How to detect all attraction basins of a function ?

Josselin Garnier and Leila Kallel

CMAP – UMR CNRS 7641 , Ecole Polytechnique,
Palaiseau 91128, France.
E-mail: *garnier@cmapx.polytechnique.fr*
E-mail: *kallel@cmapx.polytechnique.fr*

March 4, 1999

Abstract

This paper starts by a theoretical investigation of a family of landscapes characterized by the number of their local optima N and the distribution of the sizes (α_j) of their attraction basins. We then propose a practical methodology for identifying these quantities (N and (α_j) distribution) for an unknown landscape, given a random sample on that landscape and a local steepest ascent search.

This methodology applies to any landscape specified with a modification operator and provides bounds on search complexity (to detect all local optima) when using the modification operator at hand. Experiments demonstrate the efficiency of this methodology for guiding the choice of modification operators, leading to the design of problem-dependent optimization heuristics.

Introduction

In the field of stochastic optimization, two search techniques have been widely investigated during the last decade: simulated annealing [vA87] and evolutionary algorithms [Hol75, Gol89]. These algorithms are now widely recognized as methods of order zero for function optimization (no condition on function regularity). Basically, a population of candidate solutions is evolved in order to find the point that maximizes a given function. The evolution is directed by stochastic modification operators and selection scheme. However, the efficiency of these search algorithms (in terms of the time they require to reach the solution) is heavily dependent on the choice of the evolution operators, which in turn determine the *landscape* under optimization. In this context, this paper provides a new methodology allowing to investigate and to compare landscapes specified in relation to different modification operators.

Formally, a landscape can be denoted by $\mathbb{L} = (f, (\mu 1, \mu 2 \ldots))$ where f is the function to optimize and $(\mu 1, \mu 2 \ldots)$ the modification operators that are applied to elements of the search space. These operators define the *neighborhood relation* of the landscape: For example, in the binary framework (the search space is $E = \{0, 1\}^l$), the modification operator (commonly named mutation) that flips exactly one bit of each bitstring yields the Hamming neighborhood relation, where each string has l neighbors (reachable by the mutation operator in one step). The mutation that flips two bits of each string, yields a completely different neighborhood relation. The structure of the landscape, heavily depends on the neighborhood relation induced by the evolution operators, which in turn may depend on the choice of the representation (the coding of the candidate solutions into binary or gray strings for example). Hence, before the optimization process can be started, there is number of practical choices (representation and operators) that determine the landscape structure. Consequently, these choices are often crucial for the success of stochastic search algorithms.

Some research has studied how the fitness landscape structure impacts the potential search difficulties [Wei90, MdWS91, Sta95, Sta96]. This work shows that every complex fitness landscape can be represented as an **expansion of elementary landscapes** –one term in the Fourier expansion– which are easier to search in most cases. This work has been practically applied to solve a difficult NP-complete problem [SN98] (the identification of minimal finite k-state automaton for a given input-output behavior). Other theoretical studies of search feasibility consider the **whole landscape** as a tree of local optima, with a label describing the depth of the attraction basin at each node [PMV87, RT98]. Such a construction naturally describes the inclusion of the local attraction basins present in the landscape. These studies investigate tree structures that ensure a minimal correlation between the strength of the local optima and their proximity to the global optimum, with respect to an ultra-metric distance on the tree. However, from a practical point of view, the tree describing the repartition of local optima is unknown and too expensive (in terms of computational cost) to determine for a given landscape. The lack of an efficient method at reasonable cost that allows one to characterize a given landscape, motivates the construction of heuristics for problem dependent tuning of EA parameters (which determine the landscape as explained above).

A number of work are then devoted to extract *a priori* statistical information about landscape difficulty, most based on a sampling of the search space. Such information can be used to guide the choice of EA parameters. We cite, –from the field of evolutionary algorithms: Fitness Distance relations, first proposed in [JF95] and successfully used to choose problem dependent random initialization procedures [KS97a, MF98] ; Fitness Improvement of evolution operators, first proposed in [FG96], then extended and successfully used to choose binary crossover operators [KS97b] and representations [Kal98]. However, even if such heuristics can guide the –*a priori*– choice of some EA parameters, they do not give significant information about landscape structure, for instance, recent work suggests that very different landscapes (and leading to different EA behaviors) can share the same fitness distance relation [QRSS98, KNS98]. Further, the efficiency of such summary statistics is limited to the sampled regions of the space, and therefore does not necessarily help the long term convergence results (as implicitly illustrated in [KS97b] for example). This gives strong motivation for developing tools that allow one to derive a more global (beyond the sampled regions) information on the landscape at hand, relying on an implicit assumption of stationarity of the landscape. Along that line, this paper proposes a new method to identify the number and the repartition of local optima (w.r.t to a given neighborhood relation) of a given landscape. The proposed method applies to any neighborhood relation specified with a modification operator, and hence provides a practical tool to compare landscapes obtained with different operators and representations.

The framework is the following. We assume that that the search space E can be split into the partition E_1,...,E_N of subspaces which are attraction basins of local maxima m_1,...,m_N

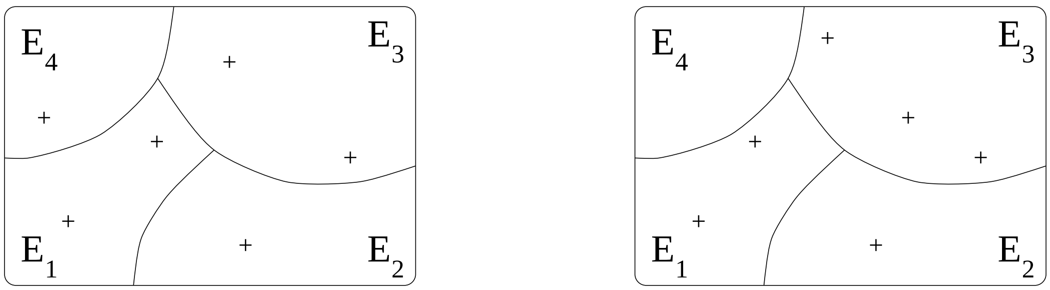


FIG 1 : *Schematic representations of the search space E with $N = 4$ attraction basins. $M = 6$ points have been randomly placed on both pictures. As a result there is at least one point in each attraction basin in the left picture, but not in the right picture.*

of the fitness function. We also assume that there exists a local search algorithm (for example a steepest ascent) which is able to find from any point of the search space the closest local maximum:

$$\Theta : \begin{cases} E & \to & \{m_1, ..., m_N\} \\ x & \mapsto & m_j \text{ if } x \in E_j \end{cases}$$

The basic problem consists in detecting all local maxima m_j. This is equivalent to find a way to put a point in all attraction basins E_j, because the local search algorithm will complete the job. We shall develop the following strategy. First we shall study the direct problem, which consists in studying the covering of the search space by a collection of points randomly distributed when the partition (E_j) is known. Second we shall deal with the inverse problem which consists in estimating the number of local maxima from information deduced from the covering.

- Direct problem (section 3): One puts M points randomly in the search space. The question is the following: Given the statistical distribution of the relative sizes of the attraction basins and their number N, what is the probability $p_{N,M}$ that at least one point lies in every attraction basin ? This probability is very important. Indeed, using the local search algorithm, it is exactly equal to the probability to detect all local maxima of the function.

- Inverse problem (section 4): The statistical distribution of the relative sizes of the attraction basins and their number are assumed to be known for computing $p_{N,M}$ in section 3. Unfortunately, this is rarely the case in practical situations, and one wants to estimate both. The strategy is to put randomly M initial points in the search space and to detect the closest local maxima by the local search algorithm. The data we collect is the set $(\beta_j)_{j\geq 1}$ of the number of maxima detected with j initial points. Of course β_0 is unknown (number of local maxima of the landscape that have not been detected). The question is the following: How can the total number of local maxima $N = \sum_{j=0}^{\infty} \beta_j$ be efficiently estimated from the set $(\beta_j)_{j\geq 1}$? A lower bound is $\bar{N} = \sum_{j=1}^{\infty} \beta_j$, but we aim at constructing a better estimator.

