Towards a New Framework for the Analysis of Genetic Algorithms

Contents

1. Introduction ............................................... 2
2. Basic definitions ............................................ 3
3. The model ................................................ 4
   3.1 The partition of populations set. ......................... 4
   3.2 The set of classes is a complete metric space .......... 5
   3.3 The set of classes is a lattice .......................... 6
   3.4 Analysis of selection ................................... 7
   3.5 Analysis of entropy ..................................... 8
4. Remarks .................................................. 10
5. Conclusions and future research ................................. 11
Towards a New Framework for the Analysis of Genetic Algorithms

José Galavíz *  Angel Kuri
Instituto de Matemáticas, UNAM  Centro de Investigación en
Área de la Investigación Científica  Computación, IPN
Circuito Exterior C.U.  Apdo. postal 75-476, Lindavista
04510, D.F. México.  07738, D.F. México.

ABSTRACT

This paper describes a formal framework for the analysis of genetic algorithms. The model is based on the idea that over the space of populations an equivalence relation can be defined, as well as a metric on the space of equivalence classes induced by this relation. On such set of classes a relation of partial order is defined. With this tools it can be proved that selection in a GA causes some kind of convergence and entropy reduction. The model is not restricted to a particular kind of selection.

Keywords: genetic algorithms, Hamming distance, partition, metric space, lattice, convergence, entropy.

1. Introduction

Genetic algorithms are heuristic search methods based on the mechanisms of natural evolution. Since its introduction by Holland in the 1960’s, genetic algorithms (GAs) have motivated several efforts to develop a formal model where they can be explained. Holland himself takes the first step in his book ([7]), and provides the well known schema theorem which was found approaching the GA via the so-called "schemas". However such theorem has recently been subject of much critical discussion, because it is based only in destructive effects of the crossover and mutation operators. Also the framework provided by Holland is restricted to a particular kind of GA with proportional selection, 1-point crossover and uniform mutation (SGA).

Other approaches have been proposed such as those based on statistical mechanics ([14],[17]), and the very successful approach based on Markov chain analysis ([13], [3],[4],[16]), and its derivations related with dynamical systems([18], [19]).
In this paper a different framework is proposed for the analysis of GAs. As a first step towards a comprehensive approach, a restricted GA with selection only is analyzed. The model presented here is more general than others because no particular kind of selection is assumed. Such operator is described as generally as possible.

The theoretical framework is based on a partition of the populations set and the definition of a metric in such space. A partial order is defined on the set of partitions in order to prove some kind of convergence, if the GA operates only by iterated application of selection. The convergence caused by selection has been analyzed as in [11].

With the tools provided by the proposed framework, an analysis of the behavior of general entropy is conducted and we show that, in general, selection forces lead to a decrease of entropy. The information theoretical approach has been presented before as in [12], but in a different way.

In the second section some concepts and notation that will be used in the rest of paper are defined. In the third section the model is described and the fundamental tools are developed, based on metric spaces and the partition of the populations set. The analysis of selection and entropy is then developed. In the fourth section some remarks about particular kinds of selection are presented and finally some conclusions and proposals for future research are made.

2. Basic definitions

In this paper we have used genetic algorithms whose individuals are binary strings or, equivalently, integers encoded in binary, of length \( l \). Every individual in a population is in \( \mathbb{B}_l = \{0, 1, \ldots, 2^l - 1\} \).

A genetic algorithm operates over populations with a finite number of such strings.

**Definition 1** The number of binary strings contained in a population is the population size.

Note that the previous definition does not consider the number of different strings. Only the number of strings, whether repeated or not, in the population.

In the following \( N \) denotes the size of finite population of binary strings. The set of all possible populations of size \( N \) of binary strings of length \( l \) is denoted by \( P_N(\mathbb{B}_l) \) or in short \( P_{N,l} \).

Any population with a finite number of individuals can be described by means of the proportion of every binary string in the population as is presented in [19].

**Definition 2** Let \( f_i \) be the number of times (frequency) that \( i \in \mathbb{B}_l \) appears in a population \( P \in P_{N,l} \). \( P \) is described by a vector
\[ P = (p_0, p_1, \ldots, p_{2^l}) \in [0,1]^{2^l}, \text{ where } p_i = f/N. \] \(P\) is the vector of proportions.

