

Theoretical Comparisons of Search Dynamics of Genetic Algorithms and Evolution Strategies

Tatsuya Okabe

Honda R&D Co., Ltd.,
Wako Research Center
1-4-1 Chuo, Wako-shi, Saitama
351-0193, Japan
tatsuya_okabe@n.w.rd.honda.co.jp

Yaochu Jin

Honda Research Institute Europe
Carl-Legien Strasse 30
63073 Offenbach am Main
Germany
yaochu.jin@honda-ri.de

Bernhard Sendhoff

Honda Research Institute Europe
Carl-Legien Strasse 30
63073 Offenbach am Main
Germany
bernhard.sendhoff@honda-ri.de

Abstract- Genetic algorithms (GAs) and evolution strategies (ESs) are two widely used evolutionary algorithms. The main differences between GAs and ESs lie in their representations and variation operators, which result in very different search dynamics. In this paper, we compare the search dynamics of GAs and ESs theoretically using a theoretical framework for analyzing the search dynamics of evolution strategies proposed in this paper and a framework for genetic algorithms we suggested in [Oka05]. Based on the theoretical analysis, interesting aspects of the search dynamics of GAs and ESs for single objective optimization are revealed. As an extension, preliminary results on the search dynamics of GAs for multi-objective optimization are also presented.

1 Introduction

Both genetic algorithms (GAs) and evolution strategies (ESs) have shown to be effective in solving parameter optimization problems¹. The main differences between GAs and ESs lie in the representation and the variation operators. Canonical GAs use bit-strings to represent parameters to be optimized, whereas ESs adopt real-valued coding². Due to their different representations and different variation operators, the way of generating offspring, in other words, the search dynamics of GAs and ESs, is also very different. While crossover is considered as the major genetic operator in GAs, mutation plays the major role in ESs.

Little work has been done on a theoretical comparison of GAs and ESs. This paper aims to understand the differences in search dynamics of GAs and ESs theoretically. It is hoped from these understandings that we are able to know better whether GAs or ESs are more likely to show good performance for a given problem. To simplify the analysis, we focus on the major genetic operator, i.e. the crossover in GAs and the mutation in ESs, without selection.

This paper is structured as follows. Section 2 introduces the related work on theoretical analysis of GAs and ESs. Section 3 revisits the theoretical framework for analyzing the search dynamics of genetic algorithms we pro-

posed in [Oka05]. Section 4 presents a theoretical framework for analyzing the dynamics of evolution strategies. The search dynamics of GAs and ESs are compared in Section 5. Although this paper mainly addresses the search dynamics of evolutionary algorithms for single objective optimization, preliminary results on the search dynamics of GAs for multi-objective optimization are also presented in Section 6. Finally, a summary of this paper is given in Section 7.

2 Related Work

A wide range of research for the dynamics analysis of evolutionary algorithms has been reported, which can roughly be divided into cumulant-based and model-based approaches.

The cumulant-based approach describes population distribution using cumulants, e.g. average, deviation, skewness, kurtosis etc., and analyzes the dynamics of the population distribution by observing the change of the cumulants. The advantage of this approach is that it becomes rather easy to observe the population dynamics, although much information will be lost by representing the population with the cumulants. The cumulant-based approach is well suited for the dynamics analysis of evolutionary algorithms in single objective optimization. However, it is insufficient for describing the dynamics of evolutionary algorithms in multi-objective optimization due to the fact that population distributions in multi-objective optimization can no longer be simplified to a normal distribution, refer to [Oka02, Oka04] for some examples.

Dynamics analysis of GAs focus on the roles of genetic operators, e.g., two-point crossover [Pru01], recombination [Pru01b], and selection schemes, such as roulette wheel and stochastic universal sampling [Rog97], generational and steady state selections [Rog99], and the stabilizing selection [Rat95].

Beyer and his colleagues have conducted systematic investigations on the dynamics of ESs [Bey01, Arn02]. On a sphere function where the objective is to minimize the distance between the current point to the optimal point, the mean of the population distribution is investigated. Based on the mean value, the progress rate and the progress gain are discussed for (μ, λ) and $(\mu + \lambda)$ evolution strategies [Bey93, Bey94, Bey95, Bey96]. The dynamics of ESs in noisy environments has been considered in [Bey00, Arn02, Bey03, Bey04]. Nürnberg and Beyer [Nur97] have

¹Evolutionary programming (EP) [Fog99] is also one of powerful methods in evolutionary computation. However, EP is beyond the scope of this paper.

²There are also variants of GAs that use a real-valued representation. These GAs are known as *Real-coded GAs*, which are essentially equivalent to ESs.

also modeled the dynamics of ESs for the optimization of traveling salesman problems and analyzed the dynamics of the tour length.

The model-based approach analyzes the dynamics of evolutionary algorithms by building a model. As far as we know, the model-based approach has mainly been used in analyzing the dynamics of GAs and no work on model-based approach to the dynamics of ESs has been reported.

One of the first models for GAs was introduced in [Gol89], where a model for canonical GAs with two-bit strings was built. Vose [Vos99] extended Goldberg's model to an arbitrary number of strings. To store the information on the population distribution, a probability vector is introduced whose element indicates the probability of a certain chromosome. In both models, the influence of genetic operators is investigated by observing the change of the probability. To address the *finite population effect* in the probability based models, *Markov chain* has also been employed to analyze population dynamics. All possible states of the population are stored in a state vector, which is changed by a transition matrix that reflects the role of crossover, mutation and selection. Modeling GAs with Markov chain is also studied in [Suz98] and convergence analysis of canonical GAs with Markov chain is investigated in [Fog92, Rud94].

A detailed review of existing theoretical work on analyzing dynamics of evolutionary algorithms is beyond the scope of this paper. Interested readers are referred to [Vos99, Bey01, Arn02, Ree03].

