

Theoretical Analysis of Biogeography Based Optimization

Aijun ZHU^{1,2,3 a}, Cong HU^{1,3}, Chuanpei XU^{1,3}, Zhi Li^{1,3}

¹School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin 541004, P.R.China

²Department of Computer Science and Engineering, Texas A&M University, College Station, TX, USA

³Guangxi Key Laboratory of Automatic Detecting Technology and Instruments, Guilin 541004, P.R.China;

^aemail: zbluebird@guet.edu.cn

Keywords: meta-heuristic; global optimization; NP hard problem

Abstract. Since the computation ability of computer improves dramatically, a lot of new meta-heuristic methods arise. All those algorithms are originated from some mechanisms in nature, and are similar in structure and widely used to solve global optimization problems. However, evolutionary 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 a Markov chain model of BBO to analyze the relationship between individual vector and PopSize, and prove that a Markov population series $\{\xi_t, t = 0, 1, 2, \dots\} (\xi_t \in \Omega)$ in BBO is an absorbing Markov chain. Convergence analysis of BBO is obtained, which is the Markov population series in BBO converge to objective subspace B_0^* with probability one.

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 global 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 and 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. When we use BBO to solve optimization problems, there are short of theoretical guarantee. Therefore, In this paper, we propose a Markov chain model of BBO to analyze the relationship between individual vector and PopSize, and prove that a Markov population series $\{\xi_t, t = 0, 1, 2, \dots\} (\xi_t \in \Omega)$ in BBO is an absorbing Markov chain. Convergence analysis of BBO is obtained, which is the Markov population series $\{\xi_t, t = 0, 1, 2, \dots, \infty\} (\xi_t \in \Omega)$ in BBO converge to objective subspace B_0^* with probability 1.

Basic of BBO

Research work about biogeography could be traced to nineteenth century, when Alfred Wallance[8] and Charles Darwin[9] started their research. In 1960s, Robert MacArthur and Edward Wilson obtained breakthrough and they published their achievement and created mathematical model of biogeography. In 2008, Dan Simon [5] proposed a new optimization method based on biogeography called BBO, which are widely applied in Engineering Optimization [10-14].

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).

Theoretical Analysis of BBO

As we create Markov chain model for BBO, we should first map the state space of a Markov chain to the operation space of BBO. If the population size of BBO is one, the state space of the Markov chain is the solution space. That means every solution is a state. If we code the solution with d bits binary system, the solution space is $S = \{1,0\}^d$. Furthermore, the number of the states is $|S| = 2^d$.

If the population size of BBO is PopSize, which is greater than one, suppose the population is a ordered and repeatable set. That means some individuals in the population could be the same. The state space of the Markov chain could be mapped to the population space $\Omega = \{0,1\}^{d \times \text{PopSize}}$. Furthermore, the number of the population states is $|\Omega| = 2^{d \times \text{PopSize}}$. If the population is an unordered and repeatable set, the state space of the Markov chain could be mapped to the population space $\Omega \subset \{0,1\}^{d \times \text{PopSize}}$.

After the state space of the Markov chain is set, arbitrary population X_x and X_y , population X_y is generated by population X_x at a probability $P(\xi_{t+1} = X_y | \xi_t = X_x)$, which is the transition matrix of the Markov chain.

Definition 1 A state S_x of a Markov chain is called an absorbing state, if it could not escape itself. That means $p_{ii} = 1$. A Markov chain is called an absorbing Markov chain, if it has at least one absorbing state and an arbitrary state could reach an absorbing state at one or several steps.

The above definition could be also stated as follows: given a Markov population series $\{\xi_t, t = 0,1,2,\dots\} (\xi_t \in \Omega)$ and a objective subspace $B_0^* \subset \Omega$, the Markov chain is an absorbing Markov chain, if

$$P(\xi_{t+1} \in B_0^* | \xi_t \in B_0^*) = 1, \quad \forall t \in \{0,1,2,\dots\}. \quad (1)$$

In BBO algorithm, there are several operations such as migration and mutation operation. Markov chain model should be created for the respective operation.

Suppose a solution space T is d bits binary system, we get the size of T is $m = 2^d$. $\forall i \in \{1,2,3,\dots,m\}$, a solution x_i , there is a population P with PopSize individuals. We define a vector $w = (w_1, w_2, \dots, w_i, \dots, w_m)$, $\forall i \in \{1,2,3,\dots,m\}$, and w_i indicates the number of x_i in the population P .

