

# Crowding under Diverse Distance Criteria for Niche Formation in Multimodal Optimization

**Fernández N., Alfonso H.**

*Proyecto UNLPAM-09/F009<sup>1</sup>*

*Departamento de Informática - Facultad de Ingeniería*

*Universidad Nacional de La Pampa*

Calle 110 esq. 9

(6360) General Pico – La Pampa – Rep. Argentina

e-mail: {fernaty, alfonsoh}@ing.unlpam.edu.ar;

Phone: (02302)422780/422372, Ext. 6302

**Gallard R.**

*Proyecto UNSL-338403<sup>2</sup>*

*Departamento de Informática*

*Universidad Nacional de San Luis*

Ejército de los Andes 950 - Local 106

5700 - San Luis

Argentina

E-mail: rgallard@unsl.edu.ar

Phone: +54 2652 420823

Fax : +54 2652 430224

## Abstract

Niche formation allows evolutionary algorithms to be used when the location and maintenance of multiple solutions appertaining to diverse areas of the phenotypic space is required. Consequently the application field can be extended to multiobjective optimization, simulation of complex systems and multimodal function optimization.

In this later case a conventional evolutionary algorithm tends to group the final population around the fittest individual. Thus, other areas of interest in the search process are lost. Niching methods permits the maintenance of solutions located around these areas of interest.

This contribution briefly describe problems preventing niche formation in conventional genetic algorithms, a crowding method for niche formation and analysis of results when optimizing two multimodal functions.

**Keywords:** Genetic algorithms, genetic diversity, genetic drift, niche formation, crowding.

---

<sup>1</sup> The Research Group is supported by the Universidad Nacional de La Pampa.

<sup>2</sup> The Research Group is supported by the Universidad Nacional de San Luis and the ANPCYT (National Agency to Promote Science and Technology).

## 1. Introduction

In nature organisms exploit the living environment divided into distinct subsets, called niches, each one performing a differentiated role in this exploitation. In this way competence is avoided allowing the formation of diverse species which conform stable populations. There, rather than contending, the species exploit separated niches (sets of environmental conditions) where other organisms have little or no interest.

In the case of evolutionary computation when optimizing multimodal functions [Deb 89], the algorithm tend to group the final population around one of the optimal (or quasi optimal) points.

The Scheme Theorem [Holland 75] ensures that the best schemes receive an exponentially increasing number of samples. Then, why this happens ? The assumptions of the Schema Theorem are based on infinite populations, but with finite population there does not exist any selective advantage for the competitive alternatives and the population will converge towards one of them. This problem is known as genetic drift [De Jong 75]: stochastic sampling errors tend to accumulate and cause that the final population converge towards one of the alternatives.

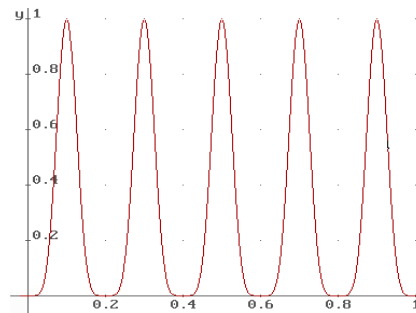
This convergence to only one of the alternatives is undesirable in multimodal optimization of real problems, because we are interested on getting information about good points and better solutions. Using the results of the theory of niche and speciation [Goldberg 89], [Perry 84], will allow us to get a more controlled competence. Caviccio [Caviccio 70] dissertation was one of the first works attempting to induce nichelike behavior in genetic algorithms search process. A mechanism called preselection was introduced. Here an offspring replaces the worst parent if his fitness is higher than that of the worst parent. As new strings tend to replace strings similar to themselves, the genetic diversity is maintained. The Caviccio's preselection technique was generalized by De Jong in a method called crowding. This technique also inserts new elements in the population by replacing similar elements. Similarity of individuals can be determined by means of a distance measure, either genotypic or phenotypic. As individuals of similar fitness are also provably residing at the same niche a weaker fitness similarity criterion can be used. Crowding methods tend to spread individuals among the most prominent peaks of the search space. To make the replacement an individual is compared to a random subpopulation of  $m$  members, where  $m$  is called the crowding factor. Then the individual with the highest similarity is replaced by the newly created string. At the first stages of the simulation this implies random selection of replacements because all individuals are likely to be equally dissimilar. As the evolution progresses more individuals are similar to one another and the replacement of individuals by similar individuals tend to maintain diversity within the population reserving room for distinct species. Other crowding methods similar in operation and behavior have proposed [Mahfoud 95], [Cedeño 94], [Harik 95].

