A Gaussian Random Field Model of Smooth Fitness Landscapes

Alberto Moraglio
Centre for Informatics and Systems
of the University of Coimbra
University of Coimbra - Polo II
Coimbra 3030-290, Portugal
moraglio@dei.uc.pt

Yossi Borenstein
Department of Computing
and Electronic Systems
University of Essex
Colchester CO4 3SQ, UK
yboren@essex.ac.uk

ABSTRACT

The smoothness of a fitness landscape, to date still an elusive notion, is considered to be a fundamental empirical requirement to obtain good performance for many existing meta-heuristics. In this paper, we suggest that a theory of smooth fitness landscapes is central to bridge the gap between theory and practice in EC. As a first step towards this theory, we formalize the notion of smooth fitness landscapes in a general setting using a Gaussian random field model on metric spaces. Then, for the specific case of the Hamming space, we show experimentally that traditional search algorithms with search operators based on this space reach better performance on smoother fitness landscapes. This shows that the formalized notion of smoothness captures the important heuristic property of its informal counterpart.

Categories and Subject Descriptors

I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search; G.1.6 [**Numerical Analysis**]: Optimization; F.2 [**Analysis of Algorithms and Problem Complexity**]

General Terms

Theory

1. INTRODUCTION

For the No Free Lunch (NFL) theorems [25] (see also [22] [9] for recent developments), the expected performance of any black-box search algorithm averaged over all functions is the same as random search. However, in practice the scenario of the NFL theorem is not realistic because search algorithms are hardly applied as black-box search algorithms. Rather, they are used as meta-heuristics (intended as algorithmic templates as opposed to ready-made search algorithms) and are adapted to the problem at hand at design time. Hence, they incorporate problem knowledge. It is not clear, however, what accounts for problem knowledge and what is the link between problem knowledge and better-than-random-search performance. In the following, we depict a scenario which links problem knowledge, solution representations, search operators, smoothness of fitness landscape, and performance. This scenario forms the context for the focus of the present paper.

A well-known general empirical heuristic to obtain good performance for existing meta-heuristics is to choose search operators associated with a neighborhood structure, or a distance, that gives rise to a fitness landscape with a smooth trend in which closer solutions are more likely to have similar fitness values [18].

The above heuristic presupposes the notion of association between distance, or neighbourhood structure, and corresponding search operator being well-defined. In previous work within the scope of evolutionary algorithms [15], we made precise this notion for mutation and crossover operators for the most frequently used solution representations using simple, representation-independent geometric definitions. We then showed experimentally that this way of associating distances and search operators is consistent with the above heuristics. This method is very general and it can be used to link distances to a variety of other search operators. For example, we have extended it to search operators used in particle swarm optimization [16].

When the designer of the algorithm knows the objective function of the problem at hand, the smoothness of the fitness landscape can be *controlled*. This could be done, for example, by choosing moves as a base for a local search algorithm which are associated to fitness landscapes which are Lipschitz continuous with a small Lipschitz constant by construction. When the problem at hand is unknown to the designer, however, the NFL theorem applies because there is no *a priori* choice of search operator which would be preferable in terms of giving rise to a smooth landscape. This perspective casts light on the nature of problem knowledge and the role of the designer. It equates problem knowledge passed by the designer to the search algorithm with the requirement of smoothness of the fitness landscape to obtain good performance for evolutionary algorithms.

This line of thinking allows us to depict a scenario which bridges the gap between theory and practice in evolutionary computation as follows. Ideally, a theory of smooth fitness landscapes would be *representation-independent*. It would express the performances of search algorithms as a function

FOGA'09, January 9–11, 2009, Orlando, Florida, USA.

of important and controllable design elements such as solution representation and search operators. These elements would be able to be expressed in the theory as variables, rather than fixed elements, in an indirect way using *variable metrics* associated with them. This theory could be used by the designer to make an informed choice on the solution representation and search operators choosing those combinations which are guaranteed by the theory to be *a priori* good. The good combinations of solution representation and search operators would be those whose associated distance gives rise to a smooth fitness landscape, in a well-defined sense, for the problem at hand. So, the theory would be able to predict the performances of a search algorithm with a given combination of solution representation and search operators from the level of smoothness of fitness landscape they induce, without the need to run any experiment.

Translating the scenario outlined above in an actual theory is, naturally, a long term project, which builds upon the geometric unification of evolutionary algorithms [15] which associates solution representation and geometric operators with the distances of the search space they induce. The underlying assumption behind the above scenario is that the smoothness of the fitness landscape is what ultimately determines good performance of traditional search algorithms. In this paper, we present a step towards this theory. We formalize the notion of smooth fitness landscape in a general way using a Gaussian Random Field (GRF) model and show experimentally that this notion of smoothness can *alone* determine average-case performance much better than random search for a number of traditional search algorithms. The GRF model is a natural choice to formalize the notion of smoothness as outlined above because it relates distance between solutions in the search space with the correlation of their fitness values.

There is some previous work on using statistical models of fitness landscapes in the context of evolutionary algorithms. Their use is quite heterogeneous. Perhaps, the first statistical model of random landscapes used in evolutionary computation is the NK-landscapes introduced by Kauffman [12]. This family of fitness landscapes is based on binary strings and the model has a parameter to tune the ruggedness of the fitness landscapes generated. Stadler has introduced a general mathematical theory for the analysis of random fitness landscapes based on algebraic combinatorics [20] [23]. There are empirical statistical measures of hardness of fitness landscapes based on correlation, such as random walk autocorrelation [24] and fitness-distance correlation [11] which, however, do not consider explicitly any underlaying statistical model. Estimation of distribution of algorithms [14] search alternating sampling and re-calibrating a statistical model of the fitness landscape. The specific model used depends heavily on the underlying solution representation. In continuous global optimization, GRF models are used to do statistical inference and determine the location of solutions whose fitness values are most likely to improve on the fitness values of already sampled solutions [10].

There is much work on GRFs in literature. However, most of the pre-existing work focuses on continuous spaces endowed with the Euclidean distance. See, for example, the book [5] for their use in geostatistics which is the major field of application of these models. GRFs can be readily generalized to general metric spaces. In this paper, we consider GRFs at this level of generality and we study the issues arising in this general setting. In particular, we show that GRFs suit combinatorial spaces as good as they suit continuous spaces.

In the experiments, we concentrate on GRFs on the Hamming space. We sample the GRF model and produce random fitness landscapes with varying level of smoothness. We then test a number of search algorithms and show that, on average, the smoother the fitness landscape, the better the performance of these algorithms. This validates experimentally the consistency of our formalized model of smoothness with the expected performance of search algorithms in full agreement with the heuristic that smoother fitness landscapes lead to better performances. This is an interesting result because it shows that, in the average-case, the smoothness of the fitness landscape is ultimately what is needed to allow the traditional search algorithms considered to perform much better than random search.

These results are average-case. We discuss how these results relate with result on a single typical fitness landscape. This will allow us to clarify the interpretation of existing empirical statistical measure of smoothness of a fitness landscape to predict its hardness and explain the origin of counter-examples to these measures.

In summary, the remainder of this paper is organized as follows. In section 2, we briefly review the geometric framework. In section 3, we present preliminary definitions and results about gaussian random fields on general metric spaces and, in section 4, we focus on the specific case of the Hamming space which is associated with the traditional crossover and mutation operators for binary strings. In section 5, we define smoothness for fitness landscapes under any metric space. In section 6, we present experimental results for varying level of smoothness. In section 7, we discuss the connection between GRFs and statistical measures hardness of fitness landscape. In section 8, we present conclusions and future work.

