

Convergence of Evolutionary Algorithms in General Search Spaces

Günter Rudolph

ICD Informatik Centrum Dortmund
Joseph-von-Fraunhofer-Str. 20
D-44227 Dortmund
rudolph@LS11.informatik.uni-dortmund.de

Abstract—This paper provides conditions under which evolutionary algorithms with an elitist selection rule will converge to the global optimum of some function whose domain may be an arbitrary space. These results generalize the previously developed convergence theory for binary and Euclidean search spaces to general search spaces.

I. INTRODUCTION

The term *evolutionary algorithm* (EA) is a collective name for those probabilistic optimization algorithms whose design is inspired by principles of biological evolution. Although particular EAs may differ considerably at a first glance, there are more similarities than differences. In fact, a general convergence theory is possible.

We shall suppose that some EA is used to minimize a real-valued objective function $f : \mathcal{M} \rightarrow \mathbb{R}$ that is bounded from below, that is, $f(x) > -\infty$ for all $x \in \mathcal{M}$. No further assumptions are imposed on the search space \mathcal{M} .

It will be shown that a specific class of evolutionary algorithms will converge to the global optimum of the optimization problem above regardless of the search space. This result generalizes results obtained for individual-based EAs in Euclidean [1; 2; 3; 4; 5; 6] and population-based EAs in binary (or finite) search spaces [7; 8; 9].

Since evolutionary algorithms are describable by Markov chains we first introduce some terminology concerning Markov chains with general state spaces [10] and establish the connection between the limiting behavior of Markov chains and the stochastic convergence of random sequences [11] leading to a precise condition for which an EA converges to the global optimum.

The main result is given in section III and discussed in section IV.

II. THE MARKOV MODEL

Let $(X_t : t \in T)$ be a family of random variables on a joint probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in a set E of a measurable space (E, \mathcal{A}) and index set T . Then the family of random variables $(X_t : t \in T)$ is called a *stochastic process* with index set T .

In general, there is no mathematical reason for restricting index set T to numerical sets. In this article, however, the index set T is identical with \mathbb{N}_0 and the indices $t \in T$ will be interpreted as points of time.

A stochastic process $(X_t : t \in T)$ with index set $T = \mathbb{N}_0$ is called a *stochastic process with discrete time*. The sequence $X_0(\omega), X_1(\omega), \dots$ is termed a *sample sequence* for each fixed $\omega \in \Omega$. The image space E of $(X_t : t \in T)$ is

called the *state space* of the process.

Let $(X_t : t \geq 0)$ be a stochastic process with discrete time on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in E of a measurable space (E, \mathcal{A}) . If for $0 < t_1 < t_2 < \dots < t_k < t$ with some $k \in \mathbb{N}$ and $A \in \mathcal{A}$

$$\mathbb{P}\{X_t \in A | X_{t_1}, X_{t_2}, \dots, X_{t_k}\} = \mathbb{P}\{X_t \in A | X_{t_k}\} \quad (1)$$

with probability 1, then $(X_t : t \geq 0)$ is called a *Markov chain*. If $\mathbb{P}\{X_{t+k} \in A | X_{s+k}\} = \mathbb{P}\{X_t \in A | X_s\}$ for arbitrary $s, t, k \in \mathbb{N}_0$ with $s \leq t$, then the Markov chain is termed *homogeneous*, otherwise *inhomogeneous*.

Let $(X_t : t \geq 0)$ be a homogeneous Markov chain on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with image space (E, \mathcal{A}) . The map $\mathbf{K} : E \times \mathcal{A} \rightarrow [0, 1]$ is termed a *Markovian kernel* or a *transition probability function* for Markov chain $(X_t : t \geq 0)$ if $\mathbf{K}(\cdot, A)$ is measurable for any fixed set $A \in \mathcal{A}$ and $\mathbf{K}(x, \cdot)$ is a probability measure on (E, \mathcal{A}) for any fixed state $x \in E$. In particular, $\mathbf{K}(x_t, A) = \mathbb{P}\{X_{t+1} \in A | X_t = x_t\}$.

