The notion of an *adaptive landscape* has proved to be a valuable concept in theoretical investigations of evolutionary change, combinatorial optimization, and the physics of disordered systems. Landscape theory has emerged as an attempt to devise suitable mathematical structures for describing the “static” properties of landscapes as well as their influence on the dynamics of adaptation. Here we focus on the connections of landscape theory and algebraic combinatorics that form the basis of spectral approach to understanding landscape structure.

1 INTRODUCTION

Evolutionary change is caused by the spontaneously generated genetic variation and its subsequent fixation by drift and/or selection. Consequently, the main focus of evolutionary theory has been to understand the genetic structure and dynamics of populations, see e.g. [101]. In recent years, however, alternative approaches have gained increasing prominence in evolutionary theory. This development has been stimulated to some extent by the application of evolutionary models to designing *evolutionary algorithms* such as Genetic Al-
algorithms, Evolution Strategies, and Genetic Programming, as well as by the

The generic structure of an evolutionary model is
\[ x' = S(x, w) \circ T(x, t), \]
where \( x \) is e.g. the vector of haplotype frequencies, and \( S(x, w) \) is a term
describing the selection forces acting on \( x \). The parameters \( w \) form the so-
called fitness function, since they can be regarded as a mapping from the
set of types into the real numbers. The second term, \( T(x, t) \), describes the
transmission processes by determining the probability of transforming one
type into another one by mutation and/or recombination [3]. Hence, evolution
models can be seen as dynamical systems of genotype frequencies which live
on an algebraic structure [96] that is determined by genetic processes such as
mutation and recombination.

Metaphorically, the dynamics of evolutionary adaptation therefore can be
seen as a walk on a landscape, where uphill moves are preferred. The realiza-
tion that the topological features of fitness landscapes crucially influence
the time-course of natural and simulated evolution led to what is now called
landscape theory. It has several roots: In evolutionary theory it can be traced
back to Wright’s ideas about adaptive landscapes, see [114, pp. 304-317], and
became important in theories of molecular evolution and the origin of life
[31, 33, 42, 48, 75, 80, 110, 126, 127] and in evolutionary computer science
[76, 77]. Similar developments exist in physics [44], where free energy land-
scape of disordered systems such as spin glasses are considered [98], and in
search theory [109]. The main challenge to landscape theory is to determine
which features of the fitness landscape determine the evolvability of the sys-
tems on the landscape.

From the mathematical point of view, a landscape consists of three ingredi-
ents: (i) a set \( V \) of “configurations” which we shall assume to be finite
but very large, (ii) a cost or fitness function \( f : V \rightarrow \mathbb{R} \) that evaluates
the configurations, and (iii) some sort of additional geometrical, topological, or
algebraic structure \( \mathcal{X} \) on \( V \) that allows us to define notions of closeness, simi-
larity, or dissimilarity among the configurations. The structure \( \mathcal{X} \), which turns
the set \( V \) into the configuration space \( (V, \mathcal{X}) \), is determined by the particular
application, e.g. a heuristic search procedure for a combinatorial optimization
problem, or by the mechanisms of mutation and recombination in biological
Evolution.

A very promising approach in landscape theory is the decomposition of
the fitness function \( f : V \rightarrow \mathbb{R} \) in terms of a basis (of the vector space \( \mathbb{R}^V \))
that is induced in some natural way by \( \mathcal{X} \). In other words, we search for
a suitable spectral theory of the combinatorial space \( (V, \mathcal{X}) \), which we then
use to “Fourier transform” \( f \) with respect to a suitable set eigenfunctions
of \( (V, \mathcal{X}) \). The resulting “Fourier coefficients”, so one hopes, will reveal the
important features of the landscape much more readily than \( f \) itself.
2 CONFIGURATION SPACES

2.1 GRAPHS, HYPERGRAPHS, P-STRUCTURES, AND FINITE TOPOLOGIES

We have argued in the introduction that the “additional structure $\mathcal{X}$ makes the boring “bag of numbers” $f : V \to \mathbb{R}$ a landscape - and an interesting mathematical object. The structure of the set $V$ is oftentimes related to, or derived from, the internal structure of the objects $x \in V$. In this section we shall explore a few possibilities of imposing structure onto the set $V$ of configurations.

2.1.1 Graphs. The simplest case is based on the notion of a move set. For each $x \in V$ we define as set $N(x)$ of “neighbors” of $x$. The elements of $N(x)$ are those configurations that can be reached in a single step starting form $x$. It will be convenient to assume $x \notin N(x)$ for all $x$ and to define $\overline{N}(x) = N(x) \cup \{x\}$. This definition allows us to regard the set

$$\mathcal{E} = \{ (x, y) \mid x \in V, y \in N(x) \}$$

(2)

as the edge set of a directed graph with vertex set $V$. Equivalently, $\mathcal{E}$ is a neighborhood relation on $V$, that is, a relation satisfying $(x, x) \notin \mathcal{E}$ for all $x \in V$.

In many cases one is interested in symmetric neighborhood relations, i.e., in move sets in which each step is “reversible” and $(x, y) \in \mathcal{E}$ implies $(y, x) \in \mathcal{E}$. We may then regard $V$ as an undirected graph with edges $\{x, y\} \in E$ iff $(x, y) \in \mathcal{E}$. The undirected graph case is by far the best studied one.

The tours of a traveling salesman problem, for example, can be encoded as the list of cities in the order in which they are visited. In other words, a particular tour is a permutation $\pi$ of the cities $\{1, \ldots, n\}$. It seems natural to make use of the fact that these permutations form the symmetric group $S_n$: choose a move set $\Omega \subset S_n$ and define that $y$ is a neighbor if $x$ if $y$ is obtained from $x$ by multiplication with an element of $t \in \Omega$, $y = xt$. Of course, we require that $\Omega$ does not contain the group identity $1$. Thus $(x, y) \in \mathcal{E}$ if and only if $x^{-1} y \in \Omega$. The resulting graph is a so-called Cayley digraph of the $S_n$. In most cases one assumes that $t \in \Omega$ implies $t^{-1} \in \Omega$, in which case the neighborhood relation is symmetric and the Cayley graph has undirected edges.

In molecular biology, we may for instance consider sequences as configurations and mutation as the move set. We have to distinguish two types of mutation: point mutations change a letter in a sequence without affecting its length. Insertions and deletions, on the other hand, change the length of a sequence. Editing operations such insertions and deletions are used for example in sequence alignment algorithms and can be generalized to trees [103, 146, 66]. Tree editing also provides a suitable analogue for mutation in Genetic Programming, see [106, 105]. Tree editing procedures are furthermore
used in phylogenetic reconstruction [35]. Editing operations related to sorting are used to deal with with genomic rearrangements [81].

2.1.2 Recombination, P-Structures, and Hypergraphs. Recombination, or crossover, is another way of imposing a sense of closeness on $V$. For strings, the meaning of cross-over is easily defined.

A **cross-over operator** is a map $\chi : V \times V \rightarrow V \times V$ with the following property: Suppose $\chi(y, z) = (u, v)$. Then for each $k$ either $y_k = u_k \wedge z_k = v_k$ or $z_k = u_k \wedge y_k = v_k$. By abuse of notation we write $x \in \chi(y, z)$ if $x = u$ or $x = v$, i.e., if $x$ is an offspring of $(y, z)$. As an immediate consequence we see that $\chi(x, x) = (x, x)$. We follow here the spirit of [77] when we regard a crossover operator as producing pairs of sequences rather than a single sequence from a pair of “ancestors”.

We write $\chi = \{k | x_k = u_k \wedge y_k = v_k \}$ and $\bar{x} = \{k | y_k = u_k \wedge x_k = v_k \}$ for the two subsets of loci (sequence positions) that are separated by the cross-over operator $\chi$. There is of course a one-to-one correspondence between a cross-over operator $\chi : V \times V \rightarrow V \times V$ and the associated set $\chi \subseteq \{1, \ldots, n\}$: the set lists exactly those loci that are inherited from the first parent $x$ by the first child $u$, see Figure 1. Analogously, $\bar{x}$ is the list of loci that the first child $u$ inherits from the second parent $y$. Of course, $\bar{x} = \{1, \ldots, n\} \setminus \chi$.

Note that ancestors and offsprings have the same length. More general types of recombination, often called **unequal crossover**, do not adhere to this restriction: here recombination events may occur between independent positions in the two sequences whence the chain lengths of the recombinants vary [130]. Crossover operators may also be defined for trees and permutations, with applications in genetic programming [88] and GAs for the traveling salesman problem [159], respectively.

A **recombination operator** (in the sense of much of the GA literature) is a family $\mathcal{F}$ of cross-over operators that act on $V \times V$ with probabilities $\pi(\chi)$. In the following we restrict ourselves to equal-length crossover on strings. The
two most important recombination operators are uniform recombination \([\infty]\), consisting of all \(2^n\) possible crossover operators, and 1-point recombination \([1]\), which contains all cross-over operators \(\chi\) for which the characteristic set is of the form \(\chi = \{1, \ldots, k\}\).

Let \(V\) be a finite set with power set \(\mathcal{P}(V)\). A \(P\)-structure \([143]\) is a pair \((V, R)\) where \(R : V \times V \to \mathcal{P}(V)\). We say that the \(P\)-structure is symmetric if \(R(x, y) = R(y, x)\) for all \(x, y \in V\). In a weighted \(P\)-structure we attach a positive weight \(H_{x, [y, z]}^\chi\) to each triple \((x, y, z)\) for which \(x \in R(y, z)\) and we set \(H_{x, [y, z]}^\chi = 0\) if \(x \notin R(y, z)\). We call \(H\) the incidence matrix of the \(P\)-structure.

In particular, there is a weighted \(P\)-structure associated in a natural way with each cross-over operator \(\chi\):

\[
R^\chi(y, z) = \{x \in V | x \in \chi(y, z)\}
\]

\[
H_{x, [y, z]}^\chi = \begin{cases} 
2 & \text{if } x = y = z \\
1 & \text{if } x \in \chi(y, z) \text{ and } y \neq z \\
0 & \text{otherwise}.
\end{cases}
\]

(3)

We observe that \(H_{x, [y, z]}^\chi > 0\) if and only if \(x\) is an offspring of \((y, z)\). The doubled weight in the “diagonal”, \(H_{x, [x, x]}^\chi = 2\), is mostly a technical convenience: It implies immediately \(\sum_x H_{x, [y, z]}^\chi = 2\), since any crossover operator produces exactly two offsprings from a pair of parents. If \(y = z\), we simply count the offspring \(y = z\) twice. The weighted \(P\)-structure associated with a recombination operator \(\mathcal{F}\) is then

\[
H = \sum_{\chi \in \mathcal{F}} \pi(\chi) H^\chi
\]

\[
R(y, z) = \bigcup_{\chi \in \mathcal{F}} R^\chi(y, z) = \{x \in V | \exists \chi \in \mathcal{F} : x \in \chi(y, z)\}
\]

(4)

The interpretation of this definition is straightforward: \(H_{x, [y, z]}^\chi\) is the chance that \(x\) is an offspring of the parents \(y\) and \(z\) under \(\mathcal{F}\)-recombination \([153]\).

