Correlated and Uncorrelated Fitness Landscapes and How to Tell the Difference

Article *in* Biological Cybernetics · September 1990

DOI: 10.1007/BF00202749

1 author:

Edward D. Weinberger
New York University
27 PUBLICATIONS **2,690** CITATIONS

Biol. Cybern. 63, 325–336 (1990)

Biological Cybernetics

Correlated and Uncorrelated Fitness Landscapes and How to Tell the Difference

E. Weinberger

Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, PA 19104, USA

Received September 30, 1989/Accepted in revised form March 7, 1990

Abstract. The properties of multi-peaked "fitness landscapes" have attracted attention in a wide variety of fields, including evolutionary biology. However, relatively little attention has been paid to the properties of the landscapes themselves. Herein, we suggest a framework for the mathematical treatment of such landscapes, including an explicit mathematical model. A central role in this discussion is played by the autocorrelation of fitnesses obtained from a random walk on the landscape. Our ideas about average autocorrelations allow us to formulate a condition (satisfied by a wide class of landscapes we call AR(1) landscapes) under which the average autocorrelation approximates a decaying exponential. We then show how our mathematical model can be used to estimate both the globally optimal fitnesses of AR(1) landscapes and their local structure. We illustrate some aspects of our method with computer experiments based on a single family of landscapes (Kauffman's "N-k model"), that is shown to be a generic AR(1) landscape. We close by discussing how these ideas might be useful in the "tuning" of combinatorial optimization algorithms, and in modelling in the experimental sciences.

Introduction

There has been considerable recent interest in considering evolution as a combinatorial optimization problem, that is, a problem in finding the best of a large, but finite number of possibilities. Biologists, such as Kauffman and Levin (1987), have embraced this paradigm in the hope that they might learn something new about evolution, and computer scientists, such as Holland (1981) and Brady (1985), hope to use evolutionary strategies in developing new methods of solving combinatorial optimization problems. Similar issues have also attracted the attention of physicists, such as Palmer (1989) and Stein (1989). They speculate that the thermodynamics of glassy systems, such as polymers and other more or less random covalent networks, is intimately related to the complex structure of the barrier heights present in the potential surface. The common denominator in all of this work is a notion that Kauffman and Levin have called a "rugged landscape.". If one is a biologist, such a landscape can be interpreted as a fitness landscape; if one is a computer scientist, the landscape is the set of allowable configurations in some optimization problem; and if one is a physicist/chemist, it is a glass and/or spin glass.

To the biologist, the notion of an adaptive landscape is not new, having been proposed in 1932 by Sewell Wright (Wright 1932). Subsequent developments in biological thinking have only reinforced the power of this idea. Molecular biology has made clear the essential discreteness of the genome, and thus the finitude of the number of possible organisms. It is, perhaps, easier to make sense of the concept of an adaptive landscape if one considers the evolution of individual molecules, rather than entire organisms; hence, Smith's notion of a "peptide space" (Smith 1970). In such a space, one arranges all peptides of a specified length in such a way that nearest neighbors differ by a single amino acid substitution at a single site. One might then define the fitness of the peptide as its ability to bind to a particular substrate, to catalyze a specific reaction, etc. The lethal nature of the sickle cell anemia mutation in humans emphasizes the fact that even a single point mutation can result in a dramatic change in fitness. Furthermore, the selection of an optimal enzyme to catalyze a particular reaction involves a host of complex tradeoffs: the enzyme must bind the substrate tightly enough so that the reaction will proceed, but not so tightly that it will not be released when the reaction is completed, the enzyme must not also catalyze competing reactions, and, of course, the enzyme must not interfere with the action of other enzymes.

Present address: Max-Planck-Institut für biophysikalische Chemie, Postfach 2841, Am Fassberg, D-3400 Göttingen-Nikolausberg, Federal Republic of Germany

It was these considerations that inspired Kauffman and Levin (1987) to draw the analogy between biological constraints and those imposed by the combinatorial optimization problems that have received much recent attention in computer science (see, for example, Garey and Johnson 1979). Perhaps the most famous example of this class of problems is the travelling salesman problem (TSP), which is to find the shortest tour through N cities from an initial city, visiting each city once, and returning at the end to the initial city. For as few as 11 cities, the number of possible tours is in the millions, growing as $((N-1)/2)!$. This is typical of combinatorial optimization: out of a finite, but extremely large set of entities, one would like to find one that is, in some well defined sense, the "fittest". For a large class of such problems (e.g. the class of NP-complete problems), it is thought that even the most efficient algorithms that solve them must require a computational effort that grows at least exponentially with some measure of the "size" of the problem (such as the number of cities that the traveling salesman must visit). Indeed, the only known method of finding the optimal traveling salesman tour is to exhaustively search through the list of possible tours, checking each to see if each is shorter (and therefore "fitter") than all of those previously encountered. Because of the extreme amounts of time required for such a search, which can easily exceed the age of the universe for even moderate N on the largest of computers, algorithms that settle for sub-optimal tours are used in practice. The effective intractability of other combinatorial optimization problems places a similar constraint on algorithms to solve these problems, as well. Typically, these algorithms use a generalization of the fact that tours that visit the cities in more or less the same order have more or less the same length. It is therefore profitable to search the vicinity of a good solution in the hopes of finding a better solution. This heuristic is incorporated into the algorithms by explicitly defining "neighborhoods", and searching from neighbor to neighbor.

Essential features of the above discussion of the travelling salesman problem are also present in our evolutionary paradigm. In the case of a 20 amino acid peptide space, there are 20^N possible proteins of length N which would need to be searched to find the "fittest" protein for a particular function. How can an evolving biological system search through such a vast number of possibilities? Obviously, it can't. Instead, it would seem that evolution, like practical combinatorial optimization algorithms, makes do with sub-optimal solutions. Indeed, it has been shown previously (Weinberger 1987a, b) that there is a detailed analogy between evolutionary optimization and the combinatorial optimization technique known as simulated annealing (Kirkpatrick et al. 1983). The "local hill climbing" heuristic described above, which, in the context of combinatorial optimization, involves transitions to randomly selected fitter neighbors, is one of the simplest of combinatorial optimization schemes.

It is useful to consider a rugged landscape as an abstract mathematical object, both because some important considerations emerge immediately, and because a framework will be provided for subsequent discussion. This we do in the next two sections, motivating our remarks by considering the two concrete examples of combinatorial optimization problems mentioned previously, the travelling salesman problem and optimization on peptide space. The second of these secions rigorously defines the notion of a "correlation length", and, more generally, an autocorrelation function for such combinatorial landscapes. We will see that this autocorrelation function is indeed a decaying exponential, as implied by the existence of a single correlation length, for the wide class of landscapes that we call AR(1) landscapes. We then use these theoretical tools to explain why the two most obvious methods of optimization on combinatorial landscapes, random search and local hill climbing, do remarkably badly in finding the point on the landscape with the globally optimal fitnesses. The next section shows how our theoretical framework can be used to predict aspects of the local structure of a landscape, illustrating our method with computer experiments based on a simple family of two amino acid peptide landscapes (the "N-k model") proposed in Kauffman et al. In this section, we also demonstrate the intriguing fact that N-k landscapes are generic members of the class of AR(1) landscapes. The final section of the paper discusses the significance of our results.

Some Thoughts about the Abstract Structure of Fitness Landscapes

In all of the applications mentioned above, one would like to know the details of specific landscapes. Unfortunately, this "best of all possible worlds" scenario is unrealistic, due to the enormous amount of experimental data required. A more realistic goal would be to gather a moderate amount of data about the landscape and infer likely statistical properties of the ensemble of landscapes that might fit the data. It is therefore appropriate to begin a mathematical description of rugged fitness landscapes by assuming that the fitnesses of its points are random variables, and inferring properties of their moments. Although this program can be carried out if the random variables have essentially any distribution, our results will be most useful if the random fitness values have a joint Gaussian distribution. In that case, the distribution is completely specified by a vector of mean fitness values and a matrix specifying the covariance of the fitnesses. It is also useful to note that any linear combination of jointly Gaussian variables is, itself, Gaussian.

