Robust design optimisation of electromagnetic devices exploiting gradient indices and Kriging

Song Xiao, Mihai Rotaru, Jan K. Sykulski

Electronics and Computer Science (ECS), University of Southampton, Southampton SO17 1BJ, UK
E-mail: jks@soton.ac.uk

Abstract: Since uncertainties in variables are unavoidable, an optimal solution must consider the robustness of the design. The gradient index approach provides a convenient way to evaluate the robustness but is inconclusive when several possible solutions exist. To overcome this limitation, a novel methodology based on the use of first- and second-order gradient indices is proposed introducing the notion of gradient sensitivity. The sensitivity affords a measure of the change in the objective function with respect to the uncertainty of the variables. A Kriging method assisted by algorithms exploiting the concept of rewards is employed to facilitate function predictions for the robust optimisation process. The performance of the proposed algorithm is assessed through a series of numerical experiments. A modification to the correlation model through the introduction of a Kriging predictor and mean square error criterion allows efficient solution of large scale and multi-parameter problems. The three-parameter version of TEAM Workshop Problem 22 has been used for illustration.

1 Introduction

The notion of robust design can be traced back to Taguchi in the 1950s who proposed the following definition: 'a product whose performance is minimally sensitive to factors causing variability (at lowest possible cost)'. His significant contribution was an introduction and specification of the orthogonal arrays as a particularly efficient approach to the design of experiments [1]. The more recent research has applied the method to the design of electromechanical devices [2] and has also focused on the related issues of optimal centring and tolerance design [3], where the optimal parameters and tolerances may be found by applying some form of a statistical procedure. The aim is a process of 'recentering' of the design factors such that the tolerances may be maximised with little or no quality losses and a solution is sought which is more compatible with the relevant manufacturing processes. Electromagnetic design generally falls into the same category of problems and searches not just for the global optimum but also some measure of its quality in case another optimum, even if only local, offers better robustness. What is somewhat special in the context of electromagnetic analysis is that accurate magnetic field simulation, essential to increase confidence in the predicted performance, requires computationally expensive numerical modelling, such as finite elements, finite differences and boundary elements. Commercial software is available for this purpose and it is extremely helpful to designers, but solution times for each field model are often long, thus repeated objective function calculations inherent to any design algorithm could make the process exceedingly time-consuming and thus impractical. This is why so much effort has gone into developing efficient design procedures based on surrogate modelling. This paper addresses such issues, both from the point of view of the actual optimisation process but primarily in the context of robust design. Some prior work is mentioned first to provide a starting point to the discussion.

The concept of the gradient index (GI) has been explored in [4–6]. The method transforms a problem into a multi-objective optimisation by concurrently minimising the function and its GI, thus forming Pareto fronts. This approach is promising but does not provide a clear answer how to select a preferred solution when the size of the uncertainty varies. Moreover, the sensitivity computations must be incorporated into the finite-element code, which may be impossible when commercial software is used. This paper offers further advances both in terms of the way in which the gradient indices are used and by improving the computational efficiency of the algorithms. The notions of the GI sensitivity and the second-order GI are introduced and explained. It is shown that when practical problems are attempted relying solely on the first-order gradient or second-order gradient for evaluation of robustness may not be enough, especially when several possible robust solutions exist which all have their first- and second-order gradients zero. Finally, rather than calculating the objective function using computationally expensive finite-element software, a Kriging prediction [7] is employed. In other words, the objective function is approximated using Kriging [8], assisted by algorithms balancing exploration and exploitation ([9, 10]) using the concept of rewards [11]. This strategy has been shown previously to be very efficient and has the advantage that it can be linked with any finite-element code, including commercial software.

2 Robust optimisation

In conventional optimisation, the minimum (maximum) of an objective function is sought while the search space is limited through a set of constrains. Once the global optimum has been found, the problem is considered to have been solved. When practical devices are designed, however, we need to recognise that almost all parameters (design variables) are subject to uncertainties (manufacturing tolerances, variation of material properties etc.) and thus not just the value but also the shape of the optimum becomes relevant in the neighbourhood of the selected design; this is demonstrated by the examples of Figs. 1 and 2. A theoretical optimum may therefore be abandoned in favour of a 'worse' but more robust design; however, the decision will depend on the size of the uncertainties involved. For this reason having a Pareto front instead of a single solution may be preferable.

2.1 Multi-objective robust optimisation using GI

Consider a commonly used one-variable test function [4–6] (see Fig. 1)

$$f(x) = 3 - \frac{3.5}{1 + (x - 5)^2}$$
$$- \frac{2.2}{1 + (x - 15)^2/10} - \frac{1.2}{1 + (x - 25)^2/30} \quad (1)$$

The uncertainties may be either specified directly (e.g. as machining tolerances, say Δ) or defined mathematically as

$$U(x) = \left\{ \xi \in R^n | x - k\sigma \le \xi \le x + k\sigma \right\} \quad (2)$$

where σ is the standard deviation of uncertain variables and k is determined by a confidence level [5].

One way of incorporating robustness into the mainstream optimisation process is by adding the GI [4] as a second objective and formulating the problem as

$$\text{Minimise } f(x) \ x \in R^n \ \left(x_L \le x \le x_U \right)$$
$$\text{Minimise GI}(x) = \max_{1 \le i \le n} \left| \partial f(x)/\partial x_i \right| \quad (3)$$
$$\text{Subject to } g_i(x) \le 0, \quad i = 1, \dots, m$$

Point A1 in Fig. 1a is the theoretical global optimum. However, any small change in the variable x results in a large variation of the objective function; thus, A1 is not a robust design and points A2 or A3 might be preferred. Importantly, it can be seen that both the values of the first-order gradient in Fig. 1a and the values of the second-order gradient in Fig. 1b, for both points A2 and A3, are very close to zero. Therefore considering only the first-order gradient or the second-order gradient to evaluate robustness of solutions might not be reliable for such cases. The final decision, however, is not straightforward and is influenced by the size of the uncertainty. The sensitivity of the gradient may thus be defined as the difference between the largest and the smallest value of the GI within the uncertainty range; as shown in Fig. 1a, the shape of this sensitivity carries useful information. The trajectory of the objective function in terms of sensitivity and objective function values is plotted in Fig. 1c. It can be seen that the traditional optimisation task of minimising a single objective function is transferred into a two-objective optimisation involving minimisation of the objective function and the sensitivity based on the GI. The theoretical most optimal solution and several critical local minima (A1–A3), together with the corresponding range restricted by specific uncertainties (U1–U5), have been outlined in Figs. 1a–c. Although A1 appears to offer a better objective function value, its sensitivity is relatively poor as compared with the other two local minima A2 and A3. The second-order gradient of the function may also be useful (Fig. 1b).

