
Optimum Tracking with Evolution Strategies

Dirk V. Arnold

dirk@cs.dal.ca

Faculty of Computer Science, Dalhousie University, Halifax, Nova Scotia, Canada
B3H 1W5

Hans-Georg Beyer

hans-georg.beyer@fh-vorarlberg.ac.at

Department of Computer Science, Vorarlberg University of Applied Sciences, Achstr. 1,
A-6850 Dornbirn, Austria

Abstract

Evolutionary algorithms are frequently applied to dynamic optimization problems in which the objective varies with time. It is desirable to gain an improved understanding of the influence of different genetic operators and of the parameters of a strategy on its tracking performance. An approach that has proven useful in the past is to mathematically analyze the strategy's behavior in simple, idealized environments. The present paper investigates the performance of a multiparent evolution strategy that employs cumulative step length adaptation for an optimization task in which the target moves linearly with uniform speed. Scaling laws that quite accurately describe the behavior of the strategy and that greatly contribute to its understanding are derived. It is shown that in contrast to previously obtained results for a randomly moving target, cumulative step length adaptation fails to achieve optimal step lengths if the target moves in a linear fashion. Implications for the choice of population size parameters are discussed.

Keywords

Genetic and evolutionary computation, evolution strategies, cumulative step length adaptation, tracking problem, dynamic optimization.

1 Introduction

Evolutionary algorithms are nature inspired heuristics for search and optimization that model the iterated interplay of variation and selection in a population of candidate solutions. Comprehensive treatments of the subject area can be found in (Bäck, 1996; Goldberg, 1989; Mitchell, 1996; Rechenberg, 1994). Reasons for the widespread use of evolutionary algorithms include their often observed robustness, the ease with which the underlying paradigm is understood and implemented, and their wide applicability. Areas of application today include management, control, design, scheduling, pattern recognition, and decision making.

In recent years, a fair amount of theoretical investigation has contributed substantially to our understanding of the dynamics of evolutionary search strategies on a variety of problem classes. For an overview, see (Beyer et al., 2002). However, most of the problem classes that have been considered are of a static nature. In contrast, many problems encountered in the computational, engineering, and biological sciences are dynamic in that the objective is not constant but varies with time. Instances of dynamic optimization problems arise for example in the context of online job scheduling, where new jobs arrive in the course of the optimization. In the engineering sciences, many control problems are of an inherently dynamic nature. An extensive list of references

concerned with dynamic optimization problems in the biological sciences has been reviewed by (Bürger, 2000).

In contrast to static optimization where the goal is to rapidly and accurately locate a (near) optimal solution, the task in dynamic optimization frequently amounts to tracking a moving target as closely as possible. Strategies for dynamic optimization need to continually adapt to changes in the environment. Evolutionary algorithms are population-based strategies that often place great emphasis on adaptability, and that are believed to be good at handling time-varying objectives. In fact, dynamic optimization is frequently cited as a prime application area for evolutionary algorithms.

The knowledge available with respect to the capabilities of evolutionary algorithms for dynamic problems is mostly of an empirical nature. An extensive survey of work concerned with evolutionary optimization in dynamic environments along with a collection of benchmark functions and a discussion of methods that have been proposed to improve the performance of evolutionary algorithms when the objective varies with time has been compiled by (Branke, 2001). More recent treatises of the area include a book by (Morrison, 2004) and a survey article concerned with evolutionary optimization in uncertain environments by (Jin and Branke, 2005). A first rigorous analysis of the performance of a $(1 + 1)$ -strategy on a discrete, dynamic objective function has been presented by (Droste, 2002). However, focus in that paper is not on the tracking behavior of the strategy but rather on the expected time required to first reach the optimum. The tracking behavior of a $(1 + \lambda)$ -strategy is subject of a recent paper by (Jansen and Schellbach, 2005) in which a one-max like dynamic problem is considered on a two-dimensional lattice. As in Droste's paper, the search space is of a discrete nature and the strategy is not adaptive.

Focus in the present paper is on the tracking behavior of adaptive evolution strategies in continuous search spaces. Among the most closely related work is a paper by (Angeline, 1997) that compares empirically the tracking performance of an evolutionary algorithm employing a form of mutative self-adaptation with that of a strategy using a simple heuristic for step length adaptation. The fitness environment considered is a three-dimensional, spherically symmetric objective function that is shifted periodically either in a random fashion or on a linear or a spherical path. Angeline observes that the self-adaptation mechanism is not without problems in the dynamic case. In that same fitness environment, (Bäck, 1998) compares different variants of mutative self-adaptation and presents evidence that seems to indicate that the lognormal self-adaptation used in evolution strategies performs better than the variant of self-adaptation commonly used in evolutionary programming. (Salomon and Eggenberger, 1997) compare empirically the performance of evolution strategies with that of a breeder genetic algorithm on the sphere, an ellipsoid, and Rastrigin's function, where the coordinates are shifted by a constant increment in every time step. The search space dimensionalities they consider for the sphere are $N = 10$ and $N = 30$. While they do not quantify the term, they find that the sensitivity to the particular implementation of the strategy and to its parameter values is much lower for the tracking task than it is in a static environment. Without providing details, they also report to have observed that recombination is not beneficial for tracking a moving target. (Weicker and Weicker, 1999) contrast self-adaptation of a single mutation strength with that of N mutation strengths and adaptation of the full mutation covariance matrix and find that in more rapidly changing environments, the adaptation of more than a single mutation strength becomes unreliable. Finally, (Weicker, 2006) has also empirically investigated the optimum tracking behavior of evolution strategies in an environment very similar to the

one considered here. The strategies under consideration employ mutative self adaptation and different selection mechanisms, but they do not make use of recombination. Based on computational experiments conducted for a large number of parameter settings, recommendations with regard to the number of offspring generated per time step and the type of selection are made.

Altogether, the above references provide a host of empirical knowledge with respect to evolutionary optimization in continuous, dynamic environments. In an attempt to establish some theoretical results to complement and explain that empirical knowledge, in (Arnold and Beyer, 2002) the performance of a multiparent evolution strategy with cumulative step length adaptation has been analyzed for a dynamic optimization problem in which the target moves in a random fashion. The tools employed in that work have also been used in an analysis of the behavior of the strategy on a static optimization problem disturbed by noise (Arnold, 2002; Arnold and Beyer, 2004). It was found that for the case of a randomly moving target, cumulative step length adaptation is able to achieve asymptotically optimal step lengths. It was noted that the case of a target that moves linearly rather than randomly is an interesting one to consider as the linear motion introduces positive correlations in the sequence of steps to be taken. As cumulative step length adaptation relies on the conjecture that consecutive steps should be uncorrelated, it appears reasonable to expect suboptimal performance of cumulative step length adaptation for the case of the linearly moving target.

