

A Framework for Controlling Sources of Inaccuracy in Gaussian Process Emulation of Deterministic Computer Experiments

Benjamin Haaland^a, Wenjia Wang^a, Vaibhav Maheshwari^b

^aGeorgia Institute of Technology, School of Industrial and Systems Engineering. Emails: bhaaland3@gatech.edu, wenjiawang@gatech.edu

^bRenal Research Institute. Email: vaibhav.maheshwari@rriny.com

Abstract

Computer experiments have become ubiquitous in science and engineering. Commonly, runs of these simulations demand considerable time and computing, making experimental design extremely important in gaining high quality information with limited time and resources. Broad principles of experimental design are proposed and justified which ensure high nominal, numeric, and parameter estimation accuracy for Gaussian process emulation of deterministic simulations. The space-filling properties “small fill distance” and “large separation distance” are only weakly conflicting and ensure well-controlled nominal, numeric, and parameter estimation error. Non-stationarity indicates a greater density of experimental inputs in regions of the input space with more quickly decaying correlation, while non-constant regression functions indicate a balancing of traditional design features with space-fillingness. This work provides robust, rigorously justified, and practically useful overarching sufficient principles for scientists and engineers selecting combinations of simulation inputs with high information content.

Keywords: Computer Experiment, Emulation, Experimental Design, Interpolation, Gaussian Process, Reproducing Kernel Hilbert Space.

1 Background

Computer experiments are complex mathematical models implemented in large computer codes used by scientists and engineers to study real systems. In many situations, the potential for actual experimentation could be very limited. For example, a computational fluid dynamics simulation could be used to compare outflow rates of various sclera flap geometries in trabeculectomy [40], mosquito population dynamics could be coupled with dengue transmission models to study urban dengue control [13], or a finite-volume radiation hydrodynamics model could be calibrated to a small physical data set, allowing exploration of input configurations for which experimental data is not available [20]. Often, a thorough exploration of the unknown simulation function or mean simulation function is wanted. However, the simulation is typically expensive enough that this exploration must be conducted very wisely. A seemingly high-quality solution is to evaluate the expensive simulation at several well-distributed data sites and then build an inexpensive approximation, or emulator, for the simulation. The accuracy of this emulator depends very strongly on the manner in which data is collected from the expensive function [35, 14, 17].

Here, we develop principles of *data collection* for Gaussian process emulation of deterministic computer experiments which are broadly applicable and rigorously justified. Three sources of inaccuracy will be considered, nominal error, numeric error, and emulation error

due to parameter estimation. Stationary and non-stationary situations, as well as regression functions, will be considered. Here, the goal is not to develop optimal experimental designs, but instead to identify features of experimental designs which *simultaneously* have high-quality nominal, numeric, and parameter estimation properties across a broad spectrum of specific situations. Several design criteria are available for computer experiments, including maximin, minimax, low-dimensional stratification, discrepancy, minimum integrated mean squared prediction error, maximum entropy, orthogonality, and others [22]. Here, we proceed by decomposing the emulation error into nominal, numeric, and parameter estimation components, then bound each source of error in turn, in terms of design properties. For the stationary covariance with constant (or null) mean situation, the developed conditions are well-aligned with the distance-based maximin and minimax criteria. Non-stationarity indicates more design locations in regions with more rapidly decaying correlation, while regression functions introduce a balancing of traditional design properties with the distance-based criteria. To the best of our knowledge, this is the first presentation of sufficient conditions for simultaneously controlling nominal, numeric, and parameter estimation error in Gaussian process emulation under stationarity, non-stationarity, and the presence of regression functions.

2 Preliminaries

Let $f : \Omega \rightarrow \mathbb{R}$, denote the function linking a computer experiment's input to its output for $\Omega \subset \mathbb{R}^d$. Ordinarily, the approximation \hat{f} to the unknown function f depends on several parameters $\boldsymbol{\vartheta}$. Here, let $\hat{f}_{\boldsymbol{\vartheta}}$ denote the *nominal* emulator at a particular value of the parameters $\boldsymbol{\vartheta}$, and $\tilde{f}_{\boldsymbol{\vartheta}}$ denote the *numeric* emulator at a particular value of the parameters $\boldsymbol{\vartheta}$. The numeric emulator represents the emulator which is calculated using floating point arithmetic, while the nominal emulator represents the idealized, exact arithmetic, version thereof. Then, for any norm $\|\cdot\|$, particular value of the parameters $\boldsymbol{\vartheta}_*$, and corresponding parameter estimate $\hat{\boldsymbol{\vartheta}}$, the normed deviation of the emulator from the computer experiment can be decomposed into *nominal*, *numeric*, and *parameter estimation* components using the triangle inequality as shown below, extending ideas in [18] to also consider inaccuracy due to parameter estimation.

$$\begin{aligned} \|f - \tilde{f}_{\hat{\boldsymbol{\vartheta}}}\| &= \|f - \hat{f}_{\boldsymbol{\vartheta}_*} + \hat{f}_{\boldsymbol{\vartheta}_*} - \hat{f}_{\hat{\boldsymbol{\vartheta}}} + \hat{f}_{\hat{\boldsymbol{\vartheta}}} - \tilde{f}_{\hat{\boldsymbol{\vartheta}}}\| \\ &\leq \underbrace{\|f - \hat{f}_{\boldsymbol{\vartheta}_*}\|}_{\text{nominal}} + \underbrace{\|\hat{f}_{\boldsymbol{\vartheta}_*} - \hat{f}_{\hat{\boldsymbol{\vartheta}}}\|}_{\text{numeric}} + \underbrace{\|\hat{f}_{\hat{\boldsymbol{\vartheta}}} - \tilde{f}_{\hat{\boldsymbol{\vartheta}}}\|}_{\text{parameter}}. \end{aligned} \quad (2.1)$$

Note that inequality (2.1) does not make any assumption about the norm or type of emulator used. It is also noteworthy that this error decomposition considers numeric error in evaluation of the interpolator but does not consider numeric error in the parameter estimation process. Explicit consideration of numeric error in parameter estimation would result in a fourth term in the error decomposition.

Consider the L_2 norm on the domain of interest Ω

$$\|g\| = \|g\|_{L_2(\Omega)} = \sqrt{\int_{\Omega} g(\mathbf{x})^2 d\mathbf{x}}. \quad (2.2)$$

For the L_2 norm (2.2) and any expectation \mathbb{E} we have

$$\mathbb{E}\|g\| = \mathbb{E}\sqrt{\int_{\Omega} g(\mathbf{x})^2 d\mathbf{x}} \leq \sqrt{\int_{\Omega} \mathbb{E}g(\mathbf{x})^2 d\mathbf{x}}, \quad (2.3)$$

by Jensen's inequality and Tonelli's theorem [6]. Applying relation (2.3) to the error decomposition (2.1), gives

$$\begin{aligned} \mathbb{E}\|f - \tilde{f}_{\hat{\boldsymbol{\vartheta}}}\| &\leq \mathbb{E}\|f - \hat{f}_{\boldsymbol{\vartheta}_*}\| + \mathbb{E}\|\hat{f}_{\hat{\boldsymbol{\vartheta}}} - \tilde{f}_{\hat{\boldsymbol{\vartheta}}}\| + \mathbb{E}\|\hat{f}_{\boldsymbol{\vartheta}_*} - \hat{f}_{\hat{\boldsymbol{\vartheta}}}\| \\ &\leq \sqrt{\int_{\Omega} \mathbb{E}\{f(\mathbf{x}) - \hat{f}_{\boldsymbol{\vartheta}_*}(\mathbf{x})\}^2 d\mathbf{x}} + \sqrt{\int_{\Omega} \mathbb{E}\{\hat{f}_{\hat{\boldsymbol{\vartheta}}}(\mathbf{x}) - \tilde{f}_{\hat{\boldsymbol{\vartheta}}}(\mathbf{x})\}^2 d\mathbf{x}} \\ &\quad + \sqrt{\int_{\Omega} \mathbb{E}\{\hat{f}_{\boldsymbol{\vartheta}_*}(\mathbf{x}) - \hat{f}_{\hat{\boldsymbol{\vartheta}}}(\mathbf{x})\}^2 d\mathbf{x}}. \end{aligned} \quad (2.4)$$

More generally, convexity of the function $t \mapsto |t|^p$ for $1 \leq p \leq \infty$ ensures that versions of inequalities (2.3) and (2.4) hold for the $L_p(\Omega)$ norms $1 \leq p \leq \infty$. Here, the *loss function*, or measure of inaccuracy, will be the left-hand-side of equation (2.4), $\mathbb{E}\|f - \tilde{f}_{\hat{\boldsymbol{\vartheta}}}\|$, for the expectation conditional on the *data*, the experimental design and corresponding output values, $(\mathbf{x}_i, f(\mathbf{x}_i))$, $i = 1, \dots, n$.

Throughout, consider a Gaussian Process (GP) model for interpolation,

$$f \sim \text{GP}(\mathbf{h}(\cdot)'\boldsymbol{\beta}, \Psi_{\boldsymbol{\theta}}(\cdot, \cdot))$$

for some fixed, known regression functions $\mathbf{h}(\cdot)$, and let $\boldsymbol{\vartheta} = (\boldsymbol{\beta}' \ \boldsymbol{\theta}')'$. The *particular* GP draw f is the *truth* one hopes to discover. It is assumed that $\Psi_{\boldsymbol{\theta}}(\cdot, \cdot)$ is a positive definite function [43]. A considerable collection of theoretical development related to GP regression is given in [37]. For a particular dataset $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ and input of interest \mathbf{x} , the *best linear unbiased predictor* (BLUP) is

$$\hat{f}_{\boldsymbol{\vartheta}}(\mathbf{x}) = \mathbf{h}(\mathbf{x})'\hat{\boldsymbol{\beta}} + \Psi_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{X})\Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} \left(f(\mathbf{X}) - \mathbf{H}(\mathbf{X})\hat{\boldsymbol{\beta}} \right), \quad (2.5)$$

where $\Psi_{\boldsymbol{\theta}}(\mathbf{A}, \mathbf{B}) = \{\Psi_{\boldsymbol{\theta}}(\mathbf{a}_i, \mathbf{b}_j)\}$ and $f(\mathbf{A}) = \{f(\mathbf{a}_i)\}$ for $\mathbf{A} = \{\mathbf{a}_i\}$ and $\mathbf{B} = \{\mathbf{b}_j\}$, $\mathbf{H}(\mathbf{X})$ has rows $\mathbf{h}(\mathbf{x}_i)'$, $\hat{\boldsymbol{\beta}} = (\mathbf{H}(\mathbf{X})'\Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1}\mathbf{H}(\mathbf{X}))^{-1} \mathbf{H}(\mathbf{X})'\Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1}f(\mathbf{X})$, and $\boldsymbol{\theta}$ equals the vector of true correlation parameters [34]. The BLUP is *best* in the sense that it minimizes the MSPE, $\mathbb{E}\{f(\mathbf{x}) - \hat{f}_{\boldsymbol{\vartheta}_*}(\mathbf{x})\}^2$ (integrated and square rooted in the first term on the right-hand-side of equation (2.4)). While the BLUP cannot actually be computed, due to unknown correlation parameters and floating point arithmetic, the parameters are commonly estimated via maximum likelihood and the BLUP's floating point approximation is taken as the estimate of the unknown function f . Here, the BLUP, as shown in (2.5), will be taken as the *nominal* emulator.

The overall approach will be to provide bounds for each of the three terms on the right-hand side of (2.4) in terms of properties of the experimental design \mathbf{X} . It will be shown that the *nominal*, *numeric*, and *parameter estimation* criteria are only weakly conflicting, and lead to broadly similar experimental designs. The remainder of this article is organized as follows. In Sections 3, 4, and 5, bounds on the nominal, numeric, and parameter estimation error, respectively, are developed. In each section, experimental design characteristics which

lead to small error bounds are discussed and a few examples are given. In Section 6, a few comparisons are presented and the implications of these broad principles are briefly discussed.

3 Nominal Error

Focusing on the first term of (2.4), the *nominal* or mean squared prediction error (MSPE) is given by [34],

$$\begin{aligned} & \mathbb{E} \left\{ f(\mathbf{x}) - \hat{f}_{\theta}(\mathbf{x}) \right\}^2 \\ &= \Psi_{\theta}(\mathbf{x}, \mathbf{x}) - \begin{pmatrix} \mathbf{h}(\mathbf{x})' & \Psi_{\theta}(\mathbf{x}, \mathbf{X}) \end{pmatrix} \begin{pmatrix} \mathbf{0} & \mathbf{H}(\mathbf{X})' \\ \mathbf{H}(\mathbf{X}) & \Psi_{\theta}(\mathbf{X}, \mathbf{X}) \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{h}(\mathbf{x}) \\ \Psi_{\theta}(\mathbf{X}, \mathbf{x}) \end{pmatrix}. \end{aligned} \quad (3.1)$$

Note that throughout this section, the unknown parameters are taken at their true values. This is not an assumption *per se*, but instead a consequence of the error decomposition (2.1). The error due to estimating the parameters is considered separately. In line with intuition, increasing the number of data points always reduces the nominal error. The proof of Proposition 3.1 is provided in Appendix A.

Proposition 3.1. *If $f \sim \text{GP}(\mathbf{h}(\cdot)'\boldsymbol{\beta}, \Psi_{\theta}(\cdot, \cdot))$, for fixed, known regression functions $\mathbf{h}(\cdot)$ and $\mathbf{X}_1 \subseteq \mathbf{X}_2$, then $\text{MSPE}_2 \leq \text{MSPE}_1$, where MSPE_1 and MSPE_2 denote the MSPE of the BLUPs based on \mathbf{X}_1 and \mathbf{X}_2 , respectively.*

Notably, this result states that the *expected* squared prediction error is always reduced by the addition of data. Certainly, for a particular combination of GP draw f and new design location(s), or for a mis-specified model, the *actual* squared prediction error at a location of interest could in fact be increased by the addition of data.

Consider controlling the inner part of the bound on the nominal error given by the first term of (2.4). The MSPE in the inner part of the nominal error is given by (3.1). Applying partitioned matrix inverse results, (3.1) can be rewritten as

$$\begin{aligned} & \mathbb{E} \left\{ f(\mathbf{x}) - \hat{f}_{\theta}(\mathbf{x}) \right\}^2 \\ &= \Psi_{\theta}(\mathbf{x}, \mathbf{x}) - \Psi_{\theta}(\mathbf{x}, \mathbf{X})\Psi_{\theta}(\mathbf{X}, \mathbf{X})^{-1}\Psi_{\theta}(\mathbf{X}, \mathbf{x}) \\ &+ \left(\mathbf{h}(\mathbf{x}) - \mathbf{H}(\mathbf{X})'\Psi_{\theta}(\mathbf{X}, \mathbf{X})^{-1}\Psi_{\theta}(\mathbf{X}, \mathbf{x}) \right)' \left(\mathbf{H}(\mathbf{X})'\Psi_{\theta}(\mathbf{X}, \mathbf{X})^{-1}\mathbf{H}(\mathbf{X}) \right)^{-1} \\ &\times \left(\mathbf{h}(\mathbf{x}) - \mathbf{H}(\mathbf{X})'\Psi_{\theta}(\mathbf{X}, \mathbf{X})^{-1}\Psi_{\theta}(\mathbf{X}, \mathbf{x}) \right). \end{aligned} \quad (3.2)$$

Initially, the uppermost terms in (3.2), which provide the MSPE for a model with mean zero or no regression functions, are bounded. We make use of the below theorem which bounds the uppermost terms of the MSPE (3.2) in terms of local bounds. The proof of Theorem 3.1 is provided in Appendix B.

Theorem 3.1. *If A_i , $i = 1, \dots, n$ is a covering of Ω in the sense that $\Omega \subseteq \cup_{i=1}^n A_i$, Ψ_{θ} is a positive definite function with $\Psi_{\theta}(\mathbf{x}, \mathbf{x}) = \sigma^2$ for all $\mathbf{x} \in \Omega$, and $\mathbf{x}_i \in A_i$ for $\mathbf{x}_i \in \mathbf{X}$, then*

$$\begin{aligned} & \sup_{\mathbf{x} \in \Omega} \Psi_{\theta}(\mathbf{x}, \mathbf{x}) - \Psi_{\theta}(\mathbf{x}, \mathbf{X})\Psi_{\theta}(\mathbf{X}, \mathbf{X})^{-1}\Psi_{\theta}(\mathbf{X}, \mathbf{x}) \\ & \leq \frac{1}{k} \left(\sigma^2 - \min_i \inf_{\mathbf{x} \in A_i} \Psi_{\theta}(\mathbf{x}_i, \mathbf{x}) \right) \left(2k - \sigma^2 + \min_i \inf_{\mathbf{x} \in A_i} \Psi_{\theta}(\mathbf{x}_i, \mathbf{x}) \right), \end{aligned}$$

where $k = n \sup_{\mathbf{u}, \mathbf{v} \in \Omega} \Psi_{\theta}(\mathbf{u}, \mathbf{v})$.

While Theorem 3.1 only assumes Ψ_{θ} is a positive definite function with $\Psi_{\theta}(\mathbf{x}, \mathbf{x}) = \sigma^2$, we examine special two cases of particular interest. The first covers many situations where stationarity is assumed, while the second considers a model of non-stationarity in correlation, adapted from [3]. We will label these cases respectively as the Stationary Model and Non-Stationary Model. While the considered model of non-stationarity is certainly not all inclusive, it forms a quality approximation in many practical situations. In the context of this model of non-stationarity, the correlation consists of more rapidly and more slowly decaying components with their weights differing across the input space. Below, $\varphi(\cdot)$ is taken to be a decreasing function of its non-negative argument.

3.1 Stationary Model

Suppose $\Psi_{\theta}(\mathbf{u}, \mathbf{v}) = \sigma^2 \varphi(\|\Theta(\mathbf{u} - \mathbf{v})\|_2)$. Theorem 3.1 can be used to write the overall bound on the uppermost terms of the MSPE (3.2) in terms of a *Voronoi* covering [2] of Ω with respect to a *Mahalanobis*-like distance [28] $d_{\Theta}(\mathbf{u}, \mathbf{v}) = \|\Theta(\mathbf{u} - \mathbf{v})\|_2$ as

$$\frac{\sigma^2}{k} \left(1 - \varphi \left(\max_i \sup_{\mathbf{x} \in V_i(\Theta)} d_{\Theta}(\mathbf{x}_i, \mathbf{x}) \right) \right) \left(2k - 1 + \varphi \left(\max_i \sup_{\mathbf{x} \in V_i(\Theta)} d_{\Theta}(\mathbf{x}_i, \mathbf{x}) \right) \right), \quad (3.3)$$

where $V_i(\Theta) = \{\mathbf{x} \in \Omega: d_{\Theta}(\mathbf{x}, \mathbf{x}_i) \leq d_{\Theta}(\mathbf{x}, \mathbf{x}_j) \forall j \neq i\}$ and $k = n\varphi(0)$.

Note that

$$\max_i \sup_{\mathbf{x} \in V_i(\Theta)} d_{\Theta}(\mathbf{x}_i, \mathbf{x}) = \sup_{\mathbf{x} \in \Omega} \min_i d_{\Theta}(\mathbf{x}_i, \mathbf{x}), \quad (3.4)$$

is the *fill distance* with respect to the distance d_{Θ} . So, the supremum of the MSPE over possible inputs, for a GP model with mean zero, can be controlled by demanding that the (potentially) non-spherical fill distance (3.4) is small. Further, the upper bound (3.3) is minimized if it is achieved uniformly for $i = 1, \dots, n$. Importantly, a uniform bound on the terms (3.3) is achieved by an experimental design \mathbf{X} for which all the $\sup_{\mathbf{x} \in V_i(\Theta)} d_{\Theta}(\mathbf{x}_i, \mathbf{x})$ are the same. That is, all the Voronoi cells have the same maximum distance with respect to d_{Θ} from their data point to their edge.

3.2 Non-Stationary Model

Suppose $\Psi_{\theta}(\mathbf{u}, \mathbf{v}) = \sigma^2 (\omega_1(\mathbf{u})\omega_1(\mathbf{v})\varphi(\|\Theta_1(\mathbf{u} - \mathbf{v})\|_2) + \omega_2(\mathbf{u})\omega_2(\mathbf{v})\varphi(\|\Theta_2(\mathbf{u} - \mathbf{v})\|_2))$. For the Non-Stationary Model, assume $\omega_1(\cdot), \omega_2(\cdot) \geq 0$ have Lipschitz continuous derivatives on Ω , $\omega_1^2(\cdot) + \omega_2^2(\cdot) = 1$, Θ_1, Θ_2 are non-singular, and $\lambda_{\max}(\Theta_1' \Xi_2' \Xi_2 \Theta_1) < 1$, where $\Xi_2 = \Theta_2^{-1}$. The final assumption can be interpreted as $\varphi(\|\Theta_2(\cdot - \cdot)\|_2)$ is narrower than $\varphi(\|\Theta_1(\cdot - \cdot)\|_2)$. Throughout, we use the notation $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ for the maximum and minimum eigenvalues of their (diagonalizable) arguments. Consider the covering of Ω , $V_i^* = V_i(\Theta_1) \cup V_i(\Theta_2)$, $i = 1, \dots, n$. Note that $V_i(\Theta_1)$ and $V_i(\Theta_2)$ often do not differ strongly. For example, if $\Theta_2 = c\Theta_1$, then $V_i(\Theta_1) = V_i(\Theta_2)$.

