

Punctuated Equilibria in Simple Genetic Algorithms for Functions of Unitation

Sangyeop Oh*

Hyunsoo Yoon†

Department of Computer Science,
Korea Advanced Institute of Science and Technology,
Advanced Information Technology Research Center,
Yeosung-goo, Daejeon, Korea

During a genetic search, the population may get stuck in a local optimum. The population can escape from this after a long duration. This phenomenon is called *punctuated equilibrium*. The punctuated equilibria observed in nature and computational ecosystems are known to be well described by diffusion equations. In this paper, simple genetic algorithms are theoretically analyzed to show that they can also be described by a diffusion equation when fitness is the function of unitation. Using theoretical results on the diffusion equation, the duration of equilibrium is shown to be exponential of such parameters as population size, $1/(\text{mutation probability})$, and potential barrier. This is corroborated by simulation results for one-dimensional bistable potential landscapes with one local optimum and one global optimum.

1. Introduction

Genetic algorithms (GAs) are optimization methods modeled from operations used during natural reproduction and natural selection [1]. Since the original idea was introduced by Holland [2], various GAs have shown practical success in various fields. Among these, the simple genetic algorithm (SGA) is the simplest one containing such essential GA operators as roulette wheel selection, simple mutation, and simple crossover.

Like many other heuristic optimization methods, SGAs have the problem of their system being trapped in a local optimum. Fortunately, there are two kinds of effects contributing to escape from the local optimum, which we will now call the *microscopic effect* and *probabilistic fluctuation*. The microscopic effect is observed when individuals escape through mutations or crossovers. This is related with the population

*Electronic mail address: syoh@camars.kaist.ac.kr.

†Electronic mail address: hyoon@camars.kaist.ac.kr.

variance while the probabilistic fluctuation indicates the fluctuation of the population mean.

When fitness is confined to the function of unitation, bistable fitness functions with one local and one global optima can be categorized into two classes according to the infinite population dynamics [3, 4]. GAs dealing with class I fitness functions converge into the global optimum for any initial population state. However, for class II fitness functions, the GA remains in the local optimum forever unless the initial ratio of the population in the global optimum is larger than a positive criterion. The equilibrium in the local optimum can be disturbed when the population size is finite. Since each GA operator has some probabilistic parameters, the result of GA operations may be different each time if the population size is finite. This induces probabilistic fluctuations and can result in punctuated equilibria.

Punctuated equilibrium is the phenomenon where a system in a metastable state shows a sudden and short transition into a more stable neighboring state after a long period of equilibrium. A *metastable state* is a local optimum beside a better local or global optimum. Punctuated equilibria are observed and analyzed in various fields such as computational ecosystems (CEs) [5], neodarwinian evolution models [6], and GAs [7]. All of these systems share the common feature of having a population of individuals driven on a fitness or potential landscape.

If there are only two types of individuals, the system state can be represented by the ratio z of a particular type of individuals. In this case, the dynamics of the CE is governed by a diffusion equation

$$dz(t) = -dtF'(z) + \varepsilon dB(t) \quad (1)$$

where F is the potential landscape, $B(t)$ the standard brownian process, and ε a small constant [5]. Using the mathematical results on the diffusion equation [8], the existence of punctuated equilibria and the duration of the metastable state can be explained. The duration is exponential in the population size and the height of the potential barrier between the metastable and the stable states.

For the neodarwinian model, the population mean \bar{x} of the individual character x is governed by equation (1) if z is replaced by \bar{x} and $(-F)$ by the fitness of \bar{x} . Thus the neodarwinian model explains the punctuated equilibrium and the exponential duration observed in natural evolution [6].

For GAs, Vose tried to explain the punctuated equilibria observed in genetic searches by showing that the local optimum state is unstable unless it has globally maximal fitness [7]. But such a quantitative feature as the duration of metastability was not considered. On the other hand, a phenomenon similar to punctuated equilibrium was analyzed for the Royal Road fitness function [9]. But the Royal Road fitness is just like a step function and hence the phenomenon, "stasis followed

by jump,” originated not from a local optimum surrounded by fitness barriers but from a long fitness plateau. Thus, the duration of the plateau is proportional to $1/(\text{population size})$ rather than exponential. The dynamics of a GA for class I fitness functions was also studied in [10, 11]. The relation between their results and ours is discussed in detail in section 5.3.

In this paper, we focus on the population mean of a phenotype as is done in the neodarwinian model since there are more than two types of individuals in general GAs. We adopt the central limit theorem [12] to show that the dynamics of the SGA can be represented by an equation whose form is similar to equation (1). Using class II functions of unitation for fitness, we can obtain more details of the equation. This enables us to use the mathematical results of equation (1) to obtain the features of the punctuated equilibria. Although the functions of unitation represent only a small class of all possible ones, they have been of considerable interest to GA researchers, since they are easy to analyze and understand [4, 13]. The theoretical results are compared with the simulation results obtained for the one-dimensional bistable potential landscapes.

In section 2, as background to the analysis, equation (1) and the CE are more precisely described. We analyze the dynamics of the SGA in section 3. The bistable fitness landscape is introduced and the simulation results are shown in section 4. In section 5 we discuss the results obtained in the previous sections focusing on the duration of metastability. The conclusion and further work are covered in section 6.

2. Background

2.1 Diffusion processes

Consider the one-dimensional diffusion process $z(t)$ satisfying equation (1) and let the bistable landscape F satisfy the following conditions.

- F is a differentiable function defined on $-\infty < z < \infty$.
- There exists $z_1 < z_2 < z_3$ such that F is strictly decreasing on $(-\infty, z_1] \cup [z_2, z_3]$ and strictly increasing on $[z_1, z_2] \cup [z_3, \infty)$.
- $F(z_1) > F(z_3)$.

Then z_1 is the metastable state and z_3 is the stable state, and between them there is a barrier at z_2 . Let $D = F(z_2) - F(z_1)$ be the height of the barrier.

According to mathematical studies of the dynamics of these bistable systems [8], if the initial state of the system is around z_1 , then punctuated

equilibrium can be observed and the duration T of metastability satisfies

$$T \propto \exp\left(\frac{2D}{\varepsilon^2}\right) \quad (2)$$

as $\varepsilon \rightarrow 0$. Also, the transition is unidirectional in the sense that the system remains in the stable state as $t \rightarrow \infty$ if there are $z_0 < z_1$ and $z_4 > z_3$ satisfying that $F(z_0), F(z_4) > F(z_2)$.

The aim of this paper is to explain the dependency of ε on SGA parameters when the SGA dynamics is described by equation (1). This will elucidate the relations between the duration T and the SGA parameters.

2.2 Computational ecosystems

The definitions of system state and state transition rule for CEs are similar to those of the GAs. In the CE [14], a population consists of N agents, and each agent chooses one of R possible strategies to get some payment. The payment function f is dependent on the chosen strategy and the population state. A population state is represented by a vector $\mathbf{n} = (n_1, n_2, \dots, n_R)$ or $\mathbf{r} = (r_1, r_2, \dots, r_R)$ where n_i is the number of agents using the strategy i and $\mathbf{r} = \mathbf{n}/N$. During a unit time, each agent has α chances on average to change its strategy to a new one according to $\rho_i(\mathbf{r})$, which is the probability that strategy i is perceived to be the best choice.

The system state is described by the population state. Let $P(\mathbf{n}, t)$ be the probability that the population state is \mathbf{n} at time t . Then the function P at a specific t describes a particular ensemble which contains all possible system states at time t . Consider a sufficiently short time interval Δt during which only a single change contributes. That is, either \mathbf{n} is invariant or else there is a single change from some strategy j to strategy i . Then, the change of the probability is

$$\frac{P(\mathbf{n}, t + \Delta t) - P(\mathbf{n}, t)}{\alpha \Delta t} = -P(\mathbf{n}, t) \sum_{i \neq j} n_i \rho_i + \sum_{i \neq j} P(\mathbf{n}^{[j,i]}, t) (n_j + 1) \rho_j^{[j,i]} \quad (3)$$

where $\mathbf{n}^{[j,i]}$ satisfies that $n_j^{[j,i]} = n_j + 1$, $n_i^{[j,i]} = n_i - 1$, and the other elements are the same as those of \mathbf{n} [14]. Also, ρ_i and $\rho_j^{[j,i]}$ are evaluated at \mathbf{n} and $\mathbf{n}^{[j,i]}$, respectively.

