Genetic Algorithm Difficulty and the Modality of Fitness Landscapes

Jeffrey Horn and David E. Goldberg
Illinois Genetic Algorithms Laboratory
University of Illinois at Urbana-Champaign
Urbana, IL 61801-2996

IlliGAL Report No. 94006
August 1994

To appear in the proceedings of the
Foundations of Genetic Algorithms (FOGA) 3 workshop
held July 30 – August 2, 1994 in Estes Park, Colorado, USA

Illinois Genetic Algorithms Laboratory (IlliGAL)
Department of General Engineering
University of Illinois at Urbana-Champaign
117 Transportation Building
104 South Mathews Avenue
Urbana, IL 61801-2996

Genetic Algorithm Difficulty and the Modality of Fitness Landscapes

Jeffrey Horn
Department of Computer Science
University of Illinois at Urbana-Champaign
117 Transportation Building
104 South Mathews Avenue
Urbana, IL 61801-2996
Phone: 217/333-2346
jeffhorn@uiuc.edu

David E. Goldberg
Department of General Engineering
University of Illinois at Urbana-Champaign
117 Transportation Building
104 South Mathews Avenue
Urbana, IL 61801-2996
Phone: 217/333-0897
deg@uiuc.edu

Abstract

We assume that the modality (i.e., number of local optima) of a fitness landscape is related to the difficulty of finding the best point on that landscape by evolutionary computation (e.g., hillclimbers and genetic algorithms (GAs)). We first examine the limits of modality by constructing a unimodal function and a maximally multimodal function. At such extremes our intuition breaks down. A fitness landscape consisting entirely of a single hill leading to the global optimum proves to be hard for hillclimbers but apparently easy for GAs. A provably maximally multimodal function, in which half the points in the search space are local optima, can be easy for both hillclimbers and GAs. Exploring the more realistic intermediate range between the extremes of modality, we construct local optima with varying degrees of "attraction" to our evolutionary algorithms. Most work on optima and their basins of attraction has focused on hills and hillclimbers, while some research has explored attraction for the GA's crossover operator. We extend the latter results by defining and implementing *maximal* partial deception in problems with k arbitrarily placed global optima. This allows us to create functions with multiple local optima attractive to crossover. The resulting maximally deceptive function has several local optima, in addition to the global optima, each with various size basins of attraction for hillclimbers as well as attraction for GA crossover. This *minimum distance* function seems to be a powerful new tool for generalizing deception and relating hillclimbers (and Hamming space) to GAs and crossover.

1 Introduction

Genetic algorithms (GAs) are robust adaptive systems that have been applied successfully to hard optimization problems, both artificial and real world. Yet GAs do fail. When and why? The question of what makes a problem hard for a GA has received a good deal of attention – and some controversy – as of late. The controversy is largely a tempest in a teapot. If we are ever to understand how hard a problem GAs can solve, how quickly, and with what reliability, we must get our hands around what "hard" is. Goldberg (1993a) suggests several quasi-separable dimensions of GA problem difficulty:

- Isolation
- Misleadingness
- Noise
- Multimodality
- Crosstalk

Important progress has been made (Goldberg, 1993b) in understanding the role of deception (the combination of solution isolation and suboptimum misleadingness), noise (both deterministic and stochastic), multimodality, and crosstalk, but a significant amount of work remains.

Some have suggested that we first create a full-fledged *description* of problem difficulty, with all the notation necessary to be complete and exact, but complex systems understanding is not achieved via this route (Goldberg, 1993a, 1993b). The more usual method is to design or *prescribe* problems that maximally but boundedly challenge a GA along one or more dimensions of problem difficulty. The work presented here continues largely in that vein by investigating deception and modality jointly. We believe that work like this ultimately will be assembled to form a rough patchquilt of models and metrics that help us quantify and analyze how hard a problem is for a particular GA. Along the way – as we have done all along – we will gather and use the knowledge of problem difficulty thus obtained.

2 Preliminaries: Landscapes and Optima

We make use of the paradigm of a *fitness landscape* (Wright, 1988), which we define here as a search space S, a metric, and scalar fitness function defined over elements of S. Assuming the goal is to maximize fitness, we can imagine the globally best solutions (the global optima, or "globals") as "peaks" in the search space. For the purposes of this paper, we define local optimality as follows. We first assume a non-negative-real valued, scalar fitness function $f(s)$ over fixed length ℓ-bit binary strings s, $f(s) \in \Re \geq 0$. Without loss of generality, we assume f is to be maximized. A *local optimum* in a discrete search space S is a point, or region, with fitness function value strictly greater than those of *all* of its nearest neighbors. By "region" we mean a set of interconnected points of equal fitness. That is, we treat as a single optimum the set of points related by the transitive closure of the nearest-neighbor relation such that all points are of equal fitness[1]. This definition allows us to include flat plateaus and ridges as single optima, and to treat a flat fitness function as having no local optima. The "nearest neighbor" relation assumes a metric on the search space, call it d, where $d(s_1, s_2) \in \Re \geq 0$ is the metric's distance between points s_1 and s_2. Then the nearest neighbors of a point s' are all points $s \in S, s \neq s'$ such that $d(s', s) \leq k$, for some neighborhood radius k. In this paper we use only *Hamming distance* (number of bit positions at which two binary strings differ) as the metric[2], and assume $k = 1$. Thus a point s with fitness $f(s)$ greater than that of all its immediate neighbors (strings differing from s in only one bit position) is a local optimum.

3 Minimum Modality Can Be Hard

At one extreme of the modality spectrum we have unimodality. Unimodal functions have only one local optimum, which is therefore the global optimum. It is well-known that such problems can be hard when the optimum is isolated, with little or no information available elsewhere in the search space. Such isolated peaks on otherwise flat fitness landscapes have been called *needle-in-a-haystack* (NIAH) problems (Goldberg, 1989a), and are clearly solvable only by enumeration of the space. But if we decrease the isolation of the single optimum, increase the size of its basin of attraction, and add information to larger portions

[1]The requirement for equality could be relaxed to allow for fitnesses within some δ of the globally optimal fitness.

[2]A recurring theme of this paper is the demonstration that Hamming distance among optima, and the shape of the Hamming space landscape in general, are quite important to GA search by crossover, contrary to recent discussions in the GA community.

of the search space, intuition tells us that the search should become easier and shorter than enumeration. In particular, if we make the entire search space a single hill, where every point is on a path to the global optimum, even simple evolutionary algorithms, such as a 1-bit hillclimber, should quickly optimize the function. But we showed otherwise in (Horn, Goldberg, and Deb, 1994).

In (Horn, Goldberg, and Deb, 1994), we constructed two different functions of ℓ-bit strings in which all points are on the path to the only optimum, but in which path length grows exponentially in ℓ. Thus even a 1-bit hillclimber is guaranteed not only to find the global optimum but to make constant progress toward it. However, for reasonably large ℓ (> 90, for example), the hillclimber, and many other local searchers, effectively will never converge to the global. Below we summarize one of the two constructions from the 1994 paper, namely the Root2path.

We choose a stepsize $k = 1$ to illustrate the construction of the Root2path. Each point on the path must be exactly $k = 1$ bit different from the point behind it and the point ahead of it on the path, while also being at least $k + 1 = 2$ bits away from any other point on the path.

The construction of the path is intuitive. If we have a Root2path P_ℓ of dimension ℓ (= number of bits), we can basically double it by moving up two dimensions to $(\ell + 2)$ as follows. Let P_ℓ be a list of binary strings representing consecutive steps on the path (e.g., $P_4 = \{0000, 0001, 0011, ...\}$). Make two copies of P_ℓ, say copy00 and copy11. Add "00" to the beginning of each string (point) in copy00, and add "11" to each point in copy11[3]. Now each point in copy00 is at least two bits different from all points in copy11. Also, copy00 and copy11 are both paths of stepsize one and of dimension $\ell + 2$. Furthermore, the endpoint of copy00 and the endpoint of copy11 differ only in their first two bit positions ("00" versus "11"). By adding a "bridge" point that is the same as the endpoint of copy00 but with "01" in the first two bit positions, we can connect the end of copy00 and the end of copy11[4]. Reversing the list in copy11, we concatenate copy00, the bridge point, and Reverse[copy11] to create the Root2path $P_{\ell+2}$ of dimension $\ell + 2$, with length essentially twice that of P_ℓ:

$$|P_{\ell+2}| = 2\,|P_\ell| + 1 \tag{1}$$

For the dimensions in between the doublings, $P_{\ell+1}$, we can simply use the path P_ℓ by adding a "0" to each point in P_ℓ. If $|P_\ell|$ is exponentional in ℓ, then $|P_{\ell+1}|$ is exponential in $\ell + 1$.