Here, take a version of the upper bound in Theorem 3.1 given by inserting a slightly reduced argument into the decreasing for $y \geq 0$ function $g(y) = \frac{\sigma^2}{k} (1 - y)(2k - 1 + y)$. First,

note that

$$\begin{aligned} & \inf_{\mathbf{x} \in V_i^*} \{ \omega_1(\mathbf{x}_i) \omega_1(\mathbf{x}) \varphi(\|\Theta_1(\mathbf{x}_i - \mathbf{x})\|_2) + \omega_2(\mathbf{x}_i) \omega_2(\mathbf{x}) \varphi(\|\Theta_2(\mathbf{x}_i - \mathbf{x})\|_2) \} \\ & \geq \inf_{\mathbf{x} \in V_i^*} \left\{ \omega_1(\mathbf{x}_i) \omega_1(\mathbf{x}) \varphi \left(\sup_{\mathbf{x} \in V_i^*} d_{\Theta_1}(\mathbf{x}_i, \mathbf{x}) \right) + \omega_2(\mathbf{x}_i) \omega_2(\mathbf{x}) \varphi \left(\sup_{\mathbf{x} \in V_i^*} d_{\Theta_2}(\mathbf{x}_i, \mathbf{x}) \right) \right\}. \end{aligned} \quad (3.5)$$

The Lipschitz derivatives of $\omega_1(\cdot)$, $\omega_2(\cdot)$ and Taylor's theorem [32] imply $\omega_1(\mathbf{x}) = \omega_1(\mathbf{x}_i) + R_1(\mathbf{x}, \mathbf{x}_i)$ and $\omega_2(\mathbf{x}) = \omega_2(\mathbf{x}_i) + R_2(\mathbf{x}, \mathbf{x}_i)$, where $|R_1(\mathbf{x}, \mathbf{x}_i)| \leq k_1 \|\mathbf{x}_i - \mathbf{x}\|_2$ and $|R_2(\mathbf{x}, \mathbf{x}_i)| \leq k_2 \|\mathbf{x}_i - \mathbf{x}\|_2$. The bound (3.5) can, in turn, be bounded below as

$$\begin{aligned} & \omega_1^2(\mathbf{x}_i) \varphi \left(\sup_{\mathbf{x} \in V_i^*} d_{\Theta_1}(\mathbf{x}_i, \mathbf{x}) \right) + \omega_2^2(\mathbf{x}_i) \varphi \left(\sup_{\mathbf{x} \in V_i^*} d_{\Theta_2}(\mathbf{x}_i, \mathbf{x}) \right) \\ & - \varphi(0) (k_1 + k_2) \max_i \sup_{\mathbf{x} \in V_i^*} \|\mathbf{x}_i - \mathbf{x}\|_2, \end{aligned} \quad (3.6)$$

where for tractability the final term in (3.6) is bounded uniformly across the design space. Next, consider an experimental design for which the bounds (3.6) are uniform over i . One might expect that regions of the design space with less weight on the global, long range, correlation $\varphi(\|\Theta_1(\cdot - \cdot)\|_2)$ and more weight on the local, short range, correlation $\varphi(\|\Theta_2(\cdot - \cdot)\|_2)$ would require more closely spaced design points, and *vice versa*. Roughly speaking this expectation holds true.

Consider two design points \mathbf{x}_i and \mathbf{x}_j along with corresponding (*union of*) Voronoi cell sizes $\sup_{\mathbf{x} \in V_i^*} d_{\Theta_1}(\mathbf{x}_i, \mathbf{x})$, $\sup_{\mathbf{x} \in V_i^*} d_{\Theta_2}(\mathbf{x}_i, \mathbf{x})$, $\sup_{\mathbf{x} \in V_j^*} d_{\Theta_1}(\mathbf{x}_j, \mathbf{x})$, and $\sup_{\mathbf{x} \in V_j^*} d_{\Theta_2}(\mathbf{x}_j, \mathbf{x})$. Suppose that the points in the input space *near* \mathbf{x}_i have more weight on the global, long range, correlation than the points in the input space *near* \mathbf{x}_j and the points in the input space *near* \mathbf{x}_j have more weight on the local, short range, correlation than the points in the input space *near* \mathbf{x}_i , in the sense that

$$\begin{aligned} & (-1)^{k-1} \omega_k(\mathbf{x}_i)^2 \left(\varphi \left(\sup_{\mathbf{x} \in V_i^*} d_{\Theta_1}(\mathbf{x}_i, \mathbf{x}) \right) - \varphi \left(\sup_{\mathbf{x} \in V_i^*} d_{\Theta_2}(\mathbf{x}_i, \mathbf{x}) \right) \right) \\ & \geq (-1)^{k-1} \omega_k(\mathbf{x}_j)^2 \left(\varphi \left(\sup_{\mathbf{x} \in V_j^*} d_{\Theta_1}(\mathbf{x}_j, \mathbf{x}) \right) - \varphi \left(\sup_{\mathbf{x} \in V_j^*} d_{\Theta_2}(\mathbf{x}_j, \mathbf{x}) \right) \right), \end{aligned} \quad (3.7)$$

for $k = 1, 2$. The $(-1)^{k-1}$ terms just mean that the direction of inequality depends on k .

Uniformity of the bounds (3.6) along with $\omega_1^2(\cdot) + \omega_2^2(\cdot) = 1$ implies

$$\begin{aligned}
& \omega_1(\mathbf{x}_i)^2 \left(\varphi \left(\sup_{\mathbf{x} \in V_i^*} d_{\Theta_1}(\mathbf{x}_i, \mathbf{x}) \right) - \varphi \left(\sup_{\mathbf{x} \in V_i^*} d_{\Theta_2}(\mathbf{x}_i, \mathbf{x}) \right) \right) \\
& - \omega_1(\mathbf{x}_j)^2 \left(\varphi \left(\sup_{\mathbf{x} \in V_j^*} d_{\Theta_1}(\mathbf{x}_j, \mathbf{x}) \right) - \varphi \left(\sup_{\mathbf{x} \in V_j^*} d_{\Theta_2}(\mathbf{x}_j, \mathbf{x}) \right) \right) \\
& = \varphi \left(\sup_{\mathbf{x} \in V_j^*} d_{\Theta_2}(\mathbf{x}_j, \mathbf{x}) \right) - \varphi \left(\sup_{\mathbf{x} \in V_i^*} d_{\Theta_2}(\mathbf{x}_i, \mathbf{x}) \right) \quad \text{and} \\
& \omega_2(\mathbf{x}_j)^2 \left(\varphi \left(\sup_{\mathbf{x} \in V_j^*} d_{\Theta_1}(\mathbf{x}_j, \mathbf{x}) \right) - \varphi \left(\sup_{\mathbf{x} \in V_j^*} d_{\Theta_2}(\mathbf{x}_j, \mathbf{x}) \right) \right) \\
& - \omega_2(\mathbf{x}_i)^2 \left(\varphi \left(\sup_{\mathbf{x} \in V_i^*} d_{\Theta_1}(\mathbf{x}_i, \mathbf{x}) \right) - \varphi \left(\sup_{\mathbf{x} \in V_i^*} d_{\Theta_2}(\mathbf{x}_i, \mathbf{x}) \right) \right) \\
& = \varphi \left(\sup_{\mathbf{x} \in V_j^*} d_{\Theta_1}(\mathbf{x}_j, \mathbf{x}) \right) - \varphi \left(\sup_{\mathbf{x} \in V_i^*} d_{\Theta_1}(\mathbf{x}_i, \mathbf{x}) \right).
\end{aligned} \tag{3.8}$$

Combining (3.7) with (3.8) gives

$$\sup_{\mathbf{x} \in V_j^*} d_{\Theta_1}(\mathbf{x}_j, \mathbf{x}) \leq \sup_{\mathbf{x} \in V_i^*} d_{\Theta_1}(\mathbf{x}_i, \mathbf{x}) \quad \text{and} \quad \sup_{\mathbf{x} \in V_j^*} d_{\Theta_2}(\mathbf{x}_j, \mathbf{x}) \leq \sup_{\mathbf{x} \in V_i^*} d_{\Theta_2}(\mathbf{x}_i, \mathbf{x}),$$

since φ is a decreasing function of its non-negative argument. That is, a uniform bound on (3.6) is achieved by an experimental design \mathbf{X} which has smaller (*union of*) Voronoi cells, with respect to either d_{Θ_1} or d_{Θ_2} , in regions with more emphasis on the local, more quickly decaying, correlation and less emphasis on the global, more slowly decaying, correlation. Note that the global ($k = 1$) and local ($k = 2$) *emphases* at \mathbf{x}_i are given concretely by

$$\omega_k(\mathbf{x}_i)^2 \left(\varphi \left(\sup_{\mathbf{x} \in V_i^*} d_{\Theta_1}(\mathbf{x}_i, \mathbf{x}) \right) - \varphi \left(\sup_{\mathbf{x} \in V_i^*} d_{\Theta_2}(\mathbf{x}_i, \mathbf{x}) \right) \right) \quad k = 1, 2.$$

The assumption that ω_1, ω_2 have Lipschitz continuous derivatives on Ω , while not overly restrictive in most practical situations, is not necessary, in principle. Without this assumption, the bilinear form on the right-hand side of (3.5) can be bounded below via one of several reverses of the Cauchy-Schwarz inequality [11]. Many of these results provide a lower bound for (3.5) in terms of a geometric mean across \mathbf{x} and \mathbf{x}_i of terms similar to (3.6), without the terms involving the Lipschitz constants, and in turn more complex development for the supremum of the uppermost terms of the MSPE (3.2). Balancing simplicity and broad applicability, these type of results are not pursued here.

3.3 Regression Functions

Subsections 3.1 (Stationary Model) and 3.2 (Non-Stationary Model) relate to the uppermost terms in equation (3.2), which without further development provides the MSPE for a model with mean zero. Now, we consider the lowermost terms in (3.2), which are relevant for Gaussian process models with a mean or non-null regression component. The lowermost

terms in (3.2) can be bounded above as

$$\begin{aligned}
& (\mathbf{h}(\mathbf{x}) - \mathbf{H}(\mathbf{X})' \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{x}))' (\mathbf{H}(\mathbf{X})' \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} \mathbf{H}(\mathbf{X}))^{-1} \\
& \quad \times (\mathbf{h}(\mathbf{x}) - \mathbf{H}(\mathbf{X})' \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{x})) \\
& \leq \|\mathbf{h}(\mathbf{x}) - \mathbf{H}(\mathbf{X})' \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{x})\|_2^2 / \lambda_{\min}(\mathbf{H}(\mathbf{X})' \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} \mathbf{H}(\mathbf{X})) \quad (3.9) \\
& \leq n \sup_{\mathbf{u}, \mathbf{v} \in \Omega} \Psi_{\boldsymbol{\theta}}(\mathbf{u}, \mathbf{v}) \|\mathbf{h}(\mathbf{x}) - \mathbf{H}(\mathbf{X})' \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{x})\|_2^2 / \lambda_{\min}(\mathbf{H}(\mathbf{X})' \mathbf{H}(\mathbf{X})).
\end{aligned}$$

The first inequality is true because $\mathbf{a}' \mathbf{B}^{-1} \mathbf{a} \leq \lambda_{\max}(\mathbf{B}^{-1}) \|\mathbf{a}\|_2^2$ and $\lambda_{\max}(\mathbf{B}^{-1}) = 1/\lambda_{\min}(\mathbf{B})$ and the second inequality is true because $\lambda_{\min}(\mathbf{A}' \mathbf{B}^{-1} \mathbf{A}) \geq \lambda_{\min}(\mathbf{A}' \mathbf{A})/\lambda_{\max}(\mathbf{B})$ and as implication (B.2) of Gershgorin's theorem [41]. Note that the term $n \sup_{\mathbf{u}, \mathbf{v} \in \Omega} \Psi_{\boldsymbol{\theta}}(\mathbf{u}, \mathbf{v})$ does not depend on the experimental design.

The components of the squared Euclidean norm $\|\mathbf{h}(\mathbf{x}) - \mathbf{H}(\mathbf{X})' \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{x})\|_2^2$ are squared errors for an interpolator of the regression functions. Intuitively, we might expect these squared interpolation errors to behave in a manner similar to the MSPE for the Gaussian process model with mean zero. In fact, it has been shown above that if the regression functions are draws from a Gaussian process with mean zero and a covariance structure as described in Subsections 3.1 and 3.2, then the expectation of these squared errors can be controlled through the experimental design as described above. That is, an experimental design which gives low nominal error in the mean zero case will also make the term $\|\mathbf{h}(\mathbf{x}) - \mathbf{H}(\mathbf{X})' \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{x})\|_2^2$ small.

Alternatively, a reproducing kernel Hilbert space (RKHS) [1, 43] may be defined as the completion of the function space spanned by $\{\Psi_{\boldsymbol{\theta}}(\mathbf{x}_i, \cdot) : \mathbf{x}_i \in \Omega\}$ with respect to the inner product $\langle \sum_i a_i \Psi_{\boldsymbol{\theta}}(\mathbf{x}_i, \cdot), \sum_j b_j \Psi_{\boldsymbol{\theta}}(\mathbf{y}_j, \cdot) \rangle = \sum_{i,j} a_i b_j \Psi_{\boldsymbol{\theta}}(\mathbf{x}_i, \mathbf{y}_j)$. Many commonly selected regression functions, for example constant, linear, polynomial, and spline, will also lie in the RKHSs induced by many of the common covariance functions. For example, the Gaussian kernel induces an RKHS of functions with infinitely many continuous derivatives and Matérn kernels induce RKHSs of functions with a fixed number of continuous derivatives. If the selected regression functions lie in the RKHS induced by the chosen covariance function, then deterministic RKHS interpolation error bounds as a decreasing function of the fill distance, such as Theorem 5.1 in [18], can be applied. Another alternative would be to choose as regression functions covariance function (half) evaluations $\{\Psi_{\boldsymbol{\theta}}(\mathbf{x}_i, \cdot) : i \in \mathcal{I}\}$ at a well-distributed set of centers \mathcal{I} . These regression functions are capable of *approximating* a broad range of mean functions and have the appealing feature that the lowermost term in the bound (3.2) is then identically zero.

The eigenvalue $\lambda_{\min}(\mathbf{H}(\mathbf{X})' \mathbf{H}(\mathbf{X}))$ has approximation

$$\lambda_{\min}(\mathbf{H}(\mathbf{X})' \mathbf{H}(\mathbf{X})) = \lambda_{\min} \left(\sum_{i=1}^n \mathbf{h}(\mathbf{x}_i) \mathbf{h}(\mathbf{x}_i)' \right) \approx n \lambda_{\min} \left(\int \mathbf{h}(\mathbf{y}) \mathbf{h}(\mathbf{y})' dF(\mathbf{y}) \right) = n s_1,$$

where F denotes the large sample distribution of the input locations \mathbf{X} , $s_1 \geq 0$, and $s_1 > 0$ unless $\mathbf{h}(\mathbf{y})' \mathbf{a} = 0$ with probability 1 with respect to the large sample distribution F for some $\mathbf{a} \neq \mathbf{0}$. The (approximate) term $s_1 = \lambda_{\min}(\int \mathbf{h}(\mathbf{y}) \mathbf{h}(\mathbf{y})' dF(\mathbf{y}))$ in the denominator of (3.9) indicates that (at least for regression functions which do not make $\mathbf{h}(\mathbf{x}) - \mathbf{H}(\mathbf{X})' \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{x}) \equiv \mathbf{0}$), the design properties implied by the mean zero development in Subsections 3.1 and 3.2 need to be balanced with traditional experimental design properties. Two common scenarios are of particular interest. First, consider a

constant regression function, a mean parameter. In this situation, $s_1 = 1$ irrespective of experimental design. Second, consider linear regression functions in each dimension in addition to the constant. If each linear function is expressed on the same scale, s_1 will be large for an experimental design with points far from $\boldsymbol{\mu} = \int \mathbf{y}dF(\mathbf{y})$ the average design value, and whose orientations $\mathbf{x}_i - \boldsymbol{\mu}$ emphasize each basis vector in an orthonormal basis of \mathbb{R}^d equally. For the common situation where $\Omega = [0, 1]^d$, s_1 will be maximized for a design with equal numbers of points in each of the corners of $[0, 1]^d$. So, high-quality experimental designs for Gaussian process models with linear regression mean components will balance the fill distance-based criteria described in Subsections 3.1 and 3.2 with the push of design points to the “corners” of Ω . Similarly, high-quality experimental designs for Gaussian process models with quadratic regression mean components will balance the fill distance-based criteria described in Subsections 3.1 and 3.2 with the push of design points to the edges and *middle* of the design space.

Example high quality 23 run experimental designs for the *nominal* situations described in Subsections 3.1 (Stationary Model), 3.2 (Non-Stationary Model), and the Stationary Model along with linear regression functions are illustrated in the left, middle, and right panels, respectively, of Figure 1. For each case, $\varphi(d) = \exp\{-d^2\}$. For the Non-Stationary Model example, $\omega_1(\mathbf{u})^2 = 1 - \|\mathbf{u}\|^2/2$, $\omega_2(\mathbf{u})^2 = \|\mathbf{u}\|^2/2$, $\Theta_1 = 1 \cdot \mathbf{I}_2$, $\Theta_2 = 10 \cdot \mathbf{I}_2$, and $\sigma^2 = 1$, while for the Stationary Model along with linear regression functions example $\Theta = 2 \cdot \mathbf{I}_2$ and $\sigma^2 = 1$. As expected, in the first panel, illustrating the stationary situation, the design points lie near a triangular lattice (subject to edge effects). Similarly, in the second panel, illustrating the non-stationary correlation situation, the design points in the upper right, where the shorter range, more quickly decaying, correlation is emphasized, are more dense than in the lower left, where the longer range, more slowly decaying, correlation is emphasized. Further, in the third panel, illustrating the impact of regression functions, the design points balance fill distance and a push towards the corners of the input space.

Finding designs which minimize (or nearly minimize) the error bounds is challenging. Here, we adopted a homotopy continuation [12] approach, which slowly transitions from an easier objective function to a more difficult objective function. Note that the respective target objectives for the stationary, non-stationary, and stationary with regression functions situations are given by minimizing equation (3.3), maximizing equation (3.6), and minimizing the sum of equations (3.3) and (3.9). For the stationary situation, the optimization routine was initialized at a triangular lattice, crudely scaled for small fill distance. The objective function is then taken as $(1 - \delta) \text{mean}_i \sup_{\mathbf{x} \in V_i(\Theta)} d_{\Theta}(\mathbf{x}_i, \mathbf{x}) + \delta \max_i \sup_{\mathbf{x} \in V_i(\Theta)} d_{\Theta}(\mathbf{x}_i, \mathbf{x})$ for $\delta \in [0, 1]$. Initially, the optimization is performed for $\delta = 0$, then $\delta = 1/K$, then $\delta = 2/K$, and so on, up to $\delta = K/K = 1$, for a moderately large number K . Notice that the *target* objective function is optimized when $\delta = 1$.

