A Generalized Crossover Operation for Genetic Algorithms

Michael Kolonko*

Institut für Mathematik, Universität Hildesheim,
Marienburger Platz 22, D-31141 Hildesheim, Germany

Abstract. The recombination of solutions (crossover) is probably the most specific operation in optimization by genetic algorithms. We consider a very general way of recombining two solutions using concepts related to orthogonal projections. This includes most of the commonly used crossover operators such as, for example, one-point or uniform crossover.

We examine symmetry properties of the operator, generalize the classical schema-oriented approach to our setting, and study the distribution of the offspring both geometrically and stochastically. In particular we show that expectation and variance of the population (defined in appropriate terms) are invariant under crossover.

It turns out that the important features of the classical crossover operators hold in much more general models, including continuous space.

1. Introduction

Genetic algorithms (GAs) adapt certain principles of natural evolution to combinatorial optimization problems. Operations like selection, crossover, and mutation are applied to a population of possible solutions. The aim is to breed an optimal or near optimal solution using the target function of the optimization problem as "fitness." See [2] and [4] for an introduction to these concepts. Though these algorithms are quite successful in many applications, the mathematical theory still seems to be incomplete. In this paper we study theoretical aspects of a general crossover operation in some detail.

In the classical model as described, for example, in [2] or [4], individuals (i.e., solutions) are represented by bitstrings. Crossover then mixes two bitstrings into a pair of new ones, usually by breaking both into two pieces and recombining them (one–point crossover). The survival of advantageous patterns (schemata) of bits is described in the so–called Schema Theorem, see

---

*Electronic mail address: `kolonko@informatik.uni-hildesheim.de`.

[4]. These concepts and their interplay with different isomorphic represen-
tations of the solutions have recently been investigated in a series of papers:
[1], [6], [7], and [8].

We introduce a more general concept of recombination of two individ-
uals x, y being elements of an arbitrary set I. First x, y are broken into
complementary fragments and then combined into two new individuals. All
that is required of these operations is that, in a sense, no information about
x, y is lost during recombination. Mathematically, this can be expressed
using concepts found in the theory of orthogonal projections. In section 2
we define these operations, study their symmetry properties, and give some
examples. In particular, we show that the usual one-point crossover and
uniform crossover of binary strings are included in our definition.

In section 3 we specialize on individuals that are elements of an euclidean
space, for example, $\mathbb{R}^K$. Here, to "break" an individual x into fragments
means to project x onto arbitrarily chosen linear subspaces and their orthog-
onal complements. This allows a geometric interpretation of the operations,
for example, we show that applying crossover to a pair (x, y) amounts to a
rotation of (x, y) around its center $(x + y)/2$.

In section 4 we examine the effect crossover has on the distribution of
the population. We give a general formula for the probability density of the
offspring given a joint density of the parents. As an application we show that
the uniform distribution is invariant under crossover and that the entropy of
the population increases, as it is known to do for the classical crossover
operators.

In section 5 we show for the euclidean case that expectation and overall
variance of the population remain invariant under crossover, though crossover
is sometimes considered as a dispersion operator.

In section 6 we introduce the concept of a schema adapted to our setting
and prove a Schema Theorem that includes the classical one.

2. The general crossover operator

Let I denote the set of individuals and let U be a space of indices (types
of crossover). We define pairs of crossover operators $C, \overline{C}$ such that $C_u(x, y)$
and $\overline{C}_u(x, y)$ are the two offspring individuals when applying a crossover of
type u to parent individuals $x, y \in I$. To be more precise, assume that for
each $u \in U$ there is a pair of mappings

$$\varphi_u : I \to I \qquad \text{and} \qquad \overline{\varphi}_u : I \to I \tag{1}$$

that break individuals into complementary "fragments" $\varphi_u(x)$ and $\overline{\varphi}_u(x)$
where we assume for ease of notation that fragments can be expressed as
elements of I.

To combine arbitrary fragments into a new individual, let $h : I \times I \to I$
and put

$$C_u(x, y) := h(\varphi_u(x), \overline{\varphi}_u(y)) \quad \text{and} \quad \overline{C}_u(x, y) := h(\varphi_u(y), \overline{\varphi}_u(x)) \tag{2}$$

such that $C, \overline{C}$ are mappings from $I \times I \times U \rightarrow I$. Let $T_u(x,y) :=$ $(C_u(x,y), \overline{C}_u(x,y))$ denote the ordered pair of offspring.

All we assume about $\varphi_u, \overline{\varphi}_u$ and h is the following condition that holds throughout this paper.

Condition 1. For all $x, y, z \in I$ and $u \in U$

$$h(\varphi_u(x), \overline{\varphi}_u(y)) = z \iff \varphi_u(x) = \varphi_u(z) \text{ and } \overline{\varphi}_u(y) = \overline{\varphi}_u(z).$$

Condition 1 guarantees that no "information" about x and y is lost during crossover.

Example 1.

(a) Binary case: For $K \in \mathbb{N}$ let $I := \{0,1\}^K = \mathbb{Z}_2 \times \ldots \times \mathbb{Z}_2$ as in the classical model of GAs. Let $\oplus$ and $\otimes$ denote the coordinatewise addition and multiplication modulo 2 (XOR and AND). Put $U := \{0,1\}^K$ and $\varphi_u(x) := u \otimes x$, $\overline{\varphi}(x) := (\mathbf{1} \oplus u) \otimes x$ and $h(x,y) := x \oplus y$. Then Condition 1 is fulfilled and

$$C_u(x,y) := (u \otimes x) \oplus ((\mathbf{1} \oplus u) \otimes y),$$
$$\overline{C}_u(x,y) := (u \otimes y) \oplus ((\mathbf{1} \oplus u) \otimes x).$$

To obtain the one–point crossover one has to restrict u to the binary representations of $2^k - 1$, with $k = 1, \ldots, K$.

(b) Product case: Let $I := \mathsf{X}_{k \in K} \Xi_k$ be an arbitrary product space, then $u \in U := \{0,1\}^K$ selects a parent for each of the coordinates $k \in K$. A formal definition can be given analogous to (a) such that Condition 1 holds. An important case is $I = \mathbb{R}^K$, here $\varphi_u(x)$ is the projection of x onto those coordinates which are marked by 1s in $u \in \{0,1\}^K$.

