A Simulated Annealing Like Convergence Theory for the Simple Genetic Algorithm

Thomas E. Davis
WL/MNGS
Eglin AFB
FI 32542-5434

Jose C. Principe
Department of Electrical Engineering
University of Florida
Gainesville, FL 32611-2024

Abstract
This paper synthizes a substantial body of work whose goal is extrapolation of the existing theoretical foundation of the simulated annealing algorithm onto a Markov chain genetic algorithm model. Some key intermediate convergence results are reported, including a genetic algorithm mutation probability control parameter sequence bound analogous to the simulated annealing temperature schedule bounds.

1 INTRODUCTION
Both simulated annealing and the genetic algorithm are stochastic relaxation search techniques suitable for application to a wide variety of combinatorial complexity non-convex optimization problems [14, 15, 1, 13, 6, 9, 10, 11]. Each produces a sequence of candidate solutions (or in the case of the genetic algorithm a sequence of populations of candidate solutions) to the underlying optimization problem, and the purpose of both algorithms is to generate sequences with probability distribution biased toward candidates which optimize the objective function.

Simulated annealing exploits an analogy of combinatorial optimization to the annealing of crystalline solids [14], in which the solid is allowed to cool very gradually from some elevated temperature and thereby allowed to relax into one of its minimum energy states. The appeal of the algorithm class is that it provides asymptotic convergence to a globally optimal solution. A substantial body of knowledge exists for describing the algorithm convergence behavior based upon the asymptotic probability distribution associated with a non-stationary Markov chain algorithm model [15]. Section 2 below is a summary of the essential features of the simulated annealing convergence theory.

The genetic algorithm emulates the evolutionary behavior of biological systems [1, 6]. It generates a sequence of populations of candidate solutions to the underlying optimization problem by employing a set of genetically inspired stochastic state transition operators to transform each population of candidate solutions into a descendnat population. A variety of distinct genetic operators are reported in the literature [1, 5, 9, 10, 11], the most important of which are (1) reproduction, (2) crossover and (3) mutation. A one, two or three operator genetic algorithm employing combinations of these operators and having specified algorithm parameters (i.e. population size, mutation and crossover probabilities), including the case in which the population size is fixed but the mutation and crossover probabilities are allowed to vary with the algorithm iteration index, is referred to herein as a simple genetic algorithm.

Other authors have employed Markov chains in examining genetic algorithm convergence behavior, notably [3] and [8] who use Markov chain methodology accompanied by approximate numerical analysis to examine genetic drift in finite population genetic algorithms, and [18] who develop their state transition matrix from the recombination and selection operators introduced in [21] and then employ it in examining steady state distributions as population size increases. Additionally, others have pursued unifying formalisms for simulated annealing and the genetic algorithm by introducing non-conventional GA operators [7, 4]. However, no convergence theory for the simple genetic algorithm comparable in scope to that of simulated annealing exists in the literature. The central theme of the work reported here is an attempt to develop a very general simple genetic algorithm model (and accompanying convergence theory) by directly extrapolating the simulated annealing convergence methodology.

An essential first step toward that goal is the development of a very general non-stationary Markov chain simple genetic algorithm model. That task is accomplished in the work synopsized here for simple genetic algorithm variants implementing three combinations of the primary operators. The resulting model provides a mechanism for employing the mutation prob-
ability parameter in a role analogous to the absolute temperature control parameter analog of simulated annealing. It is reviewed in Section 3 below.

Section 4 presents some empirical results generated from the algorithm model by computer simulation and Section 5 summarizes some key theoretical convergence results derived from the model. They include: (1) existence of a unique asymptotic probability distribution (stationary distribution) for the time-homogeneous algorithm variants which incorporate the mutation operator (the two and three-operator variants) with non-zero mutation probability, (2) formulation of the stationary distribution solution for the time-homogeneous two and three-operator algorithm variants in terms of the characteristic polynomials of matrices derived from the state transition matrix, (3) existence of a zero mutation probability limit for the time-homogeneous two and three-operator algorithm variants, (4) a mutation probability schedule bound (analogous to the annealing schedule bound of simulated annealing) sufficient for the non-stationary genetic algorithm variants to achieve (asymptotically) the limiting distribution and (5) a partially developed methodology for representing the stationary distribution components at all consistent values of mutation probability (including the zero mutation probability limit).

