46th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics & Materials Confer
18-21 April 2005, Austin, Texas

AIAA 2005-1901

A New Hybrid Updating Scheme for an Evolutionary Search Strategy using Genetic Algorithms and Kriging

Wenbin Song* and Andy J. Keane†

University of Southampton, Highfield, Southampton, SO17 1BJ, United Kingdom

This paper presents an efficient evolutionary search strategy based on design of experiments, genetic algorithms and response surface modelling. The strategy is constructed around a genetic algorithm while incorporating elements from design of experiment (DoE) and Kriging. In particular, the design points used to update the approximation model are derived from two surfaces, one is the approximation itself which provides the prediction of the function and the other is based on the error surface computed from posterior error estimates of the Kriging model. A genetic algorithm, which supports clustering, is used on both surfaces to return multiple points for parallel evaluation of the true function. A screening method is also used to remove points lying close to existing points based on the correlation coefficients between the point to be evaluated and all existing points. Numerical experiments suggest that significant improvements can be achieved using the proposed approach. Applications of the approach on engineering design problems are also studied.

I. Introduction

THE robustness in finding near-global optimal solutions of evolutionary search methods, and in particular, genetic algorithms, has attracted wide-spread use of such methods in a large number of optimisation and design problems such as engineering design, approximation, and scheduling. Improving the efficiency of evolutionary search algorithms has become a key factor in their successful application to real-world problems due to the high computational cost related to the high-fidelity simulation codes commonly used in such processes. Two categories of techniques have been proposed to tackle the efficiency issue of evolutionary search methods; the first type is focused on devising more efficient variants of the canonical algorithms,[1–3] the second type involves using approximations in lieu of the exact and often expensive function evaluations. The common feature of these two types of techniques is that they both try to reduce the number of fitness evaluations used since this is the most expensive part of the evolutionary process. As these two types of techniques can be easily combined together to further speed up the process, research effort can be focused on them separately. The focus of the current paper is on an efficient framework for combining genetic algorithms and Kriging, i.e., the focus is on the way in which genetic algorithms are used to search for near-optima and meta-models are updated to balance the need for searching for optima and building accurate approximations.

The first question is how to build approximation models based on computational simulation results. There have been various techniques developed for building meta-models; these can be broadly divided into two categories, interpolation and approximation, where the former interpolates the data points and the latter regreses them. Polynomials are often used for the later while interpolation methods include neural network methods and Kriging. Earlier work has been reported comparing the two types of methods in the context of optimisation studies. For example, comparisons between polynomial approximation and neural network and Kriging were presented by Carpenter et al.[4,5] and Giunta et al.,[6] respectively. The second question involves the use of approximation models in the optimisation study. Although these two processes are relatively independent from each other, building certain frameworks by combining the two can be beneficial and overcome some difficulties in building accurate approximate models, especially for high dimensional problems. The hybrid approach can also further improve efficiency by reducing the number of

*Research Fellow, School of Engineering Sciences, University of Southampton, Highfield, SO17 1BJ, United Kingdom, and AIAA Member.

†Professor, School of Engineering Sciences, University of Southampton, Highfield, SO17 1BJ, United Kingdom.

exact evaluations required for building sufficiently accurate metamodels. Different frameworks have been studied by a number of researchers, for example, Ong et al.[7] proposed a framework incorporating Genetic Algorithms and radial-basis function (RBF)-based local searches; Ulmer et al.[8] presented a Genetic algorithm framework based Gaussian Process using concepts similar to the expected improvements proposed by Jones et al.[9] Jin et al.[10] provided a comprehensive overview on using approximations within the evolutionary computation framework. The fundamental issue that needs to be addressed is the balance of building accurate meta-models (exploration) and locating near-global optima (exploitation). The most notable effort on this aspect is probably the use of the concept of expected improvements in choosing update points. Approaches based on Memetic Algorithms including Lamarckian learning and Baldwinian learning are also used to exploit local improvements in the global EA search. However, these efforts are more biased towards finding the optimal points, and less biased towards building accurate meta-models, which can be more important when the meta-models will be coupled with other meta-models in the context of a multidisciplinary optimisation study. Soboster and Keane[11] proposed a concept of weighted expected improvements to balance exploration and exploitation by using a weighting coefficient and also provided some useful guidelines for choosing the initial size of design of experiments and weighting coefficient values. Instead of using results obtained from searches on a weighted expected improvement surface, this work uses results obtained from parallel searches on two extreme surfaces to update the meta-models, which eliminates the computation of expected improvements. Also a screening phase is added after the search to remove those points close to existing ones by examining the correlation vectors between the candidate points and existing points in the data set.