The recombination hypergraph \(\text{imag} R\) has vertex set \(V\) and hyper-edges \(R(y, z), y, z \in V\). A spectral theory of hypergraphs is described in \([123]\). Gitschhof and Wagner \([49]\) introduced a set axioms to describe the action of recombination in terms of what we call here \(P\)-structures. In \([143, \text{Lemma C2}]\) we showed that any recombination operator forms a recombination structure if and only if the identity map on \(V \times V\) is a member of the family \(\mathcal{F}\) of crossover operators.

2.1.3 Finite Topological Spaces. The shape space of RNA secondary structure has been treated as a finite metric space, with a distance measure that is based on “structure editing” \([39, 74]\). It has become apparent, however, that distance measures of this type are not useful for explaining the features of evolutionary trajectories \([40, 41]\). In these contributions, a notion of “continuity”
is introduced and the evolutionary transitions are classified as continuous or discontinuous based on how easily one shape can be accessed from a previous one. Continuity is a topological property. Taking this idea serious one may regard shape space as (finite) topological space assuming that the “natural” topology is obtained by declaring the sets $\mathcal{N}(x)$ of structures that are accessible from $x$ as open sets. This approach will be pursued elsewhere [20]. We just mention here that finite topological spaces have a unique non-redundant basis consisting of the sets

$$B(x) = \bigcap_{y \in \mathcal{N}(x)} \overline{\mathcal{N}(y)}$$

(5)

which may be translated into the directed graph $\Gamma$ with vertex set $V$ and edges $B(x) \setminus \{x\}, x \in V$. Topological properties such as separation properties can be then expressed as graph-theoretical properties of $\Gamma$, and we are back to the graph case.

### 2.2 MATRIX REPRESENTATIONS

#### 2.2.1 Markov Chains.

Not surprisingly we shall encounter a close relationship between spectral graph theory [12, 22, 21, 23] and landscapes on graphs in the course of this survey. A graph is faithfully represented by its adjacency matrix $A$ which has the entries

$$A_{xy} = \begin{cases} 1 & \text{if } (x, y) \in \mathcal{E} \\ 0 & \text{if } (x, y) \notin \mathcal{E} \end{cases}$$

(6)

Of course, $A$ is symmetric if and only if the graph is undirected.

The most straight forward way of search on a possibly weighted (di)graph is a random walk, that is a Markov process with state space $V$. The most natural transition matrix is

$$\text{Prob}(y \to x) = S_{xy} = A_{xy}/\sum_{z \in V} A_{xz}.$$  

(7)

Such a random walk is usually called *simple* since each edge leaving $y$ is chosen with the same probability. The denominator in equ.(7) is the out-degree of vertex $x$. The matrix $S$ is the transition matrix of the random walk. Note that this is the transpose of the convention in most of the literature on Markov chains, see e.g., [11, 95]. The most important feature of random walks is the existence of a stationary distribution $\varphi$ such that $\varphi = S\varphi$ to which all initial distributions converge under fairly general conditions.

A Markov process is called reversible if its stationary distribution $\varphi$ satisfies the *balance equation* $S_{xy}\varphi(y) = \varphi(x)S_{yx}$. In particular, a simple random walk on an undirected graph is reversible. Let $P$ be the diagonal matrix with diagonal $\varphi$. If $S$ is the transition matrix of a reversible chain then
$T = P^{-1/2}SP^{1/2}$ is a bistochastic symmetric matrix. The “regularized” transition matrix $T$ still essentially describes the graph $\Gamma$ since $T_{xy} > 0$ if and only if $(x,y)$ is an edge of $\Gamma$. Since $T$ is a symmetric non-negative matrix it serves as the starting point for the spectral theory of Markov processes [11].

Let us briefly consider the case of Hamming graphs in its most general setting. The configuration space consists of “genomes” with $n$ loci (or positions) $k = 1, \ldots, n$. There are $\alpha_k$ alleles (or letters) at each position, which we denote by $x_k \in A_k = \{0,1,\ldots,\alpha_k - 1\}$. With $x_k \in A_k$ we associate the root of unity

$$\tilde{x}_k = \exp(2\pi i x_k/\alpha_k)$$

Furthermore, for $I \in \prod_k A_k$ set $\tilde{I} = \{k|I_k \neq 0\}$. For each “index” $I$ we define the generalized Walsh function

$$\varepsilon_I : \prod_k A_k \rightarrow \mathbb{C} : \varepsilon_I(x) = \prod_k x_k^{I_k} = \prod_{k \in I} \tilde{x}_k$$

We remark that $\{\varepsilon_I | I \in V\}$ is the standard Fourier basis of the Abelian group $\prod Z_{\alpha_k}$, see section 2.3.6 below. These functions are eigenvectors of the adjacency matrix of the Hamming graphs $\prod Q_{\alpha_k}$, i.e., the graphs obtained by considering point-mutations; see e.g. [135]. Note that the formal association of the index sets $I$ with the vertices in $V$ is a mere book-keeping device. If the number of alleles is the same for all loci, $\alpha_k = \alpha$, then the eigenvalue of $S$ associated with $\varepsilon_I$ is $\lambda_I = 1 - \alpha(1 - |I|)/n$.

Useful Markov processes on $V$ can be defined, however, without any reference to a graph structure. The string recombination structures introduced in section 2.1.2 may serve as an example: A cross-over walk [70, 71] on $V$ is the Markov process based on the following rule: The “father” $y$ is mated with a randomly chosen “mother” $z$. The offsprings are the “son” $x$ and the “daughter” $\tilde{x}$. The “son” $x$ becomes the “father” of the next mating. We regard the sequence of “fathers” as a random walk on $V$. It is straightforward to derive the transition matrix of this Markov process:

$$S_{xy} = \frac{1}{2} \sum_{z \in V} H_{x,y,z} \rho(z)$$

The factor 1/2 stems from the fact that the offspring $x$ is the “son” and not the “daughter” of the parents $y$ and $z$ with probability 1/2. By $\rho(z)$ we denote the probability that $z$ is the “mother” of the mating, i.e., $\rho(z)$ is the frequency of genotype $z$ in the population $\rho$ under random mating. The uniform population case $\rho(z) = 1/|V|$, which is discussed in [143, 153], is generalized to the Wright manifold $W = \{ \rho | \rho(z) = \prod_k p_k(z_k) \}$, where $p_k(a)$ denotes the frequency of allele $a$ at locus $k$, in [142]. It is not hard to verify that linkage equilibrium is maintained under recombination, i.e., that any $\rho \in W$ is a stationary distribution of $S$ as defined in equ. (10). This fact
was first proven for two alleles and arbitrary number of loci by Robbins [121]
and for multiple alleles by Bennett [9].

The “population-weighted” Walsh functions
\[
\psi_f^p(x) = \prod_{k \in I} \frac{1}{p_k(x_k)} \hat{p}_k
\]
are defined for all \( \varphi \in \mathcal{W} \). For a uniform population they coincide with the
generalized Walsh functions introduced in equ. (9). In [142] we show that the
population weighted generalized Walsh function \( \psi_f^p \) is a left eigenvector of
\( S^\chi \) with eigenvalue
\[
\lambda_f^\chi = \begin{cases} 1 & \text{if } \hat{I} = \emptyset \\
1/2 & \text{if } \emptyset \neq \hat{I} \subseteq \chi \text{ or } \emptyset \neq \check{I} \subseteq \check{\chi} \\
0 & \text{otherwise} \end{cases}
\]

This observation not only sets that stage for a spectral analysis of recombina-
tion landscapes, it also shows that recombination and mutation on strings
are compatible operations that can therefore be compared directly in a mean-
ingful way. The close relation between Hamming graphs and recombination
spaces was noted with different methods by various groups, e.g. [19, 49, 93],
see also appendix A.

2.2.2 Schrödinger Operators and Graph Laplacians. Let \( \Gamma \) be a simple graph
(without loops and multiple edges), and let \( a \) be a weight function on the edges
of \( \Gamma \), conveniently defined as \( a : V \times V \to \mathbb{R}_0^+ \) such that \( a(x, y) = a(y, x) > 0 \)
if \( \{x, y\} \in E \) and \( a(x, y) = 0 \) otherwise. We say that \( \Gamma \) is unweighted if
\( a(x, y) \in \{0, 1\} \), i.e., if \( a(x, y) = A_{xy} \). Furthermore, let \( v : V \to \mathbb{R} \) be an
arbitrary potential. The linear operator \( \mathbf{H} \) defined by the action
\[
\mathbf{H} f(x) = \sum_{y \sim x} a(x, y) [f(x) - f(y)] + v(x) f(x)
\]
is a discrete Schrödinger operator associated with \( \Gamma \) [24]. This definition
includes the transition matrices of random walks on graph discussed in the
previous section. The quantity
\[
\deg(x) = -\sum_{y : \{x, y\} \in E} \mathbf{H}_{xy}
\]
is the (generalized) degree of a vertex \( x \in V \). The degree matrix \( \mathbf{D} \) is the
diagonal matrix of the vertex degrees.

We call \( -\mathbf{\Delta} = \mathbf{D} - \mathbf{A} \) the Laplacian of the edge-weighted graph \( \Gamma \) [99,
100]. A slightly different definition is explored in [17]. Hence any Schrödinger
operator is of the form \( \mathbf{H} = -\mathbf{\Delta} + \text{diag}(v(x)) \). The Laplacian is therefore a
Schrödinger operator without potential.
The analogy between discrete and continuous Schrödinger operators is a close one because the discrete Laplacian $-\Delta$ resembles the Laplacian differential operator $\Delta$ in many ways. To see this, one introduces an arbitrary orientation on $\Gamma$ by choose one of the two vertices $u$ or $v$ of the edge $h = \{v, w\}$ as the “positive end” and the other one as the “negative end”. The matrix

$$\nabla^+_{xh} = \begin{cases} +\sqrt{a(x,y)} & \text{if } x \text{ is the positive end of } h = \{x, y\} \\ -\sqrt{a(x,y)} & \text{if } x \text{ is the negative end of } h = \{x, y\} \\ 0 & \text{otherwise} \end{cases}$$

is called the (weighted) incidence matrix of $\Gamma$. The choice of the symbol $\nabla$ is intentional. In fact, let $f : V \to \mathbb{R}$ be an arbitrary function. Then $(\nabla f)(h) = \sqrt{a(v,w)}[f(v) - f(w)]$ where $h$ is the edge $\{v, w\}$, and $v$ is the positive end of the edge $h$. This is as close to a first derivative as one can get on a graph.

Note that $1/\sqrt{a(v,w)}$ takes the role of the distance between the vertices $v$ and $w$.

