Degree of Population Diversity—A Perspective on Premature Convergence in Genetic Algorithms and its Markov Chain Analysis

Yee Leung, Yong Gao, and Zong-Ben Xu

Abstract— In this paper, a concept of degree of population diversity is introduced to quantitatively characterize and theoretically analyze the problem of premature convergence in genetic algorithms (GA's) within the framework of Markov chain. Under the assumption that the mutation probability is zero, the search ability of the GA's is discussed. It is proved that the degree of population diversity converges to zero with probability one so that the search ability of a GA decreases and premature convergence occurs. Moreover, an explicit formula for the conditional probability of allele loss at a certain bit position is established to show the relationships between premature convergence and the GA parameters, such as population size, mutation probability, and some population statistics. The formula also partly answers the questions of to where a GA most likely converges. The theoretical results are all supported by the simulation experiments.

Index Terms— Genetic algorithms, Markov chains, population diversity, premature convergence, schema.

I. INTRODUCTION

GENETIC algorithms (GA's) are search and optimization algorithms based on the principles of natural evolution [1]. In applying GA's to solve large-scale and complex real-world problems, one of the most frequent difficulties encountered is premature convergence [2], [3]. Roughly speaking, premature convergence occurs when the population in a GA reaches such a suboptimal state that most of the genetic operators can no longer produce offspring that outperform their parents [3]. Several methods have been proposed to combat premature convergence in GA's [4]–[7]. These include, for example, the restriction of the selection procedure, the mating procedure (e.g., more disruptive crossover operators, higher mutation rate, and local mating), and the modification of fitness assignment (e.g., scaling, fitness sharing). However, all these methods are heuristic in nature. Their effects vary with different problems and their implementation strategies need *ad hoc* modifications with respect to different situations.

A critical problem in studying premature convergence is the identification when it has occurred and the characterization of its extent. Srinivas and Patnaik [8], for example, use

the difference between the average and maximum fitness values as a yardstick to measure premature convergence, though not a measure of genetic diversity, and vary adaptively the crossover and mutation probabilities according to the measurement. On the other hand, the term "population diversity" has been qualitatively used in many papers to study premature convergence [9]–[11]. It is widely recognized that the decrease of population diversity leads directly to premature convergence. However, so far there exists little effort in performing quantitative analysis of population diversity, let alone to use it as a tool to prevent premature convergence.

In this paper, we formally propose a concept of degree of population diversity and quantitatively characterize and theoretically analyze the problem of premature convergence in GA's using the theory of Markov chains. Under the assumption of zero mutation probability, the search ability of GA's (in particular, the function of the crossover operator) is discussed. It is proved that the degree of population diversity converges to zero with probability one so that the search ability of genetic algorithms decreases consistently and premature convergence necessarily occurs. The relationships between premature convergence and the GA parameters such as population size, mutation probability, and relevant population statistics are also studied in light of an explicit formula for the conditional probability of allele loss at a certain gene position. The proposed formula is also employed to partly answer the question of where a GA most likely converges to.

In Section II, we introduce the population Markov chain models for the canonical genetic algorithms (CGA's) and present the strict probabilistic definitions of the genetic operators. With the aid of ergodic analysis of the population Markov chain, the reason why the mutation probability may be assumed to be zero when analyzing premature convergence is demonstrated. In Section III, we propose a concept of degree of population diversity and quantitatively characterize and theoretically analyze the problem of premature convergence in GA's within the Markov chain framework. In Section IV, a series of simulation studies is employed to substantiate our theoretical investigation. We then summarize and propose some directions for further research in Section V. To facilitate discussion, we put all proofs in the Appendix.

II. CANONICAL GENETIC ALGORITHMS AND THEIR POPULATION MARKOV CHAIN

Without loss of generality, we consider the GA's with binary string representations of length l and fixed population size

N. We further assume that the algorithms use proportional selection, one-point crossover, and bit mutation. Each individual in the population corresponds to an element of the space $S = \{0, 1\}^l$, which is called the *individual space*. The *population space* is denoted as S^N and we call S^2 the *parents space*. For the sake of convenience, we write the population $\vec{X} \in S^N$ in both the vector and matrix forms as follows:

$$\vec{X} = (X_1, X_2, \cdots, X_N)^T = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1l} \\ x_{21} & x_{22} & \cdots & x_{2l} \\ \cdots & \cdots & \cdots & \cdots \\ x_{N1} & x_{N2} & \cdots & x_{Nl} \end{pmatrix}$$

where $X_i \in S$ is the ith individual of $\vec{X}$, while x_{ij} is the jth component of X_i. The fitness function $f: S \longrightarrow R^+$ can be derived from the objective function of the optimization problem by a certain decoding rule.

A CGA can, in essence, be given as follows.
CGA

Step 1) Set $k = 0$ and generate initial population $\vec{X}(0)$.
Step 2) Independently select N pairs of individuals from the current population for reproduction.
Step 3) Independently perform crossover to the N pairs of individuals to generate N new intermediate individuals.
Step 4) Independently mutate the N intermediate individuals to get the next generation

$$\vec{X}(k+1) = [X_1(k+1), \cdots, X_N(k+1)].$$

Step 5) Stop if some stopping criterion is met. Else, set $k = k + 1$ and go to Step 2).

From the mathematical point of view, the operators are random mappings between the spaces S^N, S^2, and S. They are the analogous idealized abstractions of some of the genetic mechanisms in the evolution of natural organizms. To facilitate our later analysis, we present in the following the strict probabilistic definitions of several basic operators.

a) The proportional selection operator, $T_s: S^N \longrightarrow S^2$, selects a couple of parents from the given population for reproduction. Given the population $\vec{X}$, the probability of selecting $(X_i, X_j) \in S^2$ as the parents is

$$P\{T_s(\vec{X}) = (X_i, X_j)\} = \frac{f(X_i)}{\displaystyle\sum_{X \in \vec{X}} f(X)} \cdot \frac{f(X_j)}{\displaystyle\sum_{X \in \vec{X}} f(X)}$$
$$1 \le i \le N, 1 \le j \le N. \quad (1)$$

b) The crossover operator, $T_c: S^2 \longrightarrow S$ generates an individual from the selected parents. Given the parent $X_i = (x_{i1}, \cdots, x_{il})$, $i = 1, 2$, the probability for the one-point crossover operator to generate an individual Y is

$$P\{T_c[(X_1, X_2)] = Y\} = \begin{cases} \dfrac{k \cdot p_c}{l}, & \text{if } Y \ne X_1 \\ (1 - p_c) + \dfrac{k \cdot p_c}{l}, & \text{if } Y = X_1 \end{cases}.$$
$$(2)$$

where $0 \le p_c \le 1$ is the so-called crossover probability, $k = k(X_1, X_2, Y)$ is the number of gene positions at which successful crossovers of X_1 and X_2 can generate Y.

Remark 2.1: Throughout the present paper, we only consider the two-parents-one-child crossover operator, i.e., crossing two parents will always generate only one child (which we assume to be the first offspring). Such a scheme has also been adopted by many other theoretical works on GA's [1]. This is because only with two-parents-one-child crossover, the offspring in the next generation are conditionally independent and identically distributed. This is crucial in the derivation of some results concerning the transition probabilities of the population Markov chains of GA's.

c) The mutation operator, $T_m: S \longrightarrow S$, operates on the individual by independently perturbing each bit string in a probabilistic manner and can be specified as follows:

$$P\{T_m(X) = Y\} = p_m^{|X-Y|}(1 - p_m)^{l-|X-Y|}. \quad (3)$$

Based on the genetic operators defined above, the CGA can be represented as the following iteration of populations:

$$\vec{X}(k+1) = \{T_m^i[T_c^i(T_s^i\{\vec{X}(k)\})], i = 1, \cdots, N\}$$
$$k \ge 0 \quad (4)$$

where (T_m^i, T_c^i, T_s^i), $i = 1, \cdots, N$, are independent versions of (T_m, T_c, T_s). It is easy to see that the sequence of populations $\{\vec{X}(k), k \ge 0\}$ is a time-homogeneous Markov chain with the state space S^N (henceforth, it is called the population Markov chain). Similar to Rudolph [12], it can be proved that if $p_m > 0$, we have for any $\vec{X}, \vec{Y}$,

$$\prod_{i=1}^N P\{T_m^i[T_c^i(T_s^i(\vec{X}))] = Y_i\} > 0. \quad (5)$$

That is, $P\{\vec{X}(k+1) = \vec{Y}/\vec{X}(k) = \vec{X}\} > 0$. Therefore, the population Markov chain $\{\vec{X}(k), k \ge 0\}$ is homogeneous, irreducible, and aperiodic [13]. Hence, it can reach any state infinite times with probability one regardless of the initial state. Theoretically, this means that the CGA's will never converge and premature convergence cannot occur provided that the mutation probability is larger than zero. Practically this means that even if premature convergence can occur, it will not persist indefinitely when the mutation probability is larger than zero.

