VOL. 11, NO. 18, SEPTEMBER 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

RECONSTRUCTION OF ARABIC FONT USING ARTIFICIAL BEE COLONY ALGORITHM

Nur ‘Afifah Rusdi and Zainor Ridzuan Yahya
Institute of Engineering Mathematics, University Malaysia Perlis, Kampus Pauh Putra, Arau, Perlis, Malaysia
E-Mail: afifahrusdi@unimap.edu.my

ABSTRACT

Artificial Bee Colony (ABC) algorithm is considered as a new algorithm in the swarm intelligence family. This algorithm grabs the attention of the researchers due to its potentials in solving various types of problems. This paper presented the use of the algorithm for the purpose of reconstruction of Arabic fonts. ABC algorithm has been used to find the optimal solution of the curve fitting problem by using the cubic Bézier curve. Since the purpose of this study is to minimize the distance between boundary of the original image and parametric curve, Sum Square Error (SSE) has been used to calculate the error given by these two curves. The procedure of curve fitting process includes the following steps: boundary extraction and corner point detection, chord length parameterization and curve fitting. The finding of this study shows that the proposed method has successfully generated fitted cubic Bézier curve that resemble the boundary of original images and enhance its performance since it produce a small error.

Keywords: artificial bee colony algorithm, curve fitting, cubic bézier curve, sum square error.

INTRODUCTION

Curve fitting is vastly used in reverse engineering especially in image processing as well as computer graphics pattern recognition and computer aided geometric design (CAGD) (Rusdi and Yahya 2015). The aim of curve fitting is to find the set of data points that can accurately represent the given curve. Fitting curve to the data extracted by generic shapes has received a lot of attention of researchers due to its applications in various fields such as in engineering and medical and its demands in the industry (Sarfraz, Irshad and Hussain, 2013).

Several techniques are proposed in the literatures by using different spline models to represent the curves like B-splines (Roslan and Yahya 2015), (Sarfraz, 2013) and Bézier splines (Rusdi and Yahya 2015), (Sarfraz and Masood, 2007), (Sarfraz and Masood, 2007), (Sarfraz and Razzak, 2002), (Sarfraz and Khan, 2002). Bézier spline is widely used as a proposed curve since it is the most stable among all the polynomial-based curves used. This study explores the use of cubic Bézier curve for the purpose of reconstruction of Arabic fonts.

In an attempt to find the best solution, the usage of soft computing based optimization methods in the various types of problem shows the anticipation from the researcher to use this method especially in solving complex engineering problems. Soft computing methods do not guarantee optimal solution but require a relatively short computational time. The soft computing based optimization algorithms can be classified into two classes which are Evolutionary algorithms (EA) and swarm intelligence algorithms. Genetic algorithm (GA) is one of the Evolutionary algorithms that has been used to solve curve fitting problems (Sarfraz, Irshad, and Hussain, 2013), (Barkeshli, Mokhtari, and Amiri, 2001). Other than that, (Rusdi and Yahya, 2015), (Pandunata and Shamsuddin, 2010), (Roslan and Yahya, 2015) used Differential Evolution (DE) algorithm as their proposed method. The DE algorithm was introduced to overcome the poor local search ability of GA. These algorithms can provide better solutions in comparison to classical algorithms.

Swarm intelligence is a nature inspired algorithm which focuses on insect’s behavior. Artificial Bee Colony (ABC) algorithm is a relatively new member of swarm intelligence that tries to model natural behavior of real honey bees in food foraging (Baykasoğlu, Ozbakir, and Tapkan, 2007). According to (Kumar and Kumar, 2013), the ABC algorithm has been applied in various fields including Benchmark optimization, bioinformatics field, data mining, scheduling and engineering design and applications. (Karaboga and Basturk, 2008) presented the performance of the ABC algorithm. In their study, ABC algorithm was compared with DE, EA, and particle swarm optimization (PSO). Results showed that the ABC algorithm could be performed better than these algorithms and it can be efficiently employed to solve multimodal engineering problems with high dimensions. Not only that, (Singh, 2009) presented the use of the ABC algorithm to solve leaf-constrained minimum spanning tree problems. In his paper, the ABC algorithm was compared with three other metaheuristic approaches which are GA, ant colony optimization (ACO) and Tabu Search (TS). Based on the result, the ABC algorithm seems to work well compared to these methods since it required short computational time and it produce a quality solution. Besides that, the ABC algorithm has also been used to solve multidimensional and multimodal optimization problem (Karaboga, 2005).