There are both theoretical and simulation results showing that the marginal fitness distributions are, in fact, Gaussian for a wide class of landscapes of practical interest. (Of course, the fact that all of the marginal distributions are Gaussian is necessary, but not sufficient to conclude that a collection of random variables is jointly Gaussian. See Feller (1972) for counter-examples. However, Karlin and Taylor (1975) note that the

variables are indeed jointly Gaussian if the marginal distributions of all linear combinations of the variables are Gaussian.) In many cases, fitnesses are likely to be determined by a large number of relatively independent factors of approximately equal strength. Under surprisingly general hypothesis (even more general than the usual assumption of statistical independence), it can be shown that the distribution of the average of a large number of random variables tends to a normal distribution (see, for example, Breiman 1968). Computer simulations suggest that this situation actually occurs in a wide variety of fields. Kirkpatrick and Toulouse (1985) report that intercity distances in numerical simulations of the travelling salesman landscape are independently distributed, in which case the conditions of the central limit theorem are satisfied rigorously. As we will see in the discussion of the N-k model, the conditions for the central limit theorem are satisfied there, as well. Eigen et al. (1988) give examples of fitness landscapes – some of which are based on highly detailed models – in which normal distributions are observed. Finally, White (1984) shows graphically that "fitness" for a problem in computer design is also normally distributed, and he observes that the normal distribution is, to varying degrees of approximation, applicable to a wide class of such problems.

Kauffman and Levin studied the case in which the fitness of each point in the landscape was chosen independently from each of the others, but they noted that, in general, fitnesses of nearby points are usually correlated. Clearly, any quantification of this dependence would also imply a quantification of the "correlation structure" of the landscape. One way of thinking about these questions is to designate a starting configuration and consider the set of operations that transforms this configuration to each of the other possible configurations (assuming that such transformations exist). It is easy to verify that this set of transformations forms a mathematical group (Herstein 1964 is one of the many texts that discuss the elements of group theory). An important aspect of this theory is the decomposition of the elements of the group into combinations of a relatively small number of *generators*. The relevance of this idea to the study of our transformations is that it is usually possible to express arbitrary transformations as a succession of a few simple transformations. We can then take these transformations as the generators of the relevant group, thus exploiting regularities in the group structure. Kauffman and Levin's uncorrelated fitness landscape, which is an idealized version of peptide space, provide a particularly simple example of these ideas. They assume that only two amino acids, say, glycene and alanine, are the only possible residues of a peptide of length N. The 2^N possible sequences are then the points in the rugged landscape. It is convenient to label the amino acids with the values "1" and "0". The sequences can then be interpreted as the verticies of an N dimensional hypercube. Clearly, any sequence can be obtained from any other in this model by a finite sequence of point mutations (single bit flips), so that the group of all possible transformations is the collection of all possible point mutations. This group is generated by the N mutations that involve flipping single bits. Kauffman and Levin define sequences that differ by a single bit as "neighbors", and, in general, it is natural to identify the configurations that can be reached by a single application of a generator transformation as the "neighborhood" of the starting configuration. We will also call the set of generator transformations the *move set*, following a personal communication by Steve R. White.

It is possible to generalize the paradigm given above to the travelling salesman problem. One way to represent a TSP tour of N cities is with an $N \times N$ array of bits, $\mathscr{B} = [B_{ij}]$, such that $B_{ij} = 1$ if and only if city j follows city i in the tour. However, the move set here must preserve the special characteristics of the $\mathscr{B}$ matrix, which must always have exactly one "1" in every row and in every column. In contrast to Kauffman and Levin's model of peptide space, the order in which the generator transformations are applied to a starting travelling salesman configuration determines which of several possible configurations result. For example, suppose we start with a tour that visits Chicago first, London next, and New York third. Suppose further that our move set involves the two transformations "visit the second city first and the first city second" and "visit the thrid city first and the first city third". One way of composing the transformations leads to the tour "New York, Chicago, London", but the other leads to the tour "London, New York, Chicago". This phenomenon is well known to group theorists, who call transformations that can be interchanged *commutative*, and groups where all transformations commute *Abelian groups*.

There are several biological implications of these observations. First of all, we note that Kauffman and Levin's choice of move set is by no means the only possible choice. It is also possible to move from one vertex to another by *transposing* bit orderings, just as city orderings were transposed above. For example, starting with the vertex 000001, one might imagine interchanging the fifth and sixth bit, to obtain the vertex 000010. Mutation via transposition is well known, even if it does not occur as frequently as point mutation. These remarks also apply to the various mutations, such as insertions and deletions, that result in frame shifts. Non-commutativity of mutational move sets may also drastically reduce the number of possible sequences of mutations leading from the starting state to a desirable final state. The argument here is easiest to see if we assume the Kauffman/Levin model, but with transpositions added to the move set. There, a locally optimum sequence must be fitter than each of its N neighbors reachable by a single bit flip, and each of its neighbors reachable by interchanging any pair of bits. Although the number of these neighbors depends on the number and distribution of "1" bits in the configuration, it is clear that, for large N, the probability that a randomly chosen configuration is a local optimum is low. Thus, if a particular sequence of moves leads from the starting sequence to a local optimum, and if there is

any non-commutativity in the sequence, any reordering of the moves that leads to a different configuration will usually lead to some non-optimal configuration.

Returning to the purely theoretical, we observe that the notion of group generators allows us to define a *metric* on fitness landscapes in a natural way: the "distance" between two points on the landscape is simply the smallest number of generators ("steps") needed to move from one point to the other. This definition is a natural extension of the more familiar notion of the Hamming distance between two bit strings. This connection can be seen most easily when the generators commute (so that the entire group is abelian). Such a group is isomorphic to the direct product of cyclic groups; that is, each of its elements can be treated as an ordered "tuple" of N integers, with the group operation defined as componentwise modular addition, possible with a different modulus for each component. One set of generators is then the set of "unit vectors": $(1, 0, \ldots, 0)$, $(0, 1, \ldots, 0)$, etc. Another is the set of all vectors with exactly n 1's and $N-n$ 0's, provided that N and n are mutually prime. The point of all this, of course, is that any discussion of which nodes are adjacent, and therefore any discussion of the correlation between neighboring nodes, is based on the choice of which generators are used for the transformation group.

Measuring Correlation

Given the theoretical framework in the previous section, we can now turn to the question of how to measure correlation precisely for a given move set. We now make an important additional assumption; namely, that the landscape is *statistically isotropic*. To attach a heuristic meaning to this term, consider the sequence of fitnesses obtained by starting at a randomly chosen configuration and moving to a randomly chosen neighbor, then a randomly chosen neighbor of that neighbor, etc. A statistically isotropic landscape is one in which the statistics of this sequence of fitnesses are the same, regardless of the starting point chosen. More formally, consider a stationary random process generating a sequence of transitions between neighboring configurations in which the transitions to each of the neighbors is equiprobable. Statistically isotropic landscapes are those in which the fitness values assigned to the configurations also form a stationary random process for the assumed joint distribution of fitnesses. (Both the travelling salesman problem and the protein space problem given above satisfy this condition.) The significance of the assumption of statistical isotropy is that we can apply the powerful results of the theory of stationary random processes to the study of random walks on fitness landscapes, as will be discussed next.