The degree of variation in a population is then described by means of the following.

**Definition 3** The *variation vector* of a population is \( V(P) = (b_0, b_1, \ldots, b_{2^l}) \) where:

\[
\begin{align*}
    b_i &= \begin{cases} 
        1 & \text{if } p_i > 0 \\
        0 & \text{if } p_i = 0
    \end{cases}
\end{align*}
\]

The variation vector has an entry \(i\) equal to 1 iff the genotype \(i\) is represented by at least one instance in the population. Note that the set of all the variation vectors is equivalent to the set of integers \([0,1,\ldots,2^{2^l}-1]\) encoded in binary.

**Definition 4** The *variation coefficient* of the population \(P\), denoted by \(v(P)\) is the number of ones in \(V(P)\). Note that this is the number of different *genotypes* represented in \(P\) by at least one string, that is \(C(P)\).

**Definition 5** Let \(P\) and \(Q\) be two populations in \(P_{N,l}\). The *distance* between \(P\) and \(Q\), denoted by \(H(P,Q)\) is the number of positions where \(V(P)\) and \(V(Q)\) have different values. \(0 \leq H(P,Q) \leq 2^l\).

Note that \(H(P,Q)\) is the traditional Hamming distance as defined in [6]. In what follows \(n = 2^l\).

3 The model

3.1 The partition of populations set.

We now define a partition in the space \(P_{N,l}\). In order to do it an equivalence relation ([15], chapter 2) is defined.

**Theorem 1** \(H(P,Q) = 0\) defines an equivalence relation in \(P_{N,l}\).

Proof: \(P\) is related with \(Q\) (denoted by \(P \sim Q\)) if and only if \(H(P,Q)\).

We must prove the relation is reflexive, symmetric and transitive:

- \(P \sim P\), because the number of positions where \(V(P)\) differs from \(V(Q)\) is zero. Then \(H(P,P) = 0\)
- if \(P \sim Q\) then the number of positions where \(V(P)\) differs from \(V(Q)\) is zero then \(H(P,Q) = H(Q,P) = 0\). Therefore \(Q \sim P\)
- if \(P \sim Q\) and \(Q \sim R\) then \(H(P,Q) = H(Q,R) = 0\), the number of positions where \(V(P)\) is different from \(V(Q)\) and the number of
positions where \( V(Q) \) and \( V(R) \) are different is zero. Therefore \( V(P) = V(Q) = V(R) \), then \( H(P, R) = 0 \). That is \( P \sim R \).

The relation defined above induces a partition on \( P_{N,l} \), the space where it is defined. Every equivalence class is identified by means of the variation vector of any population in the class, given that such variation vector is the same for every population in that class. Then, definition 3 can be extended to the set of classes and we may talk about the variation vector of a class rather than the variation vector of a particular population only. In what follows \( C_i \) denotes the set of all the populations classes of strings of length \( l \).

**Definition 6** Let \( i_2 \) be the nonnegative integer \( i \) encoded in binary. The class \( C_i \in C_i \) is:

\[
C_i = \{ P \in P_{N,l} \mid V(P) = (b_0, b_1, \ldots, b_{n-1}) \text{ with } i_2 = b_0 b_1 \ldots b_{n-1} \}
\]

And

\[
V(C_i) = (b_0, b_1, \ldots, b_{n-1})
\]

is the variation vector of the class \( C_i \).

**Definition 7** The number of ones in the variation vector of the class \( C_i \in C_i \) is the variation coefficient of \( C_i \) denoted by \( v(C_i) \).

### 3.2 The set of classes is a complete metric space

*Hence, two populations* \( P \) and \( Q \) are in the same equivalence class iff they have the same genotypes. \( C_i \) denotes the class of all the populations whose variation vector corresponds to the integer \( i \) encoded in binary \( H \) now will be redefined.

Definition 8 \( H(C_i, C_j) \) is the number of positions where \( V(C_i) \) is the number of positions where \( V(C_j) \) and \( V(C_i) \) have different values.

**Theorem 2** \( H \) is a metric for \( C_i \).

*Proof*: Let \( C_i \) and \( C_j \) be two classes of populations. \( H \) is the Hamming distance between two class variation vectors \( V(C_i) \) and \( V(C_j) \), and \( H \) is a true metric as proved in [1]. Then \( H \) is a metric for \( C_i \).

