MUTATION CONTROL AND CONVERGENCE IN EVOLUTIONARY MULTI-OBJECTIVE OPTIMIZATION

Marco Laumanns[1] Günter Rudolph[2]
Hans-Paul Schwefel[2]

[1] ETH Zürich, Institut TIK, CH–8092 Zürich, Switzerland
laumanns@tik.ee.ethz.ch
[2] Universität Dortmund, Fachbereich Informatik XI, D–44221 Dortmund, Germany
{rudolph,schwefel}@ls11.cs.uni-dortmund.de

Abstract

This paper addresses the problem of controlling mutation strength in multi-objective evolutionary algorithms and its implications for the convergence to the Pareto set. Adaptive parameter control is one major issue in the field of evolutionary computation, and several methods have been proposed and applied successfully for single objective optimization problems. In this study we examine whether these results carry over to the multi-objective case and what modifications must be taken to meet the difficulties and pitfalls of conflicting objectives.

1 Introduction

The application of evolutionary algorithms to optimization problems with multiple objectives has become very popular in recent years. In order to find or to approximate the set of efficient or Pareto-optimal solutions an algorithm should ensure convergence to Pareto set and provide a 'good' distribution of solutions. This helps the decision maker to get insight into the real trade-offs between conflicting objectives.

The conceptual approaches to these tasks are mainly concerned with the evaluation of solutions for the selection operator. Fitness assignment methods based on the notion of dominance turned out to be superior to plain aggregating methods as far as the distribution of solutions is concerned. Further improvements include density based selection methods to maintain diversity and the use of elitism to speed up the search in the direction of the Pareto set and to ensure convergence properties [LZT01, RA00].

However, research focusing on the role of the variation operators in evolutionary multi-objective optimization remained rare. Virtually all implementations use standard non-adaptive operators from the single objective case. Especially the mutation operator has received little attention so far, and it is common practice to apply fixed mutation rates in binary coded representation or fixed mutation step sizes in Evolution Strategies. Theoretical considerations, however, emphasize the importance of the mutation strength for the convergence of multi-objective evolutionary algorithms (MOEAs) [Rud98, Han99]. When large search spaces are to be explored, adaptive variation operators are mandatory to achieve both a satisfactory rate of progress towards the optimum and a high precision of solutions. Thus, effective control mechanisms, which exist for the single objective case, need to be developed.

This study begins by recalling the concepts and different approaches of mutation strength control in the next section. From this, we incorporate adaptive mutation control in two different classes of MOEA: In section 4 we first use a simple Evolution Strategy with Pareto-based selection before turning to a Predator-Prey EA in section 5.

2 Concepts and Methods of Mutation Control

Choosing optimal parameter settings is considered as an important - and mainly open - 'meta-problem' of EA research. Practitioners often use rules of thumb, which are a combination of experience and extrapolation of theoretical results. One of this parameterization problems concerns the mutation strength: With too strong variation the evolution becomes a pure random search, with too weak variation no real progress can be achieved. The small region of appropriate mutation strength - sometimes referred to as the 'evolution window' - depends on the topology of the objective function which usually is unknown a priori. Thus, adaptation mechanisms for the mutation strength are a necessity for many optimization problems.

This is equally important in the case of multiple objectives as can be drawn from theoretic results. Stochastic convergence of a simple (1+1)-EA to the Pareto set of a two-objective problem has been shown in [Rud98]. However, this could only be guaranteed if the mutation step size was chose proportionally to the distance to the Pareto set. Existing approaches to mutation control can be categorized into three groups of increasing complexity:

- Predefined schedules without feedback,
- Feedback-based adaptation with (explicit) external control, and
- Self-adaptive (internal) mechanisms without external control.

Representatives of the first class are for instance time dependent schedules as applied in the Predator-Prey EA of [LRS98] in a multi-objective environment. Here, the mutation step sizes are discounted by a constant factor each time an offspring in produced. Convergence to the global Pareto set of some simple multi-objective problems can be achieved, if the initial step sizes are big enough and a 'conservative' discount factor is chosen. The disadvantages of this approach is the lack of a possibility to increase step sizes that are too small and the need for new parameters to be set.

