Balancing exploration and exploitation in robust multiobjective electromagnetic design optimisation

by

Song Xiao

Thesis for the degree of Doctor of Philosophy

June 2014
DECLARATION OF AUTHORSHIP

I, Song Xiao, declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

Balancing exploration and exploitation in robust multiobjective electromagnetic design optimisation

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. [Delete as appropriate] None of this work has been published before submission [or] Parts of this work have been published as: [please list references below]:

Signed: .................................................................................................................................................................

Date: ..................................................................................................................................................................
ABSTRACT

Design optimisation of electromagnetic and electromechanical devices is usually aided by numerical simulations, such as the finite element method, which often carry high computational costs, especially if three-dimensional transient modelling is required. Thus in addition to the task of finding the global optimum, while avoiding local minima traps, there is the additional requirement of achieving the final solution efficiently with as few objective function evaluations as possible. With this in mind several surrogate modelling techniques have been developed to replace, under controlled environment, the computationally expensive accurate field modelling by fast approximate substitutes. This thesis looks at a particular technique known as kriging, which in other applications has been demonstrated to provide accurate representations, even of complicated responses, based on a limited set of observations whilst providing an error estimate of the predictions and hence increasing the confidence in the answer. In the iterative optimisation process the critical issue is where to position the next point for evaluation to find a sensible compromise between conflicting goals to explore the search space thoroughly but at the same time exploit information already available. This thesis proposes several novel algorithms based on reinforcement learning theory using the concept of rewards for balancing exploration and exploitation automatically and adaptively. The performance of these algorithms has been assessed comprehensively using carefully selected test functions and real engineering problems (taken from TEAM workshops) and compared with the results published in literature. The kriging approach has generally been found to outperform significantly other available methods.

One of the practical limitations, however, was found to be large-scale multi-dimensional or multi-objective tasks because of the need to create special correlation matrices for the kriging predictions to work. Several techniques have been developed and implemented to alleviate such problems and control the memory space occupied by such matrices.

Finally, in practical design problems, the issue of robustness of the design has to be considered – related to manufacturing tolerances, material variability, etc – which requires the designer not only to find the theoretical optimum but also assess its quality (sensitivity) within specified uncertainties of variables. Several strategies for evaluation of design robustness assisted by kriging modelling have been developed and implemented in combination with commercial electromagnetic design software.
Acknowledgements

Thanks firstly to my parents and my finance for providing support throughout the whole process of my PhD.

Thanks to my supervisor Professor Jan Sykulski and Dr Mihai Rotaru; without their great enthusiasm, encouragement and support, this would not have happened. Meanwhile thanks to Dr Adam Sykulski for his constructive suggestions.

Thanks to my examiner Professor Alun Vaughan for his helpful advice about my research in my 9 month and transfer reports.

Thanks to Professor Paul Lewin and Dr Kevin Goddard for all their help over the past three years. A mention of thanks to the rest of the EEE group. Thanks to everyone who made the past three years fly by.

Thanks to Professor Andy Kean and Dr David Toal in the faculty of engineering for his help to my research at the final stage.
# Contents

DECLARATION OF AUTHORSHIP .............................................................................. I

ABSTRACT .................................................................................................................. II

Acknowledgements .................................................................................................... III

List of Figures .............................................................................................................. V

List of Tables ............................................................................................................... IX

List of Abbreviations ................................................................................................ X

Chapter 1 Introduction ............................................................................................... 1
  1.1. Research Background ...................................................................................... 1
  1.1.1. Research Motivation .................................................................................. 1
  1.1.2. Electromagnetic Design Problem ............................................................... 2
  1.2 Research Objectives and Aims ......................................................................... 4
  1.3 Thesis Overview ............................................................................................... 6

Chapter 2 A Review of Existing Optimisation Techniques and Robust Design
Optimization ............................................................................................................. 9
  2.1 Single-Objective Optimization ......................................................................... 9
  2.2 Multi-Objective Optimization ......................................................................... 10
  2.3 The Essential Features of Test Objective Function ......................................... 11
  2.4 An Overview of Optimisation Methodologies .................................................... 13
  2.5 Robust Design Optimization ........................................................................... 15

Chapter 3 Surrogate Modelling ............................................................................. 17
  3.1 Introduction ....................................................................................................... 17
  3.2 General Theory of Surrogate Modelling ............................................................ 18
    3.2.1 Polynomial Models .................................................................................. 18
    3.2.2 Basis Functions ...................................................................................... 19
  3.3 Kriging-Assisted Single-Objective Optimization ............................................... 20
    3.3.1 Background ........................................................................................... 20
  3.4 Design and Analysis of Computer Experiments ............................................... 21
8.1 Introduction ........................................................................................................................................ 88
8.2 Kriging with different strategies for balancing exploration and exploitation .......................... 88
  8.2.1 SMWEI with multi-weights in the pre-test ............................................................................ 89
  8.2.2 SMWEI with the strategy of adaptively tuning weights ....................................................... 91
8.3 Practical performance of kriging with EI in testing the two-variable Schwefel function.. 93
  8.3.1 The practical performance of the kriging with normal EI..................................................... 94
8.4 Practical performance of SMWEI with multi-weights in pre-test.............................................. 95
  8.4.1 Practical performance of SMWEI with the strategy of adaptively tuning weights ......... 96
8.5 A robust design ............................................................................................................................... 97
  8.5.1 Testing the algorithms for the quality of the prediction of the robustness ..................... 98
8.6 Conclusion ......................................................................................................................................... 103

Chapter 9 Kriging with Adaptive Weighted Expected Improvement with Rewards

Approach in Electromagnetic Design ................................................................................................. 104
  9.1 Introduction .................................................................................................................................... 104
  9.2 Balancing Exploration and Exploitation ...................................................................................... 104
  9.3 Numerical Experiments ................................................................................................................ 105
  9.4 TEAM Workshop Problem 22 ..................................................................................................... 112
  9.5 Results for TEAM Workshop Problem 22 ................................................................................. 113
    9.5.1 The Two-Parameter TEAM Workshop Problem 22 ......................................................... 113
    9.5.2 The Three-Parameter TEAM Workshop Problem 22 ....................................................... 115
  9.6 Conclusion ..................................................................................................................................... 120

Chapter 10 Robust Global Optimization of Electromagnetic Designs Utilizing Gradient

Indices and Kriging ............................................................................................................................... 121
  10.1 Introduction ................................................................................................................................. 121
  10.2 Robust optimisation ..................................................................................................................... 122
    10.2.1 Multi-Objective Robust Optimisation using GI ................................................................. 122
  10.3 Robust optimisation based on kriging modelling ..................................................................... 127
    10.3.1 Kriging .................................................................................................................................. 127
    10.3.2 Partitioning of Correlation Matrices .................................................................................. 129
  10.4 Application in electromagnetic design ....................................................................................... 131
    10.4.1 TEAM Problem 22 ............................................................................................................ 131
    10.4.2 Two-parameter Test Results ............................................................................................. 131
  10.5. Conclusion .................................................................................................................................. 138
Chapter 11 Further Considerations of Uncertainty in Robust Optimization of Electromagnetic Devices

11.1 Introduction ................................................................................................................. 139
11.2 Robust Optimization Algorithms exploiting Kriging Modelling ................................. 139
11.3 The Prediction of Kriging surrogate model .................................................................. 142
11.4 Average performance ................................................................................................. 143
11.5 Robust Optimization Algorithms exploiting Kriging Modelling ................................. 147
11.6 Application in Electromagnetic Design ...................................................................... 150
11.7 Conclusion .................................................................................................................. 154

Chapter 12 Conclusions & Recommendations for Further Work ........................................ 155
12.1 Conclusions ................................................................................................................ 155
12.2 Recommendations for Further Work ......................................................................... 157

Appendix A The Concepts of Multi-Objective Optimization .................................................. 159

Appendix B Conference Papers .......................................................................................... 162

Appendix C Journal Papers ................................................................................................ 190

Bibliography ....................................................................................................................... 270
List of Figures

Figure 1.1: The usage of surrogate model in optimization process. .......................... 4
Figure 2.1: Example of a feasible region and a feasible objective space in two dimensions ...................................................................................................................... 10
Figure 2.2: An example of a unimodal function ................................................................ 11
Figure 2.3: An example of a multimodal function (One-dimensional Rastrigin function) .................................................................................................................................. 12
Figure 2.4: An example of a multimodal function (One-dimensional Schwefel function) .................................................................................................................................. 12
Figure 2.5: Objective function and the global and local minima, where the grey region represents variation of design variables .......................................................................................................................... 15
Figure 4.1: Iterations of the expected improvement approach on the Schwefel test function ........................................................................................................................................... 33
Figure 4.2: Latin hypercube, n=10 ................................................................................. 35
Figure 4.3: The performance of Kriging surrogate model with EI based on only one initial sample point ......................................................................................................................... 36
Figure 4.4: (a) The objective function (0 < x1 < 27, 0 < x2 < 27); (b) The prediction performance of kriging model with seven initial sampling points; (c) The prediction performance of kriging model with only the two initial sampling points selected from the above seven sampling points; (d) The prediction performance of kriging model with two special initial sampling points .......................................................................................................................... 38
Figure 5.1: The performance of kriging model with GEI for g = 1 at the first iteration. 43
Figure 5.2: The performance of kriging model with GEI for g = 2 at the first iteration. 43
Figure 5.3: The performance of kriging model with GEI for g = 5 at the first iteration. 44
Figure 5.4: The performance of kriging model with WEI for w=0.1 at the first iteration .............................................................................................................................. 47
Figure 5.5: The performance of kriging model with WEI for w=0.5 at the first iteration .............................................................................................................................. 48
Figure 5.6: The performance of kriging model with WEI for w=0.9 at the first iteration .............................................................................................................................. 48
Figure 5.7: Performance of kriging model with WEI when $w=0.1$ .............................................. 51
Figure 5.8: Performance of kriging model with WEI when $w=0.4$ .............................................. 52
Figure 5.9: Performance of kriging model with WEI when $w=0.9$ .............................................. 52
Figure 5.10: Performance of kriging surrogated model using the WEI when $w=−0.01$ 54
Figure 5.11: Performance of kriging surrogated model using the WEI when $w=−0.1$ .... 54
Figure 5.12: Performance of kriging surrogated model using the WEI when $w=−1$ ....... 55
Figure 5.13: Performance of kriging surrogated model using the WEI when $w=−10$ . 55
Figure 5.14: Performance of WEI when $w=1.1$ ................................................................. 56
Figure 6.1: A transition graph of the reinforcement learning ............................................. 59
Figure 6.2: The transition of Adaptive Weighted Expected Improvement (AWEI)...... 62
Figure 6.3: The decision-making chart of the Adaptive Expected Improvement with the remedial measurement ........................................................................................................ 65
Figure 6.4: The Adaptive Expected Improvement for the tuneable value $\beta = 0.001$ ..... 67
Figure 6.5: The Adaptive Expected Improvement for the tuneable value $\beta = 0.05$ ...... 68
Figure 6.6: Performance of kriging model with AWEI for $\beta = 0.1$ ........................................... 68
Figure 6.7: Performance of the kriging model with AWEI for $\beta = 0.3$ ......................... 69
Figure 6.8: The best performance of AWEI when the tuneable parameter $\beta$ is equal to 0.35................................................................................................................................. 70
Figure 6.9: Performance of a kriging model with AWEI for $\beta = 0.40$ ......................... 70
Figure 6.10: Performance of a kriging model with AWEI for $\beta = 0.6$ ......................... 71
Figure 6.11: Performance of a kriging model with AWEI for $\beta = 0.8$ .......................... 71
Figure 6.12: A summary of the test results using AWEI for different values of $\beta$. ...... 72
Figure 7.1: The SMWEI for making decision based on long-term rewards for balancing exploration and exploitation................................................................. 75
Figure 7.2: The two versions of simplified surrogate model obtained by SMWEI (with smoothing treatment and without it) at the first, fifth, tenth iterations respectively ...... 80
Figure 7.3: The performance of the kriging with SMWEI when with $w_1=0.7$ and $w_2=0.3$. .................................................................................................................................................. 81
Figure 7.4: The ‘best’ performance of SMWEI with $w_1=0.7$ and $w_2=0.1$ ............... 82
Figure 7.5: The performance of the kriging with SMWEI when coping with an ‘extreme’ case ........................................................................................................................................... 84
Figure 7.6: The performance of the kriging assisted by the EI when coping with an ‘extreme’ case ...................................................................................................................... 85
Figure 7.7: The performance of the kriging with SMWEI when coping with an ‘extreme’ case ................................................................. 85
Figure 7.8: The performance of the kriging assisted by EI when coping with an ‘extreme’ case .................................................................................. 86
Figure 8.1: The SMWEI with more optional weights in pre-test for achieving the optimal weights .................................................................................................................. 91
Figure 8.2: The proposal of using the pre-test strategy of AWEI instead of the current pre-test.......................................................................................................................... 92
Figure 8.3: 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 ........................................................................................................ 93
Figure 8.4: The performance of the kriging with EI to predict the two-variable Schwefel function ......................................................................................................................... 94
Figure 8.5: The performance of the kriging assisted by SMWEI with multi-weights to predict the two-variable Schwefel function ................................................................................. 95
Figure 8.6: The performance of the kriging assisted by SMWEI with adaptively tuning weights to predict the two-variable Schwefel function ........................................................................... 97
Figure 8.7: The objective function for the robust test .............................................................................................................................. 98
Figure 8.8: (a) The performance of the kriging with EI when facing the robust problem (b) The error between the robust test function and prediction by kriging model with EI .......................................................... 99
Figure 8.9: (a) The performance of the kriging assisted by SMWEI with adaptively tuning, (b) the error between the test function and the kriging prediction assisted by SMWEI with adaptively tuning weights when facing the robust problem .......... 101
Figure 8.10: (a) The performance of the kriging assisted by SMWEI with multi-weights when facing the robust problem, (b) The error between the robust test function and prediction by the kriging model assisted SMWEI with multi-weights .............................................................. 102
Figure 8.11: The detailed situation of choosing optimal weights at each iteration ..... 102
Figure 9.1: The single-variable analytic function [57], [58] ..................................................................................................................... 106
Figure 9.2: A two-variable objective function ........................................................................................................................................ 107
Figure 10.1: Example of a robust design for a one-variable problem ................. 124
Figure 10.2: Two minima and a plateau .............................................................. 127
Figure 10.3: The correlation matrices partitioned by: (a) sampling points, .......... 130
Figure 10.4: The flowchart of the adaptive division of correlation matrices .......... 130
Figure 10.5: TEAM Workshop Problem 22 ...................................................... 132
Figure 10.6: The 3 parameter TEAM 22 problem (a) EI, (b) AWEI. ................................. 134
Figure 10.7: Objective function trajectory showing sensitivity and objective function values (OF) (a) EI, (b) AWEI. ......................................................................................... 135
Figure 10.8: The specific parameter setting and transient process of testing the TEAM 22 problem with the help of the novel partitioning scheme................................................................. 136
Figure 10.9: Monitoring of computing times and memory requirements.................. 137
Figure 11.1: (a) The constrained optimization problem, (b) The worst case optimization (WCO), (c) The worst-vertex-based WCO. .................................................................................................................. 141
Figure 11.2: (a) The objective function including a smoothly shaped local minimum, (b) the objective function including a local minimum with ripples, (c) the objective function including a non-robust local minimum with nearby ‘peak’ ................................................................. 145
Figure 11.3: (a) Analytic function (b) The kriging prediction (c) The performance of the WCO method (d) Average performance (e) The full-field solution for three objectives (f) The zoomed-in optimal part ................................................................................................................................. 149
Figure 11.4: The three objective optimization including ........................................ 152
Figure 11.5: The three objective optimization (U(J1, J2)=0.2) ........................................ 153
Figure 11.6: The three objective optimization (U(J1, J2)=0.35) ................................. 153
List of Tables

Table 4.1: (a) The setting of seven initial sampling points (b) The setting of two initial sampling points selected from the first initial setup (c) The setting of two special initial sampling points ..........................................................37
Table 4.2: The performance of the component favouring exploration ........................................39
Table 4.3: The performance of the component favouring exploitation ....................................39
Table 5.1: Value of $g$, the number of iterations of GEI on the Schwefel test function ..........45
Table 5.2: Value of $g$, as used in built-up scheme in [43]. ..........................................................45
Table 5.3: The efficiency of WEI for different values of $w$ ......................................................50
Table 5.4: Performance of WEI with the four typical negative weighting parameters ............56
Table 6.1: Performance of kriging model with AWEI for different value of tuneable parameter $\beta$ ....................................................................................................................................66
Table 8.1: The optimal weights of the kriging assisted by SMWEI with multi-weights at each iteration within the iterative process ........................................................................96
Table 8.2: Initial sampling points ..............................................................................................99
Table 9.1: Initial Sampling Points ..........................................................................................107
Table 9.2: The ‘zoom in’ strategy using kriging assisted EI for TEAM 22 tests ..................114
Table 9.3: The Three-Parameter Kriging Assisted AWEI for TEAM 22 Tests .....................115
Table 9.4: Genetic Algorithm (GA) [61]; Tabu Search (HuTS) [62]; Improved Tabu Search (ITS) [63]; Simulated Annealing Algorithm (SA) [64]; New Tabu Search (NTS) [65]; Population-based Incremental Learning (PBIL) [66] ...........................................................................................................120
Table 10.1: Values of the objective function, second-gradient and sensitivity for the three design points A1, A2 and A3 (Figure 10.2) .........................................................................................126
Table 10.2: The setting of initial sampling points ..................................................................131
Table 10.3: Performance comparison of algorithms ...............................................................131
Table 10.4: The setting of initial sampling points .................................................................133
Table 10.5: The specific definition of test .............................................................................133
Table 11.1: The common features of the two functions .......................................................145
Table 11.2: The initial setup of the prediction by the kriging model ..................................151
Table 11.3: Results for the case when uncertainty is 0.1 .....................................................152
Table 11.4: Results for the case when uncertainty is 0.2 .....................................................152
Table 11.5: Results for the case when uncertainty is 0.35 ....................................................153
List of Abbreviations

FDM Finite Difference Method

FEM Finite Element Method

BEA Boundary Element Analysis

NFL No Free Lunch

SOOP Single-Objective Problem

FEM Finite Element Method

GA Genetic Algorithm

MLE Maximum Likelihood Estimate

DACE Design and Analysis of Computer Experiments

MSE Mean Square Error

EI Expected Improvement

WEI Weighted Expected Improvement

GEI Generalized Expected Improvement

GM Global Minimum

LM Local Minimum

AWEI Adaptive Weighted Expected Improvement

AMSE Average Mean Square Error

SMWEI Surrogate Model based on Weighted Expected Improvement

RMSE Root Mean Square Error

OF Objective Function

GI Gradient Index

WCO Worst Case Optimisation
Chapter 1 Introduction

1.1. Research Background

1.1.1. Research Motivation

Electromagnetic designs almost always carry a heavy burden of high computational cost, with very few exceptions when very simplistic analytical, empirical or equivalent circuit based models are found to be adequate for performance prediction. Most of the time throughout the design process, or at least at later stages, numerical models are required to provide necessary accuracy, typically employing 3D simulations using finite element or related technique. In the optimisation part of the design routine a single objective function evaluation may require a full field solution of the entire complicated model, often transient, or even several solutions (if averaged values are needed), which may be very ‘expensive’ in terms of computing times involved. Thus it is not enough to have confidence that the algorithm finds the global optimum; for practical purposes it must do so with as few objective function calls as possible. Thus within the context of searching for the optimum (usually minimum) of a particular objective function (or functions in multiobjective problems, e.g. best performance and simultaneously minimum cost), another minimum is being sought, that is looking for a strategy which finds the optimum with a minimum use of the computationally expensive performance predicting software. To complicate things further, the issue of robustness of the design comes into consideration – related to manufacturing tolerances, material variability, etc – which requires the designer not only to find the optimum design but also know more about its ‘quality’, in other words the ‘shape’ of the objective function must be estimated. In the context of stochastic optimisation this is usually expressed in terms of a compromise between exploration (searching the uncharted space) and exploitation (using information already provided) and is often supplemented and supported by various types of surrogate models. This thesis investigates these issues and uses ‘kriging’ as the main technique for constructing the surrogate model.
1.1.2. Electromagnetic Design Problem

Electromagnetic design problems (as with all engineering design problems), may be categorised as two types [1]:

**Direct problems**, in which given the input or the cause of a phenomenon or of a process in a device, the purpose is that of finding the output or the effect [2].

**Inverse problems**, in which a design is sought which produces a specific effect, which may be measured or assumed. If the effect is measured, the problem is defined as an identification problem. But if the effect is assumed, the problem can be seen as a synthesis problem, to which a solution may not exist in practice [2].

To solve the direct problems in electromagnetics, various numerical methods are available, such as Finite Difference Method (FDM), Finite Element Method (FEM), Boundary Element Analysis (BEA) and so forth. However, normally these numerical algorithms can be computationally expensive (i.e. time-consuming), depending on the complexity of the design.

The solution of the inverse problem, and in particular the optimal design problem, is an active area of research in all branches of engineering such as mathematics, numerical analysis, software development and engineering design. The optimal design problem expresses the effects as objectives, which must be maximized or minimized, and thus is a synthesis problem. Normally it can only be achieved by solving the direct problem multiple times, which will cause higher computational cost than direct problems.

1.1.2 Solving Inverse Problems in Electromagnetic Design

It is natural to ask the question whether there is the best evolutionary algorithm that will always outperform all other methods regardless of the given optimization problem. The ‘No Free Lunch’ (NFL) theorem [3, 4] of optimization denies the possibility of the existence of such best-performed optimization algorithm and argues that, averaged over all types of problems, every algorithm performs the same [5]. However, the NFL theorem also states that it is possible to identify an algorithm which outperforms others.
over a particular domain of interest [6] (and conversely this algorithm may be outperformed by other algorithms on the problems outside this domain).

There is a wide range of algorithms which now exist for solving optimization problems. They may be categorized in different ways, for example as single-objective or multi-objective; deterministic or stochastic, greedy or cost effective and so forth. Electromagnetic design does not fall neatly into any of the above mentioned categories; moreover, depending on the complexity of the objective function, a single evaluation of objective function may take a couple of seconds or several days.

In practice, the electromagnetic design with a heavy burden of high computational cost is one of the most common challenges in modern electromagnetic designs. Most of the time during the design process, or at least at later stages, numerical algorithms are required to construct the model to provide necessary accuracy, typically employing 3D simulation using finite element or related technique. In the optimisation part of the design routine a single objective function evaluation may require a full field solution of the entire complicated model, often even several solutions (if averaged values are needed), which may be ‘expensive’ (time-consuming) as presented in the outer loop of the Figure 1.1. Thus it is not enough to have confidence that the global optimum of objective function is found by the algorithm; for practical purpose it has to do so with as few objective function calls as possible. Thus we are looking for a strategy that can find the optimum with minimum use of the computationally expensive modelling tools (such as FEM) for finding the objective function.

An alternative technique called “surrogate modelling” [7] uses a cheaper model of the true problem which is evaluated by the global optimiser. This surrogate model is constructed from an initial design of experiments based sampling of the problem and attempts to model the response of the objective function as shown in the nested loop of the Figure 1.1. The prediction of the objective function from this model is considerably cheaper than the evaluation of the true objective through, for example, a Finite Element simulation. This surrogate model is however an approximation to the true response and as such requires constant updating in regions of interest, but on the whole, considerably fewer evaluations of the true objective function are required to achieve the optimum.
As one of regression models, kriging surrogate modelling is able to approximate complex multi-dimensional functions based only on limited information and estimate the accuracy of this approximation indirectly based on the mean square error produced by the surrogate model itself, which is helpful in assisting the main decision of the optimisation process to locate to the global optimum in design space [8]. The flowchart presented below focuses specifically on the use of kriging surrogate modelling with applications in electromagnetic design.

![Flowchart](image)

Figure 1.1: The usage of surrogate model in optimization process.

### 1.2 Research Objectives and Aims

The objectives of this research were four-fold:

1. to construct the kriging surrogate model;

2. to maintain good balance between exploration and exploitation throughout the process of searching for the updated sampling points for further evaluation;

3. to attempt to mitigate the data storage burden caused by correlation matrices in the kriging model, especially when solving highly dimensional problems;
4. to consider the robustness of optimal solutions achieved by the kriging model. There is a strong correlation and even overlap between these targets.

In order to reduce the high computational cost associated with electromagnetic designs, we are looking for a surrogate model which is able to find the optimum with a minimum use of computationally expensive performance prediction software. With the help of this surrogate model, a reduction in construction effort allows designers to reallocate time to evaluate the true objective function rather than only applying the costly high fidelity simulation. The initial focus of this research was to construct the kriging surrogate model which could seamlessly link with commercial electromagnetic design software, and expand on the existing utility functions to assist the kriging surrogate model in updating the sampling points for further evaluation. The feasibility of different utility functions needs to be verified and compared with each other in terms of their capability to converge to the global optimum of the objective function.

The existing utility functions do not have the flexibility to balance exploration and exploitation; this impacts the effectiveness and efficiency of locating the global optimum throughout the optimisation process. If placing emphasis on exploitation, it may be helpful to accelerate the process of converging to the existing minimum. However it might also be risky that the kriging model is trapped by certain sub-optimal area or local minima rather than the true global minimum. If too much emphasis is on exploration, it might cause the efficiency of searching global optimum to decline. Thus a novel strategy that is able to balance exploration and exploitation adaptively and automatically along with the optimization process is essential and this was one of the aims of this project.

Within the optimization process, the response surface produced by the kriging surrogate model is updated successively in the term of updating sampling points following with certain guidance. With the sampling points increasing, the data size of the correlation matrices enlarges considerably especially when coping with the high-dimensional optimization tasks. As a critical bottleneck, the difficulty of maintaining the data of the correlation matrices to a manageable level has to be overcome.

The final part of research mainly concerns the robust design of electromagnetic devices under the inevitable perturbation produced by various elements in practical situations.
Chapter 1 Introduction

The uncertainty existing in design variables causes that in addition to pursuing the best-performed solution the robustness of solution needs to be considered as well. A reliable methodology of robustness evaluation, which is able to cooperate with the kriging surrogate model without wasting valuable function evaluations is explored and introduced.

1.3 Thesis Overview

The thesis commences with a review of the existing optimisation techniques. The concept of the optimization and a set of related generic terms will be reviewed. These concepts include the single-objective and multi-objective optimization, as well as the Pareto-optimality.

A review of the general definition of surrogate modelling and a set of frequently used basis functions forming the surrogated model are presented. The formulation process of the kriging surrogate model is explained and its mean square error is emphasised, which can be involved to inform the model where to select the next updated sampling points for further evaluate in order to locate the global optimum as well as to improve the current response.

The implementation of the kriging surrogate model assisted by one utility function called the Expected Improvement (EI) to find the global optimum of the objective function is then introduced. The exploration and exploitation terms composing the EI are tested respectively, and then either of the effects will be analysed. Finally, the importance of balancing these two terms when attempting to find the global optimum with minimum use of information from objective function is discussed.

Furthermore, two currently existing methods, named the Weighted Expected Improvement and Generalized Expected Improvement, respectively, developed from the Expected Improvement utility function, but with their own tuneable weighting parameters for pre-distributing certain weights on exploration and exploitation respectively are verified into numerical tests. Different setting of the values of the tuneable parameter is taken to observe the difference of their performance and compare their efficiency.
Having determined the impact brought from different weights on exploration and exploitation throughout the optimisation process, new algorithms are then considered. Ideas from reinforcement learning were used in combination with the weighted expected improvement method in order to automate the process of tuning the weights put on exploration and exploitation using a response from a pre-test. As the response, the reward obtained via the calculation of the variation of mean square error determines how to redistribute the weights on exploration and exploitation. Two novel algorithms possessing the ability of tuning weights on exploration and exploitation adaptively are presented and their application is demonstrated.

A three dimensional electromagnetic design problem called TEAM benchmark problem 22 is employed within this study and provides an excellent opportunity to demonstrate the effect of increasing dimensionality on the performance of an optimization, including the impact on data storage requirements. The increasing size of the correlation matrices, while the optimization is progressing, may result in a ‘combinatorial explosion’, especially when dealing with high-dimensional optimization. Thus to mitigate the burden of storing massive data, a scheme of partitioning the correlation matrices to manageable size adaptively has been considered. Then with the strategy of partitioning correlation matrices the modified kriging model can attempt to cope with the task with large-scale data especially the multi-dimensional problem as the issue of constantly accumulating data can be solved to some extent.

In practical electromagnetic designs, uncertainties existing in the design variables are normally inevitable; hence in addition to the optimal value of the objective function the robustness of the optimal solutions have to be considered. In Chapter 10 the conventional optimization process is transformed into a multi-objective optimization by concurrently maximizing the robustness of solutions. One popular methodology, called the ‘Gradient Index (GI)’, is used but then, due to certain limitations of this method, an improved version is proposed and implemented.

Furthermore, the Worst Case Optimisation method (WCO) is another popular methodology to solve robust optimization problems. As an approach for evaluating the reliability of design solutions, the WCO has the ability of predicting the worst case which might happen within the uncertain range of variables and then, through
comparison, the relatively robust solution could be picked out; the use of kriging surrogate model might be beneficial here to save computing time.

Moreover, the worst case algorithm can be improved further by considering average performance within the uncertain range. As shown in this thesis this will transform the problem into three-objective optimisation task.

Finally, the techniques developed throughout the course of this research work are combined and implemented for solving practical electromagnetic optimisation problems. The last chapter summarises the performance gains achieved by the presented research and an overview of the research followed by suggestions for the future work.
Chapter 2 A Review of Existing Optimisation Techniques and Robust Design Optimization

2.1 Single-Objective Optimization

The Single-Objective Optimization Problem (SOOP) may be stated as follows:

\[
\begin{align*}
\text{Minimize } & f(x) \\
\text{Subject to } & g_j(x) \geq 0 & j = 1,2, \ldots, J; \\
& h_k(x) = 0 & k = 1,2, \ldots, K; \\
& x_i^{(l)} \leq x_i \leq x_i^{(u)} & i = 1,2, \ldots, d;
\end{align*}
\]

This problem is said to be unconstrained if \( J = K = 0 \); otherwise it is constrained. Each design variable has been restricted in a range, which has a lower and an upper bound varying continuously (however normally the design variable may only be allowed to vary discretely [9]). Therefore in order to be feasible, a design vector needs to satisfy the \( J \) inequality and \( K \) equality constraints. This gives rise to the feasible region \( S \) which is a subset of the decision variable space \( R^d \) [10],

\[
S = \{ x \in R^d \quad s.t. \quad g(x) \geq 0 \quad \text{and} \quad h(x) = 0 \},
\]

and corresponding to the feasible region the feasible objective space \( Z \) which is a subset of the objective space \( R \),

\[
Z = \{ f(x) \in R \quad s.t. \quad x \in S \}.
\]
2.2 Multi-Objective Optimization

The typical Multi-objective optimization problem with aim of minimizing an objective function \( f_m(x) \), subject to a set of constrains \( g_j(x) \), is expressed as

\[
\text{Minimize } f_m(x) \quad m = 1, 2, \ldots, M; \\
\text{Subject to } g_j(x) \geq 0 \quad j = 1, 2, \ldots, J; \\
\]

\[
x_i^{(U)} \quad k = 1, 2, \ldots, K; \\
x_i^{(L)} \leq x_i \leq x_i^{(U)} \quad i = 1, 2, \ldots, d.
\]

In order to make \( f_m(x) \) feasible, the design vector must satisfy the \( J \) inequality and \( K \) equality constraints. The values \( x_i^{(L)} \) and \( x_i^{(U)} \) denote the lower and upper bounds of design variable vector, respectively. An example of a feasible region and feasible objective space \( Z \) is presented for a two-dimensional problem in Figure 2.1 [11]. This graphical illustration is only available when the number of objectives or dimension of the design vector is less than or equal to three, however, if \( d>3 \) such representation becomes impossible.

![Diagram](image-url)

**Figure 2.1**: Example of a feasible region and a feasible objective space in two dimensions
2.3 The Essential Features of Test Objective Function

It is essential to test an optimization algorithm well before it is used in real optimisation problems; hence the choice of an appropriate testing objective function becomes very important. Five important features of such a function have been identified in [11] and are described in the following paragraphs.

First, an important property affecting the difficulty of a global optimization problem is the degree of modality of the objective function $f$, which is the number of local minima of the objective function. For instance, the global minimum of a unimodal function can be detected by a single run of a local search routine, whereas highly multimodal functions provide a more difficult test to the optimization routine [12]. The examples of unimodal function and multimodal function are presented respectively in Figure 2.2 and Figure 2.3. The global minimum of a unimodal function with only one minimum (without other local minima) is easier to be found than the multimodal with multiple local minima.

![Figure 2.2: An example of a unimodal function](image-url)
Chapter 2 A Review of Existing Optimisation Techniques and Robust Design Optimization

Figure 2.3: An example of a multimodal function (One-dimensional Rastrigin function)

The difficulty of a global optimization problem does not only depend on the number of local minima, but also on how they are distributed in the search place. For example, functions like the Rastrigin function in Figure 2.3 appear to be easier to optimize than the Schwefel in Figure 2.4, in spite of the fact that all of these functions have a comparable number of local minima.

Figure 2.4: An example of a multimodal function (One-dimensional Schwefel function)

Besides the degree of modality of the objective function, there are still several elements required to consider such as the size of basins of attraction of local minimum, the size of improving regions (and the magnitude of oscillations), the degree of randomness in
positions of the minima and the dimension of the search space, which are demonstrated in [12] explicitly.

It has to be mentioned here that in many electromagnetic optimisation problems there is no a priori knowledge on the shape of the objective function; hence the problem of choosing the right algorithm is more difficult.

The efficiency of our model which will be illustrated later has been tested with the Schwefel Test Function. The above figure shows the one-dimensional Schwefel Test Function. As can been seen, it includes one global minimum and several local minima, which are distributed irregularly. This function provides a stern test to optimization algorithms and it has been used by many authors in testing the efficiency and robustness of their algorithms.

2.4 An Overview of Optimisation Methodologies

In general, the optimisation of any engineering design task usually commences with certain initial design solution, which could be the result of previous design process or the culmination of engineers’ test experience of what should constitute a good design. Then, based on this given information, designers would attempt to constantly improve an objective function by altering relevant parameters. This results in a gradual descent towards an optimum; however this so-called optimum is usually only local in the whole design space and while the objective function may be smaller than all other points in the vicinity, it may not be the global optimum. Thus local optimisation techniques are unlikely to find out the global optimum unless fortunately the initial sampling points are in the region of the global optimum, or the objective function is unimodal in nature. But even with such limited capability, local optimisation techniques are still a popular optimisation tool in engineering design, as they require relatively few objective function calls compared to global stochastic optimisation methods and can make effective use of gradient or curvature information existing between existing design points.

There are various kinds of local optimisation techniques throughout the literature like Nelder-Mead simplex algorithm [13], conjugate gradient methods, pure Newton methods[14], quasi-Newton methods [15], and so forth, each of which are frequently used by many authors in engineering design or mathematic area.
Compared with local optimisation methodologies, the stochastic optimisation methods are normally used as global optimisers that are applicable to nonlinear and unconstrained optimisation problem. The genetic algorithm (GA) is one of the most popular stochastic optimisation methodologies at present, developed by [16]. This algorithm is inspired by the Darwinian theory of natural selection, whereby desirable or beneficial traits become more common as a reproducing population, whereas undesirable or disadvantageous traits would be abandoned. Like most other stochastic optimisation methodologies, it employs a population that gradually evolves through a number of generations. The evolution process employed by this method mainly contains three elements, “selection”, “crossover” and “mutation”, which attempt to move the members of population gradually to the desirable regions of a design space and then converge to a global optimum. But the third procedure, “mutation”, involves the alternation of a randomly selected variable by certain proportion of the offspring and in other word it requires a random walk through the design space to guide the optimiser escape the regions containing local optimum to do some exploration. But this process is controlled by four important tuneable parameters: the size of the population, the probability of cross-over, the probability of mutation and the number of generations. And any of them greatly affects the effectiveness and efficiency of a global optimisation; unfortunately in order to achieve the best-performed values of these parameters considerable time and effort are necessary [17], [18].

In addition, the particle swarm is also one of the most popular stochastic optimisation methods developed by [19]. Like the genetic method, the particle swarm is also inspired by nature, which endeavours to model the social behaviour of a population of animals. In the process of simulation every member of the population is able to remember its best performance so far, moreover each member could know the global best performance that another member of the population had found. Like the genetic method, particle swarm do suffer from the exploitative deficiencies. They are capable of finding the region containing an optimum but not the precise position of the global optimum. Besides these two typical algorithms, there still are a number of other methods and the hybrid methods derived from them. As a whole, stochastic methods can be relied upon to search a global optimum; however they typically require a very large number of objective function calls. This may not be an issue when testing the simple analytical functions or fast low fidelity computational simulations, but in electromagnetic design
the calculation of finite element models normally are time-consuming. Thus the time for global optimisation cost by a stochastic method can be extremely significant. So a cheap surrogate model which can utilize very limited given information to find out the global optimum precisely and efficiently is essential and this technique is demonstrated from next chapter.

### 2.5 Robust Design Optimization

For an electromagnetic design problem, uncertainties in design variable such as manufacturing tolerance, variation of material properties from their nominal values are inevitable, so an optimal solution must consider the robustness of the design. Our research does not only aim to utilize the surrogate modelling techniques to solve the global optimisation problem but exploring design characteristics influenced by uncertainties in design variables also needs to be considered.

An illustration of one analytic example to be minimized is shown in Figure 2.5, meanwhile the global minimum A and two local minima B and C have been labelled. Obviously the best-performed point A can be selected as the global optimum; however the uncertainty in variable $x$ may result in a large variation of the objective function. Thus the theoretical best performance A is not a robust design and points B or C might be preferred.

![Figure 2.5: Objective function and the global and local minima, where the grey region represents variation of design variables.](image)

15
To deal with the impact produced by uncertainties in design variables, there are various kinds of techniques [20]-[31] applicable. In the local sensitivity analysis method [20], after achieving the deterministic optimum via the optimization run, the sensitivity of the objective function with respect to the optimal design in the neighbourhood of the optimum is calculated. But this procedure may be inefficient that if the optimum is not satisfying, and then a new optimization run has to be implemented. The statistical information based approach, the ‘six sigma method’, attempts to optimize both the mean value and variance of performance in order to evaluate the robustness and the average performance of designs [21]-[24]. The worst-case method (WCO) and the worst-vertex-based WCO tries to optimize the worst performance in terms of the value of both objective function and constraints under uncertainty [25]-[29]. The gradient index technique [30]-[31] transforms an optimisation problem into a multi-objective optimisation by concurrently minimising the objective function value and the corresponding gradient index (the maximum gradient of objective function with respect to design variables subject to specific uncertainties) [30]-[31]. Among these methodologies, the gradient index method and the worst-case scenario optimization method are chosen as our research targets. To reduce computing times when searching for the robust solution the surrogate modelling techniques are used again in the context of the robust design.
3.1 Introduction

Generally many stochastic methods are able to be relied upon to reach a global optimum, but they typically require enormous amount of information from the objective function to do so. This might not be a serious issue when addressing simple analytical functions or problems that require computationally simple models to be solved for objective function evaluation. However, when it comes to the highly expensive simulation, the time required by an optimizer using a stochastic method is a critical element which needs to be considered. Even with the help of the parallel computation techniques to calculate the objective function value it may be infeasible and unreliable to use a stochastic method of optimisation for some complicated problems.

In general, the problem of estimating dependencies using a set of limited observations is known as machine learning [32, 33]. This is a broad realm, which encompasses the problem of regression, where the inputs are mapped to a continuous function, classification and discrete categories [10]. Due to the general character, methods such as machine learning are used extensively in wide range of areas such as science, engineering and so forth.

Electromagnetic design almost always carries a heavy burden of high computational cost, with very few exceptions when a very simplistic analytical, empirical or equivalent circuit based model is found to be adequate for performance prediction. Most of the time throughout the design process, or at least at later stages, numerical models are required to provide necessary accuracy, typically employing 3D simulation using finite element or related techniques. In the optimisation part of the design routine a single objective function evaluation may require a full field solution of the entire complicated model, often transient, or even several solutions (if averaged values are needed), which may be very “expensive” in terms of computing times involved. It is not enough to have confidence that the algorithm finds the global optimum; for practical purpose it must do so with as few objective function calls as possible. Thus, within the context of searching for the optimum (usually minimum) of a particular objective function (or functions in
multiobjective problems, e.g. best performance and simultaneously minimum cost), another minimum is being sought, that is looking for a strategy which finds the optimum with a minimum use of the computationally expensive performance predicting software. To complicate things further, the issue of robustness of the design comes into consideration – related to manufacturing tolerances, material variability, etc. – which requires the designer not only to find the optimum design but also know more about its “quality”, in other words the “shape” of the objective function must be estimated. In the context of stochastic optimisation this is usually expressed in terms of a compromise between exploration (searching the unexplored space) and exploitation (using information already provided) and is often supplemented and supported by various types of surrogate modelling. This chapter mainly illustrates the theory of surrogate modelling, using the polynomial fitting as a start for further use of kriging surrogate model.

3.2 General Theory of Surrogate Modelling

3.2.1 Polynomial Models

To create a surrogate model the data from either physical or numerical experiments has to be fitted in some way using a mathematical description. Although there are many possible mathematical models and functions that could fit the available data the discussion in this chapter will be limited mostly to polynomial models.

The polynomial function is the simplest type of the surrogate model. The model involves only variables of the following form \( \hat{y}(x) = \beta_0 + \sum_{i=1}^{d} \beta_i x_i \), where \( \Sigma_{i=1}^{d} g_i \leq G, x_2, ..., x_d \) are the components of the d-dimensional design vector x and where \( \Sigma_{i=1}^{d} g_i \leq G, G \) being the order of the polynomial. For instance, the predicted value \( \hat{y} \) using a first order polynomial model is as follow:

\[
\hat{y}(x) = \beta_0 + \sum_{i=1}^{d} \beta_i x_i
\]  

(3.1)

Whilst using a second order polynomial model is given by

\[
\hat{y}(x) = \beta_0 + \sum_{i=1}^{d} \beta_i x_i + \sum_{j=1}^{d} \sum_{i=1}^{d} \beta_{d-1+i+j} x_i x_j,
\]  

(3.2)
where the value of the coefficients $\beta_i$ are to be determined. In general, denoting a basis of the set of all polynomials in $x$ of degree $G$ by $\{f_k(x) | k = 1, 2, ..., m\}$, the prediction made by a polynomial model may be written as

$$\hat{y}(x) = \sum_{k=1}^{m} \beta_k f_k(x)$$  \hspace{1cm} (3.3)

where the value of $\beta_k$ should be determined by fitting a polynomial such is minimizing an error function based on the observed data. One popular error function is the sum of the squares between the prediction $\hat{y}(x_k)$ and practical observations $y(x_k)$

$$E(\beta) = \frac{1}{2} \sum_{k=1}^{N} (\hat{y}(x_k) - y(x_k))^2$$  \hspace{1cm} (3.4)

Minimization of the above function is known as least-squares fitting. Nevertheless, interpolation can only be achieved when the number of coefficients $\{\beta_k | k = 1, 2, ..., m\}$ being fit, $m$, is equal to the number of observations, $N$. When polynomial models interpolate based on the observations, the over-fitting is always appearing: polynomial models are fit to the data as a result of the $\beta_i$, which minimize the error, being very large in magnitude. For alleviating this adverse factor, a penalty term is added to the error function in order to prevent a large magnitude of $\beta_i$:

$$\hat{E}(\beta) = \frac{1}{2} \sum_{k=1}^{N} (\hat{y}(x_k) - y(x_k))^2 + \frac{\lambda}{2} \| \beta \|^2.$$  \hspace{1cm} (3.5)

This technique is universal in the surrogate modelling literature, however, it should be noted that although this technique can avoid the adverse case of over-fitting, a new problem is introduced namely how to find a suitable value for new parameter $\lambda$. Furthermore, in fact through adding the penalty element for error function the model only can be interpolated for a particular number of observations.

### 3.2.2 Basis Functions

In order to generate a surrogate model that can be interpolated the use of additional basis function centred on one of the sample points is essential. Therefore the prediction of the surrogate model may be written as follow:
\[ \hat{y}(x) = \sum_{k=1}^{m} \beta_k f_k(x) + \sum_{j=1}^{n} b_j \phi(x + x_j). \] (3.6)

There are several options for \( \phi \):

\[ \phi(r) = \exp(-\beta r^2) \] (Gaussian) (3.7)
\[ \phi(r) = r \] (Linear) (3.8)
\[ \phi(r) = r^3 \] (Cubic) (3.9)
\[ \phi(r) = \sqrt{r^2 + \gamma^2} \] (Multiquadric) (3.10)
\[ \phi(r) = r^2 \ln r \] (Thin plate spline) (3.11)
\[ \phi(r) = \exp(-\sum_{i=1}^{d} \theta_i |r|^p) \] (Kriging) (3.12)

where \( r = \|x - x_i\| \), \( \gamma > 0 \) in the multiquadric case, and \( \theta_i > 0 \), \( p \in (0,2] \) in the kriging case. The functions (3.7 to 3.12) belong to a class of functions known as radial basis functions, which will be briefly discussed in next section.

