Learning environment dynamics from self-adaptation. A preliminary investigation

Amine Boumaza

Learning Environment Dynamics From Self-Adaptation

A preliminary investigation

Amine Boumaza
LORIA - MaIA
Campus Scientifique B.P. 239
54506 Vandoeuvre-les-Nancy, France
amine.boumaza@loria.fr

ABSTRACT

We present an experimental study that shows a relationship between the dynamics of the environment and the adaptation of strategy parameters. Experiments conducted on two adaptive evolutionary strategies SA-ES and CMA-ES on the dynamic sphere function, show that the nature of the movements of the function's optimum are reflected in the evolution of the mutation steps. Three types of movements are presented: constant, linear and quadratic velocity, in all, the evolution of mutation steps during adaptation reflect distinctly the nature of the movements. Furthermore with CMA-ES, the direction of movement of the optimum can be extracted.

Categories and Subject Descriptors

I.2.6 [**Learning**]: Parameter learning; G.1.6 [**Optimization**]: Stochastic programming

General Terms

Algorithms, Design, Experimentation

Keywords

Evolutionary computation, Self-adaptation, Dynamic environments

1. INTRODUCTION

Self-adaptation has become a very important property in evolutionary computation (EC). The idea of self-tuning strategy parameters by the algorithm during the search has proved very powerful and very successful on a wide range of problems [6, 5, 3]. A substantial body of work regarding self-adaptation in evolutionary computation exists, for a comprehensive overview see [12].

Although adaptation can take place at any stage of the evolution [12], the best-known examples are self-adaptation of the mutation steps. Self-adaptive mutations were originally introduced in evolutionary strategies by H.P. Schwefel [14], however extensions to other areas of evolutionary computation as evolutionary programming [9] and genetic algorithms [4] exist. In this study, we address the case of evolutionary strategies trough the examination of two self-adaptive algorithms.

In dynamic optimization problems self-adaptation plays an important role mainly because it controls the exploration and exploitation phases. In such problems, the goal is not only to find the optimum, but also track it over time [7]. Therefore, the algorithm must adjust its exploration capacity when it looses the optimum once this one moves; self-adaptation is one way of doing it. There exist many studies on self-adaption in dynamical environments see for example [7, 8, 1], however to our knowledge, there are no results that link the self-adaptation process and the dynamics of the fitness landscape.

The aim of this article is to present experimental results on the relationship between the evolution of the mutation steps and the dynamics of the search problem. Section 2 presents briefly the algorithms used in this article, whereas section 3 presents the experiments conducted and comments the results. Finally section 4 presents the discussion and conclusions.

2. SELF-ADAPTIVE MUTATIONS

The idea underlying self-adaptive mutations, is to evolve the parameters of the mutation operator to adjust the behavior of the algorithm to the environment. In most cases, these parameters represent the parameters of the normal distribution used to sample offspring. Different methods of adaptation exist depending on the type of the normal distribution. The simplest case is when this distribution is chosen isotropic (hyper-sphere), here only one step size is adapted. This model can be extended to non-isotropic distributions (ellipsoid), where each coordinate posses its own step size. Finally a further generalization can be made, where the mutation normal distribution has a full covariance matrix. In this article we will only address the case of the last two types of adaptation through two implementation the $(\mu+\lambda)$SA-ES [14] and the $(\mu/\mu,\lambda)$CMA-ES [11].

2.1 SA-ES

In this algorithm, to each coordinate corresponds a mutation step, which is usually encoded within the genotype of the individual. Steps are adapted independently in each

GECCO'05, June 25–29, 2005, Washington, DC, USA.

dimension of the search space and mutation takes place in two steps. At first the parameter vector $\overrightarrow{\sigma} = (\sigma_1, \ldots, \sigma_n)$ is mutated using a log-normal rule:

$$\tilde{\sigma_i} = \sigma_i \exp\left(\tau N(0,1) + \tau' N_i(0,1)\right) \quad \forall i \in \{0, \cdots, n\} \quad (1)$$

here, N_i are n independent realizations of a normal random variable, whereas N is a single realization common to all σ_i. τ and τ' are fixed parameters representing learning rates. In the second step, the individual $\overrightarrow{x} = (x_1, \ldots, x_n)$ is mutated by adding it to a normal vector whose components have $(\tilde{\sigma_1}, \ldots, \tilde{\sigma_n})$ as variances.

$$\tilde{x_i} = x_i + N_i\left(0, \tilde{\sigma_i}^2\right) \quad \forall i \in \{0, \cdots, n\} \quad (2)$$

2.2 CMA-ES

The CMA-ES algorithm introduced by N. Hansen [10, 11] is to this day one of the best performing algorithms based on covariance matrix adaptation. The main advantage of this algorithm compared to SA-ES is its invariance against rotations. The axes of the normal distribution are independent from the coordinate system which allows invariance against rotations of the search space.

