

**Finite Markov Chain Analysis
of
Genetic Algorithms with Niching**

Jeffrey Horn
Department of Computer Science
University of Illinois at Urbana-Champaign
Urbana, IL 61801-2996

IlliGAL Report No. 93002
February 1993

(Final version published in S. Forrest (Ed.),
*Proceedings of the Fifth International Conference
on Genetic Algorithms (ICGA 5)*
San Mateo, CA: Morgan Kaufman, 110–117)

Illinois Genetic Algorithms Laboratory
Department of General Engineering
University of Illinois at Urbana-Champaign
117 Transportation Building
104 South Mathews Avenue
Urbana, IL 61801-2996

Finite Markov Chain Analysis of Genetic Algorithms with Niching

Jeffrey Horn

Department of Computer Science
University of Illinois at Urbana-Champaign
1304 West Springfield Avenue
Urbana, IL 61801-2996
jeffhorn@uiuc.edu

Abstract

Finite, discrete-time Markov chain models of genetic algorithms have been used successfully in the past to understand the complex dynamics of a simple GA. Markov chains can *exactly* model the GA by accounting for all of the stochasticity introduced by various GA operators, such as initialization, selection, crossover, and mutation. Although such models quickly become unwieldy with increasing population size or genome length, they provide initial insights that guide our development of approximate, scalable models. In this study, we use Markov chains to analyze the stochastic effects of the “niching operator” of a *niched GA*. Specifically, we model the effect of fitness sharing on a single-locus genome. Without niching, our model is an absorbing Markov chain. With niching, we are dealing with a “quasi-ergodic” Markov chain. Rather than calculating expected times to absorption, we are interested in steady-state probabilities for positive recurrent states. Established techniques for analyzing ergodic Markov chains give us new insights into the dynamic nature of a niched GA. We explore the stability of the expected steady state distribution achieved by the niched GA. We demonstrate the “niching pressure” as a force separate from the forces of selection, drift, and mutation. Through visualization, we gain intuitions of the relationships among these separate forces. These results generalize beyond the fitness sharing algorithm to all types of GA optimization of *context dependent functions*. In any such function, the GA must find and maintain a diverse population of cooperative individuals rather than converging to the truly steady state of a uniform population.

1 Introduction

The genetic algorithm (GA) is a robust, stochastic optimization procedure that has found good solutions to some very hard problems. But GAs are also extremely complex systems. They have proven very difficult to analyze and understand. Most work in GA theory relies on approximate models whose assumptions greatly simplify the operation of the algorithm. In contrast to such approximations, Markov chains can model completely the behaviour of the GA. Unfortunately, the size of the required transition matrix grows exponentially in both the population size and the genome length to be modeled. Thus, such exact models of the GA can be used only on a small scale, or in an implicit fashion. Still, even small Markov models have led to important insights into the fundamental forces at work in the GA, such as genetic drift, selection pressure, and mutation. In turn, these insights guide the assumptions we must make for our subsequent approximate models.

Inspired by the high yield of these small models, we introduce another fundamental force to the Markov model: *niching pressure*. Niched GAs have received attention over the last five years because of their ability to find multiple good, diverse solutions. In a simple GA, (i.e., a GA without niching), selection pressure and genetic drift cause the GA to converge to a uniform population, consisting of copies of the best

solution (individual) found. With the addition of some kind of niching mechanism, such as fitness sharing, crowding, preselection, or Boltzman tournament selection, the GA tends to maintain a steady-state around a population distribution consisting of diverse, high-fitness individuals. The stability of such steady-states is difficult to ascertain. It is the goal of this study to extend the simple, but exact Markov models developed for simple GAs to gain some insight into the stability of these steady-states, the nature of the niching force, and the relationship of this force to the forces of convergence (genetic drift and selection pressure).

Although we focus on GAs with niching, the questions we attempt to answer are pertinent to any type of GA optimization of *context dependent functions*. Such functions require that an individual's fitness be evaluated in the context of the current population. That is, individuals affect each other's fitness. This is the case not only with niched GAs, but with learning classifier systems (LCS), immune system models, and artificial life. All such systems raise fundamental questions which have yet to be addressed, other than empirically:

- Can the GA achieve a steady-state distribution of the population?
- If so, how long can we expect to be in such a steady-state?
- How noisy is the steady-state?
- How long until the GA achieves it?
- What are the probabilities of convergence to competing steady-states?

We hope that a simple Markov model will get us started answering these questions analytically.

2 Background

In this section, we show how previous work has used Markov chains to model the simple GA. We then briefly review the niched GA, which has not yet been modeled by Markov chains.

2.1 Markov models of GAs

Markov chains can exactly model each generation of a GA by combining the effects of the various stochastic sources (such as initial population generation, selection, crossover, mutation, and even noisy fitness functions) into a single transition probability matrix. The major drawback to this approach is that for realistic population sizes and string lengths the transition matrices quickly become large and unwieldy. Thus, to date, the few studies that have used Markov modeling have necessarily kept to artificially small population sizes and string lengths (e.g., Goldberg & Segrest, 1987), or else have worked with matrix notation only, avoiding the generation and direct manipulation of the matrices themselves (e.g., Nix & Vose, 1992). We discuss the Goldberg and Segrest (1987) model in some detail, as it is their model for a simple GA that we extend to include niching.

2.1.1 The Goldberg and Segrest model

Goldberg and Segrest (1987) kept their model manageable by dealing with only a “single-locus” genome. That is, the genome consists of only one binary position. Thus the only possible individuals are “1” and “0”. Such a limitation allows for an intuitive numbering of states. Given a fixed population size of N , there are $(N + 1)$ possible states, where state i is the population with exactly i ones and $(N - i)$ zeros¹. Goldberg and Segrest defined a transition matrix $P[i, j]$ mapping the current state i to the next state j . Such a matrix is thus $(N + 1)$ by $(N + 1)$.

