An Analysis of Local Selection Algorithms in a Spatially Structured Evolutionary Algorithm

Jayshree Sarma
Computer Science Department
George Mason University
Fairfax, VA 22030
jsarma@gmu.edu

Kenneth De Jong
Computer Science Department
George Mason University
Fairfax, VA 22030
kdejong@gmu.edu

Abstract

A critical part of the design of effective EAs is to obtain a proper balance between exploration and exploitation. A key element of this balance is the selection algorithm used. Although we have for some time now understood the relative differences between various selection methods for standard centralized EAs, our understanding of decentralized EAs has been less complete. In this paper we present analysis tools and results which provide considerable insight into the properties of various decentralized selection methods, including the selection pressures they induce. Understanding this permits a more informed choice of selection methods when designing and implementing spatially structured EAs.

1 INTRODUCTION

With the increasing availability of parallel machines many different forms of spatially structured evolutionary algorithms (EAs) are being developed. In these EAs the population is distributed over a grid-like topology which is generally (although not necessarily) a two dimensional torus, and selection, mating, reproduction, etc. operate in a distributed fashion within local overlapping neighborhoods (Collins and Jefferson 1991; Gorges-Schleuter 1989; McInerney 1992; Spiessens and Manderick 1991).

These distributed, spatially structured EAs behave quite differently than the more familiar and better understood centralized EAs. The primary source of these differences is due to the effects of local (rather than global) selection algorithms. Understanding the effects of particular local selection algorithms will significantly increase our ability to design effective spatially structured EAs.

There are several decisions which need to be made in designing a structured EA. The first, of course, is the topology itself. This may be determined by the underlying machine architecture or by the application area itself. In this paper we focus on a single topology, the frequently used two dimensional torus. A typical example would be an $N \times N$ grid in which a single individual exists at each grid point and interacts with its neighbors on nearby grid points.

That interaction takes the form of running a local EA in parallel on each grid point for the purpose of replacing the current individual at each grid point with a new one (the next generation). This is accomplished by selecting parents from a local neighborhood and applying the usual reproductive mechanisms of crossover, mutation, etc. In order to implement such systems one must decide on the neighborhood size, it's shape, and the particular selection algorithm to be used. In the past such choices have been made primarily by experimental studies. The goal of the work presented here is to provide a more analytical basis for such decisions.

2 BACKGROUND

The first, and somewhat surprising result is that, if we use one of the standard selection methods such as fitness proportional, linear ranking, or tournament selection in the local neighborhoods, the resulting *emergent* global selection pressure on the entire population is qualitatively similar to, but quantitatively weaker than the corresponding centralized, sequential EA (De Jong and Sarma 1995). Since selection in sequential EAs is well understood theoretically (Goldberg and Deb 1991; Bäck 1995), this means that much of our understanding of the relative differences between selection methods transfers directly to these parallel models.

The fact that local selection algorithms are quantitatively weaker than their global counterparts is due to the fact that it now takes longer for good individuals to propagate throughout the population since they can only do so by spreading through adjacent neighborhoods. For a given local selection algorithm the rate of propagation is determined by both the size and the shape of the local neighborhoods relative to the total size and shape of the spatial grid. Sarma and De Jong (1996) showed that propagation rates can be determined directly by calculating the radius of a neighborhood using a distance measure adapted from spatial analysis techniques:

$$rad = \sqrt{\frac{\sum(x_i - \bar{x})^2 + \sum(y_i - \bar{y})^2}{n}}$$

where $(\bar{x}, \bar{y})$ is the center grid point of a neighborhood of size n. Selection pressure can thus be increased (decreased) by increasing (decreasing) the radius of the neighborhood relative to the radius of the total grid.

One simplified, but useful way to visualize the differences in selection pressure is to look at the growth curves of the best individual in an infinite population model in the absence of crossover or mutation. In such cases the differences in selection pressure are seen in terms of the rate at which the best individual takes over the entire population.

Figure 1 illustrates this for two selection methods: fitness proportional and linear ranking. The strongest selection pressure is obtained when using a centralized, sequential EA with a global selection pool. With distributed, parallel EAs using local selection pools (neighborhoods) the *emergent* selection pressure is weaker, but can be varied by changing the radius of the local neighborhood.

