ELSEVIER

Information Processing Letters 82 (2002) 15–22

Information Processing Letters

www.elsevier.com/locate/ipl

Order statistics and selection methods of evolutionary algorithms ☆

Erick Cantú-Paz

Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, Livermore, CA 94551, USA

Abstract

Selection methods are essential components of evolutionary algorithms (EAs). This paper reviews five popular selection methods used in EAs. The algorithms are examined using the cumulants of the fitness distribution of the selected individuals. The cumulants are calculated using order statistics. The method presented here considers finite populations of arbitrary size. The results show important differences among the selection methods considered, even when they are configured to have the same selection intensity. Published by Elsevier Science B.V.

Keywords: Analysis of algorithms; Linear ranking; Exponential ranking; Tournament selection; Boltzmann selection; Truncation selection; Selection intensity; Fitness distribution; Order statistics

1. Introduction

Evolutionary algorithms (EAs) have two basic components: selection of the fittest and randomized operators that create new solutions from the selected ones. For EAs to succeed, there must be a balance of the selection mechanism that exploits the information gathered about the problem with the genetic operators that explore new solutions (e.g., [11,7]). To achieve this balance, we must understand how selection affects the composition of the population.

Various methods have been used to quantify the effect of the selection pressure that selection algorithms exert on the population. Goldberg and Deb introduced the takeover time [8], which is the number of generations that the selection algorithm takes to reproduce a single representative of the optimal solution to occupy the entire population. High selection pressures result in short takeover times. Others quantify the selection pressure using the selection intensity [1,12,19], which is the increase of the mean fitness of the population after selection normalized by the standard deviation. A high selection intensity corresponds to a high selection pressure. The selection intensity is related to the population size [9], the optimal mutation rate, and the speed of convergence [13]. More sophisticated models of selection consider the cumulants of the distribution of fitness [14]. This approach has been applied to noisy fitness environments [15] and EAs with overlapping populations [16].

The objective of this paper is to review popular selection methods and examine them using the cumulants of the fitness distribution. The calculations consider finite populations of arbitrary size, and the results show important differences among the selection methods considered, even when they are configured to have the same intensity. The accuracy of the calculations is verified with experiments. The observations in this pa-

☆ UCRL-JC-138582. This work was performed under the auspices of the U.S. Department of Energy by University of California LLNL under contract No. W-7405-Eng-48.

E-mail address: cantupaz@llnl.gov (E. Cantú-Paz).

0020-0190/01/$ – see front matter Published by Elsevier Science B.V.
PII: S0020-0190(01)00282-4

per shed some light on reports regarding the success or failure of a particular combination of a selection algorithm and variation operators.

Blickle and Thiele [3] studied the effect of selection on the first two cumulants of the fitness distribution and considered some of the selection methods studied here. This paper extends their work by considering additional selection methods (Boltzmann, tournaments without replacement) and higher cumulants of the fitness distribution. In addition, the method presented here is easy to extend to other selection methods (such as the bi-linear selection of Voss [20]) and migration of individuals across multiple populations [5].

2. Describing the fitness distribution

The approach of this paper is to describe the distribution of fitness using its cumulants. The cumulants are related to the central moments. The rth central moment of the fitness of a population of size n is

$$\mu_r = \frac{1}{n}\sum_{i=1}^{n}(f_i - \bar{f})^r, \tag{1}$$

where f_i is the fitness of the ith individual, and

$$\bar{f} = \frac{1}{n}\sum_{i=1}^{n} f_i.$$

The first three cumulants are equal to the first central moments. The fourth cumulant is $\kappa_4 = \mu_4 - 3\mu_2^2$. The first cumulant is the mean, and the second is the variance. Sometimes the third and fourth cumulants are normalized by dividing them by $\kappa_2^{r/2}$ to obtain the skewness and kurtosis coefficients. The skewness is negative if the distribution is skewed to the left and positive if the distribution is skewed to the right. The kurtosis is negative if the distribution is flatter than a normal, and positive if the distribution is more peaked than a normal.

