

# POSTERIOR CONSISTENCY FOR GAUSSIAN PROCESS APPROXIMATIONS OF BAYESIAN POSTERIOR DISTRIBUTIONS

ANDREW M. STUART AND ARETHA L. TECKENTRUP

**ABSTRACT.** We study the use of Gaussian process emulators to approximate the parameter-to-observation map or the negative log-likelihood in Bayesian inverse problems. We prove error bounds on the Hellinger distance between the true posterior distribution and various approximations based on the Gaussian process emulator. Our analysis includes approximations based on the mean of the predictive process, as well as approximations based on the full Gaussian process emulator. Our results show that the Hellinger distance between the true posterior and its approximations can be bounded by moments of the error in the emulator. Numerical results confirm our theoretical findings.

## 1. INTRODUCTION

Given a mathematical model of a physical process, we are interested in the inverse problem of determining the inputs to the model given some noisy observations related to the model outputs. Adopting a Bayesian approach [20, 41], we incorporate our prior knowledge of the inputs into a probability distribution, referred to as the *prior distribution*, and obtain a more accurate representation of the model inputs in the *posterior distribution*, which results from conditioning the prior distribution on the observations. Since the posterior distribution is generally intractable, sampling methods such as Markov chain Monte Carlo (MCMC) [11, 14, 16, 18, 26, 34] are typically used to explore it. A major challenge in the application of MCMC methods to problems of practical interest is the large computational cost associated with numerically solving the mathematical model for a given set of the input parameters. Since the generation of each sample by the MCMC method requires a solve of the governing equations, and often millions of samples are required, this process can quickly become very costly.

This drawback of fully Bayesian inference for complex models was recognised several decades ago in the statistics literature, and resulted in key papers which had a profound influence on methodology [21, 30, 36]. These papers advocated the use of a Gaussian process surrogate model to approximate the solution of the governing equations, and in particular the likelihood, at a much lower computational cost. This approximation then results in an approximate posterior distribution, which can be sampled more cheaply using MCMC. However, despite the widespread adoption of the methodology, there has been little analysis of the effect of the approximation on posterior inference. In this work, we study this issue, focussing on

---

Received by the editor March 7, 2016 and, in revised form, September 26, 2016.

2010 *Mathematics Subject Classification.* Primary 60G15, 62G08, 65D05, 65D30, 65J22.

*Key words and phrases.* Inverse problem, Bayesian approach, surrogate model, Gaussian process regression, posterior consistency.

the use of Gaussian process emulators [7, 19, 21, 30, 32, 36, 40] as surrogate models. Other choices of surrogate models such as those described in [4, 9], generalised polynomial chaos [23, 45], sparse grid collocation [5, 22] and adaptive subspace methods [12, 13] might also be studied similarly, but are not considered here. Indeed we note that the paper [22] studied the effect, on the posterior distribution, of stochastic collocation approximation within the forward model and was one of the first papers to address such questions. That paper used the Kullback-Leibler divergence, or relative entropy, to measure the effect on the posterior, and considered finite dimensional input parameter spaces.

The main focus of this work is to analyse the error introduced in the posterior distribution by using a Gaussian process emulator as a surrogate model. The error is measured in the Hellinger distance, which is shown in [15, 41] to be a suitable metric for evaluation of perturbations to the posterior measure in Bayesian inverse problems, including problems with infinite dimensional input parameter spaces. We consider emulating either the parameter-to-observation map or the negative log-likelihood. The convergence results presented in this paper are of two types. In section 3, we present convergence results for simple Gaussian process emulators applied to a general function  $f$  satisfying suitable regularity assumptions. In section 4, we prove bounds on the error in the posterior distribution in terms of the error in the Gaussian process emulator. The novel contributions of this work are mainly in section 4. The results in the two sections can be combined to give a final error estimate for the simple Gaussian process emulators presented in section 3. However, the error bounds derived in section 4 are much more general in the sense that they apply to any Gaussian process emulator satisfying the required assumptions. A short discussion on extensions of this work related to Gaussian process emulators used in practice is included in the conclusions in section 6.

We study three different approximations to the posterior distribution. First, we consider using the mean of the Gaussian process emulator as a surrogate model, resulting in a deterministic approximation to the posterior distribution. Our second approximation is obtained by using the full Gaussian process as a surrogate model, leading to a random approximation in which case we study the second moment of the Hellinger distance between the true and the approximate posterior distribution. The uncertainty in the posterior distribution introduced in this way can be thought of representing the uncertainty in the emulator due to the finite number of function evaluations used to construct it. This uncertainty can in applications be large (or comparable) to the uncertainty present in the observations, and a user may want to take this into account to “inflate” the variance of the posterior distribution. Finally, we construct an alternative deterministic approximation by using the full Gaussian process as surrogate model, and taking the expected value (with respect to the distribution of the surrogate) of the likelihood. It can be shown that this approximation of the likelihood is optimal in the sense that it minimises the  $L^2$ -error [39]. In contrast to the approximation based on only the mean of the emulator, this approximation also takes into account the uncertainty of the emulator, although only in an averaged sense.

For the three approximations discussed above, we show that the Hellinger distance between the true and approximate posterior distribution can be bounded by the error between the true parameter-to-observation map (or log-likelihood) and its Gaussian process approximation, measured in a norm that depends on the approximation considered. Our analysis is restricted to finite dimensional input spaces.

This reflects the state-of-the-art with respect to Gaussian process emulation itself; the analysis of the effect on the posterior is less sensitive to dimension. For simplicity, we also restrict our attention to bounded parameters, i.e., parameters in a compact subset of  $\mathbb{R}^K$  for some  $K \in \mathbb{N}$ , and to problems where the parameter-to-observation map is uniformly bounded.

The convergence results on Gaussian process regression presented in section 3 are mainly known results from the theory of scattered data interpolation [28, 37, 43]. The error bounds are given in terms of the fill distance of the design points used to construct the Gaussian process emulator, and depend in several ways on the number  $K$  of input parameters we want to infer. First, when looking at the error in terms of the number of design points used, rather than the fill distance of these points, the rate of convergence typically deteriorates with the number of parameters  $K$ . Second, the proof of these error estimates requires assumptions on the smoothness of the function being emulated, where the precise smoothness requirements depend on the Gaussian process emulator employed. For emulators based on Matérn kernels [24], we require these maps to be in a Sobolev space  $H^s$ , where  $s > K/2$ . We would like to point out here that it is not necessary for the function being emulated to be in the *reproducing kernel Hilbert space* (or *native space*) of the Matérn kernel used in order to prove convergence (cf. Proposition 3.4), but that it suffices to be in a larger Sobolev space in which point evaluations are bounded linear functionals.

The remainder of this paper is organised as follows. In section 2, we set up the Bayesian inverse problem of interest. We then recall some results on Gaussian process regression in section 3. The heart of the paper is section 4, where we introduce the different approximations to the posterior and perform an error analysis. Our theoretical results are confirmed on a simple model problem in section 5, and some conclusions are finally given in section 6.

## 2. BAYESIAN INVERSE PROBLEMS

Let  $X$  and  $V$  be separable Banach spaces, and define the measurable mappings  $G : X \rightarrow V$  and  $\mathcal{O} : V \rightarrow \mathbb{R}^J$ , for some  $J \in \mathbb{N}$ . Denote by  $\mathcal{G} : X \rightarrow \mathbb{R}^J$  the composition of  $\mathcal{O}$  and  $G$ . We refer to  $G$  as the *forward map*, to  $\mathcal{O}$  as the *observation operator* and to  $\mathcal{G}$  as the *parameter-to-observation map*. We denote by  $\|\cdot\|$  the Euclidean norm on  $\mathbb{R}^n$ , for  $n \in \mathbb{N}$ . We consider the setting where the Banach space  $X$  is a compact subset of  $\mathbb{R}^K$ , for some finite  $K \in \mathbb{N}$ , representing the range of a finite number  $K$  of parameters  $u$ . The inverse problem of interest is to determine the parameters  $u \in X$  from the noisy data  $y \in \mathbb{R}^J$  given by

$$y = \mathcal{G}(u) + \eta,$$

where the noise  $\eta$  is a realisation of the  $\mathbb{R}^J$ -valued Gaussian random variable  $\mathcal{N}(0, \sigma_\eta^2 I)$ , for some known variance  $\sigma_\eta^2$ . We adopt a Bayesian perspective in which, in the absence of data,  $u$  is distributed according to a prior measure  $\mu_0$ . We are interested in the posterior distribution  $\mu^y$  on the conditioned random variable  $u|y$ , which can be characterised as follows.

**Proposition 2.1** ([41]). *Suppose  $\mathcal{G} : X \rightarrow \mathbb{R}^J$  is continuous and  $\mu_0(X) = 1$ . Then the posterior distribution  $\mu^y$  on the conditioned random variable  $u|y$  is absolutely continuous with respect to  $\mu_0$  and given by Bayes' Theorem:*

$$\frac{d\mu^y}{d\mu_0}(u) = \frac{1}{Z} \exp(-\Phi(u)),$$

where

$$(2.1) \quad \Phi(u) = \frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}(u)\|^2 \quad \text{and} \quad Z = \mathbb{E}_{\mu_0} \left( \exp(-\Phi(u)) \right).$$

We make the following assumption on the regularity of the parameter-to-observation map  $\mathcal{G}$ .

**Assumption 2.2.** *We assume that  $\mathcal{G} : X \rightarrow \mathbb{R}^J$  satisfies  $\mathcal{G} \in H^s(X; \mathbb{R}^J)$ , for some  $s > K/2$ , and that  $\sup_{u \in X} \|\mathcal{G}(u)\| =: C_{\mathcal{G}} < \infty$ .*

Under Assumption 2.2, it follows that the negative log-likelihood  $\Phi : X \rightarrow \mathbb{R}$  satisfies  $\Phi \in H^s(X)$ , and  $\sup_{u \in X} |\Phi(u)| =: C_\Phi < \infty$ . Since  $s > K/2$ , the Sobolev Embedding Theorem furthermore implies that  $\mathcal{G}$  and  $\Phi$  are continuous. Examples of model problems satisfying Assumption 2.2 include linear elliptic and parabolic partial differential equations [10, 38] and non-linear ordinary differential equations [17, 42]. A specific example is given in section 5.

Note that in Assumption 2.2, the smoothness requirement on  $\mathcal{G}$  becomes stronger as  $K$  increases. The reason for this is that in order to apply the results in section 3, we require  $\mathcal{G}$  to be in a Sobolev space in which point evaluations are bounded linear functionals. The second part of Assumption 2.2 is mainly included to define the constant  $C_{\mathcal{G}}$ , since the fact that  $\sup_{u \in X} \|\mathcal{G}(u)\|$  is finite follows from the continuity of  $\mathcal{G}$  and the compactness of  $X$ .

### 3. GAUSSIAN PROCESS REGRESSION

We are interested in using Gaussian process regression to build a surrogate model for the forward map, leading to an approximate Bayesian posterior distribution that is computationally cheaper to evaluate. Generally speaking, Gaussian process regression (or Gaussian process emulation, or kriging) is a way of building an approximation to a function  $f$ , based on a finite number of evaluations of  $f$  at a chosen set of design points. We will here consider emulation of either the parameter-to-observation map  $\mathcal{G} : X \rightarrow \mathbb{R}^J$  or the negative log-likelihood  $\Phi : X \rightarrow \mathbb{R}$ . Since the efficient emulation of vector-valued functions is still an open question [6], we will focus on the emulation of scalar-valued functions. An emulator of  $\mathcal{G}$  in the case  $J > 1$  is constructed by emulating each entry independently.

Now let  $f : X \rightarrow \mathbb{R}$  be an arbitrary function. Gaussian process emulation is in fact a Bayesian procedure, and the starting point is to put a Gaussian process prior on the function  $f$ . In other words, we model  $f$  as

$$(3.1) \quad f_0 \sim \text{GP}(m(u), k(u, u')),$$

with known mean  $m : X \rightarrow \mathbb{R}$  and two point covariance function  $k : X \times X \rightarrow \mathbb{R}$ , assumed to be positive-definite. Here, we use the Gaussian process notation as in, for example, [32]. In the notation of [41], we have  $f_0 \sim \mathcal{N}(m, C)$ , where  $m = m(\cdot)$  and  $C$  is the integral operator with covariance function  $k$  as kernel.

Typical choices of the mean function  $m$  include the zero function and polynomials [32]. A family of covariance functions  $k$  frequently used in applications are the Matérn covariance functions [24], given by

$$(3.2) \quad k_{\nu, \lambda, \sigma_k^2}(u, u') = \sigma_k^2 \frac{1}{\Gamma(\nu) 2^{\nu-1}} \left( \frac{\|u - u'\|}{\lambda} \right)^\nu B_\nu \left( \frac{\|u - u'\|}{\lambda} \right),$$

where  $\Gamma$  denotes the Gamma function,  $B_\nu$  denotes the modified Bessel function of the second kind and  $\nu, \lambda$  and  $\sigma_k^2$  are positive parameters. The parameter  $\lambda$  is

referred to as the *correlation length*, and governs the length scale at which  $f_0(u)$  and  $f_0(u')$  are correlated. The parameter  $\sigma_k^2$  is referred to as the *variance*, and governs the magnitude of  $f_0(u)$ . Finally, the parameter  $\nu$  is referred to as the *smoothness parameter*, and governs the regularity of  $f_0$  as a function of  $u$ . As the limit when  $\nu \rightarrow \infty$ , we obtain the Gaussian covariance

$$(3.3) \quad k_{\infty, \lambda, \sigma_k^2}(u, u') = \sigma_k^2 \exp\left(-\frac{\|u - u'\|^2}{\lambda^2}\right).$$

Now suppose we are given data in the form of a set of distinct *design points*  $U := \{u^n\}_{n=1}^N \subseteq X$ , together with corresponding function values

$$(3.4) \quad f(U) := [f(u^1), \dots, f(u^N)] \in \mathbb{R}^N.$$

Since  $f_0$  is a Gaussian process, the vector  $[f_0(u^1), \dots, f_0(u^N), f_0(\tilde{u}^1), \dots, f_0(\tilde{u}^M)] \in \mathbb{R}^{N+M}$ , for any set of test points  $\{\tilde{u}^m\}_{m=1}^M \subseteq X \setminus U$ , follows a multivariate Gaussian distribution. The conditional distribution of  $f_0(\tilde{u}^1), \dots, f_0(\tilde{u}^M)$ , given the values  $f_0(u^1) = f(u^1), \dots, f_0(u^N) = f(u^N)$ , is then again Gaussian, with mean and covariance given by the standard formulas for the conditioning of Gaussian random variables [32].

Conditioning the Gaussian process (3.1) on the known values  $f(U)$ , we hence obtain another Gaussian process  $f_N$ , known as the *predictive process*. We have

$$(3.5) \quad f_N \sim \text{GP}(m_N^f(u), k_N(u, u')),$$

where the predictive mean  $m_N^f : X \rightarrow \mathbb{R}$  and predictive covariance  $k_N : X \times X \rightarrow \mathbb{R}$  are known explicitly, and depend on the modelling choices made in (3.1). In the following discussion, we will focus on the popular choice  $m \equiv 0$ ; the case of a non-zero mean is discussed in Remark 3.7. When  $m \equiv 0$ , we have

$$(3.6) \quad \begin{aligned} m_N^f(u) &= k(u, U)^T K(U, U)^{-1} f(U), \\ k_N(u, u') &= k(u, u') - k(u, U)^T K(U, U)^{-1} k(u', U), \end{aligned}$$

where  $k(u, U) = [k(u, u^1), \dots, k(u, u^N)] \in \mathbb{R}^N$  and  $K(U, U) \in \mathbb{R}^{N \times N}$  is the matrix with  $ij$ th entry equal to  $k(u^i, u^j)$  [32].

There are several points to note about the predictive mean  $m_N^f$  in (3.6). First,  $m_N^f$  is a linear combination of the function evaluations  $f(U)$ , and hence a linear predictor. It is in fact the *best linear predictor* [40], in the sense that it is the linear predictor with the smallest mean square error. Second,  $m_N^f$  interpolates the function  $f$  at the design points  $U$ , since the vector  $k(u^n, U)$  is the  $n$ th row of the matrix  $K(U, U)$ . In other words, we have  $m_N^f(u^n) = f(u^n)$ , for all  $n = 1, \dots, N$ . Finally, we remark that  $m_N^f$  is a linear combination of kernel evaluations,

$$m_N^f(u) = \sum_{n=1}^N \alpha_n k(u, u^n),$$

where the vector of coefficients is given by  $\alpha = K(U, U)^{-1} f(U)$ . Concerning the predictive covariance  $k_N$ , we note that  $k_N(u, u) < k(u, u)$  for all  $u \in X$ , since  $K(U, U)^{-1}$  is positive definite. Furthermore, we also note that  $k_N(u^n, u^n) = 0$ , for  $n = 1, \dots, N$ , since  $k(u^n, U)^T K(U, U)^{-1} k(u^n, U) = k(u^n, u^n)$ .

For stationary covariance functions  $k(u, u') = k(\|u - u'\|)$ , the predictive mean is a radial basis function interpolant of  $f$ , and we can make use of results from the radial basis function literature to investigate the behaviour of  $m_N^f$  and  $k_N$  as

$N \rightarrow \infty$ . Before we do this, in subsection 3.2, we recall some results on native spaces (also known as reproducing kernel Hilbert spaces) in subsection 3.1.

