Bridging Elementary Landscapes and a Geometric Theory of Evolutionary Algorithms: First Steps

2 authors:

Marcos Diez García
University of Exeter
1 PUBLICATION   **0** CITATIONS

SEE PROFILE

Alberto Moraglio
University of Exeter
73 PUBLICATIONS   **1,146** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   Unifying Fitness Landscape Analysis Approaches of Evolutionary Algorithms View project

Bridging Elementary Landscapes and a Geometric Theory of Evolutionary Algorithms: First Steps

Marcos Diez García and Alberto Moraglio

Department of Computer Science, University of Exeter, UK
{md518,a.moraglio}@exeter.ac.uk

Abstract Based on a geometric theory of evolutionary algorithms, it was shown that all evolutionary algorithms equipped with a geometric crossover and no mutation operator do the same kind of convex search across representations, and that they are well-matched with generalised forms of concave fitness landscapes for which they provably find the optimum in polynomial time [13]. Analysing the landscape structure is essential to understand the relationship between problems and evolutionary algorithms. This paper continues such investigations by considering the following challenge: develop an analytical method to recognise that the fitness landscape for a given problem provably belongs to a class of concave fitness landscapes. Elementary landscapes theory provides analytic algebraic means to study the landscape structure [15]. This work begins linking both theories to better understand how such method could be devised using elementary landscapes. Examples on the well known One Max, Leading Ones, Not-All-Equal Satisfiability and Weight Partition problems illustrate the fundamental concepts supporting this approach.

Keywords: Discrete nodal domain · Elementary landscape · Geometric crossover · Global concavity · Laplacian · P-structure

1 Introduction

Context and Challenge. Since its early days, evolutionary computing (EC) grew rapidly and diversely, but lacking a fundamental and coherent theory [9]. Despite recent progress, still much work needs to be done regarding unification [2,14]. Fitness landscapes have been helpful in that task since they define the search space structure by the fitness function of the problem and the search operators (i.e. the neighbourhood structure), thus linking three key elements of evolutionary algorithms (EAs): problems, algorithms and performance [14]. Particularly, EAs with geometric crossover but no mutation were proved to do an abstract form of convex search [11]. Moreover, this class of EAs is well-matched with globally concave (or convex, if minimising) landscapes, for which runtime exponentially better than pure random search can provably be guaranteed [13].

This paper focuses on fitness landscapes, and extends [13], aiming at developing an analytical method to recognise if a given fitness landscape of some combinatorial problem provably matches a concave (or convex) class of landscapes in general metric spaces. This method would provide insight to determine for which classes of problems and EAs good performance can be guaranteed. We propose elementary landscapes theory [15] as a fruitful choice to develop such method, because the following reasons suggest that certain classes of elementary landscapes relate to classes of (globally) concave landscapes and that elementary landscapes theory is at our disposal to analyse recombination spaces:

– Certain elementary landscapes have all local optima clustered in a single region of the search space. For instance, the landscape of the Linear Assignment problem. Moreover, for Boolean spaces with Hamming distance, it was shown that these elementary landscapes have a unique local optimum, thus coinciding with unimodal landscapes [7,15,18].
– Many combinatorial problems (such as Graph Coloring, Symmetric Travelling Salesman and Graph Bipartitioning) have elementary landscapes with local optima grouped only in a few clusters in the search space, suggesting an approximately globally concave (or convex) structure [8,18].
– Certain recombination spaces (e.g. induced by uniform recombination) are homomorphic to mutation spaces, meaning that their associated landscapes can be compared with each other and that elementary landscapes theory extends to recombination spaces [5,18].

Scope and Contributions. This paper begins linking elementary landscapes and the geometric framework to establish firm grounds to tackle the previous challenge. The major contributions of this work are:

1. Show that from both theories it is possible to restate recombination in terms of the other, that is enabling dual geometric and algebraic views, to justify that is not futile to propose elementary landscapes as an analytic tool.
2. Examples on four well-known combinatorial problems to clarify the connection between the structure of elementary landscapes and the globally concave classes of the geometric framework.

Although these contributions may appear only loosely connected, they complement each other as we will see in the following sections, and their connection will be further developed in future work.

Organisation. Section 2 introduces basic concepts about fitness landscapes. Sections 3 and 4 present key ideas of the geometric framework and elementary landscapes theory, respectively. Section 5 presents the main result of this study. Section 6 illustrates with examples the utility of elementary landscapes to analyse concave landscapes. Section 7 summarises this work and suggests future research.

2 Fitness Landscapes

In combinatorial problems, it is natural to formalise the search space with configuration set X in terms of a move operator $m : X^n \to X^k$ mapping a vector of n parents into a vector of k offspring [16]. Mutation, acting on a single

parent, is naturally connected with a neighbourhood relation that can be described as a connected graph, which can be undirected or directed depending on whether the neighbourhood is symmetric or not. Recombination, acting on pairs of individuals, is non-trivial with different possible formalisations (e.g. using hypergraphs [5]). Besides neighbourhoods, graphs can also be described naturally using graphic distances (e.g. length of shortest paths), which directly relate to metric spaces and more generally fitness landscapes [16].

