Investigating the Influence of the Neighborhood Attraction Factor to Evolution Strategies with Neighborhood Attraction

Jutta Huhse , Andreas Zell
E-mail: [huhse|zell]@informatik.uni-tuebingen.de

Universität Tübingen, W.-Schickardt-Institut für Informatik,
Sand 1, D-72076 Tübingen, Germany

Abstract. The evolution strategy with neighborhood attraction (EN) is a new combination of self-organizing maps (SOM) and evolution strategies (ES). It adapts the neighborhood relationship known from SOM to ES individuals to concentrate them around the optimum of the problem. In this paper, detailed investigations on the influence of one of the most important EN-operators – the neighborhood attraction – were performed on a variety of well-known optimization problems. It could be shown that the parameter setting for the neighborhood attraction has a very strong influence on the convergence velocity and the robustness of the EN, and suggestions for applicable parameter settings could be made.

1 Introduction

Evolution strategies with neighborhood attraction (EN) are a combination of two different kinds of problem solvers: Evolution strategies (ES) and artificial neural networks, especially self-organizing maps (SOM). ES were developed in the late 1960s by Rechenberg and Schwefel and improved in the following time (see [6], [7] and [2]). Their main application is the optimization of real-valued multi-parameter problems. They directly use the information about the quality of a potential solution of the function to be optimized. They work on a population P of potential solutions (individuals a) by manipulating these individuals with genetic operators.

A special class of neural networks - the self-organizing maps (SOM) - were developed in the 1980s by Kohonen [5]. The neurons of a SOM are organized in neighborhood relationship, e. g. a two-dimensional grid. The idea behind EN is to transfer the neighborhood and the learning rule defined for SOM neurons onto the individuals of an ES (see figure 1) to concentrate the individuals around the optimum. In previous benchmark tests it could be shown, that this EN is able to solve difficult optimization tasks. It was also shown, that

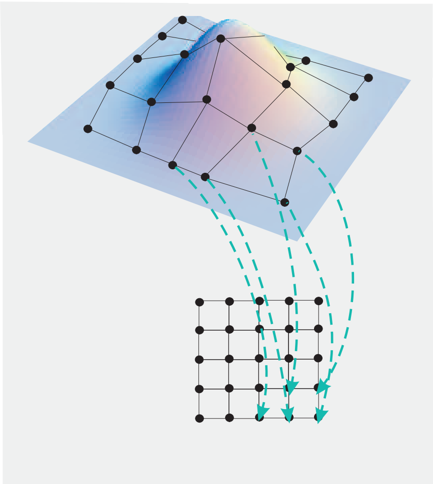

Figure 1: EN: Transfer of the SOM neighborhood onto ES individuals

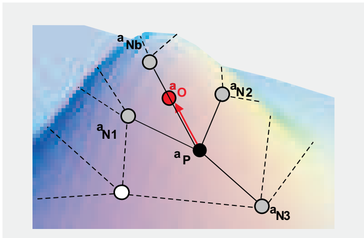

Figure 2: Neighborhood attraction in EN

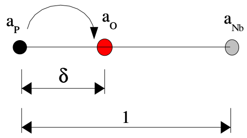

Figure 3: Attraction factor δ

the performance of the EN is equivalent to or even better than comparable conventional ES on a number of benchmark problems (cf. [4]).

In these former test series it was also found out that the convergence speed or – in the extreme case – even the success or failure of the EN depends strongly on the parameter settings for the neighborhood attraction factor δ.

In this work, this attraction factor δ is thoroughly investigated, its influence on the performance of the EN is shown, and suggestions for applicable parameter settings are made.

A short description of the EN is given in section 2. The optimization problems used as a test bed for our investigations are described in section 3. Section 4 shows the test series that were performed, and the results are discussed in section 5.

2 Basics of the Evolution Strategy with Neighborhood Attraction

The individuals, which are unordered in conventional ES, have neighborhood relations in the EN. The neighborhood between the μ parent individuals is constituted by arranging them in an orthogonal, elastic grid. As known from SOM, each individual can be identified by its fixed grid position, and two individuals are neighbors if they are directly connected on the grid.

As is customary in ES, the EN individuals are evaluated using the fitness function.

