A MULTIPOPULATION DIFFERENTIAL EVOLUTION ALGORITHM FOR MULTIMODAL OPTIMIZATION

Daniela Zaharie

Department of Computer Science, Faculty of Mathematics, West University of Timisoara
bv. V. Parvan, no. 4, 300223 Timisoara, ROMANIA
tel./fax.: +40-256-494002, e-mail:*dzaharie@info.uvt.ro*

Abstract. The aim of this work is to analyze the ability of a multipopulation differential evolution to locate all optima of a multimodal function. The exploration is assured by a controlled initialization of the subpopulations while a particular differential evolution algorithm assures the exploitation. To avoid the necessity of specifying a niche radius a multi-resolution approach is proposed. All located optima are stored in an archive that plays also the role of a communication buffer between subpopulations.

Keywords: multipopulation models, multimodal optimization., differential evolution, speciation, archive, sharing function,

1.Introduction. In practical applications of optimization techniques may appear the necessity of locating not only one optimum but also the entire set of global/local optima. Such problems are called multimodal optimization problems.

Classical Evolutionary Algorithms (EA) based on evolving a population of candidate solutions through stochastic nature-inspired operations (selection, recombination, mutation and reproduction) have been proven effective in locating the global optima. To extend this effectiveness to the multimodal optimization, different techniques that modify the behavior of classical EAs have been proposed. The main aim was to impose the formation of some so-called niches in population such that each niche could identify an optimum.

There exist two broad categories of niching techniques (Mahfoud, 1995): (i) iterative methods (Beasley et al., 1993); (ii) parallel subpopulation models (Bessaou et al., 2000), (Li et al., 2002). The niching mechanisms can be implemented within the evolutionary operators (Petrowski, 1997):

- *Selection - select the individuals which will participate to the reproduction process*. The most known technique which interferes with the selection step is the fitness sharing method based on the idea of modifying the fitness of individuals by using a so-called sharing function which depends on the distances between individuals and on a sharing parameter related to the niche radius. The derating technique used in sequential niching (Beasley, 1993) also influences the selection step.
- *Reproduction - generates new individuals from those in the current population.* An intuitive way of conserving the niches is to allow recombination only between similar individuals (belonging to the same niche). This is known as the restricted mating technique and is naturally implemented in multipopulation models.
- *Substitution - selection of survivors.* At this step the most used technique is that of crowding based on which an offspring replaces the most similar parent (Mahfoud, 1995).

These techniques have the advantage that they can be applied in combination with any EA. However the particularities of a given EA cannot be entirely neglected since sometimes EAs properties allow reducing the complexity of the niching techniques.

The Differential Evolution (DE) algorithm (Storn and Price, 1995) has been proven as an efficient and robust global optimization technique and due to its simplicity it is largely applied in design problems. DE has also been successfully extended to multiobjective optimization (Abbas, 2002), (Madavan, 2002), (Zaharie and Petcu, 2003).

The aim of this paper is to analyze the applicability of DE to multimodal optimization. In our approach we had in mind some aspects: (i) to use the particularities of DE, mainly its fast convergence and the influence of its parameters on the convergence behavior; (ii) to use few control parameters; (iii) to avoid a global processing step where the entire population is gathered (e.g. the clustering step). The search space exploration is assured by a controlled initialization of the subpopulations while a particular differential evolution algorithm assures the exploitation. To avoid the necessity of specifying a niche radius we propose the use of a multi-resolution approach: the multipopulation differential evolution is repeatedly applied for different values of a resolution factor, thus the algorithm consists in iterated search epochs (therefore it can be viewed as a hybridization between parallel and sequential niching). The optima located at each epoch are stored in an archive that plays also the role of a communication buffer between subpopulations. The redundancy is avoided by derating the initialization distribution based on the optima already found and by using a heuristic technique to eliminate the false optima (Ursem, 1999).

The paper is organized as follows. In Section 2 is presented the multipopulation differential evolution algorithm and its properties while in Section 3 the idea of the multiresolution approach and the proposed algorithm are presented. Numerical results on some test functions are presented in Section 4 and some comparisons with other multimodal optimization techniques are made. Section 5 concludes the paper.

2. The multipopulation differential evolution (MDE). The idea of dividing a population in subpopulations has been used in evolutionary algorithms with different aims: diversity preserving, parallel implementation and explicit speciation. When explicit speciation is aimed the subpopulations formation is a dynamic process of population reorganization based on distances between individuals. This process means to periodically gather all subpopulations in a single population and to split it into new subpopulations based on some clustering algorithm. The main disadvantage of such an approach is induced by the cost of the clustering process applied to the entire population.