Definition 1. A *metric space* $\mathcal{M}$ is defined as the pair (X, d), where X is a set of configurations and $d : X \times X \to \mathbb{R}$ is a metric; such that $\forall x, y, z \in X$: (I) non-negativity: $d(x, y) \geq 0$; (II) identity: $d(x, y) = 0 \iff x = y$; (III) symmetry: $d(x, y) = d(y, x)$ and (IV) triangle inequality: $d(x, y) \leq d(x, z) + d(z, y)$.

Definition 2. A *fitness landscape* $\mathcal{F}$ can be defined as the pair $(\mathcal{M}, f)$, where $\mathcal{M}$ is a metric space, and $f : X \to \mathbb{R}$ is a real-valued fitness function that indicates the fitness of each configuration in X, the set of configurations in $\mathcal{M}$.

3 Geometric Framework

Moraglio proposed a general theory of EAs, independent of the problem and representation of solutions, solely based on an axiomatic definition of distance across metric spaces [10]. This allows to formalise mutation and crossover operators in terms of metric balls and geodesic intervals, respectively [21].

Definition 3. Let (X, d) be a metric space with configuration set X and metric d. Then, a *closed ball* centred at point $x \in X$ with radius $r \in \mathbb{R}_{\geq 0}$ is defined as $B_d[x; r] := \{y \in X \mid d(x, y) \leq r\}$ and a *geodesic interval* or *metric segment* as $[x; y]_d := \{z \in X \mid d(x, z) + d(z, y) = d(x, y)\}$, where $x, y \in X$ are called the *extremes* of the segment and $d(x, y)$ its *length*.

Normally, implementations of heuristic search methods define probabilistic search operators, thus a probability distribution over the configuration set. In other words, move operators that return a subset of the possible neighbours. The geometric framework takes this into account when defining geometric operators [10].

Definition 4. Let (X, d) be a metric space and $\text{Im}[\star(\cdot)]$ denote the set of offspring produced with non-zero probability by an operator $\star$. Then, a unary operator $\mu_\varepsilon : X \to X$ is a *geometric ε-mutation* if $\forall x \in X \big(\text{Im}[\mu_\varepsilon(x)] \subseteq B_d[x; \varepsilon]\big)$, where $\varepsilon \in \mathbb{R}_{\geq 0}$ is the smallest non-negative real number for which this condition holds. When ε is not specified, $\varepsilon = 1$ is assumed. And, a binary operator $\chi : X \times X \to X$ is a *geometric crossover* if $\forall x, y \in X \big(\text{Im}[\chi(x, y)] \subseteq [x; y]_d\big)$.

Apart from its benefits regarding unification and formal design of EAs across representations, a geometric definition provides insight on the landscape properties that cause EAs to perform possibly much better than pure random search [10]. Abstract convexity in general metric spaces is one such property [21], which describes the structure induced by geometric crossovers and admits well-behaved generalisations from traditional (Euclidean) concave functions [10,11].

Definition 5. Let a metric space $\mathscr{M} := (X, d)$, with configuration set X and metric d, and a fitness function f define a fitness landscape $\mathscr{F} := (\mathscr{M}, f)$. Then, $\forall x, y, z \in X$, $\mathscr{F}$ is *quasi-concave* if $z \in [x; y]_d$ and $f(z) \geqq \min\big(f(x), f(y)\big)$; and *average-concave* if $z \sim \mathcal{U}\big([x; y]_d\big)$ and $\mathbb{E}\,[f(z)] \geqq \big(f(x) + f(y)/2\big)$. Where $\mathcal{U}$ denotes the uniform probability distribution and $\mathbb{E}$ the expectation.

4 Elementary Landscapes Theory

Based on Grover's work [6], Stadler developed an algebraic theory of fitness landscapes known as elementary landscapes (ELs) [15]. Here, regular undirected[1] graphs induced by mutation neighbourhoods are formalised using the graph or mutation Laplacian matrix $\mathbf{L}_\mu$ defined by the graph's diagonal $\mathbf{D}$ and adjacency $\mathbf{A}$ matrices:

$$-\mathbf{L}_\mu := \mathbf{D} - \mathbf{A} \ . \tag{1}$$

$$-\mathbf{L}_\mu(S_4) = \mathbf{D}(S_4) - \mathbf{A}(S_4) = \begin{pmatrix} 4 & -1 & -1 & -1 & -1 \\ -1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 1 \end{pmatrix}$$
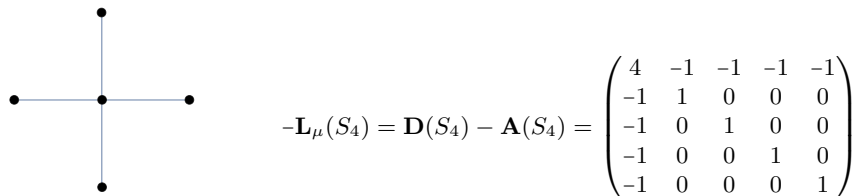
Fig. 1. Star graph S_4 (left) and its graph Laplacian matrix (right).