(c) Euclidean case: Even more generally, assume that I is a linear space with an inner product $\langle \cdot, \cdot \rangle$ and let $\{L_u | u \in U\}$ be a set of closed linear subspaces of I. Let φ_u resp. $\overline{\varphi}_u$ be the orthogonal projections of I onto L_u resp. onto the orthogonal complement $L_u^\perp$. Then with $h(v,w) := v + w$ Condition 1 holds. Note that this essentially includes (a) and (b).

Our crossover operators work independently and symmetrically on both parents. In some applications this kind of separability is not fulfilled. For example, in the travelling salesman problem there are crossover operators that select a subtour of one of the parents and then complete it by inserting the missing cities from the second parent, that is, $C_u(x,y) = h(\varphi_u(x), \overline{\varphi}_u(x,y))$, see [10] for a survey. A similar problem arises in some assignment problems, see [9].

Obviously, C has some nice symmetry properties such as $C_u(x,y) = \overline{C}_u(y,x)$. Lemma 1 shows that the pair $(C_u, \overline{C}_u)$ has even more structure, see [8] for a similar observation in the binary case.

Lemma 1. *For all $x, y \in I$ and all $u \in U$ we have the following.*

(a) $C_u(x, x) = \overline{C}_u(x, x) = x$.

(b) $C_u\left(C_u(x, y), \overline{C}_u(x, y)\right) = x$ and $\overline{C}_u\left(C_u(x, y), \overline{C}_u(x, y)\right) = y$.

(c) $T_u := (C_u, \overline{C}_u) : I \times I \to I \times I$ is a bijective self-inverse mapping, that is, $T_u(T_u(x, y)) = (x, y)$ for all $u \in U$, with $(x, y) \in I \times I$.

(d) For any $F \subset I \times I$ we have $T_u(F \cup T_u(F)) = F \cup T_u(F)$. More generally, $F \mapsto \tau_u(F) := F \cup T_u(F)$ is a hull operator, that is,

$$F \subset \tau_u(F), \quad F \subset F' \Rightarrow \tau_u(F) \subset \tau_u(F^{\cdot}) \quad \text{and} \quad \tau_u(\tau_u(F)) = \tau_u(F).$$

Proof. (a) follows from the "$\Leftarrow$" part of Condition 1.

(b) Put $v := C_u(x, y)$ and $w := \overline{C}_u(x, y)$. Then from Condition 1: $\varphi_u(v) = \varphi_u(x)$, $\overline{\varphi}_u(v) = \overline{\varphi}_u(y)$, $\varphi_u(w) = \varphi_u(y)$ and $\overline{\varphi}_u(w) = \overline{\varphi}_u(x)$. But then using (a) $x = h(\varphi_u(x), \overline{\varphi}_u(x)) = h(\varphi_u(v), \overline{\varphi}_u(w)) = C_u(C_u(x, y), \overline{C}_u(x, y))$ and similarly for y.

(c) follows from (b) and (d) from (c). ∎

3. Crossover in the euclidean case

We consider the euclidean case of Example 1(c) more closely. It seems to provide more intuitive understanding then the classical binary case and allows a geometric interpretation of crossover.

In this section let I be a linear space with an inner product $\langle \cdot, \cdot \rangle$ and the norm $\| x \| := \sqrt{\langle x, x \rangle}$. Let $\{L_u \mid u \in U\}$ be a set of closed linear subspaces. Then the orthogonal projections φ_u and $\overline{\varphi}_u$ on L_u and its orthogonal complement $L_u^\perp$ are known to exist. Note that if L_u has finite dimension then it is closed. Let $x + L_u := \{x + y \mid y \in L_u\}$ be the linear manifold spanned by $x \in I$ and subspace L_u. We use

$$C_u(x, y) := \varphi_u(x) + \overline{\varphi}_u(y) \quad \text{and} \quad \overline{C}_u(x, y) := \overline{\varphi}_u(x) + \varphi_u(y).$$

Lemma 2 shows that the pair of parents has certain invariance properties: their sum and their distance are not changed under crossover.

Lemma 2. *For any $x, y \in I, u \in U$ we have the following.*

(a) $x + y = C_u(x, y) + \overline{C}_u(x, y)$.

(b) $\| x - y \|^2 = \| C_u(x, y) - \overline{C}_u(x, y) \|^2$.

Proof. (a) From $x = \varphi_u(x) + \overline{\varphi}_u(x)$ we have at once

$$x + y = \varphi_u(x) + \overline{\varphi}_u(y) + \overline{\varphi}_u(x) + \varphi_u(y) = C_u(x, y) + \overline{C}_u(x, y).$$

(b) For any $a, b \in I$ we have

$$\langle a, b \rangle = 0 \quad (\text{i.e., } a \perp b) \quad \Longleftrightarrow \quad \| a + b \|^2 = \| a - b \|^2 . \tag{3}$$

Since $\varphi_u(x) - \varphi_u(y) \in L_u$ and $\overline{\varphi}_u(x) - \overline{\varphi}_u(y) \in L_u^{\perp}$ are orthogonal to each other we obtain

$$
\begin{aligned}
\| C_u(x,y) - \overline{C}_u(x,y) \|^2 &= \| (\varphi_u(x) - \varphi_u(y)) - (\overline{\varphi}_u(x) - \overline{\varphi}_u(y)) \|^2 \\
&= \| (\varphi_u(x) - \varphi_u(y)) + (\overline{\varphi}_u(x) - \overline{\varphi}_u(y)) \|^2 \\
&= \| x - y \|^2 . \quad \blacksquare
\end{aligned}
$$

