Chapter 22
Fitness Landscapes

Colin R. Reeves

22.1 Historical Introduction

One of the most commonly used metaphors to describe the process of heuristic methods such as local search in solving a combinatorial optimization problem is that of a *fitness landscape*. However, describing exactly what we mean by such a term is not as easy as might be assumed. Indeed, many cases of its usage in both the biological and optimization literature reveal a rather serious lack of understanding.

The landscape metaphor appears most commonly in work related to evolutionary algorithms, where it is customary to trace the usage of the term back to a paper by the population geneticist Sewall Wright (1932), although Haldane had already introduced a similar notion (Haldane 1931). The metaphor has become pervasive, being cited in many biological texts that discuss evolution.

Wright's original idea of a fitness landscape was somewhat ambiguous. It appears that what he initially had in mind concerned within-species variation where the *axes* of a search space represented unspecified gene combinations, but Dobzhansky's subsequent enthusiastic use of the metaphor (Dobzhansky 1951) seems to have established the consensus view of the axes of the search landscape as the frequency of a particular allele of a particular gene in a particular population. This can be seen in many textbooks on the subject of evolution, such as Ridley (1993). In the hands of Simpson, who seems to have thought primarily in terms of phenotypic characters (Simpson 1953), the story became even more highly developed, although even more divorced from empirical reality. Despite Wright's later attempts to clarify the situation (Wright 1967, 1988), the ambiguity remains. There is thus an interesting paradox in evolutionary biology: according to Futuyma (1998, p. 403),

[The] adaptive landscape is probably the most common metaphor in evolutionary genetic[s]

C.R. Reeves (✉)
Department of Mathematics and Control Engineering, Coventry University, Coventry, UK
e-mail: c.reeves@coventry.ac.uk

yet nobody seems sure what exactly is the reality to which the metaphor is supposed to relate! However, it remains extremely popular: Dawkins' recent book (Dawkins 1996), for example, makes considerable use of the notion, as its title *Climbing Mount Improbable* suggests.

Although we may have a vague idea of what the search space is, it is rather harder to define any axes for such a search space, as we have seen. Moreover, fitness in evolutionary biology is also a rather slippery concept. It is discussed as if there is some objective a priori measure, yet as usually defined, *fitness* concerns an organism's reproductive success, which can only be measured a posteriori.[1] Add to this the confusion over what the search space axes represent, and it becomes almost impossible to relate them to some quantifiable measure of fitness. It is thus generally dealt with by prestidigitation, and so, for all its popularity, the idea of a fitness landscape in biology is a mirage, displaying what is to a mathematician a distressing lack of rigor. Happily, some biologists agree, as in the cogent arguments against the hand-waving approach in Eldredge and Cracraft (1980). More recently, Skipper (2004) has argued that it still has a certain value in illustrating behavioral aspects of evolutionary theories, but Kaplan (2008) has suggested that it should be abandoned completely.

A more serious approach was foreshadowed by Eigen (1983) and Eigen et al. (1989). In his work on viruses, he introduced the concept of a quasi-species—a group of similar sequences. Each sequence S_k is a string of symbols drawn from some alphabet, the natural one to consider for viruses being the RNA bases adenine, cytosine, guanine and uracil—$\{A, C, G, U\}$. Differences in members of the quasi-species correspond to point mutations—replacement of one symbol by another one.

This interpretation falls somewhat short of the grand ideas in the popular biology textbooks, but it does make a formal mathematical development of the concept of a fitness landscape much more feasible, and following pioneering work by Weinberger (1990) in particular, a fairly complete formal statement of landscape theory was proposed by Stadler (1995). His paper is a sustained development and exploration of landscape concepts, requiring fairly sophisticated mathematics. Recent work has developed and explicated these ideas further (Reidys and Stadler 2002), but some quite extensive mathematical knowledge is still needed in order to appreciate it fully. In the expectation that the mathematical background of the readers of this volume will be somewhat variable, this tutorial will try to survey some of the themes most relevant to combinatorial optimization, without using advanced mathematical ideas. Some basic ideas of set theory, matrix algebra and functional analysis will be required, but the more complex ideas found in Reidys and Stadler (2002) will not be covered. Illustrative numerical examples will also be used at key points in an attempt to aid understanding.

---

[1] The *Oxford Dictionary of Biology* defines fitness as "The condition of an organism that is well adapted to its environment, as measured by its ability to reproduce itself."

22.2 Combinatorial Optimization

We can define combinatorial optimization problems as follows: we have a discrete search space $\mathcal{X}$, and a function

$$f : \mathcal{X} \mapsto \mathbb{R}.$$

The general problem is to find

$$x^* = \arg\max_{x \in \mathcal{X}} f.$$

where x is a vector of *decision variables* and f is the *objective function*. (Of course, minimization can also be the aim, but the modifications are always obvious). In the field of evolutionary algorithms, the function f is often called the *fitness*; hence the associated landscape is a *fitness landscape*. The vector $\mathbf{x}^*$ is a global optimum: that vector which is the *fittest* of all. (In some problems, there may be several global optima—different vectors of equal fitness.)

With the idea of a fitness landscape comes the idea that there are also many local optima or false peaks, in which a search algorithm may become trapped without finding the global optimum. In continuous optimization, notions of continuity and concepts associated with the differential calculus enable us to characterize quite precisely what we mean by a landscape, and to define the idea of an optimum. It is also convenient that our own experience of hill climbing in a three-dimensional world gives us analogies to ridges, valleys, basins, watersheds, etc., which help us to build an intuitive picture of what is needed for a successful search, even though the search spaces that are of interest often have dimensions many orders of magnitude higher than three.

However, in the continuous case, the landscape is determined only by the fitness function, and the ingenuity needed to find a global optimum consists in trying to match a technique to this single landscape. There is a major difference when we come to discrete optimization, which leads to mistaken *explanations* of the efficacy of this or that local search metaheuristic: sometimes it is asserted, for example, that a given technique (e.g. using a Gray code for 0–1 vectors) allows search in *more remote* parts of the landscape, when it would be more accurate to say that the landscape has been *remodeled* by changing distances. Indeed, we really should not use the term *landscape* unless we have first defined the topological relationships of the points in the search space $\mathcal{X}$. Unlike the continuous case, we have some freedom to specify these relationships, and in fact, that is precisely what we do when we decide to use a particular technique.

22.2.1 An Example

In practice, one of the most commonly used search methods for a combinatorial optimization problem is *neighborhood search*. This idea is at the root of modern

metaheuristics such as simulated annealing and tabu search—as well as being much more involved in the methodology of genetic algorithms than is sometimes realized.

