EFFICIENCY OF LOCAL SEARCH WITH MULTIPLE LOCAL OPTIMA

JOSSELIN GARNIER*† AND LEILA KALLEL*

Abstract. The first contribution of this paper is a theoretical investigation of combinatorial optimization problems. Their landscapes are specified by the set of neighborhoods of all points of the search space. The aim of the paper consists in the estimation of the number N of local optima and the distributions of the sizes (α_j) of their attraction basins. For different types of landscapes we give precise estimates of the size of the random sample that ensures that at least one point lies in each attraction basin.

A practical methodology is then proposed for identifying these quantities (N and (α_j) distributions) for an unknown landscape, given a random sample of starting points and a local steepest ascent search. This methodology can be applied to any landscape specified with a modification operator and provides bounds on search complexity to detect all local optima. Experiments demonstrate the efficiency of this methodology for guiding the choice of modification operators, eventually leading to the design of problem-dependent optimization heuristics.

Key words. Combinatorial complexity, local search, neighborhood graph, randomized starting solution

AMS subject classifications. 68R99, 90C10, 60F99

1. Introduction. In the field of stochastic optimization, two search techniques have been widely investigated during the last decade: Simulated Annealing [25] and Evolutionary Algorithms (EAs) [6, 7]. These algorithms are now widely recognized as methods of order zero for function optimization as they impose no condition on function regularity. However, the efficiency of these search algorithms, in terms of the time they require to reach the solution, is strongly dependent on the choice of the modification operators used to explore the landscape. These operators in turn determine the neighborhood relation of the landscape under optimization.

This paper provides a new methodology allowing to estimate the number and the sizes of the attraction basins of a landscape specified in relation to some modification operator. This allows one to derive bounds on the probability that one samples a point in the basin of the global optimum for example. Further, this method could be used for guiding the choice of efficient problem-dependent modification operators or representations.

Formally, a landscape can be denoted by $\mathcal{L} = (f, \mu, E)$ where f is the function to optimize and μ the modification operator that is applied to elements of the search space E. The structure of the landscape, heavily depends on the choice of the modification 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 a 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 [13, 21, 22, 26]. It is shown 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 result has been applied to

*Centre de Mathématiques Appliquées, Ecole Polytechnique, 91128 Palaiseau Cedex, France

†Corresponding author, Tel: (33).1.69.33.46.30 Fax: (33).1.69.33.30.11, E-mail: Josselin.Garnier@polytechnique.fr

solve a difficult NP-complete problem [20] (the identification of minimal finite k-state automaton for a given input-output behavior), using evolutionary algorithms. 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 [16, 19]. 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 extracting a priori statistical information about landscape difficulty, for example based on random sampling of the search space. We cite from the field of evolutionary algorithms: Fitness Distance relations, first proposed in [8] and successfully used to choose problem dependent random initialization procedures [11, 14]; Fitness Improvement of evolution operators, first proposed in [5], then extended and successfully used to choose binary crossover operators [12] and representations [9]. 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 (leading to different EA behaviors) can share the same fitness distance relation [18, 10]. 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 [12] 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 with respect 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 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. 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 corresponding local maximum:

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

The basic problem consists in detecting all local maxima m_j. This is equivalent to finding 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 4): One puts M points randomly in the search space. The question is the following: Given the statistical distribution of the relative sizes of

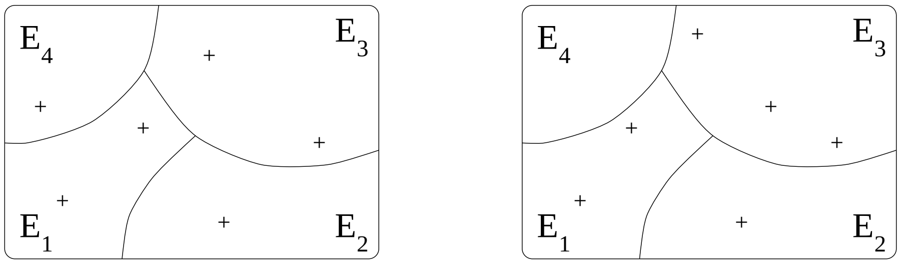

FIG. 1. *Schematic representations of the search space E with $N = 4$ attraction basins. $M = 6$ points have been randomly placed on both pictures. As a result there is at least one point in each attraction basin in the left picture, but not in the right picture, where E_4 is empty.*

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 5): 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 4. 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 corresponding 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 4 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 Section 5 is devoted to the estimation of 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 5.2 starts by considering a parametrized family of laws for the relative sizes of basins, for which it derives the corresponding covering of the search space (law of (β_j)). Then Section 5.3 comments on how these results can be practically used for characterizing the sizes of basins 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 5.2. Finally, the last part of the paper (Section 6) is devoted to some experiments that validate (Section 6.1) and illustrate (Section 6.2) 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 problems, known to be difficult to optimize for EAs. For each problem, we also compare the landscapes related to different mutation operators.

2. 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 $\mathcal{L}$, and by (α_j) the random variables describing the sizes of the attraction basins of $\mathcal{L}$ (normalized to the average size). As shown in [23, 24], a local improvement algorithm is efficient to find quickly a local optimum starting from some given point. Among the possible algorithms we present the Steepest Ascent (SA) also called optimal adjacency algorithm in [23]:

STEEPEST ASCENT ALGORITHM (SA).

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

Algorithm: Modify X by repeatedly performing the following steps:

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

Output: The point X, denoted by $\mu^{SA}(X)$.

