A Note on the Escape Probabilities for Two Alternative Methods of Selection under Gaussian Mutation

Hans-Georg Beyer
University of Dortmund
Department of Computer Science
D-44227 Dortmund
Germany
beyer@ls11.informatik.uni-dortmund.de

David B. Fogel
Natural Selection, Inc.
3333 N. Torrey Pines Ct., Suite 200
La Jolla, CA 92037
dfogel@natural-selection.com

Abstract

The probabilities for generating improved solutions under two forms of selection under Gaussian mutation are studied. The results indicate that, under some simplifying assumptions, there can be advantage to retaining offspring that are of lesser value than the parent that generates them. The limitations of the analysis are identified, as well as directions for future research.

Introduction

All methods of evolutionary computation employ a mechanism for selecting which solutions in a population should be retained to serve as the basis for generating further samples. A variety of selection procedures have been offered: proportional selection (Holland, 1975), tournament selection (Fogel, 1988; Goldberg 1990), the so-called "plus" and "comma" methods of evolution strategies (Bäck and Schwefel, 1993), and so forth. The choice of which selection mechanism to employ has traditionally been based on empirical trial-and-error or mathematical analysis of the time required for a single solution to dominate a population (takeover time) (Goldberg and Deb, 1991; Bäck, 1994; Fogel and Fogel, 1995). Although this latter analysis provides a framework for studying the stringency of different selection procedures, it does not in general provide information concerning the probability of discovering improved solutions because it omits the interaction between variation and selection (it focuses only on selection). In attempting to gain insight into the choice of selection operator, it may be more productive to move toward an analysis that explicitly incorporates the interaction of these two fundamental facets of evolutionary computation.

This paper focuses on a simplified problem of selection in the face Gaussian mutation. Two selection mechanisms are compared. The first is similar to a (1+1) method, where a parent generates an offspring and if the offspring is not better than the parent then the offspring is discarded and a new trial is again generated from the parent. This method may be used for some number of generations g. The second is similar to a (1,1) method, where a parent generates an offspring and the parent is discarded, with the offspring becoming the new parent. The method is iterated a number of times and attention is given to the probability of generating a point whose distance from the original parent is greater than or equal to a given value, say k, again after some number of generations g. This is analogous to the probability of escaping from a locally optimal well of a specific radius (new vectors that are not sufficiently distant from the parent receive a greater error score within the locally optimal well). A success is declared in the first method if any if a new vector is generated that is at least k units away from the original parent within the g trials. A success is declared in the second method if the gth trial is at least k units away from the original parent.

Analysis

Given an initial n-dimensional vector $\vec{x}^{(0)}$, a new vector $\vec{x}^{(1)}$ is created by:

$$\vec{x}^{(1)} = \vec{x}^{(0)} + \vec{N}(0,1) \tag{1}$$

where $\vec{N}(0,1)$ is a multidimensional standard normal random variable. Under the first method of selection, $\vec{x}^{(0)}$ is used as a basis for generating g new vectors using the rule in Eq. 1, giving rise to the sequence $\left\{\vec{x}^{(1)}, \vec{x}^{(2)}, ..., \vec{x}^{(g)}\right\}_R$. The subscript R refers to the sequence as being generated from a method of repeating mutation from the initial vector $\vec{x}^{(0)}$. Under the second method of selection, a sequence of g vectors is created where each vector is generated from the previously created vector:

$$\vec{x}^{(i+1)} = \vec{x}^{(i)} + \vec{N}(0,1) \tag{2}$$

giving rise to the sequence $\left\{\vec{x}^{(1)}, \vec{x}^{(2)}, ..., \vec{x}^{(g)}\right\}_S$, where the subscript S refers to the sequence as being generated under "soft" selection (offspring that have worse performance than their parent can be used to generate further offspring). The question at hand then is: what is the probability to escape from a certain k-region around the vector $x^{(0)}$ after g trials?