Our approach is based on the idea of avoiding operations involving the entire population and of using the particularities of the evolutionary algorithm, e.g. the differential evolution. The general structure of the multipopulation differential evolution (MDE) is presented in figure 1.

```
Initialization of subpopulations: $P_1,P_2,\ldots,P_s$
Repeat
    Apply DE to $P_1,P_2,\ldots,P_s$ for a given number of generations
    Apply migration
Until all subpopulations converged
Add the best element of each subpopulation to an archive
```

Figure 1. The general structure of the MDE algorithm

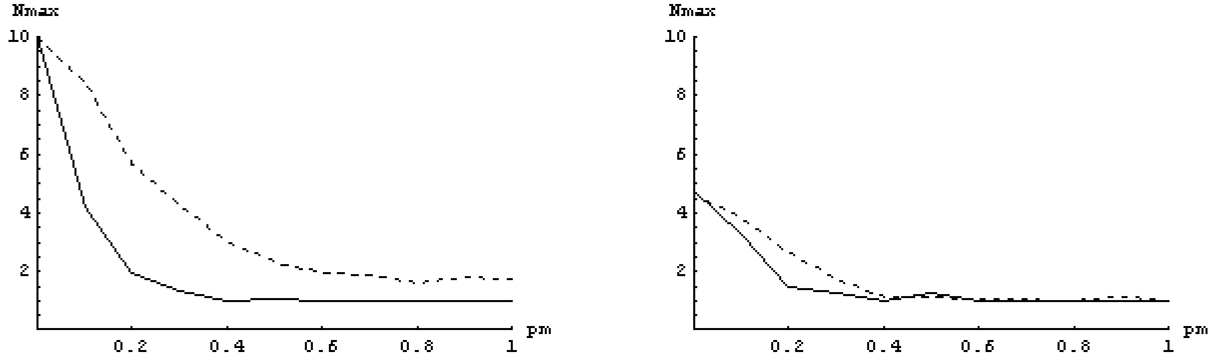
Subpopulation's initialization. All subpopulations have the same number of elements (m). Their elements are initialized in a controlled manner. Unlike the other aspects of parallel niching techniques the subpopulations initialization has attracted much less attention. However a controlled initialization could improve the ability of the algorithm to detect simultaneously many optima. In the absence of a clustering process during the evolution that would reorganize the subpopulations, the initialization step plays an important role in guiding the subpopulations toward different optima and thus reducing the search redundancy. Let us consider that the objective function domain is $D=[a,b]^n$. The initialization is based on a decomposition of D in subdomains: each subpopulation will be initialized with random elements selected from a subdomain. The subdomains are defined based on a discretization steplength called *resolution factor, r*. For instance $r=(b-a)/s^{1/n}$ and the subpopulation P_i will be initialized in $D_i=[a_1^i,b_1^i]\times[a_2^i,b_2^i]\times\ldots\times[a_n^i,b_n^i]$, where $a_j^i=a+rk_j^i, b_j^i=a_j^i+r$, with k_j^i randomly selected from $\{0, 1,\ldots, [s^{1/n}]-1\}$. For MDE initialization the values k_j^i are uniformly selected for all subpopulations and all components. After initialization the elements of a subpopulation are relatively close to each other. During the evolution, the subpopulations are *not* restricted to the subdomain affected in the initialization step, their elements are allowed to cross the subdomain boundaries. Thus during the evolution the subpopulations are not necessarily non-overlapping, therefore different subpopulations could find the same optimum. The main aim of this controlled initialization is to ensure the space exploration.

Subpopulation's DE. On each subpopulation a DE-type algorithm is applied for a given number of iterations. For the algorithm applied at the subpopulation level the emphasis is mainly on the convergence speed and not on a large explorative power. A good candidate for this task seems to be the DE2 variant (Storn and Price, 1995) characterized by the fact that for each (sub)population element, x_i, a candidate offspring, y_i, is generated based on the relation:

$$y_i^j=\begin{cases} x_*^j+F\cdot(x_{\alpha_i}^j-x_{\beta_i}^j) & \text{with probability} \quad \text{p} \\ x_i^j & \text{with probability} \quad \text{1 - p} \end{cases}, \quad j=\overline{1,n} \tag{1}$$

where x_* is the best element of the (sub)population, α_i and β_i are randomly selected without replacement from $\{1,2,\ldots,m\}$, $F>0$ is a parameter controlling the perturbation and $p\in[0,1]$ is the so-called crossover probability. During the selection process the candidate offspring, y_i, will replace its parent, x_i, if it is better (with respect to the objective function). Some theoretical remarks on the differences between DE1 and DE2 variants are presented in Appendix A.

The migration process. In the island model, after a given number of generations is started a migration process during which some elements change their subpopulations. One of the effects of the migration is that it ensures an increase of the subpopulations diversity avoiding premature convergence situations. However when the aim is to locate multiple optima, the exchange of elements between subpopulations during the migration process could direct many subpopulations toward the same optimum thus reducing the number of explored regions. To illustrate this let us consider a simple migration strategy: each element of each subpopulation is exchanged with a given migration probability, p_m, with a randomly selected element from a randomly selected subpopulation. If we denote by $\overline{x_k}$ the mean value of the elements in the subpopulation P_k (its centroid) and with $\overline{\overline{x}}$ the mean value of all population elements, then after the random migration the expected centroid of subpopulation P_k will be $(1-p_m)\overline{x_k}+p_m\overline{\overline{x}}$ (see Proposition 2 in Appendix A). Thus, after migration the subpopulations centroids migrate toward the population centroid. Since the subpopulation variance increases after migration it follows that the subpopulations will concentrate their search in the same domain reducing the probability of detecting many optima. Therefore when the aim is to find many local optima a low migration probability should be used or even no migration. The influence of the migration probability on the number of detected optima is illustrated in figure 2 for a test function having 10 unequally spaced maxima of different heights (the multi-gaussian function in Table 1). The number of detected maxima (*Nmax*) has been averaged over 30 independent runs. The figure also illustrates the influence of the multi-resolution approach presented in the next section and the difference between two subpopulation initializations: the controlled one and a simple global initialization (all subpopulations are initialized over the entire search domain).