For the base case, we use the first dimension $\ell = 1$. Here we have only two points in the search space: 0 and 1. We put them both on the Root2path for $\ell = 1$. Thus, $P_1 = \{0, 1\}$, where 0 is the beginning and 1 is the end (i.e., the global optimum).

So with every other incremental increase in dimension ℓ, we have an effective doubling of the path length. We can solve the recurrence relation in Equation 1 exactly, but it is clear that the path length increases in proportion to $2^{\ell/2}$ or $(\sqrt{2})^\ell$. Thus the path length grows exponentially in ℓ with base ≈ 1.414, although it is an ever-decreasing fraction of the space[5].

We make one end of the path the global optimum, and adjust the fitnesses of the rest of the path so that fitness decreases as we move along the path away from the global optimum and towards the beginning of the path. If we choose the all zeroes point to be the beginning of the path (rather than the end), we can easily make the rest of the search space (all points not on the path) lead to the beginning of the path. We do this by setting the fitness of offpath points to some increasing function of the number of zeros (*nilness*) in the string, such as $f(s) = f(0000...0) - u(s)$. Here $f(0000...0)$ is the fitness at the beginning of the path[6] and $u(s)$ is the *unitation* (number of ones) of string s.

In (Horn, Goldberg, and Deb, 1994) we also showed how to extend this construction to create paths of stepsize $k > 1$, where nonconsecutive points on the path are separated by at least $k + 1$ bits. Paths

[3] Thus the point "1001" in P_ℓ would become the point "001001" in copy00 and "111001" in copy11.

[4] Alternatively, we can connect the beginning of copy00 to the beginning of copy11.

[5] Thus for large ℓ, the path itself is essentially a NIAH, with a vanishing probability of a random starting point "landing" anywhere on or near the path.

[6] In (Horn, Goldberg, & Deb, 1994), we call the offpath points with their nilness function the *nilness slope*. Together the *path* plus the *slope* form the single *hill* that fills the entire search space.

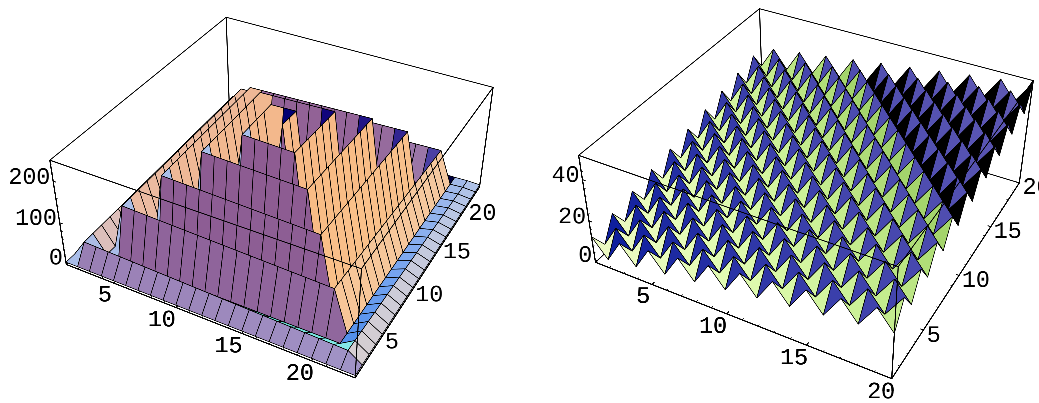

Figure 1: The two extremes of modality. *Left:* A unimodal problem, $f_{lp}(x,y)$, in which every point in the space is on an exponentially long path to the global optimum. *Right:* A maximally multimodal function, $f_{mm,easy}(x,y)$ that is easy for a GA to optimize.

constructed this way are of order $O(2^{\ell/(k+1)})$ steps in length, which is exponential in ℓ for fixed k. We then showed empirically that various types of one-bit hillclimbers do indeed take exponential time, in expectation, to climb such paths[7].

We try to visualize a long path in two dimensions in Figure 1, left. Here the fitness is a function of the integers x and y: $f_{lp}(x,y)$, for the "long path" function. As with the Root2path and Fibonacci paths constructed in (Horn, Goldberg, and Deb, 1994), the two dimensional spiral path is generated by induction on the size of the search space, in this case s^2. Here s is the integer range over which x and y each vary. For the base case $s = 1$, the single point is the global optimum, with fitness 1, and is the only point on the path. Informally, the inductive step is to take a path of dimension s, add a ring (or rather, square) of points around the outside, each with fitness 0, then add another ring of points and add them to the path, incrementing the fitness of every point on the s-dimensional path by the number of new path points added. This gives us a long path of dimension $s+2$. In other words, every other increment of s adds another ring to the spiral and pushes up the old, inner spiral to maintain the slope up to the global at the center. Thus the construction of the two dimensional spiral path has the same inductive form as those of binary-space long paths. However, in two dimensions we can actually put half the search space on the spiraling path to achieve path lengths of $O(|\text{search space}|)$. Again, we can separate non-adjacent points on the two dimensional spiral path by any number k of steps[8] by simply adding rings to the spiral every k increments of s instead of every other increment.

The long path problem is clearly and demonstrably difficult for local searchers (that is, algorithms

[7] The problem of finding a maximally long path with minimal separation has some history, and is known as the "snake-in-the-box" problem, or the design of distance preserving codes, in the literature on combinatorics (Preparata, 1974). Maximizing the length of paths with k-bits of separation is an open problem, even for $k = 1$. However, upper bounds have been found that are $< O(2^\ell)$. Thus the longest paths we can ever find will be $O(2^{\ell/c})$ for some constant $c \geq 1$. For the Root2path, $c = 2$ for $k = 1$. In (Horn, Goldberg, and Deb, 1994) we suggested the Fibonacci path, which yields a $c = 1/Log_2\Phi \approx 1.44042$ when $k = 1$, where Φ is the golden ratio.

[8] We assume the *metropolitan* or *city-block* metric for this 2-D space.

that search small neighborhoods with high probability, and larger neighborhoods with vanishingly small probability). Such algorithms include hillclimbers and the GA's mutation operator. Apparently, however, such long path problems are amenable to GA crossover. It is obvious from their inductive construction that these paths have structure. The same basic subsequences of steps are used over and over again on larger scales, resulting in fractal self-similarity. Such structure might induce building blocks exploitable by crossover (Holland, 1992). In particular, patterns such as "00" and "11" are common to all points on the (binary-space) path, and are used over and over again in the construction of the path. Although we have not yet performed a schema analysis (Bethke, 1981) of these functions[9], we have some preliminary empirical results (Horn, Goldberg, & Deb, 1994). These early results indicate that a GA with crossover alone ($p_m = 0$) outperforms $k = 1$ step hillclimbers (e.g., steepest ascent, next ascent) and several mutation alogithms (e.g., random mutation hillclimbing (RMHC)) by reaching the global optimum using several orders of magnitude fewer function evaluations. More testing and analysis is required, but if the GA is superior to hillclimbing (i.e., finds the global in $< O(2^{\ell/ck})$ time, then we have found a problem that distinguishes GAs from hillclimbers. Such a result might be of particular interest to those looking at when GAs outperform hillclimbers (Mitchell and Holland, 1993). One answer might be "on a hill".

The long path function is interesting also because it points out that the modality of a search space, if measured solely as the number of local optima, is at best a first order estimate of the amenability of the space to search by hillclimbers, mutation, or by evolutionary search algorithms in general.

4 Maximum Modality Can Be Easy

At the other extreme of the modality spectrum, what is the maximum number of local optima possible in a binary problem of size ℓ-bits? We assume minimum radius (1-bit) peaks, so that our requirement for local optimality is minimal. That is, a point is a local optimum if and only if all adjacent points have inferior fitness values. We can calculate an upper bound on the number of such optima.

