Polynomial Approximation of Survival Probabilities Under Multi-point Crossover

Sung-Soon Choi and Byung-Ro Moon

School of Computer Science and Engineering,
Seoul National University,
Seoul, 151-742 Korea
{sschoi,moon}@soar.snu.ac.kr

Abstract. We propose an analytic approach to approximate the survival probabilities of schemata under multi-point crossover and obtain its closed form. It gives a convenient way to mathematically analyze the disruptiveness of multi-point crossover. Based on the approximation, we describe a geometric property of the survival probability under multi-point crossover and show the relationship between the survival probability and the distribution of the specific symbols in schemata.

1 Introduction

Crossover is a major operator that plays an important role in the design of genetic algorithms. Discussion on the effect of crossover dates back to the Schema Theorem [9] which is one of the milestone theorems in the area of genetic algorithms. The theorem says that the probability that a schema survives through generations depends on not only its fitness but also the disruptiveness of crossover. So considerable attention has been given to estimating the disruptiveness of crossover.

A variety of crossover operators have been proposed and the disruptiveness of crossovers has also been a hot research topic. For one-dimensional chromosomes, multi-point crossover and uniform crossover are two representative crossovers that have been much studied in the area of genetic algorithms [15] [11]. Recently, several crossovers for non-linear chromosomes were proposed and studies on the disruptiveness of those crossovers were conducted [7] [1] [4] [13].

In uniform crossover, the allele at any position in an offspring is determined by the allele of the first parent with probability p or by the allele of the second parent with probability $1-p$. One property of this crossover is that it is simple to estimate the disruptiveness of the crossover; the survival probability for any order schema is easily obtained in a closed form. Based on this, Syswerda provided an initial analysis of the disruptive effects of uniform crossover for the case of $p = 0.5$ [15].

There have been more studies for the disruptiveness of multi-point crossover. De Jong [10] first observed that the defining length is not necessarily a dominating factor of a schema's survival probability when multi-point crossovers are used. He provided an exact expression for the survival probability for

the 2nd-order schemata. Later, De Jong and Spears [11] extended this to provide a detailed analysis of the survival probabilities of higher-order schemata. Bui and Moon [3] [5] proposed a new schema model which is convenient for dealing with the distribution of specific symbols within schemata. In the model, Bui and Moon investigated the characteristics of schemata that are prone to have good survival probabilities under multi-point crossovers. They also provided an exact expression for the survival probability for multi-point crossover.

Despite these results, the closed form for the survival probability of a schema under multi-point crossover has not been discovered. Due to this, the researches for multi-point crossover depended primarily on empirical analyses. The objective of this paper is to introduce an analytic approach to approximate the survival probabilities of schemata under multi-point crossover and obtain a closed form for the survival probability. As stated below, it gives a convenient way to mathematically analyze the disruptiveness of multi-point crossover and, consequently, leads to a deeper understanding of multi-point crossover.

The rest of this paper is organized as follows. In Section 2, we present some previous works on the disruptiveness of crossover, mainly, multi-point crossover. In Section 3, we propose an approach to approximate the survival probability of a schema under multi-point crossover to a polynomial related to the distribution of the specific symbols in the schema. Based on such an approximation, in Section 4, we show the convex property of the survival probability and prove the relationship between the survival probability of a schema and the distribution of the specific symbols in the schemata. Finally, we make our conclusions in Section 5.

2 Survival Probabilities

At first, we summarize basic terminologies of genetic algorithms for the completeness of the paper. A *chromosome* is a sequence of gene values. Each gene has a value from an alphabet $\mathcal{A}$. A *schema* of a chromosome of length n can be represented as an n-tuple $< s_1, s_2, \ldots, s_n >$ where $s_i \in \mathcal{A} \cup \{*\}$ for $i = 1, 2, \ldots, n$. In a schema, the symbol "$*$" represents don't-care positions and non-$*$ symbols (called *specific symbols*) specify defining positions of the schema and their corresponding gene values. The *distance* between two genes is the distance between their positions. The *defining length* of a schema is defined to be the distance from the leftmost specific symbol to the rightmost specific symbol in that schema. The *order* of a schema is the number of specific symbols in the schema.