A very sophisticated control mechanism is proposed by [Kah91]. In the adaptation rule the author distinguishes nine different cases depending on the history of the evolution. In each case the step sizes are altered differently according heuristics, some of which resemble the so-called 1/5-success rule. This rule, originally developed for a (1+1)-EA, is based on the assumption that maximum progress can be achieved through step sizes leading to a success probability of approximately 20 %. In the multi-objective case the measuring of success need to be revised, since the probability of making cooperative steps decreases rapidly near the Pareto set, even for small step sizes [Han01]. A possible remedy would be to exempt those mutations leading to incomparable solutions from the success measure, which should keep the success probability constant. However, the adaptation frequency decreases as the probability of creating incomparable solutions (no success nor failure) increases.

Finally, the self-adaptive mutative step size control of Evolution Strategies has been applied to multi-objective optimization problems by [Kur91]. In his algorithm the selection criterion changes randomly over time. Individuals are supplied with a set of step sizes for each objective function to cope with the changing environment. This additional genetic material is organized through polyploidy.

3 Problem Description and Representation of Individuals

When evolutionary algorithms are used for approximating the Pareto set of a multi-objective optimization problem (MOP), they typically face two major problems:

- convergence to the Pareto set and
- distribution of solutions.

The latter feature is created by the existence of multiple Pareto-optimal solutions in the multi-objective case, while the first point is an as important matter in single objective optimization. The convergence property can further be specified by convergence velocity and convergence reliability. For the multi-objective case this distinction should be made as well. To some extend these 'meta-problems' may also be conflicting - at least their dependencies are far from being clear - which makes the design and improvement of algorithms a complicated task. Quantitative comparisons are always based on a (performance) measure, which needs to be defined carefully. Furthermore, performance measures related to the above objectives might be incommensurable as well.

This study focuses on the problem of convergence to the Pareto set. Here, the control of the mutations strength might be very important. For a first assessment of different mutation control mechanisms two types of functions are whose components are well studied in the context of mutation control in single objective optimization: the sphere model and a multi-modal function.

$$F(x) = \begin{pmatrix} (x-c_1)^2 \\ \vdots \\ (x-c_m)^2 \end{pmatrix}, \qquad G(x) = \begin{pmatrix} \sum_{i=1}^{n-1}(-10e^{-0.2\sqrt{x_i^2+x_{i+1}^2}}) \\ \sum_{i=1}^{n}(|x_i|^{0.8}+\sin^3(x_i)) \end{pmatrix}$$

The sphere model is a unimodal function which has served as a reference in many theoretic studies of evolutionary algorithms and especially in the context of self-adaptive Evolution Strategies. This function can systematically be

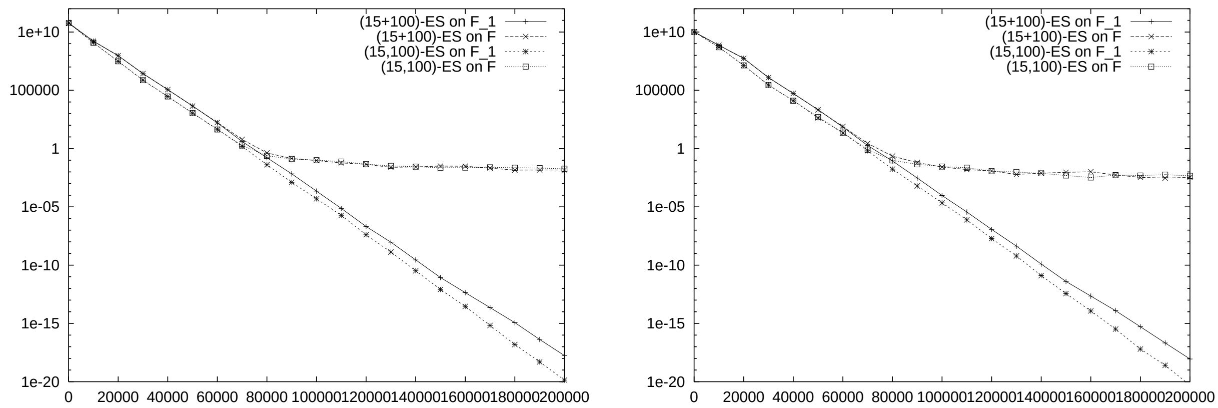

Figure 1: Median of the individuals' distances (left) and expected step length (right) over time (number of function evaluations) on F, $n = 100$, $\rho = 2$.

