Selection Intensity in Genetic Algorithms with Generation Gaps

Erick Cantú-Paz
Center for Applied Scientific Computing
Lawrence Livermore National Laboratory
Livermore, CA 94550
cantupaz@llnl.gov

Abstract

This paper presents calculations of the selection intensity of common selection and replacement methods used in genetic algorithms (GAs) with generation gaps. The selection intensity measures the increase of the average fitness of the population after selection, and it can be used to predict the number of steps until the population converges to a unique solution. The theory may help to explain the fast convergence of some algorithms with small generation gaps. The accuracy of the calculations was verified experimentally with a simple test function. The results facilitate comparisons between different algorithms, and provide a tool to adjust the selection pressure, which is indispensable to obtain robust algorithms.

1 INTRODUCTION

To maintain a constant number of individuals in their populations, genetic algorithms have a mechanism that deletes unwanted individuals to make room for the newly-created ones. Most frequently, the entire population is replaced every generation. In this case, the algorithm is called a "generational GA", and it represents an extreme case of replacement methods. In the other extreme, there are "steady-state" GAs that replace a single individual in every iteration. The fraction of the population that is replaced is controlled by a parameter called generation gap (denoted by $G \in [\frac{1}{n}, 1]$, where n is the size of the population).

Although the literature has numerous observations of the effect of the generation gap on the selection pressure, it has not been quantified accurately. The purpose of this paper is to present calculations of the selection intensity in GAs with arbitrary generation gaps. The selection intensity is the normalized increase of the average fitness of the population after selection. It can be used to predict the average fitness of the population at each iteration and the number of steps until the population converges to a unique solution. In addition, the selection intensity is related to the optimal mutation rate and population size (Mühlenbein et al., 1994; Bäck, 1996). The calculations presented here consider the selection algorithm used to choose the parents and the mechanism used to replace existing members of the population with the offspring.

The paper is organized as follows. The next section briefly reviews previous work on the analysis of overlapping populations. Section 3 defines the selection intensity and summarizes previous work on characterizing it in serial and parallel generational GAs. Section 4 has the calculations for the selection intensity of GAs with generation gaps. Experiments that verify the accuracy of the calculations are presented in section 5. Finally section 6 presents a summary and the conclusions of this study.

2 GENERATION GAPS

De Jong (1975) was the first to evaluate empirically the performance of GAs with overlapping populations. He introduced the generation gap, G, as a parameter to the GA, and found that at low values of G the algorithm had a severe loss of alleles, which resulted in poor search performance. In De Jong's algorithm, the newly created individuals replaced random members of the population. He hypothesized that the poor performance was caused by the high variance in the individuals' lifetime and the number of offspring produced. Later, De Jong and Sarma (1993) presented additional empirical evidence, and suggested alternative deletion methods to reduce the variance.

Whitley (1989) introduced GENITOR, a "steady state" GA in which the worst individual was deterministically replaced every iteration. Goldberg and Deb (1991) analyzed GENITOR, and they observed that it has a high selective pressure even when the parents were selected randomly. This suggests that the deletion of worst individuals induced a higher selection pressure than the rank-based method used to select the parents. This will be quantified in section 4.

The following deletion methods are common:

- Insert offspring at random (uniformly).
- Replace the worst individuals.
- Choose using any selection algorithm normally used to select the parents (e.g., fitness-proportional, exponential or linear ranking, tournaments, etc.).
- Delete the oldest (FIFO).
- Combinations or elitist variants of the above.

There has been considerable research on the effect of these deletion methods on the speed of convergence of GAs. For example, Syswerda (1991) compared generational and steady state GAs with fitness-proportional selection of parents and several replacement methods. Assuming an infinite population (so that effects due to small populations do not appear) and using random deletion of individuals, Syswerda showed that the generation gap had no effect on the allocation of copies to strings. However, changing the deletion strategy to least-fit, exponential ranking, or fitness-proportionate deletion caused the steady state algorithm to proceed much faster than the generational GA. Calculations presented later in this paper will confirm and quantify these observations.

