

## EFFICIENT MONTE CARLO FOR HIGH EXCURSIONS OF GAUSSIAN RANDOM FIELDS

BY ROBERT J. ADLER<sup>1</sup>, JOSE H. BLANCHET<sup>2</sup> AND JINGCHEN LIU<sup>3</sup>

*Technion, Columbia University and Columbia University*

Our focus is on the design and analysis of efficient Monte Carlo methods for computing tail probabilities for the suprema of Gaussian random fields, along with conditional expectations of functionals of the fields given the existence of excursions above high levels,  $b$ . Naïve Monte Carlo takes an exponential, in  $b$ , computational cost to estimate these probabilities and conditional expectations for a prescribed relative accuracy. In contrast, our Monte Carlo procedures achieve, at worst, polynomial complexity in  $b$ , assuming only that the mean and covariance functions are Hölder continuous. We also explain how to fine tune the construction of our procedures in the presence of additional regularity, such as homogeneity and smoothness, in order to further improve the efficiency.

**1. Introduction.** This paper centers on the design and analysis of efficient Monte Carlo techniques for computing probabilities and conditional expectations related to high excursions of Gaussian random fields. More specifically, suppose that  $f : T \times \Omega \rightarrow \mathbb{R}$  is a continuous Gaussian random field over a  $d$ -dimensional compact set  $T \subset \mathbb{R}^d$ . (Additional regularity conditions on  $T$  will be imposed below, as needed.)

Our focus is on tail probabilities of the form

$$(1.1) \quad w(b) = \mathbb{P}\left(\max_{t \in T} f(t) > b\right)$$

and on conditional expectations

$$(1.2) \quad \mathbb{E}\left[\Gamma(f) \mid \max_{t \in T} f(t) > b\right]$$

as  $b \rightarrow \infty$ , where  $\Gamma$  is a functional of the field, which, for concreteness we take to be positive and bounded.

---

Received May 2010; revised March 2011.

<sup>1</sup>Supported in part by US–Israel Binational Science Foundation, 2008262, ISF 853/10 and NSF Grant DMS-08-52227.

<sup>2</sup>Supported in part by NSF Grant DMS-09-02075, CMMI-0846816 and CMMI-1069064.

<sup>3</sup>Supported in part by Institute of Education Sciences, U.S. Department of Education, through Grant R305D100017 and NSF CMMI-1069064.

*MSC2010 subject classifications.* Primary 60G15, 65C05; secondary 60G60, 62G32.

*Key words and phrases.* Gaussian random fields, high-level excursions, Monte Carlo, tail distributions, efficiency.

While much of the paper will concentrate on estimating the exceedance probability (1.1), it is important to note that our methods, based on importance sampling, are broadly applicable to the efficient evaluation of conditional expectations of the form (1.2). Indeed, as we shall explain at the end of Section 4, our approach to efficient importance sampling is based on a procedure which mimics the conditional distribution of  $f$ , given that  $\max_T f(t) > b$ . Moreover, once an efficient (in a precise mathematical sense described in Section 2) importance sampling procedure is in place, it follows under mild regularity conditions on  $\Gamma$  that an efficient estimator for (1.2) is immediately obtained by exploiting an efficient estimator for (1.1).

The need for an efficient estimator of  $\omega(b)$  should be reasonably clear. Suppose one could simulate

$$f^* \triangleq \sup_{t \in T} f(t)$$

exactly (i.e., without bias) via naïve Monte Carlo. Such an approach would typically require a number of replications of  $f^*$  which would be exponential in  $b^2$  to obtain an accurate estimate (in relative terms). Indeed, since in great generality (see [20])  $w(b) = \exp(-cb^2 + o(b^2))$  as  $b \rightarrow \infty$  for some  $c \in (0, \infty)$ , it follows that the average of  $n$  i.i.d. Bernoulli trials each with success parameter  $w(b)$  estimates  $w(b)$  with a relative mean squared error equal to  $n^{-1/2}(1 - w(b))^{1/2}/w(b)^{1/2}$ . To control the size of the error therefore requires<sup>4</sup>  $n = \Omega(w(b)^{-1})$ , which is typically prohibitively large. In addition, there is also a problem in that typically  $f^*$  cannot be simulated exactly and that some discretization of  $f$  is required.

Our goal is to introduce and analyze simulation estimators that can be applied to a general class of Gaussian fields and that can be shown to require at most a polynomial number of function evaluations in  $b$  to obtain estimates with a prescribed relative error. The model of computation that we use to count function evaluations and the precise definition of an algorithm with polynomial complexity is given in Section 2. Our proposed estimators are, in particular, *asymptotically optimal*. (This property, which is a popular notion in the context of rare-event simulation (cf. [7, 13]) essentially requires that the second moments of estimators decay at the same exponential rate as the square of the first moments.) The polynomial complexity of our estimators requires to assume no more than that the underlying Gaussian field is Hölder continuous; see Theorem 3.1 in Section 3. Therefore, our methods provide means for efficiently computing probabilities and expectations associated with high excursions of Gaussian random fields in wide generality.

---

<sup>4</sup>Given  $h$  and  $g$  positive, we shall use the familiar asymptotic notation  $h(x) = O(g(x))$  if there is  $c < \infty$  such that  $h(x) \leq cg(x)$  for all  $x$  large enough;  $h(x) = \Omega(g(x))$  if  $h(x) \geq cg(x)$  if  $x$  is sufficiently large and  $h(x) = o(g(x))$  as  $x \rightarrow \infty$  if  $h(x)/g(x) \rightarrow 0$  as  $x \rightarrow \infty$ ; and  $h(x) = \Theta(g(x))$  if  $h(x) = O(g(x))$  and  $h(x) = \Omega(g(x))$ .

In the presence of enough smoothness, we shall also show how to design estimators that can be shown to be *strongly efficient*, in the sense that their associated coefficient of variation remains uniformly bounded as  $b \rightarrow \infty$ . Moreover, the associated path generation of the conditional field (given a high excursion) can, basically, be carried out with the same computational complexity as the unconditional sampling procedure (uniformly in  $b$ ). This is Theorem 3.3 in Section 3.

High excursions of Gaussian random fields appear in wide number of applications, including, but not limited to:

- Physical oceanography: Here the random field can be water pressure or surface temperature. See [4] for many examples.
- Cosmology: This includes the analysis of COBE and WMAP microwave data on a sphere or galactic density data; for example, [9, 25, 26].
- Quantum chaos: Here random planar waves replace deterministic (but unobtainable) solutions of Schrodinger equations; for example, the recent review [15].
- Brain mapping: This application is the most developed and very widely used. For example, the paper by Friston et al. [17] that introduced random field methodology to the brain imaging community has, at the time of writing, over 4,500 (Google) citations.

Many of these applications deal with twice differentiable, constant variance random fields, or random fields that have been normalized to have constant variance, the reason being that they require estimates of the tail probabilities (1.1) and these are only really well known in the smooth, constant (unit) variance case. In particular, it is known that, with enough smoothness assumptions,

$$\begin{aligned}
 (1.3) \quad & \liminf_{b \rightarrow \infty} -b^{-2} \log \left| P\left(\sup_{t \in T} f(t) \geq b\right) - E(\chi(\{t \in T : f(t) \geq b\})) \right| \\
 & \geq \frac{1}{2} + \frac{1}{2\sigma_c^2},
 \end{aligned}$$

where  $\chi(A)$  is the Euler characteristic of the set  $A$ , and the term  $\sigma_c^2$  is related to a geometric quantity known as the critical radius of  $T$  and depends on the covariance structure of  $f$ ; see [5, 27] for details. Since both the probability and the expectation in (1.3) are typically  $O(b^\ell \exp(-b^2/2))$  for some  $\ell \geq 0$  and large  $b$ , a more user friendly (albeit not quite as correct) way to write (1.3) is

$$(1.4) \quad P\left(\sup_{t \in T} f(t) \geq b\right) \approx E(\chi(\{t \in T : f(t) \geq b\})) \times (1 + O(e^{-cb^2}))$$

for some  $c$ .

The expectation in (1.3) and (1.4) has an explicit form that is readily computed for Gaussian and Gaussian-related random fields of constant variance (see [5, 6] for details), although if  $T$  is geometrically complicated or the covariance of  $f$  highly nonstationary there can be terms in the expectation that can only be evaluated numerically or estimated statistically; for example, [2, 28]. Nevertheless,

when available, (1.3) provides excellent approximations and simulation studies have shown that the approximations are numerically useful for quite moderate  $b$ , of the order of 2 standard deviations.

However, as we have already noted, (1.3) holds only for constant variance fields, which also need to be twice differentiable. In the case of less smooth  $f$ , other classes of results occur, in which the expansions are less reliable and, in addition, typically involve the unknown Pickands' constants; cf. [8, 24].

These are some of the reasons why, despite a well developed theory, Monte Carlo techniques still have a significant role to play in understanding the behavior of Gaussian random fields at high levels. The estimators proposed in this paper basically reduce the rare event calculations associated to high excursions in Gaussian random fields to calculations that are roughly comparable to the evaluation of expectations or integrals in which no rare event is involved. In other words, the computational complexity required to implement the estimators discussed here is similar in some sense to the complexity required to evaluate a given integral in finite dimension or an expectation where no tail parameter is involved. To the best of our knowledge these types of reductions have not been studied in the development of numerical methods for high excursions of random fields. This feature distinguishes the present work from the application of other numerical techniques that are generic (such as quasi-Monte Carlo and other numerical integration routines) and that in particular might be also applicable to the setting of Gaussian fields.

Contrary to our methods, which are designed to have provably good performance uniformly over the level of excursion, a generic numerical approximation procedure, such as quasi-Monte Carlo, will typically require an exponential increase in the number of function evaluations in order to maintain a prescribed level of relative accuracy. This phenomenon is unrelated to the setting of Gaussian random fields. In particular, it can be easily seen to happen even when evaluating a one-dimensional integral with a small integrand. On the other hand, we believe that our estimators can, in practice, be easily combined with quasi-Monte Carlo or other numerical integration methods. The rigorous analysis of such hybrid approaches, although of great interest, requires an extensive treatment and will be pursued in the future. As an aside, we note that quasi-Monte Carlo techniques have been used in the excursion analysis of Gaussian random fields in [8].

The remainder of the paper is organized as follows. In Section 2 we introduce the basic notions of polynomial algorithmic complexity, which are borrowed from the general theory of computation. Section 3 discusses the main results in light of the complexity considerations of Section 2. Section 4 provides a brief introduction to importance sampling, a simulation technique that we shall use heavily in the design of our algorithms. The analysis of finite fields, which is given in Section 5, is helpful to develop the basic intuition behind our procedures for the general case. Section 6 provides the construction and analysis of a polynomial time algorithm for high excursion probabilities of Hölder continuous fields. Finally, in Section 7, we add additional smoothness assumptions along with stationarity and explain how to

fine tune the construction of our procedures in order to further improve efficiency in these cases.

**2. Basic notions of computational complexity.** In this section we shall discuss some general notions of efficiency and computational complexity related to the approximation of the probability of the rare events  $\{B_b : b \geq b_0\}$ , for which  $P(B_b) \rightarrow 0$  as  $b \rightarrow \infty$ . In essence, efficiency means that computational complexity is, in some sense, controllable, uniformly in  $b$ . A notion that is popular in the efficiency analysis of Monte Carlo methods for rare events is weak efficiency (also known as asymptotic optimality) which requires that the coefficient of variation of a given estimator,  $L_b$  of  $P(B_b)$ , to be dominated by  $1/P(B_b)^\varepsilon$  for any  $\varepsilon > 0$ . More formally, we have:

DEFINITION 2.1. A family of estimators  $\{L_b : b \geq b_0\}$  is said to be polynomially efficient for estimating  $P(B_b)$  if  $E(L_b) = P(B_b)$ , and there exists a  $q \in (0, \infty)$  for which

$$(2.1) \quad \sup_{b \geq b_0} \frac{\text{Var}(L_b)}{[P(B_b)]^2 |\log P(B_b)|^q} < \infty.$$

Moreover, if (2.1) holds with  $q = 0$ , then the family is said to be strongly efficient.

Below we often refer to  $L_b$  as a *strongly (polynomially) efficient estimator*, by which we mean that the family  $\{L_b : b > 0\}$  is strongly (polynomially) efficient. In order to understand the nature of this definition let  $\{L_b^{(j)}, 1 \leq j \leq n\}$  be a collection of i.i.d. copies of  $L_b$ . The averaged estimator

$$\widehat{L}_n(b) = \frac{1}{n} \sum_{j=1}^n L_b^{(j)}$$

has a *relative* mean squared error equal to  $[\text{Var}(L_b)]^{1/2}/[n^{1/2}P(B_b)]$ . A simple consequence of Chebyshev’s inequality is that

$$P(|\widehat{L}_n(b)/P(B_b) - 1| \geq \varepsilon) \leq \frac{\text{Var}(L_b)}{\varepsilon^2 n P[(B_b)]^2}.$$

Thus, if  $L_b$  is polynomially efficient, and we wish to compute  $P(B_b)$  with at most  $\varepsilon$  relative error and at least  $1 - \delta$  confidence, it suffices to simulate

$$n = \Theta(\varepsilon^{-2} \delta^{-1} |\log P(B_b)|^q)$$

i.i.d. replications of  $L_b$ . In fact, in the presence of polynomial efficiency, the bound  $n = \Theta(\varepsilon^{-2} \delta^{-1} |\log P(B_b)|^q)$  can be boosted to  $n = \Theta(\varepsilon^{-2} \log(\delta^{-1}) |\log P(B_b)|^q)$  using the so-called median trick; see [23].

Naturally, the cost per replication must also be considered in the analysis, and we shall do so, but the idea is that evaluating  $P(B_b)$  via crude Monte Carlo would

require, given  $\varepsilon$  and  $\delta$ ,  $n = \Theta(1/P(B_b))$  replications. Thus a polynomially efficiently estimator makes the evaluation of  $P(B_b)$  exponentially faster relative to crude Monte Carlo, at least in terms of the number of replications.

Note that a direct application of deterministic algorithms (such as quasi-Monte Carlo or quadrature integration rules) might improve (under appropriate smoothness assumptions) the computational complexity relative to Monte Carlo, although only by a polynomial rate (i.e., the absolute error decreases to zero at rate  $n^{-p}$  for  $p > 1/2$ , where  $n$  is the number of function evaluations and  $p$  depends on the dimension of the function that one is integrating; see, e.g., [7]). We believe that the procedures that we develop in this paper can guide the construction of efficient deterministic algorithms with small relative error and with complexity that scales at a polynomial rate in  $|\log P(B_b)|$ . This is an interesting research topic that we plan to explore in the future.

An issue that we shall face in designing our Monte Carlo procedure is that, due to the fact that  $f$  will have to be discretized, the corresponding estimator  $\tilde{L}_b$  will not be unbiased. In turn, in order to control the relative bias with an effort that is comparable to the bound on the number of replications discussed in the preceding paragraph, one must verify that the relative bias can be reduced to an amount less than  $\varepsilon$  with probability at least  $1 - \delta$  at a computational cost of the form  $O(\varepsilon^{-q_0} |\log P(B_b)|^{q_1})$ . If  $\tilde{L}_b(\varepsilon)$  can be generated with  $O(\varepsilon^{-q_0} |\log P(B_b)|^{q_1})$  cost, and satisfying  $|P(B_b) - E\tilde{L}_b(\varepsilon)| \leq \varepsilon P(B_b)$ , and if

$$\sup_{b>0} \frac{\text{Var}(\tilde{L}_b(\varepsilon))}{P(B_b)^2 |\log P(B_b)|^q} < \infty$$

for some  $q \in (0, \infty)$ , then  $\hat{L}_n^l(b, \varepsilon) = \sum_{j=1}^n \tilde{L}_b^{(j)}(\varepsilon)/n$ , where the  $\tilde{L}_b^{(j)}(\varepsilon)$ 's are i.i.d. copies of  $\tilde{L}_b(\varepsilon)$ , satisfies

$$P(|\hat{L}_n^l(b, \varepsilon)/P(B_b) - 1| \geq 2\varepsilon) \leq \frac{\text{Var}(\tilde{L}_b(\varepsilon))}{\varepsilon^2 \times n \times P(B_b)^2}.$$

Consequently, taking  $n = \Theta(\varepsilon^{-2}\delta^{-1} |\log P(B_b)|^q)$  suffices to give an estimator with at most  $\varepsilon$  relative error and  $1 - \delta$  confidence, and the total computational cost is  $\Theta(\varepsilon^{-2-q_0}\delta^{-1} |\log P(B_b)|^{q+q_1})$ .

We shall measure computational cost in terms of function evaluations such as a single addition, a multiplication, a comparison, the generation of a single uniform random variable on  $T$ , the generation of a single standard Gaussian random variable and the evaluation of  $\Phi(x)$  for fixed  $x \geq 0$ , where

$$\Phi(x) = 1 - \Psi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-s^2/2} ds.$$

All of these function evaluations are assumed to cost at most a fixed amount  $\mathbf{c}$  of computer time. Moreover, we shall also assume that first- and second-order moment characteristics of the field, such as  $\mu(t) = Ef(t)$  and  $C(s, t) = \text{Cov}(f(t),$

$f(s)$ ) can be computed in at most  $\mathbf{c}$  units of computer time for each  $s, t \in T$ . We note that similar models of computation are often used in the complexity theory of continuous problems; see [29].

The previous discussion motivates the next definition which has its roots in the general theory of computation in both continuous and discrete settings [22, 29]. In particular, completely analogous notions in the setting of complexity theory of continuous problems lead to the notion of “tractability” of a computational problem [31].

