

A MARKOV CHAIN ANALYSIS OF GENETIC ALGORITHMS: LARGE DEVIATION PRINCIPLE APPROACH

JOE SUZUKI,* *Osaka University*

Abstract

In this paper we prove that the stationary distribution of populations in genetic algorithms focuses on the uniform population with the highest fitness value as the selective pressure goes to ∞ and the mutation probability goes to 0. The obtained sufficient condition is based on the work of Albuquerque and Mazza (2000), who, following Cerf (1998), applied the large deviation principle approach (Freidlin–Wentzell theory) to the Markov chain of genetic algorithms. The sufficient condition is more general than that of Albuquerque and Mazza, and covers a set of parameters which were not found by Cerf.

Keywords: Genetic algorithm; Markov chain; large deviation principle; evolutionary algorithm; Freidlin–Wentzell

2010 Mathematics Subject Classification: Primary 65C40
Secondary 65C50

1. Introduction

We aim to maximize the value of a given function $f: \{0, 1\}^* \rightarrow (0, \infty)$ using evolutionary algorithms (EAs), which are computational methods inspired by evolutionary processes. They have been utilized for more than 20 years.

EAs repeatedly update the set (population) of candidate solutions (individuals), each expressed by a binary string, using genetic operations to raise the probability of getting solutions with the highest value of the fitness function f .

Initially, EAs consisted of three genetic operators, namely selection, crossover, and mutation [9], and were called genetic algorithms (GAs). However, currently, various genetic operations are applied in the context of EAs.

In GAs, a population containing a definite number (population size) of individuals consisting of genetics is first set up. Then, individuals are selected according to probabilities that are proportional to the fitness value multiplied by the number of occurrences. This process (selection) is repeated until individuals of the current population form part of the updated population. Then, for a pair of individuals, we generate a new pair of individuals by exchanging genetics in some locations. This process (crossover) is repeated several times, and a further updated population is obtained. Determining the locations, choosing the pairs of individuals for crossover, etc., can be chosen arbitrarily. Finally, we randomly change several genetics in several individuals (mutation). The probability of mutation can also be chosen arbitrarily.

Received 28 February 2008; revision received 26 June 2010.

* Postal address: Department of Mathematics, Osaka University, Toyonaka, Osaka 560-0043, Japan.

Email address: suzuki@math.sci.osaka-u.ac.jp

Although various genetic operators may be applied in EAs, all EAs share the following properties.

- To find the individuals that optimize the fitness function.
- Their populations are states of a finite Markov chain.

However, in general, it is difficult to mathematically analyze the performance of EAs. Several results on evaluating the performance of GAs expressed by Markov chains have been reported since the early 1990s. In this paper we analyze GAs using the large deviation principle, in particular the Freidlin–Wentzell theory [8]. The basic idea was developed by Cerf [3], [4]. Similar evaluation methods have been applied not just to GAs but also to EAs, following Cerf.

Olivier Francois [7] considered an algorithm that selects a definite number of individuals and randomly changes at most one gene for each individual (mutation) with a certain probability, where the mutation probability gradually decreases as the number of generations increases. The performance of his simulated annealing type algorithm was evaluated using Cerf’s method.

Rigal and Truffet [10] considered an algorithm that generates a number, less than the population size, of individuals via mutation and selects offspring among the parents and the newly generated individuals so that the number of individuals in the next generation is equal to that in the current generation. In their algorithm, the fitness function was powered by a value (selective pressure) which increases as the number of generations increases. They proved that the stationary probability focuses on the population consisting of the best individuals as the selective pressure increases.

Intuitively, decreasing the mutation probability and increasing the selective pressure are equivalent in the sense that both decrease the temperature in statistical dynamics. For both algorithms, it is proved that the uniform populations consisting of identical individuals have positive probabilities when the temperature is low. Then, Rigal and Truffet concluded that the stationary probability focuses on uniform populations (best uniform populations), each consisting of the best individuals if the population size exceeds a definite limit. The derivation was based on the condition that the upper bound of virtual energies for the best uniform populations should be lower than the lower bound of virtual energies for the other uniform populations. We refer the reader to [2] for the details of the theoretical analysis.

