Appearing in From Animals to Animats 8: Proceedings of the Eighth International Conference on Simulation of Adaptive Behavior, July 13-17, 2004, Los Angeles, CA.

Evolving Walking: The Anatomy of an Evolutionary Search

Chad W. Seys
Dept. of Electrical Engineering and Computer Science
Case Western Reserve University
Cleveland, OH 44106
chad.seys@cwru.edu

Randall D. Beer
Dept. of Electrical Engineering and Computer Science
Dept. of Biology
Case Western Reserve University
Cleveland, OH 44106
beer@eecs.cwru.edu

Abstract

The evolution of a continuous time recurrent neural network central pattern generation for walking is characterized and found to proceed in two phases. The first phase spans the beginning of the search through the generation at which a "breakthrough" individual is discovered. The second phase proceeds from that generation forward. The first phase is most quickly completed if each succeeding population is as random as possible. Hence GA searches performed at lower mutation variances require more generations to discover a breakthrough individual than higher mutation variances searches. In the second phase the best fitness in the population most rapidly increases at low mutation variances. The role of parameter space structure in these trends will be examined.

1. Introduction

Evolutionary algorithms (EVAs) have long played an important role in adaptive behavior research (Beer & Gallagher, 1992; Cliff et al., 1993; Nolfi & Floreano, 2000). From an engineering perspective, EVAs promise an opaque design methodology whereby solutions can be produced to complex problems without necessarily understanding them. Scientifically, EVAs promise a means of concretely exploring theoretical ideas with a minimum of a priori assumptions. Both activities would greatly benefit from an improved understanding of the behavior of evolutionary search.

Within the EVA community, a great deal of work has been done on the analysis of evolutionary algorithms (van Nimwegen et al., 1997; Gao et al., 1998). However, these studies often make unrealistic assumptions (e.g., infinite population size), or are carried out with highly simplified fitness functions. It is not at all clear how results obtained from such studies apply to the evolution of controllers for situated, embodied, and dynamical agents, where the fitness spaces can be high-dimensional, discontinuous, and extremely irregular. Furthermore, an agent's fitness often depends in very indirect ways on the parameters being evolved. In a neurally-controlled agent, for example, the parameters determine neural dynamics, which interact with body dynamics and environment dynamics to produce behavior, which is then evaluated by a behavioral fitness function.

In this paper, we examine in some detail the behavior of a simple evolutionary algorithm on a walking task. A significant advantage of this task is that the neural model, body model and fitness function are well-understood (Chiel, et al., 1999; Beer et al., 1999). First, we describe the general features of evolutionary searches on this task. Then we divide the evolutionary search into two distinct phases, showing that the EVA behavior is rather different in the two phases and explaining the basis of these differences. Finally, we show how our results can be used to both predict features of the evolutionary search and to improve the performance of the EVA on this task. The paper concludes with a discussion of the broader applicability of our results.

2. Methods

The agent used in the experiments is the same as used in Beer and Gallagher (1992) and consists of a body with a single leg (of nominal length 15) with a foot, and effectors to move them. The leg is connected to the body by a joint which allows effectors to apply clockwise and counterclockwise torques which sum to produce a resultant torque which may act to move the leg through an angle of $\pm\pi/6$. If the foot is raised the opposing torques are allowed to have a maximum value of 1/40. When the foot is lowered, the resultant torque applies translational force on the body with a maximum value of 1/20. If the body has no support, the translational velocity is 0, while if the body is supported the acceleration is equal to the translational force. The leg is allowed to stretch past its nominal length (and angle) if the foot is lowered, but no forces can be applied. Finally, if the distance between the body's center of mass and foot is too great (greater than 20 along the axis in the direction of travel), the leg snaps back to nominal length and maximum clockwise angle, the body loses support, and translational velocity is 0. In these experiments an agent is always created with the leg in the fully forward (counterclockwise) position.

The effectors of the agent are controlled by a three neuron continuous time recurrent neural network (CTRNN). Each neuron *i* has the state equation

$$\frac{dy_i}{dt}=\frac{1}{\tau_i}\left(-y_i+\sum_{j=1}^{N} w_{ji}\sigma(y_j+\theta_j)\right)$$

where y_i is the state of each neuron, τ_i is the time constant, w_{ji} is the weight of the connection between the j^{th} and i^{th} neuron, θ_j is bias term, and $\sigma(x)=1/(1+e^{-x})$ is the standard logistic activation function. The neurons are named after the function of the effectors they control: Forward Swing (FS), Backward Swing (BS), and FooT (FT). The foot is down when FT's output is above 0.5 and up otherwise. When the foot is up, the output of BS and FS scale the torques swinging the leg, when down these outputs scale the translational force applied to the body.