**DEFINITION 2.2.** A Monte Carlo procedure is said to be a *fully polynomial randomized approximation scheme* (FPRAS) for estimating  $P(B_b)$  if, for some  $q, q_1, q_2 \in [0, \infty)$ , it outputs an averaged estimator that is guaranteed to have at most  $\varepsilon > 0$  relative error with confidence at least  $1 - \delta \in (0, 1)$  in  $\Theta(\varepsilon^{-q_1} \delta^{-q_2} |\log P(B_b)|^q)$  function evaluations.

The terminology adopted, namely FPRAS, is borrowed from the complexity theory of randomized algorithms for counting [22]. Many counting problems can be expressed as rare event estimation problems. In such cases it typically occurs that the previous definition (expressed in terms of a rare event probability) coincides precisely with the standard counting definition of a FPRAS (in which there is no reference to any rare event to estimate). This connection is noted, for instance, in [11]. Our terminology is motivated precisely by this connection.

By letting  $B_b = \{f^* > b\}$ , the goal in this paper is to design a class of fully polynomial randomized approximation schemes that are applicable to a general class of Gaussian random fields. In turn, since our Monte Carlo estimators will be based on importance sampling, it turns out that we shall also be able to straightforwardly construct FPRASs to estimate quantities such as  $E[\Gamma(f) | \sup_{t \in T} f(t) > b]$  for a suitable class of functionals  $\Gamma$  for which  $\Gamma(f)$  can be computed with an error of at most  $\varepsilon$  with a cost that is polynomial as function of  $\varepsilon^{-1}$ . We shall discuss this observation in Section 4, which deals with properties of importance sampling.

**3. Main results.** In order to state and discuss our main results we need some notation. For each  $s, t \in T$  define

$$\begin{aligned} \mu(t) &= E(f(t)), & \mathcal{C}(s, t) &= \text{Cov}(f(s), f(t)), \\ \sigma^2(t) &= \mathcal{C}(t, t) > 0, & r(s, t) &= \frac{\mathcal{C}(s, t)}{\sigma(s)\sigma(t)}. \end{aligned}$$

Moreover, given  $x \in \mathbb{R}^d$  and  $\beta > 0$  we write  $|x| = \sum_{j=1}^d |x_j|$ , where  $x_j$  is the  $j$ th component of  $x$ . We shall assume that, for each fixed  $s, t \in T$ , both  $\mu(t)$  and  $\mathcal{C}(s, t)$  can be evaluated in at most  $\mathbf{c}$  units of computing time.

Our first result shows that under modest continuity conditions on  $\mu, \sigma$  and  $r$  it is possible to construct an explicit FPRAS for  $w(b)$  under the following regularity conditions:

- (A1) the field  $f$  is almost surely continuous on  $T$ ;
- (A2) for some  $\delta > 0$  and  $|s - t| < \delta$ , the mean and variance functions satisfies

$$|\sigma(t) - \sigma(s)| + |\mu(t) - \mu(s)| \leq \kappa_H |s - t|^\beta;$$

- (A3) for some  $\delta > 0$  and  $|s - s'| < \delta, |t - t'| < \delta$  the correlation function of  $f$  satisfies

$$|r(t, s) - r(t', s')| \leq \kappa_H [|t - t'|^\beta + |s - s'|^\beta];$$

- (A4)  $0 \in T$ . There exist  $\kappa_0$  and  $\omega_d$  such that, for any  $t \in T$  and  $\varepsilon$  small enough,

$$m(B(t, \varepsilon) \cap T) \geq \kappa_0 \varepsilon^d \omega_d,$$

where  $m$  is the Lebesgue measure,  $B(t, \varepsilon) = \{s : |t - s| \leq \varepsilon\}$  and  $\omega_d = m(B(0, 1))$ .

The assumption is that  $0 \in T$  is of no real consequence and is adopted only for notational convenience.

**THEOREM 3.1.** *Suppose that  $f : T \rightarrow \mathbb{R}$  is a Gaussian random field satisfying conditions (A1)–(A4) above. Then, algorithm 6.1 provides a FPRAS for  $w(b)$ .*

The polynomial rate of the intrinsic complexity bound inherent in this result is discussed in Section 6, along with similar rates in results to follow. The conditions of the previous theorem are weak and hold for virtually all applied settings involving continuous Gaussian fields on compact sets.

Not surprisingly, the complexity bounds of our algorithms can be improved upon under additional assumptions on  $f$ . For example, in the case of finite fields (i.e., when  $T$  is finite) with a nonsingular covariance matrix, we can show that the complexity of the algorithm is actually bounded as  $b \nearrow \infty$ . We summarize this in the next result, whose proof, which is given in Section 5, is useful for understanding the main ideas behind the general procedure.

**THEOREM 3.2.** *Suppose that  $T$  is a finite set, and  $f$  has a nonsingular covariance matrix over  $T \times T$ . Then Algorithm 5.3 provides a FPRAS with  $q = 0$ .*

As we indicated above, the strategy behind the discrete case provides the basis for the general case. In the general situation, the underlying idea is to discretize the field with an appropriate sampling (discretization) rule that depends on the level  $b$  and the continuity characteristics of the field. The number of sampling points grows as  $b$  increases, and the complexity of the algorithm is controlled by finding a good sampling rule. There is a trade-off between the number of points sampled, which has a direct impact on the complexity of the algorithm, and the fidelity of the discrete approximation to the continuous field. Naturally, in the presence of enough smoothness and regularity, more information can be obtained with the same sample size. This point is illustrated in the next result, Theorem 3.3, which

considers smooth, homogeneous fields. Note that in addition to controlling the error induced by discretizing the field, *the variance is strongly controlled* and the *discretization rule is optimal*, in a sense explained in Section 7. For Theorem 3.3 we require the following additional regularity conditions:

(B1)  $f$  is homogeneous and almost surely twice continuously differentiable;

(B2)  $0 \in T \subset \mathbb{R}^d$  is a  $d$ -dimensional convex set with nonempty interior. Denoting its boundary by  $\partial T$ , assume that  $\partial T$  is a  $(d - 1)$ -dimensional manifold without boundary. For any  $t \in T$ , assume that there exists  $\kappa_0 > 0$  such that

$$m(B(t, \varepsilon) \cap T) \geq \kappa_0 \varepsilon^d$$

for any  $\varepsilon < 1$ , where  $m$  is Lebesgue measure.

**THEOREM 3.3.** *Let  $f$  satisfy conditions (B1) and (B2). Then Algorithm 7.3 provides a FPRAS. Moreover, the underlying estimator is strongly efficient and there exists a discretization scheme for  $f$  which is optimal in the sense of Theorem 7.4.*

The results stated in Theorem 3.3 are stronger than those in Theorem 3.1. This is because conditions (B1) and (B2) are much stronger than conditions (A1)–(A4). The structure present in Theorem 3.3 allows us to carry out a more refined complexity analysis. Using smoothness and homogeneity, the conditional distribution of the random field given a high excursion can be described quite precisely in an asymptotic sense using its derivatives. In our analysis we take advantage of such a conditional description, which is not available for Hölder continuous fields. On the other hand, it might be possible that the algorithms developed for Theorem 3.1, or closely related variations, are in fact strongly efficient for certain Hölder continuous fields. We leave this more refined analysis to future study.

**4. Importance sampling.** Importance sampling is based on the basic identity, for fixed measurable  $B$ ,

$$(4.1) \quad P(B) = \int \mathbb{1}(\omega \in B) dP(\omega) = \int \mathbb{1}(\omega \in B) \frac{dP}{dQ}(\omega) dQ(\omega),$$

where we assume that the probability measure  $Q$  is such that  $Q(\cdot \cap B)$  is absolutely continuous with respect to the measure  $P(\cdot \cap B)$ . If we use  $E^Q$  to denote expectation under  $Q$ , then (4.1) trivially yields that the random variable

$$L(\omega) = \mathbb{1}(\omega \in B) \frac{dP}{dQ}(\omega)$$

is an unbiased estimator for  $P(B) > 0$  under the measure  $Q$ , or, symbolically,  $E^Q L = P(B)$ .

An averaged importance sampling estimator based on the measure  $Q$ , which is often referred as an *importance sampling distribution* or a *change-of-measure*,

is obtained by simulating  $n$  i.i.d. copies  $L^{(1)}, \dots, L^{(n)}$  of  $L$  under  $Q$  and computing the empirical average  $\widehat{L}_n = (L^{(1)} + \dots + L^{(n)})/n$ . A central issue is that of selecting  $Q$  in order to minimize the variance of  $\widehat{L}_n$ . It is easy to verify that if  $Q^*(\cdot) = P(\cdot|B) = P(\cdot \cap B)/P(B)$ , then the corresponding estimator has zero variance. However,  $Q^*$  is clearly a change of measure that is of no practical value, since  $P(B)$ —the quantity that we are attempting to evaluate in the first place—is unknown. Nevertheless, when constructing a good importance sampling distribution for a family of sets  $\{B_b : b \geq b_0\}$  for which  $0 < P(B_b) \rightarrow 0$  as  $b \rightarrow \infty$ , it is often useful to analyze the asymptotic behavior of  $Q^*$  as  $P(B_b) \rightarrow 0$  in order to guide the construction of a good  $Q$ .

We now describe briefly how an efficient importance sampling estimator for  $P(B_b)$  can also be used to estimate a large class of conditional expectations given  $B_b$ . Suppose that a single replication of the corresponding importance sampling estimator,

$$L_b \stackrel{\Delta}{=} \mathbb{1}(\omega \in B_b) dP/dQ$$

can be generated in  $O(\log|P(B_b)|^{q_0})$  function evaluations, for some  $q_0 > 0$ , and that

$$\text{Var}(L_b) = O([P(B_b)]^2 \log|P(B_b)|^{q_0}).$$

These assumptions imply that by taking the average of i.i.d. replications of  $L_b$  we obtain a FPRAS.

Fix  $\beta \in (0, \infty)$ , and let  $\mathcal{X}(\beta, q)$  be the class of random variables  $X$  satisfying  $0 \leq X \leq \beta$  with

$$(4.2) \quad E[X|B_b] = \Omega[1/\log(P(B_b))^q].$$

Then, by noting that

$$(4.3) \quad \frac{E^Q(XL_b)}{E^Q(L_b)} = E[X|B_b] = \frac{E[X; B_b]}{P(B_b)},$$

it follows easily that a FPRAS can be obtained by constructing the natural estimator for  $E[X|B_b]$ ; that is, the ratio of the corresponding averaged importance sampling estimators suggested by the ratio in the left of (4.3). Of course, when  $X$  is difficult to simulate exactly, one must assume the bias  $E[X; B_b]$  can be reduced in polynomial time. The estimator is naturally biased, but the discussion on FPRAS on biased estimators given in Section 2 can be directly applied.

In the context of Gaussian random fields, we have that  $B_b = \{f^* > b\}$ , and one is very often interested in random variables  $X$  of the form  $X = \Gamma(f)$ , where  $\Gamma : C(T) \rightarrow \mathbb{R}$ , and  $C(T)$  denotes the space of continuous functions on  $T$ . Endowing  $C(T)$  with the uniform topology, consider functions  $\Gamma$  that are nonnegative and bounded by a positive constant. An archetypical example is given by the volume of (conditioned) high-level excursion sets with  $\beta = m(T)$  is known to satisfy (4.2).

However, there are many other examples of  $\mathcal{X}(\beta, q)$  with  $\beta = m(T)$  which satisfy (4.2) for a suitable  $q$ , depending on the regularity properties of the field. In fact, if the mean and covariance properties of  $f$  are Hölder continuous, then, using similar techniques as those given in the arguments of Section 6, it is not difficult to see that  $q$  can be estimated.

In case that  $\Gamma(f)$  is not bounded, the analysis is usually case-by-case. In particular, we need to investigate

$$E^Q(\Gamma^2(f)L_b^2) = E(\Gamma^2(f)L_b|B_b)P(B_b).$$

We provide a brief calculation for the case of the conditional overshoot, that is,  $\Gamma(f) = f^* - b$  and  $B_b = \{f^* > b\}$ . We admit the change of measure defined later in (6.5). Then, given  $\{f^* > b\}$ ,  $\Gamma^2(f)$  and  $L_b$  are negatively correlated (the higher the overshoot is, the larger the excursion set is), and we can obtain that

$$E^Q(\Gamma^2(f)L_b^2) \leq E(\Gamma^2(f)|B_b)E(L_b).$$

Conditional on the occurrence of  $\{f^* > b\}$ ,  $b\Gamma(f)$  asymptotically follows an exponential distribution. Therefore,  $E(\Gamma^2(f)|B_b) = (1 + o(1))E^2(\Gamma(f)|B_b)$ . Together with the FPRAS of  $L_b$  in computing  $P(B_b)$ ,  $\Gamma(f)L_b$  is an FPRAS to compute the conditional overshoot. The corresponding numerical examples are given in Section 8. Two key steps involve the analyses of the conditional correlation of  $\Gamma^2(f)$  and  $L_b$  and the conditional distribution of  $\Gamma(f)$  given  $B_b$ .

Thus, we have that a FPRAS based importance sampling algorithm for  $w(b)$  would typically also yield a polynomial time algorithm for functional characteristics of the conditional field given high level excursions. Since this is a very important and novel application, we devote the remainder of this paper to the development of efficient importance sampling algorithms for  $w(b)$ .

**5. The basic strategy: Finite fields.** In this section we develop our main ideas in the setting in which  $T$  is a finite set of the form  $T = \{t_1, \dots, t_M\}$ . To emphasize the discrete nature of our algorithms in this section, we write  $X_i = f(t_i)$  for  $1, \dots, M$  and set  $X = (X_1, \dots, X_M)$ . This section is mainly of an expository nature, since much of it has already appeared in [3]. Nevertheless, it is included here as a useful guide to the intuition behind the algorithms for the continuous case.

We have already noted that in order to design an efficient importance sampling estimator for  $w(b) = P(\max_{1 \leq i \leq M} X_i > b)$  it is useful to study the asymptotic conditional distribution of  $X$ , given that  $\max_{1 \leq i \leq M} X_i > b$ . Thus, we begin with some basic large deviation results.

PROPOSITION 5.1. *For any set of random variables  $X_1, \dots, X_M$ ,*

$$\max_{1 \leq i \leq M} P(X_i > b) \leq P\left(\max_{1 \leq i \leq M} X_i > b\right) \leq \sum_{j=1}^M P(X_j > b).$$

Moreover, if the  $X_j$  are mean zero, Gaussian, and the correlation between  $X_i$  and  $X_j$  is strictly less than 1, then

$$P(X_i > b, X_j > b) = o(\max[P(X_i > b), P(X_j > b)]).$$

Thus, if the associated covariance matrix of  $X$  is nonsingular,

$$w(b) = (1 + o(1)) \sum_{j=1}^M P(X_j > b).$$

PROOF. The lower bound in the first display is trivial, and the upper bound follows by the union bound. The second display follows easily by working with the joint density of a bivariate Gaussian distribution (e.g., [10, 21]) and the third claim is a direct consequence of the inclusion–exclusion principle.  $\square$

As noted above,  $\mathcal{Q}^*$  corresponds to the conditional distribution of  $X$  given that  $X^* \triangleq \max_{1 \leq i \leq M} X_i > b$ . It follows from Proposition 5.1 that, conditional on  $X^* > b$ , the probability that two or more  $X_j$  exceed  $b$  is negligible. Moreover, it also follows that

$$P(X_i = X^* | X^* > b) = (1 + o(1)) \frac{P(X_i > b)}{\sum_{j=1}^M P(X_j > b)}.$$

The following corollary now follows as an easy consequence of these observations.

COROLLARY 5.2.

$$d_{\text{TV}}(\mathcal{Q}^*, \mathcal{Q}) \rightarrow 0$$

as  $b \rightarrow \infty$ , where  $d_{\text{TV}}$  denotes the total variation norm, and  $\mathcal{Q}$  is defined, for Borel  $B \subset \mathbb{R}^M$ , as

$$\mathcal{Q}(X \in B) = \sum_{i=1}^M p(i, b) P[X \in B | X_i > b],$$

where

$$p(i, b) = \frac{P(X_i > b)}{\sum_{j=1}^M P(X_j > b)}.$$

PROOF. Pick an arbitrary Borel  $B$ . Then we have that

$$\begin{aligned} \mathcal{Q}^*(X \in B) &= \frac{P[X \in B, \max_{1 \leq i \leq M} X_i > b]}{w(b)} \leq \sum_{i=1}^M \frac{P[X \in B, X_i > b]}{w(b)} \\ &= \sum_{i=1}^M P[X \in B | X_i > b] \frac{p(i, b)}{(1 + o(1))}. \end{aligned}$$

The above, which follows from the union bound and the last part of Proposition 5.1 combined with the definition of  $\mathcal{Q}$ , yields that for each  $\varepsilon > 0$  there exists  $b_0$  (independent of  $B$ ) such that, for all  $b \geq b_0$ ,

$$\mathcal{Q}^*(X \in B) \leq \mathcal{Q}(X \in B)/(1 - \varepsilon).$$

The lower bound follows similarly, using the inclusion–exclusion principle and the second part of Proposition 5.1.  $\square$

Corollary 5.2 provides support for choosing  $\mathcal{Q}$  as an importance sampling distribution. Of course, we still have to verify that the corresponding algorithm is a FPRAS. The importance sampling estimator induced by  $\mathcal{Q}$  takes the form

$$(5.1) \quad L_b = \frac{dP}{d\mathcal{Q}} = \frac{\sum_{j=1}^M P(X_j > b)}{\sum_{j=1}^M \mathbb{1}(X_j > b)}.$$

Note that under  $\mathcal{Q}$  we have that  $X^* > b$  almost surely, so the denominator in the expression for  $L_b$  is at least 1. Therefore, we have that

$$E^{\mathcal{Q}}L_b^2 \leq \left( \sum_{j=1}^M P(X_j > b) \right)^2,$$

and by virtue of Proposition 5.1 we conclude (using  $\text{Var}_{\mathcal{Q}}$  to denote the variance under  $\mathcal{Q}$ ) that

$$\frac{\text{Var}_{\mathcal{Q}}(L_b)}{P(X^* > b)^2} \rightarrow 0$$

as  $b \rightarrow \infty$ . In particular, it follows that  $L_b$  is strongly efficient.

