WCCI 2012 IEEE World Congress on Computational Intelligence
June, 10-15, 2012 - Brisbane, Australia IEEE CEC

Landscape Characterization of Numerical Optimization Problems Using Biased Scattered Data

Mario A. Muñoz*, Michael Kirley† and Saman K. Halgamuge*
*Department of Mechanical Engineering, The University of Melbourne, Melbourne, Australia
†Department of Computing and Information Systems, The University of Melbourne, Melbourne, Australia
e-mail: mariom@student.unimelb.edu.au, {mkirley,saman}@unimelb.edu.au

***Abstract*—The characterization of optimization problems over continuous parameter spaces plays an important role in optimization. A form of "fitness landscape" analysis is often carried out to describe the problem space in terms of modality, smoothness and variable separability. The outcomes of this analysis can then be used as a measure of problem difficulty and to predict the behaviour of a given algorithm. However, the metric value estimates of the landscape characterization are dependent upon the representation scheme adopted and the sampling method used. Consequently, the development of a complete classification of problem structure and complexity has proven to be challenging. In this paper, we continue this line of research. We present a methodology for the characterization of two dimensional numerical optimization problems. In our approach, data extracted during the search process is analyzed and the dependency of the results to the nominated sampling method are corrected. We show via computational simulations that the calculated metric values using our approach are consistent with the results from random experiments. As such, this study provides a first step towards the on-line calculation of fitness landscape characterization metrics and the development of empirical performance models of search algorithms. Advances in these areas would provide answers to the algorithm selection and portfolio configuration problems.**

***Index Terms*—Data analysis, Heuristic algorithms, Information entropy, Optimization, Search problems**

I. Introduction

Search algorithms based on stochastic and heuristic rules have been used to solve a wide range of optimization problems, particularly those problems where there is neither sufficient knowledge about the objective function nor sufficient time for intense investigation of the problem [1]–[3]. However, an adequate explanation of why the algorithms were successful when tackling the given optimization problems is lacking. Subsequently, it is risky to assert that the algorithm under study would work effectively on problems beyond those in the test set. Indeed, it would be very optimistic to expect that an algorithm would work reasonably well across a wide range of problems unless some restrictions are placed [4]. Hence, it is necessary to be careful when generalizing results based on empirical tests.

Ideally, measures are required to quantify problem difficulty and to help predict whether a search algorithm might be successful on a given optimization problem. Landscape characterization refers to the development of methods, which provide metrics related to the attributes of an optimization problem such as modality, smoothness and variable separability and thus problem difficulty [5]. Landscape characterization methods have been thoroughly investigated over the past twenty years [6]–[22]. However, their use is not widespread because it is easy to find counterexamples [5]. In addition, characterization metrics typically require precise calculations using larger volumes of data [23], the results of which are dependent on the representation and the sampling method used [24]. Nevertheless, if the only accessible information available are freely chosen observations, an estimation of these metrics would be useful when characterizing the problem difficulty. We also have to acknowledge that a single measure would not be useful to detect every single attribute of the "fitness landscape".

In this paper, we present a methodology for the characterization of two dimensional numerical optimization problems. Our methodology analyses data extracted during the search process, hence it removes the necessity for a separated sampling experiment. Importantly, our approach corrects the dependency of the results from the parameters of the sampling algorithm. The estimated metrics can be used to quantify the modality, smoothness and variable scaling of the optimization problem. We show via computational simulations that our approach produces metric values that are consistent with the results from random experiments. As such, this work represents a first step towards the on-line calculation of these metrics and the development of empirical performance models of search algorithms. Advances in these areas would provide some answers to the algorithm selection and portfolio configuration problems.

The paper is organized as follows: In Section II we describe the conceptual framework for landscape characterization and analyze the reasons that make optimization problems hard. In Section III we formulate our method for landscape characterization. In Section IV we describe the experiments carried out and the obtained results. Finally, we discuss our results and propose further work in Section V.

II. Background

A. Black-box Numerical Optimization

Numerical optimization problems have a discrete domain of feasible solutions when solved in a digital computational machine. This means that both the search set, $\mathcal{X} \subset \mathbb{R}^n$, and the set of possible objective values, $\mathcal{Y} \subset \mathbb{R}$, although very large, are finite. Let f be a function such that:

$$f : \mathcal{X} \mapsto \mathcal{Y}$$

Let $\mathbf{x}_i \in \mathcal{X}$ be a point with a function value $y_i = f(\mathbf{x}_i) \in \mathcal{Y}$. In this type of problem, the pairs $(\mathbf{x}_i, y_i), i = 1, \ldots, p$ are the only accessible information and they can be chosen freely. Without loosing generality, assume that the goal is to minimize f. This means that we need to find one or more candidate solutions $\mathbf{x}_o \in \mathcal{X}, y_o = f(\mathbf{x}_o)$, such that $|y_o - y^*| \ll \delta$, where $\delta \to 0$ and $\mathbf{x}^* \in \mathbb{R}^n, y^* = f(\mathbf{x}^*)$ are the location of the global optimum and its value, which may or may not be available for representation in a digital computer. This restriction is made because, most of the time, obtaining the global optimum is neither feasible nor essential in practice [25]. It is also appropriate to assume that any *a priori* information available about the problem has been used in defining an appropriate set of variables, feasibility ranges or representation [26]. Therefore, it is acknowledged that the expertise of the designer in correctly specifying the problem cannot be replaced but complemented.

There are many methods available to search for $\mathbf{x}_o$. Ideally, the method selected does not suffer of *premature convergence*. This means that, after a bounded number of function evaluations, the method is not able to generate solutions outside a small area under examination and the solution obtained is unacceptable.

B. Search Landscapes

The search landscape is a commonly used metaphor to describe how a search method solves a particular optimization problem. In a two dimensional search space, it is possible to make an analogy of the structure of the problem to a surface composed of ridges, valleys and basins, each one of them strongly related to its neighbors. In this surface the local or global optimal points are located on the highest or deepest areas [11]. Although most practical problems will have a dimension several orders of magnitude larger than two, this metaphor helps us to understand what is needed for a successful search [27]. The choices made during the problem specification strongly influence the structure and complexity of the landscape [26].