**3.1. Native spaces of Matérn kernels.** We recall the notion of the reproducing kernel Hilbert space corresponding to the kernel  $k$ , usually referred to as the native space of  $k$  in the radial basis function literature.

**Definition 3.1.** A Hilbert space  $H_k$  of functions  $f : X \rightarrow \mathbb{R}$ , with inner product  $\langle \cdot, \cdot \rangle_{H_k}$ , is called the *reproducing kernel Hilbert space (RKHS)* corresponding to a symmetric, positive definite kernel  $k : X \times X \rightarrow \mathbb{R}$  if

- i) for all  $u \in X$ ,  $k(u, u')$ , as a function of its second argument, belongs to  $H_k$ ,
- ii) for all  $u \in X$  and  $f \in H_k$ ,  $\langle f, k(u, \cdot) \rangle_{H_k} = f(u)$ .

By the Moore-Aronszajn Theorem [3], a unique RKHS exists for each symmetric, positive definite kernel  $k$ . Furthermore, this space can be constructed using Mercer's Theorem [25], and it is equal to the Cameron-Martin space [8] of the covariance operator  $C$  with kernel  $k$ . For covariance kernels of Matérn type, the native space is isomorphic to a Sobolev space [37, 43].

**Proposition 3.2.** Let  $k_{\nu, \lambda, \sigma_k^2}$  be a Matérn covariance kernel as defined in (3.2). Then the native space  $H_{k_{\nu, \lambda, \sigma_k^2}}$  is equal to the Sobolev space  $H^{\nu+K/2}(X)$  as a vector space, and the native space norm and the Sobolev norm are equivalent.

Native spaces for more general kernels, including non-stationary kernels, are analysed in [43]. For stationary kernels, the native space can generally be characterised by the rate of decay of the Fourier transform of the kernel. The native space of the Gaussian kernel (3.3), for example, consists of functions whose Fourier transform decays exponentially, and is hence strictly contained in the space of analytic functions. Proposition 3.2 shows that as a vector space, the native space of the Matérn kernel  $k_{\nu, \lambda, \sigma_k^2}$  is fully determined by the smoothness parameter  $\nu$ . The parameters  $\lambda$  and  $\sigma_k^2$  do, however, influence the constants in the norm equivalence of the native space norm and the standard Sobolev norm.

**3.2. Radial basis function interpolation.** For stationary covariance functions  $k(u, u') = k(\|u - u'\|)$ , the predictive mean is a radial basis functions interpolant of  $f$ . In fact, it is the minimum norm interpolant [32],

$$(3.7) \quad m_N^f = \arg \min_{g \in H_k : g(U) = f(U)} \|g\|_{H_k}.$$

Given the set of design points  $U = \{u^n\}_{n=1}^N \subseteq X$ , we define the fill distance  $h_U$ , separation radius  $q_U$  and mesh ratio  $\rho_U$  by

$$h_U := \sup_{u \in X} \inf_{u^n \in U} \|u - u^n\|, \quad q_U := \frac{1}{2} \min_{i \neq j} \|u^j - u^i\|, \quad \rho_U := \frac{h_U}{q_U} \geq 1.$$

The fill distance is the maximum distance any point in  $X$  can be from  $U$ , and the separation radius is half the smallest distance between any two distinct points in  $U$ . The mesh ratio provides a measure of how uniformly the design points  $U$  are distributed in  $X$ . We have the following theorem on the convergence of  $m_N^f$  to  $f$  [27, 28, 43].

**Proposition 3.3.** *Suppose  $X \subseteq \mathbb{R}^K$  is a bounded, Lipschitz domain that satisfies an interior cone condition, and the symmetric positive definite kernel  $k$  is such that  $H_k$  is isomorphic to the Sobolev space  $H^\tau(X)$ , with  $\tau = n + r$ ,  $n \in \mathbb{N}$ ,  $n > K/2$  and  $0 \leq r < 1$ . Suppose  $m_N^f$  is given by (3.6). If  $f \in H^\tau(X)$ , then there exists a constant  $C$ , independent of  $f$ ,  $U$ , and  $N$ , such that*

$$\|f - m_N^f\|_{H^\beta(X)} \leq Ch_U^{\tau-\beta} \|f\|_{H^\tau(X)}, \quad \text{for any } \beta \leq \tau,$$

for all sets  $U$  with  $h_U$  sufficiently small.

Proposition 3.3 assumes that the function  $f$  is in the RKHS of the kernel  $k$ . Convergence estimates for a wider class of functions can be obtained using interpolation in Sobolev spaces [28].

**Proposition 3.4.** *Suppose  $X \subseteq \mathbb{R}^K$  is a bounded, Lipschitz domain that satisfies an interior cone condition, and the symmetric positive definite kernel  $k$  is such that  $H_k$  is isomorphic to the Sobolev space  $H^\tau(X)$ . Suppose  $m_N^f$  is given by (3.6). If  $f \in H^{\tilde{\tau}}(X)$ , for some  $\tilde{\tau} \leq \tau$ ,  $\tilde{\tau} = n + r$ ,  $n \in \mathbb{N}$ ,  $n > K/2$ , and  $0 \leq r < 1$ , then there exists a constant  $C$ , independent of  $f$ ,  $U$ , and  $N$ , such that*

$$\|f - m_N^f\|_{H^\beta(X)} \leq Ch_U^{\tilde{\tau}-\beta} \rho_U^{\tau-\beta} \|f\|_{H^{\tilde{\tau}}(X)}, \quad \text{for any } \beta \leq \tilde{\tau},$$

for all sets  $U$  with  $h_U$  and  $\rho_U$  sufficiently small.

We would like to point out here that, in practice, it is much more informative to obtain convergence rates in terms of the number of design points  $N$  rather than their associated fill distance  $h_U$ . This is of course possible in general, but the precise relation between  $N$  and  $h_U$  will depend on the specific choice of design points  $U$ . For uniform tensor grids  $U$ , the fill distance  $h_U$  is of the order  $N^{-1/K}$  (cf. section 5). This suggests a strong dependence on the input dimension  $K$  of the convergence rate in terms of the number of design points  $N$ .

Convergence of the predictive variance  $k_N(u, u)$  follows under the assumptions of Propositions 3.3 or 3.4 using the relation in Proposition 3.5 below. This was already noted, without proof, in [37]; we give a proof here for completeness.

**Proposition 3.5.** *Suppose  $m_N^f$  and  $k_N$  are given by (3.6). Then*

$$k_N(u, u)^{\frac{1}{2}} = \sup_{\|g\|_{H_k}=1} |g(u) - m_N^g(u)|.$$

*Proof.* For any  $u \in X$ , we have

$$\begin{aligned} \sup_{\|g\|_{H_k}=1} |g(u) - m_N^g(u)| &= \sup_{\|g\|_{H_k}=1} \left| g(u) - \sum_{j=1}^N (k(u, U)^T K(U, U)^{-1})_j g(u^j) \right| \\ &= \sup_{\|g\|_{H_k}=1} \left| \langle g, k(\cdot, u) \rangle_{H_k} - \sum_{j=1}^N (k(u, U)^T K(U, U)^{-1})_j \langle g, k(\cdot, u^j) \rangle_{H_k} \right| \\ &= \sup_{\|g\|_{H_k}=1} \left| \left\langle g, k(\cdot, u) - \sum_{j=1}^N (k(u, U)^T K(U, U)^{-1})_j k(\cdot, u^j) \right\rangle_{H_k} \right| \\ &= \|k(\cdot, u) - k(\cdot, U)^T K(U, U)^{-1} k(u, U)\|_{H_k}. \end{aligned}$$

The final equality follows from the Cauchy-Schwarz inequality, which becomes an equality when the two functions considered are linearly dependent. By Definition

3.1, we then have

$$\begin{aligned}
& \|k(\cdot, u) - k(\cdot, U)^T K(U, U)^{-1} k(u, U)\|_{H_k}^2 \\
&= \langle k(\cdot, u) - k(\cdot, U)^T K(U, U)^{-1} k(u, U), k(\cdot, u) - k(\cdot, U)^T K(U, U)^{-1} k(u, U) \rangle_{H_k} \\
&= \langle k(\cdot, u), k(\cdot, u) \rangle_{H_k} - 2 \langle k(\cdot, u), k(\cdot, U)^T K(U, U)^{-1} k(u, U) \rangle_{H_k} \\
&\quad + \langle k(\cdot, U)^T K(U, U)^{-1} k(u, U), k(\cdot, U)^T K(U, U)^{-1} k(u, U) \rangle_{H_k} \\
&= k(u, u) - 2k(u, U)^T K(U, U)^{-1} k(u, U) + k(u, U)^T K(U, U)^{-1} k(u, U) \\
&= k_N(u, u).
\end{aligned}$$

The identity which leads to the third term in the penultimate line uses the fact that  $\langle k(\cdot, u'), k(\cdot, u) \rangle_{H_k} = k(u, u')$ , for any  $u, u' \in X$ . If  $\ell(u) = K(U, U)^{-1} k(u, U)$ , then

$$\begin{aligned}
& \langle k(\cdot, U)^T K(U, U)^{-1} k(u, U), k(\cdot, U)^T K(U, U)^{-1} k(u, U) \rangle_{H_k} \\
&= \sum_{j,k} \ell_j(u) \langle k(\cdot, u^j), k(\cdot, u^k) \rangle_{H_k} \ell_k(u) \\
&= \sum_{j,k} \ell_j(u) k(u^j, u^k) \ell_k(u) \\
&= \ell(u)^T K(U, U) \ell(u) \\
&= k(u, U)^T K(U, U)^{-1} k(u, U)
\end{aligned}$$

as required. This completes the proof.  $\square$

The second string of equalities, appearing in the middle part of the proof Proposition 3.5, might appear counter-intuitive at first glance in that the left-most quantity is a norm squared of quantities which scale like  $k$ , whilst the right-most quantity scales like  $k$  itself. However, the space  $H_k$  itself depends on the kernel  $k$ , and scales inversely proportional to  $k$ , explaining that the identity is indeed dimensionally correct.

*Remark 3.6* (Exponential convergence for the Gaussian kernel). The RKHS corresponding to the Gaussian kernel (3.3) is no longer isomorphic to a Sobolev space; it is contained in  $H^\tau(X)$ , for any  $\tau < \infty$ . For functions  $f$  in this RKHS, Gaussian process regression with the Gaussian kernel converges exponentially in the fill distance  $h_U$ . For more details, see [43].

*Remark 3.7* (Regression with non-zero mean). If in (3.1) we use a non-zero mean  $m(\cdot)$ , the formula for the predictive mean  $m_N^f$  changes to

$$(3.8) \quad m_N^f(u) = m(u) + k(u, U)^T K(U, U)^{-1} (f(U) - m(U)),$$

where  $m(U) := [m(u^1), \dots, m(u^N)] \in \mathbb{R}^N$ . The predictive covariance  $k_N(u, u')$  is as in (3.6). As in the case  $m \equiv 0$ , we have  $m_N^f(u^n) = f(u^n)$ , for  $n = 1, \dots, N$ , and  $m_N^f$  is an interpolant of  $f$ . If  $m \in H_k$ , then  $m_N^f$  given by (3.8) is also in  $H_k$ , and the proof techniques in [27, 28] can be applied. The conclusions of Propositions 3.3 and 3.4 then hold, with the factor  $\|f\|$  in the error bounds replaced by  $\|f\| + \|m\|$ .



## 4. APPROXIMATION OF THE BAYESIAN POSTERIOR DISTRIBUTION

In this section, we analyse the error introduced in the posterior distribution  $\mu^y$  when we use a Gaussian process emulator to approximate the parameter-to-observation map  $\mathcal{G}$  or the negative log-likelihood  $\Phi$ . The aim is to show convergence, in a suitable sense, of the approximate posterior distributions to the true posterior distribution as the number of observations  $N$  tends to infinity. For a given approximation  $\mu^{y,N}$  of the posterior distribution  $\mu^y$ , we will focus on bounding the Hellinger distance [41] between the two distributions, which is defined as

$$d_{\text{Hell}}(\mu^y, \mu^{y,N}) = \left( \frac{1}{2} \int_X \left( \sqrt{\frac{d\mu^y}{d\mu_0}} - \sqrt{\frac{d\mu^{y,N}}{d\mu_0}} \right)^2 d\mu_0 \right)^{1/2}.$$

As proven in [15, Lemma 6.12 and 6.14], the Hellinger distance provides a bound for the total variation distance

$$d_{\text{TV}}(\mu^y, \mu^{y,N}) = \frac{1}{2} \sup_{\|f\|_\infty \leq 1} |\mathbb{E}_{\mu^y}(f) - \mathbb{E}_{\mu^{y,N}}(f)| \leq \sqrt{2} d_{\text{Hell}}(\mu^y, \mu^{y,N}),$$

and for  $f \in L^2_{\mu^y}(X) \cap L^2_{\mu^{y,N}}(X)$ , the Hellinger distance also provides a bound on the error in expected values

$$|\mathbb{E}_{\mu^y}(f) - \mathbb{E}_{\mu^{y,N}}(f)| \leq 2(\mathbb{E}_{\mu^y}(f^2) + \mathbb{E}_{\mu^{y,N}}(f^2))^{1/2} d_{\text{Hell}}(\mu^y, \mu^{y,N}).$$

Depending on how we make use of the predictive process  $\mathcal{G}_N$  or  $\Phi_N$  to approximate the Radon-Nikodym derivative  $\frac{d\mu^y}{d\mu_0}$ , we obtain different approximations to the posterior distribution  $\mu^y$ . We will distinguish between approximations based solely on the predictive mean, and approximations that make use of the full predictive process.

**4.1. Approximation based on the predictive mean.** Using simply the predictive mean of a Gaussian process emulator of the parameter-to-observation map  $\mathcal{G}$  or the negative log-likelihood  $\Phi$ , we can define the approximations  $\mu_{\text{mean}}^{y,N,\mathcal{G}}$  and  $\mu_{\text{mean}}^{y,N,\Phi}$ , given by

$$\begin{aligned} \frac{d\mu_{\text{mean}}^{y,N,\mathcal{G}}}{d\mu_0}(u) &= \frac{1}{Z_{N,\mathcal{G}}^{\text{mean}}} \exp\left(-\frac{1}{2\sigma_\eta^2} \|y - m_N^{\mathcal{G}}(u)\|^2\right), \\ Z_{N,\mathcal{G}}^{\text{mean}} &= \mathbb{E}_{\mu_0}\left(\exp\left(-\frac{1}{2\sigma_\eta^2} \|y - m_N^{\mathcal{G}}(u)\|^2\right)\right), \\ \frac{d\mu_{\text{mean}}^{y,N,\Phi}}{d\mu_0}(u) &= \frac{1}{Z_{N,\Phi}^{\text{mean}}} \exp\left(-m_N^\Phi(u)\right), \\ Z_{N,\Phi}^{\text{mean}} &= \mathbb{E}_{\mu_0}\left(\exp\left(-m_N^\Phi(u)\right)\right), \end{aligned}$$

where  $m_N^{\mathcal{G}}(u) = [m_N^{\mathcal{G}^1}(u), \dots, m_N^{\mathcal{G}^J}(u)] \in \mathbb{R}^J$ . We have the following lemma concerning the normalising constants  $Z_{N,\mathcal{G}}^{\text{mean}}$  and  $Z_{N,\Phi}^{\text{mean}}$ , which is followed by the main Theorem 4.2 and Corollary 4.3 concerning the approximations  $\mu_{\text{mean}}^{y,N,\mathcal{G}}$ ,  $\mu_{\text{mean}}^{y,N,\Phi}$ .

**Lemma 4.1.** *Suppose  $\sup_{u \in X} \|\mathcal{G}(u) - m_N^{\mathcal{G}}(u)\|$  and  $\sup_{u \in X} |\Phi(u) - m_N^\Phi(u)|$  converge to 0 as  $N$  tends to  $\infty$ , and assume  $\sup_{u \in X} \|\mathcal{G}(u)\| \leq C_{\mathcal{G}}$ . Then there exist positive constants  $C_1$  and  $C_2$ , independent of  $U$  and  $N$ , such that*

$$C_1 \leq Z_{N,\mathcal{G}}^{\text{mean}} \leq 1 \quad \text{and} \quad C_2^{-1} \leq Z_{N,\Phi}^{\text{mean}} \leq C_2.$$

*Proof.* Let us first consider  $Z_{N,\mathcal{G}}^{\text{mean}}$ . The upper bound follows from a straightforward calculation, since the potential  $\frac{1}{2\sigma_\eta^2} \|y - m_N^{\mathcal{G}}(u)\|^2$  is non-negative:

$$Z_{N,\mathcal{G}}^{\text{mean}} = \mathbb{E}_{\mu_0} \left( \exp \left( - \frac{1}{2\sigma_\eta^2} \|y - m_N^{\mathcal{G}}(u)\|^2 \right) \right) \leq \mathbb{E}_{\mu_0}(1) = 1.$$

For the lower bound, we have

$$\begin{aligned} Z_{N,\mathcal{G}}^{\text{mean}} &\geq \mathbb{E}_{\mu_0} \left( \exp \left( - \frac{1}{2\sigma_\eta^2} \sup_{u \in X} \|y - m_N^{\mathcal{G}}(u)\|^2 \right) \right) \\ &= \exp \left( - \frac{1}{2\sigma_\eta^2} \sup_{u \in X} \|y - m_N^{\mathcal{G}}(u)\|^2 \right), \end{aligned}$$

since  $\int_X \mu_0(du) = 1$ . Using the triangle inequality, the assumption  $\sup_{u \in X} \|\mathcal{G}(u)\| \leq C_{\mathcal{G}}$  and the fact that every convergent sequence is bounded, we have

$$(4.1) \quad \sup_{u \in X} \|y - m_N^{\mathcal{G}}(u)\|^2 \leq \sup_{u \in X} \|y - \mathcal{G}(u)\|^2 + \sup_{u \in X} \|\mathcal{G}(u) - m_N^{\mathcal{G}}(u)\|^2 =: -\ln C_1,$$

where  $C_1$  is independent of  $U$  and  $N$ .

