

# A dynamical model of the Simple Genetic Algorithm

ABDERRAMÁN J. C., JIMÉNEZ J. A., CUESTA P. D. and WINTER G.  
 CEANI - Departamento de Matemática Aplicada  
 Universidad de Las Palmas de Gran Canaria - ULPGC  
 Parque Científico y Tecnológico (Edificio Polivalente, 2º Planta)  
 Campus de Tafira Baja – 35017 Las Palmas – Canary Islands  
 SPAIN

*Abstract:* - A dynamical model of the Simple Genetic Algorithm, SGA, is developed. A practical improvement is introduced, related to the fundamental theorem of natural selection, Fisher's FTNS, that induces correction factors in the [Q] matrix of Nix and Vose Markov model, NV. Now, the expected waiting times, EWT, is computed easily in any situation. For the case of maximum uncertainty, it is obtained an expression for EWT that improves from the Nix and Vose model in relation to the experimental data and coincides in practice with the maximum entropy value,  $H^*$ , of the populational search space. Also, the steady state is obtained when the absorbing state, global optimum, is reached. The new absorbing model results thus integrated in more general stochastic models, as Evolutionary Algorithms, EAs, and Optimum Seeking Devices.

*Key-Words:* - Evolutionary and Genetic Algorithms, Markov Chains, Convergence, Expected Waiting Times.

## 1 Introduction

A problem related to the efficiency of SGA, is the execution time; an adequate stop criterion does not exist and it is required to find practical solutions, such as fixing a number of generations increased to be sure that the algorithm converges towards an acceptable value. Several works try this stop criterion and develop methods to obtain a suitable criterion, most of them from stochastic analysis of NV Markov model of SGA, [1]. The most known is the EWT, that calculates the average passages of time necessary to obtain a copy of the global optimum string in the population, at generation  $t_{EWT}$ , if we start from an initial population given at time  $t_0$ .

In actual cases the calculation of the EWT implies to work with equations systems of high order. Many factorials of large numbers take place, that make impossible the computational stochastic analysis.

In the NV model, when the mutation rate  $\mu = 0.5$ , the matrix elements  $q_{ij}$  only depend on the chromosomes distribution in state  $j$ . In small search spaces there is a difference respect to the value  $(\text{Card}\{Z\})^{-1}$ , the inverse of the size of all possible populations. Already in moderated spaces and especially in typical spaces used in GAs this difference is petty and tends to zero while the search space increases. It is the maximum uncertainty case.

Recently, [2], we have obtained the following result

**Proposition:** *The expected waiting times, EWT, for the case of maximum uncertainty with random initial state, from the NV Markov model of the Simple Genetic Algorithm, SGA, is*

$$EWT(J) = \frac{r+n-1}{n} \left[ 1 - \frac{1}{r^n} \sum_{i=1}^n \binom{n}{i} (r-i)^{(n-i)} \right]$$

Furthermore, in practice,  $n \ll r$ , this expression is reduced to:

$$EWT(J) \approx \frac{r-1}{n} - \frac{n-1}{r} \quad (1)$$

Being  $n$  the population size and  $r$  the cardinal of the search space,  $r = 2^l$ ,  $l$  the chromosome or string length. We compare (1) with the time for the convergence value in the Monte Carlo method,  $EWT_{MC}$  with respect to the values of the argument, search domain:

$$EWT_{MC} = \frac{\mu_n(\text{searchspace})}{\hat{1}_n(\text{populationsample})} - \frac{r}{n} \quad (2)$$

Where  $\mu_n$  is equivalent to the Lebesgue measure of the sets search space and sample, both discretas. Recalling that  $n \ll r$ , it is easy to observe that  $EWT(J) \approx EWT_{MC}$  [2]. Values for the intermediate case of chromosome length  $l = 50$  and  $n$  variable are shown.

Population size, n	EWT(J)
10	$1.13 \cdot 10^{14}$
20	$5.63 \cdot 10^{13}$
30	$3.75 \cdot 10^{13}$
40	$3.81 \cdot 10^{13}$
50	$2.25 \cdot 10^{13}$
100	$1.13 \cdot 10^{13}$
200	$5.63 \cdot 10^{12}$

Table 1. EWT(J) in Nix & Vose model of SGA for maximum uncertainty case,  $l = 50$ ,  $n$  variable.