We can obtain the relationship between w and PopSize as follows:

Proposition 1 An arbitrary population in BBO is made up of PopSize individuals which select from the search space $T = \{x_1, x_2, x_3, \dots, x_m\}$, that means $\psi: \mathbb{R}^{m \times d} \rightarrow \mathbb{R}^{\text{PopSize} \times d}$, here we get

$$\text{PopSize} = \sum_{i=1}^m w_i \quad (2)$$

where w_i indicates the number of x_i in the population P , PopSize is the population size.

Proof. Let population $P = \{y_1, y_2, \dots, y_k, \dots, y_{\text{PopSize}}\}, k \in \{1, 2, 3, \dots, \text{PopSize}\}, \forall y_k \in T$, we could find $x_c \in X$ and $x_c = y_k, c \in \{1, 2, 3, \dots, m\}$. Therefore, P is rewritten as

$$P = \{\underbrace{x_1, x_1, \dots, x_1}_{w_1}, \underbrace{x_2, x_2, \dots, x_2}_{w_2}, \dots, \underbrace{x_i, x_i, \dots, x_i}_{w_i}, \dots, \underbrace{x_m, x_m, \dots, x_m}_{w_m}\} \quad (3)$$

$\forall w_i \in \{0, 1, 2, \dots, \text{PopSize}\}$. As we know the population size of P is PopSize, the sum of $w_i \in \{0, 1, 2, \dots, \text{PopSize}\}$ is PopSize.

Proposition 2 a Markov population series $\{\xi_t, t = 0, 1, 2, \dots\} (\xi_t \in \Omega)$ in BBO is an absorbing Markov chain.

Proof. Given a Markov population series $\{\xi_t, t = 0, 1, 2, \dots\} (\xi_t \in \Omega)$ and a objective subspace $B_0^* \subset \Omega$, if the Markov chain is an absorbing Markov chain, the necessary and sufficient condition is as follows:

$$P(\xi_{t+1} \in B_0^* | \xi_t \in B_0^*) = 1, \quad \forall t \in \{0, 1, 2, \dots\}$$

First prove sufficiency: if $P(\xi_{t+1} \in B_0^* | \xi_t \in B_0^*) = 1, \quad \forall t \in \{0, 1, 2, \dots\}$ holds, then it means the population ξ_t at moment t belongs to objective subspace B_0^* . It is obvious that the population ξ_{t+1} at moment t+1 must belongs to objective subspace B_0^* . According the definition of absorbing Markov chain (definition 1), the population series $\{\xi_t, t = 0, 1, 2, \dots\} (\xi_t \in \Omega)$ in BBO is an absorbing Markov chain.

Then prove necessity: population series $\{\xi_t, t = 0, 1, 2, \dots\} (\xi_t \in \Omega)$ adopts elite strategy, that is if the population at moment t contains the optimal solution x^* , it will take the place of the worst solution and remain in the next population ξ_{t+1} at moment t+1. Evidently, the population ξ_{t+1} at moment t+1 contains the optimal solution $x^*, \xi_{t+1} \in B_0^*$. Therefore, $P(\xi_{t+1} \in B_0^* | \xi_t \in B_0^*) = 1$.

Lemma 1 Suppose a solution space with d bits binary system, $\theta_i(s) = \{j : x_j(s) = x_i(s)\}$ which is defined as Definition 2, then we get

$$(1) \theta_1(1) = \theta_2(1) = \theta_3(1) = \dots = \theta_{2^{d/2-1}}(1) = \theta_{2^{d/2}}(1).$$

$$(2) \theta_{2^{d/2+1}}(1) = \theta_{2^{d/2+2}}(1) = \theta_{2^{d/2+3}}(1) = \dots = \theta_{2^d-1}(1) = \theta_{2^d}(1).$$

Suppose a population $P = \{y_1, y_2, \dots, y_k, \dots, y_{\text{PopSize}}\}, i \in \{1, 2, 3, \dots, \text{PopSize}\}$, we order y_k according to the sequence of x_i , then we get

$$y_k = \begin{cases} x_1, & \text{if } (k = 1, 2, \dots, w_1) \\ x_2, & \text{if } (k = w_1 + 1, w_1 + 2, \dots, w_1 + w_2) \\ x_3, & \text{if } (k = w_1 + w_2 + 1, w_1 + w_2 + 2, \dots, w_1 + w_2 + w_3) \\ \vdots & \\ x_m, & \text{if } (k = (\sum_{j=1}^{m-1} w_j) + 1, (\sum_{j=1}^{m-1} w_j) + 2, \dots, (\sum_{j=1}^{m-1} w_j) + w_m) \end{cases} \quad (4)$$