The proof for  $Z_{N,\Phi}^{\text{mean}}$  is similar. For the upper bound, we use  $\int_X \mu_0(du) = 1$  and the triangle inequality to derive

$$\begin{aligned} Z_{N,\Phi}^{\text{mean}} &\leq \sup_{u \in X} \exp \left( - m_N^{\Phi}(u) \right) \leq \exp \left( \sup_{u \in X} |m_N^{\Phi}(u)| \right) \\ &\leq \exp \left( \sup_{u \in X} |\Phi(u)| + \sup_{u \in X} |\Phi(u) - m_N^{\Phi}(u)| \right). \end{aligned}$$

Since  $\sup_{u \in X} |\Phi(u)|$  is bounded when  $\sup_{u \in X} \|\mathcal{G}(u)\|$  is bounded, the fact that every convergent sequence is bounded again gives

$$\sup_{u \in X} |\Phi(u)| + \sup_{u \in X} |\Phi(u) - m_N^{\Phi}(u)| =: -\ln C_2,$$

for a constant  $C_2$  independent of  $U$  and  $N$ . For the lower bound, we note that since  $\int_X \mu_0(du) = 1$ ,

$$Z_{N,\Phi}^{\text{mean}} \geq \mathbb{E}_{\mu_0} \left( \exp \left( - \sup_{u \in X} |m_N^{\Phi}(u)| \right) \right) = \exp \left( - \sup_{u \in X} |m_N^{\Phi}(u)| \right) \geq C_2^{-1}. \quad \square$$

We would like to point out that the assumptions in Lemma 4.1 can be relaxed to assuming that the sequences  $\sup_{u \in X} \|\mathcal{G}(u) - m_N^{\mathcal{G}}(u)\|$  and  $\sup_{u \in X} |\Phi(u) - m_N^{\Phi}(u)|$  are bounded, since this is sufficient to prove the result.

We may now prove the desired theorem and corollary concerning  $\mu_{\text{mean}}^{y,N,\mathcal{G}}$  and  $\mu_{\text{mean}}^{y,N,\Phi}$ .

**Theorem 4.2.** *Under the Assumptions of Lemma 4.1, there exist constants  $C_1$  and  $C_2$ , independent of  $U$  and  $N$ , such that*

$$\begin{aligned} d_{\text{Hell}}(\mu^y, \mu_{\text{mean}}^{y,N,\mathcal{G}}) &\leq C_1 \left\| \mathcal{G} - m_N^{\mathcal{G}} \right\|_{L_{\mu_0}^2(X; \mathbb{R}^J)}, \\ \text{and } d_{\text{Hell}}(\mu^y, \mu_{\text{mean}}^{y,N,\Phi}) &\leq C_2 \left\| \Phi - m_N^{\Phi} \right\|_{L_{\mu_0}^2(X)}. \end{aligned}$$

*Proof.* Let us first consider  $\mu_{\text{mean}}^{y,N,\mathcal{G}}$ . By definition of the Hellinger distance, we have

$$\begin{aligned} 2 d_{\text{Hell}}^2(\mu^y, \mu_{\text{mean}}^{y,N,\mathcal{G}}) &= \int_X \left( \sqrt{\frac{d\mu^y}{d\mu_0}} - \sqrt{\frac{d\mu_{\text{mean}}^{y,N,\mathcal{G}}}{d\mu_0}} \right)^2 \mu_0(du) \\ &\leq \frac{2}{Z} \int_X \left( \exp\left(-\frac{1}{4\sigma_\eta^2} \|y - \mathcal{G}(u)\|^2\right) - \exp\left(-\frac{1}{4\sigma_\eta^2} \|y - m_N^{\mathcal{G}}(u)\|^2\right) \right)^2 \mu_0(du) \\ &\quad + 2 Z_{N,\mathcal{G}}^{\text{mean}} \left( Z^{-1/2} - (Z_{N,\mathcal{G}}^{\text{mean}})^{-1/2} \right)^2 \\ &=: I + II. \end{aligned}$$

For the first term, we use the local Lipschitz continuity of the exponential function, together with the equality  $a^2 - b^2 = (a-b)(a+b)$  and the reverse triangle inequality to bound

$$\begin{aligned} \frac{Z}{2} I &= \int_X \left( \exp\left(-\frac{1}{4\sigma_\eta^2} \|y - \mathcal{G}(u)\|^2\right) - \exp\left(-\frac{1}{4\sigma_\eta^2} \|y - m_N^{\mathcal{G}}(u)\|^2\right) \right)^2 \mu_0(du) \\ &\leq \int_X \left( \frac{1}{2\sigma_\eta^2} \left( \|y - \mathcal{G}(u)\|^2 - \|y - m_N^{\mathcal{G}}(u)\|^2 \right) \right)^2 \mu_0(du) \\ &= \int_X \frac{1}{4\sigma_\eta^4} \left( \|y - \mathcal{G}(u)\| + \|y - m_N^{\mathcal{G}}(u)\| \right)^2 \|\mathcal{G}(u) - m_N^{\mathcal{G}}(u)\|^2 \mu_0(du) \\ &\leq \frac{1}{4\sigma_\eta^4} \sup_{u \in X} \left( \|y - \mathcal{G}(u)\| + \|y - m_N^{\mathcal{G}}(u)\| \right)^2 \|\mathcal{G}(u) - m_N^{\mathcal{G}}(u)\|_{L_{\mu_0}^2(X; \mathbb{R}^J)}^2. \end{aligned}$$

As in equation (4.1), the first supremum can be bounded independently of  $U$  and  $N$ , from which it follows that

$$I \leq C \|\mathcal{G}(u) - m_N^{\mathcal{G}}(u)\|_{L_{\mu_0}^2(X; \mathbb{R}^J)}^2,$$

for a constant  $C$  independent of  $U$  and  $N$ . For the second term, a very similar argument, together with Lemma 4.1 and Jensen's inequality, shows

$$\begin{aligned} II &= 2 Z_{N,\mathcal{G}}^{\text{mean}} \left( Z^{-1/2} - (Z_{N,\mathcal{G}}^{\text{mean}})^{-1/2} \right)^2 \\ &\leq 2 Z_{N,\mathcal{G}}^{\text{mean}} \max(Z^{-3}, (Z_{N,\mathcal{G}}^{\text{mean}})^{-3}) |Z - Z_{N,\mathcal{G}}^{\text{mean}}|^2 \\ &= 2 Z_{N,\mathcal{G}}^{\text{mean}} \max(Z^{-3}, (Z_{N,\mathcal{G}}^{\text{mean}})^{-3}) \left( \int_X \exp\left(-\frac{1}{4\sigma_\eta^2} \|y - \mathcal{G}(u)\|^2\right) \right. \\ &\quad \left. - \exp\left(-\frac{1}{4\sigma_\eta^2} \|y - m_N^{\mathcal{G}}(u)\|^2\right) \mu_0(du) \right)^2 \\ &\leq C \|\mathcal{G}(u) - m_N^{\mathcal{G}}(u)\|_{L_{\mu_0}^2(X; \mathbb{R}^J)}^2, \end{aligned}$$

for a constant  $C$  independent of  $U$  and  $N$ .

The proof for  $\mu_{\text{mean}}^{y,N,\Phi}$  is similar. We use an identical corresponding splitting of the Hellinger distance  $d_{\text{Hell}}(\mu^y, \mu_{\text{mean}}^{y,N,\Phi}) \leq I + II$ . Using the local Lipschitz continuity

of the exponential function, we have

$$\begin{aligned} \frac{Z}{2} I &= \int_X (\exp(-\Phi(u)) - \exp(-m_N^\Phi(u)))^2 \mu_0(du) \\ &\leq (1 + \exp(\sup_{u \in X} |m_N^\Phi(u)|)) \|\Phi(u) - m_N^\Phi(u)\|_{L^2_{\mu_0}(X)}^2, \end{aligned}$$

where the first factor can be bounded independently of  $U$  and  $N$  as in Lemma 4.1. Using Lemma 4.1 and Jensen's inequality, we furthermore have

$$\begin{aligned} II &\leq 2 Z_{N,\Phi}^{\text{mean}} \max(Z^{-3}, (Z_{N,\Phi}^{\text{mean}})^{-3}) \left( \int_X \exp(-\Phi(u)) - \exp(-m_N^\Phi(u)) \mu_0(du) \right)^2 \\ &\leq C \|\Phi(u) - m_N^\Phi(u)\|_{L^2_{\mu_0}(X)}^2, \end{aligned}$$

for a constant  $C$  independent of  $U$  and  $N$ .  $\square$

We remark here that Theorem 4.2 does not make any assumptions on the predictive means  $m_N^{\mathcal{G}}$  and  $m_N^\Phi$  other than the requirement that  $\sup_{u \in X} \|\mathcal{G}(u) - m_N^{\mathcal{G}}(u)\|$  and  $\sup_{u \in X} |\Phi(u) - m_N^\Phi(u)|$  converge to 0 as  $N$  tends to  $\infty$ . Whether the predictive means are defined as in (3.6), or are derived by alternative approaches to Gaussian process regression [32], does not affect the conclusions of Theorem 4.2. Under Assumption 2.2, we can combine Theorem 4.2 with Proposition 3.3 (or Proposition 3.4) with  $\beta = 0$  to obtain error bounds in terms of the fill distance of the design points.

**Corollary 4.3.** *Suppose  $m_N^\Phi$  and  $m_N^{\mathcal{G}^j}$ ,  $j = 1, \dots, J$ , are defined as in (3.6), with Matérn kernel  $k = k_{\nu, \lambda, \sigma_k^2}$ . Suppose Assumption 2.2 holds with  $s = \nu + K/2$ , and the assumptions of Proposition 3.3 and Theorem 4.2 are satisfied. Then there exist constants  $C_1$  and  $C_2$ , independent of  $U$  and  $N$ , such that*

$$d_{\text{Hell}}(\mu^y, \mu_{\text{mean}}^{y, N, \mathcal{G}}) \leq C_1 h_U^{\nu+K/2} \quad \text{and} \quad d_{\text{Hell}}(\mu^y, \mu_{\text{mean}}^{y, N, \Phi}) \leq C_2 h_U^{\nu+K/2}.$$

If Assumption 2.2 holds only for some  $s < \nu + K/2$ , an analogue of Corollary 4.3 can be proved using Proposition 3.4 with  $\beta = 0$ . As already discussed in section 3.2, translating convergence rates in terms of the fill distance  $h_U$  into rates in terms of the number of points  $N$  typically leads to a strong dependence on the input dimension  $K$ . For uniform tensor grids  $U$ , the rates of convergence in  $N$  predicted by Corollary 4.3 are given in Table 1.

**4.2. Approximations based on the predictive process.** Alternative to the mean-based approximations considered in the previous section, we now consider approximations to the posterior distribution  $\mu^y$  obtained using the full predictive processes  $\mathcal{G}_N$  and  $\Phi_N$ . In contrast to the mean, the full Gaussian processes also carry information about the uncertainty in the emulator due to only using a finite number of function evaluations to construct it.

For the remainder of this section, we denote by  $\nu_N^{\mathcal{G}}$  the distribution of  $\mathcal{G}_N$  and by  $\nu_N^\Phi$  the distribution of  $\Phi_N$ , for  $N \in \mathbb{N} \cup \{0\}$ . We note that since the process  $\mathcal{G}_N$  consists of  $J$  independent Gaussian processes  $\mathcal{G}_N^j$ , the measure  $\nu_N^{\mathcal{G}}$  is a product measure,  $\nu_N^{\mathcal{G}} = \prod_{j=1}^J \nu_N^{\mathcal{G}^j}$ .  $\Phi_N$  is a Gaussian process with mean  $m_N^\Phi$  and covariance kernel  $k_N$ , and  $\mathcal{G}_N^j$ , for  $j = 1, \dots, J$ , is a Gaussian process with mean  $m_N^{\mathcal{G}^j}$  and

covariance kernel  $k_N$ . Replacing  $\mathcal{G}$  by  $\mathcal{G}_N$  in (2.1), we obtain the approximation  $\mu_{\text{sample}}^{y,N,\mathcal{G}}$  given by

$$\frac{d\mu_{\text{sample}}^{y,N,\mathcal{G}}}{d\mu_0}(u) = \frac{1}{Z_{N,\mathcal{G}}^{\text{sample}}} \exp\left(-\frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}_N(u)\|^2\right),$$

where

$$Z_{N,\mathcal{G}}^{\text{sample}} = \mathbb{E}_{\mu_0}\left(\exp\left(-\frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}_N(u)\|^2\right)\right).$$

Similarly, we define for the predictive process  $\Phi_N$  the approximation  $\mu_{\text{sample}}^{y,N,\Phi}$  by

$$\frac{d\mu_{\text{sample}}^{y,N,\Phi}}{d\mu_0}(u) = \frac{1}{Z_{N,\Phi}^{\text{sample}}} \exp\left(-\Phi_N(u)\right), \quad Z_{N,\Phi}^{\text{sample}} = \mathbb{E}_{\mu_0}\left(\exp\left(-\Phi_N(u)\right)\right).$$

The measures  $\mu_{\text{sample}}^{y,N,\mathcal{G}}$  and  $\mu_{\text{sample}}^{y,N,\Phi}$  are random approximations of the deterministic measure  $\mu^y$ . The uncertainty in the posterior distribution introduced in this way can be thought of representing the uncertainty in the emulator, which in applications can be large (or comparable) to the uncertainty present in the observations. A user may want to take this into account to “inflate” the variance of the posterior distribution.

Deterministic approximations of the posterior distribution  $\mu^y$  can now be obtained by taking the expected value with respect to the predictive processes  $\mathcal{G}_N$  and  $\Phi_N$ . This results in the marginal approximations

$$\begin{aligned} \frac{d\mu_{\text{marginal}}^{y,N,\mathcal{G}}}{d\mu_0}(u) &= \frac{1}{\mathbb{E}_{\nu_N^{\mathcal{G}}}(Z_{N,\mathcal{G}}^{\text{sample}})} \mathbb{E}_{\nu_N^{\mathcal{G}}}\left(\exp\left(-\frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}_N(u)\|^2\right)\right), \\ \frac{d\mu_{\text{marginal}}^{y,N,\Phi}}{d\mu_0}(u) &= \frac{1}{\mathbb{E}_{\nu_N^{\Phi}}(Z_{N,\Phi}^{\text{sample}})} \mathbb{E}_{\nu_N^{\Phi}}\left(\exp\left(-\Phi_N(u)\right)\right). \end{aligned}$$

Note that by Tonelli’s Theorem ([35], a version of Fubini’s Theorem for non-negative integrands), the measures  $\mu_{\text{marginal}}^{y,N,\mathcal{G}}$  and  $\mu_{\text{marginal}}^{y,N,\Phi}$  are indeed probability measures. It can be shown that the above approximation of the likelihood is optimal in the sense that it minimises the  $L^2$ -error [39]. In contrast to the approximation based on only the mean of the emulator, this approximation also takes into account the uncertainty of the emulator, although only in an averaged sense. The likelihood in the marginal approximations  $\mu_{\text{marginal}}^{y,N,\mathcal{G}}$  and  $\mu_{\text{marginal}}^{y,N,\Phi}$  involves computing an expectation. Methods from the pseudo-marginal MCMC literature [2] could be used within an MCMC method in this context.

Before proving bounds on the error in the marginal approximations  $\mu_{\text{marginal}}^{y,N,\mathcal{G}}$  and  $\mu_{\text{marginal}}^{y,N,\Phi}$  in section 4.2.2, and the error in the random approximations  $\mu_{\text{sample}}^{y,N,\mathcal{G}}$  and  $\mu_{\text{sample}}^{y,N,\Phi}$  in section 4.2.3, we crucially prove boundedness of the normalising constants  $Z_{N,\mathcal{G}}^{\text{sample}}$  and  $Z_{N,\Phi}^{\text{sample}}$  in section 4.2.1.

**4.2.1. Moment bounds on  $Z_{N,\mathcal{G}}^{\text{sample}}$  and  $Z_{N,\Phi}^{\text{sample}}$ .** First, we recall the following classical results from the theory of Gaussian measures on Banach spaces [1, 31].