Another example (Fig. 2) shows 'sharp' global and 'shallow' local minima with a 'plateau' with associated sensitivities and second-order gradient indices. Strictly speaking A2 is not a 'minimum', but nevertheless a possible design with attractive spread of values. Using the first-order gradient and/or the second-order gradient on their own will not identify A2 as a potential solution, whereas the use of sensitivity – as shown in Fig. 2b – allows proper judgement to be made regarding the robustness of this particular design. Table 1 shows that the sensitivity values for A2 and A3 are smaller, and thus better, than for A1. The choice between A2 and A3 will then be guided by the shape of the trajectory of Fig. 2b.

3 Robust optimisation based on Kriging modelling

3.1 Kriging

As a type of regression model, Kriging [9] is able to predict the shape of the objective function through exploiting the spatial correlation of data based only on limited information. The accuracy of this prediction can be estimated by Kriging, which may be extremely helpful when making a decision where to place the next evaluation point at any stage of the optimisation process. To accomplish this aim, Kriging needs to exploit the spatial correlation between the known points (vectors) of the objective function and all the unknown points, as well as the correlation between the known points (newly found points and initial sampling points), to build a correct surrogate model of the real objective function through interpolation. This relies on the linear regression model

$$\hat{y}(x) = \sum_{k=1}^{m} \beta_k f_k(x) + \varepsilon(x) \quad (4)$$

and the Gaussian correlation model

$$R\left(\varepsilon(x^i), \varepsilon(x^i)\right) = \prod_{k=1}^{n} e^{-\theta_k \left| x_k^i - x_k^j \right|^{p_k}} \quad (5)$$

where the global function $\sum_{k=1}^{m} \beta_k f_k(x)$ and an additive Gaussian noise $\varepsilon(x)$ are integrated to the predicted value $\hat{y}(x)$ of the objective function; the hyperparameter θ_k is the correlation among the data in k-direction and p_k determines the 'smoothness' of (5). The most popular correlation function is given by the Gaussian model where the value of p_k is simply taken as equal to 2. For a given set of data, the maximum likelihood estimation optimises the value of θ and then the correlation model is brought into the regression model to evaluate the function with the best linear unbiased predictor [8, 12].
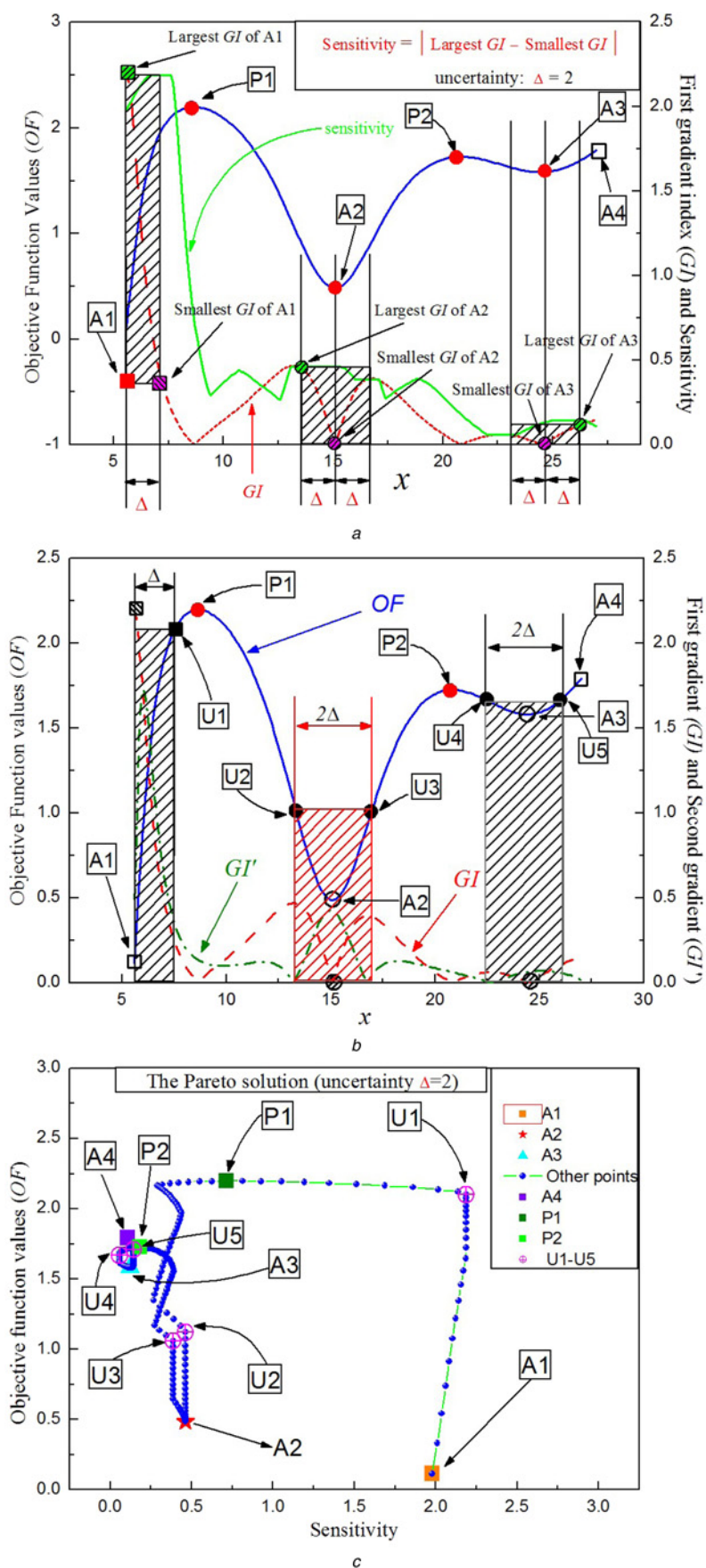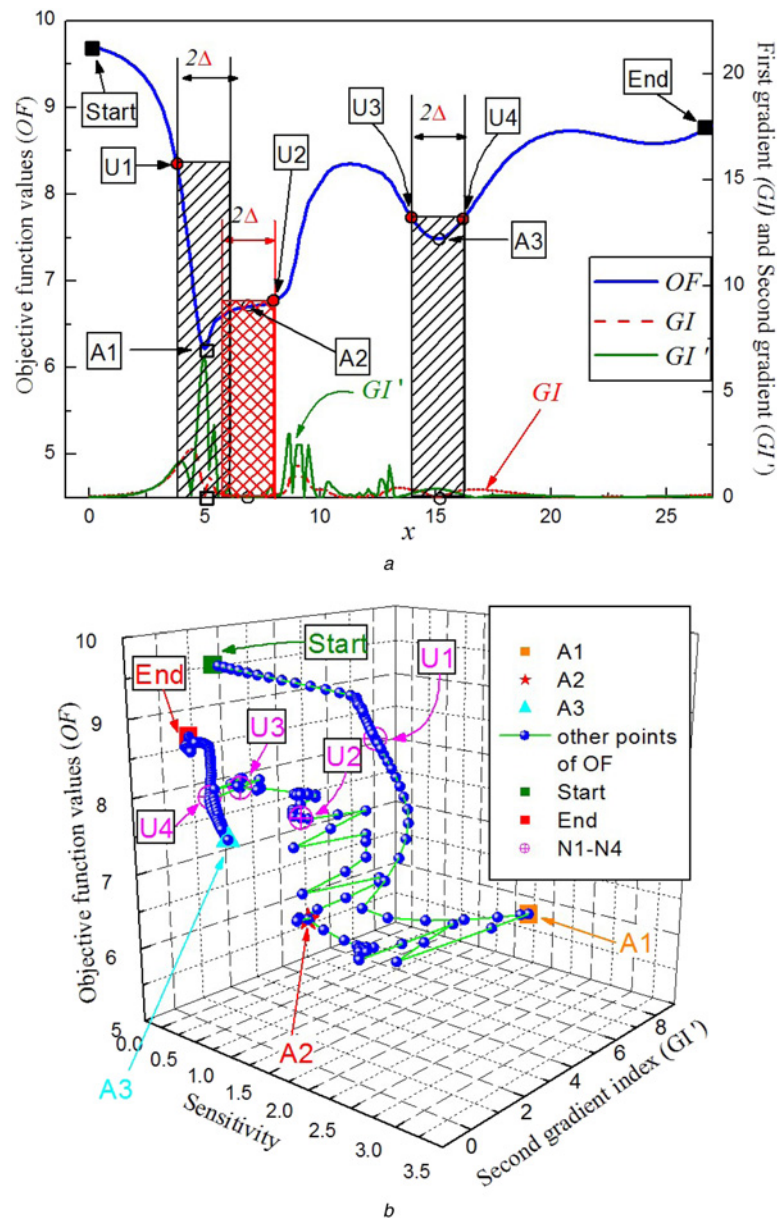
Fig. 1 *Example of a robust design for a one-variable problem*

