*I Krajowa Konferencja*
***Algorytmy Ewolucyjne***
*Murzasichle, 12-15.06.1996*

# On asymptotic properties of a selection-with-mutation operator

Kazimierz Grygiel
Institute of Informatics, Warsaw University,
Banacha 2, 02-097 Warszawa, Poland
email: grygiel@mimuw.edu.pl

May 16, 1996

**Abstract**

Asymptotic properties of a selection-with-mutation (SM) operator are investigated for the infinite population evolutionary algorithm (a special case of the SGA model proposed by Vose and others). It is shown that for any strictly positive fitness function the algorithm must converge to the unique solution, independent of the initial population, if mutation is positive. The limiting population is the eigenvector corresponding to the maximum eigenvalue of a matrix related to the SM operator.

## 1 Introduction

In a series of articles [1, 3-5] M. D. Vose and co- authors have introduced and investigated an exact mathematical model of a simple genetic algorithm. A genetic search has been considered as a homogeneous Markov chain with a state space composed of all possible populations of a given size. The transition matrix is determined by a fitness function and two genetic operators (mutation and crossover). When the mutation rate is nonzero, the chain is ergodic, and, as a consequence, it has the unique, strictly positive limiting (steady-state) probability distribution, independent of an initial state. Hence, every state of the chain (i.e., every population) will be visited infinitely often, so that the search cannot converge to any fixed population. On the other hand, with zero mutation the chain has many absorbing states (in fact, any "homogeneous" state becomes an absorbing one), and there is little hope that it will end up providing the "best" solution.

The dilemma, however, is not so serious as it could seem. GA's were invented as practical tools, and what matters from the practical point of view are the *magnitudes* of probabilities. If the limiting probability of a state is close enough to one, then the process will spend almost all its time in that state. (For the time being, we put aside the question of how long it takes to achieve such a state.)

How could we know that our GA will give high probability to a "feasible" population (i.e., one which contains the optimal individual)? It is clear that we cannot answer this question by calculating steady-state probabilities. After all, we use this algorithm just to *find* the optimal individual!

Nix and Vose [1] have argued that as population size grows to infinity, the steady state of the finite population model can give nonvanishing probability only to fixed points of an operator which determines the evolutionary path of the infinite population model. This result seems to validate the concept of an infinite population genetic algorithm and provide incentives for further investigation of that model.

In this paper a special case of an infinite population genetic algorithm is considered, with nonzero mutation and no crossover. It is shown that in this case there exists the unique limiting population vector, independent of an initial state, to which the (deterministic) search process converges. Moreover, the algebraic characterization of both the limiting population and its average fitness is given in terms of a characteristic equation of a matrix.

## 2 The Infinite Population Model

Let $\Omega$ be a *genetic space*, $\Omega = \{i : 0 \leq i \leq n-1\}$, where $n = 2^l$, $l \geq 1$. The genetic space can be thought of as the set of all binary strings of length $l$. A collection of $N$ elements (not necessarily different) of $\Omega$ is called a *population* of size $N$. Any population can be represented by its *frequency vector*

$$x^T = (x_0, x_1, \ldots, x_{n-1}) : \; x_j \geq 0, \; \sum_j x_j = 1,$$

where $x_j$ is a proportion of element $j \in \Omega$ in a population. In other words, a population vector is identical to baricentric coordinates of some point in the standard unit ($n$-1)–simplex $S^{n-1}$. A population is uniquely determined by its frequency vector $x$ and population size $N$.

This method can be used to introduce the concept of an infinite population. In the following we will identify a *population* on $\Omega$ with a point in $S^{n-1}$ (or its frequency vector).

A *discrete evolution process* is an infinite sequence

$$x^{(0)}, x^{(1)}, x^{(2)}, \ldots$$

of populations, or a motion of a point $x^{(t)}$ on the standard unit ($n$-1)–simplex, starting from an initial position $x^{(0)}$. We assume that there is some stochastic or deterministic *law of evolution* which governs this motion. In the finite case, the evolution process forms a Markov chain and the law of evolution is given by a transition probability matrix. In the infinite case, the evolution process is essentially deterministic, and much easier to analyze. Its law of evolution is given by some operator $Q : S^{n-1} \to S^{n-1}$.

In this paper, we will consider a special case of the law of evolution, determined by proportional selection and mutation. In the next section, the corresponding operators will be described, assuming the infinite model.

# 3 Selection and Mutation Operators

Let $f : \Omega \rightarrow R^+$ be a (strictly positive) fitness function. We will use the notation $f_i$ for $f(i)$ $(i \in \Omega)$ and identify function $f$ with the column vector $(f_0, \ldots, f_{n-1})^T$. Let $(x|y) = x^T y$ denote the inner product of two $n-$ vectors $x, y$.