Let p be the number of local "peaks". In an ℓ-bit problem, each peak must have exactly ℓ immediate neighbors that are not local optima (call them "nonoptima"). Thus the number of adjacent pairs of optimum-nonoptimum is $p * \ell$. That is, for a problem to have p local optima it must have $p * \ell$ different pairings of optima and adjacent nonoptima. There are exactly $2^\ell - p$ nonoptima. Each of these nonoptima can have at most ℓ adjacent optima. Thus an *upper* bound on the number of optimum-nonoptimum pairings is $(2^\ell - p) * \ell$. We can increase p until the number of optimum-nonoptimum pairs equals the maximum: $p * \ell = (2^\ell - p) * \ell$. We derive $p = 2^{\ell-1}$, which is half the size of the search space[10]. This is an upper bound on the number of local optima.

Using the concept of unitation, we can construct a function with $2^{\ell-1}$ local optima, thus showing that our upper bound is indeed the exact maximum number of local optima. A maximally multimodal function $f_{mm}(s)$ of unitation assigns a high fitness to all bit strings s of odd unitation, for example, and a low fitness to all strings of even unitation:

$$f_{mm}(s) = \begin{cases} 1 & \text{if Odd}(u(s)) \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Since all strings of odd unitation are separated from each other by strings of even unitation, odd unitation strings are indeed local optima. Strings of odd unitation occupy exactly half the search space[11].

[9] We assume familiarity with fundamental concepts of GA theory, such as building blocks and schema analysis, at least to the extent presented in (Goldberg, 1989a).

[10] In other words, the number of optima cannot exceed the number of nonoptima.

[11] It is interesting to note that maximally multimodal functions, in which half the search space are local optima, do not have to be functions of unitation but must have all of their local optima be of the same parity of unitation (i.e., either odd or even). This result can be shown by induction: Choose a local optimum. It has either odd or even unitation. The nearest neighbors are not local optima and are all of the opposite parity of unitation from the chosen optimum. Their nearest neighbors in turn must be local optima (since the function maximizes the number of optimum-nonoptimum pairings) and must be of the same parity of unitation as the first chosen optimum, and so on.

A maximally multimodal function has the maximum number of possible "attractors" on which a search algorithm might get stuck. But of course each optimum's basin of attraction is at a minimum size. To illustrate that massive multimodality *by itself* does not imply difficulty for GAs or hillclimbers in general, we can add a gentle slope to the function that leads quickly to the global optimum (all ones in this case)[12].

$$f_{mm,easy}(s) = u(s) + 2 * f_{mm}(s) \tag{3}$$

The function $f_{mm,easy}(s)$ resembles a one-max function "with bumps". A hillclimber with a non-zero probability p_{up2} of taking a step of size two or more bits uphill will climb to the global optimum in at most ℓ/p_{up2} steps[13]. For any p_{up2} that decreases no faster than linearly in ℓ, and in particular for a constant p_{up2}, the hillclimber will climb the hill in expected $O(\ell)$ steps[14].

A GA is also unlikely to have difficulty in quickly optimizing this function[15]. In addition to being amenable to search by the GA's mutation operator, the function appears easy for crossover when we apply a static analysis of schema partitions. In every schema partition, the schema containing the global optimum (the all ones schema) must have a higher average fitness than all other schemata competing in that partition. To see why the all ones schema always wins[16], we first calculate the schema average fitness for any schema as a function of the schema's unitation (number of ones in the defined bits). In a partition of order o, the fitness of a schema $\hat{s}$ is equal to the unitation of the schema ($= u(\hat{s})$) plus the average unitation of the $(\ell - o)$ undefined bit positions ($= (\ell - o)/2$), plus twice the average contribution of $f_{mm}(s)$, the parity of unitation function, which will be $2 * 1/2$:

$$\bar{f}_{mm,easy}(\hat{s}) = u(\hat{s}) + (\ell - o)/2 + 1 \tag{4}$$

So the schemata with highest unitation will always be the winners of their partition competitions.

Figure 1, right, is a visualization of a maximally multimodal function in two dimensions. Here the decision variables are the positive integers x and y, and the fitness function is

$$f_{mm,easy}(x, y) = (x + y) + \begin{cases} 10 & \text{if Even}(x + y) \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

Just as in the case of binary strings, our 2-D formulation makes half the search space local optima while providing a constant gradient pointed straight at the global (for any $\geq$ 2-bit hillclimber) as well as plenty of schema information for crossover (at least for binary coded integers).

The maximally multimodal function is interesting only because it points out that the modality of a search space, if measured solely as the number of local optima, is at best a first order estimate of GA difficulty (or difficulty for evolutionary search algorithms in general).

---

[12] Here we assume that ℓ is odd. For even ℓ, we should change $f_{mm}(s)$ to favor even unitation, thus ensuring that the all ones string is indeed a local as well as a global optimum.

[13] This is an upper bound on expectation: $E[t] \leq \ell/p_{up2}$, where t is the number of steps taken to reach the global. Note that these calculations assume nothing about the average number of function evaluations (or time involved) in taking a single step.

[14] A hillclimber with a significant p_{up2} is also expected to climb the Root2path in linear or constant time. However, Horn, Goldberg, and Deb (1994) also showed how to construct the *CubeRoot2path*, a long path for hillclimbers of step size 2. A hillclimber with significant p_{up2} but near zero p_{up3} (probability of taking a step size of three or more bits) will take an expected time exponential in ℓ to climb such a path. In other words, the intractability of the long path problem class, and the ease of the maximally multimodal (easy) problem, (for hillclimbers of stepsize k) both scale with increasing k.

[15] Although we have not yet performed the runs, we believe that the GA will perform better on the maximally multimodal (easy) function than on a long path problem of the same size ℓ. (By "better", we mean that the GA will require a smaller population size to find the global, on average.) This would be a somewhat counterintuitive result, since one might expect a single large hill to be more easily/quickly optimized (by a hillclimber or GA) than a "field" of $2^{\ell-1}$ "molehills".

[16] The function $f_{mm,easy}(s)$ also can be analyzed easily by Walsh analysis (Bethke, 1981; Goldberg, 1989b), since it can be written as a summation of just a few Walsh functions. Using the notation of Goldberg (1989b), we note that for $f_{mm,easy}(s)$ all the Walsh coefficients w_i of all orders i are zero, except for $w_0 = \lfloor \ell/2 \rfloor + 1$, $w_1' = -1$, and $w_{2^\ell} = -2$, where w_1' stands for all of the order-1 Walsh coefficients, which are identically valued in this case.

5 Intermediate Modality: Local Optima and Basins of Attraction

The results of the previous two sections remind us that the fitness landscapes of interest to us (i.e., those that challenge the GA in a realistic and general manner) have intermediate modality. But it is not clear how to add modality to the long path problem, or to reduce the modality of the maximally multimodal function. For example, we might be tempted to generalize the work on maximum modality by maximizing the number of optima that are k or more bits apart (i.e., each optimum has a local neighborhood of radius $\geq (k-1)$ bits in which it is optimal). Unfortunately, calculating the maximum number of such optima for arbitrary k is an open problem in coding theory known as *sphere packing* (Harary, Hayes, and Wu, 1988; MacWilliams and Sloane, 1977). We can calculate an upper bound by simply dividing the search space size, 2^ℓ, by the hypervolume of the *nonoverlapping* local neighborhoods of the optima, $\sum_{r=0}^{\lceil k/2 \rceil - 1} \binom{\ell}{r}$. But this results in a very loose upper bound, as evidenced by the simple case of $k = 2$ bit separation: the upper bound yields 2^ℓ while we know that the actual maximum is half that[17].

5.1 Background

Although *maximizing* the number of local optima with neighborhoods of radius k is an open problem, work has proceeded along the lines of measuring and controlling the number of optima and their basins of attraction (or the "attractiveness" of the optima). Such work can be divided into two types: that which assumes hillclimbing attraction and that which assumes GA crossover attraction. That is, most papers analyzing local optima assume one type of algorithm or the other. We give some background on both approaches, but focus on GA crossover.

5.1.1 Hillclimbing attraction

A number of recent papers define peaks (i.e., local optima together with their basins of attraction) in terms of hillclimbing. Goldberg (1991) formally defines basins of attraction to a point x^* as all points x such that a given hillclimber has a non-zero probability of climbing to within some ϵ of x^* if started within some δ-neighborhood of x. Jones and Rawlins (1993) introduce *reverse hillclimbing* and *probabilistic ascent* as techniques for defining the basin of attraction of a particular local optimum in a fitness landscape for hillclimbers with known probabilities of ascent. In the same volume, Mahfoud (1993) analyzes the performance of multimodal GAs on multi-niche problems, where each niche is a local optimum with a basin of attraction defined by the probability of a hillclimber reaching the local optimum from a point in the basin.