The paper is divided into three parts. First, section 3 addresses the direct problem of sample sizing in the case of basins of random sizes then in the case of basins of equal sizes. Second, the inverse problem of estimating the distribution of the relative basins sizes for an unknown landscape, using a random sample from the search space. This is achieved by a two step methodology: Section 4.2 starts by considering a parametrized family of laws for the relative basins sizes, for which it derives the corresponding covering of the search space (law of (β_j)). Then, section 4.3 comments on how these results can be practically used for characterizing the basins sizes of an unknown landscape. For instance, it proposes to compare the covering of an unknown landscape (given by the empirically observed (β_j) values) to the coverings studied in section 4.2. Finally, the last part of the paper (section 5) is devoted to some experiments that validate (section 5.1) and illustrate (section 5.2) the methodology: First, a landscape is purposely designed to test the reliability of the method according to the size of the random sample, and to the number of local optima (recall the theoretical results are asymptotic with respect to N and M). Second, the method is used to investigate some new landscapes (permutated onemax problems) and to compare the landscapes related to different mutation operators.

1 Notations and Definitions

Consider a fitness $f : E \to \mathbb{R}$, and a neighborhood relation induced by a modification operator μ, such that the number of different *μ-neighbors* (neighbors that can be obtained by one application of μ to x) of $x \in E$ is 'bounded'. In the following, we denote by N the **number of local optima** of L, and by (α_j) the random variables describing the **sizes of the attraction basins** of L (normalized to the average size).

Steepest Ascent Algorithm (SA):

<u>Input:</u> a fitness $f : E \to \mathbb{R}$, an operator μ and a point $X \in E$.

<u>Algorithm:</u> Modify X by repeatedly performing the following steps:

- -- Record, for all μ-neighbors of X (noted $\mu^i(X)$): $(i, f(\mu^i(X)))$
- -- Assign $X = \mu^i(X)$ where i is chosen such that $f(\mu^i(X)) - f(X)$ reaches the highest possible value (this is the steepest ascent).
- -- Stop when no strictly positive improvement in μ-neighbors fitnesses has been found.

<u>*Output:*</u> The point X, denoted $\mu^{SA}(X)$.

Attraction basin: The attraction basin of a local optimum m_j is the set of points $X_1, \ldots X_k$ of the search space such that a steepest ascent algorithm starting at X_i $(1 \leq i \leq k)$ ends at the local optimum m_j. The normalized size of the attraction basin of the local optimum m_j is then equal to $k/|E|$.

REMARKS:

1. This definition of the attraction basins yields a partition of the search space into different attraction basins, as illustrated in figure 1. The approach proposed in this paper is based on this representation of the search space into a partition of attraction basins, and could be generalized to partitions defined with alternative definitions of attraction basins.

2. In the presence of local constancy in the landscape, the above definition of the steepest ascent (and hence also the related definition of the attraction basins) is not rigorous. For instance, if the fittest neighbors of point p have the same fitness value, then the steepest ascent algorithm at point p have to make a –random or user defined– choice.

3. In the presence of a big amount of local constancy, the obtained attraction basins may vary considerably according to the steepest ascent choices in such equi-fitted areas. Nevertheless, even in the presence of local constancy, the comparison of the results (distribution of (α_j)) obtained with different steepest ascent choices, may give useful information about the landscape and guide the best elitism strategy: 'move' to fitter points, or 'move' to strictly fitter points only.

4. Note finally that to overcome the problem of local constancy, one have to find alternative definitions of the attraction basins. A different approach of the notion of attraction basins can be found in [PMV87, RT98].

2 Summary of the results

Section 3 considers the following direct problem: Given a distribution of (α_j), we propose to determine M_{min}, the minimal size of a random sample of the search space, in order to sample at least one point in each attraction basin of the landscape. Two particular cases are investigated:

1. *Deterministic configuration*: all the attraction basins have the same size ((α_j) are deterministic),
2. *Random configuration*: the sizes of the attraction basins are completely random ((α_j) are uniformly distributed).

In both configurations, we give the value of M_{min} as a function of the number of local optima N. For instance, a random sample of size $M_{min} = N(\ln N + \ln a)$ for the deterministic configuration (resp. $M_{min} = aN^2$ for the random configuration), ensures that a point is sampled in each attraction basin with probability $\exp(-1/a)$. Further, for each configuration, the variance of M is given.

Section 4 adresses the inverse problem, where the landscape is unkown. The goal is then to determine N the number of local optima of the landscape, as well as the distribution of the normalised sizes $(\alpha_j)_{j=1...N}$ of the attraction basins. In order to address this inverse problem, some direct analysis is first required: Given a distribution of $(\alpha_j)_{j=1...N}$, we propose to determine the distribution of a random family (β_j), defined as below.

• **Direct analysis:** Consider a random sample $(X_i)_{i=1...M}$ uniformly chosen in the search space. For each $i \in \{1, \dots M\}$, perform a steepest ascent starting at X_i (with the modification operator(s) at hand μ) ending at the local optimum $\mu^{SA}(X_i)$. Then define β_j as the number of local optima $(m_.)$ that have been reached by exactly j points from (X_i) (see example in Figure 2).

$$\beta_j = card\{k; card\{i; \mu^{SA}(X_i) = m_k\} = j\}.$$

The key result of proposition 4.1, gives (asymptotically with respect to N) the distribution of (β_j), for a class of parametrized distributions Law_γ for (α_j). More precisely, if $(Z_j)_{j=1...N}$

denotes a family of independent random variables with density:

$$p_\gamma(z) = \frac{\gamma^\gamma}{\Gamma(\gamma)} z^{\gamma-1} e^{-\gamma z},$$

and $\alpha_j = \frac{Z_j}{\sum_{i=1}^N Z_i}$, then the expected number $\beta_{j,\gamma} := \mathbb{E}_\gamma[\beta_j]$ is:

$$\beta_{j,\gamma} = N \left. \frac{\Gamma(j+\gamma)}{j!\Gamma(\gamma)} \frac{a^j \gamma^\gamma}{(a+\gamma)^{j+\gamma}} \right|_{a=\frac{M}{N}}$$

Moreover, the ratio $r = M/N$ is the unique solution of:

$$\frac{\sum_{j=1}^\infty \beta_{j,\gamma}}{M} = \frac{1-(1+\frac{r}{\gamma})^{-\gamma}}{r} \tag{1}$$

The latter equation is then used to find a good estimator of N, with observed values of the variables β_j, as explained below.

• **The inverse problem:** Given an unknown landscape, we then propose to characterize the distribution of (α_j) through the empirical estimation of the distribution of the random family (β_j). In fact, by construction, the distribution of (α_j) and that of (β_j) are tightly related: We experimentally determine observed values taken by (β_j) (random sampling and steepest ascent search). Then, for each γ value, we use a χ^2 test to compare the observed law for (β_j) to the law β should (theoretically) obey if the law of (α_j) were Law_γ. Naturally, we find a (possible) law for (α_j) iff one of the latter tests is positive. Otherwise, we only gain the knowledge that (α_j) does not obey the law Law_γ. Note also that the method can be used to determine sub-parts of the search space with a given distribution for (α_j). In case the law of (α_j) obeys Law_γ, Equation (1) is used to find a good estimator of N.

Last, section 5 validates the methodology of section 4, by considering known landscapes with random and deterministic basins sizes, showing that the estimations of the number of local optima N are accurate, even if M is much smaller than N. Further, we apply the methodology on unknown landscapes, and show that the *Hamming* binary and gray F1 landscapes contain much more local optima than the 3-bits-flip landscapes.

3 Direct problem

We assume that the search space E can be split into the partition E_1,...,E_N of subspaces which are attraction basins of local maxima m_1,...,m_N of the fitness function. Let us put a sample of M points randomly in the search space. We aim at computing the probabilities $p_{N,M}$ that at least one point of the random sample lies in each attraction basin. The basic result which will be applied in this section is the following.

