

A sharp minimum on the mean number of steps taken in adaptive walks

Noah A. Rosenberg*

Department of Human Genetics and Bioinformatics Program, University of Michigan, 2017 Palmer Commons, Box 2218, 100 Washtenaw Avenue, Ann Arbor, MI 48109-2218, USA

Received 12 August 2004; received in revised form 16 March 2005; accepted 22 March 2005

Available online 23 June 2005

Communicated by David Krakauer

Abstract

It was recently conjectured by H.A. Orr [2003. A minimum on the mean number of steps taken in adaptive walks. *J. Theor. Biol.* 220, 241–247] that from a random initial point on a random fitness landscape of alphabetic sequences with one-mutation adjacency, chosen from a larger class of landscapes, no adaptive algorithm can arrive at a local optimum in fewer than on average $e - 1$ steps. Here, using an example in which the mean number of steps to a local optimum equals $(A - 1)/A$, where A is the number of distinct “letters” in the “alphabet” from which sequences are constructed, it is shown that as originally stated, the conjecture does not hold. It is also demonstrated that $(A - 1)/A$ is a sharp minimum on the mean number of steps taken in adaptive walks on fitness landscapes of alphabetic sequences with one-mutation adjacency. As the example that achieves the new lower bound has properties that are not often considered as potential attributes for fitness landscapes—non-identically distributed fitnesses and negative fitness correlations for adjacent points—a weaker set of conditions characteristic of more commonly studied fitness landscapes is proposed under which the lower bound on the mean length of adaptive walks is conjectured to equal $e - 1$.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: Adaptive landscape; Gradient adaptation; *NK* model

1. Introduction

In a modern interpretation of Wright’s (1932) adaptive landscape metaphor, evolution of a population occurs on a discrete space of sequences, each of which is associated with a fitness value (Orr, 2005). The population is viewed as being located at a particular point in the space. Substitutions in the sequence then occur, each of which shifts the population to a point adjacent to its current location, so that in any given step, the population is more likely to move to neighboring points of higher fitness than to those of lower fitness. The space of sequences, together with the function that

assigns fitnesses to sequences, is termed the *fitness landscape*. The relation that describes which sequences are adjacent in the space induces a metric for the landscape. A path through the space that terminates upon reaching a local fitness optimum is termed an *adaptive walk*. If in an adaptive walk, moves from a point P always transit to the highest-fitness point among the neighbors of P , a path is a *gradient adaptive walk*.

In a recent article, Orr (2003) argued that if fitnesses are assigned randomly to sequences in such a way that all sequences have identically distributed fitness, and if adjacent points in sequence space are those that differ at exactly one position in the sequence, then for “long” sequences, the mean number of steps taken in a gradient adaptive walk from an initial sequence to a local optimum is $e - 1$. This result was obtained as an

*Tel.: +1 734 615 9556; fax: +1 734 615 6553.

E-mail address: rnoah@umich.edu.

approximation by considering infinitely long sequences. However, the approximation of $e - 1$ held well in simulated fitness landscapes with sequences of finite length.

Orr noted that in comparison with other adaptive algorithms, gradient adaptation is efficient at reaching local optima. He also observed that landscapes with identically distributed fitnesses are likely to produce shorter adaptive walks than are those in which neighboring points have positively correlated fitnesses: such correlations will cause optima to be clustered in locations that are difficult to access from points with low fitness. These observations together with the remarkably small mean number of steps for gradient adaptive walks suggested the following conjecture (Orr, 2003, p. 246): “No adaptive algorithm on any class of fitness landscape can arrive at a local optimum in fewer than a mean of $e - 1$ steps, given random starting points and long sequences.” In this statement a *class of landscapes* refers to a random field on the sequence space, so that any given landscape is a particular realization of a class of landscapes.