Although this paper does not propose any simulated annealing methods, we show that the stationary probability focuses on the best uniform populations when the mutation probability goes to 0 and the selective pressure goes to ∞ , keeping the two factors balanced. The statement is similar to Cerf, but the obtained sufficient condition is different since our derivation is based on Albuquerque and Mazza [1], who utilized the fact that the stationary probability is positive when the potential value is minimized, if a potential function exists.

This paper refines that of Albuquerque and Mazza [1] in the following senses.

- We consider the effect of crossover as well as the effects of selection and mutation.
- Our proof does not require a nongenetic operator to add a random small value to each fitness value if more than one best individual exists.
- The sufficient condition in this paper is more general than that of Albuquerque and Mazza.
- While the result in Albuquerque and Mazza holds for the order ‘mutation \rightarrow crossover \rightarrow selection’, the result in this paper holds for ‘crossover \rightarrow mutation \rightarrow selection’ as well.

Cerf conjectured that the stationary probability would not focus on the best uniform populations unless the population size exceeds a limit. However, in this paper we show that the property is true for *any* population size greater than 1 as long as the mutation probability and selective pressure are balanced properly.

2. Markov chains of GAs

Using GAs, we find individuals $i \in I := \{0, 1\}^L$, $L \geq 1$, that maximize the fitness value $f(i)$ for a given fitness function $f: I \rightarrow (0, \infty)$. In the population $S := I^M$ we keep the population size, M , greater than or equal to 2 individuals. In the sequel, we denote by $i \oplus j$ the bitwise exclusive-or operation between $i, j \in I$ and by $|k|$ the number of 1s in $k \in I$. Also, we write the Hamming distance between individuals $i, j \in I$ by $\rho(i, j) := |i \oplus j|$, and the Hamming distance between populations $x = (x_1, \dots, x_M)$ and $y = (y_1, \dots, y_M) \in S$ by $d(x, y) := \sum_{k=1}^M \rho(x_k, y_k)$.

We repeat the cycles consisting of ‘mutation \rightarrow crossover \rightarrow selection’ genetic operations in this order (if we change the order to ‘crossover \rightarrow mutation \rightarrow selection’, a similar discussion will follow).

The mutation operator randomly changes the LM bits in the current population $x \in S$ with probability (mutation probability) $0 < \mu < 1$ to give a new population $y \in S$ with probability

$$\mathcal{M}_\mu[x, y] := \mu^{d(x,y)}(1 - \mu)^{LM-d(x,y)} > 0.$$

The crossover operator changes genes i_k, j_k , $k \in \pi$, in several pairs of individuals $i = (i_1, \dots, i_L)$, $j = (j_1, \dots, j_L) \in I$ contained in the current population $x \in S$ with respect to $\pi \subseteq \{1, \dots, L\}$ to form a new population $y \in S$ with probability $\mathcal{C}[x, y]$. In particular, $\{1, \dots, k\}$ with random $1 \leq k \leq L$ is often used as π . Although the pairs of individuals may be chosen arbitrarily, crossover does not change the number of 1s in

$$\{i_{k,j}, 1 \leq j \leq M \mid x_j = (i_{1,j}, \dots, i_{k,j}, \dots, i_{L,j}), 1 \leq j \leq M, x = (x_1, \dots, x_M) \in S\}$$

for all $1 \leq k \leq L$.

We write $x \sim y$ when $\mathcal{C}[x, y] > 0$, so that $x \sim y$ implies that $y \sim x$ for any crossover. Also, we write $x \succ y$ when $\{x_1, \dots, x_M\} \supseteq \{y_1, \dots, y_M\}$ for $x = (x_1, \dots, x_M)$, $y = (y_1, \dots, y_M) \in S$, and $x_k, y_k \in I$, $k = 1, \dots, M$.

The selection operator independently selects individuals $y_k \in I$ in the current population $x \in S$ such that $x \succ y$ to form a new population $y = (y_1, \dots, y_M) \in S$ with probability

$$S_\beta[x, y] := \begin{cases} \frac{\prod_{k=1}^M f(y_k)^\beta}{[\sum_{j=1}^M f(x_j)^\beta]^M}, & x \succ y, \\ 0, & x \not\succ y, \end{cases}$$

for $\beta > 0$ (Boltzmann selection [7]).