The SA algorithm thus consists in selecting the best neighbors after the entire neighborhood is examined. An alternative algorithm, the so-called First Improvement (FI), consists in accepting the first favorable neighbor as soon as it is found, without further searching. Note that in the FI case there are extra free parameters which are the order in which the neighborhood is searched. As pointed out in [15, p. 470], the steepest ascent is often not worth the extra computation time, although it is sometimes much quicker. Nevertheless our focus in this paper is not a complete optimization of the computational time, so we let this problem as an open question.

DEFINITION 2.1. *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 from 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. 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.

3. Summary of the results. Given a distribution of (α_j), we 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

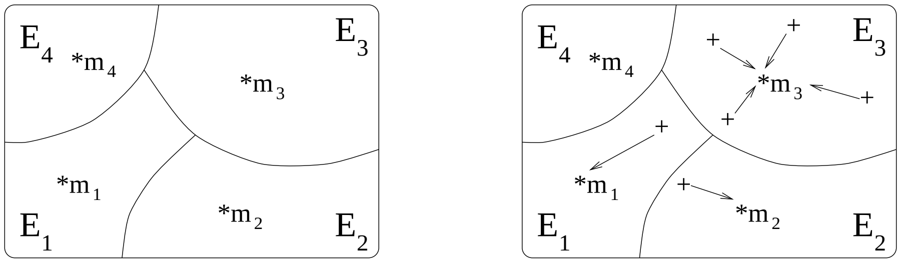

FIG. 2. *Schematic representation of the search space E with its $N = 4$ attraction basins and the 4 corresponding local maxima m_1,...,m_4. In the left picture we have put $M = 6$ points randomly chosen. We apply the search algorithm and detect 3 maxima according to the right picture, so that we have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

the deterministic configuration (resp. $M_{min} = aN^2$ for the random configuration), ensures that a point is sampled in each attraction basin with probability $\exp(-1/a)$.

We then address the inverse problem of identifying the distribution of the normalized sizes $(\alpha_j)_{j=1,\ldots,N}$ of the attraction basins, for an unknown landscape. Some direct analysis is first required as discussed below.

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

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

Proposition 5.1 gives the distribution of (β_j) for a family of parametrized distributions Law_γ for (α_j) asymptotically with respect to N and M. More precisely, if $(Z_j)_{j=1,\ldots,N}$ denotes a family of positive real-valued independent random variables with Gamma distributions whose densities are:

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

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

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

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.

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) if and only if 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, Eq. (1) is used to find a good estimator of N.
Last, Section 6 validates the methodology of Section 5, by considering known landscapes with random and deterministic sizes of basins, 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.

4. 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.

PROPOSITION 4.1. *If we denote by $\alpha_j := |E_j|/|E|$ the normalized size of the j-th attraction basin, then:*

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

Proof. Let us denote by A_j the event:

$$A_j = \{ \text{ there is no point in } E_j \}$$

The probability of the intersection of a collection of events A_j is easy to compute. For any $1 \leq i_1 < ... < i_k \leq N$, if there is one initial point chosen uniformly in E ($M = 1$), then we have

$$\mathbb{P}\left(A_{i_1} \cap ... \cap A_{i_k}\right) = 1 - \alpha_{i_1} - ... - \alpha_{i_k}$$

If there are M initial points chosen uniformly and independently in E, then:

$$\mathbb{P}\left(A_{i_1} \cap ... \cap A_{i_k}\right) = \left(1 - \alpha_{i_1} - ... - \alpha_{i_k}\right)^M$$

On the other hand, $1 - p_{N,M}$ is the probability that at least one of the attraction basin contains no point, which reads as:

$$p_{N,M} = 1 - \mathbb{P}\left(A_1 \cup ... \cup A_N\right)$$

The result thus follows from the inclusion-exclusion formula [28, Formula 1.4.1a]. ☐

Proposition 4.1 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 4.2. *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 $M = [N(\ln N + \ln a)]^1$, $a > 0$, then:

$$p_{N,M} \stackrel{N\to\infty}{\longrightarrow} \exp\left(-a^{-1}\right).$$

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

$$M_D = N\ln N - N\ln Z + o(N),$$

where Z is an exponential variable with mean 1.
An exponential variable with mean 1 is a random variable whose density with respect to the Lebesgue measure over $\mathbb{R}^+$ is $p(z) = \exp(-z)$.

Proof. The first point is a straightforward application of Proposition 4.1. It is actually referenced in the literature as the "coupon-collector" problem. The fact that $M_D/(N\ln N)$ converges in probability to 1 is also well-known. The corollary is going one step further by exhibiting the statistical distribution of $M_d - N\ln N$. Let us assume that $M = [N(\ln N + \ln a)]$ with $a > 0$. We begin by establishing an estimate of $p_{N,M,k} = C_N^k(1-k/N)^M$. First note $C_N^k = N\times(N-1)\times...\times(N-k+1)/k! \leq N^k/k!$. Second $\ln(1-x) \leq -x$ for any $0 \leq x \leq 1$ so that

$$M\ln\left(1-\frac{k}{N}\right) \leq (N\ln N + N\ln a - 1)\ln\left(1-\frac{k}{N}\right) \leq -k\ln N - k\ln a + \frac{k}{N}.$$

As a consequence, uniformly with respect to $k = 0, ..., N$:

$$\left(1-\frac{k}{N}\right)^M \leq N^{-k}a^{-k}e.$$

We thus have

$$p_{N,M,k} \leq ea^{-k}/k! \text{ for all } k \text{ uniformly with respect to } N. \tag{3}$$

