Lamarckian memetic algorithms : local optimum and connectivity structure analysis

Le, Minh Nghia; Ong, Yew-Soon; Jin, Yaochu; Sendhoff, Bernhard

2009

Le, M. N., Ong, Y., Jin, Y. & Sendhoff, B. (2009). Lamarckian memetic algorithms : local optimum and connectivity structure analysis. Memetic Computing, 1(3), 175-190. https://dx.doi.org/10.1007/s12293-009-0016-9

https://hdl.handle.net/10356/147981

https://doi.org/10.1007/s12293-009-0016-9

Noname manuscript No.
(will be inserted by the editor)

Lamarckian Memetic Algorithms: Local Optimum and Connectivity Structure Analysis

Minh Nghia Le · Yew Soon Ong · Yaochu Jin · Bernhard Sendhoff

the date of receipt and acceptance should be inserted later

Abstract Memetic Algorithms (MAs) represent an emerging field that has attracted increasing research interest in recent times. Despite the popularity of the field, we remain to know rather little of the search mechanisms of MAs. Given the limited progress made on revealing the intrinsic properties of some commonly used complex benchmark problems and working mechanisms of Lamarckian memetic algorithms in general non-linear programming, we introduce in this work for the first time the concepts of *local optimum structure* and generalize the notion of neighborhood to *connectivity structure* for analysis of MAs. Based on the two proposed concepts, we analyze the solution quality and computational efficiency of the core search operators in Lamarckian memetic algorithms. Subsequently, the structure of local optimums of a few representative and complex benchmark problems is studied to reveal the *effects of individual learning* on fitness landscape and to gain clues into the success or failure of MAs. The connectivity structure of local optimum for different memes or individual learning procedures in Lamarckian MAs on the benchmark problems is also investigated to understand the *effects of choice of memes* in MA design.

Keywords Memetic Algorithms · Lamarckian Evolution · Search Dynamics · Fitness Distance Correlation · Experimental Analysis · Numerical Optimization

M. N. Le · Y. S. Ong
Centre for Computational Intelligence, Division of Information System, School of Computer Engineering, Nanyang Technological University, Blk N4, B1a-02, Nanyang Avenue, Singapore 639798. E-mail: {lemi0005, asysong}@ntu.edu.sg

Y. Jin · B. Sendhoff
Honda Research Institute Europe GmbH, Carl-Legien-Strasse 30, 63073 Offenbach/Main, Germany. E-mail: {yaochu.jin,bernhard.sendhoff}@honda-ri.de

Nomenclature

$f(\mathbf{x})$	=	Objective or fitness function
$\mathbf{x}^*$	=	Global optimum
$\mathbf{x}^{(i)}$	=	*i-th* element of vector $\mathbf{x}$
$r^{(t)}(\mathbf{y}\|\mathbf{x})$	=	Conditional probability density function of having offspring $\mathbf{y}$ given parent $\mathbf{x}$ at generation t
$d(\mathbf{x},\mathbf{y})$	=	Euclidean distance $\|\|\mathbf{x}-\mathbf{y}\|\| = \sqrt{\sum_{i=1}^{n}(x_i-y_i)^2}$ between $\mathbf{x}$ and $\mathbf{y}$
$\mathbf{\Psi}$	=	A set of local optimums
B_v	=	Basin of attraction of local optimum $\mathbf{v}$
$p_v(\mathbf{x})$	=	Probability of converging to local optimum $\mathbf{v}$ from $\mathbf{x}$ by means of individual learning
$T(\mathbf{x},\mathbf{y})$	=	Probability of converging to local optimum $\mathbf{y}$ from $\mathbf{x}$ by means of reproduction and individual learning
$C(\mathbf{x}',\mathbf{x}'')$	=	Computational effort incurred to arrive at $\mathbf{x}''$ from $\mathbf{x}'$ by means of individual learning
$E[.\|P]$	=	Expectation of a measure conditioned to population P
C_v	=	Maximum computational effort required to converge to local optimum starting from any point within the basin of attraction B_v
n	=	Number of dimensions
N	=	Population size
$S(.)$	=	Selection operator
$R(.)$	=	Reproduction operator
$IL(.)$	=	Individual learning operator

1 Introduction

Memetic Algorithms (MA) represent an emerging field that has attracted increasing research interest in recent times, with a growing number of publications appearing in a plethora of international journals and conference proceedings. In the last five years, hundreds of publications related to the field have been listed in the *ISI Web of Knowledge* and *Scopus* databases. Within the field of computational intelligence and soft computing, special issues, special sessions and tutorials in established journals and conferences have also been dedicated to the field (1–5).