The theoretical values obtained result too large for its practical usefulness. With a manual or random search we would obtain, in average, the same results, as it is demonstrated above with the equivalence between  $EWT(J)$  and  $EWT_{MC}$ . However, the experience shows that SGA tends to give a response adapted in a number of generations in order of magnitude much less. In this situation, the only one reasonable response will be to admit that the NV Markov model, even though it is theoretically coherent, seems to be too general to model correctly the optimum performance of the SGA.

## 1.1 Mathematical Backgrounds.

In addition to the previous negative results, qualitative differences exist between NV model of SGA and the absorbing models of EAs and the Optimum Seeking Devices. These differences would not have to be presented, since the SGA is included in EAs, and both belong to Optimum Seeking Devices. These three models are introduced briefly. In 1.1.4 is presented the Fisher's FTNS Theorem, which will serve to transform NV Markov model in an absorbing Markov model, which presents more adequate characteristics.

### 1.1.1 NV Markov model of SGA

The NV dynamical model is employed to represent a SGA, consisting of a finite population, a standard representation of binary symbols, 1-point crossover, uniform mutation, and proportional selection. Elitism and other aspects used in Genetic Algorithms are not considered. Nix and Vose define two matrix operators,  $[F]$  and  $[M]$ , being  $[F]$  the fitness operator, which is determined by the fitness function. The matrix  $[M]$  is the recombination operator, depends from the strings recombination operator  $[R]$ , from  $[Z]$ , and from the permutation operator  $\sigma_i$ .

With these operators, the exact transition probabilities between states are calculated, [1]:

$$Q_{i,j} = n! \prod_{y=1}^r \frac{\left\{ M \left( \frac{Ff_i}{|Ff_i|} \right)_y \right\}^{z_{j,y}}}{z_{j,y}!} \quad (3)$$

From these transition probabilities a recurrent and aperiodic Markov chain is obtained, where there is a probability of "jump" of a state to any other. The evolution of the system tends towards a steady state, where the probability of arriving to a given state is the same, independently of the probability distribution for the initial state. In the steady state the probabilities of reaching any population are all positive, therefore, the system can ends in any population.

### 1.1.2 Convergence Theory of Evolutionary Algorithm

The SGA form part of a widest class of algorithms, the EAs. Nevertheless, the convergence theory of EAs presents a different qualitative behaviour, [3].

Presuming the use of elitist selection, incorporating a heuristic to always maintain the best solution in the population, one-point crossover and uniform mutation, form an equivalent class of all states that contain the best string for unimodal case. Then all populational states containing the best string, global optimum, are characterized as the same state, absorbing state, the Markov chain may then be written in the form:

$$[P] = \begin{array}{c|c} \mathbf{I} & \mathbf{0} \\ \hline \mathbf{R} & \mathbf{Q} \end{array}$$

Where  $[P]$  is the transition matrix,  $\mathbf{I}$  is a  $1 \times 1$  identity matrix describing the absorbing state,  $\mathbf{R}$  is strictly positive  $t \times t$  transition submatrix,  $\mathbf{Q}$  is a  $t \times t$  transition submatrix of all transient states, and  $t$  is a positive integer.

Asymptotic global convergence of the evolutionary algorithm is then transparent as every absorbing chain will reach an absorbing state; in this case, is just one such state.

**Theorem:** Given a  $(Z-J+1 \times Z-J+1)$  state transition matrix, reducible, where there is a single absorbing state and  $Z-J$  transient states describing a Markov Chain, and an initial state probability rows vector. Form the sequence:

$$x_t = \begin{cases} 1 & \text{if the absorbing state is reached} \\ 0 & \text{otherwise} \end{cases}$$

The sequence  $x_t$  converges to 1.0 with probability 1, [3].

### 1.1.3 Absorbing Model of Optimum Seeking Devices

The EAs, and consequently the SGA, are included within wide field of algorithms of different nature that are used in "Optimum Seeking Devices". The convergence theory of Optimum Seeking Devices is quite wide and complex. In elemental models, where it is observed the evolution of these devices in a qualitative way, absorbing Markov models are used, [4], as in the case of EAs. Result that both models are totally equivalent. However, the NV model of SGA has a different qualitative behaviour, its compatibility with these models would be desirable.