The analysis of schemas' survival probabilities based on defining lengths works well on 1-point crossover models, but does not generally work well on multi-point crossover models. De Jong [10] found an exact expression for the survival probabilities of 2nd-order schemata under multi-point crossover and observed that the survival probabilities of 2nd-order schemata are not necessarily affected by their defining lengths when multi-point crossovers are used. Extending this, De Jong and Spears [11] investigated the relationship between the survival probabilities and the defining lengths of higher-order schemata. (See

[14] for more detailed discussion.) Their key observation is that a schema is not disrupted when an even number (including 0) of crossover points fall between every pair of adjacent specific symbols.

Let $s_0, s_1, \ldots, s_{r-1}$ be the specific symbols of an rth-order schema from left to right. Let d_i be the distance between s_0 and s_i for $i = 1, 2, \ldots, r-1$ and let n be the length of the chromosome. They provided a recursive equation to calculate $P_{k,even}(r)$, the probability that an even number of crossover points fall between each of the defining positions of the rth-order schema (consisting of s_0 through s_{r-1}) by a k-point crossover as follows:

$$P_{k,even}(r) = \sum_{i=0}^{\lfloor k/2 \rfloor} \binom{k}{2i} \left(\frac{d_{r-1}}{n}\right)^{2i} \left(\frac{n - d_{r-1}}{n}\right)^{k-2i} P_{2i,even}(r - 1). \qquad (1)$$

The above equation assumes that the crossover points are independent of one another. They are actually dependent because no two crossover points can fall onto the same position. But this approximation causes little harm as long as $k \ll n$, which is true in most cases.

Bui and Moon [3] [5] investigated the relationship between the inner structures of schemata and their survival probabilities. They concentrated on the specific-symbol clusters and introduced a new type of schema, called *clustered schema* or *c-schema*. A *c*-schema is defined as $D_0 C_1 D_1 \cdots C_q D_q$, where $C_i \in \mathcal{A}^+, i = 1, \ldots, q$, $D_i \in \{*\}^+, i = 1, \ldots, q-1$, and $D_0 D_q \in \{*\}^*$. Let the length of the string D_i be $|D_i|$. If we define $P_k(D_0 C_1 D_1 \cdots C_q D_q)$ to be the probability that the *c*-schema $D_0 C_1 D_1 \cdots C_q D_q$ is not disrupted by a k-point crossover, then the following holds [3] [5]:

$$P_k(D_0 C_1 D_1 \cdots C_q D_q) = \sum_{i_1 + \cdots + i_q \leq \lfloor k/2 \rfloor} \frac{\binom{|D_1|+1}{2i_1} \cdots \binom{|D_{q-1}|+1}{2i_{q-1}} \binom{|D_0 D_q|}{k-2(i_1+\cdots+i_{q-1})}}{\binom{n-1}{k}}. \qquad (2)$$

Based on this equation, Bui and Moon suspected that the survival probability of a schema considerably depends on the distribution of the specific symbols. They empirically supported this to show that, the more clustered the specific symbols of a schema are, the higher survivial probability it has.

Although the equations (1) and (2) measure the survival probability of a given schema, they involve complicated summations of binomial coefficients. It seems to be impossible to mathematically handle the equations to derive some other results. On the other hand, as stated above, it is easy to obtain the closed form for the survival probability of any order schema under uniform crossover. If we denote the probability parameter by p, the survival probability of an rth-order schema under uniform crossover, $P_u(r)$, is of a simple form as follows:

$$P_u(r) = p^r + (1 - p)^r.$$

In the next section, we propose a way to approximate the survival probability of a schema under multi-point crossover and obtain its closed form.

Note that a schema can survive even when the above conditions for the equations (1) and (2) are not satisfied if all lost specific symbols in one parent are accidently recovered by the other parent. Our results can be generalized to handle the cases in the same manner as in [11]. In this discussion, we ignore such cases for simplicity.

3   Polynomial Approximations

3.1   Uniform Convergence and Linkage Distribution

Let $\langle f_n \rangle$ be a sequence of functions defined on a set X and with range in $\mathbb{R}$. We say the sequence $\langle f_n \rangle$ *uniformly converges* to the function f on the set X if and only if for each $\epsilon > 0$ there is a number N independent of x such that

$$|f_n(x) - f(x)| < \epsilon \text{ for all } x \in X \text{ and all } n > N.$$

Uniform convergence differs from ordinary pointwise convergence in that the integer N does not depend on x, although naturally it depends on ϵ. Roughly speaking, uniform convergence means that all the values of f_n on X converge to the value of f on X in the same rate, as n increases. More strictly, the uniform convergence of $\langle f_n \rangle$ to f on X implies that $\sup_X |f_n - f| \to 0$ as $n \to \infty$, and vice versa. We refer to [12] for more details.