The mathematical details of the algorithm were intentionally left out for the purpose of this article, the reader is encouraged to consult the referenced work for full details.

In CMA-ES, offspring at generation (g) are sampled from the following normal distribution:

$$X_k^{(g+1)} \sim \mathcal{N}\left(\langle x\rangle_\mu^{(g)}, {\sigma^{(g)}}^2, C^{(g)}\right) \quad (3)$$

in which $\langle x\rangle_\mu^{(g)}$ is a weighted average of the μ best individuals at generation (g), C is a positive definite matrix representing the covariance matrix, and σ is the mutation step size, note that in this case there is only one step size.

Adaptation in CMA-ES takes place in two steps: the adaptation of the covariance matrix C and the adaptation of the mutation step σ. These steps need not occur at the same time, in fact, they usually occur at different time scales.

The step size is adapted using cumulative step size adaptation (CSA) [13] which in a nutshell adapts it using information on the evolution path (a sequence of mutation steps). If consecutive mutation steps have correlated directions then fewer larger steps could have been applied to cover the same distance. Conversely, uncorrelated mutation steps are the result of large steps that make the algorithm oscillate back and forth. In a similar fashion, the covariance matrix is adapted using the evolution path and the successive differences between the mean population vectors at generations g and $g+1$.

Unlike SA-ES whose distribution ellipsoid is dependent of the coordinate system, CMA-ES allows the realization of any normal distribution. The ellipsoid has its own orthogonal basis described by the eigenvectors of the covariance matrix.

3. EXPERIMENTAL STUDY

3.1 Description of the scenario

We are interested in studying the evolution of the mutation steps over time, for that we have conducted experiments using the above described algorithms on a dynamic sphere function:

$$\mathcal{F}(X) = min \sum_{i=1}^{n} (x_i - \hat{x}_i)^2 \quad (4)$$

were $\hat{X} = (\hat{x}_1, \ldots, \hat{x}_n)$ represents the moving optimum over time. Table 1 shows the different parameters used in the experiments in which "period" represents the number of generations between the movements of the optimum. In the case of CMA-ES, we kept the default parameters suggested by N. Hansen [11]. The results illustrate the evolution of the mutation steps over time and the error to the optimum, in the case of the CMA-ES we also present the direction of the search i.e. the direction of the eigenvectors of the covariance matrix. Due to the stochastic nature of the algorithms, each experiment was conducted 1000 times and the figures represent the averaged results with their respective standard deviations. The algorithms were left evolving throughout the changes with no explicit actions taken when changes occur.

Table 1: Parameter settings for the presented experiments.

Parameters	Algorithm	
	SA-ES	CMA-ES
Dimension n	2	2
Population size μ	5	3
Offspring λ	10	6
Initial step size	5.0	0.5
τ and τ'	0.3 and 0.4	n.a.
Period (gen)	50	50

In this study, we were interested in three types of movements of the optimum: movements with constant velocity, movements with linearly increasing velocity and movements with quadratically increasing velocity.

3.2 Results and discussion

Experiments with SA-ES

Figures 1 trough 7 represent the evolution of the mutation steps and the error of the best individual over time in the case of SA-ES. In all experiments the optimum moves every 50 generations. Figure 1 represents the fluctuations of σ_0 when the optimum moved with a constant velocity in one dimension (severity of 1.0). This is the effect of self adaptation, the mutation step increases to favor exploration when the algorithm is no longer at the optimum, and then gradually decreases when it starts converging toward the optimum. Figure 2 shows the error to the optimum for that case. The interesting part about this experiment, is that the maximum value σ_0 takes remains constant throughout the movements of the optimum.

In the case where the optimum moves with linearly increasing velocity, the mutation steps behave differently, figure 3 shows the fluctuations of σ_0. In this experiment the optimum moves with linearly increasing severity i.e. the distance traveled increases linearly. Here the results show that at each movement of the optimum the mutation steps follow a similar behavior and attain a maximum value that increases linearly over time. This values reflects the search distance or the width of the normal distribution for the mutation needed to find the optimum. Figure 4 shows the error of the best individual.

