Inbreeding Properties of Geometric Crossover and Non-geometric Recombinations

Alberto Moraglio and **Riccardo Poli** [1]

Abstract. Geometric crossover is a representation-independent generalization of traditional crossover for binary strings. It is defined using the distance associated to the search space in a simple geometric way. Many interesting recombination operators for the most frequently used representations are geometric crossovers under some suitable distance. Being a geometric crossover is useful because there is a growing number of theoretical results that apply to this class of operators. To show that a given recombination operator is a geometric crossover, it is sufficient to find a distance for which offspring are in the metric segment between parents associated with this distance. However, proving that a recombination operator is not a geometric crossover requires to prove that such an operator is not a geometric crossover *under any distance*. In this paper we develop some theoretical tools to prove non-geometricity results and show that some well-known operators are not geometric.

1 Introduction

Geometric crossover and geometric mutation [2] are representation-independent search operators that generalise by abstraction many pre-existing search operators for the major representations used in EAs, such as binary strings, real vectors, permutations and syntactic trees. They are defined in geometric terms using the notions of line segment and ball. These notions and the corresponding genetic operators are well-defined once a notion of distance in the search space is defined. This way of defining search operators as function of the search space is the opposite to the standard approach [3] in which the search space is seen as a function of the search operators employed. Our new point of view greatly simplifies the relationship between search operators and fitness landscape and allows different search operators to share the same search space thereby clarifying their roles.

The paper is organized as follows. In section 2, we introduce the geometric framework and review a number of well-known recombination operators that are geometric. In section 3, we show how the geometric definition of crossover divides recombination operators into two classes: geometric crossovers and non-geometric crossovers. This classification is based only on the metric properties of the operators and is representation-independent. In section 4, we develop some general tools to prove non-geometricity of recombination operators. In section 5, we prove that a number of recombination operators for vectors of reals, permutations and syntactic trees representations are not geometric. In section 6, we claim that the class of geometric crossovers captures the essential notion of crossoverness emerged experimentally over the years and that crossover operators that are not geometric are not fully matured. Indeed, some of these operators are effectively macro mutations while others use ad hoc strategies to compensate for the bias of specific spaces. In section 7, we draw some conclusions.

[1] University of Essex, UK, email: {amoragn, rpoli}@essex.ac.uk

2 Geometric framework

2.1 Geometric preliminaries

In the following we give necessary preliminary geometric definitions and extend those introduced in [2]. For more details on these definitions see [4].

The terms *distance* and *metric* denote any real valued function that conforms to the axioms of identity, symmetry and triangular inequality. A simple connected graph is naturally associated to a metric space via its *path metric*: the distance between two nodes in the graph is the length of a shortest path between the nodes. Distances arising from graphs via their path metric are called *graphic distances*. Similarly, an edge-weighted graph with strictly positive weights is naturally associated to a metric space via a *weighted path metric*.

In a metric space (S, d) a *closed ball* is a set of the form $B(x; r) = \{y \in S | d(x, y) \leq r\}$ where $x \in S$ and r is a positive real number called the radius of the ball. A *line segment* (or closed interval) is a set of the form $[x; y] = \{z \in S | d(x, z) + d(z, y) = d(x, y)\}$ where $x, y \in S$ are called extremes of the segment. Metric ball and metric segment generalise the familiar notions of ball and segment in the Euclidean space to any metric space through distance redefinition. These generalised objects look quite different under different metrics. Notice that the notions of metric segment and shortest path connecting its extremes (*geodesic*) do not coincide as it happens in the specific case of an Euclidean space. In general, there may be more than one geodesic connecting two extremes; the metric segment is the union of all geodesics.

We assign a structure to the solution set by endowing it with a notion of distance d. $M = (S, d)$ is therefore a solution *space* and $L = (M, g)$ is the corresponding *fitness landscape*. Notice that d is arbitrary and need not have any particular connection or affinity with the search problem at hand.

2.2 Geometric crossover definition

The following definitions are *representation-independent* and, therefore, crossover is well-defined for any representation. Being based on the notion of metric segment, *crossover is only function of the metric d* associated with the search space.

