



**HAL**  
open science

# On Stochastic Fitness Landscapes: Local Optimality and Fitness Landscape Analysis for Stochastic Search Operators

Brahim Aboutaib, Sébastien Verel, Cyril Fonlupt, Bilel Derbel, Arnaud Liefoghe, Belaïd Ahiod

## ► To cite this version:

Brahim Aboutaib, Sébastien Verel, Cyril Fonlupt, Bilel Derbel, Arnaud Liefoghe, et al.. On Stochastic Fitness Landscapes: Local Optimality and Fitness Landscape Analysis for Stochastic Search Operators. PPSN 2020 - The 16th International Conference on Parallel Problem Solving from Nature, Sep 2020, Leiden, Netherlands. pp.97-110, 10.1007/978-3-030-58115-2\_7 . hal-02949069

**HAL Id: hal-02949069**

**<https://hal.science/hal-02949069>**

Submitted on 9 Sep 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# On Stochastic Fitness Landscapes: Local Optimality and Fitness Landscape Analysis for Stochastic Search Operators

Brahim Aboutaib<sup>1,2</sup>, Sébastien Verel<sup>1</sup>, Cyril Fonlupt<sup>1</sup>,  
Bilel Derbel<sup>3</sup>, Arnaud Liefoghe<sup>4</sup>, and Belad Ahiod<sup>2</sup>

<sup>1</sup> Université du Littoral Côte d’Opale, LISIC F-62100 Calais, France,  
[brahim.aboutaib@univ-littoral.fr](mailto:brahim.aboutaib@univ-littoral.fr)

<sup>2</sup> Mohammed V University in Rabat, Faculty of Science, LRIT, Morocco

<sup>3</sup> Univ. Lille, CNRS, Centrale, Inria, UMR 9189 - CRISTAL, F-59000 Lille, France

<sup>4</sup> JFLI – CNRS IRL 3527, University of Tokyo, 113-0033 Tokyo, Japan

**Abstract.** Fitness landscape analysis is a well-established tool for gaining insights about optimization problems and informing about the behavior of local and evolutionary search algorithms. In the conventional definition of a fitness landscape, the neighborhood of a given solution is a set containing nearby solutions whose distance is below a threshold, or that are reachable using a deterministic local search operator. In this paper, we generalize this definition in order to analyze the induced fitness landscape for *stochastic* search operators, that is when neighboring solutions are reachable under different probabilities. More particularly, we give the definition of a *stochastic* local optimum under this setting, in terms of a probability to reach strictly improving solutions. We illustrate the relevance of stochastic fitness landscapes for enumerable combinatorial benchmark problems, and we empirically analyze their properties for different stochastic operators, neighborhood sample sizes, and local optimality thresholds. We also portray their differences through stochastic local optima networks, intending to gather a better understanding of fitness landscapes under stochastic search operators.

**Keywords:** Combinatorial optimization · Local optimality · Fitness landscape · Stochastic search operators.

## 1 Introduction

Originally coming from evolutionary biology [23], the fitness landscape is one of the most common abstractions used to depict and analyze dynamical systems. In evolutionary computation and related stochastic optimization algorithms, a fitness landscape is the association of a search space of potential solutions, a fitness function to be optimized, and a neighborhood relation between solutions on which the optimization process is expected to move during the search process. The aim of fitness landscape analysis is twofold. The first one is to understand the relation between the geometry of the optimization problem and the search

dynamics, using the pictures of peaks, valleys, or plateaus. More recently, with the renew of machine learning techniques, the second goal uses features computed from the fitness landscape to predict algorithm performance or to select relevant algorithm components according to the optimization problem [10]. From the large number of features designed for evolutionary computation, one of the most intuitive and most important one relates to *local optimality*. A local optimum, which can be depicted as a peak of the landscape, is a solution with the best fitness value locally, among its set of neighboring solutions. This definition arises from the neighborhood relation of the fitness landscape. In combinatorial optimization, the neighborhood is a finite set, most often defined by a natural distance between solutions such as Hamming distance or Kendall distance.

However, most local search and evolutionary algorithms use stochastic local search operators, which do not always directly match with a finite set of neighboring solutions. For example, the bit-flip mutation operator conventionally used in genetic algorithms, which flips each bit independently with a given rate, does not produce a finite set of neighboring solutions with an equal probability of being reached. Similarly, hyperheuristics might combine several local search operators with different probability distributions [16]. The explored solutions at each iteration is not properly caught by any finite set. Even when the neighborhood is defined as a finite set, its cardinality might be too large for being computationally enumerated, as it is the case, for instance, in genetic programming or population-based multi-objective search [20]. Therefore, a subpart of the neighborhood is typically sampled at random or in a heuristic way, making the deterministic definition of neighborhood and local optimum less relevant.

In this paper, we extend the definition of neighborhood to stochastic local search operators, and we investigate *stochastic local optimality* in this context. The principle is to define a stochastic local optimum when the probability to strictly improve a solution by applying a stochastic operator is small, below a given threshold. The threshold can be related to the inverse of the computational budget available to find an improving solution. Intuitively, a stochastic local optimum is a solution that is difficult to improve in a reasonable computational effort using the stochastic operator, and therefore constitutes an attraction point for evolutionary or local search. By extending the fitness landscape paradigm, and in particular its neighborhood relation, to stochastic search operators, we expect the definition of local optimality to reveal new insights into the search space structure, and to allow for a better analysis of the design of algorithms based on stochastic operators. In this work, the methodology is to support the relevance of the new definition of stochastic local optimum with an experimental analysis on enumerable combinatorial optimization problems with different properties, and to show the potential additional benefit of this approach for fitness landscape analysis.

