Online Supplement for "Robust Parameter Design with Computer Experiments Using Orthonormal Polynomials"

Matthias Hwai Yong Tan
Department of Systems Engineering and Engineering Management, City University of Hong Kong

Appendix A: Proof of Lemma 1 and Proposition 1

Lemma 1: $\left\|\hat{f}^P - f\right\|_w^2 = \left(\hat{f}_0 - f_0\right)^2 + \sum_{k=1}^{d}\sum_{1\le i_1<\cdots<i_k\le d}\left\|\hat{f}_{(i_1,\ldots,i_k)} - f_{(i_1,\ldots,i_k)}\right\|_w^2$.

Proof:

$$\begin{aligned}
&\left\|\hat{f}^P - f\right\|_w^2 = \int_\chi \left[\sum_{\boldsymbol{\alpha}\in\mathbb{N}_0^d:\sum_{i=1}^d \alpha_i\le P}\beta_{\boldsymbol{\alpha}}\psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) - f(\boldsymbol{x})\right]^2 w(\boldsymbol{x})d\boldsymbol{x}\\
&= \int_\chi \left[\hat{f}_0 + \sum_{k=1}^{d}\sum_{1\le i_1<\cdots<i_k\le d}\hat{f}_{(i_1,\ldots,i_k)}(\boldsymbol{x}) - \left(f_0 + \sum_{k=1}^{d}\sum_{1\le i_1<\cdots<i_k\le d} f_{(i_1,\ldots,i_k)}(\boldsymbol{x})\right)\right]^2 w(\boldsymbol{x})d\boldsymbol{x}\\
&= \int_\chi \left[\left(\hat{f}_0 - f_0\right) + \sum_{k=1}^{d}\sum_{1\le i_1<\cdots<i_k\le d}\left(\hat{f}_{(i_1,\ldots,i_k)} - f_{(i_1,\ldots,i_k)}\right)(\boldsymbol{x})\right]^2 w(\boldsymbol{x})d\boldsymbol{x}\\
&= \int_\chi \left\{\left(\hat{f}_0 - f_0\right)^2 + \sum_{k=1}^{d}\sum_{1\le i_1<\cdots<i_k\le d}\left[\left(\hat{f}_{(i_1,\ldots,i_k)} - f_{(i_1,\ldots,i_k)}\right)(\boldsymbol{x})\right]^2\right\} w(\boldsymbol{x})d\boldsymbol{x}\\
&= \int_\chi \left(\hat{f}_0 - f_0\right)^2 w(\boldsymbol{x})d\boldsymbol{x} + \sum_{k=1}^{d}\sum_{1\le i_1<\cdots<i_k\le d}\int_\chi \left[\left(\hat{f}_{(i_1,\ldots,i_k)} - f_{(i_1,\ldots,i_k)}\right)(\boldsymbol{x})\right]^2 w(\boldsymbol{x})d\boldsymbol{x}\\
&= \left(\hat{f}_0 - f_0\right)^2 + \sum_{k=1}^{d}\sum_{1\le i_1<\cdots<i_k\le d}\left\|\hat{f}_{(i_1,\ldots,i_k)} - f_{(i_1,\ldots,i_k)}\right\|_w^2.
\end{aligned}$$

Note that we have used the fact that $\int_\chi \left(f_0 - \hat{f}_0\right)\left(f_{(i_1,\ldots,i_k)} - \hat{f}_{(i_1,\ldots,i_k)}\right)(\boldsymbol{x})w(\boldsymbol{x})d\boldsymbol{x} = 0$ and

$\int_\chi \left(f_{(i_1,\ldots,i_k)} - \hat{f}_{(i_1,\ldots,i_k)}\right)(\boldsymbol{x})\left(f_{(j_1,\ldots,j_l)} - \hat{f}_{(j_1,\ldots,j_l)}\right)(\boldsymbol{x})w(\boldsymbol{x})d\boldsymbol{x} = 0$ for $(i_1,\ldots,i_k) \ne (j_1,\ldots,j_l)$.

Proposition 1: If $\lim_{P\to\infty}\left\|\hat{f}^P - f\right\|_w^2 = 0$, then $\lim_{P\to\infty}\left(\hat{f}_0 - f_0\right)^2 = 0$ and $\lim_{P\to\infty}\left\|\hat{f}_{(i_1,\ldots,i_k)} - f_{(i_1,\ldots,i_k)}\right\|_w^2 = 0$ for all $(i_1,\ldots,i_k)$.

Proof:

Since $\left(\hat{f}_0 - f_0\right)^2 \ge 0$ and $\left\|\hat{f}_{(i_1,\ldots,i_k)} - f_{(i_1,\ldots,i_k)}\right\|_w^2 \ge 0$ for all P, it follows from Lemma 1 that

$\lim_{P\to\infty}\left(\hat{f}_0 - f_0\right)^2 = 0$ and $\lim_{P\to\infty}\left\|\hat{f}_{(i_1,\ldots,i_k)} - f_{(i_1,\ldots,i_k)}\right\|_w^2 = 0$.

Appendix B: Review of Gaussian Process Modeling

In GP modeling, the prior for the functional relationship f is given by

$$Y(\boldsymbol{x}) = \lambda + G(\boldsymbol{x}), \tag{1}$$

where $\boldsymbol{x} \in \chi$ and $G(\boldsymbol{x})$ is a zero mean stationary GP. Given any two points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, the covariance of $Y(\boldsymbol{x}_i)$ and $Y(\boldsymbol{x}_j)$ is given by $cov[Y(\boldsymbol{x}_i), Y(\boldsymbol{x}_j)] = cov[G(\boldsymbol{x}_i), G(\boldsymbol{x}_j)] = \sigma^2 R(\boldsymbol{x}_i, \boldsymbol{x}_j)$, where $R(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is the correlation function. The most commonly used correlation function is the Gaussian correlation function

$$R(\boldsymbol{x}_i, \boldsymbol{x}_j) = \prod_{k=1}^{d} \theta_k^{(x_{ik} - x_{jk})^2}, \tag{2}$$

where x_{ik} is the kth element of $\boldsymbol{x}_i$, and $\theta_k \in (0,1), k = 1, \ldots, d$.

In a computer experiment, the computer output is evaluated at n values of inputs given by the rows $\boldsymbol{x}^1, \ldots, \boldsymbol{x}^n$ of the design $\mathcal{D}$, where the choice of each $\boldsymbol{x}^i$ is restricted to χ. This yields a vector $\boldsymbol{Y}$ of observed outputs. The prior process is updated with experiment data, giving a posterior GP (Currin et al., 1991; Santner et al., 2003)

$$Y(\cdot)|(\boldsymbol{Y}, \lambda, \sigma^2, \boldsymbol{\theta}) \sim GP\big(\mu(\cdot|\lambda, \boldsymbol{\theta}), C(\cdot,\cdot|\sigma^2, \boldsymbol{\theta})\big), \tag{3}$$

with mean function $\mu(\cdot|\lambda, \boldsymbol{\theta})$ and covariance function $C(\cdot,\cdot|\sigma^2, \boldsymbol{\theta})$. The mean function is given by

$$\mu(\boldsymbol{x}|\lambda, \boldsymbol{\theta}) = \lambda + \boldsymbol{r}(\boldsymbol{x})^T \boldsymbol{R}^{-1}(\boldsymbol{Y} - \lambda \boldsymbol{1}), \tag{4}$$

where $\boldsymbol{r}(\boldsymbol{x}) = \big(R(\boldsymbol{x}, \boldsymbol{x}^1), \ldots, R(\boldsymbol{x}, \boldsymbol{x}^n)\big)^T$, and $\boldsymbol{R} = \Big(R\big(\boldsymbol{x}^i, \boldsymbol{x}^j\big)\Big)_{1 \le i \le n, 1 \le j \le n}$; the covariance function is given by

$$C\big(\boldsymbol{x}_i, \boldsymbol{x}_j \big| \sigma^2, \boldsymbol{\theta}\big) = \sigma^2 \big\{R\big(\boldsymbol{x}_i, \boldsymbol{x}_j\big) - \boldsymbol{r}(\boldsymbol{x}_i)^T \boldsymbol{R}^{-1} \boldsymbol{r}\big(\boldsymbol{x}_j\big)\big\}. \tag{5}$$

