

Finite Markov Chain Results in Evolutionary Computation: A Tour d’Horizon

Günter Rudolph

Universität Dortmund, Fachbereich Informatik XI, D-44221 Dortmund

Abstract. The theory of evolutionary computation has been enhanced rapidly during the last decade. This survey is the attempt to summarize the results regarding the limit and finite time behavior of evolutionary algorithms with finite search spaces and discrete time scale. Results on evolutionary algorithms beyond finite space and discrete time are also presented but with reduced elaboration.

Keywords: evolutionary algorithms, limit behavior, finite time behavior

1. Introduction

The field of evolutionary computation is mainly engaged in the development of optimization algorithms which design is inspired by principles of natural evolution. In most cases, the optimization task is of the following type: Find an element $x^* \in \mathcal{X}$ such that $f(x^*) \geq f(x)$ for all $x \in \mathcal{X}$, where $f : \mathcal{X} \rightarrow \mathbb{R}$ is the objective function to be maximized and \mathcal{X} the search set.

In the terminology of evolutionary computation, an *individual* is represented by an element of the Cartesian product $\mathcal{X} \times \mathcal{A}$, where \mathcal{A} is a possibly empty set collecting additional search state information. The *fitness* of an individual $(x, a) \in \mathcal{X} \times \mathcal{A}$ is given by the objective function value $f(x)$. A *population* consists of $n < \infty$ individuals and is thus an element of the product space $(\mathcal{X} \times \mathcal{A})^n$. During each iteration of an evolutionary algorithm the population is modified by a number of successive probabilistic transformations. At the beginning of each iteration the n members of the population are called the *parents* which produce $n' < \infty$ *offspring* by random variation: For each offspring two or more parents are selected and they are used to generate a preliminary offspring by *recombination*. Subsequently, each preliminary offspring is *mutated* at random yielding a final offspring. After all n' offspring have been produced in this manner the current population consist of $n + n'$ individuals. To keep the population at constant size

n , a selection method decides which parents and/or offspring will serve as parents in the next iteration. Now the process repeats until some stopping criterion is fulfilled.

Evidently, the resulting new population only depends on the state of the current population in a probabilistic manner. This fact, known as the *Markov property*, reveals that Markov processes are appropriate models for the probabilistic behavior of evolutionary algorithms. The well-developed theory of Markov processes may be divided into twelve sub-fields according to the following characteristics of the true process under consideration:

- The state space may be finite, denumerable or not denumerable.
- The evolution may happen in discrete or continuous time.
- The transition probabilities may depend on the time parameter or not.

Since the population size is finite the state space of the associated Markov process is finite (denumerable or not denumerable) if the search set is finite (denumerable or not denumerable). Therefore the Markov theory of evolutionary algorithms may be classified analogously. Most results are available for evolutionary algorithms with time-homogeneous transitions and (1) finite search space in discrete time, (2) not denumerable search space \mathbb{R}^ℓ in discrete as well as continuous time.

This survey will concentrate on the first class including the case of time-inhomogeneous transitions.

2. Limit Behavior in Finite Space and Discrete Time

Let $X_k = (X_{k,1}, X_{k,2}, \dots, X_{k,n})$ be the random population of size $n < \infty$ at step $k \geq 0$ and $F_k = \max\{f(X_{k,i}) : i = 1, \dots, n\}$ the best fitness value within the population at step $k \geq 0$. As soon as the random variable F_k attains the value of the global maximum f^* it is ensured that the population contains an individual representing the global solution of the maximization problem. Ideally, this event should happen after a finite number of steps with probability one and regardless of the initialization of the evolutionary algorithm. This desirable property can be formalized as follows.

Definition 2.1.

Let random variable $T = \min\{k \geq 0 : F_k = f^*\}$ denote the first hitting time of the global solution. An evolutionary algorithm is said to *visit the global optimum in finite time with probability one* if $\mathbf{P}\{T < \infty\} = 1$ regardless of the initialization. \square

Since it may be generally supposed that in practical implementations of evolutionary algorithms the best solution found in the course of the evolution is kept in memory, the property above guarantees that the global optimum will be found in finite time and never be lost although the population itself may loose the global solution once it was found. Thus, the property above alone does not exclude that the random sequence $(F_k : k \geq 0)$ oscillates freely without tending to a limit. There are evolutionary algorithms that show exactly such a behavior. But it can also be observed that there exist versions of evolutionary algorithms for which the random

sequence $(F_k : k \geq 0)$ “converges” to the limit f^* . Notice that the deterministic concept of the “convergence to the optimum” is not appropriate because the state transitions of an evolutionary algorithm are of stochastic nature. In order to clarify the exact semantic of a phrase like “the EA converges to the global optimum” one has at first to distinguish between the various modes of stochastic convergence [1].

Definition 2.2.

Let D_0, D_1, \dots be non-negative random variables defined on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$. The sequence $(D_k : k \geq 0)$ is said to *converge completely* to zero if $\sum_{k=0}^{\infty} \mathbf{P}\{D_k > \epsilon\} < \infty$ for any $\epsilon > 0$, to *converge with probability 1* (w.p.1) or *almost surely* (a.s.) to zero if $\mathbf{P}\{\lim_{k \rightarrow \infty} D_k = 0\} = 1$, to *converge in probability* to zero if $\mathbf{P}\{D_k > \epsilon\} = o(1)$ as $k \rightarrow \infty$ for any $\epsilon > 0$, and to *converge in mean* to zero if $\mathbf{E}[D_k] = o(1)$ as $k \rightarrow \infty$. \square

Complete convergence implies convergence with probability 1 while both convergence with probability 1 and convergence in mean implies convergence in probability. The reverse implications are wrong in general [1]. But if the sequence $(D_k : k \geq 0)$ is upper bounded by some finite constant then convergence in probability implies convergence in mean. With these definitions one can assign a rigorous meaning to the notion of the convergence of an evolutionary algorithm.

Definition 2.3.

Let $(X_k : k \geq 0)$ be the sequence of populations generated by some evolutionary algorithm and let $F_k = \max\{f(X_{k,1}), \dots, f(X_{k,n})\}$ denote the best objective function value of the population of size $n < \infty$ at generation $k \geq 0$. An evolutionary algorithm is said to *converge completely (with probability 1, in probability, in mean) to the global maximum* $f^* = \max\{f(x) : x \in \mathcal{X}\}$ of objective function $f : \mathcal{X} \rightarrow \mathbb{R}$ if the nonnegative random sequence $(D_k : k \geq 0)$ with $D_k = f^* - F_k$ converges completely (with probability 1, in probability, in mean) to zero. \square

At this point it should be noted that the property of visiting the global solution with probability one is a precondition for convergence but that the additional property of convergence does not automatically indicate any advantage with respect to finding the global solution.

2.1. Time-Homogeneous Transitions

In principle, the question whether some evolutionary algorithm will visit the global optimum in finite time and, if so, whether it will converge in some mode to the optimum or not may be answered by modeling the specific evolutionary algorithm under consideration as a finite Markov chain so that the existing powerful results from Markov chain theory [2, 3] can be exploited. Completely specified Markov chain models of a certain evolutionary algorithm were derived (apparently independently) in [4] and [5, 6]. But it is not necessary to build a quantitatively exact Markov model for each variant of an evolutionary algorithm in order to investigate the limit behavior. Instead, qualitative models are sufficient for this purpose. The idea to characterize the limit behavior of the evolutionary algorithm by the properties of the variation and selection operators was realized in [7]. This approach is adopted here. Actually, almost all results presented in this subsection are already given in [7, 8]. Subsequent publications offered some minor

extensions in case of specific combination of variation and selection operators [9, 10, 11, 12] or from a more abstract point of view [13, 14]. In retrospective, one may say that the intensive elaboration of these issues during the past years has finally led to simple proofs which do not require Markov chain theory any more. This is demonstrated next.

