Analysis of Convergence Rate for Biogeography Based Optimization

Aijun Zhu[1,2,3], Chuanpei Xu[1,3], Zhi Li[1,3,4], Cong Hu[1], Jun Wu[1], Zhenbing Liu[1]

[1]School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin
[2]Department of Computer Science and Engineering, Texas A&M University, College Station, TX, USA
[3]Guangxi Key Laboratory of Automatic Detecting Technology and Instruments, Guilin
[4]Guilin University of Aerospace Technology

Abstract With the dramatic improvement of computation ability in modern computers, a lot of new meta-heuristic methods appear. All these meta-heuristic algorithms are originated from some mechanisms in nature, which are alike in composition and extensively adopted to solve combination optimization problems. However, meta-heuristic algorithm, such as BBO, is lack of strict theory foundation and hard to be analyzed in theory, because it comes from heuristic idea and has complicated random behavior. Therefore, in this paper, we propose some theoretical analysis to lay solid foundation for BBO application. To obtain the average first hitting time (AFHT) of BBO, firstly the relationship between AFHT and convergence rate is discussed, and then the upper bound and the lower bound of convergence rate are put forward. When BBO is adopted to solve an optimization problem with problem size n and population size n, if migration operation and Mutation-Constant or Mutation-1/n operation are adopted, then the lower bound of AFHT is proposed.

Keywords Meta-Heuristic, Global Optimization, NP Hard Problem.

Introduction

Since 1960s, a lot of researchers are interested in evolutionary computation with the development of computer technology. Some famous algorithms are proposed, such as Genetic Algorithm, Evolutionary Programming, and Evolutionary Strategy. Since the computation ability of computer improves dramatically, a lot of new meta-heuristic methods arise, such as Ant Colony Optimization (ACO) [1], Particle swarm Optimization (PSO) [2], Differential Evolution (DE) [3], Grey Wolf Optimization (GWO) [4], Biogeography Based Optimization (BBO) [5], Hybrid Grey Wolf Optimization (HGWO) [6] ,et al. All those algorithms are originated from some mechanisms in nature, and are similar in structure. We call such algorithms as evolutionary algorithms. In general, evolutionary algorithms are random and heuristic optimization methods, which are widely used to solve combination optimization problems. As we use evolutionary algorithms to solve optimization problems, it is easy to add some heuristic ideas. Therefore, evolutionary algorithms demonstrate excellent performance in a lot of fields such as data mining, engineering optimization an industry design.

However, evolutionary algorithm, such as BBO, is lack of strict theory foundation and hard to be analyzed in theory [7], because it comes from heuristic idea and has complicated random behavior. In this paper, we propose some theoretical analysis to lay solid foundation for BBO application. To obtain the average first hitting time (AFHT) of BBO, firstly the relationship between AFHT and convergence rate is discussed, and then the upper bound and the lower bound of convergence rate are put forward. When BBO is adopted to solve an optimization problem with problem size n and population size n, if migration operation and Mutation-Constant or Mutation-1/n operation are adopted, then the lower bound of AFHT is proposed.

Organization of the rest of this paper is as follows. The basic of BBO is briefly introduced in Section 2. Analysis of Convergence rate is presented in Section 3. Finally, concluding remarks is given in Section 4.

Basic of BBO

Practical work about biogeography started their research in 1960s by Alfred Wallance [8] and Charles Darwin [9]. In 2008, Dan Simon [5] proposed a new optimization method based on biogeography called BBO, which are widely applied in Engineering Optimization [10-15].

The mathematical model of biogeography demonstrates the process that species migrate from one island to another. Islands or habitats that are well suited for species obtain high Habitat Suitability Index (HSI).

In BBO methodology, every individual has its own immigration rate Im and emigration rate Em. The immigration rate and emigration rate are both the function of number of species in the habitat. The immigration rate and the emigration rate are calculated as follows:

$$Em_k = E.k / n,$$ (1)

$$Im_k = I.(1 - k / n),$$ (2)

where k is the number of species in the habitat and n is the largest number of species contained in the habitat; E is the largest emigration rate when there are zero species in the habitat; I is the largest immigration rate when there are the largest number of species which can be contained in the habitat; Suppose all the habitats contain the same largest number of species.