III. DEGREE OF POPULATION DIVERSITY AND ITS MARKOV CHAIN ANALYSIS

In this section, we first propose and define a concept called "degree of population diversity" as a way to formalize the notion of population diversity, which has not been rigorously characterized in the literature. We then use the concept to study, in conjunction with the basic GA operators, the problem of premature convergence in CGA's within the Markov chain framework.

Definition 3.1: Let $\vec{X} = (X_1, \cdots, X_N) \in S^N$ be a population. The degree of population diversity of $\vec{X}$, denoted by $\lambda(\vec{X})$, is defined as the number of the components of the vector $\sum_{i=1}^N X_i$ whose values are not equal to 0 and N. Accordingly, $\beta(\vec{X}) = l - \lambda(\vec{X})$ is called the degree of matureness (or, the number of lost alleles) of the population $\vec{X}$.

If we denote $\vec{X}$ in its matrix form, then $\lambda(\vec{X})$ is just the number of columns of $\vec{X}$ whose entries take on both zero and

one as values. In particular, all the individuals of $\vec{X}$ must be identical whenever $\lambda(\vec{X}) = 0$. On the contrary, $\beta(\vec{X})$ is just the number of columns of $\vec{X}$ whose entries are either all zero or all one.

A schema $\mathbf{L}([1])$ is a hyperplane of the individual space S and can be represented as

$$\mathbf{L} = \{X = (x_1, \cdots, x_l) \in S; \, x_{i_k} = a_{i_k}$$
$$1 \leq i_k \leq l, 1 \leq k \leq K\}$$

where $K(1 \leq K \leq l)$ is called the order of $\mathbf{L}$, $\{i_1, \cdots, i_K\}$ are called the defining components (defining gene positions), and $\{a_{i_k}, 1 \leq k \leq K\}(a_{i_k} \in \{0, 1\})$ are the values of the defining components (defining alleles). To signify $\mathbf{L}$ by its defining components and their corresponding values, we may denote $\mathbf{L}$ as $\mathbf{L}(a_{i_1}, \cdots, a_{i_k})$. It is obvious that a schema of order K contains 2^{l-K} different individuals.

Definition 3.2: Let $\vec{X} = (X_1, \cdots, X_N) \in S^N$ be a population with the degree of population diversity $\lambda(\vec{X})$ and the degree of maturity $\beta(\vec{X}) = l - \lambda(\vec{X})$. Let $i_k, 1 \leq k \leq \beta(\vec{X})$ be the components at which all the individuals of $\vec{X}$ take the same values, say $a_{i_k} \in \{0, 1\}, 1 \leq k \leq \beta(\vec{X})$. We call the schema $\mathbf{L}[a_{i_1}, \cdots, a_{i_{\beta(\vec{X})}}]$ the minimum schema containing $\vec{X}$ and denote it by $\mathbf{L}[a_{i_1}, \cdots, a_{i_{\beta(\vec{X})}}; \vec{X}]$ or simply $\mathbf{L}(\vec{X})$ if there is no confusion.

For example, consider the population $\vec{X} = \{(0\,0\,0\,1), (0\,0\,1\,0), (0\,0\,1\,1), (1\,0\,1\,0)\}$. Since the individuals of $\vec{X}$ take both zero and one at the first, third, and fourth components, the degree of population diversity of $\vec{X}$ is then $\lambda(\vec{X}) = 3$. Since all the individuals of $\vec{X}$ take the same value zero at the second component, the minimum schema containing $\vec{X}$ is $\mathbf{L}(a_2; \vec{X}) = (*\,0\,*\,*)$ with $a_2 = 0$. By "minimum" we mean that if there is another schema $\mathbf{L}$ that also contains $\vec{X}$, it must be true that $\mathbf{L}(a_2; \vec{X}) \subseteq \mathbf{L}$. It is not difficult to see that there are $2^{l-K} = 2^{4-1} = 8$ different individuals in $\mathbf{L}(a_2; \vec{X}) = (*\,0\,*\,*)$.

To evaluate the effect of population diversity on CGA performance, we first assume that the mutation probability is zero. The following theorem characterizes the search ability of CGA's with the mutation probability $p_m = 0$.

Theorem 3.1: Let $\{\vec{X}(k), k \geq 0\}$ be the population Markov chain with $p_m = 0$ and let $\vec{X}(0) = \vec{X}_0$.

1) For each $Y \in \mathbf{L}[a_{i_1}, \cdots, a_{i_{\beta(\vec{X}_0)}}; \vec{X}_0]$, there exists an $n \geq 0$ such that

$$P\left\{Y \in \vec{X}(n)/\vec{X}(0) = \vec{X}_0\right\} > 0. \qquad (6)$$

2) For each $Y \notin \mathbf{L}[a_{i_1}, \cdots, a_{i_{\beta(\vec{X}_0)}}; \vec{X}_0]$ and every $n \geq 0$

$$P\left\{Y \in \vec{X}(n)/\vec{X}(0) = \vec{X}_0\right\} = 0. \qquad (7)$$

Remark 3.1: Theorem 3.1 shows that the search ability of CGA's with $p_m = 0$ is confined to the minimum schema containing the current population in which there are $2^{\lambda(\vec{X})}$ different individuals. So the larger the degree of diversity of the current population, the more the feasible solutions the CGA's can search. On the contrary, the smaller the degree of population diversity, the less the feasible solutions the CGA's

can search. In particular, when $\lambda(\vec{X}) = 0$, CGA's with $p_m = 0$ will have no ability to search for feasible solutions.

Theorem 3.1 tells us that if the global optimum lies in the minimum schema containing the initial population, it is possible for CGA's to find it. However, as demonstrated later, the selection and crossover operators have a serious effect on maturation—their employment may decrease the degree of population diversity and degrade the search ability of the CGA's. So, although the global optimal solution is in the minimum schema containing the initial population, it may be excluded outside the search range by the selection and crossover operators which are ironically searching for it. We first present a necessary lemma for the discussion.

Lemma 3.1: Let $\{\vec{X}(k), k \geq 0\}$ be the population Markov chain of a CGA with $p_m = 0$ and $\mathbf{B}$ be the set of homogenous populations, i.e., $\mathbf{B} = \{(X, X, \cdots, X); \, X \in S\}$. Then for each $n \geq 1$,

$$P\left\{\vec{X}(n) \in \mathbf{B}/\vec{X}(0) \in \mathbf{B}\right\} = 1. \qquad (8)$$

Theorem 3.2: Let $\{\vec{X}(k), k \geq 0\}$ be the population Markov chain of a CGA with $p_m = 0$ and $\mathbf{B}$ be the set of homogenous populations.

1) $\{\vec{X}(k), k \geq 0\}$ converges to $\mathbf{B}$ with probability one, i.e.,

$$P\left\{\lim_{k \to \infty} \vec{X}(k) \in \mathbf{B}\right\} = 1. \qquad (9)$$

2) The degree of diversity of the sequence of populations decreases monotonically with probability one, decreases strictly monotonically with positive probabilities, and converges to 0 with probability one. That is,

$$P\{\lambda[\vec{X}(k+1)] \leq \lambda[\vec{X}(k)]\} = 1, \qquad \forall k \geq 0, \qquad (10)$$
$$P\{\lambda[\vec{X}(k+1)] < \lambda[\vec{X}(k)]\} > 0, \qquad \forall k \geq 0 \text{ and} \qquad (11)$$
$$P\left\{\lim_{k \to \infty} \lambda[\vec{X}(k)] = 0\right\} = 1. \qquad (12)$$

Remark 3.2: Part a) of Theorem 3.2 has been proved earlier by Fogel [14].