Therefore, in this study the ABC algorithm was proposed to solve the curve fitting problems by using the cubic Bézier curve. This paper can be divided into six sections. Section 2 discussed about the cubic Bézier curve and the ABC algorithm. Then it was followed by results and discussion including boundary extraction and corner detection, chord length parameterization and elaboration on the curve fitting processes. The paper has been concluded in Section 7.

VOL. 11, NO. 18, SEPTEMBER 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

CUBIC BÉZIER CURVE

Bézier curve is defined as a parametric curve $P(t)$. This curve can be defined in any degree n with $n+1$ control points (Rusdi and Yahya, 2015). In general, the equation of Bézier curve is given by:

$$P(t) = \sum_{i=0}^{n} B_i^n(t) P_i \quad (1)$$

where

$$B_i^n(t) = \frac{n!}{(n-i)!(i)!}(1-t)^{n-i} t^i$$

P_i = Control points

From Equation (1), P_i are the control points and $B_i^n(t)$ is the blending function, namely Bernstein Polynomial. For this study, cubic Bézier curve with $n=3$ was used. It consists of four control points P_0, P_1, P_2, and P_3 as shown in Figure-2.

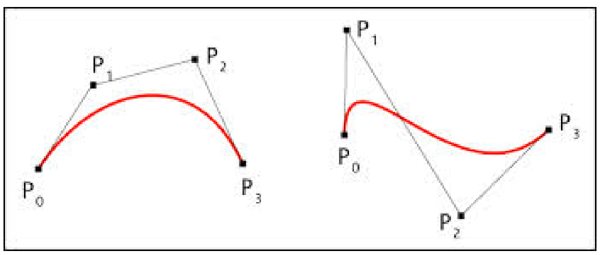


Figure-1. Examples of cubic Bézier curve.

ARTIFICIAL BEE COLONY

ABC optimization algorithm was proposed by (Karaboga, 2005), which is inspired by the principles of foraging behavior of honey bees. The aim of this algorithm is to optimize multivariable numerical functions (Karaboga and Basturk, 2008). In this algorithm, there are several mechanisms that have been used by honey bees to optimally locate food sources and to search new location. This makes them good candidates for developing a new intelligent search algorithm.

According to (Kumar and Kumar, 2013), ABC algorithm is more preferable compared to other optimization methods like DE, PSO and GA since it offers a bundle of advantages. Firstly, this algorithm has least control parameters as compared to these three methods. Apart from population size and maximum cycle numbers, DE and PSO have two other control parameters and GA has three other control parameters, meanwhile ABC algorithm only has one control parameter which is called 'limit'. This makes it more robust and less time consuming. Other than that, it is known as simple and flexible procedure. The flexibility of this method makes it easier to be hybridized with other optimization methods.

Behavior of real bees

Bee system consists of two essential components namely food sources and foragers. There are few elements that are important to ensure the value of food sources such as proximity to the nest, richness of energy and ease of extracting this energy (Baykasoğlu *et al*., 2007). Foragers can be divided into two main groups which are employed foragers or known as employed bees and unemployed foragers.

According to (Karaboga and Basturk, 2008), there are two types of unemployed foragers namely onlooker bees and scout bees. Employed bees are associated with a particular food source. They will carry information like about the specific source, its distance and direction from nest and the profitability of the food sources and share the information with other bees waiting in the hive (Karaboga and Akay, 2009). When employed bees found the food source, the bee utilizes its own capability to memorize and then immediately start exploiting the food location.

Meanwhile, onlooker bees are assigned to selecting good food sources from those founded by employed bees and finding a food source based on the information gathered by employed bees. Last but not least, scout bees are translated from few employed bees which abandon their food sources and search for new ones (Zhu and Kwong, 2010).

Artificial bee colony algorithm

The ABC algorithm consists of four main phases which are initialization phase, employed bee phase, onlooker bee phase and scout phase. This section will explain in detail the role of each phase.