A search landscape $\mathcal{L}$ for a function f can be defined as the tuple $\mathcal{L} = (\mathcal{X}, f, d)$, where d denotes a distance measure or *metric*. The existence of a distance allows the definition of neighborhood, $\mathcal{N}_\omega(\mathbf{x}_i) \subset \mathcal{X}$, which is generated by applying an operator ω to a vector $\mathbf{x}_i$ in order to transform it to a vector $\mathbf{x}_j$ as follows:

$$\mathbf{x}_j \in \mathcal{N}_\omega(\mathbf{x}_i) \Leftrightarrow d_\omega(\mathbf{x}_i, \mathbf{x}_j) < r$$

where $r \to 0$ is the radius of a hypersphere centered in $\mathbf{x}_i$. This definition is fundamental for our understanding of the landscape [26] as experience suggests that the type and size of the neighborhood is crucial in the practical performance of local search methods [14]. It is also noteworthy that the concepts of distance and neighborhood become less meaningful as n increases due to a loss of contrast between distances. This effect is known as the *curse of dimensionality* [28].

A neighborhood definition allows us to refine the concepts of local and global optima [27]. A local optima for a landscape $\mathcal{L}$ is a vector $\mathbf{x}_l \in \mathcal{X}$ such that $|y_j - y^*| > |y_l - y^*|, \forall \mathbf{x}_j \in \mathcal{N}_\omega(\mathbf{x}_l)$. The set of local optima can be denoted as $\mathcal{X}_l \subset \mathcal{X}$. The global optima for a landscape $\mathcal{L}$ is a vector $\mathbf{x}_o \in \mathcal{X}$ such that $y_o \leq y_l, \forall \mathbf{x}_l \in \mathcal{X}_l$. The set of global optima can be denoted as $\mathcal{X}_o \subseteq \mathcal{X}_l$. A *unimodal* landscape is the one such that $|\mathcal{X}_l| = 1$, while a *multimodal* landscape is the one such that $|\mathcal{X}_l| > 1$. It is not always the case that a multimodal problem poses more difficulties than a unimodal landscape [29], [30].

A closely related concept to multimodality is the one of *smoothness*, which refers to the amount of fluctuation in the landscape and can be informally classified as rugged, smooth or neutral. A *rugged* landscape exhibits large fluctuations between points, presenting several local optima and steep ascents and descents. A *neutral* landscape exhibits large flat areas or plateaus, where changes in the input do not generate significant changes in the output. These extremes make both gradient and correlation information less useful and makes it difficult to assert if one region is worth exploring. Non-smooth landscapes are usually the result of weak causality. While there is no viable method that can directly mitigate weak causality, it is often the result of a bad specification of the optimization problem [29].

Sometimes $\mathcal{X}_l$ defines a global structure, which can be exploited to find $\mathbf{x}_o$. Therefore if the structure has fluctuations or modality, the problem becomes difficult to solve [30]. It is also possible to find problems who exhibit *deceptiveness* — the global structure and gradient information leads the algorithm away from the optima, rendering the optimization algorithm less efficient than random search or exhaustive enumeration [29].

Another definition derived from the landscape metaphor is the *basin of attraction* [27]. If we consider that a local search can be thought of as a function:

$$\mu : \mathcal{X} \mapsto \mathcal{X}_l$$

where $\mathbf{x}_i$ is the initial point, $\mu(\mathbf{x}_i)$ is the local optimum that it reaches. Then, the basin of attraction of $\mathbf{x}_l$ can be defined as the set $B(\mathbf{x}_l) = \{\mathbf{x}_i : \mu(\mathbf{x}_i) = \mathbf{x}_l\}$. Hence, the search space is split into areas each with its own basin [13]. The shape and distribution of the basins are key attributes. A landscape with homogeneous basins is one such that $|B(\mathbf{x}_l)| \propto \frac{1}{y_l}, \forall \mathbf{x}_l \in \mathcal{X}_l$, a desirable characteristic [30]. Also, a landscape can have a *deterministic configuration* if its basins have the same size, (i.e., $|B(\mathbf{x}_l)| \approx |\mathcal{X}| / |\mathcal{X}_l|, \forall \mathbf{x}_l \in \mathcal{X}_l$); or a *random configuration* if its basins have different sizes [13].

Variable scaling also has an effect on the shape of the basin of attraction. If a problem behaves differently in each dimension, it might be necessary to perform small steps on some dimensions and larger ones in others. Usually this is due to anisotropic or non-spherical basins. A problem can be even more difficult to solve if the variables are non-separable, which means that the problem cannot be partitioned into subproblems of lower dimensionality each of which is easier to solve. Therefore, variations on the step size must be all encompassing.

It is possible that the search landscape possesses different properties at different locations, e.g. highly multimodal in some regions, while flat and neutral in others. Such *globally inhomogeneous* landscape would be difficult to characterize, since global analysis would generate contradictory conclusions. Hence, it is important that the developed methods are *statistically isotropic*, which means that they do not depend on the particular order of sampling nor the initial point where the search was begun [11], [19].

A problem with landscape theory is that the existent models cannot be used as framework when an algorithm uses multiple operators, or even with different representations [26]. This is due to the fact that every operator used defines its own landscape, a view which is referred to as "one operator, one landscape" [26], [31]. For example, genetic algorithms with binary representation can define up to three different landscapes: a mutation landscape, a crossover landscape and a selection landscape. All of these landscapes have different characteristics for different problems, and usually are studied independently.