We repeat here Goldberg and Segrest's calculation of the transition probabilities for their model of a single-locus, simple GA using proportionate selection with no mutation. Under proportionate selection, we choose a member k of the current population to reproduce (i.e., to be in the next population) with probability proportional to its fitness relative to total fitness of the population. Thus, with probability

¹Notice that we do not care about “location” in the population.

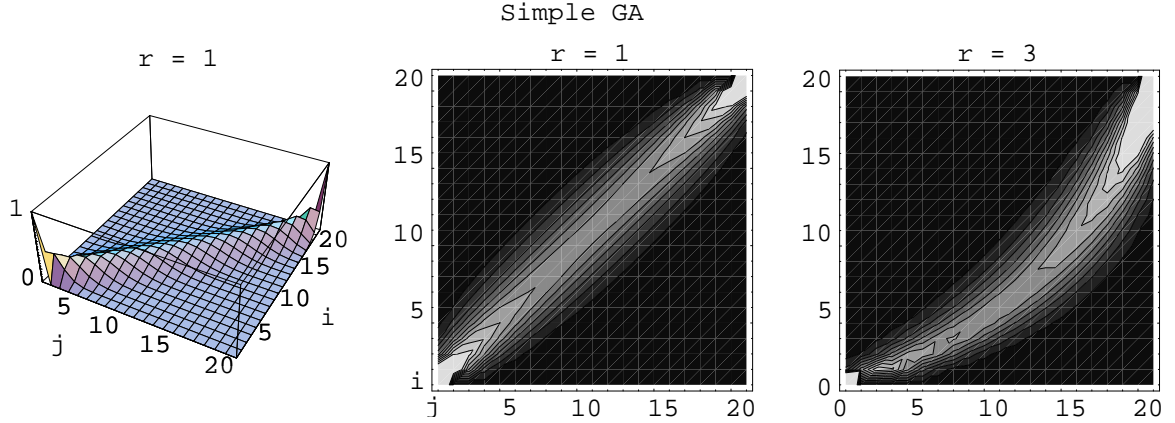


Figure 1: The transition probabilities for the simple GA. On the left is a surface plot for the case where $r = 1$ (i.e., no selection pressure), while at the center is a contour plot of the same matrix. On the right is the matrix for the case where f_1 is three times that of f_0 . “i” is the current state, while “j” is the next state. Each state corresponds to the number of “1”s in the current population. Here the population size $N = 20$. Note the absorbing states $i = 0, N$.

$f_k / \sum f$, we choose individual k (where f_k is the fitness of k and $\sum f$ is the sum of the fitnesses of all individuals in the current population).

In the *generational* GA, we replace the entire population each generation, thus making N selections per generation. In our single-locus model, we can write the denominator $\sum f$ as simply $i * f_1 + (N - i) * f_0$, where i is the number of ones in the current population, f_1 is the fitness of “1”, and f_0 is the fitness of “0”. Then the probability of choosing a one for the next generation’s population is p_1 :

$$p_1 = \frac{i * f_1}{i * f_1 + (N - i) * f_0} \quad \Rightarrow \quad p_1 = \frac{r * i}{r * i + (N - i)}$$

Where r is the ratio f_1/f_0 . The probability of choosing a zero, p_0 , can be similarly calculated:

$$p_0 = \frac{N - i}{i * r + (N - i)}$$

The probability of going from a state with i ones, and $(N - i)$ zeros, to a state with j ones is $p_{(i,j)} = \binom{N}{j} (p_1)^j (p_0)^{N-j}$. Substituting for p_1 and p_0 :

$$p_{(i,j)} = \binom{N}{j} \left(\frac{i * r}{i * r + (N - i)} \right)^j \left(\frac{N - i}{i * r + (N - i)} \right)^{N-j} \quad (1)$$

Equation 1 defines a complete $(N + 1)$ by $(N + 1)$ transition matrix for any population size N and fitness ratio r . In Figure 1, on the left, we plot the transition probabilities for a population of size 20, and a fitness ratio of $r = 1$. The center matrix is a contour plot of the surface plot at the left. The centering of the distributions on the main diagonal is due to the fact that $r = 1$. There is no selection pressure toward the all-ones or all-zeros states. With $r = 1$, Equation 1 reduces to the equation for pure *genetic drift*, a major focus of the Goldberg and Segrest study. Drift is a concern for the simple GA because it means the GA will eventually converge to a uniform population ($i = 0, N$), when there is no justification.

Another important feature to note in Goldberg and Segrest’s model of selection pressure alone is that the two states 0 and N , corresponding to all-zeros and all-ones respectively, are absorbing states, and thus have transition probability rows and columns consisting of a single spike of probability one at $i \rightarrow i$ ($p_{(i,i)} = 1$) and zeros elsewhere. Goldberg and Segrest used Equation 1 to investigate expected times to absorption for the drift case ($r = 1$). They also looked at different values of r , where the GA should converge to the better individual.

For our purposes, we are interested in visualizing the force of selection when there is a preference ($r \neq 1$), before we add niching pressure. Figure 1, right, shows the transition matrix for $r = 3$. Notice how the “ridge” of higher probabilities moves off the main diagonal when $r \neq 1$, thus favoring the higher fit individual. The presence of the ridge in the lower or upper triangles of the matrix indicates a pressure, or force, toward more or less ones, respectively. Notice how this force is consistent in always favoring the better individual. However, near the absorbing states the “preferential selection force” (due to the term r in Equation 1) is overwhelmed by the effect of drift (due to the term i).

