On the number of excursion sets of planar Gaussian fields

Dmitry Beliaev[1] · **Michael McAuley[1]** · **Stephen Muirhead[2,3]**

Abstract

The Nazarov–Sodin constant describes the average number of nodal set components of smooth Gaussian fields on large scales. We generalise this to a functional describing the corresponding number of level set components for arbitrary levels. Using results from Morse theory, we express this functional as an integral over the level densities of different types of critical points, and as a result deduce the absolute continuity of the functional as the level varies. We further give upper and lower bounds showing that the functional is at least bimodal for certain isotropic fields, including the important special case of the random plane wave.

✉ Michael McAuley
  mcauley@maths.ox.ac.uk

  Dmitry Beliaev
  belyaev@maths.ox.ac.uk

  Stephen Muirhead
  s.muirhead@qmul.ac.uk

[1] Mathematical Institute, University of Oxford, Oxford, UK

[2] Department of Mathematics, King's College London, London, UK

[3] Present Address: School of Mathematical Sciences, Queen Mary University of London, London, UK

1 Introduction

1.1 The Nazarov–Sodin constant

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a continuous stationary planar Gaussian field normalised to have zero mean and unit variance. The *nodal set* of f is the random set

$$\mathcal{N} = \left\{ x \in \mathbb{R}^2 : f(x) = 0 \right\}.$$

Let $\kappa : \mathbb{R}^2 \to [-1, 1]$ denote the covariance kernel of f, i.e. $\kappa(x) = \mathbb{E}[f(x)f(0)]$. We assume throughout that κ is C^{4+}, which ensures that almost surely f is C^{2+}. Since κ is positive definite, continuous and $\kappa(0) = 1$, by Bochner's theorem there exists a probability measure ρ such that

$$\kappa(x) = \int_{\mathbb{R}^2} e^{i \langle t, x \rangle} \, d\rho(t); \tag{1}$$

this is known as the *spectral measure* of f, and must be Hermitian (that is, $\rho(-A) = \rho(A)$ for all Borel sets A). Since the distribution of f is uniquely determined by its covariance function (Kolmogorov's theorem), (1) shows that the distribution of f is uniquely determined by ρ.

The geometric properties of $\mathcal{N}$ are of interest, in part, because in the case that f is a random eigenfunction of the Laplacian they relate to a significant conjecture in the physics literature: the Berry conjecture [4]. A summary of this conjecture and other research on this topic may be found in [17]. One of the main analytical results concerning this set, due to Nazarov and Sodin [13,15], states that the number of components of $\mathcal{N}$ in a large domain scales like the area of the domain. Specifically, if N_R denotes the number of components of $\mathcal{N}$ inside the centred ball of radius $R > 0$, then provided f is ergodic, there exists a constant $c_{LS} = c_{LS}(\rho) \geq 0$ such that

$$N_R/(\pi R^2) \to c_{LS}$$

as $R \to \infty$, where convergence occurs almost surely and in L^1. Nazarov–Sodin also obtained analogous results in higher dimensions and for Gaussian ensembles on manifolds [13]. In the case that f is not ergodic, it has been shown [10] (under the additional assumption that ρ has compact support) that the expected number of nodal components, scaled by the area, still converges, i.e.

$$\mathbb{E}[N_R]/(\pi R^2) \to c_{LS}$$

as $R \to \infty$. Further, in [10] it was also shown that among fields with compactly supported spectral measures, the constant c_{LS} varies continuously with ρ (in the weak-$*$ topology).

1.2 The main results

The first contribution of this paper is to extend the results of Nazarov–Sodin and [10] to arbitrary levels. For $u \in \mathbb{R}^2$ and $R > 0$ let $B(u, R)$ be the ball of radius R centred at u and $B(R) := B(0, R)$. Let $\mathcal{N}_\ell = \{x \in \mathbb{R}^2 : f(x) = \ell\}$ denote a level set of f and let $N_{LS,R}(\ell)$ be the number of components of $\mathcal{N}_\ell$ contained in $B(R)$ (i.e. those which intersect $B(R)$ but not $\partial B(R)$). We consider fields satisfying the following assumptions:

Conditions 1.1 *The Gaussian field f satisfies:*

1. *For some $\nu > 0$, $f \in C_{loc}^{2+\nu}(\mathbb{R}^2)$ almost surely;*
2. *$\nabla^2 f(0)$ is a non-degenerate Gaussian vector (here, and later on, we treat $\nabla^2 f$ as a vector of distinct partial derivatives (f_{xx}, f_{xy}, f_{yy}));*
3. *For any $t \in \mathbb{R}^2$, if $f(t) - f(0)$ is a non-degenerate Gaussian variable then $(f(t) - f(0), \nabla f(t), \nabla f(0))$ is a non-degenerate Gaussian vector.*

We note that these assumptions are quite minimal. The first condition holds if the covariance function κ is $C^{4+\epsilon}$ for some $\epsilon > 2\nu$, or equivalently when $\int_{\mathbb{R}^2} |\lambda|^{4+\epsilon} \, d\rho(\lambda) < \infty$. The second condition is equivalent to the support of ρ not being contained in the union of two lines through the origin. The third condition holds provided that the support of ρ is not too degenerate; in particular it holds if the support of ρ contains an open set.

Theorem 1.2 *Let f satisfy Conditions 1.1. For each $\ell \in \mathbb{R}$, there exists $c_{LS}(\rho, \ell) \geq 0$ such that*

$$\mathbb{E}[N_{LS,R}(\ell)] = c_{LS}(\rho, \ell) \cdot \pi R^2 + O(R)$$

as $R \to \infty$. The constant implied by the $O(\cdot)$ notation may depend on ρ but is independent of ℓ. If f is also ergodic, then

$$N_{LS,R}(\ell)/(\pi R^2) \to c_{LS}(\rho, \ell)$$

almost surely and in L^1.

There is also interest in studying the number of excursion sets of Gaussian fields. Let $N_{ES,R}(\ell)$ denote the number of components of $\{x \in \mathbb{R}^2 : f(x) \geq \ell\}$ contained in $B(R)$.

Theorem 1.3 *Let f satisfy Conditions 1.1. For each $\ell \in \mathbb{R}$, there exists $c_{ES}(\rho, \ell) \geq 0$ such that*

$$\mathbb{E}[N_{ES,R}(\ell)] = c_{ES}(\rho, \ell) \cdot \pi R^2 + O(R)$$

as $R \to \infty$. The constant implied by the $O(\cdot)$ notation may depend on ρ but is independent of ℓ. If f is also ergodic, then

$$N_{ES,R}(\ell)/(\pi R^2) \to c_{ES}(\rho, \ell)$$

almost surely and in L^1.

Remark 1.4 Our use of the domain $B(R)$ in the definitions of $N_{LS,R}(\ell)$ and $N_{ES,R}(\ell)$ is mainly for simplicity, and after minor modifications the proofs of Theorems 1.2 and 1.3 go through equally well for rescaled copies of any bounded reference domain $\Omega \subset \mathbb{R}^2$, provided that Ω is convex and $\partial\Omega$ is piecewise smooth (and probably more generally as well). Moreover the limiting constants c_{LS} and c_{ES} do not depend on Ω (after replacing the scaling factor πR^2 by $\mathrm{Area}(\Omega)R^2$).

Remark 1.5 Theorem 1.3 can also be applied to lower excursion sets (the components of $\{x \in \mathbb{R}^2 : f(x) \leq \ell\}$) since f has symmetric distribution and $\{x \in \mathbb{R}^2 : f(x) \leq \ell\} = \{x \in \mathbb{R}^2 : -f(x) \geq -\ell\}$. Theorem 1.3 can then be used to prove Theorem 1.2 by making use of Euler's formula to show that the number of level set components $N_{LS,R}(\ell)$ is equal to the number of upper and lower excursion set components in $B(R)$ and a bounded error term (see the proof of Lemma 2.5 for details of this argument).

The symmetry of f along with the observations in the previous remark immediately give the following corollary.

Corollary 1.6 *Let f satisfy Conditions 1.1. Then*

(1) $c_{LS}(\ell) = c_{LS}(-\ell)$ *for all $\ell \in \mathbb{R}$,*
(2) $c_{LS}(\ell) = c_{ES}(\ell) + c_{ES}(-\ell)$ *for all $\ell \in \mathbb{R}$,*
(3) $c_{LS}(0) = 2c_{ES}(0)$.

Theorems 1.2 and 1.3 are, in isolation, only a modest improvement on previous results; they could be proven by slightly adapting the analysis in [10,13]. The main novelty of our work is to relate the functionals c_{LS} and c_{ES} to the density of critical points of f of different types and at different levels. To state the relationship, we shall require, in particular, a classification of the saddle points of the field into two types:

Definition 1.7 Let x_0 be a saddle point of a function $g : \mathbb{R}^2 \to \mathbb{R}$ such that there are no other critical points at the same level as x_0. We say that x_0 is an *upper connected saddle* if it is in the closure of only one component of $\{x \in \mathbb{R}^2 : g(x) > g(x_0)\}$. Similarly, x_0 is said to be *lower connected* if it is in the closure of only one component of $\{x \in \mathbb{R}^2 : g(x) < g(x_0)\}$.

We say that a Gaussian field f satisfying Conditions 1.1 is *periodic* if there exists $x \neq 0$ with $\kappa(x) = 1$ and is *aperiodic* otherwise. A Gaussian field which is aperiodic almost surely has no two critical points at the same level (see Lemma 2.4), and we will show that for such fields, all saddle points of f satisfy exactly one of the two conditions in Definition 1.7. In the case that f is periodic, we will require a more general definition for classifying saddle points as upper or lower connected, which is given in Sect. 3.

Previous work has shown that, for sufficiently regular isotropic[1] Gaussian fields, the expected number of local maxima, local minima or saddle points with value in a certain interval can be expressed as the integral of an explicit density function [5,6].

---

[1] A Gaussian field is said to be *isotropic* if its covariance function $\kappa(x)$ can be expressed as a function of $|x|$ where $|\cdot|$ denotes the Euclidean norm.

In Sect. 3 we prove the following version of this result, which applies to more general Gaussian fields and also isolates the upper and lower connected saddles, but does not explicitly identify the densities. This proposition uses Definition 3.1 for upper and lower connected saddle points (which coincides with Definition 1.7 for aperiodic fields).

Proposition 1.8 *Let f satisfy Conditions 1.1. Then there exist non-negative functions p_{m^+}, p_{m^-}, p_{s^+}, p_{s^-} and p_s on $\mathbb{R}$ such that the following holds. Let $\Omega \subset \mathbb{R}^2$ be compact and $\partial\Omega$ have finite Hausdorff-1 measure. Let $\ell \in \mathbb{R}$ and let $N_{m^+}(\ell)$, $N_{m^-}(\ell)$, $N_{s^+}(\ell)$, $N_{s^-}(\ell)$ and $N_s(\ell)$ denote the number of local maxima, local minima, upper connected saddles, lower connected saddles and saddles of f in Ω with level above ℓ respectively. Then*

$$\mathbb{E}[N_h(\ell)] = \text{Area}(\Omega) \int_\ell^\infty p_h(x)\, dx$$

for $h = m^+, m^-, s^+, s^-, s$. Furthermore, these functions can be chosen to satisfy the relations $p_{m^+}(x) = p_{m^-}(-x)$, $p_{s^+}(x) = p_{s^-}(-x)$ and $p_{s^-} + p_{s^+} = p_s$, and such that p_{m^+}, p_{m^-} and p_s are continuous.

The main theorem of the paper gives an explicit expression for the functionals c_{LS} and c_{ES} in terms of the densities introduced in Proposition 1.8. As a result, we deduce the absolute continuity of these functionals as the level varies. In the case of spectral measures with compact support, we also show the joint continuity of these functionals with respect to both the level and the spectral measure.

For each f satisfying Conditions 1.1 we can associate a spectral measure ρ via (1). Let $\mathcal{P}_c$ denote the collection of such measures that are supported in the closure of $B(1)$. By rescaling the axes, the results we state for $\mathcal{P}_c$ can be shown to hold for any spectral measures with compact support.

Theorem 1.9 *Let f satisfy Conditions 1.1 and let p_{m^+}, p_{m^-}, p_{s^+}, p_{s^-} denote the densities specified in Proposition 1.8. Then*

$$c_{LS}(\rho, \ell) = \int_\ell^\infty p_{m^+}(x) - p_{s^-}(x) + p_{s^+}(x) - p_{m^-}(x)\, dx \tag{2}$$

and

$$c_{ES}(\rho, \ell) = \int_\ell^\infty p_{m^+}(x) - p_{s^-}(x)\, dx \tag{3}$$

and hence c_{LS} and c_{ES} are absolutely continuous in ℓ. In addition c_{LS} and c_{ES} are jointly continuous in $(\rho, \ell) \in \mathcal{P}_c \times \mathbb{R}$ where $\mathcal{P}_c$ is given the weak-$*$ topology.

Remark 1.10 Theorem 1.9 provides a new tool with which to analyse the Nazarov–Sodin constant. Since the densities p_{m^+}, p_{m^-} and p_s are in principle known by the Kac–Rice formula, our result demonstrates that the study of the Nazarov–Sodin constant can be reduced to an analysis of the density p_{s^-} (or, equivalently, $p_{s^+} = p_s - p_{s^-}$), which may be an easier quantity to handle.
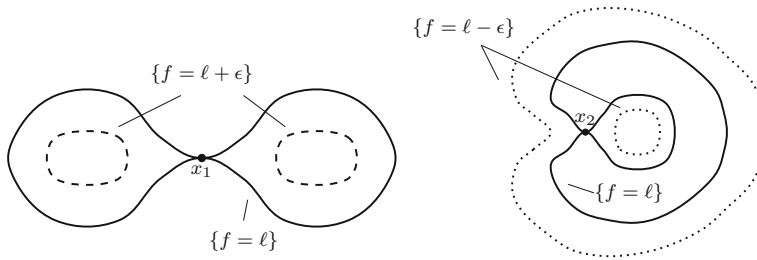
Fig. 1 The number of excursion set components increases by one on passing through the lower connected saddle x_1 and is constant on passing through the upper connected saddle x_2

Remark 1.11 Theorems 1.2, 1.3 and 1.9 can be generalised to many examples of non-Gaussian stationary random fields, since our proof requires only that the field satisfies certain topological properties almost surely and is sufficiently regular to apply the Kac–Rice formula (see the more general version of our main results stated in Propositions 2.6 and 2.7 below).

We also believe Theorems 1.2, 1.3 and 1.9 could be generalised to higher dimensions, with the analogues of (2) and (3) still valid once the saddle points defining p_{s+} and p_{s-} are replaced with critical points of index 1 and $d-1$ respectively (using a more general definition for upper and lower connected saddles). However, since some of the topological arguments in the proof increase in complexity in higher dimensions, in the interest of simplicity we do not pursue this generalisation here.

Let us mention the key intuition behind the proof of Theorem 1.9. This theorem is based on a deterministic relationship between the excursion sets and critical points of sufficiently regular planar functions which is closely related to Morse theory. The excursion set of such a function above a level deforms continuously as the level increases, provided it does not pass through a critical point. In particular, there is no change in the number of components of the excursion set. When passing through a critical point, the topology of the excursion set changes in a way that is predicted by Morse theory and depends on the index of the critical point. For local maxima and local minima, the number of components of the excursion set changes in a consistent way. For saddle points, the change in the number of components is determined by whether it is upper connected or lower connected (see Fig. 1). Ultimately, Theorem 1.9 is a probabilisitic expression, in the setting of Gaussian fields, of this deterministic relationship between excursion set components and critical points of various types.

We also briefly discuss the assumptions required for our results (i.e. Conditions 1.1). The assumption that f is almost surely $C^{2+\nu}$ is necessary to apply the topological arguments that we borrow from Morse theory. The non-degeneracy assumptions on $\nabla^2 f(0)$ and $(f(t) - f(0), \nabla f(t), \nabla f(0))$ are used with the Kac–Rice theorem (see Sect. 2) to show certain non-degeneracy properties of f, which are again necessary for our topological arguments.

It is natural, therefore, to ask whether our results still apply when these assumptions are weakened. While we suspect that this is true, we are not able to show it with our methods. On the other hand, for a certain special case of non-trivial degenerate field – namely, the case where the spectral measure ρ is supported on at most five points –

we are able to give a complete description of c_{LS} and c_{ES}, which in particular shows that the main results hold also in this case (see Sect. 1.4 below).

1.3 Bounds on c_{LS} and c_{ES}

It is possible to bound the expected number of excursion sets or level sets of stationary Gaussian fields using local estimates (i.e. estimates which depend only on the derivatives of κ at the origin). Here we outline how these estimates apply to the functionals c_{ES} and c_{LS} and how they can be better characterised by making use of our results.

Corollary 1.12 *Let f be a Gaussian field satisfying Conditions 1.1 with covariance function κ. Then for $\ell \in \mathbb{R}$*

$$c_{ES}(\ell) - c_{ES}(-\ell) = \sqrt{\det \nabla^2 \kappa(0)} \, \frac{\ell}{2\pi} \, \phi(\ell) \tag{4}$$

where ϕ denotes the standard normal probability density function.

Remark 1.13 Since $-\nabla^2 \kappa(0)$ is the covariance matrix of $\nabla f(0)$, this result implies that the difference $c_{ES}(\ell) - c_{ES}(-\ell)$ depends only on the covariance of $\nabla f(0)$ and not on the covariance of higher order derivatives of f.

Remark 1.14 This result was proven for isotropic fields in [16] using a winding number calculation. We prove this result using Theorem 1.9 which simplifies the calculation in the non-isotropic case and also highlights an interesting identity. Specifically, the proof is as follows: substituting (3) into the left hand side of (4) and using the symmetries $p_{m+}(x) = p_{m-}(-x)$, $p_{s+}(x) = p_{s-}(-x)$ and $p_{s-} + p_{s+} = p_s$, we see that

$$c_{ES}(\ell) - c_{ES}(-\ell) = \int_\ell^\infty p_{m+}(x) - p_s(x) + p_{m-}(x) \, dx$$
$$= \mathbb{E}(N_{m+}(\ell) - N_s(\ell) + N_{m-}(\ell))$$

where $N_{m+}(\ell)$, $N_s(\ell)$ and $N_{m-}(\ell)$ are the number of critical points above level ℓ as defined in Proposition 1.8 for $\Omega = B(1/\sqrt{\pi})$. Lemma 11.7.1 of [1] states that this alternating sum is precisely the right hand side of (4).