Just as in the centralized case, the growth curves generated by the local neighborhood models can be closely approximated by a logistic equation of the form:

$$P_{b,t} = \frac{1}{1 + (\frac{1}{P_{b,0}} - 1)e^{-at}}$$

where a is the growth coefficient and $P_{b,t}$ is the proportion of the best in the population at time t.

In Sarma and De Jong (1996) it was shown that, for a given selection method, there is a simple inverse exponential relationship between the growth coefficient a and the ratio of the neighborhood radius to the grid radius. Figure 2 shows this relationship for fitness proportional and linear ranking. This means that, for

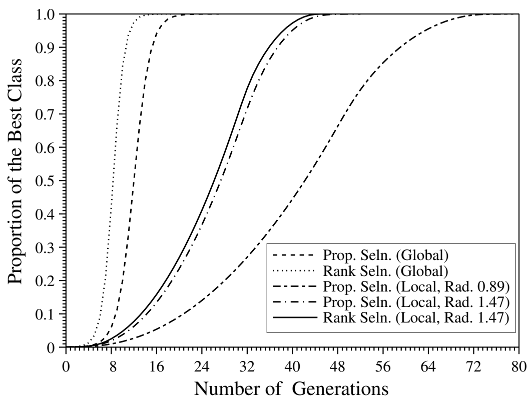

Figure 1: Growth curves for fitness proportional and linear rank selection for a global pool of size 4096 and local pools of size 5 and 13 on a 64×64 toroidal grid.

a particular selection method, we can predict growth curves (and hence selection pressure) directly from the radii ratio. Alternatively, for a given neighborhood structure one can precisely predict the effects of switching to an alternative selection method.

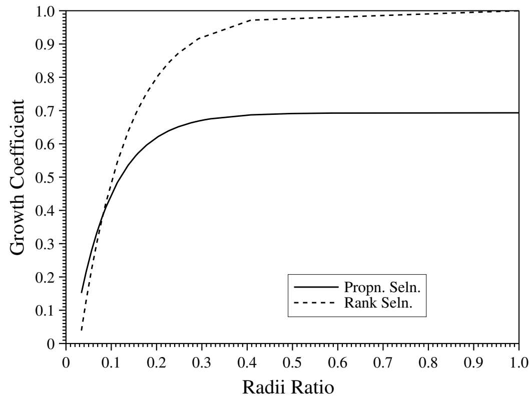

Figure 2: Growth curve coefficients for fitness proportional and linear ranking selection as a function of the radii ratios.

These results now provide the tools and the insights required for a complete analysis of local selection algorithms. Fitness proportional and ranking selection were analyzed in (Sarma and De Jong 1996). In this paper we analyze the two remaining selection methods commonly used in spatially structured EAs: local tournament and random walk selection.

3 TOURNAMENT SELECTION

Tournament selection involves randomly choosing k individuals with uniform probability from the selection pool. The individual with the highest fitness is declared the winner and selected to be a parent. We know that in standard sequential EAs using global selection pools, binary tournament selection ($k = 2$) results in the same expected selection pressure as linear rank selection and hence produces identical growth curves (Goldberg and Deb 1991). Moreover, increasing the tournament size k increases the selection pressure.

The results summarized in the previous section predict that we should observe the same effects when using tournament selection on local selection pools, and this is precisely what we see. Figure 3 illustrates these effects for a grid size of 64×64 and neighborhood size 13. The emergent growth curves of binary tournament and linear ranking selection are identical and stronger than fitness proportional selection. Also the emergent selection pressure induced by tournament selection becomes stronger when the tournament size k is increased.

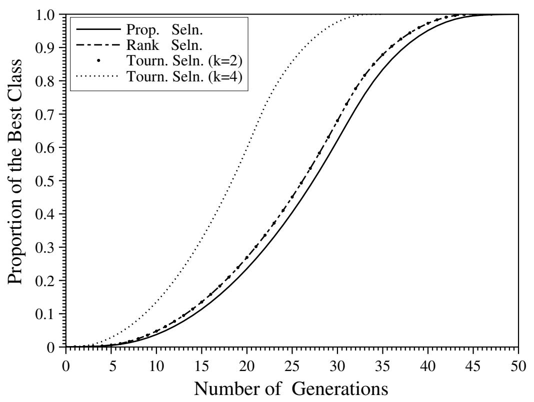

Figure 3: Emergent growth curves for fitness proportional, linear ranking, and tournament selection ($k = 2$ and $k = 4$) using a 64×64 toroidal grid and neighborhood size 13.