2.1.2 Other Markov models

Two recent papers extended the Goldberg and Segrest model by allowing genome sizes (string lengths) greater than one (Davis & Principe, 1991; Nix & Vose, 1992). With string length l , the number of possible binary strings is 2^l , and the number of possible states (distributions of a population of size N over such a 2^l gene space) is $\frac{(N+2^l-1)!}{N!(2^l-1)!}$. This number grows exponentially with either N or l , so that the transition matrix for a realistic GA implementation could not be generated, let alone manipulated and analyzed. However, Nix and Vose worked with the matrix notation directly, rather than generating the actual probabilities. By assuming infinite population sizes and/or genome lengths, they described asymptotic behaviours for a simple GA. Similarly, Davis and Principe were able to develop the outlines of a theoretical proof of convergence for a simple GA. It should also be noted that by allowing genomes of length greater than one, both groups of researchers were able to incorporate the crossover operator in their models. A single-locus genome, of course, does not support meaningful crossover. Finally, Mahfoud (1991) adopted the extended models of Nix and Vose, with some simplifying assumptions, such as partitioning the 2^l gene space into a much smaller number of equivalence classes of similar individuals. Mahfoud then analyzed the selection method known as Boltzmann tournament selection, which achieves some degree of niching.

2.2 Niche GAs

Although we are far from understanding exactly how a simple GA works, it is clear that the simple GA does not tap all the power of the schema processing going on in the selection/crossover/mutation cycle. Consequently, many extensions to the GA have been made to exploit the additional information made available with each generation’s fitness evaluations. In particular, a number of different GA enhancements are aimed at preserving the information available across a diverse population. These *niched GAs* avoid the loss of information that results from convergence of the population to a single, high-fitness individual. Rather, the niched GA seeks to maintain several subpopulations, or species, of individuals at different good solutions. When we use a niched GA we are asking the algorithm to tell us more about the fitness landscape than what the best solution is. Niched GAs are particularly useful for finding and maintaining a set of mutually supportive solutions, such as in an LCS, where the GA searches for a set of rules which together implement a successful classification strategy.

Each peak in the fitness landscape, or region of high fitness, can be considered a “niche”. We would like the niched GA to find the best niches, and fill such niches with individuals in proportion to the quality of each niche.

A number of niching mechanisms have been proposed and used over the last couple of decades. One of the earliest was Cavicchio’s *preselection* (Cavicchio, 1970; Mahfoud, 1992), in which offspring could only replace one of their parents. DeJong’s *crowding* (DeJong, 1975; Mahfoud, 1992) had the same flavor, in that new individuals replaced less-fit, but similar, solutions in the old population. *Boltzmann tournament*

selection has also been shown to have niching effects (Mahfoud, 1991), and recently *immune system models* (Smith, Forrest, & Perelson, 1992) have been gaining attention as for maintaining multiple solutions. In this paper, we limit our study to *fitness sharing*, introduced by Goldberg and Richardson (1987), studied in detail in (Deb, 1989), and used extensively in (Goldberg, Deb, & Horn, 1992).

Fitness sharing accomplishes niching by degrading the *objective* fitness (i.e., the unshared fitness) of an individual according to the presence of nearby (similar) individuals. Thus this type of sharing requires a distance metric on the phenotype or genotype of the individuals. In this study, we use the Hamming distance between the binary encodings (genotypes) of individuals. We degrade the objective fitness of an individual by first summing all of the *share values* of individuals within a fixed radius, called σ_{share} , of that individual, and then dividing the objective fitness value by this sum, which is known as the *niche count* for that individual. Thus, if two individuals, h and k , are separated by Hamming distance d_{hk} , and $d_{hk} < \sigma_{share}$, then we add a share value, $sh(d_{hk}) = 1 - (\frac{d_{hk}}{\sigma_{share}})^\alpha$ to both of their niche counts, m_h and m_k . Here, σ_{share} is the radius of our *estimated niches*. Individuals separated by σ_{share} , or more, do not degrade each other’s fitness (that is, $sh(d_{hk}) = 0$ for $d_{hk} \geq \sigma_{share}$)². For an individual h , the niche count m_h is calculated as $m_h = \sum_{k=1}^N sh(d_{hk})$, where N is the size of the population. The *shared fitness* of individual h is then given by f_h/m_h .

Sharing tends to spread the population out over multiple peaks (niches) in proportion to the height of the peaks. GAs with proportionate selection and fitness sharing have been successfully used in solving a variety of multimodal functions (Deb, 1989).

Goldberg and Richardson (1987) note that for fitness sharing to be at equilibrium, it must be true that the shared fitness values of the individuals at each local optima h are equal: $\frac{f_h}{m_h} = \frac{f_k}{m_k}, \forall h, k$. If this were not satisfied, then at least one niche would be “underfilled”. If we assume non-overlapping niches centered on local optima, then given a population size N , and knowing the number and objective fitness of the local optima f_h , the above equations allow us to calculate the expected distribution of the population (i.e., the m_h) when the niched GA is at equilibrium. For our single-locus model, with $\sigma_{share} \leq 1$, $m_1 = i$, where i is the index of the current state, and $m_0 = N - i$. The steady-state equation above then reduces to

$$\frac{f_1}{i_{eq}} = \frac{f_0}{N - i_{eq}} \quad \Rightarrow \quad i_{eq} = \frac{f_1 N}{f_0 + f_1} \quad \Rightarrow \quad i_{eq} = \frac{rN}{1 + r} \quad (2)$$

Later, we compare i_{eq} , the number of ones at equilibrium predicted by Equation 2, with the distribution shown by Markov models.

3 Modeling the niched GA

In this section we add “niching pressure” to the Goldberg and Segrest model. We start out simply, assuming perfect fitness sharing. We then look at various cases of overlapping niches.