extended to a 'multi-sphere model' by translating the optimum for each component so that the component optima span an m-dimensional space. Here we restrict ourselves to $m = 2$ objectives with $c_1 = (1, 0, 0, \ldots, 0)$ and $c_2 = (0, 1, 0, \ldots, 0)$. This function is perhaps the most simple multi-objective test function for infinite search spaces, because it is continuous, non multi-frontal, and the Pareto set can be determined analytically: It is the m-dimensional polyeder of the component optima. However, since the search space in unbounded, convergence depends crucially on the starting point and the mutation strength.

The second function under examination has been proposed by [Kur91] and used as a test function of an evolution strategy for vector optimization. It is part of van Veldhuizen's test function suite [Vel99] for MOEAs and features a disconnected Pareto sets. Both functions are defined on $\mathbb{R}^n$ advocating the use a floating point representation for the decision variables. The individual space is defined as

$$I = \mathbb{R}^n \times {\mathbb{R}^+}^{n_\sigma}$$

Thus, an individual consists of n decision variables and n_σ mutation step sizes (strategy variables). Since the component functions are symmetric, we use a single mutation step size σ only ($n_\sigma = 1$). Mutation is carried out by first multiplying the step size with an instantiation of a log-normally distributed random variable:

$$\sigma^{(t+1)} = \sigma^{(t)} \cdot \exp \mathcal{N}(0, \tau_0), \quad \tau_0 = \frac{1}{\sqrt{n}} \tag{1}$$

Thereafter, a normal-distributed random vector with zero mean and variance σ^2 is added to the individual.

4 Standard EAs with Dominance-based Selection

The first algorithm under consideration is a standard $(\mu, \kappa, \lambda, \rho)$ Evolution Strategy (ES) where μ denotes the number of possible parents, λ the number of offspring individuals, κ the maximum lifespan of individuals (measured in generations) and ρ the number of recombination partners. Fitness is assigned based on the dominance relation concerning all objective function values. Specifically, we used the individuals' dominance level within the corresponding population.[1]

4.1 Behavior on the multi-sphere model (F)

It is known from the single-objective case that standard self-adaptive evolution strategies exhibit linear order convergence on the sphere model (here: first component of F). For the multi-sphere model, however, this property seems to be valid only individuals which are 'sufficiently far away' from the Pareto-optimal set. After a period of exponentially decreasing population distance to the Pareto set, the solutions suddenly start to oscillate around a

[1] Alternatively, the dominance grade (or rank) or a combination of both could be used (to increase the discriminating power), but in our case the results are not significantly different. Usually, some additional features are included that 'punish' individual which are close to each other according to some metric to maintain diversity. In this study we concentrate on the convergence and try to avoid any side-effects caused by sophisticated fitness assignment methods.