**Proposition 4.4** (Fernique's Theorem). *Let  $E$  be a separable Banach space and  $\nu$  a centred Gaussian measure on  $(E, \mathcal{B}(E))$ . If  $\lambda, r > 0$  are such that*

$$\log \left( \frac{1 - \nu(f \in E : \|f\|_E \leq r)}{\nu(f \in E : \|f\|_E \leq r)} \right) \leq -1 - 32\lambda r^2,$$

*then*

$$\int_E \exp(\lambda \|f\|_E^2) \nu(df) \leq \exp(16\lambda r^2) + \frac{e^2}{e^2 - 1}.$$

**Proposition 4.5** (Borell-TIS inequality<sup>1</sup>). *Let  $f$  be a scalar, almost surely bounded Gaussian field on a compact domain  $T \subseteq \mathbb{R}^K$ , with zero mean  $\mathbb{E}(f(t)) = 0$  and bounded variance  $0 < \sigma_f^2 := \sup_{t \in T} \mathbb{V}(f(t)) < \infty$ . Then  $\mathbb{E}(\sup_{t \in T} f(t)) < \infty$ , and for all  $r > 0$ ,*

$$\mathbb{P}(\sup_{t \in T} f(t) - \mathbb{E}(\sup_{t \in T} f(t)) > r) \leq \exp(-r^2/2\sigma_f^2).$$

**Proposition 4.6** (Sudakov-Fernique inequality). *Let  $f$  and  $g$  be scalar, almost surely bounded Gaussian fields on a compact domain  $T \subseteq \mathbb{R}^K$ . Suppose that  $\mathbb{E}((f(t) - f(s))^2) \leq \mathbb{E}((g(t) - g(s))^2)$  and  $\mathbb{E}(f(t)) = \mathbb{E}(g(t))$ , for all  $s, t \in T$ . Then*

$$\mathbb{E}(\sup_{t \in T} f(t)) \leq \mathbb{E}(\sup_{t \in T} g(t)).$$

Using these results, we are now ready to prove bounds on moments of  $Z_{N,\mathcal{G}}^{\text{sample}}$  and  $Z_{N,\Phi}^{\text{sample}}$ , similar to those proved in Lemma 4.1. The reader interested purely in approximation results for the posterior can simply read the statements of the following two lemmas, and then proceed directly to subsections 4.2.2 and 4.2.3.

Recall that, as in (3.1),  $\Phi_0$  and  $\mathcal{G}_0^j$  denote the initial Gaussian process models for  $\Phi$  and  $\mathcal{G}^j$ , respectively, and, as in (3.5),  $\Phi_N$  and  $\mathcal{G}_N^j$  denote the conditioned Gaussian process models for  $\Phi$  and  $\mathcal{G}^j$ , respectively.

**Lemma 4.7.** *Let  $X \subseteq \mathbb{R}^K$  be compact. Suppose that  $\sup_{u \in X} \|\mathcal{G}(u) - m_N^{\mathcal{G}}(u)\|$ ,  $\sup_{u \in X} \|\Phi(u) - m_N^{\Phi}(u)\|$  and  $\sup_{u \in X} k_N(u, u)$  converge to 0 as  $N$  tends to infinity, and assume  $\sup_{u \in X} \|\mathcal{G}(u)\| \leq C_{\mathcal{G}} < \infty$ . Suppose the assumptions of the Sudakov-Fernique inequality hold, for  $g = \Phi_0$  and  $f = \Phi_N - m_N^{\Phi}$ , and for  $g = \mathcal{G}_0^j$  and  $f = \mathcal{G}_N^j - m_N^{\mathcal{G}^j}$ , for  $j \in \{1, \dots, J\}$ . Then, for any  $1 \leq p < \infty$ , there exist positive constants  $C_1$  and  $C_2$ , independent of  $U$  and  $N$ , such that for all  $N$  sufficiently large*

$$C_1^{-1} \leq \mathbb{E}_{\nu_N^{\mathcal{G}}}((Z_{N,\mathcal{G}}^{\text{sample}})^p) \leq 1 \quad \text{and} \quad 1 \leq \mathbb{E}_{\nu_N^{\mathcal{G}}}((Z_{N,\mathcal{G}}^{\text{sample}})^{-p}) \leq C_1$$

*and*

$$C_2^{-1} \leq \mathbb{E}_{\nu_N^{\Phi}}((Z_{N,\Phi}^{\text{sample}})^p) \leq C_2 \quad \text{and} \quad C_2^{-1} \leq \mathbb{E}_{\nu_N^{\Phi}}((Z_{N,\Phi}^{\text{sample}})^{-p}) \leq C_2.$$

*Proof.* We start with  $Z_{N,\mathcal{G}}^{\text{sample}}$ . Since the potential  $\frac{1}{2\sigma_{\eta}^2} \|y - \mathcal{G}_N(u)\|^2$  is non-negative and  $\int_X \mu_0(du) = 1 = \int_{C^0(X; \mathbb{R}^J)} \nu_N^{\mathcal{G}}(d\mathcal{G}_N)$ , we have for any  $1 \leq p < \infty$ ,

$$\begin{aligned} \mathbb{E}_{\nu_N^{\mathcal{G}}}((Z_{N,\mathcal{G}}^{\text{sample}})^p) &= \int_{C^0(X; \mathbb{R}^J)} \left( \int_X \exp\left(-\frac{1}{2\sigma_{\eta}^2} \|y - \mathcal{G}_N(u)\|^2\right) \mu_0(du) \right)^p \nu_N^{\mathcal{G}}(d\mathcal{G}_N) \\ &\leq 1. \end{aligned}$$

---

<sup>1</sup>The Borell-TIS inequality is named after the mathematicians Borell and Tsirelson, Ibragimov and Sudakov, who independently proved the result.

From Jensen's inequality, it then follows that

$$\mathbb{E}_{\nu_N^{\mathcal{G}}}((Z_{N,\mathcal{G}}^{\text{sample}})^{-p}) \geq (\mathbb{E}_{\nu_N^{\mathcal{G}}}((Z_{N,\mathcal{G}}^{\text{sample}})^p))^{-1} \geq 1.$$

To determine  $C_1$ , we use the triangle inequality to bound, for any  $1 \leq p < \infty$ ,

$$\begin{aligned} \mathbb{E}_{\nu_N^{\mathcal{G}}}((Z_{N,\mathcal{G}}^{\text{sample}})^{-p}) &= \int_{C^0(X;\mathbb{R}^J)} \left( \int_X \exp\left(-\frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}_N(u)\|^2\right) \mu_0(du) \right)^{-p} \nu_N^{\mathcal{G}}(d\mathcal{G}_N) \\ &\leq \int_{C^0(X;\mathbb{R}^J)} \left( \exp\left(-\frac{1}{2\sigma_\eta^2} \sup_{u \in X} \|y - \mathcal{G}_N(u)\|^2\right) \right)^{-p} \nu_N^{\mathcal{G}}(d\mathcal{G}_N) \\ &= \int_{C^0(X;\mathbb{R}^J)} \exp\left(\frac{p}{2\sigma_\eta^2} \sup_{u \in X} \|y - \mathcal{G}_N(u)\|^2\right) \nu_N^{\mathcal{G}}(d\mathcal{G}_N) \\ &\leq \exp\left(\frac{\sup_{u \in X} \|y - m_N^{\mathcal{G}}(u)\|^2}{2p^{-1}\sigma_\eta^2}\right) \\ &\quad \times \int_{C^0(X;\mathbb{R}^J)} \exp\left(\frac{\sup_{u \in X} \|\mathcal{G}_N(u) - m_N^{\mathcal{G}}(u)\|^2}{2p^{-1}\sigma_\eta^2}\right) \nu_N^{\mathcal{G}}(d\mathcal{G}_N). \end{aligned}$$

The first factor can be bounded independently of  $U$  and  $N$  using the triangle inequality, together with  $\sup_{u \in X} \|\mathcal{G}(u)\| \leq C_{\mathcal{G}}$  and  $\sup_{u \in X} \|\mathcal{G}(u) - m_N^{\mathcal{G}}(u)\| \rightarrow 0$  as  $N \rightarrow \infty$ . For the second factor, we use Fernique's Theorem (Proposition 4.4). First, we note that (using independence)

$$\begin{aligned} &\int_{C^0(X;\mathbb{R}^J)} \exp\left(\frac{\sup_{u \in X} \|\mathcal{G}_N(u) - m_N^{\mathcal{G}}(u)\|^2}{2p^{-1}\sigma_\eta^2}\right) \nu_N^{\mathcal{G}}(d\mathcal{G}_N) \\ &= \int_{C^0(X;\mathbb{R}^J)} \exp\left(\frac{\sup_{u \in X} \sum_{j=1}^J |\mathcal{G}_N^j(u) - m_N^{\mathcal{G}^j}(u)|^2}{2p^{-1}\sigma_\eta^2}\right) \nu_N^{\mathcal{G}}(d\mathcal{G}_N) \\ &\leq \int_{C^0(X;\mathbb{R}^J)} \exp\left(\sum_{j=1}^J \frac{\sup_{u \in X} |\mathcal{G}_N^j(u) - m_N^{\mathcal{G}^j}(u)|^2}{2p^{-1}\sigma_\eta^2}\right) \nu_N^{\mathcal{G}}(d\mathcal{G}_N) \\ &= \int_{C^0(X;\mathbb{R}^J)} \prod_{j=1}^J \exp\left(\frac{\sup_{u \in X} |\mathcal{G}_N^j(u) - m_N^{\mathcal{G}^j}(u)|^2}{2p^{-1}\sigma_\eta^2}\right) \nu_N^{\mathcal{G}}(d\mathcal{G}_N) \\ &= \prod_{j=1}^J \int_{C^0(X)} \exp\left(\frac{\sup_{u \in X} |\mathcal{G}_N^j(u) - m_N^{\mathcal{G}^j}(u)|^2}{2p^{-1}\sigma_\eta^2}\right) \nu_N^{\mathcal{G}^j}(d\mathcal{G}_N^j). \end{aligned}$$

It remains to show that, for  $N$  sufficiently large, the assumptions of Fernique's Theorem hold for  $\lambda = p\sigma_\eta^{-2}/2$  and a value of  $r$  independent of  $U$  and  $N$ , for  $\nu$  equal to the push-forward of  $\nu_N^{\mathcal{G}^j}$  under the map  $T(f) = f - m_N^{\mathcal{G}^j}$ . Denote by  $B_{N,r}^{\mathcal{G}^j} \subset C^0(X)$  the set of all functions  $f$  such that  $\|f - m_N^{\mathcal{G}^j}\|_{C^0(X)} \leq r$ , for some fixed  $r > 0$  and  $1 \leq j \leq J$ . Let  $\overline{\mathcal{G}_N^j} = \mathcal{G}_N^j - m_N^{\mathcal{G}^j}$ . By the Borell-TIS inequality, we have for all  $r > \mathbb{E}(\sup_{u \in X} \overline{\mathcal{G}_N^j}(u))$ ,

$$\nu_N^{\mathcal{G}^j}(\mathcal{G}_N^j : \sup_{u \in X} \overline{\mathcal{G}_N^j}(u) > r) \leq \exp\left(-\frac{(r - \mathbb{E}(\sup_{u \in X} \overline{\mathcal{G}_N^j}(u)))^2}{2\sigma_N^2}\right),$$

where  $\sigma_N^2 := \sup_{u \in X} k_N(u, u)$ . By assumption,  $\mathbb{E}_{\nu_N^j}((\overline{\mathcal{G}_N^j}(u) - \overline{\mathcal{G}_N^j}(u'))^2) \leq \mathbb{E}_{\nu_0^j}((\mathcal{G}_0^j(u) - \mathcal{G}_0^j(u'))^2)$ , and so  $\mathbb{E}(\sup_{u \in X} \overline{\mathcal{G}_N^j}(u)) \leq \mathbb{E}(\sup_{u \in X} \mathcal{G}_0^j(u))$ , by the Sudakov-Fernique inequality. We can hence choose  $r > \mathbb{E}(\sup_{u \in X} \mathcal{G}_0^j(u))$ , independent of  $U$  and  $N$ , such that the bound

$$\nu_N^{\mathcal{G}^j}(\mathcal{G}_N^j : \sup_{u \in X} \overline{\mathcal{G}_N^j}(u) > r) \leq \exp\left(-\frac{(r - \mathbb{E}(\sup_{u \in X} \mathcal{G}_0^j(u)))^2}{2\sigma_N^2}\right),$$

holds for all  $N \in \mathbb{N}$ . By assumption we have  $\sigma_N^2 \rightarrow 0$  as  $N \rightarrow \infty$ , and by the symmetry of Gaussian measures, we hence have  $\nu_N^{\mathcal{G}^j}(B_{N,r}^{\mathcal{G}^j}) \rightarrow 1$  as  $N \rightarrow \infty$ , for all  $r > \mathbb{E}(\sup_{u \in X} \mathcal{G}_0^j(u))$ . For  $N = N(p)$  sufficiently large, the inequality

$$\log\left(\frac{1 - \nu_N^{\mathcal{G}^j}(B_{N,r}^{\mathcal{G}^j})}{\nu_N^{\mathcal{G}^j}(B_{N,r}^{\mathcal{G}^j})}\right) \leq -1 - 32\lambda r^2,$$

in the assumptions of Fernique's Theorem is then satisfied, for  $\lambda = p\sigma_\eta^{-2}/2$  and  $r > \mathbb{E}(\sup_{u \in X} \mathcal{G}_0^j(u))$ , both independent of  $U$  and  $N$ . Hence,  $\mathbb{E}((Z_{N,\mathcal{G}}^{\text{sample}})^{-p}) \leq C_1(p)$ , for a constant  $C_1(p) < \infty$  independent of  $U$  and  $N$ . From Jensen's inequality, it then finally follows that

$$\mathbb{E}_{\nu_N^{\mathcal{G}}}((Z_{N,\mathcal{G}}^{\text{sample}})^p) \geq (\mathbb{E}_{\nu_N^{\mathcal{G}}}((Z_{N,\mathcal{G}}^{\text{sample}})^{-p}))^{-1} \geq C_1^{-1}(p).$$

The proof for  $Z_{N,\Phi}^{\text{sample}}$  is similar. Using  $\int_X \mu_0(du) = 1$  and the triangle inequality, we have

$$\begin{aligned} \mathbb{E}_{\nu_N^\Phi}((Z_{N,\Phi}^{\text{sample}})^p) &= \int_{C^0(X)} \left( \int_X \exp(-\Phi_N(u)) \mu_0(du) \right)^p \nu_N^\Phi(d\Phi_N) \\ &\leq \int_{C^0(X)} \exp(p \sup_{u \in X} |\Phi_N(u)|) \nu_N^\Phi(d\Phi_N) \\ &\leq \exp(p \sup_{u \in X} |m_N^\Phi(u)|) \int_{C^0(X)} \exp(p \sup_{u \in X} |\Phi_N(u) - m_N^\Phi(u)|) \nu_N^\Phi(d\Phi_N). \end{aligned}$$

The first factor can be bounded independently of  $U$  and  $N$ , since  $\sup_{u \in X} \|\mathcal{G}(u)\| \leq C_{\mathcal{G}}$  and  $\sup_{u \in X} |\Phi(u) - m_N^\Phi(u)|$  converges to 0 as  $N \rightarrow \infty$ . The second factor can be bounded by Fernique's Theorem. Using the same proof technique as above, we can show that  $\nu_N^\Phi(B_{N,r}^\Phi) \rightarrow 1$  as  $N \rightarrow \infty$  for all  $r > \mathbb{E}(\sup_{u \in X} \Phi_0(u))$ , where  $B_{N,r}^\Phi \subset C^0(X)$  denotes the set of all functions  $f$  such that  $\|f - m_N^\Phi\|_{C^0(X)} \leq r$ . Hence, it is possible to choose  $r > 0$ , independent of  $U$  and  $N$ , such that the assumptions of Fernique's Theorem hold for  $\nu$  equal to the push-forward of  $\nu_N^\Phi$  under the map  $T(f) = f - m_N^\Phi$ , for some  $\lambda > 0$  also independent of  $U$  and  $N$ . By Young's inequality, we have

$$\exp(p \sup_{u \in X} |\Phi_N(u) - m_N^\Phi(u)|) \leq \exp(\lambda \sup_{u \in X} |\Phi_N(u) - m_N^\Phi(u)|^2 + p^2/4\lambda),$$

and it follows that  $\mathbb{E}_\omega((Z_{N,\Phi}^{\text{sample}})^p) \leq C_2(p)$ , for a constant  $C_2(p) < \infty$  independent of  $U$  and  $N$ , for any  $1 \leq p < \infty$ . Furthermore, we note

$$\mathbb{E}_{\nu_N^\Phi}((Z_{N,\Phi}^{\text{sample}})^{-p}) \leq \int_{C^0(X;\mathbb{R})} \exp(p \sup_{u \in X} |\Phi_N(u)|) \nu_N^\Phi(d\Phi_N) \leq C_2(p).$$



By Jensen's inequality, we finally have  $\mathbb{E}_{\nu_N^\Phi}((Z_{N,\Phi}^{\text{sample}})^{-p}) \geq C_2(p)^{-1}$  and  $\mathbb{E}_{\nu_N^\Phi}((Z_{N,\Phi}^{\text{sample}})^p) \geq C_2(p)^{-1}$ .  $\square$

We would like to point out here that the assumption that  $\sup_{u \in X} k_N(u, u)$  converges to 0 as  $N$  tends to infinity in Lemma 4.7 is crucial in order to enable the choice of any  $1 \leq p < \infty$ . This is related to the fact that the parameter  $\lambda$  needs to be sufficiently small compared to  $\sup_{u \in X} k_N(u, u)$  in order to satisfy the assumptions of Fernique's Theorem.

In Lemma 4.7, we supposed that the assumptions of the Sudakov-Fernique inequality hold, for  $g = \Phi_0$  and  $f = \Phi_N - m_N^\Phi$ , and for  $g = \mathcal{G}_0^j$  and  $f = \mathcal{G}_N^j - m_N^{\mathcal{G}_j}$ , for  $j \in \{1, \dots, J\}$ . This is an assumption on the predictive variance  $k_N$ . In the following lemma, we prove this assumption for the predictive variance given in (3.6).