This paper is organised as follows: section two describes the rational for the Kriging modelling. Section three gives details on the hybrid update scheme. Section four illustrate the effectiveness of the frameworks using numerical test functions, and section five concludes the paper.

II. Surrogate Modeling

Let $Y(\mathbf{x})$ denote the true response of the system under study, and $\mathbf{x} = (x_1, x_2, ..., x_m)^T$ denote the vector of control variables. Sometimes the true response of the system can be represented in explicit mathematical forms, but, in most cases, the knowledge of the system is incomplete or the model is too complex to represent using explicit functions, and therefore a complex computer code is used to simulate the relationship between the responses and inputs. Whatever the case, observations can be made either through physical experiments or computer simulations at some chosen values of the design variables, often using experimental design methods. In this work, let us suppose that data has been collected at n points denoted by $\mathbf{x}^{(i)} = (x_1^{(i)}, x_2^{(i)}, ..., x_m^{(i)}), (i = 1, 2, ...n)$, and the associated responses denoted by $y^i = Y(\mathbf{x}^(i))$. Let $y(x)$ represent the approximation model. The relationship between the true response and approximation can be represented as follows:

$$Y(\mathbf{x}) = y(\mathbf{x}) + \Delta(\mathbf{x}) \tag{1}$$

The difference between the true response and approximated response, the total error or residual, is due to two types of errors, one is system error (bias error) denoted by $\epsilon(\mathbf{x})$, which exists because of the incompleteness of the models employed. The second type is random error denoted by $\delta(\mathbf{x})$, which exists because of a number of reasons such as the effect of uncontrollable factors in the physical experiments, discretization errors typically encountered in the finite element analysis and computational fluid dynamics, and round off errors, etc. Therefore the total error is the sum of these two types of errors:

$$\Delta(\mathbf{x}) = \epsilon(\mathbf{x}) + \delta(\mathbf{x}) \tag{2}$$

Random errors can usually be controlled within certain level so that the output of a deterministic simulation code can be regarded as deterministic. Therefore the same set of inputs will produce the same outputs, this partly explains why a least squares model does not always provide a reasonably good approximation to a deterministic computer simulation code. In this case, an interpolation model would be more suitable for creating approximations. Among various techniques which interpolate the data, radial basis function (RBF) and Kriging were identified by Jin et al.[10] as being able to produce better results than other methods under multiple modelling criteria. The choice of Kriging techniques in this work is due to the fact that this method not only provides an estimate of the function values but also an estimate of posterior variance, which is used to control the frequency of re-sampling. A brief description of the Kriging model is provided below and

detailed discussions can also be found in.[9] The Kriging model is here expressed as

$$y(\mathbf{x}) = \beta + Z(\mathbf{x}) \tag{3}$$

where β represents a constant term in the model, and $Z(\mathbf{x})$ is a Gaussian random process with zero mean and variance of σ^2. The covariance matrix of $Z(\mathbf{x})$ is given by

$$Cov(Z(\mathbf{x}^i, Z(\mathbf{x}^j) = \sigma^2 R(\mathbf{x}^i, \mathbf{x}^j) \tag{4}$$

where σ^2 is the variance of the stochastic process and $R(.,.)$ is a correlation function between $\mathbf{x}^i$ and $\mathbf{x}^j$. Different types of correlation function can be employed as noted in Jones et al.[9] A commonly used type of correlation function can be expressed as