3.1 Perfect Fitness Sharing

To add in the effect of niching, we must degrade each of the two fitnesses, f_0 and f_1 , by the existence of nearby individuals. We begin with the simple case of “perfect sharing”, such that the niches centered on each kind of individual do not overlap³. Thus, $\sigma_{share} \leq 1$, so that ones do not degrade f_0 and zeros do not degrade f_1 . And with $\sigma_{share} \leq 1$, we are no longer concerned with α , since for any setting of α , the contribution of an individual to its own niche count is always 1.

²The parameters α and σ_{share} are chosen by the user of the niched GA based on some knowledge of the fitness landscape to be searched. Commonly, $\alpha = 1$, yielding the *triangular sharing function*. Choosing σ_{share} is trickier, since we need to have some idea of the size and separation of our expected niches.

³We can imagine a simple biological analogue, in which isolated niches filled with identical competitors evenly split a niche’s resource among themselves.

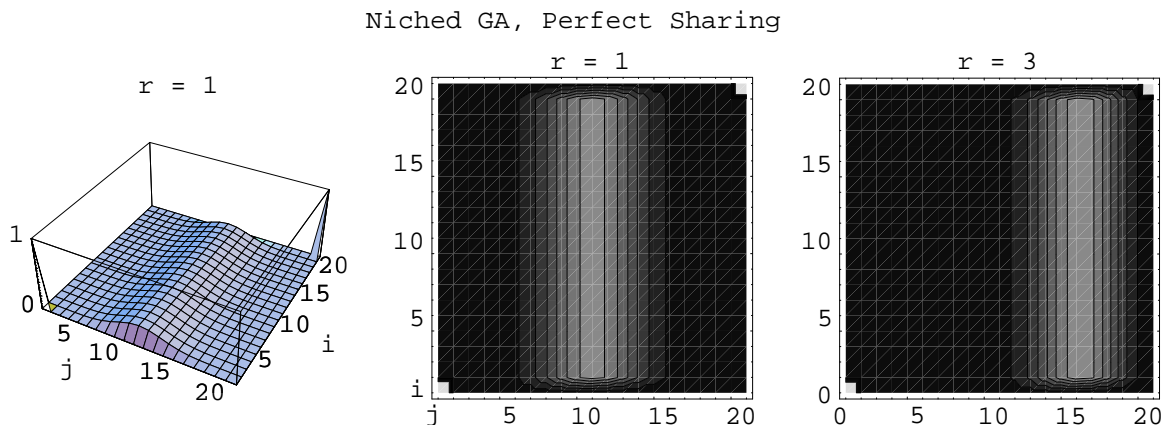


Figure 2: The transition probabilities for the niched GA (with perfect fitness sharing), population size $N = 20$. Note how the probabilities are independent of “ i ”, the current state, except when “ i ” is an absorbing state ($i = 0, N$). The left matrix assumes fitness ratio $r = 1$. In the center is a contour plot of the same matrix. On the right, f_1 is three times f_0 .

With niche counts $m_1 = i$ and $m_0 = N - i$, the shared fitness value for the ones is then f_1/i , while for the zeros it is $f_0/(N - i)$. If we substitute these shared fitness values for the objective fitness values f_1 and f_0 in Equation 1, we obtain the transition probabilities⁴:

$$p(i,j) = \binom{N}{j} * \left(\frac{r}{r+1}\right)^j * \left(\frac{1}{r+1}\right)^{N-j} \quad (3)$$

Let us see how perfect fitness sharing has affected our transition probability matrix (Figure 1). On the left of Figure 2 is the transition probability matrix for a GA with perfect fitness sharing, population size $N = 20$, and fitness ratio $r = 1$. In the center is a contour plot of the matrix at the left. Interestingly, perfect fitness sharing exactly counters the benefits of quantity bestowed by proportionate selection. Each additional copy of a zero in the population, for example, degrades f_0 but also increases the probability of selection under proportionate selection, resulting in the elimination of the current state i from the transition probability equation. So the row independence of the matrix for fitness sharing is not due to the removal of a complex force, but rather the addition of a new one (niching) that balances the old (selection/drift). For our simple model at least, perfect sharing combined with proportionate selection yields an ideal restorative force.

From Figure 2, center, we can say that in general, the effect of the niching pressure is to move the population toward the center states, where the number of ones and zeros is balanced. This is as we expect given that $r = 1$. When there are more ones than zeros we are likely to move towards states of lower numbers of ones. The reverse is true of states with high numbers of zeros. Thus we can interpret the niching force as being a stabilizing one, with a point of equilibrium far from the poles toward which selection and drift drive us. The apparent point of equilibrium is at the state $i = 10$, where half the population is ones and the other half zeros. This is as we expect given that $r = 1$. What about for other r values? At the right of Figure 2 is the matrix for $r = 3$. Note how the apparent point of equilibrium, that

⁴Note that Equation 3 cannot be used where $i = 0$ or $i = N$, since one of the quotients $\frac{f_1}{i}$ or $\frac{f_0}{(N-i)}$ does not exist there. These cases correspond to the absorbing states of all ones or all zeros. Thus for these states we already know the row of transition probabilities (all zeros except for a single 1).

is the peak of the probability ridge, is approximately where Equation 2 predicts, namely at $i = 15$. Later, we will investigate this potential empirical confirmation of niched GA theory.

3.2 Overlapping niches

It is often the case that niches overlap. In fitness sharing, we easily might estimate σ_{share} to be too large, sometimes on purpose (Goldberg, Deb, and Horn, 1992). In niched GAs and in life, while two species compete to fill a niche, the niches centered on them overlap. In this section, we examine the general case where $\sigma_{share} > 1$.

