Strategies for balancing exploration and exploitation in electromagnetic optimisation

Song Xiao, Mihai Rotaru and Jan K. Sykulski
Electronics and Computer Science (ECS), University of Southampton, Southampton, UK

Abstract

Purpose – Electromagnetic design utilising finite element or similar numerical methods is computationally expensive, thus efficient algorithms reducing the number of objective function calls to locate the optimum are sought. The balance between exploration and exploitation may be achieved using a reinforcement learning approach, as demonstrated previously. However, in practical design problems, in addition to finding the global optimum efficiently, information about the robustness of the solution may also be important. In this paper, the aim is to discuss the suitability of different search algorithms and to present their fitness to solve the optimization problem in conjunction with providing enough information on the robustness of the solution.

Design/methodology/approach – Two novel strategies enhanced by the surrogate model based weighted expected improvement approach are discussed. The algorithms are tested using a two-variable test function. The emphasis of these strategies is on accurate approximation of the shape of the objective function to accomplish a robust design.

Findings – The two novel strategies aim to pursue the optimal value of weights for exploration and exploitation throughout the iterative process for better prediction of the shape of the objective function.

Originality/value – It is argued that the proposed strategies based on adaptively tuning weights perform better in predicting the shape of the objective function. Good accuracy of predicting the shape of the objective function is crucial for achieving a robust design.

Keywords Optimisation, Kriging, Surrogate modelling, Robust design, Electromagnetics, Electromagnetic devices, Reinforcement learning, Optimization techniques, Electromagnetism

Paper type Research paper

1. Introduction

Design problems in electromagnetic devices are commonly solved using time-consuming numerical techniques, such as the finite element method. In order to relieve the heavy burden of computation in such designs, kriging has been suggested as one of the reliable surrogate models with low computational cost and good accuracy of predicting the shape of the objective function. In the optimisation task, the main target is using as few "expensive" objective function calls as possible to find the global optimum. The balance between exploration (searching the region with high uncertainty) and exploitation (searching the highly confident space) has been discussed before (Lebensztajn *et al.*, 2004; Jones *et al.*, 1998; Sobester *et al.*, 2005; Xiao *et al.*, 2012). This paper puts main emphasis on improving the existing strategies to predict the shape of the objective function as accurately as possible – in addition to locating the global optimum – in order to assess the robustness of the solution.

COMPEL: The International Journal for Computation and Mathematics in Electrical and Electronic Engineering
Vol. 32 No. 4, 2013
pp. 1176-1188

0332-1649
DOI 10.1108/03321641311317004

2. Kriging with different strategies

2.1 Kriging and different strategies

Kriging (Lebensztajn *et al.*, 2004) can exploit the spatial correlation of data in order to predict the shape of the objective function based only on limited information and estimates the accuracy of this prediction, which is helpful in assisting the main decision of the optimisation process in how to choose the next design vector for evaluation. In general, an estimate of the accuracy (called the potential error) by the kriging model is commonly used to build a range of different "utility functions" such as the expected improvement (EI) (Jones *et al.*, 1998), or weighted expected improvement (WEI) (Sobester *et al.*, 2005). The EI function is defined as:

$$\text{if} \quad s(x) > 0 \quad EI = (f_{\min} - \hat{y}(x))\Phi\left(\frac{f_{\min} - \hat{y}(x)}{s(x)}\right) + s(x)\phi\left(\frac{f_{\min} - \hat{y}(x)}{s(x)}\right) \tag{1}$$

$$\text{if} \quad s(x) = 0 \quad EI = 0 \tag{2}$$

where $\hat{y}(x)$ is the predicted value of objective function by the kriging model, and $s(x)$ is the root mean squared error in this prediction. The first term, called the Gaussian density, favours searching promising regions, whereas the second term (Gaussian distribution function) is related to exploration, which favours searching regions with high uncertainty. Finding the global optimum of the objective function is one of the significant aims for an optimization problem. In practical experiments, the exploration term performs dramatically better in terms of finding the global optimum of the objective function, while the exploitation often can only find the local minimum. Since EI applies equal weights on the two terms, it may be seen as a fixed compromise between exploration and exploitation. The WEI is derived from the EI by adding a tuneable parameter which can adjust the weights on exploration and exploitation.

As suggested by previous tests (Xiao *et al.*, 2012), the optimal choice of the weights is critical in terms of the ability of the algorithm to achieve the global optimum and doing it efficiently; unfortunately the optimal weights are normally hard to find and require numerous tests. Therefore, two novel algorithms using reinforcement learning (Sutton and Barto, 1998) called Adaptive Weighted Expected Improvement (AWEI) and Surrogate Model based WEI approach with rewards (Sykulski *et al.*, 2010) (SMWEI) (Xiao *et al.*, 2012) have been proposed to make the process of tuning weights more intelligent and self-guiding.