For both the non-stationary and stationary with regression functions situations, the objective function transition occurs both from mean to minimum/maximum and from stationary to non-stationary or with regression functions. For convenience and without loss of generality, suppose one is minimizing the maximum of local bounds. Let $Q_{\delta}(\mathbf{x}_i)$ denote an objective function *local* to data point \mathbf{x}_i , which is continuous as a function of δ and converges to the *target* local objective function as $\delta \rightarrow 1$. The objective function is then taken as $(1 - \delta) \text{mean}_i Q_{\delta}(\mathbf{x}_i) + \delta \max_i Q_{\delta}(\mathbf{x}_i)$, and is solved repeatedly for a sequence of δ

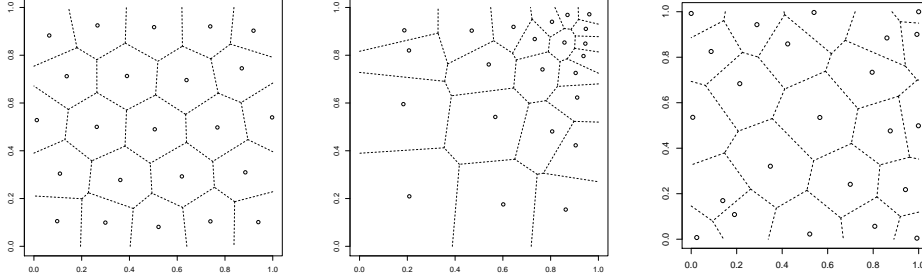


Figure 1: **Left Panel:** Nominal error design for stationary correlation. **Middle Panel:** Nominal error design for the Case 2 model of non-stationary correlation with $\varphi(d) = \exp\{-d^2\}$, $\omega_1(\mathbf{u})^2 = 1 - \|\mathbf{u}\|^2/2$, $\omega_2(\mathbf{u})^2 = \|\mathbf{u}\|^2/2$, $\Theta_1 = 1 \cdot \mathbf{I}_2$, $\Theta_2 = 10 \cdot \mathbf{I}_2$. **Right Panel:** Nominal error design for stationary correlation and a linear regression function for each dimension.

going from 0 to 1. For the non-stationary case (with minus signs to express as minization),

$$Q_\delta(\mathbf{x}_i) = -\omega_\delta^2(\mathbf{x}_i)\varphi\left(\sup_{\mathbf{x} \in V_i^*} d_{\Theta_1}(\mathbf{x}_i, \mathbf{x})\right) - (1 - \omega_\delta^2(\mathbf{x}_i))\varphi\left(\sup_{\mathbf{x} \in V_i^*} d_{\Theta_2}(\mathbf{x}_i, \mathbf{x})\right) \\ + \varphi(0)(k_1 + k_2) \max_i \sup_{\mathbf{x} \in V_i^*} \|\mathbf{x}_i - \mathbf{x}\|_2,$$

with $\omega_\delta(\mathbf{x}) = (1 - \delta) + \delta\sqrt{1 - \|\mathbf{x}\|_2^2/2}$ and Lipschitz constants $k_1 = k_2 = 1/2$, while for the non-null regression functions case,

$$Q_\delta(\mathbf{x}_i) = \frac{1}{n} \left(1 - \varphi\left(\sup_{\mathbf{x} \in V_i(\Theta)} d_\Theta(\mathbf{x}_i, \mathbf{x})\right)\right) \left(2n - 1 + \varphi\left(\sup_{\mathbf{x} \in V_i(\Theta)} d_\Theta(\mathbf{x}_i, \mathbf{x})\right)\right) \\ + \delta n \sup_{\mathbf{x} \in V_i} \left\| \mathbf{h}(\mathbf{x}) - \mathbf{H}(\mathbf{X})' \Psi_\theta(\mathbf{X}, \mathbf{X})^{-1} \Psi_\theta(\mathbf{X}, \mathbf{x}) \right\|_2^2 / \lambda_{\min}(\mathbf{H}(\mathbf{X})' \mathbf{H}(\mathbf{X})).$$

Nelder-Mead black box optimization along with penalties to enforce input space constraints was used throughout [32].

It should be noted that the approach to bounding the MSPE which was used above is not the only approach available. One possibility is to formulate the function approximation problem as n numeric integration problems [44], then apply numeric integration results [31] to provide upper bounds on the *integrated* MSPE in terms of the *star discrepancy* of the point set \mathbf{X} ,

$$D^*(\mathbf{X}) = \sup_{J = \prod_{k=1}^d [0, u_k]} \left| \frac{\#\{\mathbf{x}_i \in \mathbf{X} : J \cap \Omega\}}{n} - \frac{\text{vol } J \cap \Omega}{\text{vol } \Omega} \right|,$$

the supremum of differences between the proportion of points in a *rooted* rectangle and the proportion that are supposed to be in the rectangle under uniform measure. The fill distance can in fact be bounded above in terms of star discrepancy. It is conceivable that these type of results could be extended to the situation of non-stationarity by considering

discrepancy with respect to a more general measure. However, this line of development is not pursued here. We provide the below theorem as an indication of the type of results which are attainable. A proof of Theorem 3.2 and related development is provided in Appendix C.

Theorem 3.2. *If $f \sim \text{GP}(0, \Psi(\cdot, \cdot))$, $\Omega = [0, 1]^d$, then we have bounds on the integrated mean squared prediction error*

$$\mathbb{E}\|f - \hat{f}\|_{L_2([0,1]^d)}^2 \leq \sum_{j=n+1}^{\infty} \lambda_j + D^*(\mathbf{X})^2 \sum_{j=1}^n \mathbb{E}V(\varphi_j f)^2,$$

under the condition that $\varphi_j f$ has finite expected squared variation in the sense of Hardy and Kraus $V(\varphi_j f)$ [31], where $\Psi(\cdot, \cdot)$ has eigenvalue, eigenfunction decomposition $\Psi(\cdot, \cdot) = \sum_{i=1}^{\infty} \lambda_i \varphi_i(\cdot) \varphi_i(\cdot)$.

A similar bound can be developed in terms of the *modulus of continuity* [31] of $\varphi_j f$. See Appendix C for details.

4 Numeric Error

In Section 3, it has been shown that increasing the number of data points will decrease the nominal error. On the other hand, the numeric error can become arbitrarily large by the addition of new data sites. Here, we develop bounds on the numeric error in terms of properties of the experimental design by adapting and extending results in [15, 43, 18].

The numeric accuracy of Gaussian process emulation depends on the accuracy of floating point matrix manipulations. Floating point numbers are the rounded versions that computers perform calculations with as opposed to the targeted numbers. Commonly, computer and software have 15 digits of accuracy meaning that $\|\tilde{x} - x\|_2 / \|x\|_2 \leq 10^{-15}$, where x denotes the actual value and \tilde{x} denotes the value that the computer stores. We state a few typical assumptions on computer and software floating point accuracy and provide the following proposition relating numeric accuracy to eigenvalues of $\Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})$. Below, $\kappa(\cdot) = \lambda_{\max}(\cdot) / \lambda_{\min}(\cdot)$ denotes the condition number of its (diagonalizable) argument. The proof of Proposition 4.1 is provided in Appendix D.

Assumption 4.1. *Take $\kappa(\Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})) = r/\delta$ for $r < 1$ and*

$$\begin{aligned} \|\mathbf{h}(\mathbf{x}) - \tilde{\mathbf{h}}(\mathbf{x})\|_2 &\leq \delta \|\mathbf{h}(\mathbf{x})\|_2, \quad \|f(\mathbf{X}) - \tilde{f}(\mathbf{X})\|_2 \leq \delta \|f(\mathbf{X})\|_2, \quad \|h_j(\mathbf{X}) - \tilde{h}_j(\mathbf{X})\|_2 \leq \delta \|h_j(\mathbf{X})\|_2, \\ \|\Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X}) - \tilde{\Psi}_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})\|_2 &\leq \delta \|\Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})\|_2, \quad \text{and} \quad \|\Psi_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{X}) - \tilde{\Psi}_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{X})\|_2 \leq \delta \|\Psi_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{X})\|_2. \end{aligned}$$

Proposition 4.1. *Under Assumption 4.1, the Gaussian process BLUP (2.5) has numeric error $|\hat{f}_{\boldsymbol{\theta}}(\mathbf{x}) - \tilde{f}_{\boldsymbol{\theta}}(\mathbf{x})|$, for arbitrary parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}' \quad \boldsymbol{\theta}')'$, bounded above by*

$$\delta \|\mathbf{h}(\mathbf{x})\|_2 \|\boldsymbol{\beta}\|_2 + \frac{2\delta}{1-r} \|\Psi_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{X})\|_2 \left(\|f(\mathbf{X})\|_2 + \|\boldsymbol{\beta}\|_2 \sqrt{\sum_{j=1}^p \|h_j(\mathbf{X})\|_2^2} \right) g(\mathbf{X}, \Psi_{\boldsymbol{\theta}}), \quad (4.1)$$

where

$$g(\mathbf{X}, \Psi_{\boldsymbol{\theta}}) = \frac{\kappa(\Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})) + 1}{\lambda_{\min}(\Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X}))}.$$

For experimental designs which are not too small and have reasonable parameter estimation properties, $\|\hat{\beta}\|_2$ will be of a similar magnitude to $\|\beta\|_2$. Further, the terms $\|\Psi_{\theta}(\mathbf{x}, \mathbf{X})\|_2$, $\|f(\mathbf{X})\|_2$, and $\|h_j(\mathbf{X})\|_2$ are Monte Carlo approximations to $\sqrt{n}\|\Psi_{\theta}(\mathbf{x}, \cdot)\|_{L_2(F)}$, $\sqrt{n}\|f(\cdot)\|_{L_2(F)}$, and $\sqrt{n}\|h_j(\cdot)\|_{L_2(F)}$, respectively, with respect to the large sample distribution of the experimental design F . That is, the terms in the bound (4.1), aside from $g(\mathbf{X}, \Psi_{\theta})$, influence the numeric accuracy only weakly and vanishingly, and the bound depends on the experimental design primarily through $g(\mathbf{X}, \Psi_{\theta})$. The implication of Gershgorin's theorem [41] in (B.2) implies $g(\mathbf{X}, \Psi_{\theta})$ can be bounded in terms of the minimum eigenvalue of $\Psi_{\theta}(\mathbf{X}, \mathbf{X})$ as

$$g(\mathbf{X}, \Psi_{\theta}) \leq \frac{1}{\lambda_{\min}(\Psi_{\theta}(\mathbf{X}, \mathbf{X}))} \left(\frac{n \sup_{\mathbf{u}, \mathbf{v} \in \Omega} \Psi_{\theta}(\mathbf{u}, \mathbf{v})}{\lambda_{\min}(\Psi_{\theta}(\mathbf{X}, \mathbf{X}))} + 1 \right). \quad (4.2)$$

We adapt and generalize results from [43] in the theorem below providing a lower bound for $\lambda_{\min}(\Psi_{\theta}(\mathbf{X}, \mathbf{X}))$ and thereby an upper bound for (4.2). The proof of Theorem 4.1 is provided in Appendix E. First, a definition of the Fourier transform is provided.

Definition 4.1. For $f \in L_1(\mathbb{R}^d)$ define the Fourier transform [36]

$$\hat{f}(\omega) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} f(x) e^{-i\omega'x} dx.$$

Theorem 4.1. Suppose Φ is a positive definite, translation invariant kernel with Fourier transform $\hat{\Phi} \in L_1(\mathbb{R}^d)$. Then,

$$\sum_{j=1}^n \sum_{k=1}^n \alpha_j \alpha_k \Phi(\mathbf{x}_j - \mathbf{x}_k) \geq \Upsilon_{c_*/q(\Theta)}(\mathbf{0}) \sum_{j=1}^n \alpha_j^2 \left(1 - \frac{\Gamma^2(d/2 + 1)\pi}{18} \left(\frac{q(\Theta)}{q_j(\Theta)} \right) \left(\frac{12}{c_*} \right)^{d+1} \right),$$

for $c_* > 0$, where

$$\Upsilon_M(\mathbf{0}) \equiv \lim_{\mathbf{t} \rightarrow \mathbf{0}} \Upsilon_M(\mathbf{t}) = \frac{\hat{\Phi}_*(M)}{\Gamma(d/2 + 1)} \left(\frac{M}{2^{3/2}} \right)^d, \quad \hat{\Phi}_*(M) = \inf_{\|\omega\|_2 \leq 2M} \hat{\Phi}(\omega)$$

for $M > 0$, and the respective local and global separation distances with respect to the Mahalanobis-like distance d_{Θ} , Θ non-singular are given by

$$q_j(\Theta) = \frac{1}{2} \min_{k=1, \dots, n, k \neq j} d_{\Theta}(\mathbf{x}_j, \mathbf{x}_k) \quad \text{and} \quad q(\Theta) = \min_j q_j(\Theta).$$

This result is now applied to the Stationary Model and Non-Stationary Model (Subsections 3.1 and 3.2, respectively).

4.1 Stationary Model

Here,

$$\begin{aligned}
\lambda_{\min}(\Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})) &= \min_{\|\mathbf{a}\|_2=1} \sum_{i,j} a_i a_j \Psi_{\boldsymbol{\theta}}(\mathbf{x}_i, \mathbf{x}_j) \\
&= \sigma^2 \min_{\|\mathbf{a}\|_2=1} \sum_{i,j} a_i a_j \varphi(\|\boldsymbol{\Theta}(\mathbf{x}_i - \mathbf{x}_j)\|_2) \\
&= \sigma^2 \min_{\|\mathbf{a}\|_2=1} \sum_{i,j} a_i a_j \varphi(\|\mathbf{x}_i^* - \mathbf{x}_j^*\|_2) \\
&= \sigma^2 \min_{\|\mathbf{a}\|_2=1} \sum_{i,j} a_i a_j \Phi(\mathbf{x}_i^* - \mathbf{x}_j^*) \\
&\geq \sigma^2 \min_{\|\mathbf{a}\|_2=1} \sum_{i=1}^n a_i^2 \ell_i(\boldsymbol{\Theta}),
\end{aligned} \tag{4.3}$$

where $\Phi(\mathbf{x}^* - \mathbf{y}^*) = \varphi(\|\mathbf{x}^* - \mathbf{y}^*\|_2)$ and

$$\ell_i(\boldsymbol{\Theta}) = \Upsilon_{c_*/q(\boldsymbol{\Theta})}(\mathbf{0}) \left(1 - \frac{\Gamma^2(d/2 + 1)\pi}{18} \left(\frac{q(\boldsymbol{\Theta})}{q_i(\boldsymbol{\Theta})} \right) \left(\frac{12}{c_*} \right)^{d+1} \right). \tag{4.4}$$

The lower bound (4.3) is maximized for $\ell_i(\boldsymbol{\Theta})$ constant over i and as large as possible. This implies $q_i(\boldsymbol{\Theta}) = q(\boldsymbol{\Theta})$ for all i . Now, the lower bound depends on

$$\Upsilon_{c_*/q(\boldsymbol{\Theta})}(\mathbf{0}) = \frac{\hat{\Phi}_*(c_*/q(\boldsymbol{\Theta}))}{\Gamma(d/2 + 1)} \left(\frac{c_*/q(\boldsymbol{\Theta})}{2^{3/2}} \right)^d, \tag{4.5}$$

which, for sufficiently small $q(\boldsymbol{\Theta})$, is an increasing function of $q(\boldsymbol{\Theta})$ that approaches zero as $q(\boldsymbol{\Theta})$ approaches zero. That is, in the stationary situation, numeric accuracy is preserved for designs which are well-separated.

4.2 Non-Stationary Model

Here, assume *additionally* that $\boldsymbol{\Theta}_2 = a\boldsymbol{\Theta}_1$ for some $a > 1$. Slightly, coarsen the bounds by replacing $\Upsilon_M(\mathbf{0})$ with its monotone decreasing in M lower bound $\Upsilon_M^0 = \inf_{m \in [r_*, M]} \Upsilon_m(\mathbf{0})$, $r_* = c_*/\max_{\mathbf{x}, \mathbf{y} \in \Omega} d_{\boldsymbol{\Theta}_1}(\mathbf{x}, \mathbf{y})$. Let $\tilde{\ell}_i$ denote the coarsened version of (4.4). Then,

$$\begin{aligned}
\lambda_{\min}(\Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})) &= \min_{\|\mathbf{a}\|_2=1} \sum_{i,j} a_i a_j \Psi_{\boldsymbol{\theta}}(\mathbf{x}_i, \mathbf{x}_j) \\
&= \sigma^2 \min_{\|\mathbf{a}\|_2=1} \sum_{i,j} a_i a_j (\omega_1(\mathbf{x}_i)\omega_1(\mathbf{x}_j)\varphi(\|\boldsymbol{\Theta}_1(\mathbf{x}_i - \mathbf{x}_j)\|_2) + \omega_2(\mathbf{x}_i)\omega_2(\mathbf{x}_j)\varphi(\|\boldsymbol{\Theta}_2(\mathbf{x}_i - \mathbf{x}_j)\|_2)) \\
&\geq \sigma^2 \min_{\|\mathbf{a}\|_2=1} \sum_i a_i^2 \left(\omega_1(\mathbf{x}_i)^2 \tilde{\ell}_i(\boldsymbol{\Theta}_1) + \omega_2(\mathbf{x}_i)^2 \tilde{\ell}_i(\boldsymbol{\Theta}_2) \right),
\end{aligned} \tag{4.6}$$

where $\hat{\Phi}$ is the Fourier transform of Φ defined by $\Phi(\mathbf{x}^* - \mathbf{y}^*) = \varphi(\|\mathbf{x}^* - \mathbf{y}^*\|_2)$ in (4.4) and (4.5). The lower bound (4.6) is maximized for $\omega_1(\mathbf{x}_i)^2 \tilde{\ell}_i(\boldsymbol{\Theta}_1) + \omega_2(\mathbf{x}_i)^2 \tilde{\ell}_i(\boldsymbol{\Theta}_2)$ constant over i and as large as possible.

Consider two design points \mathbf{x}_i and \mathbf{x}_j and suppose that the points in the input space *near* \mathbf{x}_i have more weight on the global, long range, correlation than the points in the input space *near* \mathbf{x}_j and the points in the input space *near* \mathbf{x}_j have more weight on the local, short range, correlation than the points in the input space *near* \mathbf{x}_i , in the sense that

$$\begin{aligned}\omega_1(\mathbf{x}_i)^2(\tilde{\ell}_i(\Theta_2) - \tilde{\ell}_i(\Theta_1)) &\geq \omega_1(\mathbf{x}_j)^2(\tilde{\ell}_j(\Theta_2) - \tilde{\ell}_j(\Theta_1)), \\ \omega_2(\mathbf{x}_i)^2(\tilde{\ell}_i(\Theta_2) - \tilde{\ell}_i(\Theta_1)) &\leq \omega_2(\mathbf{x}_j)^2(\tilde{\ell}_j(\Theta_2) - \tilde{\ell}_j(\Theta_1)).\end{aligned}\tag{4.7}$$

Here, we consider the situation where $q_i(\Theta_1)$ and $q_i(\Theta_2)$ are *small* across i , the situation where a bound on the numeric error is most relevant. For $q(\Theta_1)$ and $q(\Theta_2)$ sufficiently small, $\tilde{\Upsilon}_{c_*/q(\Theta)}^0$ is *strictly* increasing in q . Further, the assumption $\Theta_2 = a\Theta_1$ for some $a > 1$ implies $q(\Theta_1)/q_i(\Theta_1) = q(\Theta_2)/q_i(\Theta_2)$ and $q(\Theta_2) > q(\Theta_1)$. Together, these facts imply $\tilde{\ell}_i(\Theta_2) > \tilde{\ell}_i(\Theta_1)$. Uniformity of the bounds (4.6) along with $\omega_1(\cdot)^2 + \omega_2(\cdot)^2 = 1$ gives

$$\begin{aligned}\omega_1(\mathbf{x}_i)^2(\tilde{\ell}_i(\Theta_1) - \tilde{\ell}_i(\Theta_2)) - \omega_1(\mathbf{x}_j)^2(\tilde{\ell}_j(\Theta_1) - \tilde{\ell}_j(\Theta_2)) &= \tilde{\ell}_j(\Theta_2) - \tilde{\ell}_i(\Theta_2), \\ \omega_2(\mathbf{x}_j)^2(\tilde{\ell}_j(\Theta_1) - \tilde{\ell}_j(\Theta_2)) - \omega_2(\mathbf{x}_i)^2(\tilde{\ell}_i(\Theta_1) - \tilde{\ell}_i(\Theta_2)) &= \tilde{\ell}_j(\Theta_1) - \tilde{\ell}_i(\Theta_1).\end{aligned}\tag{4.8}$$

Combining (4.7) and (4.8) with the fact that $\tilde{\ell}_i(\Theta)$ is an increasing function of $q_i(\Theta)$ for small $q_i(\Theta)$ gives

$$q_j(\Theta_1) < q_i(\Theta_1) \quad \text{and} \quad q_j(\Theta_2) < q_i(\Theta_2).$$

That is, a uniform bound on (4.6) is achieved by an experimental design \mathbf{X} which has smaller separation distance, with respect to either d_{Θ_1} or d_{Θ_2} , in regions with more emphasis on the local, more quickly decaying, correlation and less emphasis on the global, more slowly decaying, correlation. Note that in the numeric accuracy context, the global and local *emphases*, for *small* $q_i(\Theta_1)$ and $q_i(\Theta_2)$, at \mathbf{x}_i are given concretely by $\omega_1(\mathbf{x}_i)^2(\tilde{\ell}_i(\Theta_2) - \tilde{\ell}_i(\Theta_1))$ and $\omega_2(\mathbf{x}_i)^2(\tilde{\ell}_i(\Theta_2) - \tilde{\ell}_i(\Theta_1))$, respectively.