Chakraborty et al. (1996) used Markov chains to obtain the probability that a specific class of individuals takes over the population at each iteration. They considered random, worst-fit, and exponential ranking deletion, and their framework can be extended to other replacement strategies. Smith and Vavak (1999) did just that, and observed that replacing the oldest member or replacing randomly may result in loss of the optimal value. In the case of random replacement, Rudolph (1999) determined that the probability of losing the optimal individual is approximately 50%. De Jong and Sarma (1993) also observed losses of the optimal value, even when the initial population had 10% of the optimal individuals. Smith et al. noted that the loss can be corrected simply by using an elitist replacement strategy that ensures that the best individual in the current generation survives to the next. The simple correction suggests that variability in the number of offspring or the individuals' lifetime may not be the major cause of failure.

Interestingly, De Jong and Sarma (1993) end their paper noting that "...the important behavioral changes [between generational and steady state GAs] are due to the changes in the exploration/exploitation balance resulting from the different selection and deletion strategies used. This is where we should continue our analysis efforts." That is precisely the purpose of this paper: to quantify accurately the selection intensity (the exploitation part). De Jong and Sarma also question whether an algorithm that selects a block of the best individuals and replaces a block of the worst would reduce the variance without changing the selection pressure. Section 4 shows that the answer is negative, and that indeed the selection pressure changes significantly as a function of G (the size of the blocks).

3 SELECTION INTENSITY

This section briefly reviews previous work on quantifying the intensity of selection methods. In addition, this section reviews recent work that characterizes the selection intensity caused by migration of individuals between populations in parallel GAs. The next section builds on the models presented here.

3.1 SELECTION METHODS

Some common selection methods are proportionate selection (Holland, 1975), linear ranking (Baker, 1985), tournament selection (Brindle, 1981), $(\mu \overset{+}{,} \lambda)$ selection (Schwefel, 1981), and truncation selection (Mühlenbein & Schlierkamp-Voosen, 1993). In linear ranking selection, individuals are selected with a probability that is linearly proportional to the rank of the individuals in the population. The desired expected number of copies of the best (n^+) and worst ($n^- = 2 - n^+$) individuals are supplied as parameters to the algorithm. In tournament selection, s individuals are randomly sampled from the population (with or without replacement), and the best individual in the sample is selected. The process is repeated until the mating pool is filled. In $(\mu + \lambda)$ selection, λ offspring are created from μ parents, and the μ best individuals out of the union of parents and offspring are selected. In (μ, λ) selection ($\lambda \geq \mu$) the μ best offspring are selected to survive. Truncation selection selects the top $1/\tau$ of the population and creates τ copies of each individual. It is equivalent to (μ, λ) selection with $\mu = \lambda/\tau$.

Selection	Parameters	I
Tournament	s	$\mu_{s:s}$
(μ, λ)	μ, λ	$\frac{1}{\mu}\sum_{i=\lambda-\mu+1}^{\lambda}\mu_{i:\lambda}$
Linear Ranking	n^+	$(n^+ - 1)\frac{1}{\sqrt{\pi}}$
Proportional	σ^t, μ_t	σ^t/μ_t

Table 1: Selection intensity for common selection schemes.

Mühlenbein and Schlierkamp-Voosen (1993) introduced the use of the selection intensity to study the convergence of selection schemes. The selection intensity is defined as

$$I = \frac{\bar{f}^{t+1} - \bar{f}^t}{\sigma^t}, \tag{1}$$

where $\bar{f}^t = \frac{1}{n}\sum_{i=1}^{n} f_i^t$ is the mean fitness of the population at iteration t, σ^t is the standard deviation of the population, and the superscript t denotes the iteration. The numerator is called the selection differential, and is usually denoted as s^t.

The challenge to calculate the intensity of a selection method is to compute the mean fitness of the selected individuals, $\bar{f}_s^{t+1}$. This has been accomplished analytically for some common selection schemes of generational GAs. In particular, Bäck (1995) and Miller and Goldberg (1995) independently derived the selection intensity for tournament selection, and Bäck (1995) also derived I for (μ, λ) selection. Blickle and Thiele (1996) calculated the intensity of linear ranking, and Mühlenbein and Schlierkamp-Voosen (1993) calculated I for proportional selection. Table 1 contains the known selection intensities (adapted from (Miller & Goldberg, 1996)). $\mu_{i:n}$ denotes the expected value of the i-th order statistic of n samples with a unit Gaussian distribution (see equation 11). Note that I is independent of the distribution of the current population, except for proportional selection.

