A discussion of adaptive design for computer experiments

Wynn, Henry
LSE, Department of Statistics
Houghton Street
London (WC2A 2AE),UK
E-mail:h.wynn@lse.ac.uk

Youssef, Noha
LSE, Department of Statistics
Houghton Street
London (WC2A 2AE),UK
E-mail:h.wynn@lse.ac.uk

Introduction

The target of a computer experiment is to find a surrogate for a complex model or a simulator. This surrogate is numerically simpler and reduces the numerical complexity and run time. The surrogate is usually called an emulator or a meta model. A computer experiment, then, comprises n runs of this complex model (simulator) with various inputs $\mathbf{x}_{\mathcal{D}}$, where, $\mathbf{x}$ is the vector of the input variables of dimension d and $\mathcal{D}$ is a design of n inputs leading to a more efficient analysis of the model. The output, $Y(\mathbf{x})$, of the simulator can be univariate or multivariate. In this paper the concern is with the univariate and deterministic output. The output at any set of the inputs always yields the same output. There is no random error assumed for these deterministic models.

The aim is to explore different techniques to sequentially select the model based optimal design $\mathcal{D}$ for the computer experiment. The mathematical models used in modelling the output $Y(\mathbf{x})$ are described in Section 1 with emphasis on the Gaussian Process (GP) model. Section 2 illustrates the orthogonal expansion in approximating the covariance function of the GP model via the Karhunen-Loeve (K-L) expansion. Section 3 discusses the use of the Haar wavelets in the numerical solution for the K-L expansion. The hierarchical Bayesian analysis of the approximated GP model is explained in Section 3. The design problem using the Maximum Entropy Sampling (MES) criterion is demonstrated in Section 4. In addition, Section 4 introduces the new Approximate Adaptive Maximum Entropy Sampling criterion which is derived from the Bayesian analysis of the approximated model. The algorithm used to select the adaptive design using the new criterion is explained in details in Section 5. Examples on the sequential designs using the innovated criterion are given in Section 6. Conclusions are given in Section 6.

Modelling the deterministic output of a computer experiment

The main task for implementing the computer experiment is to construct an accurate and efficient model for the output $Y(\mathbf{x})$ for prediction purposes. Chen et al. (2003) and Fang et al. (2006) discuss different methods for modelling the output $Y(\mathbf{x})$. In most cases the assumed models are a linear combination of some basis functions.

Kriging is widely used in computer experiments (Cressie, 1988), (Sacks et al., 1989) and (Koehler and Owen, 1996). It models the deterministic computer output $Y(\mathbf{x})$ as a realization of a stochastic process, usually assumed to be a Gaussian Process. A Gaussian Process is defined to be a collection of a finite number of random variables with a joint Gaussian distribution.

The GP model for the output $Y(\mathbf{x})$ at the design point $\mathbf{x}$ where $\mathbf{x} \in \mathcal{D}$ and $\mathbf{x} = \{x_1, \ldots x_d\}$ of

dimension d has the following form

$$Y(\mathbf{x})_{1\times 1} = h^T(\mathbf{x})_{1\times p}\beta_{p\times 1} + Z(\mathbf{x})_{1\times 1} \tag{1}$$

with

$$\begin{aligned} h(\mathbf{x}) &= [h_1(\mathbf{x}), \cdots, h_p(\mathbf{x})]^T \\ \beta^T &= [\beta_1, \cdots, \beta_p]^T \end{aligned}$$

where $h(\mathbf{x})$'s are known fixed functions, β's are unknown coefficients and $Z(\mathbf{x})$ is a stochastic process. The stochastic process $Z(\mathbf{x})$ is usually assumed initially to be a stationary Gaussian process with a zero mean and a positive definite covariance function between $Z(\mathbf{x})$ and $Z(\mathbf{x}')$,

$$\mathrm{Cov}(Z(\mathbf{x}), Z(\mathbf{x}')) = \sigma^2 R(\mathbf{x}, \mathbf{x}') \tag{2}$$

where σ^2 is the process variance and $R(\mathbf{x}, \mathbf{x}')$ is the correlation function in $\mathbf{x}$ and $\mathbf{x}'$. In the matrix form, the model is represented as

$$Y_{n\times 1} = X_{n\times p}\beta_{p\times 1} + Z_{n\times n}$$

where Y is the vector of n observations, X is the design matrix, β is the vector of p mean parameters and Z is a vector of size n stochastic processes.