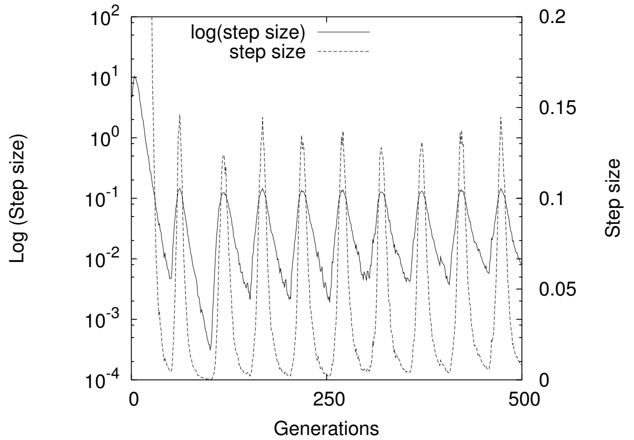

Figure 1: Evolution of the mutation step (both in log and regular scale) in the case of SA-ES when the optimum moves with constant velocity in the dimension x_0

In the third experiment conducted on the SA-ES, the optimum moved with a quadratically increasing velocity in other words the distance (severity) traveled by the optimum increased quadratically. Here again the mutation steps behave in a similar fashion with the movements of the optimum, and attain a maximal value that increase quadratically overtime. Figure 5 shows the evolution of σ_0 and figure 6 shows the error of the best individual to the optimum.

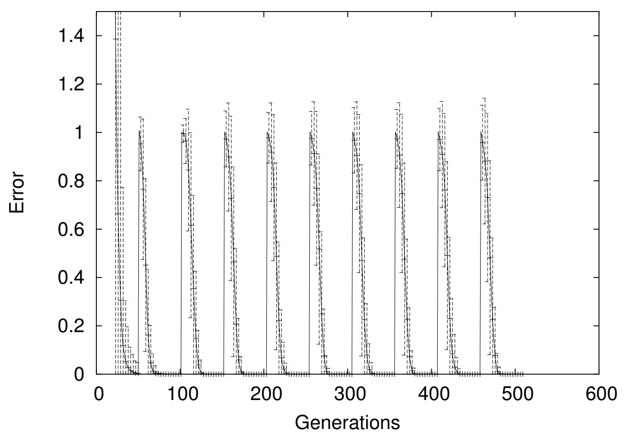

Figure 2: Evolution of the error in the case of SA-ES when the optimum moves with constant velocity in the dimension x_0

Figure 7 show a case where the optimum moves with constant speed (severity 1.0) in the first dimension and then shifts direction at generation 250 to move in both dimensions at once with a 60 degrees heading. In the first phase of this experiment, only σ_0 adapts to the changing environment, σ_1 keeps low values during this phase (no exploration needed in this dimension). However when the optimum changes direction, σ_1 starts evolving and attain higher values than σ_0 due to the fact that with a 60 degrees the optimum moves more in the second dimension than in the first one. However since the axes of the mutation distribution are linked to the coordinates system, the heading of the optimum cannot be known directly. In the next set of experiments with the CMA-ES we show that the direction of movement could be easily known.

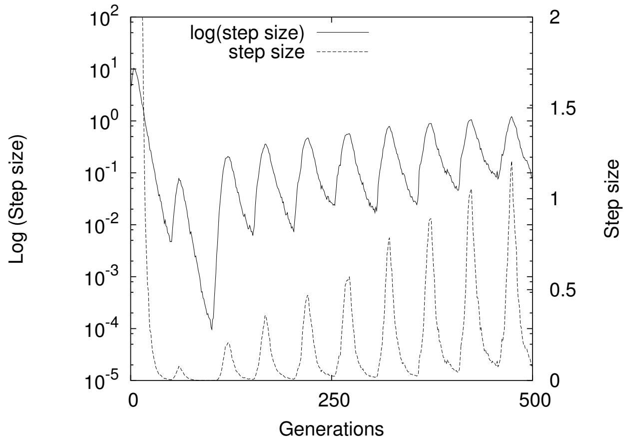

Figure 3: Evolution of the mutation step (both in log and regular scale) in the case of SA-ES when the optimum moves with linearly increasing velocity in the dimension x_0

Experiments with CMA-ES

In this part, we present the experiments conducted using the CMA-ES algorithm. In addition to the mutation steps, we show the direction of the dominant eigenvector. The dominant eigenvector is the largest vector in the eigenvector basis of the covariance matrix. These experiments show that not only the mutation step reflect the amplitude of the optimum's movements, the eigenvector follows the direction of the optimum.