Our proposed algorithm can now be summarized as follows.

ALGORITHM 5.3. There are two steps in the algorithm:

STEP (1). Simulate  $n$  i.i.d. copies  $X^{(1)}, \dots, X^{(n)}$  of  $X$  from the distribution  $\mathcal{Q}$ .

STEP (2). Compute and output

$$\widehat{L}_n = \frac{1}{n} \sum_{i=1}^n L_b^{(i)},$$

where  $L_b^{(i)} = \sum_{j=1}^M P(X_j^{(i)} > b) / \sum_{j=1}^M \mathbb{1}(X_j^{(i)} > b)$ .

Since the generation of  $L_i$  under  $\mathcal{Q}$  takes  $O(M^3)$  function evaluations we conclude, based on the analysis given in Section 2, that Algorithm 5.3 is a FPRAS with  $q = 0$ . This implies Theorem 3.2, as promised.

**6. A FPRAS for Hölder continuous Gaussian fields.** In this section we shall describe the algorithm and the analysis behind Theorem 3.1. Throughout the section, unless stated otherwise, we assume conditions (A1)–(A4) of Section 3.

There are two issues related to the complexity analysis. First, since  $f$  is assumed continuous, the entire field cannot be generated in a (discrete) computer, and so the algorithm used in the discrete case needs adaptation. Once this is done, we need to carry out an appropriate variance analysis.

Developing an estimator directly applicable to the continuous field will be carried out in Section 6.1. This construction will not only be useful when studying the performance of a suitable discretization, but will also help to explain some of the features of our discrete construction. Then, in Section 6.2, we introduce a discretization approach and study the bias caused by the discretization. In addition, we provide bounds on the variance of this discrete importance sampling estimator.

6.1. *A continuous estimator.* We start with a change of measure motivated by the discrete case in Section 5. A natural approach is to consider an importance sampling strategy analogous to that of Algorithm 5.3. The continuous adaptation involves first sampling  $\tau_b$  according to the probability measure

$$(6.1) \quad P(\tau_b \in \cdot) = \frac{E[m(A_b \cap \cdot)]}{E[m(A_b)]},$$

where  $A_b = \{t \in T : f(t) > b\}$ . The idea of introducing  $\tau_b$  in the continuous setting is not necessarily to locate the point at which the maximum is achieved, as was the situation in the discrete case. Rather,  $\tau_b$  will be used to find a random point which has a reasonable probability of being in the excursion set  $A_b$ . (This probability will tend to be higher if  $f$  is nonhomogenous.) This relaxation will prove useful in the analysis of the algorithm. Note that  $\tau_b$ , with the distribution indicated in (6.1), has a density function (with respect to Lebesgue measure) given by

$$h_b(t) = \frac{P(f(t) > b)}{E[m(A_b)]},$$

and that we also can write

$$E[m(A_b)] = E \int_T \mathbb{1}(f(t) > b) dt = \int_T P(f(t) > b) dt = m(T)P(f(U) > b),$$

where  $U$  is uniformly distributed over  $T$ .

Once  $\tau_b$  is generated, the natural continuous adaptation corresponding to the strategy described by Algorithm 5.3 proceeds by sampling  $f$  conditional on  $f(\tau_b) > b$ . Note that if we use  $\bar{Q}$  to denote the change-of-measure induced by such a continuous sampling strategy, then the corresponding importance sampling estimator takes the form

$$\bar{L}_b = \frac{dP}{d\bar{Q}} = \frac{E[m(A_b)]}{m(A_b)}.$$

The second moment of the estimator then satisfies

$$\begin{aligned}
 (6.2) \quad E\bar{Q}[(\bar{L}_b)^2] &= E(\bar{L}_b; A_b \neq \emptyset) \\
 &= E[m(A_b)]P(f^* > b)E[m(A_b)^{-1}|A_b \neq \emptyset].
 \end{aligned}$$

Unfortunately, it is easy to construct examples for which  $E\bar{Q}[(\bar{L}_b)^2]$  is infinite. For instance, consider a homogeneous and twice differentiable random field with zero mean and unit variance living on  $T = [0, 1]^d$ . Using the Slepian model, discussed in Section 7, it follows that the asymptotic distribution of the overshoot given  $\{f^* > b\}$  satisfies

$$b(f^* - b) \rightarrow S,$$

weakly as  $b \rightarrow \infty$  where  $S$  is an exponential random variable. Consequently, the distribution of  $m(A_b)$  given  $m(A_b) > 0$  satisfies

$$m(A_b) \rightarrow \kappa b^{-d} S^{d/2}$$

for some constant  $\kappa$ . Therefore, the second moment in (6.2) is infinity as long as  $d \geq 2$ . This example suggests that the construction of the change of measure needs to be modified slightly.

Extreme value theory considerations similar to those explained in the previous paragraph give that the overshoot of  $f$  over a given level  $b$  will be of order  $\Theta(1/b)$ . Thus, in order to keep  $\tau_b$  reasonably close to the excursion set, we shall also consider the possibility of an undershoot of size  $\Theta(1/b)$  right at  $\tau_b$ . As we shall see, this relaxation will allow us to prevent the variance in (6.2) becoming infinite. Thus, instead of (6.1), we shall consider  $\tau_{b-a/b}$  with density

$$(6.3) \quad h_{b-a/b}(t) = \frac{P(f(t) > b - a/b)}{E[m(A_{b-a/b})]}$$

for some  $a > 0$ . To ease on later notation, write

$$\gamma_{a,b} \triangleq b - a/b, \quad \tau_{\gamma_{a,b}} = \tau_{b-a/b}.$$

Let  $Q'$  be the change of measure induced by sampling  $f$  as follows. Given  $\tau_{\gamma_{a,b}}$ , sample  $f(\tau_{\gamma_{a,b}})$  conditional on  $f(\tau_{\gamma_{a,b}}) > \gamma_{a,b}$ . In turn, the rest of  $f$  follows its conditional distribution (under the nominal, or original, measure) given the observed value  $f(\tau_{\gamma_{a,b}})$ . We then have that the corresponding Radon–Nikodym derivative is

$$(6.4) \quad \frac{dP}{dQ'} = \frac{E[m(A_{\gamma_{a,b}})]}{m(A_{\gamma_{a,b}})},$$

and the importance sampling estimator  $L'_b$  is

$$(6.5) \quad L'_b = \frac{dP}{dQ'} \mathbb{1}(A_b \neq \emptyset) = \frac{E[m(A_{\gamma_{a,b}})]}{m(A_{\gamma_{a,b}})} \mathbb{1}(m(A_b) > 0).$$

Note that we have used the continuity of the field in order to write  $\{m(A_b) > 0\} = \{A_b \neq \emptyset\}$  almost surely. The motivation behind this choice lies in the fact that since  $m(A_{\gamma_{a,b}}) > m(A_b) > 0$ , the denominator may now be big enough to control the second moment of the estimator. In particular, we consider the homogeneous and twice differentiable field mentioned previously. Given  $m(A_b) > 0$ ,  $m(A_{\gamma_{a,b}})$  is asymptotically lower bounded by  $\kappa a^{d/2} b^{-d}$ . As we shall see, introducing the undershoot of size  $a/b$  will be very useful in the technical development both in the remainder of this section and in Section 7. In addition, its introduction also provides insight into the appropriate form of the estimator needed when discretizing the field.

6.2. *Algorithm and analysis.* We still need to face the problem of generating  $f$  in a computer. Thus we now concentrate on a suitable discretization scheme, still having in mind the change of measure leading to (6.4). Since our interest is to ultimately design algorithms that are efficient for estimating expectations such as  $E[\Gamma(f) | f^* > b]$ , where  $\Gamma$  may be a functional of the whole field, we shall use a global discretization scheme.

Consider  $U = (U_1, \dots, U_M)$  where  $U_i$  are i.i.d. uniform random variables taking values in  $T$  and independent of the field  $f$ . Set  $T_M = \{U_1, \dots, U_M\}$  and  $X_i = X_i(U_i) = f(U_i)$  for  $1 \leq i \leq M$ . Then  $X = (X_1, \dots, X_m)$  (conditional on  $U$ ) is a multivariate Gaussian random vector with conditional means  $\mu(U_i) \triangleq E(X_i | U_i)$  and covariances  $\mathcal{C}(U_i, U_j) \triangleq \text{Cov}(X_i, X_j | U_i, U_j)$ . Our strategy is to approximate  $w(b)$  by

$$w_M(\gamma_{a,b}) = P\left(\max_{t \in T_M} f(t) > \gamma_{a,b}\right) = E\left[P\left(\max_{1 \leq i \leq M} X_i > \gamma_{a,b} | U\right)\right].$$

Given the development in Section 5, it might not be surprising that if we can ensure that  $M = M(\varepsilon, b)$  is polynomial in  $1/\varepsilon$  and  $b$ , then we shall be in a good position to develop a FPRAS. The idea is to apply an importance sampling strategy similar to that we considered in the construction of  $L'_b$  of (6.5), but this time it will be conditional on  $U$ . In view of our earlier discussions, we propose sampling from  $Q''$  defined via

$$Q''(X \in B | U) = \sum_{i=1}^M p_U(i) P[X \in B | X_i > \gamma_{a,b}, U],$$

where

$$p_U(i) = \frac{P(X_i > \gamma_{a,b} | U)}{\sum_{j=1}^M P(X_j > \gamma_{a,b} | U)}.$$

We then obtain the (conditional) importance sampling estimator

$$(6.6) \quad L_b(U) = \frac{\sum_{i=1}^M P(X_i > \gamma_{a,b} | U)}{\sum_{i=1}^M \mathbb{1}(X_i > \gamma_{a,b})} \mathbb{1}\left(\max_{i=1}^M X_i > \gamma_{a,b}\right).$$

Note that the event  $\{\max_{i=1}^M X_i > \gamma_{a,b}\}$  occurs with probability 1 under  $Q''$ . Therefore, the indicator  $I(\max_{i=1}^M X_i > \gamma_{a,b})$  will be omitted when it does not cause confusion. It is clear that

$$w_M(\gamma_{a,b}) = E^{Q''} [L_b(U)].$$

Suppose for the moment that  $M$ ,  $a$  and the number of replications  $n$  have been chosen. Our future analysis will, in particular, guide the selection of these parameters. Then the procedure is summarized by the next algorithm.

ALGORITHM 6.1. The algorithm has three steps:

STEP (1). Simulate  $U^{(1)}, \dots, U^{(n)}$  which are  $n$  i.i.d. copies of the vector  $U = (U_1, \dots, U_M)$  described above.

STEP (2). Conditional on each  $U^{(i)}$ , for  $i = 1, \dots, n$ , generate  $L_b^{(i)}(U^{(i)})$  as described by (6.6) by considering the distribution of  $X^{(i)}(U^{(i)}) = (X_1^{(i)}(U_1^{(i)}), \dots, X_M^{(i)}(U_M^{(i)}))$ . Generate the  $X^{(i)}(U^{(i)})$  independently so that at the end we obtain that the  $L_b^{(i)}(U^{(i)})$  are  $n$  i.i.d. copies of  $L_b(U)$ .

STEP (3). Output

$$\widehat{L}_n(U^{(1)}, \dots, U^{(n)}) = \frac{1}{n} \sum_{i=1}^n L_b^{(i)}(U^{(i)}).$$

6.3. *Running time of Algorithm 6.1: Bias and variance control.* The remainder of this section is devoted to the analysis of the running time of the Algorithm 6.1. The first step lies in estimating the bias and second moment of  $L_b(U)$  under the change of measure induced by the sampling strategy of the algorithm, which we denote by  $Q''$ . We start with a simple bound for the second moment.

PROPOSITION 6.2. *There exists a finite  $\lambda_0$ , depending on  $\mu_T = \max_{t \in T} |\mu(t)|$  and  $\sigma_T^2 = \max_{t \in T} \sigma^2(t)$ , for which*

$$E^{Q''} [L_b(U)^2] \leq \lambda_0 M^2 P\left(\max_{t \in T} f(t) > b\right)^2.$$

PROOF. Observe that

$$\begin{aligned} E^{Q''} [L_b(U)^2] &\leq E\left(\left(\sum_{i=1}^M P(X_i > \gamma_{a,b} | U_i)\right)^2\right) \\ &\leq E\left(\left(\sum_{i=1}^M \sup_{t \in T} P(f(U_i) > \gamma_{a,b} | U_i = t)\right)^2\right) \end{aligned}$$

$$\begin{aligned} &= M^2 \max_{t \in T} P(f(t) > \gamma_{a,b})^2 \\ &\leq \lambda_0 M^2 \max_{t \in T} P(f(t) > b)^2 \\ &\leq \lambda_0 M^2 P\left(\max_{t \in T} f(t) > b\right)^2, \end{aligned}$$

which completes the proof.  $\square$

Next we obtain a preliminary estimate of the bias.

PROPOSITION 6.3. *For each  $M \geq 1$  we have*

$$\begin{aligned} |w(b) - w_M(\gamma_{a,b})| &\leq E[\exp(-Mm(A_{\gamma_{a,b}})/m(T)); A_b \cap T \neq \emptyset] \\ &\quad + P\left(\max_{t \in T} f(t) > \gamma_{a,b}, \max_{t \in T} f(t) \leq b\right). \end{aligned}$$

PROOF. Note that

$$\begin{aligned} |w(b) - w_M(\gamma_{a,b})| &\leq P\left(\max_{t \in T_M} f(t) \leq \gamma_{a,b}, \max_{t \in T} f(t) > b\right) \\ &\quad + P\left(\max_{t \in T_M} f(t) > \gamma_{a,b}, \max_{t \in T} f(t) \leq b\right). \end{aligned}$$

The second term is easily bounded by

$$P\left(\max_{t \in T_M} f(t) > \gamma_{a,b}, \max_{t \in T} f(t) \leq b\right) \leq P\left(\max_{t \in T} f(t) > \gamma_{a,b}, \max_{t \in T} f(t) \leq b\right).$$

The first term can be bounded as follows:

$$\begin{aligned} &P\left(\max_{t \in T_M} f(t) \leq \gamma_{a,b}, \max_{t \in T} f(t) > b\right) \\ &\leq E[(P[f(U_i) \leq \gamma_{a,b} | f])^M \mathbb{1}(A_b \cap T \neq \emptyset)] \\ &\leq E[(1 - m(A_{\gamma_{a,b}})/m(T))^M; A_b \cap T \neq \emptyset] \\ &\leq E[\exp(-Mm(A_{\gamma_{a,b}})/m(T)); A_b \cap T \neq \emptyset]. \end{aligned}$$

This completes the proof.  $\square$

The previous proposition shows that controlling the *relative* bias of  $L_b(U)$  requires finding bounds for

$$(6.7) \quad E[\exp(-Mm(A_{\gamma_{a,b}})/m(T)); A_b \cap T \neq \emptyset]$$

and

$$(6.8) \quad P\left(\max_{t \in T} f(t) > \gamma_{a,b}, \max_{t \in T} f(t) \leq b\right),$$

and so we develop these. To bound (6.7) we take advantage of the importance sampling strategy based on  $Q'$  introduced earlier in (6.4). Write

$$(6.9) \quad \begin{aligned} & E[\exp(-Mm(A_{\gamma_{a,b}})/m(T)); m(A_b) > 0] \\ &= E^{Q'}\left(\frac{\exp(-Mm(A_{\gamma_{a,b}})/m(T))}{m(A_{\gamma_{a,b}})}; m(A_b) > 0\right)Em(A_{\gamma_{a,b}}). \end{aligned}$$

Furthermore, note that for each  $\alpha > 0$  we have

$$(6.10) \quad \begin{aligned} & E^{Q'}\left(\frac{\exp(-Mm(A_{\gamma_{a,b}})/m(T))}{m(A_{\gamma_{a,b}})}; m(A_b) > 0\right) \\ &\leq \alpha^{-1} \exp(-M\alpha/m(T)) \\ &\quad + E^{Q'}\left(\frac{\exp(-Mm(A_{\gamma_{a,b}})/m(T))}{m(A_{\gamma_{a,b}})}; m(A_{\gamma_{a,b}}) \leq \alpha; m(A_b) > 0\right). \end{aligned}$$

The next result, whose proof is given in Section 6.4, gives a bound for the above expectation.

**PROPOSITION 6.4.** *Let  $\beta$  be as in conditions (A2) and (A3). For any  $v > 0$ , there exist constants  $\kappa, \lambda_2 \in (0, \infty)$  [independent of  $a \in (0, 1)$  and  $b$ , but dependent on  $v$ ] such that if we select*

$$\alpha^{-1} \geq \kappa^{d/\beta} (b/a)^{(2+v)2d/\beta},$$

and define  $W$  such that  $P(W > x) = \exp(-x^{\beta/d})$  for  $x \geq 0$ , then

$$(6.11) \quad \begin{aligned} & E^{Q'}\left(\frac{\exp(-Mm(A_{\gamma_{a,b}})/m(T))}{m(A_{\gamma_{a,b}})}; m(A_{\gamma_{a,b}}) \leq \alpha; m(A_b) > 0\right) \\ &\leq EW^2 \frac{m(T)}{\lambda_2^{2d/\beta} M} \left(\frac{b}{a}\right)^{4d/\beta}. \end{aligned}$$

The following result gives us a useful upper bound on (6.8). The proof is given in Section 6.5.

