MATHEMATICAL ENGINEERING
TECHNICAL REPORTS

Validity of the expected Euler characteristic heuristic

Jonathan Taylor, Akimichi Takemura
and Robert Adler

METR 2003–26 August 2003

DEPARTMENT OF MATHEMATICAL INFORMATICS
GRADUATE SCHOOL OF INFORMATION SCIENCE AND TECHNOLOGY
THE UNIVERSITY OF TOKYO
BUNKYO-KU, TOKYO 113-8656, JAPAN

WWW page: http://www.i.u-tokyo.ac.jp/mi/mi-e.htm

Validity of the expected Euler characteristic heuristic

Jonathan Taylor * Akimichi Takemura† Robert Adler‡

August, 2003

Abstract

We study the accuracy of the expected Euler characteristic approximation to the distribution of the maximum of a smooth, centered, unit variance Gaussian process f. Using a point process representation of the error, valid for arbitrary smooth processes, we show that the error is in general exponentially smaller than any terms in the approximation. We also give a lower bound on this exponential rate of decay in terms of the maximal variance of a family of Gaussian processes f^x, derived from the original process f.

1 Introduction

In this paper, we study the (arguably) well-known expected Euler characteristic approximation to

$$\mathrm{P}\left(\sup_{x\in M} f(x) \geq u\right) \tag{1}$$

where f is the restriction to M of $\widehat{f}$, a C^2 process on a C^3 manifold $\widehat{M}$, and M is an embedded piecewise C^3 submanifold of $\widehat{M}$.

When the process $\widehat{f}$ is Gaussian with zero mean and has unit variance, the expected Euler characteristic approximation is given by

$$\widehat{\mathrm{P}}\left(\sup_{x\in M} f(x) \geq u\right) = \mathrm{E}\left(\chi\left(M \cap \widehat{f}^{-1}[u,+\infty)\right)\right) = \sum_{j=0}^{\dim M} \mathcal{L}_j(M)(2\pi)^{-(j+1)/2}\int_u^\infty H_j(r)\, \mathrm{e}^{-r^2/2}\, dr \tag{2}$$

where H_j is the j-th Hermite polynomial, $\chi\left(M \cap \widehat{f}^{-1}[u,+\infty)\right)$ is the Euler characteristic of the excursion of f above the level u and the $\mathcal{L}_j(M)$ are the intrinsic volumes, or

*Department of Statistics, Sequoia Hall, Stanford, CA 94305-4065, U.S.A. e-mail: jtaylor@stat.stanford.edu

†Graduate School of Information Science and Technology, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN, e-mail: takemura@stat.t.u-tokyo.ac.jp

‡Faculty of Industrial Engineering and Management, Technion, Haifa 32000, Israel. e-mail: robert@ieadler.technion.ac.il

Lipschitz-Killing curvatures of the parameter space M [17, 19], measured with respect to a Riemannian metric induced by f, which is discussed below in Section 2. In [17] only the special case of finite Karhunen-Loève processes (see below) was treated and in [19] the case of a manifolds with smooth boundaries were treated. The result, proved in the generality above, appears in [1]. With a slight abuse of notation, we have used $\widehat{\mathrm{P}}$ to denote our approximation to (1), i.e. we are using $\widehat{\mathrm{P}}$ in the statistical sense of a "point estimator" and not as some alternative probability measure.

As noted above, the case when $\widehat{M}$ is C^∞ and $\widehat{f}$ is a centered, unit variance, finite Karhunen-Loève expansion process was studied in [17]. These assumptions imply that there exists a map $\varphi : \widehat{M} \to S(\mathbb{R}^n)$ where $S(\mathbb{R}^n)$ is the unit sphere in $\mathbb{R}^n$ and a random vector $\xi(\omega) \sim N(0, I_{n\times n})$ such that

$$\widehat{f}(x,\omega) = \langle \varphi(x), \xi(\omega)\rangle_{\mathbb{R}^n} = \sum_{j=1}^{n} \xi_j(\omega)\varphi_j(x).$$

In this setting, without loss of generality, we can assume that $\widehat{M}$ is an embedded submanifold of $S(\mathbb{R}^n)$ and φ is just the inclusion map. Using the volume of tubes approach [20, 15], it was shown in [17] that if M is a piecewise C^∞ submanifold of $S(\mathbb{R}^n)$ the error in the above approximation above is bounded by

$$\begin{aligned}\left|\mathrm{P}\left(\sup_{x\in M} f(x) \geq u\right) - \widehat{\mathrm{P}}\left(\sup_{x\in M} f(x) \geq u\right)\right| &\leq \frac{C}{\Gamma\left(\frac{n}{2}\right) 2^{(n-2)/2}} \int_{u/\cos\theta_c(M)}^{\infty} w^{n-1} \mathrm{e}^{-w^2/2}\, dw \\ &= C \times \mathrm{P}\left(\chi_n^2 \geq u^2/\cos^2\theta_c(M)\right)\end{aligned} \tag{3}$$

where $\theta_c(M)$, is a geometric quantity known as the critical radius of M [9, 10, 15, 18].

In another setting, when $\widehat{f}$ is "almost" isotropic on $\mathbb{R}^k$, then, with some additional assumptions on M (cf. Theorem 4.5.2 in [4]) Piterbarg [14] showed using the "double-sum" method that the error in using the expected Euler characteristic approximation is bounded by

$$\left|\mathrm{P}\left(\sup_{x\in M} f(x) \geq u\right) - \widehat{\mathrm{P}}\left(\sup_{x\in M} f(x) \geq u\right)\right| \leq C\mathrm{e}^{-\alpha u^2/2} \tag{4}$$

for some $\alpha > 1$, though no expression for α is given.

Note that both bounds show that the error in approximating (1) is *exponentially* smaller than all terms in the expected Euler characteristic approximation. While undeniably useful, these two situations do not cover all possibilities. Referring to (4), certainly not every smooth Gaussian process of interest is isotropic, nor are the conditions required of M easily interpretable (cf. Definition 4.5.1 in [4]). Referring to (3), while every Gaussian process does admit an *infinite* orthogonal expansion, cf. [3]

$$f(x,\omega) = \sum_{j=1}^{\infty} \xi_j(\omega)\varphi_j(x) \tag{5}$$

through its reproducing kernel Hilbert space (RKHS), it is clear that substituting $n = \infty$ into (3) is meaningless. In fact, the situation is even worse in that the two cases do not even overlap: an isotropic field restricted to a bounded domain $T \subset \mathbb{R}^k$ *cannot* have a finite Karhunen-Loève expansion [16]!

This brings us to the main result of this work, Theorem 4.3, which, when f is a constant variance Gaussian process as in the works cited above, provides bounds for the error in using the expected Euler characteristic approximation to (1). Specifically, when f has unit variance, we show that

$$\liminf_{u\to\infty} -u^{-2} \log \left| \mathrm{P}\left(\sup_{x\in M} f(x) \geq u\right) - \widehat{\mathrm{P}}\left(\sup_{x\in M} f(x) \geq u\right) \right| \geq \frac{1}{2} + \frac{1}{2\sigma_c^2(f)}. \tag{6}$$

Above, the "critical variance" $\sigma_c^2(f)$ depends on the variance of an auxiliary family of Gaussian processes $(f^x)_{x\in M}$, defined in (14) below.

An alternative approximation to (1) is to use the expected number of (extended outward) local maxima [8]. The term "approximation" is used in a somewhat loose sense, as, to the authors' knowledge, there are no generally applicable known closed form expressions for the number of extended outward local maxima of a smooth process. The only results known are ones which relate the asymptotic behaviour of the expected number of local maxima of a Gaussian field on a manifold *without* boundary (which renders the qualifier "extended outward" unnecessary) to the expected Euler characteristic approximation, cf. [8]. Nevertheless, *if* one could compute the expected number of local maxima *exactly*, as one can the expected Euler characteristic in certain cases, one might expect to get a better approximation to (1). Virtually identical arguments to those used in this paper show that when M is a manifold without boundary and f is Gaussian with constant variance, on an exponential scale the error in the approximations are equivalent, though, in the interest of brevity, we do not pursue this here. When the manifold M has boundary, the situation is more subtle and it may indeed be the case that the expected number of extended outward local maxima may be more accurate on an exponential scale.

The critical variance $\sigma_c^2(f)$ is closely related to the critical radius appearing in (3). Specifically, when f is a centered, unit variance, finite Karhunen-Loève expansion process, and M is a manifold without boundary, it is proven in Lemma 5.1 that

$$\sigma_c^2(f) = \cot^2 \theta_c(M)$$

where $\theta_c(M)$ is the critical radius of M, mentioned above and used in [18].

We note that, while (3) *is* an *explicit bound* for finite Karhunen-Loève expansion Gaussian processes, it is not sharp, nor generally applicable. In particular it depends on the (generally unknown) dimension of the sphere into which M is embedded, i.e. the dimension of the sphere in which M sits. In a companion paper [12], when f is centered, unit variance, finite Karhunen-Loève expansion Gaussian process, the asymptotic error as $u \to \infty$ is evaluated using a Laplace approximation, rather than just the exponential behaviour, which is the topic of this paper.

Our main result, Theorem 4.3, is formally an application of Theorem 3.3 to the case when f is Gaussian with constant variance. Theorem 3.3 gives a bound for the error of the expected Euler characteristic approximation for the restriction of an *arbitrary* suitably regular (cf. [2, 19]) process $\widehat{f}$ on a C^3 manifold $\widehat{M}$ to any embedded piecewise C^2 submanifold $M \subset \widehat{M}$. Theorem 3.3 is, to the authors' knowledge, the only available bound for the error in the expected Euler characteristic approximation for arbitrary, suitably regular, smooth random fields and should prove useful in studying the accuracy of the Euler characteristic approximation to non-Gaussian fields [6, 7, 21, 22]. The case of non-constant variance Gaussian processes can be handled similarly to the case of constant variance processes and is presented in [1]. The analogy to (3) for non-constant variance Gaussian fields using a variant of the volume of tubes approach is presented in [13].

Another noteworthy feature to our approach is that it is a *direct* approach to determining the error in using the Euler characteristic approximation. This should be contrasted with the bounds (3) and (4) which were both arrived at indirectly. That is, two different approaches were used to approximate the probability (1), and the terms in the approximations were shown to be the terms in (2), so bounds derived from these approaches applied to the expected Euler characteristic approximation (cf. [4]).

The proof of Theorem 3.3 depends on a point set representation for the global maximizers of $h = \widehat{h}_{|M}$, the restriction to M of a smooth deterministic function $\widehat{h} : \widehat{M} \to \mathbb{R}$, above the level u. Of course, there is a trivial, and not very useful, point set representation of the set of maximizers of h above the level u:

$$\left\{ x \in M : h(x) = \max_{y \in M} h(y),\ h(x) \geq u \right\}.$$

Lemma 2.2 gives an alternative point set representation of the maximizers using an auxiliary family of functions $(h^x)_{x \in M}$. Once we have a point set representation of the maximizers of a smooth function, we apply a "meta-theorem" for the density of point processes arising from smooth processes of [2, 4, 5, 19]. If the global maximizer of the process f is almost surely unique, the total mass of the density of the point process of maximizers above the level u is therefore just (1). The auxiliary family of processes $(f^x)_{x \in M}$ mentioned above is, in some sense, the stochastic analogue of $(h^x)_{x \in M}$ in the deterministic setting.