A *neighborhood structure* is generated by using an operator that transforms a given vector x into a new vector x'. For example, if the solution is represented by a binary vector (as is often the case for genetic algorithms, for instance), a simple neighborhood might consist of all vectors obtainable by *flipping* one of the bits. The *bit flip* neighbors of (00000), for example, would be

$$\{(10000),(01000),(00100),(00010),(00001)\}.$$

Consider the problem of maximizing a simple function

$$f(z) = z^3 - 60z^2 + 900z + 100$$

where the solution z is required to be an integer in the range $[0, 31]$. Regarding z for a moment as a continuous variable, we have a smooth unimodal function with a single maximum at $z = 10$—as is easily found by calculus. Since this solution is already an integer, this is undoubtedly the most efficient way of solving the problem.

However, suppose we chose instead to represent z by a binary vector x of length 5. By decoding this binary vector as an integer it is possible to evaluate f, and we could then use neighborhood search, for example, to search over the binary hypercube for the global optimum using some form of hill-climbing strategy.

This discrete optimization problem turns out to have four optima (three of them local) when the bit flip operator is used. If a *steepest ascent* strategy is used (i.e. the *best* neighbor of a given vector is identified before a move is made) the local optima are as shown in Table 22.1. Also shown are the *basins of attraction*—the set of initial points from which the search leads to a specified optimum. For example, if we start the search at any of the points in the first column, and follow a strict best-improvement strategy, the search will finish up at the global optimum. However, if a *first improvement* strategy is used (where the first change that leads uphill is accepted without ascertaining if a still better one exists), the basins of attraction are rather different, as shown in Table 22.2.

In fact, there are even more complications: in Table 22.2, the order of searching the components of the vector has been termed *forward* (left-to-right). If the search is made in the reverse direction (right-to-left) the basins of attraction are different, as shown in Table 22.3.

Thus, by using flipping with this binary representation, we have created local optima that did not exist in the integer version of the problem. Further, although the optima are still the same, the chances of reaching a *particular* optimum can be seriously affected by a change in hill-climbing strategy.

Moreover, the bit flip operator is not the only mechanism for generating neighbors. An alternative neighborhood could be defined as follows: for $k = 1, \ldots, 5$, flip bits $\{k, \ldots, 5\}$. Thus, the neighbors of (00000), for example, would now be

$$\{(11111),(01111),(00111),(00011),(00001)\}.$$

Table 22.1 Local optima and basins of attraction for steepest ascent with the bit flip operator in the case of a simple cubic function. The bracketed figures are the fitnesses of each local optimum

Local optimum	0 1 0 1 0	0 1 1 0 0	0 0 1 1 1	1 0 0 0 0
	(4100)	(3988)	(3803)	(3236)
Basin	0 0 0 0 0	0 0 1 0 0	0 0 1 1 0	1 0 0 0 0
	0 0 0 0 1	0 1 1 0 0	0 0 1 1 1	1 0 0 0 1
	0 0 0 1 0	1 1 1 0 0	1 0 1 1 0	1 0 0 1 0
	0 0 0 1 1		1 0 1 1 1	1 0 0 1 1
	0 0 1 0 1			1 0 1 0 0
	0 1 0 0 0			
	0 1 0 0 1			
	0 1 0 1 0			
	0 1 0 1 1			
	0 1 1 0 1			
	0 1 1 1 0			
	0 1 1 1 1			
	1 0 1 0 1			
	1 1 0 0 0			
	1 1 0 0 1			
	1 1 0 1 0			
	1 1 0 1 1			
	1 1 1 0 1			
	1 1 1 1 0			
	1 1 1 1 1			

We shall call this the CX operator, and it creates a very different landscape. In fact, there is now only a single global optimum (01010); *every* vector is in its basin of attraction. This illustrates the point that it is not merely the choice of a binary representation that generates the landscape—the search operator needs to be specified as well.

Incidentally, there are two interesting facts about the CX operator. Firstly, it is closely related to the one-point crossover operator frequently used in genetic algorithms. (CX is actually shorthand for *complementary crossover*.) Secondly, if the 32 vectors in the search space are re-coded using a *Gray* code, it is easy to show that the bit-flip neighbors of a point in Gray-coded space are identical to those in the original binary-coded space under CX. This is an example of an *isomorphism* of landscapes.[2]

---

[2] An isomorphism in mathematics refers to mappings between mathematical objects that preserve structure. It comes from the Greek *iso* (equal) and *morphe* (shape). For example, two graphs are isomorphic if there is a one-to-one mapping σ between their respective sets of vertices such that for every edge (x, y) of one graph, $(\sigma(x), \sigma(y))$ is an edge of the other.

Table 22.2 Local optima and basins of attraction for first improvement (forward search) using the bit flip operator

Local optimum	0 1 0 1 0	0 1 1 0 0	0 0 1 1 1	1 0 0 0 0
	(4100)	(3988)	(3803)	(3236)
Basin	0 0 1 0 1	0 0 1 0 0	0 0 1 1 1	0 0 0 0 0
	0 0 1 1 0	0 1 0 0 0	0 1 1 1 1	0 0 0 0 1
	0 1 0 0 1	0 1 1 0 0	1 0 1 1 1	0 0 0 1 0
	0 1 0 1 0	1 0 1 0 0	1 1 1 1 1	0 0 0 1 1
	0 1 0 1 1	1 1 0 0 0		1 0 0 0 0
	0 1 1 0 1	1 1 1 0 0		1 0 0 0 1
	0 1 1 1 0			1 0 0 1 0
	1 0 1 0 1			1 0 0 1 1
	1 0 1 1 0			
	1 1 0 0 1			
	1 1 0 1 0			
	1 1 0 1 1			
	1 1 1 0 1			
	1 1 1 1 0			

22.3 Mathematical Characterization

Now that some of the typical features of a landscape have been illustrated, we can provide a mathematical characterization. We define a landscape $\mathcal{L}$ for the function f as a triple $\mathcal{L} = (\mathcal{X}, f, d)$ where d denotes a distance measure $d : \mathcal{X} \times \mathcal{X} \rightarrow I\!\!R^+ \cup \{\infty\}$ for which it is required that, $\forall s, t, u \in \mathcal{X}$,

$$d(s,t) \geq 0; \ \ d(s,t) = 0 \Leftrightarrow s = t; \ \ d(s,u) \leq d(s,t) + d(t,u).$$

Note that we do not need to specify the representation explicitly, since this is assumed to be implied in the description of $\mathcal{X}$. We have also decided, for the sake of simplicity, to ignore questions of search strategy and other matters in the definition of a landscape, unlike the more comprehensive definition of Jones (1995), for example.

This definition says nothing about how the distance measure arises. In fact, for many cases a *canonical* distance measure can be defined. Often, this is symmetric, i.e. $d(s,t) = d(t,s) \ \forall s, t \in \mathcal{X}$, so that d also defines a *metric* on $\mathcal{X}$. This is clearly a nice property, although it is not essential.