3.2 MULTI-POPULATION GAs

Regardless of their implementation on uni- or multi-processor computers, GAs with multiple populations exhibit a different behavior than GAs with a single population. Much has been written about this, but one of the main causes of the disparity seems to be the additional selection intensity caused by choosing migrants and replacements according to their fitness (Cantú-Paz, In press).

The selection intensity caused by migration is

$$I_{mig} = I_e + I_r, \tag{2}$$

where I_e is the selection intensity caused by selecting the emigrants, and I_r is the intensity caused by selecting replacements in the receiving deme. Using δ to denote the number of neighbors of a deme (the degree of the connectivity graph) and ρ to denote the migration rate (i.e., the fraction of the population that migrates every generation), $I_e \approx \delta\phi(\Phi^{-1}(1-\rho))$ if the best individuals are selected to migrate, and $I_e = 0$ if the migrants are chosen randomly. $\phi(z) = \exp(-z^2/2)/\sqrt{2\pi}$ and $\Phi(z) = \int_{-\infty}^{z}\phi(x)dx$ are the PDF and CDF respectively of a standard Gaussian distribution with mean 0 and standard deviation of 1.

Similarly, $I_r \approx \phi(\Phi^{-1}(1-\delta\rho))$ if the worst individuals in the receiving deme are replaced by the migrants, and $I_r = 0$ if replacements are chosen randomly. We shall see in the next section that the equations for the selection intensity in GAs with overlapping populations are very similar to those above.

4 GENERATION GAPS AND SELECTION INTENSITY

To calculate the average fitness of the population in the next iteration, we take the weighted average of the individuals selected to reproduce and the individuals that were not replaced (which we call survivors):

$$\bar{f}^{t+1} = G\bar{f}_s^{t+1} + (1-G)\bar{f}_{surv}^{t+1}, \tag{3}$$

where $\bar{f}_s^{t+1}$ is the expected fitness of the selected individuals, $\bar{f}_{surv}^{t+1}$ is the expected fitness of the survivors. To simplify things, we may also write the average fitness of the population as a weighted sum:

$$\bar{f}^t = G\bar{f}^t + (1-G)\bar{f}^t. \tag{4}$$

Collecting similar terms, we can write the selection differential as

$$\begin{aligned} s^t &= \bar{f}^{t+1} - \bar{f}^t \\ &= G(\bar{f}_s^{t+1} - \bar{f}^t) + (1-G)(\bar{f}_{surv}^{t+1} - \bar{f}^t) \\ &= Gs_s^t + (1-G)s_{surv}^t, \end{aligned} \tag{5}$$

and dividing by the standard deviation we obtain the selection intensity:

$$I = GI_s + (1-G)I_{surv}. \tag{6}$$

This equation clearly shows that the selection pressure has two independent causes, namely the selection of the parents and the selection of survivors. I_s and I_{surv} take different values depending on the methods used to select parents and replacements. In the remainder of this section we examine three basic options for each.

The first option is that parents (or replacements) are chosen with one of the common selection methods. In this case, I_s (or I_{surv}) would be simply the intensity of the selection method given in table 1.

The second option is that parents are chosen randomly. In this case, the expected fitness of the selected individuals, $\bar{f}_s^{t+1}$, is equal to the mean fitness of the population before selection, $\bar{f}^t$, so there is no selection pressure ($s_s^t = 0$) and $I_s = 0$. The same argument applies when the replacements are chosen randomly, and I_{surv} would be 0.

The third option is that the best individuals are selected as parents (or that the worst are deleted). This option produces the strongest selection pressure, and the calculation of the selection intensity is more complicated.