Choosing some $K \geq 1$, we can write from Eq. (3):

$$\left|p_{N,M} - \sum_{k=0}^{K}(-1)^k p_{N,M,k}\right| \leq e\sum_{k=K+1}^{\infty}\frac{a^{-k}}{k!}. \tag{4}$$

It is easy to check that, for any fixed k: $N^{-k}\times C_N^k \to 1/k!$ and $N^k\times(1-k/N)^M \to a^{-k}$ as $N\to\infty$, so that:

$$\limsup_{N\to\infty}\left|p_{N,M} - \sum_{k=0}^{K}\frac{(-1)^k a^{-k}}{k!}\right| \leq e\sum_{k=K+1}^{\infty}\frac{a^{-k}}{k!}. \tag{5}$$

This holds true for any K, so that we take the limit $K\to\infty$ which gives the result of the second point. The third point then follows readily from the identity $\mathbb{P}(M_d \geq m) = 1 - p_{N,m-1}$. □

[1] $[x]$ stands for the integral part of the real number x.

COROLLARY 4.3. *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:*

$$p_{N,M} \stackrel{N\to\infty}{\longrightarrow} \exp\left(-a^{-1}\right).$$

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*

$$M_R = N^2 Z^{-1} + o(N^2),$$

where Z *is an exponential variable with mean* 1.

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. Consider the order statistics $(a_{\sigma(i)})_{i=,1,\dots,N-1}$ of this sample, that is to say permute the indices of these points so that $a_{\sigma(0)} := 0 \leq a_{\sigma(1)} \leq \dots \leq a_{\sigma(N-1)} \leq a_{\sigma(N)} := 1$. Denote the spacings by $\alpha_j = a_{\sigma(j)} - a_{\sigma(j-1)}$ for $j = 1, \dots, N$. Note that α_j is also called the j-th coverage. If the j-th attraction basin E_j is the interval $[a_{\sigma(j-1)}, a_{\sigma(j)})$, 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. From Eq. (2) and the relation $\sum_{j=1}^N \alpha_j = 1$ we get that:

$$p_{N,M} = \sum_{k=0}^{\infty} (-1)^k p_{N,M,k}, \tag{6}$$

$$p_{N,M,k} = \sum_{1 \leq j_1 < \dots < j_k \leq N} \mathbb{E}\left[(1 - \alpha_{j_1} - \dots - \alpha_{j_k})^M\right], \tag{7}$$

where $\mathbb{E}$ stands for the expectation with respect to $(\alpha_j)_{j=1,\dots,N}$ whose distribution is uniform over S_N. As pointed out in Ref. [28, Section 9.6a], the probability distribution of the sum of any k of the N coverages α_j is described by the repartition function given by Formula 9.6.1-[28], which shows that it admits a density $q_{N,k}(\alpha)$ with respect to the Lebesgue measure over $[0,1]$:

$$q_{N,k}(\alpha) = \frac{(N-1)!}{(k-1)!(N-k-1)!} \alpha^{k-1} (1-\alpha)^{N-1-k}.$$

We can thus write a closed form expression for $p_{N,M,k}$:

$$\begin{aligned}
p_{N,M,k} &= C_N^k \int_0^1 d\alpha q_{N,k}(\alpha)(1-\alpha)^M \\
&= C_N^k \frac{(N-1)!}{(k-1)!(N-k-1)!} \int_0^1 d\alpha \alpha^{k-1} (1-\alpha)^{M+N-1-k} && (8) \\
&= \frac{1}{k!} \frac{N!}{(N-k)!} \frac{(N-1)!}{(N-1-k)!} \frac{(M+N-1-k)!}{(M+N-1)!}. && (9)
\end{aligned}$$

We shall first prove an estimate of $p_{N,M,k}$.

Step 1. $p_{N,M,k} \leq \frac{1}{k!}\left(\frac{N^2}{M}\right)^k$.

We have $N!/(N-k)! \leq N^k$ and $(N-1)!/(N-k-1)! \leq (N-1)^k \leq N^k$. For any $k = 0, ..., N$ we also have $(M+N-1-k)!/(M+N-1)! \leq M^{-k}$. Substituting these inequalities into Eq. (9) establishes the desired estimate.

Step 2. For any fixed k, if $M = [aN^2]$ then $p_{N,M,k} \to a^{-k}/k!$ as $N \to \infty$.
On the one hand, $C_N^k \frac{(N-1)!}{(k-1)!(N-k-1)!} \times N^{-2k} \to 1/(k!(k-1)!)$. On the other hand:

$$N^{2k} \times \int_0^1 d\alpha \alpha^{k-1}(1-\alpha)^{[aN^2]+N-1-k} = \int_0^1 ds s^{k-1}\left(1 - \frac{s}{N^2}\right)^{[aN^2]+N-1-k}.$$

Since $\left(1 - \frac{s}{N^2}\right)^{[aN^2]+N-1-k}$ is bounded by 1 and converges to $\exp(-as)$ as $N \to \infty$, the dominated convergence theorem implies that:

$$N^{2k} \times \int_0^1 d\alpha \alpha^{k-1}(1-\alpha)^{[aN^2]+N-1-k} \overset{N\to\infty}{\longrightarrow} (k-1)!a^{-k},$$

which yields the result by Eq. (8).