The classical approach using the Best Linear Unbiased Prediction (BLUP) is not followed in this paper, instead the Bayesian estimation and prediction approach is adopted. This allows for the incorporation of the updated information in the model analysis at every stage of the sequential experiment.

Orthogonal expansion of a Gaussian Process

If $\{Z(\mathbf{x})\}$ is a second order stochastic process with a zero mean and a continuous correlation function, $\sigma^2 R(s,t)$, then Mercer's theorem can be exploited to represent the correlation function $R(s,t)$ as

$$R(s,t) = \sum_{i=0}^{\infty} \lambda_i \phi_i(s)\phi_i(t) \tag{3}$$

where $\{\lambda_i, i \in N\}$ and $\{\phi_i, i \in N\}$ are the non-zero eigenvalues and the eigenfunctions of $R(s,t)$ obtained by solving the integral equation known as "Fredholm integral equation"

$$\int_a^b R(s,t)\phi_i(t)dt = \lambda_i \phi_i(s). \tag{4}$$

Series (3) converges uniformly in both variables s and t.

Correspondingly, the stochastic process $Z(\mathbf{x})$ is represented using the Karhunen-Loeve expansion (K-L) by

$$Z(\mathbf{x}) = \sum_{i=1}^{\infty} \sqrt{\lambda_i}\zeta_i \phi_i(\mathbf{x}) \tag{5}$$

and

$$\zeta_i = \frac{1}{\lambda_i}\int_a^b Z(\mathbf{x})\phi_i(\mathbf{x})d\mathbf{x}. \tag{6}$$

The basis functions $\{\phi_i(\mathbf{x})\}$ and $\{\zeta_i\}$ are orthonormal in the sense that $E(\zeta_i) = 0$ and $\mathrm{Cov}(\zeta_i, \zeta_j) = \delta_{ij}$.

In practice, the expansion in (5) is truncated at q. The choice of q is done using an arbitrary threshold, q can be chosen to be the number of eigenvalues > 0.009. The high eigenvalues correspond

to eigenfunctions mostly control the correlation function. However, for a non-smooth correlation function neglecting the small eigenvalues is not desirable for accuracy reasons.

Finally, the output $Y(\mathbf{x})$ is approximated as

$$Y(\mathbf{x}) = h^T(\mathbf{x})\beta + \sum_{i=1}^{q} \sqrt{\lambda_i}\zeta_i\phi_i(\mathbf{x}) + \epsilon(\mathbf{x}) \tag{7}$$

where

$$\epsilon(\mathbf{x}) = \sum_{i\geq q+1} \sqrt{\lambda_i}\zeta_i\phi_i(\mathbf{x}). \tag{8}$$

The term $\epsilon(\mathbf{x})$ resembles the addition of a nugget effect to the deterministic model.

The final representation of the deterministic output $Y(\mathbf{x})$ is of the form

$$Y(\mathbf{x}) = h^T(\mathbf{x})\beta + \phi(\mathbf{x})^T\gamma + \epsilon(\mathbf{x}) \tag{9}$$

where β is the vector of mean parameters of size p, $\phi(\mathbf{x})$ is a $1 \times q$ matrix of eigenfunctions and γ is a $q \times 1$ vector of mean parameters where $\gamma_i = \sqrt{\lambda_i}\zeta_i$.

Numerical solution of the orthogonal expansion

Finding the eigenvalues and the eigenfunctions of the covariance kernel requires solving the Fredholm integral equation (4). The analytical solution only exists for some particular classes of the correlation functions. Therefore the numerical solution is needed when the analytical solution is not available. The expansion method known as Galerkin method is one of the numerical methods. It is used to approximate the eigenfunction by a linear combination of chosen basis functions as described in (Huang and Phoon, 2001). The accuracy of this method depends on the number of basis functions used.

The Haar wavelet basis functions can be considered as a suitable choice to approximate the eigenfunction $\phi_i(\mathbf{x})$. Phoon et al. (2002), Ramsay and Silverman (2005), Lepik and Tamme (2007), Wang (2008) and Youssef (2011) recommend the use of the Haar basis functions in the numerical solution for the linear integral equations.

The eigenfunction $\phi_i(t)$ is expressed as a linear combination of Haar orthonormal basis functions

$$\phi_i(t) = \sum_{k=1}^{M} d_{ik}\psi_k(t) = \psi(t)^T D_i = D_i^T\psi(t)$$

where D_i is a vector of unknown coefficients $\{d_{ik}\}$, $i = 1, \ldots, q$ obtained numerically via a discrete wavelet transformation.

