Metamodel–Assisted Evolution Strategies

Michael Emmerich[1], Alexios Giotis[3], Mutlu Özdemir[1],
Thomas Bäck[2], and Kyriakos Giannakoglou[3]

[1] Informatik Centrum Dortmund, Center for Applied Systems Analysis, Joseph von Fraunhoferstr. 20, D-44227 Dortmund, Germany
emmerich@icd.de

[2] NuTech Solutions GmbH
Martin Schmeißer Weg 15, D-44227 Dortmund, Germany
baeck@nutechsolutions.de
http://www.nutechsolutions.com

[3] National Technical University of Athens, Dept. of Mechanical Engineering
Athens, Greece
{agiotis,kgianna}@central.ntua.gr

Abstract. This paper presents various Metamodel–Assisted Evolution Strategies which reduce the computational cost of optimisation problems involving time–consuming function evaluations. The metamodel is built using previously evaluated solutions in the search space and utilized to predict the fitness of new candidate solutions. In addition to previous works by the authors, the new metamodel takes also into account the error associated with each prediction, by correlating neighboring points in the search space. A mathematical problem and the problem of designing an optimal airfoil shape under viscous flow considerations have been worked out. Both demonstrate the noticeable gain in computational time one might expect from the use of metamodels in Evolution Strategies.

1 Introduction

Evolution Strategies (ES) are a powerful tool for global optimisation in high-dimensional search spaces. However, their known weakness is that they require a high number of evaluations. Similar problems may be reported for other population–based methods, such as the widely used Genetic Algorithms. In optimisation problems with time–consuming evaluation software (applications in the field of aeronautics are typical examples) this renders the total CPU cost to be prohibitive for industrial use. A way for keeping this cost as low as possible is through the use of a surrogate evaluation tool, i.e. the so–called metamodel; a relevant literature survey can be found in [3], at least from the viewpoint of applications in aeronautics. The Metamodel–Assisted Evolution Strategies (MAES) can be applied for global optimisation with any time–consuming evaluation method, especially in industrial design optimisation. Extending a previous work by the same authors [5], a new enhanced metamodel is employed herein. In the course of evolution, the metamodel's role is to point to the most promising

J.J. Merelo Guervós et al. (Eds.): PPSN VII, LNCS 2439, pp. 361–370, 2002.

individuals that will be re–examined through the time–consuming evaluation software. The use of the metamodel prerequisites the existence of a database and procedure for optimally selecting the database subset to be used for its construction.

Various metamodels can be devised. In the past, Giannakoglou et al [4] utilized radial basis function networks as surrogate models. Other algorithmic variants that use metamodels in evolutionary optimisation can also be found in the literature (for instance, in Jin et al. [6]).

In the present study, we will extend the use of metamodels, as described in [5], by incorporating a local error estimation technique that enables the optimisation method to estimate the reliability of the approximated function values and to exploit further this information. The error estimation is based on the local density and clustering of points and on estimates of local correlations. A search criterion is used, which considers both the local error estimation and the fitness estimation. Furthermore, it will be demonstrated that the step-size self-adaptation is preserved despite the use of the metamodel. Unlike [5], which made use of small population size ES, in the present paper ES with large population size will be employed; to the authors experience, the latter is best suited for global optimisation.

This paper is organised as follows. The metamodel is first introduced and then the ES along with the search criterion based on fitness and error estimations are described. For the assessment of the proposed method, artificial landscapes and an airfoil optimisation problem will be analysed.

2 Kriging and Local Error Estimation

A metamodel approximates a multivariate function using points that have already been evaluated. It is a reasonable assumption to consider that the time required for building the metamodel is negligible compared to the CPU cost for an exact evaluation, at least in real world problems. Thus, a metamodel is to be considered as a fast surrogate model to the exact evaluation software.