$$R(\mathbf{x}^i, \mathbf{x}^j) = \prod_{k=1}^{n} exp(-\theta_k |x_k^i - x_k^j|^p) \tag{5}$$

where $\theta_k 0$ and $1 \leq p_k \leq 2$ are the hyperparameters. Note that the above equation asserts that there is a complete correlation of a point with itself and this correlation decreases rapidly as the two points move away from each other in the parameter space. The choice of $p_k = 2$ would provide enough flexibility for modelling smooth but highly non-linear functions for most cases. The hyperparameters θ_k are estimated by maximizing the log-likelihood function given by

$$-\frac{1}{2}[n \ln \sigma^2 + \ln |\mathbf{R}| + \frac{1}{\sigma^2}(\mathbf{y} - \mathbf{1}\beta)^T \mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\beta)] \tag{6}$$

where σ^2 and β can be derived using the following equations once the θ_k are given

$$\hat{\beta} = (\mathbf{1}^T \mathbf{R}^{-1} \mathbf{1})^{-1} \mathbf{1}^T \mathbf{R}^{-1} \mathbf{y} \tag{7}$$

$$\hat{\delta}^2 = \frac{1}{n}(\mathbf{y} - \mathbf{1}\beta)^T \mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\beta) \tag{8}$$

A numerical optimisation procedure is required to obtain the Maximum Likelihood Estimates (MLE) of the hyperparameters. Once the hyperparameters are obtained from the training data, the function value at a new point can be predicted by

$$\hat{y}(\mathbf{x}^*) = \hat{\beta} + \mathbf{r}^T \mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\hat{\beta}) \tag{9}$$

along with the posterior variance $s^2(\mathbf{x}^*)$ given by

$$s^2(\mathbf{x}^*) = \sigma^2[1 - \mathbf{r}\mathbf{R}^{-1}\mathbf{r} + \frac{(1 - \mathbf{1}^T \mathbf{R}^{-1}\mathbf{r})^2}{(\mathbf{1}^T \mathbf{R}^{-1}\mathbf{1})}] \tag{10}$$

where $\mathbf{r}(\mathbf{x}) = R(\mathbf{x}, \mathbf{x}^1), ..., R(\mathbf{x}, \mathbf{x}^n)$ is the correlation vector between the new point $\mathbf{x}$ and the training dataset. This quantity provides a good indication on the accuracy of the prediction at new points and will be used in our framework to decide whether and where further exact analyses are required. To obtain an estimate of the accuracy of the predictions of the Kriging model, a leave-one-out cross-validation procedure can be employed. The measure used in such a procedure is called the 'standardized cross validated residual' (SCVR) defined

$$SCVR_j = \frac{y(\mathbf{x}^j) - \hat{y}_{-j}(\mathbf{x}^j)}{s_{-j}(\mathbf{x}^j)} \tag{11}$$

where $\hat{y}_{-j}(\mathbf{x}^j)$ and $s_{-j}(\mathbf{x}^j)$ denotes the mean and variance computed by (9) and (10) without using the jth training data. A good predictor would mean that the Gaussian process prior is appropriate for the dataset and the majority of results will be scattered in the interval $[-3, 3]$. Plotting the values of $SCVR_j$ against the predicted function values would also provide clues and suggest solutions to problems that might exist in the model For example, if there is any linear trend in such a plot, it is sometimes possible to improve the prediction by using a log transformation.

III. Hybrid Update Strategy for Use with Genetic Algorithms and Kriging

The aim of using approximations or meta-models is to reduce the cost of finding global optimal designs. This is equivalent to reducing the number of function calls to the exact problem codes. Therefore the optimal points on a meta-model should converge to the optimal points on the exact model, which is the requirement of global convergence for meta-models. Global convergence will be naturally achieved if an exact meta-model can be built. However, due to the modelling errors inherent in the meta-models, there needs to be a balance to be made in building an accurate meta-model and locating the global optima.