Let $s_0, s_1, \ldots, s_q$ be the specific symbols of a $(q+1)^{th}$-order schema from left to right and n be the chromosome length. We denote the distance between two symbols s_i and s_j by $d(s_i, s_j)$. Let $x_i = \frac{d(s_{i-1}, s_i)}{n-1}, i = 1, \ldots, q$, which is the fraction of the distance between the adjacent specific symbols s_{i-1} and s_i over the chromosome length minus one. Then, $(x_1, \ldots, x_q)$ represents the relative distances of the specific symbols of the $(q+1)^{th}$-order schema. Note that, under multi-point crossover, the survival probability of a schema depend little on the absolute positions of the specific symbols; it depends highly on their relative positions. So the notation $(x_1, \ldots, x_q)$ is sufficient to represent the $(q+1)^{th}$-order schema in analyzing the survival probability. We call $(x_1, \ldots, x_q)$ the *linkage distribution* of the $(q+1)^{th}$-order schema. We denote $x = \sum_{i=1}^q x_i$ and call it the *linkage sum* of the schema. The linkage sum of a schema is the defining length of the schema normalized by the chromosome length. In representing a 2nd-order schema, we simply use the notation x instead of x_1.

We denote by $P_{k,n}(x_1, \ldots, x_q)$ the survival probability of a $(q+1)^{th}$-order schema with linkage distribution $(x_1, \ldots, x_q)$ in a chromosome of length n under k-point crossover. For a 2nd-order schema, we simply use the notation $P_{k,n}(x)$ as above. Note that $P_{k,n}(x_1, \ldots, x_q)$ is a function of the chromosome length n. In this section, we show that the sequence $\langle P_{k,n}(x_1, \ldots, x_q) \rangle$ with respect to n uniformly converges to a polynomial of $x_1, \ldots, x_q$ on the set $\{(x_1, \ldots, x_q)|0 < x = \sum_{i=1}^q x_i \le 1\}$ that contains the linkage distributions of the $(q+1)^{th}$-order schemata. This supports that the survival probability may be approximated to the polynomial for sufficiently large n.

3.2 2nd-Order Schemata

Let $\delta(i) = 1$ if $i = 0$, and $\delta(i) = 0$ otherwise. Then the following holds.

Theorem 1. *The sequence $\langle P_{k,n}(x) \rangle$ uniformly converges on the set $\{x | 0 < x \le 1\}$ to the kth-order polynomial of x, $P_k(x)$, with the form*

$$P_k(x) = \sum_{i=0}^{k} c_i x^i,$$

where $c_i = (-1)^i \binom{k}{i} 2^{i-1+\delta(i)}$ for $i = 0, 1, 2, \ldots$.

Proof: Let d be the defining length of a 2nd-order schema. Then,

$$P_{k,n}(x) = P_{k,n}(\frac{d}{n}) = \sum_{i=0}^{\lfloor k/2 \rfloor} \binom{d}{2i} \binom{n-1-d}{k-2i} / \binom{n-1}{k}.$$

The numerator in the right hand side can be rewritten as follows:

$$\sum_{i=0}^{\lfloor k/2 \rfloor} \binom{d}{2i} \binom{n-1-d}{k-2i} = \sum_{i=0}^{\lfloor k/2 \rfloor} \frac{d^{2i}}{2i!} \frac{(n-d)^{k-2i}}{(k-2i)!} + O(n^{k-1}).$$

Dividing both sides by $\binom{n-1}{k}$,

$$\sum_{i=0}^{\lfloor k/2 \rfloor} \binom{d}{2i} \binom{n-1-d}{k-2i} / \binom{n-1}{k} = \sum_{i=0}^{\lfloor k/2 \rfloor} \frac{k!}{2i!(k-2i)!} \frac{d^{2i}(n-d)^{k-2i}}{(n-1)^k} + O(\frac{1}{n}).$$