2. GEOMETRIC FRAMEWORK

2.1 Geometric preliminaries

In the following we give necessary preliminary geometric definitions. For more details on these definitions see [7].

The terms *distance* and *metric* denote any real valued function that conforms to the axioms of identity, symmetry and triangular inequality. A simple connected graph is naturally associated to a metric space via its *path metric*: the distance between two nodes in the graph is the length of a shortest path between the nodes. Distances arising from graphs via their path metric are called *graphic distances*. Similarly, an edge-weighted graph with strictly positive weights is naturally associated to a metric space via a *weighted path metric*.

In a metric space (S, d) a *closed ball* is a set of the form $B_d(x; r) = \{y \in S | d(x, y) \leq r\}$ where $x \in S$ and r is a positive real number called the radius of the ball. A *line segment* (or closed interval) is a set of the form $[x; y]_d = \{z \in S | d(x, z) + d(z, y) = d(x, y)\}$ where $x, y \in S$ are called extremes of the segment. Metric ball and metric segment generalize the familiar notions of ball and segment in the Euclidean space to any metric space through distance redefinition. These generalized objects look quite different under different metrics. Notice that the notions of metric segment and shortest path connecting its extremes (*geodesic*) do not coincide as it happens in the specific case of an Euclidean space. In general, there may be more than one geodesic connecting two extremes; the metric segment is the union of all geodesics.

We assign a structure to the solution set S by endowing it with a notion of distance d. $M = (S, d)$ is therefore a solution *space* (or search space) and $L = (M, g)$ is the corresponding *fitness landscape* where $g : S \to \mathbf{R}$ is the fitness function. Notice that in principle d could be arbitrary and need not have any particular connection or affinity with the search problem at hand.

2.2 Geometric crossover

The geometric framework [15] defines search operators in a *representation-independent* way expressing search operators entirely as functions of the metric d associated with the search space, using simple geometric definitions. In the following, as an example, we briefly review one of these search operators, the geometric crossover [17], which is based on the notion of metric segment.

A recombination operator OP takes parents p_1, p_2 and produces one offspring c according to a given conditional probability distribution:

$Pr\{OP(p_1, p_2) = c\} = Pr\{OP = c|P_1 = p_1, P_2 = p_2\} = f_{OP}(c|p_1, p_2)$

Definition 1. *(Image set) The* image set $Im[OP(p_1, p_2)]$ *of a genetic operator OP is the set of all possible offspring produced by OP with non-zero probability when parents are p_1 and p_2.*

Definition 2. *(Geometric crossover) A recombination operator CX is a geometric crossover under the metric d if all offspring are in the segment between its parents: $\forall p_1, p_2 \in S : Im[CX(p_1, p_2)] \subseteq [p_1, p_2]_d$*

Definition 3. *(Uniform geometric crossover) The uniform geometric crossover UX under d is a geometric crossover under d where all z laying between parents x and y have the same probability of being the offspring:*

$$\forall x, y \in S : f_{UX}(z|x, y) = \frac{\delta(z \in [x; y]_d)}{|[x; y]_d|}$$

$$Im[UX(x, y)] = \{z \in S | f_{UX}(z|x, y) > 0\} = [x; y]_d$$

where δ is a function that returns 1 if the argument is true, 0 otherwise.

A number of general properties for geometric crossover and mutation have been derived in [17].

Many pre-existing recombination operators are geometric crossovers [15], under a suitable metric. Some recombination operators are provably not geometric crossovers (under any metric). For vectors of reals, various types of blend or line crossovers, box recombinations, and discrete recombinations are geometric crossovers. For binary and multary strings (fixed-length strings based on a n symbols alphabet), all mask-based crossovers (one point, two points, n-points, uniform) are geometric crossovers. For permutations, PMX, Cycle crossover, merge crossover and others are geometric crossovers. For Syntactic trees, the family of Homologous crossovers (one-point, uniform crossover) are geometric crossovers. Recombinations for other more complicated representations such as variable length sequences, graphs, permutations with repetitions, circular permutations, sets, multisets partitions are geometric crossovers.

3. GAUSSIAN RANDOM FIELDS ON METRIC SPACES

Most of the known results on GRF are about Euclidean spaces. In this section, we will consider GRFs on general metric spaces. The reader already familiar with both subjects may move on to section 4 that presents new results for GRFs on Hamming spaces. By no means this section treats exhaustively the generalization of GRFs to metric spaces. Very important uses of GRFs are left out such as, for example, statistical spatial inference, also known as Kriging [5], and GRFs model fitting to spatial data using maximum likelihood. We will consider these aspects of GRFs in future work. The message of this section is that, ultimately GRFs are special cases of Multivariate Gaussian distributions, and therefore all methods and theory developed for them can be readily specialized for GRFs with any underlying metric. The most peculiar characteristic of the GRFs is the correlation function, which requires special attention.

3.1 Basic definitions

A random field (see [13] for an introduction) is a family $\{X_t, t \in T\}$ of random variables. The set T is called the index set of the field. No restriction is placed on the nature of T. However, important special cases are when T is the set of natural numbers (discrete random process), when T is the set of positive real numbers (continuous random process) and when T is the space of n-dimensional real vectors (continuous random field). In this paper, we will embrace a more general approach and we will consider T as a metric space endowed with a metric d. To the authors's knowledge, in literature there is no comprehensive reference on this topic. However, Adler [2] surveys some aspects of GRFs on metric spaces. The type of index set, although at first seemingly arbitrary assigned, is important because it casts an interpretation on the object the random field models and it shapes the choice of what particular families of random field to study and what properties of the random field are relevant under the specific interpretation.

A random field is completely specified by the joint probability distribution of the n random variables $X_{t_1}, \ldots, X_{t_n}$ where $T = \{t_1, \ldots, t_n\}$. Its mean value function, denoted by $m(t)$, is defined for all $t \in T$ by $m(t) = E(X_t)$, its covariance function, denoted by $C(s, t)$, is defined for all $s, t \in T$ by $C(s, t) = Cov(X_s, X_t)$, and its correlation function, denoted by $\rho(s, t)$, is defined for all $s, t \in T$ by $\rho(s, t) = Corr(X_s, X_t)$.

A Gaussian random field (GRF) is a random field in which the joint probability distribution of the random variables $\{X_t\}$ is a multivariate Gaussian distribution. A GRF is completely specified by its mean value function and its covariance function. A standardized GRF is a GRF with mean function $m(t) = 0$ and covariance function $C(t, t) = 1$ for all $t \in T$. In this case, the correlation function coincides with the covariance function and it fully characterizes the random field.

A standardized GRF is a *isotropic GRF* if its correlation function depends on the distance alone, i.e. $\rho(s, t) = \rho(d(s, t))$. The corresponding correlation function is called isotropic correlation function. Therefore, in a isotropic GRF the correlation between any two random variables of the field depends exclusively on the distance between their locations in the field.

Above, we have defined a GRF as a family of random variables. A GRF can be equivalently interpreted as a probability distribution over the space of functions whose domain is the index set T and codomain is the domain of the random variables X_t. This view is particularly suited when doing inference on GRF [19].

3.2 Correlation functions

A function $f : \mathbf{R} \to \mathbf{R}$ must be *positive semi-definite* and have $f(0) = 1$ to be a valid isotropic correlation function. These conditions together guarantee that for non-negative input the function f returns values in the range $[-1, 1]$. Positive semi-definiteness of the correlation function is required to guarantee that every random variable of the field has non-negative variance.

