Generation Gaps Revisited

Kenneth A. De Jong
George Mason University
Fairfax, VA 22030 USA
kdejong@aic.gmu.edu

Jayshree Sarma
George Mason University
Fairfax, VA 22030 USA
jsarma@cs.gmu.edu

Abstract

There has been a lot of recent interest in so-called "steady state" genetic algorithms (GAs) which, among other things, replace only a few individuals (typically 1 or 2) each generation from a fixed size population of size N. Understanding the advantages and/or disadvantages of replacing only a fraction of the population each generation (rather than the entire population) was a goal of some of the earliest GA research. In spite of considerable progress in our understanding of GAs since then, the pros/cons of overlapping generations remains a somewhat cloudy issue. However, recent theoretical and empirical results provide the background for a much clearer understanding of this issue. In this paper we review, combine, and extend these results in a way that significantly sharpens our insight.

1 INTRODUCTION

In Holland's book *Adaptation in Natural and Artificial Systems* [Holland75], he introduces and analyzes two forms of reproductive plans which have served as the basis for the field of Genetic Algorithms. The first plan, R_1, maintained a fixed size population and at each time step a single individual was selected probabilistically using payoff-proportional selection to produce a single offspring. To make room for this new offspring, one individual from the current population was selected for deletion via a uniform random distribution.

Since the *expected* lifetime of an individual in R_1 under uniform deletion in a population of size N is simply N, Holland was able to show that, if the average fitness of the population, f_t, didn't change much over the lifetime of an individual, then the expected number of offspring produced by individual i is $f(i) / f_t$, which is the basis for Holland's now famous schema theorem.

Holland also introduced a second plan, R_d, in which at each time step all individuals were deterministically selected to produce their expected number of offspring $f(i) / f_t$ in a temporary storage location and, when that process was completed, the offspring produced replaced the entire current population. Thus in R_d individuals were guaranteed to produce their expected number of offspring (within probabilistic roundoff).

At the time, the two plans were viewed as generally equivalent from a theoretical point of view. However, because of practical considerations relating to severe genetic drift in small populations and the overhead of recalculating selection probabilities, most early researchers favored the R_d approach. However, the emergence of classifier systems (e.g., [Holland78]) in which only a subset of the population is replaced each time step, and recent interest in "steady state" GAs (e.g., [Whitley88]) in which only 1 or 2 individuals are replaced each time step has again raised the issue of the advantages/disadvantages of overlapping generations. In spite of considerable progress in our understanding of GAs, the pros/cons of overlapping generations remains a somewhat cloudy issue, partly due the fact that other changes such as non-uniform deletion and ranking selection are frequently also incorporated making comparisons difficult. However, recent theoretical and empirical results provide the background for a much clearer understanding of this issue. In this paper we review, combine, and extend these results in a way that significantly sharpens our insight.

2 BACKGROUND

The earliest attempt at evaluating the properties of overlapping generations was a set of empirical studies in [DeJong75] in which a parameter G, called the generation gap, was defined to control the fraction of the population to be replaced in each generation. Thus, $G = 1.0$ corresponded to R_d and $G = 1/N$ represented R_1. These early studies suggested that any advantages that overlapping populations might have was dominated by the negative effects of genetic drift (allele loss) due to the high variance in expected lifetimes and expected number of offspring given the generally modest population sizes in use at the time (N <= 100). These effects were shown to increase in severity as G was reduced. These studies also suggested the advantages of an *implicit* generation overlap in the sense that, even with $G = 1.0$, the optimal crossover rate of 0.6 and optimal mutation rate of 0.001 identified for the test suite used meant that approximately 40% of the offspring were clones of their parents.

Early experience with classifier systems yielded quite the opposite behavior in that replacing a smaller number of classifiers was generally more beneficial than replacing a large number or possibly all of them. In this case, however, the performance drop off observed as G increased was attributed to the fact that the population as a whole represented a single solution and could not tolerate large changes in content.

