Optimal Sampling and Speed-Up for Genetic Algorithms on the Sampled OneMax Problem

Tian-Li Yu[1], David E. Goldberg[2], and Kumara Sastry[3]

Illinois Genetic Algorithms Laboratory (IlliGAL)
Department of General Engineering
University of Illinois at Urbana-Champaign
104 S. Mathews Ave, Urbana, IL 61801
{tianliyu,deg,kumara}@illigal.ge.uiuc.edu

Abstract. This paper investigates the optimal sampling and the speed-up obtained through sampling for the sampled OneMax problem. Theoretical and experimental analyses are given for three different population-sizing models: the decision-making model, the gambler's ruin model, and the fixed population-sizing model. The results suggest that, when the desired solution quality is fixed to a high value, the decision-making model prefers a large sampling size, the fixed population-sizing model prefers a small sampling size, and the gambler's ruin model has no preference for large or small sizes. Among the three population-sizing models, sampling yields speed-up only when the fixed population-sizing model is valid. The results indicate that when the population is sized appropriately, sampling does not yield speed-up for problems with subsolutions of uniform salience.

1 Introduction

Over the last few decades, significant progress has been made in the theory, design and application of genetic and evolutionary algorithms. A decomposition design theory has been proposed and *competent* genetic algorithms (GAs) have been developed (Goldberg, 1999). By competent GAs, we mean GAs that can solve hard problems quickly, reliably, and accurately. Competent GAs render problems that were intractable by first generation GAs tractable requiring only a polynomial (usually subquadratic) number of function evaluations.

However, in real-world problems, even the time required for a subquadratic number of function evaluations can be very high, especially if the function evaluation is a complex model, simulation, or computation. Therefore, GA practitioners have used a variety of *efficiency enhancement* techniques to alleviate the computational burden. One such technique is *evaluation relaxation* (Sastry and Goldberg, 2002), in which the accurate, but costly fitness evaluation is replaced by a cheap, but less accurate evaluation. Partial evaluation through sampling is an example of evaluation relaxation, and sampling has empirically shown to yield significant speed-up (Grefenstette & Fitzpatrick, 1985). Evaluation relaxation through sampling has also been analyzed by developing facetwise and dimensional models (Miller & Goldberg 1996a; Miller, 1997; Giguère & Goldberg,

E. Cantú-Paz et al. (Eds.): GECCO 2003, LNCS 2724, pp. 1554–1565, 2003.

1998). This study extends the theoretical analyses and investigates the utility of sampling as an evaluation-relaxation technique.

The objective of this paper is to extend the work of Giguère and Goldberg (1998) and incorporate the effect of sampling on both convergence time and population sizing of GAs. Specifically, we concentrate on problems with substructures of uniform salience. This paper is composed of four primary parts: (1) background knowledge including previous work and an introduction to the sampled OneMax problem, (2) the derivation of the optimal sampling sizes for the decision-making model, the gambler's ruin model, and the fixed population-sizing model, (3) empirical results, and (4) extensions and conclusions of this work.

2 Past Work

Grefenstette and Fitzpatrick (1985) achieved a great success in applying sampling techniques to the image registration problem. Their success motivated Miller and Goldberg (1996b), which gave a theoretical analysis of a related problem. They derived and empirically verified the theoretical optimal sampling size when the fitness function is clouded by an additive Gaussian noise. The sampling methods in the above two papers have a subtle but significant difference. In Grefenstette and Fitzpatrick (1985), the variance becomes zero if full sampling is used. However, in the additive Gaussian noisy OneMax problem in Miller and Goldberg (1996b), the variance of fitness function varies at the sampling size. The variance can be very small when the sampling size is large, but it will never be zero. More recently, Giguère and Goldberg (1998) investigated a sampled OneMax problem (SOM) where the fitness value is calculated by sampling. Their results showed that sampling is not really useful when the gambler's ruin population-sizing model (Harik, Cant'u-Paz, Goldberg & Miller, 1997; Miller, 1997) is adopted. Even though Giguère and Goldberg (1998) have investigated all population-sizing models discussed in this paper, they did not consider convergence time. In addition, detailed analytical models are needed for a better understanding of the sampling schemes as a technique of evaluation relaxation.