Let k be a positive integer, and let $t_i \in T$ and $c_i \in \mathbf{R}$ for $i = 1, \ldots, k$. Then a function $f : \mathbf{R} \to \mathbf{R}$ is said to be positive semi-definite on the metric space (T, d) if

$$\sum_{i=1}^{k} \sum_{j=1}^{k} c_i c_j f(d(t_i, t_j)) \geq 0$$

for any choice of k, $\{t_1, \ldots, t_k\}$ and $\{c_1, \ldots, c_k\}$.

Most of the existing theory on correlation functions deals with the special case of the Euclidean distance in $\mathbf{R}^n$, for which various families of valid correlation functions are known. When using this distance the verification of positive semi-definiteness is relatively simple in principle [1]. However, there is no guarantees that the correlation functions which are valid for the Euclidean distance remain positive semi-definite, hence valid, for other distances (see [6] for a counterexample). Verifying positive semi-definiteness of functions on families of metric spaces other than the Euclidean or even for general metric spaces is non-trivial and it is a research field with a number of open questions (see [21]). In section 4, we will introduce a method to derive valid correlation functions for GRFs based on the Hamming distance.

Perhaps the most well-known family of isotropic correlation functions is the *exponential* [19] in which the correlation between random variables in the field is always positive and it decreases with their distance r:

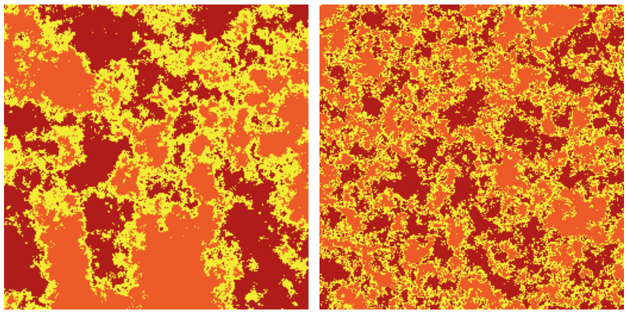
$$\rho_E(r) = \exp(-3 \cdot (r/l)^v)$$

with the positive real parameter l defining the *correlation length*, which is the distance at which there is practically no correlation and with $0 < v \leq 2$, which specifies the shape of the curve. The scaling factor '-3' is chosen so that the correlation at the correlation length is $e^{-3} \approx 0.05$.

Two important special cases of correlation functions are the *white noise*, which has correlation 1 for distance 0 and correlation 0 at any other distance, and the *everywhere constant*, which has correlation 1 for any distance. Realizations of the GRF with the white noise correlation function have no spatial dependency. Realizations of the GRF with the everywhere constant correlation function are flat and their height is determined by a single random number.

3.3 Sampling a GRF

Generating realizations of a GRF on any metric space is analogous to generating realizations for the specific case of the Euclidean space. Since a GRF is a special type of multivariate Gaussian distribution described by a correlation function, it is, therefore, sufficient to derive the covariance matrix of the multivariate Gaussian distribution from the correlation function of the random field applied to the distances between the points in the field using the specific metric, instead of using the Euclidean metric. It is then possible to use a standard method to sample the multivariate Gaussian distribution with this covariance matrix to generate a realization of the random field.



(A) smooth (B) less smooth

Figure 1: Two realizations of an isotropic GRF with exponential correlation function

The most common method to sample a multivariate Gaussian distribution is the one based on the Cholesky decomposition of the covariance matrix (see, for example [3]) which consists of obtaining a sample of an uncorrelated multivariate Gaussian distribution and then transforming it to enforce the correlation structure dictated by the covariance matrix.

Figure 1 shows two realizations of isotropic GRF on a two-dimensional grid with exponential correlation function with different correlation lengths. The colors correspond to thresholds based on three classes. The points whose values belong to the classes of the smallest, medium, largest third of all values are in orange, yellow and red, respectively. Intuition may suggest that realizations of isotropic random fields should be spatially uniform. This is incorrect. Figure 1 (A) shows clearly that, because of the stronger spatial correlation between closer points, closer points are more likely to belong to the same class of values so giving rise to a landscape with "mountain ranges" (red) and "large valleys" (orange). Figure 1 (B) shows the effect of a reduced spatial correlation: few large "mountain ranges" are replaced by many "mounts" which are separated from each other by smaller and numerous "valleys".

4. CORRELATION FUNCTIONS FOR THE HAMMING SPACE

In this section, we derive a general method to transform all valid correlation functions on the Euclidean space to valid correlation functions on the Hamming space.

A valid covariance matrix must be positive semi-definite. We say that a correlation function for a given metric space is valid if produces always a valid covariance matrix associated with any choice of points in the metric space.

There is an interesting link between Euclidean distance, positive semi-definite correlation function and positive semi-definite covariance matrix that shed light on how to obtain valid correlation functions for metric spaces other than Euclidean. We illustrate this in the following.

Let $x_1, \ldots, x_n \in \mathbf{R}^k$. The matrix D defined by $D_{ij} = ed(x_i, x_j)$ where ed denotes the Euclidean distance between two vectors is called a *Euclidean distance matrix.* Since the Euclidean distance is a metric, the matrix D satisfies some obvious properties such as $D_{ij} = D_{ji}$, $D_{ii} = 0$, $D_{ij} \geq 0$, and (from the triangle inequality) $D_{ik} \leq D_{ij} + D_{jk}$. However, these properties do not characterize completely a Euclidean distance matrix. When then is a matrix D a Euclidean distance matrix (for some points in $\mathbf{R}^k$, for some k)? A famous result [21] answers this question: D is a Euclidean distance matrix if and only if $D'_{ii} = 0$ and $x^T D' x \leq 0$ for all vectors x with $\mathbf{1}^T x = 0$ where $D'_{ij} = D^2_{ij}$.

From the definition of positive semi-definite correlation function for Euclidean GRF, by applying element-wise a positive semi-definite correlation function to the Euclidean distance matrix, we obtain the corresponding covariance matrix, which is a positive semi-definite matrix. Importantly, when the distance matrix is not Euclidean or the correlation function is not positive semi-definite on the Euclidean space, the resulting covariance matrix is not necessarily positive semi-definite, hence, not guaranteed to be valid.

DEFINITION 4. *(Isometry) Two metric spaces $M = (S, d)$ and $M' = (S', d')$ are isometric if there exists a bijective function $g : S \rightarrow S'$ which is distance-preserving: so $\forall x, y \in S : d(x, y) = d'(g(x), g(y))$. The mapping g is called isometry.*

From the definitions of isometry, of Euclidean distance matrix and of positive semi-defined function, it is clear that if a metric space M is isometric to any subspace of the Euclidean space, the distances between any set of points in M, will give rise to an Euclidean distance matrix, and consequently any valid correlation function for the Euclidean space is a valid correlation function for M. This was first noted by Schoenberg [21].

This is an interesting result because it extends the scope of the Euclidean correlation functions. In particular, any valid correlation function on the Euclidean space is a valid correlation function for any discrete space which is a subspace of the Euclidean space, like for example a two-dimensional grid.

However, this result does not extend to the Hamming space because it is known that, in the general case, the Hamming space cannot be isometrically embedded in the Euclidean space [8]. So, it is not guaranteed that if a correlation function is valid for a GRF on a Euclidean space it will also be valid for the Hamming space. To find valid correlation functions for the Hamming space we take an alternative route.

PROPOSITION 1. *The Euclidean distance between binary strings equals the square root of their Hamming distance [4].*

This fact allows us to state the next theorem.