Let $(x_1, x_2, \dots, x_n) \in \mathcal{X}^n$ denote the population of parents. An offspring is produced as follows: At first, ρ parents are selected to serve as mates for the recombination process. This operation is denoted by

$$\text{mat} : \mathcal{X}^n \rightarrow \mathcal{X}^\rho$$

where $2 \leq \rho \leq n$. These individuals are then recombined by the procedure

$$\text{reco} : \mathcal{X}^\rho \rightarrow \mathcal{X}$$

yielding a preliminary offspring. Finally, a mutation via

$$\text{mut} : \mathcal{X} \rightarrow \mathcal{X}$$

yields the complete offspring. After all m offspring have been produced in this manner the selection procedure

$$\text{sel} : \mathcal{X}^k \rightarrow \mathcal{X}^n$$

decides which offspring and possibly parents ($k \geq n$) will serve as the new parents in the next iteration. Thus, a single iteration of the evolutionary algorithm can be described as follows:

$$\begin{aligned} \forall i \in \{1, \dots, m\} : x'_i &= \text{mut}(\text{reco}(\text{mat}(x_1, \dots, x_n))) \\ (y_1, \dots, y_n) &= \begin{cases} \text{sel}(x_{\pi(1)}, \dots, x_{\pi(q)}, x'_1, \dots, x'_m) & \text{(parents and offspring)} \\ \text{sel}(x'_1, \dots, x'_m) & \text{(only offspring)} \end{cases} \end{aligned}$$

where $1 \leq q \leq n$ and $\pi(1), \dots, \pi(n)$ is a permutation of the indices $1, \dots, n$ such that $f(x_{\pi(1)}) \geq f(x_{\pi(2)}) \geq \dots \geq f(x_{\pi(n)})$. This formulation includes selection methods that choose from the offspring and a subset of parents under the restriction that the best parent is a member of this subset.

After this operational description of evolutionary algorithms one is in the position of defining some assumptions about the properties of the variation and selection operators:

- (A₁) $\forall x \in (x_1, \dots, x_n) : \mathbf{P}\{x \in \text{reco}(\text{mat}(x_1, \dots, x_n))\} \geq \delta_r > 0$.
- (A₂) For every pair $x, y \in \mathcal{X}$ there exists a finite path x_1, x_2, \dots, x_k of pairwise distinct points with $x_1 = x$ and $x_k = y$ such that $\mathbf{P}\{x_{i+1} = \text{mut}(x_i)\} \geq \delta_m > 0$ for all $i = 1, \dots, k-1$.
- (A'₂) For every pair $x, y \in \mathcal{X}$ holds $\mathbf{P}\{y = \text{mut}(x)\} \geq \delta_m > 0$.
- (A₃) $\forall x \in (x_1, \dots, x_k) : \mathbf{P}\{x \in \text{sel}(x_1, \dots, x_k)\} \geq \delta_s > 0$.
- (A₄) Let $v_k^*(x_1, \dots, x_k) = \max\{f(x_i) : i = 1, \dots, k\}$ denote the best fitness value within a population of k individuals ($k \geq n$). The selection method fulfills the condition

$$\mathbf{P}\{v_n^*(\text{sel}(x_1, \dots, x_k)) = v_k^*(x_1, \dots, x_k)\} = 1.$$

Assumption (A₁) means that every parent may be selected for mating and is not altered by recombination with minimum probability $\delta_r > 0$. Assumption (A₂) ensures that every individual can be changed to an arbitrary other individual by a finite number of successive mutations, whereas assumption (A'₂) asserts the same but within a single mutation. Assumption (A₃) guarantees that every individual competing for survival may survive with minimum probability $\delta_s > 0$, whereas assumption (A₄) makes sure that the best individual among the competitors in the selection process will survive with probability one.

Theorem 2.1.

If the assumptions (A₁), (A₂), and (A₃) are valid then the evolutionary algorithm visits the global optimum after a finite number of iterations with probability one, regardless of the initialization. If assumption (A₄) is valid additionally and the selection method chooses from parents as well as offspring then the evolutionary algorithm converges completely and in mean to the global optimum regardless of the initialization.

Proof: Let $\mathcal{X}^* = \{x \in \mathcal{X} : f(x) = f^*\}$ be the set of globally optimal solutions. Owing to assumption (A₂) there exists a finite path from an arbitrary $x \notin \mathcal{X}^*$ to some $x^* \in \mathcal{X}^*$ that can be traversed by successive mutations. Let k_x be the length of the shortest path between $x \notin \mathcal{X}^*$ to the set \mathcal{X}^* and $k^* = \max\{k_x : x \notin \mathcal{X}^*\}$.

Now consider an arbitrary parent x of some population. Assumption (A₁) ensures that this parent passes the recombination process without being altered at least with probability $\delta_r > 0$. The probability that this preliminary offspring transitions to the next point of the shortest path towards \mathcal{X}^* by mutation is guaranteed to be at least $\delta_m > 0$ by assumption (A₂). Owing to assumption (A₃) this offspring will survive the selection process at least with probability $\delta_s > 0$. Thus, the probability that parent x transitions to a parent representing the next point on the shortest path to \mathcal{X}^* is at least $\delta_r \cdot \delta_m \cdot \delta_s > 0$. A k_x -fold repetition of this argumentation shows that the probability of a transition from $x \notin \mathcal{X}^*$ into the set \mathcal{X}^* at iteration k_x is at least $(\delta_r \cdot \delta_m \cdot \delta_s)^{k_x-1} \cdot \delta_r \cdot \delta_m > 0$. Therefore it can be asserted that the probability of visiting a globally optimal solution after k^* iterations is at least $\delta = (\delta_r \cdot \delta_m \cdot \delta_s)^{k^*-1} \cdot \delta_r \cdot \delta_m > 0$ regardless of the true instantiation of $x \notin \mathcal{X}^*$. Consequently, the probability that a globally optimal solution has not been found after k iterations is at most $(1 - \delta)^{\lfloor k/k^* \rfloor}$ which converges exponentially fast to zero as $k \rightarrow \infty$. This immediately implies $\mathbf{P}\{T < \infty\}$, i.e., a global optimum will be visited for the first time after a finite number of iterations with probability one. This proves the first part of the theorem.

As for the second part, suppose that the global optimum was found for the first time at iteration k_0 . Assumption (A₄) guarantees that this offspring will be a parent of the next iteration, since neither an old parent nor another offspring can be better than this one. Thus,

$$\mathbf{P}\{F_k < f^*\} = \mathbf{P}\{f^* - F_k > 0\} = \mathbf{P}\{D_k > 0\} \leq (1 - \delta)^{\lfloor k/k^* \rfloor} \rightarrow 0$$

as $k \rightarrow \infty$. This proves convergence in probability. Since

$$\sum_{k=0}^{\infty} \mathbf{P}\{D_k > 0\} \leq \sum_{k=0}^{\infty} (1 - \delta)^{\lfloor k/k^* \rfloor} \leq \frac{1}{1 - (1 - \delta)^{1/k^*}} < \infty$$

one obtains even complete convergence to the global optimum. Finally, convergence in mean follows from convergence in probability and the fact that the sequence $(F_k : k \geq 0)$ is bounded. \square

A variation of this result is given next.

Corollary 2.1.

Theorem 2.1 remains valid if the assumptions (A_1) , (A_2) , and (A_3) are replaced by assumption (A'_2) . \square

As can be seen from assumptions (A_2) or (A'_2) , the reachability of the optimum is guaranteed solely by the properties of the mutation operators. The potential positive effects of recombination are completely neglected. Notice that an EA without recombination ($\delta_r = 1$) will always visit the optimum, whereas an EA without mutation but with a usual recombination operator does not have this guarantee [10]. This observation might have been the reason why Evans [15] suggested the following modified EA “without” mutation.

Let $\mathcal{X} = \{0, 1\}^\ell$. The population consist of $n + 2$ parents. Two distinguished parents are protected, i.e., they pass through all stages of the life cycle with probability one, but they participate in the mating and recombination process. In the initial population, the first protected individual is chosen at random and the second protected individual is set to the binary complement of the first one. Suppose that the recombination process employs *uniform crossover* of two parents, i.e., each entry of the offspring’s bit vector is independently chosen either from the first or from the second parent with the same probability. Moreover, assume that the two protected individuals can be chosen for mating with some minimum probability $\delta_c > 0$. If this event occurs then the optimal solution is assembled with probability $2^{-\ell} > 0$. Notice that this is equivalent to generating an individual uniformly at random. As a consequence, the probability to find the optimum within one step is at least $\delta_c 2^{-\ell} > 0$ and it is guaranteed that the optimum will be visited in finite time with probability one. Needless to say, the concept of protected individuals in nothing more than a disguised method of permitting mutations.