Despite the large amount of work on theoretical work on dynamics analysis of GAs and ESs, theoretical comparisons of the search dynamics of GAs and ESs have not been conducted, to the best of our knowledge. This paper suggests a framework for analyzing search dynamics of ESs and compares the dynamics of ESs with that of GAs using the theoretical framework for GAs [Oka05]. In addition, preliminary results on dynamics analysis of GAs in multi-objective optimization are also presented.

3 Dynamics of Genetic Algorithms

In this section, we briefly review the theoretical framework for analyzing dynamics of GAs suggested in [Oka05].

Let us denote the ratio between the number of individuals with the chromosome ζ against the total number of individuals in a population as $P(\zeta)$, where ζ is the index of chromosome Z that can be defined by $\zeta = \sum_{i=1}^l B(Z, i) \times 2^{l-i}$. Here, $B(Z, i)$ means the i -th component of the chromosome Z . $P(\zeta)$ is termed as the probability of chromosome ζ . Consequently, we have $\sum_{\zeta} P(\zeta) = 1$.

In the theoretical framework introduced in [Oka05], we calculate the probability of a certain chromosome at generation $(t + 1)$ according to the probability distribution at generation (t) . Consider the probability of chromosome $Z = \{\#\#\dots\#\# * \dots*\}$, of which the last i alleles are the *don't care symbol*, $*$. The symbol $\#$ denotes alleles that should keep constant. To avoid confusion, we denote Z as ζ_i^b , which can be defined by $\zeta_i^b = \sum_{j=1}^{l-i} B(Z, j) \times 2^{l-j-i}$, where, the superscript of b and the subscript of i mean that

the i loci counted *backwards* are the *don't care symbol*. The probability of $\zeta_i^b = k$, denoted by $P_i^b(k)$, can be calculated recursively from the probabilities of ζ_{i-1}^b as follows:

$$P_i^b(k) = \begin{cases} P(2k) + P(2k + 1) & \text{if } i = 1 \\ P_{i-1}^b(2k) + P_{i-1}^b(2k + 1) & \text{if } i \neq 1 \end{cases} \quad (1)$$

Similarly, the probability of $Z = \{*\dots * \#\#\dots\#\#\}$ where the first i alleles are *don't care symbol* is considered. We denote Z as $\zeta_i^f = \sum_{j=1+i}^l B(Z, j) \times 2^{l-j}$, here the superscript of f and the subscript of i mean that the *first* i loci are the *don't care symbol*. The probability of $\zeta_i^f = k$ can be calculated based on the probability of ζ_{i-1}^f as follows:

$$P_i^f(k) = \begin{cases} P(k) + P(k + 2^{l-i}) & \text{if } i = 1 \\ P_{i-1}^f(k) + P_{i-1}^f(k + 2^{l-i}) & \text{if } i \neq 1. \end{cases} \quad (2)$$

With Equations (1) and (2), the generative probability of Z , i.e., the probability of Z after crossover (denoted by $\xi(Z)$) can be calculated as follows:

$$\xi(Z) = P(Z) \times (1.0 - P_c) + \sum_{k=1}^{l-1} P_{l-k}^b(\zeta_{l-k}^b) \times P_k^f(\zeta_k^f) \times P_c \times \frac{1}{l-1}. \quad (3)$$

Here, P_c is crossover rate. The first term of Equation (3) is the probability where exactly the same chromosome, Z , is selected and no crossover occurs. The second term describes the probability where two proper chromosomes are chosen as parents and crossover occurs at the proper position. With Equation (3), the probability of all chromosome after crossover, $\xi(Z)$, can be calculated by using the probability of the chromosome before crossover.

4 Dynamics of Evolution Strategies

4.1 Definitions

Before discussing the dynamics of ESs, a number of definitions need to be introduced.

First, the probability density function (PDF) of an one-dimensional normal distribution, denoted by $\phi(x)$, can be expressed as follows:

$$\phi(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad (4)$$

where, μ and σ are the mean and the standard deviation. When $\mu = 0$ and $\sigma = 1$, this distribution is termed *Standard Normal Distribution*. We denote the PDF of the standard normal distribution by $\phi_0(x)$.

Now, we can define the cumulative density function (CDF) by:

$$\Phi(x) = \int_{s=-\infty}^{s=x} \phi_0(s) ds = \frac{1}{\sqrt{2\pi}} \int_{s=-\infty}^{s=x} \exp\left(-\frac{1}{2}s^2\right) ds. \quad (5)$$

The CDF $\Phi(x)$ describes the probability of a variable smaller than x . The integration of the PDF in the whole range, i.e., $[-\infty, \infty]$, should be 1. Thus, the following equation holds:

$$\int_{s=x}^{s=\infty} \phi_0(s) ds = 1 - \Phi(x). \quad (6)$$

We also define the *Gaussian Integral* $\Phi_0(x)$ as:

$$\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_{s=0}^{s=x} \exp\left(-\frac{1}{2}s^2\right) ds. \quad (7)$$

For analyzing the dynamics of ESs, the following *Error Function* $\text{erf}(x)$ is also used:

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_{s=0}^{s=x} \exp(-s^2) ds. \quad (8)$$

The integration of the *error function* can be expressed as:

$$\begin{aligned} \int \text{erf}(x) dx &= \int \frac{2}{\sqrt{\pi}} \int_{s=0}^{s=x} \exp(-s^2) ds dx \\ &= \frac{1}{\exp(x^2) \sqrt{\pi}} + x \text{erf}(x). \end{aligned} \quad (9)$$

Note that the *Gaussian Integral* and the *Error Function* are closely related and the following relationship should be satisfied:

$$\Phi_0(x) = \frac{1}{2} \text{erf}\left(\frac{x}{\sqrt{2}}\right), \quad \text{erf}(x) = 2\Phi_0(\sqrt{2}x). \quad (10)$$

The CDF can now be calculated from $\Phi_0(x)$ or $\text{erf}(x)$ as follows:

$$\begin{aligned} \Phi(x) &= \Phi_0(0) + \Phi_0(x) = \frac{1}{2} + \Phi_0(x) \\ \Phi(x) &= \frac{1}{2} + \Phi_0(x) = \frac{1}{2} \left(1 + \text{erf}\left(\frac{x}{\sqrt{2}}\right)\right). \end{aligned} \quad (11)$$

4.2 Analysis of ES Search Dynamics

In ESs, the mutation is carried out by adding a normally distributed random number to an objective parameter. We assume that all individuals have the same global step size σ for simplicity.