Theorem 3.2 suggests that CGA's with $p_m = 0$ converge to homogoneous populations with probability one and the convergence is monotone in terms of the degree of population diversity. The difficulty is, however, that the limiting homogeneous populations may correspond to local optimal solutions, or nonextremal solutions, as well as the global optimal solutions (that is, all of the points in the individual space S). Hence, to show the effectiveness of CGA's (especially that of the selection and crossover operators), the question of where a CGA most likely converges to should be answered. We now proceed to establish some explicit formulas concerning the conditional probabilities of the population Markov chain. These formulas will partly answer the above question and, on the other hand, will give a formula of the probability of allele loss at a certain gene position. In the following, we no longer assume that the mutation probability is zero. Let us first introduce some notations in the following definition:

Definition 3.3: Given a population $\vec{X} = (X_1, \cdots, X_N)$, $X_i = (x_{i1}, \cdots, x_{il})$, $i = 1, \cdots, N$, for any positive integer $1 \leq m \leq l$, let I_0^m and I_1^m denote the sets of indexes of all the individuals of the population $\vec{X}$ that have, respectively, a zero or one at the gene position m, that is,

$$I_0^m = \{i \in \{1, 2, \cdots, N\}; x_{im} = 0\}$$
$$I_1^m = \{i \in \{1, 2, \cdots, N\}; x_{im} = 1\}$$

and write

$$f_0^m(\vec{X}) = \sum_{i \in I_0^m} f(X_i),$$
$$f_1^m(\vec{X}) = \sum_{i \in I_1^m} f(X_i).$$

We call

$$a_m = \frac{f_0^m(\vec{X})}{\sum_{j=1}^{N} f(X_j)}$$

$$b_m = 1 - a_m$$
$$= \frac{f_1^m(\vec{X})}{\sum_{j=1}^{N} f(X_j)}$$

respectively, the fitness ratio of zero and one allele individuals at the mth gene position.

Theorem 3.3: Let $\{\vec{X}(k), k \geq 0\}$ be the population Markov chain of a CGA, then for every $1 \leq m \leq l$, we have

$$P\left\{\vec{X}(1) \text{ loses allele 1 at gene } m/\vec{X}(0) = \vec{X}\right\}$$
$$= [a_m + (1 - 2a_m)p_m]^N \qquad (13)$$
$$P\left\{\vec{X}(1) \text{ loses allele 0 at gene } m/\vec{X}(0) = \vec{X}\right\}$$
$$= [b_m + (1 - 2b_m)p_m]^N. \qquad (14)$$

The following corollary shows the relationship between premature convergence and the population size, the mutation probability, and the population statistics a_m.

Corollary 3.1: For the CGA's, the probability for allele loss to occur at a gene position (hence premature convergence at the gene position) decreases with the population size N, and increases with $|a_m - \frac{1}{2}|$ and $|p_m - \frac{1}{2}|$. Particularly, for fixed $N > 1$, the above probability attains its minimum at $a_m = \frac{1}{2}$ and $p_m = \frac{1}{2}$.

Remark 3.3: Corollary 3.1 justifies the methods of scaling to prevent premature convergence. In fact, on the condition that the average fitness of the current population is not changed, scaling will make the difference between the individual fitness and the average fitness, and hence $|a_m - \frac{1}{2}|$, even smaller.

Remark 3.4: From Theorem 3.3 and Corollary 3.1, we can see that the probability of premature convergence at a gene position is independent of the crossover probability. So the method of adapting crossover probability to prevent premature convergence presented in [8] seems to bare no theoretical support. Adapting the crossover probability can merely speed

up the search of the minimum schema containing the current population.

From Theorem 3.1, we can also get the following corollary which partly answers the question of where a CGA most likely converges to.

Corollary 3.2: Let $\{\vec{X}(k), k \geq 0\}$ be the population Markov chain with $\vec{X}(0) = \vec{X}$. Denote by $\mathbf{L}(1)$ and $\mathbf{L}(0)$ the two competing schemat

$$\mathbf{L}(1) = \{X; x_m = 1\}$$
$$\mathbf{L}(0) = \{X; x_m = 0\}, \qquad 1 \leq m \leq l$$

and assume that $a_m > b_m$. Then we have

$$P\left\{\vec{X}(1) \subset \mathbf{L}(0)/\vec{X}(0) = \vec{X}\right\} > P\left\{\vec{X}(1) \subset \mathbf{L}(1)/\vec{X}(0) = \vec{X}\right\}$$
$$\text{if } 0 \leq p_m < \frac{1}{2} \qquad (15)$$

and

$$P\left\{\vec{X}(1) \subset \mathbf{L}(0)/\vec{X}(0) = \vec{X}\right\} < P\left\{\vec{X}(1) \subset \mathbf{L}(1)/\vec{X}(0) = \vec{X}\right\}$$
$$\text{if } 1/2 < p_m \leq 1. \qquad (16)$$

We close this section by a discussion of the relationship between the fitness ratio a_m and the degree of population diversity $\lambda(\vec{X})$. This relationship is indeed the very motivation for us to introduce the concept of degree of population diversity. From the proof of Theorem 3.3, we see that a_m is in fact the probability for the mth component of an individual generated from the current population to take on zero as its value. To put it in another way, let us assume that the current population is $\vec{X}(k)$ and $X(k+1)$ is an individual of the next generation. Then, a_m is nothing but the probability for the $0 - 1$ random variable $x_m(k+1)$ to take on zero as its value. The variance of $x_m(k+1)$ is then $a_m(1 - a_m)$. Define

$$\overline{\lambda} = \sum_{m=1}^{l} a_m(1 - a_m)$$

which is a scalar characterization of the scatterness of individuals generated from the population $\vec{X}$. Define a scalar function g as

$$g(x) = \begin{cases} 1, & \text{if } 0 < x < 1 \\ x, & \text{if } x = 0 \text{ or } x = 1 \end{cases}$$

then we have

$$\lambda(\vec{X}) = \sum_{m=1}^{l} g(a_m)g(1 - a_m).$$

This means that the degree of population diversity is a rough and yet easy to manipulate two-level discrete approximation of the degree of scatter of individuals generated from the population.

IV. EXPERIMENTAL RESULTS

To substantiate our theoretical results, a series of simulations was carried out to apply CGA's with parameters N (population size), p_m (mutation probability), and p_c (crossover probability) to a function optimization problem. The aim of the empirical analysis was three-fold: first, for a specific population size

and a specific mutation probability, we evaluated changes in the degree of population diversity over time (indicated by the number of function evaluations). By varying the mutation probability, we further obtained and compared the trends of changes in diversity. The same analysis was then repeated for different population sizes to examine if there were significant differences. Second, for a fixed population size, we evaluated the performance of the CGA (indicated by its ability to search for the global optimal solution) under different mutation probabilities. Third, for a fixed mutation probability, we examined the performance of the CGA with different population sizes.

Except for p_c whose value was kept constant throughout, both N and p_m took on specified values for comparisons. The parameters of the CGA experiments were set as follows:

Population size $N = 50, 100, 200$.

Crossover probability $p_c = 0.85$.

Mutation probability $p_m = 0.000, 0.001, 0.005, 0.01, 0.50$.

The optimization problem involved the search for the global optimum of the test function, the famous Sine envelope sine wave function ([15])

$$f(x_1, x_2) = 0.5 - \frac{\sin^2 \sqrt{x_1^2 + x_2^2} - 0.5}{[1.0 + 0.001(x_1^2 + x_2^2)]^2}.$$

This is a rapidly varying mutimodal function of two variables with $x_1 = 0$, $x_2 = 0$ as the unique global optimal solution having value $f(0, 0) = 1.0$. Each variable was encoded by 22 b so that the string length of an individual was $l = 44$. Therefore, the maximal degree of population diversity was 44. Each experiment was repeated in 30 independent runs and the average was taken as the basis for evaluation.

The simulation results of changes in diversity over time are summarized in Figs. 1–3, which depict the variations of the degree of population diversity with respect to the number of function evaluations for different population sizes. The simulation results are in support of our theoretical analysis in that the rate of decrease of the degree of population diversity is in inverse proportion to the population size and is proportion to $|p_m - \frac{1}{2}|$. The larger the population size, the longer the population is able to maintain its diversity. The closer p_m is to 1/2, the higher is the degree of diversity the CGA can maintain, i.e., the slower is the decrease in diversity. In particular, the degrees of population diversity of the CGA's with different population sizes all dramatically decrease to zero when $p_m = 0$ and the most rapid of them is the CGA with the smallest population size $N = 50$. With respect to population size and CGA performance, we observe in Table I and Figs. 4 and 5 that the larger the population size, the better is the performance of the algorithms. A moderate mutation probability, e.g., $p_m = 0.01$, also contributes to good performance while too large a mutation rate degrades the performance. It is also observed that the smaller the population size, the more notable is the effect of the mutation probability on the algorithm's performance (Fig. 5). In addition, though the performance of the CGA with 0.5 mutation probability is not very bad (Fig. 4), the convergence rate is obviously slower

TABLE I
THE PERFORMANCE OF GAs WITH DIFFERENT PARAMETERS. fopt. IS THE OPTIMAL VALUE FOUND BY THE ALGORITHMS. err. IS THE SQUARE ERROR OF OPTIMAL SOLUTION

p_m		0.000	0.001	0.005	0.010	0.500
N=	fopt.	0.9216	0.9480	0.9816	0.9891	0.9793
50	err.	100.32	61.28	19.70	11.16	14.50
N=	fopt.	0.9615	0.9672	0.9865	0.9892	0.9820
100	err.	42.33	36.45	13.80	11.40	15.35
N=	fopt.	0.9838	0.9865	0.9869	0.9910	0.9822
200	err.	16.70	14.12	11.49	8.87	12.50

than the case of a lower mutation probability, e.g., $p_m = 0.005$ and $p_m = 0.01$.