The simple genetic algorithm model and convergence results presented in this paper are synthesized from [2]. Space limitations prohibit replicating them in the generality developed there.

2 SIMULATED ANNEALING

Let a combinatorial optimization problem be represented by the pair $(S, C)$ where $S$ is the problem solution space and $C$ its objective function, and assume without loss of generality that the problem requires minimization of $C$. Then, a simulated annealing algorithm executing on this problem generates a sequence of candidate solutions drawn from $S$ by employing a stochastic state transition mechanism to transform each member of the sequence into a successor. The stochastic state transformation mechanism is selected such that the resulting asymptotic probability distribution over $S$ is the Gibbs (or Boltzmann) distribution

$$P_{\text{r}}(i) = \frac{\exp\left(-C(i)/T\right)}{\sum_{j \in S} \exp\left(-C(j)/T\right)}$$

with $C(i)$ playing the role of energy function for the analogous thermodynamic system and $T$ a strictly positive algorithm parameter analogous to absolute temperature.

As the parameter $T$ in Eq. 1 approaches zero, the probability distribution approaches a limit in which all states corresponding to sub-optimal solutions have zero probability, while the states corresponding to the optimal value of $C$ are uniformly distributed, i.e.

$$\lim_{T \to 0^+} P_{\text{r}}(i) = \begin{cases} \frac{1}{|S_{\text{opt}}|} & i \in S_{\text{opt}} \\ 0 & \text{otherwise} \end{cases}$$

where $S_{\text{opt}} = \{ i \in S : C(i) = C_{\text{opt}} = C_{\text{min}} \}$. Thus, if sufficient conditions on the state transformation mechanism are invoked to ensure that Eq. 1 represents the algorithm asymptotic state probability distribution, and if the parameter $T$ goes to zero asymptotically, then the candidate solution sequence probability distribution is (asymptotically) nonzero only for optimal solutions (Eq. 2).

Since the conditional dependence of each candidate solution in the sequence upon the sequence history is equal to its conditional dependence upon its immediate predecessor, the solution sequence evolves as a Markov chain. In fact, the algorithm can be represented by the quadruple $(S, i_0, \overline{P}_T, \tau)$ where $S$ is as defined above, $i_0$ is an initial solution selected from $S$, $\overline{P}_T$ is a state transition matrix describing the stochastic state transition mechanism and $\tau = \{ T_k \}$ is a sequence of parameter values. The state transition matrix at algorithm iteration $k$ depends upon the parameter value $T_k$, so in general the Markov chain is non-stationary.

The essence of the simulated annealing convergence theory is a set of sufficient conditions on the functional form of $\overline{P}_T$ to ensure that the asymptotic probability distribution of the stationary Markov chain corresponding to every fixed, strictly positive value of $T$ is given by Eq. 1, along with sufficient conditions on $\tau$ to ensure that the non-stationary algorithm achieves the limiting distribution (Eq. 2) asymptotically. Sufficient conditions on the functional form of $P_{\text{r}}$ are established in [15, 16] and elsewhere, and are not reviewed here. The topic of interest here is sufficient conditions on $\tau$, and consequently on the non-stationary algorithm behavior. The first condition is that the sequence have limit zero. Beyond that requirement, a sufficient condition on $\tau$ is obtained by enforcing strong ergodicity on the Markov chain and deducing a corresponding parameter sequence bound. Strong ergodicity of the non-stationary chain, along with the zero control parameter sequence limit, ensures asymptotic convergence to the zero temperature limit in Eq. 2 (i.e. convergence to global optimality). The methodology is thoroughly reviewed in [15]. Among the annealing schedule bounds so deduced are those of [5] and [17], both of which have the general form $T_k \geq K/\log(k)$.

3 THE MARKOV CHAIN GENETIC ALGORITHM MODEL

The sequence of candidate solution populations produced by the simple genetic algorithm is a realization of a stochastic process with finite state space, and
the stochastic operators which emulate biological system behavior possess the property that the conditional dependence of each population upon its predecessors in the sequence is completely determined by its conditional dependence upon its immediate predecessor population. Thus, the sequence of populations evolves as a Markov chain. The work synthesized in this paper includes a Markov chain model of one, two and three-operator variants of the algorithm and exploration of the state transition behavior of each variant. The following paragraphs summarize the essential model results reported in [2].