### 3.3 Kriging-Assisted Single-Objective Optimization

#### 3.3.1 Background

In the 1960s, the French mathematician Georges Matheron, on the Master’s thesis of D. G. Krige, developed a statistical method named ‘kriging’ to interpolate the value of a random field for making predictions of unobserved location based only on limited observations. In general, kriging is combined with two different approaches: the geostatistical one and the maximum likelihood estimate (MLE) [34], both of them are based on random functions [13]. In this research, the MLE is used as it is more suitable for the design and analysis of computer experiments (DACE) [35]. The following section will present the main DACE equations, necessary for the optimizers which follow in later chapters.
3.4 Design and Analysis of Computer Experiments

3.4.1 Maximum Likelihood Estimate Approach

In this approach, firstly based on the initial samples it is possible to “control” some key parameters of the correlation function, and use the revised parameters to construct the kriging model.

The kriging as a standard regression model is defined as follows:

$$\hat{y}(x) = \sum_{k=1}^{m} \beta_k f_k(x_i) + \varepsilon(x_i) \quad (3.13)$$

The sum $\sum_{k=1}^{m} \beta_k f_k(x_i)$ including the linear combination of the value of initial sampled points $x_i$ is viewed as a global approximation to the true function. The coefficients $\beta_k$ are regression parameters, and $\varepsilon$ is an additive Gaussian noise, representing our uncertainty.

By considering two design vectors $x_i, x_j$, quite close to each other in design variable space, it can be expected that the corresponding objective function value for both of them is similar as well. The kriging model is modelled statistically by assuming that the errors $\varepsilon(x_i)$ and $\varepsilon(x_j)$ are correlated as follow:

$$R(\varepsilon(x_i)\varepsilon(x_j)) = \prod_{k=1}^{n} e^{-\theta_k|x_i-x_j|^p_k} \quad (3.14)$$

where

1) $|x_i - x_j|_k$ is the distance between $x_i$ and $x_j$ on $k$-direction

2) $\theta_k$ and $p_k$ determine how fast the correlation between design vectors drops away and the smoothness of the function in $k^{th}$ coordinate direction, respectively. There are several optional models of $R$, but normally the Gaussian model ($p_k = 2$) is selected because of its good performance in many circumstances.
The kriging model may be simplified further by replacing the polynomial in Equation (3.13) by a constant term.

\[ \hat{y}(x) = \mu + \varepsilon(x). \]  

(3.15)

In doing so, kriging model then becomes a probabilistic model which has \(2d+2\) parameters: \(\mu, \sigma^2, \theta_1, \ldots, \theta_d, p_1, \ldots, p_d\). These parameters are chosen to maximize the likelihood of the observations, which for this model is given by

\[ \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}|R|^{1/2}} \exp\left[\frac{-(y-1\mu)^T R^{-1} (y-1\mu)}{2\sigma^2}\right] \]  

(3.16)

where \(R\) is the correlation matrix, which represents the correlation between each pair of evaluated design vectors, defined as the \(n \times n\) matrix whose \(i - j\)th entry is \(R(x^i, x^j)\), \(y\) is the \(n \times 1\) vector of observed objective function values,

\[ y = [y(x^{(1)}), y(x^{(2)}), \ldots, y(x^{(n)})]^T \]  

(3.17)

where \(1\) means \(n \times 1\) vector filled with ones. In order to obtain the optimal value of parameter \(\theta\), it is more common to maximize the log of Equation 3.16, which is (ignoring constant terms)

\[ -\frac{n}{2} \log(\sigma^2) - \frac{1}{2} \log(|R|) - \frac{(y-1\bar{\mu})^T R^{-1} (y-1\bar{\mu})}{2\sigma^2} \]  

(3.18)

Based on the information provided by the sampled values, \(\theta\) are fixed meanwhile \(\mu\) and \(\sigma^2\) have an explicit expression. The estimated global model \(\hat{\mu}\) is

\[ \hat{\mu} = \frac{1^T R^{-1} y}{1^T R^{-1} 1} \]  

(3.19)

The variance \(\sigma^2\) is

\[ \hat{\sigma}^2(\theta) = \frac{(y-1\bar{\mu})^T R^{-1} (y-1\bar{\mu})}{n} \]  

(3.20)

where the hats on these two terms signify that these are the optimal values of \(\sigma\) and \(\mu\).
By substituting Equation 3.19 and Equation 3.20 into Equation 3.18, the following log-likelihood function is obtained.

\[-\frac{n}{2} \ln(\hat{\sigma}^2(\theta)) - \frac{1}{2} \ln(|R|)\]  

(3.21)

By maximizing this log-likelihood function, the optimal value of \( \theta \) can be found.

### 3.4.2 The Kriging Prediction Formula

After achieving the updated value of \( \theta \) and \( p \), an unevaluated vector \( x^* \) needs to be considered, assuming that \( y^* \) is an estimate of the function value for \( x^* \). The optimal value of \( \theta \) is obtained by maximizing the likelihood (for the observations) as explained in the former section, the ‘augmented’ log-likelihood function can be obtained by adding the predicted values \( (x^*, y^*) \) to the observations. In fact with the tuneable model parameter \( \theta \) fixed, the augmented log-likelihood function is a function of \( y^* \). As \( y^* \) varies, its value is a measure of how consistent the estimate \( (x^*, y^*) \) is with the description of variation, as determined by the observed samples. In order to achieve a good prediction of the objective function, maximizing the augmented log-likelihood is essential.

To derive the expression for \( \hat{y} \) (the optimal predicted value), it is essential to set the derivative of the augmented log-likelihood equal to zero with respect to \( y^* \). Let \( \hat{y} = (y^T, y^*)^T \) be the vector of function values composed of the observed values, \( \hat{y} \) the estimated value, and \( r \) the correlation vector, which expresses the correlation between an unevaluated design vector \( x \) and the \( n \) evaluated design vectors, defined as [10]

\[r(x) = [R(x, x^{(1)}), R(x, x^{(2)}), ..., R(x, x^{(n)})]^T.\]  

(3.22)

The correlation matrix for the augmented set as \( \hat{R} \)

\[\hat{R} = \begin{pmatrix} R & r \\ r^T & 1 \end{pmatrix}\]  

(3.23)
Chapter 3 Surrogate Modelling

The part of the augmented likelihood based on $y^*$ is

$$\frac{1}{2\hat{\sigma}^2} \left( \hat{\gamma} - 1\hat{\mu} \right) \hat{R}^{-1} \left( \hat{\gamma} - 1\hat{\mu} \right).$$

(3.24)

Substituting in for $\hat{\gamma}$ and $\hat{R}$, the above formula becomes

$$\frac{1}{2\hat{\sigma}^2} \left( \begin{array}{c} y - 1\hat{\mu} \\ y^* - \hat{\mu} \end{array} \right)^T \left( \begin{array}{cc} R & r^T \\ r & 1 \end{array} \right)^{-1} \left( \begin{array}{c} y - 1\hat{\mu} \\ y^* - \hat{\mu} \end{array} \right).$$

(3.25)

The following identity for the inverse of a partitioned matrix has to be mentioned for deriving the augmented log-likelihood formula

$$\left( \begin{array}{cc} R & r^T \\ r & 1 \end{array} \right)^{-1} = \left( \begin{array}{cc} M & -MBD^{-1} \\ -D^{-1}CM \end{array} \begin{array}{c} D^{-1} \end{array} \begin{array}{c} D^{-1}CMBD^{-1} \end{array} \right)$$

(3.26)

where

$$M = (A - BD^{-1}C^{-1})^{-1},$$

(3.27)

and the matrix may be written as

$$\left( \begin{array}{cc} R^{-1} + R^{-1}r(1 - r^T R^{-1}r)^{-1}r^T R^{-1} & -R^{-1}r(1 - r^T R^{-1}r)^{-1} \\ -(1 - r^T R^{-1}r)r^T R^{-1} & (1 - r^T R^{-1}r)^{-1} \end{array} \right).$$

(3.28)

Substituting into Equation 3.24, the partial terms of the augmented log-likelihood based on $y^*$ are

$$\left[ \frac{-1}{2\hat{\sigma}^2(1 - r^T R^{-1}r)} \right] (y^* - \hat{\mu})^2 + \left[ \frac{r^T R^{-1}(y - 1\hat{\mu})}{\hat{\sigma}^2(1 - r^T R^{-1}r)} \right] (y^* - \hat{\mu}).$$

(3.29)

As mentioned at the outset, the expression for $y^*$ is found by setting the derivative of the augmented likelihood equal to zero with respect to $y^*$ as follow:

$$\left[ \frac{-1}{\hat{\sigma}^2(1 - r^T R^{-1}r)} \right] (y^* - \hat{\mu}) + \left[ \frac{r^T R^{-1}(y - 1\hat{\mu})}{\hat{\sigma}^2(1 - r^T R^{-1}r)} \right] = 0$$

(3.30)
This derives

\[ \hat{y}(x^*) = \hat{\mu} + r^T R^{-1}(y - 1\hat{\mu}). \]  

This formula is the general expression for the kriging predictor.

As a final note, it should be explained that Equation 3.31 is combined with the general surrogate model equation, Equation 3.6. The polynomial terms in Equation 3.6 is substituted by \( \mu \), and \( b_l \) is replaced by the \( i^{th} \) element of \( R^{-1}(y - 1\hat{\mu}) \). Then the kriging predictor becomes a linear combination of polynomial terms and a set of basis functions.

### 3.4.3 The Standard Error Formula

As shall be emphasised in the following chapters, a potential error in the kriging prediction is extremely helpful for deciding where to evaluate next in the design variable space. Although this potential error is not the real error between the prediction produced by kriging model and the real objective function value, it is useful in quantifying uncertainty in kriging predictions.

Mathematically, the curvature of the augmented log-likelihood function is inversely related to our estimate of the potential error in the prediction. For example if the curvature of the augmented log-likelihood function is small, it corresponds to a low confidence in the prediction. Conversely, if the augmented log-likelihood function varies rapidly, it indicates the prediction with a high confidence.

The full derivation of the Mean Squared Error (MSE) is given in [34] and may be summarised by the following equation

\[ s^2(x) = \hat{\sigma}^2 \left[ 1 - r^T R^{-1}r + \frac{(1 - r^T R^{-1}r)}{1 - r^T R^{-1}1} \right], \quad (s(x) \text{ is the standard error}) \]  

The notion of the mean squared error and the standard error \( s(x) \) play a critical role in quantifying uncertainty in kriging predictions, as shall be illustrated further in the following chapters.
Chapter 4 Kriging – Assisted Single-Objective Optimization

4.1 Introduction

As one kind of the regression models, the kriging surrogate model can produce response surface for predicting objective function only based on extremely limited information. The accuracy of the kriging prediction is also available therefore the quality of the surrogate model can be quantified. Normally the values of the objective function are obtained via time-consuming Finite Element (FEM) model computation, therefore a budget of function evaluation should be enforced to minimise the computational time necessary for the optimization process.

How to utilize the ‘Mean Square Error (MSE)’ in selecting design vectors successively for further prediction throughout the iterative process is discussed in the chapter. Several kinds of ‘utility functions’ including mean square error have been created to assist the model in selecting the new vectors that will be used in updating the prediction. With the guidance provided by the utility function, the kriging model can improve the approximation via consistently updating the sampling points with the optimization aim of finding the global optimum. In addition to ensuring that the global optimum of the objective function can be found successfully the efficiency of the optimizer is another critical factor which has to be considered. As mentioned earlier in practical electromagnetic design each sampling point’s value is calculated by calling time-consuming FEM models, therefore a minimum number of FEM calls are also pursued.

4.2 Schwefel test function

In accordance with the fundamental demands of choosing analytic functions to numerically verify the feasibility of surrogate models described in Chapter 2, the Schwefel function is an appropriate example meeting all requirements and providing a stern test for our algorithms. The Schwefel function is deceptive in that the global minimum is geometrically distant, over the parameter space, from the next best local minima. Therefore, the search algorithms are potentially prone to convergence in the
wrong direction. Because of these properties the Schwefel function has been a popular choice in testing the robustness of optimisation algorithms. While testing using a single function may not be conclusive, the Schwefel function has in the past been found helpful when validating similar algorithms [36]-[38]; testing under practical conditions will obviously continue after the algorithm has been fully integrated into an electromagnetic design system.

The multimodal Schwefel test function is as follows [39]:

\[
f(x) = \sum_{i=1}^{d} -x_{i} \sin(\sqrt{|x_{i}|})
\]  

(4.1)

Firstly the single-variable Schwefel function \((d=1)\) in the interval \(x \in [-500, 500]\) is selected as the objective function to verify the feasibility of the kriging surrogate model.

### 4.3 Expected Improvement

The prediction made by the kriging surrogate model can be viewed as a Gaussian process \(\gamma\), while a number of updating schemes can be adopted in this process. This allows for the concept of improvement to be defined: for a single objective to be minimized, the improvement may be measured by comparing the value realized by the objective function with the current minimum of prediction. This is written as [40]:

\[
I(x) = \max(f_{\text{min}} - \gamma(x), 0).
\]  

(4.2)

The expectation of improvement normally called the expected improvement (EI) may be found by integrating it over the likelihood of achieving it given by the normal density function:

\[
E[I(x)] = \int_{0}^{\infty} \left\{ \frac{1}{\sqrt{2\pi}s(x)} \exp\left[ - \frac{(f_{\text{min}} - \hat{y}(x))^{2}}{2s^{2}(x)} \right] \right\} dl
\]

(4.3)

which is written as [2]:

\[
E[I(x)] = \begin{cases} 
(f_{\text{min}} - \hat{y}(x)) \Phi \left( \frac{f_{\text{min}} - \hat{y}(x)}{s(x)} \right) + s(x) \phi \left( \frac{f_{\text{min}} - \hat{y}(x)}{s(x)} \right) & \text{if } s(x) > 0 \\
0 & \text{if } s(x) = 0
\end{cases}
\]

(4.4)
where \( \hat{y}(x) \) is the predicted value of objective function by kriging model, given by Equation 3.31, \( f_{\text{min}} \) is the minimum value of \( y \) at the observed sample points. Besides, \( s(x) \) is the root mean squared error (standard error) in this prediction, given by Equation 3.32, \( \phi \) and \( \Phi \) are the normal density and normal distribution functions respectively.

If \( s(x) > 0 \), the EI utility function contains two independent components: the first component in Equation 4.4 is large when the kriging prediction is lower than the current minimum, and so favours searching promising area with high confidence of design variable space, whilst the second component in Equation 4.4 is large when the potential error produced by kriging model is large, thus favours searching regions with high uncertainty. The EI function can be viewed as the fixed compromise between exploration and exploitation.

In Figure 4.1, a set of continuous snapshots throughout the whole iterative process of the kriging model assisted by EI in terms of the iteration number are presented, with the six initial sample points (\( x = -450, y = 315.8177; x = -230, y = 118.6874; x = -150, y = -47.0315; x = 40, y = -1.6543; x = 160, y = -13.2233; x = 500, y = 180.5892 \)) plotted as the black points in the following graphs.

(a) Iteration 1
Chapter 4 Kriging – Assisted Single-Objective Optimization

(b) Iteration 2

(c) Iteration 3
Chapter 4 Kriging – Assisted Single-Objective Optimization

(d) Iteration 4

(e) Iteration 5
(f) Iteration 6

(g) Iteration 7
(h) Iteration 8

(i) Iteration 9
Figure 4.1: Iterations of the expected improvement approach on the Schwefel test function.
The maximum of the value of EI (the green curve lays on the bottom of coordinates in the above graphs) indicates the place where to select the updated sampling points. The history of selecting the updated sampling points each iteration is highlighted. It can be seen that the kriging model with EI requires 11 iterations to find out the global minimum of the objective function. Among these 11 iterations, at the 1\textsuperscript{st}, 3\textsuperscript{rd}, 4\textsuperscript{th}, 8\textsuperscript{th}, 9\textsuperscript{th}, 10\textsuperscript{th} and 11\textsuperscript{th} the kriging model tends to searching the region around the minimum of existing sampling points, whereas at the 2\textsuperscript{nd}, 5\textsuperscript{th}, 6\textsuperscript{th} and 7\textsuperscript{th} iteration the place for selecting updated sampling points tends to be near the unsampled region. Although the global minimum ($x = 421, y = -418.9828$) is found ultimately, the approximation of the shape of the objective function is not quite ideal in certain region. In order to verify the reliability of the kriging model with EI and the feature of the two components in EI, some extreme cases need to be involved. Here the termination criterion of the kriging model is formulated as: if the repeated sampling point is chosen as the updated sampling point for further evaluation (within certain tolerance), the kriging model will be stopped.

### 4.3.1 Latin Hypercube

The principle of kriging is based on the correlation between the known sampling points and all other design vectors in the design space, thus a certain number of sampling points are indispensable before a kriging model can even be constructed. In this section, one ubiquitously used methodology of sampling, namely the Latin Hypercube [40], [41], is discussed and the concept is presented as below.

The Latin Hypercube method is able to maintain the selected sample points to be distributed in a relatively average fashion in the searching space. As shown in Figure 4.2, a Latin square of size $n$ may be constructed by partitioning each side into $n$ intervals to create and $n \times n$ grid, and then placing a point in centre of $n$ squares in the grid, while each row and each column only contains one point [40].
4.3.2 Extreme Distribution of Initial Sample Points

It should be noted that the increasing cost of evaluating the new sampling points via FEM model as updates are added to the model might be mitigated through the proper setting of initial sampling points. This section mainly aims to observe the impact of the optimization efficiency by using different distribution of initial sampling points. As an extreme case, Figure 4.3 helps to demonstrate the reliable performance of the kriging model under the extreme distribution of the initial sampling points. To verify the reliability of the kriging model, only one sampling point \((x=0, y=0)\) is chosen as the initial setting, which is extremely limited for the prediction process. The difficulty in the setting of initial sampling points is created deliberately to examine whether the global minimum could be located precisely by the kriging model and the convergence efficiency would be impacted due to this extreme case. Fortunately excluding the unique sampling point there are 16 additional sampling points necessary to find the global minimum successfully (the number of sampling points is 17). To compare with the previous test requiring 6 initial sampling points mentioned above, the total number of sampling points used by the kriging model is the same at 17, which indicates that the distribution and the number of initial sampling points might not affect the outcome of the approximation.
In order to understand further the principle of initializing sampling points, a new analytic function is used to observe the outcome of approximation under different distribution of initial sampling points. This analytic function is defined as follow:

\[ f(x) = 10 - \sum_{i=1}^{n} \left[ \frac{3.5}{1+(x_i-5)^2} + \frac{2.2}{1+(x_i-15)^2/10} + \frac{1.2}{1+(x_i-25)^2/30} \right], \quad 0 \leq x_i \leq 27 \quad (4.5) \]

The profile of the two-variable \((n = 2)\) version Schwefel function is plotted in Figure 4.4 (a) and will be taken as the objective function to compare the approximation outcome with different distribution of sampling points.

The first test is initialized as the seven sampling points which are chosen arbitrarily; the second one only selects two sampling points from the first setup; two special points are pre-set as the initial sampling points of the third test (Table 4.1(a)-(c)). Except the different setting of initial sampling points, other elements are all set to be the same in these three tests. The tuning range of hyper-parameter \(\theta\) in the kriging model is set as \((0.1 < \theta < 300)\); the step size of these following tests is set as 0.1, while the data size of the task becomes \(271 \times 271\); the scheme of updating the sampling points is chosen to be EI.
The setting of seven initial sampling points

<table>
<thead>
<tr>
<th>$x_1$</th>
<th>10</th>
<th>10</th>
<th>5.7</th>
<th>13</th>
<th>20</th>
<th>7</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_2$</td>
<td>7</td>
<td>10</td>
<td>10</td>
<td>7</td>
<td>7</td>
<td>13</td>
<td>20</td>
</tr>
<tr>
<td>$y$</td>
<td>7.9966</td>
<td>8.1913</td>
<td>6.4292</td>
<td>7.0688</td>
<td>7.6024</td>
<td>7.0688</td>
<td>7.2927</td>
</tr>
</tbody>
</table>

(a)

The setting of two initial sampling points selected among the above initial sampling points

<table>
<thead>
<tr>
<th>$x_1$</th>
<th>20</th>
<th>27</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_2$</td>
<td>7</td>
<td>27</td>
</tr>
<tr>
<td>$y$</td>
<td>9.4336</td>
<td>7.5822</td>
</tr>
</tbody>
</table>

(b)

The special setting of two special initial sampling points

<table>
<thead>
<tr>
<th>$x_1$</th>
<th>0</th>
<th>27</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_2$</td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>$y$</td>
<td>9.4336</td>
<td>7.5822</td>
</tr>
</tbody>
</table>

(c)

Table 4.1: (a) The setting of seven initial sampling points (b) The setting of two initial sampling points selected from the first initial setup (c) The setting of two special initial sampling points.

The corresponding response surfaces produced by the kriging surrogate model with different setting of initial sampling points at the final iteration of finding out the global minimum are shown in Figure 4.4 (b)-(d). The case with seven initial sampling points listed in Table 4.1 (a) and based on this initial sampling setup the kriging model takes 324 iterations (331 sampling points) to converge to the global minimum of the objective function; while the second case (Table 4.1(b)) spends 309 iterations (311 sampling points) finding out the global minimum which shown in Figure 4.4 (c); the third test with two special sampling points (Table 4.1 (c)) requires 322 iterations (324 sampling points) to find out the global minimum shown in Figure 4.4 (d). In addition to the efficiency of optimization, the quality of response surface produced by the kriging surrogate model when finding the global optimum is also one of elements needs to be researched, especially when considering the robustness of practical designs. The total
absolute error between the approximation and the real objective function value that indicates the accuracy of approximating the shape of the objective function is measured: over the full design space the total error of these three cases is 6514.3, 7064.8 and 6724.2 respectively. From the above results, it can be found that although the first case owns more sampling points than other two cases, the efficiency of convergence is not the highest; but the accuracy of the first case is higher than the other two. The number of initial sampling points does not affect the efficiency of converging to the global optimum directly, whereas the number and the position of initial sampling points may affect the accuracy of predicting the shape of objective function.

Figure 4.4: (a) The objective function \((0 < x_1 < 27, 0 < x_2 < 27)\); (b) The prediction performance of kriging model with seven initial sampling points; (c) The prediction performance of kriging model with only the two initial sampling points selected from the above seven sampling points; (d) The prediction performance of kriging model with two special initial sampling points.
4.4 Exploration and Exploitation

As mentioned in previous section, the expected improvement utility function contains the two components favouring exploration and exploitation respectively. In order to observe the difference between these two independent components’ performance, the same objective function and initial sampling points \((x = -450, y = 315.8177; x = -230, y = 118.6874; x = -150, y = -47.0315; x = -40, y = 1.6543; x = 160, y = -13.2233; x = 500, y = 180.5892)\) are used to test them separately.

<table>
<thead>
<tr>
<th>No of iterations</th>
<th>The updated sampling points</th>
<th>No of iterations</th>
<th>The updated sampling points</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>((x=278, y=228.5880))</td>
<td>10</td>
<td>((x=-361, y=54.1057))</td>
</tr>
<tr>
<td>2</td>
<td>((x=77, y=-46.5879))</td>
<td>11</td>
<td>((x=-294, y=-291.4295))</td>
</tr>
<tr>
<td>3</td>
<td>((x=-93, y=-20.1930))</td>
<td>12</td>
<td>((x=394, y=-331.5134))</td>
</tr>
<tr>
<td>4</td>
<td>((x=119, y=118.5514))</td>
<td>13</td>
<td>((x=423, y=-418.4620))</td>
</tr>
<tr>
<td>5</td>
<td>((x=-135, y=-109.6082))</td>
<td>14</td>
<td>((x=412, y=-408.9091))</td>
</tr>
<tr>
<td>6</td>
<td>((x=43, y=-11.6456))</td>
<td>15</td>
<td>((x=500, y=-180.5892))</td>
</tr>
<tr>
<td>7</td>
<td>((x=197, y=-195.9861))</td>
<td>16</td>
<td>((x=418, y=-417.8725))</td>
</tr>
<tr>
<td>8</td>
<td>((x=217, y=-179.8598))</td>
<td>17</td>
<td>((x=423, y=-418.4620))</td>
</tr>
<tr>
<td>9</td>
<td>((x=-330, y=-208.4379))</td>
<td></td>
<td>((x=-330, y=-208.4379))</td>
</tr>
</tbody>
</table>

Table 4.2: The performance of the component favouring exploration

<table>
<thead>
<tr>
<th>No of iterations</th>
<th>The updated sampling points</th>
<th>No of iterations</th>
<th>The updated sampling points</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>((x=-120, y=-119.8985))</td>
<td>7</td>
<td>((x=-489, y=-59.6052))</td>
</tr>
<tr>
<td>2</td>
<td>((x=-105, y=-76.9258))</td>
<td>8</td>
<td>((x=18, y=16.0503))</td>
</tr>
<tr>
<td>3</td>
<td>((x=-124, y=-122.7876))</td>
<td>9</td>
<td>((x=-500, y=-180.5892))</td>
</tr>
<tr>
<td>4</td>
<td>((x=-125, y=-122.8724))</td>
<td>10</td>
<td>((x=-499, y=-169.7738))</td>
</tr>
<tr>
<td>5</td>
<td>((x=-465, y=192.6987))</td>
<td>11</td>
<td>((x=-499, y=-169.7738))</td>
</tr>
<tr>
<td>6</td>
<td>((x=-478, y=60.9874))</td>
<td></td>
<td>((x=-478, y=60.9874))</td>
</tr>
</tbody>
</table>

Table 4.3: The performance of the component favouring exploitation
The global minimum of the Schwefel test function is:

\[ \text{Min}(f(x)) = -418.9828, \text{ when } x=421. \]

The history of selecting the updated sampling points at each iteration is presented in Table 4.2 and 4.3. From the results shown in Table 4.2 and 4.4, neither the exploration or exploitation component can guide the kriging model converge to the global optimum precisely. But the best-performed solution in the test of the exploration component is much better than the one obtained from testing exploitation. The exploration component is more likely to misguide the model to converge at one of the local optima. Although any of these two independent components can hardly guide the model locate the global optimum precisely, the combination of them, the expected improvement utility function, can complete the task of finding the global optimum. Therefore the trade-off between exploration and exploitation seems critical to find the global optimum successfully and efficiently.

4.5 Conclusion

In this section one popular utility function, the Expected Improvement (EI), that is able to guide the kriging surrogate model to select updated sampling points for converging to the global optimum and evolving approximation, is introduced and verified by a series of numerical tests under different initial setting. The EI utility function contains two components favouring exploration and exploitation respectively. Through exploring the feature of these two independent components, it can be summarised as: the component favouring exploitation is helpful to guide the optimiser converge to the minimum among the existing sampling points, thus it might cause the optimiser to be trapped by a very deceptively positioned local minimum; the component favouring exploration tend to guide the model to search the unsampled region, but if emphasizing it excessively undoubtedly the efficiency of searching global optimum would be affected. Thus the importance of the trade-off between exploration and exploitation is confirmed in order to guarantee the model finding the global optimum successfully and efficiently.
Chapter 5 Balance between Exploration and Exploitation

5.1 Introduction

The previous chapter of this thesis introduces one typical utility function named Expected Improvement (EI), which is able to guide the kriging surrogate model to infill the next updated point for searching the global optimum of objective function. The best currently existing solution, the mean square error, the Gaussian density function and Gaussian distribution function are involved to construct the EI function, which consists of two components favouring searching the regions with high confidence called exploitation and favouring searching the regions with high uncertainty called exploration respectively. Through a set of practical kriging-assisted single-objective tests developed specially to assess the performance of these two terms, it has been shown that the second term representing exploration performs dramatically better in terms of finding the global optimum of the objective function, whereas the exploitation often can only find the local minimum. Since EI applies equal weights to the two terms, it may be seen as a fixed compromise between exploration and exploitation. The trade-off between exploration and exploitation is a critical issue when attempting to find the global optimum of an objective function effectively and efficiently.

The purpose of this chapter is to introduce two methodologies called Generalized Expected Improvement and Weighted Expected Improvement, which can both adjust the balance between exploration and exploitation via tuning certain relevant parameters. The optimal choice of these tuneable parameters, both in terms of the ability of the algorithm to achieve the correct answer (global minimum) and doing it efficiently (fewer iterations required), is especially important. A series of numerical tests have been done for exploring the principle of how these tuneable parameters affects the efficiency of converging to the global optimum.
5.2 Non-Target Based Tuneable Utility Function

Non-target based tuneable utility functions are characterized by allowing the balance between exploration and exploitation to be tuned through a revising parameter [10]. Two existing utility functions both based on the Expected Improvement (EI) are introduced: the Generalized Expected Improvement (GEI), and the Weighted Expected Improvement (WEI).

5.2.1 Generalized Expected Improvement

The Generalized Expected Improvement (GEI) utilizes one unique tuneable parameter $g$ to adjust the emphasis on exploration or exploitation. By defining the generalized improvement as

$$ I^g(x) = \max\{(f_{\min} - \gamma(x))^g, 0\}, \quad (5.1) $$

where $g$ is an integer, the expectation value of this, known as Generalized Expected Improvement (GEI), can be calculated as follow [42]:

$$ GE\left(I^g(x)\right) = s^g \sum_{k=0}^{g} (-1)^k \frac{g!}{k!(g-k)!} \left(\frac{f_{\min} - \gamma(x)}{s(x)}\right)^{g-k} \bar{T}_k \quad (5.2) $$

where

$$ \bar{T}_k = -\Phi\left(\frac{f_{\min} - \gamma(x)}{s(x)}\right) \left(\frac{f_{\min} - \gamma(x)}{s(x)}\right)^{k-1} + (k-1) \bar{T}_{k-2} \quad (5.3) $$

With

$$ T_0 = \Phi\left(\frac{f_{\min} - \gamma(x)}{s(x)}\right) \text{ and } T_1 = -\Phi\left(\frac{f_{\min} - \gamma(x)}{s(x)}\right) \quad (5.4) $$

If applying the higher value of $g$, the greater level of improvement is being sought. The more emphasis is placed on exploring the regions close to the minimum of existing sampling points rather than searching locally [42]. When $g = 1$ the GEI given by equation 5.2 is equivalent to EI. For comparing with the performance of the normal EI in previous chapter, the Schwefel test function and same initial setting is applied as before. The GEI method for several values of $g$ is shown for the first iteration of the
Schwefel function in Figure 5.1 to 5.3. The olive curve locating at the bottom of the graphs represents the value of GEI, furthermore the maximum of GEI that indicates the location of the updated sampling points for further evaluation.

Figure 5.1: The performance of kriging model with GEI for \( g = 1 \) at the first iteration

Figure 5.2: The performance of kriging model with GEI for \( g = 2 \) at the first iteration
When $g = 1$ in Figure 5.1, the location chosen as the updated sampling point is relatively close to the minimum of existing sampling points. With the value of parameter $g$ increasing, the location selected for the updated sampling point tends to move from the local region including the minimum of existing sampling points to the unsampled region, as the performance shown in Figure 5.2 and 5.3 (when $g = 2$ and $g = 5$ respectively). This set of performances obtained by GEI can be viewed as that the emphasis of searching new sampling points is shifted from exploitation to exploration with the tuneable parameter $g$ increasing.

In addition to exploring the principle of the tuneable parameter $g$, more tests have been done for analysing the impact from the parameter $g$ on the kriging model’s efficiency of converging to the global optimum, and the corresponding results are listed in Table 5.1. Through observing the results achieved by the five different values of $g$, the best performance appears when $g$ is set as 1; obviously the increasing value of $g$ causes the optimization efficiency to decline gradually. Due to the increasing value of $g$ the emphasis tends to be placed on exploration rather than exploitation, thus a scheme that emphasizes exploration at the beginning of test to avoid the kriging model being
‘tricked’ by a particularly poor or unlucky initial sampling points or a very deceptive local optimum; after certain iterations the emphasis can be shifted from exploration to exploitation to accelerate the process of converging to the global optimum.

<table>
<thead>
<tr>
<th>$g$</th>
<th>Number of Iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (GEI is equal to EI)</td>
<td>11 (found the global minimum)</td>
</tr>
<tr>
<td>2</td>
<td>12 (found the global minimum)</td>
</tr>
<tr>
<td>3</td>
<td>14 (found the global minimum)</td>
</tr>
<tr>
<td>4</td>
<td>14 (found the global minimum)</td>
</tr>
<tr>
<td>5</td>
<td>16 (found the global minimum)</td>
</tr>
</tbody>
</table>

Table 5.1: Value of $g$, the number of iterations of GEI on the Schwefel test function

For example in [43], there is a built-up scheme presented in Table 5.2 that attempts to construct an optimal collocation of the different value of $g$ with the aim of finding the global minimum efficiently. But unfortunately the specific setting of the combination of different values of $g$ has to depend on the experience obtained by numerous tests. This drawback is one of the reasons prompting us to create novel algorithms to control the balance between exploration and exploitation automatically.

<table>
<thead>
<tr>
<th>Iteration</th>
<th>$g$</th>
<th>Iteration</th>
<th>$g$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-4</td>
<td>20</td>
<td>20-24</td>
<td>2</td>
</tr>
<tr>
<td>5-9</td>
<td>10</td>
<td>25-34</td>
<td>1</td>
</tr>
<tr>
<td>10-19</td>
<td>5</td>
<td>$\geq 35$</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 5.2: Value of $g$, as used in built-up scheme in [43].

### 5.3 Weighted Expected Improvement

The expected improvement utility function can be seen as the compromise between exploration and exploitation, as the equal weights are distributed on the two components favouring the exploration and the exploitation respectively. The Weighted Expected Improvement (WEI) [44] is derived from EI by adding a tuneable parameter which can adjust the weights on the components favouring exploration and exploitation.
respectively, whilst the quality of the approximation of the objective function can be improved by incorporating the newly evaluated design vector at each iteration.

The WEI utility function has been proposed in [44]:

$$F[I(x)] = \begin{cases} 
  w(f_{min} - \hat{y}(x)) \Phi \left( \frac{f_{min} - \hat{y}(x)}{s(x)} \right) + (1 - w)s(x)\phi \left( \frac{f_{min} - \hat{y}(x)}{s(x)} \right) & \text{if } s(x) > 0 \\
  0 & \text{if } s(x) = 0 
\end{cases}$$

(5.5)

The WEI via the weighting parameter \( w \) is capable to adjust the balance between exploration and exploitation mechanically for searching locally or globally [45].

There are three representative values of the weighting parameter \( w \) need to be explained specifically: when \( w=1 \), that all weight is placed on the first component means heavily emphasizing exploitation; contrarily when \( w=0 \), it means overwhelmingly emphasizing exploration; when \( w=0.5 \), the WEI is equivalent to the EI.

To find out the optimal value of the weighting parameter \( w \), a set of values of weights have been tested as shown in Table 5.3. As before, the Schwefel Test Function and the same initial circumstance are pre-set here. When \( w=0.1, 0.2, 0.3, 0.4, 0.5 \) the kriging model assisted by the WEI can find the global minimum successfully (when \( w=0.5 \), the WEI is same as EI). This scope of value of the weighting parameter (0<\( w <0.5 \)) indicates that the exploration is emphasised more than exploitation; the WEI tends to search the area with high uncertainty rather than concentrating at the area near the best point of existing sample points with higher confidence. Conversely while the scope of weighting parameter is 0.5<\( w <1 \), the process of choosing updated sampling points could be seen that more emphasis is placed on exploitation than exploration. The four typical values of weighting parameter have been applied as \( w=0.6, 0.7, 0.8, 0.9 \), unfortunately the WEI fails to locate the global minimum and even lead the surrogate model trapped in unacceptable plateaus due to the termination criteria (end the model running while the repeated sampling points appear). Because of overrated emphasis placed on exploitation it causes the kriging model with WEI tends to concentrate on searching the local region around the minimum of existing sampling points. As a case of placing more emphasis on exploration than exploitation, it should be noted that when \( w=0.4 \) the WEI outperforms all of other chosen values of weighting parameter \( w \), which only spends 7
objective function evaluations to find out the global minimum successfully since the normal EI requires 11 function evaluations.

In the last section, the snapshots of the numerical tests using GEI at the first iteration with several typical values of tuneable parameter are presented for comparison of different location of selecting the updated sampling points. Likewise, three typical examples applying the WEI with different values of weighting parameter $w$ which are $w=0.1$, $w=0.5$, $w=0.9$ respectively are presented in Figure 5.4-5.6. The same test function and initial setting were applied as before. The green curve in these three graphs represents the value of WEI utility function of which the maximum indicates the place to select the updated sampling points for producing response surface next iteration. With the variation of the weight $w$, the value of the WEI varies dramatically as well, hence the location of the maximum of the WEI shifts from the region nearby the existing sampling points to the region containing rare sampling points (Figures 5.4 - 5.6).

Figure 5.4: The performance of kriging model with WEI for $w=0.1$ at the first iteration
Chapter 5 Balance between Exploration and Exploitation

Figure 5.5: The performance of kriging model with WEI for $w=0.5$ at the first iteration

Figure 5.6: The performance of kriging model with WEI for $w=0.9$ at the first iteration
Table 5.3 presents the specific number of iterations required for finding the global minimum with the weighting parameter $w$ in the range $(0 < w < 1)$. When the weighting parameter $w=0.4$, the kriging model assisted by WEI outperforms all of other chosen weights, which only takes 7 iterations to locate the global minimum. It should be noted that when the weighting parameter’s value is from 0.1 to 0.5 (EI), the kriging model is able to guarantee the global minimum can be found successfully. Whereas the kriging model is trapped at the unacceptable local minimum subject to the termination criteria mentioned before, when $w$ is from 0.6 to 0.9. The failure of finding the global optimum proves that too much emphasis on exploitation is a risky strategy. However if excessive emphasis is placed on exploration, the efficiency of optimization would be impacted. When $w=0.1, 0.2, 0.3$ the WEI performs worse than the normal EI. Due to the important turning point existing around $w=0.5$, more detailed values of the weighting parameter are tested in the range $(0.5 < w < 0.6)$. When $0 < w < 0.54$, the global minimum can be found successfully. As shown in the table somewhere between 0.54 and 0.55 there is a changeover between a regime where only a local minimum is found and values of $w$ which allow for the global minimum to be correctly identified. Commencing with $w=0.54$ to $w=0.1$, the global minimum can be found successfully by the kriging model with the assistance of the WEI.
Chapter 5 Balance between Exploration and Exploitation

<table>
<thead>
<tr>
<th>Value of weight</th>
<th>Percentage of weights on the component favouring exploration</th>
<th>Percentage of weights on the component favouring exploitation</th>
<th>The number of iterations required to find the global minimum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>100%</td>
<td>Fails</td>
</tr>
<tr>
<td>0.9</td>
<td>10%</td>
<td>90%</td>
<td>3 (finds LM)</td>
</tr>
<tr>
<td>0.8</td>
<td>20%</td>
<td>80%</td>
<td>3 (finds LM)</td>
</tr>
<tr>
<td>0.7</td>
<td>30%</td>
<td>70%</td>
<td>3 (finds LM)</td>
</tr>
<tr>
<td>0.6</td>
<td>40%</td>
<td>60%</td>
<td>3 (finds LM)</td>
</tr>
<tr>
<td>0.59</td>
<td>41%</td>
<td>59%</td>
<td>3 (finds LM)</td>
</tr>
<tr>
<td>0.58</td>
<td>42%</td>
<td>58%</td>
<td>3 (finds LM)</td>
</tr>
<tr>
<td>0.57</td>
<td>43%</td>
<td>57%</td>
<td>3 (finds LM)</td>
</tr>
<tr>
<td>0.56</td>
<td>44%</td>
<td>56%</td>
<td>9 (finds LM)</td>
</tr>
<tr>
<td>0.55</td>
<td>45%</td>
<td>55%</td>
<td>9 (finds LM)</td>
</tr>
<tr>
<td>0.54</td>
<td>46%</td>
<td>54%</td>
<td>13 (finds GM)</td>
</tr>
<tr>
<td>0.53</td>
<td>47%</td>
<td>53%</td>
<td>14 (finds GM)</td>
</tr>
<tr>
<td>0.52</td>
<td>48%</td>
<td>52%</td>
<td>11 (finds GM)</td>
</tr>
<tr>
<td>0.51</td>
<td>49%</td>
<td>51%</td>
<td>11 (finds GM)</td>
</tr>
<tr>
<td>0.5</td>
<td>50%</td>
<td>50%</td>
<td>11 (finds GM)</td>
</tr>
<tr>
<td>0.4</td>
<td>60%</td>
<td>40%</td>
<td>7 (finds GM)</td>
</tr>
<tr>
<td>0.3</td>
<td>70%</td>
<td>30%</td>
<td>12 (finds GM)</td>
</tr>
<tr>
<td>0.2</td>
<td>80%</td>
<td>20%</td>
<td>17 (finds GM)</td>
</tr>
<tr>
<td>0.1</td>
<td>90%</td>
<td>10%</td>
<td>15 (finds GM)</td>
</tr>
<tr>
<td>0</td>
<td>100%</td>
<td>0</td>
<td>Fails</td>
</tr>
</tbody>
</table>

Table 5.3: The efficiency of WEI for different values of $w$
(GM- Global Minimum; LM-Local Minimum)

The performances of three particular weights are chosen from Table 5.3 for further analysis and their history of searching the global minimum during the iterative process is shown in Figure 5.7 - 5.9 respectively. The numbered labels used in these three graphs represent the sequence of updating the sampling points throughout the whole optimization process. The same initial setting of sampling points is applied and labelled as the black dots in these graphs. When $w=0.1$ (Figure 5.7), 15 iterations (objective function evaluation) are necessary to converge at the global minimum of objective
function. But obviously the location of 11th, 13th and 14th updated sampling points is already extremely close to the global minimum. Compared with the result obtained by the normal EI, the kriging model assisted by WEI with the weighting parameter \( w = 0.1 \) spends 4 more iterations exploring the region with few sampling points, thus the approximation of the objective function’s shape is more accurate than normal EI. When \( w = 0.4 \) (Figure 5.8), the kriging with WEI emphasises slightly more on exploration and it only takes 7 iterations to converge globally. But unfortunately when \( w = 0.9 \) the kriging model is trapped in a local minimum and fails to converge to the global optimum. Although when \( w = 0.4 \) the WEI can guide the kriging model to locate the global minimum more efficiently than normal EI, the approximation of the shape of objective function in Figure 5.8 is not as good as the performance of EI in Figure 4.1. The content about the approximation of objective function’s shape will be discussed in the robust optimization later.

![Figure 5.7: Performance of kriging model with WEI when \( w = 0.1 \)](image-url)
Figure 5.8: Performance of kriging model with WEI when $w=0.4$

Figure 5.9: Performance of kriging model with WEI when $w=0.9$
At this stage, the appropriate weight plays the important role of searching the global minimum effectively and efficiently. However, it cannot guarantee that the best-performed value of weighting parameter achieved above can outperform other weights when coping with any kind of cases. The manner of adjusting the weights manually cannot satisfy the designer’s demand, so a strategy that is able to adapt to any test environment is necessary.

5.3.1 Negative Weighting Parameter

In this section, the work is mainly the preparation for building up the novel strategy which is able to adjust weights automatically. In WEI the sum of the weights distributed on the two components favouring exploitation and exploration respectively should be equal to 1. However, one new algorithm that will be illustrated in the next chapter includes an occasional appearance of negative weights during the iterative process; thus in order to understand the usage of the negative weights a set of tests has been done. The same test function and initial setting are utilized for comparing with the previous results. The cases of negative weights on exploration or exploitation can be classified as two categories: when the weighting parameter $w$ is set in the range $w \in (0,-1)$, which can be understood as excessive emphasis on exploration; whereas when $w$ is set as $w \in (1,2)$ means over emphasis on exploitation.

To verify the first one, four different negative values of the weighting parameter $w$ are chosen to test, which is $-0.01, -0.1, -1, -10$ respectively. The negative value of the weight parameter $w$ indicates that the impact of the component favouring exploitation is weakened excessively whereas the impact of exploration is amplified more. The following figures (Figures 5.10 - 5.13) present the history of selecting updated sampling points throughout the whole iterative process until the model is stopped by the termination criteria (when the repeated sampling point is found, the model will be stopped). Likewise, the sequence of updating sampling points is highlighted by the numbered label. By observing these four graphs, when $w$ is set as negative number the shape of the objective function is approximated well by the kriging with WEI, although ultimately the global minimum cannot be found precisely. But the best-performed solutions of these four cases are all particularly close to the global minimum as listed in Table 5.4. Because when $w$ is set as negative value the shape of the objective function
can be approximate well, probably the feature might be used to provide the approximation for the evaluation of robustness.

