Application of Evolution Strategy in Parallel Populations

Reinhard Lohmann
Technische Universität Berlin, Fachgebiet Bionik und Evolutionstechnik
Ackerstr. 71-76, Sekr. ACK 1, D-1000 Berlin 65

Introduction

Evolution strategy has turned out to be an advantageous optimization strategy which is applicable in a lot of different technical problems. The algorithm has been developed originally for systems described by continuously adjustable variables [1] [2], but can also be used successfully in discrete problems [3]. In the first part of the paper a few examples of applications are presented preceded by a short introduction to the algorithm of evolution strategy. The examples have been worked out at the department of "bionics and evolution techniques" (Prof. Rechenberg), TU Berlin, using single population systems.

In the second part of the paper the algorithm of evolution strategy is extended to a system of many populations. The objective of this extension is to establish a floating coordinate system which is defined by the connections between solutions of the different populations and which usually is not orthogonal. Within these flexible coordinate systems, for each population a different one, some optimization problems could be solved on which the single population system described in the first part of the paper refused to show sufficient convergence velocities. As a first approach to the multi population system some observations on a so called filter problem are reported. Experimental investigations of a second problem, a curve fitting task, show that the flexible coordinate systems in the parallel populations adapt to difficult quality functions, i.e. if the direction of highest gradient is not identical with one of the coordinate axes and if the shape of the quality function demands for extremely different stepsizes due to each variable.

Evolution strategy

According to Darwins theory about the development of species the most important features of the evolution process are inheritance, mutation and selection. Its only these properties of the biological evolution which had to be translated into mathematical terms to formulate a most general and effective optimization method, the evolution strategy by Rechenberg, 1964 [4].

The algorithm of evolution strategy can be described as follows. The individuals taking part in the evolution process usually consist of n object variables O_1, O_2, ... O_n, which are gathered in an object vector $\underline{O}$ and can continuously be adjusted. A descendant object $\underline{O}_N$ is generated by mutation of a parent object $\underline{O}_E$ according to equation 1.

$$\underline{O}_N = \underline{O}_E + S_N * \underline{Z} \tag{1}$$

S_N is the stepsize assigned to the descendant $\underline{O}_N$ as a strategy variable and $\underline{Z}$ is a random vector whose components are Gaussian distributed numbers. If there are more than one descendants the stepsize can be mutated and inherited as well. Equal proportions of λ descendants get a smaller stepsize than the parent or a greater respectively and the mutations of the stepsizes are carried out by multiplication with or division by a constant f=1.5. Thus λ descendants are generated from μ parents. A quality function Q provides for the criterion in the selection process in which either both parents and descendants can take part - (μ + λ)-evolution strategy - or the descendants only which is written as (μ , λ)-ES. The μ best objects are chosen as the parents of the next generation and the procedure is repeated.

If there are two or more parents recombination can be introduced between ρ parent objects which means that the components of the object $\underline{O}_E$ in equation 1 are composed by a random choice among the variables in the ρ parents. In that case, which is written as (μ/ρ +, λ)-ES, the stepsize is the mean value of the stepsizes of all parents involved. Note that there are no mechanisms to isolate objects from each other unless the strategy consists of (μ' + λ') parallel populations. The most general formulation of evolution strategy is given in equation 2 where γ is the number of generations during which the populations are isolated [5].

$$[\ \mu'/\rho' +,\ \lambda' (\ \mu/\rho +,\ \lambda\ )^{\gamma}\ ] \text{-ES} \tag{2}$$

A necessary demand for evolution strategy being applicable is that the quality function complies with strong causality. Strong causality means that small changes of the object variables result in comparatively small changes of the quality and big changes of variables lead to big alterations of quality. If strong causality is violated any systematical search for minima or maxima is impossible if it is based on local detection of the direction for the highest gradient of quality in the n-dimensional variable space. In fact evolution strategy belongs to the group of hill-climbing-methods although it does not depend on calculating partial derivatives [6]. That means that evolution strategy is still able to work even if the quality function is explicitly unknown.