If we assume such a landscape, each component of the vector of mean fitness values must have the same value, which is the average fitness of a randomly chosen configuration, and which is therefore easily estimated in practice. Without loss of generality but with a gain in simplicity, we assume for the rest of this section that this mean value is zero. The correlation matrix of the fitnesses, $\mathscr{C}$, can also be estimated from the autocovariance function of the fitness sequence if we describe more precisely the nature of the random walk involved. We assume that steps are made from configuration i to neighboring configuration j according to a Markov chain with stationary transition probabilities P_{ij} *that are independent of the landscape fitnesses*. (To make this crucial assumption absolutely clear, we restate our paradigm in the language of physics. The walks are now on "energy", rather than "fitness" landscapes, and the random dynamics of the walk arise from thermal motion, rather than mutation. The assumption is now that walks take place at infinite temperature, so that all energy barriers are ignored.) We further assume that the Markov chain is ergodic (i.e. that each configuration has a positive probability of being reached in finite time starting from any other configuration), so that the chain has a unique equilibrium distribution, the components of which we denote $\pi_1, \pi_2, \ldots \pi_M$, where M is the number of points on the landscape (e.g. 2^N for the N dimensional boolean hypercube). The autocovariance function, $r(s)$, of a sequence of n fitness values $F_1, F_2 \ldots F_n$ obtained by such a random walk is then approximately

$$r(s) = \sum_i \pi_i \sum_j P_{ij}^s F_i F_j ,$$

if we ignore the sample mean fitnesses, $\sum_i \pi_i F_i$, which tend to zero as N tends to infinity. Note that the function $r(s)$ depends only on s, the number of steps since the beginning of the walk, because of the statistical isotropy assumption. If we average over all landscapes in the ensemble of possible landscapes, we have

$$\begin{aligned} E[r(s)] &= \sum_i \pi_i \sum_j P_{ij}^s E[F_i F_j] \\ &= \sum_i \pi_i \sum_{d=0}^{s} \phi(d, s) R(d) \\ &= \sum_{d=0}^{s} \phi(d, s) R(d) , \end{aligned} \tag{1}$$

where $\phi(d, s)$ is the probability that the Hamming distance between configurations i and j is d after s steps and

$$R(d) = E[F_c F_{c+d}] - E[F_c] E[F_{c+d}] . \tag{2}$$

for $c, d \geqslant 0$. We can make sense of the subscripts by choosing some point as the "zero state" and using the "generator metric" defined in the previous section, to assign distances to everything else. Assuming statistical isotropy, the only thing that matters, on average, is this distance. In fact, the only thing that matters is the *relative* distance between points, so that the choice of the zero state doesn't matter, and $R(d)$ is indeed independent of c. Although it is possible for d to be strictly smaller than s, the probability that $d = s$ is $M!/[M^s(M-s)!]$, which is extremely close to 1 if s is significantly smaller than $O(N^{1/2})$. Furthermore, it follows from our definition of statistical isotropy that the

full correlation matrix, $\mathscr{C}$, of the n points we have sampled is given by

$$\mathscr{C}_{ij} = R(|i-j|)\,, \tag{3}$$

where $|i-j|$ is the distance between point i and point j.

The power of the theory of stationary time series (and therefore its utility in the present context) follows from the special form of $\mathscr{C}$. We now proceed to list some of the major results of this theory and discuss their applicability to fitness landscapes. Following Priestley (1981), we first extend and modify the definition of $R(d)$ in (2) above by replacing the index $c+d$ on the right hand side of the definition with the index $c+d$ mod n. Call the new function $R'(d)$. It is then easy to check that $R'(s) = R'(n-s)$. Because $\mathscr{C}'$, the correspondingly modified version of $\mathscr{C}$, must commute with cyclic permutation matrices, the eigenvalues of $\mathscr{C}'$ are the components of the discrete Fourier transform, $H(k)$, of $R'(s)$, given by

$$H(k) = \sum_{s=0}^{n-1} R'(s)\, \mathrm{e}^{-\mathrm{i}\omega_k s}\,,$$

where $\omega_k = 2\pi k/n$. High values of $H(k)$ represent high correlations between points n/k steps apart because of high amplitudes of the corresponding spatial frequencies.

The theory of stationary time series also includes a well developed theory of prediction, which is also at our disposal. One aspect of this theory involves estimating the value of F_s from linear combinations of previous values of F. If the F distribution is jointly Gaussian, it can be shown that the optimal linear predictor is also the optimal non-linear predictor. Even in the general case, the predictor

$$\hat{F}_s = \alpha_1 F_{s-1} + \alpha_2 F_{s-2} + \cdots + \alpha_p F_{s-p}$$

is optimal over all such linear combinations of the F's, provided that the α's satisfy the *Yule-Walker equations*,

$$\begin{aligned}
R(1) &= \alpha_1 R(0) + \alpha_2 R(1) + \alpha_3 R(2) + \cdots + \alpha_p R(p-1)\\
R(2) &= \alpha_1 R(1) + \alpha_2 R(0) + \alpha_3 R(1) + \cdots + \alpha_p R(p-2)\\
R(3) &= \alpha_1 R(2) + \alpha_2 R(1) + \alpha_3 R(0) + \cdots + \alpha_p R(p-3)\\
&\vdots\\
R(p) &= \alpha_1 R(p-1) + \alpha_2 R(p-2) + \alpha_3 R(p-3)\\
&\quad + \cdots + \alpha_p R(0)\,.
\end{aligned}$$

(Karlin and Taylor 1975). The last of the above equations can also be interpreted as recurrence relation for the values of $R(s)$, thus implicitly specifying the class of autocovariance functions, $R(s)$, for which the estimator works well. It is easy to verify that the general solution to the recurrence is of the form

$$R(s) = c_1 z_1^s + c_2 z_2^s + \cdots + c_p z_p^s\,,$$

where the c's are determined from the initial conditions and the z's are the roots of the characteristic equation

$$z^p - \alpha_1 z^{p-1} - \cdots - \alpha_{p-2} z^2 - \alpha_{p-1} z - \alpha_p = 0$$

In most cases, each of the p roots to this equation will be distinct, with $|z| < 1$, and the c's will be non-zero. The magnitudes of the z's determine the rate at which different local features of the landscape are forgotten as the walk moves away from its starting point. Of particular interest is the root with the largest magnitude, which we denote z_L. The behavior of $R(s)$ for large s will be similar to that of $c_L z_L^s = c_L \mathrm{e}^{-s/\tau}$, where $\tau = -1/\ln z_L$. Because of the rapid decay of the exponential function, points that are significantly further apart than τ steps are effectively uncorrelated and all local properties of the landscape due to proximity to the starting point are effectively forgotten. Thus, τ is the effective "correlation length" of such a landscape. Furthermore, for large s, the dynamics of the process simplify to that of the so-called AR(1) or *first order autoregressive* process, which defined by the equation

$$X_t = z_L X_{t-1} + N_t\,,$$

where N_t is a stationary sequence of uncorrelated random variables.

There are good reasons to think that the AR(1) process captures the statistics of walks on a wide class of landscapes, including both the N-k model and the travelling salesman problem, even when s is small. We defer an explicit calculation of the autocorrelation function for the N-k model to the section that discusses the model itself in more detail and focus now on the more general situation. Yet another theorem in the theory of stationary stochastic processes asserts that a stationary process that is both Gaussian and Markov *must*, like the AR(1) process with Gaussian N_t, have a decaying exponential as its autocorrelation function. Furthermore, because a Gaussian process is completely characterized by its mean and autocorrelation function, *all* such Markov processes must be AR(1). Both the random walks associated with the N-k model and the travelling salesman problem, as well as many other problems, are "effectively Markov" in the following sense: In each of these problems, the fitness of a successive point on the walk is typically a linear combination of the fitness of the previous point and a small perturbation. If successive perturbations were strictly uncorrelated with each other, then the AR(1) model would be strictly applicable. Instead, there is some small correlation between the perturbations, if only because there is a small probability of returning to the same place in the landscape and retracing at least one step in the walk. If the landscape is large, the AR(1) model might still approximate the dynamics involved. Of course, the validity of this approximation must be checked for each specific case, which is why we include the N-k autocorrelation calculation described above as a prototype of how such a verification is performed.