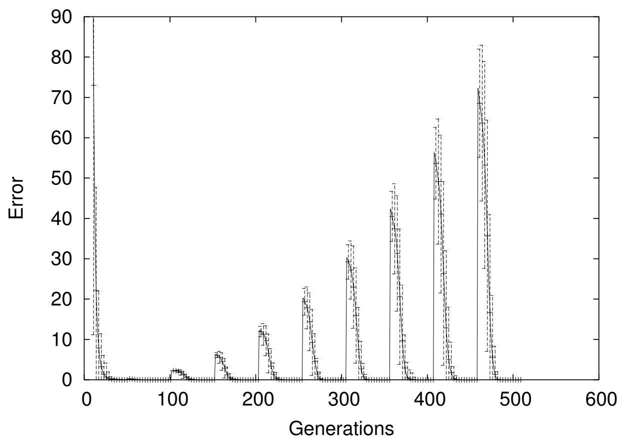

Figure 4: Evolution of the error in the case of SA-ES when the optimum moves with linearly increasing velocity in the dimension x_0

In the first experiment of this series, the optimum moves with constant velocity with severity one in the first dimension. Figure 8 illustrates the direction of the eigenvector when the optimum moves in the first dimension. It shows that the angle of the eigenvector maintains a constant direction at angle zero throughout the experiment; this corresponds to the direction in which the optimum moves. However from figure 8, we can't describe the nature of movement, we will have to inspect the evolution of the mutation step (figure 9) which behave in a similar fashion as in the case described by figure 1, where step sizes attain a constant maximum value.

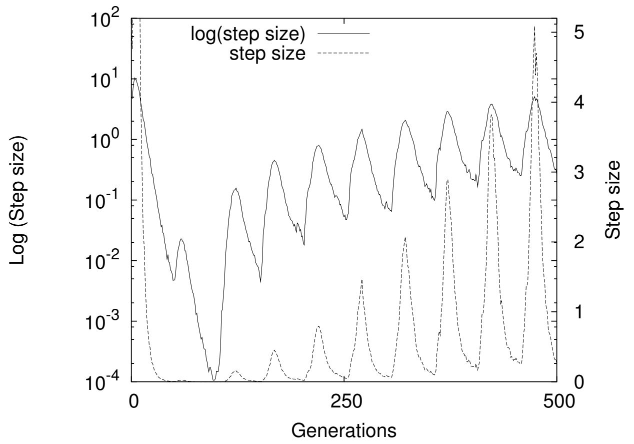

Figure 5: Evolution of the mutation step (both in log and regular scale) in the case of SA-ES when the optimum moves with quadratically increasing velocity in the dimension x_0

In the second experiment, the optimum changes direction during the evolution. In the first phase, the optimum moves on the first dimension as in the experiment above. Afterward it shifts direction by 45 degrees during the second phase. In phase three it moves back in the direction of the first dimension then sifts again in phase four to -45 degrees. Throughout all phases, the optimum maintains a constant velocity. The evolution of the direction the eigenvector takes is shown on figure 10, in which we can clearly distinguish the different phases of the movement.

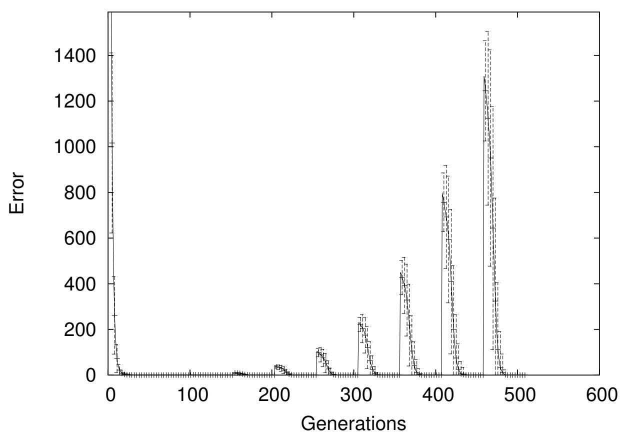

Figure 6: Evolution of the error in the case of SA-ES when the optimum moves with quadratically increasing velocity in the dimension x_0

Figures 12 and 13 illustrate the case where the optimum moves with linearly increasing velocity in both dimensions at once (45 degrees heading). They correspond respectively to the direction of the eigenvector and the evolution of the mutation step. In this case we also note that the mutation step increases linearly over time. Similarly, figures 14 and 15 show the same entities in the case were the optimum moves with quadratically increasing velocity. In this experiment the optimum moved in the first dimension.

