

A MARKOV CHAIN ANALYSIS ON GENETIC ALGORITHMS - LARGE DEVIATION PRINCIPLE APPROACH -

JOE SUZUKI,* *Osaka University*

Abstract

This paper proves that the stationary distribution over the populations in genetic algorithms focuses on the uniform populations with the highest fitness value as the selective pressure goes to infinity and the mutation probability goes to zero. The obtained sufficient condition is based on Albuquerque-Mazza (2000) who followed Cerf (1998) who initiated the large deviation principle approach (Freidlin-Wentzell theory) to the Markov chain of genetic algorithms. The sufficient condition is more general than Albuquerque-Mazza and covers a set of parameters which were not found by Cerf.

Keywords: Genetic Algorithms; Markov Chains; Large Deviation Principle; Evolutionary Algorithms; Freidlin-Wentzell;

2000 Mathematics Subject Classification: Primary 65C40

Secondary 65C50

1. Introduction

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

The 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, evolutionary algorithms 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 the EAs.

In the GAs, a population containing of a definite number (population size) of

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

individuals consisting of genetics is set up first. Then, individuals are selected according to the probabilities each of which is proportional to the fitness value multiplied by the number of occurrence. This process (selection) is repeated until individuals of the population size are contained in the updated population. Then, for a pair of individuals, we exchange genetics in some locations to replace them by the new pair. This process (crossover) is repeated several times, and a further updated population is obtained. How to decide the locations, how to choose the pairs of individuals to be crossed over, etc. can be chosen arbitrarily. Finally, we randomly change several genetics in several individuals (mutation). The probability of mutation can be chosen arbitrarily.

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

1. Aim to find the individuals that optimize the fitness function.
2. Their populations are states of a finite Markov chain.

However, in general, it is considered to be hard 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 1990's. This paper analyzes GAs using large deviation principle, in particular the Freidlin-Wentzell theory [8]. The basic idea was initiated by R. Cerf [3]. Similar evaluation methods have been applied not just to GAs but also to EAs, following Cerf.

Olivier Francois [7] considered an algorithm which selects a definite number of individuals and randomly changes at most one genetic for each individual (mutation) with a probability, where the mutation probability gradually decreases as the generation change proceeds. The performance of his simulated annealing like algorithm was evaluated by the Cerf's method.

L. Rigal and L. Truffet [10] considered an algorithm which generates individuals no more than the population size 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 generation change proceeds. They proved that the stationary probability focuses on the populations each consisting of the same best individuals as the selective pressure increases.

Intuitively, decreasing the mutation probability and increasing the selective pressure are equivalent in the sense that both of them decreases the temperature in statistical dynamics. For both algorithms, it is proved that the uniform populations consisting of same individuals have positive probabilities when the temperature is low. Then, they concluded that the stationary probability focuses on the uniform populations (best uniform populations) each consisting of best individuals if the population size exceeds a definite limit. The derivation was due to 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. Consult Olivier Catoni [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 zero and the selective pressure goes to infinity, keeping the two factors balanced. The statement is similar to Cerf, but the obtained sufficient condition is different since our derivation is based on Paul Albuquerque and Christian 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 the Albuquerque-Mazza in the following sense:

1. This paper considers the effect of crossover as well as those of selection and mutation.
2. Our proof does not require a non-genetic operator to add random small value to each fitness value if more than one best individual exists.
3. The sufficient condition in this paper is more general than that of Albuquerque-Mazza.
4. While the result in Albuquerque-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, this paper shows that

the property is true for *any* population size no less than two 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 maximizes the fitness value $f(i)$ for a given fitness function $f : I \rightarrow (0, \infty)$. We keep M (≥ 2 , population size) individuals in the population $S := I^M$. 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 ones 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 $x = (x_1, \dots, x_M), y = (y_1, \dots, y_M) \in S$ by $d(x, y) := \sum_{k=1}^M \rho(x_k, y_k)$.

We repeat the cycles consisting of genetic operations: “mutation \rightarrow crossover \rightarrow selection” in this order (even 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 obtain 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 genetics $i_k, j_k, k \in \pi$ 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 obtain 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 in an arbitrary manner, crossover does not change the numbers of ones in

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

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

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

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

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

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

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

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

which converges into the individuals $i \in I$ that maximizes $f(i)$ [9]. Although it is not easy to define GAs generally, we require only the above assumptions for GAs in this paper. Then, the probability to obtain $y \in S$ from $x \in S$ is expressed by

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

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 $0 < \mu < 1, \beta > 0$.

