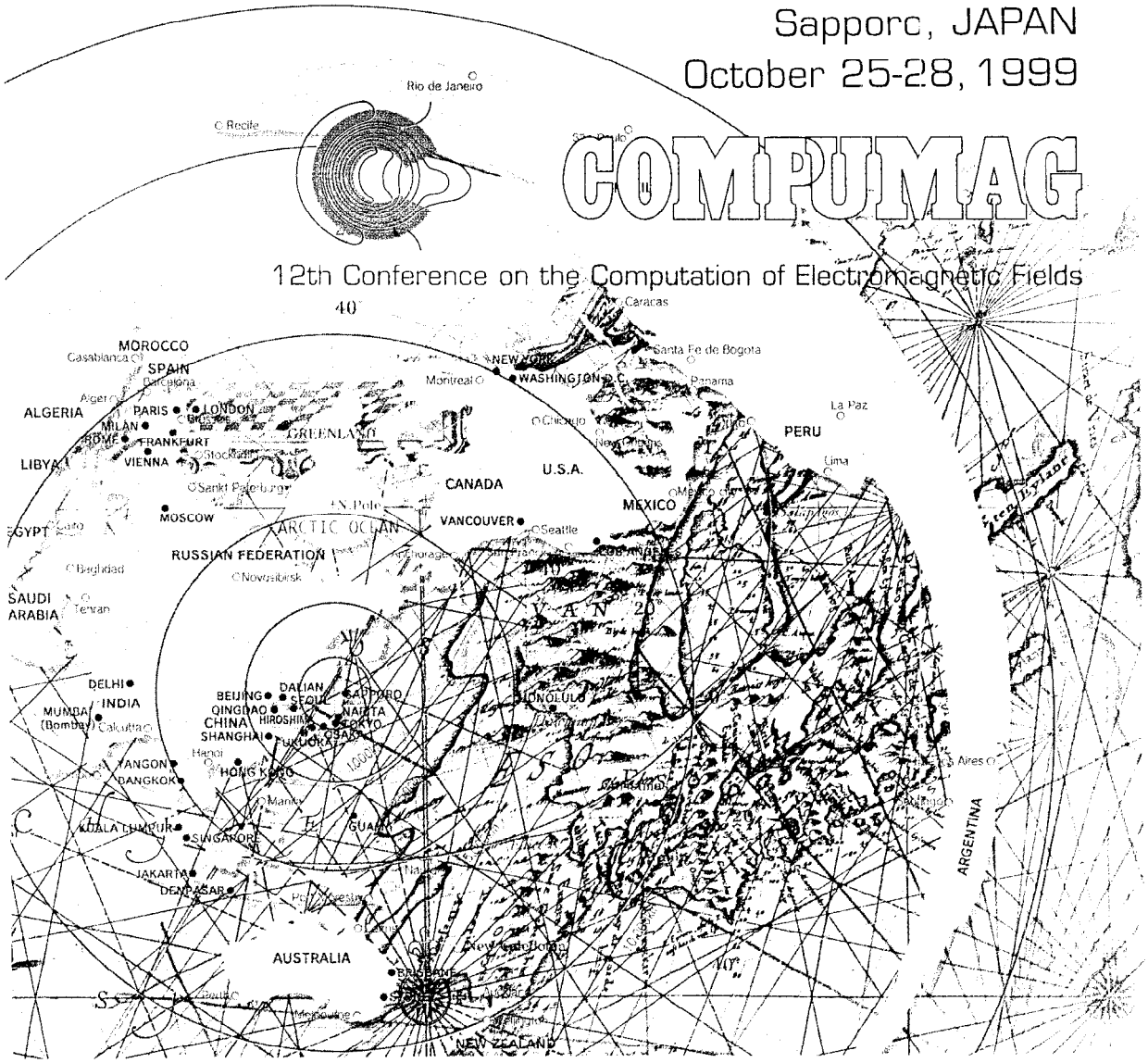
IEEE TRANSACTIONS ON

MAGNETICS

A PUBLICATION OF THE IEEE MAGNETICS SOCIETY

JULY 2000 VOLUME 36 NUMBER 4 IEMGAQ (ISSN 0018-9464)

PART I OF TWO PARTS



SELECTED PAPERS FROM THE 12TH CONFERENCE OF THE COMPUTATION OF ELECTROMAGNETIC FIELDS (COMPUMAG '99)
Sapporo, Japan, October 25–28, 1999

See p. 613, Part I, for COMPUMAG '99 Table of Contents. COMPUMAG '99 papers begin on p. 633. See Part II, p. 1962 for regular papers.