22.3.1 Neighborhood Structure

The distance measure is typically related to the neighborhood structure. Every solution $x \in \mathcal{X}$ has an associated set of *neighbors*, $N_\omega(x) \subset \mathcal{X}$, called the neighborhood of x. This neighborhood is generated by applying an operator ω to a vector s in order

Table 22.3 Local optima and basins of attraction for first improvement (reverse search) using the bit flip operator

Local optimum	0 1 0 1 0	0 1 1 0 0	0 0 1 1 1	1 0 0 0 0
	(4100)	(3988)	(3803)	(3236)
Basin	0 1 0 0 0	0 1 1 0 0	0 0 0 0 0	1 0 0 0 0
	0 1 0 0 1	0 1 1 0 1	0 0 0 0 1	1 0 0 0 1
	0 1 0 1 0	0 1 1 1 0	0 0 0 1 0	1 0 0 1 0
	0 1 0 1 1	0 1 1 1 1	0 0 0 1 1	1 0 0 1 1
			0 0 1 0 0	1 0 1 0 0
			0 0 1 0 1	1 0 1 0 1
			0 0 1 1 0	1 0 1 1 0
			0 0 1 1 1	1 0 1 1 1
				1 1 0 0 0
				1 1 0 0 1
				1 1 0 1 0
				1 1 0 1 1
				1 1 1 0 0
				1 1 1 0 1
				1 1 1 1 0
				1 1 1 1 1

to transform it into a vector t. What we may call a canonical distance measure d_ω is that induced by ω whereby

$$t \in N_\omega(s) \Leftrightarrow d_\omega(s,t) = 1.$$

The distance between non-neighbors is defined as the length of the shortest path between them (if one exists). The operator ω generally takes a parameter, which means that it is technically a one-to-many function, able to generate many neighbors from one initial vector. The size of the neighborhood will be denoted by n.

For example, if X is the binary hypercube $\{0,1\}^\ell$, the bit flip operator can be defined as

$$\phi(i) : \{0,1\}^\ell \times \mathbb{Z} \to \{0,1\}^\ell \qquad \begin{cases} z'_i = 1 - z_i \\ z'_k = z_k & \text{if } i \neq k, \end{cases}$$

where z is a binary vector of length ℓ, and i is the parameter specifying the bit to be flipped. It is clear that the distance metric induced by ϕ is the well-known Hamming distance

$$d_H(x,y) = \sum_{i=1}^{\ell} [x_i \neq y_i],$$

where the square brackets $[expr]$ denote an *indicator* function, which takes the value 1 if the logical expression *expr* is true and 0 otherwise. Thus we could describe this landscape as a Hamming landscape (with reference to its distance measure), or as the bit-flip landscape (with reference to the operator). Similarly, we can define the CX operator as

$$\gamma(i) : \{0,1\}^\ell \times \mathbb{Z} \to \{0,1\}^\ell \qquad \begin{cases} z'_k = 1 - z_k \text{ for } k \geq i \\ z'_k = z_k \qquad \text{otherwise.} \end{cases}$$

The distance measure induced here is clearly more complicated than the Hamming landscape, and cannot be described by a simple function. In both of these cases the size of the neighborhood is $n = \ell$.

As an example of an asymmetric distance measure, consider the case where X is Π_m, the space of permutations of length m, which is often relevant for scheduling problems. A familiar neighborhood is defined by the *forward shift* operator, taking two parameters in this case:

$$\mathcal{FSH}(i,j) : \Pi_m \times \mathbb{Z} \times \mathbb{Z} \to \Pi_m \qquad \begin{cases} \pi'_{k-1} = \pi_k \text{ for } j < k \leq i \\ \pi'_i = \pi_j \\ \pi'_k = \pi_k \quad \text{otherwise.} \end{cases}$$

The neighbors of (1234), for example, would be

$$\{(2134), (2314), (2341), (1324), (1342), (1243)\}.$$

Note that the size of this neighborhood is $n = \binom{m}{2}$. It is easily seen that (1234), however, is not a neighbor of $(2314), (2341)$ or (1342), so $\mathcal{FSH}$ is not symmetric. Other neighborhood operators (for example, *exchange*, where two items in the sequence are swapped) induce different distance measures, so there may be advantages in choosing operator-independent distance measures (Reeves 1999) for practical comparisons.

Distance measures may become even more complicated: for instance, in the problem of biological sequence comparison (RNA, DNA and protein sequences, see Waterman 1995), it is common to compare sequences in terms of the minimal number of genetic operations necessary to convert one string into another (the "string edit" distance). Thus, even finding the distance measure is effectively an optimization problem.

22.3.2 Local Optima

We can now give a formal statement of a fundamental property of fitness landscapes: for a landscape $\mathcal{L} = (X, f, d)$, a vector $x^o \in X$ is *locally optimal* if

$$f(x^o) > f(t) \quad \forall\, t \in N(x^o).$$

We shall denote the set of such optima as X^o, and the set of *global optima* (recall that we allow the possibility of more than one) as X^* where the vector $x^* \in X^o$ is a global optimum if

$$f(x^*) \geq f(x^o) \quad \forall\, x^o \in X^o.$$

Landscapes that have only one local (and thus also global) optimum are commonly called *unimodal*, while landscapes with more than one local optimum are said to be *multimodal*.

The number of local optima in a landscape clearly has some bearing on the difficulty of finding the global optimum. In our previous example, it is clearly more difficult to find the global optimum using bit-flipping than if we used CX. However, it is not the only indicator: in our example the steepest-ascent strategy increased the chance of finding the global optimum, since there were more initial solutions that led to the global optimum than under first-improvement.

22.3.3 Basins of Attraction

We can also now define more precisely the idea of a *basin of attraction*. Neighborhood search can be interpreted as a function

$$\mu : X \mapsto X^o,$$

where if x is the initial point, $\mu(x)$ is the optimum that it reaches. With this in mind, we can define the basin of attraction of x^o as the set

$$B(x^o) = \{x : \mu(x) = x^o\}.$$

The problem is that $B(x^o)$ is not independent of the search strategy, as the example of Sect. 22.2.1 demonstrated. In fact, it is only well defined for the case of steepest ascent. For other search strategies, such as first improvement, the order of searching may be highly influential. Our example showed that the basin of attraction of the global optimum was much larger for steepest ascent than for the other strategies, but it is possible to find examples where the opposite is the case.

22.3.4 Plateaux

The definition of a local optimum above ignored the possibility that in many problems neighboring points may have the *same* fitness value. We could ignore this by assuming that fitnesses can be *jittered* by adding a small random perturbation, but in some cases the issue causes considerable difficulty to a local search, as there may be extensive *plateaux*. A plateau $\mathcal{P}$ is a set of points $\{x_i\}$ such that $f(x_i) = c$, a constant, and if $s,t \in \mathcal{P}$, there is a path

$$s = x_1, x_2, \ldots, x_k = t \quad \text{such that } x_{i+1} \in N(x_i) \text{ and } x_i \in \mathcal{P} \quad \forall i.$$

In fact, this situation can be investigated further: in Frank et al. (1997), a whole taxonomy of plateaux is discussed, but in this brief chapter the subject will not

be explored further. Suffice it to say that understanding plateaux and methods to circumvent them is a very important part of finding efficient techniques for problems such as MAXSAT.