The proportional selection operator $\mathcal{F} : S^{n-1} \rightarrow S^{n-1}$ is now defined as

$$\mathcal{F}(x) = \frac{1}{(f|x)} Fx,$$

where $F = diag\{f_0, \ldots, f_{n-1}\}$ is an $n \times n$ diagonal matrix.

The standard interpretation of the $\mathcal{F}$ operator is as follows: if $x^{(t)}$ is a population, then $\mathcal{F}(x^{(t)})$ is a vector with components giving selection probabilities of individuals to the next population $x^{(t+1)}$ (prior to any genetic transformations). In the infinite model, these probabilities coincide (by the law of large numbers) with relative frequencies (proportions) of individuals in the new population.

We also assume the usual mutation scheme with *mutation rate* $\mu$, $0 < \mu < 1$. Let $\mu_{ij}$ be the probability of changing individual (binary string) $j \in \Omega$ into individual $i \in \Omega$ by mutation. Then

$$\mu_{ij} = \mu^{/i \oplus j/} (1 - \mu)^{l - /i \oplus j/}$$

where $\oplus$ is *component-wise exclusive-or* on binary strings and $/ \cdot /$ denotes the number of ones in a binary string.

The mutation operator $\mathcal{M} : S^{n-1} \rightarrow S^{n-1}$ is defined by the following equation:

$$\mathcal{M}(x) = Mx,$$

where $M = (\mu_{ij})$ is the *mutation matrix.*

It's clear that $\mathcal{M}(x)$ is the population resulting from population $x$ subject to the process of mutation.

The *selection-with-mutation* (SM) operator $\mathcal{Q}$ is obtained by superposition of mutation and selection operators:

$$\mathcal{Q}(x) = \mathcal{M}(\mathcal{F}(x)) = \frac{1}{(f|x)} Qx,$$

where $Q = MF$.

It is easy to see that $Q = (\mu_{ij} f_j)$ is a positive matrix (i.e., all entries of $Q$ are positive).

Observe also that (i) $M$ is a symmetric matrix and (ii) $F$ is a symmetric, positive definite matrix. The latter fact follows from the relationship

$$x^T F x = \sum_{i=0}^{n-1} f_i x_i^2 > 0$$

for all $x \in R^n, x \neq 0$, which is true under assumption that fitness function $f$ is strictly positive.

In the following, we will use some facts from matrix algebra (cf. [2]):

**Definition.** Let $K$ be a symmetric, positive definite matrix. We say that a matrix $A$ is $K$-*symmetric* if $KA = A^T K$. Two vectors $x$, $y$ are said to be $K$-*orthogonal* if $(x|Ky) = 0$.

The following theorem is a generalization of the well- known result for symmetric matrices:

**Theorem.** A $K$-symmetric matrix $A$ has a complete set of eigenvectors that can be chosen to be $K$- orthogonal in pairs. Moreover, the eigenvalues of $A$ are real.

Now, we can state the following proposition:

**Proposition 1.** The matrix $Q = MF$ is $F$- symmetric.

**Proof:**

$$FQ = F(MF) = (FM)F = (F^T M^T)F = (MF)^T F = Q^T F.$$

An immediate consequence of Proposition 1 is that the matrix $Q$ has a complete set of eigenvectors that can be chosen to be $F$-orthogonal in pairs.

# 4 Evolution of Population and the Limiting Solution

For the infinite population model and $\mathcal{Q}$ operator as the law of evolution we have

$$x^{(t)} = \mathcal{Q}^t(x^{(0)})$$

**Lemma** For any $x \in S^{n-1}$

$$\mathcal{Q}^t(x) = \frac{1}{(f|Q^{t-1}x)} Q^t x$$

**Proof** (by induction on $t$):
(i) For $t = 1$ we have

$$\mathcal{Q}(x) = \frac{1}{(f|x)} Qx$$

(ii) Assume our Lemma is true for some $t \geq 1$. Then

$$\mathcal{Q}^{t+1}(x) = \mathcal{Q}(\mathcal{Q}^t(x)) = \frac{1}{(f|\mathcal{Q}^t(x))} Q \cdot \mathcal{Q}^t(x) =$$

$$\frac{1}{(f|\frac{1}{(f|Q^{t-1}x)} Q^t x)} Q \cdot \frac{1}{(f|Q^{t-1}x)} Q^t x = \frac{1}{(f|Q^t x)} Q^{t+1} x.$$

By the Perron-Frobenius theorem for primitive (in particular, positive) matrices, there exists an eigenvalue $\lambda_0$ of $Q$ such that:

(i) $\lambda_0 > 0$,
(ii) with $\lambda_0$ can be associated the unique, strictly positive eigenvector $v^{(0)} \in S^{n-1}$,
(iii) $\lambda_0 > |\lambda|$ for any eigenvalue $\lambda \neq \lambda_0$.