The above formula could be simplified as $y_k = x_{g(k)}, k = 1, 2, 3, \dots, \text{PopSize}$

Where $g(k)$ could be expressed as

$$\begin{aligned}
g(k) &= \min(z) \\
\text{subject to } & \sum_{j=1}^z w_j \geq k.
\end{aligned} \tag{5}$$

Proposition 3 Markov transfer matrix T after one generation in BBO is an $H_{\text{PopSize}} \times H_{\text{PopSize}}$ matrix, where H_{PopSize} is

$$\begin{aligned}
H_{\text{PopSize}} &= \sum_{S'(k)} \prod_{i=0}^{\text{PopSize}} \binom{\sum_{j=0}^i k_j}{k_i} \\
S'(k) &= \left\{ k \in R^{\text{PopSize}+1} : k_i \in \{0, 1, 2, \dots, m\}, \sum_{j=0}^{\text{PopSize}} k_j = m, \sum_{j=0}^{\text{PopSize}} jk_j = \text{PopSize} \right\}.
\end{aligned} \tag{6}$$

Proof. As $\text{PopSize} = 1$, H_{PopSize} is m because there are m solutions in the solution space. The right of the above equation equals $\binom{k_0}{k_0} \cdot \binom{k_0 + k_1}{k_1}$. Because $\sum_{j=0}^{\text{PopSize}} jk_j = \text{PopSize}$, we get $0 \times k_0 + 1 \times k_1 = \text{PopSize} = 1$. Because $\sum_{j=0}^{\text{PopSize}} k_j = m$, we get $k_0 = m - 1$ and $k_1 = 1$. The right of equation (21) equals $\binom{m-1}{m-1} \cdot \binom{m}{1} = m$. As $\text{PopSize} = 1$, equation (21) holds.

Suppose equation (21) holds, as $\text{PopSize} = \text{Pop}$. Then as $\text{PopSize} = \text{Pop} + 1$, we get

$$H_{\text{Pop}+1} = \sum_{k_1 + k_2 + \dots + k_{\text{Pop}-1} + K = m} \binom{m}{k_1, k_2, \dots, k_{\text{Pop}-1}, K} \cdot \sum_{k_{\text{Pop}} + k_{\text{Pop}+1} = K} \binom{K}{k_{\text{Pop}}, k_{\text{Pop}+1}}$$

Because of $\binom{m}{k_1, k_2, \dots, k_{\text{Pop}-1}, K} = \frac{m!}{k_1! k_2! \dots k_{\text{Pop}-1}! K!}$ and $\binom{K}{k_{\text{Pop}}, k_{\text{Pop}+1}} = \frac{K!}{k_{\text{Pop}}! k_{\text{Pop}+1}!}$, we obtain

$$H_{\text{Pop}+1} = \frac{m!}{k_1! k_2! \dots k_{\text{Pop}-1}! K!} \times \frac{K!}{k_{\text{Pop}}! k_{\text{Pop}+1}!} = \frac{m!}{k_1! k_2! \dots k_{\text{Pop}-1}! k_{\text{Pop}}! k_{\text{Pop}+1}!}.$$

The above equation could be rewritten as

$$H_{\text{Pop}+1} = \binom{k_0}{k_0} \binom{k_0 + k_1}{k_1} \dots \binom{k_0 + k_1 + \dots + k_{\text{Pop}+1}}{k_{\text{Pop}+1}} = \sum_{S'(k)} \prod_{i=0}^{\text{Pop}+1} \binom{\sum_{j=0}^i k_j}{k_i}.$$

Lemma 2 The optimal value in a series of generations in BBO algorithm is non-increasing, i.e.,

$$S(\mathbf{X}(t+1)) \leq S(\mathbf{X}(t)) \tag{7}$$

Proof. During the update of every generation in BBO algorithm, the best individual x_{best} in t -th generation is stored and used to replace the worst individual in $(t+1)$ -th generation. Therefore, there is always the best individual in the $(t+1)$ -th generation, which belongs to the t -th generation.

