

Research Article

On the Convergence of Biogeography-Based Optimization for Binary Problems

Haiping Ma,^{1,2} Dan Simon,³ and Minrui Fei²

¹ Department of Electrical Engineering, Shaoxing University, Shaoxing, Zhejiang, China

² Shanghai Key Laboratory of Power Station Automation Technology, School of Mechatronic Engineering and Automation, Shanghai University, Shanghai, China

³ Department of Electrical and Computer Engineering, Cleveland State University, Cleveland, OH 44115, USA

Correspondence should be addressed to Dan Simon; d.j.simon@csuohio.edu

Received 28 January 2014; Accepted 1 May 2014; Published 22 May 2014

Academic Editor: Erik Cuevas

Copyright © 2014 Haiping Ma 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.

Biogeography-based optimization (BBO) is an evolutionary algorithm inspired by biogeography, which is the study of the migration of species between habitats. A finite Markov chain model of BBO for binary problems was derived in earlier work, and some significant theoretical results were obtained. This paper analyzes the convergence properties of BBO on binary problems based on the previously derived BBO Markov chain model. Analysis reveals that BBO with only migration and mutation never converges to the global optimum. However, BBO with elitism, which maintains the best candidate in the population from one generation to the next, converges to the global optimum. In spite of previously published differences between genetic algorithms (GAs) and BBO, this paper shows that the convergence properties of BBO are similar to those of the canonical GA. In addition, the convergence rate estimate of BBO with elitism is obtained in this paper and is confirmed by simulations for some simple representative problems.

1. Introduction

Mathematical models of biogeography describe the immigration and emigration of species between habitats. Biogeography-based optimization (BBO) was first presented in 2008 [1] and is an extension of biogeography theory to evolutionary algorithms (EAs). BBO is modeled after the immigration and emigration of species between habitats. One distinctive feature of BBO is that it uses the fitness of each candidate solution to determine its immigration and emigration rate. The immigration rate determines how likely a candidate solution is to change its decision variables, and the emigration rate determines how likely a candidate solution is to share its decision variables with other candidate solutions. Specifically, a candidate solution's emigration rate increases with fitness, and its immigration rate decreases with fitness.

Although BBO is a relatively new EA, it has demonstrated good performance on various unconstrained and constrained benchmark functions [2–5] and on real-world optimization

problems such as sensor selection [1], economic load dispatch [6], robot controller tuning [7], satellite image classification [8], and power system optimization [9]. In addition, Markov models have been derived for BBO on binary problems [10, 11]. Reference [12] discusses the conceptual, algorithmic, and performance differences between BBO and GAs using both Markov model comparisons and benchmark simulation results. These simulation and theoretical results confirm that BBO is a competitive evolutionary algorithm. But until now there have not been any theoretical results concerning its convergence properties.

We say that an optimization algorithm converges to the global optimum if the value of at least one of its candidate solutions, in the limit as the generation count approaches infinity, is equal to the global optimum of the optimization problem. Several mathematical tools have been used to analyze EA convergence [13–17]. Recent work includes the analysis of EA convergence using Markov's inequality, Chebyshev bounds, Chernoff bounds, and martingales for minimum spanning tree problems, maximum matching problems,

scheduling problems, shortest path problems, Eulerian cycle problems, multiobjective problems, and others [18–20].

Markov chain models are still some of the most frequently used methods for the analysis of EA convergence. They have been widely used in a variety of EAs, including genetic algorithms (GAs) [21–25] and simulated annealing [4, 26], to prove probabilistic convergence to the global optimum. A Markov chain is a random process which has a discrete set of possible states s_i ($i = 1, 2, \dots, T$). The probability that the system transitions from state s_i to s_j is given by p_{ij} , which is called a transition probability. The $T \times T$ matrix $\mathbf{P} = [p_{ij}]$ is called the transition matrix, where T is the total number of possible population distributions. A population distribution is a specific multiset of individuals with a cardinality that is equal to the population size. A Markov state in [11] represents a BBO population distribution. Each state represents a particular population distribution, that is, how many individuals at each point of the search space there are in the population. Probability p_{ij} is the probability that the population transitions from the i th population distribution to the j th population distribution in one generation.

This paper analyzes the global convergence properties of BBO as applied to optimization problems with binary search spaces, based on a previously derived BBO Markov model, and obtains the convergence rate estimate using homogeneous finite Markov chain properties. Section 2 gives a brief review of BBO and its Markov transition probabilities. Section 3 gives some basic definitions, obtains BBO convergence properties, and obtains the convergence rate estimate. Section 4 confirms the theory using simple numerical simulation. The convergence properties and convergence rates derived here are not surprising in view of previous EA convergence results, but this paper represents the first time that such results have been formalized for BBO. Finally, Section 5 presents some concluding remarks and directions for future work.

2. Biogeography-Based Optimization (BBO)

This section presents a review of the biogeography-based optimization (BBO) algorithm with migration and mutation (Section 2.1) and then provides a review of the Markov transition probability of BBO populations (Section 2.2).

2.1. Overview of Biogeography-Based Optimization. This section provides an overview of BBO. The review in this section is very general because it applies to optimization problems with either real domains, integer domains, binary domains, or combinations thereof.

BBO is a new optimization approach, inspired by biogeography theory, to solve general optimization problems. A biogeography habitat corresponds to a *candidate solution* to the optimization problem. A multiset of biogeography habitats corresponds to a *population* of candidate solutions. Habitat suitability index (HSI) in biogeography corresponds to the goodness of a candidate solution, which is also called *fitness* in standard EA notation. Like other EAs [27], BBO probabilistically shares information between candidate

solutions to improve candidate solution fitness. In BBO, each candidate solution is comprised of a set of features, which are also called independent variables or decision variables in the optimization literature. Note that the decision variables can be taken from sets of real numbers, integers, binary numbers, or combinations thereof, depending on the problem. Each candidate solution immigrates features from other candidate solutions based on its immigration rate and emigrates features to other candidate solutions based on its emigration rate. BBO consists of two main steps: migration and mutation.

Migration. Migration is a probabilistic operator that is intended to improve a candidate solution y_k . For each feature of a given candidate solution y_k , the candidate solution's immigration rate λ_k is used to probabilistically decide whether or not to immigrate. If immigration occurs, then the emigrating candidate solution y_j is probabilistically chosen based on the emigration rate μ_j . Migration is written as follows:

$$y_k(r) \leftarrow y_j(r), \quad (1)$$

where r is a candidate solution feature. In BBO, each candidate solution y_k has its own immigration rate λ_k and emigration rate μ_k . A good candidate solution has relatively high μ and low λ , while the converse is true for a poor candidate solution. According to [1], the immigration rate λ_k and emigration rate μ_k of the candidate solution y_k are based on linear functions and are calculated as

$$\begin{aligned} \lambda_k &= 1 - \text{fitness}(y_k) \\ \mu_k &= \text{fitness}(y_k), \end{aligned} \quad (2)$$

where fitness denotes candidate solution fitness value, which is normalized to the range $[0, 1]$. The probabilities of immigrating to y_k and of emigrating from y_k are calculated as

$$\begin{aligned} \Pr(\text{immigration to } y_k) &= \lambda_k \\ \Pr(\text{emigration from } y_k) &= \frac{\mu_k}{\sum_{j=1}^N \mu_j}, \end{aligned} \quad (3)$$

where N is the population size.

Mutation. Mutation is a probabilistic operator that randomly modifies a candidate solution feature. The purpose of mutation is to increase diversity among the population, just as in other EAs. Mutation of the i th candidate solution is implemented as shown in Algorithm 1.