**Proposition 2.** For any $x \in S^{n-1}$, a discrete evolution process

$$x, \mathcal{Q}(x), \mathcal{Q}^2(x), \ldots,$$

converges to the strictly positive limit $x^* \in S^{n-1}$, independent of $x$. The limiting population $x^*$ is an eigenvector corresponding to the maximum eigenvalue of the matrix $Q$.

**Proof**

We will use the well-known power method. Let $\{\lambda_j,\ j = 0, \ldots, n-1\}$ be the set of all eigenvalues of $Q$ and let $\{v^{(j)} :\ j = 0, \ldots, n-1\}$ be an $F$-orthogonal set of corresponding eigenvectors. (We assume, without loss of generality, that $v^{(0)} \in S^{n-1}$.)

Then, any $x \in S^{n-1}$ can be expanded as

$$x = \sum_j a_j v^{(j)}, \text{ where } a_j \in R$$

Therefore

$$Qx = \sum_j a_j Q v^{(j)} = \sum_j a_j \lambda_j v^{(j)},$$

and generally

$$Q^t x = \sum_j a_j \lambda_j^t v^{(j)}$$

Using the Lemma we have now

$$Q^t(x) = \frac{\sum_j a_j \lambda_j^t v^{(j)}}{\sum_j a_j \lambda_j^{t-1} (f|v^{(j)})} = \lambda_0 \frac{\sum_j a_j (\lambda_j/\lambda_0)^t v^{(j)}}{\sum_j a_j (\lambda_j/\lambda_0)^{t-1} (f|v^{(j)})}$$

Hence

$$\lim_{t\to\infty} Q^t(x) = \lambda_0 \frac{a_0 v^{(0)}}{a_0 (f|v^{(0)})} = \frac{\lambda_0 v^{(0)}}{(f|v^{(0)})} = x^*$$

Cancellation of the term $a_0$ above is correct because $a_0 > 0$. Indeed, $(v^{(0)}|Fx) > 0$ as $v^{(0)}$ is positive and $Fx$ is nonnegative, nonzero for $x \in S^{n-1}$. From $F$-orthogonality of eigenvectors $v^{(j)}$

$$(v^{(0)}|Fx) = \sum_j a_j (v^{(0)}|Fv^{(j)}) = a_0 (v^{(0)}|Fv^{(0)}) > 0,$$

hence $a_0 > 0$.

To complete the proof, let us observe that $x^*$ must equal $v^{(0)}$. Indeed,

$$Qx^* = \frac{\lambda_0}{(f|v^{(0)})} Qv^{(0)} = \frac{\lambda_0}{(f|v^{(0)})} \lambda_0 v^{(0)} = \lambda_0 x^*$$

Therefore, $x^*$ is an eigenvector corresponding to $\lambda_0$. Moreover, $x^* \in S^{n-1}$, as $S^{n-1}$ is a closed subset of $R^n$. Hence, $x^* = v^{(0)}$ and $(f|v^{(0)}) = \lambda_0$.

# 5 Conclusion

We investigated the asymptotic behavior of a discrete evolution process in the infinite population model. The law of evolution was determined by the selection-with-mutation (SM) operator. It has been shown that, for any initial population, there exists the unique limiting population which only depends on a fitness function and a mutation rate. The limiting population (represented by a frequency vector) can be algebraically characterised as an eigenvector $v^{(0)}$ corresponding to the maximum eigenvalue $\lambda_0$ of a matrix related to the SM operator. The average fitness of the limiting population is equal to the maximum eigenvalue, i.e. to the spectral radius of the matrix.

It is not difficult to obtain first-order approximations (with respect to the mutation rate $\mu$) for both $\lambda_0$ and $v^{(0)}$. In particular

$$\lambda_0 = (1 - l\mu) f_{max} + o(\mu),$$

where $f_{max}$ is the maximum fitness. However, the usefulness of such first-order approximations has yet to be verified in practice.

Vose and Wright [5] have investigated the composition of selection and crossover operators (the case of zero mutation). In that case, however, the limiting solution can depend on an initial population. Some experimental results have been obtained for the complete genetic algorithm (selection + mutation + crossover) [4], but there's still no general theory.

## References

1. A. E. Nix, M. D. Vose, Modeling genetic algorithms with Markov chains, *Annals of Mathematics and Artificial Intelligence* 5 (1):79-88 April 1992.

2. M. C. Pease, III. *Methods of Matrix Algebra*, Academic Press, New York and London, 1965.

3. M.D. Vose. Modeling simple genetic algorithms. In: Whitley, D. (editor) *Foundations of Genetic Algorithms - 2 (FOGA-92)*, pp. 63-74. Vail, CO, 24-29 July 1992. Morgan Kaufmann, San Mateo, CA, 1993.

4. M. D. Vose. A closer look at mutation in genetic algorithms. *Annals of Mathematics and Artificial Intelligence*, 10 (4), 1994.

5. M.D. Vose, A. D. Wright (1994). Stability of vertex fixed points and applications. (Obtained via Internet.)