This paper show some aspects of implementation and analysis of results using crowding under the three above mentioned distance criteria when optimizing two multimodal testing functions.

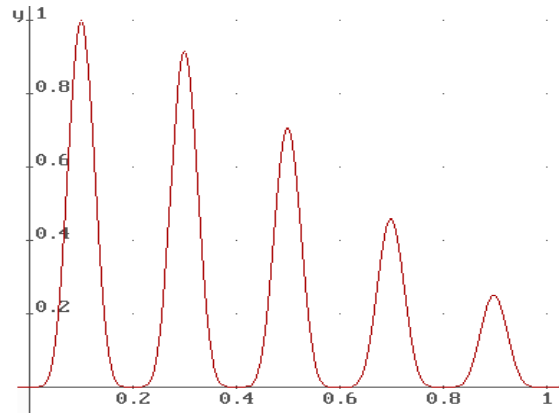
## 2. Experiment description

For this preliminary experiments we chose two simple functions

$$f_1(x) = \sin^6(5\pi x)$$



$$f_2(x) = f_1(x) + e^{(-2)\ln 2 \left[ \frac{x-0.1}{0.8} \right]^2}$$



Function  $f_1$  have a landscape showing five equal height peaks in  $[0,1]$  while function  $f_2$  shows five unequal height peaks in the same interval.

Given two solutions  $x_i$  and  $x_j$  and their corresponding encodings, of length  $l$ ,  $c_i$  and  $c_j$  and the binary string  $z = c_i \text{ XOR } c_j$ , three criteria to determine a distance measure (similarity) among individual were used:

**Phenotypic distance:** It is determined by using problem-specific knowledge of the phenotype. In this case (single variable functions) it is given by

$$d_f = |x_i - x_j|$$

**Genotypic distance:** it is the result of comparing bit-by-bit two binary encoded solutions and it is given by

$$d_g = \sum_{k=1}^l z_k$$

**Fitness distance:** it is the absolute difference of fitness among both solutions and it is given by

$$d_F = |f(x_i) - f(x_j)|$$

A genetic algorithm with a population of 200, binary encoded chromosomes of length 30, elitism and proportional selection was implemented. Crossover and mutation probabilities were set to 0.8 and 0.3<sup>3</sup> respectively. The maximum number of generation was fixed to 100 and crowding factors of 2, 3 and 4 were used.

As an indication of the algorithm performance the following relevant performance variables were chosen:

$$\mathbf{Ebest} = (\text{Abs}(\text{opt\_val} - \text{best value})/\text{opt\_val})100$$

It is the percentile error of the best found individual when compared with the known, or estimated, optimum value  $\text{opt\_val}$ . It gives us a measure of how far are we from that  $\text{opt\_val}$ .

$$\mathbf{Epop} = (\text{Abs}(\text{opt\_val} - \text{pop mean fitness})/\text{opt\_val})100$$

It is the percentile error of the population mean fitness when compared with  $\text{opt\_val}$ . It tell us how far the mean fitness is from that  $\text{opt\_val}$ .

<sup>3</sup> According to Maulding [Maulding 84] results the probability for mutation is rather high. This ensures enough genetic diversity while evolution progresses

**Niche Count:** Indicates the mean number of individuals belonging to a particular niche and gives an idea of the division of the whole population into subpopulations.

**Optimal Hits Per Niche Ratio** = (# optimal hits per niche/ # runs)100

It is the hit ratio to find the optimal solution within a niche, all over the total number of runs.

Gives the percentage of hits within each niche over all runs.