Henceforth, $\mathbf{x}_1, \ldots, \mathbf{x}_m \in \mathbb{R}^n$ will denote previously evaluated candidate solutions and $\mathbf{y} = (y_1, \ldots, y_m) := (f(\mathbf{x_1}), \ldots, f(\mathbf{x_m}))$ are results from the exact evaluations associated with each one of the aforementioned solutions. Using any interpolation method, the estimation function $\hat{f}$ returns the exact value y_i, at the data sites $\mathbf{x}_i, i = 1, \ldots, m$. In this paper, the metamodel is based on Kriging techniques, which provide estimates to the fitness values of new candidate solutions. Kriging stands for an isotropic interpolation method which can deal with irregularly distributed points in the search space. We recall that an interpolation method is called isotropic if $\hat{f}$ depends exclusively on distances $\| \mathbf{x} - \mathbf{x}_i \|$ from neighbouring points instead of the absolute value of $\mathbf{x}$ and the direction $\mathbf{x} - \mathbf{x}_i$. Kriging was originated by the mining engineer Krige, who used this method to estimate ore concentrations in gold mines. Later, Kriging was formulated rigorously by Matheron [9]. In recent years it has been used in geostatistics [12] and in metamodelling and optimisation [8,11,2]. Kriging assumes that that each measured value of the objective function f is the realisation of an isotropic n-dimensional Gaussian stochastic process with unknown mean $\beta \in \mathbb{R}$ and covariance function of the form $c(\mathbf{s}, \mathbf{t}) = \sigma^2 r_\theta(\mathbf{s}, \mathbf{t})$. Here $\sigma^2 > 0$ and θ are unknown and

$$r_\theta(\mathbf{s}, \mathbf{t}) := \exp(-\theta \parallel \mathbf{s} - \mathbf{t} \parallel^2) \tag{1}$$

Note that this kernel allows to estimate values that are lower than the minimal values of $\vec{y}$. An alternative kernel function would be $r_\theta(\mathbf{s}, \mathbf{t}) := \exp(-\theta \parallel \mathbf{s} - \mathbf{t} \parallel)$. For this function the optima of the Kriging approximation are the same as the optima in the set of measured values $\{y_1, \ldots, y_n\}$ (cf. [7]).

In order to construct an approximation by Kriging, unknown parameters (β, σ, θ) have to be estimated by the maximum likelihood method. This is done by solving a minimisation problem with a local search method.

$$\begin{aligned} n \log \hat{\sigma}^2(\hat{\theta}) + \log\det \mathbf{R}(\hat{\theta}) &\rightarrow \min \\ \hat{\beta} &= [\mathbf{I}^T \mathbf{R}(\hat{\theta}) \mathbf{I}]^{-1} \mathbf{I}^T \mathbf{R}(\hat{\theta})^{-1} \mathbf{y} \\ \hat{\sigma}^2(\hat{\theta}) &= \frac{1}{n} [\mathbf{y} - \mathbf{I}\hat{\beta}]^T \mathbf{R}(\hat{\theta})^{-1} [\mathbf{y} - \mathbf{I}\hat{\beta}] \\ \mathbf{R}(\hat{\theta}) &:= [r_{\hat{\theta}}(\mathbf{x}_i, \mathbf{x}_j)] \end{aligned} \tag{2}$$

Once $\hat{\theta}$ has been obtained, estimations of function values at new points can be computed as follows

$$\hat{f}(\mathbf{x}) := \hat{\beta} + (\mathbf{y} - \mathbf{I}\hat{\beta})^T \mathbf{R}(\hat{\theta})^{-1} \mathbf{r}(\mathbf{x}; \hat{\theta}), \text{ with } \mathbf{r}(\mathbf{x}; \hat{\theta}) := [\mathbf{r}_{\hat{\theta}}(\mathbf{x}_i, \mathbf{x})] \tag{3}$$

The mean squared error of this estimation is estimated as follows

$$\hat{\text{MSE}}(\mathbf{x}) = \sigma^2 - \sigma^2 [\mathbf{I}; r(\mathbf{x}, \hat{\theta})^T] \begin{bmatrix} 0 & \mathbf{I} \\ \mathbf{I} & \mathbf{R}(\hat{\theta}) \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{I} \\ \mathbf{r}(\mathbf{x}; \hat{\theta}) \end{bmatrix} \tag{4}$$