The mean square error (MSE) from the kriging model is used to calculate the rewards. The AWEI algorithm can tune the weights automatically based on the comparison between the potential rewards from two different weight distributions emphasising exploitation and exploration, respectively. After comparison, the weights are redistributed on the two terms of (1) to encourage exploration or exploitation depending on the results of the initial pre-test. However, AWEI only takes account of the short-term rewards at a given iteration step, whereas SMWEI can predict the cumulative rewards likely to occur in the long-term as a consequence of a particular choice of actions. Furthermore, SMWEI creates a surrogate model based on potential error and kriging prediction to use in a pre-test, rather than using the information from the time-consuming finite element modelling software. In the pre-test, two distinct weights are used – one favouring exploration and the other one exploitation – and iterations continue using the surrogate model independently in parallel until overall rewards have been found.

The optional weight with better reward of the two is then used to feed back – via the FEM module – into the main iterative loop of the design process.

2.2 SMWEI with multi-weights in the pre-test

In practical electromagnetic problems, the robustness of the design is a significant requirement that needs to be considered. Through testing it has been found that the SMWEI algorithm with certain pairs of weights in the pre-test performs better in terms of estimating the shape of the objective function, a feature which might be helpful when assessing the robustness of the design. As SMWEI is limited by the pre-set pair of weights in the pre-test, a number of experiments may be necessary to find the pair resulting in more faithful representation of the shape. As the pre-test is "cheap", more weights can be selected to broaden the base for comparisons. The new version of SMWEI with multi-weights is described in Figure 1. Because there is only one pair of weights (one emphasising exploration and the other emphasising exploitation) provided in the pre-test of SMWEI described previously in (Xiao *et al.*, 2012), if one of the rewards is not assessed properly or fails the pre-test for some other reason, there is only one "back-up" action available, that is apply EI rather than continuing comparing the rewards produced by the two weights. Thus, applying more optional weights in the pre-test will allow the comparisons to continue, even if some combinations of weights may fail in the pre-test.

2.3 SMWEI with the strategy of adaptively tuning weights

In the pre-test of SMWEI, a pair of fixed weights (one emphasising exploration and the other one exploitation) needs to be set initially; the guidelines for how to select such weights are subject to further experiments. However, the strategy of tuning the weights automatically and adaptively in the pre-test of AWEI can also be used in SMWEI in order to avoid the need for setting initial optional weights. The decision-making chart of the actual implementation is shown in Figure 1.

Because all pre-tests in SMWEI apply a "cheap" simplified surrogate model based on the specific prediction and potential error produced by kriging, the MSE might be directly used in each pre-test's remaining iterations instead of the EI. The simplified surrogate model in the pre-test, quite rough initially, is increasingly accurate as a result of adding objective function calls; therefore the MSE might guide the kriging model directly to search the region of the simplified surrogate model with high uncertainty.

3. Practical performance of the kriging with different strategies

3.1 A two-variable Schwefel test function

The efficiency of finding the global optimum using kriging with normal EI and with the other two novel strategies has been tested with the two-variable Schwefel test function (Schwefel, 1981) as an objective function in the range ($x_1 \in [-500\ 500]$, $x_2 \in [-500\ 500]$). The two-variable Schwefel test function is defined as follows ($d = 2$):

$$f(x) = \sum_{i=1}^{d} -x_i \sin\left(\sqrt{|x_i|}\right) \tag{3}$$

Figure 2 shows the contour of the two-dimensional Schwefel test function including one global minimum and several local minima, which are distributed irregularly; this function is acknowledged to provide a stern test for optimization algorithms.

