Towards a Theory of Landscapes

By

Peter F. Stadler[a,b,*]

[a]Institut für Theoretische Chemie, Universität Wien
Währingerstraße 17, A-1090 Wien, Austria
[b]Santa Fe Institute
1399 Hyde Park Rd., Santa Fe, NM 87501, USA

1. Introduction

Since Sewall Wright's seminal paper [145] the notion of a *fitness landscape* underlying the dynamics of evolutionary adaptation optimization has proved to be one of the most powerful concepts in evolutionary theory. Implicit in this idea is a collection of genotypes arranged in an abstract metric space, with each genotype next to those other genotypes which can be reached by a single mutation, as well as a value assigned to each genotype. Such a construction is by no means restricted to biological evolution; Hamiltonians of disordered systems, such as spin glasses [14, 88], and the cost functions of combinatorial optimization problems [52] have the same mathematical structure. It has been known since Eigen's [39] pioneering work on the molecular quasispecies that the dynamics of optimization on a landscape depends crucially on detailed structure of the landscapes itself. Extensive computer simulations, see, e.g., [46, 45, 2], have made it very clear that a complete understanding of the dynamics is impossible without a thorough investigation of the underlying landscape [40].

The landscapes of a number of well known combinatorial optimization problems[1] such as the Traveling Salesman Problem `TSP`, the Graph Bipartitioning Problem `GBP`, or the Graph Matching Problem `GMP` have been investigated in some detail [127, 121, 125]. The distribution of local optima and the statistical characteristics of down-hill walks have been computed for the uncorrelated landscape of the random energy model [82, 81, 42, 4]. Furthermore, two one-parameter families of tunably rugged landscapes have been studied extensively: the Nk model and its variants [71, 72, 142, 47] and the p-spin models [2, 139]. A detailed survey of a variety of model landscapes derived from folding RNA molecules into their secondary structures has been performed recently [18, 46, 45, 43, 47, 44, 62, 65, 66, 108, 109, 110, 124, 131, 132].

In this contribution I will present an overview over the mathematical techniques that have proved useful for an analysis of landscapes. This approach is based on Fourier Series defined on the highly symmetric graphs that occur as configuration spaces underlying combinatorial optimization problems and evolutionary dynamics. We will consider both stochastic models of landscapes, so-called random fields, and individual "measured" landscapes.

Section 2 deals with the structure of the configuration spaces. The mathematical language described there, mostly permutation group theory and graph theory, seems to be the appropriate basis for a general theory of (fitness) landscapes.

Section 3 discusses random fields. The theory of random fields at this points is mostly a "second order theory". That is to say, it revolves around covariance and correlation measures. The main emphasis lies here on *isotropic* random fields and their correlation functions. Among other results we give a characterization of isotropic random fields in terms of their Fourier coefficients. Then we briefly discuss the relation between the autocorrelation function of random fields and the correlation function of the time series obtained from a random walk on the random field. Furthermore we investigate superpositions and transformations of random fields.

Section 4 contains the theory of individual landscapes. The most surprising result concern the relations between "elementary landscapes",which are eigenvectors of the graph Laplacian Δ, and their autocorrelation functions. The relation of the correlation structure and the distribution of local optima of the landscape is subject of on-going research; first results are discussed in subsection 4.3. The notion of elementary landscapes leads to a refined classification scheme for landscapes. The final subsections contain material about landscapes on irregular graphs and on anisotropies in landscapes on highly regular graphs.

Section 5 provides a condensed review of RNA landscapes and a generalization of the theory of landscapes to sequence-structure mappings.

[1] See section 4.4 for details.

2. Configuration Spaces

2.1. An Example — TSP

The travelling salesman problem (TSP) [77] is a classical example of an NP-complete [52] combinatorial optimization problem. Few mathematical problems have attracted as much attention. Given a distribution of cities the task is to find the shortest tour visiting each city once and returning to the starting point with prescribed costs w_{ij} for traveling from i to j. The cost function is

$$f(\tau) = \sum_{i=1}^{n-1} w_{\tau(i)\tau(i+1)} + w_{\tau(n)\tau(1)}$$

where τ is the permutation encoding the order of the cities. The symmetric problem $w_{ij} = w_{ji}$ has applications in X-ray crystallography [15], electronics [24], and the study of protein conformations [16]. The asymmetric case [89] has applications in scheduling chemical processes or from pattern allocation problems in glass industry.

Besides the cost function there is another ingredient in the problem that is as important: the combinatorial structure of the configurations (in our case the possible tours of the salesman). Heuristic algorithms designed to find approximate solutions of the TSP use the internal structure of configurations to arrange them (hopefully) in a way which simplifies the search. Typically the search proceeds from configuration to a neighboring one.

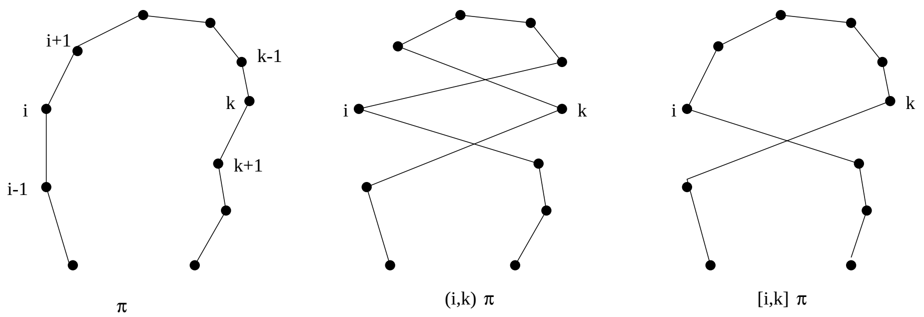


Figure 1: The two most common types of elementary moves for the TSP are transpositions, $(i,k)\pi$, and inversions, $[i,k]\pi$ of a tour π. A transposition of the form (i,i+1) is called *canonical transposition*.

What does neighboring mean in the case of the TSP? There is no unique answer, instead a few different ways of defining neighborhood come to mind. For a mathematician the most natural choice is probably to use the exchange of two cities along the tour (a transposition) as elementary "move". Alternatively one might allow only the exchange of subsequent cities; this operation is called a "canonical transposition". Lin and Kernighan [78] suggested to reverse the order of the cities in a part of the tour, this move is called inversion, reversal, or 2opt move. These three types of moves are displayed in figure 1.

All landscapes are composed of these two ingredients:

(1) a *cost function* or *fitness function* f assigning a value to each *configuration*, and

Table 1. The Correspondence of Statistical Mechanics, Combinatorial Optimization and Evolutionary adaptation.

Combinatorial Optimization	Statistical Mechanics	Folding of Biopolymers	Evolutionary Adaptation	Symbol
system size	system size	chain length	chain length	n
configuration	micro-state	structure	sequence	x
cost	energy	energy	fitness	$f(x)$
pseudo-temperature	temperature	temperature	mutation-rate	$T = -\ln q$
optimum	ground state	native state	fittest	o
instance	sample	*	*	f
move	(spin flip)†	†	mutation	

† There is no generic term for the elementary process of the dynamics
* Ideally, there is only one energy or fitness function in a given system

(2) a rule determining whether two configurations x and y are neighbors of each other.

As a consequence of (2), the set $\mathcal{C}$ of all configurations can be viewed as a graph Γ with each vertex corresponding to a configuration and an edge between any two nearest neighbors. A landscape is therefore simply a real valued function of the vertex set V of some graph Γ. A list of corresponding terms for landscapes occurring in different fields is presented in table 1.

Often we are not given a single landscape but a model containing a large number of parameters which are assigned randomly. It is customary, for instance, to consider TSPs with the cities randomly distributed in the unit square [77]. Such a model is not a landscape but consists of an entire probability space whose elements are landscapes. These so-called *random fields* will be discussed in section 3.

2.2. Groups

2.2.1. Permutation Groups

In this section we briefly recall the basic definitions of the theory of permutation groups as far as we will need them in this contribution. An excellent introduction to permutation groups is Wielandt's book [144]. Let X be a finite set and G a group acting on X, notation (G, X). The group identity will be denoted by e. The cardinality of X is the *degree* N of the permutation group. The cardinality of G itself is called the *order* of G. The *permutation matrix* associated with a group element $g \in G$ is

$$g_{ij} \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } g(i) = j \\ 0 & \text{otherwise} \end{cases}$$

In this notation we identify an element of the abstract group G with its representation as a permutation matrix. Analogously, we will write G also for the group of permutation matrices corresponding to the group elements $g \in G$.

$D(x) \stackrel{\text{def}}{=} \{y \in X | y = g(x), \forall g \in G\}$ is an *orbit* of G in X. The orbits of G partition the set X. We have $y \in D(x) \implies D(y) = D(x)$.

The *centralizer* of G in the algebra of all $N \times N$ matrices over $\mathbb{C}$, i.e.,

$$\mathbb{C}[G,X] \stackrel{\text{def}}{=} \{Q \in \mathbb{C}^{N\times N} \,|\, Qg = gQ \quad \text{for all} \quad g \in G\} ,$$

is called the *commuting algebra* of G on X. It seems to play a key role in the study of highly symmetric configuration spaces. We will also need a few simple properties of permutation groups.

Definition. Let G be a permutation group acting on X.

- G acts *transitively* in X if there is only a single orbit of G.
- G acts *regularly* on X if $g(x) = x$ for some x implies that g is the group identity.
- A subset $Y \subset X$ such that for each $g \in G$ we have either $g(Y) = Y$ or $g(Y) \cap Y = \emptyset$ is called a block. The one-point subsets of X, and X itself, clearly fulfil this requirement for all permutation groups; they are usually referred to as trivial blocks. G acts *primitively* on X if there are no non-trivial blocks.

2.2.2. Orbits, Orbitals, and Symmetry Classes

If G acts transitively on X it does not necessarily so[2] on $X \times X$. We will denote the orbits of G on $X \times X$ by $\ddot{\kappa}$, $\ddot{\mu}$, $\ddot{\nu}$, etc.

There is a one-to-one relation-ship between the orbits of G on $X \times X$ and the orbits of the *stabilizer subgroup*

$$G_0 \stackrel{\text{def}}{=} \{g \in G \,|\, g(0) = 0\}$$

of an arbitrary element $0 \in X$:

$$\begin{aligned}
\dot{\mu} &= \{v \in X \,|\, (v,0) \in \ddot{\mu}\}, \\
\ddot{\mu} &= \{\, (u,v) \in X \times X \,|\, \exists \alpha \in G : \alpha(v) \in \dot{\mu} \,\wedge\, \alpha(v) = 0 \,\} \, .
\end{aligned}$$

An *orbital* is a mapping Ψ from X into the subsets of X such that

(i) $\Psi(x)$ is a G_x-orbit $x \in X$.

(ii) $g(\Psi(x)) = \Psi(g(x))$.

Let $\Psi_\mu(x)$ denote the orbitals, i.e., $\Psi_\mu(x)$ is the orbit of the stabilizer G_x of x that corresponds to the orbit $\dot{\mu}$ of G_0, the stabilizer of the "reference vertex" 0. We have therefore $x \in \Psi_\mu(y)$ if and only if $(x,y) \in \ddot{\mu}$. We will refer to μ as the *symmetry class* to which the orbits $\dot{\mu}$ and $\ddot{\mu}$ as well as the orbital Ψ_μ belong. The mirror image of an orbital is defined by

$$\Psi^+(x) = \{g^{-1}(x) | g(x) \in \Psi(x)\},$$

it is again an orbital of G acting on X. Consequently, for each μ there is a ν such that $\Psi^+_\mu = \Psi\nu$. We write $\nu = \mu^+$.

The symmetry class corresponding to the diagonal on $X \times X$ will be denoted by $\mathbf{0}$, i.e., $\ddot{\mathbf{0}} = \{(x,x)|x \in X\}$. The number of symmetry classes is $M + 1$.

The cardinalities of the sets $\Psi_\mu(x)$ are of course independent of x. It is often called a *subdegree* of G; it will be denoted here by $|\mu|$.

[2] $g(x,y) \stackrel{\text{def}}{=} (g(x), g(y))$

2.2.3. Incidence Matrices

A central part of the theory discussed in this contribution are the matrices $R^{(\mu)}$ defined by

$$R^{(\mu)}_{xy} = \begin{cases} 1 & \text{if } (x,y) \in \ddot{\mu} \\ 0 & \text{otherwise} \end{cases} = \begin{cases} 1 & \text{if } x \in \Psi_\mu(y) \\ 0 & \text{otherwise} \end{cases}$$

Higman [59] called them *incidence matrices* of the transitive permutation group G on X. The mirror image of a symmetry class belongs to the transpose of the corresponding incidence matrix: $(R^{(\mu)})^+ = R^{(\mu^+)}$. The collection of the $M+1$ matrices $R^{(\mu)}$ will be denoted by $\mathfrak{R}$.

Definition. Let $\mathfrak{A} = \{A_0, A_1, \ldots, A_M\}$ be a collection of $N \times N$ 0-1 matrices with the following properties:

(i) $A_0 = I$, the identity matrix

(ii) $\sum_{j=0}^{M} A_j = J$, the matrix with all entries 1.

(iii) For all $0 \le i, j \le M$ we have $A_i A_j = \sum_{k=0}^{M} s_{ijk} A_k$.

(iv) For all $0 \le i \le M$ there is a j such that $A_i^+ = A_j$.

The we call $\mathfrak{A}$ an *incidence scheme.*

Properties (i), (ii), and (iii) imply that $\mathfrak{A}$ forms the basis of an algebra of dimension $M+1$; we will denote this algebra by $\langle\mathfrak{A}\rangle$.

A well known result is the following:

Theorem. Let X be a finite set and let G be a permutation group operating transitively on X. Then the collection of the matrices $\mathfrak{R}$ forms an incidence scheme. In fact, the collection $\mathfrak{R}$ generates the commuting algebra of G on X:

$$\langle\mathfrak{R}\rangle = \mathfrak{C}[G, X].$$

Proof. It is trivial that (i), (ii), and (iv) are satisfied. It is shown in [144, Thm. 28.4] and [10, Thm. 4.3.6] that the matrices $R^{(\mu)}$ form a basis of the commuting algebra $\mathfrak{C}[G, X]$, see also [59], and thus (iii) holds as well. ∎

2.2.4. Intersection Numbers and Collapsed Matrices

We define the *intersection numbers* of the orbitals Ψ_κ and Ψ_μ with respect to the orbital Ψ_ν by

$$s^{(\mu)}_{\kappa\nu} \stackrel{\text{def}}{=} |\Psi_\mu(u) \cap \Psi_\kappa(w)| \qquad \text{with} \qquad u \in \Psi_\nu(w).$$

It is easy to see that these numbers really depend only on κ, μ, and ν [59]. The *intersection matrix* of the orbital Ψ_μ is the matrix with entries $(s^{(\mu)}_{\kappa\nu})$.

The intersection numbers are of course closely related to the incidence matrices. In fact, we have the following

Lemma. $R^{(\mu^+)} R^{(\nu)} = \sum_\kappa s^{(\mu)}_{\nu\kappa} R^{(\kappa)}$.

Proof. We have $[R^{(\mu^+)}R^{(\nu)}]_{xy} = \sum_z R^{(\mu^+)}_{xz}R^{(\nu)}_{zy} = \sum_z R^{(\mu)}_{zx}R^{(\nu)}_{zy}$
$= \left|\{z|z \in \Psi_\mu(x) \wedge z \in \Psi_\nu(y)\}\right| = |\Psi_\mu(x) \cap \Psi_\nu(y)| = s^{(\mu)}_{\nu\kappa}$,
where κ is such that $x \in \Psi_\kappa(y)$, i.e., it is the coefficient of R^κ_{xy}. ∎

The intersection matrices provide an elegant means for determining if G acts primitively on X:

Lemma. [59, (4.8)] G acts primitively on X if and only if all intersection matrices $R^{(\mu)}$ are irreducible.

Higman [59] shows that the intersection matrix $S^{(\mu)}$ can be obtained from the incidence matrix $R^{(\mu)}$ by the following procedure:
Suppose the points of X are arranged according to the symmetry classes μ, and consider the corresponding blocks of $R^{(\mu)}$ (see the example below). Each block has by construction constant column sum. Construct the matrix $\hat{R}^{(\mu)}$ by replacing each block by its column sum. Using the terminology in [17] we call a matrix obtained by this procedure *collapsed*.

Theorem. (i) The collapsed incidence matrices are the intersection matrices, i.e., $\hat{R}^{(\mu)}_{\kappa\nu} = s^{(\mu)}_{\kappa\nu}$.
(ii) The intersection matrices span an algebra $\hat{\mathfrak{C}}[G,X]$ which is isomorphic to the commuting algebra $\mathfrak{C}[G,X]$.

Proof. See [59]. ∎

2.2.5. Association Schemes

Definition. Let $\mathfrak{A} = \{A_0, A_1, \ldots, A_M\}$ be an incidence scheme. $\mathfrak{A}$ is an *association scheme* if we have in addition
(v) For all $0 \le i,j \le M$ holds $A_iA_j = A_jA_i$.
If (iv) is replaced by
(iv') $A_i^+ = A_i$ we call $\mathfrak{A}$ a *symmetric association scheme.*

Commutativity (v) follows already from (i), (ii), (iii), and (iv'). Note that Godsil [54] uses the term 'association scheme' for symmetric association schemes. We use here Delsarte's [29] original terminology.

The theory of association schemes plays a fundamental role in algebraic combinatorics. Many questions concerning distance regular graphs are best discussed in this framework. It plays a fundamental role in coding theory, and is important for the theory of polynomial spaces. For a recent textbook on this topic see [54].

Theorem. The incidence matrices form an association scheme, i.e., the commuting algebra of (G,X) is commutative, if and only if the irreducible constituents of the permutation representation are inequivalent.

Proof. The argument in [59, (4.10)] is based on [144, Thm. 29.3]. ∎

We will not use the incidence matrices $R^{(\mu)}$ themselves but we will rather work with the *symmetrized incidence matrices* $\tilde{R}^{(\mu)}$ defined by

$$\tilde{R}^{(\mu)} = \begin{cases} R^{(\mu)} & \text{if} \quad \mu = \mu^+ \\ R^{(\mu)} + R^{(\mu^+)} & \text{if} \quad \mu \neq \mu^+ \end{cases}$$

Of course, these matrices are again 0-1 matrices. It will be an important question under which conditions they form again an incidence scheme. If this is the case, then they form even a symmetric association scheme which will be denoted by $\tilde{\mathfrak{R}}$. Obviously, a trivial sufficient condition is that $R^{(\mu)} = R^{(\mu^+)}$ for all symmetry classes μ, i.e., if all orbits of G are "self-paired". A less trivial condition can be obtained from the above theorem:

Lemma. If the incidence matrices $R^{(\mu)}$ form an association scheme then the symmetrized incidence matrices form a symmetric association scheme.

Proof. If the incidence matrices $R^{(\mu)}$ commute, so do the symmetrized incidence matrices $\tilde{R}^{(\mu)}$. It remains to show that $\tilde{R}^{(\mu)}\tilde{R}^{(\nu)}$ is a linear combination of symmetrized incidence matrices. The product of two symmetric matrices is symmetric if and only if they commute, thus the product $\tilde{R}^{(\mu)}\tilde{R}^{(\nu)}$ is again a symmetric matrix. We have

$$\tilde{R}^{(\mu)}\tilde{R}^{(\nu)} = \sum_{\kappa} p(\kappa)R^{(\kappa)} = (\tilde{R}^{(\mu)}\tilde{R}^{(\nu)})^{+} = \sum_{\kappa} p(\kappa)(R^{(\kappa^+)}).$$

The symmetry of the product implies $p(\kappa) = p(\kappa^+)$, i.e., the product is in fact a linear combination of the symmetrized incidence matrices. ∎

2.3. Graphs

2.3.1. Graphs and Their Associated Matrices

ADJACENCY MATRIX.

A graph consists of a set V of N vertices and a set of edges. Two vertices x and y are neighbors (one says that x and y are adjacent) if they are joined by an edge. The set of edges will be denoted by E. We will denote the set of neigbors of a vertex x by $N(x)$. The number of neighbors, $|N(x)|$ is called the *degree* of the vertex x. We will need the diagonal *vertex degree matrix* D which has the entries $D_{xy} \stackrel{\text{def}}{=} |N(x)|\delta_{xy}$, where δ_{ij} is Kronecker's symbol. A graph is *regular* if all vertices have the same degree. In this case we will use D to denote the common vertex degree, i.e., the vertex degree matrix is DI, where I is as usual the identity matrix.

A graph Γ is uniquely described by the $N \times N$ matrix A defined by

$$A_{xy} \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } y \in N(x) \\ 0 & \text{otherwise.} \end{cases}$$

It is called the *adjacency matrix* of Γ as its non-zero entries correspond to the pairs of adjacent vertices; therefore A is symmetric.

A *path* of length ℓ is a sequence of vertices $x_0, x_1, \ldots, x_\ell$ such that $\{x_{i-1}, x_i\}$ is an edge for $1 \le i \le \ell$. A well known result, see, e.g., [12] links the number of paths of given length with the adjacency matrix:

Lemma. The entry x, y in the s-th power of the adjacency matrix, A^s, equals the number of paths of length s from x to y. For later reference we note here the following

Definition. A *simple random walk* [120] on Γ starting at 0 is a Markov process with transition matrix

$$T = D^{-1}A$$

and initial condition $p_0(x) = \delta_{x,0}$.

DISTANCE MATRICES.

The notion of a path suggest a natural definition of a distance between two vertices of a graph: let $d(x,y)$ denote the minimum length of a path joining the vertices x and y. In this contribution we will consider only finite connected graphs. Thus $d(x,y)$ is finite for any two vertices. It is easy to check that this distance measure is metric on the vertex set of Γ, i.e., for all $x,y,z \in V$ the following is true:

(i) $d(x,y) = 0$ if and only if $x = y$,
(ii) $d(x,y) = d(y,x)$, and
(iii) $d(x,y) \le d(x,z) + d(z,y)$.

It is natural to consider the sets

$$N_k(x) \stackrel{\text{def}}{=\!=} \{y \in V \mid d(x,y) = k\}$$

and the *distance matrices* $A^{(k)}$ with entries

$$A^{(k)}_{xy} \stackrel{\text{def}}{=\!=} \begin{cases} 1 & \text{if} \quad d(x,y) = k \\ 0 & \text{otherwise} \end{cases}$$

Of course, $A^{(0)} = I$, the identity matrix, and $A^{(1)} = A$, the adjacency matrix.

INCIDENCE MATRIX AND GRAPH LAPLACIAN.

A second important matrix links the vertices and edges of Γ. For each edge $h = \{v,w\}$ we choose one of the two vertices as the "positive end" and the other one as the "negative end" of the edge. The choice of the orientation is completely arbitrary. The matrix

$$\nabla^+_{ij} = \begin{cases} +1 & \text{vertex } v_i \quad \text{is the positive end of edge } e_j \\ -1 & \text{vertex } v_i \quad \text{is the negative end of edge } e_j \\ 0 & \text{otherwise} \end{cases}$$

is called the *incidence matrix of* Γ.

The notion of an incidence matrix of a graph should not be confused with the incidence matrices of transitive permutation groups described above — they are unrelated. The choice of the symbol ∇ is intentional. In fact, let $f : V \to \mathbb{R}$ be an arbitrary function. Then $(\nabla f)(h) = f(v) - f(w)$ where h is the edge $\{v,w\}$, and v is the positive end of the edge h. This is as close to a differential operator as one can get on a graph.

The matrix

$$\Delta = A - D$$

is called the *Laplacian* on Γ. For regular graphs this becomes $A - DI$, where D is now the constant vertex degree of Γ. For an arbitrary function $f : V \to \mathbb{R}$ we have

$$(\Delta f)(x) = \sum_{y \in N(x)} \big(f(y) - f(x)\big).$$

The graph Laplacian shares its most important properties with the familiar Laplacian operator in $\mathbb{R}^n$ as listed in the following

Theorem. The graph Laplacian Δ has the following properties:

(i) Δ is symmetric.

(ii) Δ is non-positive definite.

(iii) Δ is singular; the eigenvector $(1,\dots,1)$ belongs to the eigenvalue $\tilde{\Lambda}_0 = 0$. If Γ is connected (as we will always assume), then Λ_0 has multiplicity 1.

(iv) $\Delta = -\nabla^+\nabla$, that is, it corresponds to "second derivatives" on the graphs.

(v) for all landscapes f and g holds *Green's formula* in the following form

$$\sum_{x\in V} f(x)(\Delta g)(x) = \sum_{x\in V} g(x)(\Delta f)(x) = -\sum_{h\in E}(\nabla f)(h)(\nabla g)(h).$$

(vi) $\langle\nabla f,\nabla g\rangle = -\langle f,\Delta g\rangle = -\langle g,\Delta f\rangle$ for arbitrary $f,g : V \longrightarrow \mathbb{R}$, where $\langle\, .\, ,\, .\,\rangle$ is the usual scalar product in $\mathbb{R}^N$.

Proof. (i) is obvious, (ii) and (iii) are well known, see, e.g., [12, 26], and (iv) is Proposition 4.8 of [12]. Green's formula (v) is checked by explicit calculation, see e.g., [126]. ∎

The graph Laplacian has received considerable attention in the theory of electrical networks, see, e.g., [12, Chap.5]. A recent book on potential theory on infinite discrete lattices is [117]. Mark Kac [68] has asked whether the knowledge of the spectrum of a Laplacian with Dirichlet boundary conditions is sufficient to completely determine the shape of its domain. Shortly after Kac's talk a series of papers, e.g., [5, 41, 7] investigated the analogous question for the graph Laplacian: Given the spectrum of Δ, how much geometric information on the underlying graph can be retrieved?

The set of eigenvalues of the adjacency matrix is called the spectrum of the graph Γ. The spectral properties of the matrices A, T, and Δ are closely related. For later reference we note the following trivial

Lemma. Let Γ be a regular graph with adjacency matrix A and vertex degree D. Then A, T, and Δ have the same eigenvectors and the corresponding eigenvalues Λ_k, λ_k, and $\tilde{\Lambda}_k$ belonging to the eigenvector ϕ_k are related via

$$\tilde{\Lambda}_k = \Lambda_k - D \qquad \text{and} \qquad \lambda_k = \Lambda_k/D.$$

2.3.2. Symmetry and Regularity of Graphs

An *automorphism* of Γ is mapping $\alpha : V \longrightarrow V$, such that $(\alpha(x),\alpha(y))$ is an edge of Γ if and only if (x,y) is an edge of Γ. The set of all automorphisms forms a group, the automorphism group $\mathbf{Aut}[\Gamma]$.

A large number of different symmetry properties have been defined for graphs. These notions can be subdivided into two classes depending on whether their definition uses properties of the automorphism group. For reference we mention here the most important ones:

REGULARITY CONDITIONS

- Γ is *regular* if all vertices have the same number D of neighbors.
- Γ is *distance degree regular* if the cardinality of the sets
$$N_k(x) \stackrel{\text{def}}{=} \{y \in V \mid d(x,y) = k\}$$
depends only on k.
- Γ is *walk regular* if the number of closed walks of length r starting at vertex x is independent of x for all r, or equivalently, if $(A^r)_{xx}$ is independent of x for all r.
- Γ is *distance regular* if the cardinalities of the intersections $N_i(u) \cap N_j(v)$ depends only on the distance $d(u,v)$. The distance matrices $A^{(k)}$ with entries $a^{(k)}_{xy} = 1$ if $d(i,j) = k$ and $a^{(k)}_{xy} = 0$ otherwise form a symmetric association scheme.
- Γ is *strongly regular* if the cardinality of the intersections $N_1(u) \cap N_1(v)$ depends only on whether u and v are adjacent or not. Usually one excludes complete graphs. Strongly regular graphs are then distance regular with diameter 2. They play a prominent role in algebraic combinatorics.
- Γ is *k-set-regular* if, given a subset Y consisting of at most k vertices, the number of vertices joined to each vertex in Y depends only on the isomorphism type of the induced subgraph of Y in Γ. Consequently, 1-set-regular is regular, and 2-set-regular is strongly regular.

TRANSITIVITY CONDITIONS

- Γ is *vertex transitive* if $\mathbf{Aut}[\Gamma]$ acts transitively on V, i.e., if for any two vertices x and y there is an automorphism α such that $\alpha(x) = y$.
- Γ is *generously vertex transitive* if $\mathbf{Aut}[\Gamma]$ acts generously transitively on V, i.e., if for any two vertices x and y there is an automorphism α such that $\alpha(x) = y$ and $\alpha(y) = x$.
- Γ is *edge transitive* if $\mathbf{Aut}[\Gamma]$ acts transitively on the set E of edges (considered as unordered pairs of vertices).
- Γ is *weakly symmetric* if it is both vertex transitive and edge transitive.
- Γ is *symmetric* if for all vertices $x, y, u, v \in V$ such that (x,y) and (u,v) are edges there is an automorphism α fulfilling $\alpha(x) = u$ and $\alpha(y) = v$.
- Γ is t-arc transitive if $\mathbf{Aut}[\Gamma]$ acts transitively on "arcs" consisting of at most t subsequent edges such that two consecutive edges in the arc are not identical. Note that arcs can contain a vertex or an edge twice, but it is not allowed to backtrack a single step. 1-arc-transitive is the same as "symmetric".
- Γ is *distance transitive* if for any four vertices $x, y, x', y' \in V$ such that $d(x,y) = d(x',y')$ there is an automorphism fulfilling $\alpha(x) = x'$ and $\alpha(y) = y'$.
- Γ is *metrically k-transitive* if for any two k-tuples of vertices $(x_1, \dots, x_k)$ and $(y_1, \dots, y_k)$ which satisfy $d(x_i, x_j) = d(y_i, y_j)$ there is an automorphism fulfilling $\alpha(x_i) = y_i$. In other words, a graph is metrically k-transitive if any isometry between sets of at most k vertices can be extended to an automorphism. Metrically 1-transitive is the same as vertex-transitive, metrically 2-transitive is the same as distance-transitive.
- Γ is *k-set-transitive* if any isomorphism between induced subgraphs of sets consisting of at most k vertices extends to an automorphism of Γ. For $k \geq 2$, k-set transitive graphs are metrically transitive with diameter 2.
- Γ is *homogeneous* if it is k-set transitive for all k.
- Γ is *metrically homogeneous* if it is metrically k-transitive for all k.