In the mutation logic (Algorithm 1), $\text{rand}(a, b)$ is a uniformly distributed random number between a and b , p_m is the mutation rate, and L_r and U_r are the lower and upper search bounds of the r th independent variable. The above logic mutates each independent variable with a probability of p_m . If mutation occurs for a given independent variable, then that independent variable is replaced with a random number within its search domain.

A description of one generation of BBO is given in Algorithm 2.

```

For each candidate solution feature  $r$  in  $y_i$ 
  If  $\text{rand}(0,1) < p_m$ 
     $y_i(r) \leftarrow \text{rand}(L_r, U_r)$ 
  End if
Next  $r$ 
    
```

ALGORITHM 1

2.2. Transition Probability of Biogeography-Based Optimization. Two main steps in BBO are significant: migration and mutation. The transition probability of one generation includes migration probability and mutation probability. Consider a problem whose search space $\{x_1, \dots, x_n\}$ is binary. The search space is the set of all bit strings x_i consisting of q bits each. Therefore, the cardinality of the search space is $n = 2^q$. Suppose that BBO is currently in the t th generation. Based on the previously derived Markov chain model for BBO [11], the probability that migration results in the k th candidate solution y_k are equal to x_i at generation $t + 1$ is given by

$$\Pr(y_{k,t+1} = x_i) = \prod_{r=1}^q \left[(1 - \lambda_k) \mathbf{1}_0(y_k(r) - x_i(r)) + \lambda_k \frac{\sum_{j \in \zeta_i(r)} v_j \mu_j}{\sum_{j=1}^n v_j \mu_j} \right], \quad (4)$$

where $\mathbf{1}_0$ is the indicator function on the set 0 (i.e., $\mathbf{1}_0(a) = 1$ if $a = 0$, and $\mathbf{1}_0(a) = 0$ if $a \neq 0$), r denotes the index of the candidate solution feature (i.e., the bit number), λ_k denotes the immigration rate of candidate solution y_k , μ_j denotes the emigration rate of candidate solution x_j , and v_j denotes the number of x_j individuals in the population. The notation $\zeta_i(r)$ in (4) denotes the set of search space indices j such that the r th bit of x_j is equal to the r th bit of x_i . That is, $\zeta_i(r) = \{j : x_j(r) = x_i(r)\}$. Note that the first term in the product on the right side of (4) denotes the probability that $y_{k,t+1}(r) = x_i(r)$ if immigration of the r th candidate solution feature does not occur, and the second term denotes the probability if immigration of the r th candidate solution feature does occur.

Example 1. To clarify the notations in (4), an example is presented. We use the notation v_j to denote the number of x_j individuals in the population. Suppose we have a two-bit problem ($q = 2$, $n = 4$) with a population size $N = 3$. The search space consists of the bit strings $x = \{x_1, x_2, x_3, x_4\} = \{00, 01, 10, 11\}$. Suppose that the individuals in the current population are $y = \{x_2, x_4, x_4\} = \{01, 11, 11\}$. Then we have $v_1 = 0$, $v_2 = 1$, $v_3 = 0$, and $v_4 = 2$. To clarify the notation $\zeta_i(r)$, we now explain how to calculate $\zeta_1(1)$. We arbitrarily number bits from left to right; that is, in any given bit string, bit 1 is the leftmost bit and bit 2 is the rightmost bit. From the definition of $\zeta_i(r)$ we see that

$$\zeta_1(1) = \{j : x_j(1) = x_1(1)\}. \quad (5)$$

Since $x_1 = 00$, we see that $x_1(1) = 0$ (i.e., the leftmost bit is 0). Then (5) can be written as

$$\zeta_1(1) = \{j : x_j(1) = 0\}. \quad (6)$$

But $x_j(1) = 0$ for $x_j \in \{00, 01\}$, which in turn indicates that $x_j(1) = x_1(1)$ for $j \in [1, 2]$; therefore, $\zeta_1(1) = \{1, 2\}$. Continuing this process, we see that

$$\begin{aligned} \zeta_1(1) &= \{1, 2\}, & \zeta_1(2) &= \{1, 3\}, \\ \zeta_2(1) &= \{1, 2\}, & \zeta_2(2) &= \{2, 4\}, \\ \zeta_3(1) &= \{3, 4\}, & \zeta_3(2) &= \{1, 3\}, \\ \zeta_4(1) &= \{3, 4\}, & \zeta_4(2) &= \{2, 4\}. \end{aligned} \quad (7)$$

Mutation. Mutation operates independently on each candidate solution by probabilistically reversing each bit in each candidate solution. Suppose that the event that each bit of a candidate solution is flipped is stochastically independent and occurs with probability $p_m \in (0, 1)$. Then the probability that candidate solution x_i mutates to become x_j can be written as

$$\Pr(x_i \rightarrow x_j) = p_m^{H_{ij}} (1 - p_m)^{q - H_{ij}} > 0, \quad (8)$$

where q is the number of bits in each candidate solution and H_{ij} represents the Hamming distance between bit strings x_i and x_j .

3. Convergence of Biogeography-Based Optimization

The previous section reviewed the BBO algorithm and its Markov model. In this section, which comprises the main contribution of this paper, we use the results of Section 2 to analyze the convergence behavior of BBO. Section 3.1 gives some basic foundations of Markov transition matrices, including notation and basic theorems that we will need later. Section 3.2 reviews previously published Markov theory as it relates to BBO. This leads to Section 3.3, which obtains some important properties of the BBO transition matrix. Section 3.4 uses transition matrix properties to analyze BBO convergence to the solution of a global optimization problem. This leads to Section 3.5, which uses the BBO convergence analysis to obtain an estimate of the convergence rate.

3.1. Preliminary Foundations of Markov Theory. A finite Markov chain is a random process which has a finite number of possible state values $S = \{s_i\}$ ($i = 1, 2, \dots, T$), where T is the total number of states, which is the cardinality $|S|$. The probability that the system transitions from state s_i to s_j at time step t is given by $p_{ij}(t)$, which is called the transition probability. The $T \times T$ matrix $\mathbf{P} = (p_{ij}(t))$ is called the transition matrix, where $p_{ij} \in [0, 1]$ for $i, j \in [1, T]$, and $\sum_{j=1}^T p_{ij} = 1$ for all i . The \mathbf{P} matrix is called *stochastic* because the elements in each row sum to 1. If the transition probability is independent of t , that is, $p_{ij}(t_1) = p_{ij}(t_2)$ for all $i, j \in [1, T]$ and for all t_1 and t_2 , then the Markov chain is said

```

For each  $y_k$ , define emigration rate  $\mu_k$  proportional to fitness of  $y_k$ ,  $\mu_k \in [0, 1]$ 
For each  $y_k$ , define immigration rate  $\lambda_k = 1 - \mu_k$ 
For each  $y_i$ 
  For each candidate solution feature  $r$ 
    Use  $\lambda_i$  to probabilistically decide whether to immigrate to  $y_i$  (see (2) and (3))
    If immigrating then
      Use  $\{\mu_k\}$  to probabilistically select the emigrating solution  $y_k$  (see (2) and (3))
       $y_i(r) \leftarrow y_k(r)$ 
    End if
  Next candidate solution feature
  Probabilistically decide whether to mutate  $y_i$ 
Next solution

```

ALGORITHM 2: One generation of the BBO algorithm. y is the entire population of candidate solutions, y_k is the k th candidate solution, and $y_k(r)$ is the r th feature of y_k .

to be *homogeneous*. Given an initial probability distribution of states $\boldsymbol{\pi}(0)$ as a row vector, the probability distribution of the Markov chain after t steps is given by $\boldsymbol{\pi}(t) = \boldsymbol{\pi}(0)\mathbf{P}^t$. Therefore, a homogeneous finite Markov chain is completely specified by $\boldsymbol{\pi}(0)$ and \mathbf{P} , and the limiting distribution as $t \rightarrow \infty$ depends on the structure of \mathbf{P} . For homogeneous finite Markov chains, we have the following two theorems [28, 29].