5.1.2 GA attraction

The literature on the attraction of peaks and regions in the landscape to the GA's crossover operator is largely based on Holland's *schema theorem* (Holland, 1992) and schema average fitness calculations (Bethke, 1981). Goldberg (1987, 1989a, 1989b, 1989c) and later others (Whitley, 1991; Homaifar, Qi, and Frost, 1991; Deb, Horn, and Goldberg, 1993) defined and constructed deceptive landscapes, in which the GA should be attracted to suboptimal local optima and led away from the global optimum. Schema analysis has also been used to construct "GA-easy" functions (Wilson, 1991) which have large basins of attraction for the global optimum [18]. More recently, Mitchell and Holland (1993) have begun to weaken the GA-easy conditions by limiting the number and order of schema partitions leading toward the global[19].

[17] It is interesting to note that sphere packing with radius $\lceil k/2 \rceil - 1$ also provides an upper bound (within a factor of two) on the path length of $(k-1)$-step long path problems, since nonconsecutive steps on the path must be at least k bits apart. However, tighter upper bounds on $(k-1)$-step paths have been found (Preparata, 1974).

[18] Like our maximally multimodal function, previous GA-easy functions contain local optima that can make the problem difficult for hillclimbing.

[19] It is interesting to note that according to our definition of local optimum, the Royal Road functions are unimodal.

We look to deception for guidelines on how to lead or mislead a GA both toward and away from multiple optima. Since a function with more than two local optima implies partial (and not full) deception[20], we first review and define partial deception.

5.1.3 Partial deception defined

One must be careful in defining partial deception. It is easy to weaken the requirements for full deception such that a GA can easily find the global optimum of some functions meeting those requirements. We briefly present the definitions we use in this paper.

A *deceptive attractor* (Whitley, 1991) is the suboptimal point D toward which a GA is (mis)led. In a particular schema partition, the schema containing the deceptive attractor is called the *deceptive schema*, while the schemata containing the global optima are called the *global schemata*. In general, a *deceptive partition* is a partition in which the deceptive schema has a high (schema average) fitness and/or all of the global schemata have low fitness. The rather vague concept of deceptive schemata "beating" global schemata, in terms of schema average fitness, has been interpreted in several different ways by different researchers. Here we list three specific definitions of interest to us (the order of labeling has only chronological meaning):

- **Type I deceptive** – global schemata lose to all other schemata (Bethke, 1981).
- **Type II deceptive** – deceptive schema wins over all other schemata (Goldberg, 1987).
- **Type III deceptive** – deceptive schema has higher fitness than all global schemata (Grefenstette, 1992).

It is not yet clear which, if any, of these definitions implies more difficulty for the GA on partially deceptive landscapes[21]. For fully deceptive functions, all partitions of order $< \ell$ are type II and III deceptive[22]. But in general (i.e., partial deception), deceptive and global schemata can be placed anywhere in the fitness ordering of a partition's schemata.

In the design of maximally misleading functions, our goal is to choose D and define the landscape such that the GA converges to D with high probability. We assume that we are given one or more globals (that is, their locations and perhaps their fitness value). In the case of a single global optimum g, the above definitions are sufficient to unambiguously identify a unique D, which is the complement of g (Whitley, 1991). We can then construct *fully deceptive* functions in which the schema containing D is the winner of *every* partition (i.e., types II and III deception) at every order up to the string length ℓ (Goldberg, 1989a, 1989b, 1990). Full deception is clearly maximally misleading to a GA. It is also clearly bimodal, with local optima[23] at D and at g. To add more optima, and basins of GA attraction, we need to define partial deception.

Deception becomes a more practical tool of GA theory when it is embedded in fitness landscapes as *partial deception*. But at the moment, we can only define partial deception vaguely as less misleading than full deception and more misleading than GA-easy. Trying to order partially deceptive functions according to some scalar measure of deception is problematic. Goldberg recognized the need to generalize full deception and did so by defining *order k full deception* (Goldberg, 1991). Homaifar, Qi, and Frost (1991) call such limited deception *reduced order k deception*. Although it has gone by other names, this kind of partial deception is most widely known as *bounded deception*. A problem has bounded deception

[20] A function with more than one global optimum cannot be misleading in all ℓ of its order one partitions. Two (or more) distinct global optima must differ in at least one bit position. The order one schema partitions defined at each of those partitions must "point to" at least one of the globals.

[21] We do note, however, that types I and II deception conditions imply type III.

[22] Such partitions can also be considered type I deceptive (at least for some of the fully deceptive functions constructed in the literature) if we discount the global optimum from schema average fitness calculations, a modification suggested and justified later in this paper.

[23] Whitley (1991) proves that D need not be a local optimum. But if it is not, then one of its immediate neighbors must be.

of order k if and only if *all* partitions of order $(k-1)$ or less are deceptive [24]. Partitions of order $\geq k$ might or might not be deceptive. Thus for $k = \ell$, bounded deception is full deception. Goldberg, Deb, and Korb (1991) construct examples of boundedly deceptive problems by concatenating some number m of fully deceptive subfunctions of length ℓ_s to get an $(m * \ell_s)$-bit function of order ℓ_s bounded deception. Such functions have 2^m local optima, one of which is global.

The order k of bounded deception establishes a partial order over functions. We can safely say that a function of order k bounded deception is at least as difficult for a GA as a function of order $k' < k$ bounded deception, all other problem dimensions (e.g., noise, crosstalk) being roughly equal.

Other attempts to define, quantify, or order the degree of partial deception have resulted in unnecessarily weakening the requirements for deception. For example, Grefenstette (1992) described an $\ell = 20$-bit function in which all partitions defined over the first ten bits lead toward the deceptive attractor (i.e., type II deception) and all partitions defined over the last ten bits lead toward the global optimum. "Despite this high level of Deception," this function was easily optimized by a GA with population size 200. However, as Goldberg pointed out with his analysis of a similarly constructed partially deceptive function (Goldberg, 1991), we do not want to call such problems highly deceptive. Any function with low order partitions that lead toward the global optimum (i.e., any function with building blocks) is amenable to GA search. Adding additional misleading bits does not necessarily make the problem any more deceptive, let alone "arbitrarily more Deceptive..." (Grefenstette, 1992), because the number of low order building blocks remains undiminished. A function of low order k bounded deception is probably more difficult for a GA than is Grefenstette's function with some arbitrarily large number of misleading bits, even though the boundedly deceptive problem will have fewer deceptive partitions overall[25].

The above example points out the importance of both number and *order* of deceptive partitions. As discussed above, order k bounded deception only allows comparisons between functions with the same number of deceptive partitions, $\binom{\ell}{o}$, at each order $o < \min(k_1, k_2)$, where k_i are the orders of bounded deception for two functions. Goldberg, Deb, and Horn found another way to construct and order partially deceptive functions without losing essential misleadingness (Goldberg, Deb, & Horn, 1992; Deb, Horn, & Goldberg, 1993). They constrained their "bipolar deceptive function" to be a function of *folded unitation* $u_{fld}(s)$, which is simply the number of ones minus the number of zeros: $u_{fld}(s) = u(s) - (\ell - u(s)) = 2\,u(s) - \ell$. This leads to a symmetric function of unitation, $f_{bip}(u_{fld}(s))$. By enforcing full deception in the composite function $f_{bip} \circ u_{fld}$, they ended up with two globals (all ones and all zeros) in the unfolded unitation space, and $\binom{\ell}{\ell/2}$ deceptive attractors (all points consisting of half ones and half zeros, for even ℓ). They also showed that the full deception enforced on the folded unitation function led to partitions in the unfolded function in which the schemata containing the *most* deceptive attractors won (type II deception). This occurred in all partitions where the schemata containing the globals (the global schemata) could be distinguished from the schemata containing the deceptive attractors. Thus, at order one, where the two schemata ...#1#... and ...#0#... compete, there was no possible distinction between the globals and the deceptive optima, and thus no preference between the schemata (i.e., they had equal fitness). But at order two, and above, the deceptive attractors could be distinguished, hence ...#01#... and ...#10#... beat ...#11#... and ...#00#... in all order two partitions.

