On the Mean of the Second Largest Eigenvalue on the Convergence Rate of Genetic Algorithms

Florian Schmitt and **Franz Rothlauf**

Working Paper 1/2001
January 2001

Working Papers in Information Systems
Editor: Prof. Dr. Armin Heinzl

University of Bayreuth
Department of Information Systems
Universitaetsstrasse 30
D-95440 Bayreuth, Germany
Phone +49 921 552807, Fax +49 921 552216
E-Mail: wi@uni-bayreuth.de
Internet: *h*ttp://wi.oec.uni-bayreuth.de

On the Mean of the Second Largest Eigenvalue on the Convergence Rate of Genetic Algorithms

Florian Schmitt
Department of Information Systems, University of Bayreuth
Universitaetsstr. 30
D-95440 Bayreuth/Germany
phone: ++49 921 55 2807
florian.schmitt@stud.uni-bayreuth.de,

Franz Rothlauf*
Department of Information Systems
University of Bayreuth
Universitaetsstr. 30
D-95440 Bayreuth/Germany
rothlauf@uni-bayreuth.de

January 22, 2001

Abstract

Genetic algorithms are sometimes disparagingly denoted as just a fancier form of a plain, stupid heuristic. One of the main reasons for this kind of critique is that users believed a GA could not guarantee global convergence in a certain amount of time.

Because the proof of global convergence of GAs using elitism has been performed elsewhere (11), in this work we want to extend previous work by J. Suzuki (13) and focus on the identification of the determinants that influence the convergence rate of genetic algorithms. The convergence rate of genetic algorithms is addressed using Markov chain analysis. Therefore, we could describe an elitist GA using mutation, recombination and selection as a discrete stochastic process. Evaluating the eigenvalues of the transition matrix of the Markov chain we can prove that the convergence rate of a GA is determined by the second largest eigenvalue of the transition matrix. The proof is first performed for diagonalizable transition matrices and then transferred to matrices in Jordan normal form.

The presented proof allows a more detailed and deeper understanding of the principles of evolutionary search. As an extension to this work we want to encourage researchers to work on proper estimations of the second largest eigenvalue of the transition matrix. With a good approximation, the convergence behavior of GAs could be described more exactly and GAs would be one step ahead on the road to a fast, reliable and widely accepted optimization method.

1 Introduction

Sometimes researchers speak disparagingly about genetic algorithms and label them to be just fancier form of a plain, simple heuristic. One main reason for this is that genetic algorithms (GA) stick to the prejudice that they are not able to guarantee convergence to the global optimum. The users do not know if the GA converges for a specific problem to the global optimum, and how much time the GA needs to converge. A closer look at genetic algorithms, however, reveals that there exists not only proof of the global convergence for genetic algorithms using elitism (11), but also some work about the convergence rate of GAs (13). The

* Also with Illinois Laboratory of Genetic Algorithms, University of Illinois at Urbana-Champaign, USA.

more complicated analysis of the convergence rate is important because a genetic algorithm which can be proven to converge, but needs infinite time for it, is not helpful for a effective use of genetic algorithms and would confirm the prejudices against GAs.

In this work we want to perform a more detailed analysis of the convergence rate using Markov chains. The Markov chain model is used for modeling a simple GA with the genetic operators selection, mutation and crossover. We investigate the determinants which the convergence rate depends on. With using some results from G. Rudolph (11) and J. Suzuki (13) we can prove that the convergence rate depends mainly on the value of the second largest eigenvalue of the transition matrix of the Markov chain. Furthermore, we transfer the results we get for diagonalizable transition matrices to the more general class of matrices in Jordan normal form. We illustrate that the proof for the convergence rate holds true for matrices in Jordan normal form, too.

The paper is structured as follows. In the following section we review some of the previous work about convergence behavior of genetic algorithms. This is followed in section 3 by presenting the requisites we want to use for our mathematical proof. We illustrate the assumptions for the genetic algorithm, present some fundamentals about the used Markov chain models, and review some properties of stochastic matrices. The section ends with a lemma about stochastic matrices from M. Iosifescu (8) and the global convergence statement from G. Rudolph (11). In section 4 we present the proof that the convergence rate is mainly determined by the second largest eigenvalue of the Markov chain describing a GA. First, the proof is performed for diagonalizable matrices and then transferred to matrices in Jordan normal form. The paper ends with concluding remarks.