So, we get

$$P_{k,n}(x) = \sum_{i=0}^{\lfloor k/2 \rfloor} \binom{k}{2i} x^{2i}(1-x)^{k-2i} + O(\frac{1}{n}), \tag{3}$$

which means that $P_{k,n}(x)$ uniformly converges to a kth-order polynomial of x. From the equation, the coefficient of x^i is

$$(-1)^i \sum_{j=0}^{\lfloor i/2 \rfloor} \binom{k}{2j} \binom{k-2j}{i-2j} = (-1)^i \sum_{j=0}^{\lfloor i/2 \rfloor} \binom{k}{i} \binom{i}{2j} = (-1)^i \binom{k}{i} 2^{i-1+\delta(i)}$$

using the fact that $\binom{r}{m} \binom{m}{k} = \binom{r}{k} \binom{r-k}{m-k}$ and the binomial theorem [8].   □

Here are some examples.

$$\begin{aligned}
P_1(x) &= -x + 1, \\
P_2(x) &= 2x^2 - 2x + 1, \\
P_3(x) &= -4x^3 + 6x^2 - 3x + 1, \\
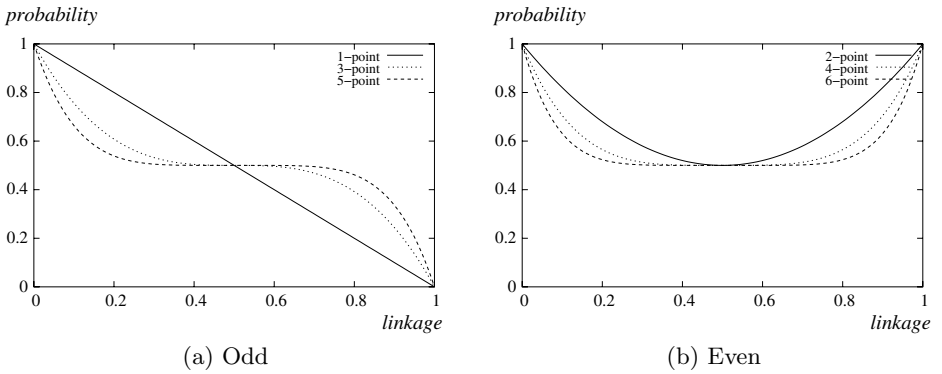P_4(x) &= 8x^4 - 16x^3 + 12x^2 - 4x + 1.
\end{aligned}$$

Fig. 1. Survival probabilities of 2nd-order schemata under k-point crossover

On the other hand, from the equation (3),

$$P_k(x) = \sum_{i=0}^{\lfloor k/2 \rfloor} \binom{k}{2i} x^{2i} (1-x)^{k-2i}$$

$$= \binom{k}{0} x^0 (1-x)^k + \binom{k}{2} x^2 (1-x)^{k-2} + \cdots + \binom{k}{2\lfloor \frac{k}{2} \rfloor} x^{2\lfloor \frac{k}{2} \rfloor} (1-x)^{k-2\lfloor \frac{k}{2} \rfloor}.$$

Then, by the binomial theorem [8], the following holds.

$$2P_k(x) = (x + (1-x))^k + (-x + (1-x))^k = 1 + (1-2x)^k.$$

Therefore, we have the following simple form.

Theorem 2.

$$P_k(x) = \frac{1 + (1-2x)^k}{2}.$$

Note that Theorem 1 can be derived directly from Theorem 2. Figure 1 shows the survival probabilities of 2nd-order schemata under k-point crossover with respect to linkage x. For odd k, the graph $y = P_k(x)$ is symmetric around the point $(x, y) = (0.5, 0.5)$, and, for even k, the graph $y = P_k(x)$ is symmetric about $x = 0.5$. These facts are easily checked from Theorem 2. We also have that, for $x \in (0, 1)$, the survival probability of a 2nd-order schema with linkage x pointwise converges to $y = \frac{1}{2}$, the same as that of uniform crossover with $p = 0.5$, as k increases.

3.3   Higher-Order Schemata

Extending Theorem 1, we have

Theorem 3. *The sequence $\langle P_{k,n}(x_1,\dots,x_q)\rangle$ uniformly converges on the set $\{(x_1,\dots,x_q)|0 < x = \sum_{i=1}^{q} x_i \leq 1\}$ to the kth-order polynomial of $x_1,\dots,x_q$, $P_k(x_1,\dots,x_q)$, with the form*

$$P_k(x_1, x_2, \dots, x_q) = \sum_{i_1=0}^{k} \sum_{i_2=0}^{k-i_1} \cdots \sum_{i_q=0}^{k-(i_1+\cdots+i_q)} c_{i_1,i_2,\dots,i_q} x_1^{i_1} x_2^{i_2} \cdots x_q^{i_q},$$

where

$$c_{i_1,i_2,\dots,i_q} = (-1)^{i_1+\cdots+i_q} \frac{(i_1+\cdots+i_q)!}{i_1!\cdots i_q!} \binom{k}{i_1+\cdots+i_q} 2^{i_1+\cdots+i_q-q+\sum_{j=1}^{q}\delta(i_j)}$$

for $i_1,\dots,i_q = 0, 1, 2, \dots$.

