

Netcrawling - Optimal Evolutionary Search with Neutral Networks

Lionel Barnett (lionelb@cogs.susx.ac.uk)

Centre for the Study of Evolution
Centre for Computational Neuroscience and Robotics
School of Cognitive and Computing Sciences
University of Sussex
Brighton BN1 9QH, UK

March 14, 2001

Abstract- Several studies have demonstrated that in the presence of a high degree of *selective neutrality*, in particular on fitness landscapes featuring *neutral networks*, evolution is qualitatively different from that on the more common model of rugged/correlated fitness landscapes often (implicitly) assumed by GA researchers. We characterise evolutionary dynamics on fitness landscapes with neutral networks and argue that, if a certain correlation-like statistical property holds, the most efficient strategy for evolutionary search is not population-based, but rather a population-of-one *netcrawler* - a variety of hill-climber. We derive quantitative estimates for expected waiting times to discovery of fitter genotypes and discuss implications for evolutionary algorithm design, including a proposal for an adaptive variant of the netcrawler.

1 Introduction

In the GA community at large there is a widely-held perception of fitness landscapes as being characteristically rugged or semi-correlated, a perception that has been reinforced by the study of “toy” problems and highly artificial multi-peaked test functions. This, coupled with historical factors within the development of GA’s as a search technique (such as the Building Block Hypothesis), have led to what might be characterised as the “Big Bang” view of evolutionary search (see Section 2.1). At Sussex University there is a tradition of the application of evolutionary techniques to decidedly real-world problems, such as on-chip hardware evolution [12, 27, 21, 29] and evolutionary robotics [4, 11, 15, 16, 26], from which has emerged a markedly different view of the nature of artificial evolutionary fitness landscapes and their concomitant evolutionary dynamics. In particular the presence and significance of *neutral evolution* has come under the research spotlight.

The major impetus for this research has come from evolutionary biology; ironically, GA research has traditionally remained somewhat isolated from its biologically-inspired origins. The work of the population biologist Motoo Kimura on

selectively neutral mutation [19, 5] initiated a debate amongst biologists which continues to this day. More recently, research into the structure of RNA secondary structure folding landscapes [8, 10, 14, 24] led to the concept of *neutral networks*. These are connected networks of genotypes which map to the same phenotype, where two genotypes are “connected” if they differ by one (or possibly a few) point mutations. It is found that the dynamics of populations of genotypes evolving on fitness landscapes featuring neutral networks differ qualitatively from population dynamics on more conventional rugged/correlated landscapes as might be encountered in either the biological or artificial evolution literature. Although the structures of fitness landscapes arising in the context of artificial evolution are ill-understood, a cursory examination of the types of genotype to phenotype mappings deployed in many real-world applications of artificial evolution (e.g. evolution of neural network controllers in robotics, on-chip electronic circuit evolution, etc.) suggests that we might expect a substantial degree of neutral mutation. Work here at Sussex suggests that this neutrality may well take the form of neutral networks as envisaged by RNA researchers.

In this paper we argue that, given large-scale neutrality, the traditional view of evolutionary dynamics may be largely irrelevant to the GA practitioner concerned with real-world problems. In Section 2 we characterise evolutionary search on fitness landscapes featuring neutral networks and address what is for the GA practitioner perhaps the most pertinent aspect of evolutionary search: how long can we expect to wait to see improvements in fitness and how should we design and tune our search algorithms so as to effect the most efficient search? In Section 3 these issues are analysed quantitatively for a class of fitness landscapes with neutral networks satisfying a statistical property that we term ϵ -*correlation*. We show that there is an optimal mutation scheme for such landscapes and conjecture that there is too an optimal evolutionary search algorithm, the *netcrawler*. Section 4 presents a summary of results.

2 Evolutionary Dynamics on Neutral Networks

2.1 Overview

Fig. 2.1 captures many of the salient features of evolutionary dynamics on a fitness landscape featuring neutral networks¹ as identified in [14, 13, 30, 2, 3]:

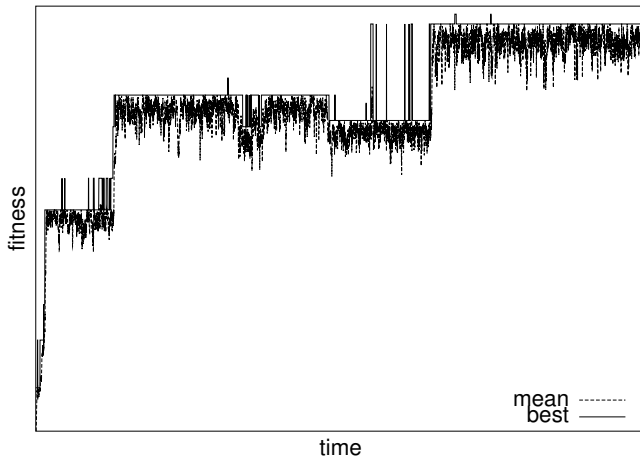


Figure 2.1: Typical evolutionary dynamics on a fitness landscape featuring neutral networks.