2 Previous work

In this section we give a short review about the two main approaches that are used for investigating the convergence behavior of genetic algorithms.

In the field of genetic algorithms and convergence behavior we could distinguish two large areas of research. The first line of research results from the theoretical investigations by J. Holland (7) and D. E. Goldberg (4) and are based on the schemata theorem and the existence of building blocks for selectorecombinative genetic algorithms. In depth work in this line of research was done by H. Mühlenbein (1), D. Thierens (15), or D.E. Goldberg (5). A comprehensive overview of convergence time complexity can be found in D. Thierens thesis (14). All these models use the notion of building blocks, and are able to describe how building blocks grow and how long it takes to overtake a population accurately.

The second line of research treats GAs as a stochastic process with special properties. Starting with D.E. Goldberg (6) using Markov chains for modeling a GA, further work was done by T.E. Davis (2), A.E. Nix (9), and M.D. Vose (16) showing that a GA using selection, mutation and crossover can be fully described by the transition matrix of a Markov chain. Markov chains were also used by G. Rudolph for the proof of the global convergence of evolutionary algorithms (11). The proof was an important step towards a better theoretical understanding of GAs. Based on the global convergence proof, J. Suzuki (13) identified the influence of the eigenvalues of the transition matrix of the Markov chain on the convergence rate of a GA. Although he gave upper and lower bounds for the convergence rate, he was not able to specify the important eigenvalue for the convergence rate exactly.

3 Preliminaries

This section provides the background that is necessary for understanding the investigations concerning the convergence analysis. We start by defining some basic properties of the used GA. This is followed by a description of the basic concepts of Markov chains and a review of the properties of stochastic matrices. The section ends with the proof of GA convergence as provided in (11).

3.1 Properties of the Genetic Algorithm

This paper deals with a kind of simple Genetic Algorithm (GA) in which the genetic operators are restricted to crossover, mutation and selection. Furthermore, the GA uses a binary representation of fixed length. The

population size of the GA is determined a priori, and the probabilities for the three operators are not equal to zero.

For our investigation we use elitism in a way that the best parent survives if it is better than the best offspring. The individual with the highest fitness among all possible individuals is denoted as *super individual*. It represents the global optimum of the problem.

Further assumptions concerning the fundamental structure of the GA are not necessary in this context. Different types of crossover and mutation operators should have no influence on the convergence behavior of GAs and the proof shown in section 4 should still hold.

3.2 Markov chain analysis

The principal behavior of a GA can be described by using the Markov chain model. Using this concept we are able to develop a convergence model for the GA.

A Markov chain is a discrete stochastic process. The behavior of the stochastic process in future states depends only on the present states, but not on the past ones. Therefore, the probabilistic motion of a Markov chain could be described by using a transition matrix P.

For homogeneous Markov chains the t-th step transition matrix P^t can be determined iteratively. The Chapman-Kolmogorov equations (compare M. Iosifescu, p.65 (8)) yield

$$P^t = \prod_t P.$$

Let p_i^t denote the probability that the Markov chain is in state i at step t. The p_i^t can be gathered in a row vector $p^t = (p_1^t, p_2^t, \ldots, p_n^t)$. The initial distribution p^0 is similarly defined. Then

$$p^t = p^0 \cdot P^t$$

for $t \geq 0$. Therefore, a homogeneous Markov chain is completely determined by the tuple (p^0, P).

The distribution p on the states of the Markov chain is called a stationary distribution, if $p\,P = p$, and is called a limit distribution, if the limit $p = p^0 \lim_{t\to\infty} P^t$ exists.

Every transition matrix of a Markov chain is stochastic. A non-negative matrix is said to be *stochastic* if all its row sums are equal to one. Further matrix classifications occurring in the following are given by G. Rudolph, p.55 (11). Stochastic matrices possess special properties:

- The eigenvalues of a stochastic matrix have modulus less or equal to 1.
- An irreducible stochastic matrix possess a simple unit eigenvalue.
- The right-hand eigenvector corresponding to a unit eigenvalue of a stochastic matrix is given by $e = (1, \ldots, 1)^T$.
- The vector p is a stationary probability vector of a stochastic matrix, if a left-hand eigenvalue corresponds to a unit eigenvalue.