The foregoing implies that, if the AR(1) process is *not* a good approximation for the statistics of the random walk, then the sequence of fitnesses induced by the random walk process is non-Markovian. This, in fact, will be the generic situation if the transition probabilities in the random walk are allowed to depend on the fitness values in the landscape. Although (1) is invalid in this case, the theory of stationary processes

still applies. An intriguing example of such a walk is the sequence of travelling salesman tour lengths obtained by running the simulated annealing algorithm at a finite temperature. There, the non-Markovian nature of the process is evident from the fact that regions of "tour space" associated with relatively short tours are preferentially populated.

In the extreme non-Markovian case, p is large, and the magnitude of the roots of the characteristic equation may cluster about 1. In this case, the autocorrelation function becomes effectively non-exponential, and, in an important special case, may decay as some inverse power of s. This case is an equivalent condition to statistical self-similarity. Heuristically, self-similarity means that every piece of the landscape has the same statistical properties as the whole. (Although the discreteness of the landscape guarantess that self-similarity must break down when the pieces are sufficiently small, landscapes of practical interest have sufficiently many points that approximate self-similarity can hold over a wide range of partitions.) We make the notion of self-similarity precise by assuming that R is "self-affine" in a sense suggested by Mandelbrot (1982). The mathematical statement of this version of self-affinity is

$$R(x) = y^s R(xy)$$

for some constant $s > 0$ and all rescaling parameters, $y > 0$. If we set $x = 1$ in this equation, we see that

$$R(1) = y^s R(y)$$

so that

$$R(x) = \text{constant} \times x^{-s}$$

for some constant $s > 0$.

We now return to the general case and discuss briefly how a knowledge of the statistical properties of a random walk on the landscape can be used to infer its other features. In principal, any property of the landscape should be recoverable from a sufficiently long random walk. Indeed, as shown in the theoretical discussion of landscapes given above, a landscape is specified uniquely by the distribution of fitnesses on it and by allowable transitions between configurations. The distribution of fitnesses can be obtained trivially from the random walk, and, because fitnesses are almost surely unique, allowable transitions can be inferred from the presence or absence of the relevant pairs of fitnesses. What is, perhaps, more to the point is that a knowledge of $R(s)$ and the assumptions of our model allow us to *predict* important local features of the landscape. For example, if the fitness of a given configuration $\mathscr{X}$ is x, the distribution of fitnesses, y, for a neighboring point, $\mathscr{Y}$, on the landscape is the conditional distribution of y, given x. Papoulis (1965) has shown that this last distribution is also normal, with mean and variance given by

$$E[y|\text{fitness of } \mathscr{X} = x] = \mu + R(1)(x - \mu)$$

and

$$\text{Var}[y|\text{fitness of } \mathscr{C} = x] = \sigma^2[1 - R^2(1)],$$

where μ and σ^2 are the marginal means and variances for *both* $\mathscr{X}$ and $\mathscr{Y}$ (The marginal fitness distribution for every point on the landscape must be the same, by the hypothesis of statistical isotropy).

Why Combinatorial Optimization Problems are Difficult

In this section, we use the results of our previous discussion to study two of the most obvious adaptation strategies, namely keeping the best of many randomly chosen fitness values and "local hill climbing". We start by presenting some mathematical results that show just how badly one does if one tries to find optima by picking points at random from the landscape and retaining the best one found so far. These results constitute a rigorous discussion of Kauffman and Levin's "long jump adaptation", in which landscape correlations are ignored. They list several ways that long jumps can arise from mutation. The most obvious of these is that even a single frame shift mutation often causes almost all of the codons downstream of the mutation to be misread and code for different amino acids. We have already argued that the randomly chosen fitnesses likely to be generated by a sequence of long jump mutations are likely to be independent Gaussian random variables. It is therefore natural to ask how the maximum of a sequence of these random variables grows as a function of the number considered. A precise answer to this question is given by the following

Theorem 1. *If X_N, with $N = 1, 2, 3, \ldots$ are independent Gaussian random variables with mean zero and variance σ^2, where σ^2 doesn't depend on N, then*

$$Pr\left\{\liminf_{N\to\infty} (2\sigma^2 \ln N)^{-1/2} \max_{k \leqslant N} X_k \geqslant 1\right\} = 1$$

and

$$Pr\left\{\limsup_{N\to\infty} (2\sigma^2 \ln N)^{-1/2} \max_{k \leqslant N} X_k \leqslant \sqrt{2}\right\} = 1.$$

Although this result seems to be well known to at least some probabilists – as per the acknowledgements – a published proof does not seem to be readily available, so we provide one in Appendix I. In less formal terms, the theorem states that the maximum of a sequence of N independently chosen Gaussian random variables will almost surely grow roughly as $\sqrt{2 \ln N \sigma^2}$. We note that, for $N = 10^3$, $\sqrt{2 \ln N} \approx 3.71$, and, for $N = 10^6$, $\sqrt{2 \ln N} \approx 5.25$. This shows how remarkably slowly the maximum of the N samples grows with N. We also note that the general subject of this theorem, that of "large deviations" of a random variable from its mean value, has received much attention over the years. The interested reader is referred to David (1970) and Gumbel (1958).

We can use variants of this result to compare what happens to the "typical" change in fitness between

neighbors as compared with the global maximum of a landscape as the size, N, of the problem increases. Typically, the fitness will be the sum of N approximately independent components of approximately constant magnitude, so that the mean fitness is roughly proportional to N. We divide each component by N to ensure that the mean of the scaled fitnesses is independent of N. This rescaling also changes the variance of the fitnesses by the factor $1/N$. We can then estimate the change in fitness between neighboring configurations $\mathscr{X}$ and $\mathscr{Y}$ by writing

$$F_{\mathscr{Y}} = R(1)F_{\mathscr{X}} + \Delta F,$$

where ΔF is a normally distributed random variable with zero mean and variance $\sigma^2[1 - R^2(1)]/N$. As we saw in the previous section, it is reasonable to take the ΔF's for different pairs of neighbors to be independent. The largest improvement we can expect from a single step is then the maximum of N independent copies of ΔF, which we estimate via

Corollary 1. *If N Gaussian random variables have zero mean and variance σ^2/N, then*

$$Pr\left\{\liminf_{N\to\infty} (2\sigma^2 \ln N/N)^{-1/2} \max_{k\leqslant N} X_k \geqslant 1\right\} = 1$$

and

$$Pr\left\{\limsup_{N\to\infty} (2\sigma^2 \ln N/N)^{-1/2} \max_{k\geqslant N} X_k \geqslant \sqrt{2}\right\} = 1.$$

The proofs of both Corollary 1 and Corollary 2 (stated below) are intimately related to that of Theorem 1, so that the are also discussed in the Appendix.

Thus, with the normalization described above, the fitness of typical single step improvements is $O(\sqrt{\ln N/N})$, which recedes to zero as the size of the landscape increases.

