

Considerations of accuracy and uncertainty with kriging surrogate models in single-objective electromagnetic design optimisation

G. Hawe and J. Sykulski

Abstract: The high computational cost of evaluating objective functions in electromagnetic optimum design problems necessitates the use of cost-effective techniques. The paper discusses the use of one popular technique, surrogate modelling, with emphasis placed on the importance of considering both the accuracy of, and uncertainty in, the surrogate model. After a brief review of how such considerations have been made in the single-objective optimisation of electromagnetic devices, their use with kriging surrogate models is investigated. Traditionally, space-filling experimental designs are used to construct the initial kriging model, with the aim of maximising the accuracy of the initial surrogate model, from which the optimisation search will start. Utility functions, which balance the predictions made by this model with its uncertainty, are often used to select the next point to be evaluated. In this paper, the performances of several different utility functions are examined, with experimental designs that yield initial kriging models of varying degrees of accuracy. It is found that no advantage is necessarily achieved through a search for optima using utility functions on initial kriging models of higher accuracy, and that a reduction in the total number of objective function evaluations can be achieved if the iterative optimisation search is started earlier with utility functions on kriging models of lower accuracy. The implications for electromagnetic optimum design are discussed.

1 Introduction

Optimisation problems in electromagnetic design are typified by features that present difficulties to most deterministic search algorithms, such as the existence of multiple local minima. Genetic algorithms (GAs), on the other hand, with their ability to search more globally, are better suited for exploring complicated objective function landscapes. The high computational cost of evaluating the objective function in such problems, however, means that direct use of a GA is often not feasible or is impractical, owing to their general requirement for a large number of objective function evaluations. Additional cost-effective techniques must be used, with the aim to make the GA require fewer evaluations of the objective function. Techniques used include hybrid algorithms [1], GAs that are specially adapted for small population sizes [2] and simplification of the problem by removal of irrelevant design variables [3]. One technique that has recently attracted significant attention, called surrogate modelling [4], is the focus of this paper.

A surrogate model is a functional relationship between the design variable space of an optimisation problem and the objective function space, which is constructed based on a set of design vectors that have their objective function values known, referred to henceforth as ‘experiments’. Once a surrogate model has been constructed, a GA can

then use it to predict fitness values for unevaluated design vectors, rather than call the true expensive objective function, thus reducing computational costs. Surrogate-assisted optimisation can therefore be viewed as consisting of two separate stages

- (i) the construction of the initial surrogate model from a set of experiments (off-line learning)
- (ii) the subsequent iterative optimisation search where, during each iteration, predictions from the most recently constructed surrogate model are used to select a design vector for evaluation. Once evaluated, this design vector is then used to construct a more accurate kriging model (on-line learning).

Accuracy and uncertainty considerations play an important role in surrogate-assisted optimisation. In the first stage, the set of experiments is typically chosen with the aim of maximising the accuracy of the initial surrogate model; this allows predictions made by the model to be used with greater confidence. Uncertainty in the surrogate model is then usually considered during the optimisation search and has an influence on the selection of where to sample next in the design variable space.

This paper includes a brief review of how accuracy and uncertainty considerations have been used with three popular surrogate models (polynomial interpolation, artificial neural networks and kriging) in electromagnetic optimum design. Although reviews of the use of different surrogate models exist in the literature, none focuses on accuracy and uncertainty considerations. The paper also comprises an investigation into how kriging model accuracy depends on features of the experimental design used, and an investigation into how the performance of different utility

functions in locating the optimum of several test functions depends upon the accuracy of the kriging model used.

It should be noted that, for the accuracy of a surrogate model to be determined, the true function it is approximating must be known, so that a comparison can be made between the two, and the relevant accuracy statistics can be computed. Real electromagnetic examples are therefore not used in this paper, as the true objective function is almost always unknown. This should not pose any concern anyway, as no problem exists that is truly representative of all electromagnetic optimum design problems; the logic is that there is as much justification in extending conclusions based on an investigation of three different test functions to real electromagnetic problems, as there is in extending conclusions based on any single electromagnetic design problem to real electromagnetic problems.

2 Use of surrogate model accuracy considerations in electromagnetic design optimisation

2.1 Polynomial approximations

Polynomial approximations are among the most common methods of surrogate model construction. Despite their popularity, however, polynomial approximations suffer in several respects. A minimum number of experiments is required in the experimental design before a polynomial model (of a given order) can be constructed. Also, it is only for this minimum number that the model will actually interpolate the experiments; use of additional points through on-line learning results in an over-determined system of equations, and the resulting polynomial surrogate model will no longer be an interpolating surface, that is the inclusion of additional points into the model does not necessarily lead to increased model accuracy. Furthermore, if uncertainty considerations are not made during the selection of where to sample next during the optimisation search, so that only the optimum of the surface is added in each iteration, not only can the model accuracy decrease, but rapid convergence to a false optimum can occur [6].

In [7], a polynomial surrogate model was used to optimise a brushless permanent magnet motor. The accuracy of the model was considered at the two main stages of the optimisation process, namely during the construction of the initial surrogate model from an experimental design, and is the choice of which points to evaluate during an iteration of the optimisation search. The experimental design was constructed so as to minimise the condition number of the matrix $[M]$ that was to be inverted to determine the polynomial coefficients; furthermore, to ensure it was interpolating (as well as keeping computational cost to a minimum), only the minimum number of points required to construct the polynomial model were used. A dynamic weighting factor was then used, as the optimisation search proceeded, to place more emphasis on the region around the predicted optimum. Then, to ensure that $[M]$ did not become ill-conditioned as the search continued, additional *learning points* were evaluated, chosen specifically so as to minimise the condition number of $[M]$. An analysis of the errors on predicted optima and learning points indicated that the inclusion of learning points was effective in improving the accuracy of the polynomial surrogate model.

2.2 Artificial neural networks

A wide range of different types of ANN exist that can be used to construct surrogate models [8]. One of the most

widely used is the radial basis function ANN [9]. An example of the use of such an ANN in electromagnetic optimum design can be found in [10], where a multiquadratic radial basis function was used as a surrogate model in the optimisation of a C-core magnet and a magnetiser. In addition to evaluation of the predicted optimum during on-line learning, design vectors in the most unexplored regions of design space were also evaluated, with the aim of avoidance of local minima. Unexplored regions of design variable space have the highest uncertainty in their values, and so the use of such points in on-line learning facilitated reduction of the uncertainty in the surrogate model as the optimisation search proceeded.