The mean and covariance functions given by (4)-(5) depend on the parameters λ, σ^2, and $\boldsymbol{\theta}$, which can be estimated from the data. One approach to estimating λ, σ^2, and $\boldsymbol{\theta}$ is the maximum likelihood method, which estimates the parameters by maximizing the likelihood $\mathcal{L}(\boldsymbol{\theta}, \lambda, \sigma^2) = [(2\pi\sigma^2)^n |\boldsymbol{R}|]^{-1/2} \exp[-(\boldsymbol{Y} - \lambda \boldsymbol{1})^T \boldsymbol{R}^{-1}(\boldsymbol{Y} - \lambda \boldsymbol{1})/(2\sigma^2)]$. It turns out that given $\boldsymbol{\theta}$, $\lambda = \hat{\lambda} = (\boldsymbol{1}^T \boldsymbol{R}^{-1} \boldsymbol{1})^{-1} \boldsymbol{1}^T \boldsymbol{R}^{-1} \boldsymbol{Y}$, and $\sigma^2 = \hat{\sigma}^2 = \big(\boldsymbol{Y} - \hat{\lambda}\boldsymbol{1}\big)^T \boldsymbol{R}^{-1}(\boldsymbol{Y} - \hat{\lambda}\boldsymbol{1})/n$ maximize

$\mathcal{L}(\boldsymbol{\theta}, \lambda, \sigma^2)$. Thus, the maximum likelihood estimate $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is obtained by maximizing

$\mathcal{L}\left(\boldsymbol{\theta}, \hat{\lambda}, \hat{\sigma}^2\right)$, which is equivalent to minimizing

$$\mathcal{K}(\boldsymbol{\theta}) = n\log(\hat{\sigma}^2) + \log(|\boldsymbol{R}|). \quad (6)$$

It is common to perform statistical inference on f using $\widehat{\boldsymbol{\theta}}$ in place of $\boldsymbol{\theta}$, $\hat{\lambda}$ in place of λ, and $\tilde{\sigma}^2 = \frac{n}{n-1}\hat{\sigma}^2$ in place of σ^2, i.e., using the GP

$$Y(\cdot)|\left(\boldsymbol{Y}, \hat{\lambda}, \tilde{\sigma}^2, \widehat{\boldsymbol{\theta}}\right) \sim GP\left(\mu\left(\cdot|\hat{\lambda}, \widehat{\boldsymbol{\theta}}\right), C\left(\cdot,\cdot|\tilde{\sigma}^2, \widehat{\boldsymbol{\theta}}\right)\right). \quad (7)$$

However, by replacing $(\lambda, \sigma^2, \boldsymbol{\theta})$ by estimates, we ignore the uncertainty in these parameters and hence may underestimate uncertainty in inferences about f.

References

1. Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D. (1991). "Bayesian Prediction of Deterministic Functions, with Applications to the Design and Analysis of Computer Experiments," *Journal of the American Statistical Association*, 86(416), 953-963.
2. Santner, T.J., Williams, B.J., and Notz, W.I. (2003). *The Design and Analysis of Computer Experiments*. New York: Springer-Verlag.

Appendix C: Comparison with Other Computational Techniques

The paper by Chen et al. (2006) gives analytical formulas for computing some of the RPD indices from the posterior mean $\mu(\boldsymbol{x}|\lambda, \boldsymbol{\theta})$ of the GP model for the case where the product Gaussian correlation function is used and $\boldsymbol{x}$ has uniformly and independently distributed components. In more general cases such as when $\boldsymbol{x}$ have independent Beta distributions (Section 6), or when the GP has product Matern correlation function, the more general formulas given by Chen et al. (2006) that involve evaluating a large number of univariate integrals need to be used. In Equation (35) of Chen et al. (2006), there are a total of nd $C_{1,il}$'s and a total of $\binom{n}{2} d$ $C_{2,ijl}$'s if the posterior mean of a GP is used as surrogate. Thus, the number of univariate integrals that need to be computed is $nd + \binom{n}{2} d$. The integrals can only be evaluated analytically when certain products of the correlation function and density for each input dimension have known

anti-derivatives. When numerical integration is employed, computation is more difficult and numerical errors are unavoidable. When the covariance function of the GP is not of product form (examples are given in Chapter 4 of Rasmussen and Williams (2006)), the formulas by Chen et al. (2006) are not applicable. In contrast, the proposed method works for almost any given densities $w_1, \dots, w_d$ and covariance function of the GP model. Moreover, it does not require the computation of any integral when the orthonormal polynomials $\{\psi_j^i : j \in \mathbb{N}_0\}$ are known (there are many classes of w_i with known ψ_j^i's (see Xiu and Karniadakis (2002)).

The method by Chen et al. (2006), if applied directly to GP models, yields estimators of RPD indices that are different from ours. Suppose we intend to estimate $\mathcal{H} = H_1(f)/H_2(f)$, where f is the computer model and $\mathcal{H}$ can represent any of the RPD indices $\mathcal{N}_{(t_1,\dots,t_l)}$, $\mathcal{A}_{(s)}$, and $\mathcal{B}_{(s)}$. Then, since the posterior GP Y is the posterior of the computer model, a Bayesian approach involves replacing f with Y and estimating $\mathcal{H}$ with its posterior mean

$$\widehat{\mathcal{H}} = E_Y\{H_1(Y)/H_2(Y)\}. \quad (1)$$

Note that the proposed method attempts to recover $\widehat{\mathcal{H}}$. On the other hand, the estimator employed by Chen et al. (2006) for $\mathcal{H}$ is

$$\widetilde{\mathcal{H}} = H_1\big(E_Y(Y)\big)/H_2\big(E_Y(Y)\big), \quad (2)$$

which is obtained by replacing $f(\boldsymbol{x})$ with $E_Y\big(Y(\boldsymbol{x})\big) = \mu\big(\boldsymbol{x}|\hat{\lambda}, \widehat{\boldsymbol{\theta}}\big)$. The estimators $\widehat{\mathcal{H}}$ and $\widetilde{\mathcal{H}}$ can yield different results.

A method that can produce valid point and interval estimates of $\mathcal{N}_{(t_1,\dots,t_l)}$, $\mathcal{A}_{(s)}$, $\mathcal{B}_{(s)}$, and $\mathcal{V}_{(t_1,\dots,t_l)}/\mathcal{V}$ is to use Oakley and O'Hagan's (2004) method for decomposing $Y(\cdot)$ into its functional ANOVA components and then simulating random realizations of the components. However, a formidable number of integrals need to be computed to obtain the mean and

covariance functions of the components. In addition, obtaining $\mathcal{N}_{(t_1,\ldots,t_l)}$, $\mathcal{A}_{(s)}$, $\mathcal{B}_{(s)}$, and $\mathcal{V}_{(t_1,\ldots,t_l)}/\mathcal{V}$ from the functional ANOVA components requires further integrations with respect to the inputs. It is not clear how the cumulative integration error can be controlled. Estimates of $\mathcal{M}$ and $\mathcal{V}$ at a given control factor setting can be obtained by simulating the GP at noise factor settings that serve as integration nodes. However, independent simulations at distinct control factor settings produce noisy and nonsmooth estimates of $\mathcal{M}$ and $\mathcal{V}$, unlike averaging $\widehat{\mathcal{M}}$ and $\hat{\mathcal{V}}$ over a sample of $\boldsymbol{\beta}$. Although stochastic approximation algorithms can be used in this case, they are less efficient than deterministic optimization algorithms, which can be employed to optimize the average of $\widehat{\mathcal{M}}$ and $\hat{\mathcal{V}}$ over a sample of $\boldsymbol{\beta}$ (see Nemirovski et al. (2009) and Wright and Nocedal (1999)). Smooth estimates of $\mathcal{M}$ and $\mathcal{V}$ can be obtained by simulating sample paths of the GP but this involves simulating from high dimensional distributions (Apley et al., 2006).