Adaptive Coupling of Differential Evolution and Multiquadrics Approximation for the Tuning of the Optimization Process

Uwe Pahner and Kay Hameyer

Abstract—**Recently, the combination of global convergent stochastic search methods with approximation schemes based on radial basis functions has been introduced. This paper presents a new approach: instead of a procedural sequencing of approximation algorithm and optimization algorithm, this optimization scheme is characterized by a direct and adaptive coupling of both algorithms. An approximation of the feasible space is constructed and updated during the progress of the evolutionary search. If the approximation fulfils particular accuracy-criteria, the evolutionary search algorithm starts sampling the approximation (indirect search) instead of directly sampling the objective function. This can lead to a significant reduction of function calls, which is desirable if the function evaluation is computational expensive (e.g. involving finite element analysis steps).**

Index Terms—**Approximation methods, numerical methods, optimization methods, search methods.**

I. Introduction

IN DIRECT search methods, the optimization algorithm samples the objective function directly, while in indirect methods the optimization algorithm is applied to an approximation of the N-dimensional feasible solution space. A trial on a fitted surface is computationally cheap if it replaces, e.g. finite element analysis. In the classical Response Surface Methodology (RSM) only one global polynomial is fitted. It becomes an optimization method if the following successive steps are performed until an accuracy criterion is met: sampling the feasibility space, constructing an approximation, finding the optimum of the approximation, constructing a new approximation closer to the optimum found. The Generalized RSM (GRSM) applies the same methodology as the RSM, except of using radial basis functions to construct the approximations [1], [2]. The presented new scheme does not have a well-defined sampling grid, it rather uses the sampling points of the evolutionary search to construct and update an approximation. Instead of constructing a highly accurate approximation in each successive step, the approximation is gradually improved during the course of the evolutionary search.

Manuscript received October 25, 1999. This work was supported by the Belgian "Fonds voor Wetenschappelijk Onderzoek Vlaanderen" and the Belgian Ministry of Scientific Research for granting the IUAP P4/20 on Coupled Problems in Electromagnetic Systems.

The authors are with the Katholieke Universiteit Leuven, Department EE (ESAT), Division ELEN, Kardinaal Mercierlaan 94, B-3001 Leuven, Belgium (e-mail: {uwe.pahner; kay.hameyer}@esat.kuleuven.ac.be).

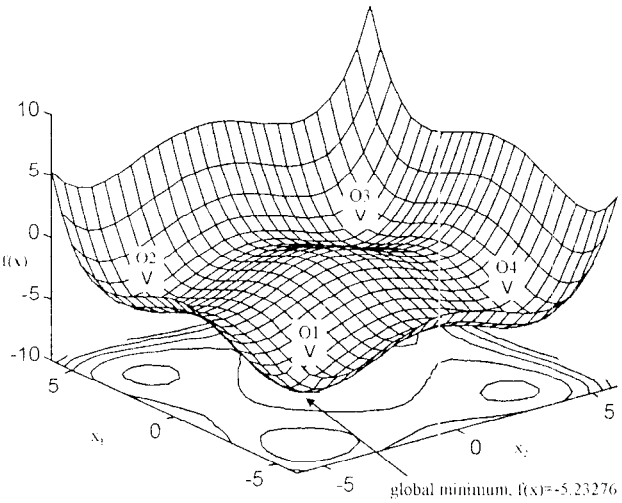


Fig. 1. Visualization of the objective function surface of the theoretical optimization problem defined by (1) and (2) with the global optimum at $x_{1,best} = x_{2,best} = -1.45377$.

