

# A Probabilistic Analysis of a Simplified Biogeography-Based Optimization Algorithm

Dan Simon\*  
Cleveland State University  
Department of Electrical and Computer Engineering  
Cleveland, Ohio 44115

February 13, 2009

## Abstract

Biogeography-based optimization (BBO) is a population-based evolutionary algorithm (EA) that is based on the mathematics of biogeography. Biogeography is the study of the geographical distribution of biological organisms. We present a simplified version of BBO and then we analyze its population using probability theory. Our analysis provides approximate values for the expected number of generations before the population's best solution improves, and the expected amount of improvement. These expected values are functions of the population size. We quantify three behaviors as the population size increases: first, the best solution in the initial randomly-generated population improves; second, the expected number of generations before improvement increases; and third, the expected amount of improvement decreases.

*Key Words* – biogeography-based optimization, evolutionary algorithms, probability, Markov analysis.

---

\*Email address d.j.simon@csuohio.edu. This work was supported by NSF Grant 0826124 in the CMMI Division of the Engineering Directorate.

# 1 Introduction

## 1.1 The Science of Biogeography

The science of biogeography can be traced to the work of 19th century naturalists such as Alfred Wallace [1] and Charles Darwin [2]. Eugene Munroe was the first to develop mathematical models of biogeography in 1948 [3, 4], and Robert MacArthur and Edward Wilson were the first to extensively develop and publicize them in the 1960's [5, 6].

Mathematical models of biogeography describe the migration, speciation, and extinction of species. Species migrate between islands. Note that the term “island” here is used descriptively rather than literally. That is, an island is any habitat that is geographically isolated from other habitats. In the classic sense of the term, an island is isolated from other habitats by water. But islands can also be habitats that are isolated by stretches of desert, rivers, impassable mountain ranges, predators, or other obstacles.

Islands that are well suited as residences for biological species are said to have a high island suitability index (ISI). Features that correlate with ISI include factors such as rainfall, diversity of vegetation, diversity of topographic features, land area, and temperature. The variables that characterize habitability are called suitability index variables (SIVs). SIVs can be considered the independent variables of the island, and ISI can be considered the dependent variable.

Islands with a high ISI tend to have a large number of species, while those with a low ISI have a small number of species. Islands with a high ISI have many species that emigrate to nearby islands. The reason that emigration occurs from high ISI islands is due to the accumulation of random effects on a large population. For example, emigration could occur as animals ride flotsam to neighboring islands, or swim to neighboring islands. In the case of insects and birds, emigration can occur by flying or being carried by the wind.

Islands with a high ISI have a low immigration rate because they are already nearly saturated with species and cannot easily support new species. Islands with a low ISI have a high immigration rate because of their sparse populations, which allows a lot of room for additional species. The immigration of new species to low ISI islands might raise the ISI of the island because the suitability of an island is proportional to its biological diversity.

## 1.2 The Application of Biogeography to Optimization

The application of biogeography to optimization was first presented in [7] and is an example of how a natural process can be modeled to solve general optimization problems. This is similar to what has occurred in the past few decades with genetic algorithms, neural networks, ant colony optimization, particle swarm optimization, and other areas of computer intelligence.

Biogeography is nature's way of distributing species, and is analogous to general problem solving. Suppose that we have some problem, and that we also have a certain number of candidate solutions. A good solution is analogous to an island with a high ISI, and a poor solution is like an island with a low ISI. High ISI solutions are more likely to share their features with other solutions, and low ISI solutions are more likely to accept shared features from other solutions. We call this new approach to problem solving biogeography-based optimization (BBO). As with every other evolutionary algorithm (EA), each solution might also have some probability of mutation, although mutation is not an essential feature of BBO.

## 1.3 Paper Goals and Overview

The goals of this paper are three-fold. Our first goal is to present an overview of the BBO algorithm, which we do in Section 2. Our second goal is to present a simplified version of BBO, which we do in Section 3. Our third goal is to use probability theory to analyze some of the mathematical properties of the simplified BBO (SBBO), which we do in the following sections.

Section 4 derives the probability that the best individual in the SBBO population improves from one generation to the next. Section 5 derives an approximate Markov transition matrix for the SBBO. A more traditional Markov analysis, along the lines of that discussed in [8] for GAs, has been presented for the regular BBO algorithm [9]. The limitations of traditional Markov analyses of EAs is that the transition matrix grows factorially with the problem size. The size of the approximate SBBO transition matrix derived in Section 5 grows only with the square of the population size. Section 6 derives the expected amount of improvement in the best individual over one SBBO generation. We give some supporting simulation results in Section 7, and provide some concluding remarks and directions for future work in Section 8.

## 2 Biogeography-Based Optimization

### 2.1 Biogeography

Figure 1 illustrates a model of species abundance in a single island. The immigration rate  $\lambda$  and the emigration rate  $\mu$  are functions of the number of species on the island.

The immigration curve shows that the maximum possible immigration rate  $I$  occurs when there are zero species on the island. As the number of species increases, the island becomes more crowded, fewer species are able to successfully survive immigration, and the immigration rate decreases. The largest possible number of species that the island can support is  $S_{max}$ , at which point the immigration rate is zero.

The emigration curve shows that if there are no species on the island then the emigration rate is zero. As the number of species increases, the island becomes more crowded, more species are available to leave the island to explore other possible habitats, and the emigration rate increases. The maximum emigration rate  $E$  occurs when the island contains the largest number of species that it can support.

