A Dual Kriging Approach With Improved Points Selection Algorithm for Memory Efficient Surrogate Optimization in Electromagnetics

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

Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K.

This paper introduces a new approach to kriging surrogate model sampling points allocation. By introducing the second (dual) kriging during the model construction, the existing sampling points are reallocated to reduce overall memory requirements. Moreover, a new algorithm is proposed for selecting the position of the next sampling point by utilizing a modified expected improvement criterion.

***Index Terms*— Global optimization, kriging, large data sets, surrogate modeling.**

I. INTRODUCTION

KRIGING offers significant advantages in computationally costly optimization by reducing the number of objective function calls. This matters, in particular, when each call entails time-consuming simulations, as 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; the amount of storage required is usually proportional to n^2 [2], where n is the number of sample points. This may become an issue in multi-parameter optimization performed on smaller computers with limited memory [3], [4]. In this paper, we look at an improved algorithm for selecting a point for evaluation and more efficient memory management. A simple criterion is proposed for removing some points without losing important information, and the concept of dual kriging is utilized. Advantages of the approach are considered.

II. MODIFIED EI SAMPLING CRITERION

Expected improvement (EI) [5] is commonly used to guide the process of selecting the next sampling point [1]. The challenge is to balance exploitation and exploration in order to avoid the model being trapped in a local optimum; moreover, the quality of the kriging prediction of the objective function's shape may be important in the context of the robustness of the design. A modification to the standard EI criterion is proposed to spread the infill (new sampling) points more efficiently.

Consider a simple case of Fig. 1 with the objective function shown by the dotted line and the range normalized between 0 and 1; the values of the function have no actual meaning in this example. The proposed sampling criterion calculates EI while taking the estimated error (the mean square error (MSE) [1]) between known sampling points into consideration

$$\text{sampling criterion} = \max\{\text{EI}\} \times \text{MSE} \times \text{weight} + \max\{\text{MSE}\} \times \text{EI} \quad (1)$$

and scaling applied to account for different component values and thus to normalize the results. Moreover, a weight is added to the estimated error provided by the kriging predictor together with the predicted value at any given point.

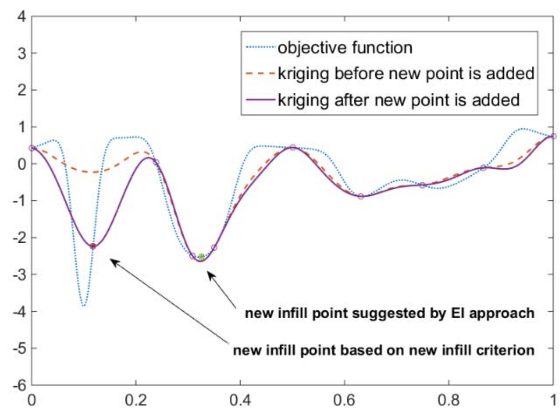

Fig. 1. Kriging model before and after a new point has been added.

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; its value decreases as the optimization process continues and the model quality improves. This average at the current iteration is calculated using a formula

$$\text{ESDA} = \frac{\text{sd}_1 + (1-\alpha)\text{sd}_2 + (1-\alpha)^2\text{sd}_3 + \cdots}{1 + (1-\alpha) + (1-\alpha)^2 + \cdots} \quad (2)$$

where α controls the weight on each standard deviation term, and $0 < \alpha < 1$; sd is the current standard deviation at a point.

The classical EI criterion itself would advocate exploitation of the area close to the recently simulated minimum (which in reality is a local minimum) and create a new point at $x = 0.347$ (see Fig. 1), whereas this could be counterproductive during the exploration stage. The modified criterion, however, proposes more exploration and places the new sampling point where a better chance of capturing the global minimum exists, as shown in Fig. 2. Moreover, it is important in the context of robust optimization to predict not just the position of the optimum but also the shape of the function near the optimum.

