Minimum Reduced Order Modelling

Robert A. Milton, Solomon F. Brown

Department of Chemical and Biological Engineering, University of Sheffield, Sheffield, S1 3JD, United Kingdom

Abstract

Keywords: Gaussian Process, Global Sensitivity Analysis, Sobol' Index, Surrogate Model

1. Introduction

The simulation of a large range of engineering systems requires the application of complex computational models. The use of these models is often computationally expensive, and can be prohibitive when attempting to use the underlying model as part of a system optimization. This is commonly mitigated by emulating the complex model response $y(\mathbf{x})$ to its M-dimensional input $\mathbf{x}$ with a surrogate.

A popular class of surrogate is Gaussian Processes (GPs) [1, 2] which are flexible, efficient, non-parametric and analytically tractable (references). Other surrogate methods include Polynomial Chaos Expansions [3, 4, 5], low-rank tensor approximations [6, 7], and support vector regression [8]. These, however, tend to lack the combination of characteristics just mentioned, which make GPs ideal for our purposes.

The development of surrogate models usually requires an exponentially growing number of output results $y(\mathbf{x})$ throughout the input space as M increases: known as the curse of dimensionality. There is therefore a significant driver for methods with which to reduce the dimensionality of the input space, and so a more efficient means of generating the emulator. One way of selecting the directions of most influence is through the application of

Email addresses: `r.a.milton@sheffield.ac.uk` (Robert A. Milton), `s.f.brown@sheffield.ac.uk` (Solomon F. Brown)

Global Sensitivity Analysis, however where these directions are not aligned with the input basis this resulting dimensionality reduction is sub-optimal (references).

As such a number of approaches to obtaining the optimal dimensionality reduction have been developed, which can broadly be catagorized through their use of different sensitivity measures. For example, [9] proposed a means of calculating this optimal reduced dimensional space, the Active Subspace, through a derivative sensitivity measure. This was found to work very effectively provided reliable derivatives are available, which may not available for complex or noisy systems. Liu and Guillas [10] recently addressed this using gradient-based kernel reduction of a GP surrogate, locating the Active Subspace via eigendecomposition of the surrogate gradient. On the other hand Minimum Average Variance Estimation (MAVE) proposed by Xia et al [11] effectively uses a variance-based sensitivity (statistical independence) as the measure of input relevance. The method therefore bears some similarity to our own, but relies on local linear kernels, whose bandwidth must be chosen carefully.

A key variance-based sensitivity measure are the Sobol' indices [12], which is one of the most widely used approaches for GSA. With the increased popularity of using GPs as surrogate models, the main disadvantage to the Sobol' method was solved [13, 14, 15]. The use of GPs provided an alternative method to estimating multidimensional integrals using Monte Carlo schemes, which required 10,000 datapoints to reach 10% precision [16]. GPs enable semi-analytic evaluation of Sobol' indices, introduced by Jin et al. [14]. Before Oakley and O'Hagan [13] used the global stochastic model of a GP, providing the calculations to produce random variables as a new sensitivity measures. Oakley and O'Hagan's [13] model allows the sensitivity indices accuracy to be analysed due to the distribution of the variables. Marrel et al. [15] extended this comparing them and building on the work from Oakley and O'Hagan [13] leading to a novel algorithm which builds confidence intervals for the Sobol' indices. Marrel et al. [15] tested both methods on toy functions providing results that show very accurate sensitivity indices and satisfactory confidence intervals from the second method. However, when the approach was illustrated on real data to provide a sensitivity analysis on radionuclide groundwater transport, it was found that the confidence intervals were inaccurate for very low indices due to overestimation of the lowest Sobol' indices.

The purpose of this work is to present a Global Sensitivity Analysis based model order reduction approach, which uses

- A GP surrogate.
- Semi-analytic Sobol' indices for the surrogate.
- Optimal dimension reduction (essentially locating the active subspace) by Sobol' index.

This paper is organised as follows: Section 2 presents a review of the concepts and measures that are used in this work. Section 3 describes the approach developed, including the details of the calculation of the Sobol' indices and of the optimization of the basis of the optimal low dimensional subspace. Section 4 presents the application of the method to a variety of test problems to assess its performance, while Section 5 summarises our findings and directions for future work.

2. Review of Gaussian Processes and Global Sensitivity Analysis

2.1. Gaussian Process Surrogate

In order to avoid the difficulty and expense in obtaining and analyzing response data from a computationally heavy model we adopt a Gaussian Process (GP) surrogate or emulator. The response $y(\mathbf{x})$ to arbitrarily fixed input is modelled as the sum $f(\mathbf{x})+e(\mathbf{x})$ of two Gaussian random variables encapsulating coherent signal and incoherent noise. The latter is characterized by a zero-mean distribution that is independent of the input:

$$e(\mathbf{x}) \sim \mathsf{N}\left[0, \sigma_{\mathbf{e}}^2\right]$$

The signal $f(\mathbf{x})$ is characterized by its covariance kernel $\sigma_{\mathbf{f}}^2 k(\mathbf{x}_n, \mathbf{x})$ which measures the similarity between inputs $\mathbf{x}_n$ and $\mathbf{x}$, and propagates any similarity to $y(\mathbf{x}_n)$ and $y(\mathbf{x})$. In the majority of applications, the kernel is naturally stationary, a function of $(\mathbf{x}-\mathbf{x}_n)$ alone. We shall further assume that the kernel is twice differentiable at its maximum $(\mathbf{x}=\mathbf{x}_n)$. Hence, the Hessian at the maximum must be symmetric negative semi-definite and therefore diagonalizes to

$$\partial_{\mathbf{x}\mathbf{x}} \log k(\mathbf{x}, \mathbf{x}) =: -\Theta^{\intercal} \Lambda^{-2} \Theta$$

When $|\mathbf{x}-\mathbf{x}_n|$ is large the kernel value is miniscule in any any relevant direction. The kernel details are therefore largely irrelevant to the response *any* time $|\mathbf{x} - \mathbf{x}_n|$ is large, advocating (if not justifying) the Taylor approximation

$$k(\mathbf{x}_n, \mathbf{x}) = \exp\left(-\frac{(\mathbf{x} - \mathbf{x}_n)^\intercal \Theta^\intercal \Lambda^{-2} \Theta (\mathbf{x} - \mathbf{x}_n)}{2} \left(1 + O(|\mathbf{x} - \mathbf{x}_n|)\right)\right)$$

The differentiability we have imposed forces the power spectrum of the signal f to decay rapidly. Modes of response oscillating rapidly with $\mathbf{x}$ are interpreted as noise by the GP, as the kernel smoothes y into f. Such regularization is often, but not always, desirable, to avoid wildly unreliable interpolation of an overfit regression.

2.2. Kernel Optimization

In order to deal with the curse of dimensionality we propose to find orthogonal rotation matrix Θ and diagonal length-scale matrix Λ which best fit observed responses $y(\mathbf{X}^\intercal)$. The largest lengthscales in Λ mark the least relevant directions that can be ignored. However, the best fit must optimize $M(M+1)/2+2$ hyperparameters simultaneously to determine $\Theta, \Lambda, \sigma_{\mathbf{f}}^2$ and $\sigma_{\mathbf{e}}^2$. Direct optimization of such a large problem may result in obtaining only local optima. Exploratory grid search is astronomically expensive $O(\exp(M(M+1)/2))$, likewise any random sampling which is not hopelessly sparse. Perhaps for these reasons, Θ has always been fixed as identity in the literature. The lengthscales comprising Λ are also usually identical, furnishing a radial basis function (RBF) kernel [1]. The few studies where Λ is not identical speak of an automatic relevance determination (ARD) kernel, with model order reduction in mind [17, 18].

2.3. Global Sensitivity Analysis

This paper aims to achieve kernel optimization indirectly, via global sensitivity analysis (GSA). The surrogate expectation

$$\mathsf{E}_y[y(\mathbf{x})] = \mathsf{E}_y[f(\mathbf{x})] =: \bar{f}(\mathbf{x})$$

has a variation (over $\mathbf{x} \in \mathbb{R}^M$) which can be apportioned by Sobol' index

$$S_{\mathbf{m}}((\Theta)_{\mathbf{m}\times\mathbf{M}}) := \frac{\mathsf{Var}_{\mathbf{x}}\left[\mathsf{E}_{\mathbf{x}}\left[\bar{f}(\mathbf{x})|(\Theta\mathbf{x})_{\mathbf{m}}\right]\right]}{\mathsf{Var}_{\mathbf{x}}\left[\bar{f}(\mathbf{x})\right]} \leq 1$$

to subspaces $(\Theta\mathbf{x})_{\mathbf{m}}$ of dimension $m \leq M$. These may be calculated analytically for the exponential quadratic kernel used here. To cure to the curse of dimensionality is to find $(\Theta)_{\mathbf{m}\times\mathbf{M}}$ such that $S_{\mathbf{m}} \approx 1$ for $m \ll M$. The rotation sub-matrix $(\Theta)_{\mathbf{M}\backslash\mathbf{m}\times\mathbf{M}}$ has a manageable number of elements if $M-m$ is small. This paper takes the most economical approach, maximizing $S_{\mathbf{M-1}}$ to eliminate $(\Theta\mathbf{x})_M$ iteratively.