Theorem 2 (see [28, page 123]). *Let \mathbf{P} be a primitive stochastic matrix of order T ; that is, all of the elements of \mathbf{P}^k are positive for some integer k . Then \mathbf{P}^k converges as $k \rightarrow \infty$ to a stochastic matrix which has all nonzero entries. That is, for all $i, j \in [1, T]$,*

$$\mathbf{P}^\infty = \lim_{k \rightarrow \infty} (p_{ij})^{(k)} = \begin{pmatrix} \pi \\ \vdots \\ \pi \end{pmatrix}_{T \times T}, \quad (9)$$

where $\boldsymbol{\pi} = (\pi_1, \dots, \pi_T)$, and $\pi_j \neq 0$ for $1 \leq j \leq T$.

Theorem 3 (see [28, page 126]). *Let \mathbf{P} be a stochastic matrix of order T with the structure $\mathbf{P} = \begin{bmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{R} & \mathbf{Q} \end{bmatrix}$, where \mathbf{C} is a primitive stochastic matrix of order m and $\mathbf{R}, \mathbf{Q} \neq \mathbf{0}$. Then \mathbf{P}^k converges as $k \rightarrow \infty$ to a stochastic matrix. That is,*

$$\mathbf{P}^\infty = \lim_{k \rightarrow \infty} (p_{ij})^{(k)} = \begin{pmatrix} \pi \\ \vdots \\ \pi \end{pmatrix}_{T \times T}, \quad (10)$$

where $\boldsymbol{\pi} = (\pi_1, \dots, \pi_m, 0, \dots, 0)$, and $\pi_j \neq 0$ for $1 \leq j \leq m < T$.

We will use these theorems in Section 3.3 to derive important properties of the BBO transition matrix and in Section 3.4 to derive BBO convergence properties.

3.2. BBO Markov Theory. In previous work [11], the transition probability of BBO with migration and mutation was obtained. This provides us with the probability $\Pr(\mathbf{u} | \mathbf{v})$ of transitioning in one generation from population vector $\mathbf{v} = [v_1, v_2, \dots, v_n]$, where v_i is the number of candidate solutions x_i in the population and n is the size of the search space, to

population vector $\mathbf{u} = [u_1, u_2, \dots, u_n]$. BBO can be described as a homogeneous finite Markov chain: the state of BBO is defined as the population vector, so the element p_{ij} of the state transition matrix \mathbf{P} is obtained by computing $\Pr(\mathbf{u} | \mathbf{v})$ for each possible \mathbf{v} and each possible \mathbf{u} . Namely, $p_{ij} = \Pr(\mathbf{u} | \mathbf{v})$ denotes the probability that the i th population vector, denoted as \mathbf{v} , transitions to the j th population vector, denoted as \mathbf{u} , where $i, j \in [1, T]$. Note that the cardinality of the state space is $|S| = T$, where T is the total number of possible populations; that is, T is the number of possible vectors \mathbf{v} and the number of possible vectors \mathbf{u} . The number T can be calculated in several different ways, as discussed in [11].

Let the $N \times n$ matrix $\mathbf{M} = (m_{kj})$ and the $n \times n$ matrix $\mathbf{U} = (u_{ij})$ be intermediate transition matrices corresponding to only migration and only mutation, respectively, where N is the population size and n is the cardinality of the search space. Note that $m_{kj} = \Pr(y_{k,t+1} = x_j) \geq 0$ and $u_{ij} = \Pr(x_j \rightarrow x_i) > 0$. That is, m_{kj} is the probability that the k th individual in the population transitions to the j th individual in the search space when only migration is considered, and u_{ij} is the probability that the j th individual in the search space transitions to the i th individual in the search space when only mutation is considered. We can use [11] to obtain the transition probability of the i th population state vector \mathbf{v} to the j th population state vector \mathbf{u} as

$$p_{ij} = \Pr(\mathbf{u} | \mathbf{v}) = \sum_{J \in Y} \prod_{k=1}^N \prod_{i=1}^n [Q_{ki}(\mathbf{v})]^{J_{ki}},$$

$$\text{where } Y \equiv \left\{ \mathbf{J} \in R^{N \times n} : J_{ki} \in \{0, 1\}, \right. \quad (11)$$

$$\left. \sum_{i=1}^n J_{ki} = 1 \forall k, \sum_{k=1}^N J_{ki} = u_i \forall i \right\},$$

where $Q_{ki}(\mathbf{v})$ is a single element of the product of \mathbf{M} and \mathbf{U} . The matrix composed of the $Q_{ki}(\mathbf{v})$ elements can be represented as $[Q_{ki}(\mathbf{v})] = \mathbf{M}\mathbf{U}^T$, where $k \in [1, N]$ and $i \in [1, n]$. $Q_{ki}(\mathbf{v})$ denotes the probability that the k th migration trial followed by mutation results in candidate solution x_i .

Note that $Q_{ki}(\mathbf{v})$ is a scalar, and transition matrix $\mathbf{P} = (p_{ij})$ is a $T \times T$ matrix, each element of which can be obtained by (11).

Example 4. Here we use a simple example based on [11] to clarify (11). Consider a simple BBO experiment in which a trial of migration and mutation can result in one of four possible outcomes $x_1, x_2, x_3,$ and x_4 with probabilities $Q_{i1}, Q_{i2}, Q_{i3},$ and $Q_{i4},$ respectively. Index i refers to the migration trial number (i.e., the ‘‘For each y_i ’’ loop in Algorithm 2). Assume that the total number of trials (i.e., the population size N) is equal to 2. Suppose that the probabilities are given as follows:

$$\begin{aligned} \text{Trial 1: } & Q_{11} = 0.1, Q_{12} = 0.3, Q_{13} = 0.5, Q_{14} = 0.1; \\ \text{Trial 2: } & Q_{21} = 0.1, Q_{22} = 0.1, Q_{23} = 0.1, Q_{24} = 0.7. \end{aligned}$$

In this example, we calculate the probability that x_1 and x_4 occur after two migration trials. In order to calculate this probability, let $\mathbf{c} = [C_1, C_2, C_3, C_4]$ denote a vector of random variables, where C_i is the total number of times that x_i occurs after two migration trials. Based on (11), we use $C_1 = 1, C_2 = 0, C_3 = 0,$ and $C_4 = 1$ to obtain

$$\Pr(C_1 = 1, C_2 = 0, C_3 = 0, C_4 = 1) = \sum_{J \in Y} \prod_{k=1}^2 \prod_{i=1}^4 [Q_{ki}]^{J_{ki}}, \quad (12)$$

where

$$Y = \left\{ \mathbf{J} \in R^{2 \times 4} : J_{ki} \in \{0, 1\}, \sum_{i=1}^4 J_{ki} = 1 \ \forall k, \sum_{k=1}^2 J_{ki} = C_i \ \forall i \right\}. \quad (13)$$

According to (13), \mathbf{J} belongs to Y if it satisfies the following conditions: (1) \mathbf{J} is a 2×4 matrix; (2) each element of \mathbf{J} is either 0 or 1; (3) the elements in each row of \mathbf{J} add up to 1; and (4) the elements in the i th column of \mathbf{J} add up to C_i .