Here, after formalizing Orr’s results on the distribution of walk lengths for the case of independent and identically distributed fitnesses, it is then found that as originally stated, the conjecture does not hold. The minimum mean length of adaptive walks is sharpened from $e - 1$ to $(A - 1)/A$, where A is the size of the “alphabet” from which “letters” in sequences are chosen. In other words, it is shown that no adaptive algorithm on any fitness landscape of alphabetic sequences can arrive at a local optimum in fewer than a mean of $(A - 1)/A$ steps, given random starting points and sequences of any length. Because no landscape has a mean adaptive walk length smaller than $(A - 1)/A$, it is also impossible for any *class* of landscapes to have a mean adaptive walk length less than $(A - 1)/A$. After this improved lower bound is obtained, an example of a class of fitness landscape that achieves the bound is presented, demonstrating that the new bound is sharp. The hypotheses of Orr’s conjecture are then weakened, giving rise to a more restricted set of conditions under which the conjectured lower bound on the mean length of adaptive walks equals $e - 1$.

2. Definitions

The mean length of adaptive walks is a property of a finite space of points together with an adjacency relation, a fitness function, and an adaptive algorithm. Of interest here is “alphabetic sequence space” with the “one-mutation” adjacency relation.

Consider an alphabet \mathbb{A} with $A \geq 2$ letters, denoted a_0, a_1, \dots, a_{A-1} . These are the letters that can potentially be part of a sequence. As an example, for DNA sequences,

$\mathbb{A} = \{\text{adenine, cytosine, guanine, thymine}\}$. Each sequence has length N , with each letter taken from the alphabet \mathbb{A} . For $N \geq 1$, the set of sequences of length N is $\mathbb{A} \times \dots \times \mathbb{A} = \mathbb{A}^N$, as each element in the alphabet can be in any position in a sequence. The number of sequences in \mathbb{A}^N is A^N . For a sequence $Y \in \mathbb{A}^N$, Y_i denotes the i th letter of the sequence. Here \mathbb{A}^N is termed the space of *alphabetic sequences* of length N for alphabet \mathbb{A} . Note that it is possible to view any space of points on which adaptive walks can occur as an alphabetic sequence space with $N = 1$ and with an exact correspondence between letters of the alphabet and points in the space.

Given a space of points S , an *adjacency relation* \bowtie is a relation between pairs of points in S that satisfies (i) for $Y, Z \in S$, $Z \bowtie Y$ if and only if $Y \bowtie Z$, and (ii) it is never true that $Y \bowtie Y$. To ensure that S with adjacency relation \bowtie is a connected graph, for this article, it is assumed that adjacency relations satisfy a third condition: (iii) for $Y, Z \in S$, if it is not true that $Y \bowtie Z$, then there must exist points $Y^{(1)}, Y^{(2)}, \dots, Y^{(k)} \in S$ such that $Y \bowtie Y^{(1)}, Y^{(1)} \bowtie Y^{(2)}, \dots, Y^{(k-1)} \bowtie Y^{(k)}$, and $Y^{(k)} \bowtie Z$. Two distinct alphabetic sequences $Y, Z \in A^N$ are *one-mutation adjacent* if $Y_i = Z_i$ for all except one value of $i \in \{1, 2, \dots, N\}$, that is, if a single mutation at one position can transform one of the sequences into the other. The one-mutation adjacency relation is denoted \sim , so that if Y and Z are one-mutation adjacent, it is written $Y \sim Z$ (or $Z \sim Y$).

Given the space of points S with adjacency relation \bowtie , let $f : \mathbb{A}^N \rightarrow [0, \infty)$ be a *fitness function*. A sequence $Y \in S$ is a *local optimum* of f if $f(Y) > \max_{Z \bowtie Y} f(Z)$. To avoid trivial cases (such as a situation in which all points have equal fitness), it is assumed for any fitness function f that if $Y, Z \in S$ and $Y \bowtie Z$, then $f(Y) \neq f(Z)$.

Given the space of points S with adjacency relation \bowtie and fitness function f , an *adaptive algorithm* Q is a procedure that, given a starting point $Y^{(0)} \in S$, chooses points $Y^{(1)}, \dots, Y^{(L)} \in S$, such that (i) $Y^{(l-1)} \bowtie Y^{(l)}$ for each $l \geq 1$, (ii) $Y^{(L)}$ is a local optimum of f , and (iii) $Y^{(l)}$ is not a local optimum of f for $l < L$. The collection $Y^{(0)}, Y^{(1)}, \dots, Y^{(L)}$ is an *adaptive walk* with length L . The *gradient adaptive algorithm*, denoted Q^* , chooses $Y^{(l)} = \operatorname{argmax}_{Z \bowtie Y^{(l-1)}} f(Z)$.

