

Research Article

Convergence Analysis of a Class of Computational Intelligence Approaches

Junfeng Chen,^{1,2,3} Jianjun Ni,¹ and Mingang Hua¹

¹ College of IOT Engineering, Hohai University, Changzhou 213022, China

² Jiangsu Key Laboratory of Power Transmission and Distribution Equipment Technology, Changzhou 213022, China

³ Changzhou Key Laboratory of Sensor Networks and Environmental Sensing, Changzhou 213022, China

Correspondence should be addressed to Junfeng Chen; chen-1997@163.com

Received 11 January 2013; Revised 25 March 2013; Accepted 13 April 2013

Academic Editor: Yang Tang

Copyright © 2013 Junfeng Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Computational intelligence approaches is a relatively new interdisciplinary field of research with many promising application areas. Although the computational intelligence approaches have gained huge popularity, it is difficult to analyze the convergence. In this paper, a computational model is built up for a class of computational intelligence approaches represented by the canonical forms of generic algorithms, ant colony optimization, and particle swarm optimization in order to describe the common features of these algorithms. And then, two quantification indices, that is, the variation rate and the progress rate, are defined, respectively, to indicate the variety and the optimality of the solution sets generated in the search process of the model. Moreover, we give four types of probabilistic convergence for the solution set updating sequences, and their relations are discussed. Finally, the sufficient conditions are derived for the almost sure weak convergence and the almost sure strong convergence of the model by introducing the martingale theory into the Markov chain analysis.

1. Introduction

Computational intelligence approaches (CIAs) are a set of nature-inspired computational methodologies to interpret and construct the processes of optimization problems solving from different biological points of view, based on which a large number of derived approaches have been developed [1]. An example of particularly successful research direction in CIAs is genetic algorithm (GA) [2] which generates solutions to optimization problems using techniques inspired by natural evolution, such as selection, crossover, and mutation. Differential evolution (DE) [3, 4], proposed by Storn and Price, is similar to GA in the sense that it uses same evolutionary operators like mutation, crossover, and selection for guiding the population towards the optimum solution. Ant colony optimization (ACO) [5] is another very popular CIA, which was initially proposed by Marco Dorigo in his Ph.D. thesis. The first ACO was aiming to search for an optimal path in a graph, based on the behavior of ants seeking a path between their colony and a source of food.

The original idea has since diversified to solve a wider class of numerical problems, and as a result, several problems have emerged, drawing on various aspects of the behavior of ants. Particle swarm optimization (PSO) [6], introduced in 1995, was intended for simulating birds flocking or fishes schooling. PSO optimizes a problem by having a population of particles and moving these particles around in the search space according to mathematical formulae over the particle's position and velocity. Innumerable variants and hybrids of these techniques have been proposed, and many more applications of CIAs to specific problems have been reported [7–9]. In the basic methodology, however, these algorithms have some features in common: (i) looking for global solutions of optimization problems by means of random search with heuristic information; (ii) promoting problem solving efficiency by multipoint search strategies; and (iii) synthesizing the information involved in a number of currently obtained candidate solutions to produce subsequent search decision. Due to the similarity, these algorithms possess several common advantages such as the robustness

against the faint mathematical descriptions and the odd mathematical properties of optimization problems, as well as the capability to attain global optima. But they also share some shortcomings. One of the most well-known disadvantages is that the algorithm parameters are determined by experience rather than theoretical guidance. Therefore, when the algorithm parameters are improperly set up, the search process is apt to sink into a stagnant status; that is, no any better candidate solutions that are superior to their parents will be yielded (or the candidate solutions converged too early, resulting in being suboptimal).

A number of theoretical studies have been conducted in order to seek the seeds of the stagnation problems. Rudolph [10] analyzed the convergence properties of the canonical GA (CGA) and proved that in the sense of probability, CGA does not converge to global optima for any selection of initial population, crossover operator, and objective function, but some variants of CGA with an elite reserve policy are shown to converge in probability. Xu et al. [11] considered a class of GA with the strategy in which parental populations are put into competition with their offspring. It is shown that a variant of CGA with such a strategy can reach global optima after a finite number of iterations in probability one. Rastegar [12] gave the lower boundaries of the probability in which the compact genetic algorithm (cGA) converges to global optima and proved that the algorithm converges almost surely to the global optima of a special class of functions as the learning rate is sufficiently small. In the studies on ACO, Gütjahr [13, 14] first proved the convergence of the graph-based ant system by means of the graph theory. It is shown that under certain conditions, the candidate solutions generated in each iteration converge to the global solution of a given problem instance in a probability that can be made arbitrarily close to one. Stützle and Dorigo [15] and Badr and Fahmy [16] studied the convergence properties of ACO with the aid of the algebraic theory, but their work was incomplete due to too many hypotheses made in their proofs. Duan et al. [17] regarded the pheromone trail vector in ACO as a Markov chain and presented a new approach for analyzing the almost sure convergence properties of ACO based on the martingale theory. However, the pheromone trail vector is not self-contained if the path information is unknown. Thus further discussions are needed for Duan's conclusions. The convergence of PSO is also an issue of concern in the work on CIAs. Clerc and Kennedy [18] analyzed a sole particle's trajectory in simplified PSO systems and led their analysis to a generalized algorithm model in which a set of coefficients are used to control the system convergence. Based on Clerc's analysis, the dynamic behavior and the convergence of the simplified (deterministic) PSO algorithm were analyzed by Trelea [19] with the aid of tools from the discrete-time dynamic system theory. Van den Bergh and Engelbrecht [20] investigated the influence of the inertia factor on particle trajectories in PSO and provided a formal proof that each particle converges to a stable point. Poli [21] presented a method that allowed one to exactly determine the sampling distribution of PSO and explained how it changes over time during stagnation for a large class of PSO variants. Liu et al. [22] proved that PSO converges to global optima in probability one based on an

iterative function model and the probabilistic theory. Chen and Jiang [23] analyzed the behavior of PSO on the facet of particle interaction and proposed a statistical interpretation of PSO in order to capture the stochastic behavior of the entire swarm. Based on a social-only model, they derived the upper and lower boundaries of the expected particle norm.

The basic methodology in the aforementioned studies is that the convergence properties are first investigated from a specific implementation of GA, ACO, or PSO and then to be extended to a wider area of the corresponding algorithm. In this paper, the authors attempt to analyze the general convergence properties of these algorithms on the basis of their common features in methodology. The main attention is paid to the characteristics of heuristic search strategies which play crucial roles in problem solving. The purpose of this paper is to have a deeper understanding on the essence of the convergence of the CIAs and, based on which, to develop more efficient algorithms for complex problems optimization.

The rest of this paper is organized as follows. In Section 2, a computational model of a class of CIAs is built up on the basis of several common features of the heuristic search strategies adopted by GA, ACO, and PSO, and a set of definitions are given to describe the external and internal changes of a solution set updating sequence in CIAs. In Section 3, the concepts of convergence in probability and almost sure convergence are defined for the solution set updating sequences generated in the search process of the computational model, and their relations are discussed. In Section 4, the conditions under which the computational model converges almost surely to a space of weakly optimal solution sets are studied based on the Markov chain properties of the solution set updating sequences. Then the almost sure convergence to a strongly optimal solution space is proved for the computational model with the aid of the martingale theory. Several conclusive remarks are given in Section 5.