Theorem 1 examines the behavior of the potential offspring of a set of parents under a fixed crossover u (see Figure 1) and of a fixed pair of parents under all possible crossovers $u \in U$ (see Figure 2). Figure 2 shows in particular that all possible offspring of parents x, y lie on

$$
S(x,y) := \{ z \in I \mid \| z - \frac{x+y}{2} \| = \| \frac{x-y}{2} \| \},
$$

that is, the surface of the smallest ball of full dimension that contains x and y. Note that starting from the viewpoint of continuous evolutionary strategies a similar result was found in [11].
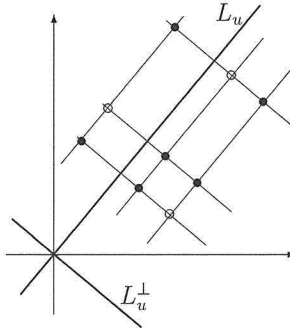

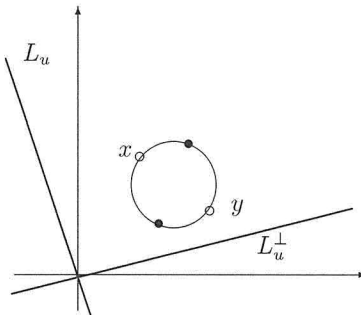Figure 1: The circles represent three parent individuals, the dots are all their possible offspring from C_u resp. $\overline{C}_u$.

Figure 2: All offspring of x, y under $T_u, u \in U$ are lying on the circle.

Theorem 1.

(a) *Let $u \in U$ be fixed and let $D \subset I$ be a set of "parents." Then*

$$D \cup \{C_u(x,y) \mid x,y \in D\} =$$
$$\bigcup\{x + L_u \mid x \in D\} \cap \bigcup\{x + L_u^\perp \mid x \in D\},$$

that is, the parents and all their possible offspring generated by crossover u occupy the points on the intersections of all linear manifolds which run through parent individuals and which are parallel to L_u or $L_u^\perp$ (see Figure 1). The same holds with C_u replaced by $\overline{C}_u$.

(b) *For any $x,y \in I$ and any $U' \subset U$*

$$\{C_u(x,y) \mid u \in U'\} \cup \{\overline{C}_u(x,y) \mid u \in U'\} \subset S(x,y).$$

Let $\{L_u \mid u \in \widehat{U}\}$ be the set of all one-dimensional linear subspaces of I. Then

$$S(x,y) = \{C_u(x,y) \mid u \in \widehat{U}\}$$

(see Figure 2), that is, applying some crossover u to a pair of parents (x,y) means to rotate (x,y) around the center $(x-y)/2$ by an angle α with

$$\cos(\alpha/2) = \frac{\| \varphi_u(x-y) \|}{\| x-y \|}.$$

Proof. (a) From $\varphi_u(z) \in L_u$ we have for any $z \in I$ $z \in x + L_u \iff \overline{\varphi}_u(z) = \overline{\varphi}_u(x)$ and similarly $z \in x + L_u^\perp \iff \varphi_u(z) = \varphi_u(x)$. Hence

$$z \in \bigcup\{x + L_u \mid x \in D\} \cap \bigcup\{x + L_u^\perp \mid x \in D\}$$
$$\iff \exists \, x,y \in D \quad z \in (x + L_u) \cap (y + L_u^\perp)$$
$$\iff \exists \, x,y \in D \quad \overline{\varphi}_u(z) = \overline{\varphi}_u(x) \text{ and } \varphi_u(z) = \varphi_u(y)$$
$$\iff z \in \{C_u(x,y) \mid x,y \in D\}.$$

(b) Assume $z \in \{C_u(x,y) \mid u \in U\}$, then there is an $u \in U$ such that $z = \varphi_u(x) + \overline{\varphi}_u(y)$. Hence using (3)

$$\| z - \frac{x+y}{2} \|^2 = \frac{1}{4} \| (z-x) + (z-y) \|^2$$
$$= \frac{1}{4} \| (\overline{\varphi}_u(y) - \overline{\varphi}_u(x)) + (\varphi_u(x) - \varphi_u(y)) \|^2$$
$$= \frac{1}{4} \| -(\overline{\varphi}_u(y) - \overline{\varphi}_u(x)) + (\varphi_u(x) - \varphi_u(y)) \|^2 = \| \frac{x-y}{2} \|^2,$$

that is, $z \in S(x,y)$. An analogous argument applies if $z \in \{\overline{C}_u(x,y) \mid u \in U\}$. On the other hand, if $z \in S(x,y)$ then we have from $\| z - (x+y)/2 \|^2 = \| (x-y)/2 \|^2$ that $\| (z-x) + (z-y) \|^2 = \| (x-z) + (z-y) \|^2$. From (3) we see that $(z-x)$ and $(z-y)$ must be perpendicular. Let $L_u := \langle z - y \rangle$ be

the one-dimensional linear subspace spanned by $(z - y)$. Then we must have $\varphi_u(z - x) = 0 = \overline{\varphi}_u(z - y)$ and hence

$$\varphi_u(z) = \varphi_u((z - x) + x) = \varphi_u(z - x) + \varphi_u(x) = \varphi_u(x)$$
$$\overline{\varphi}_u(z) = \overline{\varphi}_u((z - y) + y) = \overline{\varphi}_u(z - y) + \overline{\varphi}_u(y) = \overline{\varphi}_u(y)$$

that is, $z = C_u(x, y)$. The assertion about the rotation angle can be verified in Figure 2. ∎

For $I = \mathbb{R}^K$ one may simulate the result z of a random crossover of x and y by choosing z with a suitable distribution from $S(x, y)$. For example, if crossover uses a randomly chosen one-dimensional subspace L_u then z should be uniformly distributed on $S(x, y)$. If the L_u are chosen parallel to the coordinate axis (as in the product case of Example 1) then z should be uniformly distributed on the vertices of the K-dimensional standard cube contained in $S(x, y)$.

4. The distribution of the offspring

In this section we examine the distribution of the outcome $T_\chi(X, Y)$ of a random crossover χ applied to two randomly chosen parents X, Y. We assume that I and U are endowed with σ-algebras and that φ, $\overline{\varphi}$, and h are such that $(u, x, y) \mapsto C_u(x, y)$ and $(u, x, y) \mapsto \overline{C}_u(x, y)$ are measurable mappings. For more general readability we do not further discuss measurability questions.