The alternating sum of critical points of a function of different indices above a certain level can be used to calculate the Euler characteristic of the excursion set of the function above the level (see Chapter 9 of [1]). When working on finite subsets of the plane, boundary effects must be considered, but these become negligible as the area of the subset increases. Formally, if we let $\varphi(A)$ denote the Euler characteristic of a set A, then Theorem 11.7.2 of [1] gives an expression for the expected Euler characteristic of an excursion set on a cube (including boundary effects) which implies that

$$c_{ES}(\ell) - c_{ES}(-\ell) = \lim_{R \to \infty} \frac{1}{R^2} \mathbb{E}\left(\varphi\left(\{x \in [0, R]^2 : f(x) \geq \ell\} \right) \right)$$

Corollary 1.12 immediately gives a lower bound for $c_{ES}(\ell)$ since $c_{ES}(-\ell) \geq 0$. We now consider an upper bound on c_{LS}. This is most easily formulated when f is an isotropic Gaussian field. In this case its covariance function may be expressed as $\kappa(x) = K(|x|)$. It is shown in [5] that

$$\lambda := \frac{-\sqrt{3} K^{(2)}(0)}{\sqrt{K^{(4)}(0)}} \in \left(0, \sqrt{2}\right] \quad \text{and} \quad \eta^2 := \frac{-6 K^{(2)}(0)}{K^{(4)}(0)} \in [0, \infty) \tag{5}$$

parameterise the critical point densities p_{m+}, p_{m-} and p_s of isotropic fields in any dimension. It can also be shown that, for planar isotropic fields, if $\lambda = \sqrt{2}$ then $\kappa(x) = J_0(\sqrt{8/\eta^2}\,|x|)$, where J_0 is the 0-th Bessel function. When $\eta^2 = 8$, this particular field is known as the random plane wave (hereafter abbreviated to RPW), and is an object of great interest as it is the subject of the Berry conjecture [17].

Proposition 1.15 ([16]) *Let f be an isotropic Gaussian field satisfying Conditions* 1.1 *with covariance function $\kappa(x) = K(|x|)$. Then for $\ell \geq 0$*

$$c_{LS}(\ell) \leq \frac{\lambda^2}{\pi \eta^2} \phi(\ell) \left(\frac{2\sqrt{3 - \lambda^2}}{\lambda} \phi\left(\lambda\ell/\sqrt{3 - \lambda^2}\right) + \ell \left(2\Phi\left(\lambda\ell/\sqrt{3 - \lambda^2}\right) - 1\right)\right). \tag{6}$$

where ϕ and Φ denote the standard normal probability density and cumulative density functions respectively.

In particular, if $\lambda^2 > \frac{6e}{2e+\pi} \approx 1.9$ then by Corollary 1.12, $c_{LS}(0) < c_{LS}(1)$ *so that c_{LS} is at least bimodal (that is, it has at least two local maxima).*

Remark 1.16 This result is proven in [16] using the method of 'flip points'. Specifically, for any fixed direction u, the number of level set components at level ℓ in a finite region is bounded above by half the number of points t in the region such that $f(t) = \ell$ and $\partial_u f(t) = 0$ where ∂_u denotes the partial derivative in the direction u. The expected number of such points can be computed using the Kac–Rice formula. For isotropic fields, the choice of direction u is irrelevant. This method could also be applied to non-isotropic fields, and the bound could be optimised over the direction u, but we omit this for simplicity.

It is also possible to construct an upper bound on c_{LS} using the inequality $0 \leq p_{s-} \leq p_s$ and the densities p_{m+}, p_s and p_{m-}, which are explicitly known for isotropic fields. However this bound is larger than the bound in (6) at all levels ℓ.

Remark 1.17 Figure 2 shows lower and upper bounds for c_{ES} and c_{LS} for the RPW based on (4), (6) and the equality $c_{LS}(\ell) = c_{ES}(\ell) + c_{ES}(-\ell)$ (recall that $\lambda = \sqrt{2}$ and $\eta^2 = 8$ in this case). Although these bounds are not particularly tight for ℓ close to 0, they quickly become accurate as $|\ell|$ increases. For $\ell \geq 1$ both the upper and lower bounds on $c_{LS}(\ell)$ are within 5.1% of the true value (since the upper bound is within 5.1% of the lower bound) while for $\ell \geq 1.5$ both bounds are within 0.6% of the true value.
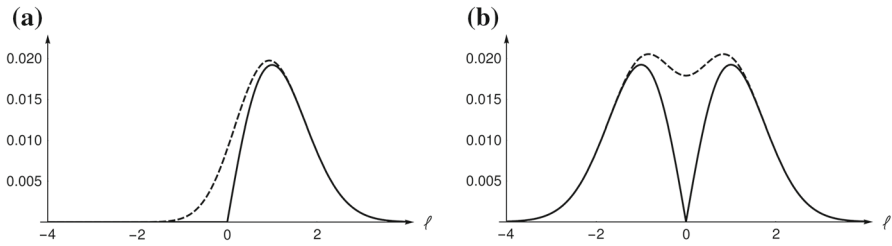
Fig. 2 Subfigures **a** and **b** show lower bounds (solid) and upper bounds (dashed) for $c_{ES}(\rho, \ell)$ and $c_{LS}(\rho, \ell)$ respectively in the RPW case, for which $\lambda = \sqrt{2}$ and $\eta^2 = 8$
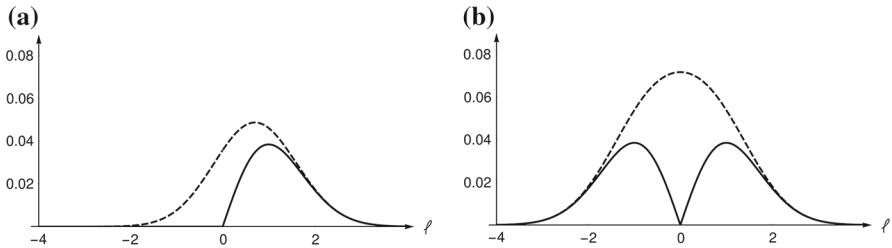

Fig. 3 Subfigures **a** and **b** show lower bounds (solid) and upper bounds (dashed) for $c_{ES}(\rho, \ell)$ and $c_{LS}(\rho, \ell)$ respectively, where ρ is the density of a standard two-dimensional Gaussian random vector (i.e. the spectral measure of the Bargmann–Fock field), for which $\lambda = 1$ and $\eta^2 = 2$

For fields with lower values of λ, the percentage difference between the upper and lower bounds on $c_{LS}(\ell)$ is a bit larger. Figure 3 shows the corresponding bounds on c_{ES} and c_{LS} for the Bargmann-Fock field: the centred, planar Gaussian field with covariance kernel $\kappa(x) = \exp(-|x|^2/2)$, for which $\lambda = 1$ and $\eta^2 = 2$. In this case for $\ell \geq 1$ the upper and lower bounds on $c_{LS}(\ell)$ are within 40% of the true value, and the corresponding accuracies for $\ell \geq 1.5$, $\ell \geq 2$ and $\ell \geq 2.5$ are 14%, 5% and 2% respectively.

Remark 1.18 Corollary 1.12 states that $c_{ES}(\rho, \ell) > 0$ for all $\ell > 0$, and so $c_{LS}(\rho, \ell) > 0$ for all $\ell \neq 0$, however the positivity of $c_{LS}(\rho, 0)$ is not addressed by this method. In [13] Nazarov and Sodin provide sufficient conditions for the positivity of $c_{LS}(\rho, 0)$ in terms of the spectral measure which cover almost all non-degenerate cases.

The only current quantitative lower bound for $c_{LS}(\rho, 0)$ has been proven for the RPW. In this case it has been shown that $c_{LS}(0) \geq 1.1 \times 10^{-5}$ (see [9]) although numerical simulations suggest that $c_{LS}(0) \approx 0.0589/(4\pi) \approx 0.00469$ (see [3] and references therein).

Although the bounds in the previous two corollaries can be proven by local methods (i.e. without using our analysis of saddle points) there is no scope for improving these estimates using the same methods. However it may be possible to get tighter bounds on c_{LS} and c_{ES} through Theorem 1.9 and a better characterisation of p_{s^-}.

Using the explicitly known densities p_{m^+}, p_{m^-} and p_s along with Theorem 1.9 allows us to derive some monotonicity properties of c_{ES} and c_{LS}.

Corollary 1.19 *Let f satisfy Conditions 1.1 and λ be defined as above. Then $c_{ES}(\rho, \ell)$ and $c_{LS}(\rho, \ell)$ are strictly decreasing in ℓ for $\ell > \sqrt{2}/\lambda$, so that any local maxima of c_{LS} must be contained in $[-\sqrt{2}/\lambda, \sqrt{2}/\lambda]$ and any local maxima of c_{ES} must be contained in $(-\infty, \sqrt{2}/\lambda]$.*

If $\lambda = \sqrt{2}$, then $c_{ES}(\ell)$ is non-decreasing on $(-\infty, 0]$ and strictly decreasing on $[1, \infty)$, so that any strict local maxima of c_{ES} must be contained in $[0, 1]$.

1.4 A special class of degenerate fields

As mentioned above, our methods do not cover the case in which ρ is supported on the union of two lines through the origin. In this section we consider a certain special class of such fields for which we can give a more or less complete description of c_{LS} and c_{ES}.

If the spectral measure ρ is supported on a single line through the origin, it is well known that f is almost surely constant in one direction and so in particular has no compact level domains. In this case, all of our results hold trivially with c_{LS}, c_{ES} and the critical point densities identically equal to zero. The simplest non-trivial case of degenerate ρ are spectral measures which are supported on four or five points. In this case, we can compute c_{ES} and c_{LS} explicitly.

Recall that a random variable Y is said to be Rayleigh distributed with parameter $\sigma^2 > 0$, denoted $Y \sim \text{Ray}(\sigma)$, if $\mathbb{P}(Y \leq x) = 1 - e^{-x^2/(2\sigma^2)}$ for all $x \geq 0$.

Proposition 1.20 *Let f be the Gaussian field with spectral measure*

$$\rho = \alpha\delta_0 + \frac{\beta}{2}(\delta_K + \delta_{-K}) + \frac{\gamma}{2}(\delta_L + \delta_{-L})$$

where $\beta, \gamma > 0$, $\alpha = 1 - \beta - \gamma \geq 0$ and $K, L \in \mathbb{R}^2$ are linearly independent. Then

$$\mathbb{E}[N_{ES,R}(\ell)] = c_{LS}(\ell) \cdot \pi R^2 + O(R)$$

and

$$\mathbb{E}[N_{LS,R}(\ell)] = c_{ES}(\ell) \cdot \pi R^2 + O(R)$$

where

$$c_{LS}(\ell) = |K \times L| \cdot \mathbb{P}(|Y_1 - Y_2| \leq \ell + X_0 \leq Y_1 + Y_2),$$
$$c_{ES}(\ell) = |K \times L| \cdot \mathbb{P}(|Y_1 - Y_2| \leq |\ell + X_0| \leq Y_1 + Y_2),$$

$\times$ denotes the cross product, $X_0 \sim \mathcal{N}(0, \alpha)$, $Y_1 \sim \text{Ray}(\sqrt{\beta})$, $Y_2 \sim \text{Ray}(\sqrt{\gamma})$ and X_0, Y_1, Y_2 are independent. Moreover the constants implied by the $O(\cdot)$ notation in these expressions are independent of ℓ.

So in particular, $c_{LS}(\ell) = 0$ if and only if $\ell = \alpha = 0$. If $c_{LS}(\ell) \neq 0$ then $N_{LS,R}(\ell)/(\pi R^2)$ converges in L^1 to a non-constant random variable and hence

does not converge a.s. to a constant, and this statement also holds for c_{ES} and $N_{ES,R}(\ell)/(\pi R^2)$.

Furthermore there exist functions p_{m^+}, p_{s^-}, p_{m^-}, p_{s^+} and p_s satisfying the conclusions of Proposition 1.8, and these are defined by

$$p_{m^+}(x) = p_{m^-}(-x) = |K \times L| \cdot p_{X_0 + Y_1 + Y_2}(x)$$
$$p_{s^-}(x) = p_{s^+}(-x) = |K \times L| \cdot p_{X_0 + |Y_1 - Y_2|}(x)$$

where p_Z denotes the probability density of a random variable Z. Therefore the equalities in the conclusion of Theorem 1.9 hold for f.

Remark 1.21 A stationary centred continuous Gaussian field is ergodic if and only if its spectral measure has no atoms [12, Section 6.1], and so the fields considered in this proposition are not ergodic. The fact that $N_{LS,R}(\ell)/(\pi R^2)$ does not generally converge to $c_{LS}(\ell)$ for these fields shows that the ergodicity requirement in the second part of Theorems 1.2 and 1.3 cannot be entirely relaxed.

Remark 1.22 It has previously been shown that if $\mathrm{spt}(\rho)$ has precisely four points then $c_{LS}(0) = 0$ (see [10]). Proposition 1.20 shows that $c_{LS}(0) > 0$ when $\mathrm{spt}(\rho)$ has five points, and that $c_{LS}(\ell) > 0$ for $\ell \neq 0$ when $\mathrm{spt}(\rho)$ has four or five points.

Remark 1.23 Figures 4 and 5 show c_{ES} and c_{LS} for different values of α, β, γ when $K = (1, 0)$ and $L = (0, 1)$. Figure 4 suggests that c_{LS} is bimodal whenever $\alpha = 0$ (we know that it is at least bimodal, since $c_{LS}(0) = 0$ and $c_{LS}(\ell) > 0$ for $\ell \neq 0$ in this case), however for $\alpha > 0$, c_{LS} can be bimodal or unimodal. Combining this observation with Corollary 1.12 raises the interesting question of determining for which Gaussian fields $c_{LS}(\ell)$ is bimodal.

Remark 1.24 Figures 4 and 5 also suggest that the derivatives of c_{ES} and c_{LS} need not be continuous at zero when $\alpha = 0$. This can be verified analytically since

$$c'_{ES}(\ell) = p_{s^-}(\ell) - p_{m^+}(\ell)$$

and these densities are explicitly known. In particular, it can be shown that $p_{Y_1 + Y_2}$ is everywhere continuous, whereas $p_{|Y_1 - Y_2|}$ has a jump discontinuity at zero (but is continuous elsewhere). So c'_{ES} is continuous except at zero. An analogous calculation permits the same conclusion for c'_{LS}. On the other hand, when $\alpha \neq 0$, c_{ES} and c_{LS} are continuously differentiable everywhere. A major unresolved question arising from our work is to determine whether c_{LS} and c_{ES} are continuously differentiable more generally (outside degenerate cases).

1.5 Summary of the rest of the paper

The remainder of the paper is organised as follows. In Sect. 2 we state the main (completely deterministic) topological lemma underlying our results (Lemma 2.5) and combine this with Proposition 1.8 to prove Theorems 1.2, 1.3 and 1.9 and their
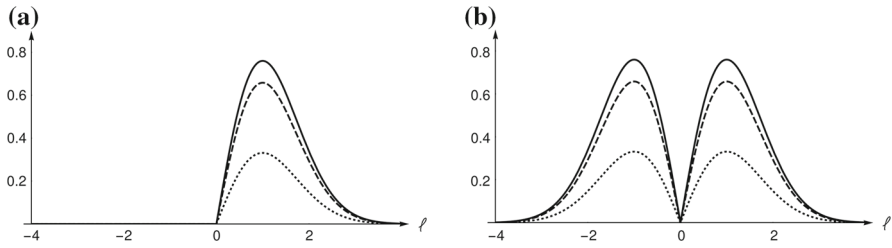
(a)

(b)

Fig. 4 The functions $c_{ES}(\ell)$ (left) and $c_{LS}(\ell)$ (right) with $\alpha = 0$ for $\beta - \gamma = 0$ (solid), $\beta - \gamma = 0.5$ (dashed) and $\beta - \gamma = 0.9$ (dotted) respectively
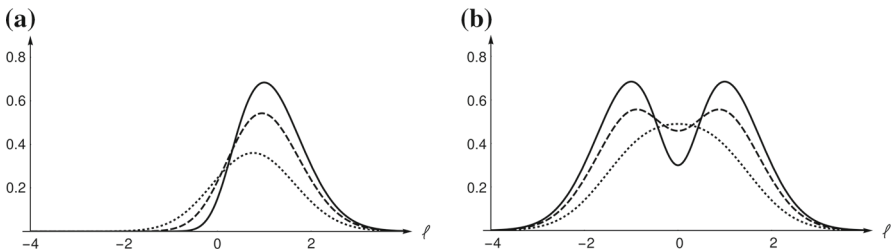
(a)

(b)

Fig. 5 The functions $c_{ES}(\ell)$ (left) and $c_{LS}(\ell)$ (right) with $\beta = \gamma$ for $\alpha = 0.1$ (solid), $\alpha = 0.3$ (dashed) and $\alpha = 0.6$ (dotted) respectively

corollaries. In Sect. 3 we give the full definition of upper/lower connected saddle points and prove Proposition 1.8. In Sect. 4 we prove Lemma 2.5 via a series of steps, and also prove Proposition 1.20, thereby completing the proof of all results in the paper.

2 Proof of the main results

The first step in proving Theorems 1.2, 1.3 and 1.9 is to translate our assumptions on the Gaussian field f into topological properties; these properties differ slightly depending on whether or not f is periodic.

Definition 2.1 Let $g : \mathbb{R}^2 \to \mathbb{R}$ be a function which is not constant in any direction; that is, there is no $v \in \mathbb{R}^2 \backslash \{0\}$ such that for all $x \in \mathbb{R}^2$ and $a \in \mathbb{R}$, $g(x) = g(x+av)$. We say that g is *doubly periodic* if there exist two linearly independent vectors $y, z \in \mathbb{R}^2$ such that for all $x \in \mathbb{R}^2$, $g(x + y) = g(x) = g(x + z)$. We say that g is *singly periodic* if it is not doubly periodic and there exists $y \in \mathbb{R}^2 \backslash \{0\}$ such that for all $x \in \mathbb{R}^2$, $g(x) = g(x + y)$. If neither of these conditions holds we say that g is *aperiodic*.

We note that this definition applies to deterministic functions. We say that a random field f is *doubly periodic* if there exist linearly independent vectors $y, z \in \mathbb{R}^2$ such that with probability one, $f(x + y) = f(x) = f(x + z)$ for all $x \in \mathbb{R}^2$. We make an analogous definition for singly periodic fields, and we say that a random field is *aperiodic* if it is neither doubly periodic nor singly periodic. The Gaussian field with spectral measure ρ is doubly periodic if and only if there exists $A \in GL(2, \mathbb{R})$ such

that $\{Ax : x \in \mathrm{spt}(\rho)\} \subset \mathbb{Z}^2$, and similarly is singly periodic if and only if it is not doubly periodic and there exists $A \in GL(2, \mathbb{R})$ such that $\{Ax : x \in \mathrm{spt}(\rho)\} \subset \mathbb{Z} \times \mathbb{R}$.