Suppose the probability that a habitat has exactly k species is Pr_k, and Pr_k changes from time t to time $t + \Delta$ as follows [5] :

$$Pr_k(t + \Delta t) = Pr_k(t).(1 - Im_k.\Delta t - Em_k.\Delta t) + Pr_{k-1}(t).Im_{k-1}.\Delta t + Pr_{k+1}(t).Em_{k+1}.\Delta t,$$ (3)

where Im_k is immigration rate when the current habitat has k species; Im_{k-1} is immigration rate when the current habitat has k-1 species; Em_k is emigration rate when the current habitat has k species; Em_{k-1} is emigration rate when the current habitat has k-1 species; Pr_{k+1} is the probability that the current habitat has exactly k+1 species; Pr_{k-1} is the probability that the current habitat has exactly k-1 species.

Suppose Δt is small enough such that we can ignore the probability of more than one emigration and immigration. Calculating the limit of equation (3) as $\Delta t \rightarrow 0$ obtain equation (4):

$$Pr_k = \begin{cases} -(Im_k + Em_k).Pr_k + Em_{k+1}. Pr_{k+1}, & k=0 \\ -(Im_k + Em_k).Pr_k + Im_{k-1}. Pr_{k-1}, + Em_{k+1}. Pr_{k+1}, & k\varepsilon[1, S_{max}-1] \\ -(Im_k + Em_k).Pr_k + Im_{k-1}. Pr_{k-1}, & k=S_{max} \end{cases}$$ (4)

where S_{max} is the maximum number of species in a single habitat, the rest are the same as equation (3). If a given solution S with k species has a low probability Pr_k, it is amazing that it is regarded as a solution. The mutation rate m_k is as follows.

$$m_k = m_{max}.(1 - Pr_k) / Pr_{max}$$ (5)

Where m_{max} is the maximum number of mutation rate.

Analysis of Convergence rate

We name a population an optimal population, which contains one optimal individual at least. There may be more than one optimal population. An objective subspace $B_0^*(\subset \Omega)$ is consists of all the optimal populations. Ω is the population space. The aim of BBO is to find a population which is in $B_0^*(\subset \Omega)$.

Definition 1 The optimal value is defined as $S(X) = \min\{\cos t(x_i), i = 1, 2, ..., NP\}$. An objective subspace

$$B_0^* = \{X \mid s(X) = \min\{\cos t(x), x \in S\}\},$$ where S is the solution space.

Definition 2 Given a Markov population series $\{\xi_t, t = 0, 1, 2, ...\}(\xi_t \in \Omega)$ and An objective subspace $B_0^* \subset \Omega$, where Ω is the population space. We define u_t as the probability of ξ_t being in the objective subspace B_0^*, i.e.,

$$u_t = \sum_{y \in B_0^*} P(\xi_t = y), \quad t = 0, 1, 2, ...$$ (6)

Definition 3 Given a Markov population series $\{\xi_t, t = 0,1,2,...\}(\xi_t \in \Omega)$ and an objective subspace $B_0^* \subset \Omega$, where Ω is the population space[20]. The Markov population series could converge to objective subspace $B_0^* \subset \Omega$, if

$$\lim_{t \to \infty} u_t = 1. \tag{7}$$

Definition 4 Given a Markov population series $\{\xi_t, t = 0,1,2,...\}(\xi_t \in \Omega)$ and an objective subspace $B_0^* \subset \Omega$, where Ω is the population space [7]. We define p_t as the convergence rate to $B_0^* \subset \Omega$ at time t, i.e.,

$$p_t = 1 - \sum_{y \in B_0^*} P(\xi_t = y), \quad t = 0,1,2,... \tag{8}$$

From the Definition 4, we could get a new relationship as follows:

$$p_t = 1 - u_t, \quad t = 0,1,2,...$$

Suppose there is a random variable ρ, which denotes the following events:

$$\rho = 0 : X_0 \in B_0^*;$$

$$\rho = 1 : X_1 \in B_0^* \wedge X_j \notin B_0^* (j = 0);$$

$$\rho = 2 : X_2 \in B_0^* \wedge X_j \notin B_0^* (\forall j \in \{0,1\});$$

$$\rho = 3 : X_3 \in B_0^* \wedge X_j \notin B_0^* (\forall j \in \{0,1,2\});$$

$$...$$

$$\rho = t-1 : X_{t-1} \in B_0^* \wedge X_j \notin B_0^* (\forall j \in \{0,1,...,t-2\});$$

$$\rho = t : X_t \in B_0^* \wedge X_j \notin B_0^* (\forall j \in \{0,1,...,t-1\});$$

The random variable ρ indicates the first hitting time, and $E[\rho]$ denotes the average first hitting time (AFHT) or expected first hitting time (EFHT).

Definition 5 The distribution function of the random variable ρ could be defined as

$$D_\rho(t) = P(\rho \le t) = P(\rho = t) + P(\rho < t). \tag{9}$$

Because the Markov chain in BBO is an absorbing Markov chain, we can obtain

$$u_t - u_{t-1} = \sum_{y \in B_0^*} P(X_t = y) - \sum_{y \in B_0^*} P(X_{t-1} = y)$$

$$= P(\rho = t). \tag{10}$$

From above equation, we can get

$$u_t = P(\rho = t) + u_{t-1} \tag{11}$$

By comparing Eq.() and Eq.(), we can obtain

$$u_t = D_\rho(t) \tag{12}$$

According to definition 4，we can obtain

$$D_\rho(t) = u_t = 1 - p_t \tag{13}$$

The following Lemma 3 is to uncover the relationship [7] between AFHT and convergence rate.

Lemma 1 Given two random variable α and β, where $\mathrm{E}[\alpha] < \infty$ and $\mathrm{E}[\beta] < \infty$, the distribution functions of α and β at time t are $D_\alpha(t)$ and $D_\beta(t)$, respectively, i.e.,

$$D_\alpha(t) = P(\alpha \le t) = \sum_{j=0}^{t} P(\alpha = j),$$

$$D_\beta(t) = P(\beta \le t) = \sum_{j=0}^{t} P(\beta = j).$$

if $\forall t \in \{0,1,2,3...\}$, $D_\alpha(t) \ge D_\beta(t)$, then

$$E[\alpha] \le E[\beta], \tag{14}$$

Where $E[\alpha] = \sum_{t=0}^{\infty} t.P(\alpha = t) < \infty$; $E[\beta] = \sum_{t=0}^{\infty} t.P(\beta = t) < \infty$.

Proof. According to the definition of expected value of random variable α, which is $E[\alpha] = \sum_{t=0}^{\infty} t.P(\alpha = t)$,

we can get

$$E[\alpha] = \sum_{t=0}^{\infty} t.P(\alpha = t)$$

$$= 0.P(\alpha = 0) + \sum_{t=1}^{\infty} t.P(\alpha = t)$$

$$= 0.\sum_{t=0}^{0} P(\alpha = 0) + \sum_{t=1}^{\infty} t.(P(\alpha \le t) - P(\alpha \le t-1))$$

$$= 0.\sum_{t=0}^{0} P(\alpha = 0) + \sum_{t=1}^{\infty} t.(\sum_{j=0}^{t} P(\alpha = j) - \sum_{j=0}^{t-1} P(\alpha = j))$$

$$= 0.D_\alpha(0) + \sum_{t=1}^{\infty} t.(D_\alpha(t) - D_\alpha(t-1))$$

$$= \sum_{t=1}^{\infty} t.(D_\alpha(t) - D_\alpha(t-1))$$

$$= \sum_{j=1}^{+\infty} \sum_{t=j}^{+\infty} (D_\alpha(t) - D_\alpha(t-1))$$

$$= \sum_{j=0}^{+\infty} \lim_{t \to +\infty} (D_\alpha(t) - D_\alpha(j))$$

$$= \sum_{j=0}^{+\infty} (1 - D_\alpha(j)).$$

In the above equations, the second and the third equations is according to the definition of distribution function.