Our contributions can be summarized as follows:

- (1) We extend the definition of fitness landscape to stochastic operators, and we propose a new definition of local optimum for stochastic local search.

- (2) We empirically show the relevance of this definition on enumerable NK-landscapes.
- (3) We show some potential interests of stochastic fitness landscapes for iterated local search, and we reveal, for the first time, the structure of local optima networks for stochastic operators.

*Outline.* The paper is organized as follows. In Section 2, we first recall necessary definitions and related works. In Section 3, we present the stochastic fitness landscape and stochastic local optimality definitions. In Section 4, we detail the experimental setup, and the results of the analysis. The discussion and conclusions close the paper in Section 5.

## 2 Preliminaries

A *fitness landscape* [14] is defined as a triplet  $(\mathcal{X}, f, \mathcal{N})$  such that  $\mathcal{X}$  is the search space of candidate solutions,  $f : \mathcal{X} \rightarrow \mathbb{R}$  is the fitness function, and  $\mathcal{N}$  is a neighborhood relation. In combinatorial optimization, the search space is a finite set. The neighborhood relation  $\mathcal{N} : \mathcal{X} \rightarrow 2^{\mathcal{X}}$  assigns a set of solutions, called *neighbors*, to any solution from the search space. Although this definition is quite general, the basic idea behind it is to define neighboring solutions in the vicinity of a solution. Standard definitions are based on a distance measure between elements from the search space, such as the Hamming distance:  $\mathcal{N}(x) = \{x' \in \mathcal{X} \mid d(x, x') \leq D\}$ , where  $d$  is a distance function and  $D$  is the radius of the neighborhood. Other standard definitions are based on a deterministic local search operator that performs a move from one solution to another:  $\mathcal{N}(x) = \{x' \in \mathcal{X} \mid \exists \theta \text{ s.t. } op_{\theta}(x) = x'\}$ , where  $op_{\theta}$  is a parametric deterministic operator such as swap or insertion. In both cases, the neighborhood relation allows one to depict the fitness landscape with peaks, plateaus, and valleys, but also to analyze the fitness landscape using tools from graph theory [14, 21].

One of the main intuitive and fundamental feature of a fitness landscape deals with local optimality, where local optima represent the peaks in the pictures collection of the fitness landscape. For a maximization problem, a *local optimum* is a solution  $x \in \mathcal{X}$  such that:

$$\forall x' \in \mathcal{N}(x), f(x') \leq f(x) \tag{1}$$

The number of local optima in the fitness landscape provides a first information about the difficulty of a combinatorial optimization problem, and about the performance of local and evolutionary search algorithms [8]. For large landscapes, different methods allow one to estimate the number of local optima using uniform random sampling, biased random sampling [1, 7], or the length of an adaptive walk before being trapped [11]. In addition to the number of local optima, the size, the distribution and the structure of local optima's basins of attraction is one major feature related to algorithm performance [5, 8], including for problems from machine learning [3]. The basin of attraction of a local optimum  $x^*$  is defined as the set of solutions from which a hill-climbing algorithm  $h$  falls into:

$\mathcal{B}(x^*) = \{y \in \mathcal{X} \mid h(y) = x^*\}$ . Depending on the pivot rule used by the hill-climber, e.g., first or best improvement, the structure of the basins is different, and so is its impact on search performance [2, 15]. Besides local optima and their basins of attraction, a complementary view of fitness landscapes is given by the so-called *local optima network* (LON) [4, 13]. In particular, the LON with escape edges [19] is defined as a weighted directed graph  $(V, E)$ , where vertices  $V$  are local optima, and there is an edge  $(x_i, x_j) \in E$  between local optima  $x_i$  and  $x_j$  if there is a solution  $y \in \mathcal{X}$  such that the distance between  $x_i$  and  $y$  is below a given threshold, and  $y$  belongs to the basin of attraction of  $x_j$ ; i.e.  $(x_i, x_j) \in E$  iff  $\exists y \in \mathcal{X}$  such that  $d(x_i, y) \leq D$  and  $y \in \mathcal{B}(x_j)$ . The weight  $w_{ij}$  of an edge  $(x_i, x_j)$  gives the ratio of such solutions  $y \in \mathcal{X}$  that satisfies the definition above from the set of solutions at distance  $D$ . Several metrics have been proposed to characterize LONs, and have been related to problem difficulty or search performance for both single-objective [18] and multi-objective optimization [6, 11].

However, although many evolutionary and local search algorithms are based on a stochastic operator, all definitions related to local optimality and fitness landscape analysis are based on a deterministic neighborhood relation, considering a set of neighbors that is often finite. We argue that this only partially reflects the properties of stochastic search operators, and we introduce the notion of stochastic local optimality and fitness landscapes in the next section.

### 3 Stochastic Fitness Landscapes and Local Optimality

A local search algorithm is based on a local search operator  $\text{op} : \mathcal{X} \rightarrow \mathcal{X}$  which moves from one solution to another solution. A stochastic local search operator defines a probability distribution over the search space. When the search space is finite, as in discrete or combinatorial optimization, a stochastic operator can be defined by the probabilities of moving from a solution  $x \in \mathcal{X}$  to a solution  $y \in \mathcal{X}$ :  $\mathbb{P}\{\text{op}(x) = y\} = p_{x \rightarrow y}$  such that  $\forall x, y \in \mathcal{X}, p_{x \rightarrow y} \geq 0$ , and  $\forall x \in \mathcal{X}, \sum_{y \in \mathcal{X}} p_{x \rightarrow y} = 1$ . When the search space is infinite, a stochastic operator can be defined with probability density functions over the search space.