The calculation of transition probabilities for $\sigma_{share} > 1$ is somewhat more complex than for $\sigma_{share} \leq 1$, but is still straightforward. The niche count for ones, m_1 , is now $i + (N - i)(1 - 1/\sigma_{share})$ instead of just i , since we have to add in the share value of each zero: $1 - 1/\sigma_{share}$. Similarly, $m_0 = N - i + i(1 - 1/\sigma_{share})$. Substituting the new degraded fitnesses f_1/m_1 and f_0/m_0 into the equations for p_1 and p_0 , dividing through by f_0 to obtain r , and after some rearranging, we obtain the new transition probability equation⁵:

$$p_{(i,j)} = \binom{N}{j} * \left(\frac{1}{1 + \frac{(N-i)(N+\frac{i-N}{\sigma_{share}})}{ir(N-\frac{i}{\sigma_{share}})}} \right)^j * \left(\frac{1}{1 + \frac{ir(N-\frac{i}{\sigma_{share}})}{(N+\frac{i-N}{\sigma_{share}})(N-i)}} \right)^{N-j} \quad (4)$$

In Figure 3 we show the transition matrices resulting from Equation 4 for a population size $N = 20$, fitness ratio $r = 1$, but for three different degrees of overlap. It is clear that for very small overlaps ($\sigma_{share} \approx 1$), the niching pressure dominates and the matrix is very similar to that of perfect sharing. As the overlap increases, the effect of niching is diminished, most especially near the absorbing states, and genetic drift comes to dominate. Even for $\sigma_{share} = 3$, our matrix appears not significantly different from the case of pure genetic drift. This gradual transition from ideal niching to genetic drift agrees with intuition. When $\sigma_{share} = 1$, Equation 4 reduces to Equation 3 for perfect sharing. As $\sigma_{share} \rightarrow \infty$, the entire search space becomes a single niche, and we approach the case of genetic drift (or selection pressure, when $r \neq 1$).

If the effect of fitness sharing degrades gracefully with increasing error in σ_{share} (overestimation), does the stability of our assumed steady state also decay gracefully? To our list of questions about the nature of the supposed equilibrium, we can add this question of how the steady state degenerates as we move away from the ideal niching situation.

3.3 Absorbing Markov chain?

Before we can talk about steady states, we must address the absorbing states. In Markov chain analysis, the only steady states of an absorbing chain are absorbing states (Hillier & Lieberman, 1990). Their existence means that eventually the niching force must be overcome, despite its strength made obvious in Figure 2. The GA will be “trapped” into convergence.

But in practice we do not wait around for the niched GA to converge to a uniform population. Implementors and users of niched GAs know that they are seeking a noisy steady-state. Often, it is a matter of watching the population distribution for many generations and deciding that a run has “converged”. Although the transition matrix tells us that this steady state cannot last, we usually do not wait around long enough to see the GA wander far from equilibrium.

So how do we reconcile our experience with our model? How can we talk about steady-states other than absorbing states when we have an absorbing Markov chain? One possible answer is to ignore the absorbing states and just analyze the transient states. In the well-known partitioning of states of an absorbing Markov chain,

$$P = \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix}$$

⁵For now we assume that $\alpha = 1$, yielding the triangular sharing function. For the single-locus model, varying α has the same effect as varying σ_{share} , since $(\frac{d_{hk}}{\sigma_{share}})^\alpha = \frac{1}{\sigma_{share}^\alpha}$ when $d_{hk} = 1$.

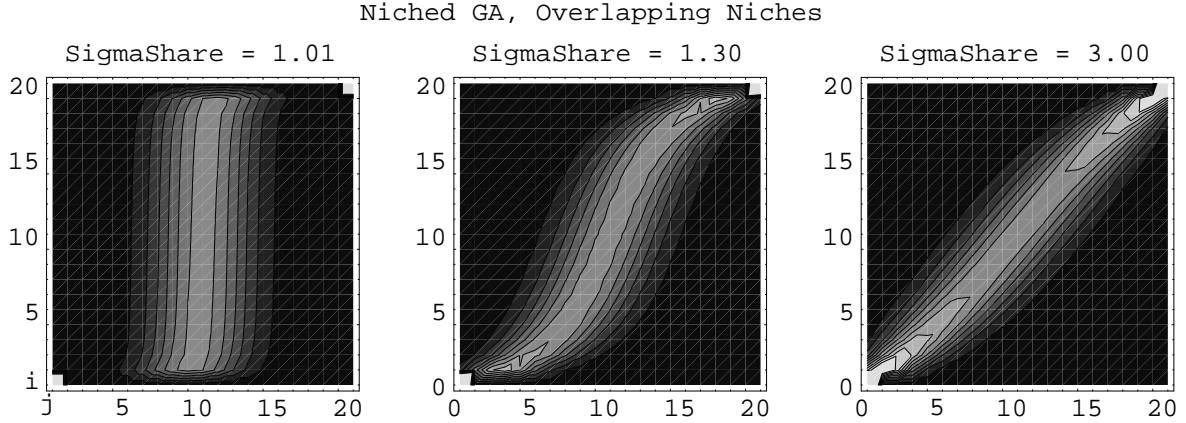


Figure 3: The transition probabilities for the niched GA with overlapping niches. The fitness ratio $r = 1$, so there is no preference toward either individual, and population size $N = 20$. Note the gradual degradation from near-perfect fitness sharing (SigmaShare = 1.01) to almost pure genetic drift (Sigma = 3.00).

we take the Q partition to be the entire matrix, ignoring $R, 0, I$, which only consist of two rows and two columns for the two absorbing states. If we normalize the Q matrix so that each row sums to 1, the resulting matrix, call it Q_{norm} , is an ergodic Markov chain and allows us to calculate steady-state probabilities for *all* (non-absorbing) states.