To formalise recombination neighbourhoods in terms of hypergraphs [5], a new concept called 'P-structure' given by $(X, \mathcal{R})$ was introduced [18], where X is a finite set with power set $\mathcal{P}(X)$ and $\mathcal{R} : X \times X \to \mathcal{P}(X)$ maps a pair of parents into the set of possible offspring. By a hypergraph we mean a graph with vertex set X and hyperedge set $\mathscr{E} := \{\mathcal{R}(x, y) \in \mathcal{P}(X) \mid \mathcal{R}(x, y) \neq \emptyset \wedge x, y \in X\}$. Furthermore, the following subclass of P-structures is defined to capture some desirable properties of recombination operators.

Definition 6. A P-structure $(X, \mathcal{R})$ is a *recombination structure* if $\forall x, y, z \in X$: (I) fix-point: $\mathcal{R}(x, x) = \{x\}$; (II) symmetry: $\mathcal{R}(x, y) = \mathcal{R}(y, x)$; (III) null-recombination: $\{x, y\} \subseteq \mathcal{R}(x, y)$; and (IV) size-monotonicity: if $z \in \mathcal{R}(x, y)$, then $|\mathcal{R}(x, z)| \leqq |\mathcal{R}(x, y)|$.

Laplacians $\mathbf{L}_\mathcal{R}$ for recombination P-structures can also be defined based on the identity matrix $\mathbf{I}$ and a generalised adjacency matrix $\mathbf{S}$ for hypergraphs [18]:

$$-\mathbf{L}_\mathcal{R} := 2\,|X|\,\mathbf{I} - \mathbf{S} \ , \tag{2}$$

---

[1] Extensions for non-regular and non-symmetric neighbourhoods are possible, but regular and symmetric ones will be assumed here.

where the square matrix $\mathbf{S}$ has entries $\mathbf{S}_{zx} := 2\sum_{y \in X} \mathbf{H}_{z,(x,y)} |\mathcal{R}(x,y)|^{-1}$, and the non-square binary incidence matrix $\mathbf{H}$ with entries $\mathbf{H}_{z,(x,y)}$ tells whether z is offspring of x and y under $\mathcal{R}$.

To better grasp the generalised adjacency matrix $\mathbf{S}$ in (2), one may consider its associated stochastic matrix $\mathbf{T}_{zx}$ describing the transition probabilities of a recombination-based random walk on a graph, where a father x is mated with a randomly chosen mother y to produce an offspring z that will be the father in the next recombination [18]. It is defined as $\mathbf{T}_{zx} := \sum_{y \in X} t_{xy}^z p_y$, where p_y is the probability of choosing y from X, and t_{xy}^z is the probability of z given x and y. $\mathbf{T}_{zx}$, $\mathbf{S}$ and $\mathbf{L}_{\mathcal{R}}$ relate as: $\mathbf{T}_{zx} = \sum_{y \in X} t_{xy}^z p_y = \sum_{y \in X} \mathbf{H}_{z,(x,y)} |\mathcal{R}(x,y)|^{-1} * \frac{1}{|X|} = \frac{1}{2|X|}\mathbf{S}_{zx} = \frac{1}{2|X|}\mathbf{L}_{\mathcal{R}} + \mathbf{I}$. Although this assumes a uniform population and that all offspring occur with equal probability, it is possible to define weights for $\mathbf{S}$ to formalise more realistic population recombination-based search algorithms [17].

In the light of Grover's work and the characterisations above, Stadler defines a non-flat fitness landscape as *elementary* if its zero-averaged fitness function (as a column vector) $\tilde{f} := f - \bar{f} = \left[f(x_1) - \bar{f}, f(x_2) - \bar{f}, \ldots, f(x_n) - \bar{f} \right]^{\mathrm{T}}$, with finite discrete domain $X = \{x_1, x_2, \ldots, x_n\}$ representing the vertex set of an underlying connected graph, is an eigenfunction of a (generalised) Laplacian matrix $\mathbf{L}$ of the graph: $\mathbf{L}\tilde{f} = \lambda\tilde{f}$; where $\lambda > 0$ is the eigenvalue and the constant $\bar{f} := \frac{1}{|X|} \sum_{x \in X} f(x)$ is the average fitness of a configuration in X [8,15]. If a landscape f is not elementary, it can always be decomposed into a linear combination of ELs f_k called its *Fourier expansion*: $f = a_0 + \sum_{k>0}^{|X|-1} a_k f_k$, where scalars a_k are the Fourier coefficients, a_0 is the average fitness and the eigenfunctions f_k have corresponding eigenvalues λ_k (in increasing order and counting multiplicities, $0 = \lambda_0 < \lambda_1 \leqq \lambda_2 \leqq \cdots \leqq \lambda_{|X|-1}$) such that $\mathbf{L}f_k = \lambda_k f_k$.