A typical example of a stochastic local search operator is the bit-flip mutation operator used in standard genetic algorithms. It flips each bit from a given bitstring of size  $n$  independently at random, with a rate  $p$ . In this case,  $p_{x \rightarrow y} = p^k(1-p)^{n-k}$ , where  $k$  is the Hamming distance between  $x$  and  $y$ . Besides, a stochastic operator can be derived from any neighborhood relation with finite support. For a given neighborhood relation  $\mathcal{N}$  such that,  $\forall x \in \mathcal{X}, |\mathcal{N}(x)| < \infty$ , a stochastic operator  $\text{op}_{\mathcal{N}}$  can be defined using a uniform random distribution over the set of neighbors:  $\forall y \in \mathcal{N}(x), \mathbb{P}\{\text{op}_{\mathcal{N}}(x) = y\} = \frac{1}{|\mathcal{N}(x)|}$ .

It is straightforward to extend the definition of a fitness landscape by replacing the neighborhood relation with a stochastic search operator.

**Definition 1.** *A stochastic fitness landscape (SFL) is a triplet  $(\mathcal{X}, \text{op}, f)$  where  $\mathcal{X}$  is the search space,  $f : \mathcal{X} \rightarrow \mathbb{R}$  the fitness function, and  $\text{op}$  is the stochastic local search operator.*

However, this definition only makes sense if it is possible to define reasonable basic features of such fitness landscapes. As pointed out in Section 2, one of the main features is the concept of local optimality. According to Eq. (1), a local optimum is a solution for which all neighbors have a lower or equal fitness value. Roughly speaking, if we translate this definition in terms of probability, the probability to reach a neighbor with a lower or equal fitness value from a local optimum is very high. However, this probability cannot always be considered as 1. Indeed, for an ergodic operator and a finite search space, there is a non-zero probability to reach the global optimum from any solution from the search space. As such, considering a probability of 1 would end up having the global optimum as the single local optimum. We thus introduce a threshold  $\epsilon > 0$  to define a stochastic local optimum. The probability to reach a neighbor with lower quality is higher than  $(1 - \epsilon)$ :  $\mathbb{P}\{f(\text{op}(x)) \leq f(x)\} \geq (1 - \epsilon)$ , or equivalently  $\mathbb{P}\{f(x) < f(\text{op}(x))\} \leq \epsilon$ .

**Definition 2.** *Given a SFL  $(\mathcal{X}, \text{op}, f)$ , and a real number  $\epsilon \geq 0$ , a solution  $x \in \mathcal{X}$  is a stochastic local optimum (SLO) at a local optimality threshold  $\epsilon$  iff  $\mathbb{P}\{f(x) < f(\text{op}(x))\} \leq \epsilon$ .*

In other words, a solution is a stochastic local optimum at threshold  $\epsilon$  when the probability to reach a strictly improving solution by applying the stochastic operator is below  $\epsilon$ . Notice that the definition is still effective when the probability to obtain the same solution is not null, i.e. when  $\mathbb{P}\{\text{op}(x) = x\} > 0$ .

Interestingly, the definitions of deterministic and stochastic local optimality can be connected. Let us consider a neighborhood relation with finite support; i.e.  $\forall x \in \mathcal{X}, |\mathcal{N}(x)| < \infty$ . A solution  $x \in \mathcal{X}$  is a deterministic local optimum under  $\mathcal{N}$  iff  $x$  is a stochastic local optimum under  $\text{op}_{\mathcal{N}}$  at threshold  $\epsilon < 1/|\mathcal{N}(x)|$ . Indeed, the probability to strictly improve a solution  $x$  is equal to  $n^+ / |\mathcal{N}(x)|$ , where  $n^+$  is the number of strictly improving neighbors. As a consequence, when  $x$  is a deterministic local optimum for  $\mathcal{N}$ , the probability to strictly improve  $x$  with  $\text{op}_{\mathcal{N}}$  is zero. Conversely, when the probability to strictly improve  $x$  is strictly below  $1/|\mathcal{N}(x)|$ , no neighbor from  $\mathcal{N}(x)$  has a higher fitness value.

In the definition of a stochastic local optimum, the threshold  $\epsilon$  is critical, and a relevant value has to be carefully chosen with respect to the considered landscape and search scenario. Intuitively, we can think of the  $\epsilon$ -value as the inverse of the expected computational budget (in terms of fitness evaluations) required to escape from a local optimum with a stochastic operator. For instance, in one scenario, if at a given step of the search process, the remaining computational budget is  $n_{eval}$ , we could define  $\epsilon = \frac{1}{n_{eval}}$ . Alternatively, in an iterated local search scenario switching between a local search operator and a perturbation operator, the threshold could be defined as the inverse of the budget dedicated to each local search run:  $\epsilon = \frac{1}{n_{ls}}$ . At last, in an evolutionary algorithm scenario for which  $\lambda$  candidate solutions are computed at each iteration, as in a  $(\mu + \lambda)$ -evolution strategy, the  $\epsilon$ -value can be set to  $\epsilon = \frac{1}{\lambda}$ . As such, the probability of strictly improving the current population is below the computational budget of one iteration of the algorithm. Furthermore, we argue that a fitness landscape analysis could actually benefit from the use of a broader range of

$\epsilon$ -values. Fitness landscape metrics such as the number of local optima or the size of the basins of attraction can be studied according to the  $\epsilon$ -value in order to show the metrics spectrum, and not only for a given accurate value of  $\epsilon$ . We illustrate this point and analyze the impact of  $\epsilon$  empirically in the next section.

