



# Quantitative measure of nonconvexity for black-box continuous functions

Kenichi Tamura<sup>a,\*</sup>, Marcus Gallagher<sup>b</sup>

<sup>a</sup> Graduate School of Systems Design, Tokyo Metropolitan University, 1-1 Minamiosawa, Hachioji, Tokyo 192-0397, Japan

<sup>b</sup> School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane 4072, Australia

## ARTICLE INFO

### Article history:

Received 7 April 2018

Revised 20 September 2018

Accepted 7 October 2018

Available online 9 October 2018

### Keywords:

Fitness landscape analysis

Metaheuristics

Black-box functions

Large problems

Complexity

Monte Carlo integral

## ABSTRACT

Metaheuristic algorithms usually aim to solve nonconvex optimization problems in black-box and high-dimensional scenarios. Characterizing and understanding the properties of nonconvex problems is therefore important for effectively analyzing metaheuristic algorithms and their development, improvement and selection for problem solving. This paper establishes a novel analysis framework called *nonconvex ratio analysis*, which can characterize nonconvex continuous functions by measuring the degree of nonconvexity of a problem. This analysis uses two quantitative measures: the *nonconvex ratio* for global characterization and the *local nonconvex ratio* for detailed characterization. Midpoint convexity and Monte Carlo integral are important methods for constructing the measures. Furthermore, as a practical feature, we suggest a rapid characterization measure that uses the local nonconvex ratio and can characterize certain black-box high-dimensional functions using a much smaller sample. Throughout this paper, the effectiveness of the proposed measures is confirmed by numerical experiments using the COCO function set.

© 2018 The Authors. Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license.

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

## 1. Introduction

Metaheuristic algorithms usually aim to solve nonconvex optimization problems in high-dimensional and black-box scenarios [20]. In general, nonconvexity accommodates much more complex structure than convexity. For this reason, various benchmark functions with nonconvexity have been used to analyze and test metaheuristic algorithms in their development, improvement and selection. Characterizing and understanding the properties of nonconvex problems is therefore important for effectively analyzing metaheuristic algorithms and improving our understanding of which algorithms are more suitable for which problem classes and why. This information leads to the development of new and improved algorithms and facilitates algorithm selection for problem solving.

The structures of nonconvex problem landscapes have typically been characterized intuitively, e.g., “big valley,” “multi-funnel,” “deceptive,” and “needle-in-a-haystack” [10]. These structures are based on our experience of landscape geometry in 1, 2 or 3 dimensions. Even complex nonconvex benchmark functions defined on generalized spaces, such as the COCO function set [3], were classified based on such characteristics and documented using one- or two-dimensional slices through the landscape (e.g., contour [3]).

\* Corresponding author.

E-mail addresses: [tamu@tmu.ac.jp](mailto:tamu@tmu.ac.jp) (K. Tamura), [marcusg@itee.uq.edu.au](mailto:marcusg@itee.uq.edu.au) (M. Gallagher).

However, metaheuristic algorithms are expected to work well in high-dimensional spaces. In general, nonconvex problems with dependencies among variables (nonseparability) become more complex and difficult to solve as the number of dimensions increases [23]. Furthermore, certain aspects of the geometry of high-dimensional spaces are not well understood. Thus, landscape features designed to capture structure based on intuition may not be sufficient to represent the diverse nonconvex structure of high-dimensional complex functions. This limitation motivates the development of new landscape features. More effective features can then be used to support algorithm selection or improvement informed by problem characteristics.

Metaheuristic algorithms are often used in black-box scenarios. In practice, this means that landscape features must be derived using only sampled information. Features that can be efficiently and reliably estimated based on small sample sizes are desirable. For instance, sample sizes ranging from  $10^2n$  to  $10^3n$  (where  $n$  is the problem dimension) are used and assumed to be 1 to 10% of the total evaluation budget available for the algorithm [12].