Theorem 2.

(a) For any $F \subset I \times I$ we have $\mathbf{P}(T_\chi(X, Y) \in F) = \mathbf{P}((X, Y) \in T_\chi(F))$.

(b) Assume that (X, Y) and χ are independent and that (X, Y) has a density p with respect to a σ-finite measure ϱ on $I \times I$. Let ϱ fulfill

$$\varrho(T_u(F)) = \varrho(F) \qquad (4)$$

for all $F \subset I \times I$ and $u \in U$. Then the following three assertions hold.

(i) $T_\chi(X, Y)$ has the conditional ϱ-density $(v, w) \mapsto p(T_u(v, w))$ given $\chi = u$, that is, for all $F \subset I \times I$

$$\mathbf{P}(T_u(X, Y) \in F) = \int_F \varrho(dv, dw)\, p(T_u(v, w)).$$

(ii) $T_\chi(X, Y)$ has the unconditional ϱ-density

$$(v, w) \mapsto \int_U \mathbf{P}_\chi(du) p(T_u(v, w)).$$

(iii) If in addition, $\varrho = \mu \times \mu$ for some measure μ on I, then $C_\chi(X, Y)$ has the conditional resp. unconditional μ-density

$$v \mapsto \int \mu(dw)\, p(T_u(v, w)) \qquad \text{resp.}$$
$$v \mapsto \int \mu(dw) \int_U \mathbf{P}_\chi(du)\, p(T_u(v, w)).$$

Proof. (a) Since $T_u = T_u^{-1}$ we have

$$\begin{aligned}
\mathbf{P}(T_\chi(X,Y) \in F) &= \int_U \mathbf{P}_\chi(du)\, \mathbf{P}[T_u(X,Y) \in F \mid \chi = u] \\
&= \int_U \mathbf{P}_\chi(du)\, \mathbf{P}[(X,Y) \in T_u(F) \mid \chi = u] \\
&= \mathbf{P}((X,Y) \in T_\chi(F)).
\end{aligned}$$

(b) From (4) we see that $\varrho(F) = \varrho(T_u(F)) = \varrho(T_u^{-1}(F))$, that is, ϱ coincides with its image ϱ_{T_u} for any $u \in U$. Hence we obtain from the independence of (X,Y) and χ, and from a simple integral transformation:

$$\begin{aligned}
\mathbf{P}(T_\chi(X,Y) \in F) &= \int_U \mathbf{P}_\chi(du)\mathbf{P}(T_u(X,Y) \in F) \\
&= \int_U \mathbf{P}_\chi(du) \int \varrho(dx,dy)\, p(x,y)\, \mathbf{1}_F(T_u(x,y)) \\
&= \int_U \mathbf{P}_\chi(du) \int \varrho_{T_u}(dv,dw)\, p(T_u^{-1}(v,w))\, \mathbf{1}_F(v,w) \\
&= \int_U \mathbf{P}_\chi(du) \int_F \varrho(dv,dw)\, p(T_u(v,w)).
\end{aligned}$$

Now (i) – (iii) follow. ∎

Equation (4) means that ϱ is invariant under the mappings T_u. The following examples show that this condition holds in some of the most important cases.

Example 2.

(a) If I is countable and ϱ is the usual counting measure we have:

$$\varrho(T_u(F)) = |T_u(F)| = |F| = \varrho(F)$$

hence (4) holds and we obtain from Theorem 2(b) part (iii)

$$\mathbf{P}(C_\chi(X,Y) = v) = \sum_{u \in U}\sum_{w \in I} \mathbf{P}(\chi = u)\mathbf{P}(X = C_u(v,w))\mathbf{P}(X = \overline{C}_u(v,w))$$

assuming that U is countable too and that X and Y are drawn independently and identically distributed.

(b) Consider the product case and assume that there are σ-finite measures μ_k on Ξ_k, $k \in K$, and let $\mu := \mathsf{X}_{k \in K}\mu_k$. Then with $A_k, B_k \subset \Xi_k$, $k \in K$, we have for any $u \in U$

$$C_u(\mathsf{X}_{k \in K} A_k \times \mathsf{X}_{k \in K} B_k) = \mathsf{X}_{k \in K} D_k^{u_k}$$

and

$$\overline{C}_u(\mathsf{X}_{k \in K} A_k \times \mathsf{X}_{k \in K} B_k) = \mathsf{X}_{k \in K} \overline{D}_k^{u_k}$$

where $D_k^1 := A_k, D_k^0 := B_k$ and $\overline{D}_k^1 := B_k, \overline{D}_k^0 := A_k$. Using Lemma 1(c) we have for $\varrho := \mu \times \mu$

$$\varrho\left(T_u\left(\underset{k\in K}{\mathsf{X}}A_k \times \underset{k\in K}{\mathsf{X}}B_k\right)\right) = \varrho\left(\left(\underset{k\in K}{\mathsf{X}}D_k^{u_k} \times \underset{k\in K}{\mathsf{X}}\overline{D}_k^{u_k}\right)\right)$$

$$= \prod_{k\in K}\mu_k(A_k) \cdot \prod_{k\in K}\mu_k(B_k)$$

$$= \varrho\left(\underset{k\in K}{\mathsf{X}}A_k \times \underset{k\in K}{\mathsf{X}}B_k\right),$$

that is, condition (4) of Theorem 2 is fulfilled.