To summarize, the experimental results reveal that the impact of population size on the CGA's performance is more favorable than that of the mutation probability. For example, as shown in Table I, the performance of the CGA with $N = 200$ and $p_m = 0.000$ is better than that of the CGA with $N = 50$ and $p_m = 0.005$. All these theoretical and empirical results suggest that a more effective method to maintain the degree of population diversity to prevent premature convergence is to increase the population size instead of to adopt a high mutation rate.

V. CONCLUSION

We have introduced in this paper a concept of degree of population diversity, and quantitatively characterized and theoretically analyzed the problem of premature convergence in CGA's using the theory of Markov chain. Under the assumption that the mutation probability is zero, we have proved that the degree of population diversity converges to zero with probability one so that the search ability of a genetic algorithm decreases over time, resulting in premature convergence. An explicit formula for the conditional probability of allele loss at a certain gene position has been established to show relationships between premature convergence and the CGA parameters—population size N, mutation probability p_m, and population statistics a_m. The formula also partly answers the question of where a genetic algorithm most likely converges to. A series of simulations have also been conducted to validate our theoretical analysis. We conclude that to prevent premature convergence, increase in population size tends to play a more important role than the variation of the mutation probability. Any effort in varying the crossover probability, unlike what has been suggested by some research in the literature, has no contribution in avoiding premature convergence.

For further research, we at least can pursue the following two lines of investigation:

1) Use the degree of population diversity as a quantitative measure to prevent premature convergence by adaptively varying population size and mutation probability. Some optimization procedures can also be explored to improve performance.
2) The degree of population diversity presented in this paper is exclusively based on the genotype of the population. Resorting to Theorem 3.3 of the present paper, as has been outlined at the end of Section III, it is
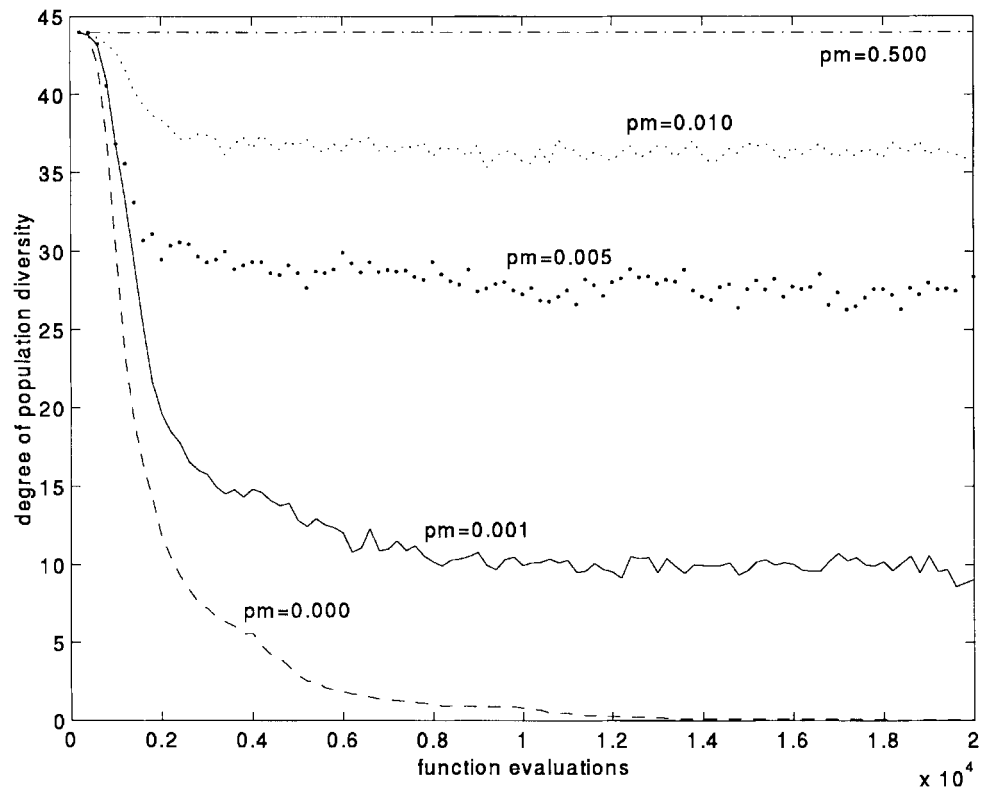
Fig. 1.   The convergence curves of population diversity as functions of the number of function evaluations with different mutation probabilities p_m and fixed population size $N = 50$.
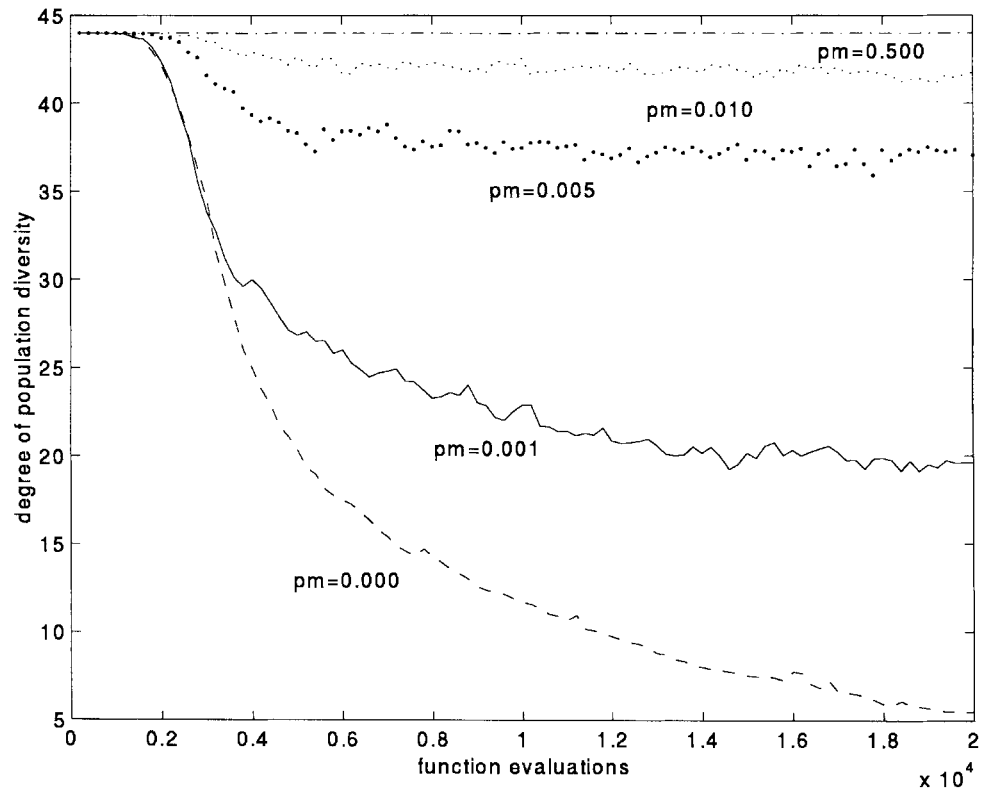

