A New Model of Simulated Evolutionary Computation—Convergence Analysis and Specifications

Kwong-Sak Leung, *Senior Member, IEEE*, Qi-Hong Duan, Zong-Ben Xu, and C. K. Wong, *Fellow, IEEE*

***Abstract*—There have been various algorithms designed for simulating natural evolution. This paper proposes a new simulated evolutionary computation model called the abstract evolutionary algorithm (AEA), which unifies most of the currently known evolutionary algorithms and describes the evolution as an abstract stochastic process composed of two fundamental operators: selection and evolution operators. By axiomatically characterizing the properties of the fundamental selection and evolution operators, several general convergence theorems and convergence rate estimations for the AEA are established. The established theorems are applied to a series of known evolutionary algorithms, directly yielding new convergence conditions and convergence rate estimations of various specific genetic algorithms and evolutionary strategies. The present work provides a significant step toward the establishment of a unified theory of simulated evolutionary computation.**

***Index Terms*—Aggregating and scattering rate, evolutionary strategy, genetic algorithm, selection intensity, selection pressure, stochastic process.**

I. Introduction

THERE has been a growing interest in algorithms that rely on an analogy to the natural evolution. The best-known examples of such simulated evolutionary algorithms (SEAs) include genetic algorithms (GAs), evolutionary programming (EP), and evolution strategies (ESs). These algorithms simulate the principle of evolution (survival of the fittest), and maintain a population of potential solutions (individuals) through repeated application of some "evolutionary" operators. They yield individuals with successively improved fitness, and converge, hopefully, to the fittest individuals representing optimum solutions. As such, SEAs can be abstractly viewed as general-purpose optimization solvers for "fitness" functions, although they were originally developed for different purposes (e.g., [1], [7], and [10]) and are now applied in different areas. They all have been successfully applied to solve global optimization problems related to some concrete multimodal, nondifferentiable, or combinatorial functions.

Manuscript received March 1, 1999; revised December 3, 1999. This work was supported by the earmarked Grant CUHK 352/96E of the Hong Kong Research Grant Council, the Hi-Tech. R&D (863-306) Project, and the Natural Science Foundation of China.

K.-S. Leung and C. K. Wong are with the Department of Computer Science and Engineering, Chinese University of Hong Kong, Shatin, N.T., Hong Kong, R.O.C.

Q.-H. Duan and Z.-B. Xu are with the Research Center for Applied Mathematics and the Institute for Information and System Sciences, Xi'an Jiaotong University, Xi'an, 710049, R.O.C.

Publisher Item Identifier S 1089-778X(01)02134-8.

The most fundamental theoretical issue related to the research and applications of SEAs is convergence, i.e., whether or not a given implementation scheme of the SEAs will converge to the global optimal solution, and how fast it will converge. This issue has been studied recently [2], [17], [19], [25] for several specific types of SEAs (such as GAs with elitist selection and ESs with specific objective functions). The purpose of the present paper is to extend these results and to establish a more general convergence theory of SEAs by developing a new unified SEA model.

Such a study is motivated by the following observations. First, the currently known SEAs such as GAs, EP, and ESs can all be explained as global population-based random search algorithms with several features in common: 1) all aim to maximize fitness of individual; 2) they yield population sequences whose elitist or average fitness is monotonically improved; and 3) they are based on the same principle of natural evolution—applying a "strive for survival" selection process and using "evolutionary" operators. These common features make it direct and possible to abstract a more general unified SEA model. Second, the convergence of SEAs should be a natural consequence of implementation of "survival of the fittest." This already has been justified in some special cases of GAs, EP, and ESs (e.g., [15], [23]). By identifying the core reason why these special implementation schemes converge, we expect to be able to find the intrinsic mechanism (general conditions) of the evolution that implies convergence of a general SEA. The present work seeks to characterize such general conditions of convergence for a general SEA, in terms of involved selection and evolution operators.

There have been several approaches to conduct convergence analyzes of SEAs. These include the simulated annealing-like (SA-like) approach, the Vose–Liepins model, and the stochastic model approach.

In the SA-like approach, an SEA is treated as an instance of the standard simulated annealing (SA) algorithm via viewing a certain parameter of the SEA as *temperature*, and convergence of the SEA is then deduced from a direct extrapolation of the existing convergence theory of SAs [2]–[4], [9]. This approach offers an important insight into SEAs in the sense that these algorithms can be regarded as also simulating certain physical principles (while simulating the biological evolution). However, since the SEA looks rather like generalization than an instance of SA algorithms, the convergence assertions that such an approach have deduced are very limited. Indeed, the existing results from this approach only pertain to several GA models (see [4]), where the mutation rate is treated as *temperature*, and therefore are required to tend to zero. Moreover, the obtained

1089–778X/01$10.00 © 2001 IEEE

results are not only based on the strong ergodicity of the GA, but also on the assumption that, for each fixed rate (*temperature*), the evolution of GA is an infinite stochastic process. The latter assumption is clearly far from practical in modeling any real implementations of GAs.

The approach based on the Vose–Liepins model is to cast an SEA (particularly, a GA), a stochastic process, as a deterministic (i.e., nonstochastic) dynamical system (more precisely, a successive approximation iterative process determined by a deterministic nonlinear mapping), then to make a correspondence between the convergence of the SEA and the convergence of the system to the equilibrium set (i.e., the fixed points of the nonlinear mapping). Such an approach was developed by Vose and Liepins [27], and then extended and applied in [12], [29], and [16]. The most attractive feature of this approach is that the model used to cast the SEA is exact and deterministic, so many functional and numerical analysis tools can be applied directly. The difficulty is, however, that the precise expression of the related nonlinear mapping can only be obtained for very few cases (e.g., a binary GA with proportional selection); even in these cases, it is not clear how the fixed-point set of the mapping is related to the global optimum that we have expected to look for. Thus, the approach may not be applied to analysis of the GAs, for instance, when the encoding is real or the selection operator is nonproportional. Even though convergence of an SEA is concluded from this approach, it is still open whether or not the algorithm converges to the expected global optimum. It is noted also that most existing analyzes based on the Vose–Liepins model, have an assumed infinite population size [27], [12], [29]. An analysis based on the Vose–Liepins model, but for finite population GAs, was developed by Nix and Vose [16]. They obtained, in particular, the upper bound on the number of iterations for several special implementation schemes of a GA to traverse all states of the search space at a certain level of confidence. The result derived was still based on ergodicity analysis of the considered algorithms.

The stochastic model analysis perhaps is the most natural and widely used approach for convergence analysis of SEAs [4], [5], [11], [18]–[20], [25]. Owing to the fact that most existing SEAs can be modeled as Markov chains in a very natural way, this approach can be applied to SEAs with any finite population based on any kind of selection mechanism and evolution operators. The most difficult issue, however, is that it is impossible or at least impractical in general to formulate the details of the related transition probability matrix and, therefore, analyzing the properties of the matrix is difficult. Almost all existing Markov chain analysis results are deduced based on the ergodicity analysis of the corresponding algorithms. Therefore, convergence is directly derived for those algorithms with an elitist selection strategy. For instances, Rudolph [18] proved that the canonical GAs (CGAs) cannot converge to the global optimum if the mutation rate is not zero, but any CGA with elitist selection converges. These results were later extended by Suzuki [25] for CGAs with a modified elitist strategy. Within an appropriate stochastic process framework, models other than the Markov chain approach for an SEA are also discussed. Rudolph applied the theory of supermartingales to ESs, and obtained a rather general convergence criterion for a class of convex fitness functions in [19] and [20], which is valid for both elitist-type and nonelitist-type ESs.

From the above brief exposition, we can conclude that almost all existing convergence results of SEAs are actually validated only for their "elitist" versions and are based on ergodicity analysis of the algorithms. However, we remark that the convergence of the elitist version of an SEA by no means implies the convergence of the SEA itself since elitist selection is only a recording strategy for search results of the algorithm, which has nothing to do with the simulating evolution process. So the convergence issue of a general SEA is still open (except for Miller [15], Suzuki [25], and Peck and Dhawan [17]). Furthermore, it should be noted that, although convergent, the elitist-type SEA does not always deserve recommendation in use. Salomon [21] noted that from a computational point of view, elitist selection, which is not a local operator, requires global communication in each generation. This requires at least $O(\ln N)$ additional operations. This additional runtime factor can cause low performance of algorithms in distributed and parallel systems without global communication or without a central processing unit. Thus, the study of convergence of the nonelitist-type SEAs is important both from the viewpoint of theoretical significance and from the perspective of parallel computation.

This paper focuses on the study of convergence for general nonelitist-type SEAs using a completely new approach. The main points behind this approach include the following.

- Any general evolution process is modeled as an abstract stochastic process that is deduced from two fundamental stochastic operators—selection and evolution operators.
- Through identifying key components of the existing SEAs, the two fundamental stochastic operators are modeled axiomatically and the functions of the operators are characterized quantitatively, more precisely, through specifying a set of six parameters describing notions such as selection pressure, selection intensities, evolution stability rate, evolution aggregating rate, and evolution scattering rate.
- Based on mathematical estimations on the probability of the population containing the global optimum (in terms of the parameters mentioned above), the convergence and convergence rate for general SEAs are treated simultaneously.

The proposed approach is not based on ergodicity analysis of the related algorithms and, therefore, could be regarded as a "nonergodicity-based" one. By this approach, we can develop an abstract general SEA model (called an abstract evolutionary algorithm—AEA), which unifies most of the currently known SEAs. Also, we establish a unified convergence theory of the nonelitist versions of various existing implementation schemes of SEAs.

The paper is organized as follows. In Section II, we develop the generalized SEA model—AEA—through describing the fundamental stochastic operators and quantitatively characterizing their functions axiomatically. Two types of useful convergence definitions of SEAs are then introduced. Based on the introduced abstract SEA model, several general convergence theorems and convergence rate estimations are established in Section III. In Section IV, a series of examples

of selection and evolution operators are provided that show, in particular, how the related parameters like selection pressure and evolution aggregating rate can be specified precisely. We also apply the established general convergence theorems to various known SEAs, showing the utility of the proposed approach. The paper is then concluded in Section V.

II. The Abstract Evolutionary Algorithm

This section presents an abstract model of the simulated evolutionary algorithm (SEA), called the abstract evolutionary algorithm (AEA). We first explain the optimization problem studied in Section II-A. Then, in Section II-B, we formulate the axioms of the fundamental evolutionary operators and the abstract evolutionary algorithm (AEA). A set of related characteristic parameters of the fundamental evolutionary operators that characterizes their functions quantitatively is also introduced here. In Section II-C, two kinds of global convergence definitions that will be used later are presented.

A. Optimization Problem

Consider the optimization problem

$$\max\{g(x);\, x \in \Omega\} \tag{1}$$

where g: $\Omega \to R$ is the fitness function and Ω is the feasible region. Without loss of generality, assume that $g(x) \geq 0$ for any $x \in \Omega$ and Ω is a discrete space.