Our GA is defined as a repetition process (generation changes) of the following three genetic operations:

- apply mutation to $x \in S$ with parameter μ to obtain $u \in S$;
- apply crossover to u to obtain $v \in S$; and
- apply selection to v with parameter β to obtain $y \in S$.

The new population consists of individuals $i \in I$ that maximize $f(i)$ [9]. Although it is not easy to define GAs generally, we require only the above assumptions for GAs in this paper.

Then, the probability of obtaining $y \in S$ from $x \in S$ is

$$Q_{\mu,\beta}[x, y] := \sum_{u,v \in S} \mathcal{M}_\mu[x, u] \mathcal{C}[u, v] \delta_\beta[v, y],$$

where the matrix $Q_{\mu,\beta} = (Q_{\mu,\beta}[x, y])_{x,y \in S}$ is the transition probability matrix of a finite Markov chain. Since the Markov chain is irreducible ($Q_{\mu,\beta}[x, y] > 0$) and acyclic, which will be defined later, the stationary probability $(q_{\mu,\beta}(x))_{x \in S}$ is unique for each $0 < \mu < 1$ and $\beta > 0$.

We define the set of uniform populations by

$$U := \{(x_1, \dots, x_M) \in S \mid x_1 = \dots = x_M\}$$

and the set of uniform populations consisting of individuals that maximize the fitness function by

$$U^* := \{(x_1, \dots, x_M) \in U \mid f(x_k) = \max_{i \in I} f(i), k = 1, \dots, M\}.$$

For stationary probabilities, several results are known.

Proposition 2.1. ([5].) *When crossover is not present, for arbitrary $\beta > 0$,*

$$\lim_{\mu \rightarrow 0} q_{\mu,\beta}(x) > 0 \implies x \in U.$$

Proposition 2.2. ([6], [11], and [12].) *When crossover is not present,*

$$\lim_{\beta \rightarrow \infty} \lim_{\mu \rightarrow 0} q_{\mu,\beta}(x) > 0 \implies x \in U^*.$$

It is not recommendable to exclude crossover to obtain successful performance, but mathematically analyzing GAs becomes hard if crossover is included.

In this paper we prove Proposition 2.2 even when crossover is present. In particular, we utilize the approach based on the Freidlin–Wentzell theory [8] and large deviation techniques to evaluate the Markov chains.

In general, when the absolute temperature $T_{\mu,\beta}$ goes to 0, the stationary probability $(q_{\mu,\beta})_{\mu,\beta}$ will focus on the set $S^* := \{x \in S \mid W(x) = \min_{y \in S} W(y)\}$ of states that minimize the virtual energy $W(x)$, $x \in S$, which will be defined later ($S^* \subseteq U$ can be proved). Therefore, it is sufficient to derive the conditions with respect to L , M , f , μ , and β for $S^* \subseteq U^*$. It was Cerf [4] who first introduced this approach. An upper bound W^- of $W(x)$, $x \in U^*$, a lower bound W^+ of $W(x)$, $x \in U \setminus U^*$, and, as a result, conditions for $W^- < W^+$ were derived in the paper.

Following Cerf, Albuquerque and Mazza [1] proved that $U^* \subseteq S^*$, under the conditions that

- there exist $0 < \varepsilon < 1$ and $\lambda > 0$ such that $\mu = \varepsilon \exp(-\lambda\beta)$;
- $\lambda > M\Delta := \max_{i,j \in I} |\log f(i) - \log f(j)|$,

by showing that $\varphi(x) := \sum_{k=1}^M -\log f(x_k)$ is a potential which will be defined later. However, in their derivation,

- a nongenetic operation of adding a small random value to each fitness value so that $f: I \rightarrow (0, \infty)$ is injective was included; and
- crossover was not considered.

In this paper, we prove that $U^* \subseteq S^*$ under $\lambda > M\Delta/(M - 1)$.