2. The CIA Model

In this section, the characteristics of solution sets representation and solution sets update in CIAs are analyzed, and based on that, the computational model of a class of the CIAs is established. Moreover, variation rate of set component and progress rate of solution set are defined, respectively, to describe the variety and the optimality of the solution sets generated in the search process of the model.

Without loss of generality, consider the optimization problem of the following type:

$$\max_x \{f(X) \mid X \in \mathbf{S}\}, \quad (1)$$

where \mathbf{S} is usually called a search (or solution) space, and $f(X)$ is an evaluation function satisfying $0 < f(X) < \infty$.

A problem-solving procedure observing the following principles is stipulated as the computational model of a class of the CIAs ("the CIA model" for short) under study in this paper.

Step 1. Transform the feasible space \mathbf{S} into the coded space of variables.

Step 2. Generate a certain number of candidate solutions to composite an initial solution set.

Step 3. Produce a new generation of solution set based on the heuristic information involved in the current solution set and corresponding evaluation values, and set the size of the solution set fixed.

Step 4. Stop the iterations and output the best candidate solution within the current solution set if a predefined termination criterion is satisfied; otherwise go back to Step 3.

Obviously, the canonical forms of GA, ACO, and PSO and most of their derived algorithms obey the aforementioned computational rules, thus can be regarded as the realization instances of the CIA model.

We consider a set of candidate solutions (“solution set” for short) $\mathbf{X} = \{X_1, X_2, \dots, X_N\} \in \mathbf{S}^N$ composed by N strings of length L in CIAs, where \mathbf{S} is the search space of candidate solutions, and $\mathbf{S}^N = \underbrace{\mathbf{S} \times \mathbf{S} \times \dots \times \mathbf{S}}_N$ is the search space of

solution sets of N th order. A candidate solution is composed of L bit numbers as its components (“solution components” for short). The candidate solution and its component have the special meaning in different CIAs. Table 1 illustrates the expressions of solution candidate and solution component in GA, ACO, and PSO. Therefore, a solution set \mathbf{X} generated by CIAs can be written in the following vector and matrix forms:

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1L} \\ x_{21} & x_{22} & \cdots & x_{2L} \\ \vdots & \vdots & \vdots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NL} \end{bmatrix} = [C_1, C_2, \dots, C_L], \quad (2)$$

where $X_i = (x_{i1}, x_{i2}, \dots, x_{iL}) \in \mathbf{S}$ is the i th candidate solution in \mathbf{X} , and x_{ij} is the j th solution component of the i th candidate solution, $i = 1, 2, \dots, N$, $j = 1, 2, \dots, L$. For the sake of convenience, the j th vector which consists of the j th components of all the candidate solutions in \mathbf{X} is called the j th set component, written as $C_j = (x_{1j}, x_{2j}, \dots, x_{Nj})^T$.

As we know, a new generation of solution set in the CIA model generally relies on the heuristic information involved in the current solution set rather than the previous solution sets. Thereby, a state transition sequence of solution set can be expressed as a Markov chain, denoted by $\{\mathbf{X}(t), t \geq 0\}$ [24]. In essence, the search procedure of the CIA model can be regarded as the updating process of solution sets expressed by

$$\mathbf{X}(t+1) = H(\mathbf{X}(t)), \quad (3)$$

where $H(\cdot)$ is the solution set updating operator or in other words, the heuristic search strategy adopted by the CIA model. It is instantiated by the evolutionary operations composed of selection, crossover, and mutation in GA; by the path search guided by pheromone updating and path

selection in ACO; and by the particle motion adjusted by velocity and position updating in PSO.

It is well known that “elite reserve” is a commonly used strategy in many realisation instances of the CIA model. In this strategy, the best candidate solution of the previous generation is always retained in the current generation, so the best candidate solution obtained so far will not be missed in the future search process. We assume that an elitist strategy is always adopted by the CIA model.

It is reasonable to estimate the evolution trend of a solution set by the change of its solution components. In order to quantify the variation of solution components, the solution set \mathbf{X} is first pretreated as a canonical form in the following way (see Figure 1). First of all, according to the element values of set component C_1 , solution set \mathbf{X} can be divided into two parts:

$$X_1^0 = \left\{ X_i \text{ satisfying } x_{i1} > \frac{\sum_{i=1}^N x_{i1}}{N}, \forall i \in \{1, 2, \dots, N\} \right\},$$

$$X_1^1 = \left\{ X_i \text{ satisfying } x_{i1} \leq \frac{\sum_{i=1}^N x_{i1}}{N}, \forall i \in \{1, 2, \dots, N\} \right\}, \quad (4)$$

where $X_1^0 \cup X_1^1 = \mathbf{X}$ and $X_1^0 \cap X_1^1 = \emptyset$.

Then, depending upon the element values of set component C_2 , X_1^0 and X_1^1 can be further decomposed into two parts, respectively, denoted by

$$X_1^0 X_2^0 = \left\{ X_i \text{ satisfying } \left(x_{i1} > \frac{\sum_{i=1}^N x_{i1}}{N}, \right. \right.$$

$$\left. \left(x_{i2} > \frac{\sum_{i=1}^N x_{i2}}{N} \right), \forall i \in \{1, 2, \dots, N\} \right\},$$

$$X_1^0 X_2^1 = \left\{ X_i \text{ satisfying } \left(x_{i1} > \frac{\sum_{i=1}^N x_{i1}}{N}, \right. \right.$$

$$\left. \left(x_{i2} \leq \frac{\sum_{i=1}^N x_{i2}}{N} \right), \forall i \in \{1, 2, \dots, N\} \right\}, \quad (5)$$

$$X_1^1 X_2^0 = \left\{ X_i \text{ satisfying } \left(x_{i1} \leq \frac{\sum_{i=1}^N x_{i1}}{N}, \right. \right.$$

$$\left. \left(x_{i2} > \frac{\sum_{i=1}^N x_{i2}}{N} \right), \forall i \in \{1, 2, \dots, N\} \right\},$$

$$X_1^1 X_2^1 = \left\{ X_i \text{ satisfying } \left(x_{i1} \leq \frac{\sum_{i=1}^N x_{i1}}{N}, \right. \right.$$

$$\left. \left(x_{i2} \leq \frac{\sum_{i=1}^N x_{i2}}{N} \right), \forall i \in \{1, 2, \dots, N\} \right\},$$

where $X_1^0 X_2^0 \cup X_1^0 X_2^1 = X_1^0$, $X_1^0 X_2^0 \cap X_1^0 X_2^1 = \emptyset$, $X_1^1 X_2^0 \cup X_1^1 X_2^1 = X_1^1$, and $X_1^1 X_2^0 \cap X_1^1 X_2^1 = \emptyset$. The process continues until the last set component C_L is treated.

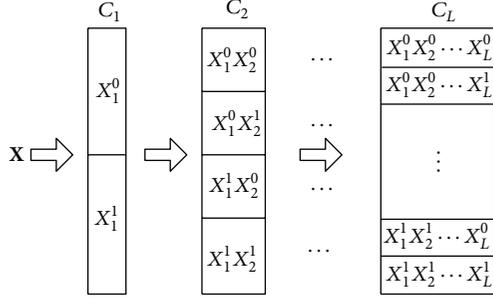


FIGURE 1: The sorting process of the solution candidates in \mathbf{X} .

TABLE 1: The expressions of solution candidate and solution component.