The ensemble mean $\langle r_i \rangle$ of the ratio r_i is represented as

$$\langle r_i \rangle = \sum_{\mathbf{n}} r_i(\mathbf{n}, t) P(\mathbf{n}, t).$$

The derivative of $\langle r_i \rangle$ is, by applying $\Delta t \rightarrow 0$ to equation (3) [14],

$$\frac{d\langle r_i \rangle}{\alpha dt} = \langle \rho_i \rangle - \langle r_i \rangle. \quad (4)$$

The detail of the derivation is shown in appendix A. Equation (4) means that the ratio r_i has a tendency to approach the conceived value ρ_i on average. This result will be used to outline the form of the potential function F in section 3.4.

3. Analysis

In this paper, the mean and variance for a random variable Z are represented as $E(Z)$ and $V(Z)$, respectively. However, to distinguish them from each other, the population mean is represented as \bar{Z} while the ensemble mean of the population character is represented as $\langle Z \rangle$.

3.1 Simple genetic algorithms

An SGA deals with a population which consists of N individuals. Each individual is a binary string of L bits and each bit has one of two values, 0 and 1; hence, there are $R = 2^L$ genotypes. The phenotype for the genotype i is x_i and the fitness f is a function on the phenotype domain. The population state is represented by $\mathbf{n} = (n_1, n_2, \dots, n_R)$ or $\mathbf{r} = (r_1, r_2, \dots, r_R)$ where n_i is the number of individuals with the genotype i and $\mathbf{r} = \mathbf{n}/N$.

The population of the next generation is produced from the current one through the following SGA operators: roulette wheel selection, simple mutation, and simple crossover [1]. After two individuals are selected from the current population by roulette wheel selections, simple mutation toggles each bit of the individuals with the probability p_m . Simple crossover is then applied to the pair with the probability p_c . In simple crossover, each individual is cut at the same point and divided into two substrings, and the second substrings are then exchanged. The cutting point is selected uniformly at random among all possible L points. After crossover, one of two children is chosen randomly and inserted into the temporary population. The temporary population is made as a copy of the old population when the new generation begins. The individual to be deleted from the temporary population is chosen uniformly at random without replacement, and is replaced by the child individual produced. Repeating this process N times, a new generation with generation gap 1 [1] is completely produced. Each child's production is independent of the others' while the replacement is dependent.

3.2 Ensemble mean part and brownian part

As defined for CEs, let $P(\mathbf{r}, t)$ be the probability that the population state is \mathbf{r} at time t for SGAs. Then P at a specific t describes a particular ensemble which contains possible system states at time t . The ensemble mean of a random variable Z is $\langle Z \rangle = \sum_{\mathbf{r}} Z_{\mathbf{r}} P(\mathbf{r}, t)$ where Z represents

a feature of the population. And let α be the average number of generations per unit time. Then the SGA changes the population state $\alpha\Delta t$ times in time interval Δt . We take sufficiently small Δt so that the $\alpha\Delta t$ is 0 or 1. Note that a CE corresponds to a generation gap $1/N$.

Since SGA operations contain random events, the phenotype of a child can be considered as a random variable X . Each production of a child is independent of the others since the generation gap is 1 and the population can be considered as a sample of size N for X . Hence, we can use the central limit theorem with the result that the population mean \bar{X} has a gaussian distribution with mean $E(X)$ and variance of approximately $V(X)/N$ [12]. If the SGA starts at time t_0 , \bar{X} changes into

$$\bar{X}(t) = \langle \bar{X}(t) \rangle + \alpha(t - t_0)G(t) \quad (5)$$

where $G(t)$ is the gaussian random variable with mean 0 and variance $V(\bar{X}(t)) = V(X)/N$. Hence $\Delta\bar{X}$ is obtained by

$$\bar{X}(t + \Delta t) - \bar{X}(t) = \langle \bar{X}(t + \Delta t) \rangle - \langle \bar{X}(t) \rangle + \alpha\Delta t\hat{G}(t) \quad (6)$$

where $\Delta t\hat{G}(t) = (t + \Delta t - t_0)G(t + \Delta t) - (t - t_0)G(t)$. Being a linear summation of gaussians, $\hat{G}(t)$ is also a gaussian with mean 0 and variance $\hat{V}(\bar{X}(t))$ which has a value between $V(\bar{X}(t))$ and $V(\bar{X}(t + \Delta t))$.

The accumulation of gaussian random variables, each of which has the variance 1, results in a standard brownian process $B(t)$ provided that the increments are mutually independent. The randomness of \hat{G} originates from the random events contained in SGA operations. These random events are mutually independent among successive generations and, hence, the \hat{G} of these generations are also mutually independent though the $\hat{V}(\bar{X})$ have specific relations. Since $\hat{G}/\sqrt{\hat{V}(\bar{X})}$ is a gaussian with mean 0 and variance 1, the change in B during Δt is $\Delta B = (\alpha\Delta t)\hat{G}/\sqrt{\hat{V}(\bar{X})}$.

On the other hand, to obtain the contribution of the term $\langle \Delta\bar{X}(t) \rangle = \langle \bar{X}(t + \Delta t) \rangle - \langle \bar{X}(t) \rangle$, we first find the function $a(\bar{X})$ which is defined by $a(\bar{X}(t_0)) = \langle \Delta\bar{X}(t_0) \rangle / \Delta t$ as $\Delta t \rightarrow 0$. Then, we replace $\bar{X}(t_0)$ with $\bar{X}(t)$. This method is valid if the following *conditions of the analysis* are satisfied [15].

- The fitness landscape should be smooth enough over \bar{X} .
- The “jump” of \bar{X} between two adjacent generations should be sufficiently small.

If these conditions are satisfied, equation (6) also becomes the stochastic differential equation

$$\frac{d\bar{X}(t)}{dt} = a(\bar{X}(t)) + \sqrt{\frac{V(X)}{N}} \frac{dB(t)}{dt} \quad (7)$$

as $\Delta t \rightarrow 0$ [15], where $\hat{V}(\bar{X}) = V(\bar{X}) = V(X)/N$ is used. If the potential function, which we will now call F , can be determined by $F'(\bar{X}) = -a(\bar{X})$, finally equation (7) becomes

$$\frac{d\bar{X}}{dt} = -F'(\bar{X}) + \sqrt{\frac{V(X)}{N}} \frac{dB}{dt} \tag{8}$$

which has the same form as equation (1).

In this derivation, instances of X were thought to be grouped into populations each of which corresponds to a population state \mathbf{r} . We can obtain the population mean from the population state as $\bar{X} = \bar{x}(\mathbf{r})$ where \bar{x} is defined by $\bar{x}(\mathbf{r}) = \sum_i x_i r_i$ and r_i is considered a random variable. Since $E(X^k) = \langle X^k \rangle$ and $\langle x^k(\mathbf{r}) \rangle = x^k(\langle \mathbf{r} \rangle)$ for an integer k ,

$$E(X) = \langle \bar{X} \rangle = \bar{x}(\langle \mathbf{r} \rangle) \tag{9}$$

and

$$V(X) = E(X^2) - E^2(X) = s_x^2(\langle \mathbf{r} \rangle) \tag{10}$$

where $s_x^2(\mathbf{r}) = \overline{x^2}(\mathbf{r}) - \bar{x}^2(\mathbf{r})$ is the population variance of the phenotype for the state \mathbf{r} . These equivalences are used in the following analysis.

■ **3.3 Functions of unitation and effects of simple genetic algorithm operators**

To apply the theory on the diffusion process of section 2.1 to equation (8), $V(X)$ needs to be constant during the running of an SGA. At worst, the duration T could be bounded by exponential boundary values if $V(X)$ is bounded in a finite range. In this section, the effects of the SGA operators are separately addressed focusing on $V(X)$. These are used to obtain the more useful result in section 3.5.

Now let the fitness f be the function of unitation. The phenotype of an individual is defined by $x_i = l_i/L$ where l_i is the number of bits with the value 1 in the genotype i . The definition of phenotype enables us to get more details on $V(X)$. In this section, we consider X in a discrete time domain. Let X at generation $(\tau + 1)$ be composed of mutually independent random variables $X_s(\tau)$, $X_m(\tau)$, and $X_c(\tau)$; $X(\tau + 1) = X_s(\tau) + X_m(\tau) + X_c(\tau)$ where X_s is the result of the selection, and X_m and X_c are the changes due to mutation and crossover, respectively. This is possible because the SGA operators can be collected and run separately for one generation. We can then consider the effects of the SGA operators, one by one.

