ON THE EXPECTED CONVERGENCE TIME OF A GENETIC ALGORITHM WITH MINIMUM PARAMETERS

MARGARITA REYES-SIERRA AND CARLOS A. COELLO COELLO
Evolutionary Computation Group at CINVESTAV-IPN (EVOCINV)
Departamento de Ing. Eléctrica Sección Computación
Av. IPN No. 2508
Col. San Pedro Zacatenco
México, D.F. 07300, MÉXICO
mreyes@computacion.cs.cinvestav.mx
ccoello@cs.cinvestav.mx

ABSTRACT
In general, the problem of characterizing the behavior of a genetic algorithm is rather complex because it varies depending on the application domain and on the parameters adopted for the genetic algorithm. In this paper, we briefly discuss some of the most representative approaches that have been previously proposed in the literature to estimate convergence times of a genetic algorithm. Then, we introduce our own approach whose analysis is based on Markov Chains. Our model is empirically validated using a genetic algorithm with minimum parameters. Our results indicate the correctness of the proposed model.

INTRODUCTION

Genetic algorithms (GAs) are a heuristic search technique inspired on natural evolution (i.e., the survival of the fittest). In its origins, the genetic algorithm (now called "classical") was applied to single-objective optimization problems (Goldberg,1989). Rudolph proved in the mid-1990s convergence of this simple GA to the global optimum of a given function, under certain conditions (Rudolph, 1994). This work gave some of the desired theoretical foundations to the behavior of GAs. This paper extends Rudolph's model so that we can estimate the expected convergence time of a genetic algorithm with minimum parameters. We also detail the general features of the transition matrix of the elitist genetic algorithm discussed in (Rudolph, 1994). The results of our theoretical model are validated with a simple example. Our empirical results indicate the validity of our theoretical model.

PREVIOUS WORK

Ankenbrandt (1991) obtained a bound for the execution time to achive convergence of a genetic algorithm using a model based on allele convergence and a relatively simple proof by induction. On the other hand, taking advantage of the fact that natural selection uses the diversity of a population to give rise to adaptation, Louis & Rawlins (Louis, 1993) developed a model based on Hamming distances between individuals in a population. They argued that, if we ignore the effects of mutation, the lack of diversity makes useless the role of natural selection. Thus, they used a measure of diversity to estimate an upper

bound on the time beyond which progress is unlikely (because of the lack of diversity). (Goldberg, 1987) presented a study similar to the work of (Louis, 1993). In this case, they estimated the convergence time of a genetic algorithm whose population consists of individuals of length 1 using a Markov chain model. Note that in the current paper we use a population consisting of individuals of length n.

MARKOV CHAINS

The basic concepts regarding Markov chains can be consulted in (Reyes, 2003). Next, we will discuss a special type of Markov chain on which the current work is based. The material of this section was obtained from (Rudolph 2000a, Iosifescu 1980).

Absorbing Chains

Consider a Markov chain with finite state space $S \neq \emptyset$. In this section the states of a Markov chain are classified based on whether it is possible to go from a given state to another state.

Definition 1. We say that the state i **leads to** state j and write $i \to j$ if and only if $p_{ij}^{k}>0$ for an $k>1$. If $i \to j$ and $j \to i$ we say that state i **communicates with** state j and write $i \leftrightarrow j$.

Beginning with the previous definition, the states are classified into "equivalence classes". Two states are in the same class if they are "communicated", i.e. if the process can go from one state to the another and viceversa. The equivalence classes are classified as **ergodic** sets (so called **recurrent**) or **transient** sets.This way, the corresponding states in those classes are called ergodic states and transient states, respectively. For each finite Markov chain there must be always at least one ergodic set; however, it is unnecessary to have transient sets. Once a chain leaves a transient set, it can never come back to it; conversely, once a chain enters into an ergodic set, it can never leave it. In particular, if an ergodic set contains just one state, this state is called an **absorbing** state, because once into it, the Markov chain will stay there forever. From the previous, we have that a state i is absorbing if and only if $p_{ii}=1$ (Kemeny, 1960). Now, a chain with transient states has a behavior such that it moves towards ergodic sets. The probability that the process lies within an ergodic set tends to 1. When all the ergodic sets are unitary, these chains are called **absorbing chains**, because they will eventually get trapped in an absorbing state. It is important to consider the canonical form of the transition matrix of a Markov chain. Let's assume that we have s transient states and r-s ergodic states, and that we cluster all the transient sets and all the ergodic sets together, the resulting form is:

$$P = \begin{pmatrix} S & O \\ R & Q \end{pmatrix}$$

The region O consists completely of zeros. The matrix $Q_{s \times s}$ represents the chain while it is in transient states, the matrix $R_{s \times r-s}$ represents the transition from the transient states to ergodic states and the matrix $S_{r-s \times r-s}$ represents the chain once that it is into an ergodic state.If we consider an absorbing chain, we have that by definition $S=I_{r-s \times r-s}$, so its canonical form is:

$$P = \begin{pmatrix} I & O \\ R & Q \end{pmatrix}$$

Definition 2.For an absorbing Markov chain we define the **fundamental matrix** to be $N = (I - Q)^{-1}$.

Definition 3. We define $\mathbf{n}_j$ to be a function whose value is the total number of times that the process is in a transient state s_j

Let **T** the set of transient states of the Markov chain. If we denote with $\mathbf{E}_i[\mathbf{n}_j]$ the expected value of $\mathbf{n}_j$ assuming that the process starts in the state s_i, we have the next result:

Theorem 1.(Kemeny, 1960) $\{\mathbf{E}_i[\mathbf{n}_j]\}=N$

Definition 4. Let **t** be a function whose value is given by the number of steps (including the initial state) in wich the process is in a transient state.

If the process starts in an ergodic state then **t**=0. If the process starts in a transient state, then **t** gives us the total number of necessary steps for reaching an ergodic state. In an absorbing chain, this is the *time to absorption*.

Let x a column vector with all entries equal to 1.

Theorem 2.(Kemeny, 1960){ $\mathbf{E}_i[\mathbf{t}]\}=N$x.

EXPECTED CONVERGENCE TIME

In (Rudolph, 1994), we can study the mathematical models of a Simple Genetic Algorithm (SGA) and an Elitist Genetic Algorithm (EGA). It is required to consult this reference to get the preliminaries required to understand the work presented here. In the previous section, we showed some results on the fundamental matrix of a Markov chain. As we saw, such a matrix can be used to compute the expected convergence time of the chain. In this section, we will apply such results to the corresponding transition matrix of the EGA. Since in Rudolph's model the matrix **P** corresponds with the populations whose super individual is the global optimum, we can consider that when the chain is in one of those states, the search process has finished. Any further change in the population can be ignored because the super individual will be no longer modified. Therefore, we can rewrite the matrix like:

$$P^{+} = \begin{pmatrix} I & O \\ R & T \end{pmatrix}$$

We can see clearly now that the Markov chain corresponding to the EGA is absorbent. According to Definition 2, the fundamental matrix that interests us in this case is: $N=(\mathrm{I}-\mathbf{T})^{-1}$. Since our objective is to know the fundamental matrix *N*, we will start by studying the structure of the block **T**.

Matrix P

In this section, we will show the elements of the **P** matrix. As we know, this matrix is the result of the product: **P=CMS**. This is the reason why the elements of each of the corresponding matrices will be specified next. The details of the previous expressions may be found in (Reyes, 2003).

Elements of the matrix of crossover. The elements of uniform crossover were modeled. For that sake, the Å operator was defined as the negation of an *or-exclusive*. Using this fact the following formula was developed (assuming a crossover percentage of 0.5):

$$c_{ij} = \prod_{q=1}^{n}(0.5)^{l}\prod_{r=1}^{l}\sum_{s=f(q)-1}^{f(q)} p_s^r(i) \oplus p_q^r(j)$$

where $f(q) = 2\left\lfloor \frac{q+1}{2} \right\rfloor$

Elements of the Mutation Matrix. The corresponding elements are:
$m_{ij}=p_m^{Hij}(1-p_m^{N-Hij})$where p_m is the mutation probability and H_{ij} is the Hamming distance between populations *i* and *j.*
Elements of the Selection Matrix.