In recent years with the increased capacity of easily available computing equipment and the desire to solve increasing larger problems, a variety of researchers have reported positive results with so called "steady state" systems involving larger population sizes in which at most 1 or 2 individuals are replaced each generation. Whitley's GENITOR system [Whitley88] is a good example of this approach. Unfortunately, other simultaneous changes such as the use of ranking selection and deletion of the worst individual makes it difficult to understand precisely where the improved performance is coming from.

This, in turn, has led to renewed attempts to analyze the behavior of these alternatives in an attempt to resolve some of the conflicting empirical results, and ultimately build better GA-based systems. Several recent results [Goldberg90, Syswerda90] have shed light on these issues and set the stage for the work presented here.

3 GENERATION GAP ANALYSIS

A recurring theme in Holland's work is the importance of a proper balance between exploration and exploitation when adaptively searching an unknown space for high performance solutions [Holland75]. Many aspects of a GA affect this balance including selection, reproduction and deletion mechanisms, genetic operators, and population size. One way of understanding the effects of overlapping generations is to focus on the changes it makes in the exploitation/exploration balance. This is can be done by imagining a system in which there are no genetic operators. Without the ability to generate any novel individuals, selection, reproduction and deletion drive the system to a state of equilibrium in which one of members of the initial population takes over the entire population. In the following sections we analyze the effects that varying G from a value of $1/N$ (replacing a single individual) to $G = 1.0$ (replacing the entire population) will have on the actual equilibrium points reached and the rates at which equilibrium is reached.

3.1 VARIANCE EFFECTS

If we let $g = G * N$ (the number of offspring produced each generation) and let $m_i(t)$ represent the number of individuals of type i in the population at time t, then we can easily express the growth equations of type i individuals as:

$$\begin{aligned} m_i(t+1) &= m_i(t) + m_i(t) * p_s(i,t) * g - m_i(t) * p_d(i,t) * g \qquad (1) \\ &= m_i(t)[\ 1 + g * (p_s(i,t) - p_d(i,t)\ )\ ] \end{aligned}$$

where $p_s(i,t)$ is the probability of selecting a type i individual at time t for cloning, and $p_d(i,t)$ the probability of deleting a type i individual at time t.

The second line of equation 1 highlights quite clearly one immediate effect of changing g (or G), namely the amount of variance expected among individual growth curves. The "steady state" value of $g = 1$ represents the case with the highest variance. As we increase g (or G), the variance decreases since the average of g samples from the probability distribution $(p_s(i,t) - p_d(i,t))$ is more likely to approximate the mean of that distribution than a single sample. In the extreme case of $g = N$, one can deterministically replace each individual by its expected offspring (as in R_d), driving the variance to its minimal value (which is always > 0 for finite populations).

This analysis immediately provides an explanation for two known phenomena. The first is that small values of g (or G) result in actual growth curves of individuals on a single GA run (without genetic operators) which vary much more widely from the expected curves than for larger g (or G) values, including cases where the best individuals disappear and the population converges to a suboptimal equilibrium point. Figures 1 and 2 illustrate this effect by plotting the averages and variances of the growth curves obtained from 100 independent simulation runs. Using a population of size 50 in which the best individual occupies 10% of the initial population and using only payoff proportional selection, reproduction and uniform deletion (no crossover or mutation), notice that for g=1 the best individuals take over the population only about 80% of the time and the growth curves exhibit much higher variance than the case of g=N.