2.3 Kriging

Kriging has recently been recognised as a useful method for surrogate model construction in electromagnetic optimum design [5]. Owing to its statistical nature, useful information can be extracted, giving an indication of model uncertainty. The efficient global optimisation (EGO) algorithm [11] uses such information to build up a utility function, known as the expected improvement, that automatically balances the objective function values predicted by the kriging model with the uncertainty in this prediction. Through optimisation of this auxiliary function, model uncertainty is automatically considered as the optimum is being searched for. Expected improvement and other utility functions are reviewed in the following section.

A variation of EGO, known as superEGO, has been used to solve two electromagnetic design problems with expensive objective functions [12], one with three design variables and one with four. Experimental designs consisting of random points sparsely distributed in design variable space were used to construct the initial kriging surrogate models, and convergence was found to occur within tens of iterations in both cases.

The simultaneous growth in the development of modern experimental designs for computer experiments and that of the use of utility functions with kriging models for optimisation have motivated this investigation into the relationship between the two areas.

3 Kriging model accuracy and experimental design

3.1 Modern experimental designs

All surrogate models are constructed from a set of experiments, and it is the purpose of design of experiments [13] to determine the set that will yield the most accurate possible surrogate model. This set will vary according to the nature of the true (unknown) function being modelled, and so generic experimental designs that have desirable properties are often used. Traditional experimental designs are increasingly being replaced by modern experimental designs [14] that have been developed for use with deterministic computer experiments. Such designs usually concentrate on the desirable properties of being uniform and space-filling, although recently the usefulness of such properties has come under scrutiny [15]. Two experimental designs are investigated in this paper

(a) Latin hypercube: a Latin square is a square with each side subdivided, so that there are N rows and N columns. N points are then positioned (at random) in N of the N^2 segments, so that only one point appears in each row and each column. A Latin hypercube [16] extends this idea to higher

dimensions. If each side of the hypercube represents a different design variable, N design vectors can be selected using this design, with the advantage that the number of experiments is independent of the number of design variables.

(b) Hammersley sequence: a Hammersley sequence provides a low-discrepancy set of experiments, in that the points are uniformly distributed in design variable space. For details of the construction of a Hammersley sequence, the reader is referred to [17].

Through consideration of the accuracy of kriging models constructed using these two experimental designs for a simple test function in the following section, two further potentially attractive features of experimental designs for the construction of kriging models have been suggested.

3.2 Potential problem for experimental designs for kriging model construction

Given a set of experiments, a kriging model is defined by a vector θ that is found by maximisation of the likelihood function [5]

$$-\frac{N \ln(\sigma^2(\theta)) + \ln(|\mathbf{R}(\theta)|)}{2} \quad (1)$$

where N is the number of experiments used to construct the kriging model, σ^2 is the variance as predicted by the kriging model, and \mathbf{R} is the matrix of correlations between the N experiments, whose i - j th entry is usually formed using the Gaussian correlation function

$$R(\mathbf{x}^i, \mathbf{x}^j) = \prod_{k=1}^n e^{-\theta_k |\mathbf{x}_k^i - \mathbf{x}_k^j|^{p_k}} \quad (2)$$

From this, it is clear that θ_k determines how rapidly the correlation is lost in the k th design variable (large values implying rapid loss in correlation), and p_k determines the smoothness of the function of the k th design variable (values near 2 denoting smooth functions, values near 0 denoting rough functions). Usually, $p_k = 2$ is used for all k , and this is the value that will be used for all of the kriging models constructed in this investigation.

The behaviour of the likelihood function determines the value obtained for the vector θ and so determines the behaviour of the kriging model that is constructed. Poor behaviour of the likelihood function is likely to result in a poor value of θ being obtained and, consequently, in an inaccurate kriging model. One common example of poor behaviour for this function is monotonicity, as this results in an extremely large value of θ being used to construct the kriging model, implying virtually no correlation between any of the design vectors, which is undesirable for most functions.

An example of this can be seen in Fig. 1, where the Latin hypercube and Hammersley sequence experimental designs have both been used to construct kriging models for the Wilkinson test function [18]

$$f(x) = 8.9248 \times 10^{-5}x - 2.18343 \times 10^{-2}x^2 + 0.998266x^3 - 1.6995x^4 + 0.2x^5 \quad (3)$$

Each experimental design results in a monotonic likelihood function (note that the negative sign of the likelihood functions plotted in the figures has been omitted, so that the likelihood functions in the figures are now to be minimised), and so, for each kriging model constructed, an extremely

large value of θ is used; hence, their inaccuracy. At this point, two potentially attractive features of experimental designs are suggested, and their effect on the behaviour of the likelihood function for constructing kriging models for (3) is examined.

3.3 Potentially attractive features of experimental designs

For an n -dimensional design variable space, the two following attractive features of experimental designs for the construction of kriging models are proposed:

- (i) the inclusion of the 2^n extreme boundary points of the design variable space; that is, for a problem with n design variables that vary as $x_i' \leq x_i \leq x_i''$, $i = 1, \dots, n$, the 2^n corners of the hypercube that bounds the feasible region in design variable space are included
- (ii) for a particular experiment x^* in the experimental design, the inclusion of n additional experiments a small distance away from x^* , in each of the n orthogonal directions around x^* .

The first of these features is proposed as kriging models are designed to interpolate points in multidimensional space. Therefore the inclusion of the extreme boundary points in the design variable space in the experimental design means that extrapolation by the kriging model is kept to a minimum. The second of these features is proposed as, because a kriging model interpolates the experiments it is constructed from, the inclusion of such points forces the gradient around the experiment x^* to be highly accurate in the constructed kriging model.

It should be noted that these features can occur naturally in some experimental designs (for example, the Hammersley sequence design will always include one of the boundary points); alternatively, these features can be combined with existing experimental designs, giving a hybrid design that has some of the features of both.

Each of these features will be combined with the Latin hypercube and Hammersley sequence designs, to give two additional variants of each: designs that have boundary points included will be referred to as type II variants, and designs that have the gradient enforced to be correct at a point will be referred to as type III variants (Type I variants will be the experimental design without either feature imposed). In addition, the effect of combining both features simultaneously with an experimental design will be investigated; these will be referred to as type IV variants.

The effect of each of these variants on the construction of kriging models for the Wilkinson test function can be seen in Fig. 2. Each variant successfully removes the monotonicity of the likelihood function, resulting in more reasonable values for θ being used to construct the kriging model. Similar results were obtained for the Latin hypercube variants. Note that, for fair comparison, the same number of experiments is used in each variant. This means, for example, that a type II Hammersley sequence of size 6 in one dimension actually uses a Hammersley sequence of size 5 (the two boundary points plus the four non-boundary points of a Hammersley sequence of size 5, giving a total of six points).

The performance of each of these experimental design variants in the construction of accurate kriging models will now be investigated for a range of test functions.