The earliest form of Memetic Algorithms (6–8) was first introduced as a marriage between population-based global search and heuristic learning, where the latter is often referred to as a meme, capable of local refinement. They are often regarded as population-based meta-heuristic search approaches, inspired by Darwin's theory of natural evolution and Dawkins' notion of meme defined as a unit of information that reproduces itself while people exchange ideas. This instantiation of MAs has materialized into the form of a hybrid global-local approach that facilitates both exploration and exploitation in the search. Up to date, many MAs have been crafted for solving real-world problems more efficiently. Such hybrid algorithms have been used successfully to solve a wide variety of engineering design problems and often shown to generate higher quality solutions more efficiently than canonical evolutionary algorithms (9–16). So far, most researchers have been concentrated on improving the algorithmic aspects of MAs (17–23). For a discussion on the different depictions of MAs inspired from Dawkins's theory of Universal Darwinism, the reader is referred to (24).

In the literature, two basic forms of individual learning schemes are often discussed, namely, *Lamarckian* and *Baldwinian* learning (25). Lamarckian learning forces the genotype to reflect the result of improvement in local search by placing the locally improved individual back into the population to compete for reproduction. Baldwinian learning, on the other hand, only alters the fitness of the individuals and the improved genotype is not encoded back into the population. It is worth noting the significant research effort in memetic algorithms is mainly spent on crafting specialized algorithms for solving a specific problem or a set of problems via empirically investigations. The reported empirical results usually demonstrate that the proposed algorithm works well on a problem of interest or a relatively small set of problems under investigation. Sometimes, the reported results may not be easily reproducible due to minute differences in the implementations that are often omitted in the published manuscripts. Despite the extensive research efforts in the field, we remain to know rather little of the mechanisms responsible for the success of MAs. Particularly, there is a lack of rigorous studies to provide understandings on the search mechanisms of MAs, limiting their credibility even though the proposed memetic algorithm shows statistical significant improvements over existing methods.

In the last two decades, a few studies to enhance the understandings of Baldwinian MAs (26–29) have been reported. However, to the best of our knowledge little progress on study of Lamarckian MA in non-linear programming has been made so far (30–32) with few exceptions (33). It is worth highlighting that to date the majority of memetic algorithms that have experienced great success on real-world problems adopt the Lamarckian learning concept in their design. Given the limited progress made on revealing the intrinsic properties of commonly used complex benchmark problems and working mechanisms of Lamarckian memetic algorithms in general non-linear programming, we introduce in this paper the concepts of *local optimum structure* and generalize the notion of neighborhood to *connectivity structure* for analysis of MAs. Based on the two proposed concepts, we analyze the solution quality and computational efficiency of the core search operators in Lamarckian memetic algorithms. For the first time, a systematic study on the local optimum structure of a few commonly used benchmark problems for continuous optimization (34) is subsequently performed to reveal the *effects of individual learning* on fitness landscape to gain clues into the success or failure of MAs. To bring about new insights into the *choice of memes* on Lamarckian MA design, we proceed further to investigate on the search mechanisms of two unique individual learning procedures in Lamarckian MAs on the benchmark problems using the concept of local optimum connectivity. These empirical studies thus serve as our initial effort to address the lack of analysis of the benchmark problems commonly used in the literature.

The paper is organized in the following manner. In Section 2, two quantitative performance metrics for unveiling search dynamics of MA, the progress rate and computational cost, are introduced and described. The stochastic selection operator of MA is analyzed based on the concept of *local optimum structure*, and reproduction and individual learning operators is also studied based on *connectivity analysis*. Case studies on the local optimum structure of five continuous optimization benchmark problems of diverse properties are presented in Section 3. Subsequently, the connectivity analysis of individual learning procedures or memes in MAs is presented in Section 4. Finally, a brief conclusion of the paper is provided in Section 5.