Theorem 1. Given a Markov population series $\{\xi_t, t = 0, 1, 2, \dots\}$ ($\xi_t \in \Omega$) in BBO algorithm and an objective subspace $B_0^* \subset \Omega$, where Ω is the population space. $B_0^* = \{Y = \{y_1, y_2, \dots, y_{NP}\} : \exists i \in \{1, 2, \dots, NP\}, y_i \in B^*\}$. The Markov population series

$\{\xi_t, t = 0, 1, 2, \dots, \infty\} (\xi_t \in \Omega)$ converge to objective subspace B_0^* with probability 1, i.e.,

$$\lim_{t \rightarrow \infty} P(\xi_t \in B_0^* | \xi_0) = 1 \quad (8)$$

Proof. Suppose x^* is the optimal solution to the cost function $\text{cost}(\cdot)$:

if $x^* \in \xi_0$, according to the Proposition 2(the Markov population series $\{\xi_t, t = 0, 1, 2, \dots\} (\xi_t \in \Omega)$ in BBO is an absorbing Markov chain), we can get $x^* \in \xi_t, t = 1, 2, 3, \dots$. Therefore, $\lim_{t \rightarrow \infty} P(\xi_t \in B_0^* | \xi_0) = 1$;

if $x^* \notin \xi_0$, suppose $\exists t_1 > 0, x^* \in \xi_{t_1}$, $\exists t_2 > 0, x^* \in \xi_{t_2}$, then $\xi_{t_1} \in B_0^*$ and $\xi_{t_2} \in B_0^*$. Therefore, $P(\xi_{t_1}, \xi_{t_2}) > 0$, $P(\xi_{t_2}, \xi_{t_1}) > 0$, we get $\xi_{t_1} \leftrightarrow \xi_{t_2}$;

if $x^* \notin \xi_0$, suppose $\exists t_1 > 0, x^* \in \xi_{t_1}$ and $\exists t_2 > 0, x^* \notin \xi_{t_2}$, then $P(\xi_{t_1}, \xi_{t_2}) = 0$, that is $\xi_{t_1} \nrightarrow \xi_{t_2}$. Because B_0^* is a normally returned, irreducible and non-cycle closed set, $\forall \xi_0, \exists \pi(Y), \pi(Y)$ is a limit probability distribution. we obtain

$$\lim_{t \rightarrow \infty} P(\xi_t = Y | \xi_0) = \begin{cases} \pi(Y), Y \in B_0^* \\ 0, Y \notin B_0^* \end{cases}$$

That is ξ_t which must be in B_0^* , therefore $\lim_{t \rightarrow \infty} P(\xi_t \in B_0^* | \xi_0) = 1$.

Conclusion

Evolutionary algorithm, such as BBO, is lack of strict theory foundation and hard to be analyzed in theory due to the complicated random behavior. Therefore, we propose a Markov chain model of BBO to analyze the relationship between individual vector and PopSize, and prove that a Markov population series in BBO is an absorbing Markov chain. Convergence analysis of BBO is obtained, which is the Markov population series in BBO converge to objective subspace B_0^* .

Acknowledgement

This work is supported by the National Natural Science Foundation of China (61561012, 61562013), Science and Technology Research Project of Guangxi Department of Education (KY2015YB110), Guangxi Key Laboratory of Automatic Detecting Technology and Instruments (YQ16110), Talent Project of Guilin university of electronic technology(UF15008Y), China Scholarship Council(201508455020). Thanks Professor Fabrizio Lombardi, who is with Northeastern University, USA, he has given some valuable suggestions. Great appreciation to Rabi N Mahapatra, who is with Texas A&M University.

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] WANG L, XU Y. An effective hybrid biogeography-based optimization algorithm for parameter estimation of chaotic systems [J]. *Expert Systems with Applications*, 38 (12): 15103–15109.
- [11] 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
- [12] Aijun Zhu. Zhi Li. Wangchun Zhu, *etal.* Design of Test wrapper scan chain based on Differential Evolution [J]. *Journal of Engineering Science and Technology Review*, 2013, 6(2): 10-14.
- [13] Aijun Zhu, Zhi Li, Chuanpei Xu, *etal.* A new wrapper scan chain balance algorithm for Intellectual Property module in SoC [J]. *Open Electrical and Electronic Engineering Journal*, 2014, 8(1): 42-49.
- [14] Aijun Zhu, Zhi Li and Chuanpei Xu, “Wrapper scan chain design algorithm for SoC test based on biogeography optimization”, *Chinese Journal of Scientific. Instrument*, 2012, 33 (12): 2774-2780.