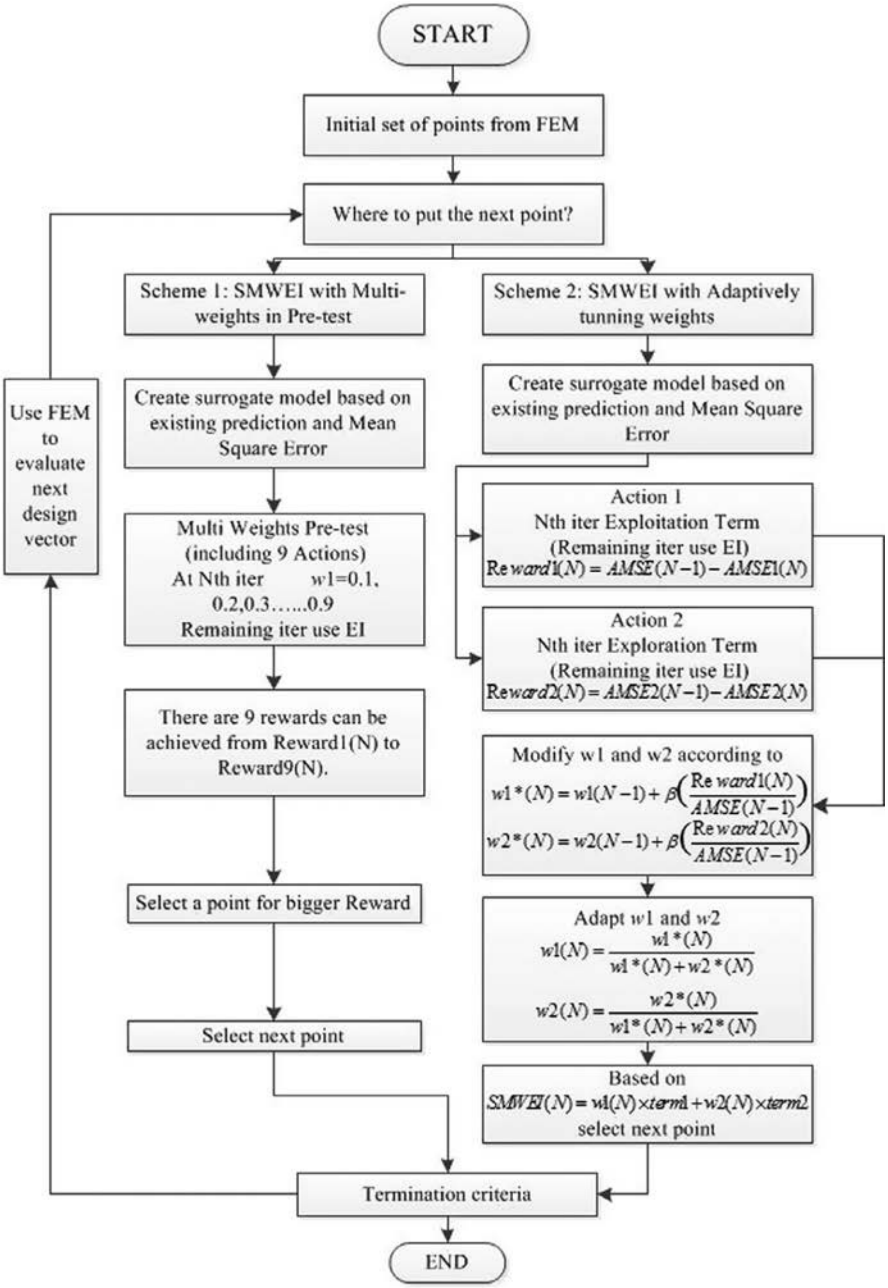


Figure 1.
The flowchart for SMWEI with multi weights in the pre-test

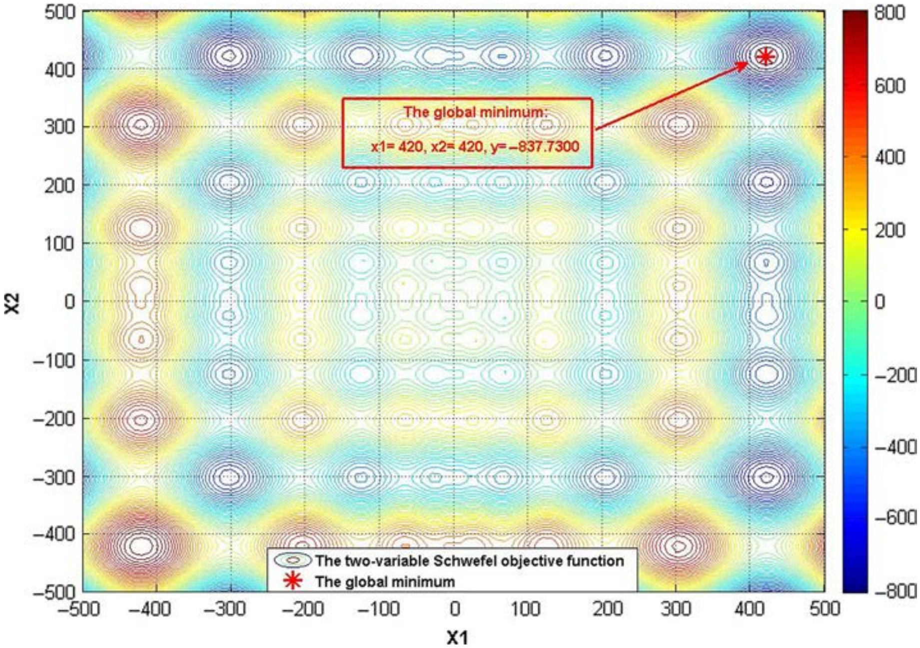


Figure 2.
The contour of the two-variable Schwefel function in the $x_1 \in [-500\ 500]$, $x_2 \in [-500\ 500]$ and $y \in [-500\ 500]$ domain

3.2 The performance of the kriging with different strategies

In order to analyse and compare the effects of applying different strategies to assist the kriging method, the two-variable Schwefel function with the initial samples imposed at $x_1 = -450,\ x_2 = -480;\ x_1 = -300,\ x_2 = 280;\ x_1 = -300,\ x_1 = -280;\ x_1 = -450,\ x_1 = 380;\ x_1 = 450,\ x_2 = -380;\ x_1 = -450,\ x_2 = -480;\ x_1 = -300,\ x_2 = -280;\ x_1 = 0, x_2 = 0$ was used to test the EI, the SMWEI with multi-weights and the SMWEI with the adaptively tuning weights algorithms. The following graphs show the "history" of how the points were added throughout the iterative process. The points found using the two different strategies used in this process – exploitation and exploration – are shown in Figures 3-5. Figure 3 shows that 12 iterations are needed to find the global minimum of the objective function when using EI.

When using the strategy of adaptively tuning the weights, a slightly better performance is obtained as only 11 iterations are needed to find the global minimum (Figure 4). In Figure 5, SMWEI with the strategy of multi-weights still needs the same number of iterations as EI to find the global minimum. The optimal weights for each iteration are presented in Table I.