The present article presents an analysis of the tracking performance of the $(\mu/\mu, \lambda)$ -ES with cumulative step length adaptation for the case of a linearly moving target. The analysis is more difficult than that of the tracking problem with random motion of the target as the linear motion of the target introduces an additional variable in the calculations. The fitness environment closely resembles that considered in (Angeline, 1997; Bäck, 1998; Salomon and Eggenberger, 1997; Weicker, 2006). However, in contrast to those references, no particular values need to be assumed for the size of the population, the speed of the target, or the search space dimensionality. Instead, scaling laws that describe the dependence of the performance of the strategy on those parameters are obtained. Such scaling laws provide a quantitative understanding of the dynamics of the evolutionary processes that makes it possible to predict optimal parameter settings for the problem class considered. We share the belief of (van Nimwegen and Crutchfield, 2001) that a general, predictive theory of the dynamics of evolutionary search can be built incrementally, starting with a quantitative analytical understanding of specific problems and then generalizing to more complex situations. The present article can be considered as a small step toward such a theory.

The remainder of this article is organized as follows. In Section 2, the $(\mu/\mu, \lambda)$ -ES with isotropic mutations and cumulative step length adaptation is briefly described. In Section 3, the tracking problem is defined and previously obtained knowledge useful for the analyses that follow is summarized. In Section 4, the behavior of the strategy is analyzed for fixed step length. The optimal step length is determined as a function of the speed of the target, the dimensionality of the search space, and the size of the population. In Section 5, the performance of the step length adaptation scheme is investigated. The step length realized by the evolution strategy is compared with the optimal step length derived in Section 4. Finally, Section 6 concludes with a brief summary, a discussion of the findings and the insights gained, and suggestions for future work.

2 The $(\mu/\mu, \lambda)$ -ES

The $(\mu/\mu, \lambda)$ -ES is a strategy for the optimization of real-valued functions $f : \mathbb{R}^N \rightarrow \mathbb{R}$ that is popular both due to its proven good performance (in static settings) and its relative mathematical tractability. For a thorough introduction to evolution strategies, see (Bäck, 1996; Rechenberg, 1994; Schwefel, 1995). An explanation of the $(\mu/\rho \nmid \lambda)$ -symbolism (of which $(\mu/\mu, \lambda)$ is an instantiation) can be found in (Beyer, 2001). The double appearance of the parameter μ indicates that recombination is global, i.e. that all parents participate in the creation of every single offspring candidate solution.

The $(\mu/\mu, \lambda)$ -ES can be considered as repeatedly updating a search point $\mathbf{x} \in \mathbb{R}^N$ (that is the centroid of the population of candidate solutions that survived the most recent round of selection) using the following four steps:

1. Generate λ offspring candidate solutions $\mathbf{y}_i = \mathbf{x} + \sigma \mathbf{z}_i$, $i = 1, \dots, \lambda$. The \mathbf{z}_i are vectors consisting of N independent, standard normally distributed components and are referred to as mutation vectors. The nonnegative quantity σ is referred to as the mutation strength and determines the step length of the strategy.
2. Determine the objective function values $f(\mathbf{y}_i)$ of the offspring candidate solutions and order the \mathbf{y}_i according to those values. After ordering, index k ; λ refers to the k th best of the λ offspring (the k th smallest for minimization; the k th largest for maximization).
3. Compute the arithmetic mean

$$\langle \mathbf{z} \rangle = \frac{1}{\mu} \sum_{k=1}^{\mu} \mathbf{z}_{k;\lambda} \quad (1)$$

of those mutation vectors that correspond to the μ best of the offspring. Vector $\langle \mathbf{z} \rangle$ is referred to as the progress vector.

4. Replace the search point by letting

$$\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} + \sigma \langle \mathbf{z} \rangle^{(t)}, \quad (2)$$

where superscripts indicate time.

Note that while generally, initialization schemes and termination criteria are important components of the algorithm, they are frequently application dependent and are irrelevant in the present context. Rather than considering them here, we refer to (Bäck, 1996) for a discussion. Also note that the distribution of mutation vectors as described above is isotropic. The restriction to isotropic mutations has been made in order to keep the analysis of the strategy's behavior tractable. Practical implementations often rely on mutation vectors that are drawn from a normal distribution with general covariance matrix \mathbf{C} rather than the unity matrix. A mechanism for the adaptation of \mathbf{C} has been proposed by (Hansen, 1998; Hansen and Ostermeier, 2001).

In real-valued search spaces, it is often necessary for the mutation strength to be adapted continuously to the local characteristics of the objective function. Information from the present and possibly from past time steps needs to be used in order to learn appropriate step sizes. A mechanism that is commonly employed for the adaptation of the mutation strength is the cumulative step length adaptation algorithm of (Ostermeier et al., 1994). It relies on the conjecture that if the mutation strength is below its

optimal value, then consecutive steps of the strategy tend to be parallel. Conversely, if the mutation strength is too high, then consecutive steps tend to be antiparallel. For optimally adapted mutation strength, the steps taken by the evolution strategy are uncorrelated. This is plausible intuitively as several steps in the same direction in search space are ideally replaced by a single longer step in that direction. Conversely, consecutive steps that nullify each other are a sign that the step length is too high. So as to be able to reliably detect parallel or antiparallel correlations between successive steps, information from a number of time steps needs to be accumulated. For the $(\mu/\mu, \lambda)$ -ES, the accumulated progress vector \mathbf{s} is defined by $\mathbf{s}^{(0)} = \mathbf{0}$ and the recursive relationship

$$\mathbf{s}^{(t+1)} = (1 - c)\mathbf{s}^{(t)} + \sqrt{\mu c(2 - c)}\langle \mathbf{z} \rangle^{(t)}, \quad (3)$$

where c is a constant determining how far back the “memory” of the accumulation process reaches. The mutation strength is updated according to

$$\sigma^{(t+1)} = \sigma^{(t)} \exp\left(\frac{\|\mathbf{s}^{(t+1)}\|^2 - N}{2DN}\right), \quad (4)$$

where D denotes a damping constant. The constants c and D are set to $1/\sqrt{N}$ and \sqrt{N} , respectively, according to recommendations made by (Hansen, 1998). From the way that mutation vectors are generated along with the choice of coefficients in Eq. (3) it can be inferred that the expected squared length of the accumulated progress vector equals the search space dimensionality N if selection is random (i.e., if the mean in Eq. (1) is computed over μ randomly chosen mutation vectors). Positive correlations in the sequence of steps lead to the squared length of the accumulated progress vector exceeding N ; negative correlations result in shorter accumulated progress vectors. The term N in the numerator of the argument to the exponential function in Eq. (4) thus ensures that step length adaptation functions as conjectured above. In case of positive correlations in the sequence of steps taken by the strategy the mutation strength is increased. The mutation strength is decreased in response to negative correlations. Note that the prescription Eq. (4) for adapting the mutation strength has been changed slightly from the prescription in the original algorithm given by (Hansen, 1998) in that here, adaptation is performed on the basis of the squared length of the accumulated progress vector rather than on its length. The difference in performance appears to be insignificant except for very small values of N while elegance in the formulation is gained by the change.