For the selection, suppose that the population at generation τ is uniformly distributed on the slope of $f(x) = gx$ so that the individuals have fitness values of $H_0, H_0 + b, H_0 + 2b, \dots, H_0 + (N - 1)b = H_0 + H_1$. In this case, $f_i = H_0 + ib$, $x_i = f_i/g$, and $r_i = 1/N$. We can find $E(X)$ and

$E(X^2)$ through simple calculation. The variance $V(X(\tau))$ is then

$$V(X(\tau)) = \frac{(N^2 - 1)b^2}{12g^2} \approx \frac{H_1^2}{12g^2} \tag{11}$$

for $N \gg 1$. The moments after the selection, $E(X_s)$ and $E(X_s^2)$, can be calculated replacing r_i by $\rho_i = r_i f_i / \bar{f}$ where $\bar{f} = H_0 + (N - 1)b/2$. Finally, the variance after roulette wheel selection is, for $N \gg 1$,

$$\begin{aligned} V(X(\tau)_s) &= \frac{(N^2 - 1)b^2}{12g^2} \left[1 - \frac{(N^2 - 1)b^2}{3(2H_0 + (N - 1)b)^2} \right] \\ &\approx \frac{H_1^2}{12g^2} \left[1 - \frac{H_1^2}{3(2H_0 + H_1)^2} \right]. \end{aligned} \tag{12}$$

The above assumptions are not satisfiable, in general, and this fact will be considered when equations (11) and (12) are used in section 3.5. The mean $E(X_s)$ is addressed to obtain the potential function in section 3.4.

Next, simple mutation is considered. When one bit is mutated, the change (e.g., -1 for $1 \Rightarrow 0$) has the probability distribution with mean $p_m(1 - 2l(t)/L)$ and variance $p_m - p_m^2(1 - 2l(t)/L)^2$. Hence, by the central limit theorem, X_m has approximately the gaussian distribution with

$$E(X_m) = p_m(1 - 2l(t)/L) = p_m(1 - 2E(X_s)) \tag{13}$$

and

$$V(X_m) = [p_m - p_m^2(1 - 2l(t)/L)^2]/L \approx p_m/L \tag{14}$$

for $p_m \ll 1$ and $L \gg 1$.

Finally, we consider the effect of crossovers. Simple crossover does not change the number of 1s within two parents. Thus \bar{X} , and furthermore $E(X)$ by equation (9), are independent of crossover. However, the evaluation of the contribution to the variance $V(X)$ is more complicated. Suppose that the genotypes of the parents are i and j and the substrings subdivided by the crossover in an individual have lengths u and $L - u$, respectively. Here u represents a random variable which has a uniform distribution with the probability $P(u) = P(L - u) = 1/L$. Moreover, let $y_{i,1}$ and $y_{i,2}$ be the number of 1s in the first and second substrings of genotype i , respectively. Then $y_{i,2} = l_i - y_{i,1}$, and $y_{i,1}$ can be considered as a random variable. The probability of $P(y_{i,1})$ is

$$P(y_{i,1}) = \binom{u}{y_{i,1}} \binom{L - u}{l_i - y_{i,1}} \binom{L}{l_i}^{-1} \tag{15}$$

According to the binomial theorem $(1 + z)^u = \sum_y \binom{u}{y} z^y$ [16], and we can obtain

$$\sum_{y_{i,1}} y_{i,1} P(y_{i,1}) = u \binom{L - 1}{l_i - 1} \binom{L}{l_i}^{-1} = \frac{ul_i}{L} = ux_i. \tag{16}$$

Next, in the same way, we can obtain

$$\begin{aligned} \sum_{y_{i,1}} y_{i,1}^2 P(y_{i,1}) &= \left[u \binom{L-2}{l_i-1} + u^2 \binom{L-2}{l_i-2} \right] \binom{L}{l_i}^{-1} \\ &= \frac{(uL - u^2)x_i + (u^2 - u)Lx_i^2}{L-1}. \end{aligned} \tag{17}$$

The detail of the derivation can be found in appendix B.

The phenotype of a child is determined by a random variable $x' = (y_{i,1} + y_{j,2})/L$. The phenotype of the other child has the same probability distribution as x' , since the function of unitation is invariant if the order is reversed for a bit string. For simplicity, we now replace $y_{i,1}$ and $y_{j,2}$ with y_1 and y_2 . Because the random variables u , x_i , x_j , y_1 , and y_2 are mutually independent, the mean value of $(y_1 + y_2)$ is determined by

$$\begin{aligned} &\sum_{y_1, x_i, y_2, x_j, u} P(u)P(x_i)P(y_1)P(x_j)P(y_2)(y_1 + y_2) \\ &= \sum_{x_i, u} P(u)P(x_i)ux_i + \sum_{x_j, u} P(u)P(x_j)(L-u)x_j \\ &= \sum_u P(u)uE(X) + \sum_u P(u)(L-u)E(X) \\ &= LE(X). \end{aligned} \tag{18}$$

And the mean of $(y_1 + y_2)^2$ is

$$\begin{aligned} &\sum_{y_1, x_i, y_2, x_j, u} P(u)P(x_i)P(y_1)P(x_j)P(y_2)(y_1 + y_2)^2 \\ &= \frac{(L+1)E(X)}{3} + \frac{L(2L-1)E(X^2)}{3} + \frac{(L^2-1)E^2(X)}{3}. \end{aligned} \tag{19}$$

Hence, the variance of x' is

$$\begin{aligned} &\frac{1}{L^2} \left[\frac{(L+1)E(X)}{3} + \frac{L(2L-1)E(X^2)}{3} + \frac{(L^2-1)E^2(X)}{3} - L^2E^2(X) \right] \\ &= \frac{(L+1)E(X)(1-E(X))}{3L^2} + \frac{(2L-1)}{3L} V(X) \\ &\simeq \frac{E(X)(1-E(X))}{3L} + \frac{2}{3} V(X) \end{aligned} \tag{20}$$

when $L \gg 1$. Finally, the variance $V(X_c)$ is

$$V(X_c) = \frac{E(X_s + X_m)(1 - E(X_s + X_m))}{3L} - \frac{V(X_s + X_m)}{3}. \tag{21}$$

For every u , the values of $y_{i,1}/u$ and $y_{j,2}/(L-u)$ may be different from x_i and x_j as the 1s can be gathered into one substring. The first term of

the right-hand side of equation (21) represents these probabilistic fluctuations and has the same form as the variance of a binomial distribution. If the binomial fluctuations are small enough, the second term causes $V(X)$ to decrease.

■ 3.4 Potential function

The precise form of the potential function is helpful in designing the simulation and in understanding the population dynamics. Consider the process where a child individual is produced by SGA operations and inserted into the population replacing an old one. When genotypes of the child and the old individual are i and j , respectively, this replacement corresponds to the changing from strategy j to i for an agent in CEs. The fitness function also corresponds to the payment function of the CE even though fitness is independent of the population state, in general. If ρ_i is interpreted as the probability that a child with the genotype i may be produced, the SGA with the generation gap $1/N$ is a special case of a CE. Even when the generation gap is 1, if \bar{X} does not change considerably for a generation, we can use equation (4) without change since α is the average number of replacements for an individual, as it is for the CE. Hence, from equation (4),

$$\begin{aligned} \frac{d\langle \bar{X}(t) \rangle}{\alpha dt} &= \sum_{i=1}^R x_i (\langle \rho_i(\mathbf{r}(t)) \rangle - \langle r_i(t) \rangle) \\ &= E(X_s + X_m + X_c) - E(X) \end{aligned} \tag{22}$$

using $\sum_{i=1}^R x_i \langle r_i(t) \rangle = \langle \bar{X} \rangle = E(X)$.