We can interpret the fitness values $f_i^t, i \in [1, n]$ as samples of random variables F_i^t with a common distribution, say $N(\bar{f}^t, \sigma^t)$, although most of the following derivation applies to other distributions. We may arrange the variables in increasing order to obtain the order statistics:

$$F_{1:n}^t \leq F_{2:n}^t \leq \dots \leq F_{n:n}^t.$$

Without loss of generality, we assume a maximization problem. The expected fitness of the Gn best individuals that are selected is

$$\bar{f}_s^{t+1} = \frac{1}{Gn} \cdot \sum_{i=n-Gn+1}^{n} \mathrm{E}(F_{i:n}^t). \tag{7}$$

The random variables can be normalized as

$$Z_{i:n} = \frac{F_{i:n}^t - \bar{f}^t}{\sigma^t},$$

and the average fitness of the selected individuals may be rewritten in terms of the normalized variables

$$\begin{aligned} \bar{f}_s^{t+1} &= \frac{1}{Gn} \sum_{i=n-Gn+1}^{n} (\mathrm{E}(Z_{i:n})\sigma^t + \bar{f}^t) \\ &= \sigma^t \cdot \frac{1}{Gn} \sum_{i=n-Gn+1}^{n} \mathrm{E}(Z_{i:n}) + \bar{f}^t. \end{aligned} \tag{8}$$

Now, we can calculate the selection differential caused by selecting the best as

$$\begin{aligned} s_s^t &= G(\bar{f}_s^{t+1} - \bar{f}^t) \\ &= \frac{1}{n} \cdot \sigma^t \cdot \sum_{i=n-Gn+1}^{n} \mathrm{E}(Z_{i:n}). \end{aligned} \tag{9}$$

Since the selection differential is $s^t = I \cdot \sigma^t$, the selection intensity in this case is

$$I_s = \frac{1}{n} \cdot \sum_{i=n-Gn+1}^{n} \mathrm{E}(Z_{i:n}). \tag{10}$$

The expected value of the i-th order statistic of a sample of size n is defined as

$$\begin{aligned} \mu_{i:n} &= \mathrm{E}(Z_{i:n}) \\ &= n\binom{n-1}{i-1} \int_{-\infty}^{\infty} z\phi(z)\Phi^{i-1}(z)[1-\Phi(z)]^{n-i} dz, \end{aligned} \tag{11}$$

where $\phi(z)$ and $\Phi(z)$ are the PDF and CDF respectively of the fitness distribution (in our case a standard Gaussian distribution with mean 0 and standard deviation of 1). The values of $\mu_{i:n}$ are computationally expensive to calculate, but for a Gaussian distribution they are tabulated for $n \leq 400$ (Harter, 1970). Nevertheless, computing the sum in equation 10 can be tedious, so we use the following approximation[1] (Burrows, 1972):

$$\sum_{i=n-Gn+1}^{n} \mu_{i:n} \approx \frac{n}{G} \phi(\Phi^{-1}(1-G)), \tag{12}$$

and therefore equation 10 can be approximated as

$$I_s \approx \frac{\phi(\Phi^{-1}(1-G))}{G}. \tag{13}$$

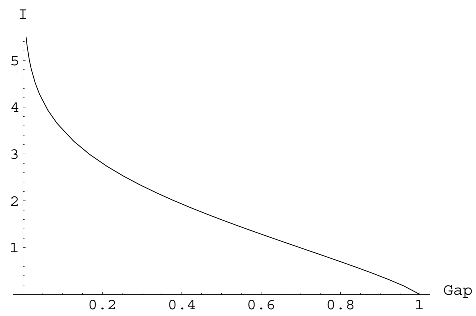
A similar derivation shows that replacing the worst individuals results in

$$I_{surv} \approx \frac{\phi(\Phi^{-1}(G)}{1-G}. \tag{14}$$

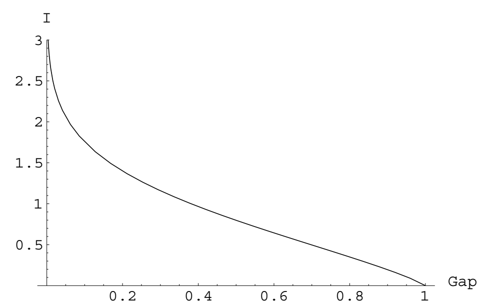
It is important to realize that the selection intensity is an adimensional quantity that does not depend on the fitness function or on the generation t. The only assumption made was that the fitness values had a normal distribution, but any other distribution may be used as long as $\mathrm{E}(F_{i:n})$ may be computed (by substituting the appropriate PDF and CDF in equation 11).