Overall, the landscape metaphor allows the description of specific attributes of the search space, which are considered influential in the performance of an optimization algorithm. These attributes can be described qualitatively in cases where the knowledge about the function is complete. To illustrate this point we present the *Comparing Continuous Optimization* (COCO) noiseless benchmark, which is a set of 24 scalable functions [32] whose well known qualitative characteristics are shown in Table I. But in cases where the only information available are the pairs $(\mathbf{x}_i, y_i)$, it is necessary to develop a set of techniques that can provide a numerical attribute. This procedure is known as landscape characterization.

III. Methodology

In this section, we describe a new approach for landscape characterization. The rationale behind our approach, is based two key issues: Firstly, it is difficult to represent all the attributes of a problem through a single measure. Secondly, although accurate estimators might be impossible to obtain, a sufficiently precise estimation might be enough to classify a problem. Hence, if a set of measures can be extracted during or after an algorithm run, and this set is related to a performance metric of the algorithm, characterization would provide some answers to the algorithm selection and tuning problems.

A first step in this endeavor is to modify currently existing methods so scattered data can be used. However, it is important to acknowledge that search algorithms are designed to explore promising areas of $\mathcal{X}$, which produces a large concentration of observations in small regions of the space. Hence, if raw data obtained during an algorithm run is used to extract any statistic measures, the results would suffer of sampling bias. This is a systematic error that causes some members of a population to be underrepresented in a sample, making impossible to infer the real characteristics of the population. We provide methods for the removal of such bias, and also for calculating the Information Content [11], [20], [21] and Basin of Attraction [10], [13], [22] measures from scattered data.

A. Removing bias from datasets

To correct the error produced by the sampling bias a weighted resample can be used. Let $\mathbf{x}_i$ be an observation in a sample from $\mathcal{X}$ of size p. Assume that each dimension of $\mathcal{X}$ has been split into α sections forming a partition of α^n bins. Let $\eta_u, u = 1, \ldots, \alpha^n$ be the number of elements in the bin u. If $\mathbf{x}_i$ is an element of the bin u, define ω_i as the weight of the point $\mathbf{x}_i$, which is calculated as follows:

$$\omega_{(i)} = \frac{1}{\eta_u}\left(1 - \frac{\eta_u}{p}\right) = \frac{p - \eta_u}{p \cdot \eta_u}$$

Now, let w_i be the empirical probability for $\mathbf{x}_i$ to be selected in the weighted resample, which is calculated by:

$$w_i = \frac{\sum_{j=1}^{i} \omega_j}{\sum_{j=1}^{p} \omega_j}$$

We select q observations from the original p using the inverse transformation method [33]. Two issues arise by using this method: first, it is possible that some elements have repeats, and second, not all the data from the original p might be selected in the resample. These issues are considered in the next sections.

B. Information content measures for continuous optimization landscapes

Information content is a demonstrated measure for ruggedness and multimodality of a function, but its implementation requires a set of observations extracted during a random walk with variable step size [20], [21]. The results show that the selected step size influences the measures obtained. To solve this limitation we devise the following procedure. Assume that the sample of q observations is ordered in a sequence of points following the rules:

- The first element of the sequence, $\mathbf{x}_1$, is randomly selected.
- The next elements are selected following the rule:

  $$\mathbf{x}_i = \left\{\mathbf{x} : \mathbf{x} \notin \{\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}\}, \mathbf{x} = \arg\min d(\mathbf{x}, \mathbf{x}_{i-1})\right\}$$

  where $d(\mathbf{x}, \mathbf{x}_{i-1}) = \|\mathbf{x} - \mathbf{x}_{i-1}\|$.

From the ordered sample we can create a symbol sequence $\Phi(\varepsilon) = \{\phi_1, \ldots, \phi_{q-1}\}$ such that $\phi_i \in \{\bar{1}, 0, 1\}$ given by $\phi_i(\varepsilon) = \Psi(i, \varepsilon)$ where:

$$\Psi(i, \varepsilon) = \begin{cases} \bar{1} & \text{if} \quad \frac{y_i - y_{i-1}}{d(\mathbf{x}_i, \mathbf{x}_{i-1})} < -\varepsilon \\ 0 & \text{if} \quad \left|\frac{y_i - y_{i-1}}{d(\mathbf{x}_i, \mathbf{x}_{i-1})}\right| \leq \varepsilon \\ 1 & \text{if} \quad \frac{y_i - y_{i-1}}{d(\mathbf{x}_i, \mathbf{x}_{i-1})} > \varepsilon \end{cases}$$

where $\varepsilon \geq 0$ is a sensitivity parameter used to control the degree of detail obtained. By including the euclidean distance

TABLE I

QUALITATIVE CLASSIFICATION OF THE COCO NOISELESS BENCHMARK FUNCTIONS, ACCORDING THE PROPERTIES OF MULTIMODALITY, GLOBAL STRUCTURE, SEPARABILITY, HOMOGENEITY, BASIN SIZES, GLOBAL-LOCAL CONTRAST AND PLATEAUS [30].

Function	Multimod.	Gl.Struct	Separable	Homogeneity	Basins	Gl.Loc.	Plateaus
1: Sphere	none	none	yes	high	none	none	none
2: Ellipsoidal separable	none	none	yes	high	none	none	none
3: Rastrigin separable	high	strong	yes	high	low	low	none
4: Büche-Rastrigin	high	strong	yes	high	medium	low	none
5: Linear Slope	none	none	yes	high	none	none	none
6: Attractive Sector	none	none	yes	medium	none	none	none
7: Step Ellipsoidal	none	none	yes	high	none	none	small
8: Rosenbrock	low	none	no	medium	low	low	none
9: Rosenbrock rotated	low	none	no	medium	low	low	none
10: Ellipsoidal high conditioned	none	none	no	high	none	none	none
11: Discus	none	none	no	high	none	none	none
12: Bent Cigar	none	none	no	high	none	none	none
13: Sharp Ridge	none	none	no	medium	none	none	none
14: Different Powers	none	none	no	medium	none	none	none
15: Rastrigin multimodal	high	strong	no	high	low	low	none
16: Weierstrass	high	medium	no	high	medium	low	none
17: Schaffer F7	high	medium	no	medium	medium	high	none
18: Schaffer F7 moderately ill-cond.	high	medium	no	medium	medium	high	none
19: Griewank-Rosenbrock	high	strong	no	high	low	low	none
20: Schwefel	medium	deceptive	no	high	low	low	none
21: Gallagher 101 Peaks	medium	none	no	high	medium	low	none
22: Gallagher 21 Peaks	low	none	no	high	medium	medium	none
23: Katsuura	high	none	no	high	low	low	none
24: Lunacek bi-Rastrigin	high	weak	no	high	low	low	none