In the initialization phase, there are few parameters that need to be set which are population sizes, maximum cycle number and limit. Initial solutions are randomly produced by Equation (2), where N is the number of employed bees, D represents the dimensional of solution space and $rand(0,1)$ is a random number within the range $[0,1]$.

$$x_i^j = x_{\min}^j + rand(0,1) \times (x_{\max}^j - x_{\min}^j)$$
$$i = 1,2,...,N \quad (2)$$
$$j = 1,2,...,D$$

Apart from that, the objective function will be calculated in this phase by using the equation specific for the problem. The fitness value *fit* of each solution *i* is obtained by using Equation (3) as below.

$$fit\,i = \begin{cases} 1/(1+fi) & if\ fi \geq 0 \\ 1+abs(fi) & if\ fi < 0 \end{cases} \quad (3)$$

where fi is the objective function value that was calculated earlier. Lastly, trial counter is set to be zero at this phase (Kiran and Gündüz, 2012). The number of trials

VOL. 11, NO. 18, SEPTEMBER 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

for releasing a food source is equal to the value of 'limit', which is an important control parameter in ABC algorithm.

The next phase of this process is called the employed bee phase. Number of employed bees is equal to the number of the population size. Employed bees go onto the food source to measure the nectar amounts. They produce a modification on the position (solution) in their memory depending on the local information and the nectar amount (fitness value) of the new source (new solution). The position of the selected neighbor food source is calculated as in Equation (4).

$$v_{i,j} = x_{i,j} + \Phi(x_{i,j} - x_{k,j}) \quad (4)$$

where k is random selected solution in neighborhood of i, j is random selected dimension of the problem and Φ is a random number in the range of $[-1,1]$. If the fitness value of new solution is better than the fitness value of old solution, the position x_i of the food source is changed to be v_i, otherwise, x_i is kept as it is. In the other word, the greedy selection mechanism is employed as the selection operation between the old and new solutions. After all employed bees completed the search process, they will share the nectar information of the food source and their position information with the onlooker bees (Dervis Karaboga and Ozturk, 2011).

At the third phase, onlooker bees were assigned to evaluate the profitability of the food source and select the highest food source. The profitability of food sources p_i is determined by:

$$p_i = \frac{fit\,i}{\sum fit\,i} \quad (5)$$

In Equation (5), $fit\,i$ refers to fitness value of each bee i and $\sum fit\,i$ is the total fitness values. After selecting one of the food sources, the position of the selected neighbor food source is calculated by using Equation (4). If the fitness value $fit\,i$ at v_i is higher than the fitness value $fit\,i$ at x_i, then x_i will be replaced by v_i, otherwise x_i is kept as it is. Note that, employed bees and onlooker bees carry out the exploitation process in the search space. Exploitation is defined as the ability to apply knowledge of the previous good solution to find a better solution (Zhu and Kwong, 2010).

The last phase of ABC algorithm is called the scout phase. At this step, if the position of x_i of the food source is not improved through the predetermined number of trials 'limit', a new food source is randomly determined by a scout bee and replaced the abandoned one. The new solution will be discovered by the scout bee by using the expression in Equation (2). Notice Phase two to Phase four are repeated until the termination criterion is satisfied or until achieved the maximum cycle number.

BOUNDARY EXTRACTION AND CORNER POINT DETECTION

The first process in reverse engineering of generic shapes is to extract the boundary and detect the corner point of the proposed shapes. The generic shapes must be represented in the form of bitmap images. For this study, there are three Arabic fonts that have been used. All these three fonts were obtained from electronic device. After getting the bitmap images, the next step is to extract the boundary of the images. Built in MATLAB function called boundaries has been used to detect the boundaries of the proposed images. The selection of the boundary points are based on their corner strength and contour fluctuations.

The next process is called corner point detection. Corner points are defined as points that partition the boundary into several segments before curve fitting. According to (Sarfraz, Irshad and Hussain, 2013), there are two main reasons of this step. One of the reason is, this process can reduce the boundary's complexity and simplifies the fitting process. Other reason is it can improve the quality of the approximation since the boundary is subdivided into smaller pieces. There are numerous algorithms that have been discussed in literatures to detect the corner points. The method proposed in (Rusdi and Yahya, 2015) was used in this paper. Extracted boundaries and detected corners are presented in Table-1.