The next result complements and partially sharpens Theorem 1 and Corollary 1.

Theorem 2.2.

An evolutionary algorithm visits the global optimum infinitely often if assumption (A'_2) or the assumptions (A_1) , (A_2) , and (A_3) are valid. If the selection method only chooses from the offspring then the sequence $(F_k : k \geq 0)$ will not converge to the global optimum, even if assumption (A_4) is valid.

Proof: The proof of Theorem 1 has already shown that conditions (A'_2) or (A_1) , (A_2) , (A_3) are sufficient to find the optimum in finite with probability one. Assume that the optimum has been visited at step k_1 for the first time and that the population loses all optimal individuals at step $k_2 > k_1$. Then the same assumptions guarantee that the optimum will be found again in finite time with probability one at step $k_3 > k_2$, and possibly lost again and found again and so forth *ad nauseam*. This proves the first part of the theorem.

As for the second part, let the selection method only choose from the offspring and assume that the optimum is contained in the current population (possibly several times). Even if all

optimal parents pass through the recombination process without being altered, assumption (A₂) as well as (A'₂) ensures that *each* individual is mutated with some minimum probability. As a consequence, there is a minimum probability $\delta_L > 0$ that all individuals being optimal before mutation are mutated to non-optimal individuals. Since the probability of this event is strictly bounded from zero, the population will loose the optimum in finite time with probability one. But the optimum will be found again, and lost again ... in short: The sequence $(F_k : k \geq 0)$ of the best fitness value within a population at step $k \geq 0$ oscillates forever preventing the property of stochastic convergence. \square

The assumptions and their implications presented so far are valid for the vast majority of evolutionary algorithms with finite search space and time-homogeneous transitions. But every conference on evolutionary computation gives birth to new versions of evolutionary algorithms that do not necessarily fit in this framework. In this case, the assumptions and proofs must be adapted. For example, it was recently shown that Corollary 1 can be generalized to situations in which the set of fitness values is only partially in lieu of totally ordered [16]. Actually, only the assumptions regarding the selection methods were generalized.

2.2. Time-Inhomogeneous Transitions

The development of evolutionary algorithms with time-inhomogeneous transitions was motivated by the observation that a specific popular evolutionary algorithm fulfilling the preconditions of Theorem 2 did apparently not converge, and by the fact that there existed convergence proofs for stochastic optimization algorithms with time-inhomogeneous transitions [4, 17]. The previous subsection has already disclosed the reason for non-convergence of the sequence $(F_k : k \geq 0)$. The optimum is found and lost infinitely often. Needless to say, stochastic convergence with time-inhomogeneous transitions also requires the precondition that the optimum will be found in finite time with probability one. But if the optimum is not guaranteed to stay in the population, then it is necessary that it will be found again. In order to prevent everlasting oscillation of the sequence $(F_k : k \geq 0)$ it must be ensured that the event of finding the optimum happens infinitely often whereas the event of loosing the optimum happens only finitely often. Actually, this is the decisive property that must be shown when proving global convergence of an evolutionary algorithm—may the transitions be time-homogeneous or not.

The Borel-Cantelli Lemma and its extension are actually sufficient to establish conditions for stochastic convergence of an evolutionary algorithm. Since EAs have the Markov property the condition is as follows: Let α_k be the probability of loosing the optimum and β_k the probability of finding the optimum at step k . If

$$\sum_k \alpha_k < \infty \quad \text{and} \quad \sum_k \beta_k = \infty$$

then the event of loosing the optimum happens finitely often with probability 1 whereas the probability of visiting the optimum happens infinitely often with probability one. Thus, the

probability of loosing the optimum must decrease faster than the probability of finding the optimum.

Davis [4, 17] tried to establish this property by introducing a time-dependent decreasing schedule for the probability of mutating an individual of an evolutionary algorithm with selection from offspring only. Clearly, a decreasing mutation probability leads to decreasing sequences of both probabilities α_k and β_k . But there is problem: The rate of decrease is of the same order for both sequences. As a consequence, both sequences either converge or diverge, i.e., either the optimum is not found with probability one or the sequence $(F_k : k \geq 0)$ oscillates forever. This observation reveals that the selection mechanism must be time-dependent.

Mahfoud and Goldberg [18] used time-homogeneous mutations fulfilling assumption (A'_2) and adopted the time-inhomogeneous selection method as it is known from *simulated annealing*. Assumption (A'_2) ensures $\beta_k = \beta > 0$ and hence the divergence of the sequence $(\beta_k : k \geq 0)$ while the simulated annealing like selection method yields the convergence of the sequence $(\alpha_k : k \geq 0)$.

Cerf employed the Freidlin/Wentzel theory of dynamical perturbed systems to prove global convergence for time-dependent schedules for recombination, mutation, and selection operators [19, 20] while Suzuki [21], Lozano et al. [22], as well as He and Kang [23] came to similar results via Markov chain theory. Since most of these results are specialized to certain combinations of variation and selection operators, it is refrained from reproducing all assumptions here.

3. Finite Time Behavior in Finite Space and Discrete Time

The examination of the finite time behavior of evolutionary algorithms cannot be treated in the same general manner as it is possible for the limit behavior. Apart from the problem type under consideration, the choice and parameterization of the variation and selection operators have a significant impact on the finite time behavior of evolutionary algorithms. As a consequence, the theoretical studies are restricted to certain problem classes and simple evolutionary algorithms yet.

Most results are available for maximizing real-valued fitness functions with domain $\mathcal{X} = \mathbb{B}^\ell = \{0, 1\}^\ell$. This problem is called the *pseudo-boolean optimization problem* [24] and it is known to be NP-hard in general [25]. But there are classes of pseudo-boolean optimization problems with reduced computational complexity.

Here, it is assumed that the time of calculating the fitness value $f(x)$ for $x \in \mathbb{B}^\ell$ is bounded by a polynomial in ℓ . Since the population size is finite, the number of fitness evaluations is an appropriate measure to assess the efficiency of an evolutionary algorithm. Evidently, for this purpose one needs a stopping rule that indicates the termination of the stochastic process. Unless there is a efficiently computable criterion to decide whether the optimum has been found or not, one has to define another stopping rule that may depend on the entire history of the process. Let \mathcal{H}_k contain the information available to the process until iteration $k \geq 0$. Then the stopping rule $\tau(\mathcal{H}_k)$ indicates termination at step $k \geq 0$ if it evaluates to 1, and continuation of the process if it evaluates to 0. Notice that a stopping rule induces a random stopping time

$S = \min\{k \geq 0 : \tau(\mathcal{H}_k) = 1\}$ in general. After these preparations one is in the position to offer a criterion for *efficient* evolutionary algorithms.

Definition 3.1.

Let $F_k^* = \max\{F_{k-1}^*, F_k\}$ for $k \geq 1$ and $F_0^* = F_0$ denote the best fitness value found until iteration $k \geq 0$. An evolutionary algorithm is said to be efficient for a problem class \mathcal{C} if $\mathbb{E}[S] \leq \text{poly}_1(\ell)$ and $\mathbb{P}\{F_S^* = f^*\} \geq 1/\text{poly}_2(\ell)$ for every instance of \mathcal{C} , where $\text{poly}_1(\cdot)$ and $\text{poly}_2(\cdot)$ are two polynomial functions of the problem dimension ℓ . \square

The association of the term “efficient” with this criterion is justified by the algorithmic technique known as *probability amplification* or *probability boosting* [26]. Suppose there exists an efficient EA for some problem class with $1/\text{poly}_2(\ell) \leq \mathbb{P}\{F_S^* = f^*\} < 1$. The probability that the optimal solution is not found after r independent runs (possibly in parallel) of the EA is at most $(1 - 1/\text{poly}_2(\ell))^r$. The choice $r = k \cdot \text{poly}_2(\ell)$ with $k \in \mathbb{N}$ leads to a total expected sequential runtime $r \mathbb{E}[S]$ which remains polynomial in ℓ , whereas the probability of not finding the optimum in r runs decreases exponentially in k .