The value of $\hat{\text{MSE}}$ depends on the correlation of the landscape as well as on the local density of points. This has been illustrated in Figure 1. From this figure it comes out that the approximation is precise at the points that have been evaluated and more precise in regions with a high point density. In this region the $\hat{\text{MSE}}$ is low. The approximation is also precise, if a point lies between two data points for which exact measurements exist. This illustrates that Kriging takes into account the clustering of points. The right part in Figure 1 f shows a sinusoidal function with doubled frequency with the same points $\mathbf{x}_i,\ i = 1, \ldots, 8$ being evaluated. In this case the correlation between neighbouring points is much weaker, which worsens the quality of the function approximation. The increased difficulty has an effect on the $\hat{\text{MSE}}$ prediction too, which is much more pessimistic in that case.

Although it has been stated that the metamodel's CPU cost could be safely neglected in real word optimisation problems with time–consuming evaluation methods, the real CPU cost of the Kriging method depends mainly on the number of evaluated sites and not so much on the dimension of the search space. In order to estimate θ, repetitive inversions of the covariance matrix $\mathbf{R}(\hat{\theta})$ are needed.

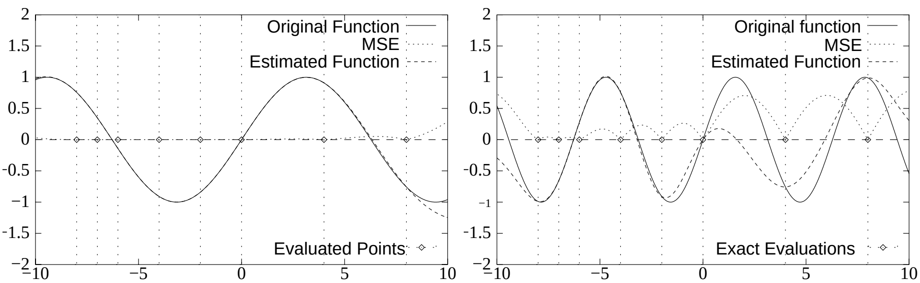

Fig. 1. Interpolation with Kriging: Original function, approximation and error-estimation for the one-dimensional function $\sin(x)$ (left) and $\sin(2x)$ (right).

The CPU cost for this inversion is in $\mathcal{O}(m^3)$ and this determines the asymptotic time complexity of the metamodelling algorithm, which is $\mathcal{O}(N_{opt}m^3 + nm^2)$ (N_{opt} is the number of iterations for the local minimisation of the MSE with $N_{opt} \approx 200$). Note that the time complexity for calculating $\hat{MSE}$ and $\hat{f}$ is in $\mathcal{O}(nm^2)$, once the metamodel has been built.

The Kriging metamodel used in this work is based on the k nearest neighbours at each point, so it will be referred to as *local metamodel*. The value of k has been set to 20. Any further increase in k seems to slightly improve the results but, at the same time, increases the computation time significantly. The Kriging algorithm is the one proposed by Padula et al. [7] with pseudo inversion of the covariance matrix. One building and evaluation of the local metamodel takes about 0.2 seconds on a Pentium III, 1GHz PC. Failed evaluations are treated by penalizing them with the worst feasible value multiplied by 10. It is also recommended to increase k for highly dimensioned search spaces or in case that the exact evaluation tool is time–consuming.

3 Metamodel Assisted Evolution Strategies

It is known that ES are powerful and robust optimisation tools. They are established as standard tools in practical optimisation. In an ES, parameter vectors (search points) $\mathbf{x} \in \mathbb{R}$ together with one or many step-size parameter(s) $\sigma \in \mathbb{R}^+$ form an individual. A set of individuals is termed a population. Modern (μ, κ, λ)-ESs usually work with increased populations. They can be easily adapted to different computing environments and representations. In this study they will be used for continuous parameter optimisation.

The (μ, κ, λ)-ES has first been applied by Schwefel [10]. Within any generation, mutation and recombination operators are applied to generate λ offspring from μ parents. In order to form a new generation, the best from the $\mu + \lambda$ individuals are selected. Individuals that exceed the age of κ generations are eliminated from the selection procedure.