3. Large deviation principle approach

We consider a general finite Markov chain $\{X_n\}_{n=0}^\infty$ with state set S and transition probability $Q \in \mathbb{R}^{\sigma \times \sigma}$, where $\sigma := |S|$. Since the sum of the row elements in the matrix Q is 1, $\lambda = 1$ is a solution of the equation $\det(\lambda I - Q) = 0$, which does not mean that the dimension of the corresponding eigenvalue space is 1. We say that the Markov chain $\{X_n\}_{n=0}^\infty$ is irreducible if all the elements become positive when the matrix Q is multiplied finitely many times. The integer $\gcd\{n \mid P(X_n = x \mid X_0 = x) > 0\}$ is said to be a cycle of $x \in S$. If all the cycles of $x \in S$ are 1, then the Markov chain $\{X_n\}$ is said to be acyclic. In general, the dimension of each eigenvalue space for the irreducible and acyclic finite Markov chain is 1 (Perron–Frobenius). Then, the stationary probability $(q(x))_{x \in S}$ of the transition probability Q is unique:

$$qQ = q, \quad \sum_{x \in S} q(x) = 1.$$

We express the set of irreducible finite Markov chains by their transition probability matrix $(Q_\alpha[x, y])_{x, y \in S}$ parameterized by $\alpha > 0$ with common state set S . We write the stationary probabilities corresponding to $(Q_\alpha[x, y])_{x, y \in S}$ by $(q_\alpha(x))_{x \in S}$, and define $q_\infty(x) := \lim_{\alpha \rightarrow \infty} q_\alpha(x)$, $x \in S$. We define the communication cost from $x \in S$ to $y \in S$ by

$$V(x \rightarrow y) := \lim_{\alpha \rightarrow \infty} -\frac{1}{\alpha} \log Q_\alpha[x, y] \in [0, \infty].$$

Also, we say that the sequence $\{(z_{i-1} \rightarrow z_i)\}_{i=1}^k$, $z_i \in S$, for any k is a path from $z_0 \in S$ to $z_k \in S$, and that each $(z_{i-1} \rightarrow z_i)$ is an edge contained in the path. If there exists only one path from each $x \in S \setminus \{y\}$ to $y \in S$ then we say that the set of edges contained in those paths is a spanning tree rooted at y and write the set of spanning trees rooted at y by $G(y)$. Let

$$V(\gamma) := \sum_{(y \rightarrow z) \in \gamma} V(y \rightarrow z), \quad \gamma \in G(x),$$

$$W(x) := \min_{\gamma \in G(x)} V(\gamma), \quad \text{and} \quad W_{\min} := \min_{x \in S} W(x).$$

It is known (Freidlin–Wentzell [8]) that, for

$$Q_\alpha(x) := \sum_{\gamma \in G(x)} \prod_{(y \rightarrow z) \in \gamma} Q_\alpha[y, z], \quad x \in S,$$

the stationary probability of the transition probability matrix $(Q_\alpha[x, y])_{x, y \in S}$ is given by

$$q_\alpha(x) = \frac{Q_\alpha(x)}{\sum_{y \in S} Q_\alpha(y)}, \quad x \in S,$$

and $\lim_{\alpha \rightarrow \infty} (-\log q_\alpha(x)/\alpha) = W(x) - W_{\min}$. Thus, for each $x \in S$, we have

$$q_\infty(x) > 0 \implies W(x) = W_{\min}. \tag{3.1}$$

Furthermore, for each $(x, y) \in S^2$ such that $V(x \rightarrow y) \in [0, \infty)$, we say that φ is a potential if $\varphi: S \rightarrow [0, \infty)$ satisfies

$$V(x \rightarrow y) - V(y \rightarrow x) = \varphi(y) - \varphi(x). \tag{3.2}$$

The proofs of the following lemmas are given in Appendix A for self-containedness.

Lemma 3.1. *If a potential $\varphi: S \rightarrow [0, \infty)$ exists, for each $x \in S$,*

$$q_\infty(x) = 0 \implies \varphi(x) > \min_{y \in S} \varphi(y).$$

Lemma 3.2. *If a subset S_- of S satisfies*

- (a) *for each $x \in S_+ := S \setminus S_-$, there exists $y \in S_-$ such that $V(x \rightarrow y) = 0$; and*
- (b) *for each pair of $x \in S_+$ and $y \in S_-$, $V(y \rightarrow x) > 0$,*

then $q_\infty(x) > 0$ implies that $x \in S_-$.