The organization of the paper follows. Sections 2 lays out the regularity conditions needed for Theorem 3.3, and reviews some notions of piecewise smooth manifolds. Theorem 3.3 is proved in Section 3, and Section 4 deals with the unit variance Gaussian case, where Theorem 4.3 is proven. We conclude in Section 5 with some examples, specifically we compute $\sigma_c^2(f)$ for stationary processes on $\mathbb{R}$ and isotropic processes on $\mathbb{R}^k$ restricted to compact convex subsets.

2 Suitably regular processes on piecewise C^2 manifolds

In this section we describe the class of processes to which Theorem 3.3 will apply. Before setting out our assumptions, we recall some basic facts about piecewise C^l ($l \geq 2$)

submanifolds of an ambient C^j $(j \geq l)$ manifold $\widehat{M}$.

A k-dimensional piecewise C^2 manifold has a unique decomposition into C^2 i-dimensional manifolds, $0 \leq i \leq k$

$$M = \bigcup_{i=0}^{k} \partial M_i.$$

Associated to every point $x \in M$, is its support cone in $\widehat{M}$

$$\mathcal{S}_x M = \Big\{ X_x \in T_x\widehat{M} : \exists\, c \in C^1\left((-\delta,\delta),\widehat{M}\right), \\ c(t) \in M \; \forall t \in [0,\delta), c(0) = x, \dot{c}(0) = X_x \Big\}.$$

In words, the support cone of M at x is the set of all directions in which a smooth curve can leave x into $\widehat{M}$, but, in an infinitesimally small time period, still remain in M.

Piecewise smooth manifolds are required to have the additional property that they are locally approximated by $\mathcal{S}_x M$ in the sense that for every $x \in M$ there exists a diffeomorphism

$$\varphi_x : T_x\widehat{M} \to \widehat{M}$$

which, when restricted to $\mathcal{S}_x M$ maps any sufficiently small neighbourhood of the origin to a neighbourhood of $x \in M$. This condition, for instance, rules out cusps in M as can occur when two manifolds intersect non-transversally.

When $\widehat{M}$ is endowed with a Riemannian metric, it is possible to define the dual cone of $\mathcal{S}_x M$. The following Riemannian metric will be essential to our analysis: an L^2 differentiable process induces a natural metric on $\widehat{M}$ given by (cf. [19])

$$\widehat{g}_x(X_x, Y_x) \triangleq \mathrm{Cov}(X_x\widehat{f}, Y_x\widehat{f}).$$

The dual cone of $\mathcal{S}_x M$, $\mathcal{S}_x M^*$, in this case called the *normal cone* in $\widehat{M}$ at $x \in M$ is defined by

$$N_x M = \mathcal{S}_x M^* = \left\{ X_x \in T_x\widehat{M} : \widehat{g}_x(X_x, Y_x) \leq 0, \forall\, Y_x \in \mathcal{S}_x M \right\}.$$

For $x \in \partial M_k = \overset{\circ}{M}$, $N_x M = T_x\overset{\circ}{M}^{\perp}$, the orthogonal complement of $T_x \partial M_k$ in $T_x\widehat{M}$.

Assumption 2.1 *We assume that f is the restriction of $\widehat{f}$ to M, where $\widehat{f}$ is a square-integrable C^2 process on $\widehat{M}$, a C^3 q-dimensional manifold, and M is an embedded piecewise C^2 k-dimensional submanifold of $\widehat{M}$. We assume that, for each i, the gradient of $f_{|\partial M_i}$ read off in some non-random orthonormal frame field $E_i = (X_1, \ldots, X_i)$ on ∂M_i satisfies the conditions of Lemma 2.5 of [19]. We denote this $\mathbb{R}^i$-valued process by $\nabla f_{|M_i,E_i}$. These conditions are satisfied in particular, if $\widehat{f}_{|\partial M_i}$ is suitably regular in the sense of [19]. We also assume that*

$$\rho(x,y) \triangleq \mathrm{Cor}\left(\widehat{f}(x), \widehat{f}(y)\right) = 1 \iff x = y.$$

2.1 A point set representation of the global maximizers of f

In this section, we derive the point set representation of the global maximizers of $f = \widehat{f}_{|M}$, the restriction of a C^2 process $\widehat{f}$ to $M \subset \widehat{M}$.

We begin with the deterministic case, stating and proving Lemma 2.2. This is the point process representation of the global maximizers of a deterministic function h, which we will extend to the stochastic setting below in Corollary 2.4.

Lemma 2.2 *Suppose $h = \widehat{h}_{|M}$, the restriction of $\widehat{h} \in C^2(\widehat{M})$. Fix $x \in \partial M_i$, and choose $\alpha^x \in C^2(M, (-\infty, 1])$ such that*

$$\alpha^x(y) = 1 \implies h(x) = h(y).$$

Then, x is a maximizer of h above the level u if, and only if

(i) $h(x) \geq u$;

(ii) $\nabla \widehat{h}(x) \in N_x M$, *that is, x is an extended outward critical point of h;*

(iii)

$$h(x) \geq \sup_{y \in M \setminus \{x\}} h^x(y)$$

where

$$h^x(y) \triangleq \begin{cases} \frac{h(y) - \alpha^x(y) h(x)}{1 - \alpha^x(y)} & \alpha^x(y) \neq 1 \\ h(y) & \alpha^x(y) = 1. \end{cases}$$

Further, if $\nabla^2 \alpha^x(x)$ is non-degenerate, and x is a critical point of $h_{|\partial M_i}$, then, for any C^1 curve $c : (-\delta, \delta) \to \partial M_i$ with $c(0) = x, \dot{c}(0) = X_x$

$$\lim_{t \to 0} \frac{h(c(t)) - \alpha^x(c(t)) h(x)}{1 - \alpha^x(c(t))} = \frac{\nabla^2 h_{|\partial M_i}(x)(X_x, X_x) - \nabla^2 \alpha^x_{|\partial M_i}(x)(X_x, X_x) h(x)}{-\nabla^2 \alpha^x_{|\partial M_i}(x)(X_x, X_x)}. \tag{7}$$

Proof:

First, we note that the condition $h(x) \geq u$ is obvious and nothing more need be said about it.

Suppose, then, that $x \in \partial M_i, 0 \leq i \leq k$ is a maximizer of h. Then, $\nabla \widehat{h}(x) \in N_x M$, otherwise there exists a direction $X_x \in \mathcal{S}_x M$ such that $\widehat{g}_x(X_x, \nabla \widehat{h}(x)) > 0$ and x cannot be a maximizer.

Because x is a maximizer, for all y such that $\alpha^x(y) < 1$ it follows that

$$\frac{h(y) - \alpha^x(y) h(x)}{1 - \alpha^x(y)} < h(x).$$

On the other hand, if $\alpha^x(y) = 1$, then, by choice of α^x, $h(y) = h(x)$ which proves that

$$h(x) \geq \sup_{y \in M \setminus \{x\}} h^x(y).$$

To prove the reverse implication, assume that x is an extended outward critical point of $h_{|\partial M_i}$ and

$$h(x) \geq \sup_{y \in M \setminus \{x\}} h^x(y).$$

Now suppose that x is not a maximizer of h, then there exists $y \in M\setminus\{x\}$ such that

$$h(x) < h(y).$$

In particular, for such a y, our choice of α^x implies that $\alpha^x(y) < 1$. It follows that

$$h(x) < \frac{h(y) - \alpha^x(y)h(x)}{1 - \alpha^x(y)}$$

which is a contradiction.

The second assertion follows from two applications of l'Hopital's rule after noting that x must be a critical point of $h_{|\partial M_i}$ as well as $\alpha^x_{|\partial M_i}$, and, at a critical point, the (Riemannian) Hessian of a function is independent of the Riemannian metric. $\square$

Corollary 2.3 *Lemma 2.2 is still true if we replace* (i) *above with*

(i′) $h(x) - \widehat{g}\left(F^x(x), \nabla h_{|\partial M_i}(x)\right) \geq u$

and redefine

$$h^x(y) \triangleq \begin{cases} \frac{h(y) - \alpha^x(y)h(x) - \widehat{g}(F^x(y) - \alpha^x(y)F^x(x), \nabla h_{|\partial M_i}(x))}{1-\alpha^x(y)} & \alpha^x(y) \neq 1 \\ h(y) & \alpha^x(y) = 1. \end{cases}$$

where F^x *is defined as*

$$F^x(y) = \sum_{j=1}^{q} \operatorname{Cov}\left(f(y), \widehat{X}_j f(x)\right) \widehat{X}_{j,x} \tag{8}$$

for some orthonormal frame field $(\widehat{X}_1, \ldots, \widehat{X}_q)$ *on* $\widehat{M}$.

Further, with h^x *redefined, we can remove the condition that* x *is a critical point in the second conclusion of Lemma 2.2, i.e. if* $\nabla^2\alpha^x(x)$ *is non-degenerate, then for any* C^1 *curve* $c : (-\delta, \delta) \to \partial M_i$ *with* $c(0) = x, \dot{c}(0) = X_x$

$$\begin{aligned} \lim_{t\to 0} & \frac{h(c(t)) - \alpha^x(c(t))h(x)}{1 - \alpha^x(c(t))} \\ &= \frac{\nabla^2 h_{|\partial M_i}(x)(X_x, X_x) - \nabla^2\alpha^x_{|\partial M_i}(x)(X_x, X_x)\left(h(x) - \widehat{g}(F^x(x), \nabla h_{|\partial M_i}(x))\right)}{-\nabla^2\alpha^x_{|\partial M_i}(x)(X_x, X_x)}. \end{aligned} \tag{9}$$

Note: Although Lemma 2.2 is completely deterministic, we bring this up here, as condition (i') corresponds to a statement about the "regression" of h onto ∇h, which will be useful when dealing with smooth processes. This redefinition of h^x, in the stochastic setting, can be thought of as a conditioned version of the original h^x, conditioned on x being a critical point.

Proof:

If x is a maximizer, it must be a critical point of $h_{|\partial M_i}$, and the terms involving $\nabla h_{|\partial M_i}(x)$ in the redefinition of h^x simply cancel out. The reverse implication is also true: if x is an extended outward critical point of h, then $\nabla h_{|\partial M_i}(x) = 0$, so we can ignore the terms involving $\widehat{X}_j h(x)$ above. $\square$

In the stochastic setting, that is when $\widehat{h}$ is replaced by a smooth process $\widehat{f}$, a natural candidate for $\alpha^x(y)$ is the partial map

$$\alpha^x(y) = \rho^x(y) \triangleq \rho(x, y) = \mathrm{Cor}(f(x), f(y)).$$

In this case, if the process f has constant variance,

$$\rho^x(y) = 1 \implies f(x) - \mathrm{E}(f(x)) = f(y) - \mathrm{E}(f(y)),$$

almost surely. Even if f does not have constant variance, the above still gives a point process representation of the maximizers of f.

