F. Nazarov, M. Sodin, Asymptotic laws for the spatial distribution and the number of connected components of zero sets of Gaussian random functions, *Zh. Mat. Fiz. Anal. Geom.*, 2016, Volume 12, Number 3, 205–278

DOI: https://doi.org/10.15407/mag12.03.205

Journal of Mathematical Physics, Analysis, Geometry
2016, vol. 12, No. 3, pp. 205–278

Asymptotic Laws for the Spatial Distribution and the Number of Connected Components of Zero Sets of Gaussian Random Functions

F. Nazarov*

Dept. of Math. Sciences, Kent State University, Kent OH 44242, USA

E-mail: nazarov@math.kent.edu

M. Sodin**

School of Math. Sciences Tel Aviv University Tel Aviv 69978, Israel

E-mail: sodin@post.tau.ac.il

Received September 2, 2015

We study the asymptotic laws for the spatial distribution and the number of connected components of zero sets of smooth Gaussian random functions of several real variables. The primary examples are various Gaussian ensembles of real-valued polynomials (algebraic or trigonometric) of large degree on the sphere or torus, and translation-invariant smooth Gaussian functions on the Euclidean space restricted to large domains.

Key words: smooth Gaussian functions of several real variables, the number of connected components of the zero set, ergodicity.

Mathematics Subject Classification 2010: 60G15.

Contents

*Supported by grants No. 2006136, 2012037 of the United States – Israel Binational Science Foundation and by U.S. National Science Foundation Grants DMS-0800243, DMS-1265623.

**Supported by grants No. 2006136, 2012037 of the United States – Israel Binational Science Foundation and by grant No. 166/11 of the Israel Science Foundation of the Israel Academy of Sciences and Humanities.

In memory of Volodya Matsaev

1. Introduction and the Main Results

The result we present has two main versions. The first one treats zero sets of smooth Gaussian functions on the Euclidean space $\mathbb{R}^m$ with translation-invariant distributions. The second version deals with parametric ensembles of smooth Gaussian functions in open domains in $\mathbb{R}^m$. We also show how to translate the second version to parametric ensembles of smooth Gaussian functions on smooth manifolds without boundary.

In Appendix A, all parts of the theory of smooth Gaussian functions needed for understanding this work are developed from scratch. Appendix B contains the proof of the Fomin–Grenander–Maruyama theorem in the multidimensional setting. None of the results in these Appendices is our own work.

1.1. The translation invariant case

Suppose $F\colon \mathbb{R}^m \to \mathbb{R}$ is a continuous Gaussian random function with translation-invariant distribution (meaning that for every $v \in \mathbb{R}^m$, the continuous Gaussian functions F and $F(\cdot + v)$ have the same distribution). Then the covariance kernel

$$K(x, y) = \mathcal{E}\{F(x)F(y)\}$$

of F depends only on the difference $x-y$ and can be written in the form $K(x,y) = k(x-y)$ where $k : \mathbb{R}^m \to \mathbb{R}$ is a symmetric positive definite function. By Bochner's theorem*, k can be written as the Fourier integral

$$k(x) = (\mathcal{F}\rho)(x) = \int\limits_{\mathbb{R}^m} e^{2\pi i(x\cdot\lambda)}\,\mathrm{d}\rho(\lambda)$$

of some finite symmetric positive Borel measure ρ on $\mathbb{R}^m$, which is called *the spectral measure* of F.

We denote by $Z(F) = F^{-1}\{0\}$ the (random) zero set of F. Let S be any bounded open convex set in $\mathbb{R}^m$ containing the origin. By $S(R)$ we denote the set $\{x \in \mathbb{R}^m \colon x/R \in S\}$. By $N_{S}(R;F)$ we denote the number of the connected components of $Z(F)$ that are contained in $S(R)$. When S is the unit ball $B = \{x \colon |x| < 1\}$, we will write $N(R;F)$ instead of $N_{B}(R;F)$.

We say that a finite complex-valued measure μ on $\mathbb{R}^m$ is *Hermitian* if for each Borel set $E \subset \mathbb{R}^m$, we have $\mu(-E) = \overline{\mu(E)}$. By $\mathcal{F}\mu$ we denote the Fourier integral of the measure μ, and by $\operatorname{spt}(\mu)$ we denote the closed support of μ.

Theorem 1. *Suppose that the spectral measure ρ of a continuous Gaussian translation-invariant function F satisfies the following conditions:*

(ρ1)

$$\int\limits_{\mathbb{R}^m} |\lambda|^4\,\mathrm{d}\rho(\lambda) < \infty;$$

(ρ2) *ρ has no atoms;*

(ρ3) *ρ is not supported on a linear hyperplane.*

Then there exists a constant $\nu \geqslant 0$ such that for every bounded open convex set $S \subset \mathbb{R}^m$ containing the origin,

$$\lim_{R\to\infty} \frac{N_{S}(R;F)}{\operatorname{vol} S(R)} = \nu \text{ almost surely} \qquad \text{and} \qquad \lim_{R\to\infty} \mathcal{E}\Big|\frac{N_{S}(R;F)}{\operatorname{vol} S(R)} - \nu\Big| = 0\,. \tag{1.1.1}$$

Furthermore, $\nu > 0$ provided that

(ρ4) *there exist a finite compactly supported Hermitian measure μ with $\operatorname{spt}(\mu) \subset \operatorname{spt}(\rho)$ and a bounded domain $D \subset \mathbb{R}^m$ such that $\mathcal{F}\mu\big|_{\partial D} < 0$ and $(\mathcal{F}\mu)(u_0) > 0$ for some $u_0 \in D$.*

1.1.1. Rôle of conditions (ρ1) – (ρ3). Condition (ρ1) guarantees that $F \in C^{2-}(\mathbb{R}^m) \stackrel{\text{def}}{=} \bigcap_{\alpha\in(0,1)} C^{1+\alpha}(\mathbb{R}^m)$. Condition ($\rho$3) says that the distribution of the gradient ∇F is non-degenerate. Together conditions (ρ1) and (ρ3) allow us to

*See [3, § 20] for the original proof or [17] for a clear self-contained exposition.

think of the zero set $Z(F)$ as a collection of pairwise disjoint smooth hypersurfaces that partition $\mathbb{R}^m$ into "nodal domains".

The translation invariance allows us to consider the probability distribution measure generated by F on an appropriate space of functions as an invariant measure with respect to the action of the abelian group $\mathbb{R}^m$ by translations $(\tau_v g)(\cdot) = g(\cdot + v)$. Condition $(\rho 2)$ ensures that this action is ergodic, which in turn implies that the limit ν is non-random.

1.1.2. Condition $(\rho 4)$. Condition $(\rho 4)$ is essentially equivalent to the possibility to deterministically create at least one bounded connected component of the zero set $Z(F)$. The measures not satisfying $(\rho 4)$ have to be very degenerate. In particular, the support of any measure not satisfying $(\rho 4)$ has to be contained in a quadratic hypersurface in $\mathbb{R}^m$. We prove this, as well as some other observations pertaining to condition $(\rho 4)$, in Appendix C..

On the other hand, the Fourier transform of the Lebesgue surface measure on the sphere centered at the origin is radial and sign changing. So if $\mathrm{spt}(\rho)$ is a sphere in $\mathbb{R}^m$ centered at the origin then $(\rho 4)$ is still satisfied.

These observations suffice to check condition $(\rho 4)$ in most interesting examples.

1.1.3. What can be said about the constant ν? Unfortunately, the proof of Theorem 1 does not provide much information about the value of the constant ν. There is a huge discrepancy between the lower bounds that can be extracted from the "barrier construction" introduced in [25] and the upper bounds obtained by computing the mean number of special points in the nodal domains or in the zero set (cf. Nastasescu's undergraduate thesis [24]).

It is worth noting that the limiting constant $\bar{\nu}$ equals the expectation $\mathcal{E}\big\{\mathrm{vol}(G_0)^{-1}\big\}$, where G_0 is the connected component of $\mathbb{R}^m \setminus Z(F)$ containing the origin (or any other given point in $\mathbb{R}^m$). The random variable $\mathrm{vol}(G_0)$ is, perhaps, even more mysterious than $N(R; F)$. Our theorem shows that $\mathcal{P}\big\{\mathrm{vol}(G_0) < +\infty\big\} > 0$, but we still do not even know how to prove that this probability is 1, not mentioning any efficient tail estimate for its distribution.

1.1.4. Further remarks about Theorem 1. Theorem 1 can be viewed as a version of the "law of large numbers" for the "connected component process" on $\mathbb{R}^m$ associated with the Gaussian function F. In most applications, one does not need as strong convergence as is guaranteed by Theorem 1 and just the convergence in probability (which is equivalent to the convergence in distribution for constant limits) is enough.

Note also that the value of the intensity $\nu(F)$ is completely determined by the covariance kernel $k(x-y)$ of F, or, which is the same, by the spectral measure ρ.

Our last remark concerns a non-degenerate linear change of variables. Let $T\colon \mathbb{R}^m \to \mathbb{R}^m$ be a non-degenerate linear operator and let $\widetilde{F}(x) = F(Tx)$. Then

$\widetilde{F}$ is also a Gaussian translation-invariant function. Moreover, for every $S \subset \mathbb{R}^m$ and $R > 0$, we have $N_{TS}(R; F) = N_S(R; \widetilde{F})$, whence,

$$\frac{\mathcal{E}N_S(R;\widetilde{F})}{\operatorname{vol} S(R)} = |\det T| \, \frac{\mathcal{E}N_{TS}(R;F)}{\operatorname{vol}(TS)(R)} \, .$$

Thus, if the intensity $\nu(F)$ exists, then so does $\nu(\widetilde{F})$, and we have the relation

$$\nu(\widetilde{F}) = |\det T| \, \nu(F) \, .$$

1.2. Parametric Gaussian ensembles

Definition 1 (parametric Gaussian ensemble)**.** *A parametric Gaussian ensemble (f_L) on an open set $U \subset \mathbb{R}^m$ (or on an m-dimensional manifold X without boundary) is any family (f_L) of continuous Gaussian functions on U (on X, respectively) indexed by some countable set of numbers $L \geqslant 1$ accumulating only at $+\infty$.*

Many interesting parametric Gaussian ensembles (in particular, two examples considered below in Sec. 2) arise from the following construction. Let X be a smooth compact m-dimensional manifold without boundary. Let $\mathcal{H}_L$ be a sequence of real finite-dimensional Hilbert spaces of continuous functions on X indexed by some scaling parameter $L \geqslant 1$ so that $\lim_{L\to\infty} \dim \mathcal{H}_L = \infty$. Since, for every $x \in X$, the point evaluation $\mathcal{H}_L \ni f \mapsto f(x)$ is a continuous linear functional on $\mathcal{H}_L$, there is a unique function $K_L^x \in \mathcal{H}_L$ such that $f(x) = \langle f, K_L^x \rangle$. The function $K_L(x, y) = K_L^x(y)$ is called *the reproducing kernel* of the space $\mathcal{H}_L$. Since, $K_L^x \in \mathcal{H}_L$, we have $K_L^x(y) = \langle K_L^x, K_L^y \rangle$, so $K_L(x, y)$ is symmetric. Now let $\{e_k\}$ be an orthonormal basis in $\mathcal{H}_L$. Then for every $f \in \mathcal{H}_L$, we have $f = \sum_k \langle f, e_k \rangle e_k$ in $\mathcal{H}_L$ and, therefore, pointwise. Thus

$$K_L(x, y) = K_L^x(y) = \sum_k e_k(x) e_k(y).$$

Consider the continuous Gaussian function

$$f_L(x) = \sum_k \xi_k e_k(x), \qquad x \in X \, ,$$

where ξ_k are independent standard real Gaussian random variables. The covariance kernel of the Gaussian function f_L equals

$$\mathcal{E}\{f_L(x) f_L(y)\} = \sum_k e_k(x) e_k(y) = K_L(x, y) \, ,$$

so it does not depend on the choice of the orthonormal basis $\{e_k\}$ and coincides with the reproducing kernel of $\mathcal{H}_L$. It follows that the distribution of $f_{_L}$ also does not depend on the choice of the basis and is completely determined by the space $\mathcal{H}_L$ itself. We shall call this continuous Gaussian function $f_{_L}$ *the continuous Gaussian function generated by* $\mathcal{H}_L$.

1.2.1. Normalization. We say that a continuous Gaussian function f on U with the covariance kernel K is *normalized* if

$$\mathcal{E}\{f(x)^2\} = K(x,x) = 1 \qquad \text{for all } x \in U\,.$$

If the random Gaussian function f is not normalized but non-degenerate (that is, $\mathcal{E}\{f(x)^2\} > 0$, or, what is the same, $\mathcal{P}\{f(x) = 0\} = 0$ for every $x \in U$), we can just replace f by $\widetilde{f}(x) = \frac{f(x)}{\sqrt{K(x,x)}}$, which will correspond to replacing the covariance kernel $K(x,y)$ by

$$\widetilde{K}(x,y) = \frac{K(x,y)}{\sqrt{K(x,x)\,K(y,y)}}\,, \tag{1.2.1.}$$

without affecting the zero set $Z(f)$ in any way.

Note that if we allow f to degenerate at some points uncontrollably, then the zero set of f may contain deterministic pieces of arbitrarily complicated structure and our talk about the asymptotic behavior of the number of nodal components of f may easily become totally meaningless. Thus,

- *we will always assume that all continuous Gaussian functions and all parametric Gaussian ensembles in this paper are normalized.*

Note that in many basic examples, including the ones we consider below in Sec. 2, the function $x \mapsto K_L(x,x)$ is constant, so the normalization of K reduces to the division by that constant.

1.2.2. Scaling and translation-invariant local limits. Let U be an open set in $\mathbb{R}^m$ and let $(f_{_L})$ be a parametric Gaussian ensemble on U. Let $K_{_L}$ be the covariance kernel of $f_{_L}$.

We define the scaled covariance kernel $K_{x,L}$ at a point $x \in U$ by

$$K_{x,L}(u,v) = K_L\big(x + \frac{u}{L}, x + \frac{v}{L}\big).$$

Note that $K_{x,L}$ is the covariance kernel of the scaled Gaussian function

$$f_{x,L}(u) = f_{_L}\big(x + \frac{u}{L}\big),$$

i.e., $K_{x,L}(u,v) = \mathcal{E}\{f_{x,L}(u) f_{x,L}(v)\}$. Note also that if $x \in U$ is fixed and $L \to \infty$, the sets $U_{x,L} = \{u \in \mathbb{R}^m \colon x + \frac{u}{L} \in U\}$ exhaust $\mathbb{R}^m$.

Definition 2 (translation-invariant limit)**.** *Let (f_L) be a parametric Gaussian ensemble on an open set $U \subset \mathbb{R}^m$ and let K_L be the covariance kernel of f_L. Let $x \in U$. We say that the scaled covariance kernels $K_{x,L}$ have a* translation-invariant limit *if there exists a continuous function $k_x \colon \mathbb{R}^m \to \mathbb{R}$ such that, for each $u, v \in \mathbb{R}^m$,*

$$\lim_{L\to\infty} K_{x,L}(u,v) = k_x(u-v)\,. \tag{1.2.2}$$

We say that the parametric Gaussian ensemble (f_L) has a translation invariant limit at the point x if there exists a translation invariant continuous Gaussian function F_x on $\mathbb{R}^m$ such that, for every finite point set $\mathcal{U} \in \mathbb{R}^m$, the finite-dimensional Gaussian vectors $f_{x,L}|_{\mathcal{U}}$ converge to $F_x|_{\mathcal{U}}$ in distribution.

We call the function F_x *the local limiting function* and its spectral measure ρ_x *the local limiting spectral measure* of the parametric Gaussian ensemble (f_L) at the point x. If a parametric Gaussian ensemble (f_L) on U has a translation invariant limit F_x at some point $x \in U$, then the scaled covariance kernels $K_{x,L}$ have a translation invariant limit as well and the limiting kernel $k_x(u-v)$ is the covariance kernel of F_x. On the other hand, without any additional assumptions, covariance kernels $K_{x,L}(u,v)$ may have a translation invariant limit $k_x(u-v)$ that corresponds to no continuous Gaussian function F. However, within the set-up considered in this paper, these notions become equivalent.

It is natural to believe that if a parametric Gaussian ensemble (f_L) on U has a translation invariant limit at every point $x \in U$, then for large L, we can count the nodal components of f_L in some open set $V \subset U$ by partitioning V into nice sets V_j of size larger than $1/L$, choosing some points $x_j \in V_j$, approximating the number of nodal components of f_L in each set V_j by the number of nodal components of F_{x_j} in $(V_j)_{x_j,L} = \bigl\{v \in \mathbb{R}^m \colon x_j + L^{-1}v \in V_j\bigr\}$, and adding all these counts up. If we are lucky enough, the nodal components of F_x may have asymptotic intensity $\bar\nu(x) = \nu(F_x)$ and then the total count we get will be typically close to

$$\sum_j \bar\nu(x_j)\operatorname{vol}(V_j)_{x_j,L} = L^m \sum_j \bar\nu(x_j)\operatorname{vol} V_j\,.$$

If we are even luckier, the quantity $\bar\nu(x)$ may depend on x in a nice enough way for the Riemann sums $\sum_j \bar\nu(x_j)\operatorname{vol} V_j$ to converge to $\int_V \bar\nu \,\mathrm{d}\operatorname{vol}$.

The formalization of this intuitive argument requires some accuracy, especially because the standard integral calculus nowadays is Lebesgue, not Riemann. The classical form of a convergence statement for integrals in the Lebesgue language is that of the dominated convergence theorem, whose general structure is

- *Given a sequence of nice objects that converge in some fairly weak and easy to check sense to some limiting object, and assuming that our pre-limiting*

objects are uniformly controlled in some way, the limiting object is nice as well, and some integral functional of the limiting object is the limit of the integral functionals of the pre-limit objects.

Our Theorem 2 will be exactly of this structure.

We have already introduced in Definition 2 the modes of convergence we will be using. Now it is time to define "controllability".

1.2.3. Uniform smoothness of covariance kernels K_L. The control we want to impose will be two-fold. First, we will need to restrict the typical speed of oscillation of the continuous Gaussian functions f_L. Some restriction of this type is inevitable because fast oscillating continuous Gaussian functions like the Brownian motion on $\mathbb{R}^1$ change sign infinitely many times near every their zero and they still have fairly decent moduli of continuity on the Hölder scale. The control we will impose will guarantee that $f_{x,L} \in C^{2-}(U) = \bigcap_{\tau\in(0,1)} C^{1+\tau}(U)$ on every compact subset of U.

For $k \geqslant 1$, by $C^{k,k}(U \times U)$ we denote the class of functions $g\colon U \times U \to \mathbb{R}$ for which all partial derivatives $\partial_x^\alpha \partial_y^\beta g(x,y)$, $|\alpha|, |\beta| \leqslant k$ (taken in any order) exist and are continuous*. For $L \geqslant 1$, a compact set $Q \subset U$, and $g \in C^{k,k}(U \times U)$, we put

$$\|g\|_{L,Q,k} \stackrel{\text{def}}{=} \max_{|\alpha|,|\beta|\leqslant k} \max_{x,y\in Q} L^{-(|\alpha|+|\beta|)} \big|\partial_x^\alpha \partial_y^\beta g(x,y)\big|.$$

When $L = 1$, we will write $\|g\|_{Q,k}$ instead of $\|g\|_{1,Q,k}$.

If the covariance kernel K of a continuous Gaussian function f on U belongs to $C^{k,k}(U\times U)$, then the semi-norms $\|K\|_{L,Q,k}$ can be computed on "the diagonal" $\alpha = \beta$ and $x = y$:

$$\|K\|_{L,Q,k} = max_{|\alpha|\leqslant k} \max_{x\in Q} L^{-2|\alpha|} \big|\partial_x^\alpha \partial_y^\alpha K(x,y)|_{y=x}\,\big|.$$

A naïve explanation to this fact comes from the Cauchy–Schwarz inequality combined with the formula

$$\partial_x^\alpha \partial_y^\beta K(x,y) = \mathcal{E}\big\{\partial_x^\alpha f(x)\, \partial_y^\beta f(y)\big\},$$

which is true in the case when the derivatives on the RHS exist and are continuous random functions. The proof of this fact for the general case will be given in Appendix A.11.

The uniform smoothness of the kernels K_L with $k \geqslant 1$ is more than enough to erase any distinction between the existence of a translation invariant limit of the kernels $K_{x,L}$ and the existence of a translation invariant limit at x of the parametric Gaussian ensemble (f_L).

*in which case, they do not depend on the order

1.2.4. Local uniform non-degeneracy of the parametric Gaussian ensemble (f_L). Our second restriction will be of the opposite character. While the local uniform smoothness guarantees that the continuous Gaussian functions f_L do not change too fast or in a too rough way, the condition we discuss in this section will ensure that f_L cannot change too slowly or in a too predictable way in any direction. Without any such restriction, there will be nothing that would prevent long regular components to prevail and, with our methods, we will either not be able to say anything at all in such case, or will just conclude that some limit is 0, which would merely mean that the particular scaling we have chosen is a wrong one for the problem. With all this in mind, let us pass to the formal definitions.

Let $K \in C^{1,1}(U \times U)$ and let C_x be the matrix with the entries

$$C_x(i,j) = \partial_{x_i}\, \partial_{y_j}\, K(x,y)|_{y=x}\,, \qquad x \in U\,.$$

If K is the covariance kernel of some C^1 Gaussian function f on U, then C_x is the covariance matrix of the Gaussian random vector $\nabla f(x)$. Assuming that $\det C_x \neq 0$, we can say that the density of the probability distribution of the random Gaussian vector $\nabla f(x)$ in $\mathbb{R}^m$ is given by

$$p(\xi) = \frac{1}{(2\pi)^{m/2}\sqrt{\det C_x}}\, e^{-\frac{1}{2}(C_x^{-1}\xi\,\xi)}\,.$$

Since in this case C_x^{-1} is positive definite, we have

$$\max_{\xi} p(\xi) = p(0) = (2\pi)^{-m/2}(\det C_x)^{-1/2}\,.$$

Definition 3 (local uniform non-degeneracy of (f_L))**.** *We say that a parametric Gaussian ensemble (f_L) on some open set $U \subset \mathbb{R}^m$ is locally uniformly non-degenerate if the corresponding kernels $K_L(x,y)$ are at least in $C^{1,1}(U \times U)$ and for every compact set $Q \subset U$,*

$$\varliminf_{L\to\infty} \inf_{x\in Q} \det C_{x,L} > 0$$

where $C_{x,L}$ is the matrix with the entries

$$C_{x,L}(i,j) = \partial_{u_i}\partial_{v_j} K_{x,L}(u,v)\big|_{u=v=0} = L^{-2}\partial_{x_i}\partial_{y_j} K_L(x,y)\big|_{y=x}\,.$$

As the argument above shows, if f_L is C^1-smooth, then our non-degeneracy condition just means that, for every compact set $Q \subset U$, there is a uniform upper bound for the densities of the distributions of all Gaussian vectors $L^{-1}\nabla f_L(x)$ with $x \in Q$.

Suppose that, for some $x \in U$, the kernels $K_{x,L}$ have a translation invariant limit and the convergence holds in the semi-norm $\|\cdot\|_{Q,1}$ for some compact set Q containing x in its interior. Then the matrix $C_{x,L}$ converges to the matrix c_x with the entries

$$c_x(i,j) = -\big(\partial_{u_i}\partial_{u_j} k_x\big)(0) = 4\pi^2 \int\limits_{\mathbb{R}^m} \lambda_i \lambda_j \, \mathrm{d}\rho_x(\lambda)$$

and we see that in this case the limiting measure ρ_x satisfies

$$\inf_{\xi \in \mathbb{S}^{m-1}} \int\limits_{\mathbb{R}^m} \big|\lambda\,\xi\big|^2 \, \mathrm{d}\rho_x(\lambda) > 0\,,$$

which means that ρ_x cannot be supported on any linear hyperplane $\big\{\lambda\colon \lambda\,\xi = 0\big\}$, i.e., condition $(\rho 3)$ is satisfied.

1.2.5. Controllability. Now we are ready to say what we mean by a locally uniformly controllable parametric Gaussian ensemble (f_L) on U.

Definition 4 (locally uniform controllability)**.** *The parametric Gaussian ensemble (f_L) on an open set $U \subset \mathbb{R}^m$ is* locally uniformly controllable *if it is locally uniformly non-degenerate and the corresponding covariance kernels K_L satisfy*

$$\overline{\lim_{L\to\infty}} \|K_L\|_{L,Q,2} < \infty$$

for every compact set $Q \subset U$.

The above considerations combined with results presented in Appendix (see A.11 and A.12) imply that

- *if the kernels K_L are locally uniformly controllable and if the scaled kernels $K_{x,L}$ have translation-invariant limits, then the limiting spectral measure ρ_x satisfies assumptions $(\rho 1)$ and $(\rho 3)$ of Theorem 1.*

1.2.6. Tameness

Definition 5 (tame ensembles)**.** *The parametric Gaussian ensemble (f_L) on an open set $U \subset \mathbb{R}^m$ is* tame *if*

(i) *it is locally uniformly controllable,*

and there exists a Borel subset $U' \subset U$ of full Lebesgue measure such that, for all $x \in U'$,

(ii) *the scaled kernels $(K_{x,L})$ have translation invariant limits;*

(iii) *the limiting spectral measure ρ_x has no atoms.*

A tame parametric Gaussian ensemble (f_L) has a translation invariant limit at every point $x \in U'$. Moreover, by Theorem 1, the point intensity $\bar\nu(x) \stackrel{\text{def}}{=} \nu(F_x)$ associated with the ensemble (f_L) is well-defined on U'.

1.3. The main result

Before we state our second main theorem, we will introduce one more object. Let U be an open set in $\mathbb{R}^m$ and let f be a continuous Gaussian function on U. We say that a (depending on the implicit probability variable ω) Borel measure n on U is *a connected component counting measure* of f if $\operatorname{spt}(n) \subset Z(f)$ and the n-mass of each connected component of $Z(f)$ equals 1. Note that we do not require the dependence of n on ω to be measurable in any sense (for this reason, we do not call n a random measure), so in the statement of the next theorem we will have to use "the upper expectation" $\mathcal{E}^*$ instead of the usual one $\mathcal{E}$.

Theorem 2. *Suppose that $(f_{_L})$ is a tame parametric Gaussian ensemble on an open set $U \subset \mathbb{R}^m$. Then*

(i) *the function $x \mapsto \bar\nu(x)$ is measurable and locally bounded in U;*
and

(ii) *for every sequence of connected component counting measures $n_{_L}$ of $f_{_L}$ and for every compactly supported in U continuous function φ, we have*

$$\lim_{L\to\infty} \mathcal{E}^*\Big\{\Big|\frac{1}{L^m}\int \varphi \,\mathrm{d}n_{_L} - \int \varphi\bar\nu \,\mathrm{d}\operatorname{vol}\Big|\Big\} = 0\,.$$

Note that the second statement of that theorem can be strengthened to

$$\lim_{L\to\infty} \mathcal{E}^*\Big\{\Big|\frac{1}{L^m}\int \varphi \,\mathrm{d}n_{_L} - \int \varphi\bar\nu \,\mathrm{d}\operatorname{vol}\Big|^q\Big\} = 0$$

for some $q = q(m) > 1$ (which tends to 1 as $m \to \infty$) without any essential change in the proof but we are not aware of any application of this stronger result for which the current version would not suffice as well.

1.4. The manifold version of Theorem 2

Theorem 2 can be transferred to parametric Gaussian ensembles on smooth manifolds without boundary. Everywhere in this section X is an m-dimensional C^2-manifold without boundary (not necessarily compact) that can be covered by countably many charts, all charts being assumed open and C^2-smooth, and $(f_{_L})$ is a parametric Gaussian ensemble on X. We start with two definitions.

Definition 6 (tame ensembles on manifolds)**.** *We say that a parametric Gaussian ensemble $(f_{_L})$ on X is* tame *if, for every chart $\pi\colon U \to X$, the parametric Gaussian ensemble $(f_{_L} \circ \pi)$ is tame on U.*

This definition implies that for every chart $\pi\colon U \to X$, the parametric Gaussian ensemble $(f_{_L} \circ \pi)$ satisfies the assumptions of Theorem 2. So the associated point intensity $\bar\nu_\pi$ belongs to $L^\infty_{\mathtt{loc}}(U)$.

Definition 7 (Volumes compatible with smooth structure)**.** *We say that a locally finite Borel positive measure* vol_X *on* X *is a volume* compatible with the smooth structure *of* X *if for every chart* $\pi : U \to X$*, the measures* $\pi_* \mathrm{vol}$ *and* vol_X *are mutually absolutely continuous and the corresponding Radon–Nikodym densities are continuous on* $\pi(U)$*.*

Of course, the main example we have in mind giving this definition is that of a smooth Riemannian manifold X and the volume generated by the Riemannian metric on X.

We also note that despite the manifold X may be endowed with no measure, the words "almost every $x \in X$" still have meaning because all push-forward measures $\pi_* \mathrm{vol}$ corresponding to various charts $\pi : U \to X$ of X are mutually absolutely continuous wherever they can be compared to each other.