Tournament selection has an advantage over linear ranking in that it does not require sorting the selection pool. This advantage is intensified when local selection is done on finely grained parallel architectures in which there are significant communication costs in accessing information about all the grid points in a neighborhood. Hence, for spatially distributed EAs one can use binary tournament selection to achieve selection pressure identical to linear ranking without the additional overhead.

4 SELECTION VIA RANDOM WALKS

One way to think of the local section methods discussed so far is that they introduce a rather crude distance bias in which individuals within the neighborhood are "visible" and those outside "invisible" from the point of view of interaction. One can imagine other forms of distance bias in which the probability of interaction decreases as a function of distance. A natural bias of this type is obtained by taking random walks of length r from a grid point (Hillis 1991; Collins 1992; McInerney 1992), and selecting the individual at the endpoint of the random walk.

This induces both a neighborhood size and shape (the set of grid points reachable in exactly r randomly chosen steps) and a distance bias on the members of the neighborhood in that random walks are more likely to terminate at grid points close to the starting point. For continuous spaces the actual probability distribution induced by a random walk is Gaussian; for discrete spaces it is binomial.

In the standard selection algorithms (fitness proportional, linear ranking, and tournament selections) discussed earlier, all the individuals in the neighborhood are equally likely to be chosen, and selection was biased only by the fitness values. In selection via random walk, individuals closer to the grid point have a higher probability of being chosen; hence now we have both a distance and a fitness bias in the selection process.

4.1 NEIGHBORHOOD SIZES AND SHAPES

The neighborhoods generated by random walk selection depends on the length of the walk and the directions traversed. If steps are allowed in only four directions (up, down, left or right), the neighborhoods induced are diamond (D) shaped. If diagonal steps are also permitted, then compact square neighborhoods (C) are generated. Figures 4 and 5 illustrate this with some typical examples. The choices regarding the length and possible directions of a random walk are generally motivated by the interconnection architecture of the underlying parallel machine. The important point here is to note that random walks of length r in eight directions result in significantly larger neighborhoods than those involving only four directions, and hence involve larger radii.

For small values of r obtaining the endpoint of a ran-

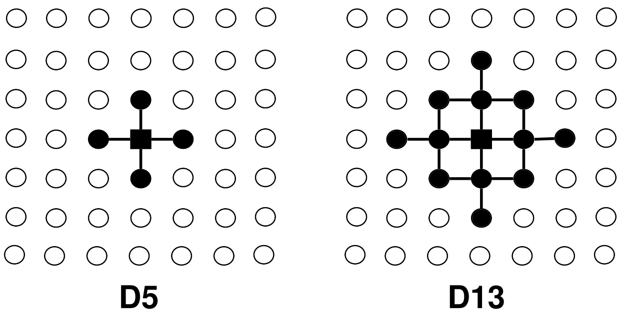

Figure 4: Neighborhood Shape D - Movement in four directions for random walk lengths 1 and 2.

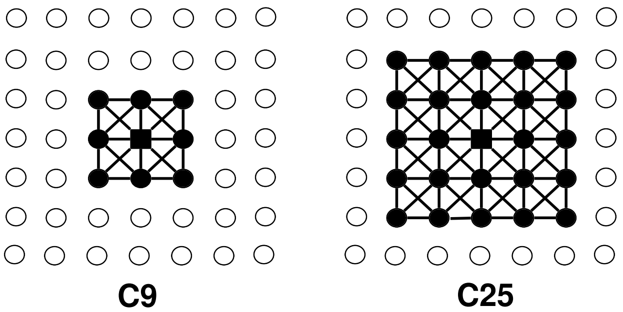

Figure 5: Neighborhood Shape C - Movement in eight directions for random walk lengths 1 and 2.

dom walk is usually accomplished by simulating a random walk. For larger values of r it is generally more efficient to obtain the endpoints by randomly sampling from the underlying induced probability distribution.

4.2 TOURNAMENT SELECTION VIA RANDOM WALKS

A natural and frequently used way to introduce a distance bias into the selection mechanism is via tournament selection using random walks. That is, rather that selecting k members uniformly from a neighborhood, k random walks are performed and the endpoints of the walks are used as the competitors of the tournament. By doing so, individuals in the neighborhood which are closer to the starting grid point are more likely to be selected than those that are further away.