between $\mathbf{x}_i$ and $\mathbf{x}_{i-1}$, the step size is taken into account in the analysis. Now the information content, which characterizes the ruggedness of the landscape is given by:

$$H(\varepsilon) = -\sum_{a \neq b} F_{ab} \log_6 F_{ab}$$

where $a, b \in \{\bar{1}, 0, 1\}$ and F_{ab} is the frequency of the substring ab in the symbol string $\Phi(\varepsilon)$. The number of possible substrings where $a \neq b$ is 3!, hence the logarithm is taken with base six so the information content is in the interval $[0, 1]$. As this is an entropy measure, $0 \log_6 0 \equiv 0$ [20].

The partial information content characterizes the modality of the landscape. If a new string $\Phi'(\varepsilon)$ is constructed from $\Phi(\varepsilon)$ by removing all the 0's and any repeated symbols. This new string would have the form of "$\ldots\bar{1}1\bar{1}1\bar{1}\ldots$". Then, the partial information content $M(\varepsilon)$ is defined as:

$$M(\varepsilon) = \frac{|\Phi'|}{q-1}$$

From this analysis, we define the following measures as points of interest [20]: Maximum information content (1) and Maximum information sensitivity (2) are intended to identify unimodal from multimodal problems; Settling sensitivity (3) is intended to identify variable scaling differences; Initial partial information (4) and Half partial information sensitivity (5) are intended to detect neutral areas. The measures also are expected to provide redundancy between each other.

$$H_{\max} = \max_{\varepsilon} \{H(\varepsilon)\} \tag{1}$$

$$\varepsilon_{\max} = \max_{\varepsilon} \{\varepsilon : H(\varepsilon) = H_{max}\} \tag{2}$$

$$\varepsilon_s = \min_{\varepsilon} \{\varepsilon : H(\varepsilon) < 0.05\} \tag{3}$$

$$M_0 = M(\varepsilon = 0) \tag{4}$$

$$\varepsilon_{0.5} = \max_{\varepsilon} \{\varepsilon : M(\varepsilon) > 0.5 M_0\} \tag{5}$$

Figure 1 shows some result of this analysis. The top figure shows through markers the location of H_{max}, $\varepsilon_{\max}$, ε_s, M_0 and $\varepsilon_{0.5}$. The bottom figure shows the results from three different functions from Table I: A unimodal-smooth function, f_1, a unimodal-neutral function, f_7, and a multimodal function, f_{24}. The solid line represents the $H(\varepsilon)$ curve and the dashed line the $M(\varepsilon)$ curve.

C. Basin of attraction measures

In this section, we use the concept of basin of attraction to extract measures from data obtained during a search algorithm run. We start with a sample of q points, which may or may not contain repeated elements. The first step is to remove any repeats from this dataset. Next, define a neighborhood structure using a mesh calculated through Delaunay triangulation. The points which are only one jump distant in the mesh are neighbors.

To start the algorithm, we select a random point in the mesh, and follow the steepest descent heuristic over the grid. If the

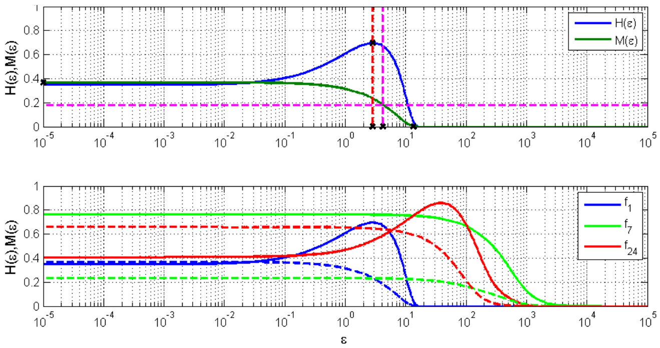


Fig. 1. Result for the information content method. (Top) Curves for the $H(\varepsilon)$ and $M(\varepsilon)$ for the f_1 function from Table I. The $\times$ marks indicate the location of H_{max}, $\varepsilon_{\max}$, ε_s, M_0 and $\varepsilon_{0.5}$. (Bottom) Curves for three functions from Table I with different qualitative characteristics.

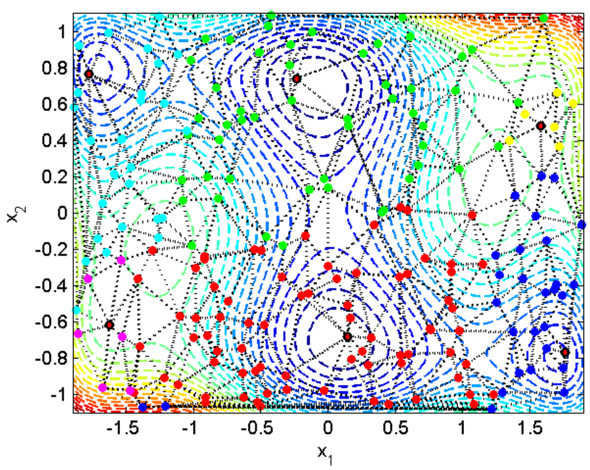


Fig. 2. Contour plot of the Six-hump camel back function, the triangulation mesh, the six effectively identified local optima and its basins of attraction.

point is a local minimum, we define the basin of attraction as the set of points in the path followed to reach the local optima. The procedure is repeated until all points have been visited at least once. To avoid additional computational effort, if the point currently under evaluation has been visited previously, we add the current path to the basin of attraction of the current point. Figure 2 shows the result of the algorithm for the well known Six-hump Camel Back function. We use this function for this example as its low modality and similar basin sizes allow the generation of a clear graphic. The algorithm identifies effectively the four local optima and two global optima. It also generates an estimation of the basin sizes.

