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ which satisfies all the assumptions of Proposition 4.1. Then the probability $P^{\pm}(\delta)$ defined in (35) satisfies*

$$\lim_{\delta \to 0} P^{\pm}(\delta) \cdot \sqrt{\frac{\det C(\delta)}{\lambda_1(\delta)^n}} = \frac{\Gamma(n/2) \cdot S_\alpha^{\pm}}{2 \cdot \pi^{n/2} \cdot (n-1)!} \cdot \left| \prod_{j=1}^{n} \bar{v}_{1j} \right|^{-1}, \tag{36}$$

where $S_\alpha^{\pm} = S_\alpha + S_{-\alpha}$, with α as in (29) and S_α as in (30). Moreover, for the special case $n = 3$ one obtains

$$S_\alpha^{\pm} = 2e^{-(\alpha_2^2 + \alpha_3^2)/2} \cdot (1 + \alpha_1^2). \tag{37}$$

PROOF. One only has to apply Proposition 4.1 twice—first with the given sign vector $(s_1, \ldots, s_n)$, and then with the sign vector $(-s_1, \ldots, -s_n)$. Notice that in the latter case, we have to use the eigenvector $-v_1(\delta)$ instead of $v_1(\delta)$, which leads to $-\alpha_k$ instead of α_k in (29); everything else remains unchanged. This immediately implies (36). As for (37), one only has to notice that

$$\int_{\alpha_1}^{\infty} e^{-s^2/2}\, ds + \int_{-\alpha_1}^{\infty} e^{-s^2/2}\, ds = \int_{-\infty}^{\infty} e^{-s^2/2}\, ds = \sqrt{2\pi}$$

and employ Remark 4.2. □

5. Sampling based on spatial correlations. The goal of this section is the proof of Theorem 1.4. To do this, we need to relate the spatial correlation function R to local probability asymptotics. For this, we use the following lemma.

LEMMA 5.1. *Consider a Gaussian random field $u : [a, b] \times \Omega \to \mathbb{R}$ satisfying* (G1) *and* (G2). *For $x \in [a, b)$ and sufficiently small values of $\delta > 0$, define the random vector $T(\delta) = (T_1(\delta), T_2(\delta), T_3(\delta))^t$ via*

$$T_1(\delta) = u(x), \qquad T_2(\delta) = u\left(x + \frac{\delta}{2}\right) \quad \text{and} \quad T_3(\delta) = u(x + \delta). \tag{38}$$

Then T is a centered Gaussian random variable with positive definite covariance matrix $C(\delta)$. Moreover, if we denote the eigenvalues of $C(\delta)$ by $0 < \lambda_1(\delta) \le \lambda_2(\delta) \le \lambda_3(\delta)$, then

$$\lambda_1(\delta) = \frac{\det \mathcal{R}(x)}{96\mathcal{R}^m_{3,3}(x)} \cdot \delta^4 + O(\delta^5),$$

$$\lambda_2(\delta) = \frac{\mathcal{R}^m_{3,3}(x)}{2R_{0,0}(x)} \cdot \delta^2 + O(\delta^3),$$

$$\lambda_3(\delta) = 3R_{0,0}(x) + O(\delta),$$

where we use the notation introduced in (3), (4) and (5). In addition, we can choose the normalized eigenvectors $v_1(\delta)$, $v_2(\delta)$ and $v_3(\delta)$ corresponding to these eigenvalues in such a way that

$$\lim_{\delta\to 0} v_1(\delta) = \frac{1}{\sqrt{6}}\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}, \qquad \lim_{\delta\to 0} v_2(\delta) = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix},$$

$$\lim_{\delta\to 0} v_3(\delta) = \frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Finally, for a C^3-function $\mu : [a, b] \to \mathbb{R}$ define the vector $\tau(\delta) = (\tau_1(\delta), \tau_2(\delta), \tau_3(\delta))^t$ via

$$\tau_1(\delta) = \mu(x), \qquad \tau_2(\delta) = \mu\left(x + \frac{\delta}{2}\right) \quad \text{and} \quad \tau_3(\delta) = \mu(x + \delta). \tag{39}$$