3. Methodology

Let $\mathbf{X}$ be the $(N \times M)$ design matrix of observed inputs eliciting the N response $y(\mathbf{X}^\intercal)$. The observations are standardized such that

$$(\mathbf{0})_{\mathbf{M}} = \mathsf{E}[\mathbf{x}_n] := N^{-1}\sum_{n=1}^{N}(\mathbf{X})^\intercal_{n\times\mathbf{M}} \quad ; \quad 1 = \mathsf{Var}[\mathbf{x}_{n\times m}] = N^{-1}\sum_{n=1}^{N}(\mathbf{X})_{n\times m}(\mathbf{X})^\intercal_{n\times m}$$

$$0 = \mathsf{E}[y(\mathbf{x}_n)] := N^{-1}\sum_{n=1}^{N}(y(\mathbf{X}^\intercal))_n \quad ; \quad 1 = \mathsf{Var}[y(\mathbf{x}_n)] = N^{-1}y(\mathbf{X}^\intercal)^\intercal y(\mathbf{X}^\intercal)$$

where boldface subscripts refer to the multi-indices

$$\emptyset =: \mathbf{0} \subseteq \mathbf{m} := (1,\ldots,m) \subseteq \mathbf{M} \tag{1}$$

which always precede superscript operations (such as transposition or inversion). For brevity, we shall admit row vector Gaussian probability densities $p((\mathbf{u})_{\mathbf{m}};(\mathbf{U})_{\mathbf{m}\times\mathbf{N}},(\Sigma)_{\mathbf{m}\times\mathbf{m}})$ such that

$$\begin{aligned}&(p((\mathbf{u})_{\mathbf{m}};(\mathbf{U})_{\mathbf{m}\times\mathbf{N}},(\Sigma)_{\mathbf{m}\times\mathbf{m}}))_{1\times n}\\ &\quad := (2\pi)^{-M/2}\left|\Sigma_{\mathbf{u}}\right|^{-1/2}\exp\left(-\frac{(\mathbf{u}-(\mathbf{U})_{\mathbf{m}\times n})^\intercal(\Sigma)^{-1}_{\mathbf{m}\times\mathbf{m}}(\mathbf{u}-(\mathbf{U})_{\mathbf{m}\times n})}{2}\right)\end{aligned} \tag{2}$$

naturally collapsing to the (scalar) normal multivariate density when $N=1$. The algebraic development which follows relies exclusively on trivial Gaussian marginalization and scaling

$$\mathbf{u} \sim \mathsf{N}[\mathbf{U},\Sigma] \Rightarrow \mathsf{E}[\mathbf{u}|(\mathbf{u})_{\mathbf{m}}] \sim \mathsf{N}[(\mathbf{U})_{\mathbf{m}},(\Sigma)_{\mathbf{m}\times\mathbf{m}}] \tag{3}$$

$$\mathbf{u} \sim \mathsf{N}[\mathbf{U},\Sigma] \Rightarrow \Theta^\intercal\mathbf{u} \sim \mathsf{N}[\Theta^\intercal\mathbf{U},\Theta^\intercal\Sigma\Theta] \tag{4}$$

together with an extremely useful product formula reported in [19]

$$\begin{aligned}&p(\mathbf{u};\mathbf{a},\mathbf{A})p(\Theta^\intercal\mathbf{u};\mathbf{b},\mathbf{B}) = p\left(\mathbf{0};(\mathbf{b}-\Theta^\intercal\mathbf{a}),(\mathbf{B}+\Theta^\intercal\mathbf{A}\Theta)\right)\\ &\quad\times p\left(\mathbf{u};(\mathbf{A}^{-1}+\Theta\mathbf{B}^{-1}\Theta^\intercal)^{-1}(\mathbf{A}^{-1}\mathbf{a}+\Theta\mathbf{B}^{-1}\mathbf{b}),(\mathbf{A}^{-1}+\Theta\mathbf{B}^{-1}\Theta^\intercal)^{-1}\right)\end{aligned} \tag{5}$$

In referring back to these formulae, remember that $\mathbf{u}, \mathbf{U}, \Sigma, \Theta$ are still quite arbitrary here (within the dictates of the minimal dimension, sign, symmetry and invertibility requirements for these formulae to make sense), and not restricted to any particular values these quantities may later take.

3.1. Gaussian Process Surrogate

Non-parametric GP regression fits signal f and noise e Gaussian processes to

$$y(\mathbf{X}^\intercal) = f(\mathbf{X}^\intercal) + e(\mathbf{X}^\intercal) \tag{6}$$

This work exclusively employs objective Bayesian priors

$$\begin{aligned} f(\mathbf{X}^\intercal) &\sim \mathsf{N}\big[(\mathbf{0})_{\mathbf{N}}, \sigma_{\mathbf{f}}^2 k(\mathbf{X}^\intercal, \mathbf{X}^\intercal)\big] \\ e(\mathbf{X}^\intercal) &\sim \mathsf{N}\big[(\mathbf{0})_{\mathbf{N}}, \sigma_{\mathbf{e}}^2 (\mathbf{I})_{\mathbf{N}\times\mathbf{N}}\big] \end{aligned}$$

built on an ARD kernel

$$k(\mathbf{x}, \mathbf{x}_n) := (2\pi)^{M/2} \, |\Lambda| \, p\big(\mathbf{x}; \mathbf{x}_n, \Lambda^2\big) \tag{7}$$

with diagonal positive definite lengthscale matrix Λ. Bayesian conditioning ultimately furnishes the predictive process

$$y(\mathbf{x}) \sim \mathsf{N}\big[\bar{f}(\mathbf{x}), \Sigma_{\mathbf{f}}(\mathbf{x}) + \sigma_{\mathbf{e}}^2\big]$$

with signal mean and variance

$$\begin{aligned} \bar{f}(\mathbf{x}) &:= \sigma_{\mathbf{f}}^2 k(\mathbf{x}, \mathbf{X}^\intercal) \mathbf{K}^{-1} y(\mathbf{X}^\intercal) \\ \Sigma_{\mathbf{f}}(\mathbf{x}) &:= \sigma_{\mathbf{f}}^2 k(\mathbf{x}, \mathbf{x}) - \sigma_{\mathbf{f}}^2 k(\mathbf{x}, \mathbf{X}^\intercal) \mathbf{K}^{-1} \sigma_{\mathbf{f}}^2 k(\mathbf{X}^\intercal, \mathbf{x}) \end{aligned} \tag{8}$$

where

$$\mathbf{K} := \sigma_{\mathbf{f}}^2 k(\mathbf{X}^\intercal, \mathbf{X}^\intercal) + \sigma_{\mathbf{e}}^2 (\mathbf{I})_{\mathbf{N}\times\mathbf{N}} \tag{9}$$

The $M+2$ hyperparameters constituting $\Lambda, \sigma_{\mathbf{f}}$ and $\sigma_{\mathbf{e}}$ are simultaneously optimized for maximium marginal likelihood $\mathsf{p}[y|\mathbf{X}^\intercal]$, using the GPy software library [20].

3.2. Global Sensitivity Analysis

Imagine a sample datum $\mathbf{u}$ is drawn from a standardized normal test distribution

$$\mathbf{u} \sim \mathsf{N}[(\mathbf{0})_{\mathbf{M}}, (\mathbf{I})_{\mathbf{M}\times\mathbf{M}}] \tag{10}$$

The datum basis is rotated to

$$\mathbf{x} =: \Theta^\intercal \mathbf{u} \tag{11}$$

eliciting the conditional surrogate responses

$$\bar{f}_{\mathbf{m}}((\mathbf{u})_{\mathbf{m}}; (\Theta)_{\mathbf{m}\times\mathbf{M}}) := \mathsf{E}\left[\bar{f}(\Theta^\intercal \mathbf{u})|(\mathbf{u})_{\mathbf{m}}\right] \tag{12}$$