What effect will this have on the overall emergent selection pressure? Our analysis tools provide a precise characterization. Introducing a distance bias has the effect of reducing the effective radius of the neighborhood. More formally:

$$rad = \sqrt{\left[\frac{\sum(x_i - \bar{x})^2 + \sum(y_i - \bar{y})^2}{n}\right] * p_i}$$

where p_i is the probability of selecting grid point i.

This weighted radius can be used to directly predict the corresponding growth coefficient using our inverse exponential mapping scheme. For example, if we were currently using standard binary tournament selection, and would like to understand the effect of adding a distance bias by switching to selection via two random walks of length r, we first compute the new effective radius using the formula above. Since binary tournament selection and rank selection have identical growth coefficients, we can use the inverse exponential relationship given in Figure 2 to predict growth curves for distance biased binary tournament selection. For example, with a grid size of 64×64 and a neighborhood size of 49 (C49) induced by random walk of length 3 in eight directions, the neighborhood radius is 2.12 and a radii ratio of 0.08. Using the predicted growth coefficient, we can compute the growth curve. This predicted growth curve is compared with the actual growth curve in Figure 6.

4.3 SELECTION VIA PATH ELITISM

Since tournament selection involving multiple random walks can be expensive to implement, a variation involving "path elitism" is sometimes used. In this form

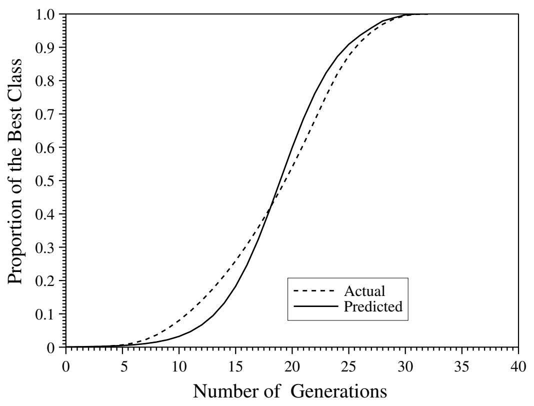

Figure 6: Predicted and actual growth curves for binary tournament selection using random walks of length 3 in 8 directions on a 64 × 64 toroidal grid.

a single random walk of length r is taken, and every individual encountered along the way participates in the tournament. This variation typically includes the individual at the starting grid point in the competition, and thus implicitly includes local elitism (Collins and Jefferson 1991; McInerney 1992).

How does such a change affect selection pressure? We note that r now specifies *both* the length of the random walk *and* the tournament size ($k = r$). Hence, we would expect the emergent selection pressure of path elitism of order r to be similar to that produced by a tournament of size r using the endpoints of random walks to select the competitors.

However, a single path generated by taking a random walk of length r is contiguous and more compact than the endpoints generated by taking r random walks, reducing the effective radius of the neighborhoods. So, for example, the effective radius of the neighborhood created by the single path of a random walk of length 3 in eight directions (C49 neighborhood) is 1.54 compared to a radius of 2.12 for the endpoint case.

Since the radius has decreased one can expect a lower selection pressure compared to the endpoint tournament selection. We can use the reduced radius to derive a growth coefficient as before. Figure 7 illustrates this by comparing the predicted and the actual growth curves for the particular path selection mechanism described above in a 64 × 64 grid. Note that the prediction is not as accurate as before. This is because we are using the growth coefficient model for non-elitist tournament selection to predict the behavior of an elitist selection method. However, simply adding local elitism to the model results in higher predicted growth rates and a better fit.

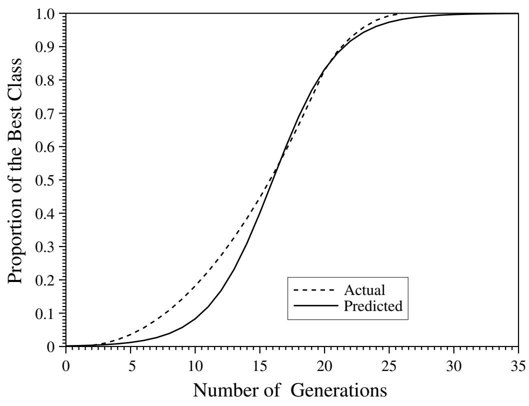

Figure 7: Predicted and actual growth curves for path elitism selection using a random walk of length 3 in 8 directions on a 64 × 64 toroidal grid 3.