The number of the basis functions M is agreed to be $M = 2^n$ orthogonal basis functions. They are constructed on $[0, 1]$ in the following way

$$\psi_1 = 1; \quad \psi_i = \psi_{j,k}(\mathbf{x}); \quad i = 2^j + k + 1; \quad j = 0, 1, \cdots, n-1; \quad k = 0, 1, \cdots, 2^j - 1,$$

where the simplest Haar wavelet basis function is defined as

$$\psi(\mathbf{x}) = \begin{cases} 1 & 0 < x < \frac{1}{2} \\ -1 & \frac{1}{2} \leq x < 1 \\ 0 & \text{otherwise} \end{cases}. \tag{10}$$

and the other Haar wavelets functions are

$$\psi_{j,k}(\mathbf{x}) = \begin{cases} 1 & k2^{-j} \leq x < 2^{-j-1} + k2^{-j} \\ -1 & 2^{-j-1} + k2^{-j} \leq x < 2^{-j} + k2^{-j} \\ 0 & \text{otherwise} \end{cases}. \tag{11}$$

Function (11) forms the family of the orthogonal Haar wavelets by shifting $\psi(\mathbf{x})$ of the function in (10). These functions are orthogonal because

$$\int_0^1 \psi_{j,k}(\mathbf{x})\psi_{m,n}(\mathbf{x})d\mathbf{x} = 2^{-j}\delta_{j,m}.$$

For a detailed explanation of computing the K-L expansion for dimension $d > 1$ using Haar wavelets see (Youssef, 2011).

Hierarchical Bayesian analysis

The hierarchical Bayesian analysis assigns hyper-priors to the model parameters in (9). As assumed, the output $Y|\beta, \Lambda, \sigma^2$ has a multivariate normal distribution

$$Y|\beta, \Lambda, \sigma^2 \sim N(X\beta, \delta\sigma^2 I) \tag{12}$$

where Λ is the prior covariance for β and δ represents the nugget effect resulting from using the K-L expansion. We assign the following natural conjugate prior distributions as

$$\begin{aligned} \beta|\Lambda, \sigma^2 &\sim N(\mu, \sigma^2\Lambda) \\ \Lambda_{p+q\times p+q}|\sigma^2 &\sim IW(\Psi, m) \\ \sigma^2 &\sim IG\left(\frac{a}{2}, \frac{b}{2}\right) \end{aligned}$$

The hyper-parameter Ψ is a matrix of the same size as Λ. The hyper-parameter m corresponds to the degrees of the Inverse-Wishart (IW) distribution. The degrees of freedom m has to be greater than the size of the matrix Λ. The mean parameters β and γ are a priori independent. The output Y depends on Λ via the dependence of β on Λ.

The main interest now is to find the marginal predictive distribution $f(Y_r|Y_n)$ where Y_n is the vector of n already observed outputs and Y_r is the vector of r unobserved outputs that correspond to a future design of r points. Finding the marginal posterior distributions for the parameters or the predictive distribution is computationally hard using the ordinary methods. Integrating over β to obtain the marginal conditional distribution of $\Lambda|Y$ is not straightforward, it requires numerical methods. Empirical Bayes estimation is a commonly used alternative in the hierarchical structure (Carlin and Louis, 1996). This helps in introducing a new criterion in the next section that adapts to the observations better, i.e. selecting the next stage design requires observing the output.

The design problem

One of the design optimality criteria used in computer experiments is the Maximum Entropy Sampling design (MES) presented in (Shewry and Wynn, 1987). In general, entropy is defined as the negative measure of information,

$$\mathrm{Ent}(Y(\mathbf{x})) = \mathrm{E}_{Y(\mathbf{x})}[-\log p(Y(\mathbf{x}))]$$

where $Y(\mathbf{x})$ is a random vector, $p(\cdot)$ is a density function of $Y(\mathbf{x})$ at the design points $\mathbf{x} = \{\mathbf{x}_i\}$, $i = 1, \ldots, n$.

Maximizing the information about the parameters is equivalent to minimizing the information for prediction at the unsampled points. In this paper, the aim is to find a sequential optimal design to maximise the information about the parameters of the truncated model in (9). The Bayesian framework is suitable for the sequential strategy of choosing the design. However, it has been shown in (Youssef, 2011) that sequential strategy does not necessarily imply adaptivity except for some criteria for the Gaussian models.