22.3.5 Graph Representation

Neighborhood structures are clearly just another way of defining a graph Γ, which can be described by its $(n \times n)$ *adjacency matrix A*. The elements of A are given by $a_{ij} = 1$ if the indices i and j represent neighboring vectors, and $a_{ij} = 0$ otherwise. For example, the graph induced by the bit flip ϕ on binary vectors of length 3 has adjacency matrix

$$
A_\phi = \begin{bmatrix}
0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 & 1 & 0
\end{bmatrix}
$$

where the vectors are indexed from 0 to 7 in the usual binary-coded integer order (i.e. $(000), (001)$, etc.). By way of contrast, the adjacency matrix for the CX operator is

$$
A_\gamma = \begin{bmatrix}
0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\
0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\
1 & 0 & 0 & 0 & 1 & 0 & 1 & 0
\end{bmatrix} .
$$

It is simply demonstrated that permuting the rows and columns so that they are in the order $0, 1, 3, 2, 7, 6, 4, 5$ reproduces the adjacency matrix A_ϕ—another way of demonstrating the isomorphism mentioned earlier. In other words,

$$
P^{-1} A_\phi P = A_\gamma
$$

where P is the associated permutation matrix of the binary-to-Gray transformation. It is also clear that the eigenvalues and eigenvectors are the same (up to a permutation).

As a final example, we may consider the adjacency matrix for $\mathcal{FSH}$ in the space Π_3:

$$A_{\mathcal{FSH}} = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix},$$

where the permutations have been indexed in lexicographic order $(123), \dots, (321)$. The lack of symmetry in the distance measure is of course reflected in an asymmetric matrix.

22.3.6 Laplacian Matrix

The *graph Laplacian* Δ is defined as

$$\Delta = A - D,$$

where D is a diagonal matrix such that d_{ii} is the degree of vertex i. Usually, these matrices are vertex-regular and $d_{ii} = n \; \forall i$, so that

$$\Delta = A - nI.$$

This notion recalls that of a Laplacian operator in the continuous domain; the effect of this matrix, applied at the point s to the fitness function f is

$$\Delta f(s) = \sum_{t \in N(s)} (f(t) - f(s)),$$

so it functions rather like a finite-differencing operator in the continuous domain. In particular, $\Delta f(s)/n$ is the average difference in fitness between the vector s and its neighbors. Grover (1992) observed that the landscapes of several combinatorial optimization problems satisfy a difference equation of the form

$$\Delta f + \frac{C}{n} f = 0,$$

where C is a problem-specific constant. This is essentially a discrete analog of the *wave equation* familiar in physics. From this it can be deduced that *all* local optima are better (i.e. larger, in the case of maximization) than the mean $\bar{f}$ over all points on the landscape. Furthermore, it can also be shown that under mild conditions on the nature of the fitness function, the time taken by neighborhood search to find a local maximum is $O(n \log_2[f_{max}/\bar{f}])$ where f_{max} is the fitness of a global maximum. (Similar results hold, mutatis mutandis, for minimization problems.)

22.3.7 Graph Eigensystem

In the usual way, we can define eigenvalues and eigenvectors of the matrices associated with the graph Γ. The set of eigenvalues is called the *spectrum* of the graph. For an $n \times n$ matrix A the spectrum is

$$\left(\lambda_0 \ \lambda_1 \ \ldots \ \lambda_{n-1} \right),$$

where λ_i is the ith eigenvalue, ranked in (weakly) descending order. Similarly, the spectrum of the Laplacian is

$$\left(\mu_0 \ \mu_1 \ \ldots \ \mu_{n-1} \right),$$

where, again, μ_i is the ith eigenvalue, ranked this time in (weakly) ascending order. For a regular connected graph it can be shown that

$$\mu_i = n - \lambda_i \ \forall i.$$

Further, from the corresponding eigenvectors the *eigenfunctions* $\{\varphi_i\}$ can be formed, and f can be expanded as

$$f(s) = \sum_i a_i \varphi_i(s).$$

This is sometimes called a Fourier expansion (Stadler and Wagner 1998).

Unfortunately, the size of these graphs rapidly becomes very large. However, graphs can often be partitioned in a way that makes it possible to reduce the scale of the problem. This enables the formation of a *collapsed matrix* $\tilde{A}$ whose eigenvalues are the *distinct* eigenvalues of A, with multiplicities given by the cardinalities of the partitions. (Relevant mathematical details may be found in the books by Biggs (1993) and Godsil (1993).) If the diameter of such a graph is δ, the number of distinct eigenvalues is only $\delta + 1$, so a considerable reduction in size is possible—at least in principle.

Similarly, the Fourier expansion can be partitioned into a sum

$$f(s) = \sum_p \beta_p \tilde{\varphi}_p(s)$$

over the distinct eigenvalues of Δ. The corresponding values

$$|\beta_p|^2 = \sum |a_k|^2$$

(where the sum is over the coefficients that correspond to the pth distinct eigenvalue) form the *amplitude spectrum*, which expresses the relative importance of different components of the landscape.

22.3.8 Elementary Landscapes

Ideally, such mathematical characterizations could be used to aid our understanding of the important features of a landscape, and so help us to exploit them in designing search strategies. One step in this direction starts from the connection between Grover's observation and Stadler's rather more obscure results (Stadler 1995). Stadler designated landscapes where f is itself an eigenfunction of the "wave equation" as *elementary* landscapes.

In Whitley et al. (2008), it is shown that from a practical point of view, an elementary landscape arises when the objective function is *partially decomposable*. This occurs when the function f is formed as a sum of subfunctions that each involve only a subset of the relevant components of the objective function. Often, such a property is already used in implementing local search: evaluating the cost of a move implied by applying the operator ω does not necessarily entail evaluating f itself. (The TSP is the most obvious case in point: a *2-opt* move disconnects the tour, but most of the inter-city distances are unchanged and do not need to be re-calculated.) In such cases, Whitley and his colleagues show, useful information can be gained from partial evaluation of neighborhoods

Another interesting implication of this analysis is for landscapes where f is *not* partially decomposable. Problems such as flowshop and jobshop scheduling, for example, may have fundamentally different performance characteristics from those problems whose neighborhoods can induce elementary landscapes.