As the two algorithms were originally designed to pursue the optimal weights by minimizing the average MSE, they have the potential of representing correctly the shape of the objective function in addition to finding the location of the global optimum efficiently. The algorithms do not stop automatically once the global minimum has been found but instead continue until the termination a criterion is triggered (finding a repeated sample point with prescribed tolerance). Hence, the kriging surrogate model might improve the prediction of the shape of the objective function further after locating the global minimum. However, the two-variable

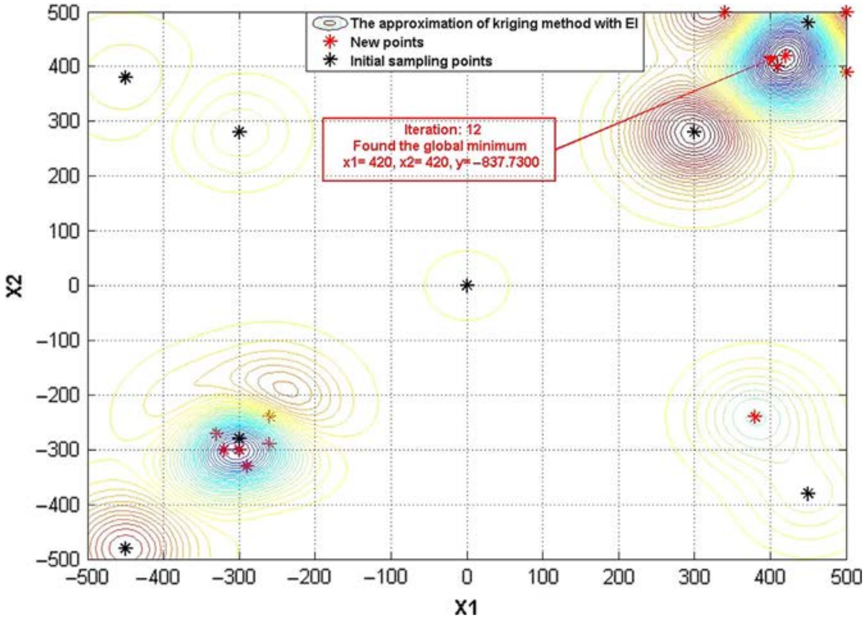
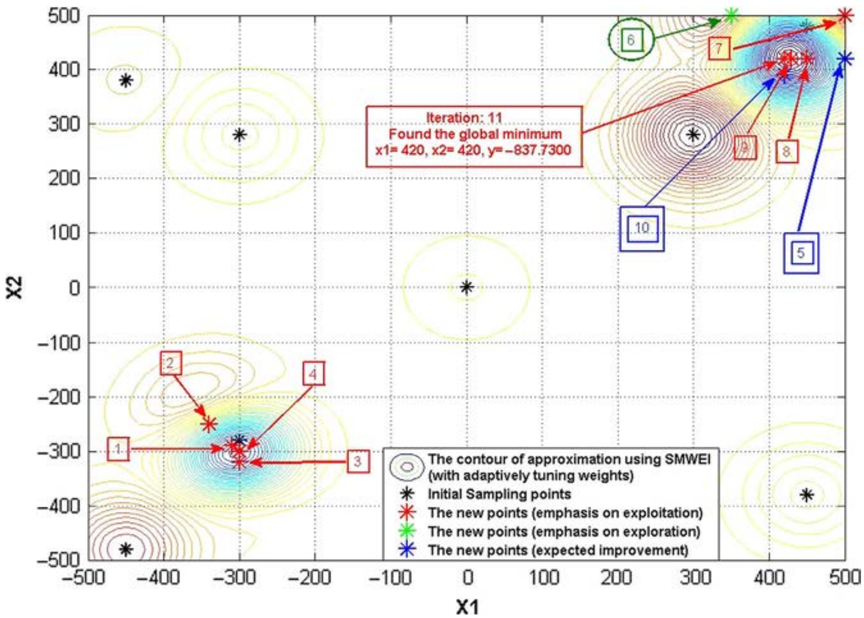
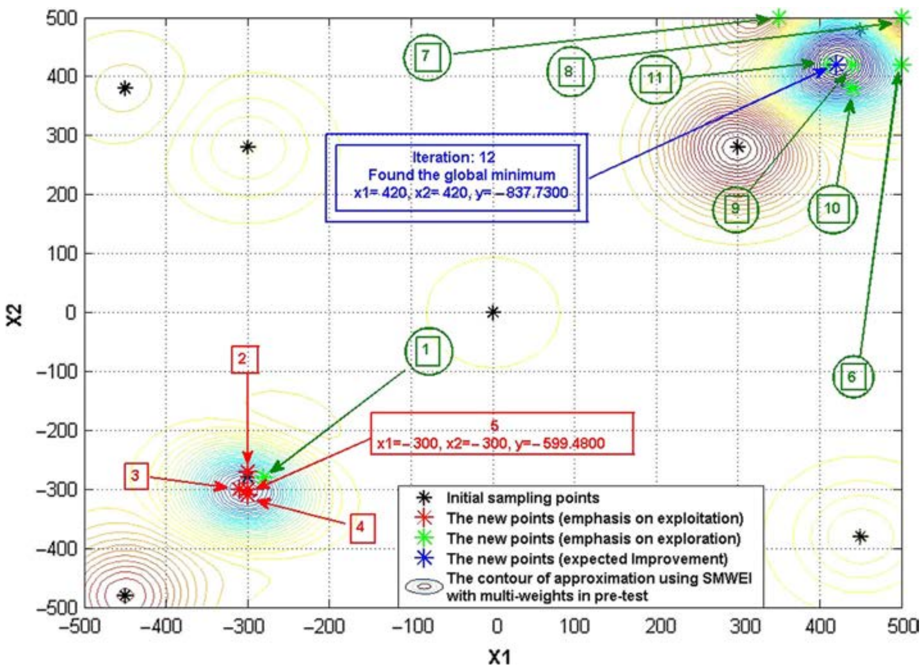