3.3 Discussion

The aim of mutation adaptation in evolutionary computing is to control the step size in order to adapt the algorithm to the fitness landscape. This adaptation process is generally expressed in the evolution of the step sizes; this evolution is characterized by two phases: exploration and exploitation. The exploration phase represents the period during which the algorithm is searching for better performance, most mutations during this period are unsuccessful and the step sizes increase. When the algorithm starts to gain performance, it enters the exploitation phase, mutations become more successful and the step size reduces as the algorithm approaches the optimum. Mutation steps increase during exploration and decrease during exploitation.

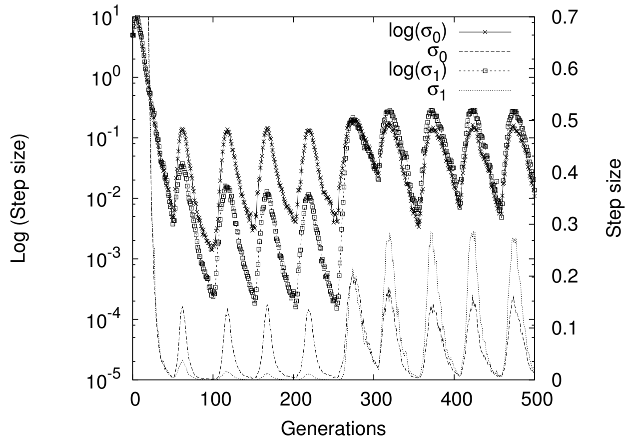

Figure 7: Evolution of the mutation steps in the case of SA-ES when the optimum moves with constant velocity with a change of direction at generation 250

In dynamical environments the algorithm is constantly switching between exploration and exploitation, and step sizes oscillate between high and low values. Once an optimum is lost, the adaptation process increases mutation steps to raise exploration until it finds it again.

In the case where the optimum moves with constant velocity, the step sizes attain a constant hight value after the exploration phase and then drop during the exploitation phase. Furthermore, in the case where the optimum moves with linearly increasing velocity, the maximum values attained by the mutation steps are not constant but increase linearly over time, which is in concordance with the movements of the optimum. Finally, when the velocity of the optimum increases quadratically, the maximum values attained by the mutation steps increase quadratically also. These results show that there exist a correspondence between the nature of the dynamics in the environment and the evolution of the step sizes. This correspondence exists with both adaptation techniques.

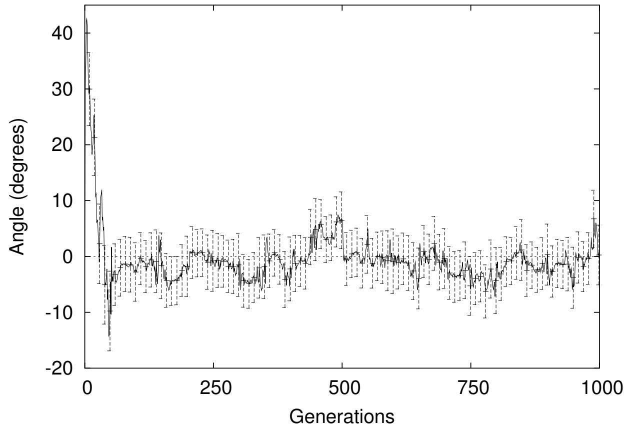

Figure 8: Direction of the dominant eigenvector in the covariance matrix over time when the optimum moves in the dimension x_0 with constant speed.

It is conceivable to exploit the information given by the step size evolution, in order to for example detect the dynamic of the environment and further adapt the algorithms to the different movements. Even though our experiments were conducted on a simple unimodal function, we believe that in multi-modal environments the evolution steps should behave similarly. In that case we expect to track the evolution of different optima at once.