ELs have interesting geometric properties. One of them, proved by Grover [6], is that all local maxima (minima) have fitness higher (lower) than or equal to the average fitness: $f(x_{min}) \leqq \bar{f} \leqq f(x_{max})$. This property is closely related to the idea of monotonic sequences of local optima (i.e. sequences of ever non-decreasing or non-increasing fitness) used, for instance, in the local optima networks model [20] to formalise the funnel structure of landscapes.

Another important aspect of ELs is their global structure. There are two major classes depending on the index p (ignoring multiplicities) of the eigenvalue: Fujiyama or single-peaked [7], and non-Fujiyama. Fujiyama are those ELs with the smallest non-zero eigenvalue (λ_1), that is $\mathbf{L}\tilde{f} = \lambda_{p=1}\tilde{f}$; characterised for having all local optima clustered in a single region of the space [18]. In particular, for Boolean spaces with Hamming distance they have a unique global optimum [15]. Non-Fujiyama are those ELs for $p > 1$ with local optima clustered in more than one region; indeed, many combinatorial problems fall here with $p = 2$ (e.g. Max Cut) [8]. More formally, by clusters we mean *discrete nodal domains* [1]. These are the maximally connected subgraphs induced by the vertex subsets $V_+ := \{x \in V \mid f(x) \geqq 0\}$ and $V_- := \{x \in V \mid f(x) \leqq 0\}$, denoting the weak positive and negative *nodal sets* respectively, for a given graph with vertex set V and eigenfunction f (i.e. the fitness function) on V. Similarly, strong

nodal sets are defined using a strict inequality instead. Note that, for a zero-averaged function, V_+ and V_- induce subgraphs separating precisely above and below average configurations. Interestingly, the number of discrete nodal domains can be upper-bounded a priori, provided that we know p, using Proposition 1, generally, and the more sharp Proposition 2 specifically for Boolean spaces; which were proved in [1,4].
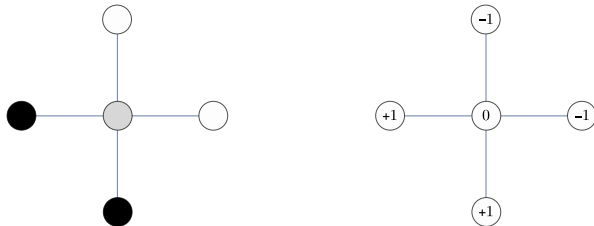

Fig. 2. An eigenfunction defined on the star graph S_4 (right), with eigenvalue $\lambda_1 = 1$ of the graph Laplacian, and its induced discrete nodal domains (left): two weak and four strong. Positive nodes in 'black', negative nodes in 'white' and zero nodes in 'grey'.

Proposition 1. Given a generalised Laplacian of a connected graph, any eigenfunction f_k corresponding to the k-th eigenvalue λ_k with multiplicity r has at most k weak and $k + r - 1$ strong nodal domains.

Proposition 2. For an eigenfunction f with eigenvalue $2p$, where p is its index ignoring multiplicities and N the number of vertices of the Boolean hypercube, the number $W(f)$ of weak nodal domains is upper-bounded as:

$$W(f) \leqq w_{N,p} = \begin{cases} p + 1 & \text{if } p = 0 \text{ or } p = 1, \\ 2\left(1 + \displaystyle\sum_{k=0}^{p/2-1} \binom{N}{2k}\right) & \text{if } p \text{ is even}, \\ 2\left(1 + \displaystyle\sum_{k=0}^{\lfloor p/2 \rfloor - 1} \binom{N}{2k+1}\right) & \text{if } p \text{ is odd}. \end{cases} \tag{3}$$

5 Main Results

To analyse the landscapes induced by geometric crossovers (Definition 4) using ELs theory, it is necessary to prove that geometric crossovers belong to the class of recombination P-structures (Definition 6). Connecting both theories (Sects. 3 and 4) can help to develop dual geometric and algebraic views with which tackle the challenge of identifying generalised concave landscape classes using ELs theory, and give insight on explaining the good performance behind convex search (Sect. 1). Next, we prove that all geometric crossovers, regardless of the metric, are recombination P-structures.

Lemma 1. Let (X, d) be any metric space. Then, $\forall x, y \in X$ and any metric d, $(X, [x, y]_d)$ is a recombination P-structure.

Proof. We need to prove that metric segments fulfil the axioms of recombination P-structures.

(I) *Fix-point.* The only possible z in $[x; x]_d = \{z \in X \mid d(x, z) + d(z, x) = d(x, x)\}$ is exactly x. Therefore, $[x; x]_d = \{x\}$.

(II) *Symmetry.* $[x; y]_d = [y; x]_d$ follows immediately from the symmetry axiom of metric segments.

(III) *Null-recombination.* $\{x, y\} \subseteq [x; y]_d$ holds by definition, since the extremes x and y of the segment are always included in the segment.