Figure 3.
The performance of the kriging with EI to predict the two-variable Schwefel function



Notes: The circle with a number means more exploration at that iteration; the double square means EI; the square with a number means more exploitation at that iteration

Figure 4.
The performance of the kriging assisted by SMWEI with adaptively tuning weights to predict the two-variable Schwefel function



Notes: The circle with a number means more exploration at that iteration; the double square means EI; the square with a number means more exploitation at that iteration

Figure 5.
The performance of the kriging assisted by SMWEI with multi-weights to predict the two-variable Schwefel function

Iteration number	x_1	x_2	y	The optimal weights
1	− 280	− 280	− 478.7100	0.3 (emphasis exploration)
2	− 300	− 270	− 478.5300	0.6 (emphasis exploitation)
3	− 310	− 300	− 593.2100	0.9 (emphasis exploitation)
4	− 300	− 310	− 593.2100	0.7 (emphasis exploitation)
5	− 300	− 300	− 599.4800	0.7 (emphasis exploitation)
6	500	420	− 238.2800	0.4 (emphasis exploitation)
7	350	500	229.8700	0.3 (emphasis exploration)
8	500	500	361.1800	0.4 (emphasis exploration)
9	440	420	− 792.6300	0.2 (emphasis exploration)
10	440	380	− 601.9300	0.4 (emphasis exploration)
11	410	420	− 822.8200	0.1 (emphasis exploration)
12	420	420	− 837.7300	0.5 (EI)

Table I.
The optimal weights of the kriging assisted by SMWEI with multi-weights at each iteration within the iterative process

Schwefel function may not be the best test for a robust design and thus a special function has been built for that purpose.

4. A robust design

A design engineer is always expected to have an appreciation of how small changes in parameters will affect the device performance (Di Barba, 2010). It may be the case, however, that even when an optimal design method is applied to a practical engineering problem, a theoretically optimal solution with excellent predicted performance in reality performs poorly when it is manufactured (Koh *et al.*, 1997), because in real-world implementations the nominal values are often subject to uncertainties or tolerances (Steiner *et al.*, 2004). The theoretically optimal values may also be affected by uncontrollable external perturbations which can result in considerable deterioration of the target performance compared with the nominal solution (Steiner *et al.*, 2004). In order to increase the reliability of the product, it is sometimes the case that the theoretically best solution which is not robust enough has to be abandoned in favour of a more robust solution which otherwise may not perform as well as under ideal circumstances. Clearly the first prerequisite for solving a robust design problem is that the surrogate model must be able to predict the shape of the objective function as well as locating the global optimum precisely.

The efficiency of the two proposed strategies in finding the global optimum has been investigated in the previous section. The new challenge is now to assess the quality of the shape representation to allow judgements to be made regarding the robustness. For this purpose a test function with two variables $F(x_1, x_2)$ has been built as plotted in Figure 6. The function has a global minimum ($x_1 = 46, x_2 = 46, y = 700$) and a local one in Region B ($x_1 \in [16\ 26], x_2 \in [16\ 26], y = 790$). Any small departure from the position of the global minimum (Point A) will result in a significant increase in the value of the objective function (making the performance of the device unacceptable), whereas in Region B the objective function is far less sensitive to changes in the two variables (x_1,x_2). Thus, compared with the "sharp" global minimum (Point A), all solutions within the marked square of Region B may be considered as robust as the practical performance of the device will be consistent even when actual dimensions change due to tolerances or the material