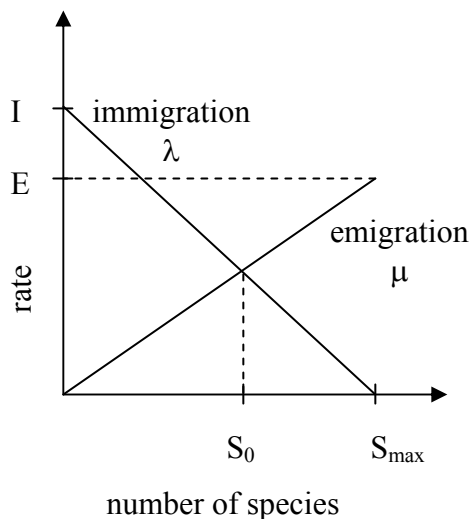


Figure 1: Linear migration curves for an island [6]. The equilibrium number of species is  $S_0$ , at which point the immigration and emigration rates are equal.

We have shown the immigration and emigration curves in Figure 1 as straight lines, but in general they might be more complicated curves. Nevertheless this simple model gives us a general description of the process of immigration and emigration. The details can be

adjusted if needed.

## 2.2 Biogeography-Based Optimization Algorithms

Suppose that we have a problem and a population of candidate solutions that can be represented as vectors. Further suppose that we have some way of assessing the goodness of the solutions. Good solutions are considered to be islands with high ISI, and poor solutions are considered to be islands with low ISI. Note that ISI is the same as “fitness” in other population-based optimization algorithms (GAs, for example). High ISI solutions are like islands with many species, and low ISI solutions are like islands with few species. We assume that each solution (island) has an identical species curve (with  $E = I$  for simplicity), but the  $S$  value represented by the solution depends on its ISI. For example,  $S_1$  in Figure 2 represents a low ISI solution while  $S_2$  represents a high ISI solution.

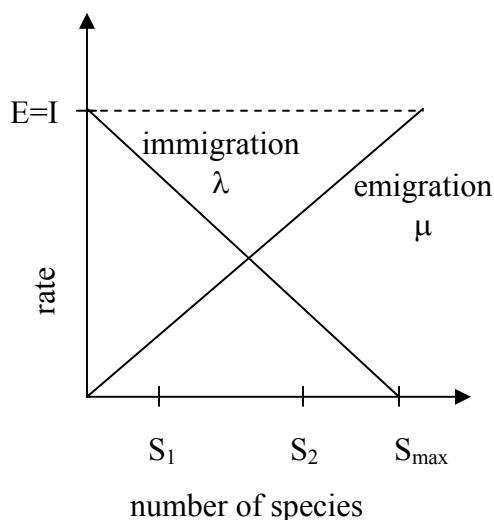


Figure 2: Illustration of two candidate solutions to some problem with symmetric immigration and emigration curves.  $S_1$  is a relatively poor solution while  $S_2$  is a relatively good solution.  $S_1$  has a high immigration and a low emigration rate, and  $S_2$  has a low immigration and a high emigration rate.

$S_1$  in Figure 2 is like a low ISI island, while  $S_2$  is like a high ISI island. The immigration rate for  $S_1$  will therefore be higher than the immigration rate for  $S_2$ . The emigration rate for  $S_1$  will be lower than the emigration rate for  $S_2$ . We use the migration rates of each solution to probabilistically share information between islands. This can be implemented in several different ways [10].

**Algorithm (1): Partial Immigration-Based BBO**

The first approach is to base migration on immigration rates for each island, and probabilistically decide whether or not to immigrate each SIV (solution feature) independently. The simulation results presented in the original BBO paper [7] were obtained with this approach. The following algorithm describes one generation of this approach, where we use the notation  $x_i(s)$  to denote the  $s$ th feature of the  $i$ th population member.

```

For each island  $x_i$ 
  For each SIV  $s$ 
    Use  $\lambda_i$  to probabilistically decide whether to immigrate to  $x_i(s)$ 
    If immigrating then
      Use  $\mu$  to probabilistically select the emigrating island  $x_j$ 
       $x_i(s) \leftarrow x_j(s)$ 
    end if
  next SIV
next island

```

**Algorithm (2): Partial Emigration-Based BBO**

The second approach is to base migration on emigration rates for each island, and probabilistically decide whether or not to immigrate each SIV independently. One generation of this approach can be described as follows.

```

For each island  $x_i$ 
  For each SIV  $s$ 
    Use  $\mu_i$  to probabilistically decide whether to emigrate from  $x_i(s)$ 
    If emigrating then
      Use  $\lambda$  to probabilistically select the immigrating island  $x_j$ 
       $x_j(s) \leftarrow x_i(s)$ 
    end if
  next SIV
next island

```

**Algorithm (3): Single Immigration-Based BBO**

The third approach is to base migration on immigration rates for each island, and probabilistically decide whether or not to immigrate a single randomly-chosen SIV. One generation of this approach, which was presented in the original BBO paper [7], is described as follows.

```

For each island  $x_i$ 
  Use  $\lambda_i$  to probabilistically decide whether to immigrate to  $x_i(s)$ 
  If immigrating then
    Pick a random SIV  $s$ 
    Use  $\mu$  to probabilistically select the emigrating island  $x_j$ 
     $x_i(s) \leftarrow x_j(s)$ 

```

```

    end if
next island

```

#### Algorithm (4): Single Emigration-Based BBO

The fourth approach is to base migration on emigration rates for each island, and probabilistically decide whether or not to emigrate a single randomly-chosen SIV. One generation of this approach can be described as follows.

```

For each island  $x_i$ 
  Use  $\mu_i$  to probabilistically decide whether to emigrate from  $x_i(s)$ 
  If emigrating then
    Pick a random SIV  $s$ 
    Use  $\lambda$  to probabilistically select the immigrating island  $x_j$ 
     $x_j(s) \leftarrow x_i(s)$ 
  end if
next island

```

The BBO migration strategy is similar to the global recombination approach of evolutionary strategies (ES) [11] in which many parents can contribute to a single offspring, but it differs in at least one important aspect. In ES global recombination is used to create new solutions, while in BBO migration is used to change existing solutions. Global recombination in ES is a reproductive process, while migration in BBO is an adaptive process that is used to modify existing islands. A quantitative comparison between BBO and ES is included in [7].