The aim is to find the global optimum of g or, more precisely, to find such $x^* \in \Omega$ such that $g(x^*) = \max\{g(x); x \in \Omega\}$. Such a task is difficult or even impossible in general. Therefore, instead of x^*, we will focus on finding a satisfactory solution of the problem (1).

Definition 2.1: A nonempty subset $B \subset \Omega$ is called a satisfactory set of the problem (1) if $g(a) > g(b)$ for any $a \in B$ and $b \in \Omega \backslash B$. The collection of all satisfactory sets of (1) is denoted by $\mathcal{T}$.

A satisfactory set B is clearly such a subset of Ω whose elements have higher fitness than any others outside B. The present paper derives the conditions under which an SEA can converge to a satisfactory set. It is evident that a satisfactory set B is the global optimum set if and only if $g(a) = g(b)$ for any a, $b \in B$. Furthermore, an algorithm that converges to any satisfactory sets must be convergent to the optimum set.

B. The Axiomatic Model of SEA: AEA

For simplicity, we treat each element of Ω as an individual. Thus, Ω is the individual space of SEA, and Ω^N is the population space whenever the population size is N. For any stochastic operator T and two populations X and Y, we use $P_T(X;\, Y)$ to denote the probability of $T(X)$ being Y.

To formulate SEAs axiomatically, we observe that for almost all known SEAs, the algorithm's mechanism of generating the next population can be abstracted as two independent procedures that can be represented by two operators: selection and evolution. The selection operator works on the fitness value of individuals in the current population, and assigns the probability of survival for each individual. Then, the evolution operator, independent of the fitness function, manages the search strategy and decides the sites of the next population based only on the current population sites. For an effective SEA, the selection and evolution operators are both designed to improve the fitness of populations generation by generation for all types of fitness functions. Inspired by this observation, we will model the general SEA as a stochastic process deduced from the two independent stochastic operators: the abstract selection and evolution operators, which are detailed, respectively, as follows.

Axiom 2.1 (Selection Operator): A stochastic function S: $\Omega^N \to \Omega^N$ is an abstract selection operator if

1) $S(X) \subset X$ for any $X \in \Omega^N$;
2) there is a positive constant p such that for each $X \in \Omega^N$ with $|X_M| < N$

$$P_S(X;\, |S(X)_M| > p + |X_M|) \neq 0;$$

3) for any fixed $p > 0$

$$P_S(X;\, |S(X)_M| > p) \geq \inf\{P_S(X;\, |S(X)_M| > p);\, |X_M| = 1\}$$

where $M = \max\{g(x);\, x \in X\}$, $Y_M = \{x \in Y;\, g(x) = M\}$ for each $Y \in \Omega^N$, $|Y|$ denotes the cardinality of the set Y, and $P[\cdot]$ is the probability.[1]

Axiom 2.1 characterizes three fundamental features of the abstract selection operator: 1) selection must be carried out within the current population; 2) selection should, with a positive probability, increase the number of higher fitness individuals whenever the current population is not degenerate (i.e., consisting of individuals of equal fitness); and 3) the more the fittest individuals exist in the current population, the greater the number of increased fittest individuals in the selected population.

Axiom 2.2 (Evolution Operator): A stochastic function E: $\Omega^N \to \Omega^N$ is an abstract evolution operator if, for each $X \in \Omega^N$ and $B \in \mathcal{T}$

a) $P_E(X;\, E(X) \cap B \neq \phi) \neq 0$ whenever $X \cap B = \phi$;
b) $P_E(X;\, E(X) \cap B = \phi) \neq 1$ whenever $X \cap B \neq \phi$; furthermore, E is an abstract strong evolution operator if E additionally satisfies
c) $P_E(X;\, |E(X) \cap B| \geq |X \cap B|) \neq 0$.

Property a) in Axiom 2.2 shows that the evolution is capable of reaching any satisfactory set from anywhere in Ω^N, whereas b) shows that, once a satisfactory set is reached (by the current search), then the evolution should, with probability 1, prevent the individuals of the population from escaping it totally. c) indicates that if E is a strong evolution operator, then it can surely increase the probability of a population to meet any fixed satisfactory set.

Fogel [8] noted that all evolutionary optimization algorithms can be described as a series of operators applied to a population of candidate solutions. Particularly, in terms of the above-introduced selection and evolution operators, an abstract SEA model—abstract evolution algorithm (AEA)—can then be defined as follows.

Axiom 2.3 (AEA): An AEA is a stochastic process, deduced from a sequence of abstract selection operators $\{S(t);\, t \geq 1\}$

[1]Here and in what follows, the notation $P_S(X;\, A)$ denotes the probability $\sum_{Y \in A} P_S(X;\, Y)$, where A is an event. Whenever no confusion is caused, we also often omit the subscript S in $P_S(X;\, A)$.

and evolution operators $\{E(t);\, t \geq 1\}$, whose population at time t is defined by

$$X^{(t)} = E(t) \circ S(t)\left(X^{(t-1)}\right)$$

where $X^{(0)}$ is the initial population randomly chosen from Ω^N.

We will show in Section IV that almost all known SEAs are concrete examples of the AEA introduced here.

To analyze such an axiomatically defined AEA, we need to introduce the related characteristic parameters of the selection and evolution operators.

Definition 2.2 (Selection Pressure and Selection Intensities): Let S be an abstract selection operator; the selection pressure P_S, the selection intensity I_S, and the uniform selection intensity U_S of S are positive real numbers such that

$$\begin{aligned} P_S &= \sup\{p \in R^+;\, P(X;\, |S(X)_M| \geq p+1) \neq 0; \\ &\qquad |X_M| = 1\} \\ I_S &= \inf\{P(X;\, |S(X)_M| \geq P_S + 1);\, |X_M| = 1\} \\ U_S &= \inf\{P(X;\, |S(X)_M| > |X_M|);\, |X_M| < N\}. \end{aligned}$$

The selection pressure provides a quantitative measure of the maximal capability of S in increasing the number of the fittest individual (note that for fair comparison purposes, we have assumed $|X_M| = 1$, which is obviously necessary). The selection intensity provides a measure of the smallest probability of S maintaining the selection pressure. Likewise, the uniform selection intensity measures the infimum probability of S uniformly increasing the number of the best individuals for any nondegenerate populations (that is, any population X such that $|X_M| < N$).

Definition 2.3 (Aggregating Rate, Scattering Rate, and Stability Rate): Let E be an abstract evolution operator; the aggregating rate A_E and the scattering rate S_E of E are positive real numbers such that

$$\begin{aligned} A_E &= \inf\{P(X;\, E(X) \cap B \neq \phi);\, X \cap B = \phi,\, B \in \mathcal{T}\} \\ S_E^{(r)} &= \sup\{P(X;\, E(X) \cap B = \phi);\, |X \cap B| \geq r,\, B \in \mathcal{T}\}. \end{aligned}$$

Moreover, if E is an abstract strong evolution operator, its stability rate ST_E is defined by

$$\begin{aligned} ST_E = \inf\{&P(X;\, |E(X) \cap B| \geq |X \cap B|); \\ &X \in \Omega^N,\, B \in \mathcal{T}\}. \end{aligned}$$

The parameters A_E, $S_E^{(r)}$, and ST_E defined above provide quantitative measures of global capability of E from the aspects, respectively, specified by a)–c) of Axiom 2.2. In particular, A_E measures the exploitation (convergence) capability of E in searching and aggregating to a fixed satisfactory set, $S_E^{(r)}$ measures the exploration capability (diversity) of E forcing a population to leave a satisfactory set and scattering it around, and ST_Ethen measures the stability degree of E preserving the satisfactory individuals found.

The significance of introducing the parameters $\{P_S, I_S, U_t, A_E, S_E^r, ST_E\}$ is not only that they provide quantitative characterizations of the fundamental selection and evolution operators of AEA, which makes it possible to develop a unified convergence theory for SEAs, but also that they offer a unified quantitative criterion for fairly comparing existing (different) SEAs. We will present a series of concrete examples to illustrate this point in Section IV.

C. Convergence Definitions

We will study two types of global convergence of AEA. The definitions are given as follows.

Definition 2.4: An AEA is said to be quasi-convergent if $\lim_{t\to\infty} P[X^{(t)} \cap B \neq \phi] = 1$ for each satisfactory set $B \in \mathcal{T}$. The AEA is said to be convergent if $\lim_{t\to\infty} P[X^{(t)} \subset B] = 1$ for each satisfactory set $B \in \mathcal{T}$.

By this definition, an AEA is quasi-convergent if and only if with probability 1 its population eventually intersects each satisfactory set. That is, the probability of a global maximum appearing in the population will tend to 1. By comparison, an AEA being convergent means that its population will be included in any satisfactory set in the end, so that the population must evolve gradually to be the degenerate one consisting of only global maximum. A convergent AEA is naturally quasi-convergent, but the reverse is not true in general.

There are other convergence notions used in literature. For instance, *almost sure convergence* [19] reads as follows. An AEA is almost surely convergent if $P[\lim_{t\to\infty} d(OPT, X^{(t)}) = 0] = 1$, where OPT is the optimal solution of (1) and $d(\cdot, \cdot)$ is the Hausdorff distance of sets on R^N. This convergence notion is slightly stronger than ours in the general case, but in the situations that either Ω is discrete or g and Ω both are continuous, it can be shown that this notion is equivalent to ours. For other convergence definitions, see [17].

III. Global Convergence for AEAs

This section presents a detailed convergence analysis of AEA. We will particularly formulate general convergence conditions, and provide convergence speed estimations of AEA in terms of the introduced characteristic parameters of the selection and evolution operators. The specifications of the results will be presented in the next section.

Let $X^{(t)}$ denote the population generated by AEA at time t through application of the selection operator $S(t)$ and evolution operator $E(t)$ according to Axiom 2.3. The selection pressure, the selection intensity, and the uniform selection intensity of $S(t)$ are denoted, respectively, by P_t, I_t, and U_t. The evolution aggregating rate, the evolution scattering rate, and the evolution stability rate of $E(t)$ are then denoted by A_t, $S_t^{(r)}$, and ST_t, respectively.

A. Quasi-Convergence

We first characterize the quasi-convergence conditions of AEA. We need the following lemma.

Lemma 3.1: Let $\{p_k;\, k > 0\}$, $\{a_k;\, k > 0\}$, and $\{b_k;\, k > 0\}$ be three nonnegative real sequences such that:

a) for each $k > 0$, $a_k, b_k, p_k \in (0, 1)$ and $p_{k+1} \leq a_k + b_k \cdot p_k$
b) $\lim_{k\to\infty}(a_k/1 - b_k) = 0$
c) $\sum_{k=1}^{\infty}(1 - b_k) = \infty$.

Then, $\lim_{k\to\infty} p_k = 0$. Furthermore, $p_n < \epsilon$ for every $n > R+T$ where R and T are integers such that $\prod_{k=R}^{R+T} b_k < \epsilon/2$ and $(a_k/1-b_k) < \epsilon/2$ for every $k \geq R$.