**PROPOSITION 6.5.** *Assume that conditions (A2) and (A3) are in force. For any  $v > 0$ , let  $\rho = 2d/\beta + dv + 1$ , where  $d$  is the dimension of  $T$ . There exist constants  $b_0, \lambda \in (0, \infty)$  [independent of  $a$  but depending on  $\mu_T = \max_{t \in T} |\mu(t)|$ ,  $\sigma_T^2 = \max_{t \in T} \sigma^2(t)$ ,  $v$ , the Hölder parameters  $\beta$  and  $\kappa_H$ ] so that for all  $b \geq b_0 \geq 1$  we have*

$$(6.12) \quad P\left(\max_{t \in T} f(t) \leq b + a/b \mid \max_{t \in T} f(t) > b\right) \leq \lambda ab^\rho.$$

Consequently,

$$\begin{aligned} &P\left(\max_{t \in T} f(t) > \gamma_{a,b}, \max_{t \in T} f(t) \leq b\right) \\ &= P\left(\max_{t \in T} f(t) \leq b \mid \max_{t \in T} f(t) > \gamma_{a,b}\right)P\left(\max_{t \in T} f(t) > \gamma_{a,b}\right) \\ &\leq \lambda ab^\rho P\left(\max_{t \in T} f(t) > \gamma_{a,b}\right). \end{aligned}$$

Moreover,

$$P\left(\max_{t \in T} f(t) > \gamma_{a,b}\right)(1 - \lambda ab^\rho) \leq P\left(\max_{t \in T} f(t) > b\right).$$

Propositions 6.4 and 6.5 allow us to prove Theorem 3.1, which is rephrased in the form of the following theorem, which contains the detailed rate of complexity and so the main result of this section.

**THEOREM 6.6.** *Suppose  $f$  is a Gaussian random field satisfying conditions (A1)–(A4) in Section 3. Given any  $v > 0$ , put  $a = \varepsilon/(4\lambda b^\rho)$  (where  $\lambda$  and  $\rho$  as in Proposition 6.5), and  $\alpha^{-1} = \kappa^{d/\beta}(b/a)^{(2+v)d/\beta}$ . Then, there exist  $c, \varepsilon_0 > 0$  such that for all  $\varepsilon \leq \varepsilon_0$ ,*

$$(6.13) \quad |w(b) - w_M(\gamma_{a,b})| \leq w(b)\varepsilon,$$

if  $M = \lceil c\varepsilon^{-1}(b/a)^{(4+4v)d/\beta} \rceil$ . Consequently, by our discussion in Section 2 and the bound on the second moment given in Proposition 6.2, it follows that Algorithm 6.1 provides a FPRAS with running time  $O((M)^3 \times (M)^2 \times \varepsilon^{-2}\delta^{-1})$ .

**PROOF.** Combining (6.3), (6.9) and (6.10) with Propositions 6.3–6.5 we have that

$$\begin{aligned} &|w(b) - w_M(\gamma_{a,b})| \\ &\leq \alpha^{-1} \exp(-M\alpha/m(T))E[m(A_{\gamma_{a,b}})] \\ &\quad + E[W^2] \frac{m(T)}{\lambda_2^{2d/\beta} M} \left(\frac{b}{a}\right)^{4d/\beta} E[m(A_{\gamma_{a,b}})] + \left(\frac{\lambda ab^\rho}{1 - \lambda ab^\rho}\right)w(b). \end{aligned}$$

Furthermore, there exists a constant  $K \in (0, \infty)$  such that

$$Em(A_{\gamma_{a,b}}) \leq K \max_{t \in T} P(f(t) > b)m(T) \leq Kw(b)m(T).$$

Therefore, we have that

$$\begin{aligned} \frac{|w(b) - w_M(\gamma_{a,b})|}{w(b)} &\leq \alpha^{-1} Km(T) \exp(-M\alpha/m(T)) \\ &\quad + E[W^2]K \frac{m(T)^2}{\lambda_2^{2d/\beta} M} \left(\frac{b}{a}\right)^{4d/\beta} + \left(\frac{\lambda ab^\rho}{1 - \lambda ab^\rho}\right). \end{aligned}$$

Moreover, since  $a = \varepsilon/(4\lambda b^\rho)$ , we obtain that, for  $\varepsilon \leq 1/2$ ,

$$\frac{|w(b) - w_M(\gamma_{a,b})|}{w(b)} \leq \alpha^{-1} K m(T) \exp(-M\alpha/m(T)) + [EW^2]K \frac{m(T)^2}{\lambda_2^{2d/\beta} M} \left(\frac{b}{a}\right)^{4d/\beta} + \varepsilon/2.$$

From the selection of  $\alpha$ ,  $M$  and  $\theta$  it follows easily that the first two terms on the right-hand side of the previous display can be made less than  $\varepsilon/2$  for all  $\varepsilon \leq \varepsilon_0$  by taking  $\varepsilon_0$  sufficiently small.

The complexity count given in the theorem now corresponds to the following estimates. The factor  $O((M)^3)$  represents the cost of a Cholesky factorization required to generate a single replication of a finite field of dimension  $M$ . In addition, the second part of Proposition 6.2 gives us that  $O(M^2\varepsilon^{-2}\delta^{-1})$  replications are required to control the relative variance of the estimator.  $\square$

We now proceed to prove Propositions 6.4 and 6.5.

6.4. *Proof of Proposition 6.4.* We concentrate on the analysis of the left-hand side of (6.11). An important observation is that conditional on the random variable  $\tau_{\gamma_{a,b}}$  with distribution

$$Q'(\tau_{\gamma_{a,b}} \in \cdot) = \frac{E[m(A_{\gamma_{a,b}} \cap \cdot)]}{E[m(A_{\gamma_{a,b}})]}$$

and, given  $f(\tau_{\gamma_{a,b}})$ , the rest of the field, namely  $(f(t) : t \in T \setminus \{\tau_{\gamma_{a,b}}\})$  is another Gaussian field with a computable mean and covariance structure. The second term in (6.10) indicates that we must estimate the probability that  $m(A_{\gamma_{a,b}})$  takes small values under  $Q'$ . For this purpose, we shall develop an upper bound for

$$(6.14) \quad P(m(A_{\gamma_{a,b}}) < y^{-1}, m(A_b) > 0 | f(t) = \gamma_{a,b} + z/\gamma_{a,b})$$

for  $y$  large enough. Our arguments proceeds in two steps. For the first, in order to study (6.14), we shall estimate the conditional mean covariance of  $\{f(s) : s \in T\}$ , given that  $f(t) = \gamma_{a,b} + z/\gamma_{a,b}$ . Then, we use the fact that the conditional field is also Gaussian and take advantage of general results from the theory of Gaussian random fields to obtain a bound for (6.14). For this purpose we recall some useful results from the theory of Gaussian random fields. The first result is due to Dudley [16].

**THEOREM 6.7.** *Let  $\mathcal{U}$  be a compact subset of  $\mathbb{R}^n$ , and let  $\{f_0(t) : t \in \mathcal{U}\}$  be a mean zero, continuous Gaussian random field. Define the canonical metric  $d$  on  $\mathcal{U}$  as*

$$d(s, t) = \sqrt{E[f_0(t) - f_0(s)]^2}$$

and put  $\text{diam}(\mathcal{U}) = \sup_{s,t \in \mathcal{U}} d(s, t)$ , which is assumed to be finite. Then there exists a finite universal constant  $\kappa > 0$  such that

$$E \left[ \max_{t \in \mathcal{U}} f_0(t) \right] \leq \kappa \int_0^{\text{diam}(\mathcal{U})/2} [\log(\mathcal{N}(\varepsilon))]^{1/2} d\varepsilon,$$

where the entropy  $\mathcal{N}(\varepsilon)$  is the smallest number of  $d$ -balls of radius  $\varepsilon$  whose union covers  $\mathcal{U}$ .

The second general result that we shall need is the so-called B–TIS (Borel–Tsirelson–Ibragimov–Sudakov) inequality [5, 12, 14].

**THEOREM 6.8.** *Under the setting described in Theorem 6.7,*

$$P \left( \max_{t \in \mathcal{U}} f_0(t) - E \left[ \max_{t \in \mathcal{U}} f_0(t) \right] \geq b \right) \leq \exp(-b^2 / (2\sigma_{\mathcal{U}}^2)),$$

where

$$\sigma_{\mathcal{U}}^2 = \max_{t \in \mathcal{U}} E[f_0^2(t)].$$

We can now proceed with the main proof. We shall assume from now on that  $\tau_{\gamma_{a,b}} = 0$ , since, as will be obvious from what follows, all estimates hold uniformly over  $\tau_{\gamma_{a,b}} \in T$ . This is a consequence of the uniform Hölder assumptions (A2) and (A3). Define a new process  $\tilde{f}$

$$(\tilde{f}(t) : t \in T) \stackrel{\mathcal{L}}{=} (f(t) : t \in T | f(0) = \gamma_{a,b} + z/\gamma_{a,b}).$$

Note that we can always write  $\tilde{f}(t) = \tilde{\mu}(t) + g(t)$ , where  $g$  is a mean zero Gaussian random field on  $T$ . We have that

$$\tilde{\mu}(t) = E \tilde{f}(t) = \mu(t) + \sigma(0)^{-2} C(0, t) (\gamma_{a,b} + z/\gamma_{a,b} - \mu(0)),$$

and that the covariance function of  $\tilde{f}$  is given by

$$C_g(s, t) = \text{Cov}(g(s), g(t)) = C(s, t) - \sigma(0)^{-2} C(0, s) C(0, t).$$

The following lemma describes the behavior of  $\tilde{\mu}(t)$  and  $C_g(t, s)$ .

**LEMMA 6.9.** *Assume that  $|s|$  and  $|t|$  small enough. Then the following three conclusions hold:*

(i) *There exist constants  $\lambda_0$  and  $\lambda_1 > 0$  such that*

$$|\tilde{\mu}(t) - (\gamma_{a,b} + z/\gamma_{a,b})| \leq \lambda_0 |t|^\beta + \lambda_1 |t|^\beta (\gamma_{a,b} + z/\gamma_{a,b}),$$

and for all  $z \in (0, 1)$  and  $\gamma_{a,b}$  large enough,

$$|\tilde{\mu}(s) - \tilde{\mu}(t)| \leq \kappa_H \gamma_{a,b} |s - t|^\beta.$$

(ii) 
$$C_g(s, t) \leq 2\kappa_H \sigma(t) \sigma(s) \{|t|^\beta + |s|^\beta + |t - s|^\beta\}.$$

(iii) 
$$D_g(s, t) = \sqrt{E([g(t) - g(s)]^2)} \leq \lambda_1^{1/2} |t - s|^{\beta/2}.$$

PROOF. All three consequences follow from simple algebraic manipulations. The details are omitted.  $\square$

PROPOSITION 6.10. For any  $v > 0$ , there exist  $\kappa$  and  $\lambda_2$ , such that for all  $t \in T$ ,  $y^{-\beta/d} \leq \frac{a^{2+v}}{\kappa b^{(2+v)}}$ ,  $a$  sufficiently small, and  $z > 0$ ,

$$P(m(A_{\gamma_{a,b}})^{-1} > y, m(A_b) > 0 | f(t) = \gamma_{a,b} + z/\gamma_{a,b}) \leq \exp(-\lambda_2 a^2 y^{\beta/d} / b^2).$$

PROOF. For notational simplicity, and without loss of generality, we assume that  $t = 0$ . First consider the case that  $z \geq 1$ . Then there exist  $c_1, c_2$  such that for all  $c_2 y^{-\beta/d} < b^{-2-v}$  and  $z > 1$ ,

$$\begin{aligned} &P(m(A_{\gamma_{a,b}} \cap T)^{-1} > y, m(A_b) > 0 | f(0) = \gamma_{a,b} + z/\gamma_{a,b}) \\ &\leq P\left(\inf_{|t| < c_1 y^{-1/d}} f(t) \leq \gamma_{a,b} | f(0) = \gamma_{a,b} + z/\gamma_{a,b}\right) \\ &= P\left(\inf_{|t| < c_1 y^{-1/d}} \tilde{\mu}(t) + g(t) \leq \gamma_{a,b}\right) \\ &\leq P\left(\inf_{|t| < c_1 y^{-1/d}} g(t) \leq -\frac{1}{2\gamma_{a,b}}\right). \end{aligned}$$

Now apply (iii) from Lemma 6.9, from which it follows that  $\mathcal{N}(\varepsilon) \leq c_3 m(T) / \varepsilon^{2d/\beta}$  for some constant  $c_3$ . By Theorem 6.7,  $E(\sup_{|t| < c_1 y^{-1/d}} f(t)) = O(y^{-\beta/(2d)} \log y)$ . By Theorem 6.8, for some constant  $c_4$ ,

$$\begin{aligned} &P(m(A_{\gamma_{a,b}} \cap T)^{-1} > y, m(A_b) > 0 | f(0) = \gamma_{a,b} + z/\gamma_{a,b}) \\ &\leq P\left(\inf_{|t| < c_1 y^{-1/d}} g(t) \leq -\frac{1}{2\gamma_{a,b}}\right) \\ &\leq \exp\left(-\frac{1}{c_4 \gamma_{a,b}^2 y^{-\beta/d}}\right) \end{aligned}$$

for  $c_2 y^{-\beta/d} < b^{-2-v}$  and  $z > 1$ .

Now consider the case  $z \in (0, 1)$ . Let  $t^*$  be the global maximum of  $f(t)$ . Then,

$$\begin{aligned} &P(m(A_{\gamma_{a,b}} \cap T)^{-1} > y, m(A_b) > 0 | f(0) = \gamma_{a,b} + z/\gamma_{a,b}) \\ &\leq P\left(\inf_{|t-t^*| < c_1 y^{-1/d}} f(t) < \gamma_{a,b}, f(t^*) > b | f(0) = \gamma_{a,b} + z/\gamma_{a,b}\right) \\ &\leq P\left(\sup_{|s-t| < c_1 y^{-1/d}} |f(s) - f(t)| > a/b | f(0) = \gamma_{a,b} + z/\gamma_{a,b}\right). \end{aligned}$$

Consider the new field  $\xi(s, t) = g(s) - g(t)$  with parameter space  $T \times T$ . Note that

$$\sqrt{\text{Var}(\xi(s, t))} = D_g(s, t) \leq \lambda_1 |s - t|^{\beta/2}.$$

Via basic algebra, it is not hard to show that the entropy of  $\xi(s, t)$  is bounded by  $\mathcal{N}_\xi(\varepsilon) \leq c_3 m(T \times T) / \varepsilon^{2d/\beta}$ . In addition, from (i) of Lemma 6.9, we have

$$|\tilde{\mu}(s) - \tilde{\mu}(t)| \leq \kappa_H \gamma_{a,b} |s - t|^\beta.$$

Similarly, for some  $\kappa > 0$  and all  $y^{-\beta/d} \leq \frac{1}{\kappa} (\frac{a}{b})^{2+v}$ ,  $a < 1$ , there exists  $c_5$  such that

$$\begin{aligned} &P(m(A_{\gamma_{a,b}} \cap T)^{-1} > y, m(A_b) > 0 | f(0) = \gamma_{a,b} + z/\gamma_{a,b}) \\ &\leq P\left(\sup_{|s-t| < c_1 y^{-1/d}} |f(s) - f(t)| > a/b | f(0) = \gamma_{a,b} + z/\gamma_{a,b}\right) \\ &= P\left(\sup_{|s-t| < c_1 y^{-1/d}} |\xi(s, t)| > \frac{a}{2b}\right) \\ &\leq \exp\left(-\frac{a^2}{c_5 b^2 y^{-\beta/d}}\right). \end{aligned}$$

Combining the two cases  $z > 1$  and  $z \in (0, 1)$  and choosing  $c_5$  large enough we have

$$\begin{aligned} &P(m(A_{\gamma_{a,b}} \cap T)^{-1} > y, m(A_b) > 0 | f(0) = \gamma_{a,b} + z/\gamma_{a,b}) \\ &\leq \exp\left(-\frac{a^2}{c_5 b^2 y^{-\beta/d}}\right) \end{aligned}$$

for  $a$  small enough and  $y^{-\beta/d} \leq \frac{1}{\kappa} (\frac{a}{b})^{2+v}$ . Renaming the constants completes the proof.  $\square$

The final ingredient needed for the proof of Proposition 6.4 is the following lemma involving stochastic domination. The proof follows an elementary argument and is therefore omitted.