(a) (b)

Figure 2. Influence of the migration probability on the number of found maxima (MDE – continuous line, multi-resolution MDE – dashed line). (a) Controlled initialization (b) Global initialization

Stopping criteria. We consider that the DE2 evolution stopped when all subpopulations *converged*. Convergence of a subpopulation means that it loosed its diversity (e.g. its variance vanished, $Var(P_i) = 0$ or it is very small). If the subpopulation approached an optimum it is a *good convergence*, otherwise it is a *premature convergence*.

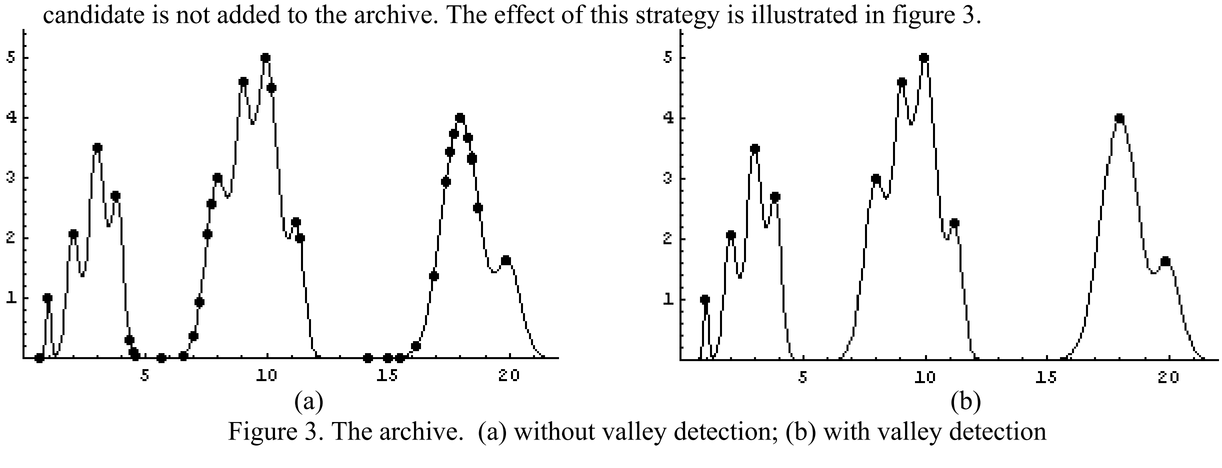
Control parameters. Preventing premature convergence of subpopulations can be assured through an adequate choice of the DE parameters, p and F. To avoid the annoying task of choosing the parameters an adaptive rule based on the idea of controlling the variance (Zaharie, 2003) can be used. The adaptation idea is to choose p and F such that the next recombination step will compensate the decrease of diversity induced by the selection step. In the case of DE2 algorithm the adaptation rule is based on the relations between the expected variance (for each of the n components) of the (sub)population after the recombination (1) is applied and the variance before recombination:

$$E(Var(Y)) = \left(2pF^2 + \frac{(1-p)^2}{m} + \frac{m-1}{m}\right)Var(x) + \frac{m-1}{m}p(1-p)(\bar{x}-x_*)^2 \qquad (2)$$

For high values of p the following approximation can be used: $E(Var(Y)) \approx (2pF^2+(m-1)/m)Var(x)$. Based on this relation, simple rules to adapt p and F can be derived. When $p \approx 1$ (fixed) the value of F corresponding to generation g can be computed by using the ratio of population variances (at component levels) $c(g)=Var(x(g-1))/Var(x(g))$ as follows: $F(g)=((c(g)-(m-1)/m)/2)^{1/2}$. For each component a specific value of F is used. These values are randomly initialized and independently adapted based on the above relation.

The archive. The best element from each subpopulation is added to the optima archive. To avoid redundancy in the archive (presence of multiple copies of the same optimum) two techniques are implemented in step 6:

- A new optimum is added to the archive only if it is sufficiently dissimilar from the already stored optima. The dissimilarity criterion is based on the euclidean distance on R^n. A candidate is added to the archive only if its distance to the other optima is greater than a threshold depending on the current resolution factor, r (e.g. r/4).
- A sufficiently dissimilar candidate is added to the archive only if it belongs to a different hill (in the case of a maximization problem) than the other elements of the archive. The decision procedure is based on the idea of hill-valley function introduced in (Ursem, 1999). To decide if there is a valley between two elements x and y a given number of convex combinations of x and y are generated. If for at least one $z=cx+(1-c)y$ the relation $f(z)<\min\{f(x),f(y)\}$ hold then one can decide that there exists a valley between x and y. If for a candidate optimum there is found an element in the archive such that no valley is detected between it and the candidate then the candidate is not added to the archive. The effect of this strategy is illustrated in figure 3.