Throughout this study we employed the mutative self adaptation suggested by Ostermeier [10] with a single global step-size. According to [10], a discrete recombination operator has been used for the object variables and an intermediate recombination for the step-size parameter. A recommended strategy variant for complex multimodal problems is the $(15, 5, 100)$-ES. Thanks to the large parent population size, the recombination operator becomes beneficial. Furthermore, the selection pressure $\lambda/\mu \approx 7$ is high enough to enable the self-adaptation of step sizes. The maximum life time of 5 generations makes the strategy robust, even in the presence of discontinuities.

In this work, the metamodel will be used to accelerate the (15, 5, 100)-ES. As in Trosset and Torczon [11], direct optimisation in continuous spaces will be combined with metamodelling techniques. A fitness criterion based on both the estimated value and the estimated local variance of the prediction model is used by setting up a search criterion to estimate the potential outcome of a computer experiment:

$$S_c(\mathbf{x}) := \hat{f}(\mathbf{x}) - w\sqrt{M\hat{S}E(\mathbf{x})}. \tag{5}$$

In the MAES, this criterion S_c is used to pre-select s $(0 < s < \lambda)$ individuals out of λ offspring, in order to evaluate them exactly. Only the individuals with the lowest values for S_c are chosen for this. The behaviour for $w = 1$ and for $w = 0$ will be studied. In the latter case, S_c reduces to $S_{c,w=0}(\mathbf{x}) = \hat{f}(\mathbf{x})$. The difference between both criteria is that with $w = 0$ the most promising candidate solutions are selected whereas with $w = 1$ this concept is extended to still unexplored search areas, by considering the estimation error and additionally selecting the individuals with a potential of good performance. Thus, it is recommended to use S_c in order to make the whole algorithm more robust.

The parameter w can be used to increase the influence of the error term. Its value should be increased in complex problems. On the other hand, the algorithm will converge faster to a local optimum if w is low. The value of $w = 1.0$ is used as default throughout this study.

The number of individuals that are selected for exact evaluations is an important parameter. About 10 exact evaluations per generation is a figure that turned out to perform well. It allows to get enough iterations for making the step-size adaptation work and also to have a sufficient amount of new information for areas of the search space, which are of interest in the forthcoming iteration. Over and above to the individuals which are pre-selected by the search criterion(s), individuals which outperform the so-far best individual are also examined through the exact evaluation tool. By this measure, the algorithm can hardly be trapped in artificial local optima.

With the proposed metamodel assisted selection scheme it is possible for the ES not only to learn from promising individuals but also to memorize and make use of search points with bad performance, by accessing the *long term memory* of the evolution's history.

In order to increase the metamodel's performance each run is started using a randomly initialised population of 100 individuals, which has been exactly evaluated. From this population the 15 best individuals have been selected in order to build the starting population for the ES.

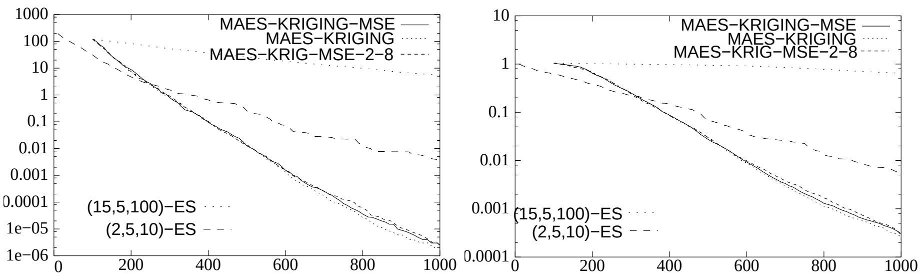

Fig. 2. Average plots of 20 runs on sphere function. Fitness value (right) and the global step-size (left) vs the number of exact evaluations.

Six different strategy variants have been compared in this study:

- $(15, 5, 100)$-ES: The canonical (15, 5, 100)-ES with random starting population of 100 individuals.
- $(2, 5, 10)$-ES: The canonical (2, 5, 10)-ES with a random starting population of 10 individuals.
- MAES-KRIGING-MSE: Metamodel–Assisted (15, 5, 100)-ES - 10 individuals are exactly evaluated per generation, selected using criterion $S_{c,w=1}$.
- MAES-KRIG-MSE-2-8: Metamodel–Assisted (15,5,100)-ES - 2 individuals are pre-selected using $S_{c,w=1}$ and 8 individuals by $\hat{f}$, per generation.
- MAES-KRIGING: Metamodel–Assisted (15,5,100)-ES - 10 individuals are pre-selected using the criterion $\hat{f}$.