PROPOSITION **3.1**
If we denote $\alpha_j := |E_j|/|E|$, *then:*

$$p_{N,M} = \sum_{k=0}^{N} (-1)^{N-k} \sum_{1\le j_1<...<j_k\le N} (\alpha_{j_1} + ... + \alpha_{j_k})^M \tag{2}$$

Proof. Let us denote by $(X_j)_{j=1,...,M}$ the collection of initial points. Since they are randomly chosen uniformly over E, we have for any $(j_1,...,j_M) \in \{1,...,N\}^M$:

$$\mathbb{P}\left(X_1 \in E_{j_1},...,X_M \in E_{j_M}\right) = \alpha_{j_1}...\alpha_{j_M}$$

By a straightforward combinatorial argument, this implies that the probability that k_1 points be in E_1, ..., and k_N points be in E_N is given by:

$$\mathbb{P}\left(k_1 \text{ points in } E_1,...,k_N \text{ points in } E_N\right) = \frac{M!}{k_1!...k_N!}\alpha_1^{k_1}...\alpha_N^{k_N}$$

where $\sum_{j=1}^N k_j = M$ and consequently:

$$p_{N,M}(\alpha) = \sum_{\substack{k_1\geq 1,...,k_N\geq 1\\ \sum k_j=M}} \frac{M!}{k_1!...k_N!}\alpha_1^{k_1}...\alpha_N^{k_N}$$

We introduce some new notations. Let N' be an integer and α' a N'-vector. If $1 \leq k_1 \neq ... \neq k_p \leq N'$, then we denote by $\check{\alpha'}^{k_1,...,k_p}$ the $(N'-p)$-vector constructed from α' by deleting the elements corresponding to the coordinates k_1,...,k_p. $\overline{\alpha'}$ is a shorthand for the sum of the coefficients $\sum_{j=1}^{N'} \alpha'_j$. Furthermore for any N'-vector α' we introduce:

$$p_{N',M}(\alpha') = \sum_{\substack{k_1\geq 1,...,k_{N'}\geq 1\\ \sum k_j=M}} \frac{M!}{k_1!...k_{N'}!}{\alpha'}_1^{k_1}...{\alpha'}_{N'}^{k_{N'}}$$

Since $\sum_{\substack{k_1\geq 0,...,k_{N'}\geq 0\\ \sum k_j=M}} \frac{M!}{k_1!...k_{N'}!}{\alpha'}_1^{k_1}...{\alpha'}_{N'}^{k_{N'}} = \overline{\alpha'}^M$, we can express $p_{N',M}(\alpha')$ as:

$$p_{N',M}(\alpha') = \overline{\alpha'}^M - \sum_{\substack{\exists j \text{ s. t. } k_j=0\\ \sum k_j=M}} \frac{M!}{k_1!...k_{N'}!}{\alpha'}_1^{k_1}...{\alpha'}_{N'}^{k_{N'}}$$

We decompose the sum in the right-hand member over all the possible subsets $\{k_1,...,k_p\}$ of $\{1,...,N'\}$ which correspond to the k_j's which are equal to 0:

$$p_{N',M}(\alpha') = \overline{\alpha'}^M - \sum_{p=1}^{N'} \sum_{1\leq k_1<..<k_p\leq N'} p_{N'-p,M}(\check{\alpha'}^{k_1,...,k_p}) \tag{3}$$

This expression is a recursive relation which allows to compute $p_{N,M}(\alpha)$ from $p_{N',M}(\alpha')$ for $N' \leq N-1$. First check that, for any scalar α we have $p_{1,M}(\alpha) = \alpha^M$ as soon as $M \geq 1$ and 0 otherwise. Now assume that $p_{N',M}(\alpha')$ is given by:

$$p_{N',M}(\alpha') = \sum_{k=0}^{N'}(-1)^k \sum_{1\leq j_1<...<j_k\leq N'} (\overline{\alpha'} - {\alpha'}_{j_1} - ... - {\alpha'}_{j_k})^M \tag{4}$$

for any $N' \leq N-1$ and for any N'-vector α'. In this sum the first term corresponding to

$k=0$ is $\overline{\alpha'}^M$. Then Eq. (3) implies that $p_{N,M}$ is given by:

$$p_{N,M}(\alpha) = \overline{\alpha}^M - \sum_{p=1}^{N} q_{p,N}(\alpha)$$
$$q_{p,N}(\alpha) := \sum_{1\le k_1<..<k_p\le N}\sum_{p'=0}^{N-p}(-1)^{p'}\sum_{1\le j_1<...<j_{p'}\le N-p}(\overline{\breve{\alpha}^{k_1,...,k_p}} - \breve{\alpha}_{j_1}^{k_1,...,k_p} - ... - \breve{\alpha}_{j_{p'}}^{k_1,...,k_p})^M$$

which can also be rewritten as:

$$p_{N,M}(\alpha) = \overline{\alpha}^M + q_{0,N}(\alpha) - \sum_{p=0}^{N} q_{p,N}(\alpha) \tag{5}$$
$$q_{p,N}(\alpha) = \frac{1}{p!}\sum_{k_1\neq..\neq k_p=1}^{N}\sum_{p'=0}^{N-p}\frac{(-1)^{p'}}{p'!}\sum_{j_1\neq...\neq j_{p'}=1}^{N-p}(\overline{\breve{\alpha}^{k_1,...,k_p}} - \breve{\alpha}_{j_1}^{k_1,...,k_p} - ... - \breve{\alpha}_{j_{p'}}^{k_1,...,k_p})^M \tag{6}$$

Since $\overline{\breve{\alpha}^{k_1,...,k_p}} = \overline{\alpha} - \alpha_{k_1} - ... - \alpha_{k_p}$, we get by grouping the terms $p+p'=p''$ together that:

$$\sum_{p=0}^{N} q_{p,N}(\alpha) = \sum_{p''=0}^{N}\sum_{k_1\neq..\neq k_{p''}=1}^{N}(\overline{\alpha} - \alpha_{k_1} - ... - \alpha_{k_{p''}})^M\left\{\sum_{p=0}^{p''}\frac{1}{p!}\frac{1}{(p''-p)!}(-1)^p\right\}$$

The sum within the brackets is the expansion of $(1-1)^{p''}/(p'')!$ which is equal to zero for $p''\ge 1$, so that:

$$\sum_{p=0}^{N} q_{p,N}(\alpha) = \overline{\alpha}^M \tag{7}$$

Furthermore, from (6),

$$q_{0,N}(\alpha) = \sum_{p'=0}^{N}\frac{(-1)^{p'}}{p'!}\sum_{j_1\neq...\neq j_{p'}=1}^{N}(\overline{\alpha} - \alpha_{j_1} - ... - \alpha_{j_{p'}})^M$$

which also reads as:

$$q_{0,N}(\alpha) = \sum_{p'=0}^{N}(-1)^{p'}\sum_{j_1<...<j_{p'}=1}^{N}(\overline{\alpha} - \alpha_{j_1} - ... - \alpha_{j_{p'}})^M \tag{8}$$

Substituting (7) and (8) into (5) establishes that (4) holds true for any N-vector α, which completes the proof of the Proposition. □

Proposition 3.1 gives an exact expression for $p_{N,M}$ which holds true whatever N, M and (α_j), but is quite complicated. The following corollaries show that the expression of $p_{N,M}$ is much simpler in some particular configurations.

Corollary **3.2**
1. If the attraction basins all have the same size $\alpha_j \equiv 1/N$ (the so-called D-configuration), then:

$$p_{N,M} = \sum_{k=0}^{N} C_N^k (-1)^k (1-k/N)^M$$

2. If moreover the numbers of attractors and initial points are large $N \gg 1$ and $M = N(\ln N + \ln a)$, $a > 0$, then:

$$p_{N,M} = e^{-a^{-1}}$$

3. Let us denote by M_D the number of points which are necessary to detect all local maxima. Then in the asymptotic framework $N \gg 1$, M_D obeys the distribution of

$$M_D = N \ln N - N \ln Z$$

where Z is an exponential variable with mean 1.

Proof. The first point is a straightforward application of Proposition 3.1. Let us assume that $N \gg 1$ and $M = N(\ln N + \ln a)$. Then $C_N^k \simeq \frac{N^k}{k!}$ and $(1-k/N)^M \simeq e^{-k(\ln N + \ln a)} = (Na)^{-k}$, which yields the second point of the corollary. The third point then follows readily. □

Corollary **3.3**
1. If the sizes of the attraction basins are random (the so-called R-configuration), in the sense that their joint distribution is uniform over the simplex of $\mathbb{R}^N$:

$$S_N := \{\alpha_i \geq 0, \sum_{i=1}^{N} \alpha_i = 1\}.$$

and the numbers of attractors and initial points are large: $N \gg 1$ and $M = N^2 a$, $a > 0$, then:

$$p_{N,M} = e^{-a^{-1}}$$

2. Let us denote by M_R the number of points which are necessary to detect all local maxima. Then in the asymptotic framework $N \gg 1$, M_R obeys the distribution of

$$M_R = N^2 Z^{-1},$$

where Z is an exponential variable with mean 1.