References

1. Apley, D.W., Liu, J., and Chen, W. (2006), "Understanding the Effects of Model Uncertainty in Robust Design with Computer Experiments," *Journal of Mechanical Design*, 128, 945-958.
2. Chen, W., Jin, R., and Sudjianto, A. (2006). "Analytical Global Sensitivity Analysis and Uncertainty Propagation for Robust Design," *Journal of Quality Technology*, 38(4), 333-348.
3. Nemirovski, A., Juditsky, A., Lan, G., and Shapiro, A. (2009). "Robust stochastic approximation approach to stochastic programming," *SIAM Journal on Optimization*, 19(4), 1574-1609.
4. Oakley, J.E. and O'Hagan, A. (2004). "Probabilistic Sensitivity Analysis of Complex Models: A Bayesian Approach," *Journal of the Royal Statistical Society (Series B)*, 66(3), 751-769.
5. Rasmussen, C.E. and Williams, C.K.I. (2006). *Gaussian processes for machine learning*. Cambridge: MIT Press.
6. Wright, S. J. and Nocedal, J. (1999). *Numerical optimization*. New York: Springer.
7. Xiu, D. and Karniadakis, G.E. (2002). "The Wiener-Askey Polynomial Chaos for Stochastic Differential Equations," *SIAM Journal on Scientific Computing*, 24(2), 619-644.

Appendix D: Convergence of WLS Method

The fact that the orthonormal polynomial model can approximate the GP model to any desired accuracy by using sufficiently large *P* and *N* shall be proven in this section. As

discussed, under some nonrestrictive conditions, the set of polynomials $\{\psi_{\boldsymbol{\alpha}} = \prod_{i=1}^{d} \psi_{\alpha_i}^{i} : \boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}_0^d\}$ form an orthonormal basis for the Hilbert space $L^2(\chi, w) = \left\{\zeta : \chi \to \mathbb{R} : \int_\chi \zeta^2(\boldsymbol{x})\, w(\boldsymbol{x}) d\boldsymbol{x} < \infty\right\}$. Recall that the posterior GP that we attempt to approximate with the orthonormal polynomial model is given by $Y(\cdot) \sim GP\big(\mu(\cdot|\lambda, \boldsymbol{\theta}), C(\cdot,\cdot|\sigma^2, \boldsymbol{\theta})\big)$, which for simplicity, we shall denote by $Y(\cdot) \sim GP\big(\mu(\cdot), C(\cdot,\cdot)\big)$ in this section. In the following, we shall assume that χ is compact and the sequence $\boldsymbol{x}_1, \dots, \boldsymbol{x}_N$ is such that $\lim_{N\to\infty} \sum_{i=1}^{N} \zeta(\boldsymbol{x}_i)/N = \int_\chi \zeta(\boldsymbol{x}) d\boldsymbol{x}$ for every continuous ζ. This assumption holds if $\chi = [-1,1]^d$ and $\boldsymbol{x}_1, \dots, \boldsymbol{x}_N$ is a quasi-Monte Carlo sequence (e.g., Sobol sequence) on χ (Theorem 2.13 in Niederreiter (1992)). We assume that w is continuous on χ and $w(\boldsymbol{x}) > 0$ for all $\boldsymbol{x} \in \chi$ except on a set $\chi' \subset \chi$ of Jordan measure zero. We also assume that $GP\big(\mu(\cdot), C(\cdot,\cdot)\big)$ is almost surely continuous on χ as can be guaranteed by choosing the prior correlation function R so that the prior for $Y(\cdot)$ is almost surely continuous. This implies that both posterior mean function $\mu : \chi \to \mathbb{R}$ and posterior covariance function $C : \chi \times \chi \to \mathbb{R}$ are continuous. Define the expectation operator $E_{\boldsymbol{x},Y}$ by $E_{\boldsymbol{x},Y}\big[\zeta\big(\boldsymbol{x}, Y(\boldsymbol{u}_1), \dots, Y(\boldsymbol{u}_k)\big)\big] = \int_\chi E_Y\big[\zeta\big(\boldsymbol{x}, Y(\boldsymbol{u}_1), \dots, Y(\boldsymbol{u}_k)\big)\big] w(\boldsymbol{x})\, d\boldsymbol{x}$, where E_Y denotes expectation with respect to the posterior distribution of $Y(\cdot)$. Then,

$$E_{\boldsymbol{x},Y}[Y^2(\boldsymbol{x})] = \int_\chi E_Y[Y^2(\boldsymbol{x})]\, w(\boldsymbol{x}) d\boldsymbol{x} = \int_\chi [\mu^2(\boldsymbol{x}) + C(\boldsymbol{x}, \boldsymbol{x})] w(\boldsymbol{x}) d\boldsymbol{x} < \infty. \quad (1)$$

By Fubini's theorem,

$$\infty > \int_\chi E_Y[Y^2(\boldsymbol{x})]\, w(\boldsymbol{x}) d\boldsymbol{x} = E_Y\left[\int_\chi Y^2(\boldsymbol{x})\, w(\boldsymbol{x}) d\boldsymbol{x}\right]. \quad (2)$$

Hence, $\int_\chi Y^2(\boldsymbol{x})\, w(\boldsymbol{x}) d\boldsymbol{x} < \infty$ almost surely. This implies that $Y(\cdot) \in L^2(\chi, w)$ almost surely.

Let $\boldsymbol{Y}^{N,1} = \big(Y(\boldsymbol{x}_1), \dots, Y(\boldsymbol{x}_N)\big)^T$, $\boldsymbol{\phi}_1 = \big(\psi_j(\boldsymbol{x}_i)\big)_{1\le i\le N, 1\le j\le m}$, and $\boldsymbol{W}_1 = \mathrm{diag}\{w(\boldsymbol{x}_1), \dots, w(\boldsymbol{x}_N)\}$. Consider the weighted least squares method, which gives

$$\boldsymbol{\beta}_{LS} = (\boldsymbol{\phi}_1^T \boldsymbol{W}_1 \boldsymbol{\phi}_1)^{-1} \boldsymbol{\phi}_1^T \boldsymbol{W}_1 \boldsymbol{Y}^{N,1}. \quad (3)$$

If $w(\boldsymbol{x}_i) = 0$, then we simply delete $\boldsymbol{x}_i$ and replace it with $\boldsymbol{x}_{i+1}$. Note that since $w(\boldsymbol{x}) = 0$ on a set χ' of Jordan measure zero, it follows that the fraction of $\boldsymbol{x}_i$'s with $w(\boldsymbol{x}_i) = 0$ tends to zero as $N \to \infty$ (see page 18 of Niederreiter (1992)). Define $\boldsymbol{\beta}_P = \int_\chi \boldsymbol{\psi}(\boldsymbol{x})^T Y(\boldsymbol{x}) w(\boldsymbol{x}) d\boldsymbol{x}$, where $\boldsymbol{\psi}(\boldsymbol{x}) = \big(\psi_1(\boldsymbol{x}), \dots, \psi_m(\boldsymbol{x})\big)$. This is the vector of coefficients obtained by projecting $Y(\cdot)$ onto the space spanned by the bases $\psi_1, \dots, \psi_m$. We shall show that $\lim_{N\to\infty} \boldsymbol{\beta}_{LS} = \boldsymbol{\beta}_P$. If we let $N \to \infty$, the element in the ith row and jth column of $\boldsymbol{\phi}_1^T \boldsymbol{W}_1 \boldsymbol{\phi}_1 / N$ converges to

$$\lim_{N\to\infty} (\boldsymbol{\phi}_1^T \boldsymbol{W}_1 \boldsymbol{\phi}_1)_{(i,j)} / N = \lim_{N\to\infty} \sum_{k=1}^N \psi_i(\boldsymbol{x}_k) \psi_j(\boldsymbol{x}_k) w(\boldsymbol{x}_k) / N = \int_\chi \psi_i(\boldsymbol{x}) \psi_j(\boldsymbol{x}) w(\boldsymbol{x}) d\boldsymbol{x}. \quad (4)$$

