## **One-Point Geometric Crossover**

Alberto Moraglio

School of Computing and Center for Reasoning, University of Kent, Canterbury, UK A.Moraglio@kent.ac.uk

**Abstract.** Uniform crossover for binary strings has a natural geometric interpretation that allows us to generalize it *rigorously* to any search space endowed with a notion of distance and any representation [6]. In this paper, we present an analogous characterization for one-point crossover and explicitly derive formally specific one-point crossovers for a number of well-known representations.

#### 1 Introduction

Geometric crossover [6] is a representation-independent formalization of crossover which requires the offspring to be in the metric segment between parents for some distance. This formalization encompasses many recombination operators across representations [5]. The geometric framework sees recombination operators acting on representations as dual and equivalent to suitably formalized versions of combination operators acting on the neighborhood structure (e.g., pathrelinking). Early studies on the relation between crossover and path-relinking in the space of permutations linked with a notion of distance were pioneered by Reeves [7]. The geometric interpretation of crossover operators is interesting as it clarifies what distance/search space is associated with them and how the search operator can be seen as navigating the search space (i.e., sampling offspring in the space-specific segment between parents). This, in turns, clarifies what the fitness landscape induced by a certain crossover operator is, and how the search of this operator relates to the "geographic" structure of the fitness landscape. Ultimately, this characterization may form the basis for a unified, representation-independent theory of evolutionary algorithms [5].

The definition of geometric crossover is rather coarse as it does not consider the actual probability of a particular offspring of being sampled. More finely grained subclasses of geometric crossover can be defined by specifying a probability distribution of the offspring over the segment. Perhaps, the most natural among such subclasses is the uniform geometric crossover, in which offspring are drawn from a uniform distribution over the metric segment between parents. In previous work [6], we have shown that all mask-based crossovers for binary strings are geometric crossover, as the offspring strings are drawn from a uniform distribution over the metric segment between parent strings under Hamming distance. This equivalence allows us to generalize naturally uniform crossover for binary strings to any representation by making use of the definition of uniform geometric crossover and replacing the Hamming distance with another distance defined on the target representation. For example, the uniform geometric crossover for the swap distance between permutations (i.e., the minimum number of swaps needed to transform a permutation in the other) requires the offspring permutations to be drawn uniformly from the metric segment between parent permutations. This is a formal specification of a recombination operator on permutations rather than the actual operator. To turn this specification into a procedural definition that tells how to manipulate parent permutations to obtain offspring permutations, one can rely on the observation that picking offspring in the segment between two parent permutations is equivalent to generating partially sorted permutations on a minimal sorting trajectories obtained while sorting one parent permutation into the other parent permutation using swaps. Therefore, this operator can be implemented by adapting a traditional sorting algorithm [5]. Importantly, the uniform crossover for binary strings and the recombination for permutations above are fundamentally the same recombination operator instantiated to two different spaces as both share exactly the same geometric definition, that of uniform geometric crossover.

One-point crossover for binary strings [2] selects a common crossover point uniformly at random on the length of the parent strings and produces two offspring by swapping the tails of the parent strings after the crossover point. Whereas it was possible to generalize uniform crossover across representations due essentially to its highly symmetric definition, the situation seems different for one-point crossover as cutting and swapping tails is an operation which relies much on the special characteristic of binary strings of being a vector. In this paper, we show that also one-point crossover has a simple characterization in geometric terms which allows us to generalize it rigorously to *any* representation. We then specify it for a number of well-known representations.

### 2 Generalization of One-Point Crossover

In the following, we generalize a version of one-point crossover for binary strings that returns only one offspring, in which the head comes from the first parent and the tail comes from the second parent. To do this, we first show how shortest paths between two binary strings in the Hamming space can be constructed (see figure 2). The sequence s of binary strings  $s_0, s_1, \ldots, s_5$  is generated from parent a and b using the bit ordering p which specifies when to exchange the bits between parents. Note that, in the sequence s, the strings  $s_1, s_2$  and  $s_3$  coincide. The bottom part of the figure reports the sequence s' obtained form s after removing repetitions.