The t -th iteration of the Markovian kernel given by

$$\mathbf{K}^{(t)}(x, A) = \begin{cases} \mathbf{K}(x, A) & , t = 1 \\ \int_E \mathbf{K}^{(t-1)}(y, A) \mathbf{K}(x, dy) & , t > 1 \end{cases}$$

describes the probability to transition to some set $A \subseteq E$ within t steps when starting from the state $x \in E$. Let $p(\cdot)$ be the initial distribution over subsets A of \mathcal{A} . Then the probability that the Markov chain is in set A at step $t \geq 0$ is determined by

$$\mathbb{P}\{X_t \in A\} = \begin{cases} p(A) & , t = 0 \\ \int_E \mathbf{K}^{(t)}(x, A) p(dx) & , t > 0 \end{cases}$$

where integration is with respect to an appropriate measure on (E, \mathcal{A}) . For example, if $E = \mathbb{R}^n$ then integration is with respect to the Lebesgue measure. If E is finite then the counting measure is appropriate and the integrals reduce to sums. Then the Markovian kernel can be described by a finite number of transition probabilities $p_{xy} = \mathbf{K}(x, \{y\}) \geq 0$ that can be gathered in a square matrix $\mathbf{P} = (p_{xy})$ with $x, y \in E$.

We are now in the position to build the Markov model of the EA: An EA consists of a population of N individuals represented by the N -tuple (x_1, \dots, x_N) with $x_i \in \mathcal{M}$ for $i = 1, \dots, N$. Thus, the state space is $E = \mathcal{M}^N$. The initial population at step $t = 0$ is chosen according to some initial distribution $p(\cdot)$ yielding the random population $X_0 = (X_{0,1}, \dots, X_{0,N})$. A population at step $t \geq 0$ is modified by so-called genetic operators that generate a

new population at step $t + 1$. The outcome only depends on the previous population and the probabilistic modifications caused by the genetic operators are described by the stochastic kernel $\mathbf{K}(\cdot, \cdot)$. Consequently, the stochastic sequence $(X_t : t \geq 0)$ is a Markov chain.

In the next section we shall investigate whether the EA is able to converge in some sense to a specific set A that is related to the globally optimal solutions of our optimization problem. It is clear that we must study the iterates of the stochastic kernel. But at first it must be made precise what we shall understand under the term *convergence*.

Since random sequences are defined on probability spaces the main difference between all modes of stochastic convergence and the convergence concept of classical analysis relies on the fact that the first must take into account the existence of a probability measure. These various modes of stochastic convergence distinguish among each other in the way in which the probability measure enters the definition. Here, only two concepts will be of interest:

Let (D_t) be a sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Then (D_t) is said to *converge completely* to 0, denoted as $D_t \xrightarrow{c} 0$, if for any $\epsilon > 0$

$$\lim_{t \rightarrow \infty} \sum_{i=1}^t \mathbf{P}\{|D_i| > \epsilon\} < \infty \quad (2)$$

and to *converge in probability* to 0, denoted as $D_t \xrightarrow{\mathbf{P}} 0$, if for any $\epsilon > 0$

$$\lim_{t \rightarrow \infty} \mathbf{P}\{|D_t| > \epsilon\} = 0. \quad (3)$$

Let $b(X_t) = \min\{f(X_{t,k}) : k = 1, \dots, N\}$ denote the best objective function value that is represented by some individual $X_{t,k}$ of population X_t at step $t \geq 0$. Provided that the random sequence $b(X_t)$ converges in some mode to the global optimum $f^* = \min\{f(x) : x \in \mathcal{M}\}$, we can be sure that the population X_t will contain better and better solutions of our optimization problem for increasing t .

Therefore we define: An EA converges to the global optimum if the random sequence $(D_t : t \geq 0)$ with $D_t = d(X_t) = b(X_t) - f^*$ converges completely to zero.

Since (3) is a necessary condition for (2) we shall first prove convergence in probability. Then we shall show that the convergence in (3) is fast enough such that the series in (2) is finite.