Figure 5.10: Performance of kriging surrogated model using the WEI when $w = -0.01$

Figure 5.11: Performance of kriging surrogated model using the WEI when $w = -0.1$
Figure 5.12: Performance of kriging surrogated model using the WEI when $w = -1$

Figure 5.13: Performance of kriging surrogated model using the WEI when $w = -10$
Chapter 5 Balance between Exploration and Exploitation

<table>
<thead>
<tr>
<th>The Value of $w$</th>
<th>The Performance of WEI</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.01</td>
<td>At 14th iteration, the best-performed point is found $(x=431, y=-406.3003)$</td>
</tr>
<tr>
<td>-0.1</td>
<td>At 13th iteration, the best-performed point is found $(x=432, y=-403.6559)$</td>
</tr>
<tr>
<td>-1</td>
<td>At 12th iteration, the best-performed point is found $(x=430, y=-406.6971)$</td>
</tr>
<tr>
<td>-10</td>
<td>At 12th iteration, the best-performed point is found $(x=433, y=-400.7649)$</td>
</tr>
</tbody>
</table>

Table 5.4: Performance of WEI with the four typical negative weighting parameters

5.3.1.2 Overemphasis on Exploitation

If the value of the weighting parameter is set in the range $w \in (1, 2)$, the weight on the component emphasising exploration should be negative. Thus the weights are placed disproportionately on exploitation and exploration, the exploitation is emphasized.
overwhelmingly which might be risky. From the 1\textsuperscript{st} to 4\textsuperscript{th} iteration the searching region tends to be around the minimum of the existing sampling points, whereas from 5\textsuperscript{th} to 8\textsuperscript{th} iteration the updated sampling points mainly appear in the unsampled region. The final three updated sampling points distributed concentrated around the minimum of existing sampling points, ultimately the kriging model is trapped in the local minimum rather than the global minimum. So to some extent, over emphasis set on exploitation might be quite risky.

5.4 Conclusion

Two methodologies called Generalized Expected Improvement and Weighted Expected Improvement respectively which are derived from the normal Expected Improvement is able to adjust the trade-off between exploration and exploitation via tuning the hyperparameters $g$ and $w$. The optimal choice of these hyperparameters determine directly to the performance of the kriging optimiser, both in terms of the ability of the kriging model to achieve the correct answer (global minimum) and doing it efficiently (fewer iterations required); unfortunately the optimal choice of these hyperparameters is normally problem dependent and thus numerous test experience is necessary. A modified strategy is required to make the method more intelligent and guide itself automatically through the optimisation process. The extreme setting of weighting parameter that might appears occasionally in the novel algorithms mentioned next chapter have been tested to verify the feasibility of it. Excessive emphasis on exploration is helpful to approximate the shape of objective function, whereas placing too much emphasis on exploitation may be a risky strategy.
Chapter 6 Reinforcement Learning Methods

6.1 Introduction

Two types of methodologies WEI and GEI derived from normal EI are able to guide the kriging model to update sampling points for further evaluation throughout the whole iterative process of searching global optimum of objective function meanwhile possessing the ability of balancing the exploration and exploitation via tuning corresponding hyperparameters. Through a series of numerical tests the feasibility of these two methodologies has been verified, moreover the importance of the optimal choice of the weighting parameter $w$ and the tuneable parameter $g$ is confirmed. Unfortunately the optimal choice of $w$ is normally problem dependent and thus a modified strategy is required to make the method more intelligent and guide itself automatically through the process. Compared with the GEI, tuning the weighting parameter to adjust the balance between exploration and exploitation for WEI is easier to be observed, thus it is chosen as the foundation for building the novel algorithm.

The idea of automatically tuning the weighting parameter $w$ in response to the environment feedback is considered. As a goal-directed learning methodology, reinforcement learning owns the ability of determining what to do next and how to map the situation to actions so as to maximize a numerical reward [46]. As the kernel of reinforcement learning, a decision maker is capable to select the action among different given options and then implement it for the next step referring to the current performance and reward raised by this action. The reward is classified into two kinds: one is current reward for short-term; whereas the other one is accumulative reward for long-term, which may be considered as the total reward through the whole optimization process. One intelligent decision maker cannot only apply the action searching the known area, the ‘exploitation’, to earn more current reward, but more importantly it also takes account of the long-term accumulative reward to select the action exploring the area with high uncertainty. Normally via the pre-estimating the impact brought from different actions, the decision maker can finally find the appropriate combination of actions in order to maximize the total rewards rather than overemphasising the short-term reward at certain steps. These following four elements mainly formulate a
reinforcement learning system: a policy (the strategy of choosing exploration and exploitation), a reward function (the approach to calculate reward), a value function (a statistical method to sum and value the reward) and a model of the environment. Figure 6.1 illuminates the basic procedure of reinforcement learning transition.

Figure 6.1: A transition graph of the reinforcement learning

The above transition graph is a helpful way to summarize the reinforcement learning process. There are four kinds of sub-elements that can be seen in the transition graph. They include action, statement, reward (advantage after taking action), and policy $\pi$. Here reward and policy need to be explained more specifically. A policy defines the way of choosing actions learning from the feedback produced by environment, for instance, one usually used policy called $\pi$-policy can decide the possibility of taking different actions based on the environment feedback at the specific statements.

Following the strategy of selecting actions the decision maker often faces the dilemma of choosing exploration or exploitation. For example normally if choosing exploration
among other actions it may not produce the best performance immediately, but probably might be beneficial for long-term advantage [47]. Different from the traditional reinforcement, the experience owned by decision maker can be obtained from testing different actions and observing their impact consistently. In our research, the action is the calling of the time-consuming FEM model, thus trying every optional action to summarise the best combination of actions would sacrifice the original advantage brought from the kriging surrogate model. Therefore, how to use the reinforcement learning theory to balance exploration and exploitation automatically and adaptively without impacting the efficiency of optimization is considered.

6.2 Using Reinforcement Learning Methods for Balancing Exploration and Exploitation

The concept of the reinforcement learning and the fundamental principle of building the reinforcement learning framework are illustrated in the last section. Numerous sequential decision-making problems can be solved by reinforcement learning methods. Due to maximizing the final cumulative reward [48], the decision maker needs to consider both short-term and long-term impact produced by actions. The idea of reinforcement learning could be used in combination with the WEI to distribute the weights on exploration and exploitation properly adapting the feedback produced by existing actions.

Although the WEI utility function has been discussed before, here it is a need to review it briefly.

\[
WEI(x) = \begin{cases} 
  w(f_{min} - \hat{y}(x)) \Phi \left( \frac{f_{min} - \hat{y}(x)}{s(x)} \right) + (1 - w) s(x) \phi \left( \frac{f_{min} - \hat{y}(x)}{s(x)} \right) & \text{if } s(x) > 0 \\
  0 & \text{if } s(x) = 0
\end{cases}
\]

(6.1)

The WEI method owns a tuneable weighting parameter (0<w<1) controls the balance between the two terms (exploration and exploitation), thus searching globally and locally. The corresponding numerical tests that have been presented in Chapter 5 confirm the importance of the optimal choice of the weights, both in terms of the ability of the algorithm to achieve the global minimum and doing this efficiently.
Unfortunately for obtaining optimal value of weighting parameter large numbers of tests needs to be done, moreover the optimal weight parameter specialized to certain task cannot guarantee that it can perform outstandingly in other cases. For example different objective functions or even same objective function but different initial setting of sampling points may cause the optimal value of weighting parameter different, thus it is unrealistic to expect that one weight value would be optimal to all possible optimisation problems (no free lunch theorem). Therefore a strategy that is able to balance exploration and exploitation to find global optimum precisely and at the same time to make efficient use of information already found is necessary now.

### 6.3 Adaptive Weighted Expected Improvement Algorithm

In this section a novel algorithm which utilizes the idea of reinforcement learning and is able to tune the weighting parameter $w$ automatically in response to the environment feedback will be introduced. In particular, the Mean Square Error (MSE) from the kriging model is used to guide the choice of the optimum weight $w$ and the concept of reward is introduced. By calculating the difference of the average MSE over the full design range between the given iterations, the rewards from two different weight distribution, one with emphasis whereas the other on exploration or exploitation are achieved. By comparing the pairs of rewards (Reward1 and Reward2) in pre-test, the weights are redistributed on the two terms which favour the exploration and exploitation respectively so that the biggest reward is achieved. The Adaptive Weighted Expected Improvement (AWEI) strategy is described as one of the possible algorithms in Figure 6.2. AWEI endeavours to encourage the alternative options emphasising exploration or exploitation depending of the results of the initial pre-test.
Here the initial setting of critical parameters should be explained further. Before the pre-test, at each iteration the alternative weights $w_1$ and $w_2$ in Figure 6.2 are calibrated as 0.5 (the Expected Improvement), and then after comparing the rewards produced by the pre-test from separated components emphasising exploration and exploitation respectively in EI utility function guide the redistribution of weights since $w_1^*$ and $w_2^*$ are intermediate variables.

The following annotations explain the specific meaning of the signs and abbreviation used in this flowchart.
\(N\): the number of iterations

**AMSE 1(N)**: the average value of mean square error of the action 1 gained from the current pre-test, which emphasis on exploitation

**AMSE 2(N)**: the average value of mean square error of the action 2 gained from the current pre-test, which emphasis on exploration

**AMSE (N-1)**: the average value of mean square error gained from the previous approximation (not the pre-test) depended on the redistributed weights

Thus, the rewards of the two pre-tests can be calculated as \(\text{reward1} = AMSE(N - 1) - A \quad E1(N)\) and \(\text{reward2} = AMSE(N - 1) - AMSE2(N)\). The estimation of reward1 and reward2 will be used to get the intermediate weight value \(w1^*\) and \(w2^*\) assisted by the parameter \(\beta\) which decides the magnitude of tuning the weighting parameters.

\[
w1^*(N) = w1(N - 1) + \beta \left( \frac{\text{Reward1}(N)}{AMSE(N-1)} \right) \quad (N \text{ is the number of iterations}) \tag{6.2}
\]

\[
w2^*(N) = w2(N - 1) + \beta \left( \frac{\text{Reward2}(N)}{AMSE(N-1)} \right) \quad (N \text{ is the number of iterations}) \tag{6.3}
\]

The Equation (6.2) and (6.3) are the approaches of calculating the intermediate weights \(w1^*(N)\) and \(w2^*(N)\) (not the final redistributed weights \(w1(N)\) and \(w2(N)\) at the \(N_{th}\) iteration), respectively. The AWEI algorithm endeavours to encourage exploration or exploitation depending on the results of the initial pre-test. The setting of the parameter \(\beta\) will be discussed in detail at the next section.

\[
w1(N) = \frac{w1^*(N)}{w1^*(N) + w2^*(N)} \quad (N \text{ is the number of iterations}) \tag{6.4}
\]

\[
w2(N) = \frac{w2^*(N)}{w1^*(N) + w2^*(N)} \quad (N \text{ is the number of iterations}) \tag{6.5}
\]

The pair of final optimal weights \(w1(N)\) and \(w2(N)\) which consist of the intermediate weights \(w1^*(N)\) and \(w2^*(N)\) will be applied to the exploration and exploitation terms in WEI at the \(N_{th}\) iteration respectively. However when coping with the practical task, this intelligent scheme of tuning weights still needs to apply certain standby actions in
order to prevent the test failure of any of the two optional weights during pre-test. There are more details about the difficulty appeared in practical experiments will be presented in next section.

6.4 Practical Performance of the Adaptive Weighted Expected Improvement

The rewards calculated by the variance of the average mean square error from two parallel actions, the ‘exploration and exploitation terms in the EI utility function’, guide the kriging surrogate model to put more emphasis on the term obtaining more rewards. However, as demonstrated in Chapter 4, the feature of the two terms favouring exploration and exploitation in EI utility function has been analysed respectively. One particular problem was identified and needed special attention. The term which encourages exploitation can sometimes cause the kriging model to stop because of choosing repeatedly the same new point for evaluation (within the specified accuracy). Here, in the pre-test, the alternative exploitation term may be extremely likely to cause the kriging model selecting the sampling points repeatedly in the area already with abundant known information, thus fails to achieve the reward from this optional pre-test. Should this happen (or should – for any other reason – one of the rewards not be assessed properly or fail), the algorithm is effectively reset and the EI function is temporarily applied to select the next point for evaluation; in the next step the algorithm reverses to the AWEI. The specific process of balancing exploration and exploitation is shown in the following decision-making chart.
In Figure 6.3, as a critical intermediate parameter, the tuneable parameter $\beta$ controls the impact brought from rewards to the present weights balancing exploration and exploitation. The tuneable parameter $\beta$ determines if the kriging surrogate model can finally converge to the global optimum successfully, thus in order to achieve an appropriate value of $\beta$ a series of numerical tests needs to be done. As illustrated in the section 4.4 of Chapter 4, along with the iterative process going the average value of the
MSE over all the design space might increase instead of declining consistently. If any of
the parallel actions in the given pre-test causes the average mean square error increases
like \( AMSE1(N) \geq A(E(N - 1) \) or \( AMSE2(N) \geq AMSE(N - 1) \), the reward
\( \ \text{reward1} = AMSE(N - 1) - AMSE1(N) \) and \( \text{reward2} = AMSE(N - 1) - AMSE2(N) \)) obtained via observing the variation of average mean square error can be
close to zero or even negative value. Then the unappropriated value of \( \beta \) may cause the
weighting parameters \( w1 \) or \( w2 \) to be zero or even negative, moreover the occasional
appearance of negative weighting parameter may affect the convergence process
negatively, which has been verified in Chapter 5. We have used the Schwefel test
function again with the same settings of initial sampling points imposed as \( x = -450, -230, -150, -40, 160, 500 \) and the tuneable parameter \( \beta \) varied in a controlled way as
follows.

<table>
<thead>
<tr>
<th>( \beta )</th>
<th>The number of iterations to find the global minimum of objective function</th>
<th>( \beta )</th>
<th>The number of iterations to find the global minimum of objective function</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001</td>
<td>16 iter</td>
<td>0.08</td>
<td>12 iter</td>
</tr>
<tr>
<td>0.005</td>
<td>16 iter</td>
<td>0.1</td>
<td>12 iter</td>
</tr>
<tr>
<td>0.01</td>
<td>16 iter</td>
<td>0.2</td>
<td>8 iter</td>
</tr>
<tr>
<td>0.05</td>
<td>12 iter</td>
<td>0.3</td>
<td>8 iter</td>
</tr>
</tbody>
</table>

Table 6.1: Performance of kriging model with AWEI for different value of tuneable parameter \( \beta \)

In Table 6.1, when \( \beta = 0.001, 0.005, 0.01, 16 \), iterations are required to find a global
minimum, thus AWEI performs slightly worse than the kriging model assisted by EI
with the same initial setting; when \( \beta = 0.05, 0.08, 0.1 \) altogether 12 iterations are needed
to find the global minimum; when \( \beta = 0.2 \) or 0.3 a better performance is observed with
eight iterations needed to find the global minimum. In order to track the history of
selecting sampling points throughout the convergence process, several tests with the
typical value of \( \beta \) have been demonstrated from Figure 6.4 to Figure 6.11.
Figure 6.4: The Adaptive Expected Improvement for the tuneable value $\beta = 0.001$

(Legend: circle with a number denotes iterations where exploration term is dominant; square with a number denotes iterations where exploitation term is dominant.).

Figure 6.4 shows an approximation result for the tuneable value $\beta = 0.001$; the sampling points labelled by the corresponding number demonstrates the history of infilling sampling points. Two distinct symbols, ‘a circle or a square’, are used to differentiate the action emphasising exploration or exploitation at the specific iteration. When $\beta = 0.001$, the kriging model requires 16 iterations to converge at the global minimum and in the mean time the two local minima ($x = -303, = -300.5 ; x = -125, x = -122.8724$) are also found. It can be found that amongst the 16 actions there are 15 of them encouraging exploration rather than exploitation, since only in the last iteration the action favouring exploitation is applied. Compared with the previously described results of the kriging assisted by normal EI, the kriging model with AWEI when $\beta = 0.001$ emphasize more exploration. Although, except for the region (-50 $\leq x \leq$ 150), in the remaining design space the quality of approximating the shape of the objective function is better than the prediction achieved by the kriging with EI. This potential feature may be utilized for the robust design in later which requires the optimizer to predict the shape of the objective function as accurately as possible rather than only to find out the global minimum successfully.
Figure 6.5: The Adaptive Expected Improvement for the tuneable value $\beta = 0.05$
(Legend: circle with a number denotes iterations where exploration term is dominant; square with a number denotes iterations where exploitation term is dominant.).

Figure 6.6: Performance of kriging model with AWEI for $\beta = 0.1$
(Legend: circle with a number denotes iterations where exploration term is dominant; square with a number denotes iterations where exploitation term is dominant.).
In order to observe the correlation between the tuneable parameter $\beta$ and the balance between exploration and exploitation, the three more typical tests ($\beta = 0.05, 0.1$ and $0.3$) have been done. When $\beta = 0.05$ and $0.1$, they both take 11 iterations to locate the global minimum; when $\beta = 0.3$, a better performance is observed that only eight is necessary to find the global minimum. Along with the increasing value of $\beta$, the number of actions favouring exploitation also increases which indicates faster convergence.

To achieve the best performed value of $\beta$, more tests are necessary. The best performance with only five iterations needed was observed when $\beta = 0.35$ and $\beta = 0.40$, as demonstrated in Figure 6.8 and Figure 6.9. Based on these results, it can be concluded roughly that in the range of $\beta$ from 0.001 to 0.35 the efficiency of finding global minimum increases meanwhile the kriging model frequently applies the action favouring exploitation; when $\beta = 0.35$ and $\beta = 0.40$ the best performance has been observed, since when $\beta$ is from 0.4 to 2 the AWEI requires more iterations than the best performance to converge at the global minimum; finally until when $\beta = 3$, the AWEI fails to find the global minimum. A summary of the results of testing the kriging
surrogate model assisted by the AWEI algorithm for different values of $\beta$ is presented in the column chart (Figure 6.12).

![Figure 6.8](image1.png)

Figure 6.8: The best performance of AWEI when the tuneable parameter $\beta$ is equal to 0.35 (Only require 5 iterations to find the global minimum) (Legend: circle with a number denotes iterations where exploration term is dominant; square with a number denotes iterations where exploitation term is dominant.).

![Figure 6.9](image2.png)

Figure 6.9: Performance of a kriging model with AWEI for $\beta = 0.40$ (Legend: circle with a number denotes iterations where exploration term is dominant; square with a number denotes iterations where exploitation term is dominant.).
Chapter 6 Reinforcement Learning Methods

Figure 6.10: Performance of a kriging model with AWEI for $\beta = 0.6$
(Legend: circle with a number denotes iterations where exploration term is dominant; square with a number denotes iterations where exploitation term is dominant.).

Figure 6.11: Performance of a kriging model with AWEI for $\beta = 0.8$
(Legend: circle with a number denotes iterations where exploration term is dominant; square with a number denotes iterations where exploitation term is dominant.).
Chapter 6 Reinforcement Learning Methods

The efficiency of finding global minimum using kriging surrogate model assisted by AWEI with different value of $\beta$

![Bar chart showing the number of iterations](chart.png)

Figure 6.12: A summary of the test results using AWEI for different values of $\beta$.

6.5 Conclusion

The importance of choosing proper weights distributed on exploration and exploitation in WEI has been confirmed form the point of view of convergence to the global optimum effectively and efficiently; however, an optimal weighting parameter is difficult to be found. Even the best performing weight under certain special situation can hardly guarantee that it will also cope with other tasks and perform equally well. Thus a better algorithm is essential which can adjust the weighting parameter in response to environment adaptively and automatically. The idea of reinforcement learning has been implemented to construct one new strategy called Adaptively Expected Improvement which can guide the kriging surrogate model how to map the situation and how to choose the next action so as to balance the exploration and exploitation automatically and intelligently. Through sensing the impact of the built-in feedback from the previous actions, this algorithm can calculate the reward via comparing the current performance and the future performance obtained from pre-tests for different optional actions. The main contribution of AWEI is that through observing the variation of the average mean square error of the overall design space the reward is
obtained and the pre-test involves the actions favouring exploration and exploitation respectively. Via testing these alternative actions in advance, the one with a better reward would be applied next. In addition, the setting of intermediate parameter $\beta$ controls the magnitude of tuning the redistributed weights, hence a number of tests have been undertaken to achieve the optimal value for further use.
Chapter 7 Surrogate Model based Weighted Expected Improvement approach with rewards

7.1 Introduction

Based on the theory of reinforcement learning, the AWEI illustrated in last chapter is able to take account of the feedback produced by the kriging model during the iterative process, and then guides the model itself redistribute the weights on the exploration and exploitation respectively [49]. The updated sampling points selected by the WEI with different weights in the pre-test of the prototype of AWEI need to call the real objective function which should be unacceptable when solving the practical problem by FEM modelling. Moreover the pre-test of AWEI only makes an anticipation of optimal weight combination for next iteration in order to obtain the largest reward [48]. It may be viewed as an emphasis of immediate interests and may not necessarily be helpful to gain the long-term benefits.

In this chapter, another new strategy is developed which is able to predict the cumulative rewards likely to occur in long-term as a consequence of a particular choice of actions. This approach also follows the ideas of reinforcement learning first introduced in [46] in the context of games theory to a well-known one-armed bandit problem [48]. Only assessing the short-term rewards may be not enough to guide the kriging model converge to the global optimum successfully, whilst the pre-test of AWEI need more information from the real objective function. But this novel strategy mainly focuses on collecting the long-term rewards/accumulative rewards at a given iteration step based on a ‘simplified surrogate model’ built especially for pre-test which only uses existing information produced by the kriging model itself rather than the information from the real objective function [48]. During the pre-test process, a pair of weights (one emphasising exploration and the other exploitation) are chosen as two optional actions in parallel for determining how to redistribute the weights on these two alternative terms. As the most critical capability, this new strategy utilizes the combination of the most recent prediction results and the corresponding mean square error to build the simplified surrogate model for testing the alternative actions preliminarily and then the rewards achieved from the pre-test would guide the optimizer.
to balance exploration and exploitation adaptively. If either of the optional actions in the pre-test fails to be tested because of the termination criteria, then the normal EI would be applied automatically as the standby approach for selecting the new sampling point for further prediction in the next iteration. The novel strategy, named the ‘Surrogate Model based on Weighted Expected Improvement approach with rewards (SMWEI)’ is fully depicted in Figure 7.1.

Figure 7.1: The SMWEI for making decision based on long-term rewards for balancing exploration and exploitation
7.2 The Surrogate Model of Objective Function for ‘Forward Prediction’

In order to evaluate the long-term rewards, a cheap surrogate model only utilizing the existing information is essential rather than requiring more information from FEM modelling. Inevitably such a pre-test will be less reliable than using real data points (which in this application, as already stressed, would be too expensive and thus unacceptable) but may result in an overall better assessment of long-term benefits of different actions than a simple one-stage algorithm developed and described earlier as the Adaptive Weighted Expected Improvement strategy. So how to build the simplified surrogate model for the pre-test of observing the impact of different optional actions becomes a critical issue.

There are of course many methods of constructing a surrogate model. We suggest, and have implemented, using the Root Mean Square Error (RMSE) already available in the kriging prediction, even though this is not the real error between the kriging approximation and the real objective function (which is of course unknown at this stage). A mean square error distributed randomly is added to the specific approximation and thus effectively we now have two kriging surrogate models simultaneously, the original one based on the most recent “real” data points, and a second one – used only for the purpose of “forward prediction” of the long-term effects of a particular action – which ultimately leads to an overall long-term “reward” of a particular action. As there are two possible actions, and they are assessed independently, we end up with two rewards; the better reward will identify the better cause of action, a new point is selected, the finite element programme executed and a new point added to the curve. This will give rise to a new surrogate model and a new “secondary” surrogate model (or rather a pair of models as there are two parallel actions); the process will continue until some termination criteria are met. The randomly distributed RMSE is added to the specific approximation and thus effectively we now have two kriging surrogate models simultaneously, the original one based on the most recent ‘real’ data points, and a second one – used only for the purpose of ‘forward prediction’ of the long-term effects of a particular action – which ultimately leads to an overall long-term ‘reward’ of a
particular action. The second simplified surrogate model for the pre-test is built in the following way:

The Simplified Surrogate model for pre-test = The most recent kriging prediction + Gaussian distributed the accessory RMSE (produced by the kriging model).

### 7.2.1 The Simplified Surrogate Model in Numerical Tests

In the previous section, the process of how to construct a simplified surrogate model for the pre-test of SMWEI was demonstrated. Because the simplified surrogate model involves both the randomly distributed RMSE and the most recent prediction result, it is found that the trajectory of this data combination looks like jagged or fluctuated as shown in Figure 7.2(a), 7.2(c), 7.2(e). If directly utilizing this original data with drastic fluctuations, the roughness of the data would easily misguide the kriging model to make mistaken decision. Therefore a smoothing treatment is applied to smooth the shape of the original data. As before, the Schwefel test function is chosen as the test function here in order to compare to the other methods. In Figure 7.2, three pairs of the data of simplified surrogate model and the data with the smoothing treatment at three different iterations, the first, fifth and tenth iteration are listed respectively. It can be found that along with the ongoing iterative process the prediction results get more accurate, whilst due to obtain more objective function evaluation the trajectory of the simplified surrogate model gradually evolves to be more similar to the real objective function relatively.
Chapter 7 Surrogate Model based Weighted Expected Improvement approach with rewards

(a) The simplified surrogate model for pre-test without smoothing treatment at the first iteration

(b) The simplified surrogate model for pre-test with smoothing treatment at the first prediction
Chapter 7 Surrogate Model based Weighted Expected Improvement approach with rewards

(c) The simplified surrogate model for pre-test without smoothing treatment at the fifth iteration

(d) The simplified surrogate model for pre-test with smoothing treatment at the fifth prediction
(e) The simplified surrogate model for pre-test without smoothing treatment at the tenth iteration

(f) The simplified surrogate model for pre-test with smoothing treatment at the tenth prediction

Figure 7.2: The two versions of simplified surrogate model obtained by SMWEI (with smoothing treatment and without it) at the first, fifth, tenth iterations respectively

\(w_1=0.7, \ w_2=0.1\)
Two weights, $w_1=0.7$ and $w_2=0.1$, have been chosen as alternative optional weights in the pre-test, favouring exploration and exploitation, respectively. During the period of pre-test, only the existing information has been used rather than calling the FEM model. Because the simplified surrogate model is computationally cheap, the observation of long-term rewards has been considered. Throughout the iterative process the number of inner iterations in the pre-test is set as ten, rather than one for observing the performance of the alternative actions in long-term, which has been applied all the following tests. The Schwefel test function and the same initial sampling points as before were used for comparison with the previous results, since in reality such initial sample points should be distributed randomly or chosen by a sampling strategy such as Latin Hypercube sampling [40]. Several pairs of specific weights were chosen to test, and the results are discussed below. Finally, as the “second” kriging surrogate model relies on a random distribution of error, all tests were conducted ten times with the same pair of values of $w_1$ and $w_2$ and the performance averaged.

Figure 7.3: The performance of the kriging with SMWEI when with $w_1=0.7$ and $w_2=0.3$.
(Legend: circle with a number denotes iterations where exploration term is dominant; square with a number denotes iterations where exploitation term is dominant.)
When $w_1=0.7$ and $w_2=0.3$ the number of iterations required to find the global minimum is between 9 and 16 during the ten tests and the average number of iterations for this case is 12. One of the results is shown in Figure 7.3 and the following comments can be made: the kriging model with SMWEI takes 11 iterations to find out the global minimum; among the 11 iterations there are 6 actions emphasizing exploration and 5 actions emphasizing exploitation; there is an interesting ‘departure’ at the tenth iteration to explore a remote region rather than the recently searched region. In addition to this setting of alternative weights, there are more tests have been done: when $w_1 = 0.6$ and $w_2 = 0.4$ the number of iterations required to find a global minimum is between 10 and 16, for $w_1 = 0.8$ and $w_2 = 0.2$ it is 10-17, finally for $w_1 = 0.9$ and $w_2 = 0.1$ it is 11-15. The average number of iterations for the pairs of values above is 13, 15 and 12, respectively, so it is quite steady and does not appear to be sensitive to the variation of the values. One particularly ideal performance appears when $w_1 = 0.7$ and $w_2 = 0.1$ as demonstrated by Figure 7.4. For this case only 6 iterations were required to find the global minimum of the objective function.

![Figure 7.4](image.png)

Figure 7.4: The ‘best’ performance of SMWEI with $w_1=0.7$ and $w_2=0.1$

(The meaning of labels is same as Figure 7.3.)
There appears to be little benefit in applying the last strategy compared with the previous AWEI algorithm, but more testing is required to draw more meaningful conclusions. In particular, it is interesting to notice that the SMWEI clearly makes a better attempt at exploring local minima – this may prove very important in the context of robust design where not only the value but also the shape of the minimum is of relevance. Thus, a strategy which explores the space more thoroughly may after all be preferable even if more expensive.

7.3 The Performance of SMWEI facing an ‘Extreme case’

To become confident about the reliability of SMWEI, the kriging model with the novel strategy needs to be taken into certain extreme environment to test. The above tests all applied the initial sampling points dispersed abundantly, however in many optimisation problems the searching space is not given or the initial sampling points may not be favourably distributed throughout the searching space. This means that the very limited local information presents substantial difficulty in the prediction process. We have therefore decided to assess the performance of SMWEI under an “extreme case” when the initial sampling points are located only in a narrow part of the searching space. Hypothetically, if all initial sampling points are close to the valley of a local minimum, the optimizer may easily be guided to converge to a local minimum rather than finding the global one successfully under the overemphasis of exploitation.

A set of corresponding test results are shown in Figures 7.5 - 7.8 to verify the reliability of SMWEI. For the first case, the initial sampling points are concentrated in the middle of the test range of the objective function, where there seems to be a plateau compared with other strongly fluctuated parts. Two distinct extreme cases of initial point distributions were set to test the robustness of the SMWEI algorithm, while at the same time comparing with the performance of a kriging model assisted by EI with the same initial test environment. The specific test results of the first ‘extreme’ case are presented in Figure 7.5 and Figure 7.6; they are the kriging models with SMWEI and EI, respectively. The initial sample points are located in a narrow region in the middle of the objective function. When \( w_1 = 0.6 \) and \( w_2 = 0.4 \), the SMWEI performs better than the EI and only requires 11 iterations to find the global minimum as compared with EI which requires 20 iterations. In Figures 7.5 - 7.8, the unfavourable distribution of initial
sample points is definitely a challenging case as the initial points give very little clue as to the optimizer. Particularly, the distribution of initial points shown in Figures 7.7 and 7.8 is likely to puzzle the kriging surrogate model to converge at the local minimum, because the initial sample points are highly concentrated on one slope of a valley of the local minimum and one of the initial sampling points is extremely close to the local minimum. Rather remarkably the SMWEI algorithm performs very robustly with only nine iterations to find the global minimum which includes three actions favouring exploration, four actions favouring exploitation and two actions applying normal EI because of the failure of the pre-test, whereas the kriging with EI needs 12 iterations.

Figure 7.5: The performance of the kriging with SMWEI when coping with an ‘extreme’ case (Legend: circle with a number denotes iterations where exploration term is dominant; square with a number denotes iterations where exploitation term is dominant.).
Chapter 7 Surrogate Model based Weighted Expected Improvement approach with rewards

Figure 7.6: The performance of the kriging assisted by the EI when coping with an ‘extreme’ case

Figure 7.7: The performance of the kriging with SMWEI when coping with an ‘extreme’ case (Legend: circle with a number denotes iterations where exploration term is dominant; square with a number denotes iterations where exploitation term is dominant.).
7.4 Conclusion

In addition to AWEI, another novel algorithm called SMWEI has been proposed, also adopting concepts of reinforcement learning, in an attempt to automatically balance exploration and exploitation in computationally expensive electromagnetic design optimization problems. It is also based on kriging surrogate modelling and uses the notion of rewards for selecting the best position of further evaluation. In the pre-test of the SMWEI method, a simplified surrogate model utilizes the combination of the most recent prediction result and the randomly distributed mean square error over the design space for test the different optional weights to achieve the best-performed one. Different setting of test condition has been tried, and the SMWEI indeed performs the high confidence of converging to the global optimum even in certain very difficult test environment. From the above test results, it is found that this novel strategy with the capability of automatically tuning the weights on exploration and exploitation can offer more reliable approximation about the shape of the objective function than the currently existing utility function. Currently, pursuing the global optimum is the unique target of the model, but if the robustness of the achieved optimum requires to be considered.
Probably this feature may be helpful for the robustness evaluation which will be demonstrated later, and this algorithm will be implemented in practical design of electromagnetic and electromechanical devices.
Chapter 8 Strategies for balancing exploration and exploitation in electromagnetic optimisation

8.1 Introduction

Balancing exploration and exploitation adaptively with automatic response of environmental feedback has been realised by using a reinforcement learning approach, as demonstrated before [50]-[52]. However, in practical design problems, in addition to finding the global optimum efficiently, the robustness of the optimal solution may also need to be considered due to the uncertainties existing in design variables. The improved kriging surrogate model with the help of the intelligent strategy of balancing exploration and exploitation is able to guarantee the high success rate of converging to the global optima as well as the relatively high accuracy of approximating the shape of objective functions, thus it is considered to provide the prediction for the robustness evaluation. The quality of the approximation of objective functions’ shape is closely related to the number of known sampling points and the distribution of them over the design space. The termination criterion of the kriging model determines the number of updated sampling points. The two currently used termination criteria are illustrated in this chapter. One two-variable numerical function including the robust and rugged area is especially designed for verifying the reliability of finding the global optimum and observing the approximation quality of the function’s shape. In addition, the specific weights distribution during the iterative process is exposed for better understanding the process of balancing the exploration and exploitation.

8.2 Kriging with different strategies for balancing exploration and exploitation

The concept of the kriging method and the currently existing methodologies for seeking the new sampling points was illustrated in detail in Chapter 4. By comparing the reliability and efficiency between different approaches of selecting sampling points, the WEI method was found to perform better than other utility functions with certain combination of the weighting parameters. Then through testing the exploration and
exploitation terms in WEI separately the importance of balancing them were confirmed. In practical experiments, the exploration term performs dramatically better in terms of finding the global optimum since the exploitation tends to search the area containing the minimum of known sampling points.

As suggested by previous tests in Chapter 6, 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 further algorithms using reinforcement learning [46] called Adaptive Weighted Expected Improvement (AWEI) and Surrogate Model based Weighted Expected Improvement approach with rewards [48] (SMWEI) [52] have been proposed to make the process of tuning weights more intelligent and self-guiding.

The Mean Square Error (MSE) produced by the kriging model is used to calculate the rewards as the basis of how to redistribute the weights on exploration and exploitation. The AWEI algorithm can tune the weights automatically based on the comparison between the potential rewards from two optional weight combinations emphasising exploitation and exploration, respectively. After comparison, the best-performed 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.

8.2.1 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 8.1. 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 [52], 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 expected improvement (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. The target of selecting the best-performed weight according to the feedback from the pre-test is to reduce the average standard error provided by kriging model itself. Whilst considering applying the kriging model dealing with the robust design problems, the critical prerequisite is that kriging can provide enough accuracy of the prediction which not only includes locating the global minimum precisely but fixes the shape of objective function as well. As the potential error, the mean square error, is able to indicate the confidence about the response surface at each point within the design range, so it is used to judge the effect of the predicting outcome. Always choosing the optimal weights, which can reduce the average mean square error fastest in the pre-test, makes the kriging model approximate the shape objective function as precisely as possible. The improved decision-making chart of the actual implementation is shown in Figure 8.1.
8.2.2 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 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 8.2.

Since all pre-tests in SMWEI apply a ‘cheap’ simplified surrogate model based on the specific prediction and potential error produced by kriging, the Mean Square Error (MSE) might be directly used in each pre-test’s remaining iterations instead of the...
Chapter 8 Strategies for balancing exploration and exploitation in electromagnetic optimisation

Expected Improvement. The simplified surrogate model in the pre-test, quite rough initially, is increasingly becoming accurate as a result of adding objective function calls; therefore the MSE might guide the kriging model directly to search the regions of the simplified surrogate model with high uncertainty. Figure 8.2 demonstrates how this strategy tunes the weights adaptively throughout the whole iterative process.

![Diagram showing the proposed pre-test strategy of AWEI](image)

**Figure 8.2:** The proposal of using the pre-test strategy of AWEI instead of the current pre-test
8.3 Practical performance of kriging with EI in testing the two-variable Schwefel function

The efficiency and effectiveness of converging at the global optimum using the kriging with normal EI and with the other two novel strategies has been tested with the two-variable Schwefel test function [39] 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 \((d=2)\):

\[
f(x) = \sum_{i=1}^{d} -x_i \sin(\sqrt{|x_i|})
\]

(8.1)

Figure 8.3 shows the contour of the two-variable Schwefel Test Function including one global minimum and several local minima, which are distributed irregularly and even the value of several local minima are specially similar with the global minimum, which might tempt the kriging model to be trapped in the local minima. Hence this function is acknowledged to provide a stern test to optimization algorithms.
8.3.1 The practical performance of the kriging with normal EI

In order to analyse and compare the effects of applying different strategies to guide the kriging method for searching the global optimum, the two-variable Schwefel function with the eight initial sampling points imposed at \( x_1 = -450, x_2 = -480; x_1 = -300, x_2 = 280; x_1 = -300, x_2 = -280; x_1 = -450, x_2 = 380; x_1 = 450, x_2 = -380; x_1 = 450, x_2 = 480; x_1 = 300, x_2 = 280; x_1 = 0, x_2 = 0 \) were used to test the EI and other novel strategies with the function of tuning weighting parameter automatically. Figure 8.4 shows that 12 iterations are necessary to find the global minimum of the objective function when using expected improvement. The labels of numbers tagged on the sampling points show the ‘history’ of how the sampling points were updated throughout the whole iterative process for searching the global minimum.

It can be seen that compared with the real objective function presented in Figure 8.3 the predicted result by kriging with EI might not approximate the shape of the Schwefel function ideally when the global optimum is found out. Because here the termination criterion is related to the surrogate model finding the location of the global minimum the surrogate model stops. In fact, when testing the practical designs, there is no information to indicate where the global minimum of the objective function is. Therefore how to stop the surrogate model – the ‘termination criterion’ – is one of our
research emphasises. Currently, there are two possible approaches that could be set as termination criteria: one is that because the surrogate model gets a certain average mean square value this could be used as an indicator to halt the iterations once a small enough mean square error is reached; the other is that iterations are terminated when repeated sampling points start to appear.

8.4 Practical performance of SMWEI with multi-weights in pre-test

In order to determine the optimal weights throughout the iterative process, more weights have been set as options in the pre-test of SMWEI. In the pre-test, each optional weight would receive feedback and then, through comparing the rewards, the optimal weight holding the largest reward from the pre-test would be applied in this iteration. Figure 8.5 shows one example applying the SMWEI with the strategy of multi-weights, and it takes the same number of updated sampling points as EI to find the global minimum. The optimal weights for each iteration are presented by different label in Table 8.1 to differentiate them depending on the emphasis on exploration or exploitation.

Figure 8.5: The performance of the kriging assisted by SMWEI with multi-weights to predict the two-variable Schwefel function (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).
Chapter 8 Strategies for balancing exploration and exploitation in electromagnetic optimisation

<table>
<thead>
<tr>
<th>Iteration number</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$y$</th>
<th>The optimal weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-280</td>
<td>-280</td>
<td>-478.7100</td>
<td>0.3 (Emphasis on Exploration)</td>
</tr>
<tr>
<td>2</td>
<td>-300</td>
<td>-270</td>
<td>-478.5300</td>
<td>0.6 (Emphasis on Exploitation)</td>
</tr>
<tr>
<td>3</td>
<td>-310</td>
<td>-300</td>
<td>-593.2100</td>
<td>0.9 (Emphasis on Exploitation)</td>
</tr>
<tr>
<td>4</td>
<td>-300</td>
<td>-310</td>
<td>-593.2100</td>
<td>0.7 (Emphasis on Exploitation)</td>
</tr>
<tr>
<td>5</td>
<td>-300</td>
<td>-300</td>
<td>-599.4800</td>
<td>0.7 (Emphasis on Exploitation)</td>
</tr>
<tr>
<td>6</td>
<td>500</td>
<td>420</td>
<td>-238.2800</td>
<td>0.4 (Emphasis on Exploration)</td>
</tr>
<tr>
<td>7</td>
<td>350</td>
<td>500</td>
<td>229.8700</td>
<td>0.3 (Emphasis on Exploration)</td>
</tr>
<tr>
<td>8</td>
<td>500</td>
<td>500</td>
<td>361.1800</td>
<td>0.4 (Emphasis on Exploration)</td>
</tr>
<tr>
<td>9</td>
<td>440</td>
<td>420</td>
<td>-792.6300</td>
<td>0.2 (Emphasis on Exploration)</td>
</tr>
<tr>
<td>10</td>
<td>440</td>
<td>380</td>
<td>-601.9300</td>
<td>0.4 (Emphasis on Exploration)</td>
</tr>
<tr>
<td>11</td>
<td>410</td>
<td>420</td>
<td>-822.8200</td>
<td>0.1 (Emphasis on Exploration)</td>
</tr>
<tr>
<td>12</td>
<td>420</td>
<td>420</td>
<td>-837.7300</td>
<td>0.5 (EI)</td>
</tr>
</tbody>
</table>

Table 8.1: The optimal weights of the kriging assisted by SMWEI with multi-weights at each iteration within the iterative process

8.4.1 Practical performance of SMWEI with the strategy of adaptively tuning weights

The performance of SMWEI with the strategy of multi-weights has been discussed in the previous section. Here, the SMWEI with the strategy of adaptively tuning weights is also tested under the same conditions (the same objective function and initial sampling points). Figure 8.6 shows the ‘history’ of how the sampling points are updated throughout the whole iterative process when applying this novel strategy. A slightly better performance is observed as only 11 iterations are taken to converge to the global minimum.
Figure 8.6: The performance of the kriging assisted by SMWEI with adaptively tuning weights to predict the two-variable Schwefel function (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)

### 8.5 A robust design

A design engineer is always expected to have an appreciation of how small perturbation in design variables will affect the device performance [53]. It may be the case, however, that even when an optimal design method is applied to a practical engineering problem, a theoretical optimum with excellent predicted performance in reality performs poorly when it is manufactured [54], because in real-world implementations the nominal values are often subject to uncertainties or tolerances existing in design variables [55]. The theoretical optimum may also be affected by uncontrollable external perturbations which can result in considerable deterioration of the target performance compared with the nominal solution [55]. In order to improve 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 providing the prediction results for robustness evaluation 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 8.7. 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 material properties are variable within prescribed limits. Thus the function of Figure 8.7 may be argued to be a possible simple representation of the robust optimization problem.

![Figure 8.7: The objective function for the robust test.](image_url)

**8.5.1 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 8.2.
These six initial sampling points have been used throughout all the robust design tests. The kriging method with EI requires 177 iterations to converge to the global minimum finally and the approximation surface after 177 iterations is shown in Figure 8.8. 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 prediction of it.

<table>
<thead>
<tr>
<th>$x_1$</th>
<th>$x_2$</th>
<th>Objective function value (y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>10</td>
<td>850</td>
</tr>
<tr>
<td>13</td>
<td>12</td>
<td>810</td>
</tr>
<tr>
<td>19</td>
<td>13</td>
<td>805</td>
</tr>
<tr>
<td>10</td>
<td>41</td>
<td>1000</td>
</tr>
<tr>
<td>41</td>
<td>10</td>
<td>1000</td>
</tr>
<tr>
<td>46</td>
<td>44</td>
<td>840</td>
</tr>
</tbody>
</table>

Table 8.2: Initial sampling points

Figure 8.8: (a) The performance of the kriging with EI when facing the robust problem (b) The error between the robust test function and prediction by kriging model with EI.
This graphical representation of the results should aid the understanding of the quality of the prediction. For example, in Figure 8.8 (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 cause the inaccuracy of the robustness evaluation around these solutions.

The SMWEI with adaptively tuning weights used the same initial setting to test. Compared with EI, the SMWEI method performs more efficient in finding the global minimum (only 85 iterations – see Figure 8.9); however, the error between the objective function and the predicted result is worse, which is a direct consequence of having fewer points available for shape representation.

In these works the Gaussian correlation model has been applied as it outperforms most of other models used with kriging when objective functions with nonlinear shape such as densely distributed valleys and hills are approximated [50]. In the robust test discussed here the test function is very simple: it has only two minima (one global optimum and one local minimum) and is flat around these two regions (Figure 8.7). Thus the Gaussian correlation model might not fix the shape of these linear regions precisely. 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 8.8b), 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 8.9(b)) whereas EI distributes the sampling point more evenly throughout the searching space therefore the error in these regions is smaller (Figure 8.8(b)).
Figure 8.9: (a) The performance of the kriging assisted by SMWEI with adaptively tuning weights when facing the robust problem, (b) The error between the test function and the kriging prediction assisted by SMWEI with adaptively tuning weights when facing the robust problem.

A similar observation can be made for the kriging model using the SMWEI with multi-weights (Figure 8.10). 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 8.10: (a) The performance of the kriging assisted by SMWEI with multi-weights when facing the robust problem, (b) The error between the robust test function and prediction by the kriging model assisted SMWEI with multi-weights.

Figure 8.11: The detailed situation of choosing optimal weights at each iteration
Figure 8.11 shows the history of the kriging assisted by SMWEI using the multi-weights as options during the pre-test process. The graph describes the variation of the optimal weights 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 mean square error more efficiently and specific to the problem being solved.