From this analysis we extract the following measures:

- Number of Local Optima (*NLO*), which is an approximation of $|\mathcal{X}_l|$ and identifies if the problem is unimodal or multimodal.
- Basin Size Ratio (*BSR*), which is calculated by (6) and it characterizes the existence of a dominant basin.

$$BSR = \frac{\max_{\mathbf{x}_l} |B(\mathbf{x}_l)|}{\min_{\mathbf{x}_l} |B(\mathbf{x}_l)|} \tag{6}$$

TABLE II

Average values of the Pearson χ^2 goodness of fit test. A one indicates a uniformly distributed sample.

No.	RS	PSO		CMA-ES	
		Biased	Debiased	Biased	Debiased
Average	1.0000	0.1717	1.0000	0.1600	0.9946

- Maximum distance between basins ($d_{\max}$), and Distance between the global optimum and the local optimum with the largest attraction basin (d_{opt}), which detect deceptive characteristics of the landscape.

IV. Experimental procedure and results

To test the performance of our methods, we use the COCO noiseless benchmark from Table I with $n = 2$ and three different sampling methods: Random sampling (RS), (1+1)CMA-ES [34], and the Particle Swarm Optimization (PSO) with Trelea's second parameter set [35]. RS is the comparison algorithm, and the other two algorithms were selected based only in their popularity. For each algorithm fifteen runs of 2000 function evaluations are made. If the algorithm converges to an optimal point during its run before completing the allocated evaluations, the algorithm is restarted at random to complete its allocation.

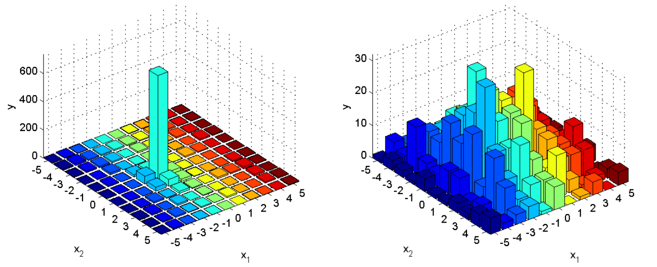
The required parameters were set as follows: the number of splits per dimension, α, was set at 50, the number of resamples was set at 100, and the number of points per resample, q, was set as 2000. For the Information content method, ε have values of 0, and logarithmically increasing from 10^{-5} to 10^{15}. The number of different values tested for ε is 1000.

A. *Resampling algorithm*

To evaluate the performance of the resampling method, we use the Pearson χ^2 goodness of fit test [36] over sample of 2000 observations in partition with $\alpha = 5$. The test is equal to one if the sample is uniformly distributed and it is zero otherwise. We carry out the experiment over the 100 samples per function and averaged the total results. To evaluate the efficiency of the method, we compare five different scenarios: RS sampling, PSO without removing the bias, PSO removing the bias, CMA-ES without removing the bias and CMA-ES removing the bias. The results, shown in Table II, confirm that applying our method produces a uniformly distributed sample.

To illustrate the effect of the resampling algorithm, we present a histogram for 1000 points extracted from the PSO dataset over the f_1 function, from biased (Fig. 3(a)) and debiased (Fig. 3(b)) datasets. It is noteworthy the difference between the two distributions, where the biased dataset has a high concentration of data in one sector. This sector is the location of the global optima found during the experiments with PSO.

Through these experiments, it is demonstrated that the resampling algorithm performs adequately. As the diversity in the dataset is higher, it is possible to obtain a uniform dataset. This step is necessary if we are to obtain adequate landscape



(a) (b)

Fig. 3. Histogram for a sample (a) Using uniform selection (b) Using weighted selection

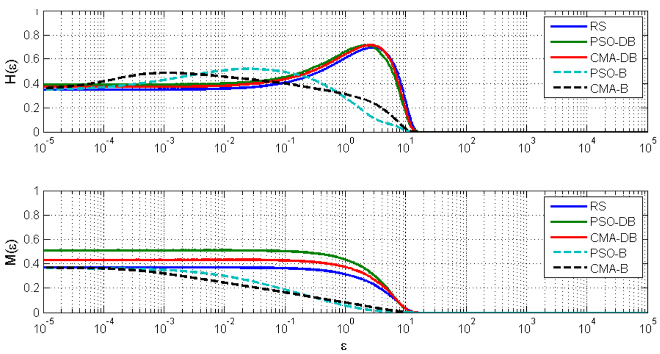


Fig. 4. Information content curves for the RS, PSO Debiased (PSO-DB), CMA-ES Debiased (CMA-DB), PSO Biased (PSO-B), and CMA-ES Biased (CMA-B) methods for the f_1 function

measures, as the bias has an effect on the results given by the measuring algorithms.

B. Sampling bias reduction

The next step is to evaluate the effect of the sampling bias in the measures. Figure 4 shows the result obtained by performing the information content analysis over 100 resamples of 2000 points for the f_1 function using RS, PSO and CMA-ES, the latter two with biased and debiased datasets. The f_1 function provides illustration of a typical case. To have a useful measure, the resulting data from the analysis must not differ significantly from the RS experiment. If they differ, it would be impossible to use the results in further analysis stages. Through visual inspection of the figure it can be noted that the biased resamples are noticeably distorted, and the difference between the results obtained with the random analysis are significant. While the results obtained with the debiased resamples are close to the RS sampling. To quantify this effect, let $\gamma_a : \mathcal{F} \mapsto \mathbb{R}$ be a landscape measurement function with sampling method a, γ_{RS} is landscape measurement with random sampling. Then, $e(\gamma_a)$ is the relative error of γ_a when compared with γ_{RS}, which is calculated with (7).

$$e(\gamma_a) = 100 \cdot \left| \frac{\gamma_{RS} - \gamma_a}{\gamma_{RS}} \right| \% \tag{7}$$

The results $e(\gamma_a)$ of this evaluation are shown in Table III, where boldfaced values indicate a relative error equal or lesser than 5%. The results show that the resampling effectivelly reduces the difference between results, with exception of M_0. Larger error in the basin of attraction measures is not as relevant, since they can be used in relative ranking. It is noteworthy that the PSO algorithm produces less biased results, when compared with (1+1)CMA-ES. This has a simple explaination. Since PSO is a population algorithm, it samples at different positions concurrently. Therefore, the distribution of the observations is closer to uniform, which was also demonstrated by the data from Table II.