LEMMA 6.11. *Let  $v_1$  and  $v_2$  be finite measures on  $\mathbb{R}$  and define  $\eta_j(x) = \int_x^\infty v_j(ds)$ . Suppose that  $\eta_1(x) \geq \eta_2(x)$  for each  $x \geq x_0$ . Let  $(h(x) : x \geq x_0)$  be a nondecreasing, positive and bounded function. Then,*

$$\int_{x_0}^\infty h(s) v_1(dx) \geq \int_{x_0}^\infty h(s) v_2(ds).$$

PROOF OF PROPOSITION 6.4. Note that

$$\begin{aligned} & E \mathcal{Q}' \left( \frac{\exp(-Mm(A_{\gamma_{a,b}})/m(T))}{m(A_{\gamma_{a,b}})}; m(A_{\gamma_{a,b}}) \in (0, \alpha); m(A_b) > 0 \right) \\ &= \int_T \int_{z=0}^\infty E \left( \frac{\exp(-Mm(A_{\gamma_{a,b}})/m(T))}{m(A_{\gamma_{a,b}})}; \right. \\ &\quad \left. m(A_{\gamma_{a,b}}) \in (0, \alpha); m(A_b) > 0 \middle| f(t) = \gamma_{a,b} + \frac{z}{\gamma_{a,b}} \right) \\ &\quad \times P(\tau_{\gamma_{a,b}} \in dt) P(\gamma_{a,b}[f(t) - \gamma_{a,b}] \in dz | f(t) > \gamma_{a,b}) \\ &\leq \sup_{z>0, t \in T} E \left( \frac{\exp(-Mm(A_{\gamma_{a,b}})/m(T))}{m(A_{\gamma_{a,b}})}; \right. \\ &\quad \left. m(A_{\gamma_{a,b}}) \in (0, \alpha); m(A_b) > 0 \middle| f(t) = \gamma_{a,b} + \frac{z}{\gamma_{a,b}} \right). \end{aligned}$$

Now define  $Y(b/a) = (b/a)^{2d/\beta} \lambda_2^{-d/\beta} W$  with  $\lambda_2$  as chosen in Proposition 6.10 and  $W$  with distribution given by  $P(W > x) = \exp(-x^{\beta/d})$ . By Lemma 6.11,

$$\begin{aligned} & \sup_{t \in T} E \left( \frac{\exp(-Mm(A_{\gamma_{a,b}})/m(T))}{m(A_{\gamma_{a,b}})}; \right. \\ &\quad \left. m(A_{\gamma_{a,b}}) \in (0, \alpha); m(A_b) > 0 \middle| f(t) = \gamma_{a,b} + \frac{z}{\gamma_{a,b}} \right) \\ &\leq E[Y(b/a) \exp(-MY(b/a)^{-1}/m(T)); Y(b/a) > \alpha^{-1}]. \end{aligned}$$

Now let  $Z$  be exponentially distributed with mean 1 and independent of  $Y(b/a)$ . Then we have (using the definition of the tail distribution of  $Z$  and Chebyshev's inequality)

$$\begin{aligned} \exp(-MY(b/a)^{-1}/m(T)) &= P(Z > MY(b/a)^{-1}/m(T) | Y(b/a)) \\ &\leq \frac{m(T)Y(b/a)}{M}. \end{aligned}$$

Therefore,

$$\begin{aligned} & E[\exp(-MY(b/a)^{-1}/m(T))Y(b/a); Y(b/a) \geq \alpha^{-1}] \\ &\leq \frac{m(T)}{M} E[Y(b/a)^2; Y(b/a) \geq \alpha^{-1}] \\ &\leq \frac{m(T)}{M\lambda_2^{2d/\beta}} \left(\frac{b}{a}\right)^{4d/\beta} E(W^2), \end{aligned}$$

which completes the proof.  $\square$

6.5. *Proof of Proposition 6.5.* We start with the following result of Tsirelson [30].

**THEOREM 6.12.** *Let  $f$  be a continuous separable Gaussian process on a compact (in the canonical metric) domain  $T$ . Suppose that  $\text{Var}(f) = \sigma$  is continuous and that  $\sigma(t) > 0$  for  $t \in T$ . Moreover, assume that  $\mu = Ef$  is also continuous and  $\mu(t) \geq 0$  for all  $t \in T$ . Define*

$$\sigma_T^2 \triangleq \max_{t \in T} \text{Var}(f(t))$$

and set  $F(x) = P\{\max_{t \in T} f(t) \leq x\}$ . Then,  $F$  is continuously differentiable on  $\mathbb{R}$ . Furthermore, let  $y$  be such that  $F(y) > 1/2$ , and define  $y_*$  by

$$F(y) = \Phi(y_*).$$

Then, for all  $x > y$ ,

$$F'(x) \leq \Psi\left(\frac{xy_*}{y}\right) \left(\frac{xy_*}{y}(1 + 2\alpha) + 1\right)(1 + \alpha),$$

where

$$\alpha = \frac{y^2}{x(x - y)y_*^2}.$$

We can now prove the following lemma.

**LEMMA 6.13.** *There exists a constant  $A \in (0, \infty)$  independent of  $a$  and  $b \geq 0$  such that*

$$(6.15) \quad P\left(\sup_{t \in T} f(t) \leq b + a/b \mid \sup_{t \in T} f(t) > b\right) \leq aA \frac{P(\sup_{t \in T} f(t) \geq b - 1/b)}{P(\sup_{t \in T} f(t) \geq b)}.$$

**PROOF.** By subtracting  $\inf_{t \in T} \mu(t) > -\infty$  and redefining the level  $b$  to be  $b - \inf_{t \in T} \mu(t)$  we may simply assume that  $Ef(t) \geq 0$  so that we can apply Theorem 6.12. Adopting the notation of Theorem 6.12, first we pick  $b_0$  large enough so that  $F(b_0) > 1/2$  and assume that  $b \geq b_0 + 1$ . Now, let  $y = b - 1/b$  and  $F(y) = \Phi(y_*)$ . Note that there exists  $\delta_0 \in (0, \infty)$  such that  $\delta_0 b \leq y_* \leq \delta_0^{-1} b$  for all  $b \geq b_0$ . This follows easily from the fact that

$$\log \mathbb{P}\left\{\sup_{t \in T} f(t) > x\right\} \sim \log \sup_{t \in T} \mathbb{P}\{f(t) > x\} \sim -\frac{x^2}{2\sigma_T^2}.$$

On the other hand, by Theorem 6.12  $F$  is continuously differentiable, and so

$$(6.16) \quad P\left\{\sup_{t \in T} f(t) < b + a/b \mid \sup_{t \in T} f(t) > b\right\} = \frac{\int_b^{b+a/b} F'(x) dx}{P\{\sup_{t \in T} f(t) > b\}}.$$

Moreover,

$$\begin{aligned} F'(x) &\leq \left(1 - \Phi\left(\frac{xy_*}{y}\right)\right) \left(\frac{xy_*}{y}(1 + 2\alpha(x)) + 1\right) \frac{y_*}{y}(1 + \alpha(x)) \\ &\leq (1 - \Phi(y_*)) \left(\frac{xy_*}{y}(1 + 2\alpha(x)) + 1\right) \frac{y_*}{y}(1 + \alpha(x)) \\ &= P\left(\max_{t \in T} f(t) > b - 1/b\right) \left(\frac{xy_*}{y}(1 + 2\alpha(x)) + 1\right) \frac{y_*}{y}(1 + \alpha(x)). \end{aligned}$$

Therefore,

$$\begin{aligned} &\int_b^{b+a/b} F'(x) dx \\ &\leq P\left(\sup_{t \in T} f(t) > b - 1/b\right) \int_b^{b+a/b} \left(\frac{xy_*}{y}(1 + 2\alpha(x)) + 1\right) \frac{y_*}{y}(1 + \alpha(x)) dx. \end{aligned}$$

Recalling that  $\alpha(x) = y^2/[x(x - y)y_*^2]$ , we can use the fact that  $y_* \geq \delta_0 b$  to conclude that if  $x \in [b, b + a/b]$ , then  $\alpha(x) \leq \delta_0^{-2}$ , and therefore

$$\int_b^{b+a/b} \left(\frac{xy_*}{y}(1 + 2\alpha(x)) + 1\right) \frac{y_*}{y}(1 + \alpha(x)) dx \leq 4\delta_0^{-8} a.$$

We thus obtain that

$$(6.17) \quad \frac{\int_b^{b+a/b} F'(x) dx}{P\{\sup_{t \in T} f(t) > b\}} \leq 4a\delta_0^{-8} \frac{P\{\sup_{t \in T} f(t) > b - 1/b\}}{P\{\sup_{t \in T} f(t) > b\}}$$

for any  $b \geq b_0$ . This inequality, together with the fact that  $F$  is continuously differentiable on  $(-\infty, \infty)$ , yields the proof of the lemma for  $b \geq 0$ .  $\square$

The previous result translates a question that involves the conditional distribution of  $\max_{t \in T} f(t)$  near  $b$  into a question involving the tail distribution of  $\max_{t \in T} f(t)$ . The next result then provides a bound on this tail distribution.

LEMMA 6.14. *For each  $v > 0$  there exists a constant  $C(v) \in (0, \infty)$  (possibly depending on  $v > 0$  but otherwise independent of  $b$ ) so that such that*

$$P\left(\max_{t \in T} f(t) > b\right) \leq C(v)b^{2d/\beta + dv + 1} \max_{t \in T} P(f(t) > b)$$

for all  $b \geq 1$ .

PROOF. The proof of this result follows along the same lines of Theorem 2.6.2 in [6]. Consider an open cover of  $T = \bigcup_{i=1}^M T_i(\theta)$ , where  $T_i(\theta) = \{s : |s - t_i| < \theta\}$ . We choose  $t_i$  carefully such that  $N(\theta) = O(\theta^{-d})$  for  $\theta$  arbitrarily small. Write

$f(t) = g(t) + \mu(t)$ , where  $g(t)$  is a centered Gaussian random field and note, using (A2) and (A3), that

$$P\left(\max_{t \in T_i(\theta)} f(t) > b\right) \leq P\left(\max_{t \in T_i(\theta)} g(t) > b - \mu(t_i) - \kappa_H \theta^\beta\right).$$

Now we wish to apply the Borel–TIS inequality (Theorem 6.8) with  $\mathcal{U} = T_i(\theta)$ ,  $f_0 = g$ ,  $d(s, t) = E^{1/2}([g(t) - g(s)]^2)$ , which, as a consequence of (A2) and (A3), is bounded above by  $C_0|t - s|^{\beta/2}$  for some  $C_0 \in (0, \infty)$ . Thus, applying Theorem 6.7, we have that  $E \max_{t \in T_i(\theta)} g(t) \leq C_1 \theta^{\beta/2} \log(1/\theta)$  for some  $C_1 \in (0, \infty)$ . Consequently, the Borel–TIS inequality yields that there exists  $C_2(v) \in (0, \infty)$  such that for all  $b$  sufficiently large and  $\theta$  sufficiently small we have

$$P\left(\max_{t \in T_i(\theta)} g(t) > b - \mu(t_i) - \kappa_H \theta^\beta\right) \leq C_2(v) \exp\left(-\frac{(b - \mu(t_i) - C_1 \theta^{\beta/(2+\beta v)})^2}{2\sigma_{T_i}^2}\right),$$

where  $\sigma_{T_i} = \max_{t \in T_i(\theta)} \sigma(t)$ . Now select  $v > 0$  small enough, and set  $\theta^{\beta/(2+\beta v)} = b^{-1}$ . Straightforward calculations yield that

$$\begin{aligned} P\left(\max_{t \in T_i(\theta)} f(t) > b\right) &\leq P\left(\max_{t \in T_i(\theta)} g(t) > b - \mu(t_i) - \kappa_H \theta^\beta\right) \\ &\leq C_3(v) \max_{t \in T_i(\theta)} \exp\left(-\frac{(b - \mu(t))^2}{2\sigma(t)^2}\right) \end{aligned}$$

for some  $C_3(v) \in (0, \infty)$ . Now, recall the well-known inequality (valid for  $x > 0$ ) that

$$\phi(x) \left(\frac{1}{x} - \frac{1}{x^3}\right) \leq 1 - \Phi(x) \leq \frac{\phi(x)}{x},$$

where  $\phi = \Phi'$  is the standard Gaussian density. Using this inequality it follows that  $C_4(v) \in (0, \infty)$  can be chosen so that

$$\max_{t \in T_i(\theta)} \exp\left(-\frac{(b - \mu(t))^2}{2\sigma(t)^2}\right) \leq C_4(v) b \max_{t \in T_i} P(f(t) > b)$$

for all  $b \geq 1$ . We then conclude that there exists  $C(v) \in (0, \infty)$  such that

$$\begin{aligned} P\left(\max_{t \in T} f(t) > b\right) &\leq N(\theta) C_4(v) b \max_{t \in T} P(f(t) > b) \\ &\leq C \theta^{-d} b \max_{t \in T} P(f(t) > b) \\ &= C b^{2d/\beta + dv + 1} \max_{t \in T} P(f(t) > b) \end{aligned}$$

giving the result.  $\square$

We can now complete the proof of Proposition 6.5.

PROOF OF PROPOSITION 6.5. The result is a straightforward corollary of the previous two lemmas. By (6.15) in Lemmas 6.13 and 6.14 there exists  $\lambda \in (0, \infty)$  for which

$$\begin{aligned} &P\left(\max_{t \in T} f(t) \leq b + a/b \mid \max_{t \in T} f(t) > b\right) \\ &\leq aA \frac{P(\max_{t \in T} f(t) \geq b - 1/b)}{P(\max_{t \in T} f(t) \geq b)} \\ &\leq aCAb^{2d/\beta + dv + 1} \frac{\max_{t \in T} P(f(t) > b - 1/b)}{P(\max_{t \in T} f(t) \geq b)} \\ &\leq aCAb^{2d/\beta + dv + 1} \frac{\max_{t \in T} P(f(t) > b - 1/b)}{\max_{t \in T} P(f(t) > b)} \\ &\leq a\lambda b^{2d/\beta + dv + 1}. \end{aligned}$$

The last two inequalities follow from the obvious bound

$$P\left(\max_{t \in T} f(t) \geq b\right) \geq \max_{t \in T} P(f(t) > b)$$

and standard properties of the Gaussian distribution. This yields (6.12), from which the remainder of the proposition follows.  $\square$

**7. Fine tuning: Twice differentiable homogeneous fields.** In the preceding section we constructed a polynomial time algorithm based on a randomized discretization scheme. Our goal in this section is to illustrate how to take advantage of additional information to further improve the running time and the efficiency of the algorithm. In order to illustrate our techniques we shall perform a more refined analysis in the setting of smooth and homogeneous fields and shall establish optimality of the algorithm in a precise sense, to described below. Our assumptions throughout this section are (B1) and (B2) of Section 3.

Let  $C(s - t) = \text{Cov}(f(s), f(t))$  be the covariance function of  $f$ , which we assume also has mean zero. Note that it is an immediate consequence of homogeneity and differentiability that  $\partial_i C(0) = \partial_{ijk}^3 C(0) = 0$ .

We shall need the following definition.

DEFINITION 7.1. We call  $\tilde{T} = \{t_1, \dots, t_M\} \subset T$  a  $\theta$ -regular discretization of  $T$  if, and only if,

$$\min_{i \neq j} |t_i - t_j| \geq \theta, \quad \sup_{t \in T} \min_i |t_i - t| \leq 2\theta.$$

Regularity ensures that points in the grid  $\tilde{T}$  are well separated. Intuitively, since  $f$  is smooth, having tight clusters of points translates to a waste of computing resources, as a result of sampling highly correlated values of  $f$ . Also, note that every

region containing a ball of radius  $2\theta$  has at least one representative in  $\tilde{T}$ . Therefore,  $\tilde{T}$  covers the domain  $T$  in an economical way. One technical convenience of  $\theta$ -regularity is that for subsets  $A \subseteq T$  that have positive Lebesgue measure (in particular ellipsoids)

$$\lim_{M \rightarrow \infty} \frac{\#(A \cap \tilde{T})}{M} = \frac{m(A)}{m(T)},$$

where here and throughout the remainder of the section  $\#(A)$  denotes the cardinality of the set  $A$ .

Let  $\tilde{T} = \{t_1, \dots, t_M\}$  be a  $\theta$ -regular discretization of  $T$ , and consider

$$X = (X_1, \dots, X_M)^T \stackrel{\Delta}{=} (f(t_1), \dots, f(t_M))^T.$$

We shall concentrate on estimating  $w_M(b) = P(\max_{1 \leq i \leq M} X_i > b)$ . The next result (which we prove in Section 7.1) shows that if  $\theta = \varepsilon/b$ , then the relative bias is  $O(\sqrt{\varepsilon})$ .

**PROPOSITION 7.2.** *Suppose  $f$  is a Gaussian random field satisfying conditions (B1) and (B2). There exist  $c_0, c_1, b_0$  and  $\varepsilon_0$  such that, for any finite  $\varepsilon/b$ -regular discretization  $\tilde{T}$  of  $T$ ,*

$$(7.1) \quad P\left(\sup_{t \in \tilde{T}} f(t) < b \mid \sup_{t \in T} f(t) > b\right) \leq c_0 \sqrt{\varepsilon} \quad \text{and} \quad \#(\tilde{T}) \leq c_1 m(T) \varepsilon^{-d} b^d$$

for all  $\varepsilon \in (0, \varepsilon_0]$  and  $b > b_0$ .

Note that the bound on the bias obtained for twice differentiable fields is much sharper than that of the general Hölder continuous fields given by (6.13) in Theorem 6.6. This is partly because the conditional distribution of the random field around local maxima is harder to describe in the Hölder continuous than in the case of twice differentiable fields. In addition to the sharper description of the bias, we shall also soon show in Theorem 7.4 that our choice of discretization is optimal in a certain sense. Finally, we point out that the bound of  $\sqrt{\varepsilon}$  in the first term of (7.1) is not optimal. In fact, there seems to be some room of improvement, and we believe that a more careful analysis might yield a bound of the form  $c_0 \varepsilon^2$ .

We shall estimate  $w_M(b)$  by using a slight variation of Algorithm 5.3. In particular, since the  $X_i$ 's are now identically distributed, we redefine  $Q$  to be

$$(7.2) \quad Q(X \in B) = \sum_{i=1}^M \frac{1}{M} P[X \in B \mid X_i > b - 1/b].$$

Our estimator then takes the form

$$(7.3) \quad \tilde{L}_b = \frac{M \times P(X_1 > b - 1/b)}{\sum_{j=1}^M \mathbb{1}(X_j > b - 1/b)} \mathbb{1}\left(\max_{1 \leq i \leq M} X_i > b\right).$$

Clearly, we have that  $E^Q(L_b) = w_M(b)$ . (The reason for subtracting the factor of  $1/b$  was explained in Section 6.1.)

ALGORITHM 7.3. Given a number of replications  $n$  and an  $\varepsilon/b$ -regular discretization  $\tilde{T}$  the algorithm is as follows:

STEP (1). Sample  $X^{(1)}, \dots, X^{(n)}$  i.i.d. copies of  $X$  with distribution  $Q$  given by (7.2).