As with other population-based algorithms, we typically incorporate elitism in order to retain the best solutions in the population. This prevents the best solutions from being corrupted by immigration. Elitism can be implemented by setting  $\lambda_i = 0$  for the  $p$  best habitats, where  $p$  is a user-selected elitism parameter.

For each of the four algorithms summarized above, mutation is typically implemented to increase the search space of the algorithm. The BBO mutation mechanism is problem-dependent, just as it is for other EAs.

### 2.3 BBO Compared to Other Evolutionary Algorithms

BBO has certain features in common with other EAs. Like genetic algorithms (GAs) [12], BBO has a way of sharing information between solutions. However, GA solutions typically die at the end of each generation, while BBO solutions survive forever, although their

characteristics change as the algorithm progresses. Although BBO is a population-based optimization algorithm it does not involve reproduction or the generation of children. This distinguishes it from reproductive strategies such as GAs and ES.

BBO also differs from ant colony optimization (ACO), which generates a new set of solutions with each iteration [13, 14]. BBO, on the other hand, maintains its set of solutions from one iteration to the next, relying on migration to probabilistically adapt those solutions.

BBO has the most in common with algorithms such as particle swarm optimization (PSO) [15, 16] and differential evolution (DE) [17]. In PSO and DE, solutions are maintained from one iteration to the next, but each solution is able to learn from its neighbors and adapt itself as the algorithm progresses. PSO represents each solution as a point in space, and represents the change over time of each solution as a velocity vector. However, PSO solutions do not change directly; it is rather their *velocities* that change, and this indirectly results in position (i.e., solution) changes. DE changes its solutions directly, but changes in a particular DE solution are based on differences between other DE solutions. Also, DE is not biologically motivated. BBO can be contrasted with PSO and DE in that BBO solutions are changed directly via migration from other solutions (islands). That is, BBO solutions directly share their attributes (SIVs) with other solutions.

It is the distinctive of each EA that determines its relative strengths and weaknesses. An important open research question is how algorithmic differences affect the performance of a given strategy on different types of problems. Theoretical and empirical investigations into this question remain to be done.

### **3 A Simplified BBO Algorithm**

A simplified BBO (SBBO) algorithm can be formulated by always using the best solution as the emigrating island, and using any other randomly chosen solution as the immigrating island. The immigrating island is chosen from a uniform probability distribution and is thus independent of fitness. This is conceptually similar to the stud GA in which the best chromosome is always chosen as one of the parents, and the other parent is chosen using standard fitness-based selection [18, 19]. In SBBO, the migration curves of Figure 2 are

modified to those of Figure 3.

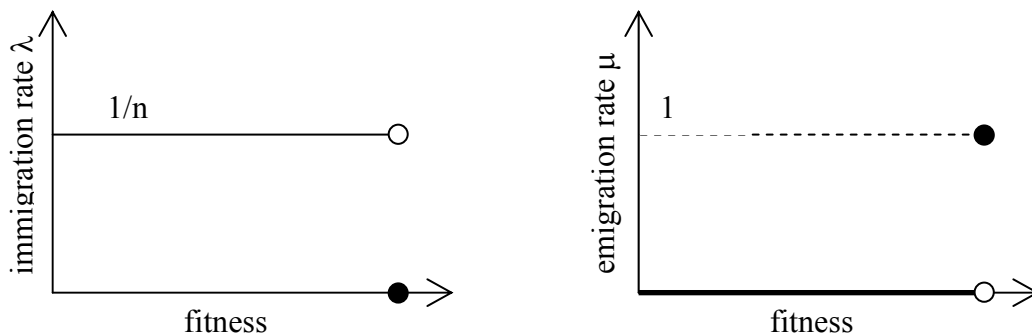


Figure 3: Illustration of SBBO migration curves in an  $n$ -island population. All solutions have an equal probability of immigration except for the most fit solution, which has a zero probability of immigration. The most fit solution has a 100% probability of emigration.

One generation of the SBBO approach can be described as follows.

- (1) Find the fittest solution. Call this solution  $x_i$ .
- (2) Pick a random SIV  $s$
- (3) Select the immigrating island  $x_j$  from a uniform probability distribution
- (4)  $x_j(s) \leftarrow x_i(s)$

One advantage of SBBO is that we do not need to evaluate all fitness values each generation. We only need to keep track of the most fit island. So after each generation we compare the new fitness of the immigrating island with the previous best solution. This requires only one fitness evaluation per generation.

## 4 Probability Analysis

From this point on we assume that we are trying to solve a minimization problem. High fitness corresponds to low cost. We try to minimize cost, which is always non-negative, and maximize fitness.