CIA model	GA	ACO	PSO
Solution candidate	Chromosome	Searching path	Multidimensional position of a particle
Solution component	Gene	Path segment	One-dimensional position of a particle

Let $T(t)$ be a descriptive function of the set component variation from the $(t-1)$ th generation to the t th generation of solution set based on the canonical form, defined by

$$T(t) = [T_1(t), T_2(t), \dots, T_L(t)] = \mathbf{X}(t) \oplus \mathbf{X}(t-1), \quad (6)$$

where the j th element of $T(t)$, $T_j(t) = [x_{1j}(t) \oplus x_{1j}(t-1), \dots, x_{Nj}(t) \oplus x_{Nj}(t-1)]^T$. The sign “ \oplus ” expresses the operator which is very similar to the logical exclusive-OR; that is, the output of two compared values is 0 only if both $x_{ij}(t-1)$ and $x_{ij}(t)$ are the same. Equation (6) implies that an element of $T(t)$ is 1 if the corresponding components of $\mathbf{X}(t)$ and $\mathbf{X}(t-1)$ have different values; otherwise the element is 0.

Based on the descriptive function $T(t)$ we can define an index to measure the evolution trend of a solution set quantitatively.

Definition 1 (variation rate of set component). Let $\{\mathbf{X}(t), t \geq 0\}$ be the solution set updating sequence generated by the CIA model, and let $\sum T_j$ be the sum of all the element values in T_j . We call $p_j(t) = \sum T_j(t)/N$ the variation rate of the j th set component of solution set \mathbf{X} evolving from $\mathbf{X}(t-1)$ to $\mathbf{X}(t)$.

From Definition 1 we know that in view of the fact that $0 \leq \sum T_j(t) \leq N$, $0 \leq p_j(t) \leq 1$, for all t . The variation rate of a set component measures the variation of the set component by calculating the percentage of its elements which change their values from the previous generation to the current generation. Thus, if $p_j(t) = 0$, then all the elements in set component C_j keep unchanged. Otherwise, if $p_j(t) = 1$, it implies that all the values in C_j have changed from $\mathbf{X}(t-1)$ to $\mathbf{X}(t)$.

Without loss of generality, it is assumed that the objective function $f(\cdot)$ of the optimization problem (1) is used as the evaluation function in the CIA model. In order to quantitatively estimate the optimality of the candidate solutions in a

solution set, it is useful to introduce the concept of progress rate of solution set as follows.

Definition 2 (progress rate of solution set). Let $\{\mathbf{X}(t), t \geq 0\}$ be the solution set updating sequence generated by the CIA model. Let $\bar{f}(\mathbf{X}(t))$ denote the mean evaluation value of the candidate solutions in solution set $\mathbf{X}(t)$, calculated by $\bar{f}(\mathbf{X}(t)) = \sum_{i=1}^N f(X_i(t))/N$; and $\hat{f}(\mathbf{X}(t))$ the maximal evaluation value of the candidate solutions in $\mathbf{X}(t)$, calculated by $\hat{f}(\mathbf{X}(t)) = \max_{i=1,2,\dots,N} f(X_i(t))$. Then $\alpha(t) = \bar{f}(\mathbf{X}(t))/\hat{f}(\mathbf{X}(t))$ is called the progress rate of solution set \mathbf{X} at the t th generation.

From Definition 2, we know that $0 < \alpha(t) \leq 1$. $\alpha(t)$ approaches 0 if most of the candidate solutions are far worse than the best candidate solution in $\mathbf{X}(t)$. On the other hand, $\alpha(t)$ approaches 1 if the evaluation values of most of the candidate solutions in $\mathbf{X}(t)$ are very close to the evaluation value of the best candidate solution.

3. The Basic Concepts on the Convergence of the CIA Model

In this section, we first present two different notions of convergence of random variables in probability theory. Then, these concepts are introduced into the solution set updating sequence in the CIA model. Finally, the relationships among the different types of convergence are discussed.

There exist several different notions of convergence of random variables, and the convergence of sequences of random variables to some limit random variable is an important concept in probability theory. We assume that $\{\delta_n, n = 1, 2, \dots\}$ is a sequence of random variables, and δ is a random variable.

Definition 3 (convergence in probability [25]). A sequence $\{\delta_n, n = 1, 2, \dots\}$ of random variables converges in probability towards δ if for all $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} P(|\delta_n - \delta| \geq \varepsilon) = 0. \quad (7)$$

Definition 4 (almost sure convergence [25]). A sequence $\{\delta_n, n = 1, 2, \dots\}$ of random variables converges almost surely or almost everywhere towards δ means that

$$P\left(\lim_{n \rightarrow \infty} \delta_n = \delta\right) = 1. \quad (8)$$

This means that the values of δ_n approach the value of δ , in the sense that events for which δ_n does not converge to δ have probability 0.

Almost sure convergence implies convergence in probability. Convergence in probability is the type of convergence established by the weak law of large numbers, while almost sure convergence is the type of convergence established by the strong law of large numbers.

The definition of convergence may be extended from random variables to more complex random elements in CIAs. Before giving the definitions of the convergence of the CIAs,

we explain two different concepts of the space of optimal solution sets.

Let the global optimal value of the evaluation function $f(X)$ be denoted by $f_{\max} = \max_{X \in \mathbf{S}} f(X)$, and let the set of the global optimal solutions of problem (1) be represented as $OC = \{X \in \mathbf{S} \mid \text{for all } Y \in \mathbf{S}, f(X) \geq f(Y)\}$. Obviously, for all $X \in OC$, we have $f(X) = f_{\max}$.

The solution set $\mathbf{X} \in \mathbf{S}^N$ satisfying $\widehat{f}(\mathbf{X}) = f_{\max}$ is called the weakly optimal solution set. All the weakly optimal solution sets produced by the CIA model in its search process constitute a space of weakly optimal solution set, denoted by $\mathbf{WOS} = \{\mathbf{X} \in \mathbf{S}^N \mid \widehat{f}(\mathbf{X}) = f_{\max}\}$. The solution set $\mathbf{X} \in \mathbf{S}^N$ satisfying $\overline{f}(\mathbf{X}) = f_{\max}$ is called the strongly optimal solution set. All the strongly optimal solution sets constitute a space of strongly optimal solution set, denoted by $\mathbf{SOS} = \{\mathbf{X} \in \mathbf{S}^N \mid \overline{f}(\mathbf{X}) = f_{\max}\}$.

From the aforementioned definitions, we know that $\exists X_i \in \mathbf{X} \in \mathbf{WOS}$, the equality $f(X_i) = f_{\max}$ holds. Otherwise, if for all $X_i \in \mathbf{X} \notin \mathbf{WOS}$, the inequality $f(X_i) < f_{\max}$ can be deduced. Similarly, if for all $X_i \in \mathbf{X} \in \mathbf{SOS}$, we have $f(X_i) = f_{\max}$, and if $\exists X_i \in \mathbf{X} \notin \mathbf{SOS}$, we can conclude that $f(X_i) < f_{\max}$. So it is easy to deduce the relationship among the space of strongly optimal solution sets, the space of weakly optimal solution sets, and the coded space of solution sets, that is, $\mathbf{SOS} \subset \mathbf{WOS} \subset \mathbf{S}^N$.

In view of the aforementioned notions, there are four kinds of convergence related to a solution set updating sequence in the CIA model.