A real valued genetic algorithm (GA) (Bäck, 1996) was used to evolve the CTRNN parameters. Each 3 neuron neural circuit was represented as 15 real values. A generation of GA search contains 100 individuals described by a point in 15 dimensions restricted to [-1,1]. The first generation of individuals is created by choosing points randomly from this space. After initialization, succeeding generations are created from the proceeding generation by copying the highest fitness individual (a.k.a. "the elite") and then by mutating parents chosen for reproduction by a linear rank based method. A parent vector is mutated into a child vector by adding a displacement vector created by randomly choosing a direction in the fifteen dimensional space and then randomly choosing a magnitude from a Gaussian distribution with a mean 0 and variance of σ^2 (also called "mutation variance" throughout this paper). If the point resulting from the addition of the parent vector and displacement vector lies outside [-1,1] then those dimensions are rounded ("clipped") to within the range.

To evaluate the fitness of an individual, the neural circuit is instantiated and the coupled neural circuit/body simulated. A neural circuit is instantiated by linearly mapping the point representing the individual in search space to CTRNN parameter space (with ranges [-16,16] for weights and biases, [0.5,10] for time constants). The motor neuron outputs are initialized to zero. The coupled neural network body are numerically integrated over 220 time steps by the forward Euler method at a step size of 0.1 and the individual's fitness is the distance covered in this time. (The actual fitness used is the average velocity, but because the amount of time is fixed it is equivalent, and easier, to think of the fitness in units of single leg insect steps.) The distance of one full step gives the agent a fitness of approximately 0.125, two steps $2*0.125=0.25$ and so on. An optimal agent (Beer et al., 1999) has time to go the distance of a little over five steps in the time allowed.

3. General Characteristics

All GA searches on this task have a common structure when best fitness per generation is plotted. Initially all searches have a best fitness of approximately 0.125. No improvement to the best fitness is made for a varying number of generations until an individual is found with a fitness greater than approximately 0.125. After this individual is discovered, there is a steady ratcheting upwards in the best fitness which eventually levels off to different levels in each search as the theoretical maximum of 0.627 (Beer et al., 1999) is approached. We call the discovery of a greater than 0.126 fitness individual "breakthrough". (As will be discussed later, 0.125 individuals are qualitatively different than 0.126 individuals.) The period of the search which occurs before breakthrough will be referred to as "phase I" and the period after breakthrough "phase II".

Figure 1 shows three searches performed nominal mutation variance (0.5) for this problem. One can see that all three searches begin with a best fitness of 0.125 and remain at this fitness for a varying number of generations until breakthrough occurs. In these searches, breakthrough occurs at generations 6, 42, and 116 as marked by the small arrows.

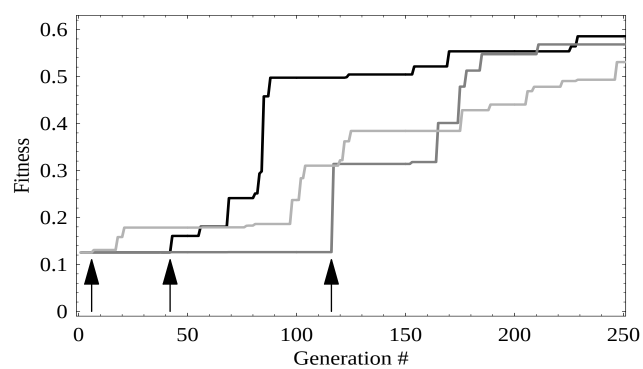

Figure 1: Best fitness by generation of three searches Small arrows indicate "breakthrough", which separates phase I from phase II.

There are parallels here to work on neutral networks. The pattern of constant fitness while a population moves over parameter space is related to the concept of neutrality, phase I is similar to to an epoch, and breakthrough is analogous to the discovery of a portal. Information on neutral networks can be found in van Nimwegen (1999).