Example high quality 23 run experimental designs for the *numeric* situations described in Subsection 4.1 (Stationary Model) and Subsection 4.2 (Non-Stationary Model) are illustrated in the left and right panels, respectively, of Figure 2. For the Stationary Model example, $\varphi(d) = \exp\{-d^2\}$, the so-called *Gaussian* correlation function. For the Non-Stationary Model example, $\varphi(d)$ is Wendland's kernel with $k = 10$ [43], $\omega_1(\mathbf{u})^2 = 1 - \|\mathbf{u}\|_2^2/2$, $\omega_2(\mathbf{u})^2 = \|\mathbf{u}\|_2^2/2$, $\Theta_1 = 0.1 \cdot \mathbf{I}_2$, and $\Theta_2 = 1 \cdot \mathbf{I}_2$. For both cases, $\sigma^2 = 1$. As expected, in the first panel, illustrating the stationary situation, the design points lie near a triangular lattice (subject to edge effects), similar to, but expanded towards the edges of the design space relative to, the nominal design. Similarly, in the second panel, illustrating the non-stationary correlation situation, the design points in the upper right-hand corner, where the shorter range, more quickly decaying, correlation is emphasized, are more dense than in the lower left-hand corner, where the longer range, more slowly decaying, correlation is emphasized. While the provided bounds hold for all $c_* > 0$, the actual value of the bounds depends on the selected value of c_* . Here, we take $c_* = 1.1 \times 12 \left(\frac{18}{\pi\Gamma^2(d/2+1)} \right)^{-1/(d+1)}$. Similarly to the nominal examples, the optimization routine was initialized at a triangular lattice, scaled to maximize the separation distance. For the stationary situation, a homotopy continuation [12] approach was used, slowly transitioning from maximizing the mean of the local bounds $\ell_i(\Theta)$ to maximizing the minimum of the local bounds. For the non-stationary situation, a homotopy continuation approach with two stages was used, first transitioning from ω_1, ω_2 constant to varying over the input space, then transitioning from maximizing the mean of the local bounds $\omega_1(\mathbf{x}_i)^2\tilde{\ell}_i(\Theta_1) + \omega_2(\mathbf{x}_i)^2\tilde{\ell}_i(\Theta_2)$ to maximizing the minimum of the local bounds. Nelder-Mead [32] black box optimization was used throughout.

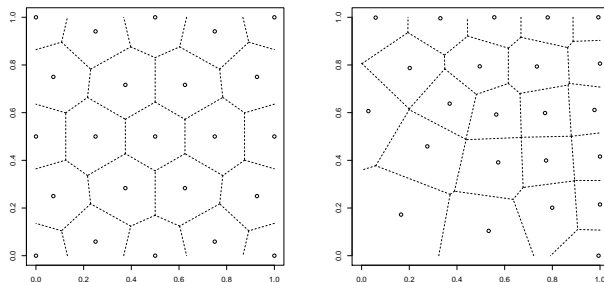


Figure 2: **Left Panel:** Numeric error design for stationary correlation. **Right Panel:** Numeric error design for the Case 2 model of non-stationary correlation with $\varphi(d)$ Wendland's kernel with $k = 10$, $\omega_1(\mathbf{u})^2 = 1 - \|\mathbf{u}\|_2^2/2$, $\omega_2(\mathbf{u})^2 = \|\mathbf{u}\|_2^2/2$, $\Theta_1 = 0.1 \cdot \mathbf{I}_2$, and $\Theta_2 = 1 \cdot \mathbf{I}_2$

5 Parameter Estimation

Consider maximum likelihood estimation and let \mathbb{E} denote the expectation conditional on \mathbf{X} and $f(\mathbf{X})$. Then, for n not too small,

$$\begin{aligned} \mathbb{E} \left\{ \hat{f}_{\vartheta_*}(\mathbf{x}) - \hat{f}_{\hat{\vartheta}}(\mathbf{x}) \right\}^2 &\approx \frac{\partial \hat{f}_{\vartheta_*}(\mathbf{x})}{\partial \vartheta_*'} \text{Var} \hat{\vartheta} \frac{\partial \hat{f}_{\vartheta_*}(\mathbf{x})}{\partial \vartheta_*} \\ &\approx \frac{\partial \hat{f}_{\vartheta_*}(\mathbf{x})}{\partial \vartheta_*'} \mathcal{I}(\vartheta_*)^{-1} \frac{\partial \hat{f}_{\vartheta_*}(\mathbf{x})}{\partial \vartheta_*}, \end{aligned} \quad (5.1)$$

where $\mathcal{I}(\vartheta_*) = \mathbb{E} \frac{\partial \ell}{\partial \vartheta_*'} \frac{\partial \ell}{\partial \vartheta_*}$ denotes the information matrix and ℓ denotes the log-likelihood of the data $f(\mathbf{X})$. Roughly, a high-quality design for parameter estimation will have $\left\| \frac{\partial \hat{f}_{\vartheta_*}(\mathbf{x})}{\partial \vartheta_*'} \right\|_2$ small and $\lambda_{\min}(\mathcal{I}(\vartheta_*))$ large. Arrange the vector of parameters as $\vartheta = (\beta' \ \theta')'$ and $\theta = (\sigma^2 \ \varrho')'$. Throughout the parameter estimation section, take

$$\Psi_{\theta}(\mathbf{u}, \mathbf{v}) = \sigma^2 \Phi_{\varrho}(\mathbf{u}, \mathbf{v}). \quad (5.2)$$

Expressions for the components of the right-hand side of (5.1) are provided in Lemma F.1 in Appendix F. These expressions are in turn used to develop the approximate upper bound for the mean squared prediction error given in Theorem 5.1. Proofs of Lemma F.1 and Theorem 5.1 are provided in Appendix F.

Theorem 5.1. *Suppose $f(\cdot) \sim \text{GP}(\mathbf{h}(\cdot)' \beta, \sigma^2 \Phi_{\varrho}(\cdot, \cdot))$ for fixed, known regression functions $\mathbf{h}(\cdot)$ and positive definite $\Phi_{\varrho}(\cdot, \cdot)$. Further, assume that the input locations \mathbf{X} have large sample distribution F . Let $\hat{\vartheta}$ denote the maximum likelihood estimator of the unknown parameters $\vartheta = (\beta' \ \sigma^2 \ \varrho')'$. Then, an approximate upper bound for*

$\mathbb{E} \left\{ \hat{f}_{\vartheta_*}(\mathbf{x}) - \hat{f}_{\hat{\vartheta}}(\mathbf{x}) \right\}^2$ *is given by*

$$\sup_{\mathbf{u}, \mathbf{v} \in \Omega} \Phi_{\varrho}(\mathbf{u}, \mathbf{v}) \left(\frac{\|\mathbf{c}_1\|_2^2}{s_1} + \frac{2 \sup_{\mathbf{u}, \mathbf{v} \in \Omega} \Phi_{\varrho}(\mathbf{u}, \mathbf{v}) \|\mathbf{c}_3\|_2^2}{s_2} \right)$$

where

$$\begin{aligned} \mathbf{c}_1 &= \mathbf{h}(\mathbf{x}) - \mathbf{H}(\mathbf{X})' \Psi_\theta(\mathbf{X}, \mathbf{X})^{-1} \Psi_\theta(\mathbf{X}, \mathbf{x}), \\ \mathbf{c}_3 &= \left(\frac{\partial \Psi_\theta(\mathbf{x}, \mathbf{X})}{\partial \boldsymbol{\rho}} - (\mathbf{I}_d \otimes \Psi_\theta(\mathbf{x}, \mathbf{X}) \Psi_\theta(\mathbf{X}, \mathbf{X})^{-1}) \frac{\partial \Psi_\theta(\mathbf{X}, \mathbf{X})}{\partial \boldsymbol{\rho}} \right) \Psi_\theta(\mathbf{X}, \mathbf{X})^{-1} \delta(\mathbf{X}), \end{aligned}$$

$s_1 = \lambda_{\min} \left(\int \mathbf{h}(\mathbf{y}) \mathbf{h}(\mathbf{y})' dF(\mathbf{y}) \right)$, and s_2 is implicitly defined in (F.15). In particular, $s_1 > 0$ unless $\mathbf{h}(\mathbf{y})' \mathbf{a} = 0$ with probability 1 with respect to F for some $\mathbf{a} \neq \mathbf{0}$ and $s_2 > 0$ unless $\frac{\partial \Phi_{\boldsymbol{\rho}}(\mathbf{x}, \mathbf{y})}{\partial \boldsymbol{\rho}'} \mathbf{a} = \Phi_{\boldsymbol{\rho}}(\mathbf{x}, \mathbf{y}) \mathbf{b}$ with probability 1 with respect to $F \times F$ for some $(\mathbf{a}' \ \mathbf{b}') \neq \mathbf{0}$.

This upper bound is approximate in the sense that for a sequence of experimental designs for which the maximum likelihood estimates converge, the probability that the upper bound is violated by more than $\varepsilon > 0$ goes to zero.

The term $\|\mathbf{c}_3\|_2^2$ admits the simple upper bound

$$\left\| \frac{\partial \Psi_\theta(\mathbf{x}, \mathbf{X})}{\partial \boldsymbol{\rho}} - (\mathbf{I}_u \otimes \Psi_\theta(\mathbf{x}, \mathbf{X}) \Psi_\theta(\mathbf{X}, \mathbf{X})^{-1}) \frac{\partial \Psi_\theta(\mathbf{X}, \mathbf{X})}{\partial \boldsymbol{\rho}} \right\|_2^2 \|\delta(\mathbf{X})\|_2^2 / \lambda_{\min} (\Psi_\theta(\mathbf{X}, \mathbf{X}))^2.$$

The term $\|\delta(\mathbf{X})\|_2 = \|f(\mathbf{X}) - \mathbf{H}(\mathbf{X})\boldsymbol{\beta}\|_2$ is an approximation to $\sqrt{n} \|f(\cdot) - \mathbf{h}(\cdot)' \boldsymbol{\beta}\|_{L_2(F)}$ with respect to the large sample distribution of the data F . Further, $\lambda_{\min} (\Psi_\theta(\mathbf{X}, \mathbf{X}))$ is well-controlled by experimental designs which maintain high-quality numeric properties. Similarly, both $\mathbf{h}(\mathbf{x}) - \mathbf{H}(\mathbf{X})' \Psi_\theta(\mathbf{X}, \mathbf{X})^{-1} \Psi_\theta(\mathbf{X}, \mathbf{x})$ and $\frac{\partial \Psi_\theta(\mathbf{x}, \mathbf{X})}{\partial \boldsymbol{\rho}} - (\mathbf{I}_u \otimes \Psi_\theta(\mathbf{x}, \mathbf{X}) \Psi_\theta(\mathbf{X}, \mathbf{X})^{-1}) \frac{\partial \Psi_\theta(\mathbf{X}, \mathbf{X})}{\partial \boldsymbol{\rho}}$ are nominal interpolation errors, respectively for the regression functions and (the transpose of) the Jacobian of $\Psi_\theta(\mathbf{X}, \mathbf{x})$ with respect to the correlation parameters. As discussed towards the end of Section 3, we expect the norms of both of these interpolation errors to behave in a manner similar to Gaussian process or RKHS interpolation. That is, the norms of both of these terms will be small for experimental designs which are high-quality with respect to nominal error.

As discussed towards the end of Section 3, s_1 will be large for sets of input locations which have good traditional experimental design properties. The term s_2 will be large for experimental designs whose system of differences $\{\mathbf{x}_i - \mathbf{x}_j\}$ make $\frac{\partial \Phi_{\boldsymbol{\rho}}(\mathbf{x}_i, \mathbf{x}_j)}{\partial \boldsymbol{\rho}}$ far from zero, *balanced* with respect to a basis of $\mathbb{R}^{\dim \boldsymbol{\rho}}$, and *not collinear* with $\Phi_{\boldsymbol{\rho}}(\mathbf{x}_i, \mathbf{x}_j)$. Consider as an example, underlying kernels which *depend only on the difference between their arguments* and are *radially decreasing* in the sense that $\Phi(\boldsymbol{\delta}_1) \geq \Phi(\boldsymbol{\delta}_2)$ if $\|\boldsymbol{\delta}_1\|_2 \leq \|\boldsymbol{\delta}_2\|_2$ with $\Phi_{\boldsymbol{\rho}}(\cdot) = \Phi(\text{diag}\{\boldsymbol{\rho}\}(\cdot))$. For radially decreasing underlying kernels Φ , the term $\frac{\partial \Phi_{\boldsymbol{\rho}}(\mathbf{x}_i - \mathbf{x}_j)}{\partial \boldsymbol{\rho}}$ is near zero if $\mathbf{x}_i - \mathbf{x}_j$ is near zero or far from zero, while the term $\frac{\partial \Phi_{\boldsymbol{\rho}}(\mathbf{x}_i - \mathbf{x}_j)}{\partial \boldsymbol{\rho}}$ has negative components if the difference $\mathbf{x}_i - \mathbf{x}_j$ is slightly beyond the location where $\Phi(\text{diag}\{\boldsymbol{\theta}\}(\cdot))$ is decreasing most rapidly along each coordinate axis, since $\frac{\partial \Phi_{\boldsymbol{\rho}}(\mathbf{x}_i - \mathbf{x}_j)}{\partial \boldsymbol{\rho}} = \text{diag}\{\mathbf{x}_i - \mathbf{x}_j\} \nabla \Phi(\text{diag}\{\boldsymbol{\rho}\}(\mathbf{x}_i - \mathbf{x}_j))$. In this situation, $\frac{\partial \Phi_{\boldsymbol{\rho}}(\mathbf{x}_i - \mathbf{x}_j)}{\partial \boldsymbol{\rho}}$ has negative components and $\Phi_{\boldsymbol{\rho}}(\mathbf{x}_i - \mathbf{x}_j)$ is a non-negative weighting function, so they could only be (nearly) collinear for experimental designs which make almost all $\frac{\partial \Phi_{\boldsymbol{\rho}}(\mathbf{x}_i - \mathbf{x}_j)}{\partial \boldsymbol{\rho}}$ near zero. Figure 3 shows $\Phi_{\boldsymbol{\rho}}(\cdot)$ and both components of $\frac{\partial \Phi_{\boldsymbol{\rho}}(\cdot)}{\partial \boldsymbol{\rho}}$ for $\Phi(\mathbf{d}) = \exp\{-\mathbf{d}'\mathbf{d}\}$ and $\boldsymbol{\rho} = (1 \ 2)'$. Pairs of points $\mathbf{x}_i, \mathbf{x}_j$ whose difference lies slightly beyond the location where $\Phi(\text{diag}\{\boldsymbol{\theta}\}(\cdot))$ is decreasing most rapidly along each coordinate axis have potential to increase eigenvalues of $\sum_{ij} \frac{\partial \Phi_{\boldsymbol{\rho}}(\mathbf{x}_i - \mathbf{x}_j)}{\partial \boldsymbol{\rho}} \frac{\partial \Phi_{\boldsymbol{\rho}}(\mathbf{x}_i - \mathbf{x}_j)}{\partial \boldsymbol{\rho}'}$. Further, $\lambda_{\min} \left(\sum_{ij} \frac{\partial \Phi_{\boldsymbol{\rho}}(\mathbf{x}_i - \mathbf{x}_j)}{\partial \boldsymbol{\rho}} \frac{\partial \Phi_{\boldsymbol{\rho}}(\mathbf{x}_i - \mathbf{x}_j)}{\partial \boldsymbol{\rho}'} \right)$ is large for sets of differences $\{\mathbf{x}_i - \mathbf{x}_j\}$ which *balance* the

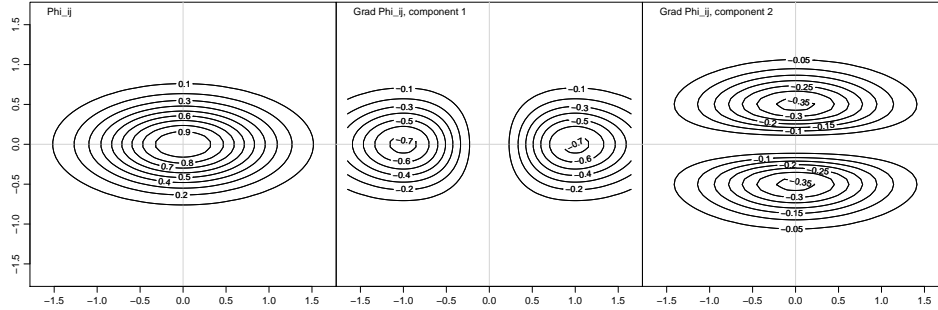


Figure 3: **Left Panel:** $\Phi_{\boldsymbol{\rho}}(\cdot)$. **Middle and Right Panels:** Components of $\frac{\partial \Phi_{\boldsymbol{\rho}}(\cdot)}{\partial \boldsymbol{\rho}}$.

differences along coordinate axes in the sense that

$$n_k \max_{\mathbf{d}} \left\{ \frac{\partial \Phi_{\boldsymbol{\rho}}(\mathbf{d})}{\partial \boldsymbol{\rho}} \right\}_k^2 \approx n_l \max_{\mathbf{d}} \left\{ \frac{\partial \Phi_{\boldsymbol{\rho}}(\mathbf{d})}{\partial \boldsymbol{\rho}} \right\}_l^2,$$

for $k, l = 1, \dots, d$ where $\{\cdot\}_k$ denotes element k of its argument and n_k denotes the number of differences (of length slightly beyond the location where $\Phi(\text{diag}\{\boldsymbol{\theta}\}(\cdot))$ is decreasing most rapidly) along coordinate axis k . In the example described above and illustrated in Figure 3, $\{\nabla \Phi(\text{diag}\{\boldsymbol{\rho}\}(\mathbf{d}))' \text{diag}\{\mathbf{d}\}\}_1^2 \approx (0.8)^2$ and $\{\nabla \Phi(\text{diag}\{\boldsymbol{\rho}\}(\mathbf{d}))' \text{diag}\{\mathbf{d}\}\}_2^2 \approx (0.4)^2$, respectively at $(\pm 1 \ 0)'$ and $(0 \ \pm 0.5)'$, so an experimental design *solely targeting the eigenvalues* of $\sum_{ij} \frac{\partial \Phi_{\boldsymbol{\rho}}(\mathbf{x}_i - \mathbf{x}_j)}{\partial \boldsymbol{\rho}} \frac{\partial \Phi_{\boldsymbol{\rho}}(\mathbf{x}_i - \mathbf{x}_j)}{\partial \boldsymbol{\rho}'}$ would have roughly n_1 differences $\mathbf{x}_i - \mathbf{x}_j = (\pm 1 \ 0)'$ and n_2 differences $\mathbf{x}_i - \mathbf{x}_j = (0 \ \pm 0.5)'$ where

$$n_1(0.8)^2 = n_2(0.4)^2 \implies n_1 = \frac{n_2}{4}.$$

Consider another example with $\Phi(\mathbf{d}) = \exp\{-\mathbf{d}'\mathbf{d}\}$, $\Phi_{\boldsymbol{\rho}}(\cdot) = \Phi(\text{diag}\{\boldsymbol{\rho}\}(\cdot))$, and $\boldsymbol{\rho} = (3 \ 3)'$. An experimental design maximizing $\lambda_{\min} \left(\sum_{ij} \frac{\partial \Phi_{\boldsymbol{\rho}}(\mathbf{x}_i - \mathbf{x}_j)}{\partial \boldsymbol{\rho}} \frac{\partial \Phi_{\boldsymbol{\rho}}(\mathbf{x}_i - \mathbf{x}_j)}{\partial \boldsymbol{\rho}'}$ is shown in the left panel of Figure 4. There are 11 points at the middle location and 3 at each peripheral location. In particular, this design is *not* space-filling. A high quality experimental design with respect to the upper-bound in Theorem 5.1 is shown in the right panel of Figure 4. The influence of $\lambda_{\min} \left(\sum_{ij} \frac{\partial \Phi_{\boldsymbol{\rho}}(\mathbf{x}_i - \mathbf{x}_j)}{\partial \boldsymbol{\rho}} \frac{\partial \Phi_{\boldsymbol{\rho}}(\mathbf{x}_i - \mathbf{x}_j)}{\partial \boldsymbol{\rho}'}$ is substantially less than the influence of the space-filling properties controlling the nominal and numeric error.

6 Discussion

Broadly applicable and rigorously justified principles of experimental design for Gaussian process emulation of deterministic computer experiments have been developed. The *space-filling* properties “small fill distance” and “large separation distance”, potentially with respect to an input space rescaling to accommodate varying rates of correlation decay depending on displacement orientation, are only weakly conflicting and ensure *well-controlled* nominal, numeric, and parameter estimation error. The presence of non-stationarity in correlation requires a higher density of input locations in regions with more emphasis on

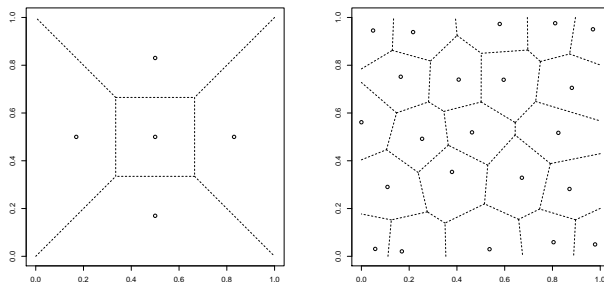


Figure 4: **Left Panel:** Experimental design maximizing $\lambda_{\min}(\mathbf{C}'_{\theta}\mathbf{C}_{\theta})$ and minimizing $\kappa(\mathbf{C}'_{\theta}\mathbf{C}_{\theta})$ for $\Phi(\mathbf{d}) = \exp\{-\mathbf{d}'\mathbf{d}\}$ and $\boldsymbol{\varrho} = (3 \ 3)'$. Note that there are 11 points at the middle location and 3 at each peripheral location. **Right Panel:** Parameter estimation error design with respect to the upper-bound in Theorem 5.1.

the local, more quickly decaying, correlation, relative to input locations in regions with more emphasis on the global, more slowly decaying, correlation. Designs of this type can potentially be constructed via homotopy continuation techniques [12]. Alternatively, techniques in [8] or [24] may be useful for constructing high-quality experimental designs in a non-stationary setting. The inclusion of regression functions results in high quality designs which balance traditional experimental design properties, targeting the variance of regression function coefficients, with space-filling properties, while consideration of error in parameter estimation results in high quality designs slightly favoring pairs of input locations having particular lengths and orientations of their displacement vector. The influence on the accuracy of emulation of regression functions and error in parameter estimation appears to be substantially less than the influence of the *space-filling* properties “small fill distance” and “large separation distance”.