### 3. Results

The following tables show the results obtained under each distance criterion on the selected testing functions. In particular  $\mu$ ,  $\sigma$  and  $\sigma/\mu$  stands for the mean, standard deviation and coefficient of deviation for the corresponding performance variables.

#### Function f1

- Ebest analysis**

Crowding Factor M		Distance criterion		
		Phenotypic	Genotypic	Fitness
2	$\mu$	0.00946293	0.05665764	0.01746724
	$\sigma$	0.02464380	0.12442613	0.03145618
	$\sigma/\mu$	2.60424475	2.19610520	1.80086757
3	$\mu$	0.00460087	0.02172401	0.01133947
	$\sigma$	0.01284428	0.05324600	0.02476983
	$\sigma/\mu$	2.79170892	2.45102101	2.18438971
4	$\mu$	0.00677591	0.04454333	0.06014901
	$\sigma$	0.01280554	0.07603844	0.17201931
	$\sigma/\mu$	1.88986104	1.70706698	2.85988583

The best mean values for Ebest are achieved under the phenotypic criterion and with crowding factor value of  $m = 3$ .

- Epop analysis**

Crowding Factor m		Distance criterion		
		Phenotypic	Genotypic	Fitness
2	$\mu$	60.3124457	65.0977623	65.7088902
	$\sigma$	2.3743500	2.7331421	3.0659769
	$\sigma/\mu$	0.0393675	0.0419851	0.0466600
3	$\mu$	57.6172093	66.017956	66.341672
	$\sigma$	3.4126921	2.9637199	3.5877803
	$\sigma/\mu$	0.0592304	0.0448926	0.0540803
4	$\mu$	56.4997877	66.596331	66.637700
	$\sigma$	4.3506215	4.1875263	3.1733967
	$\sigma/\mu$	0.0770024	0.0628792	0.0476216

Epop values vary from 56.5% for phenotypic criterion and  $m = 4$  to 66.6% for fitness criterion and  $m = 4$ , indicating a high diversity in the final population. This is an expected effect for the crowding method which distribute individuals along the phenotypic space. Also it can be observed that for any value of  $m$ , the phenotypic criterion shows the lower values while the fitness criterion shows the higher values of Epop. The relative dispersion given by the deviation coefficient is low.

- **Niche Count**

Cr. Fac. M	Niche Count				
	1st niche	2nd niche	3rd niche	4th niche	5th niche
Phenotypic					
2	53.00	34.85	29.40	37.75	45.00
3	46.35	34.95	34.00	31.55	53.15
4	49.55	33.85	31.15	34.85	50.60
Genotypic					
2	42.55	39.80	39.00	41.05	37.60
3	40.45	39.30	38.60	42.00	39.65
4	40.85	43.60	36.50	40.90	38.15
Fitness					
2	40.10	39.75	38.50	39.10	42.55
3	37.45	44.60	38.95	43.75	35.25
4	41.00	43.50	36.55	43.05	35.90

For  $f_1$  with five peaks of equal height, it is expected an equal number of individuals per niche when niching methods are used. In our experiments the genotypic criterion showed to be more adjusted to the expectancies, for any crowding factor, than the other two criteria

- **Optimal Hits**

Cr. Fac. m	Optimal Hits					
	1st niche	2nd niche	3rd niche	4th niche	5th niche	Global
Phenotypic						
2	35	15	35	20	30	80
3	45	45	15	25	40	95
4	50	35	15	5	40	80
Genotypic						
2	25	15	5	5	5	45
3	20	20	10	20	10	65
4	15	15	15	20	15	55
Fitness						
2	15	25	15	10	20	65
3	15	25	20	15	25	80
4	15	15	15	5	15	55

According to the Global column the genetic algorithm found the optimal value, 45% of the runs in the worst case under the genotypic criterion and  $m = 2$  and 95% of the runs in the best case under the phenotypic criterion and  $m = 3$ . The Global percentage accounts one hit either when only one peak was reached (by one ore many individuals) or when different peaks were reached (by one ore many individuals). Partial percentages in the niches indicates reaching of the corresponding peak by at least one individual. In general the phenotypic approach behaves better that the other two approaches when the number of optimal or near optimal values obtained are considered.

