Some Comments on Evolutionary Algorithm Theory

Ralf Salomon
AI Laboratory
Department of Computer Science
University of Zurich
Winterthurerstrasse 190
8057 Zurich, Switzerland
salomon@ifi.unizh.ch

Abstract

The development of a sound theory that predicts and verifies existing evolutionary algorithms (EA) is one of the most important research issues in the field today. In mathematical proofs, the assumption of spherical symmetry is probably one of the most widely used simplifications. This paper discusses the extent to which spherical symmetry is appropriate for certain EAs. It turns out that spherical symmetry leads to simplifications in (self-adaptive) EAs but seems inappropriate for certain genetic algorithm variants, since small mutation rates bias a search algorithm toward the coordinate axes. This paper also argues that current test suites are weak in that they do not provide problems with significant epistasis that describes the interaction between different parameters. Consequently, when using an empirical test for pushing existing theory beyond its limits, benchmark functions should include more epistatic interaction or at least should use coordinate rotations.

Keywords

Evolutionary algorithm theory, spherical symmetry, multimodal functions, performance degradation, search bias.

1. Introduction

Evolutionary algorithms (EAs), also called evolutionary computation, are a class of stochastic optimization/adaptation techniques that have been successfully applied in diverse areas, such as machine learning, combinatorial problems, VLSI design, and numerical optimization (Srinivas & Patnaik, 1994). Over time, success itself has initiated the development of various EA variants as well as several new evolutionary operators, and the efficacy of new algorithms is often examined by applying them to different benchmark functions.

The development of a sound theory that describes and predicts the behavior of EAs is one of the most important issues of current research. It is rather difficult to develop a sound EA theory, since EAs use several probabilistic operators, such as mutation, recombination, or roulette wheel selection, that require certain simplifications and lead to statistical results. Like (classical) one-point optimization techniques (e.g., gradient descent), most theories for evolution strategies (ESs), a member of EAs, assume spherical symmetry. Section 2 discusses some examples in more detail and explains why spherical symmetry can drastically simplify a theory.

The theory for genetic algorithms (GAs), another member of EAs, is traditionally based on the schema theorem and the building block hypothesis (see, e.g., Goldberg, 1989). How-

ever, recent discussion (Altenberg, 1995; Grefenstette & Baker, 1989; Vose, 1991) questions whether the schema theorem appropriately describes the behavior of GAs.

The theory for the breeder genetic algorithm (BGA), a particular GA variant proposed in Mühlenbein and Schlierkamp-Voosen (1993), is different from traditional GA theory in that it is based on the assumption of spherical symmetry. The BGA has attracted recent attention because it provides a complete theory and yields striking results on many test functions, especially on multimodal test functions. Section 3 summarizes the most important steps of the original BGA theory and discusses the fact that small mutation rates are not rotationally invariant, since a rotation does change the average progress. Section 4 then extends the n-dimensional analysis of mutation to multimodal functions and indicates that small mutation rates might increase the computational complexity up to $O(n^n)$ for finding the global optimum.

The self-adaptation mechanism for the mutation strength is a major component of ESs. Even though several mechanisms have been proposed that self-adapt the GA's mutation probability, no adaptation scheme is usually used for the adaptation of the mutation strength itself. An interesting claim of the original BGA theory is that the BGA mutation operator does not require any self-adaptation in order to yield nearly optimal progress. This issue is discussed in Section 5. Finally, Section 6 concludes with an outlook.

2. Spherical Symmetry

Spherical symmetry is probably one of the most widely used simplifications assumed in mathematical analyses. This assumption is often reasonable, especially in the neighborhood of the (global) optimum. The analysis of steepest descent algorithms is a classical example that might assume this simplification (see, e.g., Luenberger, 1984). The steepest descent algorithm calculates the gradient in the following way: $\mathbf{g} = \nabla f(\mathbf{x}) = (\partial/\partial x_1, \ldots, \partial/\partial x_n)^T f(\mathbf{x})$. This form of calculation always ensures that the gradient points in the direction of the maximal increase of the objective function $f(\mathbf{x})$. This direction is, of course, independent of the particular orientation of the coordinate system.[1] Therefore, spherical symmetry is a very convenient assumption, since the application of a coordinate rotation does not affect the gradient; without loss of generality, all points of a hypersphere can be moved to a particular point on the x_1-axis.