Fig. 2.   The convergence curves of population diversity as functions of the number of function evaluations with different mutation probabilities p_m and fixed population size $N = 100$.
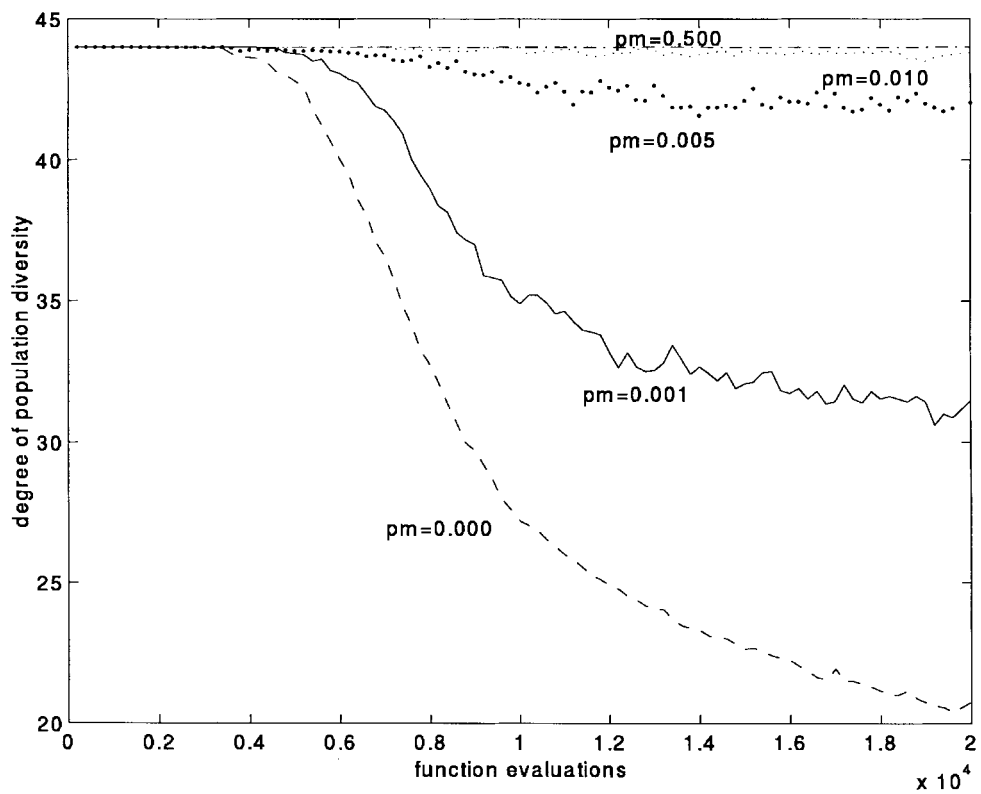
Fig. 3. The convergence curves of population diversity as functions of the number of function evaluations with different mutation probabilities p_m and fixed population size $N = 200$.
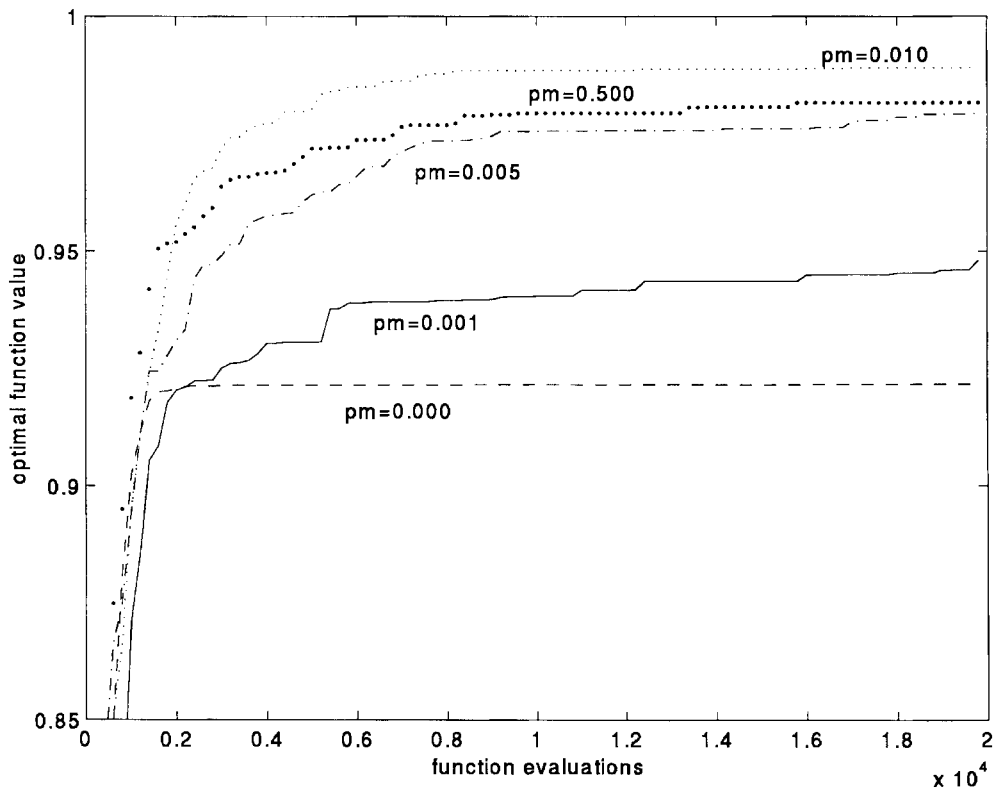

Fig. 4. The convergence curves of the optimal value found by GA's, as functions of the number of function evaluations, with different mutation probabilities p_m and fixed population size $N = 50$.
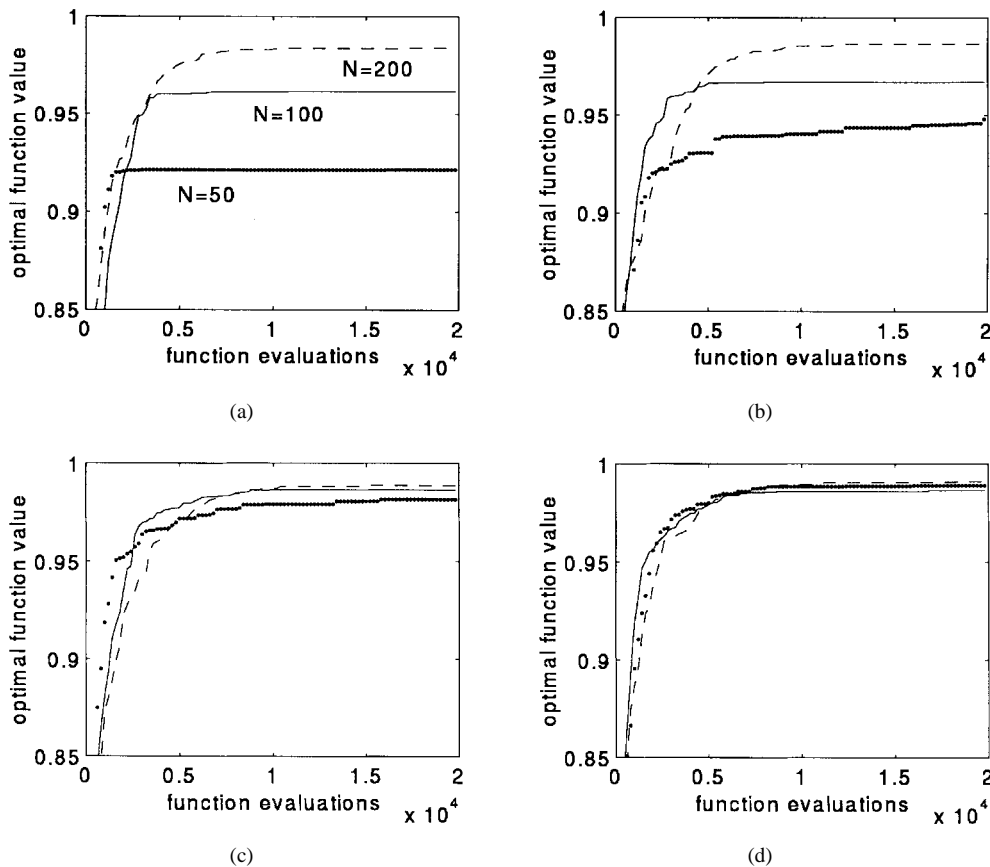
Fig. 5.   The convergence curves of the optimal value found by GA's, as functions of the number of function evaluations, with different mutation probabilities p_m and population size N. (a) $p_M = 0.000$. (b) $p_M = 0.001$. (c) $p_M = 0.005$. (d) $p_M = 0.010$.

possible to introduce a more involving concept of degree of population diversity based on both genotype and its corresponding fitness.