(c) Now consider the euclidean case with $I = \mathbb{R}^K$, $K < \infty$, and φ_u a projection on some $m(u)$-dimensional subspace L_u of $\mathbb{R}^K$. To give an explicit expression of φ_u let B_u be an orthonormal $K \times K$ matrix whose first $m(u)$ columns span L_u and let E_u resp. $\overline{E}_u$ be the projections on the first $m(u)$ resp. last $K - m(u)$ coordinates of $\mathbb{R}^K$, that is, ·

$$E_u := \begin{pmatrix} 1 & 0 & \cdots\cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & \ddots & \vdots \\ \vdots & & \ddots & 0 & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & 0 \\ 0 & \cdots\cdots & & 0 & 0 \end{pmatrix} \qquad \overline{E}_u := \begin{pmatrix} 0 & 0 & \cdots\cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & 0 & \ddots & \vdots \\ \vdots & & \ddots & 1 & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & 0 \\ 0 & \cdots\cdots & & 0 & 1 \end{pmatrix}.$$

Then, as is well known from linear algebra, $\varphi_u(x) = B_u E_u B_u' x$ and $\overline{\varphi}(y) = B_u \overline{E}_u B_u' y$ where B_u' denotes the transposed matrix. As $h(v, w) = v + w$ is symmetrical, the transformation $(x, y) \mapsto T_u(x, y)$ is given by the matrix

$$T_u = \begin{pmatrix} B_u & 0 \\ 0 & B_u \end{pmatrix} \begin{pmatrix} E_u & \overline{E}_u \\ \overline{E}_u & E_u \end{pmatrix} \begin{pmatrix} B_u' & 0 \\ 0 & B_u' \end{pmatrix}.$$

Let $\varrho := \lambda^K \times \lambda^K$ where λ^K denotes the K-dimensional Lebesgue measure on $\mathbb{R}^K$. Then we have, from the transformation theorem for $(2K)$-dimensional Lebesgue integrals, for $F \subset \mathbb{R}^K \times \mathbb{R}^K$:

$$\varrho_{T_u}(F) = \int_F |\det(T_u)| \, d(\lambda^K \times \lambda^K) = \varrho(F)$$

as $|\det(T_u)| = 1$ is easily verified. Hence, the condition of (4) holds in this case too.

As an application we consider two properties that are known to hold for classical crossover operators: that the uniform distribution is invariant under crossover and that the entropy increases.

Corollary 1. *Assume that the conditions of Theorem 2(b) hold.*

(a) *If (X, Y) is uniformly distributed over some set $F \subset I \times I$ then given $\chi = u$ the conditional distribution of the pair of offspring $T_\chi(X, Y)$ is the uniform distribution on $T_u(F)$. In particular, if X, Y are drawn independently and uniformly distributed from I then $C_\chi(X, Y)$ and $\overline{C}_\chi(X, Y)$ are also independent and uniformly distributed.*

(b) *For the entropy* $\mathbf{H}(X,Y) := -\mathbf{E}\log(p(X,Y))$ *we have (from[12]):*

  (i) $\mathbf{H}(X,Y) = \mathbf{H}(T_u(X,Y))$ *for any* $u \in U$ *and*

  (ii) $\mathbf{H}(X,Y) < \mathbf{H}(T_\chi(X,Y))$.

Proof. (a) The uniform distribution on some set F with $0 < \varrho(F) < \infty$ has the density $p(x,y) := \mathbf{1}_F(x,y)/\varrho(F)$, hence by Theorem 2(b) part (i), $T_u(X,Y)$ has density

$$
\begin{aligned}
p(T_u(x,y)) &= \varrho(F)^{-1}\mathbf{1}_F(T_u(x,y)) \\
&= \varrho(T_u^{-1}(F))^{-1}\mathbf{1}_{T_u^{-1}(F)}(x,y) \\
&= \varrho(T_u(F))^{-1}\mathbf{1}_{T_u(F)}(x,y).
\end{aligned}
$$

If I allows a uniform distribution, that is, there is a measure μ with $0 < \mu(I) < \infty$ then $p(x,y) = \mu(I)^{-2}\mathbf{1}_I(x)\mathbf{1}_I(y)$, hence for $F_1, F_2 \subset I$

$$
\begin{aligned}
\mathbf{P}(C_\chi(X,Y) &\in F_1, \overline{C}_u(X,Y) \in F_2) \\
&= \int_{F_1 \times F_2}(\mu \times \mu)(dx,dy)\int_U \mathbf{P}_\chi(du)\mu(I)^{-2}\big(\mathbf{1}_I(C_u(x,y)) \cdot \mathbf{1}_I(\overline{C}_u(x,y))\big) \\
&= \mu(I)^{-2}\mu(F_1)\mu(F_2).
\end{aligned}
$$

  (b) From Theorem 2(b) part (i) we have

$$
\begin{aligned}
\mathbf{H}(T_u(X,Y)) &= -\int \varrho(dx,dy)p(T_u(x,y))\log(p(T_u(x,y))) \\
&= -\mathbf{E}\log\big(p(T_u(T_u(X,Y)))\big) = -\mathbf{E}\log(p(X,Y)) = \mathbf{H}(X,Y).
\end{aligned}
$$

Since $x \mapsto -x\log x$ is strictly concave we have from the strict Jensen inequality

$$
\begin{aligned}
\mathbf{H}(T_\chi(X,Y)) &= -\int \varrho(dx,dy)\Big(\int_U \mathbf{P}_\chi(du)p(T_u(x,y))\Big)\log\Big(\int_U \mathbf{P}_\chi(du)p(T_u(x,y))\Big) \\
&> -\int_U \mathbf{P}_\chi(du)\int \varrho(dx,dy)\, p(T_u(x,y))\,\log(p(T_u(x,y))) \\
&= \int_U \mathbf{P}_\chi(du)\mathbf{H}(T_u(X,Y)) \\
&= \mathbf{H}(X,Y). \quad \blacksquare
\end{aligned}
$$

5. The distribution in the euclidean case

In this section we examine the stochastic effects of crossover in the euclidian case more closely. For one-point and uniform crossover it is well known that within a population represented by a random variable $X = (X_1, \ldots, X_K)$ the marginal distribution of X_i is not changed by crossover. Theorem 3 shows that this is true for the general crossover for the expected value $\mathbf{E}\, X_i$ at each coordinate $i = 1, \ldots, K$. Also the variances $\mathbf{V}(X_1), \ldots, \mathbf{V}(X_K)$ change depending on the covariance of the corresponding coordinates of $\varphi_u(X)$ and

$\overline{\varphi}_u(X)$, but the overall variance $\sum_i \mathbf{V}(X_i) = \mathbf{E} \parallel X - \mathbf{E} X \parallel^2$ remains unchanged.