As mentioned in the introduction, approaches for the development of a theory for the ES also assume spherical symmetry. Examples can be found in Rechenberg (1973) and Beyer (1993, 1995, 1996); but for ESs, three different cases should be distinguished. Normally, the ES applies $(0, \sigma)$-Gaussian mutations with a probability $p_m = 1$ for each parameter x_i. In its simplest form (Rechenberg, 1973), the ES maintains one global step size σ for each individual. Since the application of Gaussian random numbers to *all* parameters x_i distributes all offspring on a hypersphere, the behavior of the simple ES is (almost) rotationally invariant.

The situation changes, however, if an enhanced ES version uses individual step sizes σ_i for each parameter x_i. Here, all offspring are not distributed on a hypersphere but on hyperellipsoids. This technique transforms all hyperellipsoids that are *aligned* with the coordinate system into virtual hyperspheres. Obviously, this form of compensation is limited to aligned objective functions, and, therefore, this mutation operator is not rotationally invariant (see also, Ostermeier, Gawelczyk, & Hansen, 1994).

1 For non-Cartesian coordinate systems, the calculation of the gradient changes, but the result is independent of the particular coordinate system.

Table 1. A small collection of widely used test functions.

	Function	Limits	Name
1	$f_1(\mathbf{x}) = \sum_{i=1}^{3} x_i^2 = \|\|\mathbf{x}\|\|$	$-5.12 \leq x_i \leq 5.12$	Sphere
2	$f_2(\mathbf{x}) = 100(x_1^2 - x_2)^2 + (1 - x_i)^2$	$-2.048 \leq x_i \leq 2.048$	Rosenbrock's
3	$f_3(\mathbf{x}) = \sum_{i=1}^{5} \text{integer}(x_i)$	$-5.12 \leq x_i \leq 5.12$	
4	$f_4(\mathbf{x}) = \sum_{i=1}^{30} i x_i^4 + \text{Gauss}(0, 1)$	$-1.28 \leq x_i \leq 1.28$	
5	$f_5(\mathbf{x}) = 0.002 + \sum_{j=1}^{25} \frac{1}{c_j + \sum_{i=1}^{2} (x_i - a_{ij})^6}$	$-65.536 \leq x_i \leq 65.536$	Shekel's
6	$f_6(\mathbf{x}) = \sum_{i=1}^{20} \left[x_i^2 - 10\cos(2\pi x_i) + 10\right]$	$-5.12 \leq x_i \leq 5.12$	Rastrigin's
7	$f_7(\mathbf{x}) = \sum_{i=1}^{n} -x_i \sin\left(\sqrt{\lvert x_i \rvert}\right)$	$-500 \leq x_i \leq 500$	Schwefel's
8	$f_8(\mathbf{x}) = 0.00025 \sum_{i=1}^{n} x_i^2 - \prod_{i=1}^{n} \cos\left(x_i/\sqrt{i}\right) + 1$	$-600 \leq x_i \leq 600$	Griewangk's
9	$f_9(\mathbf{x}) = -20\exp\left(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}\right) - \exp\left(\frac{1}{n}\sum_{i=1}^{n} \cos(2\pi x_i)\right) + 20 + e$	$-30 \leq x_i \leq 30$	Ackley's

In its full form, the ES uses correlated mutations (Schwefel, 1995) **Mz**, with **M** denoting an n by n correlation matrix and **z** denoting a random vector with Gaussian distributed components. Since the correlation matrix **M** is updated by a self-adaptation mechanism, this form of mutation transforms any hyperellipsoid into a hypersphere. Consequently, correlated mutations are rotationally invariant, and spherical symmetry is again a convenient assumption.

Despite its convenience, the assumption of spherical symmetry limits the analysis to unimodal functions. It is desirable to extend any theory to multimodal functions, since many practical relevant fitness functions are multimodal and since EAs have been successfully applied to various multimodal functions.