THEOREM 1. *If $f(x)$ is a valid correlation function on the Euclidean space, $g(x) = f(\sqrt{x})$ is a valid correlation function on the Hamming space.*

PROOF. Let S be a set of binary strings of the same size and $D = \{ed(s_i, s_j)\}$ their Euclidean distance matrix. Therefore, the covariance matrix associated with S under Euclidean distance via the correlation function f is $C = \{f(ed(s_i, s_j))\}$. C is positive semi-definite, hence valid, by construction. Now, let $D_H = \{hd(s_i, s_j)\}$ be the Hamming distance matrix of S. The covariance matrix associated with S under Hamming distance via the correlation function g is $C_H = \{g(hd(s_i, s_j))\}$. C_H is always positive semi-definite because, since $D_H = \{ed(s_i, s_j)^2\}$, we have that $C_H = \{f(\sqrt{hd(s_i, s_j)})\} = \{f(\sqrt{ed(s_i, s_j)^2})\} = C$. Therefore, g must be a valid correlation function for the Hamming space because it always gives rise to a valid covariance matrix. $\square$

Notice that the converse of this result may not be true, since there may be correlation functions for the Hamming space that do not correspond to any valid correlation functions for the Euclidean space. This is because, whereas any adequately transformed Hamming space can be isometrically embedded in the Euclidean space, the vice versa is not true.

This theorem can be easily generalized to any metric space d obtained by an invertible distance transform t of the Euclidean distance or of a space isometric to any subspace of the Euclidean space, i.e. $d(a, b) = t(ed(a, b))$. So, if $f(x)$ is a valid correlation function on the Euclidean space, $g(x) = f(t^{-1}(x))$ is a valid correlation function of under the metric d.

Analogously to the result for the Euclidean space, a valid correlation function for the Hamming space is also a valid correlation function for any space that can be isometrically embedded in the Hamming space. A number of interesting discrete spaces have this property [4], such as, for example, the metric between sets equalling the size of their symmetric difference, the Hamming distance for strings based on non-binary alphabets, the Manhattan distance on integer vectors and the adjacent swap distance between permutations.

COROLLARY 1. *The family of the exponential correlation functions on the Euclidean space gives rise to the following family of exponential correlation functions on the Hamming space.*

$$\rho_E(r) = \exp(-3 \cdot (r/l)^v)$$

with the positive real parameter l defining the correlation length, *which is the Hamming distance at which there is practically no correlation with $0 < v \leq 1$.*

PROOF. Let us consider the family of correlation functions for the Euclidean space $f(r) = \exp(-3 \cdot (r/l')^{v'})$ with $v' \in [0, 2]$. Applying proposition 1 we obtain the following family of correlation functions for Hamming space: $g(r) = \exp(-3 \cdot (r^{v'/2}/l'^{v'}))$. Now, if we substitute $v = v'/2$ with $v \in [0, 1]$ and $l = l'^2$, we obtain $g(r) = \exp(-3 \cdot (r^v/l^v))$. The parameter l can be interpreted as correlation length for the Hamming space because for Hamming distance $r = l$ the correlation is practically zero, $g(l) = e^{-3} \approx 0.05$. $\square$

Notice that the family of exponential functions valid for the Hamming space is a subset of the family of exponential functions valid for the Euclidean space.

5. SMOOTHNESS OF FITNESS LANDSCAPES

In this section, we use the GRF model to define a notion of smooth fitness landscape which is *independent* from the specific underlying space and solution representation. In order to avoid subtle errors related to the interpretation of smoothness, it is important to differentiate between smoothness as a parameter of the statistical model, hence characterizing a probability distribution over the space of fitness landscapes, and smoothness as empirical measure obtained from a single fitness landscape. In this section, we refer to the former notion of smoothness. In section 7, we will consider the relation between the two notions and show how this distinction clarifies the origin of known problems with statistical empirical measures of hardness of fitness landscapes.

5.1 Smoothness

An isotropic GRF is completely characterized by its correlation function. If we consider two correlation functions ρ_1 and ρ_2, if ρ_1 is always above ρ_2, (i.e., $\forall r \geq 0 : \rho_1(r) \geq \rho_2(r)$) we can certainly say that the GRF G_1 characterized by ρ_1 is smoother (more precisely, it models smoother fitness landscapes) than the GRF G_2 characterized by ρ_2 because, at each distance, the correlation between fitness at that distance dictated by G_1 is stronger than the one dictated by G_2. However, when neither of the correlation functions is dominant, none of them corresponds to a smoother GRF.

In special circumstances, it is possible to characterize exactly the smoothness of a GRF by a single number [1]. The best candidate to this end is the correlation length of the correlation function,[2] as we explain in the following.

The most common families of correlation functions are continuous positive monotonic decreasing functions, like the exponential family. However, other types of correlation functions exist, which are non-monotonic and can take negative values (for an overview on correlation functions see [1]).

Different families of correlation functions are characterized by a different set of parameters. However, the correlation length parameter can be defined for all families of correlation functions which reach correlation zero or near-zero for some distance (which is the case for almost all correlation functions) and it can be always adequately re-scaled to have the same interpretation for any function, which is, to be the smallest distance at which the correlation function is zero (or approximatively zero).

For decreasing correlation functions, once we fix the family of correlation functions and all its parameters other than the correlation length parameter, the correlation length can be used to characterize exactly the level of smoothness of the field. This is because a smaller correlation length corresponds to a correlation curve that reduces more rapidly for increasing distance and reaches zero earlier (see figure 2, for the case of the exponential family of correlation functions with parameter $v = 1$). In formula, let $\rho(r, l)$ be a decreasing correlation function with distance r and correlation length l. Then, $l_1 \leq l_2 \Rightarrow \forall r : \rho(r, l_1) \leq \rho(r, l_2)$ thereby for smaller correlation length the field becomes less smooth.

[1] Although it may be easier in practice to formalize the notion of smoothness using a single number to characterize the smoothness level of a GRF, in general, this cannot be done without introducing an error, because the comparison between two numbers produces always a well-defined outcome (smoother, equally smooth or less smooth) and the situation of incomparable smoothness cannot be modeled.

[2] The parameter of smoothness suggested in this paper is closely related to the measure of smoothness proposed by Weinberger [24], which, however, is an empirical measure of smoothness and not intended as a smoothness parameter of a GRF.

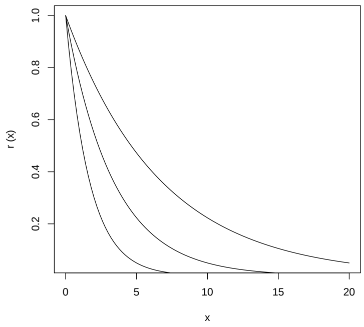


Figure 2: Exponential correlation function with correlation lengths 5, 10 and 20.

5.2 Distance re-scaling

To give a proper interpretation to the correlation length parameter in terms of its effect on the realizations of a GRF, it is important to realize that the correlation length is not invariant with respect to the re-scaling of the underlying metric space, whereas the overall smoothness of the realization it is invariant with respect to this transformation. We will illustrate this with an example.

Let us consider a realization of a GRF with exponential correlation function $\rho(r, l)$ with $v = 1$ on a two-dimensional square with side length 10 endowed with Euclidean distance generated with correlation length $l = 5$. If we re-scale of a factor 3 this realization, we obtain a new realization on a square with side length 30, which overall looks identical, except for the size, to the original realization. In particular, both realizations have the same level of smoothness. Therefore, the smoothness of the realization is invariant with distance re-scaling.

However, the re-scaling of the size of the square of a factor 3 has affected the distances between points in the square of the same factor, so that the points that before re-scaling were at a distance d after the re-scaling are at a distance $3 \cdot d$. This, in turn, has implicitly affected the correlation length of the re-scaled realization, since the correlation length is defined in terms of underlying distance, so that when the distance is re-scaled also the correlation length is re-scaled. So, the effective correlation length of the re-scaled realization is $3 \cdot l = 15$. This means that two realizations with the same level of smoothness are associated to different correlation lengths. So, the correlation length is not invariant with distance re-scaling.