**Lemma 4.8.** *Suppose the predictive variance  $k_N$  is given by (3.6). Then the assumptions of the Sudakov-Fernique inequality hold, for  $g = \Phi_0$  and  $f = \Phi_N - m_N^\Phi$ , and for  $g = \mathcal{G}_0^j$  and  $f = \mathcal{G}_N^j - m_N^{\mathcal{G}_j}$ , for  $j \in \{1, \dots, J\}$ .*

*Proof.* We give a proof for  $g = \Phi_0$  and  $f = \Phi_N - m_N^\Phi$ , the proof for  $g = \mathcal{G}_0^j$  and  $f = \mathcal{G}_N^j - m_N^{\mathcal{G}_j}$ , for  $j \in \{1, \dots, J\}$ , is identical. For any  $u, u' \in X$ , we have  $\mathbb{E}_{\nu_0^\Phi}(\Phi_0(u)) = 0 = \mathbb{E}_{\nu_N^\Phi}(\Phi_N(u) - m_N^\Phi(u))$ , and

$$\begin{aligned} & \mathbb{E}_{\nu_N^\Phi}(((\Phi_N(u) - m_N^\Phi(u)) - (\Phi_N(u') - m_N^\Phi(u'))))^2) \\ &= k_N(u, u) - k_N(u, u') - k_N(u', u) + k_N(u', u'), \\ & \mathbb{E}_{\nu_0^\Phi}((\Phi_0(u) - \Phi_0(u'))^2) = k(u, u) - k(u, u') - k(u', u) + k(u', u'). \end{aligned}$$

By (3.6), we have

$$k_N(u, u') = k(u, u') - k(u, U)^T K(U, U)^{-1} k(u', U),$$

and so

$$\begin{aligned} & \mathbb{E}_{\nu_0^\Phi}((\Phi_0(u) - \Phi_0(u'))^2) - \mathbb{E}_{\nu_N^\Phi}(((\Phi_N(u) - m_N^\Phi(u)) - (\Phi_N(u') - m_N^\Phi(u'))))^2) \\ &= (k(u, U)^T - k(u', U)^T) K(U, U)^{-1} (k(u, U) - k(u', U)) \\ &\geq 0, \end{aligned}$$

since the matrix  $K(U, U)^{-1}$  is positive definite.  $\square$

We are now ready to prove bounds on the approximation error in the posterior distributions.

**4.2.2. Error in the marginal approximations  $\mu_{\text{marginal}}^{y,N,\mathcal{G}}$  and  $\mu_{\text{marginal}}^{y,N,\Phi}$ .** We start by analysing the error in the marginal approximations  $\mu_{\text{marginal}}^{y,N,\mathcal{G}}$  and  $\mu_{\text{marginal}}^{y,N,\Phi}$ .

**Theorem 4.9.** *Under the assumptions of Lemma 4.7, there exist constants  $C_1$  and  $C_2$ , independent of  $U$  and  $N$ , such that for any  $\delta > 0$ ,*

$$\begin{aligned} d_{\text{Hell}}(\mu^y, \mu_{\text{marginal}}^{y,N,\mathcal{G}}) &\leq C_1 \left\| \left( \mathbb{E}_{\nu_N^\Phi}(\|\mathcal{G} - \mathcal{G}_N\|^{1+\delta}) \right)^{1/(1+\delta)} \right\|_{L_{\mu_0}^2(X)}, \\ d_{\text{Hell}}(\mu^y, \mu_{\text{marginal}}^{y,N,\Phi}) &\leq C_2 \left\| \mathbb{E}_{\nu_N^\Phi}(|\Phi - \Phi_N|^{1+\delta})^{1/(1+\delta)} \right\|_{L_{\mu_0}^2(X)}. \end{aligned}$$

*Proof.* We start with  $\mu_{\text{marginal}}^{y,N,\mathcal{G}}$ . By the definition of the Hellinger distance, we have

$$\begin{aligned} 2 d_{\text{Hell}}^2(\mu^y, \mu_{\text{marginal}}^{y,N,\mathcal{G}}) &= \int_X \left( \sqrt{\frac{d\mu^y}{d\mu_0}} - \sqrt{\frac{d\mu_{\text{marginal}}^{y,N,\mathcal{G}}}{d\mu_0}} \right)^2 \mu_0(du) \\ &\leq \frac{2}{Z} \int_X \left( \sqrt{\exp\left(-\frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}(u)\|^2\right)} \right. \\ &\quad \left. - \sqrt{\mathbb{E}_{\nu_N^{\mathcal{G}}} \left( \exp\left(-\frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}_N(u)\|^2\right) \right)} \right)^2 \mu_0(du) \\ &\quad + 2\mathbb{E}_{\nu_N^{\mathcal{G}}} (Z_{N,\mathcal{G}}^{\text{sample}}) \left( Z^{-1/2} - \mathbb{E}_{\nu_N^{\mathcal{G}}} (Z_{N,\mathcal{G}}^{\text{sample}})^{-1/2} \right)^2 \\ &= I + II. \end{aligned}$$

For the first term, we use the (in)equalities  $a - b = (a^2 - b^2)/(a + b)$  and  $(\sqrt{a} + \sqrt{b})^2 \geq a + b$ , for  $a, b > 0$ , to derive

$$\begin{aligned} \frac{Z}{2} I &= \int_X \left( \sqrt{\exp\left(-\frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}(u)\|^2\right)} \right. \\ &\quad \left. - \sqrt{\mathbb{E}_{\nu_N^{\mathcal{G}}} \left( \exp\left(-\frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}_N(u)\|^2\right) \right)} \right)^2 \mu_0(du) \\ &\leq \int_X \frac{\left( \exp\left(-\frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}(u)\|^2\right) - \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( \exp\left(-\frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}_N(u)\|^2\right) \right) \right)^2}{\exp\left(-\frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}(u)\|^2\right) + \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( \exp\left(-\frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}_N(u)\|^2\right) \right)} \mu_0(du) \\ &\leq \sup_{u \in X} \left( \exp\left(-\frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}(u)\|^2\right) + \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( \exp\left(-\frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}_N(u)\|^2\right) \right) \right)^{-1} \\ &\quad \times \int_X \left( \exp\left(-\frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}(u)\|^2\right) \right. \\ &\quad \left. - \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( \exp\left(-\frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}_N(u)\|^2\right) \right) \right)^2 \mu_0(du). \end{aligned}$$

For the first factor, using the convexity of  $1/x$  on  $(0, \infty)$ , together with Jensen's inequality, we have for all  $u \in X$  the bound

$$\begin{aligned} &\left( \exp\left(-\frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}(u)\|^2\right) + \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( \exp\left(-\frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}_N(u)\|^2\right) \right) \right)^{-1} \\ &\leq \exp\left(-\frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}(u)\|^2\right)^{-1} + \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( \exp\left(-\frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}_N(u)\|^2\right) \right)^{-1} \\ &\leq \exp\left(\frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}(u)\|^2\right) + \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( \exp\left(\frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}_N(u)\|^2\right) \right) \\ &\leq \exp\left(\frac{1}{2\sigma_\eta^2} \sup_{u \in X} \|y - \mathcal{G}(u)\|^2\right) + \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( \exp\left(\frac{1}{2\sigma_\eta^2} \sup_{u \in X} \|y - \mathcal{G}_N(u)\|^2\right) \right). \end{aligned}$$

As in the proof of Lemma 4.7, it then follows by Fernique's Theorem that the right-hand side can be bounded by a constant independent of  $U$  and  $N$ .

For the second factor in the bound on  $\frac{Z}{2}I$ , the linearity of expectation, the local Lipschitz continuity of the exponential function, the equality  $a^2 - b^2 = (a-b)(a+b)$ , the reverse triangle inequality, and Hölder's inequality with conjugate exponents  $p = (1 + \delta)/\delta$  and  $q = 1 + \delta$  give

$$\begin{aligned}
& \int_X \left( \exp \left( -\frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}(u)\|^2 \right) - \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( \exp \left( -\frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}_N(u)\|^2 \right) \right) \right)^2 \mu_0(du) \\
&= \int_X \left( \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( \exp \left( -\frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}(u)\|^2 \right) \right) - \exp \left( -\frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}_N(u)\|^2 \right) \right)^2 \mu_0(du) \\
&\leq 2 \int_X \left( \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( \left| \frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}(u)\|^2 - \frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}_N(u)\|^2 \right| \right) \right)^2 \mu_0(du) \\
&\leq \frac{2}{4\sigma_\eta^4} \int_X \left( \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( (\|y - \mathcal{G}(u)\| + \|y - \mathcal{G}_N(u)\|) \|\mathcal{G}(u) - \mathcal{G}_N(u)\| \right) \right)^2 \mu_0(du) \\
&\leq \frac{2}{4\sigma_\eta^4} \int_X \left( \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( (\|y - \mathcal{G}(u)\| + \|y - \mathcal{G}_N(u)\|)^{(1+\delta)/\delta} \right) \right)^{2\delta/(1+\delta)} \\
&\quad \times \left( \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( \|\mathcal{G}(u) - \mathcal{G}_N(u)\|^{1+\delta} \right) \right)^{2/(1+\delta)} \mu_0(du) \\
&\leq \frac{2}{4\sigma_\eta^4} \sup_{u \in X} \left( \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( (\|y - \mathcal{G}(u)\| + \|y - \mathcal{G}_N(u)\|)^{(1+\delta)/\delta} \right) \right)^{2\delta/(1+\delta)} \\
&\quad \times \int_X \left( \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( \|\mathcal{G}(u) - \mathcal{G}_N(u)\|^{1+\delta} \right) \right)^{2/(1+\delta)} \mu_0(du),
\end{aligned}$$

for any  $\delta > 0$ . The supremum in the above expression can be bounded by a constant independent of  $U$  and  $N$  by Fernique's Theorem as in the proof of Lemma 4.7, since  $\sup_{u \in X} \|\mathcal{G}(u)\| \leq C_G < \infty$ . It follows that there exists a constant  $C$  independent of  $U$  and  $N$  such that

$$\frac{Z}{2}I \leq C \left\| \left( \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( \|\mathcal{G}_N - \mathcal{G}\|^{1+\delta} \right) \right)^{1/(1+\delta)} \right\|_{L^2_{\mu_0}(X)}^2.$$

For the second term in the bound on the Hellinger distance, we have

$$\begin{aligned}
\frac{1}{2\mathbb{E}_{\nu_N^{\mathcal{G}}}(Z_{N,\mathcal{G}}^{\text{sample}})} II &= \left( Z^{-1/2} - (\mathbb{E}_{\nu_N^{\mathcal{G}}}(Z_{N,\mathcal{G}}^{\text{sample}}))^{-1/2} \right)^2 \\
&\leq \max(Z^{-3}, (\mathbb{E}_{\nu_N^{\mathcal{G}}}(Z_{N,\mathcal{G}}^{\text{sample}}))^{-3}) |Z - \mathbb{E}_{\nu_N^{\mathcal{G}}}(Z_{N,\mathcal{G}}^{\text{sample}})|^2.
\end{aligned}$$

Using the linearity of expectation, Tonelli's Theorem and Jensen's inequality, we have

$$\begin{aligned}
& \left| Z - \mathbb{E}_{\nu_N^{\mathcal{G}}}(Z_{N,\mathcal{G}}^{\text{sample}}) \right|^2 \\
&= \left| \int_X \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( \exp \left( -\frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}(u)\|^2 \right) - \exp \left( -\frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}_N(u)\|^2 \right) \right) \mu_0(du) \right|^2 \\
&\leq \int_X \left( \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( \exp \left( -\frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}(u)\|^2 \right) - \exp \left( -\frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}_N(u)\|^2 \right) \right) \right)^2 \mu_0(du),
\end{aligned}$$

which can now be bounded as before. The first claim of the theorem now follows by Lemma 4.7.

The proof for  $\mu_{\text{marginal}}^{y,N,\Phi}$  is similar. We use an identical corresponding splitting of the Hellinger distance  $d_{\text{Hell}}(\mu^y, \mu_{\text{marginal}}^{y,N,\Phi}) \leq I + II$ . For the first term, we have

$$\begin{aligned} \frac{Z}{2}I &= \int_X \left( \sqrt{\exp(-\Phi(u))} - \sqrt{\mathbb{E}_{\nu_N^\Phi}(\exp(-\Phi_N(u)))} \right)^2 \mu_0(du) \\ &\leq \sup_{u \in X} \left( \exp(-\Phi(u)) + \mathbb{E}_{\nu_N^\Phi}(\exp(-\Phi_N(u))) \right)^{-1} \\ &\quad \times \int_X \left( \exp(-\Phi(u)) - \mathbb{E}_{\nu_N^\Phi}(\exp(-\Phi_N(u))) \right)^2 \mu_0(du). \end{aligned}$$

The first factor can again be bounded using Jensen's inequality,

$$\begin{aligned} &\sup_{u \in X} \left( \exp(-\Phi(u)) + \mathbb{E}_{\nu_N^\Phi}(\exp(-\Phi_N(u))) \right)^{-1} \\ &\leq \exp\left(\sup_{u \in X} \Phi(u)\right) + \mathbb{E}_{\nu_N^\Phi}\left(\exp\left(\sup_{u \in X} \Phi_N(u)\right)\right), \end{aligned}$$

which as in the proof of Lemma 4.7, can be bounded by a constant independent of  $U$  and  $N$  by Fernique's Theorem. For the second factor in the bound on  $\frac{Z}{2}I$ , the linearity of expectation, Hölder's inequality with conjugate exponents  $p = (1+\delta)/\delta$  and  $q = 1 + \delta$ , and the inequality  $|\exp(x) - \exp(y)| \leq (\exp(x) + \exp(y))|x - y|$ , for all  $x, y \in \mathbb{R}$ , give

$$\begin{aligned} &\int_X \left( \exp(-\Phi(u)) - \mathbb{E}_{\nu_N^\Phi}(\exp(-\Phi_N(u))) \right)^2 \mu_0(du) \\ &= \int_X \left( \mathbb{E}_{\nu_N^\Phi}(\exp(-\Phi(u)) - \exp(-\Phi_N(u))) \right)^2 \mu_0(du) \\ &\leq \int_X \left( \mathbb{E}_{\nu_N^\Phi} \left( (\exp(-\Phi(u)) + \exp(-\Phi_N(u)))^{(1+\delta)/\delta} \right)^{\delta/(1+\delta)} \right. \\ &\quad \left. \times \mathbb{E}_{\nu_N^\Phi} \left( |\Phi(u) - \Phi_N(u)|^{1+\delta} \right)^{1/(1+\delta)} \right)^2 \mu_0(du) \\ &\leq \sup_{u \in X} \mathbb{E}_{\nu_N^\Phi} \left( (\exp(-\Phi(u)) + \exp(-\Phi_N(u)))^{(1+\delta)/\delta} \right)^{2\delta/(1+\delta)} \\ &\quad \times \left\| \mathbb{E}_{\nu_N^\Phi} (|\Phi(u) - \Phi_N(u)|^{1+\delta})^{1/(1+\delta)} \right\|_{L_{\mu_0}^2(X)}^2, \end{aligned}$$

where the first factor can be bounded independently of  $U$  and  $N$  as in Lemma 4.7.

For the second term in the bound on the Hellinger distance, the linearity of expectation, Tonelli's Theorem and Jensen's inequality give

$$\left| Z - \mathbb{E}_{\nu_N^\Phi}(Z_{N,\Phi}^{\text{sample}}) \right|^2 \leq \int_X \left( \mathbb{E}_{\nu_N^\Phi}(\exp(-\Phi(u)) - \exp(-\Phi_N(u))) \right)^2 \mu_0(du),$$

which can now be bounded as before. The second claim of the theorem then follows by Lemma 4.7.  $\square$

Similarly to Theorem 4.2, Theorem 4.9 provides error bounds for general Gaussian process emulators of  $\mathcal{G}$  and  $\Phi$ . An example of a Gaussian process emulator that satisfies the assumptions of Theorem 4.9 is the emulator defined by (3.6), however, other choices are possible. As in Corollary 4.3, we can now combine Assumption 2.2,

Theorem 4.9, and Proposition 3.3 with  $\beta = 0$  to derive error bounds in terms of the fill distance.