We first assume that the number of individuals in the parent population is m , and the location of the individuals is denoted by x_i , $\forall i = 1, 2, \dots, m$. The probability of x_i is denoted as $P(x_i)$. An illustration of the situation where 5 individuals exist in one dimensional search space is plotted in Figure 1.

Since the offspring distribution generated by the parents is also Gaussian, the PDF of offspring, $\psi(x)$, can be calculated by summing up all Gaussian distributions over the m parents:

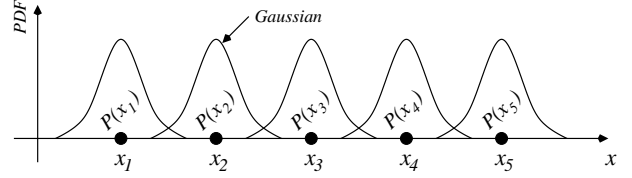


Figure 1: An illustrative ES population where the population size is 5. The probability of each individual is $P(x_i)$.

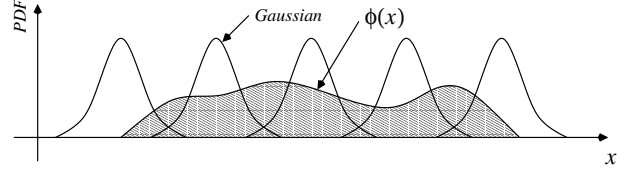


Figure 2: The PDF, $\phi(x)$ of parent when the population size is infinite.

$$\psi(x) = \sum_{i=1}^m P(x_i) \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-x_i)^2}{2\sigma^2}\right). \quad (12)$$

Although the parent individuals are located discretely, the locations of the offspring are continuous because the Gaussian distribution is continuous. Now, we assume that the number of parent individuals is infinite, and consequently, the locations of the parent become also continuous, see Figure 2. Thus, the Equation (12) can be modified as follows:

$$\begin{aligned} \psi(x) &= \lim_{m \rightarrow \infty} \sum_{i=1}^m P(x_i) \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-x_i)^2}{2\sigma^2}\right) \\ &= \int_{y=-\infty}^{y=\infty} \phi(y) \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-y)^2}{2\sigma^2}\right) dy, \end{aligned} \quad (13)$$

where, $\phi(y)$ is the PDF of the parent population.

As a simple example, we can calculate the PDF of offspring $\psi(x)$ assuming that the parents are uniformly distributed within $[x_s, x_e]$. In this case, the PDF of parents, $\phi(x)$, can be defined by:

$$\phi(x) = \begin{cases} (x_e - x_s)^{-1} & \text{if } x \in [x_s, x_e] \\ 0 & \text{otherwise} \end{cases}. \quad (14)$$

Then, the PDF of the offspring distribution after mutation can be calculated by:

$$\begin{aligned} \psi(x) &= \int_{y=x_s}^{y=x_e} \frac{1}{x_e - x_s} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-y)^2}{2\sigma^2}\right) dy \\ &= \frac{1}{x_e - x_s} \int_{y=x_s}^{y=x_e} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-y)^2}{2\sigma^2}\right) dy. \end{aligned} \quad (15)$$

If we replace $(y-x)/\sigma$ by $s = (y-x)/\sigma$ and since $dy/ds = \sigma$, we can transform the above equation into the following form:

$$\psi(x) = \frac{1}{x_e - x_s} \frac{1}{\sqrt{2\pi}} \int_{s=\frac{x_s-x}{\sigma}}^{s=\frac{x_e-x}{\sigma}} \exp\left(-\frac{1}{2}s^2\right) ds. \quad (16)$$

With the help of *Gaussian integral* $\Phi_0(x)$ or the *error function* $\text{erf}(x)$, the above equation can be further simplified:

$$\psi(x) = \frac{1}{x_e - x_s} \left\{ \Phi_0\left(\frac{x_e - x}{\sigma}\right) - \Phi_0\left(\frac{x_s - x}{\sigma}\right) \right\}, \quad (17)$$

$$\psi(x) = \frac{1}{2(x_e - x_s)} \left(\text{erf}\left(\frac{x_e - x}{\sigma\sqrt{2}}\right) - \text{erf}\left(\frac{x_s - x}{\sigma\sqrt{2}}\right) \right). \quad (18)$$

Figure 3 shows examples of the PDF of offspring distribution, $\psi(x)$, given different values of σ . Obviously, the search dynamics of the mutation, in other words, the distribution change caused by the Gaussian mutation, is quite different to that of crossover. One main difference is that the probability for offspring that are located outside $[x_s, x_e]$ is non-zero and its maximum value is $(x_s + x_e)/2$.

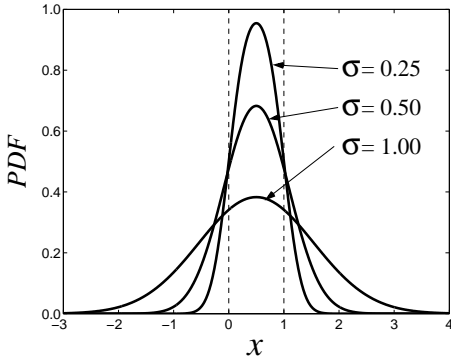


Figure 3: Examples of PDFs of offspring distribution. In these examples, $[x_s, x_e] = [0.0, 1.0]$. Offspring distributions when $\sigma = 1.0, 0.5, 0.25$ are shown.