This example shows that the interpretation of the correlation length in terms of overall smoothness of the realization has to keep into account the size of the underlying metric space. This can be done, for example, by relating correlation length with the diameter of the space (d_{max}) or alternatively with the average distance of the space. Alternatively, a notion of normalized correlation length $\hat{l} = l/d_{max}$ can be introduced, which is distance re-scaling invariant.

5.3 Separation between algorithm, search space and correlation structure

The same correlation function can be used for different underlying metric spaces provided it always produces positive semi-definite covariance matrices when used with these spaces. For example, the exponential correlation function with parameter $v = 1$ is a valid correlation function for Euclidean spaces of any dimension. The same function is also a valid correlation function for the Hamming space of any dimension. Validity consideration apart, the polymorphism of correlation functions with respect to their applicability to underlying spaces with very heterogeneous characteristics (for example, different dimensions and different underlying representation) is made possible by the interface between the underlying metric space and the correlation function. This interface is completely independent from the specific characteristics of the space and it is exclusively based on distance values, which are common to all metric spaces. This property is interesting because it leads to a natural modularity, or separation, of the correlation function C from the underlying metric space M. This connects well with the abstraction provided by the geometric framework, as follows.

The geometric framework allows us to define a formal search algorithm A which can be seen as a function of a generic metric space M. When the metric space is specified by formal composition of the formal algorithm A and the metric space M, we obtain A_M, which is a concrete search algorithm for the specific space.

We can now express, in a functional form $P(A, M, C)$, the average-case performance of a formal algorithm A specified to the metric space M on the distribution probability of the input fitness landscapes specified by the GRF with correlation function C specified to the metric space M. Notice that the performance P is a random variable with two sources of randomness since it is a function of the algorithm A, which is a randomized algorithm, and a function of the GRF specified by C, which is a random variable defined over the space of fitness landscapes.

This functional form, which separates search algorithm, space searched, and correlation profile of the fitness landscape, allows us to selectively fix two input variables and change the third one and make rigorous comparisons on the performance. So, for example, we could fix the correlation function C to, say, exponential with parameters $v = 1$ and $l = 3$, the algorithm A to a formal genetic algorithm with uniform geometric crossover, and vary the underlying metric space from Euclidean to Hamming.

6. EXPERIMENTS

In this section, we test experimentally, for the case of the Hamming space, the consistency of our formalized model of smoothness with the performance of traditional search algorithms as to whether smoother fitness landscapes lead to better average-case performances. The search algorithms considered are Local Search (LS), Randomised Local Search (RLS), a simple Genetic Algorithm (GA) and, for reference, Random Search (RS) (see Table 1).

We sampled fitness landscapes from a GRF on Hamming space with n dimensions (bit string length) as follows. Firstly, a vector containing 2^n samples of standard normally distributed values was generated. Then, a deterministic transformation was applied to it to enforce the correlation structure dictated by the correlation function of the GRF. The transformation consists of 3 steps: 1) generating the covariance matrix from the correlation function and the distance matrix between all points; 2) performing the Cholesky decomposition of the covariance matrix (that has time complexity $O(k^3)$ with $k = 2^n$); and 3) multiplying the vector of independent samples by decomposed matrix. This process grows exponentially slow with n. For this reason, we restrict the experiments to bit-strings of length 12 [3]. As correlation function family, we used the exponential with shape parameter $v = 1$.

Table 1: The algorithms used in the experiment.

Type	Algorithm
Random	**Random Search:** At time step t the solution x_t is selected uniformly at random from $\{0, 1\}^n$
Local	**Local Search:** The first solution is sampled uniformly at random. Then at time step t, $x_{t+1} = \arg\max_x \{f(x) \mid d(x, x_t) \leq 1\}$. If $x_{t+1} = x_t$ the algorithm restarts. **Randomised Local Search:** The first solution is sampled uniformly at random. Then at time step t, the algorithm chooses uniformly at random $x' \in \{x \mid d(x, x_t) = 1\}$. $x_{t+1} = \arg\max_{x \in \{x_t, x'\}} f(x)$. The algorithm restarts if there is no improvement (in fitness) within $2n$ steps.
Population	**GA, uniform crossover:** We used population size of 6, tournament of size 2, crossover applied with probability 1 and mutation with probability $1/n$. The algorithm was restarted every 20 generations.

The Hamming distance between two strings, in this case, varies from 0 to 12. In order to study the effect of varying levels of smoothness of the landscape on the performance, we generated, for each correlation length $l \in \{1, \cdots, 12\}$, 50 different landscapes for a total of $12 * 50 = 600$ landscapes. For each landscape, we run each algorithm until the optimum was found (note that as the fitness values of the landscapes are sampled from a continuous distribution, each of the landscapes has only one global optimum). This was repeated 100 times. Figures 3, 4 and 5 show the average number of fitness evaluations (including repetitions) it took GA, RLS and LS to sample the optimum for the first time. Table 2 gives the results of a pairwise t-test which tests the significance of the performance gain for adjacent correlation lengths.

We can note a clear trend for all the search algorithms: *the performance becomes better as the value of the correlation length increases.* We can also see that for low correlation lengths (1 and 2), RLS and GA perform worse than random search. This is mainly the effect of resampling. For larger correlation lengths, LS outperforms the other two search algorithms.

Interestingly, the curve of the average performance becomes flatter for $l > 7$, which is, for values of l larger than the average distance between solutions. Also, Table 2 shows that the difference in performance for successive values of l after this threshold becomes less statistically significant. We do not yet understand clearly why the performance almost

[3]However, it may become feasible to sample larger fitness landscapes employing statistical inference on GRFs and sparse approximation of GRFs. We briefly discuss this possibility in section 8

Table 2: **Pairwise significance test for the experiment. $t(i,j)$ denotes the significance test for the results obtained for correlations i and j. The check mark denotes that $t(i,j) < 0.001$. Otherwise, the actual value is given.**

	$t(1,2)$	$t(2,3)$	$t(3,4)$	$t(4,5)$	$t(5,6)$	$t(6,7)$
RLS	✓	✓	✓	✓	✓	0.04
LS	✓	✓	✓	✓	0.02	0.08
GA	✓	✓	0.01	✓	✓	0.31

	$t(7,8)$	$t(8,9)$	$t(9,10)$	$t(10,11)$	$t(11,12)$
RLS	✓	0.68	0.15	0.68	0.16
LS	0.01	0.23	0.05	0.86	0.1
GA	0.01	0.87	0.41	0.79	0.24

stops getting better when the correlation length reaches the average distance between solutions. We plan to investigate this in future research.

6.1 Decomposition of performance variance

Note that the standard deviation of the performance for each correlation length (the error bars in Figures 3, 4 and 5) is the result of two distinct sources of randomness: the randomization of the search algorithm and the probabilistic sampling of the input fitness landscape from the GRF. Interestingly, it is possible to determine the relative contributions of these two sources of randomness on the variability of the performance, as follows.

Let us fix an algorithm A and let L be a random variable over the space of fitness landscapes, which is a GRF with a fixed correlation function (in our specific experimental setting, we simply need to fix the correlation length l). The performance of A on L, $P_A(L)$, is a random variable P_A function of the random variable L. Notice that P_A is not simply a deterministic function of r.v., but it is a r.v. piloted by the outcome of another r.v. because of the randomization in the search algorithm A. There are two well-known statistical laws of general applicability that can be used to characterize the expectation and variance of P_A by decomposing them conditionally on the values of the piloting variable L. For the law of conditional expectations, the expected value of P_A is:

$$E_{P_A}[P_A] = E_L[E_{P_A}[P_A|L]].$$

For the law of total variance, the variance of P_A is:

$$Var_{P_A}(P_A) = E_L[Var_{P_A}(P_A|L)] + Var_L(E_{P_A}[P_A|L]).$$

Interestingly, the summands in the formula of the variance can be interpreted as follows. The first summand is the part of variance not caused by the variability of L (unexplained variance) and the second summand is the part of variance caused by the variability of L (explained variance). The only other source of variability of the performance, in our case, is the randomization of A, so the first summand is the part of variance caused by the variability of A.