The relations between the animals in this zoo of graph theoretical properties is shown in figure 2. In the following sections we will discuss the most important properties in a little more detail.

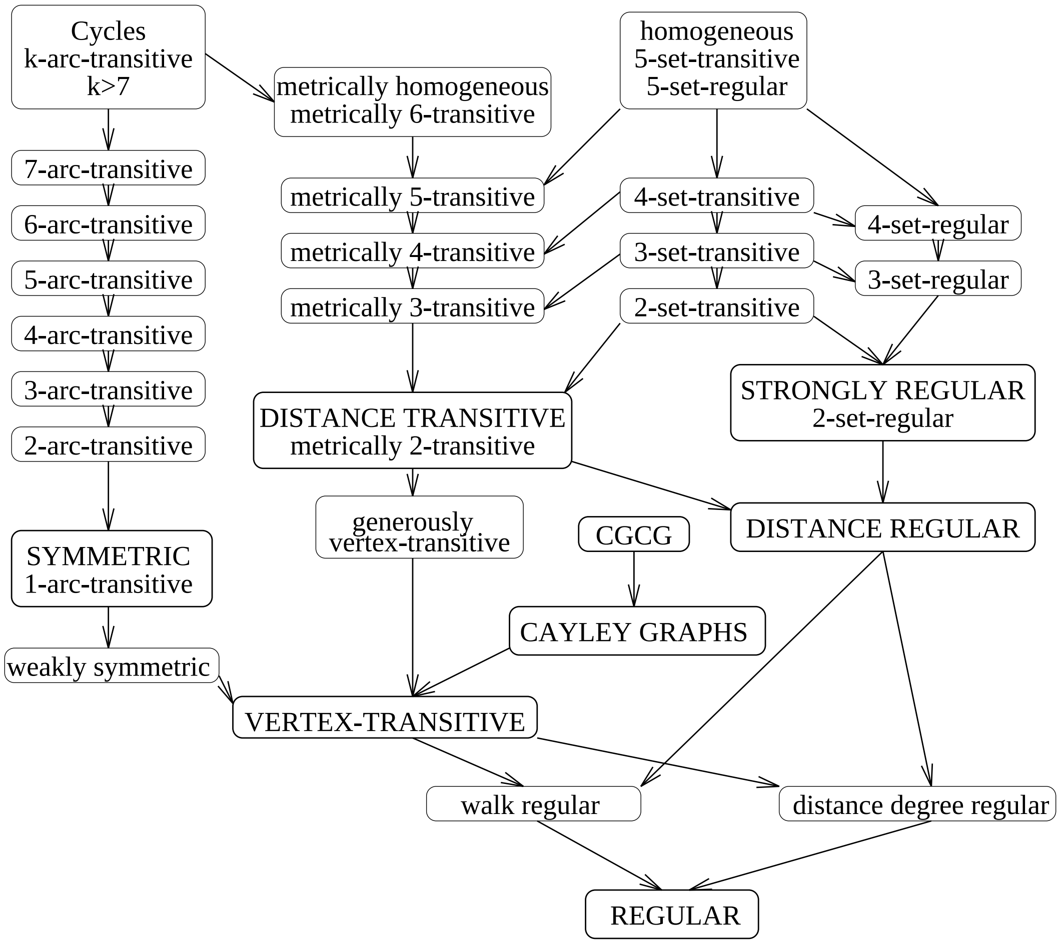


Figure 2: A zoo of symmetry properties of graphs has been studied. Arrows indicate implications; equivalent properties are shown within the same rectangle. CGCG refers to Cayley graphs with an Abelian group. Cayley graphs are discussed in section 2.3.4.

2.3.3. Vertex-Transitive Graphs

A most important class of graphs are derived from groups: Let H be a group with N elements and Φ a set of generators[3] of H such that

(i) $e \notin \Phi$, and

(ii) $x \in \Phi$ implies x^{-1} in Φ.

Then the graph $\Gamma(H, H', \Phi)$ with vertex set $V = H$ and edge set $\{(x,y)|xy^{-1} \in \Phi$ is called the Cayley *graph* of the group H with respect to the set of generators Φ.

Not all vertex transitive graphs are Cayley graphs, however. A famous counter-example is the Petersen graph, figure 2.

Sabidussi [106] gives a method for constructing all vertex transitive graphs from groups which is closely related to Cayley graphs. Let $\sim$ be an equivalence relation on the vertex set of a graph Γ. Then the graph $\Gamma/_{\sim}$ has the equivalence classes as its vertices, and two such vertices are joined by an edge if and only if there is at least one edge connecting the two classes in Γ. Now let $H' < H$

[3] Φ is a set of generators of the group H if each element $x \in G$ can be represented as a product of members of H.

be a subgroup of H, and let Φ again be a set of generators of H. The left cosets of H' induce an equivalence relation $\sim$ on H. We will denote the graph $\Gamma(H,\Phi)/_\sim$ by $\Gamma(H,H',\Phi)$. Explicitly, its vertex set and its edge set are

$$V = \{gH' \,|\, g \in H\} \quad \text{and} \quad E = \{(aH', bH') \,|\, aH' \neq bH' \wedge aH' \cap (bH'\Phi) \neq 0\}\,.$$

As an example consider the Cayley graph of the symmetric group S_n with the set of all transpositions as set of generators. The set of all rotations forms a cyclic group $\mathbb{Z}_n < S_n$. The graph $\Gamma(S_n, \mathbb{Z}_n, T)$ is a sensible choice for dealing with a traveling salesman problem, see sect. 2.1.

Theorem. (1) A graph Γ is vertex transitiv if and only if there is a group H, a subgroup $H' < H$ and a set of generators Φ fulfilling (i) and (ii) above with $\Phi \cap H' = \emptyset$ such that $\Gamma = \Gamma(H, H', \Phi)$.
(2) A graph Γ is a Cayley graph if and only if its automorphism group contains a subgroup G which acts regularly on the vertex set. Then $\Gamma = \Gamma(G,\Phi)$ with an appropriate set of generators.

Proof. The first statement is due to Sabidussi [106], the second one is proved in [11, 12]. ∎

Lateron we will need a few properties of simple random walks on vertex transitive graphs. If Γ is vertex transitive, let us define the probability $\varphi_{s\mu}$ that a simple random walk of s steps ends in a vertex contained in the symmetry class μ.

An explicit expression for $\varphi_{s\mu}$ is given in [123], see sect. 4.1.2. The relation of $\varphi_{s\mu}$ and the powers of the adjacency matrix is also of interest: for vertex transitive graphs one finds [126]

$$A^s = \sum_\mu \frac{D^s}{|\mu|}\varphi_{s\mu} R^{(\mu)} \stackrel{\text{def}}{=\!=} \sum_\mu \Theta_{s\mu} R^{(\mu)}.$$

2.3.4. Cayley Graphs

It is easy to see that a left multiplication of $x \in G$ by an arbitrary $g \in G$, i.e., the permutation action $x \mapsto gx$, is an automorphism of $\Gamma(G,\Phi)$, i.e., the group of left multiplications $\tilde{G}$ is a subgroup of $\mathbf{Aut}[\Gamma(G,\Phi)]$. By construction $\tilde{G}$ acts transitively on G. Consequently, (yz, y) belongs to the same orbit of $\tilde{G}$, and hence also to the same orbit of $\mathbf{Aut}[\Gamma(G,\Phi)]$, as (z, e), where e is the group identity which we choose as reference vertex in the Cayley graph. Setting $x = yz$ this implies

$$xy^{-1} = z \implies (x,y) \in \ddot{\mu}_z$$

where μ_z denotes the symmetry class to which z belongs. Now consider the $N \times N$ matrices defined by

$$G^{(z)}_{xy} = \begin{cases} 1 & \text{if} \quad xy^{-1} = z \\ 0 & \quad \text{otherwise} \end{cases}$$

Lemma. The matrices $G^{(z)}$ form an incidence scheme $\mathfrak{G}$ with the following properties:
(i) $G^{(e)} = I$.
(ii) $\sum_{z\in G} G^{(z)} = J$.
(iii) $G^{(r)}G^{(s)} = G^{(rs)}$.
(iv) $G^{(z^{-1})} = (G^{(z)})^+$.

Proof.(i) and (ii) are obvious. Properties (iii) and (iv) are checked by direct computation. ∎

The algebra spanned by $\mathfrak{G}$ is (isomorphic to) the *group algebra* of G. In particular, if G is a commutative group, then $\mathfrak{G}$ is an association scheme, and $\langle\mathfrak{G}\rangle$ is a commutative algebra.

The matrices $R^{(\mu)}$ are obtained from the matrix representation of the group algebra via

$$R^{(\mu)} = \sum_{z\in\dot\mu} G^{(z)} \qquad \text{and} \qquad \tilde R^{(\mu)} = \sum_{z\in(\dot\mu\cup\dot\mu^{+})} G^{(z)}.$$

An immediate consequence of this representation is the following

Lemma. Let Γ be a Cayley graph of a commutative group G. Then the symmetrized incidence matrices $\tilde R^{(\mu)}$ form a symmetric association scheme.

Proof. It suffices to observe that the matrices $R^{(\mu)}$ commute. The lemma then follows immediately from the lemma in sect. 2.2.5. ∎

The (Cartesian) product $\Gamma_1\times\Gamma_2$ of two graphs has vertex set $V(\Gamma_1\times\Gamma_2) = V(\Gamma_1)\times V(\Gamma_2)$. Two vertices (x_1,x_2) and (y_1,y_2) are connected by an edge if either (i) $x_1 = y_1$ and x_2, y_2 are adjacent in Γ_2, or (ii) $x_2 = y_2$ and x_1, y_1 are adjacent in Γ_1. It is easily checked that the product of two Cayley graphs $(G_1,\Phi_1)\times(G_2,\Phi_2)$ is the Cayley graph $(G_1\times G_2,\Phi)$, where $G_1\times G_2$ is the direct product of the two groups and $\Phi = (\Phi_1\times\{e_2\})\cup(\Phi_2\times\{e_1\})$, e_1 and e_2 being the group identities.

2.3.5. Distance Regular Graphs

A connected graph Γ is *distance regular* if for any two vertices u and v, and any two integers k and l, the numbers

$$s_{kl}^{(d)} \overset{\text{def}}{=\!=} |N_k(u)\cap N_l(v)|$$

depend only on $d \overset{\text{def}}{=\!=} d(u,v)$, that is, on the distance of the vertices u and v. The numbers $s_{kl}^{(d)}$ are known as the intersection numbers of Γ. Let x by an arbitrary reference vertex. Then

$$\varpi = (\,\{x\}, N(x), N_2(x), \dots, N_M(x)\,)$$

is called the *distance partition* of Γ centered at x, where $M = \operatorname{diam}\Gamma$, the diameter of the graph. Distance regularity strongly restricts both the algebraic and the geometric properties of a graph. The most important features are summarized below.

Lemma. Let Γ be distance regular. Then the distance matrices $A^{(d)}$ form a symmetric association scheme, and $A = A^{(1)}$ has exactly $M+1$ distinct eigenvalues.

Proof. See, e.g., [54, Chap.11]. ∎

Lemma. The sequence $k_r \overset{\text{def}}{=\!=} |N_r(x)|$ is unimodal and $k_r \le k_{M-r}$ for all $r\le M/2$. Thus the average distance $\bar d$ of two vertices is at least $\operatorname{diam}\Gamma/2$.

Proof. See, e.g., [54, Chap.11]. ∎

Distance transitivity implies distance regularity. The converse is not true. A counterexample can be found in [53], see also [12, 20c].

2.4. Fourier Series on Graphs

A series expansion in terms of a complete and orthonormal system of eigenfunctions of the Laplace operator is commonly termed *Fourier expansion*. We will adopt the same terminology here following Weinberger's paper [141]. Thus let $f : V \to \mathbb{R}$ be a real valued function on the vertex set of Γ and let $\{\theta_s\}$ denote a complete orthonormal set of eigenvectors of the graph Laplacian Δ. Then we call

$$f(x) = \sum_{y\in\Gamma} b_y \theta_y(x)$$

the *Fourier expansion* of f.

It will turn out to be convenient to label the eigenvectors θ_y by the vertices of the underlying graph Γ. This is possible because the eigensystem of the finite symmetric operator Δ is complete. In general, this labeling is arbitrary.

For Cayley graphs of commutative groups (CGCGs), however, the eigenvectors and eigenvalues have a particularly simple form. Let $x = (x_1, \dots, x_m)$ denote the component wise representation of the group element x which is obtained from the decompositions of the commutative group G into a direct product of m cyclic groups of orders N_k. The group action is represented by component-wise addition modulo N_k:

$$x \circ y = (x_1 + y_1 \bmod N_1, x_2 + y_2 \bmod N_2, \dots, x_m + y_m \bmod N_m).$$

The Fourier basis of a CGCG, which coincides with a complete set of eigenvectors since Cayley graphs are regular, are of the form [79]

$$e_g(x) = \exp\left(2\pi i \sum_k \frac{x_k g_k}{N_k}\right) ,$$

and the corresponding eigenvalues are given by

$$\Lambda_g = \sum_{x\in\Phi} e_g(x) \qquad \text{and} \qquad \tilde{\Lambda}_g = \sum_{x\in\Phi} (e_g(x) - 1).$$

Note that in this special case we have a canonical labelling of the eigenvectors. For convenience we recall a few basic properties of the functions $e_x(y)$.

(i) $e_g(x) = e_x(g)$.

(ii) $e_g(x \circ y) = e_g(x)e_g(y)$ and $e_{g\circ h}(x) = e_g(x)e_h(x)$.

(iii) $e_g(x)e^*_g(x) = 1$, where the asterisk denotes complex conjugation.

(iv) $\sum_{x\in G} e_g(x)e^*_h(x) = N\delta_{gh}$ and $\sum_{l\in G} e_l(x)e^*_l(y) = N\delta_{xy}$, where N is the total number of configurations.

These functions have the well-known form of the Fourier basis on $\mathbb{C}^n$. They are, of course, the characters of the group commutative G. Except for property (ii), which will be crucial for some results in the following, the above statements are immediate consequences of Schur's Lemma, and are therefore true for the characters of arbitrary groups, see, e.g., [113].

Since Δ is symmetric, the Fourier basis $\{\theta_y\}$ can always be chosen real valued; it is convenient, however, to allow for complex conjugate pairs of eigenvectors. In particular, we will use the following identity for functions on a CGCG:

$$f(x) = \sum_{l\in G} a_l e_l(x) = \sum_{l\in G} a_l^* e_l^*(x).$$

Since we deal with a finite vector space with a scalar product (for which we will use the notation $\langle\, .\, ,\, .\, \rangle$ in this paper), spanned by eigenvectors $\{\theta_y\}$ of the graph Laplacian, the familiar properties of Fourier series, such as *Parseval's equation*

$$||f||^2 = \langle f, f\rangle = \sum_{y\in\Gamma} \left(\frac{\langle f, \theta_y\rangle}{\langle \theta_y, \theta_y\rangle} \right)^2 ,$$

and the *mean square approximation theorem hold* for all landscapes on all connected graphs. For the convenience of the reader we recall

Proposition. [*Mean Square Approximation Theorem*]
Consider a landscape f on Γ with Fourier expansion

$$f(x) = \sum_{y\in V} a_y \theta_y(x)$$

Let X be a subset of V, let

$$g(x) = \sum_{y\in X} b_y \theta_y(x)$$

be an approximation to f. Then the squared approximation error $||f-g||^2 = \langle (f-g), (f-g)\rangle$ is minimized by choosing $b_y = a_y \equiv \langle f, \theta_y\rangle$ for all $y \in X$.

2.5. Spectral Properties of the Adjacency Matrix

2.5.1. Equitable Partitions

Definition. Let $\varpi = (\varpi_0, \varpi_1, \ldots, \varpi_M)$ be a partition of the vertex set V into the *cells* ϖ_i, $i = 0, \ldots, M$. ϖ is called *equitable* if the number of neighbors which a vertex in ν has in cell μ is independent of the choice of the vertex in ν. In other words, ϖ is equitable if for all $\mu, \nu \in \varpi$ holds

$$\breve{A}_{\mu\nu} \stackrel{\text{def}}{=} |N(y) \cap \mu| = \sum_{x\in\mu} A_{xy} \qquad \text{for all } y \in \nu.$$

We call $\breve{A}$ a *combinatorially collapsed adjacency matrix.*

There is an $(M+1)\times N$ matrix associated with each partition ϖ of V into $M+1$ cells. This matrix will also be denote by ϖ:

$$\varpi_{\mu x} \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } x \in \mu \\ 0 & \text{otherwise} \end{cases}$$

We remark that this definition is the transpose of the convention in Godsil's book [54], while it conforms the notation used by Bollobás [17]. Equitable partitions have been introduced by Schwenk [111]; more recently they have been used by Powers and coworkers as *colorations*, see, e.g., [101, 100], see also [27, Chap.4].

Theorem. Let ϖ an equitable partition of the vertex set V of a graph Γ. Then the following statements hold.

(i) $\varpi A = \breve{A}\varpi$.

(ii) If e is an eigenvector of A with eigenvalue Λ, then $\upsilon = \varpi e$ is a *right* eigenvector of $\breve{A}$ with the same eigenvalue provided $\upsilon \neq 0$. For later reference we note the explicit formula

$$\upsilon(\mu) = \sum_{x\in\mu} e(x)$$

(iii) If υ is a right eigenvector of $\breve{A}$, then

$$\rho(\mu) = \frac{1}{|\mu|}\upsilon(\mu)$$

is a left eigenvector of $\breve{A}$.

(iv) If ρ is a left eigenvector of $\breve{A}$, then $\rho\varpi$ is an eigenvector of A which is constant on all cells.

(v) $|\mu|(\breve{A}^r)_{\mu\nu} = |\nu|(\breve{A}^r)_{\nu\mu}$.

(vi) The characteristical polynomial of $\breve{A}$ divides the characteristical polynomial of A.

(vii) If Γ is connected then the A and $\breve{A}$ have the same spectral radius.

Proof. See [54, chap.5]. ∎

There is an intimate relation between the orbits of a group of automorphisms acting on V and the equitable partitions of a graph.

Lemma. Let $G \le \mathrm{Aut}[\Gamma]$ be an arbitrary group of automorphisms of Γ. Then the orbits of G form an equitable partition ϖ_G of V.

Proof. See [54, p.76]. ∎

The converse is not true. There are equitable partitions which are not induced by any automorphism group. For an example see [54, p.76]. One says that a partition ϖ' is a *refinement* of ϖ, in symbols $\varpi' \le \varpi$, if each cell of ϖ' is contained in some cell of ϖ. The equitable partitions form a lattice with respect to refinement [54, p.87].

2.5.2. Equitable Partitions with a Reference Vertex

The results (vi) and (vii) on equitable partitions can be strenghtened if ϖ fulfils a single, simple additional condition. An equitable partition ϖ which has a cell $\varpi_0 = \{u\}$ consisting of the single vertex u will be called *equitable partition with reference vertex.*

Theorem. Let ϖ be an equitable partition with reference vertex u. Then A and $\breve{A}$ have the same minimal polynomial, and Λ is an eigenvalue of A if and only if it is an eigenvalue of $\breve{A}$.

Proof. See [17, Thm.8.6]. ∎

As an example we note that the distance partitions of a distance regular graph as well as the orbits of an arbitrary stabilizer subgroup $G_x < \mathrm{Aut}[\Gamma]$ of a vertex transitive graph form an equitable partition of V with a reference vertex.

Lemma. Let ϖ be an equitable partition of a graph Γ with reference vertex 0. Then there is always a basis $\{\rho_i\}$ of left eigenvectors of the corresponding combinatorially collapsed adjacency matrix $\breve{A}$ with the following properties:

(i) The eigenvectors are normalized such that $\rho(0) = \upsilon(0) = 1$.

(ii) The left eigenvectors are orthogonal w.r.t. the scalar product $\langle a; b\rangle = \sum_{\mu} a(\mu)b(\mu)|\mu|$.

(iii) This basis is unique if all eigenvalues of $\breve{A}$ are simple.

Proof. The corresponding result in [123] is formulated for the collapsed adjacency matrix $\hat{A}$ of a vertex transitive graph, but only the properties of equitable partitions with a reference vertex are used in the course of the proof. ∎

A few remarks on distance regular graphs are in order here.

Lemma. Let Γ be distance regular and let ϖ be a distance partition with respect to an arbitrary reference vertex x. Then $\breve{A}^{(d)}_{kl} = s^{(d)}_{kl}$.

Proof. This follows immediately from the discussion in [54, Chap.11]. ∎

The lemma states that we recover the intersection numbers of a distance regular graph as the combinatorially collapsed distance matrices. This is analogous to the relation of the incidence matrices and their intersection numbers in Higman's theory, although there are no group actions involved in the case of distance regular graphs.

2.5.3. The Adjacency Algebra

The *adjacency algebra* of a graph Γ is the matrix algebra spanned by the powers of the adjacency matrix. It will be denoted by $\mathfrak{A}[\Gamma]$. Of course, this algebra is commutative. It is clear from the previous section that the properties of $\mathfrak{A}[\Gamma]$ are closely linked to the equitable partitions with reference vertices, since its dimension coincides with the degree of the minimal polynomial of A or $\breve{A}$. In particular, therefore, the dimension of $\mathfrak{A}[\Gamma]$ is bounded above by 1 plus the number of cells in an equitable partition with reference vertex. On the other hand we have the following lower bound on the dimension of the adjacency algebra:

Theorem. Let Γ be a connected graph with diameter $\mathrm{diam}\Gamma$. Then the dimension of the adjacency algebra $\mathfrak{A}[\Gamma]$ is at least $\mathrm{diam}\Gamma + 1$.

Proof. See [12, Prop.2.6]. ∎

This result suggests that equitable partitions with a reference vertex x are always refinements of the corresponding distance partition centered at the reference vertex x. We do not know whether this is actually true. Another important question is whether there is always an equitable partition with reference vertex such that all eigenvalues of $\breve{A}$ are simple. The above discussion, along with the striking similarity of the algebraic properties of distance regular graphs and the theory of incidence structures seems to suggest that it might even be possible to base the investigation of configuration spaces entirely on equitable partitions with reference vertices instead of graph automorphisms.

Here we will return, however, to the safer grounds of vertex transitive graphs, for which Higman's beautiful theory of intersection numbers is applicable, since the automorphism group of a vertex transitive graph acts by definition transitively on V. The incidence scheme $\mathfrak{R}[\Gamma]$ associated with a vertex transitive graph Γ is of course the one of $\mathbf{Aut}[\Gamma]$ acting on $V \times V$, see sect. 2.2. The algebra spanned by $\mathfrak{R}[\Gamma]$ is usually referred to as the *commuting algebra of the graph*, denoted by $\mathfrak{C}[\Gamma]$. Since the adjacency matrix A is a linear combination of the matrices $R^{(\mu)}$ we have necessarily $\mathfrak{A}[\Gamma] \subseteq \tilde{\mathfrak{R}}[\Gamma]$.

Theorem. Let Γ be vertex transitive. Then the following statements are equivalent:
(i) $\mathfrak{A}[\Gamma] = \mathfrak{C}[\Gamma]$.
(ii) All eigenvalues of the collapsed adjacency $\hat{A}$ matrix are simple.
(iii) The powers of A span $\mathfrak{C}[\Gamma]$, i.e., all matrices $R^{(\mu)}$ are polynomials in A.
(iv) The powers of $\hat{A}$ span $\hat{\mathfrak{C}}[\Gamma]$.

Proof. (i) $\Longleftrightarrow$ (ii) $\Longleftrightarrow$ (iv) follows immediately from [59, (4.12)].
(i) $\Longleftrightarrow$ (iii) The matrices $R^{(\mu)}$ form the basis of a vector space of dimension $M + 1$. We have seen in section 2.3.3 that

$$A^s = \sum_{\mu} \Theta_{s\mu} R^{(\mu)} \qquad \text{with} \qquad \Theta_{s\mu} = \frac{D^s \varphi_{s\mu}}{|\mu|}.$$

The coefficients $\Theta_{s\mu}$ form a $(M+1) \times (M+1)$ matrix which can be invertible only if the number of distinct eigenvectors of A equals the number of symmetry classes, i.e., if all eigenvalues of the collapsed adjacency matrix are simple. Since $\mathfrak{A}[\Gamma] \subseteq \mathfrak{C}[\Gamma]$ the converse is also true. Hence Θ is invertible if and only if all eigenvalues of $\hat{A}$ are simple. ∎

Property (iii) has been termend " Θ-property" in [126]. It seems to be of particular importance for the theory of landscapes and random fields. It is not known (to the author) whether the Θ-property is related to any of the well-studied graph theoretical properties of configuration spaces. Not all vertex transitive graphs fulfil (iii); An example for a vertex transitive graph without the Θ-property is the permutohedron, i.e., the Cayley $\Gamma(S_n, \mathcal{K})$ of the symmetric group with the set of canonical transpositions $(i, i+1)$ as generators.

2.6. Examples of Configuration Spaces

2.6.1. Sequence Spaces

Consider the set of all sequences of length n, which are composed of letters taken from some finite alphabet of size α. The canonical distance is given by the number of positions in which two sequences x and y differ; this metric distance measure is the Hamming distance [58]. The graph obtained from connecting two sequences if and only if their Hamming distance is $d(x,y)=1$ is called the *Hamming graph* or *sequence space* $\mathcal{Q}^n_\alpha$. The sequence space $\mathcal{Q}^n_\alpha$ is the n-fold product of the complete graph with α vertices with itself. A complete graph with α vertices can be regarded as the Cayley graph graph of an arbitrary group G of order α with $\Phi = G\setminus\{e\}$. Thus Hamming graphs are Cayley graphs of commutative groups, and they are distance transitive, see e.g., [12].

Hamming graphs constitute the most common class of configuration spaces. Their automorphism group is well known [36]

$$\mathbf{Aut}[\mathcal{Q}^n_\alpha] = S_\alpha \wr S_n = S_n \ltimes (S_\alpha)^n \,,$$

where S_m denotes the symmetric group on m elements.

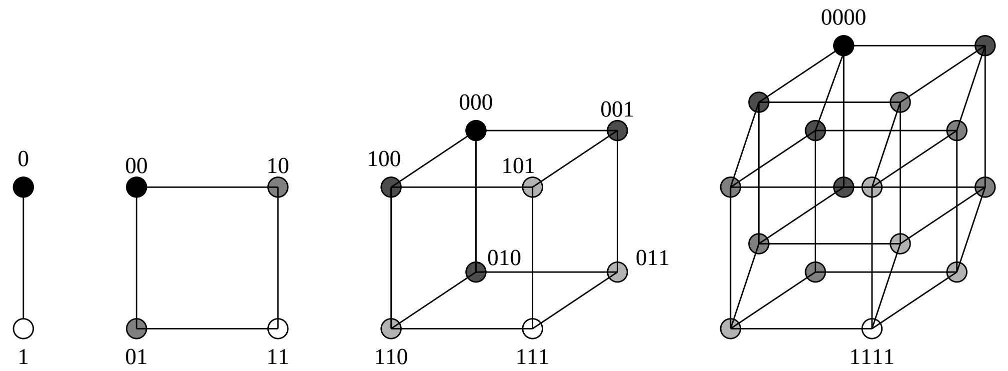


Figure 3: Hypercubes $\mathcal{Q}^n_2$ for $n=1$ through $n=4$. Vertices belonging to the same symmetry class μ are marked by a common symbol. Vertices within a symmetry class have common distance from the reference vertex $\bullet$. The hypercubes are distance transitive, since their symmetry classes are defined by the distance from the reference vertex.

The following results on the eigenvalues and eigenvectors of the sequence spaces are well known, see, e.g., [21, 37, 73, 75, 84, 105, 136].