## 4 Experimental Analysis

### 4.1 Experimental Setup

**Problem Testbed.** We consider NK-landscapes as a problem-independent model of combinatorial optimization problems defined on binary strings. NK-landscapes were proposed in [9] for constructing multi-modal fitness landscapes in a tunable way, by adjusting the epistatic (or non-linearity) degree  $K$ . Given a binary string  $x \in \{0, 1\}^N$ , its fitness  $f(x) \in [0, 1)$ , is defined as follows:

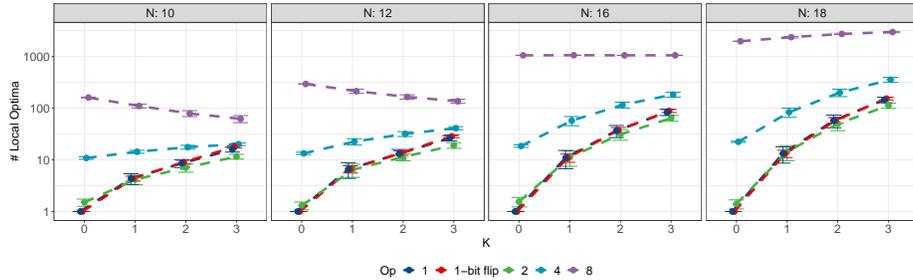
$$f(x) := \frac{1}{N} \sum_{i=1}^N f_i(x_i, x_{i_1}, \dots, x_{i_K}), \text{ where } f_i : \{0, 1\}^{K+1} \rightarrow [0, 1]: \text{ is the epistasis}$$

level of the  $i$ -th bit, its value depend on the allele at the bit  $i$  but also on the other alleles at the  $K$  other epistatic bits, and  $\{i_1, \dots, i_K\} \subset \{1, \dots, i-1, i+1, \dots, N\}$ . For each variable  $x_i$ , there exist two ways for selecting the  $K$  epistatic bits: either randomly, or by choosing the  $K$  closest ones. Beside that NK-landscapes belong to the family of NP-hard problems for  $K > 1$  [22], it has the property of modeling many interesting optimization problems [21]. Thus, it is not just a serious testbed for randomized search heuristics, but also a proxy for other combinatorial optimization problems. We shall mention that the best algorithm to tackle NK-landscapes is not our concern here. We do not consider this problem for benchmarking purposes, but rather for illustration/concept-testing purposes, given that the number of local optima is known to be closely related to the epistatic degree  $K$ .

In order to analyze the impact of different parameters on the definition of SLO, we consider enumerable instances from NK-landscapes with a bitstring length  $N \in \{10, 12, 16, 18\}$ , and an epistatic degree  $K \in \{0, 1, 2, 3\}$ . These  $K$ -values were chosen as they correspond to linear, quadratic, cubic, and quartic versions, the most recurrent problem types encountered in combinatorial optimization. We report the result over 20 different instances for each combination of  $N$  and  $K$ , that we generate with adjacent epistatic interactions.

**Stochastic Operators.** We investigate the bit-flip mutation operator as a stochastic local search operator. It flips each bit independently at random, with a rate  $p$ . We experiment different rates  $p = \frac{c}{N}$ , inversely proportional to the bitstring length  $N$ , such that  $c \in \{1, 2, 4, 8\}$ . For comparison purposes, we also consider the standard 1-bit flip neighborhood operator, which flips 1 bit precisely. Thus, the neighborhood consists of all solutions located at Hamming distance 1, and the neighborhood size is  $N$ .

**Estimating Stochastic Local Optimality.** Even for small enumerable problem instances, it is not computationally doable to enumerate all possible neighbors for each solution to compute exactly all SLO from the search space. Therefore, to estimate the improving probability given in Definition 2, we use the



**Fig. 1.** Number (average value and confidence interval, computed over 20 considered instances) of SLO with respect to the problem non-linearity  $K$ , for different problem sizes  $N$  and stochastic operators  $op$  with a bit-flip mutation rate  $c/N$ .

classical estimator of the empirical mean:  $\hat{p}^+ = \frac{\lambda^+}{\lambda}$  where  $\lambda^+$  is the number of strictly improving solution over a random sample of  $\lambda$  solutions produced by the stochastic operator. Thus, a solution is depicted as a SLO when this estimation  $\hat{p}^+$  is lower or equal to the threshold  $\epsilon$ . The quality of the estimation depends on local optimality threshold values, and the sample size  $\lambda$ . We study different values, proportional to the bitstring length:  $\epsilon \in \{1/(j \cdot N), \mid j \in \{1, 2, 4, 8\}\}$ , and  $\lambda \in \{2^i N \mid i \in \{0, \dots, 7\}\}$ .

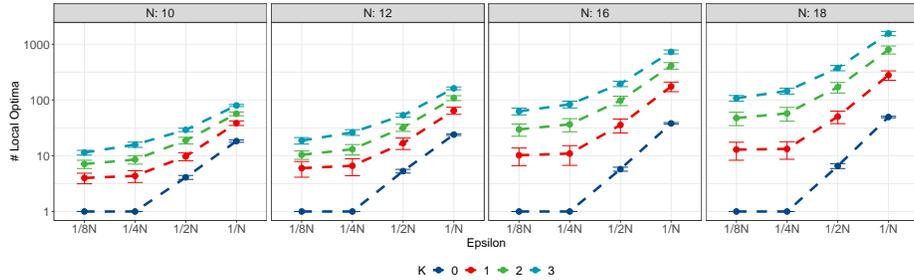
## 4.2 Experimental Results

In this section, we illustrate the relevance of the SLO definition, and we show preliminary scenarios for analyzing stochastic fitness landscapes<sup>5</sup>.