The EN-specific genetic operator – the neighborhood attraction – manipulates the EN-individuals according to one learning step in a SOM. Every individual a_P is attracted to its best neighbor a_{Nb} and thus becomes the offspring a_O (see figure 2). The object variables $\vec{x}_O$ of the offspring are calculated with equation 1, the neighborhood relations are retained unchanged.

$$\vec{x}_O = \vec{x}_P + \delta \cdot (\vec{x}_{Nb} - \vec{x}_P) \tag{1}$$

$(\vec{x}_{Nb} - \vec{x}_P)$ is the difference vector between the object variables vectors $\vec{x}_P$ of the parent and its best neighbor $\vec{x}_{Nb}$. The parameter δ (see figure 3) defines the strength of the attraction along the difference vector and $\vec{x}_O$ denotes the object variables of the offspring.

If the parent individual a_P is considered better than all its neighbors a_{Nj} ($j = 1 \ldots g$, g is the number of neighbors) a "simple conventional" mutation is performed. For details, please see [4].

3 Test Functions

The functions below have been used as a test bed to investigate the influence of δ to the EN. They are numbered according to common test functions used in e. g. [1] and [3]. All test functions are minimization problems with an optimum in $f_i(\vec{x}) = 0$. This test bed includes uni-modal (f_1, f_2, f_6, f_{15}) and multi-modal functions (f_9, f_{21}) as well as symmetric and non-symmetric functions (f_2). We chose these test functions because they were already used successfully in previous work to compare the performance of the EN to other ES.

- $f_1(\vec{x}) = \sum_{i=1}^{n} x_i^2$ (Sphere model)
- $f_2(\vec{x}) = \sum_{i=1}^{n-1} (100 \cdot (x_{i+1} - x_i^2)^2 + (x_i - 1)^2)$ (Generalized Rosenbrock's)
- $f_6(\vec{x}) = \sum_{i=1}^{n} \left(\sum_{j=1}^{i} x_j \right)^2 = x^{\mathrm{T}} \mathbf{A} x$ (Schwefel's double sum)
- $f_9(\vec{x}) = -a \cdot \exp\left(-b \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2} \right) - \exp\left(\frac{1}{n} \cdot \sum_{i=1}^{n} \cos(c \cdot x_i) \right) + a + e$
- $f_{15}(\vec{x}) = \sum_{i=1}^{n} i \cdot x_i^2$ (Weighted sphere model)
- $f_{21}(\vec{x}) = 1 + \frac{1}{d} \sum_{i=1}^{n} x_i^2 - \prod_{i=1}^{n} \cos\left(\frac{x_i}{\sqrt{i}} \right)$ (Griewangk's)

4 Test Series

For the test series, we varied the value of δ between 0.00011 and 0.1. The values of μ were set to $\mu = 2$, $\mu = 11$, and $\mu = 100$, which result in a simple connection of two individuals, a 1D-chain neighborhood and a 2D-grid neighborhood, resp., and which scale roughly with factor 10. The number of offsprings was set to the constant value $\lambda = 2$. For all test functions the dimension of the problem space was set to $d = 10$. All test series were conducted with several different seeds for each μ-δ-combination and then averaged out. The graphical representations of