The last equation is due to $\lim_{t \to +\infty} D_\alpha(t) = \lim_{t \to +\infty} \sum_{j=0}^{t} P(\alpha = j) = 1$.

By the same way, $E[\beta] = \sum_{j=0}^{+\infty} (1 - D_\beta(j))$.

To compare $E[\alpha]$ and $E[\beta]$, we subtract $E[\beta]$ with $E[\alpha]$.

$$E[\alpha] - E[\beta] = \sum_{j=0}^{+\infty} (1 - D_\alpha(j)) - \sum_{j=0}^{+\infty} (1 - D_\beta(j))$$

$$= \sum_{j=0}^{+\infty} (D_\beta(j) - D_\alpha(j))$$

$$\le 0$$

∎

The following Lemma 2 is to get the upper bound and the lower bound of convergence rate, which is applied in

Journal of Scientific and Engineering Research

discrete space[16] .

Lemma 2 Given an absorbing Markov chain $\{X_t, t = 0,1,2,...\}(X_t \in \Omega)$ and an objective subspace $B_0^* \subset \Omega$, where Ω is the population space. If there are $\{\phi_t, t = 0,1,2,...\}$ and $\{\varphi_t, t = 0,1,2,...\}$, which can satisfy the following conditions:

(1) $\displaystyle\prod_{t=0}^{+\infty}(1-\phi_t) = 0$

(2) $\displaystyle\phi_t \le \sum_{y \notin B_0^*} P(X_{t+1} \in B_0^* \mid X_t = y)\frac{P(X_t = y)}{p_t} \le \varphi_t$

Then the absorbing Markov chain can reach the objective subspace $B_0^*(\subset \Omega)$ with a convergence rate p_t which is as following

$$p_0\prod_{j=0}^{t-1}(1-\varphi_j) \le p_t \le p_0\prod_{j=0}^{t-1}(1-\phi_j) \tag{15}$$

Proof. According to the definition of p_t we can obtain

$$
\begin{aligned}
p_t - p_{t-1} &= (1-\mu_t)-(1-\mu_{t-1})\\
&= -(\mu_t - \mu_{t-1})\\
&= -(\sum_{y \in B_0^*} P(X_t = y) - \sum_{y \in B_0^*} P(X_{t-1} = y))\\
&= -P(\rho = t)\\
&= -\sum_{y \notin B_0^*} P(X_t \in B_0^* \mid X_{t-1} = y)P(X_{t-1} = y)
\end{aligned}
$$

The above equation is divided by p_{t-1}, then we get

$$\frac{p_t - p_{t-1}}{p_{t-1}} = -\sum_{y \notin B_0^*} P(X_t \in B_0^* \mid X_{t-1} = y)\frac{P(X_{t-1} = y)}{p_{t-1}}$$

Because of the condition (2), we obtain

$$
\begin{aligned}
&-\varphi_{t-1} \le \frac{p_t - p_{t-1}}{p_{t-1}} \le -\phi_{t-1}\\
\Rightarrow &-p_{t-1}\varphi_{t-1} \le p_t - p_{t-1} \le -\phi_{t-1}p_{t-1}\\
\Rightarrow &-p_{t-1}\varphi_{t-1} + p_{t-1} \le p_t \le -\phi_{t-1}p_{t-1} + p_{t-1}\\
\Rightarrow &-(1-\mu_{t-1})\varphi_{t-1} + (1-\mu_{t-1}) \le 1-\mu_t \le -\phi_{t-1}(1-\mu_{t-1}) + (1-\mu_{t-1})\\
\Rightarrow &(1-\mu_{t-1})(1-\varphi_{t-1}) \le 1-\mu_t \le (1-\mu_{t-1})(1-\phi_{t-1})\\
\Rightarrow &p_{t-1}(1-\varphi_{t-1}) \le p_t \le p_{t-1}(1-\phi_{t-1})
\end{aligned}
$$

The above inequality is unfolded with recursion, and we get

$$p_0\prod_{j=0}^{t-1}(1-\varphi_j) \le p_t \le p_0\prod_{j=0}^{t-1}(1-\phi_j).$$

∎

Through the above two lemmas, the relationship between AFHT and convergence rate is obtained. the upper and low bound of the convergence rate is also got. Based on the two lemmas [7], the following theorem gives boundary of AFHT.