At last, we can state the manifold version of Theorem 2.

Theorem 3. *Suppose that* (f_L) *is a tame parametric Gaussian ensemble on* X*. Then*

(i) *there exists a locally finite Borel non-negative measure* n_∞ *on* X *such that for every choice of connected component counting measures* n_L *of* f_L *and every function* $\varphi \in C_0(X)$*,*

$$\lim_{L\to\infty} \mathcal{E}^* \Big\{ \Big| \frac{1}{L^m} \int \varphi \, \mathrm{d}n_L - \int \varphi \, \mathrm{d}\, n_\infty \Big| \Big\} = 0 \,;$$

(ii) *for every chart* $\pi : U \to X$*, the measure* n_∞ *coincides on* $\pi(U)$ *with the push-forward* $\pi_*(\bar\nu_\pi \ \mathrm{vol})$ *where* $\bar\nu_\pi$ *is the point intensity associated with the parametric Gaussian ensemble* $f_L \circ \pi$*;*

(iii) *if* vol_X *is some volume measure compatible with the smooth structure of* X*, then* n_∞ *is absolutely continuous with respect to* vol_X*, and there exists a set* $X' \subset X$ *of full* vol_X *such that, for every* $x \in X'$*, the quantity*

$$\mathfrak{n}(x) = \bar\nu_\pi(\pi^{-1}(x)) \, \frac{\mathrm{d}\pi_* \mathrm{vol}}{\mathrm{d}\,\mathrm{vol}_X}(x)$$

is well-defined and does not depend on the choice of the chart $\pi \colon U \to X$ *with* $x \in \pi(U)$*. Moreover,*

$$\mathrm{d}n_\infty = \mathfrak{n} \, \mathrm{d}\,\mathrm{vol}_X \,.$$

The point of part (iii) is that, for vol_X-almost all $x \in X$, it allows one to compute the Radon–Nikodym derivative $\frac{\mathrm{d}n_\infty}{\mathrm{d}\,\mathrm{vol}_X}(x)$ using *any* chart containing x. In particular, nothing prevents us from choosing for each point its own individual chart.

1.4.1. How to verify tameness? Theorem 3, as stated, has an essential shortcoming: it may be somewhat unpleasant to verify tameness of (f_L) because formally it requires one to estimate various quantities in the local coordinates given by π for every chart $\pi\colon U \to X$, however weird or ugly. The next two observations (both of purely technical nature) allow one to substantially reduce this workload. Recall that an atlas on X is any family of charts $\mathbb{A} = \{\pi_\alpha : U_\alpha \to X\}_\alpha$ such that $\bigcup_\alpha \pi_\alpha(U_\alpha) = X$. Here is our first observation:

- *Suppose $\mathbb{A}$ is an atlas on X and that, for every chart $\pi_\alpha \in \mathbb{A}$, $(f_L \circ \pi_\alpha)$ is tame on U_α. Then (f_L) is tame on X.*

The possibility to check the tameness for the charts from any atlas of our choice is quite a relief. However, one unpleasant thing still remains. It may (and often does) happen that for every point $x \in X$ there is one "preferred" chart $\pi_x : U_x \to X$ covering x such that the computations in this chart are a piece of cake in any infinitesimal neighborhood of x but not quite so even a bit away from x. In this case we would strongly prefer to compute all quantities and check all conditions at x using its preferred chart π_x. However, we are still formally required to run the computations concurrently on any compact subset of any given chart using the local coordinates given by that particular chart. Our next observation takes care of this difficulty.

Definition 8. *We say that an atlas $\mathbb{A}$ of X has* uniformly bounded distortions *if there exists a constant $A > 0$ such that all partial derivatives of orders $\leqslant 2$ of all coordinate functions of all transition maps between the charts of $\mathbb{A}$ are bounded by A.*

Note that this definition doesn't require X to be uniform in any sense; rather it requires that the charts in $\mathbb{A}$ be small enough so that X doesn't show any non-trivial structure within the union of each chart with all charts it intersects, and that the chart scalings be more or less consistent with each other within small regions. Our second observation says that

- *If the atlas $\mathbb{A}$ has uniformly bounded distortions, then to check the tameness of (f_L) on X, it suffices to check the relevant conditions and uniform bounds (on compact subsets of X) for the related quantities computed in the charts (U_x, π_x) at the points $\pi_x^{-1}(x)$ only.*

These two observations may be not obvious and we will explain them more in Section 9..

Note that in our examples, we will deal with compact manifolds admitting a transitive group $\mathcal{G}$ of diffeomorphisms leaving the parametric Gaussian ensemble (f_L) under consideration invariant (meaning that for each $\mathbf{g} \in \mathcal{G}$ and each L,

the continuous Gaussian functions f_L and $f_L \circ \mathbf{g}$ have the same distribution). In such situation, all one needs is to find one chart $\pi : U \to X$ such that the atlas consisting of the charts $\mathbf{g} \circ \pi$, $\mathbf{g} \in \mathcal{G}$, has uniformly bounded distortions. Then one may fix his/her favorite point $x = \pi(u)$ in that chart, and establish all the required bounds and conditions at this single point for this single chart. All passages about "almost every x" and suprema and infima over Q in all conditions can be ignored in such setup because all the related objects and quantities do not depend on x at all.

1.5. The final remarks about Theorems 2 and 3

1.5.1. Note that the particular choice of the counting measures n_L plays no rôle. The reason is that, for large L, with high probability the overwhelming part of n_L comes from components of arbitrarily small diameter. Such components can be viewed as single points at the macroscopic level.

1.5.2. If the manifold X is compact, we can apply the conclusion of Theorem 3 to $\varphi \equiv 1$ and to obtain the asymptotics $(n_\infty(X) + o(1))L^m$ for the typical (and the mean) total number of nodal components of f_L on X as $L \to \infty$. Of course, this asymptotic law is really useful only when $n_\infty(X) > 0$. Finding an asymptotic formula (or even a decent estimate) for the variance of the total number of nodal components in such regimes remains an open problem.

1.5.3. The proof of Theorem 2 also shows that the value $\bar\nu(x)$ can be recovered as a double-scaling limit. In Lemma 12 we show that, for almost every $x \in U$ and for each $\varepsilon > 0$, we have

$$\lim_{R\to\infty} \overline{\lim_{L\to\infty}} \mathcal{P}\Big\{ \Big| \frac{N\big(x, R/L; f_L\big)}{\operatorname{vol} B(R)} - \bar\nu(x) \Big| > \varepsilon \Big\} = 0$$

where $N\big(x, \frac{R}{L}; f_L\big) = N(R, f_{x,L})$ is the number of the connected components of the zero set $Z(f_L)$ contained in the open ball centered at x of radius R/L.

1.5.4. A few words should be said about the measurability issues. While we prove every measurability result that is necessary for the completeness of the formal exposition, when possible, we circumvent this discussion by using upper integral and upper expectation instead of the usual ones. Note that the Borel measurability of similar quantities has been discussed in detail in Rozenshein's Master Thesis [26, Sec. 5].

1.6. Pertinent works

1.6.1. The earliest non-trivial lower bound for the mean number of connected components is, probably, due to Malevich. In [23], she considered a C^2-smooth

translation-invariant Gaussian random function F on $\mathbb{R}^2$ with *positive* covariance kernel decaying at a certain rate at infinity. She proved that $\mathcal{E}N(R;F)/R^2$ is bounded from below and from above by two positive constants. Her proof of the lower bound uses Slepian's inequality and probably cannot be immediately extended to models with covariance kernels that change their signs.

1.6.2. Several years ago, Bogomolny and Schmit [5] proposed a bond percolation model for the description of the zero set of the translation-invariant Gaussian function F on $\mathbb{R}^2$ whose spectral measure is the Lebesgue measure on the unit circumference. This model completely ignores slowly decaying correlations between values of the random function at different points and is very far from being rigorous. The predictions of Bogomolny and Schmit were checked by computational experiments carried out by Nastasescu [24], Konrad [18], and Beliaev and Kereta [2]. The observed value of the constant ν was very close to but still noticeably less than the Bogomolny and Schmit prediction. It would be very interesting to reveal a hidden "universality law" that provides the rigorous foundation for the work done by Bogomolny and Schmit. Note also that it is not clear whether or to what degree their approach can be extended to make reasonably accurate predictions about the behavior of nodal components of translation-invariant Gaussian functions corresponding to other spectral measures in $\mathbb{R}^2$ or in dimensions $m > 2$.

1.6.3. In [25], we showed that for the Gaussian ensemble of spherical harmonics of large degree L on the two-dimensional sphere, the total number $N(f_L)$ of connected components of $Z(f_L)$ satisfies

$$\mathcal{P}\left\{\left|L^{-2}N(f_{_L}) - \upsilon\right| > \varepsilon\right\} < C(\varepsilon)e^{-c(\varepsilon)\dim\mathcal{H}_L},$$

with some $\upsilon > 0$. The limiting function for this ensemble is the one considered by Bogomolny and Schmit. The case of higher dimension (in a slightly different setting) was treated by Rozenshein in [26]. The exponential concentration of $N(f_{_L})/L^2$ is interesting since this model has slowly decaying correlations.

We were unable to prove the exponential concentration for other ensembles considered here. The difficulty is caused by the small components, which do not exist when $f_{_L}$ is an eigenfunction of the Laplacian. Even in the univariate case, the question about the exponential concentration in Theorem 1 remains open; cf. Tsirelson's lecture notes [30].

Some lower bounds for the number of connected components of the zero set and for other similar quantities were obtained in different settings by Bourgain and Rudnick [7], Fyodorov, Lerario, Lundberg [9], Gayet and Welschinger [10, 11, 12], Lerario and Lundberg [22] using the "barrier construction" from [25].

1.6.4. Certain versions of main results of this work were presented at the St. Petersburg Summer School in Probability and Statistical Physics (June, 2012) and appeared in the lecture notes [29].

1.6.5. There have been several works of interest relying on ideas and techniques developed in this paper, among which those by Bourgain [6], Canzani and Sarnak [8], Kurlberg and Wigman [20] and Sarnak and Wigman [28] deserve special attention of the reader.

Acknowledgments. On many occasions, Boris Tsirelson helped us by providing information and references concerning Gaussian measures and measurability. Leonid Polterovich and Zeév Rudnick have read parts of a preliminary version of this work and made several valuable comments, which we took into account. We have had encouraging discussions of this work with Andrei Okounkov, Peter Sarnak, and Jean-Yves Welschinger. Alex Barnett and Maria Nastasescu showed us the beautiful and inspiring simulations. We thank them all.

2. Examples

Here, we point out two examples illustrating Theorem 3. In our examples, the manifold X has a natural Riemannian metric and a transitive group of isometric diffeomorphisms that leaves the distribution of (f_L) invariant. As discussed near the end of Sec. 1.4, this will allow us to check the conditions of Theorem 3 at just one point $x \in X$ with respect to a natural local chart associated with this point and to conclude that the limiting measure n_∞ on the manifold X is a constant multiple of the Riemannian volume on X. Moreover, since in our examples the kernels $K_{x,L}(u,v)$ converge to $k(u-v)$ uniformly with all derivatives on compact subsets of $\mathbb{R}^m \times \mathbb{R}^m$, the uniform smoothness and non-degeneracy of the kernels $K_{x,L}(u,v)$ can be derived from the corresponding properties of the limiting kernel $k(u-v)$. Passing to the limit in our examples is an elementary exercise in Taylor calculus and complex analysis. This list of examples may be continued (see [8, 20, 26, 28]) but the two ones we included into this paper should be already enough to convey its main message, which is

- *Under not unreasonably unfavorable conditions, establishing the asymptotics for the number of nodal domains for parametric Gaussian ensembles is about as easy (or, if the reader prefers, as hard) as establishing the convergence of the scaled kernels and investigating the resulting limiting processes.*

2.1. Trigonometric ensemble

Here $\mathcal{H}_n$ is the subspace of $L^2(\mathbb{T}^m)$ that consists of real-valued trigonometric polynomials

$$\operatorname{Re} \sum_{\nu \in \mathbb{Z}^m \,:\, |\nu|_\infty \leqslant n} c_\nu e^{2\pi \mathrm{i}(\nu \cdot x)}$$

in m variables of degree $\leqslant n$ in each variable. A straightforward computation shows that the corresponding normalized covariance kernel coincides with the

product of m Dirichlet's kernels:

$$K_n(x,y)=\prod_{j=1}^{m}\frac{\sin\left[\pi(2n+1)(x_j-y_j)\right]}{(2n+1)\sin\left[\pi(x_j-y_j)\right]}\,.$$

In this case, it is natural to choose the degree n as the scaling parameter L. The scaled kernels $K_{x,n}(u,v)=K_n(x+n^{-1}u,x+n^{-1}v)$ do not depend on the choice of the point $x\in\mathbb{T}^m$. They extend analytically from $\mathbb{R}^m\times\mathbb{R}^m$ to $\mathbb{C}^m\times\mathbb{C}^m$ and the extensions converge uniformly on compact subsets of $\mathbb{C}^m\times\mathbb{C}^m$ to

$$\prod_{j=1}^{m}\frac{\sin 2\pi(u_j-v_j)}{2\pi(u_j-v_j)}\,.$$

This implies the convergence with all derivatives on all compact subsets of $\mathbb{R}^m\times\mathbb{R}^m$. The limiting spectral measure ρ is the normalized Lebesgue measure on the cube $[-1,1]^m\subset\mathbb{R}^m$.

2.2. Kostlan's ensemble

In this case, $\mathcal{H}_n$ is the space of the homogeneous real-valued polynomials of degree n in $m+1$ variables restricted to the unit sphere $\mathbb{S}^m$. The scalar product in $\mathcal{H}_n$ is given by

$$\langle f,g\rangle=\sum_{|J|=n}\binom{n}{J}^{-1}f_Jg_J \tag{2.2.1}$$

where

$$f(X)=\sum_{|J|=n}f_JX^J,\quad g(X)=\sum_{|J|=n}g_JX^J,\qquad X^J=x_0^{j_0}x_1^{j_1}x_2^{j_2}\dots x_m^{j_m},$$

and

$$J=(j_0,j_1,j_2,\dots,j_m),\quad |J|=j_0+j_1+j_2+\ldots+j_m,\quad \binom{n}{J}=\frac{n!}{j_0!j_1!j_2!\dots j_m!}\,.$$

The form of the scalar product (2.2.1) comes from the complexification: extending the homogeneous polynomials f and g to $\mathbb{C}^{m+1}$, one can show that

$$\langle f,g\rangle_{\mathcal{H}_n}=c(n,m)\int\limits_{\mathbb{C}^{m+1}}f(Z)\overline{g(Z)}e^{-|Z|^2}\,\mathrm{d}\,\mathrm{vol}(Z)\,,$$

i.e., $\langle f,g\rangle_{\mathcal{H}_n}$ coincides (up to a positive factor) with the scalar product in the Fock–Bargmann space (or any other weighted L^2-space of entire functions with fast decaying radial weight).

It is known that the complexified Kostlan ensemble is *the only unitarily invariant* Gaussian ensemble of homogeneous polynomials. On the other hand, there are many other orthogonally invariant Gaussian ensembles, all of them having been classified by Kostlan [19] (see [9, Section 2] for some details).

The normalized covariance kernel of Kostlan's ensemble equals $(x \cdot y)^n$. Take $x = (0, \dots, 0, 1)$ (the "North Pole") and consider the local chart $\pi(u) = \big(u, \sqrt{1 - |u|^2}\big)$ where u runs over a small neighbourhood of the origin in $\mathbb{R}^m$. Then

$$\begin{aligned}\pi(u)\,\pi(v) &= \sum_{j=1}^m u_j v_j + \Big(1 - \sum_{j=1}^m u_j^2\Big)^{\frac{1}{2}} \Big(1 - \sum_{j=1}^m v_j^2\Big)^{\frac{1}{2}} \\ &= 1 - \frac{1}{2}\sum_{j=1}^m (u_j - v_j)^2 + O\big(|u|^4 + |v|^4\big) \text{ as } u, v \to 0\,.\end{aligned}$$

This suggests that the correct scaling in this case is $L = \sqrt{n}$ and the limiting covariance kernel is

$$\begin{aligned}\lim_{n\to\infty} \Big(\pi(n^{-\frac{1}{2}}u)\,\pi(n^{-\frac{1}{2}}v)\Big)\Big)^n &= \lim_{n\to\infty} \Big(1 - (2n)^{-1}\sum_{j=1}^m (u_j - v_j)^2 + n^{-2} O\big(|u|^4 + |v|^4\big)\Big)^n \\ &= \exp\Big\{-\frac{1}{2}\sum_{j=1}^m (u_j - v_j)^2\Big\}\,.\end{aligned}$$

The justification of the local uniform convergence with all derivatives is similar to that in the previous example, and we skip it. The limiting spectral measure is the Gaussian measure on $\mathbb{R}^m$ with the density $c_m e^{-2\pi^2|\lambda|^2}$.

An interesting feature of this example is a very rapid off diagonal decay of the covariance kernel.

3. Notation

We denote by $B(x, r)$ the open ball of radius r centered at x, $\bar{B}(x, r)$ denotes the corresponding closed ball. $B(r)$ always denotes the open ball of radius r centered at the origin.

For a closed set $\Gamma \subset \mathbb{R}^m$, we denote by $N(x, r; \Gamma)$ the number of the connected components of Γ that are contained in the open ball $B(x, r)$, and by $N^*(x, r; \Gamma)$ the number of the connected components of Γ that intersect the closed ball $\bar{B}(x, r)$. If $\Gamma = Z(f)$ is the zero set of a continuous function f, we will abuse the notation slightly and write $N(x, r; f)$ instead of $N(x, r; Z(f))$. For a bounded open convex set S and $R > 0$, we denote by $N_S(R; \Gamma)$ the number of connected components of Γ that are contained in $S(R) = \{u\colon R^{-1}u \in S\}$.

Throughout the paper, we denote by c and C various positive constants, which may depend on the dimension m and on the parameters of the Gaussian process or ensemble under consideration (the parameters in the conditions of Theorems 1 and 2) but on nothing else. The values of these constants may vary from line to line. Usually, the constants denoted by C should be thought of as large, and the constants denoted by c as small. The notation $a \lesssim b$ means that $a \leqslant C \cdot b$.

Quite frequently, we will use the smoothness class $C^{2-}(U)$ ($U \subset \mathbb{R}^m$ is an open set), which we define as

$$C^{2-}(U) = \bigcap_{0<\beta<1} C^{1+\beta}(U)\,.$$

Recall that to check that $g \in C^{1+\beta}(U)$ it suffices to show that $g \in C^1(U)$ and the first order partial derivatives $\partial_{x_i} g$ are β-Hölder functions on any closed ball $\bar{B} \subset U$.

4. Lemmata

In this section, we present several lemmas needed for the proofs of Theorems 1 and 2.

4.1. Some integral geometry

The first result is taken from [25, Claim 5.1] where it appears in a slightly different form.

Lemma 1. *Suppose $\Gamma \subset \mathbb{R}^m$ is a closed set and $S \supset B(1)$ is a bounded open convex set. Then, for $0 < r < R$,*

$$\int\limits_{S(R-r)} \frac{N(u,r;\Gamma)}{\operatorname{vol} B(r)}\,\mathrm{d}\operatorname{vol}(u) \leqslant N_S(R;\Gamma) \leqslant \int\limits_{S(R+r)} \frac{N^*(u,r;\Gamma)}{\operatorname{vol} B(r)}\,\mathrm{d}\operatorname{vol}(u)\,.$$

Note that $N(u,r;\Gamma)$ is lower semicontinuous as a function of u. Proving the Lebesgue measurability of $u \mapsto N^*(u,r;\Gamma)$ without additional assumptions on Γ may be somewhat nontrivial. However, we will apply this lemma only in the case when the set of connected components of Γ is countable. Also, replacing the integral on the RHS by the upper Lebesgue integral will not affect the argument in any way. So, we will not dwell on this particular measurability.

P r o o f. For a connected component γ of Γ, we put

$$G_*(\gamma) = \bigcap_{y\in\gamma} B(y,r), \quad G^*(\gamma) = \bigcup_{y\in\gamma} \bar{B}(y,r)\,.$$

Note that since γ is closed, $G_*(\gamma)$ is open and $G^*(\gamma)$ is closed. Also, for any $y \in \gamma$, $G_*(\gamma) \subset B(y,r) \subset G^*(\gamma)$. Hence,

$$\begin{aligned}\int\limits_{S(R-r)} N(u,r;\Gamma)\,\mathrm{d}\,\mathrm{vol}(u) &= \int\limits_{S(R-r)} \Big(\sum_{\gamma\colon \gamma\subset B(u,r)} 1\Big)\mathrm{d}\,\mathrm{vol}(u)\\ &\leqslant \int\limits_{S(R-r)} \Big(\sum_{\gamma\colon \gamma\subset S(R),\, u\in G_*(\gamma)} 1\Big)\mathrm{d}\,\mathrm{vol}(u)\\ &= \sum_{\gamma\subset S(R)} \mathrm{vol}\big(G_*(\gamma)\cap S(R-r)\big)\\ &\leqslant N_S(R;\Gamma)\ \mathrm{vol}\,B(r)\,,\end{aligned}$$

proving the left inequality.

On the other hand,

$$\begin{aligned}\int\limits_{S(R+r)} N^*(u,r;\Gamma)\,\mathrm{d}\,\mathrm{vol}(u) &= \int\limits_{S(R+r)} \Big(\sum_{\gamma\colon u\in G^*(\gamma)} 1\Big)\,\mathrm{d}\,\mathrm{vol}(u)\\ &= \sum_{\gamma} \mathrm{vol}\big(G^*(\gamma)\cap S(R+r)\big)\,.\end{aligned}$$

Since for every connected component γ having a common point y with $S(R)$, we have $B(y,r)\subset G^*(\gamma)\cap S(R+r)$, the last sum is at least $N_S(R;\Gamma)\ \mathrm{vol}\,B(r)$, so the right inequality holds as well. ∎

4.2. Stability of components of the zero set under small perturbations

If zero is not a critical value of a smooth function then the zero set of this function is stable under small perturbations. The following lemma, which quantifies this general principle, is taken from [25, Claim 4.2] where it was proven in the two-dimensional case. The proof of the general case needs no changes.

Denote by V_{+t} the open t-neighbourhood of a set $V\subset\mathbb{R}^m$.

Lemma 2. *Fix $\alpha,\beta>0$. Let F be a C^1-smooth function on an open ball $B\subset\mathbb{R}^m$ such that at every point $u\in B$, either $|F(u)|>\alpha$, or $|\nabla F(u)|>\beta$. Then each component γ of the zero set $Z(F)$ with $\mathrm{dist}(\gamma,\partial B)>\alpha/\beta$ is contained in an open "annulus" $A_\gamma\subset\gamma_{+\alpha/\beta}$ bounded by two smooth connected hypersufaces such that $F=+\alpha$ on one boundary component of A_γ, and $F=-\alpha$ on the other one. Furthermore, the "annuli" A_γ are pairwise disjoint.*

As an immediate corollary, we obtain

Lemma 3. *Under the assumptions of the previous lemma, suppose that* $G \in C(B)$ *with* $\sup_B |G| < \alpha$. *Then each component* γ *of* $Z(F)$ *with* $\operatorname{dist}(\gamma, \partial B) > \alpha/\beta$ *generates a component* $\widetilde{\gamma}$ *of the zero set* $Z(F+G)$ *such that* $\widetilde{\gamma} \subset \gamma_{+\alpha/\beta}$ *and different components* $\gamma_1 \neq \gamma_2$ *of* $Z(F)$ *generate different components* $\widetilde{\gamma}_1 \neq \widetilde{\gamma}_2$ *of* $Z(F+G)$.

4.3. Statistical independence of g and ∇g

Quite often, we will use the following well-known fact:

Lemma 4. *Suppose* $U \subset \mathbb{R}^n$ *is an open set and* $g\colon U \to \mathbb{R}$ *is a Gaussian* C^1*-function on* U *that has constant variance. Then* $g(u)$ *and its gradient* $\nabla g(u)$ *are independent for every* $u \in U$.

P r o o f. Denote by g_{u_i} the partial derivative $\partial_{u_i} g$. The covariance kernel $K(u,v) = \mathcal{E}\{g(u)g(v)\}$ is a C^1-function, and $\mathcal{E}\{g_{u_i}(u)g(u)\} = K_{u_i}(u,v)\big|_{v=u}$. Since the function $u' \mapsto K(u',u)$ attains its maximal value at $u'=u$ and is C^1-smooth, we have $K_{u_i}(u,v)\big|_{v=u} = 0$. Therefore, $\mathcal{E}\{g_{u_i}(u)g(u)\} = 0$. Since $g(u)$ and $\nabla g(u)$ are jointly Gaussian, this orthogonality implies their independence. ■

We will be using the following corollary:

Lemma 5. *Suppose* $F\colon \mathbb{R}^m \to \mathbb{R}$ *is a Gaussian random function with translation-invariant distribution whose spectral measure* ρ *satisfies conditions* (ρ*1) and* (ρ*3). Then the distribution of the Gaussian vector* $\big(F(u), \nabla F(u)\big)$ *does not degenerate.*

P r o o f of Lemma 5. By Lemma 4, $F(u)$ and $\nabla F(u)$ are independent. Hence, it suffices to show that the distribution of $\nabla F(u)$ does not degenerate. If it degenerates, then there exists a non-zero vector $v \in \mathbb{R}^m$ such that

$$0 = \mathcal{E}\{(v\,\nabla F)^2\} = 4\pi^2 \int\limits_{\mathbb{R}^m} (v\,\lambda)^2 \,\mathrm{d}\rho(\lambda)\,,$$

which is impossible since, due to condition (ρ3), the spectral measure ρ cannot be supported on a linear hyperplane. ■

5. Quantitative Versions of Bulinskaya's Lemma

5.1. Preliminaries

The purpose of this part is to show that certain "bad events" have negligibly small probability. The particular bad events we want to get rid of are the event that the random Gaussian function and its gradient are simultaneously small at some point and the event that $Z(f)$ has too many connected components.

Everywhere in this part, $B_R \subset \mathbb{R}^m$ is a fixed ball of large radius $R > 1$, $S = \partial B_R$, U is an open neighbourhood of B_{R+1}, and f is a continuous Gaussian function on U with the covariance kernel K. As usual, we will assume that the function f is normalized, that is, $\mathcal{E}|f(x)|^2 = K(x,x) = 1$, $x \in U$. We will impose certain bounds on the smoothness and non-degeneracy. These bounds are normalized versions of estimates used in the definition of controllability of parametric Gaussian ensembles. Namely, we will assume that

(i) the kernel K is $C^{2,2}(U \times U)$-smooth and

$$\max_{|\alpha| \leqslant 2} \max_{x \in \bar{B}_{R+1}} \left| \partial_x^\alpha \, \partial_y^\alpha K(x,y)|_{y=x} \right| \leqslant M < \infty \,,$$

and that

(ii) the process f is non-degenerate on U and

$$\inf_{x \in \bar{B}_{R+1}} \det C_x \geqslant \kappa > 0 \,,$$

where C_x is the covariance matrix of the Gaussian random vector $\nabla f(x)$, that is, the matrix with the entries $C_x(i,j) = \partial_{x_i} \partial_{y_j} K(x,y)|_{y=x}$.

- *Till the end of Sec. 5, the constants M and κ remain fixed and all the constants that appear in the conclusions of all results proven here may depend on M and κ.*

As shown in Appendix A.11, the smoothness assumption (i) yields that, almost surely, the process f is $C^{2-}(U)$-smooth. We will be frequently using a quantitative version of this statement, which is also given in A.11. For a closed ball $\bar{B} \subset U$, denote by $\|f\|_{\bar{B},1+\beta}$ the least N such that

$$\max_{\bar{B}} |f| \leqslant N, \quad \max_{\bar{B}} |\nabla f| \leqslant N, \quad \text{and} \quad |\nabla f(x) - \nabla f(y)| \leqslant N|x-y|^\beta \text{ for } x, y \in \bar{B}.$$

Then, for every $\beta < 1$ and every $p < \infty$,

$$\sup_{x \in \bar{B}_R} \mathcal{E}\big\{ \|f\|^p_{\bar{B},1+\beta} \big\} \leqslant C(\beta, p, M) < \infty \,.$$

5.2. The function Φ

A prominent rôle in our approach will be played by the function

$$\Phi(x) = |f(x)|^{-t} |\nabla f(x)|^{-tm} \,, \qquad x \in U, \ t \in (0,1) \,,$$

and by its spherical version

$$\Phi_S(x) = |f(x)|^{-t} |\nabla_S f(x)|^{-t(m-1)}\,, \qquad x \in S = \partial B_R,\ t \in (0,1)\,,$$

where $\nabla_S f(x)$ is the projection of the vector $\nabla f(x)$ to the tangent space to S at the point $x \in S$. The main feature of this function is that if f and ∇f (or $\nabla_S f$) are very small at two points that are close to each other (in particular, if they are small at the same point), then Φ (Φ_S correspondingly) is very large in a neighbourhood of these two points. At the same time, since f is normalized and ∇f is non-degenerate,

- *the moments $\mathcal{E}\{\Phi^q(x)\}$ and $\mathcal{E}\{\Phi_S^q(x)\}$ are bounded locally uniformly on U and uniformly on S whenever we fix $t < 1 < q$ so that $tq < 1$. Moreover, if t and q satisfying this restrictions are fixed, the suprema $\sup_{\bar{B}_{R+1}} \mathcal{E}\{\Phi^q\}$ and $\sup_S \mathcal{E}\{\Phi_S^q\}$ are bounded by constants depending only on κ.*

5.3. Almost surely, zero is not a critical value of f

As a warm up, we prove a useful qualitative result that goes back to Bulinskaya.