For our purposes, it is more natural to consider periodic functions as being defined on the torus or cylinder, and so we will specify the topological properties of Gaussian fields in terms of these domains.

For any doubly periodic function $g : \mathbb{R}^2 \to \mathbb{R}$ we can choose two linearly independent vectors $y, z \in \mathbb{R}^2$ satisfying the conditions in Definition 2.1 with minimum distance to the origin, and g is then specified entirely by its values on the parallelogram

$$P := \left\{ x \in \mathbb{R}^2 : x = ty + sz, \ t, s \in [0, 1) \right\}.$$

We call this the *associated parallelogram* for g and call y, z *periodic vectors* for g (they are not unique since we could also choose $-y, -z$). By identifying P with the 2-dimensional torus $\mathbb{T}$, we let $g_{\mathbb{T}}$ be g defined on the torus. By the choice of y, z above, when g is a stationary random field there are no two distinct points in P on which g is almost surely equal.

If g is singly periodic then we can choose $y \in \mathbb{R}^2 \backslash \{0\}$ satisfying the conditions in Definition 2.1 with minimum distance to the origin; we call this a *periodic vector* for g. By rotating the axes we may assume that g is periodic in the direction of the x-axis (i.e. $y = (y_1, 0)$) so that g is specified entirely by its values on $[0, y_1] \times \mathbb{R}$. This strip can be identified with the infinite cylinder $\mathcal{C} := S^1 \times \mathbb{R}$ under the quotient relationship that $x \sim z$ if and only if $x - z = (my_1, 0)$ for some $m \in \mathbb{Z}$. We will often work with the compact subset $\mathcal{C}(n) := [0, y_1] \times [-n, n]$ under the same quotient relationship. We let $g_{\mathcal{C}}$ and $g_{\mathcal{C}(n)}$ denote g restricted to these surfaces under the previous identification.

We next introduce some elements of Morse theory, in particular the concept of a Morse function on a manifold; here we follow [8]. We emphasise that, unless explicitly stated, a manifold M may or may not have a boundary ∂M. We also emphasise that we will only ever work with M being the plane $\mathbb{R}^2$, the closure of the disc $B(R)$, the torus $\mathbb{T}$, or the cylinders $\mathcal{C}$ and $\mathcal{C}(n)$, and so it is sufficient to bear these examples in mind.

Definition 2.2 Let M be an n-dimensional Riemannian manifold and $f \in C^2_{\mathrm{loc}}(M)$. (If $\partial M \neq \emptyset$, we take this to mean that for any coordinate chart $\mathbf{x}$, the function $f \circ \mathbf{x}^{-1}$ can be extended to a C^2 function on an open subset of $\mathbb{R}^n$. In particular this means that $f|_{\partial M}$ is twice continuously differentiable.) We say that f is *Morse* if the following hold (with any condition depending on ∂M holding implicitly if M has no boundary):

1. The critical points of f and $f|_{\partial M}$ are non-degenerate;
2. None of the critical points of f are contained in ∂M;
3. If ∇f is tangent to ∂M at p and u is a unit vector normal to ∂M at p, then $\partial_{\nabla f(p)} \partial_u f(p) \neq 0$ where $\partial_{\nabla f(p)}$ and ∂_u denote the directional derivatives in the directions $\nabla f(p)$ and u respectively;
4. The critical points of f and $f|_{\partial M}$ all occur at distinct levels.

If M is a manifold without boundary this simplifies to the requirement that all critical points of f are non-degenerate and occur at distinct levels.

We note that in the above definition $f|_{\partial M}$ is viewed as a function on the $(n-1)$-dimensional manifold ∂M so that a critical point of $f|_{\partial M}$ is not necessarily a critical point of f.

Using the definition of a Morse function, we next introduce a set of assumptions that a stationary random field must satisfy in order for our main results to hold. In the lemma that immediately follows, we will claim in particular that these assumptions are satisfied for Gaussian fields satisfying Conditions 1.1, but we have chosen to isolate these assumptions to illustrate the limited role of Gaussianity in the proof of the main results.

We say that a point $t \in \partial M$ is a *tangent point* of f if t is a critical point of the restricted function $f|_{\partial M}$ (the name comes from the fact that the level set $\{f = f(t)\}$ will be tangent to ∂M at t). We will also use the term 'tangent point' to describe a local extremum of the restriction of f to a finite union of line segments. Let $N_{\text{tang}}(A)$ denote the number of tangent points of f in A (where A is the boundary of a manifold or a finite union of line segments) and $N_{\text{crit}}(M)$ the number of critical points of f in M. We will only use these terms when A and M are compact, so that by the first point in the definition of a Morse function, the number of critical and tangent points will be finite.

Conditions 2.3 *A stationary random field f satisfies the following:*

1. *If f is aperiodic, then for each $R > 0$, $f|_{\overline{B(R)}}$ is almost surely Morse (taking a union over $R \in \mathbb{N}$, this implies also that f is almost surely Morse on $\mathbb{R}^2$);*
2. *If f is doubly periodic, then $f_{\mathbb{T}}$ is almost surely Morse;*
3. *If f is singly periodic, then for each $n \in \mathbb{N}$, $f_{\mathcal{C}(n)}$ is almost surely Morse (taking a union over $\mathbb{N}$, this implies that $f_{\mathcal{C}}$ is almost surely Morse);*
4. *If f is periodic and $L \subset \mathbb{R}^2$ is a line segment of length R, then $\mathbb{E}(N_{tang}(L)) = c_\theta R$ for a constant $c_\theta > 0$ that depends only on the direction of L;*
4. $\mathbb{E}(N_{crit}(B(R))) = cR^2$ *for a constant $c > 0$;*
5. $\mathbb{E}(N_{tang}(\partial B(R))) = O(R)$ *as $R \to \infty$.*

Lemma 2.4 *Let f satisfy Conditions 1.1. Then f satisfies Conditions 2.3.*

Proof All six conditions are established using variations of the Kac–Rice formula. This is an important tool in the study of Gaussian fields, and random functions more generally, which relates moments of the number of zeroes of a field to conditional moments of its gradient/derivative. See [14] for an introduction to the Kac–Rice formula or [1, Chapter 11] for a rigorous derivation.

(1)–(3). We first note that since $f \in C^2_{\text{loc}}(\mathbb{R}^2)$ almost surely by assumption, $f|_M \in C^2_{\text{loc}}(M)$ almost surely whenever M is $\overline{B(R)}$, $\mathbb{T}$ or $\mathcal{C}(n)$. Corollary 11.3.5 of [1] states that $f|_M$ has the first two properties of a Morse function if there exists a countable atlas $\{\mathbf{x}_i\}_{i \in I}$ such that for each $i \in I$, the covariance kernel of $f^{(i)} := f|_M \circ \mathbf{x}_i^{-1}$ is C^{4+} (on its domain of definition) and the Gaussian vectors $(\partial_x f^{(i)}(0), \partial_{xx} f^{(i)}(0), \partial_{xy} f^{(i)}(0))$ and $(\partial_y f^{(i)}(0), \partial_{yy} f^{(i)}(0), \partial_{xy} f^{(i)}(0))$ are non-degenerate (where ∂_x denotes taking the derivative with respect to the first variable, ∂_{xx} denotes taking the second derivative with respect to the first variable and we make analogous definitions for ∂_y, ∂_{yy} and ∂_{xy}). For each choice of M that we consider (i.e. $\overline{B(R)}$, $\mathbb{T}$ or $\mathcal{C}(n)$) it is clear that we
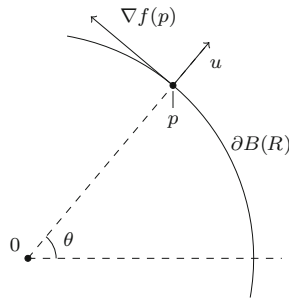
Fig. 6 If $p = (R\cos(\theta), R\sin(\theta))$ and $\nabla f(p)$ is orthogonal to $u := (\cos(\theta), \sin(\theta))$, we require that $\partial_{\nabla f(p)}\partial_u f(p) \neq 0$

can choose a finite atlas of charts $\mathbf{x}_i$ and that $f|_M \circ \mathbf{x}_i^{-1}$ will be a translation of f (restricted to some open set) for each such chart. Therefore the covariance kernel of $f^{(i)}$ is a restriction of κ, which is C^{4+} by assumption, and all that remains is to show that $(\partial_x f(t), \partial_{xx} f(t), \partial_{xy} f(t))$ and $(\partial_y f(t), \partial_{yy} f(t), \partial_{xy} f(t))$ are non-degenerate Gaussian vectors for any $t \in \mathbb{R}^2$. By assumption $\nabla f(t)$ and $\nabla^2 f(t)$ are non-degenerate Gaussian vectors for any $t \in \mathbb{R}^2$. It is a standard fact, for Gaussian fields with constant variance, that $\nabla f(t)$ and $\nabla^2 f(t)$ are independent for fixed t (see e.g. [1, Chapter 5]) and this proves the necessary non-degeneracy. (This part of the result does not depend on whether f is periodic.)

We note that $f|_{\overline{B(R)}}$ has the third property of a Morse function provided that there is no $\theta \in [0, 2\pi]$ such that

$$0 = g_1(\theta) := \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix}^T \nabla f(p)$$

$$0 = g_2(\theta) := \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix}^T \nabla^2 f(p) \begin{pmatrix} -\sin(\theta) \\ \cos(\theta) \end{pmatrix}$$

where $p = (R\cos(\theta), R\sin(\theta))$ (see Fig. 6).

If we assumed that f is C^3, then Bulinskaya's lemma would imply that no such θ exists almost surely. Since we assume only that f is $C^{2+\nu}$, we use a different argument which is adapted from the proof of [1, Lemma 11.2.11]. Since $f \in C_{\text{loc}}^{2+\nu}(\mathbb{R}^2)$ almost surely, g_2 is almost surely ν-Hölder continuous. Therefore for each $\delta > 0$ we can find $C_\delta, D_\delta > 0$ such that, with probability at least $1 - \delta$, on $\partial B(R)$ the ν-Hölder norm of g_2 is at most C_δ and each element of $\nabla^2 f$ is bounded in absolute value by D_δ; we denote this event by A_δ.

Now let I_{mj} be a collection of open intervals in $\mathbb{R}$ such that $[0, 2\pi] \subset \cup_j I_{mj}$ for each m, $\sum_j \text{Length}(I_{mj}) \to 2\pi$ as $m \to \infty$, and $\text{Length}(I_{mj}) \to 0$ uniformly in j as $m \to \infty$. Furthermore let θ_{mj} be the midpoint of each I_{mj} and define

$$E_{mj} = \left\{ \exists \theta \in I_{mj} : g_1(\theta) = g_2(\theta) = 0 \right\}.$$

On the event $A_\delta \cap E_{mj}$ we note that

$$|g_1(\theta_{mj})| \leq c_1 D_\delta \text{Length}(I_{mj}) \quad \text{and} \quad |g_2(\theta_{mj})| \leq c_1 C_\delta \text{Length}(I_{mj})^\nu$$

for some universal constant $c_1 > 0$. Therefore

$$
\begin{aligned}
\mathbb{P}(A_\delta \cap E_{mj}) \\
\leq \int_{|x| \leq c_1 D_\delta \text{Length}(I_{mj})} \mathbb{P}\left(|g_2(\theta_{mj})| \leq c_1 C_\delta \text{Length}(I_{mj})^\nu \middle| g_1(\theta_{mj}) = x\right) p_{g_1(\theta_{mj})}(x)\, dx \\
= \int_{|x| \leq c_1 D_\delta \text{Length}(I_{mj})} \mathbb{P}\left(|g_2(\theta_{mj})| \leq c_1 C_\delta \text{Length}(I_{mj})^\nu\right) p_{g_1(\theta_{mj})}(x)\, dx \qquad (7)
\end{aligned}
$$

where $p_{g_1(\theta_{mj})}$ is the probability density function of $g_1(\theta_{mj})$ and we have used the fact that $g_1(\theta_{mj})$ is independent of $g_2(\theta_{mj})$. Each $g_i(\theta)$ is a linear combination of elements of a non-degenerate Gaussian vector and so has non-zero variance, therefore on the compact region $[0, 2\pi]$ these variances are bounded away from zero. Since these are Gaussian fields, their univariate densities are bounded above by a constant times the inverse of their standard deviations, so we see that the density of each $g_i(\theta_{mj})$ is bounded above uniformly in m and j. We can therefore sum the expression on either side of (7) over j. Since $\text{Length}(I_{mj})$ converges to zero uniformly in j, so does the integrand in each term of the sum, and so we conclude that

$$\sum_j \mathbb{P}(A_\delta \cap E_{mj}) \to 0$$

as $m \to \infty$. Since this is true for any $\delta > 0$ and $\mathbb{P}(A_\delta) \geq 1 - \delta$ we conclude that there is almost surely no $\theta \in [0, 2\pi]$ such that $g_1(\theta) = g_2(\theta) = 0$, and hence f almost surely has the third property of a Morse function.

The proof that $f_{\mathcal{C}(n)}$ satisfies the third property of a Morse function is near identical. The argument above can be repeated for the boundary $S^1 \times \{-n, n\}$. Finally, $f_{\mathbb{T}}$ trivially satisfies the third property of a Morse function.

Next we consider the fourth property of a Morse function. We first suppose f is aperiodic and define $h_1 : T_k \to \mathbb{R}^5$ where $T_k = \left\{(s, t) \in \overline{B(R)}^2 \,\middle|\, |s - t| \geq 1/k\right\}$ by

$$h_1(s, t) = \begin{pmatrix} \nabla f(s) \\ \nabla f(t) \\ f(s) - f(t) \end{pmatrix}.$$

Bulinskaya's lemma (Lemma 11.2.10 of [1]) states that $\mathbb{P}(h_1^{-1}(0) = \emptyset) = 1$ provided that $h_1 \in C^1(T_k)$ almost surely and the univariate densities of h_1 are bounded uniformly in T_k. By assumption $h_1 \in C^1_{\text{loc}}(\mathbb{R}^2)$ and so, since h^1 is Gaussian, we need only show that the determinant of the covariance matrix of $(\nabla f(s), \nabla f(t), f(s) - f(t))$ is bounded away from zero on T_k. Since f is aperiodic, $f(s) - f(t)$ is a non-degenerate Gaussian variable for all $s \neq t$, and so by Conditions 1.1 (in particular using stationarity of f), the determinant of this covariance matrix is non-zero for all $s \neq t$. Since

this determinant is continuous in (s, t) it is bounded away from zero on the compact set T_k. Taking the countable union of T_k for $k \in \mathbb{N}$ shows that almost surely there are no points $s, t \in B(R)$ with $s \neq t$ such that $\nabla f(s) = \nabla f(t) = 0$ and $f(s) = f(t)$.

Let $f_\partial(\theta)$ be a parametrisation of $f|_{\partial B(R)}$. To show that f has no two tangent points at the same level and no tangent points at the level of any critical point, we apply identical arguments to the following two functions:

$$
h_2(\theta, \omega) = \begin{pmatrix} f'_\partial(\theta) \\ f'_\partial(\omega) \\ f_\partial(\theta) - f_\partial(\omega) \end{pmatrix} \quad \text{and} \quad h_3(t, \theta) = \begin{pmatrix} \nabla f(t) \\ f'_\partial(\theta) \\ f(t) - f_\partial(\theta) \end{pmatrix}.
$$

This completes the proof that $f|_{\overline{B(R)}}$ is a Morse function when f is aperiodic. When f is doubly periodic, we can repeat these arguments with

$$
T_k = \left\{ (s, t) \in P_k^2 : |s - t| \geq 1/k \right\}
$$

where $P_k = \{ ay + bz : a, b \in [0, 1 - 1/k] \}$ and y, z are the periodic vectors of f, to show that $f_{\mathbb{T}}$ is Morse. Finally in the singly periodic case we consider $C_k = [0, y_1 - 1/k] \times [-n, n]$ and $T_k = \left\{ (s, t) \in C_k^2 : |s - t| \geq 1/k \right\}$ where $y = (y_1, 0)$ is the periodic vector of f. In both of these cases, the choice of domain ensures that $f(t) - f(s)$ is non-degenerate, so that Conditions 1.1 can be used. This completes the proof of the first three points of the lemma.

(4). Let L be a line segment of length R and u a unit length vector parallel to L. Applying a Kac–Rice formula (for the particular form, see Corollary 11.2.2 of [1]) to $f|_L$ and using the stationarity of f along with the independence of $\nabla f(t)$ and $\nabla^2 f(t)$, we see that

$$
\mathbb{E}(N_{\text{tang}}(L)) = \int_L \mathbb{E}(|\partial_{uu} f(t)| \mid \partial_u f(t) = 0) p_{\partial_u f(t)}(0) \, dt
$$
$$
= R \, p_{\partial_u f(0)}(0) \mathbb{E}(|\partial_{uu} f(0)|) =: c_\theta R
$$

where $p_{\partial_u f(0)}$ is the probability density of $\partial_u f(0)$, and $c_\theta < \infty$ since $\partial_{uu} f(0)$ can be expressed as a linear combination of components of the non-degenerate Gaussian vector $\nabla^2 f(0)$ and so has finite moments of all orders.

(5). Applying the same Kac–Rice formula to f, and using the same arguments as above, shows that

$$
\mathbb{E}(N_{\text{crit}}(B(R))) = \int_{B(R)} \mathbb{E} \left(\left| \det \nabla^2 f(t) \right| \mid \nabla f(t) = 0 \right) p_{\nabla f(t)}(0) \, dt
$$
$$
= \pi R^2 p_{\nabla f(0)}(0) \mathbb{E} \left(\left| \det \nabla^2 f(0) \right| \right) =: c R^2.
$$

(6). We now explicitly parametrise $f|_{\partial B(R)}$ as $g(\theta) := f(R\cos(\theta), R\sin(\theta))$ and note that

$$g'(\theta) = -R\sin(\theta)\partial_x f(p) + R\cos(\theta)\partial_y f(p)$$
$$g''(\theta) = -R\cos(\theta)\partial_x f(p) - R\sin(\theta)\partial_y f(p)$$
$$+R^2\sin^2(\theta)\partial_{xx}f(p) - 2R^2\sin(\theta)\cos(\theta)\partial_{xy}f(p) + R^2\cos^2(\theta)\partial_{yy}f(p) \tag{8}$$

where $p = (R\cos(\theta), R\sin(\theta))$. Since $\nabla f(p)$ and $\nabla^2 f(p)$ are independent, a simple calculation shows that $(g'(\theta), g''(\theta))$ has non-degenerate distribution, so we can apply a Kac–Rice formula (specifically, Corollary 11.2.2 of [1]) to show that

$$\mathbb{E}(N_{\text{tang}}(\partial B(R))) = \mathbb{E}\left(\#\{\theta \in [0, 2\pi] \mid g'(\theta) = 0\}\right)$$
$$= \int_0^{2\pi} \mathbb{E}\left(|g''(\theta)| \mid g'(\theta) = 0\right) p_{g'(\theta)}(0)\, d\theta. \tag{9}$$

Since f is stationary

$$\text{Var}\left(g'(\theta)\right) = R^2 \mathbb{E}\left(\left(-\sin(\theta)\,\partial_x f(0) + \cos(\theta)\,\partial_y f(0)\right)^2\right).$$

The expectation on the right side of this equation is non-zero for each θ, since $\nabla f(0)$ is non-degenerate, and by compactness is bounded away from zero uniformly in θ. Therefore, by considering the density of a Gaussian random variable, there exists $c_0 > 0$ such that $p_{g'(\theta)}(0) \le c_0/R$ for all $\theta \in [0, 2\pi]$ and $R > 0$.