(a) (b)

Figure 3. The archive. (a) without valley detection; (b) with valley detection

3. A multi-resolution approach. The effectiveness of the MDE in locating multiple optima is highly dependent on the number of subpopulations. Choosing the right number of subpopulations could be difficult mainly in the case of unequally spaced optima. Even if the number of subpopulations is higher than the number of optima it is possible that some optima to be missed during a single search epoch if there are optima close to each other. To solve this problem we propose the use of repeated search epochs each searching epoch being characterized by a specific resolution factor. The same number of subpopulations is used at each epoch but the resolution factor is decreasing. This idea is somewhat similar to that used in sequential niching (Beasley et al., 1993) and in the roaming algorithm (Lung and Dumitrescu, 2003). However unlike in sequential niching many subpopulations are simultaneously evolved and unlike in the roaming algorithm it is not necessary to specify a niche radius and the subpopulation initialization is controlled. The general structure of the multi-resolution multipopulation differential evolution (MMDE) is illustrated in figure 4.

```
1: Archive initialization: $A:=\emptyset$
2: Epoch counter initialization: $e:=1$
3: Repeat
4:    Compute the resolution factor: $r_e:=(b-a)/(se)^{1/n}$
5:    (Re)initialization of $P_1,P_2, \ldots, P_s$
6:    Apply MDE (steps 2-5)
7:    Add the best elements of subpopulations to the archive
8:    $e:=e+1$
9: Until a stopping criterion is satisfied
```

Figure 4. The general structure of the multi-resolution MDE algorithm

Subpopulations (re)initialization. The main characteristic of this multi-resolution approach is the subpopulations initialization. When passing from a search epoch to another, the subpopulations are reinitialized based on a finer discretization of the domain, i.e. on a smaller resolution factor, r. A simple rule to choose the resolution factor in the epoch e is: $r_e=(b-a)/(se)^{1/n}$. For each of the s subpopulations, a subdomain is randomly chosen from a number of se possible subdomains. The elements of the subpopulation are selected using a uniform distribution only in the first epoch. In the next epochs the selection distribution is influenced by the elements already placed in the archive. To be more specific, let us consider that at epoch e the archive contains the elements $\{a_1, a_2, \ldots, a_k\}$ and the subdomain affected to subpopulation P_i is D_i. During the initialization of P_i, an element x, randomly selected from D_i is accepted with a probability P_a determined using a sharing function:

$$P_a(x)=\frac{1}{1+\sum_{i=1}^{k}\sigma(x,a_i)}, \qquad \sigma(x,a)=\begin{cases}1-\dfrac{d(x,a)}{r_e/2} & \text{if } d(x,a)<r_e/2\\ 0 & \text{otherwise}\end{cases} \tag{3}$$

The sharing function is computed with respect to the elements belonging to the current archive, the resolution factor playing the role of a niche radius. The acceptance probability is high only for elements that are not too close to the optima already placed in the archive. This idea is similar to that of derating the fitness function used in sequential niching but here one derates the acceptance probability. If when selecting a subpopulation element a given number of successive rejections occur, the subdomain is replaced with a new randomly selected one.

Archive. After each epoch the archive is updated based on the same procedure as for MDE using the current resolution factor r_e.

Stopping criterion. The iterated search can be stopped after a given number of epochs or when the archive didn't change for a given number of epochs.

Communication between subpopulations. Even if no migration is applied, the subpopulations communicate, during the initialization step, through the archive. Thus the MMDE is not only a repeated application of MDE.

Computational complexity. To analyze the complexity of MMDE we shall take into account the steps 5 (subpopulation initialization) and 7 (archive updating). Let us consider that there are s subpopulations each one with m individuals and the archive contains k elements. If the problem dimension is n, then the complexity order of step 5 is determined by the sharing function computation and it is $O(ksmn)$. The costly operations in the step 7 are: computation of the distance between the best element of each subpopulation and the archive elements ($O(ksn)$) and the hill-valley function computation which involves some $O(ks)$ objective function evaluations.

4. Numerical results. The numerical experiments have two aims: to compare the ability of MMDE to locate multiple optima with that of other multimodal optimization techniques and to analyze the influence of the number of subpopulations and of the number of search epochs on the MMDE properties.

As performance measures we used the *number of function evaluations*, the *number of found optima* and the *success rate* (defined as the percent of runs on which all desired optima have been found). All reported values are averaged over 30 independent runs. The test functions used are both one-dimensional and bi-dimensional (see Table 1). For the multi-gaussian the parameters are: a=(1,2,3,3.85,8,9,10,11,18,2), b=(0.1,0.25,0.35,0.25,0.5,0.3,0.5,0.3,0.75,0.5), c=(1,2,3.5,2.5,3,3.5,5,2,4,1.5).

Test function	*Expression*	*Domain*	*Remarks*
Shubert (1D)	$f_1(x)=\sum_{j=1}^{5} j(\cos(j+1)x+j)$	*[-30,30]*	9 global maxima+ 49 local maxima
Multi-gaussian (1D)	$f_2(x)=\sum_{j=1}^{10} c_j \exp\left(-\frac{(x-a_j)^2}{2b_j^2}\right)$	*[0,22]*	10 unevenly spaced, unequal maxima
Himmelblau (2D)	$f_3(x,y)=200-(x^2+y-11)^2-(x+y^2-7)^2$	*[-6,6]*^{2}	4 global maxima
Shubert (2D)	$f_4(x,y)=\sum_{j=1}^{5} j(\cos(j+1)x+j)\cdot\sum_{j=1}^{5} j(\cos(j+1)y+j)$	*[-10,10]*^{2}	18 global minima/maxima and 760 local minima/maxima
Schaffer (2D)	$f_5(x,y)=0.5+\frac{\sin^2(\sqrt{x^2+y^2})-0.5}{(1+0.001(x^2+y^2))}$	*[-10,10]*^{2}	1 global maxima and a infinity of local maxima
multi-peaks (2D)	$f_6(x,y)=x\sin(4\pi x)-y\sin(4\pi y+\pi)+1$	*[-2,2]*^{2}	Almost 80 maxima of unequal height

Table 1. Test functions

The influence of the population structure and of the number of search epochs are presented in Table 2 for two one-dimensional test functions and a bi-dimensional one. The other parameters of MMDE were chosen as follows: *p=1*, *F* adapted according to the adaptation rule presented in section 2, $p_m=0$ (no migration). In each subpopulation, DE stopped when the variance averaged over the subpopulation and over all components became less than 10^{-5}.