Suppose that the cost of a solution  $x_i$  is denoted as  $f(x_i)$ . We write

$$x_i = \left[ x_{i1} \quad \cdots \quad x_{is} \right] \quad (1)$$

where  $s$  is the dimension of the problem and also the number of SIVs in each island. If  $f(x_i)$  is separable then it can be written as

$$f(x_i) = f_1(x_{i1}) + \cdots + f_s(x_{is}) \quad (2)$$

If  $x_i$  is a randomly-chosen island, then  $f(x_i)$  is a random variable with mean  $\bar{f}$  and variance  $\sigma^2$ . The parameters  $\bar{f}$  and  $\sigma$  can be approximated from a population of islands  $\{x_i\}$ . In the absence of any other information, we assume that the  $f_j(\cdot)$  function values on the right side of (2) are independent identically-distributed random variables. Then the expected value and variance of each  $f_j(\cdot)$  can be written as

$$\begin{aligned}\bar{f}_j &= \bar{f}/s \\ \sigma_j^2 &= \sigma^2/s\end{aligned}\tag{3}$$

Now suppose that we migrate a single SIV from the lowest-cost island to a randomly-chosen island. The contribution of a single SIV to the cost of the immigrating island can be approximated as the random variable

$$y \sim (f_{\min}/s, \sigma^2/s)\tag{4}$$

where we are using the notation  $y \sim (\mu, \sigma^2)$  to indicate that  $y$  is a random variable with mean  $\mu$  and variance  $\sigma^2$ . If  $y$  is a uniform random variable, then it can also be written as

$$y \sim U[f_{\min}/s - \sigma\sqrt{3/s}, f_{\min}/s + \sigma\sqrt{3/s}]\tag{5}$$

where we are using the notation  $y \sim U[a, b]$  to indicate that  $y$  is a uniform random variable whose probability density function (pdf) is nonzero only from  $a$  to  $b$ .

The immigrating island can have any fitness  $f \in [f_{\min}, f_{\max}]$ . Therefore, before an SIV is replaced in the immigrating island, that SIV contributes a cost that can be approximated as the random variable

$$\begin{aligned}z &\sim (f/s, \sigma^2/s) \\ &\sim U[f/s - \sigma\sqrt{3/s}, f/s + \sigma\sqrt{3/s}]\end{aligned}\tag{6}$$

where the second expression for  $z$  applies if it is uniform. SBBO replaces a random SIV in the immigrating island with an SIV from the emigrating island. The probability that this SIV replacement results in the immigrating island's fitness becoming lower than the emigrating island's fitness can be written as

$$\begin{aligned}P_s &= Pr(z - y > f - f_{\min}) \\ &= Pr(y < z - \Delta f)\end{aligned}\tag{7}$$

where  $\Delta f$  is the difference between the total cost of the immigrating and emigrating islands before migration. Figure 4 displays this probability pictorially for two uniform random variables. The probability that the previously-best island is replaced by a new best island due to migration is equal to the probability that a random variable taken from the  $z$  distribution, minus a random variable taken from the  $y$  distribution, is greater than  $f - f_{\min}$ .

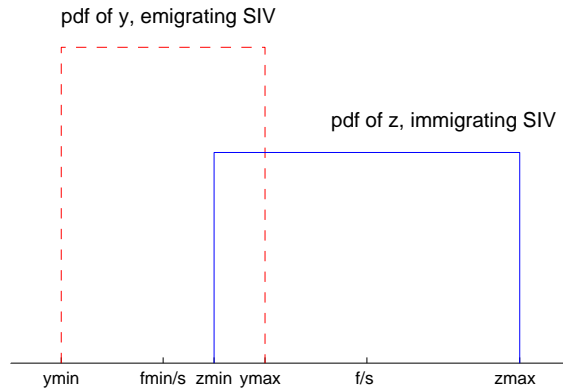


Figure 4: Sample probability distribution functions of an emigrating and immigrating SIV. The probability that the immigrating island is better than the emigrating island after migration is equal to the probability that  $z - y > f - f_{\min}$ .

Equation (7) can be written as

$$P_s = \int_{z_{\min}}^{z_{\max}} \int_{y_{\min}}^{z - \Delta f} \text{pdf}(y) \text{pdf}(z) dy dz \quad (8)$$

This expression applies to a specific immigrating island. If the immigrating island is chosen randomly, then the fitness  $f$  of the immigrating island can range from  $f_{\min}$  to  $f_{\max}$ , and (8) can be written more generally as

$$P_s = \int_{f_{\min}}^{f_{\max}} \int_{z_{\min}}^{z_{\max}} \int_{y_{\min}}^{z - (f - f_{\min})} \text{pdf}(y) \text{pdf}(z) \text{pdf}(f) dy dz df \quad (9)$$

The pdf's in the above equation are entirely problem-dependent. Given a population,  $f_{\min}$  and  $f_{\max}$  are known. The pdf of  $f$  can be approximated by generating a large enough population of solutions. Similarly, the pdf of  $y$  and  $z$  can be approximated by quantifying the cost contribution of a large enough sample of randomly-selected SIVs.

Equation (9) gives the probability that the best individual in a population is replaced after a single generation of the SBBO. This probability is approximately constant from

one generation to the next, assuming that the population does not change much between generations. So the probability that there is no improvement in the global best of the population after each generation is  $(1 - P_s)$ . Therefore, on average we expect the best solution of the population to improve after  $k$  generations, where

$$\begin{aligned} (1 - P_s)^k &= 0.5 \\ k &= \frac{\log(0.5)}{\log(1 - P_s)} \end{aligned} \quad (10)$$

As the SBBO population increases, we have a better chance of beginning the optimization process with a smaller initial cost. But a larger population also typically causes the difference  $f_{\max} - f_{\min}$  to increase, which in turn causes  $P_s$  in (7) to decrease. This causes  $k$ , the expected number of generations between improvements in (10), to increase. We see that a larger population size results in a better initial solution but a longer time between improvements in SBBO. This is illustrated qualitatively in Figure 5.