2 Lamarckian Memetic Algorithm

In modern stochastic optimization, Lamarckian memetic algorithms have manifested as population-based meta-heuristic search methods inspired by both Darwinian principles of natural evolution, Dawkins notion of a meme as a unit of cultural evolution, and Lamarck's theory of evolution. Although Lamarck's theory of evolution has generated controversies and doubts from biology, the potentials and contributions of Lamarckian learning in computational evolutionary systems have been significant (35; 36). It is worth emphasizing that most successful MAs to date are designed in the spirit of Lamarckian learning. In diverse contexts, Lamarckian memetic algorithms have also been used under the name of hybrid evolutionary algorithm, Lamarckian evolutionary algorithm, cultural algorithm or genetic local search. An outline of a basic Lamarckian MA is provided in Algorithm 1, where we can see that besides the evolutionary operators, i.e., selection, reproduction, an individual learning phase (line 7) is included to refine the individuals. Note that Lamarckian learning (line 8) allows the genotype to reflect the result of refinement through placing the improved individual back into the population to compete for reproductive opportunities (10; 13; 22).

Algorithm 1 Lamarckian Memetic Algorithm Search

```
1:  Generate an initial population
2:  while Stopping conditions are not satisfied do
3:    Evaluate all individuals in the population
4:    Select individuals for the reproduction pool Ω via selection operator S(.)
5:    for each individual in Ω do
6:      Evolve an offspring according to reproduction operators R(.)
7:      Perform Individual Learning via learning operator IL(.)
8:      Proceed in the spirit of Lamarckian learning
9:    end for
10: end while
```

2.1 Performance Metrics for Analysis of Memetic Algorithm

In this subsection, we define some metrics for measuring the search dynamics of Lamarckian MAs. In particular, we introduce two quantitative performance metrics, namely, progress rate and computational cost, for unveiling the effectiveness (solution quality) and efficiency (computational cost) of MA search, respectively. Progress rate describes the distance dynamics of the population and thus of the search in approaching the global optimum. The second performance metric, i.e., computational cost, is introduced to complement the progress rate since the former does not take into account the effort required to attain the progression of solution quality. Progress rate and computational cost together will serve as the core metrics used in this paper for a comprehensive study on the working mechanisms of Lamarckian MAs. In addition, we concentrate on the general nonlinear programming problem of the following form:

$$\textbf{Minimize/Maximize}: \quad f(\mathbf{x}), \qquad (1)$$
$$\textbf{Subjectto}: \mathbf{x}_l^{(i)} \leq \mathbf{x}^{(i)} \leq \mathbf{x}_u^{(i)}, \forall i = 1 \ldots n$$

where $\mathbf{x} \in R^n$ is the vector of design variables, and $\mathbf{x}_l$, $\mathbf{x}_u$ are vectors of lower and upper bounds, respectively.

2.1.1 Progress Rate

Progress rate, notated here as θ, provides a measure on the improvement in solution quality during the evolutionary search (37; 38). A consistent positive progress rate of several generations suggests that the evolution (as guided by the memetic algorithm) is progressing well towards the global optimum of the problem at hand. Clearly, a negative progress rate implies the search diverges away from the global optimum.

Let $\mathbf{x}^*$ denote the global optimum solution of the problem. The progress rate θ at population state $P = \{\mathbf{x}_k\}_{k=1}^N$ of size N defined as the expected change in *distance to global optimum* of the population as a result of the *stochastic* evolutionary operator(s) is given as

$$\theta = D - E[D'|P], \qquad (2)$$
$$D = \sum_{k=1}^{N} d(\mathbf{x}_k, \mathbf{x}^*)/N, \qquad D' = \sum_{k=1}^{N} d(\mathbf{x}_k^{'}, \mathbf{x}^*)/N$$

where $P' = \{\mathbf{x}_k^{'}\}_{k=1}^N$ is the next population, D and D' denote the distance to global optimum of the current and next populations, respectively.