The most important property of evolution strategy is its adaptive stepsize control. Rechenberg [7] has shown for a quadratic function, which is a most representative case for quality functions, that the rate of progress Φ (in universal notation) is rather sensible to the stepsize Δ. Only within a narrow band of stepsize, called the evolution window (see fig.1) evolution takes places with a remarkable rate of progress. Larger stepsizes are due to regression while too small stepsizes cause stagnation of the process.

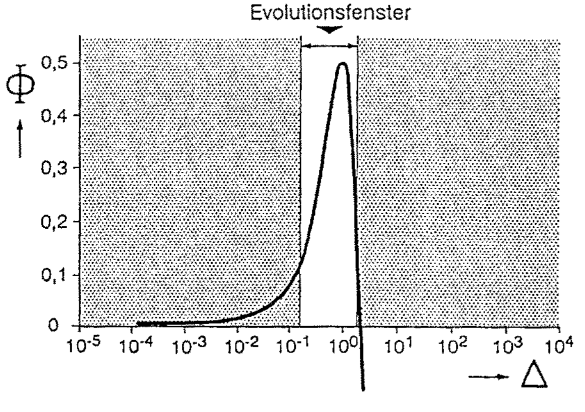


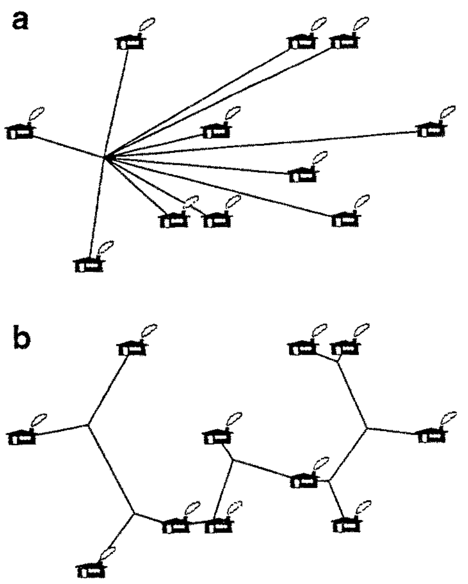
Fig. 1: Evolution window.

Applications of evolution strategy in single populations

Now the evolution strategy can be applied without great difficulties to solve different problems some of which already have been reported [5].

Supply network

The first of the examples is the design of a supply network for a number of randomly distributed houses. Whatever the houses have to be supplied with, gas, water or information, the shortest connection between each house and the source is obviously not the best solution (Fig. 2a). A good solution requires that the total length of the supply network is as small as possible if the quality of the net is regarded as the costs for the implementation. That means that $Q = \Sigma \; r_i$ has to be minimized where r_i are the lengths of single connections.



A variable system can be established by letting the pipelines branch into two and by shifting the branching points to different places. The complete set of coordinates of the branching points in the fixed network structure is the object which is exposed to evolution strategy algorithm. The result is plotted in figure 2 b where the total length of the connections is by 40% smaller than that in figure 2a.

Fig. 2: Supply network for a number of houses. a) Direct connections with the source. b) Network developed by evolution strategy.

Girder-bridge

The second example for the application of evolution strategy is a two-dimensional framework of a girder-bridge [8]. Figure 3 shows a randomly configurated construction usually used as a start for the optimization process. Dotted lines indicate tensile load, the other girders are under compressive load. Thicknesses of lines represent thicknesses of the girders as they are calculated from the load distribution in the construction.

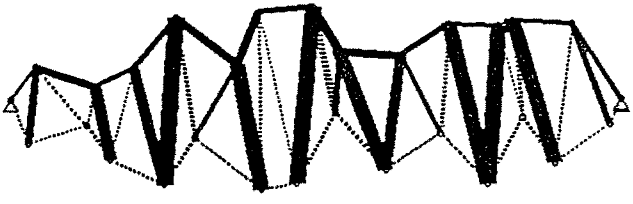


Fig. 3: Random configuration of a bridge.