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

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

and

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

respectively.

For stationary probabilities, several results have been known:

Proposition 2.1. (Davis-Princeple, 1991 [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. (Suzuki, 1998 [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.

This paper proves Proposition 2 even when crossover is present. In particular, we utilize the approach based on the Freidlin-Wentzell theory [8] in large deviation to evaluate the Markov chains.

In general, when the absolute temperature $T_{\mu, \beta}$ goes to zero, the stationary probability $(q_{\mu, \beta})_{\mu, \beta}$ will focus on the set $S^* := \{x \in S | W(x) = \min_{y \in S} W(y)\}$ 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, β for $S^* \subseteq U^*$. It was R. Cerf(1998) [3] 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-Mazza (2000) [1] proved $U^* \subseteq S^*$ under the conditions

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

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,

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

In this paper, we prove $U^* \subseteq S^*$ under $\lambda > \frac{M}{M-1}\Delta$.

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 one, $\lambda = 1$ is a solution of the equation $\det(\lambda I - Q) = 0$, which does not mean that the dimension of the corresponding eigenvalue space is one. 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 $\gcd\{n | P(X_n = x | X_0 = x) > 0\}$ is said to be a cycle of $x \in S$. If all the cycles of $x \in S$ are one, 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 one (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 $(Q_\alpha[x, y])_{x, y \in S}$ by $(q_\alpha(x))_{x \in S}$, and denote $q_\infty(x) := \lim_{\alpha \rightarrow \infty} q_\alpha(x)$, $x \in S$. 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 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

$$\begin{aligned} V(\gamma) &:= \sum_{(y \rightarrow z) \in \gamma} V(y \rightarrow z), \quad \gamma \in G(x), \\ W(x) &:= \min_{\gamma \in G(x)} V(\gamma), \end{aligned}$$

and

$$W_{min} := \min_{x \in S} W(x).$$

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

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

1. 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 that

2. $\lim_{\alpha \rightarrow \infty} -\frac{1}{\alpha} \log q_\alpha(x) = W(x) - W_{min}$.

Thus, for each $x \in S$, we have

$$q_\infty(x) > 0 \implies W(x) = W_{min}. \quad (1)$$

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

$$V(x \rightarrow y) - V(y \rightarrow x) = \varphi(y) - \varphi(x). \quad (2)$$

Proofs of the following lemmas are presented in the appendix 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

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

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

4. Main Result

We express the mutation probability by $\mu(\beta) := \epsilon \exp(-\lambda\beta)$ via one parameter β with constants $0 < \epsilon < 1$ and $\lambda > 0$, where $0 < \epsilon < 1$ is required for $0 < \mu(0) < 1$. Then, we apply the lemmas in the previous section, expressing $\{Q_{\mu, \beta}[x, y]\}_{x, y \in S}$ by $\{Q_\beta[x, y]\}_{x, y \in S}$.

We notice that any uniform population remains the same via selection and via crossover:

Lemma 4.1. 1. $x \in U, x \succ y \implies y \in U$

2. $x \in U, x \sim y \implies y \in U$

For $x = (x_1, \dots, x_M) \in S$, if we put

$$\varphi(x) := \sum_{k=1}^M -\log f(x_k) ,$$

then

$$\begin{aligned} & V(x \rightarrow y) \\ &= \lim_{\beta \rightarrow \infty} -\frac{1}{\beta} \log Q_\beta[x, y] \\ &= \lim_{\beta \rightarrow \infty} -\frac{1}{\beta} \log \left\{ \sum_{u, v \in S} \mathcal{M}_\beta[x, u] \mathcal{C}[u, v] \mathcal{S}_\beta[u, v] \right\} \\ &= \lim_{\beta \rightarrow \infty} -\frac{1}{\beta} \log \left\{ \sum_{u \sim v} C[u, v] \exp \left\{ -[\lambda d(x, u) + \varphi(y) - \min_{v \succ r} \varphi(r)] \beta + o(\beta) \right\} \right\} \\ &= \min_{u \sim v, v \succ y} \left\{ \lambda d(x, u) + \varphi(y) - \min_{v \succ r} \varphi(r) \right\} , \end{aligned} \tag{3}$$

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 $\lambda > \frac{M}{M-1} \Delta$, 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 $u \sim v, v \succ y$ minimizing $\lambda d(x, u) + \varphi(y) - \min_{v \succ r} \varphi(r)$

1. 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 $y \in U$ such that $V(x \rightarrow y) = 0$.
2. 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 the first item of Lemma 3, we have $v \in S \setminus U$. From the second item of Lemma 3, we have $u \in S \setminus U$. From $x \in U$, we have $d(x, u) > 0$, which means $V(x \rightarrow y) > 0$.