Finally, let $\mathcal{L}(S, \bowtie, f, Q)$ denote the mean length of adaptive walks on space S with adjacency relation \bowtie , fitness function f , and adaptive algorithm Q .

3. Independent and identically distributed fitnesses

Consider \mathbb{A}^N with one-mutation adjacency. Let h be a fitness function on \mathbb{A}^N that assigns independent and identically distributed fitnesses to all points. Let $p_J(N)$ denote the probability that a gradient adaptive walk in \mathbb{A}^N terminates on the J th step.

The probability that the starting sequence $Y^{(0)}$ for a gradient adaptive walk is a local optimum is the probability that it has higher fitness than all of its one-mutant neighbors, or, assuming independent and identically distributed fitnesses (Kauffman and Levin, 1987)

$$p_0(N) = \frac{1}{(A-1)N+1}. \tag{1}$$

Suppose now that $Y^{(0)}$ is not a local optimum. Consider an adaptive walk $Y^{(0)}, Y^{(1)}, \dots, Y^{(J)} \in \mathbb{A}^N$, with $Y^{(0)} \sim Y^{(1)}, \dots, Y^{(J-1)} \sim Y^{(J)}$. Let $\alpha_J(N)$ be the probability that an adaptive walk that has reached $Y^{(J)}$ continues rather than terminates. If $J \geq 1$, an adaptive walk can reach $Y^{(J)}$ only if it has reached each of $Y^{(0)}, Y^{(1)}, \dots, Y^{(J-1)}$. Thus, the probability that an adaptive walk has length greater than or equal to J , denoted $\phi_J(N)$, equals $\prod_{j=0}^{J-1} \alpha_j(N)$. For $J \geq 1$, the probability that the walk has length exactly J equals $\phi_J(N) - \phi_{J+1}(N)$, or

$$p_J(N) = [1 - \alpha_J(N)] \prod_{j=0}^{J-1} \alpha_j(N). \tag{2}$$

The quantity $\alpha_J(N)$ is the probability that at least one of the points adjacent to $Y^{(J)}$ has a higher fitness than $Y^{(J)}$ itself. Equivalently, for $J \geq 1$, it equals the probability that the point with the highest fitness among all points adjacent to any of $Y^{(0)}, \dots, Y^{(J)}$ is adjacent to $Y^{(J)}$, but is not adjacent to any of $Y^{(0)}, \dots, Y^{(J-1)}$. Because all points have independent and identically distributed fitnesses, $\alpha_J(N)$ is the ratio of the number of points adjacent to $Y^{(J)}$ but not adjacent to any of $Y^{(0)}, \dots, Y^{(J-1)}$ (and not equal to $Y^{(0)}$) and the number of points adjacent to one or more of $Y^{(0)}, \dots, Y^{(J)}$. In

other words, it is the quotient of the number of “new” points adjacent to $Y^{(J)}$ and the total number of points explored.

For small J , the ratio $\alpha_J(N)$ is straightforward to compute exactly (Table 1). For larger J , Orr (2003) used an approximation to evaluate $\lim_{N \rightarrow \infty} \alpha_J(N) = 1/(J+1)$, from which he obtained $\lim_{N \rightarrow \infty} p_J(N) = J/(J+1)!$ via Eq. (2). Here Orr’s argument is formalized to verify the result.

Proposition 1 (Orr). For each integer $J \geq 0$, $\lim_{N \rightarrow \infty} p_J(N) = J/(J+1)!$.