4. Phase I

Phase I lasts from the beginning of search until the population contains at least one individual whose fitness is greater than 0.126. The breakthrough generation distribution is very non-normal, with a strong skew toward lower generations. Figure 2 is a plot of the fraction of GA searches which have broken through at the generation

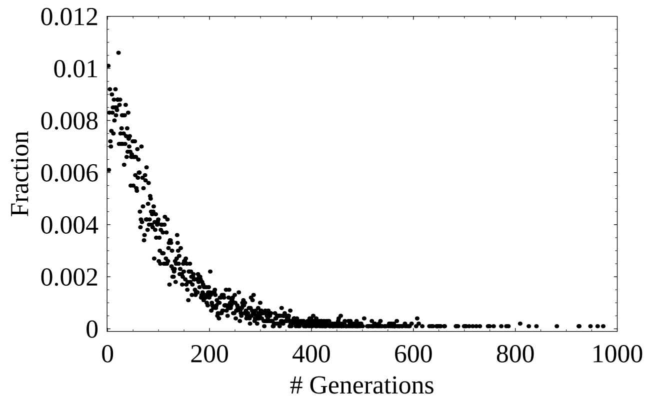

Figure 2: Fraction of breakthroughs per generation for 10,000 GA searches at nominal mutation variance.

indicated. After 250 generations slightly greater than 90% of searches have broken through, while after 1000 generations more than 99% have completed phase I.

This distribution can be fit with an exponential function $N(t) = N_0 e^{-pt}$. The exponential function is commonly used to fit random processes such as radioactive decay which occur with a constant probability per unit time. The data are converted into the appropriate form by calculating the fraction of searches yet to breakthrough at each generation. Then N(t) gives the fraction of breakthroughs at generation t, N_0 the initial fraction not broken through (should be around 1), and p the probability of breakthrough per generation. Ignoring the last 10% of searches in the tail of the distribution, the best fit equation has $N_0 = 0.984$, p = 0.00962, and a r^2 value of 0.998. The median[1] (i.e. generation at which half of all breakthroughs have taken place) is predicted to be at generation 72, which is very near the observed median of 71.

Interestingly, random sampling in "generations" of 100 individuals discovers a breakthrough individual with a median of 73 "generations". Randomly generating individuals within parameter space and checking "generations" of 100 individuals for at least one breakthrough individual essentially simulates a GA search in which one generation (of 100 individuals) is completely unrelated to the next generation. Thus, the GA search at nominal mutation variance is behaving like random sampling at the level of breakthroughs per generation.

Not only does the breakthrough event occur with random-like equal probability per generation, but the population's fitness change per generation also has analogies to random processes. To plot how the population's fitness is changing, a line is drawn from the fitness of each parent in one generation to the fitness of all that parent's children in the next generation. Figure 3 is a lineage web of the first 60 generations of search A above. One can see that each generation parents which have a

[1]The median is more stable than the mean in describing data with this distribution: the median breakthrough generation is 71 whether the GA search is carried out to 250 or 1000 generations. Setting those searches which have not broken through equal to the maximum generation, the mean is 93 for 250 generations 105 for 1000 generations.

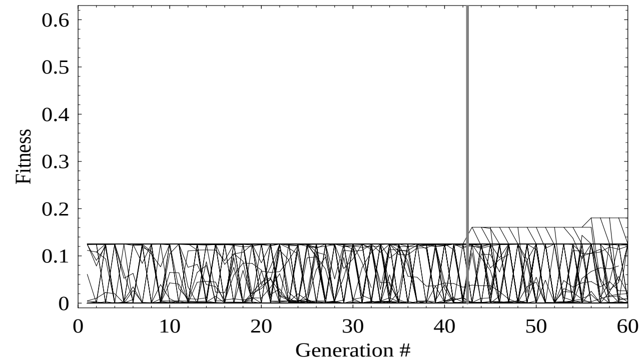

Figure 3: Phase I lineage web of search A from Figure 1. "X"s result from parents in one fitness band having children in the other fitness band – a very common occurrence. Phase I ends at generation 43 when a ≥ 0.126 individual is discovered (gray vertical line).

fitness of 0.125 have 0 fitness children as well as 0.125 fitness. Similarly 0 fitness parents often have children which have a 0.125 fitness as well as the same fitness. Graphically this appears as the two bands at 0 and 0.125 fitness connected by "X"s. By the lack of lines changing direction between 0 and 0.125 fitness it can be concluded that most individuals are either 0 or 0.125 fitness and very few individuals appear at intermediate fitness values. A plot (Figure 4) of individuals' fitness distribution in every generation of ten searches up until breakthrough confirms that a vast majority of individuals are either of 0 or 0.125 fitness. This pattern suggests a dynamic equilibrium between two relatively large fitness plateaus with relatively few intermediary fitness regions.