The selection operator adopted is the well-known proportional selection:

$$s_{ij} = \begin{cases} \frac{\prod_{k=1}^{n} f(p_k(j))}{(\sum_{k=1}^{n} f(p_k(i)))^n} & p_k(j) \in \{p_k(i) \mid r = 1,..., n\} \forall k = 1,..., n. \\ 0 & otherwise. \end{cases}$$

where *f* is the objective function (fitness).
Elements of the **P** *Matrix*. Since the matrix **P** is the product of the crossover, mutation and selection matrices, we have that:

$$p_{ij} = \sum_{p=1}^{n}(\sum_{q=1}^{n} c_{ij} m_{qp}) s_{pj}$$

such that, given the previous expressions for each of the operators, it is possible to compute all the elements p_{ij}.

Matrix E
From (Rudolph, 1994) we conclude that this matrix has exactly one 1 per row. Given a fixed population is clear that its super individual can only improve. Then, $\mathbf{E}_{11}$ is an identity matrix and the matrices $\mathbf{E}_{aa}$ (a>2) are identity matrices with some zeros in the diagonal. It should also be clear that the **E** matrix has columns of zeros: those corresponding to populations whose super individual is not the best individual contained within. Conversely, in those columns corresponding to populations whose super individual is the best in the population, it is possible to find more than one 1. Then, as we descend throughout the rows of blocks, the populations get distributed along the row, depending on the quality of their maximum individual. We will assume that within each block the populations are sorted (by sets within which order is irrelevant) based on the quality of their maximum individual.

individual	fitness	rank
00	0.5	4
01	1.0	3
10	1.5	2
11	2.0	1

Table 1: The four possible individuals.

Experiments
Further details can be consulted in (Reyes, 2003). Let us consider the simplest case that has become commonly associated with the basic (minimum) conditions of a GA. Let *l*=2 and *n*=2. In this case, each population consists of 2 individuals and each individual is of length 2. This gives us a total of $2^{2^{\prime}2}=2^4=16$ populations to consider by the GA. With the aim of being able to rank the four individuals, we defined the following function: $f(x_1x_2)= x_1+0.5x_2+0.5$. As the

function f is strictly positive for any individual, this same function was used as fitness function. In Table 1, we show the four possible individuals, their corresponding fitnesses and their rank. According to the model described in a previous section, the fundamental matrix N was obtained. The package MATHEMATICA 4.0 was used to carry out the computations corresponding to the fundamental matrix N, where $N = (I - \mathbf{T})^{-1}$. On the other hand, a GA with the conditions imposed by the model was run. The fixed parameters were: population size = 2, chromosome length = 2 and crossover probability = 1.0. Thus, the results only depend on the mutation probability (p_m). Note that the super individual (elitist individual) does not have to take part in the evolutionary process. Let g be the random variable whose value is the number of necessary iterations for the convergence of the GA. Next, we show the expected value of the variable g (E[g]) and the corresponding standard deviation (D[g]) obtained (using MATHEMATICA 4.0) by the theoretical model (TM) developed and the results obtained by the GA. In our experiments, we performed 100 runs with different random seeds. The results are shown in Table 2. In Figure 1, we show the graph of the values obtained by both methods.

p_m	GA		TM	
	E[g]	D[g]	E[g]	D[g]
0.001	345.05	540.001	514.137	660.781
0.005	90.98	131.461	103.782	132.47
0.01	40.89	54.019	52.491	66.433
0.03	13.0	16.836	18.307	22.415
0.07	5.08	7.036	8.561	9.852
0.1	4.69	6.5	6.380	7.03
0.2	2.69	3.47	3.872	3.780
0.5	1.32	1.847	2.527	1.964

Table 2: Results.