(IV) *Size-monotonicity.* To prove that if $z \in [x; y]_d$, then $\big|[x; z]_d\big| \leqq \big|[x; y]_d\big|$, we can recall that all metric segments fulfil monotonicity [21]: $\forall x, y, z \in X$ if $z \in [x; y]_d$ then $[x; z]_d \subseteq [x; y]_d$. Therefore, it follows $\big|[x; z]_d\big| \leqq \big|[x; y]_d\big|$. $\square$

Unfortunately, to prove equivalence in the other direction it is necessary to restrict recombination P-structures to those that precisely produce offspring $z \in \mathcal{R}(x, y)$ lying in the geodesic interval between parents x and y, since in general not all recombination P-structures fulfil this property [3].

Example 1. Consider $\mathcal{R}_\ell$ returning offspring lying on longest paths between parents. This can generate offspring beyond the geodesic interval between parents (i.e. in the extension ray [21]), e.g. $[000; 011]_{d_\mathrm{H}} \not\ni 111 \in \mathcal{R}_\ell(000, 011)$ where d_H is the Hamming distance. Indeed, extended line recombination, which generates offspring in the extension ray, is not a geometric crossover for any metric [12].

Definition 7. A recombination P-structure $(X, \mathcal{R})$, for a connected graph G with vertex set X, that $\forall x, y \in X$ fulfils the axiom $\mathcal{R}(x, y) \subseteq I(x, y)$, where $I(x, y)$ is the set of all shortest paths between x and y in G, is a *geometric recombination P-structure* denoted by $\mathcal{R}_\mathrm{g}$.

Theorem 1. Let (X, d) be any graphic metric space. Then, all geometric recombination P-structures $\mathcal{R}_\mathrm{g}$ are equivalent to all geometric crossovers χ defined on (X, d). That is, $\forall x, y \in X \big(\mathcal{R}_\mathrm{g}(x, y) = \chi(x, y)\big)$.

Proof. The proof follows immediately from Lemma 1, and Definitions 4 and 7 of geometric crossovers and geometric recombination P-structures respectively. $\square$

Interestingly, all recombination P-structures fulfil the inbreeding properties (Theorem 2), that is breeding between close relatives, common to all geometric crossovers and independent of the metric used, which were proposed as a test for non-geometricity of crossovers: if any of them fails, then a crossover is not geometric [12].

Theorem 2. Let $(X, \mathcal{R})$ be any recombination P-structure. Then, $\mathcal{R}$ satisfies all inbreeding properties of geometric operators: purity, convergence and partition.

Proof. Purity: The recombination of one parent with itself can only produce the parent itself. Follows immediately from the fix-point axiom of recombination P-structures: $\forall x \in X, \mathcal{R}(x, x) = \{x\}$.

Convergence: The recombination of one parent with one offspring cannot produce the other parent of that offspring, unless the offspring and the second parent coincide. We want to prove that $\forall x, y, z, s \in X$ if $z \in \mathcal{R}(x,y)$ and $s \in \mathcal{R}(x,z)$, then $s = y \implies z = y$. Let $z \in \mathcal{R}(x,y)$ such that $z \neq y$, and $s \in \mathcal{R}(x,z)$. We want to show that actually $s \neq y$ too. From the size-monotonicity axiom of recombination structures we know:

$$|\mathcal{R}(x,z)| < |\mathcal{R}(x,y)| \quad , \tag{4}$$

since $y \neq z \in \mathcal{R}(x,z)$. Besides, either $s = z$ or $s \neq z$. If $s = z$, then we know automatically that $s \neq y$, since $s = z \neq y$. If on the other hand $s \neq z$, then:

$$|\mathcal{R}(x,s)| < |\mathcal{R}(x,z)| \quad . \tag{5}$$

If s was allowed to be y, then $|\mathcal{R}(x,s)| = |\mathcal{R}(x,y)|$, however by (5):

$$|\mathcal{R}(x,s)| = |\mathcal{R}(x,y)| < |\mathcal{R}(x,z)| \tag{6}$$

thus contradicting (4). Therefore, $s \neq y$ for the $s \neq z$ case. Consequently, the only possibility left for $s = y$ is that $z = y$.

Partition: If z is a child of x and y, then the recombination of x and z, and the recombination of y and z, cannot produce a common grandchild s other than z. Another way to phrase it is that both recombinations must produce two different grandchildren when they are not z. That is, we want to prove that $\forall x, y, z, s_1, s_2 \in X$ if $z \in \mathcal{R}(x,y)$, $s_1 \in \mathcal{R}(x,z)$ and $s_2 \in \mathcal{R}(z,y)$, then $s_1 \neq s_2$. Without loss of generality, let $z \in \mathcal{R}(x,y)$ such that $z \neq y$, and assume the opposite: $s_1 = s_2$. Then, we know by the convergence property that $s_1 \neq y$. Analogously, exchanging the roles of the parents x and y using the symmetry axiom of recombination structures, we know $s_2 \neq x$. Observe, however, that there are no restrictions in having $s_1 = x$ and $s_2 = y$. But, since we assumed $s_1 = s_2$, we find the following contradictions: $x = s_1 = s_2 \neq x$, and $y = s_2 = s_1 \neq y$. Therefore, $s_1 \neq s_2$. □

Since all recombination P-structures fulfil the inbreeding properties but only geometric recombination P-structures are equivalent to geometric crossovers (Theorem 1), from Theorem 2 we conclude the following.