Corollary 2.4 *Under Assumption 2.1 the maximizers of f are almost surely isolated and the maximizers of f are the points $x \in \partial M_i, 0 \le i \le k$ such that*

- $\nabla f_{|\partial M_i}(x) = 0$;
- $f(x) - \widehat{g}\left(F^x(x), \nabla f_{|\partial M_i}(x)\right) \ge u$;
- $P^{\perp}_{T_x \partial M_i} \nabla \widehat{f}(x) \in N_x M$;
- $f(x) \ge \sup_{y \in M \setminus \{x\}} \tilde{f}^x(y)$,

where $P^{\perp}_{T_x \partial M_i}$ represents projection onto the orthogonal complement of $T_x \partial M_i$ in $T_x \widehat{M}$ and

$$\tilde{f}^x(y) \triangleq \begin{cases} \frac{f(y) - \rho^x(y) f(x) - \widehat{g}\left(F^x(y) - \rho^x(y) F^x(x), \nabla f_{|\partial M_i}(x)\right)}{1 - \rho^x(y)} & \rho^x(y) \ne 1, \\ f(y) & \rho^x(y) = 1. \end{cases} \tag{10}$$

Note: If the joint density of $\nabla \widehat{f}(x)$, read off in some orthonormal basis of $T_x \widehat{M}$, is bounded by some constant K uniformly in $x \in \widehat{M}$ then, almost surely, there will be no critical points of $f_{|\partial M_i}$ such that $P^{\perp}_{T_x \partial M_i} \nabla \widehat{f}(x) \in \partial N_x M \subset T_x \partial M_i^{\perp}$. Therefore, almost surely, all global maximizers will be such that $P^{\perp}_{T_x \partial M_i} \nabla \widehat{f}(x)$ is in the relative interior of $N_x M$ in $T_x \partial M_i^{\perp}$. The proof of this claim is similar to the proof of Lemma 2.9 and is omitted.

Proof: The only part of the argument in Lemma 2.2 that needs to be modified is what happens when $\alpha^x(y) = \rho^x(y) = 1$. In the deterministic case, we assumed that $\alpha^x(y) = 1$ implied $h(x) = h(y)$. In the random case, we know that $\alpha^x(y) = 1$ implies $f(x) - \mathrm{E}(f(x)) = (f(y) - \mathrm{E}(f(y)))\sigma(x)/\sigma(y)$ almost surely, where

$$\sigma^2(x) = \mathrm{Var}(f(x)).$$

Almost surely, then, it is still true that if x is a maximizer of f then $f(x) \geq \tilde{f}^x(y)$ for all y such that $\rho^x(y) = 1$, otherwise, x cannot be a maximizer. The reverse implication follows similarly.

$\square$

2.2 Properties of $\tilde{f}^x$

From its definition, it is clear that the process $\tilde{f}^x$ is singular near x. In this section, we study some of its properties, specifically its singularities. Our main result is the following.

Lemma 2.5 *Fix a curve $c : (-\delta, 0] \to M$ so that $c(0) = x, \dot{c}(0) = -X_x$ with $X_x \in \mathcal{S}_x M \backslash T_x \partial M_i$. Then,*

$$\lim_{t \uparrow 0} \tilde{f}^x(c(t)) = \begin{cases} -\infty & \widehat{g}\left(X_x, P^{\perp}_{T_x \partial M_i} \nabla \widehat{f}(x)\right) < 0 \\ +\infty & \widehat{g}\left(X_x, P^{\perp}_{T_x \partial M_i} \nabla \widehat{f}(x)\right) > 0 \end{cases}$$

Therefore, when $P^{\perp}_{T_x \partial M_i} \nabla \widehat{f}(x)$ is in the relative interior of $N_x M$ in $T_x \partial M_i^{\perp}$,

$$\sup_{y \in M \backslash \{x\}} \tilde{f}^x(y) < +\infty$$

Proof:
The first conclusion follows once we note that

$$\tilde{f}^x(y) = \frac{f(y) - f(x) - \widehat{g}\left(F^x(y) - F^x(x), \nabla f_{|\partial M_i}(x)\right)}{1 - \rho^x(y)} + f(x) - \widehat{g}\left(F^x(x), \nabla f_{|\partial M_i}(x)\right)$$

and for y close to x

$$\mathrm{sign}\left(f(y) - f(x) - \widehat{g}(F^x(y) - F^x(x), \nabla f_{|\partial M_i}(x))\right) = \mathrm{sign}\left(\widehat{g}\left(X_x, P^{\perp}_{T_x \partial M_i} \nabla \widehat{f}(x)\right)\right).$$

The second conclusion follows once we note that when $P^{\perp}_{T_x \partial M_i} \nabla \widehat{f}(x)$ is in the relative interior of $N_x M$

$$\widehat{g}\left(X_x, P^{\perp}_{T_x \partial M_i} \nabla \widehat{f}(x)\right) < 0$$

for all $X_x \in \mathcal{S}_x M \backslash T_x \partial M_i$. For $X_x \in T_x \partial M_i$, the second conclusion of Lemma 2.3 implies that for all curves $c : (-\delta, \delta) \to \partial M_i$ with $c(0) = x, \dot{c}(0) = X_x$

$$\lim_{t \uparrow 0} \tilde{f}^x(c(t)) < +\infty.$$

$\square$

Lemma 2.5 implies that, when $P^{\perp}_{T_x\partial M_i}\nabla\widehat{f}(x)$ is in the relative interior of N_xM the supremum of $\tilde{f}^x$ is either achieved along a curve $c : (-\delta,\delta) \to \partial M_i$, in which case it is related to the largest eigenvalue of $\nabla^2 f_{|\partial M_i}$, or it is achieved at a point in $M\backslash\partial M_i$ away from x.

We can therefore find a ball of sufficiently small radius η around x such that for all $y \in (M\backslash\partial M_i)\cap B(x,\eta)$

$$\tilde{f}^x(y) < -C$$

for some $C > 0$. Continuity (in x) of the functions $\tilde{f}^x(y)$ imply that for all z in some neighbourhood N_x of x

$$\sup_{z\in N_x}\sup_{y\in B(x,\eta)} \tilde{f}^z(y) < -C.$$

This, along with continuity of $\nabla^2 f_{|\partial M_i}(x)$ is enough to prove the following corollary.

Corollary 2.6 *Define the process*

$$\tilde{W}(x) \stackrel{\Delta}{=} \sup_{y\in M\backslash\{x\}} \tilde{f}^x(y).$$

Then, for $1 \le i \le k$ the process $\tilde{W}(x)$ is continuous on the set

$$\left\{x \in \partial M_i : P^{\perp}_{T_x\partial M_i}\nabla\widehat{f}(x) \in \overset{\circ}{N_x M}\right\}$$

and is identically infinite on the set

$$\left\{x \in \partial M_i : P^{\perp}_{T_x\partial M_i}\nabla\widehat{f}(x) \in (N_xM)^c\right\}.$$

2.3 Point process representation of $\mathrm{Diff}_{f,M}(u)$

Our assumptions allow us to use the Morse theorem of [17] to express the expected Euler characteristic of the excursions $M \cap \widehat{f}^{-1}[u,+\infty)$ as integrals over M. The formula is not new, though we repeat it here for use in deriving bounds on the error in the Euler characteristic approximation.

What is new is the exact expression in Proposition 2.8 for the supremum distribution (1).

Proposition 2.7 *Under Assumption 2.1,*

$$\begin{aligned}
&\widehat{\mathrm{P}}\left(\sup_{x\in M} f(x) \ge u\right) \\
&\quad = \mathrm{E}\left(\chi\left(M\cap\widehat{f}^{-1}[u,+\infty)\right)\right) \\
&\quad = \sum_{i=0}^{k}\int_{\partial M_i} \mathrm{E}\left(\det\left(-\nabla^2 f_{|\partial M_i,E_i}(x)\right)\mathbb{1}_{A_x^{EC}}\middle|\nabla f_{|\partial M_i,E_i}(x)=0\right)\times \\
&\qquad\qquad \varphi_{\nabla f_{|\partial M_i,E_i}(x)}(0)\, d\mathcal{H}_i(x)
\end{aligned} \tag{11}$$

where $\mathcal{H}_i$ is i-dimensional Hausdorff measure, $\varphi_{\nabla f_{|\partial M_i,E_i}(x)}$ is the density of $\nabla f_{|\partial M_i,E_i}(x)$ and

$$A_x^{EC} = \{f(x) \geq u, \nabla \widehat{f}(x) \in N_x M\}.$$

Suitable regularity of the process $\nabla f_{|\partial M_i,E_i}(x)$ implies that the maximizers of f are almost surely isolated, though it does not guarantee uniqueness. If $\tilde{W}(x)$ were continuous when restricted to ∂M_i, Assumption 2.1 would allow us to apply the general point process Lemmas 2.4 and 2.5 of [19] to the point process representation of the maximizers in Corollary 2.4. Corollary 2.6 shows that $\tilde{W}(x)$ is not continuous, but it is continuous on the open set

$$\left\{ x \in \partial M_i : P^{\perp}_{T_x \partial M_i} \nabla \widehat{f}(x) \in \overset{\circ}{N_x M} \right\}.$$

Further, we are only interested in its behaviour on this set. Straightforward modifications of the above cited Lemmas, which we omit, lead to the following representation for the supremum distribution.

Proposition 2.8 *Under Assumption 2.1 suppose f has a unique maximum, almost surely. Further suppose that, for every $x \in M$*

$$\mathrm{P}\left(\tilde{W}(x) = f(x)\right) = \mathrm{P}\left(\tilde{W}(x) = u\right) = 0.$$

Then,

$$\begin{aligned} &\mathrm{P}\left(\sup_{x \in M} f(x) \geq u\right) \\ &\quad = \sum_{i=0}^{k} \int_{\partial M_i} \mathrm{E}\left(\det\left(-\nabla^2 f_{|\partial M_i,E_i}(x)\right) \mathbb{1}_{A_x^{SUP}} \middle| \nabla f_{|\partial M_i,E_i}(x) = 0\right) \times \\ &\qquad\qquad \varphi_{\nabla f_{|\partial M_i,E_i}(x)}(0)\, d\mathcal{H}_i(x) \end{aligned} \tag{12}$$

where

$$A_x^{SUP} = \{f(x) \geq u \vee \tilde{W}(x), \nabla \widehat{f}(x) \in N_x M\}.$$

The discepancy between the expected Euler characteristic approximation and the true supremum distribution is

$$\begin{aligned} \mathrm{Diff}_{f,M}(u) &\triangleq \widehat{\mathrm{P}}\left(\sup_{x \in M} f(x) \geq u\right) - \mathrm{P}\left(\sup_{x \in M} f(x) \geq u\right) \\ &= \sum_{i=0}^{k} \int_{\partial M_i} \mathrm{E}\left(\det\left(-\nabla^2 f_{|\partial M_i,E_i}(x)\right) \mathbb{1}_{A_x^{ERR}} \middle| \nabla f_{|\partial M_i,E_i}(x) = 0\right) \times \\ &\qquad\qquad \varphi_{\nabla f_{|\partial M_i,E_i}(x)}(0)\, d\mathcal{H}_i(x) \end{aligned} \tag{13}$$

where

$$A_x^{ERR} = \{u \leq f(x) \leq \tilde{W}(x), \nabla \widehat{f}(x) \in N_x M\}.$$

Before concluding this section, we provide a lemma giving sufficient conditions for the uniqueness of the global maximum of f. The proof is omitted, as it is similar to the proof of Theorem 3.2.1 of [2].