It should be mentioned that cumulative step length adaptation is not the only mechanism conceivable for the adaptation of mutation strengths. Possible alternatives include:

Nested evolution strategies (Herdy, 1992; Rechenberg, 1994). Nested evolution strategies adjust strategy parameters such as mutation strengths by means of evolutionary optimization on a meta level. Several populations, each one with their own parameter settings, compete with each other for survival. After a number of time steps, the respective progress of the different strategies is examined. The mutation strengths of those populations that have achieved the largest progress are used as a basis for generating mutation strengths for the next round of competition by means of recombination and mutation.

Mutative self adaptation (Rechenberg, 1994; Schwefel, 1995). Mutative self-adaptation includes the mutation strengths into the optimization process at the same

hierarchical level as the object parameters of the problem. Different candidate solutions have differing mutation strengths. Assuming that favorable mutation strengths are more likely to generate successful offspring than unfavorable ones, selection of favorable mutation strengths is then a by-product of evolution.

Machine learning approaches (Ravisé and Sebag, 1996; Sebag et al., 1997). The problem of mutation strength adaptation can be understood as a learning task. An explicit memory of the evolutionary process can be created in the form of a set of rules by discriminating between successful trials and unsuccessful ones. Mutation strengths can then be adapted by means of inductive learning. (Michalski, 2000) suggests a different approach in which a machine learning system seeks reasons why certain individuals in a population are superior to others. These reasons, formulated as inductive hypotheses, are then used to generate new populations directly rather than by recombination and mutation.

In the present article, we choose to consider cumulative step length adaptation due to its relative mathematical tractability and as it is known to generate asymptotically optimal step lengths for the random tracking problem studied in (Arnold and Beyer, 2002).

3 The Tracking Problem

The purpose of this section is to introduce the tracking problem studied in the remainder of this article. In order to investigate the tracking behavior of the $(\mu/\mu, \lambda)$ -ES, we consider a dynamic version of the sphere model. The sphere model is the set of all functions $f : \mathbb{R}^N \rightarrow \mathbb{R}$ with

$$f(\mathbf{x}) = g(\|\hat{\mathbf{x}} - \mathbf{x}\|),$$

where $g : \mathbb{R} \rightarrow \mathbb{R}$ is a strictly monotonic function of the Euclidean distance $R = \|\hat{\mathbf{x}} - \mathbf{x}\|$ of a candidate solution \mathbf{x} from the target $\hat{\mathbf{x}}$. The sphere model has frequently served as a model for fitness landscapes at a stage where the population of candidate solutions is in relatively close proximity to the target and is most often studied in the limit of very high search space dimensionality. So as to study the tracking behavior of evolutionary algorithms, several authors (Angeline, 1997; Arnold and Beyer, 2002; Bäck, 1998; Salomon and Eggenberger, 1997; Weicker, 2006) have added a dynamic component to the sphere model by stipulating that the target $\hat{\mathbf{x}}$ vary with time. Several modes of motion of the target are conceivable and have been explored. Examples include random motion, linear motion, and circular motion in search space.

The case of random motion has been studied from a theoretical perspective in (Arnold and Beyer, 2002). If the target shifts such that

$$\hat{\mathbf{x}}^{(t+1)} = \hat{\mathbf{x}}^{(t)} + \delta \hat{\mathbf{z}}^{(t)},$$

where vector $\hat{\mathbf{z}}^{(t)}$ consists of N independent, standard normally distributed components that are drawn anew in every time step, then the following conclusions have been arrived at in the limit $N \rightarrow \infty$:

- The mutation strength that minimizes the distance at which the target is tracked is $\sigma = \sqrt{\mu}\delta$. As the goal is to track the target as closely as possible, this mutation strength is considered optimal. With optimal mutation strength, for the average distance from the target it follows $R/N = \delta/\sqrt{\mu}c_{\mu/\mu, \lambda}$, where $c_{\mu/\mu, \lambda}$ denotes the

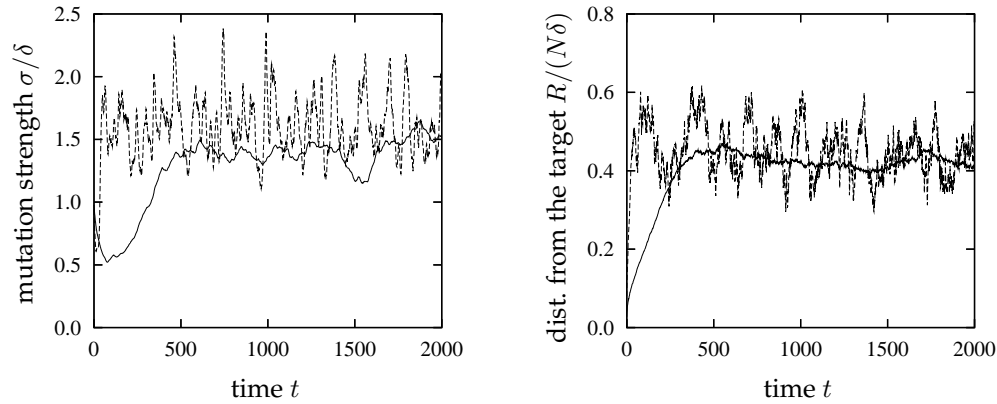


Figure 1: Mutation strength σ/δ and stationary distance from the target $R/(N\delta)$ as functions of the time t . The strategy is a $(3/3, 10)$ -ES with cumulative step length adaptation. The population is initialized to be centered at the location of the target at time step 0. The initial mutation strength is $\sigma = \delta$. The curves depict typical runs of the strategy with search space dimensionalities $N = 40$ (dashed lines) and $N = 400$ (solid lines).

$(\mu/\mu, \lambda)$ -progress coefficient defined in (Beyer, 2001). That distance is thus proportional to the speed of the target. Due to the properties of the $(\mu/\mu, \lambda)$ -progress coefficient¹, it can be decreased by increasing the population size parameters μ and λ in equal proportions.

- For optimally adapted σ , the normalized mutation strength $\sigma^* = \sigma N/R$ equals $\mu c_{\mu/\mu, \lambda}$ and therefore agrees with the optimal normalized mutation strength on the static sphere derived in (Beyer, 2001).²
- Cumulative step length adaptation successfully adapts the mutation strength and achieves the optimal setting on average.

Numerical experiments have been used to show that these predictions are reasonably accurate provided that the search space dimensionality is not too small.

In the present paper, we consider linear motion and assume that the target at time step $t + 1$ is

$$\hat{\mathbf{x}}^{(t+1)} = \hat{\mathbf{x}}^{(t)} + \delta \mathbf{v}, \quad (5)$$

where vector \mathbf{v} is independent of time and has unit length. As in the case of random motion, the factor δ is a measure for the speed of change of the objective. Figure 1 illustrates the time behavior of a $(\mu/\mu, \lambda)$ -ES with $\mu = 3$ and $\lambda = 10$ that uses cumulative step length adaptation for the tracking problem with linear dynamics of the target. Shown are measurements of typical runs of the strategy for search space dimensionalities $N = 40$ and $N = 400$. The results that are obtained for other values of μ and λ

¹The $(\mu/\mu, \lambda)$ -progress coefficient is the expectation of the μ last order statistics of a sample of λ independent, standard normally distributed random variables. It has been seen in (Beyer, 2001) that it tends to a constant limit value as μ and λ increase provided that the ratio μ/λ is kept constant.