Proof: The proof is analogous to that of Theorem 1. We omit the proof by space limitation. See [6] for details. □

Again, in an analogous way to Theorem 2, we have the following.

Theorem 4.

$$
\begin{aligned}
P_k(x_1,\dots,x_q) = {} & \frac{1}{2^q}(1 + (1-2x_1)^k + \cdots + (1-2x_q)^k \\
& + (1-2x_1-2x_2)^k + (1-2x_1-2x_3)^k + \cdots + (1-2x_{q-1}-2x_q)^k \\
& + \cdots \\
& + (1-2x_1-2x_2-\cdots-2x_q)^k).
\end{aligned}
$$

Proof: We omit the proof by space limitation. See [6] for details. □

It is clear that $P_k(x_1,\dots,x_q)$ is symmetric in $x_1,\dots,x_q$ in that $P_k(x_1,\dots,x_q)$ is left fixed by all permutations of $x_1,\dots,x_q$. Here are some examples.

$$
\begin{aligned}
P_1(x_1,\dots,x_q) &= (1-x), \\
P_2(x_1,\dots,x_q) &= (1-x)^2 + (x_1^2+\cdots+x_q^2), \\
P_3(x_1,\dots,x_q) &= (1-x)^3 + 3(1-x)(x_1^2+\cdots+x_q^2), \\
P_4(x_1,\dots,x_q) &= (1-x)^4 + 6(1-x)^2(x_1^2+\cdots+x_q^2) \\
&\quad + 6(x_1^2x_2^2 + x_1^2x_3^2 + \cdots + x_{q-1}^2x_q^2) \\
&\quad + (x_1^4+\cdots+x_q^4).
\end{aligned}
\tag{4}
$$

As shown in the proofs of the above theorems, the difference between $P_{k,n}(x_1,\dots,x_q)$ and the limiting polynomial $P_k(x_1,\dots,x_q)$ is $O(\frac{1}{n})$, which means that the approximation is acceptable even for not very large n. Consider two 8th-order schemata H_1 and H_2 in a chromosome of length n. Figure 2 shows the appearances of the two schemata, where the symbol $*$ represents don't-care
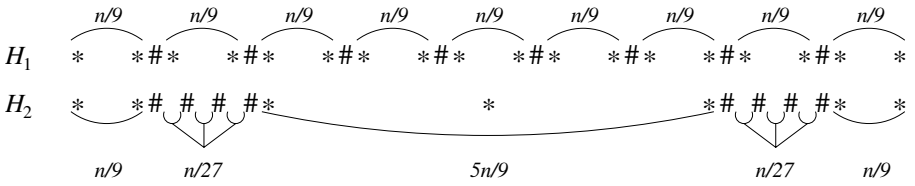
Fig. 2. Example schemata H_1 and H_2

positions and the symbol # represents specific positions. H_1 is the schema with linkage distribution $(\frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9}, \frac{1}{9})$ and H_2 is the schema with linkage distribution $(\frac{1}{27}, \frac{1}{27}, \frac{1}{27}, \frac{5}{9}, \frac{1}{27}, \frac{1}{27}, \frac{1}{27})$. As shown in the figure, the specific symbols are evenly distributed in H_1 and are more clustered in H_2. For more concrete experiments, we set the number of the don't-care symbols between adjacent specific symbols to be $\lfloor \frac{n-8}{9} \rfloor$ in H_1. And, in H_2, we set the number of the don't-care symbols between adjacent specific symbols in a cluster to be $\lfloor \frac{n-8}{27} \rfloor$ and make the defining length of H_2 be the same as that of H_1.