Lemma 2.9 *Suppose that for all pairs $1 \le i, j \le k$ and all tuples $\{(x,y) : x \in \partial M_i, y \in \partial M_j\}$ the random vector*

$$V(x,y) = \big(f(x) - f(y), \nabla f_{|\partial M_i, E_i}(x), \nabla f_{|\partial M_j, E_j}(y)\big)$$

has a density, bounded by some constant K independent of i, j. Then,

$$\mathrm{P}\left(\left\{\exists (x,y) : x \in \partial M_i, y \in \partial M_j, V(x,y) = 0 \in \mathbb{R}^{i+j+1}\right\}\right) = 0.$$

3 Bounding $\mathrm{Diff}_{f,M}(u)$

Expression (13) is an explicit formula for the error in the expected Euler characteristic approximation. Similar explicit formula can be derived for the error of the approximation based on the expected number of local maxima above the level u, though we do not pursue this here. However, as described in Section 2.2, the process $\tilde{f}^x$ is singular near x, and has infinite variance near x, which means that standard tools such as the Borell-Tsirelson inequality cannot be used to bound its supremum distribution.

To see that the process $\tilde{f}^x$ has infinite variance, assume that f is the restriction of a unit variance field with covariance function ρ. In this case, the variance of $\tilde{f}^x(y)$ is easily seen to be

$$\mathrm{Var}\left(\tilde{f}^x(y)\right) = \frac{1 + \rho(x,y)}{1 - \rho(x,y)}$$

and

$$\lim_{y \to x} \mathrm{Var}\left(\tilde{f}^x(y)\right) = \lim_{y \to x} \frac{1 + \rho(x,y)}{1 - \rho(x,y)} = +\infty.$$

Although this is somewhat worrying, in (13) we only care about large *positive* values of $\tilde{f}^x$, and, further, we only care about the behaviour of $\tilde{f}^x$ on the set

$$\{P^{\perp}_{T_x \partial M_i} \nabla \widehat{f}(x) \in N_x M\}.$$

To get around this difficulty, we introduce a process f^x in this section which has, under some conditions, finite variance and dominates $\tilde{f}^x$ on $\{P^{\perp}_{T_x \partial M_i} \nabla \widehat{f}(x) \in N_x M\}$. It is this process whose variance appears in the exponential bound for the behaviour of $\mathrm{Diff}_{f,M}(u)$ in the Gaussian case.

Obviously, the process f^x, which we define below, does not dominate the *absolute* value of $\tilde{f}^x$. Indeed, if this were true, the process f^x would have infinite variance as well.

The main goal of this section is to introduce the process f^x, after which, we arrive to Theorem 3.3 with some simple manipulations applied to expression (13).

The process f^x is defined as follows

$$f^x(y) \triangleq \begin{cases} \frac{f(y)-\rho^x(y)f(x)-\widehat{g}\left(\widehat{F}^x(y)-P_{N_xM}\widehat{F}^x(y),\nabla\widehat{f}(x)\right)}{1-\rho^x(y)} & \rho^x(y) \neq 1 \\ f(y) - \widehat{g}\left(\widehat{F}^x(y) - P_{N_xM}\widehat{F}^x(y), \nabla\widehat{f}(x)\right) & \rho^x(y) = 1 \end{cases} \tag{14}$$

where $P_{N_xM} : T_x\widehat{M} \to N_xM$ represents orthogonal projection onto N_xM and $\widehat{F}^x : M \to T_x\widehat{M}$ is given by

$$\widehat{F}^x(y) = \begin{cases} F^x(y) - \rho^x(y)F^x(x) & \rho^x(y) \neq 1 \\ F^x(y) & \rho^x(y) = 1 \end{cases} \tag{15}$$

where F^x is defined in (8).

Lemma 3.1 *On the set* $\{P^{\perp}_{T_x\partial M_i}\nabla\widehat{f}(x) \in N_xM\}$, *for every* $y \in M$

$$f^x(y) \geq \tilde{f}^x(y).$$

If $x \in \partial M_k = \overset{\circ}{M}$, *then equality holds above on the set* $\{P^{\perp}_{T_x\partial M_i}\nabla\widehat{f}(x) \in N_xM\}$.

Proof: First, we note that

$$(1-\rho^x(y)) \cdot (\tilde{f}^x(y) - f^x(y)) = \widehat{g}\left(\widehat{F}^x(y) - P_{N_xM}\widehat{F}^x(y), P^{\perp}_{T_x\partial M_i}\nabla\widehat{f}(x)\right).$$

Now, as N_xM is a convex cone, it follows that for any $Y_x \in T_x\widehat{M}$

$$Y_x - P_{N_xM}Y_x \in N_xM^* = \mathrm{Con}(\mathcal{S}_xM)$$

where N_xM^* is the dual cone of N_xM, which is just $\mathrm{Con}(\mathcal{S}_xM)$, the convex hull of $\mathcal{S}_xM$. By duality,

$$\widehat{g}\left(Y_x - P_{N_xM}Y_x, V_x\right) \leq 0$$

for every $V_x \in N_xM$. Consequently, on the set $\{P^{\perp}_{T_x\partial M_i}\nabla\widehat{f}(x) \in N_xM\}$

$$\widehat{g}\left(Y_x - P_{N_xM}Y_x, P^{\perp}_{T_x\partial M_i}\nabla\widehat{f}(x)\right) \leq 0$$

for every $Y_x \in T_x\widehat{M}$. As $\widehat{F}^x(y) \in T_x\widehat{M}$ for each y, the first conclusion now follows.

As for the second, if $x \in \partial M_k$, then $N_xM = T_x\partial M_k^{\perp}$ and $P_{N_xM}V_x = 0$ for all $V_x \in T_x\partial M_k$. Similarly,

$$\widehat{g}\left(V_x, P^{\perp}_{T_x\partial M_i}\nabla\widehat{f}(x)\right) = 0.$$

Therefore, on this set

$$\widehat{g}\left(\widehat{F}^x(y) - P_{N_xM}\widehat{F}^x(y), \nabla\widehat{f}(x)\right) = 0.$$

□

As we will see in the proof of Theorem 3.3, Lemma 3.1 provides the basic bounds for $\mathrm{Diff}_{f,M}(u)$. The following corollary to Lemma 2.2 will also be of use to us.

Corollary 3.2 *If f has unit variance, then, for any C^1 unit speed curve $c : (-\delta, \delta) \to \partial M_k$ with $c(0) = x, \dot{c}(0) = X_x$*

$$\begin{aligned}\lim_{t\to 0} f^x(c(t)) &= -\frac{\nabla^2 f_{|\partial M_k}(x)(X_x, X_x) - \nabla^2 \rho^x(x)(X_x, X_x) f(x)}{-\nabla^2 \rho^x(x)(X_x, X_x)} \\ &= -\nabla^2 f_{|\partial M_k}(x)(X_x, X_x) + f(x).\end{aligned}$$

Further,

$$\sup_{X_x \in S(T_x \partial M_k)} \left| -\nabla^2 f_{|\partial M_k}(x)(X_x, X_x) + f(x) \right| \leq \sup_{y \in M \setminus \{x\}} |f^x(y)|. \tag{16}$$

Proof: The proof is essentially just the second conclusion of Lemma 2.2, recast in the stochastic process framework. The only thing that needs to be verified is that

$$\nabla^2 \rho^x(x)(X_x, X_x) = -1,$$

but this follows from the fact that

$$\nabla^2 \rho^x(y)(X_y, Y_y) = \mathrm{Cov}\left(\nabla^2 f_{|\partial M_k}(y)(X_y, Y_y), f_{\partial M_k}(x)\right)$$

and the fact that, as a double form

$$\mathrm{Cov}\left(\nabla^2 f_{|\partial M_k}(x)(X_x, Y_x), f(x)\right) = -\widehat{g}_x(X_x, Y_x),$$

(cf. [19]). $\square$

Using the results of Lemma 3.1 we have the following theorem.

Theorem 3.3 *Under Assumption 2.1, suppose the process*

$$W(x) \triangleq \sup_{y \in M \setminus \{x\}} f^x(y) \tag{17}$$

is continuous and f has a unique maximum, almost surely. Further suppose that, for every $x \in M$

$$\mathrm{P}\left(W(x) = f(x)\right) = \mathrm{P}\left(W(x) = u\right) = 0.$$

Then,

$$\begin{aligned}|\mathrm{Diff}_{f,M}(u)| \leq \sum_{i=0}^{k} \int_{\partial M_i} \mathrm{E}\left(\left| \det\left(-\nabla^2 f_{|\partial M_i, E_i}(x)\right) \right| \mathbb{1}_{B_x^{ERR}} \middle| \nabla f_{|\partial M_i, E_i}(x) = 0 \right) \times \\ \varphi_{\nabla f_{|\partial M_i, E_i}(x)}(0)\ d\mathcal{H}_i(x)\end{aligned} \tag{18}$$

where

$$B_x^{ERR} = \{u \leq f(x) \leq W(x)\}.$$

4 Gaussian fields with constant variance

In this section, using Theorem 3.3, we derive an explicit bound for the exponential behaviour of $\mathrm{Diff}_{f,M}(u)$ when $\widehat{f}$ is a Gaussian field with constant variance, satisfying Assumption 2.1. The assumption of constant variance implies certain random variables are uncorrelated, hence independent in the Gaussian case. In particular, the assumption of constant variance implies that for $x \in \partial M_i, 0 \le i \le k$ the entire process $(f^x(y))_{y \in M\setminus\{x\}}$ is independent of $f(x)$ as well as $\nabla f_{|\partial M_i}(x)$. This allows us to remove the conditioning on $\nabla f_{|\partial M_i}(x)$ below. Once this conditioning is removed, the rest of the argument relies only on the Borell-Tsirelson inequality [3].

Our first observation is that, whether f has constant variance or not, for each $x \in M$, the process $f^x(y)$ is uncorrelated with the random vector $\nabla f_{|\partial M_i}(x)$. Hence, in the Gaussian case, $f^x(y)$ is independent of $\nabla f_{|\partial M_i}(x)$.

Lemma 4.1 *For every $x \in \partial M_i, 0 \le i \le k$ and every $y \in M\setminus\{x\}$*

$$\mathrm{Cov}\left(f^x(y), X_x f\right) = 0$$

for every $X_x \in T_x\partial M_i$.