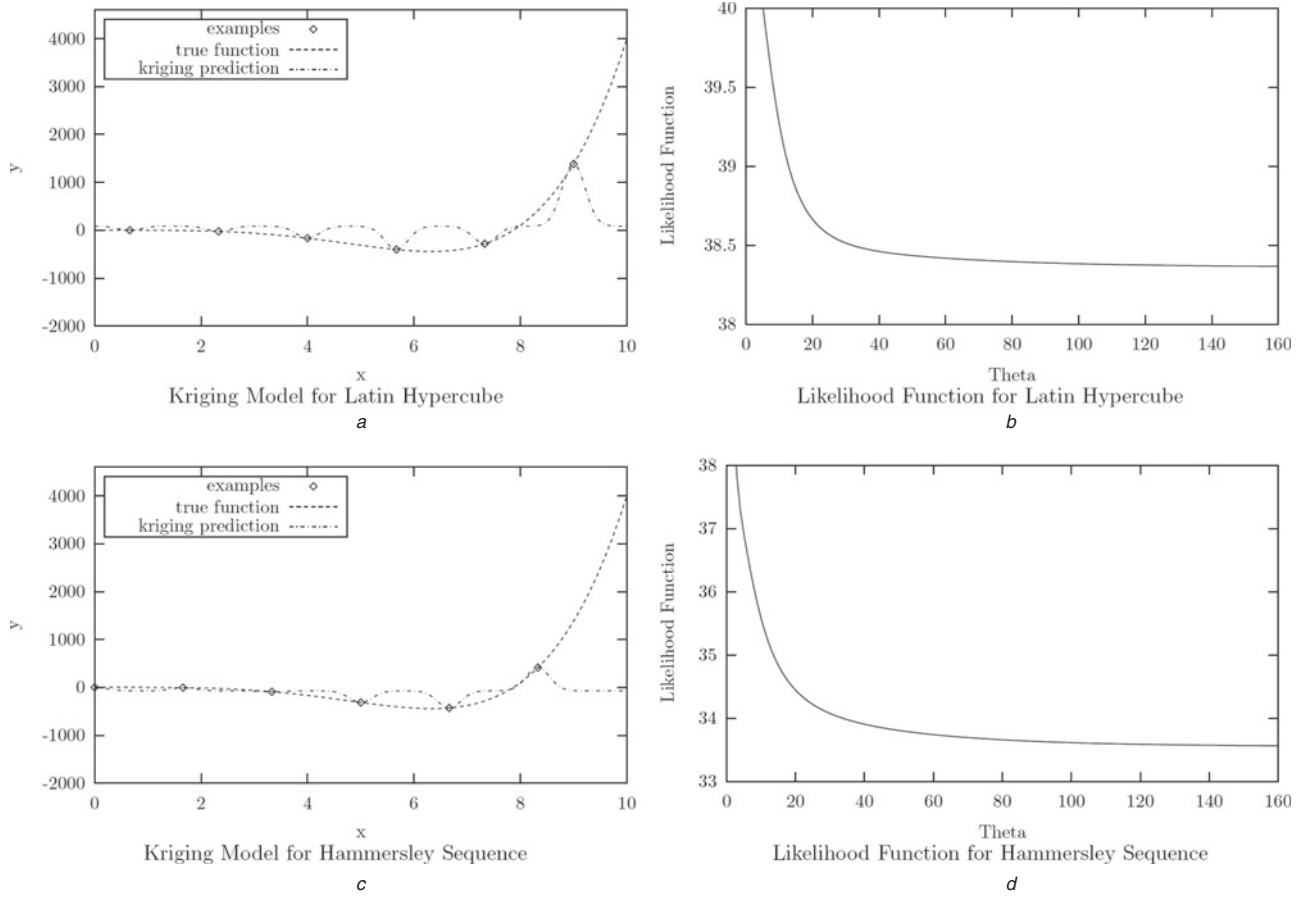


Fig. 1 Kriging models and likelihood functions for Wilkinson test function using Latin hypercube and Hammersley sequence experimental designs

3.4 Test functions

Strictly speaking, the true accuracy of a surrogate model can only be determined if the true function, which it is attempting to approximate, is also known and is available for comparison. Almost all electromagnetic optimum design problems are non-analytic, meaning it is impossible to measure the accuracy of any surrogate model constructed. For this reason, the investigation in this paper will be carried out using well-known test functions. In total, three test functions have been selected: one one-dimensional (1D) test function, known as Humps (as used in [19])

$$f(x) = 6 - \frac{1}{(x - 0.3)^2 + 0.01} - \frac{1}{(x - 0.9)^2 + 0.04} \quad (4)$$

one two-dimensional (2D) test function, known as Branin [20]

$$f(x_1, x_2) = (1 - 2x_2 + 0.05 \sin(4\pi x_2) - x_1)^2 + (x_2 - 0.5 \sin(2\pi x_1))^2 \quad (5)$$

and one three-dimensional (3D) test function, known as Hartman3 [21]

$$f(x_1, x_2, x_3) = -\sum_{i=0}^3 c_i \exp\left(-\sum_{j=0}^{n-1} A_{ij}(x_j - p_{ij})^2\right) \quad (6)$$

where the matrices A and p and the vector c are as defined in [21].

Each of these functions will be evaluated over a mesh of about $N = 1000$ points (1000 equally spaced points for (4), $32 \times 32 = 1024$ equally spaced points for (5), and

$10 \times 10 \times 10 = 1000$ equally spaced points for (6)). The kriging models are then evaluated over the same meshes for a variety of experimental designs, and compared with the true functions so that their accuracy can be determined.

3.5 Accuracy measures

A variety of methods exist for comparing the surrogate model with the true function to assess its accuracy. Three methods will be used here: the normalised maximum error NEMAX, the normalised root mean squared error NRMSE and the Kolmogorov–Smirnov statistic D . NEMAX and NRMSE are defined as

$$\text{NEMAX} = \frac{\max \{|y_i - \hat{y}_i| \mid i = 1, \dots, N\}}{\sqrt{(1/N) \sum (y_i - \bar{y})^2}} \quad (7)$$

$$\text{NRMSE} = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i)^2}} \quad (8)$$

where y_i is the true value at site i , \hat{y}_i is the value predicted by the kriging model at site i , and \bar{y} is the true mean of the points at i . Obviously, smaller values of NEMAX and NRMSE indicate greater accuracy. The Kolmogorov–Smirnov statistic D is defined as the maximum distance between the predicted and the true cumulative distribution functions, denoted $S_N(y)$ and $P_N(y)$, respectively,

$$D = \max_{-\infty < y < \infty} |S_N(y) - P_N(y)| \quad (9)$$

with smaller values of D indicating greater accuracy.