For $n = 100$, the exact values of the survival probabilities of H_1 and H_2 under 2-point crossover are 0.1270 and 0.3391, respectively.[1] Using the equation (4), we get the corresponding approximate values as follows:

$$P_2(H_1) = P_2(\frac{11}{99}, \frac{11}{99}, \frac{11}{99}, \frac{11}{99}, \frac{11}{99}, \frac{11}{99}, \frac{11}{99}) = 0.1358,$$

$$P_2(H_2) = P_2(\frac{4}{99}, \frac{4}{99}, \frac{4}{99}, \frac{53}{99}, \frac{4}{99}, \frac{4}{99}, \frac{4}{99}) = 0.3458.$$

The approximation errors of $P_2(H_1)$ and $P_2(H_2)$ are 0.0088 and 0.0067, respectively, which are fairly small. Table 1 shows how well such an approximation

Table 1. Survival probabilities of H_1 and H_2 under 2-point crossover

n	H_1			H_2		
	Exact	Approx.	Δ	Exact	Approx.	Δ
200	0.1323	0.1367	0.0044	0.3413	0.3446	0.0033
400	0.1350	0.1371	0.0021	0.3574	0.3590	0.0016
600	0.1359	0.1373	0.0014	0.3629	0.3639	0.0010
800	0.1343	0.1354	0.0011	0.3636	0.3644	0.0008
1,000	0.1349	0.1358	0.0009	0.3656	0.3663	0.0006

scales with the chromosome length n. In this table, "Exact" and "Approx." indicate the exact values and the approximate values of the survival probabilities of H_1 and H_2 under 2-point crossover, respectively. Δ indicates the difference

---

[1] We get the values, for example, using the equation (2).

between these values, i.e., the approximation error. As expected, the approximation errors seem to decrease in the rate of $O(\frac{1}{n})$. In particular, the errors for H_1 and H_2 decrease almost at the same rate as n increases, which relates to the uniform convergence of the approximation.

Despite the same defining length, on the other hand, the survival probability of H_2 is much higher than that of H_1 under 2-point crossover. As claimed in [3] and [5], this is because the distribution of the specific symbols are different in H_1 and H_2. In the next section, we show how the survival probabilities of the schemata with the same defining length relate to the distributions of the specific symbols in the schemata.

4 Convex Property of Survival Probability

In this section, we assume that the chromosome length n is sufficiently large, and that, based on the assumption, $P_k(x_1, \ldots, x_q)$ is the survival probability of a schema with linkage distribution $(x_1, \ldots, x_q)$ under k-point crossover and $\{(x_1, \ldots, x_q)|0 < x = \sum_{i=1}^{q} x_i \leq 1\}$ is the set of the linkage distributions of the $(q+1)^{th}$-order schemata.

A set S is *convex* if the line segment between any two points in S lies in S, i.e., if for any $\mathbf{x}, \mathbf{y} \in S$ and any θ with $0 \leq \theta \leq 1$ we have $\theta\mathbf{x} + (1-\theta)\mathbf{y} \in S$. Roughly speaking, a set is convex if every point in the set can be seen by every other point, along an "unobstructed" straight path between them. We call a point of the form $\theta_1\mathbf{x}_1 + \cdots + \theta_k\mathbf{x}_k$, where $\theta_1 + \cdots + \theta_k = 1$ and $\theta_i \geq 0, i = 1, \ldots, k$, a *convex combination* of the points $\mathbf{x}_1, \ldots, \mathbf{x}_k$. The *convex hull* of a set C, denoted $\mathbf{conv}C$, is the set of all convex combinations of points in C:

$$\mathbf{conv}C = \{\theta_1\mathbf{x}_1 + \cdots + \theta_k\mathbf{x}_k | \mathbf{x}_i \in C, \theta_i \geq 0, i = 1, \ldots, k, \theta_1 + \cdots + \theta_k = 1\}.$$

As the name suggests, the convex hull of a set is always convex. We let

$$S_{x,q} = \{(x_1, \ldots, x_q)| \sum_{i=1}^{q} x_i = x, x_i \geq 0, i = 1, \ldots, q\}$$

for a fixed $x \in (0, 1]$, which is the set of the linkage distributions of the $(q+1)^{th}$-order schemata with linkage sum x. Let $\mathbf{e}_i$ be the vector of dimension q all whose elements are zero except that the ith element is one. Then, it holds that $S_{x,q} = \mathbf{conv}\{x\mathbf{e}_i|i = 1, \ldots, q\}$ and, consequently, $S_{x,q}$ is a convex set.