Proof: We first note that, if $\rho^x(y) \neq 1$ then

$$\begin{aligned}
(1-\rho^x(y))\,\mathrm{Cov}\left(f^x(y), X_x f\right) &= \mathrm{Cov}\left(f(y) - \rho^x(y) f(x), X_x f\right) \; - \\
&\qquad \mathrm{Cov}\left(\widehat{g}\left(\widehat{F}^x(y) - P_{N_xM}\widehat{F}^x(y), \nabla\widehat{f}(x)\right), X_x f\right) \\
&= \mathrm{Cov}\left(f(y) - \rho^x(y) f(x), X_x f\right) - \widehat{g}\left(\widehat{F}^x(y) - P_{N_xM}\widehat{F}^x(y), X_x\right).
\end{aligned}$$

If, on the other hand, $\rho^x(y) = 1$, then

$$\begin{aligned}
\mathrm{Cov}\left(f^x(y), X_x f\right) &= \mathrm{Cov}\left(f(y), X_x f\right) \; - \\
&\qquad \mathrm{Cov}\left(\widehat{g}\left(F^x(y) - P_{N_xM}F^x(y), \nabla\widehat{f}(x)\right), X_x f\right) \\
&= \mathrm{Cov}\left(f(y), X_x f\right) - \widehat{g}\left(F^x(y) - P_{N_xM}F^x(y), X_x\right).
\end{aligned}$$

The conclusion will therefore follow once we prove, for every $y \in M$

$$\begin{aligned}
\mathrm{Cov}\left(f(y), X_x f\right) &= \widehat{g}\left(F^x(y) - P_{N_xM}F^x(y), X_x\right) \\
\mathrm{Cov}\left(f(y) - \rho^x(y) f(x), X_x f\right) &= \widehat{g}\left(\widehat{F}^x(y) - P_{N_xM}\widehat{F}^x(y), X_x\right).
\end{aligned}$$

As the two arguments are similar, we just prove the first equality. The map F^x can be decomposed as follows

$$F^x(y) = P_{T_x\partial M_i}F^x(y) + P^{\perp}_{T_x\partial M_i}F^x(y)$$

where

$$P_{T_x\partial M_i} F^x(y) = \sum_{j=1}^{i} \mathrm{Cov}\left(f(y), X_j f(x)\right) X_{j,x}$$

$$P^{\perp}_{T_x\partial M_i} F^x(y) = \sum_{j=i+1}^{q} \mathrm{Cov}\left(f(y), X_j f(x)\right) X_{j,x}$$

and the orthonormal basis $(X_{1,x}, \dots, X_{q,x})$ is chosen so that the set $(X_{1,x}, \dots, X_{i,x})$ forms an orthonormal basis for $T_x\partial M_i$ and $(X_{i+1,x}, \dots, X_{q,x})$ forms an orthonormal basis for $T_x\partial M_i^{\perp}$, the orthogonal complement of $T_x\partial M_i$ in $T_x\widehat{M}$.

Further, because $\widehat{g}\,(X_x, V_x) = 0$ for every $X_x \in T_x\partial M_i$ and $V_x \in N_x M$, it follows that

$$P_{N_x M} F^x(y) = P_{N_x M} P^{\perp}_{T_x\partial M_i} F^x(y)$$

and for every $X_x \in T_x\partial M_i$

$$\begin{aligned}
\widehat{g}\left(F^x(y) - P_{N_x M} F^x(y), X_x\right) &= \widehat{g}\left(P_{T_x\partial M_i} F^x(y), X_x\right) \; + \\
&\qquad \widehat{g}\left(P^{\perp}_{T_x\partial M_i} F^x(y) - P_{N_x M} P^{\perp}_{T_x\partial M_i} F^x(y), X_x\right) \\
&= \widehat{g}\left(P_{T_x\partial M_i} F^x(y), X_x\right) \\
&= \sum_{j=1}^{i} \mathrm{Cov}\left(f(y), X_j f(x)\right) \widehat{g}\left(X_{j,x}, X_x\right) \\
&= \mathrm{Cov}\left(f(y), X_x f\right).
\end{aligned}$$

□

As noted above, the independence between $\nabla f_{|\partial M_i}(x)$ and the process f^x allows us to remove the conditioning on $\nabla f_{|\partial M_i}(x)$ in the expression for $\mathrm{Diff}_{f,M}(u)$, whether f has constant variance or not.

Corollary 4.2 *Suppose f is a Gaussian process satisfying the conditions of Theorem 3.3. Then,*

$$\begin{aligned}
\left|\mathrm{Diff}_{f,M}(u)\right| \;\; &\leq \sum_{i=0}^{k} \int_{\partial M_i} \mathrm{E}\left(\left|\det\left(-\nabla^2 f_{|\partial M_i, E_i}(x)\right)\right| \mathbb{1}_{C_x^{ERR}}\right) \;\times \\
&\qquad\qquad\qquad \varphi_{\nabla f_{|\partial M_i, E_i}(x)}(0)\; d\mathcal{H}_i(x)
\end{aligned} \tag{19}$$

where

$$C_x^{ERR} = \{u \leq f(x) - \widehat{g}\left(P_{T_x\partial M_i} F^x(x), \nabla \widehat{f}(x)\right) \leq W(x)\}.$$

If f has constant variance, then $F^x(x) = 0 \in T_x M$ and

$$C_x^{ERR} = \{u \leq f(x) \leq W(x)\}.$$

Proof: The only thing that needs to be proven is that, in the event C_x^{ERR} the condition $\{u \leq f(x) \leq W(x)\}$ can be replaced with

$$\left\{u \leq f(x) - \widehat{g}\left(P_{T_x\partial M_i} F^x(x), \nabla \widehat{f}(x)\right) \leq W(x)\right\}$$

and the conditioning can be removed.

The reason that the above replacement is justified is that, on the set $\{\nabla f_{|\partial M_i}(x) = 0\}$

$$f(x) = f(x) - \widehat{g}\left(P_{T_x\partial M_i} F^x(x), \nabla \widehat{f}(x)\right).$$

Further, $f(x) - \widehat{g}\left(P_{T_x\partial M_i} F^x(x), \nabla \widehat{f}(x)\right)$ is independent of $\nabla f_{|\partial M_i}(x)$, and Lemma 4.1 implies that $W(x)$ is also independent of $\nabla f_{|\partial M_i}(x)$. Therefore, the conditioning on $\nabla f_{\partial M_i}(x)$ can be removed. □

We are now ready to prove the following Theorem.

Theorem 4.3 *Let $\widehat{f}$ be a Gaussian process with constant, unit variance, on $\widehat{M}$ and let $f = \widehat{f}_{|M}$ be such that f satisfies Assumption 2.1. Then,*

$$\liminf_{u\to\infty} -u^{-2} \log |\mathrm{Diff}_{f,M}(u)| \geq \frac{1}{2}\left(1 + \frac{1}{\sigma_c^2(f)}\right)$$

where

$$\sigma_c^2(f, x) \triangleq \sup_{y \in M\setminus\{x\}} \mathrm{Var}\left(f^x(y)\right). \tag{20}$$

and

$$\sigma_c^2(f) \triangleq \sup_{x\in M} \sigma_c^2(f, x) \tag{21}$$

Proof:

We must find an upper bound for (19). Writing

$$\begin{aligned}
\nabla^2 f_{|\partial M_i, E_i}(x) &= \nabla^2 f_{|\partial M_i, E_i}(x) - \mathrm{E}\left(\nabla^2 f_{|\partial M_i, E_i}(x) \middle| f(x)\right) + \mathrm{E}\left(\nabla^2 f_{|\partial M_i, E_i}(x) \middle| f(x)\right) \\
&= \nabla^2 f_{|\partial M_i, E_i}(x) + f(x) I - f(x) I
\end{aligned}$$

(cf. [19]), and applying Hölder's inequality to (19) yields, for any conjugate exponents p, q

$$\begin{aligned}
\left|\mathrm{Diff}_{f,M}(u)\right| \leq \sum_{i=0}^{k} \int_{\partial M_i} \sum_{j=0}^{i} &\mathrm{E}\left(f(x)^j \mathbb{1}_{\{f(x)\geq u\}}\right) \times \\
&\mathrm{E}\left(|\mathrm{detr}_{i-j}\left(-\nabla^2 f_{|\partial M_i, E_i}(x) - f(x) I\right)|^p\right)^{1/p} \times \mathrm{P}\left(W(x) \geq u\right)^{1/q} \, d\mathcal{H}_i(x).
\end{aligned}$$

Define

$$\mu^+ \triangleq \sup_{x\in M} \mu(x) \mathbb{1}_{\{\mu(x)\geq 0\}}, \quad \mu(x) = \mathrm{E}(f(x))$$

For

$$u \geq \sup_{x\in M} \mathrm{E}\left(\sup_{y\in M\setminus\{x\}} \left(f^x(y) - \mathrm{E}(f^x(y))\right)\right) + \mu^+$$

the Borell-Tsirelson inequality implies that

$$\mathrm{P}\left(W(x) \geq u\right) \leq 2\mathrm{e}^{-(u-\mu^+)^2/2\sigma_c^2(f,x)}.$$

Recalling that $f^x(x) = f(x)$, for such u, it also follows that

$$\mathrm{E}\left(f(x)^j \mathbb{1}_{\{f(x)\geq u\}}\right) \leq C_j u^{j-1} \mathrm{e}^{-(u-\mu^+)^2/2}.$$

Putting these facts together, for any conjugate exponents p, q

$$\left|\mathrm{Diff}_{f,M}(u)\right| \leq C_k u^{k-1} \mathrm{e}^{-\frac{(u-\mu^+)^2}{2}\left(1+\frac{1}{q\sigma_c^2(f)}\right)} \times$$
$$\sum_{i=0}^{k} \int_{\partial M_i} \sum_{j=0}^{i} \mathrm{E}\left(|\mathrm{detr}_{i-j}\left(-\nabla^2 f_{|\partial M_i, E_i}(x) - f(x)I\right)|^p\right)^{1/p} \, d\mathcal{H}_i(x).$$

The result now follows after noting that we can choose q close to 1, and $u(q)$ so that, for $u \geq u(q)$, the remaining terms are arbitrarily small logarithmically, compared to u^2.$\square$

Theorem 4.3 provides a lower bound on the exponential decay of $\mathrm{Diff}_{f,M}(u)$. The lower bound is generally tight when a maximizer of $\sigma_c^2(f)$ occurs in ∂M_k, in the sense that the term corresponding to ∂M_k in the sum defining $\mathrm{Diff}_{f,M}(u)$ in (13) is exponentially of the same order as the upper bound. It is still open as to whether the $\liminf_{u\to\infty}$ in Theorem 4.3 can be replaced with $\lim_{u\to\infty}$ as we can not rule out the possibility of some terms in the sum (13) cancel each other out, leading to a faster rate of exponential decay. Although we have not settled the issue completely, these situations seem somewhat singular.

Theorem 4.4 *Let f be as in Theorem 4.3. If*

$$\{x \in M : \exists\, y \in M\backslash\{x\}, \mathrm{Var}(f^x(y)) = \sigma_c^2(f)\} \cap \partial M_k \neq \emptyset,$$

and the metric entropy $H(f^x, \varepsilon)$ of the process f^x is bounded by $C\varepsilon^{-b}$ for some $b > 0$, uniformly in $x \in \partial M_k$, then

$$\lim_{u\to\infty} -u^{-2} \log \left| \int_{\partial M_k} \mathrm{E}\left(\det\left(-\nabla^2 f_{|\partial M_k, E_k}(x)\right) \mathbb{1}_{\{u \leq f(x) \leq W(x)\}}\right) \varphi_{\nabla f_{|\partial M_k, E_k}(x)}(0) \, d\mathcal{H}_k(x)\right|$$
$$= \frac{1}{2}\left(1 + \frac{1}{\sigma_c^2(f)}\right). \tag{22}$$

Proof:

For ease of notation, we define the contribution of $A \subset M$ to the error as

$$\mathrm{Diff}_{f,A}(u) = \sum_{i=0}^{k} \int_{\partial M_i \cap A} \mathrm{E}\left(\det\left(-\nabla^2 f_{|\partial M_i, E_i}(x)\right) \mathbb{1}_{A_x^{ERR}} \middle| \nabla f_{|\partial M_i, E_i}(x) = 0\right) \times$$
$$\varphi_{\nabla f_{|\partial M_i, E_i}(x)}(0) \, d\mathcal{H}_i(x). \tag{23}$$

Verifying the lower bound in (22), i.e. the fact that

$$\liminf_{u\to\infty} -u^{-2}\log\left(|\mathrm{Diff}_{f,\partial M_k}(u)|\right) \geq \frac{1}{2}\left(1+\frac{1}{\sigma_c^2(f)}\right)$$

follows the arguments of Theorem 4.3.