The discrete Laplacian $-\Delta$ is symmetric, non-negative definite, and singular. The eigenvector $(1, \ldots, 1)$ belongs to the eigenvalue $\lambda_0 = 0$. $\lambda_0$ has multiplicity 1 if and only if $\Gamma$ is connected. A few simple computations verify that $\Delta = -\nabla^+ \nabla$ and hence corresponds to “second derivatives” on $\Gamma$. Let $\langle \cdot, \cdot \rangle$ denote the standard scalar product on $\mathbb{R}^{|V|}$, and let $f, g : V \to \mathbb{R}$ be arbitrary landscapes. Then Green’s formula holds in the following form:

$$\langle \nabla f, \nabla g \rangle = -\langle f, \Delta g \rangle = -\langle g, \Delta f \rangle$$

Graph Laplacians appear in very diverse fields of pure and applied mathematics. Their earliest use goes back to Kirchhoff’s theory of electrical networks [84], see, e.g., [12, Chap.5].

2.2.3 Courant’s Nodal Domain Theorem. A well-known feature of Schrödinger operators on Riemannian manifolds is that the nodal domains, that is, the connected components of $M \setminus \psi^{-1}(0)$, of their eigenfunctions are severely constrained. In order to formulate Courant’s theorem for graphs, we define for any function $f : V \to \mathbb{R}$ on $\Gamma$: $\text{supp}_+(f) = \{x \in V | f(x) > 0\}$, $\text{supp}_-(f) = \{x \in V | f(x) < 0\}$, $\text{zero}(f) = \{x \in V | f(x) = 0\}$, $\text{supp}_0^+(f) = \text{supp}_+(f) \cup \text{zero}(f)$, and $\text{supp}_0^-(f) = \text{supp}_-(f) \cup \text{zero}(f)$. A (strong) nodal domain of $f$ is a maximal connected component of either $\text{supp}_+(f)$ or $\text{supp}_-(f)$. A weak nodal domain is a maximal connected component of $\text{supp}_+(f) \cup \text{zero}(f)$ or $\text{supp}_-(f) \cup \text{zero}(f)$, respectively.

Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{|V|}$ be the eigenvalues of a Schrödinger operator on $\Gamma$ with corresponding eigenvectors $\varphi_i$. Define $M(i) = \max\{k | \lambda_k = \lambda_i\}$ and $m(i) = \min\{k | \lambda_k = \lambda_i\}$. Hence, $m(i) \leq i \leq M(i)$, $M(i) = m(i) + \text{mult}(\lambda_i) - 1$, and $M(i) = M(i) = i$ if and only if $\lambda_i$ is a simple eigenvalue of $\mathbf{H}$.

The main result on discrete Schrödinger operators is the following version of Courant’s Nodal Domain Theorem, which motivates why the eigenfunctions of a Laplacian form particularly interesting basis sets for our purposes:
Let $\psi_i$ be an eigenvector of $H$ with eigenvalue $\lambda_i$. Then:

(i) There are at most $M(i)$ (strong) nodal domains of $\psi_i$.
(ii) There are at most $m(i)$ weak nodal domains of $\psi_i$.
(iii) If $\psi_i$ has $m(i) + k$, $k > 0$, (strong) nodal domains, then no two of them meet at a non-vertex point of the geometric representation of the graph $\Gamma$, and every vertex meets at least $k + 1$ (strong) nodal domains.

The proofs of these results were obtained independently by different authors [45, 24, 149], beginning with Fiedler [36] who showed that the number of components of $\text{supp}_+(\psi_i)$ is at most $M(i)$. Some closely related results on the component structure of $\text{supp}_+(\psi_i) \cup \text{supp}_-(\psi_i)$ can be found in [112].

### 2.3 SYMMETRIES, PARTITIONS, AND MATRIX ALGEBRAS

In many cases of practical interest there is a substantial amount of symmetry in the ways by which the set of configurations is constructed. Below we shall briefly explore a few approaches which have been used to exploit these regularities in search of a workable spectral theory.

#### 2.3.1 Relations and Automorphisms

A relation $\mu$ on $V$ is simply a subset $\mu \subseteq V \times V$. The adjacency relation of a graph $\Gamma$ may serve as an example. An automorphism of $\mu$ is a permutation $g \in S_{|V|}$ such that $(x, y) \in \mu$ if and only if $(g(x), g(y)) \in \mu$. The automorphisms of $\mu$ for the (permutation) group $\text{Aut}[\mu]$, the automorphism group of $\mu$. The automorphism group of a set $\mathcal{R}$ of relations on $V$ is

$$\text{Aut}[\mathcal{R}] = \bigcap_{\mu \in \mathcal{R}} \text{Aut}[\mu]$$

(16)

To each relation $\mu$ on $V$ there is an associated characteristic $|V| \times |V|$ matrix $R^{(\mu)}$ with entries $R^{(\mu)}_{xy} = 1$ if $(x, y) \in \mu$ and $R^{(\mu)}_{xy} = 0$ if $(x, y) \notin \mu$.

Let $G$ be an arbitrary permutation group acting on $V$. By $2\text{orb}(G, V)$ we denote the set of orbits of $G$ acting on $V \times V$. Of course the $2\text{orb}(G, V)$ corresponds to a partition of $V \times V$, and each element of $2\text{orb}$ may be regarded as a relation on $V$. These relation encapsulate the information about the symmetries that are most relevant for us.

A matrix representation of a finite group $G$ is a map $\rho$ from $G$ into the group of $d \times d$ invertible matrices with complex coefficients such that $\rho(gh) = \rho(g)\rho(h)$ for all $g, h$ in $G$. The permutation representation $G$ of $(G, V)$ consists of the $|V| \times |V|$ permutation matrices $G(g)$ whose non-zero entries are $G_{xy}(g) = 1$ if and only if $x = g(y)$.

A permutation group $G$ on $V$ is intimately connected with its centralizer algebra

$$\mathfrak{U} = \mathfrak{U}_C(G, V) = \{ M \in C^{|V| \times |V|} \mid \forall g \in G : MG(g) = G(g)M \}$$

(17)

The set $\mathfrak{U}$ is closed w.r.t. addition and multiplication of matrices and w.r.t. multiplication with scalars from the underlying field $C$. Its dimension (as vec-
tor space) equals the rank of its permutation group, \(\dim(\mathfrak{W}) = \text{rank}(G, V)\).

The characteristic matrices \(R^{(\mu)}\) of the orbits \(\mu \in 2\text{orb}(G, V)\) form the standard basis of the vector space \(\mathfrak{W}\). From \(R^{(\mu)} \circ R^{(\nu)} = \delta_{\mu, \nu}R^{(\kappa)}\) we see that \(\mathfrak{W}\) is also closed under component-wise (Schur or Hadamard) multiplication. Finally, \(\mathfrak{W}\) is closed under transposition since the transpose \(\mu^\top = \{(x, y) | (y, x) \in \mu\}\) of an orbit is again an orbit.

2.3.2 Coherent Algebras. A set of complex matrices that is closed under (i) scalar multiplication with complex numbers, (ii) component-wise addition, (iii) ordinary matrix multiplication, (iv) component-wise multiplication, and (v) transposition is called a coherent algebra or cellular algebra. Equivalently, a matrix algebra \(\mathfrak{W} \subseteq \mathbb{C}^{|V| \times |V|}\) is coherent if and only if it satisfies the following axioms:

(i) As a linear space over \(\mathbb{C}\), \(\mathfrak{W}\) has a basis of \(\{R^{(1)}, \ldots, R^{(r)}\}\) of 0-1 matrices.

(ii) \(\sum_{\nu=1}^{r} R^{(\nu)} = J\), the all-1 matrix.

(iii) For every \(i \in \{1, \ldots, r\}\) there is an \(i'\) such that \(R^{(i)} \circ R^{(i')} = R^{(i')} \circ R^{(i)}\).

(iv) \(I \in \mathfrak{W}\).

Sometimes coherent algebras without unity are considered, i.e., axiom (iv) is disregarded. The centralizer algebras of permutation groups form the most prominent class of coherent algebras (with identity).

Axiom (ii) above implies that the relations associated with the basis matrices \(R^{(j)}\) form a partition of \(V \times V\). Such partitions are known as coherent configurations [63, 64, 65]. Table 1 gives an overview of various properties of partitions of \(V \times V\) that are of interest in the context of landscapes. For details see e.g. [135, 136].

For each collection \(\mathcal{M} = \{M_1, \ldots, M_k\}\) of \(|V| \times |V|\) matrices there is a smallest coherent algebra \(\langle \mathcal{M} \rangle\) which is defined as the intersection of all coherent algebras that contain \(\{M_1, \ldots, M_k\}\). Since the centralizer algebra is coherent we have

\[
\langle \mathcal{M} \rangle \subseteq \mathfrak{W}(\text{Aut}[M], V)
\]  

Equality holds if and only if there is a permutation group that has \(\langle \mathcal{M} \rangle\) as its centralizer algebra [87]. The coherent algebra \(\langle \mathcal{M} \rangle\) can therefore be regarded as a “combinatorial approximation” of the centralizer algebra [34, 85]. This is of particular importance in the graph case: given the adjacency matrix \(A\) of \(\Gamma\), there is polynomial time algorithm that determines the coherent algebra \(\mathfrak{W}(\Gamma) = \langle A \rangle\), see [157, 7, 6].

Let \(\mathcal{R} = \{R^{(1)}, \ldots, R^{(r)}\}\) be the standard basis of a coherent algebra \(\mathfrak{W}\). We have \(R^{(\mu)} \circ R^{(\nu)} = p_{\mu, \nu}R^{(\kappa)}\) where intersection numbers

\[
p_{\mu, \nu} = \left| \left\{ z \in V \mid (x, z) \in \mu \land (z, y) \in \nu \right\} \right| \in \mathbb{N}_0
\]  

are the same for all pairs \((x, y) \in \kappa\). The \(r \times r\) matrices \(\hat{R}^{(\kappa)}\) with entries \(\hat{R}_{\mu, \nu}^{(\kappa)} = p_{\mu, \nu}^{(\kappa)}\) generate a matrix algebra \(\hat{\mathfrak{W}}\) that is isomorphic to \(\mathfrak{W}\) [63]. This
TABLE 1 Regularity properties of partitions.

<table>
<thead>
<tr>
<th>Property</th>
<th>i</th>
<th>ii</th>
<th>iii</th>
<th>iv</th>
<th>v</th>
<th>vi</th>
<th>vii</th>
<th>viii</th>
</tr>
</thead>
<tbody>
<tr>
<td>homogeneous</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>transitive</td>
<td></td>
<td>○</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>precaherent configuration</td>
<td></td>
<td>○</td>
<td>•</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>symmetric</td>
<td></td>
<td>○</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>class degree regular</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>homogeneous class deg. regular</td>
<td></td>
<td>○</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>coherent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>homogeneous coherent configuration</td>
<td></td>
<td>○</td>
<td>•</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>class degree regular cc</td>
<td></td>
<td>○</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>association scheme</td>
<td></td>
<td>○</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>symmetric association scheme</td>
<td></td>
<td>○</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(\mathcal{I} = \{(x, x) | x \in V\}\) is called the diagonal of \(V \times V\). The symbol \(\bullet\) indicates properties that are used for definition, while \(\circ\) marks additional properties that are implied by the definition.

observation makes coherent algebras appealing objects for our purposes because \(\mathfrak{M}\) is small enough in many cases to allow for explicit computations.