The source of these statements can be found in W. Stewart, p. 28-30 (12).

3.3 Proof of global convergence

A qualitative Markov chain model of GAs is sufficient for the global convergence proof. The following lemma is necessary for the convergence proof presented by G. Rudolph (11). We use this lemma later for the analysis of the convergence rate.

Lemma 1
Let P be a reducible stochastic matrix, where $C \in \mathbb{R}^{m\times m}$ is a primitive stochastic matrix and $R, T \neq 0$. Then

$$\begin{aligned} P^{\infty} &= \lim_{t\to\infty} P^t \\ &= \lim_{t\to\infty} \begin{pmatrix} C^t & 0 \\ \sum_{i=0}^{t-1} T^i R C^{t-i} & T^t \end{pmatrix} \\ &= \begin{pmatrix} C^{\infty} & 0 \\ R^{\infty} & 0 \end{pmatrix} \end{aligned}$$

is a stable stochastic matrix with $P^{\infty} = e\,p^{\infty}$, where $p^{\infty} = p^0 P^{\infty}$ is unique regardless of the initial distribution, and the limit distribution p^{∞} satisfies

$$p_i^{\infty} > 0 \quad \text{for } 1 \leq i \leq m \text{ and } p_i^{\infty} = 0$$

for $m < i \leq n$.

Matrix C is associated with the absorbing states of the Markov chain. For the proof see M. Iosifescu, p.126 (8). Using this lemma we can finally present the global convergence statement:

A genetic algorithm with an arbitrary initial distribution converges to the global optimum if the following assumptions are fulfilled:

- Selection chooses the best individual from parents and offspring (elitism).
- Every state is reachable from any other state.

For a detailed proof the reader is referred to G. Rudolph, chapter 5 (11).

4 Analysis of the convergence rate

For an analysis of the convergence rate, the proof of convergence, illustrated in the previous section, is a necessary condition. If we can not prove that the GA converges an investigation into convergence rate is useless. Using the convergence proof we prove in this section that the convergence rate is mainly determined by the second largest eigenvalue of the Markov chain describing a GA.

The theoretical identification of the second largest eigenvalue as the fundamental parameter influencing the convergence rate is first obtained for diagonalizable matrices. This result is then transferred to matrices in Jordan normal form, which represents a more general class of matrices.

4.1 Diagonalizable matrices

We want to start by identifying a measurement for the convergence rate of a GA. This should allow us to determine the progress of the GA's convergence.

J. Suzuki (13) measures the convergence rate as the degree in which the individual with the highest fitness in a population coincides with the super individual. Therefore, he investigates how closely the probability $\sum_{k\in X^*} p_k^n$ converges to $\sum_{k\in X^*} p_k^{\infty}$, where X^* denotes the amount of populations containing the super individual. We want to use the same convergence measurement for the following analysis of the convergence rate.

Due to the necessary assumption that the GA is globally convergent, the probability $\sum_{k\in X^*} p_k^{\infty}$ converges to one (Lemma 1). Based on Suzuki's measurement of convergence and Lemma 1 we analyze how fast the probability $\sum_{k\in X^*} p_k^n$ converges to one for a finite number of generations.

Furthermore, we use for the analysis of the convergence rate of transition matrices the classical Perron Formula (compare V. Romanovsky, chapter 1 (10)). This formula allows us to compute the powers of a square matrix P. In the case of diagonalizable matrices, the Perron Formula can be reduced to

$$P = \sum_{i=1}^{n} \lambda_i v_i u_i^T, \tag{1}$$

where λ_i is an eigenvalue of matrix P, v_i and u_i are the corresponding right and left eigenvectors. Equation 1 is called the spectral representation of a matrix. As a relevant consequence it follows

$$P^n_{k,\nu} = \sum_{i=1}^{N} w^{(i)}_{k,\nu} \lambda_i^n. \tag{2}$$

Using the previous statements we could finally formulate the theorem that the convergence rate depends on the second largest eigenvalue of the diagonalizable transition matrix:

Theorem 1
Let $C > 0$ be constant.
A constant C exists which satisfies

$$\sum_{k \in X^*} p^n_k \geq 1 - C \cdot |\lambda_2|^n, \tag{3}$$

where $|\lambda_2|$ is the second largest eigenvalue of a diagonalizable transition matrix P describing the Markov chain of a global convergent GA.