Knowledge of $\mathbf{u}$ herein ranges from totally conditional $\bar{f}_{\mathbf{M}}(\mathbf{u};\Theta) = \bar{f}(\mathbf{x})$ to unconditional ignorance $\bar{f}_{\mathbf{0}} = \mathsf{E}\left[\bar{f}(\mathbf{x})\right]$. Equations (7) to (10) enable analytic integration using Equations (3) to (5) to yield

$$\bar{f}_{\mathbf{m}}((\mathbf{u})_{\mathbf{m}}; (\Theta)_{\mathbf{m}\times\mathbf{M}}) = \frac{p((\mathbf{u})_{\mathbf{m}}; (\mathbf{F})_{\mathbf{m}\times\mathbf{N}}, (\Sigma)_{\mathbf{m}\times\mathbf{m}})}{p((\mathbf{u})_{\mathbf{m}}; (\mathbf{0})_{\mathbf{m}}, (\mathbf{I})_{\mathbf{m}\times\mathbf{m}})}\, \bar{\mathbf{f}}_0^\intercal \tag{13}$$

where the constant $\bar{\mathbf{f}}_0$ is the Hadamard (element-wise) product $\circ$ of two row vectors

$$\bar{\mathbf{f}}_0 := \sigma_{\mathbf{f}}^2 (2\pi)^{M/2} \left|\Lambda\right| p\left(\mathbf{0}; \mathbf{X}^\intercal, \Lambda^2 + \mathbf{I}\right) \circ \left(\mathbf{K}^{-1} y(\mathbf{X}^\intercal)\right) \tag{14}$$

and

$$\mathbf{F} := \Theta\left(\Lambda^2 + \mathbf{I}\right)^{-1} \mathbf{X}^\intercal \tag{15}$$

$$\Sigma := \Theta\left(\Lambda^{-2} + \mathbf{I}\right)^{-1} \Theta^\intercal \tag{16}$$

According to these formulae, the unconditional surrogate response is

$$\bar{f}_{\mathbf{0}} = \mathsf{E}\left[\bar{f}(\mathbf{x})\right] = (\mathbf{1})_{1\times\mathbf{N}}\, \bar{\mathbf{f}}_0^\intercal \tag{17}$$

which does not depend on Θ of course. Standardization of $y(\mathbf{X}^\intercal)$ instills an expectation of precisely zero here if $\mathbf{x}_n \sim \mathsf{N}[(\mathbf{0})_{\mathbf{M}}, (\mathbf{I})_{\mathbf{M}\times\mathbf{M}}]$ (which is often not exactly true).

Conditional variances may now be calculated as

$$D_{\mathbf{m}}((\Theta)_{\mathbf{m}\times\mathbf{M}}) := \mathsf{Var}\left[\bar{f}_{\mathbf{m}}((\mathbf{u})_{\mathbf{m}})\right] = \bar{\mathbf{f}}_0\, \mathbf{D}((\Theta)_{\mathbf{m}\times\mathbf{M}})\, \bar{\mathbf{f}}_0^\intercal - \bar{f}_{\mathbf{0}}^2 \tag{18}$$

where

$$(\mathbf{D}((\Theta)_{\mathbf{m}\times\mathbf{M}}))_{n\times o} := \frac{p((\mathbf{0})_{\mathbf{m}}; ((\mathbf{F})_{\mathbf{m}\times n} - (\mathbf{F})_{\mathbf{m}\times o}), 2(\Sigma)_{\mathbf{m}\times\mathbf{m}})}{p((\mathbf{0})_{\mathbf{m}}; ((\mathbf{F})_{\mathbf{m}\times n} + (\mathbf{F})_{\mathbf{m}\times o}), 2(\Psi)_{\mathbf{m}\times\mathbf{m}})} \tag{19}$$

and

$$\Psi := \Theta\left(\Lambda^{-2} + \mathbf{I}\right)^{-1} \left(2\Lambda^{-2} + \mathbf{I}\right) \Theta^\intercal \tag{20}$$

It is worth noting that the totally conditional variance $D_{\mathbf{M}}(\Theta) =: D_{\mathbf{M}}$ is independent of Θ and is ideally equal to $1 - \sigma_{\mathbf{e}}^2$.

3.3. Basis Optimization

At this point in the analysis, everything has been fixed save the rotation Θ used to determine the marginalized sampling distribution

$$\mathbf{u}_{\mathbf{m}} = (\Theta\mathbf{x})_{\mathbf{m}} = (\Theta)_{\mathbf{m}\times\mathbf{M}}(\mathbf{x})_{\mathbf{M}}$$

For the purpose of merely calculating the closed Sobol' indices

$$S_{\mathbf{m}}((\Theta)_{\mathbf{m}\times\mathbf{M}}) := \frac{D_{\mathbf{m}}((\Theta)_{\mathbf{m}\times\mathbf{M}})}{D_{\mathbf{M}}} \tag{21}$$

we may take $\Theta = (\mathbf{I})_{\mathbf{M}\times\mathbf{M}}$ and cancel it throughout Section 3.2, and our work is complete. To locate an active subspace, on the other hand, is to seek

$$(\Theta)_{\mathbf{m}\times\mathbf{M}} = \operatorname{argmax} S_{\mathbf{m}}((\Theta)_{\mathbf{m}\times\mathbf{M}}) \quad \text{such that} \quad S_{\mathbf{m}} \approx 1 \tag{22}$$

This is a fraught and onerous calculation, featuring an infinitude of optima for $m > 1$ (corresponding to alternative bases for the same active subspace). It is far more cogent and efficient to focus on the marginalized dimensions $\mathbf{M} \setminus \mathbf{m}$ via the total Sobol' index

$$S^{T}_{\mathbf{M}\setminus\mathbf{m}}((\Theta)_{\mathbf{M}\setminus\mathbf{m}\times\mathbf{M}}) := 1 - S_{\mathbf{m}}((\Theta)_{\mathbf{m}\times\mathbf{M}}) =: 1 - \frac{\tilde{D}_{\mathbf{m}}((\Theta)_{\mathbf{M}\setminus\mathbf{m}\times\mathbf{M}})}{D_{\mathbf{M}}} \tag{23}$$

This directly measures the output variance lost by ignoring the input dimensions in $\mathbf{M} \setminus \mathbf{m}$, without fussing over how the output variance is distributed over the remaining m input dimensions. This enables fast and efficient optimisation, with the flexibility to make $M - m$ as small as needed, whereas m is bounded below by the number of dimensions needed to explain most of the output variance.

It is clear from Eqs. (18), (21) and (23) that

$$\begin{aligned}\bar{\mathbf{f}}_0\, \tilde{\mathbf{D}}((\Theta)_{\mathbf{M}\setminus\mathbf{m}\times\mathbf{M}})\, \bar{\mathbf{f}}_0^{\intercal} - \bar{f}_{\mathbf{0}}^2 &:= \tilde{D}_{\mathbf{m}}((\Theta)_{\mathbf{M}\setminus\mathbf{m}\times\mathbf{M}}) \\ &:= D_{\mathbf{m}}((\Theta)_{\mathbf{m}\times\mathbf{M}}) = \bar{\mathbf{f}}_0\, \mathbf{D}((\Theta)_{\mathbf{m}\times\mathbf{M}})\, \bar{\mathbf{f}}_0^{\intercal} - \bar{f}_{\mathbf{0}}^2\end{aligned}$$

so $\mathbf{D}((\Theta)_{\mathbf{m}\times\mathbf{M}})$ must be expressable in terms of $(\Theta)_{\mathbf{M}\setminus\mathbf{m}\times\mathbf{M}}$ alone. This is achieved using Eq. (4) to write

$$(\tilde{\mathbf{D}}((\Theta)_{\mathbf{M}\setminus\mathbf{m}\times\mathbf{M}}))_{n\times o} := \frac{p\Big((\mathbf{0})_{\mathbf{m}}; \Big((\tilde{\mathbf{F}})_{\mathbf{m}\times n} - (\tilde{\mathbf{F}})_{\mathbf{m}\times o}\Big), 2(\tilde{\Sigma})_{\mathbf{m}\times\mathbf{m}}\Big)}{p\Big((\mathbf{0})_{\mathbf{m}}; \Big((\tilde{\mathbf{F}})_{\mathbf{m}\times n} + (\tilde{\mathbf{F}})_{\mathbf{m}\times o}\Big), 2(\tilde{\Psi})_{\mathbf{m}\times\mathbf{m}}\Big)} \tag{24}$$

where

$$\begin{aligned}
\tilde{\mathbf{F}} &:= \tilde{\Theta}\left(\Lambda^2+\mathbf{I}\right)^{-1}\mathbf{X}^\intercal \\
\tilde{\Sigma} &:= \tilde{\Theta}\left(\Lambda^{-2}+\mathbf{I}\right)^{-1}\tilde{\Theta}^\intercal \\
\tilde{\Psi} &:= \tilde{\Theta}\left(\Lambda^{-2}+\mathbf{I}\right)^{-1}\left(2\Lambda^{-2}+\mathbf{I}\right)\tilde{\Theta}^\intercal
\end{aligned} \tag{25}$$

and

$$\tilde{\Theta} := (\Theta)^\intercal_{\mathbf{m}\times\mathbf{m}}(\Theta)_{\mathbf{m}\times\mathbf{M}} = (\mathbf{I})_{\mathbf{m}\times\mathbf{M}} - (\Theta)^\intercal_{\mathbf{M}\backslash\mathbf{m}\times\mathbf{m}}(\Theta)_{\mathbf{M}\backslash\mathbf{m}\times\mathbf{M}} \tag{26}$$