Theorem 1 Given an absorbing Markov chain $\{X_t, t = 0,1,2,...\}(X_t \in \Omega)$ and an objective subspace $B_0^* \subset \Omega$, where Ω is the population space. If there are $\{\phi_t, t = 0,1,2,...\}$ and $\{\varphi_t, t = 0,1,2,...\}$, which can

Journal of Scientific and Engineering Research

satisfy the following conditions:

(1) $\prod_{t=0}^{+\infty}(1-\phi_t) = 0$

(2) $\phi_t \leq \sum_{y \notin B_0^*} P(X_{t+1} \in B_0^* \mid X_t = y)\dfrac{P(X_t = y)}{p_t} \leq \varphi_t$

Then the absorbing Markov chain can reach the objective subspace B_0^* with the average first hitting time AFHT, which is as following

$$p_0(\varphi_0 + \sum_{t=2}^{+\infty} t\varphi_{t-1}\prod_{j=0}^{t-2}(1-\varphi_j)) \leq \mathrm{E}[\rho] \leq p_0(\phi_0 + \sum_{t=2}^{+\infty} t\phi_{t-1}\prod_{j=0}^{t-2}(1-\phi_j)) \qquad (16)$$

Proof. According to the lemma 2, we get

$$p_t \leq p_0\prod_{j=0}^{t-1}(1-\phi_j)$$

Due to $p_t = 1 - \mu_t$, let t=0, we get $p_0 = 1 - \mu_0$.

As the μ_t is essentially the distribution of random variable ρ, which is $D_\rho(t) = \mu_t$. Then the lower bound can be described using μ_t

$$D_\rho(t) \geq \begin{cases} 1 - p_0 & t = 0 \\ 1 - p_0\prod_{j=0}^{t-1}(1-\phi_j) & t \neq 0 \end{cases}$$

Suppose there is a flabby variable ξ, which is exactly the lower bound of D_ρ. The expectation of ξ can be calculated as following.

$$\mathrm{E}[\xi] = \sum_{t=0}^{\infty} t.P(\xi = t) = 0.(1 - p_0) + 1.(p_0 - p_0(1-\phi_0))$$

$$+ \sum_{t=2}^{+\infty} t.(p_0\prod_{j=0}^{t-2}(1-\phi_j) - p_0\prod_{j=0}^{t-1}(1-\phi_j))$$

$$= \phi_0 p_0 + \sum_{t=2}^{+\infty} t.\phi_{t-1} p_0\prod_{j=0}^{t-2}(1-\phi_j)$$

$$= p_0(\phi_0 + \sum_{t=2}^{+\infty} t.\phi_{t-1}\prod_{j=0}^{t-2}(1-\phi_j))$$

Due to $D_\rho(t) \geq D_\xi(t)$ and lemma 1, we get

$$\mathrm{E}[\rho] \leq \mathrm{E}[\xi].$$

Then we get the upper bound of AFHT

$$\mathrm{E}[\rho] \leq p_0(\phi_0 + \sum_{t=2}^{+\infty} t.\phi_{t-1}\prod_{j=0}^{t-2}(1-\phi_j)).$$

There is a special case when the initial population does not contain the optimal solution. In such case, $p_0 = 1$. As the initial population in most evolutionary algorithms do not contain the optimal solution, the above equation can be simplified with $p_0 = 1$.

Similarly, we can get the lower bound of AFHT.

∎

In general, there are several mutation operation in evolutionary algorithms, such as Mutation-Constant，Mutation-1/n and Mutation-random-1.