Definition: The *image set* $Im[OP]$ of a genetic operator OP is the set of all possible offspring produced by OP with non-zero probability.

Definition: A binary operator is a geometric crossover under the metric d if all offspring are in the segment between its parents.

Definition: Uniform geometric crossover UX is a geometric crossover where all z laying between parents x and y have the same probability of being the offspring:

$$f_{UX}(z|x,y) = \frac{\delta(z \in [x;y])}{|[x;y]|}$$

$$Im[UX(x,y)] = \{z \in S | f_{UX}(z|x,y) > 0\} = [x;y].$$

A number of general properties for geometric crossover and mutation have been derived in [2].

2.3 Notable geometric crossovers

For vectors of reals, various types of blend or line crossovers, box recombinations, and discrete recombinations are geometric crossovers [2]. For binary and multary strings (fixed-length strings based on a n symbols alphabet), all mask-based crossovers (one point, two points, n-points, uniform) are geometric crossovers [2] [6]. For permutations, PMX, Cycle crossover, merge crossover and others are geometric crossovers [1] [8]. For Syntactic trees, the family of Homologous crossovers (one-point, uniform crossover) are geometric crossovers [9]. Recombinations for other more complicated representations such as variable length sequences, graphs, permutations with repetitions, circular permutations, sets, multisets partitions are geometric crossovers [10] [2] [1] [7].

3 Geometric classes of recombination operators

The definition of geometric crossover requires a distance d that respects the metric axioms. This definition allows to answer easily the question: what is the geometric crossover associated with a given distance? Indeed, it is simply matter of applying the geometric definition to the specific distance to obtain its associated geometric crossover (or, more precisely, a formal description of it).

However, other questions can be asked: given a recombination operator what is the distance under which the operator fits the definition of geometric crossover? If there is more than one distance, what is the distance that fits it best and in what sense? These are very interesting questions. In previous work we have proven that various crossover operators are geometric crossovers by using our intuition and experience to identify candidate distances that seemed appropriate, by analysing the geometric crossovers resulting from such distances, and finally by identifying the one which matched our original operator. For some nice class of distances, namely graphic distances, we have some theoretical results that help us match recombination and distance [2]. In general, however, finding a distance fitting a specific operator may be rather difficult (see [9] for an example of a geometric crossover associated to a non-graphic distance and related issues). We have done some preliminary work on finding the best fitting distance for a given geometric crossover, which will appear in a future publication.

There is a more fundamental question that can be asked: is there a distance for any recombination operator? Or more precisely: given a recombination operator, is it possible that it does not fit the definition of geometric crossover for *any distance*? As we will see, some recombination operators are not geometric crossover, or simply non-geometric, under any distance. Therefore, the definition of geometric crossover induces a natural classification of all recombination operators into two non-empty and mutually exhaustive classes of operators:

- *geometric crossover class*: for any recombination belonging to this class, there exists at least a distance conforming to the metric axioms under which such a recombination is geometric. Formally: OP is geometric crossover iff $\exists d : \forall p_1, p_2 \in S : Im[OP(p_1,p_2)] \subseteq [p_1,p_2]_d$.
- *non-geometric crossover class*: for any recombination belonging to this class, there is no distance conforming to the metric axioms under which such a recombination is geometric. Formally: OP is non-geometric crossover iff $\forall d : \exists p_1, p_2 \in S : Im[OP(p_1,p_2)] \setminus [p_1,p_2]_d \neq \emptyset$.

We want to emphasize that the adjective "geometric" before crossover denotes, not simply a different way of looking at recombination operators, but more fundamentally it denotes the membership to a certain mathematically well-defined class of operators. In section 4, we will prove some properties common to all geometric crossovers arising from the definition of geometric crossover class only.

3.1 Special classes of geometric crossover

The property of a recombination operator of being geometric depends on the *existence* of a metric under which such a recombination is geometric. If such a metric exists, the operator is geometric; if such a metric does not exist, the operator is non-geometric. Analogously, we can define some interesting special classes of geometric crossover using the existence of specific types of metric as their defining property as follows.