APPENDIX

1) Proof of Theorem 3.1:

a) Let $Y = (y_1, y_2, \cdots, y_l) \in \mathbf{L}[a_{i_1}, \cdots, a_{i_{\beta(\vec{X}_0)}}; \vec{X}_0]$ and

$$\vec{X}_0 = (X_1, X_2, \cdots, X_N)^T$$
$$= \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1l} \\ x_{21} & x_{22} & \cdots & x_{2l} \\ \cdots & \cdots & \cdots & \cdots \\ x_{N1} & x_{N2} & \cdots & x_{Nl} \end{pmatrix}.$$

Since the degree of population diversity of $\vec{X}_0$ is $\lambda(\vec{X}_0)$, we may suppose that the first $\beta(\vec{X}_0) = l - \lambda(\vec{X}_0)$ components of the individuals in $\vec{X}_0$ take the same bit value, i.e.,

$$x_{ij} = y_j, \qquad 1 \le i \le N, 1 \le j \le \beta(\vec{X}_0).$$

By the definition of degree of population diversity, for each $1 + \beta(X_0) \le j \le l$, there is an individual X_i in $\vec{X}_0$ such that its jth component x_{ij} equals to y_j.

We now proceed with the method of induction. Consider first the case of $\lambda(\vec{X}_0) = 2$. We have

$$x_{ij} = y_j, \qquad 1 \le i \le N, 1 \le j \le l - 2.$$

So there are $1 \le p, q \le N$, $p \ne q$, so that $x_{p(l-1)} = y_{l-1}$, and $x_{ql} = y_l$. Let $\vec{Z} = (Z_1, \cdots, Z_N)$ be a population such that $Z_i = X_i \, (i \ne p)$ and that Z_p is the individual generated by crossing the two individuals X_p, X_q at the lth component, i.e.,

$$Z_p = [x_{p1}, x_{p2}, \cdots, x_{p(l-1)}, x_{ql}]$$
$$= (y_1, \cdots, y_l)$$
$$= Y.$$

By the definition of the selection operator and crossover operator, we get

$$P\left\{Y \in \vec{X}(1)/\vec{X}(0) = \vec{X}(0)\right\}$$
$$\ge P\left\{\vec{X}(1) = \vec{Z}/\vec{X}(0) = \vec{X}(0)\right\}$$
$$\ge \prod_{i \ne p} \frac{f^2(X_i)}{\left[\sum_{j=1}^{N} f(X_j)\right]^2} \cdot \frac{p_c}{l} \frac{f(X_p) \cdot f(X_q)}{\left[\sum_{j=1}^{N} f(X_j)\right]^2}$$
$$> 0. \tag{17}$$

This proves the case of $\lambda(\vec{X}_0) = 2$.

Suppose that the theorem is valid for $\lambda(\vec{X}_0) = m - 1$. Consider the case of $\lambda(\vec{X}_0) = m$. Then, we have

$$x_{ij} = y_j, \qquad 1 \leq i \leq N, 1 \leq j \leq l - m$$

and there exists $1 \leq i_0 \leq N$ such that $x_{i_0 l} = y_l$. Let $\vec{Z}(1) = (Z_1, \cdots, Z_N)$ be a population such that $Z_{i_0} = X_{i_0}$ and

$$Z_i = [x_{i1}, x_{i2}, \cdots, x_{i(l-1)}, x_{i_0 l}], \qquad i \neq i_0.$$

It is easy to see that $\lambda[Z(1)] = m - 1$ and $Y \in \mathbf{L}[y_1, y_2, \cdots, y_{l-m}, y_l; \vec{Z}(1)]$. And hence

$$P\left\{\vec{X}(1) = \vec{Z}(1)/\vec{X}(0) = \vec{X}_0\right\}$$
$$\geq \prod_{i \neq i_0} \left\{ \frac{f(X_i)f(X_{i_0})}{\left[\sum_{j=1}^N f(X_j)\right]^2} \right\} \cdot \frac{p_c}{l} \frac{f^2(X_{i_0})}{\left[\sum_{j=1}^N f(X_j)\right]^2}$$
$$> 0. \tag{18}$$

By the assumption and homogeneity of the Markov chain $\{\vec{X}(k), k \geq 0\}$, there is a $n^* > 0$ such that

$$P\left\{Y \in \vec{X}(n^* + 1)/\vec{X}(1) = \vec{Z}(1)\right\} > 0.$$

Let $n = n^* + 1$, we get

$$P\left\{Y \in \vec{X}(n)/\vec{X}(0) = \vec{X}_0\right\}$$
$$\geq P\left\{Y \in \vec{X}(n^* + 1)/\vec{X}(1) = \vec{Z}(1)\right\}$$
$$\cdot P\left\{\vec{X}(1) = \vec{Z}(1)/\vec{X}(0) = \vec{X}_0\right\}$$
$$> 0. \tag{19}$$

Thus, the theorem is valid for the case $\lambda(\vec{X}_0) = m$. The proof of the first part of Theorem 3.1 is completed.

b) We write $\vec{X} \subset \mathbf{L}[a_{i_1}, \cdots, a_{i_{\beta(\vec{X}_0)}}; \vec{X}_0]$ to express that each individual of $\vec{X}$ belongs to the schema $\mathbf{L}[a_{i_1}, \cdots, a_{i_{\beta(\vec{X}_0)}}; \vec{X}_0]$, so that all the individuals of $\vec{X}$ have the same values at the components $i_1, \cdots, i_{\beta(\vec{X}_0)}$. Therefore, neither selection operator nor crossover operator can change the values at these components of the individuals in the population. So, for each $Y \in \mathbf{L}[a_{i_1}, \cdots, a_{i_{\beta(\vec{X}_0)}}; \vec{X}_0]$, we have

$$P\left\{\vec{X}(k+1) \subset \mathbf{L}[a_{i_1}, \cdots, a_{i_{\beta(\vec{X}_0)}}; \vec{X}_0]/\vec{X}(k) = Y\right\} = 1.$$

To simplify the notation, let us write $\mathbf{L}$ for

$$\mathbf{L}[a_{i_1}, \cdots, a_{i_{\beta(\vec{X}_0)}}; \vec{X}_0].$$

By the Markovian property, we get

$$P\left\{\vec{X}(n) \subset \mathbf{L}/\vec{X}(0) = \vec{X}_0\right\}$$
$$\geq \sum_{\vec{Y}(1) \subset \mathbf{L}} \cdots \sum_{\vec{Y}(n) \subset \mathbf{L}} P\left\{\vec{X}(1) = \vec{Y}(1)/\vec{X}(0) = \vec{X}_0\right\}$$
$$\cdot P\left\{\vec{X}(2) = \vec{Y}(2)/\vec{X}(1) = \vec{Y}(1)\right\} \cdots$$
$$P\left\{\vec{X}(n) = \vec{Y}(n)/\vec{X}(n-1) = \vec{Y}(n-1)\right\}$$
$$= 1.$$

Since $Y \in \vec{X}(n)$ implies $\vec{X}(n) \not\subset \mathbf{L}$, we have

$$P\left\{Y \in \vec{X}(n)/\vec{X}(0) = \vec{X}_0\right\} = 0.$$

This completes the proof.

2) Proof of Lemma 3.1: For any homogeneous population $\vec{X} \in \mathbf{B}$ and nonhomogeneous population $\vec{Y} \notin \mathbf{B}$, it follows from Theorem 3.1b) that $P\{\vec{X}(n) = \vec{Y}/\vec{X}(0) = \vec{X}\} = 0$. And hence

$$P\left\{\vec{X} \in \mathbf{B}/\vec{X}(0) \in \mathbf{B}\right\}$$
$$= 1 - P\left\{\vec{X}(n) \notin \mathbf{B}/\vec{X}(0) \in \mathbf{B}\right\}$$
$$= 1 - \sum_{\vec{Y} \notin \mathbf{B}} P\left\{\vec{X}(n) = \vec{Y}/\vec{X}(0) \in \mathbf{B}\right\}$$
$$= 1 - \sum_{\vec{Y} \notin \mathbf{B}} P\{\vec{X}(n) = \vec{Y}, \vec{X}(0) \in \mathbf{B}\}/P\{\vec{X}(0) \in \mathbf{B}\}$$
$$= 1 - \sum_{\vec{Y} \notin \mathbf{B}} \sum_{\vec{X} \in \mathbf{B}}$$
$$\cdot \frac{P\left\{\vec{X}(n) = \vec{Y}/\vec{X}(0) = \vec{X}\right\} \cdot P\{\vec{X}(0) = \vec{X}\}}{P\{\vec{X}(0) \in \mathbf{B}\}}$$
$$= 1.$$

3) Proof of Theorem 3.2:

a) Let $P\{\vec{X}, \vec{Y}\} = P\{\vec{X}(k+1) = \vec{Y}/\vec{X}(k) = \vec{X}\}$. For each $\vec{X} = (X_1, \cdots, X_N) \in S^N$, we have

$$P(\vec{X}, \mathbf{B}) := \sum_{\vec{Y} \in \mathbf{B}} P(\vec{X}, \vec{Y})$$
$$\geq \sum_{i=1}^N P\left\{\vec{X}(k+1) = (X_i, \cdots, X_i)/\vec{X}(k) = \vec{X}\right\}$$
$$\geq \sum_{i=1}^N \left\{ \frac{f^2(X_i)}{\left[\sum f(X_j)\right]^2} \right\}^N > 0.$$

Let $a = \inf_{\vec{Y} \in \mathbf{B}^c} P(\vec{Y}, \mathbf{B})$, then $0 < a < 1$. If $\vec{X}(0) \in \mathbf{B}$, it follows from Lemma 3.1 that $\vec{X}(n) \in \mathbf{B}$ for any $k \geq 1$. So the part a) of Theorem 3.2 is valid.