III. MAIN RESULT

We first show an auxiliary result:

Lemma 1 *Let $A_\epsilon = \{x \in E : d(x) < \epsilon\}$ with some $\epsilon > 0$ be the set of ϵ -optimal states. If $\mathbf{K}(x, A_\epsilon) \geq \delta > 0$ for all $x \in A_\epsilon^c = E \setminus A_\epsilon$ and $\mathbf{K}(x, A_\epsilon) = 1$ for $x \in A_\epsilon$ then*

$$\mathbf{K}^{(t)}(x, A_\epsilon) \geq 1 - (1 - \delta)^t$$

for $t \geq 1$.

Proof: (by induction)

Let $t = 1$. Then the hypothesis is true because

$\mathbf{K}^{(1)}(x, A_\epsilon) = \mathbf{K}(x, A_\epsilon) \geq 1 - (1 - \delta)^1 = \delta$. Now assume that the hypothesis is true for $t > 1$. First note that $\mathbf{K}^{(t)}(x, A_\epsilon) = 1$ for all $t \geq 1$ if $x \in A_\epsilon$. This will be used to obtain (4). Thus,

$$\begin{aligned} & \mathbf{K}^{(t+1)}(x, A_\epsilon) \\ &= \int_E \mathbf{K}^{(t)}(y, A_\epsilon) \mathbf{K}(x, dy) \\ &= \int_{A_\epsilon} \mathbf{K}^{(t)}(y, A_\epsilon) \mathbf{K}(x, dy) + \\ & \quad \int_{A_\epsilon^c} \mathbf{K}^{(t)}(y, A_\epsilon) \mathbf{K}(x, dy) \\ &= \int_{A_\epsilon} \mathbf{K}(x, dy) + \int_{A_\epsilon^c} \mathbf{K}^{(t)}(y, A_\epsilon) \mathbf{K}(x, dy) \quad (4) \\ &= \mathbf{K}(x, A_\epsilon) + \int_{A_\epsilon^c} \mathbf{K}^{(t)}(y, A_\epsilon) \mathbf{K}(x, dy) \\ &\geq \mathbf{K}(x, A_\epsilon) + [1 - (1 - \delta)^t] \int_{A_\epsilon^c} \mathbf{K}(x, dy) \\ &= \mathbf{K}(x, A_\epsilon) + [1 - (1 - \delta)^t] \mathbf{K}(x, A_\epsilon^c) \\ &= \mathbf{K}(x, A_\epsilon) + \mathbf{K}(x, A_\epsilon^c) - (1 - \delta)^t \mathbf{K}(x, A_\epsilon^c) \\ &= 1 - (1 - \delta)^t \mathbf{K}(x, A_\epsilon^c) \\ &= 1 - (1 - \delta)^t (1 - \mathbf{K}(x, A_\epsilon)) \\ &\geq 1 - (1 - \delta)^t (1 - \delta) \\ &= 1 - (1 - \delta)^{t+1}. \end{aligned}$$

Consequently, the hypothesis is true for $t \geq 1$. \square

Now we are in the position to state our main result:

Theorem 1 *An evolutionary algorithm, whose stochastic kernel satisfies the preconditions of Lemma 1, will converge to the global minimum of a real-valued function $f : \mathcal{M} \rightarrow \mathbb{R}$ with $f > -\infty$ defined on an arbitrary space \mathcal{M} , regardless of the initial distribution.*

Proof:

Let $(X_t : t \geq 0)$ be the random sequence of populations generated by the EA. To prove the theorem we have to show that the random sequence $(d(X_t) : t \geq 0)$ converges completely to zero under the preconditions of Lemma 1. To this end first note that for $t \geq 1$

$$\begin{aligned} & \mathbf{P}\{X_t \in A_\epsilon\} \\ &= \int_E \mathbf{K}^{(t)}(x, A_\epsilon) p(dx) \\ &\geq [1 - (1 - \delta)^t] \int_E p(dx) \quad \text{by Lemma 1} \\ &= 1 - (1 - \delta)^t. \end{aligned}$$