3. Small Mutation Rates and Rotational Invariance

This section is devoted to the problems that small mutation rates impose on the analysis of the average progress under the assumption of spherical symmetry. These problems can be best illustrated with an example, for which we have chosen the BGA for the following two reasons. First, the BGA yields outstanding results on many test functions, especially on multimodal test functions, such as $f_6 - f_9$ in Table 1; and second, the BGA provides an accompanying theory that assumes spherical symmetry. The theoretical problems discussed in this paper are further supported by practical evidence presented in Salomon (1996a).

Before discussing the problems on a more theoretical level, this section first outlines the BGA and its theory, as proposed in Mühlenbein and Schlierkamp-Voosen (1993). The BGA

is tailored to the optimization of continuous parameter problems. It encodes each parameter x_i as a floating-point value and implements mutation by adding or subtracting small random numbers. The BGA applies a mutation to each parameter with probability p_m. Normally, the mutation probability p_m equals $1/n$, with n denoting the number of parameters. The BGA hence mutates, on average, one parameter x_i at a time. In addition, it features different crossover operators, such as discrete recombination, extended intermediate recombination, or extended line recombination.

The BGA's mutation operator is defined as follows (we take a typical example):

$$\delta = A \sum_{i=0}^{15} \alpha_i 2^{-i}, \qquad \alpha_i \in 0, 1, \tag{1}$$

where A specifies the mutation range, which might be reduced during evolution, and each α_i is set to the value 1 with probability 1/16 and 0 with probability 15/16. On average, the BGA's mutation operator picks only one nonzero α_j, yielding $\delta = 2^{-j}$ for some j, with $0 \leq j \leq 15$. In more recent implementations (e.g., Schlierkamp-Voosen & Mühlenbein 1994), mutations are extended to continuous values as opposed to discrete values produced by Equation 1.

As already outlined, spherical symmetry is often assumed in order to simplify further theoretical considerations. The nice advantage of assuming spherical symmetry is that often *but not always* a coordinate rotation can be applied without changing the algorithm's rate of progress. The same idea is used in the basic step of the BGA proof, in which it is assumed that "the rotation does not change the average progress because we assume that the fitness function is spherically symmetric" (Mühlenbein & Schlierkamp-Voosen, 1993, p. 31).

The BGA analysis is based on the *expected progress* $E(n, r)$ for an arbitrary point $\mathbf{x} = (x_1, \ldots, x_n)$ in an n-dimensional search space with distance r to the (global) optimum. The expected progress is defined as an integral over successful mutations and originates from a product of the *progress(y)* and its *probability(y)*. The original analysis assumes "a unimodal function with spherical symmetry" (Mühlenbein & Schlierkamp-Voosen, 1993, p. 29) and that "a variable is chosen with probability $p_m = 1/n$ for mutation" (Mühlenbein & Schlierkamp-Voosen, 1993, p. 31), which means that, "on average, just one variable will be mutated." Furthermore, for the sake of simplicity, it is assumed that the (global) optimum $\mathbf{x}^0$ is located at the origin $\mathbf{0}$.

For the one-dimensional case, the BGA theory derives a normalized expected progress

$$\frac{1}{32} \leq \frac{E(1, r)}{r} \leq \frac{1}{16} \tag{2}$$

yielded by the BGA's mutation operator (Equation 1). For the n-dimensional case, the BGA theory applies a rotation of the coordinate system, since it is assumed that "the rotation does not change the average progress because we assume that the fitness function is spherically symmetric" and "therefore $E(n, r) = E(1, r)/n$" (Mühlenbein & Schlierkamp-Voosen, 1993, p. 31). Consequently, the normalized expected progress in n dimensions is estimated as follows:

$$\frac{1}{32n} \leq \frac{E(n, r)}{r} \leq \frac{1}{16n} \tag{3}$$

The critical point is, however, that in contradiction to the BGA assumption, a rotation *does change* the average progress. This is first shown with a simple example and then discussed on a more theoretical level. Let $\mathbf{x}$ be the point with coordinates $x_i = r/\sqrt{n}$, $1 \leq i \leq n$. It

certainly has distance r to the optimum, since $\|\mathbf{x}\| = \sqrt{\sum_i x_i^2} = \sqrt{n r^2/n} = r$. If applying a mutation to a randomly chosen variable x_i, the (one-dimensional) *nonnormalized* expected progress $E(1, x_i)$ is, according to Equation 2, bounded by

$$\frac{1}{32}\frac{r}{\sqrt{n}} \leq E(1, x_1) \leq \frac{1}{16}\frac{r}{\sqrt{n}} \tag{4}$$