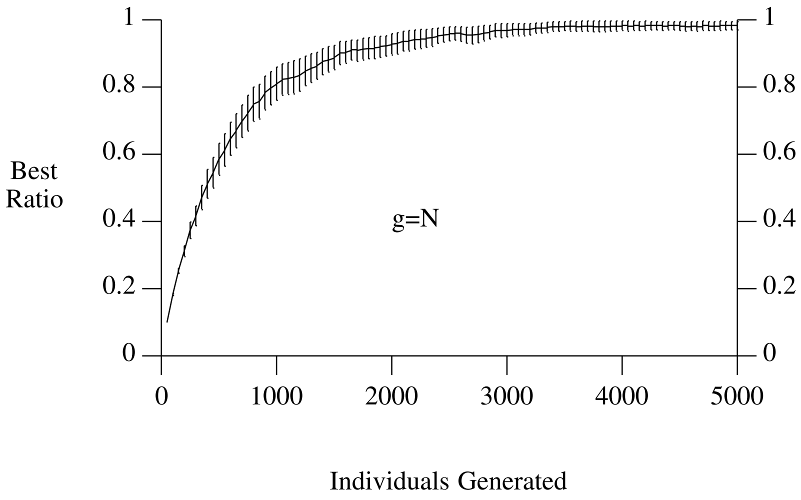

Figure 1: Mean and variance of the growth of the best with POP=50 and g=50.

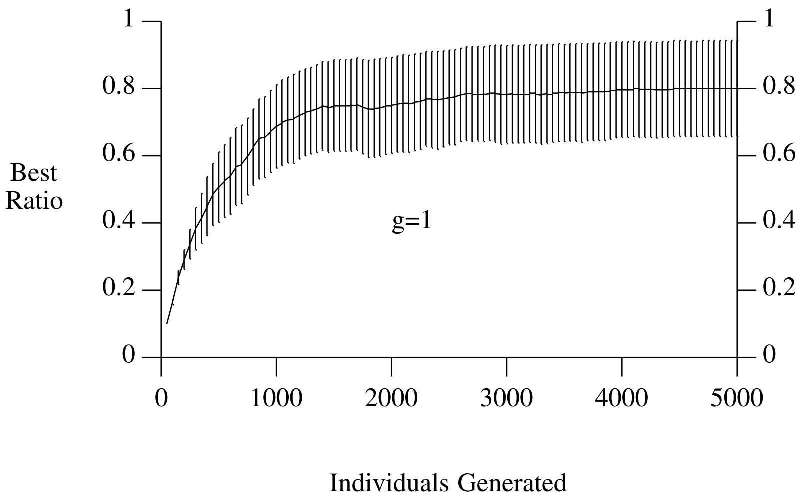

Figure 2: Mean and variance of the growth of the best with POP=50 and g=1.

A more detailed analysis of these effects can be found in [Syswerda90].

This high variance for small g values gives rise to a second phenomena: more genetic drift. With smaller population sizes the higher variance makes it is easier for alleles to disappear. Figure 3 illustrates the typical differential in allele loss rate one can see on almost any test function and smaller population sizes.

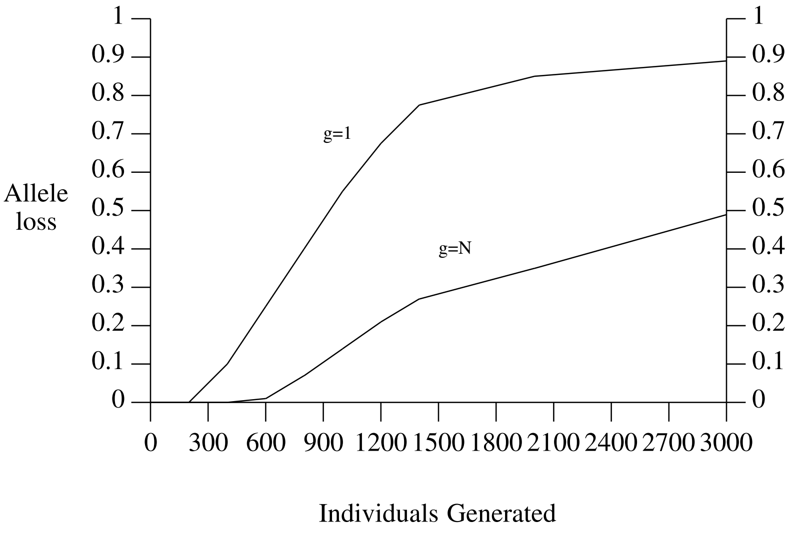

Figure 3: Allele loss for g=1 and g=N for N=50 on F1

Looked at in this way one can gain considerable insight as to why "steady state" systems (g=1) are generally run with larger population (offsetting allele loss), and prefer deletion biased towards poorer individuals (thus increasing the life expectancy of the better individuals).