The CIA model is said to converge weakly in probability if its solution set updating sequence $\{\mathbf{X}(t), t \geq 0\}$ satisfies $\lim_{t \rightarrow \infty} P\{\mathbf{X}(t) \in \mathbf{WOS}\} = 1$, for all $\mathbf{X}(0) \in \mathbf{S}^N$. The CIA model is said to converge strongly in probability if the sequence $\{\mathbf{X}(t), t \geq 0\}$ satisfies $\lim_{t \rightarrow \infty} P\{\mathbf{X}(t) \in \mathbf{SOS}\} = 1$, for all $\mathbf{X}(0) \in \mathbf{S}^N$.

The CIA model is said to weakly converge almost surely if its solution set updating sequence $\{\mathbf{X}(t), t \geq 0\}$ satisfies $P\{\lim_{t \rightarrow \infty} [\mathbf{X}(t) \in \mathbf{WOS}]\} = 1$, for all $\mathbf{X}(0) \in \mathbf{S}^N$. The CIA model is said to strongly converge almost surely if the sequence $\{\mathbf{X}(t), t \geq 0\}$ satisfies $P\{\lim_{t \rightarrow \infty} [\mathbf{X}(t) \in \mathbf{SOS}]\} = 1$, for all $\mathbf{X}(0) \in \mathbf{S}^N$.

Based on the definitions, the almost surely weak convergence of the CIA model can also be defined by $P\{\lim_{t \rightarrow \infty} \widehat{f}(\mathbf{X}(t)) = f_{\max}\} = 1$, and the almost surely strong convergence of the CIA model can also be defined by $P\{\lim_{t \rightarrow \infty} \overline{f}(\mathbf{X}(t)) = f_{\max}\} = 1$.

According to the above definitions, we can deduce the relationship among the four types of convergence.

Theorem 5. *There are the following relations among the four types of convergence of the CIA model.*

- (i) *Almost surely strong convergence \Rightarrow almost surely weak convergence \Rightarrow weak convergence in probability.*
- (ii) *Almost surely strong convergence \Rightarrow strong convergence in probability \Rightarrow weak convergence in probability.*

Proof. Based on the definition of the space of optimal solution set, we have

$$[\mathbf{X}(t) \in \mathbf{SOS}] \Rightarrow [\mathbf{X}(t) \in \mathbf{WOS}]. \quad (9)$$

From (9) we can derive that

$$P\{[\mathbf{X}(t) \in \mathbf{SOS}]\} \leq P\{[\mathbf{X}(t) \in \mathbf{WOS}]\}. \quad (10)$$

Hence we have

$$\begin{aligned} \lim_{t \rightarrow \infty} P\{[\mathbf{X}(t) \in \mathbf{SOS}]\} &\Rightarrow \lim_{t \rightarrow \infty} P\{[\mathbf{X}(t) \in \mathbf{WOS}]\}, \\ P\left\{\lim_{t \rightarrow \infty} [\mathbf{X}(t) \in \mathbf{SOS}]\right\} &\Rightarrow P\left\{\lim_{t \rightarrow \infty} [\mathbf{X}(t) \in \mathbf{WOS}]\right\}. \end{aligned} \quad (11)$$

Owing to the fact that

$$\lim_{t \rightarrow \infty} [\mathbf{X}(t) \in \mathbf{WOS}] = \bigcup_{t=1}^{\infty} \bigcap_{k=t}^{\infty} [\mathbf{X}(t) \in \mathbf{WOS}], \quad (12)$$

according to the monotonicity of probability, we obtain

$$\begin{aligned} P\left\{\lim_{t \rightarrow \infty} [\mathbf{X}(t) \in \mathbf{WOS}]\right\} &\leq \lim_{t \rightarrow \infty} P\left\{\bigcap_{k=t}^{\infty} [\mathbf{X}(t) \in \mathbf{WOS}]\right\} \\ &\leq \lim_{t \rightarrow \infty} P\{[\mathbf{X}(t) \in \mathbf{WOS}]\}. \end{aligned} \quad (13)$$

Thus, we get

$$\begin{aligned} P\left\{\lim_{t \rightarrow \infty} [\mathbf{X}(t) \in \mathbf{WOS}]\right\} &\Rightarrow \lim_{t \rightarrow \infty} P\{[\mathbf{X}(t) \in \mathbf{WOS}]\}, \\ P\left\{\lim_{t \rightarrow \infty} [\mathbf{X}(t) \in \mathbf{SOS}]\right\} &\Rightarrow \lim_{t \rightarrow \infty} P\{[\mathbf{X}(t) \in \mathbf{SOS}]\}. \end{aligned} \quad (14)$$

□

Based on Theorem 5, we know that almost surely strong convergence implies almost surely weak convergence and strong convergence in probability, and hence implies weak convergence in probability. However, the almost surely weak convergence and strong convergence in probability are not interdeducible.

4. Convergence Analysis of the CIA Model

In this section, the conditions under which the CIA model converges almost surely to the space of weakly optimal solution set are studied according to the Borel-Cantelli Lemma. And then the conditions under which the CIA model converges almost surely to the space of strongly optimal solution sets are deduced based on the submartingale convergence theorem.

Lemma 6 (Borel-Cantelli Lemma [26]). *Let $\{\xi_i, i = 1, 2, \dots\}$ be a sequence of events in a probability space and $\limsup \xi_n$*

the set of outcomes that occur infinitely many times within the infinite sequence of events (ξ_n) , then

$$\limsup_{n \rightarrow \infty} \xi_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \xi_k. \quad (15)$$

The Borel-Cantelli Lemma states the facts that:

- (i) if the sum of the probabilities of the events ξ_i , $i = 1, 2, \dots$, is finite, then the probability in which infinitely many of the events occur is zero. Namely, if $\sum_{n=1}^{\infty} P\{\xi_n\} < \infty$, then $P\{\limsup_{n \rightarrow \infty} \xi_n\} = 0$;
- (ii) if the events ξ_i , $i = 1, 2, \dots$, are independent and the sum of the probabilities of the events diverges to infinity, then the probability in which infinitely many of the events occur is one. Namely, if $\sum_{n=1}^{\infty} P\{\xi_n\} = \infty$ and the events ξ_i , $i = 1, 2, \dots$, are independent, then $P\{\limsup_{n \rightarrow \infty} \xi_n\} = 1$.

Lemma 7. Let $\{\mathbf{X}(t), t \geq 0\}$ be a solution set updating sequence generated by the CIA model and $P\{\mathbf{X}(t+1) \mid \mathbf{X}(t)\}$ the one-step transition probability of $\mathbf{X}(t)$. If the mutation rate of set components satisfies $p_j(t) > 0$, for all $j \in \{1, 2, \dots, L\}$, then $\exists c \in (0, 1)$, satisfying

- (i) $P\{\mathbf{X}(t+1) \in \mathbf{WOS} \mid \mathbf{X}(t) \in \mathbf{WOS}\} = 1$;
- (ii) $P\{\mathbf{X}(t+1) \in \mathbf{WOS} \mid \mathbf{X}(t) \notin \mathbf{WOS}\} \geq c$.

Proof. (1) Consider the case of $\mathbf{X}(t) \in \mathbf{WOS}$. By the definition in Section 3, we know

$$\widehat{f}(\mathbf{X}(t)) = f_{\max}. \quad (16)$$

Since an elite reserve strategy is adopted in the CIA model, from (16) we get

$$\widehat{f}(\mathbf{X}(t+1)) \geq \widehat{f}(\mathbf{X}(t)) = f_{\max}. \quad (17)$$

Thus $P\{\mathbf{X}(t+1) \in \mathbf{WOS} \mid \mathbf{X}(t) \in \mathbf{WOS}\} = 1$. The proof of the first part of Lemma 7 is completed.