Then, the normalized expected progress with respect to $\|\mathbf{x}\|$ is bounded by

$$\frac{1}{32\sqrt{n}} \leq \frac{E(n, \mathbf{x})}{r} \leq \frac{1}{16\sqrt{n}} \tag{5}$$

in contradiction to Equation 3. Note that Equation 5 is not to be divided by the factor n, since *all* components of $\mathbf{x}$ equal $r/\sqrt{n}$. Thus, compared to the case where the point $\mathbf{x}$ is moved to the first axis, the estimate (Equation 5) predicts an increase in the normalized expected progress, which is due to the linear superposition along the x_i-axes.

Let us now discuss the complications caused by using *small* mutation rates on a more theoretical level. For the explanation of this peculiar situation, the origin of the coordinate system is shifted to the location of the parent under consideration. Then, the probability p that a mutation places an offspring in a region Q is given by integrating the probability density $pd(x_1, \ldots, x_n)$ over that region Q:

$$p = \int_Q pd(x_1, \ldots, x_n)\, dQ \tag{6}$$

In the following, it is assumed that mutations are independently applied to each variable x_i and that the probability density $pd_i(x_i)$ is continuous. Recall from fundamental statistics that for any subregion $Q' \in R^k$, with $k < n$, the associated probability equals zero. For example, considering a two-dimensional search space, the probability would be given as

$$p = \iint_Q pd_x(x)\, pd_y(y)\, dy\, dx \tag{7}$$

Then, the probability for *any* line $l \in R^1 \subset R^2$ equals zero, since, for example,

$$\int_{x_a}^{x_a} \int_{y_0}^{y_1} pd_x(x)\, pd_y(y)\, dy\, dx = 0 \tag{8}$$

In other words, if applying mutations to all parameters with a probability $p_m = 1$, as is done in ESs, the probability is greater than zero for finite regions $Q \in R^n$ and vanishes for all subregions $Q' \in R^k$, $k < n$ (i.e., no special subregion is present).

This analysis, however, yields different results for mutation rates $p_m < 1.0$. In this case, not only a probability density $pd_{x_i}(x_i)$ is to be considered, but also a probability variable z_i, with which a mutation is applied to the variable x_i. A probability variable $z_i < 1$ gives rise to peculiar subregions $Q' \in R^k$, $k < n$, for which the probability is greater than zero. For example, for each coordinate axis (with respect to the parent under consideration), we arrive at

$$p = \int_{x_i^0}^{x_i^1} z_i \left(\prod_{j, j \neq i} 1 - z_j \right) pd(x_i)\, dx_i = \frac{c}{n} \int_{x_i^0}^{x_i^1} pd(x_i)\, dx_i \tag{9}$$

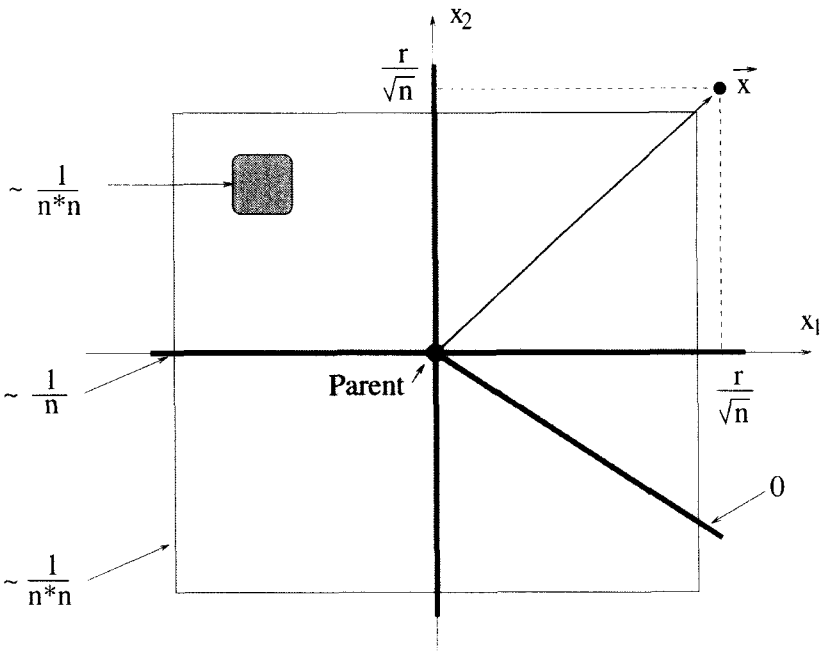

Figure 1. Different probability densities for applying a mutation when a small mutation rate $p_m = 1/n$ is assumed. See text for details.