Now we define a Cauchy sequence in the metric space \( (C_i, H) \) as in [15] and [5] (chapter 3).

**Definition 9** A sequence \{\( C \}\} in \( C_i \) is a Cauchy sequence iff for any
real number $\varepsilon > 0$ there exists a nonnegative integer $N$ such that for every $k,m > N$:

$$H(C^k, C^m) < \varepsilon$$

The previous definition provides the elements needed to demonstrate the following theorem. The definition of a complete metric space can be found in [15] and (chapter 3).

**Theorem 3** $(C, H)$ is a complete and bounded metric space.

*Proof:* In order to demonstrate completeness we need to show that any Cauchy sequence in $(C, H)$ converges.

Let $(C)^\prime$ be a Cauchy sequence, then for any $\varepsilon > 0$ there exists an integer $N$ such that for every $k,m > N$:

$$H(C^k, C^m) < \varepsilon$$

In particular for

$$\varepsilon = 1$$

There exists $N$ such that for every $k,m > N$:

$$H(C^k, C^m) < 1$$

This implies that $H(C^k, C^m) = 0$ because this is the only possible value for $H$ less than 1, then $C^k$ and $C^m$ are the same string, and this string is in $C_i$.

Since $H$ can only take values in $\{0,1,...,2^l\}$ then $H$ is a bounded metric.

**3.3 The set of classes is a lattice**

The following defines a relation of partial ordering ([10], [8]) in $C_i$.

**Definition 10** Let $C_i$ and $C_j$ be two populations in $C_i$, $C_i$ is a *reduction* of $C_j$ ($C_i \preceq C_j$) iff in every position where $V(C_i)$ has 1, $V(C_j)$ also has 1. That is, any class $C_i$ is a reduction of other class $C_j$ iff all the genotypes represented in $C_i$ are also represented in $C_j$.

With the previous definition $C_i$ constitutes a partially ordered set (poset). In fact the set is a lattice (any two classes are reduction of some other class and a class can be found that is reduction of both classes), the supremum element is $(1,...,1)$ and infimum element is $(0,...,0)$. An alternative lattice definition can be found in [10] (ch. 1). A poset $(L, \preceq)$ is a lattice iff the supremum and infimum exist for any
finite non void subset $S \subseteq L$.

Using the previously demonstrated theorem and the lattice definition mentioned above it is possible to prove the following theorem.

**Theorem 4** Let $\{C\}^i$ be an infinite sequence of elements in $\mathbb{C}_i$ such that

$$C^i \leq C^{i+1}$$

(monotonically decreasing), then $\{C\}^i$ converges to some class $l \in \mathbb{C}_i$ where $I = \inf\{C\}^i$

**Proof:** $\mathbb{C}_i$ has exactly $2^i$ elements. Hence, $\{C\}^i$ has a finite number of different terms $\Omega \subseteq \mathbb{C}_i$. By the lattice definition, $\Omega$ has an infimum element $I$ such that:

$$I \leq C \quad \forall C \in \Omega \quad \text{(3.1)}$$

Let $M$ be the smaller integer such that $C^M \rightarrow I$. Given that $\{C\}^i$ is monotonically decreasing then for every $k \geq M$ we have that $C^k = I$. Therefore, for every $\varepsilon > 0$ there exists an $N = M$ such that for every $k, r > N$

$$H(C^k, C^r) = 0 < \varepsilon$$

Consequently $\{C\}^i$ is a Cauchy sequence and converges from theorem 3.

Also, for every $\varepsilon > 0$ if $r > M$ then

$$H(I, C^r) = 0 < \varepsilon$$

and $\{C\}^i$ converges to $I$.

**3.4 Analysis of selection**

In this subsection it is shown that a genetic algorithm which operates only via selection converges. Similar result has been demonstrated before [11], using other techniques.

The algorithm converges to a class of populations rather than to a particular population. Every generation of a GA belongs to a class. Looking at the evolution of the algorithm from the class point of view, it is possible to see that such evolution converges to some particular class.

In general terms the meaning of selection can be defined as follows.
**Definition 11** Let \( P_i \in \mathbb{P}_{N,t} \) be a population in a class \( C_k \in \mathbb{C}_t \). *Selection* is a function \( S: \mathbb{P}_{N,t} \rightarrow \mathbb{P}_{N,t} \) such that:

\[
S(P_i) = P_{i+1}
\]

where \( P_{i+1} \in C_r \) and \( C_r \leq C_k \).