The eigenvalues and left eigenvectors of the collapsed adjacency matrix $\hat{A}$ of a sequence space $\mathcal{Q}^n_\alpha$ are given by

$$\Lambda_p = n(\alpha-1) - p\alpha,$$
$$\rho_p(d) = \frac{1}{(\alpha-1)^p\binom{n}{p}}\mathbf{K}^{(\alpha,n)}_p(d),$$

where $\mathbf{K}^{(\alpha,n)}_p$ are the *Krawtchouk polynomials*

$$\mathbf{K}^{(\alpha,n)}_p(d) = \sum_{j=0}^{p}(-1)^j\binom{d}{j}\binom{n-d}{p-j}(\alpha-1)^{p-j}.$$

The multiplicity of Λ_p w.r.t. the full adjacency matrix A is $m(\Lambda_p) = (\alpha-1)^p\binom{n}{p}$; all eigenvalues are simple w.r.t. the collapsed adjacency matrix $\hat{A}$. The left eigenvectors ρ_p of $\hat{A}$ fulfill the orthogonality relation

$$\langle \rho_k ; \rho_l \rangle = \frac{\kappa^n}{(\kappa-1)^k \binom{n}{k}}\, \delta_{kl}$$

with respect to the scalar product

$$\langle a ; b \rangle = \sum_{d=0}^{n} (\kappa-1)^d \binom{n}{d}\, a(d)\, b(d),$$

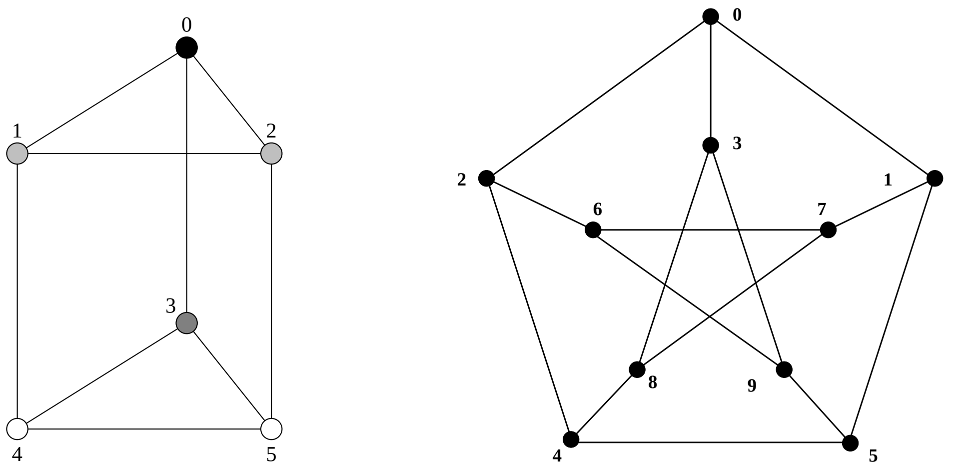
and the symmetry relation $\rho_p(d) = \rho_d(p)$.

2.6.2. Generalized Hamming Graphs

The name *Generalized Hamming Graph* is sometimes used for direct products of sequence spaces. They correspond to sequences where the alphabet is not the same for all positions. Such graphs have turned out to be useful when considering the inverse folding problem for RNA secondary structures [104, 103]: An RNA secondary structure can be partioned into the unpaired regions where each position can be realized by one of the 4 nucleotides **G**, **C**, **A**, and **U**, and the paired regions, where each base pair can be realized by one of the 6 types of base pairs (**GC**, **CG**, **AU**, **UA**, **GU**, and **GU**). The set of sequences compatible with the given structure can then be considered as the graph

$$\mathcal{C} \cong \mathcal{Q}_4^{n_u} \times \mathcal{Q}_6^{n_p},$$

where n_u and n_p are the numbers of the unpaired positions and of the base pairs, respectively.



a b

Figure 4: **a)** The trigonal prism graph is the smallest vertex transitive graph which is not edge transitive. Its vertex set decomposes into four symmetry classes which are the reference vertex $\mathbf{0} = \{0\}$ itself, its neighbors within the same triangle $\mathbf{1a} = \{1, 2\}$, its neighbor in the other triangle $\mathbf{1b} = \{3\}$, and the remaining vertices $\mathbf{2} = \{4, 5\}$ which have distance two from the reference vertex.
b) The Petersen graph is the smallest distance transitive graph which is not a Cayley graph of any group. It has three classes of vertices: the reference vertex $\mathbf{0} = \{0\}$, its neighbors $\mathbf{1} = \{1, 2, 3\}$, and all other vertices $\mathbf{2} = \{4, 5, 6, 7, 8, 9\}$.

The simplest case of a Generalized Hamming Graph which is not a sequence space is the trigonal prism $\mathcal{Q}_2 \times \mathcal{Q}_3$, see figure 4a. Its collapsed adjacency matrix is

$$\hat{A} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 2 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 2 & 1 \end{pmatrix}$$

The eigenvalues of $\hat{A}$, the corresponding *left* eigenvectors, and the multiplicities of these eigenvalues in A are

$$\begin{array}{llll} \Lambda_0 = & 3 & \rho_0 = (1,1,1,1) & m(\Lambda_0) = 1 \\ \Lambda_1 = & 1 & \rho_1 = (1,1,-1,-1) & m(\Lambda_1) = 1 \\ \Lambda_2 = & 0 & \rho_2 = (1,-\frac{1}{2},1,-\frac{1}{2}) & m(\Lambda_2) = 2 \\ \Lambda_3 = & -2 & \rho_3 = (1,-\frac{1}{2},-1,\frac{1}{2}) & m(\Lambda_3) = 2 \end{array}$$

2.6.3. Johnson Graphs

The vertices of the Johnson graph $\mathcal{J}(n,k)$ are the subsets consisting of exactly k elements from a set with n elements. Two vertices are adjacent if the corresponding subsets have $k-1$ exactly elements in common. The Johnson graphs are distance transitive, see, e.g., [54]. The graph $\mathcal{J}(n,n/2)$ for even n is the natural configuration spaces for bipartitioning problems, see sect. 4.4.2. [23, 38, 54].

2.6.4. Odd Graphs

The odd graphs are relatives of the Johnson graphs. The vertex set of $O(n)$ is the set of all $(n-1)$-subsets of a $(2n-1)$ set. Two vertices x and y in $O(n)$ are joined by an edge if the subsets corresponding to x and y are disjoint. $O(1)$ is a single vertex and $O(2)$ is the triangle graph.

The Petersen Graph $O(3)$, Figure 4b, is the smallest graph which is distance transitive but not a Cayley graph [13]. The symmetry classes are the three distance classes $\mathbf{0} = \{0\}$, $\mathbf{1} = \{1,2,3\}$, and $\mathbf{2} = \{4,5,6,7,8,9\}$, with $|\mathbf{0}| = 1$, $|\mathbf{1}| = 3$, and $|\mathbf{2}| = 6$. The collapsed adjacency matrix is

$$\hat{A} = \begin{pmatrix} 0 & 1 & 0 \\ 3 & 0 & 1 \\ 0 & 2 & 2 \end{pmatrix}$$

The eigenvalues of $\hat{A}$, the corresponding *left* eigenvectors, and the multiplicities of these eigenvalues in A are

$$\begin{array}{llll} \Lambda_0 = & 3 & \rho_0 = (1,1,1) & m(\Lambda_0) = 1 \\ \Lambda_2 = & 1 & \rho_1 = (1,\frac{1}{3},-\frac{1}{3}) & m(\Lambda_1) = 5 \\ \Lambda_3 = & -2 & \rho_2 = (1,-\frac{2}{3},\frac{1}{6}) & m(\Lambda_2) = 4 \end{array}$$

Note that the multiplicities $m(\Lambda_i)$ of the eigenvalues of A are 1, 5, and 4, while the the symmetry classes contain 1, 3, 6 vertices, respectively.

The odd graphs $O(n)$ are distance transitive, with degree n and diameter $n-1$; its complete automorphism group of is

$$\mathbf{Aut}[O(n)] \cong S_{2n-1},$$

see [12, 17d]. The spectrum of is also known for all n

$$\Lambda_k = (-1)^k(n-k) \qquad \text{with} \quad m(\Lambda_k) = \binom{2n-1}{k} - \binom{2n-1}{k-1}$$

where $0 \le k \le n-1$ indexes the distinct eigenvalues.

2.6.5. Cayley Graphs of the Symmetric Group

Let S_n denote, as usual, the symmetric group on n elements, and let $\mathcal{T}$ be the set of all transpositions (i,j). It is well known that the transpositions generate the symmetric group. The corresponding Cayley-graph $\Gamma(S_n,\mathcal{T})$ is hence a possible choice for the configuration space of a n-city travelling salesman problem. Little is known in general on these graphs. We will use them here in order to explain the definitions in the above sections on a non-trivial example. $\Gamma(S_n,\mathcal{T})$ has $N=n!$ vertices and is regular with degree $n(n-1)/2$. Furthermore they are bipartite since transpositions change the sign of a permutation.

The cases $n=1$ and $n=2$ are trivial: $\Gamma(S_1,\mathcal{T})$ is an isolated vertex, and $\Gamma(S_2,\mathcal{T})$ is the connected graph with $N=2$ vertices.

The first case of interest is $\Gamma(S_3,\mathcal{T})$, see figure 5. It has $N=6$ vertices, and its diameter is 2. It coincides with the *complete bipartite graph* $K_{3,3}$, which is known to be distance transitive. Therefore the symmetry classes are the distance classes. $K_{3,3}$ may serve as an example for the fact that Cayley graphs of non-commutative groups can coincide with Cayley graphs of commutative groups. Consider the commutative group with $N=6$ elements, namely $C_2\times C_3$, with the notation $C_2=\{e,a\}$ and $C_3=\{e,b,\bar{b}\}$. It is easy to check that $\Phi=\{a,ab,a\bar{b}\}$ generates $C_2\times C_3$, and that

$$\Gamma(C_2\times C_3,\Phi)=K_{3,3}=\Phi(S_3,\mathcal{T}).$$

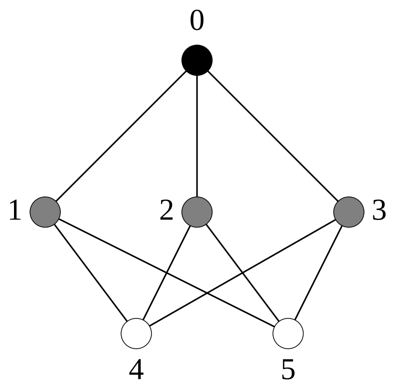


Figure 5: The Cayley graph $\Gamma(S_3,\mathcal{T})$ coincides with the distance transitive graph $K_{3,3}$, which is also obtained as $\Gamma(C_2\times C_3,\{a,ab,a\bar{b}\})$, where $C_2=\{e,a\}$ and $C_3=\{e,b,\bar{b}\}$ are the cyclic groups with 2 and 3 elements, respectively.

We emphasize at this point that the results presented in this paper depend only on the structure of the configuration space and not on the particular group representation used to construct this graph. For instance we may regard a landscape on $\Gamma(S_3,\mathcal{T})$ as a landscape on $\Gamma(C_2\times C_3,\Phi)$, i.e., as a landscape on a Cgcg, even though the physical model might force us to use the symmetric group.

The adjacency matrix and the collapsed adjacency matrix of $\Gamma(S_3,\mathcal{T})$ are

$$A=\begin{pmatrix}0&1&1&1&0&0\\1&0&0&0&1&1\\1&0&0&0&1&1\\1&0&0&0&1&1\\0&1&1&1&0&0\\0&1&1&1&0&0\end{pmatrix}\qquad\qquad \hat{A}=\begin{pmatrix}0&1&0\\3&0&3\\2&0&2\end{pmatrix}$$

The eigenvalues, the corresponding left eigenvectors of $\hat{A}$, and the multiplicities of the eigenvalues in A are

$$\begin{array}{llll} \Lambda_0 = & 3 & \rho_0 = & (1,1,1) & m(\Lambda_0)=1 \\ \Lambda_1 = & 0 & \rho_1 = & (1,0,-\frac{1}{2}) & m(\Lambda_1)=4 \\ \Lambda_2 = & -3 & \rho_2 = & (1,-1,1) & m(\Lambda_2)=1 \end{array}$$

The graph $\Gamma(S_4,\mathcal{T})$ has $N=24$ vertices; it is shown in figure 6, and its adjacency matrix is given in table 2.

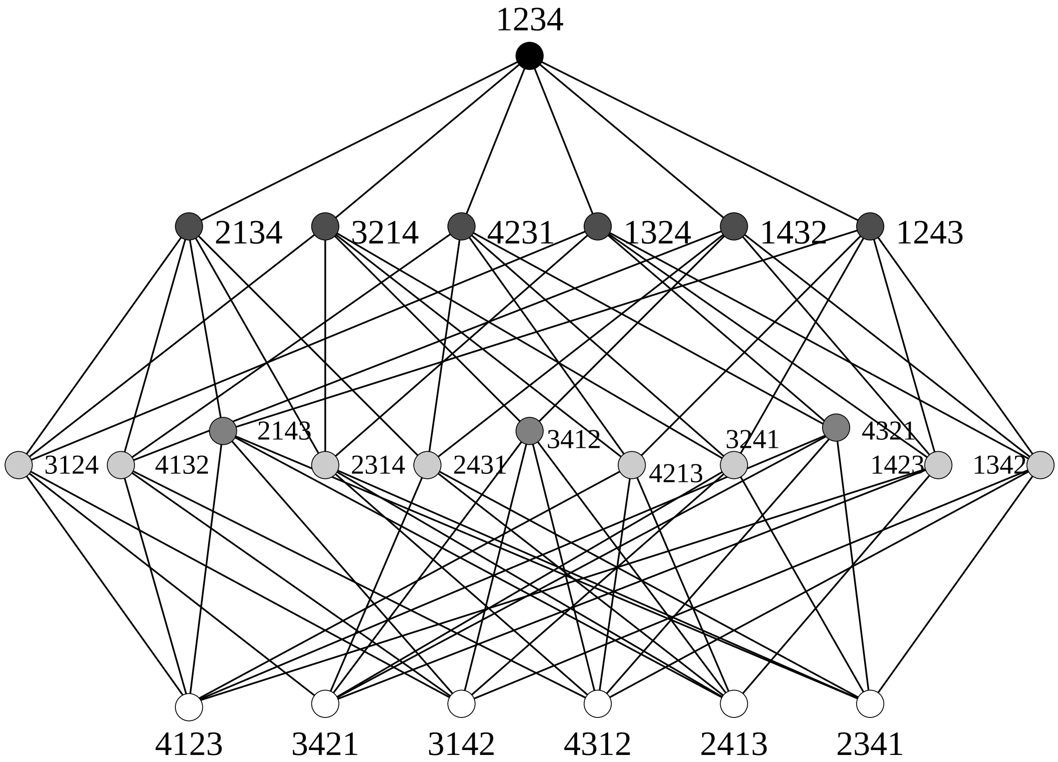


Figure 6: The Cayley graph $\Gamma(S_4,\mathcal{T})$. Symmetry classes with respect to the reference vertex (1234) are indicated by different shadings.

A close inspection of the graph shows that there are 5 symmetry classes. The collapsed adjacency matrix $\hat{A}$ of $\Gamma(S_4,\mathcal{T})$ is

$$\hat{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 6 & 0 & 2 & 3 & 0 \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 4 & 0 & 0 & 4 \\ 0 & 0 & 4 & 3 & 0 \end{pmatrix},$$

which has the following eigenvalues and left eigenvectors:

$$\begin{array}{llll} \Lambda_0 = & 6 & \rho_0 = & (1,1,1,1,1) \\ \Lambda_1 = & 2 & \rho_1 = & (1,\frac{1}{3},-\frac{1}{3},0,-\frac{1}{3}) \\ \Lambda_2 = & 0 & \rho_2 = & (1,0,1,-\frac{1}{2},0) \\ \Lambda_3 = & -2 & \rho_3 = & (1,-\frac{1}{3},-\frac{1}{3},0,\frac{1}{3}) \\ \Lambda_4 = & -6 & \rho_4 = & (1,-1,1,1,-1) \end{array}$$

The eigenvalues of $\Gamma(S_n, \mathcal{T})$ can be obtained explicitly for all n. Let $\eta = [k_1^{l_1} k_2^{l_2} \dots k_r^{l_r}] = (\eta_1, \eta_2, \dots, \eta_m)$ be a partition of n, i.e., $\sum_{i=1}^{r} l_i k_i = \sum_{j=1}^{m} \eta_j = m$, with $k_i > k_j$ and $\eta_i \geq \eta_j$ whenever $i < j$. It is well known that the conjugacy classes of the S_n correspond to the partitions of n via the cycle-decomposition of the permutations (see, e.g., Jacobson 1985, p. 48). The conjugacy classes of S_n are contained in the symmetry classes of the graph $\Gamma(S_n, \mathcal{T})$, see [126, Lem.2]. The eigenvalues of $\Gamma(S_n, \mathcal{T})$ have been obtained from the characters of the symmetric group. One finds

$$\Lambda_\eta = \frac{1}{2} \left[n + \sum_{j=1}^{m} \eta_j (1 - 2j\eta_j) \right] \,,$$

see, e.g., [34, 67, 80].

2.6.6. A Concluding Remark

In the previous sections we have outlined a broad spectrum of tools for an analysis of the graph theoretic and algebraic properties of graphs without giving "the result". The reason for this is twofold: (i) The study of the structure of the configuration spaces is interesting in itself and provides the necessary basis for a wide range of models beyond the theory of landscapes, see for instance [104, 103]. (ii) In the investigations reported in the subsequent chapters of this contribution we will frequently require certain algebraic properties of the matrices $R^{(\mu)}$, for instance, we will need for some proofs that the $\tilde{R}^{(\mu)}$ form an association scheme, and for other results we will have to require that all eigenvalues of the collapsed adjacency matrix are simple. Let me just assure the reader that all of the theory of landscapes discussed below works just fine on the most important type of configuration spaces, the Hamming graphs. On the other hand, most of it is still under construction for the more irregular classes of graphs, in particular for the Cayley graphs of the symmetric group.

3. Random Fields

3.1. Preliminaries

3.1.1. Definition

Many of the model landscapes mentioned in the introduction contain a stochastic element: a particular instance is generated by assigning a usually large number of parameters at random. This procedure suggests the following

Definition. The set $\{f : \mathcal{C} \to \mathbb{R}\}$ together with a measure $\boldsymbol{\mu}\{f\}$ forms the probability space Ξ, which we will call a *random field* on $\mathcal{C}$ [9].

The measure $\boldsymbol{\mu}$ can be recast in the form $P(c_1, c_2, \ldots, c_N)$ which is the probability that for all configurations x_i holds simultaneously $f(x_i) < c_i$, where $c_i \in \mathbb{R}$, and N is the total number of configurations. Then the *expected value* of a random variable X defined on the random field is given by $\int X\, d\boldsymbol{\mu}$ which takes the form

$$\mathcal{E}[X] = \int_{\mathbb{R}^N} X \,\mathrm{d}P(c_1, \ldots, c_N)\,.$$

We will restrict the use of the term *landscape* to mappings $f : \mathcal{C} \to \mathbb{R}$. Therefore, an element of a random field on $\mathcal{C}$ is a landscape.

3.1.2. Gaussian Random Fields

The well-known Gaussian probability density in $\mathbb{R}$ is

$$g_{m,\sigma^2}(x) \stackrel{\text{def}}{=} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-m)^2\right)\,.$$

The corresponding measure is $\boldsymbol{\nu}(m, \sigma^2) \stackrel{\text{def}}{=} g_{m,\sigma^2}\boldsymbol{\mu}_L$, where $\boldsymbol{\mu}_L$ denotes the Lebuesgue measure. The degenerate case $\boldsymbol{\nu}(m, 0)$ is by definition the Dirac measure centered at m. $\boldsymbol{\nu}(m, \sigma^2)$ is called a Gaussian measure for all m and σ. In the N-dimensional case one defines a measure $\boldsymbol{\mu}$ the be Gaussian if for any linear map $h : \mathbb{R}^N \to \mathbb{R}$ the measure $h(\boldsymbol{\mu})$ is Gaussian. It is uniquely defined by the vector $\vec{m}$ of means and the nonnegative definite covariance matrix C. We will say a random field is Gaussian if its probability measure is Gaussian, $\boldsymbol{\nu}(\vec{m}, C)$. If C is positive definite there is a probability density function

$$g_{\vec{m},C}(x) \stackrel{\text{def}}{=} \frac{1}{\sqrt{2\pi}^N} \frac{1}{\sqrt{\det\, C}} \exp\left(-\frac{1}{2}(x-\vec{m})C^{-1}(x-\vec{m})\right)$$

such that $\boldsymbol{\nu}(\vec{m}, C) = g_{\vec{m},C}\boldsymbol{\mu}_L$.

3.1.3. Averages

In many cases of practical importance, however, we are not given the random field, instead we have only a single instance (i.e., a landscape) and little or no clue from what kind of a random field it might have been drawn. This is the case, e.g., for the landscapes of RNA and protein folding, viral adaptation, and for some technical optimization problems such as the "low autocorrelated string problem" [55]. In this situation we have to be content with averages over values taken from our single instance. We will use the following notation: $\mathcal{E}[\,.\,]$ denotes averages over random fields, while angular brackets $\langle\ \rangle$ denote averages over configurations in a landscape[4]. For example we have

$$\mathrm{Var}[f(x)] = \mathcal{E}[f(x)^2] - \mathcal{E}[f(x)]^2 \qquad \text{and} \qquad \sigma^2\mathrm{var}[f] = \langle f^2\rangle - \langle f\rangle^2,$$

where the first average is taken over all sample functions of the random field f evaluated at a *fixed position* $x \in \mathcal{C}$, while the second average runs over all vertices x in a *fixed instance* f. The averages over configurations in a single landscape are of course merely well defined finite sums. In practice, however, one has to resort to Monte-Carlo sampling techniques in order to evaluate these mean values for a problem of practical interest. In particular we have

$$\langle f\rangle \stackrel{\text{def}}{=} \frac{1}{N}\sum_{x\in V} f(x),$$

which gives a precise meaning to the angular brackets.

Of course the two variances $\mathrm{Var}[f(x)]$ and σ^2 are not the same in general. This can be seen, for instance, from the following (trivial)

Example. Consider a (degenerate) Gaussian random field with $\mathcal{E}[f(x)] = 0$ for all x and $\mathcal{E}[f(x)f(y)] = 1$ for all x and y. A sample function drawn from this random field is almost always constant since the correlation of any two points is 1. The constant value $\bar{f}$ of this flat landscape, however, is drawn from a Gaussian distribution with mean zero and unit variance. In fact: $\mathrm{Var}[f(x)] = 1$ for all x, while the empirical variance measured on a sample function is almost always $\sigma^2 = 0$ since almost all sample landscapes drawn from this random field are completely flat.

If, for a given quantity X, both averages give the same result in a particular model (at least in the limit of infinite system size), one says that X is a *self-averaging* quantity [14, 88].

[4] This notation should not be confused with $\langle\,,\,\rangle$ or $\langle\,;\,\rangle$ which are used for scalar products.

3.2. Karhunen-Loève Decomposition

Recall that a random field on Γ is a probability space $\Xi = (\{f : \mathcal{C} \to \mathbb{R}\}, \boldsymbol{\mu})$. An important characteristic of Ξ is its covariance matrix $C = (C_{xy})_{x,y\in V}$, which is defined element-wise by

$$C_{xy} = \mathcal{E}[f(x)f(y)] - \mathcal{E}[f(x)]\mathcal{E}[f(y)].$$

Denote by $\{\phi_s\}$ a complete set of orthonormal eigenvectors (which exists by symmetry of C). Although $\{\phi_s\}$ can be chosen real, we will admit complex vectors as well. Then the random field f on Γ can be represented in mean square sense as

$$f(x) \doteq \sum_{y\in\Gamma} a_y \phi_y(x), \quad \text{i.e.,} \quad \mathcal{E}\left[\left(f(x) - \sum_{y\in\Gamma} a_y\phi_y(x)\right)^2\right] = 0\,.$$

This series representation is known as the *Karhunen-Loève* decomposition of the random field Ξ. It coincides for finite sets with the well known *principal component analysis* introduced by Hotelling in 1933 [64], see also [102]. Its most important property is

Theorem. [64] *The coefficients a_y are uncorrelated with respect to the measure $\boldsymbol{\mu}$, i.e., $\mathcal{E}[a_x a_y^*] = \mathcal{E}[a_x]\mathcal{E}[a_y^*]$ if $x \neq y$.*

3.3. The Markov Property

Properties related to the Markov property of time series have been first studied in the context of spatial stochastic processes. There are two non-equivalent definitions of nearest neighbor models in the literature, one due to Bartlett [6, sect. 2.2] based on conditional probabilities, and the other one due to Whittle [143] based on a particular product form of the joint probability distribution $P(\vec{c})$. These two approaches have been linked in the early seventies by the Hammersley-Clifford theorem, which we will discuss below.

Let $p(c_1, c_2, \ldots, c_N)$ be a probability density function (or, if the number of possible states c_k is countable, the probability for the $f(x_i) = c_i$ for all i). Recall that in the continuous case we have

$$p(c_1, c_2, \ldots, c_N) = \frac{\partial^N}{\partial c_1 \partial c_2 \ldots \partial c_n} P(c_1, c_2, \ldots, c_N)$$

in terms of the cummulative probabilities used above.

Definition. A random field Ξ on a graph Γ has the *Markov Property* if the conditional probability density

$$p(c_x | c_1, c_2, \ldots c_{x-1}, c_{x+1}, \ldots c_N)$$

depends for each $x \in V$ only on the neighbors of x, that is, on the values c_y of all all vertices $y \in N(x)$.

Theorem. (Hammersley-Clifford)
Let $p(\vec{c}) > 0$ be an absolutely continuous probability density function (in the continuous case, or

the probabilities of the states $\vec{c}$ in the discrete case). Then the following holds:
The random field Ξ is Markovian if and only if there is a function

$$Q(\vec{c}) \stackrel{\text{def}}{=} g_0 + \sum_{i=1}^{N} c_i g_i(c_i) + \sum_{i<j} c_i c_j g_{ij}(c_i, c_j) + \sum_{i<j<k} c_i c_j c_k g_{ijk}(c_i, c_j, c_k) + \dots$$

fulfilling $p(\vec{c}) = \exp(Q(\vec{c}))/\int \exp(Q(\vec{c}))\mathrm{d}\vec{c}$ such that $g_{i_1,i_2,\dots,i_p}$ vanishes identically whenever the subgraph of Γ induced by vertex set $\{x_{i_1}, x_{i_2}, \dots, x_{i_p}\}$ is not complete, i.e., if not all pairs of these vertices are edges in Γ.

Proof. The original Proof by Hammersley and Clifford (1971) has never been published. A detailed discussion can be found in Besag's paper [9]. A proof for discrete variables can be found in [56]. Moussouris [93] gives three different proofs and shows that the positivity condition cannot be omitted. Less general results for regular lattices are described in [3, 35, 119]. ∎

The denominator $\int \exp(Q(\vec{c}))\mathrm{d}\vec{c}$ is reminiscent of the partition function in statistical physics. A very detailed theory is available for stationary Gaussian Markov time series. It remains to be seen if a comparably rich and interesting structure can be developed also for the class of Markovian random fields on graphs. A simple result in this direction will be discussed at the end of the following section.

3.4. Isotropy

3.4.1. Characterization of Isotropic Random Fields

Definition. A random field on a graph Γ is *isotropic* if and only if

(i) $\mathcal{E}[f(x)] = a_0$ for all configurations $x \in \Gamma$.

(ii) Suppose the pairs of configurations (x, y), (u, v) belong to the same symmetry class $\ddot{\mu}$. Then $\mathcal{E}[f(x)f(y)] = \mathcal{E}[f(u)f(v)]$.

Remark. The notion of isotropy for random fields is analogous to the defintion of stationarity for stochastic processes. Following the conventions of Karlin and Taylor [69] our notion of isotropy should be called "covariance isotropic", "weakly isotropic", or "wide sense isotropic". Weinberger's [141] definition corresponds to "strictly isotropic". For a Gaussian random field the notions of (weak) isotropy and strict isotropy coincide.

Lemma. Let Γ be a vertex transitive graph. Then a random field on Γ is isotropic if and only if

(i) $\mathcal{E}[f(x)] = a_0$ for all $x \in V$, and

(ii) there exist a constant $C_{00} = \mathrm{Var}[f(x)]$ and a real valued function ρ of the symmetry classes with $\rho(0) = 1$ such that

$$C = C_{00} \sum_{\mu} \rho(\mu) R^{(\mu)}.$$

Proof. This is a simple rewriting of the above definition, observing that all entries in the covariance matrix belonging to a fixed symmetry class μ are the same and correspond to the non-zero entries of $R^{(\mu)}$. ∎

Note that by symmetry of the covariance matrix C we have $\rho(\mu) = \rho(\mu^+)$, i.e., $C = C_{00}\sum_{\tilde\mu}\rho(\tilde\mu)R^{(\tilde\mu)}$. The function $\rho(\mu)$ is called the *autocorrelation function* of the random field. It will be discussed in section 3.3.

Theorem. Let Γ be a vertex transitive graph for which the matrices $R^{(\tilde\mu)}$ form a symmetric association scheme, and let Ξ be an isotropic random field on Γ. Then the adjacency matrix A of Γ and the covariance matrix C of the random field f commute, i.e., Fourier decomposition and Karhunen-Loève decomposition coincide.

Proof. By assumption there is a commutative algebra $\langle\tilde{\mathfrak{R}}\rangle$ spanned by the matrices $R^{(\tilde\mu)}$. Since C is a linear combination of these matrices we have $C \in \langle\tilde{\mathfrak{R}}\rangle$. Of course A is contained in this algebra as well, and hence $AC = CA$. ∎

Remark. The case of CGCG is dealt with in [141]. The case of distance transitive graphs is proved in [122], and an extension to vertex transitive graphs for which the commuting algebra coincides with the adjacency algebra is discussed in [126].

As a consequence we immediately conclude that the Fourier coefficients are uncorrelated for an isotropic random field on a vertex transitive graph with the Θ-property, see sect. 2.5.3. A more detailed result can be obtained for isotropic landscapes on a CGCG.