A sufficient condition for validity is invertibility, which, using Weinstein-Aronszajn [21], may be expressed as

$$\left|(\Theta)_{\mathbf{m}\times\mathbf{m}}\right|^2 = \left|(\Theta)_{\mathbf{M}\backslash\mathbf{m}\times\mathbf{M}\backslash\mathbf{m}}\right|^2 > 0 \tag{27}$$

Taking $M-m>1$ can potentially to eliminate an $M-m$ dimensional space whose influence is even less than the $M-m$ least influential inputs found and eliminated in order. However, this does not seem worth the greatly increased complexity and infinitude of local optima introduced. We therefore take $M-m=1$ and seek

$$(\Theta)_{M\times\mathbf{M}} = \operatorname{argmin} S^T_M((\Theta)_{M\times\mathbf{M}}) \quad \text{such that} \quad S^T_M \approx 0 \tag{28}$$

to construct the unique RQ decomposition

$$\mathbf{R}\Theta = \left(\mathring{\Theta}\right)_{\mathbf{M}\times\mathbf{M}} \quad \text{where} \quad \left(\mathring{\Theta}\right)_{m\times\mathbf{M}} := \begin{cases}(\mathbf{I})_{m\times\mathbf{M}} & m<M \\ (\Theta)_{M\times\mathbf{M}} & m=M\end{cases} \tag{29}$$

for upper triangular $\mathbf{R}$. By construction, the resulting Θ is orthogonal (thanks to Eq. (27)) and has the final row $(\Theta)_{M\times\mathbf{M}}$ caclulated by Eq. (28).

$$\begin{aligned}
&\text{For } m\in\mathbf{M}-\mathbf{1} \\
\frac{\partial}{\partial(\Theta)_{M\times m}} &= \sum_{k=1}^{M-1}\sum_{l=1}^{M}\frac{\partial(\tilde{\Theta})_{k\times l}}{\partial(\Theta)_{M\times m}}\frac{\partial}{\partial(\tilde{\Theta})_{k\times l}} \\
&= -\sum_{k=1}^{M-1}(\Theta)_{M\times k}\frac{\partial}{\partial(\tilde{\Theta})_{k\times m}} - \sum_{l=1}^{M}(\Theta)_{M\times l}\frac{\partial}{\partial(\tilde{\Theta})_{m\times l}} \\
\frac{\partial}{\partial(\Theta)_{M\times M}} &= \sum_{k=1}^{M-1}\frac{\partial(\tilde{\Theta})_{k\times M}}{\partial(\Theta)_{M\times M}}\frac{\partial}{\partial(\tilde{\Theta})_{k\times M}} = -\sum_{k=1}^{M-1}(\Theta)_{M\times k}\frac{\partial}{\partial(\tilde{\Theta})_{k\times M}}
\end{aligned} \tag{30}$$

$$
\begin{aligned}
&\text{For } m \in \mathbf{M-1} \\
&(\Theta)_{M\times M} = \left(1 - \sum (\Theta)^2_{M\times m}\right)^{1/2} \\
&\frac{\partial}{\partial(\Theta)_{M\times m}} = \frac{\partial}{\partial(\Theta)_{M\times m}} - \frac{(\Theta)_{M\times m}}{(\Theta)_{M\times M}} \frac{\partial}{\partial(\Theta)_{M\times M}}
\end{aligned}
\tag{31}
$$

$$
\begin{aligned}
\frac{\partial}{\partial(\Theta)_{M\times m}} = &-(\Theta)_{M\times M} \frac{\partial}{\partial(\tilde{\Theta})_{m\times M}} \\
&+ \sum_{k=1}^{M-1} (\Theta)_{M\times k} \left(\frac{(\Theta)_{M\times m}}{(\Theta)_{M\times M}} \frac{\partial}{\partial(\tilde{\Theta})_{k\times M}} - \frac{\partial}{\partial(\tilde{\Theta})_{k\times m}} - \frac{\partial}{\partial(\tilde{\Theta})_{m\times k}} \right)
\end{aligned}
\tag{32}
$$

Algorithm 1 Summary of the basis optimization algorithm.

Input: $y(\mathbf{X}^\intercal)\colon \mathbb{R}^{M\times N} \to \mathbb{R}$
Output: $y(\mathbf{U}^\intercal)\colon \mathbb{R}^{m\times N} \to \mathbb{R}$ where $m \leq M$

loop
    Fit GP surrogate to $y(\mathbf{X}^\intercal)$, determining $\bar{f}(\mathbf{x})$ according to Section 3.1
    Swap input dimensions M and $k \leq M$ to maximise $D_{\mathbf{M-1}}((\mathbf{I})_{\mathbf{M-1}\times\mathbf{M}})$, calculated according to Section 3.2
    Optimize $(\Theta)_{M\times\mathbf{M}} = \operatorname{argmin} S^T_M((\Theta)_{M\times\mathbf{M}})$ according to Section 3.3
    if $S^T_M \not\approx 0$ **then**
        return $y(\mathbf{X}^\intercal)$
    end if
    Set Θ by RQ decomposition, according to Eq. (29)
    Update the input basis to $\mathbf{X}^\intercal \leftarrow \Theta\mathbf{X}^\intercal$
    Eliminate the final input dimension, setting $M \leftarrow M - 1$
end loop

4. Results

In this section, the method described in **??** is applied to a series of test functions to evaluate its performance. Each function takes $N \in \{100, 200, 400, 800, 1600\}$ data from a latin hypercube of $M = 5$ input dimensions. Random noise $\epsilon(\mathbf{x}) \sim \mathsf{N}[0, \sigma_\epsilon^2]$ is applied to the output, for $\sigma_\epsilon \in \{0.1, 0.05, 0.01\}$. All inputs and outputs are then standardized to mean 0, standard deviation 1 before folding. All results are calculated as the mean over two folds (each with N training data and N test data, so predictions are rigorously cross-validated.

In each case an $N \times M$ design matrix $\mathbf{X}$ is sampled from a standard normal distribution (latin hypercube). The input to the test function $f\colon [x_-, x_+]^M \to \mathbb{R}$ is generally constructed as

$$\hat{\mathbf{X}}^\intercal = (x_+ - x_-)c(\Phi \mathbf{X}^\intercal) + x_-(\mathbf{1})_{\mathbf{M}\times\mathbf{N}} \tag{33}$$

where $c\colon \mathbb{R}^M \to \mathbb{R}^M$ is the cumulative density function for M independent standard normal random variables, and Φ is a test rotation matrix. The corresponding optimal input rotation from **??** is

$$\Theta_\Pi = \begin{cases} \Theta_{\mathbf{1}} & \text{if } \Phi \text{ is identity matrix } \mathbf{1} \\ \Theta_{\mathbf{R}} & \text{if } \Phi \text{ is a random rotation matrix } \Phi_{\mathbf{R}} \end{cases} \tag{34}$$

which should recover the random rotation as

$$\Theta_{\mathbf{R}} \cong \Theta_{\mathbf{1}}\Phi_{\mathbf{R}} \tag{35}$$

However, this is congruence, not equality: different rotations might locate (excactly or nearly exactly) the same active subspace.

For each function, the intial GP fit is assessed by test statistics from independent data (from the other fold), together with errors in the calculated Sobol' indices. The latter are important as they are at the heart of subsequent calculations. The input basis is then optimized, calculating $\Theta_{\mathbf{1}}$. A reduced dimensionality $\underline{M}$ for the optimized basis is determined as

$$\min\left\{\underline{M} \le M \;\middle|\; S_{\underline{\mathbf{M}}} \ge 0.90\right\} \tag{36}$$

A GP is fit to this reduced input, and its test statistics compared with the initial GP.

The whole procedure is then repeated (with entirely fresh data) to which a random input rotation $\Phi_{\mathbf{R}}$ is applied. The input basis is optimized, calculating $\Theta_{\mathbf{R}}$, whereas the reduced dimensionality $\underline{M}$ is not re-assessed, but retained from the unrotated analysis.

Finally the rotated and unrotated active subspaces are compared for congruence, using the ordered sigular values $\Sigma_m(\mathbf{u}^\dagger)$ of

$$\mathbf{u}^\dagger = (\Theta_{\mathbf{1}}\Phi_{\mathbf{R}}\Theta_{\mathbf{R}}^\intercal)_{\underline{\mathbf{M}}\times\underline{\mathbf{M}}} \tag{37}$$

This matrix transforms the active subspace according to $\Theta_{\mathbf{R}}$ into the active subspace according to $\Theta_{\mathbf{1}}\Phi_{\mathbf{R}}$ without straying outside the union of two. The basis vector length(s) lost to the inactive subspaces in doing this is $(\mathbf{1} - \Sigma_m(\mathbf{u}^\dagger))$.

4.1. Sine Function

$$f(\hat{\mathbf{x}}) := \sin(\hat{\mathbf{x}}_1) \tag{38}$$

$$[x_-, x_+] := [-\pi, +\pi]$$
$$S_{\mathbf{1}} = 1$$

Fitting a GP to the unrotated sine function turns out to be rather straightforward. Exact Sobol' indices are recovered to within 0.1% even with only 100 training data. Optimizing the input basis has no significant impact on the Sobol' indices or cross-validation tests.