a Objective function, the GI and sensitivity
b First- and second-order gradients
c Objective function trajectory showing sensitivity and objective function values (OF)

Fig. 2 *Two minima and a plateau*

a Objective function, the first-order gradient index (GI) and the second-order gradient index (GI')
b Objective function trajectory showing sensitivity, second-order gradient index (GI') and objective function values (OF)

Along with the increase in the number of sampling points selected by Kriging during the iterations, the amount of data produced by the correlation matrices accumulates constantly throughout the optimisation process, which may become problematic especially when dealing with large-scale multi-parameter problems, leading to a 'combinatorial explosion'. In [13], we proposed a successive 'zoom-in'

Table 1 Values of the objective function, second gradient and sensitivity for the three design points A1, A2 and A3 (Fig. 2)

	Coordinates on x-axis	Objective function value	Second gradient value	Sensitivity
A1	5	6.21628	6.702025	2.2634
A2	7.8	6.735	0.5	1.47
A3	15.1	7.48699	0.42475	0.4618

strategy to alleviate the problem, where –to reduce the amount of data storage and utilise the installed physical memory capacity efficiently – the step sizes of the variable parameters involved in the design (named design vectors here) were increased while the test range reduced. However, the optimal step size is often problem dependent, thus if the 'roughness' of the initial test is set inappropriately, it is possible that certain regions of the search space containing important information (including the optimum) might be missed. Hence, what appears to be the best point found in such a search may in fact misguide the algorithm leading to erroneous results. To address this issue, an alternative strategy is pursued here where, rather than reducing the size of the problem, a more efficient handling of relevant matrices is proposed. This novel methodology divides the correlation matrices in an adaptive manner so that the physical memory available is used in an optimum way.
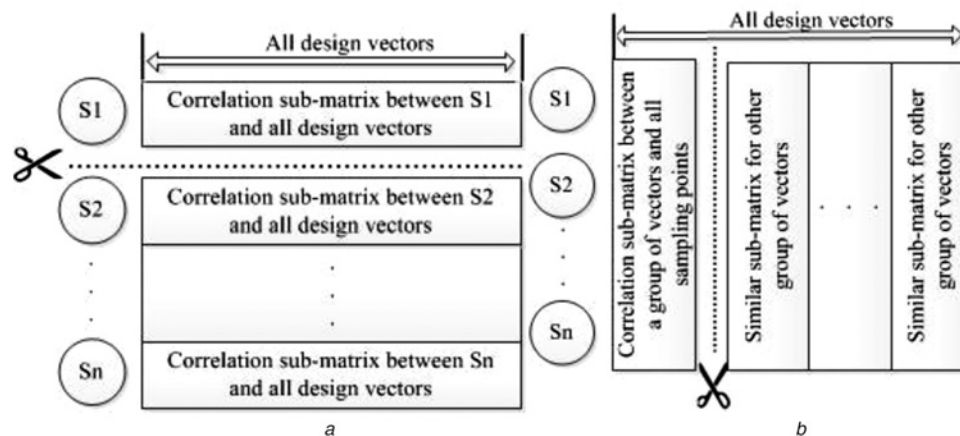
Fig. 3 *Correlation matrices partitioned by*

a Sampling points,
b Design vectors (where $S_1, S_2, \ldots, S_n$: sampling points)