Test fct.	*Supopulation size (m)*	*Number of subpopulations (s)*	*Number of Epochs* (e_{max})	*Averaged number of function evaluations*	*Averaged number of found optima*	*Success rate*
Multi gaussian	5	20	1	1525	7.53	6%
	5	10	2	1438	7.26	0%
	5	30	1	2317	8.96	36%
	5	15	2	2214	8.70	16%
	5	10	3	2213	8.66	10%
	5	**100**	**1**	**8351**	**9.93**	**97%**
	5	**20**	**5**	**8201**	**9.90**	**96%**
Shubert 1D	5	200	1	12708	55.80	10%
	5	100	2	12447	56.10	10%
	5	50	4	12510	55.66	0%
	5	10	20	12815	55.30	10%
	10	400	1	50004	57.80	80%
	10	100	4	46794	57.83	83%
	10	**50**	**8**	**47215**	**57.96**	**96%**
Shubert 2D	10	25	4	25895	16.03	16%
	10	50	4	53502	17.5	63%
	10	**50**	**10**	**146270**	**17.9**	**90%**
	20	20	4	44841	17.36	50%

Table 2. Influence of the subpopulation size, subpopulations number and the number of epochs on the success rate (for Shubert 2D, only the global optima were counted)

Results in Table 2 suggest that the number of subpopulations and of search epochs influences the behavior of the MMDE. Good results have been obtained using rather small subpopulations ($m=5$ or $m=10$). The number of subpopulations and of epochs seem to be more important than the subpopulations size. Other illustrations of MMDE behaviour for the test functions in Table 1 are presented in the Appendix B.

On the other hand we compared the results obtained using MMDE with some results reported in the literature for other multimodal optimization techniques: sequential niching (Beasley, 1993) - Table 3a, artificial immune network –opt Ai net (deCastro, 2002) -Table 3b, species conservation genetic algorithm – SCGA (Li et al., 2002) -Table 3c and an island model cooperating with speciation (Bessaou et al., 2000) – Table 3d. For MMDE only the population structure and the number of epochs were differently chosen for different test functions. Taking into account the fact that no clustering algorithm is used and no niche radius have to be specified (however we have to notice that the resolution factor depends on s and on e, thus the niche radius is indirectly specified) the preliminary experiments indicate that MMDE is a competitive candidate for multimodal optimization. MMDE slightly outperforms the sequential niching and the opt Ai-net but it does not succeed to reach the performance reported for SCGA (however the good results of SCGA were obtained based on a careful choice of the species distance) and for the island model with speciation (Bessaou et al., 2000).

MMDE($m=5,s=10,e_{max}=2$, $p_m=0$, $p=1$, *F adapted*)			*Sequential niching (Beasley, 1993) – pop.size:26*		
Test function: Himmelblau					
Averaged fct. evaluations	*Success rate*	*RMS error*	*Averaged fct. evaluations*	*Success rate*	*RMS error*
2665	96%	0.1	5500	76%	0.2

(a)

MMDE($m=5,s=50$, $e_{max}=20$, $p_m=0$, $p=1$, *F adapted*)		*Opt AI-net (de Castro, 2002) – 20 cells, 10 clones, 451 generations*	
Test function: multi-peaks			
Averaged fct. eval.	*Aver. no. of found optima*	*Averaged fct. eval.*	*Aver. no. of found optima*
76630	88.16	90200	61

(b)

MMDE($m=10,s=50$, $e_{max}=20$, $p_m=0$, $p=1$, *F adapted*)				*SCGA (Li, 2002) – pop. size:300*	
Test function: Shubert 2D					
Aver. fct. eval.	*Aver.no of optima*	*Aver. no. of global optima*	*Success rate*	*Aver. fct. eval.*	*Aver. no. of global optima*
39463	85.27	17.26	33%	35747	18

(c)

MMDE($m=40,s=10$, $e_{max}=1$, $p_m=0.5$, $p=1$, *F adapted*)		*Island model+speciation (Bessaou, 2000) m=10, s=50*	
Test function: Schaffer on $[-100,100]^2$ (aim: finding the global optima with accuracy 10^{-3})			
Aver. fct. eval.	*Success rate*	*Aver. fct. eval.*	*Success rate*
26253	90%	≈18000	100%

(d)

Table 3. A comparison between MMDE and other multimodal optimization techniques

5. Conclusions. We proposed a simple multi-resolution multipopulation DE algorithm based on the idea of avoiding any processing of the global population (e.g. clustering) and of using few control parameters. The ability of MMDE to locate multiple optima relies on a controlled initialization step and its efficiency follows from the properties of the DE2 algorithm. Preliminary numerical results suggest that MMDE is a competitive candidate for multimodal optimization.