## Function f2

- Ebest analysis

Crowding Factor m		Distance criterion		
		Phenotypic	Genotypic	Fitness
2	$\mu$	0.006271975	0.21395053	0.37232822
	$\sigma$	0.011088158	0.34319198	0.63139212
	$\sigma/\mu$	1.767889546	1.60407167	1.69579441
3	$\mu$	0.001170103	0.53243204	0.30208194
	$\sigma$	0.005477042	0.96075900	0.82907636
	$\sigma/\mu$	4.680820132	1.80447254	2.74454127
4	$\mu$	0.006429366	0.16520279	0.12629741
	$\sigma$	0.013399935	0.2905757	0.34791863
	$\sigma/\mu$	2.084176862	1.75890312	2.75475668

The best values for Ebest are again achieved under the phenotypic criterion and with crowding factor value of  $m = 3$ .

- Epop analysis

Crowding Factor m		Distance criterion		
		Phenotypic	Genotypic	Fitness
2	$\mu$	64.58129807	74.62718210	72.75620450
	$\sigma$	2.26964021	3.35972631	2.36011032
	$\sigma/\mu$	0.03514392	0.04502014	0.03243861
3	$\mu$	62.86013418	75.86109270	73.92359440
	$\sigma$	2.77918096	2.95172413	2.84588277
	$\sigma/\mu$	0.04421213	0.03890959	0.03849762
4	$\mu$	62.62528435	73.9092748	73.53783990
	$\sigma$	2.44701856	2.97971506	2.67502983
	$\sigma/\mu$	0.03907397	0.04031585	0.03637624

Epop values vary from 62.6% for phenotypic criterion and  $m = 4$  to 75.8% for genotypic criterion and  $m = 3$ , indicating a high diversity in the final population. When multimodality include peaks of different heights a greater diversity is observed in the final population when compared with functions with equal height peaks. For function f2, it can be observed that for any value of  $m$ , the phenotypic criterion shows the lower values while the genotypic criterion shows the higher values of Epop. The relative dispersion given by the deviation coefficient is also low.

- **Niche Count**

Cr. Fac. m	Niche Count				
	1st niche	2nd niche	3rd niche	4th niche	5th niche
Phenotypic					
2	74.70	45.50	29.80	22.85	27.15
3	75.10	45.90	27.80	23.60	27.60
4	75.10	43.30	28.65	24.30	28.65
Genotypic					
2	47.40	50.95	39.25	32.90	29.50
3	46.55	48.75	41.75	34.25	28.70
4	49.05	48.70	40.95	32.60	28.70
Fitness					
2	55.30	49.35	37.15	32.75	25.45
3	55.30	50.15	36.70	30.95	26.90
4	52.70	52.30	36.60	32.75	25.65

For f2 with five peaks of decreasing height, it is expected a number of individuals per niche proportional to the peak height. In our experiments the phenotypic criterion showed to be the most adjusted to the expectancies, except for the fifth niche. On the other hand under the fitness criterion a monotonic and softer decrement of the niche count is observed.

- **Optimal Hits**

Cr. Fac. m	Optimal Hits				
	1st niche	2nd niche	3rd niche	4th niche	5th niche
Phenotypic					
2	75	50	60	55	30
3	100	70	65	70	40
4	80	70	65	60	40
Genotypic					
2	15	55	45	65	70
3	25	20	65	60	50
4	30	40	45	80	55
Fitness					
2	25	45	80	55	40
3	20	30	40	55	55
4	40	40	40	50	55

Here, as the peaks are of different magnitude, the global optimum is reached when at least one individual climbs the highest (first) peak. Consequently the genetic algorithm found the optimal value, 15% of the runs in the worst case under the genotypic criterion and  $m = 2$  and 100% of the runs in the best case under the phenotypic criterion and  $m = 3$ . Percentages in the niches indicates reaching of the corresponding peak by at least one individual. In general, again, the phenotypic approach behaves better than the other two approaches when the number of optimal or near optimal values obtained are considered.