II. Theoretical Optimization Problem

In order to test the presented optimization scheme, a theoretical optimization problem is defined by [2]:

$$\text{minimize } f(\boldsymbol{x}) = \sum_{i=1}^{2} \left(-0.01\left((x_i + 0.5)^4 - 30x_i^2 - 20x_i\right)\right) \tag{1}$$

subject to the constraints:

$$-6 \le x_i \le 6, \quad i = 1, 2. \tag{2}$$

There are four local minima of approximately equal value, one of them is the global minimum. Such an objective function resembles to a large extent typical objective functions found in engineering problems, which can be characterized as locally smooth, but nonconvex.

III. Multiquadric Approximation

The new feature of the GRSM is the usage of multiquadrics to achieve a nonconvex response surface [1], [2]. This nonconvexity does not permit the usage of gradient based optimization

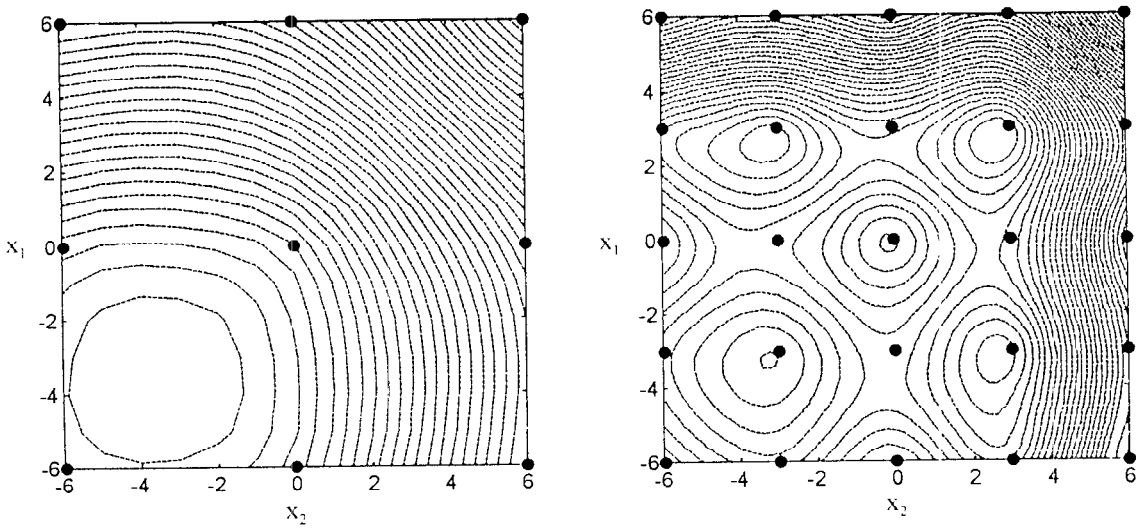


Fig. 2. (a) Multiquadric approximation based on a) full 3^N (9 samples) and (b) full 5^N-factorial (25 samples) of the analytical test function [2].

instead [1]. The GRSM uses approximations of the objective function at any point $\boldsymbol{x}_i$ of the form:

$$f(\boldsymbol{x}_i) = \sum_{j=1}^{M} c_j h(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|) \tag{3}$$

with c_j the approximation coefficients, M the number of experiments and a possible radial basis function $h(\|\boldsymbol{x} - \boldsymbol{x}_j\|)$, here chosen to be:

$$h(\|\boldsymbol{x} - \boldsymbol{x}_j\|) = \sqrt{\|\boldsymbol{x} - \boldsymbol{x}_j\|^2 + s}. \tag{4}$$

The shift factor s is a parameter defining the curvature of the approximated N-dimensional surface.

Alotto *et al.* proposes in [2] a statistical method (Bootstrapping) to defines in a near optimal way. Experiments have shown that choosings smaller than the average spacing of the sample points is sufficient for most applications. If one substitutes the interpolation condition $f(\boldsymbol{x}_i) = y_i$ in (3), the matrix equation for the unique coefficients c_j is obtained:

$$\boldsymbol{H}\boldsymbol{c} = \boldsymbol{y}. \tag{5}$$

with the coefficients of the matrix $H_{ij} = h(\|x_i - x_j\|)$. The matrix H is a full matrix with all diagonal elements equal to zero. As long as the number of sample points is relatively small (up to a few hundred points) and singularities due to duplicate points are avoided, this system can be solved by methods with a full pivoting scheme. The advantage of the multiquadrics approximation of nonconvex functions arises with higher factorial designs when using Design of Experiment (Fig. 2) or by accumulating the sample points in successive zooming steps. More detail of the original curve is present in the approximation, however at the expense of significantly more sample points.