Note from [11] that the cardinality of Y is

$$|Y| = \frac{N!}{(C_1! \cdots C_n!)}. \quad (14)$$

The number of matrices $\mathbf{J}^{(t)}$ that satisfy these conditions is calculated as $N!/(C_1! \cdots C_n!) = 2!/(1!1!0!1!) = 2,$ and the $\mathbf{J}^{(t)}$ matrices are found as follows:

$$\mathbf{J}^{(1)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{J}^{(2)} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}. \quad (15)$$

Substituting these matrices into (12) gives

$$\begin{aligned} \Pr(C_1 = 1, C_2 = 0, C_3 = 0, C_4 = 1) &= \sum_{t=1}^2 \left(Q_{11}^{J_{11}^{(t)}} Q_{12}^{J_{12}^{(t)}} Q_{13}^{J_{13}^{(t)}} Q_{14}^{J_{14}^{(t)}} \right) \times \left(Q_{21}^{J_{21}^{(t)}} Q_{22}^{J_{22}^{(t)}} Q_{23}^{J_{23}^{(t)}} Q_{24}^{J_{24}^{(t)}} \right) \\ &= Q_{11}Q_{24} + Q_{14}Q_{21} = 0.08. \end{aligned} \quad (16)$$

3.3. BBO Transition Matrix Properties. Recall that the migration probability and mutation probability can be calculated by (3) and (8), respectively: $m_{kj} = \Pr(y_{k,t+1} = x_j) \geq 0$ and $u_{ij} = \Pr(x_j \rightarrow x_i) > 0$. Therefore, \mathbf{M} is a nonnegative stochastic matrix; although it is not a transition matrix since it is not square, each row sums to 1. We also see that \mathbf{U} is a positive left stochastic matrix; that is, each of its columns sums to 1. We now present two theorems that show that there is a nonzero probability of obtaining any individual in the search space from any individual in a BBO population after migration and mutation. This means that there is a nonzero probability of transitioning from any population vector \mathbf{u} to any other population vector \mathbf{v} in one generation, which means that the BBO transition matrix is primitive.

Theorem 5. *If \mathbf{M} is a positive stochastic matrix and \mathbf{U} is a positive left stochastic matrix, then the product $\mathbf{M}\mathbf{U}^T$ is positive.*

Proof. If \mathbf{M} is positive and stochastic, then every entry of \mathbf{M} is positive; that is, $m_{kj} > 0$ for $k \in [1, N]$ and $j \in [1, n]$, and $\sum_{j=1}^n m_{kj} = 1$ for all k . Similarly, if \mathbf{U} is positive, then every entry of \mathbf{U} is positive; that is, $u_{ij} > 0$ for $i, j \in [1, n]$. Therefore, by matrix multiplication, $(\mathbf{M}\mathbf{U}^T)_{ki} = \sum_{j=1}^n m_{kj}u_{ij} > 0$ for $k \in [1, N]$ and $i \in [1, n]$. \square

Theorem 6. *The transition matrix of BBO with migration and mutation is primitive.*

Proof. From (11) we know that if $Q_{ki}(\mathbf{v}) = [\mathbf{M}\mathbf{U}^T]_{ki} > 0$ for all $k \in [1, N]$ and $i \in [1, n]$, then $p_{ij} = \sum_{J \in Y} \prod_{k=1}^N \prod_{i=1}^n [Q_{ki}(\mathbf{v})]^{J_{ki}} > 0$ for $i, j \in [1, T]$, where Y is given in (11). So the transition matrix $\mathbf{P} = (p_{ij})$ of BBO is positive. Therefore, \mathbf{P} is primitive since every positive transition matrix is primitive. \square

Corollary 7. *There exists a unique limiting distribution for the states of the BBO Markov chain. Also, the probability that the Markov chain is in the i th state at any time is nonzero for all $i \in [1, T]$.*

Proof. Corollary 7 is an immediate consequence of Theorems 2 and 6. \square

3.4. Convergence Properties of BBO. Before we obtain the convergence properties of BBO, some precise definitions of the term *convergence* are required [15]. Assume that the search space of a global optimization problem is I with cardinality $|I| = n$. Further assume that the BBO algorithm with population size N consists of both migration and mutation, as shown in Algorithm 2.

Definition 8. Let $A(t) = \{a_i(t) \mid i \in [1, N], a_i(t) \in I\}$ be the population at generation t , where N is the population size, $a_i(t)$ denotes an individual representing a candidate solution in the search space I , and $A(t)$ may contain duplicate elements; $f: I \rightarrow R$ denotes a fitness function assigning real values to individuals; $I^* = \{a^* \mid a^* = \arg \max\{f(a) \mid a \in I\}$ is a subset in the search space, each member of which has

the globally maximum fitness; and the best individuals in the population at generation t are $I^*(t) = \{a_j^*(t)\} \subset A(t)$, where $f(a_j^*(t)) \geq f(a_i(t))$ for all $a_j^*(t) \in I^*(t)$ and for all $i \in [1, N]$.

We use the notation $a^*(t)$ to denote an arbitrary element of $I^*(t)$ (i.e., one of the best individuals in the population at generation t). Because of migration and mutation, $a^*(t)$ and its fitness will change randomly over time. As $t \rightarrow \infty$, the convergence, or lack of convergence, of $a^*(t)$ to the subset I^* indicates whether or not the BBO algorithm is globally convergent. That is, BBO is said to converge if

$$\Pr\left(\lim_{t \rightarrow \infty} a^*(t) \in I^*\right) = 1 \iff \Pr\left(a^* \in \lim_{t \rightarrow \infty} A(t)\right) = 1. \quad (17)$$

Note that $a^*(t)$ is not necessarily unique. However, Definition 8 states that the BBO algorithm is globally convergent if and only if (17) holds for every $a^*(t)$. Clearly, the evolution of $a^*(t)$ is a homogeneous finite Markov chain, which we call an $a^*(t)$ -chain.

Now we sort all the states of I in order of descending fitness; that is, $I = \{I_1, \dots, I_n\}$, and $f(I_1) \geq f(I_2) \geq \dots \geq f(I_n)$. We define S as the set of indices of I ; that is, $S = \{1, 2, \dots, n\}$. Further, we define S^* as the elements $\{j\}$ of S such that $I_j \in I^*$; that is, $I_j \in I^*$ for all $j \in S^*$. This leads to the following definition.

Definition 9. Let $\hat{\mathbf{P}} = (\hat{p}_{ij})$ be the transition matrix of an $a^*(t)$ -chain, where \hat{p}_{ij} for $i, j \in [1, n]$ is the probability that $a^*(t) = I_i$ transitions to $a^*(t+1) = I_j$. The BBO algorithm converges to a global optimum if and only if $a^*(t)$ transitions from any state $i \in I$ to I^* as $t \rightarrow \infty$ with probability one, that is, if

$$\lim_{t \rightarrow \infty} \sum_{j \in S^*} (\hat{\mathbf{P}}^t)_{ij} = 1, \quad \forall i \in S. \quad (18)$$

As noted earlier, there may be more than one $a^*(t)$ -chain since more than one element of the search space may have a globally maximum fitness. Definition 9 states that the BBO algorithm converges to a global optimum if and only if (18) holds for every $a^*(t)$ -chain. Also note that $\hat{\mathbf{P}}$ depends on the other individuals in the population at generation t . Definition 9 states that the BBO algorithm converges to a global optimum if and only if (18) holds for every possible $\hat{\mathbf{P}}$ transition matrix for every $a^*(t)$ -chain.