Since $\mathbf{P}\{d(X_t) \leq \epsilon\} = \mathbf{P}\{X_t \in A_\epsilon\}$ for all $t \geq 0$ we obtain

$$\mathbf{P}\{d(X_t) > \epsilon\} = 1 - \mathbf{P}\{X_t \in A_\epsilon\} \leq (1 - \delta)^t \rightarrow 0$$

as $t \rightarrow \infty$ and hence $d(X_t) \xrightarrow{\mathbf{P}} 0$. Finally, note that

$$\sum_{t=1}^{\infty} \mathbf{P}\{d(X_t) > \epsilon\} \leq \sum_{t=1}^{\infty} (1 - \delta)^t = \frac{1 - \delta}{\delta} < \infty$$

satisfying the condition that the random sequence $(d(X_t) : t \geq 0)$ converges completely to zero. This completes the proof. \square

Thus, owing to Theorem 1 it is sufficient to investigate the properties of the stochastic kernel associated with the particular EA instead of its iterates. But which EAs possess a stochastic kernel that satisfies the preconditions of Lemma 1? This question will be discussed in the next section.

IV. DISCUSSION

In general, one iteration of an EA can be split into two phases: A modification and a selection phase. In the modification phase we subsume all those genetic operators that are used to create offspring from the parents. The selection phase represents the operation to choose offspring that will serve as parents in the next iteration. Since the selection phase begins after the modification phase is finished, both phases may be described separately by their own Markovian kernels. Thus, the Markovian kernel of the entire EA may be written as the product kernel

$$\mathbf{K}(x, A) = (\mathbf{K}_m \mathbf{K}_s)(x, A) = \int_E \mathbf{K}_m(x, dy) \mathbf{K}_s(y, A)$$

of the ‘‘modification kernel’’ \mathbf{K}_m and the ‘‘selection kernel’’ \mathbf{K}_s . We shall suppose that the selection phase is performed by an *elitist selection* operator — a notion that was introduced in [12] and can be realized in various ways. In any case, the key property of elitist selection rules is that the best individual of a population at step $t + 1$ is not worse than the best individual of a population at step $t \geq 0$.

Let us start with a very simple EA: The population only consists of a single individual. Thus, the state space E is just $E = \mathcal{M}$. Exactly one offspring is generated from the parent by mutation. We shall assume that this operation is representable by the modification (or mutation) kernel \mathbf{K}_m . If the offspring $y \in E$ is not worse than its parent $x \in E$, that is, $y \in B(x) = \{v \in E : f(v) \leq f(x)\}$, then the offspring will serve as new parent, otherwise the parent will survive. Thus, the algorithm may be written as

$$X_{t+1} = Y_t \cdot 1_{B(X_t)}(Y_t) + X_t \cdot 1_{B^c(X_t)}(Y_t) \quad (5)$$

where $1_A(x)$ denotes the indicator function for some set A and where $B^c(x) = E \setminus B(x)$. Evidently, the EA described in (5) uses an elitist selection rule. Next, we shall derive the selection kernel \mathbf{K}_s for this EA.

Since the selection kernel depends on the previous state $x \in E$ this state is attached to \mathbf{K}_s as an additional parameter. Then the selection kernel is given by

$$\mathbf{K}_s(y, A; x) = 1_{B(x)}(y) \cdot 1_A(y) + 1_{B^c(x)}(y) \cdot 1_A(x) \quad (6)$$

and may be interpreted as follows: If state $y \in E$ is better equal state x (i.e., $y \in B(x)$) and also in set A , then y transitions to set A , and more precisely to set $A \cap B(x)$, with probability one. If y is worse than x (i.e., $y \in B^c(x)$) then y is not accepted. Rather, y will transition to the

old state x with probability one. But if x was in set A then y will transition to $x \in A$ with probability one. All other cases have probability zero. Evidently, the selection kernel is purely deterministic here. Putting all together the product kernel of mutation and selection is

$$\begin{aligned} \mathbf{K}(x, A) &= \int_E \mathbf{K}_m(x, dy) \cdot \mathbf{K}_s(y, A; x) \\ &= \int_E \mathbf{K}_m(x, dy) \cdot 1_{A \cap B(x)}(y) \\ &\quad + 1_A(x) \cdot \int_E \mathbf{K}_m(x, dy) \cdot 1_{B^c(x)}(y) \\ &= \int_{A \cap B(x)} \mathbf{K}_m(x, dy) + 1_A(x) \cdot \int_{B^c(x)} \mathbf{K}_m(x, dy) \\ &= \mathbf{K}_m(x, A \cap B(x)) + 1_A(x) \cdot \mathbf{K}_m(x, B^c(x)). \end{aligned}$$