²In contrast to the mutation strength that needs to be adapted as the distance from the target changes, the normalized mutation strength is independent of the location in search space. The factor N in the normalization is required in order to ensure that normalized mutation strengths are finite as $N \rightarrow \infty$.

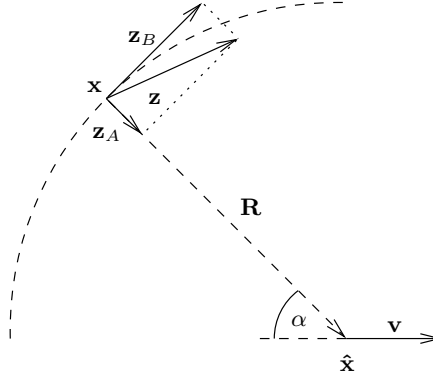


Figure 2: *Decomposition of a vector \mathbf{z} into central component \mathbf{z}_A and lateral component \mathbf{z}_B . Vector \mathbf{z}_A is parallel to $\mathbf{R} = \hat{\mathbf{x}} - \mathbf{x}$, vector \mathbf{z}_B is in the hyperplane perpendicular to that. The trailing angle α is the angle formed by vector \mathbf{R} and the direction \mathbf{v} of the motion of the target.*

qualitatively agree. It can be seen that after a transient period, a state in which the mutation strength and the distance from the target fluctuate around stationary mean values is reached. Additional measurements not included here seem to indicate that the length of the transient period is of order N . The stationary state that is reached is independent of the initialization of the strategy, justifying our decision not to discuss initialization issues. Furthermore, it can be seen that the magnitude of the fluctuations decreases with increasing search space dimensionality. It is the goal in the remainder of this paper to characterize the dependence of the stationary state on the population size parameters μ and λ , the speed δ of the target, and the search space dimensionality N .

Analyses of the behavior of evolution strategies on the sphere model rely on a decomposition of vectors that is illustrated in Figure 2. A (mutation or progress) vector \mathbf{z} originating at search space location \mathbf{x} can be written as the sum of two vectors \mathbf{z}_A and \mathbf{z}_B , where \mathbf{z}_A is parallel to $\mathbf{R} = \hat{\mathbf{x}} - \mathbf{x}$ and \mathbf{z}_B is in the hyperplane perpendicular to that. The vectors \mathbf{z}_A and \mathbf{z}_B are referred to as the central and lateral components of vector \mathbf{z} , respectively. Formally, the signed length z_A of the central component of vector \mathbf{z} can be computed as

$$z_A = \frac{\mathbf{R} \cdot \mathbf{z}}{R}. \quad (6)$$

The central component of vector \mathbf{z} is $\mathbf{z}_A = z_A \mathbf{R} / R$, the lateral component is $\mathbf{z}_B = \mathbf{z} - \mathbf{z}_A$. Due to the stochastic nature of the search process and in particular to the presence of the lateral component of mutation vectors, the population follows the target not on the straight line defined by vector \mathbf{v} but at an angle that is nonzero on average. As illustrated in Figure 2, the trailing angle α is defined as the angle formed by vector \mathbf{R} and the direction \mathbf{v} of the motion of the target in that

$$\cos \alpha = \frac{\mathbf{R} \cdot \mathbf{v}}{R}. \quad (7)$$

The analyses in Sections 4 and 5 will reveal that the average magnitude of that angle can be substantial.

Clearly, from Eqs. (2) and (5),

$$\mathbf{R}^{(t+1)} = \mathbf{R}^{(t)} + \delta \mathbf{v} - \sigma \langle \mathbf{z} \rangle^{(t)} \quad (8)$$

holds as a result of the motion of the target and the step that the strategy takes. Moreover, it is well known from (Beyer, 2001; Rechenberg, 1994) that in the limit $N \rightarrow \infty$, the expectation of the signed length of the central component of the progress vector is

$$E[\langle z_A \rangle] = c_{\mu/\mu, \lambda}. \quad (9)$$

The lateral component of the progress vector is of random direction in the hyperplane defined by normal vector $\mathbf{R} = \hat{\mathbf{x}} - \mathbf{x}$. Moreover, in the limit of infinite search space dimensionality, for its expected overall squared length

$$\frac{E[\|\langle \mathbf{z} \rangle\|^2]}{N} = \frac{1}{\mu}. \quad (10)$$

holds. The variance of $\|\langle \mathbf{z} \rangle\|^2/N$ is of order $1/N$, diminishing the relative influence of fluctuations as N tends to infinity. Notice that this effect is largely responsible for the decrease in fluctuations as N increases that had been observed in Figure 1. From Eq. (6) with Eq. (9) it follows that

$$E[\mathbf{R} \cdot \langle \mathbf{z} \rangle] = R c_{\mu/\mu, \lambda} \quad (11)$$

holds. Moreover, due to the randomness of the direction of the lateral component of the progress vector,

$$\begin{aligned} E[\langle \mathbf{z} \rangle \cdot \mathbf{v}] &= E[\langle \mathbf{z}_A \rangle \cdot \mathbf{v}] + \underbrace{E[\langle \mathbf{z}_B \rangle \cdot \mathbf{v}]}_{=0} \\ &= c_{\mu/\mu, \lambda} \cos \alpha \end{aligned} \quad (12)$$

holds as a consequence of Eqs. (6), (7), and (9).

4 Analysis for Constant Mutation Strength

Before proceeding to the analysis of the behavior of the $(\mu/\mu, \lambda)$ -ES with cumulative step length adaptation, in this section we first consider the case that the mutation strength is fixed. The strategy in combination with the dynamic objective function outlined in Section 3 forms a stochastic dynamic system. Due to the symmetries inherent in both the environment and the strategy, for fixed mutation strength, the state of that system can be described by just two variables. A possible choice for those variables are the distance R of the population centroid from the target and the cosine of the trailing angle α . Provided that the mutation strength is large enough for the strategy to be able to keep pace with the target, the system tends toward a limit state in which the state variables have time-invariant distributions.

The approach to determining the scaling behavior of the $(\mu/\mu, \lambda)$ -ES on the tracking problem with linear dynamics of the target is the same as that employed in (Arnold and Beyer, 2002) for the tracking problem with random dynamics of the target. It consists in using stationarity conditions while neglecting all fluctuations. In particular, it is assumed that the progress vector has a central component of length $c_{\mu/\mu, \lambda}$, that its lateral component is of random direction, and that its overall squared length is N/μ . As a consequence, quantities involving the progress vector can be replaced by their expectations given in Eqs. (9), (10), (11), and (12). The state variables tend to stationary limit values that can be obtained by relatively simple considerations. As fluctuations of the progress vector decrease with increasing search space dimensionality, the approach that neglects them becomes increasingly accurate with increasing N . It will be seen

that good agreement between predictions and measurements is obtained already for relatively moderate values of N .

