

On the Topology of Genetic Algorithms

David Hofmeyr

Lancaster University, Lancaster, UK

d.hofmeyr@lancaster.ac.uk

Abstract

Genetic algorithms are stochastic search heuristics which are popular for their broad applicability, especially in combinatorial search problems. The search mechanism relies on an abstraction of genetic evolution and selection as seen in nature. This paper introduces a topological structure for the search space which is consistent with existing theory and practice for genetic algorithms, namely forma analysis. A notion of convexity is defined within this context and connections between this definition and forma analysis are established. This framework provides an alternative perspective on the exploitation/exploration dilemma as well as population convergence, which relates directly to the genetic operators employed to drive the evolution process. It also provides a different interpretation of design constraints associated with genetic algorithm implementations. The intention is to provide a new analytical perspective for genetic algorithms, and to establish a connection with exact search methods through the concept of convexity.

1 Introduction

Genetic algorithms (GAs) are search techniques which fall in the broader class of evolutionary algorithms (EAs). Within EAs a population of candidate solutions for a problem is evolved through iterative production and selection. Selection bias is used to ensure that with high probability the quality of the overall population improves in each generation (round of selection). The quality of a solution is referred to as its *fitness*. GAs are specific for their use of techniques which mimic genetic mechanisms observed in nature, such as gene crossover and mutation, to produce new candidates.

A GA is thus the recursive application of so-called *genetic operators* to an evolving population $\mathcal{P} \in P_M(\mathcal{X})$, where \mathcal{X} is the space of all solutions and $P_M(\mathcal{X})$ denotes all multisets which can be constructed from elements in \mathcal{X} . A genetic operator may be seen as a random function either mapping from $P_M(\mathcal{X})$ to $P_M(\mathcal{X})$, or from a finite product, $\mathcal{X}^k, k \in \mathbb{N}$, to the solution space \mathcal{X} . In other words, if K is a genetic operator from \mathcal{X}^k to \mathcal{X} then for any $x_1, \dots, x_k \in \mathcal{X}$, $K(x_1, \dots, x_k)$ is a random variable on \mathcal{X} . Similarly $K(A)$

may be a random variable on $P_M(\mathcal{X})$ for all $A \in P_M(\mathcal{X})$. Common examples of genetic operators taking values in \mathcal{X} include *recombination*, which produces a random $x \in \mathcal{X}$, and takes as argument (usually) two members $y, z \in \mathcal{X}$, known as the parents of x , and *mutation*, which takes a single argument from \mathcal{X} and returns a random element of \mathcal{X} which tends to be similar. On the other hand, *selection* operators take arguments in $P_M(\mathcal{X})$ and return a random subset of the argument. Selection operators are used to determine which individuals are used for recombination, as well as to determine which individuals survive in the subsequent generation.

The term ‘‘Genetic Algorithm’’ was coined by John Holland in his seminal work [Holland, 1975], where the potential of GAs was described through *schema analysis* and the concept of *implicit parallelism*. Here each candidate solution is encoded as a binary string of fixed length. The positions on such a string are analogies for the gene locations along a genome.

A schema is a template for such representations. Specifically it is a subset of the space of gene representations for which certain positions in the string (gene locations) have specified values. For example, if solutions are encoded using strings of length 7, then a schema in which the first position has value a , and the fourth and fifth positions take values b and c respectively may be represented by

$$a \square \square b c \square \square \quad (1)$$

where \square 's may assume any value. The schema theorem then states that schemata with simple structure which have above average fitness will tend to proliferate [Holland, 1975]. Simple structure in this case means that few positions have specified value, and also the distance between the first and last specified position is small. With a relative abundance of high quality schemata present in a population, the search inherently operates on the schemata and not the individual positions, and hence there is an implicit reduction in the dimension of the problem [Goldberg, 1989]. Implicit parallelism in the context of GAs refers to the fact that the number of schemata effectively being processed is, in general, far greater than the number of individual candidate solutions. For practically sized problems [Holland, 1980] and [Goldberg, 1985] have shown that $\mathcal{O}(n^3)$ schemata are

present for n random binary strings.

The specificity of schema analysis, arising from the fact that solutions must be encoded as linearly arranged strings of a fixed length, has been argued as too limiting and generalisations of the associated theory have been developed by, for example [Vose, 1991], [Radcliffe, 1992], and [Greene, 2000]. Most relevant, and central to the motivation for this paper, is that of [Radcliffe, 1992], and the concept of forma analysis. Forma analysis provides a basis for the representation of candidate solutions in terms of their equivalence class memberships arising from an arbitrary collection of equivalence relations. The representation through equivalence relations provides a convenient and meaningful connection with topology through uniformities. In finite dimensional GAs it is shown that the associated topology is pseudo-metrisable, and provided the choice of equivalence relations is sufficient to describe each solution uniquely the corresponding distance function is a metric. A generalisation of convexity is defined within this topological framework, and connections with forma analysis are established. This notion of convexity is then extended to genetic operators in a way which is consistent with the design constraints given by [Radcliffe, 1994] in the context of forma analysis, and is shown to have connections with exploitation and population convergence.

The success and versatility of GAs has led to a number of explorations into their theoretical properties, including their algebraic structure [Radcliffe, 1994], the spectral properties of genetic operators [Schmitt, 2001], the geometric relationships between genetic operators and the fitness landscape [Moraglio, 2007], and the long run behaviour of GAs through fixed points of dynamical systems [Vose, 1999] and their limiting distributions as Markov chains [Nix and Vose, 1992], [Suzuki, 1995], [Schmitt, 2001] and [Schmitt, 2004]. To our knowledge this is the first paper to formally investigate the topological structure of genetic algorithms.