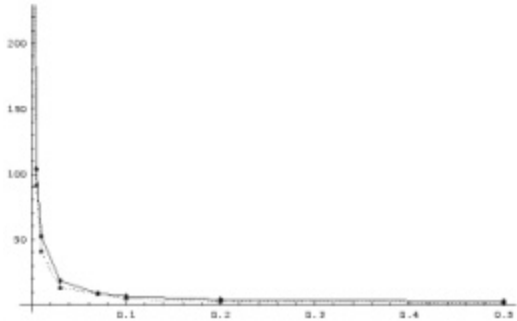

Figure 1:Graph of the values obtained by the GA and the theoretical model developed for the variable g, in terms of the probability of mutation. The continuous line corresponds with the values generated by the theoretical model and the dotted line corresponds to the values generated by the GA.

CONCLUSIONS

The current models to estimate the convergence time of a GA are quite simple and, therefore, very distant from the observed behavior of such type of

algorithm. We developed a mechanism to estimate the convergence time of a GA by means of a model based on Markov chains. Our results for the most simple case that can be considered (perhaps argueably) as realistic, led us to conclude that the proposed model is correct. Nevertheless, it is clear that in such case it was relatively easy to obtain the corresponding matrix **P**, but this is a process that will generally get more complicated as we increase the size of the population and the chromosomic length. In general, the size of the matrix has an exponential growth. Therefore, we can conclude that, from a practical point of view, Markov chains are not a recommended theoretical tool for this sort of analysis (i.e., estimation of expected convergence time). Nevertheless, other alternatives exist to which it would be possible to resort (e.g., statistical mechanics (Prugel, 1994) and geometric interpretation approaches (Vose, 1991)). Another possible alternative is to focus the analysis in a different way. For example, in (Coffey, 1999), the author presents a derivation of an upper bound on the convergence of an elitist genetic algorithm which is based on the largest eigenvalue of the corresponding transition matrix. This author also presents an analysis in which he proves that GA convergence can be accelerated when its corresponding Markov chain is of the type called "rapidly mixing chains".

ACKNOWLEDGEMENTS

The first author acknowledges support from CONACyT through a scholarship to pursue graduate studies at the Computer Science Section of the Electrical Engineering Department at CINVESTAV-IPN. The second author gratefully acknowledges support from CONACyT through project 34201-A.

REFERENCES

Ankenbrandt, C. 1991. An extension to the theory of convergence and a proof of the time complexity of genetic algorithms. In *FOGAs*, pages 53–68. Morgan Kaufmann Publishers.

Coffey, S. 1999. An Applied Probabilist's Guide to Genetic Algorithms. Master's thesis, Department of Mathematics, University of Dublin, Trinity College, Dublin, Ireland, September.

Goldberg D. E. and Segrest, P. 1987. Finite Markov chains analysis of genetic algorithms. In *Proceedings of the Second ICGA*, pages 1—8. Lawrence Erlbaum Associates.

Goldberg D. E. 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Publishing Co., Reading, Massachusetts.

Iosifescu M. 1980. *Finite Markov Processes and Their Applications*. Wiley, Chichester.

Kemeny J. G. and Snell J. L. 1960. *Finite Markov Chains*. D. Van Nostrand Company, Inc., Princeton, New Jersey.

Louis S. J.and Rawlins G. 1993. Syntactic analysis of convergence in genetic algorithms. In L. Darrell Whitley, editor, *FOGA*, pages 141– 151, Morgan Kaufmann Publishers.

Prügel-Bennett A. and Shapiro J. L. 1994. An analysis of genetic algorithms using statistical mechanics. *Physical Review Letters*, 72(9):1305–1309.

Rudolph G. 1994. Convergence properties of canonical genetic algorithms. *IEEE Transactions on Neural Networks*, 1(5):96–101.

Rudolph G. and Agapie A. 2000. Convergence properties of some multi-objective evolutionary algorithms. In *Proceedings of the CEC'2000*, pp. 1010–1016, Piscataway, New Jersey, IEEE.

Reyes M. and Coello C. 2003. On the Study of some theoretical aspects of genetic algorithms. Technical Report 02-2003 (available at http://delta.cs.cinvestav.mx/~ccoello/2003.html).

Vose M. D. 1991. Modeling simple genetic algorithms. In L. D. Whitley, editor, *FOGA 2*, pages 63–73, San Mateo California. Morgan Kaufmann Publishers