

“Optimal” Mutation Rates for Genetic Search

J. Cervantes
IIMAS

Universidad Nacional Autonoma de Mexico
Circuito Escolar, Mexico D.F. 04510
JorgeCervantesO@netscape.net

C. R. Stephens*

Instituto de Ciencias Nucleares
Universidad Nacional Autonoma de Mexico
A. Postal 70-543, México D.F. 04510
stephens@nucleares.unam.mx

ABSTRACT

Using a set of model landscapes we examine how different mutation rates affect different search metrics. We show that very universal heuristics, such as $1/N$ and the error threshold, can generally be improved upon if one has some qualitative information about the landscape. In particular, we show in the case of multiple optima (signals) how mutation affects which signal dominates and how passing between the dominance of one to another depends on the relative height and size of the peaks and their relative positions in the configuration space.

Categories and Subject Descriptors

I.2.m [Computing Methodologies]: Artificial Intelligence—*Miscellaneous*

General Terms

Performance

Keywords

Genetic Algorithms, Selection, Mutation rate, Effective Fitness, Error threshold

1. INTRODUCTION

There has always been a strong interest in Evolutionary Computation (EC) with respect to what are “optimal” parameter settings for a given class of Evolutionary Algorithm (EA) [1, 2]. In this paper we concentrate on the roles of mutation and selection, neglecting the effect of recombination, the resulting intuition also being applicable to simulated annealing type algorithms. There have been three recurring themes with respect to an “optimal” mutation rate: i) that a rate $\sim 1/N$ is preferred [5], where N is the string length;

*Part of this work was done while on sabbatical leave at the Department of Computer Science, University of Essex, Wivenhoe, CO4 5SQ.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

GECCO'06, July 8–12, 2006, Seattle, Washington, USA.
Copyright 2006 ACM 1-59593-186-4/06/0007 ...\$5.00.

ii) that a preferred rate should be dynamic [3, 4, 10, 9, 11]; and iii) that the error threshold offers guidance as to an optimal rate [6, 7, 8].

Of course, the desire has been to derive useful heuristics for setting mutation rates and, the more universal the better. The well known $1/N$ heuristic is one such. Obviously this is too universal in that it is independent of the fitness landscape. The error threshold does depend on the fitness landscape and has been argued to be associated with an optimal balance between exploration and exploitation. It also often scales with string length as $1/N$. Other work [9] has argued that an optimal mutation rate depends not only on the fitness landscape, but also on the actual population, and hence is dynamic. To have to go to such a fine-grained level is clearly not practically feasible. A possible solution is to consider a self-adaptive mutation rate [3, 9, 11]. This obviously sounds very attractive, especially in the context of dynamic fitness landscapes. However, this leaves *everything* in the hands of the algorithm. There is clearly still both theoretical and practical value in trying to develop an intuition for how mutation and selection interact in different circumstances.

In this paper we will pay more attention to trying to better understand, in a fairly generic fashion, how different fitness landscape features impinge on what is a useful mutation rate to use. In particular, we will show that heuristics, such as $1/N$ or the error threshold, are too universal. However, rather than conclude that it is a forlorn hope to try to find preferred mutation rates, other than tailor them to a very specific problem, we will show that there is much that can be done by thinking of how preferred mutation rates can depend on certain generic landscape features. We will make frequent use of the concept of effective fitness [12] to interpret our results, the effective fitness of a genotype I being defined as $f_I^{\text{eff}}(t) = (P_I(t+1)/P_I(t))\bar{f}(t)$ where $P_I(t)$ is the proportion of genotype I present at time t and which, with an evolution equation in order to specify the relation between $P_I(t+1)$ and $P_I(t)$, defines the effective fitness at t as a function of the population and genetic operators at t . The idea behind the effective fitness is that it measures the contribution of all genetic operators, not just selection, to the reproductive success of an individual.

2. METRICS AND LANDSCAPES

To investigate what are the most appropriate mutation rates we will consider the following metrics:

1. The asymptotic proportion of optimal strings as a function of the algorithm's parameters

2. The proportion of optimal strings as a function of time
3. The time (function evaluations) for the algorithm to first find the optimum
4. The probability that after being found the optimum is lost after a certain time

The first metric, 1, is a standard one associated with identifying the error threshold itself. The rest are different well-known measures of the quality of the search. As we have multiple criteria, it may be that an optimal mutation rate for one metric is highly suboptimal for another. For instance, minimizing the time to find the optimum may require a mutation rate that makes it very likely that it will be lost again very soon thereafter.

We will examine a range of different well-known “model” landscapes, taking the point of view that a principal task of search is to distinguish signal from noise by a suitable balance between exploration and exploitation. It is also important to differentiate between offline and online signals, the former being the signals of the fitness landscape itself, and the latter those signals that exist in the current state of the algorithm, e.g., in the current population. The landscapes we will consider fall into the following classes:

1. “Needle-in-a-haystack” - this is simply to make contact with known results in the context of a model where there is only one unique offline signal - the needle.
2. Degenerate needle-in-a-haystack - where for the needle phenotype there correspond more than one needle genotype. The motivation here is to see how such degeneracy affects the error threshold and “robustness”, and to understand how degeneracy can enhance both online and offline signals.
3. “Competing” needles - here we consider a non-degenerate optimal needle and a second degenerate suboptimal needle in order to see under what circumstances another suboptimal signal can help or hinder search as a function of the relative signal strengths and their relative location in the fitness landscape.
4. Multiple uncorrelated needles - The goal here is to see how the existence of many competing unrelated signals makes search more difficult. By generating the landscape by adding noise to the fitness function one can also investigate the universality of phenomena, such as the error threshold, that exist in the case of the single needle, to see if they are robust to (random) changes in the landscape.
5. Multiple correlated needles - landscapes where there are many needles that are strongly related. We will take as a concrete example a counting ones landscape.