IV. Differential Evolution Strategy

Differential evolution (DE) [3], [4] is a rather recent approach for the treatment of real-valued nonconvex optimization problems. As is typical for stochastic search algorithms, differential evolution does not require any prior knowledge of the variable space, nor of the derivatives of the objective functions toward the design variables. This algorithm is very simple, requiring only two control parameters and is inherently parallel. Differential evolution is a self-adaptive evolution scheme (σSA). Consider a N-dimensional vector of design variables x:

$$\boldsymbol{x} = (x_0, x_1, \cdots, x_{N-1})^T. \tag{6}$$

In the initial step, a population of size λ of randomly chosen designs $\boldsymbol{x}_i$ is constructed in such a way, that the initial population covers the entire parameter space uniformly. Practically, this could be achieved by defining an initial step length of design variations δ_0, being randomly applied to a given start design $\boldsymbol{x}_0$:

$$\boldsymbol{x}_{m,i}^{(0)} = \boldsymbol{x}_0 - \delta_0 + 2 \cdot \delta_0 \cdot s_i(0, 1), \tag{7}$$

with $m = 1(1)\lambda$, $i = 1(1)N$ and the randomly chosen $s_i \in [0, 1]$. During the optimization, DE generates new parameter vectors by adding the weighted difference between a defined number of randomly selected members of the previous population to another member. In its basic strategy, this is the difference of two vectors added to a third one:

$$\boldsymbol{v}_m^{(k+1)} = \boldsymbol{x}_{r1}^{(k)} + \alpha \cdot \left(\boldsymbol{x}_{r2}^{(k)} - \boldsymbol{x}_{r3}^{(k)}\right), \tag{8}$$

for $m = 1(1)\lambda$with k the generation index, $r_1, r_2, r_3 \in [1, \lambda]$, randomly chosen and mutually different and $\alpha \in \mathbb{R}$, $\alpha > 0$. To increase diversity in the population, crossover is introduced, leading to a new parameter vector of the form:

$$x_{m,i}^{(k+1)} = \begin{cases} v_{m,i}^{(k+1)} & \text{for } i = \langle n \rangle_N, \langle n+1 \rangle_N, \cdots, \langle n+L-1 \rangle_N \\ x_{m,i}^{(k)} & \text{for all other } i \subset [1, N] \end{cases}. \tag{9}$$

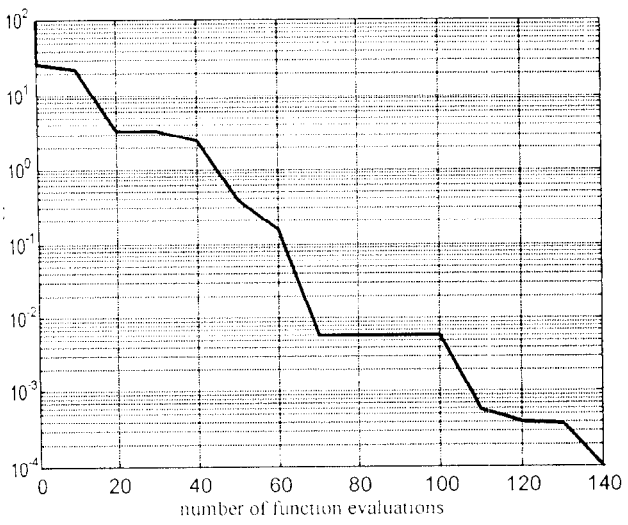


Fig. 3. Convergence of the absolute error of the best trial per iteration using differential evolution (DE/best/1-strategy with $\lambda = 10$, $\alpha = 0.5$ and $p_c = 0$).