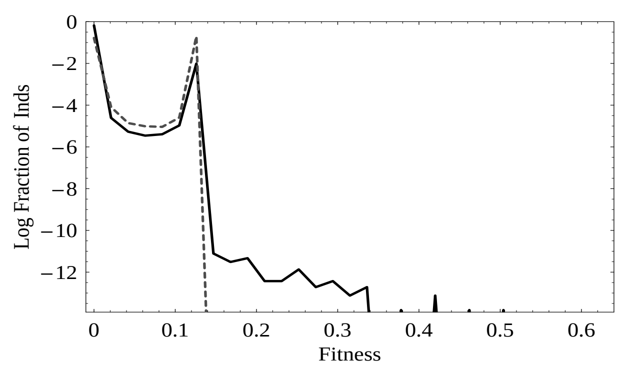

Figure 4: Distribution of fitnesses for A) GA search in phase I (dashed line) and for B) a random sample of parameter space (solid line). Random sampling and GA search at nominal mutation variance result in a similar distribution of individuals less than 0.125 fitness. Adapted from Ames (2003)

Non-oscillating neural networks are all that is required to achieve the fitnesses observed in phase I. Zero fitness is achieved simply by not moving the agent's body at all: behaviorally this can be done either by never putting the foot down to make contact with the ground or by leaving the agent's leg in the initial fully forward position. A fitness of 0.125 is given to individuals which move the distance equivalent to a full swing of the leg. Monotonically changing neural network outputs are all that is required for both of these fitnesses. Intermediary fitnesses are less common because the sigmoid activation function makes it easier for neurons to be "on" or "off" rather than an output resulting in a step midway between all or nothing.

Random sampling of individuals from parameter space have a fitness distribution (Figure 4) very similar to the GA search in phase I. (The exception is individuals greater than 0.125 fitness: there are no individuals greater than 0.125 in phase I.) The close match between the fitness distributions is most easily explained by GA search populations covering the same regions of parameter space as random sampling. This suggests that at the nominal mutation variance GA search is emulating random sampling.

Up until now it has been argued that GA search *at nominal mutation variance* has been analogous to random sampling. At lower mutation variances this analogy breaks down immediately. Figure 5 shows the median generation of breakthrough for a wide range of mutation variances compared to the median number of random individuals in "generations" of 100 needed to discover at least one breakthrough individual. At mutation variances greater than approximately 0.3 the GA search does at least as well as random sampling while at mutation variances less than approximately 0.3 the median number of generations to finding a breakthrough individual is larger than random sampling.

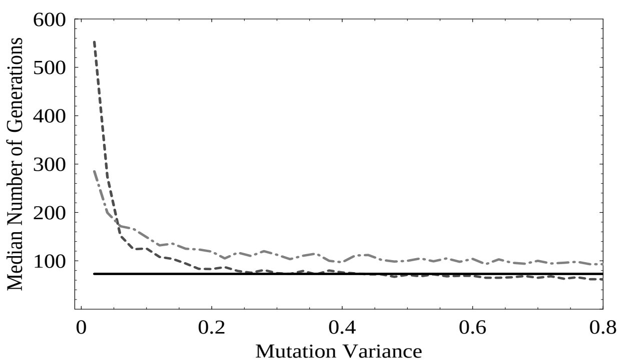

Figure 5: Median number of generations to discovery of a ≥ 0.126 fitness individual. A) Evenly dashed line is GA search, B) syncopated line is diffusion search, and C) solid line is random samples of 100 individuals. At higher mutation variances (>2) all lines converge.

Selection has a varying role over the range of mutation variances. This can be determined by running "diffusive searches" at the same mutation variances as the GA searches and comparing their median generation to breakthrough. Diffusive search is performed by turning off elitism and having each parent generate exactly one child using the same procedure as GA search.

At very low mutation variances, less than 0.05, one can see that the GA search performs more poorly than diffusive search. This is most likely due to premature convergence of the GA search.

At higher mutation variances, between 0.05 and 2, GA search does better than diffusion search. The reason for this is currently unknown: the slim majority parents of breakthrough individuals in diffusion search are nearer 0 in fitness than 0.125, eliminating the possibility that the 0.125 areas are "nearer" breakthrough areas in general. Turning off elitism and turning down the selective advantage given to higher ranked individuals eliminates this gap.