3. LANDSCAPES WITH ONE UNIQUE SIGNAL - THE ERROR THRESHOLD

An interesting phenomena, seen in the NIAH landscape, is the existence of a critical mutation rate - the error threshold - beyond which selection is effectively “absent” [13]. This is obviously important for search, as it means that, effectively, the optimum in the landscape no longer provides a reliable

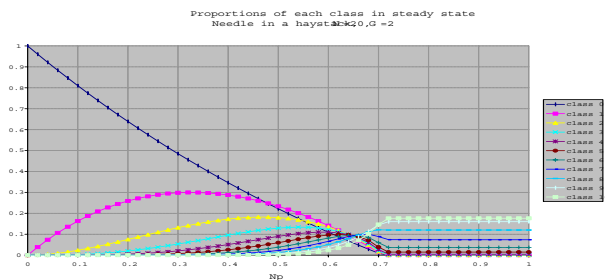


Figure 1: Graph of string proportions for NIAH for different Hamming classes versus pN for $N = 20$.

signal for the population to follow. To investigate this phenomena analytically we may integrate the infinite population dynamical equations without recombination, until an asymptotic steady state is reached. We may then graph the proportions of the different Hamming classes as a function of Np , a Hamming class being that set of genotypes that are a fixed Hamming distance from the optimal one. As an example, we take the fitness landscape to be such that the “needle” is the string of all zeros and has fitness 2, while the rest, the “hay”, have fitness 1. In Figure 1 we see such a graph for $N = 20$. With pure selection one would expect to see a preference for the needle versus the rest, but no relative preference for any non-needle versus another. As we can clearly see however, those strings which are closer in Hamming distance to the needle are favoured. This is because an optimal string more easily mutates into another string that is close by rather than one much further away. Thus, the strings that are closer in Hamming distance to the needle have a higher Effective fitness. The other notable feature of Figure 1 is the existence of a well defined mutation rate, p^* , at which the derivative of the string proportion curves appears to change “discontinuously”. Often it is said that a (second order) phase transition takes place. At this point - the error threshold - the string proportions are close to what they would be in a random population.

A theoretical value for the error threshold is posited by considering the value of p that leads to a fixed point for the needle proportion *neglecting* back mutations, i.e., that other strings can mutate into the optimal string. The corresponding expression is

$$p^* = 1 - \exp(-\ln G/N) \rightarrow \ln G/N \quad (1)$$

where $G = f_1/f_0$ is the fitness gain, f_1 being the needle fitness and f_0 that of the hay. A more phenomenological definition is that point where the second derivative of the curve of the needle proportion is a maximum, as this represents the point around which there is a “maximal” change in behaviour. In Figure 2 we see how this point changes as a function of N , for $3 \leq N \leq 30$, by plotting Np^* against N . The straight line represents the approximation given by taking the large N limit, while the (pink) line with squares represents the first equality in equation (1). Even for quite short string lengths the phenomenological (triangles/yellow) and theoretical results are quite close. The deviations for $N < 7$ arise principally from the breakdown of the linear approximation to the exponential in (1), though one can also see the effect of back mutations as the “experimental” curve for $N > 7$ lies above both theoretical approximations as back mutations make the master sequence somewhat more robust. This definition of the error threshold is valid for

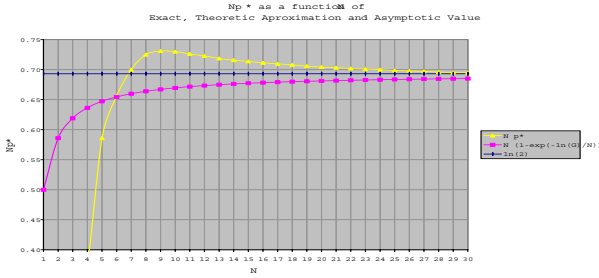


Figure 2: Graph of p^*N vs N comparing phenomenological and theoretical definitions of the error threshold.

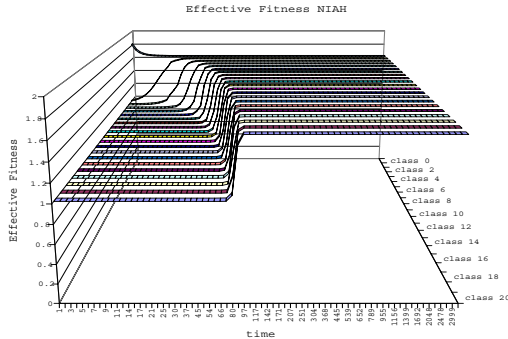


Figure 3: Graph of Effective fitness versus time for the different Hamming classes below the error threshold

both finite and infinite population models, although, due to sampling errors, in order to be able identify the point for a small population it would be necessary to have a large number of runs in order to obtain better statistics.