Proof. For $J = 0$, the result is clear from Eq. (1). Suppose now that $J \geq 1$ and $N > J$. To place bounds on $\alpha_J(N)$, note that each of $Y^{(1)}, \dots, Y^{(J)}$ differs in one sequence position from the previous point. Therefore, at most J sequence positions vary among $Y^{(0)}, Y^{(1)}, \dots, Y^{(J)}$. A point adjacent to $Y^{(J)}$ that differs from $Y^{(J)}$ in one of these J positions might be adjacent to one or more of $Y^{(0)}, \dots, Y^{(J-1)}$. However, any point adjacent to $Y^{(J)}$ that differs from $Y^{(J)}$ in one of the remaining $N - J$ positions cannot be adjacent to any of $Y^{(0)}, \dots, Y^{(J-1)}$. Thus, the number of points adjacent to $Y^{(J)}$ and not adjacent to any of $Y^{(0)}, \dots, Y^{(J-1)}$ is at least $(N - J)(A - 1)$, but at most $N(A - 1)$.

The number of points adjacent to one or more of $Y^{(0)}, \dots, Y^{(J)}$ is greater than or equal to $(J + 1)(N - J)(A - 1)$, as each of the $J + 1$ points has at least $(N - J)(A - 1)$ neighbors that cannot be adjacent to any of the other J points. The number of points adjacent to one or more of $Y^{(0)}, \dots, Y^{(J)}$ is at most $(J + 1)N(A - 1)$.

Combining the upper and lower bounds for the numerator (new points) and denominator (total points)

Table 1

The probability $p_J(N)$ for alphabet size $A \geq 2$ that a gradient adaptive walk on \mathbb{A}^N (with one-mutation adjacency and independently and identically distributed fitnesses) lasts exactly J steps (assuming $A^N > J$)

J	$\alpha_J(N)$	$\lim_{N \rightarrow \infty} \alpha_J(N)$	$p_J(N)$	$\lim_{N \rightarrow \infty} p_J(N)$
0	$\frac{(A-1)N}{(A-1)N+1}$	1	$\frac{1}{(A-1)N+1}$	0
1	$\frac{(A-1)(N-1)}{(A-1)(2N-1)+1}$	1/2	$\frac{(A-1)N}{(A-1)(2N-1)+1}$	1/2
2	$\frac{(A-1)(N-1)-1}{(A-1)(3N-2)}$	1/3	$\frac{(A-1)N(N-1)}{[(A-1)N+1](3N-2)}$	1/3
3	$\frac{(A-1)(N-1)-1}{(A-1)(4N-3)-1}$	1/4	$\frac{(A-1)^2 N(N-1)[(A-1)(N-1)-1]}{[(A-1)N+1][(A-1)(2N-1)+1][(A-1)(4N-3)-1]}$	1/8
4	$\frac{(A-1)(N-1)-1}{(A-1)(5N-4)-2}$	1/5	$\frac{(A-1)N(N-1)[(A-1)(N-1)-1]^2}{[(A-1)N+1][(A-1)(2N-1)+1](3N-2)[(A-1)(5N-4)-2]}$	1/30

Values of $p_J(N)$ are obtained from values of $\alpha_J(N)$ using Eq. (2). To compute $\alpha_J(N)$, without loss of generality, assume $\mathbb{A} = \{0, 1, \dots, A - 1\}$ and $Y^{(0)} = 000\dots 0$. The formula for $\alpha_0(N)$ follows from the fact that all points in \mathbb{A}^N have exactly $(A - 1)N$ neighbors. To compute $\alpha_1(N)$, suppose that a walk lasts at least one step. Without loss of generality, assume $Y^{(1)} = 100\dots 0$. Of the $(A - 1)N$ neighbors of $Y^{(1)}$, the $A - 1$ sequences obtained by modifying the first position of $Y^{(1)}$ include $A - 2$ neighbors of $Y^{(0)}$ as well as $Y^{(0)}$ itself. Therefore, the number of new points adjacent to $Y^{(1)}$ is $(A - 1)(N - 1)$. The total number of points explored is $[(A - 1)N + 1] + (A - 1)(N - 1)$, from which the formula for $\alpha_1(N)$ follows. To compute $\alpha_2(N)$, suppose that a walk lasts at least two steps. Without loss of generality assume $Y^{(2)} = 110\dots 0$. Of the $(A - 1)N$ neighbors of $Y^{(2)}$, the $A - 1$ sequences obtained by modifying the second position of $Y^{(2)}$ include $A - 2$ neighbors of $Y^{(1)}$ as well as $Y^{(1)}$ itself. The sequence $010\dots 0$ is a neighbor of $Y^{(0)}$. Therefore, the number of new points adjacent to $Y^{(2)}$ is $(A - 1)(N - 1) - 1$. The total number of points explored is $[(A - 1)(2N - 1) + 1] + (A - 1)(N - 1) - 1$, from which the formula for $\alpha_2(N)$ follows. With $J > 2$ the calculation is similar, except that multiple cases for $Y^{(j-1)}$ must be considered.