III. DUAL KRIGING APPROACH

The main drawback of the kriging approach is the need for correlation matrices, which may become very large and must be handled carefully, as discussed in [4], where a possible

Manuscript received June 26, 2015; revised September 4, 2015; accepted September 30, 2015. Date of publication October 5, 2015; date of current version February 17, 2016. Corresponding author: J. K. Sykulski (e-mail: jks@soton.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMAG.2015.2486522

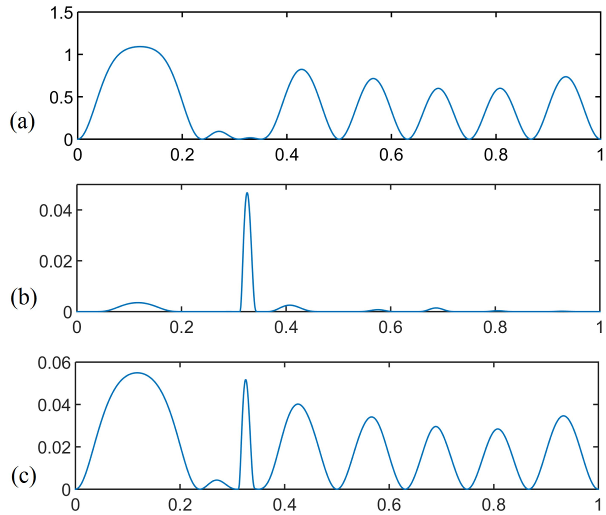

Fig. 2. Distributions of (a) MSE, (b) EI, and (c) resultant sampling criterion for the example test function.

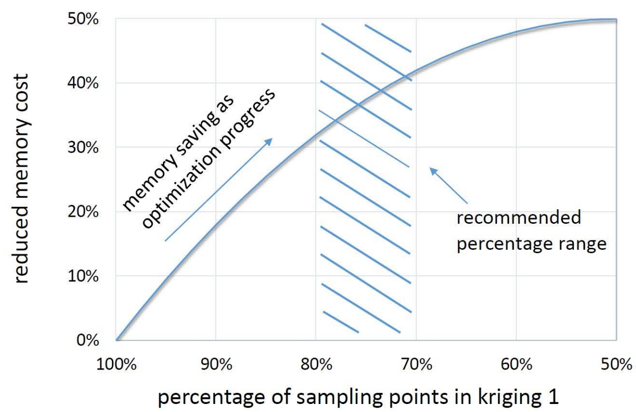

Fig. 3. Efficiency of sampling points allocation.

solution was offered. Here, we suggest a 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 need only 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 steady, or increasing slowly, while the removed points may be used to create a dual kriging model. A tradeoff exists between the gain of memory saving due to removed points and the overall saving of memory (Fig. 3). Operating at 20% to 30% of reduced number of points would retain a relative high gain of memory saving while having a large reduction on overall memory usage (32% and 42% overall memory saving at 20% and 30%, respectively).

The simple criterion for removing a point is related to the triangular area formed by connecting this point and the two neighbors, as shown in Fig. 4 (here, $x = 0.75$ can be removed). Thus, the points with the smallest associated area are removed and may be used to create a dual kriging model. The current optimum is assigned a large area and will not be removed.

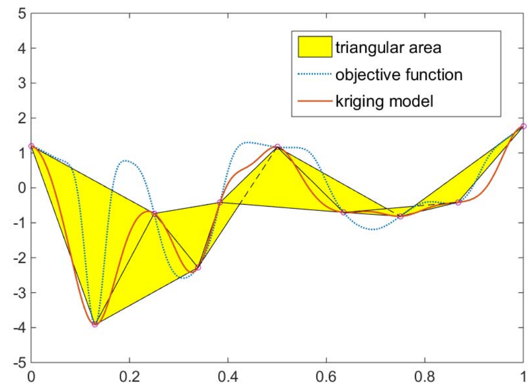

Fig. 4. Point removal criterion plotted as shaded triangular areas.

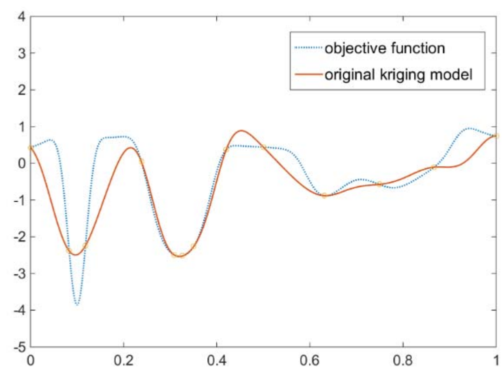

Fig. 5. Single kriging model with 13 sampling points (iterations have not completed yet).