It is natural to ask: What are the appropriate effective degrees of freedom of this model? In the absence of mutation, the two phenotypes that correspond to the needle and the hay give the most appropriate description. However, when mutation is present then the appropriate description appears to be in terms of the N Hamming classes. However, as shown by Eigen [13], the most natural description is in terms of a “quasi-species”, which is a “cloud” of genotypes centered around the needle.

In Figures 3 and 4 we see graphs of Effective fitness as a function of time for the different Hamming classes, i.e., the Effective fitness landscape, for $N = 20$ with $p = 0.01$ (below the error threshold) in Figure 3, and $p = 0.04$ (above the threshold) in Figure 4. The error threshold in this case is at $p^* = \ln 2/20 \sim 0.0347$. Below the error threshold we can see that the Effective fitness of the needle decreases in time, while that of the hay increases. Notice that the Effective fitness of the closest Hamming neighbours of the needle increases faster than that of the farther neighbours. In this sense the “wave” that ripples out from the needle is describing the evolution of the quasi-species. Asymptotically, of course, the Effective fitness landscape becomes flat, indicating that an asymptotic steady state has been reached. Above the threshold, we see that the Effective fitness landscape is essentially flat throughout the entire evolution, with only an initial transient wherein the Effective fitness of the needle decreases rapidly to that of the hay. This is a sure

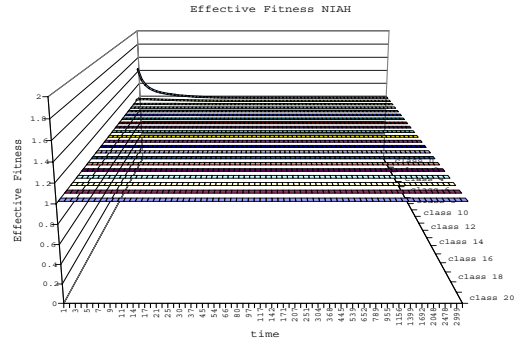


Figure 4: Graph of Effective fitness versus time for the different Hamming classes above the error threshold

indicator of the absence of selective evolution, in spite of the fact that the fitness landscape itself is certainly not flat.

So, what are the implications for search? Firstly, at the error threshold search is essentially random. Hence, due to the continuity of the problem as a function of p , this implies that, close to p^* , the search will also be almost random. For instance, for $p = 0.03$, which corresponds to a 13% difference relative to p^* , the Effective fitness of the needle quickly reduces to a value of 1.1, which is small compared to its fitness value of 2. Thus, we see that the competitive advantage of the needle is small in the vicinity of the error threshold.

Our conclusions above came from the infinite population model. In Figure 5 we see the same graph as in Figure 1, but now for a finite population and considering only the needle frequency. A 1000 runs of a plain vanilla GA were considered for each combination of population size and mutation rate. For each run, after a certain time, chosen to be in the asymptotic regime, the proportion of needles present in the population was measured. The curves rapidly asymptote to

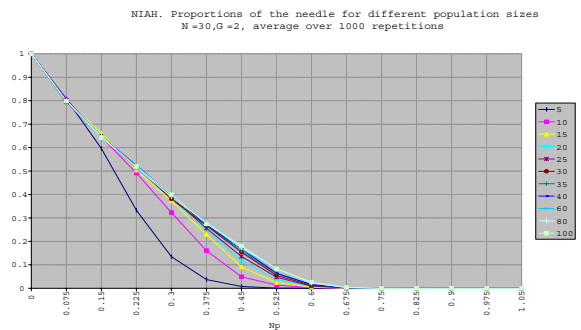


Figure 5: Graph of needle frequency as a function of Np for different population sizes.

a universal curve as a function of population size. As can be appreciated, the chief effect of a finite population is to effectively lower the error threshold, so that the system looks even more “noisy” and it is even more difficult to assure the existence of a fixed proportion of needles. Considering now our other performance metrics: we first consider the computational effort in terms of the number of fitness evaluations needed to find the needle. In Figure 6 we see how this effort varies as a function of population size and Np , for $N = 8$ and $G = 2$ in a NIAH landscape. Note that the computational effort is a minimum when $Np = 4$ which,

given that $N = 8$, shows that the optimal mutation rate is $p = 0.5$, i.e. random search. This is due to the fact that, without having the needle in the initial population, there is absolutely no online signal to direct the search in a particular direction. Note also that this optimal mutation rate is independent of population size. Thus, there is no difference between a population size of 100 over 10 generations and a population size of 1 over 1000 generations. In fact, population size plays effectively no role until low mutation rates, $< 1/N$ are reached. For $Np > 1$, essentially, there is no preservation of any string (at $Np = 1$, on average one bit changes per string), hence there is no exploitation and only exploration. This is not the case however, for mutation rates that are $< 1/N$, where a larger population size is preferable, at least if the population is chosen *randomly*. Such a random choice means that the corresponding population was generated in the same way as if a $p = 0.5$ mutation rate had been used. As random search is more efficient in this landscape, then the larger the population the more the search is in exploration rather than exploitation mode.

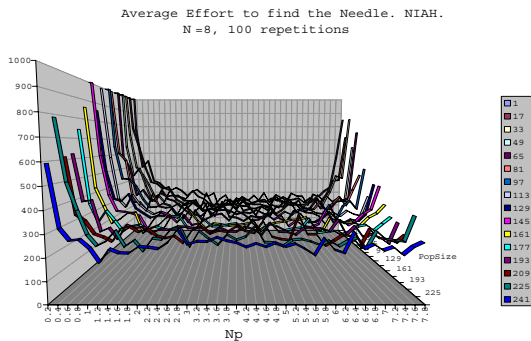


Figure 6: Graph of computational effort to find the needle versus mutation rate and population size.