Step 3. Convergence of $p_{N,M}$ when $M = [aN^2]$.
We first choose some $K \geq 1$. We have from the result of Step 1:

$$\left| p_{N,M} - \sum_{k=0}^{K} (-1)^k p_{N,M,k} \right| \leq \sum_{k=K+1}^{\infty} \frac{1}{k!}\left(\frac{N^2}{[aN^2]}\right)^k. \tag{10}$$

Substituting the result of Step 2 into Eq. (10) shows that:

$$\limsup_{N\to\infty} \left| p_{N,M} - \sum_{k=0}^{K} \frac{(-1)^k a^{-k}}{k!} \right| \leq \sum_{k=K+1}^{\infty} \frac{a^{-k}}{k!}.$$

This inequality holds true for any K, so letting $K \to \infty$ completes the proof of the corollary. □

It follows from the corollaries that about $N \ln N$ points are needed in the D-configuration to detect all maxima, while about N^2 points are needed 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. Actually it can be proved that the smallest attraction basin in the R-configuration has a relative size which obeys an exponential distribution with mean N^{-2} (for more detail about the asymptotic distribution concerning order statistics we refer to [28]). That is why a number of points of the order of N^2 is required to detect this very small basin.

Mean values. The expected value of M_D is:

$$\mathbb{E}[M_D] = N \ln N + NC + o(N),$$

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:

$$\mathbb{P}(M_R \geq N^2 a) \overset{N\to\infty}{\longrightarrow} 1 - \exp(-a^{-1}) \underset{a \gg 1}{\simeq} a^{-1}.$$

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:

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

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.

5. Inverse problem.

5.1. 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 corresponding 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 ($j \geq 1$):

$$\beta_j := \text{number of maxima detected with } j \text{ points.} \tag{11}$$

Our arguments are based upon the following observations. First note 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, we should be able to find a better estimate of N than $\bar{N}$.

5.2. 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 $(\alpha_j)_{j=1,\dots,N}$ of the attraction basins can be described by a distribution parametrized by some positive number γ as follows. Let $(Z_j)_{j=1,\dots,N}$ be a sequence of independent random variables whose common distribution has density p_γ with respect to the Lebesgue measure over $(0, \infty)$:[2]

$$p_\gamma(z) = \frac{\gamma^\gamma}{\Gamma(\gamma)} z^{\gamma-1} e^{-\gamma z}, \tag{12}$$

where Γ is the Euler's Gamma function $\Gamma(s) := \int_0^\infty e^{-t} t^{s-1} dt$. The density p_γ is the so-called Gamma density with parameters (γ, γ) [4, p. 47, Formula 2.2]. Under p_γ,

[2]If γ is a positive integer then p_γ is a negative-binomial distribution.

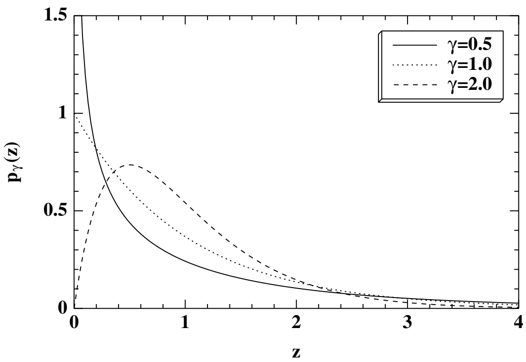

FIG. 3. *Probability density of the sizes of the attraction basins under H^γ for different γ.*

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 $(\alpha_j)_{j=1,\ldots,N}$ 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. Note that the large deviations principle (Cramer's theorem [1, Chapter 1]) applied to the sequence (Z_j) yields that for any $x > 0$ there exists $c_{\gamma,x} > 0$[3] such that:

$$\mathbb{P}_\gamma\left(\left|\frac{T_N}{N} - 1\right| \geq x\right) \leq \exp(-Nc_{\gamma,x}), \tag{13}$$

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 in Section 4 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 [17].

The important statement is the following one.

PROPOSITION 5.1. *Under H^γ the expected values $\beta_{j,\gamma} := \mathbb{E}_\gamma[\beta_j]$ of the β_j's can be computed for any N, M, and $j \leq M$:*

$$\beta_{j,\gamma} = N\frac{\Gamma(\gamma+j)}{j!\Gamma(\gamma)}\frac{\Gamma(N\gamma)}{\Gamma((N-1)\gamma)}\frac{\Gamma((N-1)\gamma+M-j)}{\Gamma(N\gamma+M)}\frac{M!}{(M-j)!}. \tag{14}$$

In the asymptotic framework $N \gg 1$, if $M = [aN]$ then $\beta_{j,\gamma}$ can be expanded as:

$$\beta_{j,\gamma} = N\frac{\Gamma(j+\gamma)}{j!\Gamma(\gamma)}\frac{a^j\gamma^\gamma}{(a+\gamma)^{j+\gamma}} + o(N). \tag{15}$$

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

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

[3]Applying the procedure described in [1] establishes that $c_{\gamma,x} = \gamma(x - 1 - \ln x)$.

where $\alpha_k = Z_k / \sum_i Z_i$ and $\mathbb{E}_\gamma$ stands for the expectation of Z_j with distribution p_γ. Accordingly, in terms of the Z_i's this expression reads:

$$\mathbb{P}_\gamma\left(j \text{ points in } E_k\right) = C_M^j \mathbb{E}_\gamma \left[\frac{Z_k^j \check{Z}_k^{M-j}}{(Z_k + \check{Z}_k)^M} \right],$$

where $\check{Z}_k = \sum_{i \neq k} Z_i$. The random variables Z_k and $\check{Z}_k$ are independent. The probability density of Z_k is p_γ given by Eq. (12). The random variable $\check{Z}_k$ is the sum of $N-1$ independent random variables with densities p_γ, so that its probability density is [4, p. 47, Formula 2.3]:

$$\check{p}_\gamma(\check{z}) = \frac{\gamma^{(N-1)\gamma}}{\Gamma((N-1)\gamma)} \check{z}^{(N-1)\gamma-1} e^{-\gamma \check{z}}.$$