4 Studies on Artificial Landscapes

In order to prove the general applicability of our approach and learn about their global and local convergence behaviour (speed, reliability), experiments on artificial landscapes have been conducted. The algorithms employed started with a randomly selected population of λ individuals, all of them exactly evaluated. The initial step-size was set to 5% of the variables range. The first test function was a simple sphere function ($\sum_{i=1}^{n} x_i^2$, $\mathbf{x} \in [-10, 10]^n \subset \mathbb{R}^n$). It has been selected to demonstrate the different local search characteristics of the strategy variants and to investigate their ability to adapt step-sizes.

The result shows that despite the large population size, it is still possible to self-adjust step-sizes by mutative self-adaptation within a comparably low number of exact evaluations. Another conclusion drawn from these computations is that the convergence behaviour is not seriously affected if the search criterion $\hat{f}$ is replaced (partly) by $S_{c,w=1}$. However, it can be seen that the strategies that only employ $\hat{f}$ as pre-selection criterion are slightly better than those also using $S_{c,w=1}$.

The second function is the multimodal Keane's Bump problem 3, which is denoted as follows:

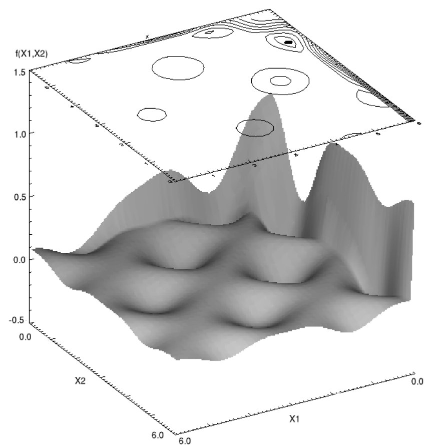

Fig. 3. Keane's function plotted in 2-dimensions for a cut of the search space. The optimum is indicated by the black spot in the upper right corner of the contour plot.

$$\min -\frac{|\sum_{i=1}^{n}(\cos^4 x_i) - 2 * \prod_{i=1}^{n}(\cos^2(x_i))|}{\sqrt{\sum_{i=1}^{n} i * x_i^2}}, \tag{6}$$

$$\prod_{i=1}^{n} x_i > 0.75, \ \sum_{i=1}^{n} x_i < \frac{15n}{2}, \ \mathbf{x} \in]0, 10[^n \subset \mathbb{R}^n$$

This problem is characterized by a nonlinear boundary and a high number of local optima. The optimal solution is located near the constraint boundary.

The global convergence behaviour of different strategies was investigated (cf. 4). The $(15, 5, 100)$-ES performs much better than the $(2, 5, 10)$-ES. This indicates that strategies with a small population size are not robust for such problems. In contrast to the previous study, it now matters which search criterion is applied. Runs using $S_{c,w=1}$ perform much better (in average) than those guided only by the function estimation with Kriging $\hat{f}$. Here we get the desired effect that the strategy concentrates not only on the most promising solutions but makes also evaluations in unexplored regions of the search space that have a high potential of containing better solutions.

In the case of complex multimodal functions it is typical that MAES based on Kriging start by yielding a wide margin and that later they are overtaken by the $(15, 5, 100)$-ES. This is contradictory to what often occurs in unimodal functions. The fact that the $(15, 5, 100)$-ES overtakes the Kriging variants in the long term, might be explained by the fact that this strategy adapts the step-size much slower and the high step-size makes it easier to escape from local optima.