Using this observation, along with the fact that $\nabla f(0)$ is independent of $\nabla^2 f(0)$, and substituting (8) into (9), we see that there exists a constant $c_1 > 0$ independent of θ and R such that

$$\mathbb{E}(N_{\text{tang}}(\partial B(R))) \le c_1 R \int_0^{2\pi} \mathbb{E}\left(|\partial_{xx}f(0)| + |\partial_{yy}f(0)| + |\partial_{xy}f(0)|\right) d\theta$$
$$+ c_1 \int_0^{2\pi} \mathbb{E}\left(|\partial_x f(0)| + |\partial_y f(0)| \mid \sin(\theta)\partial_x f(p) = \cos(\theta)\partial_y f(p)\right) d\theta$$
$$= O(R).$$

$\square$

We now introduce the deterministic relationship between level sets and critical points that is the foundation of our results. We denote the number of critical points of f in $B(R)$ of type h with level in $[\ell_1, \ell_2]$ by $N_{h,R}[\ell_1, \ell_2]$, where $h = m^+, m^-, s^+, s^-$ denotes local maxima, local minima, upper connected saddles and lower connected saddles respectively (for an aperiodic function that is Morse on $\mathbb{R}^2$, Definition 1.7 defines the upper/lower connected saddle points; the definitions in the general case will be given in Sect. 3). Let us also generalise our earlier notation: for a deterministic, planar function f, we denote the number of components of $\{f = \ell\}$ and $\{f \ge \ell\}$ in $B(R)$ by $N_{LS,R}(\ell)$ and $N_{ES,R}(\ell)$ respectively.

In the case of a singly periodic function $f : \mathbb{R}^2 \to \mathbb{R}$ we also need to define internal and external rectangular tilings of $B(x, R)$ which have a finite number of horizontal and vertical line segments as boundary. Let $(y_1, 0)$ be the periodic vector of f, and let $S(n_1, n_2) = [n_1 y_1, (n_1+1)y_1] \times [n_2, n_2+1]$ where $n_1, n_2 \in \mathbb{Z}$. We define $B_{\text{int}}(x, R)$ to be the union over $n_1, n_2 \in \mathbb{Z}$ of all $S(n_1, n_2)$ contained in $B(x, R)$. Similarly we define $B_{\text{ext}}(x, R)$ to be the union over $n_1, n_2 \in \mathbb{Z}$ of all $S(n_1, n_2)$ which intersect $B(x, R)$.

Lemma 2.5 *Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a deterministic function. Suppose first that f is aperiodic and assume that f and $f|_{\overline{B(R)}}$ are Morse. Then there exists $c > 0$, independent of f, such that for each $\ell \in \mathbb{R}$,*

$$N_{LS,R}(\ell) = N_{m^+,R}[\ell, \infty) - N_{s^-,R}[\ell, \infty) + N_{s^+,R}(\ell, \infty) - N_{m^-,R}(\ell, \infty) + \eta_{\ell,R} \tag{10}$$

and

$$N_{ES,R}(\ell) = N_{m^+,R}[\ell, \infty) - N_{s^-,R}[\ell, \infty) + \gamma_{\ell,R} \tag{11}$$

where

$$\max\{|\eta_{\ell,R}|, |\gamma_{\ell,R}|\} \le c N_{tang}(\partial B(R)). \tag{12}$$

Suppose instead that $f : \mathbb{R}^2 \to \mathbb{R}$ is doubly periodic and assume that $f_{\mathbb{T}}$ is Morse and $f|_{\partial P}$ has a finite number of local extrema. Then there exists a constant $c_f > 0$ depending only on P, such that, for each $R > 0$ and $\ell \in \mathbb{R}$, (10) and (11) hold with (12) replaced by

$$\max\{|\eta_{\ell,R}|, |\gamma_{\ell,R}|\} \le c_f \cdot \left(R + N_{tang}(\partial B(R))\right).$$

Finally, suppose that $f : \mathbb{R}^2 \to \mathbb{R}$ is singly periodic with periodic vector $(y, 0)$ and assume that $f_{\mathcal{C}(n)}$ is Morse for each $n \in \mathbb{Z}$ and that $N_{tang}(\{0\} \times [-2R, 2R]) < \infty$. Then (10) and (11) still hold with (12) replaced by

$$\max\{|\eta_{\ell,R}|, |\gamma_{\ell,R}|\} \le c_f \left(N_{tang}(\partial B_{int}(0, R)) + N_{tang}(\partial B_{ext}(0, R)) \right.$$

$$\left. + N_{crit}(B(R + r_f) \backslash B(R - r_f)) \right)$$

where c_f and r_f are constants that depend only on the periodic vector $y = (y_1, 0)$ of f (so in particular, they are independent of ℓ and R).

The proof of Lemma 2.5 is contained in Sect. 4. The error term in the above estimate can be intuitively understood as the result of boundary effects from working with the domain $B(R)$ in the definition of $N_{LS,R}$ and $N_{ES,R}$, and although it appears quite complicated (especially in the case of singly periodic functions), when applied to the Gaussian field f it is bounded in expectation by $O(R)$ as a result of Lemma 2.4.

We are now ready to state the core technical results of the paper, which are versions of Theorems 1.2, 1.3 and 1.9 that hold for arbitrary random fields (i.e. not necessarily Gaussian) using only the properties contained in Conditions 2.3.

Proposition 2.6 *Let f be a stationary random field satisfying Conditions 2.3. For each $\ell \in \mathbb{R}$*

$$\mathbb{E}\left(N_{LS,R}(\ell)\right) = c_{LS}(\ell) \cdot \pi R^2 + O(R)$$

as $R \to \infty$, where

$$c_{LS}(\ell) := \frac{1}{\pi}\mathbb{E}\left(N_{m^+,1}[\ell,\infty) - N_{s^-,1}[\ell,\infty) + N_{s^+,1}(\ell,\infty) - N_{s^-,1}(\ell,\infty)\right).$$

The constant implied by the $O(\cdot)$ notation may depend on the distribution of f but is independent of ℓ. If in addition f is ergodic, then

$$\frac{1}{\pi R^2}N_{LS,R}(\ell) \xrightarrow{L^1,a.s.} c_{LS}(\ell).$$

Proposition 2.7 *Let f be a stationary random field satisfying Conditions 2.3. For each $\ell \in \mathbb{R}$*

$$\mathbb{E}\left(N_{ES,R}(\ell)\right) = c_{ES}(\ell) \cdot \pi R^2 + O(R)$$

as $R \to \infty$, where

$$c_{ES}(\ell) := \frac{1}{\pi}\mathbb{E}(N_{m^+,1}[\ell,\infty) - N_{s^+,1}[\ell,\infty)).$$

The constant implied by the $O(\cdot)$ notation may depend on the distribution of f but is independent of ℓ. If in addition f is ergodic, then

$$\frac{1}{\pi R^2}N_{ES,R}(\ell) \xrightarrow{L^1,a.s.} c_{ES}(\ell).$$

Remark 2.8 Since f is stationary, we could equivalently define $c_{LS}(\ell)$ in Proposition 2.6 as

$$c_{LS}(\ell) = \frac{1}{\pi R^2}\mathbb{E}\left(N_{m^+,R}[\ell,\infty) - N_{s^-,R}[\ell,\infty) + N_{s^+,R}(\ell,\infty) - N_{s^-,R}(\ell,\infty)\right)$$

for any $R > 0$ or make an analogous definition for $c_{ES}(\ell)$ in Proposition 2.7.

Together with Lemma 2.4 and Proposition 1.8, these propositions imply Theorems 1.2 and 1.3, and also the first two parts of Theorems 1.9.

Proof (Proposition 2.6) As f is stationary, the expected number of critical points of f of a particular type in a domain is proportional to the area of the domain. Since the field satisfies Conditions 2.3 we can apply Lemma 2.5, and taking expectations yields

$$\mathbb{E}(N_{LS,R}(\ell)) = \pi R^2 c_{LS}(\ell) + O(R).$$

Sending $R \to \infty$ proves the first two statements of the proposition.

The remainder of the proof follows the general roadmap of the original derivation of the existence of the Nazarov–Sodin constant in [13]. Suppose that f is ergodic and let $N_{h,R}^{(u)}$ denote the number of critical points of type h in $B(u, R)$ with level in $[\ell, \infty)$ or (ℓ, ∞) for $h = m^+, s^+$ and $h = m^-, s^-$ respectively. The 'sandwich estimate' of [13] (Lemma 1) can be slightly altered to show that for any $r \in (0, R)$

$$\frac{1}{\pi R^2} \int_{B(R-r)} \frac{N_{h,r}^{(u)}}{\pi r^2} \, du \leq \frac{N_{h,R}^{(0)}}{\pi R^2} \leq \frac{1}{\pi R^2} \int_{B(R+r)} \frac{N_{h,r}^{(u)}}{\pi r^2} \, du.$$

Applying Wiener's ergodic theorem (see [13, Section 6.1]) both of these integrals converge almost surely and in L^1 to the limit $\mathbb{E}\left(N_{h,1}^{(0)}\right)/\pi$. So in particular, $N_{h,R}^{(0)}/(\pi R^2)$ has the same limit, and

$$\frac{1}{\pi R^2} \left(N_{m^+,R}^{(0)} + N_{s^+,R}^{(0)} - N_{s^-,R}^{(0)} - N_{m^-,R}^{(0)}\right) \xrightarrow{L^1} c_{LS}(\ell).$$

Applying Lemma 2.5 with the bound on $\mathbb{E}(|\eta_{\ell,R}|)$ implied by Conditions 2.3 shows that

$$\frac{1}{\pi R^2} \mathbb{E}\left(\left|N_{LS,R}(\ell) - N_{m^+,R}^{(0)} - N_{s^+,R}^{(0)} + N_{s^-,R}^{(0)} + N_{m^-,R}^{(0)}\right|\right) = o(1)$$

as $R \to \infty$. Combining these results completes the proof of L^1 convergence.

We now extend our notation by defining $N_{LS,r}^{(u)}$ to be the number of components of $\{f = \ell\}$ in $B(u, r)$. The original 'sandwich estimate' of [13] states that

$$\frac{1}{\pi R^2} \int_{B(R-r)} \frac{N_{LS,r}^{(u)}}{\pi r^2} \, du \leq \frac{N_{LS,R}^{(0)}}{\pi R^2} \leq \frac{1}{\pi R^2} \int_{B(R+r)} \frac{\tilde{N}_{LS,r}^{(u)}}{\pi r^2} \, du$$

where $\tilde{N}_{LS,r}^{(u)}$ is the number of components of $\{f = \ell\}$ which intersect $\overline{B(u, r)}$. Since

$$\tilde{N}_{LS,r}^{(u)} = N_{LS,r}^{(u)} + O\left(\#\{x \in \partial B(u, r) \mid f(x) = \ell\}\right)$$

we can rearrange this estimate as

$$\left| \frac{N_{LS,R}^{(0)}}{\pi R^2} - \frac{1}{\pi R^2} \int_{B(R)} \frac{N_{LS,r}^{(u)}}{\pi r^2} \, du \right| \leq \frac{1}{\pi R^2} \int_{B(R+r) \backslash B(R-r)} \frac{N_{LS,r}^{(u)}}{\pi r^2} \, du$$
$$+ \frac{c_0}{\pi R^2} \int_{B(R+r)} \frac{\#\{x \in \partial B(u,r) \mid f(x) = \ell\}}{\pi r^2} \, du \tag{13}$$

for some universal constant $c_0 > 0$. Applying Lemma 2.5 inside the integral term on the left hand side of (13) shows that

$$\left| \frac{N_{LS,R}^{(0)}}{\pi R^2} - \frac{1}{\pi R^2} \int_{B(R)} \frac{N_{m^+,r}^{(u)}}{\pi r^2} + \frac{N_{s^+,r}^{(u)}}{\pi r^2} - \frac{N_{s^-,r}^{(u)}}{\pi r^2} - \frac{N_{m^-,r}^{(u)}}{\pi r^2} \, du \right|$$
$$\leq \frac{1}{\pi R^2} \int_{B(R+r) \backslash B(R-r)} \frac{N_{LS,r}^{(u)}}{\pi r^2} \, du$$
$$+ \frac{c_0}{\pi R^2} \int_{B(R+r)} \frac{\#\{x \in \partial B(u,r) \mid f(x) = \ell\}}{\pi r^2} \, du \tag{14}$$
$$+ \frac{c_f}{\pi R^2} \int_{B(R)} \frac{r + N_{\text{tang}}(B(u,r))}{\pi r^2} \, du$$
$$+ \frac{c_f}{\pi R^2} \int_{B(R)} \frac{N_{\text{tang}}(B_{\text{int}}(u,r)) + N_{\text{tang}}(B_{\text{ext}}(u,r))}{\pi r^2} \, du$$
$$+ \frac{c_f}{\pi R^2} \int_{B(R)} \frac{N_{\text{crit}}(B(u, r+r_f) \backslash B(u, r-r_f))}{\pi r^2} \, du.$$

(The upper bound here is a result of adding the upper bounds on the error terms in Lemma 2.5 in the cases that f is aperiodic, doubly periodic and singly periodic respectively, so that we can deal with all three cases at once.) From Wiener's ergodic theorem, the integral term within the absolute value signs will converge almost surely to $c_{LS}(\ell)$ and the first integral term on the right hand side will converge almost surely to zero. Applying the same argument to the remaining integral terms shows that they will each converge to a constant, $\Phi_1(r)$, $\Phi_2(r)$, $\Phi_3(r)$ and $\Phi_4(r)$ respectively, where

$$\Phi_1(r) = \frac{c_0}{\pi r^2} \mathbb{E} \left(\#\{x \in \partial B(r) \mid f(x) = \ell\} \right),$$
$$\Phi_2(r) = \frac{c_f}{\pi r} + \frac{c_f}{\pi r^2} \mathbb{E} \left(N_{\text{tang}}(B(r)) \right),$$
$$\Phi_3(r) = \frac{c_f}{\pi r^2} \left(\mathbb{E} \left(N_{\text{tang}}(B_{\text{int}}(0,r)) \right) + \mathbb{E} \left(N_{\text{tang}}(B_{\text{ext}}(0,r)) \right) \right),$$
$$\Phi_4(r) = \frac{c_f}{\pi r^2} \mathbb{E} \left(N_{\text{crit}}(B(u, r+r_f) \backslash B(u, r-r_f)) \right).$$

We now fix $r > 0$ and take the limit superior of (14) as $R \to \infty$ to show that

$$\limsup_{R \to \infty} \left| \frac{N_{LS,R}^{(0)}}{\pi R^2} - c_{LS}(\ell) \right| \le \Phi_1(r) + \Phi_2(r) + \Phi_3(r) + \Phi_4(r) \qquad (15)$$

almost surely. Since the number of boundary points of f at level ℓ is deterministically bounded above by a constant times the number of tangent points, we see that $\Phi_1(r) \le c_1 \Phi_2(r)$ for some $c_1 > 0$. Conditions 2.3 imply that $\Phi_2(r) + \Phi_3(r) + \Phi_4(r) = o(1)$ as $r \to \infty$. Therefore we may take a countable sequence $r_n \to \infty$, such that (15) holds almost surely for each r_n and the right hand side of (15) becomes arbitrarily small, to show that $N_{LS,R}^{(0)}/(\pi R^2)$ converges almost surely to $c_{LS}(\ell)$. $\qquad \square$

Proof (Proposition 2.7) This follows the proof of Proposition 2.6 almost exactly. Taking expectations of the second part of Lemma 2.5, using the stationarity of f and the bound on $\mathbb{E}(N_{\text{tang}}(\partial B(R)))$ implied by Conditions 2.3 proves the first two statements of the theorem.

The proof of the remainder of the theorem follows identically by defining $N_{ES,r}^{(u)}$ as the number of components of $\{f \ge \ell\}$ contained in $B(u, r)$, $\tilde{N}_{ES,r}^{(u)}$ as the number of components of $\{f \ge \ell\}$ intersecting $\overline{B(u, r)}$ and noting that the difference in these two terms is also bounded above by the number of tangent points of $\partial B(u, r)$. $\qquad \square$

To complete the proof of Theorem 1.9, it remains only to show the joint continuity of c_{LS} and c_{ES} with respect to the level and spectral measure.

Proof (Theorem 1.9 - Joint continuity) By Prokhorov's theorem the weak-$*$ topology on $\mathcal{P}_c$ is metrisable and so we can use the sequential definition of continuity for c_{LS} and c_{ES}. Let (ρ_n, ℓ_n) be a sequence in $\mathcal{P}_c \times \mathbb{R}$ converging to (ρ, ℓ). By the triangle inequality

$$|c_{LS}(\rho_n, \ell_n) - c_{LS}(\rho, \ell)| \le |c_{LS}(\rho_n, \ell_n) - c_{LS}(\rho_n, \ell)| + |c_{LS}(\rho_n, \ell) - c_{LS}(\rho, \ell)|. \qquad (16)$$

Theorem 1.3 of [10] states that the second term on the right hand side converges to 0 with n in the special case $\ell = 0$. However the proof of this theorem can be repeated verbatim replacing the field f_ρ with $f_\rho - \ell$ to show that $|c_{LS}(\rho_n, \ell) - c_{LS}(\rho, \ell)| \to 0$.