However, apart from the implications of Grover's equation, it is possible to carry out further analytical studies only for small graphs or graphs with a special structure, as illustrated for example, in Stadler (1995). For the important case of the Hamming landscape of a binary search space, analytical results for the graph spectrum show that the eigenvectors are thinly disguised versions of the familiar Walsh functions.[3] For the case of recombinative operators the problem is considerably more complicated, and necessitates the use of *P-structures* (Stadler and Wagner 1998). The latter are essentially generalizations of graphs in which the mapping is from pairs of *parents* (x, y) to the set of possible strings that can be generated by their recombination. However, it can be shown that for some *recombination landscapes* (such as that arising from the use of uniform crossover) the eigenvectors are once more the Walsh functions. Whether this is also true in the case of one- or two-point crossover, for example, is not known, but Stadler and Wagner conjecture that it is. In view of the close relationship between the bit-flip and CX landscapes as demonstrated above, it would not be surprising if this is a general phenomenon. However, to obtain these results, some assumptions have to be made—such as a uniform distribution of parents—that are unlikely to be true in a specific finite realization of a genetic search.

---

[3] For readers who are unfamiliar with Walsh functions, they are digital analogs of trigonometrical functions, forming an orthonormal set of rectangular waveforms. An introduction to their uses in the analysis of optimization methods can be found in Reeves and Rowe (2002).

In the case of the bit-flip landscape, the distinct eigenvalues correspond to sets of Walsh coefficients of different orders, and the amplitude spectrum is exactly the set of components of the *epistasis variance* associated with other attempts to measure problem difficulty (see Reeves and Rowe 2002 for a review). For the cubic function of Sect. 22.2.1, the components of variance for the different orders of Walsh coefficients can be shown to be $(0.387, 0.512, 0.101, 0, 0)$ respectively; i.e. 61.3 % of the variation in the landscape is due to interactions of order two and three. This is consistent with the relatively poor performance of the bit-flip hill climber.

Of course, the eigenvalues and eigenvectors are exactly the same (up to a permutation) for the CX landscape of this function, and the set of values for the Walsh coefficients in the Fourier decomposition is also the same. However, the effect of the permutation inherent in the mapping from the bit flip landscape to the CX landscape is to re-label some of the vertices of the graph, and hence some of the Walsh coefficients. Thus some coefficients that previously referred to linear effects now refer to interactions, and vice versa. Taking the cubic function as example again, the components of variance or amplitude spectrum become $(0.771, 0.174, 0.044, 0.011, 0.000)$. We see that the linear effects now predominate (77.1 %), and this is consistent with the fact that the hill climber in the CX landscape always finds the optimum of this function.

Table 22.4 Illustration of the different groupings of the Walsh coefficients associated with the bit flip, CX and recombination landscapes

Index	Binary coding	Bit flip	CX	Recom	Index	Binary coding	Bit flip	CX	Recom
0	0000	0	0	0	8	1000	1	2	1
1	0001	1	1	1	9	1001	2	3	4
2	0010	1	2	1	10	1010	2	4	3
3	0011	2	1	2	11	1011	3	3	4
4	0100	1	2	1	12	1100	2	2	2
5	0101	2	3	3	13	1101	3	3	4
6	0110	2	2	2	14	1110	3	2	3
7	0111	3	1	3	15	1111	4	1	4

22.3.9 Recombination Landscapes

If we look at the recombination landscapes derived from the P-structures of Stadler and Wagner (1998), we find that once again the Walsh coefficients are obtained, but labeled in yet another way. The coefficients in the bit flip and CX landscapes are grouped according to the number of 1s in their binary- and Gray-coded index representations respectively. However, in a recombination landscape—such as that generated by 1-point crossover—it is the *separation* between the outermost 1-bits that defines the groupings. Table 22.4 shows the groupings for a 4-bit problem.

Several things can be seen from this table: firstly, the linear Walsh coefficients (and hence the linear component of epistasis variance) are the same in both the bit flip and the recombination landscapes. Secondly (as already explained), the coefficients in the CX landscape are simply a re-labeling of those in the bit flip landscape. Thirdly, the coefficients in the recombination landscape do not form a natural grouping in terms of interactions, and consequently the different components of variance for the recombination landscape do not have a simple interpretation as due to interactions of a particular order.

22.3.10 Summary

This section has set out some of the basic mathematics necessary for the analysis of landscapes. As has probably become obvious, the details can require an extensive mathematical knowledge. Furthermore, the full analysis of a particular landscape (i.e. its eigensystem) may need the gathering of a large amount of empirical information, perhaps equivalent to a complete knowledge of the fitness function at all points of the search space! Landscape analysis in such cases can be no more than an a posteriori justification (or lack of it!) for the choice of a particular neighborhood. Further discussion on some of these points may be found in Reeves and Rowe (2002).

While it is undeniably useful that we can construct mathematical techniques to help us neatly summarize certain facts about a landscape, we must also recognize that there are other features—possibly very important ones—that are not captured by these methods. In the simple example of the cubic function we have seen that the search strategy adopted can make a big difference to the likelihood of a hill climber finding the global optimum.

Mathematical analysis holds out some tantalizing prospects of future progress, but for the moment we turn to a consideration of the results of experimental work on landscapes.

22.4 Statistical Measures

If mathematical analysis of a landscape is a difficult task, then it is natural to ask if there is something we can learn about the nature of a landscape, simply from the process of searching it. Several ideas have been suggested.

22.4.1 Autocorrelation

One of the earliest attempts to obtain some statistical measure of a landscape was by Weinberger, who showed that certain quantities obtained in the course of a random walk can be useful indicators (Weinberger 1990). If the fitness of the point visited at time t is denoted by f_t, we can estimate the *autocorrelation function* (usually abbreviated to *acf*) of the landscape during a random walk of length T as

$$r_j = \frac{\sum\limits_{t=1}^{T-j} (f_t - \bar{f})(f_{t+j} - \bar{f})}{\sum\limits_{t=1}^{T} (f_t - \bar{f})^2}.$$

Here $\bar{f}$ is of course the mean fitness of the T points visited, and j is known as the *lag*. The concept of autocorrelation is of course an important one in time series analysis, but its interpretation in the context of landscapes is interesting.

For *smooth* landscapes, and at small lags (i.e. for points that are close together), the *acf* is likely to be close to 1 since neighbors are likely to have similar fitness values. However, as the lag increases the correlations will diminish. "Rugged" landscapes are informally those where close points can nevertheless have completely unrelated fitnesses, and so the *acf* will be close to zero at all lags. Landscapes for which the *acf* has significant negative values are conceptually possible, but they would have to be rather odd.

A related quantity is the *correlation length* of the landscape, usually denoted by τ. Classical time series analysis (Box and Jenkins 1970) can be used to show that the standard error of the estimate r_j is approximately $1/\sqrt{T}$, so that there is only approximately 5 % probability that $|r_j|$ could exceed $2/\sqrt{T}$ by chance. Values of r_j less than this value can be assumed to be zero. The correlation length τ is then the last j for which r_j is non-zero:

$$\tau = j : |r_{j+1}| < 2/\sqrt{T} \wedge \{|r_k| > 2/\sqrt{T} \quad \forall\, k \le j\}.$$

The *acf* and the correlation length are useful indicative measures of the ruggedness of a landscape, but they are rather crude statistics, and it is difficult to attach a great deal of meaning to their values for particular instances.