If $\mathbb{E}[S] \leq \text{poly}_1(\ell)$ and $\mathbb{P}\{F_S^* = f^*\} = 1$, then the evolutionary algorithm always gives the optimal solution. The only variation from one run to another is its random running time, whose distribution has to be studied. Theoretical work regarding evolutionary algorithms with random stopping time is rarely available. Hulin [27] suggested a stopping rule that is optimal in a certain Bayesian sense, but the goodness of this stopping rule in the sense of Definition 3.1 was studied empirically only. Aytug and Koehler [28] developed bounds on the number of iterations s required to achieve the validity of the inequality $\mathbb{P}\{F_s^* = f^*\} \geq \alpha$ for some prescribed $\alpha > 0$. Their final results were designated to be trivial since they did not take into consideration the function to be maximized. This omission finally led to an optimal parameterization under which the EA degenerates to pure random search with population size $n = 1$. As a consequence, the bound on s was not polynomial but exponential in the problem dimension ℓ . This observation reveals the necessity of restricting the analysis to certain problem classes whose special properties can be exploited in order to achieve non-trivial results.

As in the work of Aytug and Koehler [28] let the stopping rule not depend on the history of the evolutionary process. More specifically, the EA is stopped after a prescribed number s of iterations so that the random stopping time S degenerates to the constant s with $\mathbb{E}[S] = s$ and $\mathbb{V}[S] = 0$. Moreover, it is assumed that the EA only employs mutation and elitist selection. Notice that this assumption implies $F_k = F_k^*$ for all $k \geq 0$. Two types of mutations will be considered here:

- (M₁) An individual $x \in \mathbb{B}^\ell$ is mutated by drawing an index uniformly at random and inverting the associated entry in x .
- (M₂) An individual $x \in \mathbb{B}^\ell$ is mutated by inverting each entry in x independently with probability $p \in (0, 1)$.

Originally, the results to be presented shortly have been derived for an evolutionary algorithm with population size $n = 1$, which is usually termed the $(1 + 1)$ -EA. But it is easy to see that

an EA with larger population size cannot be worse than the $(1+1)$ -EA with respect to the number of iterations. Moreover, the (positive) results are based on bounds on the expected first hitting time $\mathbb{E}[T]$. The relationship between T and the criterion of Definition 3.1 is established as follows:

Since $\mathbb{P}\{F_s^* = f^*\} = \mathbb{P}\{T \leq s\}$ for every $s \geq 0$, one may use the Markov inequality to obtain $\mathbb{P}\{F_s^* \neq f^*\} = \mathbb{P}\{T > s\} \leq \mathbb{E}[T]/s$. Assume it can be shown that $\mathbb{E}[T] \leq \text{poly}_1(\ell)$ for every instance of a specific problem class \mathcal{C} , where the bound $\text{poly}_1(\ell)$ is explicitly known. If the stopping time is set to $s = c \text{poly}_1(\ell)$ with $c \geq 1 + 1/\text{poly}_2(\ell)$ for an arbitrary polynomial $\text{poly}_2(\ell)$, then

$$\mathbb{P}\{F_s^* \neq f^*\} \leq \frac{\mathbb{E}[T]}{s} = \frac{\mathbb{E}[T]}{c \text{poly}_1(\ell)} \leq \frac{\text{poly}_1(\ell)}{c \text{poly}_1(\ell)} = \frac{1}{c}$$

and finally

$$\mathbb{P}\{F_s^* = f^*\} \geq 1 - \frac{1}{c} \geq \frac{1}{\text{poly}_2(\ell) + 1}$$

for every instance of problem class \mathcal{C} . Thus, the development of a polynomial upper bound for $\mathbb{E}[T]$ is sufficient for proving the efficiency of the evolutionary algorithm for a specific problem class.

Definition 3.2.

A function $f : \mathbb{B}^\ell \rightarrow \mathbb{R}$ is said to be *modular* if $f(x \wedge y) + f(x \vee y) = f(x) + f(y)$ for all $x, y \in \mathbb{B}^\ell$. \square

It is easy to see that a function $f : \mathbb{B}^\ell \rightarrow \mathbb{R}$ is modular if and only if it is linear, i.e., $f(x) = c_0 + \sum_{i=1}^\ell c_i x_i$ with $c_i \in \mathbb{R}$. For example, the fitness function $f(x) = \ell - H(x, x^*)$ based on the Hamming distance $H(x, x^*)$ between some $x \in \mathbb{B}^\ell$ and a target pattern $x^* \in \mathbb{B}^\ell$ is linear and therefore modular. The special case with target vector $x^* = (1, \dots, 1)' \in \mathbb{B}^\ell$ is known known as the ‘counting ones problem.’ Bäck [29] and independently Mühlenbein [30] made the first steps towards an upper bound on the expected first hitting time. While Bäck derived the complete finite Markov chain model for the $(1+1)$ -EA with mutations of type (M_2) , Mühlenbein developed an approximation for the expected first hitting time under simplifying assumptions. Later it was shown [13] that Mühlenbein’s approach can be combined with Bäck’s Markov chain model to achieve the upper bound $\mathbb{E}[T] \leq \ell(\log \ell + 1) \exp(1)$ under mutations of type (M_2) with $p = 1/\ell$. In general, the results are as follows:

Theorem 3.1.

Let the fitness function $f : \mathbb{B}^\ell \rightarrow \mathbb{R}$ be modular. If the evolutionary algorithm only uses mutation and elitist selection then

- (a) $\mathbb{E}[T] \geq \ell \log \ell$ under (M_1) .
- (b) $\mathbb{E}[T] = \Omega(\ell \log \ell)$ under (M_2) with $p = 1/\ell$.
- (c) $\mathbb{E}[T] \leq \ell(\log \ell + 1)$ under (M_1) .
- (d) $\mathbb{E}[T] = O(\ell \log \ell)$ under (M_2) with $p = 1/\ell$.

Proof: For part (b) and (d) see [31], for part (c) see [13, p. 98]. As for part (a), the basic argument from the proof of part (b) may be used: In the worst case, the algorithm starts at

a point that is the binary complement of the optimum. Thus, every bit in the vector must be inverted at least once in order to reach the optimum by successive mutations.

Assume that $i < \ell$ different bits are already inverted. Then there are $\ell - i$ different bits that needs to be inverted. Since the index of the entry to be mutated is chosen uniformly at random, the probability that one of these $\ell - i$ bits will be inverted is $(\ell - i)/\ell$. The expected time that such an event occurs is just $\ell/(\ell - i)$. As soon as this event has happened, $i + 1$ bits are inverted at least once and the argumentation repeats until all ℓ bits are inverted at least once. As a consequence, the expected number of mutations required to invert each bit at least once is given by

$$\sum_{i=0}^{\ell-1} \frac{\ell}{\ell - i} = \ell \sum_{i=1}^{\ell} \frac{1}{i} \geq \ell \log \ell$$

which proves part (a) of the theorem. \square

Thus, modular functions can be efficiently maximized by an evolutionary algorithm with both versions of mutations. The next class of pseudo-boolean functions strictly includes the class of modular functions.

Definition 3.3.

A function $f : \mathbb{B}^\ell \rightarrow \mathbb{R}$ is said to be *pseudo-modular* if simultaneously

$$\begin{aligned} \min\{f(x), f(y)\} &\leq \max\{f(x \wedge y), f(x \vee y)\} \\ \max\{f(x), f(y)\} &\geq \min\{f(x \wedge y), f(x \vee y)\} \end{aligned}$$

for all $x, y \in \mathbb{B}^\ell$. \square

A non-modular instance of this class is the pseudo-boolean function

$$f(x) = \sum_{i=1}^{\ell} \prod_{j=1}^i x_j. \quad (1)$$

It is easily seen [13, p. 114] that $f(x \wedge y) = \min\{f(x), f(y)\}$ and $f(x \vee y) \geq \max\{f(x), f(y)\}$ and there are pairs (x, y) for which the inequality is strict. Thus,

$$\min\{f(x), f(y)\} = f(x \wedge y) \leq \max\{f(x), f(y)\} \leq f(x \vee y)$$

and hence

$$f(x \wedge y) = \min\{f(x \wedge y), f(x \vee y)\} \leq \max\{f(x \wedge y), f(x \vee y)\} = f(x \vee y)$$

which immediately implies the pseudo-modularity of function (1). The general problem has not been studied yet. But for this specific instance the following result is available.

Proposition 3.1.