Approximate Adaptive Maximum Entropy Sampling criterion

The target is to find the entropy of the predictive distribution which is $\mathrm{Ent}(Y_r|Y_n)$. This involves integrating over Λ, β and σ^2. The general entropy of the future observations Y_r in the Gaussian case, ignoring constants, is

$$\mathrm{Ent}(Y_r) \approx \log\det Var(Y_r) \approx \log\det(X_r\sigma^2\Lambda X_r^T + \sigma^2\delta I). \tag{13}$$

To reach the new criterion we initialise by computing the conditional posterior distribution of $\Lambda|\beta, Y_n, \sigma^2$ as

$$\begin{aligned}\pi(\Lambda|\beta,\sigma^2,Y_n) &\approx \frac{\pi(\Lambda,\beta,\sigma^2|Y_n)}{\pi(\beta,\sigma^2|Y_n)} \\ &\approx (\frac{1}{\sigma^2})^{\frac{p+q+m+1}{2}-1}\det(\Lambda)^{-\frac{p+q+m+1}{2}}\exp\left(-\frac{1}{2\sigma^2}(\mathrm{trace}(\Psi^*)\Lambda^{-1})\right) \qquad (14)\end{aligned}$$

where Ψ^* is given by $\Psi + (\beta-\mu)(\beta-\mu)^T$.

The above term Ψ^* can be considered as an update for the Inverse Wishart (IW) distribution parameter Ψ. The conditional expectation, $\mathrm{E}(\Lambda|\beta,\sigma^2,Y_n)$, is given by

$$\mathrm{E}(\Lambda|\beta,\sigma^2,Y_n) = \frac{\Psi+(\beta-\mu)(\beta-\mu)^T}{m-p-q-1} = \frac{\Psi^*}{m-p-q-1} \tag{15}$$

where m is the degrees of freedom of the IW distribution and $p+q$ is the number of mean parameters in the model. This conditional expectation is calculated to replace Λ in the right hand side of equation (13). However, this is not the proper posterior mean of Λ because we have not marginalised with respect to β and σ. Therefore, instead of carrying out the integration we replace β in equation (15) by some estimates. This is considered as an empirical Bayes step. We suggest two options for finding these estimates. The first option is to replace β by its OLS estimator

$$\hat{\beta}_{OLS} = (X_n^T X_n)^{-1} X_n^T Y_n \tag{16}$$

where X_n is the design matrix corresponds to the already selected design, and Y_n is the observed output at the chosen design points. The second option is to find the posterior mode Bayes estimator by maximising the joint posterior distribution with respect to β. In this case $\hat{\beta}$ is given by

$$\hat{\beta}_{Bayes} = \left(X_n^T\delta^{-1}X_n + \Lambda^{-1}\right)^{-1}\left(X_n^T\delta^{-1}Y_n + \Lambda^{-1}\mu\right). \tag{17}$$

The formula in equation (17) requires an estimate for Λ. This is replaced by its conditional expectation in equation (15). The updating formula becomes

$$\hat{\beta}_{Bayes} = \left(X^T\delta^{-1}X + (\frac{\Psi^*}{m-p-q-1})^{-1}\right)^{-1}\left(X^T\delta^{-1}Y_n + (\frac{\Psi^*}{m-p-q-1})^{-1}\mu\right) \tag{18}$$

Now, the criterion in equation (13) is equivalent to

$$\det\left(\sigma^2 X_r E(\Lambda|\beta,Y_n)X_r^T + \sigma^2\delta I\right),$$

and this becomes

$$\det\left(X_r\left[\frac{(\Psi+(\hat{\beta}-\mu)(\hat{\beta}-\mu)^T)}{m-p-q-1}\right]X_r^T + \delta I\right). \tag{19}$$

By expanding the formula in (19) we have three terms inside the determinant

$$\det\left(X_r\frac{\Psi}{m-p-q-1}X_r^T + X_r\frac{(\hat{\beta}-\mu)(\hat{\beta}-\mu)^T}{m-p-q-1}X_r^T + \delta I\right). \tag{20}$$

It can be noticed from equation (20) that this criterion is iterative. This implies that updating Ψ depends on updating the estimate of β. The next stage design is influenced by the previous one through the updated values of Ψ and $\hat{\beta}$. If the design is selected one point at a time then the effect of δ is negligible. Moreover, the first term in equation (20) is similar to the information matrix while the second term involves choosing the point where the difference between the two successive estimates of the mean function $X^T\beta$ is the maximum. The next section provides a detailed algorithm of selecting the design based on the criterion in equation (19).