The action of a permutation group \((G, V)\) is transitive if for all \(x, y \in V\) there is \(g \in G\) such that \(y = g(x)\). If \(g \in G\) can be chosen such that \(y = g(x)\) and \(x = g(y)\) then \((G, V)\) is generously transitive. Generously transitive permutations groups have symmetric (and therefore commutative) centralizer algebras see [63] and [160, Thm. 29.3]. The group case is summarized in Table 2.

2.3.3 Association Schemes. If the coherent algebra \(\mathfrak{M}\) with standard basis \(\mathcal{R} = \{R^{(1)}, \ldots, R^{(r)}\}\) is commutative, we obtain a symmetric association scheme \(\mathfrak{M}^\circ\) by taking as basis elements \(R^{(\mu)}\) if \(\mu = \mu^T\) and \(R^{(\mu)} + R^{(\mu^T)}\) if \(\mu \neq \mu^T\), respectively. In particular, if a graph has a generously transitive or at least multiplicity-free group of automorphisms then \(\mathfrak{M}(\Gamma) = \langle A \rangle\) is a (symmetric) association scheme.
TABLE 2  Permutation Groups and Their Centralizer Algebras

<table>
<thead>
<tr>
<th>$G$</th>
<th>$\mathfrak{A}(G, V)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>transitive</td>
<td>$\iff$ homogeneous</td>
</tr>
<tr>
<td>multiplicity free</td>
<td>$\iff$ commutative</td>
</tr>
<tr>
<td>generously transitive</td>
<td>$\iff$ symmetric</td>
</tr>
</tbody>
</table>

The situation becomes particularly simple in this case. Since $\mathfrak{W}$ is a commutative algebra of symmetric matrices, a so-called Bose-Mesner algebra [14, 25], there is a common basis $\Phi = \{\varphi_i : V \to \mathbb{C}, i = 1, \ldots, |V|\}$ of eigenvectors of all matrices $M \in \mathfrak{W}$.

The following observation is also of interest in this context [136, Lemma 9]: The coherent algebra $\mathfrak{W}[\Gamma]$ of a graph is always a refinement of the distance partition of $\Gamma$ which has the classes $\delta_d = \{(x, y) \in V \mid d(x, y) = d\}$, i.e., $A^{(d)} \in \langle A \rangle$, where $A^{(d)}$ is the characteristic matrix of $\delta_d$. The class of distance regular graphs, which contains important examples such as the Hamming graphs $Q_n$ and the Johnson graphs, is characterized by the fact that the distance partition forms a (symmetric) association scheme. These graphs have received considerable attention, see e.g., [16].

2.3.4 Adjacency Algebra and Hoffman Algebras  The adjacency algebra of a graph $\Gamma$ is the matrix algebra generated by the adjacency matrix, $\mathfrak{A}[\Gamma] = \langle A \rangle$. Clearly, $\mathfrak{A}[\Gamma] \subseteq \mathfrak{W}[\Gamma]$. Higman [64] showed that $\mathfrak{A}[\Gamma] = \mathfrak{W}[\Gamma]$ if and only if $\mathfrak{W}[\Gamma]$ is commutative, i.e., an association scheme. It is interesting to note in this context that a homogeneous coherent algebra with rank $r \leq 5$ is always commutative [64]. So-called “orbit-polynomial graphs” characterized by $\mathfrak{A}[\Gamma] = \mathfrak{W}[\text{Aut}][\Gamma, V]$ are considered in [8].

A Hoffman algebra is matrix algebra $\mathfrak{H} \subseteq \mathbb{C}^{|V| \times |V|}$ such that (i) there is a basis consisting of non-negative integer matrices and (ii) $J \in \mathfrak{H}$ [86]. This notion is of interest as a generalization of coherent algebras and since $\mathfrak{A}[\Gamma]$ is a Hoffman algebra, i.e., $J \in \mathfrak{A}[\Gamma]$ if and only if $\Gamma$ is a connected regular graph [67].

2.3.5 Equitable Partitions.  Consider the set $\mathcal{R}$ of relations associated with the coherent algebra $\mathfrak{W}[\Gamma]$ of a graph $\Gamma$. Fix $x_0 \in V$ and define

$$\mu[x_0] = \{y \in V | (x, y) \in \mu\}$$

(20)

Clearly, $\Pi(x_0) = \{\mu[x_0] | \mu \in \mathcal{R}\}$ is a partition of $V$. If $\mathfrak{W}$ is homogeneous, then $\mu[x_0]$ is non-empty for all $\mu \in \mathcal{R}$, see Table 1, i.e., $\mathcal{R}$ and $\Pi(x_0)$ have the
same number of classes for all \( x_0 \in V \). In [64] it is shown that

\[
\mathbf{R}(x) = p_{\mu,\nu} = \sum_{x \in \nu[x_0]} \mathbf{R}(x) \text{ for each } x \in \kappa[x_0]
\] (21)

Noting that the adjacency matrix \( \mathbf{A} \) of \( \Gamma \) is a sum of \( \mathbf{R}(\mu) \)-matrices, it is shown in [136] that

\[
\sum_{x \in \nu[x_0]} \mathbf{A}_{xy} = \mathbf{A}_{\mu[x_0],\nu[x_0]} = \sum_{\kappa \in \mathcal{E}} p_{\mu,\nu} = \mathbf{A}_{\mu,\nu}
\] (22)

holds for any \( y \in \nu[x_0] \). Partitions of \( V \) that satisfy the first equality in equation (22) are called equitizable. Equitizable partitions have been introduced by Schwenk [128]; more recently they have been used by Powers and coworkers as “colorations”, see, e.g., [113, 111]. In [22, Chap.4] they appear as “divisors” of graphs.

The most important property of an equitizable partition \( \Pi \) is that all eigenvalues of the collapsed adjacency matrix \( \mathbf{A} \) are also eigenvalues of \( \mathbf{A} \). If \( \Pi \) contains a class that consists of single vertex, then the minimal polynomials of \( \mathbf{A} \) and \( \mathbf{A} \) are the same [13, Thm.8.6], i.e., the relevant spectral information is already contained in \( \mathbf{A} \). More information about equitizable partitions can be found in [50, 51, 52].

2.3.6 Fourier Transform on Finite Groups and Cayley Graphs. Let \( G \) be a finite group and let \( f : G \rightarrow \mathbb{C} \). Let \( \rho \) be matrix representation of \( G \). Then

\[
\hat{f}(\rho) = \frac{1}{|G|} \sum_{g \in G} f(g) \rho(g)
\] (23)

is called the Fourier transform\(^1\) of \( f \) at \( \rho \). The Fourier transform on a complete set \( \mathcal{R} \) of irreducible representations is inverted by

\[
f(g) = \frac{1}{|G|} \sum_{\rho \in \mathcal{R}} \dim \rho \text{ Tr} \left[ \hat{f}(\rho) \rho^{-1} \right]
\] (24)

Fast Fourier Transform algorithms are known for a variety of finite groups. For a recent overview see e.g., [97, 122].

It is not surprising that the spectral properties of Cayley graphs are intimately related to the Fourier transform on the underlying group. The crucial observation is the following. Let \( \delta \) be the characteristic function of the set of generators \( \Omega \). Then \( \sqrt{|G|} \delta \Omega(\rho_{\text{reg}}) \) is the Fourier transform of \( \delta \Omega \) at the regular representation of \( G \) equals the adjacency matrix of \( \Gamma(G,\Omega) \) up to a reordering of the group elements. Its spectrum is therefore the union of the spectra of \( \sqrt{|G|} \delta \Omega(\rho_i) \) where \( \rho_i \) are the irreducible representations of \( G \). If \( \Omega \) is a union of conjugacy classes of \( G \) the situation simplifies further [29].

\(^1\)In most of the literature the normalization factor \(|G|^{-1/2} \) is omitted
The irreducible representations are all 1-dimensional if $G$ commutative. Since $G$ can be written as a direct product of cyclic groups, $G = \prod_{k=1}^{m} C_{n_k}$, the characters are

$$\chi_k(x) = \exp \left( 2\pi i \sum_{k=1}^{m} \frac{x_k g_k}{n_k} \right)$$

(25)

where we use the additive representation of $C_{n_k}$ as $\{0, 1, \ldots, n_k - 1\}$ with addition modulo $n_k$. It is not hard to verify that the characters $\chi_k$ are eigenvectors of the adjacency matrix of each Cayley graph of $G$. The corresponding eigenvalue are $\sum_{\epsilon \in \Omega} \chi_k(\epsilon)$, see e.g. [94]. The Fourier transform on $C_2^n$ is also known as the Walsh-Hadamard transform. Note the that the Boolean hypercube can be regarded as a Cayley graph on this group. An FFT algorithm for this case is due to Yates [163]. Further material about the Cayley graphs on commutative groups can be found in [1].

3 LANDSCAPES

3.1 FOURIER DECOMPOSITION AND ELEMENTARY LANDSCAPES

Having derived a set of basis functions $\{\varphi_k|V \rightarrow \mathbb{C}\}$ from the structure of a configuration space $(V, \mathcal{X})$ by means of one of the approaches outlines in the previous section, it is natural to expand the fitness function $f$ in terms of this basis:

$$f(x) = \sum_k a_k \varphi_k(x)$$

(26)

We shall use the following convention: (i) The index 0 is reserved for the “ground state”. If the basis is derived from a Laplacian, for instance, then $\varphi_0$ is constant, the associated eigenvalue is zero, and

$$a_0 = \sum_x \varphi_0(x) f(x) = |V|^{-1} \sum_x f(x)$$

(27)

Similarly, the index 0 will refer to the stationary distribution in the case of a Markov chain on $V$. (ii) The distinct eigenvalues of $-\Delta$ will be denoted by $\lambda_p$, in the Markov chain case we write $\lambda_p$. It will be convenient to define the index sets $J_p = \{ k | - \Delta \varphi_k = \lambda_p \varphi_k \}$ that collect all eigenfunctions belonging to the same (Laplacian) eigenvalue. (iii) We write $\bar{f}(x) = f(x) - a_0$. If $\varphi_0$ is constant, this is the “non-flat” part of fitness function.