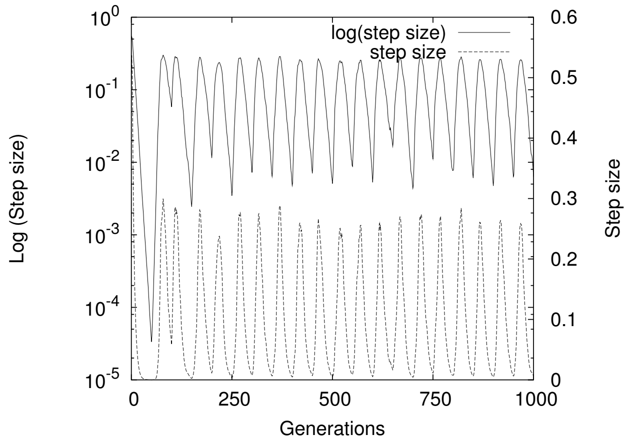

Figure 9: Mutation steps (both in log and regular scale) over time when the optimum moves in the dimension x_0 with constant speed in the case of CMA-ES.

We have also noticed in our experiments that using high severity for the optimum's movements, as in our case, the algorithm needs several generations to adapt its parameters. Experiments not presented here with 10 generations between changes in the environment, showed that the algorithm was not able to adapt its mutation steps before the change occurred. However that was not the case when the severity was low, with a small number of generations the algorithm could adapt its steps. This observation implies that, in order to extract useful results on the dynamics of the environment, we have to give the algorithm enough time to adapt. However most of the times we don't have such a possibility, since we don't control the dynamics.

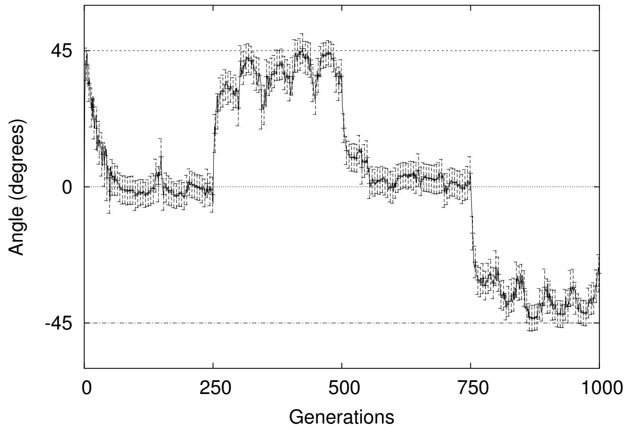

Figure 10: Direction of the dominant eigenvector in the covariance matrix over time when the optimum changes direction.

Finally, our study lacks the investigation concerning the link between the values of the mutations steps and the velocity values of the movements. We believe, that the amplitude of the step sizes could indicate the distance traveled by the optimum. In figure 7 where both mutation steps are shown, we notice that since the movement in the second dimension is greater than the movement in the first one, the steps have different amplitudes. It would be interesting to make a link between the value of the step size and the displacement.

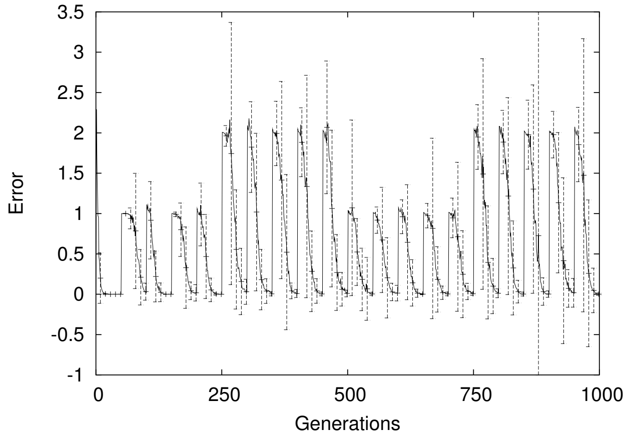

Figure 11: Evolution of the error when the optimum changes directions in the case of CMA-ES.

In experiments where the optimum shifts direction, we show how the different steps evolve. In the case of CMA-ES, the direction of the optimum's movements is learned be the covariance matrix and characterized by the principal axis of the mutation ellipsoid. This is due to the invariance of CMA-ES against rotations. This is not the case with SA-ES, since it is dependent of the coordinate system. The results show that if the optimum moves further in one dimension than in the others, the step size on that dimension attain higher values during adaptation. We believe though not shown here, that the heading can by inferred from the difference in proportions of the mutation steps. For example if the σ_0 is twice the value of σ_1, we could say that the optimum moves roughly at a 25 degrees heading.