Proof: Since $\lim_{k\to\infty}(a_k/1-b_k) = 0$, for any $\epsilon > 0$, there exists an integer K such that, whenever $k \geq K$,

$$a_k < \epsilon \cdot (1-b_k).$$

Therefore, from condition a), we have

$$\begin{aligned}(p_{k+1}-\epsilon) - b_k \cdot (p_k - \epsilon) &= p_{k+1} - b_k \cdot p_k - \epsilon \cdot (1-b_k) \\ &\leq p_{k+1} - b_k \cdot p_k - a_k \\ &\leq 0.\end{aligned}$$

That is

$$p_{k+1} - \epsilon \leq b_k \cdot (p_k - \epsilon).$$

This implies that, for any $n > K$,

$$p_n \leq \epsilon + \left(\prod_{l=K}^{n-1} b_l\right) \cdot (p_K - \epsilon).$$

Note that condition c) implies $\prod_{k=1}^{\infty} b_k = 0$, which shows that $\lim_{n\to\infty}\prod_{l=K}^{n-1} b_l = 0$ and, hence, $\limsup_{k\to\infty} p_k \leq \epsilon$. As ϵ is arbitrary and $p_k \geq 0$, therefore $\lim_{k\to\infty} p_k = 0$.

If $(a_k/1-b_k) < \epsilon/2$ for every $k \geq R$ and $\prod_{k=R}^{R+T} b_k < \epsilon/2$, replacing K and ϵ in the preceding inequality by R and $\epsilon/2$, respectively, we have $p_n \leq \epsilon/2 + (\prod_{l=R}^{n-1} b_l)\cdot(p_R - \epsilon/2) \leq \epsilon/2 + \prod_{l=R}^{n-1} b_l$. It follows from $b_l \leq 1$ that $p_n \leq \epsilon/2 + \prod_{l=R}^{R+T} b_l < \epsilon$ whenever $n > R+t$. This finishes the proof. □

By making use of the above lemma, we now can prove the following basic theorem.

Theorem 3.1 (Quasi-convergence): An AEA is quasi-convergent if it satisfies the following conditions:

1) there is an $m > 0$ such that $P_t \geq m$ for each $t > 0$
2) $\lim_{t\to\infty}(1 - I_t \cdot (1 - S_t^{(m+1)}))/A_t = 0$
3) $\sum_{t=1}^{\infty} A_t = \infty$.

In this case, for any $\epsilon > 0$, the AEA obeys the following convergence speed estimation:

$$P\left[X^{(t)} \cap B \neq \phi\right] > 1 - \epsilon$$

for any $t > R+T$, where R, T are integers such that $\prod_{t=R}^{R+T} A_t < \epsilon/2$ and $(1 - I_t \cdot (1 - S_t^{(m+1)}))/A_t < \epsilon/2$ whenever $t \geq R$.

Proof: For any fixed satisfactory set B, let p_i^t denote the probability at which there are i individuals in the intersection of $X^{(t)}$ and B, that is

$$p_i^t = P\left[\left|X^{(t)} \cap B\right| = i\right]. \tag{2}$$

Clearly, it suffices to show that $p_0^t \to 0$ as $t \to \infty$.

By Bayesian formula, we first can express

$$\begin{aligned}p_0^{t+1} &= P\left[X^{(t+1)} \cap B = \phi\right] \\ &= P\left(X^{(t)} \cap B \neq \phi;\, X^{(t+1)} \cap B = \phi\right) \\ &\quad \cdot P\left[X^{(t)} \cap B \neq \phi\right] \\ &\quad + P\left(X^{(t)} \cap B = \phi;\, X^{(t+1)} \cap B = \phi\right) \cdot p_0^t.\end{aligned}$$

Consequently

$$\begin{aligned}p_0^{t+1} &\leq P\left(X^{(t)} \cap B \neq \phi;\, X^{(t+1)} \cap B = \phi\right) \\ &\quad + \left(X^{(t)} \cap B = \phi;\, X^{(t+1)} \cap B = \phi\right) \cdot p_0^t. \qquad (3)\end{aligned}$$

Let $Y = S(t+1)(X^{(t)})$. Then, by Axiom 2.1, $Y \subset X^{(t)}$, and hence $\{Y \in \Omega^N;\, Y = S(t+1)(X^{(t)}),\, X^{(t)} \cap B = \phi\} \subset \{Y \in \Omega^N;\, Y \cap B = \phi\}$. It then follows from the definition of evolution aggregating rate (Definition 2.3) that

$$\begin{aligned}&P\left(X^{(t)} \cap B = \phi;\, X^{(t+1)} \cap B = \phi\right) \\ &= P\left(Y = S(t+1)\left(X^t\right),\, X^{(t)} \cap B = \phi;\right. \\ &\qquad \left. E(t+1)(Y) \cap B = \phi\right) \\ &\leq \sup\{P(Y;\, E(t+1)(Y) \cap B = \phi); \\ &\qquad X^t \cap B = \phi,\, Y = S(t+1)(X^t)\} \\ &\leq \sup\{P(Y;\, E(t+1)(Y) \cap B = \phi);\, Y \cap B = \phi\} \\ &= 1 - \inf\{P(Y;\, E(t+1)(Y) \cap B \neq \phi);\, Y \cap B = \phi\} \\ &= 1 - A_{t+1}. \qquad (4)\end{aligned}$$

Also, let $M = \max\{g(x);\, x \in X^{(t)}\}$ and $\overline{B} = \{x \in B;\, g(x) \geq M\}$. Then, according to the Bayesian formula, we have

$$\begin{aligned}&P\left(X^{(t)} \cap B \neq \phi;\, X^{(t+1)} \cap B = \phi\right) \\ &= P\left(|Y \cap \overline{B}| \geq m+1,\, X^{(t)} \cap B \neq \phi;\right. \\ &\qquad \left. X^{(t+1)} \cap B = \phi\right) \\ &\quad \cdot P\left(X^{(t)} \cap B \neq \phi;\, |Y \cap \overline{B}| \geq m+1\right) \\ &\quad + P\left(|Y \cap \overline{B}| \leq m,\, X^{(t)} \cap B \neq \phi;\right. \\ &\qquad \left. X^{(t+1)} \cap B = \phi\right) \\ &\quad \cdot P\left(X^{(t)} \cap B \neq \phi;\, |Y \cap \overline{B}| \leq m\right) \\ &\leq P\left(|Y \cap \overline{B}| \geq m+1,\, X^{(t)} \cap B \neq \phi;\right. \\ &\qquad \left. X^{(t+1)} \cap B = \phi\right) \\ &\quad \cdot P\left(X^{(t)} \cap B \neq \phi;\, |Y \cap \overline{B}| \geq m+1\right) \\ &\quad + P\left(X^{(t)} \cap B \neq \phi;\, |Y \cap \overline{B}| \leq m\right). \qquad (5)\end{aligned}$$

Since $Y \subset X^{(t)}$ and $\overline{B} \subset B$, the condition $|Y \cap \overline{B}| \geq m+1$ implies $X^{(t)} \cap B \neq \phi$. From the definition of the evolution scattering rate of $E(t+1)$, we have

$$\begin{aligned}&P\left(|Y \cap \overline{B}| \geq m+1,\, X^{(t)} \cap B \neq \phi;\, X^{(t+1)} \cap B = \phi\right) \\ &\leq \sup\{P(Y;\, E(t+1)(Y) \cap B = \phi);\, |Y \cap \overline{B}| \geq m+1\} \\ &= S_t^{(m+1)}. \qquad (6)\end{aligned}$$

Moreover, from the definitions of $\overline{B}$ and conditional probability, we have

$$\begin{aligned} &P\left(X^{(t)}\cap B\neq\phi;\,|Y\cap\overline{B}|\geq m+1\right)\\ &\quad=P\left(X^{(t)}\cap B\neq\phi;\,\left|\left(S(t+1)\left(X^{(t)}\right)\right)_M\right|\geq m+1\right)\\ &\quad=\sum_{|Z_M|\geq 1,\,Z\cap B\neq\phi}P(Z;\,|(S(t+1)(Z))_M|\geq m+1)\\ &\qquad\cdot\frac{P\left[X^{(t)}=Z\right]}{P\left[X^{(t)}\cap B\neq\phi\right]} \end{aligned}\tag{7}$$

which implies

$$\begin{aligned} &P\left(X^{(t)}\cap B\neq\phi;\,|Y\cap\overline{B}|\geq m+1\right)\\ &\quad\geq\inf\{P(Z;\,|(S(t+1)(Z))_M|\geq m+1);\\ &\qquad\quad|Z_M|\geq 1,\,Z\cap B\neq\phi\}\\ &\quad\geq\inf\{P(Z;\,|(S(t+1)(Z))_M|\geq m+1);\,|Z_M|\geq 1\} \end{aligned}\tag{8}$$

because of the fact that

$$\sum_{Z\cap B\neq\phi}\frac{P\left[X^{(t)}=Z\right]}{P\left[X^{(t)}\cap B\neq\phi\right]}=1.$$

By using Axiom 2.1c), the definitions of selection pressure and selection intensity (Definition 2.2) and the assumption a), the above inequality then further shows

$$\begin{aligned} &P\left(X^{(t)}\cap B\neq\phi;\,|Y\cap\overline{B}|\geq m+1\right)\\ &\quad\geq\inf\{P(Z;\,|(S(t+1)(Z))_M|\geq m+1);\,|Z_M|\geq 1\}\\ &\quad\geq\inf\{P(Z;\,|(S(t+1)(Z))_M|\geq (P_{t+1}+1));\,|Z_M|\geq 1\}\\ &\quad=I_{t+1}. \end{aligned}\tag{9}$$

Bringing (6) and the foregoing estimation (9) into (5), we obtain

$$\begin{aligned} &P\left(X^{(t)}\cap B\neq\phi;\,X^{(t+1)}\cap B=\phi\right)\\ &\quad\leq S_{t+1}^{(m+1)}\cdot P\left(X^{(t)}\cap B\neq\phi;\,|Y\cap\overline{B}|\geq m+1\right)\\ &\qquad+P\left(X^{(t)}\cap B\neq\phi;\,|Y\cap\overline{B}|\leq m\right)\\ &\quad=1-P\left(X^{(t)}\cap B\neq\phi;\,|Y\cap\overline{B}|\geq m+1\right)\\ &\qquad\cdot\left(1-S_{t+1}^{(m+1)}\right) \end{aligned}$$

that is

$$\begin{aligned} &P\left(X^{(t)}\cap B\neq\phi;\,X^{(t+1)}\cap B=\phi\right)\\ &\quad\leq 1-I_{t+1}\cdot\left(1-S_{t+1}^{(m+1)}\right). \end{aligned}\tag{10}$$

Combining the estimations (4) and (10) with (3) yields

$$p_0^{t+1}\leq 1-I_{t+1}\cdot\left(1-S_{t+1}^{(m+1)}\right)+(1-A_{t+1})\cdot p_0^t.\tag{11}$$

By applying Lemma 3.1 to the above estimation (11), it then follows that $\lim_{t\to\infty}p_0^t=0$, and the convergence speed estimation is established. This completes the proof of Theorem 3.1. □

Remark 3.1: The quasi-convergence conditions formulated in Theorem 3.1 are quite intuitive: a) says that for a quasi-convergent AEA, its selection pressure must be positive; c) says that its evolution aggregating rate may be very small (even asymptoctically tending to zero), but should not be too small or tend to zero too fast; and b) then shows that, once the selection intensity is near one (which is the case in most applications), the scattering rate should be a higher order infinitesimal of the aggregating rate. Properties b) and c) provide us with a quantitative strategy for balancing the exploration/exploitation dilemma that exists in SEAs. All of these explanations reveal that conditions a)–c) are indeed very natural.