However, for any other axis-parallel line and any slanted line, the probability is again zero. Under the condition that variables x_I, with $I = \{i_1, \ldots, i_k\}$, $k \geq 1$, are strictly mutated, Q is a region in R^k defined by the variables $x_{i_1}, \ldots, x_{i_k}$. In search space R^k, the probability $p(Q)$ for region Q is given by Equation 8. Hence, for a one-dimensional line in R^2, the probability $p(Q)$ vanishes. The probabilities for two-dimensional subregions on and off $x_i - x_j$ planes, etc., are similar to the one-dimensional case. In other words, using a mutation probability $p_m < 1$ leads to subregions $R^k, k < n$, for which the probability is strictly greater than zero, which in turn, results in a bias toward these subregions. The dimensionality k of a region $Q \in R^k$ is also called probability dimensionality.

With a commonly used mutation probability $p_m = 1/n$, $n(1/n)(1 - 1/n)^{(n-1)} \approx 1/e \approx 37\%$ of all mutations are biased toward the coordinate axes. If an implementation, as the BGA does, guarantees at least one mutation per offspring, the procedure places approximately 74% of all offspring on the coordinate axes. Then, approximately $(1/n^2)(1 - 1/n)^{(n-2)}\binom{n}{2} \approx 18\%$ offspring are placed onto the $x_i - x_j$, $i \neq j$ planes, and so forth. Only $1/(n \cdot n)$ percent of all offspring are placed in the n-dimensional search space R^n.

The theoretical results presented in this section are illustrated in Figures 1 and 2. Figure 2 illustrates a multimodal example that is further discussed in Section 4. This second example also demonstrates that a rotation can considerably influence the average progress yielded by small mutation rates. In summary, the use of small mutation rates (and not the implementation of a particular GA variant) not only causes difficulties for the theoretical analysis, but also makes the expected progress sensitive to a rotation of the coordinate system.

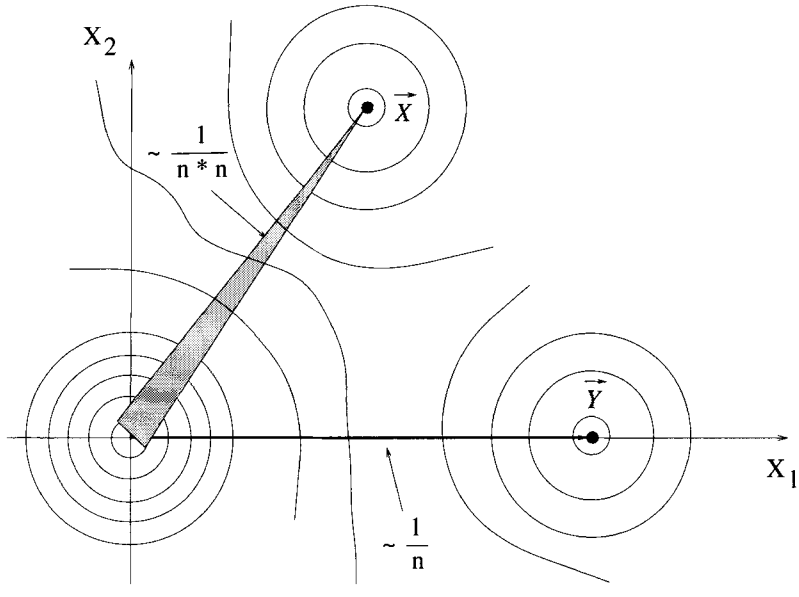

Figure 2. Scaling of the expected progress for different local optima when optimizing multimodal functions.