The remainder of this paper is as follows. In Section 2 forma analysis is discussed in greater detail. Concepts relevant in the remaining paper are also introduced. Section 3 describes the proposed topological framework and provides the main results of this paper. In Section 4 some concluding remarks are given.

2 Forma Analysis

It has been argued [Radcliffe, 1992] that unless the representation of solutions has relevance within the actual solution space, GAs can't do better than random search, and that the emergence of powerful schemata is incidental. That is, unless similarities in the representations described by schemata imply similarity within the solution space as well, the distribution of fitness across schemata is random and collections of highly fit individuals within schemata occur by chance.

A direct and principled address of this concern is given in the context of forma analysis [Radcliffe, 1992]. Here the representation is directly related to subsets of the solution space within which solutions have similar characteristics, which can

be defined in terms of equivalence relations. These equivalence relations can also be used to encode information about the fitness function used to evaluate solutions. The representation of a solution is therefore given by the collection of equivalence classes to which it belongs, under an arbitrary collection of equivalence relations over the solution space. These equivalence classes are referred to as formae, and a rigorous presentation of the algebraic structure of formae has been developed [Radcliffe, 1994].

This representation is sufficiently general to capture arguably any problem space, but provides added benefits in terms of interpretability. It also ensures the relevance of the relationship between solutions arising in the same equivalence classes, where in the case of schema analysis these relationships may be fairly arbitrary [Radcliffe, 1992].

Note that a representation via equivalence classes is a generalisation of schema analysis, since schemata can be formulated through equivalence relations. If we return to strings of length 7 then a variable template written as

$$\blacksquare \square \square \blacksquare \blacksquare \square \square \quad (2)$$

where \blacksquare 's represent specified (but variable) positions and \square 's are as before unspecified positions, then (2) represents an equivalence relation under which two solutions are equivalent iff they are equal in the first, fourth and fifth positions. The schema given in (1) is therefore an equivalence class arising from this relation.

Henceforth equivalence relations will be written as \sim , and if two solutions x and y are equivalent under \sim , we write $x \sim y$. The equivalence class of a solution x under \sim will be written $[x]_{\sim}$. Combinations of equivalence relations can combine to form smaller and smaller equivalence classes, as described in the following lemma, which can be found in any standard text on algebra.

Lemma 2.1 *Let \sim_1 and \sim_2 be two equivalence relations on a set \mathcal{X} . Then their intersection $\sim_{1,2} = \sim_1 \cap \sim_2$, defined as*

$$x \sim_{1,2} y \iff x \sim_1 y \text{ and } x \sim_2 y,$$

is an equivalence relation on \mathcal{X} and the induced equivalence classes satisfy $[x]_{\sim_{1,2}} = [x]_{\sim_1} \cap [x]_{\sim_2}$.

This leads us to consider the collection of equivalence relations which result from intersecting different combinations of relations.

Definition Let \mathcal{E} be a collection of equivalence relations on a set \mathcal{X} . Then the *generated set* of \mathcal{E} , $\mathcal{G}(\mathcal{E})$, is defined as the collection of finite intersections of members of \mathcal{E} . That is,

$$\mathcal{G}(\mathcal{E}) := \{ \cap_{i \in I} \sim_i \mid I \subset \mathcal{E}, I \text{ finite} \}.$$

The following definitions arise from [Radcliffe, 1994] and are important in the theory presented in Section 3.

Definition A collection of equivalence relations \mathcal{E} *covers* \mathcal{X} iff for any $x, y \in \mathcal{X}$ with $x \neq y$; $\exists \sim \in \mathcal{E}$ s.t. $x \not\sim y$.

Coverage can be understood as a separation of points in \mathcal{X} by the members of \mathcal{E} . Separation of points is fundamental to the structure of topological spaces, and the notion of coverage above is used to establish a metrisable topology on the search space in the following section.

Definition A collection of equivalence relations \mathcal{E} on \mathcal{X} is *orthogonal* if any collection of equivalence classes induced by different members of \mathcal{E} has nonempty intersection.

Orthogonality ensures that the mapping from representation to solution is fully defined, but is also crucial in the formulation of convexity which we propose. Convexity requires that the underlying space has no “gaps”, i.e., that generalisations of the line segments between points in the representation space are not broken by missingness in the solution space.

[Radcliffe, 1994] also provides useful characterisations of recombination operators; genetic operators which recombine the genes of a small collection of (usually two) members of \mathcal{X} . While the results of the following section may be generalised, we will limit attention to recombinations of two individuals’ genes.

Definition Let \mathcal{E} be a collection of equivalence relations on \mathcal{X} . A recombination operator, K , mapping from $\mathcal{X} \times \mathcal{X}$ to \mathcal{X} , *transmits* genes if for any pair $x, y \in \mathcal{X}$ and $\sim \in \mathcal{E}$, one has $\text{Support}(K(x, y)) \subset [x]_{\sim} \cup [y]_{\sim}$.

A *child* solution arising from K is therefore related to at least one of its parents by each element of \mathcal{E} .

Definition Let \mathcal{E} be a collection of equivalence relations on \mathcal{X} . A recombination operator, K , mapping from $\mathcal{X} \times \mathcal{X}$ to \mathcal{X} , *respects* genes if for any $\sim \in \mathcal{G}(\mathcal{E})$ and any $y \in [x]_{\sim}$ one has $\text{Support}(K(x, y)) \subset [x]_{\sim}$.