The objektive is minimization of the total weight of the bridge and the variables are the positions of the nodes connecting the girders.

Results will depend on restrictions and on the distribution of external load. Figure 4a shows a bridge for the case of evenly distributed load and top nodes kept on a horizontal line between the bearings of the bridge. A completely different shape emerges from the case of two locally applied external loads of opposite direction. Obviously there are no restrictions to the positions of nodes in the construction in figure 4b.

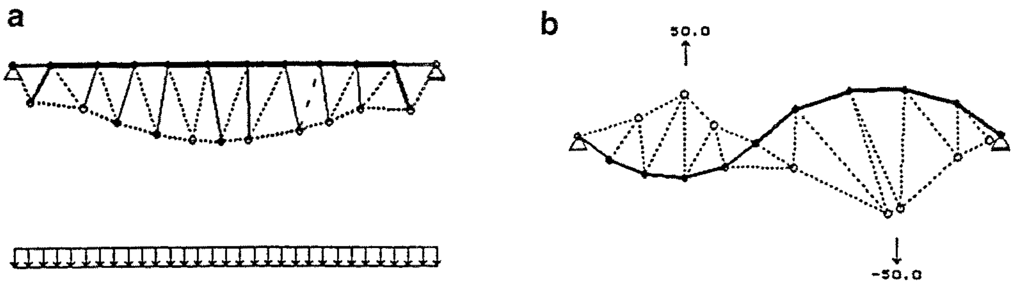
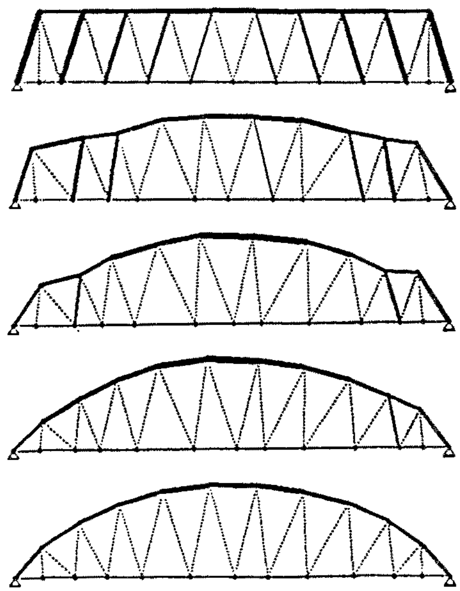


Fig. 4: Results of evolution depend on restrictions and on external load distribution.



The process of conversion of a box-like shaped bridge into a round one is shown in figure 5. For some purpose here the bottom nodes are fixed to keep on the horizontal line and external load is again evenly distributed. The conversion is propagated by nothing else than mutation and selection according to the criterion that the weight of the construction should be as smallas possible. The task was carried out by a (6/6,60)-ES, which took only 100 generations reducing the total weight of the bridge by about 50%.

Fig. 5: Conversion of a bridge. Shapes plotted after 1, 10, 20, 30 and 100 generations.

Evolution of a lense

The third example concerns the development of the eye which has often been used as an argument against Darwin's theory. Critics claim that an unlikely lot of big mutations precisely tuned to each other had to occur at the same time in the same individual so that a complicated organ like an eye could have emerged. But the evolution always takes small steps. The first one, for instance, could have been the agglomeration of some photosensitive cells followed by the formation of a small dip. That would mean for the individual concerned that it got a first poor capability of detecting the direction where light comes from. If the whole thing is covered by some transparent tissue, again step by step, that would be no mentionable improvement of those capabilities of detecting directions of light sources, but it would mean an advantageous protection of the organ. The meaning of the transparent protection tissue, however, can change. If the thickness of the transparent layer, assumed to be flat at the beginning, undergoes locally different changes this protection skin will start to focus the light on a smaller area. Each bit of such a process would improve the sensitivity to directions of light sources which is the basis for more intelligent visual information processing.