Now assume that $\ell_n \le \ell$. By the first part of Theorem 1.9 we see that

$$|c_{LS}(\rho_n, \ell_n) - c_{LS}(\rho_n, \ell)| \le (4/\pi) \mathbb{E}_{\rho_n} \left(N_{\text{crit}}^{(n)}[\ell_n, \ell] \right) \qquad (17)$$

where $N_{\text{crit}}^{(n)}[\ell_n, \ell]$ denotes the number of critical points of f_n (the Gaussian field with spectral measure ρ_n) in the circle of radius 1 with level in $[\ell_n, \ell]$. By the Kac–Rice theorem (Corollary 11.2.2 of [1])

$$\mathbb{E}_{\rho_n}\left(N_{\text{crit}}^{(n)}[\ell_n,\ell]\right) = \int_{B(1)} \mathbb{E}\left(\left|\det \nabla^2 f_n(t)\right| \mathbb{1}_{f_n(t)\in[\ell_n,\ell]} \,\Big|\, \nabla f_n(t) = 0\right) p_{\nabla f_n(t)}(0)\,dt$$

$$= \pi\, \mathbb{E}\left(\left|\det \nabla^2 f_n(0)\right| \mathbb{1}_{f_n(0)\in[\ell_n,\ell]}\right) p_{\nabla f_n(0)}(0)$$

$$\leq \mathbb{E}\left(\left|\det \nabla^2 f_n(0)\right|^2\right)^{\frac{1}{2}} \mathbb{P}\left(f_n(0) \in [\ell_n,\ell]\right)^{\frac{1}{2}} \frac{1}{2} \frac{1}{\sqrt{\det \text{Var}(\nabla f_n(0))}} \tag{18}$$

where we have used the independence of a field and its gradient at a point along with the Cauchy-Schwarz inequality. It is well known that the covariance structure of a stationary Gaussian field along with its derivatives at a point can be expressed in terms of the spectral measure (see [1, Chapter 5]). For example,

$$\text{Var}(\partial_x f_n(0)) = \int_{\mathbb{R}^2} \lambda_1^2\, d\rho_n(\lambda)$$

with similar expressions for other derivatives and covariances. Since all spectral measures we are considering are supported on $\overline{B(1)}$, the definition of weak-$*$ convergence implies that the covariance structure associated with each ρ_n at the origin converges to that of ρ. Specifically, if f denotes the field with spectral measure ρ, then

$$\text{Var}(f_n(0), \nabla f_n(0), \nabla^2 f_n(0)) \to \text{Var}(f(0), \nabla f(0), \nabla^2 f(0)).$$

Applying this to (18), we see that $\det \text{Var}(\nabla f_n(0))$ is bounded away from 0 in n. Similarly $\mathbb{E}\left(|\det \nabla^2 f_n(0)|^2\right)$ is uniformly bounded above in n. Finally we note that $\text{Var}(f_n(0))$ is uniformly bounded away from 0, so that choosing ℓ_n sufficiently close to ℓ ensures that $\mathbb{P}\left(f_n(0) \in [\ell_n,\ell]\right)$ is arbitrarily small. An identical argument works for $\ell_n \geq \ell$, and so combining these observations with (16)–(18) shows that

$$|c_{LS}(\rho_n,\ell_n) - c_{LS}(\rho,\ell)| \xrightarrow{n\to\infty} 0$$

as required. The proof of continuity of $c_{ES}(\rho,\ell)$ is almost identical. Although Theorem 1.3 of [10] is stated only for level sets, the proof of this result can be adapted to apply to excursion sets with no changes of any substance. □

Proof (Corollary 1.19) Since f is isotropic we can use the explicitly derived critical point densities for local maxima, local minima and saddle points from [5]. We recall that these are parametrised in terms of $\lambda \in (0,\sqrt{2}]$ and $\eta^2 \in [0,\infty)$ as defined by (5). If $0 < \lambda < \sqrt{2}$ then

$$p_{m+}(x) = p_{m-}(-x) = \frac{1}{\pi \eta^2}\left(\lambda^2(x^2-1)\phi(x)\Phi\left(\frac{\lambda x}{\sqrt{2-\lambda^2}}\right) + \frac{\lambda x\sqrt{2-\lambda^2}}{2\pi}e^{-\frac{x^2}{2-\lambda^2}}\right.$$

$$\left. + \frac{\sqrt{2}}{\sqrt{\pi(3-\lambda^2)}}e^{-\frac{3x^2}{2(3-\lambda^2)}}\Phi\left(\frac{\lambda x}{\sqrt{(3-\lambda^2)(2-\lambda^2)}}\right)\right)$$

$$p_s(x) = \frac{1}{\pi\eta^2}\frac{\sqrt{2}}{\sqrt{\pi(3-\lambda^2)}}e^{-\frac{3x^2}{2(3-\lambda^2)}}$$

where ϕ and Φ denote the standard normal probability density function and cumulative density function respectively. If $\lambda = \sqrt{2}$ then

$$p_{m+}(x) = p_{m-}(-x) = \frac{\sqrt{2}}{\pi^{3/2}\eta^2}\left((x^2-1)e^{-\frac{x^2}{2}} + e^{-\frac{3x^2}{2}}\right)\mathbb{1}_{x\geq 0}$$

$$p_s(x) = \frac{\sqrt{2}}{\pi^{3/2}\eta^2}e^{-\frac{3x^2}{2}}.$$

By Theorem 1.9,

$$c_{LS}(\ell+\epsilon) - c_{LS}(\ell) \leq \int_\ell^{\ell+\epsilon} p_{m-}(x) + p_s(x) - p_{m+}(x)\,dx.$$

We denote the integrand above by $I(x)$. In the case $\lambda \in (0, \sqrt{2})$, evaluating this expression using the standard Gaussian inequality $1 - \Phi(x) \leq (1/x)\phi(x)$ shows that

$$I(x) \leq \frac{1}{\pi\eta^2}\left(\frac{\sqrt{2-\lambda^2}}{\pi}\frac{1}{x}\frac{2-\lambda^2}{\lambda}e^{-\frac{x^2}{2-\lambda^2}} - \frac{\lambda^2}{\sqrt{2\pi}}(x^2-1)e^{-\frac{x^2}{2}}\right)$$

which is negative for $x > \sqrt{2}/\lambda$. For $\lambda = \sqrt{2}$ and $x > 0$

$$I(x) \leq -\frac{1}{4\pi\sqrt{2\pi}}(x^2-1)e^{-\frac{x^2}{2}}$$

which is negative for $x > 1 = \sqrt{2}/\lambda$. Similarly we have

$$c_{ES}(\ell+\epsilon) - c_{ES}(\ell) \leq \int_\ell^{\ell+\epsilon} p_s(x) - p_{m+}(x)\,dx,$$

which is less than our upper bound for $c_{LS}(\ell+\epsilon) - c_{LS}(\ell)$, and the first result follows. When $\lambda = \sqrt{2}$ and $\ell \leq 0$ we see that $p_{m+}(\ell) = 0$ and so by Theorem 1.9

$$c_{ES}(\ell+\epsilon) - c_{ES}(\ell) = \int_\ell^{\ell+\epsilon} p_{s-}(x)\,dx \geq 0.$$

Therefore c_{ES} is weakly increasing in this case.

$\square$

(a)   **(b)**
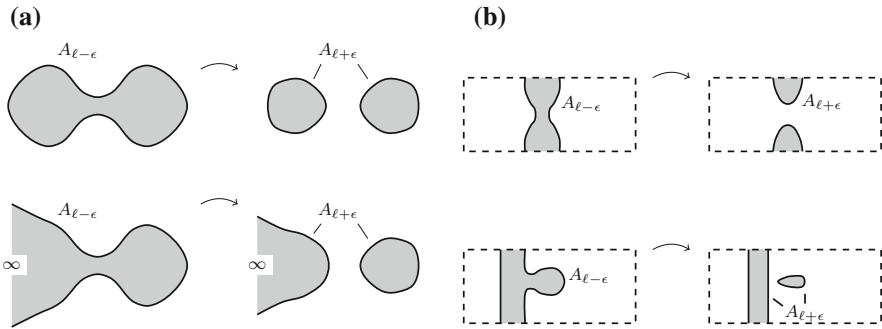

Fig. 7 Subfigures **a** and **b** give two examples of the excursion sets (shaded) at a level below and above a lower connected saddle point in $\mathbb{R}^2$ and $\mathbb{T}$ respectively (the dashed rectangles in **b** are identified with the torus by the standard quotient relation)

3 Critical point densities

In this section we prove Proposition 1.8. We begin by giving the definition for lower and upper connected saddle points in full generality. Let M be a manifold and let $A \subset M$. We say that A is *simple* if it is compact, connected and every loop in A (i.e. every continuous map $h : S^1 \to A$) is M-contractible. We will be solely interested in the case that A is a component of an excursion set of $f : M \to \mathbb{R}$. For such A, and in the case that M is simply connected (e.g. for $M = \mathbb{R}^2$ or $B(R)$), the condition of being simple is just the same as being bounded.

Definition 3.1 Let M be a 2-dimensional Riemannian manifold without boundary and let $f : M \to \mathbb{R}$ be a Morse function with a saddle point $x \in M$ such that $f(x) = \ell$. For $c \leq \ell$, let A_c denote the component of $\{f \geq c\}$ containing x and for $c > \ell$ let $A_c = A_\ell \cap \{f \geq c\}$. We say that x is *lower connected* if either of the following conditions hold for $\epsilon > 0$ sufficiently small:

1. $A_{\ell-\epsilon}$ is simple and $A_{\ell+\epsilon}$ consists of two simple components; or
2. $A_{\ell-\epsilon}$ is not simple but $A_{\ell+\epsilon}$ has a simple component.

We say that x is *upper connected* if it is a lower connected saddle for $-f$.

When f is an aperiodic Gaussian field, by Lemma 2.4 it will be Morse on $\mathbb{R}^2$ so we may take this as our Riemannian manifold and use the definition of lower/upper connected saddle points directly (we show in Sect. 4 that this definition coincides with that given in Sect. 1). If f is doubly periodic, recall that it is completely specified by its values on the associated parallelogram P which we identify with the torus. We then say that a saddle point of f is lower connected if it corresponds to a lower connected saddle of $f_\mathbb{T}$ by the definition above (see Fig. 8). Similarly if f is singly periodic we say that a saddle point of f is lower connected if it corresponds to a lower connected saddle point of $f_\mathcal{C}$. Upper connected saddle points are defined analogously.

The following lemma shows that our above definitions partition the set of saddle points in all cases of interest.
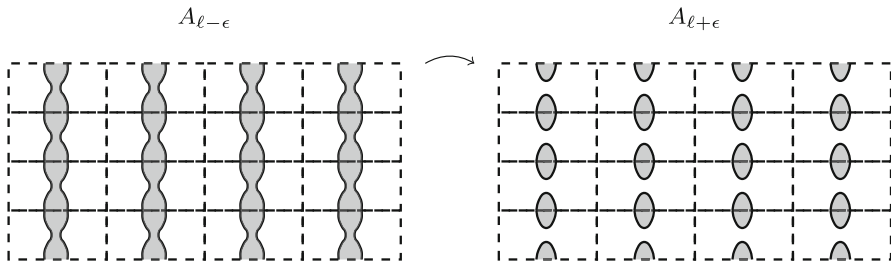
$A_{\ell-\epsilon}$ $A_{\ell+\epsilon}$

Fig. 8 Passing through the lower connected saddle points of a doubly periodic function increases the number of compact excursion sets

Lemma 3.2 *Let f be a stationary random field satisfying Conditions* 2.3. *Then with probability one all saddle points of f are either upper connected or lower connected but not both.*

This result is proven in Sect. 4. We are now ready to prove Proposition 1.8.

Proof (Proposition 1.8) Let f satisfy Conditions 1.1. We fix a compact $\Omega \subset \mathbb{R}^2$ such that $\partial\Omega$ has finite Hausdorff-1 measure and consider $N_h(\ell)$, the number of critical points of $f|_\Omega$ of type h with value greater than ℓ, where $h = m^-, s, m^+$ correspond to local minima, saddle points and local maxima respectively. If $(f(0), \nabla^2 f(0))$ has a non-degenerate distribution, then by a Kac–Rice formula (specifically, Corollary 11.2.2 of [1], which requires that $\partial\Omega$ has finite Hausdorff-1 measure)

$$
\begin{aligned}
\mathbb{E}(N_h(\ell)) &= \int_\Omega \mathbb{E}\left(\left|\det \nabla^2 f(t)\right| \mathbb{1}_{\text{Index}\nabla^2 f(t)=i,\, f(t)>\ell} \,\middle|\, \nabla f(t) = 0 \right) p_{\nabla f(t)}(0)\, dt \\
&= \text{Area}(\Omega) p_{\nabla f(0)}(0) \mathbb{E}\left(\left|\det \nabla^2 f(0)\right| \mathbb{1}_{\text{Index}\nabla^2 f(0)=i,\, f(0)>\ell} \right) \\
&= \text{Area}(\Omega) p_{\nabla f(0)}(0) \int_\ell^\infty \mathbb{E}\left(\left|\det \nabla^2 f(0)\right| \mathbb{1}_{\text{Index}\nabla^2 f(0)=i} \,\middle|\, f(0) = u \right) \phi(u)\, du
\end{aligned}
$$
(19)

where i is the index corresponding to h and ϕ denotes the standard Gaussian probability density. The third equality follows from the definition of conditioning on a Gaussian variable. If $(f(0), \nabla^2 f(0))$ has a degenerate distribution then $f(0)$ can be expressed as a linear combination of the elements of $\nabla^2 f(0)$ almost surely. Substituting in this expression for $f(0)$ allows us to apply Kac–Rice and the arguments above to derive (19) in this case too. Then, by Gaussian regression, $\mathbb{E}(|\det \nabla^2 f(0)|\mathbb{1}_{\text{Index}\nabla^2 f(0)=i} \mid f(0) = u)$ is a continuous function of u (see, e.g., Proposition 1.2 of [2]). This proves the existence and continuity of the densities p_{m^+}, p_{m^-} and p_s, and the fact that $p_{m^+}(x) = p_{m^-}(-x)$ follows from the symmetry of f.

Since $\mathbb{E}(N_s[0, \ell,]) = \int_0^\ell p_s(x)dx$, we know that $\mathbb{E}(N_s[0, \ell])$ is absolutely continuous in ℓ. It is also clear that for any $\ell_1 < \ell_2$, $\mathbb{E}(N_{s^+}[\ell_1, \ell_2]) \leq \mathbb{E}(N_s[\ell_1, \ell_2])$ so $\mathbb{E}(N_{s^+}[0, \ell])$ is absolutely continuous in ℓ. Therefore there exists a function $p_{s^+} : \mathbb{R} \to [0, \infty)$ such that

$$\mathbb{E}(N_{s+}[0, \ell]) = \int_0^\ell p_{s+}(x) \, dx.$$

Since $\mathbb{E}(N_s[0, \infty))$ is finite, the monotone convergence theorem shows that $p_{s+} \in L^1(\mathbb{R})$. By symmetry of the Gaussian distribution, and the definition of lower connected saddles, this also shows the existence of $p_{s-}(x) = p_{s+}(-x)$. The fact that $p_{s-} + p_{s+} = p_s$ follows from Lemma 3.2. □

4 Topological lemmas

In this section we prove the deterministic Lemma 2.5 using topological arguments. We also establish Lemma 3.2 and Proposition 1.20 using similar methods, thereby completing the proof of all results in the paper.

To prove Lemmas 2.5 and 3.2 we require several results from Morse theory which we now introduce. We note that several aspects of this theory require us to work with compact manifolds, and so in the aperiodic and singly periodic cases we work with $B(R)$ and $\mathcal{C}(n)$ rather than directly with $\mathbb{R}^2$ and $\mathcal{C}$; in the doubly periodic case, by contrast, we work directly with the torus $\mathbb{T}$. We also emphasise that we only ever work with the five examples of manifolds just mentioned, and so it is sufficient to have them in mind.

Recall that, for a Morse function f defined on a manifold M, the tangent points of f are defined as the critical points of $f|_{\partial M}$.

Theorem 4.1 (Theorem 7 of [8]) *Let M be a compact n-dimensional Riemannian manifold and let $f : M \to \mathbb{R}$ be a Morse function. If f has no critical or tangent points with value in $[a, b]$ then $\{f \geq b\}$ is homotopy equivalent to $\{f \geq a\}$.*

We define a k-cell to be a copy of the closed unit disc in $\mathbb{R}^k$ and temporarily denote this by B_k. If Y is a topological space then we define the following operation to be 'attaching a k-cell to Y'. First we find a continuous function $g : \partial B_k \to Y$, then we take the disjoint union $Y \sqcup B_k$ and identify each point in ∂B_k with its image under g. By attaching a 0-cell, we simply mean taking the disjoint union of Y and a single point.

Theorem 4.2 (Theorem 8 of [8]) *Let M be a compact 2-dimensional Riemannian manifold and let $f : M \to \mathbb{R}$ be a Morse function. If t is a critical point of f of index k with $f(t) = \ell$, then for $\epsilon > 0$ sufficiently small, $\{f \geq \ell - \epsilon\}$ is homotopy equivalent to $\{f \geq \ell + \epsilon\}$ with a $(2 - k)$-cell attached. If t is a tangent point and ℓ, ϵ are defined in the same way, then $\{f \geq \ell - \epsilon\}$ is homotopy equivalent to either $\{f \geq \ell + \epsilon\}$ or $\{f \geq \ell + \epsilon\}$ with a k-cell attached for some $k \in \{0, 1, 2\}$.*

These theorems are proven using methods very similar to those of the standard proofs for manifolds without boundary which can be found in almost any text on Morse theory.

We now work towards a proof of Lemma 2.5. First we will need another definition which, in the case of aperiodic or singly periodic fields, identifies a subset of the upper/lower connected saddle points which have unfavourable topological properties

(for our purposes). Recall that if M is a manifold then $A \subset M$ is said to be *simple* if A is compact, connected and every loop in A is M-contractible. In the case that $M = \mathbb{R}^2$ or $B(R)$ and A is an excursion set component, this is just the condition that A is bounded.

Definition 4.3 Let M_∞ be a Riemannian 2-manifold without boundary, with $f_\infty \in C^2_{\text{loc}}(M_\infty)$ a Morse function on M_∞, and let M be a compact submanifold of M_∞ with boundary, with $f := f_\infty|_M$ a Morse function on M. Let x_0 be a saddle point of f at level ℓ and let A_c be defined as in Definition 3.1 for f_∞. If x_0 is lower connected, then we say that it is *four-arm in M* if, for $\epsilon > 0$ sufficiently small, all of the simple components of $A_{\ell+\epsilon}$ intersect ∂M. Similarly, we say that an upper connected saddle is *four-arm in M* if it satisfies the previous condition for $-f$ (and $-\ell$).

We say that x_0 is an *infinite-four-arm saddle point* if $A_{\ell+\epsilon}$ has two unbounded components for all $\epsilon > 0$ sufficiently small.

Remark 4.4 In the proof of Lemma 3.2, we will show that when $M_\infty = \mathbb{R}^2$ (the most important case for our analysis), the level set at the level of an infinite-four-arm saddle takes the form shown in Fig. 9b. This corresponds to the way an 'infinite-four-arm event' is typically defined in the percolation literature: as two disjoint paths joining a point to infinity which are separated by two 'dual' paths joining the same point to infinity. For other choices of M_∞ (such as $M_\infty = \mathcal{C}$) the level set at the level of an infinite-four-arm saddle may look quite different, so that this terminology is less intuitive.