In summary:

- Evolution proceeds by *fitness epochs*, during which the mean fitness of the population persists close to the fitness of the fittest genotype(s) currently represented in the population.
- Transitions to higher fitness epochs are preceded by the discovery of higher fitness genotype(s) than currently reside in the population.
- Transitions may be down as well as up; the current fittest genotype(s) may be lost.
- The discovery of a higher fitness genotype does not necessarily initiate a new epoch; the new genotype may be quickly lost before it can become established in the population. This may be repeated several times.
- If a higher fitness genotype *does* initiate a fitness epoch, there is a transition period, brief compared to a typical epoch duration², during which the population mean fitness climbs to the new epoch level.

Through the work of various researchers a consistent explanation has emerged for the features presented above:

- During a fitness epoch the population is localised in sequence space, somewhat like a classical *quasi-species* [7]. The fittest genotypes reside on a neutral network, along which they *diffuse* [14, 6], until either...
 - ... a *portal* [30] to a higher fitness neutral network is

discovered, or...

- ... all genotypes on the current highest neutral network are lost due to sampling noise - this phenomenon is related to the concept of the *error threshold* [23].
- If a higher fitness genotype is discovered it will survive and drift to fixation with a certain probability [5].
- During the transient period when a higher fitness genotype is fixating, the population becomes strongly converged genetically (this phenomenon is variously known in the literature as “hitch-hiking” or the “founder effect”), as the higher fitness portal genotype and its selectively neutral mutants are strongly selected at the expense of the old population.

Now a more “traditional” view might impute a somewhat different interpretation to Fig. 2.1. It might be assumed that epochs correspond to episodes during which the population is entrapped in the vicinity of a local fitness optimum, while transitions to higher fitness levels signify discovery of higher local peaks. Broadly, the traditional view might be characterised thus: recombination assembles fitness-enhancing “building blocks” present in the population into higher fitness genotypes (the Building Block Hypothesis); mutation is merely a “background operator” to prevent total loss of genetic diversity. This process continues as long as there is sufficient genetic diversity in the population for recombination to work with. Once genetic diversity has waned (inevitably so, due to the combined effects of selection pressure and finite-population stochastic sampling) the population is deemed “converged” and no further fitness improvements are likely.

Thus it is deemed necessary to initiate the GA with a (usually large) randomly generated population - that is, with a “Big Bang” of genetic diversity for recombination to work upon. This perception goes some way to explaining the obsession of much GA research with “premature convergence” and the multitudinous schemes prevalent in the literature for the avoidance thereof. In this author’s view there are several serious flaws to this picture:

- There is scant evidence that real-world artificial evolution fitness landscapes conform to the rugged (non-neutral) picture often implicitly assumed.
- It is not clear whether “building blocks” in the conventionally understood sense will necessarily exist.
- Even if there are building blocks, serious doubt must be cast on the Building Block Hypothesis - see for instance [9]. Basically, due to the comparatively brief takeover period of any genotype with a fitness advantage (see Fig. 2.1) and the effects of hitch-hiking, multiple building blocks are never likely to be simultaneously present in a population.

¹The landscape in Fig. 2.1 is an NKp landscape [2, 3], the GA standard fitness-proportional roulette-wheel selection, with a fixed population size.

²Indeed so brief as to be virtually indiscernible on the time-scale of Fig. 2.1.

These points lead us to question the role of recombination. For if recombination is not assisting us by *assembling* building blocks what is its purpose? Might it help us *find* building blocks? In this author’s opinion there is no good reason to think so; at best recombination will function as a *macro-mutation* operator - a “leap in the dark” across the fitness landscape. Indeed, it might be remarked, in *biological* evolution the role of recombination is almost certainly *not* the assembly of putative building blocks [22], although other mechanisms such as *error repair* [20] are considered plausible.

This is not to imply that there is *no* role for recombination; indeed, GA practitioners commonly report a significant improvement in search efficacy with recombination. Nevertheless, in this paper we reverse received wisdom and view *mutation* as the driving force behind evolutionary search. If recombination has a role to play we view it as secondary (and obscure!) and thus exclude it from our current analysis.

2.2 Statistical Dynamics

Our analysis of evolutionary dynamics on fitness landscapes with neutral networks follows the model developed by van Nimwegen *et. al.* [30], termed “statistical dynamics” by obvious analogy with classical statistical mechanics. Here the fitness landscape is *coarse-grained* by decomposition into neutral networks. For populations of genotypes evolving on the landscape a *maximum entropy* approximation is made regarding the distribution of genotypes within the constituent neutral networks, thereby reducing the *state-space* of populations to a manageable number of “macroscopic” variables. As in statistical mechanics, it is difficult to predict how well this model will approximate the actual dynamics of evolving populations - it is thus judicious to test all theoretical predictions of such models against computer simulations.

All genotypes are taken to be *binary, haploid* sequences of fixed *sequence length* L . The space of all such sequences, with the graph structure induced by the adjacency of sequences connected by a single point-mutation (i.e. bit flip) defines the *sequence space* \mathcal{S} , an L -dimensional binary hypercube. A *fitness landscape* is defined to be a mapping from the sequence space to the set of real numbers (so fitness is deterministically associated with genotype; i.e. there is no “noise” on fitness evaluation). Given a fitness landscape \mathcal{L} we define two genotypes to be *connected* iff there is a sequence of fitness-preserving point-mutations taking one genotype to the other; this is evidently an equivalence relation and thus induces a partitioning of the sequence space. The *neutral networks* are defined to be the equivalence classes of this partitioning; i.e. the maximal connected subsets. We label the neutral networks Γ_i , where Γ_i has fitness w_i ; indices i, j, \dots run from 1 to N , where N is the number of neutral networks. By convention the Γ_i ’s are listed in order of ascending fitness. For simplicity, in this paper we assume that neutral network fitnesses are *strictly* increasing; i.e. $w_1 < w_2 < \dots < w_N$.