CHORD LENGTH PARAMETERIZATION

As mentioned in the previous section, cubic Bézier curve consists of four control points P_0, P_1, P_2 and P_3. This curve interpolates the two end control points P_0 and P_3, and approximate the two intermediate points P_1 and P_2. From Equation (1), when $n = 3$, the expression becomes:

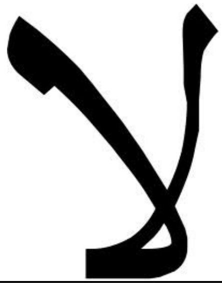
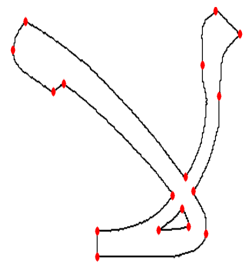
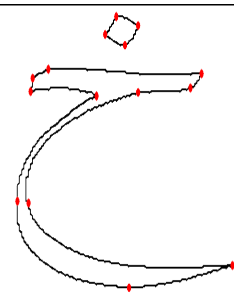
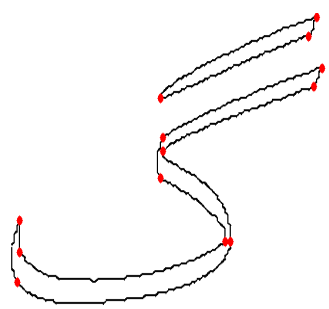
$$P(t) = B_0^3(t)P_0 + B_1^3(t)P_1 + B_2^3(t)P_2 + B_3^3(t)P_3 \quad (6)$$

The expansion of Equation (6) can be written in matrix form as shown in Equation (7).

ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

Table-1. Boundary of images and detected corner points.

Original images	Boundary of bitmap images	No. of corner points
 Lam Alif		17
 Kha		15
 Kaf		13

$$P(t)=\begin{bmatrix} t^3 & t^2 & t & 1 \end{bmatrix}\begin{bmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 3 \\ -3 & 3 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} P_0 \\ P_1 \\ P_2 \\ P_3 \end{bmatrix} \quad (7)$$

Then, chord length parameterization has been used to find the value of t via the expression in Equation (8).

$$t_i=\begin{cases} 0 & if \quad i=1 \\ \dfrac{|p_1p_2|+|p_2p_3|+...+|p_ip_{i+1}|}{|p_1p_2|+|p_2p_3|+...+|p_{n-1}p_n|} & if \quad i\le n-1 \\ 1 & if \quad i=n \end{cases} \quad (8)$$

CURVE FITTING USING ABC ALGORITHM

The climax of the reconstruction of Arabic fonts is called curve fitting. This process is done when the t value associated with each point has been calculated. At this step, the two intermediate points P_1 and P_2 were approximated by using the ABC algorithm. Then, these points will be used to find the parametric curve that give the best optimal results for given values of t.

At this phase, several parameters have been set. According to (Dervis, Karaboga and Basturk, 2007), the number of the employed bees or the onlooker bees is equal to the number of solutions in the population (population size). Hence, for this study, the population size, employed bees and onlooker bees are set to be 100 and the number of iterations has been set to be 50. Besides that, two other parameters are also defined which are dimension and limit. Since there are two points to be approximated, the dimension of this problem becomes four and the limit is set to be 200. Then, the initial solutions are obtained by

VOL. 11, NO. 18, SEPTEMBER 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

using Equation (2). Since the purpose of this study is to minimize the distance between boundary of the original image and parametric curve, Sum Square Error (SSE) was used as an objective function to calculate the error given by those two curves. The procedure of curve fitting process can be summarized in a flowchart as shown in Figure-3.

In the solution procedure, the stopping criterion has been set which is every process will be repeated ten times and the average error will be recorded. The proposed method is illustrated in Figure-1.

The demonstration of the fitted cubic Bézier curve (solid line) over boundary of original image (dotted line) is shown in Figure-4 and the error is illustrated in Table-2. As seen in Figure-4, the most of the fitted cubic Bézier curve is on the boundary of the bitmap images. Note that in this study, Douglas-Peucker algorithm has been used to minimize the distance of data. This is to ensure that the computational time can be minimized.