4. Main result

We denote the mutation probability by $\mu(\beta) := \varepsilon \exp(-\lambda\beta)$, where β is a parameter and $0 < \varepsilon < 1$ and $\lambda > 0$ are constants. Here $0 < \varepsilon < 1$ is required for $0 < \mu(0) < 1$. Then, we apply Lemmas 3.1 and 3.2, expressing $\{Q_{\mu,\beta}[x, y]\}_{x,y \in S}$ by $\{Q_\beta[x, y]\}_{x,y \in S}$.

We note that any uniform population remains the same via selection and via crossover.

Lemma 4.1. (a) *$x \in U$ and $x \succ y$ imply that $y \in U$.*

(b) *$x \in U$ and $x \sim y$ imply that $y \in U$.*

For $x = (x_1, \dots, x_M) \in S$, if we set $\varphi(x) := -\sum_{k=1}^M \log f(x_k)$, then

$$\begin{aligned} V(x \rightarrow y) &= \lim_{\beta \rightarrow \infty} -\frac{1}{\beta} \log Q_\beta[x, y] \\ &= \lim_{\beta \rightarrow \infty} -\frac{1}{\beta} \log \left\{ \sum_{u,v \in S} \mathcal{M}_\beta[x, u] \mathcal{C}[u, v] \mathcal{S}_\beta[u, v] \right\} \\ &= \lim_{\beta \rightarrow \infty} -\frac{1}{\beta} \log \left\{ \sum_{u \sim v} C[u, v] \exp\{-[\lambda d(x, u) + \varphi(y) - \min_{v \succ r} \varphi(r)]\beta + o(\beta)\} \right\} \\ &= \min_{u \sim v, v \succ y} \{ \lambda d(x, u) + \varphi(y) - \min_{v \succ r} \varphi(r) \}, \end{aligned} \tag{4.1}$$

where $o(\beta)$ is a quantity satisfying

$$\lim_{\beta \rightarrow \infty} \frac{o(\beta)}{\beta} = 0.$$

Let $\Delta := \max_{i,j \in I} |\log f(i) - \log f(j)|$. Then we have the following result.

Theorem 4.1. *Under the condition that $\lambda > M\Delta/(M - 1)$, the stationary distribution $q_\beta(x)$, $x \in S$, is asymptotically positive as $\beta \rightarrow \infty$ only if $x \in U^*$.*

Proof. We denote by $(i) \in U$ the uniform population consisting of only $i \in I$. For $x, y \in S$, we find $u, v \in S$ such that $u \sim v$ and $v \succ y$ minimize $\lambda d(x, u) + \varphi(y) - \min_{v \succ y} \varphi(r)$.

- For $x = (x_1, \dots, x_M) \in S \setminus U$, if $i \in \{x_1, \dots, x_M\}$ maximizes $f(i)$ then, for $x = u = v \succ (i) = y \in U$, we have $u \sim v, v \succ y$, and $\lambda d(x, u) + \varphi(y) - \min_{v \succ r} \varphi(r) = 0$. Thus, there exists a $y \in U$ such that $V(x \rightarrow y) = 0$.

- For $x \in U$ and $y \in S \setminus U$, in $x \rightarrow u \rightarrow v \rightarrow y$, where mutation, crossover, and selection are applied, respectively, from Lemma 3.2(a) we have $v \in S \setminus U$. From Lemma 3.2(b) we have $u \in S \setminus U$. From $x \in U$ we have $d(x, u) > 0$, which means that $V(x \rightarrow y) > 0$.

Thus, from Lemma 3.2, the dynamics in S are restricted to U .

Next, we show that $\varphi((i)) = -M \log f(i)$, $i \in I$, is a potential of U ; therefore, we can complete the proof of Theorem 4.1 using Lemma 3.1. We evaluate the communication cost

$$\bar{V}((i) \rightarrow (j)) := \min_k \min_{z_1, \dots, z_{k-1}} \sum_{i=1}^k V(z_{i-1} \rightarrow z_i)$$

from $(i) \in U$ to $(j) \in U$ with $z_0 = (i)$ and $z_k = (j)$ for some k .

Suppose that $\varphi((j)) \geq \varphi((i))$. If the population contains both i and j immediately before the last selection, then the cost excluding the last selection and the one at the last selection are no less than $\lambda\rho(i, j)$ and $\varphi((j)) - \varphi((i))$, respectively. In fact, in order to generate j , it is necessary to either generate j via mutation directly, or $k, h \in I$ via mutation and change one of them into j via crossover. However, for the latter case,

$$\rho(i, k) + \rho(i, h) \geq \rho(i, j)$$

bits are necessary to change from i to j in the population immediately before the last selection. Thus, before the last selection, the cost is at least $\lambda\rho(i, j)$.