The results presented in Theorems 3.1, 4.1, 5.1, and their subsequent discussions are generally well-aligned with *distance-based* design criteria, such as minimax (minimize the maximum distance from inputs of interest to the design) and maximin (maximize the minimum distance between design points). On their surface, the results appear to be somewhat less well-aligned with designs which emphasize *low-dimensional projections*, such as Latin hypercube [29], orthogonal array-based Latin hypercube [38], or MaxPro [25]. On the other hand, when some inputs to the unknown target function are inert, having little or no impact on the response, the presented results indicate that the corresponding correlation parameters (here described as inverse bandwidths) are near zero and high-quality designs should have space-fillingness in the relevant lower-dimensional projections. Throughout this work, it has been tacitly assumed that a relatively accurate *a priori* guess for the correlation parameters is available. Explicitly handling of *a priori* uncertainty in correlation parameters could form an important extension of this work. In the case where one expects several inert variables, a design with good projection properties might reasonably be expected.

As a brief exploration, several of the proposed *high-quality* designs were compared to one another and a spectrum of designs from the literature, in stationary, non-stationary, non-trivial regression functions, and inert inputs settings. In particular, five examples were considered, with results presented in Table 1. The *Stationary Example* follows the setting illustrated in the left panel of Figure 1, $\Psi(\mathbf{u}, \mathbf{v}) = \exp\{-\|\mathbf{u} - \mathbf{v}\|_2^2\}$. The *Non-Stationary*

Nominal Example follows the non-stationary example illustrated in the middle panel of Figure 1, $\varphi(d) = \exp\{-d^2\}$, $\omega_1(\mathbf{u})^2 = 1 - \|\mathbf{u}\|^2/2$, $\omega_2(\mathbf{u})^2 = \|\mathbf{u}\|^2/2$, $\Theta_1 = 1 \cdot \mathbf{I}_2$, $\Theta_2 = 10 \cdot \mathbf{I}_2$. The *Regression Functions Example* follows the setting illustrated in the right panel of Figure 1, $\Psi(\mathbf{u}, \mathbf{v}) = \exp\{-\|\mathbf{u} - \mathbf{v}\|_2^2\}$ and a linear regression function for each dimension. The *Non-Stationary Numeric Example* follows the non-stationary example illustrated in the right panel of Figure 2, $\varphi(d)$ Wendland’s kernel with $k = 10$, $\omega_1(\mathbf{u})^2 = 1 - \|\mathbf{u}\|^2/2$, $\omega_2(\mathbf{u})^2 = \|\mathbf{u}\|^2/2$, $\Theta_1 = 0.1 \cdot \mathbf{I}_2$, $\Theta_2 = 1 \cdot \mathbf{I}_2$. The *x₂ Inert Example* follows a setting where the unknown function does not depend on \mathbf{x}_2 , $\Psi(\mathbf{u}, \mathbf{v}) = \exp\{-(u_1 - v_2)^2\}$. Compared designs include *Nominal*, *Nominal Non-Stationary*, and *Regression* shown respectively in the left, middle, and right panels of Figure 1, *Numeric* and *Numeric Non-Stationary* shown respectively in the left and right panels of Figure 2, and *Parameter Estimation* shown in the right panel of Figure 4. Designs from the literature include random uniform design points, random Latin hypercube, maximin Latin hypercube, S-optimal [26] Latin hypercube, and MaxPro [25]. Comparison was performed by generating 500 draws from the corresponding Gaussian process with mean zero, then for each generating a random uniform design, a random Latin hypercube, a maximin Latin hypercube, an S-optimal Latin hypercube, and a MaxPro design as well as a 100 point random uniform testing set. For each Gaussian process draw emulators were built, for both fixed, known correlation parameters and estimated parameters, predictions generated on the testing set, and the *maximum* squared prediction error computed. The R packages SLHD [9], MaxPro [5], mlegp [10], and CGP [4] were used for generating Latin hypercube and MaxPro designs, fitting *stationary* Gaussian process emulators with *estimated correlation parameters*, and fitting composite Gaussian process emulators with *estimated correlation parameters*, respectively. All computation was performed in R 3.1.1 (R Core Team, Vienna, Austria).

Results are summarized in Table 1. Broadly, the proposed high-quality designs, as well as the MaxPro designs, performed well across situations, while random uniform design points, random Latin hypercube, maximin Latin hypercube, S-optimal Latin hypercube performed relatively poorly. There are a handful of deviations from and nuances to this basic pattern. First, designs which are tailor-built for non-stationarity *can* outperform space-filling designs (nominal and numeric/minimax and maximin), but do not always outperform them in practice when nominal, numeric, and potentially parameter estimation sources of inaccuracy come into play simultaneously. In particular, the influence of parameter estimation in the context of the composite Gaussian process is not clear, and could be a fruitful arena for future research. Second, in a situation where inert variables are expected, designs with very poor projection properties (such as the stationary numeric design) should be avoided. Poor performance of designs which emphasize low-dimensional projections is almost certainly *partly* due to the fact that no low-order functional ANOVA is present in the Gaussian process in the first four examples. However, lower dimensional projections of the *distance-based* criteria may be the relevant quantities. On the other hand, the results in Theorem 3.2 indicate that the nominal and parameter estimation sources of inaccuracy can be controlled via the *star discrepancy*.

This work has a several limitations. All results are in terms of *controlling* error rates with upper bounds. Actual error rates (of the nominal, numeric, or parameter estimation variety) could be substantially less in a particular situation. In a specific practical context, a sequential approach to data collection may be appropriate, see for example [23] or [16]. Further, no consideration is given to numeric error in parameter estimation and this error could be substantial, especially if the design is poor with respect to information about the parameters. However, given the secondary importance of experimental design properties

Table 1: Comparisons of *expected maximum* squared error of emulators based on nominal, numeric, parameter estimation, random uniform, random Latin hypercube, maximin Latin hypercube, S-optimal Latin hypercube, and MaxPro designs, potentially along with nominal non-stationary, regression, and numeric non-stationary, for several situations, stationary covariance, linear regression functions, non-stationary covariance, and inert prediction variables.

Stationary Example			Non-Stationary Nominal Example		
Design	True Parameters	Estimated Parameters	Design	True Parameters	Estimated Parameters
Nominal	2.54×10^{-9}	7.86×10^{-6}	Nominal	5.02	5.65
Numeric	4.48×10^{-10}	2.65×10^{-6}	Numeric	7.27	6.47
Parameter Estimation	5.56×10^{-10}	3.15×10^{-6}	Parameter Estimation	6.29	5.58
Random Uniform	4.67×10^{-6}	1.20×10^{-1}	Random Uniform	9.13	15.24
Random LH	1.81×10^{-6}	7.98×10^{-5}	Random LH	8.79	11.90
Maximin LH	5.06×10^{-8}	2.78×10^{-5}	Maximin LH	7.13	9.12
S-Optimal LH	5.50×10^{-8}	2.89×10^{-5}	S-Optimal LH	7.69	9.23
MaxPro	8.67×10^{-8}	4.58×10^{-5}	MaxPro	6.37	6.69
			Nominal Non-Stationary	4.57	10.04

Regression Functions Example			Non-Stationary Numeric Example		
Design	True Parameters	Estimated Parameters	Design	True Parameters	Estimated Parameters
Nominal	5.13×10^{-2}	1.25×10^{-2}	Nominal	7.13×10^{-1}	3.52
Numeric	8.52×10^{-3}	4.56×10^{-3}	Numeric	1.07	3.54
Parameter Estimation	1.41×10^{-2}	2.48×10^{-2}	Parameter Estimation	1.19	2.17
Random Uniform	9.00×10^{-1}	1.19	Random Uniform	6.80	1.63×10^1
Random LH	4.38×10^{-1}	8.33×10^{-2}	Random LH	5.05	1.08×10^1
Maximin LH	1.75×10^{-1}	3.69×10^{-2}	Maximin LH	2.66	7.30
S-Optimal LH	2.10×10^{-1}	3.24×10^{-1}	S-Optimal LH	3.17	7.93
MaxPro	1.20×10^{-1}	2.85×10^{-2}	MaxPro	1.38	3.53
Regression	1.43×10^{-2}	6.35×10^{-2}	Numeric Non-Stationary	9.30×10^{-1}	1.87

x_2 Inert Example		
Design	True Parameters	Estimated Parameters
Nominal	3.34×10^{-19}	2.23×10^{-5}
Numeric	3.78×10^{-19}	7.18
Parameter Estimation	2.56×10^{-19}	4.83×10^{-6}
Random Uniform	1.10×10^{-11}	3.20
Random LH	4.23×10^{-18}	1.71×10^{-3}
Maximin LH	1.33×10^{-18}	2.67×10^{-7}
S-Optimal LH	3.66×10^{-18}	2.37×10^{-5}
MaxPro	1.81×10^{-19}	2.30×10^{-8}

specific to parameter estimation, this source of error is not expected to strongly impact the error in interpolation. Also, the discussed model for non-stationarity is capable of approximating only non-constant correlation decay across the input space and, in particular, does not allow non-constant underlying variability in the Gaussian process model. However, non-constant underlying variability can be modeled as $\Psi(\mathbf{u}, \mathbf{v}) = \sigma(\mathbf{u})\sigma(\mathbf{v})\Phi(\mathbf{u} - \mathbf{v})$ and this non-stationary model behaves intuitively, with regions having more underlying variability requiring a higher density of points than regions having relatively less variability. The results follow in a manner similar to non-stationarity in correlation, although they are in fact simpler, and this development is omitted due to space constraints. Lastly, the impact on interpolator accuracy of a number important modeling and design considerations, such as low-order functional ANOVAs, orthogonality of inputs [7], or mixed categorical and quantitative input variables [33] has not been examined.

A Proof of Proposition 3.1

Express MSPE_2 from equation (3.1) in terms of partitioned matrices,

$$\text{MSPE}_2 = \Psi_{\theta}(\mathbf{x}, \mathbf{x}) - \begin{pmatrix} \mathbf{a}'_1 & \mathbf{a}'_2 \end{pmatrix} \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \end{pmatrix}, \quad (\text{A.1})$$

where

$$\begin{aligned} \mathbf{a}_1 &= \begin{pmatrix} \mathbf{h}(\mathbf{x}) \\ \Psi_\theta(\mathbf{X}_1, \mathbf{x}) \end{pmatrix}, \quad \mathbf{a}_2 = \Psi_\theta(\mathbf{X}_2^*, \mathbf{x}), \quad \mathbf{B}_{11} = \begin{pmatrix} \mathbf{0} & \mathbf{H}(\mathbf{X}_1)' \\ \mathbf{H}(\mathbf{X}_1) & \Psi_\theta(\mathbf{X}_1, \mathbf{X}_1) \end{pmatrix}, \\ \mathbf{B}_{12} &= \begin{pmatrix} \mathbf{H}(\mathbf{X}_2^*)' \\ \Psi_\theta(\mathbf{X}_1, \mathbf{X}_2^*) \end{pmatrix}, \quad \mathbf{B}_{21} = \mathbf{B}_{12}', \quad \mathbf{B}_{22} = \Psi_\theta(\mathbf{X}_2^*, \mathbf{X}_2^*), \quad \text{and} \quad \mathbf{X}_2^* = \mathbf{X}_2 \setminus \mathbf{X}_1. \end{aligned}$$

Applying partitioned matrix inverse results [19] and simplifying (A.1) gives,

$$\text{MSPE}_2 = \text{MSPE}_1 - (\mathbf{a}_2 - \mathbf{B}_{21}\mathbf{B}_{11}^{-1}\mathbf{a}_1)' \mathbf{B}_{22\cdot 1}^{-1} (\mathbf{a}_2 - \mathbf{B}_{21}\mathbf{B}_{11}^{-1}\mathbf{a}_1),$$

where $\mathbf{B}_{22\cdot 1} = \mathbf{B}_{22} - \mathbf{B}_{21}\mathbf{B}_{11}^{-1}\mathbf{B}_{12}$. Now, the proof is completed by showing that $\mathbf{B}_{22\cdot 1}$ is non-negative definite ($\mathbf{B}_{22\cdot 1} \succeq 0$). Once again applying partitioned matrix inverse results gives,

$$\begin{aligned} \mathbf{B}_{22\cdot 1} &= \Psi_\theta(\mathbf{X}_2^*, \mathbf{X}_2^*) - (\mathbf{H}(\mathbf{X}_2^*) \quad \Psi_\theta(\mathbf{X}_2^*, \mathbf{X}_1)) \begin{pmatrix} \mathbf{0} & \mathbf{H}(\mathbf{X}_1)' \\ \mathbf{H}(\mathbf{X}_1) & \Psi_\theta(\mathbf{X}_1, \mathbf{X}_1) \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{H}(\mathbf{X}_2^*)' \\ \Psi_\theta(\mathbf{X}_1, \mathbf{X}_2^*) \end{pmatrix} \\ &= \Psi_\theta(\mathbf{X}_2^*, \mathbf{X}_2^*) - \Psi_\theta(\mathbf{X}_2^*, \mathbf{X}_1)\Psi_\theta(\mathbf{X}_1, \mathbf{X}_1)^{-1}\Psi_\theta(\mathbf{X}_1, \mathbf{X}_2^*) \\ &\quad + (\mathbf{H}(\mathbf{X}_2^*) - \Psi_\theta(\mathbf{X}_2^*, \mathbf{X}_1)\Psi_\theta(\mathbf{X}_1, \mathbf{X}_1)^{-1}\mathbf{H}(\mathbf{X}_1)) (\mathbf{H}(\mathbf{X}_1)'\Psi_\theta(\mathbf{X}_1, \mathbf{X}_1)^{-1}\mathbf{H}(\mathbf{X}_1))^{-1} \\ &\quad \times (\mathbf{H}(\mathbf{X}_2^*) - \Psi_\theta(\mathbf{X}_2^*, \mathbf{X}_1)\Psi_\theta(\mathbf{X}_1, \mathbf{X}_1)^{-1}\mathbf{H}(\mathbf{X}_1))' \end{aligned}$$

The first two terms are non-negative definite because they represent a conditional variance. The third term is non-negative definite since $\mathbf{H}(\mathbf{X}_1)'\Psi_\theta(\mathbf{X}_1, \mathbf{X}_1)^{-1}\mathbf{H}(\mathbf{X}_1) \succeq 0$.

B Proof of Theorem 3.1

For $\mathbf{x}_i \in \mathbf{X}$, an arbitrary set $A \subset \Omega$, and positive definite function Ψ_θ , the uppermost terms in (3.2) can be *locally* bounded as

$$\begin{aligned} &\sup_{\mathbf{x} \in A} \Psi_\theta(\mathbf{x}, \mathbf{x}) - \Psi_\theta(\mathbf{x}, \mathbf{X})\Psi_\theta(\mathbf{X}, \mathbf{X})^{-1}\Psi_\theta(\mathbf{X}, \mathbf{x}) \\ &= \sup_{\mathbf{x} \in A} \Psi_\theta(\mathbf{x}, \mathbf{x}) - [(\Psi_\theta(\mathbf{x}, \mathbf{X}) - \Psi_\theta(\mathbf{x}_i, \mathbf{X}))\Psi_\theta(\mathbf{X}, \mathbf{X})^{-1}(\Psi_\theta(\mathbf{X}, \mathbf{x}) - \Psi_\theta(\mathbf{X}, \mathbf{x}_i)) \\ &\quad + 2\Psi_\theta(\mathbf{x}_i, \mathbf{X})\Psi_\theta(\mathbf{X}, \mathbf{X})^{-1}\Psi_\theta(\mathbf{X}, \mathbf{x}) - \Psi_\theta(\mathbf{x}_i, \mathbf{X})\Psi_\theta(\mathbf{X}, \mathbf{X})^{-1}\Psi_\theta(\mathbf{X}, \mathbf{x}_i)] \\ &= \sup_{\mathbf{x} \in A} \Psi_\theta(\mathbf{x}, \mathbf{x}) - 2\Psi_\theta(\mathbf{x}_i, \mathbf{x}) + \Psi_\theta(\mathbf{x}_i, \mathbf{x}_i) \\ &\quad - (\Psi_\theta(\mathbf{x}, \mathbf{X}) - \Psi_\theta(\mathbf{x}_i, \mathbf{X}))\Psi_\theta(\mathbf{X}, \mathbf{X})^{-1}(\Psi_\theta(\mathbf{X}, \mathbf{x}) - \Psi_\theta(\mathbf{X}, \mathbf{x}_i)) \\ &\leq \sup_{\mathbf{x} \in A} \Psi_\theta(\mathbf{x}, \mathbf{x}) - 2\Psi_\theta(\mathbf{x}_i, \mathbf{x}) + \Psi_\theta(\mathbf{x}_i, \mathbf{x}_i) - \frac{\|\Psi_\theta(\mathbf{X}, \mathbf{x}) - \Psi_\theta(\mathbf{X}, \mathbf{x}_i)\|_2^2}{\lambda_{\max}(\Psi_\theta(\mathbf{X}, \mathbf{X}))} \\ &\leq \sup_{\mathbf{x} \in A} \Psi_\theta(\mathbf{x}, \mathbf{x}) - 2\Psi_\theta(\mathbf{x}_i, \mathbf{x}) + \Psi_\theta(\mathbf{x}_i, \mathbf{x}_i) - \frac{(\Psi_\theta(\mathbf{x}_i, \mathbf{x}) - \Psi_\theta(\mathbf{x}_i, \mathbf{x}_i))^2}{n \sup_{\mathbf{u}, \mathbf{v} \in \Omega} \Psi_\theta(\mathbf{u}, \mathbf{v})}, \end{aligned} \tag{B.1}$$

where the first equality follows by cancellation of terms, the second equality follows from the fact that $\Psi_\theta(\mathbf{X}, \mathbf{X})^{-1}\Psi_\theta(\mathbf{X}, \mathbf{x}_i)$ equals the i^{th} component of the n dimensional identity, the first inequality follows from $\mathbf{a}'\mathbf{B}^{-1}\mathbf{a} \geq \lambda_{\min}(\mathbf{B}^{-1})\|\mathbf{a}\|_2^2$ and $\lambda_{\min}(\mathbf{B}^{-1}) = 1/\lambda_{\max}(\mathbf{B})$,

and the second inequality is true since a sum of squares $\|\cdot\|_2^2$ is larger than any one of its elements squared and Gershgorin's theorem [41] implies

$$\lambda_{\max}(\Psi_{\theta}(\mathbf{X}, \mathbf{X})) \leq \max_j \sum_{i=1}^n \Psi_{\theta}(\mathbf{x}_i, \mathbf{x}_j) \leq n \sup_{\mathbf{u}, \mathbf{v} \in \Omega} \Psi_{\theta}(\mathbf{u}, \mathbf{v}). \quad (\text{B.2})$$

Then for $\Psi_{\theta}(\mathbf{x}, \mathbf{x}) = \sigma^2$ across $\mathbf{x} \in \Omega$, and $k = n \sup_{\mathbf{u}, \mathbf{v} \in \Omega} \Psi_{\theta}(\mathbf{u}, \mathbf{v})$, the right-hand side of (B.1) can be rewritten as

$$\frac{1}{k} \sup_{\mathbf{x} \in A} (\sigma^2 - \Psi_{\theta}(\mathbf{x}_i, \mathbf{x})) (2k - \sigma^2 + \Psi_{\theta}(\mathbf{x}_i, \mathbf{x})). \quad (\text{B.3})$$

Expression (B.3) is a concave down quadratic in $\Psi_{\theta}(\mathbf{x}_i, \mathbf{x})$ with axis of symmetry $\sigma^2 - k$, which is ≤ 0 for $n \geq 1$. That is, (B.1) is bounded above by

$$\frac{1}{k} \left(\sigma^2 - \inf_{\mathbf{x} \in A} \Psi_{\theta}(\mathbf{x}_i, \mathbf{x}) \right) \left(2k - \sigma^2 + \inf_{\mathbf{x} \in A} \Psi_{\theta}(\mathbf{x}_i, \mathbf{x}) \right).$$