We now define a (*finite*) *population* on a sequence space \mathcal{S} to be a sequence $\mathbf{n} \equiv (n_g : g \in \mathcal{S})$ where n_g represents

the number of copies of genotype g in the population. The *population size* is given by $S \equiv \sum_{g \in \mathcal{S}} n_g$. We define an *evolutionary process* on \mathcal{S} to be a *stochastic process* [17] $\mathbf{n}(t)$ with state space the set of all populations on \mathcal{S} . In this paper all evolutionary processes will be assumed to satisfy the following:

- The time parameter t is *discrete*.
- Population size S is *fixed*.
- The process is *Markovian* [17].
- At each time step a number of genotypes are selected (independently and possibly with replacement) for *replication*. A copy of each selected genotype then *mutates* independently and is added to the population. A number of genotypes are then eliminated from the resultant (larger) population, so as to maintain a fixed population size.

An evolutionary process is *elitist* if the fittest genotypes in the population are never eliminated *en masse*.

Given a population \mathbf{n} , following our statistical mechanics analogy, we can define the corresponding “coarse-grained” population to be the vector $\mathbf{X} = (X_1, X_2, \dots, X_N)$ where $X_i \equiv \sum_{g \in \Gamma_i} n_g$ represents the number of genotypes in our population that are on the neutral network Γ_i . Population size is given by $S \equiv \sum_{i=1}^N X_i$. We shall also refer to any vector \mathbf{X} of the above form as a “population” - it should be clear from the context and notation to which manner of population we refer. Likewise, given an evolutionary process $\mathbf{n}(t)$, we derive a corresponding stochastic process $\mathbf{X}(t)$. It must be stressed that the process $\mathbf{X}(t)$ will in general *not* be Markovian - the transition probabilities from one population to the next will generally depend on the distributions of genotypes *within* the individual neutral networks Γ_i . We shall, however, approximate the evolutionary process $\mathbf{X}(t)$ with a Markov process; to do so we make a *maximum entropy approximation* that, roughly speaking, given a genotype from our population in Γ_i , that genotype is treated as if it had been chosen uniformly and at random from Γ_i . We are ultimately, however, modeling the underlying “real” evolutionary processes $\mathbf{n}(t)$, so all Monte Carlo simulations used to test results should model the full stochastic process $\mathbf{n}(t)$ rather than the coarse-grained approximation $\mathbf{X}(t)$.

2.3 Mutation Modes

All evolutionary processes considered in this paper operate via *selection* (on the basis of fitness) and *mutation*, in the sense that the Markov transition probabilities depend only on the fitnesses of sequences and the probabilities that one sequence mutates to another. Mutation is assumed independent of both genotype and locus; i.e. the probability that a point-mutation occurs at a given locus for a given genotype does not depend on the specific genotype nor on the locus under consideration. In all that follows subscripts α, β, \dots run from 0 to L . We shall consider a *mutation mode* to be defined by a probability distribution u_α , $\alpha = 0, 1, 2, \dots, L$ where u_α

is the probability that in the event of mutation of a genotype exactly α (randomly selected) loci undergo point-mutations. We define the *per-sequence mutation rate* $\mu \equiv \sum_{\alpha} \alpha u_{\alpha}$; i.e. the expected number of point-mutations per sequence. Examples of mutation modes include:

Poisson mutation: Here each locus flips independently with the same probability u . The expected number of flips per sequence is given by $\mu = Lu$. In the *long sequence length limit* $L \rightarrow \infty$, the mutation distribution approaches a Poisson distribution; i.e.:

$$u_{\alpha} = e^{-\mu} \frac{\mu^{\alpha}}{\alpha!} \quad (1)$$

Constant mutation: Here exactly μ loci undergo point-mutation; i.e.:

$$u_{\alpha} = \delta_{\alpha\mu} \quad (2)$$

To analyse our evolutionary processes we will want to know the probability m_{ij} that an (arbitrary) sequence in Γ_j which undergoes point-mutation at an (arbitrary) locus ends up in Γ_i (note the order of indices). This reflects our coarse-grained approach; in reality the probability of mutation to Γ_i will differ according as to which sequence in Γ_j we choose to mutate. We then adopt the maximum entropy approximation that m_{ij} also reflects with sufficient accuracy the probability that, if a sequence chosen arbitrarily from a *population* is in Γ_j , that sequence ends up in Γ_i after point-mutation at an arbitrary locus. At this level of analysis the (*single-locus*) *mutation matrix* $\mathbf{m} \equiv (m_{ij})$ contains all the structural information about our landscape that we require. Note that \mathbf{m} is a *stochastic* matrix; i.e. each column sums to 1. We also define the *neutrality* of the neutral network Γ_i to be $\nu_i \equiv m_{ii}$; i.e. the probability that an arbitrary point-mutation of a sequence in Γ_i leaves that sequence in Γ_i .