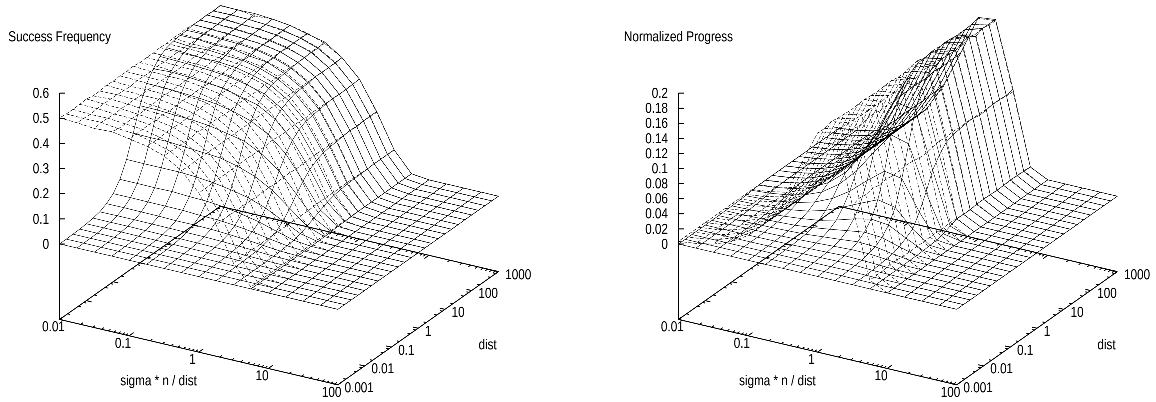

Figure 2: Estimated success probabilities (left) and normalized progress (right) for mutations (step size σ) of individuals with distance *dist* to the optimum on the SOP F_1 (upper surfaces) and the MOP F (lower surfaces), n=100.

small but fixed final distance (Fig. 1). An explanation can be deduced from Fig. 2, which gives a visual impression of the 'evolution window': For k randomly initialized individuals x_i and their mutants $mut(x_i)$, the plots show the success frequency

$$w_s = \frac{1}{k} \sum_{i=1}^{k} 1(mut(x_i) \prec x_i) \tag{2}$$

and the average (normalized) progress for successful mutations

$$\phi' = w_s \cdot n \cdot \sum_{i=1}^{k} \frac{d(x_i, X^*) - d(mut(x_i), X^*)}{d(x_i, X^*)} \tag{3}$$

over the (normalized) step size

$$\sigma' = n \cdot \frac{\sigma}{d_0} \tag{4}$$

for mutations of randomly chosen individuals with a fixed expected distance d_0 to the Pareto set. $d(x, X)$ denotes the minimum Euclidean distance of vector x to all elements of set X and X^* is the Pareto set of the given problem. It is obvious that near the Pareto set not only the success rate decreases, but also the normalized average progress for successful mutations. Thus, it is increasingly difficult to proceed closer to the Pareto set, if successes are judged via the dominance relation.

As far as the commonly accepted conditions for this kind of mutative step-size control are concerned, we found that in our case the algorithm is much less sensitive to the μ-λ ratio than expected.

4.2 Multi-modal objectives (G)

On the multi-modal function G it gets more difficult to increase the step sizes sufficiently with increasing n. At least if initial step sizes are chosen that are larger than the distance of different local optimal sets, an increase can be observed for large n, too. With too small step sizes, optimization is only performed locally, the whole search space besides the basin of attraction of the next local optimal set will not be 'visible' for the algorithm. A more annoying fact is that in spite of the raised step sizes no real progress can be observed. Instead, the step sizes start to decrease again just when they seem to be near optimal. This is supported by experiments with initial step sizes greater than the decision variables. Thus, starting with large step sizes does not help at all to overcome this algorithmic limitation.

Fig. 3 can help to explain this situation: For $n = 2$ the graphs for F and G look quite similar, only the maximal progress on F is considerably greater than on G. For both functions the region of step sizes with positive progress, lies around the order of magnitude of the distance to the Pareto set. This matches perfectly with previously derived theoretical results [Rud98]. With increasing n this difference get more apparent. For $n = 100$ the maximum progress for G is less than half of the corresponding value for F. But even more striking is the region on negative(!) average progress for successful mutations for step sizes slightly greater than the optimum. This imposes additional pressure on the step sizes to decrease besides the success probability. All in all, the combination of greater success