Theorem 10. *If the transition matrix $\hat{\mathbf{P}} = (\hat{p}_{ij})$ of an $a^*(t)$ -chain is a positive stochastic matrix, then BBO with migration and mutation does not converge to any of the global optima.*

Proof. Since every positive matrix is also a primitive one, it follows by Theorem 2 that the limiting distribution of $\hat{\mathbf{P}}$ is unique with all nonzero entries. Therefore, for any $i \in S$,

$$\begin{aligned} \lim_{t \rightarrow \infty} \sum_{j \in S^*} (\hat{\mathbf{P}}^t)_{ij} &= 1 - \lim_{t \rightarrow \infty} \sum_{j \in S-S^*} (\hat{\mathbf{P}}^t)_{ij} \\ &= 1 - \sum_{j \in S-S^*} \lim_{t \rightarrow \infty} (\hat{\mathbf{P}}^t)_{ij} \\ &= 1 - \sum_{j=|S^*|+1}^{|S|} \pi_j < 1, \end{aligned} \quad (19)$$

where we use the notation $S - S^*$ to denote all elements of S that do not belong to S^* . We see that (18) is not satisfied, which completes the proof. \square

Theorem 11. *If the transition matrix $\hat{\mathbf{P}} = (\hat{p}_{ij})$ of an $a^*(t)$ -chain is a stochastic matrix with the structure $\hat{\mathbf{P}} = \begin{bmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{R} & \mathbf{Q} \end{bmatrix}$, where \mathbf{C} is a positive stochastic matrix of order $|S^*|$, and $\mathbf{R}, \mathbf{Q} \neq \mathbf{0}$, then the BBO algorithm converges to one or more of the global optima.*

Proof. From Theorem 3, we see that, for all $i, j \in S$,

$$\hat{\mathbf{P}}^\infty = \lim_{t \rightarrow \infty} (\hat{p}_{ij})^t = \begin{pmatrix} \boldsymbol{\pi} \\ \vdots \\ \boldsymbol{\pi} \end{pmatrix}_{|S| \times |S|}, \quad (20)$$

where $\boldsymbol{\pi} = (\pi_1, \dots, \pi_{|S^*|}, 0, \dots, 0)$, $\pi_j \neq 0$ for $1 \leq j \leq |S^*|$, and $\sum_{j=1}^{|S^*|} \pi_j = 1$. It follows directly that, for any $i \in S$,

$$\lim_{t \rightarrow \infty} \sum_{j \in S^*} (\hat{\mathbf{P}}^t)_{ij} = \sum_{j \in S^*} \lim_{t \rightarrow \infty} (\hat{\mathbf{P}}^t)_{ij} = \sum_{j=1}^{|S^*|} \pi_j = 1. \quad (21)$$

We see that (18) is satisfied which completes the proof. \square

Theorems 10 and 11 can be applied directly to determine the global convergence of BBO if the structure of the transition matrix of the Markov chain can be determined, as we will show in the remainder of this section. In particular, we will formalize the observation that the transition matrix of BBO without elitism satisfies the conditions of Theorem 10 (as stated in Theorem 6). We will further show that the transition matrix of the $a^*(t)$ -chain of BBO with elitism satisfies the conditions of Theorem 11.

Elitism. We now discuss a modified BBO which uses *elitism*, an idea which is also implemented in many EAs. There are many ways to implement elitism, but here we define elitism as the preservation of the best individual at each generation in a separate partition of the population space. This enlarges the population size by one individual; the elite individual increases the population size from N to $N + 1$. However, note that the population size is still constant (i.e., equal to $N + 1$) from one generation to the next. The elite individual does not take part in recombination or mutation

but is maintained separately from the other N -members of the population. At each generation, if an individual in the N -member main population is better than the elite individual, then the elite individual is replaced with a copy of the better individual.

Relative to a standard N -member BBO population, elite BBO increases the number of possible population distributions by a factor of n , which is the search space size. That is, each possible population distribution of the N -member main population could also include one of n elite individuals. The number of possible population distributions increases by a factor of n , from T (see Section 3.1) to nT . We order these new states so that each group of n states has the same elite individual. Also, the elite individual in the m th group of n states is the m th best individual in the search space, for $m = 1, \dots, n$.

The elitist-preserving process can be represented by an upgrade transition matrix \mathbf{O} , which contains the probabilities that each population distribution of the $(N + 1)$ -member population transitions to some other population distribution after the elitist-preserving step. That is, the element in the i th row and j th column of \mathbf{O} , denoted as $O(i, j)$, is the probability that the i th population distribution transitions to the j th population distribution after the step in which the elite individual is replaced with the best individual from the N -member main population. The upgrade matrix is similar to the one in [29]; it does not include the effects of migration or mutation but only includes the elitism-preserving step. The upgrade matrix only includes the probability of changing the elite individual; it does not include the probability of changing the N -member main population, since it does not include migration or mutation. If there are no individuals in the N -member main population that are better than the elite individual, then the elite individual does not change. The structure of the upgrade matrix \mathbf{O} can be written as

$$\mathbf{O} = \begin{pmatrix} \mathbf{O}_{11} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{O}_{21} & \mathbf{O}_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{O}_{n1} & \cdots & \mathbf{O}_{n,n-1} & \mathbf{O}_{nm} \end{pmatrix}, \quad (22)$$

where each \mathbf{O}_{ij} matrix is $T \times T$, where T is the number of population distributions in an EA with a population size of N and search space cardinality of n . \mathbf{O}_{11} is the identity matrix since the first n population distributions have the global optimum as their elite individual, and the elite individual can never be improved from the global optimum. Matrices \mathbf{O}_{aa} with $a \geq 2$ are diagonal matrices composed of all zeros and ones. Since the population distributions are ordered by grouping common elite individuals and since elite individuals in the population distribution ordering are in order of decreasing fitness, the super block diagonals in \mathbf{O} are zero matrices as shown in (22); that is, there is zero probability that the i th population distribution transitions to

the j th population distribution if $i < j$. So the Markov chain of elite BBO can be described as

$$\begin{aligned} \mathbf{P}^+ &= \text{diag}(\mathbf{P}) \mathbf{O} \\ &= \begin{pmatrix} \mathbf{P} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{P} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{P} \end{pmatrix} \begin{pmatrix} \mathbf{O}_{11} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{O}_{21} & \mathbf{O}_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{O}_{n1} & \cdots & \mathbf{O}_{n,n-1} & \mathbf{O}_{nm} \end{pmatrix} \quad (23) \\ &= \begin{pmatrix} \mathbf{PO}_{11} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{PO}_{21} & \mathbf{PO}_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{PO}_{n1} & \cdots & \mathbf{PO}_{n,n-1} & \mathbf{PO}_{nm} \end{pmatrix}, \end{aligned}$$

where \mathbf{P} is the $T \times T$ transition matrix described in Section 3.1.

Example 12. To explain the update matrix \mathbf{O} described in (22), a simple example is presented. Suppose there exists a search space consisting of $n = 3$ individuals which are $x = \{x_1, x_2, x_3\}$ where the fitness of x_1 is the lowest and the fitness of x_3 is the highest. Suppose the main population size is $N = 1$, so the elitist population size is $N + 1 = 2$. So there are nine possible populations before the elitist-preserving step, and they are $\{C_1, C_2, C_3, C_4, C_5, C_6, C_7, C_8, C_9\} = \{\{x_3, x_3\}, \{x_3, x_2\}, \{x_3, x_1\}, \{x_2, x_3\}, \{x_2, x_2\}, \{x_2, x_1\}, \{x_1, x_3\}, \{x_1, x_2\}, \{x_1, x_1\}\}$. Note that the first element in each population is the elite individual and the last N element ($N = 1$ in this example) is the main population. Also note that the populations are ordered in such a way that the first three have the most fit individual as their elite individual, the next three have the second most fit individual as their elite individual, and the last three have the least fit individual as their elite individual. The update matrix \mathbf{O} is a 9×9 matrix.