Theorem. A random field on a CGCG is isotropic if and only if
(i) $\mathcal{E}[a_l] = 0$ for all $l > 0$; the coefficient a_0 is arbitrary.
(ii) The coefficients are pairwise uncorrelated, $\mathcal{E}[a_k a_l] = \mathcal{E}[|a_l|^2]\delta_{kl}$.
(iii) $\mathcal{E}[|a_l|^2] = V(\mu)$ depends only on the symmetry class μ to which l belongs.

Proof. A complete proof is given in [126] based on [141]. ∎

Note that this theorem depends on the fact the we can relate the eigenfunctions $e_p(x)$ directly with symmetry classes. This is not possible, however, for more general graphs [126]. The relation of Fourier coefficients and isotropy on more general graphs certainly deserves attention in forthcoming investigations.

3.4.2. Landscapes on the Hypercube and the p-Spin Models

The vertices of a Boolean hypercube are the sequences of length n taken from an alphabet of size $\alpha = 2$. Without loosing generality we may use the alphabet $\{+1, -1\}$. A configuration is then a string σ of "spins" $\sigma_k \in \{+1, -1\}$. An alternative encoding uses a binary string x, with $x_i \in \{0, 1\}$. The following result is well known in the literature:

Lemma. Any landscape f on the Boolean Hypercube can be written as

$$f(\sigma) = J_0 + \sum_{p=1}^{n}\sum_{i_1<i_2<\ldots<i_p} J_{i_1 i_2\ldots i_p}\sigma_{i_1}\sigma_{i_2}\ldots\sigma_{i_p}\,,$$

where the $J_{i_1 i_2\ldots i_p}$ are constants.

Proof. See, e.g., [126, 141]. ∎

The p-spin Hamiltonian

$$\mathcal{H}_p(\sigma) = \sum_{i_1<i_2<\ldots<i_p} J_{i_1 i_2\ldots i_p}\sigma_{i_1}\sigma_{i_2}\ldots\sigma_{i_p}$$

was introduced by Derrida [31] in order to bridge the gap between the SK model [116], which is the special case $p = 2$ and the random energy model [31, 30, 32, 51]. The coefficients $J_{i_1 i_2 \ldots i_p}$ are i.i.d. Gaussian random variables with mean 0 and unit variance.[5] The autocorrelation function of the p-spin Hamiltonian $\mathcal{H}_p(\sigma)$ is simply $\rho = \rho_p$ for all p [139]. Finally we remark that $\varepsilon_q(\sigma) \stackrel{\text{def}}{=} \sigma_{i_1}\sigma_{i_2}\ldots\sigma_{i_p}$ is of course only an alternative representation of the Fourier basis vectors $e_q(x)$ on the Boolean hypercube.

3.4.3. Autocorrelation Functions of Isotropic Random Fields

As usual let Γ be a vertex transitive graph. Consider a simple random walk $\{x_0, x_1, \ldots\}$ on Γ, see sect. 2.3.1. Weinberger [140] has suggested to use the time series obtained by evaluating the random field at each step in order to gather information on the random field itself. The correlation function along such a time series is defined by

$$r(s) = \frac{\mathcal{E}[f(x_{t+s})f(x_t)] - \mathcal{E}[f(x_{t+s})f(x_t)]}{\sqrt{\mathrm{Var}[f(x_{t+s})]\mathrm{Var}[f(x_t)]}} \,.$$

For isotropic random fields this reduces to

$$r(s) = \frac{\mathcal{E}[f(x_s)f(x_0)] - \mathcal{E}[f(x_0)]^2}{\mathrm{Var}[f(x_0)]} = \sum_\mu \varphi_{s\mu}\rho(\mu) \,.$$

In the light of this equation it is not surprising that there is a simple relation between the autocorrelation functions $\rho(\mu)$ and $r(s)$.

Theorem. Let Γ be a vertex transitive graph and consider an isotropic random field on Γ with correlation functions $\rho(\mu)$ and $r(s)$, respectively. Then:

(i) $r(s) = \lambda^s$ if and only if $\rho\hat{A} = \Lambda\rho$.

(ii) Let $\{\rho_p\}$ be the orthonormal system of left eigenvectors of $\hat{A}$ described in section 2.7. We have $r(s) = \sum_p b_p\lambda_p$ if and only if $\rho(\mu) = \sum_p b_p\rho_p$. Of course, such a decomposition always exists.

Proof. This theorem is the content of [123]. ∎

An immediate consequence is the following surprising relation between eigenvalues and eigenvectors of a large class of graphs.

Corollary. Let Γ be symmetric, i.e., assume that Γ is vertex transitive and that it has a symmetry class

$$\ddot{\mathbf{1}} = \{\, (x,y) \,|\, d(x,y) = 1 \,\}$$

consisting of all pairs of neighbors. Then $\rho_p(\mathbf{1}) = \lambda_p$.

Proof. See [126]. ∎

[5] In Derrida's original formulation the variance was chosen to be $S_p^2 = p!/n^p$, such that the variance of the Hamiltonian itself is independent of n.

3.4.4. Gaussian Markovian Isotropic Random Fields

A stationary Gaussian Markov process in $\mathbb{R}$, a so-called Ornstein-Uhlenbeck process or AR(1) process, has an autocorrelation function of the form $\rho(t) = \exp(-at)$. Gaussian Markovian isotropic random fields present an immediate generalization. Hence we are interested in the form of their covariance matrices. For simplicity we will assume that the Gaussian distribution is non-degenerate (i.e., that the covariance matrix C is invertible) and that the configuration space Γ is a symmetric graph.

Theorem. Let Γ be symmetric and let Ξ be a Markovian Gaussian isotropic random field on Γ. Then its covariance matrix is of the form

$$C = a_0(I + \alpha A)^{-1}$$

Proof. Consider the function $Q(\vec{c})$. Since Ξ is Gaussian we have

$$Q(\vec{c}) = -\frac{1}{2}\left((\vec{c} - \vec{m})C^{-1}(\vec{c} - \vec{m})\right)$$

where $\vec{m}$ is the vector of expected values $m_x \stackrel{\text{def}}{=} \mathcal{E}[f(x)]$. Isotropy implies $\vec{m} = m_0\mathbf{1}$. Without loosing generality we can assume therefore that $m_0 = 0$. The Markov property now requires that $(C^{-1})_{xy} = 0$ unless $x = y$ or x and y are neighbors in Γ. Since C is an element of the commuting algebra of Γ so is C^{-1}, i.e., C^{-1} must be a linear combination of the matrices $R^{(\mu)}$. Symmetry of Γ implies that the identity I and $R^{(\mathbf{1})} = A$ are the only $R^{(\mu)}$ matrices which have non-zero entries only in the diagonal and for pairs of neighbors. Thus the covariance matrix must be of the form $C^{-1} = b_0I + b_1A$. ∎

3.5. Homogeneity

The definition of isotropy becomes void if the graph Γ does not have symmetries. In fact, almost all graphs have only the trivial group of automorphisms. Nevertheless, it is possible to consider random fields and their correlation structure on such objects. To this end we introduce the following

Definition. A random field Ξ on a graph Γ is *homogeneous* if
(i) $\mathcal{E}[f(x)] = a_0$ for all $x \in V$, and
(ii) $\mathcal{E}[f(x)f(y)] = \mathcal{E}[f(u)f(v)]$ whenever $d(u, v) = d(x, y)$.

Note that isotropy and homogeneity are the same thing if Γ is distance transitive.

Homogeneous random fields play a special role on distance regular graphs. Isotropy is in general trivial on this class of graphs, since distance regular graphs need not even be vertex transitive. In particular, the main result of sect. 3.4.3 holds for homogeneous random fields on distance regular graphs. In general, however, isotropy proves to be the more powerful concept for deriving interesting theorems. In particular, the condition for obtaining exponential autocorrelation functions along simple random walks on a non distance transitive graph requires in general different covariances for pairs of vertices in different symmetry classes which may well belong to the same distance class. Thus there will in general be no homogeneous random fields with this property.

3.6. Superpositions of Random Fields

3.6.1. Definition

Note that we regard a p-spin model as random field on the Boolean Hypercube. For the next result we have to state precisely what we mean by a "sum" of random fields.

Definition. Let $\Xi_i = (\{f : V \to \mathbb{R}\}, \boldsymbol{\mu}_i)$ be random fields on Γ. The *superposition* of two random fields on Γ with fixed coefficients $a, b \in \mathbb{R}$ is the probability space

$$a\Xi_1 + b\Xi_2 \overset{\text{def}}{=\!=} (\{f : V \to \mathbb{R}\}, \boldsymbol{\mu}^*) \quad \text{with} \quad \boldsymbol{\mu}^*(f) = \int_g \boldsymbol{\mu}_1(\frac{1}{a}g)\boldsymbol{\mu}_2(\frac{1}{b}(f-g)).$$

The effect of this definition is that elements of the probability spaces Ξ_i are added as independent random variables.

Theorem. A Gaussian isotropic random field on the Boolean hypercube is a superposition of p-spin models with independent coefficients.

Proof. See [126]. ∎

3.6.2. The Random Energy Model

The results derived in the previous section allow us to calculate a "spectrum" $\Omega(\ell) = \Omega_\ell$ of the autocorrelation functions: $\rho(d) = \sum_{\ell=0}^{n} \Omega_\ell \rho_\ell(d)$.

We consider here a random field on a vertex transitive graph, where the values $f(x)$ are i.i.d. random variables (see, e.g., [50]. This Random Energy Model (REM) introduced by Derrida [31, 30, 32] has been applied to a variety of problems, in particular to the maturation of the immune system [72, 81].

Lemma. Let Γ be vertex transitive. The REM is isotropic and has autocorrelation function $\rho(\mu) = 0$ for all $\mu \neq 0$.

Proof. See [126]. ∎

As a consequence a Gaussian REM on a Boolean hypercube – that is, Derrida's original model – is a superposition of p-spin models. Derrida argued that the REM is the limit $p \to \infty$ of the p-spin models. The following theorem makes the sense in which this is true more precise.

Theorem. Let Γ be vertex transitive. Then the spectrum of the REM is

$$\Omega(\ell) = \frac{1}{\langle \rho_\ell; \rho_\ell \rangle}$$

On a Boolean hypercube $\mathbf{S}_2^n$ we have explicitly: $\Omega(p) = \dfrac{1}{2^n}\dbinom{n}{p}$.

Proof. See [126]. ∎

Hence the REM on the Boolean hypercube can be represented as a superposition of a "continuum" of p-spin models with $p \approx n/2$.

3.6.3. Nk Models

The Nk model assigns a real valued "fitness" to the bit string $\mathbf{b}$ by first assigning a real valued "fitness contribution", f_i , to the i^{th} bit, b_i, in $\mathbf{b}$. Each such assignment depends, not just on i and the value of b_i, but also on $0 \leq k < n$ other bits, which we call its "neighbors". The fitness contribution of each site is a random function, $f_i(\mathbf{b}_i)$, of the substring, $\mathbf{b}_i$, formed by the i^{th} bit and its k neighbors. $f_i(\mathbf{b}_i)$ is assigned by selecting an independent random variable from some distribution, such as the uniform or Gaussian distributions, for each of the 2^{k+1} possible values of $\mathbf{b}_i$, thus generating a "fitness table" for the i^{th} site. There is a different, independently generated table for each of the n sites. Then, given any string of n bits, the total fitness of the string, f, is defined as the average of the fitness contributions of each site; that is,

$$f(\mathbf{b}) = \frac{1}{n}\sum_{i=1}^{n} f_i(\mathbf{b}_i).$$

The Nk model comes in different "flavors" depending on the choice of neighbors influencing a particular site fitness. The simplest — but not the only — way of choosing neighbors, at least for even k, is to use the k sites adjacent to site i; that is, the bits at sites $i - k/2$ through $i + k/2$. As in the original formulation of the model, we introduce periodic boundary conditions to assign neighbors to sites i with $i \leq k/2$ and $i \geq n - k/2$. In other words, we assume that the sites are arranged in a circle, such that site n is next to site 1. (Periodic boundary conditions are chosen because they minimize chain length dependent end effects and because we are interested in bulk properties only). This choice of neighbors gives rise to a class of short range spin glasses. We will call this arrangement the "adjacent neighborhood" model, AN. Alternatively, we could assign the neighbors by randomly selecting, for each site i, k other sites. This assignment of neighbors makes the model similar to a long range, dilute spin glass. It will be referred to as "random neighborhood" model, RN. In a third variation we drop the requirement that f_i depends on bit i. This version will be called "purely random" model, PR. The Nk model reduces to the REM for $k + 1 = n$, see e.g., [70].

Lemma. The autocorrelation functions of the Nk model with random, purely random, and adjacent neighborhood, respectively are given by

$$\begin{aligned}
\rho_{RN}(d) &= (1-\frac{d}{n})(1-\frac{k}{n-1})^d \\
\rho_{PR}(d) &= (1-\frac{k+1}{n})^d \\
\rho_{AN}(d) &= 1-\frac{k+1}{n}d+\frac{1}{\binom{n}{d}}\sum_{j=1}^{\min(k,n+1-d)}(k-j+1)\binom{n-j-1}{d-2}.
\end{aligned}$$

Proof. For a proof see, e.g., the appendix in [47]. ∎

Lemma. The autocorrelation function of the adjacent neighbor Nk model, ρ_{AN}, is a polynomial of degree $k+1$. Its autocorrelation function is therefore of the form

$$\rho_{AN}(d) = \sum_{p=0}^{k+1} \Omega_p \rho_p(d)\,.$$

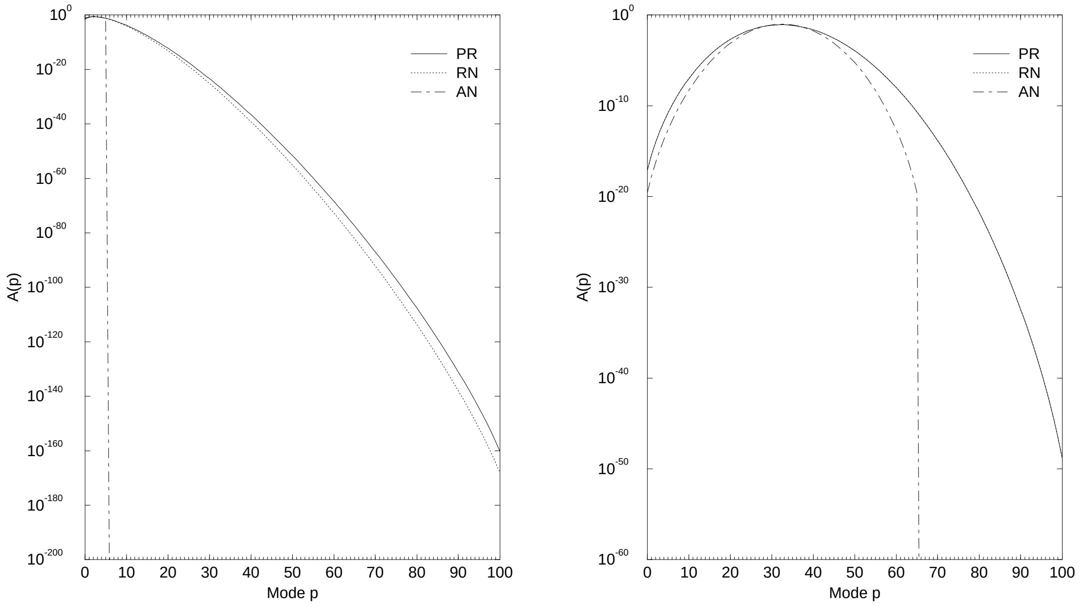


Figure 7: Spectral decompositions of Nk models on Boolean hypercubes with $n = 100$. The logarithmic plots emphasize that the two random neighborhood Nk models involve all modes $0 \leq p \leq n$ for all k while the AN models are superpositions of the modes $0 \leq p \leq k+1 \leq n$.

Proof. See [126, Thm.11]. ∎

The coefficients $\Omega(\ell)$ can be obtained explicitly for sequence spaces, if the correlation function $\rho(d)$ is known.

$$\Omega(\ell) = \frac{\langle \rho_\ell ; \rho \rangle}{\langle \rho_\ell ; \rho_\ell \rangle} = \frac{n!}{\alpha^n} \sum_{i=0}^{n} \rho(i) \sum_{j=0}^{p} \frac{(-1)^j (\kappa - 1)^{p+i-j}}{j!(i-j)!(p-j)!(n+j-i-p)!}$$

For Nk models this expression has been evaluated numerically, results are shown in figure 7.

Weinberger and Stadler [139] concluded based on a comparison of correlation lengths, that Nk models roughly resemble p-spin models with $p = (k+1)/2$. This view is supported by the spectra of the Nk models showing that the dominating modes are those centered around $\bar{\pi} = (k+1)/2$ for all three versions of the Nk model [126]. The convergence of Nk models and p-spin models is similar to the case of the p-spin models and the REM. If k scales linearly with n, i.e., if $(k+1)/n \to \chi > 0$, then we expect that the spectrum in terms of the scaled variables $\psi = p/n$ will be concentrated at $\bar{\psi} = \chi/2$ in the limit $n \to \infty$.

3.6.4. Superposition of Nk Models

We have seen that all isotropic random fields on the Boolean hypercube can be constructed by superpositions of p-spin models. Kauffman's Nk-models cannot be used for this purpose. Let $\{\mathcal{N}_k(x)\}$ denote a set of Nk models with parameter k, such that

$$\mathcal{E}[\mathcal{N}_i(x)\mathcal{N}_j(y)] - \mathcal{E}[\mathcal{N}_i(x)]\mathcal{E}[\mathcal{N}_j(y)] = 0$$

whenever $i \neq j$, i.e., we suppose that the entries in the fitness tables are uncorrelated also between the random fields. Given the construction of the Nk model this is the natural assumption. Then we have the following

Theorem. Let Ξ be an isotropic random field of the form $\Xi = \sum_{k=0}^{k_{\max}} \beta_k \mathcal{N}_k$, i.e., Ξ is a superposition of Nk models. Then $\Omega(p) > 0$ for all $p \leq k_{\max} + 1$. Furthermore $\Omega(p) > 0$ for all $p \leq n$ if at least one of the Nk models $\mathcal{N}_k$ is an RN or a PR version of the Nk model.

Proof. The result is obtained by a direct calculation of the covariance matrix, see [126]. ∎

Corollary. A p-spin model cannot be represented as a superposition of Nk models for all $p \geq 1$. The empirical correlation functions of the 1-spin model and the Nk-model with $k = 0$ coincide. The empirical correlation function of a p-spin model with $p \geq 2$ cannot be modeled by a superposition of Nk models.

3.7. Transformations

3.7.1. Operators on Landscapes and Random Fields

Let $f : \mathcal{C} \to \mathbb{R}$ be a landscape on $\mathcal{C}$ and let $F : R^{|V|} \to R^{|V|}$ be a vector field. Then we explain $F[f] : \mathcal{C} \to \mathbb{R}$ component-wise as

$$F[f](x) \stackrel{\text{def}}{=} (\, F(f(x_1), f(x_2), \ldots, f(x_N)\,)_x \,,$$

i.e., we interpret F as an operator acting on the landscape f. Before we proceed we briefly introduce the two most common examples:

(i) The averaging (or smoothing) operator Ψ is defined in the above notation as

$$\Psi[f](x) \stackrel{\text{def}}{=} \frac{1}{1 + N(x)} \left(f(x) + \sum_{y \in N(x)} f(y) \right) \;=\; \frac{1}{1 + N(x)} [(I + A)f]_x .$$

Ψ is of course a linear operator, its effect is averaging the fitness values over a ball of radius 1. Consequently it amounts to applying the matrix $(I + D)^{-1}(I + A)$ to the vector of fitness values.

(ii) Let $\beta : \mathbb{R} \to \mathbb{R}$ be an arbitrary function. Then $\beta[f]$ is explained as the composition of f and β, i.e.,

$$\beta[f](x) \stackrel{\text{def}}{=} \beta(f(x)).$$

Now let $\Xi = (\{f : \mathcal{C} \to \mathbb{R}\}, \mu)$ be a random field on $\mathcal{C}$. Then

$$F[\Xi] \stackrel{\text{def}}{=} (\{F[f] : \mathcal{C} \to \mathbb{R}\}, \mu)$$

is again a random field on $\mathcal{C}$. In the following we will briefly discuss some properties of the random fields $F[\Xi]$ which are obtained from a simple random field Ξ by means of the two types of operators described above.

3.7.2. Averaging and Iterated Smoothing Landscapes

The basic properties of linear transformations of random fields are compiled in the following

Theorem. Let C be the covariance matrix of a random field Ξ and let Ω be a linear symmetric operator.

(i) The covariance matrix of $\Upsilon = \Omega[\Xi]$ is given by $C^{\Omega} = \Omega C \Omega$.

(ii) If Ξ is isotropic and Γ is a vertex transitive graph for which the matrices $\tilde{R}^{(\mu)}$ form a symmetric association scheme, then $C^{\Omega} = (\Omega^2)C$.

(iii) If Γ is a in (ii) and $\Omega \in \mathfrak{A}[\Gamma]$ (that is, Ω is a polynomial in A), then $\Upsilon = \Omega[\Xi]$ is again isotropic on Γ and $\rho^{\Omega} = c \cdot \hat{\Omega}^2$, where the constant $c = 1/(\hat{\Omega}^2\rho)(0)$ is chosen such that $\rho^{\Omega}(0) = 1$.

(iv) Under the assumptions of (iii) holds $\rho^{\Omega}(\mu) = c \cdot \sum_{\nu} a_\nu P^2(\lambda_\nu)\rho_\nu(\mu)$, where λ_ν denotes the eigenvalue of $\hat{A}$ belonging to ρ_ν. In particular, if $\rho = \rho_j$ for some j, then we have $\rho^{\Omega} = \rho$, i.e., eigenfunctions of $\hat{A}$ remain unchanged.

Proof. See [122]. ∎

Let us now consider the smoothing operator $\Psi \stackrel{\text{def}}{=} \frac{1}{D+1}(A + E)$. Suppose Γ is distance transitive. Then we obtain as an immediate consequence of the above theorem the following explicit representation of the autocorrelation function

$$\rho^{\Psi}(d) = \frac{\sum_{k=-2}^{2} \alpha_k(d)\rho(d+k)}{\sum_{k=0}^{2} \alpha_k(0)\rho(k)}$$

where $\alpha_k(d) = (\hat{\Psi}^2)_{d,d+k}$. The coefficients have a simple combinatorial interpretation: $\alpha_k(d)$ is the number of pairs (u,v) with $u \in N(x)$, $v \in N(y)$ with $d(u,v) = d+k$ where $d(x,y) = d$ is fixed.

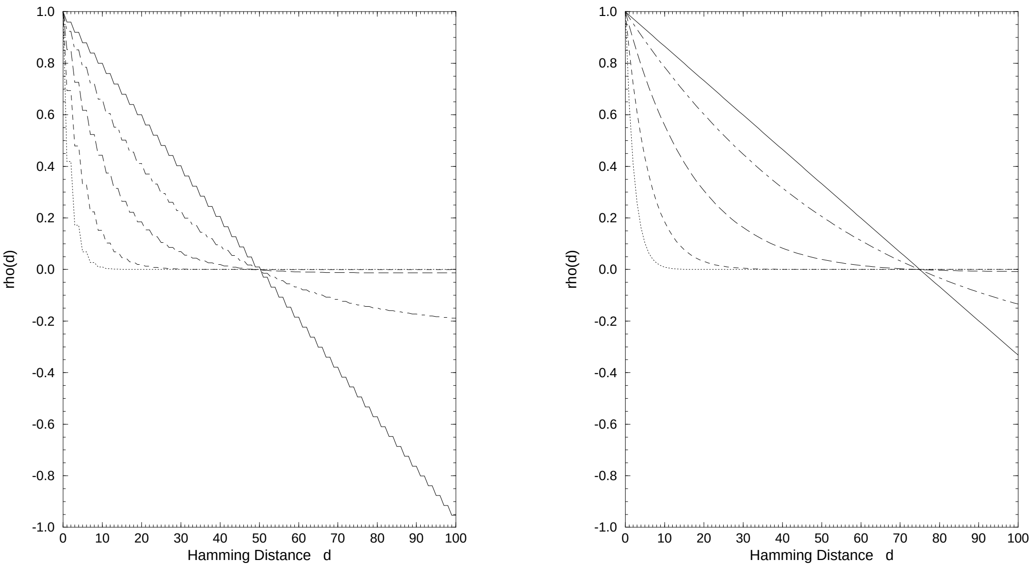
For particular graphs Γ it is fairly easy to calculate the coefficients $\alpha_k(d)$ explicitly. On Hamming graphs, for instance, one obtains [122]

$$\begin{aligned}
\alpha_{-2}(d) &= d(d-1) \\
\alpha_{-1}(d) &= 2d + (\alpha-2)(2d-1)d \\
\alpha_0(d) &= 1 + (\alpha-1)(n+2d(n-d)) + (\alpha-2)(\alpha d + d(d-1)(\alpha-2)) \\
\alpha_{+1}(d) &= 2(\alpha-1)(n-d) + (\alpha-2)(\alpha-1)(n-d)(2d+1) \\
\alpha_{+2}(d) &= (\alpha-1)^2(n-d)(n-d-1)
\end{aligned}$$

Let Ξ_0 be the random energy model. The autocorrelation function is $\rho(d) = \delta_{0,d}$. The corresponding empirical correlation function is $\hat{\rho}(0) = 1$ and $\hat{\rho}(d) = -1/(N-1)$ for $d > 0$, respectively, where N denotes the number of points in the configuration space, see sect. 4.1. Define $\Xi_q = \Psi[\Xi_{q-1}]$. Let $\hat{\rho}^{(q)}$ denote the empirical autocorrelation function expected for an instance of the random field Ξ_q. We will refer to this family of landscapes as *iterated smoothing landscapes*, ISLs. The numerical data shown in figure 8 indicate that they form indeed a family of tuneably rugged random fields, just like the Nk models. In particular, we have the following

Theorem. Let Γ be a generalized hypercube over an alphabet with $\alpha > 2$ letters. Then

$$\lim_{q\to\infty} \hat{\rho}^{(r)}(d) = 1 - \frac{1}{n}\frac{\alpha}{\alpha-1}d.$$



a) b)

Figure 8: Autocorrelation function of ISLs on generalized hypercubes with **a)** α=2, and **b)** α=4 for with n=100. The number of smoothings are **a)** r=25 (dotted), r=50 (short dashed), r=75 (long dashed), r=100 (dot-dashed), r=200 (solid), and **b)** r=50 (dotted), r=100 (short dashed), r=150 (long dashed), r=200 (dot-dashed), and r=400 (solid), respectively.

On a Boolean Hypercube ($\alpha = 2$) we have

$$\lim_{q\to\infty} \hat{\rho}^{(r)}(d) = \frac{n}{n+1}(1-\frac{2d}{n}) + \frac{1}{n+1}(-1)^d .$$

Proof. See [122]. ∎

We remark that the contribution of $\hat{\rho}_n$ decreases with the size of the configuration space. The limiting autocorrelation function corresponds to both the trivial spin glass model with Hamiltonian $\mathcal{H}(\sigma) = \sum_j A_j\sigma_j$, and to the Nk-model with $k = 0$. We also note that for $q/n \ll 1$ the landscapes are essentially uncorrelated, i.e., the correlation length is $o(n)$. Numerical estimates show that the nearest neighbor correlation $\hat{\rho}^{(q)}(1)$ depends only on the ratio q/n for large n, see [122] for details.

Instead of starting with a REM one can as well apply the averaging operator to spin glass and Nk models. For p-spin models one finds

Theorem. Let Ω be any polynomial of the adjacency matrix of the Boolean Hypercube. Then the landscapes $\mathcal{H}_p$ and $\Omega[\mathcal{H}_p]$ have the same autocorrelation functions.

Proof. See [122, Thm.14]. ∎

In contrast to the p-spin models, the landscapes of Nk type are affected by the smoothing operator.

Theorem. Let f be an arbitrary Nk landscape. Then the autocorrelation function of the landscapes $\Psi^r[f]$ converges to $\rho(d) = \xi(1-2d/n) + (1-\xi)(-1)^d$ with $0 < \xi \leq 1$ as q tends towards infinity.

Proof. See [122, Thm.15]. ∎

The variants of the Nk model do not behave identically under the action of Ψ. Whenever the autocorrelation function is a polynomial of degree less than n, then we have $\xi = 1$ in the theorem above. For exact expressions of the autocorrelation functions of various Nk models see sect. 3.6.3. The theorem holds with $\xi = 1$ for all alphabets with $\alpha \geq 3$ letters.

3.7.3. Non-Linear Transformations

Transforming a landscape or a random field with a non-linear function is important for instance in (bio)physical chemistry, since the chemical reaction rate k and activation energies $\Delta G^{\neq}$ are related by Arrhenius' law [76]: $k = A\exp(-\Delta G^{\neq}/RT)$, where R is a universal constant and T is the temperature. Consider now a fixed type of chemical reaction, say binding to a fixed target. Then $\Delta G^{\neq}$ will be sequence dependent, and thus it forms a landscape. Consequently the reaction rate constants form a landscape as well. If we assume, for simplicity, that the pre-exponential factor A is not sequence dependent, we observe immediately that both landscapes have the same geometry since the exponential function is monotonic and thus maps local optima to local optima, see also sect 4.3. We will see in the following that the correlation structure, however, does not remain unchanged.