Describing and analyzing the operation of the simple genetic algorithm is facilitated by assuming that the underlying optimization problem is defined over a bit-string solution space. This assumption is not essential and sacrifices very little generality.

Let a combinatorial optimization problem be characterized by the ordered pair \((S, R)\) where \(S = \{0, 1\}^L\) is the problem’s solution space and where \(R\) is a strictly positive reward function, and assume, with no loss of generality, that the problem requires maximization of \(R\). Also, let a simple genetic algorithm designed to execute on this problem have fixed population size \(M\), let \(i \in S\) be interpreted as an unsigned integer \((0 \leq i \leq 2^L - 1)\), and let a generation be represented by \(\overline{m} = (m(0), m(1), \ldots, m(2^L - 1))\) where \(m(i)\) is the number of occurrences of solution \(i \in S\) in the population. Thus, in the parlance of combinatorial mathematics, \(\overline{m}\) is a distribution of \(M\) non-distinct objects over \(N = 2^L\) bins [12, 19], and the set of all such distributions, \(S' = \{\overline{m}\}\), is a suitable representation of the simple genetic algorithm search space. The cardinality of \(S'\) is given by

\[
N' = \text{card}(S') = \binom{M + N - 1}{M} = \binom{M + 2^L - 1}{M}.
\]

Since both \(L\) and \(M\) are finite, so is \(N'\).

Then, if \(\overline{m}_0\) is selected as an initial population, the simple genetic algorithm can be represented by the quadruple \((S', \overline{m}_0, \overline{P}_Q, \Gamma)\) where \(\overline{P}_Q\) is a state transition matrix (analogous to \(\overline{P}_T\) of the simulated annealing model) and \(\Gamma = \overline{Q}_k\) is a sequence of parameter vectors \(\overline{Q}_k = (p_m(k), p_i(k))\). The algorithm parameters \(p_m(k)\) and \(p_i(k)\) are respectively the mutation and crossover probabilities. In the work reported here, the mutation probability sequence is employed in a role analogous to absolute temperature in simulated annealing, and consideration is limited to monotone non-increasing sequences. In general, the only limitation on the crossover probability sequence is that its values are probabilities. However, in [2], consideration is limited to constant crossover probability sequences.

The solution evolves as a sequence \(\{\overline{m}_k\}\) of states \(\overline{m}_k \in S'\) in which the conditional dependence of \(\overline{m}_{k+1}\) on the sequence history is equivalent to its conditional dependence on \(\overline{m}_k\), and thus the solution sequence is a Markov chain. In general, the chain is not time-homogeneous (i.e. it is non-stationary), however it is time-homogeneous if the parameter vectors are constant.

The functional forms of the state transition matrices corresponding to three combinations of the genetic algorithm operators are developed and their associated state behaviors explored in [2]. The first case consists of a one-operator algorithm which employs only reproduction. The second is a two-operator variant which employs reproduction with mutation. Finally a three-operator algorithm which includes crossover is developed. Following is a summary.

The proportional reproduction only model variant corresponds to the case \(\forall k : \overline{Q}_k = (0, 0)\). In this case, the conditional probability of selecting a solution \(i \in S\) from a population described by the state vector \(\overline{m} \in S'\) is

\[
\forall i \in S, \forall \overline{m} \in S' : P_1(i|\overline{m}) = \frac{n(i) \times R(i)}{\sum_{j \in S} n(j) \times R(j)}.
\]

Thus, the conditional probability of the successor generation \(\overline{m}\), given that the present generation is \(\overline{m}\), is a multinomial distribution,

\[
\forall \overline{m}, \overline{m} \in S' : P_1(\overline{m}|\overline{m}) = \frac{M!}{\prod_{i \in S} m(i)!} \times \prod_{i \in S} P_1(i|\overline{m})^m(i) = \left( \frac{M}{\overline{m}} \right) \times \prod_{i \in S} P_1(i|\overline{m})^m(i)
\]

where the symbol

\[
\left( \frac{M}{\overline{m}} \right) = \frac{M!}{\prod_{i \in S} m(i)!}
\]

denotes the indicated multinomial coefficient. The transition probability matrix of the Markov chain representing the one-operator algorithm is composed of the array of conditional probabilities defined by Eq. 4,

\[
\overline{P} = [P_1(\overline{m}|\overline{m})].
\]

Since it is independent of the sequence index (i.e. the parameter vectors are constant), the one-operator Markov chain is time-homogeneous.