Lemma 6. *Almost surely, the following assertions hold:*

(i) *zero is not a critical value of f;*

(ii) *there is no point $z \in S \cap Z(f)$ at which $\nabla_S f(z) = 0$.*

P r o o f. In the first case, we use the function Φ. Fix a compact set $Q \subset U$ and take a positive $\delta < \operatorname{dist}(Q, \partial U)$. Consider the event

$$\Omega_Q = \big\{\exists z \in Q\colon \text{such that } f(z) = 0,\ \nabla f(z) = 0\big\}$$

and take a ball $\bar{B} \subset U$ centered at z of radius less than δ. Since the function $\nabla f(x)$ is β-Hölder with every $\beta < 1$, we have, for all $x \in \bar{B}$,

$$|f(x)| \lesssim |x - z|, \quad |\nabla f(x)| \lesssim |x - z|^{\beta}\,,$$

whence $\Phi(x) \gtrsim |x - z|^{-t(1+\beta m)}$. Hence, choosing t and β so close to 1 that $t(1 + \beta m) > m$, we see that

$$\int\limits_{Q_{+\delta}} \Phi \,\mathrm{d}\,\mathrm{vol} \geqslant \int\limits_{B} \Phi \,\mathrm{d}\,\mathrm{vol} = +\infty\,.$$

Recalling that $\mathcal{E}\{\Phi(x)\}$ is uniformly bounded on $\bar{Q}_{+\delta}$ and using Fubini's theorem, we conclude that the event Ω_Q has zero probability. It remains to note that U can be covered by countably many compact subsets.

Similarly, in the second case we take Φ_S. As above, the expectation $\mathcal{E}\{\Phi_S(x)\}$ is uniformly bounded on S. Suppose that, for some $z \in S \cap Z(f)$, $\nabla_S f(z) = 0$, that is, the gradient $\nabla f(z)$ is orthogonal to the sphere S. Then, for $x \in S$, we have

$$|f(x)| \lesssim |x - z|, \qquad |\nabla_S f(x)| \lesssim |x - z|^\beta + R^{-1}|x - z| \lesssim |x - z|^\beta,$$

and, thereby, $\Phi_S(x) \gtrsim |x - z|^{-t(1+\beta(m-1))}$. Therefore, choosing t and β so close to 1 that $t(1 + \beta(m - 1)) > m - 1$, we get

$$\int_S \Phi \, \mathrm{d}\,\mathrm{vol}_S = +\infty,$$

and conclude that the event we consider has zero probability. ∎

5.4. With probability close to one, f and ∇f cannot be simultaneously small

Here, we prove a quantitative version of Lemma 6.

Lemma 7. *Given $\delta > 0$, there exists $\tau > 0$ (possibly, depending on R) such that*

$$\mathcal{P}\big\{\min_{x \in \bar{B}_R} \max\{|f(x)|, |\nabla f(x)|\} < \tau\big\} < \delta\,.$$

P r o o f. Denote by Ω_τ the event

$$\big\{\exists z \in \bar{B}_R \colon |f(z)|, |\nabla f(z)| < \tau\big\}$$

and put

$$W = 1 + \|f\|_{\bar{B}_{R+1}, 1+\beta}\,.$$

The parameter $\beta \in (0, 1)$ will be specified later. If the event Ω_τ occurs, then in the ball $B = B(z, \tau)$ with $\tau \in (0, 1)$, we have

$$|f(x)| \leqslant \tau + \tau \|f\|_{\bar{B}_{R+1}, 1+\beta} = W\tau\,,$$

and

$$|\nabla f(x)| \leqslant \tau + \tau^\beta \|f\|_{\bar{B}_{R+1}, 1+\beta} < W\tau^\beta\,.$$

Then, on Ω_τ,

$$\Phi(x) \geqslant \tau^{-t(1+\beta m)} W^{-t(1+m)}\,, \qquad \text{for } x \in B$$

and

$$\int_{B_{R+1}} \Phi \, \mathrm{d}\,\mathrm{vol} \geqslant \int_B \Phi \, \mathrm{d}\,\mathrm{vol} \geqslant c\tau^{m - t(1+\beta m)} W^{-t(1+m)}\,.$$

Therefore,

$$\mathcal{P}\{\Omega_\tau\} \leqslant C\tau^{t(1+\beta m)-m}\ \mathrm{vol}(B_{R+1})\,\mathcal{E}\Big\{W^{t(1+m)}\,\frac{1}{\mathrm{vol}(B_{R+1})}\int\limits_{B_{R+1}}\Phi\,\mathrm{d}\,\mathrm{vol}\Big\}$$

$$\leqslant C\tau^{t(1+\beta m)-m}\,\mathrm{vol}(B_{R+1})\Big(\mathcal{E}\{W^{pt(1+m)}\}\Big)^{\frac{1}{p}}\Big(\frac{1}{\mathrm{vol}(B_{R+1})}\int\limits_{B_{R+1}}\mathcal{E}\{\Phi^q\}\mathrm{d}\,\mathrm{vol}\Big)^{\frac{1}{q}},$$

with $\frac{1}{p}+\frac{1}{q}=1$. The only restriction we have is $\beta<1<q<\frac{1}{t}$. So we can take β and t so close to 1 that the exponent $t(1+\beta m)-m=t-m(1-\beta)$ remains positive. This completes the proof. ∎

5.5. General principle for estimating the number of connected components

Our next aim is to estimate how many connected components of various kinds $Z(f)$ may have. We start with "an abstract scheme", which our estimates will be based on.

Let (X,μ) be a measure space with $0<\mu(X)<\infty$, and let $X=\bigcup_j X_j$ be a cover of X with bounded covering number C_0 (that is, for every $x\in X$, $\#\{j\colon x\in X_j\}\leqslant C_0$). Let $(\Omega,\mathcal{P})$ be a probability space, and let $\big\{(Y_i(\omega),z_i(\omega))\big\}_{1\leqslant i\leqslant N(\omega)}$ be disjoint subsets of X with marked points $z_i\in Y_i$ depending on the parameter $\omega\in\Omega$. Our aim is to estimate the cardinality $N(\omega)$ of the collection $\{Y_i\}$.

Lemma 8. *Let $\Phi\colon X\to\mathbb{R}_+$ be a random function such that, for some $q>1$,*

$$\sup_X \mathcal{E}\{\Phi^q\}<\infty\,.$$

Let $\{W_j\}$ be non-negative random variables such that, for any $p<\infty$,

$$\sup_j \mathcal{E}\{W_j^p\}<\infty\,.$$

Suppose that, for every pair (i,j) with $z_i\in X_j$, we have

$$\int\limits_{X_j\cap Y_i}\Phi\,\mathrm{d}\mu\geqslant\rho\,\mu(X_j\cap Y_i)^{-\sigma}W_j^{-\eta}$$

with some $\rho,\sigma,\eta>0$. Then

$$\mathcal{E}^*\{N^q\}\leqslant(C_0C(\rho,\sigma))^q\,\mu(X)^q\,\Big[\sup_X\mathcal{E}\big\{\Phi^q\big\}\Big]^{\frac{1}{1+\sigma}}\,\Big[\sup_j\mathcal{E}\big\{W_j^{\frac{q\eta}{\sigma}}\big\}\Big]^{\frac{\sigma}{1+\sigma}}\,.$$

P r o o f of Lemma 8. Let N_j be the number of i's such that $z_i \in X_j$. Then, for at least $\frac{1}{2}N_j$ indices i with this property, we have $\mu(X_j \cap Y_i) \leqslant \frac{2}{N_j}\,\mu(X_j)$, and therefore,

$$\int\limits_{X_j \cap Y_i} \Phi \,\mathrm{d}\mu \geqslant c(\rho,\sigma) N_j^{\sigma} \mu(X_j)^{-\sigma} W_j^{-\eta}\,,$$

whence,

$$\int\limits_{X_j} \Phi \,\mathrm{d}\mu \geqslant c(\rho,\sigma) N_j^{1+\sigma} \mu(X_j)^{-\sigma} W_j^{-\eta}\,.$$

Applying Hölder's inequality with the exponents $1+\sigma$ and $\frac{1+\sigma}{\sigma}$, we get

$$\begin{aligned} N \leqslant \sum_j N_j &= \sum_j N_j \mu(X_j)^{-\frac{\sigma}{1+\sigma}} W_j^{-\frac{\eta}{1+\sigma}}\, \mu(X_j)^{\frac{\sigma}{1+\sigma}} W_j^{\frac{\eta}{1+\sigma}} \\ &\leqslant \Big(\sum_j N_j^{1+\sigma} \mu(X_j)^{-\sigma} W_j^{-\eta}\Big)^{\frac{1}{1+\sigma}} \Big(\sum_j \mu(X_j) W_j^{\frac{\eta}{\sigma}}\Big)^{\frac{\sigma}{1+\sigma}} \\ &\leqslant C(\rho,\sigma) \Big(\int\limits_X \Phi \,\mathrm{d}\mu\Big)^{\frac{1}{1+\sigma}} \Big(\sum_j \mu(X_j) W_j^{\frac{\eta}{\sigma}}\Big)^{\frac{\sigma}{1+\sigma}}\,. \end{aligned}$$

Then

$$\mathcal{E}^*\big\{N^q\big\} \leqslant C(\rho,\sigma)^q\, \mathcal{E}\Big\{\Big(\int\limits_X \Phi \,\mathrm{d}\mu\Big)^{\frac{q}{1+\sigma}} \Big(\sum_j \mu(X_j) W_j^{\frac{\eta}{\sigma}}\Big)^{\frac{q\sigma}{1+\sigma}}\Big\}$$

and, applying Hölder's inequality with the same exponents again, we obtain

$$\mathcal{E}^*\big\{N^q\big\} \leqslant C(\rho,\sigma)^q \Big[\mathcal{E}\Big\{\Big(\int\limits_X \Phi \,\mathrm{d}\mu\Big)^q\Big\}\Big]^{\frac{1}{1+\sigma}} \Big[\mathcal{E}\Big\{\Big(\sum_j \mu(X_j) W_j^{\frac{\eta}{\sigma}}\Big)^q\Big\}\Big]^{\frac{\sigma}{1+\sigma}}\,.$$

At last, using Hölder's inequality with the exponents $\frac{q}{q-1}$ and q, we get

$$\mathcal{E}\Big\{\Big(\int\limits_X \Phi \,\mathrm{d}\mu\Big)^q\Big\} \leqslant \mu(X)^{q-1} \int\limits_X \mathcal{E}\big\{\Phi^q\big\}\,\mathrm{d}\mu \leqslant \mu(X)^q \sup_X \mathcal{E}\big\{\Phi^q\big\}$$

and

$$\begin{aligned} \mathcal{E}\Big\{\Big(\sum_j \mu(X_j) W_j^{\frac{\eta}{\sigma}}\Big)^q\Big\} &= \mathcal{E}\Big\{\Big(\sum_j \mu(X_j)^{1-\frac{1}{q}} \mu(X_j)^{\frac{1}{q}} W_j^{\frac{\eta}{\sigma}}\Big)^q\Big\} \\ &\leqslant \Big[\sum_j \mu(X_j)\Big]^{q-1} \Big[\sum_j \mu(X_j)\, \mathcal{E}\big\{W_j^{\frac{q\eta}{\sigma}}\big\}\Big] \\ &\leqslant \big(C_0 \mu(X)\big)^q \sup_j \mathcal{E}\big\{W_j^{\frac{q\eta}{\sigma}}\big\}\,. \end{aligned}$$

Finally,

$$\mathcal{E}^*\{N^q\} \leqslant C(\rho,\sigma)^q \Big[\mu(X)^q \sup_X \mathcal{E}\{\Phi^q\}\Big]^{\frac{1}{1+\sigma}} \Big[\big(C_0\mu(X)\big)^q \sup_j \mathcal{E}\big\{W_j^{\frac{q\eta}{\sigma}}\big\}\Big]^{\frac{\sigma}{1+\sigma}}$$
$$\leqslant (C_0C(\rho,\sigma))^q\,\mu(X)^q \Big[\sup_{x\in X} \mathcal{E}\big\{\Phi(x)^q\big\}\Big]^{\frac{1}{1+\sigma}} \Big[\sup_j \mathcal{E}\big\{W_j^{\frac{q\eta}{\sigma}}\big\}\Big]^{\frac{\sigma}{1+\sigma}},$$

completing the proof. ∎

5.6. Components on a sphere

For the sphere $S = \partial B_R$, we denote by $\mathfrak{N}(S;f)$ the number of connected components of $S \setminus Z(f)$.

Lemma 9. *There are positive constants $C < \infty$ and $q > 1$ such that*

$$\mathcal{E}^*\{\mathfrak{N}^q(S;f)\} \leqslant CR^{q(m-1)}.$$

P r o o f of Lemma 9. We cover the sphere S with bounded covering number by closed spherical caps X_j of Euclidean radius 1, and denote by $\bar{B}_j$ the closed m-dimensional Euclidean balls of radius 1 having the same centers as X_j. The total number of the caps in the cover is $\lesssim R^{m-1}$. By Y_i we denote the connected components of $S \setminus Z(f)$. In each domain Y_i we fix a point z_i where the gradient $\nabla f(z_i)$ is directed normally to S, that is, $\nabla_S f(z_i) = 0$. The number of i's such that, for some j, $X_j \subset Y_i$ is $\lesssim R^{m-1}$. Thus, in what follows, we consider only those i's for which Y_i does not contain any X_j.

In order to apply Lemma 8 with the function Φ_S and with $W_j = \|f\|_{\bar{B}_j,1+\beta}$ we need to establish the lower bounds for the integrals

$$\int\limits_{X_j\cap Y_i} \Phi_S \,\mathrm{d}\,\mathrm{vol}_S(x) \qquad \text{with } \Phi_S = |f|^{-t}\,|\nabla_s f|^{-t(m-1)},$$

assuming that $z_i \in X_j$.

Since the sets $Y_i \cap X_j$ and $X_j \setminus Y_i$ aren't empty, the closed set $\partial Y_i \cap X_j$ is not empty too. Denote $\rho_i = \operatorname{dist}(z_i, \partial Y_i \cap X_j) \leqslant 2$ and take a closest to z_i point $p \in \partial Y_i \cap X_j$. By V_i we denote the spherical cap centered at z_i such that $p \in \partial V_i$. Note that, by the construction, $\mathrm{vol}_S(Y_i \cap X_j) \geqslant \mathrm{vol}_S(V_i \cap X_j) \gtrsim \rho_i^{m-1}$. Since $f(p) = 0$, we have

$$|f(x)| \lesssim \rho_i \,\|f\|_{\bar{B}_j,1+\beta} = \rho_i W_j\,, \qquad x \in V_i \cap X_j\,.$$

Furthermore, since $\nabla_S f(z_i) = 0$,

$$|\nabla_S f(x)| \lesssim \rho_i^\beta \,\|f\|_{\bar{B}_j,1+\beta} + \frac{\rho_i}{R}\|f\|_{\bar{B}_j,1+\beta} \lesssim \rho_i^\beta W_j\,, \qquad x \in V_i \cap X_j\,.$$

Hence, on $V_i \cap X_j$ we have

$$\Phi_S \gtrsim \rho_i^{-t(1+\beta(m-1))}\, W_j^{-tm} \gtrsim (\mathrm{vol}_S(V_i \cap X_j))^{-t(\frac{1}{m-1}+\beta)}\, W_j^{-tm}\,,$$

and

$$\int\limits_{Y_i \cap X_j} \Phi_S \,\mathrm{d}\,\mathrm{vol}_S(x) \geqslant \int\limits_{V_i \cap X_j} \Phi_S \,\mathrm{d}\,\mathrm{vol}_S(x) \gtrsim (\mathrm{vol}_S(V_i \cap X_j))^{1-t(\frac{1}{m-1}+\beta)} W_j^{-tm}\,.$$

Now, fixing the parameters t and β so close to 1 that $t(\frac{1}{m-1}+\beta) > 1$, we see that the RHS is $\geqslant (\mathrm{vol}_S(Y_i \cap X_j))^{1-t(\frac{1}{m-1}+\beta)} W_j^{-tm}$. At last, applying Lemma 8, we complete the proof. ∎

5.7. Regular components

Definition 9. *We call a connected component G of the set $U \setminus Z(f)$* regular, *if G is compactly supported in U and $\mathrm{vol}(G) < \mathrm{vol}(B(1))$. By $N_{\mathtt{reg}}(B_R; f)$ we denote the number of regular connected components G compactly contained in B_R.*

Lemma 10. *There exist constants $q > 1$ and $C < \infty$ such that*

$$\mathcal{E}^*\big\{N^q_{\mathtt{reg}}(B_R; f)\big\} \leqslant C R^{qm}\,.$$

P r o o f. The proof of this lemma follows closely that of Lemma 9. Cover the ball $\bar{B}_R$ by closed balls X_j of radius 1 with centers in B_R keeping the covering number bounded, and put $X = \bigcup_j X_j$. Then $\bar{B}_R \subset X \subset B_{R+1}$. Denote by $\{Y_i\}$ the set of regular nodal domains of f that are contained in B_R. In each domain Y_i choose a point z_i with $\nabla f(z_i) = 0$. In order to apply Lemma 8 with the function $\Phi = |f|^{-t}|\nabla f|^{-tm}$ and with $W_j = \|f\|_{\bar{B}_j, 1+\beta}$, we need to estimate from below the integrals $\displaystyle\int_{X_j \cap Y_i} \Phi \,\mathrm{d}\,\mathrm{vol}$ assuming that $z_i \in X_j$.

Since $\mathrm{vol}(Y_i) < \mathrm{vol}(X_j)$, we note again that $\partial Y_i \cap X_j \neq \varnothing$. Put $\rho_i = \mathrm{dist}(z_i, \partial Y_i \cap X_j) \leqslant 2$, take a closest to z_i point $p \subset \partial Y_i \cap X_j$, and denote $V_i = B(z_i, \rho_i)$. By the construction,

$$\mathrm{vol}(Y_i \cap X_j) \geqslant \mathrm{vol}(V_i \cap X_j) \gtrsim \rho_i^m\,.$$

Since $f(p) = 0$ and $\nabla f(z_i) = 0$, we have

$$|f(x)| \leqslant \rho_i \|f\|_{\bar{B}_j, 1+\beta} = \rho_i W_j, \qquad x \in V_i \cap X_j\,,$$

and

$$|\nabla f(x)| \lesssim \rho_i^\beta \|f\|_{\bar{B}_j, 1+\beta} = \rho_i^\beta W_j, \qquad x \in V_i \cap X_j\,.$$

Hence, on $V_i \cap X_j$,

$$\Phi \gtrsim \rho_i^{-t(1+\beta m)} W_j^{-t(1+m)} \gtrsim (\mathrm{vol}(V_i \cap X_j))^{-t(\beta+\frac{1}{m})}\, W_j^{-t(m+1)}\,,$$

and

$$\int\limits_{Y_i \cap X_j} \Phi \,\mathrm{d}\,\mathrm{vol} \geqslant \int\limits_{V_i \cap X_j} \Phi \,\mathrm{d}\,\mathrm{vol} \gtrsim (\mathrm{vol}(V_i \cap X_j))^{1-t(\beta+\frac{1}{m})}\, W_j^{-t(m+1)}\,.$$

Fixing the parameters t and β so close to 1 that $t(\beta + \frac{1}{m}) > 1$, we get

$$\int\limits_{Y_j \cap X_j} \Phi \,\mathrm{d}\,\mathrm{vol} \gtrsim (\mathrm{vol}(Y_i \cap X_j))^{1-t(\beta+\frac{1}{m})}\, W_j^{-t(m+1)}\,.$$

Finally, Lemma 8 ends the job. ∎

5.8. The moment estimate for the total number of connected components

If the function f is C^1-smooth and 0 is not a critical value, then we can bound the number of connected components γ of $Z(f)$ contained in B_R by the number of connected components G of $U \setminus Z(f)$ compactly contained in B_R. All we need for that is to note that each $\gamma \subset B_R$ is the outer boundary* of some G compactly supported in B_R and no two different connected components $\gamma \subset B_R$ of $Z(f)$ can serve as the outer boundary of the same connected component G of $U \setminus Z(f)$ simultaneously.

Thus, combining the estimate of Lemma 10 with the trivial bound

$$\#\big\{G\colon \bar{G} \subset B_R,\ \mathrm{vol}(G) \geqslant \mathrm{vol}(B(1))\big\} \leqslant R^m\,,$$

we conclude that, for some $q > 1$,

$$\mathcal{E}^*\big\{N(B_R; f)^q\big\} \lesssim R^{mq}\,.$$

If, in addition, f is non-degenerate on $S = \partial B_R$ in the sense that f and $\nabla_S f$ do not vanish simultaneously anywhere on S (due to Lemma 6 this event has probability 1), then, arguing in a similar way, we can estimate the number of connected components of $Z(f)$ intersecting S by the number of connected components of $S \setminus Z(f)$. Thus, the result of Lemma 9 can be viewed as an upper bound for the q-th moment of the number of connected components of $Z(f)$ intersecting S.

We will use these observations several times when referring to Lemmas 10 and 9 as if they were about the connected components of $Z(f)$ rather than about those of $U \setminus Z(f)$ and $S \setminus Z(f)$.

*i.e., the part of the boundary of G that bounds the unbounded connected component of $\mathbb{R}^m \setminus G$ as well

6. Proof of Theorem 1

6.1. Preliminaries

We need several basic notions from the ergodic theory. Suppose $\big(\Omega, \mathfrak{S}, \mathcal{P}\big)$ is a probability space on which $\mathbb{R}^m$ acts by measure-preserving transformations τ_v, $v \in \mathbb{R}^m$. This means that for each $v \in \mathbb{R}^m$, $\tau_v \colon \Omega \to \Omega$ is a $\mathfrak{S}$-measurable transformation, $\tau_u \circ \tau_v = \tau_{u+v}$, $\tau_{-v} = \tau_v^{-1}$, and for each $v \in \mathbb{R}^m$ and each $A \in \mathfrak{S}$, we have $\mathcal{P}(\tau_v A) = \mathcal{P}(A)$.

The following version of Wiener's ergodic theorem suffices for our purposes:

Wiener's ergodic theorem: *Suppose $\big(\Omega, \mathfrak{S}, \mathcal{P}\big)$ is a probability space on which $\mathbb{R}^m$ acts by measure-preserving transformations τ_v, $v \in \mathbb{R}^m$. Suppose that $\Phi \in L^1(\mathcal{P})$, and that the function $(v, \omega) \mapsto \Phi \circ \tau_v$ is measurable with respect to the product σ-algebra $\mathfrak{B}(\mathbb{R}^m) \times \mathfrak{S}$, where $\mathfrak{B}(\mathbb{R}^m)$ is the Borel σ-algebra generated by open sets in $\mathbb{R}^m$. Suppose that $S \subset \mathbb{R}^m$ is a bounded open convex set containing the origin. Then the limit*

$$\lim_{R \to \infty} \frac{1}{\operatorname{vol} S(R)} \int\limits_{S(R)} \Phi(\tau_v \omega) \, \mathrm{d} \operatorname{vol}(v) = \bar{\Phi}(\omega)$$

exists with probability 1 *and in $L^1(\mathcal{P})$. The limiting random variable $\bar{\Phi}$ is τ-invariant (i.e., for each $v \in \mathbb{R}^m$, $\bar{\Phi} \circ \tau_v = \bar{\Phi}$), and does not depend on the choice of the convex set S.*

This is a special case of a theorem proven in Becker [1, Theorems 2 and 3]. Note that Becker's formulation of this theorem deals with rather general increasing families (U_R) of open sets in $\mathbb{R}^m$ satisfying two conditions:

(A) *the Hardy-Littlewood maximal operator associated with the family (U_R) is of weak type* $(1, 1)$,

and

(B) *for each $t \in \mathbb{R}^m$,*

$$\lim_{R \to \infty} \operatorname{vol}((t + U_R) \triangle U_R) / \operatorname{vol}(U_R) = 0 \,,$$

where $\triangle$ denotes the symmetric difference.

In the case when S is the unit ball, condition (A) reduces to the classical Hardy–Littlewood maximal theorem, after which it remains to note that the maximal function associated with the family $S(R)$ is dominated (up to a constant factor) by the one corresponding to the unit ball. The verification of condition (B) is straightforward.

Note that Becker's presentation does not formally contain the claim that the limiting random variable Φ does not depend on the family (U_R) but in our

situation it can be easily established by applying Becker's theorem to a family U_R containing arbitrarily large homothetic images of two bounded convex sets S' and S''.

Next, recall that the action of $\mathbb{R}^m$ is called *ergodic*[⋆] if for every set $A \in \mathfrak{S}$ satisfying $\mathcal{P}((\tau_v A)\triangle A) = 0$, either $\mathcal{P}(A) = 0$, or $\mathcal{P}(A) = 1$. In the ergodic case, the limiting random variable $\bar{\Phi}$ is a constant function. Due to the $L^1(\mathcal{P})$-convergence, the value of this constant equals the expectation of Φ: $\bar{\Phi} = \mathcal{E}\{\Phi\}$.

Let $X \subset C(\mathbb{R}^m)$ be an invariant set of continuous functions (i.e., $G \in X$ implies $G \circ \tau_v \in X$ for all $v \in \mathbb{R}^m$). Let $\mathfrak{S}$ be the minimal σ-algebra on X containing all "intervals" $I(u; a, b) = \{G \in X \colon G(u) \in [a, b)\}$. Let γ be a Gaussian probability measure on $(X, \mathfrak{S})$ meaning that for every finitely many points $u_1, \ldots, u_k \in \mathbb{R}^m$, the push-forward of γ by the mapping $G \mapsto [G(u_1), \ldots, G(u_k)]$ is a (centered) Gaussian, possibly degenerate, measure on $\mathbb{R}^k$. If γ is invariant under the introduced action of $\mathbb{R}^m$ on X, then

$$\mathbb{R}^m \times X \ni (u, G) \mapsto G(u) \in \mathbb{R}$$

is a translation-invariant Gaussian function on the probability space $(X, \mathfrak{S}, \gamma)$ with continuous trajectories and continuous covariance kernel and we can talk about its spectral measure ρ.

Fomin–Grenander–Maruyama theorem: *Suppose that ρ has no atoms. Then the action of $\mathbb{R}^m$ on $(X, \mathfrak{S}, \gamma)$ by translations is ergodic.*

For the reader's convenience, we remind the proof of this theorem[⋆⋆] in AppendixB.

Now, let F be a Gaussian function on $\mathbb{R}^m$ satisfying the assumptions of Theorem 1. By the moment assumption $(\rho 1)$, with probability 1 it is C^{2-}-smooth. Hence, it generates a Gaussian measure γ_F on $(C^1(\mathbb{R}^m), \mathfrak{B}(C^1(\mathbb{R}^m)))$ where $\mathfrak{B}(C^1(\mathbb{R}^m))$ is the Borel σ-algebra generated by open sets in $C^1(\mathbb{R}^m)$. In what follows, it will be convenient to pass from $C^1(\mathbb{R}^m)$ to its subset

$$C^1_*(\mathbb{R}^m) = \{G \in C^1(\mathbb{R}^m) \colon |G| + |\nabla G| \neq 0\},$$

which consists of functions for which 0 is not a critical value. Note that $C^1_*(\mathbb{R}^m)$ is a Borel subset of $C^1(\mathbb{R}^m)$ and, by the first statement in Lemma 6,

$$\gamma_F\big(C^1(\mathbb{R}^m) \setminus C^1_*(\mathbb{R}^m)\big) = 0.$$

[⋆] a.k.a. metric-transitive

[⋆⋆] The full version of the Fomin–Grenander–Maruyama theorem states that the continuity of the spectral measure ρ is *necessary and sufficient* for the ergodicity of the action of $\mathbb{R}^m$ on $(X, \mathfrak{S}, \gamma)$ by translations. We will use (and prove) only the sufficiency part. The proof we present follows the argument for the univariate case given in [13, Section 5.10].

Furthermore, $\mathbb{R}^m$ acts on $\big(C^1_*(\mathbb{R}^m), \mathfrak{B}(C^1_*(\mathbb{R}^m)), \gamma_F\big)$ by translations and, since the distribution of F is translation invariant, the action is measure-preserving. Thus, Wiener's theorem applies in this setting. To apply the Fomin–Grenander–Maruyama theorem, we only need to note that the Borel σ-algebra $\mathfrak{B}(C^1_*(\mathbb{R}^m))$ coincides with the σ-algebra $\mathfrak{S}$ generated by the intervals $I(u; a, b)$ (see Appendix A.1.).