## 4. Conclusions

When optimizing multimodal functions conventional evolutionary algorithm tend to group the final population around one of the optimal (or quasi optimal) points.

The convergence to only one of the alternatives is unattractive when we are interested on getting information about good points scattered through the problem space.

This effect, due to genetic drift, can be partially released by considerably enlarging the population size, which implies a significant increment of computational effort.

Crowding, is a simple, convenient and inexpensive method to solve the main question: to provide niche formation.

In this work two simple multimodal functions with different landscapes were chosen to study the niching capability of the crowding method under three distinct distance criteria: phenotypic, genotypic and fitness.

In the case of f1, with five equal peaks, when looking at the quality of results the phenotypic approach prevails over the others providing higher optimal hits and minimum Ebest. When looking at population statistics the phenotypic approach provides lower diversity while higher diversity is observed under the fitness criterion. Finally when looking at niche count the genotypic criterion provides an even distribution of the population in the niches.

In the case of f2 with five peaks of decreasing height, when looking at the quality of results the phenotypic approach again prevails over the others providing minimum Ebest. When looking at population statistics the phenotypic approach provides lower diversity while higher diversity is observed under the genotypic criterion. Finally when looking at niche count the phenotypic criterion provides a number of individuals per niche nearly proportional to the peak height, and absolute differences are stronger in the first three higher peaks.

Summarizing, for both type of multimodal functions, under the chosen values of crowding factor m, experimental results point to the phenotypic criterion as the most convenient approach. This is a consequence of the analysis and its lower computational effort.

Further work will investigate variants of crowding including methods to regain diversity when it was lost by means of adaptation of parameters settings such as mutation probabilities.

## 5. Acknowledgements

We acknowledge the cooperation of the project group for providing new ideas and constructive criticisms. Also to the Universidad Nacional de San Luis, the Universidad Nacional de La Pampa, and the ANPCYT from which we receive continuous support.

## 6. References

- [Caviccio 70] Caviccio D. J., *Adaptive search using simulated evolution*, Unpublished doctoral dissertation, University of Michigan Press, Ann Arbor, 1970.
- [Cedeño 94] Cedeño W., Vemuri V. R., Slezak T., *Multiniche crowding in genetic algorithms and its application to the assembly of DNA restriction-fragments*, *Evolutionary Computation*, 2, pp 321-345, 1994.
- [De Jong 75] De Jong K. A. - *Analysis of the Behavior of a Class of Genetic Adaptive Systems* - PhD Dissertation, University of Michigan, 1975
- [Deb 89] Deb K., *Genetic algorithms in multimodal function optimization*, Master Thesis, TCGA Report 89002, University of Alabama. The Clearinghouse for genetic algorithms., 1989.
- [Goldberg 89] Goldberg D. E., Deb K., *An investigation of niche and species formation in genetic function optimization*, *Proceedings of the 3rd Int. Conf. on Genetic Algorithms*, pp 42-50. Morgan Kaufmann Publishers, Los Altos CA, 1989.



- [Harik 95] Harik G., *Finding multimodal solutions using restricted tournament selection*, Proc. of the 6th International Conference on Genetic Algorithms, pp 24-31, Pittsburgh, PA, July 1995.
- [Holland 75] Holland J. H., *Adaptation in natural and artificial systems*, University of Michigan Press, Ann Arbor, 1975.
- [Mahfoud 95] Mahfoud S. W., *Niching Methods for genetic algorithms*, Doctoral dissertation and IlliGAL report 95001, University of Illinois at Urbana Champaign, Dissertation abstracts Int. (Univ. Microfilms 9543663)
- [Maulding 84] Maulding M. L., *Maintaining Diversity in Genetic Search*, Proc. Of the National Conf. On Artificial Intelligence. Pp 247-250, 1984.
- [Perry 84] Perry Z., *Experimental study of speciation in ecological niche theory using genetic algorithms*. Doctoral dissertation, Univ. of Michigan, Dissertation abstracts Int. 45(12) 3870B.