Since mutation will in general involve more than a single point-mutation we will also want to know the probability $m_{ij}^{(\alpha)}$ that an (arbitrary) sequence in Γ_j which undergoes point-mutations at exactly α (arbitrary) loci ends up in Γ_i . We consider the α -locus mutation as a sequence of α point-mutations (taken in some arbitrary order). We have, conditioning on the first point-mutation:

$$m_{ij}^{(\alpha)} = \sum_{k=1}^N m_{ik}^{(\alpha-1)} m_{kj} \quad (3)$$

for $\alpha = 1, 2, \dots, L$. In matrix notation, setting $\mathbf{m}^{(\alpha)}$ to be the matrix $(m_{ij}^{(\alpha)})$, we have $\mathbf{m}^{(\alpha)} = \mathbf{m}^{(\alpha-1)} \cdot \mathbf{m}$, so that:

$$\mathbf{m}^{(\alpha)} = \mathbf{m}^{\alpha} \quad (\text{matrix power}) \quad (4)$$

for $\alpha = 0, 1, 2, \dots, L$. Now given a mutation mode (u_{α}) we introduce the (*general*) *mutation matrix*:

$$\mathbf{M} = \sum_{\alpha=0}^L u_{\alpha} \mathbf{m}^{\alpha} \quad (5)$$

so that M_{ij} represents the probability that a sequence in Γ_j mutates to Γ_i under the given mutation mode. Thus e.g. for Poisson mutation we have:

$$\mathbf{M} = \mathbf{M}(\mu) = e^{-\mu(\mathbf{I}-\mathbf{m})} \quad (6)$$

where \mathbf{I} is the $N \times N$ identity matrix, while for constant mutation:

$$\mathbf{M} = \mathbf{M}(\mu) = \mathbf{m}^{\mu} \quad (7)$$

2.4 Epochal Dynamics

We will now make more precise what we mean by an ‘‘epoch’’. Referring to Fig. 2.1, during an epoch (i.e. during periods when transients associated with losing the current neutral network or moving to a higher network have subsided) the evolutionary process $\mathbf{X}(t)$ is, as a Markov process, ‘‘almost’’ *stationary* [17]; roughly speaking, the probability of finding the population in a given state does not vary over time. In [30] an evolutionary process during such an episode is described as *metastable*. We shall thus say that the evolutionary process $\mathbf{X}(t)$ is in epoch n if $X_n(t) > 0$ and $X_i(t) = 0$ for $i > n$ (i.e. Γ_n is the highest-fitness neutral network currently represented in the population) and the process is metastable as described above. As an approximation we may consider an evolutionary process $\mathbf{X}(t)$ during epoch n as a (stationary) Markov process in its own right.

2.5 Waiting Time to Portal Discovery

We attempt to provide an approximate answer to the following question: given that an evolving population $\mathbf{X}(t)$ is in the n -th epoch, what is the expected waiting time until a portal genotype (i.e. one of higher fitness than currently resides in the population) is discovered? The first issue is how we should *measure* time. Given that for real-world GA’s the most time-intensive aspect of the process is likely to be fitness evaluation, it makes sense to measure time to portal discovery in terms of the number of fitness evaluations. It is furthermore supposed that we may store the fitnesses of all genotypes currently in the population; i.e. it is only necessary to evaluate fitness when a *new* genotype appears in the population. Since the only generator of new genotypes is mutation, we thus associate fitness evaluation with the occurrence of mutation. Note that time thus defined may not equate to a time step of the evolutionary process. When confusion might arise we shall make clear which time we are talking about.

Now from Section 2.4 we deduce that, given the assumed metastability of the process during epoch n , the probability of discovery of a portal to Γ_{n+1} will be approximately the same at each time step of the process; we write this (*per-time step*) *portal discovery probability* as p_n . The distribution of the waiting time T_n (i.e. the *first passage time*) to discovery

of a portal during epoch n is thus approximately *geometric* and the *mean first passage time* measured in time steps of the evolutionary process is given by:

$$\mathbf{E}(T_n) = \frac{1}{p_n} \quad (8)$$

Now in all processes considered in this paper, $v \equiv$ *number of fitness evaluations per time-step of the evolutionary process* is constant. The expected waiting time in fitness evaluations is thus given by:

$$\mathbf{E}(vT_n) = \frac{v}{p_n} \quad (9)$$

We have, however, disregarded an important issue: in Section 2.1 we noted that, if not elitist, an evolutionary process in epoch n may well “lose” Γ_n before a portal to Γ_{n+1} is discovered. How then can T_n be well-defined? We side-step the issue in this paper; the evolutionary process which will most concern us is, in fact, elitist and we may at worst have to suppose that if our algorithm is *not* elitist, then the mutation rate is low enough that the probability of losing the current fittest network before finding a portal is small. We refer the reader to [30] for more detailed analysis of this topic.