Of course, this result is intuitively obvious. We start with a landscape that has one single isolated signal and an initial population where that signal is not present. The population evolves therefore, as if it were in a completely flat landscape until it finds the optimum. In this circumstance, the optimal mutation rate is 0.5. Note that the error threshold at $Np^* = 0.69$, as well as the familiar $Np = 1$, require a computational effort which is very similar to that for $p = 0.5$. Once again, the reason is clear, because with these mutation rates there are few individuals that remain unmutated. Note that the computational effort is independent of the population size for $p = 0.5$, and quite insensitive for $pN > 1$, while for $pN < 1$ the computational effort is smaller as a function of population size. This is because the random generation of the initial population is equivalent to using a mutation rate $p = 0.5$ and therefore leads to more diversity.

Turning now to the robustness of the search, in Figure 7 we see how robust the optimum is once it has been found for $G = 2$, $N = 30$. The plot shows in what percentage of runs, as a function of population size and mutation rate (taken over 1000 repeats) the needle is present for 50 generations after being found for the first time, where here we start with one needle and “random” string proportions for the other strings. Note that, above the error threshold at $p^* = 0.69/30 = 0.023$, there is little chance of the needle surviving. More than 97% of runs show a loss of the op-

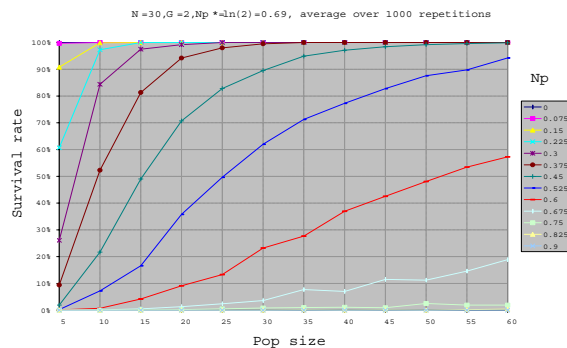


Figure 7: Graph of survival rate of the needle, mutation rate and population size.

timum irrespective of the population size, at least up to a population size of 60. Note also that, for $p < p^*/2$, even relatively small population sizes lead to high survival rates for the optimum. One can also think in terms of what population size is needed to maintain a survival rate of 80-90% for instance. In this case, a 20% increase in the mutation rate from $Np = 0.45$ to $Np = 0.53$ requires twice the population size in order to maintain the same survival rate.

The graphs certainly show how sensitive the “survival of the fittest” is to the population size and mutation rate. Note particularly, that a mutation rate of $1/N = 0.033$ would correspond to a point above the error threshold, and therefore to a very, very small survival rate for the optimum. We see then the antagonistic nature of the two requirements of minimising the time to find the optimum and maximising its survival rate, the two optimal mutation rates in this landscape being $p = 0.5$ (exploration) and $p = 0$ (exploitation) respectively. These results can also be understood qualitatively, if not quantitatively, from the infinite population model: Starting with random proportions, but *without* the needle, the proportion of needles in the next generation would be maximised by $p = 0.5$. Similarly, after finding the needle the survival rate is maximised by taking $p = 0$.

So what is an optimal mutation rate? There are clearly several notions of “optimal” if one decides to have a multi-objective criterion that includes, for example, both computational effort and survival rate. This of course relates back to the fundamental dichotomy of all search - exploration versus exploitation. We also see that heuristics, such as $1/N$ or the error threshold, can lead to a reasonably efficient search but are very suboptimal in terms of other criteria, such as robustness or online population fitness.

4. LANDSCAPES WITH ONE UNIQUE DEGENERATE SIGNAL

Having considered a landscape signal that consists of only one genotype, we now consider signals that consist of several degenerate genotypes. One motivation for this comes from requiring a better understanding of “robustness” and the role of neutral networks. For a given value of needle fitness one would expect the signal to be stronger the more needles that were involved. We will therefore first consider what happens on a landscape with several adjacent needles of equal height, where by adjacent we mean in terms of Hamming distance. This is achieved by choosing a needle position, and then placing the rest as close as possible in Hamming distance to that needle so as to form a neutral

network. In Figure 8 we show a graph of how p^*N varies as a function of the percentage of the landscape that consists of needles, i.e. the relative size of the neutral network, for $G = 2$ and the cases $N = 20$ and $N = 30$. As would be ex-

