Inbreeding Properties of Geometric Crossover and Non-geometric Recombinations

Alberto Moraglio and Riccardo Poli

Dept. of Computer Science
University of Essex, UK
{amoragn, rpoli}@essex.ac.uk

Abstract. Geometric crossover is a representation-independent generalization of traditional crossover for binary strings. It is defined in a simple geometric way by using the distance associated with the search space. Many interesting recombination operators for the most frequently used representations are geometric crossovers under some suitable distance. 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*. This casts serious doubts on the possibility to draw a clear-cut line between geometric crossovers and non-geometric crossovers. In this paper we develop some theoretical tools to resolve this issue and prove that some well-known operators are not geometric. Finally, we discuss the important implications of this.

1 Introduction

A fitness landscape [22] can be visualised as the plot of a function resembling a geographic landscape when the problem representation is a real vector. When dealing with binary strings and other more complicated combinatorial objects, e.g., permutations, the fitness landscape is better represented as a height function over the nodes of a simple graph [19], where nodes represent locations (solutions), and edges represent the relation of direct neighbourhood between solutions.

An abstraction of the notion of landscape encompassing all the previous cases is possible. The solution space is seen as a metric space and the landscape as a height function over the metric space [1]. A metric space is a set endowed with a notion of distance between elements fulfilling few axioms [3]. Specific spaces have specific distances that fulfil the metric axioms. The ordinary notion of distance associated with real vectors is the Euclidean distance, though there are other options, e.g. Minkowski distances. The distance associated to combinatorial objects is normally the length of the shortest path between two nodes in the associated neighbourhood graph [4]. For binary strings, this corresponds to the Hamming distance.

In general, there may be more than one neighbourhood graph associated to the same representation, simply because there can be more than one meaningful notion of syntactic similarity applicable to that representation [10]. For example, in the case of permutations the adjacent element swap distance and the block reversal distance are equally natural notions of distance for permutations. Different notions of similarity are possible because the same permutation (genotype) can be used to represent different types of solutions (phenotypes). For example, permutations can represent solutions of a problem where relative order is important. However, they can also be used to represent tours, where the adjacency relationship among elements is what matters.

The notion of fitness landscape is useful if the search operators employed are connected or matched with the landscape: the greater the connection the more landscape properties mirror search properties. Therefore, the landscape can be seen as a function of the search operator employed [5]. Whereas mutation is intuitively associated with the neighbourhood structure of the search space, crossover stretches the notion of landscape further leading to search spaces defined over complicated topological structures [5].

Geometric crossover and geometric mutation [9] are representation-independent search operators that generalise by abstraction many pre-existing search operators for the main 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 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.

The reminder of this paper is organized as follows. In section 2, we introduce the geometric framework. In section 3, we show that the definition of geometric crossover can be cast in two equivalent but conceptually very different forms: functional and existential. When proving geometricity the existential form is the relevant one. We use this form also to show why proving non-geometricity of an operator looks impossible. In section 4, we develop some general tools to prove non-geometricity of recombination operators. In section 5, we prove that three recombination operators for vectors of reals, permutations and syntactic trees representations are not geometric. Importantly this implies that there are two *non-empty* representation-independent classes of recombination operators: geometric crossovers and non-geometric crossovers. In section 6, we discuss the consequence of this result. In section 7, we draw some conclusions and present future work.

2 Geometric framework

2.1 Geometric preliminaries

In the following we give necessary preliminary geometric definitions and extend those introduced in [9]. 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 $B_d(x; r) = \{y \in S | d(x, y) \leq r\}$ 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 $[x; y]_d = \{z \in S | d(x, z) + d(z, y) = d(x, y)\}$ where $x, y \in S$ are called extremes of the segment. Metric ball and metric segment generalize the familiar notions of ball and segment in the Euclidean space to any metric space through distance redefinition. These generalized 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 S by endowing it with a notion of distance d. $M = (S, d)$ is therefore a solution *space* (or search space) and $L = (M, g)$ is the corresponding *fitness landscape* where $g : S \to \mathbb{R}$ is the fitness function. Notice that in principle d could be 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.