We conclude that

- *under the assumption* $(\rho 1)$ *of Theorem 1, for any random variable* $\Phi \in L^1(\gamma_F)$ *such that the function* $(v, G) \mapsto \Phi(\tau_v G)$ *is measurable, the ergodic averages*

$$(A^S_R \Phi)(G) \stackrel{\text{def}}{=} \frac{1}{\operatorname{vol} S(R)} \int\limits_{S(R)} \Phi(\tau_v G)\, \mathrm{d}\operatorname{vol}(v)$$

converge to a τ*-invariant limit* $\bar\Phi$ *with probability* 1*, as well as in* $L^1(\gamma_F)$*, as* $R \to \infty$*. Moreover, under assumption* $(\rho 2)$*, we have* $\bar\Phi = \mathcal{E}\big\{\Phi\big\}$.

We split the proof of Theorem 1 into two parts: first, we prove the convergence of $(\operatorname{vol} S(R))^{-1} N_S(R; F)$ to a limit ν. Then, assuming condition $(\rho 4)$, we show that this limit is positive.

6.2. Existence of the limit

6.2.1. The sandwich estimate for $N_S(R; G)/\operatorname{vol} S(R)$. Without loss of generality, we assume that $S \supset B(1)$. Then, the integral-geometric Lemma 1 provides us with the "sandwich estimate":

$$\frac{1}{\operatorname{vol} S(R)} \int\limits_{S(R-r)} \frac{N(v, r; G)}{\operatorname{vol} B(r)}\, \mathrm{d}\operatorname{vol}(v) \leqslant \frac{N_S(R; G)}{\operatorname{vol} S(R)} \leqslant \frac{1}{\operatorname{vol} S(R)} \int\limits_{S(R+r)} \frac{N^*(v, r; G)}{\operatorname{vol} B(r)}\, \mathrm{d}\operatorname{vol}(v)\,.$$

The difference $N^*(v, r; G) - N(v, r; G) = N^*(r; \tau_v G) - N(r; \tau_v G)$ is bounded by $\mathfrak{N}_\#(r; \tau_v G)$, where

$$\mathfrak{N}_\#(r; G) \stackrel{\text{def}}{=} \begin{cases} \mathfrak{N}(\partial B(r); G) & \text{if } G \text{ is non-degenerate on } \partial B(r), \\ +\infty & \text{otherwise,} \end{cases}$$

and $\mathfrak{N}(\partial B(r); G)$ is the number of connected components of $\partial B(r) \setminus Z(G)$. Recall that we say that G is non-degenerate on the sphere $\partial B(r)$ if G and $\nabla_{\partial B(r)} G$ do not vanish simultaneously anywhere on $\partial B(r)$.

We introduce the functionals

$$\Phi_r(G) \stackrel{\text{def}}{=} \frac{N(r;G)}{\operatorname{vol} B(r)}, \qquad \Psi_r(G) \stackrel{\text{def}}{=} \frac{\mathfrak{N}_{\#}(r;G)}{\operatorname{vol} B(r)}\,.$$

Then the sandwich estimate takes the form

$$\Big(1-\frac{r}{R}\Big)^m (A^S_{R-r}\Phi_r)(G) \leqslant \frac{N_S(R;G)}{\operatorname{vol} S(R)} \leqslant \Big(1+\frac{r}{R}\Big)^m \big[(A^S_{R+r}\Phi_r)(G) + (A^S_{R+r}\Psi_r)(G)\big]. \quad (6.2.1)$$

6.2.2. Checking measurability. We need to check that, given $r>0$, the functions

$$(v,G)\mapsto \Phi_r(\tau_v G), \quad (v,G)\mapsto \Psi_r(\tau_v G)$$

are measurable with respect to the product σ-algebra $\mathfrak{B}(\mathbb{R}^m)\times\mathfrak{B}(C^1_*(\mathbb{R}^m))$. The function $(v,G)\mapsto\tau_v G$ is a measurable (even continuous) map

$$\big(\mathbb{R}^m\times C^1_*(\mathbb{R}^m), \mathfrak{B}(\mathbb{R}^m)\times\mathfrak{B}(C^1_*(\mathbb{R}^m))\big) \to \big(C^1_*(\mathbb{R}^m), \mathfrak{B}(C^1_*(\mathbb{R}^m))\big)\,.$$

Since the composition of measurable functions is measurable, it remains to show that, given $r>0$, the functions $G\mapsto N(r,G)$ and $G\mapsto\mathfrak{N}_{\#}(r,G)$ are measurable as maps from $\big(C^1_*(\mathbb{R}^m),\mathfrak{B}(C^1_*(\mathbb{R}^m))\big)$ to $\big([0,+\infty],\mathfrak{B}([0,+\infty])\big)$.

The measurability of the map $G\mapsto N(r,G)$ follows from its lower semicontinuity on $C^1_*(\mathbb{R}^m)$. To see that $G\mapsto\mathfrak{N}_{\#}(r,G)$ is measurable, first, consider the set $\operatorname{Degen}(r)$ of functions $G\in C^1_*(\mathbb{R}^m)$ for which there exists a point $x\in\partial B(r)$ such that $\nabla G(x)$ is orthogonal to the tangent space to $\partial B(r)$ at x. This set is closed in $C^1_*(\mathbb{R}^m)$ with respect to the C^1-topology and, therefore, is $\mathfrak{B}(C^1_*(\mathbb{R}^m))$-measurable. On the other hand, our map $G\mapsto\mathfrak{N}_{\#}(r,G)$ is lower semi-continuous on $C^1_*(\mathbb{R}^m)\setminus\operatorname{Degen}(r)$.

6.2.3. Integrability. Next, we note that, for every fixed $r>0$, the functions Φ_r and Ψ_r on $C^1_*(\mathbb{R}^m)$ are γ_F-integrable. This readily follows from Lemma 10 and Lemma 9, correspondingly.

6.2.4. Proof of convergence. By the sandwich estimate ((6.2.1)), for every function $G\in C^1_*(\mathbb{R}^m)$, we have

$$\Big|\frac{N_S(R;G)}{\operatorname{vol} S(R)} - (A^S_R\Phi_r)(G)\Big| \leqslant \Big|\Big(1-\frac{r}{R}\Big)^m (A^S_{R-r}\Phi_r)(G) - (A^S_R\Phi_r)(G)\Big| + \Big|\Big(1+\frac{r}{R}\Big)^m (A^S_{R+r}\Phi_r)(G) - (A^S_R\Phi_r)(G)\Big| + \Big(1+\frac{r}{R}\Big)^m (A^S_{R+r}\Psi_r)(G)\,. \quad (6.2.2)$$

By the Wiener ergodic theorem, there exist τ-invariant functions $\bar\Phi_r$ and $\bar\Psi_r$ such that

$$\lim_{R\to\infty} A_R^S\Phi_r = \bar\Phi_r \quad \text{and} \quad \lim_{R\to\infty} A_R^S\Psi_r = \bar\Psi_r$$

both γ_F-almost everywhere and in $L^1(\gamma_F)$. Letting $R \to \infty$ on both sides of ((6.2.2)), we get

$$\overline{\lim_{R\to\infty}} \left| \frac{N_S(R;G)}{\operatorname{vol} S(R)} - (A_R^S\Phi_r)(G) \right| \leqslant \bar\Psi_r(G) \qquad \text{for } \gamma_F\text{-almost every } G, \tag{6.2.3}$$

and

$$\overline{\lim_{R\to\infty}} \int \left| \frac{N_S(R;G)}{\operatorname{vol} S(R)} - (A_R^S\Phi_r)(G) \right| \mathrm{d}\gamma_F(G) \leqslant \int \bar\Psi_r \,\mathrm{d}\gamma_F = \frac{\mathcal{E}\{\mathfrak{N}_\#(r;F)\}}{\operatorname{vol} B(r)}\,. \tag{6.2.4}$$

By Lemma 9, the RHS of (6.2.4) is $\lesssim r^{-1}$ for $r \geqslant 1$. So taking a sequence $r_k \uparrow \infty$, we observe that

$$\lim_{k\to\infty} \int \bar\Psi_{r_k} \,\mathrm{d}\gamma_F = 0\,.$$

and, consequently,

$$\inf_k \bar\Psi_{r_k} = 0 \qquad \gamma_F\text{-almost everywhere}\,.$$

Since $A_R^S\Phi_r(G)$ converge to $\bar\Phi_r$ for γ_F-almost every G, the second observation together with (6.2.3) imply that $(\operatorname{vol} S(R))^{-1} N_S(R;G)$ is Cauchy for γ_F-almost every G. Similarly, the convergence of $A_R^S\Phi_r(G)$ to $\bar\Phi_r$ in $L^1(\gamma_F)$ together with the first observation and (6.2.4) imply that $(\operatorname{vol} S(R))^{-1} N_S(R;G)$ is Cauchy in $L^1(\gamma_F)$. Thus, the limit

$$\nu \stackrel{\text{def}}{=} \lim_{R\to\infty} \frac{N_S(R;G)}{\operatorname{vol} S(R)}$$

exists γ_F-almost everywhere and in $L^1(\gamma_F)$. It follows from ((6.2.1)) that, for every $r > 0$,

$$\bar\Phi_r \leqslant \nu \leqslant \bar\Phi_r + \bar\Psi_r \qquad \gamma_f - \text{almost everywhere}\,.$$

If, in addition, the action of $\mathbb{R}^m$ on $\big(C^1_*(\mathbb{R}^m), \mathfrak{B}(C^1_*(\mathbb{R}^m)), \gamma_F\big)$ is ergodic, then $\bar\Phi_r = \mathcal{E}\{\Phi_r\}$, $\bar\Psi_r = \mathcal{E}\{\Psi_r\}$. Therefore,

$$\mathcal{E}\big\{\bar\Phi_r\big\} \leqslant \nu \leqslant \mathcal{E}\big\{\bar\Phi_r\big\} + \mathcal{E}\big\{\bar\Psi_r\big\} \qquad \gamma_f - \text{almost everywhere}\,, \tag{6.2.5}$$

whence, for every $r > 0$, γ_F-essential oscillation of ν does not exceed $\mathcal{E}\{\Psi_r\}$. Recalling that $\mathcal{E}\{\Psi_r\} \lesssim r^{-1}$ and letting $r \to \infty$, we see that ν is a (non-random) constant. This completes the proof of convergence in Theorem 1. ∎

6.3. Positivity of ν

It remains to show that condition ($\rho 4$) yields the positivity of the limiting constant ν. We prove that if assumption ($\rho 4$) holds, then $\mathcal{P}\{N(r;F) > 0\} > 0$ when r is sufficiently big. Since the LHS of estimate (6.2.5) can be rewritten as $\nu \geqslant \mathcal{E}\{N(r;F)\}/\operatorname{vol} B(r)$ for each $r > 0$, this will yield the positivity of ν.

6.3.1. A Gaussian lemma

Lemma 11. *Let μ be a compactly supported Hermitian measure with* $\operatorname{spt}(\mu) \subset \operatorname{spt}(\rho)$. *Then for each closed ball $\bar{B} \subset \mathbb{R}^m$ and for each $\varepsilon > 0$,*

$$\mathcal{P}\big\{\|F - \widehat{\mu}\|_{C(\bar{B})} < \varepsilon\big\} > 0\,.$$

P r o o f of Lemma 11. The part of the theory of continuous Gaussian functions developed in Appendix (A.7. and A.12.) yields the statement of the lemma for all measures μ absolutely continuous with respect to ρ with density

$$h \in L^2_{\mathtt{H}}(\rho) \stackrel{\text{def}}{=} \big\{g \in L^2(\rho)\colon g(-x) = \overline{g(x)} \text{ for all } x \in \mathbb{R}^m\big\}\,.$$

In the general case, we can approximate the measure μ in the weak topology by measures $\mathrm{d}\mu = h\,\mathrm{d}\rho$ with $\operatorname{spt}(h)$ contained in a fixed compact neighbourhood of $\operatorname{spt}(\mu)$. Then it remains to recall that for measures supported on a fixed compact set, the weak convergence yields locally uniform convergence of their Fourier integrals. ∎

6.3.2. Proof of the positivity of $\nu(\rho)$. We take a Hermitian compactly supported measure μ with $\operatorname{spt}(\mu) \subset \operatorname{spt}(\rho)$ and a bounded domain $D \subset \mathbb{R}^m$ so that $\widehat{\mu}\big|_{\partial D} < 0$ and $\widehat{\mu}(u_0) > 0$ for some $u_0 \in D$. Choose r so big that $\bar{D} \subset B(r)$. If $\varepsilon > 0$ is sufficiently small, then $G(u_0) > 0$ and $G\big|_{\partial D} < 0$ for every function G satisfying $\|G - \widehat{\mu}\|_{C(\bar{B}(r))} < \varepsilon$. Thus, for every such function G, the zero set $Z(G)$ has at least one connected component in D. Applying Lemma 11, we see that

$$\mathcal{P}\big\{N(r;F) > 0\big\} \geqslant \mathcal{P}\big\{\|F - \widehat{\mu}\|_{C(\bar{B}(r))} < \varepsilon\big\} > 0$$

completing the proof of Theorem 1. ∎

7. Recovering the Function $\bar{\nu}$ by a Double Scaling Limit

The proof of Theorem 2 will rely upon the following lemma, which is of independent interest. Let (f_L) be a tame parametric Gaussian ensemble, that is, an ensemble satisfying the assumptions of Theorem 2. As above, we put $f_{x,L}(u) = f(x + L^{-1}u)$ and

$$K_{x,L}(u,v) = \mathcal{E}\big\{f_{x,L}(u)f_{x,L}(v)\big\} = K_L(x + L^{-1}u, x + L^{-1}v)\,.$$

Till the end of this section, we fix a point $x \in U$ so that

$$\lim_{L\to\infty} K_{x,L}(u,v) = k_x(u-v) \qquad \text{pointwise in } \mathbb{R}^m \times \mathbb{R}^m\,,$$

where the Hermitian positive-definite function k_x is the Fourier integral of a measure ρ_x satisfying assumptions $(\rho 1)$–$(\rho 3)$. By F_x we denote the limiting Gaussian function on $\mathbb{R}^m$, and put $\nu = \bar\nu(x) = \nu(F_x)$.

Lemma 12. *For every $\varepsilon > 0$,*

$$\lim_{R\to\infty} \overline{\lim_{L\to\infty}}\, \mathcal{P}\Big\{\Big|\frac{N(R; f_{x,L})}{\operatorname{vol} B(R)} - \nu\Big| > \varepsilon\Big\} = 0\,.$$

P r o o f. Fix $R > 2$ and $\varepsilon > 0$. Our goal will be to show that, for every t,

$$\overline{\lim_{L\to\infty}}\, \mathcal{P}\Big\{N(R; f_{x,L}) > t\Big\} \leqslant \mathcal{P}\Big\{N(R+1; F_x) > t\Big\}\,,$$
$$\overline{\lim_{L\to\infty}}\, \mathcal{P}\Big\{N(R; f_{x,L}) < t\Big\} \leqslant \mathcal{P}\Big\{N(R-1; F_x) < t\Big\}\,.$$

Applying these inequalities with $t = (\nu+\varepsilon)\operatorname{vol} B(R)$ and $t = (\nu-\varepsilon)\operatorname{vol} B(R)$ respectively, and then combining the results with Theorem 1, we get the conclusion of Lemma 12. The proofs of these two relations are very similar, so we will present only the proof of the first one.

We choose a big constant M and a small constant κ so that the kernels k_x and $K_{x,L}$ (with $L \geqslant L_0$) satisfy the "(M,κ)-conditions" introduced in the beginning of Section 5.1. For the kernel k_x this is possible due to conditions $(\rho 1)$ and $(\rho 3)$. For the scaled kernels $K_{x,L}$ this is possible due to the controllability of (f_L).

Given positive constants A and a, we put

$$E(A,a) = \Big\{g \in C^1(B(R+1.1)) \colon \|g\|_{C^1(\bar B(R+1))} \leqslant A,\ \min_{\bar B(R+1)} \max\{|g|, |\nabla g|\} \geqslant a\Big\}\,.$$

Introduce the events $\Omega'_L = \big\{f_{x,L} \notin E(A,a)\big\}$ and $\Omega'' = \big\{F_x \notin E(A,a)\big\}$. By Lemma 7, the aforementioned "(M,κ)-conditions" imply that, for a given $\delta > 0$, we can make the probabilities of both events less than δ if we choose sufficiently big A and sufficiently small a. We fix a finite $a/(2A)$-net X in $\bar B(R+1)$ and denote by $E' \subset \mathbb{R}^{|X|}$ the set of traces on X of functions $g \in E(A,a)$ satisfying $N(R;g) > t$. This is a bounded subset of $\mathbb{R}^{|X|}$. Note that if $g, h \in E(A,a)$ and $|g-h| < a/2$ on X, then $|g-h| < a$ everywhere on $\bar B(R+1)$, and by Lemma 3 (applied with $\alpha = \beta = a$), $N(R+1;h) \geqslant N(R;g)$.

We fix a function $\varphi \in C_0^\infty(\mathbb{R}^{|X|})$ satisfying $0 \leqslant \varphi \leqslant 1$ everywhere, $\varphi \equiv 1$ on E' and $\varphi \equiv 0$ on $\mathbb{R}^{|X|} \setminus E'_{+a/2}$ (as usual, by E'_{+s} we denote the s-neighbourhood

of E'), and consider the finite dimensional Gaussian vectors $f_{x,L}|_X$ and $F_x|_X$. First, we note that

$$\begin{aligned}\big\{\omega\colon N(R;f_{x,L})>t\big\}&\subset\big\{\omega\colon f_{x,L}|_X\in E',\ f_{x,L}\in E(A,a)\big\}\cup\Omega'_L\\&\subset\big\{\omega\colon\varphi(f_{x,L}|_X)=1\big\}\cup\Omega'_L\,,\end{aligned}$$

whence,

$$\mathcal{P}\big\{N(R;f_{x,L})>t\big\}<\mathcal{E}\{\varphi(f_{x,L}|_X)\}+\delta\,.$$

The pointwise convergence of the scaled kernels $K_{x,L}(u,v)$ to the limiting kernel $k_x(u-v)$ yields*

$$\mathcal{E}\{\varphi(f_{x,L}|_X)\}\overset{L\to\infty}{\to}\mathcal{E}\{\varphi(F_x|_X)\}\leqslant\mathcal{P}\big\{\varphi(F_x|_X)>0\big\}$$

(in the inequality we used that $\varphi\leqslant 1$ everywhere). Now,

$$\begin{aligned}\{\omega\colon\varphi(F_x|_X)>0\}\subset\big\{\omega\colon F_x|_X\in E'_{+a/2}\big\}&\subset\big\{\omega\colon F_x|_X\in E'_{+a/2},\ F_x\in E(A,a)\big\}\cup\Omega''\\&\subset\big\{\omega\colon N(R+1;F_x)>t\big\}\cup\Omega''\,.\end{aligned}$$

In the last step we used that, by our construction, if $F_x\in E(A,a)$ and $F_x|_X\in E'_{+a/2}$, then there is a function $g\in E(A,a)$ such that $N(R,g)>t$, and $|F_x-g|<\frac{1}{2}a$ on X, whence, $N(R+1;F_x)\geqslant N(R;g)>t$. Hence,

$$\mathcal{P}\{\varphi(F_x|_X)>0\}<\mathcal{P}\big\{N(R+1;F_x)>t\big\}+\delta\,.$$

Thus, for sufficiently large L, we have

$$\begin{aligned}\mathcal{P}\big\{N(R;f_{x,L})>t\big\}&<\mathcal{E}\{\varphi(f_{x,L}|_X)\}+\delta\\&<\mathcal{P}\big\{\varphi(F_x|_X)>0\big\}+2\delta<\mathcal{P}\big\{N(R+1;F_x)>t\big\}+3\delta\,,\end{aligned}$$

completing the argument. ∎

*If ξ_L are Gaussian n-dimensional vectors and the entries of the covariance matrices K_L of ξ_L converge to the entries of the covariance matrix K of ξ, then

$$\begin{aligned}\mathcal{E}\{\varphi(\xi_L)\}=\mathcal{E}\Big\{\int\limits_{\mathbb{R}^n}\widehat{\varphi}(\lambda)e^{2\pi\mathrm{i}\lambda\cdot\xi_L}\,\mathrm{d}\lambda\Big\}&=\int\limits_{\mathbb{R}^n}\widehat{\varphi}(\lambda)\mathcal{E}\Big\{e^{2\pi\mathrm{i}\lambda\cdot\xi_L}\Big\}\,\mathrm{d}\lambda\\&=\int\limits_{\mathbb{R}^n}\widehat{\varphi}(\lambda)e^{-\pi K_L\lambda\cdot\lambda}\,\mathrm{d}\lambda\to\int\limits_{\mathbb{R}^n}\widehat{\varphi}(\lambda)e^{-\pi K\lambda\cdot\lambda}\,\mathrm{d}\lambda=\mathcal{E}\{\varphi(\xi)\},\end{aligned}$$

where the convergence holds by the dominated convergence theorem.

8. Proof of Theorem 2

It remains to tie the ends together. Let (f_L) be a tame parametric Gaussian ensemble on an open set $U \subset \mathbb{R}^m$. This implies that,

- *for every compact set $Q \subset U$, there exist constants $M < \infty$ and $\kappa > 0$ such that the covariance kernels of the functions $f_{x,L}$ on $\bar{B}(R+1)$ satisfy the (M,κ)-conditions from Sec. 5.1 whenever $x \in Q$, $R > 0$, and $L \geqslant L_0(Q,R)$.*

Fix a Borel set $U' \subset U$ of full volume on which the scaled functions $f_{x,L}$ have translation invariant limits F_x. Then, by Appendix A.12,

- *the covariance kernels $k_x(u-v)$ of the limiting functions F_x satisfy the (M,κ)-conditions whenever $x \in Q \cap U'$.*

8.1. $\bar{\nu} \in L^\infty_{\mathrm{loc}}(U)$

First, we show that $\bar{\nu}$ is locally uniformly bounded on U' and then that it is measurable.

8.1.1. Boundedness of $\bar{\nu}$. Recall that

$$\bar{\nu}(x) = \lim_{R\to\infty} \frac{\mathcal{E}\{N(R;F_x)\}}{\operatorname{vol} B(R)}\,, \qquad x \in U'\,.$$

Given any compact set $Q \subset U$, Lemma 10 implies that, for every $x \in U' \cap Q$, we have $\mathcal{E}\{N(R;F_x)\} \leqslant C(Q)\operatorname{vol} B(R)$. Thus, the function $\bar{\nu}$ is locally bounded on U'. ∎

8.1.2. Measurability of $\bar{\nu}$. Put

$$\nu_{R,L}(x,\omega) = \frac{N(R;f_{x,L})}{\operatorname{vol} B(R)}\,.$$

The function $\nu_{R,L}$ is defined on the set $U_{-(R+1)/L} \times \Omega'$, where $U_{-r} = \{x \in U\colon \operatorname{dist}(x,\partial U) > r\}$ and $\Omega' = \{\omega \in \Omega\colon f_L \in C^1_*(U)\}$, $\mathcal{P}(\Omega \setminus \Omega') = 0$. It is measurable as a composition of a lower semicontinuous mapping

$$C^1_*(B(R+1)) \ni g \mapsto \frac{N(R;g)}{\operatorname{vol} B(R)} \in \mathbb{R}\,,$$

a continuous mapping

$$U_{-(R+1)/L} \times C^1_*(U) \ni (x,g) \mapsto g_{x,L}\big|_{B(R+1)} \in C^1_*(B(R+1))\,,$$

and a measurable mapping

$$U_{-(R+1)/L} \times \Omega' \ni (x,\omega) \mapsto (x,f_L) \in U_{-(R+1)/L} \times C^1_*(U)\,.$$

Fix $x \in U'$. By Lemma 10, there exist $q > 1$ and $C < \infty$ such that

$$\int\limits_{\Omega'} \nu_{R,L}^q \, \mathrm{d}\mathcal{P} < C$$

for all sufficiently large L. Given $\varepsilon > 0$, put

$$\Omega_\varepsilon(R, L, x) = \bigl\{\omega \in \Omega' \colon |\nu_{R,L}(x, \omega) - \bar{\nu}(x)| > \varepsilon\bigr\}\,.$$

Then

$$\begin{aligned}\int\limits_{\Omega_\varepsilon} |\nu_{R,L}(x, \omega) - \bar{\nu}(x)| \, \mathrm{d}\mathcal{P}(\omega) &\leqslant \int\limits_{\Omega_\varepsilon} \nu_{R,L} \, \mathrm{d}\mathcal{P} + \bar{\nu}(x)\mathcal{P}\{\Omega_\varepsilon\} \\ &\leqslant (\mathcal{P}\{\Omega_\varepsilon\})^{1-\frac{1}{q}} \Bigl(\int\limits_{\Omega_\varepsilon} \nu_{R,L}^q \, \mathrm{d}\mathcal{P}\Bigr)^{\frac{1}{q}} + \bar{\nu}(x)\mathcal{P}\{\Omega_\varepsilon\} \leqslant C(\mathcal{P}\{\Omega_\varepsilon\})^{1-\frac{1}{q}}\,.\end{aligned}$$

Therefore,

$$\Bigl|\int\limits_{\Omega'} \nu_{R,L}(x, \omega) \, \mathrm{d}\mathcal{P}(\omega) - \bar{\nu}(x)\Bigr| \leqslant \int\limits_{\Omega'} |\nu_{R,L}(x, \omega) - \bar{\nu}(x)| \, \mathrm{d}\mathcal{P}(\omega) \leqslant \varepsilon + C(\mathcal{P}\{\Omega_\varepsilon\})^{1-\frac{1}{q}}$$

and

$$\lim_{R\to\infty} \overline{\lim_{L\to\infty}} \Bigl|\int\limits_{\Omega'} \nu_{R,L}(x, \omega) \, \mathrm{d}\mathcal{P}(\omega) - \bar{\nu}(x)\Bigr| \leqslant \varepsilon + C \lim_{R\to\infty} \overline{\lim_{L\to\infty}} (\mathcal{P}\{\Omega_\varepsilon\})^{1-\frac{1}{q}}\,.$$

By Lemma 12, the double limit on the RHS vanishes, so

$$\lim_{R\to\infty} \overline{\lim_{L\to\infty}} \Bigl|\int\limits_{\Omega'} \nu_{R,L}(x, \omega) \, \mathrm{d}\mathcal{P}(\omega) - \bar{\nu}(x)\Bigr| = 0\,.$$

It follows from here that the function $\bar{\nu}(x)$ can be represented as, say,

$$\bar{\nu}(x) = \lim_{R\to\infty} \overline{\lim_{L\to\infty}} \int\limits_{\Omega'} \nu_{R,L}(x, \omega) \, \mathrm{d}\mathcal{P}(\omega)\,.$$

Since the functions $\nu_{R,L}(x, \omega)$ are non-negative and measurable in (x, ω), their integrals with respect to ω over a fixed set Ω' are also measurable as functions of $x \in U'$. Thus, the function $\bar{\nu}$ is also measurable. ∎

8.2. Towards the proof of Theorem 2: another sandwich estimate

Without loss of generality we assume that the continuous compactly supported function φ in the assumptions of Theorem 2 is non-negative. We denote $Q = \operatorname{spt}(\varphi)$. Fix $\delta > 0$ such that $Q_{+4\delta} \subset U$ and put $Q_1 = Q_{+\delta}$, $Q_2 = Q_{+2\delta}$. For $x \in Q_1$, let

$$\varphi_-(x) = \min_{\bar{B}(x,\delta)} \varphi, \quad \varphi_+(x) = \max_{\bar{B}(x,\delta)} \varphi\,.$$

Note that

$$\varphi_-(x) \leqslant \varphi(y) \leqslant \varphi_+(x)$$

whenever $x \in Q_1$, $y \in B(x,\delta)$.

Fix the parameters D, R, L so that $1 < D < R < \delta L$. We have

$$L^{-m} \int\limits_U \varphi \,\mathrm{d}n_L = \int\limits_{Q_1} \Big(\int\limits_{B(x,R/L)} \frac{\varphi(y)\,\mathrm{d}n_L(y)}{\operatorname{vol} B(R)} \Big) \mathrm{d}\operatorname{vol}(x)\,,$$

whence,

$$\int\limits_{Q_1} \varphi_-(x)\, \frac{n_L(B(x,R/L))}{\operatorname{vol} B(R)}\, \mathrm{d}\operatorname{vol}(x) \leqslant L^{-m} \int\limits_U \varphi \,\mathrm{d}n_L \leqslant \int\limits_{Q_1} \varphi_+(x)\, \frac{n_L(B(x,R/L))}{\operatorname{vol} B(R)}\, \mathrm{d}\operatorname{vol}(x)\,. \quad (8.2.1)$$

Since the total n_L-mass of each connected component of $Z(f_L)$ equals 1, the LHS of (8.2.1) cannot be less than

$$\int\limits_{Q_1} \varphi_-(x) \nu_{R,L}(x,\omega)\, \mathrm{d}\operatorname{vol}(x),$$

where, as above, $\nu_{R,L}(x,\omega) = (\operatorname{vol} B(R))^{-1} N(R; f_{x,L})$.