Let $C_{\chi,1}(X,Y), \ldots, C_{\chi,K}(X,Y)$ be the coordinates of $C_\chi(X,Y)$ and similarly $C_{u,i}(X,Y), \varphi_{u,i}$ and $\overline{\varphi}_{u,i}$ for $i = 1, \ldots, K$.

Theorem 3. *Let $I = \mathbb{R}^K$ and assume that $X = (X_1, \ldots, X_K)$ and $Y = (Y_1, \ldots, Y_K)$ are two independent and identically distributed random vectors. Let (X,Y) and χ be independent.*

(a) $\mathbf{E} X = \mathbf{E} C_\chi(X,Y)$, *that is,* $\mathbf{E} X_i = \mathbf{E} C_{\chi,i}(X,Y)$ *for all* $1 \le i \le K$.

(b) *For all $1 \le i \le K$ and $u \in U$ we have*

$$\mathbf{V}(C_{u,i}(X,Y)) = \mathbf{V}(X_i) - 2 \operatorname{cov}(\varphi_{u,i}(X), \overline{\varphi}_{u,i}(X))$$

and

$$\sum_{i=1}^K \mathbf{V}(C_{\chi,i}(X,Y)) = \sum_{i=1}^K \mathbf{V}(X_i).$$

Proof. (a) We have, since X, Y are identically distributed,

$$\begin{aligned}
\mathbf{E} C_{u,i}(X,Y) &= \mathbf{E}(\varphi_{u,i}(X) + \overline{\varphi}_{u,i}(Y)) && (5)\\
&= \mathbf{E} \varphi_{u,i}(X) + \mathbf{E} \overline{\varphi}_{u,i}(Y) \\
&= \mathbf{E} \varphi_{u,i}(X) + \mathbf{E} \overline{\varphi}_{u,i}(X) \\
&= \mathbf{E} X_i.
\end{aligned}$$

Hence $\mathbf{E} C_{\chi,i}(X,Y) = \int_U \mathbf{P}_\chi(du) \mathbf{E} C_{u,i}(X,Y) = \mathbf{E} X_i$.

(b) We suppress the index u in $\varphi_{u,i}$ for better readability. We have from straightforward calculations

$$\begin{aligned}
\mathbf{E} C_{u,i}(X,Y)^2 &= \mathbf{E}[\varphi_i(X) + \overline{\varphi}_i(Y)]^2 \\
&= \mathbf{E}[\varphi_i(X)^2] + 2\mathbf{E}\,\varphi_i(X)\mathbf{E}\,\overline{\varphi}_i(Y) + \mathbf{E}[\overline{\varphi}_i(Y)^2] \\
&= \mathbf{E}[\varphi_i(X)^2] + 2\mathbf{E}\,\varphi_i(X)\mathbf{E}\,\overline{\varphi}_i(X) + \mathbf{E}[\overline{\varphi}_i(X)^2] \\
&= \mathbf{E}[\varphi_i(X) + \overline{\varphi}_i(X)]^2 - 2\mathbf{E}[\varphi_i(X)\overline{\varphi}_i(X)] + 2\mathbf{E}\,\varphi_i(X)\mathbf{E}\,\overline{\varphi}_i(X) \\
&= \mathbf{E} X_i^2 - 2[\mathbf{E}(\varphi_i(X)\overline{\varphi}_i(X)) - \mathbf{E}\,\varphi_i(X)\mathbf{E}\,\overline{\varphi}_i(X)] \\
&= \mathbf{E} X_i^2 - 2 \operatorname{cov}(\varphi_i(X), \overline{\varphi}_i(X)).
\end{aligned}$$

Hence the first assertion of (b) follows from (5). For the second assertion it suffices to show $\sum_{i=1}^K \operatorname{cov}(\varphi_{u,i}(X), \overline{\varphi}_{u,i}(X)) = 0$ for all $u \in U$ as we have

$$\begin{aligned}
\sum_{i=1}^K \mathbf{V}(C_{\chi,i}(X,Y)) &= \sum_{i=1}^K \left(\int_U \mathbf{P}_\chi(du) \mathbf{E} C_{u,i}(X,Y)^2 - (\mathbf{E} C_{\chi,i}(X,Y))^2 \right) \\
&= \sum_{i=1}^K \left(\mathbf{E} X_i^2 - 2 \int_U \mathbf{P}_\chi(du) \operatorname{cov}(\varphi_{u,i}(X), \overline{\varphi}_{u,i}(X)) - (\mathbf{E} X_i)^2 \right) \\
&= \sum_{i=1}^K \mathbf{V}(X_i) - 2 \int_U \mathbf{P}_\chi(du) \sum_{i=1}^K \operatorname{cov}(\varphi_{u,i}(X), \overline{\varphi}_{u,i}(X)).
\end{aligned}$$

We use the representation of $\varphi_u(X)$ from Example 2(c), that is, $\varphi_u(x) = B_u E_u B'_u x$ and $\overline{\varphi}(y) = B_u \overline{E}_u B'_u y$ where B_u is an orthogonal matrix whose first $m(u)$ columns span L_u and whose last $K - m(u)$ columns span $L_u^\perp$. Note that $B_u \overline{E}_u B'_u$ and $B_u E_u B'_u$ are symmetric. Let $[B_u E_u B'_u]_i$ denote the ith row of $B_u E_u B'_u$, then

$$
\begin{aligned}
\mathrm{cov}(\varphi_{u,i}(X), \overline{\varphi}_{u,i}(X)) &= \mathrm{cov}([B_u E_u B'_u]_i X \ , \ [B_u \overline{E}_u B'_u]_i X) \\
&= [B_u E_u B'_u]_i \ \mathrm{cov}(X, X) \ ([B_u \overline{E}_u B'_u]_i)' \\
&= [B_u E_u B'_u]_i \ \mathrm{cov}(X, X) \ [B_u \overline{E}_u B'_u]_i'.
\end{aligned}
$$