Accordingly:

$$\mathbb{P}_\gamma\left(j \text{ points in } E_k\right) = C_M^j \int_0^\infty dz \int_0^\infty d\check{z} p_\gamma(z) \check{p}_\gamma(\check{z}) \frac{z^j \check{z}^{M-j}}{(z+\check{z})^M}.$$

By the change of variables $u = z/(z+\check{z})$ and $v = z + \check{z}$ we get:

$$\mathbb{P}_\gamma\left(j \text{ points in } E_k\right) = C_M^j \int_0^1 du \int_0^\infty dv v^{N\gamma-1} e^{-\gamma v} u^{\gamma+j-1} (1-u)^{(N-1)\gamma+M-j-1}.$$

The integral with respect to v is straightforward by definition of the Gamma function. The integral with respect to u can be obtained via tabulated formulae [4, p. 47, formula 2.5]. This gives the explicit formula (14) for $\beta_{j,\gamma}$ since

$$\mathbb{E}_\gamma\left[\beta_j\right] = \sum_{k=1}^N \mathbb{P}_\gamma\left(j \text{ points in } E_k\right).$$

If $N \gg 1$ and $M = [aN]$, then we have $N^{-\gamma} \times \Gamma(N\gamma)/\Gamma((N-1)\gamma) \to \gamma^\gamma$, $N^{\gamma+j} \times \Gamma((N-1)\gamma + M - j)/\Gamma(N\gamma + M) \to 1/(\gamma+a)^{j+\gamma}$, and $N^{-j} \times M!/(M-j)! \to a^j$ as $N \to \infty$. This proves the asymptotic formula (15). $\square$

In particular, the distribution of the β_j's under the D-configuration is Poisson in the asymptotic framework $N \gg 1$:

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

while it is geometric under the R-configuration:

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

From Eq. (15) we can deduce that the following relation is satisfied by the ratio $r = M/N$:

$$\frac{\sum_{j=1}^\infty \beta_{j,\gamma}}{M} = \frac{1-(1+\frac{r}{\gamma})^{-\gamma}}{r}. \tag{16}$$

5.3. Estimator of the number of local maxima. We have now sufficient tools to exhibit a good estimator of the number of local maxima. We remind the reader of the problem at hand. We assume that some algorithm is available to determine from any given point the corresponding local maximum. We choose randomly M points in the search space and detect the corresponding local maxima. We thus obtain a set of values $(\beta_j)_{j\geq 1}$ as defined by (11). 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 statistics used to compare observed and expected results is the so-called χ^2 goodness of fit test [27, 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 $\Omega := \{j \in \mathbb{N}, \beta_j \geq 1\}$ is the set of the indices j for which $\beta_j \geq 1$. 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 statistics. A general result states 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:

$$\gamma_0 = \text{argmin}\,\{T_\gamma, \gamma > 0\}\,. \tag{17}$$

Furthermore, we can estimate the accuracy of the configuration H^{γ_0} by referring T_{γ_0} to tables of the χ^2-distribution for $\text{card}(\Omega) - 1$ 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^{γ_0} is the closest distribution of the real one.

REMARK. The distribution theory of χ^2 goodness of fit statistic can be found in [3, 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 power in the tail regions is then lost, where differences are more likely to show up.

Defining γ_0 as (17), we denote by $\bar{\beta}$ the quantity:

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

From Eq. (16), under H^{γ_0} the ratio $\alpha = M/N$ is the unique solution of:

$$\bar{\beta} = \frac{1 - (1 + \frac{r}{\gamma_0})^{-\gamma_0}}{r}. \tag{18}$$

Consequently, once we have determined γ_0, formula (18) is a good estimator of the ratio $\alpha = M/N$, hence N.

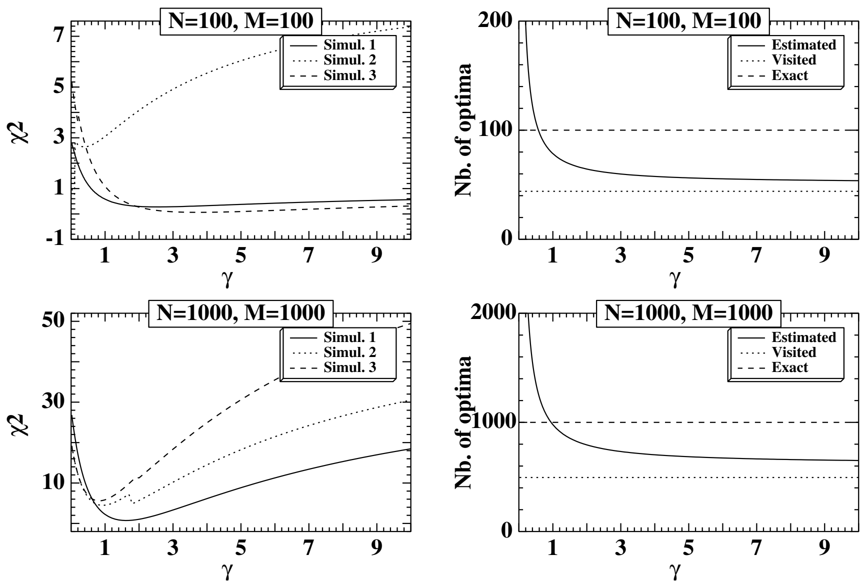

FIG. 4. *Basins with random uniform sizes: The left figures plot the χ^2 test results (i.e. the values of T_γ) comparing the empirically observed β distribution to the family of γ-parametrized distributions. The right figures plot for different γ values the estimation of the number of local optima computed by Eq. (18). These estimations are very robust (only one estimation is plotted) and are accurate for $\gamma_0 = 1$. The same figures also show the visited numbers of optima actually visited by the steepest ascent ($\bar{N} = \sum_{j=1}^{\infty} \beta_j$). The numerical simulations exhibit unstable results for the χ^2 test for small N values and $M = N$.*