The last part of the development can easily be simulated on a computer. Thicknesses d_i of the transparent body in figure 6 are the variables of the objekt which is calculated as a pile of prisms deflecting the in coming light beams. The quality function is the sum of squared distances q_k between point P and the points on the photosensitive layer where the deflected beams hit.

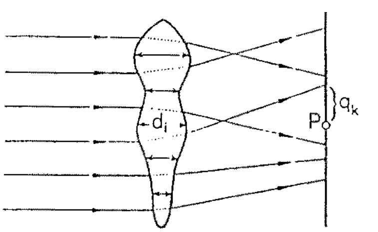


Fig. 6: Deformable transparent body subjected to evolution strategy.

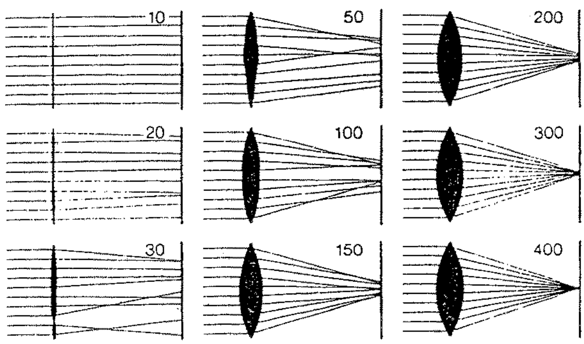


Fig. 7: Stages in the evolution of a lense.

Minimal quality is the selection criterion. Figure 7 shows a few intermediate stages of the design. the last one is a collecting lens which most of vertebrate eyes are equipped with. I think these are some arguments against Darwin's critics.

Difficult problems

The three examples of applications described above are all good-natured with respect to the algorithm of evolution strategy. The quality functions comply with the necessary demand for strong causality and there is no need for giving individual stepsizes to each variable: Convergence velocity seems to be sufficient.

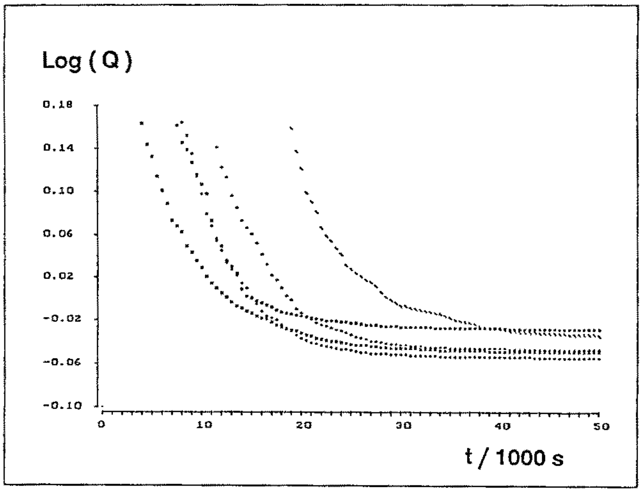
But things are not easy like that in any application problem. If the path of highest gradient in the quality function is a very narrow ridge and, in addition, if the direction of the ridge deviates from the direction of a single variable, only appropriate transformations of the system can help to provide for sufficient convergence velocities as has been pointed out by Schwefel [9],[6].

Observations on the "filter problem"

A superficial look at a problem to be solved shows in these cases that convergence slows down, that the mutated stepsize becomes too small and that the solution seems to be trapped at different places in variable space which could depend on the start configuration. A cheap interpretation is the suspicion that local minima in the quality function were responsible for the failure of the optimization algorithm.