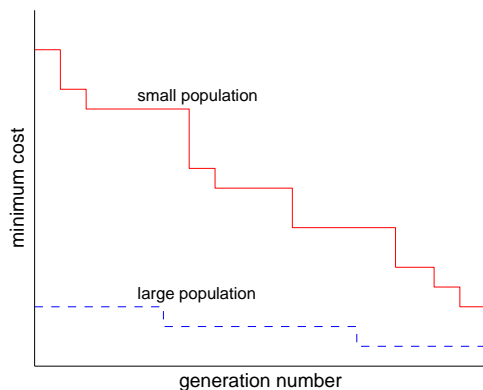


Figure 5: Illustration of the effect of population size on SBBO. A large population has a better initial solution, but more generations between improvements, and smaller improvements. A small population has a poorer initial solution, but fewer generations between improvements, and larger improvements.

Small populations collapse to a single solution more quickly than large populations. This is intuitively obvious and is discussed quantitatively in the next section. As illustrated in Figure 5, improvements are not only more often for smaller populations, but also larger in magnitude. This is explored quantitatively in Section 6.

## 5 Markov Analysis of the Simplified BBO

We define the Markov states of the  $n$ -island SBBO as follows.

- State  $i \in [1, n]$ : The  $i$ th island is the most fit population member and the immigrating island has not yet been selected.
- State  $k \in [n + (i - 1)(n - 1) + j]$  where  $i \in [1, n]$  and  $j \in [1, n - 1]$ : The  $i$ th island is the most fit population member and island  $m_{ij}$  has been selected for immigration, where

$$m_{ij} = \begin{cases} j & j < i \\ j + 1 & j \geq i \end{cases} \quad (11)$$

We see that an  $n$ -member population has  $n^2$  states.

As a simple illustrative example, consider a three-member population. The nine Markov states consist of the following, where we use  $I$  to denote the immigrating island.

1. The first island is the most fit and  $I$  has not yet been selected.
2. The second island is the most fit and  $I$  has not yet been selected.
3. The third island is the most fit and  $I$  has not yet been selected.
4. The first island is the most fit and the second island has been selected for immigration.
5. The first island is the most fit and the third island has been selected for immigration.
6. The second island is the most fit and the first island has been selected for immigration.
7. The second island is the most fit and the third island has been selected for immigration.
8. The third island is the most fit and the first island has been selected for immigration.
9. The third island is the most fit and the second island has been selected for immigration.

The Markov transition probabilities can be summarized as follows.

- If the SBBO is in state  $i \in [1, n]$  then we have an equally likely probability of transitioning to state  $n + i(n - 1) + j$  for each  $j \in [1, n - 1]$ . The probability of transitioning to each of these states is  $1/(n - 1)$ .
- If the SBBO is in state  $n + i(n - 1) + j$  for some  $i \in [1, n]$  and some  $j \in [1, n - 1]$ , then we have a probability  $P_s$  of transitioning to state  $m_{ij}$ , and a probability  $(1 - P_s)$  of transitioning to state  $i$ .  $P_s$  is given in (9) and is the probability that migration to island



state transitions. We use the symbol  $T$  to denote this transition matrix.

$$T = \sum_{j=1}^n P_j L_{n-1,n}^{(j)} \quad (15)$$

**Theorem 1**  $T$  is a regular transition matrix for  $P_s \in (0, 1)$ .

**Proof:** From (15) we see that  $T$  is a sum of  $n$  matrices. If  $P_s \in (0, 1)$ , then each element in the  $j$ th column of the  $j$ th matrix in the sum is greater than zero. So when these matrices are added together, the result is a matrix that contains all nonzero elements.

QED

Theorem 1 can also be proven with a more intuitive argument.  $T_{ij}$  is the probability that island  $j$  is replaced by island  $i$  as the most fit island after one migration. But the probability that island  $j$  migrates to island  $i$  is nonzero,  $1/(n-1)$  to be exact, and the probability that this migration results in island  $i$  replacing island  $j$  as the most fit island is given by the nonzero quantity of (9).

By expanding the sum in (15) it can be seen that

$$T = \frac{1}{n-1} (NI_n + P_s 1_n) \quad (16)$$

where  $N = n - 1 - nP_s$ ,  $I_n$  is the  $n \times n$  identity matrix, and  $1_n$  is the  $n \times n$  matrix that contains all ones.

**Theorem 2**

$$\begin{aligned} T^k &= \frac{1}{(n-1)^k} (NI_n + P_s 1_n)^k \\ &= \frac{1}{(n-1)^k} \left( N^k I_n + P_s \sum_{j=1}^k n^{k-j} N^{j-1} P_s^{k-j} \binom{k}{j-1} 1_n \right) \end{aligned} \quad (17)$$

where  $k$ -choose- $j$  is denoted as

$$\binom{k}{j} = \frac{k!}{j!(k-j)!} \quad (18)$$

Note that computing the first expression above requires about  $\log_2 k$  matrix multiplications, but computing the second expression does not require any matrix multiplications.