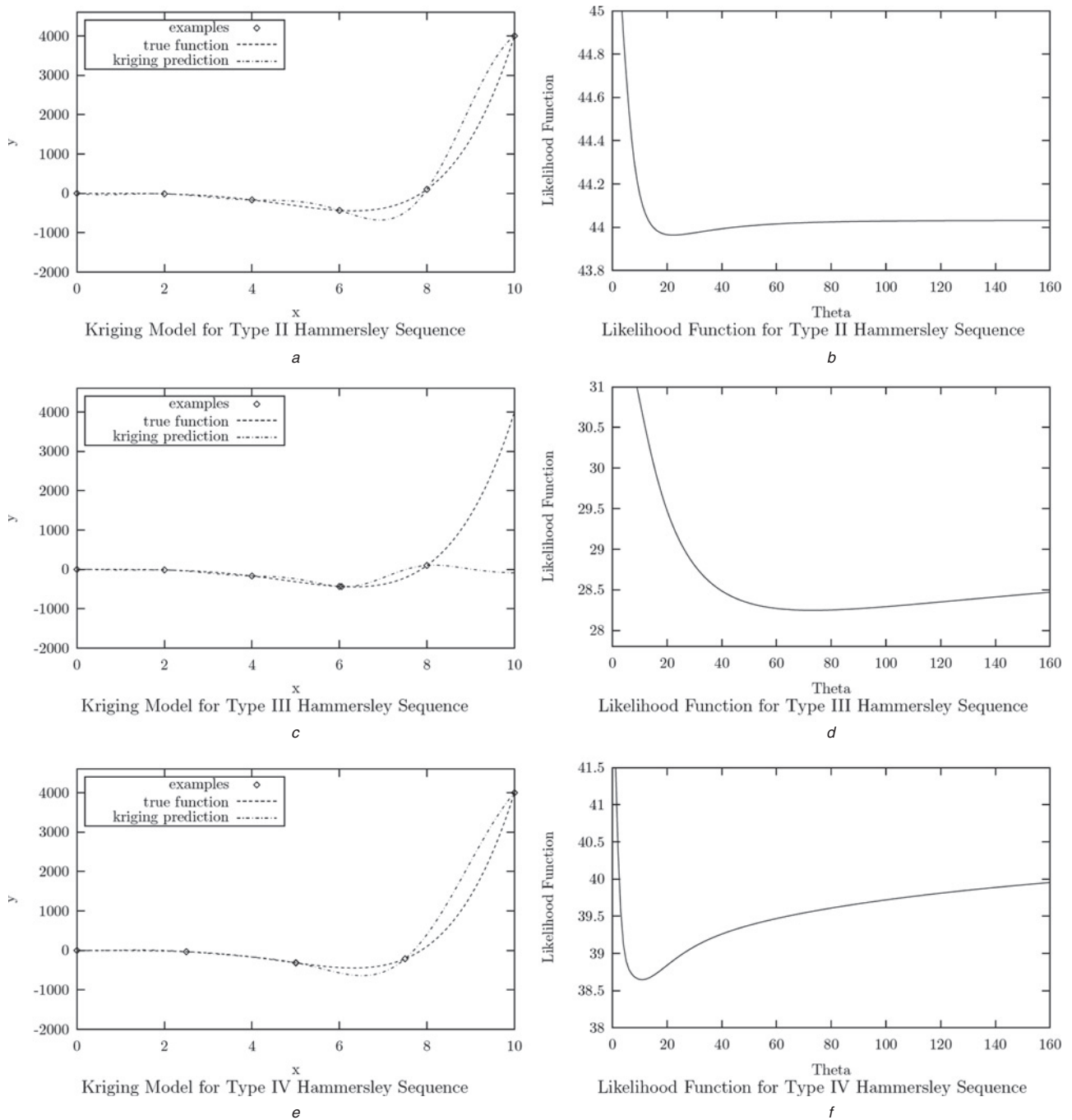


Fig. 2 Kriging models and likelihood functions for Wilkinson test function using variants of Hammersley sequence experimental design

3.6 Surrogate model accuracy dependence on experimental design

Each of the four variants of the Latin hypercube and the Hammersley sequence was used to construct initial kriging models for each of the three test functions, for varying experimental design sizes. The accuracy of the kriging models constructed was then assessed using each of the measures given above. The results are shown in Figs. 3–5.

The results for the 1D Humps problem showed that there is no significant difference in performance between the two main types of experimental design. This is to be expected in one-dimensional problems: for an experimental design size of n , when the 1D design variable space is divided into n equally sized intervals, the Latin hypercube will place one point at random in each interval, whereas the Hammersley sequence will place one point at the start of each interval. As would

be expected, the trend within each experimental design is that accuracy increases as the experimental design size increases, but this increase is perhaps not as significant as we might expect. For small experimental design sizes, some increase in accuracy resulted from the imposition of the suggested features; however, no general increase in accuracy was found in experimental designs of greater size.

Results for the 2D Branin function showed clearer trends. The simultaneous forcing of the gradient to be correct at a point and inclusion of the boundary points yielded significant increases in accuracy by all measures for both experimental design types, particularly the Hammersley sequence design. The enforcing of a correct gradient at a point gave increases in accuracy when combined alone with the Hammersley sequence design; however, its performance, when combined with the Latin hypercube, did not always result in an increase in accuracy.