We note that roughly $2^{N-\tau}$ points on the landscape are effectively uncorrelated so that we can study the behavior of the global maximum by using the same methods as above. Of particular interest is the following

Corollary 2. *If the Gaussian random variables have zero mean and variance σ^2/N, then*

$$Pr\left\{\liminf_{N\to\infty} \left(2\ln 2\sigma^2 \frac{N}{N-\tau}\right)^{-1/2} \max_{k\leqslant 2^{N-\tau}} X_k \geqslant 1\right\} = 1$$

and

$$Pr\left\{\limsup_{N\to\infty} \left(2\ln 2\sigma^2 \frac{N}{N-\tau}\right)^{-1/2} \max_{k\leqslant 2^{N-\tau}} X_k \leqslant \sqrt{2}\right\} = 1.$$

Although this estimation technique can be applied to *any* landscape in which a correlation length can be defined as above (even landscapes with non-commuting transformation groups, such as the travelling salesman problem), these bounds are directly applicable to landscapes such as peptide space, in which there are 2^N points. There, we can conclude that, even in correlated peptide landscapes with fitnesses scaled as above, the fitness of the global optimum stays roughly constant, but the typical improvement per step decreases to zero as N increases whenever a correlation length can be defined. Of course, Kauffman and Levin's uncorrelated version of peptide space is included as a special case in which we can take $\tau = 0$. It then follows that the globally maximal fitness in such a space, assuming that fitnesses are sampled from a normal distribution with zero mean and variance σ^2/N, must lie between $\sqrt{2\ln 2\sigma} \approx 1.177\sigma$ and $2\sqrt{\ln 2\sigma} \approx 1.665\sigma$.

Application to the *N-k* Model

In this section, we apply our theory to the N-k model, a simple mathematical model of a class of rugged landscapes introduced in Kauffman et al. (1988). This model assigns a fitness to an N residue peptide in protein space by first assigning a "fitness contribution", $f_i \in [0, 1]$, to the ith amino acid in the chain. Each amino acid fitness assignment depends, not just on the value of the ith amino acid, but also $k < N$ other amino acids. Once more, we make the simplifying assumption that each residue is one of only two possible amino acids, so that the N residue peptide can be treated as a string of N boolean bits. Thus, the fitness contribution of each residue, a function of $k + 1$ of the bits, can have 2^{k+1} possible values, one for each of the possible combinations of the $k + 1$ bits. The fitness contributions associated with each of these combinations is assigned by selecting an independent random variable from the uniform distribution on [0, 1]; that is, a "randomly chosen" number between zero and one. This generates a "fitness table" for the ith amino acid. There is a different, independently generated table for each of the N amino acids. Then, given any "string" of N amino acids, the total fitness of the string, F, is defined as the average of the fitness contributions of each part, (i.e. each of the f_i's) each in the context of the k others which impinge upon it. The use of the uniform distribution in assigning the fitness contributions can be interpreted as either an admission of ignorance of the true nature of the complex couplings between residues or as an attempt to capture the typical statistical properties of a wide class of landscapes with k interconnections per residue.

One other aspect of the N-k model must be specified; namely, which amino acid residues are coupled to which. (For the purpose of the model, two residues are coupled if one of them appears in the other's fitness table.) The simplest – but not the only – coupling is to take the k amino acid residues that affect residue i to be its nearest neighbors; that is, if k is even, the amino acids in residues $i - k/2$ thru residues $i + k/2$. We must then assign couplings for sites at the ends of the string, that is, sites i with $i \leqslant k/2$ and $i \geqslant N - k/2$. One way to do this is to introduce periodic boundary conditions. In other words, we assume that the sites are arranged in a circle, such that site N is next to site 1. Under this assumption, if $k = 2$, site N has neighbors $N - 1$ and 1, site 1 has neighbors N and 2, and the general site i has

neighbors $N-i-k/2 \bmod N \cdots N+i-k/2 \bmod N$. This assignment of the coupling gives rise to a class of statistical mechanical models known as short range spin glasses. (See Binder and Young (1986) for a general discussion of spin glasses.) Alternatively, we could assign the coupling by randomly selecting, for each site i, k other sites on the peptide to be used in forming the index to the ith fitness table. This assignment of couplings makes the model similar to a long range, dilute spin glass.

The N-k model affords a "tuneably rugged" fitness landscape, since tuning k alters how rugged the landscape is. This can be seen from the following discussion: For $k=0$, each site is independent of all other sites. Either the bit value 0 or the bit value 1 is almost surely "fitter" than the other; hence, a single specific sequence comprised of the fitter bit value in each position is almost surely the single, global optimum in the fitness landscape. Any other string is suboptimal, and lies on a connected walk via 1-mutant fitter variants to the global optimum by flipping bits from less fit to more fit values. The length of the walk is just the Hamming distance from the initial string to the global optimum. For a randomly chosen initial string, half the bits will be in their less fit state, hence the expected walk length is just $N/2$. A transition to a one mutant neighbor (i.e. the flip of a single bit) typically alters fitness by an amount $O(1/N)$. In contrast, the fully connected N-k model yields a completely random fitness landscape. For $k=N-1$, the fitness contribution of each site depends on all of the other sites because the "context" for each of the $N-1$ other bits is changed when even a single bit is flipped. In this case, therefore, the fitness of each N bit string is assigned a fitness that is statistically independent of its neighbors. As was shown in Kauffman and Levin (1987); Weinberger (1988); and Macken and Perelson (1989), such random landscapes have very many local optima, walks to optima are short ($O(\ln N)$), and only a small fraction of local optima are accessible from any initial string. Thus adaptive walks vary dramatically as the ruggedness of the landscape varies.

We now show that the sequence of fitnesses generated by a random walk on the N-k landscape can be approximated by an AR(1) process, as claimed above. We let f_j^s, with $-\infty < s < \infty$, be the site fitness of the jth site at step s, so that

$$F_s = \frac{1}{N}\sum_j f_j^s \tag{4}$$

is the fitness at step s. In the process of making step s, and thus flipping one of the bits in the N bit string, we change the context of other bits in the string, and thus the values of the site fitnesses associated with these bits. When neighborhoods are chosen via physical adjacency along the bit string, a single bit flip changes exactly $k+1$ site fitnesses. When the neighbors for each bit are randomly chosen sites along the string, the number of site fitnesses changed is a random variable with expectation $k+1$ and fluctuations of $O(\sqrt{k})$. We therefore expect that our conclusions do not depend on whether adjacent or random neighborhoods are chosen, and thus we denote the set of site fitnesses that are changed at step s as v_s in both cases. It is easy to check that the correlation between neighboring points in the landscape is $z=(N-(k+1)/N)$, motivating the representation

$$F_s = zF_{s-1} + \Delta F_s ,$$

where

$$\Delta F_s = \frac{1}{N}\sum_{j\in v_s}(f_j^s - f_j^{s-1} + F_{s-1}) .$$

What needs to be verified is that correlations in ΔF_s are sufficiently small that $z^s \approx R(s)$. To do this, we repeat the same steps used in Karlin and Taylor (1975) in deriving $R(s)=z^s$ in the strict AR(1) case and estimate the error introduced by the small correlations in the ΔF's. The first part of their derivation can be carried through verbatim, to conclude that

$$E[F_s F_{t+s}] = z^t + \lim_{l\to\infty}\sum_{\gamma=0}^{l-1}\sum_{\delta=0}^{t-1} z^{\gamma+\delta} E[\Delta F_{s-\gamma}\Delta F_{s+t-\delta}], \tag{5}$$

assuming that the f's have zero mean and unit variance. We continue to make this simplifying assumption throughout this derivation, because the same results obtain in the general case. If the ΔF's are strictly uncorrelated, the sum vanishes, and Karlin and Taylor's result follows. In the present case, the sum does not vanish, but does remain small, as we will now verify. We have

$$\begin{aligned}
&E[\Delta F_{s-\gamma}\Delta F_{s+t-\delta}] \\
&\quad = \frac{1}{N^2}E\left[\left\{\sum_{j\in v_{s-\gamma}}(f_j^{s-\gamma} - f_j^{s-\gamma-1} + F_{s-\gamma})\right\}\right. \\
&\qquad \left.\times\left\{\sum_{j\in v_{s+t-\delta}}(f_j^{s+t-\delta} - f_j^{s+t-\delta-1} + F_{s+t-\delta})\right\}\right] \\
&\quad = \frac{(k+1)^2}{N^2}E[F_{s-\gamma}F_{s+t-\delta}] \\
&\qquad + \frac{k+1}{N^2}E\left[F_{s-\gamma}\sum_{j\in v_{s+t-\delta}}(f_j^{s+t-\delta} - f_j^{s+t-\delta-1})\right] \\
&\qquad + \frac{k+1}{N^2}E\left[F_{s+t-\delta}\sum_{j\in v_{s-\gamma}}(f_j^{s-\gamma} - f_j^{s-\gamma-1})\right] \\
&\qquad + \frac{1}{N^2}E\left[\left\{\sum_{j\in v_{s-\gamma}}(f_j^{s-\gamma} - f_j^{s-\gamma-1})\right\}\right. \\
&\qquad \left.\times\left\{\sum_{j\in v_{s+t-\delta}}(f_j^{s+t-\delta-1} - f_j^{s+t-\delta})\right\}\right]
\end{aligned}$$