A recombination operator OP takes parents p_1, p_2 and produces one offspring c according to a given conditional probability distribution:

$$Pr\{OP(p_1, p_2) = c\} = Pr\{OP = c | P_1 = p_1, P_2 = p_2\} = f_{OP}(c | p_1, p_2)$$

Definition 1. *(Image set) The* image set $Im[OP(p_1, p_2)]$ *of a genetic operator OP is the set of all possible offspring produced by OP with non-zero probability when parents are p_1 and p_2.*

Definition 2. *(Geometric crossover) A recombination operator CX is a geometric crossover under the metric d if all offspring are in the segment between its parents:* $\forall p_1, p_2 \in S : Im[CX(p_1, p_2)] \subseteq [p_1, p_2]_d$

Definition 3. *(Uniform geometric crossover) The uniform geometric crossover* UX *under* d *is a geometric crossover under* d *where all* z *laying between parents* x *and* y *have the same probability of being the offspring:*

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

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

where δ *is a function that returns 1 if the argument is true, 0 otherwise.*

A number of general properties for geometric crossover and mutation have been derived in [9].

2.3 Notable geometric crossovers

For vectors of reals, various types of blend or line crossovers, box recombinations, and discrete recombinations are geometric crossovers [9]. 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 [9, 13]. For permutations, PMX, Cycle crossover, merge crossover and others are geometric crossovers [10, 11]. For Syntactic trees, the family of Homologous crossovers (one-point, uniform crossover) are geometric crossovers [12]. Recombinations for other more complicated representations such as variable length sequences, graphs, permutations with repetitions, circular permutations, sets, multisets partitions are geometric crossovers [15, 9, 10, 14].

2.4 Geometric crossover landscape

Since our geometric operators are representation-independent, one might wander as to the usefulness of the notion of geometricity and geometric crossovers in practical applications. To see this, it is important to understand the difference between problem and landscape.

Geometric operators are defined as functions of the distance associated to the search space. However, the search space does not come with the problem itself. The problem consists only of a fitness function to optimize, that defines what a solution is and how to evaluate it, but it does not give any structure on the solution set. The act of putting a structure over the solution set is part of the search algorithm design and it is a designer's choice. A fitness landscape is the fitness function plus a structure over the solution space. So, for each problem, there is one fitness function but as many fitness landscapes as the number of possible different structures over the solution set. In principle, the designer could choose the structure to assign to the solution set completely independently from the problem at hand. However, because the search operators are defined over such a structure, doing so would make them decoupled from the problem, hence turning the search into something very close to random search.

In order to avoid this one can exploit problem knowledge in the search. This can be achieved by carefully designing the connectivity structure of the fitness landscape. That is, the landscape can be seen as a knowledge interface between algorithm and problem [10]. In [10] we discussed three heuristics to do so in such a way to aid the evolutionary search performed by geometric crossover: i) pick a crossover associated to a good mutation, ii) build a crossover using a neighbourhood structure based on the small-move/small-fitness-change principle, and iii) build a crossover using a distance that is relevant for the solution interpretation. Once this is done, problem knowledge can be exploited by search operators to perform better than random search, even if the search operators are problem-independent (as in the case of geometric crossover and mutation). Indeed, by using these heuristics, we have *designed* very effective geometric crossovers for N-queens problem [11], TSP [11] [10], Job Shop Scheduling [11], Protein Motifs discovery [20], Graph Partitioning [6], Sudoku [16] and Finite State Machines [7].

3 Interpretations of the definition of geometric crossover

In section 2, we have defined geometric crossover as function of the distance d of the search space. In this section we take a close look at the meaning of this definition *when the distance d is not known*. We identify three fundamentally different interpretations of understanding the definition of geometric crossover. Interestingly it will become evident that there is an inherent element of self-reference in the definition of geometric crossover. We show that proving that a recombination operator is non-geometric may be impossible.