**Theorem 1.** Every bit ordering generates a shortest path between a and b in the Hamming space. All shortest paths can be generated by using all possible bit orderings and there are HD(a, b)! distinct shortest paths between a and b.

*Proof.* We have a few remarks that will lead to prove this theorem. First, the sequence s' forms a path in the Hamming space as two consecutive strings in the sequence differ in exactly one bit. This is

Fig. 1. Example of shortest path generation.

a shortest path as its length equals the Hamming distance between a and b. This holds for any choice of the parent strings and bit ordering. Second, the fact that in the sequence s there are subsequences of repeated elements has its origin in the fact that at some positions the strings a and b do not differ. When the contents at those positions are exchanged the newly generated string equals the previous string in the sequence. So, if we restrict the bit ordering that generates the sequence to those positions in which a and b differ, we can generate the sequence s' directly without repetitions. By definition, the number of positions in which a and b differ is HD(a, b), so the reduced bit ordering p' needs to be an order of HD(a, b) elements. Third, distinct reduced bit orderings produce distinct sequences. Forth, all bit orderings defines all ways of how to apply sequentially all the differences between aand b to a to be turned into b. Hence, all reduced bit orderings, that is, all orders of application of these differences, account for all shortest paths between a and b. So we have HD(a, b)! distinct shortest paths between a and b.

**Corollary 1.** All possible offspring of one-point crossover for binary strings are on a single geodesic (shortest path) in the Hamming space between parents.

*Proof.* All the offspring generated by one-point crossover can be generated from the left to the right by exchanging one by one the contents of the two parents at each position. This is the same sequence obtained by applying the bit ordering (1 ... n) to the two parents where n is the length of the parents.

The corollary makes precise and proves an idea of Whitley [9] who noticed earlier that the offspring of one-point crossover forms paths in the Hamming space between parents. Since the notion of geodesic is well-defined in every metric space, the previous theorem allows us to generalize one-point crossover to any metric space, as follows.

# **Definition 1.** (One-point geometric crossover) In one-point geometric crossover all offspring are on a single geodesic between parents under some distance.

Notice that in general there may be more than one geodesic between two points (parents). The definition leaves deliberately the geodesic unspecified, but it requires that all offspring are on it. The reason we do not specify a specific geodesic is that since they are all indistinguishable from a distance viewpoint, we cannot specify anyone in particular using only the distance. Only using extrainformation based on the underlying representation, we can refer to one geodesic in particular.

This necessary indetermination in the definition of one-point geometric crossover makes it an "over-generalization" of the one-point crossover for binary strings, as explained as follows. There are three distinct types of indetermination in this definition: (i) the specific geodesic is not specified, (ii) the specific probability distribution of the offspring on the geodesic is not specified and, (iii) the specific distance is not specified. Therefore, even when both the probability distribution of the offspring is fixed (e.g., uniform probability on the geodesic) and the operator is instantiated to a specific search space (i.e., the distance is fixed), the indetermination about the specific geodesic makes the one-point operator non-uniquely determined. This is unlike the case of uniform geometric crossover which is unique with respect to its underlying metric space, as the segment between two points, unlike a geodesic, is uniquely determined by these points.

As a consequence of this indetermination, when the geometric one-point crossover is instantiated to the space of binary strings under Hamming distance, it gives rise to a family of crossover operators, one operator for each geodesic between parents, rather than only to the original one-point crossover, which corresponds to a specific geodesic. This family of operators is completely characterized by generating offspring on shortest paths using any possible bit ordering. Every crossover of this family corresponds to a bit ordering, and the strings returned by the application of this ordering to the parent strings correspond to the offspring set of that specific crossover operator (for those parent strings).

When a space allows for a family of one-point geometric crossovers, rather than a single one, they are all indistinguishable from a distance viewpoint. So all are the "right" one-point geometric crossover for the specific space. We will see that in some spaces one member of this family may be preferable to others for reasons linked with the specific character of the underlying representation.