3.2 Partitioning of correlation matrices

In general, the known sampling points selected throughout the iterative process of prediction only take account of a very limited part of the full design vectors; hence, the correlation matrix between the existing sampling points only is unlikely to cause memory problems. On the other hand, the correlation matrix between the known points and all the design vectors may be very large and keeps growing with the increasing number of updated sampling points as iterations progress. For problems with several variables and a large number of potential design vectors, the correlation matrix can grow uncontrollably and may result in 'combinatorial explosion'. Should this happen, a part of data might need to be transferred to a hard disk, which could cause the iterative process to slow down considerably. Therefore a strategy of partitioning the correlation matrices into manageable size is proposed and illustrated in the flowchart sketched in Fig. 3; it should be noted, however, that unordered (or random) splitting could cause unnecessary calling of the correlation model

subprogram leading to increased computing times, hence the need for a 'strategy'. The scheme proposed here adjusts the size of the sub-elements into which the correlation matrices are split adaptively to make full use of the available memory while minimising the number of calls of the correlation model. Two schemes have been employed and operate throughout the process as explained in Fig. 3, one based on partitioning in terms of the sampling points and the other related to the design vectors, for Kriging prediction and calculating mean-square error, respectively. Two alternative switching criteria have been implemented. The first is memory related: if the memory occupied by the predictor involving the correlation matrices exceeds the available memory of a specific computer, a modified strategy of partitioning matrices is applied instead of the original method of producing correlation matrices. However, as other background processes may simultaneously be taking place the available memory is never fixed. Consequently, another criterion has also been implemented related to the average time taken by a single iteration; should this time suddenly start to increase the switching is triggered and the partitioning matrices scheme is activated; it then continues throughout the remaining iterations.

A careful balance is therefore maintained between preventing the correlation matrices to grow uncontrollably while monitoring simultaneously the computing times. The operation of the scheme is problem related but also depends on the actual computer used, so it is impossible to provide strict guidelines regarding memory limits. The example of Fig. 9, however, does show some quantitative details for a specific case and particular computer implementation.

To verify the viability of the proposed methodology, the TEAM problem 22 has been attempted [14]. Without the strategy of Fig. 4, the Kriging algorithm failed because of the memory size problem when tested on a smaller computer.

4 Application to electromagnetic design

4.1 TEAM Problem 22

The full description of the TEAM benchmark problem 22 (superconducting magnetic energy storage system) may be found in [14]. The target is to achieve an arrangement of the two superconducting coils such that the stored energy
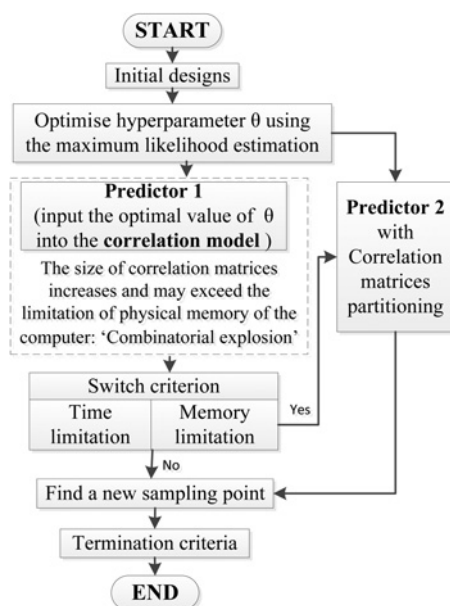
Fig. 4 *Flowchart of the adaptive division of correlation matrices*

Table 2 Setting of initial sampling points

R_2, m	2.7	2.9	3.0	3.3	2.6	3.4	2.6	3.4
h_2, m	0.744	1.304	1.64	2.088	0.408	0.408	2.2	2.2

within the system is $E_{ref} = 180$ MJ while a minimal stray field B_{stray} is obtained. The objective function is defined as

$$\text{OF} = \frac{B_{stray}^2}{B_{norm}^2} + \frac{|E - E_{ref}|}{E_{ref}} \qquad (6)$$

where $B_{norm} = 3\,\mu T$ and $B_{stray}^2 = \left[\sum_{i=1}^{22} \left|B_{stray,i}\right|^2\right] \Big/ 22$, subject to geometrical and 'quench' constraints.

4.2 Two-parameter test results

Before addressing the three-parameter problem, for the purpose of demonstrating typical shapes of the objective function, one of the variables has been fixed ($d_2 = 0.394$ m), while R_2 and h_2 are varied. Usually the initial sampling points are selected using Latin Hypercube [15]; however,

Table 3 Performance comparison of algorithms

Algorithm	R_2, m	$h_2/2$, m	Best OF	Iterations
RBF	3.06	0.236	0.088	240
AWEI (Kriging)	3.08	0.239	0.089	38

Radial basis functions (RBF) [16]; AWEI (Kriging) [10].

for this test eight initial sampling points were used as shown in Table 2. Table 3 compares one typical result from literature with our AWEI algorithm (Kriging with adaptive weighted expected improvement) [9, 10], whereas Fig. 5 demonstrates the convergence process of AWEI.