The expected first hitting time of the $(1 + 1)$ -EA with fitness function (1) is bounded by

- (a) $\mathbb{E}[T] \leq \ell^2$ under (M_1) ,
- (b) $\mathbb{E}[T] \leq \ell^2 (\exp(1) - 1)$ under (M_2) with $p = 1/\ell$.

Proof: See [13], p. 105 for part (a) and p. 102 for part (b). \square

Hammer et al. [32] presented a hierarchy of classes of pseudo-boolean functions that strictly include each other. Therefore it may be useful to investigate the most general class of this hierarchy since the existence of a polynomial bound on the expected first hitting would imply that this bound is valid for all problem classes within this hierarchy. This top hierarchy class, which includes injective pseudo-modular functions, is characterized as follows.

Definition 3.4.

An injective function $f : \mathbb{B}^\ell \rightarrow \mathbb{R}$ is called *unimax* if there is a unique locally maximizing point $x^* \in \mathbb{B}^\ell$.

In general, an element $x^* \in \mathbb{B}^\ell$ is a locally maximizing point of a pseudo-boolean function $f : \mathbb{B}^\ell \rightarrow \mathbb{R}$ if $f(x^*) \geq f(x)$ for all $x \in \mathbb{B}^\ell$ with Hamming distance $H(x, x^*) = 1$ [24, p. 135]. If the function $f(\cdot)$ is injective then the inequality is actually always strict. This leads to subtle differences between the various definitions especially in the case of pseudo-boolean functions that are termed *unimodal*. Even worse, these subtle differences in the definitions may have a huge impact on the extent of the resulting class and its complexity. For example, Rudolph [33, 13] shows that the definition of unimodality used in [34, 35] includes the class of surjective boolean functions $f : \mathbb{B}^\ell \rightarrow \mathbb{B}$ for which the unique satisfying truth assignment is sought for. In contrast, the definitions given in [32, 33, 13] (albeit slightly different) ensure that for every $x \in \mathbb{B}^\ell$ there exists a path along adjacent vertices of the hypercube \mathbb{B}^ℓ with increasing function values leading to the global optimum. This property is also valid for unimax functions, even if they are not injective.

In any case, the problem of maximizing unimax functions is challenging task for it is known [36] that the decision version of this problem is in $\text{NP} \cap \text{CO-NP}$ but it is unknown whether the the optimization problem can be solved in polynomial time or not. The difficulty associated with this class is caused by the existence of problem instances for which the only increasing path along the vertices of \mathbb{B}^ℓ may be exponentially long. Examples for such instances have been constructed in [32, 35]. Although the general case remains an open field of research there is a noteworthy result.

Proposition 3.2.

Let the unimax fitness function $f : \mathbb{B}^\ell \rightarrow \mathbb{R}$ be the long “Root2”-path problem given in [35]. The expected first hitting time of the $(1 + 1)$ -EA can be bounded by

- (a) $\mathbb{E}[T] \geq 3 \cdot \ell \cdot 2^{(\ell-1)/2} - 2\ell$ under (M_1) ,
- (b) $\mathbb{E}[T] \leq (\ell^3 - \ell) \exp(1)/2$ under (M_2) with $p = 1/\ell$,

if the EA starts at the bottom of the increasing path.

Proof: See [33] or [13], section 5.1.2.2. □

Evidently, it is not hopeless to tackle unimax problems with evolutionary algorithms provided they employ mutations of type (M_2) with $p = 1/\ell$. The poor performance under mutations of type (M_1) is due to the fact that the $(1 + 1)$ -EA only can move along the path of adjacent vertices, whereas type (M_2) offer the chance of taking shortcuts along the path by inverting several bits simultaneously. In fact, since an exponentially long path must be folded several

times to fit into the box \mathbb{B}^ℓ it may be speculated that the bit patterns of elements on the path are “relatively regular.” This regularity might be exploited by a population of individuals that produce offspring by some kind of crossover/recombination. Horn et al. [35] provide empirical evidence that this conjecture is not too far-fetched.

Other hierarchies of classes of pseudo-boolean function are presented in Crama [37]. Here, the top hierarchy classes are known to be solvable in polynomial time. But already at a low level of these hierarchies there is a problem class that reveals some limitations of evolutionary algorithms.

Definition 3.5.

A function $f : \mathbb{B}^\ell \rightarrow \mathbb{R}$ is called *almost-positive* if the coefficients of all nonlinear terms are non-negative. \square

An instance of this problem class is the pseudo-boolean function

$$f(x) = \ell - \sum_{i=1}^{\ell} x_i + (\ell + 1) \prod_{i=1}^{\ell} x_i. \quad (2)$$

Proposition 3.3.

The expected first hitting time of the $(1 + 1)$ -EA with fitness function (2) can be bounded by

- (a) $\mathbb{E}[T] = \infty$ under (M_1) , unless being started at the optimum;
- (b) $\mathbb{E}[T] \geq \ell^\ell$ under (M_2) with $p = 1/\ell$ with worst starting point;
- (c) $\mathbb{E}[T] \geq [(\ell + 1)^\ell - 1]/2^\ell$ under (M_2) with $p = 1/\ell$ and random starting point.

Proof: See [13], pp. 116–117. \square

It should be noted that lower bounds on the expected first hitting time that are exponential in ℓ do generally not imply the absence of a bound $\mathbb{P}\{F_s^* = f^*\} \geq 1/\text{poly}(\ell)$ for some stopping time $s = \text{poly}(\ell)$. For this particular problem, however, this implication is unfortunately true.

Two points in conclusion: First, the behavior of evolutionary algorithms can be studied numerically provided that the transition matrix of the associated Markov chain is known [38, 39]. But this approach is manageable only for moderately large populations and problem dimensions because of the exponentially growing transition matrices and the inevitable rounding errors during the calculations. Moreover, this approach does not lead to theoretical results as they were presented here. The size of the state space and hence the transition matrices can be decreased considerably by a technique that is called *lumping* or *grouping* of states [2]. This technique in conjunction with a modification of the original Markov chain to a simpler Markov chain with either better or worse performance than the original one is extensively exploited in the proofs cited in this section.

Second, Vitányi [40] suggested to seek for evolutionary algorithms whose associated Markov chain is *rapidly mixing*, i.e., Markov chains that quickly approach their stationary distribution. Assume that the Markov chain reaches stationarity up to some small $\epsilon > 0$ in a number of iterations that is a polynomial in ℓ . If the probability of being in an optimal state is larger than $1/\text{poly}(\ell)$ then the associated evolutionary algorithm is efficient in the sense of Definition

3.1. Although the toy problem dedicated to demonstrate this “paradigm” is not convincing (since pure random search works equally well for this problem) the idea itself deserves closer examination.

4. Beyond Finite State Space and Discrete Time

4.1. Infinite Population Size

Even if the search set \mathcal{X} is finite, the state space of the Markov chain becomes infinitely large as soon as the population is assumed to be infinitely large. Since the point of view from an infinitely large population implies that the *de facto* probabilistic trajectory of the evolutionary algorithm is appropriately described by the iteration of the expected one-step transitions, it is clear that these models may show a qualitative behavior that is completely different to the true behavior of an evolutionary algorithm with a finite population. Vitányi [40] offers a nice example of this fact. But under certain conditions, these models can lead to considerable insight into the dynamics of evolutionary algorithms [41].

The popularity of infinite population models with finite search space [5, 42, 43, 44, 45, 46, 47, 48, 49, 50] is probably due to the fact that an infinite population in conjunction with the smoothing effect of the expectation operator leads to a non-linear *deterministic* dynamical system with *continuous* state space. Although such dynamical systems are easier to characterize than those with probabilistic and discontinuous state transitions, much work has been devoted to the determination of these systems’ fixed points and their stability [51, 52, 53, 54, 55, 56, 57].

The models taken from quantitative genetics [58, 59] essentially also assume infinite populations. And not surprisingly, infinite population models are also present in non-finite search spaces [60, 61].

4.2. Continuous Space and Discrete Time

In case of search space $\mathcal{X} = \mathbb{R}^\ell$, the convergence theory of stochastic optimization algorithms resembling a $(1+1)$ -EA dates back to the mid-60s [62] and was further developed in [63, 64, 65, 66, 67] under various assumptions. Interestingly, these authors did not use the theory of Markov processes for their proofs. Instead, their proofs are solely based on the Borel-Cantelli Lemma and its extensions. The same observation can be made in the proof of Born [68], who developed a population-based evolutionary algorithm with “genetic load” and proved its convergence to the optimum with probability one. The analysis of population-based algorithms relying on mutation and selection from a Markovian point of view was exhaustively treated in [69, chapter 5]. The extension of this theory in case of additional recombination was considered in [70, 71].