The population $C_1 = \{x_3, x_3\}$ transitions to the population $C_1 = \{x_3, x_3\}$ with probability 1; that is, $O(1, 1) = 1$. Population C_1 cannot transition to any other population $C_i (i \neq 1)$; that is, $O(1, i) = 0$ for $i \neq 1$. Similarly, population $C_2 = \{x_3, x_2\}$ transitions to C_2 with probability 1 since the elite x_3 is better than the main-member population x_2 ; therefore, $O(2, 2) = 1$, and $O(2, i) = 0$ for $i \neq 2$. Continuing with this reasoning, we obtain the \mathbf{O} matrix as follows:

$$\begin{aligned} \mathbf{O} &= \begin{bmatrix} \mathbf{O}_{11} & \mathbf{O}_{12} & \mathbf{O}_{13} \\ \mathbf{O}_{21} & \mathbf{O}_{22} & \mathbf{O}_{23} \\ \mathbf{O}_{31} & \mathbf{O}_{32} & \mathbf{O}_{33} \end{bmatrix} \\ &= \begin{bmatrix} 1 & & & & & & & & \\ 0 & 1 & & & & & & & \\ 0 & 0 & 1 & & & & & & \\ 1 & 0 & 0 & 0 & & & & & \\ 0 & 0 & 0 & 0 & 1 & & & & \\ 0 & 0 & 0 & 0 & 0 & 1 & & & \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & & \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad (24) \end{aligned}$$

where each \mathbf{O}_{ij} matrix is 3×3 and the blank elements are 0.

Now we consider the convergence of the $a^*(t)$ -chain, which is the sequence of elite individuals in the elite BBO algorithm. If the elite individual is equal to the global optimum, we call this an absorbing state of the $a^*(t)$ -chain. Recall that the elite individual in elite BBO can only be replaced by one with better fitness. Therefore, the $a^*(t)$ -chain of elite BBO contains three classes of states: (1) at least one absorbing state, (2) nonabsorbing states which transition to absorbing states in one step, and (3) nonabsorbing states which transition to nonabsorbing states in one step. So the transition matrix $\hat{\mathbf{P}}$ of the $a^*(t)$ -chain, which we introduced in (18)–(21), can be written as

$$\hat{\mathbf{P}} = \begin{pmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{R} & \mathbf{Q} \end{pmatrix}, \quad (25)$$

where \mathbf{I}_k is a $k \times k$ unit matrix corresponding to optimal individuals (k is the number of optima), \mathbf{R} is a matrix of order $(|S| - k) \times k$ corresponding to nonabsorbing states that transition to absorbing states ($|S|$ is the cardinality of the state space S , so $|S| - k$ is the number of nonabsorbing states), and \mathbf{Q} is a matrix of order $(|S| - k) \times (|S| - k)$ corresponding to nonabsorbing states that transition to nonabsorbing states. The matrix $\hat{\mathbf{P}}$ of (25) has the same structure as the matrix $\hat{\mathbf{P}}$ in Theorem 11. It follows from Theorem 11 that the $a^*(t)$ -chain of elite BBO is globally convergent.

These results are similar to the canonical GA [29], which is proven to never converge to the global optimum, but elitist variants of which are proven to converge to the global optimum. We sum up these results in the following corollary.

Corollary 13. *BBO with migration and mutation does not converge to any of the global optima, but elite BBO, which preserves the best individual at each generation, converges to the global optimum.*

Proof. This is an immediate consequence of Theorems 6 and 10 (the nonconvergence of BBO without elitism), Theorem 11 (the convergence of BBO with elitism), and the discussion above. \square

3.5. Convergence Rate. The previous subsection analyzed the convergence properties of elite BBO, and this subsection discusses its convergence rate. The transition matrix of elite BBO after t steps can be found from (25) as follows:

$$\hat{\mathbf{P}}^t = \begin{pmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{N}_t \mathbf{R} & \mathbf{Q}^t \end{pmatrix}, \quad (26)$$

where $\mathbf{N}_t = \mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \dots + \mathbf{Q}^{t-1} = (\mathbf{I} - \mathbf{Q})^{-1}(\mathbf{I} - \mathbf{Q}^t)$. If $\|\mathbf{Q}\| < 1$ the limiting distribution of the Markov chain of BBO can be found from $\hat{\mathbf{P}}^\infty$, which can be written as

$$\hat{\mathbf{P}}^\infty = \lim_{t \rightarrow \infty} \hat{\mathbf{P}}^t = \begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ (\mathbf{I} - \mathbf{Q})^{-1} \mathbf{R} & \mathbf{0} \end{bmatrix}. \quad (27)$$

Modified BBO with elitism has been proven to converge to a global optimum in the previous subsection, and there exists a limiting distribution $\boldsymbol{\pi}^* = \boldsymbol{\pi}(\mathbf{0})\hat{\mathbf{P}}^\infty$, where $\boldsymbol{\pi}(\mathbf{0}) = [\pi_1(\mathbf{0}), \dots,$

$\pi_k(\mathbf{0}), \pi_{k+1}(\mathbf{0}), \dots, \pi_{|S|}(\mathbf{0})] = [\bar{\pi}_1(\mathbf{0}), \bar{\pi}_2(\mathbf{0})]$ (recall that k is the number of global optima). The convergence rate estimate of elite BBO can be obtained as follows.

Theorem 14. *If $\|\mathbf{Q}\| = \rho < 1$, the convergence rate of elite BBO satisfies $\|\boldsymbol{\pi}(\mathbf{t}) - \boldsymbol{\pi}^*\| \leq O(\rho^t)$.*

Proof. Consider

$$\begin{aligned} & \|\boldsymbol{\pi}(\mathbf{t}) - \boldsymbol{\pi}^*\| \\ &= \|\boldsymbol{\pi}(\mathbf{0})\hat{\mathbf{P}}^t - \boldsymbol{\pi}(\mathbf{0})\hat{\mathbf{P}}^\infty\| \\ &= \|\boldsymbol{\pi}(\mathbf{0}) \left[\begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{N}_t \mathbf{R} & \mathbf{Q}^t \end{bmatrix} - \begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ (\mathbf{I} - \mathbf{Q})^{-1} \mathbf{R} & \mathbf{0} \end{bmatrix} \right]\| \\ &= \left\| \begin{bmatrix} \bar{\pi}_1(\mathbf{0}) & \bar{\pi}_2(\mathbf{0}) \end{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ (\mathbf{N}_t - (\mathbf{I} - \mathbf{Q})^{-1}) \mathbf{R} & \mathbf{Q}^t \end{bmatrix} \right\| \\ &= \|\bar{\pi}_2(\mathbf{0}) [(\mathbf{N}_t - (\mathbf{I} - \mathbf{Q})^{-1}) \mathbf{R} \quad \mathbf{Q}^t]\| \\ &\leq \|\bar{\pi}_2(\mathbf{0})\| (\|\mathbf{N}_t - (\mathbf{I} - \mathbf{Q})^{-1}\| \cdot \|\mathbf{R}\| + \|\mathbf{Q}^t\|) \\ &= \|\bar{\pi}_2(\mathbf{0})\| (\|(\mathbf{I} - \mathbf{Q})^{-1}(\mathbf{I} - \mathbf{Q}^t) - (\mathbf{I} - \mathbf{Q})^{-1}\| \\ &\quad \cdot \|\mathbf{R}\| + \|\mathbf{Q}^t\|) \\ &= \|\bar{\pi}_2(\mathbf{0})\| (\|(\mathbf{I} - \mathbf{Q})^{-1} - (\mathbf{I} - \mathbf{Q})^{-1}\mathbf{Q}^t - (\mathbf{I} - \mathbf{Q})^{-1}\| \\ &\quad \cdot \|\mathbf{R}\| + \|\mathbf{Q}^t\|) \\ &\leq \|\bar{\pi}_2(\mathbf{0})\| (\|-(\mathbf{I} - \mathbf{Q})^{-1}\| \cdot \|\mathbf{R}\| + 1) \|\mathbf{Q}^t\| \\ &\leq \|\bar{\pi}_2(\mathbf{0})\| \left(\frac{\|\mathbf{R}\|}{1 - \|\mathbf{Q}\|} + 1 \right) \|\mathbf{Q}\|^t = O(\rho^t). \end{aligned} \quad (28)$$