The set of states which represent uniform populations (i.e. the states \(\overline{m}_A \in S'_A \subset S'\) in which one component is \(M\) and all others are zero) are absorbing states of the Markov chain, because for any such state, \(P_1(\overline{m}_A|\overline{m}_A) = 1\) and the state cannot be escaped. \(N = 2^L\) such uniform population states exist, one associated with each \(i \in S\). It follows from Eq. 3 and 4 that \(\forall \overline{m} \in S' : P_1(\overline{m}_A|\overline{m}) > 0\), and consequently there are exactly \(N = 2^L\) absorbing
states. The corresponding rows of $\overline{P}$ contain 1 in the principal diagonal location and 0 elsewhere,

$$\forall \overline{m}_A \in S'_A : P_1(\overline{m}|\overline{m}_A) = \begin{cases} 1 & \overline{m} = \overline{m}_A \\ 0 & \overline{m} \in S' - \{\overline{m}_A\} \end{cases}$$  \hspace{1cm} (6)

It follows that the probability distribution given by the $N' \times 1$ vector $\overline{\pi}_{\overline{m}_A}$ whose $\pi_{\overline{m}_A} \in S'_A$ component is one is invariant with respect to the iteration index in the sense that

$$\overline{\pi}_{\overline{m}_A}^T \overline{P} = \overline{\pi}_{\overline{m}_A}^T.$$

Any such probability distribution is called a stationary distribution, or invariant distribution, of the Markov chain. In the case of the one-operator algorithm variant the stationary distribution is not unique because the corresponding vector for any of the $N = 2^L$ absorbing states satisfies the invariance requirement, as does any probability vector of the form $q = \sum_{\overline{m}_A \in S'_A} \pi_{\overline{m}_A} \overline{\pi}_{\overline{m}_A}$ where $\sum_{\overline{m}_A \in S'_A} \pi_{\overline{m}_A} \geq 0$ and $\sum \pi_{\overline{m}_A} = 1$.

The two-operator algorithm composed of reproduction and mutation corresponds to the case $V_k : Q_k = (p_m(k), 0)$ where $0 < p_m(k) < 1$. Results analogous to Eq. 3-5 are obtainable for this algorithm variant as follows. Let $P_2(i|\overline{m})$ and $P_2(\overline{m}|\overline{m})$ be the conditional probability of the two-operator algorithm corresponding to the one-operator distributions defined by Eq. 3 and 4. Then, $P_2(i|\overline{m})$ can be expressed as a sum over all $j$ of the corresponding $P_1(j|\overline{m})$ times a factor which accounts for the probability of the mutation event required to transform $j$ into $i$. This probability can be expressed as $p_m^{H(i,j)}(1-p_m)^{L-H(i,j)}$ where $H(i,j) = H(j,i)$ is the Hamming distance of the pair $i, j$, and thus $P_2(i|\overline{m})$ can be written as

$$\forall i \in S, \forall \overline{m} \in S' : \quad P_2(i|\overline{m}) = \sum_{j \in S} P_m^{H(i,j)}(1-p_m)^{L-H(i,j)} \times P_1(j|\overline{m})$$

$$= \frac{1}{(1+\alpha)^L} \times P_1(j|\overline{m}) \times \alpha^{H(i,j)}$$

where

$$\alpha = \frac{p_m}{1-p_m}. \hspace{1cm} (8)$$

The two-operator analog of Eq. 4 is

$$\forall \overline{m}, \overline{m}_A \in S' : \quad P_2(\overline{m}|\overline{m}_A) = \left( \frac{M}{\overline{m}} \right) \times \prod_{i \in S} P_2(i|\overline{m}_A)^{\overline{m}(i)}. \hspace{1cm} (9)$$

The admissible range of $p_m$ is $0 < p_m < 1$, and consequently that of $\alpha$ is $0 < \alpha < \infty$. The transition probability matrix of the Markov chain representing the two-operator algorithm is composed of the array of conditional probabilities defined by Eq. 9, i.e.

$$\overline{P} = [P_2(\overline{m}|\overline{m}_A)]. \hspace{1cm} (10)$$

Since the elements of $\overline{P}$ depend on $\alpha$ (and hence by Eq. 8 on $p_m(k)$), the two-operator Markov chain is generally not time-homogeneous. It is time-homogeneous if the mutation probability is fixed. It also follows from Eq. 7 and 9 that

$$\lim_{\alpha \to 0^+} P_2(i|\overline{m}) = P_1(i|\overline{m})$$

and

$$\lim_{\alpha \to 0^+} P_2(\overline{m}|\overline{m}) = P_1(\overline{m}|\overline{m}) \hspace{1cm} (11)$$