3.1 Functional interpretation

Geometric crossover is function of a *generic distance*. If one considers a specific distance one can obtain a specific geometric crossover for that distance by functional application of the definition of geometric crossover to this distance. This approach is particularly useful when the specific distance is firmly rooted in a solution representation (e.g., edit distances) because the specification of the definition of geometric crossover to the distance acts as a formal recipe that indicates how to manipulate the syntax of the representation to produce offspring from parents. This is a general and powerful way to get new geometric crossover for any type of solution representation. For example, given the Hamming distance on binary string by functional application of the definition of geometric crossover we obtain the family of mask-based crossover for binary strings. In particular, by functional application of the definition of uniform geometric crossover one obtains the traditional uniform crossover for binary strings.

3.2 Abstract interpretation

The second use of the definition of geometric crossover does not require to specify any distance. In fact we do apply the definition of geometric crossover to a generic distance. Since the distance is a metric that is a mathematical object defined axiomatically, the definition of geometric crossover becomes an axiomatic object as well. This way of looking at the definition of geometric crossover is particularly useful when one is interested in deriving general theoretical results that hold for geometric crossover under any specific metric. We will use this abstract interpretation in section 4 to prove the inbreeding properties that are common to all geometric crossovers.

3.3 Existential interpretation

The third way of looking at the definition of geometric crossover becomes apparent when the distance d is not known and we want to find it. This happens when we want to know whether a recombination operator RX, defined operationally as some syntactic manipulation on a specific representation, is a geometric crossover and for what distance. This question hides an element of self-reference of the definition of geometric crossover. In fact what we are actually asking is: given that *the geometric crossover is defined over the metric space it induces by manipulating the candidate solutions*, what is such a metric space for RX if any?

The self-reference arises from the fact that the definition of geometric crossover applies at two distinct levels at the same time: (a) at a representation level, as a manipulation of candidate solutions, and (b) at a geometric level, on the underlying metric space based on a geometric relation between points. This highlights the inherent *duality* between these two worlds: they are based on the *same* search space seen from opposite viewpoints, from the representation side and from the metric side.

Self-referential statements can lead to paradoxes. Since the relation between geometric crossover and search space is what ultimately gives to it all its advantages, it is of fundamental importance to make sure that this relation sits on a firm ground. So, it is important to show that the definition of geometric crossover does not lead to any paradox. We show in the following that the element of self-reference can be removed and the definition of geometric crossover can be cast in existential terms making it paradox-free.

A non-functional definition of geometric crossover is the following: a recombination operator RX is a geometric crossover if the induced search space is a metric space on which RX can be defined as geometric crossover using the functional definition of geometric crossover. This is a self-referential definition. If a recombination operator does not induce any metric space on which it can be defined as geometric crossover, then it is a non-geometric crossover.

We can remove the element of self-reference from the previous definition and cast it in an existential form: a recombination RX is a geometric crossover if for any choice of the parents all the offspring are in the metric segment between them for some metric.

The existential definition is equivalent to the self-referential definition because if such a metric exists the operator RX can be defined as geometric crossover on such a space. On the other hand, if an operator is defined on a metric space as geometric crossover in a functional form, such a space exists by hypothesis and offspring are in the segment between parents under this metric by definition.

3.4 Geometric crossover classes

The functional definition of geometric crossover induces a natural existential classification of all recombination operators into two classes of operators:

- *geometric crossover class* $\mathcal{G}$: a recombination OP belongs to this class if there exists at least a distance d under which such a recombination is geometric: $OP \in \mathcal{G} \iff \exists d : \forall p_1, p_2 \in S : Im[OP(p_1, p_2)] \subseteq [p_1, p_2]_d$.
- *non-geometric crossover class* $\bar{\mathcal{G}}$: a recombination OP belongs to $\bar{\mathcal{G}}$ if there is no distance d under which such a recombination is geometric: $OP \in \bar{\mathcal{G}} \iff \forall d : \exists p_1, p_2 \in S : Im[OP(p_1, p_2)] \setminus [p_1, p_2]_d \neq \emptyset$.