Note that elite BBO is guaranteed to converge to a global optimum regardless of the initial state. In addition, note that we can improve the convergence rate bound by decreasing the parameter ρ . That is, reducing the number of nonabsorbing states which transition to other nonabsorbing states can accelerate the convergence of elite BBO. In spite of differences between GAs and BBO [12], we see from Theorem 14 that the convergence rate of BBO with elitism is very similar to that of GAs [22, Theorem 5]. \square

4. Simulation Results

Theorem 14 gives the upper bound of the convergence rate estimate of elite BBO. In this section, we use simulation experiments to confirm this theorem. Note that in (28) the parameter ρ is a norm: $\rho = \|\mathbf{Q}\|$. Here we define $\|\cdot\|$ as the infinity norm $\|\cdot\|_\infty$; that is, $\|\mathbf{Q}\|_\infty = \max_i(\sum_{j=1}^n q_{ij})$, where q_{ij} is the element in the i th row and the j th column of matrix \mathbf{Q} . Now note that the transition matrix $\hat{\mathbf{P}}$ in (25) can be obtained from (11) and (23) using elementary matrix transformations. We can thus use Theorem 11 to check for BBO convergence, and we can use Theorem 14 to estimate

TABLE 1: Convergence rate comparisons for the three-bit unimodal one-max problem f_1 .

Mutation rate	Theoretical analysis				Average generation number using simulation	
	Parameter ρ		Convergence time t		BBO	GA
	BBO	GA	BBO	GA		
0.1	0.87	0.85	99.20	90.45	87.58	82.39
0.01	0.68	0.73	35.82	44.16	42.12	50.18
0.001	0.30	0.44	11.47	16.81	11.63	17.37

The table shows the convergence time t in seconds, the corresponding ρ based on (11) and (28), and the generation number of first finding an all-optimal population using BBO and GA, averaged over 25 Monte Carlo simulations.

the convergence rate of BBO. That is, we define $\boldsymbol{\pi}(\mathbf{t}) - \boldsymbol{\pi}^*$ as the error between a BBO population distribution and a distribution that includes at least one optimal solution. We then define the convergence criteria as an arbitrarily small error (e.g., $\|\boldsymbol{\pi}(\mathbf{t}) - \boldsymbol{\pi}^*\| = 10^{-6}$). We can then estimate the time t to convergence from (28) as follows:

$$t \approx \log_{\rho} 10^{-6}. \quad (29)$$

Test functions are limited to three-bit problems with a search space cardinality of eight and a population size of four. The three fitness functions that we examine are

$$\begin{aligned} f_1 &= (1 \ 2 \ 2 \ 3 \ 2 \ 3 \ 3 \ 4), \\ f_2 &= (4 \ 2 \ 2 \ 3 \ 2 \ 3 \ 3 \ 4), \\ f_3 &= (4 \ 1 \ 1 \ 2 \ 1 \ 2 \ 2 \ 3), \end{aligned} \quad (30)$$

where f_1 is a unimodal one-max problem, f_2 is a multimodal problem, and f_3 is a deceptive problem. Note that all three problems are to be maximized. Fitness values are listed in binary order, so the first element of each fitness function corresponds to the bit string 000, the second element corresponds to the bit string 001, and so on. For the BBO parameters, we use a maximum immigration rate and maximum emigration rate of 1, and we use linear migration curves as described in (2). We test elite BBO with three different mutation rates which are applied to each bit in each individual at each generation: 0.1, 0.01, and 0.001. Note that we do not test with a zero-mutation rate because the theory in this paper requires that the mutation rate be positive (see Theorem 5). Convergence is not guaranteed unless the mutation rate is positive.

Numerical calculations show that the transition matrices for these three problems satisfy the convergence conditions of Theorem 11, which indicates that the BBO algorithm converges to one or more of the global optima. As a heuristic test of Theorem 14, we use simulations to record the generation number of first obtaining a population in which all individuals are optimal, and all results are computed from 25 independent runs. Tables 1, 2, and 3 show comparisons of the theoretical convergence time t , the corresponding parameter

TABLE 2: Convergence rate comparison for the three-bit multimodal problem f_2 .

Mutation rate	Theoretical analysis				Average generation number using simulation	
	Parameter ρ		Convergence time t		BBO	GA
	BBO	GA	BBO	GA		
0.1	0.61	0.59	27.95	25.16	26.09	24.87
0.01	0.11	0.17	6.26	7.95	6.17	7.72
0.001	0.04	0.05	4.29	4.76	4.18	4.63

The table shows the convergence time t in seconds, the corresponding ρ based on (11) and (28), and the generation number of first finding an all-optimal population using BBO and GA, averaged over 25 Monte Carlo simulations.

TABLE 3: Convergence rate comparison for the three-bit deceptive problem f_3 .

Mutation rate	Theoretical analysis				Average generation number using simulation	
	Parameter ρ		Convergence time t		BBO	GA
	BBO	GA	BBO	GA		
0.1	0.96	0.95	338.43	270.13	289.55	256.37
0.01	0.48	0.52	18.82	21.49	20.65	22.44
0.001	0.25	0.29	9.97	11.14	9.92	11.08

The table shows the convergence time t in seconds, the corresponding ρ based on (11) and (28), and the generation number of first finding an all-optimal population using BBO and GA, averaged over 25 Monte Carlo simulations.

ρ , and the generation number of first finding an all-optimal population, averaged over 25 independent simulations.

Tables 1–3 show time to convergence and time to finding an optimum, for both BBO and GA. The table confirms the statement following Theorem 14 that the convergence behavior of BBO is similar to that of GA. The tables show that GA converges slightly faster than BBO for high mutation rates, but BBO converges slightly faster for low mutations, and this latter behavior is more important in practice because low mutations rates provide faster convergence.

Several things are notable about the results in Tables 1–3. First, the mutation rate affects the convergence rate of BBO. For all test problems, the convergence rate improves when the mutation rate decreases. We can accelerate the convergence of BBO by decreasing the mutation rate. This may provide practical guidance for BBO tuning for real-world problems. Second, by analyzing the relationship of the parameter ρ and the convergence time t in Tables 1–3, we see that the convergence time t is exponentially related to the parameter ρ , as predicted by Theorem 14. Third, the theoretical results and simulation results match well for most of the test problems, which confirms the convergence rate estimate provided by Theorem 14.