**Number of Stochastic Local Optima.** Fig. 1 shows the number of SLO over all instances according to the degree of non-linearity  $K$ , the problem dimension  $N$ , and different mutation rates  $p$ . The threshold is set to  $\epsilon = 1/(4N)$  (other values are reported below). In order to increase the estimation accuracy, the sample size is set to the largest value  $\lambda = 128N$ . Except for a very large mutation rate  $p = 8/N$ , and a small problem dimension  $N < 16$ , the expected number of SLO seems to increase fast with the degree of non-linearity  $K$ . It also increases with the problem size  $N$ . As expected, the trend follows the number of local optima for the classical 1 bit-flip neighborhood. These first results show that the proposed definition of SLO make sense w.r.t. the multimodality of NK-landscapes. The number of SLO is also impacted by the mutation rate, and a more precise analysis is detailed below.

**Local Optimality Threshold.** In Fig. 2, we report the number of SLO as a function of the local optimality threshold  $\epsilon$ . The sample size is set to the largest value  $\lambda = 128N$ , and the bit-flip mutation rate is set to the typical setting of  $p = 1/N$ . The number of SLO increases with  $\epsilon$  for all problem instances.

<sup>5</sup> code, and data are available on <https://gitlab.com/b.aboutaib/slo>



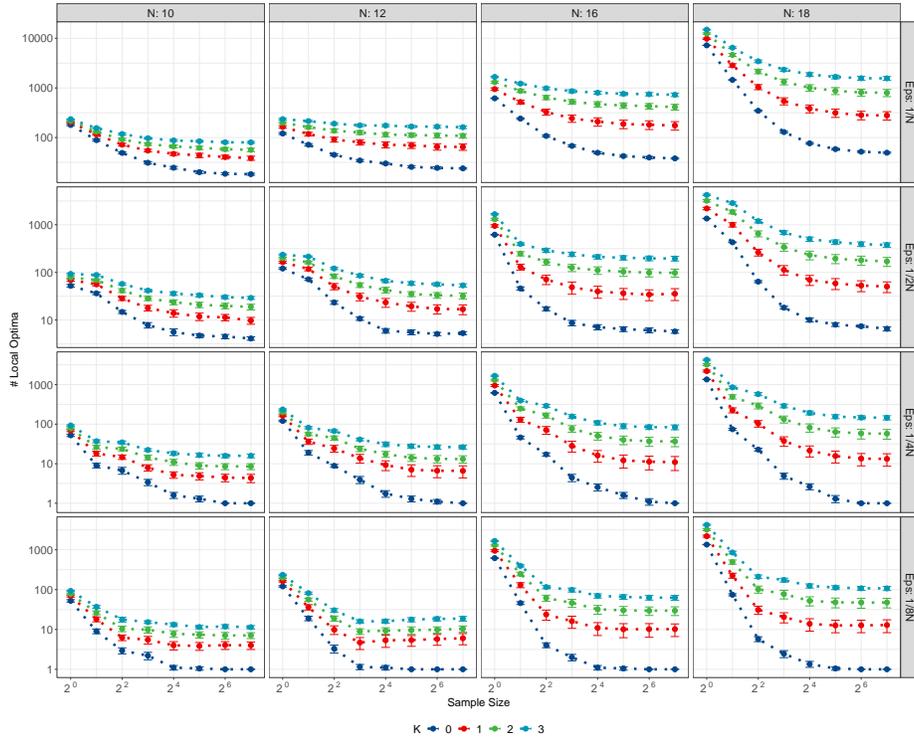
**Fig. 2.** Number (average value and confidence interval) of SLO with respect to the local optimality threshold  $\epsilon$ , for different instances (size  $N$  and non-linearity  $K$ ).

Indeed, a small threshold allows fewer solutions to be a SLO. Let us remind that the extreme setting of  $\epsilon = 0$  implies that a single solution is a SLO: the global optimum. Notice that the number of SLO increases with the degree of non-linearity  $K$ , and the problem dimension  $N$  for all investigated  $\epsilon$ -values. For large  $\epsilon$ -values, the number of SLO is larger than for the standard 1 bit-flip neighborhood. In this case, the inverse of the  $\epsilon$ -value is larger than  $1/N$ .

**Impact of the Sample Size.** We now investigate how the sample size impacts the estimation of the number of SLO. From a statistical point of view, a large sample size gives a better estimation of the improving probability considered in Definition 2. Therefore, a better estimation of the number of SLO is expected. Fig. 3 shows the estimated number of SLO as a function of the sample size for different values of  $N$ ,  $K$ , and  $\epsilon$ . The estimated number decreases with the sample size to converge toward the number of SLO. The empirical mean estimator of improving probability tends to overestimate the number of SLO when the sample size is too small with respect to the  $\epsilon$ -value. At a first sight, a sample size of about 2 or 3 times the inverse of  $\epsilon$  seems to provide a fair estimation. However for a given sample size, the estimated number of SLO increases with non-linearity  $K$ , and problem dimension  $N$ . Further theoretical studies should allow us to improve this first empirical finding.