A number of methods can be used when it comes to updating the Kring models with added data points evaluated using exact objective functions. The most straightforward technique is to use the best points found on the Kriging model. This approach depends on the quality of the search on the Kriging, and the process may miss the global optima as it lacks the ability to identify promising but unsearched areas due to prediction errors involved in the Kriging. A more promising method is to update the Kriging model with points where the expected improvement (EI) is large compared to existing exact solutions, as shown by Jones et al. [9]. However, the update procedure in these two approaches is very much separated from the overall process of identifying the global optimum. Therefore, almost any optimisation method can be applied in the search for optima on Kriging or expected improvements, although the Branch-and Bound method is suggested by Jones et al.[9] in the search for EI.

Compared to above approaches, several other frameworks have been presented which involve strong coupling of the Kriging with Genetic algorithms due to the robustness of GAs and its ability to locate near-global optimal solutions. This paper adopts GAs as the basic framework and also the search algorithm used in the search on the Krig model, leading to a two-level search strategy. The searches on the Krig are carried out twice: the first to locate optimal objective function positions, and the second to locate positions of maximum posterior error. This is the reason why Krig is chosen here as the approximated objective function and its posterior error can be obtained in one go. Suggested points from these two GA searches are evaluated in parallel after being congregated and compared with the available exact solutions in order to remove overlapping points to further improve the efficiency. Adding these newly available exact points to the metamodel achieves the following two aims simultaneously: reduced approximation errors and improved search efficiency. The framework is shown in Fig.

BEGIN

Generate the initial population using design of experiments, the population size will be 10n;

Evaluate the 10n points using high fidelity analysis codes, and store objective function values

Build the initial response surface model using kriging, the hyperparemeters are tuned using GA/DHC (Dynamic Hill Climber) code;

While (number of exact evaluations less than the number requested/affordable)

Search the response surface model, return the best n points in terms of the best objective function on the response surface, and also search the response surface error model, return n points in terms of the maximum posterior prediction errors

Remove those points from combined set of design points based on correlation criteria or Euclidean distance criteriathis will produce the candidate points for exact evaluations

Evaluate these points using exact codes and place these points into database, otherwise if the aggregated data set is empty, terminate the search.

Update the response surface by updating the hyperparameters of the Kriging model

End While

END

Figure 1. Hybrid update scheme using genetic algorithm and Kriging

IV. Numerical Experiments

Numerical experiments have been carried out on two multimodal test functions, which are defined in Table 1. These two test functions share the common feature of point symmetry, so to avoid this feature being exploited, a non-symmetric design space with respect to the symmetry point is used in the search. In addition, these functions have known global optimum at $\mathbf{x} = 0$. Also both functions have a large number of local optimum points within the search range defined in the table, which makes it even more challenging to build an accurate global approximation model and locate the near global optimum.

Table 1. Test functions and Variable Bounds

Name	Function	Bound
Ackley	$f(\mathbf{x}) = 20 - 20\exp(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}) + e - \exp(\frac{1}{n}\sum_{i=1}^{n}\cos(2\pi x_i))$	$[-5, 30]^n$
Griewank	$f(\mathbf{x}) = \frac{1}{4000}\sum_{i=1}^{n} x_i^2 - \prod_{i=1}^{n}\cos(\frac{x_i}{\sqrt{i}}) + 1$	$[-10, 600]^n$

A two dimensional Ackley function is first used in the experiments, as the two dimensional problem can be readily visualized to illustrate the effect of the current strategy, as shown in Fig. 2 and Fig. 3. It can be seen that without increasing the total number of function evaluations (which is often a rough measurement of the computational cost), a more accurate global approximation can be achieved, and therefore a higher probability of locating the near-global optimum at lower computational cost is provided. As discussed earlier, the basic issue in optimisation problems using surrogate models in the balance between achieving an accurate surrogate model and locating the near global optimum. Depending on the type of surrogate models used, the computational effort required to achieve a good approximation model varies. In the case of krig models and the current scheme, tuning the hyperparameters is quite an expensive process, especially when the number of sample points grows. The search traces for the standalone GA and current scheme on these two test functions are shown in Fig. 4