5 Comparison of Dynamics

5.1 Dynamics of GAs

In this section, we study the crossover dynamics of binary-coded GAs and Gray-coded GAs. We compare the *Linkage Equilibrium* [Oka05], i.e., the converged probability distribution when the number of generation approaches to infinite. Three initial parent distributions are used, which are illustrated in Figures 4. In these case studies, we use 15 bits for both binary and Gray coding.

Figures 5(a), (c) and (e) show the crossover dynamics of binary-coded GAs, while Figures 5(b),(d) and (f) the dynamics of Gray-coded GAs using our theoretical framework. In all the figures, the dotted lines denote the parent probability distribution and the solid lines the offspring probability distribution.

To verify the theoretical results, a GA of population size 100 and chromosome length 15 is run for 1000 times. The average converged probability distributions are plotted in Figure 6. It is noticed that the empirical results agree with the theoretical results very well.

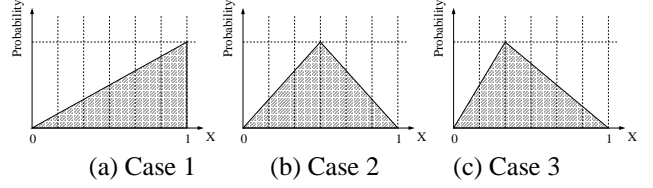


Figure 4: Examples of parent distributions used in comparing dynamics of binary-coded GAs and Gray-coded GAs. Case 1: Probability increases linearly. Case 2: Probability increases linearly until $x = 0.5$ and decreases linearly. Case 3: Probability increases linearly until $x = 1/3$ and decreases linearly.

From these figures, it can be seen that binary-coded GAs with one-point crossover has a higher search probability in the areas near the two boundaries. This may imply that binary GAs are more suited for problems whose optimal solutions are located near the boundaries of the search space, see Figures 5 (a) and (b). On the contrary, Gray-coded GAs search more often in the middle of the coded parameter space, compare Figures 5 (c) with (d) or (e) with (f).

5.2 Dynamics of ES

We now investigate the dynamics of ESs resulting from the Gaussian mutation using Equation (13). We again use three initial PDFs, $\phi(x)$, as the parent distribution. Unlike in Section 5.1, probability density functions are used instead of the probability of all possible chromosomes due to the continuity of the search space in ESs. In all three cases, the initial probability density outside $[0,1]$ is 0.

Case 1 : In this case, the upper bound of the initial search space has the highest probability and the lower bound has the lowest probability density, which can be described as follows:

$$\phi(x) = 2x \quad (0 \leq x \leq 1). \quad (19)$$

Case 2 : The probability density on both upper and lower bounds is 0, and the highest probability density is in the center of the initial search space.

$$\phi(x) = \begin{cases} 4x & (0 \leq x \leq 0.5) \\ 4 - 4x & (0.5 \leq x \leq 1) \end{cases} \quad (20)$$

Case 3 : Similar to Case 2, both upper and lower bounds have the lowest probability density, however, the highest density is located at $x = \frac{1}{3}$.

$$\phi(x) = \begin{cases} 6x & (0 \leq x \leq \frac{1}{3}) \\ 3 - 3x & (\frac{1}{3} \leq x \leq 1) \end{cases} \quad (21)$$

Given the above three initial PDFs, we can calculate the PDFs of offspring with Equation (13):

Case 1 :

$$\psi(x) = 2\sigma \frac{1}{\sqrt{2\pi}} \left\{ -\exp\left(-\frac{1}{2}\left(\frac{1-x}{\sigma}\right)^2\right) \right.$$

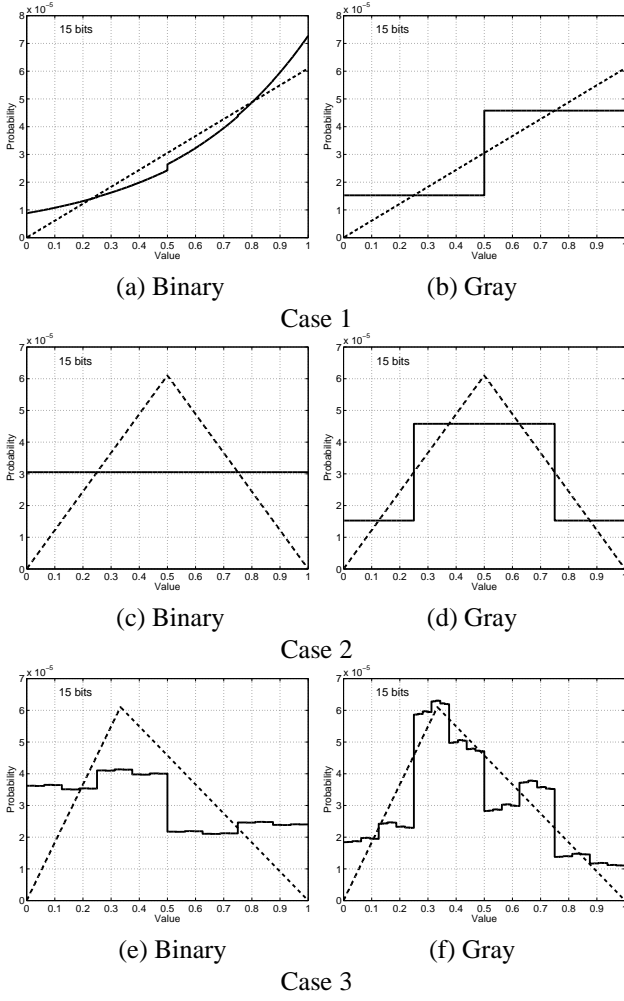


Figure 5: Theoretical results on the converged probability distribution of GAs with a chromosome length of 15. (a) and (b) for Case 1, (c) and (d) for Case 2, and (e) and (f) for Case 3.

$$\begin{aligned}
 & + \exp\left(-\frac{1}{2}\left(\frac{-x}{\sigma}\right)^2\right) \Big\} \\
 & + x \left(\operatorname{erf}\left(\frac{1-x}{\sqrt{2}\sigma}\right) - \operatorname{erf}\left(\frac{-x}{\sqrt{2}\sigma}\right) \right). \quad (22)
 \end{aligned}$$