The last expression is just the ith diagonal element of the matrix $B_u E_u B'_u \ \mathrm{cov}(X, X) \ B_u \overline{E}_u B'_u$. Now using the well known-rule $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ for the trace $\mathrm{tr}(A) = \sum_{i=1}^{K} a_{i,i}$ we have

$$
\begin{aligned}
\sum_{i=1}^{K} \mathrm{cov}(\varphi_{u,i}(X), \overline{\varphi}_{u,i}(X)) &= \mathrm{tr}(B_u E_u B'_u \ \mathrm{cov}(X, X) \ B_u \overline{E}_u B'_u) \\
&= \mathrm{tr}(\ \mathrm{cov}(X, X) \ B_u \overline{E}_u B'_u \ B_u E_u B'_u) \\
&= \mathrm{tr}(\ \mathrm{cov}(X, X) \ B_u \overline{E}_u E_u B'_u) \\
&= 0,
\end{aligned}
$$

where we used the fact that B_u is orthogonal and $\overline{E}_u E_u = 0$. ∎

6.   Schemata

In this section we sketch a concept of schemata that generalizes the classical notion and is compatible with our concept of crossover.

For any $a \in I, u \in U$ define

$$
\Phi_u(a) := \varphi_u^{-1} \circ \varphi_u(a) = \{z \in I \mid \varphi_u(z) = \varphi_u(a)\}
$$

and similarly

$$
\overline{\Phi}_u(a) := \overline{\varphi}_u^{-1} \circ \overline{\varphi}_u(a).
$$

We shall call $\Phi_u(a)$ the schema defined by u and a. Note that in our definition the connection between schemata and crossover is more obvious than in the classical notation using wildcard symbols (e.g., [2]).

Example 3.

(a) In the binary case, let φ_u be as in Example 1(a). Then $\Phi_u(a)$ is the classical schema with fixed positions a_k where $u_k = 1$ and with wildcard symbols $'\#'$ where $u_k = 0$.

(b) In the euclidean case schemata are linear manifolds $\Phi_u(a) = \varphi_u(a) + L_u^\perp$ as was noted in the proof of Theorem 1(b). In the case $I = \mathbb{R}^K$ we may consider an orthonormal basis $B := \{b_1, \ldots, b_K\}$ of $\mathbb{R}^K$. For $U := \{0, 1\}^K$ and $u \in U$ let $L_u := \langle b_i \mid u_i = 1, 1 \leq i \leq K \rangle$ be the linear subspace spanned by those b_i with $u_i = 1$. Then $L_u^\perp = \langle b_i \mid u_i = 0, 1 \leq i \leq K \rangle$ and φ_u is the projection on the coordinates selected by u. This allows a straightforward generalization of (a).

For a generalization of the classical Schema Theorem in [4] we first have to consider the interplay of crossover and schemata. Lemma 3 describes the nondisruption of schemata during crossover.

Lemma 3. *Let $u, u' \in U$ and $y, a \in I$.*

(a) *If $\varphi_u \circ \varphi_{u'} = \varphi_u$ then $C_{u'}(\cdot, y)$ maps $\Phi_u(a)$ into itself.*

(b) *If $\varphi_u \circ \overline{\varphi}_{u'} = \varphi_u$ then $\overline{C}_{u'}(\cdot, y)$ maps $\Phi_u(a)$ into itself.*

Proof. Let $x \in \Phi_u(a)$ and $\varphi_u \circ \varphi_{u'} = \varphi_u$, then from Condition 1

$$\varphi_u(C_{u'}(x, y)) = \varphi_u(h(\varphi_{u'}(x), \overline{\varphi}_{u'}(y))) = \varphi_u \circ \varphi_{u'}(h(\varphi_{u'}(x), \overline{\varphi}_{u'}(y)))$$
$$= \varphi_u \circ \varphi_{u'}(x) = \varphi_u(x) = \varphi_u(a)$$

and similarly for (b). ■

Example 4.

(a) In the euclidean case we have $\varphi_u \circ \varphi_{u'} = \varphi_u \iff L_u \subset L_{u'}$ and $\varphi_u \circ \overline{\varphi}_{u'} = \varphi_u \iff L_u \subset L_{u'}^{\perp}$.

(b) In the binary case with one–point crossover we have $\varphi_u \circ \varphi_{u'} = \varphi_u$ if and only if $u \leq u'$, where u' is restricted to the binary representations of $2^k - 1, 1 \leq k \leq K$. In other words, $\varphi_u \circ \varphi_{u'} = \varphi_u$ (resp. $\varphi_u \circ \overline{\varphi}_{u'} = \varphi_u$) holds if and only if the crossover point encoded in u' (i.e., the leftmost 1) falls left (resp. right) of all fixed positions of the schema $\Phi_u(a)$.

Theorem 4 gives a lower bound on the survival probability of a schema during selection according to fitness and crossover (we do not take mutation into account). It is thus a Schema Theorem for our setting.

Let the current population be represented by a random variable X with values in I and a μ-density q on I. Let $f : I \to \mathbb{R}$ be a fitness function with $\mathbf{E}\, f(X) = \int_I \mu(dx) q(x) f(x) < \infty$. Then parents selected proportional to their fitness are two independent identically distributed random variables $\widehat{X}, \widehat{Y}$ with μ-densities $x \mapsto \widehat{q}(x) := q(x) f(x) / \mathbf{E}\, f(X)$. The result of selecting parents by fitness and applying a randomly chosen crossover χ is given by $C_\chi(\widehat{X}, \widehat{Y})$ resp. $\overline{C}_\chi(\widehat{X}, \widehat{Y})$.