(2) Consider the case of $\mathbf{X}(t) \notin \mathbf{WOS}$. Also by the definition in Section 3, we know

$$\widehat{f}(\mathbf{X}(t)) < f_{\max}. \quad (18)$$

If $\mathbf{X}(t+1) \in \mathbf{WOS}$, then $\exists X_i(t+1) \in \mathbf{X}(t+1)$, $i \in \{1, 2, \dots, N\}$, satisfying

$$\widehat{f}(\mathbf{X}(t+1)) = f(X_i(t+1)) = f_{\max}. \quad (19)$$

When $\widehat{f}(X_i(t+1)) = f_{\max}$, we have

$$P\{X_i(t+1) \mid \mathbf{X}(t) \notin \mathbf{WOS}\} \geq \prod_{j=1}^L p_j(t) > 0. \quad (20)$$

Thus $\exists c \in (0, 1)$, where $c = \min P\{X_i(t+1) \mid \mathbf{X}(t) \notin \mathbf{WOS}\}$, such that

$$\begin{aligned} P\{\mathbf{X}(t+1) \in \mathbf{WOS} \mid \mathbf{X}(t) \notin \mathbf{WOS}\} \\ = \max P\{X_i(t+1) \in \mathbf{OC} \mid \mathbf{X}(t) \notin \mathbf{WOS}\} \geq c. \end{aligned} \quad (21)$$

The proof of the second part of Lemma 7 is completed. \square

Theorem 8. Let $\{\mathbf{X}(t), t \geq 0\}$ be a solution set updating sequence generated by the CIA model and $\mathbf{X}(0)$ an initial solution set. $\{\mathbf{X}(t), t \geq 0\}$ is weakly convergent almost surely if for all $j \in \{1, 2, \dots, L\}$ and $t > 0$, the inequality $p_j(t) > 0$ holds.

Proof. Consider the case of $\mathbf{X}(0) \in \mathbf{WOS}$. By the definition in Section 3, we know

$$\widehat{f}(\mathbf{X}(0)) = f_{\max}. \quad (22)$$

Since the CIA model is assumed to adopt an elite reserve strategy, we get

$$\widehat{f}(\mathbf{X}(t)) \geq \widehat{f}(\mathbf{X}(t-1)) \geq \dots \geq \widehat{f}(\mathbf{X}(0)). \quad (23)$$

Combining (23) with (22) yields

$$\widehat{f}(\mathbf{X}(t)) = \widehat{f}(\mathbf{X}(t-1)) = \dots = \widehat{f}(\mathbf{X}(0)) = f_{\max}. \quad (24)$$

Thus, the theorem is valid for $\mathbf{X}(0) \in \mathbf{WOS}$.

Consider the case of $\mathbf{X}(0) \notin \mathbf{WOS}$. Let $p_T = P\{\mathbf{X}(T) \notin \mathbf{WOS}\}$, then we have

$$p_T = \begin{cases} 0, & \text{if } \exists t \in \{0, 1, 2, \dots, T\}, \mathbf{X}(t) \in \mathbf{WOS}, \\ p_T', & \text{if } \forall t \in \{0, 1, 2, \dots, T\}, \mathbf{X}(t) \notin \mathbf{WOS}. \end{cases} \quad (25)$$

According to Lemma 7, we obtain

$$\begin{aligned} p_T' &= P\{\mathbf{X}(t) \notin \mathbf{WOS} \mid \mathbf{X}(t-1) \notin \mathbf{WOS}, \forall t = 1, 2, \dots, T\} \\ &= P\{\mathbf{X}(T) \notin \mathbf{WOS} \mid \mathbf{X}(T-1) \notin \mathbf{WOS}\} \\ &\quad \dots P\{\mathbf{X}(1) \notin \mathbf{WOS} \mid \mathbf{X}(0) \notin \mathbf{WOS}\} \\ &\leq (1-c)^T. \end{aligned} \quad (26)$$

From (25) and (26) and considering the fact that $0 < c < 1$, we have

$$\sum_{T=1}^{\infty} p_T \leq \sum_{T=1}^{\infty} (1-c)^T = \frac{1-c}{c} < \infty. \quad (27)$$

Based on (27), the first part of Lemma 6 holds. Then we have

$$P\left\{\bigcap_{t=1}^{\infty} \bigcup_{T=t}^{\infty} \mathbf{X}(T) \notin \mathbf{WOS}\right\} = 0. \quad (28)$$

Thus $P\{\lim_{t \rightarrow \infty} \widehat{f}(\mathbf{X}(t)) = f_{\max}\} = 1$. \square

The sequence of the maximal evaluation values and the sequence of the average evaluation values of the solution sets generated by the CIA model will be used to describe the evolution process of the solution sets for further discussion. The martingale theory is introduced into the Markov chain analysis of the solution set updating sequence instead of the traditional ergodic theory. The conditions under which the CIA model converges almost surely to the space of strongly optimal solution sets are deduced based on the submartingale convergence theorem and the related conclusion in Section 2.

The concept of submartingale on a random sequence $\{\beta(k), \gamma(k), k \geq 0\}$ can be found in [26] in the effort to construct the convergence criteria for random sequences; that is, a sequence $\{\beta(k), k \geq 0\}$ is said to be a submartingale with respect to another sequence $\{\gamma(k), k \geq 0\}$ if for $\{\beta(k), \gamma(k), k \geq 0\}$ the following conditions are satisfied: (i) $E|\beta(k)| < \infty$; (ii) $\beta(k)$ is the function of $\gamma(0), \gamma(1), \dots, \gamma(k)$; and (iii) $E[\beta(k+1) | \gamma(0), \gamma(1), \dots, \gamma(k)] \geq \beta(k)$. In particular, a sequence $\{\beta(k), k \geq 0\}$ is called the sequence of submartingale if for all $k \geq 0$, $\beta(k)$ is an integrable random variable satisfying $E[\beta(k+1) | \beta(0), \beta(1), \dots, \beta(k)] \geq \beta(k)$.

Based on the concept of submartingale and Doob's upcrossing inequality, the following submartingale convergence theorem was deduced for random sequences.

Lemma 9 (submartingale convergence theorem [26]). *If $\{\beta(k), k \geq 0\}$ is a sequence of submartingale with $\sup_k E|\beta(k)| < \infty$, then $\beta(\infty) = \lim_k \beta(k)$ almost surely, and $\beta(\infty)$ is integrable; that is, $P(\lim_{k \rightarrow \infty} \beta(k) = \beta(\infty)) = 1$ and $E|\beta(\infty)| < \infty$.*

The search process of the CIA model can be regarded as the progress of the maximal evaluation values of the solution sets generated in the process, expressed by the sequence $\{\hat{f}(\mathbf{X}(t)), t \geq 0\}$; or the progress of the average evaluation values of the solution sets, expressed by the sequence $\{\bar{f}(\mathbf{X}(t)), t \geq 0\}$. Both of the sequences are Markov chains. Thus, the convergence analysis of the CIA model can be firstly transferred to the submartingale analysis of the maximal evaluation value sequence $\{\hat{f}(\mathbf{X}(t)), t \geq 0\}$ and then be further transferred to the submartingale analysis of the average evaluation value sequence $\{\bar{f}(\mathbf{X}(t)), t \geq 0\}$.