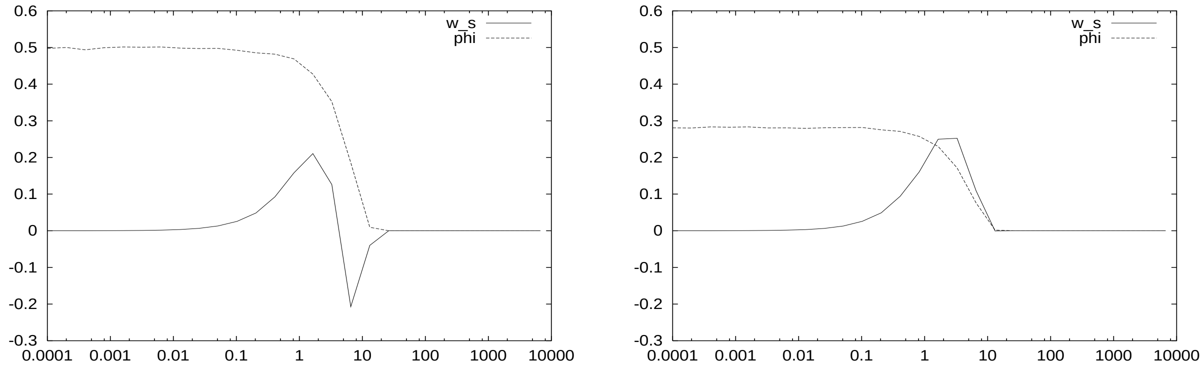

Figure 3: Success frequency (w_s) and average normalized progress (ϕ') in case of success for mutations with different σ values (normalized), G, $n = 100$. Average distance to the Pareto set 10^{10} (left) and 10 (right).

probabilities for smaller step sizes, the small average progress in case of optimal step sizes and the negative progress for larger step sizes seem to hinder an adequate preservation of step sizes around the optimal value. Of course, this assumption in only based on numerical experiment and lacks a theoretical validation still.

5 An Evolutionary Predator-Prey Model

The predator-prey model of [LRS98] consists of a spatially distributed population. Predator individuals move across the spatial structure according to a random walk and perform selection: They delete the worst prey individual of their neighborhood according to their associated objective function. This is a population-based single criterion selection in contrast to the Pareto-based selection applied in the Pareto ES of the previous section. Empty spaces are refilled by the best of λ individuals created by recombination (in case of $\rho \geq 2$ or duplication ($\rho = 1$) of ρ randomly chosen individuals of a given neighborhood.

In the original study two different step size rules were applied: fixed step sizes of different magnitude and monotonously decreasing step sizes according to a given schedule (see section 2). Now, we extend the model by presenting two additional adaptation rules that are capable of increasing the step sizes if necessary.

5.1 Standard mutative self-adaptation

Though the predator-prey model can be regarded as a special $(\mu+1)$-EA, we try the standard mutative self-adaptation from the previous section first. Here, two necessary conditions for the applicability of this kind of step size control are not fulfilled: the existence of a 'comma' strategy (which ensures the deletion of bad adapted individuals) and the production of an offspring surplus to cope with the optimal success probability of less than $1/2$.

The results on the multi-sphere model are surprisingly good. Compared to a (100, ∞, 100, 1)-ES progress is even better in respect to the number of function evaluations. In addition, the mere stagnation of the Pareto ES during the third phase of the run can be avoided, and convergence is still visible near the Pareto set. This may be explained by the superior success probability of single criterion selection in the vicinity of the Pareto set, where is gets increasingly difficult to make cooperative steps from one mutation alone. On G, however, the results are as bad as before using the Pareto ES. We conclude that the same mechanisms come into play than cause both the Pareto ES and the predator-prey model difficulty to keep the step sizes to a sufficiently high value. In this algorithm, recombination can help as well, but an appropriate selective pressure on the step sizes must be guaranteed through higher numbers of offspring.

5.2 Self-adaptation through recombination frequency

We now propose another method of self-adapting step sizes which combines the implicit evaluation of good 'inner models' through fitness evaluation and selection with a fixed but flexible schedule depending on the recombination frequency. With this schedule all individuals use bigger step sizes for the early offspring and smaller for later ones. Thus there will hopefully be a pressure on the step sizes to increase in all stages of the evolution. The mutation of the step sizes is done deterministically according to the rule

$$\sigma^{(t+1)} = \gamma^{d-d_0} \cdot \sigma^{(t)}, \quad \gamma \in]0,1[\tag{5}$$

where d denotes the number of descendants an individual has produced so far. The 'delay' parameter d_0 determines the number of descendants that must be created before smaller step sizes are passed on to the offspring. With d_0 the intensity of testing bigger step sizes can be controlled, whereas γ can be interpreted as an 'adaptation rate'. This schedule is supposed to be very robust concerning high d_0 values: For step sizes just one order of magnitude greater than the optimal ones the success rate rapidly goes to zero. Hence, it should only be a matter of time until appropriate step sizes will be found.