6. Experiments. Given a landscape $\mathcal{L}$, the following steps are performed in order to identify a possible law for the number and sizes of the attraction basins of $\mathcal{L}$, 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 from each X_i up to $\mu^{SA}(X_i)$.*
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 %.

6.1. Experimental validation. The results obtained in Section 5 are asymptotic with respect to the number of local optima N and 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 sizes 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 4-5 and 6. Samples with smaller sizes than those shown in these figures yield β_j values which are not rich enough to allow a significant

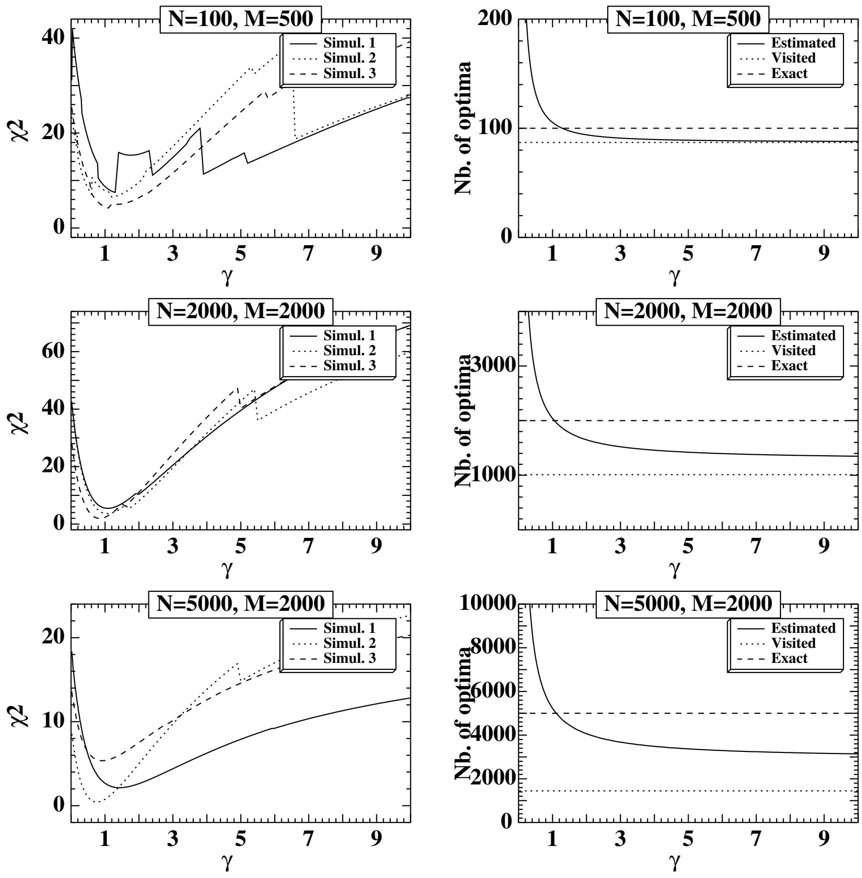

Fig. 5. *The same as in Figure 4 with different values for N and M. Stable results are obtained when N increases and M is bounded ($M \leq min(2000, 3N)$ here). The estimation of N corresponding to the smallest χ^2 value ($\gamma = 1$) is very accurate.*

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). 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: The numerical simulations exhibit unstable results for the χ^2 test for $M = N$ and small N values (Figures 4). When N increases and M is bounded ($M \leq min(2000, 3N)$ in the experiments), results become stable and accurate (Figures 5). 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 6). This situation is even more striking in the experiments of the following section on Baluja F1 problems.

6.2. 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 [2], for a 1-bit-flip and 3-bit-flips neighborhood relations.