Let us explain the importance of four-arm saddles. If f is aperiodic, then its lower connected saddle points are defined so as to correspond to an increase in the number of excursion set components in $\mathbb{R}^2$. However if such a saddle point is four-arm in $\overline{B(R)}$, then this increase cannot be observed from inside $\overline{B(R)}$ since the excursion sets that are created when passing through the saddle intersect $\partial B(R)$ (see Fig. 9). The case when f is singly periodic is similar (in the case that f is doubly periodic we do not have to worry about such saddles). Infinite-four-arm saddle points will be relevant when we prove Lemma 3.2. Fortunately we can control the number of four-arm saddles which occur, in terms of the boundary behaviour of f.

Lemma 4.5 *Let M_∞ be a Riemannian 2-manifold without boundary and $f_\infty \in C^2_{loc}(M_\infty)$. Let M be a compact submanifold of M_∞ with boundary and $f := f_\infty|_M$ be Morse. Let $N_{4\text{-arm}}(M)$ be the number of saddle points of f_∞ contained in M which are four-arm in M or infinite-four-arm. Then*

$$N_{4\text{-arm}}(M) \leq 3N_{tang}(\partial M).$$

Intuitively this bound follows because as we raise the level past a saddle which is four-arm in $\overline{B(R)}$ or an infinite-four-arm saddle we separate two components of $\{f \geq \ell\} \cap \partial B(R)$ (i.e. the boundary components are no longer in the same excursion set component). The total number of such separations which may occur is bounded above by the number of boundary excursion components at different levels, which in turn is bounded by the number of tangent points. This is formalised in the proof below.
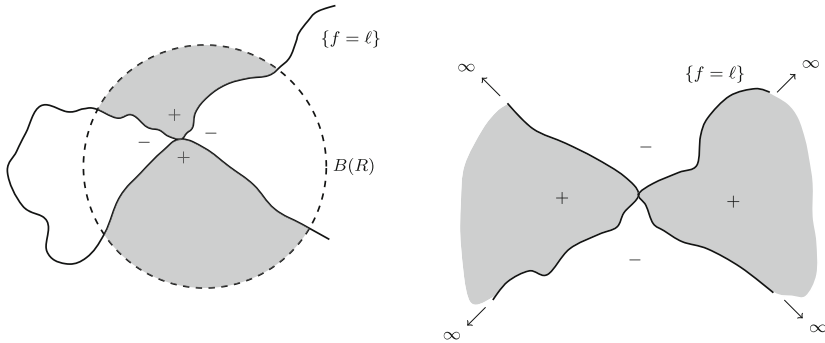
(a) An upper connected saddle that is four-arm in $\overline{B(R)}$.   **(b)** An infinite-four-arm saddle point in $\mathbb{R}^2$.

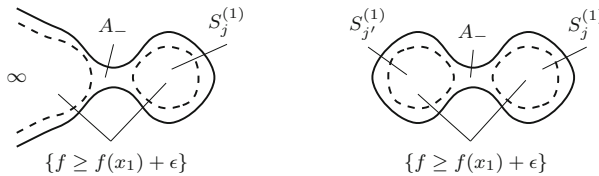**Fig. 9** Different types of four-arm saddle points

Fig. 10 Since x_1 is a lower connected saddle point, A_- may contain either one (left) or two (right) simple component of $\{f \geq f(x_1) + \epsilon\}$

Proof We describe an algorithm which will create an injective mapping from the lower connected saddle points which are four-arm in M to the tangent points of f in ∂M. Let $x_1, \ldots, x_m$ denote all such saddle points arranged in order of increasing level, i.e. $f(x_i) < f(x_{i+1})$. We fix $\epsilon > 0$ such that there are no tangent points with value in $[f(x_i) - \epsilon, f(x_i) + \epsilon]$ for any i and these intervals are disjoint for different i. For each $i = 1, \ldots, m$, we define $S_1^{(i)}, \ldots, S_{m_i}^{(i)}$ to be the simple components of $\{f \geq f(x_i) + \epsilon\}$ which intersect ∂M.

We proceed by induction. Let A_- denote the component of $\{f \geq f(x_1) - \epsilon\}$ containing x_1. Since x is lower connected and four-arm in M, A_- will contain some $S_j^{(1)}$ (it may contain two such elements: see Fig. 10), and so we make a preliminary association between x_1 and $S_j^{(1)}$.

We now assume inductively that for each $x_1, \ldots, x_n$ we have either a preliminary association with some $S_j^{(n)}$ (which differs across points) or a final association to some tangent point with value less than $f(x_n)$.

We consider x_{n+1} and let A_- denote the component of $\{f \geq f(x_{n+1}) - \epsilon\}$ containing x_{n+1}. First suppose that A_- is contained in some $S_j^{(n)}$ which has a preliminary association with some x_i (where $i \leq n$). Then A_- must be simple, as it is a subset of $S_j^{(n)}$, and so, by the definition of a lower connected four-arm saddle point, this means that A_- must contain two different components $S_p^{(n+1)}$ and $S_q^{(n+1)}$. We then make a preliminary association between x_{n+1} and $S_p^{(n+1)}$ and make a new preliminary

association between x_i and $S_q^{(n+1)}$ (so we remove the old association between x_i and $S_j^{(n)}$). If A_- is contained in some $S_j^{(n)}$ which does not have a preliminary association or if A_- is not contained in any $S_j^{(n)}$ then we choose some $S_p^{(n+1)}$ contained in A_- (which must exist since x_{n+1} is four-arm in M) and make a preliminary association between this and x_{n+1}.

Now suppose that x_i, where $i \leq n$, has a preliminary association with $S_j^{(n)}$. If the maximum level of any point in $S_j^{(n)}$ is less than $f(x_{n+1})$ then we make a final association between x_i and the highest such point in $\partial M \cap S_j^{(n)}$, which is by definition a tangent point. We also remove the preliminary association between x_i and $S_j^{(n)}$. Otherwise $S_j^{(n)}$ must be a superset of $S_k^{(n+1)}$ for some k, and we then make a new preliminary association between x_i and $S_k^{(n+1)}$. We perform this process for all preliminary associations which remain after the step described in the previous paragraph, which completes the inductive step. This algorithm will cease after the m-th step, at which stage if a point x_i has a preliminary association with some $S_j^{(n+1)}$ we make a final association between x_i and the highest tangent point of $S_j^{(n+1)}$.

The end result of this algorithm is an injective mapping, defined by the final associations, from the set of lower connected saddles which are four-arm in M to the tangent points of f. An identical argument applied to $-f$ creates such a map for the the upper connected saddles which are four-arm in M. A very similar algorithm can be applied to the infinite-four-arm saddles of f, except that the inductive step is slightly simpler. Combining these results proves the lemma. ☐

Our next result is a preliminary version of Lemma 2.5 for Morse functions on manifolds without boundary. This setting allows us to prove the desired relationship between the number of excursion set components and critical points in the simpler case that all critical points have different levels.

Lemma 4.6 *Let M_∞ be a Riemannian 2-manifold without boundary and $f \in C_{loc}^2(M_\infty)$ be a Morse function. Let M be a compact submanifold of M_∞ and $f := f_\infty|_M$ be Morse. For $\ell \in \mathbb{R}$, let $N_{simple}(\ell)$ be the number of simple components of $\{f \geq \ell\}$ which do not intersect the boundary of M and let $N_{m^+}(\ell)$ and $N_{s^-}(\ell)$, be the number of local maxima and lower connected saddle points respectively of f with level in $[\ell, \infty]$. Then*

$$N_{simple}(\ell) = N_{m^+}(\ell) - N_{s^-}(\ell) + \zeta$$

where $|\zeta| \leq 5N_{tangent}$. So in particular, if M has no boundary then $\zeta = 0$.

The proof involves applying the Morse theorems to each critical/tangent point to derive the change in the number of simple components at the corresponding levels and then summing these changes.

Proof First suppose that f has no critical or tangent points with level in $[a, b]$. By Theorem 4.1, $\{f \geq b\}$ is a deformation retract of $\{f \geq a\}$ under some map $h : \{f \geq$

$a\} \times [0, 1] \to \{f \geq a\}$. In particular, this is also true for each component of $\{f \geq b\}$ and the component of $\{f \geq a\}$ in which it is contained. Since h is a homotopy, the number of simple components is the same in each set. If A is a component of $\{f \geq a\}$ which intersects ∂M, then we claim that $A \cap \{f \geq b\}$ also intersects ∂M. To see why, suppose A intersects ∂M but $A \cap \{f \geq b\}$ does not. Then by considering the infimum ℓ^* of $\ell \in [a, b]$ such that $A \cap \{f \geq \ell\}$ does not intersect ∂M and taking a sequence of points in $A \cap \{f \geq \ell^* - 1/n\} \cap \partial M$ we see that $A \cap \{f \geq \ell^*\}$ contains a tangent point at level $\ell^* \in [a, b]$, which is a contradiction. Combining all these observations, we see that N_{simple} is constant on intervals which contain no critical or tangent points.

Now let x be a critical or tangent point of f at level c and take $\epsilon > 0$ small enough to apply Lemma 4.2. We consider in turn the different types of critical or tangent point and calculate $N_{\text{simple}}(c - \epsilon) - N_{\text{simple}}(c + \epsilon)$. Let $A_{c-\epsilon}$ denote the component of $\{f \geq c - \epsilon\}$ containing x and $A_{c+\epsilon} = A_{c-\epsilon} \cap \{f \geq \ell + \epsilon\}$. We note that to determine $N_{\text{simple}}(c - \epsilon) - N_{\text{simple}}(c + \epsilon)$ it is enough to consider $A_{c-\epsilon}$ and $A_{c+\epsilon}$, since by the arguments in the previous paragraph, the number of simple components of $\{f \geq \ell\}$ in $M \setminus A_{c-\epsilon}$ which do not intersect ∂M will be constant as ℓ varies in $[c - \epsilon, c + \epsilon]$.

If x is a local maximum, then by Lemma 4.2, $A_{c-\epsilon}$ is homotopy equivalent to $A_{c+\epsilon}$ with a 0-cell attached. So in particular $\{f \geq c - \epsilon\}$ has one more component than $\{f \geq c + \epsilon\}$, and this extra component is M-contractible. Since the extra component contains a local maximum, (which cannot be in ∂M) for $\epsilon > 0$ sufficiently small, this component is disjoint from ∂M and so $N_{\text{simple}}(c - \epsilon) - N_{\text{simple}}(c + \epsilon) = 1$.

If x is a local minimum, then $A_{c-\epsilon}$ is homotopy equivalent to $A_{c+\epsilon}$ with a 2-cell attached. Attaching a 2-cell does not change the number of components of $\{f \geq c+\epsilon\}$, and does not affect whether the component it is attached to is simple or not. Once again since the local minimum is not in ∂M, restricting ϵ sufficiently small ensures that the attached 2-cell does not intersect ∂M and so $N_{\text{simple}}(c - \epsilon) - N_{\text{simple}}(c + \epsilon) = 0$.

Next we suppose that x is a saddle point, so that by Lemma 4.2, $A_{c-\epsilon}$ is homotopy equivalent to $A_{c+\epsilon}$ with a 1-cell attached. So in particular, $A_{c+\epsilon}$ consists of either one or two components. First we note that if $A_{c-\epsilon}$ is simple and does not intersect ∂M, then any components of $A_{c+\epsilon}$ have both these properties, therefore $N_{\text{simple}}(c - \epsilon) - N_{\text{simple}}(c+\epsilon) \leq 0$. If $A_{c-\epsilon}$ is simple but intersects ∂M, then by repeating the sequential argument in the first paragraph of this proof we see that $A_{c+\epsilon}$ must intersect ∂M and so cannot consist of two simple components disjoint from the boundary. Furthermore, if $A_{c-\epsilon}$ is not simple, then by the above homotopy, $A_{c+\epsilon}$ cannot consist of two simple components. Combining these two observations show that $N_{\text{simple}}(c-\epsilon) - N_{\text{simple}}(c+\epsilon) \geq -1$. So the number of simple components disjoint from ∂M is either constant or increases by one on passing through the saddle point. If x is not a lower connected saddle, then by definition $N_{\text{simple}}(c-\epsilon) - N_{\text{simple}}(c+\epsilon) = 0$, if x is a lower connected saddle which is not four-arm in M, then $N_{\text{simple}}(c-\epsilon) - N_{\text{simple}}(c+\epsilon) = -1$ and if x is a lower connected saddle which is four-arm in M, then $N_{\text{simple}}(c-\epsilon) - N_{\text{simple}}(c+\epsilon) = 0$.

Finally let x be a tangent point, so that $A_{c-\epsilon}$ is homotopy equivalent to $A_{c+\epsilon}$ with a k-cell attached for some $k \in \{0, 1, 2\}$. In each case, $A_{c+\epsilon}$ has at most two simple components disjoint from ∂M, so $|N_{\text{simple}}(c - \epsilon) - N_{\text{simple}}(c + \epsilon)| \leq 2$.

Now suppose that there are no critical or tangent points at level ℓ. Since $N_{\text{simple}}(a) \to 0$ as $a \to \infty$, we see that $N_{\text{simple}}(\ell)$ equals the sum of the finite

number of jumps $N_{\text{simple}}(c-\epsilon) - N_{\text{simple}}(c+\epsilon)$ at each level c with a critical/tangent point. This sum equals the number of local maxima of f above level ℓ minus the corresponding number of lower connected saddles with an error ζ bounded in absolute value by $N_{4\text{-arm}}(M) + 2N_{\text{tang}}(M)$. By Lemma 4.5, $|\zeta| \leq 5N_{\text{tang}}(M)$.

Finally, we note that $\{f \geq \ell\} = \cap_{\epsilon > 0}\{f \geq \ell - \epsilon\}$ and that the intersection of a decreasing family of compact, connected sets is connected, so that $\{f \geq \ell\}$ and $\{f \geq \ell - \epsilon\}$ have the same number of components for ϵ sufficiently small. Repeating the earlier argument for tangent points shows that these sets have the same number of components which do not intersect ∂M. Then, if f has a critical point x at level ℓ, applying the Morse lemma on a neighbourhood of x and Theorem 4.1 outside this neighbourhood shows that $N_{\text{simple}}(\ell - \epsilon) = N_{\text{simple}}(\ell)$. (If no such point exists, we simply apply Theorem 4.1 to M.) Therefore we can apply the above arguments to the level $\ell - \epsilon$ such that there are no critical or tangent points with level in $[\ell - \epsilon, \ell)$ to prove the lemma. $\qquad\square$

With the above preliminary lemma, the proof of Lemma 2.5 in the case of aperiodic functions is straightforward.

Proof (Lemma 2.5 in the case of aperiodic functions) Let f be an aperiodic function satisfying the assumptions of Lemma 2.5. Applying Lemma 4.6 with $M = \overline{B(R)}$ and $M_\infty = \mathbb{R}^2$ gives exactly the stated relationship for excursion sets. To complete the proof, we prove a corresponding relationship for level sets. We fix a level ℓ and let $f_R = f|_{\overline{B(R)}}$, we construct a graph on the vertex set

$$V := \{\text{Components of } \{f_R \geq \ell\}\} \cup \{\text{Components of } \{f_R \leq \ell\}\}$$

by declaring two vertices to be joined by an edge if they have non-empty intersection. Clearly the graph is bipartite and each edge corresponds to a component of $\{f = \ell\}$. This graph is acyclic, and so by Euler's formula

$$\begin{aligned}\#\{\text{Components of } \{f_R = \ell\}\} = &\#\{\text{Components of } \{f_R \geq \ell\}\} \\ &+ \#\{\text{Components of } \{f_R \leq \ell\}\} - 1\end{aligned}$$

The number of components of $\{f_R = \ell\}$ which intersect $\partial B(R)$ is bounded above by the number of tangent points of f in $\partial B(R)$, and the same bound holds for the components of $\{f_R \geq \ell\}$ and $\{f_R \leq \ell\}$. Therefore we can express the equation above as

$$N_{LS,R}(f, \ell) = N_{ES,R}(f, \ell) + N_{ES,R}(-f, -\ell) + \eta^{(1)}(\ell)$$

where $|\eta^{(1)}(\ell)| \leq 4N_{\text{tang}}(\partial B(R))$. Applying the first part of this lemma to each of the $N_{ES,R}$ terms here then completes the proof. $\qquad\square$

For the periodic cases, the argument is a little more technical since we cannot apply Lemma 4.6 to $B(R)$ directly. Instead, we tile $B(R)$ with translated parallelograms or rectangles, apply Lemma 4.6 to $f_{\mathbb{T}}$ or $f_{\mathcal{C}(n)}$ on each translated domain and aggregate the results.

Proof (Lemma 2.5 **in the case of periodic functions)** Let f be a doubly or singly periodic function satisfying the assumptions of Lemma 2.5. It suffices to prove the excursion set relationship, since the level set relationship will follow by the same argument as in the aperiodic case.

Suppose f is doubly periodic with periodic vectors $y, z \in \mathbb{R}^2$ and associated parallelogram P. By assumption, $f_{\mathbb{T}}$ is a Morse function almost surely. Suppose that $\{f_{\mathbb{T}} \geq \ell\}$ has N_{simple} simple components.

Let A be a component of $\{f \geq \ell\}$ and A' be the corresponding component of $\{f_{\mathbb{T}} \geq \ell\}$. If A is compact, then clearly A' cannot contain a non-$\mathbb{T}$-contractible loop, and so it is simple. Since $f|_{\partial P}$ has a finite number of local extrema, $\{f \geq \ell\} \cap \partial P$ has a finite number of components. Therefore if A intersects the boundary of a translated parallelogram $P + n_1 y + n_2 z$ where $n_1, n_2 \in \mathbb{Z}$ then it must contain the translation of one of these boundary components. So if A is unbounded, it must contain a path joining two translated versions of the same boundary component, which implies that A' is not simple.