Using Eqs. (7), (8), and (12) it follows that at time step $t + 1$

$$\begin{aligned} R^{(t+1)} \cos \alpha^{(t+1)} &= \mathbf{R}^{(t+1)} \cdot \mathbf{v} \\ &= \left(\mathbf{R}^{(t)} + \delta \mathbf{v} - \sigma \langle \mathbf{z} \rangle^{(t)} \right) \cdot \mathbf{v} \\ &\simeq R^{(t)} \cos \alpha^{(t)} + \delta - \sigma c_{\mu/\mu, \lambda} \cos \alpha^{(t)}. \end{aligned}$$

As stationarity of the state variables implies the stationarity of their product, it follows

$$\cos \alpha \simeq \frac{\delta}{\sigma c_{\mu/\mu, \lambda}} \quad (13)$$

for the stationary cosine of the trailing angle.

Similarly, the square of the distance from the population centroid to the target at time step $t + 1$ is

$$\begin{aligned} R^{(t+1)2} &= \mathbf{R}^{(t+1)} \cdot \mathbf{R}^{(t+1)} \\ &= \left(\mathbf{R}^{(t)} + \delta \mathbf{v} - \sigma \langle \mathbf{z} \rangle^{(t)} \right) \cdot \left(\mathbf{R}^{(t)} + \delta \mathbf{v} - \sigma \langle \mathbf{z} \rangle^{(t)} \right) \\ &\simeq R^{(t)2} + \delta^2 + \frac{N\sigma^2}{\mu} + 2\delta R^{(t)} \cos \alpha^{(t)} - 2\sigma R^{(t)} c_{\mu/\mu, \lambda} - 2\delta \sigma c_{\mu/\mu, \lambda} \cos \alpha^{(t)}, \end{aligned}$$

where Eqs. (7), (8), (10), (11), and (12), have been used. Stationarity of the distance from the target means that $R^{(t+1)} = R^{(t)} = R$. Rearranging terms and using Eq. (13) to replace $\cos \alpha$ yields

$$2R \left(\sigma c_{\mu/\mu, \lambda} - \frac{\delta^2}{\sigma c_{\mu/\mu, \lambda}} \right) \simeq \frac{N\sigma^2}{\mu} \left(1 - \frac{\mu\delta^2}{N\sigma^2} \right).$$

In order for the strategy to be able to track the target, the mutation strength must be large enough to satisfy $\delta/\sigma = \mathcal{O}(1)$. Therefore, the second term in the parentheses on the right hand side vanishes compared to the first as $N \rightarrow \infty$ and it follows

$$R \simeq \frac{N\sigma^3 c_{\mu/\mu, \lambda}}{2\mu(\sigma^2 c_{\mu/\mu, \lambda}^2 - \delta^2)} \quad (14)$$

for the stationary distance from the population centroid to the target.

The quality of the approximation to the stationary state thus derived is illustrated in Figure 3. Predictions from Eqs. (13) and (14) are compared with empirical measurements of runs of evolution strategies. For those measurements, it is irrelevant whether results from one run are used or several runs are averaged, provided that the length of the runs is sufficient. For the experiments reported here, the length of the runs was chosen to ensure that the standard deviation of the measurements is below the size of the crosses in Figure 3. It can be seen from the figure that for search space dimensionality $N = 40$, the qualitative dependence of the performance of the strategy on the mutation strength and the target speed is reflected properly. The deviations that can be observed are attributable to the fluctuations inherent in the runs that have been neglected in the calculations. For search space dimensionality $N = 400$, very good quantitative agreement of the predictions with the empirical measurements can be observed.

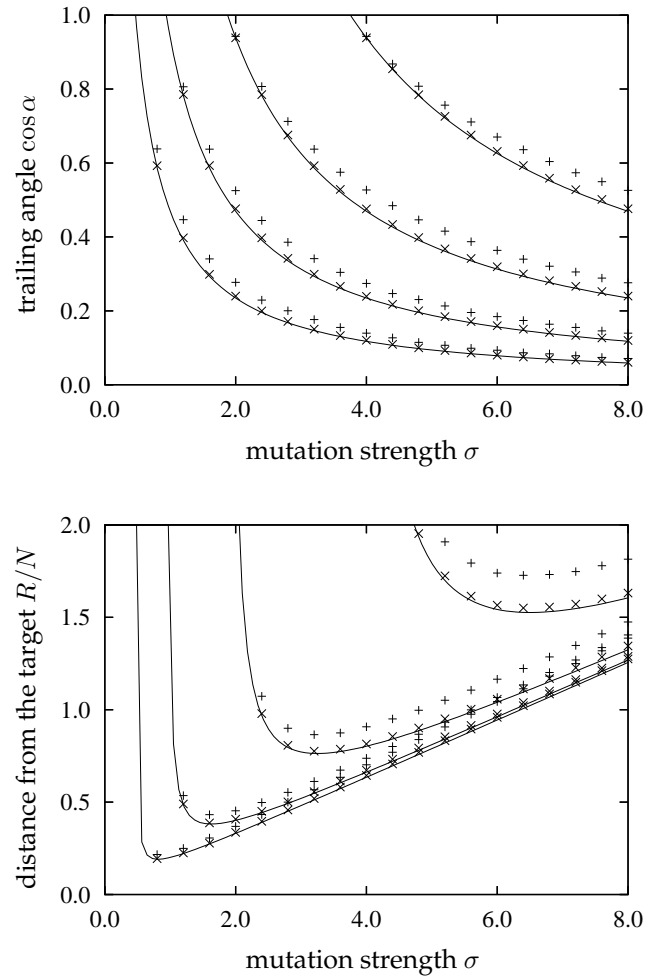


Figure 3: Cosine of the trailing angle and distance from the target as functions of mutation strength σ . The target moves linearly with, from bottom to top, speeds $\delta = 0.5, 1.0, 2.0$, and 4.0 . The strategy is a $(3/3, 10)$ -ES. The solid lines are predictions from Eqs. (13) and (14), the points mark empirical measurements for search space dimensionalities $N = 40$ (+) and $N = 400$ (x).

Some important conclusions with regard to the scaling behavior of the strategy can be drawn from Eqs. (13) and (14). First, from Eq. (14) it can be seen that a stationary state is reached only if $\sigma > \delta/c_{\mu/\mu,\lambda}$. For smaller mutation strengths, the strategy is not able to keep pace with the target and the distance from the population centroid to the target increases indefinitely. Second, Eq. (14) can be used to obtain the optimal mutation strength, i.e. the mutation strength that minimizes the distance to the target. Computing the derivative with respect to σ and determining the root yields optimal mutation strength

$$\sigma \simeq \frac{\sqrt{3}\delta}{c_{\mu/\mu,\lambda}}. \quad (15)$$

Reinserting this result in Eq. (14), the resulting distance from the target is

$$R \simeq \frac{3\sqrt{3}N\delta}{4\mu c_{\mu/\mu,\lambda}^2}. \quad (16)$$

Therefore, combining Eqs. (15) and (16), we have

$$\sigma^* = \sigma \frac{N}{R} \simeq \frac{4}{3}\mu c_{\mu/\mu,\lambda} \quad (17)$$

for the optimal normalized mutation strength. Comparison with the corresponding result from (Arnold and Beyer, 2002) reveals that this differs by a factor of 4/3 from the optimal mutation strength for both the static case and the case of a randomly moving target. As in the case of a linearly moving target the cosine of the trailing angle is consistently greater than zero, the strategy should ideally operate with a larger mutation strength.