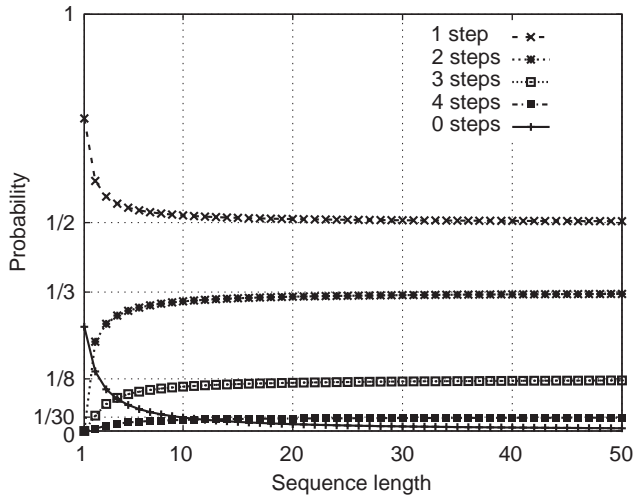


Fig. 1. As functions of sequence length (N), the probabilities $p_J(N)$ of gradient adaptive walks on \mathbb{A}^N (with one-mutation adjacency, independently and identically distributed fitnesses, and alphabet size $A = 4$) of lengths $J = 0, 1, 2, 3$, and 4 steps.

of $\alpha_J(N)$, for $1 \leq J < N$, we have

$$\frac{N - J}{(J + 1)N} \leq \alpha_J(N) \leq \frac{N}{(J + 1)(N - J)}. \tag{3}$$

Note also from Eq. (1) that

$$\alpha_0(N) = \frac{(A - 1)N}{(A - 1)N + 1}. \tag{4}$$

Combining Eqs. (2)–(4),

$$\begin{aligned} \frac{N - J - 1}{N - J} \frac{(A - 1)N}{(A - 1)N + 1} \frac{J}{(J + 1)!} \frac{N!}{(N - J)!N^J} &\leq p_J(N) \\ &\leq \frac{N + 1}{N} \frac{(A - 1)N}{(A - 1)N + 1} \frac{J}{(J + 1)!} \frac{(N - J)!N^J}{N!}. \end{aligned} \tag{5}$$

Finally, $p_J(N)$ is bounded between two functions whose limits both equal $J/(J + 1)!$ as $N \rightarrow \infty$. \square

The convergence of $p_J(N)$ to $J/(J + 1)!$ occurs rapidly, so that values of $p_J(N)$ are quite close to their large- N limits even for small sequence lengths (Fig. 1).

4. General landscapes

It is now shown that for general f and Q , $\mathcal{L}(\mathbb{A}^N, \sim, f, Q) \geq (A - 1)/A$. The proof proceeds first by demonstrating that if M is the number of local optima in \mathbb{A}^N , then no adaptive algorithm can produce adaptive walks of mean length less than $(A^N - M)/A^N$. Next, it is shown that with the adjacency relation \sim , the maximum number of local optima in \mathbb{A}^N equals A^{N-1} .

Lemma 2. Given that the number of local optima of a fitness function f on the alphabetic sequence space \mathbb{A}^N equals M , for any adjacency relation \triangleright and any adaptive

algorithm Q , the mean length of an adaptive walk from a random starting point in \mathbb{A}^N , $\mathcal{L}(\mathbb{A}^N, \triangleright, f, Q)$, is greater than or equal to $(A^N - M)/A^N$.