We can tile $\mathbb{R}^2$ with the translated parallelograms $P + n_1 y + n_2 z$ where $n_1, n_2 \in \mathbb{Z}$. We associate each bounded component A of $\{f \geq \ell\}$ to a particular translation $P + n_1 y + n_2 z$ of P such that $(P + n_1 y + n_2 z) \cap A \neq \emptyset$, N_{simple} components are mapped to each translated parallelogram and if A is associated with P, then $A + n_1 y + n_2 z$ is associated with $P + n_1 y + n_2 z$. If A is a component of $\{f \geq \ell\}$ contained in $B(R)$ then it must be associated with a parallelogram which intersects $B(R)$, therefore

$$N_{ES,R}(\ell) \leq N_{\text{simple}} \cdot \#\{\text{Translations of } P \text{ in } B(R+d)\} \qquad (20)$$

where $d = \text{diam}(P)$. Similarly, each compact component of $\{f \geq \ell\}$ which is associated with a parallelogram inside $B(R)$ must be in $B(R)$ unless it intersects $\partial B(R)$, but the number of such intersections is bounded above by the number of tangent points of f to $B(R)$. Therefore

$$N_{\text{simple}} \cdot \#\{\text{Translations of } P \text{ in } B(R)\} - N_{\text{tang}}(\partial B(R)) \leq N_{ES,R}(\ell) \qquad (21)$$

The number of translated parallelograms contained in the ball of a given radius can be approximated by a generalisation of Gauss' circle problem. It is shown in [11] that

$$\#\{\text{Translations of } P \text{ in } B(R)\} = \frac{\pi R^2}{\text{Area}(P)} + o(R)$$

as $R \to \infty$. So combining this with (20) and (21)

$$N_{ES}(R, \ell) = N_{\text{simple}} \cdot \frac{\pi R^2}{\text{Area}(P)} + O(R + N_{\text{tang}}(\partial B(R)))$$

as $R \to \infty$.

Now suppose that $\{f_{\mathbb{T}} \geq \ell\}$ contains m local maxima. Reasoning as above, it is clear that

$$m \cdot \#\{\text{Translations of } P \text{ in } B(R)\} \leq N_{m^+, R}[\ell, \infty)$$
$$\leq m \cdot \#\{\text{Translations of } P \text{in } B(R+d)\}$$

and so

$$N_{m^+, R}[\ell, \infty) = m \cdot \frac{\pi R^2}{\text{Area}(p)} + O(R)$$

with an analogous result holding for lower connected saddles. Applying Lemma 4.6 to $f_{\mathbb{T}}$ (i.e. with the setting $M_{\infty} = M = \mathbb{T}$) then shows that

$$N_{ES, R}(\ell) = N_{m^+, R}[\ell, \infty) - N_{s^-, R}[\ell, \infty)) + O(R + N_{\text{tang}}(\partial B(R)))$$

as $R \to \infty$ where the constant in the $O(\cdot)$ notation is independent of ℓ, as required (although it may depend on d).

Now suppose that f is singly periodic with periodic vector $y = (y_1, 0)$ and define $\mathcal{R}(n) = [ny_1, (n+1)y_1) \times \mathbb{R}$. By assumption f almost surely has a finite number of tangent points in $[0, y_1] \times \{n_1\}$ or $\{n_2 y_1\} \times [-R, R]$ for any $n_1, n_2 \in \mathbb{Z}$ or $R > 0$. Repeating the argument from the doubly periodic case then shows that the compact components of $\{f \geq \ell\}$ correspond precisely to the simple components of $\{f_C \geq \ell\}$.

As before, we can associate each compact component of $\{f \geq \ell\}$ with a translated rectangle $\mathcal{R}(n)$ for $n \in \mathbb{Z}$, in such a way that if A is a compact component of $\{f \geq \ell\}$ which is mapped to $\mathcal{R}(n)$ then A intersects $\mathcal{R}(n)$, and if $A' = A + y$ then A' maps to $\mathcal{R}(n+1)$.

We recall our earlier definitions: if $S(n_1, n_2) = [n_1 y_1, (n_1 + 1)y_1] \times [n_2, n_2 + 1]$ where $n_1, n_2 \in \mathbb{Z}$, then we define $B_{\text{int}}(x, R)$ to be the union over $n_1, n_2 \in \mathbb{Z}$ of all $S(n_1, n_2)$ contained in $B(x, R)$. Similarly we define $B_{\text{ext}}(x, R)$ to be the union over $n_1, n_2 \in \mathbb{Z}$ of $S(n_1, n_2)$ which intersect $B(x, R)$. We now fix $R > 0$ and define m_n for $n \in \mathbb{Z}$ to be the largest integer such that

$$[ny_1, (n+1)y_1] \times [-m_n, m_n] \subset B(R)$$

and note that under the standard quotient relationship, this rectangle can be identified with $C(m_n)$.

Let $N_{\text{simple}}(f_{C(m_n)}, \ell)$ denote the number of simple components of $\{f|_{C(m_n)} \geq \ell\}$ which do not intersect $\partial C(m_n)$. By the mapping described above, there will be at least $N_{\text{simple}}(f_{C(m_n)}, \ell)$ compact components of $\{f \geq \ell\}$ which are associated with $\mathcal{R}(n)$ and intersect $[ny_1, (n+1)y_1] \times [-m_n, m_n]$. If any of these components intersect $\partial B_{\text{int}}(0, R)$, then they will do so at distinct components of $\{f|_{\partial B_{\text{int}}(0,R)} \geq \ell\}$ and so in particular the number of such components intersecting the boundary is at most the number of tangent points on $\partial B_{\text{int}}(0, R)$. Therefore

$$N_{ES}(R, \ell) \geq \sum_n N_{\text{simple}}(f_{\mathcal{C}(m_n)}, \ell) - N_{\text{tang}}(\partial B_{\text{int}}(0, R))$$

We can then apply Lemma 4.6 to $f_{\mathcal{C}(m_n)}$ (i.e. with the setting $M_\infty = \mathcal{C}$ and $M = \mathcal{C}(m_n)$) for each n to get this inequality in terms of local maxima and lower connected saddles points. Let $r = \max\{\sqrt{2}, y_1\}$. Then since $B_{\text{int}}(0, R)$ covers $B(R - r)$ we see that

$$N_{ES}(R, \ell) \geq N_{m^+}(R, \ell) - N_{s^-}(R, \ell) - 6N_{\text{tang}}(\partial B_{\text{int}}(R)) - N_{\text{crit}}(B(R) \backslash B(R - r)).$$

This gives the required lower bound for $N_{ES}(R, \ell)$. The upper bound follows from a very similar argument. First we define M_n to be the largest integer such that

$$[ny_1, (n + 1)y_1] \times [M_n - 1, M_n] \subset B_{\text{ext}}(0, R)$$

so that $B(R)$ is covered by the finite union of rectangles $[ny_1, (n+1)y_1] \times [-M_n, M_n]$. Then

$$N_{ES}(R, \ell) \leq \sum_n N_{\text{simple}}(f_{\mathcal{C}(M_n)}, \ell)$$

and by applying Lemma 4.6 to each cylinder we see that

$$N_{ES}(R, \ell) \leq N_{m^+}(R, \ell) - N_{s^-}(R, \ell) + 5N_{\text{tang}}(\partial B_{\text{ext}}(R)) + N_{\text{crit}}(B(R + r) \backslash B(R))$$

as required.   □

We next complete the proof of Lemma 3.2. Again we shall separate the argument into the aperiodic case and the periodic cases, so that the reader interested only in the aperiodic case can access the simplest version of the argument. Since in this case we work only with the manifolds $\mathbb{R}^2$ and $B(R)$, we replace the condition that an excursion set component A is simple with the equivalent condition that it is bounded.

Proof (Lemma 3.2 ***in the case of aperiodic fields)*** Let f be an aperiodic stationary field satisfying Conditions 2.3. Let $g : \mathbb{R}^2 \to \mathbb{R}$ be a Morse realisation of f and assume that $g|_{\overline{B(n)}}$ is Morse for all $n \in \mathbb{N}$. Let x be a lower connected saddle of g at level ℓ.

The first step is to show that, with probability one, x is not also an upper connected saddle of g. For $\epsilon > 0$ sufficiently small, let $A_{\ell-\epsilon}$ be the component of $\{g \geq \ell - \epsilon\}$ containing x and $A_{\ell+\epsilon} = A_{\ell-\epsilon} \cap \{g \geq \ell + \epsilon\}$. From the definition of a lower connected saddle point, we have three cases to consider. In the first two cases we use the fact that g is Morse to (deterministically) rule out the possibility of x being upper connected; we show that the third case occurs with zero probability and can therefore be neglected.

1)$A_{\ell-\epsilon}$ **is unbounded and** $A_{\ell+\epsilon}$ **has one bounded and one unbounded component.** Let S denote the bounded component of $A_{\ell+\epsilon}$. We start by choosing n sufficiently large that $\overline{B(n)}$ contains a neighbourhood of S and a neighbourhood of x. We let $A'_{\ell-\epsilon}$ be the component of $\{g|_{\overline{B(n)}} \geq \ell - \epsilon\}$ containing x and $A'_{\ell+\epsilon}$ denote $A'_{\ell-\epsilon} \cap \{g \geq \ell + \epsilon\}$. We can apply Theorem 4.2 to $-g$ to deduce that $\overline{B(n)} \cap \{g \leq \ell + \epsilon\}$ is homotopy

equivalent to $\overline{B(n)} \cap \{g \leq \ell - \epsilon\}$ with a 1-cell, denoted γ, attached. By reducing ϵ we can ensure that $A'_{\ell+\epsilon}$ has two components, one of which is bounded, and so both end-points of γ must be contained in the same component of $\overline{B(n)} \cap \{g \leq \ell - \epsilon\}$. This will in turn be contained in a component of $\{g \leq \ell - \epsilon\}$ which we denote by B. Clearly $B \cup \gamma$ is bounded if and only if B is, therefore $\overline{B(n)} \cap \{g \leq \ell - \epsilon\}$ and $\overline{B(n)} \cap \{g \leq \ell + \epsilon\}$ have the same number of bounded components. This holds for any n sufficiently large (possibly after reducing ϵ) but if x were upper connected we could find m large enough that $\overline{B(m)} \cap \{g \leq \ell - \epsilon\}$ has one more bounded component than $\overline{B(m)} \cap \{g \leq \ell + \epsilon\}$ (this argument is stated more formally in the proof of Lemma 4.6). Therefore x is not an upper connected saddle point.

2) $A_{\ell-\epsilon}$ is bounded and $A_{\ell+\epsilon}$ has two bounded components. The arguments in the previous case are also valid in this case where S is chosen to be either of the two components of $A_{\ell+\epsilon}$.

3) $A_{\ell-\epsilon}$ is unbounded and $A_{\ell+\epsilon}$ is bounded. We will show that when f is aperiodic, it almost surely has no saddle points of this type. We fix $R > 0$ and let $g_R = g|_{B(R)}$ which we assume is Morse. For $y \in \mathbb{R}^2$ we define $A_{y,-\epsilon}$ to be the component of $\{g \geq g(y) - \epsilon\}$ containing y and $A_{y,\epsilon} = A_{y,-\epsilon} \cap \{g \geq g(y) + \epsilon\}$. Let $x_1, \ldots, x_n$ be the saddle points of g_R for which $A_{x_i,-\epsilon}$ is unbounded but $A_{x_i,\epsilon}$ is bounded for all $\epsilon > 0$ sufficiently small. We then choose a fixed ϵ sufficiently small that this condition holds for each i and such that g_R has no other critical points and no tangent points in $\{g_R(x_i) - 3\epsilon \leq g_R \leq g_R(x_i) + 3\epsilon\}$ for any i. This ensures the intervals $[g(x_i) - \epsilon, g(x_i) + \epsilon]$ are non-overlapping for $i = 1, \ldots, n$. Finally we assume that the x_i are ordered so that $g(x_i) < g(x_j)$ for $i < j$.

Since $A_{x_i,-\epsilon}$ is unbounded for each i, we see that

$$B_i^- := A_{x_i,-\epsilon} \cap \partial B(R)$$

is non-empty for each i. Since there are no tangent points with level in $[g_R(x_i) - \epsilon, g_R(x_i) + \epsilon]$, we can repeat the arguments in the proof of Lemma 4.6 to show that

$$B_i^+ := A_{x_i,\epsilon} \cap \partial B(R)$$

is non-empty and has the same number of components as B_i^- for each $i = 1, \ldots, n$ (see Fig. 11). Suppose there exists a point $y \in B_i^- \cap B_j^-$ for some $i < j$. Since $A_{x_j,-\epsilon}$ is unbounded and path connected, there exists an unbounded path started at y and contained in $\{g \geq g(x_j) - \epsilon)\}$. Since $y \in A_{x_i,\epsilon}$ which is path connected, this path must also be contained in $A_{x_i,\epsilon}$ which is bounded. This contradiction implies that $B_1^-, \ldots, B_n^-$ are disjoint. Since each of these sets must have a local maximum, we see that $n \leq N_{\text{tang}}(\partial B(R))$. Since f is stationary, we know that the expected number of lower connected saddle points x contained in $B(R)$ for which $A_{x,-\epsilon}$ is unbounded and $A_{x,\epsilon}$ is bounded for $\epsilon > 0$ sufficiently small is $c_0 R^2$ for some $c_0 \geq 0$. However by the argument just given, this number is bounded above by the expected number of tangent points of f to $\partial B(R)$ which is $O(R)$ by Conditions 2.3. Hence $c_0 = 0$ so f almost surely has no saddle points of this type.
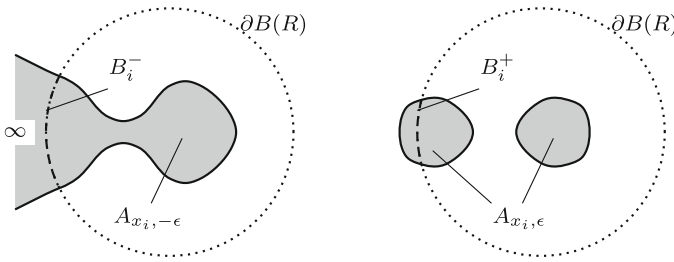
Fig. 11 We show that almost surely an unbounded excursion set component of f cannot shrink into a bounded excursion set by passing through a saddle point
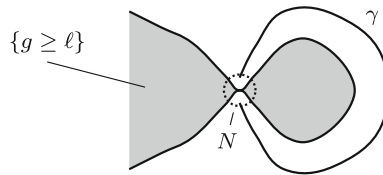

Fig. 12 The existence of the path γ implies that x is a lower connected saddle point

We have shown that the set of lower connected saddles of f is almost surely disjoint from the set of upper connected saddles. Using the same notation as above, we will now show that any saddle point of g must be lower connected or upper connected, completing the proof of the lemma.

For a small neighbourhood N of x, $\{g > \ell\} \cap N$ and $\{g < \ell\} \cap N$ each have two components. Since g is Morse, it has no other critical points at level ℓ and so the level set $\{g = \ell\}$ consists of non-intersecting curves except for the component which contains x. Since g has no infinite-four-arm saddles there are two mutually exclusive possibilities to consider; either there exists a path in $\{g < \ell\}$ joining the two components of $\{g < \ell\} \cap N$ or there exists a path in $\{g > \ell\}$ joining the two components of $\{g > \ell\} \cap N$.

We now show that in the first case x will be lower connected. By symmetry, this will imply that in the second case, x is upper connected and so complete the proof for aperiodic fields. We fix a path γ in $\{g < \ell\}$ connecting the components of $\{g < \ell\} \cap N$. From the existence of γ it is clear that $A_\ell \backslash \{x\}$ has two components B_1, B_2 and that without loss of generality, B_1 is bounded (see Fig. 12).

For $\epsilon > 0$ sufficiently small, $A_{\ell+\epsilon}$ therefore has at least one bounded component $B_1 \cap \{g \geq \ell+\epsilon\}$. If $A_{\ell-\epsilon}$ is compact for some $\epsilon > 0$ then clearly B_2 is also bounded and so $A_{\ell+\epsilon}$ has two bounded components so x is lower connected. If $A_{\ell-\epsilon}$ is unbounded for arbitrarily small ϵ we can assume that $A_{\ell+\epsilon}$ has an unbounded component, by the argument given above, and so x is again lower connected. We note that this also proves the equivalence of Definition 3.1 in the aperiodic case and Definition 1.7. □

Proof (Lemma 3.2 in the case of periodic fields) The arguments here are very similar to those in the aperiodic case (i.e. replacing the condition of boundedness with the more general condition of simplicity). Let f be a stationary field satisfying Conditions 2.3. Let $g : M \to \mathbb{R}$ be a Morse realisation of f restricted to M, where M is one of $\mathbb{T}$ or
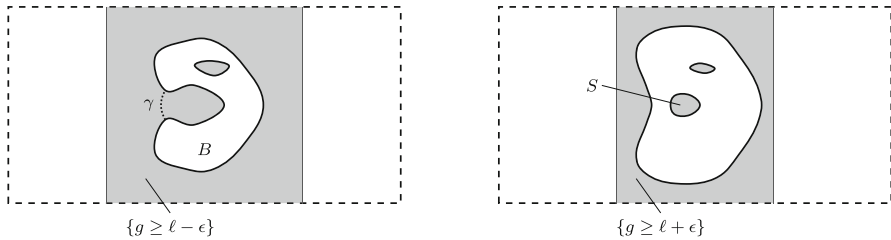
Fig. 13 An example of the excursion sets at a level below (left) and above (right) a saddle point at level ℓ for which $A_{\ell-\epsilon}$ is not simple and $A_{\ell+\epsilon}$ has one simple and one non-simple component when $M = \mathbb{T}$

$\mathcal{C}$ depending on whether f is doubly periodic or singly periodic. We also assume that $g|_{\mathcal{C}(n)}$ is Morse for all $n \in \mathbb{N}$ in the singly periodic case. Let x be a lower connected saddle of g at level ℓ.

Again the first step is to show that, with probability one, x is not an upper connected saddle of g. For $\epsilon > 0$ sufficiently small, let $A_{\ell-\epsilon}$ be the component of $\{g \geq \ell - \epsilon\}$ containing x and $A_{\ell+\epsilon} = A_{\ell-\epsilon} \cap \{g \geq \ell + \epsilon\}$. Again we have three cases to consider, the first two of which are proven deterministically:

1) $A_{\ell-\epsilon}$ **is not simple and** $A_{\ell+\epsilon}$ **has one simple and one non-simple component.** Let S denote the simple component of $A_{\ell+\epsilon}$. If f is singly periodic, we set $M_c = \mathcal{C}(n)$ where n is sufficiently large that M_c contains a neighbourhood of S and a neighbourhood of x. If f is doubly periodic we take $M_c = M = \mathbb{T}$. We define $A'_{\ell-\epsilon}$ to be the component of $\{g|_{M_c} \geq \ell - \epsilon\}$ containing x and $A'_{\ell+\epsilon} := A'_{\ell-\epsilon} \cap \{g|_{M_c} \geq \ell + \epsilon\}$. By Theorem 4.2 applied to $-g$ we deduce that $M_c \cap \{g \leq \ell + \epsilon\}$ is homotopy equivalent to $M_c \cap \{g \leq \ell - \epsilon\}$ with a 1-cell, denoted γ, attached. By reducing ϵ we can ensure that $A'_{\ell+\epsilon}$ has two components, one of which is simple, and so both end-points of γ must be contained in the same component of $M_c \cap \{g \leq \ell - \epsilon\}$. This will then be contained in a component of $\{g \leq \ell - \epsilon\}$ which we denote by B. If B is not simple, then clearly $B \cup \gamma$ is not simple. Suppose that B is simple but $B \cup \gamma$ is not simple, then $B \cup \gamma$ must contain a loop, denoted η, which is not M-contractible and in particular η must intersect γ. However B is path connected and $B \cup \gamma$ 'surrounds' a region which is homotopy equivalent to S and so must be simple (see Fig. 13 for an example of these sets in the doubly periodic case).