In order to estimate the RHS of (8.2.1), we cover Q_2 by $\simeq \operatorname{vol}(Q_2)\big(\frac{L}{D}\big)^m$ open balls of diameter D/L. Denote by $\{S_j\}$ the collection of boundary spheres of these balls. Due to the second statement in Lemma 6, with probability 1 there is no point x such that, for some j, $x \in S_j \cap Z(f_L)$ and $\nabla_{S_j} f_L(x) = 0$. Under this non-degeneracy condition, the number of connected components of $Z(f_L)$ that intersect the sphere S_j is bounded by the number $\mathfrak{N}(S_j; f_L)$ of connected components of $S_j \setminus Z(f_L)$. Denote by n_L^* the part of the component counting measure n_L supported on the connected components of $Z(f_L)$ intersecting at least one of the spheres S_j. Since every other component of $Z(f_L)$ intersecting a

ball $B(x, R/L)$ centered at $x \in Q_1$ is contained in $B(x, (R+D)/L)$, we see that the RHS of (8.2.1) does not exceed

$$\Big(\frac{R+D}{R}\Big)^m \int\limits_{Q_1} \varphi_+(x)\nu_{R+D,L}(x,\omega)\,\mathrm{d}\,\mathrm{vol}(x) + \int\limits_{Q_1} \varphi_+(x)\,\frac{n_L^*(B(x,R/L))}{\mathrm{vol}\,B(R)}\,\mathrm{d}\,\mathrm{vol}(x)\,.$$

By Fubini, the second integral on the RHS is bounded by $(\max_U \varphi)L^{-m}n_L^*(Q_2)$. In turn, $n_L^*(Q_2) \leqslant \sum_j \mathfrak{N}(S_j; f_L)$ with probability 1. Thus, for almost every ω, we have

$$\begin{aligned}\int\limits_{Q_1} \varphi_-(x)\nu_{R,L}(x,\omega)\,\mathrm{d}\,\mathrm{vol}(x) &\leqslant L^{-m}\int\limits_U \varphi\,\mathrm{d}n_L \\ &\leqslant \big(1+\tfrac{D}{R}\big)^m \int\limits_{Q_1} \varphi_+(x)\nu_{R+D,L}(x,\omega)\,\mathrm{d}\,\mathrm{vol}(x) + (\max_U \varphi)\,L^{-m}\sum\nolimits_j \mathfrak{N}(S_j; f_L)\,.\end{aligned}$$

8.3. Completing the proof of Theorem 2

To juxtapose the integrals

$$L^{-m}\int\limits_U \varphi\,\mathrm{d}n_L \quad \text{and} \quad \int\limits_U \varphi\bar{\nu}\,\mathrm{d}\,\mathrm{vol}\,,$$

we note that, since pointwise $\varphi_+ \leqslant \varphi + \omega_\varphi(\delta)$, where ω_φ is the modulus of continuity of φ, we have

$$\begin{aligned}\int\limits_U \varphi\bar{\nu}\,\mathrm{d}\,\mathrm{vol} &\geqslant \int\limits_U \varphi_+\bar{\nu}\,\mathrm{d}\,\mathrm{vol} - \omega_\varphi(\delta)\|\bar{\nu}\|_{L^\infty(Q_1)}\,\mathrm{vol}(Q_1) \geqslant \big(1+\tfrac{D}{R}\big)^m \int\limits_U \varphi_+\bar{\nu}\,\mathrm{d}\,\mathrm{vol} \\ &\quad - \big[\big(1+\tfrac{D}{R}\big)^m - 1\big](\max_U \varphi)\|\bar{\nu}\|_{L^\infty(Q_1)}\,\mathrm{vol}(Q_1) - \omega_\varphi(\delta)\|\bar{\nu}\|_{L^\infty(Q_1)}\,\mathrm{vol}(Q_1)\,,\end{aligned}$$

whence, for almost every ω,

$$\begin{aligned}L^{-m}\int\limits_U \varphi\,\mathrm{d}n_L - \int\limits_U \varphi\bar{\nu}\,\mathrm{d}\,\mathrm{vol} &\leqslant 2^m\,(\max_U \varphi)\int\limits_{Q_1} |\nu_{R+D,L}(x,\omega) - \bar{\nu}(x)|\,\mathrm{d}\,\mathrm{vol}(x) \\ &\quad + \big[\big(1+\tfrac{D}{R}\big)^m - 1\big](\max_U \varphi)\|\bar{\nu}\|_{L^\infty(Q_1)}\,\mathrm{vol}(Q_1) + \omega_\varphi(\delta)\|\bar{\nu}\|_{L^\infty(Q_1)}\,\mathrm{vol}(Q_1) \\ &\quad + (\max_U \varphi)\,L^{-m}\sum\nolimits_j \mathfrak{N}(S_j; f_L)\,.\end{aligned}$$

The matching lower bound is similar but somewhat simpler: for almost every ω, we have

$$L^{-m}\int\limits_U \varphi\,\mathrm{d}n_L - \int\limits_U \varphi\bar\nu\,\mathrm{d}\,\mathrm{vol} \geqslant -(\max_U \varphi)\int\limits_{Q_1} |\nu_{R,L}(x,\omega)-\bar\nu(x)|\,\mathrm{d}\,\mathrm{vol}(x) - \omega_\varphi(\delta)\|\bar\nu\|_{L^\infty(Q_1)}\,\mathrm{vol}(Q_1)\,.$$

Gathering the upper and lower bounds and taking the upper expectation, we obtain

$$\begin{aligned}\mathcal{E}^*\Big|L^{-m}\int\limits_U \varphi\,\mathrm{d}n_L - \int\limits_U \varphi\bar\nu\,\mathrm{d}\,\mathrm{vol}\Big| &\leqslant 2^m(\max_U\varphi)\int\limits_{Q_1}\mathcal{E}\{|\nu_{R+D,L}(x)-\bar\nu(x)|+|\nu_{R,L}(x)-\bar\nu(x)|\}\,\mathrm{d}\,\mathrm{vol}(x)\\ &\quad+(\max_U\varphi)\,L^{-m}\sum\nolimits_j \mathcal{E}^*\big\{\mathfrak{N}(S_j;f_L)\big\}\\ &\quad+(\max_U\varphi)\,\|\bar\nu\|_{L^\infty(Q_1)}\,\mathrm{vol}(Q_1)\big[\big(1+\tfrac{D}{R}\big)^m-1\big]\\ &\quad+2\omega_\varphi(\delta)\|\bar\nu\|_{L^\infty(Q_1)}\,\mathrm{vol}(Q_1)\,.\end{aligned}$$

It remains to estimate the terms on the RHS.

Fix $\varepsilon>0$ and choose δ so small that $\omega_\varphi(\delta)<\varepsilon$. This takes care of the last term on the RHS. To treat the second term we use Lemma 9, which yields $\mathcal{E}^*\big\{\mathfrak{N}(S_j;f_L)\big\}\lesssim D^{m-1}$ uniformly in j. Therefore,

$$L^{-m}\sum\nolimits_j \mathcal{E}^*\big\{\mathfrak{N}(S_j;f_L)\big\}\lesssim L^{-m}\ \mathrm{vol}(Q_2)(L/D)^m\,D^{m-1}\lesssim D^{-1}\,\mathrm{vol}(Q_2)\,.$$

Let $U'\subset U$ be a Borel subset of full volume on which the scaled functions $f_{x,L}$ have translation invariant limits. The functions $\bar\nu$ and $\mathcal{E}\big\{\nu_{R,L}\big\}$ are locally uniformly bounded on U' by a constant independent of R and L. Let

$$\eta_R(x)\stackrel{\mathrm{def}}{=}\overline{\lim_{L\to\infty}}\,\mathcal{E}\big|\nu_{R,L}(x)-\bar\nu(x)\big|\,.$$

The function η_R is uniformly bounded on $Q_1\cap U'$ by a constant independent of R. Then, applying Fatou lemma, we obtain

$$\overline{\lim_{L\to\infty}}\ \mathcal{E}^*\Big|L^{-m}\int\limits_U \varphi\,\mathrm{d}n_L - \int_U \varphi\bar{\nu}\,\mathrm{d}\,\mathrm{vol}\Big|$$
$$\leqslant C(\varphi,Q)\Big(\int\limits_{Q_1}\big[\eta_{R+D}(x)+\eta_R(x)\big]\,\mathrm{d}\,\mathrm{vol}(x)+\|\bar{\nu}\|_{L^\infty(Q_1)}\big[\big(1+\tfrac{D}{R}\big)^m-1+\varepsilon\big]+D^{-1}\Big).$$

For any $x\in U'$, we have $\eta_R(x)\to 0$ as $R\to\infty$. Using the dominated convergence theorem, we get

$$\overline{\lim_{L\to\infty}}\ \mathcal{E}^*\Big|L^{-m}\int\limits_U \varphi\,\mathrm{d}n_L - \int\limits_U \varphi\bar{\nu}\,\mathrm{d}\,\mathrm{vol}\Big| \leqslant C(\varphi,Q)\big(\varepsilon\|\bar{\nu}\|_{L^\infty(Q_1)}+D^{-1}\big).$$

Letting $\varepsilon\to 0$ and $D\to\infty$, we finish off the proof of Theorem 2. ∎

9. The Manifold Case. Proof of Theorem 3

9.1. Smooth Gaussian functions and their covariance kernels under C^2-changes of variable

Suppose that U, V are open subsets of $\mathbb{R}^m$ and $\psi\colon V\to U$ is a C^2-diffeomorphism. Suppose that $f\colon U\to\mathbb{R}$ is a continuous Gaussian function on U with a $C^{2,2}$ covariance kernel $K(x,y)$. Then $f\circ\psi$ is a continuous Gaussian function on V with the reproducing kernel $\widetilde{K}(x,y)=K(\psi(x),\psi(y))$. Note that for every pair of the multi-indices α,β, the mixed partial derivative $\partial_x^\alpha\,\partial_y^\beta\,\widetilde{K}(x,y)$ is a linear combination of finitely many expressions of the kind

$$\big[\partial_x^{\alpha'}\,\partial_y^{\beta'}\,K\big](\psi(x),\psi(y))\,Q_{\alpha',\beta'}\,,$$

where α', β' are multi-indices with $1\leqslant|\alpha'|\leqslant|\alpha|$, $1\leqslant|\beta'|\leqslant|\beta|$, and $Q_{\alpha',\beta'}$ is a certain polynomial expression of partial derivatives of order at most $\max(|\alpha|,|\beta|)$ of coordinate functions of ψ. In particular, if K is $C^{k,k}$-smooth, then so is $\widetilde{K}$. Since the maxima of higher order derivatives in the definition of the norm $\|K\|_{L,Q,k}$ are multiplied by higher negative powers of L, we conclude that for every compact $Q\subset V$ and $L\geqslant 1$,

$$\|\widetilde{K}\|_{L,Q,k}\leqslant C(\psi,Q,k)\,\|K\|_{L,\psi(Q),k}\,,$$

where $C(\psi,Q,k)$ is some factor depending on $\max\limits_{|\gamma|\leqslant k}\max\limits_{Q}|\partial^\gamma\psi|$.

Next, let C_x^K be the matrix with the entries $C_x^K(i,j)=\partial_{x_i}\,\partial_{y_j}\,K(x,y)$. Then

$$\det C_x^{\widetilde{K}}=(\det(D\psi)(x))^2\,\det C_{\psi(x)}^K\,.$$

One immediate consequence of these observations is that

- *the local controllability of K can be verified after any C^2-change of variables ψ, and moreover, the corresponding constants at x will change only by bounded factors depending on the first and second derivatives of ψ and ψ^{-1} at x and $\psi(x)$ respectively.*

Now, let us see what the C^2-change of variable ψ does to translation invariant scaling limits. Let $z = \psi(z') \in U$. Assume that we have a sequence of kernels K_L such that the corresponding scaled kernels $K_{z,L}(u, v) = K_L(z + L^{-1}u, z + L^{-1}v)$ converge to $k(u - v)$, where k is a continuous function on $\mathbb{R}^m$. Assume that for some $r > 0$, there is a closed ball $\bar{B} = \bar{B}(z, r) \subset U$ such that

$$\sup_{L>1} L^{-1} \max_{\bar{B}\times\bar{B}} (|\nabla_x K_L| + |\nabla_y K_L|) = \mathfrak{M} < \infty\,.$$

Let $u', v' \in \mathbb{R}^m$. Then, for sufficiently large L, we have

$$\left|\psi(z' + L^{-1}u') - \psi(z') - \tfrac{1}{L}(D\psi)(z')u'\right| \leqslant C(\psi)\, L^{-2}|u'|^2\,,$$

and similarly for v', where the constant $C(\psi)$ depends only on the bounds for the second partial derivatives of ψ in an arbitrarily small (but fixed) neighbourhood of z'. Moreover, if u' and v' are fixed and L is large, then the points

$$\psi(z' + \tfrac{1}{L}u'), \quad z + \tfrac{1}{L}(D\psi)(z')u'\,,$$

together with similar two points taken with v' instead of u', belong to the ball B. So we obtain

$$\begin{aligned}\left|\widetilde{K}(z' + \tfrac{1}{L}u', z' + \tfrac{1}{L}v') - K(z + \tfrac{1}{L}(D\psi)(z')u', z + \tfrac{1}{L}(D\psi)(z')v')\right| \\ \leqslant L\,\mathfrak{M}\,C(\psi)L^{-2}(|u'|^2 + |v'|^2) \to 0, \qquad \text{as } L \to \infty\,.\end{aligned}$$

Since $K(z + \frac{1}{L}(D\psi)(z')u', z + \frac{1}{L}(D\psi)(z')v')$ converge to $k((D\psi)(z')(u' - v'))$, we conclude that $\widetilde{K}(z' + L^{-1}u', z' + L^{-1}v')$ converge to $\widetilde{k}(u' - v')$, where $\widetilde{k}(u') = k((D\psi)(z')(u'))$.

Since a non-degenerate linear change of variable on the space side corresponds to a non-degenerate linear change of variable and renormalization on the Fourier side, the spectral measures ρ and $\widetilde{\rho}$, corresponding to k and $\widetilde{k}$, respectively, do or do not have atoms simultaneously. This shows that *the Gaussian parametric ensembles f_L on U and $\widetilde{f}_L = f_L \circ \psi$ on $V = \psi^{-1}U$ are or aren't tame simultaneously.* Finally, the corresponding limiting Gaussian functions F_z and $\widetilde{F}_{z'}$ are related by $\widetilde{F}_{z'} = F_z \circ (D\psi)(z')$, whence, $\bar{\nu}_{\widetilde{F}_{z'}}(z') = |\det(D\psi)(z')|\,\bar{\nu}_{F_z}(z)$.

9.2. Possibility to verify tameness in charts

From the previous discussion it becomes clear why it suffices to check that $f_L \circ \pi_\alpha$ is tame on U_α for some atlas $\mathbb{A} = (U_\alpha, \pi_\alpha)$ to be sure that $f_L \circ \pi$ is tame on U for any chart $\pi\colon U \to X$. Indeed, take any compact $Q \subset \pi(U)$ and cover it by a finite union of open charts $\bigcup_j \pi_{\alpha_j}(U_{\alpha_j})$. Then we can choose compact sets $Q_j \subset Q \cap \pi_{\alpha_j}(U_{\alpha_j})$ so that $\bigcup_j Q_j = Q$. However, on each Q_j the computations of all relevant quantities in the charts (U, π) and $(U_{\alpha_j}, \pi_{\alpha_j})$ give essentially the same results (up to bounded factors) because all partial derivatives of order 1 and 2 of the transition mappings $\pi_{\alpha_j}^{-1} \circ \pi$ and $\pi^{-1} \circ \pi_{\alpha_j}$ are bounded on $\pi^{-1}(Q_j)$ and $\pi_{\alpha_j}^{-1}(Q_j)$, respectively.

If the atlas $\mathbb{A}$ has uniformly bounded distortions, our observations show that for every point $x \in X$, *all* computations in *all* charts (U, π) of $\mathbb{A}$ such that $x \in \pi(U)$ yield essentially the same results. Thus, for every point $x \in X$, we can compute the relevant quantities in its own chart from $\mathbb{A}$ (the most convenient one) without affecting the existence of uniform bounds for them, but, of course, affecting the best possible values of those bounds.

9.3. Completing the proof of Theorem 3

Take two charts $\pi_1\colon U_1 \to X$ and $\pi_2\colon U_2 \to X$ and consider the corresponding Gaussian parametric ensembles $f_{1,L} = f_L \circ \pi_1$ and $f_{2,L} = f_L \circ \pi_2$ on U_1 and U_2 respectively. For every $x \in \pi(U_1) \cap \pi(U_2) \subset X$, we have

$$\bar\nu_1(\pi_1^{-1}(x)) = \bigl|\det\bigl([D(\pi_2^{-1}\pi_1)](\pi_1^{-1}(x))\bigr)\bigr|\, \bar\nu_2(\pi_2^{-1}(x))$$

in the sense that if one side is defined, then so is the other and the equality holds. Therefore, the push-forwards $(\pi_1)_*(\bar\nu_1 \,\mathrm{d}\,\mathrm{vol})$ and $(\pi_2)_*(\bar\nu_2 \,\mathrm{d}\,\mathrm{vol})$ coincide on $\pi(U_1) \cap \pi(U_2)$, which allows us to define a Borel measure n_∞ on X unambiguously and to justify the formula for its density with respect to any volume vol_X on X compatible with the smooth structure.

The only thing that remains to do to establish Theorem 3 as stated, is to show that

$$\lim_{L\to\infty} \mathcal{E}^*\Bigl\{\Bigl|L^{-m}\int_X \varphi \,\mathrm{d}n_L - \int_X \varphi \,\mathrm{d}n_\infty\Bigr|\Bigr\} = 0\,.$$

The standard partition of unity argument allows us to reduce the problem to the case when the support of the test function φ is contained in one chart $\pi(U)$. Hence, the desired result would be an immediate consequence of Theorem 2 applied to the pull-back measures $\pi^* n_L$ and the test-function $\varphi \circ \pi$, if not for one minor nuisance: the pull-back to U of a component counting measure of (f_L) on X by the chart mapping π may fail to be a component counting measure of $f_L \circ \pi$

on U because the connected components of $Z(f_L)$ on X that stretch outside $\pi(U)$ may get truncated or split into several pieces when mapped to U by π^{-1}. So the pull-back $\pi^* n_L$ may have mass less than 1 on some connected components of $f_L \circ \pi$ that stretch to the boundary of U. We circumvent this difficulty by noticing that the closed support $\operatorname{spt}(\varphi \circ \pi)$ is contained in U. Thus, if we "beef up" the measure of each "defective" component γ by adding an appropriate positive multiple of a point mass at any point $u \in \gamma \setminus \operatorname{spt}(\varphi \circ \pi)$, the pull-back $\pi^* n_L$ will turn into a component counting measure n'_L but the total integral of $\varphi \circ \pi$ will not be affected in any way. Now we can just apply Theorem 2 to n'_L instead of $\pi^* n_L$ and reach the desired conclusion. ∎

Appendices

A. Smooth Gaussian Functions

In this appendix, we collect well-known facts about smooth Gaussian functions, which have been used throughout this paper. Our smooth Gaussian functions will be defined on open subsets of $\mathbb{R}^m$. For a topological space X, by $\mathfrak{B}(X)$ we denote the Borel σ-algebra generated by all open subsets of X. As everywhere else in the paper, all Hilbert spaces are real and separable and all Gaussian random variables have zero mean.

A.1. The space $C^k(V)$

Let $V \subset \mathbb{R}^m$ be an open set. For $k \in \mathbb{Z}_+$, we denote by $C^k(V)$ the space of C^k-smooth functions on V. The topology in $C^k(V)$ is generated by the seminorms*

$$\|g\|_{Q,k} = \max_{Q} \max_{|\alpha| \leqslant k} \left|\partial^\alpha g\right|$$

where Q runs over all compact subsets of V. If Q_n is an increasing sequence of compact subsets of V such that every compact set $K \subset V$ is contained in each Q_n with $n \geqslant n_0(K)$, then the *countable* family of the seminorms $\|g\|_{Q_n,k}$, $n = 1, 2, \dots$, gives the same topology, so $C^k(V)$ is metrizable. Since it is separable as well, every open set in $C^k(V)$ can be written as a countable union of "standard neighbourhoods"

$$\mathbb{B}(Q, g_0, \varepsilon) = \big\{g \in C^k(V) \colon \|g - g_0\|_{Q,k} < \varepsilon\big\}.$$

We will need two simple lemmas.

*The reader should be aware that the same notation was used in the main text for the seminorm in $C^{k,k}(V \times V)$. This should not lead to a confusion because the functions measured in these seminorms have different numbers of variables.

Lemma A.1. *The Borel σ-algebra $\mathfrak{B} = \mathfrak{B}(C^k(V))$ coincides with the least σ-algebra on $C^k(V)$ containing all "intervals" $I(x;a,b) = \{g \in C^k(V)\colon a \leqslant g(x) < b\}$, i.e., $\mathfrak{B}$ is generated by point evaluations $g \mapsto g(x)$.*

P r o o f of Lemma A.1. Denote by $\mathfrak{B}'$ the least σ-algebra on $C^k(V)$ containing all intervals $I(x;a,b)$. We need to show that the σ-algebras $\mathfrak{B}$ and $\mathfrak{B}'$ coincide. Since the mapping $C^k(V) \ni g \mapsto g(x) \in \mathbb{R}$ is continuous and, thereby, measurable, every interval $I(x;a,b)$ is Borel, that is $\mathfrak{B}' \subset \mathfrak{B}$.

To show that $\mathfrak{B} \subset \mathfrak{B}'$, it suffices to check that every standard neighourhood $\mathbb{B}(Q, g_0, \varepsilon)$ belongs to $\mathfrak{B}'$, or, which is the same, that the mapping $C^k(V) \ni g \mapsto \|g - g\|_{Q,k}$ is $\mathfrak{B}$-measurable. Since for every fixed $x \in V$ and every multiindex α with $|\alpha| \leqslant k$, the mapping $g \mapsto \partial^\alpha g(x)$ can be represented as a pointwise (on $C^k(V)$) limit of finite linear combinations of point evaluations, it is measurable as well. It remains to note that

$$\|g - g_0\|_{Q,k} = \sup_{x \in Q'} \max_{|\alpha| \leqslant k} \bigl|\partial^\alpha g(x) - \partial^\alpha g_0(x)\bigr|,$$

where Q' is any countable dense (in Q) subset of Q. ∎

Lemma A.2. *$C^k(V)$ is a Borel subset of $C(V)$.*

P r o o f of Lemma A.2. Take any function $\varphi_1 \in C_0^\infty(B)$, where B is the unit ball in $\mathbb{R}^m$, put $\varphi_j = j^m \varphi(jx)$ and consider the sequence of continuous mappings

$$C(V) \ni g \mapsto g * \varphi_j \in C^k(V_{-1/j}).$$

Note that $g \in C^k(V)$ if and only if $g * \varphi_j$ converge in C^k uniformly on every compact set $Q \subset V$. Taking a countable exhaustion Q_n of V and choosing $j(n)$ so that $Q_n \subset V_{-1/j(n)}$, we get the representation

$$C^k(V) = \bigcap_{q \geqslant 1} \bigcap_{n \geqslant 1} \bigcup_{j > j(n)} \bigcap_{s', s'' > j} \bigl\{g \in C(V)\colon \|g * \varphi_{s'} - g * \varphi_{s''}\|_{j, Q_n} < \tfrac{1}{q}\bigr\}.$$

Clearly, the RHS is Borel in $C(V)$ (since each "basic set" on the RHS is open in $C(V)$). ∎

A.2. The definition and basic properties of C^k-smooth Gaussian functions

Definition A.1. *Let $(\Omega, \mathfrak{S}, \mathcal{P})$ be a probability space. The function $f\colon V \times \Omega \to \mathbb{R}$ is a Gaussian function on V if*

(i) *for each $x \in V$, the mapping $\omega \mapsto f(x,\omega)$ is measurable as a mapping from $\bigl(\Omega, \mathfrak{S}\bigr)$ to $\bigl(\mathbb{R}, \mathfrak{B}(\mathbb{R})\bigr)$;*

(ii) *for each finite set of points* $x_1, \dots, x_n \in V$ *and for each* $c_1, \dots, c_n \in \mathbb{R}$, *the sum* $\sum_j c_j f(x_j, \omega)$ *is a Gaussian random variable (maybe, degenerate).*

Let $k \in \mathbb{Z}_+$. *The Gaussian function* f *is called* C^k-smooth *(or just* C^k*) if*

(iii) *for almost every* $\omega \in \Omega$, *the function* $x \mapsto f(x, \omega)$ *belongs to the space* $C^k(V)$.

Removing a subset of zero probability from Ω, we may (and will) just demand that the function $x \mapsto f(x, \omega)$ is in $C^k(V)$ for all $\omega \in \Omega$.

Every C^k-Gaussian function f generates two mappings

$$\Omega \ni \omega \mapsto f(\,\cdot\,, \omega) \in C^k(V) \quad \text{and} \quad V \ni x \to f(x, \,\cdot\,) \in L^2(\Omega, \mathcal{P})\,.$$

With some abuse of notation, we denote these mappings by the same letter f.

Lemma A.3. *Suppose that* f *is a* C^k*-smooth Gaussian function on* V. *Then*

(a) *the mapping* $f\colon \big(V \times \Omega, \mathfrak{B}(V) \times \mathfrak{S}\big) \to \big(\mathbb{R}, \mathfrak{B}(\mathbb{R})\big)$ *is measurable;*

(b) *the mapping* $f\colon \big(\Omega, \mathfrak{S}\big) \to \big(C^k(V), \mathfrak{B}(C^k(V))\big)$ *is measurable;*

(c) *the mapping* $f\colon V \to L^2(\Omega, \mathcal{P})$ *is* C^k*-smooth.*

P r o o f of Lemma A.3.

(a) We partition V into countably many Borel sets V_j of diameter $\leqslant 1/n$ each, fix an arbitrary collection of points $x_j \in V_j$, and define a function $f_n\colon (V \times \Omega) \to \mathbb{R}$ by

$$f_n(x, \omega) = f(x_j, \omega) \qquad \text{for } x \in V_j\,.$$

The mappings $f_n\colon \big(V \times \Omega, \mathfrak{B}(V) \times \mathfrak{S}\big) \to \big(\mathbb{R}, \mathfrak{B}(\mathbb{R})\big)$ are measurable and $f = \lim\limits_{n\to\infty} f_n$ pointwise on $V \times \Omega$.

(b) It is an immediate consequence of Lemma A.1 combined with fact that, for every $x \in V$, the mapping $\omega \mapsto f(x, \omega)$ is measurable.

(c) Recall that if a sequence $\xi_n\colon \Omega \to \mathbb{R}$ of Gaussian random variables converges pointwise to ξ, then ξ is also a Gaussian random variable. It follows that for every multiindex α with $|\alpha| \leqslant k$, the mapping $(x, \omega) \mapsto \partial^\alpha f(x, \omega)$ is a continuous Gaussian function. Since the pointwise convergence of Gaussian random variables yields convergence in in $L^2(\Omega, \mathcal{P})$, we see that the mapping

$$V \ni x \mapsto \partial^\alpha f(x, \,\cdot\,) \in L^2(\Omega, \mathcal{P})$$

is continuous and gives the corresponding partial derivative of the mapping $V \ni x \mapsto f(x, \,\cdot\,)$ considered as a function on V with values in $L^2(\Omega, \mathcal{P})$. ∎

Definition A.2. *Let f be a C^k Gaussian function on V. Let $\gamma_f \stackrel{\text{def}}{=} f_*\mathcal{P}$ be the push-forward of the probability measure $\mathcal{P}$ to $C^k(V)$ by f. We say that two C^k Gaussian functions f_1 and f_2 are* equivalent *if $\gamma_{f_1} = \gamma_{f_2}$. We do not distinguish between equivalent Gaussian functions.*

In principle, we can forget about the original probability space $(\Omega, \mathfrak{S}, \mathcal{P})$ and consider the probability space $\big(C^k(V), \mathfrak{B}(C^k(V)), \gamma_f\big)$ and the mapping

$$V \times C^k(V) \ni (x, g) \mapsto g(x) \in \mathbb{R}$$

instead. We can go one step further and remove any Borel subset of γ_f-measure 0 from $C^k(V)$ in this representation.

A.3. Positive-definite kernels

Let f be a C^k Gaussian function on V. Let $K_f(t,s) \stackrel{\text{def}}{=} \mathcal{E}\big\{f(t)f(s)\big\}$ be the corresponding covariance kernel. It is a positive-definite symmetric function* on $V \times V$. The function f is uniquely determined by K_f up to equivalence. Indeed, since a Gaussian distribution in $\mathbb{R}^n$ is determined by its covariance matrix, this fact is evident for the sets of the form

$$S = \big\{g \in C^k(V) \colon \big(g(x_1), \ldots, g(x_n)\big) \in B\big\}$$

where $x_1, \ldots, x_n \in V$ and $B \in \mathfrak{B}(\mathbb{R}^n)$. The general case follows immediately because the fact that $\mathfrak{B}(C^k(V))$ is generated by point evaluations implies that every set $S \in \mathfrak{B}(C^k(V))$ can be approximated by sets of such kind up to an arbitrary small γ_f-measure.