B. Convergence

Now, we further show when a quasi-convergent AEA becomes convergent.

Theorem 3.2 (Convergence): A quasi-convergent AEA is convergent if, for each $t\geq 1$, $E(t)$ is an abstract strong evolution operator and $\lim_{t\to\infty}U_t\cdot ST_t=1$. In this case, for any satisfactory set B, the convergent speed can be estimated as

$$\begin{aligned} &P\left[X^{(t+N)}\subset B\right]\\ &\quad\geq\left(\prod_{s=t+1}^{t+N}ST_s\cdot U_s\right)\cdot P\left[X^{(t)}\cap B\neq\phi\right]. \end{aligned}$$

Proof: As in the proof of Theorem 3.1, we let $p_i^t=P[|X^{(t)}\cap B|=i]$ for any satisfactory set $B\in\mathcal{T}$. It is observed first that the quasi-convergence of the AEA implies that $\lim_{t\to\infty}P[X^{(t)}\cap B\neq\phi]=1$ and, therefore, $\lim_{t\to\infty}p_0^t=0$. We now proceed by showing that for any $l=1,2,\cdots,N$, the following equality holds:

$$\lim_{t\to\infty}\sum_{i=l}^{N}p_i^t=1\tag{12}$$

which then yields the convergence of the AEA. Equation (9) is trivially true for $l=1$ since $\lim_{t\to\infty}p_0^t=0$ and $\sum_{i=0}^{N}p_i^t=1$. If we suppose that (9) holds for any integer up to $l>0$, then, by the definitions of the uniform selection intensity and the evolution stability rate, we find

$$\begin{aligned} &\sum_{i=l+1}^{N}p_i^{t+1}\\ &\quad\geq P\left[\left|X^{(t+1)}\cap B\right|\geq l+1\right]\\ &\quad\geq P\left[\left|X^{(t+1)}\cap B\right|\geq(l+1)\,\middle|\,\left|X^{(t)}\cap B\right|\geq l\right]\\ &\qquad\cdot P\left[\left|X^{(t)}\cap B\right|\geq l\right]\\ &\quad\geq P\left[\left|X^{(t+1)}\cap B\right|\geq(l+1)\,\middle|\,\left|X^{(t)}\cap B\right|\geq l\right]\sum_{i=l}^{N}p_i^t \end{aligned}$$

and

$$\begin{aligned} &P\left[\left|X^{(t+1)}\cap B\right|\geq(l+1)\,\middle|\,\left|X^{(t)}\cap B\right|\geq l\right]\\ &\quad\geq P\left[|E(t+1)(Y)\cap B|\geq|Y\cap B|\,|\,|Y\cap B|\geq l+1,\right.\\ &\qquad\left.\left|X^{(t)}\cap B\right|\geq l\right]\\ &\qquad\cdot P\left[|Y\cap B|>\left|X^{(t)}\cap B\right|\,\middle|\,\left|X^{(t)}\cap B\right|\geq l\right]\\ &\quad\geq\inf\{P[|E(t+1)(Y)\cap B|\geq|Y\cap B|];\\ &\qquad Y\in\Omega^N,B\in\mathcal{T}\} \end{aligned}$$

$$\cdot \inf\{P[|S(t+1)(X)\cap B| > |X\cap B|]; X\in\Omega^N, B\in\mathcal{T}\} = ST_{t+1}\cdot U_{t+1}$$

where, as before, $Y = S(t+1)(X^{(t)})$. This then implies

$$1 \geq \sum_{i=l+1}^{N} p_i^{t+1} \geq (ST_{t+1}\cdot U_{t+1})\sum_{i=l}^{N} p_i^t. \tag{13}$$

By the assumptions, it then follows that $\lim_{t\to\infty}\sum_{i=l+1}^{N} p_i^{t+1} = 1$. That is, the equality (12) holds for any $l = 1, 2, \cdots, N$ as claimed.

From (9) and (10), we have

$$P\left[X^{(t+N)}\subset B\right] = p_N^{(t+N)} \geq \prod_{s=t+1}^{t+N}(ST_s\cdot U_s)\cdot P\left[X^{(t)}\cap B\neq\phi\right]$$

which yields $\lim_{t\to\infty} P[X^{(t)}\subset B] = 1$ and the estimation required. This finishes the proof of Theorem 3.2. □

Remark 3.2: Theorem 3.2 can be also explained very intuitively: since the AEA is quasi-convergent, there is at least one individual in the AEA population which is in any given satisfactory set B from a sufficiently large t onwards. The selection operator gradually increases the number of these satisfactory individuals, and the strong evolution operator then ensures that these individuals become more and more stable (i.e., it becomes more and more difficult for them to leave the population). As a result, as the AEA population evolves, it must be totally contained in the satisfactory set and, hence, the convergence of the AEA follows.

IV. Examples and Applications

This section is devoted to specifications of the theories developed in previous sections to various known SEAs. In Sections IV-A and B, the characteristic parameters of various widely used selection operators and evolution operators are evaluated, respectively. The evaluation results obtained are then used in Section IV-C to derive the convergence and convergence speed estimations of genetic algorithms. In Section IV-D, Theorems 3.1 and 3.2 are specified to a class of evolutionary strategies.

A. Selection Operators

Assume that $X\in\Omega^N$ is a population, $X = (x_1, x_2, \cdots, x_N)$ with each $x_i\in\Omega\,(i = 1, 2, \cdots, N)$ being an individual. Given a selection operator S: $\Omega^N\to\Omega^N$, we denote $Y = S(X)$ and write $Y = (y_1, y_2, \cdots, y_N)$.

Example 4.1 (Proportional Selection [14]): By this selection rule, the selected population Y is created from N independent random experiments. In the ith experiment, the individual $x\in X$ is selected according to the probability

$$P[y_i = x] = \frac{\sigma(g(x))}{\sum_{y\in X}\sigma(g(y))}$$

where g is the fitness (or objective) function and σ: $R\to R^+$ is a gauge (namely, a strictly increasing positive function). Let

$$\rho = \max\left\{\frac{\sigma(g(x))}{\sigma(g(y))};\ \sigma(g(x)) < \sigma(g(y)),\ x, y\in\Omega\right\}. \tag{14}$$

Then the selection pressure P_S, the selection intensity I_S, and the uniform selection intensity U_S of S are as follows:

$$P_S = N-1;\quad I_S = (1+(N-1)\rho)^{-N}$$
$$U_S \geq (1+(N-1)\rho)^{-N}.$$

Proof: We can assume that $|X_M| = 1$, and x_1 is the unique fittest individual. Since Y is created from N independent random experiments, we have

$$P(|X_M| = 1;\ |Y_M| \geq (N-1)+1) = \alpha^N > 0$$

where

$$\alpha = \frac{\sigma(g(x_1))}{\sum_{i=1}^{N}\sigma(g(x_i))}.$$

Therefore, $P_S = N-1$. Noting that (14) implies

$$\alpha \geq \inf_{X\in\Omega^N}\frac{\sigma(g(x_1))}{\sum_{i=1}^{N}\sigma(g(x_i))} = \inf_{X\in\Omega^N}\frac{1}{1+\sum_{i=2}^{N}\frac{\sigma(g(x_i))}{\sigma(g(x_1))}} = (1+(N-1)\rho)^{-1}$$

we obtain $I_S = (1+(N-1)\rho)^{-N}$. To estimate the uniform selection intensity U_S, we let $|X_M| < N$ and assume that β is the probability of that an individual x_i in X_M is selected in one independent experiment. Then it follows that

$$\beta = \sum_{x_i\in X_M}\frac{\sigma(g(x_i))}{\sum_{j=1}^{N}\sigma(g(x_j))} \tag{15}$$

$$= \sum_{x_i\in X_M}\frac{1}{|X_M| + \sum_{y\notin X_M}\frac{\sigma(g(y))}{\sigma(g(x_i))}}. \tag{16}$$

By Newton's formula, this then implies that, in the N independent experiments, the probability of more than $|X_M|$ individuals in X_M being selected into Y is

$$\sum_{i=|X_M|+1}^{N} C_N^i\beta^i(1-\beta)^{N-i} \geq C_N^N\beta^N = \beta^N.$$

That is

$$P(|X_M| < N;\ |Y_M| > |X_M|) \geq \beta^N. \tag{17}$$

TABLE I
THE CHARACTERISTIC PARAMETERS OF SELECTION OPERATORS; P_S: SELECTION PRESSURE, I_S: SELECTION INTENSITY, AND U_S: UNIFORM SELECTION INTENSITY

Selection Operator	P_S	I_S	U_S
elitist	1	1	1
linear ranking	$N-1$	$(\frac{1}{N}+\frac{N-1}{2}q)^N$	$\geq (\frac{1}{N}+\frac{N-1}{2}q)^N$
nonlinear ranking	$N-1$	q^N	$\geq q^N$
tournament	$s-1$	1	1
proportional	$N-1$	$(1+(N-1)\rho)^{-N}$	$\geq (1+(N-1)\rho)^{-N}$
Boltzmann	1	$(1+e^{-\frac{\Delta}{T}})^{-2}$	$(1+e^{-\frac{\Delta}{T}})^{-2}$

From (11) and (13), β attains its infimum $|X_M|+(N-|X_M|)\rho$. Consequently, (17) implies

$$U_S \geq \min_{1\geq n\geq N-1}\left(\frac{n}{n+(N-n)\rho}\right)^N = \left(\frac{1}{1+(N-1)\rho}\right)^N. \qquad \square$$

There are many different selection operators that are similar to proportional selection in the sense that they all generate the selected population Y by N independent random experiments and in each experiment, the fittest individual in X is chosen with a fixed positive probability. Consequently, selection pressures and selection intensities of these selection operators can be computed in the same way as that in the above example. Some results, for instance, are listed as follows.