N	σ_ϵ (%)	$S_{\mathbf{1}}$ (%)		
100	14.20	10.52	99.95	99.90
100	7.04	0.27	100.00	100.00
100	1.41	0.77	100.00	100.00
200	13.99	2.98	100.00	100.00
200	7.05	21.10	100.00	99.99
200	1.42	35.84	100.00	100.00
400	14.00	31.44	100.00	99.99
400	7.06	5.76	100.00	100.00
400	1.41	5.96	100.00	100.00
800	13.94	12.03	99.99	100.00
800	7.05	29.41	100.00	100.00
800	1.41	0.40	100.00	100.00
1600	13.99	26.94	100.00	100.00
1600	7.05	2.68	100.00	100.00
1600	1.41	0.29	100.00	100.00

Table 1: Closed Sobol' indices calculated from initial (left sub-column) and optimized (middle sub-column) GPs, for the randomly rotated sine function. For comparison the Sobol' indices for the optimized GP on the unrotated sine function are shown in the rightmost sub-column.

Turning immediately to the sine function with randomly rotated inputs, the first closed Sobol' index is shown in Table 1. In all cases, the random rotation is effectively reversed within just 5 iterations of input basis optimization, as indicated by recovering virtually exact Sobol' indices. This is confirmed by the cross-validation predictive performance of the optimized

GP, shown in Table 2. It may be seen that the predictive uncertainty $\sigma_{f(\mathbf{x})}$ closely replicates the noise σ_ϵ inherent in the test function, and the outliers are close to the 5% expected of the normal distribution. Reducing dimensionality by fitting a GP to just the principal direction of the optimized basis barely affects the predictive performance.

N	σ_ϵ (%)	$\sigma_{f(\mathbf{x})}$ (%)			Outliers (%)			RMSE (%)		
100	13.74	29.68	14.92	15.10	3.50	6.00	6.00	7.49	3.94	3.91
100	7.04	23.32	7.70	7.70	3.50	7.00	7.00	5.54	1.98	1.98
100	1.47	33.86	1.70	1.70	9.50	7.50	7.50	10.22	0.45	0.45
200	14.40	27.71	15.10	15.10	4.25	4.50	4.50	6.99	4.02	4.02
200	6.93	16.53	7.15	7.15	7.25	4.25	4.25	5.30	1.83	1.83
200	1.41	23.97	1.76	1.77	7.50	6.00	6.25	7.31	0.47	0.47
400	14.12	20.72	14.33	14.33	3.88	4.50	4.62	4.84	3.73	3.73
400	7.06	17.68	7.32	7.32	4.88	4.25	4.25	4.66	1.80	1.80
400	1.43	18.13	1.61	1.61	2.88	5.75	5.88	4.72	0.38	0.38
800	13.78	21.64	13.86	13.89	4.62	4.50	4.31	5.84	3.47	3.47
800	7.15	15.91	7.31	7.31	4.12	4.62	4.62	4.11	1.81	1.81
800	1.42	4.72	1.44	1.44	4.44	4.44	4.44	1.39	0.36	0.36
1600	14.00	17.78	14.07	14.07	4.84	4.59	4.59	4.47	3.53	3.53
1600	7.04	11.44	7.07	7.07	3.94	4.81	4.81	3.02	1.77	1.77
1600	1.41	2.09	1.43	1.43	4.34	4.53	4.53	0.50	0.35	0.35

Table 2: Predictive performance of initial GPs (left sub-columns) after optimizing the input basis (middle sub-columns) and reducing dimensionality (right sub-columns), for the sine function with randomly rotated inputs. Three measures are shown: the GPs' predictive standard deviation $\sigma_{f(\mathbf{x})}$, the percentage of observations outside $\pm 2\sigma_{f(\mathbf{x})}$, and the Root Mean Square Error.

As expected, the active subspace measures $\Sigma_1(\mathbf{u}^\dagger)$ are all close to 100%, confirming that the random rotation has been reversed.

Noise	N	$\Sigma_1(\mathbf{u}^\dagger)$
13.74	100	99.91
7.04	100	99.87
1.47	100	100.00
14.40	200	99.93
6.93	200	99.98
1.41	200	99.90
14.12	400	99.97
7.06	400	99.97
1.43	400	99.98
13.78	800	100.00
7.15	800	99.97
1.42	800	99.95
14.00	1600	100.00
7.04	1600	100.00
1.41	1600	100.00

Table 3: The active subspace measures $\Sigma_m(\mathbf{u}^\dagger)$ for the sine function, comparing optimization of unrotated and randomly rotated inputs.

4.2. Decoupled Ishigami Function

$$f(\mathbf{x}) := \left(1 + b\mathbf{x}_3^4\right)\sin(\mathbf{x}_1) + a\sin^2(\mathbf{x}_2) \tag{39}$$

$$\begin{aligned} a &= 2.0 \quad ; \quad b = 0 \\ S_{\mathbf{1}} &= 0.5 \quad ; \quad S_{\mathbf{2}} = 1 \end{aligned}$$

The predictive performance is shown in Table 5.

N	σ_ϵ (%)	$S_{\mathbf{1}}$ (%)		$S_{\mathbf{2}}$ (%)	
100	10.19	75.06	75.33	99.94	99.94
100	4.77	69.53	70.76	97.02	97.82
100	1.00	67.94	70.77	96.88	100.00
200	10.11	50.48	50.48	99.99	99.99
200	5.19	49.81	49.82	100.00	100.00
200	0.99	49.96	49.96	100.00	100.00
400	10.09	49.89	50.23	100.00	100.00
400	4.98	50.18	50.15	100.00	100.00
400	1.03	50.21	50.21	100.00	100.00
800	9.79	50.06	50.22	100.00	100.00
800	4.98	50.24	50.38	100.00	100.00
800	1.03	50.10	50.26	100.00	100.00
1600	10.04	50.09	50.09	100.00	100.00
1600	5.08	49.87	50.13	100.00	100.00
1600	1.01	50.10	50.10	100.00	100.00

Table 4: Closed Sobol' indices calculated from initial (left sub-column) and optimized (right sub-column) GPs, for the decoupled Ishigami function.

N	σ_ϵ (%)	$\sigma_{f(\mathbf{x})}$ (%)			Outliers (%)			RMSE (%)		
100	10.19	46.27	46.31	24.76	6.50	7.00	5.00	13.35	13.33	7.09
100	4.77	37.48	37.69	21.37	7.00	7.50	7.00	10.21	10.34	6.58
100	1.00	36.60	37.06	21.29	9.00	8.50	11.50	11.55	11.55	7.70
200	10.11	14.72	14.75	17.29	5.25	5.50	4.75	4.71	4.72	5.56
200	5.19	10.53	10.53	12.52	4.75	4.50	7.75	3.81	3.82	5.38
200	0.99	4.11	4.11	6.33	6.25	6.25	4.75	2.51	2.51	3.38
400	10.09	12.67	12.73	14.31	3.75	3.75	3.62	3.48	3.47	4.04
400	4.98	7.40	7.37	8.80	3.25	3.50	3.50	2.41	2.42	3.17
400	1.03	2.71	2.71	3.56	4.50	4.50	5.12	1.80	1.80	2.21
800	9.79	11.01	11.01	11.62	3.88	3.88	3.62	2.84	2.84	3.11
800	4.98	5.75	5.74	6.14	4.19	4.19	4.56	1.55	1.55	1.78
800	1.03	1.82	1.82	2.29	4.00	4.06	3.62	1.37	1.37	1.67
1600	10.04	10.64	10.64	10.89	4.81	4.78	5.09	2.71	2.71	2.85
1600	5.08	5.68	5.68	5.97	4.16	4.16	4.16	1.65	1.65	1.90
1600	1.01	1.34	1.34	1.51	4.53	4.53	4.69	0.61	0.61	0.82

Table 5: Predictive performance of initial GPs (left sub-columns) after optimizing the input basis (middle sub-columns) and reducing dimensionality (right sub-columns), for the decoupled Ishigami function. Three measures are shown: the GPs' predictive standard deviation $\sigma_{f(\mathbf{x})}$, the percentage of observations outside $\pm 2\sigma_{f(\mathbf{x})}$, and the Root Mean Square Error.