Since the reward function, $R$, is strictly positively by hypothesis, and since $\forall i, j \in S : 0 \leq R(i,j) \leq L$, it follows that for $\alpha$ in the range $0 < \alpha \leq 1$, which corresponds to $0 < p_m \leq 1/2$, then

$$\alpha^L \sum_{j \in S} n(j) \times R(j) \leq \sum_{j \in S} n(j) \times R(j) \times \alpha^{H(i,j)}$$

and consequently from Eq. 3 and 7 that

$$\forall i \in S, \forall \overline{m} \in S' : \quad \left( \frac{\alpha}{1+\alpha} \right)^L \leq P_2(i|\overline{m}) \leq \left( \frac{1}{1+\alpha} \right)^L. \hspace{1cm} (12)$$

Using Eq. 12 in Eq. 9 yields

$$\forall \overline{m}, \overline{m}_A \in S' : \quad \left( \frac{M}{\overline{m}} \right) \times \left( \frac{\alpha}{1+\alpha} \right)^{ML} \leq P_2(\overline{m}|\overline{m}_A) \leq \left( \frac{M}{\overline{m}} \right) \times \left( \frac{1}{1+\alpha} \right)^{ML}. \hspace{1cm} (13)$$

The lower bound in Eq. 13 can be used with the Perron-Frobenius theorem, which is fundamental to the study of non-negative matrices in general and stochastic state transition matrices in particular [20], to establish that the time-homogeneous two-operator algorithm variant possesses a unique stationary distribution [2]. That is, the vector $\overline{\pi}_0$ given by

$$\overline{\pi}_0^T \overline{P} = \overline{\pi}_0^T$$

$$\overline{\pi}_0^T \overline{I} = 1$$

and having strictly positive components exists and is unique. Further, $\overline{\pi}_0$ is the asymptotic state probability distribution of the corresponding Markov chain. It follows from Eq. 3 and Eq. 7-10 that it is completely determined by the objective function and the algorithm parameters, and is independent of the algorithm initial population, $\overline{m}_0$. The time-homogeneous two-operator stationary distribution is analogous to the asymptotic distribution in Eq. 1 for the simulated annealing algorithm.

The three-operator simple genetic algorithm corresponds to the case $V_k : Q_k = (p_m(k), p_s(k))$ with both
\( p_m(k) \) and \( p_e(k) \) non-zero. Results analogous to Eq. 7-11 are obtained in [2] by defining a crossover operator function which is similar in character to the Hamming distance function employed for the two-operator case. Further, the bounds in Eq. 12 and 13 apply without alteration, a consequence of which is that a unique stationary distribution exists for the three-operator algorithm also (Eq. 14). The details are omitted here in the interest of brevity.

The unique stationary distribution of the time-homogeneous two and three-operator algorithm variants (for all strictly positive values of the parameter \( \alpha \)) and the fact that both algorithms degenerate into absorbing state behavior at zero mutation probability (Eq. 11 and its three-operator counterpart) suggests a mechanism for adapting the simulated annealing convergence theory onto the non-stationary simple genetic algorithm model. Some theoretical consequences of its behavior derived from the model and reported in [2] are summarized in Section 5 below, but first some empirical data generated from a computer simulation of the model is presented.

4 SOME EMPIRICAL RESULTS

This section reports the results of some computer simulations based upon the simple genetic algorithm Markov chain model summarized above. The results reported here concern the converged limiting (mutation probability approximately zero) stationary distribution of the three-operator algorithm variant executing on a four bit optimization problem at two selected values of the population size parameter, \( M \). These results are a very small sample of the data reported in [2].

The underlying optimization problem for the data presented here is defined by the reward function presented in Figure 1. The solution state which maximizes the reward value is the \( i \in S \) represented by the decimal integer value 12.