Example 4.2 (Linear Ranking Selection [14]): The selected population Y is generated from N independent random experiments. Suppose $X = (x_1, \cdots, x_N)$ which is ranked so that $g(x_i) \geq g(x_j)$ whenever $i < j$. Then, in each experiment, x_r is selected from X with probability

$$P[y_i = x_r] = p + (N-r)q$$

where p, q are the adjustable parameters such that $((N-1)N/2)q + Np = 1$. The operator has the characteristics

$$P_S = N-1; \quad I_S = \left(\frac{1}{N}+\frac{N-1}{2}q\right)^N$$
$$U_S \geq \left(\frac{1}{N}+\frac{N-1}{2}q\right)^N.$$

Example 4.3 (Nonlinear Ranking Selection [14]): Different from the linear ranking selection, the nonlinear ranking selection selects x_r from X with probability $q(1-q)^{(r-1)}$. The operator then has the characteristics

$$P_S = N-1; \quad I_S = q^N; \quad U_S \geq q^N.$$

The subsequent examples will be other types of selection operators: they all generate the selected population Y in the way that the individual of Y is selected (with a certain positive probability) successively from the best one of several adjacent individuals in X.

Example 4.4 (Tournament Selection [15]): In this selection mechanism, for each $1 \leq i \leq N$, y_i is selected as the fittest individual among s individuals adjacent to x_i (where s is the so-called tournament size). The operator has the characteristics

$$P_S = s-1; \quad I_S = 1; \quad U_S = 1.$$

Proof: Without loss of generality, we assume that x_1 is the unique fittest individual in X. Then, by the tournament selection rule, $y_1 = x_1$ and $y_{N-s+2} = \cdots = y_N = x_1$ since, for any $N-s+i$ $(2 \leq i \leq s)$, the s individuals adjacent to x_{N-s+i} must contain x_1. Thus, $P(|X_M| = 1; |Y_M| \geq (s-1)+1) = 1$. That is, $P_S = s-1$ and $I_S = 1$ follow. It is obvious that $P(|X_M| < N; |Y_M| > |X_M|) = 1$. So we have $U_S = 1$. $\square$

Analogously, we have the following.

Example 4.5 (Elitist Selection [14]): Suppose that x_i is the best individual in X. The selection forces the fittest individual x_i into the selected population to replace the worst individual in X. So its characteristic parameters are

$$P_S = 1; \quad I_S = 1; \quad U_S = 1.$$

Example 4.6 (Boltzmann Selection [13]): In this rule, for each $1 \leq i \leq N$, y_i is selected from two individuals in X, say, a and b, which are adjacent to x_i. Suppose that $g(a)$ is more than $g(b)$. The probability of selecting a is then given by

$$P[y_i = a] = \frac{1}{1+e^{(g(b)-g(a))/T}}$$

where T is a temperature parameter. Let $\Delta = \min\{g(a) - g(b); g(a) > g(b), a, b \in \Omega\}$. This operator has the following characteristics:

$$P_S = 1; \quad I_S = \left(1+e^{-(\Delta/T)}\right)^{-2}; \quad U_S = \left(1+e^{-(\Delta/T)}\right)^{-2}.$$

The above listed characteristic parameter evaluation results of the selection operators are summarized in Table I.

B. Evolution Operators

We first briefly review the notions of mutation, crossover, and combination operators used in canonical GAs.

Let $X = \{x_1, x_2, \cdots, x_N\}$ be a population with each $x_i = (x_{i1}, x_{i2}, \cdots, x_{iL}) \in \{0, 1\}^L = \Omega$. The mutation M is an operator from Ω^N to Ω^N, which satisfies

$$M(X) = (M(x_1), \cdots, M(x_N)), \qquad \text{for any } x \in \Omega^N$$

where, for a given *mutation rate* p_m, $y_i = M(x_i)$ will take its jth component $y_{ij} = 1-x_{ij}$ with probability p_m, and $y_{ij} = x_{ij}$ with probability $1-p_m$.

The crossover operator C: $\Omega^N \to \Omega^N$ is defined as follows. For any $X \in \Omega^N$, $Y = C(X) = \{y_1, \cdots, y_N\}$ is created by N independent experiments. For each $1 \le i \le N$, two parent individuals, say, a and b, are first randomly sampled from X and then the probability of y_i being a random sample between a and b equals to $1-p_c$, where p_c is called *crossover rate*. With the probability p_c, y_i is created from a and b through a more complicated implementation such as the following.

- *One-Point Crossover:* If k is a random sample from $\{0, 1, 2, \cdots, N\}$. Then the jth component of y_i is taken as the jth bit of a when j is more than k; otherwise, it is taken as the jth component of b.
- *Two-Point Crossover:* If k, l are two random samples from $\{0, 1, 2, \cdots, N\}$, then the jth component of y_i is taken as the corresponding component of a when j is located between k and l; otherwise, it is taken as the corresponding component of b.
- *Uniform Crossover:* The jth component of y_i is a random sample from the corresponding components of a and b.

The combination operator (M, C) is the composition of the mutation M and the crossover C. It generates a population through two successive steps: first, an intermediate population is created by the crossover C and then, the mutation M is applied to the intermediate population, yielding the final population.

Example 4.7 (Mutation Operator): If M is a mutation operator with mutation rate $p_m < 1/2$, then its evolution aggregating rate, the scattering rate, and the stability rate are given, respectively, by

$$\begin{aligned} A_E &= 1 - \left(1 - p_m^L\right)^N \\ S_E^{(r)} &= (1-(1-p_m)^L)^r \left(1-p_m^L\right)^{N-r} \\ ST_E &\ge (1-p_m)^{LN}. \end{aligned}$$

(Note: by complicated calculation, we can also prove that the stability rate in this case is given by $(1-p_m)^{L(N-1)} + (N-1)p_m^L(1-p_m)^{L(N-1)}(1-(1-p_m)^L$.)

Proof: By the definition, $A_E = \inf\{P(X; M(X) \cap B \neq \phi); X \cap B = \phi, B \in \mathcal{T}\}$. This infimum will be attained at the most diffcult case when $B = \{\text{“}00 \cdots 0\text{”}\}$ and each individual in X is “$11 \cdots 1$.” In this case, A_E is equal to the probability of at least one individual “$11 \cdots 1$” becoming “$00 \cdots 0$.” Since the probability of one individual “$11 \cdots 1$” becoming “$00 \cdots 0$” is p_m^L, we find $A_E = 1 - (1-p_m^L)^N$.

To compute $S_E^{(r)} = \sup\{P(X; M(X) \cap B = \phi); |X \cap B| \ge r, B \in \mathcal{T}\}$, we let $|X \cap B| = r_1$ and notice that this supremum will occur in the easiest case: $B = \{\text{“}00 \cdots 0\text{”}\}$, and X contains r_1 individuals “$00 \cdots 0$” and $N - r_1$ individuals “$11 \cdots 1$.” In this case, $S_E^{(r)}$ is the probability of: 1) at least one bit of each individual “$00 \cdots 0$” is transformed into 1 and 2) no individual “$11 \cdots 1$” is converted to “$00 \cdots 0$.” Consequently, $S_E^{(r)} \le (1-(1-p_m)^L)^{r_1}(1-p_m^L)^{N-r_1}$. This function is monotonically decreasing in r_1 whenever $p_m \le 1/2$. So we obtain $S_E^{(r)} = (1-(1-p_m)^L)^r(1-p_m^L)^{N-r}$.

The stability rate of M is defined as $ST_E = \inf\{P(X; |M(X) \cap B| \ge |X \cap B|); X \in \Omega^N, B \in \mathcal{T}\}$. Despite the fact that it can be computed precisely, we only notice here that this value is clearly more than the probability that every individual in X remains unchanged under the mutation that is given by $(1-p_m)^{LN}$. Consequently, $ST_M \ge (1-p_m)^{LN}$. □

Example 4.8 (Combination Operator): The combination operator E composed of crossover and mutation has the characteristic parameters

$$\begin{aligned} A_E &= 1 - \left(1 - p_m^L\right)^N \\ S_E^{(N)} &= (1-(1-p_m)^L)^N \\ ST_E &\ge (1-p_m)^{LN}(1-p_c)^N. \end{aligned}$$

Proof: We obtain the aggregating rate and the stability rate by the same arguments as that used in the mutation case. The scattering rate $S_E^{(N)}$ will be attained in the easiest case when $B = \{\text{“}00 \cdots 0\text{”}\}$ and all N individuals of X are “$00 \cdots 0$.” Let a and b be two random samples from X (therefore, $a = b =$ “$00 \cdots 0$”). Thus, c, the intermediate individual created from a and b under crossover, is “$00 \cdots 0$” also. Consequently, the probability of the final individual not being “$00 \cdots 0$” is the probability of the individual “$00 \cdots 0$” not preserving itself under mutation, which is given by $1-(1-p_m)^L$. Therefore, $S_E^{(N)} = (1-(1-p_m)^L)^N$. □

Remark 4.1: The precise evaluation of the evolution scattering rate $S_E^{(r)}(0 < r < N)$ of the combination operators is also possible, but it is extremely complicated. The presentation of such a general scattering rate expression is, therefore, omitted here. Any reader interested in the expression can, however, calculate it by carefully distinguishing among different crossover manners (say, one-point, two-point, or uniform crossover).

C. Convergence of Typical Genetic Algorithms

The examples presented in the last two subsections now can be combined directly with Theorems 3.1 and 3.2 to yield convergence and convergence speed estimations of various canonical GAs.

Theorem 4.1: Assume that the GAs are defined by a series of the proportional selections $\{S(t)\}$ (with the gauge function σ_t) and the combination operators $\{M(t), C(t)\}$ (with the mutation rate $p_m^{(t)}$ and the crossover rate $p_c^{(t)}$). Corresponding to (11), define

$$\rho_t = \max\left\{\frac{\sigma_t(g(x))}{\sigma_t(g(y))};\ \sigma_t(g(x)) < \sigma_t(g(y)),\ x, y \in \{0,1\}^N\right\}.$$

If

1) $\rho_t = o(1/t)$
2) $p_m(t) = O(\sqrt[L]{1/t})$
3) $N > L$

then the genetic algorithms are quasi-convergent and for any $B \in \mathcal{T}$

$$P\left[X^{(t)} \cap B \neq \phi\right] \ge 1 - O\left(\left(\frac{1}{t}\right)^{N/(N+1)}\right).$$

Proof: According to Examples 4.1 and 4.8, the GAs in this case are special AEAs, with the selection pressure $P_t = N-1$ and the selection intensity $I_t = (1+(N-1)\rho_t)^{-N} =$

$1 - o(1/t)$ for each $t > 0$. We also know that $A_t = 1 - (1 - p_m^L)^N = 1 - (1 - O(1/t))^N = O(N/t)$ and $S_t^{(N)} = (1 - (1 - O(\sqrt[L]{1/t}))^L)^N = O(L^N t^{-(N/L)})$, which implies that $\sum_{t=1}^{\infty} A_t = \infty$ and

$$\begin{aligned}\frac{1}{A_t}\left(1 - I_t\left(1 - S_t^{(N)}\right)\right) &= \frac{1 - I_t}{A_t} + I_t\frac{S_t^{(N)}}{A_t} \\ &= o(1) + O\left(\frac{1}{N}L^N t^{1-(N/L)}\right) \to 0.\end{aligned}$$

According to Theorem 3.1, the algorithms are quasi-convergent. Moreover, let $K = t^{(N/N+1)}$; we then have $1 - I_K(1 - S_K^{(N)}) \leq 1 - I_K = o((1/t)^{(N/N+1)})$ and $\prod_{j=K}^{t}(1 - A_j) = \prod_{j=K}^{t}(j - 1/j)^N = (K/t)^N = (1/t)^{(N/N+1)}$. By Theorem 3.1, the convergence speed estimation of $X^{(t)}$ follows also. □

Theorem 4.1 can be applied directly to the simulated annealing-like genetic algorithms. For instance, the global annealing selection introduced in [32] corresponds to the proportional selection operator with time-variable guage function $\sigma_t(x) = \exp[-(x/T_t)]$, where T_t is the annealing temperature. Such a selection operator combined with the combination operator defines the GA which is called the global annealing genetic algorithm. In this case

$$\rho_t = e^{-(\Delta/T_t)}$$

with $\Delta = \min\{g(a) - g(b); g(a) > g(b), a, b \in \Omega\}$ and obviously, $\rho_t = o(t^{-1})$ amounts to $T_t = o(1/\ln(t))$. Thus, Theorem 4.1 implies the following corollary.