The brackets $\langle\rangle_N$ denote the modulo function with modulus N. The index n is a randomly chosen integer from the interval $[1, N]$. L defines the number of parameters which are to be exchanged and is taken from the interval $[1, N]$. L is chosen in such a way, that the probability $P(L \geq \nu) = (p_c)^{\nu-1}$, $\nu > 0$, with the crossover probability $p_c \in [0, 1]$. The new parameter vector $\boldsymbol{x}_m^{(k+1)}$ is checked for violation of any constraints. If it violates one of the constraints, this parameter vector is rejected and a new vector is generated. The selection process has now similarity to a tournament selection process. If this resulting design vector yields a better value of the objective function than its predecessor $\boldsymbol{x}_m^{(k)}$, the new design replaces the old one in the population. If not, the old vector is retained. Several variants of this algorithm can be defined, depending on the choice of the vector to be perturbed, the number and choice of parameter vectors considered for the computation of the difference vector and the crossover method [4]. A good choice for technical problems is a strategy that increases the greediness of the algorithm by using the best parameter vector from the previous population:

$$\boldsymbol{v}_m^{(k+1)} = \boldsymbol{x}_{best}^{(k)} + \alpha \cdot \left(\boldsymbol{x}_{r1}^{(k)} - \boldsymbol{x}_{r2}^{(k)} \right). \tag{10}$$

The L_2-norm of the difference vector between each population member and its predecessor is taken as the stopping criterion. The experiments conducted in the scope of this work have indicated that the (DE/best/1)-strategy defined by (10) is favored for most technical problems [5]. The two remaining strategy parameters can be chosen as $\alpha = 0.5$ and $\gamma = 0.9$. The influence of the population size has been found to be less critical than in other σSA schemes. A minimum of $\lambda > 15$ should be chosen always if $N > 2$. However, problems with up to 25 parameters have been successfully solved with population sizes between 30–40 [5], [8]. The results of applying DE to the test problem (1), (2) are presented in Fig. 3. The strategy using $\lambda = 10$ is successful

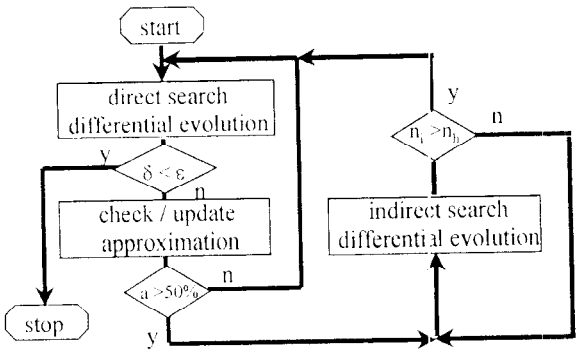


Fig. 4. Flowchart of the basic steps of the new method.

V. Adaptive Coupling

The basic idea of the new optimization method is to combine the above mentioned features of both methods to reduce the overall optimization time. The basic steps of this novel scheme are (Fig. 4):

1) Start an evolution strategy optimization using the objective function (direct search iteration).
2) Construct a multiquadrics approximation $f'(\boldsymbol{x})$ after each iteration, including only points within a defined radius from the active optimum. This radius is a function of the active step length, ensuring an automatic contraction of the active approximation space.
3) Compute the next evolutionary iteration using the objective function, but determine the predicted value from the approximation $f'(\boldsymbol{x})$ as well.
4) Record each point within a maximal radius $k \cdot \delta$ from the active optimum and construct an updated approximation $f'(\boldsymbol{x})$. If more than a defined ratio of predicted experiments are accepted during an iteration, go to step 5, otherwise return to step 3.
5) Start a new evolutionary iteration, but now using the approximation function $f'(\boldsymbol{x})$ only (indirect search iteration).
6) Depending on the acceptance ratio determined in step 4, continue an adaptive number of evolutionary iterations using the approximation $f'(\boldsymbol{x})$ only.
7) Stop if the stopping criterion is fulfilled, if not return to step 3 and start the direct search again.

Three levels of adaptivity determine the algorithm:

1) The contraction or zooming of the approximated region is adaptive to the progress of the optimization by considering a search space with a maximum radius of $k\delta^{(I)}$. The factor k is empirically chosen as:

$$k = k_1(\alpha) \cdot 10.0 \tag{11}$$

with α the step length factor of the evolution strategy.