8.6 Conclusion

In this chapter, the preparation work for the robustness evaluation commences from improving the strategies which has been mentioned in Chapter 6 and Chapter 7, because 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. 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 specially devised robust test function and shown to perform better than the traditional utility function. Through analysing the error between the prediction provided by the kriging model and the real objective function value, the predicting accuracy of the kriging locating around global minimum and local minima has been convinced primarily. Both of these improved strategies will be implemented in practical design systems, especially for the purpose of electromagnetic robust design optimization. More details about testing the practical problem will be provided in the next chapter.
Chapter 9 Kriging with Adaptive Weighted Expected Improvement with Rewards Approach in Electromagnetic Design

9.1 Introduction

The previous chapter demonstrated the importance of finding the trade-off between exploration and exploitation while utilizing the kriging model to approximate the shape of objective function for the further evaluation of robustness. The Adaptive Weighted Expected Improvement with Rewards [48] (AWEI) [52] employed to distribute the weights on exploration and exploitation is investigated further, extended to real design problems. In this chapter, the AWEI also applies the scheme of building the ‘cheap’ simplified surrogate model for pre-test during the iterative process rather than directly obtain the information from objective functions. The Schwefel function with different dimensions and the TEAM Workshop problem 22 [56] are chosen to verify the feasibility of this novel algorithm.

9.2 Balancing Exploration and Exploitation

Two algorithms inspired by reinforcement learning [46], the Adaptive Weighted Expected Improvement (AWEI) and the Surrogate Model based Weighted Expected Improvement approach utilizing the concept of rewards [48] (SMWEI), were proposed in the previous chapter to make the process of tuning the weighting parameters more intelligent and self-guiding. During the pre-test stage AWEI applies a ‘cheap’ simplified surrogate model relying on a specific prediction and the potential error produced by kriging [50]. A pair of fixed weights (one emphasizing exploration and the other exploitation) are initially set at equal values. The Mean Square Error from the kriging model is used to calculate the rewards in the pre-test stage. After comparing the rewards obtained from the pre-test, the weights are redistributed on the terms favouring exploration and exploitation respectively. The weight with a better reward of the two is then used to feed back – via the FEM module – into the main iterative loop of the design process.
Although finding the global optimum is often sufficient, this may not be enough when a robust design is considered under certain uncertainties of variables, as information about the shape of the objective function around local minima in addition to the global minimum is also relevant. The quality of the approximation of the objective function is the most important prerequisite when the robustness of the solution is based on a kriging model rather than the real function. In the next section the performance of the AWEI method will be assessed in the context of finding the global optimum concurrently with obtaining a good quality function prediction for the search space of interest. A two-variable analytical test function will be used for illustrative purposes, which has been applied by many authors to discuss robust optimization.

9.3 Numerical Experiments

To explain the concepts we use a simple single-variable function shown graphically in Figure 9.1. There are two local minima (B and C) and one global minimum (A). When considering practical factors such as uncertainty of the variables due to manufacturing tolerances, non-uniform material properties or imperfect control of operating conditions, the theoretically best-performing point A may not be a practical optimal solution. The final judgment will be influenced by the margin of the uncertainty and by how much the objective function may change when the design variables assume limiting values. For a minor uncertainty, even under extreme (but still small) departure of the parameter from its nominal value, the objective function may still be acceptable – even if worse than at the theoretical ‘best’ – and thus even a ‘sharp’ optimum like A may be judged as practical. But under increased uncertainty the confidence in the final performance may be assured only by a much more ‘shallow’ minimum, such as B, or even C. Thus the preferred practical design may be selected away from the theoretical global optimum. The proposed methodology of how to evaluate robustness will be explained in the next two chapters.
Various methods of assessing the robustness have been proposed in literature. In [58], the initial optimization problem is transformed into a multi-objective optimization where both the objective function and the gradient index of this function are minimised simultaneously. We take a similar approach; however, we calculate the gradient index using the predicted value of the objective function rather than the function itself. The motivation behind such an approach is to reduce the computational effort by avoiding the costly FEM (or similar) numerical solution. The specific approach of calculating the sensitivity of solutions based on gradient index will be presented in the next Chapter, this chapter mainly places emphasis on the tests about the accuracy of the prediction obtained by kriging model with the adaptively weighted expected improvement. For this methodology to work it is very important to have good prediction for the objective function. The quality of the prediction will now be investigated with the aid of a test function defined as

\[
f(x) = 10 - \sum_{i=1}^{n} \left[ \frac{3.5}{1+(x_i-5)^2} + \frac{2.2}{1+(x_i-15)^2/10} + \frac{1.2}{1+(x_i-25)^2/30} \right], \quad (9.1)
\]

in the range \(5.6 \leq x_i \leq 27\). The two-variables (\(n = 2\)) version is plotted in Figure 9.2 and will be used to assess AWEI and compare its performance against standard EI method.
Chapter 9 Kriging with Adaptive Weighted Expected Improvement with Rewards Approach in Electromagnetic Design

Figure 9.2: A two-variable objective function

<table>
<thead>
<tr>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$y$</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>7</td>
<td>7.9966</td>
<td>20</td>
<td>7</td>
<td>7.6024</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>8.1913</td>
<td>7</td>
<td>13</td>
<td>7.0688</td>
</tr>
<tr>
<td>5.7</td>
<td>10</td>
<td>6.4292</td>
<td>25</td>
<td>20</td>
<td>7.2927</td>
</tr>
<tr>
<td>13</td>
<td>7</td>
<td>7.0688</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 9.1: Initial Sampling Points

(a) 20th iteration

107
Chapter 9 Kriging with Adaptive Weighted Expected Improvement with Rewards Approach in Electromagnetic Design

(b) 60th iteration

(c) 120th iteration

(d) 180th iteration
Figure 9.3: The iterative process of Kriging assisted by EI (prediction and error)
Chapter 9 Kriging with Adaptive Weighted Expected Improvement with Rewards Approach in Electromagnetic Design

(a) 20th iteration

(b) 60th iteration

(c) 120th iteration
For both the EI and AWEI tests, seven initial sampling points were used (as in Table 9.1). For better observation of variation throughout the whole iterative process, while applying EI and AWEI respectively, several prediction results and the corresponding errors between the prediction and the real objective function at certain iteration are presented as above. Kriging assisted EI needs 324 iterations to find the global minimum and the predicted shape is demonstrated in Figure 9.3, with the error calculated against
the actual objective function shown in Figure 9.4. The AWEI strategy is more efficient in finding the global minimum, requiring 211 iterations (Figure 9.4), but the final approximation is not as good as that resulting from EI as the number of available points is less. Nonetheless, in some regions, e.g. $5 \leq x_1 \leq 27$, $2.5 \leq x_2 \leq 7.5$ or $2.5 \leq x_1 \leq 7.5$, $5 \leq x_2 \leq 27$, AWEI outperforms EI (Figure 9.3).

9.4 TEAM Workshop Problem 22

The TEAM Workshop Problem 22 (superconducting magnetic energy storage system) in Figure 9.5 has been used to test the efficiency of the approximation model as well as the ability of coping with the multi-variable optimization problem which might result in the heavy burden of the correlation matrices with numerous data. The whole system including two concentric coils carrying current with opposite direction offer the opportunity to store a significant amount of energy under superconducting conditions while maintaining the stray field within certain limits [56]. The goal of the optimization task is to achieve the stored energy of $E_{\text{ref}}=180\text{MJ}$ with a minimal stray field $B_{\text{stray}}$ (evaluated along 22 equidistant points along line $a$ and line $b$ in Figure 9.5 as small as possible). The objective function is defined as

$$\text{OF} = \frac{B_{\text{stray}}^2}{B_{\text{norm}}^2} + \frac{|E-E_{\text{ref}}|}{E_{\text{ref}}}, \quad (9.2)$$

where $B_{\text{norm}}=3\mu\text{T}$ and $B_{\text{stray}}^2 = \frac{\sum_{i=1}^{22} |B_{\text{stray},i}|^2}{22}$, subject to some geometrical and ‘quench’ constraints. As constraints, the superconducting coils should not violate the quench condition which links together the value of the current density and the maximum value of magnetic flux density as follow:

$$g_i(x) = |I_i| + 6.4 \cdot |B_{m,i}| - 54.0 \leq 0, \quad i = 1,2 \quad (9.3)$$

The specific tests especially considering constraints will be demonstrated in Chapter 11. The TEAM 22 problem is classified into the three-parameter case and the eight-parameter case. The dimensionality of test and the setting of step size with respect to each parameter are directly relevant to the data produced by the correlation matrices in kriging model. Considering the difficulty of storing data while solving multi-variable
optimization problem, the tests specialized to the TEAM 22 problem commence from the tests with lower dimensionality to ones with higher dimensionality.

Figure 9.5: Configuration of TEAM Workshop Problem 22

9.5 Results for TEAM Workshop Problem 22

In previous chapters, the kriging surrogate model assisted by different methodologies for searching the global optimum of objective function has been tested by the one-variable and two-variable numerical functions in order to verify its feasibility. As one of the popular electromagnetic optimisation problem that has been tested by many authors, the TEAM Workshop Problem 22 is chosen to verify our surrogate models’ effectiveness when coping with practical cases. Here the 3 parameter TEAM Workshop problem 22 is tested, although initially – for the purpose of demonstrating typical shapes of the objective function – one of the variables is fixed \( R_2=3.08 \text{m} \), while \( d_2 \) and \( h_2 \) varied. One of the challenges in creating the kriging correlation matrix is the ‘combinatorial explosion’ when many parameters are used [59] and we propose to use a successive ‘zoom in’ strategy to cope with this issue; a two-variable version implemented in the context of the TEAM Problem 22 is tried first. The initial setting of test range and step size is described in the following Table:
<table>
<thead>
<tr>
<th>Test range</th>
<th>The first rough stage</th>
<th>The second stage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$h_2 [m]$</td>
<td>$d_2 [m]$</td>
</tr>
<tr>
<td>Min</td>
<td>0.408</td>
<td>0.1</td>
</tr>
<tr>
<td>Max</td>
<td>2.2</td>
<td>0.4</td>
</tr>
<tr>
<td>Step size</td>
<td>0.056</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Table 9.2: The ‘zoom in’ strategy using kriging assisted EI for TEAM 22 tests

The initial sampling points would normally be selected using the Latin Hypercube [60]; in our tests we fixed the positions at $(h_2=0.744m, d_2=0.13m)$, $(h_2=1.304, d_2=0.22m)$, $(h_2=1.64m, d_2=0.40m)$ and $(h_2=2.088m, d_2=0.37m)$ to facilitate comparisons. Kriging with EI required 25 iterations to complete the first ‘rough’ stage, and further 29 iterations to find the best-performing solution $(h_2=0.478m, d_2=0.394m, OF=0.0874)$ (Figure 9.6). For kriging assisted AWEI the relevant iteration numbers are 21 and 23 (Figure 9.7). The ‘history’ of the EI and AWEI strategies may be followed on the two figures.

Figure 9.6: The two-parameter SMES problem tested by kriging with EI.
9.5.2 The Three-Parameter TEAM Workshop Problem 22

The initial settings for the 3 parameter test are presented in Table 9.3. For kriging with EI, 79 sampling points were created in the first stage, with further 132 points in the zoomed-in region; in the AWEI case, 156 sampling points were followed by 167 points in the second stage (Figure 9.8). The total numbers quoted include the initial sampling points needed by the kriging model. It is no longer possible to conveniently display the objective function itself for three parameters.

<table>
<thead>
<tr>
<th>Test range</th>
<th>The first rough stage</th>
<th>The second stage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R_2 [m]$</td>
<td>$h_2 [m]$</td>
</tr>
<tr>
<td>Min</td>
<td>2.6</td>
<td>0.408</td>
</tr>
<tr>
<td>Max</td>
<td>3.4</td>
<td>2.2</td>
</tr>
<tr>
<td>Step size</td>
<td>0.1</td>
<td>0.056</td>
</tr>
</tbody>
</table>

Table 9.3: The Three-Parameter Kriging Assisted AWEI for TEAM 22 Tests
Chapter 9 Kriging with Adaptive Weighted Expected Improvement with Rewards Approach in Electromagnetic Design

(a) The 20th iteration (The first stage test)

(b) The 80th iteration (The first stage test)

(c) The 156th iteration (The first stage test)
Chapter 9 Kriging with Adaptive Weighted Expected Improvement with Rewards Approach in Electromagnetic Design

(d) The 175\textsuperscript{th} iteration (The second stage test)

(e) The 235\textsuperscript{th} iteration (The second stage test)

(d) The 323\textsuperscript{th} iteration (The second stage test)

Figure 9.8: Kriging with AWEI applied to three-parameter SMES problem
Chapter 9 Kriging with Adaptive Weighted Expected Improvement with Rewards Approach in Electromagnetic Design

(a) The 20th iteration (The first stage test)

(b) The 60th iteration (The first stage test)

(c) The 132th iteration (The first stage test)
In order to provide better clarity of results and present the history of finding the optimum throughout the whole iterative process, a series of graphs in Figure 9.8 and Figure 9.9 present the sampling process of the kriging model assisted by EI and AWEI methods respectively. It can be seen that in the graphs along the left hand side the blue cube indicates the test range and the black star points are initial sampling points (to facilitate comparison both tests apply the same initial sampling points). Due to utilizing the zoom-in strategy, the green start points represent the sampling points selected within the first stage of the test while the red points represent the sampling points in the second stage. The bar graphs mark the best solution found at specific iteration. To demonstrate the advantages of the kriging assisted EI and AWEI approaches over other well-known stochastic optimization methods, when solving the three-parameter TEAM 22 problem, Table 9.4 has been compiled using available published data [67]. It is encouraging from this comparison that kriging assisted EI and AWEI both offer significant advantages in
terms of much reduced number of computationally expensive function calls to achieve required level of accuracy. However, those gains are somewhat offset by the need to create the kriging correlation matrix. Finally, the robustness of the solution requires further consideration and will be studied in the following sections.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>$R_2$(m)</th>
<th>$d_2$(m)</th>
<th>$h_2/2$(m)</th>
<th>$OF$</th>
<th>No. of Iters</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA</td>
<td>3.040</td>
<td>0.389</td>
<td>0.240</td>
<td>0.134</td>
<td>2400</td>
</tr>
<tr>
<td>HuTS</td>
<td>3.080</td>
<td>0.380</td>
<td>0.246</td>
<td>0.089</td>
<td>3821</td>
</tr>
<tr>
<td>ITS</td>
<td>3.100</td>
<td>0.388</td>
<td>0.240</td>
<td>0.098</td>
<td>1824</td>
</tr>
<tr>
<td>SA</td>
<td>3.078</td>
<td>0.390</td>
<td>0.237</td>
<td>0.098</td>
<td>5025</td>
</tr>
<tr>
<td>NTS</td>
<td>3.080</td>
<td>0.370</td>
<td>0.254</td>
<td>0.089</td>
<td>1800</td>
</tr>
<tr>
<td>PBIL</td>
<td>3.110</td>
<td>0.421</td>
<td>0.241</td>
<td>0.101</td>
<td>3278</td>
</tr>
<tr>
<td>El(Kriging)</td>
<td>3.090</td>
<td>0.394</td>
<td>0.236</td>
<td>0.0875</td>
<td>211</td>
</tr>
<tr>
<td>AWEI(Kriging)</td>
<td>3.090</td>
<td>0.400</td>
<td>0.232</td>
<td>0.0875</td>
<td>323</td>
</tr>
</tbody>
</table>

Table 9.4: Genetic Algorithm (GA) [61]; Tabu Search (HuTS) [62]; Improved Tabu Search (ITS) [63]; Simulated Annealing Algorithm (SA) [64]; New Tabu Search (NTS) [65]; Population-based Incremental Learning (PBIL) [66].

9.6 Conclusion

The balance between exploration and exploitation in the design algorithm is controlled via the kriging model assisted by the Adaptive Weighted Expected Improvement method utilizing the ‘rewards’ produced by the kriging model. But when solving tasks with the large number of points over the design space, as iterative process progresses the increasing number of updated sampling points may cause the size of correlation models increase dramatically. Thus the storage of correlation matrices may become an issue, especially when handling multi-parameter design problems. To avoid problems associated with combinatorial explosion, a ‘zoom in’ strategy is applied, which can maintain the size of data produced by the correlation model within an affordable range via redefining the test range and the test step size. But the optimal setting of step size and test range is often problem dependent and even inappropriate setting would cause the region containing important information to be ignored. Thus more advanced methods with the capability of mitigating the burden of data storage is necessary.
Chapter 10 Robust Global Optimization of Electromagnetic Designs Utilizing Gradient Indices and Kriging

10.1 Introduction

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.

This chapter mainly describes the process of formulating the methodology for sensitivity analysis while considering uncertainty existing in design variables. First, the concept of the Gradient Index (GI) has been explored in [57], [68] and [69]. The method transforms a problem into a multi-objective optimisation by concurrently minimising the function and its gradient index, 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. Here we attempt to provide significant further advances both in terms of the way in which the gradient indices are used but also by improving the computational efficiency of the algorithms. The notions of the gradient index sensitivity and the second-order gradient index 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 [5] is employed. In other words, the objective function is approximated using kriging [50], assisted by algorithms balancing exploration and exploitation ([52], [70]) using the concept of rewards [48]. 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.

10.2 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 Figure 10.1 and 10.2. 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.

10.2.1 Multi-Objective Robust Optimisation using GI

Consider a commonly used one-variable test function [57], [65] and [69] (see Figure 10.1)

\[ f(x) = 3 - \frac{3.5}{1+(x-5)^2} - \frac{2.2}{1+(x-15)^2/10} - \frac{1.2}{1+(x-25)^2/30} \]  \hspace{1cm} (10.1)
Chapter 10 Robust Global Optimization of Electromagnetic Designs Utilizing Gradient Indices and Kriging

(a) 

Largest $GI$ of A1
Sensitivity = \[ \frac{\text{Largest } GI - \text{Smallest } GI}{\text{uncertainty}} = \lambda = 2 \]

Objective Function Values ($OF$)

First gradient index ($GI$) and Sensitivity

(b) 

Objective Function Values ($OF$)

First gradient ($GI$) and Second gradient ($GI'$)
Figure 10.1: Example of a robust design for a one-variable problem
(a) Objective function, the gradient index and sensitivity, (b) First and second order gradients, (c) Objective function trajectory showing sensitivity and objective function values (OF).

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

\[ U(x) = \{ \xi \in \mathbb{R}^n | x - k \sigma \leq \xi \leq x + k \sigma \} \]

(10.2)

where σ is standard deviation of uncertain variables and k is determined by a confidence level [65].

One way of incorporating robustness into the mainstream optimisation process is by adding the gradient index [57] as a second objective and formulating the problem as

\[
\text{Minimise} \quad f(x) \quad x \in \mathbb{R}^n \quad (x_L \leq x \leq x_U) \\
\text{Minimise} \quad GI(x) = \max_{1 \leq i \leq n} \left| \frac{\partial f(x)}{\partial x_i} \right| \\
\text{Subject to} \quad g_i(x) \leq 0, i = 1, ..., m
\]

(10.3)
Point A1 in Figure 10.1(a) 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 Figure 10.1(a) and the values of the second-order gradient in Figure 10.1(b), 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 Figure 10.1(a) 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 Figure 10.1(c). 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 gradient index.

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 Figures 10.1(a)-(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 (Figure 10.1b).

Another example (Figure 10.2) 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 Figure 10.2(b) – allows proper judgement to be made regarding the robustness of this particular design. Table 10.1 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 Figure 10.2(b).
Table 10.1: Values of the objective function, second-gradient and sensitivity for the three design points A1, A2 and A3 (Figure 10.2).
Figure 10.2: 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, the second-order gradient index (GI’) and objective function values (OF).

10.3 Robust optimisation based on kriging modelling

10.3.1 Kriging

As a type of regression model, kriging [50] 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), in order to build a correct surrogate model of
the real objective function through interpolation. This relies on the linear regression model

$$\hat{y}(x) = \sum_{k=1}^{m} \beta_k f_k(x) + \epsilon(x)$$  \hspace{1cm} (10.4)

and the Gaussian correlation model

$$R(\epsilon(x^i), \epsilon(x^j)) = \prod_{k=1}^{n} e^{-\theta_k |x_k^i - x_k^j|^p_k}$$  \hspace{1cm} (10.5)

where the global function $\sum_{k=1}^{m} \beta_k f_k(x)$ and an additive Gaussian noise $\epsilon(x)$ are integrated to the predicted value $\hat{y}(x)$ of the objective function; the hyperparameter $\theta_k$ is the correlation amongst 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 optimizes the value of $\theta$ and then the correlation model is brought into the regression model to evaluate the function with the best linear unbiased predictor [50], [71].

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 [72] we proposed a successive ‘zoom in’ strategy to alleviate the problem, where – in order to reduce the amount of data storage and utilize the installed physical memory capacity efficiently – the step sizes of design vectors 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 methodology divides the correlation matrices in an adaptive manner so that the physical memory available is used in an optimum way.
10.3.2 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 Figure 10.3; 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 in order to make full use of the available memory while minimising the number of calls of the correlation model. Two schemes have been implemented and operate throughout the process as explained in Figure 10.3, either partitioning in terms of the sampling points or via the design vectors – the choice is governed by considerations of optimal memory management and will depend on the physical size of the memory available. 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 from growing 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 Figure 10.9, however, does show some quantitative details for a specific case and particular computer implementation.

Figure 10.3: The correlation matrices partitioned by: (a) sampling points, (b) design vectors

Figure 10.4: The flowchart of the adaptive division of correlation matrices.
To verify the viability of the proposed methodology, the TEAM problem 22 has been attempted [56]. Without the strategy of Figure 10.4, the kriging algorithm faces a new problem that the accumulating data in correlation matrices along with the increase of sampling points is easy to cause our personal computers’ physical memory to explode especially when copying with the large-scale multi-variable task.

### 10.4 Application in electromagnetic design

#### 10.4.1 The Test Results of two-parameter TEAM 22 problem

In last chapter, the addressing the three parameter problem, for the purpose of demonstrating typical shapes of the objective function, one of the variables has been fixed ($D_2=0.394m$), while $R_2$ and $h_2$ are varied. Usually the initial sampling points are selected using Latin Hypercube [60]; however, for this test eight initial sampling points were used as shown in Table 10.2. Table 10.3 compares one typical result from literature with our AWEI algorithm (kriging with Adaptive Weighted Expected Improvement) [52], [70], while Figure 10.5 demonstrates the convergence process of AWEI.

<table>
<thead>
<tr>
<th>$R_2$</th>
<th>2.7</th>
<th>2.9</th>
<th>3.0</th>
<th>3.3</th>
<th>2.6</th>
<th>3.4</th>
<th>2.6</th>
<th>3.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_2$</td>
<td>0.744</td>
<td>1.304</td>
<td>1.64</td>
<td>2.088</td>
<td>0.408</td>
<td>0.408</td>
<td>2.2</td>
<td>2.2</td>
</tr>
</tbody>
</table>

Table 10.2: The setting of initial sampling points.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>$R_2$(m)</th>
<th>$h_2/2$(m)</th>
<th>Best OF</th>
<th>Iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>RBF</td>
<td>3.06</td>
<td>0.236</td>
<td>0.088</td>
<td>240</td>
</tr>
<tr>
<td>AWEI (Kriging)</td>
<td>3.08</td>
<td>0.239</td>
<td>0.089</td>
<td>38</td>
</tr>
</tbody>
</table>

Table 10.3: Performance comparison of algorithms.
Radial basis functions (RBF) [11]; AWEI (Kriging) [70].
Chapter 10: Robust Global Optimization of Electromagnetic Designs Utilizing Gradient Indices and Kriging

Figure 10.5: TEAM Workshop Problem 22

(a) Prediction by kriging with AWEI
(b) Sensitivity with respect to $R_2$ and $H_2$,
(c) Objective function trajectory showing sensitivity and objective function values (OF).

$3 \leq R_2 \leq 3.3$, $0.408 \leq h_2 \leq 0.5$, $D_2=0.394$, other parameters fixed

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

10.4.2 Three-parameter Test Results

The full 3 parameter TEAM 22 problem [56] is potentially a challenge to the kriging method because of the ‘combinatorial explosion’ associated with setting up the correlation model, as explained in Section 10.3 – thus the savings due to 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 (with small physical memory space) or a laptop is used for simulation. The previously reported ‘zoom-in’ strategy [72] 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 10.3. The initial sampling points are set as in Table 10.4, more information provided in Table 10.5.

Table 10.4: The setting of initial sampling points.

<table>
<thead>
<tr>
<th></th>
<th>( R_2 ) (m)</th>
<th>( h_2 ) (m)</th>
<th>( d_2 ) (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1</td>
<td>3.1</td>
<td>0.576</td>
<td>0.32</td>
</tr>
<tr>
<td>Sample 2</td>
<td>3</td>
<td>0.408</td>
<td>0.3</td>
</tr>
<tr>
<td>Sample 3</td>
<td>3.2</td>
<td>0.744</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Table 10.5: The specific definition of test.

<table>
<thead>
<tr>
<th></th>
<th>( R_2 ) (m)</th>
<th>( h_2 ) (m)</th>
<th>( d_2 ) (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test range</td>
<td>[2.6 3.4]</td>
<td>[0.408 2.2]</td>
<td>[0.1 0.4]</td>
</tr>
<tr>
<td>Step size</td>
<td>0.01</td>
<td>0.014</td>
<td>0.003</td>
</tr>
<tr>
<td>Number of steps</td>
<td>81</td>
<td>129</td>
<td>101</td>
</tr>
</tbody>
</table>

A comparison between the results obtained with the ‘zoom in’ strategy and results using ‘the adaptive correlation matrices division’ for this example are presented in Figure 10.6. With the ‘zoom in’ strategy, the kriging assisted by 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 specific range of each test stage. Via applying the scheme of partitioning the correlation matrices, the issue of combinatorial explosion is overcome throughout the process of infilling the sampling points. 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 (Figure 10.6b). But the main benefit of the modified correlation model is that the
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’.

Figure 10.6: The 3 parameter TEAM 22 problem (a) EI, (b) AWEI.
Figure 10.7: Objective function trajectory showing sensitivity and objective function values (OF)
(a) EI, (b) AWEL.
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 Figures 10.7a and 10.7b, respectively. Both best solutions, in terms of the objective function value and the sensitivity, are shown in Figure 10.7. 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.

Figure 10.8: The specific parameter setting and transient process of testing the TEAM 22 problem with the help of the novel partitioning scheme.
In order 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 Figure 10.9. 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 Figure 10.8. The decision if the correlation matrix needs to
be partitioned follows from two considerations: limiting memory occupied by the optimiser or limiting the time spent on 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. In the particular case, the testing environment provided approximately 11 GB available memory, which was set as the triggering value. However, due to 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 Figure 10.9 – at some point a marked increase could be observed when the 11GB memory limit had not 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.

10.5. Conclusion

Finding the global optimum of an objective function may not be sufficient; the robustness of the design is also an important consideration in practical electromagnetic optimization problems. As the uncertainties of design variables are given, the difference between the highest and the lowest gradient within the corresponding uncertain range of each solution can be used to evaluate the sensitivity of the solution during the process of searching robust optimum. Because based on very limited information from objective function the improved kriging surrogate model can approximate the shape of the objective function relatively accurately, the prediction produced by the kriging model is used to provide the information for further robustness evaluation. But when coping with the large-scale multi-variable task, the classic kriging model can be inefficient due to the dramatically increasing data in correlation matrices. The strategy of splitting correlation matrices can mitigate the burden of data storage to some extent. Instead of storing the full-version correlation matrices, the split sub-matrix with affordable data size is used as substitute. It makes that the kriging model can take the challenge of dealing with larger scale multi-variable electromagnetic design problems via adjusting the size of sub-matrix.
Chapter 11 Further Considerations of Uncertainty in Robust Optimization of Electromagnetic Devices

11.1 Introduction

In this chapter, the worst-case optimisation (WCO) and the worst-vertex-based WCO are proposed to evaluate the robustness of both performance and constraints under uncertainty. The worst-case optimization (WCO) method has been selected to evaluate the accuracy of the prediction of the objective function provided by an improved kriging model, implemented for the sake of reducing computational effort associated with direct application of information produced by a time-consuming Finite Element Method (FEM). However, in applying this approach, some shortcomings of the WCO approach have been identified associated with the inability to assess the performance variation under conditions of uncertainty. Therefore the concept of average performance evaluation has been suggested as an improved measure of robustness.

11.2 Robust Optimization Algorithms exploiting Kriging Modelling

When designing real devices, many design variables have to be subject to specific uncertainties. Hence the assessment of the influence caused by these uncertainties on the performance becomes essential in practical problems. Algorithms applying different strategies to evaluate robustness, such as the sensitivity analysis [74], the worst-case optimization method [20], [55], [75]-[77], and the mean value and variance of performance [21], [78]-[79], have been developed to assist in the design tasks. In [80] a multi-objective optimization method, which included sensitivity analysis using gradient index, was developed and demonstrated.

A widely used approach to evaluate the reliability of a robust solution is the worst-case method. This technique can maintain a certain level of robustness by avoiding solutions that may push the function into unfeasible region when searching for the optimum, as shown in Figure 11.1(a). A set of typical examples is shown in Figure 11.1(a) to
illustrate the principle of the worst-case optimization technique. The theoretical optimum A, for example, may be abandoned in favour of a ‘worse’ solution B because the actual design A’, which accounts for the uncertainties of the variables, might violate constraints and enter unfeasible area. The uncertainties have been defined mathematically in equation (10.2) [68]. The uncertainties may also be specified directly (e.g. as machining tolerances, say Δ). As an algorithm which can predict the worst scenario considering the uncertainties, as well as constraints with respect to specific designs, the worst case optimization (WCO) method [55], [76] and [77] may be applied to analyse the reliability of the solution as follows

\[
\text{Minimize } f_w(x) \equiv \max_{\xi \in \Omega(x)} f(\xi)
\]

Subject to \( g_{w,i}(x) = \max_{\xi \in \Omega(x)} g(x) \leq 0, \ i = 1, \cdots, m \) \hspace{1cm} (11.1)

The worst values of the objective function and the \( i \)-th constraint function are chosen to substitute the original values of the nominal design \( x \).
Numerical methods, such as finite elements, are often used when searching for the worst objective function value under imposed constraints which may be an extremely time-consuming process. Compare with the worst case optimization, the worst-vertex-based WCO (W-WCO) \[55\] require fewer FEM callings; this algorithm only needs to observe the vertices within the region restricted by uncertainties rather than evaluating every design value. For example in the problem illustrated in Figure 11.1(b), in addition to \(x\) there are 8 more points evaluation (4 points located at the vertices and the other 4 points located at the middle of the four edges of the square) required, located at the corners and the middle of the specified boundary. However, in certain cases assessing only these 8 points might still not be sufficient. Figure 11.1(c) illustrates such a case were a large variation of the function will not be identified by the W-WCO method. But the WCO
requires evaluating numerous sampling points’ objective function value in the uncertain range of each solution, which undoubtedly is a heavy computational burden.

One of the possibilities is to employ the improved kriging method, assisted by a set of strategies capable of balancing exploration and exploitation [52], [72] using the concept of rewards [48], which can be used to predict the objective function value instead of directly calculating it using computationally expensive FEM models. Based on such a kriging prediction, the worst-case method can be directly implemented. In other words, the WCO method uses the predicted information rather than the expensive FEM models. The accuracy of the predicted objective function using the improved kriging model has been considered in [80]. Hence here the suitability of directly using WCO with the predicted function model is discussed and demonstrated.

11.3 The Prediction of Kriging surrogate model

Both the worst case and the worst-vertex-based WCO methods for robustness evaluation, introduced in the last section, would involve unreasonable calculation cost if based entirely on simulation results from FEM models. Especially the worst case approach needs all information in the uncertain region restricted by the given uncertainties with respect to each solution in the design space. On the other hand, the cost of the worst-vertex-based WCO is comparatively cheaper, with only the 9 vertices and the solution itself which is under test necessary instead of the full information with respect to each solution’s uncertain region. Thus to solve the difficulty of the excessive calculation cost, the prediction provided by the kriging surrogate model is employed.

But when solving large multi-parameter optimization problems, kriging has some inherent limitations making the implementation difficult due to the data space occupied by correlation matrices. In particular, for multi-parameter problems, the correlation model built by the kriging algorithm can grow very fast resulting in a ‘combinatorial explosion’ of correlation data filling very quickly the memory of standard computer workstations. As a result the process can become slow and inefficient. To solve this issue a scheme that adaptively partitions the correlation matrices has been developed. Using this approach the size of the data is managed to use efficiently the available memory throughout the iterative process of kriging. The scheme mentioned above has
another advantage as the kriging predictor and Mean Squared Error are being calculated at the same time hence more computational time is saved in the process of building the kriging model. Therefore this modification makes kriking suitable for solving multi-parameter optimization problems and could be linked with WCO which needs detailed data to work effectively.

### 11.4 Average performance

As explained in the previous sections, the WCO method can be used to find robust solutions for a particular problem once the constraints and the uncertainties of the variables have been defined. In this section, however, we address the extreme case depicted in Figure 11.2; it is argued that WCO on its own is not sufficient to find a reliable and robust solution. To deal with such a situation the concept of an average performance has been suggested and will now be explained. Figures 11.2(a) and 11.2(b) show two similar functions, otherwise identical, except the region around the points A2 and A4, respectively. If WCO is used to find a robust solution using the same uncertainties $\Delta$ for the case depicted in Figure 11.2(a), point A2 will be found, whereas for the case shown in Figure 11.2(b) point A4 is likely to be returned as a robust solution. Apparently it can be seen that in Figure 11.2 with the same uncertainty of variable $x$ the worst case values of points A2, A4 and A6 are identical, but the variation of the objective function in these region is different. Presumably if the uncertainty of variable satisfies the normal distribution as presented in formula (10.2), the point A2 is more robust than the other two. Thus only comparing the worst case a solution might not be enough to determine that a solution is robust or non-robust. To resolve such problems the concept of average performance within the uncertainty range is introduced.
Chapter 11 Further Considerations of Uncertainty in Robust Optimization of Electromagnetic Devices

(a)

(b)
Figure 11.2: (a) The objective function including a smoothly shaped local minimum, (b) the objective function including a local minimum with ripples, (c) the objective function including a non-robust local minimum with nearby 'peak'.

(subject to uncertainty $\Delta$)

The initial idea of evaluating average performance was to simply calculate the average value of all the potential perturbed values in the uncertain region with respect to the solution. This, however, proved infeasible, as although the shape of the two objective functions is clearly different, they share some common characteristics as explained in Table 11.1:

<table>
<thead>
<tr>
<th>Global minimum (A1, A3)</th>
<th>Local minimum (A2, A4)</th>
<th>Uncertainty $\Delta$</th>
<th>Worst case (A1, A3)</th>
<th>Worst case (A2, A4)</th>
<th>Average value (A2, A4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x=18$, $y=1.0216$</td>
<td>$x=55$, $y=2.12657$</td>
<td>15</td>
<td>$x=3$, $y=6.557$</td>
<td>$x=55$, $y=2.12657$</td>
<td>2.9556</td>
</tr>
</tbody>
</table>

Table 11.1: The common features of the two functions
By visual inspection we can see that the variation around the solution A4 is much more intense than around A2, but the average values within the shaded areas with middle-points A2 and A4 for the given uncertainty are the same and equal to 2.9556. This means that the average value criterion may not be useful in assessing the variation of the function close to the point of interest. As another particulate case, likewise the Figure 11.2(c) shows the objective function which is similar to the above two functions except the uncertain region around A6. Obviously the worst case value of A6 is worse than the two forward cases; however the average in the uncertain scope of A6 is exactly same as the two ones. It proves that the average value criterion is unable to inform us the robustness of solution. Hence an average value of the gradient index (GI) [58] has been introduced as an alternative way of assessing the average performance. The average value of the gradient index is calculated as:

$$\text{Minimize } \frac{\sum_{x_i=x_L}^{x_U} \max \left( \frac{\partial f(x)}{\partial x_i} \right) }{n}$$

where $x_i \in \mathbb{R}^2$ ($x_L \leq x_i \leq x_U$) is the $i$-dimensional design variable vector with lower and upper bounds $x_L$ and $x_U$, respectively, and $n$ is the sum of the design vectors. The sum of the maximum gradients is divided by the total number of design vectors as shown in the Formula (11.3). The average value of Gradient Index for the first case is 0.0594; while in the second case it is 0.1416 and three times larger than the first one. Therefore the average gradient index could be used as a more reliable criterion to evaluate the average performance. This criterion can therefore be combined with the WCO method to resolve difficult problems such as the one described by Figure 11.2(b). By generalizing this methodology it can be argued that a robust optimization problem can be transformed into a three-objective optimization problem defined as

$$\text{Minimize } f(x)$$

$$\text{Minimize } f_w(x) \equiv \max_{\xi \in U(x)} f(\xi)$$

$$\text{Subject to } g_{w,i}(x) = \max_{\xi \in U(x)} g(\xi) \leq 0, \; i = 1, \ldots, m$$

$$\text{Minimize } \frac{\sum_{x_i=x_L}^{x_U} \max \left( \frac{\partial f(x)}{\partial x_i} \right) }{n}$$
11.5 Robust Optimization Algorithms exploiting Kriging Modelling

In [80] the robustness of an optimal solution was evaluated using the gradient index, where the task of robust optimization was transformed to a two-objective optimization. One objective was to minimize the difference between the absolute value of the largest and the smallest gradients within the uncertain range, called the ‘sensitivity’, while minimizing the objective function that was the second objective. Although the sensitivity calculated by the gradient index method is able to provide information on the rate of change of the objective function, the WCO method can also be employed to obtain similar information. However, as shown in the previous section, the WCO method has some limitations; especially for extreme cases (Figure 11.2(b)). The average performance assessment described above can thus be added to the WCO method to improve the overall reliability of the result. To verify the concept and to analyse further the average performance criterion, two problems have been tested and the results are reported below.
Figure 11.3: (a) Analytic function (b) The kriging prediction (c) The performance of the WCO method (d) Average performance (e) The full-field solution for three objectives (f) The zoomed-in optimal part.

First, the two-variable analytic function (8) depicted in Figure 11.3(a), which was also used to assess the accuracy of the improved kriging model, has been tested.

\[
f(x) = 10 - \sum_{i=1}^{n} \left[ \frac{3.5}{1 + (x_i - 5)^2} + \frac{2.2}{1 + (x_i - 15)^2 / 10} + \frac{1.2}{1 + (x_i - 25)^2 / 30} \right] \quad (0 < x_i < 27)
\]  

(11.8)
Kriging with adaptive weighted expected improvement (AWEI) [72] has provided an approximation of this analytic function with a test step size of 0.1. Within 85 iterations, the kriging model can find the global minimum successfully; however – for a better approximation of the shape of the objective function – the model continued to run until the 285th iteration (Figure 11.3(b)).

The uncertainty with respect to variables is set as \( U(x_1, x_2) = 0.5 \), and the WCO method is used to obtain the surface of the worst case for each solution as shown in Figure 11.3(c). For this case the ‘best’ solution shifts from the theoretical optimum \((x_1=5, x_2=3, y=2.443)\) to the location \((x_1=5.1, x_2=5.1, y(\text{worst case})=3.7639)\) which provides a more robust result, for the given conditions. If the values of the uncertainties were to keep increasing, up to a certain extent, ultimately the robust optimum would thoroughly shift from the sharp global minimum to one of the preferable local minima with higher robustness. Figure 11.4(d) depicts the average gradient index values in the search space. Finally, the full-scale \((0 < x_i < 27)\) optimal solutions including all three objectives have been presented in Figure 11.3(e). For clearer presentation of the pareto front, the full-scale version is zoomed in Figure 11.3(f). Two typical pareto solutions are labelled in the zoomed-in graph: solution N1 delivers a more optimal value of the prediction of the objective value, while N2 offers a relatively better average performance.

### 11.6 Application in Electromagnetic Design

In addition to the numerical test above, the proposed WCO procedure was tested using a practical electromagnetic optimisation problem, namely, a multi-objective version of the TEAM 22 benchmark problem [56]. The specific description of TEAM 22 benchmark problem has been presented in the Chapter 9. The approach taken here combines WCO method with kriging and commercially available FEM based software. A 2D model of the TEAM 22 problems is solved throughout this procedure. The three parameter case of TEAM 22, which includes three geometric variables \( R_2, H_2 \) and \( D_2 \), while \( R_1, H_1 \) and \( D_1 \) are fixed, has been tried under different settings of uncertainties. The uncertainties are assumed to exist in the current densities \( J(J_1, J_2) \) of the two coils, because normally they are limited within certain range by a current controller for compensating perturbation. A constraint is imposed that the superconducting coils should not violate
the quench condition which links together the value of the current density and the maximum value of magnetic flux density as follows

\[ g_i(x) = |J_i| + 6.4 \cdot |B_{m,i}| - 54.0 \leq 0, \quad i = 1, 2 \quad (11.9) \]

Three different tests were performed for this 3-parameter TEAM 22 problem. The initial data set-up of step size, the number of steps and the test range, for the three tests are listed in Table 11.2 respectively. One of the differences between the three sets is the size of the uncertainty existing in the current densities \( J (J_1, J_2); \) the first set had the uncertainty set as 0.1, the second test 0.2, while in the third test 0.35 was used. Another difference was in the number of iterations that were used to approximate the objective function by the kriging model with AWEI [80]. For the first data set (uncertainty 0.1) kriging generated 185 sampling points, for the second case (uncertainty 0.2) kriging produced 257 sampling points, while for the third set (uncertainty 0.35) only 188 points were necessary. The results returned by the WCO, coupled with kriging, for these three cases are summarised in Tables 11.3 to 11.5. Figures 11.4, 11.5 and 11.6 show both the full scale and zoomed-in versions of the pareto fronts obtained for the three cases. In the figures and tables, solutions P1, P4 and P7 refer to the global optimum; P3 and P6 have the largest value of ‘worst case’ performance; while P2, P5 and P8 describe the solutions with the best value of average gradient index.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Uncertainties</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R_2 (m) )</td>
<td>( h_2 /2(m) )</td>
</tr>
<tr>
<td>Lower bound</td>
<td>3.03</td>
</tr>
<tr>
<td></td>
<td>3.13</td>
</tr>
<tr>
<td>Upper bound</td>
<td>0.01</td>
</tr>
<tr>
<td>No of steps</td>
<td>11</td>
</tr>
</tbody>
</table>

Table 11.2: The initial setup of the prediction by the kriging model
Chapter 11 Further Considerations of Uncertainty in Robust Optimization of Electromagnetic Devices

<table>
<thead>
<tr>
<th></th>
<th>$R_2$ (m)</th>
<th>$h_2/2$ (m)</th>
<th>$d_2$ (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>3.04</td>
<td>0.492</td>
<td>0.397</td>
</tr>
<tr>
<td>P2</td>
<td>3.09</td>
<td>0.464</td>
<td>0.382</td>
</tr>
<tr>
<td>P3</td>
<td>3.03</td>
<td>0.562</td>
<td>0.397</td>
</tr>
</tbody>
</table>

Table 11.3: Results for the case when uncertainty is 0.1

![Pareto Solution of 3-Objective Optimization](image)

Figure 11.4: The three objective optimization including (Worst case(WC), Average gradient index performance(AVGI), the prediction of objective functions(Ob)) ($U(J_1, J_2)=0.1$)

<table>
<thead>
<tr>
<th></th>
<th>$R_2$ (m)</th>
<th>$h_2/2$ (m)</th>
<th>$d_2$ (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P4</td>
<td>3.11</td>
<td>0.492</td>
<td>0.376</td>
</tr>
<tr>
<td>P5</td>
<td>3.09</td>
<td>0.478</td>
<td>0.385</td>
</tr>
<tr>
<td>P6</td>
<td>3.03</td>
<td>0.562</td>
<td>0.397</td>
</tr>
</tbody>
</table>

Table 11.4: Results for the case when uncertainty is 0.2
Figure 11.5: The three objective optimization ($U(J_1, J_2)=0.2$)

<table>
<thead>
<tr>
<th></th>
<th>$R_2$ (m)</th>
<th>$h_2/2$ (m)</th>
<th>$d_2$ (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P7</td>
<td>3.07</td>
<td>0.492</td>
<td>0.382</td>
</tr>
<tr>
<td>P8</td>
<td>3.03</td>
<td>0.562</td>
<td>0.397</td>
</tr>
</tbody>
</table>

Table 11.5: Results for the case when uncertainty is 0.35

Figure 11.6: The three objective optimization ($U(J_1, J_2)=0.35$)
Unlike other reported work [52] that uses a stochastic optimization method to find the global optimum and then employs Monte Carlo method to explore the space around the global minimum, combined with WCO and the gradient index to judge the robustness of the solution, the method introduced here takes a holistic approach and explores the whole searching space. The kriging model allows comparison amongst several local minima (maxima) that may be more robust than the global optimum. Another major advantage of the procedure proposed in this work is the fact that it can be linked with any commercial electromagnetic design software giving more freedom to the designer.