Adaptive Maximum Entropy Sampling algorithm

This algorithm describes the steps to find the design $\mathcal{D}$ sequentially one point at a time using the new criterion in (19)

Initialisation Assume a candidate set E of all eligible points, a sample size s, estimates for the covariance parameters associated with the chosen covariance function, $Y_r = \emptyset$ and $\mathcal{D} = \emptyset$ and set the counter $k = 0$.

Step 1 Find the eigenvalues $\lambda_1, \lambda_2, \ldots \lambda_q$ and eigenfunctions of the corresponding covariance using the K-L expansion via the Haar wavelet basis and assume δ to be a small scalar, for example 0.0003

Step 2 Assuming a constant mean β in addition to the mean parameters obtained using the eigenvalues of the K-L expansion then the vector of the hyper-parameters of the mean parameters μ is

$$\mu^{(0)} = (\mu_0, \mu_1, \ldots, \mu_q)$$

where $\mu_0^{(0)} = 0$, $\mu_i = \sqrt{\lambda_i}\varepsilon_i$, $i = 1, \ldots, q$ and $\varepsilon \sim N(0, \sigma^2)$.

Step 3 Set the parameters of the IW such that $\psi^{(0)} = I_{q+p\times q+p}$ or it can be Λ the diagonal matrix with $\{0.001, \lambda_1, \ldots, \lambda_q\}$ and m the degrees of freedom to be less than $p+q$. Also, set $\beta^{(0)} = \mu^{(0)}$ and assume $\sigma^2 = 1$. These values are chosen arbitrarily depending on the aim of the study.

Step 4 Since the sample size at each stage is one point at a time, then the selected point is the one at which the criterion value is the maximum. The criterion $\mathcal{C}$ at the k^{th} iteration is

$$\mathcal{C}^{(k+1)} = x^{(k)^T}\Psi^{(k)}x^{(k)} + x^{(k)^T}(\hat{\beta}^{(k)} - \mu^{(k)})(\hat{\beta}^{(k)} - \mu^{(k)})^T x^{(k)} + \delta.$$

If $k > 0$ then $\mu^{(k)} = \hat{\beta}^{(k-1)}$.

Step 5 Once the design point $\mathcal{D}_k$ is chosen we observe the output at this design point. Set $k = k+1$, $Y_r = Y_r \bigcup Y^{(k)}$, $E = E \setminus \mathcal{D}_k$ and $\mathcal{D} = \mathcal{D} \bigcup \mathcal{D}_k$. Compute $\hat{\beta}^{(k)}$ using either equation (16) or equation (17) and $\hat{\Psi}^{(k)}$ and hence $E(\Lambda|\beta, \sigma^2, Y)^{(k)} = \frac{\hat{\Psi}^{(k)}}{m-q-p-1}$.

Step 6 Repeat Steps 4 and 5 till we get the design $\mathcal{D}$ of size s.

Example on Approximate Adaptive Maximum Entropy Sampling design

This example aims at studying whether the design selected is adaptive to the observed data. This can be illustrated by choosing a simulator function and comparing the design points locations

with the shape of the simulator function. We choose, as a simple example, the simulator function used by (Saltelli, 2000) in two dimensions

$$Y(\mathbf{x}) = \frac{(x_2 + \frac{1}{2})^4}{(x_1 + \frac{1}{2})^2}$$

with the power exponential covariance function

$$\text{Cov}(Z(\mathbf{x}), Z(\mathbf{x}')) = \sigma^2 \exp(-\theta |x_{ki} - x_{kj}|^{q_k}) \tag{21}$$