4. Rotational Sensitivity in Multimodal Functions

Clearly, any optimization/adaptation algorithm is not very interesting if it is limited to simple spherical fitness functions. Therefore, the research community has designed more complex multimodal functions, such as $f_6 - f_9$, which are widely used as benchmark tests for evolutionary algorithms. Evolutionary algorithms have been applied very successfully to these test functions, as is nicely illustrated by the BGA.

When applied to the multimodal test functions summarized in Table 1, the BGA yields the same $O(n \ln n)$ complexity as is derived for unimodal, spherically symmetrical test functions. With such encouraging results, one is tempted to empirically generalize the theory to multimodal test functions as well, even though the assumptions of the theory do not hold for these test cases. It is common practice to test an algorithm against test cases, for which the theory has not been designed, in order to examine further application areas and initiate the development of a better theory. This section discusses reasons why GAs are successful at these special test cases and why GAs can exhibit a drastic performance loss under a rotation of the coordinate system.

Figure 2 depicts a hypothetical multimodal function with the global optimum at the origin. Assume a parent at the local optimum **y**, which itself is located on the x_1-axis. In this situation, the normalized expected progress is clearly smaller than in the unimodal case, since the region Q that yields progress is smaller than in the unimodal case (compare also Equation 6). In the n-dimensional case, however, the normalized expected progress still scales with $1/n$; the rates of progress differ only in some constant factors.

Now, assume a parent at the local optimum **x** (Fig. 2, upper right). As can be seen from Figure 2, the expected normalized progress is zero when applying a mutation to only one variable x_1 or x_2. To obtain a progress $E(n, r) > 0$, mutations have to be simultaneously applied to both variables x_1 *and* x_2. In this case, the normalized expected progress rapidly decreases and scales with $1/n^2$. In the general case where *all* variables x_i have to be simulta-

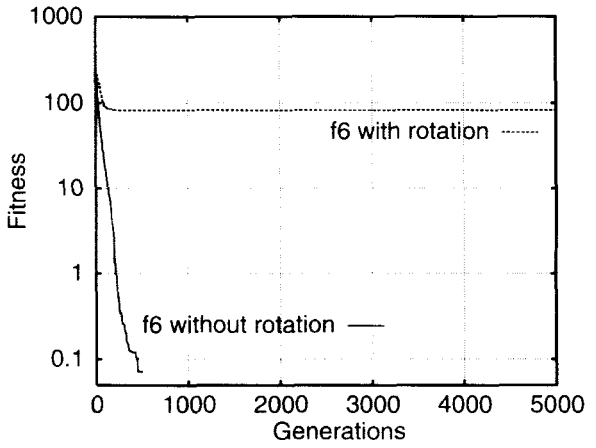

Figure 3. The BGA performance on Rastrigin's function f_6 with and without a rotation of the coordinate system.

neously mutated, the normalized expected progress obtained by mutations alone decreases to $E(n, r) \approx 1/n^n$, which is very small. The discussed performance loss associated with point X clearly illustrates that a rotation can influence the average progress significantly, as previously noted.

The performance loss discussed in this section on a theoretical level has already been observed in experiments presented in Salomon (1996a). In these experiments, a rotation of the coordinate system is applied. Rotation means that an orthogonal matrix **M** is applied so that $f(\mathbf{Mx})$ is to be calculated. The transformation matrix **M** is a pure rotation and, thus, does not change the function's structure. Results presented in Salomon (1996a) show that such a rotation results in a severe performance loss of the BGA and that this performance loss is observable even when using uniform recombination. A typical example, in which the BGA has to optimize function f_6, is depicted in Figure 3.

The performance loss observable under rotation suggests that current test suites are not adequate for a generalization of EA theory beyond the assumptions. It is encouraging that some algorithms yield the same computational complexity when applied to multimodal functions, but from an engineering point of view, the performance of a robust algorithm should not be degraded by a rotation. Current test suites are weak in that they do not provide problems with significant epistasis, which describes the interaction between different parameters. In most test functions, all parameters x_i are independent of each other. Only functions f_2 and f_8 contain epistasis to a certain degree. Due to the second term, epistasis in f_8 seems to be very high. It seems that all parameters influence each other. However, even the parameters of function f_8 can be optimized almost independently. Once a few parameters x_i are close to the global or a local optimum, the second term vanishes, and the function value is dominated by the first term, which is linear separable.