Any shared information in the parent solutions, which is expressible in terms of the formae, will remain in all of their children arising from K .

Definition Let \mathcal{E} be a collection of equivalence relations on \mathcal{X} . A recombination operator, K , mapping from $\mathcal{X} \times \mathcal{X}$ to \mathcal{X} , *assorts* genes if for any $\sim_1, \sim_2 \in \mathcal{G}(\mathcal{E})$ and $x, y \in \mathcal{X}$ with $[x]_{\sim_1} \cap [y]_{\sim_2} \neq \emptyset$, it is possible to produce (possibly over multiple generations) $z \in [x]_{\sim_1} \cap [y]_{\sim_2}$ using only x, y and their descendents resulting from applications of K .

The assorting property of K means that any combination of genes in a pair of solutions can arise in their descendants. Alternatively, the entirety of their *dynastic potential*, defined below, can be explored using only the operator K .

Definition Let \mathcal{E} be a collection of equivalence relations on \mathcal{X} and let $L \in P_M(\mathcal{X})$. The *dynastic potential* of L , denoted $\Gamma(L)$, is defined as all members of \mathcal{X} which are related to a member of L by each $\sim \in \mathcal{E}$. Formally,

$$\Gamma(L) := \{y \in \mathcal{X} \mid \forall \sim \in \mathcal{E} \exists x \in L \text{ s.t. } y \sim x\}.$$

In the following section these properties of recombination operators will be discussed theoretically in the context of convexity, establishing connections between these definitions and the concepts of exploitative search and population convergence. Conditions are also described which permit the designation of $\Gamma(L)$ as the potential child solutions arising from L , their potential “dynasty”.

3 The Topology of Genetic Algorithms

This section introduces a topological structure for the solution space of GAs. The equivalence relation formulation adopted

in forma analysis provides a convenient and meaningful link with topology through uniform structures.

A uniformity \mathcal{V} is a filter on the product $\mathcal{X} \times \mathcal{X}$, i.e., a non-empty collection of subsets of $\mathcal{X} \times \mathcal{X}$ which does not contain the empty set and is stable under finite intersection and supersets, which satisfies the following. Each *entourage* $V \in \mathcal{V}$ contains the diagonal of \mathcal{X} , $\{(x, x) \mid x \in \mathcal{X}\} \subset V$. For each $V \in \mathcal{V}$ there is a $U \in \mathcal{V}$ s.t. the square of U , $U \cdot U := \{(x, y) \mid \exists z \in \mathcal{X} \text{ with } (x, z), (z, y) \in U\}$, is contained in V . Finally, if $V \in \mathcal{V}$, then the inverse of V , $V^{-1} := \{(x, y) \mid (y, x) \in V\}$, is in \mathcal{V} .

\mathcal{V} generates a topology on \mathcal{X} in which the basic neighbourhoods of a point $x \in \mathcal{X}$ are given by $\{V(x) \mid V \in \mathcal{V}\}$ where, $V(x) := \{y \in \mathcal{X} \mid (x, y) \in V\}$. A basis for a uniformity is simply a filter base on $\mathcal{X} \times \mathcal{X}$ for which the generated filter is a uniformity. The following lemma can be found in any extensive text on general topology.

Lemma 3.1 *Let \mathcal{E} be a non-empty collection of equivalence relations on a set \mathcal{X} . Then $\mathcal{B}_{\mathcal{E}} := \{V_{\sim} \mid \sim \in \mathcal{G}(\mathcal{E})\}$, where*

$$V_{\sim} = \{(x, y) \mid x \sim y\},$$

is a basis for a uniformity on \mathcal{X} , say $\mathcal{V}_{\mathcal{E}}$.

We now introduce some of the properties of this topology, in relation to GAs and forma analysis. Firstly, it can be shown that in most practical cases the resulting topology is metrisable. In addition, in the case of representation by binary strings, the metric is equivalent to the hamming distance function.

Lemma 3.2 *Let the conditions of Lemma 3.1 hold. If \mathcal{E} is finite then the topology generated by $\mathcal{V}_{\mathcal{E}}$ is pseudo-metrisable.*

Proof Consider the function

$$d(x, y) = |\{\sim \in \mathcal{E} \mid x \not\sim y\}|,$$

where $|\cdot|$ is the cardinality operator. It is easy to verify that d is a pseudo-metric. To see that the topologies are equivalent, consider that a point $x \in \mathcal{X}$ has a smallest pseudo-metric neighbourhood described by $\{y \in \mathcal{X} \mid x \sim y \forall \sim \in \mathcal{E}\} = \{y \in \mathcal{X} \mid d(x, y) = 0\}$. Similarly x has a smallest uniform neighbourhood described by $V_{\bigcap_{\sim \in \mathcal{E}} \sim}(x)$, which is plainly the same set. Thus every pseudo-metric neighbourhood contains a uniform neighbourhood, and vice versa, and hence the topologies are equivalent. ■

Corollary 3.3 *Let the conditions of Lemma 3.1 hold. If \mathcal{E} is finite and covers \mathcal{X} then $\mathcal{V}_{\mathcal{E}}$ is metrisable.*

Proof Let d be defined as in the proof of Lemma 3.2. Since \mathcal{E} covers \mathcal{X} we have $\forall x, y \in \mathcal{X}, x \neq y, \exists \sim \in \mathcal{E} \text{ s.t. } x \not\sim y$ and so

$$d(x, y) = |\{\sim \in \mathcal{E} \mid x \not\sim y\}| > 0.$$

Therefore d is a metric. ■

We can thus endow the space with a topology that conveniently coincides with the theory of forma analysis, and the basic neighbourhoods of a point are merely the formae to which it belongs.