A function $f : \mathbb{R}^n \to \mathbb{R}$ is *convex* if the domain of f is a convex set and, for all $\mathbf{x}, \mathbf{y}$ in the domain and θ with $0 \leq \theta \leq 1$, we have

$$f(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \leq \theta f(\mathbf{x}) + (1-\theta)f(\mathbf{y}). \tag{5}$$

Geometrically, this inequality means that the line segment between $(\mathbf{x}, f(\mathbf{x}))$ and $(\mathbf{y}, f(\mathbf{y}))$ lies above the graph of f. A function f is *strictly convex* if strict
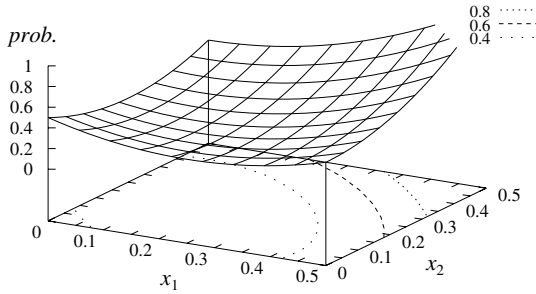
Fig. 3. The strict convexity of $P_3(x_1, x_2, x_3)$ with $x = 0.5$ under 3-point crossover

inequality holds in the inequality (5) whenever $\mathbf{x} \neq \mathbf{y}$ and $0 < \theta < 1$. A non-negative weighted sum of convex functions is convex. Similarly, a nonnegative, nonzero weighted sum of strictly convex functions is strictly convex. We refer to [2] for more details.

From Theorem 2, $P_k(x) = \frac{1+(1-2x)^k}{2}$. Since $P_k''(x) = 2k(k-1)(1-2x)^{k-2}$, $P_k(x)$ is strictly convex for even k. We define $P_0(x_1, \dots, x_q) = 1$. Then, in the same manner as in [11], we get the following recurrence which we omit the proof in this paper:

$$P_k(x_1, \dots, x_q) = \sum_{i=0}^{\lfloor k/2 \rfloor} x^{2i}(1-x)^{k-2i} P_{2i}(\frac{x_1}{x}, \dots, \frac{x_{q-1}}{x}).$$

Hence, by induction, $P_k(x_1, \dots, x_q)$ is strictly convex for even k and any $q \geq 1$, since a nonnegative weighted sum of strictly convex functions is also strictly convex. For odd k and $q \geq 2$, for the same reason, $P_k(x_1, \dots, x_q)$ is strictly convex in the set $S_{x,q}$ for a fixed $x \in (0, 1]$. Therefore, we get the following.

Theorem 5. *For even k, $P_k(x_1, \dots, x_q)$ is strictly convex in $\mathbb{R}^q$. And, for any odd k and $q \geq 2$, $P_k(x_1, \dots, x_q)$ is strictly convex in the set $S_{x,q}$ for a fixed $x \in (0, 1]$.*

Figure 3 shows the survival probabilities of the 4th-order schemata (x_1, x_2, x_3) with linkage sum $x = 0.5$ under 3-point crossover. In the figure, we plot the graph of $y = P_3(x_1, x_2, x_3)$ with respect to x_1 and x_2 since x_3 is determined by x_1 and x_2. The strict convexity of the survival probability is clear from the graph. Note that $y = P_3(x_1, x_2, x_3)$ is not convex any longer if x is not fixed.

An important result for convexity is that a strictly convex function has one global minimum and has at least one global maximum at the boundary of the domain. In particular, if its domain is the convex hull of a set, the function can have global maxima only at the points in the set. The following theorem shows how the survival probability depends on the distribution of the specific symbols of the schema.

Theorem 6. *Over the $(q + 1)^{th}$-order schemata with linkage distribution $(x_1, \dots, x_q)$ and linkage sum $x \in (0, 1]$, the survival probability is minimized*
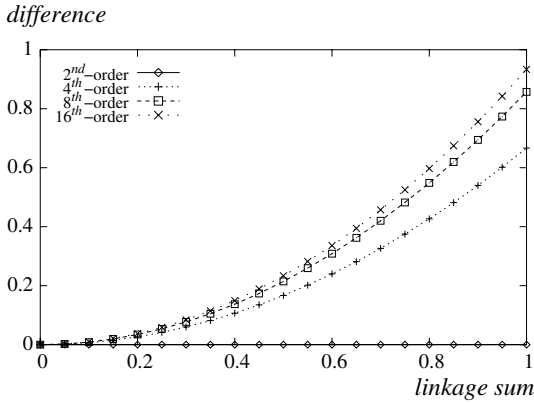
difference

Fig. 4. The difference between the maximum and minimum values of the survival probabilities under 2-point crossover

when the variance of x_i's is minimized, and maximized when the variance is maximized.