Table 3 gives the fraction of the total variance of P_A which is caused by the variability of L. That is, for algorithm A and GRF with correlation length l, Table 3 reports the value $Var_L(E_{P_A}[P_A|L])/Var_{P_A}(P_A)$ where $E_{P_A}[P_A|L]$ is the average performance of the 100 runs of A on the same fitness landscape, $Var_L(E_{P_A}[P_A|L])$ is the variance of the averages over 50 landscapes sampled at random from GRF with correlation length l and $Var_{P_A}(P_A)$ is the total variance of the performance.

For example, the value 0.01 (Random Search, $l = 1$) should be interpreted as follows: 99% of the variability in the performance of random search in our experiments can be attributed to A, the randomization of the search algorithm, and only 1% to the actual landscape sampled from L. This is not surprising for random search, which is invariant to any property of the fitness landscape other than the number of optima (the theoretical value for random search is 0).

For the other search algorithms, the variance of the performance that is caused by the variability of L does not exceed 0.25. This means that most of the variability in the performance is caused by the search algorithms. Comparing the three, the variance due to the algorithm is more dominant in the case of GA. This seems reasonable as there are more sources of randomness in GA than in the other search algorithms. The difference between RLS and LS can be explained in a similar way. Finally, note that in these experiments, the more randomized the algorithm the less efficient it is.

Table 3: The variance in performance of Random Search (RS), Randomised Local Search (RLS), Local Search (LS) and GA which can be explained by the distribution induced by the GRF with correlation length $l = i$.

	$l=1$	$l=2$	$l=3$	$l=4$	$l=5$	$l=6$
RS	0.01	0.01	0.01	0.01	0.01	0.01
RLS	0.08	0.10	0.12	0.13	0.12	0.12
LS	0.16	0.18	0.16	0.17	0.16	0.24
GA	0.01	0.04	0.05	0.07	0.07	0.08

	$l=7$	$l=8$	$l=9$	$l=10$	$l=11$	$l=12$
RS	0.01	0.01	0.01	0.01	0.01	0.01
RLS	0.09	0.12	0.17	0.12	0.13	0.16
LS	0.17	0.14	0.22	0.17	0.16	0.22
GA	0.09	0.08	0.11	0.10	0.10	0.11

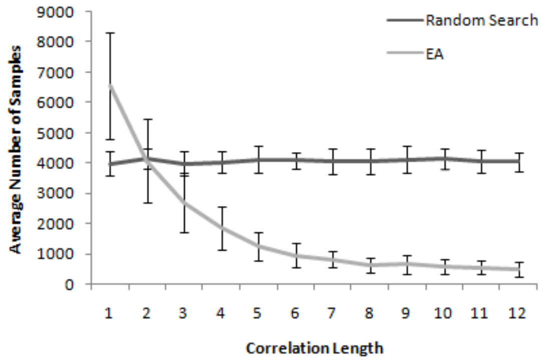


Figure 3: The average number of fitness evaluation as a function of the correlation length. The error bars represent the standard deviation of the 50 different instances.

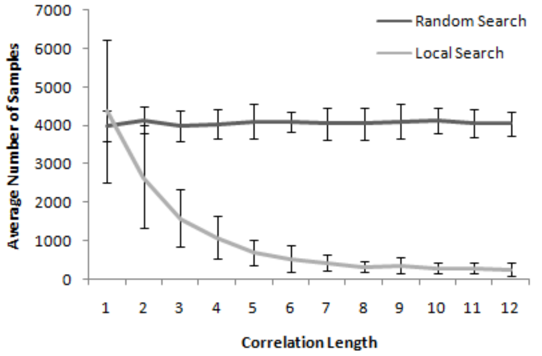


Figure 4: The average number of fitness evaluation as a function of the correlation length. The error bars represent the standard deviation of the 50 different instances.

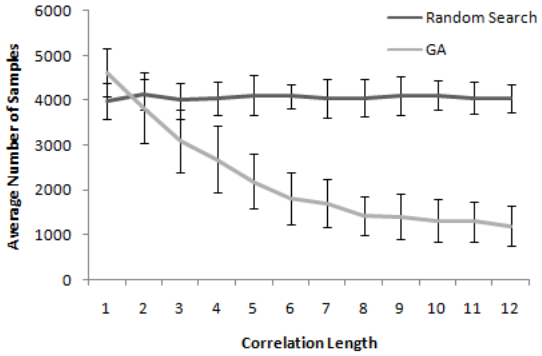


Figure 5: The average number of fitness evaluation as a function of the correlation length. The error bars represent the standard deviation of the 50 different instances.

7. EMPIRICAL SMOOTHNESS AND LANDSCAPE HARDNESS

In section 6, we have seen that there is consistency between the value of the correlation length parameter of the GRF and the average-case performance of traditional search algorithms: larger correlation length corresponds to better performance. It is well-known, however, that there are counter-examples on using empirical measures of smoothness as measures of hardness of a fitness landscape. In this section, we will point out the origin of these counter-examples.

From an instance of fitness landscape we can derive the empirical correlation function, also known as correlogram [5]. The correlogram can be used to estimate (maximum likelihood estimation) the correlation function of the most likely isotropic GRF model of which the landscape at hand can be considered a realization. The empirical correlation length obtained from the correlogram can be then used to estimate the correlation length parameter of the GRF. Now the question is: can we reliably use the estimated correlation function and correlation length as measure of hardness of the fitness landscape for traditional search algorithms? This is a legitimate question which *seems* consistent with the claim that the correlation length parameter of the GRF is the main determinant of average-case performance. However, if we do so, we make three assumptions, which, if not satisfied, will be the cause of counter-examples to the use of empirical correlation length as a measure of hardness:

1. we assume that there is consistency between the correlation length parameter and performance. Although, our experiments corroborated this claim, there is no proof yet.

2. we assume good-fit of the fitness landscape at hand to the isotropic GRF estimated from it. However, this may not be always the case.

3. we assume that the average-case performance of a traditional search algorithm on the GRF can be used as an estimation of the performance of the search algorithm on a "typical" fitness landscape drawn from that GRF.

We believe that assumptions 1 and 3 are essentially correct. However, they need proofs or, at least, further experimental study. In the following, we will show what can go wrong with assumption 2.

Firstly, we need a goodness-of-fit test that indicates the probability that a fitness landscape can be thought of as being a realization of a given GRF. The outcome of the goodness-of-fit test can be fit, mis-fit (probability < 0.05), or over-fit (probability > 0.95). Only fitness landscapes which fit the GRF model can be considered as valid realizations of the model. Mis-fitting landscapes are not consistent with the GRF model. Over-fitting landscapes are too consistent and hence lack of necessary variability to be considered typical outcomes of random sampling of the GRF.

In geo-statistics, there are a number of statistical tests that checks whether various "geographical" features of a given landscape are compatible with the features required by the GRF [5]. However, we would like to have a more general test which does not consider individually various aspects of a landscape, but rather tells us whether the entire landscape fit the GRF model, and to what degree. To do so, we propose the following test.