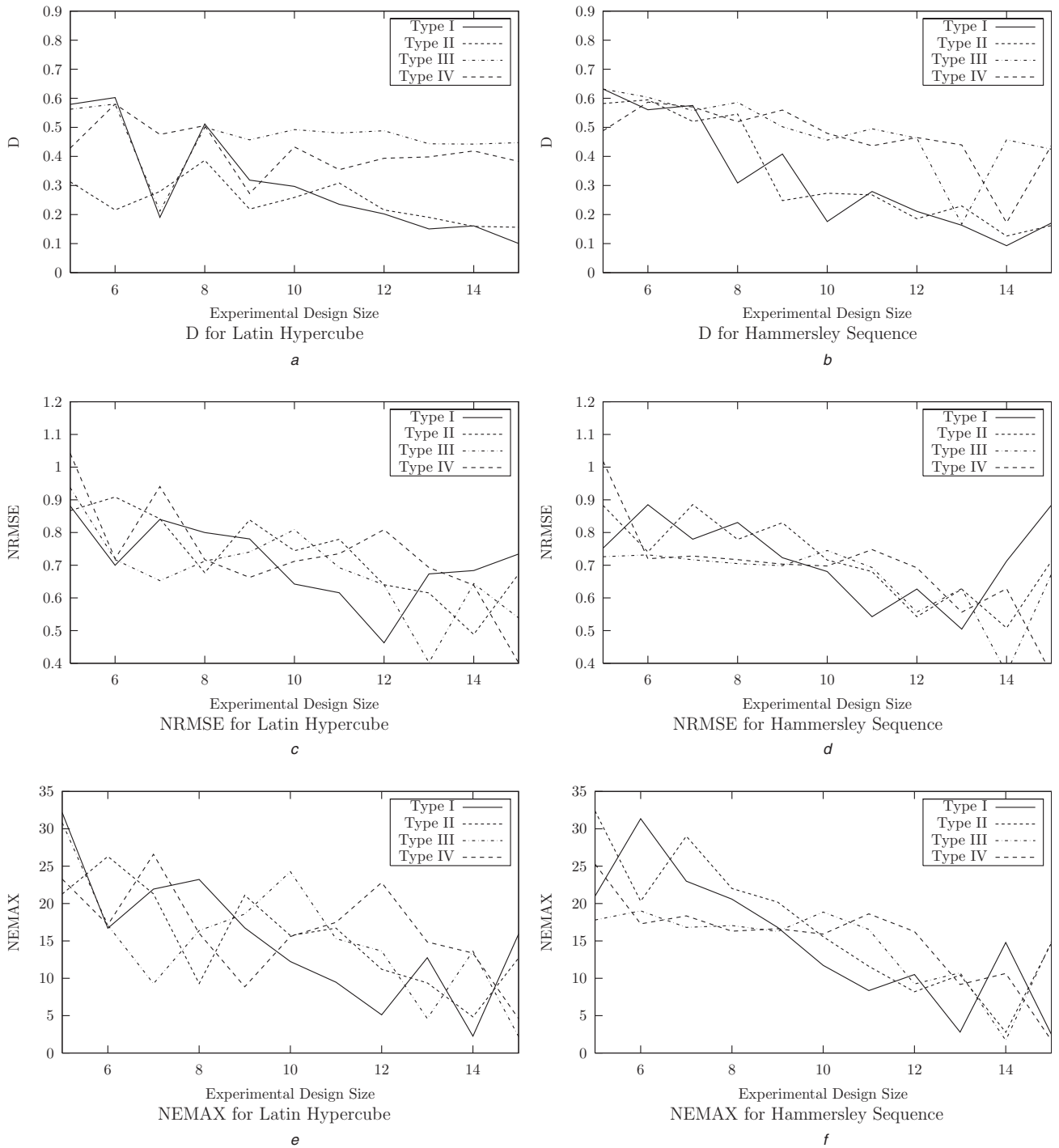


Fig. 3 Accuracy measures (Kolmogorov-Smirnoff statistic D , NRMSE and NEMAX) for kriging models constructed using different experimental designs for Humps test function

Results for the 3D Hartman3 function showed that inclusion of the boundary points in either design in general did not give an increase in accuracy. However, the combination of either or both of the suggested features with the Hammersley sequence design led to an increase in accuracy for smaller design sizes. The forcing of the gradient to be correct at a point in the kriging model tended to result in an increase in accuracy, although this gain in accuracy lessened as the experimental design size grew.

In general, the only feature of experimental designs that consistently resulted in an increase in surrogate model accuracy was an increase in the number of experiments used. However, as the task of optimisation is not accurately to approximate an unknown function, but to locate the minimum of it using as few experiments as possible, the merit of these extra

experiments at such an early stage of the optimisation procedure is questionable, given that utility functions exist that search for the minimum of an unknown function while accounting for the uncertainty in its approximation. The following section begins with a review of such utility functions, before continuing to investigate how their performance depends on the accuracy of the kriging model yielded by the experimental design.

4 Utility functions in kriging surrogate models

4.1 False optima and the need for utility functions

The choice of where to sample next in design variable space based solely on the values predicted by a surrogate model brings with it the possibility of being trapped by false

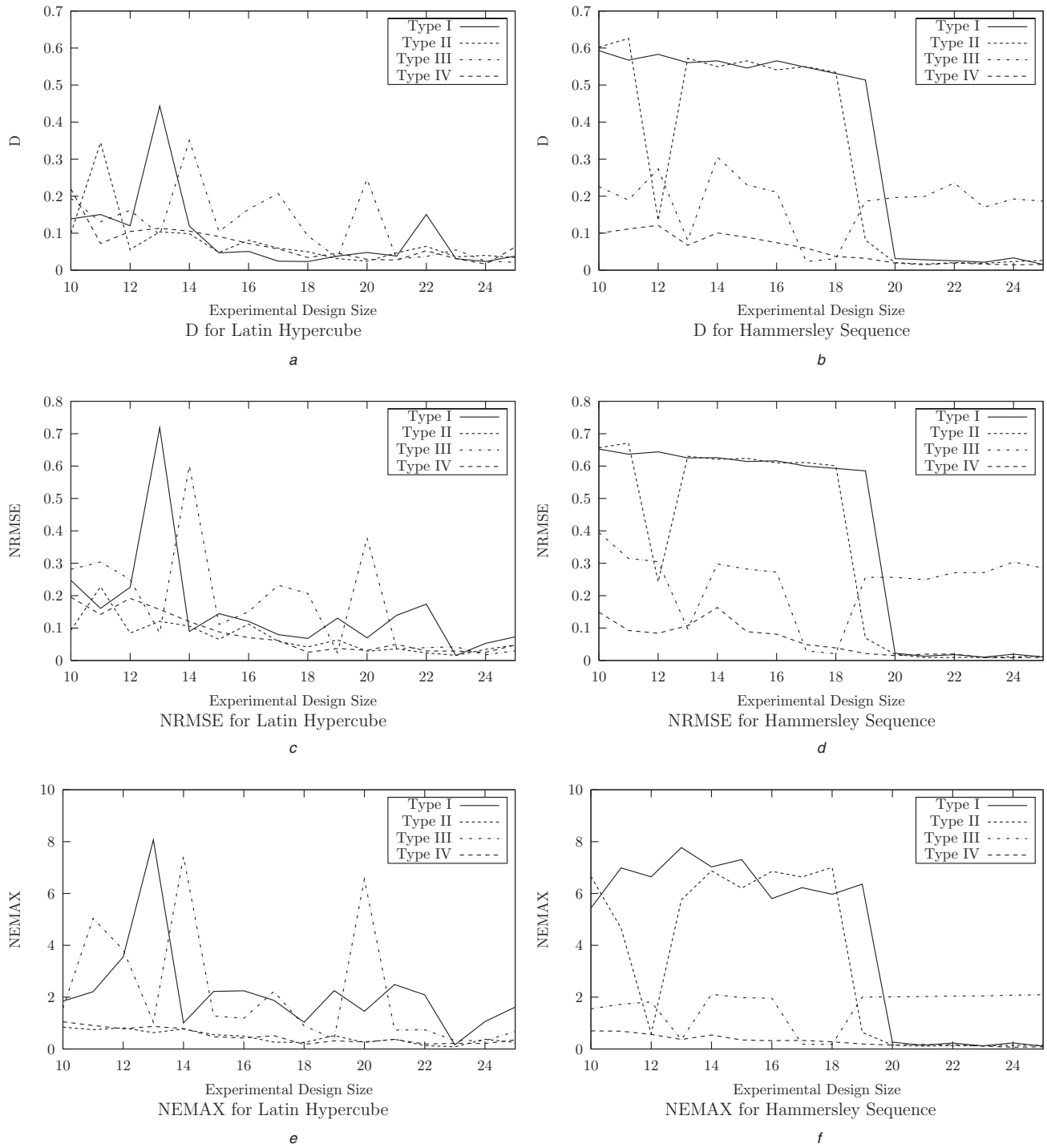


Fig. 4 Accuracy measures (Kolmogorov-Smirnoff statistic D , NRMSE and NEMAX) for kriging models constructed using different experimental designs for Branin test function