Turning to the upper bound in (22), the expression $\mathrm{Diff}_{f,\partial M_k}(u)$ can be expressed as a sum of k terms

$$\mathrm{Diff}_{f,\partial M_k}(u) = \sum_{j=0}^{k}\int_{\partial M_k} \mathrm{E}\Bigg(f(x)^j \mathrm{detr}_{k-j}\left(-\nabla^2 f_{|\partial M_k,E_k}(x) + f(x)I\right) \times \\ \mathbb{1}_{\{u\leq f(x)\leq W(x)\}}\Bigg)\varphi_{\nabla f_{|\partial M_k,E_k}(x)}(0)\, d\mathcal{H}_k(x).$$

Using relation (16) of Corollary 3.2, for u large enough and $1\leq j\leq k$, it follows that

$$\begin{aligned}
&\mathrm{E}\left(f(x)^{k-j}\mathrm{detr}_j\left(-\nabla^2 f_{|\partial M_k,E_k}(x) + f(x)I\right)\mathbb{1}_{\{W(x)\geq f(x)\geq u\}}\middle| f(x)\right)\\
&\quad\leq f(x)^{k-j}\mathrm{E}\left(\sup_{y\in M\setminus\{x\}}|f^x(y)|^j\mathbb{1}_{\{W(x)\geq f(x)\geq u\}}\middle| f(x)\right)\\
&\quad\leq f(x)^{k-j}\mathrm{E}\left(\sup_{y\in M\setminus\{x\}}|f^x(y)|^j\mathbb{1}_{\{\sup_{y\in M\setminus\{x\}}|f^x(y)|\geq f(x)\}}\middle| f(x)\right)\mathbb{1}_{\{f(x)\geq u\}}\\
&\quad\leq C\times f(x)^{k-j}\left(\int_{f(x)}^{\infty} jr^{j-1}r^{b-1}\mathrm{e}^{-r^2/2\sigma_c^2(f,x)}\,dr\right)\mathbb{1}_{\{f(x)\geq u\}}\\
&\quad\leq C\times f(x)^{k+b-3}\mathrm{e}^{-f(x)^2/2\sigma_c^2(f,x)}\mathbb{1}_{\{f(x)\geq u\}}\\
&\quad\leq C\times f(x)^{k+b}\mathrm{e}^{-f(x)^2/2\sigma_c^2(f,x)}\mathbb{1}_{\{f(x)\geq u\}}\\
&\quad\leq C\times \mathrm{E}\left(f(x)^{k-\eta}\mathbb{1}_{\{W(x)\geq f(x)\geq u\}}\middle| f(x)\right)
\end{aligned}$$

for some $\eta>0$. Above, we have used the fact that f has constant variance only when we assert the independence of $W(x)$ and $f(x)$. In going from the third to the fourth line above, as well as from the sixth to the seventh line we have used Theorem 5.3 of [3], and our assumption on the metric entropy of the processes f^x.

Therefore, for u large enough,

$$C_1 \leq \frac{\int_{\partial M_k}\mathrm{E}\left(f(x)^k\mathbb{1}_{\{u\leq f(x)\leq W(x)\}}\right)\, d\mathcal{H}_k(x)}{\mathrm{Diff}_{f,\partial M_k}(u)} \leq C_2,$$

and

$$\lim_{u\to\infty} -u^{-2}\log\left(\mathrm{E}\left(f(x)^k\mathbb{1}_{\{u\leq f(x)\leq W(x)\}}\right)\, d\mathcal{H}_k(x)\right) = \lim_{u\to\infty} -u^{-2}\log\left(\mathrm{Diff}_{f,\partial M_k}(u)\right)$$

as long as the limit on the left hand side exists.

Theorem 5.3 of [3] implies that, for each $\eta > 0$

$$
\begin{aligned}
C \times u^{2k-2} \mathrm{e}^{-u^2\left(1+1/\sigma_c^2(f,x)\right)/2} &\geq \mathrm{E}\left(f(x)^k \mathbb{1}_{\{u \leq f(x) \leq W(x)\}}\right) \\
&\geq C(\eta) \times u^{2k-2-\eta} \mathrm{e}^{-u^2\left(1+1/\sigma_c^2(f,x)\right)/2}
\end{aligned}
$$

and a standard Laplace approximation implies that, for η small enough,

$$
\begin{aligned}
&\lim_{u\to\infty} -u^{-2} \log\left(\int_{\partial M_k} u^{2k-2} \mathrm{e}^{-u^2\left(1+1/\sigma_c^2(f,x)\right)/2} d\mathcal{H}_k(x)\right) \\
&\quad = \lim_{u\to\infty} -u^{-2} \log\left(\int_{\partial M_k} u^{2k-2-\eta} \mathrm{e}^{-u^2\left(1+1/\sigma_c^2(f,x)\right)/2} d\mathcal{H}_k(x)\right) \\
&\quad = \frac{1}{2}\left(1 + \frac{1}{\sigma_c^2(f)}\right).
\end{aligned}
$$

We have now established (22). $\square$

Remark: The polynomial form of the entropy assumed above can be weakened, as long as it is logarithmically small enough to be killed by the factor u^{-2}.

5 Examples

In this section, we compute $\sigma_c^2(f)$ for some simple examples, strengthening earlier results of [11] and [14]. Before turning to the examples, however, we discuss the relation between $\sigma_c^2(f)$ and the critical radius of a tube around M when f is assumed to be centered with unit variance. Specifically, we describe the geometry of the situation in the case of "global overlap." That is, when the supremum

$$
\sigma_c^2(f) = \sup_{x\in M} \sup_{y\in M\setminus\{x\}} \mathrm{Var}(f^x(y))
$$

is achieved at a point (x^*, y^*), in which case the point (x^*, y^*) must be strictly off the diagonal.

5.1 Geometric picture in the case of global overlap

Here, we describe the notion of "global overlap" and describe the geometry of the process f near points (x^*, y^*) achieving the critical variance $\sigma_c^2(f)$. Roughly speaking, this situation occurs when M, the parameter space of f "wraps around itself" and, for some $x \in M$ there is a point $y \in M$ that is close to x in the L^2-metric but far in terms of geodesic distance from x. To describe the geometry involved in this situation we turn to spherical geometry in $\tilde{H}_f$ the RKHS of f. Recall that $\tilde{H}_f$ is defined by the reproducing kernel condition

$$
\langle R(s,\cdot), R(t,\cdot)\rangle_{\tilde{H}_f} = R(t,s)
$$

and there exists an isometry that maps $\tilde{H}_f$ onto the linear span H_f of $\{f(x), x \in M\}$ in $L^2(\Omega, \mathcal{F}, \mathrm{P})$. Without loss of generality, then, we can describe the geometry in terms of H_f. Let $\Psi : M \to H_f \subset L^2(\Omega, \mathcal{F}, \mathrm{P})$ denote the map

$$x \mapsto f(x).$$

If $S(H_f)$ is the unit sphere in H_f, and f is centered with unit variance then $\Psi(M) \subset S(H_f)$, and our standing assumptions, namely that f is C^2 and $\rho(x, y) \neq 1$ if $x \neq y$, imply that Ψ is a piecewise C^2 embedding. Further, it is not hard to see that

$$\Psi_*(X_x) = X_x f$$

so the tangent space $T_{f(x)}\Psi(M)$ is spanned by $(X_{1,x}f, \ldots, X_{k,x}f)$ for some basis $\{X_{1,x}, \ldots, X_{k,x}\}$ of T_xM.

We denote the orthogonal complement of $T_{f(x)}\Psi(M)$ in $T_{f(x)}H_f$ by $T^{\perp}_{f(x)}\Psi(M)$, and the orthogonal complement of $T_{f(x)}\Psi(M)$ in $T_{f(x)}S(H_f)$ by $N_{f(x)}\Psi(M)$.

Given a point $f(x)$ and a unit normal vector $v_{f(x)} \in N_{f(x)}\Psi(M)$ we denote the geodesic, in $S(H_f)$, originating at $f(x)$ in the direction $v_{f(x)}$ by $c_{f(x),v_{f(x)}}$. That is,

$$c_{f(x),v_{f(x)}}(t) = \cos t \cdot f(x) + \sin t \cdot v_{f(x)}, \qquad 0 \leq t < \pi.$$

As discussed in [18], up to a certain point along $c_{f(x),v_{f(x)}}$, the points on $c_{f(x),v_{f(x)}}$ metrically project uniquely to x. That is, for t small enough, the unique point on $\Psi(M)$ closest to $c_{f(x),v_{f(x)}}(t)$ is $f(x)$. We denote the largest t for which this is true by $\theta(f(x), v_{f(x)})$ and we call it the *local critical radius (angle) at $f(x)$ in the direction $v_{f(x)}$*. Taking the supremum over all directions $v_{f(x)} \in S(N_{f(x)}\Psi(M))$ we obtain $\theta(f(x))$, the *local critical radius at $f(x)$*

$$\theta_c(f(x)) = \inf_{v_{f(x)} \in S(N_{f(x)}\Psi(M))} \theta(f(x), v_{f(x)}) \tag{24}$$

and the global critical radius

$$\theta_c = \theta_c(\Psi(M)) = \inf_{x \in M} \theta_c(f(x)). \tag{25}$$

When M has no boundary, the relation between these critical angles, and $\sigma_c^2(f, x)$ in (20) is given in the following lemma.

Lemma 5.1 *Suppose $f = \widehat{f}_{|M}$ is the restriction of a centered, unit variance Gaussian process on $\widehat{M}$, to M, a piecewise C^2 k-dimensional submanifold of $\widehat{M}$. Suppose that $\Psi : \widehat{M} \to H_f$ is a C^2 embedding. Then, for all $x \in \partial M_k$*

$$\sigma_c^2(f, x) = \cot^2(\theta(x)).$$

Proof: Without loss of generality we can assume that M has no boundary setting $\widehat{M} = M$ and $\widehat{f} = f$. In this case, $N_xM = \{0\}$, and because f has constant variance

$F^x(x) = 0$. Putting these facts together show that

$$\begin{aligned}
f^x(y) &= \frac{f(y) - \rho^x(y)f(x) - \widehat{g}(\widehat{F}^x(y), \nabla \widehat{f}(x))}{1 - \rho^x(y)} \\
&= \frac{f(y) - \rho^x(y)f(x) - \sum_{i=1}^k \mathrm{Cov}(f(y), X_i f(x)) X_i f(x)}{1 - \rho^x(y)}
\end{aligned}$$

for some orthonormal frame field $(X_1, \ldots, X_k)$ on M.