**Mutation Rate vs. Local Optimality Threshold.** In Fig. 4, we report the number of SLO as a function of the bit-flip mutation rate  $p$ , for different  $\epsilon$ -values and different degrees of non-linearity  $K$ . The problem dimension is set to  $N = 18$ , and the largest sample size of  $\lambda = 128 N$  is used. In order to better appreciate the trend, additional rate values are experimented:  $p = c/N$ , with  $c \in \{0.5, 1, \dots, 8\}$ . The local optimality threshold is set to  $\epsilon \in \{1/N, 1/(2N), 1/(4N), 1/(8N)\}$ .

All curves have a convex shape. As such, given a threshold value  $\epsilon$ , there is a bit-flip mutation rate that minimizes the expected number of SLO, which does not map to the extreme bounds of the domain. In other words, the lowest number of SLO is reached at a particular trade-off point between low and high bit-flip mutation rates. For example, for  $K = 2$  and  $\epsilon = 1/(2N)$ , the mutation

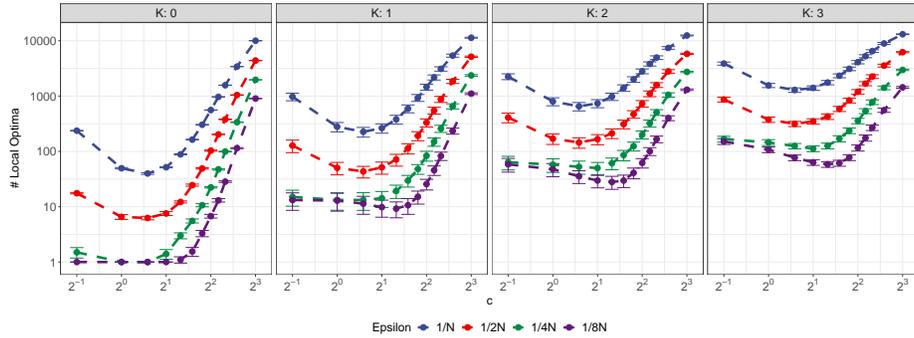


**Fig. 3.** Number (average value and confidence interval) of SLO with respect to the sample size, for different instances (size  $N$  and non-linearity  $K$ ) and local optimality threshold  $\epsilon$ . The mutation rate is set to  $p = 1/N$  (i.e.,  $c = 1$ ).

rate that minimizes the number of SLO is  $1.5/N$ , but changing  $\epsilon$  to  $1/(8N)$  shifts this mutation rate to  $2.5/N$ . Interestingly, this observation suggests that there exists an accurate mutation rate that reduces the multimodality of the stochastic fitness landscape, and that does *not* correspond to the largest bit-flip mutation rate. To the best of our knowledge, this is the first fitness landscape analysis that brings an understanding of a relevant mutation rate for a stochastic operator. A high mutation rate results in a landscape with many local optima. Hence, a local search would be easily stuck, and would not benefit from potential local improvements. A low mutation rate also induces a large number of local optima. However, by contrast, it shall be understood as a lack of exploration with respect to the local optimality threshold. Remember that a single feature is not able to explain all facets of search difficulty, and other metrics from the fitness landscape would be required to have a better global picture.

### 4.3 Stochastic Fitness Landscape Analysis

**Iterated Local Search.** In this section, we illustrate the potential usefulness of a stochastic fitness landscape analysis for the design of search algorithms based

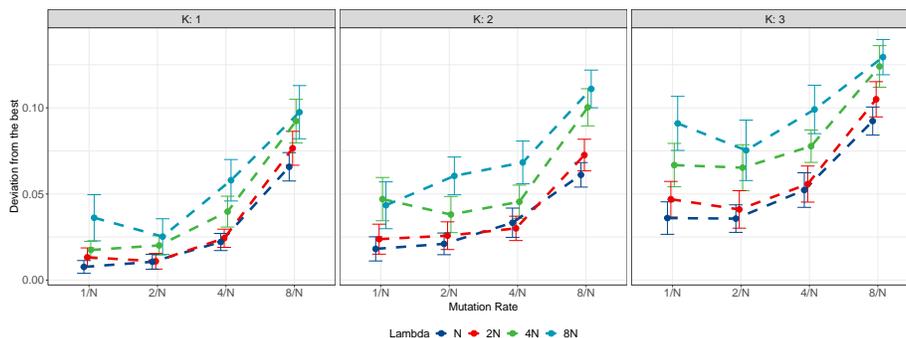


**Fig. 4.** Number (average value and confidence interval) of SLO with respect to the bit-flip mutation rate  $c/N$ , for different instances (non-linearity  $K$ ) and local optimality threshold  $\epsilon$  (see legend). The problem size is  $N = 18$ .

on stochastic local search operators. We consider the Iterated Local Search (ILS) framework [12] as a case study. ILS aims at escaping from poor local optima based on a perturbation mechanism followed by a local search procedure. Whenever the local search falls into a local optimum, it is perturbed by means of random modifications to obtain a new (inferior) solution from which another local search round starts. This process is iterated until the computational budget is exhausted. In the following, we analyze the performances of ILS on NK-landscapes. In particular, we perform 30 independent ILS executions on a randomly generated instance with  $N = 18$  and  $K \in \{1, 2, 3\}$ . We set the perturbation rate to 0.3; that is, each bit is flipped with a rate of 0.3. The local search components considered within the ILS is a first-improvement hill-climbing algorithm, where at most  $\lambda$  solutions are produced at each step by means of a stochastic bit-flip mutation with a rate of  $c/N$ . If there is no strict improvement, the current solution is considered as a stochastic local optimum, and a perturbation is performed for further iterations. The maximum number of fitness evaluations for the ILS is set to  $10^4$ .