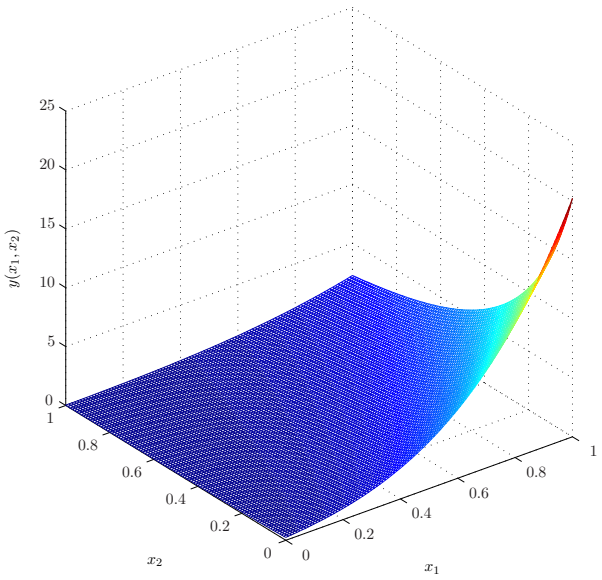
where $\hat{(}y(\mathbf{x})) = X^T \hat{\beta}$. The simulator shape is shown in Figure 1(a). Estimates of the covariance parameters are available from a previous study, see Table . The truncated K-L expansion at $q = 6$ is obtained using $M = 256$ Haar wavelets basis functions. For simplicity reasons, the process variance σ^2 is assumed to be known and equals to one. The nugget is assumed to be a small scalar, $\delta = 0.0003$. The design selection is implemented sequentially one point at a time. To observe the adaptivity and the improvement in the prediction accuracy we calculate the Empirical Root Mean Squared Error (ERMSE). Hence, a validation sample of size n_v is required to calculate the ERMSE. The validation sample is chosen to be a Halton sequence of size 30. The ERMSE is then calculated as

$$ERMSE = \sqrt{\frac{1}{n_v} \sum_{i=1}^{n_v} (\hat{y}(\mathbf{x}_i) - y(\mathbf{x}_i))^2}.$$

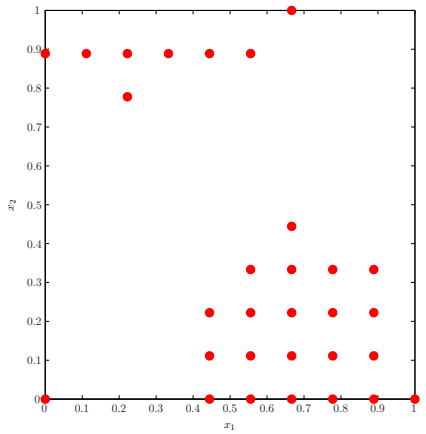
The design is obtained as described in (Youssef, 2011). The algorithm depends on finding a value for $\hat{\beta}$ and then substituting in the updating formulae to estimate Ψ^* at every stage and then find the design point. In practice, two designs are obtained using both estimators of $\hat{\beta}$ in (16) and in (17). Using both estimators for $\hat{\beta}$ has proved that the designs are adaptive to the observations. The sequential design obtained by using the estimator in (16) is shown in Figure 1(b) while the one obtained using the estimator in (17) is shown in Figure 1(c). The ERMSE for both designs is proved to decrease by observing more data. By comparing the design points locations in Figure 1(b) to the shape of the simulator in Figure 1(a) we can observe that the design points are more located in the regions where the simulator function has some peaks. The design has fewer points where the curve is flat and has no peaks or no fluctuations. The same behaviour is noticed for the design obtained using the Bayes estimator in (17), see Figure 1(c). Both designs have a decreasing ERMSE values, see Figure 1(d) and Figure 1(e). The ERMSE curve in both figures starts with high values then decreases when the sample size increases with some irregular changes in the middle. This implies that the chosen design is adaptive to the data.

$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{q}_1$	$\hat{q}_2$
2.093	1.547	1.981	1.981

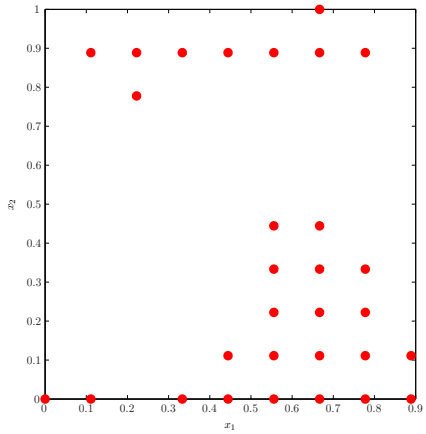
Table 1: Estimates of the exponential covariance parameters



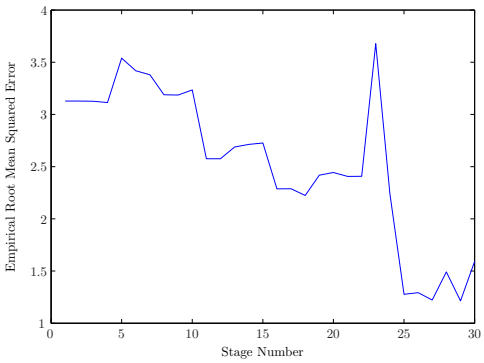
(a) The Saltelli's simulator .