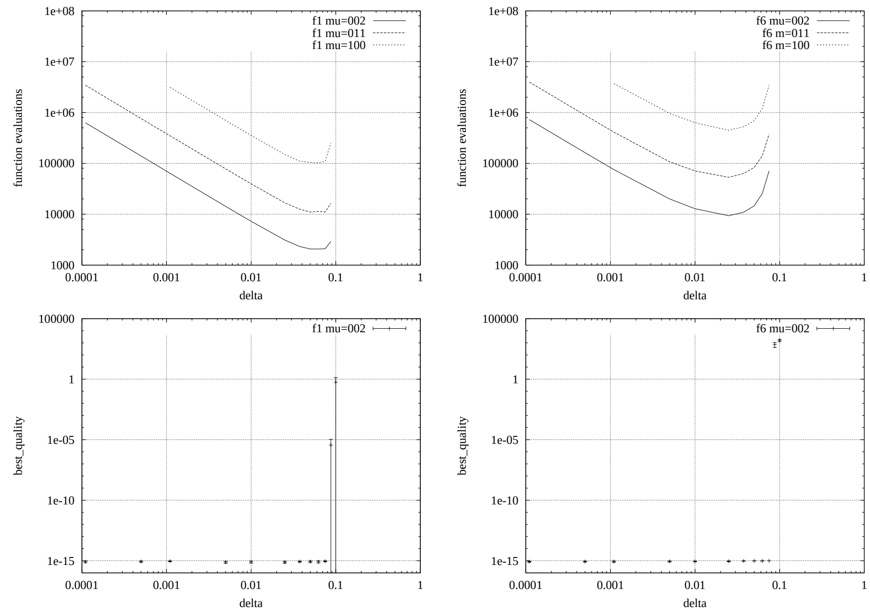

Figure 4: Tests with functions f_1 and f_6: Number of function evaluations and result quality (for $\mu = 2$) in dependency of δ

the results of the test series can be seen in figures 4, 5 and 6. The graphs show the following: The upper graphs show the correlation between δ (abscissa) and the number of function evaluations (ordinate) which were necessary to find the optimum of the test function. For these graphs, only converged runs were regarded. The lower graphs show for the different values of δ the average and the standard deviation of the best function values reached (ordinate). Here, all test runs were considered. These graphs are indicators of the robustness influenced by δ. Only the results for $\mu = 2$ and $\mu = 100$ are shown (figure 6), the results for $\mu = 11$ lie fairly in the middle of those.

5 Results

The graphical representation of the test series shows the reciprocal relation between δ and the number of function evaluations (i. e. the performance of the strategy).

It can be seen that for uni-modal functions, the strategy might be sped up without any loss of reliability by tuning δ to its optimal value of about 0.02 - 0.05 (see functions f_1, f_6, f_{15}). The value of μ has almost no influence on the robustness; therefore only the results for $\mu = 2$ are shown here (lower graphs in figures 4, 5). But μ has a proportional influence on the speed of the strategy, what can be seen in the upper graphs.

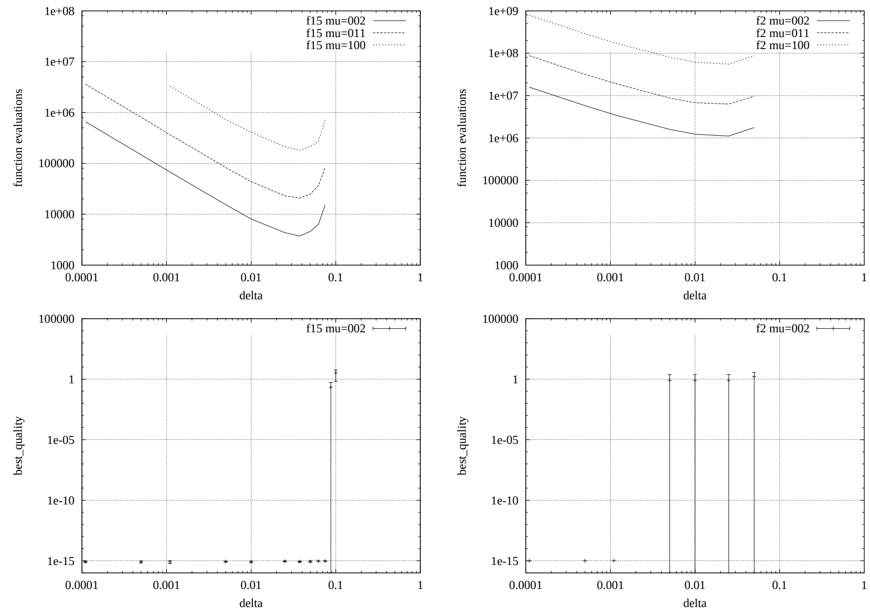

Figure 5: Tests with functions f_{15} and f_2: Number of function evaluations and result quality (for $\mu = 2$) in dependency of δ

For multi-modal functions it is more difficult to choose an optimal δ. The convergence velocity increases with δ like for uni-modal functions (upper graphs in figure 6), but the robustness decreases (lower graphs in figure 6). Thus, the choice of a smaller δ is the slower but safer way to optimize a multi-modal function. Another way to increase the robustness is to increase μ, which can be seen for the functions f_9 and f_{21} in figure 6. The two lower graphs show that for $\mu = 100$ the optimum is more reliably found than for $\mu = 2$. This holds also for function f_2, which is uni-modal but very difficult to solve.