Definition 6 if every bit in the solution can mutate with constant probability, such mutation is called Mutation-Constant. (if the candidate solution is binary, then every bit can mutate with probability p_c ; if the candidate solution is coded with d-nary, then every gene can mutate with probability $\dfrac{p_c}{(d-1)}$)

Definition 7 if every bit in the solution can mutate with probability 1/n, such mutation is called **Mutation-**1/n**. (** if the candidate solution is binary, then every bit can mutate with probability 1/n; if the candidate solution is coded with d-nary, then every gene can mutate with probability $\dfrac{1}{n(d-1)}$)

Definition 8 if a gene in the solution is selected randomly, and would mutate, such mutation is called Mutation-random-1. (if the candidate solution is binary, then every bit can mutate with probability 1/n; if the candidate solution is coded with d-nary, then every gene can mutate with probability $\dfrac{1}{n(d-1)}$)

If the population size is one, then the population space equals the solution space; If a candidate solution is coded with d-nary, both the population space and the solution space are d^n .

If the population size is greater than one, then the population space does not equal the solution space; If a candidate solution is coded with d-nary the solution space are d^n . While the population space is associated with the definition of population, such as unordered population and ordered population.

Definition 9 if a candidate solution is denoted as (.)，while a population is denoted as {.}. if two populations with the same solution but different arranged order are equivalent, such population is called unordered population.

Definition 10 if a candidate solution is denoted as (.)，while a population is denoted as {.}. if two populations with the same solution but different arranged order are not equivalent, such population is called ordered population.

To distinguish unordered population with ordered population, the following example is shown. If the candidate solution is coded by binary, then suppose the problem size is three and the population size is 4. there are two population: $POP_1 = \{(000),(111),(101),(001)\}$, $POP_2 = \{(000),(111),(001),(101)\}$. If POP_1 and POP_2 are unordered populations, then they are equivalent. Otherwise, they are not equivalent.

A population space with problem size n and population size PopSize is provided with $\begin{pmatrix} d^n + \text{PopSize} - 1 \\ \text{PopSize} \end{pmatrix}$ different populations in the case of unordered population. Otherwise, it has $d^{n \times \text{PopSize}}$ different populations in the case of ordered population.

Proposition 1 if BBO is adopted to solve a optimization problem with problem size n and population size n and there is no optimal solution in the initial population, then the following lower bound is obtained in the case of migration operation and mutation operation independently

(1) if migration operation and Mutation-Constant operation are adopted, then the lower bound of AFHT is as follows:

$$\text{E}[\rho] = O(\max(\frac{1}{n} \times \frac{d-1}{p_c(1-p_c)^{n-1}}, \frac{1}{n} \times \frac{1}{\zeta(k)} \times \frac{1}{(1-\zeta(k))^{n-1}})) \,, \tag{17}$$

Where p_c is a constant; n is dimension number of a candidate solution; d is the code number of every gene of a candidate solution (d is an integer and $d \geq 2$); $\zeta(k)$ is the normalized immigration rate; where k is the number of species in the habitat .

(2) if migration operation and Mutation-1/n operation are adopted, then the lower bound of AFHT is as follows:

$$\text{E}[\rho] = O(\frac{d^n}{n^2}) \,, \tag{18}$$

Where n is dimension number of a candidate solution; d is the code number of every gene of a candidate solution (d is an integer and $d \geq 2$).