Next, we observe that if g is a continuous Gaussian function on V with $K_g = K_f$, then g is equivalent, as a continuous Gaussian function, to the Gaussian function $\widetilde{f} \colon \Omega \stackrel{f}{\to} C^k(V) \hookrightarrow C(V)$. The function $\widetilde{f}$ generates a measure $\gamma_{\widetilde{f}}$ on $C(V)$:

$$\gamma_{\widetilde{f}}(S) = \gamma_f(S \cap C^k(V))\,, \qquad S \in \mathfrak{B}(C(V))\,.$$

Furthermore,

$$K_{\widetilde{f}}(x,y) = \mathcal{E}\big\{\widetilde{f}(x)\widetilde{f}(y)\big\} = \mathcal{E}\big\{f(x)f(y)\big\} = K_f(x,y) = K_g(x,y)\,.$$

Therefore, by the previous remark, $\gamma_{\widetilde{f}} = \gamma_g$. In this situation, almost surely, g is a C^k Gaussian function. Indeed, by Lemma A.2, $C(V) \setminus C^k(V) \in \mathfrak{B}(C(V))$, whence,

$$\gamma_g(C(V) \setminus C^k(V)) = \gamma_{\widetilde{f}}(C(V) \setminus C^k(V)) = \gamma_f(\varnothing) = 0\,.$$

*That is, the symmetric matrix $\big(K_f(x_i, x_j)\big)_{i,j=1}^n$ is positive definite for every choice of $x_1, \ldots, x_n \in V$.

That is, *any continuous Gaussian function whose covariance kernel coincides with the one of a C^k Gaussian function, almost surely, is a C^k function itself.*

Also observe that since the mapping

$$V \ni x \mapsto f(x, \cdot) \in L^2(\Omega, \mathcal{P})$$

is C^k, the partial derivative $\partial_x^\alpha \partial_y^\beta K_f(x, y)$ exists and is continuous on $V \times V$ for any multiindices α, β with $|\alpha|, |\beta| \leqslant k$. Moreover,

$$\partial_x^\alpha \partial_y^\beta K_f(x, y) = \mathcal{E}\big\{\partial_x^\alpha f(x)\, \partial_y^\beta f(y)\big\}.$$

A.4. From positive-definite kernels to reproducing kernel Hilbert spaces

In this section, we shall only assume that we are given a continuous positive-definite symmetric kernel K on $V \times V$. We shall describe a canonical construction of the Hilbert space $\mathcal{H} = \mathcal{H}(K)$ of continuous functions on V such that K is the *reproducing kernel* in that space, that is, for every $g \in \mathcal{H}$ and every $x \in V$, we have $g(x) = \langle g, K_x \rangle_{\mathcal{H}}$ where $K_x(y) = K(x, y)$.

Consider the linear space $\mathcal{L}$ of all mappings $\mathfrak{h}\colon V \to \mathbb{R}$ such that $\mathfrak{h}(x) \neq 0$ only for finitely many $x \in V$. Define the semi-definite scalar product on $\mathcal{L}$ by

$$\langle \mathfrak{h}_1, \mathfrak{h}_2 \rangle = \sum_{x,y \in V} K(x, y)\, \mathfrak{h}_1(x) \mathfrak{h}_2(y)$$

(this sum is actually finite). Since K is positive-definite, we have $\langle \mathfrak{h}, \mathfrak{h} \rangle \geqslant 0$ for every $\mathfrak{h} \in \mathcal{L}$. Define the Hilbert seminorm on $\mathcal{L}$ by $\|\mathfrak{h}\| = \sqrt{\langle h, h \rangle}$. Then $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ become a nondegenerate scalar product and the associated Hilbert norm on $\mathcal{L}/\mathcal{L}_0$ where $\mathcal{L}_0$ is the linear subspace of $\mathcal{L}$ consisting of all $\mathfrak{h} \in \mathcal{L}$ with $\|\mathfrak{h}\| = 0$. Let H be the Hilbert space completion of the pre-Hilbert space $\big(\mathcal{L}/\mathcal{L}_0, \langle \cdot, \cdot \rangle\big)$.

For $x \in V$, denote by h_x the vector in H corresponding to the function

$$\mathfrak{h}_x(y) = \begin{cases} 0, & y \neq x, \\ 1, & y = x. \end{cases}$$

Note that $\|h_x - h_y\|^2 = K(x, x) + K(y, y) - 2K(x, y) \to 0$ as $y \to x$, so the mapping $V \ni x \mapsto h_x \in H$ is continuous. Since $\operatorname{span}\{h_x\colon x \in V\}$ is dense in H by construction and since for every countable dense subset $V' \subset V$, the set $\{h_x\colon x \in V'\}$ is dense in $\{h_x\colon x \in V\}$, H is separable.

Now, define a linear map $\Phi\colon H \to C(V)$ by $\Phi[h](x) = \langle h, h_x \rangle$, $h \in H$. If $\Phi[h] = 0$, then $\langle h, h_x \rangle = 0$ for all $x \in V$, so $h = 0$. Thus, we can identify H with a linear subspace $\mathcal{H} = \Phi(H)$ of $C(V)$. Note also that

$$\Phi[h_x](y) = \langle h_x, h_y \rangle = K(x, y) = K_x(y),$$

so h_x is identified with K_x. Transferring the scalar product from H to $\mathcal{H}$, we turn $\mathcal{H}$ into a Hilbert space of continuous functions on V with the reproducing kernel K.

Observe, finally, that such a Hilbert space is unique. Indeed, if $\mathcal{H}_1 \subset C(V)$ is another Hilbert space of continuous functions with the same reproducing kernel K, then the linear span $\mathcal{H}_0$ of the functions K_x, $x \in V$, is contained and dense in $\mathcal{H}_1$ with respect to the Hilbert norm in $\mathcal{H}_1$ (because if $g \in \mathcal{H}_1$ is orthogonal to all K_x in $\mathcal{H}_1$, then $g(x) = \langle g, K_x \rangle_{\mathcal{H}_1} = 0$ for all $x \in V$, whence, $g = 0$) and for every pair of functions

$$g_1 = \sum_{\texttt{finite}} a_x K_x, \quad g_2 = \sum_{\texttt{finite}} b_y K_y$$

in $\mathcal{H}_0$, we have

$$\langle g_1, g_2 \rangle_{\mathcal{H}_1} = \sum_{x,y} K(x,y)\, a_x b_y = \langle g_1, g_2 \rangle_{\mathcal{H}}\,.$$

Thus the identity mapping $\mathcal{H}_0 \to \mathcal{H}_0$ can be extended to a bijective isometry $\mathcal{H} \to \mathcal{H}_1$. Let now $g' \in \mathcal{H}_1$ be the image of $g \in \mathcal{H}$ under this isometry. Then

$$g'(x) = \langle g', K_x \rangle_{\mathcal{H}_1} = \langle g, K_x \rangle_{\mathcal{H}} = g(x)\,, \quad x \in V\,,$$

so $\mathcal{H}_1$ and $\mathcal{H}$ consist of exactly the same functions on V and are endowed with the same scalar product. ∎

We end this section with a useful observation. Let $\{e_k\}$ be an arbitrary orthonormal basis in $\mathcal{H}$. For every $g \in \mathcal{H}$, we put $\widehat{g}(k) = \langle g, e_k \rangle_{\mathcal{H}}$. Then the Fourier series $\sum_k \widehat{g}(k) e_k$ converges to g in $\mathcal{H}$. For every $y \in V$, we have

$$\begin{aligned}\Big| g(y) - \sum_{1 \leqslant k \leqslant N} \widehat{g}(k) e_k(y) \Big| &= \Big| \Big\langle g - \sum_{1 \leqslant k \leqslant N} \widehat{g}(k) e_k, K_y \Big\rangle \Big| \\ &\leqslant \Big\| g - \sum_{1 \leqslant k \leqslant N} \widehat{g}(k) e_k \Big\|_{\mathcal{H}} \|K_y\|_{\mathcal{H}} \to 0 \quad \text{as } N \to \infty\,.\end{aligned}$$

Since $\|K_y\|_{\mathcal{H}} = \sqrt{K(y,y)}$ is a continuous function of y on V, this yields the locally uniform convergence of the Fourier series $\sum_k \widehat{g}(k) e_k$ to g.

Taking $g = K_x$ and observing that $\langle K_x, e_k \rangle = e_k(x)$, we conclude that for every $x, y \in V$, we have

$$\sum_k e_k(x) e_k(y) = K(x,y)\,.$$

A.5. Canonical series representation of continuous Gaussian functions

Let H_0 be any Gaussian subspace of $L^2(\Omega, \mathcal{P})$ and let $V \ni x \mapsto f_x \in H_0$ be a continuous mapping such that for every $x \in V$, the random variable f_x is Gaussian. The corresponding covariance kernel $K(x, y) = \mathcal{E}\{f_x f_y\} = \langle f_x, f_y \rangle_{L^2(\Omega,\mathcal{P})}$ is also continuous. Let H be the closed linear span of $\{f_x\}_{x\in V}$ in $L^2(\Omega, \mathcal{P})$. It is a Gaussian subspace of $L^2(\Omega, \mathcal{P})$. For $h \in H$, define $\Phi[h](x) = \langle h, f_x \rangle_{L^2(\Omega,\mathcal{P})}$. Note that $\Phi[h] \in C(V)$ and $\Phi[h] = 0$ if and only if $h = 0$. Also, $\Phi[f_x] = K(x, \cdot\,) = K_x$. Thus, $\mathcal{H} = \{\Phi[h] \colon h \in H\}$ is a linear subspace of $C(V)$ and if we endow it with the scalar product $\langle \Phi[h], \Phi[h'] \rangle_{\mathcal{H}} = \langle h, h' \rangle_{L^2(\Omega,\mathcal{P})}$, it will become a Hilbert space $\mathcal{H}(K)$ of continuous functions with the reproducing kernel K.

Now, take any orthonormal basis $\{e_j\}$ in $\mathcal{H}$ and choose $\xi_j \in H$ such that $e_j = \Phi[\xi_j]$. Note that

$$\langle \xi_i, \xi_j \rangle_{L^2(\Omega,\mathcal{P})} = \langle e_i, e_j \rangle_{\mathcal{H}} = \begin{cases} 0, & i \neq j\,, \\ 1, & i = j\,, \end{cases}$$

so ξ_j are orthogonal and, thereby, independent standard Gaussian. For every $x \in V$, we have

$$f_x = \sum_j \langle f_x, \xi_j \rangle\, \xi_j = \sum_j \Phi[\xi_j](x) \xi_j = \sum_j e_j(x) \xi_j\,.$$

The upshot is that,

- *given any Gaussian subspace $H_0 \subset L^2(\Omega, \mathcal{P})$, any continuous mapping $x \mapsto f_x$ from V to H_0, and any orthonormal basis e_j in the reproducing kernel Hilbert space $\mathcal{H}(K)$, where $K(x, y) = \mathcal{E}\{f(x) f(y)\}$, we can define independent standard Gaussian real variables on $(\Omega, \mathfrak{S}, \mathcal{P})$ such that $f_x = \sum_j \xi_j\, e_j(x)$ for all $x \in V$.*

Assume now that we start with a continuous Gaussian function f with some underlying probability space $(\Omega, \mathfrak{S}, \mathcal{P})$. Applying the above construction to the induced mapping $x \mapsto f_x = f(x, \cdot\,)$, we get

$$f(x, \omega) = \sum_j \xi_j(\omega)\, e_j(x) \quad \text{in } L^2(\Omega, \mathcal{P}) \text{ for all } x \in V\,. \tag{A.1}$$

Implementing ξ_j as some everywhere defined functions on Ω and taking into account that $L^2(\Omega, \mathcal{P})$-convergence yields convergence in probability, we have, in particular, that

$$\text{for every } x \in V, \quad \sum_{j=1}^{n} \xi_j(\omega)\, e_j(x) \to f(x) \quad \text{in probability as } n \to \infty\,. \tag{A.2}$$

Now, put $X_j = \xi_j e_j(x)$, $S_n = \sum_{j=1}^n X_j$ and note that for every compact $Q \subset V$, the random variables X_j, S_n and $S = f$ can be viewed as random vectors in the Banach space $C(Q)$. The random variables X_j are symmetric and independent, and (A.2) means that for every point evaluation functional $z_x \in C(Q)^*$ given by $\langle z_x, g\rangle = g(x)$ for $x \in Q$, we have $\langle z, S_n\rangle \to \langle z_x, S\rangle$ in probability. By the classical Ito–Nisio theorem, which we will recall in the next section, the series $\sum_j X_j$ converges to S in $C(Q)$. Thus,

- *the canonical series representation* (A.1) *actually converges in* $C(V)$.

A.6. The Ito–Nisio theorem

Let $\mathfrak{X}$ be a separable Banach space. An $\mathfrak{X}$-valued random variable on a probability space $(\Omega, \mathfrak{S}, \mathcal{P})$ is just a measurable mapping from $(\Omega, \mathfrak{S}, \mathcal{P})$ to $(\mathfrak{X}, \mathfrak{B}(\mathfrak{X}))$. Everywhere below, X_j is a sequence of independent $\mathfrak{X}$-valued random variables, S is an $\mathfrak{X}$-valued random variable on the same probability space, and $S_n = \sum_{j\leqslant n} X_j$. We denote by $\|\cdot\|$ the norm in $\mathfrak{X}$, and by $\langle\cdot,\cdot\rangle$ the natural coupling of the dual space $\mathfrak{X}^*$ and $\mathfrak{X}$.

First, we recall a classical

P. Levý's lemma. *If S_n converges to S in probability, then S_n converges to S almost surely.*

P r o o f of Levý's lemma. We will check that, for almost every $\omega \in \Omega$, $S_n(\omega)$ is a Cauchy sequence. Take $\varepsilon \in (0, \frac{1}{2})$ and take m so large that

$$\mathcal{P}\big\{\|S_k - S\| \geqslant \tfrac{1}{2}\,\varepsilon\big\} < \tfrac{1}{2}\,\varepsilon\,, \qquad k \geqslant m\,.$$

Then take any positive integer $n > m$. For $k = m, \dots, n$, let

$$A_k = \big\{\omega \in \Omega\colon \|S_\ell - S_m\| \leqslant 2\varepsilon \text{ for all } \ell = m, \dots, k-1, \text{ but } \|S_k - S_m\| > 2\varepsilon\big\}.$$

Note that the events A_k are disjoint, and each A_k is independent of $S_n - S_k$. Also, if $\omega \in A_k$ and $\|S_n - S_k\| \leqslant \varepsilon$, then

$$\|S_n - S_m\| \geqslant \|S_k - S_m\| - \|S_n - S_k\| > 2\varepsilon - \varepsilon = \varepsilon\,.$$

Furthermore, since

$$\big\{\|S_k - S_n\| > \varepsilon\big\} \subset \big\{\|S_k - S\| > \tfrac{1}{2}\,\varepsilon\big\} \bigcup \big\{\|S_n - S\| > \tfrac{1}{2}\,\varepsilon\big\}\,,$$

we have

$$\mathcal{P}\big\{\|S_k - S_n\| \leqslant \varepsilon\big\} = 1 - \mathcal{P}\big\{\|S_k - S_n\| > \varepsilon\big\} > 1 - \varepsilon\,.$$

Therefore,

$$\begin{aligned}(1-\varepsilon)\,\mathcal{P}\{\max_{m\leqslant k\leqslant n}\|S_k-S_m\|>2\varepsilon\}&=(1-\varepsilon)\sum_{k=m}^{n}P(A_k)\\&<\sum_{k=m}^{n}\mathcal{P}\{\|S_n-S_k\|\leqslant\varepsilon\}\,\mathcal{P}(A_k)\\&=\sum_{k=m}^{n}\mathcal{P}\{A_k\text{ and }\|S_n-S_k\|\leqslant\varepsilon\}\\&\leqslant\mathcal{P}\{\|S_n-S_m\|>\varepsilon\}\leqslant\varepsilon\,.\end{aligned}$$

Since n and ε are arbitrary, we see that, with probability 1, S_n is Cauchy, so it converges almost surely. Clearly, the almost sure limit and the limit in probability must be the same. ∎

We call a subset $Z\subset\mathfrak{X}^*$ *normalizing* if it is countable and $\|x\|=\sup\{\langle z,x\rangle\colon z\in Z\}$ (then automatically Z is contained in the unit ball of $\mathfrak{X}^*$). Now, we can state the part of the Ito–Nisio theorem that we need:

Ito–Nisio theorem. *Suppose that the random variables X_j are symmetric and that there exists a normalizing set $Z\subset\mathfrak{X}^*$ such that $\langle z,S_n\rangle\to\langle z,S\rangle$ in probability for every $z\in Z$. Then $S_n\to S$ in $\mathfrak{X}$ almost surely.*

P r o o f. By P. Levý's lemma, it is enough to show that S_n converges to S in probability. First of all, note that the Borel σ-algebra $\mathfrak{B}(\mathfrak{X})$ coincides with the σ-algebra $\mathfrak{B}'(\mathfrak{X})$ generated by the events $\{x\colon\langle z,x\rangle\in[a,b)\}$, $z\in Z$, $a,b\in\mathbb{R}$. Indeed, for every $x_0\in\mathfrak{X}$, the mapping $x\mapsto\|x-x_0\|=\sup_Z\|\langle z,x\rangle-\langle z,x_0\rangle\|$ is $\mathfrak{B}'(\mathfrak{X})$-measurable. Thus, every open ball in $\mathfrak{X}$ is $\mathfrak{B}'(\mathfrak{X})$-measurable. Since $\mathfrak{X}$ is separable, every open set in $\mathfrak{X}$ is $\mathfrak{B}'(\mathfrak{X})$-measurable, so $\mathfrak{B}(\mathfrak{X})\subset\mathfrak{B}'(\mathfrak{X})$. The inverse inclusion is obvious.

Next, we show that S_n and $S-S_n$ are independent for every n. We need to check that

$$\mathcal{P}\{S_n\in C_1,S-S_n\in C_2\}=\mathcal{P}\{S_n\in C_1\}\,\mathcal{P}\{S-S_n\in C_2\}$$

for every $C_1,C_2\in\mathfrak{B}(\mathfrak{X})$. Since $\mathfrak{B}(\mathfrak{X})=\mathfrak{B}'(\mathfrak{X})$, it suffices to check this for the events of the form

$$C=\big\{\big(\langle z_1,x\rangle,\,\dots\,,\langle z_q,x\rangle\big)\in B\big\},\quad B\in\mathfrak{B}(\mathbb{R}^q),\ z_1,\dots,z_q\in Z\,,$$

in which case it follows from the independence of S_m-S_n and S_n for $m>n$ and the fact that, for $m\to\infty$,

$$\big(\langle z_1,S_m-S_n\rangle,\,\dots\,,\langle z_q,S_m-S_n\rangle\big)\to\big(\langle z_1,S-S_n\rangle,\,\dots\,,\langle z_q,S-S_n\rangle\big)\quad\text{in probability}\,.$$

For a set $\mathfrak{X}' \subset \mathfrak{X}$, denote

$$\mathfrak{X}'_{+\varepsilon} = \bigcup_{x\in\mathfrak{X}'} B(x,\varepsilon)\,.$$

We claim that *for every finite set $\mathfrak{X}' \subset \mathfrak{X}$, there exists a finite "separating set" of functionals $Z' \subset Z$ such that*

$$\max_{z\in Z'} |\langle z, y'-y''\rangle| > \varepsilon\,, \quad \text{whenever } y', y'' \in \mathfrak{X}'_{+\varepsilon} \text{ and } \|y'-y''\| > 8\varepsilon\,.$$

Indeed, consider all differences $x'-x''$ with $x', x'' \in \mathfrak{X}'$ and for each of them choose $z = z(x', x'') \in Z$ such that $|\langle z, x'-x''\rangle| \geqslant \frac{1}{2}\,\|x'-x''\|$. Since $y', y'' \in \mathfrak{X}'_{+\varepsilon}$, we can find $x', x'' \in \mathfrak{X}'$ so that $\|x'-y'\|, \|x''-y''\| < \varepsilon$. Then $\|x'-x''\| > 6\varepsilon$. Taking $z = z(x', x'')$, we get

$$|\langle z, y'-y''\rangle| > |\langle z, x'-x''\rangle| - 2\varepsilon \geqslant 3\varepsilon - 2\varepsilon = \varepsilon\,,$$

proving the claim.

Now, comes the crux of the proof. Suppose that A and B are $\mathfrak{X}$-valued independent random variables and A is symmetric. Then, for every finite $\mathfrak{X}' \subset \mathfrak{X}$ and every $\varepsilon > 0$, we can write

$$\mathcal{P}\big\{A \notin \big(\tfrac{1}{2}(\mathfrak{X}'-\mathfrak{X}')\big)_{+\varepsilon}\big\} \leqslant \mathcal{P}\big\{A+B \notin \mathfrak{X}'_{+\varepsilon}\big\} + \mathcal{P}\big\{-A+B \notin \mathfrak{X}'_{+\varepsilon}\big\} = 2\,\mathcal{P}\big\{A+B \notin \mathfrak{X}'_{+\varepsilon}\big\}\,.$$

The inequality here is due to the observation that

$$\text{if } a, b \in \mathfrak{X} \quad \text{and } a+b, -a+b \in \mathfrak{X}'_{+\varepsilon}, \quad \text{then } a \in \big(\tfrac{1}{2}(\mathfrak{X}'-\mathfrak{X}')\big)_{+\varepsilon}\,.$$

The equality follows at once from the symmetry of A and the independence of A and B.

To finish the proof, we take $\varepsilon > 0$ and let x_1, x_2, ... be a countable dense set in $\mathfrak{X}$. Put $\mathfrak{X}'_N = \{x_1, \ldots, x_N\}$. Since $(\mathfrak{X}'_N)_{+\varepsilon} \uparrow \mathfrak{X}$ as $N \to \infty$, we have $\mathcal{P}\{S \notin (\mathfrak{X}'_N)_{+\varepsilon}\} < \varepsilon$ for large enough N. We fix such N and, to simplify notation, let $\mathfrak{X}' = \mathfrak{X}'_N$. Since S_n and $S - S_n$ are independent, and S_n is symmetric, we can use them as A and B in the argument above and get $\mathcal{P}\big\{S_n \notin \big(\frac{1}{2}(\mathfrak{X}'-\mathfrak{X}')\big)_{+\varepsilon}\big\} < 2\varepsilon$ for all n. Let $Z' \subset Z$ be a finite separating set for $\mathfrak{X}' \cup \frac{1}{2}(\mathfrak{X}'-\mathfrak{X}')$. Then for every n,

$$\begin{aligned}
&\mathcal{P}\big\{\|S_n - S\| > 8\varepsilon\big\} \\
&\quad\leqslant \mathcal{P}\big\{S_n \notin \big(\tfrac{1}{2}(\mathfrak{X}'-\mathfrak{X}')\big)_{+\varepsilon}\big\} + \mathcal{P}\big\{S \notin \mathfrak{X}'_{+\varepsilon}\big\} + \mathcal{P}\big\{\max_{z\in Z'} |\langle z, S_n\rangle - \langle z, S\rangle| > \varepsilon\big\} \\
&\qquad\leqslant 3\varepsilon + \sum_{z\in Z'} \mathcal{P}\big\{|\langle z, S_n\rangle - \langle z, S\rangle| > \varepsilon\big\}\,.
\end{aligned}$$

Since each term in the finite sum on the RHS tends to 0 as $n \to \infty$ and ε can be taken as small as we want, the desired convergence in probability follows. ∎

A.7. The local behavior of continuous Gaussian functions

Suppose that f is a continuous Gaussian function on V with the covariance kernel K. As before, we denote by γ_f the corresponding Gaussian measure on $C(V)$. The (closed) set $S(f)$ of functions $g \in C(V)$ for which $\mathcal{P}\{f \in U\} = \gamma_f(U) > 0$ for every open neighbourhood U of g in $C(V)$ is called the *topological support of the measure* γ_f.

The following lemma gives a simple and useful description of the topological support of γ_f:

Lemma A.4. $S(f) = \mathrm{Clos}_{C(V)}\, \mathcal{H}(K)$.

P r o o f of Lemma A.4. First, we show that for every $g \in \mathcal{H}(K)$, every compact $Q \subset V$, and every $\varepsilon > 0$, we have $\mathcal{P}\{\|f - g\|_{C(Q)} < \varepsilon\} > 0$. We choose an orthonormal basis $\{e_j\}$ in $\mathcal{H}(K)$ so that $g = te_1$ for some $t \in \mathbb{R}$, and represent f as $\sum_j \xi_j e_j$ where ξ_j are independent Gaussian random variables. Since, by the Ito–Nisio theorem, the series converges in $C(Q)$ with probability 1, there exists $N = N(\varepsilon)$ such that

$$\|\sum_{j>N} \xi_j e_j\|_{C(Q)} = \|f - \sum_{j\leqslant N} \xi_j e_j\|_{C(Q)} < \tfrac{1}{2}\,\varepsilon$$

with positive probability. Next, we choose η so small that

$$\eta \sum_{j\leqslant N} \|e_j\|_{C(Q)} < \tfrac{1}{2}\,\varepsilon\,.$$

Now, suppose that

$$\|\sum_{j>N} \xi_j e_j\|_{C(Q)} < \tfrac{1}{2}\,\varepsilon\,, \quad \xi_1 \in (t-\eta, t+\eta)\,, \quad \text{and} \quad \xi_2, \ldots, \xi_n \in (-\eta, \eta)\,.$$

Then

$$\begin{aligned}\|f - g\|_{C(Q)} &\leqslant |\xi_1 - t|\,\|e_1\|_{C(Q)} + \sum_{2\leqslant j\leqslant N} |\xi_j|\,\|e_j\|_{C(Q)} + \big\|\sum_{j>N} \xi_j e_j\big\|_{C(Q)} \\ &\leqslant \eta \sum_{j\leqslant N} \|e_j\|_{C(Q)} + \big\|\sum_{j>N} \xi_j e_j\big\|_{C(Q)} < \varepsilon\,.\end{aligned}$$

Hence,

$$\begin{aligned}\mathcal{P}\{\|f - g\|_{C(Q)} < \varepsilon\} &\geqslant \mathcal{P}\big\{\|\sum_{j>N} \xi_j e_j\|_{C(Q)} < \tfrac{1}{2}\,\varepsilon\big\} \\ &\quad \times \mathcal{P}\big\{\xi_1 \in (t-\eta, t+\eta),\ \xi_2, \ldots, \xi_N \in (-\eta, \eta)\big\} > 0\,.\end{aligned}$$

Thus, $\mathcal{H}(K) \subset S(f)$. Since $S(f)$ is closed in $C(V)$, we get $S(f) \supset \mathrm{Clos}_{C(V)}\,\mathcal{H}(K)$.

To show the converse, assume that $g \in C(V)$ and

$$\mathcal{P}\{\|f-g\|_{C(Q)} < \tfrac{1}{2}\,\varepsilon\} = p > 0\,.$$

We fix an orthonormal basis $\{e_j\}$ in $\mathcal{H}(K)$ and choose N so large that

$$\mathcal{P}\{\|f - \sum_{j\leqslant N} \xi_j e_j\|_{C(Q)} > \tfrac{1}{2}\,\varepsilon\} < \tfrac{1}{2}\,p\,.$$

Then

$$\mathcal{P}\{\|g - \sum_{j\leqslant N} \xi_j e_j\|_{C(Q)} < \varepsilon\} \geqslant p - \tfrac{1}{2}\,p > 0\,.$$

Since

$$\sum_{j\leqslant N} \xi_j e_j \in \mathcal{H}(K)\,,$$

we conclude that the ε-neighbourhood of g in $C(Q)$ intersects $\mathcal{H}(K)$. Since ε and Q are arbitrary, we see that $S(f) \subset \mathrm{Clos}_{C(V)}\,\mathcal{H}(K)$. ∎

A.8. Fernique's theorem

The next result we state was proven by Fernique and independently by Landau and Shepp. It allows one to pass from some very weak estimates for various norms and semi-norms of Gaussian functions to almost as strong bounds for tails as possible in principle.

Fernique's theorem. *Let X be a random variable with values in a Banach space $\mathfrak{X}$, and let $\{\varphi_j\} \subset \mathfrak{X}^*$ be an at most countable set of linear functionals on $\mathfrak{X}$ such that, for every choice of finitely many φ_j's, the joint distribution of $\varphi_j(X)$ is Gaussian. Suppose that, for some $\lambda > 0$ and $\mu < \frac{1}{2}$,*

$$\mathcal{P}\big\{\sup_j |\varphi_j(X)| \geqslant \lambda\big\} \leqslant \mu\,. \tag{A.1}$$

Then, for all $t \geqslant 1$,

$$\mathcal{P}\big\{\sup_j |\varphi_j(X)| \geqslant \lambda t\big\} \leqslant e^{-at^2} \tag{A.2}$$

with a positive constants a depending only on μ.