As has been shown in Salomon (1996b), the $O(n \ln n)$ complexity is due to the fact that the parameters of the test functions are independent of each other and that this independence can be exploited by a mutation rate $p_m = 1/n$. Furthermore, the analysis presented in Salomon (1996b) shows that *any* mutation operator yields an $O(n \ln n)$ complexity as long as it uses a mutation rate of $p_m = 1/n$ and a mutation range A that is greater than or equal to the search space when applied to *any* function with independent parameters.

The results on multimodal functions presented in this section severely limit (to a rather

"polar" set of functions) the generalizations (Mühlenbein & Schlierkamp-Voosen, 1993, pp. 46–47) that "these results [i.e., $O(n \ln n)$] can be predicted by the BGA theory," since "all test functions have a global structure that makes them easy to optimize," and "... the BGA solves some of the most popular test functions in global optimization in $O(n \ln n)$ function evaluations," which "demonstrates that these test functions are not as difficult to optimize as was formerly believed." Rather, these results indicate the need for new test functions that feature sufficient parameter interaction in order to generalize existing theory. At least rotation should become a standard procedure in EA benchmark test suites.

As indicated above, the expected progress $E(n, r)$ can decrease to $1/n^n$, which is equivalent to an $O(n^n)$ complexity. This very high complexity might be reduced if an additional recombination operator could exploit some global regularities. However, in the general case, recombination cannot relieve this problem. We cannot but conclude that the BGA cannot solve intrinsic NP-hard problems in a complexity smaller than $O(n^n)$. When applied to multimodal functions, the BGA yields fast convergence only if the parameters are independent of each other. In that case, a mutation probability $p_m = 1/n$ exploits this independence.

5. Self-Adaptive Mutation Operators

The importance of designing a good or appropriate mutation operator can be seen from the literature, which provides many examples of tailored mutation operators. For ESs, the self-adaptation of the mutation strength is very important, since a good rate of progress critically depends on an appropriate step size σ (Beyer, 1996; Rechenberg, 1973).

So far, GA theory has focused on optimal mutation rates (see, e.g., Bäck, 1993); self-adaptive mutation operators have been widely neglected. An interesting claim presented in Mühlenbein and Schlierkamp-Voosen, 1993, p. 33 is that the BGA mutation operator as defined in Equation 1 is optimal modulo a constant factor, since it does not require any dynamical adaptation mechanism in order to yield nearly optimal performance. This claim is the result of an analysis (Mühlenbein & Schlierkamp-Voosen, 1993) that compares different mutation operators.

The comparison consists of the BGA mutation operator using a mutation probability $p_m = 1/n$ and uniformly distributed as well as normally distributed mutations applied with a mutation probability $p_m = 1$. However, this comparison does not give any evidence that a particular mutation operator *or* a particular mutation probability is better than the other. Indeed, when assuming a mutation probability $p_m = 1/n$, one immediately obtains $E(n, r) = 0.25/n$ for uniformly distributed mutations and $E(n, r) = 0.1/n$ for normally distributed mutations, respectively. That means, however, that neither mutation operator requires any dynamical adaptation of any step size as long as a small mutation probability $p_m \approx 1/n$ is used.

Theoretical considerations presented in Salomon (1996b) suggest that any of these mutation operators yield an $O(n \ln n)$ complexity on widely used test functions. This result implies that the mutation operator can be designed according to the special requirements given in a particular optimization problem without losing the scaling behavior.

6. Discussion

This paper has discussed the use of spherical symmetry as a simplification of mathematical proofs. It turned out that this simplification is appropriate for many algorithms but seems inappropriate for algorithms using small mutation rates, since small mutation rates bias the

search toward a grid that is aligned with the coordinate system. In such cases, a rotation of the coordinate system does influence the average progress, and therefore, the assumption of spherical symmetry is not a simplification. Hence, in contradiction to some existing theory, the analysis of such biased search algorithms is much harder than currently realized. It should be kept in mind that despite its convenience, the assumption of spherical symmetry also limits the analysis to unimodal functions.