Note that the highest value of I_{surv} (or I_s) is at $G = 1/n$, and it can be of considerable magnitude. For example, for $n = 256$, $I_{surv} = 2.96$, and for $n = 1000$, $I_{surv} = 3.36$. So, even if the parents are selected randomly, replacing the worst individuals causes a considerable selection pressure. This is consistent with Goldberg and Deb's (1991) calculations of GENITOR, and Chakraborty et al.'s (1996) Markov chains analysis.

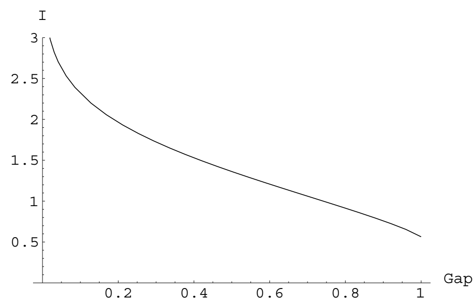
[1] In his study of (μ, λ) selection, Bäck (1995) shows that for $n > 50$ the approximation is indistinguishable from the real values.



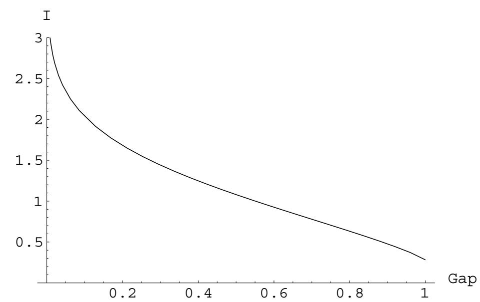
(a) Select best, delete worst



(b) Select best, delete randomly (or select randomly and delete worst, or (μ, λ))



(c) Tournaments ($s = 2$), delete worst



(d) Ranking ($n^+ = 1.5$), delete worst

Figure 1: Selection intensity of different selection and replacement strategies varying the generation gap.

Figure 1 has plots of the selection intensity of algorithms with different methods to select the parents and the replacements. To make the graphs, $\Phi^{-1}(x)$ was calculated numerically using Mathematica 3.0 as $\sqrt{2}$ `InverseErf[0,2x-1]`.

The plots show that the combination of selecting the best individuals as parents and deleting the worst individuals has the highest selection intensity. In the two cases when the best individuals are selected, there is no selection pressure at $G = 1$, because the algorithm simply copies the entire population. In addition, the selection intensity when the best are selected and replacement is random is identical to (μ, λ) selection with $G = \mu/\lambda$. In the case of tournaments and linear ranking with random deletion, the graphs would be horizontal lines at 0.5642 and 0.2820, respectively. This is consistent with Syswerda's (1991) observations that with random deletion the generation gap does not affect the selection pressure.

We must be cautious when comparing algorithms with the same selection intensity, because they are not equivalent in all aspects. The selection intensity only considers the change of the population's mean fitness over time, and ignores the higher moments of the distribution. Blickle and Thiele (1996) made an analysis of the variance of several (generational) selection methods, and Rogers and Prügel-Bennett (1999) have a detailed analysis of the first four moments of a roulette-wheel algorithm that uses Boltzmann weights. Different selection algorithms impact the higher moments in their own particular ways and may affect the quality of the solutions found. A reasonable heuristic is that given a choice between algorithms with the same selection intensity, we should prefer the one that produces the highest variance of fitness (Bäck, 1995).

Another aspect that we must take into consideration when comparing GAs is that the convergence time of a GA is inversely proportional to the selection intensity. For example, Rogers and Prügel-Bennett (1999) observed that they could replicate the dynamics of a generational GA with a steady state GA using half of the function evaluations. Their results can be explained quite easily when we take into consideration that in their steady state GA both the parents and the replacements were selected according to their fitness, effectively doubling the selection intensity. Although it may be tempting to use higher selection intensities, in some cases the algorithm may converge too fast to reach satisfactory solutions (this is sometimes called 'premature convergence'). Using small populations exacerbates this problem, and may have contributed to the poor performance of the GAs in De Jong's empirical studies with small generation gaps.