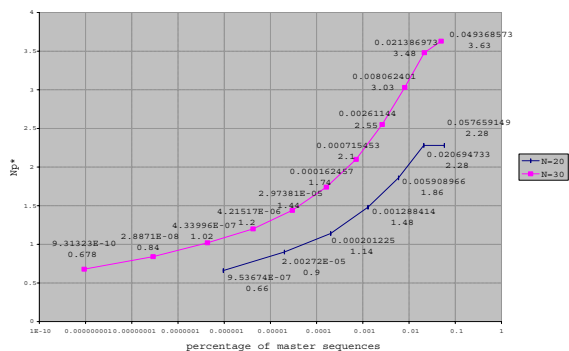


Figure 8: Graph of p^* against “signal strength” for a “degenerate” NIAH landscape

pected, increasing the number of optima increases the error threshold as the optimum is now more robust. Interestingly, we see that the relationship in terms of p^* is “universal”, in that the two curves essentially coincide, the offset between them being due to the presence of N . Thus, no matter the string length the same change of “robustness”, as proxied by the percentage of needles, gives rise to the same change in the error threshold. In this sense one can think of the percentage of needles as a good measure of signal strength. Note that above about 5% redundancy the error threshold disappears, in that there is no longer a peak in the derivative. In Figure 9 we consider the case of $N = 20$, $G = 2$, where any string with either 20 or 19 zeros is optimal. We can see that the Effective fitness of the peak associated with 20 zeros is higher than those peaks associated with 19 zeros. This is due to the fact that one-bit mutations of the all zeros string leaves it as an optimal string with 19 zeros without any decrease in fitness. For optimal strings with 19 zeros though, due to the fact that one-bit mutations can take them to non-optimal strings with 18 zeros, their Effective fitness is reduced as they are less robust.

5. LANDSCAPES WITH MORE THAN ONE SIGNAL

We now consider the case when there is more than one

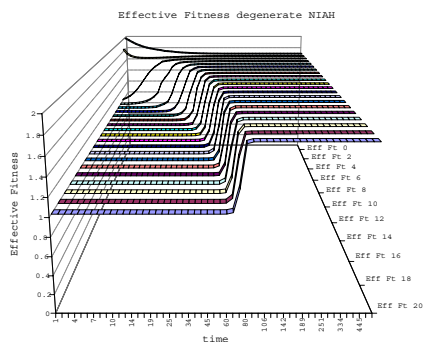


Figure 9: Graph of Effective fitness as a function of time for a degenerate NIAH landscape.

signal, in particular, a landscape where there is a peak corresponding to a single optimal genotype and a second sub-optimal peak that consists of more than one degenerate genotype. The quality of the search will depend on the strengths of the two signals from the optimum and sub-optimum and on the spatial relationship between them. The strength roughly depends on the height and width of the two signals. Here, we first consider the case where the suboptimal genotypes are as distant as possible from the optimal, i.e. clustered around the antipodal point of the hypercube to where the optimum is located. In this landscape an interesting phenomenon occurs in the asymptotic behaviour relative to the NIAH case. As can be deduced in the infinite population model, for low mutation rates, there is a preference for the optimum. However, at a certain mutation rate, p_f , (not the error threshold) there is a sudden change as a function of p , such that above this value the suboptimum is preferred. Finally, at an even higher mutation rate, one reaches the error threshold beyond which there is no preference for any string. In Figure 10 we see for $N = 20$ the temporal development, as found by integrating the equations of the infinite population model for a mutation rate below p_f , of the different Hamming classes relative to the optimum. The fitness of the optimum is 2, that of the suboptimum is 1.95, while the rest have fitness 1. The width of each colour band represents the proportion present of the corresponding Hamming class. The wider bands in the middle stems from the fact that we used a random initial population and therefore the proportion of Hamming class n is determined by a binomial factor ${}^N C_n$. Note that

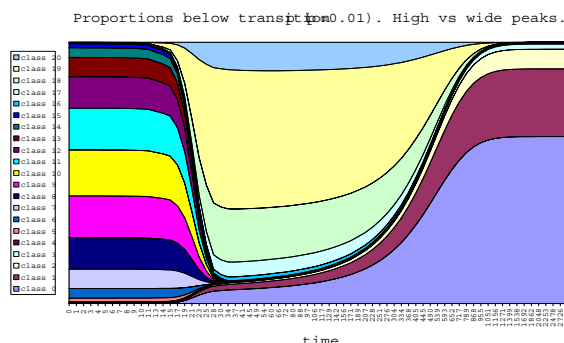


Figure 10: Graph of Hamming class proportions as a function of time in two-peak landscape ($p < p_f$).

although there is a substantial transient, wherein the sub-optimum is represented by a much higher proportion than the optimum, eventually the latter dominates. In Figure 11 we see the corresponding behaviour when $p > p_f$. Now we see that the suboptimum (Hamming class 20 and class 19) dominates the evolution. The implication for search is clear: p_f gives a threshold above which the Effective signal strength of the suboptimum is higher than the optimum. What is more, p_f is much less than the corresponding error threshold. In this case search in the vicinity of the error threshold would be completely inadequate. One can think of these graphs as representing a competition between two quasi-species: one associated with the optimum and another associated with the sub-optimum. Below p_f , initially the suboptimal quasi-species dominates only to lose out in the long run to the optimal quasi-species. On the contrary, for

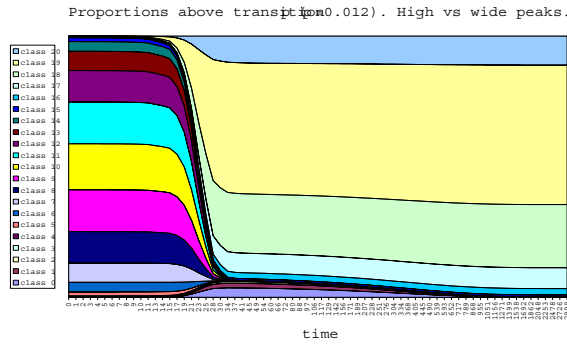


Figure 11: Graph of Hamming class proportions as a function of time in two-peak landscape ($p > p_f$).