5 Analysis for Adaptive Mutation Strength

For a randomly moving target, cumulative step length adaptation has in (Arnold and Beyer, 2002) been seen to be able to generate asymptotically optimal mutation strengths. On the static sphere model considered in (Arnold, 2002; Arnold and Beyer, 2004), mutation strengths generated using cumulative step length adaptation are by a factor of $\sqrt{2}$ larger than optimal due to the fact that the distance to the target varies and the adaptation mechanism lags behind. In the present section, we study the performance of cumulative step length adaptation for the case of the linearly moving target.

For fixed mutation strength, we have seen in the previous section that a system state is fully characterized by two state variables. The distance from the target and the cosine of the trailing angle were identified as a possible choice for the state variables. When using cumulative step length adaptation, a total of three variables beyond those used in Section 4 are required to describe the state of the system. While again several choices are conceivable, one possibility is to consider the mutation strength σ along with the signed length

$$s_A = \frac{\mathbf{R} \cdot \mathbf{s}}{R} \quad (18)$$

of the central component of the accumulated progress vector and that vector's overall squared length $\|\mathbf{s}\|^2$. In analogy to Eq. (12),

$$\begin{aligned} \mathbb{E}[\mathbf{s} \cdot \langle \mathbf{z} \rangle] &= \mathbb{E}[s_A \langle z_A \rangle] + \underbrace{\mathbb{E}[\mathbf{s}_B \cdot \langle \mathbf{z}_B \rangle]}_{=0} \\ &= s_A c_{\mu/\mu,\lambda} \end{aligned} \quad (19)$$

holds due to Eq. (9) and the randomness of the direction of the lateral component of progress vectors. Note that the same choice of variables has been made in (Arnold, 2002; Arnold and Beyer, 2002; Arnold and Beyer, 2004) for the case of the static sphere model and the sphere model with random dynamics of the target.

As in Section 4, we neglect fluctuations of the progress vector by replacing random variables with their expected values and make use of the stationarity of the limit values of the state variables. The inner product $\mathbf{s} \cdot \mathbf{v}$ at time step $t + 1$ is

$$\begin{aligned}\mathbf{s}^{(t+1)} \cdot \mathbf{v} &= \left((1 - c)\mathbf{s}^{(t)} + \sqrt{\mu c(2 - c)}\langle \mathbf{z} \rangle^{(t)} \right) \cdot \mathbf{v} \\ &\simeq (1 - c)\mathbf{s}^{(t)} \cdot \mathbf{v} + \sqrt{\mu c(2 - c)}c_{\mu/\mu, \lambda} \cos \alpha^{(t)},\end{aligned}$$

where Eqs. (3) and (12) have been used. From the stationarity of the state variables follows the stationarity of the product $\mathbf{s} \cdot \mathbf{v}$ and therefore with Eq. (13) to replace $\cos \alpha$

$$\mathbf{s} \cdot \mathbf{v} \simeq \sqrt{\frac{\mu(2 - c)}{c}} \frac{\delta}{\sigma} \quad (20)$$

for the stationary value of the inner product of vectors \mathbf{s} and \mathbf{v} .

Similarly, using Eq. (18), at time step $t + 1$,

$$\begin{aligned}R^{(t+1)}s_A^{(t+1)} &= \mathbf{R}^{(t+1)} \cdot \mathbf{s}^{(t+1)} \\ &= \left(\mathbf{R}^{(t)} + \delta \mathbf{v} - \sigma^{(t)}\langle \mathbf{z} \rangle^{(t)} \right) \cdot \left((1 - c)\mathbf{s}^{(t)} + \sqrt{\mu c(2 - c)}\langle \mathbf{z} \rangle^{(t)} \right) \\ &\simeq (1 - c)R^{(t)}s_A^{(t)} + (1 - c)\sqrt{\frac{\mu(2 - c)}{c}} \frac{\delta^2}{\sigma^{(t)}} - (1 - c)\sigma^{(t)}c_{\mu/\mu, \lambda}s_A^{(t)} \\ &\quad + \sqrt{\mu c(2 - c)}R^{(t)}c_{\mu/\mu, \lambda} + \sqrt{\mu c(2 - c)} \frac{\delta^2}{\sigma^{(t)}} - \sqrt{\mu c(2 - c)} \frac{N\sigma^{(t)}}{\mu},\end{aligned}$$

where Eqs. (3), (8), (10), (11), (12), (13), (19), and (20) have been used. Assuming stationarity, i.e. $R^{(t+1)}s_A^{(t+1)} = R^{(t)}s_A^{(t)} = Rs_A$, and rearranging terms yields

$$s_A \left(c + (1 - c) \frac{\sigma c_{\mu/\mu, \lambda}}{R} \right) \simeq \sqrt{\mu c(2 - c)} \left(\frac{\delta^2}{cR\sigma} + c_{\mu/\mu, \lambda} - \frac{N\sigma}{R\mu} \right).$$

As seen in Section 4, $\sigma/R = \mathcal{O}(1/N)$ holds. Therefore, the second term in the parentheses on the left hand side vanishes compared to the first as $N \rightarrow \infty$. Moreover, as in order for the strategy to be able to track the target $\delta/\sigma = \mathcal{O}(1)$ must hold, the first term in the parentheses on the right hand side vanishes compared to the third. Rearranging the remaining terms and using Eq. (14) yields

$$\begin{aligned}s_A &\simeq \sqrt{\frac{\mu(2 - c)}{c}} \left(c_{\mu/\mu, \lambda} - \frac{N\sigma}{R\mu} \right) \\ &\simeq \sqrt{\frac{\mu(2 - c)}{c}} \left(\frac{2\delta^2}{c_{\mu/\mu, \lambda}\sigma^2} - c_{\mu/\mu, \lambda} \right)\end{aligned} \quad (21)$$

for the stationary value of the signed length of the central component of the accumulated progress vector.

Likewise, the squared length of the accumulated progress vector at time step $t + 1$ is

$$\begin{aligned}\|\mathbf{s}^{(t+1)}\|^2 &= \left((1-c)\mathbf{s}^{(t)} + \sqrt{\mu c(2-c)}\langle \mathbf{z} \rangle^{(t)} \right) \cdot \left((1-c)\mathbf{s}^{(t)} + \sqrt{\mu c(2-c)}\langle \mathbf{z} \rangle^{(t)} \right) \\ &= (1-c)^2 \|\mathbf{s}^{(t)}\|^2 + 2(1-c)\sqrt{\mu c(2-c)}\mathbf{s}^{(t)} \cdot \langle \mathbf{z} \rangle^{(t)} + \mu c(2-c) \|\langle \mathbf{z} \rangle^{(t)}\|^2 \\ &\simeq (1-c)^2 \|\mathbf{s}^{(t)}\|^2 + 2(1-c)\sqrt{\mu c(2-c)}c_{\mu/\mu,\lambda}s_A^{(t)} + c(2-c)N,\end{aligned}$$

where Eqs. (3), (10), and (19) have been used. Demanding stationarity and using Eq. (21) yields

$$\|\mathbf{s}\|^2 \simeq 2\frac{1-c}{c}c_{\mu/\mu,\lambda} \left(\mu c_{\mu/\mu,\lambda} - \frac{\sigma N}{R} \right) + N \quad (22)$$

for the stationary squared length of the accumulated progress vector.