3 Sampled OneMax Problem (SOM)

Before starting the derivation of our models of computational requirement, let us first define the sampled OneMax problem (SOM). The SOM is basically the OneMax or the counting ones problem, except the fitness value is computed by sampling without replacement.

$$F_n(\bar{x}) = \frac{l}{n}(\Sigma_{i \in S}\, x_i), \tag{1}$$

where $\bar{x}$ is a chromosome (a binary string in SOM), x_i is the value of the i-th gene in the chromosome (0 or 1), n is the sampling size ($0 < n \leq l$), S is a subset of $\{1, 2, \cdots, l\}$ with a restriction that $|S| = n$, and l is the chromosome length. The term $\frac{l}{n}$ is just for normalization so that the expectation of the fully sampled fitness $F(\bar{x}) = \Sigma_{i=1}^{l} x_i$ is the same as the expectation of F_n for all n. The variance of the noise introduced by sampling can be expressed as follows (Giguère & Goldberg, 1998):

$$\sigma_n^2 = \frac{l^2 p(t)(1 - p(t))}{n} \frac{l - n}{l - 1}, \tag{2}$$

where $p(t)$ is the performance model defined in Thierens and Goldberg (1994). Initially, $p(0) = 0.5$. The number of function evaluations is defined as how many bits the GA has sampled before convergence. We choose the SOM because (1) it is linear and easy to analyze, and (2) the SOM is considered as a GA-easy problem, the speed-up obtained for this problem should give an idea about how sampling acts on other problems.

4 Optimal Sampling for the Sampled OneMax Problem (SOM)

This section gives theoretical and empirical analyses of the optimal sampling sizes for three different population-sizing models: the decision-making model, the gambler's ruin model, and the fixed population-sizing model for the SOM.

This section starts by deriving the number of function evaluations required for the SOM as a function of the sampling size. Then with the number of function evaluation model derived, the optimal sampling size is derived for each of the three population sizing models. Finally, experimental results are shown to verify the optimal sampling sizes derived.

In this section, we fix the solution quality (the average fitness of the population) to $(l - 1)$, where l is the chromosome length, and try to minimize the number of function evaluations through sampling. The desired solution quality is set so high that the convergence time model is valid. Nearly full convergence is one of the presumptions of the derivations of the convergence time models in Sastry and Goldberg (2002).

4.1 Model for Number of Function Evaluations for the SOM

Now let us derive the model for number of function evaluations for the SOM. Sastry and Goldberg (2002) indicated the time to convergence for those problems with uniformly-scaled building blocks (BBs) corresponds to the squared root of the variance of the fitness function:

$$t_{conv} = \frac{\ln(l)}{I} \sqrt{\sigma_f^2 + \sigma_n^2}, \tag{3}$$

where I is selection intensity (Mühlenbein & Schlierkamp-Voosen, 1993). For binary tournament selection, $I = 1/\sqrt{\pi} \sim 0.5642$ (Bäck, 1994). By Eq. (2) and approximating σ_f^2 by the initial variance $lp(0)(1-p(0))$, t_{conv} can be rewritten as

$$t_{conv} = C_1\sqrt{\frac{l^2}{n} - 1}, \tag{4}$$

where $C_1 = \frac{\ln(l)}{I} \cdot \sqrt{\frac{l}{l-1}p(0)(1-p(0))}$.

The number of function evaluations is given by $n_{fe} = nGN$, where n is the sampling size, G is the number of generations, and N is the population size. By substituting G with t_{conv} obtained in Eq. (4), the number of function evaluations can be expressed as

$$n_{fe} = C_1 \cdot nN\sqrt{\frac{l^2}{n} - 1}. \tag{5}$$

Now that we have the model for the number of function evaluations for the SOM, we are ready to derive the optimal sampling size for different population-sizing models.