The modeling and analysis of non-elitist evolutionary algorithms in terms of supermartingales (another family of stochastic processes) was suggested in [72] and extensively used in [13]; even certain EAs for multi-objective optimization can be analyzed in this manner [73]. Although this approach only offers sufficient conditions for global convergence with probability one it may

have a pleasant by-product: Under certain conditions one immediately obtains bounds on the average convergence velocity. Actually, it often suffices to determine the average convergence velocity. For, if the EA converges in mean with a certain minimum rate then the supermartingale convergence theorem implies convergence with probability one (see [13] for a more detailed discussion).

In the light of this observation, the much earlier established results regarding average convergence rates are graded up instantly. For example, Rechenberg [74] proved a linear average convergence rate towards the minimum of an ℓ -dimensional equally scaled paraboloid for the $(1+1)$ -EA with normally distributed mutations. Schumer and Steiglitz [75] proved the same in case of mutations that are uniformly distributed on the surface of an ℓ -dimensional hyperball. These results were generalized for strongly convex objective functions under spherically symmetric mutation distributions [76, 77, 78] and mutations distributions with independent marginals [79].

Linear convergence rates can also be shown for EAs with a single parents but multiple offspring [80]. This result was later successively generalized in various directions [81, 82, 83, 84] accompanied by a steady simplification of the mathematical apparatus. The convergence rate of evolutionary algorithms with multiple parents and offspring were investigated by Rechenberg [85], Beyer [86, 87], and Rudolph [13] under various variation operators (including recombination).

So far, the publications regarding average convergence rates tacitly presupposed that the EA is able to determine the Euclidean length of the gradient at arbitrary positions of the search space in order to adjust the mutation distributions appropriately—an assumption that is usually not justified in practice. In real world evolutionary algorithms this task is accomplished by a mechanism termed ‘self-adaptation’ (see e.g. [88]), but, despite first steps in this direction [89], a mathematically rigorous proof of this property is still pending.

4.3. Continuous Space and Time

Evolutionary algorithms with continuous time may be interpreted as follows: The period between the completion of two successive reproduction cycles is interpolated. In this case, the truly discrete time process is embedded into continuous time. Then, combining the theory of analysis and probability, one shows that a convergent subsequence may be extracted and identifies the limit of the sequence as a solution of an ordinary differential equation. This general approach was developed by Ljung [90] and subsequently extended by many others. The monograph by Benveniste et al. [91] offers a detailed account on this method. So-called *stochastic approximation algorithms* have been extensively studied in this manner. Yin et al. [92, 93] showed that some versions of evolutionary algorithms may be formulated as stochastic approximation algorithms so that only few modifications to the existing theory for stochastic approximations algorithms led to various conditions for global convergence with probability one or weaker modes.

Ebeling and various coauthors [94, 95, 96, 97, 98] took a more abstract point of view. They specified continuous-time models via differential equations that should behave similar to existing

evolutionary algorithms. For example, the *Fisher-Eigen equation*

$$\frac{\partial}{\partial t} P(x, t) = (f(x) - \langle f \rangle) P(x, t) + D \Delta P(x, t) \quad (3)$$

where $D > 0$ is a diffusion constant, Δ the Laplace operator, and

$$\langle f \rangle(t) = \frac{\int f(x) P(x, t) dx}{\int P(x, t) dx}$$

the average of the population's fitness at generation $t \geq 0$, models proportional selection by the first term in equation (3) and mutations via a Gaussian diffusion process by the second term. Under certain conditions on the fitness function $f(\cdot)$ the solution of equation (3) permits the determination of the limit ($t \rightarrow \infty$) of the population's distribution $P(x, t)$ at time $t \geq 0$. A more thorough presentation of this approach was given by Aßelmeier [99]. Although this approach principally presupposes an infinitely large population, the dynamics of $P(x, t)$ often characterizes the dynamics of the discrete time EA surprisingly well.

5. Concluding Remarks

This survey is an attempt to summarize finite Markov chains results in the field of evolutionary computation. The main focus was devoted to the limit and finite time behavior of evolutionary algorithms. But there are of course other theoretical questions that might be of interest. Moreover, there are also other techniques and tools that have been applied to examining specific properties of evolutionary algorithms. Apart from Markov chains one can find methods like schema analysis, dimensional analysis, orthogonal functions analysis, as well as approaches via quantitative genetics, statistical physics, and quadratical dynamical systems (see [100] for a brief introduction and key references).

But despite these enhanced activities during the last decade we currently have not more than a first layer of a theoretical foundation of evolutionary computation. Needless to say, there are many challenges that must be encountered in future—but their coverage would not only be a contribution to the field of evolutionary computation but also to the more general field of the analysis of complex systems.

Acknowledgment

This survey was made in the course of the *Collaborative Research Center “Design and Management of Complex Technical Processes and Systems by Means of Computational Intelligence Methods” (SFB 531)*. Financial support by the *Deutsche Forschungsgemeinschaft (DFG)* is gratefully acknowledged.

References

- [1] E. Lukacs. *Stochastic Convergence*. Academic Press, New York, 2nd edition, 1975.
- [2] M. Iosifescu. *Finite Markov Processes and Their Applications*. Wiley, Chichester, 1980.
- [3] E. Seneta. *Non-negative Matrices and Markov Chains*. Springer, New York, 2nd edition, 1981.
- [4] T. E. Davis. *Toward an extrapolation of the simulated annealing convergence theory onto the simple genetic algorithm*. PhD thesis, University of Florida, Gainesville, 1991.
- [5] M. D. Vose and G. E. Liepins. Punctuated equilibria in genetic search. *Complex Systems*, 5(1):31–44, 1991.
- [6] A. E. Nix and M. D. Vose. Modeling genetic algorithms with Markov chains. *Annals of Mathematics and Artificial Intelligence*, 5:79–88, 1992.
- [7] A. E. Eiben, E. H. L. Aarts, and K. M. van Hee. Global convergence of genetic algorithms: A markov chain analysis. In H.-P. Schwefel and R. Männer, editors, *Parallel Problem Solving from Nature*, pages 4–12. Springer, Berlin and Heidelberg, 1991.
- [8] E. H. L. Aarts, A. E. Eiben, and K. H. van Hee. A general theory of genetic algorithms. Technical Report 89/08, Eindhoven University of Technology, Department of Computer Science, 1989.
- [9] S. Arunkumar and T. Chockalingam. Genetic search algorithms and their randomized operators. *Computers and Mathematics with Applications*, 25(5):91–100, 1993.
- [10] D. B. Fogel. Asymptotic convergence properties of genetic algorithms and evolutionary programming: Analysis and experiments. *Cybernetics and Systems*, 25(3):389–407, 1994.
- [11] G. Rudolph. Convergence properties of canonical genetic algorithms. *IEEE Transactions on Neural Networks*, 5(1):96–101, 1994.
- [12] J. Suzuki. A markov chain analysis on simple genetic algorithms. *IEEE Transactions on Systems, Man, and Cybernetics*, 25(4):655–659, 1995.
- [13] G. Rudolph. *Convergence Properties of Evolutionary Algorithms*. Kovač, Hamburg, 1997.
- [14] A. Agapie. Genetic algorithms: Minimal conditions for convergence. In J. K. Hao, E. Lut-ton, E. Ronald, M. Schoenauer, and D. Snyers, editors, *Artificial Evolution: Third European Conference; selected papers / AE '97*, pages 183–193. Springer, Berlin, 1998.
- [15] I. K. Evans. Enhancing recombination with the complementary surrogate genetic algorithm. In *Proceedings of the 4th IEEE International Conference on Evolutionary Computation (ICEC'97)*, pages 97–102. IEEE Press, Piscataway (NJ), 1997.
- [16] G. Rudolph. Evolutionary search for minimal elements in partially ordered finite sets. In V. W. Porto, N. Saravanan, D. Waagen, and A. E. Eiben, editors, *Evolutionary Programming VII, Proceedings of the 7th Annual Conference on Evolutionary Programming*, pages 345–353. Springer, Berlin, 1998.
- [17] T. E. Davis and J. Principe. A markov chain framework for the simple genetic algorithm. *Evolutionary Computation*, 1(3):269–288, 1993.
- [18] S. W. Mahfoud and D. E. Goldberg. Parallel recombinative simulated annealing: A genetic algorithm. *Parallel Computing*, 21(1):1–28, 1995.