Now assume that $\vec{X}(0) \notin \mathbf{B}$ and define

$$T = \inf\{k \geq 1; \vec{X}(k) \in \mathbf{B}\}.$$

For any $k \geq 1$, we get

$$P\{T = k\} = \sum_{\vec{Y}_1, \cdots, \vec{Y}_{k-1} \notin \mathbf{B}} P(\vec{X}, \vec{Y}_1) P(\vec{Y}_1, \vec{Y}_2) \cdots$$
$$P(\vec{Y}_{k-2}, \vec{Y}_{k-1}) P(\vec{Y}_{k-1}, \mathbf{B})$$
$$\leq (1-a)^{k-1}.$$

It follows that $E(T) = \sum_{k=1}^{\infty} k P\{T = k\} < +\infty$, where $E(T)$ is the expectation of the random variable T. And hence $P\{T < +\infty\} = 1$. From Lemma 3.1 and the definition of T we know that $\lim_k \vec{X}(k) \in \mathbf{B}$ whenever $T < +\infty$. Therefore, we get $P\{\lim_k \vec{X}(k) \in \mathbf{B}\} = 1$ as what we intend to prove.

b) Denote by $\mathbf{L}(\vec{X})$ the minimal schema containing $\vec{X}$. By virtue of the formula of total probability, we have

$$P\{\lambda[\vec{X}(k+1)] \leq \lambda[\vec{X}(k)]\}$$
$$= \sum_{\vec{X} \in S^N} P\Big\{ \vec{X}(k+1) \in \mathbf{L}(\vec{X}) \times \mathbf{L}(\vec{X})$$
$$\times \cdots \times \mathbf{L}(\vec{X}) / \vec{X}(k) = \vec{X} \Big\} P\{\vec{X}(k) = \vec{X}\}$$
$$= 1$$

which proves (10). Since for each $\vec{X} \in S^N$,

$$P\Big\{ \lambda[\vec{X}(k+1)] < \lambda[\vec{X}(k)] / \vec{X}(k) = \vec{X} \Big\}$$
$$= P\Big\{ \lambda[\vec{X}(k+1) \leq \lambda[\vec{X}(k)] - 1 / \vec{X}(k) = \vec{X} \Big\}$$
$$\geq P\Big\{ \lambda \vec{X}(k+1) = 0 / \vec{X}(k) = \vec{X} \Big\}$$
$$\geq \left\{ \frac{f^2(X_1)}{\left[\sum_{i=1}^{N} f(X_i) \right]^2} \right\}^N > 0$$

then an application of the formula of total probability yields the desired result (11).

To get (12), it should be observed that $\lambda(\vec{Y}) = 0$ for each $\vec{Y} \in \mathbf{B}$. It follows that $\lim_k \vec{X}(k) \in \mathbf{B}$ implies $\lim_k \lambda[\vec{X}(k)] = 0$, and hence, $P\{\lim_k \lambda[\vec{X}(k)] = 0\} = 1$. The proof is completed.

4) Proof of Theorem 3.3: We only prove (13) and the proof of (14) is similar. Let $x_{ij}(1)$ be the jth component of the ith individual of $\vec{X}(1)$ and $Y = (y_1, \cdots, y_l)$ the random individual generated by the selection and crossover operator from $\vec{X}(0)$. Then, for each i

$$P\Big\{ x_{im}(1) = 0 / \vec{X}(0) = \vec{X} \Big\}$$
$$= P\Big\{ y_m = 0 / \vec{X}(0) = \vec{X} \Big\} \cdot (1 - p_m)$$
$$+ P\Big\{ y_m = 1 / \vec{X}(0) = \vec{X} \Big\} \cdot p_m$$
$$= P\Big\{ y_m = 0 / \vec{X}(0) = \vec{X} \Big\} \cdot (1 - p_m)$$
$$+ \Big(1 - P\Big\{ y_m = 0 / \vec{X}(0) = \vec{X} \Big\} \Big) \cdot p_m. \quad (20)$$

By the definition of the selection and crossover operators, we get

$$P\Big\{ y_m = 0 / \vec{X}(0) \Big\}$$
$$= \sum_{i,j \in I_0^m} \frac{f(X_i) f(X_j)}{\left[\sum_{n=1}^{N} f(X_n) \right]^2}$$
$$+ \sum_{i \in I_0^m} \sum_{j \in I_1^m} \frac{f(X_i) f(X_j)}{\left[\sum_{n=1}^{N} f(X_n) \right]^2} \frac{l-m}{l}$$
$$+ \sum_{i \in I_0^m} \sum_{j \in I_1^m} \frac{f(X_i) f(X_j)}{\left[\sum_{n=1}^{N} f(X_n) \right]^2} \frac{m(1-p_c)}{l}$$
$$+ \sum_{i \in I_1^m} \sum_{j \in I_0^m} \frac{f(X_i) f(X_j)}{\left[\sum_{n=1}^{N} f(X_n) \right]^2} \frac{m p_c}{l}$$
$$= \frac{\left[\sum_{i \in I_0^m} f(X_i) \right]^2}{\left[\sum_{n=1}^{N} f(X_n) \right]^2}$$
$$+ \sum_{i \in I_0^m} \sum_{j \in I_1^m} \frac{f(X_i) f(X_j)}{\left[\sum_{n=1}^{N} f(X_n) \right]^2}$$
$$= a_m^2 + a_m(1 - a_m)$$
$$= a_m.$$

The four terms in the right hand side of the second equation correspond to the four possible events of selection and crossover all of which result in an individual with the mth component being zero. The first one is the case that the mth components of the selected parents are both zero and crossover is arbitrary and immaterial. The second and third terms correspond to the situation that the first parent is selected with the mth component zero and the second parent has one at the mth component. The difference is that the second term corresponds to the case that the crossover point is selected after the mth component and crossover either occurs or does not occur (i.e., having crossover or not is immaterial), while the third one is the situation where the crossover point is before the mth component and crossover does not take place. The fourth term captures the case that the first parent has one at its mth component and the second parent has zero at the same position, and crossover occurs with the crossover point being selected before the mth component.

Therefore, for each i

$$P\Big\{ x_{im}(1) = 0 / \vec{X}(0) = \vec{X} \Big\} = a_m(1 - p_m) + (1 - a_m) p_m$$
$$= a_m + (1 - 2a_m) p_m$$

Since the N individuals of $\vec{X}(1)$ are independently generated with the same manner, we get

$$
\begin{aligned}
&P\left\{\vec{X}(1)\,\text{loses allele 1 at gene m}/\vec{X}(0)=\vec{X}\right\} \\
&= P\left\{x_{im}(1)=0,\, 1\le i\le N/\vec{X}(0)=\vec{X}\right\} \\
&= \prod_{i=1}^{N} P\left\{x_{im}(1)=0/\vec{X}(0)=\vec{X}\right\} \\
&= [a_m+(1-2a_m)p_m]^N.
\end{aligned}
$$

This completes the proof.

5) Proof of Corollary 3.1: From Theorem 3.3, we know that

$$
\begin{aligned}
&P\left\{\vec{X}(1)\,\text{loses an allele at gene m}/\vec{X}(0)=\vec{X}\right\} \\
&= [a_m+(1-2a_m)p_m]^N + [b_m+(1-2b_m)p_m]^N \\
&= [a_m+(1-2a_m)p_m]^N + [1-a_m-(1-2a_m)p_m]^N \\
&:= f(a_m, p_m, N). \tag{21}
\end{aligned}
$$

Since $0 \le a_m+(1-2a_m)p_m \le 1$ for each a_m and p_m, the function $f(a_m, p_m, N)$ decreases with N.