For this classification to be meaningful we need these two classes to be non-empty. In previous work we proved that a number of recombination operators are geometric crossovers so $\mathcal{G}$ is not empty. What about $\bar{\mathcal{G}}$? To prove that this class is not empty we have to prove that at least one recombination operator is non-geometric. However, as we illustrate below this is not easy.

Let us first illustrate how one can prove that a recombination operator RX is in $\mathcal{G}$. We will use the self-referential definition of geometric crossover. The procedure is the following: guess a candidate distance d, then prove that all offspring of all possible pairs of parents are in the metric segment associated with d. If this is true then the recombination RX is geometric crossover under the distance d *because the operator RX can be defined as a geometric crossover on this space*. If the distribution of the offspring in the metric segments under d is uniform, RX is the uniform geometric crossover for the metric d *because the operator RX can be defined as the (unique) geometric uniform crossover on this space*. If one finds that some offspring are not in the metric segment between parents under the initially guessed distance d then the operator RX cannot be defined as geometric crossover over this space. However, this does not imply $RX \in \bar{\mathcal{G}}$ because there may exist another metric d' that fits RX and *makes it definable* as a geometric crossover on d'. So, one has to guess a new candidate distance for RX and start all over again until a suitable distance is found.

Although we developed some heuristics for the selection of a candidate distance, in general proving that a recombination operator is geometric may be quite hard (see for example [12] where we considered homologous crossover for GP trees). Nonetheless, the approach works and, in previous work, we proved that a number of recombination operators for the most frequently used representations are geometric crossover under suitable distances.

It is evident, however, that the procedure just described cannot be used to prove that a given recombination operator RX is non-geometric. This is because we would need to test and exclude all possible distances, which are infinitely many, before being certain that RX is not geometric. Clearly, this is not possible.

In the next section we build some theoretical tools based on the abstract interpretation of the definition of geometric crossover to prove non-geometricity in a more straightforward way.

4 Inbreeding properties of geometric crossover

How could we actually prove non-geometricity? From the definition of geometric crossover based on a generic notion of distance (abstract interpretation), we could derive metric properties that are common to the class of all geometric crossovers and that could be tested without making explicit use of the distance. Any reference to the distance needs necessarily to be excluded from these properties because what in fact we need to test is the existence of an underlying distance behind a given recombination operator hence we cannot assume the existence of one a priori. So the first requirement is that these properties derive from the metric axioms but cannot be about distance. A second requirement is generality: these properties need to be representation-independent so that recombination for any solution representation can be tested. A third and last requirement is that these properties need to be independent from the specific probability distribution with which offspring are drawn from the segment between the parents. In particular they must encompass also geometric crossovers where offspring are drawn from only part of the segment. If necessary properties satisfying these requirements existed, testing a recombination operator for non-geometricity would become straightforward: if such operator does not have a property common to all geometric crossovers it is automatically non-geometric. Fortunately, properties of this type do exists. They are the inbreeding properties of 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 [21] 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 the operator 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 $P1$.

Theorem 2. *(Property of Convergence) If the operator 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 for any offspring o of parents s_1 and s_2 we have $d(s_1, o) + d(o, s_2) = d(s_1, s_2)$. If one can produce parent s_2 by recombining s_1 and o, it must be also true that $d(s_1, o) = d(s_1, s_2) + d(s_2, o)$. By substituting this last expression in the former one we have: $d(s_1, s_2) + d(s_2, o) + d(o, s_2) = d(s_1, s_2)$, which implies $d(o, s_2) = 0$ and $s_2 = o$ for any metric. □

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 $P1$.

```
(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)
```

Fig. 1. Inbreeding diagrams.