3.2 LOWER VARIANCE DELETION STRATEGIES

The previous section raises the interesting question as to whether one could identify alternate implementations of steady state systems (g=1) which have lower variance, and thus reduce the need for larger population sizes or the need to introduce a deletion bias towards less fit individuals. The analysis above suggests two alternative deletion strategies which have the right theoretical properties and deserve some experimental evaluation: FIFO deletion, and excess offspring deletion.

With FIFO deletion the population is simply a first-in-first-out queue with new individuals added to one end and deleted individuals removed from the other end. The variance reduction effect is obvious in that the variance in individual lifetimes is reduced to zero. Figure 4 illustrates this by showing the effects of rerunning the Figure 2 experiments with FIFO deletion rather than uniform random deletion.

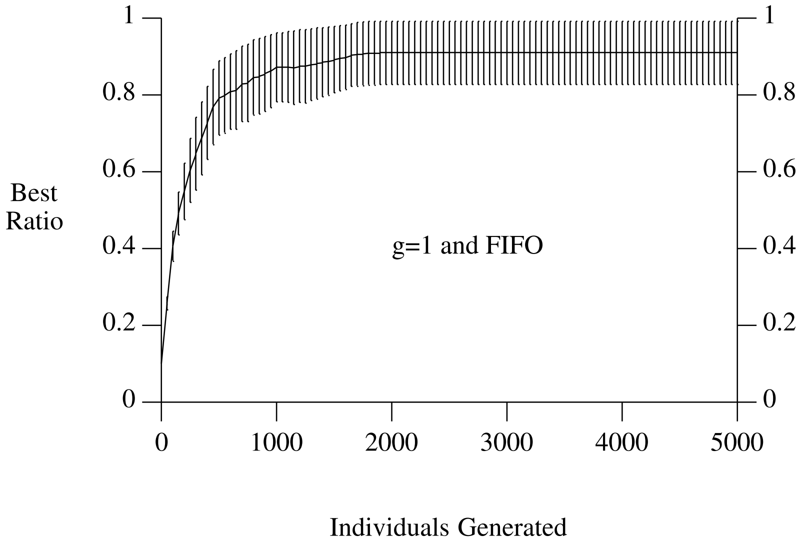

Figure 4: Mean and variance of the growth of the best with g=1 and FIFO deletion.

The variance remaining is due to the differences between the actual and expected number of offspring generated. This suggests an alternate deletion strategy in which one monitors the "excess offspring", defined as the actual number of offspring produced minus the expected number of offspring. If we bias deletion towards those with excess offspring, the effect will be to reduce the variance in the actual number of offspring produced. Care has to be taken in the implementation of this strategy in much the same spirit as Baker's analysis of selection schemes in order not to introduce unintended additional biases.

3.3 COMPOUNDING EFFECTS

One of the strong intuitive motivations for small g (or G) values is that getting good individuals back in the selection pool earlier leads to the possibility of a compounding effect due to earlier exploitation. The analysis goes something like this. If we assume payoff-proportional selection and uniform deletion, then $p_s(i,t) = f(i) \,/\, \sum f(j)$ and $p_d(i,t) = 1/N$. The change in the number of individuals of type i in one generation as specified in equation (1) above becomes:

$$m_i(t+1) = m_i(t)[1 + g * (\frac{f(i)}{\sum f(j)} - \frac{1}{N})]$$

$$= m_i(t)[1 + \frac{g}{N} * (\frac{f(i)}{\bar{f}_t} - 1)]$$

After N such iterations (generations) we have:

$$m_i(t+N) = m_i(t) * \prod_{j=0}^{N-1}[1 + \frac{g}{N} * (\frac{f(i)}{\bar{f}_{t+j}} - 1)]$$