If the population does not contain i but contains j immediately before the last selection, at least one j should be generated from i and the $M - 1$ individuals are not i , so that the cost should be at least $\lambda\rho(i, j) + \lambda(M - 1)$. On the other hand, if all the individual are j immediately before the last selection, no cost occurs for the last selection. Thus, the total cost is at most $M\lambda\rho(i, j)$. Hence, we have

$$\begin{aligned} & \min\{\lambda\rho(i, j) + \varphi((j)) - \varphi((i)), \lambda[\rho(i, j) + M - 1]\} \\ & \leq \bar{V}((i) \rightarrow (j)) \\ & \leq \min\{\lambda\rho(i, j) + \varphi((j)) - \varphi((i)), M\lambda\rho(i, j)\}, \end{aligned}$$

which means that, from $\lambda > M\Delta/(M - 1)$,

$$\bar{V}((i) \rightarrow (j)) = \lambda\rho(i, j) + \varphi((j)) - \varphi((i)). \tag{4.2}$$

Suppose that $\varphi((j)) < \varphi((i))$. The cost of the last selection can be at least 0, and the total becomes

$$\bar{V}((i) \rightarrow (j)) = V((i) \rightarrow (j)) = \lambda\rho(i, j). \tag{4.3}$$

From (4.2) and (4.3), for both cases we have $\bar{V}((i) \rightarrow (j)) - \bar{V}((j) \rightarrow (i)) = \varphi((j)) - \varphi((i))$, which means that $\varphi((i))$, $i \in I$, is a potential of U .

5. Concluding remarks

In case the order is ‘crossover \rightarrow mutation \rightarrow selection’ instead of ‘mutation \rightarrow crossover \rightarrow selection’, if we replace (4.1) by

$$V(x \rightarrow y) = \min_{u>v, v\sim y} \left\{ \lambda d(x, u) + \varphi(v) - \min_{u>r} \varphi(r) \right\}$$

under the condition that $\lambda > M\Delta/(M - 1)$, a statement identical to Theorem 4.1 is obtained.

Cerf [4] proved that $U^* \subseteq S^*$ for the region

$$\left\{ (M, \lambda) \mid M \geq \frac{\lambda L + (L - 1)\Delta}{\min\{\delta, \lambda\}} \right\}$$

when no crossover is present in GAs (see Theorem 4 of [4]), where $\delta := \min_{i,j \in I} |\log f(i) - \log f(j)|$, which requires

$$M \geq L + (L - 1) \frac{\Delta}{\delta}$$

for all λ .

In this paper we claim that such a range includes

$$\left\{ (M, \lambda) \mid \lambda > \frac{M}{M - 1} \Delta \right\},$$

which means that the property is satisfied when λ is large enough, even if $M = 2$.

Appendix A

A.1. Proof of Lemma 3.1

For $x, y \in S$, let $R[x, y]: G(x) \rightarrow G(y)$ be the mapping from path y through x in each $\gamma \in G(x)$ to the one x through y in $R[x, y](\gamma) \in G(y)$. Thus, for $z_0, \dots, z_k \in S$, $(z_k \rightarrow z_{k-1}), \dots, (z_1 \rightarrow z_0) \in \gamma \in G(z_0)$ implies that

$$R[z_0, z_k](\gamma) = \gamma \cup \{(z_0 \rightarrow z_1), \dots, (z_{k-1} \rightarrow z_k)\} \setminus \{(z_k \rightarrow z_{k-1}), \dots, (z_1 \rightarrow z_0)\} \in G(z_k).$$

In particular, from (3.2),

$$\begin{aligned} V(R[z_0, z_k](\gamma)) &= V(\gamma) + \sum_{i=1}^k V(z_{i-1} \rightarrow z_i) - \sum_{i=1}^k V(z_i \rightarrow z_{i-1}) \\ &= V(\gamma) + \sum_{i=1}^k \{\varphi(z_i) - \varphi(z_{i-1})\} \\ &= V(\gamma) + \varphi(z_k) - \varphi(z_0). \end{aligned}$$