In Fig. 5, we report the relative deviation (to be minimized) from the best known solution of the final fitness value obtained by the ILS for different mutation rates,  $\lambda$ -values, and degrees of non-linearity. For any  $\lambda$ -value, there is a mutation rate that maximizes the expected ILS performance. Interestingly, in most of  $\lambda$  scenarios, this value corresponds, to the mutation rate that minimizes the number of SLO, as disclosed above. This illustrates that analyzing the stochastic fitness landscape of the target problem might actually provide insightful information about the suitable configuration of stochastic local search algorithms.

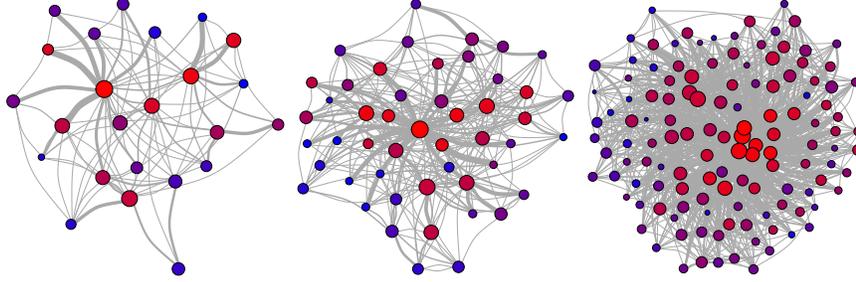
**Stochastic Local Optima Network.** In order to complement our analysis of stochastic fitness landscapes, we now naturally extend the concept of Local Optima Network (LON) [4, 13] to stochastic local optima. We define the Stochastic LON (SLON) as a graph where nodes are *stochastic* local optima, and edges represent the pairwise connections between stochastic local optima with respect



**Fig. 5.** Deviation of solution quality, from the best known solution, (average value and confidence interval) obtained by ILS with respect to the bit-flip mutation rate, for different instances (non-linearity  $K$ ) and sample sizes  $\lambda$ . The problem size is  $N = 18$ .

to another stochastic operator. More particularly, an edge  $(x_i, x_j)$  is defined and weighted so as to render the probability of reaching SLO  $x_j$  from SLO  $x_i$ . In this work, escape edges [19], as defined in Section 2, are computed according to a bit-flip mutation operator.

We generate a SLON for an NK-landscape with  $N = 18$  and  $K = 2$ . The SLO is defined by the bit-flip stochastic operator with three different mutation rates  $p \in \{1/N, 2/N, 3/N\}$ , and a stochastic operator for escape edges with a mutation rate of  $3/N$ . The sample size to estimate SLO and edges is set to  $\lambda = 4N$ . Fig. 6 shows the obtained SLON for the same NK-landscape. The node colors indicate the fitness value: the redder, the better. The node size is logarithmically proportional to the size of the corresponding basin of attraction. The edge size is linearly proportional to its weight (self-loops are omitted to improve readability). We cannot report the SLON and related metrics for all instances due to space restriction. However, they are consistent with the visual impression of Fig. 6. For the considered sample size  $\lambda$ , the main observations are as follows: the number of SLO increases with the mutation rate, the density of edges decreases with the mutation rate, and so does the self-loop weights  $w_{ii}$ . The feature from standard (deterministic) LON that is the most correlated with the performance of ILS is known to be the average distance between local optima and the global optimum [13]. The distance between two nodes  $i$  and  $j$  is defined as the inverse of the weight  $w_{ij}$ . Over all instances with  $N = 18$  and  $K = 2$ , the expected average distance to the global optimum is 8.48, 7.67, and 9.80, respectively, for mutation rates  $1/N$ ,  $2/N$ , and  $3/N$ . This metric suggests that the stochastic fitness landscape corresponding to a mutation rate of  $2/N$  is ‘easier’ to search than the one corresponding to a mutation rate of  $1/N$  or  $3/N$ . This observation is in line with the performance of ILS reported in Fig. 5. Indeed, the SLON for  $p = 1/N$  has a lower number of SLO, but nodes with higher fitness values seem to be clustered, which corresponds to a funnel structure [17]. By contrast, for  $p = 2/N$ , despite a larger number of SLO, paths to the global optimum seem



**Fig. 6.** Stochastic local optima networks for an NK-landscape with  $N = 18$  and  $K = 2$ . From left to right, the mutation rate is:  $p = 1/N$ ,  $2/N$ , and  $3/N$ , respectively.

to be shorter, and then more likely to happen during the search process of ILS. For  $p = 3/N$ , we infer that the huge number of SLO decreases the probability of reaching the global optimum by following a path on the network, thus inhibiting the performance of ILS.

## 5 Discussion and Further Considerations

In this paper, we defined and analyzed fitness landscapes based on stochastic search operators. Based on this definition, we empirically investigated enumerable instances from NK-landscapes. More particularly, we studied the number of stochastic local optima as a preliminary feature of such stochastic fitness landscapes, showing the relevance of stochastic local optima when measuring the multimodality of stochastic fitness landscapes. We also studied the underlying stochastic local optima networks. We found out that there is a critical region in the stochastic operator setting (the mutation rate) in which stochastic local optima are more scarce. A proper setting within this region would make the computational effort to solve the problem much more effective and would result in better solution quality.