Considering only roulette wheel selection, $\rho_i(\mathbf{r}(t)) = f(x_i)r_i/\bar{f}$ where $\bar{f} = \sum_{i=1}^R f(x_i)r_i$. Whether or not a particular individual has the genotype i is a random event of a Bernoulli distribution and individual selections are independent of each other since the generation gap is 1. The number n_i then has a binomial distribution and, for a sufficiently large N , the ratio $r_i = n_i/N$ can be approximated by a gaussian distribution whose variance is proportional to $1/N$. This implies that when the population size N is large enough, r_i is contained in a narrow band around $\langle r_i \rangle$; hence, we can adopt mean field approximation. Mean field approximation works very well for a situation where each individual interacts with each of the other individuals through global parameters and the population size is large [17]. Hence, the selection part of equation (22) is

$$E(X_s) - E(X) = \sum_{i=1}^R x_i \left(\frac{f(x_i)\langle r_i \rangle}{\bar{f}\langle \mathbf{r} \rangle} - \langle r_i \rangle \right) = \frac{(\overline{x - \bar{x}})f}{\bar{f}} \tag{23}$$

where the population means are taken over $\langle \mathbf{r} \rangle$. When the conditions of section 3.2 are satisfied and $V(X)$ is considerably small, f can be linearly

expanded as follows:

$$f(x) \approx f(\bar{x}) + (x - \bar{x})f'(\bar{x}). \tag{24}$$

Replacing f in equation (23) with equation (24),

$$\sum_{i=1}^R x_i \left(\frac{f(x_i)\langle r_i \rangle}{\bar{f}(\langle r \rangle)} - \langle r_i \rangle \right) \approx s_X(\langle r \rangle) \frac{f'(\bar{x}(\langle r \rangle))}{\bar{f}(\bar{x}(\langle r \rangle))} \tag{25}$$

where $\bar{f} = f(\bar{x})$ by equation (24).

Mutation and crossover, however, are highly dependent on the definitions of the phenotype and fitness function. We use the fitness function of unitation; hence, the effect of crossovers to $\langle \bar{X} \rangle$ is zero, as mentioned in section 3.3. The effect of mutations to $\langle \bar{X} \rangle$ can be included in equation (22) using equation (13), which results in

$$\begin{aligned} \frac{d\langle \bar{X}(t) \rangle}{\alpha dt} &= (E(X_s) - E(X)) + E(X_m) \\ &\approx s_X(\langle r \rangle) \frac{f'(\bar{x}(\langle r \rangle))}{\bar{f}(\bar{x}(\langle r \rangle))} + p_m(1 - 2\bar{x}(\langle r \rangle)) \end{aligned} \tag{26}$$

where $E(X_s)$ is replaced by $E(X) = \bar{x}(\langle r \rangle)$ for $p_m \ll 1$.

If $t = t_0$ then $\bar{x}(\langle r \rangle) = \langle \bar{X} \rangle = \bar{X}$ resulting in

$$a\langle \bar{X}(t_0) \rangle = \frac{d\langle \bar{X}(t) \rangle}{dt} = \alpha \left[V(X) \frac{f'(\bar{X}(t_0))}{\bar{f}(\bar{X}(t_0))} + p_m(1 - 2\bar{X}(t_0)) \right] \tag{27}$$

where $s_X(\langle r \rangle)$ is replaced by $V(X(t_0))$ using equation (10). Assuming that $V(X)$ is independent of \bar{X} , we obtain the outline of F

$$F(\bar{X}) \approx -\alpha [V(X) \log f(\bar{X}) + p_m(\bar{X} - \bar{X}^2)] \tag{28}$$

using the definition $F'(\bar{X}) = -a(\bar{X})$.

■ 3.5 Ensemble variance

Boundedness and the general features of ensemble variance $V(X)$ are addressed in this section. If the population approaches the potential barrier, H_0 of equation (12) decreases. This implies that $(V(X) - V(X_s))$ increases due to equations (11) and (12) while $V(X_m)$ does not change. The result is, therefore, a smaller $V(X)$ in the next generation. Thus the ensemble variance is bounded by V_e :

$$V(X) < V_e \tag{29}$$

for $V(X)$ in the basin of attraction of the metastable state.

We now find the ensemble variance V_e in the metastable state, temporarily ignoring crossover. While in the metastable state, $V(X)$ is approximately constant over time t and the decrement due to selections

and the increment due to mutations cancel out; $V(X) - V(X_s) = V(X_m)$ where $V(X) = V_e$. Notice that the decrement due to selection is independent of N and p_m , as mentioned in the simplified calculation of section 3.3. Hence, from equations (11), (12), and (14), we can obtain

$$V_e = \frac{H_1^2}{12g^2} = \frac{3(2H_0 + H_1) p_m}{H_1^2} \frac{1}{L}. \quad (30)$$

The linear approximation of the fitness function in section 3.3 is acceptable when the conditions of section 3.2 are satisfied. However, we cannot assert that equation (30) is exactly correct, since the assumption of a population having a uniform distribution over a linear fitness slope is not justifiable. What we can learn from equation (30) are just a few generalizable features of V_e . Even in general cases, the distribution of a population can be approximated by an accumulation of mutually independent uniform distributions each of which has its own H_0 and H_1 . Hence, the independence of N and the proportionality to p_m remain unaffected for various distributions of populations. Finally, we obtain the *general features of V_e without crossover* as follows.

1. V_e is independent of N .
2. $V_e \propto p_m$.

The general features remain nearly unaffected even when we consider crossover. Once $V(X)$ and $E(X)$ are determined, the change of $V(X)$ due to crossover is independent of N , as shown in equation (21). Hence the first feature does not change. The second feature needs to be modified a little since it is influenced by equation (21). The value of $E(X)$ in the metastable state, which we will now call E_e , is determined by $a(E_e) = 0$ from equation (27). However, it is very complicated to calculate E_e exactly since V_e is a function of p_m and E_e . When $p_m = 0$, E_e is determined by $f'(E_e) = 0$ for which the fitness is maximal. If $p_m > 0$, E_e moves toward $1/2$ due to the mutation. Provided the domain of p_m is small enough, E_e can be approximated by a linear function of p_m . This is not an unusual assumption because p_m itself is very small, in general. Furthermore, if fitness is maximal at $x \approx 0$, E_e is proportional to p_m and, also, so small that $E_e(1 - E_e) \approx E_e$. The terms of $V(X_c)$ are then proportional to either p_m or $V(X_s)$, where $V(X_s)$ is also proportional to V_e . Hence, the *general features of V_e with crossover* are modified as follows.

1. V_e is independent of N .
2. $V_e \propto p_m$ for the small domain of p_m , when the fitness is maximal at $x \approx 0$.

Consider again now the process where the next generation is produced by applying SGA operators separately: $V(X(\tau)) \Rightarrow V(X_s(\tau)) \Rightarrow$

$V(X_s(\tau) + X_m(\tau)) \Rightarrow V(X_s(\tau) + X_m(\tau) + X_c(\tau)) = V(X(\tau + 1))$. As mentioned in section 3.3, every step of this process is independent of N . Hence $V(X(\tau + 1))$ is independent of N if $V(X(\tau))$ is independent of N . Furthermore, suppose that $V(X(\tau)) \propto p_m$. The analyses of the SGA operators in section 3.3 then result in $V(X(\tau + 1)) \propto p_m$.

The metastable state is frequently reached by the population until the transition to the stable state occurs. Thus, the values of $V(X)$ are independent of N and proportional to p_m in the basin of attraction of the metastable state. Let V_* be the representative value of the ensemble variances until the transition occurs such that

$$T \propto \exp\left(\frac{2DN}{V_*}\right). \quad (31)$$

This equation originates from equation (2) where $\varepsilon = \sqrt{V(X)/N}$ by equations (1) and (8). Then $V_* = K_m p_m$ where K_m is a function of δ . This δ is a measure of the depth of the fitness barrier and defined in section 4. Finally, the duration is

$$T \propto \exp\left(\frac{2DN}{K_m p_m}\right). \quad (32)$$

Note that D is also a function of δ .

4. Simulation

4.1 Bistable landscapes

Punctuated equilibria can be observed if F has a landscape which satisfies the conditions of section 2.1. In addition, if f satisfies the conditions of section 3.2, the theoretical analysis of the previous section can be applied. Considering these conditions, the fitness function is defined by

$$f(x) = \begin{cases} 1 - 3\delta x & \text{if } 0 \leq x < 1/3 \\ 1 - 2\delta + 3\delta x & \text{else if } 1/3 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (33)$$

where δ is the depth of the barrier in the f landscape. Hence f consists of two linear slopes whose gradients have the same absolute values.