Note that having two distinct globals precludes full deception and therefore requires some kind of partial deception at best (or worst!). By constraining the function to be of folded unitation, Goldberg, Deb, and Horn were able to make the overall function maximally deceptive by enforcing full deception in the folded function. They were then able to examine its implications in the unfolded space. The result was a function that had deceptive attractors located maximally far from *both* globals, and in which the deceptive schemata won in all partitions (type II), up to order ℓ, in which it was possible to distinguish globals from deceptives.

[24] Deceptive partition here means that the winner of a partition is the schema containing D, (i.e., type II deception).

[25] It is interesting to note that Grefenstette's (1992) function is not only unimodal by our definitions, but is also, like the long path problem, a single hill to the global optimum. All points in the search space lie on stepsize-1 paths to the global.

5.2 Maximal bi-global deception

We first generalize the work of Goldberg, Deb, and Horn on bipolar deceptive functions to the case of bi-global deceptive functions. We do not assume a function of unitation, folded or not, nor do we assume bipolarity (i.e., where the two globals are full complements of each other). We only assume two globals arbitrarily placed in the space. Rather than first choosing the deceptive attractor, as is usually done, we will instead try to *maximize deception* and see what deceptive attractor emerges. We use such terms as "maximal deception" loosely at first, and define them rigorously later.

Let the two arbitrarily chosen globals in an ℓ-bit problem be $G = \{g_1, g_2\}$, and let the distance between them be d_G bits. How do we choose a deceptive attractor D that maximizes deceptive partitions? An upper bound on the number of possibly deceptive partitions is the number of partitions in which we can *distinguish* D from the G. We shall call these simply *resolvable partitions*. We now try to place D so as to maximize the number of resolvable partitions at every order o.

We first note that there are $(\ell - d_G)$ bits in which the two globals agree. Intuitively, our deceptive attractor should disagree with both globals in these bit positions to maximize the number of resolvable partitions. For each of the $(\ell - d_G)$ bits of agreement between the globals, setting the corresponding bit position in D to be the complement of the globals' bit setting gives one more bit position in which deceptive schemata can be distinguished from global schemata. Thus giving D the complement of the $(\ell - d_G)$ global bit settings only increases the number of resolvable partitions at every order. We therefore assume such complementary bit settings for the $(\ell - d_G)$ bit positions of D (in which $g_1$and g_2 agree) and next consider only how to set the d_G bit positions in which the globals disagree.

Let d_1^D be the Hamming distance from D to g_1. Then $d_G - d_1^D$ is the distance from D to g_2. At any order o schema partition there are $\binom{d_G}{o}$ partitions[26]. For exactly $\binom{d_1^D}{o}$ of these partitions, we cannot distinguish D from g_2, since all o defined bits are chosen from the d_1^D bits that distinguish D from g_1 and hence are positions at which D and g_2 are in agreement. Similarly, there are exactly $\binom{d_G - d_1^D}{o}$ partitions in which we cannot distinguish D from g_1. So the total number of resolvable partitions, $N_{rp}(o)$ at order o is

$$N_{rp}(o) = \binom{d_G}{o} - \left[\binom{d_1^D}{o} + \binom{d_G - d_1^D}{o} \right]. \tag{6}$$

To maximize N_{rp} we must minimize the sum of the two binomial coefficients in the square brackets above. Since $\binom{d_1^D}{o}$ is a strictly increasing function of d_1^D, and $\binom{d_G - d_1^D}{o}$ is strictly decreasing in d_1^D, the minimum of their sum will occur when $d_1^D = d_G - d_1^D \Rightarrow d_1^D = d_G/2$, for even d_G, and $d_1^D = \lfloor d_G/2 \rfloor$ or $\lceil d_G/2 \rceil$, for odd d_G. That is, we will have the maximum possible number of resolvable partitions when we our deceptive attractor lies equally distant from g_1 and g_2, which means it is *maximally minimally distant* from G. Here we define *minimal distance* from a point D to a set G of k points as the minimum of the distances from D to each point $g_i \in G$, $1 \leq i \leq k$. The *maximally minimally distant point* from a set G is simply the point[27] in the space with the greatest minimal distance to G.

Note that we have maximized the resolvable partitions in the sense that we have the maximum number of resolvable partitions at every order. Assume that we can make all resolvable partitions deceptive, in at least one of the three senses defined earlier (we show that we can in the next section). We conjecture that a function f_1 with more deceptive partitions at every order $o < \ell$ than another function f_2 is more misleading to a GA, all other problem dimensions being roughly equal. Thus we are suggesting another

[26] We assume throughout this paper that $\binom{n}{m} = 0$ for $m > n$.

[27] In the bi-global case there will always be $\binom{d_G}{d_G/2}$ such points for even d_G, and $2 * \binom{d_G}{\lfloor d_G/2 \rfloor}$ such points for odd d_G. For the purposes of this paper, we assume that when the maximally minimal distance criterion yields a *set* of points, then we arbitrarily choose a single deceptive attractor from that set. We could also choose the entire set or some subset to be the *set* of attractors. Our deception analyses and results would not change, and we would have multiple deceptive optima with increased "carrying capacity" for niched GAs (Goldberg, Deb, & Horn, 1992). But such considerations are beyond the scope of this paper.

relation that induces a partial ordering on the space of partially deceptive functions, just as Goldberg's bounded deception does. And just as bounded deception has full deception at the extreme, our ordering has a maximum for a given placement of globals. A *maximally* deceptive function has no fewer deceptive partitions at each order than any other deceptive function possible given the set of globals. When the global set consists of only one global, the maximally deceptive function is a fully deceptive function. Otherwise, it is partially deceptive[28].

5.3 A Function to Meet the Bi-global Deceptive Conditions

In the section above we showed that to maximize the number of resolvable partitions at every order in a bi-global problem, we should choose the deceptive attractor to be the point that is maximally minimally distant from the two globals. Such a result is meaningless if we cannot define a function that is actually misleading to a GA in those resolvable partitions. In other words, is it possible to make all the resolvable partitions deceptive? Or does this lead to too many constraints? In this section we construct a function that satisfies our deceptive conditions in all of the resolvable partitions.

The use of maximal minimal distance to G in choosing our deceptive attractor suggests the use of such a function as the fitness function itself. That is, let the fitness of a point/string s be its Hamming distance to the nearest global. Let us call such a minimum distance function simply $f_{md}(s)$:

$$f_{md}(s) = \min_{\forall g \in G} H(s, g) \tag{7}$$

where $H(s,g)$ is the Hamming distance from a point s to the global g. At the globals themselves, of course, we substitute some globally optimal value f_{max} to get a function of minimum distance "plus globals":

$$f_{mdG}(s) = \begin{cases} f_{max} & \text{if } s \in G \\ f_{md}(s) & \text{otherwise} \end{cases} \tag{8}$$

The min-dist function, with globals, $f_{mdG}(s)$, has some very interesting properties. Like trap and unitation functions, it is easy to define, is readily visualized, has mostly linear gradients, and is amenable to schema analysis[29]. Our first observation is that the max-min-dist deceptive attractor is clearly a local optimum of $f_{mdG}(s)$ and indeed must be the global optimum of $f_{md}(s)$. This is true by definition of D as the point of *maximal* min-dist.

Next we perform a modified schema analysis of $f_{mdG}(s)$ looking for deceptive partitions. Our modification to regular schema analysis is this: we ignore the global optima. That is, we assume their fitness is 0 by simply analyzing $f_{md}(s)$ rather than $f_{mdG}(s)$. Adding the globally optimal values f_{max} to our schema fitness calculations would complicate them. The only point of doing so would be to find constraints on f_{max} in order to meet deception conditions on schema average fitnesses. We are not very interested in such constraints, since it really doesn't matter how large f_{max} is, as long as it is globally optimal. The value of the global optima do not enter into the GA's schema processing until a global is found, at which time the search is over and we are no longer interested in GA schema processing. We are modeling GA performance in the *search* for global optima; that is, during the generations preceding the discovery of a global.