Figures 2 and 3 represent converged three-operator stationary distribution results at population sizes \( M = 6 \) and \( M = 7 \) respectively, and with crossover probability one. These results are for extremely small \( \alpha \) (approaching zero). As shown in [2] and reviewed in Section 5 below, only the states corresponding to uniform populations (one-operator absorbing states) have non-zero probability in the \( \alpha \rightarrow 0^+ \) limit. Consequently, only the final probabilities for the uniform population states are displayed, with each such state indexed by the decimal integer value corresponding to the solution it represents. These results are generated by iteratively multiplying a starting probability vector by the state transition matrix until convergence (within a small tolerance) is obtained. In the two cases presented, 20 and 22 iterations respectively are required.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure1.png}
\caption{Four-bit Reward Function}
\end{figure}

A mildly surprising result suggested by these data is that the \( \alpha \rightarrow 0^+ \) limiting value of the stationary distribution is non-zero for all possible uniform states. This behavior, which is confirmed by theoretical results developed in [2] and reviewed in Section 5 below, has some consequences for the attempt to extrapolate the simulated annealing convergence theory onto the non-stationary simple genetic algorithm. It means that the algorithm cannot guarantee convergence to a globally optimal solution as does the simulated annealing algorithm, at least not for finite population size. However, as suggested by the data plotted in Figure 4, it may be possible to approach the desired limiting behavior as closely as required. That figure plots the computed limiting distribution entropy versus the algorithm population size parameter. The plots suggest that the entropy can be reduced arbitrarily close to zero. Results developed in [2] reinforce this premise, at least for the two-operator case.

5 SYNOPSIS OF KEY THEORETICAL RESULTS

Some key theoretical results required for extrapolating the simulated annealing convergence theory onto the simple genetic algorithm are obtained in [2] from the Markov chain model reviewed in Section 3 above. Fundamental to those developments is the formulation of the existence argument for the two and three-operator unique stationary distributions (Eq. 14) into a solution which allows a substantial degree of insight into the functional form of their components. The stationary distribution components are shown to be rational functions of both the objective function and the algorithm parameters. The rational functions are expressed in terms of the characteristic polynomials of a
Figure 2: Three-Operator Stationary Distribution Limit \((M = 6)\)

\[
M = 6 \\
N' = 54264
\]

Figure 3: Three-Operator Stationary Distribution Limit \((M = 7)\)

\[
M = 7 \\
N' = 170554
\]

Figure 4: Three-Operator Limit Distribution Entropy vs \(M\)

set of matrices, each of which is derived from the state transition matrix by setting a selected row to zero. That is,

\[
q(m) = \frac{\phi_{m}(\lambda)}{\sum_{s \in S'} \phi_{s}(\lambda)}_{\lambda=1}
\]  \hspace{1cm} (15)

where \(q(m)\) is the \(m\) component of \(\phi_{s}\) and where \(\phi_{m}(\lambda)\) is the characteristic polynomial of the matrix \(P_{m}\) derived from \(P\) by setting row \(m\) in \(P\) to the vector \(0^{T}\).

In terms of extrapolating the simulated annealing convergence theory onto the simple genetic algorithm model, the zero mutation probability limit of Eq. 15 is of extreme interest. Unfortunately, at \(a = 0\), Eq. 15 yields an indeterminate form because the characteristic polynomials of the matrices \(P_{m}\) are all identically zero at zero mutation probability. That is, 1 is an eigenvalue of the zero mutation probability limit of each \(P_{m}\). (The Eq. 15 result is developed on the premise that \(a > 0\).) Nevertheless, by exploiting the functional form of the components of the elements in the \(P_{m}\) matrices, [2] succeeds in translating Eq. 15 into a form which is determinate at \(a = 0\). The result is expressed in terms of matrix counterparts of the \(P_{m}\) derived by coalescing the absorbing state rows of \(P_{m}\) (i.e. the rows whose zero mutation probability limit is given by Eq. 6 and 11 for the two-operator case) into the set of adjacent states reachable by a single-bit mutation event. The result is of the form

\[
\lim_{\alpha \to 0^+} q_{\alpha}(m) = \begin{cases} 
\frac{\phi_{m}^{s'}(\lambda)}{\sum_{s'_{A} \in S'} \phi_{s'_{A}}^{s'}(\lambda)}_{\lambda=1} & m = m_{A} \in S'_{A} \\
0 & m \in S' - S'_{A}
\end{cases}
\]  \hspace{1cm} (16)

where each of the \(\phi_{m}^{s'}(\lambda)\) is non-zero when evaluated at \(\lambda = 1\) and each has the same algebraic sign.
Thus, the zero mutation probability limit of the time-
homogeneous two and three-operator stationary dis-
tributions exists. Further, it has strictly positive com-
ponents corresponding to each of the one-operator ab-
sorbing states and zero components elsewhere, con-
fiming the empirical result noted in Section 4.