Similarly, if we let $N \to \infty$, the ith element of $\boldsymbol{\phi}_1^T \boldsymbol{W}_1 \boldsymbol{Y}^{N,1} / N$ converges to

$$\lim_{N\to\infty} (\boldsymbol{\phi}_1^T \boldsymbol{W}_1 \boldsymbol{Y}^{N,1})_{(i)} / N = \lim_{N\to\infty} \sum_{k=1}^N \psi_i(\boldsymbol{x}_k) Y(\boldsymbol{x}_k) w(\boldsymbol{x}_k) / N = \int_\chi \psi_i(\boldsymbol{x}) Y(\boldsymbol{x}) w(\boldsymbol{x}) d\boldsymbol{x}. \quad (5)$$

The integral in (5) exists because $Y(\cdot) \in L^2(\chi, w)$ almost surely. Thus, we have

$$\lim_{N\to\infty} (\boldsymbol{\phi}_1^T \boldsymbol{W}_1 \boldsymbol{\phi}_1)^{-1} = \boldsymbol{I} \text{ and } \lim_{N\to\infty} \boldsymbol{\beta}_{LS} = \int_\chi \boldsymbol{\psi}(\boldsymbol{x})^T Y(\boldsymbol{x}) w(\boldsymbol{x}) d\boldsymbol{x} = \boldsymbol{\beta}_P. \quad (6)$$

The average error $E_{\boldsymbol{x},Y}\{[Y(\boldsymbol{x}) - \boldsymbol{\psi}(\boldsymbol{x})\boldsymbol{\beta}_{LS}]^2\}$ in approximating $Y(\boldsymbol{x})$ with the weighted least squares estimator $\boldsymbol{\psi}(\boldsymbol{x})\boldsymbol{\beta}_{LS}$ is bounded above by a term due to projecting Y onto the space spanned by the bases $\psi_1, \dots, \psi_m$ and a term due to approximating the projection with the weighted least squares estimator:

$$E_{\boldsymbol{x},Y}\{[Y(\boldsymbol{x}) - \boldsymbol{\psi}(\boldsymbol{x})\boldsymbol{\beta}_{LS}]^2\} = E_{\boldsymbol{x},Y}\{[Y(\boldsymbol{x}) - \boldsymbol{\psi}(\boldsymbol{x})\boldsymbol{\beta}_P + \boldsymbol{\psi}(\boldsymbol{x})\boldsymbol{\beta}_P - \boldsymbol{\psi}(\boldsymbol{x})\boldsymbol{\beta}_{LS}]^2\}$$

$$\leq 2E_{\boldsymbol{x},Y}\{[Y(\boldsymbol{x}) - \boldsymbol{\psi}(\boldsymbol{x})\boldsymbol{\beta}_P]^2\} + 2E_{\boldsymbol{x},Y}\{[\boldsymbol{\psi}(\boldsymbol{x})(\boldsymbol{\beta}_P - \boldsymbol{\beta}_{LS})]^2\}. \quad (7)$$

The term $E_{\boldsymbol{x},Y}\{[Y(\boldsymbol{x}) - \boldsymbol{\psi}(\boldsymbol{x})\boldsymbol{\beta}_P]^2\}$ is due to projecting $Y(\cdot)$ onto the space spanned by the bases $\psi_1, \dots, \psi_m$. It depends on P but not on N. Because $Y(\cdot) \in L^2(\chi, w)$ almost surely, $\lim_{P\to\infty}(E_{\boldsymbol{x}}\{[Y(\boldsymbol{x}) - \boldsymbol{\psi}(\boldsymbol{x})\boldsymbol{\beta}_P]^2\}) = 0$ for almost all $Y(\cdot)$. We also have

$$E_{\boldsymbol{x}}\{[Y(\boldsymbol{x}) - \boldsymbol{\psi}(\boldsymbol{x})\boldsymbol{\beta}_P]^2\} = \sum_{i=m+1}^\infty \beta_i^2 \leq \sum_{i=1}^\infty \beta_i^2 = E_{\boldsymbol{x}}[Y^2(\boldsymbol{x})], \quad (8)$$

where $\beta_1, \beta_2, \ldots$ are the coefficients of the expansion of $Y(\cdot)$ in terms of the polynomial bases in $\{\psi_{\boldsymbol{\alpha}} = \prod_{i=1}^{d} \psi_{\alpha_i}^{i} : \boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}_0^d\}$, i.e., $\beta_i = \int_{\chi} \psi_i(\boldsymbol{x}) Y(\boldsymbol{x}) w(\boldsymbol{x}) d\boldsymbol{x}$. Since $E_{\boldsymbol{x}}[Y^2(\boldsymbol{x})]$ does not depend on P and $E_Y\{E_{\boldsymbol{x}}[Y^2(\boldsymbol{x})]\} < \infty$, it follows by the dominated convergence theorem (page 133 of Resnick (1999)) that

$$\lim_{P\to\infty} E_{\boldsymbol{x},Y}\{[Y(\boldsymbol{x}) - \boldsymbol{\psi}(\boldsymbol{x})\boldsymbol{\beta}_P]^2\} = \lim_{P\to\infty} E_Y(E_{\boldsymbol{x}}\{[Y(\boldsymbol{x}) - \boldsymbol{\psi}(\boldsymbol{x})\boldsymbol{\beta}_P]^2\})$$

$$= E_Y(\lim_{P\to\infty} E_{\boldsymbol{x}}\{[Y(\boldsymbol{x}) - \boldsymbol{\psi}(\boldsymbol{x})\boldsymbol{\beta}_P]^2\}) = E_Y(0) = 0. \quad (9)$$

The term $E_{\boldsymbol{x},Y}\{[\boldsymbol{\psi}(\boldsymbol{x})(\boldsymbol{\beta}_P - \boldsymbol{\beta}_{LS})]^2\}$ is due to approximating the projection of $Y(\cdot)$ with the weighted least squares estimator. We shall show that for fixed P, $\lim_{N\to\infty} E_{\boldsymbol{x},Y}\{[\boldsymbol{\psi}(\boldsymbol{x})(\boldsymbol{\beta}_P - \boldsymbol{\beta}_{LS})]^2\} = E_{\boldsymbol{x},Y}\{\lim_{N\to\infty}[\boldsymbol{\psi}(\boldsymbol{x})(\boldsymbol{\beta}_P - \boldsymbol{\beta}_{LS})]^2\} = 0$. Note that

$$[\boldsymbol{\psi}(\boldsymbol{x})(\boldsymbol{\beta}_{LS} - \boldsymbol{\beta}_P)]^2 \le \|\boldsymbol{\psi}(\boldsymbol{x})\|_2^2 \|\boldsymbol{\beta}_{LS} - \boldsymbol{\beta}_P\|_2^2 \le \|\boldsymbol{\psi}(\boldsymbol{x})\|_2^2 (2\|\boldsymbol{\beta}_{LS}\|_2^2 + 2\|\boldsymbol{\beta}_P\|_2^2). \quad (10)$$

Because the ψ_i's are orthonormal,

$$E_{\boldsymbol{x}}(\|\boldsymbol{\psi}(\boldsymbol{x})\|_2^2) = m < \infty. \quad (11)$$

In addition, $\|\boldsymbol{\beta}_P\|_2^2 \le \sum_{i=1}^{\infty} \beta_i^2 = E_{\boldsymbol{x}}[Y^2(\boldsymbol{x})]$, which implies that

$$E_Y(\|\boldsymbol{\beta}_P\|_2^2) \le E_{\boldsymbol{x},Y}[Y^2(\boldsymbol{x})] < \infty. \quad (12)$$