One of these "intreatable" applications has been investigated in more detail. The task was to find 24 coefficients O_i in a local filter for a particular visual system. The filter detects the edge length in binary pictures which depends usually on the rotation of the pattern. The objective was to minimize rotation dependance of the edge length which is caused by the raster of the screen. The quality function is given in equation 3 with 24 variables O_i and B*24 constant numbers $m_{i,b}$ calculated from the set of B=72 test pictures. The objektive of the task is minimization of quality.

$$Q = 1/B \sum_{b=1}^{B} [L - (\sum_{i=0}^{24} O_i * m_{i,b})]^2 \tag{3}$$



The filter problem has been attacked in a single population with different types of strategies, (1,40)-ES, (4,40)-ES, (8,40)-ES, (8/4,40)-ES and (8/8,40)-ES. Quality curves of 5 of these experiments are plotted versus execution time in figure 8.

Fig. 8: Quality versus time for 5 independant evolution strategy experiments.

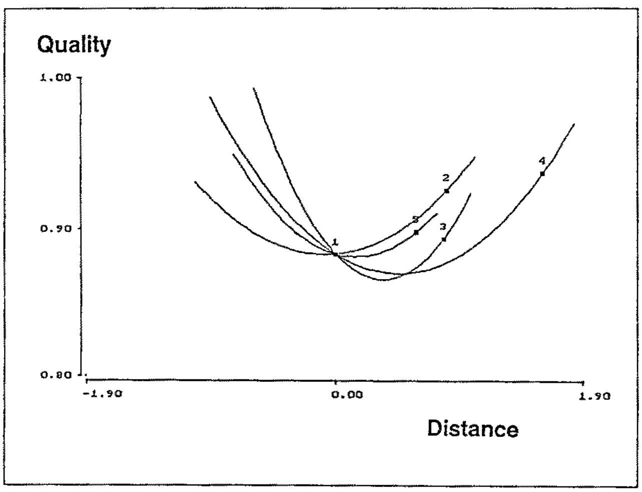


Fig. 9: Quality versus euclidian distance on 24-dimensional straight lines between solution 1 from fig.8 to the other independant four solutions Nr. 2,3,4,5.

All processes slow down and seem to converge to different levels of quality which suggests the assumption of local minima. But this assumption turns out to be false if one has a look at the quality along particular directions. This has been done for the five solutions from figure 8 by detecting the quality on straight lines in the directions between solution 1 and the others. These curves are plotted in figure 9 versus euclidian distance from the point representing solution 1. Obviously none of the 5 solutions is a local minimum.

These particular directions are very unlikely to be discovered by the usual mutation procedure according to equation 1. This is demonstrated in figure 10a through 10d. Figure 10a shows the quality along the direction from F1 to F2 referring to the solutions 1 and 3 in figure 9. The curve is a parabel with a rather small curvature at its center. Note that the vertical line starting from F1 actually is a choice of 3 parabels representing quality on random directions in the variable space. The parabels can be recognized only in the subsequent magnifications of the area around F1 in the diagram 10a (see fig. 10b-10d).