N	σ_ϵ (%)	$S_{\mathbf{1}}$ (%)			$S_{\mathbf{2}}$ (%)		
100	10.19	29.00	99.83	75.33	32.27	100.00	99.94
100	4.77	3.17	98.88	70.76	85.58	100.00	97.82
100	1.00	5.77	95.34	70.77	14.72	100.00	100.00
200	10.11	22.72	72.51	50.48	36.51	79.71	99.99
200	5.19	0.23	49.30	49.82	14.39	54.95	100.00
200	0.99	0.15	72.59	49.96	4.08	76.76	100.00
400	10.09	0.85	51.47	50.23	13.19	54.77	100.00
400	4.98	10.44	48.65	50.15	16.46	62.05	100.00
400	1.03	1.71	47.79	50.21	2.70	50.56	100.00
800	9.79	2.49	49.50	50.22	29.38	52.33	100.00
800	4.98	11.00	49.08	50.38	11.90	55.63	100.00
800	1.03	13.64	47.75	50.26	14.80	50.02	100.00
1600	10.04	19.62	50.73	50.09	34.44	61.96	100.00
1600	5.08	1.03	50.14	50.13	23.19	52.04	100.00
1600	1.01	1.29	49.84	50.10	2.83	52.82	100.00

Table 6: Closed Sobol' indices calculated from initial (left sub-column) and optimized (middle sub-column) GPs, for the randomly rotated decoupled Ishigami function. For comparison the Sobol' indices for the optimized GP on the unrotated Ishigami function are shown in the rightmost sub-column.

N	σ_ϵ (%)	$\sigma_{f(\mathbf{x})}$ (%)			Outliers (%)			RMSE (%)		
100	9.80	71.00	65.49	68.28	4.50	6.00	8.00	19.45	19.69	19.52
100	4.84	75.36	70.59	63.27	0.50	0.00	6.00	18.52	17.82	17.47
100	0.93	73.65	66.58	69.50	4.50	3.00	2.50	19.72	17.81	18.51
200	9.92	68.51	65.62	70.84	6.50	3.00	3.25	18.76	16.79	18.62
200	5.22	67.33	54.93	74.32	4.75	4.50	0.50	17.21	13.83	18.67
200	0.98	71.56	67.24	71.14	2.25	1.75	2.00	18.51	17.51	18.47
400	10.15	65.91	63.34	72.22	1.88	3.00	0.25	15.67	15.48	18.71
400	4.93	36.60	30.44	64.65	4.25	3.12	1.25	9.41	7.98	16.07
400	1.03	64.36	58.27	74.30	2.88	3.12	0.38	14.98	13.97	18.55
800	10.14	50.82	47.74	72.24	3.50	2.81	0.19	12.71	12.05	18.13
800	4.78	36.66	21.99	65.75	5.19	5.25	0.19	10.01	6.45	16.34
800	1.00	52.30	51.96	71.18	4.56	4.06	0.19	13.26	13.22	17.77
1600	10.00	24.67	22.93	63.78	3.34	4.00	1.34	6.62	6.11	15.79
1600	4.98	36.89	26.35	70.16	4.03	4.38	0.34	9.55	7.13	17.58
1600	1.01	38.28	17.33	69.72	3.94	5.28	0.19	10.27	5.43	17.38

Table 7: Predictive performance of initial GPs (left sub-columns) after optimizing the input basis (middle sub-columns) and reducing dimensionality (right sub-columns), for the decoupled Ishigami function with randomly rotated inputs. Three measures are shown: the GPs' predictive standard deviation $\sigma_{f(\mathbf{x})}$, the percentage of observations outside $\pm 2\sigma_{f(\mathbf{x})}$, and the Root Mean Square Error.

N	σ_ϵ (%)	$\Sigma_1(\mathbf{u}^\dagger)$	$\Sigma_2(\mathbf{u}^\dagger)$
100	9.80	29.11	99.50
100	4.84	50.46	99.55
100	0.93	41.32	99.78
200	9.92	87.65	99.79
200	5.22	79.94	99.93
200	0.98	67.85	99.30
400	10.15	67.81	99.88
400	4.93	93.87	99.99
400	1.03	63.52	99.95
800	10.14	70.93	100.00
800	4.78	88.92	99.99
800	1.00	52.23	99.98
1600	10.00	92.87	100.00
1600	4.98	68.48	99.99
1600	1.01	78.48	99.99

Table 8: The active subspace measures $\Sigma_m(\mathbf{u}^\dagger)$ for the decoupled Ishigami function, comparing optimization of unrotated and randomly rotated inputs.

4.3. Ishigami Function

$$f(\mathbf{x}) := \left(1 + b\mathbf{x}_3^4\right)\sin(\mathbf{x}_1) + a\sin^2(\mathbf{x}_2) \tag{40}$$

$$a = 7.0 \quad ; \quad b = 0.1$$
$$S_{\mathbf{1}} = 0.3139 \quad ; \quad S_{\mathbf{2}} = 0.7563 \quad ; \quad S_{\mathbf{3}} = 1$$

N	σ_ϵ (%)	$S_{\mathbf{1}}$ (%)		$S_{\mathbf{2}}$ (%)		$S_{\mathbf{3}}$ (%)	
100	2.76	44.61	50.91	71.99	82.50	99.99	99.98
100	1.38	67.88	71.17	69.57	95.71	96.79	98.23
100	0.28	27.42	49.90	78.07	79.29	99.96	98.69
200	2.71	44.58	47.76	68.52	84.71	100.00	100.00
200	1.28	42.34	50.44	68.78	84.26	100.00	97.99
200	0.27	31.23	42.35	75.29	73.74	100.00	100.00
400	2.63	30.98	43.91	75.55	75.55	100.00	100.00
400	1.32	31.29	44.51	76.38	76.39	100.00	100.00
400	0.27	31.11	44.25	76.01	76.07	100.00	100.00
800	2.64	31.23	44.01	76.25	75.01	100.00	98.95
800	1.37	31.34	44.41	76.03	76.06	100.00	100.00
800	0.27	31.55	44.21	76.03	76.01	100.00	100.00
1600	2.65	31.39	44.33	75.85	75.85	100.00	100.00
1600	1.37	31.46	44.51	76.00	74.40	100.00	99.17
1600	0.27	31.56	44.03	75.75	74.70	100.00	98.32

Table 9: Closed Sobol' indices calculated from initial (left sub-column) and optimized (right sub-column) GPs, for the Ishigami function.

N	σ_ϵ (%)	$\sigma_{f(\mathbf{x})}$ (%)			Outliers (%)			RMSE (%)		
100	2.76	52.36	52.96	55.07	5.00	5.00	5.50	16.71	17.20	17.89
100	1.38	74.27	74.06	71.08	4.00	3.00	3.00	18.98	18.84	18.11
100	0.28	32.21	41.55	64.15	7.00	10.50	5.50	8.83	13.19	17.21
200	2.71	45.80	45.59	35.70	3.00	3.75	2.25	12.94	13.03	10.55
200	1.28	41.65	46.21	39.81	4.75	5.00	5.25	11.62	12.79	12.86
200	0.27	18.06	18.32	33.59	4.50	6.75	2.75	6.95	7.35	10.53
400	2.63	12.54	12.55	20.38	5.38	5.50	5.00	5.07	5.06	8.24
400	1.32	10.58	10.58	18.80	4.75	4.88	2.62	4.43	4.43	7.73
400	0.27	9.10	9.12	17.57	9.38	9.12	3.38	5.01	5.01	7.69
800	2.64	7.68	9.59	17.87	3.88	6.00	6.56	4.00	4.72	6.99
800	1.37	6.11	6.16	10.85	4.44	4.00	4.00	3.57	3.58	5.92
800	0.27	4.45	4.45	9.06	4.31	4.88	3.81	2.48	2.47	5.02
1600	2.65	5.16	5.20	7.55	4.19	4.50	4.25	2.02	2.02	3.24
1600	1.37	3.89	4.95	12.81	5.25	5.50	5.19	2.28	2.61	5.12
1600	0.27	2.24	3.44	13.51	4.22	4.78	5.94	1.39	1.69	5.05

Table 10: Predictive performance of initial GPs (left sub-columns) after optimizing the input basis (middle sub-columns) and reducing dimensionality (right sub-columns), for the Ishigami function. Three measures are shown: the GPs' predictive standard deviation $\sigma_{f(\mathbf{x})}$, the percentage of observations outside $\pm 2\sigma_{f(\mathbf{x})}$, and the Root Mean Square Error.

N	σ_ϵ (%)	$S_{\mathbf{1}}$ (%)			$S_{\mathbf{2}}$ (%)			$S_{\mathbf{3}}$ (%)		
100	2.76	0.43	40.74	50.91	19.67	52.66	82.50	55.73	79.44	99.98
100	1.38	14.82	56.23	71.17	32.05	69.44	95.71	50.55	84.24	98.23
100	0.28	18.85	62.59	49.90	34.92	80.17	79.29	48.68	96.43	98.69
200	2.71	3.84	53.48	47.76	12.55	74.54	84.71	28.00	86.01	100.00
200	1.28	2.99	49.59	50.44	10.59	70.37	84.26	14.39	77.61	97.99
200	0.27	4.26	49.31	42.35	60.09	59.98	73.74	71.74	72.13	100.00
400	2.63	7.19	38.91	43.91	7.69	46.30	75.55	13.51	61.63	100.00
400	1.32	0.71	45.30	44.51	10.00	59.85	76.39	10.81	85.86	100.00
400	0.27	3.15	45.06	44.25	9.30	58.94	76.07	17.50	72.51	100.00
800	2.64	0.98	43.17	44.01	5.31	60.45	75.01	10.95	75.89	98.95
800	1.37	0.16	45.12	44.41	10.98	56.70	76.06	15.47	68.68	100.00
800	0.27	0.19	43.80	44.21	11.48	57.90	76.01	42.21	75.55	100.00
1600	2.65	8.84	43.41	44.33	10.45	53.98	75.85	28.13	67.80	100.00
1600	1.37	3.54	44.01	44.51	13.25	56.53	74.40	15.51	73.77	99.17
1600	0.27	7.99	43.49	44.03	10.85	49.90	74.70	24.93	58.30	98.32

Table 11: Closed Sobol' indices calculated from initial (left sub-column) and optimized (middle sub-column) GPs, for the randomly rotated Ishigami function. For comparison the Sobol' indices for the optimized GP on the unrotated Ishigami function are shown in the rightmost sub-column.