#### 3 Euclidean and Manhattan Spaces

Since in the Euclidean space (endowed with Euclidean distance) a segment comprises a single geodesic, in this space there is only one possible specification of one-point geometric crossover. Both uniform one-point geometric crossover, in which offspring are drawn uniformly at random on a geodesic, and uniform geometric crossover when specified for the Euclidean space pick points uniformly at random on the only geodesic. So they are equivalent.

In the 2-D Manhattan space (endowed with Manhattan distance), a segment is a rectangle and its endpoints are two diagonally opposite corners. In higher dimensions, a segment is a hyper-rectangle. For this specific space, uniform geometric crossover picks uniformly offspring in the hyper-rectangle. In this space, a segment comprises infinitely many geodesics linking two points. The geodesics between two points are all monotonic curves joining the two points. So in the Manhattan space there are infinitely many one-point crossovers, one for each monotonic curve connecting the two parents.

#### 4 Permutations

In previous work [5], we have shown that PMX, Cycle Crossover, Merge Crossover and others are geometric crossovers. In the introduction, we mentioned that geometric crossovers for permutations are naturally associated with sorting algorithms, giving rise to sorting crossovers [5]. In the following, we consider two different perspectives on one-point geometric crossover for permutations: sorting crossovers and cut-and-fill crossovers.

Let us consider sorting crossovers. Given two permutations, and a move on permutation, e.g., swap of two elements, deterministic sorting algorithms sort the elements of one permutation into the order of the elements of the other permutation always on the same minimal sorting trajectory out of all the possible minimal sorting trajectories. A minimal sorting trajectory corresponds to a geodesic, i.e., a shortest path, on the metric space induced by the sorting move (e.g., the swap move induces a space on permutations endowed with the swap distance). So, the associated deterministic sorting crossovers pick offspring always on the same geodesic between two parents. Hence, deterministic sorting crossovers are one-point geometric crossovers. In randomized sorting algorithms, the sorting trajectory is still minimal but non-deterministic. The sorting crossovers based on randomized sorting algorithms, although being geometric crossovers, are not one-point geometric (under the space induced by the sorting move)  $^1$  because for two given permutations (parents), different applications of the sorting crossover may return partially ordered permutations (offspring) belonging to different sorting trajectories (geodesics).

Let us now consider cut-and-fill crossovers [1], which are intuitive extensions of one-point crossover for permutations, as follows. If the one-point crossover for binary strings is applied directly on permutations, the offspring so obtained are not permutations. So, this operator cannot be applied as it is but it can be easily adapted: the first parent is cut at a crossover point and the part before the cutting point is passed to the offspring as in the traditional one-point crossover; the second part is then filled in using the order in the second parent avoiding elements already present in the offspring before the crossover point. We call this crossover insertion cut-and-fill crossover. Figure 4 (left) shows an example of this crossover. P1 and P2 are the parent permutations and O is the offspring permutation. The vertical bar in P1 indicates the crossover point, and the dashes indicate elements of parent P2 whose relative order is preserved in the offspring O. The insertion cut-and-fill crossover is one-point geometric because it is equivalent to a sorting crossover based on the insertion move: it is like sorting parent P2 into parent P1using the insertion sort algorithm and stopping it when all the elements before the crossover point are sorted.

<sup>&</sup>lt;sup>1</sup> Proving that a recombination operator is not a one-point geometric crossover requires showing it for any choice of the underlying distance, and not only for a specific distance. See Moraglio's PhD thesis [5] for how to prove this type of general negative results.

P1:	a b c d e f	P1: a b c d e f
P2:	 cdbfea	P2: c d b f e a
0:	abcdfe	adbfec
		abdfec
		0: abcfed

 $Fig. \ 2. \ Insertion \ cut-and-fill \ crossover \ (left) \ and \ swap \ cut-and-fill \ crossover \ (right).$ 