4.2 Optimal Sampling for the Decision-Making Model

The decision-making model (Goldberg, Deb, & Clark, 1992) is given by

$$N = \Gamma(\sigma_f^2 + \sigma_n^2), \tag{6}$$

where N is the population size, Γ is the population coefficient defined in Miller & Goldberg (1996b), σ_f^2 is the initial population fitness variance, and σ_n^2 is the variance of the fitness noise.

By Eq. (2) and substituting N into Eq. (5), n_{fe} for the decision-making model can be written as

$$n_{fe} = C_2 \cdot n\left(\frac{l^2}{n} - 1\right)^{\frac{3}{2}}, \tag{7}$$

where $C_2 = \frac{\ln(l)}{I}\Gamma\left(\frac{l}{l-1}p(0)(1-p(0))\right)^{\frac{3}{2}}$.

Equation (7) has a minimum at $n = l$, which means no sampling at all (Fig. 1). This is not surprising because the population sizing model is known as an overestimation than needed for the desired solution quality. When sampling is adopted, the noise introduced makes the population size even larger than needed according to the decision-making model. As a result, the larger population size results in a larger number of function evaluations.

4.3 Optimal Sampling for the Gambler's Ruin Model

The gambler's ruin model (Harik, Cant'u-Paz, Goldberg & Miller, 1997; Miller, 1997) is given by

$$N = \Gamma'(\sqrt{\sigma_f^2 + \sigma_n^2}), \tag{8}$$

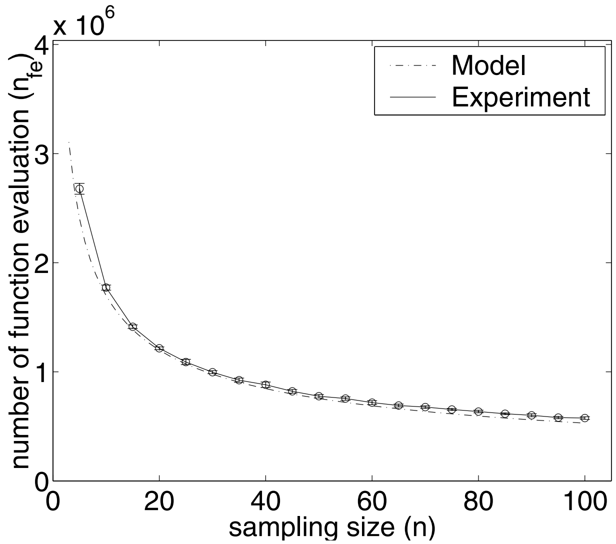

Fig. 1. The relationship between n_{fe} and n for the decision-making model

where N, σ_f^2, and σ_n^2 are defined the same as those in Eq. (6), and Γ' is another coefficient defined as

$$\Gamma' = -\frac{2^{k_{max}-1}\sqrt{\pi}\ln(\psi)}{d_{min}}. \tag{9}$$

k_{max} is an estimate of the maximal length of BBs. ψ is the failure rate, defined as the probability that a particular partition in the chromosome that fails to converge to the correct BBs. In other words, $(1-\psi)$ is the expected proportion of the correct BBs in an individual. d_{min} is an estimate of the minimal signal difference between the best and the second best BB. In other words, d_{min} is the smallest BB signal that GAs can detect. In OneMax domain, both k_{max} and d_{min} are 1, which yields a simpler form for Γ':

$$\Gamma' = -\sqrt{\pi}\ln(\psi) \tag{10}$$

By a similar algebraic process as in the previous subsection, the number of function evaluations is expressed as

$$n_{fe} = C_3 \cdot (l^2 - n), \tag{11}$$

where $C_3 = \frac{\ln(l)}{I}\Gamma'\left(\frac{l}{l-1}p(0)(1-p(0))\right)$.