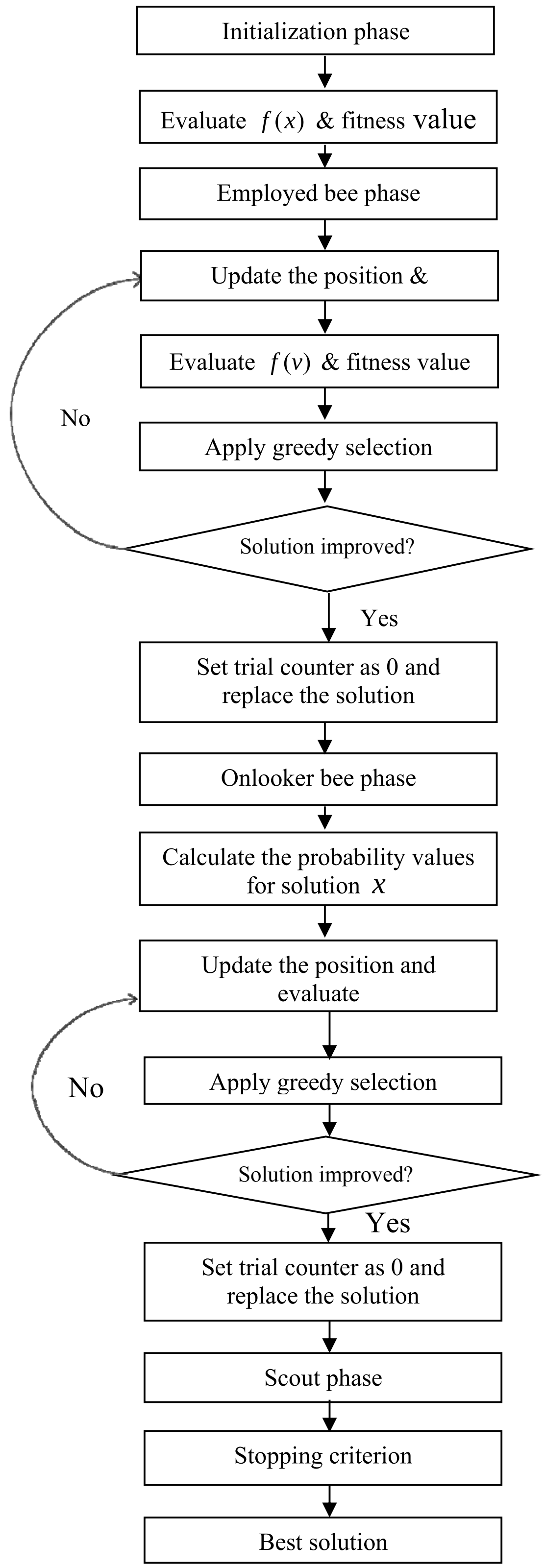


Figure-2. Flowchart for the ABC algorithm.

ARPN Journal of Engineering and Applied Sciences

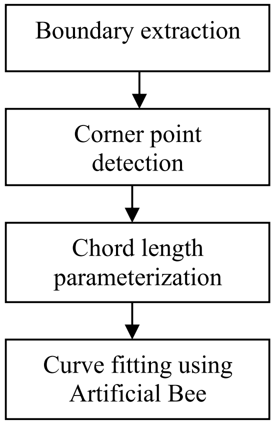


Figure-3. Flow chart of proposed algorithm.

Table-2. Sum square error.

Images	SSE
Lam Alif	34.1060
Kha	18.9343
Kaf	23.1268

CONCLUSIONS

This paper presented an efficient algorithm for the approximation of boundary of Arabic fonts by using the parametric cubic Bézier curve for curve fitting. Although the ABC algorithm is considered as a new algorithm in the swarm intelligence family but this algorithm already proves that it can produce a better solutions. Douglas-Peucker algorithm was used in this study in order to improve the performance of this method. This algorithm was used to reduce the distance of data at each curve and this study has shown that the method proposed can be used since it produce a small error.