11.7 Conclusion

As the methodologies that can evaluate the reliability of solution, the worst-case optimisation (WCO) and the worst-vertex-based WCO are discussed in this chapter. Compared with the worst-vertex-based WCO, the WCO selected more sampling points within uncertain range of each solution, which consequently provides a more precise evaluation of a solution’s reliability. This is, however, more time-consuming. Fortunately the kriging surrogate model is capable of providing prediction of objective function with relatively high accuracy based on very little information from the computationally expensive finite element modelling. The kriging model is utilized to provide information for the further reliability assessment by the WCO method. Due to the consideration of average performance of solutions within their uncertain range respectively, the WCO is enhanced by cooperating with the concept of average gradient index performance. Using this approach a conventional optimization problem, with constrains and uncertainties in variables, has been transformed into three-objective optimization with a relevant pareto front. The proposed robust optimisation algorithms assisted by the improved kriging model have been verified by both numerical tests and a practical electromagnetic design problem described by TEAM 22 benchmark problem.
Chapter 12 Conclusions & Recommendations for Further Work

12.1 Conclusions

Kriging, an efficient surrogate modelling technique popular in other fields and applications but not previously used much in the context of electromagnetic design optimisation, was selected as the main topic of this thesis. The main motivation was that this approach can predict a response surface of the objective function on the basis of limited information and at the same time estimates the accuracy of this prediction. Thus time consuming function evaluations employing computationally expensive finite element or similar numerical modelling can be replaced by kriging predictions in order to select the most promising point for further evaluation. Such decisions require a careful balance between exploration and exploitation. The focus of this project was to develop and investigate alternative approaches to this decision making process.

The tasks were phrased as research aims and objectives specified in Section 1.2; the remaining Chapters have demonstrated that all these objectives have been accomplished. Several novel strategies based on reinforcement learning theory have been put forward, implemented and thoroughly investigated. Tests have been conducted using popular benchmark functions as well as practical engineering problems taken from TEAM workshop; comparisons were made with the results of other methods published in literature. Overall, it was found that kriging offers a very efficient way of assisting the design process, in particular when the main modelling technique carries heavy computational burden. In most cases kriging was found to significantly outperform all other methods.

However, it was found that large-scale tasks, multi-objective and dealing with many variables, may lead to ‘combinatorial explosion’ when necessary correlation matrices are established between sample points and all design vectors. In order to alleviate such problems two techniques have been developed: successive ‘zoom in’ and partitioning of the correlation matrices; both were found to be helpful in reducing the data storage requirements.
Finally, the design optimisation process was considered from the point of view of robustness of the final solution in response to changes in parameters (e.g. due to tolerances). The use of gradient indices and the worst case methodology have been explored and combined with the kriging approach.

The main contributions of the thesis can be identified as follows.

- The importance of balancing exploration and exploitation to achieve convergence to the global optimum effectively has been confirmed via a series of tests.

- Techniques from reinforcement learning were employed to introduce tuning weights to balance exploration and exploitation automatically in response to the feedback produced by a kriging surrogate model. A novel method named ‘Adaptive Weighted Expected Improvement with Rewards’ was shown to be able to learn from the experience of trying the exploration and exploitation separately and then determine the distribution of weights accordingly.

- A pre-test utilizing only a combination of predicted results and the mean square error, which is computationally cheap, has been developed and shown to be helpful also for long-term decisions. Another novel method called ‘Surrogate Model based Weighted Expected Improvement approach with rewards’, applying reinforcement learning based the improved pre-test strategy, has been proposed; it attempts to capture the optimal weights combination at each iteration of the optimisation process.

- To mitigate the issues caused by the accumulation of data of correlation matrices due to the increase of updated sampling process, a scheme of adaptive partitioning of these matrices was introduced to the kriging surrogate model, especially in high-dimensional tasks.

- Several methods have been investigated with regard to the robustness of the design. First, the Gradient Index method was evaluated, but due to its limitations a modified method has been developed to evaluate the sensitivity of each solution obtained by the kriging surrogate model. The Worst Case Optimisation method has also been explored to address with the constrained optimization task.
with uncertainties in design variables and an assessment of average performance added to the algorithm to make it more reliable.

**12.2 Recommendations for Further Work**

The general academic aim of this thesis was the creation of an optimization tool for electromagnetic design problems based on the kriging method. It has been shown that this approach is capable of shortening considerably the time necessary to find an optimum solution as the whole algorithm makes use of “surrogate” objective functions instead of “directly computed” values obtained from expensive FEM models.

In the context of robust optimisation, the evaluation of the influences caused by uncertainties in variables has also been considered, with emphasis on incorporating kriging predictions into the process. The sensitivity analysis has been developed converting the optimization problem into a multi-objective one through adding another target to minimize the gradient index. The Worst Case and related methods have been used to optimize the reliability of designs.

All these methodologies can be researched and improved further. The following suggestions can be put forward for future consideration:

1. The pre-test of the SMWEI, AWEI and their derivative strategies is constructed by the randomly distributed mean square error and the kriging approximation. But because of this built-in randomness there is an element of unpredictability of the performance. Perhaps a better, more accurate, simplified model could be constructed to improve the efficiency of the kriging predictions.

2. When dealing with multi-variable problems, creating kriging correlation matrices may cause a ‘combinatorial explosion’ of associated data. Two techniques to alleviate this problem have been pursued in this thesis, the ‘zoom-in’ and ‘partitioning’, but neither is completely satisfactory. There is therefore scope for further research into more efficient ways of creating and storing the ‘bid data’ resulting from such correlation matrices. This is a big problem in its own right, not treated fully in this project – although it was identified as a potential problem while working on some of the problems reported here – and venturing into the big mathematical area of statistics.
3. The context of robust optimisation has been covered quite thoroughly, but further work is required to consider other related issues, such as probabilistic and non-probabilistic cases, local versus global models, but also matters associated with manufacturing processes.
Appendix A The Concepts of Multi-Objective Optimization

Figure 1: Ideal and utopian objective vector.

Definition 1 (Ideal Objective Vector). The $m$-th component of the ideal objective vector $z^*$ is the constrained minimum solution of the following problem:

\[
\begin{align*}
\text{Minimize } & f_m(x) \\
\text{Subject to } & x \in S \\
& \text{for } m = 1, \ldots, M.
\end{align*}
\]

Definition 2 (Utopian Objective Vector). A utopian objective vector $z^{**}$ has each of its components marginally smaller than that of the ideal objective vector:

\[
z_{i}^{**} = z_{i}^{*} - \varepsilon_i
\]

with $\varepsilon_i > 0$, $i = 1, \ldots, M$.

Both of these special solutions appear in Figure 2.2. Here, it should be mentioned that the utopian objective vector is not a feasible solution, whereas the ideal objective vector is feasible.

Definition 3 (Dominance). For any two solutions $x^{(1)}$ and $x^{(2)} \in S$, $x^{(1)}$ is said to dominate $x^{(2)}$ if and only if both the following conditions are true:
1. \( f_i(x^{(1)}) \leq f_i(x^{(2)}) \) for all \( i = 1, 2, ..., M \)

2. \( f_i(x^{(1)}) < f_i(x^{(2)}) \) for at least one \( i \in \{1, 2, ..., M\} \)

Thus, a solution \( x^{(1)} \) is said to dominate another solution \( x^{(2)} \) if and only if it is better in at least one of the objectives, and it is no worse in all other objectives. In this sense, \( x^{(1)} \) is a better solution than \( x^{(2)} \). We say that \( x^{(1)} \) is non-dominated by \( x^{(2)} \).

**Definition 4 (Non-dominated set).** Among a set of solutions \( P \subseteq S \), the non-dominated set of solutions \( P' \) are those that are not dominated by any member of the set \( P \).

Out of a set of feasible solutions \( P \) to the problem of Equation 2.4, the solutions which may be considered optimal from the set are those in the non-dominated set. When the set \( P \) of feasible solutions is the entire feasible region \( S \), then the non-dominated set \( P' \) is called the (global) Pareto-optimal set.

**Definition 5 ((Global) Pareto-optimal set).** The non-dominated set of the entire feasible search space \( S \) is the global Pareto-optimal set.

The global Pareto-optimal set is not dominated by any other solutions in \( S \), and so is the solution to Equation 2.4. As well as the global Pareto-optimal set, local Pareto-optimal sets may also exist:

**Definition 6 ((Local) Pareto-optimal set).** A set of solution vectors \( L \) is a local Pareto-optimal set if \( \exists \delta > 0 \) such that \( \forall x \in L, \ x \) is non-dominated in \( L \cap B(x, \delta) \).

The set of values in the objective space which correspond to the solutions in the global (local) Pareto-optimal set is called the global (local) Pareto-optimal front.

**Definition 7 ((Global) Pareto-optimal Front).** The image of the global Pareto-optimal set in the feasible objective space is the global Pareto-optimal front.

**Definition 8 ((Local) Pareto-optimal Front).** The image of a local Pareto-optimal set in the feasible objective space is a local Pareto-optimal front.

The Pareto-optimal front for an objective space \( Z \) is highlighted in Figure 2.6 and includes a local Pareto-optimal front as well.
Appendix A The Concepts of Multi-Objective Optimization

Figure 2: Global and local Pareto-optimal fronts

**Definition 9 (Strong Dominance).** A solution \( x^{(1)} \) strongly dominates a solution \( x^{(2)} \) if
\[
f_i(x^{(1)}) < f_i(x^{(2)}) \quad \text{for} \quad i = 1, \ldots, M.
\]

Thus if a solution \( x^{(1)} \) is better than another solution \( x^{(2)} \) in all objectives, then \( x^{(1)} \) is called to strongly dominate \( x^{(2)} \).

**Definition 10 (Weakly non-dominated set).** Among a set of solutions \( P \in S \), the weakly non-dominated set of solutions \( P' \) are those that are not strongly dominated by any other member of the set \( P \).

When the set \( P \) of feasible solutions is the entire feasible region, then the non-dominated set is called the weakly Pareto optimal set.
Appendix B Conference Papers


EXPLORATION VERSUS EXPLOITATION USING KRIGING SURROGATE MODELLING IN ELECTROMAGNETIC DESIGN

Song Xiao, Mihai Rotaru and Jan K. Sylwiński

School of Electronics and Computer Science (ECS), University of Southampton, Southampton, UK
xxlg10@ecs.soton.ac.uk, mr@ecs.soton.ac.uk, jks@soton.ac.uk

Abstract — This paper discusses the use of kriging surrogate modelling in multiobjective design optimisation in electromagnetics. The importance of achieving appropriate balance between exploration and exploitation is emphasised when searching for the global optimum. It is argued that this approach will yield a procedure to solve time consuming electromagnetic design problems efficiently and will also assist the decision making process to achieve a robust design of practical devices considering tolerances and uncertainties.

Introduction

Electromagnetic design almost always carries a heavy burden of high computational cost, with very few exceptions when a very simplistic analytical, empirical or equivalent circuit based model is found to be adequate for performance prediction. Most of the time throughout the design process, or at least at later stages, numerical models are required to provide necessary accuracy, typically employing 3D simulation using finite element or related technique. In the optimisation part of the design routine a single objective function evaluation may require a full field solution of the entire complicated model, often transient, or even several solutions (if averaged values are needed), which may be very ‘expensive’ in terms of computing time involved. Thus it is not enough to have confidence that the algorithm finds the global optimum; for practical purposes it must do so with as few objective function calls as possible. Thus within the context of searching for the optimum (usually minimum) of a particular objective function (or functions in multiobjective problems, e.g. best performance and simultaneously minimum cost), another minimum is being sought, that is looking for a strategy which finds the optimum with a minimum use of the computationally expensive performance predicting software. To complicate things further, the issue of robustness of the design comes into consideration — related to manufacturing tolerances, material variability, etc. — which requires the designer not only to find the optimum design but also know more about its “quality”; in other words the “shape” of the objective function must be estimated. In the context of stochastic optimisation this is usually expressed in terms of a compromise between exploration (searching the unexplored space) and exploitation (using information already provided) and is often supplemented and supported by various types of surrogate modelling. This paper investigates these issues and uses “kriging” as the main technique for constructing the surrogate model.

Kriging and the Utility Functions

Kriging [1] can predict the shape of the objective function based only on limited information and estimates the accuracy of this prediction; this is helpful in assisting the main decision of the optimisation process where to put the next point for evaluation. A ‘utility function’ is usually constructed, based on the predicted error, which may seamlessly adjust the way of searching between the regions with confidence and uncertainty. Thus providing an efficient and robust way to achieve a balance between exploration of unknown regions with degree of uncertainty and exploitation of attractive areas with high confidence is imperative.
A brief overview of one-stage kriging methodology is first given. The method exploits the spatial correlation of data in order to build interpolation; hence the correlation function is very important. We use the standard linear regression (1) and the correlation is modelled as (2):

\[ \hat{y}(x) = \sum_{i=1}^{n} \beta_i f_i(x) + \epsilon(x) \]  

\[ R(e(x'), e(x'')) = \prod_{i=1}^{n} \exp\left[-\|x' - x''\|^2 / p_i^2\right] \]  

where the global function \( \sum_{i=1}^{n} \beta_i f_i(x) \) and an additive Gaussian noise \( \epsilon(x) \) are integrated to the predicted value \( \hat{y}(x) \) of the objective function. \( \hat{R} \) is the correlation amongst the data in a direction and \( p_i \) determines the ‘smoothness’ of (2). The most popular correlation function is given by the Gaussian model where the value of \( p_i \) is simply taken as equal to 2.

In general, the ‘expected improvement’ utility function, based on the potential error predicted by a kriging model, is commonly used to select multiple design vectors for evaluation. The Expected Improvement Function (EI) [2] is defined as

\[ \text{EI}(f(x)) = \begin{cases} (f_{\text{min}} - \hat{y}(x)) \phi\left( \frac{f_{\text{min}} - \hat{y}(x)}{\sigma(x)} \right) + \sigma(x) \phi\left( \frac{f_{\text{min}} - \hat{y}(x)}{\sigma(x)} \right) & \text{if } s(x) > 0 \\ 0 & \text{if } s(x) = 0 \end{cases} \]  

where \( \hat{y}(x) \) is the objective function value of \( x \) as predicted by the kriging model, given by equation (1), \( s(x) \) is the root mean squared error in this prediction, and \( \phi\left( \frac{f_{\text{min}} - \hat{y}(x)}{\sigma(x)} \right) \) and \( \phi\left( \frac{f_{\text{min}} - \hat{y}(x)}{\sigma(x)} \right) \) are Gaussian density function and Gaussian distribution function, respectively.

The Expected Improvement Function (EI) may be viewed as a fixed compromise between exploration and exploitation: when the \( s(x) \) operator given by the Kriging method is positive, the first term of equation (3) favours searching the promising regions with high confidence, whereas the second term in the same equation favours searching the regions with high uncertainty. Through use of a practical kriging-assisted single-objective tests developed specially to assess the performance of these terms, it has been shown that the second term representing exploration performs dramatically better in terms of finding the global optimum of the objective function, whereas the exploitation often can only find the local minimum. Since EI applies equal weights to the two terms, it may be seen as a fixed compromise between exploration and exploitation.

The balance between exploration and exploitation is a critical issue when attempting to find the global optimum of an objective function. The Weighted Expected Improvement (WEI) [3] is derived from EI by adding a tuneable parameter which can adjust the weights on exploration and exploitation, whilst the quality of the approximation of the objective function can be improved by incorporating the newly evaluated design vector at each iteration. The WEI utility function used in this work may be written as

\[ \text{WEI}(f(x)) = \begin{cases} (w f_{\text{min}} - \hat{y}(x)) \phi\left( \frac{f_{\text{min}} - \hat{y}(x)}{\sigma(x)} \right) + (1-w) \sigma(x) \phi\left( \frac{f_{\text{min}} - \hat{y}(x)}{\sigma(x)} \right) & \text{if } s(x) > 0 \\ 0 & \text{if } s(x) = 0 \end{cases} \]  

where the tuneable parameter \( w \ (0 < w < 1) \) controls the balance between the two terms (exploration and exploitation), therefore searching globally and locally [3]. The efficiency of the kriging with WEI has been tested with the Schwefel test function [4] as an objective function in the interval \([500\ 500]\) for different values of \( w \). The multi-dimensional Schwefel test function is defined as

\[ f(x) = \sum_{i=1}^{d} -x_i \sin(\sqrt{|x_i|}) \]  

In the one-dimensional case used here, \( d=1 \). We have studied the performance of different algorithms using the Schwefel function throughout the tests. While testing using a single function may not be conclusive, the Schwefel function has in the past been found helpful when testing similar algorithms.
Appendix B Conference Papers

It has been found that the kriging model assisted by WEI when $w < 0.55$ 1 can only find a local minimum; this is perhaps not surprising as a strong weight has been applied to the term favouring exploitation. When $w = 0.54$ the kriging model is able to find the global minimum. Notable values of $w$ are $w = 1$, which puts all emphasis on exploitation, $w = 0$, which focuses on exploration, and $w = 0.5$, which makes the algorithm equivalent to EI. Table 1 summarizes the results of our tests.

<table>
<thead>
<tr>
<th>Value of weight</th>
<th>Number of iterations</th>
<th>Value of weight</th>
<th>Number of iterations</th>
<th>Value of weight</th>
<th>Number of iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Fails</td>
<td>0.57</td>
<td>3 (finds LM)</td>
<td>0.5 (EI)</td>
<td>11 (finds GM)</td>
</tr>
<tr>
<td>0.9</td>
<td>3 (finds LM)</td>
<td>0.56</td>
<td>9 (finds LM)</td>
<td>0.4</td>
<td>7 (finds GM)</td>
</tr>
<tr>
<td>0.8</td>
<td>3 (finds LM)</td>
<td>0.55</td>
<td>9 (finds LM)</td>
<td>0.3</td>
<td>12 (finds GM)</td>
</tr>
<tr>
<td>0.7</td>
<td>3 (finds LM)</td>
<td>0.54</td>
<td>13 (finds GM)</td>
<td>0.2</td>
<td>17 (finds GM)</td>
</tr>
<tr>
<td>0.6</td>
<td>3 (finds LM)</td>
<td>0.53</td>
<td>14 (finds GM)</td>
<td>0.1</td>
<td>15 (finds GM)</td>
</tr>
<tr>
<td>0.59</td>
<td>3 (finds LM)</td>
<td>0.52</td>
<td>11 (finds GM)</td>
<td>0</td>
<td>Fails</td>
</tr>
<tr>
<td>0.58</td>
<td>3 (finds LM)</td>
<td>0.51</td>
<td>11 (finds GM)</td>
<td>0</td>
<td>Fails</td>
</tr>
</tbody>
</table>

Table 1: Performance of WEI for $w$ between 0 and 1 (LM — local minimum, GM — global minimum).

In order to understand better the effects of $w$ more tests in the range of 0.5 to 0.6 were done. As shown in the table somewhere between 0.54 and 0.55 there is a changeover between a regime where only a local minimum is found and values of $w$ which allow for the global minimum to be correctly identified. Too much emphasis on exploitation is a risky strategy. Equal weights ($w = 0.5$ as in EI) are "safe", but not optimal in a sense that there is a value of $w$ around 0.4 which can provide an answer with fewer iterations (7 instead of 11). Figure 1 shows a snapshot position after the global minimum has been found after 11 iterations (using EI) and after 7 iterations (using $w = 0.4$). For both cases the same six initial points were used (in practice their positions may be selected randomly) required before a particular EI or WEI strategy can be applied. The graph also shows the "history" of how the points were added throughout the iterative process. Both strategies successfully find the global minimum and the quality of the final answer is comparable, but WEI with $w = 0.4$ is more efficient.

![Figure 1: The performance of the kriging model with WEI ($w = 0.4$) and EI (the single square with a number indicates the iteration number of kriging with EI; the double square with a number shows the iteration number of kriging with WEI ($w = 0.4$)).](image-url)
Adaptive Weighted Expected Improvement

The experiments of the previous section demonstrated the importance of the optimal choice of the weights, both in terms of the ability of the algorithm to achieve the correct answer (global minimum) and doing it efficiently (fewer iterations required); unfortunately the optimal choice of weights is normally problem dependent and thus a modified strategy is required to make the method more intelligent and guide itself automatically through the process.

Reinforcement learning is a goal-directed learning approach to what to do next and how to map the situation to actions so as to maximize a numerical reward [6]. In this paper we propose to automatically tune the weighting parameter $w$ in response to the environment feedback. In particular, the Mean Square Error (MSE) from the kriging model is used to guide the choice of the optimum weight $w$ and the concept of an award is introduced. Thus the algorithm calculates the average value of the MSE of every predicted point and uses these values as the basis of calculating the potential rewards. Then, after comparing the rewards from different weight distributions, the weights are redistributed on the two terms which control the exploration and exploitation so that the biggest reward is achieved. The Adaptive Weighted Expected Improvement (AWEI) strategy is described as one of the possible algorithms in Fig. 2. AWEI endeavours to encourage exploration or exploitation depending on the results of the initial pre-test, one with emphasis on exploration and another on exploitation. Two rewards (Reward1 and Reward2) are calculated and compared; on the basis of this comparison $w$ is then chosen to encourage either exploration or exploitation.

![Flowchart](image)

Figure 2: The decision-making chart for different strategies of balancing exploration and exploitation; term1 favours exploration while term2 favours exploitation; AMSE = average mean square error
Practical Performance of the Adaptive Weighted Expected Improvement

Several tests using the AWEI assisted kriging model for different values of \( \beta \) (a selectable parameter as described in Fig. 2) have been undertaken to assess its performance. However, one particular problem was identified and needed special attention. The term which encourages exploitation can sometimes cause the kriging model to stop because of choosing repeatedly the same new point for evaluation (within the specified accuracy). Should this happen (or should – for any other reason – one of the rewards not be assessed properly or fail), the algorithm is effectively reset and the Expected Improvement (EI) function is temporarily applied to select the next point for evaluation; in the next step the algorithm reverses to the AWEI. We have used the Schwefel test function again with the initial sample points imposed as \( x = -450, -230, -150, -40, 160, 500 \) and the tunable parameter \( \beta \) varied in a controlled way. When \( \beta = 0.001, 0.005, 0.01 \) the model fails to find the global minimum; when \( \beta = 0.05, 0.08, 0.1 \) altogether 12 iterations are needed to find the global minimum; when \( \beta = 0.2 \) or 0.3 a better performance is observed with 8 iterations needed to find the global minimum. As demonstrated by Fig. 3, however, the best performance with only 5 iterations needed was observed when \( \beta = 0.35 \). Compared to the previously described WEI, the AWEI is more flexible thanks to the built-in feedback that uses the reward scheme to make decisions on how to adapt the EI function.

![Graph](image)

Figure 3: The performance of the kriging model with AWEI for \( \beta = 0.35 \)
(the circle with a number means more exploitation at that iteration;
the square with a number means more exploitation at that iteration).

Surrogate Model based Weighted Expected Improvement approach with rewards

The AWEI is based on reinforcement learning and takes account of the feedback, which in turn uses predicted uncertainty gained from the kriging model to make a decision as a trade-off between exploitation and exploration driven by the amount of reward resulting from each action [7]. The AWEI consistently selects the action which yields the largest average reward [8] at each step of the iterative process based on the best information available, which may not necessarily be accurate or reliable. So although optimal in short term the selected action may not always be beneficial in long term. In the third strategy developed and presented in this paper an attempt is made to predict the cumulative rewards likely to occur on long terms as a consequence of a particular choice of actions. This approach follows the ideas first introduced recently in [8] in the context of games theory to a well known one-armed bandit problem. This approach requires predicting the long term rewards, rather than short term at a given iteration step, which necessitates some estimation of the long term
Appendix B Conference Papers

consequences of the actions selected. A simple (but very inefficient in the context of electromagnetic design problems) approach would involve continuing iterations independently (in parallel) for the two initially selected (and at that point fixed) weight functions using WEI until the kriging process stops in either of the tests (because of repeating the point for evaluation) and then using the most recent calculated value of the rewards to select the more promising action. This strategy has the advantage of assessing long term benefits (rather than immediate ones) but can only be applied to problems where objective function evaluation is ‘cheap’ (in terms of computing times) - as was indeed the case in the original paper [8]. However, it appears that the main concept can still be useful if supplemented by another modification to the algorithm with the aid of surrogate modelling. Thus rather than using the ‘expensive’ model (typically the time consuming finite element field modelling software) we can create a simplified surrogate model based on existing data points and continue the parallel search for the global minimum of the surrogate model – which will be a very quick process and thus not adding noticeably to the overall computing times – before the rewards for the two alternative actions are compared and the final decision is made regarding the location of the next point for evaluation. We then use finite element (or similarly ‘expensive’ software) to get a new ‘reliable’ point, update the surrogate model and continue iterations.

Thus a particular contribution of this paper to the concept of predicting the likely long term potential rewards before the next point is evaluated is the addition of surrogate modelling so that the ‘forward prediction’ is cheap; inevitably such a prediction will be less reliable than using real data points (which in this application, as already stressed, would be too expensive and thus unacceptable) but may result in an overall better assessment of long term benefits of different actions than a simple one-stage algorithm developed and described earlier as the Adaptive Weighted Expected Improvement strategy.

There are of course many methods of constructing a surrogate model. We suggest, and have implemented, using the root mean square error already available in the kriging prediction, even though this is not the real error between the kriging approximation and the real objective function (which is of course unknown at this stage). A random error is added to the calculated error and thus effectively we now have two kriging surrogate models simultaneously, the original one based on the most recent ‘real’ data points and a second one – used only for the purpose of ‘forward prediction’ of the long term effects of a particular action – which ultimately leads to an overall long term ‘reward’ of a particular action. As there are two possible actions and they are assessed independently we end up with two rewards; the better reward will identify the better cause of action, a new point is selected, the finite element programme executed and a new point added to the curve. This will give rise to a new surrogate model and a new ‘secondary’ surrogate model (or rather a pair of models as there are two parallel actions); the process will continue until some termination criteria are met. The flowchart of the decision making process can be easily followed in Fig. 2.

As before the Schwefel test function was chosen to test the Surrogate Model based Weighted Expected Improvement approach with rewards (SMWEI). The choice of the values $w_1$ and $w_2$ is a matter of further experiments in order to generate some guidelines about how to select the initial values. Moreover, as the two actions are independent the weights $w_1$ and $w_2$ do not actually have to add up to 1, although most of the testing assumed that they do. Finally, as the ‘secondary’ kriging surrogate model relies on a random distribution of error, all tests were conducted ten times with the same pair of values of $w_1$ and $w_2$ and performance averaged. Throughout the testing the same initial sample points were assumed to allow comparison, but in reality each point may be distributed randomly or using one of the accepted strategies such as a Latin Hypercube sampling. Some results will now be discussed. When $w_1=0.6$ and $w_2=0.4$ the number of iterations required to find a global minimum is between 10 and 16, when $w_1=0.7$ and $w_2=0.3$ it is between 9 and 16, for $w_1=0.8$ and $w_2=0.2$ it is 10 to 17, finally for $w_1=0.9$ and $w_2=0.1$ it is 11 to 15. The average number of iterations for the pairs of values above is 13, 12, 15 and 12, respectively, so it is quite steady and does not appear to be sensitive to the variation of the values. Figure 4 shows a particular set of results for $w_1=0.7$ and $w_2=0.3$; there is an interesting ‘departure’ at iteration 10 to explore a remote region before returning to the global minimum at iteration 11. Finally, Fig. 5 demonstrates a particularly successful case when $w_1=0.7$ and $w_2=0.1$ where only 6 iterations were required.
Figure 4: The performance of the kriging surrogate model with SMWEI (the circle with a number means more exploration at that iteration; the square with a number means more exploitation at that iteration).

Figure 5: The 'best' performance of SMWEI with $w_1=0.7$ and $w_2=0.1$ (the circle with a number means more exploration at that iteration; the square with a number means more exploitation at that iteration).

There appears to be little benefit in applying the last strategy compared with the previous AWEI algorithm, but more testing is required to draw more meaningful conclusions. In particular, it is interesting to notice that the SMWEI clearly makes a better attempt at exploring local minima – this may prove very important in the context of robust design where not only the value but also the shape of the minimum is of relevance. Thus a strategy which explores the space more thoroughly may after all be preferable even if more expensive. At this stage the three strategies are considered as alternative and all are available in the flowchart of Fig. 2. An attempt will be made in the future to provide guidelines about how to select one strategy for the given problem in hand.
Finally, a test was conducted to see how the algorithm performs when the initial points are not distributed favourably, for example if positioned as in Fig. 6. This is clearly a challenging case as the initial points give very little clue as to the real shape of the objective function. Rather remarkably the SMWEI algorithm performs very robustly with only 9 iterations required to find the global minimum, whereas kriging with EI needs 12 iterations. Other tests of that nature were equally encouraging.

**Conclusion**

Two novel algorithms have been proposed, both adopting concepts of reinforcement learning, in an attempt to automatically balance exploration and exploitation in computationally expensive electromagnetic design optimisation problems. Both are based on kriging surrogate modelling and use the notion of rewards for selecting the best position of the next point for evaluation. The one-stage algorithm appears to perform very efficiently in terms of its ability to find a global minimum, whereas the strategy based on two kriging surrogate models and forward performance prediction offers more reliable information about the shape of the objective function. Both algorithms will be implemented in practical design of electromagnetic and electromechanical devices.

**References**

STRATEGIES FOR BALANCING EXPLORATION AND EXPLOITATION IN ELECTROMAGNETIC OPTIMISATION

Seng Xiao, Mihai Ristoiu and Jan K. Sykulski
Electronics and Computer Science (ECS), University of Southampton, Southampton, UK
sx100@ecs.soton.ac.uk, mrj@ecs.soton.ac.uk, jks@ecs.soton.ac.uk

Abstract — The paper focuses on the advantages and drawbacks of different strategies which may be used to assist kriging surrogate modelling with the purpose of selecting multiple design vectors for evaluation when stepping forward in optimisation routines. The combined effort includes the efficiency of finding the global optimum but also the quality of the approximation of the shape of the objective function; the latter may be used to make judgements about the robustness of the optimised design.

I. 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 to use an '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 [1-4]. 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.

II. KRIGING WITH DIFFERENT STRATEGIES

A. Kriging and different strategies

Kriging [1] 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 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 Function (EI) [2], or Weighted Expected Improvement (WEI) [3]. The EI function is defined as

\[
EI = \left( f_{\text{opt}} - f(x) \right) \Phi \left( \frac{f_{\text{opt}} - f(x)}{\sigma(x)} \right) + \sigma(x) \phi \left( \frac{f_{\text{opt}} - f(x)}{\sigma(x)} \right)
\]

where \( f_{\text{opt}} \) is the predicted maximum value of the objective function by the kriging model, and \( \sigma(x) \) 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 an objective function is one of the desired 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 term 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 EI by adding a tunable parameter which can adjust the weights on exploration and exploitation.

As suggested by previous tests [4], the optimal choice of the weights is known as critical in terms of the ability of the algorithm to achieve 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 [5] called Adaptive Weighted Expected Improvement (AWEI) and Surrogate Model based Weighted Expected Improvement approach with rewards (SWEI) [6] have been proposed to make the process of tuning weights more intelligent and self-regulating.

The Mean Square Error (MSE) from the kriging model is used to calculate the rewards. The AWEI can tune the weights automatically based on the comparison between the potential rewards from two different weight distributions emphasizing exploitation and exploration, respectively. After comparison, the weights are redistributed on the two terms of Eq. (1) to encourage exploration or exploitation depending on the results of the initial pre-test. However, the AWEI only takes account of the short term rewards at a given iteration step, whereas the SWEI can predict the cumulative rewards likely to occur in long term as a consequence of a particular choice of actions. Furthermore, the SWEI 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 modeling 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 optimal 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.

B. The SWEI with Multi-weights in 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 in the SWEI algorithm with certain pairs of weights in the pre-test performs better in terms of estimating the shape of the objective function, a
Appendix B Conference Papers

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 a 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 Fig. 1. In the pre-tests, if one of the rewards is not assessed properly or fails, the remaining rewards still participate in the comparison until an action with the biggest reward is chosen.

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

C. The 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 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 Fig. 1.

Because all pre-tests in SMWEI apply ‘cheap’ simplified surrogate model based on the specific prediction and potential error produced by kriging, the Mean Square Error might be directly used in each pre-test’s remaining iterations instead of the Expected Improvement. The simplified surrogate model is 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.

III. RESULTS

Fig. 2(a) shows one of the test results to approximate the two variables Schwefel function [7], which is a two-objective task. The global minimum has been found after 9 iterations, but the quality of the shape representation of the objective function is poor in Fig. 2(b). Complete results will be reported in the full version.

IV. CONCLUSIONS

It is argued that a multi-reward scheme based on a surrogate model may provide the best prediction of long-term benefits for achieving best balance between exploration and exploitation.

REFERENCES

Appendix B Conference Papers

Adaptive Weighted Expected Improvement with Rewards Approach in Kriging Assisted Electromagnetic Design

Song Xiao, Mihai Rotaru and Jan K. Sykulski, Fellow IEEE
Electronics and Computer Science (ECS), University of Southampton
Southampton, SO17 1BJ, UK
jkv@soton.ac.uk

Abstract—The paper explores a kriging surrogate modelling combined with expected improvement approach for the design of electromagnetic devices. A novel approach based on the concept of rewards is proposed and demonstrated in the context of TEAM Workshop Problem 22. Balancing exploration and exploitation is emphasized and robustness of the design considered.

I. INTRODUCTION

The paper builds on our previous publication [1] where Adaptive Weighted Expected Improvement (AWEI) was first introduced and the concept of rewards [2] implemented and tested using the Schwefel test function [3]. In this paper the technique is extended to real design problems and illustrated on the TEAM Workshop Problem 22 [4]. The algorithm exploits kriging [1, 5] as an efficient tool to balance exploration with exploitation.

II. BALANCING EXPLORATION AND EXPLOITATION

The strategy for balancing exploration and exploitation has been explained in [1]; a simplified decision-making chart is shown in Fig. 1 with two rewards computed to select the more promising weighting factor for the Expected Improvement (EI) algorithm. Note that to achieve a robust design the shape of the objective function may be helpful, not just the optimum.

Fig. 1. Simplified decision-making chart for AWEI balancing exploration exploitation (for full version see [1]).

III. RESULTS FOR TEAM WORKSHOP PROBLEM 22

The Superconducting Magnetic Energy Storage (SMES) device is described in [4]; in this digest we show the results for two and three parameter problem respectively. The objective function as defined in [4] is computed for two variables (Fig. 2 - contour plot). One of the challenges in the creation of the kriging correlation matrix is the ‘combinatorial explosion’ when many parameters are used. The technique proposed here to cope with this situation is a successive ‘zoom in’ strategy. Our algorithm starts first in searching the whole domain (black * Fig. 3), however after a defined number of runs it “zooms” in around the region where the optimum value is located (red + Fig. 2). Using this strategy the algorithm finds the optimum in a relative low number of runs without overwhelming the computer resources. This approach has been successful for three parameter problem as shown in Fig. 3.

Fig. 2. The 2 Parameter SMES problem tested by kriging with EI

Fig. 3. The 2 Parameter SMES problem tested by kriging with EI (blue cube: rough searching range, red cube ‘zoom in’ range)

IV. CONCLUSIONS

Exploration and exploitation in the design algorithm are controlled via a kriging model, the Weighted Expected Improvement strategy and the use of ‘rewards’ with robust design also considered. To avoid problems associated with combinatorial explosion a ‘zoom in’ strategy is proposed.

V. REFERENCES

Appendix B Conference Papers

Robust Global Optimization of Electromagnetic Designs
Utilizing Gradient Indices and Kriging

S. Xiao, M. Rotaru and J. K. Sykulski, Fellow, IEEE

Electronics and Computer Science (ECS), University of Southampton, Southampton, SO17 1BJ, UK

Since uncertainties in variables are inevitable an optimal solution should consider the robustness of the design. A methodology based on the use of first and second-order gradient indices is proposed introducing the notion of gradient sensitivity. A kriging method assisted by algorithms exploring the concept of rewards is employed to facilitate function predictions for the robust optimization process. The performance of the proposed algorithms is assessed through a series of numerical experiments. A modification to the correlation model by introduction of a kriging predictor and Mean Square Error allows efficient solution of large scale and multi-parameter problems. The three parameter version of TEAM Workshop Problem 22 has been used for illustration.

Index Terms—Design optimization, Pareto optimization, kriging, TEAM Workshop Problem 22.

I. INTRODUCTION

The concept of the Gradient Index (GI) is explored in [1]-[3]. The method transforms a problem into a multi-objective optimization by concurrently minimizing the function and its gradient index, thus forming pareto fronts. This approach is promising but creates a challenge how to select a preferred solution when 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. In this paper similar ideas are pursued but important changes are introduced. First, the concept of the gradient index sensitivity is introduced and explained. Secondly, the second-order gradient index is used to assist the process. Finally, rather than calculating the objective function using computationally costly finite element software, a kriging prediction is employed. In other words, the objective function is approximated using kriging [4] assisted by algorithms balancing exploration and exploitation ([5], [6]) using the concept of rewards [7]. This strategy has been shown previously to be very efficient and has the advantage that it can be linked with any finite element software.

II. ROBUST OPTIMIZATION

In conventional optimization the minimum (maximum) of an objective function is sought while the search space is limited through a set of constraints. Once the global optimum has been found the problem is considered to have been solved. When practical devices are designed, however, we need to recognize 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.

A. Multi-Objective Robust Optimization using GI

Consider a one-variable test function [1]-[3] (see Fig. 1):

\[ f(x) = \frac{3.5}{1 + (x-5)^2} + \frac{1}{1 + (x-15)^2} + \frac{1}{1 + (x-25)^2} \]

Fig. 1. Example of a robust design for a one-variable problem:
(a) Objective function, the gradient index and sensitivity,
(b) First and second order gradients,
(c) Pareto solution.

Manuscript received June 25, 2013. Corresponding author: J. K. Sykulski (e-mail: jks@ecs.soton.ac.uk).

Digital Object Identifier inserted by IEEE
The uncertainties may be either specified directly (e.g. as machining tolerances, say $\delta$) or defined mathematically as

$$t(x) = \{ x \in \mathbb{R}^n | k \sigma < x \leq x + k \sigma \}$$

(2)

where $\sigma$ is standard deviation of uncertain variables and $k$ is determined by a confidence level [2].

One way of incorporating robustness into the mainstream optimisation process is by adding the gradient index [1] as a second objective and formulating the problem as

Minimise $f(x)$ where $x \in \mathbb{R}^n$ and $k \leq x \leq k + x$.

Minimise $f(x)$ where $x \in \mathbb{R}^n$ and $k \leq x \leq k + x$.

(3)

Subject to $g_i(x) \leq 0, i = 1, \ldots, m$.

Point $A_1$ in Fig. 1(a) is the theoretical global optimum. However, any small change in the variable $x$ results in a large variation of the objective function, thus $A_1$ is not a robust design and $A_2$ or $A_3$ might be preferred. The final decision, however, is not straightforward and is influenced by the size of the uncertainty. The sensitivity of the gradient may also be defined as the difference between the smallest and the largest value of the $G_1$ within the uncertainty range, as shown in Fig. 1(b). The shape of this sensitivity carries useful information. The second order gradient of the function may also be useful (Fig. 1b). Another example (Fig. 2) shows a "sharp" global and "shallow" local minima and a "plateau" with associated sensitivities and second order gradient indices.

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 optimization process, which may become problematic especially when dealing with large-scale multi-parameter problems, leading to a "combinatorial explosion." In [11] we proposed a successive "zoom in" strategy to solve the problem, where - in order to reduce the amount of data storage and utilise the installed physical memory capacity efficiently - the step sizes of design vectors 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 improperly, 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.

B. 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 these are unlikely to cause the memory problems. On the other hand, the correlation matrix between the known points and all the design vectors is very large and increases with every iteration. For problems with several variables the correlation matrix can grow very quickly and may result in 'combinatorial explosion'. Therefore a new 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 in order
to make full use of the available memory while minimizing
the number of calls of the correlation model. Two schemes
have been implemented and operate throughout the process as
explained in Fig. 3, either partitioning in terms of the sampling
cases or via the design vectors – the choice is governed by
considerations of optimal memory management and will
depend on the physical size of the memory available. Another
benefit is that the processes of calculating the kriging predictor
and the Mean Square Error are combined. Hence more time is
saved while the kriging model is being built.

![Diagram](image)

Fig. 3. The correlation matrices partitioned by: (a) sampling points,
(b) design vectors, (where S1, S2, ..., Sn: sampling points).

To verify the viability of the proposed methodology, the
TEAM problem 22 has been attempted [8]. Without the
strategy of Fig. 4, the kriging algorithm failed due to the
memory size problem when tested on a smaller computer.

Fig. 4. The flowchart of the adaptive division of correlation matrices.

IV. APPLICATION IN ELECTROMAGNETIC DESIGN

A. TEAM Problem 22

The full description of the TEAM benchmark problem 22
(superconducting magnetic energy storage system) may be
found in [8]. The target is to achieve an arrangement of the
two superconducting coils such that the stored energy within
the system is \( E_{\text{ref}} = 180 \text{MJ} \) while a minimal stray field \( B_{\text{new}} \)

\[
\text{where } B_{\text{new}} = 3 \text{pT, and } B_{\text{new}}^2 = \frac{\Delta E}{E_{\text{ref}}} \text{, subject to}
\]
geometrical and ‘quench’ constraints.

B. 2 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 \( (D_2 = 0.394 \text{m}) \),
while \( R_2 \) and \( h_2 \) are varied. Usually the initial sampling points
are selected using Latin Hypercube [10]; however, for this test
eight initial sampling points were used as shown in Table I.
Table II compares one typical result from literature with our
AWEI algorithm (kriging with Adaptive Weighted Expected
Improvement) [5], [6], while Fig. 5 demonstrates the
convergence process of AWEI.

![Graph](image)

Fig. 5. TEAM Workshop Problem 22 (a) Prediction by kriging with AWEI
\((3 \leq R_2 \leq 3.3, 0.48 \leq h_2 \leq 0.5, D_2 = 0.394, \) other parameters fixed),
(b) Sensitivity with respect to \( R_2 \) and \( h_2 \), (c) Pareto solution.

The uncertainties are pre-defined as \( R_2 - 0.02 < \xi(R_2) < \)
\( R_2 + 0.02 \), \( h_2 - 0.01 < \xi(h_2) < h_2 + 0.01 \). whilst the
increments with respect to \( R_2 \) and \( h_2 \) for calculating first-order
gradients are set as \( dx(R_2) = 0.002, dx(h_2) = 0.01 \).

C. 3 Parameter Test Results

The full 3 parameter TEAM 22 problem [8] is potentially a
challenge to the kriging method because of the ‘combinatorial
explosion’ associated with setting up the correlation model, as
explained in Section III – thus the savings due to avoidance of
computationally expensive finite element simulations may be
lost, or even overtaken, by the excessive time required by
Appendix B Conference Papers

The uncertainties for $R_2$, $h_2$ and $d_2$ have been pre-set as $R_2 = 0.03 < (R_3) < R_4 + 0.03$, $h_2 = 0.402 < (h_3) < R_5 + 0.03$, $d_2 = 0.009 < (d_3) < d_4 + 0.009$. The differences $(d_2)$ required by the gradient calculation with respect to these three parameters were set as their relevant step sizes. The pareto solutions obtained using kriging with EI and AWEI are presented in Figs. 7a and 7b, respectively. Both base 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.

V. 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.

REFERENCES


CONSIDERATIONS OF UNCERTAINTY IN ROBUST OPTIMISATION OF ELECTROMAGNETIC DEVICES

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

Electronics and Computer Science (ECS), University of Southampton
Southampton, SO17 1BJ, UK, flos@soton.ac.uk

Abstract – Due to unavoidable uncertainties related to material properties and manufacturing processes, the robustness of the optimal solution must be considered when designing electromagnetic devices. In this paper, the worst-case optimisation (WCO) and the worst-case-based WCO are proposed to evaluate the robustness of both performance and constraints under uncertainty. To reduce computing times when searching for the robust solution a predicted objective function is used, obtained with the help of a kriging algorithm which explores the searching space using the concept of rewards. Finally, to avoid some of the shortcomings of WCO, the concept of average performance evaluation is developed.

Introduction

In electromagnetic design, uncertainties in design variables are inevitable, thus the ability to evaluate the robustness is critical while pursuing the theoretical optimum. This is particularly true when constrained optimisation is considered, as illustrated later by an example involving a quenching condition for a superconducting material, in order to maintain the solutions within the feasible region when perturbation occurs. The worst-case optimization (WCO) method has been selected to evaluate the accuracy of the prediction of the objective function provided by an improved kriging model, implemented for the sake of reducing computational effort associated with direct application of information produced by a time-consuming Finite Element Method (FEM). However, in this study some shortcomings of the WCO approach have been identified associated with the inability to assess the performance variation under conditions of uncertainty. Therefore the concept of average performance evaluation was suggested as an improved measure of robustness.