Proof of Theorem 1
If $k \in X^*$, then can be followed:

$$\begin{aligned} P^n_{k,\nu} &= 0 \qquad \nu \neq k, \\ P^n_{k,\nu} &= 1 \qquad \nu = k. \end{aligned}$$

Using equation (2) the following equation holds for $k \notin X^*$

$$P^n_{k,\nu} = \sum_{i=1}^{N} w^{(i)}_{k,\nu} \lambda_i^n.$$

P is a stochastic matrix. Therefore, the largest eigenvalue of P is equal to unity. The simple unit eigenvalue is identified as the absorbing state of the corresponding Markov chain. Furthermore,

$$w^{(1)}_{k,\nu} = (v_1 u_1^T)_{k,\nu} = e \cdot {p^\infty}^T = \begin{pmatrix} p^\infty_1 & \cdots & p^\infty_N \\ \vdots & & \vdots \\ p^\infty_1 & \cdots & p^\infty_N \end{pmatrix}$$

holds because of the properties implied by a stochastic matrix (compare subsection 3.2). As P describes a Markov chain, we get

$$\begin{aligned} p^n_\nu &= \sum_{k=1}^{N} p^0_k P^n_{k,\nu} \\ &= \sum_{k=1}^{N} p^0_k \sum_{i=1}^{N} w^{(i)}_{k,\nu} \lambda_i^n. \end{aligned}$$

Therefore, $\sum_{\nu \notin X^*} p^n_\nu =$

$$\begin{aligned} &= \sum_{\nu \notin X^*} \sum_{k=1}^{N} p^0_k \sum_{i=1}^{N} w^{(i)}_{k,\nu} \lambda_i^n \\ &= \sum_{\nu \notin X^*} \sum_{k=1}^{N} p^0_k \left[\lambda_1^n w^{(1)}_{k,\nu} + \sum_{i=2}^{N} w^{(i)}_{k,\nu} \lambda_i^n \right] \\ &= \sum_{\nu \notin X^*} \sum_{k=1}^{N} \left[p^0_k w^{(1)}_{k,\nu} + p^0_k \sum_{i=2}^{N} w^{(i)}_{k,\nu} \lambda_i^n \right] \\ &= \sum_{\nu \notin X^*} \sum_{k=1}^{N} p^0_k p^\infty_\nu + \left[\sum_{\nu \notin X^*} \sum_{k=1}^{N} p^0_k \textstyle\sum_{i=2}^{N} w^{(i)}_{k,\nu} \right] |\lambda_2|^n \end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{\nu \notin X^*} \sum_{k=1}^{N} p_k^0 \, p_\nu^\infty \; + \; C \cdot |\lambda_2|^n \\
&= \sum_{\nu \notin X^*} p_\nu^\infty \sum_{k=1}^{N} p_k^0 \; + \; C \cdot |\lambda_2|^n \\
&= \sum_{\nu \notin X^*} p_\nu^\infty \cdot 1 \; + \; C \cdot |\lambda_2|^n \\
&= C \cdot |\lambda_2|^n,
\end{aligned}
$$

where $C > 0$ and $p_\nu^\infty = 0$ for $\nu \notin X^*$(Lemma 1). Using the complementary probability the proof is completed.

q.e.d.

As a result, the convergence rate of the corresponding Markov chain is mainly determined by the second largest eigenvalue of the diagonalizable transition matrix.

4.2 Matrices in Jordan normal form

In the previous subsection we used diagonalizable matrices for our proof, because they have properties that can be used advantageously. However, in general the fundamental assumptions for the class of diagonalizable matrices are very restrictive. An arbitrary matrix is often not diagonalizable. Hence, we analyze in this subsection the more general class of matrices in Jordan normal form[1] and illustrate how the proof from the previous subsection can be transferred to matrices in Jordan normal form.