References

[1] H.A. Abbass; *The Self-Adaptive Pareto Differential Evolution Algorithm*, Proceedings of 2002 Congress on Evolutionary Computation, pp. 831-836, 2002.

[2] D. Beasley, D.R. Bull, R.R. Martin, *A Sequential Niche Technique for Multimodal Function Optimization*, Evolutionary Computing, 1(2), pp. 101-125, 1993.

[3] M. Bessaou, A. Pétrowski, P. Siarry; *Island Model Cooperating with Speciation for Multimodal Optimization*, Parallel Problem Solving from Nature (PPSN 2000), Paris, France, September 16-20, 2000.

[4] L.N. de Castro, J. Timmins; *An Artificial Immune Network for Multimodal Function Optimization*, Proceedings of 2002 Congress on Evolutionary Computation, 2002.

[5] J.P. Li, M.E. Balasz, G.T. Parks, P.J. Clarkson; *A Species Conserving Genetic Algorithm for Multimodal Function Optimization*, Evolutionary Computation, 10(3), pp. 207-234, 2002.
[6] R.I.Lung, D. Dumitrescu; *Roaming Optimization: a new Evolutionary Techique*, in D.Petcu et al. (eds*),* Proc. of SYNASC'03, Timisoara, october 2003, pp. 149-156, 2003.
[7] N.K. Madavan; *Multiobjective Optimization using a Pareto Differential Evolution Approach*, in IEEE Proc. Of Congress on Evolutionary Computation (CEC'2002), vol. 1, pp. 1145-1150, 2002.
[8] S.W. Mahfoud; *A Comparison of Parallel and Sequential Niching Methods*, Proc. of 6th International Conference on Genetic Algorithms, pp. 136-143, 1995.
[9] A. Pétrowski; *A New Selection Operator Dedicated to Speciation*, In Th. Baeck, ed., Proceedings of the Seventh International Conference on Genetic Algorithms (ICGA '97) Morgan Kaufmann, San Francisco, CA, 1997.
[10] R. Storn, K. Price; *Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces*, Technical Report TR-95-012, ICSI, 1995.
[11] R.K. Ursem, R.: *Multinational Evolutionary Algorithms*, In Proc. of Congress of Evolutionary Computation, vol.3, Pp 1633-1640, 1999.
[12] D. Zaharie; *Control of Population Diversity and Adaptation in Differential Evolution Algorithms*, in R. Matoušek, P. Ošmera (eds.), Proc. of Mendel 2003, 9th Internat. Conference on Soft Computing, Brno, june 2003, pp. 41-46, 2003.
[13] D. Zaharie, D. Petcu; *Adaptive Pareto Differential Evolution and its Parallelization*, 5th Int. Conf. on Parallel Processing and Applied Mathematics - PPAM 2003, Czestochowa, Poland, sept. 7-10, 2003, LNCS 3019, pp 261-268, 2004.

Appendix A. *Some theoretical remarks.*

Each subpopulation, $P=\{x_1, \ldots, x_m\}\subset R^n$, of MDE can be characterized by its mean value $\bar{x}=\frac{1}{m}\sum_{i=1}^{m} x_i$ (the subpopulation's centroid) , by its best value x_* and by its variance $Var(P)=\frac{1}{m}\sum_{i=1}^{m}(x_i-\bar{x})^T(x_i-\bar{x})$. The subpopulation structure is influenced by the evolutionary operators (DE recombination and selection) and by the migration process (if it is applied). These transformations are based on some random elements thus the obtained (sub)populations can be modeled as populations of random variables. Therefore one can compute only the expected values of the mean and/or variance of the transformed populations. In the following we will denote by *E(.)* the expected values computed with respect to the random elements involved in the transformations.

Let us analyze first the influence of recombination. We consider a general form of DE-recombination:

$$Y_i^j=\begin{cases}\lambda x_*^j+(1-\lambda)x_{\gamma_i}^j+F\cdot(x_{\alpha_i}^j-x_{\chi_i}^j) & \text{with probability } p\\ x_i^j & \text{with probability } 1-p\end{cases} \qquad i=\overline{1,m},\quad j=\overline{1,n} \tag{A1}$$

with $\lambda\in[0,1]$, α_i, β_i and γ_i are values randomly selected without replacement from $\{1,2,\ldots,m\}$ and $F>0$. For $\lambda=0$ one obtains the classical DE1 algorithm while $\lambda=1$ leads to particular variant of DE2 algorithm. Since the transformation is applied in a similar manner to all n components we can analyze, without loss of generality, the case of one component (thus we will skip the upper index, j).

Our aim is to compare the DE1 and DE2 recombination with respect to their influence on the subpopulation mean value and variance.

Proposition 1. *Let Y^λ be the population obtained after applying the recombination rule (A1). Then the following statements hold:*

(i) $E(\overline{Y^\lambda})=\lambda p x_*+(1-p\lambda)\bar{x}$

(ii) If $p\geq\frac{m-1}{m}$ then $E(Var(Y^0))\geq E(Var(Y^1))$

Proof. By ignoring the upper index, j, relation (A1) becomes

$$Y_i^\lambda=(\lambda x_*+(1-\lambda)x_{\gamma_i}+F\cdot(x_{\alpha_i}-x_{\beta_i}))\cdot 1_{R_i}+x_i\cdot 1_{\overline{R_i}} \tag{A2}$$

where 1_{R_i} and $1_{\overline{R_i}}$ are the indicator functions of the random events corresponding to the cases when the recombination is applied (R_i) or is not applied ($\overline{R_i}$), respectively. Thus $P(R_i)=p$ and $P(\overline{R_i})=1-p$.