Theorem 3. *(Property of Partition) If the operator RX is geometric and* $\boldsymbol{c}$ *is a child of* $\boldsymbol{a}$ *and* $\boldsymbol{b}$*, then the recombination of* $\boldsymbol{a}$ *with* $\boldsymbol{c}$ *and the recombination of* $\boldsymbol{b}$ *with* $\boldsymbol{c}$ *cannot produce a common grandchild* $\boldsymbol{e}$ *other than* $\boldsymbol{c}$*.*

Proof: We have that $c \in [a, b]$, $e \in [a, c]$ and $e \in [b, c]$, from which it follows that $d(a, c) + d(c, b) = d(a, b)$, $d(a, e) + d(e, c) = d(a, c)$ and $d(b, e) + d(e, c) = d(b, c)$.

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 $d(a,e) + d(b,e) \geq d(a,b)$ and, so, the previous equation implies $d(e,c) = 0$ and $e = c$. $\square$

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.

Geometric crossovers whose offspring cover completely the segments between their parents (complete geometric crossovers) have 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

Since the inbreeding properties of geometric crossover are related with forma analysis [18] we briefly explain this relation.

Radcliffe developed a theory [18] 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.

It is important to highlight that geometric crossover theory and forma analysis overlap but they are not isomorphic. This becomes clear when we consider what schemata for geometric crossover are. In forma theory, the recombination operators introduced by Radcliffe "respect" formae: offspring must belong to the same formae both parents belong to. A natural generalization of schemata for geometric crossover in this sense are (metric) convex sets: offspring in the line segment between parents belong to all convex sets common to their parents. So *geometric crossover induces a convexity structure over the search space.* A convexity structure is not the same thing as an equivalence relation: convex sets, like equivalence classes, cover the entire space but unlike them convex sets do not partition the search space because they overlap. Interestingly, convex sets seen as schemata naturally unify the notions of inheritance and fitness landscape.

A further advantage of geometric crossover over forma theory is that whereas it is rather easy to define and deal with distances for complex representations such as trees and graphs (using edit distances) it is much harder to use equivalence classes.

5 Non-geometric crossovers

In the followingnwe use the properties of purity, convergence and partition to prove the non-geometricity of three important recombination operators: extended line recombination, Koza's subtree swap crossover and Davis's Order Crossover (see, for example, [2] 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 be two parents, and o the offspring lying in the extension line beyond p_1. 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 may produce offspring trees different from it. □

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

Proof: The *convergence property does not hold* in the counterexample in Figure 2 where the last offspring coincides with parent 2. □

```
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
```

Fig. 2. Counterexample to the geometricity of order crossover.

What is knowing that these operators are *not* geometric good for? The first implication is that, when an operator is proven to be non-geometric, one is not tempted to try to prove its geometricity with yet another distance.

A second immediate and fundamental consequence of knowing that an operator is non-geometric is that since it is not associable with any metric it is not associable with any simple fitness landscape defined as a height function on a metric space in a simple way. This is bad news for non-geometric crossovers because the alternative to a simple fitness landscape with a simple geometric interpretation is a complex topological landscape with hardly any interpretation for what is really going on.

This leads us to a third very important practical consequence. Performance-wise, just knowing that a recombination operator is geometric or non-geometric cannot tell us anything about its performance. The no free lunch theorem rules. However, as a rule-of-thumb we know that when the fitness landscape associated with a geometric crossover is smooth, the geometric crossover associated with it is likely to perform well. This is fundamental for crossover design because the designer studying the objective function can identify a metric for the problem at hand that gives rise to a smooth fitness landscape and then he/she can pick the geometric crossover associated with this metric. This is a good way to embed problem knowledge in the search. However, since is inherently linked to the existence of a distance function associated with a recombination operator, non-geometric crossovers cannot make use of this strategy.

The forth and last consequence of the mere existence of some non-geometric operators is that this implies the existence of two separate classes of operators. We state this in the following as a theorem. This is an important step when developing a theory of geometric crossover because it allows to meaningfully talk about geometric crossover in general without the need to specify the distance associated with it. We discuss the implication of this in the next section.

Theorem 7. *(Existence of non-geometric crossover) The class of non-geometric crossover is not empty. Hence the space of recombination operator is split into two proper classes: geometric and non-geometric crossover.*

6 Discussion

In this section we discuss the implications of the theorem of the existence of non-geometric crossover.