Before proceeding to analyze Q_{norm} , we must justify “chopping off” the absorbing state rows and columns. As an intuitive argument for ignoring the absorbing states, we look at the expected time to absorption. Absorption time is equivalent to the expected first passage time for the absorbing states. Expected absorption times for niched GAs should be much longer than those for the simple GA.

We calculate expected times to absorption from the Q matrix as in (Goldberg & Segrest, 1987). In short, we calculate the visitation matrix $V = (I - Q)^{-1}$, where “ I ” is the identity matrix and “ -1 ” is the inverse operator for matrices, and assume a uniformly randomly generated initial population. Goldberg and Segrest calculated absorption times for a number of scenarios. For our purpose of comparison, we recall only their results for the case where $N = 20$, $r = 1$, and no mutation. In a simple GA, Goldberg and Segrest found a linear relationship for absorption time as a function of population size⁶. We show this relationship as the bottom-most plot in Figure 4.

We calculate the expected absorption times for perfect fitness sharing, also using $N = 20$, $\sigma_{share} = 1$, no mutation, and random initial populations. In Figure 4, these times are plotted for three different values of r . Note how even with $r = 3$, where the fitness of “1” is three times that of “0”, absorption time appears to grow exponentially with population size. Although this growth will decrease with very high or very low values of r , it appears that this growth is always exponential. Just looking at Figure 4, it is apparent that as population size increases, absorption time approaches infinity.

This is not, however, a rigorous justification for using Q_{norm} . Fortunately, the problem of calculating steady state probabilities for “near-ergodic” absorbing Markov chains has received some attention in the applied probability literature. In (Darroch & Seneta, 1965), the authors examine several alternative approximations to the “effectively ultimate distribution”. A thorough treatment of their work in relation to

⁶Note that for any $r \neq 1$, the expected time to absorption (by either absorbing state) will be shorter, assuming a random initial population. So absorption times for $r = 1$ are upper bounds for all r .

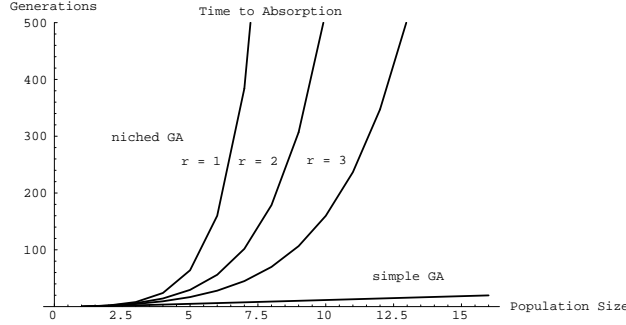


Figure 4: Expected times to absorption, where time is measured by generations, for various population sizes. We assume $\sigma_{share} = 1$, and random initial populations. Note how absorption times grow linearly with population size for the simple GA, while such growth is apparently exponential for GAs with fitness sharing, even with a significantly superior global optimum ($r = 3$).

our matrix is beyond the scope of this paper. The interested reader is invited to compare two of their “best” candidate approximations, namely the *stationary conditional distribution*, \vec{v} , and the *quasi-stationary probability*, $w_j v_j$, to the steady states we obtain for Q_{norm} . The stationary conditional distribution for Q is the left eigenvector, \vec{v} , of Q corresponding to the maximum modulus eigenvalue λ_{max} . The quasi-stationary probability is the product of the right and left eigenvectors, \vec{w} and \vec{v} , of Q corresponding to λ_{max} .

We have compared the steady state vector $\vec{\Pi}$ of Q_{norm} (which we show how to calculate in the next section) to \vec{v} and $w_j v_j$. Briefly, we found that for perfect sharing all three distributions are exactly the same. As σ_{share} increases beyond 1, the three diverge. Their convergence at $\sigma_{share} = 1$ follows from the row independence of Q for perfect sharing. $\vec{\Pi}$ for a row identical matrix is simply the row vector. The left eigenvector of λ_{max} is the same row vector, while the corresponding right eigenvector \vec{w} is the all ones vector, so that $w_j v_j = v_j, \forall j$. The divergence of the three distributions appears to grow very slowly with increasing σ_{share} , although this growth should be more carefully studied and bounded.

Since $\vec{\Pi}$ is in close agreement with the measures from Darroch and Seneta, and because it is much more quickly calculated than are eigenvectors, we use them in our study of niched GA steady states.

3.4 Ergodic Markov chain

We now have an irreducible Markov chain, Q_{norm} , with all ergodic states, 1 through $(N-1)$. Calculation of the steady-state probabilities is straightforward. We seek the vector of steady state probabilities $\vec{\Pi} = \{\pi_1, \pi_2, \dots, \pi_{N-1}\}$, where π_j is the steady state probability for state j . To find $\vec{\Pi}$, we solve⁷ ($\vec{\Pi} Q = \vec{\Pi}$) for $\vec{\Pi}$. We can analyze the vector $\vec{\Pi}$ to help us understand the behaviour of our niched GA at steady state. Figure 5 plots the steady state distributions for a population size $N = 10$, fitness ratio $r = 1$, and four different values of σ_{share} . As expected, all four distributions are symmetric about the equilibrium point of $i = 5$. We note that even with perfect fitness sharing, our steady state for this small population size is fairly noisy. And as σ_{share} increases, the steady state distribution flattens. We could perform a number of statistical analyses of the distribution, estimating the total expected amount of time spent within a certain range of states, or the expected recurrence time for returning to a particular level of ones in the population.

⁷For a unique solution, we must also use the fact that $\sum_{i=1}^{N-1} \pi_i = 1$.