The uncertainties are predefined as $R_2 - 0.02 < \xi(R_2) < R_2 + 0.02$, $h_2 - 0.01 < \xi(h_2) < h_2 + 0.01$, while the increments with respect to R_2 and h_2 for calculating first-order gradients are set as $dx(R_2) = 0.002$, $dx(h_2) = 0.01$.
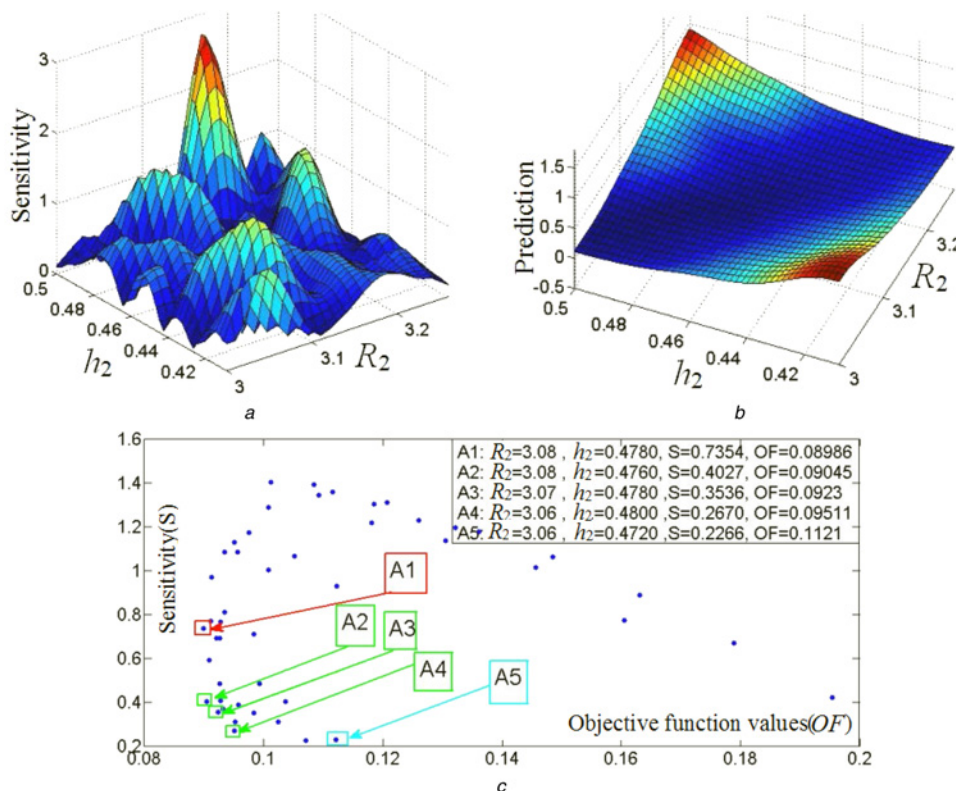
4.3 Three-parameter test results

The full three-parameter TEAM 22 problem [14] is potentially a challenge to the Kriging method because of the 'combinatorial explosion' associated with setting up the correlation model, as explained in Section 3 – thus the savings because of avoidance of the computationally expensive finite-element simulations may be lost, or even overtaken, by the excessive time required by the model if a less powerful computer is used for simulation. The previously reported 'zoom-in' strategy [13] to deal with this issue has some drawbacks and could result in loss of accuracy. This was the motivation behind the new approach based on adaptive correlation matrices division described in Section 3. The initial sampling points are set as in Table 4, whereas the settings for the test ranges and step sizes

Fig. 5 *TEAM Workshop Problem 22*

a Prediction by Kriging with AWEI ($3 \leq R_2 \leq 3.3$, $0.408 \leq h_2 \leq 0.5$, $d_2 = 0.394$, other parameters fixed)
b Sensitivity with respect to R_2 and h_2
c Objective function trajectory showing sensitivity and objective function values (OF)

Table 4 Setting of initial sampling points

	R_2, m	h_2, m	d_2, m
sample 1	3.1	0.576	0.32
sample 2	3	0.408	0.3
sample 3	3.2	0.744	0.4

Table 5 Specific definition of the test as used for constructing the Kriging model

	R_2, m	h_2, m	d_2, m
test range	[2.6 3.4]	[0.408 2.2]	[0.1 0.4]
step size	0.01	0.014	0.003
number of steps	81	129	101

shown in Table 5 are initialised to preserve the test conditions originally suggested in [14]. Please note that the relatively small imposed step sizes are probably unnecessary – and relaxing them would have made the problem easier and computationally more efficient to solve – but they have been kept here to allow for direct comparisons.

A comparison between the results obtained with the 'zoom-in' strategy and those using 'the adaptive correlation matrices division' for this example are presented in Fig. 6.
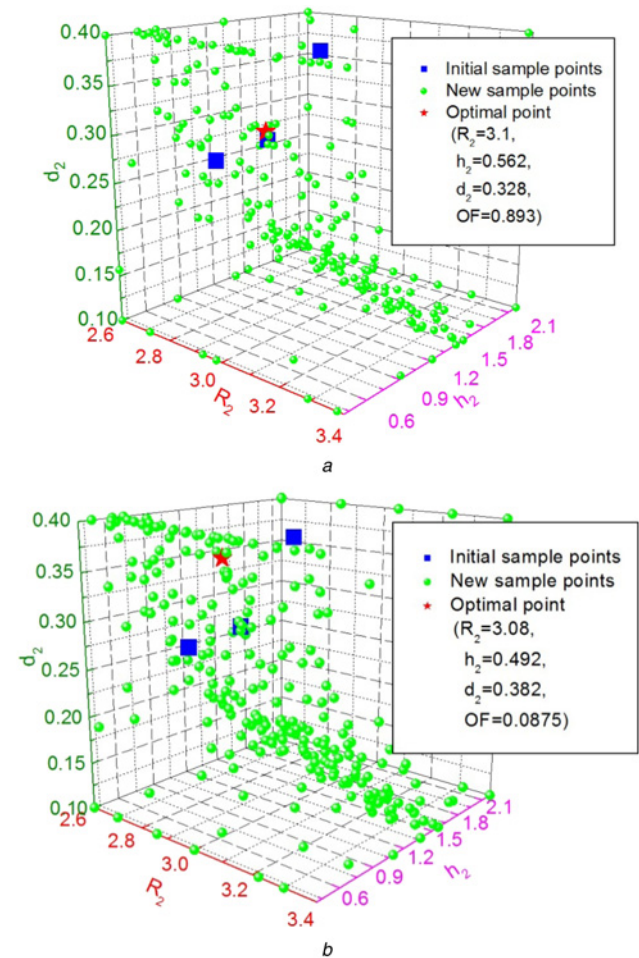

Fig. 6 *Three-parameter TEAM 22 problem*
a EI
b AWEI

Table 6 Performance comparison of algorithms for TEAM 22 problem

Algorithm	R_2, m	d_2, m	$h_2/2$, m	OF	No. of FEM calls
GA	3.040	0.386	0.240	0.134	2400
HuTS	3.080	0.380	0.246	0.089	3821
ITS	3.100	0.388	0.240	0.098	1824
SA	3.078	0.390	0.237	0.098	5025
NTS	3.080	0.370	0.254	0.089	1800
PBIL	3.110	0.421	0.241	0.101	3278
Kriging (EI)	3.1	0.328	0.281	0.0893	276
Kriging (AWEI)	3.08	0.382	0.246	0.0875	242

Genetic algorithm (GA) [17]; Tabu search (HuTS) [18]; improved Tabu search (ITS) [19]; simulated annealing algorithm (SA) [20]; new Tabu search (NTS) [21]; population-based incremental learning (PBIL) [22].