Proof. Because of the independent operation of migration and mutation, suppose X_{t+1}^{Q} is the population after migration operation and X_{t+1}^{M} is the population after migration mutation operation. The following is obtained:

$$\sum_{y \notin B_0^*} P(X_{t+1} \in B_0^* \mid X_t = y) \frac{P(X_t = y)}{p_t}$$

$$= \sum_{y \notin B_0^*} P(X_{t+1}^{Q} \in B_0^* \mid X_t = y) \frac{P(X_t = y)}{p_t}$$

$$+ \sum_{y \notin B_0^*} P(X_{t+1}^{M} \in B_0^* \mid X_t = y) \frac{P(X_t = y)}{p_t}$$

According to the features of mutation operation, we have

$$\sum_{y \notin B_0^*} P(X_{t+1}^{M} \in B_0^* \mid X_t = y) \frac{P(X_t = y)}{p_t} \leq \begin{cases} n \times \dfrac{p_c(1-p_c)^{n-1}}{d-1}, \text{Mutation-Constant} \\[2em] n^2 \left(1 - \left(\dfrac{1}{e}\right)\right)\left(\dfrac{1}{d^n}\right), \text{Mutation-1/n} \end{cases}$$

As to the migration operation, the analysis is as follows. Compared to the optimal solution, the candidate solution is provided with only one gene which is not the same. In such case, only one gene is needed to immigrate and other genes keep the same. Then, the maximum probability for migration operation to get the optimal solution is as follows

$$P(X_{t+1}^{Q} \in B_0^* \mid X_t = y) \leq 1 - \left(1 - \lambda_k'(1-\lambda_k')^{n-1}\right)^n ,$$

where $\lambda_k' = \lambda_l + (\lambda_u - \lambda_l)(\lambda_k - \lambda_{min})/(\lambda_{max} - \lambda_{min})$ and $\lambda_l = 0, \lambda_u = 1$; λ_{max} and λ_{min} are constants; $\lambda_k = I(1 - k/n\max)$, $I = 1$ and $n\max$ the largest number of species contained in the habitat; k is the number of species in the habitat. After that, $\lambda_k' = (1 - k/n\max - \lambda_{min})/(\lambda_{max} - \lambda_{min})$. Because of $n\max$, λ_{max} and λ_{min} are all constants, then λ_k' is denoted by a function $\lambda_k' = \zeta(k)$:

$$\sum_{y \notin B_0^*} P(X_{t+1}^{Q} \in B_0^* \mid X_t = y) \frac{P(X_t = y)}{p_t}$$

$$\leq \sum_{y \notin B_0^*} \left(1 - \left(1 - \zeta(k)(1-\zeta(k))^{n-1}\right)^n\right) \frac{P(X_t = y)}{p_t}$$

$$= \left(1 - \left(1 - \zeta(k)(1-\zeta(k))^{n-1}\right)^n\right) \frac{\sum_{y \notin B_0^*} P(X_t = y)}{p_t}$$

$$\left(1 - \left(1 - \zeta(k)(1-\zeta(k))^{n-1}\right)^n\right)$$

$$\approx n \times \zeta(k)(1-\zeta(k))^{n-1}$$

Therefore, in the case of Mutation-Constant we have

$$\sum_{y \notin B_0^*} P(X_{t+1} \in B_0^* \mid X_t = y) \frac{P(X_t = y)}{p_t}$$

$$\leq n \times \frac{p_c(1-p_c)^{n-1}}{d-1} + n \times \zeta(k)(1-\zeta(k))^{n-1}$$

Suppose $\varphi_t = n \times \dfrac{p_c(1-p_c)^{n-1}}{d-1} + n \times \zeta(k)(1-\zeta(k))^{n-1}$, according to theorem 1, the lower bound of AFHT is as follows

$$E[\rho] = \left(n \times \frac{p_c(1-p_c)^{n-1}}{d-1} + n \times \zeta(k)(1-\zeta(k))^{n-1} \right)^{-1}$$

$$= O(\max(\frac{1}{n} \times \frac{d-1}{p_c(1-p_c)^{n-1}}, \frac{1}{n} \times \frac{1}{\zeta(k)} \times \frac{1}{(1-\zeta(k))^{n-1}})).$$

While in the case of Mutation-Constant we have

$$\sum_{y \notin B_0^*} P(\boldsymbol{X}_{t+1} \in B_0^* \mid \boldsymbol{X}_t = y) \frac{P(\boldsymbol{X}_t = y)}{p_t}$$

$$\leq n^2 \left(1 - \left(\frac{1}{e} \right) \right) \left(\frac{1}{d^n} \right) + n \times \zeta(k)(1-\zeta(k))^{n-1}.$$