We begin our schema analysis by assuming two globals g_1 and g_2. As before, we ignore the bit positions in which g_1 and g_2 agree, since considering the fitness contributions of these bits does not change the ranking of schemata within a competition partition. To see why this is so, remember that the fitness of a schema $\hat{s}$ of $f_{md}(s)$ is the average distance to the nearest global over all strings instantiating $\hat{s}$. Thus the

[28]We note that the bipolar deceptive function (Goldberg, Deb, & Horn, 1992; Deb, Horn, & Goldberg, 1993) is also a maximally deceptive function by our definition above. And a maximally deceptive function for two ($k = 2$) bipolar globals, with the further restriction to symmetric or *folded* unitation, is *bipolar deceptive* according to Goldberg, Deb, and Horn (1992), and Deb, Horn, and Goldberg (1993).

[29]The function $f_{mdG}(s)$ can be seen as a special case of trap functions (Deb and Goldberg, 1991) or as a generalization of Whitley's fully deceptive function (Whitley, 1991) to multiple globals.

bit positions at which g_1 and g_2 agree always add the same amount to the average fitness of each schema within a particular partition, regardless of the schema.

We assume ℓ bit positions in which g_1 and g_2 differ. We now calculate exact schema average fitnesses for any schema in an order o partition. Let $\hat{s}$ be a schema in the given order o partition, and let $d_1^{\hat{s}}$ be the distance from $\hat{s}$ to global g_1, making $(o - d_1^{\hat{s}})$ the distance from $\hat{s}$ to g_2. Note that $0 \leq d_1^{\hat{s}} \leq o$. To calculate the average fitness of $\hat{s}$ in $f_{md}(s)$, we add up the fitnesses of all strings contained in $\hat{s}$ and divide by the number of such strings:

$$\bar{f}_{md}(\hat{s}) = \frac{\sum_{s \in \hat{s}} f_{md}(s)}{2^{\ell - o}} \tag{9}$$

We concentrate now on the sum in the numerator, since that is what will order the average fitnesses of schemata in the partition. The $2^{\ell-o}$ strings in the summation can be divided into two groups, those that are closest to g_1 and those that are closest to g_2. Let d be the number of bit positions from the $\ell - o$ undefined bit positions at which a particular string s disagrees with g_1. That is, the total distance from s to g_1 is $d + d_1^{\hat{s}}$. When this total distance is less than half of ℓ, then s is closer to g_1 than to g_2, and the fitness of s is its distance to g_1: $d + d_1^{\hat{s}}$. There are exactly $\binom{\ell-o}{d}$ strings in $\hat{s}$ that are d bits different from g_1 in the $\ell - o$ undefined bits of the partition. Furthermore, g_1 will be the closest global to s for $d = 0$ up to $d + d_1^{\hat{s}} = \lfloor \ell/2 \rfloor \Rightarrow d \leq \lfloor \ell/2 \rfloor - d_1^{\hat{s}}$. Thus the total contribution of strings near g_1 to the average fitness of $\hat{s}$ is

$$\sum_{d=0}^{\lfloor \ell/2 \rfloor - d_1^{\hat{s}}} (d + d_1^{\hat{s}}) \binom{\ell-o}{d} \tag{10}$$

We calculate a similar sum for the remaining points, which are all closer to g_2. These points have fitness $(\ell - (d + d_1^{\hat{s}}))$, which is their distance to g_2, and they occur when $(\lfloor \ell/2 \rfloor - d_1^{\hat{s}} + 1) \leq d \leq (\ell - o)$. Adding these two sums together, we get the exact schema average fitness of any schema $\hat{s}$ in an order o partition of $f_{md}(s)$:

$$2^{\ell-o} \bar{f}_{md}(\hat{s}) = \sum_{d=0}^{\lfloor l/2 \rfloor - d_1^{\hat{s}}} (d + d_1^{\hat{s}}) \binom{\ell-o}{d} + \sum_{d=\lfloor \ell/2 \rfloor - d_1^{\hat{s}}+1}^{\ell-o} (\ell - d - d_1^{\hat{s}}) \binom{\ell-o}{d} \tag{11}$$

We can show that this sum increases as $d_1^{\hat{s}}$ approaches $d_1^{\hat{s}} = \lfloor o/2 \rfloor$ from $d_1^{\hat{s}} = 0$ and as it approaches $d_1^{\hat{s}} = \lceil o/2 \rceil$ from $d_1^{\hat{s}} = o$. Thus the more maximally minimally distant a *schema* is from the schemata containing the globals, the higher its average fitness. In particular, the maximum of $\bar{f}_{md}(\hat{s})$ occurs at $d_1^{\hat{s}} = \lceil o/2 \rceil$ and $d_1^{\hat{s}} = \lfloor o/2 \rfloor$.

This last result has several important implications for $f_{md}(s)$. First, it means that in all the resolvable partitions, the schema containing the deceptive attractor D has greater average fitness than the schemata containing the globals. Since $f_{md}(s)$ maximizes the number of resolvable partitions at every order, it also therefore maximizes the number of partitions where the global schemata lose to the deceptive schema (type III deceptive partitions). Second, in all but the order one and order ℓ partitions, the global schemata lose to all other schemata[30]. Thus $f_{md}(s)$ is maximally deceptive according to type I partition deception.

Third, the above result implies that the winning schema (i.e., superior to all others) in every partition is the schema that is maximally minimally different from the global schemata. If the winning schema contains the deceptive attractor D, the partition is type II deceptive. How many type II deceptive partitions can a deceptive attractor possibly have, at a given order o? The analysis is similar to our previous analysis. Let d_1^D be the distance from a deceptive attractor D to global optimum g_1 with ℓ-bits of separation between g_1 and g_2. Assume an even partition order o. We can choose half the order o bit positions, from among the d_1^D bit positions in which D and g_1 differ, in exactly $\binom{d_1^D}{o/2}$ ways. We can choose the other $o/2$ bits, from among the $\ell - d_1^D$ bit positions that differ from g_2, in exactly $\binom{\ell - d_1^D}{o/2}$ ways. So there are $\binom{d_1^D}{o/2}\binom{\ell - d_1^D}{o/2}$

[30] At order one (for the ℓ positions in which the globals differ) both competing schemata are global schemata.

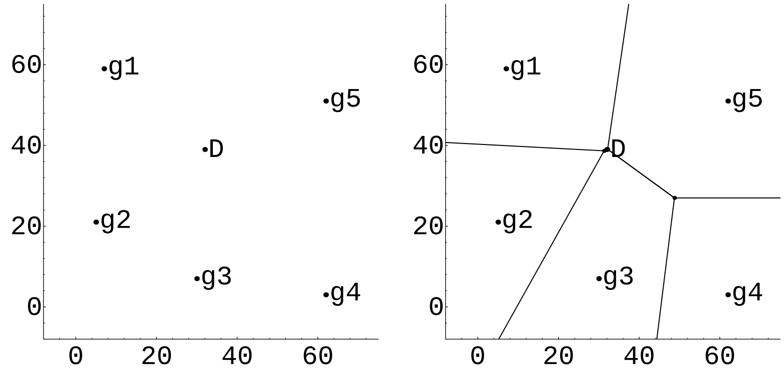

Figure 2: *Left:* A two dimensional problem with $k = 5$ globals at $G = \{(7,59),(5,21),(30,7),(62,3),(62,51)\}$, and a maximally minimally distant point at $D = (32,39)$. *Right:* The Voronoi diagram of the set G.

partitions of even order o in which the deceptive schema wins. Similarly, for odd order o partitions this number is $\binom{d_1^D}{\lfloor o/2 \rfloor}\binom{\ell - d_1^D}{\lceil o/2 \rceil} + \binom{d_1^D}{\lceil o/2 \rceil}\binom{\ell - d_1^D}{\lfloor o/2 \rfloor}$. It is clear that the number of type II deceptive partitions is maximized when $d_1^D = \ell - d_1^D$, $\Rightarrow d_1^D = \ell/2$ for even ℓ (and either of $\lfloor \ell/2 \rfloor$ or $\lceil \ell/2 \rceil$, for odd ℓ). Thus the choice of D as the deceptive attractor maximizes the number of type II deceptive partitions at every order. The function $f_{md}(s)$ is therefore maximally deceptive according to all three types of partition deception we defined.