With the 'zoom-in' strategy, the Kriging assisted by expected improvement (EI) and AWEI took 211 iterations and 323 iterations, respectively. However, the number of iterations is not uniform as it depends on the choice of the
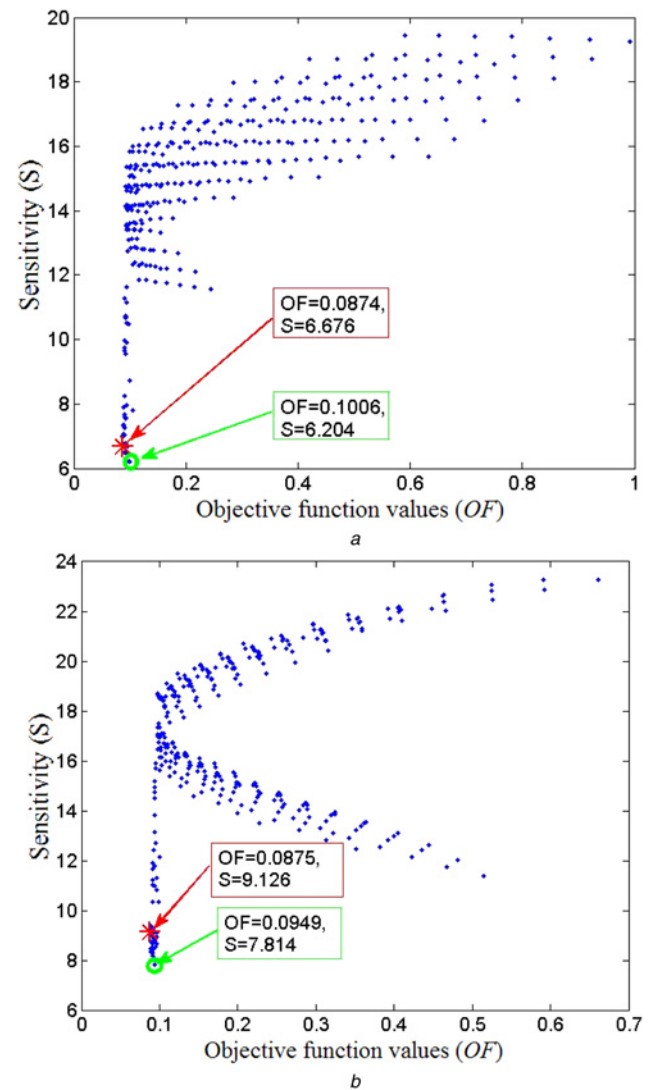

Fig. 7 *Objective function trajectory showing sensitivity and objective function values (OF)*
*a* EI
b AWEI

specific range of each test stage. By applying the scheme of partitioning the correlation matrices, the issue of combinatorial explosion is alleviated to a great extent – without a loss of important information – by the sheer fact of the sub-matrices being of much smaller size. With the help of this scheme, the Kriging with EI needed 276 iterations to find the global optimum. Although it has been noted that Kriging with EI performs slightly worse than the 'zoom-in' strategy, the entire design space has now been explored. The AWEI using the novel scheme performs better, as it took 242 iterations rather than 323 to find the optimum (Fig. 6b). The comparisons of efficiency of convergence (number of finite element method (FEM) calls) and the quality of the optimum (value of OF) are presented in Table 6. Kriging with either EI or AWEI performs far more efficiently than other currently available methodologies. Kriging guided by AWEI performs slightly better than Kriging with EI, both in terms of efficiency and the quality of the optimum. However, the main benefit of the modified correlation model is that the Kriging model is now able to cope with problems that theoretically could be of any size without the complications brought about by the 'combinatorial explosion'.

The uncertainties for R_2, h_2 and d_2 have been pre-set as $R_2 - 0.03 < \xi(R_2) < R_2 + 0.03$, $R_2 - 0.042 < \xi(h_2) < R_2 + 0.042$, $d_2 - 0.009 < \xi(d_2) < d_2 + 0.009$. The differences $d(x)$ required by the gradient calculation with respect to these three parameters were set as their relevant step sizes. The objective function trajectory in terms of sensitivity and objective function values (OF) obtained using Kriging with EI and AWEI are presented in Figs. 7a and b, respectively. Both best solutions, in terms of the objective function value and the sensitivity, are shown in Fig. 7. The graphs focus around the optimum values of the function. It can be seen that both EI and AWEI produce similar results and for the

given set of uncertainties the 'less optimal' solution seems to be more robust than the optimum one.