We now proceed to show that $E_Y(\|\boldsymbol{\beta}_{LS}\|_2^2) < \infty$. First, we argue that $\sup_{\boldsymbol{x}\in\chi}|Y(\boldsymbol{x}) - \mu(\boldsymbol{x})|$ has moments of all order. Due to the continuity of μ and C, and the compactness of χ, we must have $m_\chi = \sup_{\boldsymbol{x}\in\chi}|\mu(\boldsymbol{x})| < \infty$ and $\sigma_\chi^2 = \sup_{\boldsymbol{x}\in\chi} C(\boldsymbol{x}, \boldsymbol{x}) < \infty$. Moreover, the compactness of χ and the almost sure continuity of $Y(\cdot)$ implies that $\sup_{\boldsymbol{x}\in\chi}|Y(\boldsymbol{x})| < \infty$ almost surely. Then, by Theorem 2.1 of Adler (1990),

$$P\left(\left|\sup_{\boldsymbol{x}\in\chi}[Y(\boldsymbol{x}) - \mu(\boldsymbol{x})]\right| > K\right) \le 2\exp\left[-0.5(K - m_S)^2/\sigma_\chi^2\right], \quad (13)$$

for all $K > 0$, where $m_S = E\{\sup_{\boldsymbol{x}\in\chi}[Y(\boldsymbol{x}) - \mu(\boldsymbol{x})]\}$ is finite. Note that $-Y(\cdot) + \mu(\cdot)$ and $Y(\cdot) - \mu(\cdot)$ have the same distribution. Thus, $E\{\sup_{\boldsymbol{x}\in\chi}[-Y(\boldsymbol{x}) + \mu(\boldsymbol{x})]\} = m_S$. Since

$\sup_{\boldsymbol{x}\in\chi}|Y(\boldsymbol{x})-\mu(\boldsymbol{x})| = \max\{\sup_{\boldsymbol{x}\in\chi}[Y(\boldsymbol{x})-\mu(\boldsymbol{x})], \sup_{\boldsymbol{x}\in\chi}[-Y(\boldsymbol{x})+\mu(\boldsymbol{x})]\}$, we have

$$P\left(\left|\sup_{\boldsymbol{x}\in\chi}|Y(\boldsymbol{x})-\mu(\boldsymbol{x})|\right| > K\right)$$

$$\leq P\left(\left|\sup_{\boldsymbol{x}\in\chi}[Y(\boldsymbol{x})-\mu(\boldsymbol{x})]\right| > K\right) + P\left(\left|\sup_{\boldsymbol{x}\in\chi}[-Y(\boldsymbol{x})+\mu(\boldsymbol{x})]\right| > K\right)$$

$$\leq 4\exp\left[-\frac{0.5(K-m_S)^2}{\sigma_\chi^2}\right]. \tag{14}$$

This implies that $\sup_{\boldsymbol{x}\in\chi}|Y(\boldsymbol{x})-\mu(\boldsymbol{x})|$ has moments of all orders. In particular, we have

$$E_Y\left[\sup_{\boldsymbol{x}\in\chi}Y^2(\boldsymbol{x})\right] \leq E_Y\left[\left(\sup_{\boldsymbol{x}\in\chi}|Y(\boldsymbol{x})-\mu(\boldsymbol{x})|+m_\chi\right)^2\right] < \infty. \tag{15}$$

Since $\lim_{N\to\infty}\boldsymbol{\phi}_1^T\boldsymbol{W}_1\boldsymbol{\phi}_1/N = \boldsymbol{I}$, we must have $\|(\boldsymbol{\phi}_1^T\boldsymbol{W}_1\boldsymbol{\phi}_1/N)^{-1}\|_2^2 \leq a^{-1}$ for some $0 < a < 1$ and all N large enough. Since the largest eigenvalues of $\boldsymbol{\phi}_1^T\boldsymbol{W}_1\boldsymbol{\phi}_1/N$ and $\boldsymbol{W}_1^{1/2}\boldsymbol{\phi}_1\boldsymbol{\phi}_1^T\boldsymbol{W}_1^{1/2}/N$ are identical, we must have $\left\|\boldsymbol{\phi}_1^T\boldsymbol{W}_1^{1/2}/N^{1/2}\right\|_2^2 \leq b$ for some $b > 1$ and all N large enough. Let $\omega = \sup_{\boldsymbol{x}\in\chi}w(\boldsymbol{x}) < \infty$. It follows that

$$\|\boldsymbol{\beta}_{LS}\|_2^2 = \|(\boldsymbol{\phi}_1^T\boldsymbol{W}_1\boldsymbol{\phi}_1/N)^{-1}(\boldsymbol{\phi}_1^T\boldsymbol{W}_1\boldsymbol{Y}^{N,1}/N)\|_2^2$$

$$\leq \|(\boldsymbol{\phi}_1^T\boldsymbol{W}_1\boldsymbol{\phi}_1/N)^{-1}\|_2^2\left\|\boldsymbol{\phi}_1^T\boldsymbol{W}_1^{1/2}/N^{1/2}\right\|_2^2\left\|\boldsymbol{W}_1^{1/2}\boldsymbol{Y}^{N,1}/N^{1/2}\right\|_2^2$$

$$\leq a^{-1}b\sup_{\boldsymbol{x}\in\chi}w(\boldsymbol{x})Y^2(\boldsymbol{x})$$

$$\leq a^{-1}b\omega\left(\sup_{\boldsymbol{x}\in\chi}|Y(\boldsymbol{x})-\mu(\boldsymbol{x})|+m_\chi\right)^2. \tag{16}$$

By (10) and (16), we obtain

$$[\boldsymbol{\psi}(\boldsymbol{x})(\boldsymbol{\beta}_{LS}-\boldsymbol{\beta}_P)]^2 \leq \|\boldsymbol{\psi}(\boldsymbol{x})\|_2^2\left(2a^{-1}b\omega\left(\sup_{\boldsymbol{x}\in\chi}|Y(\boldsymbol{x})-\mu(\boldsymbol{x})|+m_\chi\right)^2 + 2\|\boldsymbol{\beta}_P\|_2^2\right) \tag{17}$$

We want to show that the expectation of the rightmost term in (17) is finite. By (11), (12), and (15), we have

$$E_{\boldsymbol{x},Y}\left[\|\boldsymbol{\psi}(\boldsymbol{x})\|_2^2\left(2a^{-1}b\omega\left(\sup_{\boldsymbol{x}\in\chi}|Y(\boldsymbol{x})-\mu(\boldsymbol{x})|+m_\chi\right)^2 + 2\|\boldsymbol{\beta}_P\|_2^2\right)\right]$$

$$= 2E_{\boldsymbol{x}}(\|\boldsymbol{\psi}(\boldsymbol{x})\|_2^2)\left\{E_Y\left[a^{-1}b\omega\left(\sup_{\boldsymbol{x}\in\chi}|Y(\boldsymbol{x})-\mu(\boldsymbol{x})|+m_\chi\right)^2\right] + E_Y(\|\boldsymbol{\beta}_P\|_2^2)\right\} < \infty. \tag{18}$$

Since $[\boldsymbol{\psi}(\boldsymbol{x})(\boldsymbol{\beta}_{LS}-\boldsymbol{\beta}_P)]^2 \le \|\boldsymbol{\psi}(\boldsymbol{x})\|_2^2\left(2a^{-1}b\omega\left(\sup_{\boldsymbol{x}\in\chi}|Y(\boldsymbol{x})-\mu(\boldsymbol{x})|+m_\chi\right)^2+2\|\boldsymbol{\beta}_P\|_2^2\right)$ for all N large enough and the right-hand-side does not depend on N, we can apply the dominated convergence theorem to obtain

$$\lim_{N\to\infty}E_{\boldsymbol{x},Y}\{[\boldsymbol{\psi}(\boldsymbol{x})(\boldsymbol{\beta}_P-\boldsymbol{\beta}_{LS})]^2\}=E_{\boldsymbol{x},Y}\{\lim_{N\to\infty}[\boldsymbol{\psi}(\boldsymbol{x})(\boldsymbol{\beta}_P-\boldsymbol{\beta}_{LS})]^2\}=0. \quad (19)$$

The results stated in (7), (9), and (19) implies that

$$\lim_{P\to\infty}\lim_{N\to\infty}E_{\boldsymbol{x},Y}\{[Y(\boldsymbol{x})-\boldsymbol{\psi}(\boldsymbol{x})\boldsymbol{\beta}_{LS}]^2\}=0. \quad (20)$$