Lemma 10. *WOS is a closed set.*

Proof. Let $\{\mathbf{X}(t), t \geq 0\}$ be a solution set updating sequence generated by the CIA model. Since the elite reserve strategy is adopted in the CIA model, we have $\hat{f}(\mathbf{X}(t+1)) \geq \hat{f}(\mathbf{X}(t))$. On the other hand, according to the definition in Section 3, for all $\mathbf{Y} \in \mathbf{WOS}$ and for all $\mathbf{Z} \notin \mathbf{WOS}$, we have $\hat{f}(\mathbf{Y}) > \hat{f}(\mathbf{Z})$. Thus $\mathbf{X}(t+1) \neq \mathbf{Z}$ if $\mathbf{X}(t) = \mathbf{Y}$, which results in the conclusion that $P(\mathbf{X}(t+1) = \mathbf{Z} | \mathbf{X}(t) = \mathbf{Y}) = 0$. Then, for all $\mathbf{Y} \in \mathbf{WOS}$, we have

$$\sum_{\mathbf{Z} \notin \mathbf{WOS}} P(\mathbf{X}(t+1) = \mathbf{Z} | \mathbf{X}(t) = \mathbf{Y}) = 0, \quad (29)$$

or equivalently

$$\sum_{\mathbf{Z} \in \mathbf{WOS}} P(\mathbf{X}(t+1) = \mathbf{Z} | \mathbf{X}(t) = \mathbf{Y}) = 1. \quad (30)$$

Therefore, **WOS** is a closed set. \square

According to Lemma 10, we can prove that the maximal evaluation value sequence of solution set is a submartingale with respect to the corresponding solution set updating sequence.

Lemma 11. *The maximal evaluation value sequence $\{\hat{f}(\mathbf{X}(t)), t \geq 0\}$ generated by the CIA model is a submartingale with respect to the solution set updating sequence $\{\mathbf{X}(t), t \geq 0\}$.*

Proof. According to the non-after-effect property of Markov chain, we can conclude that

$$\begin{aligned} E[\hat{f}(\mathbf{X}(t+1)) | \mathbf{X}(0), \mathbf{X}(1), \dots, \mathbf{X}(t)] \\ = E[\hat{f}(\mathbf{X}(t+1)) | \mathbf{X}(t)], \end{aligned} \quad (31)$$

based on which we only need to prove that for all $\mathbf{Y} \in \mathbf{S}^N$ the following inequality is true:

$$E[\hat{f}(\mathbf{X}(t+1)) | \mathbf{X}(t) = \mathbf{Y}] \geq \hat{f}(\mathbf{X}(t)), \quad \forall t \geq 0. \quad (32)$$

According to the theory of random process [26], we have

$$\begin{aligned} E[\hat{f}(\mathbf{X}(t+1)) | \mathbf{X}(t) = \mathbf{Y}] \\ = \sum_{\mathbf{Z} \in \mathbf{S}^N} \hat{f}(\mathbf{Z}) P(\mathbf{X}(t+1) = \mathbf{Z} | \mathbf{X}(t) = \mathbf{Y}) \\ = \sum_{\mathbf{Z} \in \mathbf{WOS}} \hat{f}(\mathbf{Z}) P(\mathbf{X}(t+1) = \mathbf{Z} | \mathbf{X}(t) = \mathbf{Y}) \\ + \sum_{\mathbf{Z} \notin \mathbf{WOS}} \hat{f}(\mathbf{Z}) P(\mathbf{X}(t+1) = \mathbf{Z} | \mathbf{X}(t) = \mathbf{Y}) \\ \geq \sum_{\mathbf{Z} \in \mathbf{WOS}} \hat{f}(\mathbf{Z}) P(\mathbf{X}(t+1) = \mathbf{Z} | \mathbf{X}(t) = \mathbf{Y}). \end{aligned} \quad (33)$$

For the maximal evaluation value sequence $\{\hat{f}(\mathbf{X}(t)), t \geq 0\}$, we have

$$\hat{f}(\mathbf{X}(t+1) = \mathbf{Z}) \geq \hat{f}(\mathbf{X}(t) = \mathbf{Y}), \quad \forall \mathbf{Z} \in \mathbf{WOS}. \quad (34)$$

Combining (34) with (33) yields

$$\begin{aligned} \sum_{\mathbf{Z} \in \mathbf{WOS}} \hat{f}(\mathbf{Z}) P(\mathbf{X}(t+1) = \mathbf{Z} | \mathbf{X}(t) = \mathbf{Y}) \\ \geq \sum_{\mathbf{Z} \in \mathbf{WOS}} \hat{f}(\mathbf{Y}) P(\mathbf{X}(t+1) = \mathbf{Z} | \mathbf{X}(t) = \mathbf{Y}) \\ = \hat{f}(\mathbf{Y}) \sum_{\mathbf{Z} \in \mathbf{WOS}} P(\mathbf{X}(t+1) = \mathbf{Z} | \mathbf{X}(t) = \mathbf{Y}). \end{aligned} \quad (35)$$

By Lemma 10, we obtain

$$\sum_{\mathbf{Z} \in \mathbf{WOS}} P(\mathbf{X}(t+1) = \mathbf{Z} | \mathbf{X}(t) = \mathbf{Y}) = 1. \quad (36)$$

From (35) and (36) it results in

$$\sum_{\mathbf{Z} \in \mathbf{WOS}} \hat{f}(\mathbf{Z}) P(\mathbf{X}(t+1) = \mathbf{Z} | \mathbf{X}(t) = \mathbf{Y}) \geq \hat{f}(\mathbf{Y}). \quad (37)$$

Equation (37) implies that $E[\hat{f}(\mathbf{X}(t+1)) | \mathbf{X}(t) = \mathbf{Y}] \geq \hat{f}(\mathbf{X}(t))$. \square

Based on Lemma 11, we can associate the average evaluation value sequence of solution set with the progress rate of solution set defined in Section 2 to obtain the following corollary.

Corollary 12. *The average evaluation value sequence $\{\bar{f}(\mathbf{X}(t)), t \geq 0\}$ is a submartingale with respect to the solution set updating sequence $\{\mathbf{X}(t), t \geq 0\}$ if the corresponding progress rate of the solution set satisfies $\alpha(t+1) \geq \alpha(t)$, for all $t \geq 0$.*

Proof. Based on Lemma 11, it can be obtained that

$$E[\hat{f}(\mathbf{X}(t+1)) | \mathbf{X}(0), \mathbf{X}(1), \dots, \mathbf{X}(t)] \geq \hat{f}(\mathbf{X}(t)), \quad \forall t \geq 0. \quad (38)$$

It follows that if $\alpha(t+1) \geq \alpha(t) > 0$, for all $t \geq 0$, we have

$$\alpha(t+1) E[\hat{f}(\mathbf{X}(t+1)) | \mathbf{X}(0), \mathbf{X}(1), \dots, \mathbf{X}(t)] \geq \alpha(t) \hat{f}(\mathbf{X}(t)). \quad (39)$$

By Definition 2, (39) implies

$$E[\bar{f}(\mathbf{X}(t+1)) | \mathbf{X}(0), \mathbf{X}(1), \dots, \mathbf{X}(t)] \geq \bar{f}(\mathbf{X}(t)). \quad (40)$$

Thus $\{\bar{f}(\mathbf{X}(t)), t \geq 0\}$ is a submartingale with respect to $\{\mathbf{X}(t), t \geq 0\}$. \square

Based on Lemma 11 and its corollary, the sufficient conditions can be obtained for the strong convergence of the solution set updating sequence generated by the CIA model.