At very high mutation variances (greater than 2) GA, diffusion, and random sampling all have nearly the same median generation of breakthrough. This is not surprising because as mutation variance is increased each generation becomes less related to the previous and approximates a random sample.

The pattern of performance of GA search in phase I search and random sampling of individuals lead to some conclusions about the structure of parameter/fitness space. There are vast plateaus of fitness in parameter space at 0 and 0.125. Above 0.125 fitness parameter space areas are at least 5 orders of magnitude less common. Because of a lack of fitness gradients in this parameter/fitness terrain, search in phase I does well at high mutation variances by emulating random sampling.

5. Phase II

Phase II begins when a "breakthrough" individual greater than or equal to 0.126 fitness is found. A breakthrough individual must have a neural network whose outputs are not monotonicly changing because a fitness of greater than or equal to 0.126 is achievable only by covering a distance equivalent to more than one full swing of the leg. In other words the agent must repeat some movements such as move the leg both backward and forward and both raise and lower it's foot (at the appropriate time).

Phase II has a well defined beginning, but not a well defined end because it is very difficult to find an individual with the theoretical maximum fitness by GA search. A lower, more easily achieved fitness which qualitatively remains a good walker is desirable for purposes of measuring the elapsed number of generations between breakthrough and “end of phase II”. Individuals with a fitness of 0.5 show no obvious behavioral deficits (i.e. seem to walk without difficulty) and are not too difficult for the GA to find.

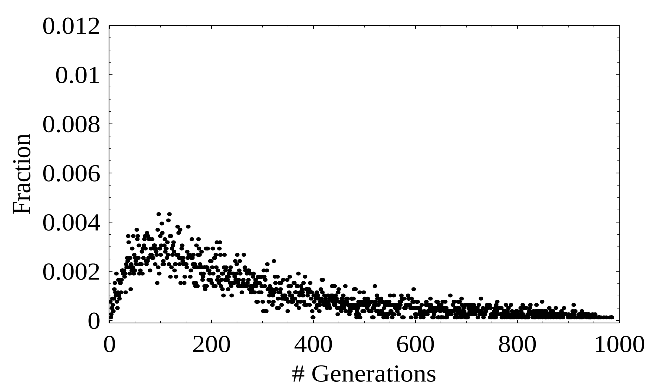

Figure 6: Fraction of 10,000 GA searches which have an elapsed number of generations between breakthrough and end of phase II.

Using these criteria for the beginning and end of phase II a plot (Figure 6) can be made of the elapsed number of generations. In contrast to the breakthrough distribution in the previous section, this distribution initially rises to peak around 100 generations elapsed and a descending phase for the remainder of the distribution Approximately 55% of searches are able to complete phase II when allowed 250 generations after breakthrough and slightly less than 80% if allowed 1000 generations.

The exponential function does not fit this distribution as nicely as the phase I breakthrough distribution. Following the same procedure as outlined in phase I, the best fit equation has $N_0 = 1.12$, $p = 0.00385$, and a r^2 value of 0.999. The median is predicted to be at generation 180, which is not near the observed median of 213.

Lineage webs of phase II (e.g. Figure 7) allow us to examine patterns in the population fitness change between breakthrough and 0.5 fitness. The majority of the population continues to exist throughout the search in a band of fitness from 0 to 0.125. A very small fraction of the population has a fitness of greater than 0.125.

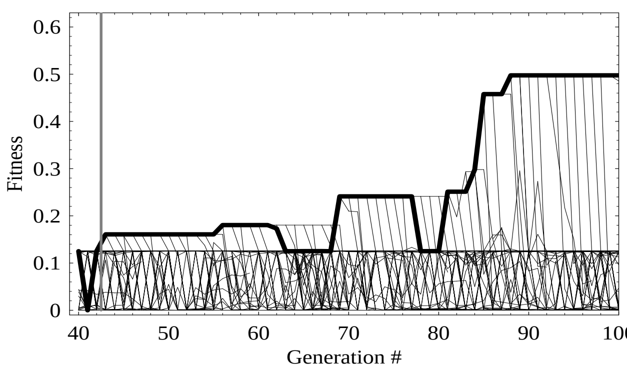

Figure 7: Phase II lineage web of search A from Figure 1. The elite individual increases in fitness every generation. The lineage which leads to the best individual in the last generation loses and gains fitness. Phase II begins at breakthrough (vertical gray line).