Remark: A construction of the R-configuration is the following. Assume that the search space E is the interval $[0,1)$. Choose $N-1$ points $(a_i)_{i=,1,\dots,N-1}$ uniformly over $[0,1]$ and independently. Re-index these points so that $a_0 := 0 \leq a_1 \leq \dots \leq a_{N-1} \leq a_N := 1$. Denote the spacings by $\alpha_j = a_j - a_{j-1}$ for $j = 1,\dots,N$. If the j-th attraction basin E_j is the interval $[a_{j-1}, a_j)$, then the sizes $(\alpha_j)_{j=1,\dots,N}$ of the attraction basins $(E_j)_{j=1,\dots,N}$ obey a uniform distribution over the simplex S_N.
Proof. In these conditions, we get from Eq.(2) and the relation $\sum_{j=1}^{N} \alpha_j = 1$ that:

$$p_{N,M} = \sum_{k=0}^{N} (-1)^k \sum_{1 \leq j_1 < \dots < j_k \leq N} \mathbb{E}\left[(1 - \alpha_{j_1} - \dots - \alpha_{j_k})^M\right] \quad (9)$$

where $\mathbb{E}$ stands for the expectation with respect to $(\alpha_j)_{j=1,\dots,N}$ whose distribution is uniform over S_N.
Let us introduce the order statistics associated with the sequence α, i.e. the unique N-vector $(\tilde{\alpha}_i)_{i=1,\dots,N}$ which satisfies:

$$\{\tilde{\alpha}_1, \dots, \tilde{\alpha}_N\} = \{\alpha_1, \dots, \alpha_N\}$$
$$\tilde{\alpha}_1 \leq \dots \leq \tilde{\alpha}_N$$

For any fixed k, the k-vector $(\tilde{\alpha}_i)_{i=1,\dots,k}$ admits the following distribution as N goes to infinity [Wil62, Section 9.6]:

$$\tilde{\alpha}_j = \frac{1}{N^2} \sum_{i=1}^{j} Z_i \; j = 1, \dots, k$$

where the family $(Z_i)_{i=1,\dots,k}$ consists of independent random variables with exponential density and mean 1.
Obviously we can substitute $\tilde{\alpha}$ for α in (9). From the fact that $\tilde{\alpha}_1$ is of order N^{-2} we guess that the number of points which are necessary to detect all attraction basins is of order N^2. Thus, with $M = aN^2$,

$$(1 - \tilde{\alpha}_j)^M \simeq \exp - \left(a \sum_{i=1}^{j} Z_i \right)$$

Since the Z_i's are independent, the expectation reads as (for $N \gg 1$):

$$\mathbb{E}\left[(1 - \tilde{\alpha}_j)^M\right] \simeq \mathbb{E}\left[\exp(-aZ_1)\right]^j = \left(\frac{1}{1+a}\right)^j$$

and the second term ($k = 1$) in the sum of the right-hand member in Eq. (9) reads as:

$$\sum_{j=1}^{N} \mathbb{E}\left[(1 - \alpha_j)^M\right] = \frac{1+a}{a}$$

More generally, if $j_1 < \dots < j_k$:

$$(1 - \tilde{\alpha}_{j_1} - \dots - \tilde{\alpha}_{j_k})^M \simeq \exp -a \left(k \sum_{i=1}^{j_1} Z_i + (k-1) \sum_{i=j_1+1}^{j_2} Z_i + \dots + \sum_{i=j_{k-1}+1}^{j_k} Z_i \right)$$

Taking the expectation:

$$\mathbb{E}\left[(1 - \tilde{\alpha}_{j_1} - \dots - \tilde{\alpha}_{j_k})^M\right] \simeq \left(\frac{1}{1+ka}\right)^{j_1} \left(\frac{1}{1+(k-1)a}\right)^{j_2 - j_1} \dots \left(\frac{1}{1+a}\right)^{j_k - j_{k-1}}$$

and summing (we have $\sum_{j=1}^{\infty} 1/(1+r)^j = 1/r$) :

$$\sum_{j_1 < \dots < j_k} \mathbb{E}\left[(1 - \alpha_{j_1} - \dots - \alpha_{j_k})^M\right] = \frac{a^{-k}}{k!}$$

Substituting into Eq. (9) completes the proof of the corollary. □

It follows from the corollaries that one needs about $N \ln N$ points in the D-configuration to detect all maxima, while one needs about N^2 points to expect the same result in the R-configuration. This is due to the fact that there exists very small attraction basins in the R-configuration. In the proof we state the result that the smallest attraction basin in the R-configuration has a relative size which obeys an exponential distribution with mean N^{-2}. That is why one needs of the order of N^2 points to detect this very small basin.
Mean values. The expected value of M_D is exactly:

$$\mathbb{E}[M_D] = N \ln N + NC$$

where C is the Euler's constant whose value is $C \simeq 0.58$. The expected value of M_R/N^2 is equal to infinity. This is due to the fact that the tail corresponding to exceptional large values of M_R is very important:

$$\mathbb{P}(M_R \geq N^2 a) = 1 - \exp(-a^{-1}) \underset{a \gg 1}{\simeq} a^{-1}$$

Standard deviations. The normalized standard deviation, which is equal to the standard deviation divided by the mean, of the number of points necessary to detect all local maxima in the D-configuration is equal to:

$$\sigma_D := \frac{\sqrt{\mathbb{E}[M_D^2] - \mathbb{E}[M_D]^2}}{\mathbb{E}[M_D]} = \frac{\pi}{\sqrt{6}(\ln N + C)}$$

which goes to 0 as $N \to \infty$, which proves in particular that $M_D/(N \ln N)$ converges to 1 in probability. This is of course not surprising. The D-configuration has a deterministic environment, since all basins have a fixed size, so that we can expect an asymptotic deterministic behavior. The situation is very different in the R-configuration which has a random environment, and it may happen that the smallest attraction basin be much smaller than its expected size N^{-2}. That is why the fluctuations of M_D, and especially the tail corresponding to exceptional large values, are very important.

4 Inverse problem

4.1 Formulation of the problem

We now focus on the inverse problem. We look for the number N of local maxima of the fitness function and also some pieces of information on the distribution of the sizes of the corresponding attraction basins. We assume that we can use an algorithm that is able to associate to any point of the search space the closest local maximum. In order to detect all local maxima, we should apply the algorithm to every point of the search space. Nevertheless this procedure is far too long since the search space has a large cardinality. Practically we shall apply the algorithm to M points that will be chosen randomly in the search space E. The result of the search process can consequently be summed up by the following set of observed values ($j \geq 1$):

$$\beta_j := \text{number of maxima detected with } j \text{ points} \tag{10}$$

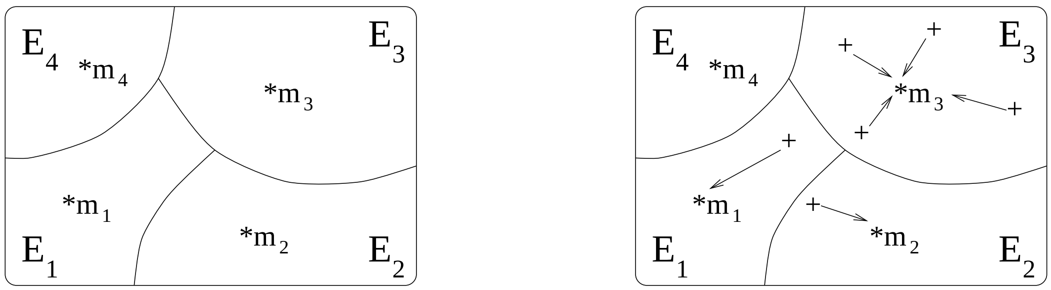


FIG 2 : *Schematic representation of the search space E with its $N = 4$ attraction basins and the 4 corresponding local maxima $m_1,...,m_4$. In the left picture we have put $M = 6$ points randomly chosen. We apply the search algorithm and detect 3 maxima according to the right picture, so that we have $\beta_1 = 2$, $\beta_2 = 0, \beta_3 = 0$, $\beta_4 = 1$, and $\beta_j = 0$ for $j \geq 5$.*

Our arguments are based upon the following observations. First notice that $\bar{N} := \sum_{j=1}^{\infty} \beta_j$ is the number of detected maxima. It is consequently a lower bound of the total number of local maxima N, but a very rough estimate in the sense that it may happen that many maxima are not detected, especially those whose attraction basins are small. Besides $\bar{N}$ represents less information than the complete set $(\beta_j)_{j\geq 1}$. By a clever treatment of this information, one should be able to find a better estimate of N than $\bar{N}$.