In order to establish connections with exact search methods through convexity it is necessary to define convexity within

the current context. Convexity in exact search and optimisation is ostensibly an algebraic property, yet in essence it is a geometric concept. It should therefore be defined within topological spaces, and specifically those wherein structure obeys rigidity, namely metrisable topologies.

An existing definition of convexity in general metric spaces already exists [Blumenthal, 1953], yet this definition is not consistent with the algebraic formulation used in search and optimisation. Here a set C is convex in a metric space (\mathcal{X}, d) iff whenever $x, y \in C$ with $x \neq y$ there exists $z \in C$ distinct from both x and y satisfying $d(x, y) = d(x, z) + d(z, y)$. To see that this is not consistent with the algebraic definition of convexity it is sufficient to consider the rational numbers, \mathbb{Q} , in the space (\mathbb{R}, d_u) , where d_u is the Euclidean metric.

A preferable definition for the current investigation is as follows.

Definition Let (\mathcal{X}, d) be a metric space. A set $C \subset \mathcal{X}$ is *convex with respect to d* iff whenever $x, y \in C$ and $\exists z \in \mathcal{X}$ s.t. $d(x, y) = d(x, z) + d(z, y)$, then $z \in C$.

We will refer to the condition $d(x, y) = d(x, z) + d(z, y)$ as “ z is d -between x and y ”. It is straightforward to check that this definition is consistent with the algebraic notion of convexity.

With this definition we can show that the dynastic potential of a set is equivalent to its convex hull, which is useful for understanding population convergence within the proposed framework. For brevity in the remaining paper we will occasionally not make an explicit distinction between a multiset and the set containing its unique elements. For example, we may state that $\Gamma(L) = L$, where $\Gamma(L)$ by definition is a set and hence this statement is not strictly true whenever L contains repeated elements. What is meant by this equality is $\Gamma(L) \subset L$ and $L \subset \Gamma(L)$ for the natural interpretation of inclusion for multisets.

Lemma 3.4 *Let the conditions of Lemma 3.1 hold. If \mathcal{E} is finite and orthogonal and covers \mathcal{X} then the dynastic potential of a set $L \in P_M(\mathcal{X})$ is equal to its convex hull with respect to the forma based topology.*

Proof Plainly $L \subset \Gamma(L)$. To show that $\Gamma(L)$ is convex, consider $x, y \in \Gamma(L)$ and suppose $\exists z \in \mathcal{X}$ s.t. $d(x, y) = d(x, z) + d(z, y)$. Suppose now that $z \notin \Gamma(L)$, then $\exists \sim \in \mathcal{E}$ s.t. $z \not\sim w \forall w \in L$. But $x, y \in \Gamma(L) \Rightarrow \exists w_x, w_y \in L$ s.t. $x \sim w_x$ and $y \sim w_y$, thus $z \not\sim x$ and $z \not\sim y$ and so $\{\sim \in \mathcal{E} | z \not\sim x\} \cap \{\sim \in \mathcal{E} | z \not\sim y\} \neq \emptyset$. Now observe that

$$\begin{aligned} & (x \sim z \cap y \sim z \Rightarrow x \sim y) \\ & \Rightarrow \{\sim \in \mathcal{E} | x \sim z\} \cap \{\sim \in \mathcal{E} | y \sim z\} \subset \{\sim \in \mathcal{E} | x \sim y\} \\ & \Rightarrow \{\sim \in \mathcal{E} | x \not\sim y\} \subset \{\sim \in \mathcal{E} | z \not\sim x\} \cup \{\sim \in \mathcal{E} | z \not\sim y\} \end{aligned}$$

But since the two sets on the right hand side have non empty intersection, and \mathcal{E} is finite, the size of the left hand side must be strictly less than the sum of those on the right, i.e.

$$d(x, y) < d(x, z) + d(z, y)$$

a contradiction. Therefore we must have $z \in \Gamma(L)$.

Now, consider some convex set C with $L \subset C$. Take $z \in \Gamma(L)$. Suppose $|\mathcal{E}| = n$ and impose some arbitrary ordering of

its elements so that $\mathcal{E} = \{\sim_i | i = 1, \dots, n\}$. Since $z \in \Gamma(L)$ we have $\forall i \in \{1, \dots, n\}; \exists x_i \in L$ s.t. $z \sim_i x_i$. For integer $k \leq n$ let $\mathcal{E}_k := \{\sim_i | i = 1, \dots, k\}$. By induction on these indexed subsets of \mathcal{E} we will show that $\forall k \exists c_k \in C$ s.t. $\forall \sim \in \mathcal{E}_k, c_k \sim z$.

For $k=1$ take $c_1 = x_1$ as defined above and the result holds. Suppose the inductive hypothesis holds for some $\mathcal{E}_k, 1 \leq k < n$. We have $c_k, x_{k+1} \in C$. Take c_{k+1} to be any member of the equivalence class formed by the intersection of equivalence classes of c_k with respect to $\sim_1, \dots, \sim_k, \sim_{k+2}, \dots, \sim_n$ and of x_{k+1} with respect to \sim_{k+1} , which is non-empty since \mathcal{E} is orthogonal. It should be plain to see that $d(c_k, x_{k+1}) = d(c_k, c_{k+1}) + d(c_{k+1}, x_{k+1})$ and so $c_{k+1} \in C$ and $c_{k+1} \sim z \forall \sim \in \mathcal{E}_{k+1}$ as required.