- [19] R. Cerf. The dynamics of mutation-selection algorithms with large population sizes. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 32(4):455–508, 1996.
- [20] R. Cerf. A new genetic algorithm. *Annals of Applied Probability*, 6(3):778–817, 1996.
- [21] J. Suzuki. A further result on the markov chain model of genetic algorithm and its application to a simulated annealing-like strategy. *IEEE Transactions on Systems, Man, and Cybernetics—Part B*, 28(1):95–102, 1998.
- [22] J. A. Lozano, P. Larrañaga, M. Graña, and F. X. Albizuri. Genetic algorithms: Bridging the convergence gap. *Theoretical Computer Science*, 214, 1999 (in press).
- [23] J. He and K. Kang. On the convergence rates of genetic algorithms. *Theoretical Computer Science*, 214, 1999 (in press).
- [24] P. L. Hammer and S. Rudeanu. *Boolean Methods in Operations Research and Related Areas*. Springer, Berlin and Heidelberg, 1968.
- [25] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. Freeman, New York, 1979.
- [26] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, New York (NY), 1995.
- [27] M. Hulin. An optimal stop criterion for genetic algorithms: A Bayesian approach. In T. Bäck, editor, *Proceedings of the 7th International Conference on Genetic Algorithms (ICGA'97)*, pages 135–142. Morgan Kaufmann, San Francisco (CA), 1997.
- [28] H. Aytug and G. J. Koehler. Stopping criteria for finite length genetic algorithms. *INFORMS Journal on Computing*, 8(2):183–191, 1996.
- [29] T. Bäck. The interaction of mutation rate, selection, and self-adaptation within a genetic algorithm. In R. Männer and B. Manderick, editors, *Parallel Problem Solving from Nature, 2*, pages 85–94. North Holland, Amsterdam, 1992.
- [30] H. Mühlenbein. How genetic algorithms really work I: Mutation and hillclimbing. In R. Männer and B. Manderick, editors, *Parallel Problem Solving from Nature, 2*, pages 15–25. North Holland, Amsterdam, 1992.
- [31] S. Droste, T. Jansen, and I. Wegener. A rigorous complexity analysis of the $(1 + 1)$ evolution strategy for separable functions with boolean inputs. In *Proceedings of the 5th IEEE International Conference on Evolutionary Computation (ICEC'98)*, pages 499–504. IEEE Press, Piscataway (NJ), 1998.
- [32] P. L. Hammer, B. Simeone, T. M. Liebling, and D. de Werra. From linear separability to unimodality: a hierarchy of pseudo-boolean functions. *SIAM Journal of Discrete Mathematics*, 1(2):174–184, 1988.
- [33] G. Rudolph. How mutation and selection solve long path-problems in polynomial expected time. *Evolutionary Computation*, 4(2):195–205, 1996.
- [34] A. N. Antamoshkin, V. N. Saraev, and E. S. Semenko. Optimization of unimodal monotone pseudoboollean functions. *Kybernetika*, 26(5):432–441, 1990.
- [35] J. Horn, D. E. Goldberg, and K. Deb. Long path problems. In Y. Davidor, H.-P. Schwefel, and R. Männer, editors, *Parallel Problem Solving from Nature, 3*, pages 149–158. Springer, Berlin and Heidelberg, 1994.

- [36] D. S. Johnson, C. H. Papadimitriou, and M. Yannakakis. How easy is local search? *Journal of Computer and System Sciences*, 37(1):79–100, 1988.
- [37] Y. Crama. Recognition problems for special classes of polynomials in 0–1 variables. *Mathematical Programming*, 44:139–155, 1989.
- [38] K. A. De Jong, W. M. Spears, and D. F. Gordon. Using markov chains to analyze GAFOs. In L. D. Whitley and M. D. Vose, editors, *Foundations on Genetic Algorithms, 3*, pages 115–137. Morgan Kaufmann, San Fransisco (CA), 1995.
- [39] W. M. Spears and K. A. De Jong. Analyzing GAs using markov chain models with semantically ordered and lumped states. In R. K. Belew and M. D. Vose, editors, *Foundations of Genetic Algorithms, 4*, pages 85–100. Morgan Kaufmann, San Francisco (CA), 1997.
- [40] P. Vitányi. Genetic fitness optimization using rapidly mixing Markov chains. In S. Arikawa and A. K. Sharma, editors, *Proceedings of the 7th International Workshop on Algorithmic Learning Theory (ALT '96)*, pages 67–82. Springer, Berlin, 1996.
- [41] E. van Nimwegen, J. P. Crutchfield, and M. Mitchell. Statistical dynamics of the royal road genetic algorithm. *Theoretical Computer Science*, 214, 1999 (in press).
- [42] Y. Rabinovich and A. Wigderson. An analysis of a simple genetic algorithm. In R. K. Belew and L. B. Booker, editors, *Proceedings of the Fourth Conference on Genetic Algorithms*, pages 215–221. Morgan Kaufmann, San Mateo, 1991.
- [43] M. Srinivas and L. M. Patnaik. Binomially distributed populations for modelling GAs. In S. Forrest, editor, *Proceedings of the Fifth International Conference on Genetic Algorithms*, pages 138–145. Morgan Kaufmann, San Mateo (CA), 1993.
- [44] U. K. Chakraborty and D. G. Dastidar. Using reliability analysis to estimate the number of generations to convergence in genetic algorithms. *Information Processing Letters*, 46:199–209, 1993.
- [45] H. Dawid. A markov chain analysis of genetic algorithms with a state dependent fitness function. *Complex Systems*, 8(6):407–417, 1994.
- [46] A. Prügel-Bennett and J. L. Shapiro. Analysis of genetic algorithms using statistical mechanics. *Physical Review Letters*, 72(9):1305–1309, 1994.
- [47] B. Goertzel. A convergence theorem for the simple GA with population size tending to infinity. In *Proceedings of the IEEE 2nd International Conference on Evolutionary Computation, Vol. 1*, pages 74–79. IEEE Press, 1995.
- [48] S. Voget. A central limit theorem for the population process of genetic algorithms. *Complex Systems*, 9(4):267–285, 1995.
- [49] S. Voget. Theoretical analysis of genetic algorithms with infinite population size. *Complex Systems*, 10(3):167–183, 1996.
- [50] J. Shapiro and A. Prügel-Bennett. Genetic algorithm dynamics in a two-well potential. In R. K. Belew and M. D. Vose, editors, *Foundations of Genetic Algorithms, 4*, pages 101–116. Morgan Kaufmann, San Francisco (CA), 1997.
- [51] M. D. Vose and A. H. Wright. Stability of vertex fixed points and applications. In L. D. Whitley and M. D. Vose, editors, *Foundations on Genetic Algorithms, 3*, pages 103–113. Morgan Kaufmann, San Fransisco (CA), 1995.