**Proof:** It can be seen from (16) that (17) holds for  $k = 1$ . Suppose that (17) holds for some value of  $k$ . Then

$$T^{k+1} = \frac{1}{(n-1)^{k+1}} \left( N^k I_n + P_s \sum_{j=1}^k n^{k-j} N^{j-1} P_s^{k-j} \binom{k}{j-1} 1_n \right) (NI_n + P_s 1_n)$$

$$\begin{aligned}
&= \frac{1}{(n-1)^{k+1}} \left( N^{k+1} I_n + P_s \sum_{j=1}^k n^{k-j} N^j P_s^{k-j} \binom{k}{j-1} 1_n + N^k P_s 1_n + \right. \\
&\quad \left. P_s^2 \sum_{j=1}^k n^{k-j+1} N^{j-1} P_s^{k-j} \binom{k}{j-1} 1_n \right) \\
&= \frac{1}{(n-1)^{k+1}} \left( N^{k+1} I_n + P_s \sum_{j=1}^k n^{k-j} N^j P_s^{k-j} \binom{k}{j-1} 1_n + \right. \\
&\quad \left. P_s \sum_{j=1}^{k+1} n^{k-j+1} N^{j-1} P_s^{k-j+1} \binom{k}{j-1} 1_n \right) \\
&= \frac{1}{(n-1)^{k+1}} \left( N^{k+1} I_n + P_s \sum_{j=1}^k n^{k-j} N^j P_s^{k-j} \binom{k}{j-1} 1_n + \right. \\
&\quad \left. P_s \sum_{j=1}^{k+1} n^{k-j} N^j P_s^{k-j} \binom{k}{j} 1_n + n^k P_s^{k+1} 1_n \right) \\
&= \frac{1}{(n-1)^{k+1}} \left( N^{k+1} I_n + P_s \sum_{j=1}^k n^{k-j} N^j P_s^{k-j} \left[ \binom{k}{j-1} + \binom{k}{j} \right] 1_n + \right. \\
&\quad \left. n^k P_s^{k+1} 1_n \right) \tag{19}
\end{aligned}$$

Recall from [20, p. 69] that

$$\binom{k}{j-1} + \binom{k}{j} = \binom{k+1}{j} \tag{20}$$

Use this in (19) to obtain

$$\begin{aligned}
T^{k+1} &= \frac{1}{(n-1)^{k+1}} \left( N^{k+1} I_n + P_s \sum_{j=1}^k n^{k-j} N^j P_s^{k-j} \binom{k+1}{j} 1_n + n^k P_s^{k+1} 1_n \right) \\
&= \frac{1}{(n-1)^{k+1}} \left( N^{k+1} I_n + P_s \sum_{j=0}^k n^{k-j} N^j P_s^{k-j} \binom{k+1}{j} 1_n \right) \\
&= \frac{1}{(n-1)^{k+1}} \left( N^{k+1} I_n + P_s \sum_{j=1}^{k+1} n^{k+1-j} N^{j-1} P_s^{k+1-j} \binom{k+1}{j-1} 1_n \right) \tag{21}
\end{aligned}$$

So we have proved (17) by induction.

QED

**Theorem 3** *SBBO reaches a population with uniform fitness as the number of generations approaches infinity.*

**Proof:** Equation (17) shows that the diagonal elements of  $T^k$  are all equal, and the off-diagonal elements are all equal. The difference between a diagonal element and an off-diagonal element is  $((n - 1 - nP_s)/(n - 1))^k$ . This difference approaches 0 as  $k \rightarrow \infty$ . We therefore have

$$\lim_{k \rightarrow \infty} T^k = 1_n/n \quad (22)$$

In other words, as the number of generations approaches infinity, each island in the population has an equal probability of being the most fit member of the population. This means that each island is equally fit.

QED

## 6 Expected Improvement in the Simple BBO

Section 4 gave the probability of improvement in one SBBO generation. This section discusses the expected improvement in the best population member given that immigration results in an improvement of the best individual.

Suppose that the best individual has cost  $f_{\min}$ , and it emigrates an SIV to an individual with cost  $f$ . The improvement in the cost of the immigrating individual is denoted as  $\Delta f$ , which we consider to be positive if the cost decreases. Before migration, the replaced SIV in the immigrating island contributed a cost  $z$  to the island as given in (6). The emigrating SIV contributes a cost  $y$  as given in (4). The change in cost  $\Delta f$  due to immigration is therefore a random variable approximately given as

$$\begin{aligned} \Delta f &= z - y \\ &\sim (f/s - f_{\min}/s, 2\sigma^2/s) \\ &\sim (\mu_1, \sigma_1^2) \end{aligned} \quad (23)$$

where  $\mu_1$  and  $\sigma_1$  are auxiliary variables defined by the above equation. We want to compute the expected value of cost improvement, given that improvement occurred due to migration.

This can be written as

$$\begin{aligned} E(\Delta f | f - \Delta f < f_{\min}) &= E(\Delta f | \Delta f > f - f_{\min}) \\ &= \int_{f-f_{\min}}^{f_{\max}} \Delta f \text{pdf}'(\Delta f) d\Delta f \end{aligned} \quad (24)$$

The upper limit in the integral is  $f_{\max}$  because we are assuming that  $f(x) \geq 0$  for all  $x$ , which implies that  $\Delta f \leq f_{\max}$ . The function  $\text{pdf}'(\Delta f)$  is the same as the pdf given in (23), except it is scaled so that it has an area of one between  $f - f_{\min}$  and  $f_{\max}$ .