We have thus shown by induction that $\exists c \in C$ s.t. $c \sim z \forall \sim \in \mathcal{E}$ and since \mathcal{E} covers \mathcal{X} we must have $c = z$. We thus have that $\Gamma(L) \subset C$ and hence is the smallest convex set containing L . ■

We can further describe the geometry of convex sets through their extreme points. An extreme point of a convex set, C , is any $x \in C$ s.t. $C \setminus \{x\}$ is convex. Notice that when looking at extreme points it is more interesting to consider subsets of \mathcal{X} , and not multisets.

Lemma 3.5 *Let the conditions of Lemma 3.4 hold. Let $C \subset \mathcal{X}$ be convex. Then $x \in C$ is an extreme point of C iff $\forall y \in C \setminus \{x\} \exists \sim \in \mathcal{E}$ s.t. $x \not\sim y \forall \sim' \in \mathcal{E} \setminus \{\sim\}$.*

Proof First suppose x is an extreme point of C . Suppose though that $\exists y \in C \setminus \{x\}, \sim_1, \sim_2 \in \mathcal{E}, \sim_1 \neq \sim_2$, with $x \not\sim_1 y$ and $x \not\sim_2 y$. Since \mathcal{E} is orthogonal $\exists w_1, w_2 \in \mathcal{X}$ s.t. $w_1 \sim_1 y, w_1 \sim x \forall \sim \in \mathcal{E} \setminus \{\sim_1\}$ and $w_2 \sim_2 y, w_2 \sim x \forall \sim \in \mathcal{E} \setminus \{\sim_2\}$. By design w_1 and w_2 are d -between x and y and therefore $w_1, w_2 \in C$ since C is convex. However, we have $\{\sim \in \mathcal{E} | w_1 \not\sim w_2\} = \{\sim_1, \sim_2\}$, and therefore $d(w_1, w_2) = 2 = d(w_1, x) + d(x, w_2)$, and hence x is d -between w_1 and w_2 . Therefore $C \setminus \{x\}$ cannot be convex, a contradiction. Therefore $\forall y \in C \setminus \{x\} \exists \sim \in \mathcal{E}$ s.t. $x \not\sim y \forall \sim' \in \mathcal{E} \setminus \{\sim\}$.

Suppose in the converse that $\forall y \in C \setminus \{x\} \exists \sim \in \mathcal{E}$ s.t. $x \not\sim y \forall \sim' \in \mathcal{E} \setminus \{\sim\}$. If it were that $C \setminus \{x\}$ were not convex, then $\exists y, z \in C$ s.t. x is d -between y and z . Then $\exists \sim_1, \sim_2 \in \mathcal{E}, \sim_1 \neq \sim_2$ with $x \not\sim_1 y$ and $x \not\sim_2 z$. Notice that by assumption we have $x \sim y \forall \sim \in \mathcal{E} \setminus \{\sim_1\}$ and $x \sim z \forall \sim \in \mathcal{E} \setminus \{\sim_2\}$. Again, since \mathcal{E} is orthogonal, $\exists w \in \mathcal{X}$ s.t. $w \sim y \forall \sim \in \mathcal{E} \setminus \{\sim_2\}$ and $w \sim z$. So w is d -between y and z by design, and hence $w \in C$, since C is convex. But $w \not\sim_1 x$ and $w \not\sim_2 x$, a contradiction of the assumption. Therefore $C \setminus \{x\}$ is convex. ■

Convexity in genetic algorithms can also be used to describe exploitative search, as well as to investigate population convergence, as discussed in the sequel. To establish these connections, the definition of convexity is first extended to genetic operators.

Note that given an exploitative recombination operator, it is fairly straightforward to obtain an associated explorative operator as follows. If K is an arbitrary genetic operator mapping $\mathcal{X} \times \mathcal{X}$ to \mathcal{X} , then we may define an explorative operator K^e by $K^e(x, y) = K(K(x, y), X)$, where X is a random variable on \mathcal{X} . The degree of exploration can be controlled through the variability of the random variable X .

The converse of defining exploitative operators using only explorative ones, is not so straightforward. We therefore focus only on exploitation in the remainder, which we introduce through convex operators.

Definition A genetic operator K mapping $\mathcal{X} \times \mathcal{X}$ to \mathcal{X} is *convex* iff whenever $C \subset \mathcal{X}$ is convex we have

$$x_1, x_2 \in C \Rightarrow \text{Support}(K(x_1, x_2)) \subset C.$$

We may also define this property for genetic operators mapping from $P_M(\mathcal{X})$ to $P_M(\mathcal{X})$ in the obvious way; $A \in P_M(C) \Rightarrow \text{Support}(K(A)) \subset P_M(C)$, where $P_M(C)$ is now the collection of multisets arising from elements in C .

We may also think of recombination operators as being in some sense affine. In general we think of a function being affine if it is restricted to some subspace of the space of solutions. A similar consideration can be made for recombination operators.

