Chapter 8

The Utility of Equilibrium (Applications)

In this chapter we present three applications of fitness sharing to artificial and real-world problems. The purpose of these demonstrations is to present a range of very different problem areas within which sharing can be of benefit:

- multimodal function optimization (a pipeline network design problem),
- multi objective decision making (a groundwater remediation problem), and
- layout and packing (an artificial one-dimensional layout problem).

Although the theoretical results developed in previous chapters are not ready for direct use in practical problem solving here, some earlier discoveries and insights will be made vivid here (e.g., level-three cooperation, the improvement of search by niching, etc.). The power, robustness, and flexibility of sharing are brought out by these experiments. These empirical data will serve at the very least to inspire and guide further work on the nature of niching.

8.1 Function Optimization: (Search and Multimodality)

In this section we present an application of fitness sharing to multimodal function optimization. It is well-known that this problem has many local optima, several of which are of high merit.

(New and better optima are found every few years; see below.) Fitness sharing is applied here with two goals in mind:

1. find multiple optima, and
2. improve the search for the very best optimum.

We are really demonstrating two capabilities of niching here, the ability to find and maintain several good optima, and the ability to improve our chances of finding the very best solution. While the second goal is generally assumed (i.e., we always want to find the best solution on a single-objective optimization problem), the first goal might or might not be present. In many cases, such as on the NYC tunnels problem below, we are interested in multiple optima because (a) we would like to find several alternative solutions of near or equal (high) utility, and/or (b) we are interested in seeing the first best, second best, third best, etc., for historical reasons, or perhaps to raise our confidence in the thoroughness of the GA search. Whatever the exact motivation, the following application of fitness sharing to the NYC tunnels problem demonstrates improved search *and* the ability to find and keep all previous best-known solutions (i.e., local optima), simultaneously.

8.1.1 The New York City Tunnels Problem

Our objective function is an open problem in water distribution pipeline design, known as the *New York City tunnels problem* (or simply "the NYC problem"). It was developed by Schaake and Lai (1969). Murphy, Simpson, and Dandy (1993) successfully applied a GA to the problem to find three new optima, each superior to the previous known best solution by Morgan and Goulter (1985). For details of the problem, the reader is referred to (Murphy, Simpson, & Dandy, 1993). We provide a brief description below.

The problem is to redesign a pipeline network for distribution of fresh water throughout Manhattan. The goal is to minimize the cost of installing new pipes while meeting minimum water pressure requirements at sixteen end nodes. The old network consists of twenty one large diameter underground pipes. Each of these pipes can be "duplicated", that is a new pipe can be installed parallel to the existing one, thereby increasing capacity on that "link". For each possible duplication, fifteen choices of new pipe diameter are available. These fifteen choices, combined with the sixteenth choice of not duplicating the pipe (pipe diameter 0, if you will), allows us to code each of the 21 decision variables in four bits, resulting in a total chromosome length of $\ell = 21 * 4 = 84$ bits. Thus the search space is size $2^{84} \approx 1.934 \times 10^{25}$. The NYC problem is still an open one.

The fitness function for the NYC problem is computationally costly, as it involves a simulation of the candidate network's performance under a fixed load to determine pressure at each of the nodes. Pressure head constraint violations are penalized by a high cost added to the total cost of installing the specified duplicate pipes. The larger the pipe diameters chosen for the duplicates, the greater the pressures but the more expensive the installation. The output of the fitness function is a total cost, including installation and penalties. But the final costs we report below are only the installation cost component, since that is the objective function by which we compare solutions found by different algorithms. The objective is to minimize this cost while meeting the pressure head constraints.

8.1.2 The Need for Niching

In 1985, Morgan and Goulter found a new optimum costing $\approx$ \$39.20 million. Murphy, Simpson, and Dandy (1993) used a GA with exponential scaling of the fitness function to find three new optima, worth $\approx \$38.80, \$39.06$, and \$39.17 million, (call them optima 1, 2, and 3, respectively), all superior to the 1985 solution. A fourth optimum, at $\approx$ \$39.22 million (optimum 4) was also found[1], and a fifth optimum of interest was found at \$39.28 million (optimum 5). However, the authors note that the GA *usually* converged to optima 2 and 3, the second and third known best, which are very similar (both genotypically and phenotypically), being slight variations on the same basic network design. Optima 1, 4, and 5, however, were found less often. These optima are very similar to each other, and very different from optima 2 and 3.