Corollary 4.1: If: a) $T_t = o(1/\ln(t))$, b) $p_m^{(t)} = O(\sqrt[L]{1/t})$, and c) $N > L$, then the global annealing genetic algorithm is quasi-convergent and, moreover, for any $B \in \mathcal{T}$

$$P\left[X^{(t)} \cap B \neq \phi\right] \geq 1 - O\left(\left(\frac{1}{t}\right)^{N/(N+1)}\right).$$

Analogously to Theorem 4.1, the following convergence of GAs can be verified.

Theorem 4.2: The AEA made of a series of the nonlinear selection and the combination operators is quasi-convergent if: 1) $q_t = 1 - o(1/t)$, where q_t is the parameter involved in nonlinear ranking selection operator at time t; 2) $p_m^{(t)} = O(\sqrt[L]{1/t})$; and 3) $N > L$.

Theorem 4.3: An AEA deduced by the tournament selections and the mutation is convergent if the mutation rate at time t satisfies $p_m^{(t)} = O(\sqrt[L]{1/t})$ and the tournament size $S \geq L$.

Proof: We can prove the quasi-convergence by using a similar argument as that in Theorem 4.1. Moreover, according to Example 4.7, the evolution stability rate of the AEA at time t is more than $(1 - p_m^{(t)})^{LN} = (1 - O(\sqrt[N]{1/t}))^{LN} \to 1 (t \to \infty)$ and the uniform selection rate of the AEA at time t is one. So, by Theorem 3.2, the algorithm is convergent also. □

Remark 4.2: As remarked in the introduction of Section I, the convergence of GAs has been studied in recent years. Nevertheless, most of the results are established for the "elitist-type" (that is, with the strategy of recording the elitist individual up to the current population and showing the convergence of the elitist individuals) and the time-independent genetic operators, and are based on ergodicity analysis. The convergence of GAs derived in this subsection are clearly all for the "nonelitist type" and the time-dependent genetic operators, and are based on nonergodicity analysis. Therefore, the results obtained provide not only new findings on convergent conditions, but also a new methodology of convergence analysis of GAs.

For GAs composed of proportional selection and combination operators, their "nonelitist" versions are known not to converge [18] when the operators are time independent. Theorem 4.1 reveals the convergent conditions of such GAs when the time-varying genetic operators are used. As a special case, Zhang *et al.* [32] showed that the global annealing GA is convergent if the "temperature" $T_t \to 0$ and the selection takes from the "parent–children" mixed populations. Corollary 4.1 in this section, however, shows that if the selection is natural [that is, only from the parent population), the algorithm will also be convergent provided the temperature declines as low as $O(1/\ln(t))$. This result coincides surprisingly with the standard requirement on the declining rate of temperature in the simulated annealing algorithms. For GAs with tournament selection, Miller and Goldberg [15] proved convergence of the algorithm, but under the assumption that each generation of population obeys the normal distribution. This assumption is obviously impractical. Theorem 4.3 in this section establishes the convergence of such GAs without supposing such an impractical condition.

In all of the convergence conditions listed in Theorems 4.1–4.3, the mutation rates are all asked to decrease in order of $O(\sqrt[L]{1/t})$. When the encoding length L is sufficiently large, $\sqrt[L]{1/t} \approx 1$. This suggests that, in this case, the constant mutation rate can actually be applied. This explains why the GAs with a constant mutation rate often are also convergent in practical applications.

It is worthwhile to note the condition "$N > L$" listed in Theorems 4.1–4.3. In our developed methodology of convergence analysis, this condition seems crucial and necessary. Note that it is a long-standing problem how the involved genetic parameters (like population size N, encoding length L, crossover rate p_c, mutation rate p_m, etc.) are reasonably set in a GA so as to yield optimal performance. Although the present research is by no means tackling this difficult problem, the condition "$N > L$" found in this paper can, however, shed some light on the final solution of this problem.

D. Evolution Strategies

We now consider the applications of the general theories developed in this paper to evolution strategies (ESs).

As a typical example, let us consider the (μ, λ)-ESs, which consist of a selection operator and an evolution operator. The initial population X contains λ individuals. To generate the next population, the best $\mu(\mu < \lambda)$ individuals are first chosen from these λ individuals, and each chosen individual will create λ/μ offspring. The resultant $\mu \cdot (\lambda/\mu) = \lambda$ offspring consist of the selected population $Y = S(X)$. The next population Z is then

generated from $S(X)$ by adding to each selected individual a random perturbation, say, as in [30] and [31]

$$Z_i = Y_i + \gamma_i N(0, \xi_i)$$

where $N(0, \xi_i)$ is the normal distribution or, as in [19]

$$Z_i = Y_i + p_t(L \circ u) + q_t(L_t \circ u) \tag{18}$$

where $L \circ u$ is a uniform distribution on $[-L, L]$, $L_t \circ u$ is the uniform distribution on $[-L_t, L_t]$ and p_t, q_t, L_t are adjustable parameters.

Example 4.9 ((μ, λ)-ES Selection): The (μ, λ)-ES has a selection pressure $P_S = \lambda/\mu - 1$, a selection intensity $I_S = 1$, and a uniform selection intensity $U_S = 1$.

Proof: Let $X = \{x_1, x_2, \cdots, x_\lambda\} \in \Omega^\lambda$ and $|X_M| = 1$ (without loss of generality, we assume that $x_1 \in X$ is the fittest). Then, by the definition, $x_1 \in S(X)$, and it gets λ/μ copies. All of these copies are clearly the fittest individuals in $S(X)$. This implies that

$$P(|X_M| = 1;\ |S(X)_M| \geq [(\lambda/\mu - 1) + 1]) = 1.$$

This means that $P_S = \lambda/\mu - 1$ and $I_S = 1$. It is noted that, if $|X_M| < \lambda$, $X_M \cap S(X)_M$ must contain at least $\min\{\lambda, |X_M|(\lambda/\mu)\}$ elements. So $|S(X)_M| > |X_M|$ since $\lambda/\mu > 1$. Thus, $U_S = 1$ follows. □

Evolution strategies are known to work on a continuous feasible region Ω (as usual, a bounded subset of R^n), so they are mainly used on optimization problems of continuous functions. To apply our theory to this case, we need to discretilize the continuous region Ω and, instead of yielding exact solutions, we consider convergence of the algorithm to any ϵ-approximate solutions. Thus, we first introduce the following definition.

Definition 4.1: A subset of the form $\sigma = \{(x_1, \cdots, x_n) \in R^n$: $k_i\epsilon < x_i \leq (k_i+1)\epsilon\}$ (where $\epsilon > 0$ is real number and k_i is an integer) is called a cell with mesh ϵ. A set of cells $\Sigma = \{\sigma^\alpha\}$ is said to be an ϵ -covering of Ω if

1) $\sigma^\alpha \cap \sigma^\beta = \phi$ if $\alpha \neq \beta$
2) $\Omega \subset \cup\{\sigma;\ \sigma \in \Sigma\}$
3) any $\sigma^\alpha \in \Sigma$ has a uniform mesh ϵ.

Given an ϵ-covering of Ω, the intersection set

$$\Omega_\epsilon = \{D \subset R^n\text{: there is } \sigma \in \Sigma \text{ such that } D = \sigma \cap \Omega\}$$

is called an ϵ -discretization of Ω.

Our purpose is to analyze how (μ, λ)-ES evolves from one cell in Ω_ϵ into another cell, and if it eventually stabilizes into the cell containing an optimal solution of the problem (1). To simplify the exposition, we assume that the fitness function has a unique global maximum. Under the assumption, the optimal set is a cell in Ω_ϵ; hence, without any loss of generality, we can only consider the satisfactory set B, which is identical to a cell in Ω_ϵ. However, the following propositions hold for any fitness function. We will give the full proof in another paper.

Example 4.10 ((μ, λ)-ES Evolution Operator): Let E: $\Omega \to \Omega$ be the componentwise defined (μ, λ)-ES evolution operator and define

$$p^l = \inf\{P(x;\ E(x) \in A);\ x \in B,\ A,\ B \in \Omega_\epsilon,\ A \neq B\} \tag{19}$$

$$p^u = \sup\{P(x;\ E(x) \in A);\ x \in B,\ A,\ B \in \Omega_\epsilon,\ A \neq B\}. \tag{20}$$

If p^l, $p^u \leq 1/2$, then the evolution aggregating rate and the scattering rate of E, respectively, satisfy the following estimations:

$$A_E \geq p^l(1 - p^u)^{\lambda - 1}$$
$$S_E^{(r)} \leq (p^u)^r(1 - p^l)^{\lambda - r}(|\Omega_\epsilon| - 1)^\lambda.$$

Proof: For a satisfactory set B, to evaluate $A_E = \inf\{P(X;\ E(X) \cap B \neq \phi);\ X \cap B = \phi,\ X \in \Omega_\epsilon^\lambda,\ B \in T(\Omega_\epsilon)\}$, we can let $X = \{X_1, \cdots, X_\lambda\}$ and $X_i \notin B\,(i = 1, 2, \cdots, \lambda)$. Then, in this case

$$\begin{aligned} P[E(X) \cap B \neq \phi] &\geq P[E(X_1) \in B]P[E(X_i) \notin B;\ i \geq 2] \\ &\geq p^l(1 - p^u)^{\lambda - 1}. \end{aligned}$$

This implies that $A_E \geq p^l(1 - p^u)^{\lambda - 1}$. If, in this case, X contains $s\,(s \geq r)$ individuals in B and other $\lambda - s$ individuals are outside B (that is, $|X \cap B| = s \geq r$), then each individual outside B remains unchanged with the probability less than $(1 - p^l)$. Consequently, we find

$$\begin{aligned} S_E^{(r)} &= \sup\{P(X;\ E(X) \cap B = \phi); \\ &\qquad |X \cap B| \geq r,\ X \in \Omega_\epsilon^\alpha,\ B \in \Omega_\epsilon\} \\ &\leq \max_{s \geq r}\{(|\Omega_\epsilon| - 1)^s(p^u)^s(1 - p^l)^{\lambda - s}\} \\ &\leq (|\Omega_\epsilon| - 1)^r(p^u)^r(1 - p^l)^{\lambda - r} \end{aligned}$$

where we have used the assumption $p^u, p^l \leq 1/2$ to deduce the last inequality. This justifies Example 4.10. □

Applying Theorem 3.1 and Examples 4.9 and 4.10, we can directly get the convergence of (μ, λ)-ESs. Assume that the (μ, λ)-ESs employ the time-dependent evolution operator E_t at time t. Correspondingly, the probability bounds p_t^l and p_t^u can be defined as in (19) and (20).