Turning to the picture in terms of geodesics, fix $f(x)$ and $v_{f(x)}$. Suppose that, for a certain t the point $c_{f(x),v_{f(x)}}(t)$ does not metrically project to $f(x)$. This implies there is a point $f(y) \in \Psi(M)$ such that

$$\begin{aligned}
d(c_{f(x),v_{f(x)}}(t), f(y)) &= \cos^{-1}\left(\langle c_{f(x),v_{f(x)}}(t), f(y)\rangle_{H_f}\right) \\
&< d(c_{f(x),v_{f(x)}}(t), f(x)) \\
&= \cos^{-1}\left(\langle c_{f(x),v_{f(x)}}(t), f(x)\rangle_{H_f}\right).
\end{aligned}$$

Alternatively,

$$\begin{aligned}
\langle c_{f(x),v_{f(x)}}(t), f(y)\rangle_{H_f} &= \cos t \cdot \langle f(x), f(y)\rangle_{H_f} + \sin t \cdot \langle v_{f(x)}, f(y)\rangle_{H_f} \\
&= \cos t \cdot \rho(x, y) + \sin t \cdot \langle v_{f(x)}, f(y)\rangle_{H_f} \\
&> \langle c_{f(x),v_{f(x)}}(t), f(x)\rangle_{H_f} \\
&= \cos t.
\end{aligned}$$

After a little rearranging, we see that this is true if and only if

$$\cot t > \frac{\langle v_{f(x)}, f(y)\rangle_{H_f}}{1 - \rho(x, y)}.$$

Therefore,

$$\cot \theta(f(x), v_{f(x)}) = \sup_{y \in M \setminus \{x\}} \frac{\langle v_{f(x)}, f(y)\rangle_{H_f}}{1 - \rho(x, y)}.$$

Taking the supremum over all $v_{f(x)} \in N_{f(x)}\Psi(M)$ we see that

$$\begin{aligned}
\cot \theta(f(x)) &= \sup_{v_{f(x)} \in N_{f(x)}\Psi(M)} \sup_{y \in M \setminus \{x\}} \frac{\langle v_{f(x)}, f(y)\rangle_{H_f}}{1 - \rho(x, y)} \\
&= \sup_{y \in M \setminus \{x\}} \frac{\|P_{N_{f(x)}\Psi(M)} f(y)\|_{H_f}}{1 - \rho(x, y)}
\end{aligned}$$

where $P_{N_{f(x)}\Psi(M)}$ represents orthogonal projection onto $N_{f(x)}\Psi(M)$. Therefore,

$$\cot^2 \theta_c(f(x)) = \sup_{y \in M \setminus \{x\}} \frac{\|P_{N_{f(x)}\Psi(M)} f(y)\|^2_{H_f}}{(1 - \rho(x, y))^2}$$

and it remains to show that

$$\|P_{N_{f(x)}\Psi(M)}f(y)\|^2_{H_f} = \mathrm{Var}\left(f(y) - \rho(x,y)f(x) - \sum_{i=1}^k \mathrm{Cov}(f(y), X_i f(x))X_i f(x)\right).$$

This, however, follows from the fact that $N_{f(x)}\Psi(M)$ is the orthogonal complement (in $T_{f(x)}H_f$) of the subspace $L_x = \mathrm{span}\{f(x), X_1 f(x), \ldots, X_k f(x)\}$, and the fact that

$$\begin{aligned}\rho(x,y)f(x) + \sum_{i=1}^k \mathrm{Cov}(f(y), X_i f(x))X_i f(x) &= \langle f(x), f(y)\rangle_{H_f} f(x) + \sum_{i=1}^k \langle f(y), X_i f(x)\rangle_{H_f} X_i f(x) \\ &= P_{L_x} f(y).\end{aligned}$$

□

We now consider the case when M is a manifold without boundary and the supremum in

$$\cot^2(\theta_c) = \sup_{x\in M}\sup_{y\in M\setminus\{x\}} \frac{\|P_{N_{f(x)}\Psi(M)}f(y)\|^2_{H_f}}{(1-\rho(x,y))^2} \tag{26}$$

is achieved at some (x^*, y^*). Thinking of the critical variance as the cotangent of some critical distance on M, then, geometrically, the tube of radius θ_c should self-intersect itself along a geodesic from x^* to y^* in such a way that the tube, viewed locally from the point x^* shares a hyperplane with the tube viewed locally from the point y^*. Alternatively, at the point of self-intersection the outward pointing unit normal vectors should be pointing in opposite directions.

The simplest way of seeing this is to think of M as just two points $\{p_1, p_2\}$ in $\mathbb{R}^2$. In this case the tube of radius r around M consists of the union of two discs of radius r. When $r = d(p_1,p_2)/2$ the two discs self-intersect at exactly one point, and the unit normal vectors are pointing in opposite directions.

We can make this statement precise in the following lemma.

Lemma 5.2 *Suppose (x^*, y^*) achieve the supremum in (26). Then,*

$$f(y^*) = \cos(2\theta_c)\cdot f(x^*) + \sin(2\theta_c)\cdot v^*_{f(x^*)}$$

where

$$v^*_{f(x^*)} = \frac{P_{N_{f(x)}\Psi(M)}f(y^*)}{\|P_{N_{f(x)}\Psi(M)}f(y^*)\|_{H_f}}$$

is the direction of the geodesic between $f(x^*)$ and $f(y^*)$. For any $X_{x^*} \in T_{x^*}M$ and any $X_{y^*} \in T_{y^*}M$

$$\begin{aligned}\mathrm{Cov}(X_{x^*}f, f(y^*)) &= \langle \Psi_*(X_{x^*}), f(y^*)\rangle_{H_f} = 0\\ \mathrm{Cov}(X_{y^*}f, f(x^*)) &= \langle \Psi_*(X_{y^*}), f(x^*)\rangle_{H_f} = 0.\end{aligned}$$

Further, the partial map

$$\rho^{x^*}(y) \triangleq \rho(x^*, y)$$

has a local maximum at y^*, and the partial map

$$\rho^{y^*}(x) \triangleq \rho(x, y^*)$$

has a local maximum at x^*.

Proof: If (x^*, y^*) achieve the supremum in (26) then there exists a point z equidistant from $f(x^*)$ and $f(y^*)$ and unit vectors $v_{f(x^*)} \in N_{f(x^*)}\Psi(M)$ and $w_{f(y^*)} \in N_{f(y^*)}\Psi(M)$ such that

$$z = \cos\theta_c \cdot f(x^*) + \sin\theta_c \cdot v_{f(x^*)} = \cos\theta_c \cdot f(y^*) + \sin\theta_c w_{f(y^*)}.$$

It is not *a priori* obvious that $w_{f(y^*)} \in N_{f(y^*)}\Psi(M)$. However, if this were not the case, then exists $t \in M$, $t \neq y$ such that $d(t, z) < \theta_c$. This contradicts the assumption that (x^*, y^*) achieve the supremum in (26) and the critical radius is θ_c. The proof that this is a contradiction is left to the reader, though it follows similar lines to the argument at the end of this proof.

To prove the first claim, therefore, it is enough to show that the unit tangent vectors $\dot{c}_{f(x^*),v_{f(x^*)}}(\theta_c)$ and $\dot{c}_{f(y^*),w_{f(y^*)}}(\theta_c)$ at z satisfy

$$u = \dot{c}_{f(x^*),w_{f(x^*)}}(\theta_c) + \dot{c}_{f(y^*),w_{f(y^*)}}(\theta_c) = 0 \in T_z H_f.$$

Suppose then that $u \neq 0$. A simple calculation shows that

$$\langle u, f(x^*)\rangle_{H_f} = \langle u, f(y^*)\rangle_{H_f} \quad = \frac{\cos(2\theta_c) - \langle f(x^*), f(y^*)\rangle_{H_f}}{\sin\theta_c}.$$

If $u \neq 0$, then $f(x^*), f(y^*)$ and z are not on the same geodesic and the triangle inequality implies that

$$\cos(2\theta_c) < \langle f(x^*), f(y^*)\rangle_{H_f}$$

which implies

$$\langle u, f(x^*)\rangle_{H_f} = \langle u, f(y^*)\rangle_{H_f} < 0.$$

Consider the geodesic originating at z in the direction $u^* = u/\|u\|_{H_f}$

$$c_{z,u^*}(t) = \cos t \cdot z + \sin t \cdot u^*.$$

Assuming $u^* \neq 0$,

$$\langle \dot{c}_{z,u^*}(0), f(y^*)\rangle_{H_f} = \langle \dot{c}_{z,u^*}(0), f(x^*)\rangle_{H_f} = \langle u^*, f(x^*)\rangle < 0.$$

This implies that for sufficiently small $|s|$, $s < 0$

$$\langle f(x^*), c_{z,u^*}(s)\rangle_{H_f} = \langle f(y^*), c_{z,u^*}(s)\rangle_{H_f} > \cos\theta_c.$$

For such an s, there exist distinct points $\hat{x}(s)$ and $\hat{y}(s)$ such that

$$d(c_{z,u^*}(s), \hat{x}(s)) < \theta_c, \qquad d(c_{z,u^*}(s), \hat{y}(s)) < \theta_c$$

such that the geodesic connecting $\hat{x}(s)$ and $c_{z,u^*}(s)$ is normal to M at $\hat{x}(s)$, and the geodesic connecting $\hat{y}(s)$ and $c_{z,u^*}(s)$ is normal to M at $\hat{y}(s)$.

Without loss of generality we assume that

$$d(c_{z,u^*}(s), \hat{x}(s)) \leq d(c_{z,u^*}(s), \hat{y}(s)).$$

In this case, the geodesic connecting $\hat{y}$ to $c_{z,u^*}(s)$ is no longer a minimizer of distance once it passes the point $c_{z,u^*}(s)$, which is of distance strictly less than θ_c from $\hat{y}(s)$. That is, for some point $\hat{z}(s)$ along the geodesic connecting $\hat{y}(s)$ and $c_{z,u^*}(s)$, beyond $c_{z,u^*}(s)$ but of distance strictly less than θ_c from $\hat{y}(s)$, there exist points in M strictly closer to $\hat{z}(s)$ than $\hat{y}(s)$. We therefore have a contradiction, as the critical radius of $\Psi(M)$ is θ_c.

To prove the second claim, we note that $f(y^*)$ is a linear combination of $f(x^*)$ and $v_{f(x^*)}$ which are perpendicular to every vector in $T_{f(x^*)}\Psi(M)$. Similarly, $f(x^*)$ is a linear combination of $f(y^*)$ and $w_{f(y^*)}$ which are perpendicular to every vector in $T_{f(x^*)}\Psi(M)$. The fact that the partial maps $\rho^{x^*}(y)$ and $\rho^{y^*}(x)$ have local maxima follows from the same contradiction argument used above.

□

5.2 Centered stationary processes on $[0, T]$

In this section $f(x), x \in [0, T]$, is assumed to be a centered C^2 stationary process with unit variance and covariance function R. As in [11] we change the time scale so that $\mathrm{Var}(\dot{f}(x)) = -\ddot{R}(0) = 1$.