If (23) is a Gaussian pdf, then (24) can be written as

$$\begin{aligned} E(\Delta f | \Delta f > f - f_{\min}) &= \int_{f-f_{\min}}^{f_{\max}} \frac{x}{c_x \sigma_1 \sqrt{2\pi}} \exp\left(\frac{-(x - \mu_1)^2}{2\sigma_1^2}\right) dx \\ c_x &= \frac{1}{2} \left[ \text{erf}\left(\frac{f_{\max} - \mu_1}{\sigma_1 \sqrt{2}}\right) - \text{erf}\left(\frac{f - f_{\min} - \mu_1}{\sigma_1 \sqrt{2}}\right) \right] \\ \text{erf}(x) &= \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt \end{aligned} \quad (25)$$

Equation (25) is for a specific value of  $f$ . Since the immigrating island is randomly selected, we take the expected value of (25) with respect to  $f$  to obtain

$$E(\Delta f | \Delta f > f - f_{\min}) = \int_{f_{\min}}^{f_{\max}} \int_{f-f_{\min}}^{f_{\max}} \frac{x}{c_x \sigma_1 \sqrt{2\pi}} \exp\left(\frac{-(x - \mu_1)^2}{2\sigma_1^2}\right) dx \text{pdf}'(f) df \quad (26)$$

The function  $\text{pdf}'(f)$  is the pdf of  $f$ , but scaled so that it has an area of one between  $f_{\min}$  and  $f_{\max}$ . If  $f$  has a mean of  $\bar{f}$  and a variance of  $\sigma^2$ , as discussed in the text following (2), and if  $f$  is Gaussian, then (26) can be written as

$$\begin{aligned} E(\Delta f | \Delta f > f - f_{\min}) &= \int_{f_{\min}}^{f_{\max}} \int_{f-f_{\min}}^{f_{\max}} \frac{x}{c_x c_y \sigma_1 \sigma_2 \pi} \exp\left(\frac{-(x - \mu_1)^2}{2\sigma_1^2}\right) \exp\left(\frac{-(f - \bar{f})^2}{2\sigma^2}\right) dx df \\ c_y &= \frac{1}{2} \left[ \text{erf}\left(\frac{f_{\max} - \bar{f}}{\sigma \sqrt{2}}\right) - \text{erf}\left(\frac{f_{\min} - \bar{f}}{\sigma \sqrt{2}}\right) \right] \end{aligned} \quad (27)$$

## 7 Simulation Results

The SBBO algorithm was simulated on some benchmarks in order to check the theoretical results of the preceding sections. We used the Rastrigin, Schwefel 2.26, Sphere, Step, and Griewank (sometimes spelled Griewangk) benchmark functions, each with dimension four. More information about these functions, including their domains, can be found in [11, 21, 22]. Recall that the approximations in the preceding sections assumed that the cost function was separable. The Rastrigin, Schwefel 2.26, Sphere, and Step functions are separable, but the Griewank function is not separable. Given a uniform randomly generated population, the cost functions of these benchmarks are approximately Gaussian.

For each function, we ran 100 Monte Carlo SBBO simulations with a population size of 4, a population size of 16, and a population size of 64. We then obtained the average number of generations that was required to improve the best population member. We also obtained the average improvement of the best population member. We compared these numbers with the expected number of generations before improvement as given in (10), and the expected improvement magnitude as given in (27). Table 1 shows the results. Note that (10) and (27) are calculated on the basis of pdf's,  $f_{\min}$ ,  $f_{\max}$ , and  $\bar{f}$ . These quantities are not available analytically, and so they must be approximated on the basis of random populations. Therefore 10,000 random populations were used to calculate (9), (10), and (27) in Table 1.

Benchmark Function	Population Size	$P_s$	Expected # gens.		Expected improvement	
		Eq. (9)	Eq. (10)	Simulation	Eq. (27)	Simulation
Rastrigin	4	0.102	6.4	6.2	6.7	5.1
	16	0.048	14.0	15.8	5.3	5.0
	64	0.020	34.5	40.1	4.0	3.7
Schwefel 2.26	4	0.121	5.4	7.8	127.2	34.6
	16	0.053	12.7	15.8	108.7	3.9
	64	0.021	32.2	54.3	95.3	0.9
Sphere	4	0.076	8.7	5.2	4.9	4.9
	16	0.037	18.6	16.3	3.3	2.1
	64	0.021	31.9	36.7	0.6	0.8
Step	4	0.076	8.8	4.1	1968.0	1441.4
	16	0.038	18.0	14.8	1286.6	889.0
	64	0.022	30.9	45.2	270.6	259.4
Griewank	4	0.073	9.1	5.7	17.7	10.8
	16	0.037	18.3	17.1	10.8	7.7
	64	0.022	31.5	35.1	1.9	3.6

Table 1: Theoretical and simulation results. Simulation results are based on 100 Monte Carlo runs.

There are several observations that can be made from Table 1. First, for a given benchmark,  $P_s$  as calculated from (9) decreases with the population size. This confirms the intuitive result that the probability per generation of finding a new solution decreases as the population increases. Second, the expected number of generations before finding a new solution as calculated from (10) matches reasonably well with the simulation results. As expected, the number of generations before improvement increases as the population size increases. Third, the expected improvement amount from (27) matches reasonably well with the simulation results. As shown qualitatively in Figure 5, the average improvement amount

decreases as the population size increases.

Although the correlation between the theoretical results and simulation results is clear from Table 1, there are still significant differences between the two sets of results. This is due to the approximations that were used in obtaining (10) and (27). These approximations include the assumption of separability as seen in (2), the assumption of identical pdf's for each SIV as seen in (3), and the assumption of normality in Section 6. However, in spite of these approximations, the simulation and theoretical results in Table 1 are still close enough that the theory can be used to qualitatively predict SBBO performance.