There is an important observation: The structure of kernel \mathbf{K} would remain valid for a population-based EA with a special version of elitist selection and if the mutation kernel \mathbf{K}_m is replaced by the corresponding modification kernel. To see this let $E = \mathcal{M}^N$ with arbitrary \mathcal{M} and recall the definition of map $b : E \rightarrow \mathbb{R}$ that extracts the objective function value of the best individual from a population. Then the set of states better equal state x can be re-defined via $B(x) = \{y \in E : b(y) \leq b(x)\}$.

What happens with the selection kernel? If $y \in E$ is in $B(x) \cap A$ the population transitions to A . If $y \notin B(x)$ then the best individual of population y is worse than the best individual of population x . If the entire population is rejected then the structure of the population-based kernel is identical to the individual-based kernel above. But under usual elitist selection the best individual is re-inserted — somehow — into population y yielding $y' = e(x, y) \in B(x)$. Here the map $e : E \times E \rightarrow E$ encapsulates the method to re-insert the best individual of $x \in E$ into y . Consequently, the selection kernel becomes

$$\mathbf{K}_s(y, A; x) = 1_{B(x) \cap A}(y) + 1_{B^c(x)}(y) \cdot 1_A(x) \cdot 1_A(e(x, y))$$

leading to

$$\begin{aligned} \mathbf{K}(x, A) &= \mathbf{K}_m(x, B(x) \cap A) \\ &\quad + 1_A(x) \cdot \int_{B^c(x)} \mathbf{K}_m(x, dy) \cdot 1_A(e(x, y)). \quad (7) \end{aligned}$$

The integral in (7) is unpleasant. But if this Markovian kernel is restricted to the set A_ϵ of ϵ -optimal solutions, it will shrink to a very simple expression. To this end consider EAs whose Markovian kernel is represented by (7): If $A_\epsilon \subset B(x)$ then $x \notin A_\epsilon$, $A_\epsilon \cap B(x) = A_\epsilon$ and $\mathbf{K}(x, A_\epsilon) = \mathbf{K}_m(x, A_\epsilon)$. If $B(x) \subseteq A_\epsilon$ then $x \in A_\epsilon$, $A_\epsilon \cap B(x) = B(x)$ and

$$\begin{aligned} \mathbf{K}(x, A_\epsilon) &= \mathbf{K}_m(x, B(x)) + \int_{B^c(x)} \mathbf{K}_m(x, dy) \cdot 1_{A_\epsilon}(e(x, y)) \\ &= \mathbf{K}_m(x, B(x)) + \mathbf{K}_m(x, B^c(x)) = 1 \end{aligned}$$

since $\epsilon(x, y) \in B(x) \subseteq A_\epsilon$. Therefore the Markovian kernel restricted to set A_ϵ is

$$K(x, A_\epsilon) = K_m(x, A_\epsilon) \cdot 1_{A_\epsilon}(x) + 1_{A_\epsilon}(x)$$

satisfying the preconditions of Theorem 1 if $K_m(x, A_\epsilon) \geq \delta > 0$ for all $x \in A_\epsilon$.

As a consequence, we must guarantee that the set A_ϵ can be reached from everywhere out of A_ϵ with some minimal probability δ that is strictly bounded from zero and may depend on $\epsilon > 0$. It is clear that this property must be realized by the genetic operators represented by the modification phase. In most cases the modification phase can be split further into the crossover and mutation operation. We assume that the mutation operation can be described by the Markovian kernel $K_m(x, A)$ whereas the crossover operation is represented by the kernel $K_c(x, A)$. Thus, the modification kernel is now interpreted as the product kernel of the crossover kernel K_c and the mutation kernel K_m .