5.4 Generalization to k globals

We would like to immediately generalize our results to the case of k globals placed arbitrarily. That is, given any set G of k globals in an ℓ-bit problem, we should place the deceptive attractor(s) at those points maximally minimally distant from the entire set G in order to maximally mislead the GA. Unfortunately the analysis used above becomes much more complicated when applied to k globals. With more than two globals we lose the symmetry of bit position agreement/disagreement (where agreement with g_1 at a position means disagreement with g_2 at that position, and where being d_1^D bits away from g_1 means being $\ell - d_1^D$ bits away from g_2).

While we are continuing to explore the generalization to k globals, we can present some intriguing results so far. It is instructive to visualize the spatial relationship of the deceptive attractor D (the max-min-dist point) to the set G of k globals. We therefore return to our use of two dimensional analogues of our binary search spaces and functions. In Figure 2 we show $k = 5$ global optima located at grid positions $G = \{(7,59),(5,21),(30,7),(62,3),(62,51)\}$. Here we assume two integer-valued decision variables $\{x, y\}$ each taking values in the range (0..63). The maximally minimally distant point from G is $D = (32,39)$, as shown on the left of Figure 2. Interestingly, this point lies at the intersection of Voronoi lines in the Voronoi diagram of G, shown on the right of Figure 2. Since the lines in a Voronoi diagram divide the space into neighborhoods of elements of G and their nearest neighbors, and thus lie on points equidistant from the nearest two globals, D must lie at an intersection of at least three Voronoi lines.

To better illustrate the relationship between the Voronoi diagram and our maximum minimum distance criterion, we redefine f_{md} to be a function of integers x and y rather than of binary strings:

$$f_{md}(x,y) = \min_{\forall (x_g,y_g) \in G} \sqrt{(x_g - x)^2 + (y_g - y)^2} \tag{12}$$

Noting that the maximum value is $f_{md}(D) \approx 32.0156$, we choose f_{max} to be 34.

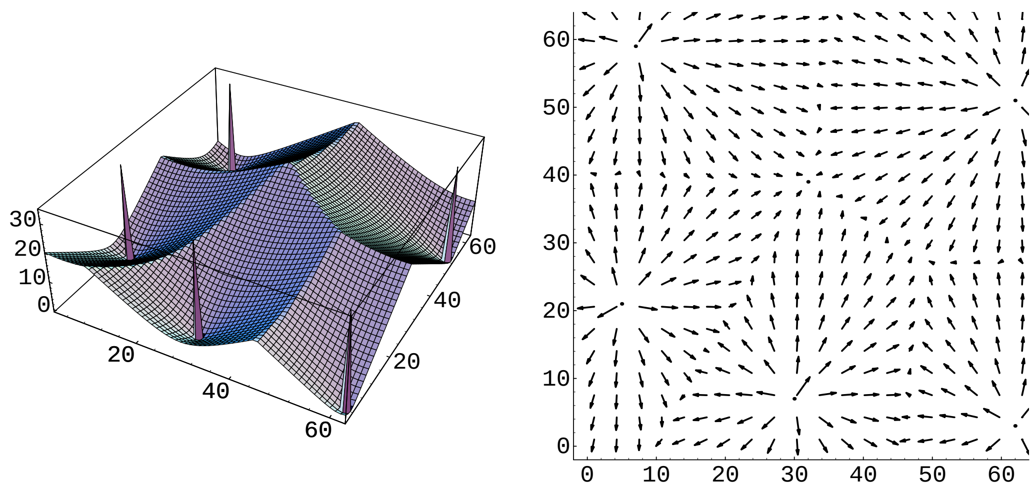

Figure 3: *Left:* A surface plot of $f_{mdG}(x,y)$. *Right:* This function is difficult for a hillclimber because all gradients lead away from the nearest global and, eventually, to a local optimum, as shown in this vector plot of $f_{mdG}(x,y)$.

$$f_{mdG}(x,y) = \begin{cases} 34 & \text{if } (x,y) \in G \\ f_{md}(x,y) & \text{otherwise} \end{cases} \tag{13}$$

We plot $f_{mdG}(x,y)$, on the left of Figure 3, for the five globals. We note how the ridges correspond to the lines of the Voronoi diagram, and how each local optimum (except the globals) occurs at intersections of such lines. By definition of $f_{md}(x,y)$ and Voronoi diagrams, this must be the case. Thus the min-dist function $f_{md}(x,y)$ leads to an interesting, rugged, multimodal landscape with apparent basins of attraction for hillclimbers. To see that there are indeed basins of attraction for the local optima, we plot the local gradients of $f_{md}(x,y)$ on the right of Figure 3. As expected, these gradients point away from the nearest globals and towards the nearest ridge. We would expect a simple hillclimber to follow the local gradients up to the nearest ridge and thence to a local optimum. Finally we note that the max-min-point D seems to have the largest basin of attraction (of all optima shown) for a simple hillclimber.

But again, to be of interest in the study of GA difficulty and multimodality, the min-dist function $f_{mdG}(s)$ with k globals must be misleading to the GA's crossover operator as well as to mutation. The schema analysis of $f_{mdG}(s)$ is complicated by the interactions of the multiple globals. We can however make a quick observation that indicates we are headed in the right direction (in order to make the GA head in the wrong direction!).

Assume any order o resolvable partition in an ℓ-bit problem with k globals. Let D be our choice of deceptive attractor. Let $\hat{s}_D$ be the schema containing the deceptive attractor and let $\hat{s}_i$ be the schema containing[31] global g_i, $1 \le i \le k$. We define $\vec{d^D} = \{d_1^D, d_2^D, ..., d_k^D\}$ to be the vector of distances from the deceptive schema $\hat{s}_D$ to each of the global schemata $\hat{s}_i$. Thus d_i^D, where $(0 \le d_i^D \le o)$, is the Hamming distance from $\hat{s}_D$ to $\hat{s}_i$, defined only over the o fixed bit positions. We also define similar distance vectors for each of the global schemata $\hat{s}_i$: $\vec{d^i} = \{d_1^i, d_2^i, ..., d_k^i\}$, for $(i \in 1..k)$ where d_j^i is the Hamming distance $(0 \le d_j^i \le o)$ between $\hat{s}_i$ and $\hat{s}_j$. Finally, for each possible setting s of the $(\ell - o)$ bits undefined in this partition, we associate another vector of distances to the globals: $\vec{d_s} = \{d_1^s, d_2^s, ..., d_k^s\}$, where d_i^s is the

[31] Some schemata may contain more than one global. This does not affect our analysis.

number of bits ($0 \leq d_i^s \leq \ell - o$), in which s differs from global g_i over the ($\ell - o$) bits undefined by the partition (but defined by s).

To calculate the average fitness of a global schema $\hat{s}_i$, we generate the $2^{\ell-o}$ substrings s in the hyperplane defined by $\hat{s}_i$. For each s we calculate its distance vector $\vec{d_s}$ and add to it the distance vector $\vec{d^i}$ for $\hat{s}_i$. We then take the minimum component of the vector resulting from this summation as the distance to the nearest global from s. Summing the minimum components of these summed distance vectors over all $s \in \hat{s}_i$ and dividing by $2^{\ell-o}$ gives the average schema fitness[32] $\bar{f}_{md}(\hat{s}_i)$. We calculate $\bar{f}_{md}(\hat{s}_D)$ similarly, using $\vec{d_s}$ and $\vec{d^D}$.

Now if $\vec{d^D} \geq \vec{d^i}$ in all components (that is $d_j^D \geq d_j^i$, $\forall j \in 1..k$), for a particular global g_i, then clearly $\bar{f}_{md}(\hat{s}_D) > \bar{f}_{md}(\hat{s}_i)$, and the deceptive schema will beat the ith global schema. If $\vec{d^D}$ is superior (i.e., greater in all k components) to all k of the $\vec{d^i}$, then the deceptive schema will be superior to all the global schemata in that partition. Such a partition would thus be type III deceptive. To maximize the number of such partitions, we should try to increase the number of partitions in which the deceptive schema is as different as possible from all of the global schemata. Note that $\vec{d^D} \geq \vec{d^i}$ is only a *sufficient* condition for type III deception in a partition. Thus the number of partitions satisfying $\vec{d^D} \geq \vec{d^i}$ is only a lower bound on the number of type III deceptive partitions. However, this sufficient condition is met more often (i.e., at more partitions at every order) by the max-min-dist point D than by any other string. This in turn suggests that the max-min-dist point D will turn out to be the most attractive local optimum for GA crossover on the min-dist function with k globals.