3 ϵ -Correlated Landscapes

We now make some structural assumptions about our fitness landscape \mathcal{L} . Specifically, we assume that the probability of a point-mutation taking a sequence to a *higher fitness* neutral network is very small compared to the probability of it being neutral or of reducing fitness. We shall actually go further than this and assume that the only non-negligible fitness-increasing point-mutations are those to the *next-highest* network. More precisely:

Definition: We say that a fitness landscape is ϵ -correlated iff there exists an ϵ with $0 < \epsilon \ll 1$ such that the point-mutation matrix takes the form:

$$\mathbf{m} = \begin{pmatrix} \nu_1 & & & & & \\ \epsilon_1 & \nu_2 & & & & * \\ & \epsilon_2 & \nu_3 & & & \\ & & \mathbf{o}(\epsilon) & \ddots & & \ddots \\ & & & & \epsilon_{N-1} & \nu_N \end{pmatrix} \quad (10)$$

for some ϵ_i with $0 < \epsilon_i \leq \epsilon$ for $i = 1, 2, \dots, N-1$. ■

Here $*$ represents the transition probabilities from higher to lower fitness networks, while ϵ_i is the probability of *back-mutation* from Γ_i to Γ_{i+1} under point-mutation. The ν_i and $*$ terms are *not* taken to be $\ll 1$. With ϵ -correlation we work throughout to $\mathbf{o}(1)$ in ϵ ; i.e. we neglect all $\mathbf{o}(\epsilon)$ terms.

This property is related to the degree of (genotype-fitness) correlation present in our landscape; i.e. the degree to which sequences nearby in sequence space are likely to be of similar

fitness. Thus there is a small but non-zero probability that a point-mutation from any neutral network leads to a (probably small) increase in fitness, while the probability of a larger fitness increase is of a smaller order of magnitude. Note that ϵ -correlation is a rather stringent condition - it is certainly not to be claimed as a property that might generally be assumed of fitness landscapes arising in the context of artificial (or indeed natural) evolution. We remark, however, that if there is *no* correlation then no search technique is likely to be more efficient than random search; this is a form of “no free lunch theorem” [32]. We are thus obliged to assume *some* correlation. Note also that correlation and neutrality are (in a sense that may be made quite precise) *statistically independent* [2, 3, 25]; that is, we should not assume that neutrality implies, nor precludes, correlation. Furthermore, it is reasonable to suppose that the higher up (in fitness) our landscape we are, the more rare are those point-mutations taking us higher still; and that point-mutations leading to *large* fitness increases will be rarer than those (already rare) point-mutations leading to *small* fitness increases. We should be aware, however, that it is quite possible that even if there is a reasonable degree of correlation there may still exist sub-optimal neutral networks for which *no* fitness-increasing point-mutations exist; such networks may be thought of as the neutral analogues of isolated sub-optimal fitness peaks in the standard theory of rugged landscapes [31, 18]. Although ϵ -correlation effectively rules out the existence of sub-optimal neutral networks there is some evidence to suggest that such networks may not necessarily be common; studies of RNA folding landscapes in particular have demonstrated various *percolation*-like properties of neutral networks [8, 10, 13, 24] which suggest that almost every network approaches to within a few point-mutations of almost every other network.

For the remainder of this section we assume that our landscape satisfies the ϵ -correlation property (10) and we neglect terms of $\mathbf{o}(\epsilon)$. Now, given a mutation mode (u_α) as in Section 2.3 let us define π_n to be the probability that an (arbitrary) individual sequence in Γ_n discovers a portal to Γ_{n+1} under mutation; i.e. $\pi_n = M_{n+1,n}$ where \mathbf{M} is given by (5). From (10), we can calculate that to $\mathbf{o}(1)$ in ϵ :

$$m_{n+1,n}^\alpha = \epsilon_n \begin{cases} \frac{\nu_n^\alpha - \nu_{n+1}^\alpha}{\nu_n - \nu_{n+1}} & \nu_n \neq \nu_{n+1} \\ \alpha \nu_n^{\alpha-1} & \nu_n = \nu_{n+1} \end{cases} \quad (11)$$

so we find e.g. that for Poisson mutation in the long sequence length limit:

$$\pi_n(\mu) = \epsilon_n \begin{cases} \frac{e^{-\mu(1-\nu_n)} - e^{-\mu(1-\nu_{n+1})}}{\nu_n - \nu_{n+1}} & \nu_n \neq \nu_{n+1} \\ \mu e^{-\mu(1-\nu_n)} & \nu_n = \nu_{n+1} \end{cases}$$

as a function of per-sequence mutation rate μ , while for constant mutation:

$$\pi_n(\mu) = m_{n+1,n}^\mu \quad (12)$$

3.1 Optimal Search

We are now in a position to state the principal results of this paper:

Proposition 1: On an ϵ -correlated fitness landscape, of all possible mutation modes, that yielding the maximum value for π_n is given by constant mutation with per-sequence mutation rate:

$$\mu = (\text{nearest integer to}) \begin{cases} -\frac{\ln(-\ln \nu_n) - \ln(-\ln \nu_{n+1})}{\ln \nu_n - \ln \nu_{n+1}} & \nu_n \neq \nu_{n+1} \\ -\frac{1}{\ln \nu_n} & \nu_n = \nu_{n+1} \end{cases} \quad (13)$$