Robust Optimization Algorithms exploiting kriging modelling

In general, seeking the minimum (maximum) of an objective function, while the search space is restricted by certain constraints, is the aim of conventional optimization

\[
\text{Minimize } f(x) \quad \text{Subject to } \quad g_i(x) \leq 0, \quad i = 1, \ldots, m
\]  

(1)

When designing real devices, however, many design variables are subject to specific uncertainties (manufacturing tolerances, variation of material properties, etc.). Hence the assessment of the influence caused by these uncertainties on the performance becomes essential in practical design problems. As a consequence, finding a theoretical optimum may not be sufficient and the robustness of the theoretical optimal solution needs to be considered as well. Algorithms applying different strategies to evaluate robustness, such as the sensitivity analysis [1], the worst-case optimization method [2-6], and the mean value and variance of performance [7-9], have been developed to assist in the design tasks. In our previous work [10] a multi-objective optimization method, which included sensitivity analysis using gradient index, was developed and demonstrated.
Appendix B Conference Papers

A widely used approach to evaluate the reliability of a robust solution is the worst-case method. This technique can maintain a certain level of robustness by avoiding solutions that may push the function into infeasible region when searching for the optimum, as shown in Fig. 1(a). A set of typical examples is shown in Fig. 1(a) to illustrate the principle of the worst-case optimization technique. The theoretical optimum A, for example, may be abandoned in favour of a ‘worse’ solution B because the actual design A', which accounts for the uncertainties of the variables, might violate constraints and enter infeasible area. The uncertainties can be defined mathematically as

\[ u(x) = \{ \xi \in \mathbb{R}^n | x - ka \leq \xi \leq x + ka \} \]  

where \( \sigma \) is standard deviation of uncertain variables and \( k \) is determined by confidence level [11]. The uncertainties may also be specified directly (e.g. as machining tolerances, say A'). As an algorithm which can predict the worst scenario considering the uncertainties, as well as constraints with respect to specific designs, the worst case optimization (WCO) method [2-5] may be applied to analyse the reliability of the solution as follows

\[
\text{Minimize } f_w(x) = \max_{\xi \in u(x)} f(\xi) \quad \text{Subject to } g_{\xi i}(x) = \max_{\xi \in u(x)} g_i(x) \leq 0, i = 1, \ldots, m
\]  

The worst values of the objective function and the \( i \)-th constraint function are chosen to substitute the original values of the nominal design \( x \).

Numerical methods, such as finite elements, are often used when searching for the worst objective function value under imposed constraints which may be an extremely time-consuming process. To reduce the computational burden the worst_vertex-based WCO (W-WCO) [4] was proposed; this algorithm only needs to observe the vertices within the region restricted by uncertainties rather than evaluating every design value. For example, in the problem illustrated in Fig. 1(b), in addition to \( x \) there are 8 more points required, located at the corners and the middle of the specified boundary. However, in certain cases assessing only these 8 points might still not be sufficient. Fig. 1(c) illustrates such a case were a large variation of the function will not be identified by the W-WCO method. Therefore finding a balance between mitigating the heavy computational burden of the original worst-case method and pursuing more detailed evaluation is the main issue to be addressed.

One of the possibilities is to employ the improved kriging method, assisted by a set of strategies capable of balancing exploration and exploitation [12, 13, 17] using the concept of rewards [14], which can be used to predict the objective function value instead of directly calculating it using computationally expensive FEM models. Based on such a kriging prediction, the worst-case method can be directly implemented. In other words, the WCO method uses the predicted information rather than the expensive FEM models. The accuracy of the predicted objective function using the improved kriging model has been considered in [10]. In this paper the suitability of directly using WCO with the predicted function model is discussed and demonstrated.
Kriging surrogate model

As a kind of regression model, kriging [15] is able to predict the shape of the objective function via spatial correlation of data using limited information. The accuracy of this prediction can also 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 optimization 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). In order to build a correct surrogate model of the real objective function through interpolation. This relies on the linear regression model

\[
\hat{y}(x) = \sum_{k=1}^{n} \beta_k f_k(x) + \varepsilon(x)
\]  

(4)

and the Gaussian correlation model

\[
R \left( \varepsilon(x'), \varepsilon(x'') \right) = \prod_{k=1}^{n} e^{-\theta \| x' - x'' \|^2}.
\]

(5)

where the global function \( \sum_{k=1}^{n} \beta_k f_k(x) \) and an additive Gaussian noise \( \varepsilon(x) \) are integrated to the predicted value \( \hat{y}(x) \) of the objective function; \( \theta \) is the correlation amongst the data in \( k \)-direction and \( \rho \), determines the ‘smoothness’ of \( \varepsilon \). The most popular correlation function is given by the Gaussian model where the value of \( \rho \) is simply taken as equal to 2. For a given set of data, the maximum likelihood estimation optimizes the value of \( \theta \) and then the correlation model is brought into the regression model to evaluate the function with the best linear unbiased predictor [4].

Although kriging can potentially solve large multi-parameter optimization problems, it has some inherent limitations making the implementation difficult. In particular, for multi-parameter problems, the correlation model built by the kriging algorithm can grow very fast resulting in a ‘combinatorial explosion’ of correlation data filling very quickly the memory of standard computer workstations. As a result the process can become slow and inefficient. To solve this issue a scheme that adaptively partitions the correlation matrices was developed [10]. Using this approach the size of the data is managed to use efficiently the available memory throughout the iterative process of kriging. The scheme mentioned above has another advantage as the kriging predictor and Mean Squared Error are being calculated at the same time hence more computational time is saved in the process of building the kriging model. Therefore this modification makes kriging suitable for solving multi-parameter optimization problems and could be linked with WCO which needs detailed data to work effectively.

Average performance

As explained in the previous sections, the WCO method can be used to find robust solutions for a particular problem once the constrains and the uncertainties of the variables have been defined. In this section, however, we address the extreme case depicted in Fig. 2, it is argued that WCO on its own is not sufficient to find a reliable and robust solution. To deal with such a situation the concept of an average performance has been suggested and will now be explained. Figures 2(a) and 2(b) show two similar functions, otherwise identical, except the region around the points A2 and A4, respectively. If WCO is used to find a robust solution using the same uncertainties \( \Delta \) for the case depicted in Fig. 2(a), point A2 will be found, whereas for the case shown in Fig. 2(b) point A4 is likely to be returned as a robust solution. It is clear from this example that A4 is by no means a robust solution and although this may be considered a somewhat extreme situation it illustrates the fact that the WCO algorithm cannot differentiate between a robust and non-robust solution in such cases. To resolve such problems the concept of average performance within the uncertainty range is introduced.

The initial idea of evaluating average performance was to simply calculate the average value of all the potential perturbed values in the uncertain region with respect to the solution. This, however, proved infeasible, as although the shape of the two objective functions is clearly different, they share some common characteristics as explained in Table 1.
Appendix B Conference Papers

Fig. 2. (a) The objective function with a robust local minimum subject to uncertainty $\Delta$
(b) the objective function with a non-robust local minimum subject to uncertainty $\Delta$

<table>
<thead>
<tr>
<th>Global minimum</th>
<th>Local minimum</th>
<th>Uncertainty</th>
<th>Worst case</th>
<th>Worst case</th>
<th>Average value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x(1, A1)$</td>
<td>$x(55, A4)$</td>
<td>$\Delta$</td>
<td>$x(15, A3)$</td>
<td>$x(55, A4)$</td>
<td>$2.9556$</td>
</tr>
</tbody>
</table>

Table 1: The common features of the two functions

By visual inspection we can see that the variation around the solution $A4$ is much more intense than around $A2$, but the average value within the shaded areas with middle-points $A2$ and $A4$ for the given uncertainty are the same and equal to $2.9556$. This means that the average value criterion may not be useful in assessing the variation of the function close to the point of interest. Hence an average value of the gradient index (GII) [16] has been introduced as an alternative way of assessing the average performance. The average value of the gradient index is calculated as:

$$\text{Minimize } \text{Average } G(x) = \frac{\sum_{x \in X} \max \left( \frac{\partial f(x)}{\partial x_i} \right)}{n}$$

(6)

where $x \in \mathbb{R}^2$ ($x_1 \leq x \leq x_2$) is the $i$-dimensional design variable vector with lower and upper bounds $x_1$ and $x_2$, respectively, and $n$ is the sum of the design vectors. The sum of the maximum gradients is divided by the total number of design vectors. The average first-order gradient for the first case is then found to be 0.0594, while in the second case it is three times larger 0.1416. Therefore the average gradient index could be used as a more reliable criterion to evaluate the average performance. This criterion can therefore be combined with the WCO method to resolve difficult problems such as the one described by Fig. 2(b). By generalizing this methodology it can be argued that a robust optimization problem can be transformed into a three-objective optimization problem defined as

$$\begin{align*}
\text{Minimize } & f(x) \\
\text{Minimize } & f_i(x) = \max_{x \in U(i)} f(x) \quad \text{Subject to } g_{i, j}(x) = \max_{x \in U(j)} g(x) \leq 0, i = 1, \ldots, m \\
\text{Minimize } & \text{Average } G(x) = \frac{\sum_{x \in X} \max \left( \frac{\partial f(x)}{\partial x_i} \right)}{n}
\end{align*}$$

Robust Optimization Algorithms exploiting kriging modelling

In our previous work [10] the robustness of an optimal solution was evaluated using the gradient index, where the task of robust optimization was transformed into a two-objective optimization. One objective was to minimize the difference between the absolute value of the largest and the smallest
Appendix B Conference Papers

gradients within the uncertain range, called the ‘sensitivity’, while minimizing the objective function that was the second objective. Although the sensitivity calculated by the gradient index method is able to provide information on the rate of change of the objective function, the WCO method can also be employed to obtain similar information. However, as shown in the previous section, the WCO method has some limitations, especially for extreme cases (Fig. 2(b)). The average performance assessment described above can thus be added to the WCO method to improve the overall reliability of the result. To verify the concept and to analyze further the average performance criterion, two problems have been tested and the results are reported below.

Fig. 3. (a) Analytic function (b) The kriging prediction (c) The performance of the WCO method (d) Average performance (e) The full-field solution for three objectives (f) The zoomed-in optimal part
First, the two-variable analytic function (8) depicted in Fig. 3(a), which was also used to assess the accuracy of the improved kriging model, has been tested.

\[
f(x) = 10 - \sum_{k=1}^{2} \left[ \frac{\sqrt{1+\left(x_1-x_2\right)^2}}{1+\left(x_1-x_2\right)^2} \right] \quad (0 < x_1 < 27) \tag{8}
\]

Kriging with adaptive weighted expected improvement (AWEI) [16] has provided an approximation of this analytic function with a test step size of 0.1. Within 85 iterations, the kriging model can find the global minimum successfully; however, for a better approximation of the shape of the objective function - the model continued to run until the 285th iteration (Fig. 3(b)).

The uncertainty with respect to variables is set as \( u(x_1, x_2) = 0.5 \), and the WCO method is used to obtain the surface of the worst case for each solution as shown in Fig. 3(c). For this case the "best" solution shifts from the theoretical optimum \( (x_1=5, x_2=3, y=2.843) \) to the location \( (x_1=5.1, x_2=3.1, y=9.033) \) which provides a more robust result for the given conditions. If the values of the uncertainties were to keep increasing, up to a certain extent, ultimately the robust optimum would thoroughly shift from the sharp global minimum to one of the preferable local minima with higher robustness. Figure 3(d) depicts the average gradient index values in the search space. Finally, the full-scale \( (0 < x_i < 27) \) optimal solutions including all three objectives have been presented in Fig. 3(e).

For clearer presentation of the pareto front, the full-scale version is zoomed in Fig. 3(f). Two typical pareto solutions are labelled in the zoomed-in graph: solution N1 delivers a more optimal value of the prediction of the objective value, while N2 offers a relatively better average performance.

### Application to Electromagnetic Design

The second example tested with the proposed WCO procedure involves a multi-objective version of the TEAM 22 benchmark problem [18]. The full description of the TEAM problem 22 may be found elsewhere and will not be repeated here. The target for this problem is to achieve an arrangement of the two superconducting coils such that the stored energy within the system is \( E_{\text{st}} \leq 1000 \text{MJ} \), while a minimal stray field \( B_{\text{mag}} \) is maintained. The objective function is defined as

\[
OF = \frac{B_{\text{mag}}}{B_{\text{norm}}} = \frac{B_{\text{mag}}}{B_{\text{ref}}} \tag{9}
\]

where \( B_{\text{norm}} = 3T \) and \( B_{\text{ref}} = \frac{\sum_{i=1}^{2} |B_{\text{mag},i}|}{22} \), subject to geometrical and "quench" constraints. The approach taken here combines WCO method with kriging and commercially available FEM based software. A 2D model of the TEAM 22 problems is solved throughout this procedure. The three parameter case of TEAM 22, which includes three geometric variables \( R_0, H_0 \), and \( D_0 \), while \( R_i, H_i \), and \( D_i \), are fixed, has been tried under different settings of uncertainties. The uncertainties are assumed to exist in the current densities \( J_i \) of the two coils, because normally they are limited within certain range by a current controller for compensating perturbation. A constraint is imposed that the superconducting coils should not violate the quench condition which links together the value of the current density and the maximum value of magnetic flux density as follows

\[
g_i(x) = |J_i| + 6.4 \cdot |H_{\text{mag}}| - 54.0 \leq 0, \quad i = 1, 2 \tag{10}
\]

Three tests were performed for this TEAM 22 problem. The initial data set-up, as well as the number of steps, for the three tests is listed in Table 2. One of the differences between the three sets is the size of the uncertainty; the first set had the uncertainty set as 0.1, the second test 0.2, while in the third test 0.35 was used. Another difference was in the number of iterations that were used to approximate the objective function by the kriging model with AWEI [10]. For the first data set (uncertainty 0.1) kriging generated 185 sampling points, for the second case (uncertainty 0.2) kriging produced 257 sampling points, while for the third set (uncertainty 0.35) only 188 points were necessary. The results returned by the WCO, coupled with kriging, for these three cases are summarised in Tables 3 to 5. Figures 4, 5, and 6 show both the full scale and zoomed-in versions of the pareto fronts obtained for the three cases. In the figures and tables, solutions P1, P4 and P7 refer to the global optimum; P3 and
P6 have the ‘worst case’ performance; while P2, P5 and P8 describe the best average gradient index solutions.

<table>
<thead>
<tr>
<th>Three variables</th>
<th>Uncertainties</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_e$ (m)</td>
<td>$h_e/2$ (m)</td>
</tr>
<tr>
<td>Lower bound</td>
<td>3.63</td>
</tr>
<tr>
<td>Upper bound</td>
<td>3.13</td>
</tr>
<tr>
<td>Step size</td>
<td>0.61</td>
</tr>
<tr>
<td>No of steps</td>
<td>11</td>
</tr>
</tbody>
</table>

Table 2: The initial setup of the prediction by the kriging model

<table>
<thead>
<tr>
<th>P1</th>
<th>P2</th>
<th>P3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_e$ (m)</td>
<td>3.04</td>
<td>3.09</td>
</tr>
<tr>
<td>$h_e/2$ (m)</td>
<td>0.492</td>
<td>0.464</td>
</tr>
<tr>
<td>$d_e$ (m)</td>
<td>0.397</td>
<td>0.382</td>
</tr>
</tbody>
</table>

Table 3: Results for the case when uncertainty is 0.1

Fig. 4.

The three objective optimization including
(Worst case (WC), Average gradient index performance (AVGI), the prediction of objective functions (Ob))
($U(J_e, J_0)$=0.1)

<table>
<thead>
<tr>
<th>P4</th>
<th>P5</th>
<th>P6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_e$ (m)</td>
<td>3.11</td>
<td>3.09</td>
</tr>
<tr>
<td>$h_e/2$ (m)</td>
<td>0.492</td>
<td>0.478</td>
</tr>
<tr>
<td>$d_e$ (m)</td>
<td>0.376</td>
<td>0.385</td>
</tr>
</tbody>
</table>

Table 4: Results for the case when uncertainty is 0.2
Appendix B Conference Papers

Fig. 5. The three objective optimization (U(z_1, z_2)=0.2)

Fig. 6. The three objective optimization (U(z_1, z_2)=0.35)

Unlike other reported work [16] that uses a stochastic optimization method to find the global optimum and then employs Monte Carlo method to explore the space around the global minimum, combined with WCO and the gradient index to judge the robustness of the solution, the method introduced here takes a holistic approach and explores the whole searching space. The kriging model allows comparison amongst several local minima (maxima) that may be more robust than the global...
optimum. Another major advantage of the procedure proposed in this work is the fact that it can be linked with any commercial electromagnetic design software giving more freedom to the designer.

**Conclusions**

The worst-case method (WCD) assisted by the prediction provided by the kriging model with AWEI has been proposed to solve robust optimization problems considering uncertainties of variables. The particular contribution of this paper is in using kriging prediction, rather than the computationally expensive finite element modelling, in assessing the robustness of the final design. The second contribution involves enhancing the worst case methodology through introducing the concept of average gradient index performance. Using this approach a conventional optimization problem, with constraints and uncertainties in variables, has been transformed into three-objective optimization with a relevant pareto front. The proposed algorithms have been verified by both numerical tests and a practical electromagnetic design problem described by TEAM 22 benchmark.

**References**


Correlation matrices in kriging assisted optimisation of electromagnetic devices

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

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

Keywords: Design optimisation, pareto optimum, kriging.

Abstract
Kriging surrogate modelling offers efficient decision making assistance on where to place the next point for evaluation during optimisation. This is particularly helpful in the design of electromagnetic devices where computationally expensive finite element modelling needs to be used. The disadvantage, however, is that correlation matrices are required which, for problems with many design variables and multiple objectives, may grow in size leading to the need for page swapping and slowing down of what in principle should be a very fast process. This paper addresses the issue of how to exploit the structure of these matrices to reduce computational burden.

1 Introduction
Kriging [1, 2] predicts the shape of the objective function by considering the spatial correlation of data based on limited information, thus offers efficient and inexpensive surrogate to replace the computationally demanding numerical simulation (such as finite elements). The accuracy of the prediction can be estimated by the mean square error in kriging to assist in a decision on where to place the next evaluation point during optimisation iterations. The spatial correlation exists between the known points (vectors) of the objective function and all the unknown points, as well as amongst the known points (newly found points and initial sampling points). The way of calculating and storing the kriging correlation matrices was suggested in [2] and is shown schematically in Figure 1.

2 The storage issue of correlation matrices
Although theoretically kriging could be used for any type and size of optimization, care must be taken with large problems (many variables and multi-objective) as correlation matrices can grow rapidly. This phenomenon, dubbed ‘combinatorial explosion’ [2], may slow up the kriging process so severely that using a surrogate model no longer offers time savings. The issue is of course related to the size of the available computer memory and the way in which ‘page swapping’ is organised. In our previous work [2] we proposed a ‘zooming’ strategy to limit the searching space while preserving the accuracy of the surrogate model. In our more recent work a strategy of partitioning the correlation matrices was suggested [3]. In this paper we investigate the issue further by considering the structure of the correlation matrices.

3 The structure of the correlation matrices
It has been observed that the sub-matrices created when new sampling points are added during iterations have similarities that can be exploited to reduce the memory requirements. In particular, the ‘shapes’ of the distribution of values in a given row of the matrix will be preserved (a ‘benchmark’) but shifted depending on the position of the point. Thus instead of creating and storing the whole matrix it may be sufficient to duplicate information. This is best explained using examples.

3.1 The single-variable numerical test
The single-variable Schwefel function [4] was first used

\[ f(x) = \sum_{i=1}^{d} -x_i \sin(\sqrt{|x_i|}) \]  

(1)

where \(d=2\), with six initial sampling points \(-500, -210, -250, -40, 160, 500\), paying attention that the first one was on the edge of the region. A full correlation sub-matrix was then created for this point and the distribution is shown in Figure 2a. It was then observed that all points have a similar distribution, with Figure 2b giving an example for the second point. It is therefore possible to set up a ‘standard’ sub-matrix (as depicted in Figure 3) and then shift it using a ‘window’ to appropriate position depending on the location of the point under consideration (effectively a ‘cut and paste’ technique), hence use storage space far more economically.

Figure 1: Kriging and partitioning of the correlation matrix, on the right the suggested improvements.
Appendix B Conference Papers

3.2 The two-variable numerical test

The following analytical function was then used:

\[ y = 10 - \sum_{i=1}^{n} \frac{1}{1 + (x_i - 2.5)^2} \]  \hspace{1cm} (2)

with two variables \((n=2)\) in the range \(0 \leq x_i \leq 27\). The point \((x_1=0, x_2=0, y=0.45)\) was selected to create a standard correlation sub-matrix to be used by all remaining and new points, as illustrated graphically in Figure 4.

![Figure 4: Building and duplicating sub-matrices.](image)

In order to assess achievable savings in computer memory and associated computing times, the graphs in Figure 5 have been produced, showing and comparing three cases: 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 described previously in [3].

![Figure 5: Memory requirements (a) and computing times (b).](image)

The test demonstrates that significant savings can be made, if required, of necessary memory requirements, by applying the concept of duplicating the standard correlation sub-matrix. This can be combined with the idea of matrix partitioning.

The added bonus is that the computing times can also be reduced, especially as the iterations progress, although in the case of matrix partitioning this benefit may be lost somewhat towards the end of the iterative process.

4 Conclusion

The structure of the correlation matrices associated with the kriging model may be exploited to avoid uncontrolled growth of such matrices for large problems to avoid the need for page swapping which will inevitably lead to slowing up of the process. In particular, it is possible to duplicate and shift data while storing only a small subset of necessary information. More description and tests will be provided in the full paper.

References


Six Sigma Quality Approach to Robust Optimisation

Song Xiao, Yinjiang Li, Mihai Rotaru and Jan K. Sykulski, Fellow, IEEE
Electronics and Computer Science, University of Southampton, Southampton, SO17 1BJ, UK.
jks@ecs.soton.ac.uk

Abstract—The paper explores the use of multi-objective robust optimization methodology based on standard deviation and mean value of objective function under uncertainty of variables. A modified kriging model to balance exploration and exploitation is employed to facilitate the objective function prediction.

Index Terms—Standard deviation, mean value, kriging.

1. INTRODUCTION

The Six Sigma techniques for process improvement and to aid business strategies were proposed in mid 80s and recently used in the context of quality manufacturing [1, 2]. In this paper we examine this approach as an alternative to the Worst Case method, aided by an improved kriging surrogate model, applied to robust electromagnetic design [3, 4]. The standard deviation and the mean value of objective function are exploited and the algorithm illustrated using a test function.

II. MULTI-OBJECTIVE ROBUST OPTIMISATION BASED ON STANDARD DEVIATION AND MEAN VALUE

In robust optimisation the objective function $f(x)$ is to be minimised subject to constraints and uncertainties; the latter may be specified directly (e.g. as machining tolerances, say $\Delta$) or defined mathematically as

$$U(x) = \{ x \in \mathbb{R} | \mu - \kappa \sigma \leq x \leq \mu + \kappa \sigma \}$$

where $\sigma$ is standard deviation of uncertain variables and $\kappa$ is determined by a confidence level [5]. In the Six Sigma Quality formulation the measure of dispersion is classified into six sigma (standard deviation) levels [1], while the optimisation problem can be reformulated as

Minimise $\mu$ and $\sigma$

where $\sigma$ is the standard deviation indicating the intensity of variation due to the uncertainty of variables and $\mu$ the mean value defining the average performance within the uncertain range. The usefulness of this strategy has been investigated with the aid of a test function [6, 7] defined as

$$y = 10 - 5 \times (e^{-2.2 \frac{(x_1 - 0.5)^2}{1 + (x_2 - 0.5)^2}} + e^{-2.2 \frac{(x_1 - 0.5)^2}{1 + (x_2 - 0.5)^2}} + e^{-2.2 \frac{(x_1 - 0.5)^2}{1 + (x_2 - 0.5)^2}})$$

in the range $0 \leq x \leq 27$. The single-variable ($n=1$) version is plotted in Fig. 1 with the corresponding standard deviation for the assumed uncertainty $\sigma=0.5$. Point A1 is the theoretical global optimum but, considering the perturbation caused by uncertainty, is not as robust as A2. The trajectory of the objective function in terms of standard deviation and mean value is plotted in Fig 2 with corresponding points A1, A2 and P1 to P6 transferred from Fig. 1. Thus it is the closeness of P5 and P6 to A2 which makes this solution attractive. A kriging assisted surrogate model with improved strategy for balancing exploration and exploitation has been employed [7] and a simplified decision-making chart is shown in Fig. 3.

REFERENCES

Appendix C Journal Papers


C4. Song Xiao, M. Rotaru, and J. K. Sykulski, “Adaptive weighted expected improvement with rewards approach in kriging assisted electromagnetic design,” has been accepted by IET.


C7. Song Xiao, M. Rotaru, and J. K. Sykulski, “Correlation matrices in kriging assisted optimisation of electromagnetic devices,” has been accepted by IET.
Exploration versus exploitation using kriging surrogate modelling in electromagnetic design

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

Abstract
Purpose – Design optimisation of electromagnetic devices is computationally expensive as use of finite element or similar codes is normally required. Thus, one of the objectives is to have efficient algorithms minimising the number of necessary function calls. In such algorithms a balance between exploration and exploitation needs to be found not to miss the global optimum but at the same time to make efficient use of information already found. This paper is a contribution to the search of such efficient algorithms.

Design/methodology/approach – This paper discusses the use of kriging surrogate modelling in multiobjective design optimisation in electromagnetics. The investigation relies on the use of special test functions.

Findings – The importance of achieving appropriate balance between exploration and exploitation is emphasised when searching for the global optimum. New strategies are proposed using kriging.

Originality/value – It is argued that the proposed approach will yield a procedure to solve time consuming electromagnetic design problems efficiently and will also assist the decision making process to achieve a robust design of practical devices considering tolerances and uncertainties.

Keywords Optimisation techniques, Electromechanical devices, Magnetic devices, Kriging, Surrogate modelling, Robust design, Electromagnetics

Paper type Research paper

1. Introduction
Electromagnetic design almost always carries a heavy burden of high computational cost, with very few exceptions when a very simplistic analytical, empirical or equivalent circuit based model is found to be adequate for performance prediction. Most of the time throughout the design process, or at least at later stages, numerical models are required to provide necessary accuracy, typically employing 3D simulation using finite element or related technique. In the optimisation part of the design routine a single objective function evaluation may require a full field solution of the entire complicated model, often transient, or even several solutions (if averaged values are needed), which may be very ‘expensive’ in terms of computing times involved. Thus, it is not enough to have confidence that the algorithm finds the global optimum; for practical purposes it must do so with as few objective function calls as possible. Thus, within the context of searching for the optimum (usually minimum) of a particular objective function (or functions in multiobjective problems, e.g. heat performance and simultaneously minimum cost), another minimum is being sought, that is looking for a strategy which finds the optimum with a minimum use of the computationally expensive performance predicting software. To complicate things further, the issue of robustness of the design comes into consideration – related to manufacturing
Appendix C Journal Papers

II. Kriging and the utility functions

Kriging (Lebensstajn et al., 2004) can predict the shape of the objective function based only on limited information and estimates the accuracy of this prediction; this is helpful in assisting the main decision of the optimisation process where to put the next point for evaluation. A "utility function" is usually constructed, based on the predicted error, which may seamlessly adjust the way of searching between the regions with confidence and uncertainty. Thus, providing an efficient and robust way to achieve a balance between exploration of unknown regions with degree of uncertainty and exploitation of attractive areas with high confidence is imperative.

A brief overview of one-stage kriging methodology is first given. The method exploits the spatial correlation of data in order to build interpolation; hence the correlation function is very important. We use the standard linear regression (1) and the correlation is modelled as (2):

\[ \hat{y}(x) = \sum_{k=1}^{m} \beta_k f_k(x) + e(x) \]  

\[ R(e(x'), e(x'')) = \prod_{k=1}^{n} e^{-\theta_k |x'_k - x''_k|^p} \]  

where the global function \( \sum_{k=1}^{m} \beta_k f_k(x) \) and an additive Gaussian noise \( e(x) \) are integrated to the predicted value \( \hat{y}(x) \) of the objective function. \( \theta_k \) is the correlation amongst the data in \( k \)-direction and \( p \) determines the "smoothness" of equation (2). The most popular correlation function is given by the Gauss model where the value of \( \theta_k \) is simply taken as equal to 2.

In general, the "expected improvement (EI)" utility function, based on the potential error predicted by a kriging model, is commonly used to select multiple design vectors for evaluation. The EI function (Jones et al., 1998) is defined as:

\[ EI(x) = \begin{cases} (f_{\text{min}} - \hat{y}(x))\psi \left( \frac{f_{\text{min}} - \hat{y}(x) - s(x)}{s(x)} \right) + s(x)\phi \left( \frac{f_{\text{min}} - \hat{y}(x) - s(x)}{s(x)} \right) & \text{if } s(x) > 0 \\ 0 & \text{if } s(x) = 0 \end{cases} \]  

where \( \hat{y}(x) \) is the objective function value of \( x \) as predicted by the kriging model, given by equation (1), \( s(x) \) is the root mean squared error in this prediction, and \( \psi((f_{\text{min}} - \hat{y}(x))/s(x)) \) and \( \phi((f_{\text{min}} - \hat{y}(x))/s(x)) \) are Gaussian density function and Gaussian distribution function, respectively.
The EI function may be viewed as a fixed compromise between exploration and exploitation: when the \( v(x) \) operator given by the kriging method is positive, the first term of equation (3) favours searching the promising regions with high confidence, whereas the second term in the same equation favours searching the regions with high uncertainty. Through a set of practical kriging-assisted single-objective tests developed specially to assess the performance of these two terms, it has been shown that the second term representing exploration performs dramatically better in terms of finding the global optimum of the objective function, whereas the exploitation often can only find the local minimum. Since EI applies equal weights to the two terms, it may be seen as a fixed compromise between exploration and exploitation.

The balance between exploration and exploitation is a critical issue when attempting to find the global optimum of an objective function. The weighted expected improvement (WEI) (Schober et al., 2005) is derived from EI by adding a tunable parameter which can adjust the weights on exploration and exploitation, whilst the quality of the approximation of the objective function can be improved by incorporating the newly evaluated design vector at each iteration. The WEI utility function used in this work may be written as:

\[
WEI'[f(x)] = \begin{cases} 
  w(f_{\text{max}} - f(x))\Phi\left(\frac{f(x) - f_{\text{max}}}{\sigma}\right) & \text{if } s(x) > 0 \\
  (1 - w)s(x)\phi\left(\frac{f_{\text{max}} - f(x)}{\sigma}\right) & \text{if } s(x) = 0 \\
  0 & \text{if } s(x) < 0
\end{cases}
\]

where the tunable parameter \( w \) (0 < w < 1) controls the balance between the two terms (exploration and exploitation), therefore searching globally and locally (Schober et al., 2005). The efficiency of the kriging with WEI has been tested with the Schwefel test function (Picheny et al., 2010) as an objective function in the interval \([-500,500]\) for different values of \( w \). The multi-dimensional Schwefel test function (Schwefel, 1981) is defined as:

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

In the one-dimensional case used here, \( d = 1 \). We have studied the performance of different algorithms using the Schwefel function throughout the tests. Schwefel's function is deceptive in that the global minimum is geometrically distant, over the parameter space, from the next best local minima. Therefore, the search algorithms are potentially prone to convergence in the wrong direction. Because of these properties, the Schwefel function has been a popular choice in testing the robustness of optimization algorithms. While testing using a single function may not be conclusive, the Schwefel function has in the past been found useful when validating similar algorithms (Pietak, 2010; Chen, 2009; Vakil-Baghmisheh and Salim, 2010); testing under practical conditions will obviously continue after the algorithm has been fully integrated into an electromagnetic design system. It has been found that the kriging model assisted by WEI (when \( w \in (0.5,1) \)) can only find a local minimum; this is perhaps not surprising as a strong weight has been applied to the term favouring exploitation. When \( w \in (0,0.5) \) the kriging model is able to find the global minimum. Notable values
of \( w \) are \( w = 1 \), which puts all emphasis on exploitation, \( w = 0 \), which focuses on exploration, and \( w = 0.5 \), which makes the algorithm equivalent to EI. Table I summarizes the results of our tests.

In order to understand better the effects of \( w \) more tests in the range of 0.5-0.6 were done. As shown in the table somewhere between 0.54 and 0.55 there is a changeover from a regime where only a local minimum is found and values of \( w \) which allow for the global minimum to be correctly identified. Thus, too much emphasis on exploitation is a risky strategy. Equal weights (\( w = 0.5 \) as in EI) are “safe”, but not optimal in a sense that there is a value of \( w \) around 0.4 which can provide an answer with fewer iterations (seven instead of 11). Figure 1 shows a snapshot position after the global minimum has been found after 11 iterations (using EI) and after seven

<table>
<thead>
<tr>
<th>Value of weight</th>
<th>Number of iterations</th>
<th>Value of weight</th>
<th>Number of iterations</th>
<th>Value of weight</th>
<th>Number of iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Fails</td>
<td>0.57</td>
<td>3 (finds LM)</td>
<td>0.5 (EI)</td>
<td>11 (finds GM)</td>
</tr>
<tr>
<td>0.9</td>
<td>3 (finds LM)</td>
<td>0.56</td>
<td>9 (finds LM)</td>
<td>0.4 (EI)</td>
<td>7 (finds GM)</td>
</tr>
<tr>
<td>0.8</td>
<td>3 (finds LM)</td>
<td>0.55</td>
<td>9 (finds LM)</td>
<td>0.3 (EI)</td>
<td>12 (finds GM)</td>
</tr>
<tr>
<td>0.7</td>
<td>3 (finds LM)</td>
<td>0.54</td>
<td>13 (finds GM)</td>
<td>0.2 (EI)</td>
<td>17 (finds GM)</td>
</tr>
<tr>
<td>0.6</td>
<td>3 (finds LM)</td>
<td>0.53</td>
<td>14 (finds GM)</td>
<td>0.1 (EI)</td>
<td>15 (finds GM)</td>
</tr>
<tr>
<td>0.58</td>
<td>3 (finds LM)</td>
<td>0.52</td>
<td>11 (finds GM)</td>
<td>0 (EI)</td>
<td>Fails</td>
</tr>
</tbody>
</table>

Table I. Performance of WEI for \( w \) between 0 and 1

Notes: LM – local minimum; GM – global minimum

Figure 1. The performance of the kriging model with WEI \( (w = 0.4) \) and EI

Notes: The single square with a number indicates the iteration number of kriging with EI; the double square with a number shows the iteration number of kriging with WEI \( (w = 0.4) \)
iterations (using \( e = 0.4 \)). For both cases the same six initial points were used (in practice their positions may be selected randomly) required before a particular EI or WEI strategy can be applied. The graph also shows the “history” of how the points were added throughout the iterative process. Both strategies successfully find the global minimum and the quality of the final answer is comparable, but WEI with \( e = 0.4 \) is more efficient.

### III. Adaptive weighted expected improvement

The experiments of the previous section demonstrated the importance of the optimal choice of the weights, both in terms of the ability of the algorithm to achieve the correct answer (global minimum) and doing it efficiently (fewer iterations required); unfortunately the optimal choice of \( e \) is normally problem dependent and thus a modified strategy is required to make the method more intelligent and guide itself automatically through the process.

Reinforcement learning is a goal-directed learning approach to what to do next and how to map the situation to actions so as to maximize a numerical reward (Sutton and Barto, 1998). In this paper we propose to automatically tune the weighting parameter \( e \) in response to the environment feedback. In particular, the mean square error (MSE) from the kriging model is used to guide the choice of the optimum weight \( e \) and the concept of an award is introduced. Thus, the algorithm calculates the average value of the MSE of every predicted point and uses these values as the basis of calculating the potential rewards. Then, after comparing the rewards from different weight distributions, the weights are redistributed on the two terms which control the exploration and exploitation so that the biggest reward is achieved. The adaptive weighted expected improvement (AWEI) strategy is described as one of the possible algorithms in Figure 2. AWEI endeavours to encourage exploration or exploitation depending on the results of the initial pre-test, one with emphasis on exploration and another on exploitation. Two rewards (Reward1 and Reward2) are calculated and compared; on the basis of this comparison \( e \) is then chosen to encourage either exploration or exploitation.

### IV. Practical performance of the adaptive weighted expected improvement

Several tests using the AWEI assisted kriging model for different values of \( \beta \) as a selectable parameter as shown in Figure 3 have been undertaken to assess its performance. However, one particular problem was identified and needed special attention. The term which encourages exploitation can sometimes cause the kriging model to stop because of choosing repeatedly the same new point for evaluation (within the specified accuracy). Should this happen (or should – for any other reason – one of the rewards not be assessed properly or fail), the algorithm is effectively reset and the EI function is temporarily applied to select the next point for evaluation; in the next step the algorithm reverses to the AWEI. We have used the Schwefel test function again with the initial sample points imposed as \( x = -450, -220, -150, -40, 100, 500 \) and the tuneable parameter \( \beta \) varied in a controlled way. When \( \beta = 0.001, 0.005, 0.01 \) the model fails to find the global minimum; when \( \beta = 0.05, 0.08, 0.1 \) altogether 12 iterations are needed to find the global minimum; when \( \beta = 0.2 \) or 0.3 a better performance is observed with eight iterations needed to find the global minimum. As demonstrated by Figure 3, however, the best performance with only five iterations needed was observed when \( \beta = 0.38 \). Compared with the previously described WEI,
Figure 2. The decision-making chart for different strategies of balancing exploration and exploitation.

Notes: Term 1 favours exploration while term 2 favours exploitation; AMSE – average mean square error.

the AWEI is more flexible thanks to the built-in feedback that uses the reward scheme to make decisions on how to adapt the EI function.

V. Surrogate model based weighted expected improvement approach with rewards

The AWEI is based on reinforcement learning and takes account of the feedback, which in turn uses predicted uncertainty gained from the kriging model to make a decision as a trade-off between exploitation and exploration driven by the amount of reward resulting from each action (Pavlidis et al., 2009). The AWEI consistently selects the action which yields the largest average reward (Sykulski et al., 2010) at each step of the iterative process based on the best information available, which may not necessarily be accurate or reliable. So although optimal in short term the selected action may not always be beneficial in long term. In the third strategy developed and presented in this paper an attempt is made to predict the cumulative rewards likely to occur on long terms as a
consequence of a particular choice of actions. This approach follows the idea first introduced recently in Sykulski et al. (2010) in the context of games theory to a well known one-armed bandit problem. This approach requires predicting the long term rewards, rather than short term at a given iteration step, which necessitates some estimation of the long term consequences of the actions selected. A simple but very inefficient in the context of electromagnetic design problems) approach would involve continuing iterations independently (in parallel) for the two initially selected (and at that point fixed) weight functions using WEL until the kriging process stops in either of the tests (because of repeating the point for evaluation) and then using the most recent calculated value of the rewards to select the more promising action. This strategy has the advantages of assessing long term benefits (rather than immediate ones) but can only be applied to problems where objective function evaluation is “cheap” (in terms of computing times) – as was indeed the case in the original paper (Sykulski et al., 2010). However, it appears that the main concept can still be useful if supplemented by another modification to the algorithm with the aid of surrogate modelling. Thus, rather than using the “expensive” model (typically the time consuming finite element field modelling software) we can create a simplified surrogate model based on existing data points and continue the parallel search for the global minimum of the surrogate model which will be a very quick process and thus not adding noticeably to the overall computing times – before the rewards for the two alternative actions are compared and the final decision is made regarding the location of the next point for evaluation. We then use finite element or similarly “expensive” software to get a new “robust” result, update the surrogate model and continue iterations.

Thus, a particular contribution of this paper to the concept of predicting the likely long term potential rewards before the next point is evaluated is the addition of surrogate modelling so that the “forward prediction” is cheap; inevitably such a prediction will be

---

**Figure 3.**
The performance of the kriging model with AWEI for $\alpha = 0.3$.
Appendix C Journal Papers

COMPEL
31.5

1548

less reliable than using real data points (which in this application, as already stressed, would be too expensive and thus unacceptable) but may result in an overall better assessment of long term benefits of different actions than a simple one-stage algorithm developed and described earlier as the AWEI strategy.

There are of course many methods of constructing a surrogate model. We suggest, and have implemented, using the root mean square error already available in the kriging prediction, even though this is not the real error between the kriging approximation and the real objective function (which is of course unknown at this stage). A mean square error distributed randomly is added to the specific approximation and thus effectively we now have two kriging surrogate models simultaneously, the original one based on the most recent "real" data points, and a second one — used only for the purpose of "forward prediction" of the long term effects of a particular action — which ultimately leads to an overall long term "reward" of a particular action. As there are two possible actions and they are assessed independently we end up with two rewards; the better reward will identify the better cause of action, a new point is selected, the finite element programme executed and a new point added to the curve. This will give rise to a new surrogate model and a new "secondary" surrogate model (or rather a pair of models as there are two parallel actions); the process will continue until some termination criteria are met. The flowchart of the decision making process can be easily followed in Figure 2.

As before the Schwefel test function was chosen to test the surrogate model based weighted expected improvement (SMWEI) approach with rewards. The choice of the values \( w_1 \) and \( w_2 \) is a matter of further experiments in order to generate some guidelines about how to select the initial values. Moreover, as the two actions are independent the weights \( w_1 \) and \( w_2 \) do not actually have to add up to 1, although most of the testing assumed that they do. Finally, as the "second" kriging surrogate model relies on a random distribution of error, all tests were conducted ten times with the same pair of values of \( w_1 \) and \( w_2 \) and performance averaged. Throughout the testing, the same initial sample points were assumed to allow comparison, but in reality such points may be distributed randomly or using one of the accepted strategies such as a Latin Hypercube sampling. Some results will now be discussed. When \( w_1 = 0.6 \) and \( w_2 = 0.4 \) the number of iterations required to find a global minimum is between 10 and 16, when \( w_1 = 0.7 \) and \( w_2 = 0.3 \) it is between 9 and 16, for \( w_1 = 0.8 \) and \( w_2 = 0.2 \) it is 10-17, finally for \( w_1 = 0.9 \) and \( w_2 = 0.1 \) it is 11-15. The average number of iterations for the pairs of values above is 13, 12, 15 and 12, respectively, so it is quite steady and does not appear to be sensitive to the variation of the values. Figure 4 shows a particular set of results for \( w_1 = 0.7 \) and \( w_2 = 0.3 \), there is an interesting "departure" at iteration 8 to explore a remote region before returning to the global minimum at iteration 9. Finally, Figure 5 demonstrates a particularly successful case when \( w_1 = 0.7 \) and \( w_2 = 0.1 \) where only six iterations were required.

There appears to be little benefit in applying the last strategy compared with the previous AWEI algorithm, but more testing is required to draw more meaningful conclusions. In particular, it is interesting to notice that the SMWEI clearly makes a better attempt at exploring local minima — this may prove very important in the context of robust design where not only the value but also the shape of the minimum is of relevance. Thus, a strategy which explores the space more thoroughly may after all be preferable even if more expensive. At this stage the three strategies are considered as alternative and all are available in the flowchart of Figure 2. An attempt will be
Note: The circle with a number means more exploration at that iteration; the square with a number means more exploitation at that iteration.

made in the future to provide guidelines about how to select one strategy for the given problem in hand.

Finally, a test was conducted to see how the algorithm performs when the initial points are not distributed favourably, for example if positioned as in Figure 6. This is
Notes: A circle with a number means more exploration, whereas a square means more exploitation at that iteration; a double square means using EI at that iteration.

VI. Conclusion

Two novel algorithms have been proposed, both adopting concepts of reinforcement learning, in an attempt to automatically balance exploration and exploitation in computationally expensive electromagnetic design optimisation problems. Both are based on kriging surrogate modelling and use the notion of rewards for selecting the best position of the next point for evaluation. The one-stage algorithm appears to perform very efficiently in terms of its ability to find a global minimum, whereas the strategy based on two kriging surrogate models and forward performance prediction offers more reliable information about the shape of the objective function. Both algorithms will be implemented in practical design of electromagnetic and electromechanical devices.

References


Appendix C Journal Papers


About the authors

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

Milind Rotaru has been a Lecturer of Applied Electromagnetics at University of Southampton since 2007. Prior to this, he was a Senior Research Engineer with the Institute of Microelectronics, Singapore, working on signal and power integrity issues on electronic packaging. His research interests are in computational electromagnetics, design and optimisation of electromechanical devices.

Jan K. Skulski is Professor of Applied Electromagnetics at the University of Southampton, UK. His research interests are 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 200 scientific papers. He is founding Secretary of International Comprising 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 the International Steering Committee 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.eecs.soton.ac.uk/info/people/jks). Jan K. Skulski is the corresponding author and can be contacted at: jks@eecs.soton.ac.uk

To purchase reprints of this article please e-mail: reprints@emeraldinsight.com
Or visit our web site for further details: www.emeraldinsight.com/reprints
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 using 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 optimisation 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.
2. Kriging with different strategies

2.1 Kriging and different strategies

Kriging (Lehmanstjern 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:

$$\text{EI} = \begin{cases} 
    f(x) - \hat{y}(x) \Phi \left( \frac{f_{\text{min}} - \hat{y}(x)}{s(x)} \right) + \hat{y}(x) \phi \left( \frac{f_{\text{min}} - \hat{y}(x)}{s(x)} \right) & \text{if } s(x) > 0 \\
    0 & \text{if } s(x) = 0
\end{cases}$$

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 optimisation 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 (Szkulski 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 PEM 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-test 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 for each emphasizing exploration and the other emphasizing 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 “backup” 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 for emphasizing exploration and the other for 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 \((d = 2)\):

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

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.
Appendix C Journal Papers

![Flowchart Illustrating Exploration and Exploitation](image)

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

### Exploration and exploitation
3.2 The performance of the kriging with different strategies

In order to analyze and compare the effects of applying different strategies to assist the kriging method, the two-variable Schwefel function with the initial samples imposed at $x_1 = -450, x_2 = 450, x_1 = -300, x_2 = 250, x_1 = -300, x_2 = -450, x_1 = 300, x_2 = 450, x_1 = -300, x_2 = -450, x_1 = 250, x_2 = 450, x_1 = 0, x_2 = 0$ 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 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
Appendix C Journal Papers

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

**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.
COMPEL
32,4

1182

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

Table 1. The optimal weights of the kriging assisted by SMWEI with multi-weights at each iteration within the iterative process.