One possibility to calculate the expected value of the cumulants is to integrate over all possible populations after selection [18]. Our approach is different: we calculate the expected fitness of each individual, and then use Eq. (1) to obtain the cumulants. The critical observation is that we may interpret the fitness values f_i as samples of random variables F_i with a common distribution. We may arrange the random variables in non-decreasing order to obtain the order statistics: $F_{1:n} \leqslant F_{2:n} \leqslant \cdots \leqslant F_{n:n}$. Without loss of generality, in the remainder we assume a maximization task, and we normalize the random variables as

$$Z_{i:n} = \frac{F_{i:n} - \overline{F}}{\sigma_F}. \tag{2}$$

The expected value of the ith order statistic is

$$\begin{aligned}\mu_{i:n} &= \mathrm{E}(Z_{i:n}) \\ &= n\binom{n-1}{i-1}\int_{-\infty}^{\infty} z\phi(z)\big[\Phi(z)\big]^{i-1} \\ &\qquad\cdot\big[1-\Phi(z)\big]^{n-i}\,\mathrm{d}z,\end{aligned} \tag{3}$$

where $\phi(z)$ and $\Phi(z)$ are the PDF and CDF of fitness, respectively.

Each individual has a certain probability p_i of being selected that depends on the selection method. Using these probabilities the mean of the population after selection is

$$\overline{Z}_{\mathrm{sel}} = \sum_{i=1}^{n} p_i\mu_{i:n}, \tag{4}$$

with $\sum_{j=1}^{n} p_j = 1$. The equation above gives the selection intensity of the selection method. The expected rth moment after selection is

$$\mu_r^{\mathrm{sel}} = \sum_{i=1}^{n} p_i\big(\mu_{i:n} - \overline{Z}_{\mathrm{sel}}\big)^r. \tag{5}$$

This equation can be used to describe the effects of selection on the population if the probabilities p_i are known and $\mu_{i:n}$ can be calculated.

3. Selection methods

This section presents calculations that assume that the fitness has a normal distribution with zero mean and unit standard deviation. Other distributions can be used, as long as we can calculate the expected value of the order statistics by substituting the corresponding PDF and CDF in Eq. (3). Using the normal distribution allows us to verify the results with previous publications that made the assumption of normality. In addition, we can use an approximation introduced by Harter [10] to calculate rapidly the expected value of the order statistics of normal distributions.

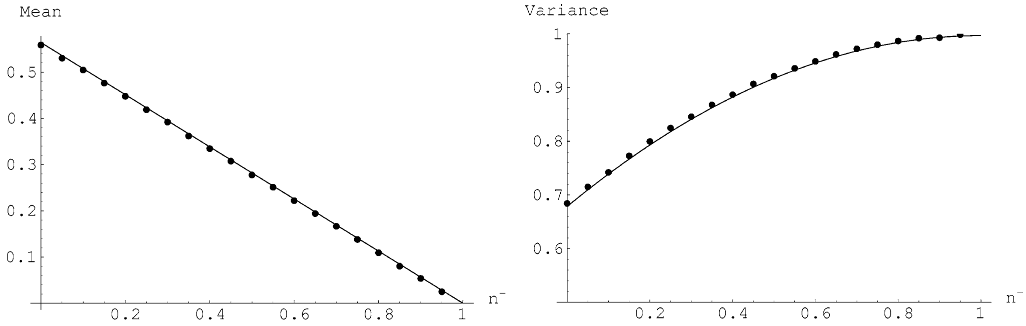

Fig. 1. Mean and variance of fitness after linear ranking selection varying the expected number of copies assigned to the worst individual, n^-.

Simple experiments were used to verify the accuracy of the theoretical calculations. The fitness function was a $l = 500$ bits onemax $F = \sum_{j=1}^{l} x_j$, where $x_j \in \{0, 1\}$ are the individual bits in the chromosome. The population was initialized randomly, so F has a binomial distribution. The fitness values were normalized to obtain a distribution with zero mean and unit standard deviation, which approximates the assumed unit Gaussian. The experiments used populations with $n = 1000$ individuals, and were repeated 1000 times with different random seeds. Using a large population size was intentional to test whether Harter's approximation was sufficiently accurate for our purposes, even if we exceeded the recommended maximum sample size of 400 [10]. In the graphs, the experimental results are presented by dots with error bars that represent 95% confidence intervals, but in most cases the error bars are not visible, confirming that the experiments and the theory match closely. Tournament selection was implemented directly from its description, and in the other experiments the individuals were chosen using stochastic universal sampling [2].

3.1. Linear ranking

In linear ranking selection, individuals are sorted according to their fitness, and the probability of being selected is linearly proportional to their rank, i, in the population [2]:

$$p_i = \left(n^- + \frac{n^+ - n^-}{n-1}(i-1)\right)\bigg/ Z, \tag{6}$$

where the desired number of copies of the best (n^+) and worst (n^-) individuals are parameters of the algorithm; $Z = n$ is a normalizing factor; and to ensure that that the probabilities add to 1, $n^- + n^+$ must equal 2.