We write $a = a_m+(1-2a_m)p_m$ to simplify the notation. The partial differential of $f(a_m, p_m, N)$ with respect to a_m can be given as

$$
\begin{aligned}
\frac{\partial f}{\partial a_m} &= Na^{N-1}(1-2p_m) - N(1-a)^{N-1}(1-2p_m) \\
&= N(1-2p_m)[a^{N-1}-(1-a)^{N-1}].
\end{aligned}
$$

It can be shown that, for each $0 \le a_m \le 1$, $\partial f/\partial a_m \le 0$ whenever $0 \le a_m \le \frac{1}{2}$ and $\partial f/\partial a_m \ge 0$ whenever $\frac{1}{2} \le a_m \le 1$. That is to say that $f(a_m, p_m, N)$ is decreasing when $a_m \le \frac{1}{2}$ and increasing when $a_m \ge \frac{1}{2}$. It follows that the probability for an allele loss to occur at a gene position increases with $|a_m - \frac{1}{2}|$. Similarly, we can prove that the probability for allele loss to occur at a gene position increases with $|p_m - \frac{1}{2}|$.

Solving the equations $\partial f/\partial a_m = 0$ and $\partial f/\partial p_m = 0$, we get an unique solution $p_m = \frac{1}{2}$, $a_m = \frac{1}{2}$. Hence, for fixed N, $f(a_m, p_m, N)$ takes its minimum at $p_m = \frac{1}{2}$ and $a_m = \frac{1}{2}$. The proof is completed.

6) Proof of Corollary 3.2: It is obvious that $\vec{X}(1) \subset \mathbf{L}(0)$ [or $\vec{X}(1) \subset \mathbf{L}(1)$] if and only if $\vec{X}(1)$ lose the allele 1 (respectively 0) at gene m. So

$$
\begin{aligned}
P\left\{\vec{X}(1) \subset \mathbf{L}(0)/\vec{X}(0)=\vec{X}\right\} &= [a_m+(1-2a_m)p_m]^N \\
P\left\{\vec{X}(1) \subset \mathbf{L}(1)/\vec{X}(0)=\vec{X}\right\} &= [b_m+(1-2b_m)p_m]^N.
\end{aligned}
$$

If $a_m > b_m$, simple calculation shows that

$$
\begin{aligned}
&a_m+(1-2a_m)p_m > b_m+(1-2b_m)p_m \\
&\quad \text{if } 0 \le p_m < \tfrac{1}{2} \\
&a_m+(1-2a_m)p_m < b_m+(1-2b_m)p_m \\
&\quad \text{if } \tfrac{1}{2} < p_m \le 1.
\end{aligned}
$$

The corollary then follows.

REFERENCES

[1] J. H. Holland, *Adaptation in Natural and Artificial Systems.* Ann Arbor, MI: Univ. Michigan Press, 1975.

[2] L. Davis, Ed., *Handbook of Genetic Algorithms.* New York: Van Nostrand Reinhold, 1991.

[3] D. B. Fogel, "An introduction to simulated evolutionary optimization," *IEEE Trans. Neural Network,* vol. 5, pp. 3–14, 1994.

[4] H. G. Cobb and J. J. Grefenstette, "Genetic algorithms for tracking changing environment," in *Proc. 5th Int. Conf. Genetic Algorithms,* S. Forrest, Ed. San Mateo, CA: Morgan Kaufmann, 1993, pp. 523–530.

[5] K. A. De Jong, *An Analysis of the Behavior of a Class of Genetic Adaptive Systems,* Doctoral dissertation, Dept. Comput. Commun. Sci., Univ. Michigan, Ann Arbor, 1975.

[6] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning.* New York: Addison-Wesley, 1989.

[7] J. J. Grefenstette, "Genetic algorithms for changing environments," in *Parallel Problem Solving from Nature 2,* R. Manner and B. Manderick, Eds. Amsterdam: North Holland, 1992, pp. 137–144.

[8] M. Srinivas and L. M. Patnaik, "Adaptive probabilities of crossover and mutation in genetic algorithms," *IEEE Trans. Syst., Man, Cybern.,* vol. 24, pp. 656–667, 1994.

[9] D. C. Mattfeld, H. Kopfer, and C. Bierwirth, "Control of parallel population dynamics by social-like behavior of GA-individuals," in *Parallel Problem Solving From Nature-PPSN III,* Y. Davidor, H.-P. Schwefel, and R. Manner, Eds. Berlin: Spinger-Verlag, 1994, pp. 16–25.

[10] R. E. Smith, S. Forrest, and A. S. Perelson, "Population diversity in an immune system model: Implications for genetic search," in *Foundations of Genetic Algorithms 2,* L. D. Whitley, Ed. San Mateo, CA: Morgan Kaufmann, 1993, pp. 153–166.

[11] Y. Yoshida and N. Adachi, "A diploid genetic algorithm for preserving population diversity-pseudo-meiosis GA," in *Parallel Problem Solving From Nature-PPSN III,* Y. Davidor, H.-P. Schwefel, and R. Manner, Eds. Berlin: Spinger-Verlag, 1994, pp. 36–45.

[12] G. Rudolph, "Convergence analysis of canonical genetic algorithms," *IEEE Trans. Neural Networks,* vol. 5, pp. 96–101, 1994.

[13] M. Iosifescu, *Finite Markov Processes and Their Applications.* New York: Wiley, 1980.

[14] D. B. Fogel, "Asymptotic convergence properties of genetic algorithms and evolutionary programming: Analysis and experiments," *Cybern. Syst.,* vol. 25, no. 3, pp. 389–407, 1994.

[15] J. D. Schaffer, R. A. Caruana, L. J. Eshelman, and R. Das, "A study of control parameters affecting online performance of genetic algorithms for function optimization," in *Proc. 3rd Int. Conf. Genetic Algorithms,* J. D. Schaffer, Ed. San Mateo, CA: Morgan Kaufmann, 1989, pp. 51–60.

[16] A. E. Eiben, E. H. L. Aarts, and K. M. van Hee, "Global convergence of genetic algorithms: A Markov chain analysis," in *Parallel Problem Solving from Nature,* H. P. Schwefel and R. Manner, Eds. Berlin: Springer-Verlag, 1991, pp. 4–12.

Yee Leung received the B.S.Sc. degree in geography from the Chinese University of Hong Kong, in 1972, the M.A. degree in 1974 and the M.S. degree in engineering, and the Ph.D. degree in geography, both in 1977, from The University of Colorado.

He is currently Professor of geography and Director of The Centre for Environmental Studies at The Chinese University of Hong Kong. He has published three monographs and more than 70 articles in international journals and book chapters. His areas of specialization covers the development and application of intelligent spatial decision support systems, spatial optimization, fuzzy sets and logic, neural networks, and evolutionary computation.

Dr. Leung serves in the editorial boards of several international journals and is council member of several Chinese professional organizations.

Yong Gao was born on August 1, 1966, in China. He received the B.S. and M.S. degrees in applied mathematics from Xi'an Jiaotong University, Xi'an, China, in 1987 and 1990, respectively.

Since 1994, he has been an Associate Professor with the Institute for Information and System Sciences and the Center for Applied Mathematics Research, Faculty of Science, Xi'an Jiaotong University. From 1995 to 1996, he was a Research Associate at the Chinese University of Hong Kong. His current research interests include genetic algorithms, probability methods for machine learning and artificial neural networks, fractal image compression, and the theory of random sets.

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

In 1988, he was Postdoctoral Fellow at the University of Strathclyde, U.K. From 1992 to 1994, February to August 1995, and October 1996 to April 1997, he worked as Research Fellow at the Department of Information Engineering, the Institute of Sciences and Technologies of The Chinese University of Hong Kong, and the Department of Computing, The Hong Kong Polytechnic University, Hong Kong. He has been with the Faculty of Science, Xi'an Jiaotong University, since 1982, where he was promoted first as Associate Professor in 1987 and then as Full Professor in 1991. He is currently a Director of the Institute for Information and System Science, Xi'an Jiaotong University. He has published two monographs and more than 80 papers in international journals on the areas of nonlinear functional analysis, geometry of Banach spaces, optimization methods, genetic algorithms, and neural networks. His current research interests include mathematical foundation of AI, neural networks, evolutionary computation, and nonlinear functional analysis.

Dr. Xu is a member of the New York Academy of Science, the International Mathematicians 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 in 1991), and he received the Distinguished Researcher Award several times. He serves in the editorial board of several Chinese professional journals.