Definition: A geometric crossover is *graphic*, if there exists a graphic distance for which it is geometric. Otherwise, the crossover will be said *non-graphic*.

Notice that the fact that a geometric crossover is geometric under a non-graphic distance does not imply that it is non-graphic. Indeed, it may also exist a graphic distance for which this operator is geometric, making it graphic.

Definition: A geometric crossover is *complete* if there exists a metric d such that the set of all the offspring coincides with the segment between two parents. A geometric crossover is *incomplete* if such a metric does not exist.

If a geometric crossover is geometric under a metric for which the set of all offspring does not coincides with the segment between parents (incomplete under such a metric), it does not imply that the crossover is incomplete. This is because there may exist another metric that fits the crossover under which the geometric crossover is complete, making it complete.

Definition: A geometric crossover is *uniform* if it is complete and the probability of picking any offspring is uniform. If the crossover is complete but the probability is not uniform then it is *non-uniform*.

4 Inbreeding properties of geometric crossover

Geometric interval spaces based on metric spaces [11] connect very naturally with the notion of geometric crossover. There is a wealth of results for these spaces that can be transferred easily to geometric crossover.

In the following we introduce three fundamental properties of geometric crossover arising only from its axiomatic definition (metric axioms), hence valid for any distance, probability distribution and any underlying solution representation. These properties of geometric crossover are simple properties of geometric interval spaces adapted to the geometric crossover.

The properties proposed are based on inbreeding (breeding between close relatives) using geometric crossover and avoid explicit reference to the solution representation.

In section 5, we will make good use of these properties to prove some non-geometricity results.

Theorem 1 *Property of Purity*
If RX is geometric then the recombination of one parent with itself can only produce the parent itself.

proof: If RX is geometric there exists a metric d such that any offspring o belongs to the segment between parents s_1, s_2 under metric d: $d(s_1, o) + d(o, s_2) = d(s_1, s_2)$. When the parents coincide, $s = s_1 = s_2$, we have: $d(s, o) + d(o, s) = d(s, s)$ hence for symmetry and identity axioms of metric $d(s, o) = 0$ for any metric. For the identity axiom this implies $o = s$.

Inbreeding diagram of the property of purity (see Fig. 1(a)): when the two parents are the same $P1$, their child C must be $P1$.

```
(a) Purity:

P1 --- P1
    |
    C=P1


(b) Convergence:

P1 -------- P2
      |
      C --- P2 (C!=P1)
          |
          G!=P1


(c) Partition:

P1 --------- P2
       |
P1 --- C --- P2
    |      |
    G1 != G2 (G1!=C or G2!=C)
```

Figure 1. Inbreeding diagrams.

Theorem 2 *Property of Convergence*
If RX is geometric then the recombination of one parent with one offspring cannot produce the other parent of that offspring unless the offspring and the second parent coincide.

proof: If RX is geometric there exists a metric d such that any offspring o belongs to the segment between parents s_1, s_2 under metric d: $d(s_1, o) + d(o, s_2) = d(s_1, s_2)$. If one can produce parent s_2 by recombining s_1 and o, it must be also true that: $d(s_1, s_2) + d(s_2, o) = d(s_1, o)$. By substituting this last expression in the former one we have: $d(s_1, s_2) + d(s_2, o) + d(o, s_2) = d(s_1, s_2)$ hence simplifying $d(o, s_2) = 0$ for any metric. For the identity axiom this implies $s_2 = o$.

Inbreeding diagram of the property of convergence (see Fig. 1(b)): two parents $P1$ and $P2$ produce the child C. We consider a C that does not coincide with $P1$. The child C and its parent $P2$ mate and produce a grandchild G. The property of convergence states that G can never coincide with $P1$.