**Corollary 4.10.** *Suppose  $\mathcal{G}_N$  and  $\Phi_N$  are defined as in (3.6), with Matérn kernel  $k = k_{\nu, \lambda, \sigma_k^2}$ . Suppose Assumption 2.2 holds with  $s = \nu + K/2$ , and the assumptions of Proposition 3.3 and Theorem 4.9 are satisfied. Then there exist constants  $C_1, C_2, C_3$ , and  $C_4$ , independent of  $U$  and  $N$ , such that*

$$d_{\text{Hell}}(\mu^y, \mu_{\text{marginal}}^{y, N, \mathcal{G}}) \leq C_1 h_U^{\nu+K/2} + C_2 h_U^\nu$$

and

$$d_{\text{Hell}}(\mu^y, \mu_{\text{marginal}}^{y, N, \Phi}) \leq C_3 h_U^{\nu+K/2} + C_4 h_U^\nu.$$

*Proof.* We give the proof for  $\mu_{\text{marginal}}^{y, N, \mathcal{G}}$ ; the proof for  $\mu_{\text{marginal}}^{y, N, \Phi}$  is similar. Using Theorem 4.9, Jensen's inequality and the triangle inequality, we have

$$\begin{aligned} d_{\text{Hell}}(\mu^y, \mu_{\text{marginal}}^{y, N, \mathcal{G}})^2 &\leq C \left\| \left( \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( \|\mathcal{G} - \mathcal{G}_N\|^2 \right) \right)^{1/2} \right\|_{L_{\mu_0}^2(X)}^2 \\ &= C \int_X \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( \|\mathcal{G}(u) - \mathcal{G}_N(u)\|^2 \right) \mu_0(du) \\ &\leq 2C \int_X \|\mathcal{G}(u) - m_N^{\mathcal{G}}(u)\|^2 \mu_0(du) \\ &\quad + 2C \int_X \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( \|m_N^{\mathcal{G}}(u) - \mathcal{G}_N(u)\|^2 \right) \mu_0(du). \end{aligned}$$

The first term can be bounded by using Assumption 2.2, Proposition 3.2, and Proposition 3.3,

$$\begin{aligned} \int_X \|\mathcal{G}(u) - m_N^{\mathcal{G}}(u)\|^2 \mu_0(du) &= \int_X \sum_{j=1}^J (\mathcal{G}^j(u) - m_N^{\mathcal{G}^j}(u))^2 \mu_0(du) \\ &\leq C h_U^{2\nu+K} \sum_{j=1}^J \|\mathcal{G}^j\|_{H^{\nu+K/2}(X)}^2, \end{aligned}$$

for a constant  $C$  independent of  $U$  and  $N$ . The second term can be bounded by using Assumption 2.2, Propositions 3.2, 3.3, 3.5, the linearity of expectation, and the Sobolev Embedding Theorem

$$\begin{aligned} \int_X \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( \|m_N^{\mathcal{G}}(u) - \mathcal{G}_N(u)\|^2 \right) \mu_0(du) &= \int_X \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( \sum_{j=1}^J (m_N^{\mathcal{G}^j}(u) - \mathcal{G}_N^j(u))^2 \right) \mu_0(du) \\ &= J \int_X k_N(u, u) \mu_0(du) \\ &\leq J \sup_{u \in X} \sup_{\|g\|_{H_k}=1} |g(u) - m_N^g(u)|^2 \\ &\leq C h_U^{2\nu}, \end{aligned}$$

for a constant  $C$  independent of  $U$  and  $N$ . The claim of the corollary then follows.  $\square$

If Assumption 2.2 holds only for some  $s < \nu + K/2$ , an analogue of Corollary 4.10 can be proved using Proposition 3.4 with  $\beta = 0$ .

Note that the term  $h_U^\nu$  appearing in the bounds in Corollary 4.10 corresponds to the error bound on  $\|k_N^{1/2}\|_{L^2(X)}$ , which does not appear in the error bounds for  $\mu_{\text{mean}}^{y,N,\mathcal{G}}$  and  $\mu_{\text{marginal}}^{y,N,\Phi}$  analysed in Corollary 4.3. Due to the supremum over  $g$  appearing in the expression for  $k_N(u, u)$  in Proposition 3.5, we can only conclude on the lower rate of convergence  $h_U^\nu$  for  $\|k_N^{1/2}\|_{L^2(X)}$ . This result appears to be sharp, and the lower rate of convergence  $\nu$  is observed in some of the numerical experiments in section 5 (cf. Figures 3 and 4).

4.2.3. *Error in the random approximations  $\mu_{\text{sample}}^{y,N,\mathcal{G}}$  and  $\mu_{\text{sample}}^{y,N,\Phi}$ .* We have the following result for the random approximations  $\mu_{\text{sample}}^{y,N,\mathcal{G}}$  and  $\mu_{\text{sample}}^{y,N,\Phi}$ .

**Theorem 4.11.** *Under the Assumptions of Lemma 4.7, there exist constants  $C_1$  and  $C_2$ , independent of  $U$  and  $N$ , such that for any  $\delta > 0$ ,*

$$\begin{aligned} \left( \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( d_{\text{Hell}}(\mu^y, \mu_{\text{sample}}^{y,N,\mathcal{G}})^2 \right) \right)^{1/2} &\leq C_1 \left\| \left( \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( \|\mathcal{G} - \mathcal{G}_N\|^{2+\delta} \right) \right)^{1/(2+\delta)} \right\|_{L_{\mu_0}^2(X)}, \\ \left( \mathbb{E}_{\nu_N^{\Phi}} \left( d_{\text{Hell}}(\mu^y, \mu_{\text{sample}}^{y,N,\Phi})^2 \right) \right)^{1/2} &\leq C_2 \left\| \left( \mathbb{E}_{\nu_N^{\Phi}} \left( |\Phi - \Phi_N|^{2+\delta} \right) \right)^{1/(2+\delta)} \right\|_{L_{\mu_0}^2(X)}. \end{aligned}$$

*Proof.* We start with  $\mu_{\text{sample}}^{y,N,\mathcal{G}}$ . By the definition of the Hellinger distance and the linearity of expectation, we have

$$\begin{aligned} \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( 2 d_{\text{Hell}}(\mu^y, \mu_{\text{sample}}^{y,N,\mathcal{G}})^2 \right) &= \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( \int_X \left( \sqrt{\frac{d\mu^y}{d\mu_0}} - \sqrt{\frac{d\mu_{\text{sample}}^{y,N,\mathcal{G}}}{d\mu_0}} \right)^2 \mu_0(du) \right) \\ &\leq \frac{2}{Z} \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( \int_X \left( \exp\left(-\frac{1}{4\sigma_\eta^2} \|y - \mathcal{G}(u)\|^2\right) - \exp\left(-\frac{1}{4\sigma_\eta^2} \|y - \mathcal{G}_N(u)\|^2\right) \right)^2 \mu_0(du) \right) \\ &\quad + 2 \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( |Z_{N,\mathcal{G}}^{\text{sample}}| Z^{-1/2} - (Z_{N,\mathcal{G}}^{\text{sample}})^{-1/2} |^2 \right) \\ &=: I + II. \end{aligned}$$

For the first term, Tonelli's Theorem, the local Lipschitz continuity of the exponential function, the equality  $a^2 - b^2 = (a-b)(a+b)$ , the reverse triangle inequality, and Hölder's inequality with conjugate exponents  $p = (1+\delta)/\delta$  and  $q = 1+\delta$  give

$$\begin{aligned} \frac{Z}{2} I &= \int_X \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( \left( \exp\left(-\frac{1}{4\sigma_\eta^2} \|y - \mathcal{G}(u)\|^2\right) - \exp\left(-\frac{1}{4\sigma_\eta^2} \|y - \mathcal{G}_N(u)\|^2\right) \right)^2 \right) \mu_0(du) \\ &\leq \frac{1}{\sigma_\eta^2} \int_X \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( (\|y - \mathcal{G}(u)\|^2 - \|y - \mathcal{G}_N(u)\|^2)^2 \right) \mu_0(du) \\ &\leq \frac{1}{\sigma_\eta^2} \int_X \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( (\|y - \mathcal{G}_N(u)\| + \|y - \mathcal{G}(u)\|)^2 \|\mathcal{G}_N(u) - \mathcal{G}(u)\|^2 \right) \mu_0(du) \\ &\leq \frac{1}{\sigma_\eta^2} \int_X \left( \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( (\|y - \mathcal{G}(u)\| + \|y - \mathcal{G}_N(u)\|)^{2(1+\delta)/\delta} \right) \right)^{\delta/(\delta+1)} \\ &\quad \times \left( \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( \|\mathcal{G}(u) - \mathcal{G}_N(u)\|^{2(1+\delta)} \right) \right)^{1/(1+\delta)} \mu_0(du) \end{aligned}$$

$$\leq \frac{1}{\sigma_\eta^2} \sup_{u \in X} \left( \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( (\|y - \mathcal{G}(u)\| + \|y - \mathcal{G}_N(u)\|)^{2(1+\delta)/\delta} \right) \right)^{\delta/(\delta+1)} \\ \times \int_X \left( \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( \|\mathcal{G}(u) - \mathcal{G}_N(u)\|^{2(1+\delta)} \right) \right)^{1/(1+\delta)} \mu_0(du).$$

for any  $\delta > 0$ . The supremum in the above bound can be bounded independently of  $U$  and  $N$  by Fernique's Theorem as in the proof of Lemma 4.7. It follows that there exists a constant  $C$  independent of  $U$  and  $N$  such that

$$\frac{Z}{2} I \leq C \left\| \left( \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( \|\mathcal{G}_N - \mathcal{G}\|^{2(1+\delta)} \right) \right)^{1/2(1+\delta)} \right\|_{L_{\mu_0}^2(X)}^2.$$

For the second term in the bound on the Hellinger distance, we have

$$\frac{1}{2} II = \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( |Z^{-1/2} - (Z_{N,\mathcal{G}}^{\text{sample}})^{-1/2}|^2 \right) \\ \leq \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( Z_{N,\mathcal{G}}^{\text{sample}} \max(Z^{-3}, (Z_{N,\mathcal{G}}^{\text{sample}})^{-3}) |Z - Z_{N,\mathcal{G}}^{\text{sample}}|^2 \right).$$

By Jensen's inequality and the same argument as above, we have

$$|Z - Z_{N,\mathcal{G}}^{\text{sample}}|^2 = \left| \int_X \left( \exp \left( -\frac{1}{4\sigma_\eta^2} \|y - \mathcal{G}(u)\|^2 \right) \right. \right. \\ \left. \left. - \exp \left( -\frac{1}{4\sigma_\eta^2} \|y - \mathcal{G}_N(u)\|^2 \right) \right) \mu_0(du) \right|^2 \\ \leq \frac{1}{\sigma_\eta^4} \int_X (\|y - \mathcal{G}_N(u)\| + \|y - \mathcal{G}(u)\|)^2 \|\mathcal{G}_N(u) - \mathcal{G}(u)\|^2 \mu_0(du).$$

Together with Tonelli's Theorem and Hölder's inequality with conjugate exponents  $p = (1 + \delta)/\delta$  and  $q = 1 + \delta$ , we then have

$$\frac{1}{2} II \leq \frac{1}{\sigma_\eta^4} \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( Z_{N,\mathcal{G}}^{\text{sample}} \max(Z^{-3}, (Z_{N,\mathcal{G}}^{\text{sample}})^{-3}) \right. \\ \left. \times \int_X (\|y - \mathcal{G}_N(u)\| + \|y - \mathcal{G}(u)\|)^2 \|\mathcal{G}_N(u) - \mathcal{G}(u)\|^2 \mu_0(du) \right) \\ = \frac{1}{\sigma_\eta^4} \int_X \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( Z_{N,\mathcal{G}}^{\text{sample}} \max(Z^{-3}, (Z_{N,\mathcal{G}}^{\text{sample}})^{-3}) \right. \\ \left. \times (\|y - \mathcal{G}_N(u)\| + \|y - \mathcal{G}(u)\|)^2 \|\mathcal{G}_N(u) - \mathcal{G}(u)\|^2 \mu_0(du) \right) \\ \leq \frac{1}{\sigma_\eta^4} \sup_{u \in X} \left( \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( (Z_{N,\mathcal{G}}^{\text{sample}})^{(1+\delta)/\delta} \max(Z^{-3}, (Z_{N,\mathcal{G}}^{\text{sample}})^{-3})^{(1+\delta)/\delta} \right. \right. \\ \left. \left. \times (\|y - \mathcal{G}(u)\| + \|y - \mathcal{G}_N(u)\|)^{2(1+\delta)/\delta} \right) \right)^{\delta/(\delta+1)} \\ \int_X \left( \mathbb{E}_{\nu_N^{\mathcal{G}}} \left( \|\mathcal{G}_N(u) - \mathcal{G}(u)\|^{2(1+\delta)} \right) \right)^{1/(1+\delta)} \mu_0(du),$$

for any  $\delta > 0$ . The supremum in the bound above can be bounded independently of  $U$  and  $N$  by Lemma 4.7 and Fernique's Theorem. The first claim of the theorem then follows.

The proof for  $\mu_{\text{sample}}^{y,N,\Phi}$  is similar. Using an identical corresponding splitting of the Hellinger distance  $\mathbb{E}_{\nu_N^{\Phi}} \left( 2 d_{\text{Hell}}^2(\mu^y, \mu_{\text{sample}}^{y,N,\Phi}) \right) \leq I + II$ , we bound the first term by

Tonelli's Theorem, Hölder's inequality with conjugate exponents  $p = (1 + \delta)/\delta$  and  $q = 1 + \delta$ , and the inequality  $|\exp(x) - \exp(y)| \leq (\exp(x) + \exp(y))|x - y|$ , for all  $x, y \in \mathbb{R}$ ,

$$\begin{aligned}
 \frac{Z}{2}I &= \int_X \mathbb{E}_{\nu_N^\Phi} \left( (\exp(-\Phi(u)/2) - \exp(-\Phi_N(u)/2))^2 \right) \mu_0(du) \\
 &\leq \int_X \mathbb{E}_{\nu_N^\Phi} \left( |\exp(-\Phi(u)) + \exp(-\Phi_N(u))|^2 |\Phi(u) - \Phi_N(u)|^2 \right) \mu_0(du) \\
 &\leq \int_X \mathbb{E}_{\nu_N^\Phi} \left( |\exp(-\Phi(u)) + \exp(-\Phi_N(u))|^{2(1+\delta)/\delta} \right)^{\delta/(1+\delta)} \\
 &\quad \times \mathbb{E}_{\nu_N^\Phi} \left( |\Phi(u) - \Phi_N(u)|^{2(1+\delta)} \right)^{1/(1+\delta)} \mu_0(du) \\
 &\leq \sup_{u \in X} \mathbb{E}_{\nu_N^\Phi} \left( (\exp(-\Phi(u)) + \exp(-\Phi_N(u)))^{2(1+\delta)/\delta} \right)^{\delta/(1+\delta)} \\
 &\quad \times \left\| \mathbb{E}_{\nu_N^\Phi} \left( |\Phi(u) - \Phi_N(u)|^{2(1+\delta)} \right)^{1/(2(1+\delta))} \right\|_{L^2_{\mu_0}(X)}^2,
 \end{aligned}$$

where the first factor can be bounded independently of  $U$  and  $N$  as in Lemma 4.7. For the second term, we have as before

$$\frac{1}{2}II \leq \mathbb{E}_{\nu_N^\Phi} \left( Z_{N,\Phi}^{\text{sample}} \max(Z^{-3}, (Z_{N,\Phi}^{\text{sample}})^{-3}) |Z - Z_{N,\Phi}^{\text{sample}}|^2 \right)$$

and

$$\begin{aligned}
 |Z - Z_{N,\Phi}^{\text{sample}}|^2 &= \left| \int_X (\exp(-\Phi(u)) - \exp(-\Phi_N(u))) \mu_0(du) \right|^2 \\
 &\leq \int_X (\exp(-\Phi(u)) + \exp(-\Phi_N(u)))^2 (\Phi(u) - \Phi_N(u))^2 \mu_0(du).
 \end{aligned}$$

Together with Tonelli's Theorem and Hölder's inequality with conjugate exponents  $p = (1 + \delta)/\delta$  and  $q = 1 + \delta$ , we then have

$$\begin{aligned}
 \frac{1}{2}II &\leq \mathbb{E}_{\nu_N^\Phi} \left( Z_{N,\Phi}^{\text{sample}} \max(Z^{-3}, (Z_{N,\Phi}^{\text{sample}})^{-3}) \right. \\
 &\quad \times \left. \int_X (\exp(-\Phi(u)) + \exp(-\Phi_N(u)))^2 (\Phi(u) - \Phi_N(u))^2 \mu_0(du) \right) \\
 &= \int_X \mathbb{E}_{\nu_N^\Phi} \left( Z_{N,\Phi}^{\text{sample}} \max(Z^{-3}, (Z_{N,\Phi}^{\text{sample}})^{-3}) \right. \\
 &\quad \times \left. (\exp(-\Phi(u)) + \exp(-\Phi_N(u)))^2 (\Phi(u) - \Phi_N(u))^2 \right) \mu_0(du) \\
 &\leq \left( \mathbb{E}_{\nu_N^\Phi} \left( (Z_{N,\Phi}^{\text{sample}})^{(1+\delta)/\delta} \max(Z^{-3}, (Z_{N,\Phi}^{\text{sample}})^{-3})^{(1+\delta)/\delta} \right. \right. \\
 &\quad \times \left. \left. \sup_{u \in X} \left( \exp(-\Phi(u)) + \exp(-\Phi_N(u)) \right)^{2(1+\delta)/\delta} \right) \right)^{\delta/(\delta+1)} \\
 &\quad \times \int_X \left( \mathbb{E}_{\nu_N^\Phi} \left( \|\Phi(u) - \Phi_N(u)\|^{2(1+\delta)} \right) \right)^{1/(1+\delta)} \mu_0(du),
 \end{aligned}$$



for any  $\delta > 0$ . The first expected value in the bound above can be bounded independently of  $U$  and  $N$  by Lemma 4.7. The second claim of the theorem then follows.  $\square$

Similarly to Theorems 4.2 and 4.9, Theorem 4.11 provides error bounds for general Gaussian process emulators of  $\mathcal{G}$  and  $\Phi$ . As a particular example, we can take the emulators defined by (3.6). We can now combine Assumption 2.2, Theorem 4.11, and Proposition 3.3 with  $\beta = 0$  to derive error bounds in terms of the fill distance.