The test is based on reversing the construction of a correlated realization L of a GRF G by filtering from L the correlation structure dictated by G and then test whether the filtered realization L' can be seen as a sample of an uncorrelated GRF, test which can be done in a simple way. This test relies on the fact that, since the transformation T_G that enforces the correlation structure dictated by G on an uncorrelated realization L' to construct a correlated realization L is deterministic and bijective, the goodness-of-fit of L with respect to G is the same as the goodness-of-fit of L' with respect to an uncorrelated GRF. Then, to test whether L' is a realization of an uncorrelated GRF is sufficient to test whether the set of values of L', irrespectively of their locations, can be thought as a set of values obtained by sampling a standardized Gaussian distribution. To do so, we can use any of the many statistical tests available for this task. We used the Pearson chi-square normality test. The reason we are allowed to neglect altogether the locations of the values of L' in the test is that any spatial arrangement of the values, including those which present regularities, has the same probability (density) of being generated as the outcome of the sampling of an uncorrelated GRF as any other configuration.

In the following, we describe an experiment that shows what may happen when assumption 2 above is not met, which is, when the empirical correlation length is used as a predictor of performance of a mis-fitting fitness landscape.

Firstly, we sample a fitness landscape L from a smooth GRF model G. In our experiment, we use the exponential correlation function with $v = 1$ and correlation length $l = 12$ on the Hamming space on 12 dimensions (associated with binary strings of 12 bits).

If we measure the goodness-of-fit of the landscape L with respect to the model G, we should obtain a good-fit, since the fitness landscape L is a typical instance of the GRF G by construction. In our experiment, we have obtained a p-value = 0.5553 ($P = 50.9258$), which denotes a good-fit as expected.

Next, we run a traditional search algorithm on the fitness landscape L and measure its performance in terms of number of fitness evaluations taken to find the optimum. Since with our settings the fitness landscape is very smooth ($l = 12$), we expect the search algorithm to perform much better than random search on this fitness landscape. This expectation is based on the assumption that the average-case performance of a given GRF estimates well the performance of a traditional search algorithm on a typical instance of fitness landscape of that GRF (we assume assumption 3 above being correct). In our experiment, we run three search algorithms – Randomised Local Search, Local Search and GA – on the fitness landscape L. Their performance (average on 100 runs) are reported in figure 6 on the left side (regular landscape). As a reference, random search would require in average 4096 fitness evaluations to find the optimum.

At this point, we could compute the empirical correlogram of the fitness landscape L (and its empirical correlation length) and derive the maximum likelihood isotropic GRF model $\hat{G}$ that best-fits the fitness landscape L. Since L is a typical instance of G, the GRF $\hat{G}$ will be identical or a very good approximation of the original GRF G used to generate the fitness landscape L.

Let us now consider a new fitness landscape L' obtained by swapping the locations of the maximum and minimum in L.

The empirical correlogram of the new fitness landscape L' is the same or imperceptibly different from the one of the original fitness landscape L because, by construction, the empirical correlogram is not affected from changes to only few points in the fitness landscapes. Therefore, the maximum likelihood isotropic GRF model $\hat{G'}$ that best-fits the new fitness landscape L' will be the same as $\hat{G}$ since it is derived using the the same empirical correlation function. In turn, $\hat{G'}$ will be identical or a very good approximation of the original GRF G.

At this point, we might be tempted to conclude that, since both fitness landscapes L and L' are associated with the same maximum likelihood GRF model G, the performance of traditional search algorithms on both fitness landscapes should be similar or equal, and in particular, the average-case performance of traditional search algorithms on the GRF model G should be a good estimate of their performance on both landscapes L and L' (assuming assumption 3 above being correct).

However, while drawing this conclusion, we are implicitly assuming that the fitness landscape L' fits well the GRF G (we are implicitly assuming assumption 2). We are not considering the possibility that, even if G is the isotropic GRF which fits best L', the fit could still be an arbitrarily bad mis-fit because, in fact, L' may be not compatible with the statistical regularities characterizing *any* isotropic GRF model. To find out whether this is the case, we can measure the goodness-of-fit of the fitness landscape L' to the GRF G. We have tested the goodness-of-fit of the new fitness landscape with respect to the GRF and obtained a p-value $< 2.210^{-16}$ ($P = 201.7988$). This is an extremely small p-value which can be interpreted as the probability of the fitness landscape L' of being sampled from the GRF G. So, we have an extremely clear mis-fitting.

The intuitive reason behind the clear mis-fit is that, in a strongly correlated GRF, the maximum, with very high probability, is close to points with near-maximal fitness or very high fitness, and, analogously, the minimum is close to points with near-minimal fitness. By swapping the location of the extreme points, whereas the fitness of all other points in the landscape are consistent with a strongly correlated GRF, the fitness of the extreme points are strongly inconsistent with it. Importantly, there is no choice of correlation function that can make an isotropic GRF to be consistent with such an arrangement of fitness. So, even the best possible fit, it is, in fact, a mis-fit.

The fitness landscape L' by construction should be highly deceptive for traditional search algorithms. To test this, we run on L' the same three search algorithms we run on L. Their performance (average on 100 runs) are reported in figure 6 on the right side (deceptive landscape). The performance obtained by these algorithms are worse than random search.

Since the landscapes L and L' have the same empirical correlogram but very different hardness for traditional search algorithms, this experiment tells us that the empirical correlogram cannot be used as a reliable performance predictor. Importantly, the experiment elicits the reason behind this counter-example, which is, that the GRF model from which the landscape L' is most likely to be drawn from cannot be used reliably as a basis to estimate the performance of traditional search algorithms on L' when L' does not fit well *any* GRF model. There is also another type of counter-examples to using the empirical correlogram to estimate the hardness of a fitness landscape for traditional search algorithms which are caused by the over-fitting of a fitness landscape to the GRF model from which it is most likely to be drawn from.

Importantly, the counter-examples to using the empirical correlogram as a measure of hardness of a fitness landscape are not counter-examples to the consistency between correlation length as determinant of average-case performance of a GRF (assumption 1).

Perhaps, a valuable lesson that we can learn from this section is that naked empirical statistical measures on a fitness landscape used to predict performance are doomed to fail (to have counter-examples) if they are not considered within the context of an underlying statistical model of fitness landscapes which defines their scope of applicability.

It might be argued that the isotropic GRF model is too

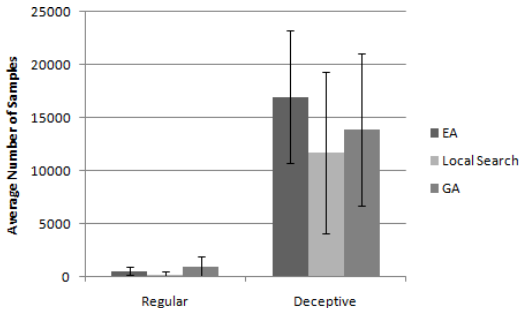


Figure 6: The performance of RLS, Local Search and GA on a *regular* GRF with $l = 12$ and on a *deceptive* landscape which was obtained by swapping the maximum and minimum fitness values of the original landscape.

restrictive to encompass all those fitness landscapes which could be reasonably understood as "smooth". There are two points to make on this issue. Firstly, the isotropic GRF model is actually more expressive than what intuition may suggest at first. Only further research will be able to reveal the extent to which isotropic GRFs are adequate to model "smooth" fitness landscapes arising from real optimisation problems. Secondly, if the isotropic GRF model turns out to be too restrictive, more general random field models will be considered. This, however, does not make the idea of considering hardness measures in the context of an underlying statistical model any less valid.