Then

$$\tau(\delta) \cdot v_1(\delta) = \frac{\mathcal{R}^m_{3,1}(x)\mu(x) - \mathcal{R}^m_{3,2}(x)\mu'(x) + \mathcal{R}^m_{3,3}(x)\mu''(x)}{4\sqrt{6}\mathcal{R}^m_{3,3}(x)} \cdot \delta^2 + O(\delta^3),$$

$$\tau(\delta) \cdot v_2(\delta) = \frac{R_{1,0}(x)\mu(x) - R_{0,0}(x)\mu'(x)}{\sqrt{2}R_{0,0}(x)} \cdot \delta + O(\delta^2),$$

$$\tau(\delta) \cdot v_3(\delta) = \sqrt{3} \cdot \mu(x) + O(\delta).$$

PROOF. Due to our assumptions on u, the vector $T(\delta)$ is normally distributed with mean $0 \in \mathbb{R}^3$ and covariance matrix $C(\delta) \in \mathbb{R}^{3\times 3}$ given by

$$C(\delta) = \begin{pmatrix} r(0,0) & r(0,\delta/2) & r(0,\delta) \\ r(0,\delta/2) & r(\delta/2,\delta/2) & r(\delta/2,\delta) \\ r(0,\delta) & r(\delta/2,\delta) & r(\delta,\delta) \end{pmatrix},$$

where we use the abbreviation

$$r(\delta_1, \delta_2) = R(x + \delta_1, x + \delta_2).$$

For $(\delta_1, \delta_2) \to 0$, the function r can be expanded as

$$\begin{aligned} r(\delta_1, \delta_2) = R_{0,0}(x) + R_{1,0}(x)\delta_1 + R_{1,0}(x)\delta_2 + \frac{R_{2,0}(x)}{2}\delta_1^2 + R_{1,1}(x)\delta_1\delta_2 \\ + \frac{R_{2,0}(x)}{2}\delta_2^2 + \frac{R_{3,0}(x)}{6}\delta_1^3 + \frac{R_{2,1}(x)}{2}\delta_1^2\delta_2 + \frac{R_{2,1}(x)}{2}\delta_1\delta_2^2 \\ + \frac{R_{3,0}(x)}{6}\delta_2^3 + O(|(\delta_1, \delta_2)|^4), \end{aligned}$$

where the $R_{k,\ell}$ where defined in (3). Furthermore, (G2) implies that we have the strict inequalities

$$R_{0,0}(x) > 0, \qquad \mathcal{R}^m_{3,3}(x) > 0 \quad \text{as well as} \quad \det \mathcal{R}(x) > 0.$$

These strict inequalities ensure that in all of the expansions derived below the leading order coefficients are positive.

Using the above expansion of r, the determinant of the covariance matrix $C(\delta)$ of the random vector $T(\delta)$ can be written as

$$\det C(\delta) = \tfrac{1}{64} \cdot \det \mathcal{R}(x) \cdot \delta^6 + O(\delta^7),$$

that is, the covariance matrix is positive definite for sufficiently small $\delta > 0$. Furthermore, by applying the Newton polygon method [26, 28] to the characteristic polynomial $\det(C(\delta) - \lambda I)$ it can be shown that in the limit $\delta \to 0$ the three eigenvalues $\lambda_k(\delta)$, for $k = 1, 2, 3$, of $C(\delta)$ are given by the expansions in the formulation of Lemma 5.1.

We now turn our attention to the asymptotic statements concerning the eigenvectors of the covariance matrix. According to the form of $C(\delta)$, we have

$$\lim_{\delta \to 0} C(\delta) = \begin{pmatrix} R_{0,0}(x) & R_{0,0}(x) & R_{0,0}(x) \\ R_{0,0}(x) & R_{0,0}(x) & R_{0,0}(x) \\ R_{0,0}(x) & R_{0,0}(x) & R_{0,0}(x) \end{pmatrix},$$

where the limit has a double eigenvalue 0, as well as the simple eigenvalue $3R_{0,0}(x)$ with normalized eigenvector $(1, 1, 1)^t/3^{1/2}$. Due to standard results on the perturbation of simple eigenvalues and corresponding eigenvectors [30], this implies that $v_3(\delta)$ can be chosen as in the formulation of the lemma.