As an example we consider here the correlation function of an exponential transformation $F(z) = \exp(qz)$ of a Gaussian isotropic random field with covariance matrix C. It is convenient to define the correlation coefficient $\rho(x,y) \stackrel{\text{def}}{=} \text{var}[f]^{-1}C_{xy}$. The covariance of the transformed random field is given by

$$C^*_{xy} = \mathcal{E}[F[f(x)]F[f(x)]] - \mathcal{E}[F[f(x)]]\mathcal{E}[F[f(x)]]\,.$$

For Gaussian random fields these expectation values are simple Gaussian integrals which can be evaluated explicitly. The details can be found in [131]. The final result is

$$\rho^*(x,y) = \frac{a^{\rho(x,y)} - 1}{a-1}\,,$$

where $a = e^{q^2}$ depends on the parameter q in the transformation. The function $f(t) = \frac{a^t-1}{a-1}$ is concave for all t and all $a > 1$, i.e. all $q \neq 0$, and furthermore $f(0) = 0$ and $f(1) = 1$. It follows immediately that $|\rho(x,y)| > |\rho^*(x,y)|$ whenever $\rho(x,y) \neq 0$, 1, or -1.

Suppose the correlation function is of the form $\rho(d) = 1 - \frac{1}{\ell}d + o(d)$, where ℓ is the correlation lenght, which will be discussed in sect. 4.1.4 in detail. A simple calculation then shows that $\rho^*(d)$ has the same form, with correlation length

$$\ell^* = \frac{a-1}{a\log a}\,\ell \approx \frac{1}{q^2}\,\ell$$

for large values of q. Thus the a-dependent pre-factor becomes very small for large absolute value of q. Thus an exponential transformation can lead to a random field with very small correlation without changing the rank order of the fitness values, and thus without disturbing the distribution of local optima.

4. Landscapes

Let us now consider individual landscapes instead of random fields. We will suppose that we have a means of computing (or measuring) the value of the landscape $f : V \to \mathbb{R}$ for all configurations $x \in V$. This means that we have now a fixed matrix of interaction coefficients in spin glass model or a fixed matrix of bond lengths for a TSP. Models for biologically important landscapes will be discussed in section 5.

4.1. Empirical Correlation Functions

4.1.1. Characterization of Empirical Correlation Functions

The empirical autocorrelation function of a landscape on a vertex transitive graph is defined as [43, 47, 44, 109]

$$\hat{\rho}(\mu) \stackrel{\text{def}}{=\!=} \frac{\langle f(x)f(y)\rangle_{(x,y)\in\ddot{\mu}} - \langle f\rangle^2}{\langle f^2\rangle - \langle f\rangle^2}.$$

Of course, the autocorrelation function is invariant under the transformation $f \to f - \langle f\rangle$. Therefore we can assume without loosing generality $\langle f\rangle = 0$ for the reminder of this chapter. Under this assumption we may rewrite the definition above in the form

$$\hat{\rho}(\mu) = \frac{1}{|\mu|}\frac{\langle f, R^{(\mu)}f\rangle}{\langle f, f\rangle},$$

where $\langle\, .\, ,\, .\, \rangle$ denotes as usual the standard scalar product on Euclidean vector spaces. Of course we have $\hat{\rho}(\mu) = \hat{\rho}(\mu^+)$ in complete analogy with the correlation functions of isotropic random fields. In fact, the above definition tacidly assumes that the landscape is "isotropic" in a certain sense, since we average over entire symmetry classes.

The appropriate definition of a correlation function on a distance regular graph is analogous to homogeneous landscapes:

$$\hat{\rho}_h(d) \stackrel{\text{def}}{=\!=} \frac{\langle f(x)f(y)\rangle_{d(x,y)=d}}{\langle f^2\rangle} = \frac{1}{|N_d|}\frac{\langle f, A^{(d)}f\rangle}{\langle f, f\rangle},$$

where we have again used $\langle f\rangle = 0$. Note that for distance transitive graphs both definitions coincide.

In the following we characterize the set of all possible empirical autocorrelation functions for sufficiently symmetric configuration spaces.

Theorem. Let $\{\rho_\mu\}$ be an orthogonal basis of left eigenvectors of the collapsed adjacency matrix of a CGCG Γ. Then $\hat{\rho}$ is an autocorrelation function of a landscape on Γ if and only if

$$\hat{\rho} = \sum_\mu \Omega_\mu \rho_\mu, \quad \Omega_0 = 0, \quad \Omega_\mu \geq 0, \quad \sum_\mu \Omega_\mu = 1.$$

The coefficients Ω_μ can be obtained from the Fourier coefficients of f:

$$\Omega_\kappa = \frac{\sum_{g\in\kappa} |a_g|^2}{\sum_{g\in G} |a_g|^2}.$$

Proof. See [126, Thm.5]. ∎

As an immediate consequence, all empirical autocorrelation functions fulfil $\sum_\mu \hat{\rho}(\mu)|\mu| = 0$.

4.1.2. Simple Random Walks on Landscapes

Before we proceed to landscapes on more general graph let us consider the empirical correlation function along a simple random walk. The geometric relaxation of such a random walk is conveniently described by the probabilities $\varphi_{s\kappa}$, defined in section 2.3.3, that a random walks of length s ends in symmetry class κ. On a vertex transitive graph one obtains [123]

$$\varphi_{s\kappa} = \sum_{\mu} \frac{|\kappa|\,\rho_\mu(\kappa)}{\langle \rho_\mu ; \rho_\mu \rangle}\,\lambda_\mu^s .$$

An analogous formula holds for distance regular graphs, with the distance classes replacing the symmetry classes. For instance, on a sequence space $\mathcal{Q}_\alpha^n$ we obtain explicitly

$$\varphi_{sd} = \sum_{p=0}^{n} \mathbf{K}_p^{(\alpha,n)}(d) \frac{(\alpha-1)^p \binom{n}{p}}{\alpha^n} \left(1 - \frac{\alpha}{\alpha-1}\frac{p}{n}\right)^s .$$

A simple random walk $\{x_0, x_1, \dots\}$ induces a time series $\{f(x_0), f(x_1), \dots\}$ on a landscape f. The autocorrelation function of this time series is defined by

$$\hat{r}(s) \stackrel{\text{def}}{=\!=} \frac{\langle f(x_t) f(x_{t+s})\rangle_t - \langle f\rangle^2}{\langle f^2\rangle - \langle f\rangle^2} .$$

Without loosing generality we assume as above that $\langle f\rangle = 0$. Clearly the denominator is just the variance of the landscape, $\sigma^2 = \langle f^2\rangle$. The landscape autocorrelation $\hat{\rho}(\mu)$ and the random walk autocorrelation are related by

$$\hat{r}(s) = \sum_{\kappa} \varphi_{s\kappa}\hat{\rho}(\kappa) ,$$

see, e.g.,[140]. The relation between the correlation functions $\hat{\rho}$ and $\hat{r}$ is completely analogous to the relation of ρ and r in section 3.4.3. In particular, we have the following

Theorem. Let Γ be a CGCG with an adjacency matrix with eigenvalues $D\lambda_\mu$ and corresponding left eigenvectors ρ_μ. Then $\hat{r}(s)$ is an autocorrelation function of a simple random walk of a landscape defined on Γ if and only if

$$\hat{r}(s) = \sum_{\mu} \Omega_\mu \lambda_\mu^s, \qquad \Omega_0 = 0, \qquad \Omega_\mu \geq 0, \qquad \sum_{\mu} \Omega_\mu = 1.$$

The coefficients are the same as for the correlation function $\hat{\rho}$.

Proof. See [126, Thm.6]. ∎

The eigenvalues of the transition matrix T are explicitly known for sequence spaces:

$$\lambda_d = \frac{\Lambda_d}{D} = 1 - \frac{d}{n}\frac{\alpha}{\alpha-1} .$$

4.1.3. General Vertex Transitive Graphs

Let us denote the set of indices of eigenvectors θ_p belonging to the same eigenvalue $\tilde{\Lambda}_i$ of Δ by $L(i)$. Recall that a vertex transitive graph has the Θ-property if and only if all eigenvalues of its collapsed adjacency matrix are simple

Definition. Let Γ be vertex transitive with the Θ-property, and let $\{\theta_y\}$ be a Fourier basis. Then define

$$\omega_i(\mu) = \frac{1}{|\mu|} \frac{\langle \theta_y \, , R^{(\mu)} \theta_y \rangle}{\langle \theta_y \, , \theta_y \rangle} \, ,$$

where $y \in L(i)$.

Theorem. Let f be a landscape on a vertex transitive graph Γ which has the Θ-property, or a distance regular graph. Then its empirical autocorrelation function is of the form

$$\hat{\rho}(\mu) = \sum_i \left[\frac{\sum_{q \in L(i)} a_q^2}{\sum_{q \neq 0} a_q^2} \right] \omega_i(\mu)$$

Proof. See [126, Thm.7] for the case of vertex transitive graphs with the Θ-property. The case of distance regular graphs can be dealt with analogously. ∎

For Cayley graphs of commutative groups we have of course $\omega_i \equiv \rho_i$ for all i. Furthermore, it can be shown that this is also true for a number of examples including the Petersen graph and the Cayley graph $\Gamma(S_5, \mathcal{T})$, see [126]. This, and the very form of the above decomposition, strongly suggests the following

Conjecture. Let Γ be a vertex transitive graph with the Θ-property. Then $\omega_i \equiv \rho_i$.

We have not been able to find a counter example to this conjecture.

4.1.4. Correlation Length

In order to easily compare different landscapes, or random fields, it is desirable to reduce the information contained in the correlation functions to single number. Most of the emprical correlation functions $\hat{r}(s)$ which have been computed so far are at least approximately decayin exponentials of the form

$$r(s) = \exp(-s/\ell) \, ,$$

where the parameter ℓ is the *correlation length.* Numerically, two approaches have been persued in different papers for estimating ℓ, thereby deliberately neglecting potential devaitions of $\hat{r}(s)$ from the exponential form. In most cases ℓ is estimated by interpolation from $\hat{r}(\ell_\iota) \stackrel{\text{def}}{=} 1/e$ [43, 47, 44, 18, 131]. Alternatively, one may use linear regression for fitting

$$-\ln \hat{r}(s) \stackrel{\text{def}}{=} \frac{1}{\ell_{lr}} s + c_0 ,$$

sometimes with the $c_0 = 0$. This approach is used for instance in [140]. Of course, these estimates can differ significantly if there are substantial deviations from an exponential function.

A more elegant definition relies only on the nearest neighbor correlation $r(1)$, using

$$\ell \stackrel{\text{def}}{=} -\frac{1}{\ln r(1)} .$$

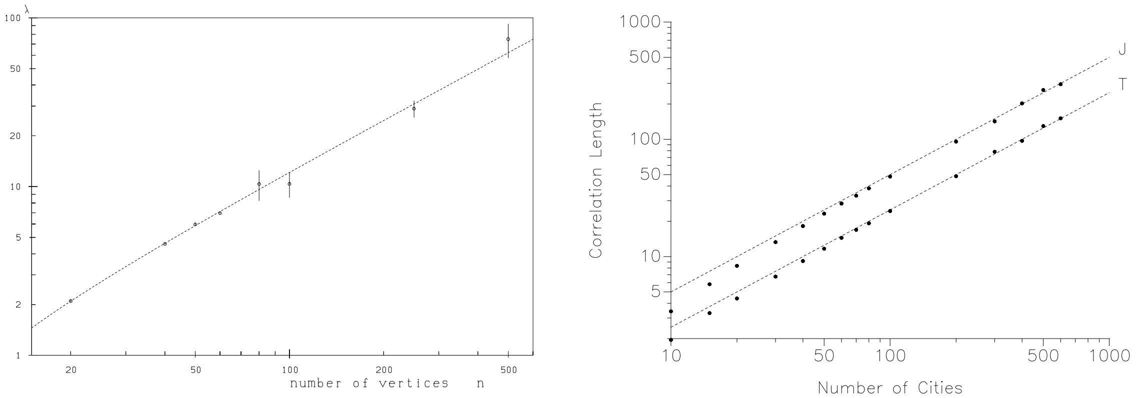


Figure 9: The correlation lengths of most model landscapes scales linearly with the system size. The examples shown here are: a) Graph Bipartitioning Problem, and b) Travelling Salesman Problem with transpositions (T) and inversions (J), respectively.

The correlation length — calculated by any of tha above numerical procedures — depends linearly on the system size n in almost all model landscapes that have been investigated so far [43, 127, 121, 125, 140], see figure 9. The only notable exception occurs for asymetric TSPs with inversions [127]. This strongly suggests to introduce the *scaled correlation length* $\xi \stackrel{\text{def}}{=} \ell/n$. It allows to compare entire families of landscapes, as it does not depend on the size n in general.

Alternative definitions of a scaled correlation function emphasize the geometry of the configuration space graph Γ by comparing ℓ to the diameter of Γ , $\xi' \stackrel{\text{def}}{=} \ell/\text{diam}\Gamma$, or to the average distance $\bar{d}$ of randomly chosen vertices in Γ, $\xi'' \stackrel{\text{def}}{=} \ell/\bar{d}$, respectively. We have thus

$$\xi = \frac{n}{\text{diam}\Gamma}\xi' = \frac{n}{\bar{d}}\xi'' .$$

The quantities relating the different definitions are, in most cases, numerical constants close to 1. For instance, we have for Hamming graphs

$$\frac{n}{\text{diam}\Gamma} = 1 \qquad \text{and} \qquad \frac{n}{\bar{d}} = \frac{\alpha}{\alpha - 1} ,$$

while we both quantities are $1 - \mathcal{O}(\frac{1}{n})$ for the Cayley graphs of the symmetric group with both transpositions and inversions as move sets. The scaled correlation length will play an important role for the estimates of the number of local optima discussed in sect 4.3.2.

4.2. Elementary Landscapes

Definition. Let Γ be a graph, Δ the Laplacian of Γ and f a landscape on Γ. We say f is *elementary* if

$$\Delta f + K'(f - \bar{f}) = 0 ,$$

where $\bar{f} = \langle f \rangle = 1/N \cdot \sum_{x\in\Gamma} f(x)$ is the average value of f, and K' is a constant.

Grover [57] observed that the landscapes of a number of classical combinatorial optimization problems are of this form, see Table 2. In order to keep the notation consistent with Grover's work for regular graphs we introduce $K = K'/D$ where D is as usual the vertex degree. The above relation for f is the discrete analogue of the Helmholtz equation, or reduced wave equation, [90]. An equivalent interpretation is that f is an eigenfunction of the Laplacian operator. It is not surprising that this property is closely related to the properties of the empirical autocorrelation function of the landscapes:

Theorem. Let Γ be a CGCG with vertex degree D. Then the following statements are equivalent:
(i) f is an elementary landscape with parameter K.
(ii) The empirical autocorrelation function $\hat{\rho}$ of f is a left eigenvector of the collapsed adjacency matrix $\hat{A}$ with eigenvalue $\Lambda = D\lambda = D(1-K) = D - K'$.
(iii) The empirical autocorrelation function $\hat{r}(s)$ of f measured along a simple random walk on Γ is $\hat{r}(s) = \lambda^s$.

Proof. Again we refer to [126]. ∎

Table 2. Elementary Landscapes.

<table>
<tr><th>Problem</th><th>Move Set</th><th>K</th><th>$1-\lambda$</th></tr>
<tr><td>NAES</td><td>Hamming</td><td colspan="2">$\frac{4}{n}$</td></tr>
<tr><td>p-spin</td><td>Hamming</td><td colspan="2">$\frac{2p}{n}$</td></tr>
<tr><td>WP</td><td>Hamming</td><td colspan="2">$\frac{4}{n}$</td></tr>
<tr><td>GC</td><td>Hamming</td><td colspan="2">$\frac{2\alpha}{(\alpha-1)n}$</td></tr>
<tr><td>GBP</td><td>Exchange</td><td>$\frac{8}{n}-\frac{8}{n^2}$</td><td>$\frac{8}{n}-\frac{8}{n^2}$</td></tr>
<tr><td>symmetric TSP</td><td>Transposition</td><td>$\frac{4}{n}$</td><td>$\sim\frac{4}{n}$</td></tr>
<tr><td></td><td>Inversions</td><td>$\frac{2}{n-1}$</td><td>$\sim\frac{2}{n}$</td></tr>
<tr><td>GMP</td><td>Transposition</td><td>$\frac{4}{n}$</td><td>$\sim\frac{4}{n}$</td></tr>
</table>

The configuration space of the GBP is the Johnson graph $\mathcal{J}(n, n/2)$.
The values K for NAES (Non-All-Equal-Satisfyability), WP (Weight Partition), GC (Graph Coloring), GBP (Graph Bipartitioning), and TSP (Traveling Salesman Problem) are taken from [57]. The value of K for the GMP (Graph Matching Problem) is derived in [126]. The values of λ for the GBP and the GMP problem are taken from [125] and [121], respectively; the remaining values are from [47]. The optimization mentionened above are discussed in detail in section 4.4.

This theorem is a central result in the theory of landscapes. While we have a complete proof only for the special case of Cayley graphs of commutative groups, there is ample evidence that it holds for a much broader class of configuration spaces. In particular it holds true whenever $\rho_i = \omega_i$ for all eigenvalues Λ_i.

A related result linking elementary landscapes to properties of their correlation functions on a larger class of graphs is given below: Let us call $r(1)$ the *empirical nearest neighbor correlation* of the landscape. If Γ is symmetric, that is, if all pairs of neighbors belong to the same orbit, then $\hat{r}(1) = \hat{\rho}(\mathbf{1})$.

Lemma. Let f be an elementary landscape on a symmetric graph Γ. Then $\hat{r}(1) = 1 - K$, with $-DK$ being the eigenvalue of the graph Laplacian corresponding to the landscape.

Proof. See [126, Lemma 12]. ∎

4.3. Local Optima

4.3.1. Local Optima and the Eigenstructure of the Laplacian

Local optima are the very feature of a landscape that makes it rugged. An understanding of the distribution of local optima is thus of utmost importance for the understanding of a landscape. The eigenstructure of the graph Laplacian offers a promising formalism for relating the correlation structure of a landscape to the geometry of its local optima. In this section we will briefly describe a few results along these lines.

The geometric structure of an elementary landscape is closely related to the constant K', i.e., the eigenvalue of the graph Laplacian to which f belongs. The solutions of the *Laplace equation* $\Delta f = 0$ form the *harmonic functions* on Γ. It is well known that there are no non-trivial harmonic functions on finite connected graphs (see, e.g., [117]. The harmonic functions correspond to the flat landscapes.

Let f be an arbitrary landscape on Γ. Let

$$V_+ = \{x \in V | f(x) \geq 0\} \qquad \text{and} \qquad V_- = \{x \in V | f(x) \leq 0\}$$

be the set of all vertices on which f is non-negative or non-positive, respectively. Let Γ_+ and Γ_- be the corresponding induced subgraphs of Γ. The connected components of Γ_+ and Γ_- are called the *nodal domains* of f. Clearly the geometry of the nodal domains is a very important characteristic of the landscape f.

The second largest eigenvalue $\tilde{\Lambda}_1$ of the graph Laplacian, which is negative for all connected graphs, is often called the *algebraic connectivity* of Γ. Theorem 2.5.7 in [22] states that if f is an eigenvector of Δ with eigenvalue $\tilde{\Lambda}_1$, then both Γ_+ and Γ_- are connected, i.e., there are exactly two nodal domains. Following a suggestion by Kauffman we term this type of landscapes "Fujijama" landscapes as they consist of only a single "mountain" Γ_+. On a Boolean hypercube we have a much stronger result: An elementary landscape with eigenvalue $\tilde{\Lambda}_1$ is of the form

$$f(\sigma) = J_0 + \sum_{k=1}^{n} J_k \sigma_k ,$$

and thus there is a unique local maximum $\sigma_{\max}$ and a unique local minimum $\sigma_{\min}$, provided the coefficients J_k are all non-zero. Note also $\sigma_{\min} = -\sigma_{\max}$ in this case. It is unknown if a similar result holds on other graphs as well.

For Riemannian manifolds even more is known. *Courant's Nodal Domain Theorem* states that if all eigenvalues of the Laplacian on a Riemannian manifold are ordered as $0 = \tilde{\Lambda}_1 \geq \tilde{\Lambda}_2 \geq \ldots$, then the number of nodal domains of an eigenfunction θ_k belonging to $\tilde{\Lambda}_k$ is at most k; see, e.g., the book by Chavel [25]. We expect that an analogous theorem holds on graphs. Results along these lines will be presented elsewhere.

Closely related to these observations is theorem 6 of Grover [57]. He showed that the local optima have a characteristic distribution on elementary landscapes: Let $z_{\min}$ and $z_{\max}$ be a local minimum and a local maximum, respectively. Then

$$f(z_{\min}) \leq \bar{f} \leq f(z_{\max}),$$

i.e., if f is an eigenvector of Δ then all local maxima are in Γ_+ and all local minima are in Γ_-. Grover also gives bounds on the length of adaptive walks for elementary landscapes.

4.3.2. The Number of Local Optima

One of the most important characteristics of a landscape is the number $\mathcal{N}$ of local optima. Palmer [97] used an at-least-exponetial increase of $\mathcal{N}$ with the system size n as definition of "rugged" landscape. Local optima are obstacles for heuristic optimization algorithms. It is a very resonable conjecture, therefore, that optimization by a general algorithm should be easier if there are fewer local optima, or probably equivalently, if the nearest neighbor correlation is larger.

It is easy to calculate the number of local optima in two extreme cases: the random energy model, and the completely correlated landscape (in which fitness is added over independent contributions from each element in a sequence). In the completely correlated case, at least in the case of the Boolean hypercube, there is one single optimum. In a random energy model on graph with vertex degree D with N vertices there are on average

$$\mathcal{E}[\mathcal{N}] = \frac{1}{D+1}N$$

local optima. An extension of these results to landscape with a small correlations can be found in [127].

Estimates for the number of local optima are also known for the Nk-model [142, 141, 70], Sherrington-Kirkpatrick model and related short-range 2-spin models [19, 20, 28, 33, 134] for the Travelling Salesman Problem with Transpostion metric [127] and for RNA free energy landscapes [47]. However, there is no routine method for deriving the number of local optima and for relating this number to the corrlation structure of the landscape. In most cases, the number of local increases exponentially with the system size n.

Let us first consider landscapes and random fields on Hamming graphs. Here we have $N = n^{\alpha}$. In statistical mechanics on considers traditionally the parameters

$$A \stackrel{\text{def}}{=} \lim_{n\to\infty} \frac{1}{n} \ln \mathcal{E}[\mathcal{N}] \qquad \text{and} \qquad A^* \stackrel{\text{def}}{=} \lim_{n\to\infty} \frac{1}{n}\mathcal{E}[\ln \mathcal{N}].$$

It will be convenient to introduce the quantity $P \stackrel{\text{def}}{=} -\lim_{n\to\infty} \frac{1}{n} \ln p_{lo}$, which is related to A via $P = \ln\alpha - A$. Note that we have $\mathcal{N} \approx A^n$ and $p_{lo} \approx P^{-n}$ for large enough systems.

Weinberger [142] gives the following estimate for the probability $p_{lo} = \mathcal{N}/N$ of finding a local optimum in an Nk model for medium and large values of k:

$$p_{lo,Nk} \approx [(\kappa-1)(k+1)]^{-\frac{n}{k+1}}.$$

In terms of the scaled correlation length $\xi \stackrel{\text{def}}{=} \ell/n$ this equation becomes

$$P \approx \xi[\ln(\alpha-1) - \ln\xi].$$

An interesting variation of the Nk model has been introduced recently by Perelson and Macken [98] for modelling the affinity maturation of antibodies by somatic hypermutations. In the *Block Model* a sequence x over an alphabet with α letter is devided into B independent blocks of size B/N^6. The fitness $f(x)$ is defined as the sum of the fitnesses of the individual blocks, which in

[6] In [98] the blocks are not required to have equal size. We impose this restriction here in order to simplify the formalism.

turn are taken to be i.i.d. random numbers drawn from some distribution with finite variance. The nearest neighbor correlation is given by $r(1) = 1 - 1/B$, implying that the correlation length is approximately $\ell \approx B$ provided $B \gg 1$. The interesting feature of the block model is that the number of local optima can be easily estimated [98]. One obtains

$$A \approx \frac{1}{n} \ln \mathcal{E}[\mathcal{N}] \approx \ln \alpha - \frac{B}{n} \ln \Big((\alpha - 1) \frac{n}{B} \Big)$$

and, using $\xi \approx B/n$ for not too few blocks, this translates to

$$A \approx \ln \alpha - \xi [\ln(\alpha - 1) - \ln \xi] \,,$$

and consequently, the probability for hitting a local optimum at random fulfils

$$P \approx \xi [\ln(\alpha - 1) - \ln \xi] \,.$$

This estimate coincides with Weinberger's [142] result for the Nk models.

Using the TAP approach [135] one can estimate the quantities A and A^*. Not unexpectedly, short-rangle spin glasses have more local optima than longe range spin glasses, even the Hamiltonian is of the same form

$$\mathcal{H}(\sigma) = \sum_{(i,j)} J_{ij} \sigma_i \sigma_j$$

where the sum is taken over all pairs (i, j) of *neighboring* spins [14]. Bray and Moore [20] show that to a first approximation the number of local optima depends on the coordination number z of the spin lattice (i.e., $z = 2$ for a chain of spins, $z = 4$ for a square lattice of Ising spins, and $z = n - 1$ for the SK model.)

$$A(z) \approx A^*(z) \approx 0.1992 + 0.0565/z$$

Exact expressions are available for the 1D case, the linear spin chain:

$$A_{1D} = 4/\pi \approx 0.2415 \qquad \text{and} \qquad A^*_{1D} = \ln 2/3 \approx 0.2310.$$

These z-expansion is surprisingly accurate: $A(2) \approx 0.2320$. Note that all these 2-spin models have the same empirical correlation function $\hat{\rho}_2$ and thus also the same correlation length $\xi = 1/4$. The numerical values for A from different models, estimated with different approaches are collected in table 3.

Table 3. Estimated values of A for landscapes with $\xi = 1/4$ on Boolean hypercubes.

Model	Method	A	Ref
2-spin SK	TAP	0.1992	[20]
2-spin chain		0.2410	[33]
Nk $k = 3$	direct	0.3466	[142]
Block Model	direct	0.3466	[98]
-	*	0.1308	

The last line is obtained by assuming one local optima in a patch with radius given by the correlation length.

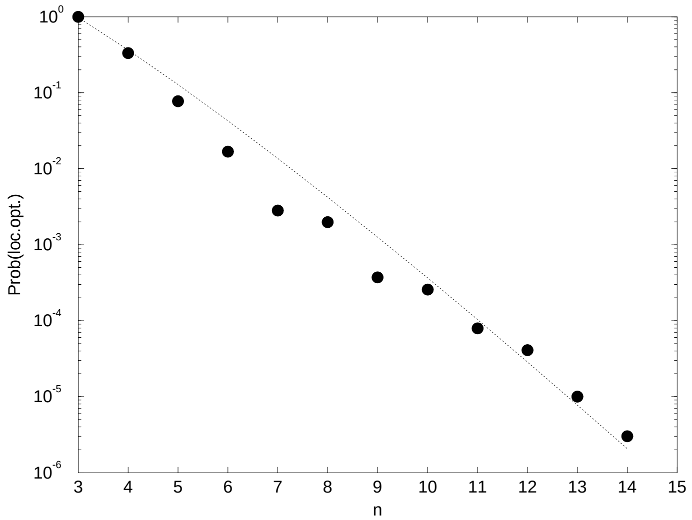


Figure 10: Number of local optima of a TSP, with neighborhood defined by transpositions, as a function of system size. The dotted line corresponds to one local optimum in a patch with a radius of one correlation length.

A rough estimate for the number of local optima on landscapes with a Gaussian distribution of values has been proposed in [127]. One concludes immediately from the discussion in sect. 3.7.2 that an assumption on the distribution function is in fact necessary. In a fully correlated landscape, i.e., correlation length $\ell \approx \max d(x,y)$ one expects to find a single optimum. Let $B(r)$ denote the number of vertices in a neighbourhood with radius r around an arbitrary configuration. We expect then, that there are roughly $\mathcal{O}(1)$ local optima in a patch with radius ℓ, i.e., $p_{lo} \approx 1/B(\ell)$. For Hamming graphs this estimate becomes explicitly [109]

$$P = \xi[\ln(\alpha-1) - \ln\xi] - (1-\xi)\ln(1-\xi).$$

It deviates from the above estimates by the term $(1-\xi)\ln(1-\xi)$. See also table 3 for a comparison of numerical values for landcsapes with $\xi = 1/4$ on a Boolean hypercube.

The parameters A and P are not useful in cases where the number and/or the probability of local optima is not exponential. The TSP with transposition metric provides such an example. For landscapes on $\Gamma(S_n, \mathcal{T})$ one estimates

$$p_{lo} \approx \frac{1}{\ell!}\left(\frac{e^2}{2\xi^2}\right)^{-\ell} \approx \frac{1}{\Gamma(n\xi)}\left[(e^2/2\xi^2)^{\xi}\right]^{-n} ,$$

see [83]. For the TSP with transpositions we have $\xi = 1/4$, i.e.,

$$p_{lo} \approx \text{const}\frac{2.773^{-n}}{(n/4)!}$$

This is in very good agreement with the numerical data [127], see also figure 10.