Proof: For a general mutation mode (u_α), we have $\pi_n = \sum_{\alpha=0}^L u_\alpha m_{n+1,n}^\alpha$. We need to find the maximum value for π_n as a function of u_0, u_1, \dots, u_L under the constraints $0 \leq u_\alpha \leq 1 \forall \alpha$ and $\sum_{\alpha=0}^L u_\alpha = 1$. Now π_n being linear in the u_α 's describes a hyper-plane in u -space; we have to find the maximum "height" of this hyper-plane over the simplex described by the constraints on the u_α 's. It is clear that (barring any "degeneracies" among the coefficients $m_{n+1,n}^\alpha$) the maximum must lie above a "corner" of the simplex; i.e. a point where all the u_α 's are zero except for one, $\alpha = \mu$ say, for which $u_\mu = 1$; i.e. the mutation mode is constant with per-sequence mutation rate μ . From (12) $\pi_n(\mu)$ is given by (11) with $\alpha = \mu$; the result follows by differentiating $\pi_n(\mu)$ with respect to μ , setting the resulting expression to zero and solving for μ . Note that the degenerate cases arise where two (or more) of the coefficients $m_{n+1,n}^\alpha$ coincide; in these cases the μ of (13) still yields a (now non-unique) maximum for π_n . ■

Fig. 3.1 plots the optimal mutation rate μ of Proposition 1 as a function of ν_n, ν_{n+1} .

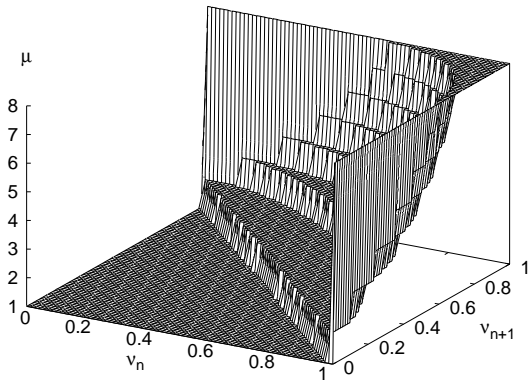


Figure 3.1: The μ of Proposition 1 plotted against ν_n, ν_{n+1}

The next proposition is of a more conjectural nature - the

"proof" supplied is by no means rigorous. Firstly, we define the *netcrawler* evolutionary process:

Definition: The *netcrawler* process operates as follows: population size = 1. At each time step the current (single) genotype replicates and the copy mutates according to the mutation mode. If the mutant offspring is less fit than its parent it is eliminated; otherwise the parent is eliminated. ■

We note that this algorithm is almost identical to the *Random Mutation Hill Climber* (RMHC) presented in [9], the only difference being that the RMHC only ever flips one (random) bit at each step. We avoid the term "hill-climber" to emphasise that, in the presence of neutral networks, the netcrawler spends most of its time not climbing hills, but rather performing *neutral walks* [13]. It is *elitist*, the number of fitness evaluations per time step is 1 and we always have $p_n = \pi_n$ so that the expected waiting time on Γ_n is just $\frac{1}{\pi_n}$.

Proposition 2: On an ϵ -correlated fitness landscape, given a mutation mode, the most efficient (fixed-population) evolutionary process is the netcrawler.

Informal Proof: Suppose a population of fixed size S evolving according to some evolutionary process under a particular mutation mode is in epoch n . From (10) it follows that for $i < n$ $M_{n+1,i}$ will be $\mathcal{O}(\epsilon)$ and hence negligible. There is thus no point in maintaining genotypes in the population on Γ_i for $i < n$. Recalling our definition of an evolutionary process (Section 2.2) our process should thus, at the elimination stage, eliminate all genotypes which have mutated to Γ_i for $i < n$. At the start of each time step, then, the entire population will be on Γ_n .

Now let π_n be as before the probability that (under the operant mutation mode) a single genotype mutates from Γ_n to Γ_{n+1} . The probability that (during the mutation phase) some mutant finds a portal to Γ_{n+1} is thus $p_n \leq 1 - (1 - \pi_n)^R$ where R is the number of replicants; equality occurs when the probabilities of the individual mutants finding a portal are *independent*. Selection for replication should thus be *without replacement*; for if a genotype is selected more than once its mutant offspring will be genetically correlated and their probabilities of finding a portal thus not independent.

Our evolutionary process during epoch n now looks like the following: at each time step a certain number of the S genotypes in our population is selected for replication/mutation. Suppose firstly that no portal is discovered. The mutants that fall off Γ_n are eliminated. The remaining population consists of the original population plus those mutants that have stayed on Γ_n . Now if our aim is to keep genetic correlations to a minimum, then we don't wish to leave (correlated) parent-offspring pairs in the population. On the other hand, we don't wish to eliminate a stay-on mutant offspring while retaining its un-mutated parent, since we would thus end up "anchoring" our search to the limited region of Γ_n

neighbouring our current population; in effect we would like to maximise *drift* of our population on Γ . We should thus retain stay-on mutants and eliminate their un-mutated parents.

Effectively, our evolutionary process now comprises S independent netcrawlers, some selection of which are “activated” at each time step. But if during some time-step a mutant discovers a portal to Γ_{n+1} then we must initiate a new epoch. Since the remaining $S - 1$ genotypes now have a negligible chance of finding a portal to Γ_{n+2} all S members of the population must become copies of the newly discovered portal genotype. But this again introduces genetic correlations between members of the population. To minimise this effect we need only choose our population size to be $S = 1$; indeed there is no loss in search efficiency by reducing the population size, since the netcrawlers are independent. Thus the most efficient evolutionary process for the given mutation mode comprises a single netcrawler. ■

To test our propositions extensive simulations (not presented here) were carried out on *Royal Road* landscapes [9, 30], which are indeed ϵ -correlated. As predicted by Proposition 2, the netcrawler consistently outperformed various population-based GA’s without recombination (including fitness-proportional roulette-wheel selection and several rank-based tournament algorithms) for Poisson and constant mutation modes. Waiting times to portal discovery were predicted with reasonable accuracy and in particular the optimum (constant) mutation rate was correctly predicted by (13) of Proposition 1.