C. Measure results and analysis

The measurements obtained through RS are shown in Table IV. To interpret this table it is convenient to use the classification in Table I and the descriptions by Finck, *et al.* [37]. Functions f_1 and f_5 are similar according to their qualitative characteristics and present close values for $\varepsilon_{\max}$, ε_s, M_0, and $\varepsilon_{0.5}$. While f_2 has the same qualitative characteristics, this function has a large anisotropic basin of attraction that forms a narrow valley, and it also possesses minor smooth local irregularities. The existence of the anisotropic basin produces differences between displacements in different dimensions. This difference is expressed by higher values of $\varepsilon_{\max}$ and ε_s. Non smooth problems are usually represented by values of $H_{\max}$ higher than 0.7, although an exception is function f_5. In particular, Functions f_{16}, f_{19}, f_{23} and f_{24} which can be considered almost noisy present the highest values of $H_{\max}$. Function f_7 is the only one that presents plateaus, which is detected by low $\varepsilon_{\max}$ and M_0, and can be identified as non smooth by $H_{\max}$. The existence of anisotropic basins of attraction, such as those presented in functions f_2, f_6, f_8 and f_9 is detected by a high value of $\varepsilon_{0.5}$. Function f_{12}, which has the most anisotropic narrow basin of attraction has the highest value of them all. Functions f_8 and f_9, f_2 and f_{10}, and f_3 and f_{15} are rotated versions of themselves, and as such they share similar measurements values.

Unimodal functions with isotropic basin of attraction such as f_1, f_5 and f_{14} are easily identified by the *NLO* measure. Unimodal functions with anisotropic basins are identified as having several local optima, such as f_2, f_{19}, f_{10} and f_{12}. This effect is a result of the meshing procedure, which divides the unique basin into sections. However, since the number of divisions is low, it is possible to avoid confusing this type of problems with medium modality problems, such as Functions f_{20} and f_{22}, or highly multimodal problems such as Functions f_3, f_4, f_{19} and f_{24}. Function f_7, which is the only one with neutral areas, is identified as having several local optima. Because the local search is unable to move after it hits a neutral area, this is considered the end of the search. Hence, this increases the number of local optima found.

Some clues about the distribution of the basin sizes can be obtained by analyzing both *NLO* and *BSR*. If a space has relatively uniform basins then $BSR \to 1$. If this ratio becomes smaller, e.g. on f_{20}, it is possible that the algorithm explores one area of the search space, as its basin is more attractive than others. If a highly multimodal problem presents a large

TABLE III
RELATIVE ABSOLUTE ERROR BETWEEN RS AGAINST PSO AND CMA-ES SAMPLING, FOR THE MEASURING TECHNIQUES DEVELOPED.

Algorithm		$H_{\max}$	$\log_{10}\varepsilon_{\max}$	$\log_{10}\varepsilon_s$	M_0	$\log_{10}\varepsilon_{0.5}$	*NLO*	*BSR*	$d_{\max}$	d_{opt}
PSO	Bias	14.1%	59.7%	13.2%	12.3%	47.0%	372.4%	43.0%	18.5%	52.1%
	Debias	**3.4%**	5.3%	**2.6%**	20.9%	**4.3%**	21.7%	23.8%	**4.6%**	26.0%
CMA-ES	Bias	20.5%	221.4%	16.8%	24.1%	93.7%	1455.5%	48.1%	14.5%	37.7%
	Debias	**3.8%**	**4.5%**	**2.7%**	20.5%	**4.7%**	20.8%	21.3%	5.8%	24.7%

TABLE IV
VALUES FOR THE INFORMATION CONTENT MEASURES FOR THE COCO BENCHMARK FUNCTIONS AND RANDOM SAMPLING

Fcn No.	$H_{\max}$	$\log_{10}\varepsilon_{\max}$	$\log_{10}\varepsilon_s$	M_0	$\log_{10}\varepsilon_{0.5}$	*NLO*	*BSR*	$d_{\max}$	d_{opt}
1	0.699	0.445	1.112	0.369	0.611	1.000	1.000	0.000	0.000
2	0.644	6.244	6.924	0.373	6.347	14.070	0.076	0.673	0.232
3	0.828	1.571	2.414	0.529	1.764	114.850	0.011	0.853	0.286
4	0.805	1.602	2.799	0.510	1.810	101.130	0.008	0.810	0.206
5	0.737	0.686	0.996	0.364	0.772	1.000	1.000	0.000	0.000
6	0.585	4.551	5.626	0.365	4.629	3.390	0.246	0.220	0.011
7	0.765	-5.000	3.206	0.234	2.322	429.010	0.018	0.966	0.091
8	0.577	2.695	4.847	0.378	3.352	19.610	0.019	0.474	0.090
9	0.584	2.726	4.941	0.369	3.255	16.070	0.018	0.407	0.111
10	0.633	6.236	7.202	0.371	6.375	14.670	0.053	0.688	0.192
11	0.617	6.297	7.216	0.370	6.427	14.860	0.034	0.701	0.256
12	0.492	5.704	10.619	0.371	8.141	14.140	0.034	0.667	0.190
13	0.741	2.121	2.483	0.370	2.219	14.410	0.065	0.698	0.157
14	0.599	-0.123	1.999	0.367	0.876	1.590	0.747	0.021	0.009
15	0.763	1.656	3.358	0.491	1.898	104.840	0.007	0.836	0.263
16	0.811	2.015	3.181	0.569	2.284	177.150	0.024	0.957	0.286
17	0.724	1.418	4.043	0.635	2.353	248.380	0.026	0.953	0.161
18	0.708	2.572	5.127	0.657	3.312	253.970	0.023	0.961	0.172
19	0.851	1.483	2.756	0.619	1.738	208.040	0.014	0.938	0.231
20	0.569	3.672	4.689	0.385	3.752	21.530	0.005	0.369	0.119
21	0.738	0.797	1.974	0.461	1.104	79.220	0.011	0.879	0.253
22	0.704	0.916	2.194	0.424	1.289	56.310	0.015	0.866	0.369
23	0.872	1.895	3.177	0.672	2.146	286.910	0.040	0.964	0.379
24	0.861	1.580	2.766	0.659	1.821	280.400	0.039	0.962	0.284