In fitness landscape analysis, a number of features have been developed for black-box functions, mostly on discrete domains [15]. Correlation length, calculated from a random walk autocorrelation function, measures the ruggedness of a landscape [19,24]. The fitness distance correlation [5] quantifies the relation between the fitness of a collection of solutions and their distances to the global optimum. Information content, which is an entropic measure of the time series by a random walk, characterizes the ruggedness of the landscape [22]. Partial information content relates to the modality encountered on the landscape path sampled by a random walk [22]. Information stability is the highest fitness difference between neighboring points reached in the time series by a random walk on a landscape. These techniques, however, have been applied to continuous domain functions by changing the distances used [9,10,12]. In addition, certain function analysis measures originate from continuous domains. The dispersion evaluates the average distance between pairs of high-quality solutions in a sample [7]. The length scale, which is a local measure of how much the objective function changes with respect to the distance between two sampled solutions, and its density entropy called the length scale entropy is used for function characterization [9].

It is possible for a measure to fail to characterize and classify target functions distinctively. In such cases, we can apply multiple measures and use machine learning techniques to analyze the resultant data of the measures. Examples applied to the COCO function set are as follows. In [9], the above six measures—the correlation length, the dispersion, the fitness distance correlation, the information content, the partial information content, and the information stability—were used with the t-SNE algorithm. This framework, called a feature ensemble method, was originally proposed in [18]. In [8], many measures relating convexity, level set, local search, curvature and so on were used with a random forest algorithm and a multi-objective optimization method. These approaches can work well, but the procedures tend to be complicated.

This paper establishes a novel and simple analysis framework, called *nonconvex ratio analysis*, which characterizes nonconvex continuous functions in high-dimensional and black-box scenarios by measuring the degree of nonconvexity. This method can be carried out using a sample of points from a problem landscape. This analysis is based on two measures defined in Section 2: the *nonconvex ratio* for global characterization and the *local nonconvex ratio* for detailed characterization of a problem landscape. In constructing the measures, midpoint convexity and Monte Carlo integral theory play important roles. Using midpoint convexity is a novel idea that can determine nonconvexity more easily than the original definition of convexity. Monte Carlo integral theory [14] contributes to the efficiency and effectiveness of the methodology. The effectiveness and efficiency of characterization are confirmed in Section 3 with numerical experiments on the COCO function set. Through the experiments, interesting characteristics related to nonconvexity are revealed. In Section 4, nonconvex ratio analysis is compared with the above landscape analysis methods based on both the classification results for the COCO functions and their definitions, which are shown in Appendix A. Furthermore, in Section 5, potential applications of nonconvex ratio analysis are discussed on a general level. Then, as a practical technique useful for the applications, we suggest a rapid characterization measure using the local nonconvex ratio, which can characterize certain black-box high-dimensional functions with much smaller samples. The efficiency and effectiveness of the proposed rapid measure is observed for functions in the COCO set. In Section 6, we conclude by summarizing the contributions of the paper and noting future research directions.

## 2. Definition of nonconvex ratio analysis

In this section, we establish a novel analysis framework, called *nonconvex ratio analysis*, which characterizes nonconvex continuous functions in high-dimensional and black-box scenarios by measuring the degree of nonconvexity. This analysis introduces two measures, the *nonconvex ratio* and the *local nonconvex ratio*, which can evaluate the degree of nonconvexity of a problem from different viewpoints.

### 2.1. Approach for quantitative measurement of nonconvexity

Nonconvexity is one of the most challenging properties in continuous optimization problems. The possibility of multiple local minima and basins of attraction, plateaus, ridges and other related features can lead to (intuitively) very complex landscapes. As a simple example, one-dimensional functions of the form  $f(x) = ax^2 + c \sin(bx)$  can produce diverse landscapes depending on the values of the parameters ( $a$ ,  $b$ ,  $c$ ). Specific examples are shown in Fig. 1, where Function B has higher frequency noise, Function C has higher amplitude noise, and Function D has higher frequency and higher amplitude noises than Function A.

Download English Version:

<https://daneshyari.com/en/article/11008030>

Download Persian Version:

<https://daneshyari.com/article/11008030>

[Daneshyari.com](https://daneshyari.com)