The One-Dimensional Case

First, consider the one-dimensional case. For the method of repeating mutation from the initial parent (a scalar), let:

$$= \left\{x^{(1)}, x^{(2)}, ..., x^{(g)}\right\}_R \tag{3}$$

$$= \left\{x^{(0)} + N_1(0,1), x^{(0)} + N_2(0,1), ..., x^{(0)} + N_g(0,1)\right\} . \tag{4}$$

Let

$$x^{(i)} = x^{(0)} + N_i(0,1) \tag{5}$$

be the ith element in , and define the quantity:

$$x_R \quad = \max_i \left\{\left| x^{(0)} - x^{(i)} \right|\right\} . \tag{6}$$

The probability that the repeated generation method will escape from a k-region around $\vec{x}^{(0)}$ is the complement of the probability that the value x_R will be less than k:

$$
\begin{aligned}
P_R \quad &= 1 - P\left(x_R < k\right) && (7)\\
&= 1 - P\left(\left| N(0,1) \right| < k\right)^g && (8)\\
&= 1 - \left[\int_{-k}^{k} \frac{1}{\sqrt{2\pi}} \exp(\frac{-1}{2} t^2) dt\right]^g && (9)\\
&= 1 - \left[\phi\left(k\right) - \phi\left(-k\right)\right]^g && (10)\\
&= 1 - \left[2\phi\left(k\right) - 1\right]^g . && (11)
\end{aligned}
$$

For the method of soft selection, the element:

$$
x^{(g)} = x^{(0)} + \sum_{i=1}^{g} N_i(0,1) . \qquad (12)
$$

The scalar $x^{(g)}$ is a sum of g independent $N(0,1)$ random variables, therefore it has distribution:

$$
x^{(g)} \sim N\left(x^{(0)}, g\right), \qquad (13)
$$

where g is the variance and $x^{(0)}$ is the mean. To escape the *k*-region we must consider:

$$
\begin{aligned}
P_S \quad &= P\left(\left| x^{(0)} - x^{(g)} \right| \geq k\right) && (14)\\
&= P\left(\left| N(0,g) \right| \geq k\right) && (15)\\
&= 1 - P\left(\left| N(0,g) \right| < k\right) && (16)\\
&= 1 - \left[\int_{-k}^{k} \frac{1}{\sqrt{2\pi g}} \exp(-\frac{1}{2}\frac{t^2}{g}) dt\right] && (17)\\
&= 1 - \left[\phi\left(\frac{k}{\sqrt{g}}\right) - \phi\left(\frac{-k}{\sqrt{g}}\right)\right] && (18)\\
&= 2\left[1 - \phi\left(\frac{k}{\sqrt{g}}\right)\right] . && (19)
\end{aligned}
$$

For sufficiently large k and $kg^{-0.5}$, using the relationship that for large x:

$$
\phi\left(x\right) \simeq 1 - \frac{1}{\sqrt{2\pi}} \frac{\exp(-\frac{1}{2}x^2)}{x} \qquad (20)
$$

asymptotically:

$$P_R \quad = 1 - \left[2\phi(k) - 1\right]^g \tag{21}$$

$$\simeq 1 - \left[1 - \frac{2}{\sqrt{2\pi}} \frac{\exp\left(-\frac{1}{2}k^2\right)}{k}\right]^g \tag{22}$$

By the binomial theorem:

$$= 1 - \left[1 - g\frac{2}{\sqrt{2\pi}} \frac{\exp\left(-\frac{1}{2}k^2\right)}{k} + \ldots\right] \tag{23}$$

$$\simeq g\sqrt{\frac{2}{\pi}} \frac{\exp\left(-\frac{1}{2}k^2\right)}{k}. \tag{24}$$

Likewise:

$$P_S \quad = 2\left[1 - \phi\left(\frac{k}{\sqrt{g}}\right)\right] \tag{25}$$

$$\simeq 2 - 2\left[1 - \frac{1}{\sqrt{2\pi}} \frac{\exp\left(-\frac{1}{2}\frac{k^2}{g}\right)}{k/\sqrt{g}}\right] \tag{26}$$

$$= \sqrt{\frac{2}{\pi}}\sqrt{g} \frac{\exp(-\frac{1}{2}\frac{k^2}{g})}{k}. \tag{27}$$