5 EXPERIMENTS

This section presents experimental results that verify the accuracy of the calculations of the previous section. The experiments use a $l = 500$ bit OneMax function, $F = \sum_{i=1}^{l} x_i$, where $x_i \in \{0, 1\}$ are the individual bits in the chromosome. For this problem, Mühlenbein and Schlierkamp-Voosen (1993) showed that with an initial random population the number of generations until convergence is given by Gen $\approx \frac{\pi}{2}\frac{\sqrt{l}}{I}$. This result assumed an infinite binomially distributed population and that the algorithm converges to the global optimum. Although this equation is an approximation, other studies have used it successfully to predict the number of generations until convergence (e.g., Blickle and Thiele (1996), Miller and Goldberg (1996)). We use it here to test the accuracy of our calculations of the selection intensity (equation 6). The number of iterations until convergence is converted to generations (that process n individuals) by multiplying by G.

In the experiments, the population size is $n = 500$ individuals, which is sufficient to ensure convergence to the optimum in all cases. The GA uses uniform crossover with probability 1.0 and no mutation. Crossover was applied five times to obtain a population that approximates a binomial distribution (Mühlenbein & Schlierkamp-Voosen, 1993). The results are the average of 30 independent runs for each parameter setting.

Figure 2 compares the theoretical predictions with experimental results. The graphs show the number of generations until convergence using best-fit selection, pairwise tournaments, and linear ranking with $n^+ = 1.5$. Both random and worst-fit deletion were used. Additional experiments with $n^+ = 2$ yielded the same results as pairwise tournaments, as was expected because the two algorithms have the exact same effect on the fitness distribution of the population (same selection intensity and selection variance) (Blickle & Thiele, 1996).

6 CONCLUSIONS

This paper presented calculations of the selection intensity of genetic algorithms with arbitrary generation gaps. We found that the selection intensity can be of considerable magnitude with small generation gaps, and that it decreases monotonically as the gap becomes larger. The accuracy of the theory was verified experimentally, and it was used as a possible explanation for previous observations reported by others. The resulting equations are similar to those that model the selection intensity of migration in multi-population GAs. This suggests the possibility of exchanging ideas and analysis techniques to further advance our understanding of the two types of algorithms.

Future work should consider the effect of selection on the higher moments of the distribution of fitness. This is important because algorithms with the same selection intensity may reduce the variance (diversity) of the population in different ways and may also change the shape of the distribution. Studying these effects may help to design recombination or mutation operators that balance the effects of selection.