### 1.1.4 Fisher's FTNS Theorem of Natural Selection

It is shown a result obtained in Populational Genetics, that it will be fundamental to approach the stochastic model to the real behaviour of the SGA.

In this simplified model, the biological population has individuals with chromosomes of size  $l$ , and its alleles with two possible symbols as a binary alphabet.

Let  $z(t)$  be the populational state at any time  $t$ , and let  $\langle f(z(t)) \rangle$  be, denoted here as  $\langle f \rangle$ , the average fitness of such state. An estimate from the variation of average fitness to each passage of time can be obtained in a general way and independently of the fitness, [5]:

$$\frac{d\langle f \rangle}{dt} = \frac{2}{\langle f \rangle} \mathbf{s}^2(f_i) \quad (4)$$

This result is known as “Fundamental Theorem of Natural Selection”, or Fisher’s FTNS Theorem. Sometimes, it is written in a different form with a quantity called the additive genetic variance,  $\mathbf{s}_g^2$ , in place of  $2\mathbf{s}^2(f_i)$ .

The FTNS Theorem is only an approximation for models of biological populations with discrete generations, but it provides valuable qualitative insight on how natural selection works. It tells us that  $\langle f \rangle$  should always increase (or at least not decrease) with time since  $d\langle f \rangle/dt$  is a non-negative quantity. Moreover, it says that evolution will be faster when there is more variability as measured by  $\mathbf{s}^2(f_i)$ .

For its totally general construction, this simple version of the FTNS theorem can be applied to any population that present a chromosomes structure and use a binary alphabet. This happens in the populations that SGA improves.

## 2 The NV - FTNS Absorbing Model of SGA

The Nix & Vose – Fundamental Theorem of Natural Selection absorbing model, NV – FTNS, is shown. It was introduced previously of empirical way, [6], through a practical postulate on the evolution of the average fitness of the sampling population, related to the Schema Theorem. With the introduction of FTNS Theorem, the postulate is rigorous and becomes an immediate consequence of Fisher’s FTNS Theorem.

*Let  $z(t)$  be the sampling state at step  $t$ , and  $\langle F(z(t)) \rangle$  the average fitness of such state. Then, the only transitions permitted in an S.G.A. are those that accomplish:  $\langle F(z(t+1)) \rangle \geq \langle F(z(t)) \rangle$ , in other words, the probabilities of transition states of a greater fitness to a smaller one are forbidden.*

The incorporation of this correction factor in the NV model can be added directly in the matrix elements of transition probabilities  $q_{i,j}$ , using the unitary function of Heaviside

$$U_j(i) = \begin{cases} 1 & \text{if } i \leq j \\ 0 & \text{if } i > j \end{cases} \quad (5)$$

Therefore:  $a_{i,j} = q_{i,j} U_{\langle F(j) \rangle}(\langle F(i) \rangle)$  (6)

Introducing these new elements on the matrix we obtain the matrix  $[A]$ , that presents multitude of zero elements. The matrix must be renormalized and the states ordering by average fitness, [7], ordering the states from great to smaller fitness. With this election  $[A]$  takes the form of a regular or canonical absorbing matrix. It is observed that all transient states have not null probabilities of ending in the absorbing state:

$$[A] = \begin{bmatrix} I & \mathbf{0} \\ \mathbf{R} & Q^* \end{bmatrix}$$

In this case,  $[I]$  is the scalar 1,  $\mathbf{0}$  is the transposed null vector of dimension  $Z-1$ ,  $\mathbf{R}$  the column vector of the same dimension and  $Q^*$  is the squared matrix of transition between transient states of dimension  $Z-1 \times Z-1$ . Here, the matrix  $Q^*$  is a lower triangular matrix, favoring the computational implementation of its stochastic analysis.

### 2.1 Expected Waiting Times, EWT.

If  $P_{i,0} = (\text{Card}\{Z\})^{-1}$  are assigned as initial probabilities,  $\forall i$ ; we are approximating to the general case. The average solution that is obtained assigning equal initial probabilities to all the states, approximately the same that in the general multinomial case for large search spaces, actual cases.

From the canonical form of the matrix  $[A]$ , the fundamental matrix  $[F]$  is obtained by induction. From  $[F]$  the general formulation for the EWT in the NV – FTNS model is obtained, [6].