That is: selection can not increase the number of genotypes represented in any population. The result of applying selection to any population is other population with, at most, the same set of genotypes.

**Definition 12** Let \( \Phi: \mathbb{P}_{N,t} \rightarrow \mathbb{C}_t \) be a function that assigns to every population \( \mathbb{P} \) the class where \( \mathbb{P} \) lives. That is, if \( \mathbb{P} \in C_k \) then \( \Phi(\mathbb{P}) = C_k \).

The iterated application of selection in any population induces a monotonously decreasing sequence (as defined in [15], chapter 3) of the classes of such populations. Given that the set of classes with \( H \) is a complete metric space, such sequence converges in \( \mathbb{C}_t \). \( S^k(\mathbb{P}) \) denotes the k-th iterated application of selection to the population \( \mathbb{P} \). Now the set of tools needed to prove following theorem is complete, analogous to the respective theorem for \( \mathbb{R} \) (theo. 3.14 [15]):

**Theorem 5** Let \( \mathbb{P}_0 \in \mathbb{P}_{N,t} \) be an initial population. The sequence:

\[
\Phi(S^k(\mathbb{P}_0)) \leq \ldots \leq \Phi(S^{2}(\mathbb{P}_0)) \leq \Phi(S(\mathbb{P}_0)) \leq \Phi(\mathbb{P}_0)
\]

converges to some class \( C \in \mathbb{C}_t \).

**Proof:** It, follows from 4.

### 3.5 Analysis of entropy

With the convergence result just proved it is possible to analyze the behavior of entropy in the generic evolution.

**Definition 13** Let \( \mathbb{P} = (p_0, p_1, \ldots, p_{n-1}) \) be a population in \( \mathbb{P}_{N,t} \). The entropy of \( \mathbb{P} \), denoted by \( H(\mathbb{P}) \) is:

\[
H(\mathbb{P}) = \sum_{i=0}^{n-1} p_i \log_2 \left( \frac{1}{p_i} \right)
\]

The entropy of a population with only one genotype is 0 because such genotype \((i)\) has \( p_i = 1 \) and \( p_j = 0 \) for every \( j \neq i \).

In a population of size \( N \), with \( k \leq n \) equally probable genotypes, the probability of each genotype is:
The entropy of a population with the previous condition is maximum:

\[ H(P) = k \frac{1}{k} \log_2 (k) = \log_2 (k) \quad (3.2) \]

Then the greatest entropy of any population in \( P_{N,j} \) is \( l \) because the total amount of possible genotypes is \( 2^l \). In summary:

\[ 0 \leq H(P) \leq l \quad \forall \ P \in P_{N,j} \]

**Theorem 6** Let \( P \in P_{N,j} \) be a population Whose variation coefficient is \( v(P) \). Then:

\[ H(P) \leq \log_2 (v(P)) \]

**Proof**: The entropy is maximum when all the genotypes are equi-probable, applying the equation 3.2 with \( k = v(P) \) the desired result is obtained.

It is useful to extend the definition of population entropy to define a class entropy in \( C_i \).

**Definition 14** Let \( C \in C_i \) be a class of populations. The entropy of \( C \) is defined as:

\[ H(C) = \max \{ H(P) \mid P \in C \} \]

From this definition and theorem 6 the following corollary can be formulated:

**Corollary 7** \( H(C) = \log_2 (v(C)) \).

**Proof**: \( H(C) = \max \{ H(P) \mid P \in C \} \)

Every population \( P \in C \) has the same variation coefficient. Then, from theorem 6:

\[ H(C) = \log_2 (v(P)) \quad \forall \ P \in C \]
and the variation coefficient of every population in \( C \) is the variation coefficient of \( C \) itself.