Of the individuals with a fitness greater than 0.125, most of them are descendants of the first breakthrough individual: of the occurances of the elite individual being replaced by a higher fitness individual, less than 4% of the time the individual is unrelated to the elite. This is not too surprising given that during the course of each search the fitness of the elite individual progressively increases above breakthrough fitness, making its replacement more and more difficult.

The descendants of the elite quickly diffuse away from the high fitness regions. Most children quickly join the 0 to 0.125 fitness band of the population. Less than than 6% of the elite's descendants ever achieve a fitness higher than the elite. Of those, the more closely related the descendant is, the more likely it is to replace the elite. Of those descendants that do replace the elite, 70% of the time it is a child, 13% a grandchild, 6% a great-grandchild, 3% a great^2-grandchild, 2% a great^3-grandchild, and less than 1% for all others. In Figure 7 the thicker line traces the lineage which produces the elite individual in the last generation of the search. Along this lineage one can see instances of both the elite's child (generation 56) and of great-grandchildren of varying degrees (generation 68, 81) replacing the elite.

A smaller mutation variance more effectively explores the region of parameter space which contains the above breakthrough fitness individuals. A survey of GA searches at various mutation variance was performed to characterize the median elapsed number of generations between the discovery of a breakthrough individual and the discovery of an individual with a fitness greater than or equal to 0.5 (Figure 8). The median elapsed number of generations increases with increasing mutation variance. The best mutation variance for completion of phase II is approximately 0.05, $^1/_{10}$ the magnitude of the nominal mutation variance taking roughly $^3/_4$ as many generations. The sensitivity of elapsed number of generations is not as great as in phase I. This may be because the distribution of displacements between parent and child are Gaussian even at high mutation variances most of the parent-child displacements remain small.

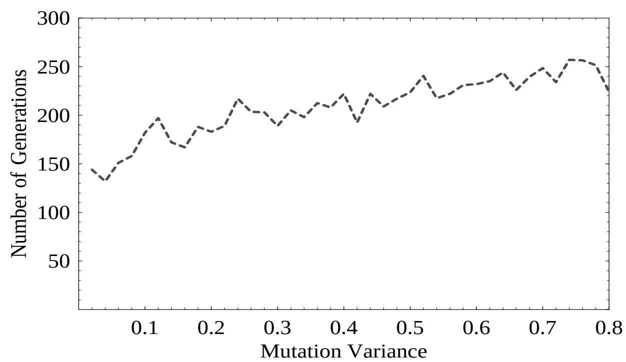

Figure 8: Mean number of generations between discovering a ≥ 0.126 and then a ≥ 0.5 fitness individual. The elapsed number of generations slowly increases with increasing mutation variance.

The improvement at low mutation variances is due to more distantly related descendants replacing their elite parents. For 0.05 mutation variance around 8% of the elite's descendants eventually attain a higher fitness than the elite compared to less than 6% for 0.5 mutation variance. Of these, fewer children replace the elite than at the nominal mutation variance (57% versus 66%), while more distantly related descendants replace the elite at a higher rate. The low mutation variance appears to allow descendants to jostle around longer in the vicinity of the high fitness region.

Taken together, phase II is best characterized as a GA search, not randomly searching large, flat fitness plateaus, but instead more like hill-climbing a smaller parameter/fitness region with a gradient. The rise and fall of the generations elapsed from breakthrough to 0.5 reflects the variation in distance between the breakthrough individual and the greater than 0.5 region over searches. Finally, at lower mutation variances the elite's descendants are given more of a chance to find higher fitness regions rather than diffuse off into the low fitness plateaus that occupy the majority of parameter space.