Suppose $\varphi_t = n^2 \left(1 - \left(\dfrac{1}{e} \right) \right) \left(\dfrac{1}{d^n} \right) + n \times \zeta(k)(1-\zeta(k))^{n-1}$, according to theorem 1, the lower bound of AFHT is as follows

$$E[\rho] = \left(n^2 \left(1 - \left(\frac{1}{e} \right) \right) \left(\frac{1}{d^n} \right) + n \times \zeta(k)(1-\zeta(k))^{n-1} \right)^{-1}$$

$$= O(\frac{d^n}{n^2}).$$

∎

Conclusions
In this paper, we propose some theoretical analysis to lay solid foundation for BBO application. To obtain the average first hitting time (AFHT) of BBO, firstly the relationship between AFHT and convergence rate is discussed, and then the upper bound and the lower bound of convergence rate are put forward. When BBO is adopted to solve an optimization problem with problem size n and population size n, if migration operation and Mutation-Constant or Mutation-1/n operation are adopted, then the lower bound of AFHT is proposed.

References
[1].   M. Dorigo, M. Birattari, T. Stutzle. Ant colony optimization, *Comput Intell Magaz*, 2006, 1: 28-39
[2].   J. Kennedy, R. Eberhart, Particle swarm optimization. *IEEE International Conference on Neural Networks*, 1995: 1942-1948.
[3].   R. Storn, K. Price. Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization,* 1997, 11(4): 341-359.
[4].   .S. Mirjalili, S. M. Mirjalili, A. Lewis. Grey Wolf Optimizer. *Advances in Engineering Software*, 2014, 69(3): 46-61.
[5].   Simon D. Biogeography-based optimization. IEEE Trans Evol Comput, 2008, 12(6):702–713.

[6].   Aijun Zhu, Chuanpei Xu, Zhi Li, *etal*. Hybridizing Grey Wolf Optimization with Differential Evolution for global optimization and test scheduling for 3D stacked SoC . *Journal of Systems Engineering and Electronics,* 2015, 26(2): 317-328.

[7].   Yu Y. Theoretical analysis of evolutionary computation and learning algorithm. *Nanjin university doctor degree* thesis, 2011: 1–50

[8].   Wallace A. The Geographical Distribution of Animals (Two Volumes)．Boston, MA:. Adamant Media Corporation, 2005:1-26

[9].   Darwin C. The Origin of Species [M]. New York: Gramercy, 1995:12-36.

[10]. LI X T, YIN M H. Hybrid Differential Evolution with Biogeography-Based Optimization for Design of a Reconfigurable Antenna Array with Discrete Phase Shifters . International Journal of Antennas and Propagation, 2011: 1-12.

[11]. WANG L, XU Y. An effective hybrid biogeography-based optimization algorithm for parameter estimation of chaotic systems. Expert Systems with Applications, 38 (12): 15103–15109.

[12]. LI X T, WANG J Y, ZHOU J P, et al. A perturb biogeography based optimization with mutation for global numerical optimization. Applied Mathematics and Computation, 218 (2): 598–609

[13]. Aijun Zhu. Zhi Li. Wangchun Zhu, *etal*. Design of Test wrapper scan chain based on Differential Evolution. Journal of Engineering Science and Technology Review, 2013, 6(2): 10-14.

[14]. Aijun Zhu, Zhi Li, Chuanpei Xu, *etal*. A new wrapper scan chain balance algorithm for Intellectual Property module in SoC. Open Electrical and Electronic Engineering Journal, 2014, 8(1): 42-49.

[15]. Aijun Zhu, Chuanpei Xu, Zhi Li, *etal*. Three-dimension Test wrapper design based on Multi-objective Cuckoo Search. Open Cybernetics and Systemics Journal, 2014, 8(1):104-110.

[16]. He J, Yao X. Drift analysis and average time complexity of evolutionary algorithms [J].Artificial Intelligence, 2001, 127(1):57–85.