The minimum of Eq. (11) occurs at $n = l$, which means, again, no sampling at all (Fig. 2). The gambler's ruin model still prefers a larger sampling size, but

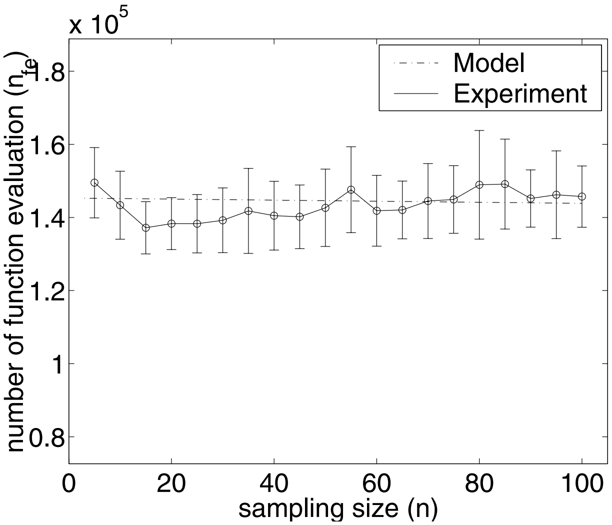

Fig. 2. The relationship between n_{fe} and n for the gambler's ruin model

only with a slight preference. This can be shown by comparing the case of $n = 1$ and $n = l$.

$$\frac{n_{fe}(n=1)}{n_{fe}(n=l)} = \frac{l^2 - 1}{l^2 - l} = \frac{l+1}{l} \tag{12}$$

For a 100-bit SOM problem, the difference of n_{fe} is only 1%, which is so small that can be neglected compared with the approximations in the derivations. As a result, for the gambler's ruin model, sampling does not make much difference. The conclusion agrees with Giguère and Goldberg (1998).

4.4 Fixed Population-Sizing Model

The fixed population-sizing model is not an accurate model of GAs since it does not account for the effects of problem size and noise. However, it is widely adopted in real-world applications, such as in Grefenstette and Fitzpatrick (1985). Therefore, it is worthwhile to investigate this model as well. In this section, the assumption is that the fixed population size is larger than needed so that the GA can converge to an $(l-1)$ solution quality. This assumption is needed for applying the convergence time model. Since the population size is larger than needed, sampling should obtain speed-up. In addition, we still fix the desired solution quality to be $(l-1)$ because we stop the GA when a $(l-1)$ solution quality is reached.

The number of function evaluations for a fixed population size is given by

$$n_{fe} = C_4 \cdot n \left(\frac{l^2}{n} - 1 \right)^{\frac{1}{2}}, \tag{13}$$

where $C_4 = \frac{\ln(l)}{I} N \left(\frac{l}{l-1} p(0)(1-p(0)) \right)^{\frac{1}{2}}$.

Equation (13) has a minimum at $n = 1$. If the overhead (α) is taken into account, and the cost of sampling one bit is β, the total run duration is expressed as

$$T = C_4 \cdot (\alpha + n\beta) \left(\frac{l^2}{n} - 1 \right)^{\frac{1}{2}} \tag{14}$$

For large l, $\frac{l^2}{n} \gg 1$, Eq. (14) can be approximated as

$$T = C_4 \cdot l(\alpha n^{-1/2} + \beta n^{1/2}) \tag{15}$$

By differentiating Eq. (15) by n, and then setting it to be zero, the minimum is found at

$$n_{op} = \frac{\alpha}{\beta} \tag{16}$$

It is interesting to compare Eq. (16) with what Miller and Goldberg (1996b) got ($\sqrt{\frac{\alpha}{\beta}}\sqrt{\frac{\sigma_n^2}{\sigma_f^2}}$). The term σ_n in Miller and Goldberg's result vanishes because now σ_n is controlled by the sampling size n. If the constant term is ignored, the result here is the square of Miller and Goldberg's optimal sampling size.