optima, points that are the optima of the surrogate model, but not of the true objective function. Even in the case of non-pathological examples, premature convergence to false optima can occur when only the minimum of the surrogate model is chosen for evaluation by the optimisation algorithm [6]. Instead, the uncertainty of the predictions made by the surrogate model should also be taken into consideration, so that design vectors that have a high uncertainty in their predicted value are also attractive candidates to the optimisation algorithm for sampling, in addition to design vectors with low uncertainty and attractive objective function values.

Various utility functions, which balance the prediction made by a kriging model with the uncertainty in this prediction, have been constructed precisely for this purpose, and are now briefly discussed.

4.2 Expected improvement

Suppose, for a design vector x , a kriging model predicts a value \hat{y} and a root mean squared error s . Let y_{\min} be the minimum objective function value so far sampled. Then, the expected improvement utility function [11] is defined as

$$EIF(I(x)) = \begin{cases} (y_{\min} - \hat{y})\Psi\left(\frac{y_{\min} - \hat{y}}{s}\right) & \text{if } s > 0 \\ + s\psi\left(\frac{y_{\min} - \hat{y}}{s}\right) & \\ 0 & \text{if } s = 0 \end{cases} \quad (10)$$

where Ψ is the standard Gaussian density function, and ψ is the standard Gaussian distribution function. Equation (10) is

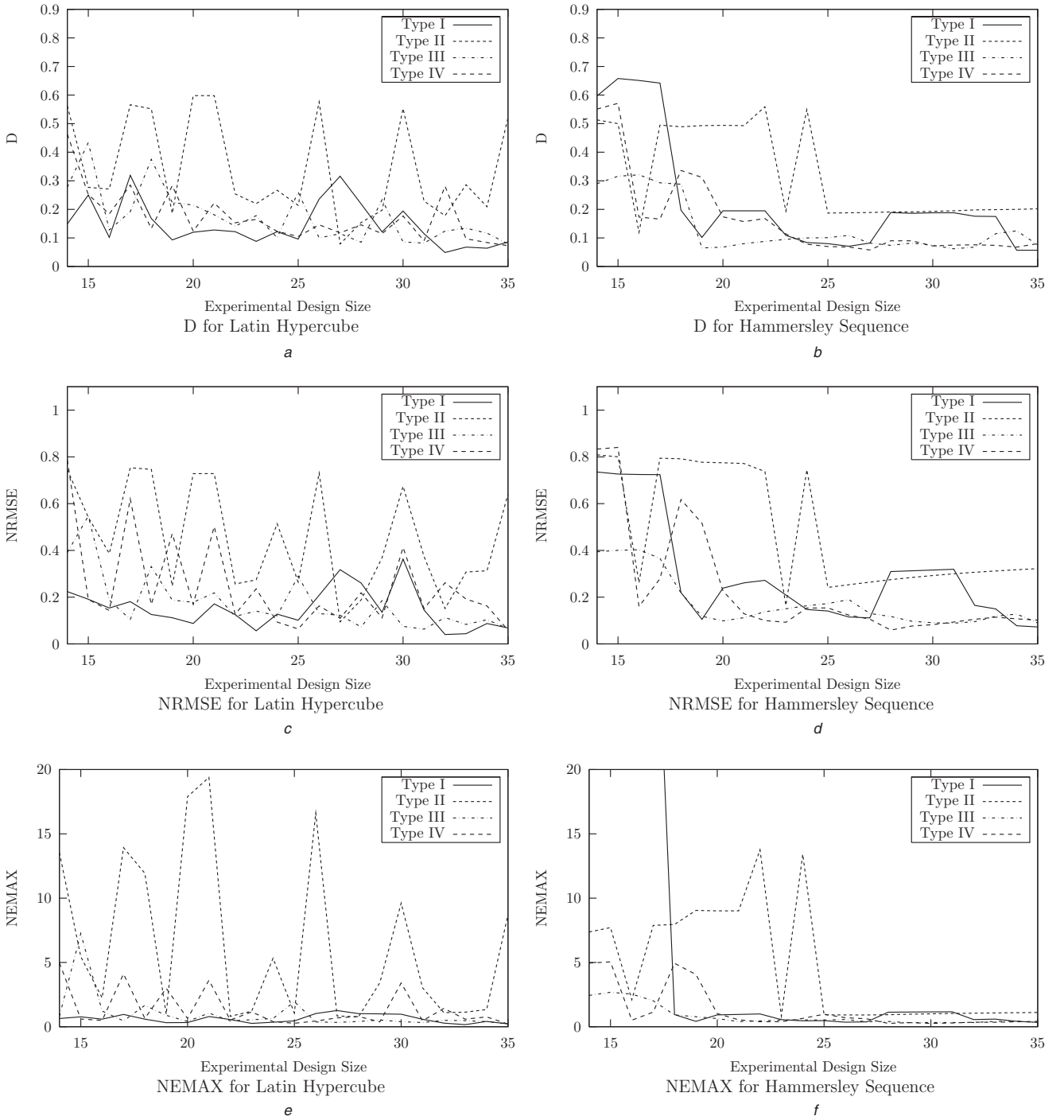


Fig. 5 Accuracy measures (Kolmogorov-Smirnoff statistic D , NRMSE and NEMAX) for kriging models constructed using different experimental designs for Hartman3 test function

composed of two terms: the first places emphasis on a local search around the current minimum, whereas the second places emphasis on design vectors with high uncertainty in their predicted values. Thus the expected improvement is a fixed compromise between a local search around the current minimum and a global search in the regions of design variable space that are of high uncertainty and are possibly modelled less accurately.