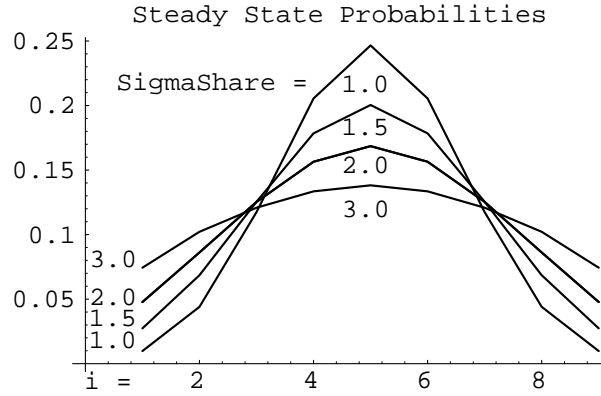


Figure 5: Steady state probability distributions, $\tilde{\Pi}$, for a population size $N = 10$, fitness ratio $r = 1$, and various degrees of overlapping niches: $\sigma_{share} = 1.0, 1.5, 2.0, 3.0$. Note the approximate normal distribution, and how its variance increases with σ_{share} .

In particular, these distributions can be approximated by a normal distribution, with mean approximately equal to the number of ones at equilibrium. We could then use the variance, σ^2 , as an inverse measure of the stability of the steady state. Clearly, σ^2 is directly related to σ_{share} .

For now, we use the steady state distribution to calculate the *expected number of ones*, $E[i]$, in the population at steady state, and compare this to the equilibrium prediction from Equation 2. We can think of $E[i]$ as the number of ones we expect to see in the population at any distant point in the future. To calculate the expected number of ones (in other words, the expected state, given our indexing of states by the number of ones in the population), we simply multiply the steady state probability for each state by the index for that state and sum over all states $E[i] = \sum_{i=1}^{N-1} i * \pi_i$.

Figure 6 plots $E[i]$ for perfect sharing, population size $N = 20$ but over different values of r , the fitness ratio. Interestingly, these values agree exactly with the values predicted by Equation 2. So the niched GA as modeled by the ergodic Markov chain achieves exactly the steady-state sought and predicted by fitness sharing theory, assuming perfect sharing. In addition to this confirmation, we also now have information on the error in this expectation (that is, the noisiness of the steady-state), which is provided by $\tilde{\Pi}$.

Although not shown in Figure 6, we have also calculated $E[i]$ for overlapping niches. We've found that for $r \neq 1$, as σ_{share} increases, the resulting $E[i]$ diverges from the equilibrium prediction. Also, for any value of $\sigma_{share} > 1$, increasing r while holding σ_{share} constant also increases the difference between $E[i]$ and the equilibrium prediction. Intuitively, as the effect of fitness sharing is diminished, the relative attraction, due to drift, of the nearest absorbing state skews the steady state distribution. More simply, the steady state degrades and allows the asymmetric forces of selection pressure and drift to be felt.

4 Extensions

All of the extensions considered in previous Markov chain analyses of simple GAs could be made to the model for niched GAs, such as adding mutation, lengthening the genome, and using other selection algorithms. It is particularly important to model other niching algorithms, to allow us to generalize these results to context-dependent function optimization.

Lengthening the genome even by a few bits quickly leads to very large matrices, but would allow us more niches, with varied separation. This in turn would allow us to study the tradeoff of quality versus quantity: finding solutions that are good *and* diverse. (Deb, Horn, & Goldberg, 1992; Goldberg, Deb, & Horn, 1992) have shown that the location, separation, quality, and quantity of local optima (niches) can make life difficult for a niched GA with fixed- σ_{share} fitness sharing. In addition, a longer genome would allow us to add the crossover operator to our model.

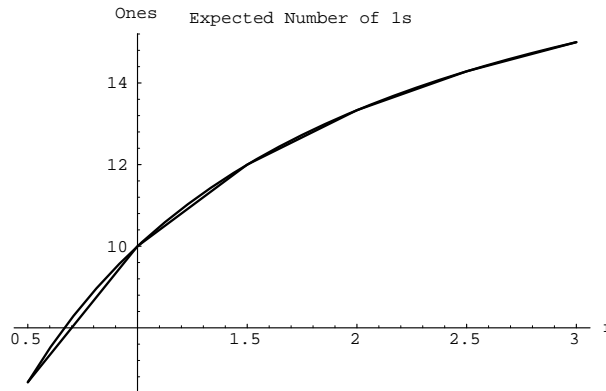


Figure 6: Expected number of ones for various values of r , with $N = 20$. The crooked line connects points at which $E[z]$ was evaluated using the Markov model, while the smooth line is the equilibrium prediction i_{eq} from Equation 2. Note the exact agreement for the plotted points, and presumably for all points in between.

But the first order of business should be the definition of a *steady state characteristic*, call it Π_c . Clearly, the variance of the steady state distribution, the time to absorption, the deviation from equilibrium prediction, and the time to convergence to steady state are all related. All of these measures “suffer” as we increase r or σ_{share} , or decrease the population size N . It seems worthwhile to seek a Π_c in terms of which all of the above measures and effects could be expressed. Such a Π_c would enable us to characterize and compare all steady states for a niched GA, and might help us define what “convergence” to a steady state means.

Finally, from the graphs of the transition matrices, it appears that we can visualize the separate, fundamental forces of genetic drift (one component of the overall selection force), preferential selection (the other component), and niching pressure. The arrows in Figure 7 indicate our speculations of the general effects of these three forces. The three forces appear to be orthogonal, suggesting that they might be represented by orthogonal tensors. Again, the next step in investigating these intuitions is to look for the same fundamental forces in other combinations of selection and niching algorithms.