4.5 Experimental Results

Experimental results are obtained for the SOM problem with chromosome length $l = 100$. Binary tournament selection and uniform crossover are adopted. For all experiments, the crossover probability (p_c) is one and the mutation probability (p_m) is zero, which means, crossover always takes place and there is no mutation. All experimental results are averaged over 50 independent runs. Sastry and Goldberg (2002) used the variance of fitness of the initial population to estimate σ_f^2. This is, of course, an overestimation. In fact, the variance of fitness becomes smaller and smaller during the runs of a GA. For the OneMax problem, it almost becomes zero when the average fitness converges to $l-1$ bits ($p(t) \simeq 1$). Therefore, for a tighter estimation, $p(t) = 0.75$ (half convergence) is used in the calculation of σ_f^2 and σ_n^2.

Figures 1, 2, and 3 show the relationship of n_{fe} versus n for the decision-making model, the gambler's ruin model, and the fixed population-sizing model, respectively. In Fig. 1, the population size is obtained from $N = 8(\sigma_f^2 + \sigma_n^2)$ (Goldberg, Deb, & Clark, 1992). Figure 2 uses $N = 9.4\sqrt{\sigma_f^2 + \sigma_n^2}$ (Miller &

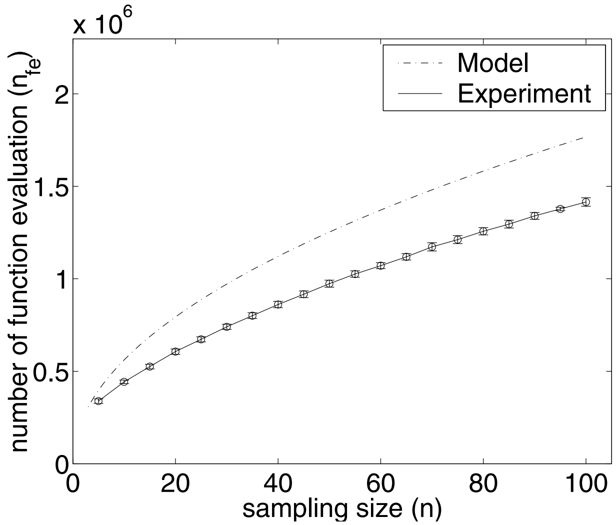

Fig. 3. The relationship between n_{fe} and n for fixed population $N = 500$. $\frac{\alpha}{\beta} = 0$

Goldberg, 1996b). In Fig. 2, the experimental results show slight minima in-between $N = 10$ and $N = 30$. It agrees with the observation in Giguère and Goldberg, 1998, which has not been explained by mathematical models so far. The fixed population size is set to be $N = 500$ and the results are shown in Fig. 3. The fixed population size is set so large to prevent failure of convergence. Finally, the total run duration for $\alpha/\beta = 20$ is shown in Fig. 4.

The experimental results agree remarkably well with the models derived in the previous section. The model of number of function evaluations for the decision-making model especially matches experimental results. The model for fixed population size overestimates somewhat the number of function evaluations required. Nevertheless, as Fig. 4 indicates, our model accurately predicts the optimal sampling size.

5 Apparent Speed-Up

From Sect. 4, the following key observations can be made:

1. Sampling does not yield speedup when the population is sized appropriately by either using the decision-making model or the gambler's ruin model. Furthermore, when the population is sized according to the decision-making model, the optimal solution is to sample all the bits, that is, $n = l$. On the other hand, when the population is sized according to the gambler's ruin model, there is no preference for a specific sampling size. That is, the same number of function evaluations is required when any valid sampling size is used.

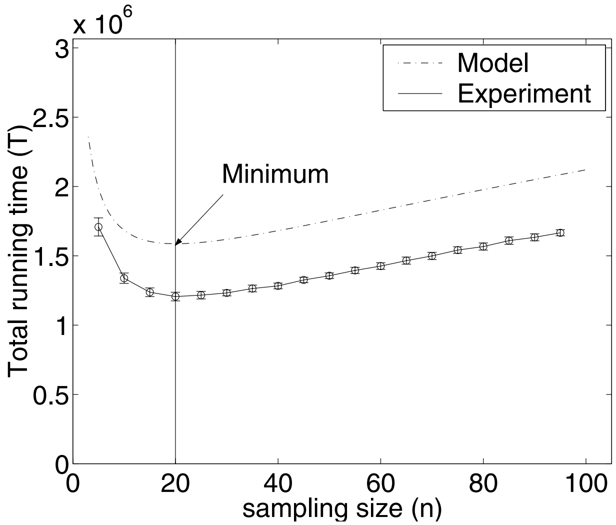

Fig. 4. The relationship between T and n for fixed population $N = 500$. $\frac{\alpha}{\beta} = 20$