The transition matrix P in Jordan normal form describing the Markov chain of a global convergent GA can be written as (see F. Gantmacher (3))

$$
P \; = \; \begin{pmatrix} 1 & & & 0 \\ & J_2 & & \\ & & \ddots & \\ 0 & & & J_m \end{pmatrix},
$$

where the J_i are called Jordan Blocks with the following shape

$$
J_i \; = \; \begin{pmatrix} \lambda_i & & & 0 \\ 1 & \lambda_i & & \\ & \ddots & \ddots & \\ 0 & & 1 & \lambda_i \end{pmatrix}.
$$

The λ_i are the eigenvalues of P (compare M. Iosifescu, p. 50-51 (8)). The Jordan Blocks have the important property

$$
J_i \; = \; \lambda_i \cdot E \; + \; \underbrace{\begin{pmatrix} 0 & & & 0 \\ 1 & 0 & & \\ & \ddots & \ddots & \\ 0 & & 1 & 0 \end{pmatrix}}_{\mathcal{U}},
$$

where matrix $\mathcal{U}$ is called nilpotent. A non-negative square matrix $\mathcal{U}$ is defined to be *nilpotent* , if $\exists k \in \mathbb{N}$ holds $\mathcal{U}^k = 0$.

$$
P^n \; = \; \begin{pmatrix} 1 & & & 0 \\ & J_2 & & \\ & & \ddots & \\ 0 & & & J_m \end{pmatrix}^n \; =
$$

[1] All matrices above $\mathbb{C}$ have a Jordan normal form.

$$= \begin{pmatrix} 1 & & & 0 \\ & J_2^n & & \\ & & \ddots & \\ 0 & & & J_m^n \end{pmatrix}$$

reveals that the unit eigenvalue does not affect the convergence rate. Therefore, J_i^n , $i = 2 \dots m$, are the remaining parameters that have to be analyzed. With using

$$\mathcal{U} = \begin{pmatrix} 0 & & & 0 \\ 1 & 0 & & \\ & \ddots & \ddots & \\ 0 & & 1 & 0 \end{pmatrix}$$

the Jordan blocks become

$$\begin{aligned} J_i^n &= [\lambda_i \cdot E + \mathcal{U}]^n \\ &= \lambda_i^n \cdot E + n\,\lambda_i^{n-1} \cdot E \cdot \mathcal{U} \\ &+ \dots + n\,\lambda_i \cdot E \cdot \mathcal{U}^{n-1} + \mathcal{U}^n. \end{aligned}$$

As mentioned before, matrix $\mathcal{U}$ is nilpotent. This yields

$$\begin{aligned} J_i^n &= \lambda_i^n \cdot E + n\,\lambda_i^{n-1} \cdot E \cdot \mathcal{U} \\ &+ \dots + \frac{n\,(n-1)\,\dots\,(n-k)}{(k-1)!}\,\lambda_i^{n-k-1} \cdot E \cdot \mathcal{U}^{k-1} \\ &\leq C \cdot \lambda_i^n \cdot E \qquad \text{because } |\lambda_i| < 1, \end{aligned}$$

for $C > 0$ and $k \in \mathbb{N}$. Hence we obtain the final result:

$$J_i^n \leq C \cdot \lambda_i^n \cdot E$$

for $C > 0$ and $k \in \mathbb{N}$.

Interpreting the results, reveals that the J_i^n are upper bounded by λ_i^n for $i = 2 \dots m$, where λ_i is the corresponding eigenvalue to J_i. The behavior of P^n mainly depends on the size of the second largest eigenvalue λ_2, because in comparison to λ_2^n the powers of the other eigenvalues can be neglected.

The obtained result extends the proof for diagonalizable transition matrices and confirms that the second largest eigenvalue is also the important parameter for the long term behavior of matrices in Jordan normal form.

Finally, we want to note that in general the transition matrix of a Markov chain is not in Jordan normal form, but using standard matrix transformations it always can be transformed into it. Although, we assume that the transformation of an arbitrary transition matrix into a matrix in Jordan normal form does not modify the statements about convergence rate, the formal proof for this is still open.