V. Engine Nacelle Shape Optimisation

Next an aero engine nacelle shape optimisation problem is used to illustrate the effectiveness of the scheme. The engine nacelle geometry is defined in ProEngineer with around 40 parameters, here six parameters that define the shape of top lip profile plus the scarf angle parameter are used to formulate a seven parameter problem. The aim is to study the aerodynamic effect of top lip profile under different scarf angles. The parameters are listed in Table 2, along with parameter ranges and reference values. The meshing package Gambit and flow solver Fluent are used in the study. First, an appropriate mesh density is determined based on the accuracy of the solution and computational time it requires. Here it was decided based on the number of available processors and licenses: the ideal situation using 30 fluent jobs on 8 processors means that each calculation can be finished within 10 hours, and around 200 jobs should be finished within a week if jobs are run immediately after placed into the queue. In practice, it has taken a little more than three weeks to finish 200 calculations.

The geometries and pressure distributions for the base design and final result achieved using 200 calculations are presented in Fig. 5.

VI. Conclusions

An efficient evolutionary framework using GAs and Kriging is presented in this paper. Two types of updates are incorporated in the process to improve the robustness of locating global optima within a limited computational budget. The approach is validated using mathematical test functions and applied to an aero engine nacelle optimisation problem using solutions obtained from a computational fluid dynamics (CFD) calculation.