Corollary 1. The inbreeding properties of geometric crossovers are necessary but not sufficient conditions to determine if a crossover is geometric.

This resolves the question of whether the inbreeding properties are sufficient or not for geometricity, which was left open in [12]. In the light of this result, it would be worth identifying additional inbreeding properties that when considered jointly would be sufficient to guarantee that a crossover is geometric.

6 Discrete Nodal Domains for Uniform Recombination in Boolean Spaces with Hamming Distance: Examples

Section 5 justified that proposing ELs theory to analyse the abstract concave classes of the geometric framework is not futile by showing that geometric

crossovers are a specific case of recombination P-structures, and that certain recombination P-structures coincide with geometric crossovers; thus motivating a dual interpretation on landscapes induced by recombination. This section clarifies, with illustrative examples, how discrete nodal domains (Sect. 4) help analyse the structure of landscapes induced by uniform recombination in Boolean spaces with Hamming distance in particular, and how they link to global concavity. The examples considered are on two artificial problems, One Max and Leading Ones, and two NP-complete problems [6]:

Not-All-Equal Satisfiability (NAES). An instance is a set of clauses and its fitness is the number of satisfied clauses. A clause consists of three literals (i.e. a binary variable or its complement) not containing simultaneously a binary variable and its complement, and it is satisfied if there are at least two distinct literals such that one is 0 and the other is 1.

Weight Partition (WP). An instance is a vector of n weights w_i corresponding to objects o_i, for $1 \leq i \leq n$. The problem consist in finding an assignment $s_i \in \{-1, 1\}$ of weights such that $f(w_1, \ldots, w_n) = \left(\sum_{i=1}^{n} w_i s_i\right)^2$ is minimised.

Figure 3, obtained using `Mathematica`[2], summarises key aspects of the problems' landscape structure. Although the examples considered are for the three-dimensional Boolean hypercube embedded in the hypergraph induced by uniform recombination [18], we have observed equivalent results in higher dimensions.

Next, we see how the global landscape structure could be inferred analytically. For ELs this is possible by knowing the corresponding eigenvalues, since from their indexes we can tell whether they are Fujiyama or non-Fujiyama and also upper-bound the number weak nodal domains using Proposition 2 (Sect. 4). Then, we will explain why Leading Ones poses certain problems for this analytical approach.

Eigenvalues of elementary landscapes. To find them we need to know the uniform recombination Laplacian $\mathbf{L}_\mathcal{R}$ for the three-dimensional hypercube (see Lemma C7 [18]): $(\mathbf{L}_\mathcal{R})_{xy} = 2(2^3)\mathbf{I}_{xy} - \left(2\,(3/2)^3\,3^{-d_\mathrm{H}(x,y)}\right)$. We observe that this Laplacian has eigenvalues $\{0, 8, 8, 8, 12, 12, 12, 14\}$ (in increasing order and starting at index 0); and that One Max is elementary for $\lambda_1 = 8$, and NAES and WP for $\lambda_2 = 12$: $\mathbf{L}_\mathcal{R}\tilde{f}_\mathrm{OneMax} = 8 * \tilde{f}_\mathrm{OneMax}$, $\mathbf{L}_\mathcal{R}\tilde{f}_\mathrm{NAES} = 12 * \tilde{f}_\mathrm{NAES}$ and $\mathbf{L}_\mathcal{R}\tilde{f}_\mathrm{WP} = 12 * \tilde{f}_\mathrm{WP}$.

One Max, NAES and WP. From the eigenvalues we find that One Max is a Fujiyama EL ($p = 1$), that is with a unique global optimum and single-peaked. This agrees with the fact that $f_\mathrm{OneMax}(x_1, \ldots, x_n) \coloneqq \sum_{i=1}^{n} x_i$ is a linear pseudo-Boolean function whose Fourier expansion using Walsh functions has only non-zero terms in the 0-th and 1-st orders thus elementary for $p = 1$ [19]. Whereas NAES and WP are non-Fujiyama ELs of order $p = 2$. These results are a consequence of Corollaries 2 and 3 in [18], proving that eigenfunctions of the mutation Laplacian (1) are also eigenfunctions of the uniform recombination Laplacian (2). Besides, one notes that the eigenvalue ($\lambda_2 = 12$) for NAES
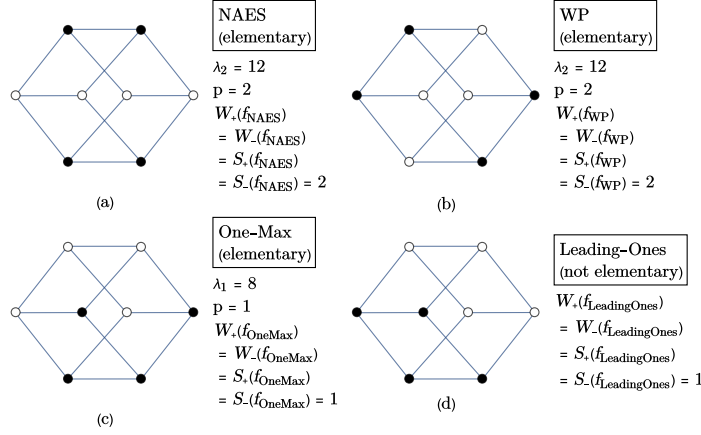