If we assume for the time being that $f(i) / \bar{f}_t$ is relatively constant, we can simplify this to:

$$m_i(t+N) = m_i(t) * [1 + \frac{g}{N} * (\frac{f(i)}{\bar{f}} - 1)]^N$$

which one can argue grows faster than the case for g=N which is simply:

$$m_i(t+N) = m_i(t) * \frac{f(i)}{\bar{f}_t}$$

Unfortunately, the assumption that $f(i) / \bar{f}_t$ is relatively constant is a bad one and leads incorrectly to the "early compounding" intuition that many of us have held in the past. In fact, the growth curves for both cases (g=1 and g=N) turn out to be theoretically identical. The effect of increasing the number of copies of an above average individual leads to a corresponding decrease in its fitness relative to the population average, resulting in an ever decreasing compounding factor rather than a constant one. This is nicely illustrated in [Syswerda90] for a population of size 100 by computing and plotting the recurrence relations for g=1 and g=100.

We can obtain more formal insight into this phenomena by noting that the analysis in [Goldberg90] extends to the case when $g < N$. Goldberg and Deb suggest that we imagine individuals of type i reproducing at an unconstrained rate given by:

$$m_i(t+1) = m_i(t) * f(i)$$

Then the proportion of the unbounded total population taken up by individuals of type i is:

$$\begin{aligned} p_i(t+1) &= \frac{m_i(t+1)}{\sum m_j(t+1)} \\ &= \frac{m_i(t) * f(i)}{\sum m_j(t) * f(j)} \end{aligned} \tag{2}$$

Note that we could control the number of individuals produced each time step by uniformly changing the growth rates to

$$m_i(t+1) = m_i(t) * \frac{f(i)}{k_t}$$

of each subpopulation. However, such a change has no effect on proportion equation (2) above since the k_t appearing in the numerator and denominator cancel out. Hence, the theoretical growth curves (expected values) are identical for any setting of g (or G), and in particular for the cases of g=1 and g=N.

4 WHAT ABOUT REAL GAs?

The growth curve analysis of the previous section describes the behavior of a simplified environment in which crossover and mutation are inactive and quantities like $f(i)$ are exact and constant. In real GAs, of course, crossover and mutation play a critical role in the exploration/exploitation balance, and $f(i)$ represent changing estimates of hyperplane averages. Currently, our mathematical tools are inadequate to formally extend the results of the previous section to this more realistic environment.

It is not difficult, however, to generate empirical evidence to support the hypothesis that the results do indeed extend to real GAs. If this hypothesis is true, the effects of decreasing the generation gap G on actual GA systems should produce no significant increase in selection pressure (due to compounding effects). One should see only an increase in variance on individual runs and the side effects of the resulting increased allele loss.

Figures 5 and 6 show typical performance curves for the two extreme values (g=1 and g=N) for population sizes of 50 and 200. All curves are averages of 50 runs. Note how the behavioral differences diminish when the effects due to variance are reduced by increasing population size and by averaging over a sufficient number of trials.