Fig. 1 shows the first two cumulants of the selected individuals varying n^-. The plots show that linear ranking is very weak: at its strongest setting the selection intensity is $1/\sqrt{\pi} = 0.5642$ [3]. Section 4 shows that at all possible settings of n^-, the higher cumulants remain close to their original values.

3.2. Exponential ranking

An alternative to the weak linear ranking is to assign survival probabilities to the sorted individuals using an exponential function [2]:

$$p_i = \frac{c^{n-i}}{Z}, \tag{7}$$

where $c \in [0, 1]$ is a parameter of the algorithm. Since

$$Z = \sum_{j=1}^{n} c^{n-j} = \frac{c-1}{c^n - 1},$$

p_i can be simplified to

$$p_i = \frac{c-1}{c^n - 1} c^{n-i}. \tag{8}$$

Fig. 2(left) shows that for most of the range of c, exponential ranking has a much higher selection intensity than linear ranking. However, for values of c that produce a high selection intensity, exponential ranking also results in a distribution with low variance. A steep loss of variance may be problematic because it suggests that the population loses diversity very rapidly,

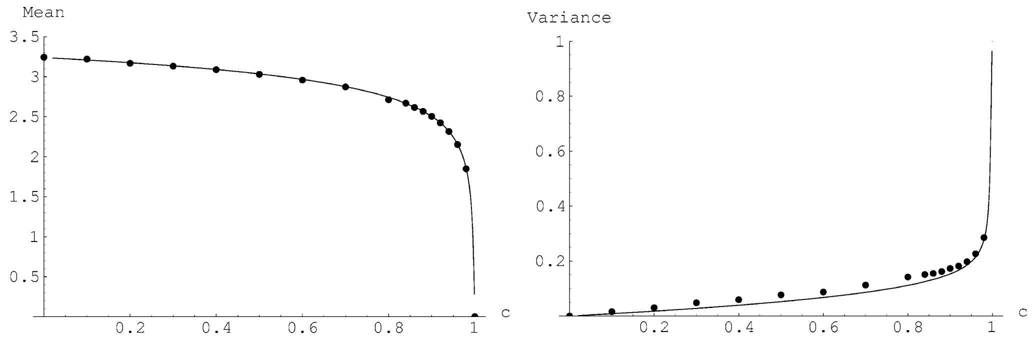

Fig. 2. Mean and variance of fitness after exponential ranking selection varying the parameter c.

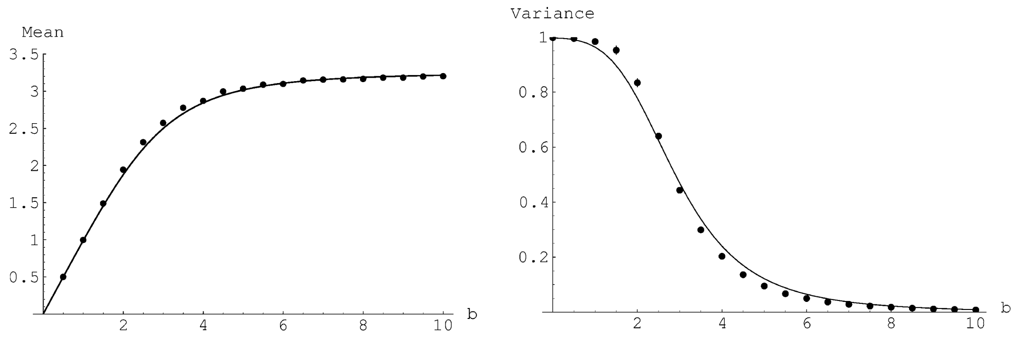

Fig. 3. Mean and variance of fitness after Boltzmann selection varying the parameter β.

making the algorithm more likely to converge to a suboptimal solution unless the variation operators reintroduce sufficient diversity. This may explain why exponential ranking is commonly used with values of c close to 1.

3.3. Boltzmann selection

In this form of selection, the probability of selection is [6]:

$$p_i = \frac{\exp(\beta f_i)}{Z}, \tag{9}$$

where β controls the selection intensity, and $Z = \sum_{j=1}^{n} \exp(\beta f_j)$. Rogers and Prügel-Bennett [16] determined analytically that for weak settings the selection intensity is approximated by β. This can be confirmed in Fig. 3(left), where the plot appears linear for $\beta < 2$. The mean fitness of the selected individuals does not increase much after $\beta > 3$, but the variance decreases significantly. Thus, it may not be advisable to use Boltzmann selection with $\beta > 3$, unless it is used with aggressive mutation or recombination that reintroduce some diversity.