Case 2 :

$$\begin{aligned}
 \psi(x) = & 4\sigma \frac{1}{\sqrt{2\pi}} \left\{ \exp\left(-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right) \right. \\
 & + \exp\left(-\frac{1}{2}\left(\frac{1-x}{\sigma}\right)^2\right) - 2 \exp\left(-\frac{1}{2}\left(\frac{x-0.5}{\sigma}\right)^2\right) \Big\} \\
 & + \operatorname{erf}\left(\frac{1-x}{\sqrt{2}\sigma}\right) - \operatorname{erf}\left(\frac{-x}{\sqrt{2}\sigma}\right) \\
 & + 2\left(x - \frac{1}{2}\right) \left\{ 2 \operatorname{erf}\left(\frac{0.5-x}{\sqrt{2}\sigma}\right) - \operatorname{erf}\left(\frac{-x}{\sqrt{2}\sigma}\right) - \operatorname{erf}\left(\frac{1-x}{\sqrt{2}\sigma}\right) \right\} \quad (23)
 \end{aligned}$$

Case 3 :

$$\begin{aligned}
 \psi(x) = & \sigma \frac{1}{\sqrt{2\pi}} \left\{ -9 \exp\left(-\frac{1}{2}\left(\frac{\frac{1}{3}-x}{\sigma}\right)^2\right) \right. \\
 & + 3 \exp\left(-\frac{1}{2}\left(\frac{1-x}{\sigma}\right)^2\right) + 6 \exp\left(-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right) \Big\}
 \end{aligned}$$

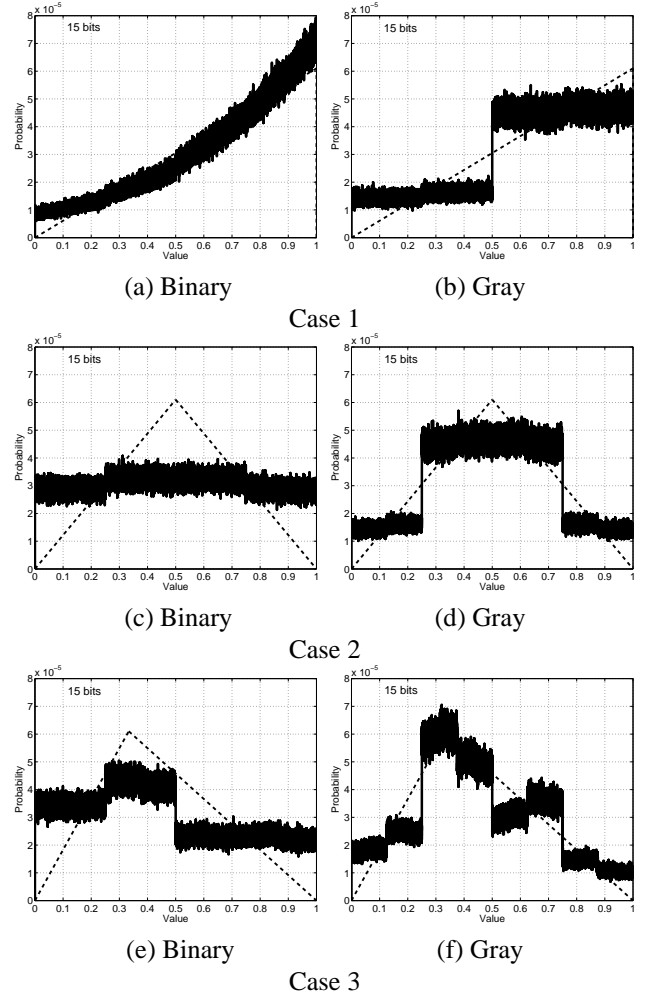


Figure 6: Empirical results of the converged probability distribution with a chromosome length of 15. (a) and (b) for Case 1, (c) and (d) for Case 2 and (e) and (f) for Case 3.

$$\begin{aligned}
 & + \frac{3}{2} \operatorname{erf}\left(\frac{1-x}{\sqrt{2}\sigma}\right) - \frac{3}{2} \operatorname{erf}\left(\frac{\frac{1}{3}-x}{\sqrt{2}\sigma}\right) \\
 & + \frac{3}{2} x \left\{ 3 \operatorname{erf}\left(\frac{\frac{1}{3}-x}{\sqrt{2}\sigma}\right) - \operatorname{erf}\left(\frac{1-x}{\sqrt{2}\sigma}\right) - 2 \operatorname{erf}\left(\frac{-x}{\sqrt{2}\sigma}\right) \right\} \quad (24)
 \end{aligned}$$

The PDFs of the population for 2 different σ s are plotted in Figure 7. We can see that the probability density is very similar to the initial one given a small step-size σ . We also notice from Figures 7 (a) and (b) that the highest probability density is shifted from the boundary ($x = 1$) towards the center of the search space even if the step-size is small. The same trend can be also seen in Figures 7 (e) and (f) but in Figures 7 (c) and (d). Thus, it is not the most efficient for ESs when the optimum is located near the boundaries of the search space.

Unlike GAs, the probability density outside the initial range is larger than 0, particularly when the step-size is large. This means that ES will search outside the initial range, while GAs can never search outside the coded search space. This topic will be discussed further in the next section.