2) The acceptance of the approximation check is determined based on the variance of the objective function value of the iteration underlying the active approximation:

$$\left(\quad m \quad / \quad , \quad m \quad \right)^2$$

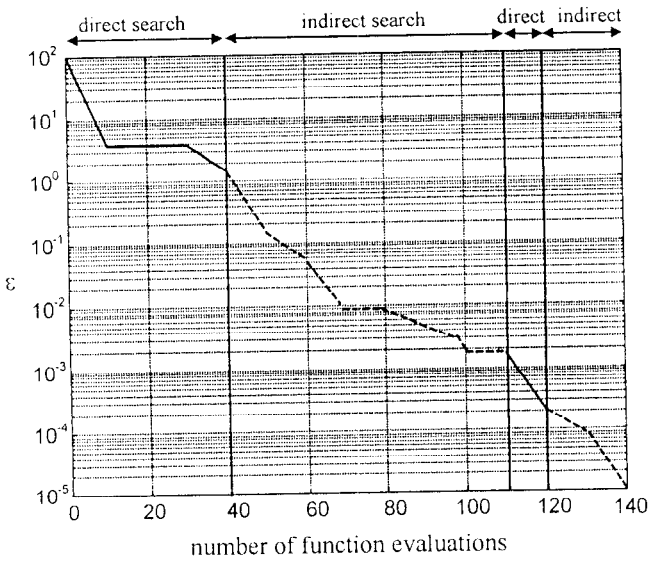


Fig. 5. Typical convergence of the error using the proposed novel method.

with $n_l = 1(1)\lambda$, λ the population size of the evolutionary search.

3) The number of indirect search iterations only depends on the acceptance ratio of the active approximation. A higher acceptance ratio allows a larger number of iterations on $f'(\boldsymbol{x})$. Tests have yield the following determination of the number of indirect search iterations n_h:

$$n_h = \left(\frac{\frac{n_a}{\lambda} - 0.5}{0.5} \right)^2 \cdot 10 + 1 \qquad (13)$$

with n_a the number of accepted trials per iteration.

The contraction of the approximation region is primarily necessary to reduce the number of unknowns in the matrix equation (5). Furthermore, when the evolutionary algorithm returns from an indirect search iteration, the selection and mutation generates new population members entirely based on members originating from an approximated function. Due to the approximation error, the approximated function might return lower values then the global minimum of the real objective function. The greedy criteria inherent to all evolutionary algorithms could cause these approximated results to survive. To prevent this development, all population members of the last approximated iterations must be mortal within the first direct sampling iteration. This causes the new method to require slightly more steps than the classical DE in order to converge toward the optimum. However, a large number of steps are taken on the computationally inexpensive approximation surface.

The performance of the scheme is demonstrated on a test example (Fig. 5). Using the same DE-strategy settings as for the direct approach, only 43% of the direct objective function evaluations are required (Figs. 3, 5). A remarkable feature of the proposed algorithm is that the efficiency depends on the curvature of the objective function. Tests with simple second order feasible surfaces have shown reductions of objective function calls of up to 80%.