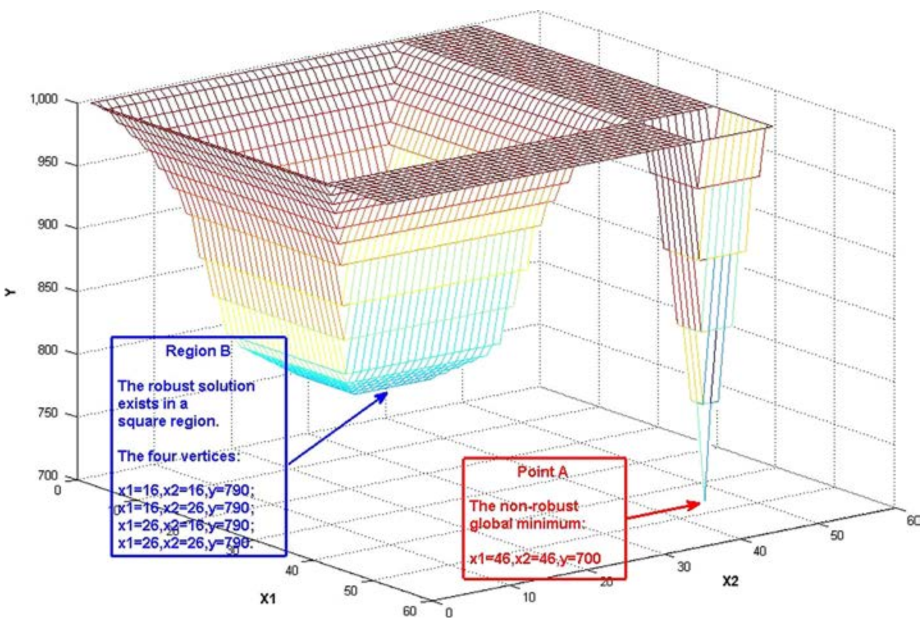


Figure 6.
The objective function for the robust test

properties are variable within prescribed limits. Thus, the function of Figure 6 may be argued to be a possible simple representation of the robust design problem.

4.1 Testing the algorithms for the quality of the prediction of the robustness

In order to compare the predictions of different strategies, the kriging method with EI has been tested first. Six initial sample points have been chosen as shown in Table II.

These six initial points have been used in all the robust design tests. The kriging method with EI needs 177 iterations to find the global minimum and the predicted function shape after 177 iterations is shown in Figure 7. The picture on the right hand side shows the error between the actual robust test function and the prediction by kriging with EI. The error is calculated as the absolute value of the difference between the value of the objective function and the value of the approximation.

This graphical representation of the results should aid the understanding of the quality of the prediction. For example, in Figure 7(b), the error around the boundary of Region B containing the robust solution indicates that the quality of the shape prediction is not that accurate at that region. The rugged shape of the prediction by EI around Region B is therefore likely to misguide the judgement regarding the robust solution.

The test of SMWEI with adaptively tuning weights used the same initial conditions. Compared with EI, the SMWEI approach is more efficient in finding the global minimum (only 85 iterations – see Figure 8); however, the error between the objective

Table II.
Initial sampling points

x_1	x_2	Objective function value (y)
10	10	850
13	12	810
19	13	805
10	41	1,000
41	10	1,000
46	44	840

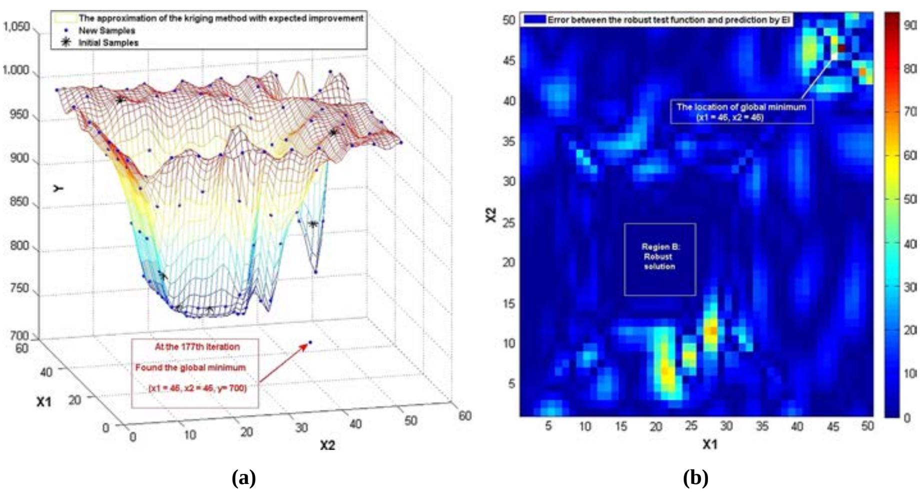


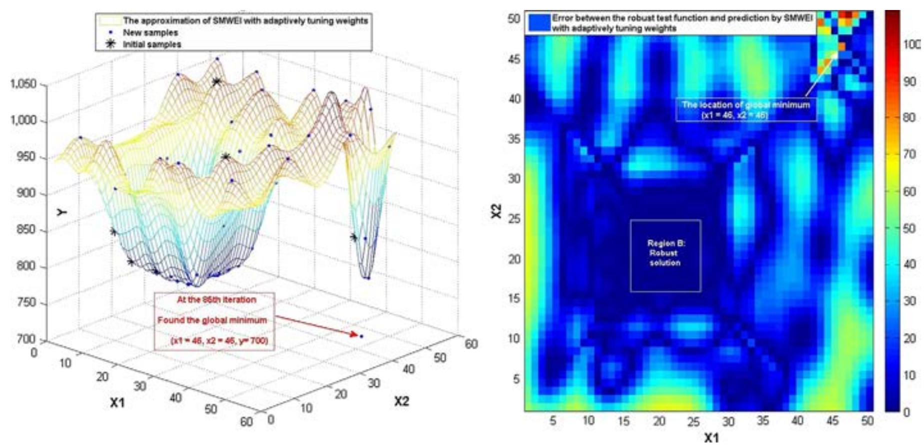
Figure 7.
(a) The performance of the kriging with EI when facing the robust problem and (b) the error between the robust test function and prediction by kriging model with EI

function and the predicted function is worse, which is a direct consequence of having fewer points available for shape representation.