It should also now be clear that we can predict the effects on selection pressure when varying the parameters of path elitism. If we increase r, we increase the neighborhood size and its radius, resulting in increased selection pressure. If we reduce the number of legal moves from eight to four, the neighborhood size and its radius is decreased, resulting in a reduction in selection pressure. For example, a random walk of length 2 in eight directions (neighborhood C25) has an effective radius of 1.25 while a random walk of the same length 2 in four directions (neighborhood D13) has an effective radius of only 1.04. Hence the selection pressure induced by neighborhood C25 should be stronger than that of neighborhood D13. This is illustrated in Figure 8.

5 CONCLUSION

A critical part of the design of effective EAs is to obtain a proper balance between exploration and exploitation. A key element of this balance is the selection algorithm used. Although we have for some time now understood the relative differences between various selection methods for our traditional, centralized EAs, our understanding of decentralized EAs has been less complete. However, the analysis tools and the results presented here provide considerable insight into the properties of various decentralized selection methods, including their emergent selection pressure. Understanding this permits a more informed choice of selection methods when designing and implementing

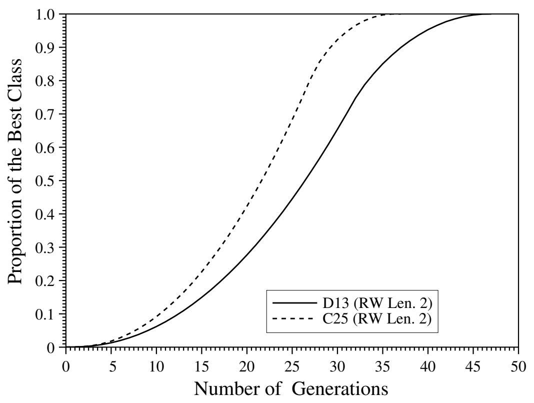

Figure 8: Emergent growth curves for path elitism using a 32×32 toroidal grid and different random walk lengths.

spatially structured EAs.

In summary, the emergent selection pressure of a local selection algorithm is qualitatively identical to its global counterpart, but quantitatively weaker because of the additional time required to propagate the best individuals throughout a two dimensional grid. For a given selection method, we can quantitatively predict the selection pressure in terms of a single parameter: the ratio of the radius of the local neighborhood to the radius of the two dimensional grid, resulting in precise characterizations of fitness proportional, ranking, tournament, and random walk selection methods.

References

Bäck, T. (1995). *Evolutionary Algorithms in Theory and Practice.* New York: Oxford University Press.

Collins, R. J. (1992). *Studies in Artificial Evolution.* Ph. D. thesis, University of California, Los Angeles.

Collins, R. J. and D. R. Jefferson (1991). Selection in massively parallel genetic algorithms. In *Proceedings of the Fourth International Conference on Genetic Algorithms*, San Diego, CA, pp. 249–256. Morgan Kaufmann.

De Jong, K. and J. Sarma (1995). On decentralizing selection algorithms. In *Proceedings of the Sixth International Conference on Genetic Algorithms*, Pittsburgh, PA, pp. 17–23. Morgan Kaufmann.

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

Gorges-Schleuter, M. (1989). ASPARGOS an asynchronous parallel genetic optimization strategy. In *Proceedings of the Third International Conference on Genetic Algorithms*, Fairfax, VA, pp. 422–427. Morgan Kaufman.

Hillis, D. W. (1991). Co-evolving parasites improve simulated evolution as an optimization procedure. In C. G. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen (Eds.), *Proceedings of Artificial Life II*, pp. 313–323. Addison-Wesley.

McInerney, J. (1992). *Biologically Influenced Algorithms and Parallelism in Non–linear Optimization.* Ph. D. thesis, University of California, San Diego.

Sarma, J. and K. De Jong (1996). An analysis of the effects of neighborhood size and shape on local selection algorithms. In *Proceedings Parallel Problem Solving from Nature - PPSN IV*, Berlin, Germany, pp. 236–244. Springer-Verlag.

Spiessens, P. and B. Manderick (1991). A massively parallel genetic algorithm: Implementation and first analysis. In *Proceedings of the Fourth International Conference on Genetic Algorithms*, San Diego, CA, pp. 279–287. Morgan Kaufmann.