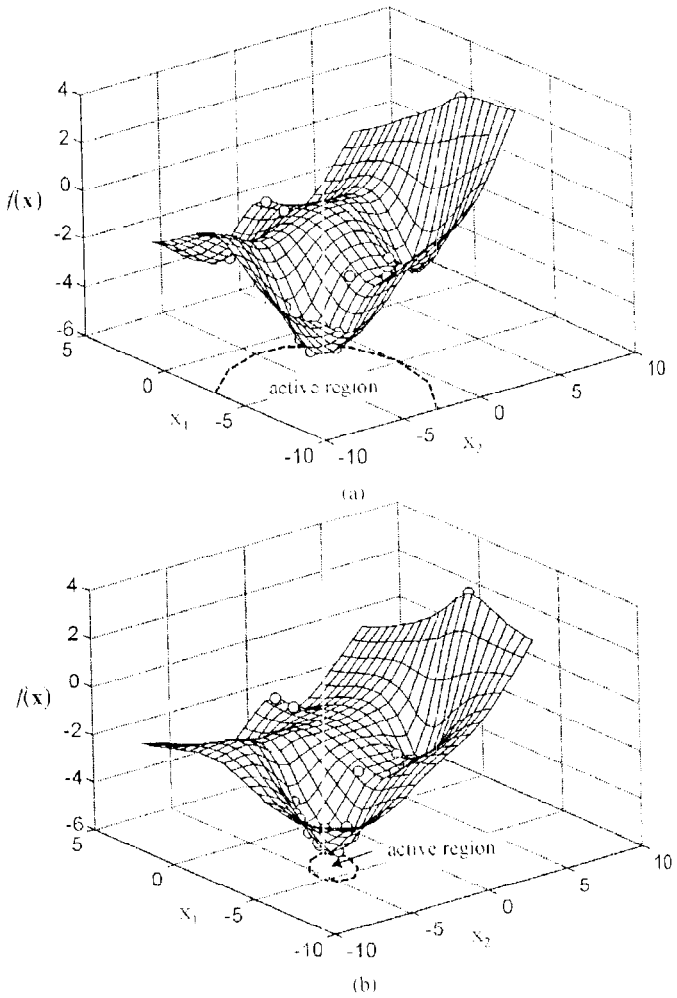


Fig. 6. Visualizations of the updated approximations after (a) 2 and (b) 4 iterations of the differential evolution algorithm ($\lambda = 10$) with the active search region indicated [8].

Further studies are required to increase the robustness of the algorithm by finding better criteria for the adaptivity of the algorithm. Some possible treatments of constraints have been outlined for the GRSM by Ebner [7]. So far, the new method performs with good results only for low dimensional optimization problems. This has been reported for the GRSM as well and remains an active research topic [7].

VI. Conclusions

A new optimization scheme has been introduced, featuring an adaptive coupling of the differential evolution strategy and multiquadric function approximation. A remarkable reduction of computationally expensive objective function calls is the result. The three levels of adaptivity provide control over progress dependent accuracy and computational expense of the entire approach.

References

[1] P. G. Alotto, A. Caiti., G. Molinari, and M. Repetto, "A multiquadrics-based algorithm for the acceleration of simulated annealing optimization procedures," *IEEE Trans. Magn.*, vol. 32, no. 3, pp. 1198–1201, 1996.

[2] P. G. Alotto, M. Gaggero, G. Molinari, and M. Nervi, "A design of experiment and statistical approach to enhance the generalized response surface method in the optimization of multiminima problems," *IEEE Trans. Magn.*, vol. 33, no. 2, pp. 1896–1899, 1997.

[3] R. Storn and K. Price, "Minimizing the real functions of the ICEC'96 contest by differential evolution," in *Proceedings of the International Conference on Evolutionary Computation*, Nagoya, Japan, 1996.

[4] ——, "Differential evolution—A simple and efficient adaptive scheme for global optimization over continuous spaces," ICSI, http://http.icsi.berkeley.edu/~storn/litera.html, 1992.

[5] U. Pahner, R. Mertens, H. De Gersem, R. Belmans, and K. Hameyer, "A parametric finite element environment tuned for numerical optimization," *IEEE Trans. Magn.*, vol. 34, no. 5, pp. 2936–2939, Sept. 1998.

[6] U. Pahner, "A General Design Tool for the Numerical Optimization of Electromagnetic Energy Transducers," Ph.D. dissertation, KU Leuven ESAT/ELEN, May 6, 1998.

[7] Th. Ebner, Ch. Magele, B. R. Brandstätter, M. Luschin, and P. G. Alotto, "Approximation of the objective function: multiquadrics versus neural networks," in *Proc. of the 8th IGTE Symposium on Numerical Field Calculation in Electrical Engineering*, Graz, Austria, Sept. 21–24, 1998, pp. 26–32.

[8] K. Hameyer and R. Belmans, *Numerical Modeling and Design of Electrical Machines and Devices*. Ashurst, UK: WIT Press, 1999, pp. 206, 215–218, 224–228.