In order to determine the asymptotic behavior of the eigenvector corresponding to λ_1, we consider the adjoint of the covariance matrix, whose expansion is given by

$$\operatorname{adj} C(\delta) = \frac{\mathcal{R}^m_{3,3}(x)}{4} \cdot \begin{pmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{pmatrix} \cdot \delta^2 + O(\delta^3).$$

The constant coefficient matrix has the double eigenvalue 0, as well as the positive eigenvalue 6 with associated unnormalized eigenvector $(1, -2, 1)^t$. Since

the eigenspace for the largest eigenvalue of the adjoint matrix coincides with the eigenspace for the eigenvalue $\lambda_1(\delta)$ of $C(\delta)$, the simplicity of these eigenvalues shows that we can choose a normalized eigenvector $v_1(\delta)$ for $\lambda_1(\delta)$ with $v_1(\delta) \to (1, -2, 1)^t/6^{1/2}$ for $\delta \to 0$. Finally, the orthogonality of the three eigenvectors shows that we can choose a normalized eigenvector $v_2(\delta)$ for $\lambda_2(\delta)$ with $v_2(\delta) \to (1, 0, -1)^t/2^{1/2}$ for $\delta \to 0$.

We now turn our attention to the asymptotics of the inner products $\tau(\delta) \cdot v_k(\delta)$. Since μ is a C^3-function, we can write

$$\tau(\delta) = \mu(x) \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \frac{\mu'(x)}{2} \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} \cdot \delta + \frac{\mu''(x)}{8} \begin{pmatrix} 0 \\ 1 \\ 4 \end{pmatrix} \cdot \delta^2 + O(\delta^3),$$

and this representation immediately furnishes $\tau(\delta) \cdot v_3(\delta) \to 3^{1/2} \cdot \mu(x)$ as $\delta \to 0$. The statements concerning $\tau(\delta) \cdot v_1(\delta)$ and $\tau(\delta) \cdot v_2(\delta)$ are more involved, and rely on expansions of the eigenvectors in terms of δ.

As for the first eigenvector, write $v_1(\delta) = (v_{1,1}(\delta), v_{1,2}(\delta), v_{1,3}(\delta))^t$, and consider the functions

$$w_{1,k}(\delta) = -\frac{2}{\sqrt{6} \cdot v_{1,2}(\delta)} \cdot v_{1,k}(\delta) \qquad \text{for } k = 1, 2, 3.$$

Then the vector $w_1(\delta) = (w_{1,1}(\delta), w_{1,2}(\delta), w_{1,3}(\delta))^t$ is defined for sufficiently small $\delta > 0$, and for these δ we have

$$w_{1,2}(\delta) = -\tfrac{2}{\sqrt{6}} \quad \text{as well as} \quad \big(C(\delta) - \lambda_1(\delta) I\big) w_1(\delta) = 0.$$

Using the abbreviation $C(\delta) = (c_{i,j}(\delta))_{i,j=1,2,3}$, the latter system is equivalent to

$$\big(c_{1,1}(\delta) - \lambda_1(\delta)\big) w_{1,1}(\delta) + c_{1,3}(\delta) w_{1,3}(\delta) = \tfrac{2}{\sqrt{6}} \cdot c_{1,2}(\delta),$$

$$c_{3,1}(\delta) w_{1,1}(\delta) + \big(c_{3,3}(\delta) - \lambda_1(\delta)\big) w_{1,3}(\delta) = \tfrac{2}{\sqrt{6}} \cdot c_{3,2}(\delta),$$

which immediately implies

$$w_{1,1}(\delta) = \frac{2}{\sqrt{6}} \cdot \frac{(c_{3,3}(\delta) - \lambda_1(\delta)) c_{1,2}(\delta) - c_{3,2}(\delta) c_{1,3}(\delta)}{(c_{1,1}(\delta) - \lambda_1(\delta))(c_{3,3}(\delta) - \lambda_1(\delta)) - c_{1,3}(\delta) c_{3,1}(\delta)},$$

$$w_{1,3}(\delta) = \frac{2}{\sqrt{6}} \cdot \frac{(c_{1,1}(\delta) - \lambda_1(\delta)) c_{3,2}(\delta) - c_{1,2}(\delta) c_{3,1}(\delta)}{(c_{1,1}(\delta) - \lambda_1(\delta))(c_{3,3}(\delta) - \lambda_1(\delta)) - c_{1,3}(\delta) c_{3,1}(\delta)}.$$