Though they work with the same number of iterations the $(2, 5, 10)$-ES leads to very bad results in the Keane function problem. This should be noticed in contrast to what is often believed, viz. that only small populations lead to good results when working with time consuming evaluations. This is certainly applica-

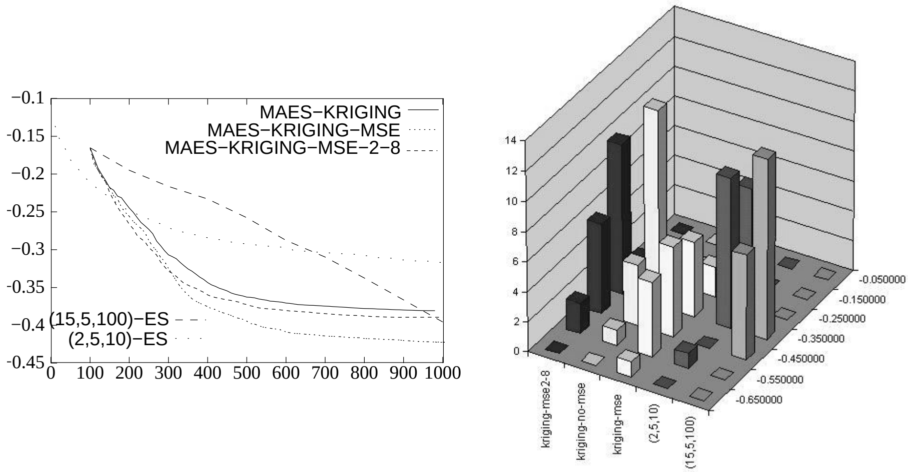

Fig. 4. Results on 20 dimensional Keane Function using the Kriging approximator: Averaged fitness histories for 20 runs (left) and a histogram for the best found values in the 20 runs after 1000 exact evaluations (right) are plotted.

ble for simple convex and unimodal functions. But EAs are not intended to solve problems, where the function topology is simple and for which other (faster) optimization tools can be recommended. ES are used in real-world application with highly nonlinear and multimodal characteristics.

5 Airfoil Shape Optimisation

This problem deals with the redesign of a 2D airfoil (the starting profile is the well known *RAE 2822*) in order to minimize the drag coeficient (C_D) and maximize the lift coefficient (C_L) at certain flow conditions. These are: $Re = 6.2\times10^6$, $M_\infty = 0.75$, $\alpha_\infty = 2.734^o$ where Re stands for the freestream Reynolds number based on the chord length, M_∞ for the freestream Mach number and α_∞ for the freestream flow angle. The transition was fixed on both sides at 3% of the chord.

For the parameterization of each shape one circle for the leading edge and two Bezier curves with five control points each have been used. The leading and trailing edge positions were fixed and the total number of design parameters was equal to 22.

The simulation tool was M. Drela's MSES analysis software [1] which at the aforementioned flow conditions yields $C_L = 0.748$, $C_D = 0.0235$. One evaluation takes about $1min$ on a Pentium III 1GHz PC. The cost function was defined as

$$F = C_D + \frac{10}{C_L} \tag{7}$$

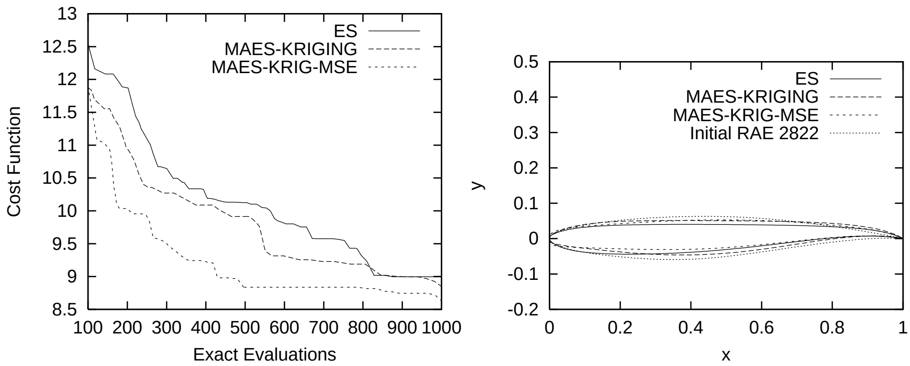

Fig. 5. Left: Average convergence histories of 3 runs of the airfoil shape optimization problem with the ES and MAES using one and both selection criteria. Right: Initial airfoil profile and the optimal ones computed using ES and MAES.