<table>
<thead>
<tr>
<th>Iteration number</th>
<th>x1</th>
<th>x2</th>
<th>y</th>
<th>The optimal weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-280</td>
<td>-380</td>
<td>-478.7100</td>
<td>0.3 (emphasis exploration)</td>
</tr>
<tr>
<td>2</td>
<td>-300</td>
<td>-270</td>
<td>-478.5300</td>
<td>0.65 (emphasis exploration)</td>
</tr>
<tr>
<td>3</td>
<td>-350</td>
<td>-280</td>
<td>-563.2100</td>
<td>0.8 (emphasis exploration)</td>
</tr>
<tr>
<td>4</td>
<td>-280</td>
<td>-210</td>
<td>-552.2100</td>
<td>0.7 (emphasis exploration)</td>
</tr>
<tr>
<td>5</td>
<td>-300</td>
<td>-390</td>
<td>-655.6400</td>
<td>0.7 (emphasis exploration)</td>
</tr>
<tr>
<td>6</td>
<td>340</td>
<td>420</td>
<td>-528.2600</td>
<td>0.4 (emphasis exploration)</td>
</tr>
<tr>
<td>7</td>
<td>350</td>
<td>500</td>
<td>239.8700</td>
<td>0.3 (emphasis exploration)</td>
</tr>
<tr>
<td>8</td>
<td>500</td>
<td>500</td>
<td>306.1800</td>
<td>0.4 (emphasis exploration)</td>
</tr>
<tr>
<td>9</td>
<td>440</td>
<td>420</td>
<td>-472.6300</td>
<td>0.6 (emphasis exploration)</td>
</tr>
<tr>
<td>10</td>
<td>440</td>
<td>280</td>
<td>-801.5300</td>
<td>0.4 (emphasis exploration)</td>
</tr>
<tr>
<td>11</td>
<td>410</td>
<td>420</td>
<td>-822.8300</td>
<td>0.1 (emphasis exploration)</td>
</tr>
<tr>
<td>12</td>
<td>420</td>
<td>420</td>
<td>-857.7700</td>
<td>0.5 (UI)</td>
</tr>
</tbody>
</table>

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.

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 \((x_1 = 45, x_2 = 46, y = 700)\) and a local one in Region B \((x_1 \in [16, 26], x_2 \in [16, 26], y = 700)\). 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 performance.

![Figure 6. The objective function for the robust test](image)

探究与开发
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 SMWEl with adaptively tuning weights used the same initial conditions. Compared with EI, the SMWEl approach is more efficient in finding the global minimum (only 85 iterations – see Figure 8); however, the error between the objective

<table>
<thead>
<tr>
<th>$x_1$</th>
<th>$x_2$</th>
<th>Objective function value ($y$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>10</td>
<td>850</td>
</tr>
<tr>
<td>13</td>
<td>12</td>
<td>810</td>
</tr>
<tr>
<td>19</td>
<td>13</td>
<td>805</td>
</tr>
<tr>
<td>19</td>
<td>41</td>
<td>1,000</td>
</tr>
<tr>
<td>41</td>
<td>10</td>
<td>1,000</td>
</tr>
<tr>
<td>10</td>
<td>44</td>
<td>840</td>
</tr>
</tbody>
</table>

Table II. Initial sampling points

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 predictions 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 (Lebensztein 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.

![Figure 8](image.jpg)

(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


Further reading

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.

Miha 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 Computation 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 Committee 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

To purchase reprints of this article please e-mail: reprints@emeraldinsight.com
Or visit our website for further details: www.emeraldinsight.com/reprints
Appendix C Journal Papers

Adaptive Weighted Expected Improvement With Rewards Approach in Kriging Assisted Electromagnetic Design

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

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

The paper explores kriging surrogate modelling combined with expected improvement approach for the design of electromagnetic devices. A novel algorithm based on the concept of rewards is proposed, tested and demonstrated in the context of TEAM Workshop Problem 22. Balancing exploration and exploitation is emphasized and robustness of the design considered.

Index Terms—Computational electromagnetics, design optimization, pareto optimization, topology.

I. INTRODUCTION

The paper builds on our previous publication [1] where Adaptive Weighted Expected Improvement (AWEI) was first introduced and the concept of rewards [2] implemented and tested using the Schwerfel test function [2]. In this paper the technique is investigated further, extended to real design problems and illustrated using the TEAM Workshop Problem 22 [4]. The algorithm exploits kriging [1], [5] as an efficient tool to balance exploration with exploitation.

II. BALANCING EXPLORATION AND EXPLOITATION

A. Kriging

As a kind of regression model, kriging [5] is able to exploit the spatial correlation of data in order to predict the shape of the objective function based on limited information. Moreover, it can estimate the accuracy of this prediction, which may be extremely helpful in assisting the main decision of any optimization process, namely where the next evaluation point (vector) should be located.

Kriging exploits the spatial correlation of data in order to build interpolation; therefore the correlation function is a critical element. This relies on the linear regression model (1) and the Gaussian correlation model (2)

\[
\hat{y}(x) = \sum_{k=1}^{m} h_k f_k(x) + \epsilon(x)
\]

\[
R(x_1, x_2) = \prod_{k=1}^{n} \exp \left(-\frac{\lVert x_1 - x_2 \rVert^2}{\theta_k^2} \right)
\]

where the global function \( \sum_{k=1}^{m} h_k f_k(x) \) and an additive Gaussian noise \( \epsilon(x) \) are integrated to the predicted value \( \hat{y}(x) \) of the objective function; \( h_k \) is the correlation amongst the data in \( k \)-direction and \( \theta_k \) determines the ‘smoothness’ of (2). The most popular correlation function is given by the Gaussian model where the value of \( \theta_k \) is simply taken as equal to 2. For a given set of data, the maximum likelihood estimation optimizes the value of \( \theta \) and then the correlation model is brought into the regression model to evaluate the function with the best linear unbiased predictor [5].

B. Utility Functions

The ‘expected improvement’ utility function [6], based on a potential error—the ‘Mean Square Error’ (MSE) produced by the kriging model—is commonly used to select multiple designs for further evaluation. Here the optimization problem is viewed as minimizing the objective function. The Expected Improvement (EI) function is defined as

\[
EI = \begin{cases} 
\frac{f_{\text{max}} - \hat{y}(x)}{\sigma(x)} + \phi \left( \frac{f_{\text{max}} - \hat{y}(x)}{\sigma(x)} \right) 
\sigma(x) & \text{if } \sigma(x) > 0 \\
0 & \text{if } \sigma(x) = 0
\end{cases}
\]

where \( \hat{y}(x) \) is the value of objective function predicted by the kriging model, and \( \sigma(x) \) is the root mean squared error in this prediction. The first term of (3), including Gaussian density, favors “exploration”—searching the most promising regions (high confidence); while the second term, containing the Gaussian distribution, favors “exploitation”—searching the regions that have high uncertainty. It has been found [2] that in practical cases exploration performs dramatically better in terms of finding the global optimum, whereas exploitation often causes the kriging model to stop around a local minimum. The termination criteria used by the kriging model are often based on finding repeatedly the same sampling point within the prescribed tolerance, thus balancing exploration and exploitation is vital. The EI algorithm distributes the weights equally between the two terms and can be seen as a fixed compromise between exploration and exploitation. In order to study the effect of different weights, the Weighted Expected Improvement (WEI) [6] was derived from EI by adding a tuneable weighting parameter. Through a set of experiments it was shown that by changing the value of the tuneable parameter the efficiency of finding the global minimum can be affected. However, the optimal value of the weight is hard to find and usually problem dependent. Hence a modified strategy is required to design a more intelligent method which could guide itself automatically through the whole sequential process using an iterative process.

C. Adaptive Weighted Expected Improvement With Reward Approach (AWEI)

Two algorithms inspired by reinforcement learning [7], the Adaptive Weighted Expected Improvement (AWEI) and the Surrogate Model based Weighted Expected Improvement-based approach utilizing the concept of rewards [2] (SMWEI), were proposed previously [1] to make the process of tuning weights
Appendix C Journal Papers

Fig. 1. The decision-making chart for balancing exploration and exploitation using AWEI algorithm.

more intelligent and self-guiding. The decision-making chart of AWEI is shown in Fig. 1. Compared with the original [1] this version is enhanced by the addition of a pre-test stage during which a "cheap" simplified surrogate model is applied built on a specific prediction and the potential error produced by kriging. A pair of fixed weights (one emphasizing exploration and the other exploitation) are initially set at equal values. The Mean Square Error from the kriging model is used to calculate the rewards. After comparing the rewards obtained from the pre-test, the weights are redistributed on the terms favoring exploration or exploitation. The weight with a better reward of the two is then used to feed back—via the FEM module—into the main iterative loop of the design process.

Although finding the global optimum is often sufficient, this may not be enough when a robust design is considered, as information about the shape of the objective function and position and shape of local minima is also relevant. The quality of the approximation of the objective function is the most important prerequisite when the robustness of the solution is based on a kriging model rather than the real function. In the next section the performance of the AWEI method will be assessed in the context of finding the global optimum concurrently with obtaining a good quality function prediction for the search space of interest.

III. NUMERICAL EXPERIMENTS

To explain the concepts we use a simple single-variable function shown graphically in Fig. 2. There are two local minima (B and C) and one global minimum (A). When considering practical factors such as uncertainty of the variables due to manufacturing tolerances, non-uniform material properties or imperfect control of operating conditions, the theoretically best-performing point A may not be a practical optimal solution. The final judgment will be influenced by the margin of the uncertainty and how much the objective function may change when the design variables assume limiting values. For a minor uncertainty, even under extreme (but still small) departure of the parameter from its nominal value, the objective function may still be acceptable—even if worse than at the theoretical “best”—and thus even a “sharp” optimum like A may be judged as practical. But under increased uncertainty the confidence in the final performance may be assured only by a much more “shallow” minimum, such as B, or even C. Thus the preferred practical design may be selected away from the theoretical global optimum.

Various methods of assessing the robustness have been proposed in literature. In [9], the initial optimization problem is transformed into a multi-objective optimization where both the objective function and the gradient index of this function are minimized simultaneously. We take a similar approach; however, we calculate the gradient index using the predicted value of the objective function rather than the function itself. The motivation behind such an approach is to reduce the computational effort by avoiding the costly FEM (or similar) numerical solution. For this methodology to work it is very important to have good prediction for the objective function. The quality of the prediction will now be investigated, initially with the aid of a popular test function [8], [9] defined as

\[
f(x) = 10 - \sum_{i=1}^{n} \left[ \frac{3.5}{1 + (x_i - 5)^2} + \frac{2.2}{1 + (x_i - 15)^2} + \frac{1.1}{1 + (x_i - 75)^2/30} \right]
\]

in the range \(5.6 \leq x_i \leq 27\). The two-variables \((n = 2)\) version is plotted in Fig. 3 and will be used to assess AWEI and compare its performance against standard EI method.
Appendix C Journal Papers

Fig. 4. Kriging assisted EI: (a) function after 324 iterations, (b) error.

For both the EI and AWEI tests, seven initial sampling points were used (as in Table I). Kriging assisted EI needs 324 iterations to find the global minimum and the predicted shape is demonstrated in Fig. 4(a), with the error calculated against the actual objective function shows in Fig. 4(b). The AWEI strategy is more efficient in finding the global minimum, requiring 211 iterations (Fig. 5), but the final approximation is not as good as that resulting from EI as the number of available points is less. Nonetheless, in some regions, e.g., $5 \leq x_2 \leq 27$, $-2.3 \leq x_1 \leq 2.5$ or $x_1 \leq x_2 \leq 7.5$, $5 \leq x_1 \leq 7.5$, AWEI outperforms EI (Fig. 5(b)).

IV. RESULTS FOR TEAM WORKSHOP PROBLEM 22

The full description of the TEAM 22 Benchmark problem (superconducting magnetic energy storage system) may be found in [10]. The goal of the optimization task is to achieve the stored energy of $E = 180$ MeJ with a minimal stray field $B_{stray}$. The objective function is defined as

$$\text{OF} = \frac{B_{stray}^2}{B_{red}^2} = \frac{|E - E_{ref}|}{E_{ref}}$$

(6)

where $B_{stray} = B_2 \mu T$ and $B_{red}^2 = (\sum_{j=1}^{22} (B_{stray,j})^2)/(22)$, subject to some geometrical and "quench" constraints. We consider here the 3 parameter problem, although initially—for the purpose of demonstrating typical shapes of the objective function—one of the variables is fixed ($B_2 = 3.08$ m), while $b_2$ and $h_2$ varied. One of the challenges in creating the kriging correlation matrix is the "combinatorial explosion" when many parameters are used [11] and we propose to use a successive "zoom in" strategy to cope with this issue; a two-stage version implemented in the context of the TEAM Problem 22 is described in Table II.

The initial sampling points would normally be selected using the Latin Hypercube [12]; in our tests we fixed the positions at

Fig. 5. Kriging assisted AWEI: (a) function after 211 iterations, (b) error.

<table>
<thead>
<tr>
<th>TABLE I</th>
<th>INITIAL SAMPLING POINTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>$x_2$</td>
</tr>
<tr>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>10</td>
<td>19</td>
</tr>
<tr>
<td>7</td>
<td>19</td>
</tr>
<tr>
<td>15</td>
<td>7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE II</th>
<th>TWO-STAGE STRATEGY USING KRIJING ASSISTED EI FOR TEAM 22 TESTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td>The first stage: $b_2 = 0.188$, $a_2 = 0.35$</td>
</tr>
<tr>
<td></td>
<td>The second stage: $b_2 = 0.188$, $a_2 = 0.35$</td>
</tr>
<tr>
<td>Test range</td>
<td>Min: 0.108, 0.1, 0.4, 0.65, 0.4</td>
</tr>
<tr>
<td></td>
<td>Max: 0.776, 0.4, 0.4</td>
</tr>
</tbody>
</table>

Fig. 6. Two-parameter SMES problem tested by kriging with EI.

($h_2 = 0.741$ m, $d_2 = 0.13$ m), ($h_2 = 1.304$, $d_2 = 0.23$ m)
($h_2 = 1.64$ m, $d_2 = 0.40$ m) and ($h_2 = 2.088$ m, $d_2 = 0.37$ m)

to facilitate comparisons. Kriging with EI required 25 iterations to complete the first "rough" stage, and further 20 iterations to
find the best-performing solution \((b_9 = 0.478\ m, d_9 = 0.395\ m, \Omega^* = 0.0874)\) (Fig. 6). For kriging assisted AWEI the relevant iteration numbers are 21 and 23 (Fig. 7). The "history" of the EI and AWEI strategies may be followed on the two figures.

The initial settings for the 3 parameter test are presented in Table III. For kriging with EI, 79 sampling points were created in the first stage, with further 132 points in the zoom-in region; in the AWEI case, 156 sampling points were followed by 167 points in the second stage (Fig. 8). The total numbers quoted include the initial sampling points needed by the kriging model. It is no longer possible to conveniently display the objective function itself for three parameters.

To demonstrate the advantages of the proposed methods over other well-known stochastic algorithms, Table IV has been compiled using available published data [19]. It is clear that kriging assisted EI and AWEI both offer significant advantages in terms of much reduced number of computationally expensive function calls to achieve required level of accuracy. However, those gains are somewhat offset by the need to create the kriging correlation matrix. Finally, the robustness of the design requires further consideration and will be studied in the context of more relevant test cases. A particular challenge is to select test functions representative of the problems encountered in electromagnetic design, although the prime concern remains to keep the number of function calls low.

V. CONCLUSION

Exploration and exploitation in the design algorithm are controlled via a kriging model, the Weighted Expected Improvement strategy and the use of "rewards", with robust design also considered. To avoid problems associated with combinatorial explosion a "zoom in" strategy is proposed.

REFERENCES


Robust Design Optimisation of Electromagnetic Devices
Exploiting Gradient Indices and Kriging

S. Xiao, M. Rotaru and J. K. Sykulski
Electronics and Computer Science (ECS), University of Southampton, Southampton, SO17 1BJ, UK

Key words: Design optimisation, Pareto optimisation, Kriging, TEAM Workshop Problem 22

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 concept of the Gradient Index (GI) has been explored in [1]-[3]. The method transforms a problem into a multi-objective optimisation by concurrently minimising the function and its gradient index, 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 significant further advances both in terms of the way in which the gradient indices are used but also by improving the computational efficiency of the algorithms. The notions of the gradient index sensitivity and the second-order gradient index 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 [4] is employed. In other words, the objective function is approximated using kriging [5], assisted by algorithms balancing exploration and exploitation ([6], [7]) using the concept of rewards [8]. 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 constraints. 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 [1]-[3] (see Fig. 1)

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

(a)
Fig. 1. Example of a robust design for a one-variable problem
(a) Objective function, the gradient index and sensitivity,
(b) First and second order gradients,
(c) Objective function trajectory showing sensitivity and objective function values (OF).

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

$$U(x) = \{ x \in \mathbb{R} | x - \delta \leq x \leq x + \delta \}$$

where $\sigma$ is standard deviation of uncertain variables and $\delta$ is determined by a confidence level [2].
One way of incorporating robustness into the mainstream optimisation process is by adding the gradient index \([1]\) as a second objective and formulating the problem as

\[
\text{Minimise } f(x) \quad x \in \mathbb{R}^n \quad (x_L \leq x \leq x_U)
\]

\[
\text{Minimise } G_l(x) = \max_{i = 1,...,m} |\partial f(x)/\partial x_i|
\]

(3)

Subject to \[g_i(x) \leq 0, \quad i = 1,...,m\]

Point A1 in Fig. 1(a) 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. 1(a) and the values of the second-order gradient in Fig. 1(b), 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. 1(a) 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. 1(c). 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 gradient index. 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. 1(a)-(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. 2(b) – 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. 2(b).

<table>
<thead>
<tr>
<th>Coordinates on x axis</th>
<th>Objective function value</th>
<th>Second gradient value</th>
<th>Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>5</td>
<td>6.21628</td>
<td>6.702025</td>
</tr>
<tr>
<td>A2</td>
<td>7.8</td>
<td>6.735</td>
<td>0.5</td>
</tr>
<tr>
<td>A3</td>
<td>15.1</td>
<td>7.48699</td>
<td>0.42475</td>
</tr>
</tbody>
</table>

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

3. Robust optimisation based on kriging modelling

3.1 Kriging

As a type of regression model, kriging [6] 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), in order to build a correct surrogate model of the real objective function through interpolation. This relies on the linear regression model

$$y(x) = \sum_{k=1}^{n} \beta_k f_k(x) + \epsilon(x)$$

(4)

and the Gaussian correlation model

$$R(e(x^i), e(x^j)) = \prod_{k=1}^{n} e^{-\frac{|x^i_k - x^j_k|^2}{\theta_k^2}}.$$ 

(5)

where the global function $\sum_{k=1}^{n} \beta_k f_k(x)$ and an additive Gaussian noise $\epsilon(x)$ are integrated to the predicted value $\hat{y}(x)$ of the objective function; the hyperparameter $\theta_k$ is the correlation amongst the data in $k$-direction and $\rho_k$ determines the 'smoothness' of (5). The most popular correlation function is given by the Gaussian model where the value of $\rho_k$ is simply taken as equal to 2. For a given set of data, the maximum likelihood estimation optimizes the value of $\theta$ and then the correlation model is brought into the regression model to evaluate the function with the best linear unbiased predictor (5), [9].

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' strategy to alleviate the problem, where – in order to reduce the amount of data storage and utilize the installed physical memory capacity efficiently – the step sizes of design vectors 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.

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. 2; 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 in order 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.

Fig. 3. The correlation matrices partitioned by: (a) sampling points, (b) design vectors, (where S1, S2, …, S6 sampling points).
To verify the viability of the proposed methodology, the TEAM problem 22 has been attempted [10]. Without the strategy of Fig. 4, the kriging algorithm failed due to the memory size problem when tested on a smaller computer.

4. Application in 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 [10]. The target is to achieve an arrangement of the two superconducting coils such that the stored energy within the system is $E_{\text{sta}} = 1800$ MJ while a minimal stray field $B_{\text{str}}$ is obtained. The objective function is defined as

$$OF = \frac{B_{\text{str}}}{B_{\text{tori}}} + \frac{|E_{\text{ref}}|}{E_{\text{ref}}},$$

(6)

where $B_{\text{str}} = 3p$ T and $B_{\text{tori}}^2 = \frac{2L_1 L_2 |B_{\text{str}}|}{22}$, 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 ($D_2 = 0.394m$), while $B_2$ and $K_2$ are varied. Usually the initial sampling points are selected using Latin Hypercube [12]; however, 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) [6], [7], while Fig. 5 demonstrates the convergence process of AWEI.
Table 2. The setting of initial sampling points.

<table>
<thead>
<tr>
<th>$R_2$</th>
<th>2.7</th>
<th>2.9</th>
<th>3.0</th>
<th>3.3</th>
<th>3.4</th>
<th>2.6</th>
<th>3.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_2$</td>
<td>0.744</td>
<td>1.304</td>
<td>1.64</td>
<td>2.088</td>
<td>0.408</td>
<td>0.408</td>
<td>2.2</td>
</tr>
</tbody>
</table>

Table 3. Performance comparison of algorithms: Radial basis functions (RBF) [11]; AWEI (Kriging) [7].

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>$R_{m}$</th>
<th>$h_{m}$</th>
<th>Best OF</th>
<th>Iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>RBF</td>
<td>3.06</td>
<td>0.236</td>
<td>0.088</td>
<td>246</td>
</tr>
<tr>
<td>AWEI (Kriging)</td>
<td>3.08</td>
<td>0.239</td>
<td>0.089</td>
<td>38</td>
</tr>
</tbody>
</table>

Fig. 5: TEAM Workshop Problem 22
(a) Prediction by kriging with $\text{AWEI (3 \leq R_2 \leq 3.3, 0.408 \leq h_2 \leq 0.5, D_2=0.394, other parameters fixed)}$.
(b) Sensitivity with respect to $R_2$ and $h_2$.
(c) Objective function trajectory showing sensitivity and objective function values (OF).

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

IET Review Copy Only
4.5 Three-parameter Test Results

The full 3 parameter TEAM 22 problem [10] 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 due to 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, more information provided in Table 5.

<table>
<thead>
<tr>
<th>Sample 1</th>
<th>$R_2$ (m)</th>
<th>$h_2$ (m)</th>
<th>$d_2$ (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 2</td>
<td>3</td>
<td>0.408</td>
<td>0.3</td>
</tr>
<tr>
<td>Sample 3</td>
<td>3.2</td>
<td>0.744</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Table 4. The setting of initial sampling points.

<table>
<thead>
<tr>
<th>Test range</th>
<th>$R_2$ (m)</th>
<th>$h_2$ (m)</th>
<th>$d_2$ (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step size</td>
<td>0.01</td>
<td>0.014</td>
<td>0.003</td>
</tr>
<tr>
<td>Number of steps</td>
<td>81</td>
<td>129</td>
<td>101</td>
</tr>
</tbody>
</table>

Table 5. The specific definition of test.

A comparison between the results obtained with the ‘zoom in’ strategy and results using ‘the adaptive correlation matrices division’ for this example are presented in Fig. 6. With the ‘zoom in’ strategy, the kriging assisted by 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 specific range of each test stage. Via applying the scheme of partitioning the correlation matrices, the issue of combinatorial explosion is overcome throughout the process of infilling the sampling points. 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). But 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’.
Fig. 6. The 3 parameter TEAM 22 problem (a) EI, (b) AWEL.
Appendix C Journal Papers

IET Science, Measurement & Technology

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

The uncertainties for $R_2$, $h_2$ and $d_2$ have been pre-set as $R_2 = 0.03 \times \xi (R_2) < R_2 + 0.03$, $R_2 = 0.042 < \xi (R_2) < R_2 + 0.042$, $d_2 = 0.009 < \xi (d_2) < d_2 + 0.009$. The differences $\Delta(x)$ 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 7b, 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.

Fig. 8. The specific parameter setting and transient process of testing the TEAM 22 problem with the help of the novel partitioning scheme.

IET Review Copy Only
In order 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 memory occupied by the optimiser or limiting the time spent on 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. In the particular case, the testing environment provided approximately 11 GB available memory, which was set as the triggering value. However, due to 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 11GB memory limit had not 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.

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 predictions 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.

References
Appendix C Journal Papers


Title:
Considerations of uncertainty in robust optimisation of electromagnetic devices

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

Affiliation:
Electronics and Computer Science (ECS), University of Southampton, Southampton, SO17 1BJ, UK

Corresponding author:
Professor Jan K. Sykulski
Electronics and Computer Science (ECS)
University of Southampton
Southampton
SO17 1BJ
United Kingdom

Tel: +44 2380 593448
Fax: +44 2380 593709
E-mail: jks@soton.ac.uk

Abstract – Due to unavoidable uncertainties related to material properties and manufacturing processes, the robustness of the optimal solution must be considered when designing electromagnetic devices. In this paper, the worst-case optimisation (WCO) and the worst-vertex-based WCO are proposed to evaluate the robustness of both performance and constraints under uncertainty. To reduce computing times when searching for the robust solution a predicted objective function is used, obtained with the help of a kriging algorithm which explores the searching space using the concept of rewards. Finally, to avoid some of the shortcomings of WCO, the concept of average performance evaluation is developed.

Keywords: robust optimisation, CAD in electromagnetics, kriging.
Appendix C Journal Papers

1. Introduction

In electromagnetic design, uncertainties in design variables are inevitable, thus the ability to evaluate the robustness is critical while pursuing the theoretical optimum. This is particularly true when constrained optimization is considered, as illustrated later by an example involving a quenching condition for a superconducting material, in order to maintain the solutions within the feasible region when perturbation occurs. The worst-case optimization (WCO) method has been selected to evaluate the accuracy of the prediction of the objective function provided by an improved kriging model, implemented for the sake of reducing computational effort associated with direct application of information produced by a time-consuming Finite Element Method (FEM). However, in this study some shortcomings of the WCO approach have been identified associated with the inability to assess the performance variation under conditions of uncertainty. Therefore the concept of average performance evaluation was suggested as an improved measure of robustness.

2. Robust Optimization Algorithms exploiting kriging modelling

In general, seeking the minimum (maximum) of an objective function, while the search space is restricted by certain constraints, is the aim of conventional optimization

\[
\text{Minimize } f(x) \quad \text{Subject to } g_i(x) \leq 0, \quad i = 1, \ldots, m
\]  

(1)

When designing real devices, however, many design variables are subject to specific uncertainties (manufacturing tolerances, variation of material properties, etc.). Hence the assessment of the influence caused by these uncertainties on the performance becomes essential in practical design problems. As a consequence, finding a theoretical optimum may not be sufficient and the robustness of the theoretical optimal solution needs to be considered as well. Algorithms applying different strategies to evaluate robustness, such as the sensitivity analysis [1], the worst-case optimization method [2-6], and the mean value and variance of performance [7,9], have been developed to assist in the design tasks (see also [19]). In our previous work [10] a multi-objective optimization method, which included sensitivity analysis using gradient index, was developed and demonstrated.

A widely used approach to evaluate the reliability of a robust solution is the worst-case method. This technique can maintain a certain level of robustness by avoiding solutions that may push the function into unfeasible region when searching for the optimum, as shown in Fig. 1(a). A set of typical examples is shown in Fig. 1(a) to illustrate the principle of the worst-case optimization technique. The theoretical optimum A, for example, may be abandoned in favour of a ‘worse’ solution B because the actual design A’, which accounts for the uncertainties of the variables, might violate constraints and enter unfeasible area. The uncertainties can be defined mathematically as

\[
U(x) = \{ \xi \in \mathbb{R}^n | x - k\sigma \leq \xi \leq x + k\sigma \}
\]  

(2)

where \(\sigma\) is standard deviation of uncertain variables and \(k\) is determined by confidence level [11]. The uncertainties may also be specified directly (e.g., as machining tolerances, say \(\Lambda\)). As an algorithm which can
Appendix C Journal Papers

predict the worst scenario considering the uncertainties, as well as constraints with respect to specific designs, the worst case optimization (WCO) method [2-5] may be applied to analyse the reliability of the solution as follows

\[
\text{Minimize } f_{\text{w}}(x) = \max_{\xi \in \Omega(x)} f(\xi) \quad \text{Subject to } g_{\text{w}}(x) = \max_{\xi \in \Omega(x)} g(x) \leq 0, i = 1, \ldots, m
\]  

(3)

The worst values of the objective function and the \(i\)-th constraint function are chosen to substitute the original values of the nominal design \(x\).

Take it in Fig. 1.

Numerical methods, such as finite elements, are often used when searching for the worst objective function value under imposed constraints which may be an extremely time-consuming process. To reduce the computational burden the worst-vertex-based WCO (W-WCO) [4] was proposed; this algorithm only needs to observe the vertices within the region restricted by uncertainties rather than evaluating every design value. For example, in the problem illustrated in Fig. 1(b), in addition to \(x\) there are 8 more points required, located at the corners and the middle of the specified boundary. However, in certain cases assessing only these 8 points might still not be sufficient. Fig. 1(c) illustrates such a case where a large variation of the function will not be identified by the W-WCO method. Therefore finding a balance between mitigating the heavy computational burden of the original worst-case method and pursuing more detailed evaluation is the main issue to be addressed.

One of the possibilities is to employ the improved kriging method, assisted by a set of strategies capable of balancing exploration and exploitation [12, 13, 17] using the concept of rewards [14], which can be used to predict the objective function value instead of directly calculating it using computationally expensive FEM models. Based on such a kriging prediction, the worst-case method can be directly implemented. In other words, the WCO method uses the predicted information rather than the expensive FEM models. The accuracy of the predicted objective function using the improved kriging model has been considered in [10]. In this paper the suitability of directly using WCO with the predicted function model is discussed and demonstrated.

3. Kriging surrogate model

As a kind of regression model, kriging [15] is able to predict the shape of the objective function via spatial correlation of data using limited information. The accuracy of this prediction can also 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 optimization 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), in order to build a correct surrogate model of the real objective function through interpolation. This relies on the linear regression model

---

236
and the Gaussian correlation model

\[
\hat{y}(x) = \sum_{k=1}^{m} \beta_k f_k(x) + \epsilon(x)
\]

(4)

where the global function \(\sum_{k=1}^{m} \beta_k f_k(x)\) and an additive Gaussian noise \(\epsilon(x)\) are integrated to the predicted value \(\hat{y}(x)\) of the objective function; \(\theta_i\) is the correlation amongst the data in \(k\)-direction and \(\rho_i\) determines the ‘smoothness’ of (5). The most popular correlation function is given by the Gaussian model where the value of \(\rho_i\) is simply taken as equal to 2. For a given set of data, the maximum likelihood estimation optimizes the value of \(\theta\) and then the correlation model is brought into the regression model to evaluate the function with the best linear unbiased predictor [4].

Although kriging can potentially solve large multi-parameter optimization problems, it has some inherent limitations making the implementation difficult. In particular, for multi-parameter problems, the correlation model built by the kriging algorithm can grow very fast resulting in a ‘combinatorial explosion’ of correlation data filling very quickly the memory of standard computer workstations. As a result, the process can become slow and inefficient. To solve this issue a scheme that adaptively partitions the correlation matrices was developed [10]. Using this approach the size of the data is managed to use efficiently the available memory throughout the iterative process of kriging. The scheme mentioned above has another advantage as the kriging predictor and Mean Squared Error are being calculated at the same time hence more computational time is saved in the process of building the kriging model. Therefore, this modification makes kriging suitable for solving multi-parameter optimization problems and could be linked with WCO which needs detailed data to work effectively.

4. Average performance

As explained in the previous sections, the WCO method can be used to find robust solutions for a particular problem once the constrains and the uncertainties of the variables have been defined. In this section, however, we address the extreme case depicted in Fig. 2; it is argued that WCO on its own is not sufficient to find a reliable and robust solution. To deal with such a situation the concept of an average performance has been suggested and will now be explained. Figures 2(a) and 2(b) show two similar functions, otherwise identical, except the region around the points A2 and A4, respectively. If WCO is used to find a robust solution using the same uncertainties \(\Delta\) for the case depicted in Fig. 2(a), point A2 will be found, whereas for the case shown in Fig. 2(b) point A4 is likely to be returned as a robust solution. It is clear from this example that A4 is by no means a robust solution and although this may be considered a somewhat extreme situation it illustrates the fact that the WCO algorithm cannot differentiate between a robust and non-robust solution in such cases. To resolve such problems the concept of average performance within the uncertainty range is introduced.
The initial idea of evaluating average performance was to simply calculate the average value of all the potential perturbed values in the uncertain region with respect to the solution. This, however, proved infeasible, as although the shape of the two objective functions is clearly different, they share some common characteristics as explained in Table 1.

Take in Fig. 2.

Take in Table 1.

By visual inspection we can see that the variation around the solution A4 is much more intense than around A2, but the average values within the shaded areas with middle-points A2 and A4 for the given uncertainty are the same and equal to 2.9556. This means that the average value criterion may not be useful in assessing the variation of the function close to the point of interest. Hence an average value of the gradient index (GI) [16] has been introduced as an alternative way of assessing the average performance. The average value of the gradient index is calculated as:

\[
\text{Minimize } \quad \text{Average } GI(x) = \frac{1}{n} \sum_{x_i \in x_L} \max \left| \frac{\partial f(x)}{\partial x_i} \right|/n
\]

where \(x_i \in R^2 \quad (x_L \leq x_i \leq x_U)\) is the \(i\)-dimensional design variable vector with lower and upper bounds \(x_L\) and \(x_U\), respectively, and \(n\) is the sum of the design vectors. The sum of the maximum gradients is divided by the total number of design vectors. The average first-order gradient for the first case is then found to be 0.0594, while in the second case it is three times larger 0.1416. Therefore the average gradient index could be used as a more reliable criterion to evaluate the average performance. This criterion can therefore be combined with the WCO method to resolve difficult problems such as the one described by Fig. 2(b). By generalizing this methodology it can be argued that a robust optimization problem can be transformed into a three-objective optimization problem defined as

\[
\begin{align*}
\min f(x) & \\
\text{Minimize } f_w(x) = \max_{\beta \in \Omega} f(\xi), \quad & \text{Subject to } g_w(x) = \max_{\beta \in \Omega} g(x) \leq 0, i = 1, \ldots, m
\end{align*}
\]

\[
\text{Minimize } \quad \text{Average } GI(x) = \frac{1}{n} \sum_{x_i \in x_L} \max \left| \frac{\partial f(x)}{\partial x_i} \right|/n
\]

5. Robust Optimization Algorithms exploiting kriging modelling

In our previous work [10] the robustness of an optimal solution was evaluated using the gradient index, where the task of robust optimization was transformed into a two-objective optimization. One objective was to minimize the difference between the absolute value of the largest and the smallest gradients within the uncertain range, called the 'sensitivity', while minimizing the objective function that was the second objective. Although the sensitivity calculated by the gradient index method is able to provide information on the rate of change of the objective function, the WCO method can also be employed to obtain similar
information. However, as shown in the previous section, the WCO method has some limitations, especially for extreme cases (Fig. 2(b)). The average performance assessment described above can thus be added to the WCO method to improve the overall reliability of the result. To verify the concept and to analyse further the average performance criterion, two problems have been tested and the results are reported below.

Take in Fig. 3.

First, the two-variable analytic function (8) depicted in Fig. 3(a), which was also used to assess the accuracy of the improved kriging model, has been tested.

\[
    f(x) = 10 - \sum_{i=1}^{n} \left[ \frac{1.5}{1 + (x_2 - y)^2} + \frac{1.2}{1 + (x_2 - y)^2 / 10} + \frac{1.2}{1 + (x_2 - y)^2 / 30} \right] (0 < x_2 < 27)
\]  

(8)

Kriging with adaptive weighted expected improvement (AWEI) [16] has provided an approximation of this analytic function with a test step size of 0.1. Within 85 iterations, the kriging model can find the global minimum successfully; however – for a better approximation of the shape of the objective function – the model continued to run until the 285\textsuperscript{th} iteration (Fig. 3(b)).

The uncertainty with respect to variables is set as \(U(x_1, x_2) = 0.5\), and the WCO method is used to obtain the surface of the worst case for each solution as shown in Fig. 3(c). For this case the “best” solution shifts from the theoretical optimum \((x_1=5, x_2=3, y=2.443)\) to the location \((x_1=5.1, x_2=5.1, y_{\text{worst case}}=3.7639)\) which provides a more robust result, for the given conditions. If the values of the uncertainties were to keep increasing, to a certain extent, ultimately the robust optimum would thoroughly shift from the sharp global minimum to one of the preferable local minima with higher robustness. Figure 3(d) depicts the average gradient index values in the search space. Finally, the full-scale \((0 < x_1 < 27)\) optimal solutions including all three objectives have been presented in Fig. 3(e). For clearer presentation of the pareto front, the full-scale version is zoomed in Fig. 3(f). Two typical pareto solutions are labelled in the zoomed-in graph: solution N1 delivers a more optimal value of the prediction of the objective value, while N2 offers a relatively better average performance.

6. Application to Electromagnetic Design

The second example tested with the proposed WCO procedure involves a multi-objective version of the TEAM 22 benchmark problem [18]. The full description of the TEAM problem 22 may be found elsewhere and will not be repeated here. The target for this problem is to achieve an arrangement of the two superconducting coils such that the stored energy within the system is \(E_{\text{ref}}=180\text{MJ}\) while a minimal stray field \(B_{\text{mxy}}\) is maintained. The objective function is defined as

\[
    \text{OF} = \frac{B_{\text{mxy}}}{B_{\text{mref}}} + \frac{E-E_{\text{ref}}}{E_{\text{ref}}}
\]

(9)

where \(B_{\text{mxy}} = 3\text{pT}\) and \(B_{\text{mref}} = \frac{20^{16}B_{\text{max}}}{E_{\text{ref}}}\), subject to geometrical and ‘quench’ constraints. The approach taken here combines WCO method with kriging and commercially available FEM based software. A 2D model of the TEAM 22 problems is solved throughout this procedure. The three parameter case of TEAM
Appendix C Journal Papers

22, which includes three geometric variables $R_s$, $H_i$ and $D_n$, while $R_e$, $H_f$ and $D_l$ are fixed, has been tried under different settings of uncertainties. The uncertainties are assumed to exist in the current densities $J_i(J_e, J_f)$ of the two coils, because normally they are limited within certain range by a current controller for compensating perturbation. A constraint is imposed that the superconducting coils should not violate the quench condition which links together the value of the current density and the maximum value of magnetic flux density as follows

$$g_i(x) = |J_i| + 6.4 \cdot |B_{ni}| - 54.0 \leq 0, \quad i = 1, 2 \quad (10)$$

Three tests were performed for this TEAM 22 problem. The initial data set-up, as well as the number of steps, for the three tests is listed in Table 2. One of the differences between the three sets is the size of the uncertainty; the first set had the uncertainty set as 0.1, the second test 0.2, while in the third test 0.35 was used. Another difference was in the number of iterations that were used to approximate the objective function by the kriging model with AWEI [10]. For the first data set (uncertainty 0.1) kriging generated 185 sampling points, for the second case (uncertainty 0.2) kriging produced 257 sampling points, while for the third set (uncertainty 0.35) only 188 points were necessary. The results returned by the WCO, coupled with kriging, for these three cases are summarised in Tables 3 to 5. Figures 4, 5 and 6 show both the full scale and zoomed-in versions of the pareto fronts obtained for the three cases. In the figures and tables, solutions $P_1$, $P_4$ and $P_7$ refer to the global optimum; $P_3$ and $P_6$ have the ‘worst case’ performance; while $P_2$, $P_5$ and $P_8$ describe the best average gradient index solutions.

Take in Table 2.

Take in Table 3.

Take in Fig. 4.

Take in Table 4.

Take in Fig. 5.

Take in Table 5.

Take in Fig. 6.

Unlike other reported work [16] that uses a stochastic optimization method to find the global optimum and then employs Monte Carlo method to explore the space around the global minimum, combined with WCO and the gradient index to judge the robustness of the solution, the method introduced here takes a holistic approach and explores the whole searching space. The kriging model allows comparison amongst several local minima (maxima) that may be more robust than the global optimum. Another major advantage of the
procedure proposed in this work is the fact that it can be linked with any commercial electromagnetic design software giving more freedom to the designer.

7. Conclusions

The worst-case method (WCO) assisted by the prediction provided by the kriging model with AWEm has been proposed to solve robust optimization problems considering uncertainties of variables. The particular contribution of this paper is in using kriging prediction, rather than the computationally expensive finite element modelling, in assessing the robustness of the final design. The second contribution involves enhancing the worst case methodology through introducing the concept of average gradient index performance. Using this approach a conventional optimization problem, with constraints and uncertainties in variables, has been transformed into three-objective optimization with a relevant pareto front. The proposed algorithms have been verified by both numerical tests and a practical electromagnetic design problem described by TEAM 22 benchmark.

References


---

**Tables**

**Table 1**

<table>
<thead>
<tr>
<th>Global minimum (A1 and A3)</th>
<th>Local minimum (A2 and A4)</th>
<th>Uncertainty Δ</th>
<th>Worst case (A1 and A3)</th>
<th>Worst case (A2 and A4)</th>
<th>Average value (A2 and A4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>x=18, y=1.0216</td>
<td>x=55, y=2.12657</td>
<td>15</td>
<td>x=3, y=6.557</td>
<td>x=55, y=2.12657</td>
<td>2.9556</td>
</tr>
<tr>
<td>y=2.12657</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 2
The initial setup of the prediction by the kriging model

<table>
<thead>
<tr>
<th></th>
<th>Three variables</th>
<th>Uncertainties</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$R_1$ (m)</td>
<td>$h_z/2$ (m)</td>
</tr>
<tr>
<td><strong>Lower</strong></td>
<td>3.03</td>
<td>0.211</td>
</tr>
<tr>
<td><strong>bound</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Upper</strong></td>
<td>3.13</td>
<td>0.281</td>
</tr>
<tr>
<td><strong>bound</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Step size</strong></td>
<td>0.01</td>
<td>0.007</td>
</tr>
<tr>
<td><strong>No of steps</strong></td>
<td>11</td>
<td>11</td>
</tr>
</tbody>
</table>

### Table 3
Results for the case when uncertainty is 0.1

<table>
<thead>
<tr>
<th></th>
<th>$R_1$ (m)</th>
<th>$h_z/2$ (m)</th>
<th>$d_z$ (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>P1</strong></td>
<td>3.04</td>
<td>0.492</td>
<td>0.397</td>
</tr>
<tr>
<td><strong>P2</strong></td>
<td>3.09</td>
<td>0.464</td>
<td>0.382</td>
</tr>
<tr>
<td><strong>P3</strong></td>
<td>3.03</td>
<td>0.562</td>
<td>0.397</td>
</tr>
</tbody>
</table>

### Table 4
Results for the case when uncertainty is 0.2

<table>
<thead>
<tr>
<th></th>
<th>$R_1$ (m)</th>
<th>$h_z/2$ (m)</th>
<th>$d_z$ (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>P4</strong></td>
<td>3.11</td>
<td>0.492</td>
<td>0.376</td>
</tr>
<tr>
<td><strong>P5</strong></td>
<td>3.09</td>
<td>0.478</td>
<td>0.385</td>
</tr>
<tr>
<td><strong>P6</strong></td>
<td>3.03</td>
<td>0.562</td>
<td>0.397</td>
</tr>
</tbody>
</table>

### Table 5
Results for the case when uncertainty is 0.35

<table>
<thead>
<tr>
<th></th>
<th>$R_1$ (m)</th>
<th>$h_z/2$ (m)</th>
<th>$d_z$ (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>P7</strong></td>
<td>3.07</td>
<td>0.492</td>
<td>0.382</td>
</tr>
<tr>
<td><strong>P8</strong></td>
<td>3.03</td>
<td>0.562</td>
<td>0.397</td>
</tr>
</tbody>
</table>
Appendix C Journal Papers

Figure captions

Fig. 1.
(a) The constrained optimization problem
(b) The worst case optimization (WCO)
(c) The worst-vertex-based WCO

Fig. 2.
(a) The objective function with a robust local minimum subject to uncertainty $\Delta$
(b) The objective function with a non-robust local minimum subject to uncertainty $\Delta$

Fig. 3.
(a) Analytic function
(b) The kriging prediction
(c) The performance of the WCO method
(d) Average performance
(e) The full-field solution for three objectives
(f) The zoomed-in optimal part

Fig. 4.
The three objective optimization including
(Worst case (WC), Average gradient index performance (AVGI), the prediction of objective functions (Ob))
$U(J_1, J_2) = 0.1$

Fig. 5.
The three objective optimization ($U(J_1, J_2) = 0.2$)

Fig. 6.
The three objective optimization ($U(J_1, J_2) = 0.35$)
Figures

Fig. 1. a

Fig. 1. b

Fig. 1. c
Fig. 2. a

Fig. 2. b

Fig. 3. a
Appendix C Journal Papers

Fig. 3. b

Fig. 3. c

Fig. 3. d
Fig. 3. c

Fig. 3. f
Appendix C Journal Papers

Fig. 4.