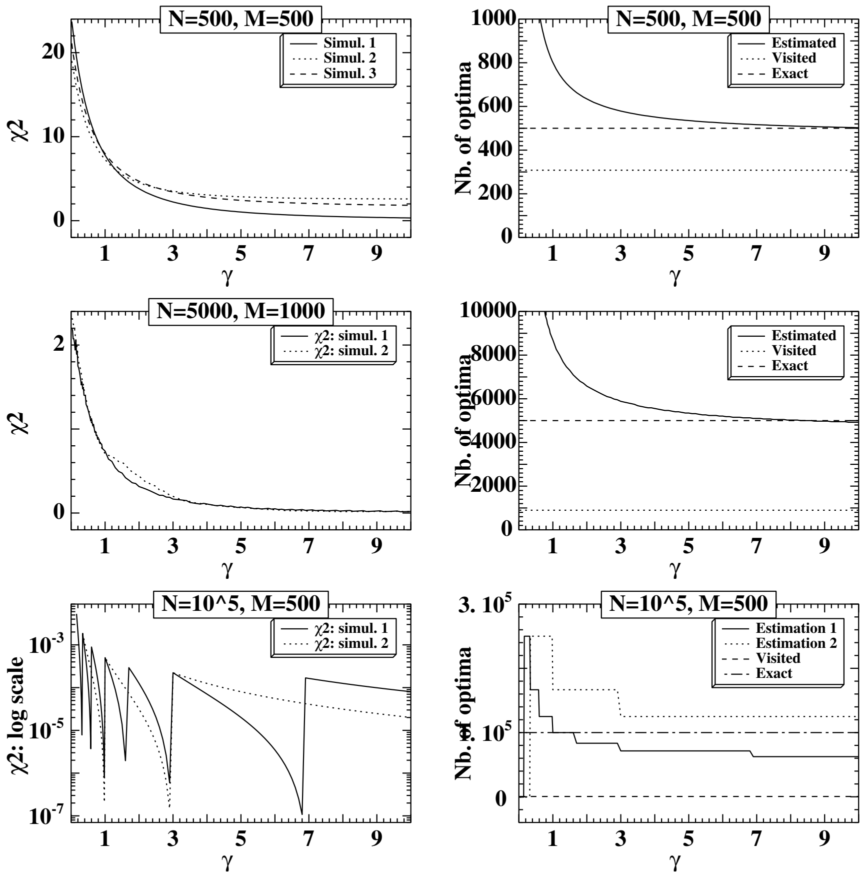

FIG. 6. *Basins with Deterministic equal sizes: The χ^2 results are stable for smaller sample sizes than those of the random configuration. The bottom figures correspond to the case $N = 10^5$ and $M = 500$, where the χ^2 test is not significant, yet the predicted number of local optima is very accurate! With 500 initial points, 497 local optima have been visited, while there are actually 10^5 optima. Yet, formula (18) is able to estimate the true number with an error of less than 30% when the adequate γ value is used.*

Gray-coded Baluja F1 functions. Consider the function of k variables $(x_0, \ldots, x_{k-1})$, with $x_i \in [-2.56, 2.56]$ [2]:

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

It reaches its maximum value among 10^7 at point $(0, \ldots, 0)$. The Gray-encoded F1g and binary F1b versions, with respectively 2, 3 and 4 variables encoded on 9 bits are considered. This encoding consequently corresponds to the binary search space with $l = 9k$.

Considering the 1-bit-flip mutation (Hamming landscape), Figure 7 shows that 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 8), the estimated number of local optima drops significantly for both problems: less than 250 for both binary and gray landscapes, whereas the Hamming landscape

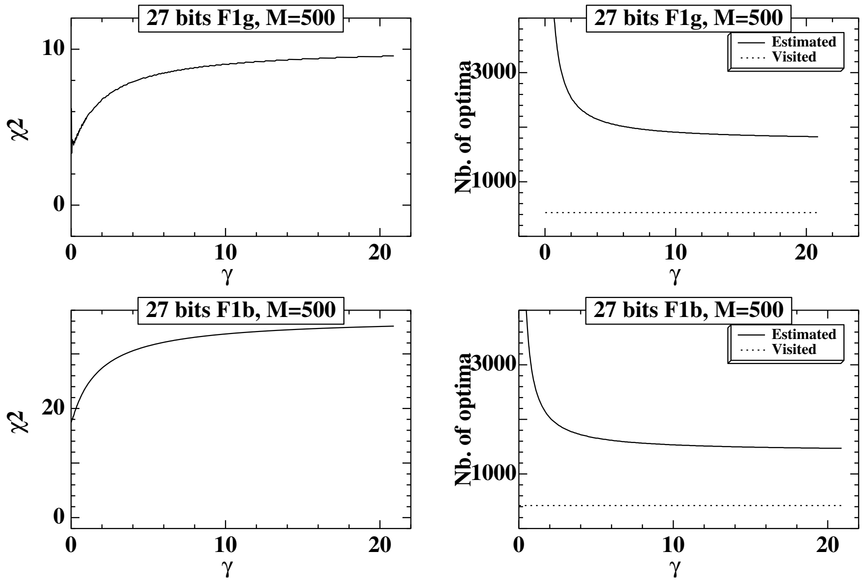

FIG. 7. *The difficult Baluja 27-bits F1 gray (F1g) and binary (F1b) landscapes with a 1-bit-flips mutation. Experiments with samples of sizes* $M = 2000$ *and* $M = 5000$ *show the same results for the* χ^2 *test, and the corresponding estimations of the number of local maxima converge to a stable value around* 4000.

contains thousands of local optima (Figure 7).

Experiments at problem sizes $l = 18$ and $l = 36$ have been carried out in addition to the plotted ones ($l = 27$), 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 $\ell = 18$, there are less than 10 local optima in the 3-bit-flips landscape versus hundreds in the Hamming landscape. When $\ell = 36$, the estimations for the Hamming landscape show about two times more local optima for the gray than for the binary encoding (resp. 45 000 and 25 000). Still for $\ell = 36$, but for the 3-bit-flips landscape, the estimated number of local optima drops respectively to 1400 and 1000.

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 6, generate a random sample whose 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 sizes of the basins are close to random). Then a simple steepest ascent starting from each point of the sample, ensures that the global optimum is found with probability $\exp(-1/a)$.

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 4.3 and the estimation of $N = 250$ in the experiments of Figure 8).