2. When the population size is fixed arbitrarily, and usually to a large number (fixed population-sizing model), then speed-up can be obtained through sampling. The optimal sampling size in this case is given by Eq. (16).

Therefore, this section focuses on the fixed population-sizing model and investigates the speed-up obtained through sampling. Since the speed-up is gained only when the population is not sized appropriately, it is called *apparent speed-up.*

Equation (15) can be rewritten as a function of n as following:

$$T(n) = C_4 \cdot l(\alpha n^{-1/2} + \beta n^{1/2}). \tag{17}$$

The speed-up gained is

$$SP = \frac{T(n = l)}{T(n = n_{op})}, \tag{18}$$

where n_{op} is given by Eq. (16). With some algebraic simplifications, the speed-up can be expressed as a function of $\frac{\alpha}{\beta}$:

$$SP(\frac{\alpha}{\beta}) = \frac{1}{2}\left[l^{-\frac{1}{2}}\left(\frac{\alpha}{\beta}\right)^{\frac{1}{2}} + l^{\frac{1}{2}}\left(\frac{\alpha}{\beta}\right)^{-\frac{1}{2}}\right]. \tag{19}$$

Note that Eq. (19) is only valid when $1 \leq \frac{\alpha}{\beta} \leq l$, because $1 \leq n_{op} \leq l$. For $\frac{\alpha}{\beta} < 1$, the speed-up is $SP(1)$, and for $\frac{\alpha}{\beta} > l$, the speed-up is $SP(l) = 1$, which means no real speed-up is gained.

Figure 5 shows the relationship between the speed-up gained and $\frac{\alpha}{\beta}$ for a 100-bit SOM. The experiments were done using the same parameter settings

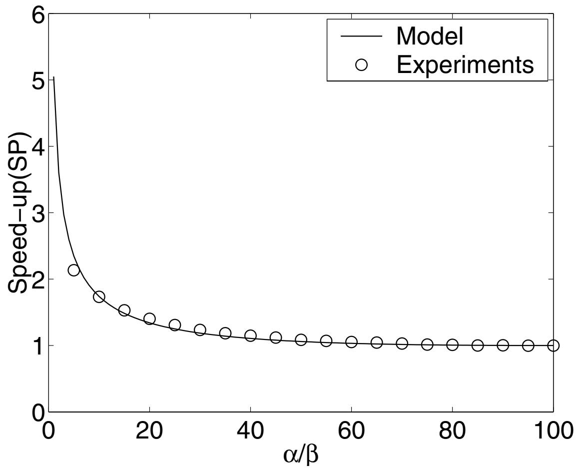

Fig. 5. The speed-up gained through sampling for a 100-bit SOM using a fixed population size 500

in the previous section. Again, all experimental results are averaged over 50 independent runs. The experimental results agree with the model derived. When the overhead (α) is relatively small, a higher speed-up is gained. When the overhead is relatively large, as one can expect, the speed-up becomes smaller. As an extreme case, when the overhead is so large that $\frac{\alpha}{\beta} > l$, sampling will not speed up the GA at all.

6 Future Work

This paper has analyzed sampling as an evaluation-relaxation technique on problems with substructures of uniform-salience. There are a number of avenues for future work which are both interesting as well as practically significant. Some of them are listed below.

- Further investigation needs to be done to bound the effectiveness of sampling schemes by considering an array of adversarial problems that contain one or more facets of problem difficulty such as deception, noise, and scaling.
- In this paper, the assumption that the GA converges to a high solution quality is needed for applying the convergence time model. Additional modeling and analyses are required to understand the time-critical condition where the number of function evaluations is limited.
- The long term goal of this work is to better understand the limit of sampling schemes applied to GAs: where and when sampling gives speed-up (or not).