The connection between cut-and-fill crossover and sorting crossover suggests that more types of cut-and-fill crossovers can be defined depending on the type of move used as base of the sorting. So we can define a swap cut-and-fill crossover that sorts parent P2 into parent P1 using selection sort (swap-based minimal sort algorithm) and stopping it when the elements before the crossover point are sorted. Since swap cut-and-fill crossover is based on a deterministic sorting algorithm, it is a one-point geometric crossover. Figure 4 (right) gives an example of swap cut-and-fill crossover. Using the same parents and the same crossover point as for the insertion cut-and-fill crossover we obtain a different offspring (the dashes indicate the elements of parent P2 that have been swapped to match the order of parent P1). If, as a base of the cut-and-fill crossover, we use the adjacent swap move that is associated with the bubble sort algorithm, we obtain again the insertion cut-and-fill crossover. This happens because, apart from the number of moves required (an insertion is equivalent to a sequence of adjacent swaps), after ordering the elements of parent P2 into the order of parent P1 up to the crossover point the order of the remaining elements is the same as when using insertions or adjacent swaps.

#### 5 Genetic Programming Trees

For GP trees, there are two recombination operators that can be thought as extensions of one-point crossover for binary strings to GP trees: Koza's subtree swap crossover [3] and one-point homologous crossover [4]. Koza's subtree swap crossover is not a geometric crossover [5]. So it is not a one-point geometric crossover either because this is a subclass of geometric crossover. Homologous one-point crossover aligns parent trees at the root and then cut-and-swap subtrees at the same position in the two parents. The family of homologous crossovers for GP trees [4] are geometric crossover sunder Structural Hamming distance [5], so also one-point homologous crossover is. However, one-point homologous crossover is not a one-point geometric crossover [5]. After these two negative results, one may wonder how one-point geometric crossovers for GP trees look like. In the following we first present a theorem that helps detecting whether a crossover operator is a one-point geometric crossover. Then we consider a family of crossovers for GP trees that are a subclass of mask-based homologous crossover. Since all mask-based homologous crossovers for GP trees are geometric also this family of crossovers is geometric. Then we show that all these crossovers are one-point geometric.

**Theorem 2.** The points  $o_1, o_2, \ldots, o_n \in [a, b]_d$  belong to a single geodesic g linking a and b in the metric space d iff they can be ordered as  $\bar{o}_1, \bar{o}_2, \ldots, \bar{o}_n$  such that  $d(a, \bar{o}_1) + d(\bar{o}_1, \bar{o}_2) + \ldots + d(\bar{o}_{n-1}, \bar{o}_n) + d(\bar{o}_n, b) = d(a, b)$ .

*Proof.* If such an order exists the length of the path g connecting a and b passing thought  $o_1, o_2, \ldots, o_n$  is d(a, b). So g is a shortest path linking a and b. If such an order does not exist, for each order of  $\bar{o}_1, \bar{o}_2, \ldots, \bar{o}_n$  we have  $d(a, \bar{o}_1) + d(\bar{o}_1, \bar{o}_2) + \ldots + d(\bar{o}_{n-1}, \bar{o}_n) + d(\bar{o}_n, b) > d(a, b)$  for the triangular inequality. Then the path g connecting a and b passing thought  $o_1, o_2, \ldots, o_n$  is larger than d(a, b). So g is not a shortest path linking a and b. Since  $o_1, o_2, \ldots, o_n \in [a, b]_d$ , for each point there exists a geodesic linking a and b passing for that point. Therefore since there is no geodesic between a and b passing through all points  $o_1, o_2, \ldots, o_n$  they must belong to distinct geodesics.

**Definition 2.** (Ordered homologous crossover family) Let us define a total order on the nodes of the common region of two parent trees. The order is a (deterministic) function of the two parents. The offspring that the two parents can produce are those obtained, exchanging in the parents the node at position 1 in the order, plus those obtained by exchanging simultaneously the nodes at positions 1 and 2, plus those obtained by exchanging simultaneously nodes at positions 1, 2 and 3 and so on.