**Corollary 8** Let \( P, Q \in P_{n, l} \) be two populations such that:

\[
S(P) = Q
\]

then:

\[
H(\Phi(Q)) \leq H(\Phi(P))
\]

**Proof:** Since (def. 11)

\[
\Phi(Q) \preceq \Phi(P)
\]

then the number of ones in \( V(Q) \) is less or equal that the number of ones of \( V(P) \) and:

\[
\log_2(v(\Phi(Q))) \leq \log_2(v(\Phi(P)))
\]

From the previous corollary:

\[
H(\Phi(Q)) = \log_2(v(\Phi(Q)))
\]

and

\[
H(\Phi(P)) = \log_2(v(\Phi(P)))
\]

So:

\[
H(\Phi(Q)) \leq H(\Phi(P))
\]

From which:

**Theorem 9** Let \( P_0 \in P_{n, l} \) be an initial population. The sequence:

\[
\{H(\Phi(S^k(P_0)))\}
\]

is monotonically decreasing, bounded and convergent.

**Proof:** The monotony and decreasing features are obvious from corollary 8. The lower bound is 0 and the upper bound is the entropy of the class where the initial population is. This sequence is in \( \mathbb{R} \), which is a complete metric space, thus it converges to some real number.

4. **Remarks**

It is now possible to analyze some less general kinds of selection. In [2] several classifications for the selection operators are presented. The
first classification divides selection operators in dynamic and static. In the second these are divided in extinctive and preservative. In the third one, selection with elitism is considered versus pure selection. Finally, generational versus steady state selection is considered. Here the focus is in the first two of such classifications.

If genotype-static selection is considered (every genotype has a generation independent probability of selection) then, after the first step the variation vector of the population remains constant and the entropy of the class where every generation "lives" is also constant. Every generation will have the same genotypes as its precedent generation.

If genotype-preservative selection is considered (every genotype in the population has a non-zero selection probability) then, as in the case discussed before, the class where every generation "lives" is the same and has a fixed entropy.

Although if a genotype-extinctive selection strategy is considered (in every generation at least one genotype disappears), then the entropy of the class where each generation "lives" tends to zero because the number of genotypes represented in population is monotonously decreasing in a strict sense. In this case there is a unique genotype present in every generation. That is, the intersection of all the generations is only one genotype (g). Since in every generation one genotype is destroyed, if we assume that the population is uniformly distributed over the genotypes that still live then the proportion of genotype g at time t is:

\[ p_{g,t} = \frac{1}{v(P_0) - t} \]

if

\[ 0 < t < v(P_0) \]

where \( P_0 \) is the initial population.

This may be rewritten as:

\[ p_{g,t} = \frac{1}{2^H(P_0)} \] (4.4)

Equations 4.3 and 4.4 are conceptually analogous to the equation \((1 + c)^t\) that appears in [9] and explains the exponential sampling of "better" schemas (those which have a fitness \( f + fc \) where \( f \) is the mean value of fitness in the generation and \( c \) is a positive constant).

5. Conclusions and future research

The framework shown here is a new and useful approach for the
analysis of GAs. With the tools provided by such a framework some features of selection have been proved. Also the analysis is more general than the one reached by the use of the other methods already discussed.

An essential feature of selection that is present in all the different kinds of this operator has been isolated: selection cannot increase the number of different genotypes present in a population. This is a sufficient condition that assures the convergence of a GA with selection only.

Also, the behavior of entropy has been analyzed along with the iterated application of selection to an initial population. It can be concluded that, from a global point of view, selection reduces entropy. Entropy can increase from one generation to the next but, in the long run, entropy will decrease and converge to a limit as close to zero as small is the number of different genotypes present in populations.

To demonstrate the usefulness of this framework it remains to apply it to all the other common genetic operators: mutation and crossover. Currently only some intuitive features have been proved. But is possible to speculate that if this method is applied to mutation, for example, the result will be such that, in the limit, the sequence of classes where the generations are, will tend to be in a neighborhood of some other class. The radius of such neighborhood will be greater in direct proportion to the probability of mutation and will also be affected by the distribution of genotypes present in the population.

The behavior of GAs is the result of the struggle two opposite forces: the contractive force of selection and the expansive one of the mutation. If we consider the effect of mutation on each genotype the operator can be expressed as a stochastic matrix that corresponds to an ergodic Markov chain whose invariant distribution vector is the uniformly distributed population. From the point of view purported in this paper, mutation tends to maximize the entropy of population.

We still need to find some parameters that characterize the behavior of GAs. Such parameters must reflect the conflict mentioned above and must also measure the degree of instability, the size of attractors and the chaotic potential of a GA.

References