22.4.2 Number of Optima

Although it is not the full story, the number of local optima is widely acknowledged as a very important factor in how easy or difficult it is to find a global optimum of a landscape, and is clearly much more directly relevant for a particular instance than the correlation measures. Recently, some attempts have been made (Reeves 2001; Eremeev and Reeves 2002, 2003; Reeves and Eremeev 2004) to obtain direct estimates of the number of optima using statistical principles.

It is assumed that a heuristic search method can be restarted many times using different initial solutions. Given the landscape framework we have discussed above, by randomly generating initial solutions, we will sample many basins of attraction. Of course, this will be evident by the number of *different* final solutions $\{x^o\}$ that are found. Suppose this number is k, and the number of restarts is $r(\geq k)$. Various statistical models may be used in order to estimate the number of optima ν.

22.4.2.1 Waiting-Time Model

We can ask for the distribution of the waiting-time to find all optima. If r exceeds k substantially, this fact can be used to estimate the probability that all optima have been found. This would also imply, a fortiori, that the global optimum had been found, and thus provides us with an objective confidence level concerning the quality of the best solution obtained.

22.4.2.2 Counting Model

In the event—unfortunately, a common one—that k is not much smaller than r, it is unlikely that we have seen many of the optima. However, a counting model can be used to estimate the value of ν, in a similar way to those used by ecologists to estimate the size of an unknown animal population. This can be quite illuminating in showing the differences between landscapes generated by different neighborhood operators.

22.4.2.3 Non-parametric Estimates

Fairly restrictive assumptions are needed in order to obtain tractable statistical models of landscapes. Where these estimates can be checked against actuality (by enumerating all points on a landscape), it appears that the effect of these assumptions is to produce negatively biased estimates—i.e. the estimate of ν is consistently smaller than the true value. Removing the assumptions by creating more general models would probably be impossible, so some *non-parametric* approaches have been explored, and found to provide useful estimates of ν, although the problem of negative bias remains, albeit at a lower level. All these models are summarized in Reeves and Eremeev (2004).

22.5 Empirical Studies

Besides explicit statistical models of landscape features, several empirical studies have been aimed at providing some idea of the "big picture". Although multi-dimensional fitness landscapes have few similarities with "real" 3D landscapes,

certain empirical findings can be interpreted sensibly in terms of characteristics of real landscapes, which provides us with some insights into ways we can approach hard optimization problems.

One of the most interesting observed properties of fitness landscapes has been seen in many different studies: it is a feature of Kauffman's well-known *NK* landscapes (Kauffman 1993),[4] and it also appears in many examples of combinatorial optimization problems, such as the traveling salesman problem (Boese et al. 1994; Lin 1965), graph partitioning (Merz and Freisleben 1998), and flowshop scheduling (Reeves 1999).

In the first place, such studies have repeatedly found that, on average, local optima are very much closer to the global optimum than are randomly chosen points, and closer to each other than random points would be. That is, the distribution of local optima is not *isotropic*; rather, they tend to be clustered in a *central massif* (or—if we are minimizing—a *big valley*). This can be demonstrated graphically by
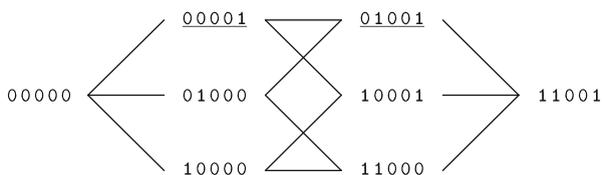

Fig. 22.1 The diagram shows the set of paths that could be traced between the parents 00000 and 11001. Only those intermediate vectors indicated by *underlines* can be generated by one-point crossover, but all can be generated by uniform crossover

plotting a scatter graph of fitness against distance to the global optimum. Secondly, if the basins of attraction of each local optimum are explored, size is quite highly correlated with quality: the better the local optimum, the larger is its basin of attraction. (If true, this also impinges on the estimation problem we discussed in the previous section: although there is a negative bias in the estimate of ν, the *big valley* phenomenon implies that it is only the small basins and low-quality optima that we are missing.)

Of course, there is no guarantee that this property holds in any particular case, but it provides an explanation for the success of *perturbation* methods (Johnson 1990; Martin et al. 1992; Zweig 1995) which currently appear to be the best available for the traveling salesman problem. It is also tacitly assumed by such methods as simulated annealing and tabu search, which would lose a great deal of their potency if local optima were isotropically distributed.

---

[4] In Kauffman's notation, N is the length of a binary string, and K is the maximum number of genes that are allowed to interact with any other; e.g. if $K = 1$, each gene can interact with just one other. There are several different ways in which the sets of interacting genes can be chosen, but essentially they turn out to make little difference.

22.5.1 Practical Applications

These studies have suggested a starting point for the development of new heuristic search algorithms, such as the *adaptive multi-start* algorithm of Boese et al. (1994). As a more recent example, we shall consider the *path tracing* algorithms introduced in Reeves and Yamada (1998) and Yamada and Reeves (1998), which can be motivated either as a use of the idea of a landscape, or in terms of extending the boundaries of evolutionary algorithms.

If we consider the case of crossover of vectors in $\{0, 1\}^{\ell}$, it is easily seen that any *child* produced from two *parents* will lie on a path that leads from one parent to another. Figure 22.1 demonstrates this fact.

In an earlier paper (Reeves 1994), such points were described as intermediate vectors. In other search spaces, the distance measure may be more complicated, but the principle is still relevant. Crossover is re-interpreted as finding a point lying *between* two parents in some landscape in which we hope the big valley conjecture is true. This *path-tracing crossover* was implemented for both the makespan and the flowsum versions of the flowshop sequencing problem; Fig. 22.2 shows in a two-dimensional diagram the idea behind it, while full details can be found in Reeves and Yamada (1998, 1999).

In this way, the concept of recombination can be fully integrated with traditional neighborhood search methods, and the results obtained for flowshop instances (see Reeves and Yamada 1998 and Yamada and Reeves 1998 for details) have been gratifyingly good. For the makespan problem, embedded path tracing helped the GA to achieve results of outstandingly high quality: several new best solutions were discovered for well known benchmarks. For the flowsum version, optimal solutions are not known, but the path-tracing GA consistently produced better solutions than other proposed techniques.

This idea has also recently been applied to multi-constrained knapsack problems (Levenhagen et al. 2001), where the need was confirmed for a *big valley* structure in order to benefit from this approach.