The non-zero value of the uniform population states
in the zero mutation probability limiting stationary distribution means that the limiting condition neces-
sary for extending the simulated annealing global op-
timality result does not obtain. However, the con-
verged distribution entropy data presented in Section
4, and some further theoretical results developed in [2]
suggest that the desired limiting behavior can be ap-
proached by adjusting the algorithm population size
parameter, at least for the two-operator algorithm
variant.

Sufficient conditions on the mutation probability con-
trol parameter to ensure that the non-stationary algo-
rithm achieves the limiting distribution asymptotically
are also developed in [2]. The methodology consists of
exploiting the form of the matrices whose characteris-
tic polynomials are represented in Eq. 15 and 16 to
deduce some results concerning continuity of the sta-
nationary distribution components and their derivatives,
and using those results along with the lower bound
in Eq. 13 to replicate the methodology by which the
simulated annealing temperature schedule bounds in
[5] and [17] are developed. The product of that effort
is a monotonic lower bound on the sequence \{pm(k)\}
which, along with the condition that the sequence
has limit zero, is sufficient to ensure that the non-
stationary simple genetic algorithm attains (asymptot-
ically) the limit distribution described in Eq. 16.

The bound is given by

\[ p_m(k) \geq \frac{1}{2} k^{-\theta/\lambda^2} \]

It is asymptotically superior to the \( K/\log(k) \) bounds
for the simulated annealing algorithm.

Also reported in [2] is a partially developed frame-
work for representing the stationary distribution compo-
nents at all consistent fixed mutation probability pa-
rameter values, including the zero mutation proba-
bility limit. The development is very extensive and
is currently incomplete. However, it provides sub-
stantial insight into the functional form of the sta-
nationary distribution components. For example, the
two-operator stationary distribution component cor-
responding to the uniform population state for solu-
tion \( i \in S \) includes the numerator polynomial factor
\( R(i)^M \). Thus, the ratio of the two-operator stationary
distribution components corresponding to the uniform
population states for solutions \( i, j \in S \) includes the factor
\( [R(i)/R(j)]^M \) which, if \( R(i) < R(j) \), can be forced
towards zero by selecting \( M \) sufficiently large.

6 CONCLUSIONS AND FUTURE DIRECTION

The work reported in [2] and synopsized here attempts
to provide an algorithm model and accompanying con-
vergence theory for the simple genetic algorithm com-
parable in scope to that which exists for simulated
annealing. It includes a very general non-stationary
Markov chain model for algorithm variants incorpo-
rating one, two and three-operator combinations of
the three primary GA operators, and some key in-
termediate theoretical results derived from the model.
Specific theoretical results include (1) stationary dis-
tribution existence for the time-homogeneous two and
three-operator variants, (2) existence of a zero muta-
tion probability limiting stationary distribution and
(3) sufficient conditions, analogous to the simulated
annealing temperature schedule bounds, on the muta-
tion probability parameter sequence to ensure that the
non-stationary algorithm achieves the limiting distri-
bution.

The major incomplete task in extrapolating the sim-
ulated annealing convergence theory consists of reduc-
ing the stationary distribution existence results to an
explicit solution in terms of the objective function and
algorithm parameters (or at least to a result from
which usable bounding values can be obtained). A
substantial amount of effort is devoted to that task
in [2] and considerable progress obtained, along with
a proposed direction for continuing the work begun
there.

Acknowledgements

The authors wish to acknowledge the generous support of
the USAF Wright Laboratories to this work.

References

[1] L. Davis, editor. Genetic Algorithms and Sim-
ulated Annealing. Morgan Kaufman Publishers,

[2] T. E. Davis. Toward an Extrapolation of the Sim-
ulated Annealing Convergence Theory onto the
Simple Genetic Algorithm. PhD thesis, Univer-

Class of Genetic Adaptive Systems. PhD thesis,
University of Michigan, Ann Arbor, MI, 1975.

Global Convergence of Genetic Algorithms: On
Infite Markov Chain Analysis. First Inter-
national Workshop on Problem Solving from Na-
ture, Dortmund, Germany, 1990.

Gibbs Distributions and the Bayesian Restoration