If $\Omega \subseteq \cup_{i=1}^n A_i$ and $\mathbf{x}_i \in A_i$, then

$$\begin{aligned} & \sup_{\mathbf{x} \in \Omega} \Psi_{\theta}(\mathbf{x}, \mathbf{x}) - \Psi_{\theta}(\mathbf{x}, \mathbf{X}) \Psi_{\theta}(\mathbf{X}, \mathbf{X})^{-1} \Psi_{\theta}(\mathbf{X}, \mathbf{x}) \\ & \leq \frac{1}{k} \left(\sigma^2 - \min_i \inf_{\mathbf{x} \in A_i} \Psi_{\theta}(\mathbf{x}_i, \mathbf{x}) \right) \left(2k - \sigma^2 + \min_i \inf_{\mathbf{x} \in A_i} \Psi_{\theta}(\mathbf{x}_i, \mathbf{x}) \right). \end{aligned}$$

C Proof of Theorem 3.2

If $f \sim \text{GP}(0, \Psi(\cdot, \cdot))$, then f can be represented as $f = \sum_{i=1}^{\infty} \langle f, \varphi_i \rangle \varphi_i$, where $\langle \cdot, \cdot \rangle$ is the inner product in $L_2([0, 1]^d)$, $\langle f, \varphi_i \rangle \stackrel{\text{ind.}}{\sim} \mathcal{N}(0, \lambda_i)$, and $\Psi(\cdot, \cdot)$ has eigenvalue, eigenfunction decomposition $\Psi(\cdot, \cdot) = \sum_{i=1}^{\infty} \lambda_i \varphi_i(\cdot) \varphi_i(\cdot)$. Further, $\hat{f}(\mathbf{x}) = \Psi(\mathbf{x}, \mathbf{X}) \Psi(\mathbf{X}, \mathbf{X})^{-1} f(\mathbf{X})$ minimizes the mean squared prediction error (MSPE) $\mathbb{E} \left(f(\mathbf{x}) - \hat{f}(\mathbf{x}) \right)^2$ over functions of the data \mathbf{X} , $f(\mathbf{X})$. Consider another predictor based on the data

$$\hat{f}_*(x) = \sum_{i=1}^n \hat{u}_i \varphi_i(x), \quad \hat{u}_i = \frac{1}{n} \sum_{j=1}^n \varphi_i(x_j) f(x_j).$$

Then, the *integrated* MSPE of \hat{f} is bounded above by the *integrated* MSPE of \hat{f}_* ,

$$\begin{aligned} \mathbb{E} \|f - \hat{f}_*\|_{L_2([0, 1]^d)}^2 &= \mathbb{E} \left\| \sum_{i=1}^{\infty} \langle f, \varphi_i \rangle \varphi_i(\cdot) - \sum_{i=1}^n \hat{u}_i \varphi_i(\cdot) \right\|_{L_2([0, 1]^d)}^2 \\ &= \sum_{j=n+1}^{\infty} \lambda_j + \sum_{j=1}^n \mathbb{E} \left(\int_{[0, 1]^d} \varphi_j(x) f(x) dx - \frac{1}{n} \sum_{j=1}^n \varphi_j(x_j) f(x_j) \right)^2. \end{aligned}$$

Respective Koksma-Hlawka and modulus of continuity bounds on the error rate for numeric integration [31] give the bounds

$$\left| \int_{[0,1]^d} \varphi_i(x) f(x) dx - \frac{1}{n} \sum_{j=1}^n \varphi_i(x_j) f(x_j) \right| \leq V(\varphi_i f) D^*(X),$$

$$\left| \int_{[0,1]^d} \varphi_i(x) f(x) dx - \frac{1}{n} \sum_{j=1}^n \varphi_i(x_j) f(x_j) \right| \leq 4\omega(\varphi_i f, D^*(X)^{1/d}),$$

where $V(\cdot)$ denotes the total variation in the sense of Hardy and Kraus of f , $D^*(X)$ denotes the star discrepancy, and ω denotes the modulus of continuity [31].

D Proof of Proposition 4.1

We will make use of the following lemma on the accuracy of floating point matrix inversion which is a combination of Lemmas 2.7.1 and 2.7.2 in [15].

Lemma D.1. *Suppose $\mathbf{Ax} = \mathbf{b}$ and $\tilde{\mathbf{A}}\tilde{\mathbf{x}} = \tilde{\mathbf{b}}$ with $\|\mathbf{A} - \tilde{\mathbf{A}}\|_2 \leq \delta\|\mathbf{A}\|_2$, $\|\mathbf{b} - \tilde{\mathbf{b}}\|_2 \leq \delta\|\mathbf{b}\|_2$, and $\kappa(\mathbf{A}) = r/\delta < 1/\delta$ for some $\delta > 0$. Then, $\tilde{\mathbf{A}}$ is non-singular,*

$$\frac{\|\tilde{\mathbf{x}}\|_2}{\|\mathbf{x}\|_2} \leq \frac{1+r}{1-r}, \quad \text{and} \quad \frac{\|\mathbf{x} - \tilde{\mathbf{x}}\|_2}{\|\mathbf{x}\|_2} \leq \frac{2\delta}{1-r}\kappa(\mathbf{A}),$$

where $\kappa(\mathbf{A}) = \|\mathbf{A}\|_2\|\mathbf{A}^{-1}\|_2$.

Additionally, note that for (conformable) \mathbf{a} , $\tilde{\mathbf{a}}$, \mathbf{b} , and $\tilde{\mathbf{b}}$,

$$\begin{aligned} \left| \mathbf{a}'\mathbf{b} - \tilde{\mathbf{a}}'\tilde{\mathbf{b}} \right| &= \left| \mathbf{a}'(\mathbf{b} - \tilde{\mathbf{b}}) - (\tilde{\mathbf{a}} - \mathbf{a})'\tilde{\mathbf{b}} \right| \\ &\leq \left| \mathbf{a}'(\mathbf{b} - \tilde{\mathbf{b}}) \right| + \left| (\tilde{\mathbf{a}} - \mathbf{a})'\tilde{\mathbf{b}} \right| \leq \|\mathbf{a}\|_2\|\mathbf{b} - \tilde{\mathbf{b}}\|_2 + \|\tilde{\mathbf{a}} - \mathbf{a}\|_2\|\tilde{\mathbf{b}}\|_2. \end{aligned} \tag{D.1}$$

The inner portion of the numeric error in the second term in (2.4) can be bounded as follows. Here, and throughout, $\mathbf{A}^{-1}\mathbf{b}$ and $\tilde{\mathbf{A}}^{-1}\tilde{\mathbf{b}}$ denote the solutions to the linear systems $\mathbf{Ax} = \mathbf{b}$ and $\tilde{\mathbf{A}}\tilde{\mathbf{x}} = \tilde{\mathbf{b}}$, respectively, as opposed to the actual matrix multiplication. The *hats* on parameter estimates are suppressed for simplicity. In fact, all the below results hold for an arbitrary, fixed parameter or parameter estimate.

$$\begin{aligned} &\left| \hat{f}_\theta(\mathbf{x}) - \tilde{f}_\theta(\mathbf{x}) \right| \\ &= \left| \left(\mathbf{h}(\mathbf{x}) - \tilde{\mathbf{h}}(\mathbf{x}) \right)' \beta - \left(\Psi_\theta(\mathbf{x}, \mathbf{X}) \Psi_\theta(\mathbf{X}, \mathbf{X})^{-1} \mathbf{H}(\mathbf{X}) - \tilde{\Psi}_\theta(\mathbf{x}, \mathbf{X}) \tilde{\Psi}_\theta(\mathbf{X}, \mathbf{X})^{-1} \tilde{\mathbf{H}}(\mathbf{X}) \right) \beta \right. \\ &\quad \left. + \left(\Psi_\theta(\mathbf{x}, \mathbf{X}) \Psi_\theta(\mathbf{X}, \mathbf{X})^{-1} f(\mathbf{X}) - \tilde{\Psi}_\theta(\mathbf{x}, \mathbf{X}) \tilde{\Psi}_\theta(\mathbf{X}, \mathbf{X})^{-1} \tilde{f}(\mathbf{X}) \right) \right| \\ &\leq \|\mathbf{h}(\mathbf{x}) - \tilde{\mathbf{h}}(\mathbf{x})\|_2 \|\beta\|_2 + \left| \Psi_\theta(\mathbf{x}, \mathbf{X}) \Psi_\theta(\mathbf{X}, \mathbf{X})^{-1} f(\mathbf{X}) - \tilde{\Psi}_\theta(\mathbf{x}, \mathbf{X}) \tilde{\Psi}_\theta(\mathbf{X}, \mathbf{X})^{-1} \tilde{f}(\mathbf{X}) \right| \\ &\quad + \left\| \Psi_\theta(\mathbf{x}, \mathbf{X}) \Psi_\theta(\mathbf{X}, \mathbf{X})^{-1} \mathbf{H}(\mathbf{X}) - \tilde{\Psi}_\theta(\mathbf{x}, \mathbf{X}) \tilde{\Psi}_\theta(\mathbf{X}, \mathbf{X})^{-1} \tilde{\mathbf{H}}(\mathbf{X}) \right\|_2 \|\beta\|_2 \\ &= \|\mathbf{h}(\mathbf{x}) - \tilde{\mathbf{h}}(\mathbf{x})\|_2 \|\beta\|_2 + \left| \Psi_\theta(\mathbf{x}, \mathbf{X}) \Psi_\theta(\mathbf{X}, \mathbf{X})^{-1} f(\mathbf{X}) - \tilde{\Psi}_\theta(\mathbf{x}, \mathbf{X}) \tilde{\Psi}_\theta(\mathbf{X}, \mathbf{X})^{-1} \tilde{f}(\mathbf{X}) \right| \\ &\quad + \sqrt{\sum_{j=1}^p \left(\Psi_\theta(\mathbf{x}, \mathbf{X}) \Psi_\theta(\mathbf{X}, \mathbf{X})^{-1} h_j(\mathbf{X}) - \tilde{\Psi}_\theta(\mathbf{x}, \mathbf{X}) \tilde{\Psi}_\theta(\mathbf{X}, \mathbf{X})^{-1} \tilde{h}_j(\mathbf{X}) \right)^2} \|\beta\|_2, \end{aligned} \tag{D.2}$$

where $h_j(\mathbf{X})$ and $\tilde{h}_j(\mathbf{X})$ denote the j^{th} regression function evaluated at \mathbf{X} and its floating point approximation, respectively. Let $\mathbf{u} = \Psi_{\theta}(\mathbf{X}, \mathbf{X})^{-1}f(\mathbf{X})$, $\tilde{\mathbf{u}} = \tilde{\Psi}_{\theta}(\mathbf{X}, \mathbf{X})^{-1}\tilde{f}(\mathbf{X})$, $\mathbf{v}_j = \Psi_{\theta}(\mathbf{X}, \mathbf{X})^{-1}h_j(\mathbf{X})$, and $\tilde{\mathbf{v}}_j = \tilde{\Psi}_{\theta}(\mathbf{X}, \mathbf{X})^{-1}\tilde{h}_j(\mathbf{X})$. Then, (D.2) along with inequality (D.1) implies

$$\begin{aligned} & \left| \hat{f}_{\theta}(\mathbf{x}) - \tilde{f}_{\theta}(\mathbf{x}) \right| \\ & \leq \|\mathbf{h}(\mathbf{x}) - \tilde{\mathbf{h}}(\mathbf{x})\|_2 \|\beta\|_2 + \|\Psi_{\theta}(\mathbf{x}, \mathbf{X})\|_2 \|\mathbf{u} - \tilde{\mathbf{u}}\|_2 + \|\Psi_{\theta}(\mathbf{x}, \mathbf{X}) - \tilde{\Psi}_{\theta}(\mathbf{x}, \mathbf{X})\|_2 \|\tilde{\mathbf{u}}\|_2 \\ & \quad + \sqrt{\sum_{j=1}^p \left(\|\Psi_{\theta}(\mathbf{x}, \mathbf{X})\|_2 \|\mathbf{v}_j - \tilde{\mathbf{v}}_j\|_2 + \|\Psi_{\theta}(\mathbf{x}, \mathbf{X}) - \tilde{\Psi}_{\theta}(\mathbf{x}, \mathbf{X})\|_2 \|\tilde{\mathbf{v}}_j\|_2 \right)^2} \|\beta\|_2. \end{aligned} \quad (\text{D.3})$$

Under Assumption 4.1, Lemma D.1 can be applied to (D.3) to obtain

$$\begin{aligned} & \left| \hat{f}_{\theta}(\mathbf{x}) - \tilde{f}_{\theta}(\mathbf{x}) \right| \\ & \leq \delta \|\mathbf{h}(\mathbf{x})\|_2 \|\beta\|_2 + \|\Psi_{\theta}(\mathbf{x}, \mathbf{X})\|_2 \frac{2\delta}{1-r} \kappa(\Psi_{\theta}(\mathbf{X}, \mathbf{X})) \|\mathbf{u}\|_2 + \delta \|\Psi_{\theta}(\mathbf{x}, \mathbf{X})\|_2 \frac{1+r}{1-r} \|\mathbf{u}\|_2 \\ & \quad + \sqrt{\sum_{j=1}^p \left(\|\Psi_{\theta}(\mathbf{x}, \mathbf{X})\|_2 \frac{2\delta}{1-r} \kappa(\Psi_{\theta}(\mathbf{X}, \mathbf{X})) \|\mathbf{v}_j\|_2 + \delta \|\Psi_{\theta}(\mathbf{x}, \mathbf{X})\|_2 \frac{1+r}{1-r} \|\mathbf{v}_j\|_2 \right)^2} \|\beta\|_2. \end{aligned}$$

Note that $\|\mathbf{u}\|_2 = \|\Psi_{\theta}(\mathbf{X}, \mathbf{X})^{-1}f(\mathbf{X})\|_2 \leq \|\Psi_{\theta}(\mathbf{X}, \mathbf{X})^{-1}\|_2 \|f(\mathbf{X})\|_2 = \|f(\mathbf{X})\|_2 / \lambda_{\min}(\Psi_{\theta}(\mathbf{X}, \mathbf{X}))$. Similarly, $\|\mathbf{v}_j\|_2 \leq \|h_j(\mathbf{X})\|_2 / \lambda_{\min}(\Psi_{\theta}(\mathbf{X}, \mathbf{X}))$. Using these facts along with $r < 1$ and grouping terms gives the proposition.

E Proof of Theorem 4.1

For a continuous, positive definite, translation invariant kernel Φ which has Fourier transform $\hat{\Phi} \in L_1(\mathbb{R}^d)$,

$$\begin{aligned} \sum_{j=1}^n \sum_{k=1}^n \alpha_j \alpha_k \Phi(\mathbf{x}_j - \mathbf{x}_k) &= (2\pi)^{-d/2} \sum_{j=1}^n \sum_{k=1}^n \alpha_j \alpha_k \int_{\mathbb{R}^d} e^{i\omega'(\mathbf{x}_j - \mathbf{x}_k)} \hat{\Phi}(\omega) d\omega \\ &= (2\pi)^{-d/2} \int_{\mathbb{R}^d} \left| \sum_{j=1}^n \alpha_j e^{i\omega' \mathbf{x}_j} \right|^2 \hat{\Phi}(\omega) d\omega, \end{aligned} \quad (\text{E.1})$$

for $\alpha \in \mathbb{R}^n$, $\mathbf{x}_i \in \mathbb{R}^d$. Representation (E.1) implies that a lower bound for $\sum_{j,k} \alpha_j \alpha_k \Phi(\mathbf{x}_j - \mathbf{x}_k)$ is provided by $\sum_{j,k} \alpha_j \alpha_k \Upsilon(\mathbf{x}_j - \mathbf{x}_k)$, where Υ has $\hat{\Upsilon}(\omega) \leq \hat{\Phi}(\omega)$. Consider Υ_M with

$$\hat{\Upsilon}_M(\omega) = \frac{\hat{\Phi}_*(M) \Gamma(d/2 + 1)}{2^d M^d \pi^{d/2}} (\chi_M * \chi_M)(\omega),$$

where $M > 0$, $\hat{\Phi}_*(M) = \inf_{\|\omega\|_2 \leq 2M} \hat{\Phi}(\omega)$, $\chi_M(\omega) = 1$ for $\|\omega\|_2 \leq M$ and 0 otherwise, and $*$ denotes the convolution operator

$$(f * g)(\mathbf{x}) = \int_{\mathbb{R}^d} f(\mathbf{y}) g(\mathbf{x} - \mathbf{y}) d\mathbf{y}.$$

For $\|\boldsymbol{\omega}\|_2 > 2M$, $\hat{\Upsilon}_M(\boldsymbol{\omega}) = 0 \leq \hat{\Phi}(\boldsymbol{\omega})$. On the other hand, for $\|\boldsymbol{\omega}\|_2 \leq 2M$,

$$\begin{aligned}\hat{\Upsilon}_M(\boldsymbol{\omega}) &= \frac{\hat{\Phi}_*(M)\Gamma(d/2+1)}{2^d M^d \pi^{d/2}} \int_{\mathbb{R}^d} \chi_M(\mathbf{t}) \chi_M(\boldsymbol{\omega} - \mathbf{t}) d\mathbf{t} \\ &\leq \frac{\hat{\Phi}_*(M)\Gamma(d/2+1)}{2^d M^d \pi^{d/2}} \text{vol } B(\mathbf{0}, 2M) = \hat{\Phi}_*(M) \leq \hat{\Phi}(\boldsymbol{\omega}),\end{aligned}$$

where $B(\mathbf{0}, 2M) = \{\|\boldsymbol{\omega} - \mathbf{0}\|_2 \leq 2M\}$ denotes a ball of radius $2M$ centered at the origin. So, $\hat{\Upsilon}(\boldsymbol{\omega}) \leq \hat{\Phi}(\boldsymbol{\omega})$ for all $\boldsymbol{\omega} \in \mathbb{R}^d$. The candidate Υ can be recovered from the inverse Fourier transform $(\hat{\Upsilon}_M)^\vee$

$$\begin{aligned}\Upsilon_M(\mathbf{t}) &= \frac{\hat{\Phi}_*(M)\Gamma(d/2+1)}{2^d M^d \pi^{d/2}} (\chi_M * \chi_M)^\vee(\mathbf{t}) \\ &= \frac{\hat{\Phi}_*(M)\Gamma(d/2+1)}{2^d M^d \pi^{d/2}} (2\pi)^{d/2} ((\chi_M)^\vee(\mathbf{t}))^2 \\ &= \frac{\hat{\Phi}_*(M)\Gamma(d/2+1)}{2^{d/2}} \|\mathbf{t}\|_2^{-d} J_{d/2}^2(M\|\mathbf{t}\|_2),\end{aligned}$$

where J_ν is a Bessel function of the first kind. A proof of the final equality is given as Lemma 12.2 in [43]. Define $\Upsilon_M(\mathbf{0})$ as

$$\Upsilon_M(\mathbf{0}) \equiv \lim_{\mathbf{t} \rightarrow \mathbf{0}} \Upsilon_M(\mathbf{t}) = \frac{\hat{\Phi}_*(M)}{\Gamma(d/2+1)} \left(\frac{M}{2^{3/2}} \right)^d.$$

This limit follows from the Taylor series representation of the Bessel function [42].

Now, a lower bound on the quadratic form involving Υ is developed.

$$\begin{aligned}\sum_{j=1}^n \sum_{k=1}^n \alpha_j \alpha_k \Upsilon_M(\mathbf{x}_j - \mathbf{x}_k) &= \sum_{j=1}^n \alpha_j^2 \Upsilon_M(\mathbf{0}) + \sum_{j \neq k} \alpha_j \alpha_k \Upsilon_M(\mathbf{x}_j - \mathbf{x}_k) \\ &\geq \sum_{j=1}^n \alpha_j^2 \Upsilon_M(\mathbf{0}) - \sum_{j \neq k} |\alpha_j| |\alpha_k| \Upsilon_M(\mathbf{x}_j - \mathbf{x}_k) \\ &\geq \sum_{j=1}^n \alpha_j^2 \Upsilon_M(\mathbf{0}) - \frac{1}{2} \sum_{j \neq k} (\alpha_j^2 + \alpha_k^2) \Upsilon_M(\mathbf{x}_j - \mathbf{x}_k) \\ &= \sum_{j=1}^n \alpha_j^2 \Upsilon_M(\mathbf{0}) - \sum_{j=1}^n \alpha_j^2 \sum_{k=1, k \neq j}^n \Upsilon_M(\mathbf{x}_j - \mathbf{x}_k) \\ &= \sum_{j=1}^n \alpha_j^2 \left(\Upsilon_M(\mathbf{0}) - \sum_{k=1, k \neq j}^n \Upsilon_M(\mathbf{x}_j - \mathbf{x}_k) \right).\end{aligned}\tag{E.2}$$

Each $\sum_{k=1, k \neq j}^n |\Upsilon(\mathbf{x}_j - \mathbf{x}_k)|$ can be bounded in terms of the separation distances

$$q_j = \frac{1}{2} \min_{k=1, \dots, n, k \neq j} \|\mathbf{x}_j - \mathbf{x}_k\|_2 \quad \text{and} \quad q = \min_j q_j.$$

For $m \in \mathbb{N}$, let

$$E_{jm} = \{\mathbf{x} \in \mathbb{R}^d : mq_j \leq \|\mathbf{x}_j - \mathbf{x}\|_2 < (m+1)q_j\}.$$

Then, every \mathbf{x}_k , $k \neq j$ is contained in exactly one E_{jm} . Further, every $B(\mathbf{x}_k, q)$ is essentially disjoint and completely contained in

$$\{\mathbf{x} \in \mathbb{R}^d : mq_j - q \leq \|\mathbf{x}_j - \mathbf{x}\|_2 < (m+1)q_j + q\}.$$

So, each E_{jm} can contain no more than

$$\frac{((m+1)q_j + q)^d - (mq_j - q)^d}{q^d} = \left(\frac{(m+1)q_j}{q} + 1\right)^d - \left(\frac{mq_j}{q} - 1\right)^d$$

data points. We now make use of the following lemma.