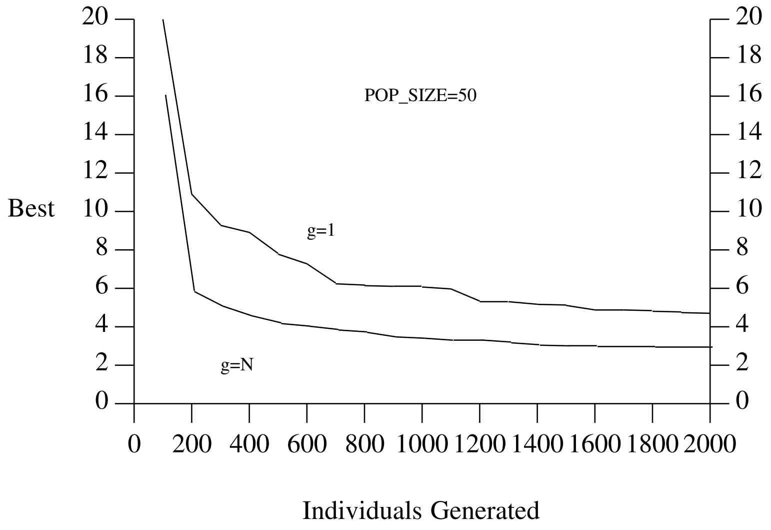

Figure 5: Performance Curves for g=1 and g=N for POP=50 on F5

5 CONCLUSIONS AND FURTHER WORK

These observations leave us with two important insights into the role of generation gaps. First, as we have seen, increasing G reduces the variance seen on individual GA runs which can be an important issue for an end user of GA-based systems. It suggests avoiding small G values in conjunction with small populations if the usual form of uniform random deletion is used. The analysis also suggests two alternate deletion strategies which could reduce variance and permit the use of steady-state systems with smaller populations.

Second, the analysis emphasizes that the important behavioral differences between current steady-state and generational systems have little if anything to do with their choice of generation gap size. In that sense, the terms "generational" and "steady-state" don't appear to be all that significant as semantic categories. Rather, the important behavioral changes are due to the changes in the exploration/exploitation balance

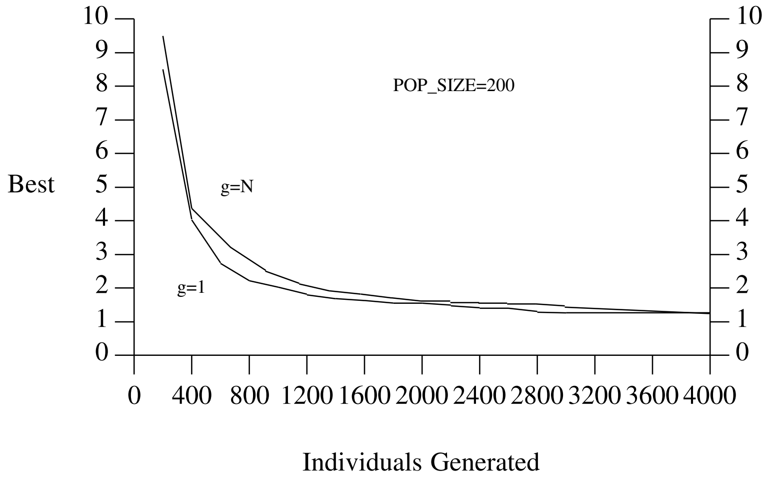

Figure 6: Performance Curves for g=1 and g=N for POP=200 on F5

resulting from the different selection and deletion strategies used. This is where we should continue to press our analysis efforts.

Note that, while a steady-state implementation of rank-based selection and worst-biased deletion is in some sense natural, one could imagine a system with $g > 1$ in which groups of individuals are generated before they are inserted back into the ranked population and the worst deleted. This raises the interesting question as to whether the effects of increasing g in such systems would also result in variance reduction without changing the selection pressure.

References

De Jong, Kenneth A. (1975). *An Analysis of the Behavior of a Class of Genetic Adaptive Systems,* Doctoral Thesis, Department of Computer and Communication Sciences, University of Michigan, Ann Arbor.

Goldberg, D. & Deb, K. (1990). A Comparative Analysis of Selection Schemes Used in Genetic Algorithms, *Proceedings of the Foundations of Genetic Algorithms Workshop*, Indiana, July 1990.

Holland, John H. (1975). *Adaptation in Natural and Artificial Systems,* The University of Michigan Press.

Holland, John H. & Reitman, Judith S. (1978). Cognitive systems based on adaptive algorithms. In D. A. Waterman & F. Hayes-Roth (Eds.), *Pattern-directed inference systems*. New York: Academic Press.

Syswerda, Gilbert. (1990). A Study of Reproduction in Generational and Steady-State Genetic Algorithms, *Proceedings of the Foundations of Genetic Algorithms Workshop*, Indiana, July 1990.

Whitley, D. & Kauth, J., (1988). Genitor: A Different Genetic Algorithm, *Proceedings 4th Rocky Mountain Conference on Artificial Intelligence*, Denver, 1988.