Very recently, a detailed examination of TSP instances (Hains et al. 2011) has shown that the big valley structure breaks down around high-quality local optima,
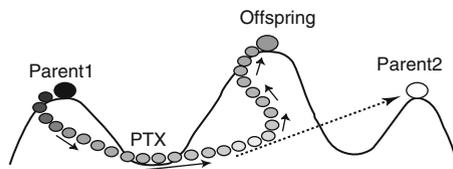

Fig. 22.2 Path tracing crossover combined with local search: a path is traced from one parent in the direction of the other. In the *middle* of the path, solutions may be found that are not in the basins of attractions of the parents. A local search can then exploit this new starting point by climbing to the *top* of a hill (or the *bottom* of a valley, if it is a minimization problem)—a new local optimum. The acronym PTX signifies *path-tracing crossover*

with multiple *funnel* structures appearing. Further investigation leads to new ideas about search operators for such instances, although whether this extends to other COPs is not yet known.

22.6 Promising Areas for Future Application

Finally, we should remark that several interesting future research questions are suggested. On the theoretical side, a deeper knowledge of the connections between algebra and graph theory may provide further useful analytical results. For example, it would be useful to have analytical results for all the common operators in permutation spaces analogous to those derived for the simpler case of binary strings. As mentioned in Sect. 22.3.4, real fitness landscapes for a number of combinatorial optimization problems have to cope with extensive plateaus. Measuring the extent and effects of such formations also needs further study, as does the characterization of basins of attraction. Some promising ideas based on the notion of a *barrier tree* have been put forward by Stadler and colleagues in Flamm et al. (2002), which are developed further in Hallam and Prugel-Bennet (2005).

Building on such notions, it would be helpful if we could provide a formal definition of what it means for a *big valley* structure to exist, and how it could be related to mathematical constructs associated with neighborhood structures. Does the big valley exist almost everywhere? If not, can we define classes of problems and neighborhood structures for which it does not occur? Further empirical analysis, such as that described by Levenhagen et al. (2001), Watson et al. (2002), and Hains et al. (2011), should be of considerable assistance in suggesting fruitful avenues to explore.

More generally, it is clear that crude correlation measures can only be a general guide to the nature of a landscape instance, and we need to find better ways of characterizing landscapes from empirical measurements. Some suggestions have been made in Reeves (2004) for further work in this direction.

In the area of implementation, it is important to see if we can further refine the path tracing methodology and its integration into heuristic search methods such as evolutionary algorithms. Also, the methodological developments pioneered in Reeves (2001), Eremeev and Reeves (2002, 2003), and Reeves and Eremeev (2004) for deducing properties of an instance of a landscape from the results of heuristic search offer the possibility of making principled probability statements about the quality of solutions obtained.

22.7 Tricks of the Trade

Mathematical analysis of landscapes is generally possible only for small problems, and then can only really be useful as an a posteriori validation (or questioning) of the decisions already made. However, empirical analysis is relatively easy and may provide some useful insights.

Correlation analysis can be a helpful indicator of the type of landscape with which we are dealing. Typically this proceeds by making a random walk on the landscape for several thousand steps and collecting data on fitness. The resultant *time series* can be analyzed with standard statistical tools. The drawback of this approach is that even when it is complete, knowing how smooth or rugged the landscape is from the perspective of a random walk does not help very much in deciding which heuristic search method to adopt. Further, much computation has been carried out yet the search for an optimum has not even started!

For those wishing to make use of empirical landscape analysis as part of a general research program, it should be realized that much of the necessary information is inherently generated in the course of applying a heuristic search method to a combinatorial optimization problem. Of course, if a single run is all that is used, nothing much can be gleaned, but if independent restarts or a Metropolis-type search are used, it becomes possible to collect statistics and make use of them.

The existence of a *big valley* is usually an encouraging feature, and requires little checking. Assuming the global optimum is unknown it will not be possible to do a complete analysis, but useful information can be gained by computing the distance of each local optimum from the best local optimum, and plotting this against their corresponding differences in fitness. A strong correlation is indicative of a *big valley*, and motivates the application of metaheuristics that perform intensive searches in the region of *good* local optima.

If every local optimum ever found is distinct, not much more can be done, but if it is noticed that specific local optima are being detected multiple times, it becomes possible to provide indications of solution quality, using statistical estimation tools based on the waiting-time or counting models mentioned above. For low values of the ratio k/r (see above), it may even be possible to provide a (probabilistic) guarantee that the global optimum has indeed been found.

22.8 Conclusion

This chapter has reviewed and discussed in some detail the basic mathematical theory and methods associated with the concept of a fitness landscape. While these methods can be very useful in enhancing our understanding of evolutionary algorithms, it has been emphasized that they cannot provide a complete explanation for the performance of a specific algorithm on their own—even in the case of very simple functions. Secondly, and more briefly, some empirically determined properties of many search landscapes have been described, and one approach whereby such properties can be exploited has been outlined.

As our understanding of the nature of fitness landscapes and how to exploit them develops, this promises to become an important area of research into the theory and application of heuristic search.

Sources of Additional Information

- For technical and theoretical analysis, there are many papers associated with Peter Stadler and his co-workers. The paper of Reidys and Stadler (2002) is perhaps the most readily accessible and recent treatment of theoretical properties of landscapes, although the seminal work is still Stadler (1995). Many of these papers can be found on the University of Leipzig website[5] and also at the Santa Fe Institute.[6]
- Several papers give a general low-tech introduction to landscapes—for example Reeves (1999, 2000)—as does the chapter in Reeves and Rowe (2002). For ideas relating to elementary landscapes, see Barnes et al. (2003), Dimova et al. (2005), and Whitley et al. (2008).
- For correlation analysis, Weinberger (1990) is still a major source of information, supplemented by more recent work in papers by Stadler and co-workers (see the Vienna website); another useful reference is Hordijk (1996).
- For work relating to the *big valley* and its exploitation, there are several important papers: Boese et al. (1994), Reeves and Yamada (1998), Merz and Freisleben (1998), and Reeves (1999); a chapter by Reeves and Yamada in Corne et al. (1999) is also an accessible introduction.
- The statistical approach to estimation of landscape properties is described in a series of papers (Eremeev and Reeves 2002, 2003; Reeves and Eremeev 2004), and its extension to the use of the Metropolis algorithm is considered in Reeves and Aupetit-Bélaidouni (2004).