STEP (2). Compute and output

$$\hat{L}_n = \frac{1}{n} \sum_{i=1}^n \tilde{L}_b^{(i)},$$

where

$$\tilde{L}_b^{(i)} = \frac{M \times P(X_1 > b - 1/b)}{\sum_{j=1}^M \mathbb{1}(X_j^{(i)} > b - 1/b)} \mathbb{1}\left(\max_{1 \leq i \leq M} X_j^{(i)} > b\right).$$

Theorem 7.5 later guides the selection of  $n$  in order to achieve a prescribed relative error. In particular, our analysis, together with considerations from Section 2, implies that choosing  $n = O(\varepsilon^{-2}\delta^{-1})$  suffices to achieve  $\varepsilon$  relative error with probability at least  $1 - \delta$ .

Algorithm 7.3 improves on Algorithm 6.1 for Hölder continuous fields in two important ways. The first aspect is that it is possible to obtain information on the size of the relative bias of the estimator. In Proposition 7.2, we saw that in order to overcome bias due to discretization, it suffices to take a discretization of size  $M = \#\tilde{T} = \Theta(b^d)$ . That this selection is also asymptotically optimal, in the sense described in the next result, will be proven in Section 7.1.

THEOREM 7.4. Suppose  $f$  is a Gaussian random field satisfying conditions (B1) and (B2). If  $\theta \in (0, 1)$ , then, as  $b \rightarrow \infty$ ,

$$\sup_{\#\tilde{T} \leq b^{\theta d}} P\left(\sup_{t \in \tilde{T}} f(t) > b \mid \sup_{t \in T} f(t) > b\right) \rightarrow 0.$$

This result implies that the relative bias goes to 100% as  $b \rightarrow \infty$  if one chooses a discretization scheme of size  $O(b^{\theta d})$  with  $\theta \in (0, 1)$ . Consequently,  $d$  is the smallest power of  $b$  that achieves any given bounded relative bias, and so the suggestion above of choosing  $M = O(b^d)$  points for the discretization is, in this sense, optimal.

The second aspect of improvement involves the variance. In the case of Hölder continuous fields, the ratio of the second moment of the estimator and  $w(b)^2$  was shown to be bounded by a quantity that is of order  $O(M^2)$ . In contrast, in the context of smooth and homogeneous fields considered here, the next result shows that this ratio is bounded uniformly for  $b > b_0$  and  $M = \#\tilde{T} \geq cb^d$ . That is, the variance remains strongly controlled.

**THEOREM 7.5.** *Suppose  $f$  is a Gaussian random field satisfying conditions (B1) and (B2). Then there exist constants  $c$ ,  $b_0$  and  $\varepsilon_0$  such that for any  $\varepsilon/b$ -regular discretization  $\tilde{T}$  of  $T$  we have*

$$\sup_{b > b_0, \varepsilon \in [0, \varepsilon_0]} \frac{E^Q \tilde{L}_b^2}{P^2(\sup_{t \in T} f(t) > b)} \leq \sup_{b > b_0, \varepsilon \in [0, \varepsilon_0]} \frac{E^Q \tilde{L}_b^2}{P^2(\sup_{t \in \tilde{T}} f(t) > b)} \leq c$$

for some  $c \in (0, \infty)$ .

The proof of this result is given in Section 7.2. The fact that the number of replications remains bounded in  $b$  is a consequence of the strong control on the variance.

Finally, we note that the proof of Theorem 3.3 follows as a direct corollary of Theorem 7.5 together with Proposition 7.2 and our discussion in Section 2. Assuming that placing each point in  $\tilde{T}$  takes no more than  $\mathbf{c}$  units of computer time, the total complexity is, according to the discussion in Section 2,  $O(nM^3 + M) = O(\varepsilon^{-2}\delta^{-1}M^3 + M)$ . The contribution of the term  $M^3 = O(\varepsilon^{-6d}b^{3d})$  comes from the complexity of applying Cholesky factorization, and the term  $M = O(\varepsilon^{-2d}b^d)$  corresponds to the complexity of placing  $\tilde{T}$ .

**REMARK 7.6.** Condition (B2) imposes a convexity assumption on the boundary of  $T$ . This assumption, although convenient in the development of the proofs of Theorems 7.4 and 7.5, is not necessary. The results can be generalized, at the expense of increasing the length and the burden in the technical development, to the case in which  $T$  is a  $d$ -dimensional manifold satisfying the so-called Whitney conditions [5].

The remainder of this section is devoted to the proof of Proposition 7.2, Theorems 7.4 and 7.5.

**7.1. Bias control: Proofs of Proposition 7.2 and Theorem 7.4.** We start with some useful lemmas, for all of which we assume that Conditions (B1) and (B2) are satisfied. We shall also assume that the global maximum of  $f$  over  $T$  is achieved, with probability one, at a single point in  $T$ . Additional conditions under which this will happen can be found in [5] and require little more than the nondegeneracy of the joint distribution of  $f$  and its first- and second-order derivatives. Of these lemmas, Lemma 7.7, the proof of which we defer to Section 7.3, is central to much of what follows. However, before we state it we take a moment to describe Palm measures, which may not be familiar to all readers.

**7.1.1. Palm distributions and conditioning.** It is well known that one needs to be careful treating the distributions of stochastic processes at random times. For a simple example, in the current setting, consider the behavior of a smooth

stationary Gaussian process  $f$  on  $\mathbb{R}$  along with its derivative  $f'$ . If  $t \in \mathbb{R}$  is a fixed point,  $u > 0$ , and we are given that  $f(0) = 0$  and  $f(t) = u$ , then the conditional distribution of  $f'(t)$  is still Gaussian, with parameters determined by the trivariate distribution of  $(f(0), f(t), f'(t))$ . However, if we are given that  $f(0) = 0$ , and that  $t > 0$  is the *first* positive time that  $f(t) = u$ , then  $t$  is an upcrossing of the level  $u$  by  $f$ , and so  $f'(t)$  must be positive. Thus it cannot be Gaussian. The difference between the two cases lies in the fact that in the first case  $t$  is deterministic, while in the second it is random.

We shall require something similar, conditioning on the behavior of our (Gaussian) random fields in the neighborhood of local maxima. Since local maxima are random points, given their positions the distribution of the field is no longer stationary nor, once again, even Gaussian. We often shall assume for that a local maximum is at the origin. This, however, amounts to saying that the point-process induced by the set of local maxima is Palm stationary (as opposed to space stationary) and therefore we must then use the associated Palm distribution; the precise conditional distribution of the field given the value of the local maximum at the origin is given in Lemma 7.11. The precise distribution is given in Lemma 7.11.

The theory behind this goes by the name of *horizontal-vertical window conditioning* and the resulting conditional distributions are known as *Palm distributions*. Standard treatments are given, for example, in [1, 6, 18, 19, 21]. To differentiate between regular and Palm conditioning, we shall denote the latter by  $\|_{\mathcal{P}}$ .

We can now set up two important lemmas which tell us about the behavior of  $f$  in the neighborhood of local and global maxima. Proofs are deferred until Section 7.3. First, we provide some notation.

Let  $\mathcal{L}$  be the (random) set of local maxima of  $f$ . That is, for each  $s$  in the interior of  $T$ ,  $s \in \mathcal{L}$  if and only if

$$(7.4) \quad \nabla f(s) = 0 \quad \text{and} \quad \nabla^2 f(s) \in \mathcal{N},$$

where  $\mathcal{N}$  is the set of negative definite matrices, and  $\nabla^2 f(s)$  is the Hessian matrix of  $f$  at  $s$ . For  $s \in \partial T$ , similar constraints apply and are described in the proof of Lemma 7.7. Then we have:

LEMMA 7.7. *Let  $\mathcal{L}$  be the set of local maxima of  $f$ . For any  $a_0 > 0$ , there exists  $c^*$ ,  $\delta^*$ ,  $b_0$  and  $\delta_0$  (which depend on the choice of  $a_0$ ), such that for any  $s \in \mathcal{L}$ ,  $a \in (0, a_0)$ ,  $\delta \in (0, \delta_0)$ ,  $b > b_0$ ,  $z > b + a/b$*

$$(7.5) \quad P\left(\min_{|t-s| < \delta ab^{-1}} f(t) < b \parallel_{\mathcal{P}} f(s) = z\right) \leq c^* \exp\left(-\frac{\delta^*}{\delta^2}\right).$$

LEMMA 7.8. *Let  $t^*$  be the point in  $T$  at which the global maximum of  $f$  is attained. Then, with the same choice of constants as in Lemma 7.7, for any  $a \in (0, a_0)$ ,  $\delta \in (0, \delta_0)$  and  $b > b_0$ ,*

$$P\left(\min_{|t-t^*| < \delta ab^{-1}} f(t) < b \parallel_{\mathcal{P}} f(t^*) > b + a/b\right) \leq 2c^* \exp\left(-\frac{\delta^*}{\delta^2}\right).$$

7.1.2. *Back to the proofs.* The following lemma gives a bound on the density of  $\sup_{t \in T} f(t)$ , which will be used to control the size of overshoot beyond level  $b$ .

LEMMA 7.9. *Let  $p_{f^*}(x)$  be the density function of  $\sup_{t \in T} f(t)$ . Then there exists a constant  $c_{f^*}$  and  $b_0$  such that*

$$p_{f^*}(x) \leq c_{f^*} x^{d+1} P(f(0) > x)$$

for all  $x > b_0$ .

PROOF. Recalling (1.3), let the continuous function  $p^E(x)$ ,  $x \in \mathbb{R}$ , be defined by the relationship

$$E(\chi(\{t \in T : f(t) \geq b\})) = \int_b^\infty p^E(x) dx,$$

where the left-hand side is the expected value of the Euler–Poincaré characteristic of  $A_b$ . Then, according to Theorem 8.10 in [8], there exists  $c$  and  $\delta$  such that

$$|p^E(x) - p_{f^*}(x)| < cP(f(0) > (1 + \delta)x)$$

for all  $x > 0$ . In addition, thanks to the result of [5] which provides  $\int_b^\infty p^E(x) dx$  in closed form, there exists  $c_0$  such that, for all  $x > 1$ ,

$$p^E(x) < c_0 x^{d+1} P(f(0) > x).$$

Hence, there exists  $c_{f^*}$  such that

$$p_{f^*}(x) \leq c_0 x^{d+1} P(f(0) > x) + cP(f(0) > (1 + \delta)x) \leq c_{f^*} x^{d+1} P(f(0) > x)$$

for all  $x > 1$ .  $\square$

The last ingredients required to provide the proof of Proposition 7.2 and Theorem 7.4 are stated in the following result, adapted from Lemma 6.1 and Theorem 7.2 in [24] to the twice differentiable case.

THEOREM 7.10. *There exists a constant  $H$  (depending on the covariance function  $C$ ), such that*

$$(7.6) \quad P\left(\sup_{t \in T} f(t) > b\right) = (1 + o(1))Hm(T)b^d P(f(0) > b)$$

as  $b \rightarrow \infty$ .

Similarly, choose  $\delta$  small enough so that  $[0, \delta]^d \subset T$ , and let  $\Delta_0 = [0, b^{-1}]^d$ . Then there exists a constant  $H_1$  such that

$$(7.7) \quad P\left(\sup_{t \in \Delta_0} f(t) > b\right) = (1 + o(1))H_1 P(f(0) > b)$$

as  $b \rightarrow \infty$ .

We now are ready to provide the proof of Proposition 7.2 and Theorem 7.4.

PROOF OF PROPOSITION 7.2. The fact that there exists  $c_1$  such that

$$\#(\tilde{T}) \leq c_1 m(T) \varepsilon^{-d} b^d$$

is immediate from assumption (B2). Therefore, we proceed to provide a bound for the relative bias. Note first that elementary conditional probability manipulations yield that, for any  $\varepsilon > 0$ ,

$$\begin{aligned} &P\left(\sup_{t \in \tilde{T}} f(t) < b \mid \sup_{t \in T} f(t) > b\right) \\ &\leq P\left(\sup_{t \in T} f(t) < b + 2\sqrt{\varepsilon}/b \mid \sup_{t \in T} f(t) > b\right) \\ &\quad + P\left(\sup_{t \in \tilde{T}} f(t) < b \mid \sup_{t \in T} f(t) > b + 2\sqrt{\varepsilon}/b\right). \end{aligned}$$

By (7.6) and Lemma 7.9, there exists  $c_2$  such that, for large enough  $b$ , the first term above can be bounded by

$$P\left(\sup_{t \in T} f(t) < b + 2\sqrt{\varepsilon}/b \mid \sup_{t \in T} f(t) > b\right) \leq c_2 \sqrt{\varepsilon}.$$

Now take  $\varepsilon < \varepsilon_0 < \delta_0^2$  where  $\delta_0$  is as in Lemmas 7.7 and 7.8. Then, applying (7.5), the second term can be bounded by

$$\begin{aligned} &P\left(\sup_{t \in \tilde{T}} f(t) < b \mid \sup_{t \in T} f(t) > b + 2\sqrt{\varepsilon}/b\right) \\ &\leq P\left(\sup_{|t-t^*| < 2\varepsilon b^{-1}} f(t) < b \mid \sup_{t \in T} f(t) > b + 2\sqrt{\varepsilon}\right) \\ &\leq 2c^* \exp(-\delta^* \varepsilon^{-1}). \end{aligned}$$

Hence, there exists a  $c_0$  such that

$$P\left(\sup_{t \in \tilde{T}} f(t) < b \mid \sup_{t \in T} f(t) > b\right) \leq c_2 \sqrt{\varepsilon} + 2c^* e^{-\delta^*/\varepsilon} \leq c_0 \sqrt{\varepsilon}$$

for all  $\varepsilon \in (0, \varepsilon_0)$ .  $\square$

PROOF OF THEOREM 7.4. We write  $\theta = 1 - 3\delta \in (0, 1)$ . First note that, by (7.6),

$$P\left(\sup_T f(t) > b + b^{2\delta-1} \mid \sup_T f(t) > b\right) \rightarrow 0$$

as  $b \rightarrow \infty$ . Let  $t^*$  be the position of the global maximum of  $f$  in  $T$ . According to the exact Slepian model in Section 7.3 and an argument similar to the proof of

Lemmas 7.7 and 7.8

$$(7.8) \quad P\left(\sup_{|t-t^*|>b^{2\delta-1}} f(t) > b \mid b < f(t^*) \leq b + b^{2\delta-1}\right) \rightarrow 0$$

as  $b \rightarrow \infty$ . Consequently,

$$P\left(\sup_{|t-t^*|>b^{2\delta-1}} f(t) > b \mid \sup_T f(t) > b\right) \rightarrow 0.$$

Let

$$B(\tilde{T}, b^{2\delta-1}) = \bigcup_{t \in \tilde{T}} B(t, b^{2\delta-1}).$$

We have

$$\begin{aligned} &P\left(\sup_{\tilde{T}} f(t) > b \mid \sup_T f(t) > b\right) \\ &\leq P\left(\sup_{|t-t^*|>b^{2\delta-1}} f(t) > b \mid \sup_T f(t) > b\right) \\ &\quad + P\left(\sup_{\tilde{T}} f(t) > b, \sup_{|t-t^*|>b^{2\delta-1}} f(t) \leq b \mid \sup_T f(t) > b\right) \\ &\leq o(1) + P\left(t^* \in B(\tilde{T}, b^{2\delta-1}) \mid \sup_T f(t) > b\right) \\ &\leq o(1) + P\left(\sup_{B(\tilde{T}, b^{2\delta-1})} f(t) > b \mid \sup_T f(t) > b\right). \end{aligned}$$

Since  $\#\tilde{T} \leq b^{(1-3\delta)d}$ , we can find a finite set  $T' = \{t'_1, \dots, t'_l\} \subset T$  and let  $\Delta_k = t'_k + [0, b^{-1}]$  such that  $l = O(b^{(1-\delta)d})$  and  $B(\tilde{T}, b^{2\delta-1}) \subset \bigcup_{k=1}^l \Delta_k$ . The choice of  $l$  only depends on  $\#\tilde{T}$ , not the particular distribution of  $\tilde{T}$ . Therefore, applying (7.7),

$$\sup_{\#\tilde{T} \leq b^{\theta d}} P\left(\sup_{B(\tilde{T}, b^{2\delta-1})} f(t) > b\right) \leq O(b^{(1-\delta)d})P(f(0) > b).$$

This, together with (7.6), yields

$$\sup_{\#\tilde{T} \leq b^{\theta d}} P\left(\sup_{B(\tilde{T}, b^{2\delta-1})} f(t) > b \mid \sup_T f(t) > b\right) \leq O(b^{-\delta d}) = o(1)$$

for  $b \geq b_0$ , which clearly implies the statement of the result.  $\square$

7.2. *Variance control: Proof of Theorem 7.5.* We proceed directly to the proof of Theorem 7.5.