Thus, from Lemma 2, the dynamics in S is restricted to U .

Next, we show that $\varphi((i)) = -M \log f(i)$, $i \in I$, is a potential of U , which means from Lemma 1 to complete proof of Theorem 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 $z_0 = (i)$, $z_k = (j)$ for some k .

Suppose $\varphi((j)) \geq \varphi((i))$.

If the population contains both of i, 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 either to generate j via mutation directly, or to generate $k, h \in I$ 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 have 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 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 from $\lambda > \frac{M}{M-1}\Delta$,

$$\bar{V}((i) \rightarrow (j)) = \lambda\rho(i, j) + \varphi((j)) - \varphi((i)). \quad (4)$$

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

$$\bar{V}((i) \rightarrow (j)) = V((i) \rightarrow (j)) = \lambda\rho(i, j). \quad (5)$$

From (4)(5), for the 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 (3) by

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

under the condition $\lambda > \frac{M}{M-1} \Delta$, a statement identical to Theorem 1 is obtained.

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

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

when no crossover is present in GAs (Theorem 4 [3]), 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 .

This paper claims that such a range includes

$$\{(M, \lambda) | \lambda > \frac{M}{M-1} \Delta\} ,$$

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

Appendix

Proof of Lemma 1

For $x, y \in S$, let $R[x, y] : G(x) \rightarrow G(y)$ be the mapping from the 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 $z_0, \dots, z_k \in S$,

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

In particular, from (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

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

in other words,

$$W(x) = W_{\min} \iff \varphi(x) = \min_{y \in S} \varphi(y) . \quad (6)$$

Lemma 1 is obtained from (1)(6).

Proof of Lemma 2

From Freidlin-Wentzell [8],

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

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

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

$y \in S_+$.

From the assumption of Lemma 2, $V(x \rightarrow y) > 0$, $x \in S_-, y \in S_+$. Therefore, for each $y \in S_+$, there exists $x \in S_-$ such that $W(y) > W(x)$, and $W(y) \neq W_{\min}$ follows.

From (1), $q_\infty(y) = 0$ follows, which means $x \in S_-$ 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*, Ed. W.N. Martin and W.M. Spears, Morgan Kaufmann Publishers, San Francisco, p. 227-240.
- [2] Olivier Catoni (1997). Simulated Annealing Algorithms and Markov chains with Rare Transitions, *Seminaire de Probabilites XXXIII*, volume 709 of Lecture Notes in Mathematics
- [3] Cerf, R. (1998). Asymptotic convergence of genetic algorithms. *Adv. Appl. Prob.* 30, 521–550.
- [4] Cerf, R. (1996). The dynamics of mutation-selection algorithms with large population sizes. *Annles de l 'Institut Henri Poincare* 32(4), 455?508.
- [5] Davis, T.E. and Principe, J.C. (1991). A simulated annealing-like convergence theory for the simple genetic algorithm, in: R.K. Belew, L.B. Booker (Eds.), *Proc. Fourth Internat. Conf. on Genetic Algorithms '91*, Morgan Kaufmann Publishers, San Francisco, CA, USA, pp. 174-181.
- [6] DeSilva, C. and Suzuki, J. (1995) On the Stationary Distribution of GAs with Positive Crossover Probability, *GECCO*, Washington DC.
- [7] O. Francois (2002). Global optimization with exploration/selection algorithms and simulated annealing, *The Annals of Applied Probability*, vol 12, pp. 248-271, 2002.
- [8] Freidlin, M.I. and Wentzell, A.D. (1984). *Random Perturbations of Dynamical Systems*, Springer-Verlag.
- [9] GOLDBERG, DAVID E. (1988) *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison Wesley.
- [10] L. Rigal and L. Truffet (2007). A new genetic algorithm specifically based on mutation and selection, *Adv. in Appl. Probab.* Volume 39, Number 1, 141-161
- [11] Suzuki, J. (1995). A Markov Chain analysis of simple genetic algorithms, *IEEE Trnasactions on Systems, Man, and Cybernetics*, vol. 25, no. 4, p. 655, 1995.
- [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 Transactions on Systems, Man and Cybernatics*, 28(1)