Expanding the right-hand sides now furnishes

$$w_1(\delta) = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} + \frac{\sqrt{6}}{24} \cdot \frac{\mathcal{R}^m_{3,2}(x)}{\mathcal{R}^m_{3,3}(x)} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \cdot \delta + \begin{pmatrix} w_{1,1,2} \\ 0 \\ w_{1,3,2} \end{pmatrix} \cdot \delta^2 + O(\delta^3),$$

with

$$w_{1,1,2} + w_{1,3,2} = \frac{1}{4\sqrt{6}} \cdot \frac{\mathcal{R}^m_{3,1}(x)}{\mathcal{R}^m_{3,3}(x)}.$$

This finally implies

$$\tau(\delta)\cdot w_1(\delta)=\frac{\mathcal{R}^m_{3,1}(x)\mu(x)-\mathcal{R}^m_{3,2}(x)\mu'(x)+\mathcal{R}^m_{3,3}(x)\mu''(x)}{4\sqrt{6}\mathcal{R}^m_{3,3}(x)}\cdot\delta^2+O(\delta^3),$$

and together with

$$\tau(\delta)\cdot v_1(\delta)=\frac{-\sqrt{6}v_{1,2}(\delta)}{2}\cdot\tau(\delta)\cdot w_1(\delta)\quad\text{and}\quad\lim_{\delta\to0}\frac{-\sqrt{6}v_{1,2}(\delta)}{2}=1$$

this establishes the asymptotic behavior of $\tau(\delta)\cdot v_1(\delta)$.

Finally, we turn our attention to the second eigenvector. Following our above approach, we write $v_2(\delta)=(v_{2,1}(\delta),v_{2,2}(\delta),v_{2,3}(\delta))^t$, and consider the functions

$$w_{2,k}(\delta)=\frac{1}{\sqrt{2}\cdot v_{2,1}(\delta)}\cdot v_{2,k}(\delta)\qquad\text{for } k=1,2,3.$$

Then the vector $w_2(\delta)=(w_{2,1}(\delta),w_{2,2}(\delta),w_{2,3}(\delta))^t$ is defined for sufficiently small $\delta>0$, and for these δ we have

$$w_{2,1}(\delta)=\tfrac{1}{\sqrt{2}}\quad\text{as well as}\quad\big(C(\delta)-\lambda_2(\delta)I\big)w_2(\delta)=0.$$

Using again the abbreviation $C(\delta)=(c_{i,j}(\delta))_{i,j=1,2,3}$, the latter system is equivalent to

$$\begin{aligned}\big(c_{2,2}(\delta)-\lambda_2(\delta)\big)w_{2,2}(\delta)+c_{2,3}(\delta)w_{2,3}(\delta)&=-\tfrac{1}{\sqrt{2}}\cdot c_{2,1}(\delta),\\ c_{3,2}(\delta)w_{2,2}(\delta)+\big(c_{3,3}(\delta)-\lambda_2(\delta)\big)w_{2,3}(\delta)&=-\tfrac{1}{\sqrt{2}}\cdot c_{3,1}(\delta),\end{aligned}$$

which immediately implies

$$\begin{aligned}w_{2,2}(\delta)&=-\frac{1}{\sqrt{2}}\cdot\frac{(c_{3,3}(\delta)-\lambda_2(\delta))c_{2,1}(\delta)-c_{2,3}(\delta)c_{3,1}(\delta)}{(c_{2,2}(\delta)-\lambda_2(\delta))(c_{3,3}(\delta)-\lambda_2(\delta))-c_{2,3}(\delta)c_{3,2}(\delta)},\\ w_{2,3}(\delta)&=-\frac{1}{\sqrt{2}}\cdot\frac{(c_{2,2}(\delta)-\lambda_2(\delta))c_{3,1}(\delta)-c_{3,2}(\delta)c_{2,1}(\delta)}{(c_{2,2}(\delta)-\lambda_2(\delta))(c_{3,3}(\delta)-\lambda_2(\delta))-c_{2,3}(\delta)c_{3,2}(\delta)}.\end{aligned}$$