Lemma E.1. For $d \in \mathbb{N}$ and $q_j \geq q > 0$,

$$\left(\frac{(m+1)q_j}{q} + 1\right)^d - \left(\frac{mq_j}{q} - 1\right)^d \leq (3q_j/q)^d m^{d-1}.$$

Proof. Take $d = 1$, then

$$\left(\frac{(m+1)q_j}{q} + 1\right) - \left(\frac{mq_j}{q} - 1\right) = 2 + q_j/q \leq 3q_j/q.$$

Now, assume the result is true for $1 \leq d_* < d$. Let $c = 3q_j/q$. Then,

$$\begin{aligned} & \left(\frac{(m+1)q_j}{q} + 1\right)^{d-1} - \left(\frac{mq_j}{q} - 1\right)^{d-1} \leq c^{d-1} m^{d-2} \\ \implies & \left(\frac{(m+1)q_j}{q} + 1\right)^d - \left(\frac{(m+1)q_j}{q} + 1\right) \left(\frac{mq_j}{q} - 1\right)^{d-1} \leq \left(\frac{(m+1)q_j}{q} + 1\right) c^{d-1} m^{d-2} \\ \implies & \left(\frac{(m+1)q_j}{q} + 1\right)^d - \left(\frac{mq_j}{q} - 1\right)^d \leq \left(\frac{(m+1)q_j}{q} + 1\right) c^{d-1} m^{d-2} + \left(\frac{q_j}{q} + 2\right) \left(\frac{mq_j}{q} - 1\right)^{d-1}. \end{aligned}$$

The proof is completed by showing that the right-hand side of the final inequality is bounded above by $c^d m^{d-1}$. The right-hand side can be represented in terms of c as

$$\begin{aligned} & \left(\frac{(m+1)c}{3} + 1\right) c^{d-1} m^{d-2} + \left(\frac{c}{3} + 2\right) \left(\frac{mc}{3} - 1\right)^{d-1} \\ & = c^d m^{d-1} \left[\frac{(m+1)}{3m} + \frac{1}{mc} + \left(\frac{1}{3} + \frac{2}{c}\right) \left(\frac{1}{3} - \frac{1}{mc}\right)^{d-1} \right] \\ & \leq c^d m^{d-1} \left[\frac{1}{3} + \frac{1}{3m} + \frac{1}{mc} + \left(\frac{1}{3} - \frac{1}{mc}\right)^{d-1} \right] \\ & \leq c^d m^{d-1} \left[\frac{1}{3} + \frac{1}{3m} + \frac{1}{mc} + \frac{1}{3} - \frac{1}{mc} \right] \\ & = c^d m^{d-1} \left[\frac{2}{3} + \frac{1}{3m} \right] \leq c^d m^{d-1}, \end{aligned}$$

where the first inequality is true because $1/3 + 2/c \leq 1$ and the second inequality is true because $(1/3 - 1/(mc))^{d-1}$ is a decreasing function of $d \geq 2$. \square

Lemma E.1 implies that each E_{jm} contains no more than $(3q_j/q)^d m^{d-1}$ points. Note that on E_{jm} , $\Upsilon(\mathbf{x}_j - \mathbf{x}_k)$ is bounded above as

$$\begin{aligned}
\Upsilon(\mathbf{x}_j - \mathbf{x}_k) &= \frac{\hat{\Phi}_*(M)\Gamma(d/2 + 1)}{2^{d/2}} \|\mathbf{x}_j - \mathbf{x}_k\|_2^{-d} J_{d/2}^2(M\|\mathbf{x}_j - \mathbf{x}_k\|_2) \\
&\leq \frac{\hat{\Phi}_*(M)\Gamma(d/2 + 1)}{2^{d/2}} \|\mathbf{x}_j - \mathbf{x}_k\|_2^{-d} \frac{2^{d+2}}{M\pi\|\mathbf{x}_j - \mathbf{x}_k\|_2} \\
&= \Upsilon_M(\mathbf{0}) \frac{\Gamma^2(d/2 + 1)}{\pi} \left(\frac{4}{M\|\mathbf{x}_j - \mathbf{x}_k\|_2} \right)^{d+1} \\
&\leq \Upsilon_M(\mathbf{0}) \frac{\Gamma^2(d/2 + 1)}{\pi} \left(\frac{4}{Mmq_j} \right)^{d+1},
\end{aligned} \tag{E.3}$$

where the first inequality follows from the Bessel function bound provided in Lemma 3.3 of [30]. Combining Lemma E.1 with (E.3) gives

$$\begin{aligned}
\sum_{k=1, k \neq j}^n \Upsilon(\mathbf{x}_j - \mathbf{x}_k) &\leq \sum_{m=1}^{\infty} \Upsilon_M(\mathbf{0}) \frac{\Gamma^2(d/2 + 1)}{\pi} \left(\frac{4}{Mmq_j} \right)^{d+1} (3q_j/q)^d m^{d-1} \\
&= \Upsilon_M(\mathbf{0}) \frac{\Gamma^2(d/2 + 1)\pi}{18} \left(\frac{q}{q_j} \right) \left(\frac{12}{Mq} \right)^{d+1},
\end{aligned}$$

where the equality follows from the fact that $\sum_{m=1}^{\infty} m^{-2} = \pi^2/6$. Now, taking $M = c_*/q$ and referring back to (E.2), the quadratic form can be bounded as

$$\sum_{j=1}^n \sum_{k=1}^n \alpha_j \alpha_k \Phi(\mathbf{x}_j - \mathbf{x}_k) \geq \Upsilon_{c_*/q}(\mathbf{0}) \sum_{j=1}^n \alpha_j^2 \left(1 - \frac{\Gamma^2(d/2 + 1)\pi}{18} \left(\frac{q}{q_j} \right) \left(\frac{12}{c_*} \right)^{d+1} \right).$$

The stated version of the theorem follows by applying the previous development to the transformed space, $\mathbf{v} \mapsto \mathbf{v}^* = \Theta \mathbf{v}$.

F Proof of Theorem 5.1

First, we develop expressions for the components of the approximate mean squared prediction error in Lemma F.1 below.

Lemma F.1. *If $f \sim \text{GP}(\mathbf{h}(\cdot)'\boldsymbol{\beta}, \Psi_{\boldsymbol{\theta}}(\cdot, \cdot))$, for fixed, known regression functions $\mathbf{h}(\cdot)$, and $\Psi_{\boldsymbol{\theta}}(\cdot, \cdot)$ as defined in (5.2), then*

$$\frac{\partial f_{\boldsymbol{\theta}}(\mathbf{x})}{\partial \boldsymbol{\theta}'} \mathcal{I}(\boldsymbol{\theta})^{-1} \frac{\partial f_{\boldsymbol{\theta}}(\mathbf{x})}{\partial \boldsymbol{\theta}} = \mathbf{c}'_1 \mathcal{I}_{11}^{-1} \mathbf{c}_1 + \mathbf{c}'_3 (\mathcal{I}_{33} - \mathcal{I}_{32} \mathcal{I}_{22}^{-1} \mathcal{I}_{23})^{-1} \mathbf{c}_3, \tag{F.1}$$

for $\mathbf{x} \in \mathbb{R}^d$ where

$$\begin{aligned}
\mathbf{c}_1 &= \mathbf{h}(\mathbf{x}) - \mathbf{H}(\mathbf{X})' \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{x}), \\
\mathbf{c}_3 &= \left(\frac{\partial \Psi_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{X})}{\partial \boldsymbol{\rho}} - (\mathbf{I}_d \otimes \Psi_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{X}) \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1}) \frac{\partial \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})}{\partial \boldsymbol{\rho}} \right) \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} \delta(\mathbf{X}), \\
\mathcal{I}_{11} &= \mathbf{H}(\mathbf{X})' \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} \mathbf{H}(\mathbf{X}), \\
\mathcal{I}_{22} &= \frac{n}{2\sigma^4}, \\
\mathcal{I}_{32} &= \frac{1}{2\sigma^4} \mathbf{C}'_{\boldsymbol{\theta}} \text{vec } \Phi_{\boldsymbol{\rho}}(\mathbf{X}, \mathbf{X})^{-1}, \\
\mathcal{I}_{33} &= \frac{1}{2\sigma^4} \mathbf{C}'_{\boldsymbol{\theta}} (\Phi_{\boldsymbol{\rho}}(\mathbf{X}, \mathbf{X})^{-1} \otimes \Phi_{\boldsymbol{\rho}}(\mathbf{X}, \mathbf{X})^{-1}) \mathbf{C}_{\boldsymbol{\theta}}, \\
\mathbf{C}_{\boldsymbol{\theta}} &= \frac{\partial(\text{vec } \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X}))}{\partial \boldsymbol{\rho}'},
\end{aligned}$$

where $\delta(\mathbf{X}) = f(\mathbf{X}) - \mathbf{H}(\mathbf{X})\boldsymbol{\beta}$.

Proof. Up to an additive constant, the log-likelihood is

$$\ell = -\frac{1}{2} \log \det \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X}) - \frac{1}{2} (f(\mathbf{X}) - \mathbf{H}(\mathbf{X})\boldsymbol{\beta})' \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} (f(\mathbf{X}) - \mathbf{H}(\mathbf{X})\boldsymbol{\beta}).$$

Throughout Appendix F, we will use matrix differentiation, see for example [27]. Then, the vector of derivatives of the emulator with respect to the unknown parameter values $\frac{\partial \hat{f}_{\boldsymbol{\theta}}(\mathbf{x})}{\partial \boldsymbol{\theta}}$ has block components

$$\begin{aligned}
\mathbf{c}_1 &= \frac{\partial \hat{f}_{\boldsymbol{\theta}}(\mathbf{x})}{\partial \boldsymbol{\beta}} = \frac{\partial}{\partial \boldsymbol{\beta}} \{ \mathbf{h}(\mathbf{x})' \boldsymbol{\beta} + \Psi_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{X}) \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} (f(\mathbf{X}) - \mathbf{H}(\mathbf{X})\boldsymbol{\beta}) \} \\
&= \mathbf{h}(\mathbf{x}) - \mathbf{H}(\mathbf{X})' \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{x}), \\
\mathbf{c}_2 &= \frac{\partial \hat{f}_{\boldsymbol{\theta}}(\mathbf{x})}{\partial \sigma^2} = 0.
\end{aligned}$$

Developing an expression for $\frac{\partial \hat{f}_{\boldsymbol{\theta}}(\mathbf{x})}{\partial \boldsymbol{\rho}}$ is more complex and broken into a few parts. Let $\delta(\mathbf{X}) = f(\mathbf{X}) - \mathbf{H}(\mathbf{X})\boldsymbol{\beta}$. Then,

$$\mathbf{c}_3 = \frac{\partial \hat{f}_{\boldsymbol{\theta}}(\mathbf{x})}{\partial \boldsymbol{\rho}} = \left(\frac{\partial \Psi_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{X})}{\partial \boldsymbol{\rho}} \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} + (\mathbf{I}_d \otimes \Psi_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{X})) \frac{\partial \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1}}{\partial \boldsymbol{\rho}} \right) \delta(\mathbf{X}). \quad (\text{F.2})$$

Note that,

$$\begin{aligned}
\mathbf{0} &= \frac{\partial \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X}) \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1}}{\partial \boldsymbol{\rho}} \\
&= \frac{\partial \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})}{\partial \boldsymbol{\rho}} \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} + (\mathbf{I}_d \otimes \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})) \frac{\partial \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1}}{\partial \boldsymbol{\rho}}.
\end{aligned}$$

So,

$$\frac{\partial \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1}}{\partial \boldsymbol{\rho}} = -(\mathbf{I}_d \otimes \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1}) \frac{\partial \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})}{\partial \boldsymbol{\rho}} \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} \quad (\text{F.3})$$

Plugging (F.3) into equation (F.2) gives the third block component

$$\begin{aligned} \mathbf{c}_3 &= \frac{\partial \hat{f}_\theta(\mathbf{x})}{\partial \boldsymbol{\varrho}} \\ &= \left(\frac{\partial \Psi_\theta(\mathbf{x}, \mathbf{X})}{\partial \boldsymbol{\varrho}} - (\mathbf{I}_d \otimes \Psi_\theta(\mathbf{x}, \mathbf{X}) \Psi_\theta(\mathbf{X}, \mathbf{X})^{-1}) \frac{\partial \Psi_\theta(\mathbf{X}, \mathbf{X})}{\partial \boldsymbol{\varrho}} \right) \Psi_\theta(\mathbf{X}, \mathbf{X})^{-1} \delta(\mathbf{X}). \end{aligned}$$

Now, we develop an expression for $\mathcal{I}(\boldsymbol{\vartheta}_*)$. First,

$$\begin{aligned} \frac{\partial \ell}{\partial \boldsymbol{\beta}} &= \mathbf{H}(\mathbf{X})' \Psi_\theta(\mathbf{X}, \mathbf{X})^{-1} (f(\mathbf{X}) - \mathbf{H}(\mathbf{X})\boldsymbol{\beta}), \\ \frac{\partial \ell}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} (f(\mathbf{X}) - \mathbf{H}(\mathbf{X})\boldsymbol{\beta})' \Phi_\theta(\mathbf{X}, \mathbf{X})^{-1} (f(\mathbf{X}) - \mathbf{H}(\mathbf{X})\boldsymbol{\beta}). \end{aligned}$$

The derivative of ℓ with respect to $\boldsymbol{\varrho}$ can be broken into three parts via the chain rule,

$$\frac{\partial \ell}{\partial \boldsymbol{\varrho}} = \frac{\partial(\text{vec } \Psi_\theta(\mathbf{X}, \mathbf{X}))'}{\partial \boldsymbol{\varrho}} \underbrace{\frac{\partial(\text{vec } \Psi_\theta(\mathbf{X}, \mathbf{X})^{-1})'}{\partial \text{vec } \Psi_\theta(\mathbf{X}, \mathbf{X})}}_A \underbrace{\frac{\partial \ell}{\partial \text{vec } \Psi_\theta(\mathbf{X}, \mathbf{X})^{-1}}}_B. \quad (\text{F.4})$$

Let

$$\mathbf{C}_\theta = \frac{\partial(\text{vec } \Psi_\theta(\mathbf{X}, \mathbf{X}))'}{\partial \boldsymbol{\varrho}'}. \quad (\text{F.5})$$

Parts A and B can be treated in turn. Consider part A . Similarly to (F.3),

$$\begin{aligned} \mathbf{0} &= \frac{\partial(\text{vec } \mathbf{I}_n)'}{\partial \text{vec } \Psi_\theta(\mathbf{X}, \mathbf{X})} = \frac{\partial(\text{vec } \Psi_\theta(\mathbf{X}, \mathbf{X})^{-1} \Psi_\theta(\mathbf{X}, \mathbf{X}))'}{\partial \text{vec } \Psi_\theta(\mathbf{X}, \mathbf{X})} \\ &= \frac{\partial(\text{vec } \Psi_\theta(\mathbf{X}, \mathbf{X})^{-1})'}{\partial \text{vec } \Psi_\theta(\mathbf{X}, \mathbf{X})} (\Psi_\theta(\mathbf{X}, \mathbf{X}) \otimes \mathbf{I}_n) \\ &\quad + \frac{\partial(\text{vec } \Psi_\theta(\mathbf{X}, \mathbf{X}))'}{\partial \text{vec } \Psi_\theta(\mathbf{X}, \mathbf{X})} (\mathbf{I}_n \otimes \Psi_\theta(\mathbf{X}, \mathbf{X})^{-1}) \\ &\implies \frac{\partial(\text{vec } \Psi_\theta(\mathbf{X}, \mathbf{X})^{-1})'}{\partial \text{vec } \Psi_\theta(\mathbf{X}, \mathbf{X})} = -(\Psi_\theta(\mathbf{X}, \mathbf{X})^{-1} \otimes \Psi_\theta(\mathbf{X}, \mathbf{X})^{-1}). \end{aligned} \quad (\text{F.6})$$

Next, consider part B ,

$$\frac{\partial \ell}{\partial \text{vec } \Psi_\theta(\mathbf{X}, \mathbf{X})^{-1}} = \frac{1}{2} [\text{vec } \Psi_\theta(\mathbf{X}, \mathbf{X}) - (f(\mathbf{X}) - \mathbf{H}(\mathbf{X})\boldsymbol{\beta}) \otimes (f(\mathbf{X}) - \mathbf{H}(\mathbf{X})\boldsymbol{\beta})]. \quad (\text{F.7})$$

Equations (F.5), (F.6), and (F.7), and can be plugged into equation (F.4) to give

$$\begin{aligned} \frac{\partial \ell}{\partial \boldsymbol{\varrho}} &= -\frac{1}{2} \mathbf{C}'_\theta (\Psi_\theta(\mathbf{X}, \mathbf{X})^{-1} \otimes \Psi_\theta(\mathbf{X}, \mathbf{X})^{-1}) [\text{vec } \Psi_\theta(\mathbf{X}, \mathbf{X}) - (f(\mathbf{X}) - \mathbf{H}(\mathbf{X})\boldsymbol{\beta}) \otimes (f(\mathbf{X}) - \mathbf{H}(\mathbf{X})\boldsymbol{\beta})] \\ &= -\frac{1}{2} \mathbf{C}'_\theta [\text{vec } \Psi_\theta(\mathbf{X}, \mathbf{X})^{-1} - \Psi_\theta(\mathbf{X}, \mathbf{X})^{-1} \delta(\mathbf{X}) \otimes \Psi_\theta(\mathbf{X}, \mathbf{X})^{-1} \delta(\mathbf{X})]. \end{aligned}$$

So, the information matrix has block components

$$\begin{aligned}
\mathcal{I}_{11} &= \mathcal{I}(\boldsymbol{\beta}, \boldsymbol{\beta}) = -\mathbb{E} \frac{\partial^2 \ell}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = \mathbf{H}(\mathbf{X})' \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} \mathbf{H}(\mathbf{X}), \\
\mathcal{I}_{21} &= \mathcal{I}(\sigma^2, \boldsymbol{\beta}) = -\mathbb{E} \frac{\partial^2 \ell}{\partial \sigma^2 \partial \boldsymbol{\beta}'} = \frac{1}{\sigma^4} \mathbb{E}(f(\mathbf{X}) - \mathbf{H}(\mathbf{X})\boldsymbol{\beta})' \Phi_{\boldsymbol{\rho}}(\mathbf{X}, \mathbf{X})^{-1} \mathbf{H}(\mathbf{X}) = \mathbf{0}', \\
\mathcal{I}_{31} &= \mathcal{I}(\boldsymbol{\rho}, \boldsymbol{\beta}) = -\mathbb{E} \frac{\partial^2 \ell}{\partial \boldsymbol{\rho} \partial \boldsymbol{\beta}'} \\
&= \mathbf{C}'_{\boldsymbol{\theta}} (\Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} \mathbf{H}(\mathbf{X}) \otimes \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} \mathbb{E}(f(\mathbf{X}) - \mathbf{H}(\mathbf{X})\boldsymbol{\beta})) = \mathbf{0}, \\
\mathcal{I}_{22} &= \mathcal{I}(\sigma^2, \sigma^2) = -\mathbb{E} \frac{\partial^2 \ell}{\partial \sigma^2 \partial \sigma^2} \\
&= -\frac{n}{2\sigma^4} + \frac{1}{\sigma^6} \mathbb{E}(f(\mathbf{X}) - \mathbf{H}(\mathbf{X})\boldsymbol{\beta})' \Phi_{\boldsymbol{\rho}}(\mathbf{X}, \mathbf{X})^{-1} (f(\mathbf{X}) - \mathbf{H}(\mathbf{X})\boldsymbol{\beta}) \\
&= -\frac{n}{2\sigma^4} + \frac{1}{\sigma^6} \text{trace } \Phi_{\boldsymbol{\rho}}(\mathbf{X}, \mathbf{X})^{-1} \sigma^2 \Phi_{\boldsymbol{\rho}}(\mathbf{X}, \mathbf{X}) = \frac{n}{2\sigma^4}, \\
\mathcal{I}_{32} &= \mathcal{I}(\boldsymbol{\rho}, \sigma^2) = -\mathbb{E} \frac{\partial^2 \ell}{\partial \boldsymbol{\rho} \partial \sigma^2} \\
&= \frac{1}{2} \mathbf{C}'_{\boldsymbol{\theta}} \mathbb{E} \left(-\frac{1}{\sigma^4} \text{vec } \Phi_{\boldsymbol{\rho}}(\mathbf{X}, \mathbf{X})^{-1} + \frac{2}{\sigma^6} \Phi_{\boldsymbol{\rho}}(\mathbf{X}, \mathbf{X})^{-1} \delta(\mathbf{X}) \otimes \Phi_{\boldsymbol{\rho}}(\mathbf{X}, \mathbf{X})^{-1} \delta(\mathbf{X}) \right) \\
&= \frac{1}{2} \mathbf{C}'_{\boldsymbol{\theta}} \left(-\frac{1}{\sigma^4} \text{vec } \Phi_{\boldsymbol{\rho}}(\mathbf{X}, \mathbf{X})^{-1} + \frac{2}{\sigma^4} \text{vec } \Phi_{\boldsymbol{\rho}}(\mathbf{X}, \mathbf{X})^{-1} \right) \\
&= \frac{1}{2\sigma^4} \mathbf{C}'_{\boldsymbol{\theta}} \text{vec } \Phi_{\boldsymbol{\rho}}(\mathbf{X}, \mathbf{X})^{-1}.
\end{aligned}$$