8. CONCLUSIONS AND PLAN OF FUTURE WORK

We have put forward that a general theory of smooth fitness landscapes is the key to bridge the gap between theory and practice in EC because it would allow the designer of the search algorithm to use the knowledge of the problem at hand to do an informed choice of design elements, such as solution representation and relative search operators, to induce a smooth fitness landscape which, by the theory, would guarantee good performance. Importantly, this theory should be very expressive and should allow us to specify solution representation and search operators as parameters and see how changing the values of these parameters impact the performance. A theory, at this level of abstraction, could be built by using as starting point the geometric framework which associates distances to search operators for any solution representation and that treats specific distances as parameters.

As a first step towards this theory, we have used a GRF model to define formally the notion of smooth fitness landscape in a very general setting, without making any reference to the specific underlying space or solution representation. On the way, we have also shown that GRF can be readily generalized to general metric spaces and, in particular, that they can be used for combinatorial spaces as good as they are being used in continuous spaces. For the specific case of the Hamming space, we have derived a general method to obtain valid correlation functions from Euclidean correlation functions. In the experiments, we have shown that the formalized notion of smoothness captures the heuristic property of its informal counterpart, so that traditional search algorithms perform better on smoother fitness landscapes. We have also discussed the relation between measure of hardness of the fitness landscape and our notion of smoothness.

Much work has to be done, if we are to make the proposed theory a reality.

An important practical requirement is to verify to what extent fitness landscapes associated to well-known real-world problems fit the GRF model, in other words, to verify whether they can be treated as if they were realizations of GRF.

A natural continuation of the work presented in this paper is to consider GRFs on more challenging solution representations such as permutations or even Genetic Programming trees, derive valid correlation functions for these spaces, and verifying empirically if our notion of smoothness is consistent with performances.

The notion of smoothness introduced in this paper is decoupled from the specific metric space, in particular it does not depend on its dimension. This would allow us to study experimentally how traditional search algorithms scale on fitness landscapes of increasing dimension with fixed smooth correlation structure. This is interesting because it would link smoothness of fitness landscapes with the average-case computational complexity associated with it.

Initially, we are planning to make more experiments for the Hamming space with more than 12 dimensions. Unfortunately, the computational complexity of generating smooth fitness landscapes grows exponentially with the number of dimensions. This makes it practically infeasible to generate larger fitness landscapes. However, we have preliminary experiments on using spatial statistical inference on GRF to generate fitness landscapes dynamically while being visited by the search algorithms. This makes it possible to run experiments on larger landscapes because only a fractions of all solutions are in fact generated.

In this paper, we have put forward the view of a statistical model of fitness landscape as an *abstraction*, in which the only relevant characteristic affecting the average-case performance of traditional search algorithms is the smoothness of a fitness landscape. All the other "geographic" characteristics of a fitness landscape, such as multi-modality, neutrality, barriers, and so forth and so on, are not explicitly taken into account in the model and are either subsumed in the notion of smoothness (for example, smoother landscapes have larger mountain ranges) or averaged out of existence and may manifest themselves only as deviation from the average in the form of noise. This is a methodological choice, not a rough simplification. It could be seen as a form of statistical *coarse graining* grounded on the assumption that the smoothness of a landscape is the most relevant characteristic for the expected performance of traditional search algorithms. Importantly, smoothness can be *controlled* at the time of design of the search algorithm, when the problem is known. This is what ultimately would make the theory-to-come relevant to practice. Clarified the methodological aspect of focusing on smoothness only, it would be interesting, however, to study the geographic characteristics of this class of fitness landscapes to have a more concrete grasp on them.

The above mentioned points are preliminary to the theory. The most challenging task will be to derive the theory itself, which is, a close formula that relates a formal algorithm A, a metric space M and a correlation function C to the average-case performance obtained with these parameters. As a first step towards this theory, we will derive such a result for the specific case of the (1+1)-Evolutionary algorithm on the Hamming space with the family of exponential correlation functions. We will then attempt to generalize this result to general metric spaces by replacing the Hamming distance with a generic metric.

There is an alternative and very attractive use of an explicit GRF model of fitness landscapes. Using spatial statistical inference, given the specific class of problems considered as a prior distribution, it would be possible to induce the most rational search algorithm to find the optimum with the minimum number of fitness evaluations. This is already a reality for the case of continuous optimization using an Euclidean GRF models [10] (response surface methods). We will extend this method to combinatorial spaces and attempt a generalization to generic metric spaces.

9. REFERENCES

[1] P. Abrahamsen. A review of gaussian random fields and correlation functions (1997). Technical report, Tech. Rep. 917, Norwegian Computing Center, 1997.

[2] R. J. Adler and J. Taylor. *Random Fields and Geometry*. Springer, 2007.

[3] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[4] G. Cormode. *Sequence Distance Embeddings*. PhD thesis, University of Warwick, 2003.

[5] N. A. C. Cressie. *Statistics for Spatial Data (revised edition)*. Wiley, 1993.

[6] F. C. Curriero. *The use of non-Euclidean distances in geostatistics*. PhD thesis, Kansas State University, 1996.

[7] M. Deza and M. Lauren. *Geometry of cuts and metrics*. Springer, 1991.

[8] P. Enflo. On the nonexistence of uniform homeomorphisms between lp-spaces. *Ark. Mat.*, 8:103–105, 1969.

[9] C. Igel and M. Toussaint. A no-free-lunch theorem for non-uniform distributions of target functions. *Journal of Mathematical Modelling and Algorithms*, 3:313–322, 2004.

[10] D. R. Jones. A taxonomy of global optimization methods based on response surfaces. *Journal of Global Optimization*, 21:4:345–383, 2001.

[11] T. Jones and S. Forrest. Fitness distance correlation as a measure of problem difficulty for genetic algorithms. In *Proceedings of the Sixth International Conference on Genetic Algorithms*, 1995.

[12] S. A. Kauffman and S. Levin. Towards a general theory of adaptive walks on rugged landscapes. *Journal of Theoretical Biology*, 128:11–45, 1987.

[13] D. Khoshnevisan. *Multiparameter Processes - An Introduction to Random Fields*. Springer, 2002.

[14] P. Larranaga and J. A. Lozano. *Estimation of Distribution Algorithms*. Springer, 2001.

[15] A. Moraglio. *Towards a geometric unification of evolutionary algorithms*. PhD thesis, University of Essex, 2007.

[16] A. Moraglio, C. D. Chio, and R. Poli. Geometric particle swarm optimization. In *European Conference on Genetic Programming*, pages 125–136, 2007.

[17] A. Moraglio and R. Poli. Topological interpretation of crossover. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 1377–1388, 2004.

[18] P. M. Pardalos and M. G. C. Resende, editors. *Handbook of Applied Optimization*. Oxford University Press, 2002.

[19] C. E. Rasmusen and C. Williams. *Gaussian Processes for Machine Learning*. the MIT Press, 2006.

[20] C. M. Reidys and P. F. Stadler. Combinatorial landscapes. *SIAM Review*, 44:3–54, 2002.

[21] I. J. Schoenberg. Metric spaces and positive definite functions. *Transaction American Mathematical Society*, 44:522–536, 1938.

[22] C. Scumacher, M. D. Vose, and L. D. Whitley. The no free lunch and problem description length. In *Proceedings of the Genetic and Evolutionary Computation Conference*, 2001.

[23] P. F. Stadler and R. Happel. Random field models for fitness landscapes. *Journal of'Mathematical Biology*, 38:435–478, 1999.

[24] E. D. Weinberger. Correlated and uncorrelated fitness landscapes and how to tell the difference. *Biological Cybernetics*, 63:325–336, 1990.

[25] D. H. Wolpert and W. G. Macready. No free lunch theorems for optimization. *IEEE Transaction on Evolutionary Computation*, 1(1):67–82, 1996.