**Corollary 4.12.** *Suppose  $\mathcal{G}_N$  and  $\Phi_N$  are defined as in (3.6), with Matérn kernel  $k = k_{\nu, \lambda, \sigma_k^2}$ . Suppose Assumption 2.2 holds with  $s = \nu + K/2$ , and the assumptions of Proposition 3.3 and Theorem 4.11 are satisfied. Then there exist constants  $C_1, C_2, C_3$ , and  $C_4$ , independent of  $U$  and  $N$ , such that*

$$d_{\text{Hell}}(\mu^y, \mu_{\text{marginal}}^{y, N, \mathcal{G}}) \leq C_1 h_U^{\nu+K/2} + C_2 h_U^\nu$$

and

$$d_{\text{Hell}}(\mu^y, \mu_{\text{marginal}}^{y, N, \Phi}) \leq C_3 h_U^{\nu+K/2} + C_4 h_U^\nu.$$

*Proof.* The proof is similar to that of Corollary 4.10, exploiting that for a Gaussian random variable  $X$ , we have  $\mathbb{E}((X - \mathbb{E}(X))^4) = 3(\mathbb{E}((X - \mathbb{E}(X))^2))^2$ .  $\square$

If Assumption 2.2 holds only for some  $s < \nu + K/2$ , an analogue of Corollary 4.12 can be proved using Proposition 3.4 with  $\beta = 0$ .

Furthermore we have the following result on a generalised total variation distance [33], defined by

$$d_{\text{gTV}}(\mu^y, \mu_{\text{sample}}^{y, N, \mathcal{G}}) = \sup_{\|f\|_{C^0(X)} \leq 1} \left( \mathbb{E}_{\nu_N^{\mathcal{G}}}(|\mathbb{E}_{\mu^y}(f) - \mathbb{E}_{\mu_{\text{sample}}^{y, N, \mathcal{G}}}(f)|^2) \right)^{1/2},$$

for  $\mu_{\text{sample}}^{y, N, \mathcal{G}}$ , and defined analogously for  $\mu_{\text{sample}}^{y, N, \Phi}$ .

**Theorem 4.13.** *Under the Assumptions of Lemma 4.7, there exist constants  $C_1$  and  $C_2$ , independent of  $U$  and  $N$ , such that for any  $\delta > 0$*

$$d_{\text{gTV}}(\mu^y, \mu_{\text{sample}}^{y, N, \mathcal{G}}) \leq C_1 \left\| \left( \mathbb{E}_{\nu_N^{\mathcal{G}}}(\|\mathcal{G} - \mathcal{G}_N\|^{2+\delta}) \right)^{1/(2+\delta)} \right\|_{L_{\mu_0}^2(X)},$$

$$d_{\text{gTV}}(\mu^y, \mu_{\text{sample}}^{y, N, \Phi}) \leq C_2 \left\| \left( \mathbb{E}_{\nu_N^{\Phi}}(\|\Phi - \Phi_N\|^{2+\delta}) \right)^{1/(2+\delta)} \right\|_{L_{\mu_0}^2(X)}.$$

*Proof.* We give the proof for  $\mu_{\text{sample}}^{y,N,\Phi}$ ; the proof for  $\mu_{\text{sample}}^{y,N,\mathcal{G}}$  is identical. By definition, we have

$$\begin{aligned}
 d_{\text{gTV}}(\mu^y, \mu_{\text{sample}}^{y,N,\Phi}) &= \sup_{\|f\|_{C^0(X)} \leq 1} \left( \mathbb{E}_{\nu_N^\Phi} (|\mathbb{E}_{\mu^y}(f) - \mathbb{E}_{\mu_{\text{sample}}^{y,N,\Phi}}(f)|^2) \right)^{1/2} \\
 &= \sup_{\|f\|_{C^0(X)} \leq 1} \left( \mathbb{E}_{\nu_N^\Phi} \left( \left| \int_X f(u) (\exp(-\Phi(u)) Z^{-1} \right. \right. \right. \\
 &\quad \left. \left. \left. - \exp(-\Phi_N(u)) (Z_{N,\Phi}^{\text{sample}})^{-1} \right) \mu_0(du) \right|^2 \right) \right)^{1/2} \\
 &\leq \left( \mathbb{E}_{\nu_N^\Phi} \left( \left| \int_X (\exp(-\Phi(u)) Z^{-1} - \exp(-\Phi_N(u)) (Z_{N,\Phi}^{\text{sample}})^{-1}) \mu_0(du) \right|^2 \right) \right)^{1/2} \\
 &\leq \frac{2}{Z} \left( \mathbb{E}_{\nu_N^\Phi} \left( \int_X |\exp(-\Phi(u)) - \exp(-\Phi_N(u))|^2 \mu_0(du) \right) \right)^{1/2} \\
 &\quad + \left( \mathbb{E}_{\nu_N^\Phi} \left( (Z_{N,\Phi}^{\text{sample}})^2 |Z^{-1} - (Z_{N,\Phi}^{\text{sample}})^{-1}|^2 \right) \right)^{1/2} \\
 &=: I + II.
 \end{aligned}$$

The terms  $I$  and  $II$  can be bounded by the same arguments as the terms  $I$  and  $II$  in the proof of Theorem 4.11, by noting that  $|Z^{-1} - (Z_{N,\Phi}^{\text{sample}})^{-1}|^2 \leq \max(Z^{-4}, (Z_{N,\Phi}^{\text{sample}})^{-4}) |Z - Z_{N,\Phi}^{\text{sample}}|^2$ .  $\square$

## 5. NUMERICAL EXAMPLES

We consider the model inverse problem of determining the diffusion coefficient of an elliptic partial differential equation (PDE) in divergence form from observation of a finite set of noisy continuous functionals of the solution. This type of equation arises, for example, in the modelling of groundwater flow in a porous medium. We consider the one-dimensional model problem

$$(5.1) \quad -\frac{d}{dx} \left( \kappa(x; u) \frac{dp}{dx}(x; u) \right) = 1 \quad \text{in } (0, 1), \quad p(1; u) = p(0; u) = 0,$$

where the coefficient  $\kappa$  depends on parameters  $u = \{u_j\}_{j=1}^K \in [-1, 1]^K$  through the linear expansion

$$\kappa(x; u) = \frac{1}{100} + \sum_{j=1}^K \frac{u_j}{200(K+1)} \sin(2\pi jx).$$

In this setting the forward map  $G : [-1, 1]^K \rightarrow H_0^1(D)$ , defined by  $G(u) = p$ , is an analytic function [10]. Since the observation operator  $\mathcal{O}$  is linear and bounded, Assumption 2.2 is satisfied for any  $s > K/2$ .

Unless stated otherwise, throughout this section we will approximate the solution  $p$  by standard, piecewise linear, continuous finite elements on a uniform grid with mesh size  $h = 1/32$ . The corresponding approximate forward map, denoted by  $G_h$ , is also an analytic function of  $u$  [10], and Assumption 2.2 is satisfied for any  $s > K/2$  also for  $G_h$ . By a slight abuse of notation, we will denote the posterior measure corresponding to the forward map  $G_h$  by  $\mu^y$ , and use this as our reference measure. The error induced by the finite element approximation will be ignored.

As prior measure  $\mu_0$  on  $[-1, 1]^K$ , we use the uniform product measure  $\mu_0(du) = \bigotimes_{j=1}^K \frac{du_j}{2}$ . The observations  $y$  are taken as noisy point evaluations of the solution,  $y_j = p(x_j; u^*) + \eta_j$  with  $\eta \sim \mathcal{N}(0, I)$  and  $\{x_j\}_{j=1}^J$  evenly spaced points in  $(0, 1)$ . To generate  $y$ , the truth  $u^*$  was chosen as a random sample from the prior, and the solution  $p$  was approximated by finite elements on a uniform grid with mesh size  $h^* = 1/1024$ .

The emulators  $\mathcal{G}_N$  and  $\Phi_N$  are computed as described in section 3.2, with mean and covariance kernel given by (3.6). In the Gaussian process prior (3.1), we choose  $m \equiv 0$  and  $k = k_{\nu,1,1}$ , a Matérn kernel with variance  $\sigma_k^2 = 1$ , correlation length  $\lambda = 1$  and smoothness parameter  $\nu$ .

For a given approximation  $\mu^{y,N}$  to  $\mu^y$ , we will compute twice the Hellinger distance squared:

$$2d_{\text{Hell}}(\mu^y, \mu^{y,N})^2 = \int_{[-1,1]^K} \left( \sqrt{\frac{d\mu^y}{d\mu_0}}(u) - \sqrt{\frac{d\mu^{y,N}}{d\mu_0}}(u) \right)^2 d\mu_0(u).$$

The integral over  $[-1, 1]^K$  is approximated by a randomly shifted lattice rule with product weight parameters  $\gamma_j = 1/j^2$  [29]. The generating vector for the rule used is available from Frances Kuo's website (<http://web.maths.unsw.edu.au/~fkuo/>) as "lattice-39102-1024-1048576.3600". For the marginal and random approximations, the expected value over the Gaussian process is approximated by Monte Carlo sampling, using the MATLAB command `mvnrnd`.

For the design points  $U$ , we choose a uniform tensor grid. In  $[-1, 1]^K$ , the uniform tensor grid consisting of  $N = N_*^K$  points, for some  $N_* \in \mathbb{N}$ , has fill distance  $h_U = \sqrt{K}(N_* - 1)^{-1}$ . In Table 1, we give the convergence rates in  $N$  for  $\sup_{u \in X} |f(u) - m_N^f(u)|^2$  and  $\|f - m_N^f\|_{L^2(X)}^2$  predicted by Proposition 3.3.

TABLE 1. Convergence rates in  $N$  predicted by Proposition 3.3 for uniform tensor grids.

$\sup_{u \in X}  f(u) - m_N^f(u) ^2$					$\ f - m_N^f\ _{L^2(X)}^2$				
$\nu$	$K$				$\nu$	$K$			
	1	2	3	4		1	2	3	4
1	2	1	0.67	0.5	1	3	2	1.7	1.5
5		5	3.3		5		6	4.3	

TABLE 2. Observed convergence rates in  $N$  of  $d_{\text{Hell}}(\mu^y, \mu^{y,N,\mathcal{G}})^2$ , as shown in Figures 1, 3 and 4.

$\mu_{\text{mean}}^{y,N,\mathcal{G}}$			$\mu_{\text{marginal}}^{y,N,\mathcal{G}}$			$\mu_{\text{sample}}^{y,N,\mathcal{G}}$		
$\nu \backslash K$	2	3	$\nu \backslash K$	2	3	$\nu \backslash K$	2	3
1	2.6	2.4	1	2.6	2.2	1	2.3	1.7
5	6.2	4.5	5	6.2	4.6	5	6.1	4.4

TABLE 3. Observed convergence rates in  $N$  of  $d_{\text{Hell}}(\mu^y, \mu^{y,N,\Phi})^2$ , as shown in Figures 1, 3 and 4.

$\mu_{\text{mean}}^{y,N,\Phi}$			$\mu_{\text{marginal}}^{y,N,\Phi}$			$\mu_{\text{sample}}^{y,N,\Phi}$		
$\nu \backslash K$	2	3	$\nu \backslash K$	2	3	$\nu \backslash K$	2	3
1	2.5	2	1	1.8	1.1	1	1.1	0.76
5	5.4	3.8	5	4.9	3.2	5	4.9	3.3

TABLE 4. Observed convergence rates in  $N$  of  $d_{\text{Hell}}(\mu^y, \mu^{y,N,\mathcal{G}})^2$  and  $d_{\text{Hell}}(\mu^y, \mu^{y,N,\Phi})^2$ , as shown in Figure 2.

$\mu_{\text{mean}}^{y,N,\mathcal{G}}$					$\mu_{\text{mean}}^{y,N,\Phi}$				
$\nu \backslash K$	1	2	3	4	$\nu \backslash K$	1	2	3	4
1	4.1	2.7	2.3	2.3	1	4	2.7	2.1	1.9

**5.1. Mean-based approximations.** In Figure 1, we show  $2d_{\text{Hell}}(\mu^y, \mu_{\text{mean}}^{y,N,\mathcal{G}})^2$  (left) and  $2d_{\text{Hell}}(\mu^y, \mu_{\text{mean}}^{y,N,\Phi})^2$  (right), for a variety of choices of  $K$  and  $\nu$ , for  $J = 1$ . For each choice of the parameters  $K$  and  $\nu$ , we have as a dotted line added the least squares fit of the form  $C_1 N^{-C_2}$ , for some  $C_1, C_2 > 0$ , and indicated the rate  $N^{-C_2}$  in the legend. The observed rates  $C_2$  are also summarised in Tables 2 and 3. By Corollary 4.3, we expect to see the faster convergence rates in the right panel of Table 1. For convenience, we have added these rates in parentheses in the legends in Figure 1. For  $\mu_{\text{mean}}^{y,N,\mathcal{G}}$ , we observe the rates in Table 1, or slightly faster. For  $\mu_{\text{mean}}^{y,N,\Phi}$ , we observe rates slightly faster than predicted for  $\nu = 1$ , and slightly slower than predicted for  $\nu = 5$ . Finally, we remark that though the convergence rates of the error are slightly slower for  $\mu_{\text{mean}}^{y,N,\Phi}$ , the actual errors are smaller for  $\mu_{\text{mean}}^{y,N,\Phi}$ .

In Figure 2, we again show  $2d_{\text{Hell}}(\mu^y, \mu_{\text{mean}}^{y,N,\mathcal{G}})^2$  (left) and  $2d_{\text{Hell}}(\mu^y, \mu_{\text{mean}}^{y,N,\Phi})^2$  (right), for a variety of choices of  $K$ , with  $J = 15$  and  $\nu = 1$ . The observed convergence rates are summarised in Table 4. We again observe convergence rates slightly faster than the rates predicted in the right panel of Table 1. As in Figure 1, we observe that the errors in  $\mu_{\text{mean}}^{y,N,\Phi}$  are smaller, though the rates of convergence are slightly faster for  $\mu_{\text{mean}}^{y,N,\mathcal{G}}$ .

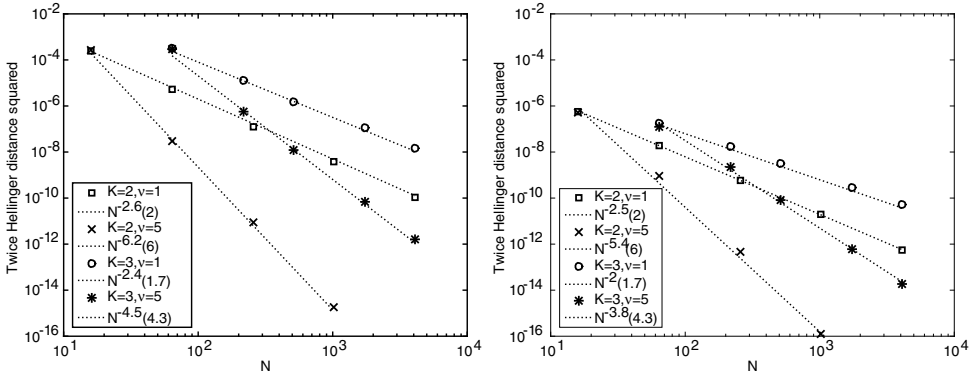


FIGURE 1.  $2d_{\text{Hell}}(\mu^y, \mu_{\text{mean}}^{y,N,\mathcal{G}})^2$  and  $2d_{\text{Hell}}(\mu^y, \mu_{\text{mean}}^{y,N,\Phi})^2$  (left and right, respectively), for a variety of choices of  $K$  and  $\nu$ , for  $J = 1$ .

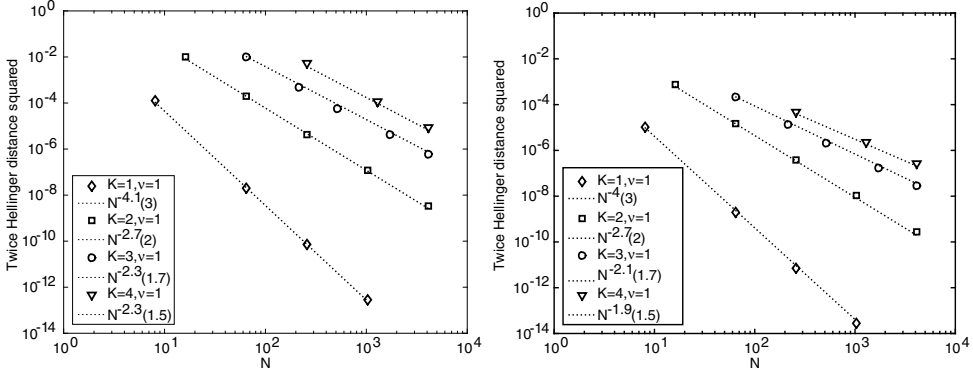


FIGURE 2.  $2d_{\text{Hell}}(\mu^y, \mu_{\text{mean}}^{y,N,\mathcal{G}})^2$  and  $2d_{\text{Hell}}(\mu^y, \mu_{\text{mean}}^{y,N,\Phi})^2$  (left and right, respectively), for a variety of choices of  $K$  and  $\nu = 1$ , for  $J = 15$ .