PROOF OF THEOREM 7.5. Note that

$$\begin{aligned}
 & \frac{E^Q \tilde{L}_b^2}{P^2(\sup_{t \in T} f(t) > b)} \\
 &= \frac{E(\tilde{L}_b)}{P^2(\sup_{t \in T} f(t) > b)} \\
 &= \frac{E(M \times P(X_1 > b - 1/b) / \sum_{j=1}^n \mathbb{1}(X_j > b - 1/b); \max_j X_j > b)}{P^2(\sup_{t \in T} f(t) > b)} \\
 &= \left( E \left( \frac{M \times P(X_1 > b - 1/b)}{\sum_{j=1}^n \mathbb{1}(X_j > b - 1/b)}; \max_j X_j > b, \sup_{t \in T} f(t) > b \right) \right) \\
 &\quad \times \left( P^2(\sup_{t \in T} f(t) > b) \right)^{-1} \\
 &= \left( E \left( \frac{M P(X_1 > b - 1/b) \mathbb{1}(\max_j X_j > b)}{\sum_{j=1}^n \mathbb{1}(X_j > b - 1/b)} \middle| \sup_{t \in T} f(t) > b \right) \right) \\
 &\quad \times \left( P(\sup_{t \in T} f(t) > b) \right)^{-1} \\
 &= E \left( \frac{M \mathbb{1}(\max_j X_j > b)}{\sum_{j=1}^n \mathbb{1}(X_j > b - 1/b)} \middle| \sup_{t \in T} f(t) > b \right) \\
 &\quad \times \frac{P(X_1 > b - 1/b)}{P(\sup_{t \in T} f(t) > b)}.
 \end{aligned}$$

The remainder of the proof involves showing that the last conditional expectation here is of order  $O(b^d)$ . This, together with (7.6), will yield the result. Note that for any  $A(b, \varepsilon)$  such that  $A(b, \varepsilon) = \Theta(M)$  uniformly over  $b$  and  $\varepsilon$ , we can write

$$\begin{aligned}
 & E \left( \frac{M \times \mathbb{1}(\max_j X_j > b)}{\sum_{j=1}^M \mathbb{1}(X_j > b - 1/b)} \middle| \sup_{t \in T} f(t) > b \right) \\
 & \leq E \left( \frac{M \times \mathbb{1}(\sum_{j=1}^M \mathbb{1}(X_j > b - 1/b) \geq M/A(b, \varepsilon))}{\sum_{j=1}^M \mathbb{1}(X_j > b - 1/b)} \middle| \sup_{t \in T} f(t) > b \right) \\
 (7.9) \quad & + E \left( \frac{M \times \mathbb{1}(1 \leq \sum_{j=1}^M \mathbb{1}(X_j > b - 1/b) < M/A(b, \varepsilon))}{\sum_{j=1}^M \mathbb{1}(X_j > b - 1/b)} \middle| \right. \\
 & \qquad \qquad \qquad \left. \sup_{t \in T} f(t) > b \right).
 \end{aligned}$$

We shall select  $A(b, \varepsilon)$  appropriately in order to bound the expectations above. By Lemma 7.8, for any  $4\varepsilon \leq \delta \leq \delta_0$ , there exist constants  $c'$  and  $c'' \in (0, \infty)$ , such that

$$\begin{aligned}
 c^* \exp\left(-\frac{\delta^*}{\delta^2}\right) &\geq P\left(\min_{|t-t^*| \leq \delta/b} f(t) < b - 1/b \mid \sup_{t \in T} f(t) > b\right) \\
 (7.10) \qquad &\geq P\left(\sum_{j=1}^M \mathbb{1}(X_j > b - 1/b) \leq c' \delta^d / \varepsilon^d \mid \sup_{t \in T} f(t) > b\right) \\
 &\geq P\left(\frac{M}{\sum_{j=1}^M \mathbb{1}(X_j > b - 1/b)} \geq \frac{b^d c''}{\delta^d} \mid \sup_{t \in T} f(t) > b\right).
 \end{aligned}$$

The first inequality is an application of Lemma 7.8. The second inequality is due to the fact that for any ball  $B$  of radius  $4\varepsilon$  or larger,  $\#\tilde{T} \cap B \geq c^d \varepsilon^{-d}$  for some  $c' > 0$ . Inequality (7.10) implies that for all  $x$  such that  $b^d c'' / \delta_0^d < x < b^d c'' / [4^d \varepsilon^d]$ , there exists  $\delta^{**} > 0$  such that

$$P\left(\frac{M}{\sum_{j=1}^M \mathbb{1}(X_j > b - 1/b) b^d} \geq x \mid \sup_{t \in T} f(t) > b\right) \leq c^* \exp(-\delta^{**} x^{2/d}).$$

Now let  $A(b, \varepsilon) = b^d c'' / (4^d \varepsilon^d)$  and observe that by the second result in (7.1) we have  $A(b, \varepsilon) = \Theta(M)$  and, moreover, that there exists  $c_3$  such that the first term on the right-hand side of (7.9) is bounded by

$$\begin{aligned}
 (7.11) \qquad &E\left(\frac{M \times \mathbb{1}(\sum_{j=1}^M \mathbb{1}(X_j > b - 1/b) \geq M/A(b, \varepsilon))}{\sum_{j=1}^M \mathbb{1}(X_j > b - 1/b)} \mid \sup_T f(t) > b\right) \\
 &\leq c_3 b^d.
 \end{aligned}$$

Now we turn to the second term on the right-hand side of (7.9). We use the fact that  $M/A(b, \varepsilon) \leq c''' \in (0, \infty)$  (uniformly as  $b \rightarrow \infty$  and  $\varepsilon \rightarrow 0$ ). There exist  $c_4$  and  $c_5$  such that, for  $\varepsilon \leq \delta_0/c_4$ ,

$$\begin{aligned}
 (7.12) \qquad &E\left(\frac{M \mathbb{1}(1 \leq \sum_{j=1}^M \mathbb{1}(X_j > b - 1/b) < M/A(b, \varepsilon))}{\sum_{j=1}^M \mathbb{1}(X_j > b - 1/b)} \mid \sup_T f(t) > b\right) \\
 &\leq MP\left(\sum_{j=1}^n \mathbb{1}\left(X_j > b - \frac{1}{b}\right) < c''' \mid \sup_T f(t) > b\right) \\
 &\leq MP\left(\min_{|t-t^*| \leq c_4 \varepsilon/b} f(t) < b - 1/b \mid \sup_T f(t) > b\right) \\
 &\leq c_1 m(T) b^d \varepsilon^{-d} \exp\left(-\frac{\delta^*}{c_4^2 \varepsilon^2}\right) \leq c_5 b^d.
 \end{aligned}$$

The second inequality holds from the fact that if  $\sum_{j=1}^M \mathbb{1}(X_j > b - \frac{1}{b})$  is less than  $c'''$ , then the minimum of  $f$  in a ball around the local maximum and of radius

$c_4\varepsilon/b$  must be less than  $b - 1/b$ . Otherwise, there are more than  $c'''$  elements of  $\tilde{T}$  inside such a ball. The last inequality is due to Lemma 7.8 and Theorem 7.2.

Putting (7.11) and (7.12) together we obtain, for all  $\varepsilon/b$ -regular discretizations with  $\varepsilon < \varepsilon_0 = \min(1/4, 1/c_4)\delta_0$ ,

$$\begin{aligned} \frac{E^Q \tilde{L}_b^2}{P(\sup_{t \in T} f(t) > b)} &\leq E \left( \frac{M \mathbb{1}(\max_j X_j > b)}{\sum_{j=1}^M \mathbb{1}(X_j > b - 1/b)} \Big| \sup_{t \in \tilde{T}} f(t) > b \right) \\ &\quad \times \frac{P(X_1 > b - 1/b)}{P(\sup_{t \in T} f(t) > b)} \\ &\leq (c_3 + c_5) \frac{b^d P(X_1 > b - 1/b)}{P(\sup_{t \in T} f(t) > b)}. \end{aligned}$$

Applying now (7.6) and Proposition 7.2, we have that

$$P\left(\sup_{t \in T} f(t) > b\right) < \frac{P(\sup_{t \in \tilde{T}} f(t) > b)}{1 - c_0\sqrt{\varepsilon}},$$

and we have

$$\sup_{b > b_0, \varepsilon \in [0, \varepsilon_0]} \frac{E^Q \tilde{L}_b^2}{P(\sup_{t \in \tilde{T}} f(t) > b)} < \infty$$

as required.  $\square$

*7.3. Remaining proofs.* We start with the proof of Lemma 7.7. Without loss of generality, we assume that the random field of that result has mean zero and unit variance. However, before getting into the details of the proof of Lemma 7.7, we need a few additional lemmas, for which we adopt the following notation: Let  $C_i$  and  $C_{ij}$  be the first- and second-order derivatives of  $C$ , and define the vectors

$$\begin{aligned} \mu_1(t) &= (-C_1(t), \dots, -C_d(t)), \\ \mu_2(t) &= \text{vech}((C_{ij}(t), i = 1, \dots, d, j = i, \dots, d)). \end{aligned}$$

Let  $f'(0)$  and  $f''(0)$  be the gradient and vector of second-order derivatives of  $f$  at 0, where  $f''(0)$  is arranged in the same order as  $\mu_2(0)$ . Furthermore, let  $\mu_{02} = \mu_{20}^\top$  be a vector of second-order spectral moments and  $\mu_{22}$  a matrix of fourth-order spectral moments. The vectors  $\mu_{02}$  and  $\mu_{22}$  are arranged so that

$$\begin{pmatrix} 1 & 0 & \mu_{02} \\ 0 & \Lambda & 0 \\ \mu_{20} & 0 & \mu_{22} \end{pmatrix}$$

is the covariance matrix of  $(f(0), f'(0), f''(0))$ , where  $\Lambda = (-C_{ij}(0))$ . It then follows that

$$\mu_{2 \cdot 0} = \mu_{22} - \mu_{20}\mu_{02}$$

be the conditional variance of  $f''(0)$  given  $f(0)$ . The following lemma, given in [6], provides a stochastic representation of the  $f$  given that it has a local maxima at level  $u$  at the origin. We emphasize that, as described above, the conditioning here is in the sense of Palm distributions. The resultant conditional, or “model” process (7.13) is generally called a *Slepian process*.

LEMMA 7.11. *Given that  $f$  has a local maximum with height  $u$  at zero (an interior point of  $T$ ), the conditional field is equal in distribution to*

$$(7.13) \quad f_u(t) \triangleq uC(t) - W_u\beta^\top(t) + g(t).$$

$g(t)$  is a centered Gaussian random field with covariance function

$$\gamma(s, t) = C(s - t) - (C(s), \mu_2(s)) \begin{pmatrix} 1 & \mu_{02} \\ \mu_{20} & \mu_{22} \end{pmatrix}^{-1} \begin{pmatrix} C(t) \\ \mu_2^\top(t) \end{pmatrix} - \mu_1(s)\Lambda^{-1}\mu_1^\top(t),$$

and  $W_u$  is a  $\frac{d(d+1)}{2}$  random vector independent of  $g(t)$  with density function

$$(7.14) \quad \psi_u(w) \propto |\det(r^*(w) - u\Lambda)| \exp(-\frac{1}{2}w^\top \mu_{2,0}^{-1}w) \mathbb{1}(r^*(w) - u\Lambda \in \mathcal{N}),$$

where  $r^*(w)$  is a  $d \times d$  symmetric matrix whose upper triangular elements consist of the components of  $w$ . The set of negative definite matrices is denoted by  $\mathcal{N}$ . Finally,  $\beta(t)$  is defined by

$$(\alpha(t), \beta(t)) = (C(t), \mu_2(t)) \begin{pmatrix} 1 & \mu_{02} \\ \mu_{20} & \mu_{22} \end{pmatrix}^{-1}.$$

The following two technical lemmas, which we shall prove after completing the proof of Lemma 7.7, provide bounds for the last two terms of (7.13).

LEMMA 7.12. *Using the notation in (7.13), there exist  $\delta_0, \varepsilon_1, c_1$  and  $b_0$  such that, for any  $u > b > b_0$  and  $\delta \in (0, \delta_0)$ ,*

$$P\left(\sup_{|t| \leq \delta a/b} |W_u\beta^\top(t)| > \frac{a}{4b}\right) \leq c_1 \exp\left(-\frac{\varepsilon_1 b^2}{\delta^4}\right).$$

LEMMA 7.13. *There exist  $c, \tilde{\delta}$  and  $\delta_0$  such that, for any  $\delta \in (0, \delta_0)$ ,*

$$P\left(\max_{|t| \leq \delta a/b} |g(t)| > \frac{a}{4b}\right) \leq c \exp\left(-\frac{\tilde{\delta}}{\delta^2}\right).$$

PROOF OF LEMMA 7.7. Using the notation of Lemma 7.11, given any  $s \in \mathcal{L}$  for which  $f(s) = b$ , we have that the corresponding conditional distribution of  $s$  is that of  $f_b(\cdot - s)$ . Consequently, it suffices to show that

$$P\left(\min_{|t| \leq \delta a/b} f_b(t) < b - a/b\right) \leq c^* \exp\left(-\frac{\delta^*}{\delta^2}\right).$$

We consider first the case for which the local maximum is in the interior of  $T$ . Then, by the Slepian model (7.13),

$$f_b(t) = bC(t) - W_b\beta^\top(t) + g(t).$$

We study the three terms of the Slepian model individually. Since

$$C(t) = 1 - t^\top \Lambda t + o(|t|^2),$$

there exists a  $\varepsilon_0$  such that

$$bC(t) \geq b - \frac{a}{4b}$$

for all  $|t| < \varepsilon_0\sqrt{a}/b$ . According to Lemmas 7.12 and 7.13, for  $\delta < \min(\varepsilon_0/\sqrt{a_0}, \delta_0)$ ,

$$\begin{aligned} &P\left(\min_{|t| \leq \delta a/b} f_u(t) < b - a/b\right) \\ &\leq P\left(\max_{|t| < \delta a/b} |g(t)| > \frac{a}{4b}\right) + P\left(\sup_{|t| \leq \delta a/b} |W_b\beta^\top(t)| > \frac{a}{4b}\right) \\ &\leq c \exp\left(-\frac{\tilde{\delta}}{\delta^2}\right) + c_1 \exp\left(-\frac{\varepsilon_1 b^2}{\delta^4}\right) \\ &\leq c^* \exp\left(-\frac{\delta^*}{\delta^2}\right) \end{aligned}$$

for some  $c^*$  and  $\delta^*$ .

Now consider the case for which the local maximum is in the  $(d - 1)$ -dimensional boundary of  $T$ . Due to convexity of  $T$  we can assume, without loss of generality, that the tangent space of  $\partial T$  is generated by  $\partial/\partial t_2, \dots, \partial/\partial t_d$ , the local maximum is located at the origin and  $T$  is a subset of the positive half-plane  $t_1 \geq 0$ . That these arguments do not involve a loss of generality follows from the arguments on pages 192–291 of [5], which rely on the assumed stationarity of  $f$  (for translations) and the fact that rotations, while changing the distributions, will not affect the probabilities that we are currently computing.

For the origin, positioned as just described, to be a local maximum it is necessary and sufficient that the gradient of  $f$  restricted to  $\partial T$  is the zero vector, the Hessian matrix restricted to  $\partial T$  is negative definite and  $\partial_1 f(0) \leq 0$ . Applying a version of Lemma 7.11 for this case, conditional on  $f(0) = u$  and 0 being a local maximum, the field is equal in distribution to

$$(7.15) \quad uC(t) - \tilde{W}_u\beta^\top(t) + \mu_1(t)\Lambda^{-1}(Z, 0, \dots, 0)^T + g(t),$$

where  $Z \leq 0$  corresponds to  $\partial_1 f(0)$  and it follows a truncated (conditional on the negative axis) Gaussian random variable with mean zero and a variance parameter which is computed as the conditional variance of  $\partial_1 f(0)$  given

$(\partial_2 f(0), \dots, \partial_d f(0))$ . The vector  $\tilde{W}_u$  is a  $(d(d + 1)/2)$ -dimensional random vector with density function

$$\tilde{\psi}_u(w) \propto |\det(\tilde{r}^*(w) - u\bar{\Lambda})| \exp(-\frac{1}{2}w^\top \mu_{2,0}^{-1}w) \mathbb{1}(\tilde{w}^* - u\bar{\Lambda} \in \mathcal{N}),$$

where  $\bar{\Lambda}$  is the second spectral moment of  $f$  restricted to  $\partial T$ , and  $\tilde{r}^*(w)$  is the  $(d - 1) \times (d - 1)$  symmetric matrix whose upper triangular elements consist of the components of  $w$ . In the representation (7.15) the vectors  $\tilde{W}_u$  and  $Z$  are independent. As in the proof of Lemma 7.12, one can show that a similar bound holds, albeit with different constants. Thus, since  $\mu_1(t) = O(t)$ , there exist  $c''$  and  $\delta''$  such that the third term in (7.15) can be bounded by

$$P\left[\max_{|t| \leq \delta a/b} |\mu_1(t)\Lambda^{-1}(Z, 0, \dots, 0)^T| \geq a/(4b)\right] \leq c'' \exp\left(-\frac{\delta''}{\delta^2}\right).$$

Consequently, we can also find  $c^*$  and  $\delta^*$  such that the conclusion holds, and we are done.  $\square$

**PROOF OF LEMMA 7.8.** Recall that  $t^*$  is the unique global maximum of  $f$  in  $T$ . Writing Palm probabilities as a ratio of expectations, as explained in Section 7.1.1, and using the fact that  $t^* \in \mathcal{L}$ , we immediately have

$$(7.16) \quad \begin{aligned} &P\left(\min_{|t-t^*| < \delta ab^{-1}} f(t) < b \parallel_{\mathcal{P}} f(t^*) > (b + a/b)\right) \\ &\leq \frac{E(\#\{s \in \mathcal{L} : \min_{|t-s| < \delta ab^{-1}} f(t) < b, f(s) > b + a/b\})}{E(\#\{s \in \mathcal{L} : f(s) > (b + a/b), s = t^*\})}. \end{aligned}$$

Writing

$$N_b = \#\{s \in \mathcal{L} : f(s) > (b + a/b)\},$$

it is standard fare that, for the random fields of the kind we are treating,

$$E(N_b) = (1 + o(1))P(N_b = 1)$$

for large  $b$ ; for example, Chapter 6 of [1] or Chapter 5 of [6].