References

Barnes JW, Dimova B, Dokov SP, Solomon A (2003) The theory of elementary landscapes. Appl Math Lett 16:37–343

Biggs NL (1993) Algebraic graph theory. Cambridge University Press, Cambridge

Boese KD, Kahng AB, Muddu S (1994) A new adaptive multi-start technique for combinatorial global optimizations. Oper Res Lett 16:101–113

Box GEP, Jenkins GM (1970) Time series analysis, forecasting and control. Holden Day, San Francisco

Corne DA, Dorigo M, Glover F (eds) (1999) New methods in optimization. McGraw-Hill, London

Dawkins R (1996) Climbing mount improbable. Viking, London

Dimova B, Barnes JW, Popova E (2005) Arbitrary elementary landscapes and AR(1) processes. Appl Math Lett 18:287–292

Dobzhansky T (1951) Genetics and the origin of species. Columbia University Press, New York

---

[5] www.bioinf.uni-leipzig.de/~studla/Publications/index.html

[6] www.santafe.edu/research/publications/

Eigen M (1993) Viral quasispecies. Sci Am 269:32–39

Eigen M, McCaskill J, Schuster P (1989) The molecular quasi-species. Adv Chem Phys 75:149–263

Eldredge N, Cracraft J (1980) Phylogenetic patterns and the evolutionary process. Columbia University Press, New York

Eremeev AV, Reeves CR (2002) Non-parametric estimation of properties of combinatorial landscapes. In: Gottlieb J, Raidl G (eds) Applications of evolutionary computing. LNCS 2279. Springer, Berlin, pp 31–40

Eremeev AV, Reeves CR (2003) On confidence intervals for the number of local optima. In: Raidl G et al (eds) Applications of evolutionary computing. LNCS 2611. Springer, Berlin, pp 224–235

Flamm C, Hofacker IL, Stadler PF, Wolfinger MT (2002) Barrier trees of degenerate landscapes. Z Phys Chem 216:155–173

Frank J, Cheeseman P, Stutz J (1997) When gravity fails: local search topology. J Artif Intell Res 7:249–281

Futuyma DJ (1998) Evolutionary biology. Sinauer Associates, Sunderland

Godsil CD (1993) Algebraic combinatorics. Chapman and Hall, London

Grover LK (1992) Local search and the local structure of *NP*-complete problems. Oper Res Lett 12:235–243

Hains DR, Whitley LD, Howe A (2011) Revisiting the big valley search space structure in the TSP. J Oper Res Soc 62:305–312

Haldane JBS (1931) A mathematical theory of natural selection, Part VI: metastable populations. Proc Camb Phil Soc 27:137–142

Hallam J, Prugel-Bennett A (2005) Large barrier trees for studying search. IEEE Trans Evol Comput 9:385–397

Hordijk W (1996) A measure of landscapes. Evol Comput 4:335–360

Johnson DS (1990) Local optimization and the traveling salesman problem. In: Goos G, Hartmanis J (eds) Automata, languages and programming. LNCS 443. Springer, Berlin, pp 446–461

Jones TC (1995) Evolutionary algorithms, fitness landscapes and search. Doctoral dissertation, University of New Mexico, Albuquerque

Kaplan J (2008) The end of the adaptive landscape metaphor? Biol Phil 23:625–638

Kauffman S (1993) The origins of order: self-organization and selection in evolution. Oxford University Press, New York

Levenhagen J, Bortfeldt A, Gehring H (2001) Path tracing in genetic algorithms applied to the multiconstrained knapsack problem. In: Boers EJW et al (eds) Applications of evolutionary computing. Springer, Berlin, pp 40–49

Lin S (1965) Computer solutions of the traveling salesman problem. Bell Syst Tech J 44:2245–2269

Martin O, Otto SW, Felten EW (1992) Large step Markov chains for the TSP incorporating local search heuristics. Oper Res Lett 11:219–224

Merz P, Freisleben B (1998) Memetic algorithms and the fitness landscape of the graph bi-partitioning problem. In: Eiben AE et al (eds) Proceedings of the PPSN V, Amsterdam. Springer, Berlin, pp 765–774

Reeves CR (1994) Genetic algorithms and neighbourhood search. In: Fogarty TC (ed) Evolutionary computing: AISB workshop selected papers. Springer, Berlin, pp 115–130

Reeves CR (1999) Landscapes, operators and heuristic search. Ann Oper Res 86:473–490

Reeves CR (2000) Fitness landscapes and evolutionary algorithms. In: Fonlupt C et al (eds) 4th European conference on artificial evolution. LNCS 1829. Springer, Berlin, pp 3–20

Reeves CR (2001) Direct statistical estimation of GA landscape features. In: Martin WN, Spears WM (eds) Foundations of genetic algorithms 6. Morgan Kaufmann, San Francisco, pp 91–107

Reeves CR (2004) Partitioning landscapes. Available online at http://www.dagstuhl. de/04081/Talks/

Reeves CR, Aupetit-Bélaidouni MM (2004) Estimating the number of solutions for SAT problems. In: Yao X et al (eds) Proceedings of the PPSN VIII, Birmingham. LNCS 3242. Springer, Berlin, pp 101–110

Reeves CR, Eremeev AV (2004) Statistical analysis of local search landscapes. J Oper Res Soc 55:687–693

Reeves CR, Rowe JE (2002) Genetic algorithms—principles and perspectives. Kluwer, Norwell

Reeves CR, Yamada T (1998) Genetic algorithms, path relinking and the flowshop sequencing problem. Evol Comput 6:45–60

Reeves CR, Yamada T (1999) Goal-oriented path tracing methods. In: Corne DA, Dorigo M, Glover F (eds) New methods in optimization. McGraw-Hill, London

Reidys CM, Stadler PF (2002) Combinatorial landscapes. SIAM Rev 44:3–54

Ridley M (1993) Evolution. Blackwell, Oxford

Simpson GG (1953) The major features of evolution. Columbia University Press, New York

Skipper R (2004) The heuristic role of Sewell Wright's 1932 adaptive landscape diagram. Philos Sci 71:1176–1188

Stadler PF (1995) Towards a theory of landscapes. In: Lopéz-Peña R et al (eds) Complex systems and binary networks. Springer, Berlin, pp 77–163

Stadler PF, Wagner GP (1998) Algebraic theory of recombination spaces. Evol Comput 5:241–275

Waterman MS (1995) Introduction to computational biology. Chapman and Hall, London

Watson J-P, Barbalescu L, Whitley LD, Howe AE (2002) Contrasting structured and random permutation flow-shop scheduling problems: search-space topology and algorithm performance. INFORMS J Comput 14:98–123

Weinberger ED (1990) Correlated and uncorrelated landscapes and how to tell the difference. Biol Cybern 63:325–336

Whitley LD, Sutton AM, Howe AE (2008) Understanding elementary landscapes. In: Proceedings of the GECCO 2008, Atlanta. ACM, New York, pp 585–592

Wright S (1932) The roles of mutation, inbreeding, crossbreeding and selection in evolution. In: Jones D (ed) Proceedings of the 6th international congress on genetics, Ithaca, vol 1, pp 356–366

Wright S (1967) Surfaces of selective value. Proc Nat Acad Sci 102:81–84

Wright S (1988) Surfaces of selective value revisited. Am Nat 131:115–123

Yamada T, Reeves CR (1998) Solving the C_{sum} permutation flowshop scheduling problem by genetic local search. In: Proceedings of the CEC 1998. IEEE, Piscataway, pp 230–234

Zweig G (1995) An effective tour construction and improvement proedure for the traveling salesman problem. Oper Res 43:1049–1057