Table 2 shows the number of generations before SBBO reaches a uniform population. Theorem 3 predicted that SBBO would reach a uniform population, and Table 2 confirms this. In agreement with intuition, the number of generations until uniformity is proportional to the population size.

Benchmark Function	Population Size	Number of Generations
Rastrigin	4	36
	16	303
	64	1677
Schwefel 2.26	4	33
	16	311
	64	1671
Sphere	4	35
	16	292
	64	1732
Step	4	35
	16	313
	64	1659
Griewank	4	30
	16	288
	64	1673

Table 2: Number of generations before a uniform population is reached, based on the average of 100 Monte Carlo simulations.

In theory, the number of generations until uniformity is reached could be predicted by Theorem 3. This theorem gives us the ratio of the largest and smallest elements of the Markov transition matrix, so when that ratio falls below some infinitesimal threshold after a certain number of generations, we can assume that the population is uniform and no more improvement is possible. In practice, however, this is difficult to calculate because Theorem 3

assumes that  $P_s$  is constant, whereas  $P_s$  actually changes with each generation. The use of the Markov theory in Section 5 to predict the number of generations until uniformity is reached has been left for future work.

The code that was used to generate these results is available at <http://academic.csuohio.edu/simond/bbo/simplified>.

## 8 Conclusion

A simplified BBO (SBBO) algorithm has been presented in this paper. SBBO has been analyzed using probability theory in order to find three related quantities: the probability per generation that its population optimum improves, the state transition matrix of the algorithm, and the expected amount of improvement in the population optimum. These quantities are inexact in light of approximations that were made, but they match simulation results reasonably well.

Further work could focus on reducing the approximation errors that were used to obtain the theoretical results in this paper. In this paper we assumed normal distributions for the fitness and for the contribution of each parameter to total fitness. But if an empirical distribution were available for a given optimization problem, then the results in this paper could be generalized by fitting a sum of Gaussians to the given distribution [23].

## References

- [1] A. Wallace, *The Geographical Distribution of Animals* (Adamant Media Corporation, 2006), first published in 1876.
- [2] C. Darwin, *The Origin of Species* (Gramercy, 1995), first published in 1859.
- [3] E. Munroe, *The geographical distribution of butterflies in the WestIndies*, PhD Dissertation, Cornell University, Ithaca, New York, 1948.
- [4] E. Munroe, *The size of island faunas*, Seventh Pacific Science Congress of the Pacific Science Association, Volume IV: Zoology, Auckland, New Zealand, pp. 52-53, 1953.
- [5] R. MacArthur and E. Wilson, *An equilibrium theory of insular zoogeography*, *Evolution* (17) pp. 373-387, 1963.
- [6] R. MacArthur and E. Wilson, *The Theory of Biogeography* (Princeton University Press, 1967).
- [7] D. Simon, *Biogeography-based optimization*, *IEEE Transactions on Evolutionary Computation* (12) pp. 702-713, December 2008.
- [8] C. Reeves and J. Rowe, *Genetic Algorithms: Principles and Perspectives* (Kluwer, 2003).
- [9] D. Simon, M. Ergezer, and D. Du, *Markov Analysis of Biogeography-Based Optimization*, submitted for publication, available at <http://academic.csuohio.edu/simond/bbo/markov>.
- [10] M. Ergezer and D. Simon, *Biogeography-based optimization algorithms and performance tradeoffs*, submitted for publication.
- [11] T. Back, *Evolutionary Algorithms in Theory and Practice* (Oxford University Press, 1996).
- [12] D. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning* (Addison-Wesley, 1989).
- [13] M. Dorigo and T. Stutzle, *Ant Colony Optimization* (The MIT Press, 2004).
- [14] M. Dorigo, L. Gambardella, M. Middendorf, and T. Stutzle (eds.), *Special section on ant colony optimization*, *IEEE Transactions on Evolutionary Computation* (6), August 2002.

- [15] R. Eberhart, Y. Shi, and J. Kennedy, *Swarm Intelligence* (Morgan Kaufmann, 2001).
- [16] R. Eberhart and Y. Shi (eds.), Special issue on particle swarm optimization, *IEEE Transactions on Evolutionary Computation* (8), June 2004.
- [17] R. Storn, System design by constraint adaptation and differential evolution, *IEEE Transactions on Evolutionary Computation* (3) pp. 22-34, April 1999.
- [18] W. Khatib and P. Fleming, The stud GA: A mini revolution?, in *Parallel Problem Solving from Nature* (A. Eiben, T. Back, M. Schoenauer, and H. Schwefel, eds.), Springer, 1998, pp. 683-691.
- [19] V. Silva, W. Khatib, and P. Fleming, Performance optimization of gas turbine engine, *Engineering Applications of Artificial Intelligence* (18) pp. 575-583 2005.
- [20] C. Chuan-Chong and K. Khee-Meng, *Principles and Techniques in Combinatorics*, World Scientific, 1992.
- [21] X. Yao, Y. Liu, and G. Lin, Evolutionary programming made faster, *IEEE Transactions on Evolutionary Computation* (3) pp. 82-102, July 1999.
- [22] Z. Cai and Y. Wang, A multiobjective optimization-baed evolutionary algorithm for constrained optimization, *IEEE Transactions on Evolutionary Computation* (10) pp. 658-675, December 2006.
- [23] D. Alspach, Gaussian sum approximations in nonlinear filtering and control, *Information Sciences* (7) pp. 271-290, 1974.