Therefore, for  $b$  large enough,

$$\frac{E(\#\{s \in \mathcal{L} : f(s) > (b + a/b)\})}{E(\#\{s \in \mathcal{L} : f(s) > (b + a/b), s = t^*\})} < 2.$$

Substituting this into (7.16) yields, for any  $s \in \mathcal{L}$ ,

$$\begin{aligned} &P\left(\min_{|t-t^*| < \delta ab^{-1}} f(t) < b \parallel_{\mathcal{P}} f(t^*) > b + a/b\right) \\ &\leq 2P\left(\min_{|t-s| < \delta ab^{-1}} f(t) < b \parallel_{\mathcal{P}} f(s) > b + a/b\right) \\ &\leq 2c^* \exp(-\delta^*/\delta^2), \end{aligned}$$

where the second inequality follows from (7.5), and we are done.  $\square$

We complete the paper with the proofs of Lemmas 7.12 and 7.13.

**PROOF OF LEMMA 7.12.** It suffices to prove the lemma for the case  $a = 1$ . Since

$$f_u(0) = u = u - W_u \beta^\top(0) + g(0),$$

and  $W_u$  and  $g$  are independent,  $\beta(0) = 0$ . Furthermore, since  $C'(t) = O(t)$  and  $\mu'_2(t) = O(t)$ , there exists a  $c_0$  such that  $|\beta(t)| \leq c_0|t|^2$ . In addition,  $W_u$  has density function proportional to

$$\psi_u(w) \propto |\det(r^*(w) - u\Lambda)| \exp(-\frac{1}{2}w^\top \mu_{2,0}^{-1}w) \mathbb{1}(w^* - u\Lambda \in \mathcal{N}).$$

Note that  $\det(r^*(w) - u\Lambda)$  is expressible as a polynomial in  $w$  and  $u$ , and there exists some  $\varepsilon_0$  and  $c$  such that

$$\left| \frac{\det(r^*(w) - u\Lambda)}{\det(-u\Lambda)} \right| \leq c,$$

if  $|w| \leq \varepsilon_0 u$ . Hence, there exist  $\varepsilon_2, c_2 > 0$ , such that

$$\psi_u(w) \leq \tilde{\psi}(w) := c_2 \exp(-\frac{1}{2}\varepsilon_2 w^\top \mu_{2,0}^{-1}w)$$

for all  $u \geq 1$ . The right-hand side here is proportional to a multivariate Gaussian density. Thus,

$$P(|W_u| > x) = \int_{|w|>x} \psi_u(w) dw \leq \int_{|w|>x} \tilde{\psi}_u(w) dw = c_3 P(|\tilde{W}| > x),$$

where  $\tilde{W}$  is a multivariate Gaussian random variable with density function proportional to  $\tilde{\psi}$ . Therefore, by choosing  $\varepsilon_1$  and  $c_1$  appropriately, we have

$$P\left(\sup_{|t| \leq \delta/b} |W_u \beta^\top| > \frac{1}{4b}\right) \leq P\left(|W_u| > \frac{b}{c_0^2 \delta^2}\right) \leq c_1 \exp\left(-\frac{\varepsilon_1 b^2}{\delta^4}\right)$$

for all  $u \geq b$ .  $\square$

**PROOF OF LEMMA 7.13.** Once again, it suffices to prove the lemma for the case  $a = 1$ . Since

$$f_b(0) = b = b - W_b \beta^\top(0) + g(0),$$

the covariance function  $(\gamma(s, t) : s, t \in T)$  of the centered field  $g$  satisfies  $\gamma(0, 0) = 0$ . It is also easy to check that

$$\partial_s \gamma(s, t) = O(|s| + |t|), \quad \partial_t \gamma(s, t) = O(|s| + |t|).$$

Consequently, there exists a constant  $c_\gamma \in (0, \infty)$  for which

$$\gamma(s, t) \leq c_\gamma(|s|^2 + |t|^2), \quad \gamma(s, s) \leq c_\gamma |s|^2.$$

We need to control the tail probability of  $\sup_{|t| \leq \delta/b} |g(t)|$ . For this it is useful to introduce the following scaling. Define

$$g_\delta(t) = \frac{b}{\delta} g\left(\frac{\delta t}{b}\right).$$

Then  $\sup_{|t| \leq \delta/b} g(t) \geq \frac{1}{4b}$  if and only if  $\sup_{|t| \leq 1} g_\delta(t) \geq \frac{1}{4\delta}$ . Let

$$\sigma_\delta(s, t) = E(g_\delta(s), g_\delta(t)).$$

Then,

$$\sup_{s \in \mathbb{R}} \sigma_\delta(s, s) \leq c\gamma.$$

Because  $\gamma(s, t)$  is at least twice differentiable, applying a Taylor expansion we easily see that the canonical metric  $d_g$  corresponding to  $g_\delta(s)$  (cf. Theorem 6.7) can be bounded as follows:

$$\begin{aligned} d_g^2(s, t) &= E(g_\delta(s) - g_\delta(t))^2 \\ &= \frac{b^2}{\delta^2} \left[ \gamma\left(\frac{\delta s}{b}, \frac{\delta s}{b}\right) + \gamma\left(\frac{\delta t}{b}, \frac{\delta t}{b}\right) - 2\gamma\left(\frac{\delta s}{b}, \frac{\delta t}{b}\right) \right] \\ &\leq c|s - t|^2 \end{aligned}$$

for some constant  $c \in (0, \infty)$ . Therefore, the entropy of  $g_\delta$ , evaluated at  $\tilde{\delta}$ , is bounded by  $K\tilde{\delta}^{-d}$  for any  $\tilde{\delta} > 0$  and with an appropriate choice of  $K > 0$ . Therefore, for all  $\delta < \delta_0$ ,

$$P\left(\sup_{|t| \leq \delta/b} |g(t)| \geq \frac{1}{4b}\right) = P\left(\sup_{|t| \leq 1} g_\delta(t) \geq \frac{1}{4\delta}\right) \leq c_d \delta^{-d-\eta} \exp\left(-\frac{1}{16c_\gamma \delta^2}\right)$$

for some constant  $c_d$  and  $\eta > 0$ . The last inequality is a direct application of Theorem 4.1.1 of [5]. The conclusion of the lemma follows immediately by choosing  $\tilde{c}$  and  $\tilde{\delta}$  appropriately.  $\square$

**8. Numerical examples.** In this section, we provide four examples which indicate how well the techniques we have suggested actually work in practice.

The first treats a random field for which the tail probability is in a closed form. This is simply to confirm that the estimates yielded from the algorithm are reasonable.

EXAMPLE 8.1. Let  $f(t) = X \cos t + Y \sin t$  and  $T = [0, 3/4]$  where  $X$  and  $Y$  are i.i.d. standard Gaussian. We compute  $P(\sup_T f(t) > b)$ . This probability is known in closed form (cf. [5]) and is given by

$$(8.1) \quad P\left(\sup_{0 \leq t \leq 3/4} f(t) > b\right) = 1 - \Phi(b) + \frac{3}{8\pi} e^{-b^2/2}.$$

Table 1 shows the (remarkably accurate) simulation results.

TABLE 1

Simulation results for the cosine process. All results are based on  $10^3$  independent simulations. The "True value" is computed using (8.1) The computation time for each estimate is less than one second. The lattice size is  $3b$

$b$	True value	Est.	Std. er.
3	3.12E-03	3.13E-03	8.43E-05
5	8.8E-07	8.6E-07	2.27E-08
10	3.83E-23	3.81E-23	8.88E-25

The remaining examples treat more interesting random fields for which  $T$  is a two-dimensional square.

EXAMPLE 8.2. Consider the smooth homogenous random field on  $T = [0, 1]^2$  with mean zero and covariance function

$$C(t) = e^{-|t|^2}.$$

Table 2 shows the simulation results of the excursion probabilities  $P(\sup_T f(t) > b)$  and expected overshoots  $E(\sup_T f(t) - b | \sup_T f(t) > b)$ . The results are based on 1,000 independent simulations by setting the tuning parameter  $a = 1$ . The size of discretization and CPU time are also reported.

EXAMPLE 8.3. Consider the continuous, but nondifferentiable, and nonhomogenous random field on  $T = [0, 1]^2$  with

$$\mu(t) = 0.1t_1 + 0.1t_2 C(s, t) = e^{-|t-s|^2}.$$

Table 3 shows the simulation results of excursion probabilities  $P(\sup_T f(t) > b)$  and expected overshoots  $E(\sup_T f(t) - b | \sup_T f(t) > b)$ . The simulation setting is the same as that in Example 8.2.

EXAMPLE 8.4. Consider the smooth random field living on  $T = [0, 1]^2$  with

$$\mu(t) = 0.1t_1 + 0.1t_2 C(t) = e^{-|t|^{1/4}}.$$

TABLE 2

Simulation results in Example 8.2

$b$	$P(\sup_T f(t) > b)$		$E(\sup_T f(t) - b   \sup_T f(t) > b)$		Lattice size	CPU time
	Est.	St. d.	Est.	St. d.		
3	1.1E-02	3.8E-04	0.30	1.5E-02	10 by 10	6 sec
4	3.3E-04	1.2E-05	0.25	1.3E-02	15 by 15	53 sec
5	4.3E-06	1.6E-07	0.19	1.0E-02	15 by 15	45 sec

TABLE 3  
Simulation results for Example 8.3

$b$	$P(\sup_T f(t) > b)$		$E(\sup_T f(t) - b   \sup_T f(t) > b)$		Lattice size	CPU time
	Est.	St. d.	Est.	St. d.		
3	1.4E-02	5.0E-04	0.32	1.6E-02	10 by 10	6 sec
4	5.3E-04	1.9E-05	0.25	1.3E-02	15 by 15	40 sec
5	7.2E-06	2.6E-07	0.20	9.8E-03	15 by 15	56 sec

Table 4 shows simulation results for the excursion probabilities  $P(\sup_T f(t) > b)$  and the expected overshoots  $E(\sup_T f(t) - b | \sup_T f(t) > b)$ . The simulation setting is the same as that in Example 8.2.

Although we have given rigorous results regarding discretization parameters, in practice we choose the lattice size sufficiently large so that the bias was inconsequential in comparison to the estimated standard deviation. We achieved this by increasing the lattice size until the change of the estimate was small enough relative to the estimated standard deviation.

Note that, for all the examples, the relative error does not increase as the level increases and the exceedance probability tends to zero as long as the lattice size also increases. This is in line with the theoretical results of the paper.

Another empirical finding is that the computational burden increases substantially with lattice size, although the algorithm has been proven to be of polynomial complexity. This complexity is mainly from the Cholesky decomposition of large covariance matrices. While this is a problem common to all discrete simulation algorithms for random fields, we nevertheless plan to look at this efficiency issue in future work.

**Acknowledgment.** We are grateful to a referee and an Associate Editor for helpful comments and suggestions.

TABLE 4  
Simulation results for Example 8.4

$b$	$P(\sup_T f(t) > b)$		$E(\sup_T f(t) - b   \sup_T f(t) > b)$		Lattice size	CPU time
	Est.	St. d.	Est.	St. d.		
3	1.5E-02	5.8E-04	0.33	1.5E-02	15 by 15	58 sec
4	6.4E-04	3.1E-05	0.25	1.4E-02	15 by 15	44 sec
5	1.3E-05	6.9E-07	0.21	1.3E-02	25 by 25	600 sec

## REFERENCES

- [1] ADLER, R. J. (1981). *The Geometry of Random Fields*. Wiley, Chichester. MR0611857
- [2] ADLER, R. J., BARTZ, K. and KOU, S. (2010). Estimating curvatures in the Gaussian kinematic formula. Unpublished manuscript, Harvard Univ.
- [3] ADLER, R. J., BLANCHET, J. and LIU, J. (2008). Efficient simulation for tail probabilities of Gaussian random fields. In WSC'08: *Proceedings of the 40th Conference on Winter Simulation* 328–336. Winter Simulation Conference, Austin, TX.
- [4] ADLER, R. J., MÜLLER, P. and ROZOVSKIĬ, B., eds. (1996). *Stochastic Modelling in Physical Oceanography. Progress in Probability* 39. Birkhäuser, Boston, MA. MR1383868
- [5] ADLER, R. J. and TAYLOR, J. E. (2007). *Random Fields and Geometry*. Springer, New York. MR2319516
- [6] ADLER, R. J., TAYLOR, J. E. and WORSLEY, K. (2011). Applications of random fields and geometry: Foundations and case studies. Available at <http://webee.technion.ac.il/people/adler/hrf.pdf>.
- [7] ASMUSSEN, S. and GLYNN, P. W. (2007). *Stochastic Simulation: Algorithms and Analysis. Stochastic Modelling and Applied Probability* 57. Springer, New York. MR2331321
- [8] AZAÏS, J.-M. and WSCHBOR, M. (2009). *Level Sets and Extrema of Random Processes and Fields*. Wiley, Hoboken, NJ. MR2478201
- [9] BARDEEN, J. M., BOND, J. R., KAISER, N. and SZALAY, A. S. (1986). The statistics of peaks of Gaussian random fields. *The Astrophysical Journal* 304 15–61.
- [10] BERMAN, S. M. (1992). *Sojourns and Extremes of Stochastic Processes*. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA. MR1126464
- [11] BLANCHET, J. H. (2009). Efficient importance sampling for binary contingency tables. *Ann. Appl. Probab.* 19 949–982. MR2537195
- [12] BORELL, C. (1975). The Brunn–Minkowski inequality in Gauss space. *Invent. Math.* 30 207–216. MR0399402
- [13] BUCKLEW, J. A. (2004). *Introduction to Rare Event Simulation*. Springer, New York. MR2045385
- [14] CIREL'SON, B. S., IBRAGIMOV, I. A. and SUDAKOV, V. N. (1976). Norms of Gaussian sample functions. In *Proceedings of the Third Japan–USSR Symposium on Probability Theory (Tashkent, 1975). Lecture Notes in Math.* 550 20–41. Springer, Berlin. MR0458556
- [15] DENNIS, M. R. (2007). Nodal densities of planar Gaussian random waves. *Eur. Phys. J.* 145 191–210.
- [16] DUDLEY, R. M. (1973). Sample functions of the Gaussian process. *Ann. Probab.* 1 66–103. MR0346884
- [17] FRISTON, K. J., WORSLEY, K. J., FRACKOWIAK, R. S. J., MAZZIOTTA, J. C. and EVANS, A. C. (1994). Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping* 1 214–220.
- [18] KALLENBERG, O. (1986). *Random Measures*, 4th ed. Akademie Verlag, Berlin. MR0854102
- [19] KALLENBERG, O. (2002). *Foundations of Modern Probability*, 2nd ed. Springer, New York. MR1876169
- [20] LANDAU, H. J. and SHEPP, L. A. (1970). On the supremum of a Gaussian process. *Sankhyā Ser. A* 32 369–378. MR0286167
- [21] LEADBETTER, M. R., LINDGREN, G. and ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer, New York. MR0691492
- [22] MITZENMACHER, M. and UPFAL, E. (2005). *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge Univ. Press, Cambridge. MR2144605
- [23] NIEMIRO, W. and POKAROWSKI, P. (2009). Fixed precision MCMC estimation by median of products of averages. *J. Appl. Probab.* 46 309–329. MR2535815

- [24] PITERBARG, V. (1995). *Asymptotic Methods in the Theory of Gaussian Processes*. Amer. Math. Soc., Providence, RI.
- [25] SHANDARIN, S. F. (2002). Testing non-Gaussianity in cosmic microwave background maps by morphological statistics. *Mon. Not. R. Astr. Soc.* **331** 865–874.
- [26] SHANDARIN, S. F., FELDMAN, H. A., XU, Y. and TEGMARK, M. (2002). Morphological measures of non-Gaussianity in cosmic microwave background maps. *Astrophys. J. Suppl.* **141** 1–11.
- [27] TAYLOR, J., TAKEMURA, A. and ADLER, R. J. (2005). Validity of the expected Euler characteristic heuristic. *Ann. Probab.* **33** 1362–1396. [MR2150192](#)
- [28] TAYLOR, J. E. and WORSLEY, K. J. (2007). Detecting sparse signals in random fields, with an application to brain mapping. *J. Amer. Statist. Assoc.* **102** 913–928. [MR2354405](#)
- [29] TRAUB, J. F., WASILKOWSKI, G. W. and WOŹNIAKOWSKI, H. (1988). *Information-Based Complexity*. Academic Press, Boston, MA. [MR0958691](#)
- [30] TSIREL'SON, V. S. (1975). The density of the maximum of a Gaussian process. *Theory Probab. Appl.* **20** 847–856.
- [31] WOŹNIAKOWSKI, H. (1996). Computational complexity of continuous problems. Technical report, Columbia Univ.

R. J. ADLER  
FACULTY OF ELECTRICAL ENGINEERING  
TECHNION  
HAIFA  
ISRAEL 32000  
E-MAIL: [robert@ee.technion.ac.il](mailto:robert@ee.technion.ac.il)  
URL: [webee.technion.ac.il/people/adler](http://webee.technion.ac.il/people/adler)

J. H. BLANCHET  
DEPARTMENT OF INDUSTRIAL ENGINEERING  
AND OPERATIONS RESEARCH  
COLUMBIA UNIVERSITY  
340 S.W. MUDD BUILDING  
500 W. 120 STREET  
NEW YORK, NEW YORK 10027  
USA  
E-MAIL: [jose.blanchet@columbia.edu](mailto:jose.blanchet@columbia.edu)  
URL: [www.ieor.columbia.edu/fac-bios/blanchet/faculty.html](http://www.ieor.columbia.edu/fac-bios/blanchet/faculty.html)

J. LIU  
DEPARTMENT OF STATISTICS  
COLUMBIA UNIVERSITY  
1255 AMSTERDAM AVENUE ROOM 1030  
NEW YORK, NEW YORK 10027  
USA  
E-MAIL: [jcliu@stat.columbia.edu](mailto:jcliu@stat.columbia.edu)  
URL: [stat.columbia.edu/~jcliu](http://stat.columbia.edu/~jcliu)