The above description of the structure of parameter space can be shown to exist for lower than 15 dimensional slices through parameter/fitness space. A two dimensional slice of parameter space is made by instantiating individuals

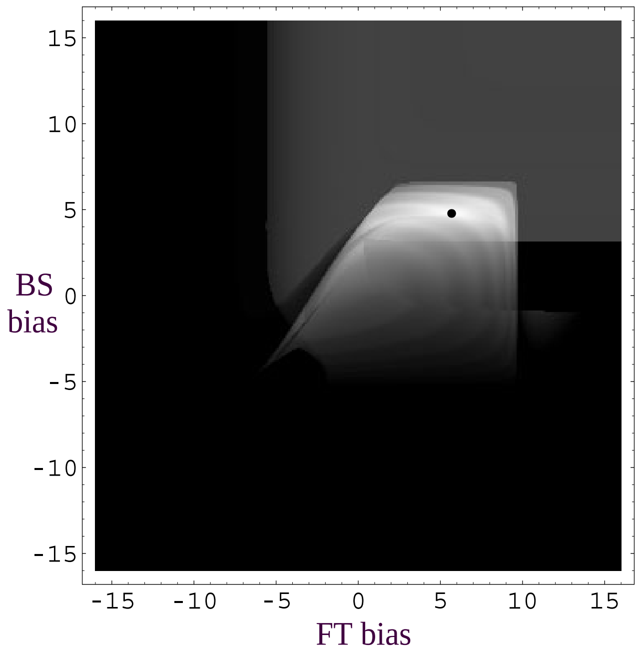

Figure 9: A two dimensional slice through fifteen dimensional parameter space. Dot is individual found by GA search. Gray scale is log of fitness. Zero fitness – black, 0.125 fitness – large dark gray area in upper right, lighter is higher fitness.

all of whose parameters are fixed but two. The resulting fitness of the individuals is plotted in gray scale at the coordinate containing the values of the two varying parameters. Most randomly chosen slices through parameter space will show fitnesses no greater than 0.125. If one centers a slice on a high fitness individual one gets a view of parameter space such as in Figure 9. The dot is a high fitness individual (found by GA search) from which the slice was extended by varying the biases of the foot and back swing motor neurons. One can see large flat expanses of zero fitness (black) and 0.125 fitness (gray area in upper right). It is these regions that the population occupies during phase I. Phase II begins when an individual discovers a greater than 0.125 point in centrally located "ridged" region of parameter space. (Note that a less than 0.125 individual on the ridged region will not have a reproductive advantage due to individual in the higher fitness 0.125 region.) The descendants of this breakthrough individual slowly ratchet (when using elitism) or climb their way up to the highest fitness areas.

6. Applications

Typically single leg stepper searches are run at a single mutation variance. In these searches, mutation variance is a balance between a high mutation variance and a low mutation variance. The high mutation variance does well in the first part of the search when the population is initially randomly scattered on the low fitness plateaus which make up the majority of parameter space. The low mutation variance does well exploring the smaller regions of higher fitness parameter space.

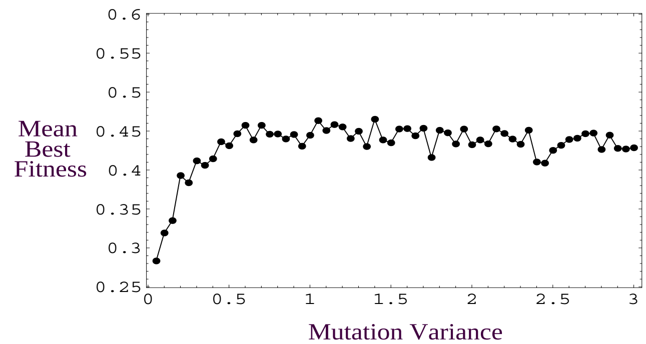

Figure 10: Averaged best (last generation) fitnesses over a range of mutation variances.

The current results can be applied towards explaining the findings of Mathayomchan (2002) whose examination of mean best fitness over a range of mutation variances was used to determined the nominal mutation variance for the single leg stepper problem. In these experiments the best fitness of each 250 generation search was recorded and averaged. The searches were performed on a variety of mutation variances. The mean best fitness was found to be very low at low mutation variances and increases steeply with increasing mutation variance until it levels off at approximately 0.6 (Figure 10). The current results show that the low mean best fitness at small mutation variances is due to the search terminating at generation 250 without having discovered an above 0.126 individual. Note that at high mutation variances the mean best fitness apparently remains unchanged. This may be explainable by the relatively more robust performance of phase II with respect to mutation variance.

If one is not restricted to only one mutation variance one can decrease the median number of generations from the beginning of search to end of search. With a single mutation variance of 0.5, the median generation for a complete search is approximately 400 generations. If two mutation variances are used with a switch from one to the other at breakthrough, then one can do better. Choosing 0.5 mutation variance until breakthrough and then 0.05 thereafter results in a median of approximately 300 generations for a savings of 100 generations.

