MATHEMATICAL ANALYSIS OF MUTATION OPERATOR
AND ITS IMPROVED STRATEGY
IN GENETIC ALGORITHMS*

Zhang Liangjie  Mao Zhihong  Li Yanda
(Dept. of Automation, Tsinghua Univ., Beijing, 100084)

Abstract   This paper analyzes the optimization problem of mutation probability in genetic
algorithms by applying the definition of i-bit improved sub-space. Then fuzzy reasoning technique
is adopted to determine the optimal mutation probability in different conditions. The superior
convergence property of the new method is evaluated by applying it to two simulation examples.

Key words   Genetic algorithm(GA); i-bit improved sub-space; Fuzzy reasoning

I. Introduction

Genetic algorithms (GA's) are global and adaptive optimization techniques based on
the mechanics of natural selection and natural genetics. They are often used to tackle the
static optimization problems of the type:

$$\min \{ f(C) | C \in IB^N \}$$  (1)

assuming that for \( \forall C \in IB^N = \{0,1\}^N \), \( 0 < f(C) < \infty \) and \( f(C) \neq \text{const.} \). GA's include the
steps that are very similar to biological evolution, such as selection, reproduction, crossover,
mutation, etc.. These steps are called genetic operators. By the operator of mutation,
GA's may reach each point in the solution space with non-zero probability, and they may
converge to the global optimum if the best solution of a generation is always maintained in
the "offspring" [1]. Moreover, in a generation of \( G \) "gene-strings", the number of schemata
that GA's deal with is \( O(G^3) \), so they are a kind of implicit parallelisms [2]. GA's have
successfully been applied to various domains, such as machine learning, pattern recognition,
network training [3-5], fuzzy logic controller design [6], etc..

However, further research is required to overcome the disadvantages in canonical GA's [7].
For example, in canonical GA's, there are no universal principles to determine the popula-
tion size and the mutation probability \( (P_m) \); there are no suitable rules to select the crossover
point in one-point crossover way, and so on. All this restricts the convergence rate of GA's.
This paper makes mathematical analysis on the mutation operator in GA's and proposes a
strategy to improve it.

II. Analysis of Mutation Operator and Its Improved Strategy

Mutation operates independently on each individual by perturbing each bit with some
probability. The event that each bit of an individual is flipped is stochastically independent
and occurs with probability \( P_m \) (mutation probability). In canonical GA's, \( P_m \) is constant,

---

*Supported by the Climbing Program — National Key Project for Fundamental Research in China, Grant NSC92097
and at present there is no satisfied principle to determine a suitable $P_m$. From many simulation experiments, we find that high evolution rate requires varied $P_m$ in different conditions. So in this section we will adopt fuzzy reasoning technique to direct a varied $P_m$ based on a mathematical analysis of $P_m$.

Considering a string $C_0$, suppose that it turns to $C_1$ after mutation. Obviously, a great probability with which $C_1$ is “better” than $C_0$ will help the evolution performs quickly. So in Eq.(1), we use $P(f(C_1) < f(C_0))$ as a standard to evaluate the evolution rate. Before a further discussion of this probability, let us first see two definitions.

**Definition 1** The Hamming distance between two binary strings $C_1, C_2 \in \{0,1\}^N$ is defined as

$$H(C_1, C_2) = \sum_{i=1}^{N} |B_i(C_1) - B_i(C_2)|$$

where $B_i(C)$ denotes the code (0 or 1) of the $i$-th bit in string $C$, $i = 1, 2, \ldots, N$.

**Definition 2** In Eq.(1), define the following classes to string $C_0$, $\Omega_0^{(1)}, \Omega_0^{(2)}, \ldots, \Omega_0^{(i)}, \ldots$, where $\Omega_0^{(i)} = \{C | C \in \{0,1\}^N, f(C) < f(C_0), H(C, C_0) = i\}$. $\Omega_0^{(i)}$ is called the $i$-bit improved sub-space of $C_0$, $\bigcup_{i=1,2,\ldots} \Omega_0^{(i)}$ is called the improved space of $C_0$ and is denoted as $\bigcup_i \Omega_0^{(i)}$ for simplification.

Obviously, if $C_1 \in \bigcup_i \Omega_0^{(i)}$, then $C_1$ is “better” than $C_0$; otherwise there is no improvement from $C_0$ to $C_1$. The probability that $C_1$ is “better” than $C_0$ is equal to the probability that $C_1 \in \bigcup_i \Omega_0^{(i)}$, i.e.,

$$P(f(C_1) < f(C_0)) = P(C_1 \in \bigcup_i \Omega_0^{(i)})$$

The selection of $P_m$ should makes this probability as great as possible. That is the principle to direct the varied $P_m$. From $C_0$ to $C_1$, there are $H(C_0, C_1)$ bits performing mutation, so, $P(C_0 \rightarrow C_1) = P_m^{H(C_0, C_1)}(1 - P_m)^{N - H(C_0, C_1)}$, thus

$$P(C_1 \in \Omega_0^{(i)}) = \frac{|\Omega_0^{(i)}|}{P_m^i(1 - P_m)^{N - i}}$$

where $|\Omega_0^{(i)}|$ is the number of elements in $\Omega_0^{(i)}$. Since $\Omega_0^{(i)} \cap \Omega_0^{(j)} = \emptyset, i \neq j$, we have

$$P(C_1 \in \bigcup_i \Omega_0^{(i)}) = \sum_i |\Omega_0^{(i)}|P_m^i(1 - P_m)^{N - i}$$

**Theorem** $P(C_1 \in \Omega_0^{(i)})$ reaches its maximum at $P_m = i/N$.

**Proof**

From Eq.(4), we have

$$\ln P(C_1 \in \Omega_0^{(i)}) = \ln |\Omega_0^{(i)}| + \ln P_m + (N - i)\ln(1 - P_m)$$

let $\partial(\ln P(C_1 \in \Omega_0^{(i)}))/\partial P_m = i/P_m - (N - i)/(1 - P_m) = 0$, then $P_m = i/N$.

Since $\partial^2(\ln P(C_1 \in \Omega_0^{(i)}))/\partial^2 P_m = -i/P_m^2 - (N - i)/(1 - P_m)^2 < 0$, $\ln P(C_1 \in \Omega_0^{(i)})$ is convex when $P_m \in (0,1)$ and reaches its maximum at the stationary point. So $P(C_1 \in \Omega_0^{(i)})$ reaches its maximum at $P_m = i/N$. Q.E.D.

From many stimulation experiments, we find: in the late period of iterations, if $P_m$ is too small, the search will not converge rapidly, and on the other hand, if $P_m$ is selected big, the search will not converge easily because there are several bits performing mutation simultaneously in a string; at the moment when the search falls into a trap of local, if $P_m$ is selected small now, it will be difficult to break away from the local optimum. So