This paper has also argued that an extension of an existing theory by empirical tests on appropriate test suites is very important to obtain further progress. Current test suites, however, are not adequate, since they do not provide problems with sufficient epistatic parameter interaction. Therefore, at least the application of rotations should become standard procedure in evolutionary benchmark tests.

Furthermore, for multimodal test functions, it has been indicated that the expected progress yielded by small mutation rates might scale with $1/n^n$, yielding a computational complexity of $O(n^n)$. A key issue of future research should be the analysis of whether additional recombination operators can substantially reduce the $O(n^n)$ computational complexity of finding the optima of general multimodal functions.

Acknowledgments

This work was supported in part by a Human Capital and Mobility fellowship of the European Union, grant number ERBCHBICT941266. The author thanks Leo van Hemmen for editorial assistance and critical reading of earlier versions of the manuscript, and Rolf Pfeifer as well as Reinhard Riedl for detailed discussions, especially on mathematical details. Special thanks are due to the referees for their patience and efforts toward a clearer presentation.

References

Altenberg, L. (1995). The schema theorem and the Price's theorem. In L. D. Whitley & M. D. Vose (Eds.), *Foundations of genetic algorithms 3* (pp. 23–49). San Mateo, CA: Morgan Kaufmann.

Bäck, T. (1993). Optimal mutation rates in genetic search. In S. Forrest (Ed.), *Proceedings of the Fifth International Conference on Genetic Algorithms* (pp. 2–8). San Mateo, CA: Morgan Kaufmann.

Beyer, H.-G. (1993). Toward a theory of evolution strategies: Some asymptotical results from the $(1 \pm \lambda)$-theory. *Evolutionary Computation 1*(2):165–188.

Beyer, H.-G. (1995). Toward a theory of evolution strategies: The (μ, λ)-theory. *Evolutionary Computation 2*(4):381–407.

Beyer, H.-G. (1996). Toward a theory of evolution strategies: Self-adaptation. *Evolutionary Computation 3*(3):311–347.

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization and machine learning.* Reading, MA: Addison-Wesley.

Grefenstette, J. J., & Baker, J. E. (1989). How genetic algorithms work: A critical look at implicit parallelism. In J. D. Schaffer (Ed.), *Proceedings of the International Conference on Genetic Algorithms ICGA3* (pp. 20–27). San Mateo, CA: Morgan Kaufmann.

Luenberger, D. G. (1984). *Linear and nonlinear programming.* Reading, MA: Addison-Wesley.

Mühlenbein, H., & Schlierkamp-Voosen, D. (1993). Predictive models for the breeder genetic algorithm I. *Evolutionary Computation 1*(1):25–50.

Ostermeier, A., Gawelczyk, A., & Hansen, N. (1994). A derandomized approach to self-adaptation of evolution strategies. *Evolutionary Computation 2*(4):369–380.

Rechenberg, I. (1973). *Evolutionsstrategie.* Stuttgart: Frommann-Holzboog Verlag.

Salomon, R. (1996a). Performance degradation of genetic algorithms under coordinate rotation. In L. J. Fogel, P. J. Angeline, & T. Bäck (Eds.), *Proceedings of the Fifth Annual Conference on Evolutionary Programming*. Cambridge, MA: MIT Press.

Salomon, R. (1996b). Implicit independence assumptions: A notorious problem for genetic algorithms. In P. G. Anderson & K. Warwick (Eds.), *Proceedings of the International ICSC Symposia on Intelligent Industrial Automation and on Soft Computing (SOCO'96)* (pp. B 93–B 99). Canada/Switzerland: ICSC Academic Press.

Schlierkamp-Voosen, D., & Mühlenbein, H. (1994). Strategy adaptation by competing subpopulations. In Y. Davidor, H.-P. Schwefel, & R. Männer (Eds.), *Parallel Problem Solving from Nature (PPSN III)* (pp. 199–208). Berlin: Springer-Verlag.

Schwefel, H. P. (1995). *Evolution and optimum seeking*. New York: John Wiley.

Srinivas, M., & Patnaik, L. (1994). Genetic algorithms: A survey. *IEEE Computer 27*(6), 17–26.

Vose, M. D. (1991). Generalizing the notion of schema in genetic algorithms. *Artificial Intelligence 50*, 385–396.