value of d_{opt} then it is possible that the problem has some deceiving characteristics.

The sampling bias imposed by the algorithms would change significantly our conclusions. If we consider $H_{\max}$, and the error imposed by the bias was incremental, this would mean that smooth functions such as f_1, f_2 and f_5, could be classified as rugged. However, if the error was decremental, highly rugged functions such as f_{23} and f_{24} would be classified as smooth. The change in the position of $\varepsilon_{\max}$ also would indicate a possible anisotropic problem for function f_1.

These results show that the method can be used to identify effectively ruggedness and neutrality in the functions. The inspection of the structure of the function allows the correlation between some values and the shape of the basin of attraction.

V. DISCUSSION AND FURTHER WORK

Developing a complete method to characterize a numerical optimization problems is a complex task. A key element of this endeavor is to acknowledge that a single measure cannot encapsulate all characteristics that make a problem difficult. Hence, it is necessary to develop alternative measures tailored for specific characteristics. In this work, we have demonstrated that it is possible to extract such measures, not only from random data, but also from data extracted during an algorithm run. Although it is necessary to pre-process the information to avoid sampling bias, this eliminates the need for additional experimentation.

In this paper, we have presented a new landscape characterization method for numerical optimization problems as a proof of concept. However, to create a complete characterization method there are several issues to be addressed. Firstly, we acknowledge that the COCO Benchmark does not have sufficient types of problems. Specifically, there are not sufficient problems with neutral areas. The methods developed in this paper detect neutrality with a low M_0, low $\varepsilon_{\max}$, and high *NLO* for the Step Ellipsoidal function. At this point, it is not possible to state definitely that all neutral problems have the similar measurement values. The inclusion of more test functions to the database would be ideal. Secondly, it is necessary to scale the methods to higher dimensions as practical problems have dimensionality much higher than two. Some difficulties to be addressed for this purpose are the selection of a structure for neighborhood definition and the limited number available observations. Thirdly, it is necessary to develop other measures that identify characteristics not evaluated by this paper. Variable separability is one of those characteristics which has major influence in the performance of an algorithm. However, as the analysis methods become more detailed, some measures might become redundant and unnecessary to calculate. The pruning of those measures would make the calculations more effective.

Finally, we expect to predict the performance of search algorithms in a function given a set of characteristics. For that it is possible to define a function $h : \mathcal{C} \mapsto \mathcal{P}$, where $\mathcal{C}$ is the space of characteristics of the problem and $\mathcal{P}$ is the space of performance of the algorithm. A nonlinear estimator of this function might be able to predict the possible performance of an algorithm if the training examples are sufficient and the input variables relevant. We will devote our further work to explore all of these ideas.

References

[1] T. Jansen, "On classifications of fitness functions," University of Dortmund, Tech. Rep. CI-76/99, November 1999.

[2] C. Blum and X. Li, "Swarm intelligence in optimization," in *Swarm Intelligence: Introduction and Applications*, ser. Nat. Comput. Ser. Springer, 2008, pp. 43–85.

[3] K. De Jong, "Parameter setting in EAs: a 30 year perspective," in *Parameter Setting in Evolutionary Algorithms*, ser. Stud. Comput. Intell., F. Lobo, C. Lima, and Z. Michalewicz, Eds. Springer, 2005, vol. 54, pp. 1–18.

[4] D. Wolpert and W. Macready, "No free lunch theorems for optimization," *IEEE Trans. Evol. Comput.*, vol. 1, no. 1, pp. 67–82, April 1997.

[5] B. Naudts and L. Kallel, "A comparison of predictive measures of problem difficulty in evolutionary algorithms," *IEEE Trans. Evol. Comput.*, vol. 4, no. 1, pp. 1–15, April 2000.

[6] E. Weinberger, "Correlated and uncorrelated fitness landscapes and how to tell the difference," *Biol. Cybern.*, vol. 63, pp. 325–336, 1990.

[7] P. Stadler, "Landscapes and their correlation functions," *J. Math. Chem.*, vol. 20, pp. 1–45, 1996.

[8] B. Naudts, D. Suys, and A. Verschoren, "Epistasis as a basic concept in formal landscape analysis," in *Proceedings of the 7th International Conference on Genetic Algorithms*, T. Bäck, Ed. Morgan Kaufmann, July 1997, pp. 65–72.

[9] S. Rochet, G. Venturini, M. Slimane, and E. El Kharoubi, "A critical and empirical study of epistasis measures for predicting ga performances: A summary," in *Third European Conference on Artificial Evolution*, 1998, pp. 275–285.

[10] J. Garnier and L. Kallel, "How to detect all attraction basins of a function?" in *Theoretical Aspects of Evolutionary Computation*, ser. Nat. Comput. Ser. Springer, 2000, pp. 343–365.

[11] V. Vassilev, T. Fogarty, and J. Miller, "Information characteristics and the structure of landscapes," *Evol. Comput.*, vol. 8, no. 1, pp. 31–60, 2000.