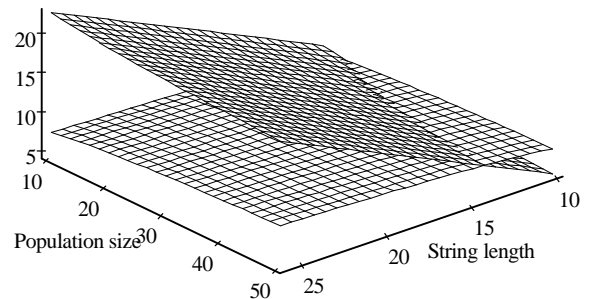


Fig. 1.  $\log_2$  EWT(J) of the NV model, inclined plane, and the NV - FTNS model of the SGA.

**Theorem:** [6]. *The EWT(J) of the Absorbing Markov Chain model for the SGA,  $n \ll r$ , in the maximum uncertainty situation and initial population of equal probability is:*

$$EWT(J) \approx \frac{\ln \Gamma \left( \binom{n+r-1}{r-1} - 1 \right)}{\binom{n+r-1}{r-1}} \quad (7)$$

with  $\Gamma(n) = (n-1)!$ , the Euler's Gamma function. This Equation can be approximated by:

$$EWT(J) \approx \ln \left( \binom{n+r-1}{r-1} \right) - 1 = H^* - 1 \quad (8)$$

For the maximum uncertainty situation, the EWT(J) in the NV - FTNS model coincides in practice with the value of the maximum entropy,  $H^*$ , from the populational search space.

## 2.2 Steady State

Using this model is immediate to find the steady state from a classic theorem of absorbing states for the limit matrix.

**Theorem:** Let  $[A]$  be the transition matrix of an absorbing Markov chain and  $[A]$  is in the canonical form, then exists a limit matrix  $[A]^\infty$ , such that  $[A]^k$  tends to  $[A]^\infty$  when  $k$  grows, where:

$$[A]^\infty = \begin{bmatrix} I & 0 \\ FR & 0 \end{bmatrix}$$

With  $[F] = ([I] - [Q^*])^{-1}$ , the fundamental matrix, [4].

The limit distribution is that the global optimum has been obtained, the system has entered the absorbing state and the stochastic process ends. Explicitly:

$$[A]^\infty = \begin{bmatrix} I & 0 \\ I & 0 \end{bmatrix}$$

In an efficient S.G.A., we arrive at the global optimum with  $n$  equal copies and the algorithm ends. The NV - FTNS model coincides with the real cases.

## 2.2 Experimental trials

We had contrasted initially the EWT in some test functions that are presented in the table 2, ordered of minor to greater hardness. Functions  $[F_3]$  and  $[F_4]$  are the same, except the parameters values.

FUNCTION	ANALITIC EXPRESSION
$[F_1]$	$(x-1)^2 - (x-1) + 3$
$[F_2]$	$\text{Int}(x+1)$
$[F_3]$	$4(x-a)(b-x)/(b-a)^2$ $a = 2.6 \times 10^8$ ; $b = 76870912$
$[F_4]$	$4(x-a)(b-x)/(b-a)^2$ $a = 0$ ; $b = 536870910$

Table 2. Functions used in the calculation of EWT.

It was calculated the EWT(J) for the four indicated functions, launching in each one of them a statistics sample of 50 launches with different random initial populations, except  $[F_4]$  with 100 launchings. Uniform mutation, with  $\mu = 0.5$ , case of maximum uncertainty, and 1-point crossover,  $\chi = 1$ . The outcomes are shown in the table 3. Despite these outcomes are preliminary, they seem to demonstrate a very good prediction with the same order of magnitude, through the NV - FTNS model.

FUNCTION	EWT(J)
$[F_1]$	61,4
$[F_2]$	67,8
$[F_3]$	168,5
$[F_4]$	1693,0
Average	$5,0 \cdot 10^2$
NV Model	$3,5 \cdot 10^7$
NV-FTNS	$5,5 \cdot 10^2$

Table 3. Comparison of EWT(J) with  $n = 30$ ,  $l = 30$

In figure 2 the results obtained in all cases, fixed  $n$  and  $l$ , as well as the case of  $n$  variables are shown. It has been drawn the  $\log_2$  EWT(J) to visualize all cases in a same graph. The thick line gives the theoretical values for the EWT(J) from the NV model, (1), the thin line shows the theoretical results from the NV - FTNS absorbing model, (7), (8), the circles give the experimental outcomes of the table 6 for  $l = 30$  and  $n$  variables. A cross represents the average result of the table 3.