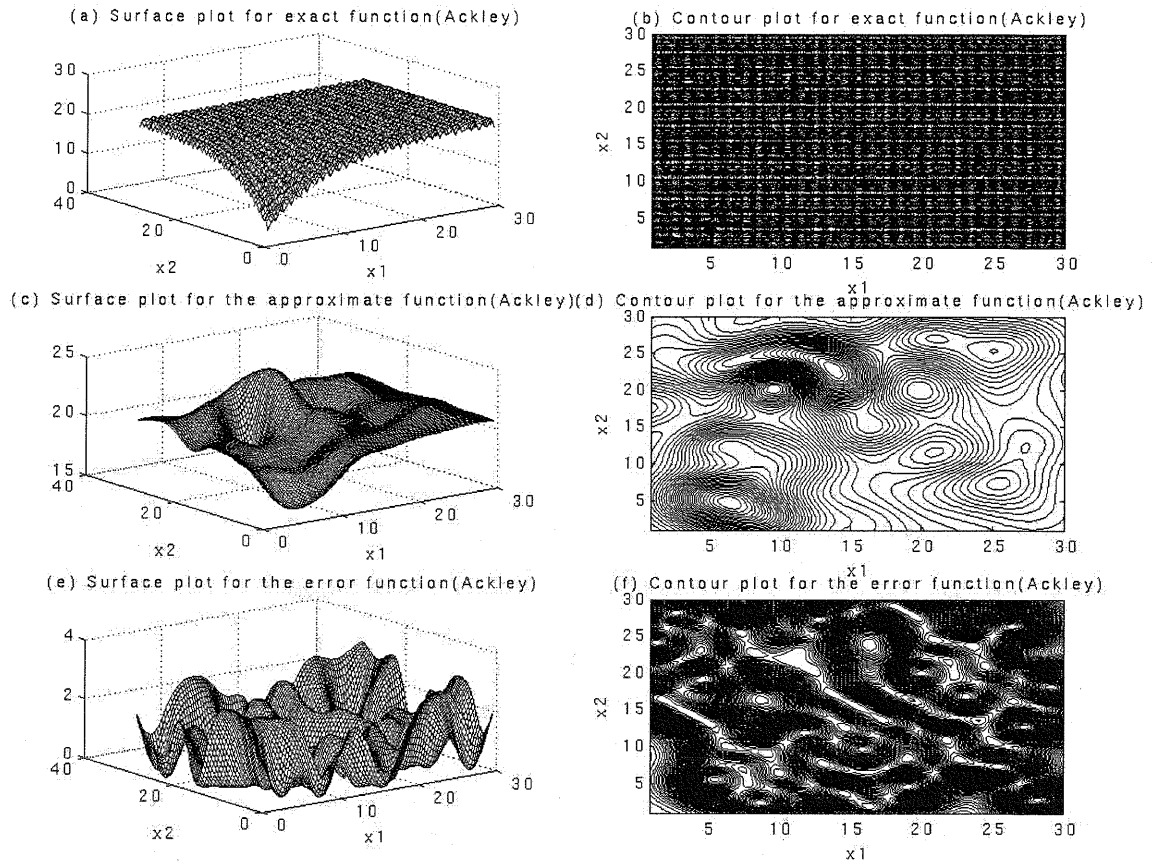


Figure 2. Surface and Contour plots of Ackley function using 40 DOE points. (a) Exact surface (b) Contour of exact function (c) Approximated surface (d) Contour of approximated surface (e) Error surface (f) Contour of error surface

Table 2. Design Variables for Engine Nacelle Geometry

Variables	x_l	x_0	x_u	Description
Scarf angle	-10	-5	25	Negative scarf angle (deg.)
Teaxis	5	12	20	Axial coordinate of top external profile (mm)
Telater	5	10	20	Radial coordinate of top external profile (mm)
Tiaxis	1.5	2	2.5	Ratio of top inner profile coordinate in axial direction against radial direction
Tilater	1	1.34	1.6	Coefficient used to determine top inner profile coordinate in lateral direction
Var_d225	22	25.4	28	Radial control length of top lip profile (mm)
Var_d226	22	25.4	28	Axial control length of top lip profile (mm)