2.1.2 Computational Cost

When studying the dynamics of an algorithm, besides the progress rate of a search, the other important issue is how fast the algorithm in finding the global optimum. Thus computational cost represents another core metric that is crucial for studying the performance of Lamarckian MA. Considering the selection, reproduction and individual learning components of a Lamarckian memetic algorithm, it is noted that the computational effort incurred by the two former evolutionary operators do not vary much throughout the entire search. Further, among the three search operators, the individual learning phase consumes the most computational effort, while selection operator incurs the least.

2.2 Local Optimum Structure and Connectivity Analysis

In continuous parametric domain, the solution of interest lies in a stationary point $\mathbf{x}$, where $\nabla F(\mathbf{x}) = 0$ or $\nabla F(\mathbf{x}) < \epsilon$, with ϵ denoting some arbitrarily small value. Stationary points exist in the form of minima, maxima or saddle points. In Lamarckian MAs, individual learning procedures based on numerical solvers[1] have been well established to generate precise stationary points efficiently (i.e., 1^{st} order necessary condition). Practically, if $\mathbf{\Psi}$ is the complete local optimum set of a problem, all subsequent populations of Lamarckian MA search, except the initial population, are subsets of $\mathbf{\Psi}$. The study on the search dynamics of Lamarckian MAs thus evolves around the set of local optimum solutions. For this reason, it is crucial to uncover the properties and characteristics of the local optimum set $\mathbf{\Psi}$ when analyzing Lamarckian MAs. In the rest of this section, we study the search mechanisms of Lamarckian MAs on general non-linear optimization problems by analyzing the *local optimum structure* and *connectivity properties* of the set $\mathbf{\Psi}$.

Lamarckian memetic algorithms are composed of three core operators, namely, selection (line 4), reproduction and individual learning (lines 6 and 7), see Algorithm 1. Here we study the working mechanisms of the algorithms by analyzing the progress rate and computational cost of these core operators. In what follows, Section 2.2.1 presents the study on stochastic selection operator using *local optimum structure*, and the study on reproduction and individual learning operators using *connectivity analysis* is subsequently presented in Section 2.2.2.

2.2.1 Local Optimum Structure

The search dynamics and performance of Lamarckian MAs depend on the properties of the problem fitness landscape. Since the search operators of Lamarckian MAs evolve around the set of local optimum solutions $\mathbf{\Psi}$, local optimum structure or the distribution of local optimum solutions represents one of the key property of the problem landscape to bring new insights into the search algorithm. Before studying on the role of stochastic selection operator in Lamarckian MAs, we introduce the term of *constructive* and *obstructive* local optimum structure and outline some basic definitions and theorems used in our analysis.

Let $P = \{\mathbf{x}_k \in \mathbf{\Psi}\}_{k=1}^N$ denotes the initial population of N locally optimum individuals and $P_S = S(P) = \{\mathbf{x}_k^{'} \in \mathbf{\Psi}\}_{k=1}^N$ denotes the resultant population of optimum individuals after undergoing the stochastic selection operation.

[1] These numerical solvers are commonly categorized into three groups according to the type of derivatives used in the search (39): (1) Zero-order methods, also known as direct search techniques, use the objective function values or the relative rank of objective values in the move operator to decide on the search structure. (2) First-order methods employ both function values and first (partial) derivative vector to decide its moves. (3) Second-order Newton methods, on the other hand, make use of the objective function values, first derivative vector as well as the second derivative Hessian matrix. However, due to the difficulty of obtaining Hessian matrix, a class of methods named as *Quasi-Newton* attempts to approximate Hessian matrix through the first-order derivatives. Davidon-Fletcher-Powell (DFP) and Broyden-Fletcher-Goldfarb-Shanno (BFGS) are two examples of Quasi-Newton methods.

Definition 1: The selection progress rate of the search (θ_S) is defined as the expected change in *distance to global optimum* as a result of the selection operator $S(.)$.

$$\begin{aligned}\theta_S(P) &= D - E[D'|P] \\ &= \frac{1}{N} \times \sum_{k=1}^{N} d(\mathbf{x}_k, \mathbf{x}^*) - \frac{1}{N} \times \sum_{k=1}^{N} E[d(\mathbf{x}_k^{'}, \mathbf{x}^*)|P]\end{aligned} \tag{3}$$

Definition 2: A *'constructive'*/ *'obstructive'* local optimum structure exhibits a property that the *'better'*/ *'poorer'* fitness quality a local optimum has, the *closer* it is to the global optimum.