Decreasing the mutation variance over time is reminscent to the strategies of simulated annealing in which a "temperature" analogous to mutation variance is changed using an annealing schedule. The annealing schedules usually decrease over the long term, letting solutions settle to local minima, often with punctuation with increases in temperature whose goal is to break the solution out of local minima. (Kirkpatrick et al. 1983, Salamon et al. 2002) In the case of the single leg stepper problem, the switch to lower mutation variance is made at a specific point in the search using problem specific knowledge.

It may be possible reduce the computing time required for the search by using the observation that after breakthrough occurs, most of the population is not needed for the continued improvement in the fitness of the individuals. Specifically, descendants of the breakthrough individual have a much higher probability of finding a new high fitness than a random individual and unrelated individuals may be safely ignored. All of the relatives of the breakthrough individual may not even be needed: If a child does not become the elite, it's lineage diffuses via mutation away from the region. With passing generations the children may eventually become equivalent to a unrelated individual. If this happens they can also be eliminated from the population. Thus it may be possible to speed the search by reducing the number of individuals in the population during phase II.

7. Conclusions

The layout of parameter/fitness space for the single leg stepper problem is such that one mutation variance does not explore all features equally well. A larger mutation variance does best in the initial search because high fitness regions are such a small proportion of parameter space that a random search is needed rather than a gradient following search. Once an above breakthrough fitness region is found a lower mutation variance does better because children are more likely to be created on nearby above breakthrough fitness regions.

These findings apply to many systems with complex dependencies among parameters. In these cases the intersection of parameter values which "work" is often very small. Instead, parameter space may be occupied by "background" fitnesses which are easily achievable by chance and supply no information as to the location of the higher fitness regions of interest. Under these conditions random search along with problem specific knowledge might be used to discover the higher fitness regions of interest followed by GA search to tune the initial solution.

Acknowledgments

This research has been supported in part by grant EIA-0130773 from the NSF and a NSF IGERT grant in neuromechanics at CWRU.

Thanks to the Beer Nuts for giving helpful suggestions on the revisions of this paper.

References

Ames, J.C. (2003) *Design methods for pattern generation circuits*, Master thesis, Department of Electrical Engineering and Computer Science, Case Western Reserve University.

Bäck, T. (1996) *Evolutionary Algorithms in Theory and Practice*, Oxford University Press

Beer, R.D. and Gallagher, J.C. (1992) Evolving dynamical neural networks for adaptive behavior. *Adaptive Behavior* **1**(1):91-122.

Beer, R.D. (1995) On the dynamics of small continuous-time recurrent neural networks. *Adaptive Behavior* **3** (4):471-511.

Beer, R.D., Chiel, H.J. and Gallagher, J.C. (1999) Evolution and analysis of model CPGs for walking II. General principles and individual variability. *J. Computational Neuroscience* **7**(2):119-147.

Chiel, H.J., Beer, R.D. and Gallagher, J.C. (1999) Evolution and analysis of model CPGs for walking I. Dynamical modules. *J. Computational Neuroscience* **7**(2):99-118.

Cliff, D., Harvey, I., and Husbands, P. (1993) Explorations in evolutionary robotics. *Adaptive Behavior* **2**(1):73-110.

Kirkpatrick, S., Gelatt, Jr., C. D., Vecchi, M.P. (1983) Optimization by Simulated Annealing, *Science* 220 (4598): 671-680.

van Nimwegen, E., Crutchfield, J.P., and Mitchell, M. (1997) Finite Populations Induce Metastability in Evolutionary Search. *Physics Letters A* **229**(2), 144-150.

van Nimwegen, E. (1999) *The Statistical Dynamics of Epochal Evolution*. PhD thesis, Bioinformatics Group, University of Utrecht.

Gao, Y., Qi, X., and Palmieri, F. (1998) Comments on theoretical analysis of evolutionary algorithms with an infinite population size in continuous space I. basic properties of selection and mutation. *IEEE Trans. Neural Netw.* **9**(2), 341-343.

Mathayomchan, B. and Beer, R.D. (2002) Center-crossing recurrent neural networks for the evolution of rhythmic behavior. *Neural Computation* **14**:2043-2051.

Nolfi, S., Floreano, D., (2000) *Evolutionary Robotics*, MIT Press.

Salamon, P., Sibani, P., Frost, R., (2002) *Facts, Conjectures, and Improvements for Simulated Annealing*, Society for Industrial & Applied Mathematics.