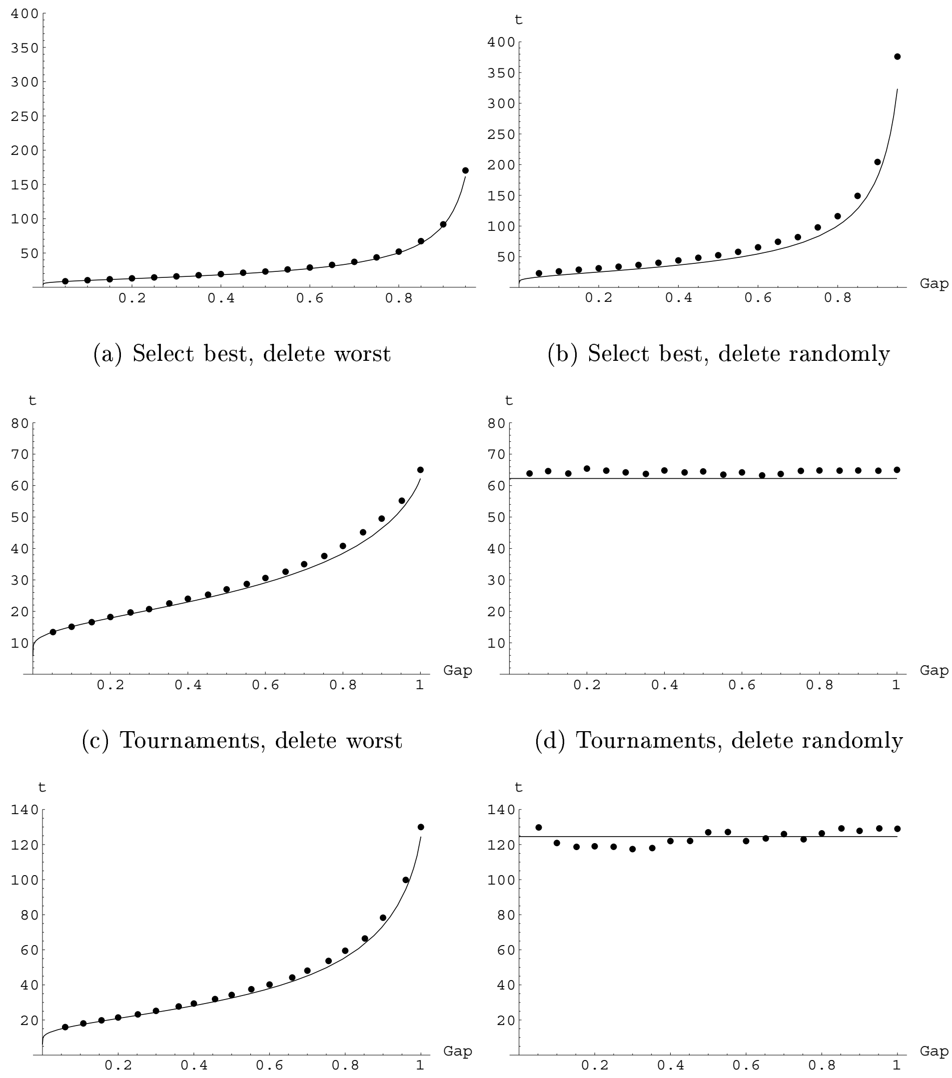
It is well known that algorithms with higher selection pressure need larger populations to succeed (Mühlenbein & Schlierkamp-Voosen, 1993; Harik et al., 1999). This introduces a tradeoff because higher selection pressures result in faster convergence, but larger populations require more computations. The tradeoff suggests that there is an optimal population size and selection pressure that minimize the total computational work. Future work along these lines may produce a framework that relates selection intensity, population size, and solution quality.

Such a framework would be very useful for the design of faster and more reliable evolutionary algorithms. Besides facilitating comparisons between different algorithms, and providing a convenient tool to adjust the selection pressure, the results of this paper would be a critical component of the framework.

Acknowledgments

Portions of this work were completed while the author was a student at the Illinois Genetic Algorithms Laboratory of the University of Illinois at Urbana-Champaign. While at Illinois, the work was sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant number F49620-97-1-0050. Research funding for this project was also provided by a grant from the US Army Research Laboratory Program, Cooperative Agreement DAAL01-96-2-003.

UCRL-JC-137169. This work was performed under the auspices of the U.S. Department of Energy by University of California Lawrence Livermore National Laboratory under contract no. W-7405-Eng-48.



(a) Select best, delete worst (b) Select best, delete randomly

(c) Tournaments, delete worst (d) Tournaments, delete randomly

(e) Ranking, delete worst (f) Ranking, delete randomly

Figure 2: Generations until convergence using different selection and replacement strategies and varying the generation gap. The lines are the theoretical predictions using equation 6, and the dots are experimental results. Note the different scales of the graphs.

References

Bäck, T. (1995). Generalized convergence models for tournament- and (μ, λ)-selection. In Eschelman, L. (Ed.), *Proceedings of the Sixth International Conference on Genetic Algorithms* (pp. 2–8). San Francisco, CA: Morgan Kaufmann.

Bäck, T. (1996). *Evolutionary Algorithms in Theory and Practice.* New York, NY: Oxford University Press.

Baker, J. E. (1985). Adaptive selection methods for genetic algorithms. In Grefenstette, J. J. (Ed.), *Proceedings of an International Conference on Genetic Algorithms and Their Applications* (pp. 101–111). Hillsdale, NJ: Lawrence Erlbaum Associates.