6.1 Possibility of a general theory of evolutionary algorithms

Not being able to prove non-geometricity at least for a single recombination operator would leave us in a rather unpleasant situation because we would not be able to qualify the word "geometric" before "crossover": what does "geometric crossover" without specifying a distance mean? Is it a synonym of all recombination operators, hence an empty word, or it defines a proper subclass of recombination operators defined on their metric property? This is a fundamental question because either answers have a critical impact on the possibility of a general theory of geometric crossover and of a programme of unification of evolutionary algorithms. We illustrate this issue in the following.

Evolutionary computation theory is fragmented and one of the main reasons is that there is not a unified way to deal with different solution representations which has led to the development of significantly different theories for different flavors of evolutionary algorithms. Once a general mathematical framework encompassing all representations is available, it will be possible to accommodate, blend and generalize pre-existing isolated theories. Geometric crossover and geometric mutation show that this common mathematical framework is possible so opening the way to a really general representation-independent theory of evolutionary algorithms. But, would such a general theory be able to tell us anything meaningful or only trivialities encompassing all operators could be derived?

A theory of all operators is an empty theory because the performance of an EA derives from how its way of searching the search space is matched with some properties of the fitness landscape. Without restricting the class of operators to a proper subset of all possible operators, there is no common behavior, hence there is no common condition on the fitness landscape to be found to guarantee better than random search performance. This is just another way of stating the NFL theorem. So a theory of all operators is necessarily a theory of random search in disguise. Hence, if the the definition of geometric crossover encompasses all operators, it would be futile to pursue a general theory of geometric crossover.

In previous work we have found that many "real-world" recombinations, those used in everyday practice, turned out to be geometric. Without being able to prove the existence of some non-geometric crossovers there are two alternative explanations for this happening: (a) the geometric crossover definition is a tautology and the theory built on it a theory of everything hence an empty theory or (b) if there are non-geometric crossovers, this is hardly a coincidence and the class of geometric crossover indeed captured the essence of the class of "real-word" recombinations.

Theorem 7 is therefore foundational because it implies that the true explanation is (b). This has two consequences: first, a general theory of geometric crossover makes sense because it is not a theory of random search in disguise. Second, it begs for an explanation on the reason why the definition of geometric crossover captures the notion of "real-world" recombinations.

6.2 Why are "real-world" recombinations geometric?

Arguably, the most common search operators must be good *in practice* because they have emerged from experimental work done by generations of practitioners over decades and have survived a fierce competition against other operators. So, it is reasonable to ask what is common to most of these successful recombination operators, if anything, and infer that this commonality must be related with the reason of their success. So, what is the underlying regularity, or the "law of nature", linking them? They are geometric. This answer would be empty if geometricity were a tautology, if every recombination were geometric given some suitable distance. Since geometricity is a clear-cut property and not a tautology, the class of geometric crossover must have some *practical* fundamental advantage over the complementary class. It is reasonable to conjecture that this is linked with the fact that geometric crossover allows very easily to embed problem knowledge in the search from the knowledge of the objective function.

The three operators considered in section 5, that have been shown to be non-geometric, are indeed very rare exceptions of non-geometric "real-world" operators. Why do not 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 [8]. 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[2], for example).

7 Conclusions and future work

In this paper we have shown that the abstract definition of geometric crossover induces two *non-empty* representation-independent classes of recombination operators: geometric crossovers and non-geometric crossovers. This is a fundamental result that put a programme of unification of evolutionary algorithms and a general representation-independent theory of recombination operators on a firm ground.

Because of the peculiarity of the definition of geometric crossover, proving non-geometricity of a recombination operator, hence the existence of the non-geometric crossover class, is a task that at first looks impossible because one needs to show that the recombination considered is not geometric under *any distance*. However taking advantage of the different possible ways of looking at the definition of geometric crossover we have been able to develop some theoretical tools to prove non-geometricity in a straightforward way. We have then used these tools to prove the non-geometricity of three well-known operators for real vectors, permutations, and syntactic trees representations. Thereby proving the fundamental result.

