

# A Mathematical Introduction to Fitness Landscapes

Lenwood S. Heath

A fitness landscape is a conceptual framework within which to imagine the evolution of a species as driven by mutations. One first considers  $n$  alleles,  $n \geq 1$ , where each allele represents a locus in the genome of the species that varies. It is customary to abstract to binary alleles, where each allele has values from  $\Sigma = \{0, 1\}$  and we can represent any allele combination for the species as a binary string  $g \in \Sigma^n$ . A *genotype* of the species is simply an element  $g \in \Sigma^n$ . The *Hamming distance*  $h(g_1, g_2)$  between two genotypes  $g_1$  and  $g_2$  is the number of alleles in which they differ. The *landscape* of interest is simply the set  $\Sigma^n$  of all genotypes. A *fitness function*  $f$  assigns a number to each genotype, that is,  $f : \Sigma^n \rightarrow \mathbb{R}$ . The *fitness*  $f(g)$  of  $g$  is intended to give some indication of the reproductive success potential of the genotype, where a higher fitness value indicates higher success potential. The suggestion is that a mutation that results in a genotype of higher fitness than that of the original genotype is likely to be accepted and to survive in the population of the species.

A typical assumption within the conceptual framework is that only one mutation occurs in a genotype at a time. This makes possible a nice graph model of a fitness landscape. For  $n \geq 1$ , the *binary hypercube*  $\mathbb{B}^n = (V, E)$  is the undirected graph with vertex set  $V = \Sigma^n$  and edge set  $E = \{(g_1, g_2) \in V \mid h(g_1, g_2) = 1\}$ . For the purposes of studying evolution, paths in  $\mathbb{B}^n$  that proceed from vertices of lower fitness to vertices of higher fitness are preferred, and vertices that are local maxima with respect to fitness are thought of as peaks in the fitness landscape from which no further evolution is possible. Sometimes in the literature, paths that follow the fitness function only up are called *accessible paths*. For two genotypes  $g_1$  and  $g_2$ , there may be no accessible paths from  $g_1$  to  $g_2$ , exactly one accessible path from  $g_1$  to  $g_2$ , or many accessible paths from  $g_1$  to  $g_2$ . In the literature, much effort has been put into studying such paths.

A final component of the conceptual framework is a number of proposed models for the fitness function that might characterize evolution in a particular circumstance. The simplest model is the *House of Cards (HoC) model* that assigns fitness to genotypes independently from some probability distribution. To assign more realistic characteristics to the fitness function, *epistasis* is introduced, which is some nontrivial interaction among one or more alleles with respect to fitness. For example, if some allele goes from 0 to 1, it may make a different allele more favorable to go from 1 to 0. The *NK model* directly addresses epistasis. In this model,  $L = n$  is the number of alleles, and  $K$ , with  $0 \leq K \leq L - 1$  is a parameter setting the maximum number of alleles that can interact with any particular allele. Hence, setting  $K = 1$  allows only pairwise interactions. Within a set of  $K + 1$  interacting alleles, fitness is assigned at random, so  $K = L - 1$  corresponds to the HoC model. An *additive model* corresponds to  $K = 0$ , that is, no epistasis. A final model of interest is the *Random Mount Fuji (RMF) model* that combines an HoC model with an additive model to control a parameter called *ruggedness*, which is somewhat associated with the number of peaks and valleys in the fitness landscape. With the addition of fitness models, the conceptual framework of fitness landscapes provides a convenient means to discuss evolution within a species.