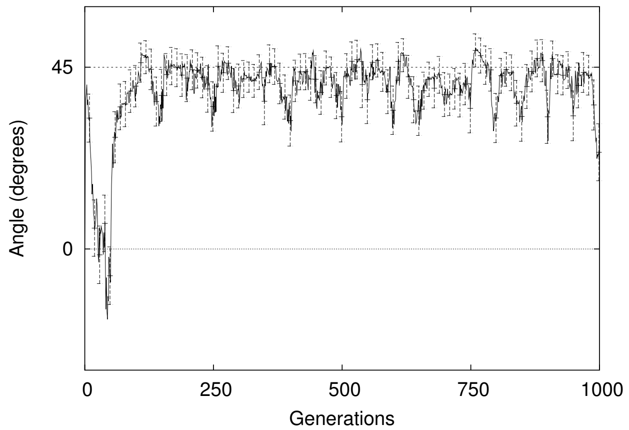

Figure 12: Direction of the dominant eigenvector in the covariance matrix over time when the optimum moves with linearly increasing velocity in both dimensions.

Theoretical results [2] showed that the convergence of the step sizes is linear in log scale, this property is demonstrated here, the step size in all experiments converge lineally. Although the theoretical results apply to the $(1, \lambda)$ strategy, it is interesting to notice that they are verified in the case where the population size is greater than one.

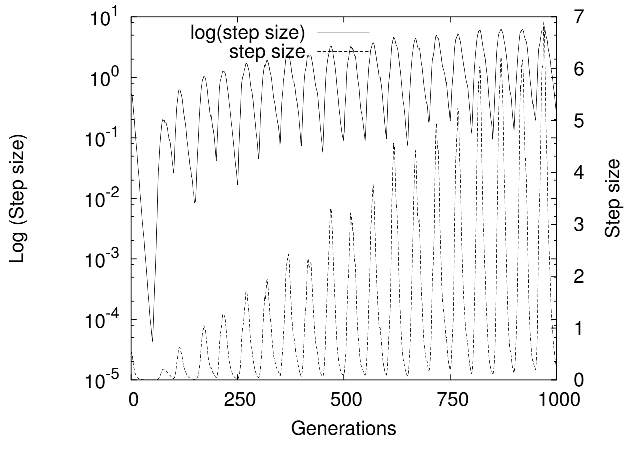

Figure 13: Evolution of the step size (both in log and regular scale) over time in CMA-ES when the optimum moves with linearly increasing velocity in both dimensions.

4. CONCLUSIONS

In this article we showed through experimentation that there exists a relationship between the movements of the dynamic sphere model and the evolution of the self-adapted mutation steps. Experiments on both SA-ES and CMA-ES algorithms show that the nature of the movement of the optimum is reflected in the evolution of the mutation steps. In the case of CMA-ES we have shown that not only the nature of the movements is reflected in the mutation step, the direction of movement of the optimum is learned by the covariance matrix. The dominant eigenvector has the same direction as the moving optimum.

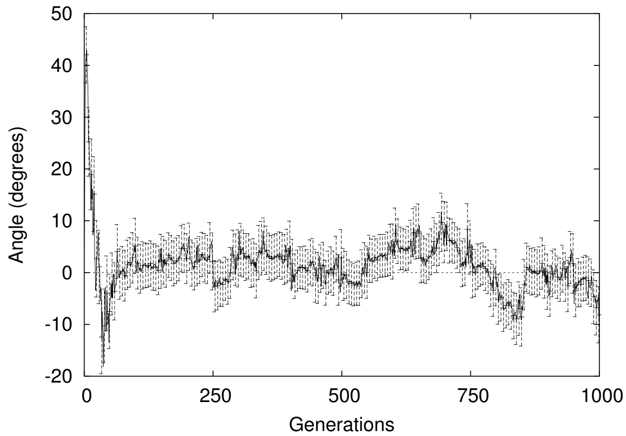

Figure 14: Direction of the dominant eigenvector in the covariance matrix over time when the optimum moves with quadratically increasing velocity.