A rigorous theory linking the correlation structure of Gaussian landscape with the number $\mathcal{N}$ of local optima is still missing. The example of the 2-spin models above shows, furthermore, that there cannot be a simple equation linking the correlation length with the number of local optima.

On the other hand, the fairly small numerical range of A for various 2-spin models leaves at least room for rough estimates.

The distribution of local optima cannot be studied directly for large values of n since p_{lo} decreases exponentially in most cases. Some information of local optima, and on the distribution their basins of attraction can be obtained from adaptive and gradient walks, see, e.g., [70]. A statement closely related to the crude estimate for $\mathcal{N}$ is that the length of a gradient walk should be approximately equal to the correlation length, while adaptive walks should be longer by a factor that depends on the structure of the configuration space only. Numerical evidence for the Nk model and RNA landscapes is consistent with this statement.

The statistics of walks and local optima of the random energy model is known in detail [72, 81, 4, 42]. This landscape is qualitatively different from the correlated landscapes discussed above in that the walk lengths are $\mathcal{O}(\log n)$ as opposed to $\mathcal{O}(n)$, and the probaility of hitting a local optimum at random is $p_{\text{l.o.}} = 1/(1+D)$, as apposed to exponentially decreasing. Here D is, as usual, the vertex degree of the a configuration, i.e., the number of nearest neighbors.

4.4. Some Example Landscapes

4.4.1. Traveling Salesman Problems

The Traveling Salesman Problem `TSP` has already been discussed in the introductory section 2.1. Here we will focus on a few more formal aspects. Recall that the cost of tour π is given by

$$f(\pi) = \sum_j w_{\pi(j)\pi(j-1)}\,,$$

where the indices are interpreted modulo n. An arbitrary matrix W can be uniquely decomposed into its symmetric component $W^\sigma = (W + W^+)/2$ and its antisymmetric component $W^\alpha = (W - W^+)/2$; it will be convenient to define

$$\begin{aligned} f^\sigma(\pi) &\stackrel{\text{def}}{=} \sum_j w^\sigma_{\pi(j)\pi(j-1)} = \frac{f(\pi)+f^*(\pi)}{2}, \\ f^\alpha(\pi) &\stackrel{\text{def}}{=} \sum_j w^\alpha_{\pi(j)\pi(j-1)} = \frac{f(\pi)-f^*(\pi)}{2}. \end{aligned}$$

Note that f^σ and f^α can be viewed as cost functions of TSPs with "distance matrices" W^σ and W^α, respectively.

Theorem 13. Both f^σ and f^α are elementary landscapes on the Cayley graphs of the symmetric group with the transpositions and the inversions as generators, respectively. In particular we have for transpositions

$$\Delta f^\sigma + 2(n-1)(f^\sigma - \bar{f}) = 0 \qquad\qquad \Delta f^\alpha + 2nf^\alpha = 0\,,$$

and for inversions (2opt moves, Lin & Kernighan 1965) we find

$$\Delta f^\sigma + n(f^\sigma - \bar{f}) = 0 \qquad\qquad \Delta f^\alpha + \frac{n(n+1)}{2} f^\alpha = 0.$$

The landscape of a TSP with transpositions or inversions is elementary if and only if W is either symmetric or antisymmetric.

Proof. The rather tedious computations which are based on [57] can be found in [126, Thm.13]. ∎

Asymmetric TSPs hence provide an example of fairly simple composite landscapes. They consist of two modes corresponding to the symmetric and the antisymmetric part of the distance matrix W, respectively. It is also interesting to note in this context that canonical transpositions $(i, i+1)$ do not lead to an elementary landscape.

Consequently, the nearest neighbor correlations of the symmetric and antisymmetric components of a TSP with transpositions are $r(1) = 1 - 4/n$ and $r(1) = 1 - 4/(n-1)$, respectively, i.e., very similar. In fact, numerical estimates [121] had been consistent with $r(1) \sim 1 - 4/n$ for large n in both cases. In the case of inversions we have a symmetric mode with nearest neighbor correlation $r(1) = 1 - 2/(n-1)$ and an antisymmetric contribution with a vanishingly small contribution $r(1) = -2/(n-1) \sim 0$.

It is interesting to correlate these values of $r(1)$ with known facts about the performance of heuristic optimization algorithm, in particular with the simulated annealing. It has been observed by serveral authors that simulated annealing on symmetric TSP is much more effective when reversals instead of transpositions are used as move set, see, e.g., the books by Aarts and Korst [1] or Otten and vanGinneken [96]. Furthermore, Miller and Pekny [89] have observed that reversals are a remarkably bad move set for asymmetric TSPs. These observations are in accordance with the conjecture that landscape with smoother correlation functions have fewer local optima and are thus easier to optimize on, *cp.* [121]. In particular the difference between symmetric and asymmetric TSPs when reversals are used is easily explained in these terms: while for the symmetric TSP the landscape is as smooth as possible, it is completely rugged for the antisymmetric case.

Numerical data by Stadler and Schnabl [121] indicate that the correlation functions $\hat{r}(s)$ of both the symmetric and the antisymmetric components should be of the form $r(1)^s$, we do not, however, have a proof for this conjecture for all n.

4.4.2. Graph Bipartitioning Problem

Consider a graph G with an even number n of vertices and a symmetric matrix H of edge weights. A configuration is a partition of the vertex set into two subsets A and B of equal size. The cost function is

$$f([A;B]) = \sum_{i\in A}\sum_{j\in B} H_{ij},$$

i.e., the total weight of edges connecting the two subsets. Usually one is interested in minimizing the total edge weight $f([A;B])$ connecting the two subsets A and B, see e.g., [49]. As a close relative of the Sherrington-Kirkpatrick model the *Graph Bipartitioning Problem*, `GBP`, has received considerable attention.

The usual move set for this problem is exchanging a vertex in A with a vertex in B. The resulting graph is the Johnson graph $\mathcal{J}(n, n/2)$, see sect. 2.6.3. The random field associated with the `GBP` is constructed by assigning the edge weights $H_{ij} = H_{ji}$ independently and identically distributed.

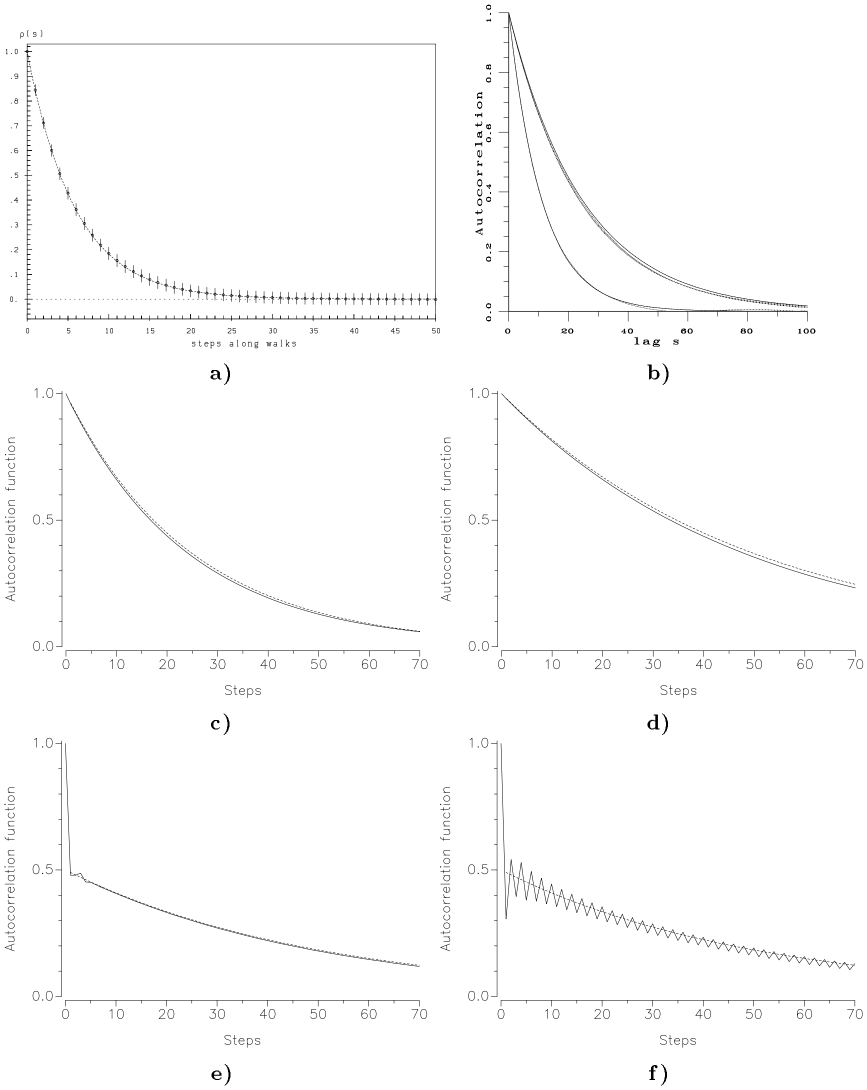


Figure 11: Examples of autocorrrelation functions.
a) GBP, b) graph matching (upper curves) and low autocorrelated binary string problem, c) TSP with transposition metric, d) symmetric TSP with inversions, e) asymmetric TSP with inversions, f) asymmetric TSP with a restricted set of inversions as move set.

Stadler and Happel [125] have shown by explicit calculations that the correlation functions of this random field are

$$\begin{aligned} \rho(d) &= 1 - 8\frac{d}{n}\left(1 - \frac{1}{n-2}\right) + 16\left(\frac{d}{n}\right)^2\left(1 - \frac{1}{n-2}\right) \\ r(s) &= \left(1 - \frac{8}{n} + \frac{8}{n^2}\right)^s , \end{aligned}$$

irrespective of the particular distribution from which the H_{ij}'s are drawn. Our results imply that $\rho(d)$ is an eigenvector with eigenvalue $r(1)$ of the Johnson graph $\mathcal{J}(n, n/2)$.

Grover [57] found that each instance of the GBP is an elementary landscape with $K = 8/n - 8/n^2$, see also table 3. Hence we find in fact the expected result $K = 1 - \lambda$. We take this as one further hint that the theorems proved in the previous sections for CGCG are in fact true for a much larger class of configuration spaces.

4.4.3. The Low Autocorrelated Binary String Problem

The `LABSP` [55, 8] consists of finding binary strings σ over the alphabet $\{-1,+1\}$ with low aperiodic off-peak autocorrelation $R_k(\sigma) = \sum_{i=1}^{N-k} \sigma_i\sigma_{i+k}$ for all lags k. These strings have technical applications such as the synchronization in digital communication systems and the modulation of radar pulses. The quality of a string σ is measured by the fitness function

$$f(\sigma) = \sum_{k=1}^{n-1} R_k(\sigma)^2.$$

In most of the literature on the `LABSP` the *merit factor* $F(\sigma) = n^2/(2f(\sigma))$ is used. This transformation gives rise to a non-Gaussian distribution of merit factors, see [8] for details.

It can be shown [126, Lemma 16] that the landscape f of the `LABSP` can be written as

$$f(\sigma) = a_0 + \sum_{k=1}^{\lceil\frac{n}{2}\rceil-1} \sum_{i=1}^{n-2k} 2\varepsilon_{i,i+k}(\sigma) + \sum_{k=1}^{n-1}\sum_{i=1}^{n-1} \sum_{j\neq i-k,i,i+k} \varepsilon_{i,i+k,j,j+k}(\sigma)\,,$$

and thus the empirical autocorrelation functions of the `LABS` are

$$\begin{aligned}
\hat\rho(d) &= [1-\mathcal{O}(1/n)]\rho_4(d) + \mathcal{O}(1/n)\rho_2(d),\\
\hat r(s) &= [1-\mathcal{O}(1/n)]\left(1-\frac{8d}{n}\right)^s + \mathcal{O}(1/n)\left(1-\frac{4d}{n}\right)^s.
\end{aligned}$$

The landscape of the `LABPS` is thus not elementary, it consists of a superposition of two modes, namely $p=2$ and $p=4$. The smoother $p=2$ contribution becomes negligible for large n, so that f behaves for long strings almost like an elementary $p=4$ landscape. This fact explains why the `LABPS` has been found to be much harder for simulated annealing than, say, the SK spin glass [8]. The empirical autocorrelation function $\hat r(s)$ has been computed numerically [121] based on the merit factor F. The numerical estimate for the correlation length

$$\ell \stackrel{\text{def}}{=} -\frac{1}{\ln r(1)} \approx 0.1230\cdot n - 0.983$$

is in excellent agreement with the asymptotic value $\ell = n/8 + \mathcal{O}(1)$ obtained above.

4.4.4. Not-All-Equal-Satisfiability

Consider a vector of n binary variable. A literal is a variable or its complement. A clause is a set of three literals that does not contain both a variable and its complement. A clause is said to be satisfied if at least on literal is 0 and at least one literal is 1. An instance of the *Not-All-Equal-Satisfiability Problem* (`NAES`) is given by a set of c clauses; the cost function is the number of non-satisfied clauses. The move set is defined by flipping the value of a single variable, thus the configuration space is the Boolean Hypercube. It is shown in [57] that `NAES` has an elementary landscape, see also table 3.

4.4.5. Graph Coloring

An instance of a graph coloring problem GC consists of a graph $G(V, E)$ and a number α of colors. A configuration x is a α-coloring of the vertex set V, i.e., an assignment of one color $x(p)$ to each vertex p of the graph. The cost function is the number edges $(p, q) \in E$ for which both incident vertices p and q have the same color:

$$f(x) = \sum_{(p,q)\in E} \delta_{x(p)x(q)}.$$

A move is the replacement of one color by another one at single vertex. The configuration spaces are thus the General Hamming graphs, i.e., sequence spaces over the alphabet of the α colors. The landscape of a GC is elementary [57], see table 3.

4.4.6. Weight Partition

Given a string of n "spins" $x = (x_i) \in \{-1, +1\}^n$ and corresponding weights w_i, the cost function is given by

$$f(x) = \left(\sum_{i=1}^{n} w_i x_i\right)^2 .$$

The move set is given by flipping a single spin, hence the configuration space is again a hypercube. The landscape of a Weight Partition Problem WP is elementary [57], see table 3.

4.4.7. Graph Matching Problem

Given a graph G with n vertices and a symmetric matrix W of edge weights, the task of the Graph Matching Problem GMP is to partition the graph into $n/2$ pairs of vertices such that the sum of the edge weights corresponding to these pairs is optimal. A convenient encoding of the problem is the following. Let $\pi \in S_n$ be a permutation of the vertices. We assume that the vertices are arranged such that $[\pi(2k-1), \pi(2k)]$ form a pair. The cost function is hence

$$f(\pi) = \sum_{k=1}^{n/2} W_{\pi(2k-1),\pi(2k)}.$$

Again, the configuration space is the symmetric group, and hence the set of all transpositions forms a canonical move set. The disadvantage of this encoding is, however, a very large degree of redundancy. In fact, $(n/2)!2^{n/2}$ permutations represent a single matching [74]. As the models discussed above the landscape of the GMP is elementary, see [126, Thm.14] and table 3. Numerical estimates of correlation functions can be found in [121].

4.5. Classification of Landscapes

4.5.1. Fractal Landscapes

Gregory Sorkin [118] proposed the following definition of a fractal fitness landscape which is built on the idea of fractional Brownian motion [137].

Definition. A landscape is *fractal* if $\langle ||f(x) - f(y)||^2 \rangle \propto d^{2h}(x,y)$, where the two configurations x and y with normally distributed fitnesses $f(x)$ and $f(y)$, respectively, are separated by a distance $d(x,y)$.

As a practical matter, the distance between configurations for any finite configuration space is bounded by the diameter of the configuration space diam $\mathcal{C}$, so that Sorkin's definition makes sense only for distances small compared to this upper bound. Sorkin's definition can be reformulated in terms of autocorrelation functions:

$$1 - \hat{\rho}(d) = (1 - r(1))d^{2h}.$$

For simplicity we will assume in this section that the autocorrelation functions are given in terms of distances, otherwise we may work with the corresponding random walk version. The details of this step are explained in the following section 4.7. For distances d small compared to the correlation length ℓ, almost all of the landscapes listed in table 4, as well as most of the RNA landscapes described in section 5, are of the form

$$\rho(d) = 1 - \frac{d}{\ell} + \dots,$$

and therefore approximately satisfy Sorkin's definition with $h = 1/2$.

4.5.2. Classification by the "Form" of the Correlation Functions

Since strictly speaking all landscapes are finite object a classification beyond the distinction elementary/non-elementary seems to be hard to achieve. Instead of considering a single landscape, we assume therefore that we are given a family $\{f_n\}$ of landscapes (or random fields) on a graphs Γ_n. The index $n = 1, 2, \dots$ is a "natural" measure of the size of the configuration space. For sequence spaces, for instance, n is simply the chain length. Now consider the correlation functions $\rho_{[n]}(d)$ and $r_{[n]}(s)$ of the landscape f_n. We define the scaled autocorrelation functions $\zeta(h)$ and $z(h)$ for all rational numbers h by

$$\zeta(h) = \lim_{n\to\infty} \rho_{[n]}(nh) \qquad\qquad z(h) = \lim_{n\to\infty} r_{[n]}(nh)\,,$$

respectively. Note that ζ and z do not necessarily exist! An example is the family of p-spin models with $p = n$ for which we find $\rho_{[n]}(n/2) = \pm 1$ depending on whether $n \equiv 0 \bmod 4$ or $n \equiv 2 \bmod 4$.

Based on the "shape" of $\zeta(h)$ or $z(h)$, which the same, Weinberger and Stadler [139] proposed a classification of landscapes. This work was motivated by the analogy to the classification of stationary continuous-time stochastic processes in terms of their autocorrelation functions.

I $\zeta(h)$ is discontinuous for $h = 0$. These landscapes are *extremely rugged.* (A continous time stochastic process with an autocorrelation function of this type is not even continuous.) Two subclasses can be distinguished:

Table 4. Combinatorial optimization problems and their autocorrelation functions.

Name	Metric	$\rho(d)$	$r(s)$	ℓ	diamΓ	Ref.
REM	any	$\delta_{0,d}$	see 3.6.2	0	n	–
s.TSP	Tr.	?	$\sim e^{-4s/n}$	$n/4$	$n-1$	[127]
	2opt	?	$\sim e^{-2s/n}$	$n/2$	$n/2 \ldots n-1$	[127]
	c.Tr.	?	$\sim e^{-2s/n}$	$n/2$	$\frac{n(n-1)}{2}$	[139]
a.TSP	Tr.	?	$\sim e^{-4s/n}$	$n/4$	$n-1$	[127]
	2opt	?	$\sim \frac{1}{2}(\delta_{0,s}+e^{-2s/n})$	–	–	[127]
	c.Tr.	?	$\sim e^{-3s/n}$	$n/3$	$\frac{n(n-1)}{2}$	[139]
GM	Tr.	?	$\sim e^{-4s/n}$	$n/4$	$n-1$	[121]
GBP	Ex.	$1-\frac{n-1}{n-2}[8\frac{d}{n}-16(\frac{d}{n})^2]$	$(1-\frac{8}{n}+\frac{8}{n^2})^s$	$(n-3)/8$	$n/2$	[125]
LAS	Ham.	?	$\sim e^{-10s/n}$	$n/10$	n	[121]
rnd.Nk	Ham.	$(1-\frac{d}{n})(1-\frac{k}{n-1})^d$	$\sim(1-\frac{k+1}{n})^s$	$n/(k+1)$	n	[142,47]
p.r.Nk	Ham.	$(1-\frac{k+1}{n})^d$	$\sim(1-\frac{k+1}{n})^s$	$n/(k+1)$	n	
adj.Nk	Ham.	see 3.6.3	$\sim(1-\frac{k+1}{n})^s$	$n/(k+1)$	n	[142,47]
p-Spin	Ham.	$1-\frac{2}{\binom{n}{p}}\sum{}'\binom{d}{j}\binom{n-d}{p-j}$	$\sim e^{-2ps/n}$	$n/(2p)$	n	[139]
SK	Ham.	$1-\frac{n}{n-1}[4\frac{d}{n}-4(\frac{d}{n})^2]$	$(1-\frac{4}{n})^s$	$n/4$	n	[139]

$\sum'$ for the p-spin model denotes the sum over all odd j subject to the restriction $j>\min(d,p)$.
REM is Derrida's random energy model; s.TSP and a.TSP denote symmetric and asymmetric travelling salesman problems [77], GM is the graph matching problem [87]. The corresponding metrics are transpositions (Tr.), 2opt moves and canonical transpositions (c.Tr). GBP is the graph bipartitioning problem [49], its metric (Exc) is derived from exchanging a pair of objects. LAS stands for the low autocorrelated string problem [55, 8]. The Sherrington-Kirkpatrick spinglass [116] is the special case $p=2$ of the p-Spin model [31, 30, 32] introduced in [2] as a model for a rugged landscape in evolutionary optimization. The abbreviations rnd.Nk, p.r.Nk, and adj.Nk refer to random neighbour, purely random and adjecent neighbour Nk-model, resp. Here the canonical metric is the Hamming metric.

Ia $\zeta(h) = \delta(h)$ (or more generally consists of a finite sum of delta functions). This subclass may be divided further depending on the behaviour of the unscaled correlation function:

- The unscaled correlation length ℓ diverges when diam $\mathcal{C} \to \infty$. An example of this subclass is the TSP with canonical transpositions defining the neighborhood. The correlation length is $\mathcal{O}(n)$, whereas the diameter of the landscape is $\mathcal{O}(n^2)$ where n is the number of cities; see table 4.
- The unscaled correlation length ℓ is uniformly bounded from above. Examples of this subclass are the random energy model and (probably) certain RNA landscapes [43].

Ib $\zeta(h)$ does not vanish in a neighbourhood of $h = 0$ but the autocorrelation function is bounded from above by some constant $\gamma < 1$ for $h \neq 0$ in a neighborhood of $h = 0$, i.e.,

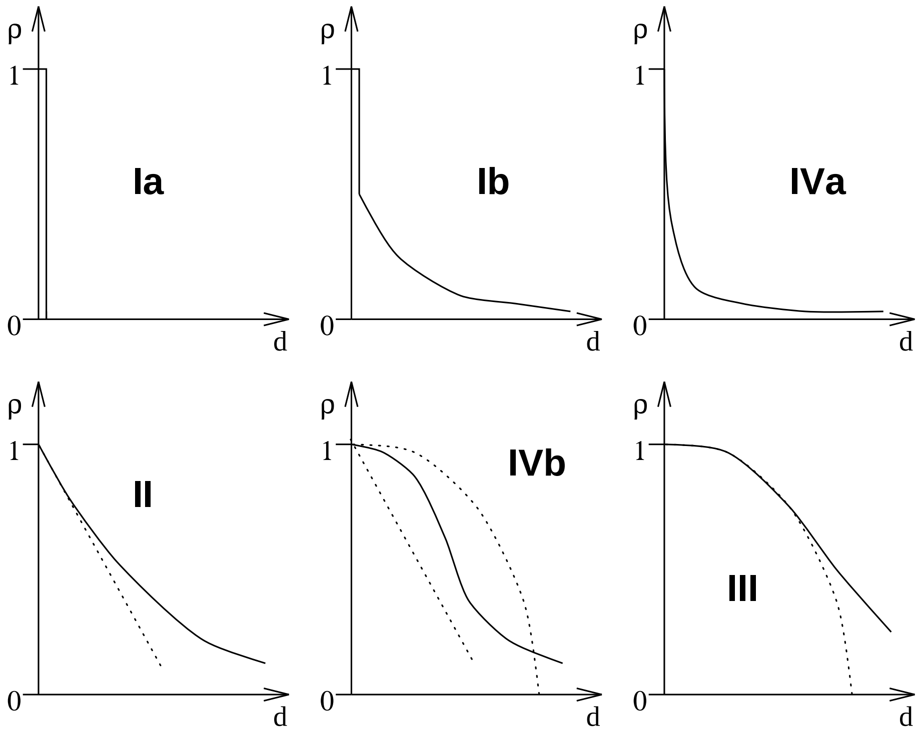


Figure 12: Classification of landscapes (and random fields) according to the "shape" of their correlation functions. For details see text.

there is a finite size jump at $h = 0$, but the landscape is not uncorrelated. An example is provided by the asymmetric TSP with 2-opt moves, see figure 11.

II $\zeta(h) = 1-\alpha h+\mathcal{O}(h^2)$, with a constant α of order 1. Examples are almost all entries in table 4. Class **II** landscapes are "trivial" random fractals with $h = \frac{1}{2}$. Continuous-time stochastic processes with an autocorrelation function of this type are continous but not differentiable; hence these landscapes are "rugged".

III $\zeta(h) = 1 - \beta h^2 + o(h^2)$, with some constant β of order 1. In contrast to all previous cases the autocorrelation function is differentiable at $h = 0$. Correlation functions of this type belong to differentiable stochastic processes in the continuous-time case. These landscapes form the second class of trivial fractals, corresponding formally to Sorkin's definition of fractals with $h = 1$. These are the smoothest possible landscapes, since any function $f(x) = 1 - \mathcal{O}(x^{2+q})$ with $q > 0$ cannot be autocorrelation function of a stochastic process.

IV $\zeta(h) = 1 - c_p h^p + o(h^p)$, with $p \neq 1, 2$, correspond to non-trivial random fractals. Such landscapes have been constructed by Sorkin on fractal configuration spaces, but have not (yet?) been found in "real life" examples.

4.5.3. Classification Based on the Characterization of Correlation Function

The characterization of autocorrelation functions has direct implications on the *classification* of autocorrelation function. We will now relate the classes defined in the previous subsection with the spectral decomposition of the autocorrelation function described in section 4.1.

On the sequence spaces we can investigate the behavior of $\zeta(h)$ for small h explicitly. Consider

$$\zeta(\frac{1}{n}) = 1 - \frac{1}{n}\frac{\alpha}{\alpha-1}\sum_{p=0}\Omega_p \cdot p =: 1 - \frac{1}{n}\frac{\alpha}{\alpha-1}\bar{\pi}\,.$$

The parameter $\bar{\pi}$ can be interpreted as the "average mode" of the landscape. We expect similar relations on other configuration spaces as well. The form of ζ clearly depends on the scaling of $\bar{\pi}$ with the system size parameter (chain length) n. First we note that $1 \le \bar{\pi} \le n$ holds for all landscapes on sequences spaces. If $\bar{\pi} = \mathcal{O}(n)$ then $\zeta(\frac{1}{n})$ does not converge to 1 and hence ζ cannot be not continuous. The initial slope of ζ can be estimated as $\partial\zeta/\partial h|_{0+} = -\frac{\alpha}{\alpha-1}\bar{\pi}$, if it exists at all. Hence we have an asymptotically finite slope only if $\bar{\pi} = \mathcal{O}(1)$.

Theorem. The smoothest landscape on a sequence space has autocorrelation function

$$\rho_1(d) = 1 - \frac{\alpha}{\alpha-1}\frac{d}{n}.$$

Proof. See [126]. ∎

Table 5. Tentative Classification of Landscapes.

	Class I	rugged "Fractals"	Class II	smooth "Fractals"	Class III
ζ	not continuous	$\zeta = 1 - c\lvert h\rvert^\gamma$ $0<\gamma<1$	$\zeta = 1 - c\lvert h\rvert$	$\zeta = 1 - c\lvert h\rvert^\gamma$ $1<\gamma<2$	$\zeta = 1 - c\lvert h\rvert^2$
$\bar{\pi}$	$\bar{\pi} = \mathcal{O}(n)$	$\bar{\pi} \sim o(n)$, $\bar{\pi} \to \infty$	$\bar{\pi} = \mathcal{O}(1)$	do not exist on $\mathcal{Q}_\alpha^n$	
	Class A	Class B	Class C		

This reasoning suggests that there should be three different classes of landscapes:

A $\bar{\pi} = \mathcal{O}(n)$ implies $\lim_{n\to\infty}\zeta(\frac{1}{n}) < 1$ and hence the asymptotic autocorrelation function is discontinuous at the origin.

B $\bar{\pi} \to \infty$ but $\bar{\pi} = o(n)$ allows ζ to be continuous with infinite initial slope.

C $\bar{\pi} \sim \mathcal{O}(1)$ implies an autocorrelation function which decays continuously with finite slope.

We do *not* claim that there is one-to-one correspondence between the two classification schemes, which are compared in table 5. One of the reasons for this is that we have no general conditions for the existence of ζ.

4.6. Landscapes on Irregular Graphs

4.6.1. Correlation Functions

Let us use the notation $D_h(x) \stackrel{\text{def}}{=} |N_h(x)|$ for the distance degrees, i.e., for the number of vertices in distance h from a vertex x. Suppose Γ be a graph with N vertices and M_h unordered pairs of vertices with mutual distance h. Of course M_1 is the number of edges. Interpreting D_h as a landscape we obtain immediately the balance equation

$$\sum_x D_h(x) = \langle D_h \rangle N = 2M_h .$$

On non-regular graphs we have essentially two ways of defining a correlation function.