4.3 Generalised expected improvement

The expected improvement can be generalised to give what is known as the generalised expected improvement utility

function [22]

$$GEIF[I^g(x)] = s^g \sum_{k=0}^g (-1)^k \left(\frac{g!}{k!(g-k)!} \right) u^{g-k} T_k \quad (11)$$

where

$$T_k = -\phi(u)u^{k-1} + (k-1)T_{k-2} \quad (12)$$

with

$$T_0 = \Phi(u) \quad (13)$$

$$T_1 = -\phi(u) \quad (14)$$

Higher levels of the integer g mean that higher levels of improvement in the current minimum are being sought,

and so more emphasis is placed on searching in areas with more uncertainty in their predicted values. Levels of $g > 2$ place more emphasis on searching regions of high uncertainty than the expected improvement function, whereas $g = 1$ is equivalent to the expected improvement function.

4.4 Weighted expected improvement

The weighted expected improvement utility function uses a weighting parameter w to determine whether emphasis is placed on searching areas of high uncertainty or around the current minimum [23]

$$WEIF(I(x)) = \begin{cases} w(y_{\min} - \hat{y})\Psi\left(\frac{y_{\min} - \hat{y}}{s}\right) \\ + (1 - w)s\psi\left(\frac{y_{\min} - \hat{y}}{s}\right) & \text{if } s > 0 \\ 0 & \text{if } s = 0 \end{cases} \quad (15)$$

Note that values of w less than 0.5 place more emphasis on searching regions of high uncertainty in design variable space than the expected improvement function, whereas values greater than 0.5 place more emphasis on searching around the current minimum. The value of $w = 0.5$ is equivalent to the expected improvement function.

The expected improvement, generalised expected improvement (with $g = 3$) and weighted expected improvement (with $w = 1$) will each be used to investigate how the consideration of uncertainty in the search for an optimum point is affected by the accuracy of the initial kriging model.

4.5 Utility function performance using experimental designs of varying effectiveness

For each test function, four kriging models, two of high accuracy and two of low accuracy, were used with each of the utility functions to locate the optimum to within a certain tolerance. An additional kriging model, constructed using only two design points selected at random, and so naturally of low accuracy, was also used. The results are shown in Tables 1–3.

The models are separated in each table according to their accuracy (with the model constructed from just two points labelled as Basic). The number of iterations taken by each utility function to locate the optimum within the required tolerance is given, with the figures in parentheses giving the total number of function evaluations needed, that is the number of iterations taken by the utility function plus the number of experiments used to construct the kriging model. In each case, the stopping criteria used the relative

tolerance, defined as

relative tolerance

$$= 100 \times \left(\frac{\text{sampled objective function value} - \text{true minimum}}{\text{true minimum}} \right) \quad (16)$$

A 1% relative tolerance was used for the Humps and Hartman3 test functions, and a 2% relative tolerance was used for the Branin test function.

Two main points can be made regarding the results

- (a) Although, in general, fewer iterations may be required when more accurate kriging models are used, this is not always the case.
- (b) Even when fewer iterations are required by the utility functions when more accurate kriging models are used, this can come at a cost of a higher total number of objective function evaluations needed.

The first point can be explained as follows: experimental designs may be found that will yield surrogate models which approximate the true function to a high degree of accuracy; however, unless the true function being modelled is actually known and available for comparison, the true accuracy will not be known, and so the kriging model constructed will have an uncertainty associated with it, no matter how accurate it actually is. It is this uncertainty that the utility function considers when selecting where to sample next. If the true function were available for comparison, so that the true accuracy of the kriging model could be determined, as in this investigation, then, through the construction of a perfectly accurate kriging model, any subsequent considerations of uncertainty would become redundant. Therefore the use of utility functions with highly accurate kriging models does not necessarily yield an advantage over the use of utility functions with less accurate initial kriging models, as the utility function considers there to be uncertainty present in both models. Nevertheless, the overall trend is that fewer iterations are required by utility functions if they are used with kriging models of higher accuracy.

The second point is more obvious. Although utility functions did show an overall trend in locating the optimum of the test functions in fewer iterations, when the kriging model they started with was of higher accuracy, the higher accuracy in the kriging model usually came at the expense of more objective function evaluations being performed in the experimental design stage, and this extra cost was not always justified. Each objective function evaluation in the experimental design stage is only justified if it

Table 1: Number of iterations required by WEIF, EIF and GEIF utility functions to locate global minimum within 1% relative tolerance for Humps test function, using kriging models of varying degrees of accuracy

	Kriging model accuracy			WEIF ($w = 1.0$)	EIF	GEIF ($g = 3$)
	D	NRMSE	NEMAX			
More accurate	0.1592	0.4882	4.8411	4 (18)	2 (16)	3 (17)
	0.1611	0.6836	2.2516	3 (17)	3 (17)	4 (18)
Less accurate	0.5791	0.8810	32.1664	4 (9)	10 (15)	7 (12)
	0.5117	0.7998	23.2252	8 (16)	19 (27)	19 (27)
Basic	0.7803	0.9170	15.4387	8 (10)	8 (10)	>30 (>32)

Table 2: Number of iterations required by WEIF, EIF and GEIF utility functions to locate global minimum within 2% relative tolerance for Branin test function, using kriging models of varying degrees of accuracy

	Kriging model accuracy			WEIF ($w = 1.0$)	EIF	GEIF ($g = 3$)
	D	NRMSE	NEMAX			
More accurate	0.0684	0.2072	3.4690	6 (23)	27 (44)	18 (35)
	0.1133	0.1176	1.6552	>30 (>43)	9 (22)	17 (30)
Less accurate	0.6113	0.6578	3.9200	15 (27)	14 (26)	27 (39)
	0.6328	0.6815	5.6191	13 (21)	21 (29)	30 (38)
Basic	0.5596	0.6885	27.4274	30 (32)	17 (19)	>30 (>32)

Table 3: Number of iterations required by WEIF, EIF and GEIF utility functions to locate global minimum within 1% relative tolerance for Hartman3 test function, using kriging models of varying degrees of accuracy

	Kriging model accuracy			WEIF ($w = 1.0$)	EIF	GEIF ($g = 3$)
	D	NRMSE	NEMAX			
More accurate	0.0800	0.1406	0.4749	1 (26)	1 (26)	2 (27)
	0.1730	0.2101	1.1150	4 (21)	2 (19)	11 (28)
Less accurate	0.6690	0.7579	10.7701	13 (18)	10 (15)	16 (21)
	0.6420	0.7239	39.9308	21 (38)	5 (22)	7 (24)
Basic	0.5650	0.7574	13.1347	>30 (>32)	30 (32)	11 (13)

means that at least one less iteration will be required by the utility function to locate the optimum.