Thus, for $x, y \in S$,

$$\begin{aligned} W(y) &= \min_{\delta \in G(y)} V(\delta) \\ &= \min_{\gamma \in G(x)} V(R[x, y](\gamma)) \\ &= \min_{\gamma \in G(x)} (V(\gamma) + \varphi(y) - \varphi(x)) \\ &= W(x) + \varphi(y) - \varphi(x), \end{aligned}$$

which means that

$$W(x) - \min_{y \in S} W(y) = \varphi(x) - \min_{y \in S} \varphi(y).$$

In other words,

$$W(x) = W_{\min} \iff \varphi(x) = \min_{y \in S} \varphi(y). \tag{A.1}$$

Lemma 3.1 follows from (3.1) and (A.1).

A.2. Proof of Lemma 3.2

From Freidlin–Wentzell [8],

- $(W(x))_{x \in S_-}$ is computed from the partial graph consisting of the vertexes in S_- ;
- letting

$$\bar{V}(x \rightarrow y) := \min_k \min_{z_1, \dots, z_{k-1}} \sum_{i=1}^k V(z_{i-1} \rightarrow z_i)(x = z_0, y = z_k),$$

we have

$$W(y) = \min_{x \in S_-} \{W(x) + \bar{V}(x \rightarrow y)\}, \quad y \in S_+.$$

From the assumption of Lemma 3.2, $V(x \rightarrow y) > 0$, $x \in S_-$, $y \in S_+$. Therefore, for each $y \in S_+$, there exists an $x \in S_-$ such that $W(y) > W(x)$, and $W(y) \neq W_{\min}$ follows. From (3.1), $q_\infty(y) = 0$ follows, which means that $x \in S_-$ if $q_\infty(x) > 0$.

References

- [1] ALBUQUERQUE, P. AND MAZZA, C. (2001). Mutation-selection algorithms: A large deviation approach. In *Foundations of Genetic Algorithms-6*, eds W. N. Martin and W. M. Spears, Morgan Kaufmann, San Francisco, pp. 227–240.
- [2] CATONI, O. (1997). Simulated annealing algorithms and Markov chains with rare transitions. *Séminaire de Probabilités XXXIII*. Lecture Notes Math. **709**, Springer, Berlin, pp. 69–119.
- [3] CERF, R. (1996). The dynamics of mutation-selection algorithms with large population sizes. *Ann. Inst. H. Poincaré Prob. Statist.* **32**, 455–508.
- [4] CERF, R. (1998). Asymptotic convergence of genetic algorithms. *Adv. Appl. Prob.* **30**, 521–550.
- [5] DAVIS, T. E. AND PRINCIPE, J. C. (1991). A simulated annealing-like convergence theory for the simple genetic algorithm. In *Proc. 4th Internat. Conf. Genetic Algorithms*, eds R. K. Belew and L. B. Booker, Morgan Kaufmann, San Mateo, CA, pp. 174–181.
- [6] DE SILVA, U. C. AND SUZUKI, J. (1995). On the stationary distribution of GAs with positive crossover probability. In *Proc. Genetic Evolutionary Comput. Conf. (GECCO 2005)*, ACM, Washington, DC, pp. 1147–1151.
- [7] FRANCOIS, O. (2002). Global optimization with exploration/selection algorithms and simulated annealing. *Ann. Appl. Prob.* **12**, pp. 248–271.
- [8] FREIDLIN, M. I. AND WENTZELL, A. D. (1984). *Random Perturbations of Dynamical Systems*. Springer, New York.
- [9] GOLDBERG, D. E. (1988) *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, Reading, MA.
- [10] RIGAL L. AND TRUFFET L. (2007). A new genetic algorithm specifically based on mutation and selection. *Adv. Appl. Prob.* **39**, 141–161.
- [11] SUZUKI, J. (1995). A Markov chain analysis of simple genetic algorithms. *IEEE Trans. Systems Man Cybernetics* **25**, 655–659.
- [12] SUZUKI, J. (1998). A further result on the Markov chain model of genetic algorithms and its application to a simulated annealing-like strategy. *IEEE Trans. Systems Man Cybernetics* **28**, 95–102.