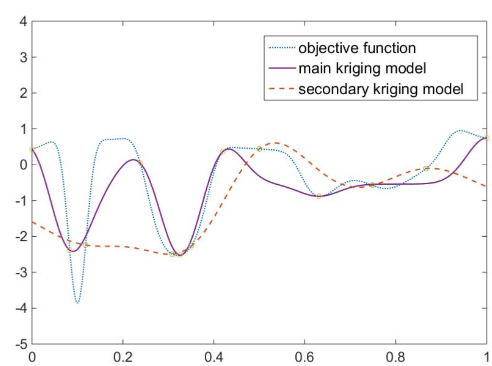

Fig. 6. Main kriging model (seven sampling points) and the secondary kriging model (six sampling points).

Consider the test function of Fig. 1 after 13 iterations and its kriging prediction, as shown in Fig. 5 (note that the iterations have not completed yet). At this stage, six points have been removed from the main kriging to create the secondary kriging. The modified model contains fewer but more relevant points, as shown in Figs. 5 and 6. In the example, a saving of 49% in memory has been achieved by reducing the correlation matrix.

IV. EI Sampling Criterion Aided by Dual Kriging

While applying dual kriging can potentially save computer memory by as much as 50% compared with the single kriging, for the same number of sampling points, its major challenge lies in the selection of the location of the new infill point. If only the main kriging model was to be considered, there would be a risk of a new location to be at or close to

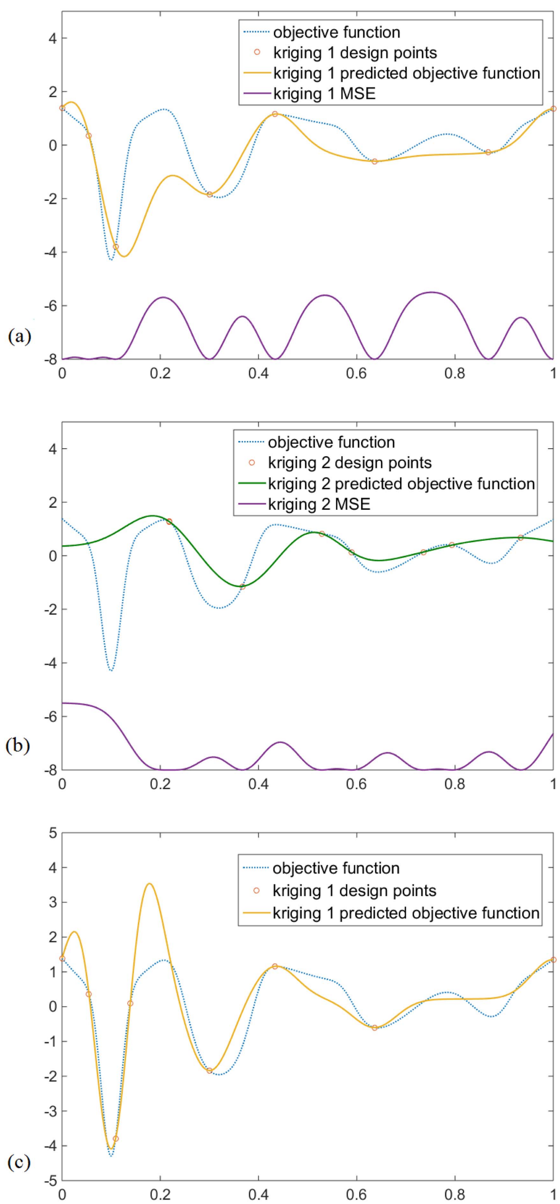

Fig. 7. (a) Main kriging and MSE. (b) Dual kriging and MSE. (c) Predicted function after a new infill point is added and the old point removed.

a point that has just been removed. In order to inhibit such an unwelcome scenario, both the kriging models and thus a modified MSE defined by a product of the principal and dual MSEs are used; consequently, the EI uses all points and no information is lost. This is illustrated in Fig. 7 showing the 15th iteration, with the point adding/removal process triggered after the eighth point had been added; thus, there are eight points in the main (1) and seven in the dual (2) kriging models. The modified sampling criterion based on the combined MSE offers a better prediction than the original single kriging approach.