Definition. Let $f : V\Gamma \to \mathbb{R}$ be a non-constant landscape on a connected graph Γ with $N > 1$ vertices. Then we define

$$\begin{aligned} \hat{\rho}(d) &\stackrel{\text{def}}{=} \frac{\langle f(x)f(y)\rangle_{d(x,y)=d} - \langle f\rangle^2}{\langle f^2\rangle - \langle f\rangle^2} \\ \hat{\rho}_*(d) &\stackrel{\text{def}}{=} 1 - \frac{\langle (f(x)-f(y))^2\rangle_{d(x,y)=d}}{\langle (f(x)-f(y))^2\rangle_{(x,y)\in V\times V}} \end{aligned}$$

Both definitions of correlations have their advantages in certain contexts. $\hat{\rho}$ is the usual definition which have already encountered for distance regular graphs. $\hat{\rho}_*$ on the other hand has the same flavour as the usual correlation coefficient. Since it depends only on squared differences it will form the starting point for a generalization of correlation measures to non-numeric functions in sect. 5.2., see also [47, 44]. In fact, two definitions coincide provided the configuration space is sufficiently symmetric:

Theorem. Let Γ be distance degree regular. Then $\hat{\rho} \equiv \hat{\rho}_*$

Proof. See [47]. ∎

Definition. Let f and g be two landscapes on Γ. We define the *covariance* of f and g by

$$\text{cov}[f,g] \stackrel{\text{def}}{=} \langle fg\rangle - \langle f\rangle\langle g\rangle .$$

Note that this conforms the usual definition of a covariance.

For each non-constant landscape on a connected graphs we define the function

$$\Psi(h) \stackrel{\text{def}}{=} \frac{\text{cov}[D_h,\varphi_f]}{\langle D_h\rangle\langle\varphi_f\rangle} \qquad \text{where} \qquad \varphi_f(x) \stackrel{\text{def}}{=} f^2(x) - \langle f\rangle^2 .$$

With this definition we can prove an interested generalization of the above relation between $\hat{\rho}$ and $\hat{\rho}_*$.

Theorem. $\rho_*(h) - \rho(h) = -\Psi(h)$ for all connected graphs and all non-constant landscapes.

Proof. The proof proceeds by a straight forward (although rather tedious) direct evaluation of both sides of the equation. We omit the details here. ∎

Note that $\Psi(0) = 0$ for all graphs, and $\Psi(1) = 0$ holds for all landscapes on regular graphs. We suspect that landscapes for which Ψ vanishes for all distances h will play a special role on irregular graphs. We propose to call such landscapes *uniform*. Note also that by replacing the landscape averages by expectation values one obtains an analogous formula relating the correlation function $\rho(d)$ on a random field and it "squared difference" counterpart $\rho_*(d)$.

4.6.2. An Upper Bound on the Nearest Neighbor Correlation

Theorem. Let $\tilde{\Lambda}_1$ be the second eigenvector of the Laplacian Δ of Γ, let f be an arbitrary landscape on Γ with correlation function $\hat{\rho}_*$. Then

$$\hat{\rho}_*(1) \leq 1 - \frac{1}{\langle D \rangle}\tilde{\Lambda}_1.$$

Proof. We start with an observation by Biggs [12, p.58], namely

$$\tilde{\Lambda}_1 \leq \sum_{\{x,y\}\in E} (f(x) - f(y))^2 \Big/ \sum_{x\in V\Gamma} f^2(x)$$

This can be rewritten as $\tilde{\Lambda}_1 \leq \frac{M}{N\sigma^2}\langle (f(x)-f(y))^2\rangle_{d(x,y)=1}$. Multiplying by $2N/M = 1/\langle D\rangle$ yields

$$\lambda_1 = \frac{\tilde{\Lambda}_1}{\langle D \rangle} = \frac{\langle (f(x) - f(y))^2\rangle_{d(x,y)=1}}{2\sigma^2} = 1 - \hat{\rho}_*(1).$$

A simple rearrangement completes the proof. ∎

Corollary. If Γ is regular with vertex degree D, then we have

$$\hat{\rho}(1) = \hat{\rho}_*(1) \leq \frac{\Lambda_1}{D} = \lambda_1,$$

where Λ_1 is the second largest eigenvector of the adjacency matrix A and λ_1 is the second largest eigenvector of the transition matrix of a simple random walk on Γ [123, 122].

Proof. This follows immediately from the definition $\Delta = D - A$ and the fact that D is a multiple of the identity matrix in case of regular graphs. ∎

This theorem effectively gives an upper bound on the nearest neighbor correlation of arbitrary landscapes on arbitrary configuration spaces. It provides also an alternative proof for the theorem in section 4.5.3.

4.7. Empirical Anisotropy

It is crucial for an understanding of the dynamics of a (heuristic) optimization procedure to know whether the landscape looks like a typical instance of an isotropic random field or whether there are significant anisotropies. The great importance of the ridge-like anisotropies for the dynamics of evolutionary adaptation, for instance, is discussed in [40].

The random field definition of isotropy becomes useless if we are given only a single instance. It is clear that we can only base our approach on a comparison of samples taken from different regions of the configuration space $\mathcal{C}$. One has to be careful when choosing these samples, however, since we have to expect uncontrolled effects on the sample statistics if we average over set of configurations with unequal geometries. For example, the mean fitness along a "straight line" in configuration space will in general be quite different from the mean fitness of a configuration and its neighbors. Recently, a collection $\mathcal{B}$ of test sets with the following four properties has been proposed as a suitable sampling strategy [124]

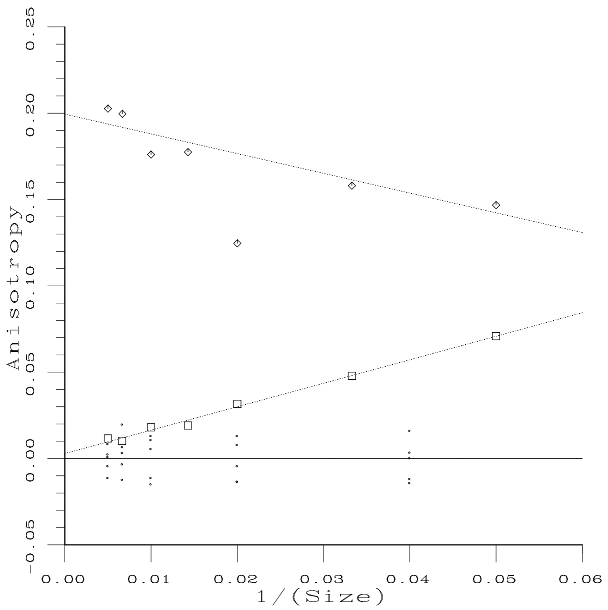


Figure 13: Empirical coefficient of anisotropy as a function of chain length for three model landscapes.
(·) Sherrington-Kirkpatrik spin glass with $2n$ spins. A reasonable collection of test sets is obtained by fixing one halve of the spins within each given test set, which is thus a hypercube of dimension n. The corresponding random field is is isotropic (see sect 3.4.1). The numerical data indicate that the instances of the SK model are indeed empirically isotropic.
In contrast the two examples of RNA free energy landscapes (see section 5 for details) exhibit anisotropies. The RNA free energy landscape of the natural four-letter alphabet **GCAU** and the landscape of the artificial **GCXK** alphabet differ in the logic and strenght of the possible base pairs. In both cases the test sets are constructed as follow: A_σ is the set of all sequences that have **G** or **C** at each position i with $\sigma_i = 1$ and **A** or **U** (or **X** or **K**, respectively) on the $\sigma_i = 0$ positions. The binary string σ thus characterizes the "slice" A_σ (which is a Boolean Hypercube of dimension n). The data show that the anisotropies vanish asymptotically for the artificial **GCXK** alphabet (□) which has two equivalent types of base pairs (**GC** and **XK**). In constrast, we find a large amount of anisotropy for the natural alphabet, for which the base pair **AU** is much weaker than **GC** and there is the possibility of forming **GU** pair (◇).

(i) $\mathcal{B}$ is a partition of the configuration space $\mathcal{C}$.
(ii) $A \in \mathcal{B}$ is a connected subgraph of $\mathcal{C}$.
(iii) Any two subgraphs $R, S \in \mathcal{B}$ are isomorphic, i.e., they have the same geometry.
(iv) $1 \ll |A| \ll |\mathcal{C}|$, for all $A \in \mathcal{B}$.

The first three requirements make sure that we take fair samples. Note that these conditions are quite restrictive. The last condition ensures that we have enough samples. The basic definition in [124] is the following:

Definition. A value landscape is *empirically isotropic* if for any family $\mathcal{B}$ of test sets holds

$$\left\langle \big(f(x) - f(y)\big)^2 \right\rangle_\mu^A = \left\langle \big(f(x) - f(y)\big)^2 \right\rangle_\mu ,$$

i.e., the average of the squared fitness difference of pairs of configurations belonging to a given symmetry class μ will be the same whether it is measured in single test set or over the entire configuration space. By replacing the symmetry classes by distance classes one obtains a completely analogous definition for *empirically homogeneous* landscapes.

The average correlation of pairs of configurations in a test set, $\bar{\rho}_{\mathcal{B}}$, will not vanish in general. This reflects that the test sets A are much smaller than the entire configuration space $\mathcal{C}$. It can be calculated from the empirical autocorrelation functions discussed earlier in this chapter:

$$\bar{\rho}_{\mathcal{B}} = \sum_{\mu} p_A(\mu)\hat{\rho}(\mu) \qquad \text{or} \qquad \bar{\rho}_{\mathcal{B}} = \sum_{d} p_A(d)\hat{\rho}_h(d)$$

where $p_A(\mu)$ and $p_A(d)$ are the probability that a randomly chosen pair of configurations from the test set A belongs to the symmetry class μ and to the distance class d, respectively.

Now consider the variance $\mathrm{var}_{\mathcal{B}}[\langle f\rangle_A]$, measured over all test sets $A \in \mathcal{B}$, of the average fitness values $\langle f\rangle_A$ within the test sets A. The main technical result in [124] is the following

Theorem. Let f be an empirically isotropic (empirically homogeneous) landscape, let $\mathcal{B}$ be a collection of test sets fulfilling (i) through (iv), and let $\bar{\rho}_{\mathcal{B}}$ be defined as above via the symmetry classes (distance classes) of the configuration space. Then

$$\mathrm{var}_{\mathcal{B}}[\langle f\rangle_A] = \sigma^2(1-\bar{\rho}_{\mathcal{B}})\,,$$

where σ^2 is as usual the variance of all fitness values.

Proof. See [124]. ∎

It seems natural to measure the anisotropy of a landscape by the extent to which this relation is violated. We therefore call the dimensionless parameter

$$\alpha_{\mathcal{B}} = \frac{\mathrm{var}_{\mathcal{B}}[\langle f\rangle_A]}{\sigma^2} - \bar{\rho}_{\mathcal{B}}$$

the *coefficient of anisotropy* with respect to the partition $\mathcal{B}$. From a theoretical point of view it is tempting to define the "true anisotropy" of a landscape as the maximum value of $\alpha_{\mathcal{B}}$ from all partitions $\mathcal{B}$. From a practical point of view, however, one would need prohibitively large computer resources to actually compute $\max_{\mathcal{B}} \alpha_{\mathcal{B}}$. A simple application is shown as figure 13.

The main advantage of this approach is that it is as close as possible to the definitions for random fields. Moreover, all quantities have the flavor of second moments and can be computed quite efficiently. Its major disadvantages are the dependence of the numerical values on the choice of the collection of test sets $\mathcal{B}$, and the fact that it gives only global information, thereby neglecting the overwhelming importance of the geometry of the anisotropies.

5. Biological Landscape: RNA

5.1. RNA Secondary Structures

5.1.1. Structure Formation

RNA molecules spontaneously form three-dimensional structures by folding the sequences in aqueous solutions, which contain appropriate concentrations of structure stabilizing divalent cations like Mg^{2+}, and have appropriate ionic strength, pH and temperature. The major driving force for structure formation is Watson-Crick base pairing (**G**$\equiv$**C**, and **A**$=$**U**) mediated by partial intramolecular complementarity of sequences, as well as **G**$-$**U** base pair formation. Other intermolecular forces and the interaction with the aqueous solvent shape the spatial structure of an RNA molecule. Small and medium size RNA molecules (with chain lengths $n < 200$) form equilibrium structures which are independent of the mechanism of folding and thus are completely determined by the sequence and the environmental conditions. Structure formation of large RNA molecules appears to be controlled at least in part by the kinetics of folding process.

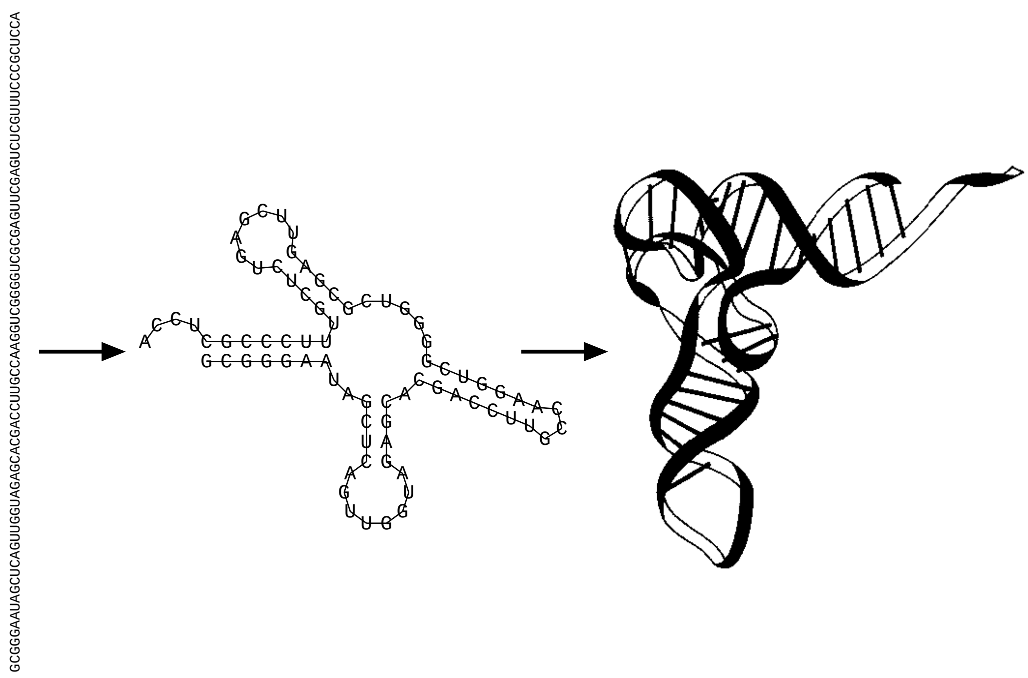


Figure 14: Folding of an RNA sequence into its spatial structure. The process is partitioned into two phases: in the first phase only the Watson-Crick-type base pairs are formed (which constitute the major fraction of the free energy), and in the second phase the actual spatial structure is built by folding the planar graph into a three-dimensional object. The example shown here is phenylalanyl-transfer-RNA (t-RNA $^{\rm phe}$) whose spatial structure is known from X-ray crystallography.

Formation of spatial structures of RNA molecules is commonly partitioned into two steps (Figure 14). The first step comprises conventional base pair formation (**AU**, **GC**, **GU**) and yields the so-called secondary structure of the RNA molecule, see below. In the second step the base pairing pattern is converted into the 3D structure.

5.1.2. Structure Prediction

Secondary structures are defined as listings of the Watson-Crick-type base pairs in the actual structure which fulfill the "no-knot condition" and may therefore be represented as planar graphs. Two features of RNA secondary structures will be important later on. They are discrete by definition (two bases either form pair or they don't), and they are composed of largely independent structural elements:

- *stacks*, i.e., double-helical regions,
- *loops*, i.e., unpaired regions enclosed by stacks (the degree of a loop is the number of stacks attached to it; consequently, a hairpin loop has degree 1, bulges and interior loops have degree 2, and all loops with degrees larger than 2 are multi-loops), and
- *external elements* are strains of unpaired bases that are not part of a loop (they are either joints, which connect components, or free ends).

There are several reasons for considering the secondary structure as a crude first approximation to the spatial structure of the RNA.

- Conventional base pairing and base pair stacking cover the major part of the free energy of folding.
- Secondary structures are used successfully in the interpretation of RNA function and reactivity.
- Secondary structures are conserved in evolutionary phylogeny.

The statistical investigation of RNA based landscapes requires the knowledge of several hundred thousand values derived from RNA structures. These data are not available at present, neither through experimental measurement nor through computation of the 3D-structures (which are highly time consuming and still unreliable). Restriction to secondary structures as a crude approximation to real structures, however, renders computation possible.

Table 6. Folding Algorithms.

Algorithm	ψ	Abbr.	Remark	Reference
deterministic				
Minimum Free Energy	$-$	`MFE`	fast	[147, 148]
Kinetic Folding	$+$	`KIN`	fast	[85]
5'-3' Folding	$+$	`5-3`	fast	[130]
Partition Function	$-$	`PF`	ensemble	[86]
Maximum Matching	$-$	`MM`	unrealistic	[94]
stochastic				
Simulated Annealing	$+$	`SA`	very slow	[91, 92]

$^{\psi}$ Pseudo-knots can be included. The major problem with the prediction of pseudo-knots is, however, the lack of sufficient experimental energy parameters.

A variety of computer programs predicting RNA secondary structures have been published. A very brief overview is given in Table 6. Two public domain packages for RNA folding are currently available by anonymous `ftp`: Zuker's `mfold` [146] and the `Vienna RNA Package` [61], the latter providing a variety of different algorithms for structure prediction and structure comparison [62].

5.2. RNA Free Energy Landscapes

The most obvious biophysical quantity to investigate is simple the stability of a (secondary) structure, that is, its free energy of structure formation. The RNA free energy landscapes may serve as a prototype for biophysically interesting landscapes despite that fact that there is not evolutionary process that would optimize the free energy of structure formation. It is has been found recently, for instance, that the landscapes of activation energies for the melting of secondary structure closely resemble the corresponding minimum free energy landscapes [131]. Free energy landscapes have been computed with a variety of algorithms (see table 6) for a number of different alphabets (see table 7).

Table 7. Scaled correlation lenght ℓ^* for RNA free energy landscapes.

	ΔG				ΔG^{ne}		Model†	
Alphabet	`MFE`	`MFE[s]`$^+$	`PF`	`KIN`	Melt.	Rec.	D	A
GC	0.0857	0.0768	0.127	0.0448	0.099	0.030	$\leq 3.5^*$	$\leq 3^*$
AU	0.0600							
GU	0.0514							
AUI	0.0532							
AGC	0.0982							
AUC	0.0856							
AUG	0.1191							
UGC	0.3850							
GCAU	0.2627	0.2976	0.243	0.1734	0.268	0.055	0.2	$\leq 15^*$
GCAU$^{\%}$	0.2044							
GCXK	0.1070				0.125	0.070		
ABCDEF	0.1182							

All data refer to T=37°C.

$^+$ [s] refers to Salser's old parameter set [107], all other data have been obtained with a recently updated paramter set [48].

* A linear dependence on the chain length could not be verified. The bounds correspond to chainlenght $n \leq 120$.

† These landscapes have been used to model replication and degradation in the computer simulations [46, 45].

% **GU** pairs forbidden.

Correlation lengths of free energies are essentially linear functions of the chain length n. Numerical data of scaled correlation lengths are compiled in table 7. The base pairing alphabet has remarkably strong influence on the correlation length: the correlation lenght for **AUGC** landscapes are much longer than for pure **GC** or pure **AU**-sequences. Detailed information on the RNA free energy landscapes can be found in [43, 47, 44].

Table 8. Decomposition of RNA Free Energy Landscapes.

n	a_6			a_1		a_7
	GCAU	**GCXK**	**GC**	**GCAU**	**GCXK**	**GC**
20	0.68	0.86		0.32	0.14	
30	0.65	0.88		0.35	0.12	
50	0.57	0.93	0.73	0.43	0.07	0.27
70	0.57	0.93		0.43	0.07	
100	0.53	0.96		0.47	0.04	

As an application of the Fourier-theory of landscapes we briefly discuss here the decomposition of RNA free energy landscapes into their modes. The correlation data are taken from [47]. The data are fitted reasonably well by by a two-mode model of the form $\hat{\rho} = a\rho_k + (1-a)\rho_l$. The coefficient $0 \le a \le 1$ and the modes k and l cannot be computed directly due to the fairly large amount of noise in the empirical correlation data. We therefore had to resort to a least square fitting procedure that takes the non-negativity of the coefficient(s) explicitly into account and uses only a small number of non-zero coefficients. We do not expect the RNA free energy landscapes to be *exactly* superpositions of only two modes, we rather expect a broad peak of the spectrum concentrated at about $\ell = 5 \ldots 8$. Using more than two modes increases the quality of the fits only slightly. Furthermore the fits with more than two coefficients do not compare well when independent data sets are used. Numerical data are given in table 8. We conclude hence that our numerical estimates of the correlation functions $\hat{\rho}(\mu)$ are not accurate enough (approximate noise level $\pm 5\%$) for obtaining a more detailed decomposition of the RNA free energy landscapes.

We find a pronounced difference between the natural **GCAU** alphabet and an artificial alphabet **GCXK** with two base pairs, **GC** and **XK**, of equal strength. While the latter is represented very well by the $\ell = 6$ mode, we obtain a superposition of the modes $\ell = 1$ and $\ell = 6$ in case of the natural alphabet. This does not come as a surprise, however. There is an essentially linear increase of the free energy with increasing **GC** content which accounts for the $\ell = 1$ mode. (Recall that $\ell = 1$ corresponds to the additive fitness model, or 'Fujijama' landscape). The order of the dominating modes allows one to draw conclusions on the approximate size of the building blocks underlying the landscape: in case of RNA free energies the data indicate that on average a nucleotide interacts strongly with about 5 other position, leading to a building block size of 6. This is a perfectly reasonable number given the energy model [148] for RNA secondary structure formation: a major contribution of the energy comes from stacking of base pairs, hence we expect at least the pairing partners and their neighbors to be important.

5.3. Combinatory Maps

5.3.1. The Metric Space of RNA Secondary Structures

The dichotomy between genotypes and phenotypes is basic to biological evolution. Molecular biology has shown that the genetic information for the unfolding of organisms, which represent the phenotypes, is contained in the DNA, which is the genotype. RNA molecules in cell-free evolution experiments are both genotypes and phenotypes. Following Sol Spiegelman we consider the spatial structure of the RNA molecule as its phenotype. Indeed, this structure is evaluated by the selection process. The structure of the RNA molecule acts thus as mediator between the sequence and the scalar property that is considered in the (fitness) landscape. The relation between sequences and structures thus seems to be basic to all landscapes of biopolymers.

The *shape space* $\mathcal{Y}$ is defined as the set of all possible RNA secondary structures. The notion of a shape space was used previously in theoretical immunology for the set of all structures presented by all possible antigens [99, 112]. Structures are non scalar objects, and there is no "structure landscape". In particular, the notion of a (local) optimum does not make sense in this context, since the structures do not admit an order.

On the other hand, we can at least define a notion of similarity between secondary structures. RNA secondary structures can be represented uniquely as rooted planar trees [114, 115, 44]. The conversion assigns an internal node to each base pair and a leaf to each unpaired digit. (An extra root is conveniently added in order to prevent the formation of a forest.) An alternative tree representation, known as homeomorphically irreducible trees, is obtained by converting connected unpaired regions and connected stacked regions, respectively, into single nodes, that can additionally be given a weight corresponding to the size of the structural element. Comparison of trees by means of edit distances is a well known technique in computer science [133, 95]. This method provides a metric distance measure between trees, and thus also between secondary structures. An alternative method for the comparison of RNA secondary structures is based on the so-called mountain representation [63]. The secondary structures are encoded as linear strings with balanced parentheses representing the base pairs, and some other symbol coding for unpaired positions. Distances are computed by direct end-to-end alignment of these strings. It has been shown that the results reported in this contribution are almost independent of the actual choice of the folding algorithm, the parameter set for the folding and the distance measures for secondary structures [132]. In this section we will use the notation $D(a,b)$ for a distance between secondary structures; thus $(\mathcal{Y},D)$ is a metric space.

Definition. A map $f:\mathcal{C}\to\mathcal{Y}$ from the configuration space $\mathcal{C}$ into another metric space $\mathcal{Y}$ is called a *combinatory map*. A landscape is special combinatory map with $\mathcal{Y}=\mathbb{R}$. The conceptual difference is that landscapes have the order of the (fitness) values as an additional feature.

5.3.2. Correlation in Combinatory Maps

A theory of combinatory maps analogous to the theory of landscapes discussed in the previous sections can be constructed based on based on the definition of ρ_* in sect. 4.6.1 by replacing the squared differences by squared distances in $\mathcal{Y}$. It has been proposed in [47, 44]

Definition. The correlation function of a combinatorial map and a "combinatorial random field", respectively, are given by

$$\begin{aligned}
\hat{\rho}(\mu) &= 1-\frac{\langle D^2(f(x),f(y))\rangle_{(x,y)\in\mu}}{\langle D^2(f(p),f(q))\rangle} \\
\rho(\mu) &= 1-\frac{\mathcal{E}[D^2(f(x),f(y))]}{\mathcal{E}[D^2(f(p),f(q))]} \qquad \text{for } (x,y)\in\ddot{\mu} \text{ and } (p,q)\in v\times V
\end{aligned}$$

The fact that there are no mean values suggests that we say a combinatorial random field is isotropic if $\mathcal{E}[D^2(f(x),f(y))]$ is constant on a symmetry class μ. The notion of empirical isotropy also carries over: the squared differences are simply replaced by the corresponding squared distances in sect. 4.7. In the same spirit we may define the unscaled and scaled structure correlation lengths ℓ and ξ, respectively.

The correlation length may serve as a statistical measure of the hardness of optimization problems for a particular heuristic algorithm. The structure correlation length allows for a similar interpretation. The shorter the correlation length, the more likely is a structural change occurring as a consequence of mutation. The correlation length thus measures stability against mutation.

More detailed information on a combinatory map can be obtained from the conditional probabilities

$$P(D,\mu) \;=\; \text{Prob}\left\{\, D(f(x),f(y)=D \;\middle|\; (x,y)\in\mu \,\right\}.$$

that the images of two randomly chosen configuration have a distance D in shape space *given* that the pair of configurations (x,y) belongs to the symmetry class μ in configuration space. We call $P(D|\mu)$ a *density surface*. The autocorrelation function can be calculated from a density surface using

$$\begin{aligned}\langle D^2(f(x),f(y))\rangle_{(x,y)\in\mu} &= \sum_D D^2\, P(D|\ddot{\mu}) \\ \langle D^2(f(p),f(q))\rangle_{p,q\in V} &= \sum_\mu p(\mu)\langle D^2(f(x),f(y))\rangle_{(x,y)\in\ddot{\mu}}\,.\end{aligned}$$

The density surfaces contain of course more information than the autocorrelation function. For instance, the probability for finding a neutral neighbor, i.e., for finding a mutant that folds into the same structure, is given by $p_{\mathrm{NN}} = P(0|1)$.

5.3.3. The Combinatory Maps of RNA Secondary Structures

The number of secondary structures which are acceptable as minimum free energy structures of RNA molecules can be computed from combinatorics of structural elements [60, 109, 128, 138]. It is much smaller than the number of different sequences, since we have

$$S_n \sim 1.485\times n^{-3/2}(1.849)^n$$

different secondary structures for κ^n sequences, i.e., there are even less structures than there are binary sequences.

The mapping from sequence space into shape space cannot be inverted: many sequences have to fold into the same secondary structure. We find that the frequency distribution of sequences folding into a common structure follows a generalized Zipf's law

$$\phi(r) = a(r+b)^{-c}$$

where r is the rank of a secondary structure S by the number of sequences folding into this particular structure and $\phi(r)$ is the fraction of sequences folding into S. The parameters b and c describe the number of very frequent structures and the fall-off of the tail of very rare structures, respectively. a is a normalization factor. The frequency distribution of structures has a very sharp peak: relatively few structures are very common, many structures are rare and play no statistically significant role.

An almost linear increase of the strucure correlation length ℓ with chain length is observed for RNA sequences. Substantial differences are found in the correlation lengths derived from different base pairing alphabets. In particular, the structures of natural **AUGC** sequences are much more stable against mutation than pure **GC**-sequences or pure **AU**-sequences.

It provides a plausible explanation for the use of two base pairs in nature: optimization in an RNA world with only one base pair would be very hard, and the base pairing probability in sequences with three base pairs is rather low and hence most random sequences of short chain lengths ($n<50$) do not form thermodynamically stable structures. The choice of two base pairs thus appears to be a compromise between stability against mutation and thermodynamic stability. An alternative explanation for the usage of two base pairs in nature was published recently [129]: the current alphabet is understood to be optimal in an RNA world where replication fidelity decreases and

Table 9. Asymptotic behaviour of correlation lenght for RNA secondary structurs.

	ΔG			
Alphabet	`MFE`	`MFE[s]`$^+$	`PF`	`KIN`
GC	0.0222	0.0245	~ 0.02	0.0224
AU	0.0296			
GU	0.0479			
AUI	0.0299			
AGC	0.1414			
AUC	0.0792			
AUG	0.0768			
UGC	0.3135			
GCAU	0.0524	0.0414	~ 0.15	0.0616
GCAU$^{\%}$	0.0659			
GCXK	0.0491			
ABCDEF	0.0793			

$^+$ See table 8.
$^{\%}$ **GU** pairs forbidden.

catalytic efficiency increases with alphabet size. Both hypothesis are experimentally testable and hence we may expect a decision in favor of one of the two alternatives in the future.