Theorem 13. *Let $\{\mathbf{X}(t), t \geq 0\}$ be a solution set updating sequence generated by the CIA model and $p_j(t)$ the variation rate of set component $C_j(t)$ in solution set $\mathbf{X}(t)$. $\{\mathbf{X}(t), t \geq 0\}$ is strongly convergent almost surely if Corollary 12 holds and $\sum_{t=1}^{\infty} \prod_{j=1}^L p_j(t) = \infty$.*

Proof. Let the complement of the set of the globally optimal solutions be denoted by OC^C and $\{M_t, t \geq 0\}$ a sequence of random events that at least a part of the candidate solutions in solution set $\mathbf{X}(t)$ belongs to OC^C ; that is, $M_t = \{\mathbf{X}(t) \cap OC^C \neq \emptyset\} = \{\bar{f}(\mathbf{X}(t)) \neq f_{\max}\}$. Similarly, let M_t^C be the exclusive event sequence of M_t ; that is, there is no any candidate solution in solution set $\mathbf{X}(t)$ belonging to OC^C , expressed by $M_t^C = \{\mathbf{X}(t) \cap OC^C = \emptyset\} = \{\bar{f}(\mathbf{X}(t)) = f_{\max}\}$. Let $P\{M_t\}$ and $P\{M_t^C\}$ be the probability measures on M_t and M_t^C , respectively.

From Corollary 12, we know that $\{\bar{f}(\mathbf{X}(t)), t \geq 0\}$ is a submartingale with respect to $\{\mathbf{X}(t), t \geq 0\}$. Then according to the properties of submartingale, it is known that

$$f_{\max} \geq E(\bar{f}(\mathbf{X}(t+1))) \geq E(\bar{f}(\mathbf{X}(t))). \quad (41)$$

According to (41) and the properties of expectation, we get

$$\int_{M_t^C} \bar{f}(\mathbf{X}(t+1)) dP = f_{\max} \cdot P\{M_t^C\}. \quad (42)$$

In addition, in view of the fact that $0 \leq \bar{f}(\mathbf{X}(t)) \leq f_{\max}$ and according to the properties of expectation, we have

$$\sup_t E(\bar{f}(\mathbf{X}(t))) \leq \sup_t f_{\max} < \infty. \quad (43)$$

Then according to Lemma 9, (43) results in the conclusion that $\exists \mathbf{X}(\infty) = \lim_t \mathbf{X}(t)$, $P\{\lim_{t \rightarrow \infty} \mathbf{X}(t) = \mathbf{X}(\infty)\} = 1$. In other words, $\{\mathbf{X}(t), t \geq 0\}$ is almost surely convergent.

In the following we further prove that the average evaluation value sequence of $\{\mathbf{X}(t), t \geq 0\}$ converges to f_{\max} almost surely. From (42) it can be seen that $\bar{f}(\mathbf{X}(t+1)) = f_{\max}$ if the event M_t^C occurs. Namely, M_t^C surely triggers off the event M_{t+1}^C . Thus the following conclusion holds for the event sequence $\{M_t^C, t \geq 0\}$:

$$M_t^C \subset M_{t+1}^C. \quad (44)$$

On the other hand, for any solution sets \mathbf{Y} and \mathbf{Z} , if $\mathbf{Y} \cap OC^C \neq \emptyset$, then $\exists Y_k \in \mathbf{Y}$, $f(Y_k) < f_{\max}$; and if $\mathbf{Z} \cap OC^C = \emptyset$, then for all $Z_k \in \mathbf{Z}$, $f(Z_k) = f_{\max}$. Hence the following inequality holds:

$$\bar{f}(\mathbf{Z}) \geq \bar{f}(\mathbf{Y}) + \frac{\delta}{N}, \quad (45)$$

where $\delta = \min(f_{\max} - f(Y_k))$ and N is the size of solution set.

In addition, by the properties of conditional expectation we have

$$\begin{aligned} & E(\bar{f}(\mathbf{X}(t+1))) - E(\bar{f}(\mathbf{X}(t))) \\ &= E(E(\bar{f}(\mathbf{X}(t+1)) | \mathbf{X}(t))) - E(\bar{f}(\mathbf{X}(t))) \\ &= \sum_{\mathbf{Y}} P\{\mathbf{X}(t) = \mathbf{Y}\} \cdot \sum_{\mathbf{Z}} P\{\mathbf{Y}, \mathbf{Z}\} \cdot (\bar{f}(\mathbf{Z}) - \bar{f}(\mathbf{Y})) \\ &\geq \sum_{\mathbf{Y} \cap OC^C \neq \emptyset} P\{\mathbf{X}(t) = \mathbf{Y}\} \cdot \sum_{\mathbf{Z}} P\{\mathbf{Y}, \mathbf{Z}\} \cdot (\bar{f}(\mathbf{Z}) - \bar{f}(\mathbf{Y})) \end{aligned} \quad (46)$$

and from the definition of M_t , we have

$$\sum_{\mathbf{Y} \cap OC^C \neq \emptyset} P\{\mathbf{X}(t) = \mathbf{Y}\} = P\{M_t\}. \quad (47)$$

Moreover, by the definition in Section 3, we know that

$$\sum_{\mathbf{Z}} P\{\mathbf{Y}, \mathbf{Z}\} \geq \prod_{j=1}^L p_j(t). \quad (48)$$

Substituting (45), (47), and (48) into (46), we obtain

$$E(\bar{f}(\mathbf{X}(t+1))) - E(\bar{f}(\mathbf{X}(t))) \geq \frac{\delta}{N} P\{M_t\} \cdot \prod_{j=1}^L p_j(t). \quad (49)$$

Now calculating the sum for both sides of the inequality in (49) from $t = 1$ to t , it yields

$$\begin{aligned} f_{\max} &\geq E(\bar{f}(\mathbf{X}(t+1))) - E(\bar{f}(\mathbf{X}(1))) \\ &\geq \frac{\delta}{N} \sum_{k=1}^t P\{M_k\} \cdot \prod_{j=1}^L p_j(k). \end{aligned} \quad (50)$$

Note that (50) holds for any t . Then let $t \rightarrow \infty$ in (50); we have

$$\sum_{t=1}^{\infty} P\{M_t\} \cdot \prod_{j=1}^L p_j(t) \leq \frac{N}{\delta} f_{\max} < \infty. \quad (51)$$

It follows that $P\{M_t\} \rightarrow 0$ and $P\{M_t^C\} \rightarrow 1$ if $\sum_{t=1}^{\infty} \prod_{j=1}^L p_j(t) = \infty$, which further leads to $P\{M_{t+1}^C\} \rightarrow 1$ according to (44). Thus we have

$$P\left\{\lim_{t \rightarrow \infty} M_t^C\right\} = P\left\{\lim_{t \rightarrow \infty} \bar{f}(\mathbf{X}(t)) = f_{\max}\right\} = 1. \quad (52)$$

□

Remark 14. The method shown in this paper has a certain similarity to methods reported in [11] for GA and in [17] for ACO. More specifically, the elitist strategy has the same benefits as the strategy of the parents population competing with their offsprings, and the condition of variation rates of set component satisfying $\sum_{t=1}^{\infty} \prod_{j=1}^L p_j(t) = \infty$ in Theorem 13 is expected to act in a similar manner as the condition of minimum mutation probability of population space satisfying $\sum_{n=1}^{\infty} r_n = \infty$ in [11]. However, the method reported in [11] only applies to GA rather than a class of CIAs. Furthermore, the variation rates of set component are measurable while minimum mutation probability of population space is not. Duan's convergence proof of ant colony algorithm is also based on the Markov chain and the martingale theory, but they ignore important information [17]. Namely, the vector of pheromone values $\{\tau(n), n \geq 1\}$, without the best-found path $\hat{\omega}(n-1)$, is not a Markov process and cannot be defined in a self-contained way. However, the stochastic process with states $\{X_n = (\tau(n), \hat{\omega}(n-1)), n \geq 1\}$ is an inhomogeneous Markov process in discrete time [14].