To demonstrate the performance of the proposed partitioning scheme, two tests – both applying EI and AWEI – have been undertaken and the results of the monitoring of memory savings and associated computing times for a specific computer are presented in Fig. 9. The size of the full correlation matrix, the reduced size using the idea of a standard sub-matrix and the joint usage of the sub-matrix and matrix partitioning are described in Fig. 8. The decision if the correlation matrix needs to be partitioned follows from two considerations: limiting the memory occupied by the optimiser or minimising the time required by each iteration. When one of the limitations is exceeded, the normal method of producing correlation matrices is replaced by the modified scheme of partitioning matrices. Once the criteria for switching on the partitioning scheme are activated, the correlation matrices partitioning into sub-matrices – both in terms of design vectors and sampling points – are triggered. If splitting in terms of design vectors, the size of the full version correlation matrix is $81 \times 129 \times 101 \times n$ (where n is the total number of existing sampling points), that is divided into 36 sub-matrices (35 sub-matrices with the size $30\,000 \times n$ and 1 sub-matrix with the size of $5349 \times n$). As mentioned before in Section 4.3, under normal design practice the number of steps could probably be made smaller making the problem go away or at least less severe. In the particular case, the testing environment provided ~11 GB available memory, which was set as the triggering value. However, because of other simultaneous processes, the effective available memory was smaller, although difficult to predict. Thus, the time for a single iteration was also monitored and – as shown in Fig. 9 – at some point a marked increase could be observed when the 11 GB memory limit had not
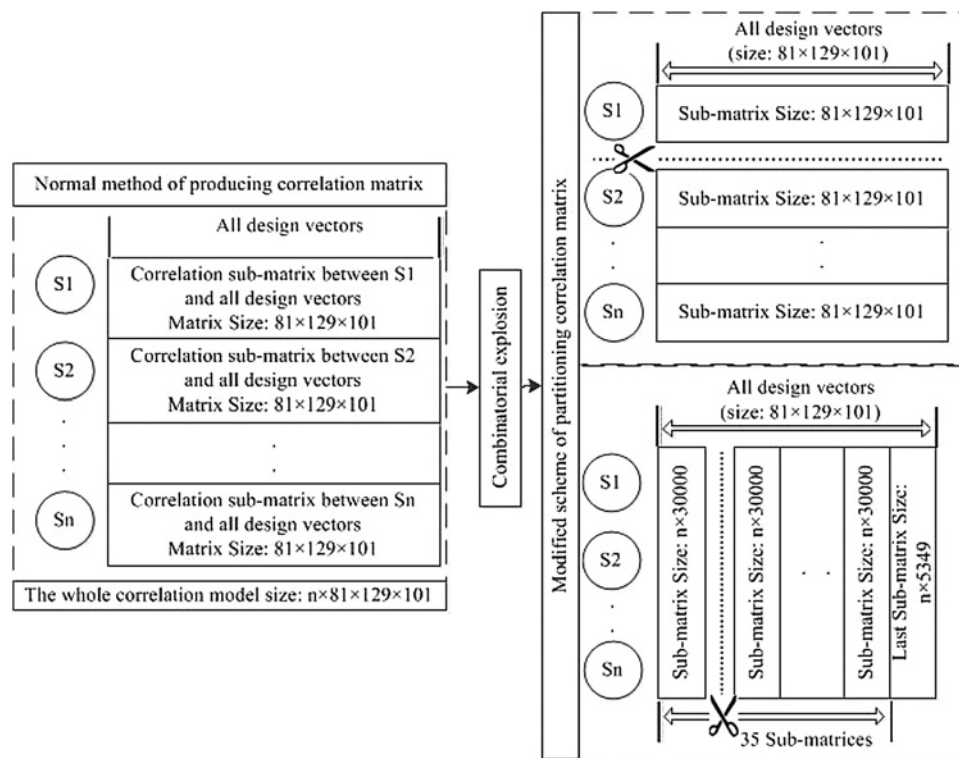
Fig. 8 *Specific parameter setting and transient process of testing the TEAM 22 problem with the help of the novel partitioning scheme*
S: sampling points, *n*: the number of existing sampling points
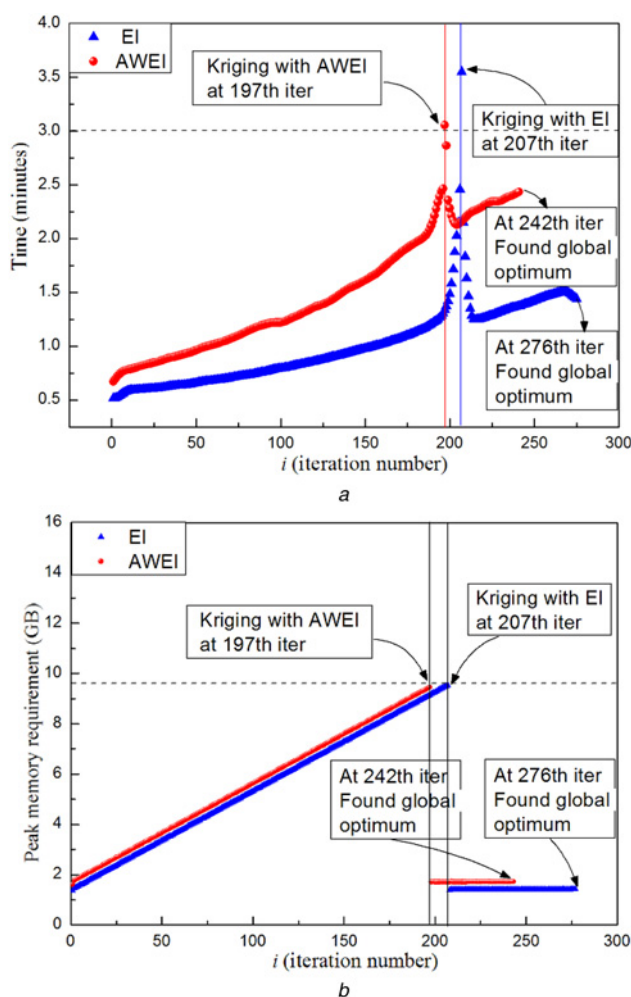
Fig. 9 *Monitoring of computing times and memory requirements*

a Computing time of each iteration
b Memory requirements

yet been reached; in this particular case, the actual memory usage was around 9.6 GB. Thus, in cases when background (or other) processes may be memory 'hungry', this additional iteration time constraint is clearly helpful. Overall, setting the memory limit somewhat below the 'theoretical maximum' might be advisable, but monitoring the computing times also useful.

Finally, although this is not the focus of this paper, it has been reported before that Kriging offers a particularly efficient way of surrogate modelling with the number of necessary function calls significantly less than when using other methods; this was demonstrated for example by Table IV in [13] and confirmed by Table 6 in Section 4.3 of this paper. Thus although in principle other algorithms could also be used to achieve the same goal, there seems to be a good reason to continue using Kriging. However, the issue of the size of the correlation matrices – and associated time required for their creation at each iteration step – is an important hurdle in implementation of the method especially on smaller computers. Some ideas about how to overcome this difficulty have been put forward here, but will be mostly pursued in future work. Notwithstanding, it should be noted that when using a computer with limited memory there will be difficulties associated with running the numerical simulation software too, which the Kriging model is trying to replace, so the issues of limited available memory are not unique to Kriging formulation; they must,

however, be addressed. One possible way forward for much larger practical problems with many variables is to explore the sensitivity of the objective function to changes in all parameters and focus on a subset of such parameters; this idea is pursued in the follow-up work [23].