Working with a (15,5,100) strategy, average convergence histories are shown in fig.5 (left) with the conventional ES and MAES using one and both selection criteria. Each curve is the average of three optimization tasks. This plot indicates that the MAES technique is capable of reducing the computing cost compared to the conventional ES. On the right part of figure 5, the initial *RAE 2822* and the new optimal profiles are illustrated.

6 Conclusions

The use of metamodels in the context of ES–based optimisation algorithms was proved to offer economy in computing time. This economy results from the fewer exact evaluations that this method requires. It proved advantageous to use the metamodel not only for predicting the fitness value of new individuals but also for guessing the error associated with these predictions. In particular, the function estimation contributes mostly to reduction of the computational time whereas the error estimation helps to increase the global convergence reliability in complex multimodal problems. It was also proved that the metamodel does not harm the self-adaptivity properties of the method and that populations of increased size with good exploration capabilities can be used with low computing cost.

Acknowledgement

The support from bilateral Personnel Exchange Programme between Greece and Germany (IKYDA 2000) is acknowledged.

References

1. M. Drela and M.B. Giles. Viscous–Inviscid Analysis of Transonic and Low Reynolds Number Airfoils. *AIAA Journal*, 25 (10):1347–1355, 1987.
2. M.A. El-Beltagy, P.B. Nair, and A.J. Keane. Metamodelling Techniques for Evolutionary Optimisation of Computationally Expensive Problems: Promises and Limitations. In A.E. Eiben M.H. Garzon V. Honavar M. Jakiela W. Banzhaf, J. Daida and R.E. Smith, editors, *Proc. of GECCO, Int'l Conf. on Genetic and Evolutionary Computation, Orlando 1999*, pages 196–203. Morgan Kaufman, 1999.
3. K.C. Giannakoglou. Design of optimal aerodynamic shapes using stochastic optimization methods and computational intelligence. *Progress in Aerospace Sciences*, (38(1)):43–76, 2002.
4. K.C. Giannakoglou, A.P. Giotis, and M. Karakasis. Low-cost genetic optimization based on inexact pre-evaluations and the sensitivity analysis of design parameters. *Inverse Problems in Engineering*, (9):389–412, 2001.
5. A. Giotis, M. Emmerich, B. Naujoks, K. Giannakoglou, and Th. Bäck. Low cost stochastic optimisation for engineering applications. In *Proc. Int'l Conf. Industrial Applications of Evolutionary Algorithms, EUROGEN2001, Athens, GR, Sept. 2001*, Barcelona, 2001. CIMNE.
6. Y. Jin, M. Olhofer, and B. Sendhoff. Managing Approximation Models in Evolutionary Aerodynamic Design Optimisation. In *CEC 2001 Int'l Conference on Evolutionary Computation, Las Vegas*, volume 1, pages 592–599, Piscataway NJ, 2001. IEEE Press.
7. A. Padula. Interpolation and pseudorandom function generators. Senior honors thesis, University, Dept. of Computational and Applied Mathematics, Rice University, Houston, TX, 2000.
8. A. Ratle. Accelerating the convergence of evolutionary algorithms by fitness landscape approximations. In A.E. Eiben, Th. Bäck, M. Schönauer, and H.-P. Schwefel, editors, *Parallel Problem Solving by Nature*, volume V of *LNCS*, pages 87–96, Berlin, 1998. Springer-Verlag.
9. J. Sacks, W.J. Welch, W.J. Mitchell, and H.-P. Wynn. Design and analysis of computer experiments. *Statistical Science*, (4):409–435, 2000.
10. H.-P. Schwefel. *Evolution and Optimum Seeking*. Wiley, 1995.
11. M.W. Trosset and V. Torczon. Numerical optimization using computer experiments. Technical report, Institute for Computer Applications in Science and Engineering ICASE TR 9738, NASA Langley Research Center, Hampton Virgina, 1997.
12. H. Wackernagel. *Multivariate Geostatistics*. Springer Verlag, Berlin, 1998.