$p > p_f$, the suboptimal dominates throughout.

Having seen what happens when there is a competing signal to the optimum that is maximally distant from it, we now ask how the signal competition changes as a function of where the suboptimum is placed. In Figure 12 we consider a landscape where there is a needle of fitness 2 and a suboptimal signal consisting of two adjacent genotypes of fitness 1.9. As we can see the mutation rate above which the suboptimal peak dominates decreases as a function of the Hamming distance between the peaks. Thus, the fur-

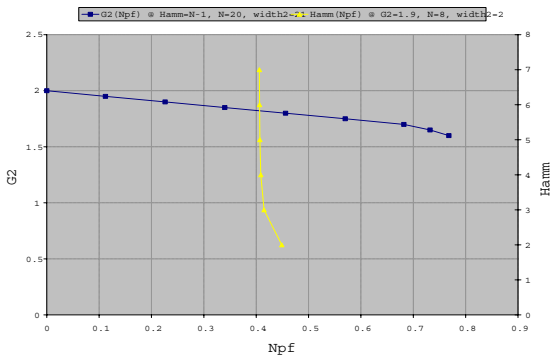


Figure 12: Graph of the transition mutation rate p_f as a function of Hamming distance (circles) and fitness difference (squares) between optimal and suboptimal peaks for $N = 8$ and $N = 20$ respectively.

ther apart are the two signals the more “deceptive” is the suboptimum in how it affects the search. An even more striking effect can also be seen in Figure 12, where we plot p_f as a function of the fitness difference between the optimum and suboptimum. Given that $p_f \rightarrow 0$ as the fitness differential decreases we see clear evidence of the potential inadequacy of the notion that search in the vicinity of the error threshold is optimal, or even preferred.

In terms of our other search metrics, in Figure 13 we consider the average effort to find the needle in a landscape with a single optimum and a single suboptimum of fitness 2 situated at the antipodal point to the optimum. As $N = 8$, a mutation rate of $1/2$ corresponds to $Np = 4$. Comparing with Figure 6 we can see that the existence of the subop-

timium makes the search more difficult for mutation rates $< 1/2$. On the contrary, for very high mutation rates we see that the search is more efficient. In fact, the best mutation rate is about $1 - 1/N$. It is clear why a very high mutation rate is useful - because the optimum and the suboptimum are anti-correlated, once the suboptimum is located, the search will be directed by mutation towards the optimum. The fact that the optimal mutation rate isn’t too close to one is because the search has to first find the suboptimum. Thus, this mutation rate is optimal because it allows initially for a better exploration of the suboptimum and then a better exploitation of the information provided by the suboptimum. Also, here, unlike the single signal case, a larger population helps. It is interesting to see such interesting effects result from the existence of only one single deceptive suboptimum.

Average Effort to find the Needle. NIAH with opposite subneedle. $N = 8$, 100 repetitions

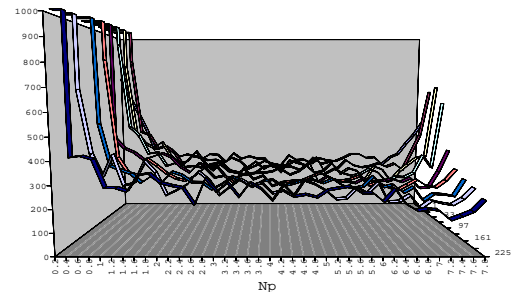


Figure 13: Graph of average effort to find the needle versus Np and population size for $N = 8$ in a landscape with one subneedle at antipode of the needle.

In the above we can see the existence of a preferred mutation rate that is very far from the usual $1/N$ or error threshold heuristics. This mutation rate results from there being an anti-correlation between the optima and suboptima, i.e. the suboptimum is deceptive in that it leads the search away from the region of interest, at least for small mutation rates. If there exists a correlation rather than an anti-correlation however, one might expect to see cooperation rather than deception. To see this, consider Figure 14 which is a graph of the frequency over a thousand runs with which the optimum is found in a system with an optimal needle and a suboptimal one, but now located at a fixed Hamming distance from the optimal. In this case, $N = 30$ and the distance between the optima is 3. The initial condition here was to have all the population at the suboptimum. This is to mimic what happens in a lot of search problems, where there is a form of punctuated evolution, the population forming a type of quasi-species around a suboptimum and searches from there. The experiments were run for a maximum of 5000 generations. As we can see, once again, there is a preferred mutation rate, $Np = 0.375$, that maximises the success rate of the search and that is quite different from the usual heuristics at 0.69 (error threshold) and 1 ($1/N$ heuristic).