5 Conclusion

Finding a global optimum may not be sufficient when a robust design is desired. If uncertainties of the variables are defined, the difference between the highest and the lowest gradient can be used to evaluate the sensitivity of the solution in the search for a robust result. Assisted by the Kriging model, the evaluation of sensitivity can rely on the Kriging prediction rather than the computationally expensive finite-element models, but the classic Kriging model can be inefficient when multi-variable problems are solved on smaller computers. The strategy based on splitting the correlation matrices makes the process much more effective for large-scale multi-parameter electromagnetic design problems.

6 References

1 Weng, W.C., Yang, F., Demir, V., Elsherbeni, A.: 'Optimization using Taguchi method for electromagnetic applications'. Proc. European Conf. Antennas and Propagation 'EuCAP 2006', (ESA SP-626), Nice, France, 6–10 November 2006
2 Kim, K.S., Lee, S.J., Cho, S.G., et al.: 'Multi-response Taguchi robust design of back electromotive force and cogging torque considering the manufacturing tolerance for electric machine'. 13th Int. Conf. Optimization of Electrical and Electronic Equipment (OPTIM), Brasov, Romania, 24–26 May 2012
3 González, I., Sánchez, I.: 'Optimal centering and tolerance design for correlated variables', *Int. J. Adv. Manuf. Technol.*, 2013, **66**, pp. 1499–1510
4 Ren, Z.Y., Pham, M.T., Song, M.H., Kim, D.H., Koh, C.S.: 'A robust global optimisation algorithm of electromagnetic devices utilizing gradient index and multi-objective optimisation method', *IEEE Trans. Magn.*, 2011, **47**, (5), pp. 1254–1257
5 Ren, Z.Y., Zhang, D.H., Koh, C.S.: 'Reliability evaluation of the electromagnetic device based on the second order sensitivity analysis'. Sixth Int. Conf. Electromagnetic Field Problems and Applications (ICEF), Dalian, China, 2012
6 Ren, Z.Y., Pham, M.T., Koh, C.S.: 'Robust global optimisation of electromagnetic devices with uncertain design parameters: comparison of the worst-case optimisation methods and multi-objective optimisation approach using gradient index', *IEEE Trans. Magn.*, 2013, **49**, (2), pp. 851–859
7 Sykulski, J.K.: 'New trends in optimisation in electromagnetics'. IET Seventh Int. Conf. Computation in Electromagnetics (CEM 2008), 2008, pp. 44–49
8 Lebensztajn, L., Marreto, C.A.R., Costa, M.C., Coulomb, J-L.: 'Kriging: a useful tool for electromagnetic device optimisation', *IEEE Trans. Magn.*, 2004, **40**, (2), pp. 1196–99
9 Xiao, S., Rotaru, M., Sykulski, J.K.: 'Exploration versus exploitation using Kriging surrogate modelling in electromagnetic design', *COMPEL*, 2012, **31**, (5), pp. 1541–51
10 Xiao, S., Rotaru, M., Sykulski, J.K.: 'Strategies for balancing exploration and exploitation in electromagnetic optimisation', *COMPEL*, 2013, **32**, (4), pp. 1176–1188
11 Sykulski, A., Adams, N., Jennings, N.: 'On-line adaptation of exploration in the one-armed bandit with covariate problem'. Proc. Ninth Int. Conf. Machine Learning and Applications, 2010
12 Hawe, G., Sykulski, J.K.: 'Considerations of accuracy and uncertainty with Kriging surrogate models in single-objective electromagnetic design optimisation', *IET Sci. Meas. Technol.*, 2007, **1**, (1), pp. 37–47
13 Xiao, S., Rotaru, M., Sykulski, J.K.: 'Adaptive weighted expected improvement with rewards approach in Kriging assisted electromagnetic design', *IEEE Trans. Magn.*, 2013, **49**, (5), pp. 2057–2060
14 Alotto, P.G., Baumgartner, U., Freschi, F., et al.: 'TEAM workshop problem 22', http://www.compumag.org/jsite/images/stories/TEAM/problem22.pdf
15 Santner, T.J., Williams, B.J., Notz, W.I.: 'The design and analysis of computer experiments' (Springer, 2003)

16 Ho, S.L., Yang, S.Y., Ni, P.H., Wong, K.F.: 'An adaptive interpolating MLS based response surface model applied to design optimisations of electromagnetic devices', *IEEE Trans. Magn.*, 2007, **43**, (4), pp. 1593–1596

17 Holland, J.H.: 'Adaption in natural and artificial system' (MIT Press, Cambridge, 1975)

18 Hu, N.: 'Tabu search method with random moves for globally optimal design', *Int. J. Numer. Methods Eng.*, 1992, **35**, pp. 1055–1071

19 Ho, S.L., Yang, S.Y., Ni, G.Z., Wong, H.C.: 'An improved Tabu search for the global optimizations of electromagnetic devices', *IEEE Trans. Magn.*, 2001, **37**, (5), pp. 3570–3574

20 Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P.: 'Optimization by simulated annealing', *Science*, 1983, **220**, pp. 671–680

21 Hajji, O., Brisset, S., Brochet, P.: 'A new Tabu search method for optimization with continuous parameters', *IEEE Trans. Magn.*, 2004, **40**, (2), pp. 1184–1187

22 Yang, S.Y., Bai, Y.N., Zhang, G.H., Vecchi, L.Wu.: 'An improved population-based incremental learning method for inverse problems'. Automation Congress, 2008, pp. 1–4

23 Xiao, S., Rotaru, M., Sykulski, J.K.: 'Correlation matrices in Kriging assisted optimisation of electromagnetic devices', *IET Sci. Meas. Technol.*, 2014, doi: 10.1049/iet-smt.2014.0194.