We believe that the analysis and methodology presented this paper should carry over or can be extended in a straightforward manner on most of the above issues.

7 Conclusions

This paper has studied the optimal sampling and the speed-up obtained through sampling for the sampled OneMax problem (SOM). Based on Sastry and Goldberg, 2002, facetwise models for solution quality as a function of sampling size were derived for three population-sizing models, namely, the decision-making model (Goldberg, Deb, & Clark, 1992), the gambler's ruin model (Harik, Cantú-Paz, Goldberg & Miller, 1997; Miller, 1997),and the fixed population-sizing model. Then the speed-up for the fixed population-sizing model is analyzed and empirically verified. The optimal sampling size and speed-up obtained by sampling are analyzed under the scenario: we fix the solution quality to a very high value and our goal is to obtain it with minimum number of function evaluations. Each of the models was verified with empirical results.

When the desired solution quality is fixed, the results suggest that the decision-making model prefers a larger sampling size, and that the fixed population sizing model prefers a smaller sampling size. Sampling does not make much difference for the gambler's ruin model.

The results show that sampling does not give speedup for problems with subsolutions of uniform salience, if the population is sized appropriately to handle both the stochastic decision making and the noise introduced by sampling. On the other hand, if the population size is fixed without accounting for either decision-making or sampling, then the results presented in this paper show that sampling does indeed yield speed-up and an optimal sampling size exists.

Acknowledgement. This work was sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant F49620-00-0163, the National Science Foundation under grant DMI-9908252. The U.S. Government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright notation thereon.

The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Office of Scientific Research, the National Science Foundation, or the U.S. Government.

References

Bäck, T. (1994). Selective pressure in evolutionary algorithms: A characterization of selection mechanisms. *Proceedings of the First IEEE Conference on Evolutionary Computation*, 57–62.

Giguère, P., & Goldberg, D.E. (1998). Population sizing for optimum sampling with genetic algorithms: A case study of the Onemax problem. In *Genetic Programming 98* (pp. 496–503).

Goldberg, D.E. (1999). The race, the hurdle, and the sweet spot: Lessons from genetic algorithms for the automation of design innovation and creativity. *Evolutionary Design by Computers*, 105–118.

Goldberg, D.E., Deb, K., & Clark, J.H. (1992). Genetic algorithms, noise, and the sizing of populations. *Complex Systems*, *6*, 333–362.

Grefenstette, J.J., & Fitzpatrick, J.M. (1985). Genetic search with approximate function evaluations. In *Proceedings of an International Conference on Genetic Algorithms and Their Applications* (pp. 112–120).

Harik, G., Cantú-Paz, E., Goldberg, D.E., & Miller, B.L. (1997). The gambler's ruin problem, genetic algorithms, and the sizing of populations. In *Proceedings of 1997 IEEE International Conference on Evolutionary Computation* (pp. 7–12). Piscataway, NJ: IEEE.

Miller, B.L. (1997). *Noise, sampling, and efficient genetic algorithms.* doctoral dissertation, University of Illinois at Urbana-Champaign, Urbana. Also IlliGAL Report No. 97001.

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

Miller, B.L., & Goldberg, D.E. (1996b). Optimal sampling for genetic algorithms. *Proceedings of the Artificial Neural Networks in Engineering (ANNIE '96) conference*, *6*, 291–297.

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

Sastry, K., & Goldberg, D.E. (2002). Genetic algorithms, efficiency enhancement, and deciding well with differing fitness variances. *Proceedings of the Genetic and Evolutionary Computation Conference*, 528–535. (Also IlliGAL Report No. 2002002).

Thierens, D., & Goldberg, D.E. (1994). Convergence models of genetic algorithm selection schemes. In *Parallel Problem Solving fron Nature, PPSN III* (pp. 119–129).