Having considered two signals and shown how an optimal mutation rate depends not only on the strengths of the two signals, but also on their locations and the state of the population, we now consider a multi-peak landscape - a “noisy” NIAH landscape, where a random number chosen from a

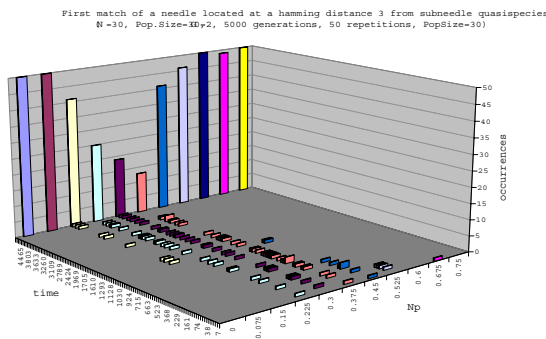


Figure 14: Graph of success frequency as a function of p and t in a two needle landscape with $N = 30$.

uniform distribution between $[-k, k]$, where k is the maximum noise level, was added to the fitness of the hay. This landscape would be very similar to that of a Kauffman NK landscape with $K = N - 1$, the only difference being that one single genotype has a fitness which is not chosen randomly. Essentially, one may think of this landscape as generically representing the common class of very rugged landscapes. As the noise is tunable we also chose this model to exhibit the universality of the error threshold phenomenon. In Figure 15, we see a graph of the needle frequency as a function of p and the amplitude of the noise distribution for $N = 8$. In this case, each integration of the exact dynamical equation, for a given p is done with the same fixed, random seed. In this case we are seeing how the needle proportion varies according to the strength of the noise for a fixed distribution of “deceptive” peaks, rather than seeing how it depends on how the particular location of the peaks changes. We can clearly see that the higher the noise level the lower the error threshold. The error threshold for zero noise, i.e., the NIAH, is $p^* = 0.09$, while for a noise level of 0.9 it is less than half that. Thus, to have a fixed proportion of optimal strings in the asymptotic population one requires quite different mutation rates in a very rugged landscape versus a smooth one with no fitness gradient. So we see that a single

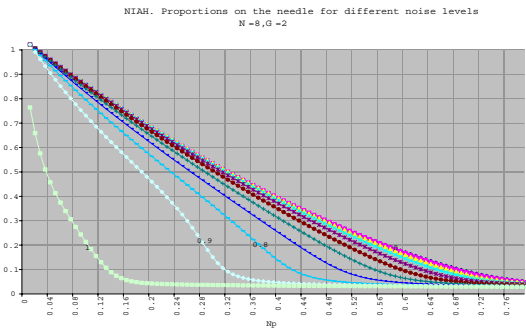


Figure 15: Graph of steady state needle proportion as a function of mutation rate for a noisy NIAH landscape with variable (left) and fixed (right) random seed.

error threshold exists not only for the NIAH, but also for any model that is “similar” to it. Of course, defining “similar” is not a simple matter. Here, we are restricting to a class of landscapes we may term - noisy NIAH. Basically, any

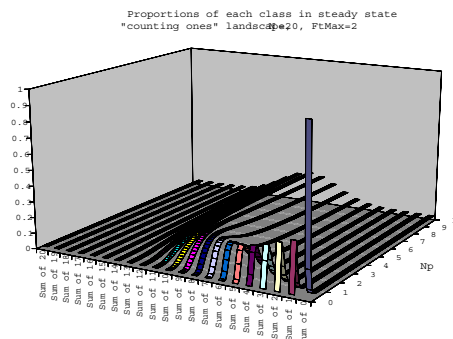


Figure 16: Graph of string proportions for counting ones for different Hamming classes versus pN for $N = 20$.

landscape where there is no correlation between the other regions of the landscape will lead to the same phenomenon. It is only when the fluctuations in the noise can cause the appearance of a stronger signal that the error threshold will disappear. This will generally occur when a new, higher needle is generated.

Finally, we consider what happens in the case of counting ones. Due to space considerations we only consider the analog of Figure 1 as seen in Figure 16 below. Unlike the NIAH in this case there is no well defined point at which the behaviour of the different curves changes simultaneously. Rather, taking our phenomenological definition of the error threshold, we see that each Hamming class seems to exhibit its own threshold with the thresholds increasing the larger the Hamming distance. This can be understood intuitively from our previous results taking the curve for the optimum as a baseline. So why does the curve for Hamming distance one exhibit a higher error threshold? Think of the set of strings at Hamming distance one as forming a degenerate suboptimum. As their fitness is lower than that of the optimum one would expect the error threshold to be less. However, as there are 19 corresponding states and not just one, one would expect them to exhibit much more robustness, and that is the reason why the error threshold is higher. The same arguments apply for the other Hamming classes.

6. CONCLUSIONS

In choosing a suitable mutation rate for an EA, there are three standard approaches: find an optimal rate experimentally; use a standard heuristic; or use a self-adaptive algorithm. Each has its advantages and disadvantages. Here, we have tried to see to what extent the standard heuristics give a good estimate of an optimal rate, and where there are limitations to try to go beyond them. As has been noted before, what is optimal depends very much on what one requires of one’s search algorithm - find the optimum with a minimum of effort, maintain a high proportion of optimal strings etc. We considered several performance metrics in the context of several model fitness landscapes, the idea behind the latter being to model the competition between different online and offline signals that affect the performance of any search algorithm. In particular, we considered what happens in the case of a unique signal, a degenerate signal, two competing signals and multiple competing signals.

We first analysed the concept of the error threshold, giving a phenomenological definition that was in good agreement

with approximate theoretical predictions. We saw in the context of the NIAH how the error threshold increases with the fitness differential between the needle and the hay, and decreases as a function of decreasing population size and/or string length. It was seen that search in the vicinity of the error threshold was quite close to random, as was evidenced by the fact that the Effective fitness of the needle was very close to that of the hay there. In this particular landscape the separation between exploration and exploitation is total - optimal search requiring a mutation rate of 0.5 until the optimum is found then zero afterwards in order to preserve the optimum in the population. Thus, the standard heuristics work well in this landscape in terms of the exploration phase, but not the exploitation phase.