**5.2. Marginal approximations.** In Figure 3, we show  $2d_{\text{Hell}}(\mu^y, \mu_{\text{marginal}}^{y,N,\mathcal{G}})^2$  (left) and  $2d_{\text{Hell}}(\mu^y, \mu_{\text{marginal}}^{y,N,\Phi})^2$  (right), for a variety of choices of  $K$  and  $\nu$ , for  $J = 1$ . For each choice of the parameters  $K$  and  $\nu$ , we have again added the least squares fit of the form  $C_1 N^{-C_2}$ , and indicated the rate  $C_2$  in the legend. The observed rates  $C_2$  are also summarised in Tables 2 and 3. By Corollary 4.10, we expect the error to be the sum of two contributions, one of which decays at the rate indicated in the left panel of Table 1, and another which decays at the rate indicated by the right panel of Table 1. For convenience, we have added these rates in parentheses in the legends in Figure 3. For  $\mu_{\text{marginal}}^{y,N,\mathcal{G}}$ , we observe the faster convergence rates in the right panel of Table 1, although a closer inspection indicates that the convergence is slowing down as  $N$  increases. For  $\mu_{\text{marginal}}^{y,N,\mathcal{G}}$ , the observed rates are somewhere between the two rates predicted by Table 1.

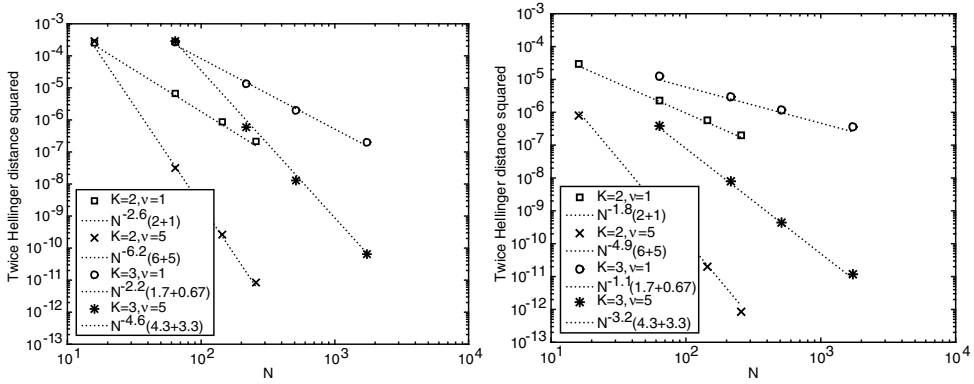


FIGURE 3.  $2d_{\text{Hell}}(\mu^y, \mu_{\text{marginal}}^{y,N,\mathcal{G}})^2$  and  $2d_{\text{Hell}}(\mu^y, \mu_{\text{marginal}}^{y,N,\Phi})^2$  (left and right, respectively), for a variety of choices of  $K$  and  $\nu$ , for  $J = 1$ .

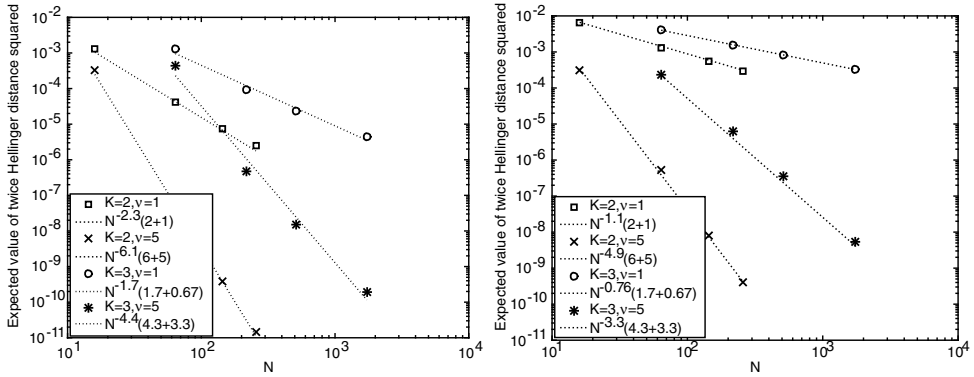


FIGURE 4.  $2\mathbb{E}_{\nu^{\mathcal{G}}}(d_{\text{Hell}}(\mu^y, \mu_{\text{sample}}^{y,N,\mathcal{G}})^2)$  and  $2\mathbb{E}_{\nu^{\Phi}}(d_{\text{Hell}}(\mu^y, \mu_{\text{sample}}^{y,N,\Phi})^2)$  (left and right, respectively), for a variety of choices of  $K$  and  $\nu$ , for  $J = 1$ .

**5.3. Random approximations.** In Figure 4, we show  $2\mathbb{E}_{\nu^{\mathcal{G}}}(d_{\text{Hell}}(\mu^y, \mu_{\text{sample}}^{y,N,\mathcal{G}})^2)$  (left) and  $2\mathbb{E}_{\nu^{\Phi}}(d_{\text{Hell}}(\mu^y, \mu_{\text{sample}}^{y,N,\Phi})^2)$  (right), for a variety of choices of  $K$  and  $\nu$ , for  $J = 1$ . For each choice of the parameters  $K$  and  $\nu$ , we have again added the least squares fit of the form  $C_1 N^{-C_2}$ , and indicated the rate  $C_2$  in the legend. The observed rates  $C_2$  are also summarised in Tables 2 and 3. By Corollary 4.12, we expect the error to be the sum of two contributions, as for the marginal approximations considered in the previous section, and the corresponding rates from Table 1 have been added in parentheses in the legends. For  $\mu_{\text{sample}}^{y,N,\mathcal{G}}$ , we again observe the faster convergence rates in the right panel of Table 1, but the convergence again seems to be slowing down as  $N$  increases. For  $\mu_{\text{marginal}}^{y,N,\mathcal{G}}$ , the observed rates are very close to the slower rates in the left panel of Table 1.

## 6. CONCLUSIONS AND FURTHER WORK

Gaussian process emulators are frequently used as surrogate models. In this work, we analysed the error that is introduced in the Bayesian posterior distribution when a Gaussian process emulator is used to approximate the forward model, either in terms of the parameter-to-observation map or the negative log-likelihood. We showed that the error in the posterior distribution, measured in the Hellinger distance, can be bounded in terms of the error in the emulator, measured in a norm dependent on the approximation considered.

An issue that requires further consideration is the efficient emulation of vector-valued functions. A simple solution, employed in this work, is to emulate each entry independently. In many applications, however, it is natural to assume that the entries are correlated, and a better emulator could be constructed by including this correlation in the emulator. Furthermore, there are still a lot of open questions about how to do this optimally [6]. Also the question of scaling the Gaussian process methodology to high dimensional input spaces remains open. The current error bounds from scattered data approximation employed in this paper feature a strong dependence on the input dimension  $K$ , yielding poor convergence estimates in high dimensions.

Another important issue is the selection of the design points used to construct the Gaussian process emulator, also known as experimental design. In applications where the posterior distribution concentrates with respect to the prior, it might be more efficient to choose design points that are somehow adapted to the posterior measure instead of space-filling designs that have a small fill distance. For example, we could use the sequential designs in [39]. It would be interesting to prove suitable error bounds in this case, maybe using ideas from [44].

In practical applications of Gaussian process emulators, such as in [19], the derivation of the emulator is often more involved than the simple approach presented in section 3. The hyper-parameters in the covariance kernel of the emulator are often unknown, and there is often a discrepancy between the mathematical model of the forward map and the true physical process, known as model error. These are both important issues for which the assumptions in our error bounds have not yet been verified.

## REFERENCES

- [1] R. J. Adler, *The Geometry of Random Fields*, John Wiley & Sons, Ltd., Chichester, 1981. Wiley Series in Probability and Mathematical Statistics. MR611857
- [2] C. Andrieu and G. O. Roberts, *The pseudo-marginal approach for efficient Monte Carlo computations*, Ann. Statist. **37** (2009), no. 2, 697–725, DOI 10.1214/07-AOS574. MR2502648
- [3] N. Aronszajn, *Theory of reproducing kernels*, Trans. Amer. Math. Soc. **68** (1950), 337–404. MR0051437
- [4] S. R. Arridge, J. P. Kaipio, V. Kolehmainen, M. Schweiger, E. Somersalo, T. Tarvainen, and M. Vauhkonen, *Approximation errors and model reduction with an application in optical diffusion tomography*, Inverse Problems **22** (2006), no. 1, 175–195, DOI 10.1088/0266-5611/22/1/010. MR2194190
- [5] I. Babuška, F. Nobile, and R. Tempone, *A stochastic collocation method for elliptic partial differential equations with random input data*, SIAM Rev. **52** (2010), no. 2, 317–355, DOI 10.1137/100786356. MR2646806
- [6] I. Bilonis, N. Zabaras, B. A. Konomi, and G. Lin, *Multi-output separable Gaussian process: Towards an efficient, fully Bayesian paradigm for uncertainty quantification*, Journal of Computational Physics **241** (2013), 212–239.

- [7] N. Bliznyuk, D. Ruppert, C. Shoemaker, R. Regis, S. Wild, and P. Mugunthan, *Bayesian calibration and uncertainty analysis for computationally expensive models using optimization and radial basis function approximation*, J. Comput. Graph. Statist. **17** (2008), no. 2, 270–294, DOI 10.1198/106186008X320681. MR2439960
- [8] V. I. Bogachev, *Gaussian Measures*, Mathematical Surveys and Monographs, vol. 62, American Mathematical Society, Providence, RI, 1998. MR1642391
- [9] T. Bui-Thanh, K. Willcox, and O. Ghattas, *Model reduction for large-scale systems with high-dimensional parametric input space*, SIAM J. Sci. Comput. **30** (2008), no. 6, 3270–3288, DOI 10.1137/070694855. MR2452388
- [10] A. Cohen, R. Devore, and C. Schwab, *Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDE's*, Anal. Appl. (Singap.) **9** (2011), no. 1, 11–47, DOI 10.1142/S0219530511001728. MR2763359
- [11] P. R. Conrad, Y. M. Marzouk, N. S. Pillai, and A. Smith, *Asymptotically exact MCMC algorithms via local approximations of computationally intensive models*, J. Amer. Statist. Assoc. **111** (2016), 1591–1607.
- [12] P. G. Constantine, *Active Subspaces: Emerging Ideas for Dimension Reduction in Parameter Studies*, SIAM Spotlights, vol. 2, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2015. MR3486165
- [13] P. G. Constantine, E. Dow, and Q. Wang, *Active subspace methods in theory and practice: applications to kriging surfaces*, SIAM J. Sci. Comput. **36** (2014), no. 4, A1500–A1524, DOI 10.1137/130916138. MR3233940
- [14] S. L. Cotter, G. O. Roberts, A. M. Stuart, and D. White, *MCMC methods for functions: modifying old algorithms to make them faster*, Statist. Sci. **28** (2013), no. 3, 424–446, DOI 10.1214/13-STS421. MR3135540
- [15] M. Dashti and A.M.Stuart, *The Bayesian Approach to Inverse Problems*, Handbook of Uncertainty Quantification (R. Ghanem, D. Higdon, and H. Owhadi, eds.), Springer.
- [16] M. Girolami and B. Calderhead, *Riemann manifold Langevin and Hamiltonian Monte Carlo methods*, J. R. Stat. Soc. Ser. B Stat. Methodol. **73** (2011), no. 2, 123–214, DOI 10.1111/j.1467-9868.2010.00765.x. MR2814492
- [17] M. Hansen and C. Schwab, *Sparse adaptive approximation of high dimensional parametric initial value problems*, Vietnam J. Math. **41** (2013), no. 2, 181–215, DOI 10.1007/s10013-013-0011-9. MR3089816
- [18] W. K. Hastings, *Monte Carlo sampling methods using Markov chains and their applications*, Biometrika **57** (1970), no. 1, 97–109, DOI 10.1093/biomet/57.1.97. MR3363437
- [19] D. Higdon, M. Kennedy, J. C. Cavendish, J. A. Cafo, and R. D. Ryne, *Combining field data and computer simulations for calibration and prediction*, SIAM J. Sci. Comput. **26** (2004), no. 2, 448–466, DOI 10.1137/S1064827503426693. MR2116355
- [20] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, Applied Mathematical Sciences, vol. 160, Springer-Verlag, New York, 2005. MR2102218
- [21] M. C. Kennedy and A. O'Hagan, *Bayesian calibration of computer models*, J. R. Stat. Soc. Ser. B Stat. Methodol. **63** (2001), no. 3, 425–464, DOI 10.1111/1467-9868.00294. MR1858398
- [22] Y. Marzouk and D. Xiu, *A stochastic collocation approach to Bayesian inference in inverse problems*, Commun. Comput. Phys. **6** (2009), no. 4, 826–847, DOI 10.4208/cicp.2009.v6.p826. MR2672325
- [23] Y. M. Marzouk, H. N. Najm, and L. A. Rahn, *Stochastic spectral methods for efficient Bayesian solution of inverse problems*, J. Comput. Phys. **224** (2007), no. 2, 560–586, DOI 10.1016/j.jcp.2006.10.010. MR2330284
- [24] B. Matérn, *Spatial variation*, 2nd ed., Lecture Notes in Statistics, vol. 36, Springer-Verlag, Berlin, 1986. MR867886
- [25] J. Mercer, *Functions of positive and negative type, and their connection with the theory of integral equations*, Philosophical Transactions of the Royal Society of London, Series A **209** (1909), 415–446.
- [26] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *Equation of state calculations by fast computing machines*, J. Chemical Physics **21** (1953), 1087.
- [27] F. J. Narcowich, J. D. Ward, and H. Wendland, *Sobolev bounds on functions with scattered zeros, with applications to radial basis function surface fitting*, Math. Comp. **74** (2005), no. 250, 743–763, DOI 10.1090/S0025-5718-04-01708-9. MR2114646



- [28] F. J. Narcowich, J. D. Ward, and H. Wendland, *Sobolev error estimates and a Bernstein inequality for scattered data interpolation via radial basis functions*, Constr. Approx. **24** (2006), no. 2, 175–186, DOI 10.1007/s00365-005-0624-7. MR2239119
- [29] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*, CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 63, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992. MR1172997
- [30] A. O'Hagan, *Bayesian analysis of computer code outputs: a tutorial*, Reliability Engineering & System Safety **91** (2006), no. 10, 1290–1300.
- [31] G. Da Prato and J. Zabczyk, *Stochastic Equations in Infinite Dimensions*, Encyclopedia of Mathematics and its Applications, vol. 44, Cambridge University Press, Cambridge, 1992. MR1207136
- [32] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, Adaptive Computation and Machine Learning, MIT Press, Cambridge, MA, 2006. MR2514435
- [33] P. Rebeschini and R. van Handel, *Can local particle filters beat the curse of dimensionality?*, Ann. Appl. Probab. **25** (2015), no. 5, 2809–2866, DOI 10.1214/14-AAP1061. MR3375889
- [34] C. P. Robert and G. Casella, *Monte Carlo statistical methods*, Springer Texts in Statistics, Springer-Verlag, New York, 1999. MR1707311
- [35] W. Rudin, *Principles of Mathematical Analysis*, Second edition, McGraw-Hill Book Co., New York, 1964. MR0166310
- [36] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, *Design and analysis of computer experiments*, Statist. Sci. **4** (1989), no. 4, 409–435. MR1041765
- [37] M. Scheuerer, R. Schaback, and M. Schlather, *Interpolation of spatial data—a stochastic or a deterministic problem?*, European J. Appl. Math. **24** (2013), no. 4, 601–629, DOI 10.1017/S0956792513000016. MR3082868
- [38] Cl. Schillings and Ch. Schwab, *Sparsity in Bayesian inversion of parametric operator equations*, Inverse Problems **30** (2014), no. 6, 065007, 30, DOI 10.1088/0266-5611/30/6/065007. MR3224127
- [39] M. Sinsbeck and W. Nowak, *Sequential design of computer experiments for the solution of Bayesian inverse problems with process emulators*, SIAM/ASA Journal on Uncertainty Quantification, to appear
- [40] M. L. Stein, *Interpolation of Spatial Data: Some Theory for Kriging*, Springer Series in Statistics, Springer-Verlag, New York, 1999. MR1697409
- [41] A. M. Stuart, *Inverse problems: a Bayesian perspective*, Acta Numer. **19** (2010), 451–559, DOI 10.1017/S0962492910000061. MR2652785
- [42] W. Walter, *Ordinary differential equations*, Graduate Texts in Mathematics, vol. 182, Springer-Verlag, New York, 1998. MR1629775
- [43] H. Wendland, *Scattered Data Approximation*, Cambridge Monographs on Applied and Computational Mathematics, vol. 17, Cambridge University Press, Cambridge, 2005. MR2131724
- [44] Z. M. Wu and R. Schaback, *Local error estimates for radial basis function interpolation of scattered data*, IMA J. Numer. Anal. **13** (1993), no. 1, 13–27, DOI 10.1093/imanum/13.1.13. MR1199027
- [45] D. Xiu and G. E. Karniadakis, *Modeling uncertainty in flow simulations via generalized polynomial chaos*, J. Comput. Phys. **187** (2003), no. 1, 137–167, DOI 10.1016/S0021-9991(03)00092-5. MR1977783

MATHEMATICS INSTITUTE, ZEEMAN BUILDING, UNIVERSITY OF WARWICK, COVENTRY, CV4 7AL, ENGLAND

*Current address:* Computing and Mathematical Sciences, Caltech, Pasadena, California 91125

*E-mail address:* [astuart@caltech.edu](mailto:astuart@caltech.edu)

MATHEMATICS INSTITUTE, ZEEMAN BUILDING, UNIVERSITY OF WARWICK, COVENTRY, CV4 7AL, ENGLAND

*Current address:* School of Mathematics, James Clerk Maxwell Building, University of Edinburgh, EH9 3FD, Edinburgh, Scotland

*E-mail address:* [a.teckentrup@ed.ac.uk](mailto:a.teckentrup@ed.ac.uk)