Expanding the right-hand sides now furnishes

$$w_2(\delta)=\frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\-1\end{pmatrix}+\begin{pmatrix}0\\w_{2,2,1}\\w_{2,3,1}\end{pmatrix}\cdot\delta+O(\delta^2)$$

with

$$w_{2,2,1}+w_{2,3,1}=\frac{1}{\sqrt{2}}\cdot\frac{R_{1,0}(x)}{R_{0,0}(x)}.$$

This finally implies

$$\tau(\delta)\cdot w_2(\delta)=\frac{R_{1,0}(x)\mu(x)-R_{0,0}(x)\mu'(x)}{\sqrt{2}R_{0,0}(x)}\cdot\delta+O(\delta^2)$$

and together with

$$\tau(\delta)\cdot v_2(\delta)=\sqrt{2}v_{2,1}(\delta)\cdot\tau(\delta)\cdot w_2(\delta)\quad\text{and}\quad\lim_{\delta\to 0}\sqrt{2}v_{2,1}(\delta)=1$$

this establishes the asymptotic behavior of $\tau(\delta)\cdot v_2(\delta)$. □

After these preparations, we are finally in a position to prove our second main result. As mentioned in Section 1, this result provides a general means for determining the location of sampling points of random fields in such a way that the topology of the underlying nodal sets is correctly recognized with the largest probability. In addition, the sampling density can readily be determined from derivatives of the spatial correlation function of the random field.

PROOF OF THEOREM 1.4. Due to our assumptions, the random variable $u(x,\cdot):\Omega\to\mathbb{R}$ is normally distributed with mean 0 and its variance $R_{0,0}(x)$ is positive for each $x\in[a,b]$ due to (G2). This immediately implies (A1). Furthermore, (A2) follows readily from [1], Theorem 3.2.1. Thus, in order to apply Theorem 1.2 we only have to verify (A3).

For this, we apply Corollary 4.3 with $n=3$ and sign vector $(s_1,s_2,s_3)=(1,-1,1)$. Fix $x\in[a,b)$ and consider the δ-dependent three-dimensional random vector $T(\delta)$ defined in (38). Then according to Lemma 5.1, this random vector satisfies all of the assumptions of Proposition 4.1 and Corollary 4.3 with

$$\det C(\delta)=\frac{1}{64}\cdot\det\mathcal{R}(x)\cdot\delta^6+O(\delta^7)\quad\text{and}\quad\lambda_1(\delta)=\frac{\det\mathcal{R}(x)}{96\mathcal{R}^m_{3,3}(x)}\cdot\delta^4+O(\delta^5)$$

as well as

$$\alpha_1=\frac{\mathcal{R}^m_{3,1}(x)\mu(x)-\mathcal{R}^m_{3,2}(x)\mu'(x)+\mathcal{R}^m_{3,3}(x)\mu''(x)}{\mathcal{R}^m_{3,3}(x)^{1/2}\det\mathcal{R}(x)^{1/2}},$$

$$\alpha_2=\frac{R_{1,0}(x)\mu(x)-R_{0,0}(x)\mu'(x)}{R_{0,0}(x)^{1/2}\mathcal{R}^m_{3,3}(x)^{1/2}},$$

$$\alpha_3=\frac{\mu(x)}{R_{0,0}(x)^{1/2}}.$$

Applying Corollary 4.3, we then obtain

$$\lim_{\delta\to 0}\bigl(p_{+1}(x,\delta)+p_{-1}(x,\delta)\bigr)\cdot\sqrt{\frac{\det C(\delta)}{\lambda_1(\delta)^3}}=\frac{3\sqrt{6}}{4\pi}\cdot(1+\alpha_1^2)\cdot e^{-(\alpha_2^2+\alpha_3^2)/2},$$

where we used the formula for $S^{\pm}_{\alpha}$ given in (37). In combination with the above expansions for $\det C(\delta)$ and $\lambda_1(\delta)$, this limit furnishes

$$p_{+1}(x,\delta)+p_{-1}(x,\delta)=\frac{(1+\alpha_1^2)\cdot e^{-(\alpha_2^2+\alpha_3^2)/2}}{64\pi}\cdot\frac{\det\mathcal{R}(x)}{\mathcal{R}^m_{3,3}(x)^{3/2}}\cdot\delta^3+O(\delta^4).$$

Thus, assumption (A3) is satisfied with $\mathcal{C}_0(x) = 3\mathcal{C}(x)/4$, and Theorem 1.4 follows now immediately from Theorem 1.2. □

6. Concluding remarks. At first glance, the title of this paper may appear somewhat misleading or more ambitious than the results delivered. After all, the techniques of proof are based on classical probabilistic arguments. However, the results are new and the examples of Section 2 demonstrate that they have interesting nonintuitive implications.