We also verified the theoretical results empirically. The

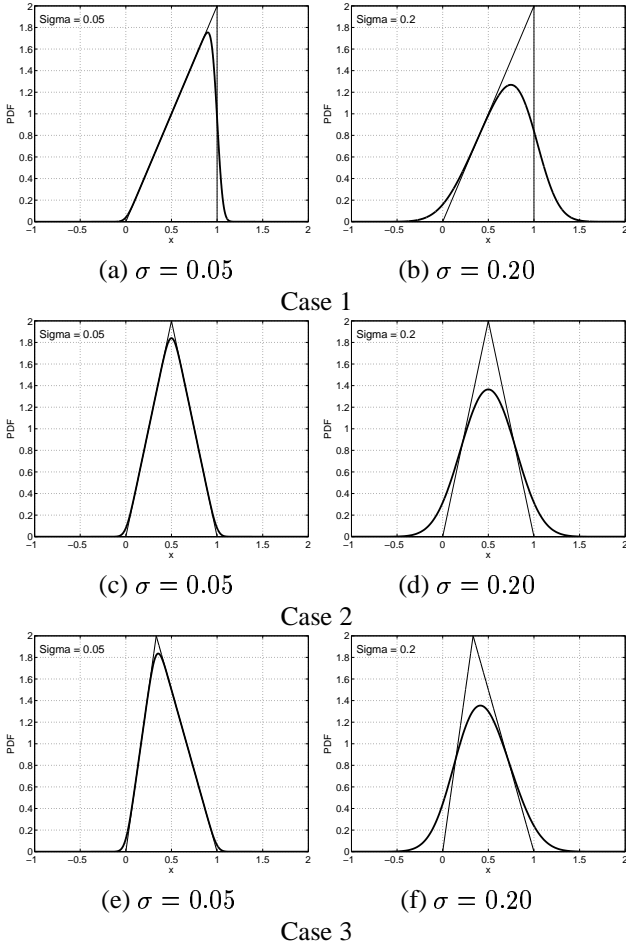


Figure 7: Probability density of ES using mutation. The initial parental distribution is denoted with thinner lines.

empirical results are completely the same as the theoretical results. Therefore, the empirical results are not plotted in the figures.

5.3 Comparing GAs and ESs

One main difference in search dynamics between GAs and ESs is that ESs have a non-zero search probability outside the initial range, while the search space of GAs is implicitly, but strictly bounded by the initial coding range.

Assuming that the parents are uniformly distributed within $[x_s, x_e]$, we can calculate the probability with which an offspring will be located outside $[x_s, x_e]$ using Equation (13). We denote this probability as $\Psi(x|x \notin [x_s, x_e])$:

$$\begin{aligned} \Psi(x|x \notin [x_s, x_e]) &= \Psi(x|x < x_s) + \Psi(x|x_e < x) \\ &= \int_{x=-\infty}^{x=x_s} \psi(x)dx + \int_{x=x_e}^{x=\infty} \psi(x)dx. \end{aligned} \quad (25)$$

The following equation can be derived regarding the probability $\Psi(x)$:

$$\Psi(x|x < x_s) = \Psi(x|x_e < x). \quad (26)$$

A proof of the equation is given in Appendix A. Using Equation (26), we can obtain the probability at which offspring is outside $[x_s, x_e]$:

$$\begin{aligned} \Psi(x|x \notin [x_s, x_e]) &= \frac{1}{x_e - x_s} \int_{x=x_e}^{x=\infty} \left(\operatorname{erf}\left(\frac{x_e - x}{\sigma\sqrt{2}}\right) - \operatorname{erf}\left(\frac{x_s - x}{\sigma\sqrt{2}}\right) \right) dx \\ &= \frac{1}{x_e - x_s} \int_{x=-\infty}^{x=x_s} \left(\operatorname{erf}\left(\frac{x_e - x}{\sigma\sqrt{2}}\right) - \operatorname{erf}\left(\frac{x_s - x}{\sigma\sqrt{2}}\right) \right) dx. \end{aligned} \quad (27)$$

We now present some examples of the probability at which the ESs search the space outside the initial range $[x_s, x_e]$ given different step-sizes. The results are shown in Table 1. By defining a new variable $\sigma/(x_e - x_s)$, we can plot the relationship between the probability $\Psi(\cdot)$ and $\sigma/(x_e - x_s)$, see Figure 8. From the figure, we can see that the search of ESs will be local if the step-size σ is small. When the step-size increases, a more global search will be performed in that the probability at which ESs search outside the initial range becomes larger.

Table 1: Examples of the probability out of $[x_s, x_e]$.

σ	$[x_s, x_e] = [0.0, 1.0]$ $\Psi(x x \notin [x_s, x_e])$	σ	$[x_s, x_e] = [0.0, 2.0]$ $\Psi(x x \notin [x_s, x_e])$
0.01	0.007979	0.02	0.007979
0.10	0.079789	0.20	0.079789
0.25	0.199468	0.50	0.199468
0.50	0.390452	1.00	0.390452
0.75	0.534821	1.50	0.534821
1.00	0.631254	2.00	0.631254
1.50	0.743468	3.00	0.743468
2.00	0.804583	4.00	0.804583
4.00	0.900781	8.00	0.900781

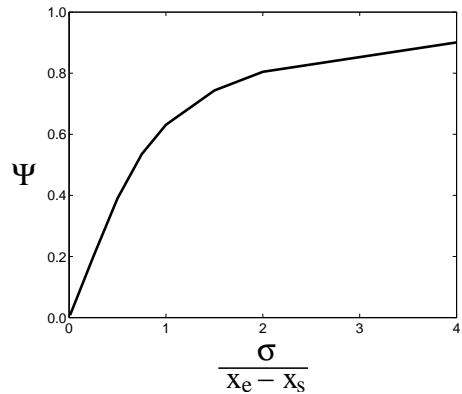


Figure 8: The probability at which ESs search the region outside $[x_s, x_e]$ vs. $\sigma/(x_e - x_s)$.

Comparing the search dynamics of GAs and ESs, we can draw the following conclusions:

- GAs are not able to search the space outside the initial coding range³, thus GAs will completely fail if the

³Even if the mutation is used, GAs cannot search the space outside the initial coding range.

optimum is not located in the coded range.

- ESs can search region outside the initial range with a probability depending on the step-size. The advantage is that ESs are still able to find an optimal solution if the optimum is not in the initial area. The disadvantage is, however, when the optimum is in the initial range and if a large step-size is used, a part of the search efforts will be “wasted”.