4.2 Analysis

The key point is that the distribution of the set β_j is closely related to the distribution of the sizes of attraction basins. Let us assume that the relative sizes $(\alpha_j)_{j=1,...,N}$ of the attraction basins can be described by a distribution parametrized by some positive number γ as follows. Let $(Z_j)_{j=1,...,N}$ be a sequence of independent random variables whose common distribution has density p_γ with respect to the Lebesgue measure over $(0, \infty)$:[1]

$$p_\gamma(z) = \frac{\gamma^\gamma}{\Gamma(\gamma)} z^{\gamma-1} e^{-\gamma z}, \tag{11}$$

where Γ is the so-called Euler's Gamma function $\Gamma(s) := \int_0^\infty e^{-t} t^{s-1} dt$. Under p_γ, the expected value of Z_1 is 1 and its standard deviation is $1/\sqrt{\gamma}$. In the following we shall say that we are under H^γ if the relative sizes of the attraction basins $(\alpha_j)_{j=1,...,N}$ can be described as $(Z_1/T_N, ..., Z_N/T_N)$ where $T_N := \sum_{j=1}^N Z_j$ and the distribution of Z_j has density p_γ. Notice that the large deviations principle (Cramer's theorem [Aze78, Chapter 1]) applied to the sequence (Z_j) yields that for any $x > 0$ there exists $c_{\gamma,x} > 0$[2] such that:

$$\mathbb{P}_\gamma \left(\left| \frac{T_N}{N} - 1 \right| \geq x \right) \leq \exp(-N c_{\gamma,x}) \tag{12}$$

which shows that, in the asymptotic framework $N \gg 1$, the ratio Z_j/N stands for the relative size α_j up to a negligible correction. The so-called D and R configurations described

[1] If γ is a positive integer then p_γ is a negative-binomial distribution.
[2] Applying the procedure described in [Aze78] shows that $c_{\gamma,x} = \gamma(x - 1 - \ln x)$.

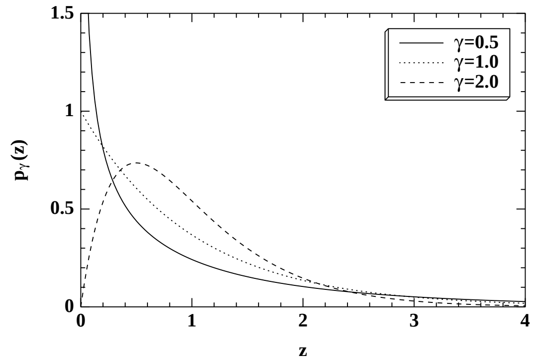


FIG 3 : *Probability density of the sizes of the attraction basins under H^γ for different γ.*

in Section 3 are particular cases of this general framework:
- For $\gamma = \infty$, $Z_j \equiv 1$ and $T_N = N$, so that we get back the deterministic D-configuration.
- For $\gamma = 1$, the Z_j's obey independent exponential distributions with mean 1, and the family $\alpha_j = Z_j/T_N$ obeys the uniform distribution over S_N [Pyk65].
The important statement is the following one.

PROPOSITION **4.1**
Under H^γ, the expected values $\beta_{j,\gamma} := \mathbb{E}_\gamma[\beta_j]$ of the β_j's can be computed and are given by:

$$\beta_{j,\gamma} = N \left. \frac{\Gamma(j+\gamma)}{j!\Gamma(\gamma)} \frac{a^j \gamma^\gamma}{(a+\gamma)^{j+\gamma}} \right|_{a=\frac{M}{N}} \tag{13}$$

Proof. Under H^γ, the probability that j of the M points lie in the k-th attraction basin can be computed explicitly:

$$\mathbb{P}_\gamma \left(j \text{ points in } E_k \right) = C_M^j \mathbb{E}_\gamma \left[\alpha_k^j (1-\alpha_k)^{M-j} \right]$$

where $\alpha_k = Z_k / \sum_i Z_i$ and $\mathbb{E}_\gamma$ stands for the expectation of Z_j with distribution p_γ. From Eq. (12), if $N \gg 1$ we can substitute N for $\sum_j Z_j$ so that:

$$\mathbb{P}_\gamma \left(j \text{ points in } E_k \right) = C_M^j N^{-j} \mathbb{E}_\gamma \left[Z_k^j (1 - Z_k/N)^{M-j} \right]$$

If $N \gg 1$ and $M = aN$, then we have:

$$\mathbb{P}_\gamma \left(j \text{ points in } E_k \right) = \frac{a^j}{j!} \mathbb{E}_\gamma \left[Z_k^j e^{-aZ_k} \right]$$

By computing the expectation and summing over $k = 1, ..., N$, we finally get (13). □

In particular, the distribution of the β_j's under the D-configuration is Poisson:

$$\beta_{j,\infty} = N e^{-M/N} \frac{1}{j!} \left(\frac{M}{N} \right)^j$$

while it is geometric under the R-configuration:

$$\beta_{j,1} = N \frac{1}{1+\frac{M}{N}} \left(\frac{\frac{M}{N}}{1+\frac{M}{N}} \right)^{j}$$

From Eq. (13) one can deduce that the following relation is satisfied by the ratio $r = M/N$:

$$\frac{\sum_{j=1}^{\infty} \beta_{j,\gamma}}{M} = \frac{1-(1+\frac{r}{\gamma})^{-\gamma}}{r} \tag{14}$$

4.3 Estimator of the number of local maxima

We have now sufficient tools to exhibit a good estimator of the number of local maxima. We remember the reader with the problem at hand. We assume that some algorithm is available to determine from any given point the closest local maximum. We choose randomly M points in the search space and detect the corresponding closest local maxima. We thus obtain a set of values $(\beta_j)_{j\geq 1}$ as defined by (10). We can then determine from the set of values $(\beta_j)_{j\geq 1}$ which configuration H^{γ_0} is the most probable, or at least which H^{γ_0} is the closest configuration of the real underlying distribution of the relative sizes of the attraction basins. The statistic used to compare observed and expected results is the so-called χ^2 goodness of fit test [Wet72, Section 8.10], which consists first in calculating for each γ:

$$T_\gamma := \sum_{j\in\Omega} \frac{(\beta_j - \beta_{j,\gamma})^2}{\beta_{j,\gamma}}$$

where $\Omega \subset \mathbb{N}$ is the set of the indices j for which $\beta_j \geq 1$:

$$\Omega := \{j, \beta_j \geq 1\}$$

Obviously a large value for T_γ indicates that the corresponding $\beta_{j,\gamma}$ are far from the observed ones, that is to say H^γ is unlikely to hold. Conversely, the smaller T_γ, the more likely H^γ holds true. In order to determine the significance of various values of $T_.$, we need the distribution of the statistic. A general result says that if the hypothesis H^{γ_0} does hold true, then the distribution of T_{γ_0} is approximatively the so-called χ^2-distribution with degrees of freedom equal to the cardinality of the set Ω minus 1. Consequently we can say that the closest configuration of the real underlying distribution of the relative sizes of the attraction basins is H^{γ_0} where γ_0 is given by:

$$\gamma_0 = \text{argmin}\,\{T_\gamma, \gamma > 0\} \tag{15}$$

Furthermore, one can estimate the accuracy of the configuration H^{γ_0} by referring T_{γ_0} to tables of the χ^2-distribution for card$(\Omega) - 1$ degrees of freedom. A value of T_{γ_0} much larger than the one indicated in the tables mean that none of the configurations H^γ hold true. Nevertheless, H_0^γ is the closest distribution of the real one.

REMARK: The distribution theory of χ^2 goodness of fit statistic can be found in [Cra46, Chapter 30]. The result is in any case approximate, and all the poorer as they are many expected $\beta_{j,\gamma}$ less than five. These cases must be avoided by combining cells. But one then loose power in the tail regions, where differences are more likely to show up.