The first term is $O(k^2/N^3)$ because $E[F_{s-\gamma}F_{s+t-\delta}] < E[F_s^2] = 1/N$. For the next term, we see that

$$\begin{aligned}
&E\left[F_{s-\gamma}\sum_{j\in v_{s+t-\delta}}(f_j^{s+t-\delta} - f_j^{s+t-\delta-1})\right] \\
&\quad = \frac{1}{N}E\left[\sum_{j\in v_{s+t-\delta}} f_j^{s-\gamma}(f_j^{s+t-\delta} - f_j^{s+t-\delta-1})\right]
\end{aligned}$$

because $f_j^{s-\gamma}$ is statistically independent of the terms in parentheses if $k \notin \nu_{s+t-\delta}$. Thus, this term is $O(k^2/N^3)$, and, by a similar simplification, so is the next term. Finally, the last term is $O(k^2/N^3)$, because $\nu_{s-\gamma}$ and $\nu_{s+t-\delta}$ overlap with probability $(2k+1)/N$. If they fail to overlap, the last term is zero. If they do overlap, the expectation is at most $k+1$. Armed with these estimates, we have, for the sum in (5),

$$\lim_{t\to\infty} \sum_{\gamma=0}^{t-1} \sum_{\delta=0}^{t-1} z^{\gamma+\delta} E[\Delta F_{s-\gamma} \Delta F_{s+t-\delta}]$$
$$\leqslant O(k^2/N^3) \left(\sum_{\gamma=0}^{\infty} z^{\gamma} \right)^2$$
$$= O(k^2/N^3) \left(\frac{1}{1-z} \right)^2$$
$$= O(1/N)$$

Figure 1 shows the $r(s)$ function for 2048 step unbiased random walks on N-k landscapes. Here, $N = 96$, and $k = 2$, 8, and 48, and the "neighbors" upon which each site fitness depends are chosen randomly. These autocorrelations are estimated using the standard estimation techniques described in Priestley (1981). We chose to present semilog plots to emphasize the roughly exponential decay of the autocorrelation function in all cases. Furthermore, the autocorrelation function drops off more rapidly as k increases. This figure, as well as similar results obtained by choosing adjacent sites as neighbors, provides clear confirmation, both of the intuition behind our definition of correlation and the prediction that N-k landscapes should have exponentially decaying autocorrelation functions independent of the details of neighborhood assignment.

Furthermore, the measured correlation lengths from both adjacent and random landscapes, 25, 11, and 1.4, for $k = 2$, 8, and 48, are in general agreement with the correlation length prediction made above, $\tau = -1/\ln(1-(k+1)/N)$, which yields predictions of 31, 11, and 1.4 for the respective k values. Note that agreement is excellent except when $k = 2$, when two of the approximations used in estimating τ break kown: the ΔF's are no longer roughly Gaussian, and the correlation length is longer than $\sqrt{N}$ thus invalidating the approximation $r(s) \approx R(s)$. As the $k = 48$ case suggests, the autocorrelation function for the $k = 95$ case, one of Kauffman and Levin's "uncorrelated" landscapes (simulated, but not shown), is approximately a δ function, as we would expect.

With these results in hand, it is now a simple matter to substantiate the claim made above that N-k landscapes are generic members of the class of AR(1) landscapes: it is clear that such a landscape is completely characterized by the joint probability distribution of the fitnesses of each of its points. For AR(1) landscapes, this must be a multivariate Gaussian distribution with a covariance matrix given by (3), where $R(s)$ is determined by (2). The only free parameters in an AR(1) landscape are therefore the mean and variance of each of its points, which, by the assumption of statistical isotropy, must be the same for all points, and the rate at which $R(s)$ decays. It is clear that the parameters of the N-k landscape – k and the mean and variance of the site fitnesses – can be chosen such that the free parameters of the AR(1) landscape can assume arbitrary values.

In the previous section, we estimated the distribution of fitter neighbors of a configuration with fitness x.

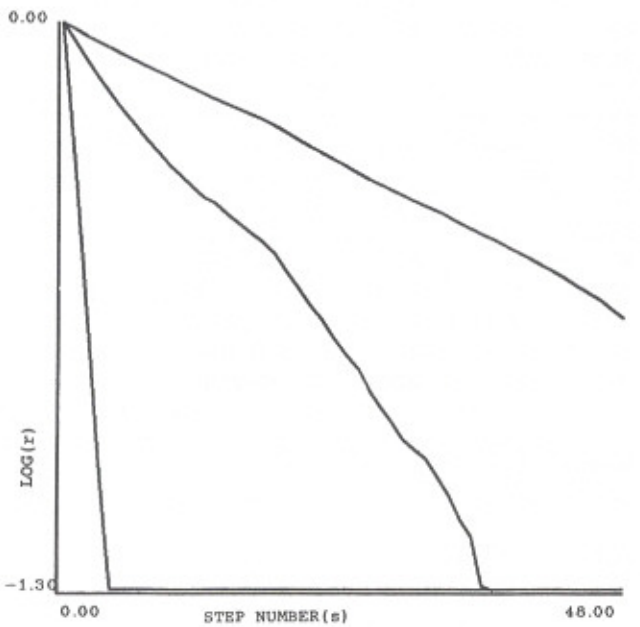

Fig. 1. Semi-log plot of autocorrelations of random walks on N-k landscapes for $N = 96$ and $k = 2$, 8, and 48. "Neighbors" of each site are chosen randomly with equal probability for all of the sites on the string

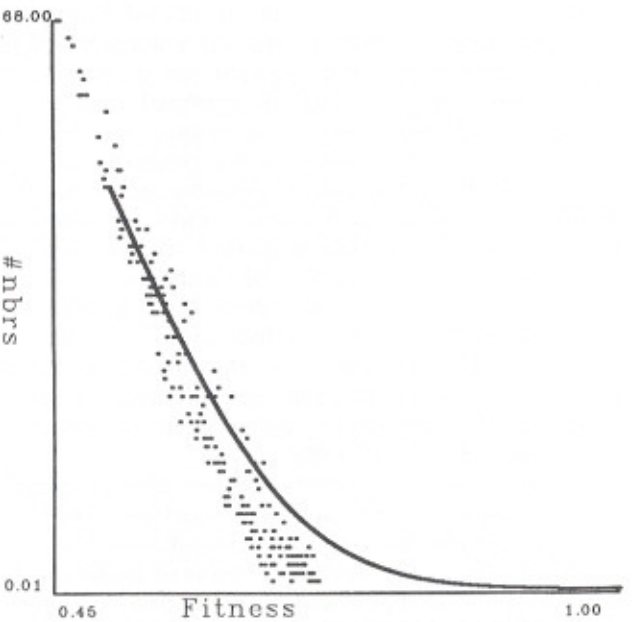

Fig. 2. Number of neighboring configurations that are fitter than an N-k configuration with a specified fitness. Points along the solid line are analytical predictions; unconnected points are from computer simulations

As a crude check of these results, we compare our estimates of the mean number of fitter neighbors of various configurations with computer simulation data, shown in Fig. 2. The simulations randomly selected a starting configuration, and then "walked uphill", by randomly selecting a *fitter* neighbor of the starting configuration, randomly selecting a fitter neighbor of the neighbor, etc. until a local optimum was attained. The coordinates of the dots in the figures represent the number of fitter neighbors as a function of the fitness attained, as compared with the predicted values indicated by the solid line. For $k \ll N$ and for larger values of k and relatively low fitness values, excellent agreement obtains. For $k = 48$, the k value shown in the figure, some deviation at high fitness values is expected because the Gaussian approximation to the distribution of fitnesses itself breaks down for the high fitnesses involved.