8.1.3 Niching Results

A niched GA with sharing, however, proves to be much more likely to locate all five top optima in a single run. Figure 8.1 illustrates a comparative run between a GA with sharing and one without. For both cases, the GA parameters other than sharing are the same: a simple, generational GA with population size $N = 6000$, probability of single point crossover $p_c = 0.8$,

[1]Although optimum 4 is slightly more expensive than Morgan and Goulter's solution, the latter was shown by Murphy, Simpson, and Dandy (1993) to be "barely" infeasible.

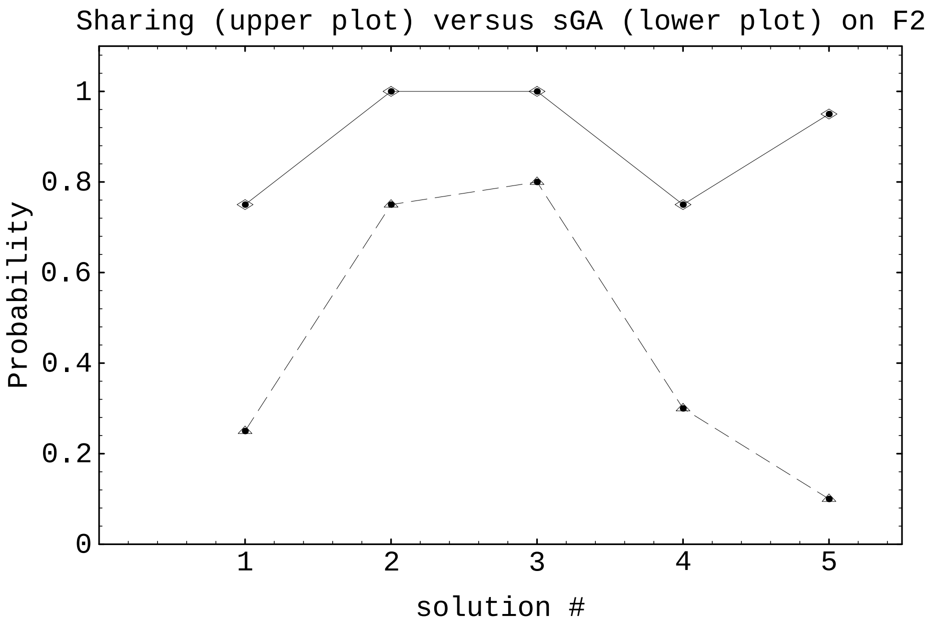

Figure 8.1: Holding all other GA parameters constant (e.g., N, β_{sh}, p_c, etc.), adding sharing increases the likelihood of locating any of the top five known solutions of problem F2. *Probability* is the percentage of 20 different trials in which a particular solution appeared at least once within the first 200 generations.

no mutation ($p_m = 0$), and the fitness function is raised to the power $\beta_{sh} = 20$ (prior to sharing). The GA with sharing uses genotypic comparisons with $\sigma_{sh} = 7$ bits. We run the GA with and without sharing for twenty different trials (i.e., different random initial populations), each time terminating the run at 200 generations. We then examine every individual created during a trial and record the number of trials (out of 20) in which a particular optimum (1-5) is found (i.e., appears in the population at any of the 200 generations). Calling this frequency of appearance "probability", we plot the sampled probabilities of finding the top five known optima in Figure 8.1. Thus each plotted point (k, p) is the percentage p of the twenty trials in which optimum k is found.

The upper curve in Figure 8.1 plots the performance of sharing, while the lower curve plots the performance of the same GA with sharing turned off. It is clear that sharing exhibits a much higher probability of success in locating *any* of the five optima, but especially the more "difficult" optima (1, 4 and 5).