Thus, to achieve $E_{\boldsymbol{x},Y}\{[Y(\boldsymbol{x})-\boldsymbol{\psi}(\boldsymbol{x})\boldsymbol{\beta}_{LS}]^2\}<\varepsilon$, we can first choose P large enough so that $2E_{\boldsymbol{x},Y}\{[Y(\boldsymbol{x})-\boldsymbol{\psi}(\boldsymbol{x})\boldsymbol{\beta}_P]^2\}<\varepsilon/2$ and then choose N large enough so that $2E_{\boldsymbol{x},Y}\{[\boldsymbol{\psi}(\boldsymbol{x})(\boldsymbol{\beta}_P-\boldsymbol{\beta}_{LS})]^2\}<\varepsilon/2$. A practical implementation of this procedure is to increment N for fixed P until changes in $100E[Q(LS,P,N)]/ETSS$ (see Section 5) are small, and then increase P before incrementing N again. The procedure can be terminated if $100E[Q(LS,P,N)]/ETSS$ is sufficiently small. In the proposed procedure for the WLS method, we simply set $N=2m$ and do not increase it unless $\boldsymbol{\phi}_1$ is not of full column rank. This is because approximation accuracy is frequently found to be much more dependent on P than on N when $N\ge 2m$.

Cohen et al. (2013) and Migliorati et al. (2014) have studied the approximation of a function f via least squares estimation of the model coefficients. However, their results cannot be applied directly to the problem in this paper. Both papers study regression with data from fixed f and derive error bounds by taking expectation with respect to the design, which is assumed to be a random sample from w. Cohen et al. (2013) assumes that f is bounded and the least squares predictions are truncated whereas Migliorati et al. (2014) mainly study one-dimensional polynomial approximations. Interestingly, the theoretical prediction of Cohen et al. (2013), Chkifa et al. (2013), and Migliorati et al. (2014) that $N/\log(N)$ should scale with m^2 when w is uniform is contradicted by numerical experiments reported in Migliorati et al. (2013)

and Chkifa et al. (2013), which demonstrate that faster convergence is achieved when N scales linearly with m. These numerical experiments lend support to the choice of $N = 2m$ in the WLS and IWLS algorithms given in Section 5.

References

1. Adler, R. J. (1990). *An introduction to continuity, extrema, and related topics for general Gaussian processes*. Hayward: IMS.
2. Chkifa, A., Cohen, A., Migliorati, G., Nobile, F., and Tempone, R. (2013). "Discrete least squares polynomial approximation with random evaluations; application to parametric and stochastic elliptic PDEs," *EPFL-MATHICSE report*, 35-2013.
3. Cohen, A., Davenport, M.A., and Leviatan, D. (2013). "On the Stability and Accuracy of Least Squares Approximations," *Foundations of Computational Mathematics*, 13, 819-834.
4. Migliorati, G., Nobile, F., von Schwerin, E., and Tempone, R. (2014). "Analysis of Discrete L^2 Projection on Polynomial Spaces with Random Evaluations," *Foundations of Computational Mathematics*, ahead of print.
5. Migliorati, G., Nobile, F., von Schwerin, E., and Tempone, R. (2013). "Approximation of quantities of Interest in stochastic PDEs by the random discrete L^2 projection on polynomial spaces," SIAM Journal of Scientific Computing, 35(3), A1440-A1460.
6. Niederreiter, H. (1992). *Random number generation and quasi-Monte Carlo methods*. Philadelphia: SIAM.
7. Resnick, S. I. (1999). *A probability path*. Boston: Birkhäuser.

Appendix E: Estimation of Response Mean and Variance

Consider the toy problem with $f(\boldsymbol{x}) = \cos[3(x_1 + 1) + 1.5(x_2 + 1)], \boldsymbol{x} \in [-1,1]^2$, x_1 is a control factor, x_2 is a noise factor and $\boldsymbol{x}$ is uniformly distributed on $[-1,1]^2$. We use the maximin Latin hypercube design shown in Figure 1 to fit a GP model. The maximum likelihood estimate of the correlation parameters $\boldsymbol{\theta}$ is $\widehat{\boldsymbol{\theta}} = (0.2566, 0.6628)$ and the posterior GP has excellent prediction accuracy and prediction credible interval coverage.

The WLS and IWLS methods are employed to construct orthonormal polynomial models, where the $\psi_{\boldsymbol{\alpha}}$'s are tensor products of orthonormal Legendre polynomials because w is the uniform density on $[-1,1]^2$. The WLS method gives a model of degree 8, with $N = 90$ and $100E[Q(LS, P, N)]/ETSS = 0.0461$; the IWLS method gives a model of degree 8, with $N = 90$ and $100E[Q(ILS, P, N)]/ETSS = 0.0082$. A total of 10,000 values of $\boldsymbol{\beta}_{LS}$ and $\boldsymbol{\beta}_{ILS}$ are

simulated. Using these values, we estimate the mean function $\mathcal{M}$ by taking the average of the