A reasonable question is why were these results not discovered sooner. We believe that the answer comes from the fact that we are approaching the problem of optimal sampling from the point of view of trying to obtain topological information. This point of view had been taken previously in the work of Adler and Taylor [1, 2]. Their main focus, however, was the estimation of excursion probabilities, that is, the likelihood that a given random function exceeds a certain threshold. In [1, 2], it is shown that such excursion probabilities can be well-approximated by studying the geometry of random sub- or super-level sets of random fields. More precisely, it is shown that the expected value of the Euler characteristic of super-level sets approximates excursion probabilities for large values of the threshold, and that it is possible to derive explicit formulas for the expected values of the Euler characteristic and other intrinsic volumes of nodal domains of random fields.

All of the above results concern the intrinsic volumes of the nodal domains—which are additive set functionals, and therefore computable via local considerations alone [20, 25]. In contrast, in previous work [14] we have demonstrated that the homological analysis of patterns of nodal sets can uncover phenomena that cannot be captured using for example only the Euler characteristic. The more detailed information on the geometry of patterns encoded in homology is an inherently global quantity and cannot be computed through local considerations alone. On the other hand, recent computational advances allow for the fast computation of homological information based on discretized nodal domains. For this reason, we focus on the interface between the discretization and the underlying nodal domain, rather than the homology of the nodal domain directly, and then quantify the likelihood of error in the probabilistic setting. In this sense, our approach complements the above-mentioned results on the geometry of random fields by Adler and Taylor [1, 2].

Given the current activity surrounding the ideas of using topological methods for data analysis and remote sensing [8, 9, 15], we believe the importance of this perspective will grow. Thus, the title of our paper is chosen in part to encourage the interested reader to consider the natural generalizations of this work to higher-dimensional domains where the question becomes one of optimizing the homology of the generalized nodal sets in terms of homology computed using a complex derived from a nonuniform sampling of space.

REFERENCES

[1] ADLER, R. J. (1981). *The Geometry of Random Fields*. Wiley, Chichester. MR611857

[2] ADLER, R. J. and TAYLOR, J. E. (2007). *Random Fields and Geometry*. Springer, New York. MR2319516

[3] BAUER, H. (1996). *Probability Theory*. *de Gruyter Studies in Mathematics* **23**. de Gruyter, Berlin. MR1385460

[4] BHARUCHA-REID, A. T. and SAMBANDHAM, M. (1986). *Random Polynomials*. Academic Press, Orlando, FL. MR856019

[5] COURANT, R. and HILBERT, D. (1953). *Methods of Mathematical Physics*. *Vol. I*. Interscience Publishers, New York. MR0065391

[6] CRAMÉR, H. and LEADBETTER, M. R. (2004). *Stationary and Related Stochastic Processes*. Dover, Mineola, NY. Reprint of the 1967 original. MR2108670

[7] DAY, S., KALIES, W. D., MISCHAIKOW, K. and WANNER, T. (2007). Probabilistic and numerical validation of homology computations for nodal domains. *Electron. Res. Announc. Amer. Math. Soc.* **13** 60–73 (electronic). MR2320683

[8] DE SILVA, V. and GHRIST, R. (2007). Coverage in sensor networks via persistent homology. *Algebr. Geom. Topol.* **7** 339–358. MR2308949

[9] DE SILVA, V. and GHRIST, R. (2007). Homological sensor networks. *Notices Amer. Math. Soc.* **54** 10–17. MR2275921

[10] DUNNAGE, J. E. A. (1966). The number of real zeros of a random trigonometric polynomial. *Proc. London Math. Soc.* (3) **16** 53–84. MR0192532