Comparing P_S with P_R:

$$\frac{P_S}{P_R} \simeq \frac{1}{\sqrt{g}} \exp\left[-\frac{1}{2}k^2\left(\frac{1}{g} - 1\right)\right] \tag{28}$$

$$\therefore \left(\frac{P_{soft}}{P_{rep}}\right)^2 \simeq \frac{1}{g}\exp\left[k^2\left(1 - \frac{1}{g}\right)\right]. \tag{29}$$

This implies that P_S will be greater than P_R when:

$$\frac{1}{g}\exp\left(k^2\left(1 - \frac{1}{g}\right)\right) > 1 \tag{30}$$

$$\exp\left(k^2\left(1 - \frac{1}{g}\right)\right) > g \tag{31}$$

$$k^2\left(1 - \frac{1}{g}\right) > \ln g \tag{32}$$

$$k > \sqrt{\frac{\ln g}{1 - \frac{1}{g}}}, \tag{33}$$

for $kg^{-0.5}$ sufficiently large (note: Eq.33 does not provide a sharp bound for the case $kg^{-0.5} < \infty$). Figure 1 shows the relationship between the two selection methods for some values of g and k ranging over [0,5].

The n-Dimensional Case

For the n-dimensional case, the general escape probability cannot be easily solved because it depends on the distributions of the basins of attraction, but if attention is given to a spherical escape domain, then for the case of repeated samples from the initial parent vector, since all of the components of the offspring vectors are independent and identically distributed (iid) as Gaussian random variables, the length of a single mutation has a χ_n-distribution, with probability density function $P_{\chi_n}(l)$. Therefore:

$$P_R = 1 - \left[P\left(\chi_n < k\right)\right]^g \tag{34}$$

$$= 1 - \left[\int_0^k P_{\chi_n}(l)dl\right]^g. \tag{35}$$

Alternatively then:

$$P_R = 1 - \left[P\left(\chi_n^2 < k^2\right)\right]^g \tag{36}$$

because

$$\chi < k \Rightarrow \chi^2 < k^2 \tag{37}$$

Thus:

$$P_R = 1 - \left[\int_0^{k^2} P_{\chi_n^2}(l)dl\right]^g. \tag{38}$$

In contrast, for the method of soft selection, the one-dimensional case transfers easily to the n-dimensional case because the ith component of $\vec{x}^{(g)}$ varies as $N\left(\vec{x}_i^{(0)}, g\right)$. This gives:

$$P_S = P\left(\left\| \vec{x}^{(0)} - \vec{x}^{(g)} \right\| \geq k\right) \tag{39}$$

$$= P\left(\left\| \vec{N}(0, g) \right\| \geq k\right) \tag{40}$$

$$= P\left(\left\| \vec{N}(0, 1) \right\| \geq \frac{k}{\sqrt{g}}\right) \tag{41}$$

$$= P\left(\chi_n \geq \frac{k}{\sqrt{g}}\right) \tag{42}$$

$$= 1 - P\left(\chi_n < \frac{k}{\sqrt{g}}\right) \tag{43}$$

$$= 1 - \int_0^{\frac{k}{\sqrt{g}}} P_{\chi_n}(l)dl. \tag{44}$$

Alternatively:

$$P_S \qquad = 1 - P\left(\chi_n^2 < \frac{k^2}{g}\right) \tag{45}$$

$$= 1 - \int_0^{\frac{k^2}{g}} P_{\chi_n^2}(l)dl. \tag{46}$$

We succeeded in showing that for the asymptotic case $n \to \infty$ that there is always a $k > \sqrt{ng}$ such that $P_S > P_R$ holds. But this result is of limited interest. Eqs. 38 and 46 can be compared for some selected values of n, g, and k. Figure 2 shows the relationship between the two escape probabilities for n = 10 or 100, g = 2 or 10, as k varies from zero to relatively large values. The escape probability for soft selection is observed to be at least as great as that for hard selection, and the difference between the two escape probabilities can be significant for large values of k.

Discussion

Early efforts in both evolution strategies and evolutionary programming relied on a strictly nonregressive evolution that retained the best available solutions at each generation (Fogel, 1964; Rechenberg, 1965). But it was acknowledged early in both communities that there may be a benefit to retaining offspring of lesser worth (Fogel et al., 1966, p. 21, and the comma selection of evolution strategies, see Bäck and Schwefel (1993) for a review). This possibility has been argued in many other references as well (e.g., Galar, 1985, 1989; and others). The current analysis shows this advantage mathematically, albeit for a simple case.