[12] A. Eremeev and C. Reeves, "Non-parametric estimation of properties of combinatorial landscapes," in *Applications of Evolutionary Computing*, ser. Lect. Notes Comput. Sc. Springer, 2002, vol. 2279, pp. 31–40.

[13] J. Garnier and L. Kallel, "Efficiency of local search with multiple local optima," *SIAM J. Discrete Math.*, vol. 15, no. 1, pp. 122–141, 2002.

[14] A. Eremeev and C. Reeves, "On confidence intervals for the number of local optima," in *Applications of Evolutionary Computing*, ser. Lect. Notes Comput. Sc. Springer, 2003, vol. 2611, pp. 115–115.

[15] P. Merz, "Advanced fitness landscape analysis and the performance of memetic algorithms," *Evol. Comput.*, vol. 12, no. 3, pp. 303–325, 2004.

[16] C. Reeves and A. Eremeev, "Statistical analysis of local search landscapes," *J. Oper. Res. Soc.*, vol. 55, no. 7, pp. 687–693, 2004.

[17] W. Hordijk and S. Kauffman, "Correlation analysis of coupled fitness landscapes," *Complexity*, vol. 10, no. 6, pp. 41–49, 2005.

[18] B. Qu and B. Xu, "On epistasis for measure of genetic algorithm hardness," in *Fuzzy Information Processing Society, 2006. NAFIPS 2006. Annual meeting of the North American*, june 2006, pp. 73 –76.

[19] H. Richter, "Coupled map lattices as spatio-temporal fitness functions: Landscape measures and evolutionary optimization," *Phys. Nonlinear Phenom.*, vol. 237, no. 2, pp. 167–186, 2008.

[20] K. Steer, A. Wirth, and S. Halgamuge, "Information theoretic classification of problems for metaheuristics," in *Proceedings of Simulated Evolution and Learning 2008*, ser. Lect. Notes Comput. Sc., vol. 5361. Springer, 2008, pp. 319–328.

[21] K. Malan and A. Engelbrecht, "Quantifying ruggedness of continuous landscapes using entropy," in *Proceedings of the 2009 IEEE Congress on Evolutionary Computation (CEC2009)*, may 2009, pp. 1440 –1447.

[22] P. Caamaño, A. Prieto, J. Becerra, F. Bellas, and R. Duro, "Real-valued multimodal fitness landscape characterization for evolution," in *Neural Information Processing. Theory and Algorithms*, ser. Lect. Notes Comput. Sc., K. Wong, B. Mendis, and A. Bouzerdoum, Eds., vol. 6443. Springer, 2010, pp. 567–574.

[23] J. He, C. Reeves, C. Witt, and X. Yao, "A note on problem difficulty measures in black-box optimization: Classification, realizations and predictability," *Evol. Comput.*, vol. 15, no. 4, pp. 435–443, 2007.

[24] G. Merkuryeva and V. Bolshakovs, "Structural analysis of benchmarking fitness landscapes," *Scientific Journal of Riga Technical University. Computer Sciences*, vol. 42, pp. 81–86, 2010.

[25] N. Hansen, "Variable metrics in evolutionary computation," Faculty Habilitation. Laboratoire de Recherche en Informatique – Université Paris-Sud, September 2009. [Online]. Available: http://www.lri.fr/~hansen/hansen-habil-manu.pdf

[26] T. Jones, "Evolutionary algorithms, fitness landscapes and search," Ph.D. dissertation, University of New Mexico, 1995.

[27] C. Reeves, "Fitness landscapes," in *Search Methodologies*, E. Burke and G. Kendall, Eds. Springer, 2005, pp. 587–610.

[28] M. Houle, H.-P. Kriegel, P. Kröger, E. Schubert, and A. Zimek, "Can shared-neighbor distances defeat the curse of dimensionality?" in *Scientific and Statistical Database Management*, ser. Lect. Notes Comput. Sc., M. Gertz and B. Ludäscher, Eds., vol. 6187. Springer, 2010, pp. 482–500.

[29] T. Weise, M. Zapf, R. Chiong, and A. Nebro, "Why is optimization difficult?" in *Nature-Inspired Algorithms for Optimisation*, ser. Stud. Comput. Intell., R. Chiong, Ed. Springer, 2009, vol. 193, pp. 1–50.

[30] O. Mersmann, M. Preuss, and H. Trautmann, "Benchmarking evolutionary algorithms: Towards exploratory landscape analysis," in *Proceedings of Parallel Problem Solving from Nature (PPSN XI)*, ser. Lect. Notes Comput. Sc., vol. 6238. Springer, 2010, pp. 73–82.

[31] Y. Borenstein and R. Poli, "Information landscapes," in *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO2005)*. ACM, June 2005, pp. 1515–1522.

[32] N. Hansen, A. Auger, S. Finck, and R. Ros, "Real-parameter black-box optimization benchmarking BBOB-2010: Experimental setup," INRIA, Tech. Rep. RR-7215, September 2010.

[33] L. Devroye, *Non-Uniform Random Variate Generation*. Springer, 1986.

[34] C. Igel, T. Suttorp, and N. Hansen, "A computational efficient covariance matrix update and a (1+1)-CMA for evolution strategies," in *GECCO '06: Proceedings of the 8th annual conference on Genetic and evolutionary computation*. New York, NY, USA: ACM, 2006, pp. 453–460.

[35] I. Trelea, "The particle swarm optimization algorithm: convergence analysis and parameter selection," *Inform. Process Lett.*, vol. 85, no. 6, pp. 317–325, 2003.

[36] R. Plackett, "Karl pearson and the chi-squared test," *Int. Stat. Rev.*, vol. 51, no. 1, pp. 59–72, 1983.

[37] S. Finck, N. Hansen, R. Ros, and A. Auger, "Real-parameter black-box optimization benchmarking 2010: Presentation of the noiseless functions," University of Applied Science Vorarlberg and Université-Paris Sud, Tech. Rep., 2010.