Here, we present Fernique's original proof, which is short and elegant. Landau and Shepp [21] gave a different proof based on the Gaussian isoperimetry. The

advantage of the latter proof is that it does not need a priori assumption $\mu < \frac{1}{2}$ and gives the optimal RHS of (A.2), which is $\Phi\big(t\Phi^{-1}(\mu)\big)$, where

$$\Phi(s) = \sqrt{\frac{2}{\pi}} \int\limits_s^\infty e^{-x^2/2}\,\mathrm{d}x\,.$$

P r o o f of Fernique's theorem. Without loss of generality, we assume that $\lambda = 1$. Let $\Omega_n(t)$ be the event $\big\{\sup_{1\leqslant j\leqslant n} |\varphi_j(X)| > t\big\}$, $n \in \mathbb{N} \cup \{\infty\}$. We need to estimate $\mathcal{P}\{\Omega_\infty(t)\}$ assuming that $\mathcal{P}\{\Omega_\infty(1)\} \leqslant \mu$. Since $\Omega_n(t) \subset \Omega_m(t)$ for $n \leqslant m$, and $\Omega_\infty(t) = \bigcup_{n\geqslant 1} \Omega_n(t)$, it will suffice to prove estimate (A.2) for every finite n with constants A and a independent of n.

In what follows, we fix $n \in \mathbb{N}$, and put $\varphi = \big(\varphi_1(X), \dots, \varphi_n(X)\big)$. This is a finite-dimensional Gaussian random vector. We let $\|\varphi\| \stackrel{\text{def}}{=} \max\limits_{1\leqslant j\leqslant n} |\varphi_j(X)|$. Fernique's proof is based on the following classical observation: *if ψ is an n-dimensional Gaussian vector, which has the same distribution as φ and which is independent of φ, then $\frac{1}{\sqrt{2}}(\varphi + \psi)$ and $\frac{1}{\sqrt{2}}(\varphi - \psi)$ are two Gaussian vectors, which have the same distribution as φ and which are independent of each other.* The proof of this statement reduces to a routine verification of coincidence of all relevant covariance matrices.

Now, take $t > 0$ and $\tau > 0$ and write

$$\begin{aligned}
\mathcal{P}\big\{\|\psi\| \leqslant \tau\big\}\,\mathcal{P}\big\{\|\varphi\| > t\big\} &= \mathcal{P}\big\{\|\tfrac{1}{\sqrt{2}}(\varphi - \psi)\| \leqslant \tau\big\}\,\mathcal{P}\big\{\|\tfrac{1}{\sqrt{2}}(\varphi + \psi)\| > t\big\} \\
&= \mathcal{P}\big\{\|\tfrac{1}{\sqrt{2}}(\varphi - \psi)\| \leqslant \tau,\ \|\tfrac{1}{\sqrt{2}}(\varphi + \psi)\| > t\big\} \\
&\leqslant \mathcal{P}\big\{\|\varphi\| > \tfrac{1}{\sqrt{2}}(t - \tau),\ \|\psi\| > \tfrac{1}{\sqrt{2}}(t - \tau)\big\} \\
&= \big(\mathcal{P}\big\{\|\varphi\| > \tfrac{1}{\sqrt{2}}(t - \tau)\big\}\big)^2.
\end{aligned}$$

Letting $\tau = 1$ and recalling that $\mathcal{P}\big\{\|\psi\| \leqslant 1\big\} \geqslant 1 - \mu$, we get

$$\mathcal{P}\big\{\|\varphi\| > t\big\} \leqslant \frac{1}{1-\mu}\Big(\mathcal{P}\big\{\|\varphi\| > \tfrac{1}{\sqrt{2}}(t - \tau)\big\}\Big)^2.$$

Put $p(t) = \mathcal{P}\big\{\|\varphi\| > t\big\}$. This is a non-increasing function of t, which satisfies

$$\begin{aligned}
&p(t) \leqslant \frac{1}{1-\mu} p^2\big(\tfrac{1}{\sqrt{2}}(t-1)\big)\,, \qquad t \geqslant 1\,, \\
&p(1) \leqslant \mu\,.
\end{aligned}$$

Let

$$t_k = \frac{(\sqrt{2})^{k+1} - 1}{\sqrt{2} - 1}\,, \qquad k \geqslant 0\,,$$

that is, $t_0 = 1$, and $t_k = \frac{1}{\sqrt{2}}(t_{k+1} - 1)$. Then, by induction on k, we have

$$p(t_k) \leqslant (1 - \mu) \left(\frac{\mu}{1 - \mu}\right)^{2^k}.$$

Since

$$t_{k+1}^2 = \Big[\frac{(\sqrt{2})^{k+2} - 1}{\sqrt{2} - 1}\Big]^2 < \frac{2^{k+2}}{(\sqrt{2} - 1)^2} = \frac{4}{(\sqrt{2} - 1)^2}\, 2^k\,,$$

we see that for $t_k \leqslant t \leqslant t_{k+1}$,

$$p(t) \leqslant p(t_k) < e^{-at^2} \quad \text{with } a = \frac{1}{4}\,(\sqrt{2} - 1)^2 \log \frac{\mu}{1 - \mu} > 0\,,$$

completing the proof. ∎

A.9. Kolmogorov's theorem

Here, we formulate a version of the classical Kolmogorov's theorem for $C^{k,k}$ kernels. Let $k \in \mathbb{N}$ and let, as before, $V \subset \mathbb{R}^m$ be an open set.

Definition A.3. *We say that a symmetric function $K\colon V \times V \to \mathbb{R}$ belongs to $C^{k,k}(V \times V)$ if all partial derivatives of K including at most k differentiations in x variables and at most k differentiations in y variables exist and are continuous on $V \times V$ (in which case, the order of differentiations does not matter and we can denote these derivatives $\partial_x^\alpha \partial_y^\beta K(x, y)$ as usual).*

Kolmogorov's theorem. *Let $k \in \mathbb{N}$. Suppose that $K\colon V \times V \to \mathbb{R}$ is a positive definite symmetric function of class $C^{k,k}(V \times V)$ and, in addition, that*

$$N_{V,k}(K) \stackrel{\text{def}}{=} \max_{|\alpha|,|\beta| \leqslant k} \sup_{x,y \in V} \left|\partial_x^\alpha\, \partial_y^\beta K(x, y)\right| < \infty\,.$$

Then there exists a (unique up to an equivalence) C^{k-1} Gaussian function f on V with the covariance kernel K.

Moreover, for every $\gamma \in (0, 1)$ and every closed ball $\bar{B} \subset V$, we have

$$\mathcal{E}\big\{\|f\|_{\bar{B},\, k-1+\gamma}\big\} \leqslant C(\bar{B}, V, k, \gamma)\, \sqrt{N_{V,k}(K)}\,.$$

Note that since every compact set $Q \subset V$ can be covered by a finite union of closed balls contained in V, the latter estimate immediately implies that, for any compact set $Q \subset V$,

$$\mathcal{E}\big\{\|f\|_{Q,\,k-1}\big\} \leqslant C(Q,V,k)\,\sqrt{N_{V,k}(K)}\,.$$

The same is true for the Hölder norm, but the cover should be chosen carefully so that any two sufficiently close points $x, y \in Q$ are covered by a single ball. Then the resulting bound on Q depends on both the bounds on the balls and the geometry of the cover. We will never need to estimate the Hölder norms on any compact set other than a ball, so we will not go into the details here.

It is also worth noting that in the assumptions of Kolmogorov's theorem we use that $N_{V,k}(K) < \infty$ instead of the more natural for a function defined on an open set assumption that $N_{Q,k}(K) < \infty$ for every compact set $Q \subset V$. This allows us to reduce the number of nested compact sets we need to choose before doing any estimate. Of course, this replacement it is harmless.

A.10. Proof of Kolmogorov's theorem

To prove Kolmogorov's theorem we will use a "convolution approach". As far as high order derivatives are concerned, this approach allows one to pass to the limits in a family of covariance kernels easier than the more usual approach based on nets (see, for instance, [15, Sec. 3.1]).

We split the proof into several steps.

A.10.1. As we have seen in A.4., there exists a separable Hilbert space $\mathcal{H}$ and a continuous mapping $V \ni x \mapsto f_x \in \mathcal{H}$ such that $K(x,y) = \langle f_x, f_y\rangle$. Without loss of generality, we assume that $\mathcal{H}$ is a Gaussian subspace of $L^2(\Omega,\mathcal{P})$, where $(\Omega,\mathfrak{S},\mathcal{P})$ is a probability space. Our first task is to implement the mapping $x \mapsto f_x$ as a $\mathfrak{B}(V)\times\mathfrak{S}$-measurable function of (x,ω).

We start with implementing each f_x as an everywhere defined function on Ω. Then we pick a compact exhaustion Q_n of V, a sequence $\varepsilon_n > 0$ with $\sum_n \varepsilon_n^2 < \infty$, and choose $\rho_n > 0$ so small that $\|f_x - f_y\|^2_{L^2(\Omega,\mathcal{P})} < \varepsilon_n^2$ for all $x \in Q_n$, $y \in V$ with $|x-y| < \rho_n$. We fix a countable partition of V into Borel sets $V_{j,n}$ of diameter less than ρ_n each, choose some point $x_{j,n}$ in every $V_{j,n}$ and put $f_n(x,\omega) = f_{x_{j,n}}(\omega)$ if $x \in V_j$. Then, for every $x \in Q_n$, $\|f_n - f_x\|^2_{L^2(\Omega,\mathcal{P})} < \varepsilon_n^2$, so for each $t > 0$, we have

$$\mathcal{P}\big\{|f_n(x,\omega) - f_x(\omega)| > t\big\} < t^{-2}\,\varepsilon_n^2\,.$$

Since $\sum_n \varepsilon_n^2 < +\infty$ and Q_n exhaust V, the functions $f_n(x,\,\cdot\,)$ converge to f_x both $\mathcal{P}$-almost surely and in $L^2(\Omega,\mathcal{P})$.

Let now $E = \{(x,\omega)\colon \lim\limits_{n\to\infty} f_n(x,\omega) \text{ exists}\}$. Since f_n is $\mathfrak{B}(V)\times\mathfrak{S}$-measurable, so is E. Also, for every $x \in V$, we have $\mathcal{P}\{\omega\colon (x,\omega) \notin E\} = 0$. Thus,

$$f(x,\omega) \stackrel{\text{def}}{=} \begin{cases} \lim_{n\to\infty} f_n(x,\omega), & (x,\omega) \in E, \\ 0, & \text{otherwise} \end{cases}$$

is a measurable representation of the mapping $x \mapsto f_x$.

A.10.2. Denote $F_\omega(x) = f(x,\omega)$. By Fubini, for every compact $Q \subset V$,

$$\mathcal{E}\Big\{\int\limits_Q |F_\omega|^2 \,\mathrm{d}\,\mathrm{vol}\Big\}$$

$$= \int\limits_Q \|f_x\|^2_{L^2(\Omega,\mathcal{P})} \,\mathrm{d}\,\mathrm{vol}(x) \leqslant \max_{x\in Q} K(x,x)\ \mathrm{vol}\, Q \leqslant N_{V,k}(K)^2\ \mathrm{vol}\, Q\,.$$

Thus $F_\omega \in L^2_{\mathtt{loc}}(V)$ for every $\omega \in \Omega_1 \subset \Omega$ with $\mathcal{P}(\Omega_1) = 1$. Replacing $f(x,\omega)$ by $f(x,\omega)\mathbb{1}_{\Omega_1}(\omega)$, we will assume that $f(x,\omega)$ is such that $F_\omega \in L^2_{\mathtt{loc}}(V)$ for all $\omega \in \Omega$.

A.10.3. Next, we note that for every $\varphi \in C^\infty_0(B(r))$, the convolution in the x variable

$$(f *_x \varphi)(x,\omega) \stackrel{\text{def}}{=} (F_\omega * \varphi)(x)$$

is a C^k (actually, C^∞) Gaussian function on V_{-r} for every $\omega \in \Omega$. The only non-trivial part of this claim is the Gaussian distribution property. To see it, observe that, as an element of $\mathcal{H} \subset L^2(\Omega,\mathcal{P})$,

$$(f *_x \varphi)(x,\,\cdot\,) = \int\limits_{B(r)} f_{x+y}\,\varphi(y)\,\mathrm{d}\,\mathrm{vol}(y)\,.$$

The integral on the RHS can be understood as the usual Riemann integral of a continuous $L^2(\Omega,\mathcal{P})$-valued function, and hence, it can be approximated in $L^2(\Omega,\mathcal{P})$ by finite Riemann sums $\sum_j c_j f_{x+y_j} \in \mathcal{H}$ and, therefore, lies in $\mathcal{H}$ itself. In what follows, we write $f * \varphi$ instead of $f *_x \varphi$ and view $f * \varphi$ as a random Gaussian function.

A.10.4. We shall need a few estimates for $f*\varphi$ and its derivatives $\partial^\alpha(f*\varphi) = f * \partial^\alpha\varphi$ for $|\alpha| \leqslant k-1$. First of all, by Fubini,

$$\mathcal{E}\big\{\big[\partial^\alpha(f*\varphi)(z)\big]^2\big\} = \iint\limits_{B(r)\times B(r)} K(z+x,z+y)\partial^\alpha\varphi(x)\partial^\alpha\varphi(y)\,\mathrm{d}\,\mathrm{vol}(x)\,\mathrm{d}\,\mathrm{vol}(y)$$

$$= \iint\limits_{B(r)\times B(r)} \partial^\alpha_x\partial^\alpha_y K(z+x,z+y)\varphi(x)\varphi(y)\,\mathrm{d}\,\mathrm{vol}(x)\,\mathrm{d}\,\mathrm{vol}(y)\,.$$

The expression on the right is trivially bounded by $\|\varphi\|_{L^1}^2 N_{V,k}(K)$.

If, in addition, the function φ has zero integral mean, we can improve our trivial bound to $Cr^2 \|\varphi\|_{L^1}^2 N_{V,k}(K)$. To see this, we put

$$E_\alpha(z;x,y) = \partial_x^\alpha \partial_y^\alpha K(z+x,z+y) - \partial_x^\alpha \partial_y^\alpha K(z,z+y) - \partial_x^\alpha \partial_y^\alpha K(z+x,z) + \partial_x^\alpha \partial_y^\alpha K(z,z)$$

and note that by "bilinear" Lagrange's Mean-Value Theorem, $|E_\alpha(z;x,y)| \leqslant Cr^2 N_{V,k}(K)$. Then, writing

$$\partial_x^\alpha \partial_y^\alpha K(z+x,z+y) = -\partial_x^\alpha \partial_y^\alpha K(z,z) + \partial_x^\alpha \partial_y^\alpha K(z,z+y) + \partial_x^\alpha \partial_y^\alpha K(z+x,z) + E_\alpha(z;x,y)$$

and integrating in x and y against $\varphi(x)\varphi(y)\,\mathrm{d\,vol}(x)\,\mathrm{d\,vol}(y)$, we obtain

$$\mathcal{E}\big\{\big[\partial^\alpha (f*\varphi)(z)\big]^2\big\} = \iint\limits_{B(r)\times B(r)} E_\alpha(z;x,y)\,\mathrm{d\,vol}(x)\,\mathrm{d\,vol}(y) \leqslant Cr^2 \|\varphi\|_{L^1}^2 N_{V,k}(K)\,.$$

A.10.5. We shall also need the following "entropy bound":

Lemma A.5 (entropy bound)**.** *Let $r > 0$. Let g be a continuous Gaussian function on V and ψ be any $C_0^\infty(B(r))$-function. Then $g*\psi$ is a continuous Gaussian function on V_{-r} and for every two compact sets $Q, Q' \subset V$ such that $Q_{+r} \subset Q'$, we have*

$$\mathcal{E}\big\{\|g*\psi\|_{C(Q)}\big\} \leqslant 5\|\psi\|_{L^1} \sqrt{1+\log\frac{\|\psi\|_{L^\infty}\operatorname{vol} Q'}{\|\psi\|_{L^1}}}\ \sqrt{\sup_{Q'} \mathcal{E}\{|g|^2\}}\,.$$

P r o o f of the entropy bound. Without loss of generality, we assume that $\sup_{Q'} \mathcal{E}\{|g|^2\} = 1$. Then for every $x \in Q'$, $g(x)$ is a Gaussian random variable with $\mathcal{E}\{g(x)^2\} \leqslant 1$. Hence,

$$\mathcal{E} e^{\frac{1}{4}g(x)^2} \leqslant \frac{1}{\sqrt{2\pi}} \int\limits_{\mathbb{R}} e^{\frac{1}{4}x^2}\, e^{-\frac{1}{2}x^2}\,\mathrm{d}x = \sqrt{2}\,.$$

Take $\rho \geqslant \sqrt{2}$. Noting that the function $\rho e^{-\frac{1}{4}\rho^2}$ decreases on $[\sqrt{2},+\infty)$, we estimate the convolution by

$$\begin{aligned}(g*\psi)(x) &= \int\limits_{B(r)} g(x+y)\psi(y)\,\mathrm{d\,vol}(y)\\ &= \int\limits_{B(r)\cap\{|g|\leqslant\rho\}} g(x+y)\psi(y)\,\mathrm{d\,vol}(y) + \int\limits_{B(r)\cap\{|g|>\rho\}} g(x+y)\psi(y)\,\mathrm{d\,vol}(y)\\ &\leqslant \rho\|\psi\|_{L^1} + \rho e^{-\frac{1}{4}\rho^2}\|\psi\|_{L^\infty} \int\limits_{Q'} e^{\frac{1}{4}g^2}\,\mathrm{d\,vol}\,,\end{aligned}$$

so

$$\mathcal{E}\|g*\psi\|_{C(Q)} \leqslant \rho\big[\|\psi\|_{L^1} + e^{-\frac{1}{4}\rho^2}\|\psi\|_{L^\infty}\sqrt{2}\,\mathrm{vol}\,Q'\big].$$

Taking

$$\rho = 2\sqrt{1+\log\frac{\|\psi\|_{L^\infty}\,\mathrm{vol}\,Q'}{\|\psi\|_{L^1}}}$$

(which is $\geqslant 2$ because $\|\psi\|_{L^1} \leqslant \|\psi\|_{L^\infty}\,\mathrm{vol}\,B(r) \leqslant \|\psi\|_{L^\infty}\,\mathrm{vol}\,Q'$) and using that $2(1+\sqrt{2}) < 5$, we get the desired bound. ∎

A.10.6. Now, we fix $\varphi \geqslant 0$ in $C_0^\infty(B(1))$ with $\int \varphi\,\mathrm{d}\,\mathrm{vol} = 1$. For $r > 0$, let $\varphi_r(x) = r^{-m}\varphi(r^{-1}x)$ and note that $\|\varphi_r\|_{L^1} = 1$ and $\|\varphi_r\|_{L^\infty} \leqslant Cr^{-m}$ for all $r > 0$. Take a sequence $r_j = 2^{-j-1}$ and put $f_j = f * \varphi_{r_j} * \varphi_{r_j}$. Then f_j are C^k Gaussian functions on V_{-2r_j}, and $f_j(x,\,\cdot\,) \to f_x$, as $j \to \infty$, in $L^2(\Omega,\mathcal{P})$ for all $x \in V$. Next, we fix a closed ball $\bar{B} = \bar{B}(x,r) \subset V$ and choose j_0 so large that $\bar{B}(x, r+2r_{j_0}) \subset V$.

Consider the series

$$f_{j_0} + \sum_{j\geqslant j_0}(f_{j+1} - f_j)\,. \tag{A.1}$$

If we show that for every α with $|\alpha| \leqslant k-1$, the expression

$$\mathcal{E}\|\partial^\alpha f_{j_0}\|_{C(\bar{B})} + \sum_{j\geqslant j_0}\mathcal{E}\|\partial^\alpha f_{j+1} - \partial^\alpha f_j\|_{C(\bar{B})}$$

is bounded by $C(\bar{B}, j_0)\,\sqrt{N_{V,k}(K)}$, and that for every α with $|\alpha| = k-1$ and every $\gamma \in (0,1)$, the expression

$$\mathcal{E}\|\partial^\alpha f_{j_0}\|_{\bar{B},\gamma} + \sum_{j\geqslant j_0}\mathcal{E}\|\partial^\alpha f_{j+1} - \partial^\alpha f_j\|_{\bar{B},\gamma}$$

is bounded by $C(\bar{B}, j_0, \gamma)\,\sqrt{N_{V,k}(K)}$, then we will be done because then the series (A.1) will converge in $C^{k-1}(V)$ almost surely, its sum will be a Gaussian function f with the covariance kernel K, and the desired bounds for $\mathcal{E}\|f\|_{\bar{B},k-1+\gamma}$ will hold as well.

A.10.7. For a multi-index α with $|\alpha| \leqslant k$, we write $\partial^\alpha f_{j_0} = \partial^\alpha(f*\varphi_{r_{j_0}})*\varphi_{r_{j_0}}$ and note that the function $g = \partial^\alpha(f*\varphi_{r_{j_0}})$ satisfies $\mathcal{E}\{g(x)^2\} \leqslant N_{V,k}(K)$. So Lemma A.5 yields the bound $\mathcal{E}\|\partial^\alpha f_{j_0}\|_{C(\bar{B})} \leqslant C(\bar{B}, j_0)\,\sqrt{N_{V,k}(K)}$. The interesting part is $\mathcal{E}\|\partial^\alpha f_{j+1} - \partial^\alpha f_j\|_{C(\bar{B})}$. If $|\alpha| \leqslant k-1$, writing

$$\partial^\alpha f_{j+1} - \partial^\alpha f_j = \partial^\alpha(f*(\varphi_{r_{j+1}} - \varphi_{r_j}))*(\varphi_{r_{j+1}} + \varphi_{r_{j+1}})$$

applying the entropy bound with $g = \partial^\alpha(f*(\varphi_{r_{j+1}} - \varphi_{r_j}))$ and $\psi = \varphi_{r_{j+1}} + \varphi_{r_j}$, and recalling that, by A.10., $\mathcal{E}\{|g|^2\} \leqslant Cr_j^2\,N_{V,k}(K)$, we see that

$$\mathcal{E}\big\{\|\partial^\alpha(f_{j+1} - f_j)\|_{C(\bar{B})}\big\} \leqslant C\,r_j\,\sqrt{1+\log(Cr_{j+1}^{-m})}\,\sqrt{N_{V,k}(K)}$$

with some $C = C(\bar{B})$. Since $\sum_j r_j \sqrt{1 + \log(Cr_{j+1}^{-m})} < \infty$, this takes care of the first of the series in A.10. including the uniform norms of the derivatives of f of order up to $k-1$.

A.10.8. To get convergence of the series

$$\sum_{j \geqslant j_0} \mathcal{E}\big\{\|\partial^\alpha f_{j+1} - \partial^\alpha f_j\|_{\bar{B},\gamma}\big\} \tag{A.2}$$

for a multi-index α with $|\alpha| = k-1$, we need the bound for $\mathcal{E}\|\nabla\partial^\alpha f_{j+1} - \nabla\partial^\alpha f_j\|_{C(\bar{B})}$. Note that despite we still have convolutions with mean zero functions in the representation of $\nabla\partial^\alpha f_j$, we cannot use our trick from A.10. because the kernel smoothness is totally exhausted. Thus, we can use only the trivial estimate from A.10. without the factor r_j, and the entropy bounds yields

$$\mathcal{E}\big\{\|\nabla\partial^\alpha f_{j+1} - \nabla\partial^\alpha f_j\|_{C(\bar{B})}\big\} \leqslant C\sqrt{1 + \log(Cr_{j+1}^{-m})}\,\sqrt{N_{V,k}(K)}\,.$$

There is no hope to choose r_j so that these terms will form a convergent series, so there is no chance to show on this way that the $k-1$-st order derivatives are Lipschitz. Fortunately, we do not need that much. All we really need is Hölder continuity.

Using a classical trick, we observe that for any function h that is C^1 in some neighbourhood of $\bar{B}$, and for any two points $x, y \in \bar{B}$, we have[⋆]

$$|h(x)-h(y)| \leqslant \min\big[2\|h\|_{C(\bar{B})}, \|\nabla h\|_{C(\bar{B})}|x-y|\big] \leqslant 2^{1-\gamma}\|h\|_{C(\bar{B})}^{1-\gamma}\,\|\nabla h\|_{C(\bar{B})}^{\gamma}|x-y|^{\gamma}.$$

By Hölder's inequality,

$$\mathcal{E}\Big\{\|\partial^\alpha f_{j+1} - \partial^\alpha f_j\|_{C(\bar{B})}^{1-\gamma}\|\,\|\nabla\partial^\alpha f_{j+1} - \nabla\partial^\alpha f_j\|_{C(\bar{B})}^{\gamma}\Big\}$$
$$\leqslant \Big(\mathcal{E}\|\partial^\alpha f_{j+1} - \partial^\alpha f_j\|_{C(\bar{B})}\Big)^{1-\gamma}\,\Big(\mathcal{E}\|\nabla\partial^\alpha f_{j+1} - \nabla\partial^\alpha f_j\|_{C(\bar{B})}\Big)^{\gamma},$$

and, by the entropy bound, the RHS is

$$\lesssim r_j^{1-\gamma}\big(1 + \log(Cr_{j+1}^{-m})\big)\,\sqrt{N_{V,k}(K)}\,.$$

Hence, the series (A.2) converges and the proof of Kolmogorov's theorem is complete. ∎

[⋆]We use the inequality $\min(a,b) \leqslant a^{1-\gamma}b^{\gamma}$ valid for positive a and b and for $\gamma \in (0,1)$.

A.11. Remarks to Kolmogorov's theorem

A.11.1. Kolmogorov's theorem, as stated and proved, allows us to estimate $\mathcal{E}\|f\|_{\bar{B},k-1+\gamma}$. However, applying then Fernique's theorem, we immediately see that, in assumptions of Kolmogorov's theorem, we can estimate any moment $\mathcal{E}\|f\|^p_{\bar{B},k-1+\gamma}$ we want (that would be exactly as much as we use in this paper), and even prove that the distribution tail $\mathcal{P}\{\|f\|_{\bar{B},k-1+\gamma} > t\}$ at $t \to +\infty$ is Gaussian with controllable bounds.

Indeed, we take $\mathfrak{X} = C^{k-1}(\bar{B})$, $X = f$, fix a countable dense set $B' \subset \bar{B}$, and put

$$\varphi_{\alpha,x}(f) = \partial^\alpha f(x), \qquad |\alpha| \leqslant k-1,\ x \in B',$$
$$\varphi_{\alpha,\gamma,x}(f) = \frac{\partial^\alpha f(x) - \partial^\alpha f(y)}{|x-y|^\gamma}, \qquad |\alpha| = k-1,\ x,y \in B', x \neq y.$$

Note that this is a countable system of linear functionals $\{\varphi_j\} \subset \mathfrak{X}^*$ satisfying the assumptions of Fernique's theorem, and that $\|f\|_{\bar{B},k-1+\gamma} = \sup_j |\varphi_j(f)|$. By Kolmogorov's theorem, there exists a positive constant $\lambda = \lambda(\bar{B}, V, k, \gamma)$ such that

$$\mathcal{P}\{\|f\|_{\bar{B},k-1+\gamma} > \lambda\sqrt{N_{V,k}(K)}\} < \tfrac{1}{4}.$$

Then Fernique's theorem tells us that

$$\mathcal{P}\{\|f\|_{\bar{B},k-1+\gamma} > t\,\lambda\sqrt{N_{V,k}(K)}\} < e^{-at^2}, \qquad t \geqslant 1,$$

whence,

$$\mathcal{P}\{\|f\|_{\bar{B},k-1+\gamma} > t\} < C(B,V,k,\gamma)e^{-c(B,V,k,\gamma)t^2/N_{V,k}(K)}, \qquad t > 0. \tag{A.1}$$

In particular,

$$\mathcal{E}\{\|f\|^p_{\bar{B},k-1+\gamma}\} \leqslant C(B,V,k,\gamma)\,N^{p/2}_{V,k}(K).$$

It is worth mentioning that one can also arrive at estimate (A.1) directly after a certain modification of the proof of Kolmogorov's theorem we gave.

A.11.2. We have to distinguish between C^k Gaussian functions on U and Gaussian functions with $C^{k,k}(U \times U)$ covariance kernels: the former are always the latter but, in general, not vice versa. However, by Kolmogorov's theorem, the continuous Gaussian functions with $C^{k,k}(U \times U)$ covariance kernels fail to be in C^k themselves just barely: they all are in $C^{k-}(U) = \bigcap_{0<\gamma<1} C^{k-1+\gamma}(U)$.

A.11.3. The "convolution approach" to Kolmogorov's theorem allows one to approximate Gaussian functions of finite smoothness by C^∞ ones. This approximation can be used to establish some properties of the kernel.