Theorem 3 *Property of Partition*
*If RX is geometric then the two recombinations, the first of parent **a** with a child **c** of **a** and **b**, and the second of parent **b** with the same child **c**, cannot produce a common grandchild **e** other than **c**.*

proof:
$c \in [a, b] \rightarrow d(a, c) + d(c, b) = d(a, b)$
$e \in [a, c] \rightarrow d(a, e) + d(e, c) = d(a, c)$
$e \in [b, c] \rightarrow d(b, e) + d(e, c) = d(b, c)$
Substituting the last two expressions in the first one we obtain:
$d(a, e) + d(e, c) + d(b, e) + d(e, c) = d(a, b)$
Notice that two terms in the left-hand are greater or equal to the right-hand:
$d(a, e) + d(b, e) \geq d(a, b)$
Hence:
$2d(e, c) = 0$ implies $e = c$.

Inbreeding diagram of the property of partition (see Fig. 1(c)): two parents $P1$ and $P2$ produce the child C. The child C mates with both its parents, $P1$ and $P2$, producing grandchildren $G1$ and $G2$, respectively. We consider the case in which at least one grandchildren is different from C. The property of partition states that $G1$ and $G2$ can never coincide.

Complete crossover has a larger set of properties including extensiveness ($a, b \in Im(UX(a, b))$) and symmetry ($Im(UX(a, b)) = Im(UX(b, a))$), which however, are not common to all geometric crossovers.

4.1 Relation with forma analysis

Radcliffe developed a theory [14] of recombination operators starting from the notion of forma that is a representation-independent generalization of schema. A forma is an equivalence class on the space of chromosomes induced by a certain equivalence relation. Radcliffe describes a number of important formal *desirable properties* that a recombination operator should respect to be a good recombination operator. These properties are representation-independent and are stated as requirements on how formae should be manipulated by recombination operators.

Geometric crossover, on the other hand, is formally defined geometrically using the distance associated with the search space. Unlike Radcliffe's properties, the inbreeding properties of geometric crossover, are not desired properties but are properties that are *common* to all geometric crossovers and derive logically from its formal definition only. In future work, we will provide a more detailed analysis of the connections between Radcliffe's theory and the theory of geometric crossover.

5 Notable non-geometric crossovers

To prove non-geometricity, one has to prove that the recombination operator is *non-geometric under any metric* and *any probability distribution* over *any subset* of the segment. This could be difficult. The properties of purity, convergence and partition apply to any geometric crossover. If a recombination fails to fulfil any of these property then it is not a geometric crossover (under any distance).

These properties, therefore, are necessary conditions for geometricity and therefore can be used to check if a recombination operator is non-geometric. In the following we use them to prove the non-geometricity of three important recombination operators: extended line recombination, Koza's subtree swap crossover and Davis's Order Crossover (see [12] for a description of these operators).

Theorem 4 *Extended line recombination is not a geometric crossover.*

proof: The convergence property fails to hold. Let p_1 and p_2 the two parents, and o the offspring lying in the extension line beyond p_1. Now it is easy to see that using the extension line recombination on o and p_2, one can obtain p_1 as offspring.

Theorem 5 *Koza's subtree swap crossover is not a geometric crossover.*

proof: The property of purity fails to hold. Subtree swap crossover applied to two copies of the same parent tree may produce offspring trees different from it.

Theorem 6 *Davis's Order Crossover is non-geometric.*

proof: We prove the theorem by showing that the *convergence property of geometric crossover does not hold* in the following counterexample.

```
Parent 1 : 12.34.567
Parent 2 : 34.56.127
Section  : --.34.---
Available elements in order: 12756

Offspring: 65.34.127
Parent 3 := Offspring

Parent 3 : 6534.12.7
Parent 1 : 1234.56.7
Section  : ----.12.-
Available elements in order: 73456

Offspring: 3456.12.7
Offspring = Parent 2
```

Since the last offspring coincides with parent 2 the monotone property does not hold, hence Order Crossover is not geometric.

6 Are "real-world" recombinations geometric?

We have seen in section 3 that the geometric definition of crossover induces a well-defined bipartition of all recombination operators: geometric crossovers and non-geometric crossovers. Starting from the definition of geometric crossover, in previous work we have started building a general, representation-independent theory that applies to all geometric crossovers. This theory, for its own axiomatic foundation, applies to geometric crossovers and does not apply to non-geometric crossovers. The significance of such a theory is, therefore, conditional to the fact that interesting recombination operators are geometric crossovers. So, a very important question is: are "real-world" recombinations, those used everyday in practice, geometric? To answer this question we have started a unification programme to show that most of the pre-existing crossover operators for major representations fit the geometric definition. The three operators considered in section 5, that have been shown to be non-geometric, are indeed very rare exceptions. So the theory of geometric crossover has a considerable scope and a real applicability.