Fig. 3. Strongly positive ('black dots') and negative ('white dots') discrete nodal domains for the (a) Not-All-Equal satisfiability (NAES), (b) Weight Partitioning (WP), (c) One Max and (d) Leading Ones problems on the three-dimensional hypercube. W_+, S_+, W_- and S_- denote the number of weakly positive, strongly positive, weakly negative and strongly negative domains respectively; and p denotes the order of the eigenvalue (λ_p) for each of the eigenfunctions f_{NAES}, f_{WP}, f_{OneMax} and $f_{\text{LeadingOnes}}$.

and WP does not coincide with the eigenvalue ($\lambda_2 = 4$) obtained for the mutation Laplacian [8]. Nevertheless, that is expected: two spaces may have identical eigenfunctions but different eigenvalues, possibly affecting the correlation between fitness values (i.e. ruggedness) of the landscape [18]. Thus, two landscapes may have identical number of nodal domains (i.e. same global structure) and different ruggedness.

Number of weak nodal domains. Proposition 2 provides upper-bounds for the number of weak nodal domains; for One Max, NAES and WP we have: $W_+(\tilde{f}_{\text{OneMax}}) + W_-(\tilde{f}_{\text{OneMax}}) = 1 + 1 \leqq w_{2^3,1} = 2$, $W_+(\tilde{f}_{\text{NAES}}) + W_-(\tilde{f}_{\text{NAES}}) = 2 + 2 \leqq w_{2^3,2} = 4$ and $W_+(\tilde{f}_{\text{WP}}) + W_-(\tilde{f}_{\text{WP}}) = 2 + 2 \leqq w_{2^3,2} = 4$. This means that One Max has two weak nodal domains corresponding to two major clusters of local optima, separated by the hyperplane described by the average fitness. NAES and WP have at most four weak nodal domains and thus we cannot expect more than four clusters. Particularly, for $p = 1$ and $p = 2$, the upper-bounds remain constant regardless of the dimension (Table 4.1 in [1]).

Leading Ones. We observed in `Mathematica` that Leading Ones has the same number of discrete nodal domains as One Max by computing them. From Fig. (3d) one may think that Leading Ones is an EL for $p = 1$; however, it is not elementary neither for recombination nor mutation neighbourhoods, but a sum of n elementary landscapes. The reason is that $f_{\text{LeadingOnes}}(x_1, \ldots, x_n) := \sum_{i=1}^{n} \prod_{j=1}^{i} x_j$ is a k-bounded pseudo-Boolean function [19] that is a sum of Walsh functions where the bitwise non-linearity is at most k bits. For Leading Ones $k = n$, thus the highest order is $p = n$ (i.e. the bit string length). This means that we cannot use Propositions 1 and 2 to know a priori the number of

discrete nodal domains, since in general they do not hold when the landscape is a sum of ELs. Finding those cases where they hold is an open problem [1].

Concluding, discrete nodal domains are a promising analytic tool with which tackle the challenge of identifying landscape classes, because they capture key information of the landscape structure and can be related to ELs via eigenvalues and eigenfunctions. Then, what the abstract concave classes would correspond to? Previous observations suggest Fujiyama ELs ($p = 1$), since they overlap with single-peaked landscapes. For instance, One Max can be shown to be average-concave [13] and we now know that One Max is Fujiyama. Further research will confirm whether this intuition is correct.

7   Conclusions

The geometric framework has been successful in finding proper subclasses of EAs and fitness landscapes with provably polynomial runtime guarantees, however missing analytical means for landscape structure analysis. Besides, ELs theory provides a fine-grained analysis of the landscape structure in terms of its spectrum (i.e. eigenvalues and eigenfunctions) and discrete nodal domains, but lacking EAs dynamics modelling and runtime results. The ultimate goal of this research is precisely unifying these two frameworks, aiming at developing complementary geometric and algebraic views on fitness landscape analysis, to better understand when and why EAs perform well.

This paper took the first steps towards such goal by putting together for the first time both frameworks and highlighting their key ideas. First, proving that all geometric crossovers can be conceived as recombination P-structures, and that there exists a specific subclass of these that match geometric crossovers (Sect. 5). Then, illustrating and clarifying with examples how discrete nodal domains could help to identify concave landscape classes for problems that are elementary, by analysing their spectrum (Sect. 6). The next important steps would be to formalise the concave classes in terms of discrete nodal domains and convex search in ELs theory. Also, it would be worth researching whether landscape funnels can be formalised in terms of discrete nodal domains.