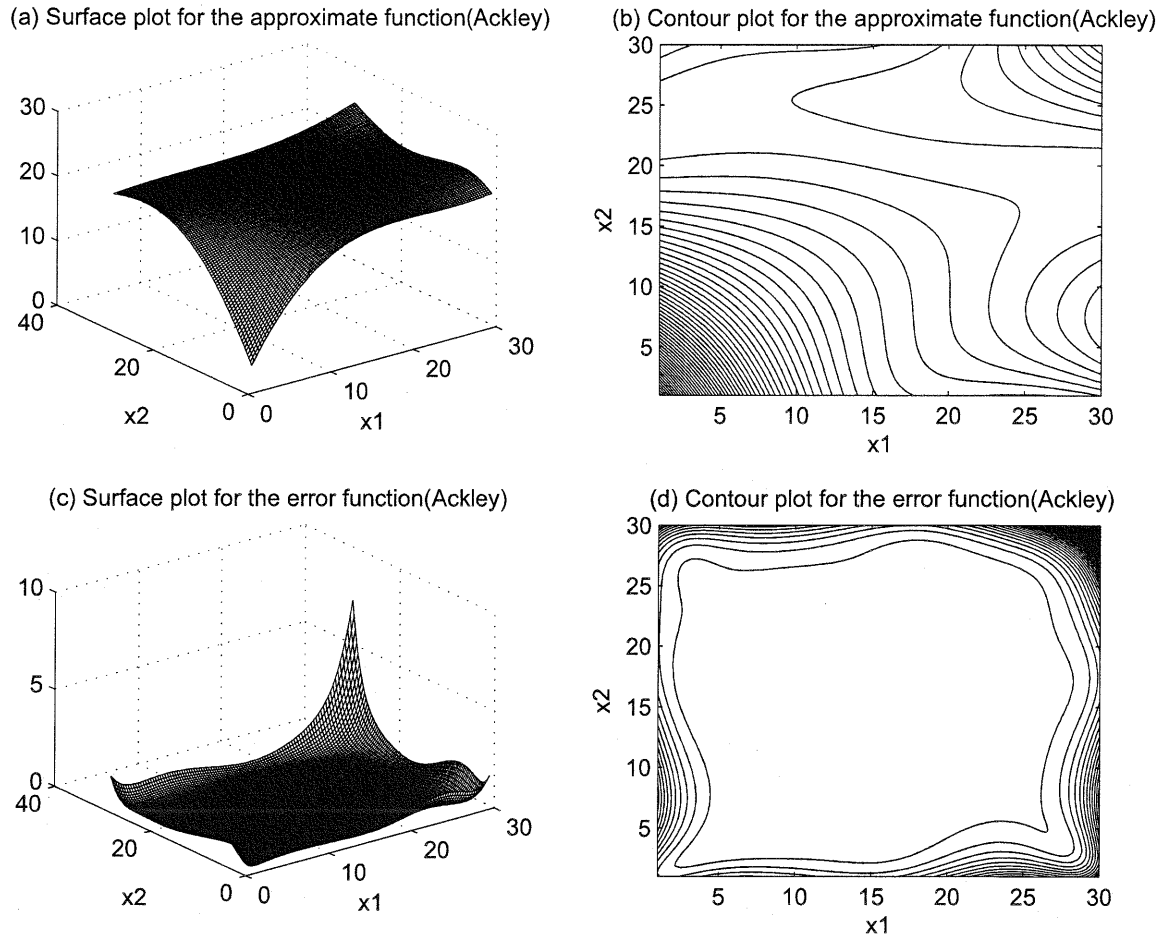
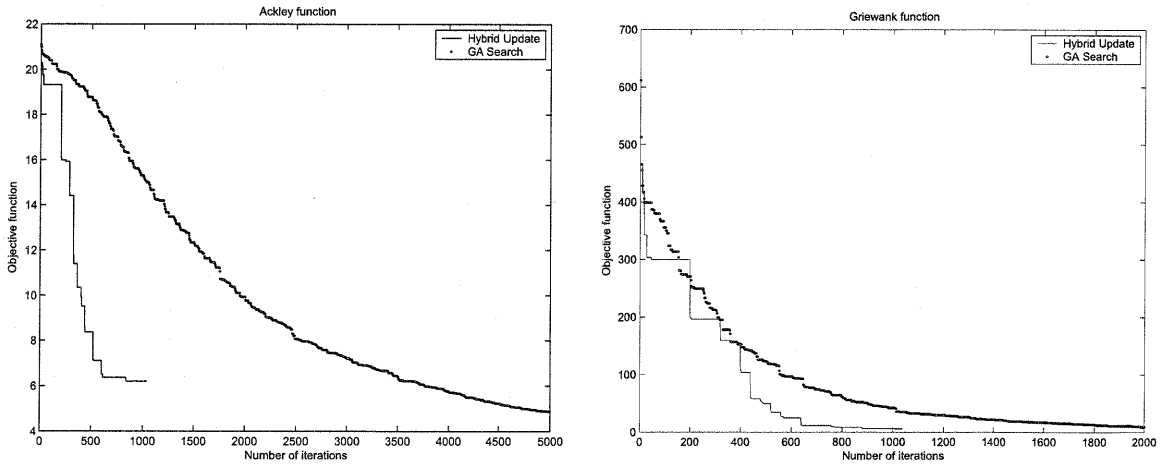
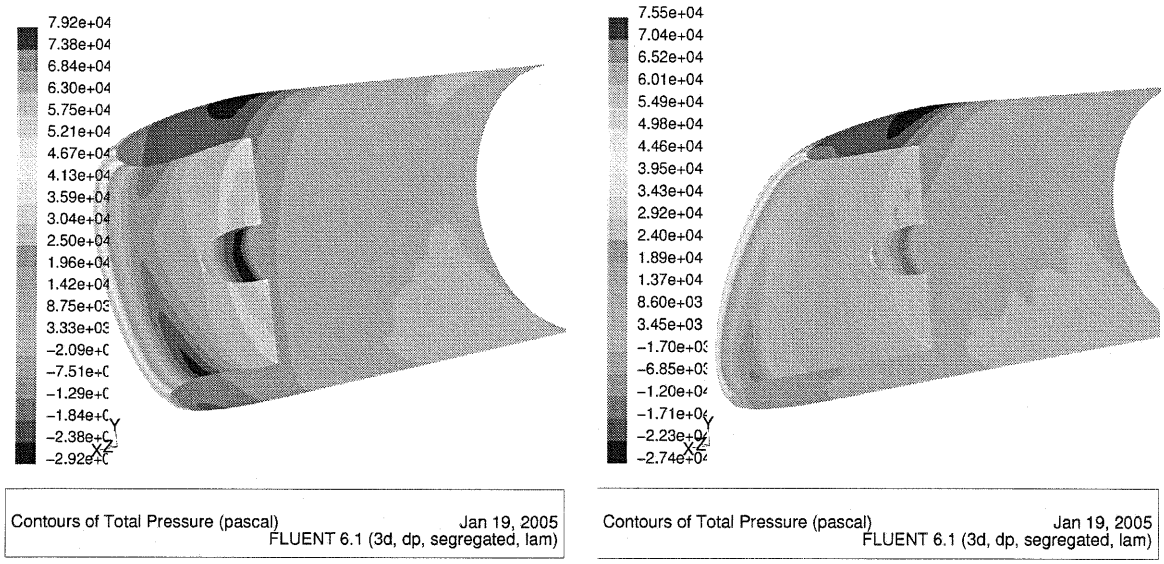