This analysis is of limited value because the soft selection method only uses the final vector in a sequence of g trials. Any reasonable evolutionary algorithm would evaluate the intermediate vectors as well, and therefore the escape probability for soft selection is bounded below by the analysis presented here. Certainly, the assumption of a hyperspherical k-region from which to escape is also a limiting factor. Future efforts may include more general conditions for generating improved solutions, as well as more realistic selection mechanisms.

References

T. Bäck (1994) "Selective pressure in evolutionary algorithms," *Proceedings of the 1994 IEEE International Conference on Evolutionary Computation*, IEEE Press, N , pp. 57-62.

T. Bäck and H.-P. Schwefel (1993) "An overview of evolutionary algorithms for parameter optimization," *Evolutionary Computation*, **1**:1, 1-24.

D.B. Fogel (1988) "An evolutionary approach to the traveling salesman problem," *Biological Cybernetics*, **60**:2, 139-144.

G.B. Fogel and D.B. Fogel (1995) "Continuous evolutionary programming: Analysis and experiments," *Cybernetics and Systems*, **26**, 79-90.

L.J. Fogel (1964) "On the organization of intellect," Doctoral Dissertation, UCLA.

L.J. Fogel, A.J. Owens, and M.J. Walsh (1966) *Artificial Intelligence Through Simulated Evolution*, John Wiley, N .

R. Galar (1985) "Handicapped individua in evolutionary processes," *Biological Cybernetics*, **53**, 1-9.

R. Galar (1989) "Evolutionary search with soft selection," *Biological Cybernetics*, **60**, 357-364.

D.E. Goldberg (1990) "A note on Boltzmann tournament selection for genetic algorithms and population-oriented simulated annealing," *Complex Systems*, **4**, 445-460.

D.E. Goldberg and K. Deb. (1991) "A comparative analysis of selection schemes used in genetic algorithms," *Foundations of Genetic Algorithms*, G.J.E. Rawlins (ed.), Morgan Kaufmann, San Mateo, CA, pp. 69-93.

J.H. Holland (1975) Adaptation in Natural and Artificial Systems, Univ. Mich. Press, Ann Arbor.

I. Rechenberg (1965) "Cybernetic solution path of an experimental problem," Royal Aircraft Establishment, Library Translation No. 1122, Ministry of Aviation, Farnborough Hants.

Figure 1. The escape probabilities for the two selection methods as the number of iterations (g) and the k-region vary. The curves illustrate the probabilities for the cases of g = 2, 5, 10, and 100 for both the repeated selection and soft selection methods as k varies from zero to five. To compare the escape probabilities for each method, choose a particular value of g and locate the intersection of the two relevant curves. Values of k greater than the point of intersection will be more easily escaped from using soft selection.