Fig. 5.
Appendix C Journal Papers

Fig. 6.
Six Sigma Quality Approach to Robust Optimisation

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

Electronics and Computer Science, University of Southampton, Southampton, SO17 1BJ, UK

In electromagnetic design, uncertainties in design variables are inevitable; thus in addition to pursuing the theoretical optimum of the objective function the evaluation of robustness of the optimum solution is also critical. Several methodologies exist to tackle robust optimisation, such as Worst Case Optimisation and Gradient Index; this paper investigates the use of standard deviation and mean value of objective function under uncertainty of variables. A modified kriging model with the ability of balancing exploration and exploitation is employed to facilitate the objective function prediction. Two TEAM benchmark problems are solved using different methodologies to compare the advantages and disadvantages of different robust optimisation approaches.

Index Terms — Worst Case Optimisation, Gradient Index, Six Sigma Quality Approach, standard deviation, mean value, kriging.

I. INTRODUCTION

In practical electromagnetic problems design variables are often subject to tolerances or uncertainties; various ways of assessing performance variation under uncertain conditions have been tried, most popular being the Worst Case Optimisation (WCO) [1]-[4], the Gradient Index (GI) [5], [6] and the Six Sigma Quality (SSQ) [7], [8]. Their performance is compared in this paper and a technique which combines the WCO and GI with with the Six Sigma technique is introduced and discussed. The Six Sigma techniques for process improvement and to aid business strategies were proposed in mid 80s and recently used in the context of quality manufacturing [7], [8]. The combined algorithm proposed here utilizes the relatively accurate and cheap prediction provided by the modified kriging surrogate model which is able to balance exploration and exploitation adaptively [9], [10]. The algorithm is verified against a difficult test function and two TEAM benchmark problems.

II. ROBUST OPTIMISATION

In conventional optimisation finding the minimum (maximum) of the objective function is normally set as the only task while the search space if limited through constrain. When practical devices are designed, however, most parameters (design variables) are subject to uncertainties due to manufacturing tolerances, variation of material properties, etc., and their influence on performance needs to be known. Thus in addition to finding the theoretical optimum its robustness may need to be assessed, often quantitatively.

A. Robust Optimisation

To illustrate the importance of robustness an example will be shown using the following test function

\[ y = 10 - \sum_{i=1}^{3} \left( \frac{x_i}{1+|x_i|^{2/5}} + \frac{x_i}{1+|x_i|^{1/5}} + \frac{x_i}{1+|x_i|^{1/3}} \right) \]  \hspace{1cm} (1)  

where the range 0 \( \leq x \leq \pm 1 \) has been considered.

The objective function (1) is to be minimised subject to constraints and uncertainties; the latter may be specified directly (e.g. as machining tolerances, say \( \Delta \)) or defined mathematically as

\[ U(x) = \{ \xi \in \mathbb{R}^n | x - k \xi \leq \xi \leq x + k \xi \} \]  \hspace{1cm} (2)  

where \( \sigma \) is standard deviation of uncertain variables and \( k \) is determined by a confidence level [5]. The single-variable (n=1) version is plotted in Fig. 1 with the corresponding standard deviation for the assumed uncertainty \( \Delta \xi = 1.5 \).

In practical cases the design vector is often constrained, as in Fig. 1. Point A1 is the theoretical global optimum, but after considering the uncertainty clearly offers inferior robustness compared with A2, even if the latter is only a local minimum. The infeasible regions 3.5\( \leq x \leq 5.5 \), and \( 15.5 < x < 21.0 \) are restricted by constraints on the variable x. Moreover, there may exist an imposed requirement for the objective function not to exceed a certain value, as depicted by the horizontal dashed line at \( y = 8.35 \), further reducing the quality of the solution given by A1. Overall, reliable ways of making a judgement about the robustness are required.

Fig. 1. Objective function and the standard deviation (\( \Delta \xi = 1.5 \)).

B. Worst Case Optimization

The Worst Case Optimization (WCO) [11]-[14] is a popular approach which can predict the worst scenario considering uncertainties and constraints with respect to specific designs. It analyses the reliability of the solution as follows

\[ \text{Minimize } f_{wco}(x) = \max_{\xi \in \Xi} (f(x)) \]  \hspace{1cm} (3)  

Subject to \( g_{wco}(x) = \max_{\xi \in \Xi} g(x) \leq 0, i = 1, \ldots, m \)
Appendix C Journal Papers

For a given uncertainty of a variable (or a set of variables), the worst value of the objective function in a given range (surrounding a selected point x) is used instead of the original value at x. Thus the worst possible performance in the vicinity of an optimum is considered directly. Moreover, solutions that may push the function into infeasible region restricted by constraints may be avoided. The results of the worst case analysis of the objective function (2) (of Fig. 1) are shown graphically in Fig. 2, under the uncertainty A(x)=1.5. Clearly A2 offers superior performance in terms of the worst case.

![Graph showing worst case performance](image)

Comparing only the worst case, however, means that the average variance within each range is ignored. Fig. 3 shows two functions, otherwise identical, except for the region around the points A2 and A4, respectively, with more details in Table I. The shape of the functions is clearly different, but all the ‘descriptors’ – global minimum (GM), local minimum (LM), worst case (WC) and even average value (AV) – are the same. Thus the WCO, even with the additional measure of average value, may not be reliable.

![Graph showing example variation of the objective function close to a local minimum](image)

### Table I

<table>
<thead>
<tr>
<th>Function</th>
<th>GM(A1, A2)</th>
<th>LM(A2, A4)</th>
<th>WC(A2, A4)</th>
<th>AV(A2, A4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>x=15, y=1.022</td>
<td>55, y=2.127</td>
<td>70, y=4.478</td>
<td>2.9556</td>
<td></td>
</tr>
</tbody>
</table>

![Table showing common features of the functions](image)

The WCO may, however, be utilized to map the probability of constraint violation. The probability of constraint violation for the function of Fig. 1, which is the measure of how likely the constraint violation may occur out of all the potential possible values in the uncertain range, is plotted in Fig. 4. A1 has a slightly lower probability of 32%, whereas A2 has probability of 35%. The worst case approach could therefore be used to evaluate the probability of constraint violation for each solution in the proposed robust optimization strategy.

### C. Gradient Index (GI)

Another way of incorporating robustness into the mainsteam optimization process is by adding the gradient index [5] as a second objective and formulating the problem as

\[
\text{Minimise } f(x) \quad x \in R^n (x_i \leq x \leq x_u)
\]

\[
\text{Minimise } G(x) = \max_{i=1}^{m} |\frac{\partial f(x)}{\partial x_i}| \quad \text{(4)}
\]

Subject to

\[
g_i(x) \leq 0, \quad i = 1, \ldots, m
\]

The values of the first-order gradient for points A1 and A2 are both close to zero. Another point, A3, from Fig. 1 appears to offer a better objective function value than A2, but its robustness is poor. Thus only minimising the first-order gradient might not offer reliable criteria.

![Graph showing first- and second-order gradient indices of the objective function](image)

Moreover, the size of the uncertainty matters. It is helpful to define the sensitivity of the gradient as the difference between the largest and the smallest value of the GI within the uncertainty range; the shape of this sensitivity carries useful information as shown in Fig. 5. However, the assessment of average performance variation should be carried out too.

![Graph showing sensitivity of the objective function](image)

### D. Six Sigma Quality Method

Both methodologies discussed above have limitations. The Six Sigma Quality (SSQ) method proposed in this paper can provide a reliable evaluation as well as assessing the average performance. With SSQ formulation the measure of dispersion is classified into six sigma (standard deviation) levels [7], while the optimisation problem can be reformulated as

\[
\text{Minimise } \mu_y \text{ and } \sigma_y
\]

(5)

where \( \sigma_y \) is the standard deviation indicating the intensity of variation due to the uncertainty of variables and \( \mu_y \) the mean value defining the average performance within the uncertain range. The two parameters, \( \sigma_y \) and \( \mu_y \), for the test function (2) are depicted in Figs. 7 and 8, respectively. The trajectory of the objective function in terms of standard deviation and mean value is plotted in Fig. 9 with reference to characteristic points from Fig. 1. The standard deviation of A2 at 0.0302 is less than for A1 at 0.1499 implying better robustness.
Appendix C Journal Papers

III. ROBUST OPTIMISATION EXPLOITING KRINGING

A. Kriging

The methods discussed in Section II were combined with a kriging assisted surrogate model [10], [11], considering both unconstrained and constraint optimisation. For the unconstrained case, once the prediction of the objective function is provided by the kriging model, the sensitivity can be evaluated using either the gradient difference or the Six Sigma Quality method. The latter can also assess the average performance. The sensitivity is then used to gauge the robustness of the solution. In the case of constraint optimisation, once again the kriging provides predictions of the objective function but also of the constrained values. The worst case method is applied to compute the probability of constraint violation and then either the gradient index or the six sigma algorithm is used once more to calculate the sensitivity to learn about the robustness. The proposed methodology is verified using two TEAM problems and against published results as described below.

B. TEAM 22 Problem

The first example involves a multi-objective version of the TEAM 22 problem, described fully in [15]. The target is an arrangement of two superconducting coils yielding the stored energy of 180 MJ while a minimal stray field $B_{\text{stray}}$ should be maintained. The objective function is defined as

$$ \text{OF} = \frac{B_{\text{stray}}}{B_{\text{form}}} \frac{E_{\text{ext}}}{E_{\text{int}}} \frac{1}{\text{Rect}} \text{,}$$

where $B_{\text{stray}}$ is the stray field and $B_{\text{form}}$, $E_{\text{ext}}$, and $E_{\text{int}}$ are defined subject to geometrical and 'quench' constraints. The three parameters case, which includes three geometric variables $R_2$, $H_2$ and $D_2$, while $R_1$, $H_1$ and $D_1$ are fixed, has been tried under different uncertainties, which can exist in the geometric variables or the current densities in the coils. The uncertainties have been set as $R_2 - 0.03 < R_2 < R_2 + 0.03$, $R_3 - 0.042 < R_3 < R_3 + 0.042$, $d_2 - 0.009 < d_2 < d_2 + 0.009$. A constraint was imposed that the superconducting coils should not violate the quench condition, which links the value of the current density and the maximum value of magnetic flux density as

$$ g_i(x) = |i| + 6.4 \cdot |B_{\text{mag}}| - 54.0 \leq 0, \quad i = 1, 2 \text{ (7)} $$

### TABLE II

<table>
<thead>
<tr>
<th>R2 (m)</th>
<th>h2 (m)</th>
<th>d2 (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.100</td>
<td>0.300</td>
<td>0.200</td>
</tr>
</tbody>
</table>

Test range: [2.6, 3.4], [0.4, 2.2], [0.0, 0.4]

Step size: 0.01, 0.01, 0.003

Number of steps: 81, 120, 101

### TABLE III

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>$R_2$ (m)</th>
<th>$d_2$ (m)</th>
<th>$g_i$ (m)</th>
<th>$OF$</th>
<th>Iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA</td>
<td>3.045</td>
<td>0.386</td>
<td>0.240</td>
<td>0.134</td>
<td>2400</td>
</tr>
<tr>
<td>HooTS</td>
<td>3.080</td>
<td>0.380</td>
<td>0.246</td>
<td>0.089</td>
<td>321</td>
</tr>
<tr>
<td>ITS</td>
<td>3.100</td>
<td>0.388</td>
<td>0.240</td>
<td>0.098</td>
<td>1824</td>
</tr>
<tr>
<td>SA</td>
<td>3.079</td>
<td>0.390</td>
<td>0.237</td>
<td>0.098</td>
<td>5052</td>
</tr>
<tr>
<td>NTS</td>
<td>3.080</td>
<td>0.370</td>
<td>0.254</td>
<td>0.089</td>
<td>1600</td>
</tr>
<tr>
<td>PBI</td>
<td>3.110</td>
<td>0.421</td>
<td>0.241</td>
<td>0.101</td>
<td>3728</td>
</tr>
<tr>
<td>FSKir(g)</td>
<td>3.090</td>
<td>0.394</td>
<td>0.236</td>
<td>0.0875</td>
<td>211</td>
</tr>
<tr>
<td>AWEI[Kir(g)]</td>
<td>3.090</td>
<td>0.400</td>
<td>0.232</td>
<td>0.0875</td>
<td>323</td>
</tr>
</tbody>
</table>

Genetic Algorithm (GA) [17]; Tabu Search (HooTS) [18]; Improved Tabu Search (ITS) [19]; Simulated Annealing Algorithm (SA) [20]; New Tabu Search (NTS) [21]; Population-based Incremental Learning (PBI) [22].

Fig. 11. TEAM 22 results: (a) kriging with EL (b) kriging with AWEI (c) prediction of objective function, average performance and standard deviation). More information about A1 to A4 in Table IV (SA: Sensitivity assessment; C: Constraint violation; Pr(C): Probability of constraint violation)

### TABLE IV

<table>
<thead>
<tr>
<th>Points</th>
<th>$R_2$ (m)</th>
<th>$h_2$ (m)</th>
<th>$d_2$ (m)</th>
<th>SA</th>
<th>C</th>
<th>Pr(C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3.04</td>
<td>0.246</td>
<td>0.397</td>
<td>6.232</td>
<td>-0.436</td>
<td>0%</td>
</tr>
<tr>
<td>A</td>
<td>3.06</td>
<td>0.232</td>
<td>0.397</td>
<td>6.654</td>
<td>-1.2073</td>
<td>0%</td>
</tr>
<tr>
<td>A</td>
<td>3.08</td>
<td>0.246</td>
<td>0.397</td>
<td>6.105</td>
<td>-0.5107</td>
<td>0%</td>
</tr>
<tr>
<td>A</td>
<td>3.03</td>
<td>0.281</td>
<td>0.397</td>
<td>6.147</td>
<td>-1.9416</td>
<td>0%</td>
</tr>
</tbody>
</table>

The initial sampling points would normally be selected using the Latin Hypercube [17]; in our tests we fixed the positions at $(R_2=2.7m, h_2=0.748m, d_2=0.13m)$, $(R_2=2.9m, h_2=1.304m, d_2=0.227m)$, $(R_2=3.0m, h_2=1.46m, d_2=0.37m)$, and $(R_2=3.3m, h_2=2.088m, d_2=0.37m)$ to facilitate comparisons between the kriging model with different strategies of updating sampling points.

C. TEAM 25 Problem

A model of a disc press with an electromagnet for producing anisotropic permanent magnets is chosen as a second example.
The objective function is evaluated at specific points as
\[ OF = \sum_{i=1}^{n} \left( B_{c_{i,calc}} - B_{c_{i,req}} \right)^2 + \left( B_{g_{i,calc}} - B_{g_{i,req}} \right)^2 \]
where coil means calculated and reqi required. There are three initial sampling points which are chosen randomly.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Min</th>
<th>Max</th>
<th>Step Size (mm)</th>
<th>No of Steps</th>
</tr>
</thead>
<tbody>
<tr>
<td>R_1</td>
<td>5</td>
<td>9.4</td>
<td>0.1</td>
<td>45</td>
</tr>
<tr>
<td>R_2</td>
<td>12.6</td>
<td>18</td>
<td>0.1</td>
<td>55</td>
</tr>
<tr>
<td>R_3</td>
<td>14</td>
<td>45</td>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td>R_4</td>
<td>4</td>
<td>19</td>
<td>0.5</td>
<td>31</td>
</tr>
</tbody>
</table>

The objective specification of TEAM 25 is provided in Table V.

### References

Correlation matrices in kriging assisted optimisation of electromagnetic devices

Song Xiao, Mihai Rotaru and Jan K. Sykulski

Electronics and Computer Science (ECS), University of Southampton, Southampton, SO17 1BJ, UK

ikss@ecs.soton.ac.uk

Keywords: Kriging, correlation matrices, multi-parameter optimisation, average gradient

Abstract

Kriging surrogate modelling facilitates efficient decision making about where to place the next point for evaluation during optimisation. This is particularly helpful in the design of electromagnetic devices where computationally expensive numerical field modelling needs to be used. The disadvantage, however, is that correlation matrices are required which, for problems with many design variables and multiple objectives, may grow in size leading to the need for page swapping and thus slowing down of what in principle should be a very fast process. In this paper several methodologies to reduce the computational resources required in such problems are proposed. The efficiency of the proposed approach is demonstrated using an example of a large multi-parameter optimisation problem where kriging coupled with the average gradient value method is employed.

1 Introduction

Kriging [1-4] predicts the shape of the objective function by considering the spatial correlation of data based on limited information, thus offers efficient and inexpensive surrogate to replace the computationally demanding numerical simulation (such as finite elements). The accuracy of the prediction can be estimated by the mean square error in kriging to assist in a decision where to place the next evaluation point during optimisation iterations. The spatial correlation exists between the known points (vectors) of the objective function and all the unknown points, as well as amongst the known points (newly found points and initial sampling points). The way of calculating and storing the kriging correlation matrices was suggested in [5] and is shown schematically in the box marked M1 in Figure 1.

This relies on the linear regression model

\[ \hat{y}(x) = \sum_{i=1}^{n} \beta_i f_i(x) + \epsilon(x) \]  

(1)

and the Gaussian correlation model
Appendix C Journal Papers

\[ R \left( e(x^i), e(x^j) \right) = \left[ \prod_{k=1}^{N} e^{-\theta_k d_k} \right]^{P_k} \]

where the global function \( \sum_{i=1}^{N} \beta_i f_i(x) \) and an additive Gaussian noise \( \epsilon(x) \) are integrated to the predicted value \( \hat{y}(x) \) of the objective function; \( \theta_k \) is the correlation amongst the data in \( k \)-direction and \( p_k \) determines the 'smoothness' of (2). The most popular correlation function is given by the Gauss model where the value of \( p_k \) is simply taken as equal to 2. For a given set of data, the maximum likelihood estimation optimizes the value of \( \theta \) and then the correlation model is brought into the regression model to evaluate the function with the best linear unbiased predictor [1]. Although theoretically kriging could be used for any type and size of optimization, care must be taken with large problems (many variables and multi-objective) as correlation matrices can grow rapidly.

2 The storage issue of correlation matrices

With the increase in the number of sampling points selected by kriging during the iterations, the amount of data produced by the correlation matrices (especially the correlation matrices between the known sampling points and the unknown points of objective functions), shown as M1 in Figure 1, accumulates constantly throughout the optimization process, which may become problematic especially when coping with large-scale multi-parameter tasks, arising to a 'combinatorial explosion' [5]. In our previous work [6] a successive 'zoom in' strategy to alleviate the problem was proposed, where -- in order to reduce the amount of data storage and utilize the installed physical memory capacity efficiently -- the step sizes of design vectors 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. How to utilize the test experience obtained from the initial-stage tests, to guide the model to define the test range and specific step size flexibly in each dimension, is discussed in this paper, with the aim to guarantee the surrogate model converging to the global optimum or, at worst, the region nearby the global optimum.

There are two ways in which the correlation matrices can be divided into sub-elements that can be manipulated in an efficient way, by partitioning the correlation matrix in terms of sampling points or via the design vectors, as shown in box M2 of Figure 1. The splitting of the correlation matrix by sampling points, which is only adding sub-matrices between each sampling point and all design vectors, rather than inserting the full-version correlation matrix into calculation, is applied during the process of producing the predicted response surface. But during the process of estimating the mean square error in the predictor, a unique way of partitioning the
correlation matrix is by dividing it into sub-matrix via the design vector. This partitioning approach can be time-consuming, however, because each cut requires rebuilding the full correlation matrix for ‘grabbing’ a slice of the sub-matrix. Using only one of these approaches may not give the expected computational improvement, thus both schemes in M2 have been used for different purpose, one for producing response surface, the other for producing mean square error. Together they can maintain correlation matrices at acceptable sizes that should not exceed the available memory space and thus manage the available RAM more efficiently.

Figure 1: The schemes of producing correlation matrices

3 The structure of the correlation matrices

It has been observed that the sub-matrices created when new sampling points are added during iterations have similarities that can be exploited to reduce the memory requirements. In particular, the ‘shape’ of the distribution of values in a given row of the matrix will be preserved (a ‘benchmark’) but shifted depending on the position of the point. Thus instead of creating and storing the whole matrix it may be sufficient to duplicate information. This is best explained using examples, therefore in the following sections several examples are discussed and analysed, starting with a single variable problem, gradually moving to multi-variable problems up to eight variables. The single and two variables problems are exemplified using analytical functions, whereas the multi variables problems are illustrated with the help of TEAM 22 (SMES Optimisation Benchmark) and TEAM 25 (Optimisation of Die Press Model) electromagnetic optimisation problems.
3.1 The single-variable numerical test

The single-variable Schwefel function [7] was first used

\[ f(x) = \sum_{i=1}^{n} -x_{i} \sin(\sqrt{|x_{i}|}) \]  

(3)

where \( n=2 \), with six initial sampling points \( x=(-500, -230, -250, -40, 160, 500) \), paying attention that the first one was on the edge of the region. A full correlation sub-matrix was then created for this point and the distribution is shown in Figure 2a. It was then observed that all points have a similar distribution, with Figure 2b giving an example for the second point. It is therefore possible to set up a ‘standard’ sub-matrix (as depicted in Figure 3) and then shift it (using a ‘window’) to appropriate position depending on the location of the point under consideration (effectively a ‘cut and paste’ technique), hence use storage space far more economically.

![Figure 2: Correlation sub-matrices for (a) point 1, (b) point 2.](image)

![Figure 3: A ‘standard’ sub-matrix and associated ‘window’.](image)

3.2 The two-variable numerical test

The following analytical function was used and tested for the two variable case

\[ y = 10 - \sum_{i=1}^{n} \left[ \frac{5\times 5}{1+(x_{1}-5)^2} + \frac{2.2}{1+(x_{2}-15)^2/10} + \frac{2.2}{1+(x_{2}-25)^2/300} \right] \]  

(4)
Although this function can have \( n \) variables, for the example discussed here \( n \) was set to two (\( n = 2 \)) in the range \( 0 \leq x \leq 27 \). The point \( (x_1=0, x_2=0, y_0=0.43) \) was selected to create a standard correlation sub-matrix to be used by all remaining and new points, as illustrated graphically in Figure 4.

![Figure 4: Building and duplicating sub-matrices.](image)

In order to assess achievable savings in computer memory and associated computing times, the graphs in Figure 5 have been produced, showing and comparing three cases: 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 described previously in [5].

![Figure 5: (a) Memory requirements, and (b) computing times.](image)
The test demonstrates that significant savings can be made, if required, of necessary memory requirements, by applying the concept of duplicating the standard correlation sub-matrix. This can be combined with the idea of matrix partitioning. The added bonus is that the computing times can also be reduced, especially as the iterations progress, although in the case of matrix partitioning this benefit may be lost somewhat towards the end of the iterative process. When applying the correlation matrix partitioning scheme, the splitting by the design vectors requires more time to produce the correlation matrix repeatedly for creating a sub-matrix; thus the sub-matrix duplication assisted by correlation matrix partitioning may ultimately require more computing time.

4 TEAM 22

4.1 TEAM Problem 22

The full description of the TEAM benchmark problem 22 (superconducting magnetic energy storage system) may be found in [8]. The target is an arrangement of the two superconducting coils such that the stored energy within the system is \( E_{opt} = 180MJ \) while a minimal stray field \( B_{stray} \) is present. The objective function is defined as

\[
OF = \frac{E_{max}}{E_{ref}} + \frac{|E_{ref} - E_{opt}|}{E_{ref}},
\]

where \( B_{stray} = 3\mu T \) and \( E_{max} = \frac{\text{V}\cdot\text{amp}\cdot\text{s}^2}{22} \), subject to geometrical and ‘quench’ constraints.

4.2 The 3-parameter Test Results

The full 3 parameter TEAM 22 problem with standard test settings [8] 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 due to 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 initial sampling points were chosen differently to our previously reported test [9]; in particular the value at the lower bound of the three variable’s test range was selected specifically as the first sampling point to create a standard correlation sub-matrix for duplication, instead of repeatedly calling the correlation function (2), which would be an inefficient and time consuming process. The rest of the sampling points are randomly chosen as summarised in Table 1. The efficiency of the algorithm and the memory usage were compared for the three methods M1, M2 and M3 of Figure 1, where M1 and M2 have already been explained, whereas M3, as a combination of the other two methods, is similar to M2, but with initial duplication of the standard sub-matrix before partitioning is applied. With the prescribed ranges and step sizes the total size of the problem is \( 81\times129\times81 \) as the standard test size shown in [8], whereas the size of the sub-matrices needs
to be set less than the physical memory available in the computer but also depends on the number of sampling points; in our tests this has been set to \( n_s \times 5 \times 10^3 \), where \( n_s \) is the number of existing sampling points.

<table>
<thead>
<tr>
<th>( R_i ) (m)</th>
<th>( h_i ) (m)</th>
<th>( d_i ) (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1</td>
<td>2.6</td>
<td>0.408</td>
</tr>
<tr>
<td>Sample 2</td>
<td>3.4</td>
<td>2.2</td>
</tr>
<tr>
<td>Sample 3</td>
<td>3.2</td>
<td>0.744</td>
</tr>
</tbody>
</table>

Table 1. The setting of initial sampling points.

<table>
<thead>
<tr>
<th>( R_i ) (m)</th>
<th>( h_i ) (m)</th>
<th>( d_i ) (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test range</td>
<td>[2.6 3.4]</td>
<td>[0.408 2.2]</td>
</tr>
<tr>
<td>Step size</td>
<td>0.01</td>
<td>0.014</td>
</tr>
<tr>
<td>Number of steps</td>
<td>81</td>
<td>129</td>
</tr>
</tbody>
</table>

Table 2. The specific definition of test.

Only the calculation time during the prediction stage and the memory usage are different when using these three different approaches of correlation matrices construction, since the number of updated sampling points required for converging to the global optimum is the same. A comparison between the results obtained with our kriging assisted algorithm, using the three schemes of building correlation matrices, and some published results using other methods is presented in Table 3. The slight difference in the results obtained here as compared with our previous work [9] is due to the different initial settings used. The kriging assisted by EI took 214 iterations to locate the optimum, including the initial sampling points, thus the total number of calls of the FEM software was 217, which outperforms other algorithms.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>( R_i ) (m)</th>
<th>( d_i ) (m)</th>
<th>( h_i/2 ) (m)</th>
<th>( OF )</th>
<th>No of FEM calls</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA</td>
<td>3.040</td>
<td>0.386</td>
<td>0.240</td>
<td>0.134</td>
<td>2400</td>
</tr>
<tr>
<td>HuTS</td>
<td>3.080</td>
<td>0.380</td>
<td>0.246</td>
<td>0.089</td>
<td>3821</td>
</tr>
<tr>
<td>ITS</td>
<td>3.100</td>
<td>0.388</td>
<td>0.240</td>
<td>0.098</td>
<td>1824</td>
</tr>
<tr>
<td>SA</td>
<td>3.078</td>
<td>0.390</td>
<td>0.237</td>
<td>0.098</td>
<td>5025</td>
</tr>
<tr>
<td>NTS</td>
<td>3.080</td>
<td>0.370</td>
<td>0.254</td>
<td>0.089</td>
<td>1800</td>
</tr>
<tr>
<td>PBIL</td>
<td>3.110</td>
<td>0.421</td>
<td>0.241</td>
<td>0.101</td>
<td>3278</td>
</tr>
<tr>
<td>Kriging</td>
<td>3.09</td>
<td>0.349</td>
<td>0.267</td>
<td>0.08778</td>
<td>217</td>
</tr>
</tbody>
</table>

Table 3: Performance Comparison Algorithms for TEAM 22 Problem
Appendix C Journal Papers

Genetic Algorithm (GA) [10]; Tabu Search (HTS) [11]; Improved Tabu Search (ITS) [12]; Simulated Annealing Algorithm (SA) [13]; New Tabu Search (NTS) [14]; Population-based Incremental Learning (PBIL) [15].

Figure 6: (a) computing time (b) peak memory requirement (obtained by different correlation matrix building methods).

But the main purpose of this test was to observe calculation times and peak memory requirements when different schemes were used for building correlation matrices. From Figure 6 it can be seen that although the memory usage can be maintained below certain size through applying the correlation matrix partitioning method, computing times are longer during the prediction process because of the way of partitioning the matrices via design vectors. On the other hand, when the correlation matrix partitioning method is combined with the sub-matrix duplication scheme, the kriging model performs the optimisation task efficiently and, as shown in Figure 6(a), outperforms the other two algorithms. Initially the sub-matrix duplication method, which requires storing the benchmark standard sub-matrix for duplicating the other correlation matrices, has a larger memory usage, but after approximately 25 iterations, for this specific TEAM 22 problem, the speed of memory increase reduces quite dramatically, as shown in Figure 6(b). As the iterations continue there is a linear increase of memory requirements for algorithms M1 and M2, which eventually overtake the memory requirements for the M3 approach. Although for this case the size of the problem was not exceptionally large, and was comfortably handled by a PC with a 16GB of installed RAM, it is obvious that for larger problems using M3 would be advantageous.

4.3 The 8-parameter TEAM 22

The 8-parameter TEAM 22 problem is a challenging test and the fact that there are very few results published in literature is a clear indication of the difficulties. Several observations can be made before attempting to solve this problem using the proposed algorithm. It is obvious that a large memory will be necessary to store the
response surface produced by kriging, as well as the correlation matrices used during the prediction stage. The size of these matrices is directly linked to the number of variables, hence in this case of 8 variables and using modest 10 steps for each variable vector, a $10^9$ parameter set needs to be stored. Because the kriging predictor produces the response surface at each iteration, and it is of the same size as the task itself ($10^9$), a computer with 9GB available memory is bound to be challenged. Moreover, background processes are likely to reduce available memory for the optimizer due to dynamic allocation. Adding correlation matrices to this data will only make the problem worse. Thus the correlation matrix partitioning and/or sub-matrix duplication schemes may need to be supplemented by some form of a ‘zoom in’ strategy, which has been used successfully before. However, the difficulty of this strategy is how to define the optimal step size, as this may determine whether the algorithm will converge to the right answer. To alleviate this difficulty it is proposed that a sensitivity analysis in each dimension is conducted before deciding which of the parameters should be using a finer step size. We have decided to use an average value of the gradient in each dimension to guide the algorithm when choosing the variables that need a finer resolution. The average value of the gradient is thus calculated after each iteration and a variable that needs finer resolution may change in the next step. Table 4 presents the specific settings at the initial stage.

<table>
<thead>
<tr>
<th>$R_1$(m)</th>
<th>$R_2$(m)</th>
<th>$h_1/2$ (m)</th>
<th>$h_2/2$ (m)</th>
<th>$d_1$(m)</th>
<th>$d_2$(m)</th>
<th>$J_1$(A/mm$^2$)</th>
<th>$J_2$(A/mm$^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min</td>
<td>1.0</td>
<td>1.8</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>10</td>
<td>-30</td>
</tr>
<tr>
<td>Max</td>
<td>4.0</td>
<td>5.0</td>
<td>1.8</td>
<td>1.8</td>
<td>0.8</td>
<td>30</td>
<td>-10</td>
</tr>
<tr>
<td>No of steps</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Step size</td>
<td>0.6</td>
<td>0.64</td>
<td>0.34</td>
<td>0.34</td>
<td>0.14</td>
<td>0.14</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 4. The specific definition of 8-parameter TEAM 22 test at the initial stage.

After the first-stage test, in order to determine how to resize the parameters at the next stage, an average value of the gradient with respect to each dimension was calculated. The values obtained from the tests at different stages are listed in Table 6. In this particular case the parameters $h_1$, $h_2$, $d_1$, $d_2$, seem to have higher values of the average gradient, indicating higher sensitivity, and should therefore be focused on in the following tests by refining the increment (increasing the number of steps) in stages two, three, and so on. All this is illustrated in Table 6.
### Appendix C Journal Papers

<table>
<thead>
<tr>
<th>Test stage</th>
<th>$D(x)$</th>
<th>$h_1$</th>
<th>$h_2$</th>
<th>$d_1$</th>
<th>$d_2$</th>
<th>$J_1$</th>
<th>$J_2$</th>
<th>$\text{A/mm}^2$</th>
<th>$\text{A/mm}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>First-stage test</td>
<td>AG</td>
<td>0.0046</td>
<td>6.433e⁻⁴</td>
<td>0.0184</td>
<td>0.0446</td>
<td>0.2168</td>
<td>0.2168</td>
<td>3.056e⁻³</td>
<td>3.499e⁻³</td>
</tr>
<tr>
<td>Second-stage test</td>
<td>AG</td>
<td>0.0148</td>
<td>0.0541</td>
<td>0.0947</td>
<td>0.1024</td>
<td>0.8546</td>
<td>0.8415</td>
<td>4.232e⁻³</td>
<td>4.279e⁻³</td>
</tr>
<tr>
<td>Third-stage test</td>
<td>AG</td>
<td>0.0375</td>
<td>0.0231</td>
<td>2.736</td>
<td>1.743</td>
<td>0.476</td>
<td>0.257</td>
<td>4.572e⁻³</td>
<td>2.1755e⁻³</td>
</tr>
</tbody>
</table>

Table 5. The average value of gradient of 8-parameter TEAM 22 test. (D: , AG: Average Gradient)

<table>
<thead>
<tr>
<th>Test stage</th>
<th>$R_1$</th>
<th>$R_2$</th>
<th>$h_1$</th>
<th>$h_2$</th>
<th>$d_1$</th>
<th>$d_2$</th>
<th>$J_1$</th>
<th>$J_2$</th>
<th>$\text{A/mm}^2$</th>
<th>$\text{A/mm}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Second-stage test</td>
<td>1.0</td>
<td>2.8</td>
<td>1.8</td>
<td>3.72</td>
<td>0.2</td>
<td>2.24</td>
<td>0.38</td>
<td>0.36</td>
<td>0.03</td>
<td>0.38</td>
</tr>
<tr>
<td>No of Steps</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Third-stage test</td>
<td>1.0</td>
<td>1.6</td>
<td>3</td>
<td>3</td>
<td>21</td>
<td>21</td>
<td>29</td>
<td>29</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>No of Steps</td>
<td>7</td>
<td>1</td>
<td>41</td>
<td>41</td>
<td>21</td>
<td>21</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6. The average value of gradient of 8-parameter TEAM 22 test.

<table>
<thead>
<tr>
<th>Test stage</th>
<th>$R_1$</th>
<th>$R_2$</th>
<th>$h_1$</th>
<th>$h_2$</th>
<th>$d_1$</th>
<th>$d_2$</th>
<th>$J_1$</th>
<th>$J_2$</th>
<th>$\text{A/mm}^2$</th>
<th>$\text{A/mm}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimum solution</td>
<td>$R_1=1$, $R_2=1$, $h_1=0.2$, $h_2=0.52$, $d_1=0.24$, $J_1=30$, $J_2=20$, $O_f=0.9392$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No of FEM calls</td>
<td>183</td>
<td>17</td>
<td>29</td>
<td>220</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7. The optimal solution found in the four tests and the corresponding iterations spent.
(The units of the 8 parameters are as same as shown in Table 6; O_f: the objective function value)
### Appendix C Journal Papers

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>( R_1 ) (m)</th>
<th>( R_2 ) (m)</th>
<th>( h_1 ) (m)</th>
<th>( h_2 ) (m)</th>
<th>( d_1 ) (m)</th>
<th>( d_2 ) (m)</th>
<th>( J_1 ) (A/mm(^2))</th>
<th>( J_2 ) (A/mm(^2))</th>
<th>Objective function</th>
<th>No of FEM calls</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSO</td>
<td>1.0000</td>
<td>2.2647</td>
<td>1.1076</td>
<td>1.7766</td>
<td>0.5225</td>
<td>0.3442</td>
<td>28.1779</td>
<td>-5.4921</td>
<td>2.3363</td>
<td>-6000</td>
</tr>
<tr>
<td>Q-PSO</td>
<td>2.2947</td>
<td>2.6126</td>
<td>1.0764</td>
<td>2.2704</td>
<td>0.3967</td>
<td>0.2040</td>
<td>30</td>
<td>-21.293</td>
<td>0.5735</td>
<td>-6000</td>
</tr>
<tr>
<td>E-QPSO</td>
<td>1.0000</td>
<td>1.8000</td>
<td>2.0616</td>
<td>3.6000</td>
<td>0.5155</td>
<td>0.2851</td>
<td>19.9975</td>
<td>-6.3571</td>
<td>1.1730</td>
<td>-6000</td>
</tr>
<tr>
<td>GSA</td>
<td>1.939</td>
<td>2.823</td>
<td>1.130</td>
<td>1.101</td>
<td>0.399</td>
<td>0.195</td>
<td>22.5</td>
<td>-22.5</td>
<td>0.00517</td>
<td>17150</td>
</tr>
<tr>
<td>ES</td>
<td>1.990</td>
<td>2.931</td>
<td>1.293</td>
<td>0.940</td>
<td>0.290</td>
<td>0.188</td>
<td>26.6</td>
<td>-26.6</td>
<td>0.00489</td>
<td>4200</td>
</tr>
<tr>
<td>SAAR</td>
<td>1.694</td>
<td>2.907</td>
<td>1.609</td>
<td>0.882</td>
<td>0.323</td>
<td>0.207</td>
<td>20.9</td>
<td>-20.9</td>
<td>0.0110</td>
<td>14000</td>
</tr>
<tr>
<td>CGM</td>
<td>1.836</td>
<td>2.762</td>
<td>1.178</td>
<td>1.001</td>
<td>0.395</td>
<td>0.214</td>
<td>22.5</td>
<td>-22.5</td>
<td>0.0248</td>
<td>200</td>
</tr>
<tr>
<td>Kriging</td>
<td>1</td>
<td>1.8</td>
<td>1.56</td>
<td>1.39</td>
<td>0.4</td>
<td>0.15</td>
<td>30</td>
<td>-30</td>
<td>0.05361</td>
<td>499</td>
</tr>
<tr>
<td>Standard</td>
<td>1.296</td>
<td>1.8</td>
<td>2.178</td>
<td>3.026</td>
<td>0.583</td>
<td>0.195</td>
<td>16.955</td>
<td>-18.91</td>
<td>0.0018</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 8. Performance comparison between different algorithms

PSO: Particle Swarm Optimisation [16], Q-PSO: Quantum-behaved Particle Swarm Optimisation [16-18], E-QPSO: QPSO using the exponential probability distribution [19], GSA: Global Search Algorithm [20], ES: Evolution Strategy [20], SAAR: Simulated Annealing Algorithm [20], CGM: Conjugate Gradient Method [20].

Through analysing the sensitivity of the first-stage rough test, it has been determined that the test range of \( R_1 \) and \( R_2 \) needs to be contracted in order to save computer memory for the other four more sensitive parameters.

Moreover, the sensitivity of the current density settings \( J_1 \) and \( J_2 \) is extremely low, so during the three tests the test ranges and step sizes for currents have not been changed. The correlation matrix partitioning strategy has been applied and the full correlation matrix divided into sub-matrices of the size \( n_s \times 3 \times 10^6 \) where \( n_s \) is the number of sampling points.

The results presented here were acquired applying the following methodology: if available physical memory was larger than requested by the problem the normal method M1 was applied. A switch that allows moving to the correlation matrix partitioning method was implemented; the switching criterion was controlled by the time consumed to construct the correlation matrix. If the time used to build the correlation matrix was larger than a certain threshold, the switching was triggered and the partitioning scheme activated, which then continued throughout the remaining iterations.

In terms of the quality of the answer itself, compared with other methods and the standard answer provided by [8], with the help of the zoom-in strategy based on the evaluation of the prediction’s average gradient value the kriging algorithm seems to perform much more efficiently than the Partial Swarm and the relevant mutated methods with only 449 iterations required to converge to the optimum, compared to around estimated 6,000 needed with other methods. The other three algorithms GSA, ES, and SA can find better solution than the one achieved by kriging but with more numbers of FEM calls.
Appendix C Journal Papers

except the Conjugate Gradient Method (CGM), which performs more efficiently and accurately. Although these three methods finally converge to the better optimum, they still require dramatically more numbers of FEM calls. The best-performed method CGM spends fewer FEM calls finding out better solution than kriging. Although the quality of the solution obtained by kriging was marginally worse than the standard answer, it still performs better within the extremely limited number of FEM calls. If the optimum in Table 8 obtained by kriging does not satisfy designer, there could be more detailed tests done for further searching. If the answer provided by our scheme is not up to full satisfaction of the designer, more tests can be conducted regarding the sensitivity in other directions.

5 TEAM 25

A model of a die press with an electromagnet for producing anisotropic permanent magnets is chosen as a second example [21]. The shape of the die is to be set up in such a way that magnetic flux density components $B_x$ and $B_y$ should be the same and equal to $0.35\cos(\theta) T$ along a circle line in 10 measurement points for $0 < \theta < 45$ and $\theta = 0.01175\pi$. The problem has four design parameters $R_i$, $L_0$, $L_3$, and $L_4$ specified in Table 9. The objective function is evaluated at specific points as

$$OF = \sum_{i=1}^{10} ((B_{x,\text{calc}} - B_{x,\text{req}})^2 + (B_{y,\text{calc}} - B_{y,\text{req}})^2)$$

where \( calc \) means calculated and \( req \) required.

Three initial sampling points have been chosen as $(R_i=5.0, L_0=12.6, L_3=14, L_4=4.0)\), $(R_i=7.0, L_0=15.8, L_3=20, L_4=15)$, $(R_i=8.5, L_0=17, L_3=31, L_4=8)$. The first sampling point has been chosen so as to allow for creating the standard matrix for duplicating correlation matrices.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Test Range (mm)</th>
<th>Step Size (mm)</th>
<th>No of Steps</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_i$ (mm)</td>
<td>Min 5 Max 9.4</td>
<td>0.1</td>
<td>45</td>
</tr>
<tr>
<td>$L_0$ (mm)</td>
<td>12.6 Max 18</td>
<td>0.1</td>
<td>55</td>
</tr>
<tr>
<td>$L_3$ (mm)</td>
<td>14 Max 45</td>
<td>1</td>
<td>32</td>
</tr>
<tr>
<td>$L_4$ (mm)</td>
<td>4 Max 19</td>
<td>0.5</td>
<td>31</td>
</tr>
</tbody>
</table>

Table 9: The specific definition of TEAM 25

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>$OF(\times 10^5 T^2)$</th>
<th>$R_i$ (mm)</th>
<th>$L_0$ (mm)</th>
<th>$L_3$ (mm)</th>
<th>$L_4$ (mm)</th>
<th>No of FEM Calls</th>
</tr>
</thead>
<tbody>
<tr>
<td>HuTS</td>
<td>0.5009</td>
<td>7.3780</td>
<td>14.613</td>
<td>14.371</td>
<td>14.201</td>
<td>1580</td>
</tr>
</tbody>
</table>
Appendix C Journal Papers

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NTS</td>
<td>0.6482</td>
<td>7.4337</td>
<td>14.732</td>
<td>14.428</td>
<td>14.237</td>
<td>575</td>
<td></td>
</tr>
<tr>
<td>Kriging</td>
<td>0.4527</td>
<td>7.2</td>
<td>14.1</td>
<td>14</td>
<td>14.5</td>
<td>241</td>
<td></td>
</tr>
</tbody>
</table>

Table 9: Performance comparison of algorithms for TEAM 25 Problem

Universal tabu search (UTS) [22], otherwise as in Table 3

![Graph: Peak Memory Requirement vs Iteration](image)

Figure 7: The time consumed and the peak memory requirement at each iteration

As before, during the first stage of the test (as shown in Figure 7) the normal approach M1 has been applied until the switching criteria (when the peak memory requirement exceeds 9GB) is triggered and then Method M3 is applied from the 61st iteration onwards. Although the theoretical physical memory size is 16 GB, with the consideration of the impact from other background processes the upper limitation for the predictor was set as 9GB. The kriging model spends 238 iterations converging to the optimum, which performs more efficient than other methodologies in Table 9.

6 Conclusion

Kriging assisted optimisation has once again been demonstrated to perform efficiently, both in terms of a better quality of the final answer but also – and most importantly – the much smaller number of necessary function calls involving computationally expensive numerical modelling, such as finite elements in electromagnetic designs. However, some of the time gains could be lost due to the need to construct memory hungry correlation matrices required in the kriging approach. This will not matter when using large computers, but as design tasks are increasingly solved on small computers or laptops the issue of efficient management of computer memory needs to be addressed and this has been the focus of this paper.

267
The structure of the correlation matrices associated with the kriging model may be exploited to avoid uncontrolled growth of such matrices for large problems to avoid the need for page swapping which will inevitably lead to slowing down of the process. In particular, it is possible to duplicate and shift data while storing only a small subset of necessary information. For large-scale multi-variable tasks, a zoom-in strategy based on sensitivity evaluation of each variable may be implemented in association with the proposed novel strategies for creating correlation matrices. With the help of these methods, the issue of memory storage can be mitigated to significant extent.

References


Bibliography


[57] Z. Y. Ren, M. T. Pham, M. H. Song, D. H. Kim, and C. S. Koh. A robust global optimization algorithm of electromagnetic devices utilizing gradient index and