The three-bit problems analyzed above are small, and the results do not tell us how convergence rates scale with the problem dimension. Also, the transition matrix grows faster than exponentially with the problem dimension and the population size [11], so realistically sized problems cannot be directly analyzed with the methods in this paper. However, our methods could be used to study the effect of BBO tuning parameters on small problems, which could provide guidance for larger, real-world problems. Also, similar population distributions could be grouped into the same Markov model state to reduce the transition matrix dimension of large problems to a manageable size [16], which could make the methods in this paper practical for realistically sized problems.

5. Conclusion

In this paper we modeled BBO as a homogeneous finite Markov chain to study convergence, and we obtained new theoretical results for BBO. The analysis revealed that BBO with only migration and mutation does not converge to the global optimum. However, an elite version of BBO that maintains the best solution in the population from one generation to the next converges to a population subset containing at least one globally optimal solution. In other words, BBO with elitism will converge to the global optimum in any binary optimization problem.

In addition, an upper bound for the BBO convergence rate was obtained in Theorem 14. We used simulations to confirm this theorem for a unimodal one-max problem, a multimodal problem, and a deceptive problem. The results in this paper are similar to those of the canonical GA [29], and so our results are not surprising, but this paper represents the first time that such results have been formalized for BBO.

The results in this paper are limited to binary problems but can easily be extended to discrete problems with any finite alphabet. This is due to the simple fact that any discrete problem can be reduced to a binary problem.

For future work there are several important directions. First, it is of interest to study how to improve BBO convergence time and robustness based on these results. The second important direction for future work is to study the asymptotic convergence of variations of BBO, including partial emigration-based BBO, total immigration-based BBO, and total emigration-based BBO [12]. The theorems in this paper provide the foundation to study these variations, so we do not need additional theoretical tools to analyze their convergence. The third important direction for future work is to develop hybrid BBO, which combines BBO with other EAs, and study their convergence behaviors using the theory presented here. Finally, it would be of interest to extend these results to continuous optimization problems, which are the types of problems to which real-world implementations of BBO are typically applied.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This material was supported in part by the National Science Foundation under Grant no. 0826124, the National Natural Science Foundation of China under Grant nos. 61305078 and 61179041, and the Shaoxing City Public Technology Applied Research Project under Grant no. 2013B70004. The authors are grateful to the reviewers for suggesting improvements to the original version of this paper.

References

- [1] D. Simon, "Biogeography-based optimization," *IEEE Transactions on Evolutionary Computation*, vol. 12, no. 6, pp. 702–713, 2008.
- [2] D. Du, D. Simon, and M. Ergezer, "Biogeography-based optimization combined with evolutionary strategy and immigration refusal," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC '09)*, pp. 1023–1002, San Antonio, Tex, USA, October 2009.
- [3] M. Ergezer, D. Simon, and D. Du, "Oppositional biogeography-based optimization," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC '09)*, pp. 1035–1040, San Antonio, Tex, USA, October 2009.
- [4] M. Lundy and A. Mees, "Convergence of an annealing algorithm," *Mathematical Programming*, vol. 34, no. 1, pp. 111–124, 1986.
- [5] H. Ma and D. Simon, "Blended biogeography-based optimization for constrained optimization," *Engineering Applications of Artificial Intelligence*, vol. 24, no. 3, pp. 517–525, 2011.
- [6] A. Bhattacharya and P. K. Chattopadhyay, "Biogeography-based optimization for different economic load dispatch problems," *IEEE Transactions on Power Systems*, vol. 25, no. 2, pp. 1064–1077, 2010.
- [7] P. Lozovyy, G. Thomas, and D. Simon, "Biogeography-based optimization for robot controller tuning," in *Computational Modeling and Simulation of Intellect: Current State and Future Perspectives*, B. Igel'nik, Ed., pp. 162–181, IGI Global, Hershey, Pa, USA, 2011.
- [8] V. Panchal, P. Singh, N. Kaur, and H. Kundra, "Biogeography based satellite image classification," *International Journal of Computer Science and Information Security*, vol. 6, no. 2, pp. 269–274, 2009.
- [9] R. Rarick, D. Simon, F. E. Villaseca, and B. Vyakaranam, "Biogeography-based optimization and the solution of the power flow problem," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC '09)*, pp. 1029–1034, usa, October 2009.
- [10] D. Simon, M. Ergezer, and D. Du, "Population distributions in biogeography-based optimization algorithms with elitism," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC '09)*, pp. 1017–1022, San Antonio, Tex, USA, October 2009.
- [11] D. Simon, M. Ergezer, D. Du, and R. Rarick, "Markov models for biogeography-based optimization," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 41, no. 1, pp. 299–306, 2011.
- [12] D. Simon, R. Rarick, M. Ergezer, and D. Du, "Analytical and numerical comparisons of biogeography-based optimization and genetic algorithms," *Information Sciences*, vol. 181, no. 7, pp. 1224–1248, 2011.

- [13] 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 (Dortmund, 1990)*, vol. 496 of *Lecture Notes in Computer Science*, pp. 4–12, Springer, New York, NY, USA, 1991.
- [14] C. Grinstead and J. Snell, *Introduction to Probability*, American Mathematical Society, Providence, RI, USA, 1997.
- [15] G. Guo and S. Yu, "The unified method analyzing convergence of genetic algorithms," *Control Theory & Applications*, vol. 18, no. 3, pp. 443–446, 2001.
- [16] C. Reeves and J. Rowe, *Genetic Algorithms: Principles and Perspectives*, vol. 20, Kluwer Academic Publishers, Norwell, Mass, USA, 2003.
- [17] L. Wang, *Intelligent Optimization Algorithms with Applications*, Springer, New York, NY, USA, 2001.
- [18] A. Auger and B. Doerr, *Theory of Randomized Search Heuristics*, World Scientific, Singapore, 2011.
- [19] T. Jansen, *Analyzing Evolutionary Algorithms*, Springer, New York, NY, USA, 2013.
- [20] F. Neumann and C. Witt, *Bioinspired Computation in Combinatorial Optimization*, Springer, New York, NY, USA, 2010.
- [21] T. Davis and J. Principe, "A Markov chain framework for the simple genetic algorithm," *Evolutionary Computation*, vol. 1, no. 3, pp. 269–288, 1993.
- [22] J. He and L. Kang, "On the convergence rates of genetic algorithms," *Theoretical Computer Science*, vol. 229, no. 1-2, pp. 23–39, 1999.
- [23] J. Horn, "Finite Markov chain analysis of genetic algorithms with niching," in *Proceedings of the International Conference on Genetic Algorithms*, pp. 1–13, San Francisco, Calif, USA, 1993.
- [24] A. Nix and M. Vose, "Modeling genetic algorithms with Markov chains," *Annals of Mathematics and Artificial Intelligence*, vol. 5, no. 1, pp. 79–88, 1992.
- [25] J. Suzuki, "Markov chain analysis on simple genetic algorithms," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 25, no. 4, pp. 655–659, 1995.
- [26] J. Suzuki, "A further result on the Markov chain model of genetic algorithms and its application to a simulated annealing-like strategy," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 28, no. 1, pp. 95–102, 1998.
- [27] X. Yao, Y. Liu, and G. Lin, "Evolutionary programming made faster," *IEEE Transactions on Evolutionary Computation*, vol. 3, no. 2, pp. 82–102, 1999.
- [28] M. Iosifescu, *Finite Markov Processes and their Applications*, John Wiley & Sons, New York, NY, USA, 1980.
- [29] G. Rudolph, "Convergence analysis of canonical genetic algorithms," *IEEE Transactions on Neural Networks*, vol. 5, no. 1, pp. 96–101, 1994.



Hindawi

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