Lov Grover and others [18, 57, 136] observed that $\bar{f}$ is in many cases an eigenfunction of the graph Laplacian $-\Delta$, see Table 3 for a list of examples. We say that $f$ is elementary w.r.t. $-\Delta$ if $\bar{f}$ is an eigenfunction of $-\Delta$ with an eigenvalue $\lambda_p < 1$. In [142] this notion is extended to calling $f$ elementary...
TABLE 3  Elementary Landscapes

<table>
<thead>
<tr>
<th>Problem</th>
<th>Graph</th>
<th>$D$</th>
<th>$\lambda$</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-spin glass</td>
<td>$Q^n_p$</td>
<td>$n$</td>
<td>$2p$</td>
<td>$p$</td>
</tr>
<tr>
<td>NAES$^{(1)}$</td>
<td>$Q^n_p$</td>
<td>$n$</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Weight Partitioning</td>
<td>$Q^n_p$</td>
<td>$n$</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Graph $\alpha$-Coloring</td>
<td>$Q^n_{\alpha}$</td>
<td>$(\alpha - 1)/n$</td>
<td>$2\alpha$</td>
<td>2</td>
</tr>
<tr>
<td>XY-spin glass for $\alpha &gt; 2$:</td>
<td>$C^n_{\alpha}$</td>
<td>$(\alpha - 1)/n$</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>TSP symmetric</td>
<td>$\Gamma(S_n, T)$</td>
<td>$n(n - 1)/2$</td>
<td>2$(n - 1)$</td>
<td>2</td>
</tr>
<tr>
<td>$\Gamma(S_n, J)$</td>
<td>$n(n - 1)/2$</td>
<td>$n$</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>TSP antisymmetric</td>
<td>$\Gamma(A_n, C_n)$</td>
<td>$n(n - 1)(n - 2)/6$</td>
<td>$(n - 1)(n - 2)$</td>
<td>?</td>
</tr>
<tr>
<td>$\Gamma(S_n, T)$</td>
<td>$n(n - 1)/2$</td>
<td>$2n$</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>$\Gamma(S_n, J)$</td>
<td>$n(n - 1)/2$</td>
<td>$n(n + 1)/2$</td>
<td>$O(n)$</td>
<td></td>
</tr>
<tr>
<td>TSP Graph Matching</td>
<td>$\Gamma(S_n, T)$</td>
<td>$n(n - 1)/2$</td>
<td>$2(n - 1)$</td>
<td>2</td>
</tr>
<tr>
<td>Graph Bipartitioning</td>
<td>$J(n, n/2)$</td>
<td>$n^2/4$</td>
<td>$2(n - 1)$</td>
<td>2</td>
</tr>
</tbody>
</table>

w.r.t. a random walk transition operator iff $S\tilde{f} = \lambda_p \tilde{f}$ with an eigenvalue $\lambda_p < 1$.

If $f$ is elementary, then $\tilde{f}$ satisfies the conditions of Courant’s nodal domain theorem, see 2.2.3. Elementary landscapes thus can be expected to have few nodal domains if they belong to a small Laplacian eigenvalue (or to an eigenvalue of Markov transition matrix close to 1), while landscapes that are far away from the ground state will in general have many nodal domains. Such landscapes will appear “rugged”. Grover [57] showed that

$$f(\hat{x}_{\min}) \leq a_0 \leq f(\hat{x}_{\max})$$

(28)

where $\hat{x}_{\min}$ and $\hat{x}_{\max}$ are arbitrary local minima and maxima, respectively. This maximum principle shows that elementary landscapes are well-behaved: There are no local optima with worse than average fitness. We shall return to local optima as a measure of ruggedness in section 4.4.

In section 2.2.1 we have seen that $p$-spin (or Walsh-) functions are the eigenfunctions not only of mutation operators but also of recombination operators. Indeed, there is an intriguing relationship between elementary landscapes for string recombination and schemata sensu Holland [2, 10, 68, 69], see also Appendix A. Each recombination-elementary landscape corresponds to a partitioning of the set of strings. Each equivalence class in this partitioning is a schema in the sense of Holland and all the schema which make up this partitioning have the same positions fixed. An elementary landscape in this context is a landscape which assumes that only the fixed positions in the schema actually influence fitness. This was first noted by Weinberger in his seminal paper on Fourier and Taylor series of fitness landscapes [155]. In
[143, 153] it is shown rigorously that this is a legitimate way of decomposing
the configuration space of string recombination.

3.2 CORRELATION MEASURES

3.2.1 Random Walk Autocorrelation Functions The ruggedness of a landscape
is most easily quantified by measuring the correlation of fitness values in
“neighboring” positions. Weinberger [154, 155] suggested the following
procedure. Given a Markov process on \( V \), we sample the fitness values \( f(x^{(t)}) \),
interpret them as a time series, and compute the autocorrelation function of
this time series. Let \( T \) be the transition matrix of a such reversible Markov
process with stationary distribution \( \varphi_0 \). We define the scalar product

\[
\langle f, g \rangle_{\varphi_0} = \sum_{x \in V} f(x)\varphi_0(x)g^*(x)
\]

where \( a^* \) denotes the complex conjugate of \( a \). The (expected) autocorrelation
function along a \( T \)-random walk on \( V \) is then

\[
r(t) = \left( \sum_{x \in V} |\hat{f}(x)|^2 \varphi_0(x) \right)^{-1} \sum_{y \in V} \hat{f}(x)(T^t)_{xy}\hat{f}^*(y)\varphi_0(y) = \frac{\langle \hat{f}, T^t\hat{f} \rangle_{\varphi_0}}{\langle \hat{f}, \hat{f} \rangle_{\varphi_0}} \tag{30}
\]

Expanding \( f \) w.r.t. eigenvectors of \( T \) it can be shown [136] that

\[
r(t) = \sum_{p \neq 0} B_p \lambda_p^t \quad \text{with} \quad B_p = \frac{\sum_{i \in J_p} |a_i|^2}{\sum_{i \neq 0} |a_i|^2}. \tag{31}
\]

Thus a landscape \( f \) is elementary w.r.t. a transition operator \( T \) if and only if
the “random walk” autocorrelation function is exponential, \( r(t) = \lambda_p^t \). In this
case the the order \( p \) indicates to which eigenvalue (not counting multiplicities) \( \hat{f} \) belongs. On a Boolean hypercube \( \mathcal{Q}_p^2 \) we have eigenfunctions of the
form \( \sum_I a_I \prod_{k \in I} x_k \) where \( p = |I| \) is constant. These are exactly Derrida’s
[26] \( p \)-spin models. The order of the elementary landscape thus equals the
“interaction order” of the underlying spin glass model.

3.2.2 Amplitude Spectra Equ.(31) decomposes non-elementary landscapes in
a natural way into a superposition of elementary ones. The amplitudes \( B_p \)
measure the relative variance contributions of the different eigenspaces (or
“modes”). Instead of the random walk correlation function \( r(t) \) we can therefore
use the amplitude spectrum \( B_p, p \geq 1 \), as a measure for the ruggedness of
a landscape. In many cases it is much easier to interpret than the correlation
function, see e.g. [136, 46, 72]. This technique was applied successfully to re-
alistic landscapes such as those arising from RNA folding, see Figure 2 for an
example. The RNA secondary structure folding model is described in detail
in Peter Schuster’s contribution to this book.
3.2.3 Distance Correlation Functions  Most “early” work on RNA landscapes, e.g. [42, 145] uses a different type of correlation measure based on the Hamming distance. In [136, 139] a more general version starting with a collection of relations on \( V \) is introduced. Let \( \mu \) be a relation on \( V \). Then we set

\[
\varrho(\mu) = \frac{|V|^2 \sum_{(x,y) \in \mu} (f(x) - \bar{f})(f(y) - \bar{f})}{|\mu| \sum_{x,y \in V} (f(x) - \bar{f})(f(y) - \bar{f})}
\]  

(32)

where \( \bar{f} = |V|^{-1} \sum_{x} f(x) \). Thus \( \varrho(\mu) \) is the variation of points of vertices within a relation \( \mu \) compared to the variance of \( f \) over all configurations \( x \in V \). On Hamming graphs, for instance, it is natural to consider the distance classes, i.e., \((x,y) \in \mu_d \) if and only if \( d_H(x,y) = d \), a predefined value. Such distance-dependent correlation functions have been considered also for some combinatorial optimization problems [4, 5, 131, 138]. Given a partition of \( V \times V \), we may of course regard \( \varrho \) as a function of the classes of this parti-
4 RANDOM LANDSCAPES

4.1 BASIC DEFINITIONS

In many cases, for instance in applications to spin glasses, the definition of the landscape contains a number random parameters. We therefore define random landscapes as elements of an appropriate probability space, following the presentation in [118].

Let $V$ be a finite set and let $W$ be a predicate of landscapes $f : V \to \mathbb{R}$. A random $W$-landscape over $V$ is the probability space
\[ \Omega = \{ f : V \to \mathbb{R} \mid f \text{ has property } W \}, \]
where $\mathcal{A}$ is a $\sigma$-field and $\mu : \mathcal{A} \to [0,1]$ a measure. Let $\xi : \Omega \to \mathbb{R}$ be an $\Omega$-random variable; we denote expectation value and variance of $\xi$ by $\mathbb{E}[\xi]$ and $\mathbb{V}[\xi]$, respectively. In particular, we shall write $\mathbb{E}[f(x)]$ for the expected value of $f(x)$, i.e., of $\mathbb{E}[f]$ evaluated at $x \in V$.

The covariance matrix $\mathbf{C}$ of the random landscape $\Omega$, is given componentwise by
\[ C_{xy} = \mathbb{E}[f(x)f(y)] - \mathbb{E}[f(x)]\mathbb{E}[f(y)] \]
Clearly, $\mathbf{C}$ is a symmetric non-negative definite $|V| \times |V|$ matrix. Taking the set of all maps, $\{f : V \to \mathbb{R}\}$, as basis space of the probability space $\Omega$, a basis is formed by a set of orthonormal eigenvectors $\{\psi_k\}$ of the covariance matrix $\mathbf{C}$. An expansion of the form
\[ f(x) = \sum_k b_k \psi_k(x) \]
is known as Karhunen-Loève series or principal component decomposition. The symbol \( \hat{=} \) denotes equality almost surely. The importance of equ.(35) comes from the following classical result [73]: Then the coefficients $\{b_k\}$ in equ.(35) are uncorrelated random variables satisfying
\[ \text{Cov}[b_k, b_l] = \theta_k \delta_{kl}, \quad 1 \leq k, l \leq |V|, \]
where $\theta_k = \mathbb{V}[b_k]$ is the eigenvalue of $\mathbf{C}$ belonging to the eigenvector $\psi_k$.

A random landscape is elementary if $\mathbb{E}[r(t)] = \lambda^*_p$ or, equivalently, if $\mathbb{E}[B_p] > 0$ only for a single mode $p > 0$. While Ising spin glasses, TSPs with
random coefficients and other random parameter variants of combinatorial optimization problems are elementary, this is not true for Kauffman’s Nk models [139]. As a consequence, there are many landscapes that cannot be constructed as a superposition of Nk-models, for instance the Sherrington-Kirkpatrick spin glass [129], see [135, 60, 61] for a detailed discussion of the amplitude spectrum $E[B_n]$ of Nk models.