Developing a formula for $\mathcal{I}(\boldsymbol{\rho}, \boldsymbol{\rho})$ is more complex and broken into parts.

$$\begin{aligned}
\mathcal{I}(\boldsymbol{\rho}, \boldsymbol{\rho}) &= -\mathbb{E} \frac{\partial^2 \ell}{\partial \boldsymbol{\rho} \partial \boldsymbol{\rho}'} \\
&= \frac{1}{2} \mathbb{E} (\mathbf{I}_d \otimes [(\text{vec } \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1})' - (\delta(\mathbf{X})' \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} \otimes \delta(\mathbf{X})' \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1})]) \frac{\partial \mathbf{C}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\rho}} \quad (\text{F.8}) \\
&\quad + \frac{1}{2} \mathbb{E} \left(\frac{\partial (\text{vec } \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1})'}{\partial \boldsymbol{\rho}} - \frac{\partial (\delta(\mathbf{X})' \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} \otimes \delta(\mathbf{X})' \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1})}{\partial \boldsymbol{\rho}} \right) \mathbf{C}_{\boldsymbol{\theta}}.
\end{aligned}$$

Note that the expectation of the first term in (F.8) is zero, since

$$\begin{aligned}
&\mathbb{E}(\delta(\mathbf{X})' \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} \otimes \delta(\mathbf{X})' \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1}) \\
&= \mathbb{E}(\text{vec } (\Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} \delta(\mathbf{X}) \delta(\mathbf{X})' \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1}))' = (\text{vec } \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1})'.
\end{aligned}$$

So,

$$\begin{aligned}
\mathcal{I}(\boldsymbol{\rho}, \boldsymbol{\rho}) &= \frac{1}{2} \left(-\mathbf{C}'_{\boldsymbol{\theta}} (\Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} \otimes \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1}) \right. \\
&\quad \left. - \mathbb{E} \frac{\partial (\text{vec } (\Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} \delta(\mathbf{X}) \delta(\mathbf{X})' \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1}))'}{\partial \boldsymbol{\rho}} \right) \mathbf{C}_{\boldsymbol{\theta}}. \quad (\text{F.9})
\end{aligned}$$

The expectation in (F.9) is

$$\begin{aligned}
& \mathbb{E} \frac{\partial (\text{vec} (\Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} \delta(\mathbf{X}) \delta(\mathbf{X})' \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1}))'}{\partial \boldsymbol{\rho}} \\
&= \frac{\partial (\text{vec} (\Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1}))'}{\partial \boldsymbol{\rho}} (\mathbb{E} \delta(\mathbf{X}) \delta(\mathbf{X})' \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} \otimes \mathbf{I}_n) \\
&\quad + \frac{\partial (\text{vec} (\Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1}))'}{\partial \boldsymbol{\rho}} (\mathbf{I}_n \otimes \mathbb{E} \delta(\mathbf{X}) \delta(\mathbf{X})' \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1}) \\
&= -2\mathbf{C}'_{\boldsymbol{\theta}} (\Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} \otimes \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1}).
\end{aligned} \tag{F.10}$$

Plugging (F.10) into (F.9) gives

$$\mathcal{I}_{33} = \mathcal{I}(\boldsymbol{\rho}, \boldsymbol{\rho}) = \frac{1}{2} \mathbf{C}'_{\boldsymbol{\theta}} (\Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} \otimes \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1}) \mathbf{C}_{\boldsymbol{\theta}}.$$

Using partitioned matrix inverse results [19] and noting that \mathbf{c}_2 , \mathcal{I}_{21} , \mathcal{I}_{12} , \mathcal{I}_{31} , and \mathcal{I}_{13} are matrices of zeros gives (F.1). \square

Now, the expressions in Lemma F.1 are used to prove Theorem 5.1. The first term on the right-hand side of (F.1) can be bounded above as

$$\begin{aligned}
\mathbf{c}'_1 \mathcal{I}_{11}^{-1} \mathbf{c}_1 &= \mathbf{c}'_1 (\mathbf{H}(\mathbf{X})' \Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} \mathbf{H}(\mathbf{X}))^{-1} \mathbf{c}_1 \\
&\leq \frac{\lambda_{\max}(\Psi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X}))}{\lambda_{\min}(\mathbf{H}(\mathbf{X})' \mathbf{H}(\mathbf{X}))} \|\mathbf{c}_1\|_2^2 \leq \frac{n \sup_{\mathbf{u}, \mathbf{v} \in \Omega} \Psi_{\boldsymbol{\theta}}(\mathbf{u}, \mathbf{v})}{\lambda_{\min}(\mathbf{H}(\mathbf{X})' \mathbf{H}(\mathbf{X}))} \|\mathbf{c}_1\|_2^2.
\end{aligned}$$

The eigenvalue $\lambda_{\min}(\mathbf{H}(\mathbf{X})' \mathbf{H}(\mathbf{X}))$ has approximation

$$\lambda_{\min}(\mathbf{H}(\mathbf{X})' \mathbf{H}(\mathbf{X})) = \lambda_{\min} \left(\sum_{i=1}^n \mathbf{h}(\mathbf{x}_i) \mathbf{h}(\mathbf{x}_i)' \right) \approx n \lambda_{\min} \left(\int \mathbf{h}(\mathbf{y}) \mathbf{h}(\mathbf{y})' dF(\mathbf{y}) \right) = n s_1, \tag{F.11}$$

where F denotes the large sample distribution of the input locations \mathbf{X} , $s_1 \geq 0$, and $s_1 > 0$ unless $\mathbf{h}(\mathbf{y})' \mathbf{a} = 0$ with probability 1 with respect to the large sample distribution F for some $\mathbf{a} \neq 0$. Giving approximate upper bound to the first term on the right-hand side of (F.1)

$$\mathbf{c}'_1 \mathcal{I}_{11}^{-1} \mathbf{c}_1 \leq \frac{\sup_{\mathbf{u}, \mathbf{v} \in \Omega} \Psi_{\boldsymbol{\theta}}(\mathbf{u}, \mathbf{v})}{s_1} \|\mathbf{c}_1\|_2^2, \tag{F.12}$$

where s_1 is implicitly defined in (F.11) and the probability of the inequality being violated by more than $\varepsilon > 0$ goes to zero as $n \rightarrow \infty$. The second term on the right-hand side of the approximate parameter estimation error expression (F.1) has

$$\mathbf{c}'_3 (\mathcal{I}_{33} - \mathcal{I}_{32} \mathcal{I}_{22}^{-1} \mathcal{I}_{23})^{-1} \mathbf{c}_3 \leq \|\mathbf{c}_3\|_2^2 / \lambda_{\min}(\mathcal{I}_{33} - \mathcal{I}_{32} \mathcal{I}_{22}^{-1} \mathcal{I}_{23}). \tag{F.13}$$

Note that

$$\begin{aligned}
& \mathcal{I}_{33} - \mathcal{I}_{32} \mathcal{I}_{22}^{-1} \mathcal{I}_{23} \\
&= \frac{1}{2\sigma^4} \mathbf{C}'_{\boldsymbol{\theta}} \left((\Phi_{\boldsymbol{\rho}}(\mathbf{X}, \mathbf{X})^{-1} \otimes \Phi_{\boldsymbol{\rho}}(\mathbf{X}, \mathbf{X})^{-1}) - \frac{1}{n} (\text{vec} \Phi_{\boldsymbol{\rho}}(\mathbf{X}, \mathbf{X})^{-1}) (\text{vec} \Phi_{\boldsymbol{\rho}}(\mathbf{X}, \mathbf{X})^{-1})' \right) \mathbf{C}_{\boldsymbol{\theta}}.
\end{aligned}$$

The matrix inside the quadratic form has eigenvector $\mathbf{u}_1 = \text{vec } \Phi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X}) / \|\text{vec } \Phi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})\|_2$ with corresponding eigenvalue 0. So, the minimum eigenvalue of the above expression can be bounded below by

$$\begin{aligned} & \frac{1}{2\sigma^4} \lambda_{\min}(\mathbf{C}'_{\boldsymbol{\theta}}(\mathbf{I}_{n^2} - \mathbf{u}_1 \mathbf{u}'_1) \mathbf{C}_{\boldsymbol{\theta}}) \\ & \times \lambda_2 \left((\Phi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} \otimes \Phi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1}) - \frac{1}{n} (\text{vec } \Phi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1}) (\text{vec } \Phi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1})' \right), \end{aligned}$$

where $\lambda_2(\cdot)$ denotes the *second* smallest eigenvalue of its argument. By Weyl's theorem [21], the second smallest eigenvalue of the perturbed matrix can be bounded below by

$$\lambda_{\min}((\Phi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1} \otimes \Phi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})^{-1})) = 1/\lambda_{\max}(\Phi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X}))^2 \geq 1/(n \sup_{\mathbf{u}, \mathbf{v} \in \Omega} \Phi_{\boldsymbol{\theta}}(\mathbf{u}, \mathbf{v}))^2. \quad (\text{F.14})$$

Further,

$$\begin{aligned} & \mathbf{C}'_{\boldsymbol{\theta}}(\mathbf{I}_{n^2} - \mathbf{u}_1 \mathbf{u}'_1) \mathbf{C}_{\boldsymbol{\theta}} \\ & = \sigma^4 \left[\sum_{i,j} \frac{\partial \Phi_{\boldsymbol{\theta}}(\mathbf{x}_i, \mathbf{x}_j)}{\partial \boldsymbol{\theta}} \frac{\partial \Phi_{\boldsymbol{\theta}}(\mathbf{x}_i, \mathbf{x}_j)}{\partial \boldsymbol{\theta}'} \right. \\ & \quad \left. - \frac{1}{\|\text{vec } \Phi_{\boldsymbol{\theta}}(\mathbf{X}, \mathbf{X})\|_2^2} \left(\sum_{i,j} \frac{\partial \Phi_{\boldsymbol{\theta}}(\mathbf{x}_i, \mathbf{x}_j)}{\partial \boldsymbol{\theta}} \Phi_{\boldsymbol{\theta}}(\mathbf{x}_i, \mathbf{x}_j) \right) \left(\sum_{i,j} \frac{\partial \Phi_{\boldsymbol{\theta}}(\mathbf{x}_i, \mathbf{x}_j)}{\partial \boldsymbol{\theta}'} \Phi_{\boldsymbol{\theta}}(\mathbf{x}_i, \mathbf{x}_j) \right)' \right] \\ & \approx n^2 \sigma^4 \left[\int \frac{\partial \Phi_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{y})}{\partial \boldsymbol{\theta}} \frac{\partial \Phi_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{y})}{\partial \boldsymbol{\theta}'} dF \times F(\mathbf{x}, \mathbf{y}) \right. \\ & \quad \left. - \frac{1}{\|\Phi_{\boldsymbol{\theta}}\|_{L_2(F \times F)}^2} \left(\int \frac{\partial \Phi_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{y})}{\partial \boldsymbol{\theta}} \Phi_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{y}) dF \times F(\mathbf{x}, \mathbf{y}) \right) \left(\int \frac{\partial \Phi_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{y})}{\partial \boldsymbol{\theta}'} \Phi_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{y}) dF \times F(\mathbf{x}, \mathbf{y}) \right)' \right] \\ & \succeq n^2 \sigma^4 s_2, \end{aligned} \quad (\text{F.15})$$

where $F \times F$ denotes the *product measure* [6]. Applying a version of the Cauchy-Schwarz inequality for random vectors in $L_2(F \times F)$, provides $s_2 \geq 0$ and $s_2 > 0$, unless $\frac{\partial \Phi_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{y})}{\partial \boldsymbol{\theta}'} \mathbf{a} = \Phi_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{y}) \mathbf{b}$ with probability 1 with respect to the large sample distribution $F \times F$ for some $(\mathbf{a}' \ \mathbf{b}') \neq \mathbf{0}$ [39]. Combining the bounds in equations (F.13), (F.14), and (F.15), gives

$$\mathbf{c}'_3 (\mathcal{I}_{33} - \mathcal{I}_{32} \mathcal{I}_{22}^{-1} \mathcal{I}_{23})^{-1} \mathbf{c}_3 \leq \frac{2 \sup_{\mathbf{u}, \mathbf{v} \in \Omega} \Phi_{\boldsymbol{\theta}}(\mathbf{u}, \mathbf{v})^2}{s_2} \|\mathbf{c}_3\|_2^2, \quad (\text{F.16})$$

with s_2 implicitly defined in equation (F.15) and the probability of the inequality being violated by more than $\varepsilon > 0$ going to zero as $n \rightarrow \infty$. Combining the approximate bounds (F.12) and (F.16) gives the result.

Acknowledgements

The authors gratefully acknowledge funding from NSF DMS-1621722.

References

- [1] Aronszajn, N. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc.* **68** 337–404.
- [2] Aurenhammer, F. (1991). Voronoi Diagrams – A Survey of a Fundamental Geometric Data Structure. *ACM Comp. Surv.* **23** 345–405.
- [3] Ba, S. and Joseph, V. R. (2012). Composite Gaussian process models for emulating expensive functions. *Ann. Appl. Stat.* **6** 1838–1860.
- [4] Ba, S. and Joseph, V. R. (2014). CGP: Composite Gaussian process models. <https://CRAN.R-project.org/package=CGP>
- [5] Ba, S., and Joseph, V. R. (2015). MaxPro: Maximum projection designs. <https://cran.r-project.org/package=MaxPro>
- [6] Bartle, R. G. (1995). *The Elements of Integration and Lebesgue Measure*. Wiley, New York.
- [7] Bingham, D., Sitter, R. R., and Tang, B. (2009). Orthogonal and nearly orthogonal designs for computer experiments. *Biometrika* **96** 51–65.
- [8] Bowman, V. E. and Woods, D. C. (2013). Weighted space-filling designs. *Journal of Simulation* **7** 249–263.
- [9] Ba, S. (2015). SLHD: Maximin-Distance (Sliced) Latin Hypercube Designs. <https://CRAN.R-project.org/package=SLHD>.
- [10] Dancik, G. M. (2013). mlegp: Maximum Likelihood Estimates of Gaussian Processes. <http://CRAN.R-project.org/package=mlegp>.
- [11] Dragomir, S. S. (2015). Reverses of Schwarz inequality in inner product spaces and applications. *Mathematische Nachrichten*, **7**, 730–742.
- [12] Eaves, B. C. (1972). Homotopies for computation of fixed points *Math. Prog.* **3** 1–22.
- [13] Ellis A. M., Garcia A. J., Focks D. A., Morrison A. C., and Scott T. W. (2011). Parameterization and sensitivity analysis of a complex simulation model for mosquito population dynamics, dengue transmission, and their control. *The American Society of Tropical Medicine and Hygiene*, **85** 257–264.
- [14] Fang, K. T., Li, R. and Sudjianto, A. (2006). *Design and Modeling of Computer Experiments*. Chapman and Hall, Florida.
- [15] Golub, G. H. and Van Loan, C. F. (1989). *Matrix Computations*. 2nd ed. Johns Hopkins University Press, Baltimore.
- [16] Gramacy, R. B. and Lee, H. K. (2009). Adaptive design and analysis of supercomputer experiments. *Technometrics* **51** 130–145.
- [17] Haaland, B. (2010). Statistical Approaches for Computer Experiments, Management of Temperature Control Systems, and Prediction of Dose-Responses. Ph.D. Dissertation. University of Wisconsin at Madison, Madison, WI, USA. AAI3437057.

- [18] Haaland, B. and Qian, P. Z. G. (2011). Accurate emulators for large-scale computer experiments. *Ann. Statist.* **39** 2974–3002.
- [19] Harville, D. A. (2008). *Matrix Algebra From a Statistician's Perspective*. Springer, New York.
- [20] Holloway, J. P., Bingham D., Chou C.-C., Doss F., Drake R. P., Fryxell B., Grosskopf M., van der Holst B., Mallick B. K., McClarren R., Mukherjee A., Nair V., Powell K. G., Ryu D., Sokolov I., Toth G., and Zhang Z. (2011). Predictive modeling of a radiative shock system. *Reliability Engineering & System Safety* **96** 1184–1193.
- [21] Ipsen, I. C. F. and Nadler, B. (2009). Refined perturbation bounds for eigenvalues of Hermitian and non-Hermitian matrices. *SIAM Journal on Matrix Analysis and Applications* **31**, 40–53.
- [22] Jin, R., Chen, W., and Sudjianto, A. (2005). An efficient algorithm for constructing optimal design of computer experiments. *Journal of Statistical Planning and Inference* **134** 268–287.
- [23] Jones, D. R., Schonlau, M., and Welch, W. J. (1998). Efficient global optimization of expensive black-box functions. *Journal of Global Optimization* **1** 455–492.
- [24] Joseph, V. R., Dasgupta, T., Tuo, R., and Wu, C. J. (2015). Sequential exploration of complex surfaces using minimum energy designs. *Technometrics* **57** 64–74.
- [25] Joseph, V. R., Gul, E., and Ba, S. (2015). Maximum projection designs for computer experiments. *Biometrika*: asv002.
- [26] Lauter, E. (1974). Experimental design in a class of models. *Statistics: A Journal of Theoretical and Applied Statistics* **5**, 379–398.
- [27] Magnus, J. R. and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics (revised)*. New York: Wiley.
- [28] Mahalanobis, P. C. (1936). On the generalised distance in statistics. *Proc. Nat. Inst. Sci. India* **2** 49–55.
- [29] McKay, M. D., Conover, W. J., and Beckman, R. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21** 239–245.
- [30] Narcowich, F. J. and Ward, J. D. (1992). Norm estimates for the inverses of a general class of scattered-data radial-function interpolation matrices. *Journal of Approximation Theory* **69** 84–109.
- [31] Niederreiter, H. (1992). *Random Number Generation and Quasi-Monte Carlo Methods*. Society for Industrial and Applied Mathematics, Philadelphia.
- [32] Nocedal, J. and Wright, S. J. (1999). *Numerical Optimization (Vol. 2)*. New York: Springer.
- [33] Qian, Peter. Z. G., Wu, H., and Wu, C. F. J. (2012). Gaussian process models for computer experiments with qualitative and quantitative factors.” *Technometrics*.

- [34] Sacks, J., William, J. W., Mitchell, T. J., and Wynn, H. P. (1989). Design and analysis of computer experiments. *Statist. Sci.* **4** 409–423.
- [35] Santner, T. J., Williams, B. J., and Notz, W. I. (2003). *The Design and Analysis of Computer Experiments*. Springer, New York.
- [36] Stein, E. M. and Weiss, G. (1971). *Introduction to Fourier Analysis on Euclidean Spaces*. Princeton University Press, New Jersey.
- [37] Stein, M. L. (2012). *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media, Chicago.
- [38] Tang, B. (1993). Orthogonal array-based Latin hypercubes. *J. Amer. Statist. Assoc.* **88** 1392–1397.
- [39] Tripathi, G. (1999). A matrix extension of the Cauchy-Schwarz Inequality. *Economics Letters* **63** 1–3.
- [40] Tse K. M., Lee H. P., Shabana N., Loon S. C., Watson P. G., and Thean S. Y. (2011). Do shapes and dimensions of scleral flap and sclerostomy influence aqueous outflow in trabeculectomy? A finite element simulation approach. *British Journal of Ophthalmology*, **96** 432–437.
- [41] Varga, R. S. (2004). *Geršgorin and His Circles*. Springer-Verlag, Berlin.
- [42] Watson, G. N. (1995). *A treatise on the theory of Bessel functions*. Cambridge University Press, New York.
- [43] Wendland, H. (2005). *Scattered Data Approximation*. Cambridge University Press, New York.
- [44] Wozniakowski, H. (1992). Average Case Complexity of Linear Multivariate Problems: I. Theory. *Journal of Complexity* **8** 337–372.