6 Preliminary Results on Multi-Objective Optimization

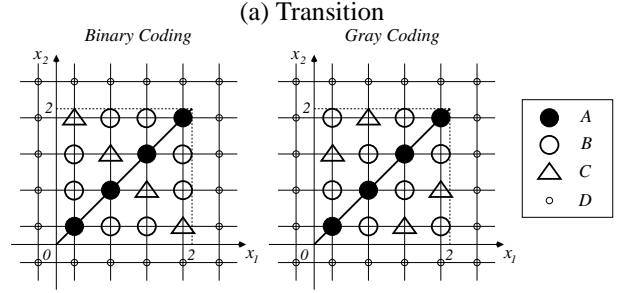
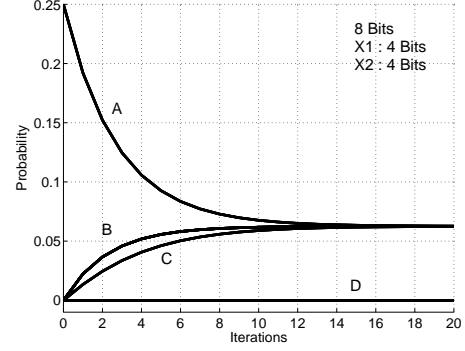
Our discussions so far are limited to single objective optimization. However, Equation (3) for GAs and Equation (13) for ESs are still valid for multi-objective optimization.

As an illustrative example, Equation (3) is applied to a widely used multi-objective test function, SCH1 ($n = 2$) [Coe01, Deb01], to examine the probability transition without selection. The SCH1 test function can be described as follows:

$$\begin{aligned} f_1(x) &= \left(\sum_{i=1}^2 x_i^2 \right) / 2 \\ f_2(x) &= \left(\sum_{i=1}^2 (x_i - 2)^2 \right) / 2 \end{aligned} \quad (28)$$

We assume that the initial parent population is distributed uniformly on the Pareto front, $x_1 = x_2, x_1, x_2 \in [0, 2]$. For this analysis, we use 8 bits for 2 design parameters, each having 4 bits. The probability transition of all possible chromosomes ($2^8 (= 256)$) is shown in Figure 9 (a). We observed that 4 various transition curves exist, which are labeled as A, B, C and D. Out of the 256 chromosomes, the transition probability of 4 chromosomes can be described by curve A, 8 by curve B, another 4 by curve C, and the rest 240 chromosomes by curve D. The corresponding locations of the different types of chromosomes in the parameter space after convergence are shown in Figure 9 (b), where both the results on binary coding and Gray coding are presented. In Figure 9 (b), the filled circles denoted by A are the initial parent individuals. After convergence, individuals (chromosomes) located on positions A, B, and C have the same probability, and the probability at which individuals are located at position D remains 0. In other words, the final population are distributed uniformly in $[0, 2]$, even though the initial population are distributed uniformly on the Pareto front. This observation is true for GAs with either binary coding or the Gray coding. Furthermore, the dynamics of the binary coding and the Gray coding are rather similar on this case.

An important remark that can be made from the above simple example is that in multi-objective optimization, the population cannot be regarded as converged even if all individuals are located on the Pareto front. The reason is that in multi-objective optimization, offspring generated from two Pareto-optimal solutions by crossover can be located far from the Pareto front.



(b) Points A – D in (a)

Figure 9: GA dynamics on SCH1. (a) Transition of the probabilities. 256 chromosomes show 4 different transition characteristics, A, B, C, and D. (b) The corresponding locations of 4 types of solutions on the $x_1 - x_2$ space. The left figure shows the solutions of the binary-coded GA and the right the Gray-coded GAs.

7 Summary

This paper suggests a theoretical framework for dynamics analysis of ESs using probability density. Based on this theory and the theory for analyzing GA dynamics suggested in [Oka05], we compared the search dynamics of binary-coded GAs, Gray-coded GAs and ESs theoretically. The theoretical results are also verified by empirical studies. We reveal very distinct search dynamics of binary-coded GAs, Gray-coded GAs and ESs. These differences in search dynamics imply that in general, binary-coded GAs are more suited for problems of which the optimum is located within initial range and is close to the boundaries. For Gray-coded GAs, the rough range of the optimum should also be known so that it can be encoded in the initial range. Besides, Gray-coded GAs have a higher probability to find the optimum if it is located near the middle of the initial range. The dynamics of ESs is more or less similar to that of Gray-coded GAs in that ESs tend to search the areas in the middle of the initial range more often, however, ESs are still able to find the optimum even if the optimum is not located in the initial range.

Acknowledgment

The authors would like to thank E. Körner and A. Richter for their support and M. Olhofer for his comments on this work. The first author would also like to thank T. Arima and