A random landscape is pseudo-isotropic [139] if there are constants $a_0, v$, and $w$ such that for all $x \in V$ holds (i) $E[f(x)] = a_0$, (ii) $V[f(x)] = v^2$, and (iii) $|V|^{-1} \sum_{y \in V} C_{xy} = w$. Pseudo-isotropy is a fairly weak regularity property that is satisfied by many random landscape models of practical importance, see Table 4.

4.2 ADDITIVE RANDOM LANDSCAPES

Many important random landscapes can be written as a sum of components with random coefficients. More precisely, let $M$ be finite index set, let $c_j, j \in M$ be independent, real valued random variables over appropriate probability spaces $\Omega_j = (\mathbb{R}, A_j, \mu_j)$, and let $\Theta = \{\vartheta_j : V \to \mathbb{R} | j \in M\}$ be a family of real valued functions on $V$. An additive random landscape (arl) is the probability space $(\Omega_V, \otimes_j A_j, \otimes_j \mu_j)$ with

$$\Omega_V = \{f : V \to \mathbb{R} | f(x) = \sum_{j=1}^{M} c_j \vartheta_j(x)\}.$$  \hspace{1cm} (37)

In other words, the random landscape is constructed as a linear combination of non-random landscapes $\vartheta_j$ with independent random coefficients $c_j$.

In particular, any Gaussian random landscape is additive: Using the Karhunen-Loève decomposition, equ.(35), any random landscape can be written as linear combination with uncorrelated random coefficients; uncorrelated Gaussian random variables are independent.

The most important additive random landscapes exhibit further regularities: An arl is uniform if and only if (i) the random variables $c_i, i \in M$, are i.i.d. and (ii) there exist constants $a, b \in \mathbb{R}$ such that $\sum_{x \in V} \vartheta_i(x) = |V|a$ and $\sum_{x \in V} \vartheta_i^2(x) = |V|b$. A uniform random landscape is strictly uniform if there exist constants $d, e \in \mathbb{R}$ such that $\sum_j \vartheta_i(x) = d$ and $\sum_j \vartheta_i^2(x) = e$. In [118] we show that a uniform random landscape is pseudo-isotropic if and only if at least one of the following two conditions is satisfied: (i) $F$ is strictly uniform, or (ii) $a = 0, E[c_i] = 0$, and there is a constant $e \in \mathbb{R}$ such $\sum_i \vartheta_i^2(x) = e$ for all $x \in V$.

4.3 ISOTROPY AS A MAXIMUM ENTROPY CONDITION

Uniformity and pseudo-isotropy are still rather weak properties. In [135, 139] the notion of an isotropic random landscape was introduced as a “statistically symmetric model”, that is, as a random landscape with a covariance matrix
TABLE 4  Examples of Additive Random Landscapes.

The component landscapes \( \psi_i \) and the index set \( M \), equ. (37), are listed together with information whether the models are uniform (U), strictly uniform (S), or pseudo-isotropic (P). As in Table 1, properties that are implied by stronger ones are shown as o.

<table>
<thead>
<tr>
<th>Model</th>
<th>Component Landscapes and Index set</th>
<th>U</th>
<th>S</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ising spin glass</td>
<td>( \psi_i(x) = \prod_{k \in I} x_k ) ( I \subseteq {1, \ldots, n} ) as above with (</td>
<td>I</td>
<td>= 2 )</td>
<td>o</td>
</tr>
<tr>
<td>SK model</td>
<td></td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>NK Landscapes</td>
<td>see [118]</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Graph Bipartitioning</td>
<td>( \psi_{ij}([A, B]) = \begin{cases} 1 &amp; \text{if } {i, j} \not\subseteq A, B, i &lt; j \ 0 &amp; \text{otherwise} \end{cases} )</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Asymmetric TSP</td>
<td>( \psi_{kl}(t) = \sum_i \delta_{k, r(i)} \delta_{l, r(i-1)} ) ( k \neq l )</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
</tbody>
</table>

that shares the symmetries of the underlying configuration space. More precisely, a random landscape is isotropic w.r.t. a partition \( \mathcal{R} \) of \( V \times V \) if there are constants \( a_0 \) and \( s \) and a function \( c : \mathcal{R} \to \mathbb{R} \) such that (i) \( \mathbb{E}[f](x) = a_0 \) and \( \mathbb{V}[f](x) = s^2 \) for all \( x \in V \), and (ii) \( C_{xy} = c(\mu) \) for all \( (x, y) \in \mu \), i.e., the covariance matrix \( C \) is constant on the classes \( \mu \in \mathcal{R} \).

The notion of isotropy for random landscapes is the analogue of stationarity for stochastic processes. Following the conventions of Karlin and Taylor [78] our notion of isotropy would be called “covariance isotropic”, “weakly isotropic”, or “wide sense isotropic”. For a Gaussian random landscape the notions of (weak) isotropy and strict isotropy coincide of course.

Not surprisingly, a useful theory does not arise by considering arbitrary partitions \( \mathcal{R} \), see Table 1. Isotropy is a stronger concept than pseudo-isotropy only if \( \mathcal{R} \) is sufficiently regular. Transitivity, for instance, ensures that the classes of \( \mathcal{R} \) are large enough to be interesting. In [139, thm.4] the following result is proved: Let \( \mathcal{F} \) be isotropic w.r.t. a homogeneous class degree regular partition \( \mathcal{R} \) of \( V \times V \). Then \( \mathcal{F} \) is pseudo-isotropic. Furthermore, suppose \( \mathbb{E}[f(x)] = a_0 \) for all \( x \in V \). Then the random landscape is isotropic w.r.t. a homogeneous coherent configuration if and only if \( C \in \langle \mathcal{R} \rangle \) [139].

If \( A \) is the adjacency matrix of an undirected graph (or more generally, the a symmetric transition matrix of a Markov process on \( V \) then we say that a random landscape is \(*\)-isotropic w.r.t. \( A \) if \( \mathbb{E}[f(x)] = a_0 \) and \( C \in \langle A \rangle \). For association schemes (such as those arising from distance regular graphs including the hypercube) isotropy and \(*\)-isotropy equivalent. In [139] we show that a random landscape is \(*\)-isotropic if and only the Fourier coefficients (w.r.t. an orthonormal basis of eigenvectors of \( A \)) satisfy: (i) \( \mathbb{E}[a_k] = 0 \) for \( k \neq 0 \), (ii) \( \text{Cov}[a_k, a_j] = \delta_{k,j} \mathbb{V}[a_k] \), and (iii) \( \mathbb{V}[a_k] = \mathbb{V}[a_j] \) if \( k, j \in J_p \). These conditions mean that the Fourier coefficients are uncorrelated and that they have the same mean and variance whenever they belong to the same mode.
(eigenspace of $\mathbf{A}$). Hence Fourier and Karhunen-Loève series coincide for *-isotropic landscapes.

For a random landscape with measure $\mu$ we define the entropy

$$ S = - \int \mu(f) \ln \mu(f) df $$

(38)

In appendix B we review some well-known properties of the entropy functional. In particular, $S$ can be decomposed into a "homogeneous" part $S_{\sigma^2}$ that only depends on $\sigma^2 = \text{Tr} \mathbf{C}$, the total variance of landscape, a terms $S_{\mathbf{C}}$ that depends only on the variations among the eigenvalues of $\mathbf{C}$, and a third term that measures the effect of deviations from the normal distribution given a fixed covariance matrix $\mathbf{C}$. Given $\mathbf{C}$, a random landscape maximizes entropy if and only if its Gaussian.

In the following we assume a Gaussian random landscape. Now suppose the values $\beta_p = \sum_{k \in L} \mathbb{E}[a_k] \sigma_k$ are prescribed, too. It follows from the discussion in Appendix B that the entropy is maximized if and only if the covariance matrix restricted to $J_p$ is a multiple of the identity, i.e., iff $\mathbb{E}[a_k]$ if constant on $J_p$. Given its amplitude spectrum, a random landscape therefore maximized entropy if and only if it is Gaussian and *-isotropic. This is of practical interest since the class *-isotropic models (on their natural configuration spaces) includes among others Derrida’s $p$-spin Hamiltonians, the graph-bipartitioning problem, and the TSP.

Most variants of Kauffman’s Nk-model, the XY-Hamiltonians, short-range Ising models, or the Graph-Matching Problem are not isotropic. This has important implications for the structure of these landscapes, as we shall see below.

### 4.4 LOCAL OPTIMA

Palmer [108] used the existence of a large number of local optima to define ruggedness. We say that $x \in V$ is a local minimum of the landscape $f$ if $f(x) \leq f(y)$ for all neighbors $y$ of $x$. The use of $\leq$ instead of $<$ is conventional [82, 124]; it does not make a significant difference for spin glass models. Local maxima are defined analogously. The number $\mathcal{N}$ of local optima of a landscape, however, is much harder to determine than its autocorrelation function $r(s)$ or its correlation length

$$ \ell = \sum_{k=0}^{\infty} r(s) = \sum_{p \neq 0} \frac{B_p}{1 - \lambda_p} $$

(39)

As it appears that $\mathcal{N}$ and $\ell$ are two sides of the same coin we search for a connection between the two quantities. In a random landscape setting it is customary to determine $\mathbb{E}[\ln \mathcal{N}]$. The only known case in which $\mathbb{E}[\ln \mathcal{N}] \neq \ln \mathbb{E}[\mathcal{N}]$ is the linear spin chain [27].
For the case of short range spin glasses, in which only a small number \( z \) of coupling constants \( J_{ij} \) are non-zero for any given spin \( i \), a slightly larger number of local optima has been found \([15, 147]\) than for the long-range Sherrington-Kirkpatrick model \([129]\). Since all Ising models have the same correlation length \( \ell = n/4 \) \([156, 133]\) but somewhat different values of \( \mathcal{N} \), we cannot hope for a general, exact formula relating \( \mathbb{E}[\ln \mathcal{N}] \) and \( \mathbb{E}[\ell] \).

From the maximum entropy interpretation of isotropy, however, we know that the expected density of meta-stable states in an isotropic Gaussian random landscape is determined completely by the expected correlation function \( \mathbb{E}[r(s)] \) because such a model simply does not contain any further information. In the case of an elementary isotropic random landscape the correlation length \( \ell \) already determines \( r(s) \) and hence there must be a direct relationship between \( \mathbb{E}[\ell] \) and the expected number of meta-stable states \( \mathbb{E}[\ln \mathcal{N}] \). Its functional form will of course depend on the geometric properties of \( \Gamma \).