Proof. For any adaptive algorithm, the length of an adaptive walk is 0 if it begins at a local optimum, and at least 1 otherwise. Thus, the mean length of an adaptive walk from a random starting point is $\mathcal{L} \geq [M \cdot 0 + (A^N - M) \cdot 1]/A^N$. \square

Lemma 3. For any fitness function f , with adjacency relation \sim for the alphabetic sequence space \mathbb{A}^N , the number M of local optima in \mathbb{A}^N is less than or equal to A^{N-1} .

Proof. Suppose $M \geq A^{N-1} + 1$. There are A^{N-1} possible values for the first $N - 1$ letters of a sequence in \mathbb{A}^N . Therefore, there exist two local optima Y and Z for which $Y_1 = Z_1, \dots, Y_{N-1} = Z_{N-1}$. Y and Z are then one-mutation adjacent, contradicting the definition of local optimum. \square

Proposition 4. For any fitness function f and any adaptive algorithm Q , with adjacency relation \sim for \mathbb{A}^N , the mean length of an adaptive walk from a random starting point in \mathbb{A}^N is greater than or equal to $(A - 1)/A$.

Proof. By Lemmas 2 and 3, $\mathcal{L}(\mathbb{A}^N, \sim, f, Q) \geq (A^N - M)/A^N \geq (A^N - A^{N-1})/A^N = (A - 1)/A$. \square

To demonstrate that $(A - 1)/A$ is a sharp lower bound, it remains to produce an example of a fitness function and adaptive algorithm on \mathbb{A}^N (with adjacency relation \sim), for which $\mathcal{L} = (A - 1)/A$. Let $C_{a_k} : \mathbb{A}^N \rightarrow \{0, 1, \dots, N\}$ be the function that counts the number of times that the letter $a_k \in \mathbb{A}$ appears in a sequence. That is, for $Y \in \mathbb{A}^N$, $C_{a_k}(Y) = \sum_{i=1}^N \delta_{Y_i, a_k}$, where δ_{Y_i, a_k} is the Kronecker delta (1 if $Y_i = a_k$, 0 otherwise). Suppose $U : \mathbb{A}^N \rightarrow (0, 1)$ is a uniform random variable, and suppose $g : \mathbb{A}^N \rightarrow \{0, 1, \dots, (A - 1)N\}$ is a function so that for $Y \in \mathbb{A}^N$, $g(Y) = \sum_{k=0}^{A-1} k C_{a_k}(Y)$. Let $f^* : \mathbb{A}^N \rightarrow (0, A)$ be given by $f^*(Y) = g(Y) \pmod{A} + U(Y)$. Then f^* is an instance of a random field on \mathbb{A}^N ; in other words, together with \mathbb{A}^N , f^* produces a specific instance of a statistical class of landscapes. This class of landscapes, denoted F , encompasses all possible realizations of f^* on the sequence space \mathbb{A}^N . The following proposition shows that any fitness function $f^* \in F$, together with the gradient adaptive algorithm Q^* , produces adaptive walks on \mathbb{A}^N with mean length $(A - 1)/A$.

Proposition 5. For any fitness function $f^* \in F$, $\mathcal{L}(\mathbb{A}^N, \sim, f^*, Q^*) = (A - 1)/A$.

Lemma 6. If Y is a randomly chosen point in \mathbb{A}^N , and $k \in \{0, 1, \dots, A - 1\}$, then $\mathbb{P}[g(Y) \pmod{A} = k] = 1/A$.

Proof. The result is clear for $N = 1$, as each letter in \mathbb{A} appears in exactly one sequence. Suppose the result holds

for sequences of length N . For any sequence $Y \in \mathbb{A}^N$, there are N sequences in \mathbb{A}^{N+1} obtained by appending a letter to Y . Choosing one of these N sequences at random, say Z , then $\mathbb{P}[g(Z)(\text{mod } A) = k] = \mathbb{P}[[g(Y) + g(Z_{N+1})](\text{mod } A) = k]$, as g for a sequence is easily shown to be the sum of the values of g for disjoint component subsequences. $\mathbb{P}[[g(Y) + g(Z_{N+1})](\text{mod } A) = k] = \mathbb{P}[g(Y)(\text{mod } A) = [k - g(Z_{N+1})](\text{mod } A)]$. Because $[k - g(Z_{N+1})](\text{mod } A) \in \{0, 1, \dots, A - 1\}$, the inductive hypothesis applies and $\mathbb{P}[g(Z)(\text{mod } A) = k] = 1/A$. \square

Proof of Proposition 5. Choose a random point $Y \in \mathbb{A}^N$ and consider two cases: (i) $g(Y)(\text{mod } A) = A - 1$, and (ii) $g(Y)(\text{mod } A) \neq A - 1$.