(i) Since $E(x_{\alpha_i}) = E(x_{\beta_i}) = E(x_{\gamma_i}) = \bar{x}$ it follows that

$$E(Y_i^{\lambda}) = p(\lambda x_* + (1-\lambda)\bar{x}) + (1-p)\bar{x} = \lambda p x_* + (1-p\lambda)\bar{x}.$$

Thus $E(\overline{Y^{\lambda}}) = \frac{1}{m}\sum_{i=1}^{m} E(Y_i^{\lambda}) = \lambda p x_* + (1-p\lambda)\bar{x}\cdot$

(ii) From (Zaharie, 2003) it follows:

$$E(Var(Y^{\lambda})) = \left(2pF^2 + \frac{(1-p)^2}{m} + \frac{m-1}{m}(p(1-\lambda)^2 + (1-p))\right)Var(x) + \frac{m-1}{m}\lambda^2 p(1-p)(\bar{x} - x_*)^2$$

Thus

$$E(Var(Y^0)) - E(Var(Y^1)) = p\frac{m-1}{m}Var(x) - p(1-p)\frac{m-1}{m}(\bar{x} - x_*)^2 = p\frac{m-1}{m}\left(\frac{1}{m}\sum_{x_i \neq x_*}(x_i - \bar{x})^2 + (p - \frac{m-1}{m})(\bar{x} - x_*)^2\right)$$

From this relation it is obvious that $E(Var(Y^0)) \geq E(Var(Y^1))$ if *p≥(m-1)/m*. □

The property proved above suggests that the DE1 recombination has a higher explorative power than DE2 recombination. If we consider that a subpopulation converged when its variance became small enough, taking into account the fact that selection usually decreases the population variance it follows that DE2 converges faster than DE1 variant. This remark motivates the use of DE2 variant for multimodal optimization.

Another aspect that we analyze here is the influence of the migration process on the ability of the algorithm to detect multiple optima. Let $P_k = \{x_1^k, \ldots, x_m^k\}$ be the *k*th subpopulation and let us consider a simple random migration: each element of P_k can be replaced with a migration probability p_m with a randomly selected element belonging to an arbitrary subpopulation. Let $P_k' = \{Y_1^k, \ldots, Y_m^k\}$ be the *k*th subpopulation after migration.

Proposition 2. *If* $\overline{x^k}$ *denotes the mean value of the subpopulation* P_k *and* $\bar{\bar{x}}$ *is the mean value of the entire population then by applying a random migration the expected mean value of subpopulation* P_k *satisfies:*

$$E(\overline{Y^k}) = (1-p_m)\overline{x^k} + p_m\bar{\bar{x}} \tag{A3}$$

Proof. From the particularities of random migration it follows that

$$Y_i^k = x_i^k \cdot 1_{\overline{M}} + \sum_{l=1}^{s}\sum_{j=1}^{m} x_j^l \cdot 1_{M_{lj}} \tag{A4}$$

where M_{lj} is the event of replacing x_i^k with x_j^l (*j*th element of *l*th subpopulation) and $\overline{M}$ is the non-migration event. Since in the migration process each subpopulation and each subpopulation's element is selected based on a uniform distribution it follows that $P(M_{lj}) = p_m/(ms)$ and $P(\overline{M}) = 1 - p_m$. Thus

$$E(Y_i^k) = (1-p_m)x_i^k + \frac{p_m}{ms}\sum_{l=1}^{s}\sum_{j=1}^{m} x_j^l = (1-p_m)x_i^k + p_m\bar{\bar{x}}$$

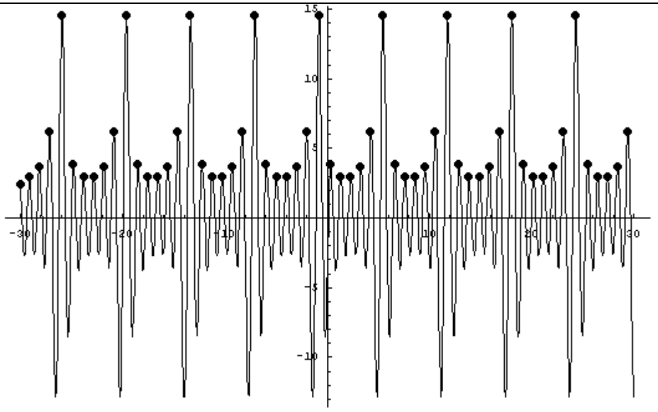
and

$$E(\overline{Y^k}) = \frac{1}{m}\sum_{i=1}^{m} E(Y_i^k) = (1-p_m)\overline{x^k} + p_m\bar{\bar{x}}$$