5 Conclusions

In this paper we extended the work of Goldberg and Segrest (1987) by modeling a niched GA using finite, discrete-time Markov chains. Although our model was very simple, with only a single-locus genome, we gained several insights into the nature of the force exerted by the fitness sharing mechanism. In particular, we saw how exactly the niching force of fitness sharing balanced the preference and drift effects of proportionate selection. One major effect of fitness sharing is to create exponentially long waiting times to absorption, inducing a quasi-steady-state away from the absorbing states. We also saw how the steady-state predictions from fitness sharing theory accurately predicted the steady state or point of equilibrium, but gave no indication of how noisy or unsteady this state actually is. However, the steady-state probabilities given by our Markov chain, which are approximately normally distributed, provide the data for studying the stability of such dynamic equilibria. For example, we observed evidence for a single characteristic measure of stability by noting the effects of niching parameters on the variance of the steady state distribution. Finally, we had some success in isolating and visualizing some of the fundamental forces at work in a niched GA. It is important to find ways of breaking down the complex dynamics of a simple or niched GA into components that can be better understood in isolation, before studying them in context.

A final comment on methodology: we should build these small, exact models because we can. The larger, more realistic exact models can never be created. Instead, we must learn how to build approximate models for long genomes and large populations. But insights from the tiny models are important to that

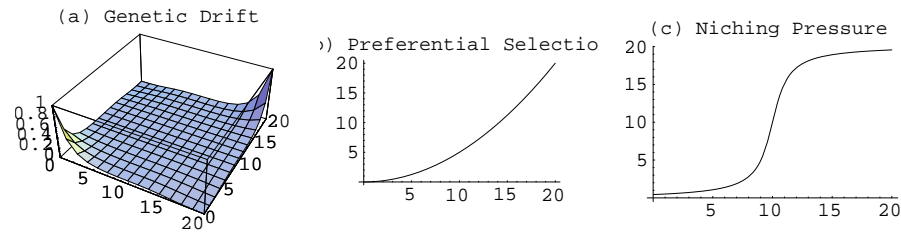


Figure 7: General effect of the forces of (a) genetic drift, (b) preferential selection, and (c) niching, on the matrix of transition probabilities for the niched GA. The exact nature of these forces needs to be investigated.

effort. After all, if these small, exact models continue to surprise us, we are probably not yet ready to make the correct, simplifying assumptions necessary for constructing approximate models of real GAs.

6 Acknowledgements

I thank Mary A. Johnson, Department of Mechanical and Industrial Engineering at the University of Illinois at Urbana-Champaign (UIUC), for pointing me toward the work on quasi-stationary Markov chains. I am also grateful to David E. Goldberg, Department of General Engineering at UIUC, for emphasizing extinction times and for introducing the notion of a characteristic measure of stability. This work was supported by NASA under contract number NGT-50873.

References

- Cavichio, D. J. (1970). *Adaptive search using simulated evolution*. Unpublished doctoral dissertation, University of Michigan, Ann Arbor.
- Darroch, J. N., & Seneta, E. (1965). On quasi-stationary distributions in absorbing discrete-time finite Markov chains. *Journal of Applied Probability*, pp. 88-100.
- Davis, T. E., & Principe, J. C. (1991). A simulated annealing-like convergence theory for the simple genetic algorithm. *Proceedings of the Fourth International Conference on Genetic Algorithms*, pp. 174-181.
- Deb, K. (1989). Genetic algorithms in multimodal function optimization. *Masters thesis, TCGA Report No. 89002*. University of Alabama.
- Deb, K., Horn, J., & Goldberg, D. E. (1992). Multimodal deceptive functions. *IlliGAL Report No. 92006 (April 1992)*, and submitted to *Complex Systems* (August 1992).

- De Jong, K. A. (1975). An analysis of the behavior of a class of genetic adaptive systems. (Doctoral dissertation, University of Michigan). *Dissertation Abstracts International*, 36(10), 5140B. (University Microfilms No. 76-9381).
- Goldberg, D. E. (1989). *Genetic algorithms in search, optimization, and machine learning*. Reading, MA: Addison-Wesley.
- Goldberg, D. E., Deb, K., & Horn, J. (1992). Massive multimodality, deception, and genetic algorithms. *Parallel Problem Solving From Nature*, 2. North-Holland, 1992, pp. 37-46.
- Goldberg, D. E., & Richardson, J. J. (1987). Genetic algorithms with sharing for multimodal function optimization. *Genetic algorithms and their applications: Proceedings of the Second International Conference on Genetic Algorithms*, pp. 41-49.
- Goldberg, D. E., & Segrest, P. (1987). Finite Markov chain analysis of genetic algorithms. *Genetic algorithms and their applications: Proceedings of the Second International Conference on Genetic Algorithms*, pp. 1-8.
- Hillier, F. S., & Lieberman, G. J. (1990). *Introduction to Operations Research, Fifth Edition*. McGraw-Hill Publishing Company, 1990.
- Mahfoud, S. W. (1991). Finite Markov chain models of an alternative selection strategy for the genetic algorithm. *IlliGAL Report No. 91007*.
- Mahfoud, S. W. (1992). Crowding and preselection revisited. *Parallel Problem Solving From Nature*, 2. North-Holland, 1992, pp. 27-36.
- Nix, A. E., & Vose, M. D. (1992). Modeling genetic algorithms with Markov chains. *Annals of Mathematics and Artificial Intelligence*, 5(1).
- Smith, R. E., Forrest, S., & Perelson, A. S. (1992). Searching for diverse, cooperative populations with genetic algorithms. *TCGA Report No. 92002*. University of Alabama, Tuscaloosa.