- (i) $g(Y)(\text{mod } A) = A - 1$. Then $f^*(Y) > A - 1$. Consider $Z \in \mathbb{A}^N$ with $Z \sim Y$, $Y_j = a_p$, $Z_j = a_q$, and $p \neq q$. Then $g(Z) = g(Y) + q - p$, as Z contains one more a_q and one less a_p than does Y . Consequently, $f^*(Z) = [g(Y) + q - p](\text{mod } A) + U(Z)$. Because $q \neq p$, $[g(Y) + q - p](\text{mod } A) \leq A - 2$, and because $U(Z) < 1$, $f^*(Z) < A - 1$. Thus Y has higher fitness than any point one-mutation adjacent to it, and is a local optimum.
- (ii) $g(Y)(\text{mod } A) \neq A - 1$. Then $f^*(Y) < A - 1$. Consider $Z \in \mathbb{A}^N$ with $Z \sim Y$, $Y_j = a_p$, $Z_j = a_q$, and $q = [p - 1 - g(Y)](\text{mod } A)$. Note that $Z_j \neq Y_j$, as $g(Y)(\text{mod } A) \neq A - 1$. Then $g(Z)(\text{mod } A) = g(Y)(\text{mod } A) + [p - 1 - g(Y)](\text{mod } A) - p(\text{mod } A) = A - 1$. Using case (i), Z is a local optimum. Such an optimum can be obtained for any j , so that Y is one-mutation adjacent to N local optima. In a gradient adaptive walk, a move from Y proceeds to the point with the highest fitness among these N optima.

Applying Lemma 6 with case (i), the probability that an adaptive walk begins at a local optimum is $1/A$. If the adaptive walk does not begin at a local optimum, using case (ii), it reaches one in its first step. Therefore, $\mathcal{L}(\mathbb{A}^N, \sim, f^*, Q^*) = (1/A) \cdot 0 + (1 - 1/A) \cdot 1$. \square

This example can be illustrated with the special case $\mathbb{A} = \{0, 1\}$ (Fig. 2). Length- N sequences with an odd number of ones have fitness in $(1, 2)$, and those with an even number of ones have fitness in $(0, 1)$. Any “odd” sequence is a local optimum, because any mutation converts it to an “even” sequence. Any “even” sequence is one step away from N local optima, because all sequences one-mutation adjacent to it are odd. Because half of all sequences are even and half are odd, the mean number of steps to a local optimum is $1/2$.

5. Orr’s conjecture

A tighter lower bound of $(A - 1)/A$ for the mean length of adaptive walks on fitness landscapes of

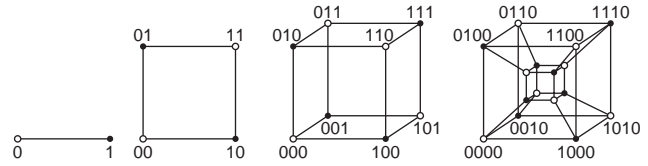


Fig. 2. Local optima for a fitness function $f^* \in F$ on spaces \mathbb{A} , \mathbb{A}^2 , \mathbb{A}^3 , and \mathbb{A}^4 , with adjacency relation \sim , for $\mathbb{A} = \{0, 1\}$. Shaded circles denote local optima. The sequence represented by a point P in the inner cube of \mathbb{A}^4 is obtained from the sequence for the point to which P is adjacent in the outer cube by replacing the last 0 with 1.

alphabetic sequences with one-mutation adjacency has been found and has been shown to be achieved under the gradient adaptive algorithm by a specific class of landscapes. However, as there may still exist conditions on fitness landscapes that guarantee that the minimum mean length of adaptive walks is no less than $e - 1$, it is instructive to consider three assumptions that, when stated explicitly, produce an emendation to Orr’s conjecture for the lower bound.