V. More on the Selection of Infill Points

The use of EI for selecting infill points is widely accepted in surrogate modeling, despite some drawbacks as reported, for example, in [12]; further modifications and improvements have been suggested. But there is one particular difficulty rarely mentioned but worth emphasizing. In order to reliably locate the point with maximum EI, all points within the design space need to be considered, requiring a calculation of the predicted objective function and the corresponding MSE at all points. This may seem straightforward for 1-D problems, but could create another instance of combinatorial explosion in more practical multi-dimensional problems. Take four dimensions, for example, and a step of 1/100 in each direction: this would require 10^8 calculations to find the maximum EI, which even for very fast surrogate models would create a computational challenge. A large step of say 1/10 would be easy to handle but unlikely to capture the actual optimum. In Fig. 8, a case is depicted where even a relatively small step of 1/50 might result in the algorithm missing the largest EI and hence the global optimum at $x = 0.131$ would be overlooked.

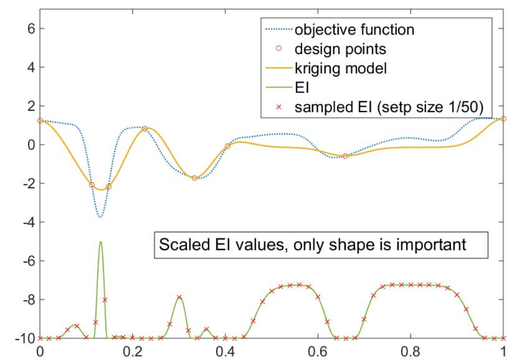

Fig. 8. Illustration of the limitation of using a fixed step size.

As an alternative to an exhaustive search using a predefined step size, the task of finding the maximum sampling criterion value could be treated as a small optimization problem in its own right. Any optimization method could be applied for that purpose and would effectively eradicate the step size problem mentioned above. A genetic algorithm has been utilized to search for the infill criterion function in Section VI.

Finally, the algorithms described so far can be augmented by local routines, e.g., a new point could be added between any two adjacent existing points and the location of the maximum EI estimated using a local gradient-based method.

VI. Team Problem 25

To illustrate the proposed optimization methodology in the context of electromagnetic design, the TEAM problem 25 has been studied [13], which is a die press with an electromagnet for the orientation of magnetic powder, used to produce an anisotropic permanent magnet. The aim is to optimize the shape to minimize a particular objective as specified in the problem description. Each function evaluation requires a full finite-element (FE) solution of a non-linear problem, which is computationally inefficient if used in combination with any optimization method, especially if extensive design space exploration is required. This is, therefore, a very appropriate practical case to be studied.

In Fig. 9, $R1$, $L2$, $L3$, and $L4$ are the design parameters to be optimized so that the objective function W is minimized

$$W = \sum_{i=1}^{n}\{(B_{\mathrm{xip}} - B_{\mathrm{xio}})^2 + (B_{\mathrm{yip}} - B_{\mathrm{yio}})^2\} \tag{3}$$

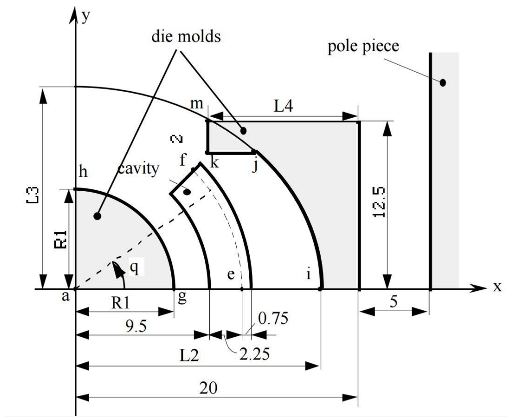

Fig. 9. Model of the die press with an electromagnet [13].

TABLE I

PARAMETER CONSTRAINTS

Variable	R1	L2	L3	L4
Lower-boundary (mm)	5	12.6	14	4
Upper-boundary (mm)	9.4	18	45	19

TABLE II

COMPARISON OF PERFORMANCE

Algorithms	Optimal value (10^{-4})	R1 (mm)	L2 (mm)	L3 (mm)	L4 (mm)	No of function calls
GA	2.686	7.2996	14.174	14.001	14.326	3421
SA	1.622	7.2252	14.322	14.110	14.306	2145
HuTS	0.500	7.3780	14.613	14.371	14.204	1580
UTS	1.050	7.5487	14.908	14.506	14.416	931
NTS	0.648	7.4337	14.732	14.428	14.237	575
Kriging EI	0.452	7.2	14.1	14	14.5	265
Kriging AWEI	0.412	7.2	14	14	14.5	214
Dual kriging	0.323	7.007	13.891	14.035	14.270	242

Genetic algorithm GA [6]; Stimulated Annealing SA [7]; Tabu Search HuTS [8]; Universal Tabu search [9]; New Tabu Search NTS [10]; Kriging EI and Kriging AWEI (with specified step size) [11];

where B_x and B_y are the x- and y-components of magnetic flux density at points along the curve ef, while subscripts p and o denote the calculated and desired values, respectively. The constraints are listed in Table I and results in Table II.