Fix a point $x \in (0, T)$, in which case the process f^x is given by

$$f^x(y) = \frac{f(y) - R(x-y)f(x) - \dot{R}(x-y)\dot{f}(x)}{1 - R(x,y)}$$

and the critical variance at x is given by

$$\begin{aligned}
\sigma_c^2(f, x) &= \sup_{y \in [0,T]\setminus\{x\}} \frac{\mathrm{Var}(f(y) \mid f(x), \dot{f}(x))}{(1 - R(x-y))^2} \\
&= \sup_{\substack{-x \leq t \leq T-x \\ t \neq 0}} \frac{1 - R(t)^2 - \dot{R}(t)^2}{(1 - R(t))^2}.
\end{aligned}$$

The critical variance in the interior is

$$\sigma_c^2(f, (0,T)) = \sup_{x \in (0,T)} \sigma_c^2(f, x) = \sup_{0<t<T} \frac{1 - R(t)^2 - \dot{R}(t)^2}{(1 - R(t))^2}. \tag{27}$$

A local version of critical variance $\sigma_{c,loc}^2(x, (0,T))$ is obtained by letting $t \to 0$ in (27), i.e.

$$\sigma_{c,loc}^2(x, (0,T)) = \lim_{t \to 0} \frac{1 - R(t)^2 - \dot{R}(t)^2}{(1 - R(t))^2} = 1 + \frac{4}{3}R^{(4)}(0),$$

where the last conclusion follows from some simple calculus. An alternative interpretation of the quantity $\sigma^2_{c,loc}(x,(0,T))$, also obtained by simple calculus is the following

$$\sigma^2_{c,loc}(x,(0,T)) = \mathrm{Var}\left(\ddot{f}(x)\middle|f(x)\right).$$

The local critical variance at the end points $\{0,T\}$ need slightly more attention. We consider the point $x=0$ without loss of generality. The normal cone at $x=0$ is

$$N_0[0,T] = \bigcup_{c\leq 0} c\frac{d}{dx} \subset T_0\mathbb{R}.$$

For ease of notation, we just write N_0 for $N_0[0,T]$.

The projection onto N_0 is just

$$P_{N_0}\left(a\frac{d}{dx}\right) = \mathbb{1}_{\{a\leq 0\}}\left(a\frac{d}{dx}\right).$$

As

$$\mathrm{Cov}\left(\dot{f}(0),f(y)\right) = -\dot{R}(y)$$

the process $f^0(y)$ is given by

$$f^0(y) = \frac{f(y)-R(y)f(0)-\mathbb{1}_{\{\dot{R}(y)\leq 0\}}\dot{R}(y)\dot{f}(0)}{(1-R(y))^2}.$$

The local critical variance at $x=0$ is given by

$$\sigma_c^2(f,0) = \sup_{0\leq t\leq T}\frac{1-R(t)^2-\dot{R}(t)^2+\max(\dot{R}(t),0)^2}{(1-R(t))^2} \geq \sup_{0\leq t\leq T}\frac{1-R(t)^2-\dot{R}(t)^2}{(1-R(t))^2}. \tag{28}$$

The inequality above implies that $\sigma_c^2(f,0)\geq\sigma_c^2(f,y),\ 0\leq y\leq T$. Therefore the critical variance $\sigma_c^2(f)$ is attained at the end points $t=0,T$.

Note that this does not exclude the case that the critical radius is also attained in the interior $0<t<T$. Under some circumstances, the critical variance is attained everywhere, as demonstrated in the following lemma.

Lemma 5.3 *Suppose f is a centered, unit variance C^2 stationary process on $\mathbb{R}$ such that $-\ddot{R}(0)=1$ and $\dot{R}(t)\leq 0$ for all $t\geq 0$. Then*

$$\sigma_c^2(f) = \sigma^2_{c,loc}(f,x) = \mathrm{Var}\left(\ddot{f}(x)\middle|f(x)\right).$$

Proof:

As noted above, we just have to compute $\sigma_c^2(f,0)$. Because $\dot{R}(t)\leq 0$ we see that

$$\sigma_c^2(f,0) = \sup_{0\leq t\leq T}\frac{1-R(t)^2-\dot{R}(t)^2}{(1-R(t))^2}.$$

Now suppose that this supremum is achieved. Lemma 5.2 implies that for any t that achieves this supremum, $\dot{R}(t) = 0$, and t is a local maximum of $R(t)$. Let

$$T^* = \left\{ t > 0 \middle| \dot{R}(t) = 0, R(t) = \cos(2\theta_c) \right\}.$$

As T^* is a closed set there exists a minimum value of T^*:

$$t^* = \min T^* > 0.$$

Because t^* is a local maximum of R, there exits $\epsilon > 0$ such that

$$R(t) \leq R(t^*), \qquad \forall t \in (t^* - \epsilon, t^*).$$

But $R(t)$ is assumed to be non-increasing and we have

$$R(t) \geq R(t^*), \qquad \forall t < t^*.$$

Therefore

$$R(t) = R(t^*), \qquad \forall t \in (t^* - \epsilon, t^*).$$

This, however, contradicts the minimality of t^*. $\square$.

5.3 Isotropic fields on $\mathbb{R}^m$ with monotone covariance functions

In this section, we compute the critical variance for centered, unit variance isotropic processes restricted to a compact, convex set M. In particular, we prove the following proposition, which shows that the exponential behaviour of $\mathrm{Diff}_{f,M}(u)$ for such processes is determined solely by the conditional variance of the second derivative, given the field.

Proposition 5.4 *Let $\widehat{f}$ be an isotropic process on $\mathbb{R}^m$ that induces the standard metric on $\mathbb{R}^m$ satisfying Assumption 2.1 and let $f = \widehat{f}_{|M}$ be the restriction of $\widehat{f}$ to a compact, convex set M with piecewise smooth boundary. If the covariance function*

$$R(\|x\|) = \mathrm{Cov}(f(x+t), f(t))$$

is monotone non-increasing, then the critical variance is attained locally and is given by

$$\sigma_c^2(f) = \mathrm{Var}\left(\frac{\partial^2 f}{\partial t_1^2}(0) \middle| f(0) \right).$$

Proof: Fix $t \in \overset{\circ}{M}$. Similar computations to those in Section 5.2 show that

$$\mathrm{Var}(f^t(s)) = \frac{1 - R^2(\|s-t\|) - \dot{R}^2(\|s-t\|)}{(1 - R(\|s-t\|))^2}.$$

Because R is assumed monotone non-increasing, the arguments used in Lemma 5.3 imply that

$$\sup_{s \in M \setminus \{t\}} \mathrm{Var}(f^t(s)) = \mathrm{Var}\left(\frac{\partial^2 f}{\partial t_1^2}(0) \middle| f(0) \right).$$

We therefore turn to the boundary ∂M, assumed to be piecewise smooth.

Fix $t \in \partial M$ and $s \neq t$. Let X_t^s be the unit vector in $T_t\mathbb{R}^m$ in the direction $s-t$. Since M is convex, $X_t^s \in \mathcal{S}_t M$ and

$$X_t^s \parallel \widehat{F}^t(s) \in \mathcal{S}_t M$$

so that $P_{N_x M}\widehat{F}^t(s) = 0$. Therefore,

$$\mathrm{Var}(f^t(s)) = \frac{1 - R^2(\|s-t\|) - \dot{R}^2(\|s-t\|)}{(1 - R(\|s-t\|))^2}.$$

We see that we do not incur any additional penalty for the local critical radius at the boundary points if M is convex. Again, the arguments of Lemma 5.3 imply that

$$\sigma_c^2(f,t) = \mathrm{Var}\left(\frac{\partial^2 f}{\partial t_1^2}(0)\middle| f(0)\right).$$

The conclusion now follows. □

References

[1] R. Adler and J. E. Taylor. *Random fields and their geometry.* Birkhäuser, Boston, 2004. To appear.

[2] R. J. Adler. *The Geometry of Random Fields.* John Wiley & Sons, Chichester, 1981.

[3] R. J. Adler. *An Introduction to Continuity, Extrema and Related Topics for General Gaussian Processes*, volume 12 of *IMS Lecture Notes-Monograph Series.* Institute of Mathematical Statistics, Hayward, CA., 1990.

[4] R. J. Adler. On excursion sets, tube formulae, and maxima of random fields. *The Annals of Applied Probability*, 10(1):1–74, 2000.

[5] D. R. Brillinger. On the number of solutions of systems of random equations. *Ann. Math. Statist.*, 43:534–540, 1972.

[6] J. Cao and K. Worsley. The geometry of correlation fields with an application to functional connectivity of the brain. *Ann. Appl. Probab.*, 9(4):1021–1057, 1999.

[7] J. Cao and K. J. Worsley. The detection of local shape changes via the geometry of Hotelling's T^2 fields. *Ann. Statist.*, 27(3):925–942, 1999.

[8] C. Delmas. An asymptotic expansion for the distribution of the maximum of a class of Gaussian fields. *C. R. Acad. Sci. Paris Sér. I Math.*, 327(4):393–397, 1998.

[9] H. Federer. Curvature measures. *Trans. Amer. Math. Soc.*, 93:418–491, 1959.

[10] S. Johansen and I. M. Johnstone. Hotelling's theorem on the volume of tubes: some illustrations in simultaneous inference and data analysis. *Ann. Statist.*, 18(2):652–684, 1990.

[11] M. F. Kratz and H. Rootzén. On the rate of convergence for extremes of mean square differentiable stationary normal processes. *J. Appl. Probab.*, 34(4):908–923, 1997.

[12] S. Kuriki, A. Takemura, and J. Taylor. Asymptotic evaluation of the remainder of tube formula approximation to tail probabilities. 2003. In preparation.

[13] S. Kuriki and J. Taylor. The tube method for Gaussian fields with inhomogeneous mean and/or variance – smooth and piecewise smooth cases. 2003. In preparation.

[14] V. I. Piterbarg. *Asymptotic methods in the theory of Gaussian processes and fields.* American Mathematical Society, Providence, RI, 1996. Translated from the Russian by V. V. Piterbarg, Revised by the author.

[15] J. Sun. Tail probabilities of the maxima of Gaussian random fields. *Ann. Probab.*, 21(1):34–71, 1993.

[16] A. Takemura and S. Kuriki. Some results on geometry of isotropic and spherically isotropic smooth gaussian fields. In preparation.

[17] A. Takemura and S. Kuriki. On the equivalence of the tube and Euler characteristic methods for the distribution of the maximum of Gaussian fields over piecewise smooth domains. *Ann. of Appl. Prob.*, 12(2):768–796, 2002.

[18] A. Takemura and S. Kuriki. Tail probability via the tube formula when the critical radius is zero. *Bernoulli*, 9(3):535–558, 2003.

[19] J. Taylor and R. J. Adler. Euler characteristics for Gaussian fields on manifolds. *Annals of Probability*, 31(2):533–563, 2003.

[20] H. Weyl. On the volume of tubes. *Amer. J. Math.*, 61:461–472, 1939.

[21] K. J. Worsley. Local maxima and the expected Euler characteristic of excursion sets of χ^2, F and t fields. *Adv. in Appl. Probab.*, 26(1):13–42, 1994.

[22] K. J. Worsley. Boundary corrections for the expected Euler characteristic of excursion sets of random fields, with an application to astrophysics. *Adv. in Appl. Probab.*, 27(4):943–959, 1995.