To obtain the $e - 1$ bound, it was assumed in Orr’s article that sequences are “long.” The purpose of this assumption was twofold. First, long sequences are required in order for the overlap of neighborhoods of the points in the adaptive walk to be small in comparison with the total number of neighbors that the points have. Second, if sequences are long, for fitness functions in which the limit as $N \rightarrow \infty$ of the proportion of points in the space that are local optima equals zero, the probability that the starting point of an adaptive walk is a local optimum is negligible. For the counterexamples in Proposition 5, however, although sequences can have arbitrary lengths, the proportion of local optima in the space is a constant that does not depend on N . Thus, the “long sequences” assumption does not take on the role of reducing the probability of starting at a local optimum.

Orr also implicitly assumed that fitnesses are identically distributed. However, in the counterexample, fitnesses are not identically distributed, as the space is partitioned into disjoint sections such that points in different sections do not have identical distributions.

Finally, Orr restricted his discussion to scenarios in which adjacent points had independent or positively correlated fitnesses, such as Nk -landscapes (Kauffman and Weinberger, 1989; Palmer, 1991). In the counterexample, however, fitnesses are in fact negatively correlated.

Consequently, it seems sensible to amend Orr’s conjecture. Recall that for any $N \geq 1$, with one-mutation adjacency and any fitness function, no adaptive algorithm starting at a random point in \mathbb{A}^N can arrive at a local optimum in fewer than a mean of $(A - 1)/A$ steps. A revised conjecture is as follows:

For any $\varepsilon > 0$, with one-mutation adjacency and any fitness function that satisfies (i)–(iii), N can be chosen sufficiently large so that no adaptive algorithm starting at a random point in \mathbb{A}^N can arrive at a local optimum in fewer than a mean of $e - 1 - \varepsilon$ steps.

- (i) The limit as $N \rightarrow \infty$ of the proportion of points in \mathbb{A}^N that are local optima, or $\lim_{N \rightarrow \infty} p_0(N)$ equals zero.
- (ii) The fitnesses of all points in \mathbb{A}^N are identically distributed.
- (iii) The correlation coefficient for the fitnesses of two randomly chosen adjacent points is nonnegative.

The Nk -landscapes satisfy the hypotheses of the conjecture, as for these landscapes, (i)–(iii) hold. Therefore it seems likely that the $e - 1$ bound applies to these landscapes.

Note that it has been assumed throughout that the space being considered is \mathbb{A}^N with one-mutation adjacency. It is conceivable that lower bounds on walk lengths would be still smaller than $(A - 1)/A$ if an adjacency relation permitting a greater number of local optima to exist in the sequence space were used.

Acknowledgements

The author is supported by a Burroughs Wellcome Fund Career Award in the Biomedical Sciences. Peter Calabrese and Vincent Plagnol provided helpful comments about Proposition 1, and Lilach Hadany, Allen Orr, and two anonymous reviewers made suggestions that have led to the improvement of the manuscript.

References

- Kauffman, S., Levin, S., 1987. Towards a general theory of adaptive walks on rugged landscapes. *J. Theor. Biol.* 128, 11–45.
- Kauffman, S.A., Weinberger, E.D., 1989. The NK model of rugged fitness landscapes and its application to maturation of the immune response. *J. Theor. Biol.* 141, 211–245.
- Orr, H.A., 2003. A minimum on the mean number of steps taken in adaptive walks. *J. Theor. Biol.* 220, 241–247.
- Orr, H.A., 2005. The genetic theory of adaptation: a brief history. *Nature Rev. Genet.* 6, 119–127.
- Palmer, R., 1991. Optimization on rugged landscapes. In: Perelson, A., Kauffman, S. (Eds.), *Molecular Evolution on Rugged Landscapes*. Addison-Wesley, Redwood City, CA, pp. 3–23.
- Wright, S., 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proc. 6th Int. Cong. Genet.* 1, 356–366.