Conclusions and Directions for Future Research

In this last section, we discuss some intriguing implications of our notion of correlation. Perhaps the most obvious of these is the importance of corrleation in determining significant features of AR(1) landscapes. Besides the explicit results obtained above, our conclusion that N-k landscapes are generic members of the class of AR(1) landscapes strongly suggests that properties of N-k landscapes will generalize to this larger class. This is fortunate, because N-k landscapes are relatively easy to study, at least in the case of adjacent neighborhoods. We should note here that a relatively complete characterization of the local properties of N-k landscapes for fixed k and asymptotically large N has been accomplished by Isopi et al. who show that the expected number of loal optima in an N-k landscape grows exponentially with N, that there average length of uphill walks from a randomly chosen starting point grows linearly with N, that the expected fitnesses of local optima are asymptotically constant, and that the fluctuation about this average as N increases is $O(1/N)$.

Another structural feature of landscapes is the degree to which one can, in Kauffman and Levin's words, "jump beyond the correlation length of the landscape." For example, consider the N-k landscape for $k = 2$ and $N = 96$. Our previous results show that the smallest correlation between two configurations is approximately .13. This suggests that the landscape has a relatively few peaks of high fitness embedded in a *massif central*, a conclusion that has been confirmed by other means (Kauffman 1989).

We speculate that correlation is also likely to play a role in the study of diffusion on rugged landscapes, a topic that includes not only physical diffusion, but also "diffusion" due to the stochastic effects of mutations in evolutionary landscapes, and combinatorial optimization algorithms that simulate either physical diffusion or evolution. Up until now, we have focussed on the properties of random walks in which the transition probabilities are independent of the fitness differences between neighbors. In contrast, diffusion processes involve biased random walks in which the transition probabilities explicitly depend on these fitness differences, giving rise to diffusive barriers that are crossed with probability $e^{-\beta \Delta F}$, where ΔF is the fitness difference between neighbors, and β is a parameter that determines the strength of the stochastic component of the motion. Our results allow us to make the estimate

$$\Delta F \approx \sigma \sqrt{1 - R^2(1)}\,;$$

that is, ΔF should be on the order of the fluctuations in the conditional fitness distribution of the neighbors that was discussed in Sect. 3. Much of the computational effort involved in an optimization algorithm that is based on a diffusive strategy is devoted to crossing these barriers. Thus, this formula substantiates Stuart Kauffman's intuition (private communication) that it is difficult to optimize on a landscape with a short correlation length, because each point contains relatively little information regarding the fitness of surrounding points. Furthermore, this formula quantifies the degree to which evolution or a human algorithm designer can increase the speed of optimization by choosing among various possible neighborhood definitions.

Although it would seem that AR(1) landscapes represent a large class of fitness landscapes, we mention in passing that a still more general class of landscapes can be generated by specifying the whole of $R(s)$, rather than merely the decay parameter of an assumed decaying exponential $R(s)$. Indeed, we have already seen that, if the landscape is self-similar in the sense described previously, $R(s)$ is polynomial. This more general model, like the AR(1) model, uses easily measured, *global* properties of the landscape to estimate various local properties.

From a scientific, as opposed to a mathematical point of view, our efforts have three benefits. First, for those landscapes that are well approximated by the Gaussian model, the model drastically reduces the amount of experimental data that needs to be collected. Second, measurements of the autocorrelation function, as described above, should serve as a means of classifying different types of landscapes as a prelude to a theory of which optimization strategies work best on which of these types. Finally, and perhaps most interesting, is the possibility of inferring something of the detailed structure of the landscape in the form of a Fourier-like decomposition, as follows: Suppose each point on the landscape can be represented as a string of the form $B = b_1 b_2 \cdots b_N$, where b_i is either 0 or 1. The fitness at each point can then be interpreted as a mapping $F(B)$ from N bit strings to real numbers. It is easy to shown that *any* such function can be written as

$$F(B) = \alpha_0 + \sum_i \alpha_i b_i + \sum_{ij} \alpha_{ij} b_i b_j + \sum_{ijk} \alpha_{ijk} b_i b_j b_k + \cdots + \alpha_{12\ldots N} b_1 b_2 \cdots b_N\,.$$

where the α's are real numbers and each sum is taken over all possible distinct combinations of b's (of which there is only one involving $b_1, b_2, \ldots, b_N$). It is possible

to use the mean and autocorrelation function of the landscape to estimate the mean and correlation matrix of the α's, but this is properly the subject of another paper.

Another direction for future research is the possibility of a connection between the ideas presented in this paper and the theory of NP-completeness. Consider first the problem of optimization on an N-k landscape (e.g. Can a bit string be found with a fitness above a certain value?), which, for fixed k, random neighborhoods, and increasing N is probably NP-complete. However, our results make clear why N-k optimization cannot be accomplished in polynomial time: if k is sufficiently large, there are exponentially many configurations whose fitnesses are almost completely uncorellated, so that N-k optimization is essentially optimization on a list of exponentially many, independently distributed random real numbers. Presumably, any algorithm that would compute any property of the maximum of this list would have to examine each of the exponentially many numbers on the list.

Acknowledgements. It is a pleasure to thank Stuart Kauffman for useful discussions and to thank the Department of the Navy for partial support under ONR grant N00014-85-K-0258 PRIME P04, while the author was affilated with the University of Pennsylvania. The author also thanks the Max Planck Gesellschaft for support while completing the manuscript. Henry P. McKean asked for a proof of Theorem 1 as a question on a take-home mid-term examination in his course in probability offered in the Fall 1981 term at the Courant Institute of Mathematical Sciences in New York City.

Appendix 1

The theorem is proven by an appeal to the first Borel-Cantelli lemma, which states that events from an infinite list of possibilities can occur only finitely often if the probabilities of the different events on the list have a finite sum. We also need the fact that

$$\begin{aligned}\mathscr{P}\{X_n \leqslant x\} &= \int_{-\infty}^{x} \frac{e^{-x^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2 x}}\,dx \\ &= 1 - \frac{e^{-x^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2 x}}(1+o(x^{-1})).\end{aligned}$$

Both the Borel-Cantelli lemma and the estimate of the integral above are proven in Feller (1965).

The proof of the first result proceeds by showing first that the sequence of probabilities

$$\mathscr{P}\left\{\max_{k\leqslant N} X_k \leqslant (1-\delta)\sqrt{2\sigma^2 \ln N}\right\}$$

has a finite sum for all δ such that $0<\delta<1$. It will then follow from the first Borel-Cantelli lemma that event $\max_{k\leqslant N} X_k \leqslant (1-\delta)\sqrt{2\sigma^2 \ln N}$ will almost surely occur only finitely many times. The first result follows upon observing that δ can be chosen to be arbitrarily close to zero.