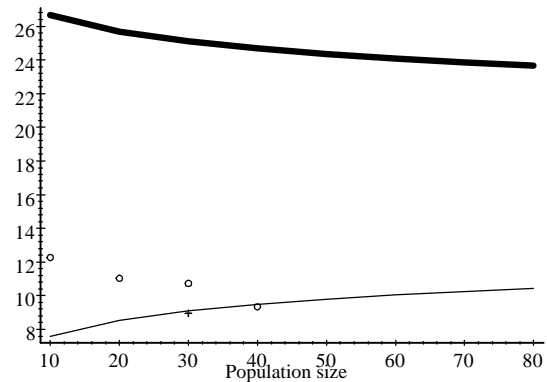


Fig. 2. Comparison of the  $\log_2$  EWT(J) for  $l = 30$ , between NV model and the NV - FTNS model.

## 3 Conclusions

The difference observed between the theoretical forecasts and the experimental results, [2], and the different qualitative behaviour between the NV model of SGA and other models more general of EAs, carry us to improve the NV model. After developing an

absorbing model, introduced through a practical postulate, [6], we have found its theoretical justification in the FTNS Theorem of Populational Genetics. It is the NV - FTNS absorbing model, which gives more close results to the experimental data and is compatible with more general models of EAs or Optimum Seeking Devices.

In the dynamical of the SGA, the sample population is approximated to the global optimum when the average fitness improves, as asserts the FTNS theorem. Its introduction in the dynamical model has achieved the equivalence among the SGA and other models more general as EAs or Optimum Seeking Devices. Thus, in maximum uncertainty case the value of the average entropy of the system coincides with the EWT, which SGA delays in obtaining for the first time in the sample population, the best individual from the search space.

The NV - FTNS model predicts values at the same order of magnitude that the experimental outcomes obtained.

In other hand, an important aspect is the appearance of meta-stable states. In the cases where it has not been reached the global optimum, nearby states to it have been obtained, that probably in an infinite time, will jump to the optimum state. These close states to the optimum can be designed as meta-stable states, their lifetime in occasions is enormously large and the algorithm seems not to evolve towards the optimum in a finite time. These meta-stable states different from the optimum have been already visualized. For example in [7] they are observed in the temporary evolution of the Markov chain, with the ordering by probability mass, as these states increased their statistic mass with time. It would be very important for the theoretical development of the S.G.A., the study of the characteristics that influence the appearance and maintenance of the meta-stable states, such as the codification, population diversity, evolutionary operators, etc. For example, it is possible to develop a heuristic that reduces notably the harmful effects of meta-stable states.

#### References:

- [1] Vose, M. D. *The Simple Genetic Algorithm. Foundations and Theory*. Bradford Book. MIT Press. Cambridge. MA. 1999
- [2] Cuesta, P.D.; Abderramán, J.C.; Jiménez, J.A. Galván, B. and Winter, G., Towards a Stop Criterion for Simple Genetic Algorithm, SGA, *European Congress on Computational Methods in Applied Sciences and Engineering, ECCOMAS 2000*, Barcelona. Spain. 2000.
- [3] Fogel, D.B. *Evolutionary Computation. Toward a New Philosophy of Machine Intelligence*. Nueva York: IEEE Press125. 2000
- [4] Taylor, H. M. and Karlin, S. *An Introduction to Stochastic Modeling*. Third Edition. Academic Press, CA. 1998.
- [5] Fisher, R. A., *The Genetical Theory of Natural Selection*, Oxford University Press, UK, 1999.
- [6] Cuesta, P.D.; Abderramán, J.C.; Jiménez, J.A. and Winter, G. Practical Modeling of Simple Genetic Algorithm via deterministic paths by Absorbing Markov Chains. *Genetic and Evolutionary Computation Conference, GECCO-2000*. Las Vegas Nevada. 2000
- [7] De Jong, K, W. Spears. Analyzing GAs using Markov models with semantically ordered and lumped states. *Foundations of Genetic Algorithms Workshop*, Morgan Kaufmann, pp. 85-100. San Mateo CA. 1996.