Stadler and Schnabl \([141]\) conjectured that \( \mathbb{E}[\ln \mathcal{N}] \) can be estimated as follows: For a typical elementary landscape we expect that the correlation length \( \ell \) gives a good description of its structure because the landscape does not have any other distinctive features. By construction \( \ell \) determines the size of the mountains and valleys. As there are many directions available at each configuration we expect there are only very few meta-stable states besides the summit of each of these \( \ell \)-sized mountains – almost all of the configurations will be saddle points with at least a few superior neighbors. We measure \( \ell \) along a random walk but the radius \( R(\ell) \) of a mountain is more conveniently described in terms of the distance between vertices on \( \Gamma \). Here \( R(\ell) \) is the average distance that is reached by the random walk in \( \ell \) steps. With the notation \( B(R) \) for the number of vertices contained in a ball of radius \( R \) in \( \Gamma \) we expect approximately \( |V'|/B(R(\ell)) \) local optima.

As an example we consider a comparison of the correlation length conjecture with an exact computation based on the TAP equations \([148]\) for Derrida’s \( p \)-spin Hamiltonian \([26]\). The TAP approach yields \([56, 120]\):

\[
\lim_{n \to \infty} \frac{1}{n} \ln \mathbb{E}[\mathcal{N}] = \alpha(0) = \ln 2 - \frac{\delta(p)^2}{2(p-1)} + \ln \Phi(\delta(p)) \quad \text{with}
\]

\[
\Phi(x) = (1 + \text{erf}(\delta(p)/\sqrt{2}))/2 \\
\delta(p) = \frac{p-1}{\sqrt{2\pi}} \frac{\exp(-\delta(p)^2/2)}{\Phi(\delta(p))}
\]

(40)

The last equation must be solved numerically for \( \delta(p) \).

An explicit evaluation of the correlation length conjecture for the \( p \)-spin Hamiltonian \([146]\) yields

\[
\alpha(0) = \ln 2 + (1 - \zeta) \ln(1 - \zeta) + \zeta \ln \zeta \quad \text{where}
\]

\[
\zeta = \frac{1}{2} \left(1 - e^{-1/p}\right).
\]

(41)
Equ. (41) compares very well with data numerical simulations for \( p = 2, 3, \ldots, 6 \) [140]. Figure 3 shows that equ. (41) is in excellent agreement with the TAP result in equ.(40). It is interesting to note that even for very large \( p \) there is a good qualitative agreement between the value of \( \alpha^* = \ln 2 - \alpha(0) \) obtained by the two methods. We find \( \alpha^*_{\text{TAP}} \sim p^{-1} \ln p \) and \( \alpha^*_{\text{cl}} \sim (1/2p) \ln p \).

The correlation length conjecture works very well for isotropic random landscapes on various other configuration space besides Boolean hypercubes [141, 89]. On the other hand, the correlation length conjecture yields sometimes very poor estimates if the landscapes deviate significantly from isotropy [46].

### 4.5 Neutrality

We say two configurations \( x, y \in V \) are neutral if \( f(x) = f(y) \). We colloquially refer to a landscape as “neutral” if a substantial fraction of adjacent pairs of configurations are neutral. This should not be confused with the flat landscape, in which \( f \) is constant. Extensive computer simulations, based on RNA secondary structures [125, 58, 59], have revealed that neutrality plays an important role in understanding the dynamics of RNA evolution [75, 119, 117].

Kimura proposed a theory of biological evolution that focuses exclusively on the aspects of neutrality [83] by assuming a flat fitness landscape. Very recently, landscapes with a large degree of neutrality have also been described in computational models such as cellular automata [71], for the mapping of sequences in combinatorial random structures [116], and in the context of sequential dynamical systems [90].

Neutral landscapes are discussed in more detail in Christian Reidys’ contribution to this book. Here we restrict ourselves to indicating how a rigorous investigation of neutrality can be linked to the techniques described above. In the case of additive random landscapes, a promising starting point is provided...
by the random variables \( X_{\{y,x\}}(f) \) which take the value 1 if \( f(x) = f(y) \) and 0 otherwise.

The number of neutral neighbors of a configuration \( x \in V \) is then

\[
\nu_x(f) = \sum_{y \in \mathcal{N}(x)} X_{\{y,x\}}(f) \tag{42}
\]

The following parameters have turned out to be particularly important for understanding the neutrality in an additive random landscape:

\[
c_x(y) = |\{ j \in M \mid \vartheta_j(x) \neq \vartheta_j(y) \}| \\
w_x(y', y'') = |\{ j \in M \mid \vartheta_j(x) \neq \vartheta_j(y') \wedge \vartheta_j(x) \neq \vartheta_j(y'') \}| \tag{43}
\]

where \( x \in V \) is an arbitrary vertex and \( y, y', y'' \in \mathcal{N}(x) \). Note that \( c_x(y) \) and \( w_x(y', y'') \) only depend on the properties of the component landscapes \( \vartheta_k \) but not on the distribution of the coefficients \( c_k \).

In [118] explicit expressions for the mean and variance of neutrality are derived for the simplest possible case, namely for a distribution of the coefficients \( c_j \) satisfying

\[
\mu\{c_j = \xi\} = \begin{cases} 
\mu_0 > 0 & \text{if } \xi = 0 \\
0 & \text{otherwise} 
\end{cases} \tag{44}
\]

For any additive random landscape with coefficients \( c_i \) satisfying equ.(44) we obtain

\[
\mathbb{E}[\nu_x] = \sum_{y \in \mathcal{N}(x)} \mu_0 c_{\mathcal{N}(x)} \\
\mathbb{V}[\nu_x] = \sum_{y', y'' \in \mathcal{N}(x)} \mu_0 c_{\mathcal{N}(x)} \left[ \mu_0 w_x(y', y'') - 1 \right] \tag{45}
\]

More explicitly, the expected number of neutral neighbors of a \( p \)-spin landscape is therefore \( \mathbb{E}[\nu] = n \mu_0 \left( \frac{z^n}{z^n - 1} \right) \). Depending on \( \mu_0 \), the expected fraction of vanishing interaction coefficients, the fraction of neutral mutations \( \mathbb{E}[\nu]/n \) may take any value between 0 and 1. This fact is independent of the order \( p \) of the spin glass. Thus ruggedness (as measured by \( p \)) and neutrality (as measured by \( \mu_0 \)) are independent properties of (random) landscapes.

In many spin glass models the spins are arranged on a finite-dimensional lattice. Hence each spin has only a finite number of other spins to which it is coupled in such a short range spin glass. All but \( O(n) \) coefficients therefore vanish and we have \( \mu_0 \approx 1 - z/n^{p-1} \), where \( z > 0 \) is a parameter determined by the connectivity of the lattice. The fraction of neutral spin flips is constant in such systems, \( \mathbb{E}[\nu]/n \approx e^{-z} \), see [118] for more details. We remark further that the fraction of neutral mutations is the crucial input parameter for random graph models of neutral landscapes [47, 115, 119], see also Chr. Reidy’s contribution to this book.
The study of neutrality in more general classes of random landscapes requires the determination of the distributions of the random variables

$$\delta(x, x') := \sum_{i: \vartheta_i(x) \neq \vartheta_i(x')} c_i \left[ \vartheta_i(x) - \vartheta_i(x') \right]$$

(46)

Since the $c_i$ are by definition independent in an additive landscape, we have to compute the convolutions

$$g_{x,x'}(\delta) = \star_{i: \vartheta_i(x) \neq \vartheta_i(x')} \rho_i \left( \vartheta_i(x) - \vartheta_i(x') \right)$$

(47)

where $\star_j$ denotes convolution of all functions indexed by $j$, $g_{x,x'}(\delta)$ is the density of the values of $\delta(x, x')$, and $\rho_i(\cdot)$ is the density function of $c_i$, which in this case has to be evaluated with the argument $c_i/[\vartheta_i(x) \neq \vartheta_i(x')]$. Then we have

$$\text{Prob}[X_{x'}(x) = 1] = \lim_{\epsilon \to 0} \int_{-\epsilon}^{\epsilon} g_{x,x'}(\delta) d\delta .$$

(48)

If $\{i: \vartheta_i(x) \neq \vartheta_i(x')\} = \emptyset$, $x$ and $x'$ are neutral for any distribution of the $c_i$. This is observed for instance in the graph matching problem [132].

We conclude from equ.(48) that a continuous density $g_{x,x'}$, which necessarily arises if the individual densities $\rho_i$ are continuous, does not lead to neutrality. Neutrality hence depends on a “discrete” contribution to the probability densities of the coefficients $c_k$, $k \in M$. Indeed, only these discrete components influence neutrality. In practice, evaluation of the convolution (47) therefore boils down to a combinatorial exercise, as for instance in the case of the integer-valued NK model proposed by Newman and Engelhardt [102].

Finally we remark that $g(.)$ would also be the appropriate starting point for a theory of nearly neutral landscapes [104], in which the condition such as $|f(x) - f(x')| < \varepsilon$ for some finite $\varepsilon > 0$ could replace the condition $f(x) = f(x')$.

5 DISCUSSION

The exposition above has been focussed almost entirely on the “static” properties of a landscape. A mathematical language has been introduced that allows us to view a cost or fitness function as it is seen by a search operator. This formalism lends a precise meaning to notions such as ruggedness, neutrality, or isotropy. Intuitively, the dynamics of (evolutionary) adaptation and the performance of optimization heuristics should be determined by exactly these properties.

The spectral approach described here has been useful in distinguishing different types of landscapes. RNA folding landscapes, for instance, are very
different from spin glasses in both ruggedness and neutrality. A more detailed analysis of the distribution of the Fourier coefficients that belong to a particular mode might help to understand and quantify the structure of anisotropies.

Dynamics on landscapes, unfortunately, is much less understood at present. Apart from a few global results such as the “No Free Lunch Theorem” [162] and detailed studies on very simple landscapes such as [150], very few exact results are known. Various dynamical phenomena have been described for special classes of landscapes. There is an error-threshold limiting the mutation rate in biological evolution [32] which is well understood at least on landscapes with a few peaks. A tunnel effect was described between two separated peaks [30]. On the other hand, a diffusion-like process is observed on landscapes with a high degree of neutrality [75, 117] similar to the situation in a flat landscape [28]. A theory that could treat all these aspects within a common formalism, however, is still missing.