Theorem 4. *Let $X, \widehat{X}, \widehat{Y}$ be given as above. For $u \in U$ fixed let $\widehat{U} := \{u' \in U \mid \varphi_u \circ \varphi_{u'} = \varphi_u \text{ or } \varphi_u \circ \overline{\varphi}_{u'} = \varphi_u\}$. Then for any $a \in I$*

$$\mathbf{P}\Big(C_\chi(\widehat{X}, \widehat{Y}) \in \Phi_u(a) \quad \text{or} \quad \overline{C}_\chi(\widehat{X}, \widehat{Y}) \in \Phi_u(a)\Big)$$
$$\geq \mathbf{P}(X \in \Phi_u(a))\, \mathbf{P}(\chi \in \widehat{U})\, \frac{\mathbf{E}\,[f(X) \mid X \in \Phi_u(a)]}{\mathbf{E}\, f(X)}.$$

Proof. From Lemma 3(b) we see that $\widehat{U}$ is the set of crossovers u' such that schema $\Phi_u(a)$ is preserved either by $C_{u'}$ or $\overline{C}_{u'}$.

$$\mathbf{P}\Big(C_\chi(\widehat{X},\widehat{Y}) \in \Phi_u(a) \quad \text{or} \quad \overline{C}_\chi(\widehat{X},\widehat{Y}) \in \Phi_u(a)\Big)$$

$$= \int_U \mathbf{P}_\chi(du') \int_I \mu(dy)\,\widehat{q}(y)\,\mathbf{P}\Big(C_{u'}(\widehat{X},y) \in \Phi_u(a)$$

$$\text{or} \quad \overline{C}_{u'}(\widehat{X},y) \in \Phi_u(a)\Big)$$

$$\geq \int_{\widehat{U}} \mathbf{P}_\chi(du') \int_I \mu(dy)\,\widehat{q}(y)\,\mathbf{P}\Big(C_{u'}(\widehat{X},y) \in \Phi_u(a)$$

$$\text{or} \quad \overline{C}_{u'}(\widehat{X},y) \in \Phi_u(a)\Big)$$

$$\geq \int_{\widehat{U}} \mathbf{P}_\chi(du') \int_I \mu(dy)\,\widehat{q}(y)\,\mathbf{P}(\widehat{X} \in \Phi_u(a))$$

$$= \mathbf{P}(\chi \in \widehat{U})\,\mathbf{P}(\widehat{X} \in \Phi_u(a)).$$

Now the assertion follows as

$$\mathbf{P}(\widehat{X} \in \Phi_u(a)) = \int_{\Phi_u(a)} \mu(dx)\widehat{q}(x)$$

$$= \frac{\int_{\Phi_u(a)} \mu(dx)q(x)f(x)}{\mathbf{E}\,f(X)}$$

$$= \frac{\mathbf{E}\,[f(X) \mid X \in \Phi_u(a)]}{\mathbf{E}\,f(X)} \cdot \mathbf{P}(X \in \Phi_u(a)). \ \blacksquare$$

Note that the conditional expected fitness $\mathbf{E}\,[f(X) \mid X \in \Phi_u(a)]$ is the correct expression for what is sometimes referred to as the average fitness or average utility of the schema $\Phi_u(a)$. Theorem 4 shows that the lower bound of the survival probability is increasing with the relative average fitness of the schema and with the probability of nondisruption. The usefulness of schema theorems is often discussed in literature (e.g., [3], [5]). Our aim here was to show that we need much less structure for the crossover operator than is usually assumed in this context.

Example 5. For the binary case with one-point crossover we have $\mathbf{P}(\chi = u') = 1/K$ for any $u' \in U'$ where U' is the set of all binary representations of $2^k - 1, k = 1, \ldots, K$. We get the classical Schema Theorem from Theorem 4 as (see Example 4(b))

$$\mathbf{P}(\chi \in \widehat{U}) = \mathbf{P}(\varphi_u \circ \varphi_\chi = \varphi_u \text{ or } \varphi_u \circ \overline{\varphi}_\chi = \varphi_u)$$

$$= \mathbf{P}(\chi \geq u \quad \text{or} \quad \chi < u) = 1 - \frac{d(u)}{K}$$

where $d(u)$ is the defining length of schema $\Phi_u(a)$, that is, the maximal distance of ones in u. For the case $q(x) \equiv 2^{-K}$, which is often assumed, we obtain

$$\mathbf{E}\,[f(X) \mid X \in \Phi_u(a)] = \frac{\sum_{x \in \Phi_u(a)} f(x)}{|\Phi_u(a)|}.$$

References

[1] Battle, D. L. and M. D. Vose, "Isomorphisms of Genetic Algorithms," *Artificial Intelligence* **60** (1993) 155–165.

[2] Goldberg, D. E., *Genetic Algorithms in Search, Optimization, and Machine Learning* (Addison-Wesley, Reading, 1989).

[3] Grefenstette, J. J. and J. E. Baker, "How Genetic Algorithms Work: A Critical Look at Implicit Parallelism," *Proceedings of the Third International Conference on Genetic Algorithms*, (1989) 20–27.

[4] Holland, J. H., *Adaptation in Natural and Artificial Systems*, (MIT Press, Ann Arbor, 1975).

[5] Levenick, J. R., "Holland's Schema Theorem Disproved?" *Journal on Experimental and Theoretical Artificial Intelligence* **2** (1990) 179–183.

[6] Liepins, G. E. and M. D. Vose, "Representational Issues in Genetic Algorithms," *Journal on Experimental and Theoretical Artificial Intelligence* **2** (1990) 101–115.

[7] Liepins, G. E. and M. D. Vose, "Polynomials, Basis Sets, and Deceptiveness in Genetic Algorithms," *Complex Systems* **5** (1991) 45–61.

[8] Liepins, G. E. and M. D. Vose, "Characterizing Crossover in Genetic Algorithms," *Annals of Mathematics and Artificial Intelligence* **5** (1992) 27–34.

[9] Nissen, V., "A New Efficient Evolutionary Algorithm for the Quadratic Assignment Problem," *Operations Research Proceedings 1992*, edited by Hansmann, K.-W., *et. al.* (Springer-Verlag, Berlin, 1993).

[10] Oliver, I. M., D. J. Smith and J. R. C. Holland, "A Study of Permutation Crossover Operators on the Traveling Salesman Problem," *Proceedings of the Second International Conference on Genetic Algorithms*, (1987) 224–230.

[11] Rechenberg, I., *Evolutionsstrategien '94.* (frommann-holzboog, Stuttgart, 1994), chapter 11.

[12] Voget, S., "Aspekte genetischer Algorithmen: mathematische Modellierung und Einsatz in der Fahrplanerstellung," doctoral dissertation, University of Hildesheim, 1995.