Let $\mathbf{\Psi} = \{\mathbf{x}_1, \mathbf{x}_2, \dots \mathbf{x}_M\}$, where $|\mathbf{\Psi}| = M$, denote the finite set of local optimum solutions. Let d_i and f_i be the abbreviations for $d(\mathbf{x}_i, \mathbf{x}^*)$ and $f(\mathbf{x}_i)$, respectively. Without loss of generality, members of set $\mathbf{\Psi}$ are sorted in an ascending order of distance to the global optimum $d(\mathbf{x}_i, \mathbf{x}^*)$ such that $0 < d_1 \le d_2 \le d_3 \dots \le d_M$. For a maximization problem, our definition implies $f_1 \ge f_2 \ge f_3 \dots \ge f_M$ for *constructive* local optimum structure and $f_1 \le f_2 \le f_3 \dots \le f_M$ for *obstructive* correlated structure of local optimums.

Examples of problems with a fitness landscape imbued with properties of *constructive* and *obstructive* correlated local optimum structure are depicted in Figures 1(a) and 1(b), respectively.

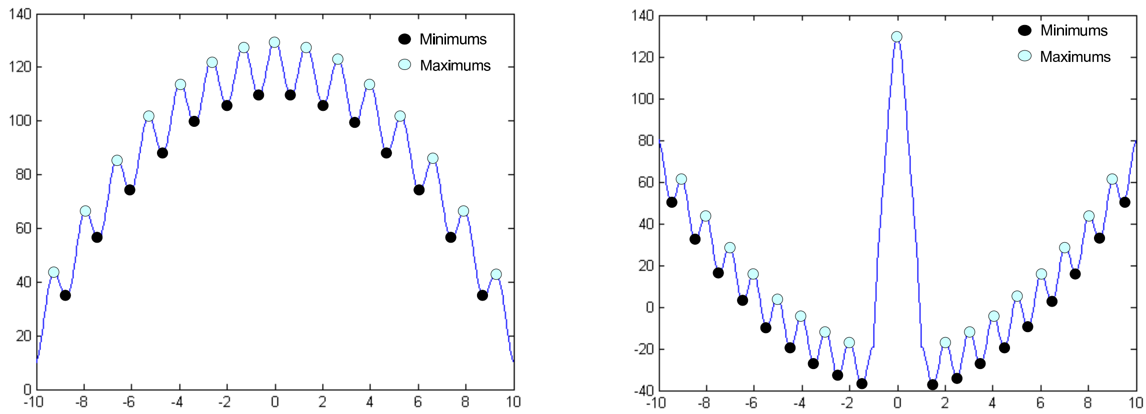
Chebyshev's sum inequality theorem: if two sequences, $\{f_j\}_{j=1}^N$ and $\{d_k\}_{k=1}^N$, are of the same order, $f_1 \le f_2 \le \dots \le f_N$ and $d_1 \le d_2 \le \dots \le d_N$, then

$$(\sum_{j=1}^{N} f_j)(\sum_{k=1}^{N} d_k) \le N \times \sum_{u=1}^{N} f_u d_u$$

Similarly, if two sequences are of an inverted order, $f_1 \ge f_2 \ge \dots \ge f_N$ and $d_1 \le d_2 \le \dots \le d_N$, then

$$(\sum_{j=1}^{N} f_j)(\sum_{k=1}^{N} d_k) \ge N \times \sum_{u=1}^{N} f_u d_u$$

In what follows, we present our analysis on the search mechanisms of the stochastic selection operator in Lamarckian MAs based on the local optimum structure of a problem landscape. Considering, for example, the stochastic fitness-proportional selection scheme (40), which represents one of the commonly used selection operators in evolutionary computation. The probability of individual $\mathbf{x}_k \in P = \{\mathbf{x}_k\}_{k=1}^N$ to survive in the reproduction pool of a Lamarckian MA with stochastic fitness-proportional selection operator can be derived as $\frac{f_k}{\sum_{j=1}^{N} f_j}$. The selection progress rate then be-