This, as stated in the introduction, is only a preliminary study, further experiments should be conducted on different fitness models and higher dimensions than presented here. It is also important to experiment on the multi-modal case, where several moving optima need to be detected. Finally, it is interesting to see such relationships between the dynamics of the search space and the evolution of the mutation parameters; however what would be more interesting and by the same means challenging, is to use the information learned from the adaptation in a meaningful manner, say for example privilege certain directions of the search space, or tune the internal parameters to the specific dynamics of the problem.

5. ACKNOWLEDGMENTS

I would like to warmly thank Nikolaus Hansen for providing me with an implementation of the CMA-ES algorithm and for his useful comments.

6. REFERENCES

[1] P. J. Angeline. Tracking extrema in dynamic environments. In P. J. Angeline, R. G. Reynolds, J. R. McDonnell, and R. Eberhart, editors, *Proceedings of the 6th International Conference on Evolutionary Programming*, volume 1213 of *Lecture Notes in Computer Science*, Indianapolis, Indiana, USA, April 13-16 1997. Springer Verlag.

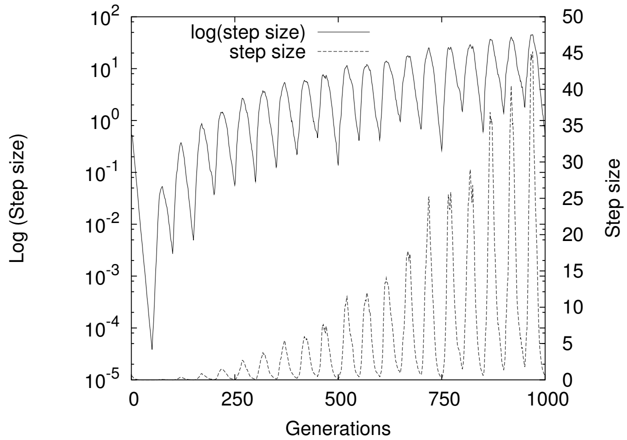

Figure 15: Evolution of the step size (both in log and regular scale) over time when the optimum moves with quadratically increasing velocity in the case of CMA-ES.

[2] A. Auger. Convergence results for (1,λ)-SA-ES using the theory of ϕ-irreducible markov chains. *Theoretical Computer Science*, 2004. In press.

[3] T. Bäck and H. P. Schwefel. Evolution strategies. In J. Périaux and G. Winter, editors, *Genetic Algorithms in Engineering and Computer Science*, chapter 6-7. John Wiley and Sons, 1995.

[4] T. Bäck. Self-adaptation in genetic algorithms. In F. J. Varela and P. Bourgine, editors, *Proc. of the 1st European Conf. on Artificial Life*, pages 227–235, Cambridge, MA, 1992. MIT Press.

[5] H.-G. Beyer. Evolution strategies: A comprehensive introduction. *Natural Computing*, 1(1):3–52, 2002.

[6] H.-G. Beyer and K. Deb. On self-adaptive features in real-parameter evolutionary algorithms. *IEEE Transactions on Evolutionary Computation*, 5(3):250–270, 2001.

[7] J. Branke. *Evolutionary Optimization in Dynamic Environments*. Kluwer, 2001.

[8] J. Branke and W. Wang. Theoretical analysis of simple evolution strategies in quickly changing environment. In E. Cantu-Paz et al., editor, *Proceedings of the Genetic and Evolutionary Conference*, pages 537–548. Springer, 2003.

[9] L.J. Fogel, P.J. Angeline, and D.B. Fogel. An evolutionary programming approach to self-adaptation on finite state machines. In J. McDonnell, R. Reynolds, and D. Fogel, editors, *In Evolutionary Programming IV: Proceedings of the Fourth Annual Conference on Evolutionary Programming.*, pages 355–366. MIT Press, 1995.

[10] N. Hansen, S.D. Müller, and P. Koumoutsakos. Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES). *Evolutionary Computation*, 11(1):1–18, 2003.

[11] N. Hansen and A. Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159–195, 2001.

[12] R. Hinterding, Z.Michalewicz, and A.E. Eiben. Adaptation in evolutionary computation: A survey. In *Proceedings of the 4th IEEE International Conference on Evolutionary Computation*, pages 65–69, 1997.

[13] A. Ostermeier, A. Gawelczyk, and N. Hansen. A derandomized approach to self-adaptation of evolution strategies. *Evolutionary Computation*, 2(4):369–38, 1994.

[14] H. P. Schwefel. *Numerical Optimisation of Computer Model*. Wiley, 1981.