Defining γ_0 as (15), we denote by $\bar{\beta}$ the quantity:

$$\bar{\beta} = \frac{\sum_{j=1}^{\infty} \beta_j}{M}.$$

From Eq. (14), under H^{γ_0} the ratio $\alpha = M/N$ is the unique solution of:

$$\bar{\beta} = \frac{1 - (1 + \frac{r}{\gamma_0})^{-\gamma_0}}{r} \tag{16}$$

Consequently, once we have determined γ_0, formula (16) is a good estimator of the ratio $\alpha = M/N$, hence N.

5 Experiments

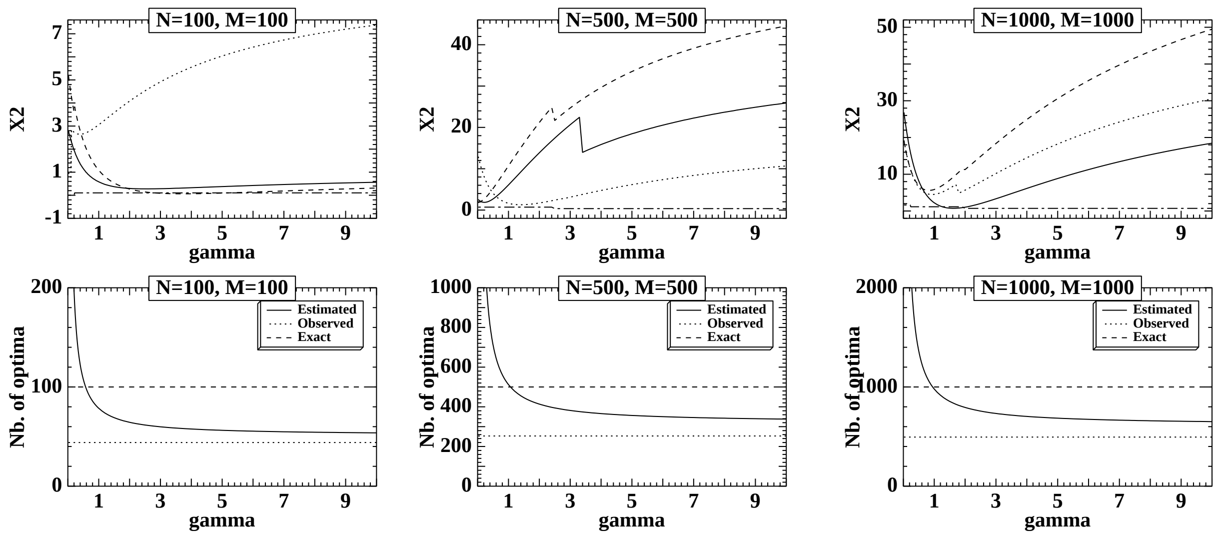
Given a landscape Ł, the following steps are performed in order to identify a possible (or closest) law for the number and size of the attraction basins of Ł, among the family of laws Law_γ studied above.

1. `choose a random sample` $(X_i)_{i=1,\dots,M}$ `uniformly in` E.
2. `perform a steepest ascent starting at each` X_i `ending at` $\mu^{SA}(X_i)$.
3. `compute` β_j, `defined as the number of local optima reached by exactly` j `initial points` X_i.
4. `compare the observed law of` β `to the laws of` $\beta(\gamma)$ `for different` γ `values, using the` χ^2 `test.`

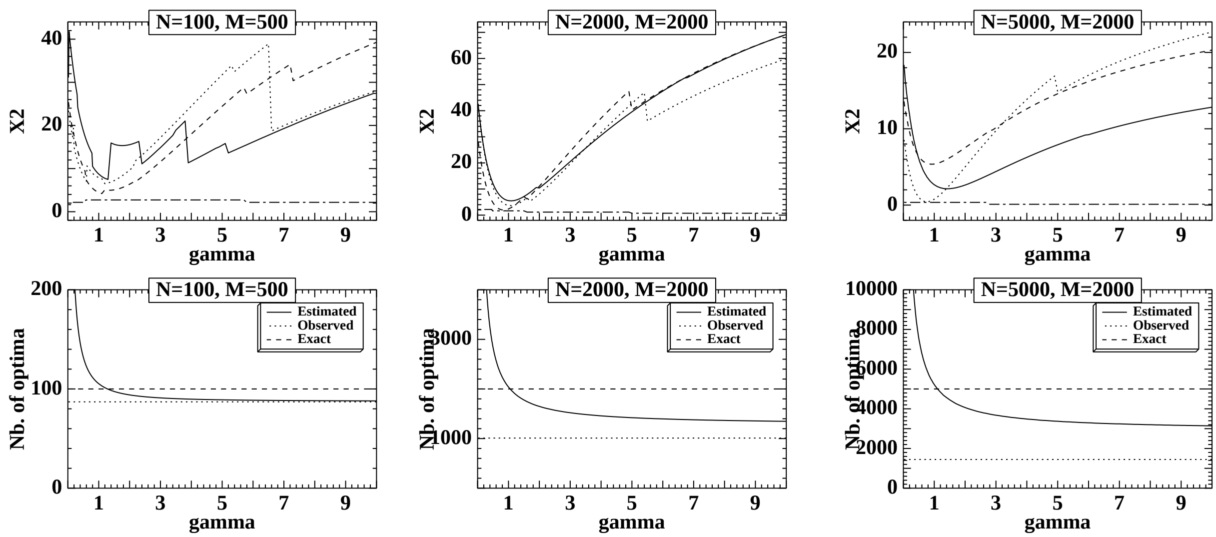
To visualize the comparison of the last item, we propose to plot the obtained χ^2 value for different γ values. We also plot the corresponding χ^2 value below which, the test is positive with a confidence of 95 %.

5.1 Experimental validation

The results obtained in section 4 are asymptotic with respect to the number of local optima N and in the size of the random sample M. Hence before, the methodology can be applied, some experimental validation is required in order to determine practical values for M and N for which the method is reliable. This is achieved by applying the methodology to determine the distribution of (α_j) (normalized size of the attraction basins) in two known purposely constructed landscapes: the first contains basins with random sizes, the second contains basins with equal sizes. Results are plotted in Figures (4) and (5). Samples with smaller size than those shown in these figures yield β_j values which are not rich enough to allow a significant χ^2 test comparison. For instance, the χ^2 test requires that observed β_j are non-null for some $j > 1$ at least (some initial points are sampled in the same attraction basin).



Unstable results *for the χ^2 test: Typical χ^2 results for small N values and $M = N$.*



Stable results *are obtained when N increases and M is bounded ($M \leq min(2000, 3 * N)$ here). The estimation of N corresponding to the smallest χ^2 value is very accurate.*

FIG 4 : **Basins with random uniform sizes***: The upper plots give the χ^2 test results comparing the empirically observed β distribution to the family of γ-parametrized distributions. The plots of the second row, give for different γ values, the estimation of the number of local optima (computed by Eq. 16). These estimations are very robust (only one estimation is plotted) and are accurate for $\gamma_0 = 1$. The same plot shows the* **observed** *number of optima, actually visited by the steepest ascent ($\bar{N} = \sum_{j=1}^{\infty} \beta_j$).*

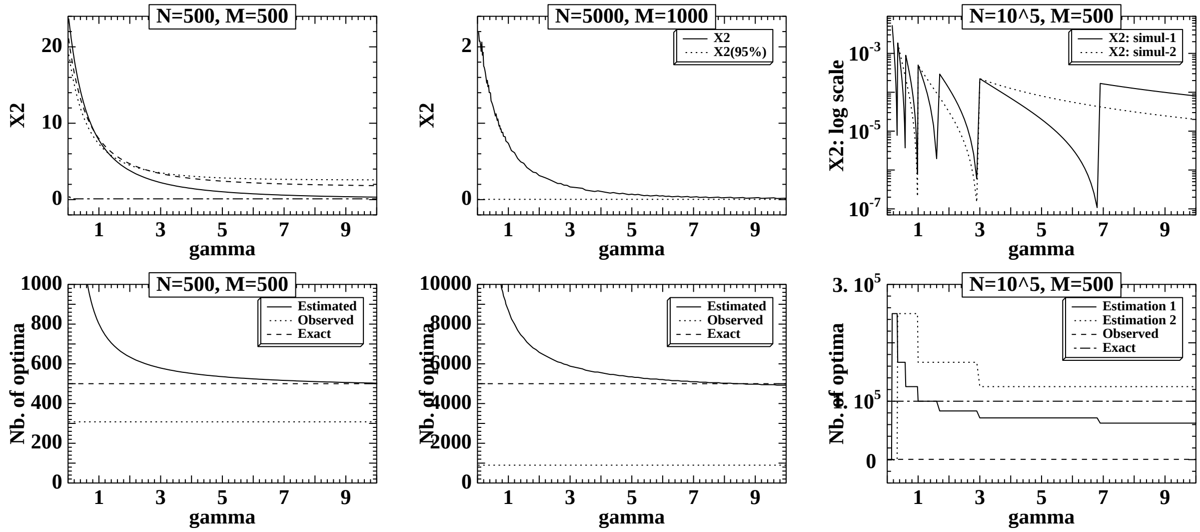


FIG 5 : **Basins with Deterministic equal sizes**: *The χ^2 results are stable for smaller sample sizes than those of the random configuration. The rightmost plot shows a case where the χ^2 test is not significant, yet the predicted number of local optima is very accurate! With 500 initial points, 497 local optima have been observed, while there are actually 10^5 optima. Yet, formula 16 is able to estimate the true number with an error of less than 30% when the adequate γ value is used.*