In the simulation, we use several guidelines for the domain of the SGA parameters as follows. First, the length L of an individual string should be large enough so that f can be approximated to be continuous and that the central limit theorem can be applied to obtain equations (13) and (14). Second, the mutation probability p_m must satisfy $p_m \ll 1$ since approximations to the first order of p_m are used in the analysis. When this condition is satisfied, the domain of p_m is also so small that it can justify the second general feature of V_e with the crossover in section 3.5.

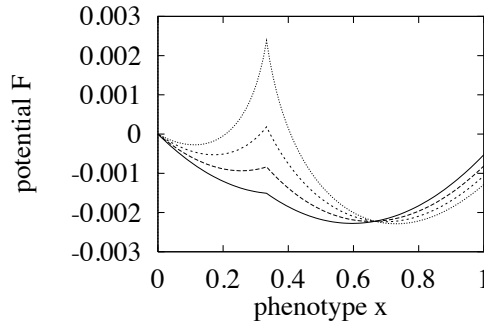


Figure 1. Potential functions obtained from equation (28) for various fitness barrier depths. The unit of the y-axis is α . From down to up, $\delta = 0.3, 0.5, 0.7,$ and 0.9 respectively. The other parameters are set by $s_x^2 = 0.002$ and $p_m = 0.01$.

Third, the population size N should be large enough to apply the central limit theorem in section 3.2. We take samples of size $N \gtrsim 30$.

Typical potential landscapes are shown in Figure 1. Notice that there is only one minimum in the potential landscape when δ is small. For the analysis to be applicable, the potential landscape needs to be bistable. Thus we choose δ within $(0.4, 1]$ for the simulation.

4.2 Punctuated equilibria

Figure 2 shows typical punctuated equilibria observed in the running of the SGA with $p_c = 0$ and $p_c = 1$ respectively. The population starts from the state $\bar{x} = 0$, converges quickly into the metastable state, and shows perturbations around it. After a long duration, \bar{x} suddenly transits the potential barrier and then fluctuates around the stable state.

Beginning with the population state of $\bar{x} = 0$, we record the duration T of metastability *versus* some parameters, where T is defined as the number of generations until the transition occurs. The parameters considered are the population size N , the mutation probability p_m , and the fitness barrier depth δ . Figures 3(a), (b), and (d) show that T is a rapidly increasing function of N and δ . Figure 3(c) shows that T is a rapidly decreasing function of p_m . Figure 3 also shows that crossover makes the duration longer. The interpretation of Figure 3 is discussed in detail in section 5.1.

5. Discussion

5.1 Duration of metastability

Figures 3(a) and (b) support the theoretical result that T is exponential of the population size N by the linear feature of the log-scale graphs for $p_c = 1$. However, exponentiality is weakened for $p_c = 0$ bending down

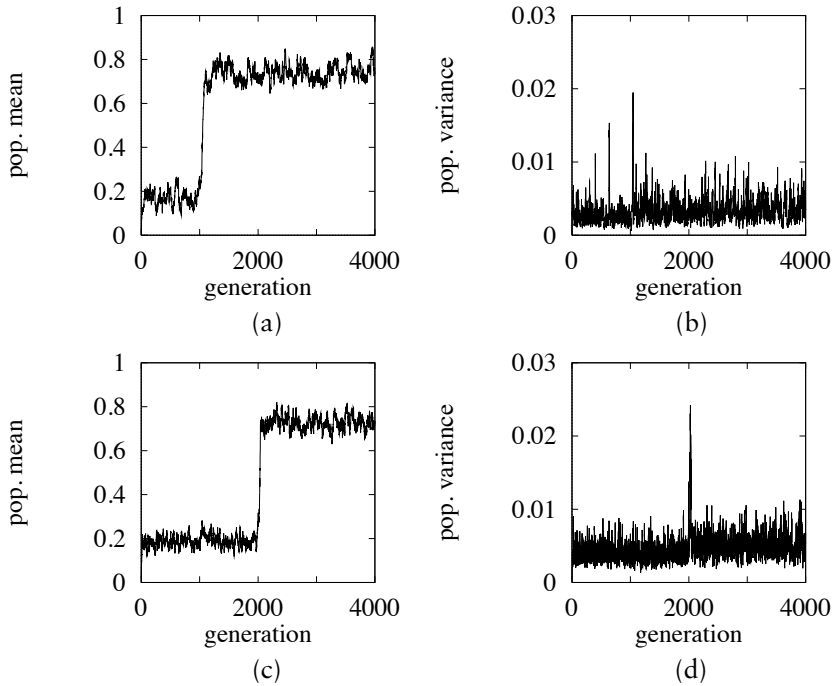


Figure 2. Punctuated equilibria on the graph of population mean \bar{x} of phenotype. Graphs (a) and (b) represent \bar{x} and s_x^2 respectively, when $L = 30$, $N = 40$, $p_c = 0$, $p_m = 0.013$, and $\delta = 0.7$. Graphs (c) and (d) represent \bar{x} and s_x^2 respectively, when $L = 30$, $N = 40$, $p_c = 1.0$, $p_m = 0.02$, and $\delta = 0.7$.

the graph in the domain of $N \gtrsim 60$. The theoretical analysis of this paper assumes that the population can be considered macroscopically as a rigid ball rolling on a solution space. This assumption breaks down when the population size is so large that it is comparable to the size of the basin of attraction of the metastable state, which is $1024 = 2^{L/3}$ for the simulation. For a large N , there may appear more individuals in the stable area by mutation thus accelerating the transition. This is the microscopic effect of the SGA operators. However, when $p_c = 1$, the individuals in the stable area are immediately eliminated by crossover, as discussed in section 5.2 and, hence, the microscopic effect is considerably weakened. Hence, exponentiality can be more explicitly observed for $p_c = 1$ even when N is considerably large.

For the mutation probability p_m , the linear feature of the log-scale graph of Figure 3(c) corresponds with equation (32). However, linearity is a little disturbed when p_m is large.

For Figure 3(d), equation (32) cannot be directly used since the barrier depth δ is the quantity of the f landscape. The potential barrier is

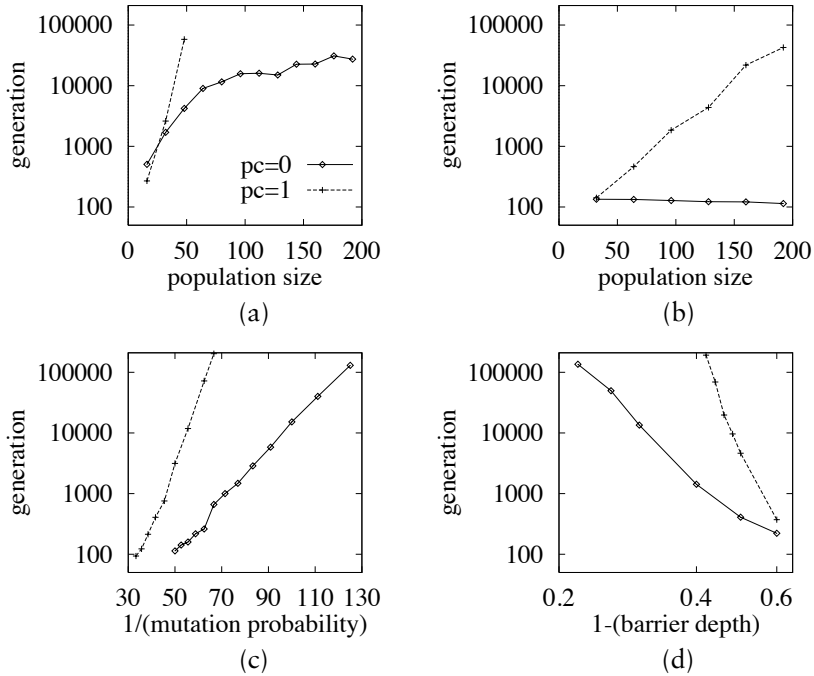


Figure 3. The duration T of metastability *versus* some parameters. Each point is an average of 100 runs. Default parameter values are $L = 30$, $N = 40$, $p_m = 0.01$, and $\delta = 0.7$ except that $p_m = 0.012$ and $p_m = 0.022$ for (a) and (b) respectively.