If the mutation kernel is strictly bounded from zero for each A_ϵ , i.e., $K_m(x, A_\epsilon) \geq \delta(\epsilon) > 0$ for each population $x \in E$ and each subset $A_\epsilon \subseteq \mathcal{A}$, then the product kernel $(K_c K_m)(x, A)$, describing the joint probabilistic behavior of the crossover and mutation operation, is also strictly bounded from zero for A_ϵ because

$$\begin{aligned} (K_c K_m)(x, A_\epsilon) &= \int_E K_c(x, dy) K_m(y, A_\epsilon) \\ &\geq \delta \int_E K_c(x, dy) \\ &= \delta \cdot K_c(x, E) = \delta > 0 \end{aligned}$$

for each $x \in E$ where $\delta = \delta(\epsilon)$. Evidently, this guarantees that arbitrary sets A_ϵ with $\epsilon > 0$ are reachable in one step from any $x \in E$ at least with probability $\delta > 0$. Consequently, the Markovian kernel of the entire EA meets the requirements of Theorem 1 and we may state the following sufficient condition:

Theorem 2 *An evolutionary algorithm with elitist selection and a mutation kernel $K_m(x, A)$ that is strictly bounded from zero for each $x \in E$ and $A_\epsilon \in \mathcal{A}$ will converge to the global minimum of a real-valued function $f : \mathcal{M} \rightarrow \mathbb{R}$ with $f > -\infty$ defined on an arbitrary space \mathcal{M} . \square*

This result reveals that it can be sufficient to look at the mutation and selection operators separately. Since the selection operator is independent from the search space \mathcal{M} and elitism is easy to check, we only have to investigate the mutation operator to detect potential difficulties in ensuring the bounded positiveness of the mutation kernel for sets A_ϵ .

We shall postulate that the mutation operator has the following working scheme: Each individual is modified at random independently.

Let $M_\epsilon = \{x \in \mathcal{M} : f(x) - f^* \leq \epsilon\}$ for $\epsilon > 0$ be the set of ϵ -optimal solutions. If the probability that an individual is mutated to an individual in the set $M_\epsilon \subseteq \mathcal{M}$ is larger equal $\beta = \beta(\epsilon) > 0$, then the probability that a population

of N individuals enters the set $A_\epsilon \in \mathcal{A}$ is at least $\delta = 1 - (1 - \beta)^N > 0$. In this case we may restrict our attention to mutation at the individuals' level. This works well for finite sample spaces E . For example, take $\mathcal{M} = \mathbb{B}^\ell = \{0, 1\}^\ell$ so that $E = \mathbb{B}^{\ell \cdot N}$. For this search space \mathcal{M} an individual $x \in \mathbb{B}^\ell$ is mutated almost always by inverting each entry x_i independently with some probability $q \in (0, 1)$. It is easy to see that the probability to mutate x to x' is larger equal $\min\{q^\ell, (1 - q)^\ell\} = \beta > 0$ independent from the choices for $x, x' \in \mathbb{B}^\ell$ (here β does not depend on ϵ). Since $\beta > 0$ implies $\delta > 0$, Theorem 2 guarantees global convergence of EAs with elitist selection and usual mutation for pseudo-boolean optimization problems. This result subsumes the specialized results given in [7; 8; 9].

Now consider problems with search space $\mathcal{M} = \mathbb{R}^\ell$. Here, the mutation of some individual $x \in \mathbb{R}^\ell$ is realized by adding a random vector $z = \sigma S z'$ with $\sigma > 0$, where z' is a ℓ -dimensional random vector with support \mathbb{R}^ℓ and expectation $E[z'] = 0$, and where S is a matrix with $\det(S) = 1$ such that $S^T S$ is positive definite with full rank. To ensure that the probability that individual x mutates to $x + z = x' \in M_\epsilon$ is larger than zero we have to exclude functions f with isolated global solutions $x^* \in \mathbb{R}^\ell$. But this is not enough to obtain a lower bound for the probability to mutate to A_ϵ . Additional restrictions are necessary.

Suppose that $K_c(x, A) = 1_A(x)$ is the identity kernel, i.e., the EA only employs mutation. Moreover, assume that the best individual at step $t \geq 0$ is mutated with probability one and that M_ϵ is bounded for each $\epsilon > 0$. We now restrict our attention to the best individual, because whenever the best individual enters M_ϵ with probability $\beta > 0$ then the population will also enter A_ϵ at least with probability $\beta = \delta > 0$.