In case all initial points are sampled in different attraction basins the χ^2 test comparison is not significant. These experiments give practical bounds on the sample sizes (in relation to the number of local optima) for which the methodology is reliable (Figure 4). Further, we demonstrate that the estimation of number of local optima is accurate, even when initial points visit a small number of attraction basins of the landscape (Figure 5).

5.2 The methodology at work

Having seen that the methodology is a powerful tool, provided that the information obtained for β is rich enough, we apply it to investigate the landscape structure of the difficult gray and binary coded F1 Baluja problems [Bal95], for a 1-bit-flip and 3-bit-flips neighborhood relations. At problem size $\ell = 27$, considering the 1-bit-flip mutation (Figure 6), the distribution of the sizes of the basins is closer to the random configuration than to the deterministic one, and that the estimated number of local optima is similar for the binary and gray codings. On the other hand, considering the 3-bit-flip mutation (Figure 7), the landscape structure is completely different. The estimated number of local optima drops significantly for both problems: less than 250 for both binary and gray landscapes, whereas the Hamming landscape contains thousands of local optima (Figure 6).

Gray-coded Baluja F1 functions: Consider the function of k variables $(x_0, \dots x_{k-1})$, with $x_i \in [-2.56, 2.56]$ [Bal95]:

$$F1(\vec{x}) \quad = \quad \frac{100}{10^{-5} + \sum_{i=0}^{k-1} |y_i|}, y_0 = x_0 \text{ and } y_i = x_i + y_{i-1} \text{ for } i = 1, \dots, k-1$$

It reaches its maximum value among 10^7 at point $(0, \dots 0)$. The Gray-encoded F1g and binary F1b versions, with resp. 2, 3 and 4 variables encoded on 9 bits each are considered.

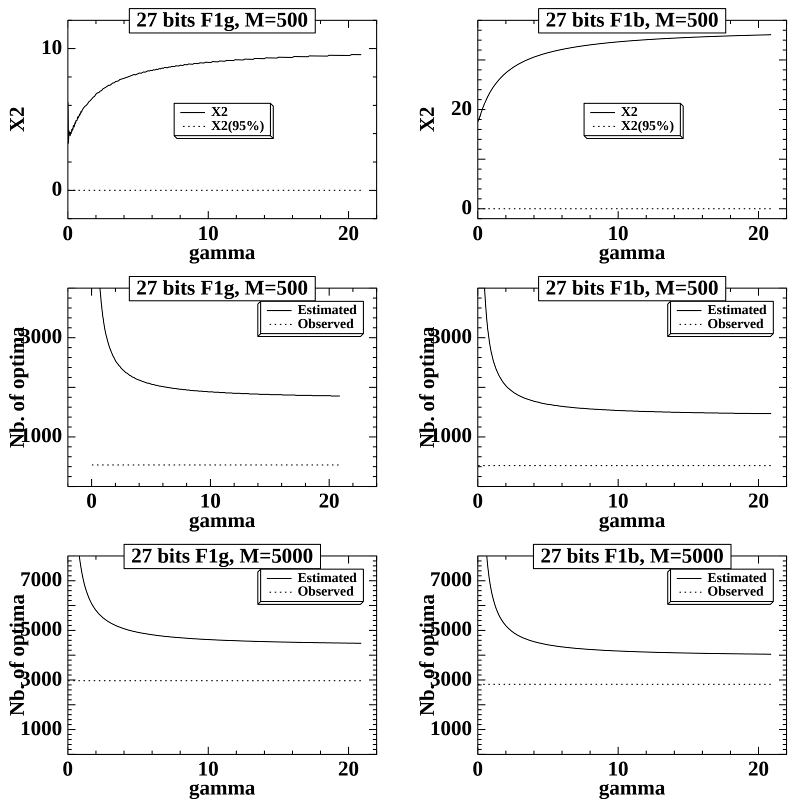


FIG 6 : *The difficult baluja 27-bits F1 gray (F1g) and binary (F1b) landscapes with a 1-bit-flips mutation. Experiments with larger sample sizes, show the same results for the* χ^2 *test, but show that the estimated number of local optima increases and stabilizes when sample sizes is larger than 2000. An example is shown in the third line with 5000 size sample.*

Experiments at problem sizes 18, 27 and 36 have been carried out, leading to similar results for both F1g and F1b problems: The number of local optima of the 3-bit-flips landscape is significantly smaller than that of the Hamming landscape. For example, when $\ell = 18$ (resp. 36), the number of local optima drops from hundreds to less than 10 (resp. from more than 25 000 to less than 1400).
A final remark is that at 18 and 27 problem sizes, the estimated number of local optima is close for binary and gray encodings; Whereas for $\ell = 36$, there are more (estimated) Hamming local optima for the gray (45 000) than for the binary (25 000) encoding.

A new optimization heuristic: A simple search strategy for solving difficult problems naturally follows from the methodology presented in this paper.

Once the number N and distribution of the attraction basins is estimated following the guidelines summarized in the beginning of section 5, generate a random sample which size is set to $N(\ln N + \ln a)$ if the sizes of the basins are close to the deterministic configuration (resp. aN^2 if the basins sizes are close to random). Then a simple steepest ascent starting at each point of the sample, ensures that the global optimum is found with probability $\exp(-1/a)$.

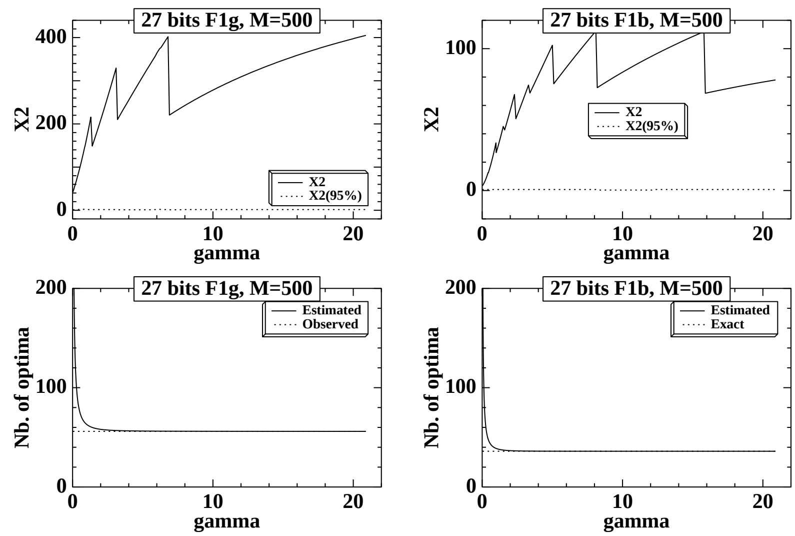


FIG 7 : *The difficult baluja 27-bits F1 gray (F1g) and binary (F1b) landscapes with a 3-bit-flips mutation: the number of local optima drops significantly compared to the Hamming 1-bit-flip landscape. These results are confirmed by experiments using a samples of size 5000.*

In the 27-bits F1 problem, this heuristic demonstrates to be very robust and efficient in solving the problem with the 3-bits-flip operator. Using a 3-bits-flip mutation steepest ascent, an initial random sample of 5 points (versus 100 with 1-bit-flip mutation) is enough to ensure that one point at least lies in the global attraction basin (experiments based on 50 runs)! This is due to the fact that the basin of the global optimum is larger than the basins of the other local optima. In order to detect all attraction basins, we can estimate the required sample size to 62500 ($250 * 250$ using corollary 3.3 and the estimation of $N = 250$ in the experiments of Figure 7).