Theorem 4.4: Assume that the AEA is specialized by the (μ, λ)-ESs. Then the AEA is quasi-convergent to an ϵ-approximate solution of (1) if

a) $\sum_{t=1}^\infty p_t^l(1 - p_t^u)^{\lambda - 1} = \infty$
b) $\lim_{t \to \infty}((p_t^u)^{(\lambda/\mu)}/(p_t^l(1 - p_t^u)^{\lambda - 1})) = 0$.

The proof of Theorem 4.4 follows directly from Theorem 3.1 and, therefore, is omitted.

Theorem 4.4 can be applied to derive the convergence of various specific (μ, λ)-ESs. As an illustration, let us consider the (μ, λ)-ESs, where the time-dependent evolution operators E_t are defined as in (15).

Corollary 4.2: For the (μ, λ)-ES specified in (15), it will be quasi-convergent to an ϵ-approximation solution of (1) if

1) $L_t \to 0$; $q_t \to 0 (t \to \infty)$
2) $p_t = O(1/t)$
3) $(q_t/L_t) = O((1/t)^r)$ with some $r \in ((\mu/\lambda), 1)$.

Proof: We need to compute the probability bounds p_t^l and p_t^μ defined by (19) and (20) [with E_t specified as in (18)]. To this end, we take $\Omega_\epsilon = \{[-L, L] \cap (i\epsilon, (i + 1)\epsilon],\ i \in Z\}$ to be the ϵ discretization of $[-L, L]$, where L/ϵ is assumed to be an integer [i.e., $(L/\epsilon) \in Z$] and Z denotes the integer set.

For any two disjoint cells $(a_1, b_1]\,(a_2, b_2] \in \Omega_\epsilon$ from the definition (15), we calculate as shown in (21). Assumptions a)–c) imply $\lim_{t\to\infty}(q_t/L_t)/(p_t/(L-L_t)) = \infty$. We can assume $(q_t/L_t) > ((p_t/L-L_t))$ without loss of generality. From (18), this then yields

$$p_t^l = \inf\{P[E_t(x) \in A \,|\, x \in B];\ A \neq B,\ A, B \in \Omega_\epsilon\} = \begin{cases} p_t \dfrac{\epsilon}{2(L-L_t)}, & \text{if } (L-L_t) > \dfrac{\epsilon}{2} \\ p_t \dfrac{\epsilon}{2(L-L_t)} + q_t \dfrac{\epsilon}{2(L_t)}, & \text{otherwise} \end{cases} \tag{22}$$

and

$$p_t^u = \sup\{P[E_t(x) \in A \,|\, x \in B];\ A \neq B,\ A, B \in \Omega_\epsilon\} = \begin{cases} q_t \dfrac{\epsilon}{2L_t}, & \text{if } L - L_t \geq \epsilon \\ p_t \dfrac{\epsilon}{2(L-L_t)} + q_t \dfrac{\epsilon}{2(L_t)}, & \text{otherwise.} \end{cases} \tag{23}$$

Since ϵ is fixed and $L_t \to 0$ $(t \to 0)$, (19) and (20) imply that there is a $T > 0$ such that, as long as $t \geq T$

$$p_t^l = p_t \frac{\epsilon}{2(L-L_t)} \quad \text{and} \quad p_t^u = q_t \frac{\epsilon}{2L_t}.$$

By the assumptions, it then follows that

$$\sum_{t=1}^{\infty} p_t^l (1-p_t^u)^{\lambda-1} = \sum_{t=1}^{\infty} O\left(\frac{1}{t}\right) O(1) = \infty$$

and

$$\frac{(p_t^u)^{\lambda/\mu}}{p_t^l (1-p_t^u)^{\lambda-1}} = \left(\frac{\epsilon}{2}\right)^{\lambda/\mu-1} \left(\frac{q_t}{L_t}\right)^{\lambda/\mu} \left(\frac{L-L_t}{p_t(1-p_t^u)^{\lambda-1}}\right) = \left(\frac{\epsilon}{2}\right)^{\lambda/\mu-1} \frac{O\left(\frac{1}{t}\right)^{r(\lambda/\mu)} (L-L_t)}{O\left(\frac{1}{t}\right)} \to 0 \quad (t \to \infty)$$

(note that $r(\lambda/\mu) > 1$). Thus, conditions a)–b) of Theorem 4.4 are satisfied; consequently, the conclusion of Corollary 4.2 follows from Theorem 4.4. □

Remark 4.3: Rudolph [19], [20] has proved the convergence of the (μ, λ)-ESs for a class of convex fitness functions. Corollary 4.2 here provides a significant extension of these results in the sense that the convergence now has been proved for every fitness function. In addition, it should be observed that Theorem 4.4 can also be applied directly to study convergence of other specific (μ, λ)-ESs (say, those defined in [31] and [30]). This is omitted because of the limitation of length of the paper.

We have verified the quasi-convergence of the (μ, λ)-ESs via discretization approximation of Ω. This makes it direct and easy to follow the same argument as that in the analysis of GAs. However, this discretization procedure is by no means imperative for application of the developed theories to the (μ, λ)-ESs. The quasi-convergence of the (μ, λ)-ESs can actually be deduced by applying the theories without such discretization procedure but with some more sophisticated mathematical estimations.

V. Conclusion

We have proposed an axiomatic model of simulated evolutionary computation to unify various known evolutionary algorithms like GAs, ESs, and EP. With the new model, a novel convergence analysis and convergence rate estimation methodology is developed, which is not based on the usual ergodicity analysis, and could be regarded as a nonergodicity approach. The effectiveness and usefulness of the new model and the methodology are demonstrated through establishment of a generic convergence theory of the model and successful applications of the theory to various concrete GAs and ESs.

The main thread of the introduced axiomatic model is to cast the complicated evolution procedure from one generation to another generation as a composition of two independent operators: the selection, and the evolution operators. Each of these operators models an independent mechanism of evolution at a level of abstraction. The selection operator simulates natural selection and the evolution operator mimics the reproduction of the natural populations. The main benefit of such an abstraction is that with this, most of the currently known GAs, ESs, and EP can be unified and expressed as a stochastic process composed of the two operators, with each taking a specific form, which then makes it possible to develop a unified theory of the various evolutionary algorithms. Moreover, there is no complex interaction between the evolutionary operators, unlike that in many analyses on some specific evolutionary algorithms, which also makes it easier to conduct a precise and detailed analysis of the algorithms.

With the proposed model, an evolutionary algorithm differs from another only in the selection and evolution operators involved. This provides a mathematical basis for unifying the field of evolutionary computation. We have introduced the related characteristic parameters of the selection and evolution operators—selection pressure, selection intensity, evolution aggregating rate, evolution scattering rate, and evolution stability rate, which quantitatively measure their functions and properties. In terms of these introduced characteristic parameters, several general convergence (or quasi-convergence) conditions are formulated, and the corresponding convergence speeds are estimated.

$$P[E_t(x) \in (a_2, b_2] \,|\, x \in (a_1, b_1]] = \begin{cases} p_t \dfrac{\epsilon}{2(L-L_t)}, & \text{if } \min\{|a_2 - x|,\ |b_2 - x|\} \geq L_t \\ q_t \dfrac{\epsilon}{2(L_t)}, & \text{if } \max\{|a_2 - x|,\ |b_2 - x|\} \leq L_t \\ p_t \dfrac{\epsilon}{2(L-L_t)} + q_t \dfrac{\epsilon}{2(L_t)}, & \text{otherwise} \end{cases} \tag{21}$$

The current approach is based on a direct and precise analysis of the evolutionary operators, but not on the ergodicity analysis of the algorithms as is commonly adopted in the literature. Therefore, the major advantage of this approach is that the theories so deduced are not only of generality (particularly, they are valid for the "nonelitist" type and the "time-dependent" evolutionary algorithms), but also very convenient to specify for any concrete algorithms (which only needs to calculate the related characteristic parameters of the selection and evolution operators). A large set of specification examples in GAs and ESs presented in Section IV has justified this advantage.

There exist many opportunities of further research. For instance, we have the following.

- To refine the established theories for specific types of fitness functions. The convergence results we have developed are all based on the worst case analysis, and for all fitness functions. This makes the results very general, but on the other hand, these general results have not explored thoroughly and characterized completely the performance of SEAs for certain specific types of fitness functions (e.g., the nondeceptive functions, convex functions, etc.). It is of significance to further refine the application of Theorems 3.1 and 3.2 to such specific types of fitness functions.
- To abstract more general mechanism of selection and evolution operators and to integrate them into a more universal evolution computational model. Particularly, it should be observed that the introduced AEA model in the present paper requires essentially a certain kind of full connectivity (i.e., the evolution should, with a positive probability, carry each individual to every state of the feasible region). This is an implicit limitation of the present developed theory. How this implicit limitation can be resolved or relaxed deserves further investigation.
- To apply the developed AEA model in the prediction of the convergence of SEAs with noisy fitness evaluation. In most challanging and practical cases, SEAs are always implemented in a noisy environment (in particular, with noisy or inexact fitness evaluation). An elitist-type SEA in a noisy environment may report an incorrect maximum fitness if it happens that a fitness value that is extremely distorted by noise is assigned to some individuals. Similarly, the known theories based on an ergodicity analysis of SEA cannot predict the convergence of an SEA in a noisy environment correctly because they have not taken the effects of noise into consideration. The developed AEA model and convergence theories in this paper can, however, apply directly to analysis of the SEAs with such noisy environment.
- To devise more efficient new SEAs based on the theoretical inspiration of the developed AEA framework. The known SEA implementation schemes are almost all based on simulation of certain natural, biological evolution mechanism. While useful and fundamental, such a pure biological simulation approach has been subjected to the thorough understanding of natural evolution, of which many fundamental matters still need to be clarified. On the other hand, the proposed AEA model in this paper is an attempt at modeling SEAs in a pure mathematical axiomization approach, which provides not only a unifying mathematical model of various biological simulation-based SEA schemes, but also the possibility of devising SEA implementation schemes beyond pure biological simulation. It can be expected, in particular, that some more efficient new SEA schemes can be developed through constructing more promising selection and evolution operators (at least in a mathematical sense) based on the developed axiomization framework.