For example, we can define a total order on the common region by numbering its nodes starting from the root and then visiting and numbering successive nodes of the common region using a breath-first strategy. Then generating uniformly a random number k between 1 and the number of nodes in the common region and exchanging in the two parents all nodes up to k. We term this crossover breathfirst homologous geometric crossover. One could change the numbering strategy with a depth-first or bottom-up or any other strategy that visits all nodes of a tree in a systematic and deterministic way and obtain new geometric crossover belonging to the ordered homologous crossover family. Figure 5 illustrates the breath-first ordered homologous crossover. The crossover mask tells for each position of the common region from which of the parents to take the node or subtree to pass to the offspring. The crossover mask on the bottom left is valid because the nodes from 1 to 5 marked with 'X' will be passed to the offspring from parent 1; the node from 6 to 10 marked with 'Y' will be passed to the offspring from parent 2. The crossover mask on the right is invalid for the breath-first ordered homologous crossover because the numbering of the nodes passed to the offspring from parent 1 (marked with 'X') is not an uninterrupted sequence (since X-marked nodes are: 1, 2, 4, 7 and 8).

#### **Theorem 3.** Ordered homologous crossovers are one-point geometric crossovers.

*Proof.* We prove it by showing that the offspring  $o_1, o_2, \ldots, o_n$  generated respectively by exchanging nodes at position 1, and at positions 1 and 2, and at positions 1, 2 and 3 and so on respect the condition of theorem 2 to be on a geodesic. It is easy to see that the sequence of offspring is a cumulative sequence of independent syntactic differences. Since the metric SHD is a weighted Hamming distance it is an additive distance on the contribution of independent syntactic differences. So we have  $d(a, b) = d(a, o_1) + d(o_1, b)$  and  $d(a, b) = d(a, o_1) + d(o_1, o_2) + d(o_2, b)$  and  $d(a, b) = d(a, o_1) + d(o_1, o_2) + d(o_2, 0) + d(o_3, b)$  and so on. So we have  $d(a, o_1) + d(o_1, o_2) + \dots + d(o_{n-1}, o_n) + d(o_n, b) = d(a, b)$ .



**Fig. 3.** Breath-first ordered homologous crossover for GP trees: (top) two parent trees P1 and P2; (center left) their associated hyperschema H(P1,P2) with nodes numbered in breath-first order; (center right) all the potential offspring applying homologous crossover to parents P1 and P2 (the part in bold means alternative content of the tree; in this case there are 5 independent binary alternatives, resulting in 32 possible offspring); (bottom left) a valid crossover mask for the breath-first ordered homologous crossover, and (bottom right) an invalid one.

The application of one-point geometric crossover to GP trees is instructive because it shows that operators that would have been intuitively understood as reasonable extensions of one-point geometric crossover for binary strings to GP trees, in fact, are not one-point geometric crossovers. This point deserves some attention. In intuitive terms, one-point geometric crossover generalizes the aspect of traditional one-point crossover for binary strings that the offspring must form a chain of gradual syntactic changes leading from one parent to the other parent. In particular, the aspect of the traditional one-point crossover that adjacent syntactic elements (adjacent loci in the string) are more likely to be passed together to the offspring is not captured by the generalization. This is essentially because this property is very specific of the binary string representation and cannot be defined in general geometric terms. One might enforce the adjacency property choosing a specific one-point geometric crossover out of the many possible for the representation at hand, if some notion of adjacency can be defined for the specific representation. For example, the breath-first order crossover introduced above has a property of syntactic adjacency as it can be understood as slicing the tree incrementally starting from the root (see also figure 5). Although this property makes this operator more in the spirit of one-point for binary strings, it is not more one-point geometric than any other operator belonging to the ordered homologous crossover family that does not have such a syntactic adjacency property.

#### 6 Variable-Length Sequences

In previous work [5], we have introduced a class of alignment-based homologous operators for variable-length sequences in which parent sequences are aligned optimally on their contents before exchanging genetic material using a crossover mask on the alignment. This is a closer model of biological recombination at molecular level than traditional crossovers for binary strings. Then, we proved that this class of operators are geometric crossovers under edit distance for sequences. One-point homologous crossover for sequences is an operator belonging to the class of alignment-based homologous operators in which the crossover masks on the alignment are the traditional one-point mask of the one-point crossover for binary strings. The following result shows that one-point homologous crossover for sequences is a one-point geometric crossover.