Proof: We use the convexity of the survival probability and Lagrange multipliers. We omit the proof by space limitation. See [6] for details.   □

From Theorem 4, the maximum value of the survival probabilities of $(q+1)^{th}$-order schemata under k-point crossover is $P_k(0,\dots,0,x) = \frac{2^{q-1}+(1-2x)^k 2^{q-1}}{2^q} = \frac{1+(1-2x)^k}{2} = P_k(x)$. The minimum value of the survival probability is $P_k(\frac{x}{q},\dots,\frac{x}{q}) = \frac{1}{2^q q^k}\sum_{i=0}^{q}\binom{q}{i}(q-2xi)^k$. Figure 4 shows the difference between the maximum and minimum values of the survival probabilities under 2-point crossover with respect to linkage sum x. It is seen that the difference is fairly large for the higher-order schemata with relatively large defining lengths.

5   Conclusion

We proposed an analytic approach to approximate the survival probabilities of schemata under multi-point crossover. By adapting the approach, we obtained a closed form for the survival probabilities of schemata, which can be calculated in polynomial time with the distribution of the specific symbols. Such an approximation provides with a convenient way to analyze the disruptiveness of multi-point crossover mathematically. This also enables us to understand multi-point crossover more deeply. It becomes also possible to investigate more complex schemata in genetic studies.

We showed the convex property of the survival probabilities under multi-point crossover. We then proved how the survival probability of a schema relates to the distribution of the specific symbols in the schema. This confirms the

previous results that were empirically shown: the more clustered specific symbols a schema has, the higher survival probability it has.

We are currently working on investigating another properties of multi-point crossover based on the approximation. We are going to applying such an approximation approach to other crossovers to analyze the survival probabilities under the crossovers.

References

1. C. Anderson, K. Jones, and J. Ryan. A two-dimensional genetic algorithm for the ising problem. *Complex Systems*, 5:327–333, 1991.
2. S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2003.
3. T. N. Bui and B. R. Moon. Analyzing hyperplane synthesis in genetic algorithms using clustered schemata. In *Parallel Problem Solving from Nature*, pages 108–118, 1994.
4. T. N. Bui and B. R. Moon. On multi-dimensional encoding/crossover. In *Sixth International Conference on Genetic Algorithms*, pages 49–56, 1995.
5. T. N. Bui and B. R. Moon. GRCA: A hybrid genetic algorithm for circuit ratio-cut partitioning. *IEEE Trans. on CAD*, 17(3):193–204, 1998.
6. S. S. Choi and B. R. Moon. An analytic approach for approximation of survival probability under multi-point crossover. In preparation, 2004.
7. J. Cohoon and D. Paris. Genetic placement. In *IEEE International Conference on Computer-Aided Design*, pages 422–425, 1986.
8. R. L. Graham, D. E. Knuth, and O. Patashnik. *Concrete Mathematics, Second Edition*. Addison Wesley, 1994.
9. J. Holland. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, 1975.
10. K. A. De Jong. *An Analysis of the Behavior of a Class of Genetic Adaptive Systems*. PhD thesis, University of Michigan, Ann Arbor, MI, 1975.
11. K. A. De Jong and W. M. Spears. A formal analysis of the role of multi-point crossover in genetic algorithms. *Annals of Mathematics and Artificial Intelligence Journal*, 5:1–26, 1992.
12. W. Rudin. *Principles of Mathematical Analysis, Third Edition*. McGraw-Hill, 1976.
13. D. I. Seo and B. R. Moon. Voronoi quantized crossover for traveling salesman problem. In *Genetic and Evolutionary Computation Conference*, pages 544–552, 2002.
14. W. M. Spears. *The Role of Mutation and Recombination in Evolutionary Algorithms*. PhD thesis, George Mason University, Fairfax, Virginia, 1998.
15. G. Syswerda. Uniform crossover in genetic algorithms. In *Third International Conference on Genetic Algorithms*, pages 2–9, 1989.