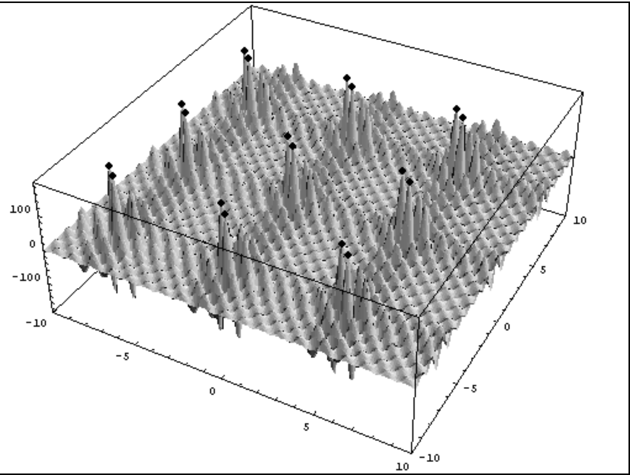
□

This result suggests that by migration the subpopulations centroids migrate toward the population centroid. On the other hand the random migration usually increases the subpopulations variances. Therefore when multiple optima have to be located the migration may not be helpful.

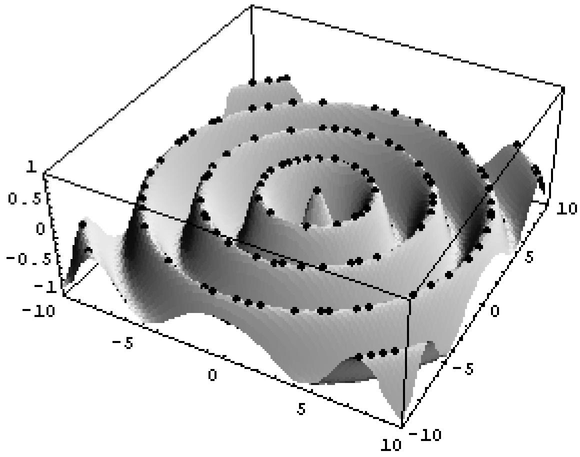
Appendix B. Illustration of search landscapes of some test functions and of the optima detected by MMDE



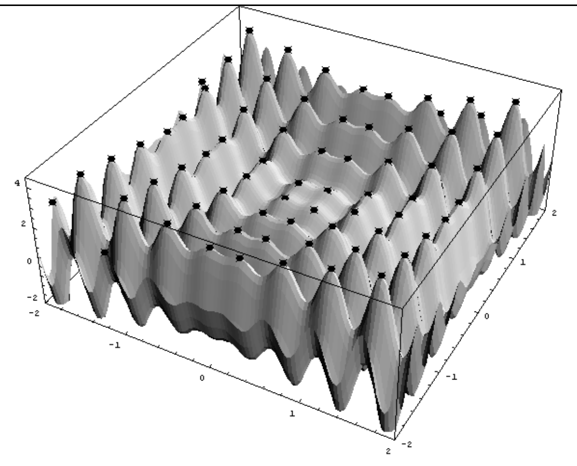
Test function: Shubert 1D
Optima searched: global and local maxima
MMDE parameters: $m=5$, $s=10$, $e=20$, $p=1$, F adapted, $p_m=0$, function evaluations: 12815



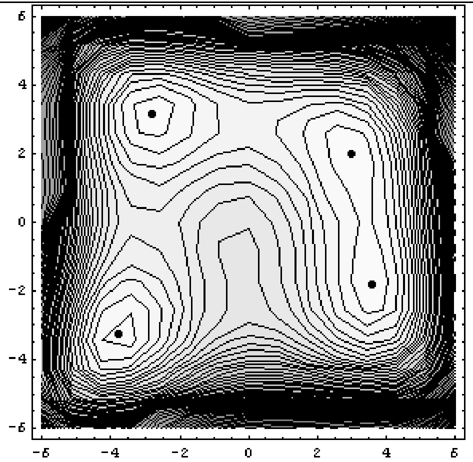
Test function: Shubert 2D
Optima searched: global maxima
MMDE parameters: $m=10$, $s=25$, $e=4$, $p=1$, F adapted, $p_m=0$, function evaluations: 25895



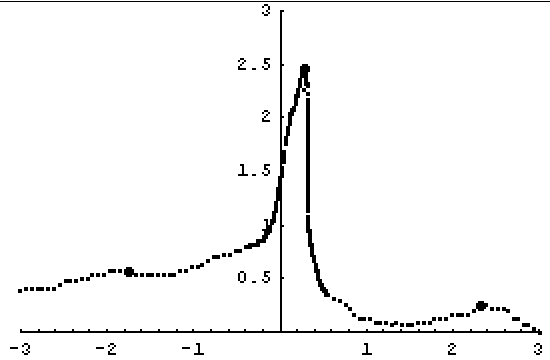
Test function: Schaffer
Optima searched: global and local maxima
MMDE parameters: $m=10$, $s=20$, $e=10$, $p=1$, F adapted, $p_m=0$, function evaluations: *65778*



Test function: multi-peaks (Castro, 2002)
Optima searched: global and local maxima
MMDE parameters: $m=5$, $s=50$, $e=20$, $p=1$, F adapted, $p_m=0$, function evaluations: *76578*



Test function: Himmelblau
Optima searched: global and local maxima
MMDE parameters: $m=5$, $s=4$, $e=5$, $p=1$, F adapted, $p_m=0$, function evaluations: *1841*



Test function: DTA (Differential Thermal Analysis) curve
Optima searched: significant peaks
MMDE parameters: $m=5$, $s=5$, $e=5$, $p=1$, F adapted, $p_m=0$, function evaluations: *1635*