In order to gain more information on the relation between RNA sequences and structures an inverse folding algorithm which determines the sequences that share the same minimum free energy secondary structure was conceived and applied to a variety of different structures [109]. One finds that sequences folding into the same secondary structure are, in essence, randomly distributed. Since there are relatively few common structures and the sequences folding into the same structure are randomly distributed in sequence space, all common structrues are found in relatively small patches of sequence space. For natural **AUGC**-sequences of chain length $n=100$, for instance, we can expect all common structures to be found in a sphere of radius $h=16$ in Hamming distance. There are as many as 6.2×10^{25} sequences in such a ball. Although this number is large, it is nothing compared to the total number of sequences of this chain length: $4^{100} \approx 1.6 \times 10^{60}$.

On the other hand there is a high probability for finding neutral neighbors. For the biophysical alphabet one finds, for instance, $p_{\mathrm{NN}} \approx 0.4$ for long enough chains. Hence the sequences folding into a given target structure are not isolated in sequence space. In order to refine our understanding of their geometrical arrangement in sequence space a computer experiment was carried out allowing for an estimate of the diameter of connected sets of neutral sequences. We search for "neutral paths" through sequence space. The Hamming distance from the reference increases monotonously along such a neutral path but the structure remains unchanged. A neutral path ends when no further neutral sequence is found in the neighbourhood of the last sequence. Biophysics forces us to allow for a modification of the notion of neighborhood; we allow point mutations as well as the replacement of a base pair by another type of base pair. The length $\mathcal{L}$ of a path is the Hamming distance between the reference sequence and the last sequence, and hence a lower bound on the diameter of the connected "neutral network". Clearly, a neutral path cannot be longer than the chain length, $\mathcal{L} \leq n$. It turns out that a fraction of as many as 20% of all neutral walks have length $\mathcal{L} = n$ for **AUGC sequences** for length $n = 100$. They lead through the entire sequence space to a sequence differing in all positions from the reference but still sharing its structure. In shape spaces derived from binary sequences almost no neutral path reaches the complementary sequence.

This is certainly a consequence of the symmetry of the binomial distribution: there are very few sequences in the error classes $n-1$, $n-2$, etc., and it is unlikely that we find one among them which folds precisely into the same structure as the reference sequence. Still the average lenght $\mathcal{L}$ is much larger than the average distance of two randomly chosen sequences, $\mathcal{L} \gg n/2$. For details see [109].

The union of all neutral paths forms a dense neutral network in the example considered here. This, of course, need not be the case in general: we may have short neutral paths confined to small disjoint regions in sequence space. We may expect a chracteristic change in the hardness of the optimization problem depending on whether the network of neutral paths is below or above a percolation threshold. It is worth noting that neutral nets are not a peculiarity of the few most frequent structures. Even the rarest structures we were able to find give rise to networks that reach way beyond the average distance of random sequences [110].

Random combinatory maps with prescribed correlation structure are not as easy to construct as random fields. As a first step one considers only the pre-image of a certain structure, i.e., the set of all sequences folding into the same secondary structure. This set forms a sub-graph of the configuration space. The structure of such random sub-graphs with a prescribed probability for neutral neighbors, that is, with a prescribed vertex degree has been investigated in detail [104]. In particular, explicit expressions for the fraction of neutral mutations such that the subgraph is dense and connected have been obtained for Hamming graphs. A random sub-graph of a Hamming graph with a alphabet of size α is dense and connected in the limit of large systemsize n if and only if

$$p_{NN} \geq 1 - \sqrt[\alpha-1]{\alpha^{-1}}\,.$$

6. Conclusions

The basic ingredient of the theory of landscapes presented in this contribution is the notion of Fourier series defined on arbitrary graphs. The detailed mathematical form of these series depends crucially on the geometry of the configuration space on which the landscape, or a random field model of a landscape, is defined in that the basis functions are eigenvectors of the graphs Laplacian. Landscapes which are such eigenvectors are termed elementary as they play a special role in the theory. Surprisingly, we find that almost all the well known combinatorial optimization problems (with the exception of the asymmetric TSP) are elementary. The main subject of this contribution is therefore a theory of elementary landscapes and random fields.

It turns out the correlation structure of landscapes and random fields is closely linked to these generalized Fourier series, and therefore to the spectrum of the graph Laplacian. In particular the geometry of local optima can be linked to the graph Laplacian. Isotropic random fields are characterized in terms of their Fourier coefficients. All landscapes on the Boolean hypercube can be represented as a superposition of p-spin Hamiltonians, and isotropic Gaussian landscapes correspond exactly to a superposition of Derrida's version of p-spin models: the coupling coefficients are uncorrelated and the variance of the coefficients depends only on the order p of the coupling. The Fourier expansion of (fitness) landscapes can be used to find the set of all possible correlation functions; more precisely, the set of all autocorrelation functions forms a simplex spanned by suitably normalized left eigenvectors of the collapsed adjacency matrix of the configuration space. The corresponding eigenvalue determines the decay of the correlation function along a simple random walk on the landscape.

The formalism derived in this contribution suggests to approximate a given landscape by a superposition of statistical models that are elementary landscapes (which are hopefully easier to understand). The coefficients of such a superposition can be obtained by observing that the correlation function is a superposition of the correlation functions of the elementary landscapes under consideration. The latter can be computed analytically for a number of problems, since they are the left eigenvectors of the collapsed adjacency matrix of the underlying configuration space. Instead of the correlation function we can then use the coefficients of this decomposition as a means of characterizing a landscape. On a Boolean hypercube, for instance, this spectrum can be interpreted directly as the relative importance of p-ary (spin) interactions.

As a biological application we have studied RNA free energy landscapes. It turns out that the natural **GCAU** landscape can be fitted reasonably well by a superposition of a 1-spin and a 6-spin model with equal amplitude, while the **GC** and **GCXK** landscapes can be approximated by a 6-spin model. We do not expect that RNA landscapes are elementary, we rather expect a distribution of p-spin models centered around $p = 6$. The numerical correlation data are to noisy, however, to permit a higher resolution, and we have to be content with the approximate location of the peak.

An extension of the notion of fitness landscapes to general sequence structure relations is outlined very briefly in the last sections. Based on an alternative form the autocorrelation one defines a generalized correlation function for mappings from one metric space into another one. Whether an anlogy of the Fourier theory of landscapes is possible also for such "combinatory maps" and their random field models is unknown at present.

Acknowledgements

Fruitful discussions with Andreas Dress, Walter Fontana, Ivo Hofacker, Stuart Kauffman, Christian Reidys, Peter Schuster, and Günter Wagner are gratefully acknowledged.

References

[1] E. Aarts and J. Korst. *Simulated Annealing and Boltzman Machines.* J. Wiley & Sons, New York, 1990.

[2] C. Amitrano, L. Peliti, and M. Saber. Population dynamics in a spin-glass model of chemical evolution. *J. Mol. Evol.*, 29:513–525, 1989.

[3] M. Averintsev. On a method of describing complete parameter fields. *Problemy Peredaci Informatsii*, 6:100–109, 1970.

[4] P. Bak, H. Flyvbjerg, and B. Lautrup. Coevolution in a rugged fitness landscape. *Phys.Rev. A[15]*, 46:6724–6730, 1992.

[5] G. Baker. Drum shapes and isospectral graphs. *J.Math.Phys.*, 7:2238, 1966.

[6] M. Bartlett. *An Introduction to Stochastic Processes.* Cambridge University Press, Cambridge UK, 1955.

[7] C. Benson and J. Jacobs. On hearing the shape of combinatorial drums. *J.Comb.Theory(B)*, 13:170–178, 1972.

[8] J. Bernasconi. Low autocorrelation binary sequences: statistical mechanics and configuration space analysis. *J.Physique*, 48:559–567, 1987.

[9] J. Besag. Spatial interactions and the statistical analysis of lattice systems. *Amer. Math. Monthly*, 81:192–236, 1974.

[10] N. Biggs. *Finite Groups of Automorphisms*, volume 6 of *London Mathematical Society Lecture Notes.* Cambridge University Press, Cambridge UK, 1971.

[11] N. Biggs. *Algebraic Graph Theory.* Cambridge University Press, Cambridge UK, 1st edition, 1974.

[12] N. L. Biggs. *Algebraic Graph Theory.* Cambridge University Press, Cambridge UK, 2nd edition, 1994.

[13] N. L. Biggs and A. White. *Permutation Groups and Combinatorial Structures.* Cambridge University Press, Cambridge UK, 1979.

[14] K. Binder and A. P. Young. Spin glasses: Experimental facts, theoretical concepts, and open questions. *Rev.Mod.Phys.*, 58:801–976, 1986.

[15] R. G. Bland and D. F. Shallcross. Large traveling salesmen problems arising from experiments in x-ray crystallography. *Oper.Res.Lett.*, 8:125–128, 1988.

[16] H. Bohr and S. Brunak. Travelling salesman approach to protein conformation. *Complex Systems*, 3:9–28, 1990.

[17] B. Bollobás. *Graph Theory – An Introductory Course.* Springer-Verlag, New York, 1979.

[18] S. Bonhoeffer and P. F. Stadler. Errortreshold on complex fitness landscapes. *J. Theor. Biol.*, 164:359–372, 1993.

[19] A. Bray and M. Moore. Metastable states in spin glasses. *J.Phys.C:Solid St.Phys.*, 13:L469–L476, 1980.

[20] A. Bray and M. Moore. Metastable states in spin glasses with short-ranged interactions. *J.Phys.C:Solid St.Phys.*, 14:1313–1327, 1981.

[21] A. Brouwer, A. Cohen, and A. Neumaier. *Distance-regular Graphs.* Springer Verlag, Berlin, New York, 1989.

[22] R. Brualdi and H. Ryser. *Combinatorial Matrix Theory.* Cambridge Univ. Press, Cambridge UK, 1991.

[23] P. Cameron and J. vanLint. *Designs, Graphs, Codes, and their Links*, volume 22 of *London Math. Soc. Student Texts.* Cambridge University Press, Cambridge UK, 1991.

[24] D. Chan and D. Mercier. IC insertion: An application of the TSP. *Int.J.Prod.Res.*, 3:9–28, 1989.

[25] I. Chavel. *Eigenvalues in Riemannian Geometry.* Academic Press, Orlando Fl., 1984.

[26] D. Cvetković, M. Doob, and H. Sachs. *Spectra of Graphs – Theory and Applications.* Academic Press, New York, 1980.

[27] D. Cvetković, M. Doob, and H. Sachs. *Spectra of Graphs – Theory and Applications*, volume New York. Academic Press, 1980.

[28] C. De Dominicis, M. Gabay, T. Garel, and H. Orland. White and weighted averages over solutions of the Thouless Anderson Palmer equations for the Sherrington Kirkpatrick spin glass. *J.Physique*, 41:923–930, 1980.

[29] P. Delsarte. *An algebraic approach to association schemes of coding theory*, volume 10 of *Phillips Research Reports Supplements.* Phillips, 1973.

[30] B. Derrida. The random energy model. *Phys.Rep.*, 67:29–35, 1980.

[31] B. Derrida. Random energy model: Limit of a family of disordered models. *Phys.Rev.Lett.*, 45:79–82, 1980.

[32] B. Derrida. Random-energy model: An exactly solvable model of disorderes systems. *Phys. Rev. B*, 24(5):2613–2626, 1981.

[33] B. Derrida and E. Gardner. Metastable states of a spin glass chain at 0 temperature. *J.Physique*, 47:959–965, 1986.

[34] P. Diaconis and M. Shahshahani. Generating a random permutation with random transpositions. *Z. Wahrscheinlichkeitsth. verw. Gebiete*, 57:159–179, 1981.

[35] R. Dobrushin. The description of a random field by means of its conditional probabilities, and conditions of its regularities. *Th. Prob. & Appl.*, 13:197–224, 1968.

[36] A. Dress and D. Rumschitzki. Evolution on sequence space and tensor products of representation spaces. *Acta Appl.Math.*, 11:103–111, 1988.

[37] C. Dunkl. A Krawtchouk polynomial addition theorem and wreath products of symmetric groups. *Indiana Univ.Math.J.*, 25:335–358, 1976.

[38] C. Dunkl. Orthogonal functions on some permutation groups. In D. Ray-Chaudhuri, editor, *Proceeding of Symposia in Pure Mathematic Vol. 34*, New York, 1979. American Mathematical Society.

[39] M. Eigen. Selforganization of matter and the evolution of biological macromolecules. *Die Naturwissenschaften*, 10:465–523, 1971.

[40] M. Eigen, J. McCaskill, and P. Schuster. The molecular Quasispecies. *Adv. Chem. Phys.*, 75:149 – 263, 1989.

[41] M. Fisher. On hearing the shape of a drum. *J.Comb.Theory*, 1:105–125, 1966.

[42] H. Flyvbjerg and B. Lautrup. Evolution in a rugged fitness landscape. *Phys. Rev. A*, 46:6714–6723, 1992.

[43] W. Fontana, T. Griesmacher, W. Schnabl, P. Stadler, and P. Schuster. Statistics of landscapes based on free energies, replication and degredation rate constants of RNA secondary structures. *Monatsh. Chemie*, 122:795–819, 1991.

[44] W. Fontana, D. A. M. Konings, P. F. Stadler, and P. Schuster. Statistics of rna secondary structures. *Biochemistry*, 33:1389–1404, 1993.

[45] W. Fontana, W. Schnabl, and P. Schuster. Physical aspects of evolutionary optimization and adaption. *Physical Review A*, 40(6):3301–3321, 1989.

[46] W. Fontana and P. Schuster. A computer model of evolutionary optimization. *Biophysical Chemistry*, 26:123–147, 1987.

[47] W. Fontana, P. F. Stadler, E. G. Bornberg-Bauer, T. Griesmacher, I. L. Hofacker, M. Tacker, P. Tarazona, E. D. Weinberger, and P. Schuster. RNA folding and combinatory landscapes. *Phys. Rev. E*, 47(3):2083 – 2099, 1993.

[48] S. M. Freier, R. Kierzek, J. A. Jaeger, N. Sugimoto, M. H. Caruthers, T. Neilson, and D. H. Turner. Improved free-energy parameters for predictions of RNA duplex stability. *Proc. Natl. Acad. Sci., USA*, 83:9373–9377, 1986.

[49] Y. Fu and P. W. Anderson. Application of statistical mechanics to NP-complete problems in combinatorial optimization. *J.Phys.A:Math.Gen.*, 19:1605–1620, 1986.

[50] C. Gardiner. *Handbook of Stochastic Methods*. Springer-Verlag, Berlin, 2nd edition, 1990.

[51] E. Gardner and B. Derrida. The probability distribution of the partition function of the random energy model. *J.Phys. A*, 22:1975–1982, 1989.

[52] M. Garey and D. Johnson. *Computers and Intractability. A Guide to the Theory of $\mathcal{NP}$ Completeness*. Freeman, San Francisco, 1979.

[53] G.M. Adel'son-Velskii *et al.* Example of a graph without a transitive automorphism group. *Soviet. Math. Dokl.*, 10:440–441, 1969. Russian.

[54] C. D. Godsil. *Algebraic Combinatorics*. Chapman & Hall, New York, 1993.

[55] M. Golay. Sieves for low-autocorrelation binary sequences. *IEEE Trans.Inform.Th.*, IT-23:43–51, 1977.

[56] G. Grimmet. A theorem about random fields. *Bull. London Math. Soc.*, 5:81–85, 1973.

[57] L. Grover. Local search and the local structure of NP-complete problems. *Oper.Res.Lett.*, 12:235–243, 1992.

[58] R. W. Hamming. Error detecting and error correcting codes. *Bell Syst.Tech.J.*, 29:147–160, 1950.

[59] D. G. Higman. Intersection matrices for finite permutation groups. *J. Algebra*, 6:22–42, 1967.

[60] I. Hofacker, P. Schuster, and P. Stadler. Combinatorics of secondary structures. *submitted to SIAM J. Disc. Math.*, 1993.

[61] I. L. Hofacker, W. Fontana, P. F. Stadler, L. S. Bonhoeffer, M. Tacker, and P. Schuster. `Vienna RNA Package`. `pub/RNA/ViennaRNA-1.03 @ ftp.itc.univie.ac.at`. (Public Domain Software).

[62] I. L. Hofacker, W. Fontana, P. F. Stadler, S. Bonhoeffer, M. Tacker, and P. Schuster. Fast folding and comparison of RNA secondary structures. *Monatsh. Chemie*, 125(2):167–188, 1994.

[63] P. Hogeweg and B. Hesper. Energy directed folding of RNA sequences. *Nucleic acids research*, 12:67–74, 1984.

[64] H. Hotelling. Analysis of a complex of statistical variables into principal components. *J.Educ.Psych.*, 24:417–441 and 498–520, 1933.

[65] M. Huynen and P. Hogeweg. Pattern generation in molecular evolution. Exploitation of the variation in RNA landscapes. *J.Mol.Evol.*, 39:71–79, 1994.

[66] M. A. Huynen, P. F. Stadler, and W. Fontana. Evolution of RNA and the Neutral Theory. SFI Preprint # 95-01-006, 1995.

[67] R. Ingram. Some characters of the symmetric group. *Proc.Amer.Math.Soc.*, 1:358–369, 1950.

[68] M. Kac. Can you hear the shape of a drum. *Am.Math.Monthly*, 73(4):1–23, 1966.

[69] S. Karlin and H. Taylor. *A first course in stochastic processes.* Academic Press, New York, 1975.

[70] S. Kauffman. *The Origin of Order.* Oxford University Press, New York, Oxford, 1993.

[71] S. A. Kauffman and S. Levin. Towards a general theory of adaptive walks on rugged landscapes. *J. Theor. Biol.*, 128:11, 1987.

[72] S. A. Kauffman and E. D. Weinberger. The n-k model of rugged fitness landscapes and its application to maturation of the immune response. *J. Theor. Biol.*, 141:211, 1989.

[73] T. Koornwinder. Krawtchouk polynomials. A unification of two different group theoretic interpretations. *SIAM J.Math.Anal.*, 13:1011–1023, 1982.

[74] B. Krakhofer. Local optima in landscapes of combinatorial optimization problems. Master's thesis, University of Vienna, Dept. of Theoretical Chemistry, 1995.

[75] M. Krawtchouk. Sur une gènèralisation des polynomes d'Hermite. *Comptes Rendus*, 189:620–622, 1929.

[76] K. Laidler. *Chemical Kinetics.* Harper, New York, 3rd edition, 1992.

[77] E. L. Lawler, J. K. Lenstra, A. H. G. R. Kan, and D. B. Shmoys. *The Traveling Salesman Problem. A Guided Tour of Combinatorial Optimization.* John Wiley & Sons, 1985.

[78] S. Lin and B. Kernighan. An effective heuristic algorithm for the traveling salesman problem. *Oper.Res.*, 21:498–516, 1965.

[79] L. Lovász. Spectra of graphs with transitive groups. *Periodica Math.Hung.*, 6:191–195, 1975.

[80] I. MacDonald. *Symmetric Functions and Hall Polynomials.* Oxford Univ. Press, Oxford UK, 1979.

[81] C. Macken, P. Hagan, and A. Perelson. Evolutionary walks on rugged landscapes. *SIAM J.Appl.Math.*, 51:799–827, 1991.

[82] C. A. Macken and A. S. Perelson. Protein evolution on rugged landscapes. *Proc.Natl.Acad.Sci.USA*, 86:6191–6195, 1989.

[83] C. A. Macken and P. F. Stadler. Rugged landscapes. To appear in SFI summerschool volume 1993, 1995.

[84] F. MacWilliams and N. Sloane. *The Theory of Error-Correcting Codes.* North-Holland, Amsterdam, New York, Oxford, Tokyo, 1991.

[85] H. M. Martinez. An RNA folding rule. *Nucl.Acid.Res.*, 12:323–335, 1984.

[86] J. S. McCaskill. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, 29:1105–1119, 1990.

[87] M. Mézard and G. Parisi. Replicas and optimization. *J.Physique Lett.*, 46:L771–L778, 1986.

[88] M. Mézard, G. Parisi, and M. Virasoro. *Spin Glass Theory and Beyond.* World Scientific, Singapore, 1987.

[89] D. Miller and J. F. Pekny. Exact solution of large asymmetric traveling salesman problems. *Science*, 251:754–761, 1991.

[90] W. Miller Jr. *Symmetry and Separation of Variable*, volume 4 of *Enceyclopedia of Mathematics and its Applications.* Cambridge Univ. Press, Cambridge, UK, 1984.

[91] A. A. Mironov, L. P. Dyakonova, and A. E. Kister. A kinetic approach to the prediction of RNA secondary structures. *Journal of Biomolecular Structure and Dynamics*, 2(5):953, 1985.

[92] A. A. Mironov and A. E. Kister. RNA secondary structure formation during transcription. *J. of Biomolecular Structure and Dynamics*, 4:1–9, 1986.

[93] J. Moussouris. Gibbs and Markov systems with constraints. *J. Stat. Phys.*, 10:11–33, 1974.

[94] R. Nussinov, G. Piecznik, J. R. Griggs, and D. J. Kleitman. Algorithms for loop matching. *SIAM J. Appl. Math.*, 35(1):68–82, 1978.

[95] K. Ohmori and E. Tanaka. A unified view on tree metrics. In G. Ferrate, editor, *Syntactic and Structural Pattern Recognition*, pages 85–100, Berlin, Heidelberg, 1988. Springer-Verlag.

[96] R. Otten and L. vanGinneken. *The Annealing Algorith.* Kluwer Acad.Publ., Boston, 1989.

[97] R. Palmer. Optimization on rugged landscapes. In A. S. Perelson and S. A. Kauffman, editors, *Molecular Evolution on Rugged Landscapes: Proteins, RNA, and the Immune System*, pages 3–25. Addison Wesley, Redwood City, CA, 1991.

[98] A. S. Perelson and C. A. Macken. Protein evolution on partially correlated landscapes. Santa Fe Institute Preprint 94-11-060.

[99] A. S. Perelson and G. Oster. Theoretical studies of clonal selection: Minimal antibody repertoire size and reliability of self/ non-self discrimination. *Journal of Theoretical Biology*, 81:645–670, 1979.

[100] D. Powers. Eigenvectors of distance-regular graphs. *SIAM J. Matrix Anal. Appl.*, 9:399–407, 1988.

[101] D. Powers and M. Sulaiman. The walk partition and colorations of a graph. *Linear Algebra Appl.*, 48:145–159, 1982.

[102] C. Rao. *Linear Statistical Interference and Its Applications.* Wiley, New York, 2nd edition, 1973.

[103] C. Reidys and C. Forst. Replication on neutral networks in rna induced by rna secondary structures. Preprint, 1994.

[104] C. Reidys, P. Schuster, and P. F. Stadler. Generic properties of combinatory maps and application on RNA secondary structures. Preprint, 1995.

[105] D. Rumschitzky. Spectral properties of eigen's evolution matrices. *J.Math.Biol.*, 24:667–680, 1987.

[106] G. Sabidussi. Vertex transitive graphs. *Mh. Math*, 68:426–438, 1964.

[107] W. Salser. Globin messenger RNA sequences — analysis of base-pairing and evolutionary implications. *Cold Spring Harbour Symp. Quant. Biol.*, 42:985, 1977.

[108] P. Schuster. Complex optimization in an artificial RNA world. In D. Farmer, C. Langton, S. Rasmussen, and C. Taylor, editors, *Artificial Life II*, volume XII of *SFI Studies in the Science of Complexity*, pages 277–291. Addison-Wesley, 1992.

[109] P. Schuster, W. Fontana, P. F. Stadler, and I. L. Hofacker. From sequences to shapes and back: A case study in RNA secondary structures. *Proc.Roy.Soc.Lond.B*, 255:279–284, 1994.

[110] P. Schuster and P. F. Stadler. Landscapes: Complex optimization problems and biopolymer structures. *Computers Chem.*, 18:295–314, 1994.

[111] A. Schwenk. Computing the characteristic polynomial of a graph. In *Graphs and Combinatorics*, volume 406 of *Lecture Notes in Mathematics*, pages 153–162. Springer-Verlag, Berlin, 1974.

[112] L. A. Segel and A. P. Perelson. Computations in shape space: a new approach to immune network theory. In *Theoretical Immunology. Part Two*, pages 321 – 343. Addison-Wesley, Redwood City (Cal.), 1988.

[113] J.-P. Serre. *Linear Representations of Finite Groups.* Springer-Verlag, New York, Heidelberg, Berlin, 1977.

[114] B. A. Shapiro. An algorithm for comparing multiple RNA secondary stuctures. *CABIOS*, 4(3):387–393, 1988.

[115] B. A. Shapiro and K. Zhang. Comparing multiple RNA secondary structures using tree comparisons. *CABIOS*, 6:309–318, 1990.

[116] D. Sherrington and S. Kirkpatrick. Solvable model of a spin-glass. *Physical Review Letters*, 35(26):1792 – 1795, 1975.

[117] P. M. Soardi. *Potential Theory on Infinite Networks*, volume 1590 of *Lecture Notes in Mathematics.* Springer-Verlag, Berlin, 1994.

[118] G. B. Sorkin. Combinatorial optimization, simulated annealing, and fractals. Technical Report RC13674 (No.61253), IBM Research Report, 1988.

[119] F. Spitzer. Markov random fields and gibbs ensembles. *Amer. Math. Monthly*, 78:142–154, 1971.

[120] F. Spitzer. *Principles of Random Walks.* Springer-Verlag, New York, 1976.

[121] P. F. Stadler. Correlation in landscapes of combinatorial optimization problems. *Europhys. Lett.*, 20:479–482, 1992.

[122] P. F. Stadler. Linear operators on correlated landscapes. *J.Physique*, 4:681–696, 1994.

[123] P. F. Stadler. Random walks and orthogonal functions associated with highly symmetric graphs. *Disc. Math.*, 1994. in press.

[124] P. F. Stadler and W. Grüner. Anisotropy in fitness landscapes. *J. Theor. Biol.*, 165:373–388, 1993.

[125] P. F. Stadler and R. Happel. Correlation structure of the landscape of the graph-bipartitioning-problem. *J. Phys. A.: Math. Gen.*, 25:3103–3110, 1992.

[126] P. F. Stadler and R. Happel. Canonical approximation of landscapes. Submitted to J.Stat.Phys., 1994.

[127] P. F. Stadler and W. Schnabl. The landscape of the traveling salesman problem. *Phys. Letters A*, 161:337–344, 1992.

[128] P. Stein and M. Waterman. On some new sequences generalizing the Catalan and Motzkin numbers. *Discrete Mathematics*, 26:261–272, 1978.

[129] E. Szathmáry. Four letters in the genetic alphabet: a frozen evolutionary optimum? *Proc.Roy.Soc. London B*, 245:91–99, 1991.

[130] M. Tacker. *Robust Properties of RNA Secondary Structure Folding Algorithms.* PhD thesis, University of Vienna, 1993.

[131] M. Tacker, W. Fontana, P. Stadler, and P. Schuster. Statistics of RNA melting kinetics. *Eur. J. Biophys.*, 23:29–38, 1994.

[132] M. Tacker, P. F. Stadler, E. G. Bornberg-Bauer, I. L. Hofacker, and P. Schuster. Robust properties of RNA secondary structure folding algorithms. In preparation, 1995.

[133] K. Tai. The tree-to-tree correction problem. *J. ACM*, 26:422–433, 1979.

[134] F. Tanaka and S. Edwards. Analytic theory of the ground state properties of a spin glass: I. Ising spin glass. *J.Phys.F:Metal Phys.*, 10:2769–2778, 1980.

[135] D. Thouless, P. Anderson, and R. Palmer. *Phil.Mag.*, 35:593, 1977.

[136] J. vanLint. *Introduction to Coding Theory.* Springer-Verlag, New York, 1982.

[137] R. Voss. Characterization and measurement of random fractals. *Physical Scripta*, T13:257–260, 1986.

[138] M. S. Waterman. Secondary structure of single - stranded nucleic acids. *Studies on foundations and combinatorics, Advances in mathematics supplementary studies, Academic Press N.Y.*, 1:167 – 212, 1978.

[139] E. Weinberger and P. Stadler. Why some fitness landscapes are fractal. *J. Theor. Biol.*, 163:255–275, 1993.

[140] E. D. Weinberger. Correlated and uncorrelated fitness landscapes and how to tell the difference. *Biol.Cybern.*, 63:325–336, 1990.

[141] E. D. Weinberger. Fourier and Taylor series on fitness landscapes. *Biological Cybernetics*, 65:321–330, 1991.

[142] E. D. Weinberger. Local properties of Kauffman's N-k model: A tunably rugged energy landscape. *Phys. Rev. A*, 44(10):6399–6413, 1991.

[143] P. Whittle. Stochastic processes in several dimensions. *Bull. Int. Statist. Inst.*, 40:974–994, 1963.

[144] H. Wielandt. *Finite Permutation Groups.* Academic Press, New York, 1964.

[145] S. Wright. The roles of mutation, inbreeding, crossbreeeding and selection in evolution. In D. F. Jones, editor, *Int. Proceedings of the Sixth International Congress on Genetics*, volume 1, pages 356–366, 1932.

[146] M. Zuker. `mfold-2.0.pub/mfold.tar.Z @ nrcbsa.bio.nrc.ca`. (Public Domain Software).

[147] M. Zuker. The use of dynamic programming algorithms in RNA secondary structure prediction. In M. S. Waterman, editor, *Mathematical Methods for DNA Sequences*, pages 159–184. CRC Press, 1989.

[148] M. Zuker and D. Sankoff. RNA secondary structures and their prediction. *Bull.Math.Biol.*, 46(4):591–621, 1984.

Table of Contents