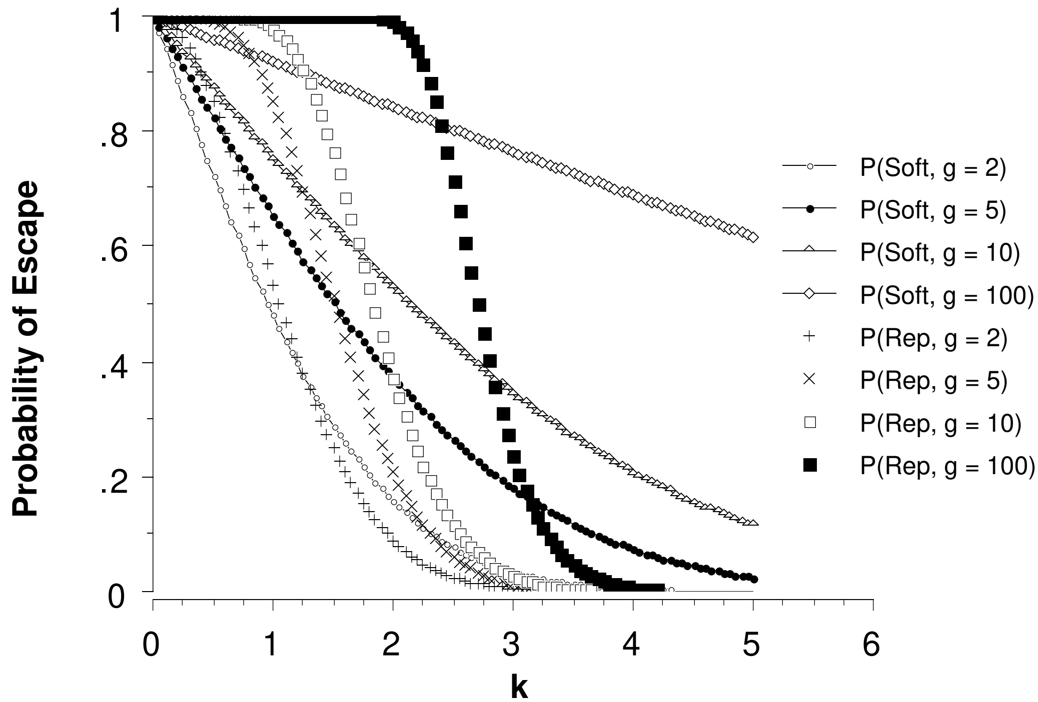
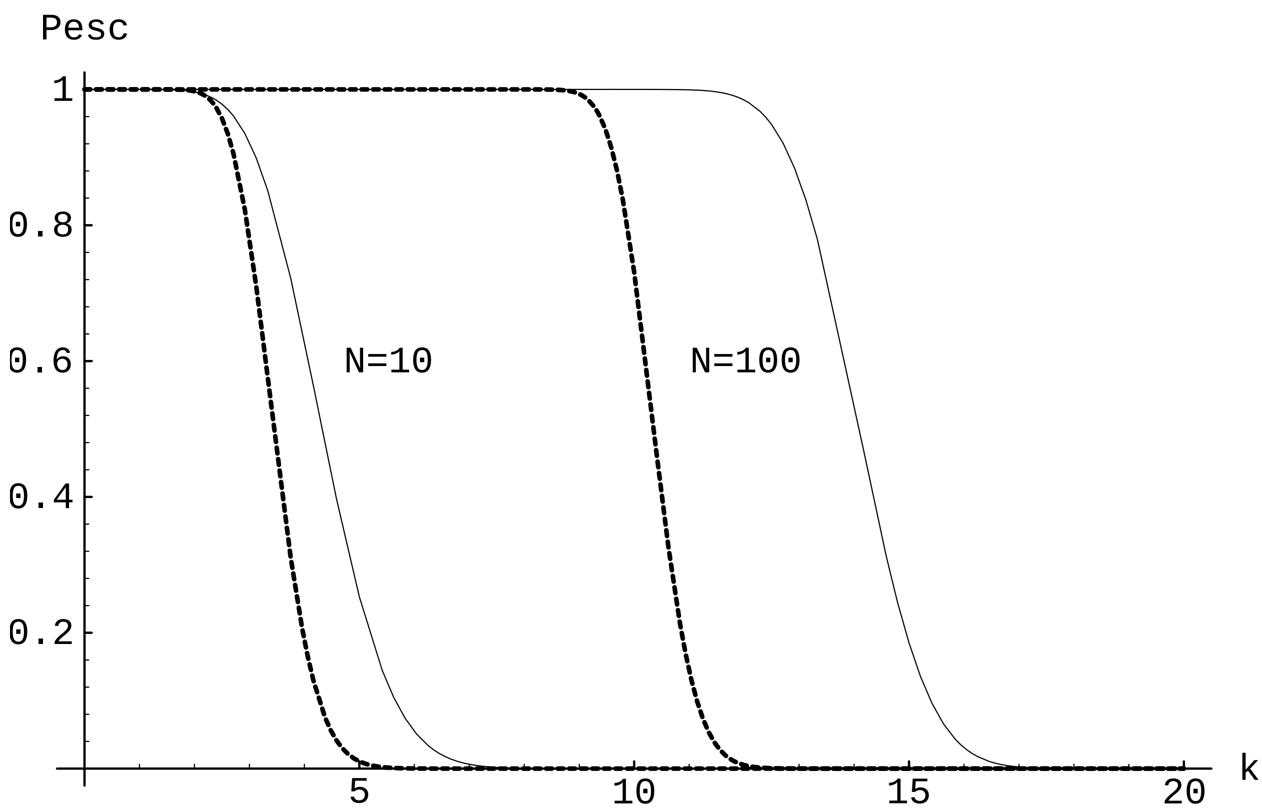
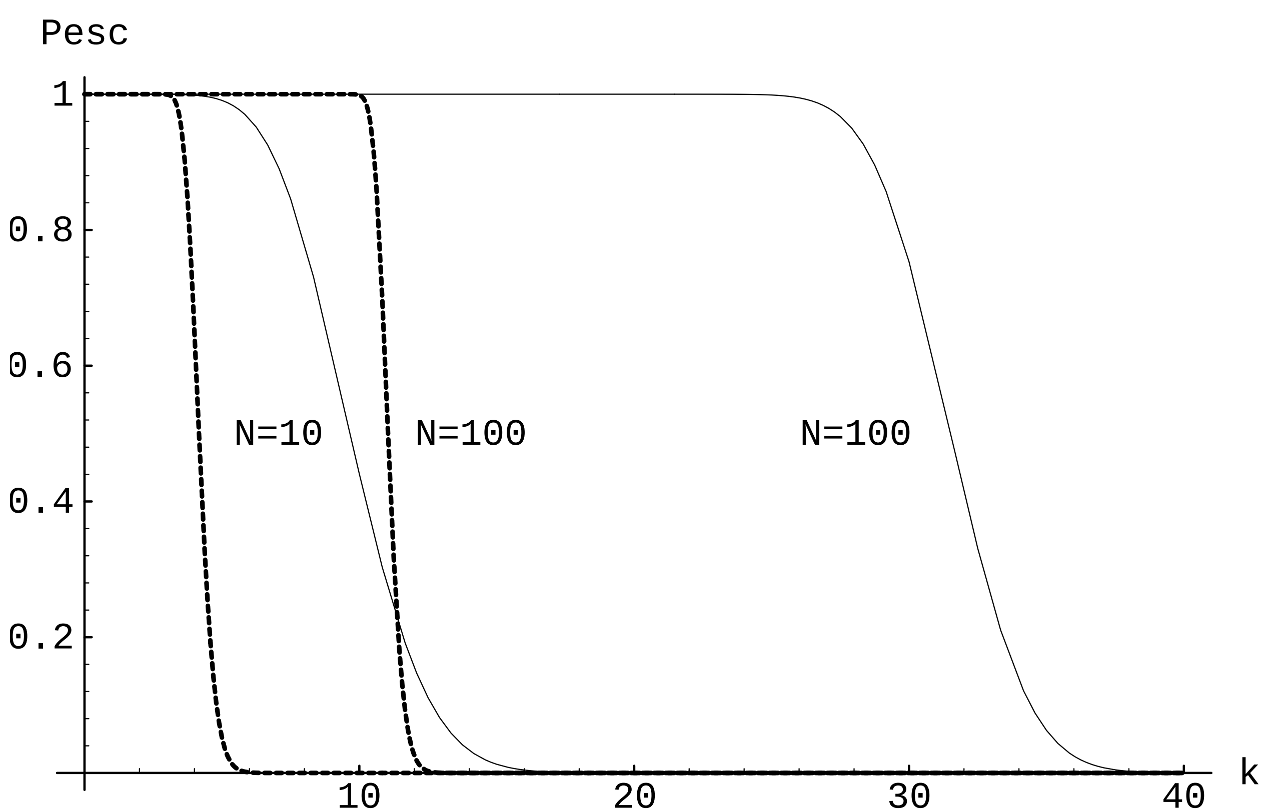


Figure 2. The escape probabilities for hard selection (bold-dashed line) and soft selection (thin solid line) for n = 10 or 100 dimensions as a function k for the cases of (a) g = 2, and (b) g = 10. The probability of escaping a local optimum under the assumptions given in the text is seen to be greater when using soft selection.


Pesc
1
0.8
0.6
0.4
0.2
N=10
N=100
k
5
10
15
20


Pesc
1
0.8
0.6
0.4
0.2
N=10
N=100
N=100
k
10
20
30
40