References

1. Bıyıkoğlu, T., Leydold, J., Stadler, P.F.: Laplacian Eigenvectors of Graphs: Perron-Frobenius and Faber-Krahn Type Theorems, Lecture Notes in Mathematics, vol. 1915. Springer-Verlag (2007)
2. Borenstein, Y., Moraglio, A. (eds.): Theory and Principled Methods for the Design of Metaheuristics. Natural Computing Series, Springer (2014)
3. Changat, M., Narasimha-Shenoi, P.G., Nezhad, F.H., Kovše, M., Mohandas, S., Ramachandran, A., Stadler, P.F.: Topological Representation of the Transit Sets of k-Point Crossover Operators. arXiv:1712.09022 (2017)
4. Davies, E.B., Gladwell, G.M., Leydold, J., Stadler, P.F.: Discrete nodal domain theorems. Linear Algebra and its Applications **336**(1), 51–60 (2001), Elsevier Science Inc.

5. Gitchoff, P., Wagner, G.P.: Recombination Induced Hypergraphs: A New Approach to Mutation-Recombination Isomorphism. Complexity **2**(1), 37–43 (1996), John Wiley & Sons, Inc.
6. Grover, L.K.: Local search and the local structure of NP-complete problems. Operations Research Letters **12**(4), 235–243 (1992), Elsevier Science Publishers B. V.
7. Kauffman, S.A.: The Origins of Order: Self-organization and Selection in Evolution. Oxford University Press, New York, New York, USA (1993)
8. Klemm, K., Stadler, P.F.: Rugged and Elementary Landscapes. In: Borenstein, Y., Moraglio, A. (eds.) Theory and Principled Methods for the Design of Metaheuristics, Natural Computing Series, vol. 378, pp. 41–61. Springer (2014)
9. Menon, A. (ed.): Frontiers of Evolutionary Computation, Genetic Algorithms and Evolutionary Computation, vol. 11. Springer US (2004)
10. Moraglio, A.: Towards a Geometric Unification of Evolutionary Algorithms. Doctoral thesis, University of Essex, Essex, UK (November 2007)
11. Moraglio, A.: Abstract Convex Evolutionary Search. In: Proceedings of the 11th Workshop on Foundations of Genetic Algorithms. pp. 151–162. FOGA '11, ACM, Schwarzenberg, Austria (2011)
12. Moraglio, A., Poli, R.: Inbreeding Properties of Geometric Crossover and Non-geometric Recombinations. In: Proceedings of the 9th International Conference on Foundations of Genetic Algorithms. pp. 1–14. FOGA '07, Springer, Mexico City, Mexico (2007)
13. Moraglio, A., Sudholt, D.: Principled Design and Runtime Analysis of Abstract Convex Evolutionary Search. Evolutionary Computation **25**(2), 205–236 (2017), MIT Press
14. Richter, H., Engelbrecht, A. (eds.): Recent Advances in the Theory and Application of Fitness Landscapes, Emergence, Complexity and Computation, vol. 6. Springer-Verlag, 1st edn. (2014)
15. Stadler, P.F.: Towards a Theory of Landscapes. In: López-Peña, R., Waelbroeck, H., Capovilla, R., García-Pelayo, R., Zertuche, F. (eds.) Complex Systems and Binary Networks, Lecture Notes in Physics, vol. 461, pp. 78–163. Springer (1995)
16. Stadler, P.F.: Fitness landscapes. In: Lässig, M., Valleriani, A. (eds.) Biological Evolution and Statistical Physics, Lecture Notes in Physics, vol. 585, pp. 183–204. Springer-Verlag (2002)
17. Stadler, P.F., Seitz, R., Wagner, G.P.: Population Dependent Fourier Decomposition of Fitness Landscapes over Recombination Spaces: Evolvability of Complex Characters. Bulletin of Mathematical Biology **62**(3), 399–428 (2000), Society for Mathematical Biology
18. Stadler, P.F., Wagner, G.P.: Algebraic Theory of Recombination Spaces. Evolutionary Computation **5**(3), 241–275 (1998), MIT Press
19. Sutton, A.M., Whitley, D., Howe, A.E.: Mutation Rates of the (1+1)-EA on Pseudo-Boolean Functions of Bounded Epistasis. In: Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation. pp. 973–980. GECCO '11, ACM, New York, New York, USA (2011)
20. Thomson, S.L., Daolio, F., Ochoa, G.: Comparing Communities of Optima with Funnels in Combinatorial Fitness Landscapes. In: Proceedings of the Genetic and Evolutionary Computation Conference. pp. 377–384. GECCO '17, ACM, New York, New York, USA (2017)
21. van de Vel, M.L.J.: Theory of Convex Structures. North-Holland Mathematical Library, North-Holland (1993)