Therefore η is homotopy equivalent to a loop which is contained in B and so B is not simple, which is a contradiction. We have shown that $B \cup \gamma$ is simple if and only if B is, therefore $M_c \cap \{g \leq \ell - \epsilon\}$ and $M_c \cap \{g \leq \ell + \epsilon\}$ have the same number of simple components. This holds for any M_c sufficiently large (possibly after reducing ϵ) but if x were upper connected we could find M_c such that $M_c \cap \{g \leq \ell - \epsilon\}$ has one more simple component than $M_c \cap \{g \leq \ell + \epsilon\}$ (again, we note that this argument is stated more formally in the proof of Lemma 4.6). We conclude that x is not an upper connected saddle point.

2) $A_{\ell-\epsilon}$ **is simple and** $A_{\ell+\epsilon}$ **has two simple components.** Once again, the arguments in the previous case are also valid in this case where S is chosen to be either of the two components of $A_{\ell+\epsilon}$.
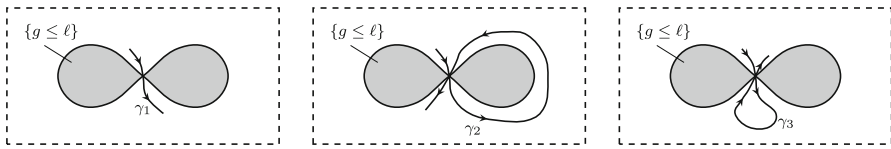
Fig. 14 The paths γ_1 and γ_2 cut through x but γ_3 does not

3) $A_{\ell-\epsilon}$ **is not simple and** $A_{\ell+\epsilon}$ **is simple (in particular, connected).** We can repeat the argument given in this section of the proof for aperiodic fields to show that if f is singly periodic then $A_{\ell-\epsilon}$ must be bounded. Clearly in the doubly periodic case, $A_{\ell-\epsilon} \subset \mathbb{T}$ must also be bounded. We therefore choose a compact domain $M_c \subset M$ of the form $\mathbb{T}$ or $\mathcal{C}(n)$ for some n which contains a neighbourhood of $A_{\ell-\epsilon}$. Let $B_1, \ldots, B_n$ denote the components of $\{g \leq \ell + \epsilon\} \cap M_c$ which intersect $A_{\ell+\epsilon}$. Since $A_{\ell+\epsilon}$ is simple, we know that at most one of these sets is not simple and the remainder are. (One of the sets will 'surround' $A_{\ell+\epsilon}$ which in turn will 'surround' the remaining sets.) Without loss of generality, we assume B_1 is the 'surrounding' set, which may not be simple. By Theorem 4.2, $A_{\ell-\epsilon}$ is homotopy equivalent to $A_{\ell+\epsilon}$ with a 1-cell, denoted γ attached. Clearly γ is contained in B_1 and B_1 is not simple, otherwise $A_{\ell+\epsilon} \cup \gamma$ would be simple. If $M = \mathcal{C}$ then $B_1 \backslash \gamma$ has two components and each of these components contains a loop which is not $\mathcal{C}$-contractible (since $A_{\ell+\epsilon} \cup \gamma$ is compact and so separated from ∞ and $-\infty$ by $B_1 \backslash \gamma$). If $M = \mathbb{T}$ then $B_1 \backslash \gamma$ may have one or two components, in either case each component will contain a non-$\mathbb{T}$-contractible loop (to see this, consider a path on either side of γ which then traverses the boundary of $A_{\ell+\epsilon}$). This means that passing through the saddle point x can only create non-simple components of $\{g \leq \ell - \epsilon\}$, so x is not upper connected.

This completes the proof that the sets of upper and lower connected saddle points of f are almost surely disjoint. We continue to use the notation defined above, and we now show that any saddle point of g must be lower connected or upper connected. As in the aperiodic case, Lemma 4.5 and Conditions 2.3 allow us to conclude that f almost surely has no infinite four-arm saddles.

Suppose that f is doubly periodic so that g is defined on the torus $\mathbb{T}$. We fix a small neighbourhood N of x such that $\{g > \ell\} \cap N$ has precisely two components denoted N_1, N_2. If $\gamma : [0, 1] \to \mathbb{T}$ is a path contained in $\{g > \ell\} \cup \{x\}$, we say that γ *cuts through* x if $x \in \gamma[0, 1]$ and for every $t \in (0, 1)$ such that $\gamma(t) = x$, for $\epsilon > 0$ sufficiently small $\gamma((t - \epsilon, t)) \subset N_1$ and $\gamma((t, t + \epsilon)) \subset N_2$ (intuitively, this means that the image of γ just before hitting x is always on the same side of the saddle point; see Fig. 14). We make an analogous definition for paths contained in $\{g < \ell\} \cup \{x\}$.

First we suppose that there exists a $\mathbb{T}$-contractible loop γ in $\{g < \ell\} \cup \{x\}$ which cuts through x. Since x is a saddle point, it is clear that $A_{\ell+\epsilon}$ must have two components, one of which is surrounded by γ and so is simple. If $A_{\ell-\epsilon}$ is simple, then both components of $A_{\ell+\epsilon}$ must be simple. If $A_{\ell-\epsilon}$ is not simple, we know from Lemma 4.6 that $A_{\ell+\epsilon}$ can contain at most one simple component. In either case, we see that x is lower connected. By symmetry, x is upper connected if there exists a $\mathbb{T}$-contractible loop in $\{g > \ell\} \cup \{x\}$ cutting through x.

Now we suppose that all loops cutting through x are non-$\mathbb{T}$-contractible. We define C_+ to be the union of $\{x\}$ and the components of $\{g > \ell\}$ which x is in the closure of. We define C_- to be the analogous set for $\{g < \ell\}$. At least one of C_+ or C_- must contain a non-contractible loop cutting through x (in order to stop any paths in C_+ cutting through x from joining to form a loop, there must be a loop in $\mathbb{T}\backslash C_+$ cutting through x which blocks them, and since x is the only critical point at level ℓ, such a 'blocking' loop can be found in C_-). We also note that both C_+ and C_- must contain non-$\mathbb{T}$-contractible loops (if C_+ was simple, then we could find a $\mathbb{T}$-contractible loop in C_- cutting through x, and this argument is symmetric).

Suppose that γ_1 is a loop contained in C_+ which cuts through x. If every non-$\mathbb{T}$-contractible loop in C_+ intersects x, then $C_+\backslash\{x\}$ is simple, so $A_{\ell+\epsilon}$ is simple and hence x is lower connected. Therefore we may assume that C_+ contains a non-$\mathbb{T}$-contractible loop γ_2 which does not intersect x. Let η_1 and η_2 be two loops which generate the fundamental group of $\mathbb{T}$, we denote the concatenation of n copies of η_1 and m copies of η_2 by $n\eta_1 + m\eta_2$. We suppose $\gamma_1 \simeq n_1\eta_1 + m_1\eta_2$ and $\gamma_2 \simeq n_2\eta_1 + m_2\eta_2$ where $\simeq$ denotes being $\mathbb{T}$-homotopic. It is known (see [7, Section 1.2.3]) that any loop homotopic to $n_1\eta_1 + m_1\eta_2$ must intersect any loop homotopic to $n_2\eta_1 + m_2\eta_2$ unless $n_1m_2 = n_2m_1$.

If $n_1m_2 \neq n_2m_1$ then we know that any non-contractible loop in C_- must intersect either γ_1 or γ_2. Since $C_- \cap C_+ = \{x\}$ clearly such a loop must intersect γ_1 at x. By the previous paragraph this means that $C_-\backslash\{x\}$ is simple and so x is upper connected.

If $n_1m_2 = n_2m_1$ then by choosing a path ξ in $C_+\backslash\{x\}$ which joins γ_1 to γ_2 we consider the concatenated loop

$$n_2\gamma_1 + \xi + n_1(-\gamma_2) + (-\xi)$$

where $-$ denotes inverting the direction of the path. This path is contained in C_+ and cuts through x (since γ_1 does but γ_2 and ξ do not hit x) and since $n_1m_2 = n_2m_1$ this loop is $\mathbb{T}$-contractible. However this contradicts the above supposition, so this case is not possible. This completes the proof in the doubly periodic case. The proof in the singly periodic case is simply a repetition of parts of the proof for aperiodic and doubly periodic fields, so we omit it. □

All that remains is to complete the calculations for the special class of degenerate fields.

Proof (Proposition 1.20) We recall that a random variable Y is Rayleigh distributed with parameter $\sigma > 0$ if

$$\mathbb{P}(Y \leq x) = 1 - e^{-x^2/(2\sigma^2)}$$

for all $x \geq 0$, and we denote this as $Y \sim \text{Ray}(\sigma)$. If $Y \sim \text{Ray}(\sigma)$ and θ is an independent uniform-$[0, 2\pi]$ random variable then it is well known that $Y\cos(\theta) \sim \mathcal{N}(0, \sigma^2)$.

Let f be the Gaussian field with spectral measure

$$\rho = \alpha\delta_0 + \frac{\beta}{2}(\delta_K + \delta_{-K}) + \frac{\gamma}{2}(\delta_L + \delta_{-L})$$

where $\alpha + \beta + \gamma = 1$ so that the covariance function of f is

$$\kappa(x) = \alpha + \beta \cos(2\pi K \cdot x) + \gamma \cos(2\pi L \cdot x).$$

Then f has the representation

$$f(x) = X_0 + Y_1 \cos(2\pi K \cdot x + \theta_1) + Y_2 \cos(2\pi L \cdot x + \theta_2) \qquad (22)$$

where $X_0 \sim \mathcal{N}(0, \alpha)$, $Y_1 \sim \text{Ray}(\sqrt{\beta})$, $Y_2 \sim \text{Ray}(\sqrt{\gamma})$, the θ_i are uniformly distributed on $[0, 2\pi]$ and all of these random variables are independent. (By well known properties of the Rayleigh distribution, the field defined by the right hand side of (22) will have Gaussian finite dimensional distributions and simple calculations show that the covariance function of this field is κ.)

Let

$$u = \frac{1}{K_1 L_2 - K_2 L_1} \begin{pmatrix} L_2 \\ -L_1 \end{pmatrix} \qquad v = \frac{1}{K_1 L_2 - K_2 L_1} \begin{pmatrix} K_2 \\ -K_1 \end{pmatrix}$$

and $P = \{tu + sv : t, s \in [0, 1]\}$. Then P is the parallelogram associated with f and we consider $f_{\mathbb{T}}$ to be the restriction of f to P when P is identified with the two dimensional torus. By rotating the axes, we may assume that $K_1 > 0$ (since this does not affect the definition of c_{LS}). Some basic calculations show that, on the event

$$\{Y_1 \neq 0\} \cap \{Y_2 \neq 0\} \cap \{Y_1 \pm Y_2 \neq 0\},$$

$f_{\mathbb{T}}$ has four critical points which occur at different levels and are all non-degenerate. Therefore $f_{\mathbb{T}}$ is almost surely a Morse function. Moreover, the critical points of $f_{\mathbb{T}} - X_0$ occur at levels $-Y_1 - Y_2$, $-|Y_1 - Y_2|$, $|Y_1 - Y_2|$, $Y_1 + Y_2$. Clearly the critical points at levels $Y_1 + Y_2$ and $-Y_1 - Y_2$ are a local maximum and local minimum respectively. We can use Lemma 4.6 to characterise the other two critical points and the number of simple components of $\{f_{\mathbb{T}} - X_0 \geq \ell\}$, denoted $N_{\text{simple}}(\ell)$, at different levels. Specifically, since $\{f_{\mathbb{T}} - X_0 \geq Y_1 + Y_2\} = \emptyset$ and $\{f_{\mathbb{T}} - X_0 \geq -Y_1 - Y_2\} = \mathbb{T}$, we see that $N_{\text{simple}}(\ell) = 0$ whenever $|\ell| > Y_1 + Y_2$. Then, since the critical points at level $Y_1 + Y_2$ and $-Y_1 - Y_2$ must be a local maximum and local minimum respectively, applying Lemma 4.6 shows that $N_{\text{simple}}(\ell)$ does not change as ℓ passes through $-Y_1 - Y_2$ and decreases by one as ℓ passes through $Y_1 + Y_2$. Therefore

$$N_{\text{simple}}(\ell) = \begin{cases} 0, & \text{if } \ell > Y_1 + Y_2, \\ 1, & \text{if } \ell \in (|Y_1 - Y_2|, Y_1 + Y_2), \\ 0, & \text{if } \ell \in (-Y_1 - Y_2, -|Y_1 - Y_2|), \\ 0, & \text{if } \ell < -Y_1 - Y_2. \end{cases}$$

Now assume that $\theta_1 = \theta_2 = 0$. Then by traversing the parallelogram P across one of its edges (depending on which of Y_1, Y_2 is bigger) we can find a closed path which is not $\mathbb{T}$-contractible, on which $f_{\mathbb{T}} - X_0$ is bounded below by $|Y_1 - Y_2|$, so in particular

is positive. This shows that $\{f_{\mathbb{T}} - X_0 \geq 0\}$ has a non-simple component (for general values of θ_1, θ_2, such a path will also exist, but it will be translated). Since $\{f_{\mathbb{T}} - X_0 \geq |Y_1 - Y_2|\}$ consists of a single simple component, this implies that the critical point at level $|Y_1 - Y_2|$ must be a lower connected saddle point. Then by Lemma 4.6 we see that

$$N_{\text{simple}}(\ell) = \begin{cases} 1, & \text{if } \ell \in (|Y_1 - Y_2|, Y_1 + Y_2], \\ 0, & \text{otherwise.} \end{cases}$$

It is then clear that

$$p_{m+}(x) = \frac{1}{\text{Area}(P)} p_{X_0 + Y_1 + Y_2}(x) \quad \text{and} \quad p_{s-}(x) = \frac{1}{\text{Area}(P)} p_{X_0 + |Y_1 - Y_2|}(x).$$

Arguments given in the proof of Lemma 2.5 for periodic functions show that

$$N_{ES,R}(\ell) = \mathbb{1}_{\ell - X_0 \in (|Y_1 - Y_2|, Y_1 + Y_2]} \cdot \frac{\pi R^2}{\text{Area}(P)} + O(R + N_{\text{tang}}(\partial B(R))).$$

Arguing as in the proof of Lemma 2.4 it can be shown that $\mathbb{E}(N_{\text{tang}}(\partial B(R)) = O(R)$ as $R \to \infty$. Then for $\ell \in \mathbb{R}$

$$\mathbb{E}\left(\left| \frac{N_{ES,R}(\ell)}{\pi R^2} - \frac{1}{\text{Area}(P)} \mathbb{1}_{\ell - X_0 \in (|Y_1 - Y_2|, Y_1 + Y_2]} \right| \right) = O\left(\frac{1}{R} \right),$$

so that $N_{ES,R}(\ell)/(\pi R^2)$ converges in L^1 to the random variable $\frac{1}{\text{Area}(P)} \mathbb{1}_{\ell - X_0 \in (|Y_1 - Y_2|, Y_1 + Y_2]}$ which is non-deterministic provided $\alpha > 0$ or $\ell > 0$. In particular, this convergence shows that

$$c_{ES}(\ell) := \lim_{R \to \infty} \frac{1}{\pi R^2} \mathbb{E}(N_{ES,R}(\ell)) = \frac{1}{\text{Area}(P)} \mathbb{P}(\ell - X_0 \in (|Y_1 - Y_2|, Y_1 + Y_2]).$$

We can repeat these arguments for the sum of upper and lower excursion sets to derive the corresponding expression for $c_{LS}(\ell)$. $\qquad\square$

References

1. Adler, R.J., Taylor, J.E.: Random Fields and Geometry. Springer, Berlin (2009)
2. Azaïs, J.M., Wschebor, M.: Level Sets and Extrema of Random Processes and Fields. Wiley, Hoboken (2009)
3. Beliaev, D., Kereta, Z.: On the Bogomolny–Schmit conjecture. J. Phys. A Math. Theor. **46**(45), 455003 (2013)
4. Berry, M.V.: Regular and irregular semiclassical wavefunctions. J. Phys. A Math. General **10**(12), 2083–2091 (1977)
5. Cheng, D., Schwartzman, A.: Expected number and height distribution of critical points of smooth isotropic Gaussian random fields. Bernoulli **24**(4B), 3422–3446 (2018)
6. Dennis, M.R.: Nodal densities of planar Gaussian random waves. Eur. Phys. J. Spec. Top. **145**(1), 191–210 (2007)
7. Farb, B., Margalit, D.: A Primer on Mapping Class Groups (pms-49). Princeton University Press, Princeton (2011)
8. Handron, D.G.: Generalized billiard paths and Morse theory for manifolds with corners. Topol. Its Appl. **126**(1–2), 83–118 (2002)
9. Ingremeau, M., Rivera, A.: A lower bound for the Bogomolny–Schmit constant for random monochromatic plane waves. Math. Res. Lett. **26**(4), 1179–1186 (2019). https://doi.org/10.4310/MRL.2019.v26.n4.a9
10. Kurlberg, P., Wigman, I.: Variation of the Nazarov–Sodin constant for random plane waves and arithmetic random waves. Adv. Math. **330**, 516–552 (2018)
11. Lax, P.D., Phillips, R.S.: The asymptotic distribution of lattice points in Euclidean and non-Euclidean spaces. J. Funct. Anal. **46**(3), 280–350 (1982)
12. Nazarov, F., Sodin, M.: On the number of nodal domains of random spherical harmonics. Am. J. Math. **131**(5), 1337–1357 (2009)
13. Nazarov, F., Sodin, M.: Asymptotic laws for the spatial distribution and the number of connected components of zero sets of Gaussian random functions. J. Math. Phys. Anal. Geom. **12**(3), 205–278 (2016)
14. Nicolaescu, L.: On the Kac–Rice formula. Unpublished (2014)
15. Sodin, M.: Lectures on random nodal portraits. Probab. Stat. Phys. St. Petersburg **91**, 395–422 (2016)
16. Swerling, P.: Statistical properties of the contours of random surfaces. IRE Trans. Inf. Theory **8**(4), 315–321 (1962)
17. Wigman, I.: On the nodal lines of random and deterministic Laplace eigenfunctions. Spectr. Geom. **84**, 285–297 (2012)