Because the X_k's are independent and identically distributed, the probability of the event $\max_{k\leqslant N} X_k \leqslant (1-\delta)\sqrt{2\sigma^2 \ln N}$ is the Nth power of the probability that any one of the X's satisfies the inequality, so that

$$\begin{aligned}&\mathscr{P}\left\{\max_{k\leqslant N} X_k \leqslant (1-\delta)\sqrt{2\sigma^2 \ln N}\right\} \\ &= \left[\int_{-\infty}^{(1-\delta)\sqrt{2\sigma^2 \ln N}} \frac{e^{-x^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}}\,dx\right]^N \\ &= \left[1 - \frac{e^{(1-\delta)^2 \ln N}}{(1-\delta)\sqrt{4\pi \ln N}}(1+o(1/\sqrt{\ln N}))\right]^N \\ &= \left[1 - \frac{1}{(1-\delta)N^{1-2\delta+\delta^2}\sqrt{4\pi \ln N}} \times (1+o(1/\sqrt{\ln N}))\right]^N \\ &\leqslant \exp\left[-\frac{N}{(1-\delta)N^{1-2\delta+\delta^2}\sqrt{4\pi \ln N}}(1+o(1/\sqrt{\ln N}))\right] \\ &\leqslant \exp\left[-\frac{N^{2\delta-\delta^2}}{\sqrt{4\pi(1-\delta)\ln N}}\right].\end{aligned}$$

The first inequality simply reflects the fact that $1-y<e^{-y}$ for small positive y, and the second reflects the fact that, for sufficiently large N,

$$(1+o(1/\sqrt{\ln N}))(1-\delta) < \sqrt{1-\delta}$$

Because

$$\frac{N^{2\delta-\delta^2}}{\sqrt{4\pi(1-\delta)\ln N}} > 2 \ln N$$

for fixed δ and sufficiently large N, the sequence of probabilities must eventually become bounded above by a constant multiple of $1/N^2$, and must therefore have a finite sum.

The second result follows by showing that the sequence of probabilities

$$\begin{aligned}&\mathscr{P}\left\{\max_{k\leqslant N} X_k \geqslant 2(1+\delta)\sigma\sqrt{\ln N}\right\} \\ &\quad = 1 - \mathscr{P}\left\{\max_{k\leqslant N} X_k \leqslant 2(1+\delta)\sigma\sqrt{\ln N}\right\}\end{aligned}$$

has a finite sum for all δ such that $0<\delta<1$. Once more,

$$\begin{aligned}&\mathscr{P}\left\{\max_{k\leqslant N} X_k \leqslant 2(1+\delta)\sigma\sqrt{\ln N}\right\} \\ &= \left[\int_{-\infty}^{2(1+\delta)\sigma\sqrt{\ln N}} \frac{e^{-x^2/2}}{\sqrt{2\pi\sigma^2}}\,dx\right]^N \\ &= \left[1 - \frac{e^{-2(1+\delta)^2 \ln N}}{(1+\delta)\sqrt{8\pi \ln N}}(1+o(1/\sqrt{\ln N}))\right]^N \\ &= \left[1 - \frac{1}{(1+\delta)N^{2(1+\delta)^2}\sqrt{8\pi \ln N}}(1+o(1/\sqrt{\ln N}))\right]^N \\ &> \exp -\frac{N}{(1+\delta)N^{2(1+\delta)^2}\sqrt{2\pi \ln N}}(1+o(1/\sqrt{\ln N})),\end{aligned}$$

where the inequality follows from the fact that $e^{-x}<1-x/2$. The fact that $1-e^{-x}<x$ allows us to

conclude that the sequence of probabilities is at most $O(1/(N^{1+2\delta+2\delta 2}\sqrt{\ln N})$, which has a convergent sum.

The corollaries also follow immediately from the above, when it is observed that the argument above holds when σ^2 is replaced by σ^2/N and N is replaced by $2^{N-\tau}$ in the appropriate places.

References

Binder K, Young A (1986) Spinglasses: experimental facts, theoretical concepts, and open questions. Rev Mod Phys 54:801

Brady R (1985) Optimization strategies gleaned from biological evolution. Nature 317:804

Breiman L (1968) Probability. Addison–Wesley, Reading, Mass

David H (1970) Order statistics. Wiley, New York

Eigen M, McCaskill J, Schuster P (1988) Molecular quasi-species. J Phys Chem 92:6881

Feller W (1965) An introduction to probability theory and its applications, vol 1, 3rd edn. Wiley, New York

Feller W (1972) An introduction to probability theory and its applications, vol 2, 2nd edn. Wiley, New York

Garey M, Johnson D (1979) Computers and intractability: a guide to the theory of incomputability. Freeman, San Francisco

Gumbel E (1985) Statistics of extremes. Colorado University, New York

Herstein I (1964) Topics in algebra. Blaisdell, New York

Holland J (1981) Genetic algorithms and adaptation. Technical Report # 34, University of Michigan Cognitive Sciences Department Ann Arbor, Mich

Karlin S, Taylor H (1975) A first course in stochastic processes. Academic Press, New York

Karlin S, Taylor H (1981) A second course in stochastic processes. Academic Press, New York

Kauffman S, Levin S (1987) Towards a general theory of adaptive walks on rugged landscapes. J Theor Biol 128:11–45

Kauffman S, (1989) Complex systems. In: Stein DL (ed) Proceedings of the 1988 Summer School on Complex Systems in Santa Fe, NM. Santa Fe Institute Studies in the Sciences of Complexity, Addison-Wesley, Reading, Mass

Kauffman S, Weinberger E, Perelson A (1988) Maturation of the immune response via adaptive walks on affinity landscapes. In: Perelson AS (ed) Theoretical immunology, part I. Santa Fe Institute Studies in the Sciences of Complexity. Addison-Wesley, Reading, Mass

Kirkpatrick S, Gelatt C, Vecci M (1983) Optimization by simulated annealing. Science 220:671–680

Kirkpatrick S, Toulouse G (1985) Configuration space analysis of Travelling Salesman Problems. J Phys 46:1277

Macken, C, Perelson, A(1989) Protein evolution on rugged landscapes. Proc Natl Acad Sci USA 86:6191–6195

Mandelbrot B (1982) The fractal geometry of nature. Freeman, New York

Palmer RG (1989) Complex systems. In: Stein DL (ed) Proceedings of the 1988 Summer School on Complex Systems in Santa Fe, NM. Santa Fe Institute Studies in the Sciences of Complexity. Addison-Wesley, Reading, Mass

Papoulis A (1965) Probability, random variables, and stochastic processes. McGraw-Hill, New York

Priestley M (1981) Spectral analysis and time series. Academic Press, London

Smith JM (1970) Natural selection and the concept of a protein space. Nature 225:563

Stein DL (1989) Complex systems. In: Stein DL (ed) Proceedings of the 1988 Summer School on Complex Systems in Santa Fe, NM. Santa Fe Institute Studies in the Sciences of Complexity. Addison-Wesley, Reading, Mass

Weinberger E (1987a) A stochastic generalization of Eigen's model of natural selection. PhD Thesis, Courant Institute of Mathematical Sciences, New York, NY Available from University Microfilms, Ann Arbor, Mich

Weinberger E (1987b) A model of natural selection that exhibits a dynamic phase transition. J Stat Phys 49:1011–1028

Weinberger E (1988) A more rigorous derivation of some results on rugged fitness landscapes. J Theor Biol 134:125–129

White SR (1984) Concepts of scale in simulated annealing. Proc ICCD

Wright S (1932) The roles of mutation, inbreeding, crossbreeding and selection in evolution. In: Proceedings 6th Congress on Genetics 1:356

Dr. Ed Weinberger
Max-Planck-Institut für biophysikalische Chemie
Postfach 2841
Am Fassberg
D-3400 Göttingen-Nikolausberg
Federal Republic of Germany