Definition A genetic operator K mapping $\mathcal{X} \times \mathcal{X}$ to \mathcal{X} is called *affine* at x in the direction of $y \in \mathcal{X}$ if $\exists \sim \in \mathcal{E}$ with $\Gamma(\{x, y\}) \neq \{x, y\}$ s.t.

$$\mathbb{P}(K(x, z) \sim x) = 0 \quad \forall z \in [y]_{\sim}.$$

Affine operators may be locally restricted to a subspace, e.g. $[y]_{\sim}$ in the above definition. In addition, such operators may be viewed in the context of gene dominance, where in the above definition x is dominated by $[y]_{\sim}$. Although the definition of affine does not provide a direct analogue for affine functions, it admits a natural definition of strict convexity.

Definition A genetic operator K mapping $\mathcal{X} \times \mathcal{X}$ to \mathcal{X} is *strictly convex* iff K is convex and nowhere affine.

Such an operator is therefore convex, and cannot become stuck within some subspace of \mathcal{X} . While it is clear that strict convexity is a desirable property for recombination operators, it is not sufficient for effective search of the overall GA. We therefore also define a stronger condition as follows.

Definition A genetic operator K mapping $\mathcal{X} \times \mathcal{X}$ to \mathcal{X} is *strongly non-affine* if $\forall x, y \in \mathcal{X}$ the following holds. For all $z \in \Gamma(\{x, y\})$ s.t. $d(x, z) = 1, \exists w \in \text{Support}(K(x, y))$ s.t. $z \in \text{Support}(K(x, w))$.

Definition A genetic operator K mapping $\mathcal{X} \times \mathcal{X}$ to \mathcal{X} is *strongly convex* iff K is convex and strongly non-affine.

The definition of strong convexity does not lend itself immediately to operators mapping $P_M(\mathcal{X})$ to $P_M(\mathcal{X})$. In this context we define a genetic operator K to be strongly convex if it is convex and whenever $A = \{x_1, \dots, x_k\} \in P_M(\mathcal{X})$, for $k > 1$, we have $\min\{\mathbb{P}(x_i \notin K(A)), \mathbb{P}(x_i \in K(A))\} > 0 \quad \forall i \in \{1, \dots, k\}$.

These definitions of convexity, strict convexity and strong convexity have direct interpretations in the context of the desirable properties for recombination discussed by [Radcliffe, 1992], which are discussed below.

Lemma 3.6 *Let \mathcal{E} be a collection of equivalence relations on \mathcal{X} and let K be a recombination operator from $\mathcal{X} \times \mathcal{X}$ to \mathcal{X} . Then K transmits genes iff it is convex.*

Proof Suppose that K transmits genes. Observe that if for $x, y, z \in \mathcal{X}$ we have $\forall \sim \in \mathcal{E}$ either $z \sim y$ or $z \sim x$, then

$$\{\sim \in \mathcal{E} | z \not\sim x\} \cap \{\sim \in \mathcal{E} | z \not\sim y\} = \emptyset.$$

Moreover, in the converse of the proof of Lemma 3.4

$$\begin{aligned} & (x \sim y \Rightarrow x \sim z \cap y \sim z) \\ & \Rightarrow \{\sim \in \mathcal{E} | x \sim y\} \subset \{\sim \in \mathcal{E} | x \sim z\} \cap \{\sim \in \mathcal{E} | y \sim z\} \\ & \Rightarrow \{\sim \in \mathcal{E} | z \not\sim x\} \cup \{\sim \in \mathcal{E} | z \not\sim y\} \subset \{\sim \in \mathcal{E} | x \not\sim y\} \end{aligned}$$

In all we have that the union of two sets which have empty intersection is contained in another set, and thus the sum of their sizes must be less than or equal that of the size of the other. In other words

$$d(x, z) + d(z, y) \leq d(x, y)$$

but since d is a pseudo metric, we must have

$$d(x, z) + d(z, y) = d(x, y).$$

Thus gene transmissive operators generate children which are d -between their parents, and thus the children must lie in any convex set containing them.

In the converse, suppose that K is convex and take $x, y \in \mathcal{X}$. Then $\text{Support}(K(x, y)) \subset \Gamma(\{x, y\})$, since $\Gamma(\{x, y\})$ is convex. For each $\sim \in \mathcal{E}$ we therefore have $\text{Support}(K(x, y)) \subset [x]_{\sim} \cup [y]_{\sim}$, and hence K transmits genes. ■

Lemma 3.7 *Let \mathcal{E} be a collection of equivalence relations on \mathcal{X} and let K be a recombination operator mapping $\mathcal{X} \times \mathcal{X}$ to \mathcal{X} . If K transmits and assort genes then it is strictly convex.*

Proof We have seen that gene transmission implies convexity, which leaves us with showing that the addition of assortment gives us strict convexity. Suppose that K is a convex operator but is not strictly convex, i.e., it is affine at some $x \in \mathcal{X}$. Then $\exists y \in \mathcal{X}, \sim \in \mathcal{E}$ with $\Gamma(\{x, y\}) \neq \{x, y\}$ s.t. $\mathbb{P}(K(x, z) \sim x) = 0 \quad \forall z \in [y]_{\sim}$. Notice that $\Gamma(\{x, y\}) \neq \{x, y\} \Rightarrow \exists \sim_1, \sim_2 \in \mathcal{E}$ with $x \not\sim_1 y$ and $x \not\sim_2 y$. Now, since $\mathbb{P}(K(x, y) \sim x) = 0$ we must have $x \not\sim y$. Thus $\exists \sim' \in \mathcal{E}$ with $\sim' \neq \sim$ and $x \not\sim' y$ and $x \not\sim' y$. Now, by definition we have $[x]_{\sim} \cap [y]_{\sim'} \cap \Gamma(\{x, y\}) \neq \emptyset$, and this intersection contains neither x nor y . Because K is affine at x with respect to y and \sim we know that no future generations arising from x, y and their descendants can be related to x by \sim , since x is dominated by $[y]_{\sim}$, and thus no member of the above intersection can occur. However assortment guarantees such an individual can occur. Therefore if K is anywhere affine, it cannot assort, and the result follows. ■