N	σ_ϵ (%)	$\sigma_{f(\mathbf{x})}$ (%)			Outliers (%)			RMSE (%)		
100	2.61	77.47	69.52	78.43	6.50	12.50	5.00	22.62	20.80	22.25
100	1.46	69.42	61.53	86.66	8.50	13.50	5.50	21.90	20.76	23.64
100	0.26	78.86	72.37	76.27	10.00	7.50	6.50	24.85	22.26	22.90
200	2.63	75.42	56.79	82.32	4.25	7.75	4.75	19.62	17.43	22.04
200	1.31	61.03	42.96	74.33	5.75	6.50	7.50	17.57	12.51	19.54
200	0.27	51.09	40.96	73.04	7.25	10.75	8.00	17.53	14.19	21.60
400	2.74	61.52	44.23	70.38	5.62	8.00	6.00	16.24	13.51	20.15
400	1.38	60.09	25.79	52.92	5.75	7.00	7.75	16.19	8.51	15.25
400	0.27	53.06	31.91	72.56	6.00	9.50	5.75	15.66	10.78	18.04
800	2.64	37.29	21.33	55.14	7.00	6.25	6.88	12.19	7.68	14.97
800	1.34	37.33	25.15	65.26	7.06	7.69	6.94	11.92	9.46	18.33
800	0.28	31.46	18.00	61.22	5.38	7.06	6.44	10.15	6.78	17.16
1600	2.67	42.56	17.42	64.48	4.25	5.66	6.66	11.79	5.45	16.03
1600	1.34	24.41	14.63	57.32	4.88	5.72	6.22	8.28	5.20	14.19
1600	0.27	32.89	18.45	69.06	6.69	7.88	7.16	10.63	6.97	17.85

Table 12: Predictive performance of initial GPs (left sub-columns) after optimizing the input basis (middle sub-columns) and reducing dimensionality (right sub-columns), for the Ishigami function with randomly rotated inputs. Three measures are shown: the GPs' predictive standard deviation $\sigma_{f(\mathbf{x})}$, the percentage of observations outside $\pm 2\sigma_{f(\mathbf{x})}$, and the Root Mean Square Error.

N	σ_ϵ (%)	$\Sigma_1(\mathbf{u}^\dagger)$	$\Sigma_2(\mathbf{u}^\dagger)$	$\Sigma_3(\mathbf{u}^\dagger)$
100	2.61	67.14	92.67	100.00
100	1.46	24.85	80.24	100.00
100	0.26	59.12	93.67	100.00
200	2.63	73.91	99.27	100.00
200	1.31	33.05	97.77	100.00
200	0.27	49.88	96.75	100.00
400	2.74	86.15	91.47	100.00
400	1.38	88.95	99.46	100.00
400	0.27	37.99	99.45	100.00
800	2.64	79.69	96.10	100.00
800	1.34	72.23	99.86	100.00
800	0.28	78.19	99.95	100.00
1600	2.67	21.13	98.78	100.00
1600	1.34	39.74	98.74	100.00
1600	0.27	56.47	81.35	100.00

Table 13: The active subspace measures $\Sigma_m(\mathbf{u}^\dagger)$ for the Ishigami function, comparing optimization of unrotated and randomly rotated inputs.

4.4. Sobol' G Function

$$f(\mathbf{x}) := \prod_{i=1}^{D} \frac{|4\mathbf{x}_i - 2| + \mathbf{a}_i}{1 + \mathbf{a}_i} \tag{41}$$

$$\mathbf{a}_i = (i-1)/2$$

$$S_{\mathbf{1}} = 0.3575 \quad ; \quad S_{\mathbf{2}} = 0.8342 \quad ; \quad S_{\mathbf{3}} = 0.9930 \quad ; \quad S_{\mathbf{4}} = 0.9999 \quad ; \quad S_{\mathbf{5}} = 1$$

N	σ_ϵ (%)	$S_{\mathbf{1}}$ (%)		$S_{\mathbf{2}}$ (%)		$S_{\mathbf{3}}$ (%)	
100	11.02	27.76	39.04	78.19	71.33	96.94	93.04
100	5.11	36.44	38.78	84.64	84.18	99.87	99.86
100	1.07	35.45	36.75	82.04	79.12	98.78	99.79
200	10.01	36.82	39.13	87.00	87.06	99.97	99.97
200	5.15	36.72	39.37	84.46	83.39	100.00	99.12
200	0.99	38.45	40.77	85.19	80.68	99.96	99.85
400	10.39	36.03	36.23	84.38	84.34	99.98	99.98
400	5.22	36.74	36.89	83.07	83.19	99.99	100.00
400	1.00	36.09	37.13	84.21	84.36	99.97	99.97
800	10.23	36.18	37.19	84.66	84.68	99.95	99.95
800	5.27	36.07	37.16	84.25	84.30	100.00	100.00
800	1.06	36.55	36.75	84.27	84.22	99.68	99.67
1600	10.24	35.82	36.99	84.39	84.41	100.00	100.00
1600	5.20	35.79	37.08	84.48	84.47	99.75	99.84
1600	1.03	36.57	36.57	83.83	83.83	99.43	99.43

Table 14: Closed Sobol' indices calculated from initial (left sub-column) and optimized (right sub-column) GPs, for the Sobol' G function.

N	σ_ϵ (%)	$\sigma_{f(\mathbf{x})}$ (%)			Outliers (%)			RMSE (%)		
100	11.02	39.25	44.59	53.36	19.50	17.50	6.50	18.07	18.49	17.29
100	5.11	31.77	31.92	33.55	15.50	16.00	12.00	17.21	17.25	16.98
100	1.07	32.45	33.57	36.96	14.50	11.00	6.50	20.14	17.84	17.45
200	10.01	28.04	28.00	30.06	9.75	9.25	5.50	12.05	12.05	11.64
200	5.15	24.86	25.82	29.90	7.75	10.25	6.25	9.61	10.31	10.48
200	0.99	22.35	23.57	25.72	13.25	14.75	10.50	11.56	12.08	11.77
400	10.39	25.22	25.23	25.83	6.38	6.38	7.50	9.79	9.80	10.29
400	5.22	20.99	21.03	22.09	6.75	6.62	6.38	9.47	9.42	9.31
400	1.00	17.89	17.89	19.29	8.75	8.50	8.75	8.10	8.08	8.21
800	10.23	20.62	20.62	21.21	5.81	5.88	5.44	6.97	6.97	7.03
800	5.27	17.07	17.11	17.79	7.75	8.38	6.88	7.97	7.96	8.04
800	1.06	15.34	15.36	16.31	10.06	10.06	8.50	7.56	7.51	7.52
1600	10.24	17.97	17.97	18.54	5.94	5.88	5.97	6.63	6.63	6.85
1600	5.20	14.38	14.94	15.23	6.59	5.75	6.62	6.36	6.36	6.48
1600	1.03	12.32	12.32	13.52	8.41	8.31	8.31	6.33	6.32	6.53

Table 15: Predictive performance of initial GPs (left sub-columns) after optimizing the input basis (middle sub-columns) and reducing dimensionality (right sub-columns), for the Sobol' G function. Three measures are shown: the GPs' predictive standard deviation $\sigma_{f(\mathbf{x})}$, the percentage of observations outside $\pm 2\sigma_{f(\mathbf{x})}$, and the Root Mean Square Error.