$D = F(1/3) - F(x_e)$. Assuming that $V(X)$ is almost insensitive to δ , which can be confirmed experimentally through the simulation, the major contribution of δ to T is from the term $F(1/3)$. Hence, the duration is, from equation (28), $T \sim \exp(-K_\delta \log(1-\delta)) = (1-\delta)^{-K_\delta}$ for $\delta \geq 0.5$ where K_δ is a constant. This is a very rough approximation but corresponds well with the log-log graph of Figure 3(d) which is linear in the domain of $(1 - \delta) \lesssim 0.5$. When δ is small, D is also so small that there is no effective barrier between the stable and metastable domains, breaking down the condition of bistability of the potential landscape.

5.2 Effect of crossovers for the functions of unitation

In a bistable problem using the fitness function of unitation, individuals can be divided into two types according to which basin of attraction they belong to: A - and B -type, respectively, in a metastable and stable area. Let r_A and f_A be the ratio and the average fitness of A -type individuals respectively, and r_B and f_B be defined for B -types, likewise. If the population contains only A -type individuals, the B -type cannot be

produced by selection only. Suppose that the probability of the B -type individual being produced by mutation or crossover is so small that $r_B = 0$ for most generations. Once r_B becomes positive at generation τ , the r_B of the next generation is

$$r_B(\tau + 1) \approx \frac{f_B}{\bar{f}} r_B(\tau) - \frac{f_B}{\bar{f}} r_B(\tau) \frac{f_A}{\bar{f}} r_A(\tau) \quad (34)$$

where $\bar{f} = f_A r_A + f_B r_B$, because the term of $r_A^2(\tau)$ and the contribution of the mutation is negligibly small. This equation implies that if one A -type and one B -type parent crossover, the children would be around the barrier and then quickly eliminated through succeeding selections. That is, crossover not only encourages the appearance of the B -type but also suppresses it provided that fitness is the function of unitation. The condition in order for r_B to decrease is, from equation (34),

$$f_B \sqrt{r_B} < \bar{f}. \quad (35)$$

This is also a sufficient condition for crossover to remove the rival. Since we have assumed that $r_B \ll 1$, this condition is satisfied even when f_B is considerably larger than f_A .

For the simulation, the effects of the SGA operations are shown in Table 1. The returns to the domain of $[1/3, 7/15]$ are mainly due to crossover while selection plays a major role in returning to the domain of $[0, 1/3)$. As a whole, crossover makes a major contribution to removing the appearance of the B -type. This also explains the longer durations of the graphs for $p_c = 1$ in Figure 3.

If the children of crossover are not too inferior, suppression may be replaced by enhancement. Crossover is highly dependent on the definitions of the phenotype and the fitness function. We discuss the range of fitness functions to which the analysis of this paper can be applied in section 5.3.

■ 5.3 Range of fitness functions

When fitness is the function of unitation, bistable problems can be categorized into two classes according to the infinite population dynamics [3, 4]. Let the ratio of the population in the global optimum be r_G . Because of the crossover effect discussed in section 5.2, the subpopulation in the global optimum decreases unless r_G is larger than a criterion r_c .

Class I consists of the fitness functions for which $r_c = 0$. In this case, the population transits the fitness barrier quickly, once an elite individual is produced in the global optimum. Let the number of generations necessary to produce the first elite in the global optimum be called T_1 . Then $T_1 \approx T$ for class I fitness functions. When the population size is infinite, the elite is produced immediately in the global optimum and T_1 is 1. However, if the population size is finite, T_1 may be much longer.

Escape by	Number of escapes	Number of returns by		
		selection	crossover	mutation
selection	0	0	0	0
mutation	333.9	217.8	108.2	7.4
crossover	190.5	110.8	59.7	19.5
sum	524.4	328.6	167.9	26.9

(a)

Escape by	Number of escapes	Number of returns by		
		selection	crossover	mutation
selection	0	0	0	0
mutation	12.5	4.8	6.8	0.5
crossover	8.8	2.9	3.7	1.6
sum	21.3	7.7	10.5	2.1

(b)

Table 1. The number of escapes from and returns to the local optimum due to SGA operations in a potential landscape. The individual with the maximal phenotype in the population is traced after each SGA operation. It sometimes escapes (from the local optimum) and returns (to the local optimum) across the particular phenotype criterion due to each operator. The criterion is 1/3 for (a), and 7/15 for (b). The numbers of escapes and returns due to each GA operation is counted until the transition occurs, and then averaged over 100 runs. Parameter values are $L = 30$, $N = 40$, $p_c = 1$, $p_m = 0.02$, and $\delta = 0.7$.

This *finite population effect* on T_1 was studied in [11] of which the analysis is based on the GA dynamics developed by [18–20]. According to their results, as the population grows large, the population variance increases and, hence, T_1 decreases. They also calculated *correlations* in the genotype, which is a measure of the similarity within population. When the population size is large, the correlation is small and, hence, crossover makes the population variance large. The effect of crossover is consistent with the theoretical proof of [10] and the simulation result of Table 1 where the number of escapes is contributed by crossover. However, these finite population effects are relatively small for class II fitness functions. This explains why the smaller population is more likely to escape from the local optimum of a class II function, even though the population variance is smaller according to those effects.

Class II is composed of fitness functions for which $0 < r_c < 1$. When the population size is infinite, the ratio r_G changes almost deterministically. As $r_G = 0$ initially, the population remains in the local optimum

forever during which the GA operators balance each other in equilibrium. This equilibrium can be disturbed if the population size is finite. Since each GA operator has some probabilistic parameters, the result of GA operations may be different each time if the population size is finite. This induces probabilistic fluctuations and can punctuate the equilibrium. In this case, when the population size is small, the disturbance is large and T decreases to a finite value. The major disturbance is not from the population variance but from the variance of the population mean. Thus the effect of the finite population is seemingly opposite to that of class I.

Since the stasis observed during a GA run for a class I function does not originate from an equilibrium between GA operators, we can bring the focus on class II functions when the punctuated equilibrium is studied. Among them, when r_c grows closer to 1, the analysis of this paper is better applied. It is not simple to calculate r_c exactly for a given fitness function. But, if the barrier is high and wide, and the difference in the fitness between local and global optima is not too large, the fitness function can be expected to belong to class II.

The relation between T and the population size N obtained in this paper is consistent with the result about natural genetic drift from which the term “punctuated equilibrium” has originated. The effect of finite population and crossover to the population variance can be found in Figure 4. When the populations are small there is a little increment in the population variance, as shown in graphs for which $N = 32$ and $N = 64$. However, comparing the graphs for which $N = 64$ and $N = 320$, we can see that the increment of the population variance is nearly ignorable when the populations are large. When crossover is used, the population variance increases a little, but this effect is also negligibly small when the populations are large, as shown by the graphs for which $N = 320$. The more significant change is diminution of fluctuations in the graphs as the population increases. This implies diminution of $\langle s_x^2 \rangle$ in equilibrium and is related with diminution of $\langle \bar{X} \rangle$ caused by the central limit theorem [12]. For the central limit theorem to be applied, we need $N \gtrsim 30$ and this too is one of the reasons why the finite population effect is not significantly large. Of course, to obtain the more accurate analysis, the finite population effect must be considered in the future.