3.4. Truncation selection

Truncation selection deterministically selects the top τ individuals in the population [13]. We can cast this deterministic method into our probabilistic framework by defining p_i as

$$p_i = \begin{cases} 0 & \text{if } i < n - \tau, \\ 1/\tau & \text{if } i \geqslant n - \tau. \end{cases} \tag{10}$$

Fig. 4 shows the mean and variance of the selected individuals varying τ. Choosing a small value of τ increases the mean fitness of the population significantly, but it also reduces the variance rapidly.

It is common in evolution strategies to set τ close to 0.15 (1/7) [17] and use aggressive mutation and recombination operators. In contrast, in genetic algo-

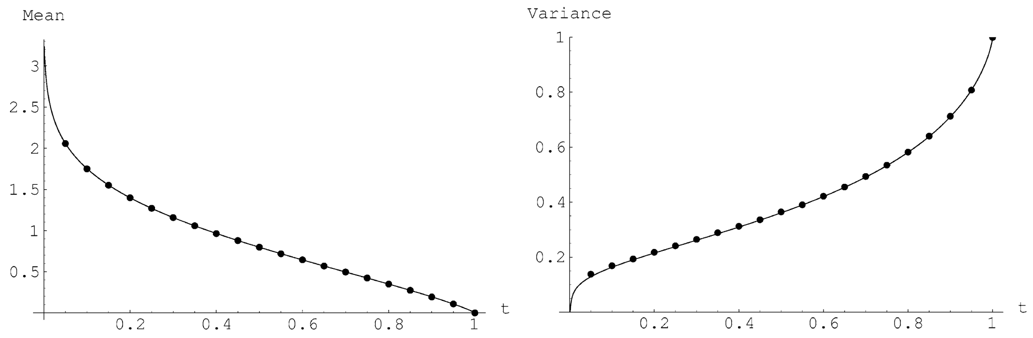

Fig. 4. Mean and variance of fitness after truncation selection varying the selection threshold τ.

rithms truncation selection is used with weaker settings (e.g., $\tau \approx 0.5$ [13]) and less aggressive operators.

3.5. *Tournament selection*

In tournament selection the best from a random sample of s individuals is chosen for the next generation [4]. The samples may be drawn with or without replacement.

An individual will win a tournament only if its fitness is greater than the fitness of the other $s-1$ competitors. If sampling with replacement, the probability that a randomly chosen individual has a lower fitness than the ith individual is the probability that a uniform random number in $[1, n]$ is less than or equal to i:

$$P(f_j \leqslant f_i) = P(j \leqslant i) = \frac{i-1}{n-1}.$$

In a tournament with s participants, the probability that all of the opponents of the ith individual are ranked lower than i is $P(j \leqslant i)^{s-1}$, so the probability that the ith individual survives is

$$p_i = \left(\frac{i-1}{n-1}\right)^{s-1} \Big/ Z, \tag{11}$$

where

$$Z = \sum_{j=1}^{n} \left(\frac{j-1}{n-1}\right)^{s-1}.$$

Note that the worst individual never survives, and the best individual wins in all the tournaments it participates in. However, when sampling with replacement it is possible that the best individual will never be selected to participate in a tournament.

If sampling without replacement, the probability that exactly $x = s-1$ individuals are chosen among those with lower fitness than individual i has a hypergeometric distribution:

$$P(x = s-1) = \frac{\binom{i-1}{s-1}\binom{n-i-1}{0}}{\binom{n-1}{s-1}} = \frac{(i-1)!(n-s)!}{(n-1)!(i-s)!},$$

and normalizing by dividing over the sum of all possibilities gives the probability that the ith individual is selected as

$$p_i = \frac{(i-1)!(n-s)!}{(n-1)!(i-s)!} \Big/ Z. \tag{12}$$

When $s = 2$ the probability of surviving the tournaments is identical when sampling with and without replacement. The difference becomes larger with larger tournaments and smaller populations, but in many practical situations the differences are so small that they may be considered negligible. It is known that the distribution of individuals selected by pairwise tournaments ($s = 2$) and linear ranking with $n^- = 0$ are identical [3,8], and this can be appreciated by comparing Figs. 5 and 1.