We then considered a degenerate NIAH showing a universal relation between the error threshold and the fraction of optimal configurations in the fitness landscape. We showed that although the optimal strings were equally fit there were differences in their Effective fitness, strings further away from non-optimal strings being more robust. We also saw that, beyond a redundancy of about 5%, the error threshold disappeared completely.

Turning then to landscapes with two competing signals, we considered first an optimal signal consisting of one unique string and a second suboptimal one consisting of two adjacent but less fit strings. We showed that there exists a particular mutation rate, p_f , above which the suboptimal string is preferred asymptotically, leading to higher proportions. This mutation rate is generally less than the error threshold and hence is of greater relevance. Even below it there is a substantial part of the evolution wherein the suboptimum is preferred. We saw that p_f decreases as the Hamming distance between the optimum and suboptimum increases. We also showed that it decreases quite strongly as a function of the fitness differential between the optima and suboptima. We showed then that an optimal mutation rate for minimising computational effort was close to one $\sim (1 - 1/N)$. We then demonstrated that for an optimum and suboptimum quite close together, where the initial condition is to have all the population in the suboptimal state, that there existed an optimal mutation rate for maximising the success rate of the algorithm and that it was substantially less than $1/N$ or the error threshold. We then turned to a multi-signal landscape showing that when the suboptima are uncorrelated that the error threshold for the optimum is substantially lower. Finally, taking as an example of a strongly correlated landscape counting ones, we showed that each Hamming class relative to the optimum exhibits its own error threshold, there being an increase as a function of the Hamming distance from the optimum.

So, generally, the $1/N$ and error threshold heuristics are too universal. Canonically, they are too high for search in landscapes with multiple peaks. Uncorrelated landscapes require a higher mutation rate than correlated ones, while landscapes with anti-correlation, i.e., with “deceptive” peaks, require even higher rates. Of course, one can argue that a complicated landscape can exhibit all these features. Very often though, a punctuated evolution takes place later on in the search process where an entire population moves quickly from a region of one type to another. In this sense we believe that a two signal model can tell you a lot about how selection and mutation interact in the search process. Generally, a mutation rate well below the $1/N$ and error thresholds is

commendable, especially in the case where there are peaks that are quite close in fitness value.

Acknowledgments

Part of this work was carried out while CRS was on sabbatical leave at the University of Essex. CRS wishes to thank the EPSRC for support (grant number GR/T2461/01), as well as DGAPA, UNAM, Conacyt project 41221-F and the UNAM macroproyecto “Tecnologías para la Universidad de la Información y la Computación”. JC thanks the PAEP, UNAM for support to visit the University of Essex and CONACyT for a doctoral fellowship.

7. REFERENCES

- [1] J.J. Grefenstette, Optimisation of Control Parameters for Genetic Algorithms, *IEEE Trans SMC*, 16(1), 122-128 (1986).
- [2] T. Bäck, Optimal Mutation Rates in Genetic Search. In *Proceedings of ICGA 5*, ed. S. Forrest, 2-8, Morgan Kaufmann (1993).
- [3] T. Bäck, Self-adaptation in Genetic Algorithms, In *Proceedings of 1st European Conference on Artificial Life*, 263-271 MIT Press (1991).
- [4] J. Hesser and R. Männer, Towards an Optimal Mutation Probability for Genetic Algorithms, *LNCS 496*, ed. H.P. Schwefel, Springer Verlag (1991).
- [5] H. Mühlenbein, How Genetic Algorithms Really Work: Mutation and Hill Climbing, *PPSN II*, ed.s B. Männer and R. Manderick, 15-25, North Holland (1992).
- [6] G. Ochoa, I. Harvey and H. Buxton, Error Thresholds and their Relation to Optimal Mutation Rates. In D. Floreano, J-D. Nicoud, and F. Mondada (Eds.) *ECAL'99, Lecture Notes in Artificial Intelligence 1674*, pp 54-63, Springer-Verlag, Berlin (1999).
- [7] G. Ochoa, Setting the Mutation Rate: Scope and Limitations of the $1/L$ Heuristics, *GECCO-2002*, pp 315-322, Morgan Kaufmann, San Francisco, CA (2002).
- [8] G. Ochoa, Error Thresholds in Genetic Algorithms. *Evolutionary Computation*, MIT Press (in press).
- [9] C.R. Stephens, I. Garcia Olmedo, J. Mora Vargas and H. Waelbroeck, Self-Adaptation in Evolving Systems, *Artificial Life*, Vol. 4, Issue 2, 183-201 (1998).
- [10] T.C. Fogarty, Varying the Probability of Mutation in the Genetic Algorithm, *ICGA 3*, ed. J.D. Schaffer, 104-109, Morgan-Kaufmann (1989).
- [11] J.E. Smith and T.C. Fogarty, Self Adaptation of Mutation Rates in a Steady State Genetic Algorithm, *Proceedings of CEC*, 318 - 323, IEEE Press (1996).
- [12] C. R. Stephens and J. Mora Vargas, Effective Fitness as an Alternative Paradigm for Evolutionary Computation I: General Formalism, *Genetic Programming and Evolvable Machines Volume 1*, 363-378 (2000).
- [13] M. Eigen and P. Schuster, *The Hypercycle: A Principal of Natural Self-Organization*, Springer-Verlag (1979).