6 Conclusions

In this study we examined the problem of controlling mutation strength in multi-objective evolutionary algorithms. The results show that the standard self-adaptive evolution strategies have difficulties to converge to the Pareto set due to the low success probability of Pareto based selection. Alternative selection methods like the predator-prey approach discussed here are a possibility to overcome this limitation. A disadvantage of this method, however, is the lower rate of convergence due to the less efficient single-criterion selection. It may be speculated whether a hybrid method using Pareto-based selection for a fast but rough localization of the Pareto set and single-criterion selection for an accurate approximation would be appropriate.

Another possible way to improve the convergence properties of self-adaptive MOEAs might be a combination of low selection pressure (to diminish the danger of premature convergence) and elitism (to prevent possible divergence caused by increasing mutation step sizes and low selection pressure). Also the topic of diversity needs a thorough investigation in connection with convergence in self-adaptive MOEAs. Finally, it should be emphasized that these results are entirely based on numerical experiments and empirical evaluation. Theoretical analysis could fundamentally improve the understanding of the effects of these heuristic methods.

References

[Han99] Thomas Hanne. On the convergence of multiobjective evolutionary algorithms. *European Journal Of Operational Research*, 117(3):553–564, 1999.

[Han01] Thomas Hanne. Global multiobjective optimization with evolutionary algorithm s: Selection mechanisms and mutation control. In E. Zitzler et al., editor, *Evolutionary Multi-criterion Optimization (EMO 2001), Proc.*, Lecture Notes in Computer Science Vol.1993, pages 197–212, Berlin, 2001. Springer-Verlag.

[Kah91] J. Kahlert. *Vektorielle Optimierung mit Evolutionsstrategien und Anwendungen in der Regelungstechnik*. VDI Verlag, Duesseldorf, 1991.

[Kur91] Frank Kursawe. A variant of evolution strategies for vector optimization. In H.-P. Schwefel and R. Männer, editors, *Parallel Problem Solving from Nature*, pages 193–197, Berlin, 1991. Springer.

[LRS98] Marco Laumanns, Günter Rudolph, and Hans-Paul Schwefel. A spatial predator-prey approach to multi-objective optimization: A preliminary study. In Agoston E. Eiben et al., editor, *Parallel Problem Solving from Nature (PPSN-V)*, pages 241–249, Berlin, 1998. Springer.

[LZT01] Marco Laumanns, Eckart Zitzler, and Lothar Thiele. On the effects of archiving, elitism, and density based selection in evolutionary multi-objective optimization. In E. Zitzler et al., editor, *Evolutionary Multi-criterion Optimization (EMO 2001), Proc.*, Lecture Notes in Computer Science Vol.1993, pages 181–196, Berlin, 2001. Springer-Verlag.

[RA00] G. Rudolph and A. Agapie. Convergence properties of some multi-objective evolutionary algorithms. In *Congress on Evolutionary Computation (CEC 2000)*, volume 2, pages 1010–1016, Piscataway, NJ, 2000. IEEE Press.

[Rud98] Günter Rudolph. On a multi-objective evolutionary algorithm and its convergence to the pareto set. In *IEEE Int'l Conf. on Evolutionary Computation (ICEC'98)*, pages 511–516, Piscataway, 1998. IEEE Press.

[Vel99] D. A. Van Veldhuizen. *Multiobjective Evolutionary Algorithms: Classifications, Analyses, and New Innovations*. PhD thesis, Graduate School of Engineering of the Air Force Institute of Technology, Air University, June 1999.