Let the best individual of the initial population be in M_γ for some $\gamma > \epsilon > 0$. Since M_γ is bounded and cannot be left by the best individuals of subsequent generations there exists a lower bound

$$\beta = \min\{\mathbf{P}\{X + Z = X' \in M_\epsilon \mid X = x\} : x \in M_\gamma\} > 0$$

for the probability that the best individual will enter M_ϵ and convergence to the global optimum is guaranteed. This result was given in [1; 3; 4] for $N = 1$ while [5; 6] offered a slight generalization that is postponed for while.

Now suppose that $K_c(x, A)$ is not the identity kernel. Then the construction above does not work, because the crossover operation may push all individuals out of M_γ and it is not obvious how to specify the set M_η with some $\eta \geq \gamma$ that cannot be left in general. The situation changes if the crossover operation is applied with some probability $p_c \in (0, 1)$, because the best individual may pass through the crossover operation without being affected with some fixed probability. Although this decreases the bound β derived in case of the identity kernel, the argumentation given there remains valid.

Next, let $p_c = 1$. If there is some minimum probability that the crossover operator leaves the best individual unaltered then the argumentation also remains valid. In fact,

many crossover operators possess this property.

Summing up: In general, it is not sufficient to check the mutation kernel for positiveness and the selection kernel for elitism to fulfill the preconditions of Lemma 1 and Theorem 1 in general search spaces. But it seems possible to derive conditions for further genetic operators and problem classes leading to more powerful results than Theorem 2.

V. OUTLOOK

A slight generalization of Lemma 1 is possible. Similar to [2; 5; 6] we could replace the constant $\delta > 0$ by a sequence of constants $\delta_t \geq 0$ that must have certain properties.

Our general goal, however, is to derive conditions on kernels for specific operators that imply global convergence, are easy to check and are general enough to be useful for a broad class of evolutionary algorithms. This approach appears promising since it is much easier to analyze the behavior of single genetic operators than the complicated interactions between them.

REFERENCES

- [1] M. Driml and O. Hanš. On a randomized optimization procedure. In J. Kožešnik (ed.), *Transactions of the 4th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes (held at Prague 1965)*, pp. 273–276. Czechoslovak Academy of Sciences, Prague, 1967.
- [2] L.P. Devroye. On the convergence of statistical search. *IEEE Transactions on Systems, Man, and Cybernetics*, 6(1):46–56, 1976.
- [3] U.G. Oppel and M. Hohenbichler. Auf der Zufallssuche basierende Evolutionsprozesse. In B. Schneider and U. Ranft, (eds.), *Simulationsmethoden in der Medizin und Biologie*, pp. 130–155. Springer, Berlin, 1978.
- [4] J. Born. *Evolutionsstrategien zur numerischen Lösung von Adaptationsaufgaben*. Doctoral thesis, Humboldt-Universität, Berlin, 1978.
- [5] F.J. Solis and R.J.-B. Wets. Minimization by random search techniques. *Mathematics of Operations Research*, 6:19–30, 1981.
- [6] J. Pintér. Convergence properties of stochastic optimization procedures. *Mathematische Operationsforschung und Statistik, Series Optimization*, 15:405–427, 1984.
- [7] D.B. Fogel. Asymptotic convergence properties of genetic algorithms and evolutionary programming: Analysis and experiments. *Cybernetics and Systems*, 25(3):389–407, 1994.
- [8] G. Rudolph. Convergence properties of canonical genetic algorithms. *IEEE Transaction on Neural Networks*, 5(1):96–101, 1994.
- [9] J. Suzuki. A markov chain analysis on simple genetic algorithms. *IEEE Transactions on Systems, Man, and Cybernetics*, 25(4):655–659, 1995.
- [10] E. Nummelin. *General irreducible Markov chains and non-negative operators*. Cambridge University Press, Cambridge, 1984.

[11] E. Lukacs. *Stochastic Convergence*. Academic Press, New York, 2nd edition, 1975.

[12] K. De Jong. *An analysis of the behavior of a class of genetic adaptive systems*. PhD thesis, University of Michigan, 1975.