N	σ_ϵ (%)	$S_{\mathbf{1}}$ (%)			$S_{\mathbf{2}}$ (%)			$S_{\mathbf{3}}$ (%)		
100	11.02	3.17	49.47	39.04	21.17	72.03	71.33	25.74	83.55	93.04
100	5.11	4.33	39.60	38.78	26.78	64.08	84.18	42.45	83.42	99.86
100	1.07	14.25	43.01	36.75	33.13	56.03	79.12	46.89	72.78	99.79
200	10.01	15.37	39.04	39.13	21.38	51.18	87.06	30.83	78.02	99.97
200	5.15	0.62	43.22	39.37	35.85	50.93	83.39	56.13	62.86	99.12
200	0.99	19.49	41.03	40.77	30.87	63.12	80.68	64.40	81.73	99.85
400	10.39	0.18	39.00	36.23	5.37	53.91	84.34	10.47	70.06	99.98
400	5.22	8.18	38.68	36.89	15.42	58.85	83.19	18.49	79.23	100.00
400	1.00	1.74	37.61	37.13	2.49	51.89	84.36	21.67	72.71	99.97
800	10.23	17.11	37.28	37.19	64.76	66.82	84.68	78.64	80.70	99.95
800	5.27	3.61	37.62	37.16	24.77	60.52	84.30	27.95	78.46	100.00
800	1.06	6.59	38.50	36.75	23.39	51.09	84.22	30.08	69.38	99.67
1600	10.24	1.67	37.30	36.99	15.00	51.96	84.41	23.31	71.69	100.00
1600	5.20	5.42	37.12	37.08	12.63	60.68	84.47	28.41	80.59	99.84
1600	1.03	0.46	36.74	36.57	7.48	50.30	83.83	26.29	69.24	99.43

Table 16: Closed Sobol' indices calculated from initial (left sub-column) and optimized (middle sub-column) GPs, for the randomly rotated Sobol' G function. For comparison the Sobol' indices for the optimized GP on the unrotated Sobol' G function are shown in the rightmost sub-column.

N	σ_ϵ (%)	$\sigma_{f(\mathbf{x})}$ (%)			Outliers (%)			RMSE (%)		
100	9.69	56.69	53.39	66.34	8.50	9.50	7.50	19.84	20.64	20.63
100	5.31	58.73	53.55	63.07	12.50	16.00	6.50	21.36	21.14	19.31
100	1.05	55.76	47.48	62.54	15.00	11.00	9.50	19.62	17.52	21.10
200	11.21	50.29	44.70	61.41	11.75	13.25	8.75	17.06	17.34	19.68
200	5.28	49.00	48.78	70.07	9.00	9.50	7.75	16.72	16.73	20.73
200	1.10	46.33	43.39	57.67	11.25	8.25	10.50	15.81	13.79	18.85
400	10.51	38.74	38.88	63.28	6.38	6.25	7.25	12.85	13.26	18.34
400	5.05	35.32	31.95	52.97	8.50	8.38	7.88	12.61	11.90	15.66
400	1.05	33.24	32.86	54.23	8.62	7.13	9.50	12.57	12.29	15.90
800	10.10	27.12	27.03	49.80	5.75	5.44	6.69	9.22	9.28	14.26
800	4.97	26.70	24.11	53.04	7.88	9.75	5.94	9.54	9.27	14.37
800	1.01	29.47	27.20	56.37	8.31	9.81	8.12	10.16	10.19	16.67
1600	10.03	27.97	24.61	56.99	5.81	6.81	6.03	8.86	8.06	14.84
1600	5.21	26.23	22.64	47.89	6.53	7.72	7.09	9.21	8.35	13.39
1600	1.06	23.51	22.50	57.52	7.06	8.12	6.31	8.81	8.57	15.33

Table 17: Predictive performance of initial GPs (left sub-columns) after optimizing the input basis (middle sub-columns) and reducing dimensionality (right sub-columns), for the Sobol' G function with randomly rotated inputs. Three measures are shown: the GPs' predictive standard deviation $\sigma_{f(\mathbf{x})}$, the percentage of observations outside $\pm 2\sigma_{f(\mathbf{x})}$, and the Root Mean Square Error.

N	σ_ϵ (%)	$\Sigma_1(\mathbf{u}^\dagger)$	$\Sigma_2(\mathbf{u}^\dagger)$	$\Sigma_3(\mathbf{u}^\dagger)$
100	9.69	27.43	96.49	100.00
100	5.31	59.84	97.92	100.00
100	1.05	57.92	96.17	100.00
200	11.21	72.26	92.40	100.00
200	5.28	61.27	99.03	100.00
200	1.10	84.88	99.02	100.00
400	10.51	76.58	94.78	100.00
400	5.05	64.18	97.78	100.00
400	1.05	55.92	99.80	100.00
800	10.10	93.04	99.63	100.00
800	4.97	29.41	98.23	100.00
800	1.01	42.15	99.70	100.00
1600	10.03	40.14	99.87	100.00
1600	5.21	47.85	98.56	100.00
1600	1.06	64.66	92.88	100.00

Table 18: The active subspace measures $\Sigma_m(\mathbf{u}^\dagger)$ for the Sobol' G function, comparing optimization of unrotated and randomly rotated inputs.

5. Conclusion

References

[1] J. Sacks, W. J. Welch, T. J. Mitchell, H. P. Wynn, Design and analysis of computer experiments, Statistical Science 4 (4) (1989) 409–423.
URL `http://www.jstor.org/stable/2245858`

[2] C. E. Rasmussen, C. K. I. Williams, Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning series), The MIT Press, 2005.

[3] R. G. Ghanem, P. D. Spanos, Spectral techniques for stochastic finite elements, Archives of Computational Methods in Engineering 4 (1) (1997) 63–100. `doi:10.1007/BF02818931`.
URL `https://doi.org/10.1007/BF02818931`

[4] D. Xiu, G. E. Karniadakis, The wiener–askey polynomial chaos for stochastic differential equations, SIAM Journal on Scientific Computing 24 (2) (2002) 619–644. `doi:10.1137/s1064827501387826`.

[5] D. Xiu, Numerical Methods for Stochastic Computations: A Spectral Method Approach, Princeton University Press, 2010.

[6] M. Chevreuil, R. Lebrun, A. Nouy, P. Rai, A least-squares method for sparse low rank approximation of multivariate functions, SIAM/ASA Journal on Uncertainty Quantification 3 (1) (2015) 897–921. arXiv:http://arxiv.org/abs/1305.0030v2, doi:10.1137/13091899X.

[7] K. Konakli, B. Sudret, Global sensitivity analysis using low-rank tensor approximations, Reliability Engineering & System Safety 156 (2016) 64–83. doi:10.1016/j.ress.2016.07.012.

[8] C. Cortes, V. Vapnik, Support-vector networks, Machine Learning 20 (3) (1995) 273–297. doi:10.1007/bf00994018.

[9] P. G. Constantine, E. Dow, Q. Wang, Active subspace methods in theory and practice: Applications to kriging surfaces, SIAM Journal on Scientific Computing 36 (4) (2014) A1500–A1524. arXiv:1304.2070, doi:10.1137/130916138.

[10] X. Liu, S. Guillas, Dimension Reduction for Gaussian Process Emulation: An Application to the Influence of Bathymetry on Tsunami Heights, SIAM-ASA Journal on Uncertainty Quantification 5 (1) (2017) 787–812. doi:10.1137/16M1090648.

[11] Y. Xia, H. Tong, W. K. Li, L.-X. Zhu, An adaptive estimation of dimension reduction space, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64 (3) (2002) 363–410. doi:10.1111/1467-9868.03411.

[12] I. M. Sobol, Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates, Mathematics and Computers in Simulation 55 (2001) 271–280.

[13] J. E. Oakley, A. O'Hagan, Probabilistic sensitivity analysis of complex models: a bayesian approach, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 66 (3) (2004) 751–769. doi:10.1111/j.1467-9868.2004.05304.x.

[14] R. Jin, W. Chen, A. Sudjianto, Analytical metamodel-based global sensitivity analysis and uncertainty propagation for robust design, SAE

Transactions Journal of Materials & Manufacturing (2004). doi:10.4271/2004-01-0429.

[15] A. Marrel, B. Iooss, B. Laurent, O. Roustant, Calculations of sobol indices for the gaussian process metamodel, Reliability Engineering & System Safety 94 (3) (2009) 742–751. doi:10.1016/j.ress.2008.07.008.

[16] B. Lamoureux, N. Mechbal, J. R. Massé, A combined sensitivity analysis and kriging surrogate modeling for early validation of health indicators, Reliability Engineering and System Safety 130 (2014) 12–26. doi:10.1016/j.ress.2014.03.007.
URL http://dx.doi.org/10.1016/j.ress.2014.03.007

[17] D. P. Wipf, S. Nagarajan, A new view of automatic relevance determination, in: Proceedings of the 20th International Conference on Neural Information Processing Systems, NIPS'07, Curran Associates Inc., 2007, pp. 1625–1632.
URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.143.8009{&}rep=rep1{&}type=pdf

[18] R. M. Neal, Bayesian Learning for Neural Networks, Springer New York, 1996.

[19] C. E. Rasmussen, Some useful gaussian and matrix equations (2016).
URL http://mlg.eng.cam.ac.uk/teaching/4f13/1617/gaussian%20and%20matrix%20equations.pdf

[20] SheffieldML, GPy: A gaussian process framework in python.
URL https://sheffieldml.github.io/GPy/

[21] T. Tao, Matrix identities as derivatives of determinant identities.
URL https://terrytao.wordpress.com/2013/01/13/matrix-identities-as-derivatives-of-determinant-identities/