Lemma 3.8 *Let \mathcal{E} be a finite, orthogonal and cover \mathcal{X} and let K be a recombination operator mapping $\mathcal{X} \times \mathcal{X}$ to \mathcal{X} . If K is strongly non-affine then it assort genes.*

Proof Take $x, y \in \mathcal{X}$ and $\sim_1, \sim_2 \in \mathcal{G}(\mathcal{E})$ s.t. $[x]_{\sim_1} \cap [y]_{\sim_2} \neq \emptyset$. Let $\mathcal{E}_x \subset \mathcal{E}$ be such that $(\bigcap_{\sim \in \mathcal{E}_x} \sim) = \sim_1$. Assign an arbitrary ordering of the elements of $\mathcal{E}_x = \{\sim_1^x, \dots, \sim_{|\mathcal{E}_x|}^x\}$. Since \mathcal{E} is orthogonal $\exists w_1 \in \mathcal{X}$ s.t. $w_1 \sim_1^x x$ and $w \sim y \quad \forall \sim \in \mathcal{E} \setminus \{\sim_1^x\}$. Therefore $w_1 \in \Gamma(\{x, y\})$ and $d(y, w_1) \leq 1$. If $d(y, w_1) = 0$ then since \mathcal{E} covers \mathcal{X} we know $w_1 = y$. Otherwise, since K is strongly non-affine $\exists z_1 \in \text{Support}(K(x, y))$ s.t. $w_1 \in \text{Support}(K(y, z_1))$. Therefore w_1 can be made using only

x, y and their descendants. Now suppose that for $1 \leq k < |\mathcal{E}_x|$ there exists a $w_k \in \mathcal{X}$ satisfying $w_k \sim_x^i x$ for $i=1, \dots, k$ and $w_k \sim y \forall \sim \in \mathcal{E} \setminus \{\sim_x^1, \dots, \sim_x^k\}$ which can be made using only x, y and their descendants. Similar to above $\exists w_{k+1} \in \mathcal{X}$ s.t. $w_{k+1} \sim_x^i x \forall i=1, \dots, k$ and $w_{k+1} \sim y \forall \sim \in \mathcal{E} \setminus \{\sim_x^1, \dots, \sim_x^{k+1}\}$. Then $w_{k+1} \in \Gamma(\{w_k, y\})$ and $d(w_k, w_{k+1}) \leq 1$. As before we can conclude that w_{k+1} can be made using only x, y and their descendants. By induction we can find a $w_1 \in [x]_{\sim_1}$ using only x, y and their descendants. Similarly we can construct a $w_2 \in [y]_{\sim_2}$ which remains in $[x]_{\sim_1}$. Notice that any overlap in the relations, \sim , contributing to both \sim_1 and \sim_2 , must satisfy $x \sim y$, and so these do not present a problem for the above argument. ■

Note that we do not have all recombination operators that assort genes being strongly non-affine. This is because the definition of assortment allows an arbitrary delay until a specific combination of genes can arise, while in the case of strongly non-affine operators an upper bound on the time by which any combination can be produced can easily be derived.

We can use these results to discuss exploitation and convergence in the context of the proposed framework. Convex genetic operators can be seen as exploitative in the sense that they exploit the genetic information present within the population, given by its dynastic potential.

Lemma 3.9 *Let \mathcal{E} be a finite collection of equivalence relations which is orthogonal and covers \mathcal{X} . If all genetic operators are convex then the collection of descendants of a multiset $L \in P_M(\mathcal{X})$ is contained within their dynastic potential.*

Proof The result follows immediately from the fact that $L \subset \Gamma(L)$, and $\Gamma(L)$ is convex. ■

Strongly convex operators also have the property of being able to access all elements of $\Gamma(L)$, for any starting population L , through their gene assorting property. It is also this ability which leads to population convergence. Population convergence is an important consideration in GA methodology. Early convergence does not allow an adequate search of the associated problem domain, yet ultimately the strong gene combinations which have been discovered need to be exploited. Strongly convex operators may be seen as convergent operators, as follows.

Lemma 3.10 *Let \mathcal{E} be a finite collection of equivalence relations which is orthogonal and covers \mathcal{X} . Let L_i be the population present in a genetic algorithm at generation i and assume that $\{|L_i|\}_{i \in \mathbb{N}}$ is bounded. If a fixed and finite collection of strongly convex genetic operators are employed in each generation then with probability 1,*

$$\lim_{n \rightarrow \infty} \text{Diam}(L_n) = 0,$$

where $\text{Diam}(A) = \max\{d(x, y) | x, y \in A\}$, $A \in P_M(\mathcal{X})$. Moreover the expected time to convergence is finite.

Proof In the interest of space only an outline is provided here. Let C_i be the child solutions in generation i . By convexity of the operators we know $C_i \subset \Gamma(L_i) \Rightarrow L_{i+1} \subset \Gamma(L_i) \Rightarrow \text{Diam}(L_{i+1}) \leq \text{Diam}(L_i)$ with probability 1.