Blickle, T., & Thiele, L. (1996). A comparison of selection schemes used in evolutionary algorithms. *Evolutionary Computation*, *4*(4), 361–394.

Brindle, A. (1981). *Genetic algorithms for function optimization.* Unpublished doctoral dissertation, University of Alberta, Edmonton, Canada.

Burrows, P. (1972). Expected selection differentials for directional selection. *Biometrics*, *28*, 1091–1100.

Cantú-Paz, E. (In press). Migration policies, selection pressure, and parallel evolutionary algorithms. *Journal of Heuristics*.

Chakraborty, U. K., Deb, K., & Chakraborty, M. (1996). Analysis of selection algorithms: A markov chain approach. *Evolutionary Computation*, *4*(2), 133–167.

De Jong, K. A. (1975). *An analysis of the behavior of a class of genetic adaptive systems.* Doctoral dissertation, University of Michigan, Ann Arbor. (University Microfilms No. 76-9381).

De Jong, K. A., & Sarma, J. (1993). Generation gaps revisited. In Whitley, L. D. (Ed.), *Foundations of Genetic Algorithms 2* (pp. 19–28). San Mateo, CA: Morgan Kaufmann.

Goldberg, D. E., & Deb, K. (1991). A comparative analysis of selection schemes used in genetic algorithms. In Rawlins, G. J. E. (Ed.), *Foundations of Genetic Algorithms* (pp. 69–93.) San Mateo, CA: Morgan Kaufmann.

Harik, G., Cantú-Paz, E., Goldberg, D. E., & Miller, B. L. (1999). The gambler's ruin problem, genetic algorithms, and the sizing of populations. *Evolutionary Computation, 7*(3), 231–254.

Harter, H. L. (1970). *Order statistics and their use in testing and estimation.* Washington, D.C.: U.S. Government Printing Office.

Holland, J. H. (1975). *Adaptation in natural and artificial systems.* Ann Arbor, MI: University of Michigan Press.

Miller, B. L., & Goldberg, D. E. (1995). Genetic algorithms, tournament selection, and the effects of noise. *Complex Systems*, *9*(3), 193–212.

Miller, B. L., & Goldberg, D. E. (1996). Genetic algorithms, selection schemes, and the varying effects of noise. *Evolutionary Computation*, *4*(2), 113–131.

Mühlenbein, H., & Schlierkamp-Voosen, D. (1993). Predictive models for the breeder genetic algorithm: I. Continuous parameter optimization. *Evolutionary Computation*, *1*(1), 25–49.

Mühlenbein, H., & Schlierkamp-Voosen, D. (1994). The science of breeding and its application to the breeder genetic algorithm (BGA) *Evolutionary Computation*, *1*(4), 335–360.

Rogers, A., & Prügel-Bennett, A. (1999). Modelling the dynamics of a steady state genetic algorithm. In *Foundations of Genetic Algorithms 5* (pp. 57–68). San Francisco, CA: Morgan Kaufmann.

Rudolph, G. (1999). Takeover times and probabilities of non-generational selection rules. Technical Report No. CI-78/99. University of Dortmund.

Schwefel, H. (1981). *Numerical optimization of computer models.* Chichester: John Wiley and Sons.

Smith, J., & Vavak, F. (1999). Replacement strategies in steady state genetic algorithms: Static environments. In *Foundations of Genetic Algorithms 5* (pp. 219–234). San Francisco, CA: Morgan Kaufmann.

Syswerda, G. (1991). A study of reproduction in generational and steady-state genetic algorithms. In Rawlins, G. J. E. (Ed.), *Foundations of Genetic Algorithms* (pp. 94–101). San Mateo, CA: Morgan Kaufmann.

Whitley, D. (1989). The GENITOR algorithm and selective pressure: Why rank-based allocation of reproductive trials is best. In Schaffer, J. D. (Ed.), *Proceedings of the Third International Conference on Genetic Algorithms* (pp. 116–121). San Mateo, CA: Morgan Kaufmann.