[11] EDELMAN, A. and KOSTLAN, E. (1995). How many zeros of a random polynomial are real? *Bull. Amer. Math. Soc.* (*N.S.*) **32** 1–37. MR1290398

[12] FARAHMAND, K. (1998). *Topics in Random Polynomials*. *Pitman Research Notes in Mathematics Series* **393**. Longman, Harlow. MR1679392

[13] GAMEIRO, M., MISCHAIKOW, K. and KALIES, W. (2004). Topological characterization of spatial-temporal chaos. *Phys. Rev. E* (3) **70** 035203, 4. MR2129999

[14] GAMEIRO, M., MISCHAIKOW, K. and WANNER, T. (2005). Evolution of pattern complexity in the Cahn–Hilliard theory of phase separation. *Acta Materialia* **53** 693–704.

[15] GHRIST, R. (2008). Barcodes: The persistent topology of data. *Bull. Amer. Math. Soc.* (*N.S.*) **45** 61–75 (electronic). MR2358377

[16] KAC, M. (1943). On the average number of real roots of a random algebraic equation. *Bull. Amer. Math. Soc.* **49** 314–320. MR0007812

[17] KAC, M. (1949). On the average number of real roots of a random algebraic equation. II. *Proc. London Math. Soc.* (2) **50** 390–408. MR0030713

[18] KACZYNSKI, T., MISCHAIKOW, K. and MROZEK, M. (2004). *Computational Homology*. *Applied Mathematical Sciences* **157**. Springer, New York. MR2028588

[19] KAHANE, J.-P. (1985). *Some Random Series of Functions*, 2nd ed. *Cambridge Studies in Advanced Mathematics* **5**. Cambridge Univ. Press, Cambridge. MR833073

[20] KLAIN, D. A. and ROTA, G.-C. (1997). *Introduction to Geometric Probability*. Cambridge Univ. Press, Cambridge. MR1608265

[21] KRISHAN, K., GAMEIRO, M., MISCHAIKOW, K., SCHATZ, M., KURTULDU, H. and MADRUGA, S. (2007). Homology and symmetry breaking in Rayleigh–Benard convection: Experiments and simulations. *Phys. Fluids* **19** 117105.

[22] MARCUS, M. B. and PISIER, G. (1981). *Random Fourier Series with Applications to Harmonic Analysis*. *Annals of Mathematics Studies* **101**. Princeton Univ. Press, Princeton, NJ. MR630532

[23] MISCHAIKOW, K. and WANNER, T. (2007). Probabilistic validation of homology computations for nodal domains. *Ann. Appl. Probab.* **17** 980–1018. MR2326238

[24] ROZANOV, Y. A. (1998). *Random Fields and Stochastic Partial Differential Equations*. *Mathematics and Its Applications* **438**. Kluwer Academic, Dordrecht. MR1629699

[25] SANTALÓ, L. A. (2004). *Integral Geometry and Geometric Probability*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2162874

[26] SIDOROV, N., LOGINOV, B., SINITSYN, A. and FALALEEV, M. (2002). *Lyapunov–Schmidt Methods in Nonlinear Analysis and Applications*. *Mathematics and Its Applications* **550**. Kluwer Academic, Dordrecht. MR1959647

[27] TORQUATO, S. (2002). *Random Heterogeneous Materials*: *Microstructure and Macroscopic Properties*. *Interdisciplinary Applied Mathematics* **16**. Springer, New York. MR1862782

[28] VAĬNBERG, M. M. and TRENOGIN, V. A. (1974). *Theory of Branching of Solutions of Nonlinear Equations*. Noordhoff International Publishing, Leyden. MR0344960

[29] VANMARCKE, E. (1983). *Random Fields*. MIT Press, Cambridge, MA. MR761904

[30] WILKINSON, J. H. (1988). *The Algebraic Eigenvalue Problem*. Oxford Univ. Press, New York. MR950175

DEPARTMENT OF MATHEMATICS
RUTGERS UNIVERSITY
PISCATAWAY, NEW JERSEY 08854
USA
E-MAIL: mischaik@math.rutgers.edu

DEPARTMENT OF MATHEMATICAL SCIENCES
GEORGE MASON UNIVERSITY
FAIRFAX, VIRGINIA 22030
USA
E-MAIL: wanner@math.gmu.edu