Since a fixed collection of operators is employed we know $|C_i|$ is bounded and hence $|L_i \cup C_i|$ is bounded. Moreover, strong convexity of recombination operators and boundedness of $|L_i|$ ensures that $\mathbb{P}(\text{Diam}(C_i) < \text{Diam}(L_i))$ is bounded away from zero uniformly in i for which $\text{Diam}(L_i) > 0$. The number of pairs $x, y \in L_i \cup C_i$ s.t. $d(x, y) = \text{Diam}(L_i)$ is therefore bounded and the probability that all such pairs come from L_i (and none from C_i) is bounded away from zero uniformly in i for which $\text{Diam}(L_i) > 0$.

Suppose the number of pairs defining the diameter of the population is bounded above by M . Then since selection is strongly convex we have $\mathbb{P}(\text{Diam}(L_i) < \text{Diam}(L_{i+M})) > 0$ whenever $\text{Diam}(L_i) > 0$. By the finiteness of \mathcal{E} and the boundedness of $|L_i|$ we know $\mathbb{P}(\text{Diam}(L_i) < \text{Diam}(L_{i+M})) > \epsilon$, uniformly in i for which $\text{Diam}(L_i) > 0$, for some $\epsilon > 0$. Since the diameter is integer valued, the result follows. ■

Note that elitist selection, which is guaranteed to keep the fittest individual in the population, is not strongly convex. However, the above proof can easily be modified to allow elitism since it relies on pairs of individuals being separated by selection, rather than specific singletons being removed.

A practical corollary of the above, is that if the same conditions hold, except that only at least one operator taking values in \mathcal{X} is strongly convex, then if the remaining operators taking values in \mathcal{X} are applied in each generation with probabilities converging to zero, the diameter of the population converges to zero in probability. The result follows along similar lines, but relies on the fact that the number of generations between applications of the operators which are not strongly convex increases without bound with probability 1. Therefore between each such application, convergence as in Lemma 3.10 occurs. Moreover, the proportion of the generations between these applications which is required for convergence tends to zero with probability 1. Therefore, the proportion of generations in which the population is not converged tends to zero with probability 1.

4 Conclusions

Genetic algorithms are highly versatile search heuristics with broad applicability in a number of difficult problems and problem types. The effectiveness of GAs has resulted in a number of studies into their theoretical properties. The current work introduces a new theoretical perspective on GAs which is consistent with forma analysis. This new approach utilises the equivalence relation basis of forma analysis to generate a topology on the search space via uniformities. The topology is metrisable by a distance function which coincides with the hamming distance in the canonical binary GA described by [Holland, 1975].

The new framework is extremely general, yet provides in addition to the basic topology, a notion of convexity for the discrete context of GAs. Through convexity the concepts of exploitation and population convergence can be investigated.

We hope that this new perspective will permit the transfer of some existing theory from convex analysis to improve the theory and methodology of GAs in the future.

Acknowledgements

The author acknowledges support from the EPSRC funded EP/H023151/1 STOR-i centre for doctoral training.

References

- [Blumenthal, 1953] Leonard Mascot Blumenthal. *Theory and applications of distance geometry*, volume 347. Oxford, 1953.
- [Goldberg, 1985] David E Goldberg. *Optimal initial population size for binary-coded genetic algorithms*. Clearinghouse for Genetic Algorithms, Department of Engineering Mechanics, University of Alabama, 1985.
- [Goldberg, 1989] David E Goldberg. Genetic algorithms in search, optimization, and machine learning. *Addison Wesley*, 1989.
- [Greene, 2000] William A Greene. A non-linear schema theorem for genetic algorithms. In *GECCO*, pages 189–194. Citeseer, 2000.
- [Holland, 1975] John H Holland. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, MI, 1975.
- [Holland, 1980] John H Holland. Adaptive algorithms for discovering and using general patterns in growing knowledge bases. *International Journal of Policy Analysis and Information Systems*, 4(3):245–268, 1980.
- [Moraglio, 2007] Alberto Moraglio. Towards a geometric unification of evolutionary algorithms. 2007.
- [Nix and Vose, 1992] Allen E Nix and Michael D Vose. Modeling genetic algorithms with markov chains. *Annals of mathematics and artificial intelligence*, 5(1):79–88, 1992.
- [Radcliffe, 1992] Nicholas J Radcliffe. Non-linear genetic representations. In *Parallel Problem Solving from Nature*, volume 2, pages 261–270, 1992.
- [Radcliffe, 1994] Nicholas J Radcliffe. The algebra of genetic algorithms. *Annals of Mathematics and Artificial Intelligence*, 10(4):339–384, 1994.
- [Schmitt, 2001] Lothar M Schmitt. Theory of genetic algorithms. *Theoretical Computer Science*, 259(1):1–61, 2001.
- [Schmitt, 2004] Lothar M Schmitt. Theory of genetic algorithms ii: models for genetic operators over the string-tensor representation of populations and convergence to global optima for arbitrary fitness function under scaling. *Theoretical Computer Science*, 310(1):181–231, 2004.
- [Suzuki, 1995] Joe Suzuki. A markov chain analysis on simple genetic algorithms. *Systems, Man and Cybernetics, IEEE Transactions on*, 25(4):655–659, 1995.
- [Vose, 1991] Michael D Vose. Generalizing the notion of schema in genetic algorithms. *Artificial Intelligence*, 50(3):385–396, 1991.
- [Vose, 1999] Michael D Vose. *The simple genetic algorithm: foundations and theory*, volume 12. MIT press, 1999.