The first five results in Table II are taken from the literature, the two kriging values are from our previous publications, and the dual kriging approach is shown in the bottom row. In general, kriging outperforms other methods in terms of finding the optimum, but mainly because of savings in computing times, measured in the number of required FE calculations. The dual kriging marginally required more iterations, but has produced a slightly better result. Without considering the numerical errors due to the different electromagnetic solvers, all the kriging models are similar and superior to other algorithms. The different numerical methods used will only have a marginal effect on the total number of function calls and it may be concluded that the kriging will almost always be computationally more efficient.

The dual kriging algorithm was triggered after 120 FE calls, with one sampling point removed, when a new one was added

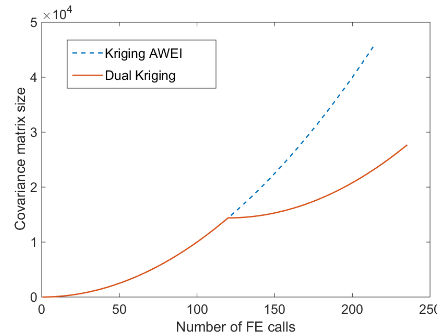

Fig. 10. Memory requirements for the covariance matrix at each iteration.

in subsequent iterations. The sizes of the covariance matrix in the kriging model for the standard AWEI [11] routine and the new dual algorithm are compared in Fig. 10 as optimization progresses. Memory savings on this occasion reached 36%.

VII. CONCLUSION

A modified infill sampling criterion is proposed and can be applied to both the single and dual kriging methods. A simple but efficient automatic exploration versus exploitation adjustment based on model quality is developed. The limitations of an exhaustive search are noted, and a global optimization-aided EI search algorithm was introduced. The dual kriging method is verified against a test function and an electromagnetic TEAM problem 25, with a 36% saving in the covariance matrix size.

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, May 2013.

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

[3] N. Cressie and G. Johannesson, "Fixed rank kriging for very large spatial data sets," *J. Roy. Statist. Soc. B (Statist. Methodol.)*, 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 Sci., Meas. Technol.*, vol. 9, no. 2, pp. 189–196, 2015.

[5] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *J. Global Optim.*, vol. 13, no. 4, pp. 455–492, 1998.

[6] J. H. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis With Applications to Biology, Control, and Artificial Intelligence*. Cambridge, MA, USA: MIT Press, 1975.

[7] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[8] N. Hu, "Tabu search method with random moves for globally optimal design," *Int. J. Numer. Methods Eng.*, vol. 35, no. 5, pp. 1055–1070, 1992.

[9] T. J. Santner, B. J. Williams, and W. I. Notz, *The Design and Analysis of Computer Experiments*. New York, NY, USA: Springer-Verlag, 2003.

[10] O. Hajji, S. Brisset, and P. Brochet, "A new Tabu search method for optimization with continuous parameters," *IEEE Trans. Magn.*, vol. 40, no. 2, pp. 1184–1187, Mar. 2004.

[11] S. Xiao, M. Rotaru, and J. K. Sykulski, "Six sigma quality approach to robust optimization," *IEEE Trans. Magn.*, vol. 51, no. 3, Mar. 2015, Art. ID 7201304.

[12] A. I. J. Forrester and D. R. Jones, "Global optimization of deceptive functions with sparse sampling," in *Proc. 12th AIAA/ISSMO Multidiscipl. Anal. Optim. Conf.*, Victoria, BC, Canada, Sep. 2008, pp. 1–15.

[13] N. Takahashi, K. Ebihara, K. Yoshida, T. Nakata, K. Ohashi, and K. Miyata, "Investigation of simulated annealing method and its application to optimal design of die mold for orientation of magnetic powder," *IEEE Trans. Magn.*, vol. 32, no. 3, pp. 1210–1213, May 1996.