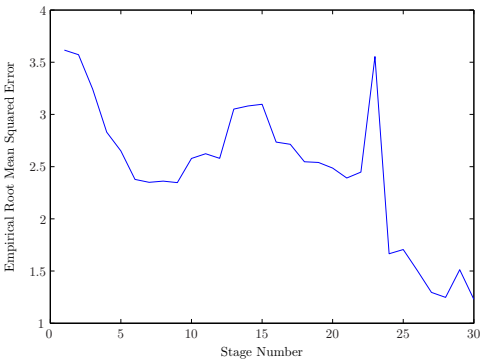
(b) A 30 point design using the approximate adaptive MES using the OLS estimator for β.



(c) A 30 point design using the approximate adaptive MES using the Bayes estimator for β.



(d) Plot of the Empirical Root Mean Squared Error against the stage number using the OLS estimator for β.



(e) Plot of the Empirical Root Mean Squared Error against the stage number using the Bayes estimator for β.

Figure 1: Figures for Example 1

Conclusions

New methods to reduce the complexity of the analysis of a Gaussian process are investigated in this study. This includes the use of the K-L expansion to approximate the covariance function of the process. Haar wavelet basis functions are the preferred methods to approximate the eigenfunctions when the analytical solution is not available. The proceeding analysis is directed to find a fully adaptive procedure which uses the past output values as well as the input values. This is achieved by having a hyper-parameter (hierarchical) structure on the model. An Inverse Wishart is assumed as a prior distribution on the covariance matrix. To avoid integration an empirical Bayes solution is adopted. This empirical Bayes step leads to find an adaptive criterion called an Approximate Adaptive Maximum Entropy Sampling criterion. Examples are given to demonstrate the adaptivity of the criterion.

REFERENCES (RÉFERENCES)

1 Chen, V. C. P., Tsui, K.-L., Barton, R. R., and Allen, J. K. (2003). A review of design and modelling in computer experiments. In: *Statistics in industry*, volume **22** of *Handbook of Statist.*, pages 231261. North-Holland.

2 Cressie, N. (1988). Spatial prediction and ordinary kriging. *Math. Geol.*, 20(4):405 421. New advances in geostatistics (Redwood City, CA, 1987).

3 Fang, K.-T., Li, R., and Sudjianto, A. (2006). *Design and modelling for computer experiments*. Computer Science and Data Analysis Series. Chapman and Hall/CRC, Boca Raton, FL.

4 Koehler, J. R. and Owen, A. B. (1996). Computer experiments. In: *Design and analysis of experiments*, volume **13** of *Handbook of Statist.*, pages 261308. North- Holland.

5 Lepik, U. and Tamme, E. (2007). Solution of nonlinear Fredholm integral equations via the Haar wavelet method. *Proc. Estonian Acad. Sci. Phys. Math.*, **56(1)**:1727.

6 Phoon, K. K., Huang, S. P., and Quek, S. T. (2002). Simulation of second-order processes using Karhunen-Loeve expansion. *Comput. and Structures*, **80(12)**:1049 1060.

7 Ramsay, J. O. and Silverman, B. W. (2005). *Functional data analysis*, (second ed.). Springer Series in Statistics. Springer.

8 Sacks, J., Schiller, S. B., and Welch, W. J. (1989a). Designs for computer experiments. *Technometrics*, 31(1):4147.

9 Saltelli, A. (2000). Fortune and future of sensitivity analysis. In: *Sensitivity analysis*, Wiley Ser. Probab. Stat., pages 421447. Wiley.

10 Shewry, M. C. and Wynn, H. P. (1987). Maximum entropy sampling. *Journal of Apply Statistics*, **14**:165170.

11 Wang, L. (2008). *Karhunen Loeve Expansion and its Applications*. PhD thesis, London School of Economics.

12 Youssef, N. (2011). *Optimal Experimental Design for Computer Experiments*. PhD thesis, London School of Economics.