3.2 The Adaptive Netcrawler

Propositions 1 and 2 of Section 3.1 suggest the following *adaptive netcrawler* for landscapes which we know, or at least suspect, of being the ϵ -correlated: we deploy a netcrawler with the constant mutation mode. We would then like to be able to find the optimal per-sequence mutation rate (13) for the current fittest neutral network Γ , say. This involves knowledge of the neutrality ν of Γ and also the neutrality ν' , say, of the *next highest* network Γ' , say. Now we can estimate ν “on the fly”: suppose that μ is our current per-sequence mutation rate. Then during the course of the progress of our netcrawler on Γ , we simply log the fraction of mutations, ω say, which leave us on Γ . The neutrality ν of Γ is then approximately $\omega^{1/\mu}$. However, we have no way of knowing the neutrality ν' of Γ' . A reasonable guess might be that ν' is close to ν . Indeed, examination of π_n shows that if ν' is not too different from ν , the value for π_n obtained from (12) by setting $\nu' = \nu$ is not too different to the actual optimum π_n . According to (13) then, we should re-adjust our mutation rate to $-1/\ln \nu \approx -\mu/\ln \omega$. Curiously this implies that, *whatever* the neutrality of Γ , we should ultimately see a fraction $\omega = 1/e \approx 0.36788$ of neutral mutations if our netcrawler is optimised by the above scheme. This result echoes Rechenberg’s “1/5 success rule” for *Evolution Strategies* [1], to which our netcrawler bears some resemblance.

4 Conclusions

Fitness landscapes for many real-world artificial evolution problems are likely to feature large-scale neutrality. We have argued that on such landscapes the conventional view of evolutionary search associated with rugged non-neutral landscapes is misleading and inappropriate. Instead we have characterised evolutionary dynamics with neutral networks in terms of drift along networks punctuated by transitions between networks. We have also argued that the role of recombination is obscure and may amount to little more than macro-mutation. Given this scenario, the perceived problem of premature convergence disappears and search for high-fitness genotypes is dominated by waiting for drift and mutation to find portals to higher neutral networks.

We introduced the statistical property of ϵ -correlation to describe landscapes with neutral networks for which higher networks are “accessible” only from the current network. For such landscapes we have shown (Proposition 1) that there is an optimal mutation mode and mutation rate and conjectured (Proposition 2) that there is also an optimal evolutionary search strategy which is not population-based but rather a form of hill-climber which we have dubbed the *netcrawler*. We have also proposed an *adaptive* variant of the netcrawler which gathers statistical information about the landscape as it proceeds and uses this information to self-optimize. It should be remarked in this context that a major motivation for the research presented in this paper was a series of experiments by Thompson and Layzell in on-chip electronic circuit design by evolutionary methods [28], during which an algorithm almost identical to our netcrawler was used with some success. The mutation rate deployed in these experiments, albeit chosen on heuristic grounds, in fact turns out to be the optimum predicted in this paper given the (estimated) neutrality inherent in the problem.

Finally, we remark that work in progress by the author suggests that the principal results presented in this paper may still obtain under considerably less stringent assumptions than ϵ -correlation.

Acknowledgements

The author would like to thank Inman Harvey, Adrian Thompson and Nick Jakobi for helpful s.

Bibliography

- [1] Back, T., Hoffmeister, F., and Schwefel, H.-P. (1991). A survey of evolution strategies. In Belew, L. B. and Booker, R. K., editors, *Proc. 4th Int. Conf. on Genetic Algorithms*, pages 2–9. Morgan Kaufmann, San Diego, CA.
- [2] Barnett, L. (1997). Tangled webs: Evolutionary dynamics on fitness landscapes with neutrality. Master’s thesis, COGS, University of Sussex. Avail-