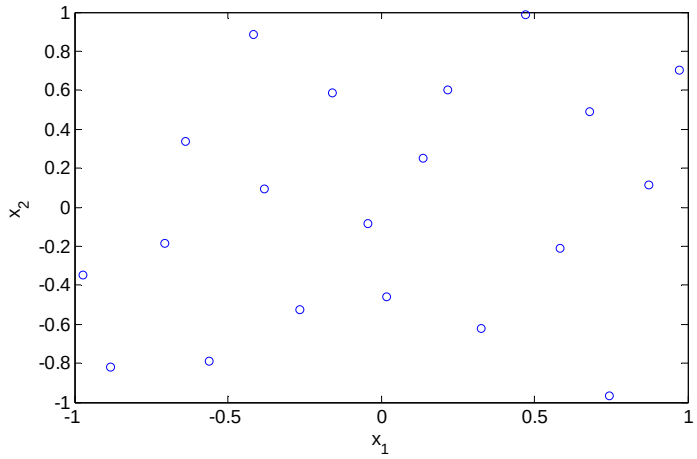

Figure 1: Maximin Latin Hypercube Design for Example

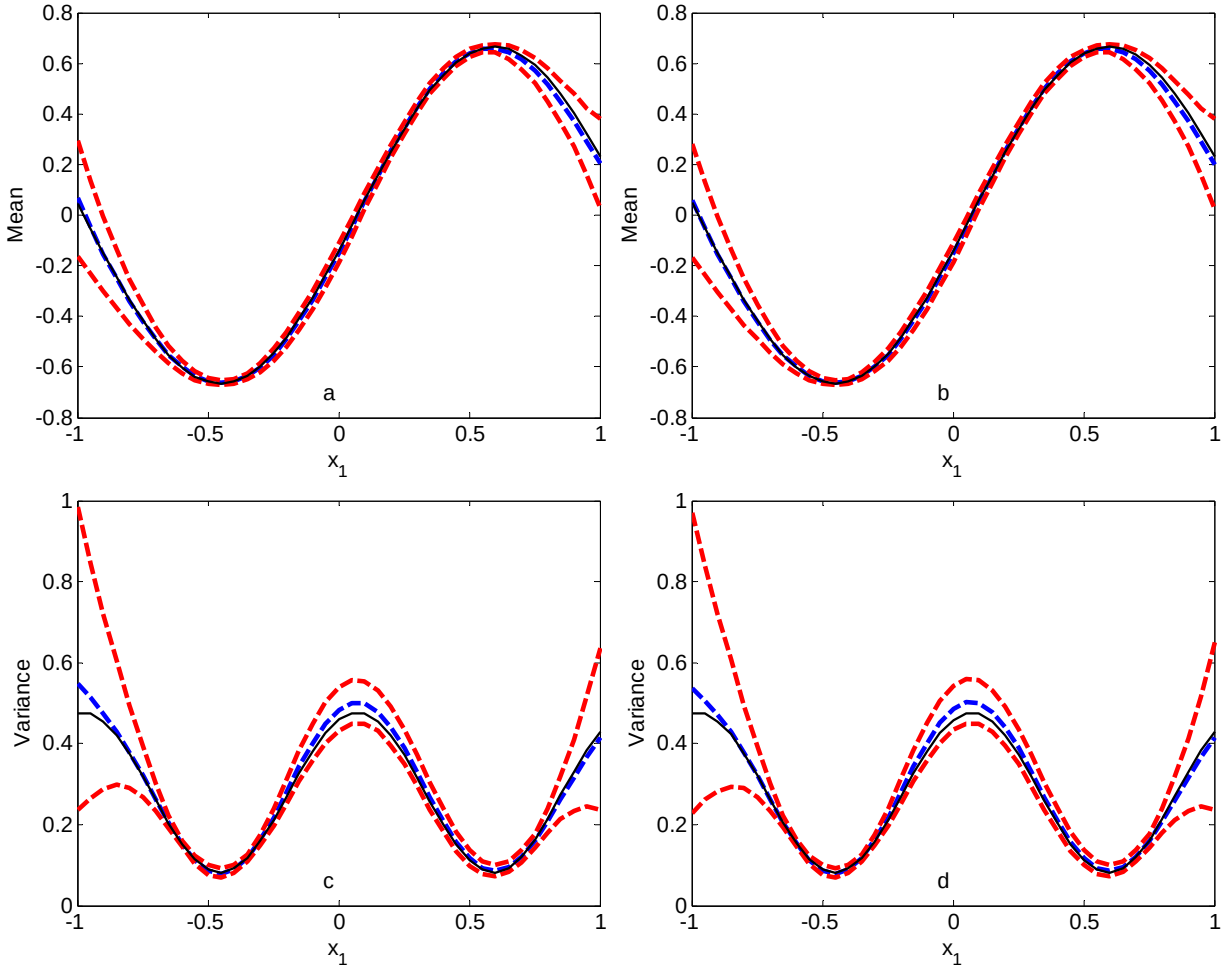

Figure 2: Point and Interval Estimates (Dashed Lines) and True Value (Solid Line) of Response Mean Obtained with WLS Method (a) and IWLS Method (b); Point and Interval Estimates (Dashed Lines) and True Value (Solid Line) of Response Variance Obtained with WLS Method (c) and IWLS Method (d).

10,000 $\widehat{\mathcal{M}}$'s, which we denote by $E(\widehat{\mathcal{M}}|\boldsymbol{Y})$ and variance function $\mathcal{V}$ by taking the average of the 10,000 $\widehat{\mathcal{V}}$'s, which we denote by $E(\widehat{\mathcal{V}}|\boldsymbol{Y})$. We also construct 98% credible intervals by using the 1% and 99% sample percentiles. The posterior mean, upper credible limit, and lower credible limit of $\widehat{\mathcal{M}}$ and $\widehat{\mathcal{V}}$ are plotted in Figure 2 as dashed lines. We also compute $\mathcal{M}$ and $\mathcal{V}$ by drawing 100,000 random samples of x_2 and calculating the mean and variance of f using these samples for each x_1. The $\mathcal{M}$ and $\mathcal{V}$ functions (or, more precisely, accurate approximations of the functions) are plotted as solid lines in Figure 1. By comparing Figures 1a and 1b, we see that the WLS and IWLS methods produce indistinguishable estimates of $\mathcal{M}$ that closely approximate and sandwich the true mean function. Similar comments apply to estimates of the variance function in Figures 1c and 1d. The wide credible intervals for the variance function near $x_1 = -1$ are due to the absence of design points in the upper left corner of Figure 1.

Appendix F: Estimation of RPD Indices for Moderately High Dimension Functions

F.1 Nondifferentiable Function

In this section, we give an example of RPD index estimation for a nondifferentiable and fairly high dimension function. We consider the case where $c = 4$, $d = 10$, $\boldsymbol{x}$ is uniformly distributed on $[-1,1]^{10}$ and the true function is a sum of two Sobol functions (Sudret, 2008): $f(\boldsymbol{x}) = |8x_1 - 6||8x_5 - 6| + |8x_2 - 6||8x_6 - 6|$.

The variance components of the Sobol function are known (see Sudret (2008)). Based on the known variance components, we obtain $\mathcal{A}_{(1)} = \mathcal{A}_{(2)} = 1/14, \mathcal{A}_{(3)} = \mathcal{A}_{(4)} = 0$ and $\mathcal{B}_{(1)} = \mathcal{B}_{(2)} = 3/14, \mathcal{B}_{(3)} = \mathcal{B}_{(4)} = 0$. We employ a maximin Latin hypercube design of size $n = 200$ to fit the GP model with Gaussian correlation function. Note that this GP model is actually inappropriate for modeling f because f is not differentiable while the GP model produces smooth sample paths. Using a target accuracy of $T = 25\%$, both WLS and IWLS methods give a

model of degree 4 with $N = 2002$, and actual achieved accuracies of about 20%. Posterior mean and 98% credible intervals given by both WLS and IWLS methods are shown in Table 1. Estimates obtained with the CJS formula and true values for $\mathcal{A}_{(s)}$ and $\mathcal{B}_{(s)}$ are also given in the table. The point estimates of $\mathcal{A}_{(s)}$ obtained by the proposed method are biased upwards from the true values and not as close to the true values as the point estimates obtained with the CJS formula. Moreover, the interval estimates for the nonzero $\mathcal{A}_{(1)}$ and $\mathcal{A}_{(2)}$ obtained with the IWLS method do not contain the true values. However, point and interval estimates of $\mathcal{A}_{(s)}$ and $\mathcal{B}_{(s)}$ given by the proposed method provide a reasonably accurate picture of the contribution of each control factor to control by noise interactions (measured by the $\mathcal{A}_{(s)}$'s) and variation in the mean (measured by the $\mathcal{B}_{(s)}$'s). In obtaining Table 1, computation with the CJS formula took around 350 seconds whereas computation with the WLS/IWLS method took around 150 seconds.

Table 1: Posterior Mean and 98% Credible Intervals, CJS Point Estimates, and True Values of $\mathcal{A}_{(s)}$ and $\mathcal{B}_{(s)}$

	s	WLS			IWLS			CJS Formula	True Value
		LCL	Mean	UCL	LCL	Mean	UCL		
$\mathcal{A}_{(s)}$	1	0.0799	0.1118	0.1504	0.0830	0.1134	0.1502	0.0701	0.0714
	2	0.0698	0.0964	0.1289	0.0721	0.0962	0.1265	0.0560	0.0714
	3	0.0193	0.0284	0.0411	0.0227	0.0316	0.0434	0.0060	0.0000
	4	0.0123	0.0171	0.0237	0.0161	0.0214	0.0285	0.0011	0.0000
$\mathcal{B}_{(s)}$	1	0.1512	0.2024	0.2563	0.1514	0.1985	0.2466	0.2360	0.2143
	2	0.1333	0.1835	0.2413	0.1356	0.1840	0.2372	0.2073	0.2143
	3	0.0030	0.0096	0.0212	0.0029	0.0085	0.0185	0.0047	0.0000
	4	0.0016	0.0040	0.0091	0.0016	0.0040	0.0089	0.0004	0.0000

F.2 Smooth Function

This section gives an example of RPD index estimation for a smooth and fairly high dimension function. We consider the case where $c = 2$, $d = 10$, $\boldsymbol{x}$ is uniformly distributed on $[-1,1]^{10}$ and the function is the Tilden function (Saltelli et al., 2000) multiplied with $1 + x_2 x_3$ and appended with five inert inputs $x_6, \ldots, x_{10}$:

$f(\boldsymbol{x}) = (1 + x_2x_3)/[2u_1u_4exp(-u_2/u_5) + 1/u_3]$,

where $u_1 = 13465000(x_1 + 1) + 8970000, u_2 = 250(x_2 + 1) + 100, u_3 = 0.1(x_3 + 1) + 1, u_4 = (4.5 \times 10^{-7})(x_4 + 1) + 1 \times 10^{-7}, u_5 = 50(x_5 + 1) + 250$. We employ a maximin Latin hypercube design $\mathcal{D}_{100}$ of size $n = 100$ to fit the GP model with Gaussian correlation function. Using a target accuracy of $T = 10\%$, both WLS and IWLS methods give a model of degree 4 with $N = 2002$, and actual achieved accuracies of about 5%. Posterior mean and 98% credible intervals given by both WLS and IWLS methods are shown in Table 2. Estimates obtained with the CJS formula are also given in the table. For all $\mathcal{A}_{(s)}$ except $\mathcal{A}_{(1)}$ and all $\mathcal{B}_{(s)}$, the discrepancies between point estimates obtained with the WLS/IWLS and CJS methods are small relative to the values of the point estimates. We perform Monte Carlo simulation to estimate all Sobol indices of f using the sobol program in the sensitivity package in R and two random samples of inputs values of size 50,000. To obtain more accurate results, we remove all five inert inputs from f. The simulation requires 1.6×10^6 evaluations of f. All third and higher order indices have small or negative estimates. Thus, we assume that they are negligible. Based on the first and second order Sobol indices, we compute the estimates of $\mathcal{A}_{(s)}$ and $\mathcal{B}_{(s)}$ shown in Table 2. Since the 95% bootstrap intervals of the Sobol indices have widths of 0.03-0.04, the Monte Carlo estimates of $\mathcal{A}_{(s)}$ and $\mathcal{B}_{(s)}$ are accurate. The Monte Carlo estimates indicate that the point estimates of all $\mathcal{A}_{(s)}$ and $\mathcal{B}_{(s)}$ except $\mathcal{B}_{(2)}$ obtained with the proposed method and the CJS formula deviate slightly from the true values.

To illustrate the importance of interval estimates of RPD indices, we randomly choose 40 runs from $\mathcal{D}_{100}$ and utilize the resulting design, which we denote by $\mathcal{D}_{40}$, to fit the GP model. Using a target accuracy of $T = 10\%$, both WLS and IWLS methods give a model of degree 3 with $N = 572$, and actual achieved accuracies of about 1.4%. Table 3 gives the posterior mean

and 98% credible intervals of $\mathcal{A}_{(s)}$ and $\mathcal{B}_{(s)}$ obtained with the WLS and IWLS methods, and also point estimates obtained with the CJS formula. The results in Table 3 clearly indicate the inadequacy of design $\mathcal{D}_{40}$. The point estimates of $\mathcal{A}_{(s)}$ and $\mathcal{B}_{(s)}$ are inaccurate and misleading. In particular, the point estimates of $\mathcal{A}_{(2)}$ are too small while those of $\mathcal{B}_{(1)}$ are too large. Thus, the existence of some very strong interactions between control factor x_2 and the noise factors is not indicated by the point estimates of $\mathcal{A}_{(2)}$ while the point estimates of $\mathcal{B}_{(1)}$ falsely suggest that control factor x_1 has an influence on the mean that is as strong as x_2. However, the wide interval estimates of $\mathcal{A}_{(s)}$ and $\mathcal{B}_{(s)}$ given by the proposed method warn us about the inaccuracy of the estimates and inadequacy of the design.

Table 2: Posterior Mean and 98% Credible Intervals, CJS Point Estimates, and Monte Carlo Estimates of $\mathcal{A}_{(s)}$ and $\mathcal{B}_{(s)}$, Design $\mathcal{D}_{100}$

	s	WLS			IWLS			CJS Formula	Monte Carlo
		LCL	Mean	UCL	LCL	Mean	UCL		
$\mathcal{A}_{(s)}$	1	0.0196	0.0337	0.0539	0.0201	0.0342	0.0539	0.0192	0.0501
	2	0.1550	0.2070	0.2661	0.1579	0.2084	0.2695	0.1953	0.2364
$\mathcal{B}_{(s)}$	1	0.0730	0.1098	0.1500	0.0740	0.1090	0.1480	0.1186	0.1332
	2	0.1404	0.1967	0.2567	0.1413	0.1977	0.2577	0.2136	0.2024

Table 3: Posterior Mean and 98% Credible Intervals, CJS Point Estimates, and Monte Carlo Estimates of $\mathcal{A}_{(s)}$ and $\mathcal{B}_{(s)}$, Design $\mathcal{D}_{40}$

	s	WLS			IWLS			CJS Formula	Monte Carlo
		LCL	Mean	UCL	LCL	Mean	UCL		
$\mathcal{A}_{(s)}$	1	0.0246	0.0524	0.0953	0.0252	0.0523	0.0942	0.0349	0.0501
	2	0.0538	0.0931	0.1486	0.0549	0.0937	0.1452	0.0730	0.2364
$\mathcal{B}_{(s)}$	1	0.1411	0.2181	0.2983	0.1422	0.2174	0.2965	0.2375	0.1332
	2	0.1083	0.1962	0.2943	0.1127	0.1965	0.2931	0.2117	0.2024

The 98% credible intervals of $\mathcal{A}_{(s)}$ and $\mathcal{B}_{(s)}$ in Table 2 are quite wide. Thus, we may want to perform additional experiment runs to improve the accuracy of the GP model and RPD estimates. Utilizing a 10,000 point subsequence of the Matlab Sobol sequence (the first 1000 points of the sequence are skipped and every 101th point is retained) as candidate set, we

sequentially add a candidate point with the largest minimum distance to the design points until the design size is 130. The resulting design, which we denote by $\mathcal{D}_{130}$, is employed to fit the GP model. Using a target accuracy of $T = 10\%$, both WLS and IWLS methods give a model of degree 4 with $N = 2002$, and actual achieved accuracies of about 3.5%. Table 4 presents the posterior mean and 98% credible intervals of $\mathcal{A}_{(s)}$ and $\mathcal{B}_{(s)}$ obtained with both the WLS and IWLS methods, and also point estimates obtained with the CJS formula. We see that the point estimates in Table 4 have better accuracy compared to the point estimates in Table 2. For example, the point estimate of $\mathcal{B}_{(1)}$ given by the WLS method is now about 13.4% instead of 11.0%, which is closer to the Monte Carlo estimate of 13.3%. The widths of the credible intervals in Table 4 are also somewhat smaller than the widths of the intervals in Table 2.

Table 4: Posterior Mean and 98% Credible Intervals, CJS Point Estimates, and Monte Carlo Estimates of $\mathcal{A}_{(s)}$ and $\mathcal{B}_{(s)}$, Design $\mathcal{D}_{130}$

	s	WLS			IWLS			CJS Formula	Monte Carlo
		LCL	Mean	UCL	LCL	Mean	UCL		
$\mathcal{A}_{(s)}$	1	0.0346	0.0494	0.0665	0.0342	0.0483	0.0647	0.0388	0.0501
	2	0.2096	0.2453	0.2849	0.2086	0.2443	0.2832	0.2369	0.2364
$\mathcal{B}_{(s)}$	1	0.1053	0.1340	0.1638	0.1066	0.1338	0.1631	0.1384	0.1332
	2	0.1502	0.1870	0.2251	0.1532	0.1882	0.2265	0.1938	0.2024

It is observed in Tables 1-4 and Section 6 that point estimates of $\mathcal{A}_{(s)}$ obtained with the proposed method are always larger than point estimates obtained with the CJS formula. This seems to be due to the fact that the posterior sample paths of the GP tend to be more oscillatory than its posterior mean. The more oscillatory sample paths tend to have larger interaction components than the posterior mean. Since $\mathcal{A}_{(s)}$ is the sum of many high order interactions, its posterior mean tends to be larger than the value of $\mathcal{A}_{(s)}$ for the posterior mean of the GP. However, in all cases, discrepancies between point estimates of all RPD indices $\mathcal{A}_{(s)}$ and $\mathcal{B}_{(s)}$ given by the CJS formula and WLS/IWLS method are not large.

References

1. Saltelli, A., Chan, K., and Scott, M. (eds) (2000). *Sensitivity Analysis*. New York: Wiley.
2. Sudret, B. (2008). "Global Sensitivity Analysis Using Polynomial Chaos Expansions," *Reliability Engineering and System Safety*, 93, 964-979.