5. Conclusions

CIAs is a relatively new interdisciplinary field of research, which has gained huge popularity in these days. One of the most important problems obsessing the researchers in the area of CIAs is the premature convergence or the stagnation of an algorithm. The solution to this problem is not only helpful to the setup of the parameters in an existing algorithm for fast convergence to global optima but also related to how to construct a more effective and efficient algorithm for complex problems optimization. In this paper, we first analyze the common characteristics shared by GA, ACO, and PSO, and based on that, a computational model is constructed for a class of CIAs. Two quantification indices, that is, the variation rate and the progress rate, are subsequently defined to estimate the evolution trend of the solution sets generated by CIAs. Secondly, we define four different notions of convergence of a solution set updating sequence generated by the CIA model, and their relations are discussed. Thirdly, a weak convergence criterion is presented by associating the almost surely weak convergence with the positive definiteness of the variation rate of set component. Finally, the conditions for strong convergence of the CIAs are derived based on submartingale convergence theorem. Two

sufficient conditions are obtained for the almost surely strong convergence: (i) the progress rate of solution set does not diminish in the search process; and (ii) the summation of the Cartesian products of the variation rates of set component for all solution sets is infinite.

It is obvious that the conditions described in Theorems 8 and 13 are sufficient but not necessary; thus, there exists the possibility that these theorems are too conservative to be used in practice. Moreover, a fast convergence rate is absolutely critical for a wide range of applications of CIAs, but it cannot be estimated by using the results presented in this paper. Further research work will be conducted by the authors in an effort to solve these interesting and important problems.

Acknowledgments

This paper was supported in part by the National Natural Science Foundation of China (Grant no. 61203365), the Jiangsu Province Natural Science Foundation (Grant no. BK2012149), and the Open Issues Foundation of Jiangsu Key Laboratory of Power Transmission and Distribution Equipment Technology (Grant no. 2010JSSPD06).

References

- [1] L. Bianchi, M. Dorigo, L. M. Gambardella, and W. J. Gutjahr, "A survey on metaheuristics for stochastic combinatorial optimization," *Natural Computing*, vol. 8, no. 2, pp. 239–287, 2009.
- [2] D. E. Goldberg, *Genetic Algorithms in Search Optimization and Machine Learning*, Addison-Wesley Publishing, New York, NY, USA, 1989.
- [3] R. Storn and K. V. Price, "Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces," *Journal of Global Optimization*, vol. 11, no. 4, pp. 341–359, 1997.
- [4] S. Das and P. N. Suganthan, "Differential evolution: a survey of the state-of-the-art," *IEEE Transactions on Evolutionary Computation*, vol. 15, no. 1, pp. 4–31, 2011.
- [5] M. Dorigo, *Optimization, learning and natural algorithms [Ph.D. thesis]*, Politecnico di Milano, Milan, Italy, 1992.
- [6] R. C. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *Proceedings of the 6th International Symposium on Micromachine and Human Science (MHS '95)*, Nagoya, Japan, 1995.
- [7] C. Blum and A. Roli, "Metaheuristics in combinatorial optimization: overview and conceptual comparison," *ACM Computing Surveys*, vol. 35, no. 3, pp. 268–308, 2003.
- [8] W. Deng, R. Chen, B. He et al., "A novel two-stage hybrid swarm intelligence optimization algorithm and application," *Soft Computing*, vol. 16, no. 10, pp. 1707–1722, 2012.
- [9] Y.-T. Kao and E. Zahara, "A hybrid genetic algorithm and particle swarm optimization for multimodal functions," *Applied Soft Computing Journal*, vol. 8, no. 2, pp. 849–857, 2008.
- [10] G. Rudolph, "Convergence analysis of canonical genetic algorithms," *IEEE Transactions on Neural Networks*, vol. 5, no. 1, pp. 96–101, 1994.
- [11] Z. B. Xu, Z. K. Nie, and W. X. Zhang, "Almost sure strong convergence of a class of genetic algorithms with parent-offspring competition," *Acta Mathematicae Applicatae Sinica*, vol. 25, no. 1, pp. 167–175, 2002.

- [12] R. Rastegar, "On the optimal convergence probability of univariate estimation of distribution algorithms," *Evolutionary Computation*, vol. 19, no. 2, pp. 225–248, 2011.
- [13] W. J. Gütjahr, "A graph-based ant system and its convergence," *Future Generation Computer Systems*, vol. 16, no. 9, pp. 873–888, 2000.
- [14] W. J. Gütjahr, "ACO algorithms with guaranteed convergence to the optimal solution," *Information Processing Letters*, vol. 82, no. 3, pp. 145–153, 2002.
- [15] T. Stützle and M. Dorigo, "A short convergence proof for a class of ant colony optimization algorithms," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 4, pp. 358–365, 2002.
- [16] A. Badr and A. Fahmy, "A proof of convergence of ant algorithms," *Information Sciences*, vol. 160, no. 1–4, pp. 267–279, 2004.
- [17] H. Duan, D. Wang, and X. F. Yu, "Markov chains and martingale theory based convergence proof of ant colony algorithm and its simulation platform," in *Proceedings of the 6th World Congress on Intelligent Control and Automation (WCICA '06)*, pp. 3057–3061, Dalian, China, June 2006.
- [18] M. Clerc and J. Kennedy, "The particle swarm-explosion, stability, and convergence in a multidimensional complex space," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 1, pp. 58–73, 2002.
- [19] I. C. Trelea, "The particle swarm optimization algorithm: convergence analysis and parameter selection," *Information Processing Letters*, vol. 85, no. 6, pp. 317–325, 2003.
- [20] F. van den Bergh and A. P. Engelbrecht, "A study of particle swarm optimization particle trajectories," *Information Sciences*, vol. 176, no. 8, pp. 937–971, 2006.
- [21] R. Poli, "Dynamics and stability of the sampling distribution of particle swarm optimisers via moment analysis," *Journal of Artificial Evolution and Applications*, vol. 2008, Article ID 761459, 10 pages, 2008.
- [22] H. B. Liu, A. Abraham, and V. Snasel, "Convergence analysis of swarm algorithm," in *Proceedings of the World Congress on Nature & Biologically Inspired Computing (NaBIC '09)*, pp. 1714–1719, Coimbatore, India, 2009.
- [23] Y.-P. Chen and P. Jiang, "Analysis of particle interaction in particle swarm optimization," *Theoretical Computer Science*, vol. 411, no. 21, pp. 2101–2115, 2010.
- [24] W. X. Zhang and Y. Liang, *Mathematical Foundation of Genetic Algorithms*, Xi'an Jiaotong University Press, Xi'an, China, 2003.
- [25] R. M. Dudley, *Real Analysis and Probability*, Cambridge University Press, Cambridge, UK, 1992.
- [26] Y. L. Ling, *Applied Stochastic Processes*, Tsinghua University Press, Beijing, China, 2002.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