6. Conclusion and future work

The punctuated equilibria observed in natural evolution are well modeled by a stochastic process in which a rigid ball is driven downward along a bistable potential landscape with small perturbations. The ball represents the population mean \bar{x} of the individual character x which is regarded as an instance of a random variable X . The perturbations are measured by a constant ε . When the ball is captured in the metastable

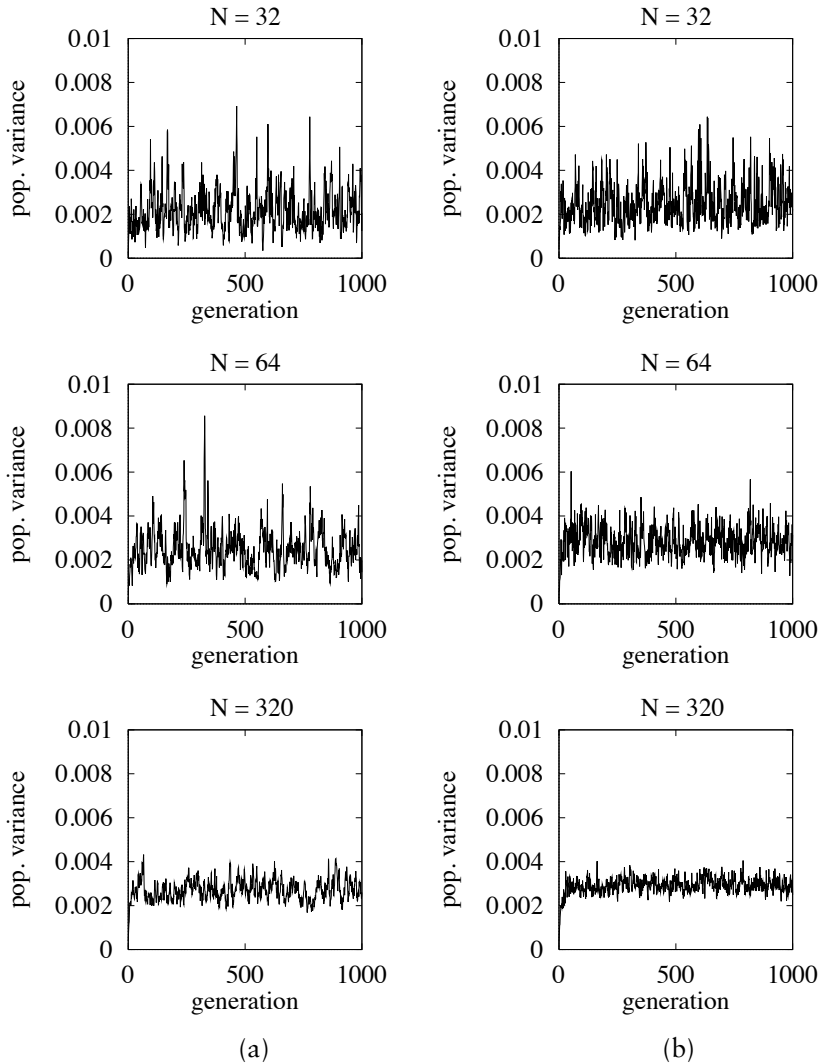


Figure 4. Population variance s_x^2 of a typical SGA run in equilibrium for various population sizes. Default parameter values are $L = 30$, $p_m = 0.01$, and $\delta = 0.7$ except that $p_c = 0$ and $p_c = 1$ for (a) and (b) respectively.

state, the perturbation enables the ball to escape out of the basin of attraction of the metastable state after a long duration of equilibrium. The duration T is proportional to $\exp(1/\varepsilon^2)$ according to mathematical studies on this stochastic process [8].

To apply the theory of punctuated equilibria to SGAs, a few conditions need to be satisfied. First, the conditions of section 3.2 are

necessary to make a continuous model from a discrete SGA. Next, the fitness functions of unitation are used to approximate the population to a rigid ball by suppressing the microscopic effect.

Under these conditions, our theoretical analysis achieved several results. First, $\varepsilon^2 \approx V(X)/N$ when the population size N is large enough to apply the central limit theorem. Second, the ensemble variance $V(X)$ is upper-bounded by V_e which is the value of $V(X)$ in the metastable state. Third, V_e is independent of N and proportional to the mutation probability p_m . Next, $V(X)$ is also independent of N and satisfies $V(X) \propto p_m$ after the metastable state is reached and before escape occurs. Finally, we obtained $T \propto \exp(N/p_m)$.

In the simulation, the duration T , which is defined by the number of generations until the transition occurs, was recorded *versus* N and p_m . The simulation result was observed to be compatible with the theoretical result. The relation between T and the depth of the fitness barrier was also observed to be consistent with a rough theoretical calculation.

If we consider the microscopic effect only, it is expected that duration shortens as population size enlarges. We ignored the microscopic effect in this paper since that is relatively small for punctuated equilibria. But, to obtain a more accurate analysis, the microscopic effect must be considered in the future. We suggest that dividing a large population into mutually isolated subpopulations can be a strategy for obtaining both microscopic and macroscopic effects. This can be one of the theoretical bases for the good performance of parallel GAs. On the other hand, equation (32) confirms the fact that larger mutation probability and smaller fitness barrier depth encourage the global optimization of SGAs. This reminds us of the basic idea of simulated annealing [21].

The results obtained in this paper could possibly be extended in various ways. First, the results on duration could be expanded to cases where F has more than two peaks, or x is multidimensional with some restrictions added to the potential landscape [8, 22]. Also, since GAs have direct relations with CEs, results obtained from studies on CEs could be applied to GAs. These include the issues of cooperation, competition, chaos, and delayed global information [17, 23].

Acknowledgment

This work was supported by the Korea Science and Engineering Foundation (KOSEF) through the Advanced Information Technology Research Center (AITRC).

Appendix

A. Derivative of $\langle r_i \rangle$

The change of the probability during sufficiently small time Δt is

$$\begin{aligned} & \frac{P(\mathbf{n}, t + \Delta t) - P(\mathbf{n}, t)}{\alpha \Delta t} \\ &= -P(\mathbf{n}, t) \sum_{i \neq j} n_j \rho_i + \sum_{i \neq j} P(\mathbf{n}^{[j,i]}, t) (n_j + 1) \rho_i^{[j,i]}. \end{aligned} \tag{A.1}$$

By applying $\Delta t \rightarrow 0$, we obtain

$$\frac{dP(\mathbf{n}, t)}{\alpha dt} = -P(\mathbf{n}, t) \sum_{i \neq j} n_j \rho_i + \sum_{i \neq j} P(\mathbf{n}^{[j,i]}, t) (n_j + 1) \rho_i^{[j,i]}. \tag{A.2}$$

Now we derive $d\langle n_k \rangle / dt$ using this equation. Since n_k itself is independent of t ,

$$\frac{d\langle n_k \rangle}{dt} = \frac{d}{dt} \sum_{\mathbf{n}} P(\mathbf{n}, t) n_k = \sum_{\mathbf{n}} \frac{dP(\mathbf{n}, t)}{dt} n_k. \tag{A.3}$$

The derivative of $\langle n_k \rangle$ is, replacing dP/dt ,

$$\begin{aligned} \frac{d\langle n_k \rangle}{\alpha dt} &= - \sum_{\mathbf{n}} n_k P(\mathbf{n}, t) \sum_{i \neq j} n_j \rho_i + \sum_{\mathbf{n}} n_k \sum_{i \neq j} P(\mathbf{n}^{[j,i]}, t) (n_j + 1) \rho_i^{[j,i]} \\ &= \sum_{i \neq j} \left\{ - \sum_{\mathbf{n}} P(\mathbf{n}, t) n_j \rho_i n_k + \sum_{\mathbf{n}} P(\mathbf{n}^{[j,i]}, t) (n_j + 1) \rho_i^{[j,i]} n_k \right\}. \end{aligned} \tag{A.4}$$

The terms satisfying $i \neq k$ and $j \neq k$ are counterbalanced by each other. The remaining terms are

$$\begin{aligned} \frac{d\langle n_k \rangle}{\alpha dt} &= \sum_{k \neq j} \left\{ - \sum_{\mathbf{n}} P(\mathbf{n}, t) n_j \rho_k n_k + \sum_{\mathbf{n}} P(\mathbf{n}^{[j,k]}, t) (n_j + 1) \rho_k^{[j,k]} n_k \right\} \\ &\quad + \sum_{i \neq k} \left\{ - \sum_{\mathbf{n}} P(\mathbf{n}, t) n_k^2 \rho_i + \sum_{\mathbf{n}} P(\mathbf{n}^{[k,i]}, t) (n_k + 1) \rho_i^{[k,i]} n_k \right\} \\ &= \sum_{k \neq j} \left\{ - \sum_{\mathbf{n}} P(\mathbf{n}, t) n_j \rho_k n_k + \sum_{\mathbf{n}} P(\mathbf{n}, t) \rho_k n_j (n_k + 1) \right\} \\ &\quad + \sum_{i \neq k} \left\{ - \sum_{\mathbf{n}} P(\mathbf{n}, t) n_k^2 \rho_i + \sum_{\mathbf{n}} P(\mathbf{n}, t) \rho_i n_k (n_k - 1) \right\} \\ &= \sum_{k \neq j} \sum_{\mathbf{n}} P(\mathbf{n}, t) \rho_k n_j - \sum_{i \neq k} \sum_{\mathbf{n}} P(\mathbf{n}, t) \rho_i n_k \end{aligned}$$

$$\begin{aligned}
 &= \sum_{\mathbf{n}} P(\mathbf{n}, t) \left(\sum_j \rho_k n_j - \sum_i \rho_i n_k \right) \\
 &= \sum_{\mathbf{n}} P(\mathbf{n}, t) (N\rho_k - n_k) \\
 &= N\langle \rho_k \rangle - \langle n_k \rangle.
 \end{aligned} \tag{A.5}$$