In this work we have used the Gaussian correlation model as it outperforms most of other models used with kriging when objective functions with densely distributed valleys and hills are approximated (Lebensztajn *et al.*, 2004). In the robust test discussed here the test function is very simple: it has only two minima and is flat around these two regions (Figure 6). Due to the nature of the Gaussian correlation model the error introduced in the approximation of the flat regions of the function with a reduced number of sampling points can be relatively large. Although the error is reduced dramatically as the number of sampling points in the flat zones increases (Figure 7(b)), having such extra points might be otherwise useless as these regions are of no interest to the optimization routine. The advantage of SMWEI with adaptively tuning weights over the EI strategy is also apparent when the local minimum plateau is considered. The important areas of the searching space are much better approximated by the SMWEI as this strategy concentrates more sampling points around these areas (Figure 8(b)) whereas EI distributes the sampling points more evenly throughout the searching space therefore the error in these regions is larger (Figure 7(b)).

A similar observation can be made for the kriging model using the SMWEI with multi-weights (Figure 9). A somewhat better distribution of sampling points results in improved approximations of the two minima and fewer iterations for finding the global minimum as compared with EI, although – as mentioned before – the flat regions of the objective functions are not well approximated due to the Gaussian correlation model and the number of sample points in these areas.

Figure 10 shows the history of the kriging assisted by SMWEI using the multi-weights process. The graph describes the variation of the optimal weight at each iteration. Rather than applying equal weights of EI within the whole predicting process, the SMWEI with multi-weights has a more flexible and adaptive approach in choosing the best-performing weights in the pre-test. This method makes the process of reducing the average MSE more efficient and specific to the problem being solved.



(a) (b)

Figure 8.
(a) The performance of the kriging model assisted by SMWEI with adaptively tuning weights when facing the robust problem and (b) the error between the robust test function and prediction by the kriging model assisted by SMWEI with adaptively tuning weights when facing the robust problem

5. Conclusion

As the main target of this paper a robust design problem has been considered. The critical prerequisite for solving the robust problem is providing accurate prediction of the shape of the objective function. Two novel strategies for selecting weights in the pre-test have been proposed in order to balance exploration and exploitation.

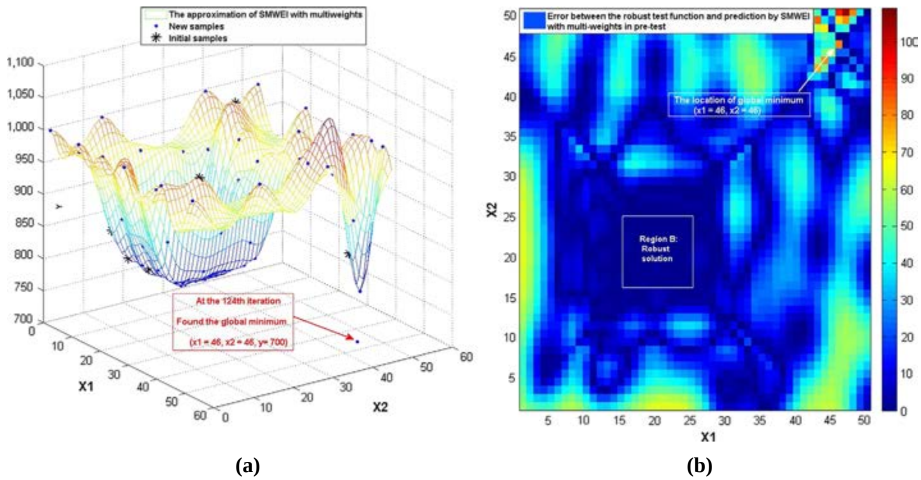


Figure 9.
(a) The performance of the kriging assisted by SMWEI with multi-weights when facing the robust problem and (b) the error between the robust test function and prediction by the kriging model assisted SMWEI with multi-weights