J. Takado for their continuous support.

Bibliography

- [Arn02] Arnold, D.V., *Noisy Optimization with Evolution Strategies*, Kluwer Academic, 2002.
- [Bey93] Beyer, H.-G., Towards a Theory of 'Evolution Strategies'. Some Asymptotical Results from the $(1, +\lambda)$ -Theory, *Evolutionary Computation*, 1(2), pages 165–188, 1993.
- [Bey94] Beyer, H.-G., Toward a Theory of Evolution Strategies: The (μ, λ) -Theory, *Evolutionary Computation*, 2(4), pages 381–407, 1994.
- [Bey95] Beyer, H.-G., Toward a Theory of Evolution Strategies: On the Benefits of Sex - the $(\mu/\mu, \lambda)$ Theory, *Evolutionary Computation*, 3(1), pages 81–111, 1995.
- [Bey96] Beyer, H.-G., Toward a Theory of Evolution Strategies: Self-Adaptation, *Evolutionary Computation*, 3(3), pages 311–347, 1996.
- [Bey00] Beyer, H.-G. and Arnold, D.V., Fitness Noise and Localization Errors of the Optimum in General Quadratic Fitness Models, *2nd Evonet Summer School on Theoretical Aspects of Evolutionary Computing*, pages 111–136, 2000.
- [Bey01] Beyer, H.-G., *The Theory of Evolution Strategies*, Springer, 2001.
- [Bey03] Beyer, H.-G. and Arnold D.V., The steady state behavior of $(\mu/\mu_I, \lambda)$ -ES on ellipsoidal fitness models disturbed by noise, *GECCO 2003*, pages 307–328, 2003.
- [Bey04] Beyer, H.-G., Actuator noise in recombinant evolution strategies on general quadratic fitness models, *GECCO 2004, LNCS 3102*, pages 654–665, 2004.
- [Coe01] Coello, C. A. C., Van Veldhuizen, D. A. and Lamont, G. B., *Evolutionary Algorithms for Solving Multi-Objective Problems*, Kluwer Academic Publishers, 2001.
- [Deb01] Deb, K., *Multi-Objective Optimization using Evolutionary Algorithms*, John Wiley & Sons, LTD., 2001.
- [Fog92] Fogel, D. B., *Evolving Artificial Intelligence*, Ph.D. Dissertation, Univ. of California, 1992.
- [Fog99] Fogel, D. B., *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence*, Wiley-IEEE Press, 1999.
- [Gol89] Goldberg, D. E., *Genetic Algorithms in Search, Optimization and machine Learning*, Addison Wesley, 1989.
- [Nur97] Nürnberg, H. -T. and Beyer, H. -G., The Dynamics of Evolution Strategies in the Optimization of Traveling Salesman Problems, *Evolutionary Programming VI: Proceedings of the Sixth Annual Conference on Evolutionary Programming*, pages 349–359, 1997.
- [Oka02] Okabe, T., Jin, Y. and Sendhoff, B., On the Dynamics of Evolutionary Multi-Objective Optimisation, *Proceedings of Genetic and Evolutionary Computation Conference (GECCO-2002)*, pages 247–255, 2002.
- [Oka04] Okabe, T., *Evolutionary Multi-Objective Optimization*, Shaker Verlag, 2004.
- [Oka05] Okabe, T., Jin, Y. and Sendhoff, B., A New Approach to Dynamics Analysis of Genetic Algorithms without Selection, *Congress on Evolutionary Computation (CEC-2005)*, (Accepted).
- [Pru01] Prügel-Bennett, A., Modelling Crossover-Induced Linkage in Genetic Algorithms, *IEEE Transactions on Evolutionary Computation*, 5(4), pages 376–387, 2001.
- [Pru01b] Prügel-Bennett, A., Modelling Genetic Algorithm Dynamics, *Theoretical Aspects of Evolutionary Computing*, pages 59–85, 2001.
- [Rat95] Rattray, L. M., The Dynamics of a Genetic Algorithm under Stabilizing Selection, *Complex Systems*, 9, pages 213–234, 1995.
- [Ree03] Reeves, C. R. and Rowe, J. E., *Genetic Algorithms - Principles and Perspectives: A Guide to GA Theory*, Kluwer Academic Publishers, 2003.
- [Rog97] Rogers, A. and Prügel-Bennett, A., The Dynamics of a Genetic Algorithm on a Model Hard Optimization Problem, *Complex Systems*, 11(6), pages 437–464, 1997.
- [Rog99] Rogers, A. and Prügel-Bennett, A., Modelling the Dynamics of a Steady State Genetic Algorithm, *Proceedings of Foundations of Genetic Algorithms 5 (FOGA-5)*, pages 57–68, 1999.
- [Rud94] Rudolph, G., Convergence Analysis of Canonical Genetic Algorithms, *IEEE Transactions on Neural Networks*, 5(1), pages 96–101, 1994.
- [Suz98] Suzuki, J., A Further Result on the Markov Chain Model of Genetic Algorithms and Its Application to a Simulated Annealing-Like Strategy, *IEEE Transactions Systems, Man, Cybernetics*, pages 95–102, 1998.
- [Vos99] Vose, M. D., *The Simple Genetic Algorithm: Foundations and Theory*, The MIT Press, 1999.

A Proof of Equation (26)

Define $f(x)$ as:

$$f(x) \equiv \operatorname{erf}\left(\frac{x_e - x}{\sigma\sqrt{2}}\right) - \operatorname{erf}\left(\frac{x_s - x}{\sigma\sqrt{2}}\right). \quad (29)$$

We have:

$$\begin{aligned} & f\left(\frac{x_s + x_e}{2} + a\right) - f\left(\frac{x_s + x_e}{2} - a\right) \\ &= \operatorname{erf}\left(\frac{x_e - (x_s + x_e)/2 - a}{\sigma\sqrt{2}}\right) - \operatorname{erf}\left(\frac{x_s - (x_s + x_e)/2 - a}{\sigma\sqrt{2}}\right) \\ & \quad - \operatorname{erf}\left(\frac{x_e - (x_s + x_e)/2 + a}{\sigma\sqrt{2}}\right) + \operatorname{erf}\left(\frac{x_s - (x_s + x_e)/2 + a}{\sigma\sqrt{2}}\right) \\ &= \operatorname{erf}\left(\frac{x_e/2 - x_s/2 - a}{\sigma\sqrt{2}}\right) - \operatorname{erf}\left(\frac{x_s/2 - x_e/2 - a}{\sigma\sqrt{2}}\right) \\ & \quad - \operatorname{erf}\left(\frac{x_e/2 - x_s/2 + a}{\sigma\sqrt{2}}\right) + \operatorname{erf}\left(\frac{x_s/2 - x_e/2 + a}{\sigma\sqrt{2}}\right) \\ &= 0 \end{aligned} \quad (30)$$

since $\operatorname{erf}(x) = -\operatorname{erf}(-x)$. Thus, function $f(x)$ is symmetric to the center $(x_s + x_e)/2$, where $\Psi(x | x < x_s)$ integrates $f(x)$ from $x = -\infty$ to $x = x_s$, and $\Psi(x | x_e < x)$ from $x = x_e$ to $x = \infty$.

(Q.E.D)