4. Comparing the selection methods

Although it is convenient to compare selection methods by their takeover time or selection intensity, these may not be sufficiently complete representations. To understand why and to facilitate comparisons, Figs. 6 and 7 plot the second and third cumulants of the selected individuals vs. the selection intensity.

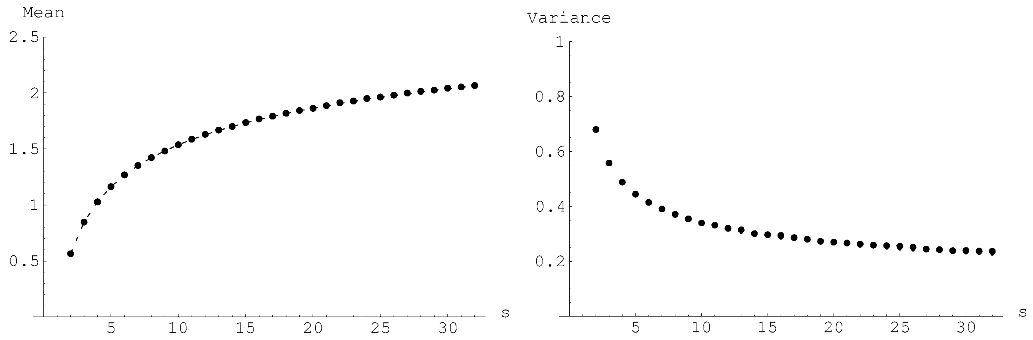

Fig. 5. Mean and variance of fitness after tournament selection without replacement varying the tournament size s from 2 to 32.

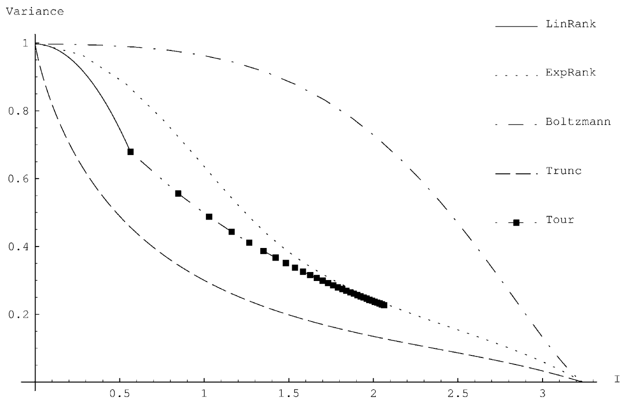

Fig. 6. Comparison of the variance of fitness varying the selection intensity.

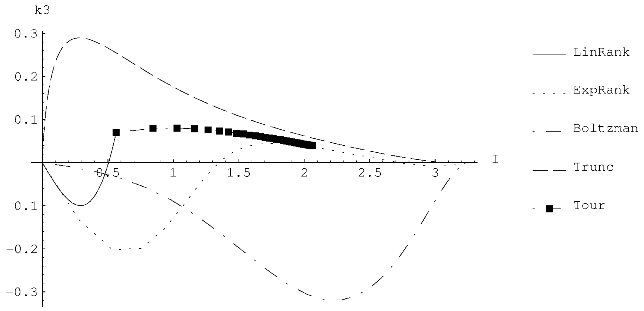

Fig. 7. Comparison of the third cumulant of the fitness distribution varying the selection intensity.

Fig. 6 shows that for equivalent selection intensities, Boltzmann selection preserves the original fitness variance more than the other methods, while truncation selection results in the largest reduction of variance. The fitness variance is a good indicator of population diversity in some cases, especially on those where the fitness is determined by an additively-decomposable function where the components are scaled equally, like in the onemax function used in the experiments. Past experience suggests that it is a good heuristic to prefer the selection method that preserves the greater diversity (e.g., [1]), but we must be cautious before dismissing the other methods. Severely reducing the diversity may be useful if the variation operators produce very diverse individuals, which is the case in many successful applications of evolution strategies.

Fig. 7 shows that starting from a unit normal, some methods always produce a distribution that is skewed either to the left ($\kappa_3 < 0$) or to the right ($\kappa_3 > 0$), but for linear and exponential ranking the direction of the skewness depends on the selection intensity. The sign of the third moment is interesting, because it clearly shows a bias in the way the individuals are sampled. A population with positive skewness has fewer samples of the fitter individuals than a distribution with negative skewness. One could argue that having many highly-fit individuals is an advantage, because it is more likely that they will remain intact after mutation or recombination. However, it is not clear that this is an advantage in all problems or with all variation operators. Similar arguments could be made for preferring flat or peaky distributions. These biases should be taken into consideration when choosing or designing variation operators to avoid unbalanced combinations with selection methods, which may result in an inadequate search of the solution space.