(a) Problem Landscape with Constructive Local Optimum Structure

(b) Problem Landscape with Obstructive Local Optimum Structure

Fig. 1 Illustrations of ‘constructive’/ ‘obstructive’ landscapes in maximization problem

comes

$$\begin{aligned}\theta_S(P) &= \frac{\sum_{k=1}^{N} d_k}{N} - \frac{1}{N} \times \sum_{k=1}^{N} E[d(\mathbf{x}_k^{'}, \mathbf{x}^*)|P] \\ &= \frac{\sum_{k=1}^{N} d_k}{N} - \frac{1}{N} \times N \times (\frac{\sum_{u=1}^{N} f_u d_u}{\sum_{j=1}^{N} f_j}) \\ &= \frac{\sum_{k=1}^{N} d_k}{N} - \frac{\sum_{u=1}^{N} f_u d_u}{\sum_{j=1}^{N} f_j} \\ &= \frac{(\sum_{j=1}^{N} f_j)(\sum_{k=1}^{N} d_k) - N \times \sum_{u=1}^{N} f_u d_u}{N \sum_{j=1}^{N} f_j} \end{aligned} \quad (4)$$

Without loss of generality, based on Equation (4), it is possible to infer the progress rate of the stochastic selection operator on maximization problems with fitness landscape containing a constructive or obstructive local optimum structure using the Chebyshev’s sum inequality theorem. For a maximization problem fitness landscape containing a *constructive* correlated local optimum structure, the selection progress rate is thus positive, i.e., $\theta_S(P) \geq 0, \forall P$,

Proof: Note that on constructive local optimum structure, two sequences $\{f_j\}_{j=1}^N$ and $\{d_k\}_{k=1}^N$ are of an inverted order. According to the Chebyshev’s sum inequality theorem, we have

$$\begin{aligned}&(\sum_{j=1}^{N} f_j)(\sum_{k=1}^{N} d_k) \geq N \times \sum_{u=1}^{N} f_u d_u \\ &(\sum_{j=1}^{N} f_j)(\sum_{k=1}^{N} d_k) - N \times \sum_{u=1}^{N} f_u d_u \geq 0\end{aligned} \quad (5)$$

As the left-hand side of Inequality (5) is the numerator of $\theta_S(P)$, refer to Equation (4), we have $\theta_S(P) \geq 0$

Conversely, for a maximization problem fitness landscape containing an *obstructive* correlated local optimum structure, the selection progress rate is thus shown negative, i.e., $\theta_S(P) \leq 0, \forall P$,

Proof: Note that on obstructive local optimum structure, two sequences $\{f_j\}_{j=1}^N$ and $\{d_k\}_{k=1}^N$ are of the same order. According to the Chebyshev’s sum inequality theorem, we have

$$\begin{aligned}&(\sum_{j=1}^{N} f_j)(\sum_{k=1}^{N} d_k) \leq N \times \sum_{u=1}^{N} f_u d_u \\ &(\sum_{j=1}^{N} f_j)(\sum_{k=1}^{N} d_k) - N \times \sum_{u=1}^{N} f_u d_u \leq 0\end{aligned} \quad (6)$$

As the left-hand side of Inequality (6) is the numerator of $\theta_S(P)$, refer to Equation (4), we have $\theta_S(P) \leq 0$.

To summarize, the stochastic selection operator is shown to have an effect of “pulling” the MA population towards the global optimum on problem landscapes imbued with a *constructive* correlated local optimum structure, thus advancing the search towards the global optimum. In contrast, the stochastic selection operator exhibits an effect of “pushing” the population away from the global optimum on problem landscapes having an *obstructive* correlated optimum structure.

2.2.2 Local Optimum Connectivity

The previous section has focused on the analysis of the MA selection operator based on the notions of correlated optimum structure and selection progress rate θ_S. In what follows, we proceed to study the synergy between the reproduction and individual learning operators of Lamarckian MAs based on the properties of the local optimum connectivity. We begin with introducing the notion of *'constructive'/ 'obstructive'* local optimum connectivity, and then outline some definitions used in our analysis.

