A Dual Kriging Approach with Improved Points Selection Algorithm for Memory Efficient Surrogate Optimisation in Electromagnetics

Yinjiang Li, Song Xiao, Mihai Rotaru, and Jan K. Sykulski, *Fellow, IEEE*

Electronics and Computer Science, University of Southampton, Southampton, United Kingdom, jks@soton.ac.uk

The paper introduces a new approach to kriging surrogate model sampling points allocation. By introducing a second (dual) kriging during the model construction process the existing sampling points are reallocated to reduce overall memory requirements. Moreover, a new algorithm is suggested for selecting the position of the next sampling point by utilising a modified Expected Improvement criterion.

***Index Terms*—Kriging, global optimisation, surrogate modelling, large datasets.**

I. Introduction

KRIGING offers significant advantages in computationally expensive optimisation as a number of necessary objective function calls may be reduced. This matters in particular when each call entails time consuming simulations, such as frequently encountered in the design of electromagnetic devices [1]. Large data sets, however, tend to be produced by correlation matrices which arise when kriging models are produced and the amount of storage required is usually proportional to n^2 [2], where n is the number of sample points; this problem may become acute in the case of multi-parameter optimisation when performed on computers with limited memory [3], [4].

II. A Modified EI Sampling Criterion

Expected Improvement (EI) [5] is commonly used to guide the process of selecting the next point for evaluation (often with modifications [1]). The challenge is to balance exploitation and exploration in order to avoid the kriging model being trapped in a local optimum; moreover, the quality of the kriging prediction of the shape of the objective function may also be important in the context of the robustness of the design. In this paper we suggest a modification to the standard EI criterion with the aim to spread the 'infill' (new sampling) points more efficiently throughout the design space. Consider a simple illustration in Fig. 1 where the dotted line is the actual objective function. The range has been normalised between 0 and 1 while the values of the objective function have no actual meaning in this example.

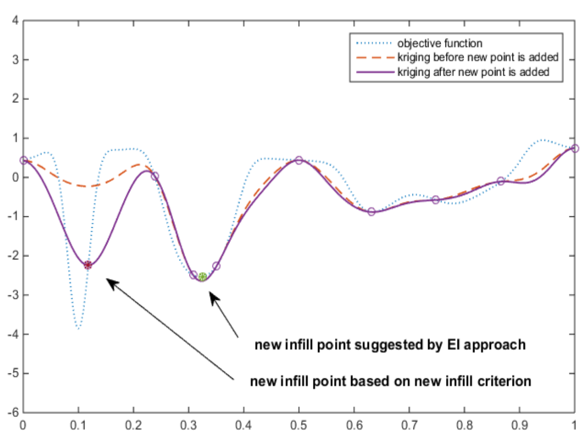

Fig. 1. The kriging model before and after a new point is added.

The proposed sampling criterion calculates EI while taking the estimated error (the 'Mean Square Error' MSE [1]) between known sampling points into consideration

$$\begin{aligned} &\mathit{sampling\ criterion} \\ &\quad = \max\{EI\} \times \mathrm{MSE} \times \mathit{weight} + \max\{\mathrm{MSE}\} \times EI \end{aligned} \tag{1}$$

whereas scaling has been applied to account for different values of components and thus normalise the result. Moreover, a weight is added to the estimated error. The estimated error is provided by the kriging predictor together with the predicted value at any given point.

The weight term is the ratio of the exponentially weighted standard deviation (between infill points and their previously predicted values) average and the uniformly weighted standard deviation average. This value decreases as optimisation process continues and the model quality increases. The exponentially weighted standard deviation average at the current iteration is calculated using a formula

$$ESDA = \frac{stdev_1 + (1-\alpha)stdev_2 + (1-\alpha)^2 stdev_3 + \cdots}{1 + (1-\alpha) + (1-\alpha)^2 + \cdots} \tag{2}$$

where α determines the weight on each standard deviation term, and $0 < \alpha < 1$; $stdev_1$ is the current standard deviation term.

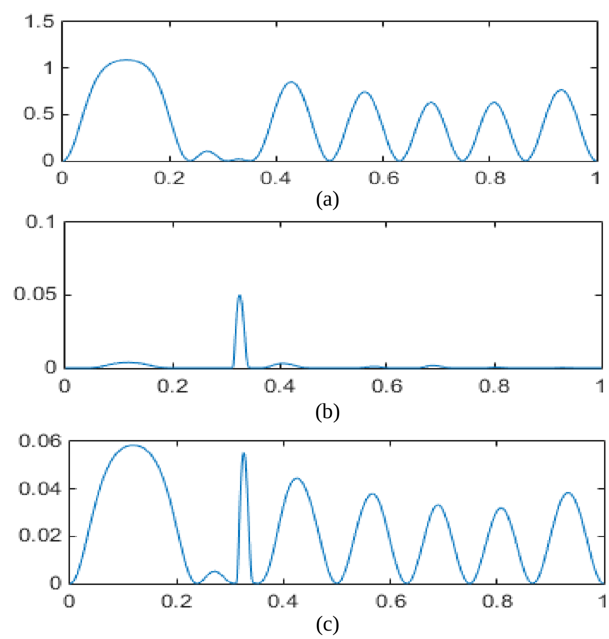

Fig. 2. The distributions of: (a) MSE, (b) Expected Improvement, and (c) the resultant sampling criterion for the example test function.