5. Conclusions

This paper presented an overview and comparison of some selection algorithms and a method to calculate their effect on the cumulants of the fitness distribution. The comparisons show that Boltzmann selection is very promising, because for a wide range of settings it preserves the variance and the shape of the original distribution more than other methods with the same selection intensity.

We must emphasize that there is no single best selection method. The success of evolutionary algorithms depends on the balance between selection and the variation operators. It is quite possible that a selection method that works well in combination with certain operators on a particular problem may be a poor choice in a different setting.

References

[1] T. Bäck, Generalized convergence models for tournament- and (μ, λ)-selection, in: Proceedings of the 6th International Conference on Genetic Algorithms, Morgan Kaufmann, San Francisco, CA, 1995, pp. 2–8.

[2] J.E. Baker, Adaptive selection methods for genetic algorithms, in: Proceedings of an International Conference on Genetic Algorithms and Their Applications, Lawrence Erlbaum Associates, Hillsdale, NJ, 1985, pp. 101–111.

[3] T. Blickle, L. Thiele, A comparison of selection schemes used in evolutionary algorithms, Evolutionary Comput. 4 (4) (1996) 361–394.

[4] A. Brindle, Genetic algorithms for function optimization, Unpublished doctoral dissertation, University of Alberta, Edmonton, AB, 1981.

[5] E. Cantú-Paz, Efficient and Accurate Parallel Genetic Algorithms, Kluwer Academic Publishers, Boston, MA, 2000.

[6] M. de la Maza, B. Tidor, An analysis of selection procedures with particular attention paid to proportional and Boltzmann selection, in: Proceedings of the 5th International Conference on Genetic Algorithms, Morgan Kaufmann, San Mateo, CA, 1993, pp. 124–131.

[7] D.E. Goldberg, The race, the hurdle, and the sweet spot: Lessons from the genetic algorithms for the automation of design innovation and creativity, in: Evolutionary Design by Computers, Morgan Kaufmann, San Mateo, CA, 1999.

[8] D.E. Goldberg, K. Deb, A comparative analysis of selection schemes used in genetic algorithms, Foundations of Genetic Algorithms 1 (1991) 69–93.

[9] G. Harik, E. Cantú-Paz, D.E. Goldberg, B.L. Miller, The gambler's ruin problem, genetic algorithms, and the sizing of populations, Evolutionary Comput. 7 (3) (1999) 231–253.

[10] H.L. Harter, Order Statistics and Their Use in Testing and Estimation, U.S. Government Printing Office, Washington, DC, 1970.

[11] J.H. Holland, Genetic algorithms and the optimal allocation of trials, SIAM J. Comput. 2 (2) (1973) 88–105.

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

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

[14] A. Prügel-Bennett, J.L. Shapiro, An analysis of a genetic algorithm using statistical mechanics, Phys. Rev. Lett. 72 (9) (1994) 1305–1309.

[15] M. Rattray, J. Shapiro, Noisy fitness evaluation in genetic algorithms and the dynamics of learning, in: Foundations of Genetic Algorithms 4, Morgan Kaufmann, San Francisco, CA, 1997, pp. 117–139.

[16] A. Rogers, A. Prügel-Bennett, Modelling the dynamics of a steady state genetic algorithm, in: Foundations of Genetic Algorithms 5, Morgan Kaufmann, San Francisco, CA, 1999, pp. 57–68.

[17] H.P. Schwefel, Numerical Optimization of Computer Models, John Wiley and Sons, Chichester, 1981.

[18] J. Shapiro, A. Prügel-Bennett, M. Rattray, A statistical mechanical formulation of the dynamics of genetic algorithms, in: Evolutionary Computing: AISB Workshop 1994, Springer, Berlin, 1994, pp. 17–27.

[19] D. Thierens, Selection schemes, elitist recombination, and selection intensity, in: Proceedings of the 7th International Conference on Genetic Algorithms, Morgan Kaufmann, San Francisco, CA, 1997, pp. 152–159.

[20] M.S. Voss, C.M. Foley, Evolutionary algorithm for structural optimization, in: Proceedings of the Genetic and Evolutionary Computation Conference, 1999, pp. 678–685.