Finally, according to Eq. (4), the stationary mutation strength, i.e. the mutation strength for which cumulative step length adaptation does not affect a change, is that for which $\|\mathbf{s}\|^2 = N$ holds. From Eq. (22) it thus follows

$$\sigma^* = \sigma \frac{N}{R} \simeq \mu c_{\mu/\mu,\lambda} \quad (23)$$

for the normalized mutation strength that cumulative mutation strength adaptation realizes for the tracking problem with linear dynamics of the target. Using Eq. (14) to eliminate the distance from the target shows that the corresponding stationary mutation strength is

$$\sigma \simeq \frac{\sqrt{2}\delta}{c_{\mu/\mu,\lambda}}, \quad (24)$$

and that the resulting distance between the population centroid and the target is

$$R \simeq \frac{\sqrt{2}N\delta}{\mu c_{\mu/\mu,\lambda}^2}. \quad (25)$$

Figure 4 illustrates the quality of the approximation. While again for $N = 40$ some deviations can be observed, the agreement is very good for $N = 400$. Comparison of Eqs. (24) and (25) with the results Eqs. (15) and (16) for the optimal mutation strength derived in Section 4 shows that the mutation strength that cumulative step length adaptation realizes is by a factor of $\sqrt{2/3} \approx 0.82$ smaller than what it would ideally be, and that the distance between the population centroid and the target is by a factor of $4\sqrt{2}/(3\sqrt{3}) \approx 1.09$ larger than it would be if the optimal mutation strength were achieved. Table 1 summarizes the findings and compares them with the corresponding results for the random dynamics case.

6 Discussion and Conclusions

In this paper, the performance of the $(\mu/\mu, \lambda)$ -ES with cumulative step length adaptation has been analyzed for a tracking problem with linear dynamics of the target. It has been seen that the adaptation of the step length works in the sense that a mutation strength that ensures that the target can be tracked is realized. The mutation strength in Eq. (24) exceeds the minimum mutation strength $\sigma = \delta/c_{\mu/\mu,\lambda}$ required to keep pace with the target that was obtained in Section 4. However, it has also been seen that the mutation strength that is realized is below the mutation strength given in Eq. (15) that

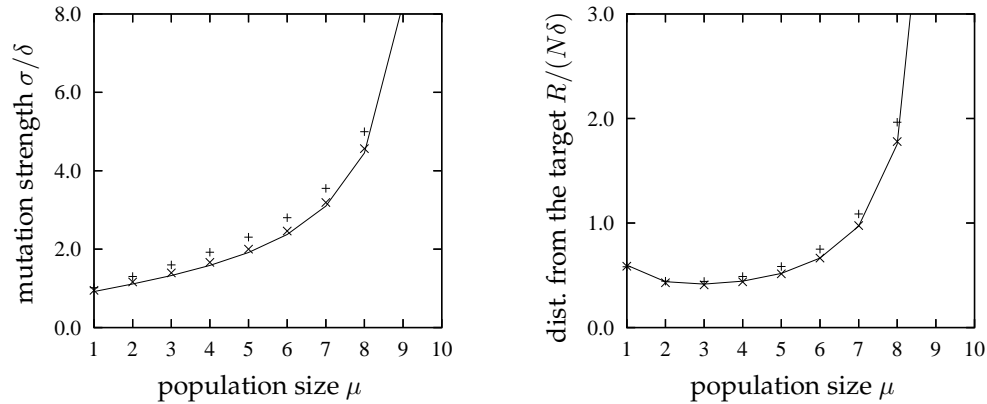


Figure 4: Mutation strength σ/δ and stationary distance from the target $R/(N\delta)$ as functions of the population size μ . The strategy is a $(\mu/\mu, 10)$ -ES. The solid lines are obtained from Eqs. (24) and (25). The points mark empirical measurements for search space dimensionalities $N = 40$ (+) and $N = 400$ (x).

was found to be optimal. An explanation for that behavior can be found from Eq. (24). The mutation strength that cumulative step length adaptation realizes on the tracking problem is the same as the target mutation strength (i.e., the mutation strength that the strategy seeks to attain) on the static sphere. While for the tracking problem with random dynamics of the target that mutation strength is optimal, linear dynamics of the target ideally call for somewhat larger mutation strengths. The strategy does not anticipate changes of the objective and thus realizes the mutation strength that is statically optimal. However, it has been seen that the increase in distance at which the target is tracked that results from the use of a suboptimal mutation strength is below 10% and thus relatively minor. Our concern voiced in (Arnold and Beyer, 2002) that the correlations in the sequence of steps taken that result from the linear motion of the target could lead to a degradation of the performance of cumulative step length adaptation has proven largely groundless at least in the limit of infinite search space dimensionality.

It has furthermore been seen that for the problem considered, the distance at which the target is tracked depends on the search space dimensionality, the speed of the target, and the population size parameters that the strategy employs. In order to minimize that distance, the population size parameters must be chosen such that the denominator of the fraction in Eq. (25) is maximized. Interestingly, that denominator equals the progress rate of the $(\mu/\mu, \lambda)$ -ES on the static sphere derived in (Beyer, 2001; Rechenberg, 1994). Recommendations with respect to the choice of population size parameters that have been made for the static sphere thus also hold for the dynamic tracking problem.

The results derived in this paper apply to the $(1, \lambda)$ -ES (that operates with a population of size one and therefore without recombination) by virtue of the specialization $\mu = 1$. It can be seen from Eq. (25) that using a population size of $\mu > 1$ has the effect of reducing the distance at which the target is tracked by a factor of $c_{1,\lambda}^2/(\mu c_{\mu/\mu,\lambda}^2)$. For example, for $\lambda = 10$ and $\mu = 3$, that factor is roughly 0.7. For $\lambda = 100$ and $\mu = 30$ the factor is approximately 0.16. Interestingly, and in contrast to the random dynamics case, the closer proximity to the target is not achieved by means of a substantially

	random dynamics	linear dynamics
optimal	$\sigma = \sqrt{\mu}\delta$ $R = \frac{N\delta}{\sqrt{\mu}c_{\mu/\mu,\lambda}}$	$\sigma = \frac{\sqrt{3}\delta}{c_{\mu/\mu,\lambda}}$ $R = \frac{3\sqrt{3}N\delta}{4\mu c_{\mu/\mu,\lambda}^2}$
realized	same as above	$\sigma = \frac{\sqrt{2}\delta}{c_{\mu/\mu,\lambda}}$ $R = \frac{\sqrt{2}N\delta}{\mu c_{\mu/\mu,\lambda}^2}$

Table 1: Comparison of the performances of the $(\mu/\mu, \lambda)$ -ES with cumulative step length adaptation for the random dynamics and the linear dynamic tracking problems.

larger mutation strength. Notice however that the reduction in distance from the target results in a larger *normalized* mutation strength.