Figure 3. Surface and Contour plots of Ackley function using 10 initial DOE points plus 30 update points. (a) Approximated surface (b) Contour of approximated surface (c) Error surface (d) Contour of error surface



(a) Ackley (b) Griewank

Figure 4. Search histories: a comparison between a canonical GA and hybrid scheme on two test functions



(a) Base

(b) Final

Figure 5. Geometry and pressure distribution of base design and search result

Acknowledgement

The work described here is supported by the UK e-Science Pilot project: Grid-Enabled Optimisation and Design Search for Engineering (Geodise) (UK EPSRC GR/R67705/01). Financial support from EPSRC is fully acknowledged.

References

[1]Xu Z.-B., Leung, K.-S., Liang Y., and Leung Y., Efficiency speed-up strategies for evolutionary computation: fundamentals and fast-GAs, Applied Mathematics and Computation Vol. 142 341-388, 2003.

[2]Salami M. and Hendtlass T. A fast evaluation strategy for evolutionary algorithms. Applied Soft Computing, 2:156–173, 2003.

[3]Branke J. and Schmidt C., Fast convergence by means of fitness estimation. Soft Computing Journal, 2003.

[4]Potter M.A. and De Jong K.A., A cooperative Coevolutionary Approach to Function Optimisation, The Third Parallel Problem Solving From Nature, Jerusalem, Israel, pp. 249-257, 1994.

[5]Carpenter W. and Barthelemy J.-F., A comparison of polynomial approximation and artificial neural nets as response surface. Technical report 92-2247, AIAA, 1992

[6]Giunta A.A. and Watson L., A comparison of approximation modelling techniques: Polynomial versus interpolating models. Technical report 98-4758, AIAA, 1998.

[7]Ong Y.S., Nair P.B., and Keane A.J., Evolutionary optimization of computationally expensive problems via surrogate modeling. AIAA Journal, 41(4):687–696, 2003.

[8]Ulmer H., Streicher F., and Zell A., Model-assisted steady-state evolution strategies. In Proceedings of Genetic and Evolutionary Computation Conference, LNCS 2723, pages 610-621, 2003

[9]Jones, D.R.; Schinlau, M.; Welch, W. J., Efficient Global Optimisation of Expensive Black-box Functions, Journal of Global Optimization, Vol. 13, 1998, pp. 455-492.

[10]Jin Y., A comprehensive survey of fitness approximation in evolutionary computation. Soft Computing Journal, 2004.

[11]Sobester A., Leary, S. and Keane A.J., On the Design of optimization Strategies Based on Global Response Surface Approximation Models, Journal of Global Optimization, 2004.

[12]Sacks, J., William J. Welh, Joby J. Mitchell and Henry P. Wynn, "Design and analysis of computer experiments", Statistical Science, Vol. 4, No. 4, pp409-435, 1989.

[13]Keane, A.J., OPTIONS Design Exploration System, http://www.soton.ac.uk/~ajk, 2004