Let $\mathbf{x}^{'}$ denote the offspring of $\mathbf{x}$ through reproduction, i.e., $\mathbf{x}^{'} = R(\mathbf{x})$, and $\mathbf{x}^{''}$ denote the resulting offspring of $\mathbf{x}$ after undergoing reproduction and individual learning operators, i.e., $\mathbf{x}^{''} = IL(R(\mathbf{x}))$. Further,

$$E[\triangle d(\mathbf{x},\mathbf{x}^{''})|P] = E[d(\mathbf{x},\mathbf{x}^{*}) - d(\mathbf{x}^{''},\mathbf{x}^{*})|P]$$

represents the expected change in distance to global optimum of individual $\mathbf{x}$.

Definition 4: Progress rate θ_{R-IL} and computational cost C are defined for reproduction & individual learning operators as the expected change in *distance to global optimum* and *incurred computational cost*, respectively.

Definition 5: Local optimum $\mathbf{u}$ is connected to local optimum $\mathbf{v}$ if and only if $\mathbf{v}$ is reachable from $\mathbf{u}$, i.e., the probability of $\mathbf{v} = IL(R(\mathbf{u}))$ is non-zero, or $P(\mathbf{v} = IL(R(\mathbf{u}))) > 0$.

Definition 6: The connectivity of a local optimum $\mathbf{x}$ is *constructive* if $E[\triangle d(\mathbf{x},\mathbf{x}^{''})|P] \geq 0$ and *obstructive* if $E[\triangle d(\mathbf{x},\mathbf{x}^{''})|P] < 0$. In other words, a *'constructive'/ 'obstructive'* connectivity indicates that a local optimum connects at a high probability to other local optimum solutions of *'improved'/ 'inferior'* quality.

Examples on *constructive* and *obstructive* connectivity of local optimum solutions are depicted in Figure 2(a) and 2(b), respectively. Note that the 'solid'/ 'dotted' lines in the figures refers to the 'high'/ 'low' probability connections of local optimums.

Here, the connectivity of local optimum solutions is modeled as a directed graph $\mathbf{G} = (\mathbf{V},\mathbf{E})$. Vertex $\mathbf{V}$ represents the local optimum solutions (i.e. $\mathbf{V} = \mathbf{\Psi}$). A directed edge $e_{\mathbf{u},\mathbf{v}}$ represents that $\mathbf{v}$ is reachable from $\mathbf{u}$ via the reproduction $R(.)$ and individual learning $IL(.)$ operators. Figure 3 depicts possible connections between a local optimum $\mathbf{u}$ to other local optimum solutions of a improved or inferior quality, as illustrated by vertex $\mathbf{v}_1, \mathbf{v}_2$ and $\mathbf{v}_3, \mathbf{v}_4$, respectively.

Weight $w_{\mathbf{u},\mathbf{v}}$ of a directed edge $e_{\mathbf{u},\mathbf{v}}$ is defined as the transition probability $T(\mathbf{u},\mathbf{v})$ of reaching $\mathbf{v}$ from $\mathbf{u}$, which is given by

$$w_{\mathbf{u},\mathbf{v}} = T(\mathbf{u},\mathbf{v}) = \int_{\mathbf{z}} r^{(t)}(\mathbf{z}|\mathbf{u}) \times p_{\mathbf{v}}(\mathbf{z})d\mathbf{z}, \quad (7)$$

where $r^{(t)}(\mathbf{z}|\mathbf{u})$ denotes the conditional probability density function of the offspring $\mathbf{z} = R(\mathbf{u})$ and $p_{\mathbf{v}}(\mathbf{z})$ denotes the probability of $\mathbf{v} = IL(\mathbf{z})$. From this point onwards, time index t of $r^{(t)}(\mathbf{y}|\mathbf{x})$ is omitted for the sake of conciseness. Note that if the *basin of attraction* $B_{\mathbf{v}}$ of local optimum $\mathbf{v}$ is defined as $B_{\mathbf{v}} = \{\mathbf{z} : p_{\mathbf{v}}(\mathbf{z}) > 0\}$, as $p_{\mathbf{v}}(\mathbf{z}) \in \{1, 0\}$