More fundamentally we can put forward the hypothesis that since established pre-existing operators have emerged from experimental work done by generations of practitioners over decades, geometric crossover compresses in a simple class an empirical phenomenon. Or in other words, the geometric crossover definition captures a law of nature. In this perspective, it is reasonable to ask why the recombination operators presented in section 5 are non-geometric. Why don't they conform to the geometricity-law?

Extended line crossover: line crossover (that is geometric) is biased toward the center of the space. The extended line are there to compensate for such a bias.

Subtree swap crossover: Koza's crossover is strongly suspected to be equivalent to subtree mutation [13]. Many researchers do not see it as a crossover and propose new form of operator that require alignments on contents or positional alignment before recombinations. Interestingly, these two variations would transform Koza's crossover into a geometric crossover.

Order crossover: Order crossover was a first attempt to recombine permutations preserving common order of the parents. However common order is not preserved all the times in this operator. Interestingly, all operator that preserve perfectly common order are provably geometric (such as merge recombination[12], for example).

7 Conclusions

In this paper we have shown that the abstract definition of geometric crossover induces two representation-independent classes of recombination operators: geometric crossovers and non-geometric crossovers. Proving non-geometricity of a recombination operator it is a non-trivial task because one needs to show that the recombination considered is not geometric under *any distance*. We have developed some theoretical tools to prove non-geometricity. We have then used these tools to prove the non-geometricity of three well-known operators for real vectors, permutations, and syntactic trees representations. We have argued that geometric crossover subsumes the notion of crossoverness emerged experimentally over the years, and that the non-geometric operators considered in this paper are accidentally non-geometric.

REFERENCES

[1] A. Moraglio, R. Poli. Topological crossover for the permutation representation. In *Workshop on theory of representations at GECCO 2005*.

[2] A. Moraglio, R. Poli. Topological interpretation of crossover. In *Proceedings of GECCO 2004*, pages 1377–1388.

[3] T. Jones. Evolutionary Algorithms, Fitness Landscapes and Search. *PhD dissertation, University of New Mexico, 1995*.

[4] M. Deza, M. Laurent. Geometry of cuts and metrics. *Springer, 1991*.

[5] P.M. Pardalos, M.G.C. Resende. Handbook of Applied Optimization. *Oxford University Press, 2002*.

[6] A. Moraglio, R. Poli. Product Geometric Crossover. *Proceedings of PPSN 2006 (to appear)*.

[7] A. Moraglio, R. Poli. Geometric Crossover for Sets, Multisets and Partitions. *Proceedings of PPSN 2006 (to appear)*.

[8] A. Moraglio, R. Poli. Geometric Crossover for the Permutation Representation. *Technical Report CSM-429, 2005*.

[9] A. Moraglio, R. Poli. Geometric Landscape of Homologous Crossover for Syntactic Trees. *Proceedings of CEC 2005*, pages 427–434.

[10] A. Moraglio, R. Poli., R. Seehuus. Geometric Crossover for Biological Sequences. *Proceedings of EuroGP 2006*, pages 121–132.

[11] M. L. J. Van de Vel. Theory of Convex Structures. *North-Holland, 1993* .

[12] T. Bäck and D. B. Fogel and T. Michalewicz. Evolutionary Computation 1: Basic Algorithms and Operators. *Institute of Physics Publishing, 2000*.

[13] S. Luke, L. Spector. A Revised Comparison of Crossover and Mutation in Genetic Programming. *Proceedings of Genetic Programming Conference, 1998*.

[14] N. Radcliffe. Equivalence Class Analysis of Genetic Algorithms. *Journal of Complex Systems, 1991*, vol. 5, pages 183–205.