Finally, it is important to reemphasize that all findings reported in this paper have been derived under relatively specific conditions, and that they represent but a small step on the way to a predictive theory of the dynamics of evolutionary search. The range of problems considered here is broad in that both linear and random modes of motion are now understood, and that it only needs to be assumed that objective function values increase with increasing distance from the target. It is limited in that the analysis is not applicable to environments that are not spherically symmetric, and to dynamic optimization problems other than tracking problems. Similarly, in a sense, the range of strategies considered here is larger than what can be considered in experimental studies as no specific assumptions with regard to the setting of the population size or other parameters of the algorithms need to be made. It is limited mostly by the restriction to isotropic mutation distributions, and more work is required in order to understand the behavior of covariance matrix adaptation algorithms such as that by (Hansen and Ostermeier, 2001) in dynamic environments. A further avenue for future research is to consider forms of recombination other than the global intermediate variant. It is conceivable that dynamic optimization may benefit from populations that are spread out in search space rather than being contracted to a point in every time step, and it remains to be seen whether the gain in diversity can compensate for the loss of information from all parents when doing recombination.

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) under grant Be1578/6-3 and as part of the Collaborative Research Center *Computational Intelligence* (SFB 531) located at the University of Dortmund. Support from the Natural Sciences and Engineering Research Council of Canada (NSERC) is also acknowledged.

References

- Angeline, P. J. (1997). Tracking extrema in dynamic environments. In Angeline, P. J. et al., editors, *Proc. of the Sixth International Conference on Evolutionary Programming — EP'97*, pages 335–345. Springer Verlag, Heidelberg.

- Arnold, D. V. (2002). *Noisy Optimization with Evolution Strategies*. Genetic Algorithms and Evolutionary Computation Series. Kluwer Academic Publishers, Boston, MA.
- Arnold, D. V. and Beyer, H.-G. (2002). Random dynamics optimum tracking with evolution strategies. In Merelo, J. J. et al., editors, *Parallel Problem Solving from Nature — PPSN VII*, pages 3–12. Springer Verlag, Heidelberg.
- Arnold, D. V. and Beyer, H.-G. (2004). Performance analysis of evolutionary optimization with cumulative step length adaptation. *IEEE Transactions on Automatic Control*, 49(4):617–622.
- Bäck, T. (1996). *Evolutionary Algorithms in Theory and Practice*. Oxford University Press, New York.
- Bäck, T. (1998). On the behavior of evolutionary algorithms in dynamic environments. In *Proc. of the 1998 International Conference on Evolutionary Computation*, pages 446–451. IEEE Press, Piscataway, NJ.
- Beyer, H.-G. (2001). *The Theory of Evolution Strategies*. Natural Computing Series. Springer Verlag, Heidelberg.
- Beyer, H.-G., Schwefel, H.-P., and Wegener, I. (2002). How to analyze evolutionary algorithms. *Theoretical Computer Science*, 287(1):101–130.
- Branke, J. (2001). *Evolutionary Optimization in Dynamic Environments*. Genetic Algorithms and Evolutionary Computation Series. Kluwer Academic Publishers, Boston, MA.
- Bürger, R. (2000). *The Mathematical Theory of Selection, Recombination, and Mutation*. John Wiley & Sons, Chichester.
- Droste, S. (2002). Analysis of the $(1 + 1)$ EA for a dynamically changing objective function. In *Proc. of the 2002 Congress on Evolutionary Computation*. IEEE Press, Piscataway, NJ.
- Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison Wesley, Reading, MA.
- Hansen, N. (1998). *Verallgemeinerte individuelle Schrittweitenregelung in der Evolutionsstrategie*. Mensch & Buch Verlag, Berlin.
- Hansen, N. and Ostermeier, A. (2001). Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159–195.
- Herdy, M. (1992). Reproductive isolation as a strategy parameter in hierarchically organized evolution strategies. In Männer, R. and Manderick, B., editors, *Parallel Problem Solving from Nature — PPSN II*, pages 207–217. Elsevier, Amsterdam.
- Jansen, T. and Schellbach, U. (2005). Theoretical analysis of a mutation-based evolutionary algorithm for a tracking problem in the lattice. In Beyer, H.-G. et al., editors, *Proc. of the Genetic and Evolutionary Computation Conference — GECCO 2005*, pages 841–848. ACM Press, New York.
- Jin, Y. and Branke, J. (2005). Evolutionary optimization in uncertain environments — A survey. *IEEE Transactions on Evolutionary Computation*, 9(3):303–317.
- Michalski, R. S. (2000). Learnable evolution model: Evolutionary processes guided by machine learning. *Machine Learning*, 38(1-2):9–40.
- Mitchell, M. (1996). *An Introduction to Genetic Algorithms*. MIT Press, Cambridge, MA.
- Morrison, R. W. (2004). *Designing Evolutionary Algorithms for Dynamic Environments*. Natural Computing Series. Springer Verlag, Heidelberg.
- Ostermeier, A., Gawelczyk, A., and Hansen, N. (1994). Step-size adaptation based on non-local use of selection information. In Davidor, Y., Schwefel, H.-P., and Männer, R., editors, *Parallel Problem Solving from Nature — PPSN III*, pages 189–198. Springer Verlag, Heidelberg.

- Ravisé, C. and Sebag, M. (1996). An advanced evolution should not repeat its past errors. In Saïta, L., editor, *Proc. of the 13th International Conference on Machine Learning*, pages 400–408. Morgan Kaufmann Publishers, San Francisco, CA.
- Rechenberg, I. (1994). *Evolutionsstrategie '94*. Frommann-Holzboog Verlag, Stuttgart.
- Salomon, R. and Eggenberger, P. (1997). Adaptation on the evolutionary time scale: A working hypothesis and basic experiments. In Hao, J.-K. et al., editors, *Proc. of the Third Conference on Artificial Evolution — EA'97*, pages 251–262. Springer Verlag, Heidelberg.
- Schwefel, H.-P. (1995). *Evolution and Optimum Seeking*. Wiley, New York.
- Sebag, M., Schoenauer, M., and Ravisé, C. (1997). Inductive learning of mutation step-size in evolutionary parameter optimization. In Angeline, P. J. et al., editors, *Proc. of the Sixth International Conference on Evolutionary Programming — EP'97*, pages 247–261. Springer Verlag, Heidelberg.
- van Nimwegen, E. and Crutchfield, J. P. (2001). Optimizing epochal evolutionary search: Population-size dependent theory. *Machine Learning*, 45(1):77–114.
- Weicker, K. (2006). An empirical investigation of optimum tracking with evolution strategies. In *Proc. of the 9th Online World Conference on Soft Computing and Industrial Applications — WSC9*. Springer Verlag, Heidelberg.
- Weicker, K. and Weicker, N. (1999). On evolution strategy optimization in dynamic environments. In *Proc. of the 1999 Congress on Evolutionary Computation*, pages 2039–2046. IEEE Press, Piscataway, NJ.