Let us emphasize that the proposed definition of local optimality and fitness landscape for stochastic operators is not entitled to any particular problem class. Moreover, although we exhaustively enumerated the search space of NK-landscapes in order to avoid any bias in our current analysis, this is obviously not practical for large-scale optimization problems. As such, we plan to investigate sampling procedures that will enable studying the stochastic fitness landscape of other academic and real-world optimization problems. We hope that the proposed definition will enable analyzing better many recurrent problems and optimization algorithms, such as population-based evolutionary, estimation of distribution, or genetic programming algorithms.

### Acknowledgements

We are very thankful to the CALCULCO center of Université du Littoral Côte d’Opale for providing computational resources used in this paper.

## References

1. Alyahya, K., Rowe, J.E.: Simple random sampling estimation of the number of local optima. In: International Conference on Parallel Problem Solving from Nature. pp. 932–941. Springer (2016)
2. Basseur, M., Goëffon, A.: Climbing combinatorial fitness landscapes. *Applied Soft Computing* **30**, 688–704 (2015)
3. Bosman, A.S., Engelbrecht, A., Helbig, M.: Visualising basins of attraction for the cross-entropy and the squared error neural network loss functions. *Neurocomputing* (2020)
4. Chicano, F., Daolio, F., Ochoa, G., Verel, S., Tomassini, M., Alba, E.: Local optima networks, landscape autocorrelation and heuristic search performance. In: International Conference on Parallel Problem Solving from Nature. pp. 337–347. Springer (2012)
5. Elorza, A., Hernando, L., Mendiburu, A., Lozano, J.A.: Estimating attraction basin sizes of combinatorial optimization problems. *Progress in Artificial Intelligence* **7**(4), 369–384 (2018), <https://doi.org/10.1007/s13748-018-0156-6>
6. Fieldsend, J.E., Alyahya, K.: Visualising the landscape of multi-objective problems using local optima networks. In: Proceedings of the Genetic and Evolutionary Computation Conference Companion. pp. 1421–1429 (2019)
7. Hernando, L., Mendiburu, A., Lozano, J.A.: An evaluation of methods for estimating the number of local optima in combinatorial optimization problems. *Evolutionary computation* **21**(4), 625–658 (2013)
8. Hernando, L., Mendiburu, A., Lozano, J.A.: Anatomy of the attraction basins: Breaking with the intuition. *Evolutionary computation* **27**(3), 435–466 (2019)
9. Kauffman, S.A.: The origins of order: Self-organization and selection in evolution. OUP USA (1993)
10. Liefvooghe, A., Daolio, F., Verel, S., Derbel, B., Aguirre, H., Tanaka, K.: Landscape-aware performance prediction for evolutionary multi-objective optimization. *IEEE Transactions on Evolutionary Computation* (2019)
11. Liefvooghe, A., Derbel, B., Verel, S., López-Ibáñez, M., Aguirre, H., Tanaka, K.: On pareto local optimal solutions networks. In: International Conference on Parallel Problem Solving from Nature. pp. 232–244. Springer (2018)
12. Lourenço, H.R., Martin, O.C., Stützle, T.: Iterated local search: Framework and applications. In: *Handbook of Metaheuristics*, pp. 363–397. Springer (2010)
13. Ochoa, G., Verel, S., Daolio, F., Tomassini, M.: Local optima networks: A new model of combinatorial fitness landscapes. In: *Recent advances in the theory and application of fitness landscapes*, pp. 233–262. Springer (2014)
14. Stadler, P.F.: Fitness landscapes. In: Lässig, M., Valleriani (eds.) *Biological Evolution and Statistical Physics. Lecture Notes Physics*, vol. 585, pp. 187–207. Springer-Verlag, Heidelberg (2002)
15. Tari, S., Basseur, M., Goëffon, A.: Worst improvement based iterated local search. In: *European Conference on Evolutionary Computation in Combinatorial Optimization*. pp. 50–66. Springer (2018)
16. Thierens, D.: An adaptive pursuit strategy for allocating operator probabilities. In: *Proceedings of the 7th annual conference on Genetic and evolutionary computation*. pp. 1539–1546 (2005)
17. Thomson, S.L., Daolio, F., Ochoa, G.: Comparing communities of optima with funnels in combinatorial fitness landscapes. In: *Proceedings of the Genetic and Evolutionary Computation Conference*. pp. 377–384 (2017)

18. Thomson, S.L., Ochoa, G., Verel, S., Veerapen, N.: Inferring future landscapes: Sampling the local optima level. *Evolutionary Computation* pp. 1–22 (2020)
19. Verel, S., Daolio, F., Ochoa, G., Tomassini, M.: Local optima networks with escape edges. In: *International Conference on Artificial Evolution (Evolution Artificielle)*. pp. 49–60. Springer (2011)
20. Verel, S., Ochoa, G., Tomassini, M.: Local optima networks of nk landscapes with neutrality. *IEEE Transactions on Evolutionary Computation* **15**(6), 783–797 (2011)
21. Weinberger, E.D.: Local properties of kauffmans n-k model: A tunably rugged energy landscape. *Physical Review A* **44**(10), 6399 (1991)
22. Wright, A.H., Thompson, R.K., Zhang, J.: The computational complexity of nk fitness functions. *IEEE Transactions on Evolutionary Computation* **4**(4), 373–379 (2000)
23. Wright, S.: The roles of mutation, inbreeding, crossbreeding, and selection in evolution. In: *Proceedings of the Sixth International Congress of Genetics* 1. pp. 356–366 (1932)