In future work, we will start constructing a general theory of evolutionary algorithms based on the abstract interpretation of the definition of geometric crossover. So this theory will be able to describe the generic behavior of all evolutionary algorithms equipped with a generic geometric crossover. We anticipate that this is a form of convex search. The next step will be to understand under what exact condition on the fitness landscape, for what general class of fitness landscape, this way of searching delivers good performance.

References

1. T. Bäck, U. Hammel and H. Schwefel. Evolutionary computation: Comments on the history and current state. *IEEE Transactions on Evolutionary Computation*, 1(1):3–17, April 1997.
2. T. Bäck, D. B. Fogel and T. Michalewicz. Evolutionary Computation 1: Basic Algorithms and Operators. *Institute of Physics Publishing, 2000.*
3. L. Blumental and K. Menger. Studies in geometry. *Freeman and Company, 1970.*
4. M. Deza and M. Laurent. Geometry of cuts and metrics. *Springer, 1991.*
5. T. Jones. Evolutionary Algorithms, Fitness Landscapes and Search. *PhD dissertation, University of New Mexico, 1995.*
6. Y.H. Kim, Y. Yoon, A. Moraglio and B.R. Moon. Geometric crossover for multiway graph partitioning. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 1217–1224, 2006.
7. S. Lucas, A. Moraglio, H. Li, R. Poli and Q. Zhang. Alignment Crossover for Finite State Machines. *(in preparation).*
8. S. Luke and L. Spector. A Revised Comparison of Crossover and Mutation in Genetic Programming. *Proceedings of Genetic Programming Conference*, 1998.
9. A. Moraglio and R. Poli. Topological interpretation of crossover. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 1377–1388, 2004.
10. A. Moraglio and R. Poli. Topological crossover for the permutation representation. In *Workshop on theory of representations at Genetic and Evolutionary Computation Conference*, 2005.
11. A. Moraglio and R. Poli. Geometric Crossover for the Permutation Representation. *Technical Report CSM-429, University of Essex, 2005.*
12. A. Moraglio and R. Poli. Geometric Landscape of Homologous Crossover for Syntactic Trees. *Proceedings of IEEE Congress of Evolutionary Computation*, pages 427–434, 2005.
13. A. Moraglio and R. Poli. Product Geometric Crossover. *Proceedings of Parallel Problem Solving from Nature*, pages 1018–1027, 2006.
14. A. Moraglio and R. Poli. Geometric Crossover for Sets, Multisets and Partitions. *Proceedings of Parallel Problem Solving from Nature*, pages 1038–1047, 2006.
15. A. Moraglio, R. Poli and R. Seehuus. Geometric Crossover for Biological Sequences. *Proceedings of Genetic Programming Conference*, pages 121–132, 2006.
16. A. Moraglio, J. Togelius and S. Lucas. Product Geometric Crossover for the Sudoku Puzzle. *Proceedings of IEEE Congress of Evolutionary Computation*, 470–476, 2006.
17. P.M. Pardalos and M.G.C. Resende. Handbook of Applied Optimization. *Oxford University Press, 2002.*
18. N. Radcliffe. Equivalence Class Analysis of Genetic Algorithms. *Journal of Complex Systems, 1991*, vol. 5, pages 183–205.
19. C.M. Reidys and P.F. Stadler. Combinatorial landscapes. *SIAM Review*, 44:3–54, 2002.
20. R. Seehuus and A. Moraglio. Geometric Crossover for Protein Motif Discovery. In *Workshop on theory of representations at Genetic and Evolutionary Computation Conference*, 2006.
21. M.L.J. Van de Vel. Theory of Convex Structures. *North-Holland, 1993* .
22. S. Wright. The roles of mutation, inbreeding, crossbreeding and selection in evolution. In D. F. Jones, editor, *Proceedings of the Sixth International Congress on Genetics*, volume 1, pages 356–366, 1932.