Using this idea, we will show now that every semi-norm $\|K\|_{Q,k}$ of a positive definite $C^{k,k}(U \times U)$ kernel can be read from the "diagonal":

$$\max_{|\alpha|,|\beta|\leqslant k} \max_{x,y\in Q} \left|\partial_x^\alpha \partial_y^\beta K(x,y)\right| = \max_{|\alpha|\leqslant k} \max_{x\in Q} \left|\partial_x^\alpha \partial_y^\alpha K(x,y)\big|_{y=x}\right| .$$

Indeed, if $|\alpha|, |\beta| \leqslant k-1$, then we can write

$$\left|\partial_x^\alpha \partial_y^\beta K(x,y)\right|^2 = \left|\mathcal{E}\left\{\partial^\alpha f(x)\, \partial^\beta f(y)\right\}\right|^2$$

$$\leqslant \mathcal{E}\left\{[\partial^\alpha f(x)]^2\right\} \mathcal{E}\left\{[\partial^\beta f(y)]^2\right\} = \left(\partial_x^\alpha \partial_y^\alpha K(x,y)\big|_{x=y}\right) \left(\partial_x^\beta \partial_y^\beta K(x,y)\big|_{y=x}\right)$$

for the C^{k-} Gaussian function f with the covariance kernel K, thus estimating the off-diagonal values by the square root of the product of the two corresponding diagonal ones. We cannot do the same estimate directly for the highest order derivatives, but we can consider the convolutions $f * \varphi$ that are infinitely smooth and get the inequality

$$\left|\partial_x^\alpha \partial_y^\beta K_\varphi(x,y)\right|^2 \leqslant \left(\partial_x^\alpha \partial_y^\alpha K_\varphi(x,y)\big|_{x=y}\right) \left(\partial_x^\beta \partial_y^\beta K_\varphi(x,y)\big|_{y=x}\right) \tag{A.2}$$

for the corresponding covariance kernels

$$K_\varphi(x,y) = \iint K(x+x', y+y')\varphi(x')\varphi(y')\, \mathrm{d}\,\mathrm{vol}(x')\mathrm{d}\,\mathrm{vol}(y') .$$

Taking $\varphi_1 \in C_0(B(1))$ and $\varphi(x) = \varphi_r(x) = r^{-m}\varphi_1(r^{-1}x)$, we can pass to the limit

$$\partial_x^\alpha \partial_y^\beta K_{\varphi_r}(x,y) \to \partial_x^\alpha \partial_y^\beta K(x,y) \qquad \text{as } r \to 0 ,$$

for any $|\alpha|, |\beta| \leqslant k$, $x, y \in U$, we conclude that (A.2) holds for K as well.

Of course, here one can also work with the kernel directly, approximating the derivatives by finite difference ratios and passing to the limit in some inequalities for long sums.

A.11.4. The convolutions also facilitate convergence: if the kernels $K_\ell \in C^{k,k}(U \times U)$ are uniformly bounded on compact subsets of $U \times U$ and converge pointwise to some kernel K on $U \times U$, then $(K_\ell)_\varphi \to K_\varphi$ in $C^\infty(U_{-r} \times U_{-r})$ for any $\varphi \in C_0^\infty(B(r))$. If we know, in addition, that for $|\alpha|, |\beta| \leqslant k$, the partial derivatives $\partial_x^\alpha \partial_y^\beta K_\ell(x,y)$ are uniformly locally bounded as well, we can use elementary analysis to show that $K \in C^{k-1,k-1}(U \times U)$ and $\partial_x^\alpha \partial_y^\beta K_\ell(x,y) \to \partial_x^\alpha \partial_y^\beta K(x,y)$ for $|\alpha|, |\beta| \leqslant k-1$ uniformly on compact subsets of $U \times U$. However, in general, it is impossible to conclude that $K \in C^{k,k}(U \times U)$. Surprisingly, this conclusion holds if the limiting kernel K is translation invariant, i.e., $K(x,y) = \kappa(x-y)$ for some $\kappa\colon \mathbb{R}^m \to \mathbb{R}$. This will be shown in the next section.

A.12. Translation-invariant Gaussian functions

A continuous Gaussian function on $\mathbb{R}^m$ is translation-invariant if its covariance kernel $K(x,y)$ depends on $x-y$ only, i.e., $K(x,y)=\kappa(x-y)$ for some continuous positive definite $\kappa\colon \mathbb{R}^m \to \mathbb{R}$. In this case, κ can be written as a Fourier integral of some finite symmetric positive Borel measure ρ on $\mathbb{R}^m$, i.e.,

$$\kappa(x) = \int\limits_{\mathbb{R}^m} e^{2\pi \mathrm{i}(\lambda\, x)}\, \mathrm{d}\rho(\lambda)\,.$$

Consider the Hilbert space $L^2_{\mathtt{H}}(\rho)$ of all Hermitean ($h(-x)=\overline{h(x)}$) functions $h\colon \mathbb{R}^m \to \mathbb{C}$ with $\int |h|^2\,\mathrm{d}\rho < \infty$. The standard $L^2(\rho)$ scalar product $\langle h_1, h_2\rangle = \int h_1 \bar{h}_2\,\mathrm{d}\rho$ is real on $L^2_{\mathtt{H}}(\rho)$. Also, for every $x\in\mathbb{R}^m$, the function $f_x(\lambda)=e^{2\pi\mathrm{i}(\lambda\, x)}$ belongs to $L^2_{\mathtt{H}}(\rho)$ and $\langle f_x, f_y\rangle = \kappa(x-y)$. Finally, the linear span of the functions f_x is dense in $L^2_{\mathtt{H}}(\rho)$. Indeed, if $h\in L^2_{\mathtt{H}}(\rho)$, then $\Phi[h](x)=\langle h, f_x\rangle$ is the Fourier transform of the finite Borel measure $h\,\mathrm{d}\rho$. Hence, it vanishes identically only if $h=0$ ρ-a.e.. Bringing all these observations together, we conclude that

- *the Hilbert space $\mathcal{H}(K)$ coincides with the Fourier image $\mathcal{F}\,L^2_{\mathtt{H}}(\rho)$.*

Now, we discuss the smoothness properties of translation invariant Gaussian functions and covariance kernels. First of all, note that if $K(x,y)=\kappa(x-y)$, then

$$\partial_x^\alpha\, \partial_y^\beta\, K(x,y) = (-1)^{|\beta|}\big(\partial^{\alpha+\beta}\kappa\big)(x-y)\,.$$

Thus, *K is in $C^{k,k}(\mathbb{R}^m\times\mathbb{R}^m)$ if and only if $\kappa\in C^{2k}(\mathbb{R}^m)$, that is, if and only if,*

$$\int\limits_{\mathbb{R}^m} |\lambda|^{2k}\,\mathrm{d}\rho(\lambda) < \infty\,. \tag{A.1}$$

We end this section with a curious and quite useful observation:

- *if a sequence of positive definite kernels $K_\ell \in C^{k,k}(U_\ell, U_\ell)$ with U_ℓ exhausting $\mathbb{R}^m$ has a pointwise translation invariant limit $\kappa(x-y)$ and $\partial_x^\alpha\,\partial_y^\alpha\, K_\ell(x,y)\big|_{x=y=0}$ and stays bounded for $|\alpha|\leqslant k$, then $\kappa\in C^{2k}(\mathbb{R}^m)$.*

P r o o f. For $\varphi\in C_0^\infty(\mathbb{R}^m)$, $\varphi(-x)=\varphi(x)$ and put $K_\varphi(x,y)=(\kappa*\varphi*\varphi)(x-y)$. Since $\kappa=\mathcal{F}\rho$ implies that $\kappa*\varphi*\varphi=\mathcal{F}\rho_\varphi$, where $\mathrm{d}\rho_\varphi=\widehat{\varphi}^{\,2}\,\mathrm{d}\rho$, we see that

$$\begin{aligned}(-1)^k \sum_{|\alpha|=k} \partial^{2\alpha}(\kappa*\varphi*\varphi)(0) &= (2\pi)^{2k} \sum_{|\alpha|=k} \int\limits_{\mathbb{R}^m} \lambda_1^{2\alpha_1}\,\ldots\,\lambda_m^{2\alpha_m}\,\mathrm{d}\rho_\varphi(\lambda)\\ &= (2\pi)^{2k} \int\limits_{\mathbb{R}^m} |\lambda|^{2k}\,\mathrm{d}\rho_\varphi(\lambda)\,.\end{aligned}$$

If we know in advance that $\kappa \in C^{2k}(\mathbb{R}^m)$, then the quantities $\partial^{2\alpha}(\kappa * \varphi * \varphi)(0)$, $|\alpha| = k$, are uniformly bounded when φ runs over even non-negative C_0^∞ functions supported on a small ball centered at the origin and normalized by $\int\limits_{\mathbb{R}^m} \varphi \, \mathrm{d}\,\mathrm{vol} = 1$. Then, taking as before, $\varphi_r(x) = r^{-m}\varphi(r^{-1}x)$, letting $r \to 0$, and applying Fatou's lemma, we get

$$\int\limits_{\mathbb{R}^m} |\lambda|^{2k} \, \mathrm{d}\rho(\lambda) \leqslant \varliminf_{r\to 0} \int\limits_{\mathbb{R}^m} |\lambda|^{2k} \, \mathrm{d}\rho_{\varphi_r}(\lambda) < \infty \,.$$

Now, observe that the quantities $\partial^{2\alpha}(\kappa * \varphi * \varphi)(0)$ stay uniformly bounded even if $\kappa(x-y)$ is a pointwise limit of $C^{k,k}$ positive definite symmetric kernels $K_\ell(x,y)$ that are defined only in a neighbourhood of the origin in $\mathbb{R}^m \times \mathbb{R}^m$ and that have uniformly bounded derivatives $\partial_x^\alpha \, \partial_y^\alpha \, K_\ell(x,y)\big|_{x=y=0}$. So, in this case, we still get

$$\int\limits_{\mathbb{R}^m} |\lambda|^{2k} \, \mathrm{d}\rho(\lambda) < \infty \,,$$

completing the proof of our observation. ∎

B. Proof of the Fomin–Grenander–Maruyama Theorem

Assuming that ρ has no atoms, we need to show that if $A \in \mathfrak{S}$ is a set satisfying $\gamma\big((\tau_v A) \triangle A\big) = 0$ for every $v \in \mathbb{R}^m$, then $\gamma(A)$ is either 0 or 1. As before, we use the notation $(\tau_v G)(u) = G(u+v)$, where $v \in \mathbb{R}^m$ and $G \in X$.

Since $\mathfrak{S}$ is generated by the intervals $I(u;a,b)$, given $\varepsilon > 0$, we can take finitely many points $u_1, \dots, u_n \in \mathbb{R}^m$ and a Borel set $B \subset \mathbb{R}^n$ so that $\gamma\{A \triangle P\} < \varepsilon$, where

$$P = P(u_1, \dots, u_n; B) \stackrel{\mathrm{def}}{=} \big\{G \in X \colon (G(u_1), \dots, G(u_n)) \in B\big\} \,.$$

Without loss of generality, we may assume that the distribution of the Gaussian vector $\big(G(u_1), \dots, G(u_n)\big)$ is non-degenerate*. In this case, we can write

$$\gamma\big(P(u_1, \dots, u_n; B)\big) = (2\pi)^{-n/2} \, \big(\det \Lambda\big)^{-\frac{1}{2}} \int\limits_B e^{-\frac{1}{2}(\Lambda^{-1} t\, t)} \, \mathrm{d}\,\mathrm{vol}(t) \,,$$

*Otherwise one of the values, say, $G(u_n)$, is a linear combination of other values with probability 1. If $G(u_n) = \sum_{j=1}^{n-1} c_j G(u_j)$ is such a representation, then

$$\gamma\big(\big\{G \in X \colon (G(u_1), \dots, G(u_n)) \in B\big\} \triangle \big\{G \in X \colon (G(u_1), \dots, G(u_{n-1})) \in B'\big\}\big) = 0 \,,$$

where $B' = \big\{(t_1, \dots, t_{n-1}) \in \mathbb{R}^{n-1} \colon \big(t_1, \dots, t_{n-1}, \sum_{j=1}^{n-1} c_j t_j\big) \in B\big\}$ is a Borel set in $\mathbb{R}^{n-1}$, so we can remove the point u_n from the consideration at no cost.

where $\Lambda = \big(k(u_i - u_j)\big)_{i,j=1}^n$ is the covariance matrix of the vector $\big(G(u_1), \dots, G(u_n)\big)$. As before, we denote by k the Fourier integral of the spectral measure ρ.

Since $\tau_v P = P(u_1 + v, \dots, u_n + v; B)$, we have

$$P \cap \tau_v P = P(u_1, \dots, u_n, u_1 + v, \dots, u_n + v; B \times B)\,.$$

Then

$$\gamma\big(P \cap \tau_v P\big) = (2\pi)^{-n} \big(\det \widetilde{\Lambda}\big)^{-\frac{1}{2}} \int\limits_{B \times B} e^{-\frac{1}{2}(\widetilde{\Lambda}^{-1}(v)\widetilde{t}\,\widetilde{t})}\, \mathrm{d\,vol}(\widetilde{t}\,)$$

where

$$\widetilde{\Lambda}(v) = \begin{pmatrix} \Lambda & \Theta(v) \\ \Theta^*(v) & \Lambda \end{pmatrix} \qquad \text{with } \Theta_{i,j}(v) = k(u_i - v - u_j)\,.$$

Note that the matrix $\widetilde{\Lambda}(v)$ is invertible and $\big(\widetilde{\Lambda}(v)\big)^{-1}$ is close to $\begin{pmatrix} \Lambda^{-1} & 0 \\ 0 & \Lambda^{-1} \end{pmatrix}$ if $\|\Theta(v)\|$ is small enough.

Next, we observe that we can choose a sequence $v_\ell \in \mathbb{R}^m$ so that $\|\Theta(v_\ell)\| \to 0$ as $\ell \to \infty$. Indeed, letting $\Delta = \max_{i,j} |u_i - u_j|$, we have

$$\frac{1}{\operatorname{vol} B(R)} \int\limits_{B(R)} \sum_{i,j} k(u_i - v - u_j)^2\, \mathrm{d\,vol}(v) \leqslant \frac{n^2}{\operatorname{vol} B(R)} \int\limits_{B(R+\Delta)} k^2\, \mathrm{d\,vol}\,,$$

while by Wiener's theorem [16, VI.2.9], the absence of atoms in ρ is equivalent to

$$\lim_{R \to \infty} \frac{1}{\operatorname{vol} B(R)} \int\limits_{B(R)} k^2\, \mathrm{d\,vol} = 0\,.$$

Then, using the dominated convergence theorem, we conclude that

$$\lim_{\ell \to \infty} \gamma\big(P \cap \tau_{v_\ell} P\big) = \gamma\big(P\big)^2\,.$$

Recalling that $A \cap \tau_{v_\ell} A = A$ up to γ-measure 0, we obtain

$$\gamma(A) = \gamma(A \cap \tau_{v_\ell} A) \leqslant \gamma(P \cap \tau_{v_\ell} P) + 2\varepsilon \overset{\ell \to \infty}{\to} \gamma(P)^2 + 2\varepsilon \leqslant \big[\gamma(A)\big]^2 + 2\varepsilon\,.$$

Since $\varepsilon > 0$ is arbitrary, we conclude that $\gamma(A) \leqslant \gamma(A)^2$, whence $\gamma(A) = 0$ or $\gamma(A) = 1$. ∎

C. Condition ($\rho 4$)

Here, we collect several observations that, in many instances, help to verify condition ($\rho 4$). Recall that this condition asserts that

- *there exist a finite compactly supported Hermitian measure μ with* $\operatorname{spt}(\mu) \subset \operatorname{spt}(\rho)$ *and a bounded domain $D \subset \mathbb{R}^m$ such that $\mathcal{F}\mu\big|_{\partial D} < 0$ and $(\mathcal{F}\mu)(u_0) > 0$ for some $u_0 \in D$.*

Throughout this section, we assume that condition ($\rho 3$) is satisfied, that is, that the measure ρ is not supported on a hyperplane in $\mathbb{R}^m$.

C.1. Quadratic hypersurface criterion

The support of any measure ρ <u>not</u> satisfying condition ($\rho 4$) must be contained in a quadratic variety $A\lambda\,\lambda = b$, where A is an $m \times m$ symmetric matrix and $b \in \mathbb{R}^m$.

P r o o f. Suppose that $\operatorname{spt}(\rho)$ is not contained in any quadratic variety of the above form. Then $\frac{1}{2}m(m+1)+1$-dimensional vectors

$$v(\lambda) = \big\{1,\ \lambda(i)\lambda(j)\colon 1 \leqslant i \leqslant j \leqslant m\big\}, \quad \lambda \in \operatorname{spt}(\rho),$$

span $\mathbb{R}^{\frac{1}{2}m(m+1)+1}$ (here $\lambda(i)$ denotes the i-th coordinate of λ). Then we can create two finite linear combinations of cosines:

$$f(x) = \sum_{\lambda\in\operatorname{spt}(\rho)} a_\lambda \cos\big(2\pi\lambda\, x\big), \quad g(x) = \sum_{\lambda\in\operatorname{spt}(\rho)} b_\lambda \cos\big(2\pi\lambda\, x\big),$$

such that

$$f(0) = 1, \quad (D^2 f)(0) = 0\,,$$

and

$$g(0) = 0, \quad (D^2 g)(0) = I\,,$$

where $D^2 f$ is the matrix with the entries $\partial^2_{x_i\,x_j} f$ and I is the unit matrix. Note that we also automatically have $Df(0) = Dg(0) = 0$, Then the function $h = \varepsilon^2 f - g$ will satisfy $h(0) = \varepsilon^2$ and $h(x) < 0$ on $\big\{|x| = 2\varepsilon\big\}$, provided that ε is small enough. ∎

C.2. Pjetro Majer's interior point criterion

The next observation is due to Pietro Majer.

Let the interior of the convex hull of $\operatorname{spt}(\rho)$ *contain a point from* $\operatorname{spt}(\rho)$. *Then condition ($\rho 4$) is satisfied.*

In particular, condition ($\rho 4$) is satisfied when $0 \in \operatorname{spt}(\rho)$.

P r o o f. Let υ be such a point. Since υ lies in the interior of the convex hull of $\operatorname{spt}(\rho)$, there are $\lambda_1, \dots, \lambda_n \in \operatorname{spt}(\rho)$ that span the whole space $\mathbb{R}^m$, such that

$$\upsilon = \sum\nolimits_i t_i\lambda_i\,, \qquad t_i \geqslant 0,\ \sum\nolimits_i t_i = \alpha < 1\,.$$

Consider the function

$$f(x) = \sum\nolimits_i b_i \cos\big(2\pi\lambda_i\, x\big) - \cos(2\pi\upsilon\, x)$$

with $b_i = \alpha t_i + n^{-1}(1-\alpha^2+\varepsilon)$, where $\varepsilon > 0$. Then, for $x \to 0$,

$$f(x) = \Big[\sum\nolimits_i b_i - 1\Big] - 2\pi^2 \Big[\sum\nolimits_i b_i (\lambda_i\, x)^2 - (\upsilon\, x)^2\Big] + o(|x|^2)\,.$$

In particular,

$$f(0) = \sum\nolimits_i b_i - 1 = \varepsilon > 0\,.$$

Next, we note that

$$\begin{aligned}(\upsilon\, x)^2 = \Big(\sum\nolimits_i t_i \lambda_i\, x\Big)^2 &= \Big(\sum\nolimits_i t_i^{1/2} t_i^{1/2} \lambda_i\, x\Big)^2 \\ &\leqslant \Big(\sum\nolimits_i t_i\Big)\Big(\sum\nolimits_i t_i (\lambda_i\, x)^2\Big) = \alpha \sum\nolimits_i t_i (\lambda_i\, x)^2\,.\end{aligned}$$

Now, suppose that x belongs to the non-degenerate ellipsoid

$$E = \Big\{ \sum\nolimits_i (\lambda_i\, x)^2 = \frac{\varepsilon n}{\pi^2(1-\alpha^2)} \Big\}\,.$$

Since $\lambda_1, \dots, \lambda_n$ span $\mathbb{R}^n$, we have

$$|x|^2 = O(\varepsilon)\,, \qquad \varepsilon \to 0,\ x \in E\,.$$

Therefore, for $x \in E$ and $\varepsilon \to 0$, we have

$$\begin{aligned} f(x) &\leqslant \varepsilon - 2\pi^2 \sum\nolimits_i (b_i - \alpha t_i)(\lambda_i\, x)^2 + o(\varepsilon) \\ &= \varepsilon - \frac{2\pi^2(1-\alpha^2+\varepsilon)}{n} \sum\nolimits_i (\lambda_i\, x)^2 + o(\varepsilon) < \varepsilon - 2\varepsilon + o(\varepsilon) < 0\,, \end{aligned}$$

completing the proof. ∎

C.3. Analytic closure criterion

Our last observation is that

- the requirement $\mathrm{spt}(\mu) \subset \mathrm{spt}(\rho)$ in condition $(\rho 4)$ can be relaxed to the requirement $\mathrm{spt}(\mu) \subset \mathrm{spt}_{\mathtt{r.a.}}(\rho)$ where $\mathrm{spt}_{\mathtt{r.a.}}(\rho)$ is the intersection of all real-analytic varieties containing $\mathrm{spt}(\rho)$.

Note that every quadratic variety is an analytic variety as well, so if $\mathrm{spt}(\rho) \subset V$ then $\mathrm{spt}_{\mathtt{r.a.}}(\rho) \subset V$ too. Sometimes, $\mathrm{spt}_{\mathtt{r.a.}}(\rho)$ is much larger that $\mathrm{spt}(\rho)$ and satisfy the assumption of C.2. (or some other condition sufficient for establishing $(\rho 4)$ without $\mathrm{spt}(\rho)$ doing so). For instance, suppose that $m = 2$, $S \subset \mathbb{R}^2$ is the unit circumference, and $\mathrm{spt}(\rho) \subset S$ is an infinite set. Since infinite subsets of S

are uniqueness sets for real-analytic functions on S, we see that $\operatorname{spt}_{\mathtt{r.a.}}(\rho) = S$. Then, taking $\mu = m_1$ (the Lebesgue measure on S), we conclude that condition $(\rho 4)$ is satisfied.

P r o o f. Let $Q \subset \mathbb{R}^m$ be a compact set. Consider two linear subspaces of the space $C(Q)$ of *real-valued* continuous functions on Q:

$$X = \big\{\mathcal{F}\mu\colon \mu \text{ is Hermitian, compactly supported}, \operatorname{spt}(\mu) \subset \operatorname{spt}(\rho)\big\}$$

and

$$X_{\mathtt{r.a.}} = \big\{\mathcal{F}\mu\colon \mu \text{ is Hermitian, compactly supported}, \operatorname{spt}(\mu) \subset \operatorname{spt}_{\mathtt{r.a.}}(\rho)\big\}.$$

We need to show that the $C(Q)$-closure of X contains $X_{\mathtt{r.a.}}$. We will be using a simple duality argument. Suppose that a signed measure ν supported by Q annihilates X, that is,

$$\int_Q \big(\mathcal{F}\mu\big)\,\mathrm{d}\nu = 0 \qquad \text{for all admissible } \mu\,.$$

Taking $\mu = \frac{1}{2}(\delta_\lambda + \delta_{-\lambda})$, $\lambda \in \operatorname{spt}(\rho)$, we find that the cosine-transform

$$\big(\mathcal{C}\nu\big)(\lambda) = \int_Q \cos\big(2\pi\lambda\, x\big)\,\mathrm{d}\nu(x)$$

vanishes on $\operatorname{spt}(\rho)$. However, $\mathcal{C}\nu$ is an entire function, which is real on $\mathbb{R}^m$. Hence, if it vanishes on $\operatorname{spt}(\rho)$, it must also vanish on $\operatorname{spt}_{\mathtt{r.a.}}(\rho)$. Therefore, the measure ν annihilates the subspace $X_{\mathtt{r.a.}}$ as well. ∎

References

[1] *M.E. Becker*, Multiparameter Groups of Measure-Preserving Tranformations: a Simple Proof of Wiener's Ergodic Theorem. — *Ann. Prob.* **9** (1981), 504–509.

[2] *D. Beliaev and Z. Kereta*, On Bogomolny–Schmit Conjecture. — *J. Phys. A: Math. Theor.* **46** (2013), 45503. `arXiv:1310.2747`

[3] *S. Bochner*, Vorlesungen Uber Fouriersche Integrale. Leipzig, 1932.

[4] *V. Bogachev*, Gaussian Measures. *Mathematical Surveys and Monographs* **62** AMS, Providence, RI, 1998.

[5] *E. Bogomolny and C. Schmit*, Percolation Model for Nodal Domains of Chaotic Wave Functions. — *Phys. Rev. Lett.* **88** (2002), 114102. `arXiv:nlin/0110019v1`; Random Wavefunctions and Percolation, *J. Phys. A* **40** (2007), 14033–14043. `arXiv:0708.4335v1`

[6] *J. Bourgain*, On Toral Eigenfunctions and the Random Wave Model. — *Israel J. Math.* **201** (2014), 611–630. arXiv:1303.2881.

[7] *J. Bourgain and Z. Rudnick*, Restriction of Toral Eigenfunctions to Hypersurfaces and Nodal Sets. — *Geom. Funct. Anal.* **22** (2012), 878–937. arXiv:1105.0018v1.

[8] *Y. Canzani and P. Sarnak*, On the Topology of the Zero Sets of Monochromatic Random Waves. arXiv:1412.4437.

[9] *Y. Fyodorov, A. Lerario, and E. Lundberg*, On the Number of Connected Components of Random Algebraic Hypersurfaces. — *J. Geom. Phys.* **95** (2015), 1–20. arXiv:1404.5349.

[10] *D. Gayet and J.-Y. Welschinger*, Lower Estimates for the Expected Betti Numbers of Random Real Hypersurfaces. — *J. London Math. Soc. (2)* **90** (2014), 105–120. arXiv:1303.3035

[11] *D. Gayet and J.-Y. Welschinger*, Expected Topology of Random Real Algebraic Submanifolds. *J. Inst. Math.* Jussieu (2015). arXiv:1307.5287.

[12] *D. Gayet and J.-Y. Welschinger*, Universal Components of Random Zero Sets. arXiv:1503.01582.

[13] *U. Grenander*, Stochastic Processes and Statistical Inference. — *Ark. Mat.* **1** (1950), 195–277.

[14] *S. Janson*, Gaussian Hilbert Spaces. Cambridge Univ. Press, Cambridge, 1997.

[15] *M. Hairer*, An Introduction to Stochastic PDEs. arXiv:0907.4178

[16] *Y. Katznelson*, An Introduction to Harmonic Analysis. Cambridge Univ. Press, Cambridge, 2004.

[17] *H. Komatsu*, Prof. Yosida's Proof of the Plancherel and the Bochner Theorems for Locally Compact Abelian Groups. — *J. Math. Kyoto Univ.* **41** (2001), 739–747.

[18] *K. Konrad*, Asymptotic Statistics of Nodal Domains of Quantum Chaotic Billiards in the Semiclassical Limit. Senior Thesis, Dartmouth College, 2012.

[19] *E. Kostlan*, On the Expected Number of Real Roots of a System of Random Polynomial Equations. *Foundations of Computational Mathematics* (Hong Kong, 2000), 149–188, World Sci. Publ., River Edge, New York, 2002.

[20] *P. Kurlberg and I. Wigman*, Non-universality of the Nazarov–Sodin Constant. — *C. R. Math. Acad. Sci. Paris* **353** (2015), 101–104. arXiv:1406.7449

[21] *H.J. Landau and L.A. Shepp*, On the Supremum of a Gaussian Process. — *Sankhy, Ser. A* **32** (1970), 369–378.

[22] *A. Lerario and E. Lundberg*, Statistics on Hilbert's Sixteenth Problem. Int. Math. Res. Notices (2015). arXiv:1212.3823

[23] *T.L. Malevich*, Contours That Arise When the Zero Level is Crossed by Gaussian Fields. — *Izv. Akad. Nauk UzSSR Ser. Fiz.-Mat. Nauk* **16** (1972), 20–23.

[24] *M. Nastasescu*, The Number of Ovals of a Real Plane Curve. Senior Thesis, Princeton, 2011.

[25] *F. Nazarov and M. Sodin*, On the Number of Nodal Domains of Random Spherical Harmonics. — *Amer. J. Math.* **131** (2009), 1337–1357. arXiv:0706.2409v1

[26] *Y. Rozenshein*, The Number of Nodal Components of Arithmetic Random Waves. Master Thesis, Tel Aviv University, 2015. arXiv:1604.00638

[27] *P. Sarnak*, Letter to B. Gross and J. Harris on Ovals of Random Plane Curves, May 2011. http://publications.ias.edu/sarnak/paper/510

[28] *P. Sarnak and I. Wigman*, Topologies of Nodal Sets of Random Band Limited Functions. arXiv:1312.7858

[29] *M. Sodin*, Lectures on Random Nodal Portraits. In: Probability and Statistical Physics in St. Petersburg. Proc. Symp. Pure Math. 91. Amer. Math. Soc., Providencs RI, 2016.
Available online http://www.math.tau.ac.il/ sodin/SPB-Lecture-Notes.pdf

[30] *B. Tsirelson*, Gaussian Geasures. *Lecture Notes*. Tel Aviv University, Fall 2010.
http://www.tau.ac.il/~tsirel/Courses/Gauss3/main.html