The classical Expected Improvement criterion itself would advocate exploitation of the area close to the recently simulated minimum (which in reality is only a local minimum) and create a new point at x=0.347 as shown in Fig. 1, whereas this could be counterproductive during the exploration stage. The modified criterion, however, proposes more exploration and positions the new sampling point elsewhere with better chance of capturing the global minimum, as illustrated. Moreover, it is important in the context of robust optimisation to predict not just the position of the optimum but also the shape of the objective function near the optimum.

III. A Dual Kriging Approach

The main drawback of the kriging approach is the need for creating correlation matrices which, especially in the case of multi-parameter problems, may become very large and thus need to be handled efficiently. This has been pointed out in our recent publication [4] and a possible solution was offered. Here we suggest a different (or complementary) approach resulting from an argument that once the surrogate model is advanced and the shape of the objective function is reasonably accurately predicted we really only need some sampling points, especially those close to the areas considered as potentially attractive. Thus as the total number of sampling points increases, and we are getting to the memory limit of the computer, in order to avoid computationally time consuming 'memory management' (e.g. page swapping) we may instead 'remove' some of the less attractive points in an attempt to keep the total number of points constant, or increasing slowly, while the removed points may be used to create a 'dual' kriging model. At the same time it should be noted that – as shown in Fig. 3 – the memory saving is biggest when we operate at roughly between 20% and 30% of reduced number of points (shaded area). For example, at 20% (that is the original kriging model preserving 80% of points) the memory saving is 32%.

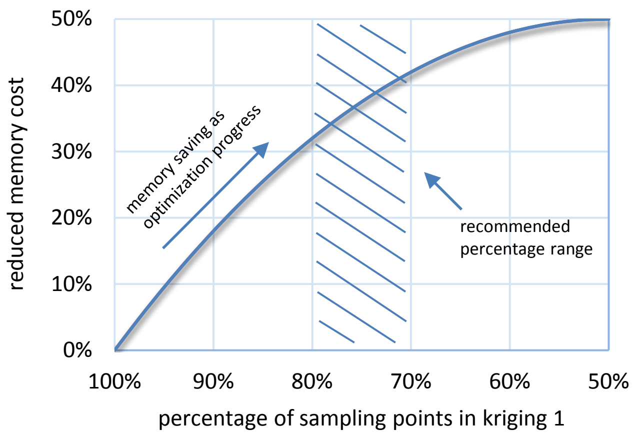

Fig. 3. Efficiency of sampling points allocation.

Consider the test function of Fig. 1 after 13 iterations and its kriging prediction as shown in Fig. 4 (note that iterations have not completed yet). The criterion for removal of a point is related to the triangular area formed by connecting this point and the two neighbours. Then the points with the smallest associated area are removed and in fact used to create a dual kriging model. The modified model now contains fewer but more important points, as illustrated by Fig. 5. In the example a saving of 49% of memory has been achieved.

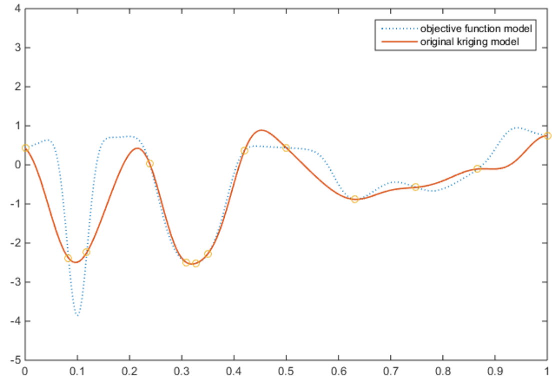

Fig. 4. A single kriging model with 13 sampling points (iterations incomplete).

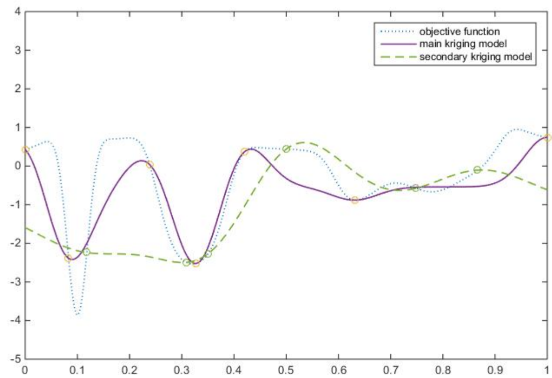

Fig. 5. The main kriging model (7 sampling points) and the secondary kriging model (6 sampling points).

The two modifications to the kriging surrogate modelling put forward in this paper address both issues of better prediction of the shape of the objective function (important for robust design) and achieving improved efficiency in handling correlation matrices for larger problems where the available computer memory may be limited. In the full version the ways of using the dual kriging model will be explained (as it preserves some useful information) and the algorithm will be illustrated using a practical electromagnetic design problem.

References

[1] S. Xiao, M. Rotaru, and J. K. Sykulski, "Adaptive Weighted Expected Improvement with Rewards Approach in Kriging Assisted Electromagnetic Design," *IEEE Trans. Magn.*, vol. 49, no. 5, pp. 2057–2060, 2013.

[2] T. Romary, "Incomplete Cholesky decomposition for the kriging of large datasets," *GeoENV 2012*, Spain, p. 259, 2012.

[3] N. Cressie and G. Johannesson, "Fixed rank kriging for very large spatial data sets," *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, vol. 70, no. 1, pp. 209–226, 2007.

[4] S. Xiao, M. Rotaru, and J. K. Sykulski, "Correlation matrices in kriging assisted optimisation of electromagnetic devices," *IET Science, Measurement and Technology*, doi: 10.1049/iet-smt.2014.0194, pp. 1–8, 2014.

[5] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient Global Optimization of Expensive Black-Box Functions," *Journal of Global Optimization*, vol. 13, pp. 455–492, 1998.