ACKNOWLEDGMENTS


APPENDIX

A SCHEMATA AND DECEPTIVENESS

A.1 Introduction. Walsh functions and “schemata” have been used extensively in the analysis of GA behavior, see e.g. [53, 54, 55, 92, 107, 37, 151, 152]. A schema is simply a hyperplane in sequence space. It is defined by the set $H$ of “fixed” bits and their values $h_i$, $i \in H$. In symbols

$$\mathcal{H} = H[h] = \{x \in V | \forall i \in H : x_i = h_i\}.$$  \hspace{1cm} (49)

For a discussion of the Schema Theorem and the Building Block Hypothesis we refer to the literature [2, 10, 68, 43, 69, 144]. Instead, we briefly consider a few properties of landscapes that a naturally defined in terms of schemata. For simplicity we restrict ourselves to landscapes on the set of binary strings of length $n$. Notion such as “local optimum” in the following subsections consequently refer to the graph structure of the Boolean hypergraph. Schemata and Walsh functions are linked by means of

$$f(\mathcal{H}) = \sum_{I \subseteq H} a_I \varepsilon(h) \quad \text{and} \quad \var(\mathcal{H}) = \sum_{I \subseteq H} \left\{ \sum_{K \subseteq H} a_K a_{K \triangle I} \right\} \varepsilon_I(h)$$  \hspace{1cm} (50)
where $I \bigtriangleup J$ denotes the symmetric difference of the sets $I$ and $J$. Note that these quantities are superpositions of Walsh functions with index set $I \subseteq H$ evaluated at the fixed bits of the schema. It is interesting that $f(H)$ depends only on Fourier coefficients $a_I$ with $I \subseteq H$, while $\text{var}(H)$ depends only on coefficients with $I \not\subseteq H$.

Let $Q$ be any property of a landscape $f$ on $V$. We say that $f$ is robustly $Q$, if there is an $\epsilon > 0$ such that any landscape $g$ satisfying $|f(x) - g(x)| < \epsilon$ for all $x \in V$ also has property $Q$. The condition $|f(x) - g(x)| < \epsilon$ may be replaced by $|a_k - \hat{a}_k| < \epsilon$ for the Fourier coefficients of $f$ and $g$, respectively. The results in following sections have not been published before; nevertheless we omit their (rather simple) proofs in this survey.

### A.2 Funnels

A landscape is called a **funnel** if there is a string $h^*$ such that $H \subseteq H'$ implies $f(H[h^*]) \leq f(H'[h^*])$ for all $H \subseteq \{1, \ldots, n\}$. If $f$ is a funnel with peak $h^*$ then (i) $f(h^*) \geq f$ and (ii) $h^*$ is a local maximum on $Q^n_2$. It is not hard to construct simple examples of funnels with just $n = 3$ bits showing that the peak of a funnel need not be globally optimal, and that there may be multiple local optima. With some more work it is also possible to show that a landscape is a robust funnel if and only if $H \subset H'$ implies $f(H[h^*]) < f(H'[h^*])$ for all $H$. The peak $h^*$ of a robust funnel is unique.

### A.3 GA-Easy Functions

Let us call a function **GA-easy** if there is a global optimum $x^*$ such that $f(H[x^*]) \geq f(H[x])$ for all $x \in V$ and all $H \subseteq [n]$. In the GA literature a more common definition of easy is what we call **robustly GA-easy**, namely a function $f$ with a global optimum $x^*$ satisfying $f(H[x^*]) > f(H[x])$ for all $x \in V$ and all $H \subseteq [n]$ for which $H[x^*] \neq H[x]$. By setting $H = \{1, \ldots, n\}$ we see that the global optimum of a robustly GA-easy function is unique. Robustly GA-easy is called "fully easy" in [91]. We prefer to say that $f$ is **fully GA-easy** if for each global optimum $x^*$, for each $H \subseteq [n]$, and for each $x \in V$ holds $f(H[x^*]) \geq f(H[x])$. Naturally, a fully GA-easy function is GA-easy, but the converse is not true. Note also that "robustly fully GA-easy" is the same as robustly GA-easy. A short computation shows that a (robustly) GA-easy function with global optimum $x^*$ is a (robust) funnel with peak $x^*$.

Linear functions $f$ are of course GA-easy. However, linear functions do not form a generic class of landscapes in the sense that linearity is not a robust property. Wilson [161] introduced a slightly larger class of landscapes: Given a string $x \in V$ let $\tilde{x}$ be a string satisfying $\tilde{x}_k = x_k$ if $f(x) > f(x^{(k)})$ and $\tilde{x}_k = \bar{x}_k$ if $f(x) < f(x^{(k)})$. That is, $\tilde{x}$ is obtained from $x$ by keeping the best bit among all 1-error mutants in each position. A function $f$ on $V$ is **bit setting optimizable** (b.s.o.) if $\tilde{x}$ is a global optimum for each $x \in V$. Of course linear functions are b.s.o., and the set of robustly b.s.o. functions is nonempty: all sufficiently small perturbations of linear functions are b.s.o. It can be shown that a (robustly) b.s.o. function is (robustly) GA-easy. It is shown in [161] that the converse is not true for all $n \geq 3$. 
A.4 Deceptive Functions  The literature on deceptive functions uses a variety of slightly different notions of deceptiveness (and sometimes does not even precisely define the notion at all). In a deceptive landscape an optimal schema of some size is “contradicted” by one of its sub-schemata. Intuitively, this is just the converse of GA-easy. Following Whitley [158] we use the following formal definition: A landscape \( f \) is **deceptive** if there are vertices \( x, y \in V \) and index sets \( H \subset K \subset \{1, \ldots, n\} \) with the following properties: (i) \( K[x] \neq K[y] \), (ii) \( f(H[x]) \geq f(H[z]) \) for all \( z \in V \) with \( H[x] \neq H[z] \), and (iii) \( f(K[y]) > f(K[z]) \) for all \( z \in V \) with \( K[y] \neq K[z] \). As expected, it can be shown that a GA-easy function \( f \) is not deceptive. However, the converse is not true since a “symmetric function”, i.e., a function fulfilling \( f(x) = f(\bar{x}) \) where \( \bar{x} \) is the complement of \( x \) is never deceptive according to Whitley’s definition. An example of a symmetric function that is not not GA-easy can be constructed e.g. on \( \mathbb{Q}_2 \). We will say that \( f \) is **GA-hard** if it is not GA easy.

We say that \( f \) is **weakly deceptive** if there are vertices \( x, y \in V \) and index sets \( H \subset K \subset \{1, \ldots, n\} \) such that (i) \( f(H[x]) \geq f(H[z]) \) for all \( z \in V \) and \( f(H[y]) > f(H[z]) \) for all \( z \in V \) with \( f(K[y]) > f(K[z]) \). The deceptiveness is of course weakly deceptive. Since symmetric functions can be weakly deceptive, the converse is not true in general.

Let \( \Omega \) denote the set of global optima. For each \( x^* \in \Omega \) and each index set \( H \) we define the set of vertices that belong to a \( H \)-schema that is superior to \( H[x^*] \):

\[
\Psi(x^*, H) = \{ y | f(H[y]) > f(H[x^*]) \text{ and } \forall z : f(H[y]) \geq f(H[z]) \}. \quad (51)
\]

Clearly \( x^* \notin \Psi(x^*, H) \). The following propositions are easily verified:

(i) \( f \) is GA-hard iff for each \( x^* \in \Omega \) there is an index set \( H \) such that \( \Psi(x^*, H) \) is non-empty.

(ii) \( f \) is not fully GA-easy iff there is a \( x^* \in \Omega \) and an index set \( H \) such that \( \Psi(x^*, H) \) is non-empty.

(iii) \( f \) is weakly deceptive iff there is a \( x^* \in \Omega \) and an index set \( H \) such that \( \Psi(x^*, H) \setminus \Omega \) is non-empty.

Both weak deceptiveness and GA-hardness imply that \( f \) is not fully GA-easy. If \( f \) has a unique global optimum, however, then “weakly deceptive”, “GA-hard”, and not “fully GA-easy” are equivalent properties.

### B: MAXIMUM ENTROPY CONDITIONS

It is well known the Gaussian distributions maximize entropy. The proof for the one-dimensional case can be found e.g. in [62, prop. 1.15]. For the convenience of the reader a short proof of the general case is included here as it is not readily accessible in the literature.

The starting point is the following inequality that holds for arbitrary probability spaces:

\[
\int p(x) \ln p(x) dx - \int p(x) \ln q(x) dx \geq 0 \quad (52)
\]
Equality holds in equ.(52) holds if and only if \( p = q \) almost everywhere. Let \( \mathbf{C} \) be the covariance matrix of \( p \). We assume that \( \mathbf{C} \) is invertible. Without loosing generality we furthermore assume \( \mathbb{E}[x] = 0 \). Substituting the Gaussian distribution

\[
q(x) = \frac{1}{(2\pi)^{V/2}\sqrt{\det \mathbf{C}}} \exp \left( -\frac{1}{2} x^T \mathbf{C}^{-1} x \right)
\]

equ.(52) translates into a general inequality for the the entropy of \( p \):

\[
S \leq \frac{|V|}{2} \ln(2\pi) + \frac{1}{2} \ln \det \mathbf{C} + \frac{1}{2} \int_{\mathbb{R}^n} (x \mathbf{C}^{-1} x)p(x)dx
\]

The integral is a simple constant independent of \( p \) as the following computation shows

\[
\int_{\mathbb{R}^n} \sum_{k,l}(\mathbf{C}^{-1})_{kl}x_kx_lp(x)dx = \sum_{k,l}(\mathbf{C}^{-1})_{kl} \int_{\mathbb{R}^n} x_kx_lp(x)dx = \\
\sum_{k,l}(\mathbf{C}^{-1})_{kl} \mathbf{C}_{kl} = \sum_k (\mathbf{C}^{-1})_{kk} = \text{Tr}\mathbf{I} = |V|.
\]

Let \( \{\lambda_k, k = 1, \ldots, |V|\} \) be the eigenvalues of \( \mathbf{C} \). Since \( \mathbf{C} \) is invertible by assumption we have \( \lambda_k > 0 \) for all \( k \). Using \( \sigma^2 = \text{Tr}\mathbf{C} = \sum_k \lambda_k \) we obtain

\[
S \leq S_C = \frac{1}{2} |V| \ln \frac{2\pi e |V|}{|V|} + \frac{1}{2} \sum_k \ln \frac{\lambda_k |V|}{\sigma^2}
\]

It is easy to verify that \( S_C \) is indeed the entropy of a Gaussian distribution with covariance matrix \( \mathbf{C} \).

The two terms in equ.(54) allow for a direct interpretation. The Gaussian entropy \( S_C \) attains its maximum subject to a given variance \( \sigma^2 \) if and only if \( \lambda_k = \sigma^2/|V| \), in which case the second term vanishes. We may this split the entropy of a random landscape into three contributions

\[
S = S_{\sigma^2} + \Delta S_C + \Delta S_{ng}
\]

where \( \Delta S_{ng} = S - S_C \) is the entropy loss due to deviations from a Gaussian distribution, \( S_{\sigma^2} \) is the maximal entropy with given variance \( \sigma^2 \), and \( \Delta S_C \), the second term in equ.(54), measures the entropy loss due to variations in the spectrum of \( \mathbf{C} \). In particular, whenever there are correlations between different vertices, then \( \mathbf{C} \) is non-diagonal and hence \( \Delta S_C < 0 \). More precisely, \( \Delta S_C = 0 \) if and only if the corresponding Gaussian random landscape is i.i.d.

REFERENCES


[36] Spectral Landscape Theory