Recapitulating, it can be stated that EN is equivalent or even better than comparable other ES in solving difficult optimization tasks, and the neighborhood attraction parameter δ can be used to tune the convergence velocity and the reliability of the EN.

References

[1] Thomas Bäck. A user's guide to genesys 1.0. Technical report, University of Dortmund, Department of Computer Science

[2] Thomas Bäck, Ulrich Hammel, and Hans-Paul Schwefel. Evolutionary computation. Comments on the history and current state. In *IEEE Transactions on Evolutionary Computation*, volume 1, pages 3–17, April 1997.

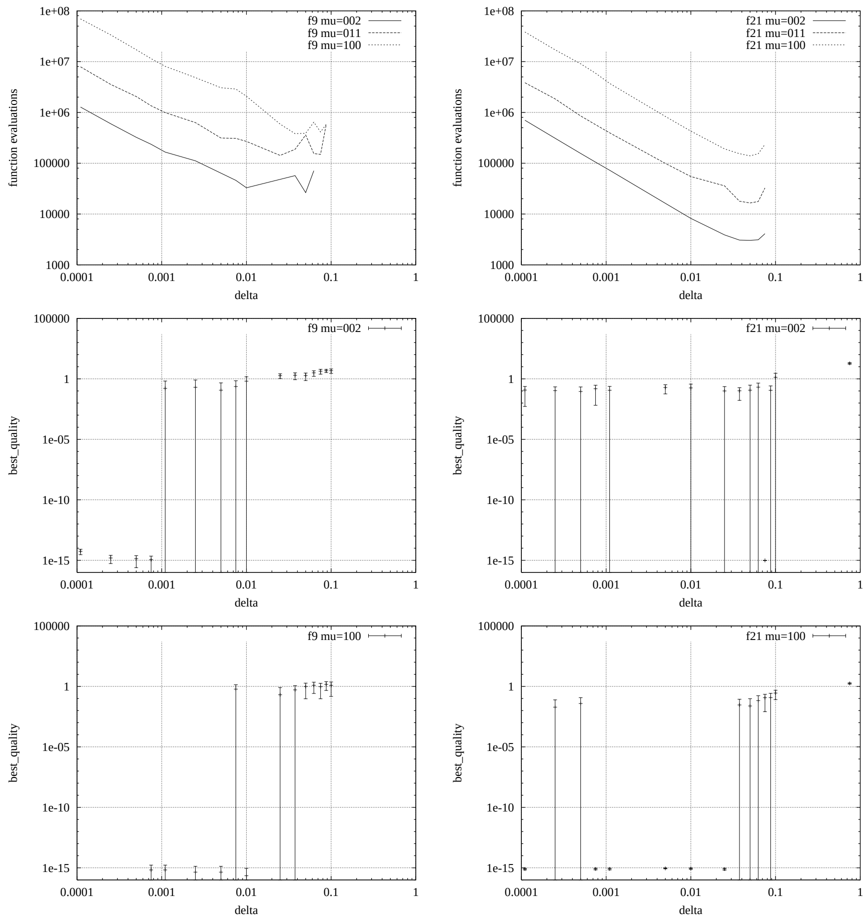

Figure 6: Tests with functions f_9 and f_{21}: Number of function evaluations and result quality (middle for $\mu = 2$; third row for $\mu = 100$) in dependency of δ

[3] Kenneth de Jong. An analysis of the behaviour of a class of genetic adaptive systems. Master's thesis, University of Michigan, 1975.

[4] Jutta Huhse and Andreas Zell. Evolution strategy with neighborhood attraction. In *Proceedings of the Second ICSC Symposium on Neural Computation (NC'2000)*, pages 363–369, 2000.

[5] T. Kohonen. *Self-Organizing Maps.* Springer Verlag, 1995.

[6] Ingo Rechenberg. *Evolutionsstrategie '94.* frommann-holzboog, Stuttgart, 1994.

[7] Hans-Paul Schwefel. *Evolution and Optimum Seeking.* John Wiley and Sons, New York, 1995.