The $f_{mdG}(s)$ function, although general with respect to the number and placement of globals, can be generalized further. We could make each global more attractive by simply increasing the radii $r_{g,i}$ of each global g_i. Thus $f_{md}(s)$ would be defined as before *outside* of the $r_{g,i}$-bit neighborhoods of each global. Within each neighborhood, however, the $f_{mdG}(s)$ could be a plateau of optimal fitness, or a hill leading to the global. Alternatively, we could simply concatenate m copies of $f_{mdG}(s)$, each with k globals. This would yield k^m global optima for the full function. Goldberg, Deb, and Horn (1992) used such a construction to build a "massively multimodal" and deceptive function from five copies of their six bit bipolar deceptive function, resulting in 32 global optima among over five million local optima. Five million is approximately 1/2% of the 2^{30} total size of the search space, and is thus within two orders of magnitude of the absolute maximum number of local optima (50%). Considering that the problem was of intermediate difficulty to a simple GA[33], this concatenation of subfunctions generated a surprisingly large number of optima[34].

6 Conclusions

Modality by itself, if defined solely as the number of local optima, actually tells us little about the difficulty of searching a space. Similarly, full deception and GA-easiness on uni- or bimodal landscapes tell us only about the extreme boundaries of GA success and failure. But maximally deceptive multimodal functions allow us to embed deception in much more rugged and general landscapes, and to define arbitrarily sized and spaced local optima and their basins of attraction to a GA. However, generalizing the definitions of full deception in order to characterize partial deception is tricky. It is all too easy to lose the essential quality of misleadingness. We have introduced a general method of relaxing the deception conditions that partially orders problem spaces according to GA misleadingness. We can now make more connections between crossover's search of hyperplanes and the population's movement over the fitness landscape. For example, in the past we have generally chosen the deceptive attractor D first, and then defined deceptive

[32] Note that we again use $f_{md}(s)$ for schema fitness calculations, rather than $f_{mdG}(s)$, effectively ignoring the globals.

[33] A GA with sufficiently large population could find a global, but with smaller populations converged to a local suboptimum.

[34] Early experiments, in which five copies of a ten bit $f_{mdG}(s)$ subfunction with $k = 5$ randomly placed global optima are concatenated to form an $\ell = 50$ bit problem whose fitness is the sum of the five subfunction fitnesses, similar to (Goldberg, Deb, & Horn, 1992), indicate that the problem is indeed misleading for a simple GA, with most runs converging to deceptive attractors (local optima) unless sufficiently large populations are used to overcome the deception (Goldberg, Deb, & Clark, 1992). These results will be reported in detail in a future publication.

partitions in terms of D. Here we saw how maximizing the number of deceptive partitions of all orders leads to a unique choice of deceptive local optimum, since the number of deceptive partitions increases with distance to the set of globals. Finally, we defined a simple, general, deceptive multimodal problem, the min-dist function, that seems to relate the attraction of crossover to interesting geometric features of the landscape, such as Voronoi diagrams and Delaunay triangularizations, in addition to local optima, ridges, and the landscape's gradient field. We must define landscape characteristics important to crossover or we will continue to look at fitness surfaces from a hillclimber's rather limited point of view.

Acknowledgments

We thank Georges Harik, Kaitlin Sherwood, Joseph C. Culberson, Terry Jones, and two anonymous (FOGA 3) referees for their comments and suggestions. The first author acknowledges support provided by NASA under Contract NGT-50873. The second author acknowledges support provided by AFOSR under Grant F49620-94-1-0103 and by the US Army under Contract DASG60-90-C-0153.

References

Bethke, A. D. (1981). *Genetic Algorithms as Function Optimizers.* Doctoral dissertation, University of Michigan at Ann Arbor.

Deb, K., Horn, J., & Goldberg, D. E. (1993). Multimodal deceptive functions. *Complex Systems.* 7 (1993), 131–153.

Goldberg, D. E. (1987). Simple genetic algorithms and the minimal deceptive problem. In L. Davis, ed., *Genetic Algorithms and Simulated Annealing.* London: Pitman Publishing, 74–88.

Goldberg, D. E. (1989a). *Genetic Algorithms in Search, Optimization, and Machine Learning.* Reading, MA: Addison–Wesley.

Goldberg, D. E. (1989b). Genetic algorithms and Walsh functions: part I, a gentle introduction. *Complex Systems.* 3 (1989), 129–152.

Goldberg, D. E. (1989c). Genetic algorithms and Walsh functions: part II, deception and its analysis. *Complex Systems.* 3 (1989), 153–171.

Goldberg, D. E. (1991). Construction of high-order deceptive functions using low-order Walsh coefficients. *Annals of Mathematics and Artificial Intelligence, 5.* 35–48.

Goldberg, D. E. (1993a). Making genetic algorithms fly: a lesson from the Wright brothers. *Advanced Technology for Developers.* Vol. 2, February. 1–8.

Goldberg, D. E. (1993b). Genetic and evolutionary algorithms come of age. *Communications of the Association for Computing Machinery.*

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

Goldberg, D. E., Deb, K., & Horn, J. (1992). Massive multimodality, deception, and genetic algorithms. In R. Maenner, B. Manderick, ed.s, *Parallel Problem Solving From Nature, 2.* North-Holland, 37–46.

Goldberg, D. E., Deb, K., & Korb, B. (1991). Don't worry, be messy. *Proceedings of the Fourth International Conference on Genetic Algorithms.* Belew, R. K., & Booker, L. B., (ed.s), Morgan Kaufmann, 24–30.

Grefenstette, J.J. (1992). Deception considered harmful. *Foundations of Genetic Algorithms, 2.* Whitley, L. D., (ed.), Morgan Kaufmann, 75–91.

Harary, F., Hayes, J. P., & Wu, H-J. (1988). A survey of the theory of hypercube graphs. *Computational Mathematical Applications*, 15:4, 277–289.

Holland, J. H. (1992). *Adaptation in natural and artificial systems* (2nd ed.). Cambridge, MA: MIT Press.

Homaifar, A., Qi, X., & Foster, J. (1991). Analysis and design of a general GA deceptive problem. *Proceedings of the Fourth International Conference on Genetic Algorithms.* Belew, R. K., & Booker, L. B., (ed.s), Morgan Kaufmann, 196–203.

Horn, J., Goldberg, D. E., & Deb, K. (1994). Long path problems. To appear in *Parallel Problem Solving from Nature, 3; Springer-Verlag Lecture Notes in Computer Science.*

Jones, T., & Rawlins, G. J. E. (1993). Reverse hillclimbing, genetic algorithms and the busy beaver problem. In S. Forrest, ed., *Proceedings of the Fifth International Conference on Genetic Algorithms.* Forrest, S. (ed.), Morgan Kaufmann, 70–75.

MacWilliams, F. J., & Sloane, N. J. A. (1977). *The Theory of Error Correcting Codes.* Amsterdam, New York: North-Holland.

Mahfoud, S. W. (1993). Simple analytical models for genetic algorithms for multimodal function optimization. *Proceedings of the Fifth International Conference on Genetic Algorithms.* Forrest, S. (ed.), Morgan Kaufmann, 643.

Mitchell, M., Forrest, S., & Holland, J. H. (1991). The royal road for genetic algorithms: fitness landscapes and GA performance. *Toward a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life.* Varela, J., & Bourgine, P. (ed.s), The MIT Press, 245–254.

Mitchell, M., & Holland, J. H. (1993). When will a genetic algorithm outperform a hill climbing? *Proceedings of the Fifth International Conference on Genetic Algorithms.* Forrest, S. (ed.), Morgan Kaufmann, 647.

Preparata, F. P. (1974). Difference-preserving codes. *IEEE Transactions on Information Theory*, IT-20:5, September, 1974, 643–649.

Whitley, D. L. (1991). Fundamental principles of deception in genetic search. *Foundations of Genetic Algorithms.* Rawlins, G. J. E. (ed.), Morgan Kaufmann, 221–241.

Wilson, S. W. (1991). GA-easy does not imply steepest-ascent optimizable. *Proceedings of the Fourth International Conference on Genetic Algorithms.* Belew, R. K., & Booker, L. B. (ed.s), Morgan Kaufmann, 85–89.

Wright, S. (1988). Surfaces of selective value revisited. *American Naturalist 131*, 115–123.