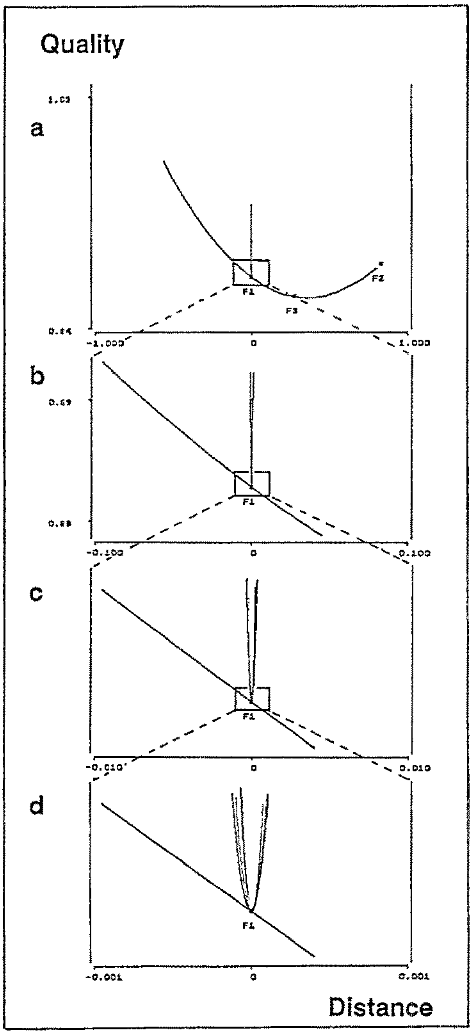


Fig. 10: a) Comparison of quality function in a particular direction with those in random directions. b-d) subsequent magnifications of 10a by factors 10.

The random parabels show much higher curvatures at their centers and we understand why the single population strategies stagnate: Stepsizes always adapt to the highest curvature of a quality function which means here that the stepsizes become very small [2]. Unless one succeeds to replace the stepsize S_N in equation 1 by an appropriate matrix, which transforms the mutation hypersphere into an adapted hyperellipsoid, all extremely unisotropic problems will keep effectively unapproachable by usual evolution strategies and, of course, by other optimization methods [6].

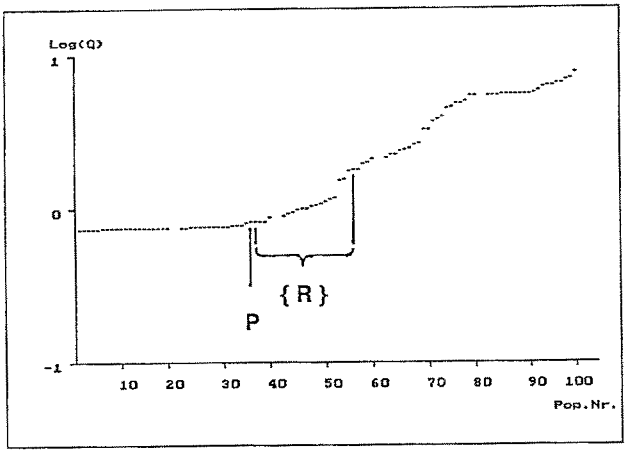
These observations give a first hint how a system can be established which developes the necessary transformation without explicit knowledge of the quality function. The extension of evolution strategy to a system of many isolated and parallel populations enables to extract the transformation from the directions between the different populations.

Multi population evolution strategy

The extension of evolution strategy is performed by establishing a system of, for example, 100 independent populations. In each population a usual (1,10)-ES is performed.

This does not make any sense unless some information is exchanged between the populations. Therefore the populations first are sorted after each generation according to their qualities. With the assumptions that objects within a certain quality range are next neighbours in the variable space and, further, that the pathway of highest gradient is a rather straight line it should be advantageous to use the vectors between neighboured populations as a coordinate system replacing the cartesian system. The number of adjacent populations installing the noncartesian system for a population P is here limited to 20 (see figure 11). For each descendant in population P a random partner R_i from the allowed set { R } is chosen and the new variables are calculated according to equation 4.

$$\underline{Q}_{PN} = \underline{Q}_P + S_N * (\underline{Q}_R - \underline{Q}_P) * Z \qquad (4)$$



The enherited stepsize S_N is treated as described above. Z is a randomly chosen number from a Gaussian distributed set of numbers. The worst 20 populations are treated with respect to the original coordinate system.

Fig. 11: Sorted set of populations. Connections between P and R_i define the coordinate system for population P which is changed a little by each generation.