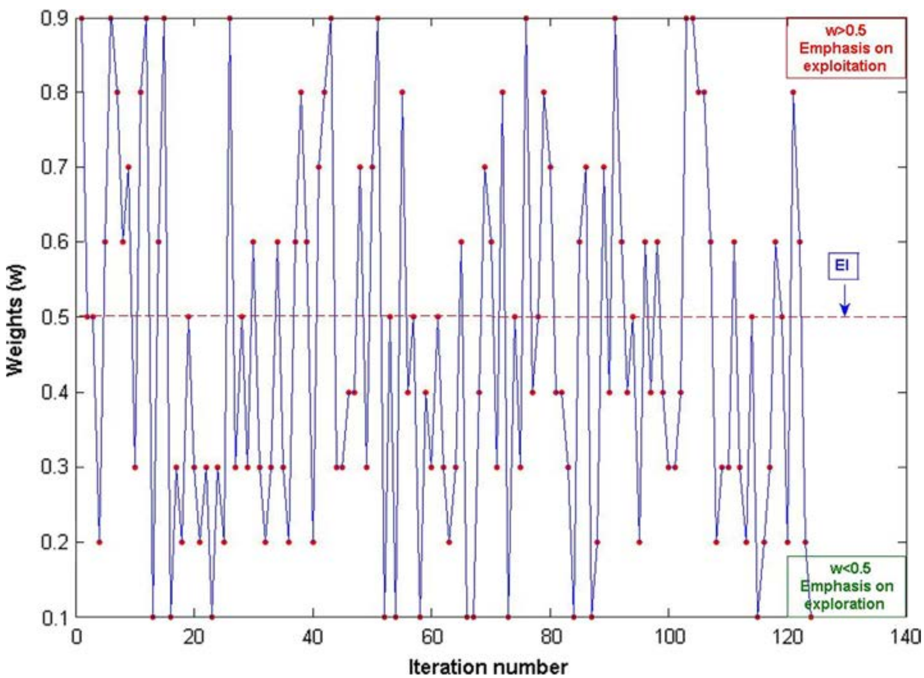


Figure 10.
The detailed situation of choosing optimal weights at each iteration

Both algorithms are based on kriging surrogate modelling and use the notion of rewards from the kriging model itself for better prediction of the shape of the objective while finding the global optimum efficiently. The algorithms have been tested using a two-variable Schwefel function and a specially devised robust test function and are shown to perform better than the traditional utility function. Both will be implemented in practical design systems, especially for the purpose of electromagnetic robust design optimization.

References

Di Barba, P. (2010), *Multiobjective Shape Design in Electricity and Magnetism*, Springer, London.

Jones, D.R., Schonlau, M. and Welch, W.J. (1998), "Efficient global optimization of expensive black-box functions", *Journal of Global Optimization*, Vol. 13, pp. 455-492.

Koh, C.S., Yoon, H.S., Nam, K.W. and Choi, H.S. (1997), "Magnetic pole shape optimization of permanent magnet motor for reduction of coging torque", *IEEE Transactions on Magnetics*, Vol. 33 No. 2, pp. 1822-1827.

Lebensztajn, L., Marreto, C.A.R., Costa, M.C. and Coulomb, J.-L. (2004), "Kriging: a useful tool for electromagnetic device optimization", *IEEE Transactions on Magnetics*, Vol. 40 No. 2, pp. 1196-1199.

Schwefel, H.P. (1981), *Numerical Optimization of Computer Models*, Wiley, New York, NY.

Sobester, A., Leary, S.J. and Keane, A.J. (2005), "On the design of optimization strategies based on global response surface approximation models", *Journal of Global Optimization*, Vol. 33, pp. 31-59.

Steiner, G., Weber, A. and Magele, C. (2004), "Managing uncertainties in electromagnetic design problems with robust optimization", *IEEE Transactions on Magnetics*, Vol. 40 No. 2, pp. 1094-1099.

Sutton, R. and Barto, A. (1998), *Reinforcement Learning: An Introduction*, The MIT Press, Cambridge, MA.

Sykulski, A., Adams, N. and Jennings, N. (2010), "On-line adaptation of exploration in the one-armed bandit with covariate problem", *Proc. of 9th Int. Conf. on Machine Learning and Applications*.

Xiao, S., Rotaru, M. and Sykulski, J.K. (2012), "Exploration versus exploitation using kriging surrogate modelling in electromagnetic design", *COMPEL – The International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, Vol. 31 No. 5, pp. 1541-1551.

Further reading

Chen, S. (2009), "Locus swarms – a new multi-optima search technique", *IEEE Congress on Evolutionary Computation*, pp. 1745-1752.

About the authors

Song Xiao is registered as a PhD student at the Electronics and Computer Science (ECS), University of Southampton, Southampton, UK. His work focuses on general methods of efficient design optimisation of electromechanical devices using kriging based approach.

Mihai Rotaru has been a Lecturer of Applied Electromagnetics at University of Southampton since 2007. Prior to this, he was a Senior Research Engineer with Institute of Microelectronics,

Singapore, working on signal and power integrity issues on electronic packaging. His research interests are in computation electromagnetics, design and optimization of electromagnetic devices.

Jan K. Sykulski is Professor of Applied Electromagnetics at the University of Southampton, UK. His personal research is in development of fundamental methods of computational electromagnetics, power applications of high temperature superconductivity, simulation of coupled field systems and design and optimisation of electromechanical devices. He has published over 340 scientific papers. He is founding Secretary of International Compumag Society, Visiting Professor at universities in Canada, France, Italy, Poland and China, Editor of *IEEE Transactions on Magnetics*, Editor-in-chief of *COMPEL* (Emerald) and member of International Steering Committees of several international conferences. He is Fellow of IEEE (USA), Fellow of the Institution of Engineering and Technology (IET), Fellow of the Institute of Physics (IoP), Fellow of the British Computer Society (BCS) and has an honorary title of Professor awarded by the President of Poland. Personal web page: www.ecs.soton.ac.uk/info/people/jks. Jan K. Sykulski is the corresponding author and can be contacted at: jks@soton.ac.uk