ACKNOWLEDGMENT

This study was supported by short-term research grants 2015 (9001-00445). The author is very grateful for the valuable comments and suggestions given by the referees to improve the quality of the paper.

REFERENCES

Barkeshli, K., Mokhtari, M. and Amiri, M. 2001. Image reconstruction of impenetrable cylinders using cubic B-splines and genetic algorithms. In Antennas and Propagation Society International Symposium, 2001. IEEE, 2, pp. 686-689.

Baykasoglu, A., Ozbakir, L. and Tapkan, P. 2007. Artificial bee colony algorithm and its application to generalized assignment problem. Swarm Intelligence: Focus on Ant and particle swarm optimization, 113-144.

Karaboga, D. and Akay, B. 2009. A comparative study of Artificial Bee Colony algorithm. Applied Mathematics and Computation, 214(1), pp. 108-132.

Karaboga, D. and Basturk, B. 2007. A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (ABC) algorithm. Journal of Global Optimization, 39(3), pp. 459-471.

Karaboga, D., & Basturk, B. 2008. On the performance of artificial bee colony (ABC) algorithm. Applied Soft Computing Journal, 8(1), pp.687–697.

Karaboga, D. and Ozturk, C. 2011. A novel clustering approach: Artificial Bee Colony (ABC) algorithm. Applied Soft Computing Journal, 11(1), pp. 652-657.

Karaboga, D. 2005. An idea based on honey bee swarm for numerical optimization (Vol. 200). Technical report-tr06, Erciyes university, engineering faculty, computer engineering department.

Kiran, M. S. and Gündüz, M. 2012. A novel artificial bee colony-based algorithm for solving the numerical optimization problems. International Journal of Innovative Computing, Information and Control, 8(9), pp.6107-6121.

Kumar, B. and Kumar, D. 2013. A review on Artificial Bee Colony algorithm Artificial bee colony algorithm, 2(3), pp. 175-186.

Rusdi, N.A. and Yahya, Z.R. 2015. Pembinaan semula Fon Arab menggunakan Lengkung Bézier Kuartik. Sains Malaysiana, Accepted.

Rusdi, N.A. and Yahya, Z.R. 2015. Reconstruction of generic shape with cubic Bézier using least square method. International Conference on Mathematics, Engineering & Industrial Applications 2014 (ICoMEIA 2014), AIP Publishing,1660,pp.050004.

Roslan, N. and Yahya, Z.R. 2015. Reconstruction of egg shape using B-spline. International Conference on Mathematics, Engineering and Industrial Applications 2014 (ICoMEIA 2014), AIP Publishing,1660,pp.050016.

Roslan, N. and Yahya, Z.R. 2015. Pembinaan semula Fon dengan Bézier Kuartik Menggunakan Evolusi Pembezaan. Sains Malaysiana, Accepted.

Sarfraz, M. 2013. Designing And Reengineering Objects Using Rational Quadratic Splines. In The 6th International Conference on Information Technology (ICIT 2013).

Sarfraz, M. and Khan, M. a. 2002. Automatic outline capture of Arabic fonts. Information Sciences, 140(3-4), pp. 269-281.

Sarfraz, M. and Khan, M. a. 2004. An automatic algorithm for approximating boundary of bitmap characters. Future Generation Computer Systems, 20(8), pp.1327-1336.

Sarfraz, M. and Masood, a. 2007. Capturing outlines of planar images using Bézier cubics. Computers and Graphics (Pergamon), 31(5), pp. 719-729.

Sarfraz, M. and Razzak, M. F. a. 2002. An algorithm for automatic capturing of the font outlines. Computers and Graphics (Pergamon), 26(5), pp.795-804.

Sarfraz, M., Irshad, M. and Hussain, M. Z. 2013. Reverse Engineering of Planar Objects Using GAs. Sains Malaysiana, 42(8), pp.1167-1179.

Singh, A. 2009. An artificial bee colony algorithm for the leaf-constrained minimum spanning tree problem. Applied Soft Computing, 9(2), pp. 625-631.

Zhu, G. and Kwong, S. 2010. Gbest-guided artificial bee colony algorithm for numerical function optimization. Applied Mathematics and Computation, 217(7), pp. 3166-3173.