We finally obtain, dividing by N ,

$$\frac{d\langle r_k \rangle}{\alpha dt} = \langle \rho_k \rangle - \langle r_k \rangle. \tag{A.6}$$

B. Derivation of equations (16) and (17)

The probability of $P(y_{i,1})$ is

$$P(y_{i,1}) = \binom{u}{y_{i,1}} \binom{L-u}{l_i - y_{i,1}} \binom{L}{l_i}^{-1}.$$

According to the binomial theorem $(1+z)^u = \sum_y \binom{u}{y} z^y$ [16], we can obtain

$$uz(1+z)^{u-1} = \sum_y y \binom{u}{y} z^y. \tag{B.1}$$

Multiplying by $(1+z)^{L-u}$,

$$\begin{aligned}
 uz(1+z)^{u-1}(1+z)^{L-u} &= uz(1+z)^{L-1} \\
 &= uz \sum_y \binom{L-1}{y} z^y \\
 &= \sum_{y_1, y_2} y_1 \binom{u}{y_1} \binom{L-u}{y_2} z^{y_1+y_2}.
 \end{aligned} \tag{B.2}$$

Extracting terms of order z^l ,

$$uz \binom{L-1}{l-1} z^{l-1} = \sum_{y_1+y_2=l} y_1 \binom{u}{y_1} \binom{L-u}{y_2} z^l. \tag{B.3}$$

Letting $z = 1$,

$$u \binom{L-1}{l-1} = \sum_{y_1} y_1 \binom{u}{y_1} \binom{L-u}{l-y_1}. \tag{B.4}$$

We can now obtain

$$\sum_{y_{i,1}} y_{i,1} P(y_{i,1}) = u \binom{L-1}{l_i-1} \binom{L}{l_i}^{-1} = \frac{ul_i}{L} = ux_i. \tag{B.5}$$

Next, in the same way, we can obtain

$$uz(1+uz)(1+z)^{u-2} = \sum_y y^2 \binom{u}{y} z^y. \quad (\text{B.6})$$

Multiplying by $(1+z)^{L-u}$,

$$\begin{aligned} uz(1+uz)(1+z)^{u-2}(1+z)^{L-u} &= uz(1+uz)(1+z)^{L-2} \\ &= uz(1+uz) \sum_y \binom{L-2}{y} z^y \\ &= \sum_{y_1, y_2} y_1^2 \binom{u}{y_1} \binom{L-u}{y_2} z^{y_1+y_2}. \end{aligned} \quad (\text{B.7})$$

Extracting terms of order z^l ,

$$\begin{aligned} uz \binom{L-2}{l-1} z^{l-1} + u^2 z^2 \binom{L-2}{l-2} z^{l-2} \\ = \sum_{y_1+y_2=l} y_1^2 \binom{u}{y_1} \binom{L-u}{y_2} z^l. \end{aligned} \quad (\text{B.8})$$

Setting $z = 1$,

$$u \binom{L-2}{l-1} + u^2 \binom{L-2}{l-2} = \sum_{y_1} y_1^2 \binom{u}{y_1} \binom{L-u}{l-y_1}. \quad (\text{B.9})$$

Hence, we can obtain

$$\begin{aligned} \sum_{y_{i,1}} y_{i,1}^2 P(y_{i,1}) &= \left[u \binom{L-2}{l_i-1} + u^2 \binom{L-2}{l_i-2} \right] \binom{L}{l_i}^{-1} \\ &= \frac{(uL - u^2)x_i + (u^2 - u)Lx_i^2}{L-1}. \end{aligned} \quad (\text{B.10})$$

References

- [1] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning* (Addison-Wesley, New York, 1989).
- [2] J. H. Holland, *Adaptation in Natural and Artificial Systems* (The MIT Press, Cambridge, MA, second edition, 1992).
- [3] D. E. Goldberg, "Simple Genetic Algorithms and the Minimal, Deceptive Problem," in *Genetic Algorithms and Simulated Annealing* (Morgan Kaufmann, Los Altos, CA, 1987).
- [4] J. Shapiro and A. Prügel-Bennett, "Genetic Algorithm Dynamics in a Two-well Potential," in *Foundations of Genetic Algorithms Four*, edited by R. K. Belew and M. D. Vose (Morgan Kaufmann, San Francisco, 1997).

- [5] H. A. Ceccatto and B. A. Huberman, "Persistence of Nonoptimal Strategies," *Proceedings of the National Academy of Sciences of the United States of America*, **86** (1989) 3443–3446.
- [6] C. M. Newman, J. E. Cohen, and C. Kipnis, "Neo-darwinian Evolution Implies Punctuated Equilibria," *Nature*, **315** (May 1985) 400–401.
- [7] M. D. Vose and G. E. Liepins, "Punctuated Equilibria in Genetic Search," *Complex Systems*, **5** (1991) 31–44.
- [8] C. Kipnis and C. M. Newman, "The Metastable Behavior of Infrequently Observed, Weakly Random, One-dimensional Diffusion Processes," *SIAM Journal on Applied Mathematics*, **45**(6) (1985) 972–982.
- [9] E. van Nimwegen, J. P. Crutchfield, and M. Mitchell, "Statistical Dynamics of the Royal Road Genetic Algorithm," Santa Fe Institute Working Paper 97-04-035 (1997).
- [10] T. Jansen and I. Wegener, "On the Analysis of Evolutionary Algorithms: A Proof that Crossover Really Can Help," in *Proceedings of the Seventh Annual European Symposium on Algorithms (ESA'99)* edited by J. Nešetřil (Springer, Berlin, 1999).
- [11] A. Rogers and A. Prügel-Bennett, "The Dynamics of a Genetic Algorithm on a Model Hard Optimization Problem," *Complex Systems*, **11** (1997) 437–464.
- [12] R. L. Scheaffer and J. T. McClave, *Probability and Statistics for Engineers* (PWS-KENT, Boston, 1990).
- [13] M. Srinivas and L. M. Patnaik, "On Modeling Genetic Algorithms for Functions of Unitation," *IEEE Transactions on Systems, Man, and Cybernetics*, **26**(6) (1996) 809–821.
- [14] B. A. Huberman and T. Hogg, "The Behavior of Computational Ecologies," in *The Ecology of Computation* edited by B. A. Huberman (North-Holland, Amsterdam, 1988).
- [15] N. V. Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier Science, Amsterdam, 1981).
- [16] R. C. Bose and B. Manvel, *Introduction to Combinatorial Theory* (John Wiley & Sons, New York, 1984).
- [17] J. O. Kephart, T. Hogg, and B. A. Huberman, "Dynamics of Computational Ecosystems," *Physical Review A*, **40**(1) (1989) 404–421.
- [18] A. Prügel-Bennett and J. L. Shapiro, "An Analysis of Genetic Algorithms Using Statistical Mechanics," *Physical Review Letters*, **72** (1994) 1305–1309.
- [19] A. Prügel-Bennett and J. L. Shapiro, "The Dynamics of a Genetic Algorithm for Simple Random Ising Systems," *Physica D*, **104** (1997) 75–114.

- [20] M. Rattray, “The Dynamics of a Genetic Algorithm Under Stabilizing Selection,” *Complex Systems*, **9**(3) (1995) 213–234.
- [21] E. Aarts and J. Corst, *Simulated Annealing and Boltzmann Machines* (John Wiley & Sons, New York, 1989).
- [22] A. Galves, E. Olivieri, and M. E. Vares, “Metastability for a Class of Dynamical Systems Subject to Small Random Perturbations,” *The Annals of Probability*, **15**(4) (1987) 1288–1305.
- [23] T. Hogg and B. A. Huberman, “Controlling Chaos in Distributed Systems,” *IEEE transactions on Systems, Man, and Cybernetics*, **21**(6) (1991) 1325–1332.