As the overall aim in single-objective optimisation is to locate the global minimum using as few objective function evaluations as possible, the most attractive experimental design is not the design that yields the most accurate kriging model, but instead the design that leads to the construction of a kriging model of such an accuracy that, when the number of additional objective function evaluations needed by the utility function to locate the optimum is added to its size, this number (the total number of objective function evaluations) is minimised. For the three test functions given, this was achieved not by use of an experimental design that yielded an accurate model, but instead by use of designs that yielded models of lower accuracy. In fact, for two of the three test functions, the lowest total number of objective function iterations required to locate the optimum to the required tolerance was achieved with a trivial experimental design of just two random points. When the overall aim is to locate the optimum of a function using a surrogate model in as few iterations as possible, it would appear that it is advantageous to start the iterative search early with a relatively inaccurate model.

5 Conclusions

Surrogate-assisted optimisation can be viewed as consisting of two stages: construction of the initial surrogate model through off-line learning, followed by an iterative optimisation search with on-line learning. Typically, an experimental design is used to construct the initial kriging model, with the aim of yielding the most accurate surrogate model possible, and utility functions with in-built uncertainty considerations are then used to search for the optimum.

The performances of several utility functions were investigated on three test functions, using experimental designs that yielded kriging models of high accuracy, and they were compared with the performances when experimental designs were used that yielded kriging models of lower accuracy. When the aim is to reduce the overall number

of objective function evaluations required to locate the optimum (which is normally the case in practical design optimisation), it was found that no advantage is necessarily achieved by the use of experimental designs that yield more accurate surrogate models, and that a saving in total iterations can even sometimes be made if the utility function is allowed to build up the model entirely by itself as it searches for the optimum through on-line learning. For scenarios where objective functions are expensive to compute, such as in electromagnetic design, such a reduction in objective function evaluations is highly desirable.

6 References

- Canova, A., Freschi, F., and Repetto, M.: 'Hybrid method coupling AIS and zeroth order deterministic search', *COMPEL Int. J. Comput. Math. Electr. Electron. Eng.*, 2005, **24**, (3), pp. 784–795
- Di Barba, P., Farina, M., and Savini, A.: 'Multiobjective design optimization of real-life devices in electrical engineering: a cost-effective approach'. (Lecture Notes in Computer Science, (Springer, 2001), pp. 560–573
- Aslett, R., Buck, R.J., Duvall, S.G., Sacks, J., and Welch, W.J.: 'Circuit optimization via sequential computer experiments: design of an output buffer', *Appl. Stat.*, 1998, **47**, pp. 31–48
- Queipo, N.E., Haftka, R.T., Shyy, W., Goel, T., Vaidyanathan, R., and Tucker, P.K.: 'Surrogate-based analysis and optimization', *Progr. Aerosp. Sci.*, 2005, **41**, pp. 1–28
- Lebensztajn, L., Marretto, C.A.R., Costa, M.C., and Coulomb, J.-L.: 'Kriging: a useful tool for electromagnetic devices optimization', *IEEE Trans. Magn.*, 2004, **40**, (2), pp. 1196–1199
- Jones, D.R.: 'A taxonomy of global optimization methods based on response surfaces', *J. Glob. Optim.*, 1998, **13**, pp. 455–492
- Sykulski, J.K., Al-Khoury, A.H., and Goddard, K.F.: 'Minimal function calls approach with on-line learning and dynamic weighting for computationally intensive design optimization', *IEEE Trans. Magn.*, 2001, **37**, (5), pp. 3423–3426
- Haykin, S.: 'Neural networks: a comprehensive foundation' (Pearson, 1999)
- Coulomb, J.-L., Kobetski, A., Costa, M.C., Marechal, Y., and Jonsson, U.: 'Comparison of radial basis function approximation techniques', *COMPEL Int. J. Comput. Math. Electr. Electron. Eng.*, 2003, **22**, (3), pp. 616–629
- Farina, M., and Sykulski, J.K.: 'Comparative study of evolution strategies combined with approximation techniques for practical

- electromagnetic optimization problems', *IEEE Trans. Magn.*, 2001, **37**, (5), pp. 3216–3220
- 11 Jones, D.R., Schonlau, M., and Welch, W.J.: 'Efficient global optimization of expensive black-box functions', *J. Glob. Optim.*, 1998, **13**, pp. 455–492
 - 12 Siah, E.S., Sasena, M., Volakis, J.L., Papalambros, P.Y., and Wiese, R.W.: 'Fast parameter optimization of large-scale electromagnetic objects using DIRECT with kriging metamodeling', *IEEE Trans. Microw. Theory Tech.*, 2004, **52**, (1), pp. 276–285
 - 13 Montgomery, D.: 'Design and analysis of experiments' (John Wiley and Sons, 2001)
 - 14 Kleijnen, J.P.C., Sanchez, S.M., Lucas, T.W., and Cioppa, T.M.: 'A user's guide to the brave new world of designing simulation experiments', *INFORMS J. Comput.*, 2005, **17**, (3), pp. 263–289
 - 15 Liu, L., and Wakeland, W.: 'Does more uniformly distributed sampling generally lead to more accurate prediction in computer experiments?'. Proc. 2005 Winter Simulation Conf. 2005, Kuhl, M.E., Steiger, N.M., Armstrong, F.B., and Joines, J.A. (Eds.), pp. 2561–2571
 - 16 McKay, M.D., Conover, W.J., and Beckman, R.J.: 'A comparison of three methods for selecting values of input variables in the analysis of output from a computer code', *Technometrics*, 1979, **21**, pp. 239–245
 - 17 Kalagnanam, J.R., and Diwekar, U.M.: 'An efficient sampling technique for off-line quality control', *Technometrics*, 1997, **39**, (3), pp. 308–319
 - 18 Wilkinson, J.H.: 'Rounding errors in algebraic processes' (Prentice-Hall, Englewood Cliffs, NJ, 1963)
 - 19 'Matlab, computation High-performance numeric guide visualization software. Reference Guide' (The Math Works Inc., Natick, MA, 2004)
 - 20 Dixon, L.C.W., and Szego, G.P.: 'The global optimisation problem: an introduction' in Dixon, L.C.W., and Szego, G.P. (Eds.): 'Towards global optimisation' (North-Holland, Amsterdam, 1978), Vol. 20, pp. 1–15
 - 21 Hartman, J.K. Tech. Rept. NP55HH72051A, Naval Postgraduate School, Monterey, CA, 1972
 - 22 Schonlau, M., Welch, W.J., and Jones, D.R.: 'Global versus local search in constrained optimization of computer models', *IMS Lecture Notes*, 1998, **34**, pp. 11–25
 - 23 Sobester, A., Leary, S.J., and Keane, A.J.: 'On the design of optimization strategies based on global response surface approximation models', *J. Glob. Optim.*, 2005, **33**, pp. 31–59