- [52] M. D. Vose and A. H. Wright. Simple genetic algorithms with linear fitness. *Evolutionary Computation*, 2(4):347–368, 1994.
- [53] A. H. Wright and Vose D. Finiteness of the fixed point set for the simple genetic algorithm. *Evolutionary Computation*, 3(3):299–309, 1995.
- [54] G. J. Koehler. A proof of the Vose–Liepins conjecture. *Annals of Mathematics and Artificial Intelligence*, 10(4):409–422, 1994.
- [55] S. Bhattacharyya and G. J. Koehler. An analysis of non-binary genetic algorithms with cardinality 2^v . *Complex Systems*, 8(4):227–256, 1994.
- [56] G. J. Koehler. Diagonalizing the simple GA mixing matrix. In R. K. Belew and M. D. Vose, editors, *Foundations of Genetic Algorithms, 4*, pages 5–51. Morgan Kaufmann, San Francisco (CA), 1997.
- [57] A. H. Wright and G. Bidwell. A search for counterexamples to two conjectures on the simple genetic algorithm. In R. K. Belew and M. D. Vose, editors, *Foundations of Genetic Algorithms, 4*, pages 73–84. Morgan Kaufmann, San Francisco (CA), 1997.
- [58] H. Mühlenbein and D. Schlierkamp-Voosen. Predictive models for the breeder genetic algorithm I: Continuous parameter optimization. *Evolutionary Computation*, 1(1):25–49, 1993.
- [59] H. Mühlenbein and D. Schlierkamp-Voosen. The science of breeding and its application to the breeder genetic algorithm (BGA). *Evolutionary Computation*, 1(4):335–360, 1993.
- [60] S. M. Ermakov and A. A. Zhiglyavskij. On random search for a global extremum. *Theory of Probability and its Applications*, 28(1):136–141, 1983.
- [61] X. Qi and F. Palmieri. Theoretical analysis of evolutionary algorithms with infinite population size in continuous space, part I: basic properties. *IEEE Transactions on Neural Networks*, 5(1):102–119, 1994.
- [62] M. Driml and O. Hanš. On a randomized optimization procedure. In J. Kožešnik, editor, *Transactions of the 4th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes (held at Prague 1965)*, pages 273–276. Czechoslovak Academy of Sciences, Prague, 1967.
- [63] L. P. Devroye. On the convergence of statistical search. *IEEE Transactions on Systems, Man, and Cybernetics*, 6(1):46–56, 1976.
- [64] U. G. Oppel and M. Hohenbichler. Auf der Zufallssuche basierende Evolutionsprozesse. In B. Schneider and U. Ranft, editors, *Simulationsmethoden in der Medizin und Biologie*, pages 130–155. Springer, Berlin, 1978.
- [65] K. Marti. On accelerations of the convergence in random search methods. *Methods of Operations Research*, 37:391–406, 1980.
- [66] F. J. Solis and R. J.-B. Wets. Minimization by random search techniques. *Mathematics of Operations Research*, 6:19–30, 1981.
- [67] J. Pintér. Convergence properties of stochastic optimization procedures. *Mathematische Operationsforschung und Statistik, Series Optimization*, 15:405–427, 1984.
- [68] J. Born. *Evolutionsstrategien zur numerischen Lösung von Adaptationsaufgaben*. Dissertation A, Humboldt-Universität, Berlin, 1978.

- [69] A. A. Zhigljavsky. *Theory of global random search*. Kluwer, AA Dordrecht, The Netherlands, 1991.
- [70] C. C. Peck and A. P. Dhawan. Genetic algorithms as global random search methods: An alternative perspective. *Evolutionary Computation*, 3(1):39–80, 1995.
- [71] G. Rudolph. Convergence of evolutionary algorithms in general search spaces. In *Proceedings of the Third IEEE Conference on Evolutionary Computation*, pages 50–54. IEEE Press, Piscataway (NJ), 1996.
- [72] G. Rudolph. Convergence of non-elitist strategies. In *Proceedings of the First IEEE Conference on Evolutionary Computation, Vol. 1*, pages 63–66. IEEE Press, Piscataway (NJ), 1994.
- [73] G. Rudolph. On a multi-objective evolutionary algorithm and its convergence to the pareto set. In *Proceedings of the 1998 IEEE International Conference on Evolutionary Computation*, pages 511–516. IEEE Press, Piscataway (NJ), 1998.
- [74] I. Rechenberg. *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. Frommann-Holzboog Verlag, Stuttgart, 1973.
- [75] M. A. Schumer and K. Steiglitz. Adaptive step size random search. *IEEE Transactions on Automatic Control*, 13:270–276, 1968.
- [76] G. Rappl. *Konvergenzraten von Random Search Verfahren zur globalen Optimierung*. Dissertation, Hochschule der Bundeswehr München, Germany, 1984.
- [77] G. Rappl. On linear convergence of a class of random search algorithms. *Zeitschrift für angewandte Mathematik und Mechanik (ZAMM)*, 69(1):37–45, 1989.
- [78] G. Rudolph. Local convergence rates of simple evolutionary algorithms with cauchy mutations. *IEEE Transactions on Evolutionary Computation*, 1(4):249–258, 1998.
- [79] G. Rudolph. Convergence rates of simple evolutionary algorithms under factorizing mutation distributions. In J. K. Hao, E. Lutton, E. Ronald, M. Schoenauer, and D. Snyers, editors, *Artificial Evolution: Third European Conference; selected papers / AE '97*, pages 275–285. Springer, Berlin, 1998.
- [80] H.-P. Schwefel. *Evolutionsstrategie und numerische Optimierung*. Dissertation, Technische Universität Berlin, 1975.
- [81] A. Scheel. *Beitrag zur Theorie der Evolutionsstrategie*. Dissertation, TU Berlin, Berlin, 1985.
- [82] T. Bäck, G. Rudolph, and H.-P. Schwefel. Evolutionary programming and evolution strategies: Similarities and differences. In D. B. Fogel and W. Atmar, editors, *Proceedings of the 2nd Annual Conference on Evolutionary Programming*, pages 11–22. Evolutionary Programming Society, La Jolla (CA), 1993.
- [83] H.-G. Beyer. Toward a theory of evolution strategies: Some asymptotical results from the $(1 + \lambda)$ -theory. *Evolutionary Computation*, 1(2):165–188, 1993.
- [84] G. Rudolph. Convergence rates of evolutionary algorithms for a class of convex objective functions. *Control and Cybernetics*, 26(3):375–390, 1997.
- [85] I. Rechenberg. *Evolutionsstrategie '94*. Frommann-Holzboog Verlag, Stuttgart, 1994.

- [86] H.-G. Beyer. Toward a theory of evolution strategies: The (μ, λ) -theory. *Evolutionary Computation*, 2(4):381–407, 1994.
- [87] H.-G. Beyer. Toward a theory of evolution strategies: On the benefits of sex — the $(\mu/\mu, \lambda)$ theory. *Evolutionary Computation*, 3(1):81–111, 1995.
- [88] T. Bäck and H.-P. Schwefel. An overview of evolutionary algorithms for parameter optimization. *Evolutionary Computation*, 1(1):1–23, 1993.
- [89] H.-G. Beyer. Toward a theory of evolution strategies: Self-adaptation. *Evolutionary Computation*, 3(3):311–347, 1995.
- [90] L. Ljung. Analysis of recursive stochastic algorithms. *IEEE Transactions on Automatic Control*, 22:551–575, 1977.
- [91] A. Benveniste, M. Métivier, and P. Priouret. *Adaptive Algorithms and Stochastic Approximations*. Springer, Berlin and Heidelberg, 1990.
- [92] G. Yin, G. Rudolph, and H.-P. Schwefel. Establishing connections between evolutionary algorithms and stochastic approximation. *Informatica*, 6(1):93–116, 1995.
- [93] G. Yin, G. Rudolph, and H.-P. Schwefel. Analyzing $(1, \lambda)$ evolution strategy via stochastic approximation methods. *Evolutionary Computation*, 3(4):473–489, 1995.
- [94] R. Feistel and W. Ebeling. Models of Darwinian processes and evolutionary principles. *BioSystems*, 15:291–299, 1982.
- [95] W. Ebeling and A. Engel. Models of evolutionary systems and their application to optimization problems. *Systems Analysis, Modelling, Simulation*, 3(5):377–385, 1986.
- [96] T. Aßelmeier, W. Ebeling, and H. Rosé. Analytical and numerical investigations of evolutionary algorithms in continuous space. In H.-M. Voigt, W. Ebeling, I. Rechenberg, and H.-P. Schwefel, editors, *Parallel Problem Solving From Nature — PPSN IV*, pages 112–121. Springer, Berlin, 1996.
- [97] T. Aßelmeyer and W. Ebeling. Unified description of evolutionary strategies over continuous parameter spaces. *BioSystems*, 41(3):167–178, 1997.
- [98] T. Aßelmeier, W. Ebeling, and H. Rosé. Evolutionary strategies of optimization. *Physical Review E*, 56(1):1171–1180, 1997.
- [99] T. Aßelmeier. *Schrödinger-Operatoren und Evolutionäre Strategien*. Dissertation, Humboldt-Universität zu Berlin, Institut für Physik, 1997.
- [100] A. E. Eiben and G. Rudolph. Theory of evolutionary algorithms: A birds eye view. *Theoretical Computer Science*, 214, 1999 (in press).