- able for download from <ftp://ftp.cogs.susx.ac.uk/pub/users/inmanh/lionelb/FullDiss.ps.gz>.
- [3] Barnett, L. (1998). Ruggedness and neutrality - the NKp family of fitness landscapes. In Adami, C., Belew, R. K., Kitano, H., and Taylor, C., editors, *ALIFE VI, Proceedings of the Sixth International Conference on Artificial Life*, pages 18–27. The MIT Press.
- [4] Cliff, D., Husbands, P., and Harvey, I. (1993). Evolving visually guided robots. In J.-A. Meyer, H. Roitblat, S. W., editor, *From Animals to Animats 2: Proc. of the Second Intl. Conf. on Simulation of Adaptive Behavior, (SAB92)*, pages 374–383. MIT Press/Bradford Books, Cambridge MA.
- [5] Crow, J. F. and Kimura, M. (1970). *An Introduction to Population Genetics Theory*. Harper and Row, New York.
- [6] Derrida, B. and Peliti, L. (1991). Evolution in a flat fitness landscape. *Bull. Math. Biol.*, 53(3):355–382.
- [7] Eigen, M., McCaskill, J., and Schuster, P. (1989). The molecular quasispecies. *Adv. Chem. Phys.*, 75:149–263.
- [8] Fontana, W., Stadler, P. F., Bornberg-Bauer, E. G., Griesmacher, T., Hofacker, I. L., Tacker, M., Tarazona, P., Weinberger, E. D., and Schuster, P. (1993). RNA folding and combinatorial landscapes. *Phys. Rev. E*, 47(3):2083–2099.
- [9] Forrest, S. and Mitchell, M. (1993). Relative building block fitness and the building block hypothesis. In Whitely, D., editor, *Foundations of Genetic Algorithms 2*. Morgan Kaufmann, San Mateo, CA.
- [10] Forst, C. V., Reidys, C., and Weber, J. (1995). Evolutionary dynamics and optimization: Neutral networks as model-landscapes for RNA secondary-structure folding-landscapes. *Lecture notes in Artificial Intelligence*, 929: Advances in Artificial Life.
- [11] Harvey, I. (1994). Evolutionary robotics and saga: the case for hill crawling and tournament selection. In Langton, C., editor, *Artificial Life III, Santa Fe Institute Studies in the Sciences of Complexity, Proc.*, volume XVI, pages 299–326. Addison Wesley.
- [12] Harvey, I. and Thompson, A. (1996). Through the labyrinth evolution finds a way: A silicon ridge. In *Proc. 1st Internat. Conf. Evol. Sys.: From Biology to Hardware (ICES 96)*. Springer Verlag.
- [13] Huynen, M. A. (1996). Exploring phenotype space through neutral evolution. *J. Mol. Evol.*, 63:63–78.
- [14] Huynen, M. A., Stadler, P. F., and Fontana, W. (1996). Smoothness within ruggedness: The role of neutrality in adaptation. *Proc. Natl. Acad. Sci. (USA)*, 93:397–401.
- [15] Jakobi, N., Husbands, P., and Harvey, I. (1995). Noise and the reality gap: The use of simulation in evolutionary robotics. In Moran, F., Moreno, A., Merelo, J. J., and Chacon, P., editors, *Proc. 3rd European Conference on Artificial Life, Springer-Verlag, Lecture Notes in Artificial Intelligence 929*, volume XVI, pages 704–720. Springer Verlag.
- [16] Jakobi, N. and Quinn, M. (1998). Some problems and a few solutions for open-ended evolutionary robotics. In Husbands, P. and Meyer, J.-A., editors, *Proceedings of Evorob98*. Springer Verlag.
- [17] Karlin, S. and Taylor, H. M. (1975). *A first course in Stochastic Processes (2nd ed.)*. Academic Press (NY).
- [18] Kauffman, S. A. (1993). *The Origins of Order - Organization and Selection in Evolution*. Oxford University Press, New York.
- [19] Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press.
- [20] Kondrashov, A. S. (1982). Selection against harmful mutations in large sexual and asexual populations. *Genet. Res.*, 40:325–332.
- [21] Layzell, P. (1998). The evolvable motherboard. Cognitive Science Research Paper 479, University of Sussex, School of Cognitive and Computing Sciences.
- [22] Maynard Smith, J. (1978). *The Evolution of Sex*. Cambridge University Press, Cambridge, UK.
- [23] Nowak, M. and Schuster, P. (1989). Error thresholds of replication in finite populations: Mutation frequencies and the onset of Muller’s ratchet. *J. Theor. Biol.*, 137:375–395.
- [24] Reidys, C., Stadler, P. F., and Schuster, P. (1997). Generic properties of combinatorial maps: Neutral networks of RNA secondary structures. *Bull. Math. Biol.*, 59(2):339–397.
- [25] Reidys, C. M. and Stadler, P. F. (1998). Neutrality in fitness landscapes. working paper 98-10-089, Santa Fe Institute.
- [26] Smith, T. M. C., Philippides, A., and O’Shea, M. (2001). Not measuring evolvability: Initial exploration of an evolutionary robotics search space. In *Proceedings of the 2001 Congress on Evolutionary Computation: CEC2001 (submitted)*. IEEE, Korea.
- [27] Thompson, A. (1996). Silicon evolution. In Koza, J. R., Goldberg, D. E., Fogel, D. B., and Riolo, R. L., editors, *Genetic Programming 1996: Proc. 1st Annual Conf. (GP96)*, pages 444–452. Cambridge, MA: MIT Press.
- [28] Thompson, A. and Layzell, P. (2000). Evolution of robustness in an electronics design. In Miller, J., Thompson, A., Thomson, P., and Fogarty, T., editors, *Proc. 3rd Int. Conf. on Evolvable Systems (ICES2000): From biology to hardware*, volume 1801 of LNCS, pages 218–228. Springer-Verlag.
- [29] Thompson, A., Layzell, P., and Zebulum, R. S. (1999). Explorations in design space: Unconventional electronics design through artificial evolution. *IEEE Trans. Evol. Comp.*, 3(3):167–196.
- [30] van Nimwegen, E., Crutchfield, J. P., and Mitchell, M. (1998). Statistical dynamics of the royal road genetic algorithm. working paper 97-04-35, Santa Fe Institute.
- [31] Weinberger, E. D. (1990). Correlated and uncorrelated fitness landscapes and how to tell the difference. *Biol. Cybern.*, 63:325–336.
- [32] Wolpert, D. H. and Macready, W. G. (1995). No free lunch theorems for search. technical report SFI-TR-95-02-010, Santa Fe Institute.