5.3 Discussion

The efficiency of the proposed method is certainly dependent on the class of laws of (α_j) (sizes of the attraction basins) for which the distribution of β is known. We have chosen a very particular family of distributions p_γ for representing all possible distributions for the relative sizes of attraction basins. The constraints for this choice are twofold and contradictory. On the one hand, a large family of distributions is required to be sure that at least one of them is sufficiently close of the observed repartition (β_j). On the other hand, if we choose an over-large family, then we need a lot of parameters to characterize the distributions. It is then very difficult to estimate all these parameters and consequently to decide which distribution is the closest of the observed one. That is why the choice of the family is very delicate and crucial. We feel that the particular family p_γ that has been chosen (11) fulfills determinant conditions. First it contains two very natural distributions, the so-called D and R configurations that we have studied with great detail. Second it is characterized by a single parameter easy to estimate. Third it contains distributions with a complete range of variances, from 0 (the D-configuration) to infinity, by going through 1 (the R-configuration).

On the other hand, th experiments carried out in this paper, point out the need of refining the family of studied landscapes for 'close' to random configurations (random basin sizes). Along that line, an alternative choice for the parametrized set of distributions (α_j) could be $\alpha_j = Z_j / \sum_{i=1}^N Z_i$ where Z_j are independent and identically distributed with one of the distributions of the bidimensional family $p_{\gamma,\delta}(.)$, $\gamma > 0$, $\delta > 0$:

$$p_{\gamma,\delta}(z) = \frac{\delta^\gamma}{\Gamma(\gamma)} z^{\delta-1} e^{\gamma z}.$$

The parameter δ characterizes the distribution of the sizes of the small basins, since $p_{\gamma,\delta}(z) \sim z^{\delta-1}$ as $z \to 0$, while γ characterizes the distribution of the sizes of the large basins, since the decay of $p_{\gamma,\delta}(z)$ as $z \to \infty$ is essentially governed by $e^{-\gamma z}$. This family presents more diversity than the family $p_\gamma(.)$ we have considered in section 4.2. The expected value of β_j is under $p_{\gamma,\delta}$:

$$\beta_{j,\gamma,\delta} = N \frac{\Gamma(j+\gamma)}{j!\Gamma(\gamma)} \frac{a^j \delta^\gamma}{(a+\delta)^{j+\gamma}} \bigg|_{a=M/N}.$$

The method of estimating the number of local minima described in section 4.3 can then be applied with this family.

A first application of the methodology presented here is to compare landscapes obtained when different operators are used (k-bit flip binary mutations for different k values for example). However, the complexity of this method is directly related to size of the neighborhood of a given point. Hence, its practical usefulness to study k-bit-flip landscapes is limited when k value increases. Hence, it seems most suited to investigate different representations. Its extension to non-binary representations is straightforward, provided that a search algorithm that leads to the closest local optimum can be provided for each representation. Further, this methodology can be used to determine sub-parts of the search space, such that (α_j) follow a particular law, hence guiding a hierarchical search in different subparts of the space.

Note finally that the distribution of the basins sizes does not fully characterize landscape difficulty. Depending on the relative position of the attraction basins, the search still may range from easy to difficult. The question of how different landscapes can share the same distribution of the attraction basins, is left unanswered here. Additional information is necessary to compare landscapes difficulty. Further work may address such issues to extract additional significant information in order to guide the choice or the design of problem dependent operators and representations.

References

[Aze78] R. Azencott. Grandes déviations et applications. In P. L. Hennequin, editor, *Proceedings of Ecole d'Eté de Probabilités de Saint-Flour*. Springer-Verlag, 1978.

[Bal95] S. Baluja. An empirical comparison of seven iterative and evolutionary function optimization heuristics. Technical Report CMU-CS-95-193, Carnegie Mellon University, 1995.

[Cra46] H. Cramer. *Mathematical methods of statistics.* Princeton University Press, New Jersey, 1946.

[FG96] D. B. Fogel and A. Ghozeil. Using fitness distributions to design more efficient evolutionary computations. In T. Fukuda, editor, *Proceedings of the Third IEEE International Conference on Evolutionary Computation*, pages 11–19. IEEE, 1996.

[Gol89] D. E. Goldberg. *Genetic algorithms in search, optimization and machine learning.* Addison Wesley, 1989.

[Hol75] J. H. Holland. *Adaptation in natural and artificial systems.* University of Michigan Press, Ann Arbor, 1975.

[JF95] T. Jones and S. Forrest. Fitness distance correlation as a measure of problem difficulty for genetic algorithms. In L. J. Eshelman, editor, *Proceedings of the* 6^{th} *International Conference on Genetic Algorithms*, pages 184–192. Morgan Kaufmann, 1995.

[Kal98] L. Kallel. *Convergence des algorithmes génétiques : aspects spatiaux et temporels.* PhD thesis, Ecole Polytechnique – Palaiseau – France, 1998.

[KNS98] L. Kallel, B. Naudts, and M. Schoenauer. On functions with a fixed fitness distance relation. Technical Report 399, CMAP, Ecole Polytechnique, 1998.

[KS97a] L. Kallel and M. Schoenauer. Alternative random initialization in genetic algorithms. In Th. Bäeck, editor, *Proceedings of the* 7^{th} *International Conference on Genetic Algorithms*, pages 268–275. Morgan Kaufmann, 1997.

[KS97b] L. Kallel and M. Schoenauer. A priori predictions of operator efficiency. In J.-K. Hao, E. Lutton, E. Ronald, M. Schoenauer, and D. Snyers, editors, *Artificial Evolution'97*, LNCS. Springer Verlag, October 1997.

[MdWS91] B. Manderick, M. de Weger, and P. Spiessens. The genetic algorithm and the structure of the fitness landscape. In R. K. Belew and L. B. Booker, editors, *Proceedings of the* 4^{th} *International Conference on Genetic Algorithms*, pages 143–150. Morgan Kaufmann, 1991.

[MF98] P. Merz and B. Freisleben. Memetic algorithms and the fitness landscape of the graph bi-partitioning problem. In A.E. Eiben, Th. Bäeck, M. Schoenauer, and H.-P. Schwefel, editors, *Proceedings of the* 4^{th} *Conference on Parallel Problems Solving from Nature*, volume 1498 of *LNCS*, pages 765–774. Springer Verlag, 1998.

[PMV87] G. Parisi, M. Mezard, and M.A. Virasoro. *Spin glass theory and beyond.* World Scientific, 1987.

[Pyk65] R. Pyke. Spacings. *J. Roy. Statis. Soc.*, 27:395–436, 1965.

[QRSS98] R.J. Quick, V.J. Rayward-Smith, and G.D. Smith. Fitness distance correlation and ridge functions. In A.E. Eiben, Th. Bäeck, M. Schoenauer, and H.-P. Schwefel, editors, *Proceedings of the* 4^{th} *Conference on Parallel Problems Solving from Nature*, volume 1498 of *LNCS*, pages 77–86. Springer Verlag, 1998.

[RT98] R. Rammal and Toulouse. *Ultrametricity for Physicists.* Reviews of Modern Physics, 198.

[SN98] V. Slavov and N.I. Nikolaev. Genetic algorithms, fitness sublandscapes and subpopulations. In W. Banzhaf and C. Reeves, editors, *Foundations of Genetic Algorithms 5.* Morgan Kaufmann, 1998.

[Sta95] P. Stadler. *Complex systems and binary networks*, chapter Towards a theory of landscapes, pages 77–163. Springer Verlag, 1995.

[Sta96] P. Stadler. Landscapes and their correlation functions. *Journal of Mathematical chemistry*, 20:1–45, 1996.

[vA87] P. J. van Laarhoven and E. H. L. Aarts. *Simulated Annealing: Theory and Applications*. Kluwer Academic Press, Dordrecht, The Netherlands, 1987.

[Wei90] E.D. Weinberger. Correlated and uncorrelated fitness landscapes and how to tell the difference. *Biological Cybernetics*, 63:325–336, 1990.

[Wet72] G. B. Wetherill. *Elementary statistical methods*. Chapman and Hall, London, 1972.

[Wil62] S. S. Wilks. *Mathematical statistics*. Wiley, New York, 1962.