5 Conclusion

After a short review of two different approaches to the analysis of the convergence behavior of genetic algorithms (building block oriented versus Markov chain models) we present some requisites we need for our investigation into the convergence rate of genetic algorithms. In section 3 we present some principles of Markov chain models, and review some properties of stochastic matrices. As an investigation into the convergence rate of genetic algorithms only makes sense, if it can be proven that a GA converges to the global optimum, we review the convergence proof from G. Rudolph (11). Using this proof and some of the work by J. Suzuki (13) we can extend the existing convergence models and prove in the following section that the convergence rate of genetic algorithms, modeled with Markov chains is determined by the second largest eigenvalue of the characteristic transition matrix of the Markov chain. Finally, we transfer the results we get for diagonalizable matrices to matrices in Jordan normal form.

This paper extends existing models about convergence rate of GAs (13) and proves that the convergence rate depends on the second largest eigenvalue of the diagonalizable transition matrix. The theoretical analysis of the convergence rate is based on the existing evidence of global convergence, which are based on Markov chains. The obtained results allow a more detailed and deeper theoretical understanding of the principles of evolutionary search. We hope that the results could inspire researchers to put the focus of research more on the underlying theoretical principles and not to focus only on practical applications of GAs.

As a straightforward extension of this work we want to encourage researchers to work on the estimation of the second largest eigenvalue of the transition matrix. A proper approximation can give us information about the optimal choice of GA parameters like mutation and crossover probability or selection pressure. With that knowledge we could use a GA more efficiently, and we would be one step ahead on our long road to the development of competent GAs that are able to solve problems of bounded complexity autonomously, fast and reliably.

References

[1] H. Asoh and H. Mühlenbein. On the mean convergence time of evolutionary algorithms without selection and mutation. *Parallel Problem Solving from Nature- PPSN III*, pages 88–97, 1994.

[2] T. E. Davis and J. C. Principe. A markov chain framework for the simple genetic algorithm. *Evolutionary Computation*, 1(3):269–288, 1993.

[3] F. Gantmacher. *Matrizentheorie*. Berlin, Springer Verlag, 1986.

[4] D. E. Goldberg. *Genetic algorithms in search, optimization and machine learning.* Reading Addison-Wesley, 1989.

[5] D. E. Goldberg. The race, the hurdle, and the sweet spot. In P. J. Bentley, editor, *Evolutionary Design by Computers*, pages 105–118. Morgan Kaufmann, San Francisco, CA, 1999.

[6] D. E. Goldberg and P. Segrest. Finite Markov chain analysis of genetic algorithms. *Proceedings of the Second International Conference on Genetic Algorithms*, pages 1–8, 1987.

[7] J. H. Holland. *Adaptation in Natural and Artificial Systems.* The MIT Press, Cambridge, Massachusetts, 2. edition, 1998.

[8] M. Iosifescu. *Finite Markov Processes and their applications.* Chichester, John Wisly and Sons, 1980.

[9] A. E. Nix and M. D. Vose. Modelling genetic algorithms with markov chains. *Annals of Mathematics and Artifical Intellignece*, 2:79–88, 1992.

[10] V. Romanovsky. *Discrete Markov chains.* Wolters-Noordhoff Publishing, Groningen, 1978.

[11] G. Rudolph. *Convergence properties of evolutionary algorithms.* Kovač, Hamburg, 1997.

[12] W. J. Stewart. *Introduction to the Numerical Solution of Markov chains.* Princeton University Press, Princeton, New Jersey, 1994.

[13] J. Suzuki. A markov chain analysis on simple genetic algorithms. *IEE Transactions on Systems, Man and Cybernetics*, 25(4):655–659, 1995.

[14] D. Thierens. *Analysis and design of genetic algorithms.* Katholieke Universiteit Leuven, Leuven, Belgium, 1995.

[15] D. Thierens and D. E. Goldberg. Convergence models of genetic algorithm selection schemes. *Parallel Problem Solving from Nature- PPSN III*, pages 119–129, 1994.

[16] M. D. Vose and G. E. Liepens. Punctuated equilibria in genetic algorithms. *Complex Systems*, 5:31–44, 1991.