ACKNOWLEDGMENT

The authors would like to thank the editors and the anonymous reviewers for their insightful comments and helpful suggestions that have improved the paper greatly.

REFERENCES

[1] T. Bäck, U. Hammel, and H.-P. Schwefel, "Evolutionary computation: Comments on the history and current state," *IEEE Trans. Evol. Comput.*, vol. 1, pp. 3–17, 1997.
[2] R. Cerf, "Asymptotic convergence of a genetic algorithm," *Comptes Rendus Acad. Sci. Paris*, ser. I, vol. 319, pp. 271–276, 1996.
[3] ——, "The dynamics of mutation selection algorithms with large population sizes," *Ann. Inst. Henri Poincaré Prob. Statist.*, vol. 32, pp. 455–508, 1996.
[4] T. E. Davis and J. C. Principe, "A simulated annealing like convergence theory for the simple genetic algorithm," in *Proc. 4th Conf. Genetic Algorithms*, R. K. Belew and L. B. Booker, Eds., San Mateo, CA, 1991, pp. 174–181.
[5] H. Dawid, "A Markov chain analysis of genetic algorithms with a state dependent fitness function," *Complex Syst.*, vol. 8, pp. 407–417, 1994.
[6] K. A. De Jong, "Genetic algorithm: A 10-year perspective," in *Proc. 1st Int. Conf. Genetic Algorithm*, J. J. Grefenstette, Ed. Hillsdale, NJ: Lawrence, Erlbaum, 1985, pp. 169–177.
[7] D. B. Fogel, "An introduction to simulated evolutionary optimization," *IEEE Trans. Neural Networks*, vol. 5, pp. 3–14, 1994.
[8] ——, "Using fitness distributions to design more efficient evolutionary computations," in *Proc. 1996 IEEE Conf. Evolutionary Computation*, Nagoya, Japan, 1996, pp. 11–19.
[9] O. François, "Convergence in simulated evolution algorithms," *Complex Syst.*, vol. 10, pp. 311–319, 1996.
[10] J. H. Holland, *Adaptation in Natural and Artificial Systems*. Ann Arbor, MI: Univ. Michigan Press, 1975.
[11] J. Horn, "Finite Markov chain analysis of genetic algorithms with niching," in *Proc. 5th Int. Conf. Genetic Algorithms*, 1993, pp. 110–117.
[12] G. J. Koehler, "A proof of the Vose–Liepins conjecture," *Ann. Math. Artif. Intell.*, vol. 10, pp. 409–422, 1994.
[13] S. W. Mahfoud, "Finite Markov chain models of an alternative selection strategy for genetic algorithm," *Complex Syst.*, vol. 7, pp. 155–170, 1993.
[14] Z. Michalewicz, *Genetic Algorithm + Data Structures = Evolution Programs*. London, U.K.: Springer, 1992.
[15] B. L. Miller and D. E. Goldberg, "Genetic algorithm, tournament selection, and the effects of noise," *Complex Syst.*, vol. 9, pp. 193–212, 1995.
[16] A. E. Nix and M. D. Vose, "Modeling genetic algorithms with Markov chains," *Ann. Math. Artif. Intell.*, vol. 5, pp. 79–88, 1992.
[17] C. C. Peck and A. P. Dhawan, "Genetic algorithms as global random search methods: An alternative perspective," *Evol. Comput.*, vol. 3, no. 1, pp. 39–80, 1995.
[18] G. Rudolph, "Convergence analysis of canonical genetic algorithms," *IEEE Trans. Neural Network*, vol. 5, pp. 96–101, 1994.
[19] ——, "Convergence of non-elitist strategies," in *Proc. 1st IEEE Conf. Evolutionary Computation*. Piscataway, NJ: IEEE, 1994, pp. 63–66.
[20] ——, "Convergence rates of evolutionary algorithms for a class of convex object function," *Contr. Cybern.*, vol. 26, no. 3, pp. 375–390, 1997.
[21] R. Salomon, "Raising theoretical questions about the utility of genetic algorithms," in *Evolutionary Programming VI*, P. J. Angeling, R. Eberhart, R. G. Reynolds, and J. McDonnell, Eds. Berlin, Germany: Springer-Verlag, 1997, vol. 1213, pp. 275–184. (Lecture Notes, Comput. Sci.).

[22] W. M. Spears and K. A. De Jong, "An analysis of multi-point crossover," in *Foundations of Genetic Algorithms*, G. Rawlins, Ed. San Mateo, CA: Morgan Kaufmann, 1991, pp. 301–315.
[23] M. Srinivas and L. M. Patnaik, "On modeling genetic algorithms for functions of unitation," *IEEE Trans. Syst., Man., Cybern.*, vol. 26, pp. 809–821, 1996.
[24] G. Syswerda, "Uniform crossover in genetic algorithm," in *Proc. Int. Conf. Genetic Algorithms*, 1989, pp. 2–9.
[25] J. Suzuki, "A Markov chain analysis on simple genetic algorithms," *IEEE Trans. Syst., Man., Cybern.*, vol. 25, no. 4, pp. 655–659, 1995.
[26] S. Voget, "A central limit theorem for the population process of genetic algorithm," *Complex Syst.*, vol. 9, pp. 267–285, 1995.
[27] M. D. Vose and G. E. Liepins, "Punctuated equilibria in genetic search," *Complex Syst.*, vol. 5, pp. 31–44, 1991.
[28] D. Williams, *Probability with Martingales*, Cambridge: Cambridge Univ. Press, 1991.
[29] A. H. Wright and M. D. Vose, "Finiteness of the fixed point set for the simple genetic algorithm," *Evol. Comput.*, vol. 3, no. 3, pp. 299–309, 1996.
[30] X. Yao and Y. Liu, "Fast evolution strategies," *Evol. Comput.*, vol. 26, no. 3, pp. 467–496, 1997.
[31] X. Yao, Y. Liu, and G. Liu, "Evolutionary programming made faster," *IEEE Trans. Evol. Comput.*, vol. 3, pp. 82–102, 1999.
[32] J. S. Zhang, Z. B. Xu, and Y. Leung, "Global annealing genetic algorithm and its convergence analysis," *Sci. China (E)*, vol. 40, no. 4, pp. 414–424, 1997.

Kwong-Sak Leung (S'79–M'80–SM'89) received the B.Sc. (Eng.) and Ph.D. degrees from the University of London, Queen Mary College, U.K., in 1977 and 1980, respectively.

He was a Senior Engineer on contract Research and Development at ERA Technology and later joined the Central Electricity Generating Board to work on nuclear power station simulators in England. He has been with the Computer Science and Engineering Department, Chinese University of Hong Kong in 1985, where he is currently Professor and Chairman of the Department. He has authored or coauthored over 130 papers and two books in fuzzy logic and evolutionary computation. His research interests are soft computing, including evolutionary computation, probabilistic search, information fusion and data mining, fuzzy data, and knowledge engineering.

Dr. Leung has been Chair and Member of many program and organizing committees of international conferences and is on the Editorial Board of "Fuzzy Sets and Systems." He is a Chartered Engineer and a Member of the Institute of Electrical Engineers and the ACM.

Qi-Hong Duan was born on September 9, 1974, in Xi'an, China. He received the Master's degree in statistics from Xi'an Jiaotong University, Xi'an, China, in 1996. He is currently working toward the Ph.D. degree at the same university

His research interests include the theory and applications of random sets and evolutionary computation.

Mr. Duan was awarded the Shaanxi Outstanding Science Thesis Award in 1997.

Zong-Ben Xu received the M.S. degree in mathematics and the D.Sc. degree in applied mathematics from Xi'an Jiaotong University, Xi'an, China, in 1981 and 1987, respectively.

In 1988, he was Postdoctoral Fellow at the University of Strathclyde, U.K. From 1992 to 1998, he was a Research Fellow at the Department of Information Engineering, the Institute of Science and Technologies, the Department of Computer Science of the Chinese University of Hong Kong, and the Department of Computing, The Hong Kong Polytechnic Universtiy and City University of Hong Kong. He has been with the Faculty of Science, Xi'an Jiaotong University, Xi'an, China, since 1982, where he was an Associate Professor in 1987, a Full Professor in 1991, and is currently the Dean of Faculty of Science and the Director of the Institute for Information and System Science. He has authored or coauthored two monographs and more than 80 papers in international journals in the areas of nonlinear functional analysis, geometry of Banach spaces, optimization methods, genetic algorithms, and neural netowrks. His current research interests include mathematical foundation of artificial intelligence, neural networks, evolutionary computation, and nonlinear functional analysis.

Dr. Xu is a member of the New York Academy of Science, the International Mathematics Union, the American Mathematical Society, and the Chinese Mathematical Society. He holds a Ph.D. degree with outstanding achievements (awarded by Chinese State Education Commission and Academic Degree Commission of the Chinese Council 1991), and received the Distinguished Researcher Award several times. He serves in the editorial board of seven Chinese professional journals.

C. K. Wong (M'71–SM'78–F'85) received the B.A. degree in mathematics from the University of Hong Kong in 1965 and the M.A. and Ph.D. degrees in mathematics from Columbia University, New York, NY, in 1966 and 1970, respectively.

He joined the IBM T. J. Watson Research Center in 1969 as a Research Staff Member and was Manager of the VLSI Design Algorithms group from 1985 to 1995. He was Visiting Professor of Computer Science at University of Illinois, Urbana, from 1972 to 1973 and at Columbia University from 1978 to 1979. Since 1995, he has been a Professor of Computer Science and Engineering at The Chinese University of Hong Kong and was Chairman of the Department of Computer Science and Engineering from 1995 to 1997. He was an Editorial Board Member of *Fuzzy Sets and Systems* from 1978 to 1996, founding Editor-in-Chief of *Algorithmica*, an Advisory Board Member of *ACM Journal of Experimental Algorithmics*, and is on the editorial boards of *Networks* and *The Journal of Fuzzy Mathematics*. He holds four U.S. patents and has authored or coauthored more than 200 papers and two books: *Algorithmic Studies in Mass Storage Systems* (Rockville, MD: Computer Science, 1983) and *An Introduction to VLSI Pysical Design* (New York: McGraw-Hill, 1996). His current research is on VLSI design algorithms.

Dr. Wong is a Fellow of the Association for Computing Machinery (ACM). He received an Outstanding Invention Award, an Outstanding Technical Achievement Award, and four Invention Achievement Awards from IBM. He also received a Best Paper Award for his work on FPGA design at the 1995 IEEE International Conference on Computer Design: VLSI in Computers and Processors. He was Chair of the IEEE Computer Society Technical Committee on VLSI from 1990 to 1991 and Editor of IEEE TRANSACTIONS ON COMPUTERS from 1982 to 1985 and an Associate Editor of IEEE TRANSACTIONS ON VLSI SYSTEMS from 1993 to 1997.