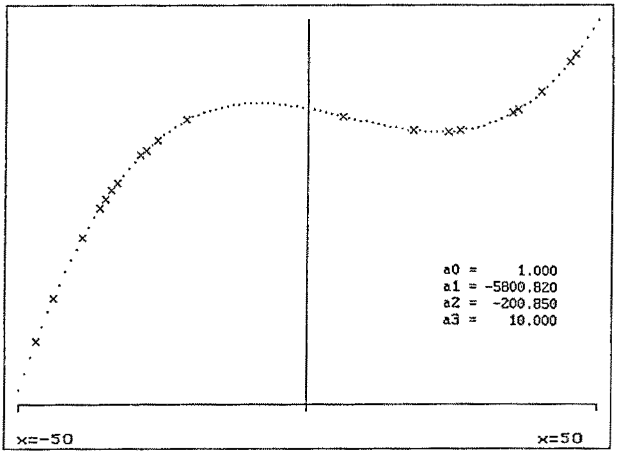
The selection procedure has been designed in accordance with selection rules in football for example. The set of 100 populations is devided into leagues of twenty members. Sorting now takes place only within one league and after each generation there is a competition between the worst of the first league and the leader of the second league possibly exchanging them. With a certain probability the worst of a league is eliminated. Then the following populations close up and at the very last position a completely new object is created. Thus this strategy can be written according to formula 5 where terms concerning the population level depend on certain conditions expressed by x.

$$[\, 95_{+x} + 5_{-x}(1, 10)^1 \,] - ES \qquad (6)$$

Adapting transformation of the mutation sphere

That far we have the description of the multi population strategy. Now the question occurs whether this system actually performs a transformation of the mutation sphere which adapts to the particular shape of the quality function. Some experimental evidence can be given from investigations on a simple curve fitting task. The 20 points to be fit by a third degree polynomial expression are calculated from equation 6 and plotted as crosses in figure 12.

$$y = a_0 + a_1 * x^1 + a_2 * x^2 + a_3 * x^3; \qquad (6)$$



Pretending the coefficients were unknown we have a 4-dimensional minimization problem with variables a_i. The quality function is given in (7) where ys is the fixed coordinate of a point j and y the value calculated by equation 6 with respect to a particular object vector $\underline{a}$.

$$Q = \sum_{j=1}^{P} (y_j - ys_j)^2 \qquad (7)$$

Fig. 12: Solution for a third degree polynomial fitting 20 points.

Application of a (1,30)-ES does not solve this problem within reasonable time (see curve A in figure 13). A short examination of equations 6 and 7 shows that in first approximation the direction of highest gradient is identical with one of the coordinates. The vektor $\underline{c}$ of appropriate variances can be calculated from information about the 20 points and the structure of equation 6.

$$\underline{c}_{calc.} = (1.0 \ , \ 0.0308 \ , \ 0.00086 \ , \ 0.00023) \qquad (8)$$

If this transformation is applied to $\underline{Z}$ in equation 1 convergence velocity is accellerated drastically in a single population experiment (curve B in figure 13). Convergence velocity of an adapting system which does not use explicit information about the quality function is expected between those of A and B. Curve C in figure 13 is the result of the multi population experiment. In this particular case we can calculate the transformation which is hidden in the 4-dimensional directions between the first population and the 10 following ones. Each connecting vector is normalized and the components converted into their absolute values. The mean value of all 10 vectors gives us the adapted version of the adapted variance vector:

$$\underline{c}_{adapt.} = (0.99922 \ , 0.03941 , 0.00147 \ , \ 0.000038) \qquad (9)$$

This result is rather similar to that in equation 8 and we conclude that the system in fact performs the required adapting coordinate transformation due to the particular quality function.

There remains the question whether the multi population system copes also with the covariances if the coordinate system describing the quality function is rotated. First evidence to this question is given by an experiment plotted as curve D in figure 13. Here the quality function has been rotated by 20° in the 0-1-plane, by 40° in the 1-2-plane and by 60° in the 1-3-plane. Convergence velocity does not depend on different rotations of the coordinate system.