**Theorem 4.** One-point alignment-based homologous crossover is one-point geometric crossover under edit distance.

*Proof.* An optimal edit transcript *T* contains a smallest set *E* of edit moves to transform parent *u* in parent *v*. A mask *m* selects a subset of edit moves  $E_m \subseteq E$  from the transcript *T* to apply to *u* and produces the offspring *z*. *z* is on a shortest path for the geometricity of homologous crossover. Any homologous crossover operator for which all offspring are generated employing a set of masks  $m_i$  that forms a total order (when understood as vectors) generates offspring on a single shortest path between parents. This is because: (i) the sets of edit moves  $E_{m_i}$  corresponding to the masks  $m_i$  can be totally ordered under inclusion; (ii) the contribution of each edit move to the distance between parents is independent and additive; (iii) hence, when considered in this order,  $E_{m_i}$  generate a sequence of offspring  $z_i$  on a single shortest path between parents that incrementally leads from *u* to *v*. One-point alignment-based homologous crossover uses crossover masks that form a total order: (00...0) < (10...0) < (11...0) < ... < (11...1). So all its offspring are on a single shortest path between parents. Hence it is one-point geometric.

#### 7 Sets

In previous work [5], we have seen that there is a duality between the geometric crossover under Hamming distance for binary strings and the geometric crossover under ins/del edit distance for sets. In fact, these two crossovers, although being based on two different solution representations, are associated to isomorphic metric spaces (via the set indicator function), hence they are completely equivalent. Using the duality, in the following we will show the equivalent for sets of the one-point crossover for binary strings. Let us consider two binary strings of length 5,  $p_1 = 11000$  and  $p_2 = 01110$ . Let  $U = \{a, b, c, d, e\}$  be the universal set. The corresponding sets of  $p_1$  and  $p_2$  are  $s_1 = \{a, b\}$  and  $s_2 = \{b, c, d\}$ . After removing repetitions, the geodesic path of the traditional one-point crossover (i.e.,

with bit ordering 12345) applied to parents  $p_1$  and  $p_2$  is 11000, 01000, 01100, 01110. The corresponding sequence of offspring sets is  $\{a, b\}, \{b\}, \{b, c\}, \{b, c, d\}$ . Notice that this sequence gradually transforms  $s_1$  into  $s_2$  a move at a time using the ins/del edit move. This is the interpretation of one-point crossover for sets.

#### 8 Conclusions

One-point geometric crossover clarifies and makes rigorous the notion of onepoint crossover across representations and formalizes the intuition behind it. It can be used to generate new one-point crossovers for new representations in a formal way without involving ad-hoc adaptations of the original concept. We have derived specific one-point crossovers for a number of well-known representations. Some of the derived operators correspond to pre-existing operators, others are new operators. Few pre-existing operators, which were conceived as analogues of the traditional one-point crossover for other representations, are not one-point geometric crossovers. In future work, we will derive properties common to all one-point crossovers, test the new operators experimentally and extend the geometric framework with the generalization of multi-point crossover.

#### References

- 1. A. Eiben and J. Smith. *Introduction to Evolutionary Computing*. Springer, Natural Computing Series, 2003.
- J. H. Holland. Adaptation in Natural and Artificial Systems. University of Michigan Press, 1975.
- J. R. Koza. Genetic Programming: On the Programming of Computers by Means of Natural Selection. The MIT Press, 1992.
- 4. W. Langdon and R. Poli. *Foundations of Genetic Programming*. Springer-Verlag, 2002.
- 5. A. Moraglio. Towards a geometric unification of evolutionary algorithms. PhD thesis, University of Essex, 2007.
- A. Moraglio and R. Poli. Topological interpretation of crossover. In Proceedings of the Genetic and Evolutionary Computation Conference, pages 1377–1388, 2004.
- 7. C. R. Reeves and T. Yamada. Genetic algorithms, path relinking, and the flowshop sequencing problem. *Evolutionary Computation*, 6(1):45–60, 1998.
- 8. G. Sywerda. Uniform crossover in genetic algorithms. In *Proceedings of the third* international conference on Genetic algorithms, 1989.
- D. Whitley. A genetic algorithm tutorial. Journal of Statistics and Computing, 4:65–85, 1994.