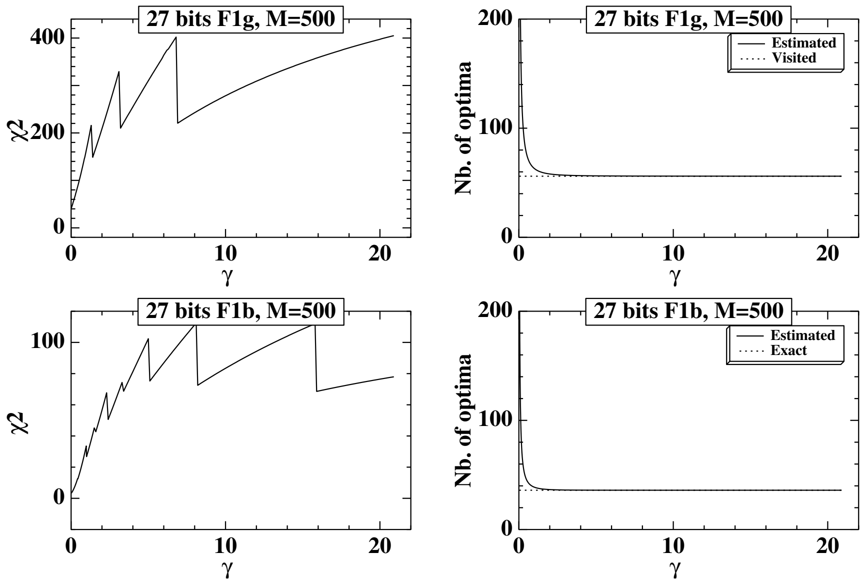

FIG. 8. *The difficult Baluja 27-bits F1 gray (F1g) and binary (F1b) landscapes with a 3-bit-flips mutation: the number of local optima drops significantly compared to the Hamming 1-bit-flip landscape. These results are confirmed by experiments using samples of sizes* $M = 2000$ *and* $M = 5000$ *which give the same estimation for the number of local optima.*

6.3. Discussion. This paper provides a new methodology allowing to estimate the number and the sizes of the attraction basins of a landscape specified in relation to some modification operator. This allows one to derive bounds on the probability that one samples a point in the basin of the global optimum for example. Further, it allows to compare landscapes related to different modification operators or representations, as illustrated with the Baluja problem.

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 parameters and consequently to decide which distribution is the closest to 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 (12) 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).

However, the experiments with the Baluja problem, appeal for refining the class of laws of (α_j) around basins with random sizes. We may propose $\alpha_j = Z_j / \sum_{i=1}^{N} Z_i$ where Z_j are independent and identically distributed with one of the distributions of the bidimensional family $p_{\gamma,\delta}(.)$, $\gamma > 0$, $\delta > 0$:

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

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 $e^{-\gamma z}$. The density $p_{\gamma,\delta}$ is the so-called Gamma density with parameters (γ, δ) [4, p. 47, Formula 2.2]. This family presents more diversity than the family $p_\gamma(.)$ we have considered in Section 5.2. The expected value of β_j is under $p_{\gamma,\delta}$:

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

The method of estimating the number of local minima described in Section 5.3 could then be applied with this family.

To apply our method we have also made a crucial choice which consists in executing the local search algorithm from randomly distributed points. We do so because we have no a priori information on the landscape at hand. However, assume for a while that we have some a priori information about the fitness function, for instance its average value. Consequently we could hope that starting with points whose fitnesses are better than average would improve the detection of the local maxima. Nevertheless extensive computer investigations of some particular problems have shown that this is not the case [15, p. 456], possibly because a completely random sampling of starting points allows one to get a wider sample of local optima.

A first application of the methodology presented in this paper 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 corresponding 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) obey a particular law, hence guiding a hierarchical search in different subparts of the space.

Note finally that the distributions of the sizes of basins do not fully characterize landscape difficulty. Depending on the relative position of the attraction basins, the search still may range from easy to difficult. 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

[1] R. Azencott, *Grandes déviations et applications*, in Proceedings of Ecole d'Eté de Probabilités de Saint-Flour, P. L. Hennequin, ed., Springer-Verlag, 1978, pp. 1-176.

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

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

[4] W. Feller, *An introduction to probability theory and its applications*, Wiley, New York, 1971.

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

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

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

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

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

[10] L. Kallel, B. Naudts, and M. Schoenauer, *On functions with a fixed fitness distance relation*, Congress on Evolutionary Computation, IEEE, in press.

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

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

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

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

[15] C. H. Papadimitriou and K. Steiglitz, *Combinatorial optimization: Algorithms and complexity*, Prentice-Hall, Inc., Englewoods Cliffs, NJ, 1982.

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

[17] R. Pyke, *Spacings*, J. Roy. Statis. Soc., **27** (1965), pp. 395-436.

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

[19] R. Rammal, G. Toulouse, and M. A. Virasoro, *Ultrametricity for Physicists*, Reviews of Modern Physics **58** (1986), pp. 765-788.

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

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

[22] P. Stadler, *Landscapes and their correlation functions*, Journal of Mathematical chemistry **20** (1996), pp. 1-45.

[23] C. A. Tovey, *Local improvement on discrete structures*, in Local search and combinatorial optimization, E. Aaarts and J. K. Lenstra, eds., Wiley, New York, 1997.

[24] C. A. Tovey, *Hill climbing with multiple local optima*, SIAM J. Alg. Disc. Math. **6** (1985), pp. 384-393.

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

[26] E. D. Weinberger, *Correlated and uncorrelated fitness landscapes and how to tell the difference*, Biological Cybernetics **63** (1990), pp. 325-336.

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

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