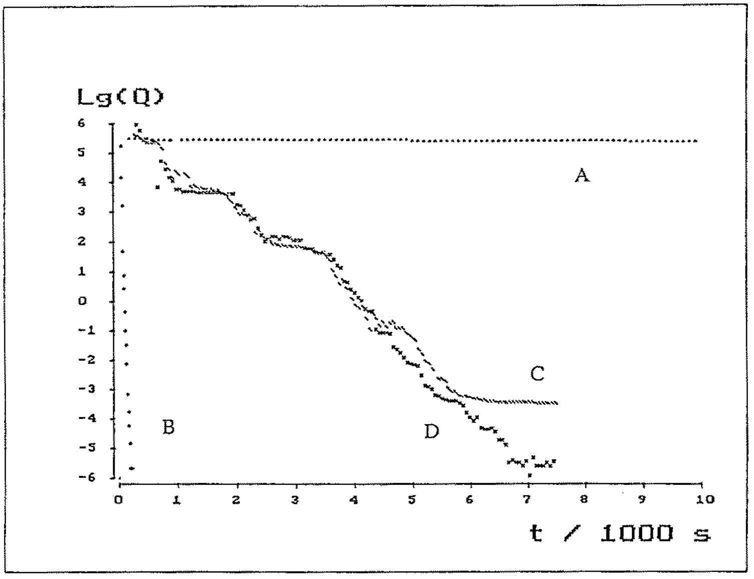


Figure 13: Quality versus time for different experiments; explanations in the text.

Conclusion

It seems that the multi population evolution strategy is a promising approach to an important class of optimization problems in which convergence velocities can be improved drastically. Despite some suggesting experimental results more theoretical evidence is still necessary. The system does not work sufficiently on "easy" problems like those reported in the first part of the paper. Therefore adaptive mechanisms have to be developed which control the type of evolution strategy most sensible to the type and complexity of the application problem.

Acknowledgement: I like to thank Mr. Herdy who contributed to this paper with important ideas and advice.

Literature

[1] Rechenberg, I. (1973): Evolutionsstrategie - Optimierung technischer Systeme nach Prinzipien der biologischen Evolution. Stuttgart: Frommann-Holzboog.

[2] Schwefel, H.-P. (1977): Numerische Optimierung von Computer-Modellen mittels der Evolutionsstrategie. Basel und Stuttgart: Birkhäuser.

[3] Herdy, M. (1990): Application of the Evolutionsstrategie to Discrete Optimization Problems. In this volume.

[4] Rechenberg, I. (1964): Cybernetic solution path of an experimental problem. Royal Aircraft Establishment, Library Translation 1122, Farnborough.

[5] Rechenberg, I. (1989): Evolutionsstrategie - Optimierung nach Prinzipien der biologischen Evolution. In: Albertz, J. (Hrsg.): Evolution und Evolutionsstrategien in Biologie, Technik und Gesellschaft. Band 9 der Schriftenreihe FREIE AKADEMIE. Hofheim.

[6] Schwefel, H.-P. (1989): "Natürliche Intelligenz" in evolutionären Systemen. In: Albertz, J. (Hrsg.): Evolution und Evolutionsstrategien in Biologie, Technik und Gesellschaft. Band 9 der Schriftenreihe FREIE AKADEMIE. Hofheim.

[7] Rechenberg, I. (1984): The evolution strategy - a mathematical model of Darwinian Evolution. In: Synergetics - from microscopic to macroscopic order. Springer series in synergetics, 22, Springer, Berlin-Heidelberg.

[8] Eggert, H. (1989): Gewichts- und Durchbiegungsminimierung von Fachwerkträgern durch Anwendung der Evolutionsstrategie. Diplomarbeit am Fachgebiet Bionik und Evolutionstechnik, TU Berlin, WS 1989/90.

[9] Schwefel, H.-P. (1981): Numerical Optimization of Computer Models. Chichester: John Wiley & Sons.