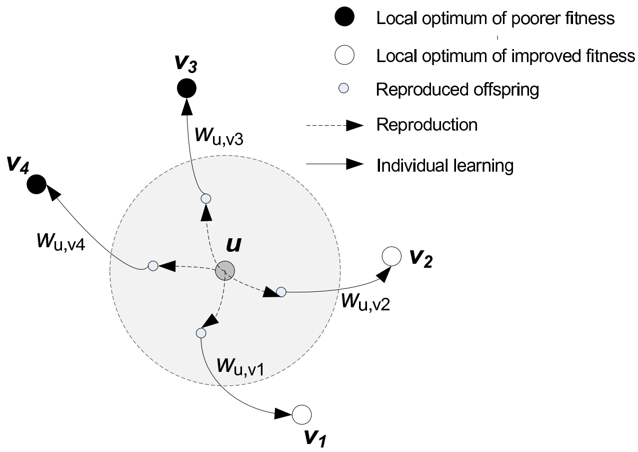


Fig. 3 Connectivity of local optimums

for deterministic individual learning methods, Equation (7) becomes

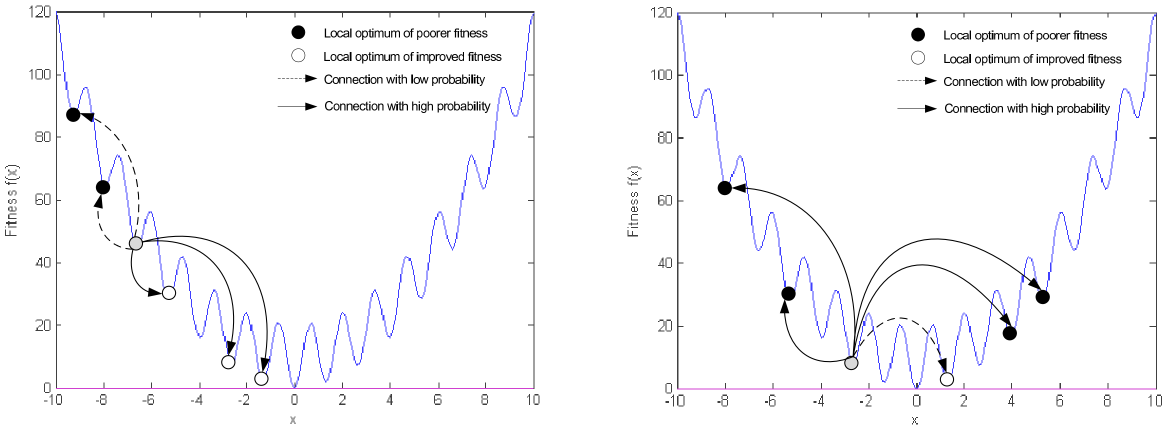
$$w_{\mathbf{u},\mathbf{v}} = T(\mathbf{u},\mathbf{v}) = \int_{B_v} r(\mathbf{z}|\mathbf{u})d\mathbf{z} \quad (8)$$

i.e., the transition probability $T(\mathbf{u},\mathbf{v})$ is an integration of $r(\mathbf{z}|\mathbf{u})$ across the basin of attraction $B_{\mathbf{v}}$.

In the following, we present our analysis on the search mechanisms of the reproduction & deterministic individual learning operators in Lamarckian MAs using the concept of local optimum connectivity. Through reproduction and individual learning, a population $P = \{\mathbf{x}_k\}_{k=1}^{N}$ would become $P^{''} = \{\mathbf{x}_k^{''}\}_{k=1}^{N}$. Note that the computational effort $C(\mathbf{x},\mathbf{x}^{''})$ required to arrive at $\mathbf{x}^{''}$ from $\mathbf{x}$ for a given precision ϵ can be defined as $\alpha + C(\mathbf{x}^{'},\mathbf{x}^{''})$ [2]. The expected progress rate θ_{R-IL} and computational cost C for population $P = \{\mathbf{x}_k\}_{k=1}^{N}$ can then be derived as

$$\begin{aligned}\theta_{R-IL} &= D - E[D''|P]\\ &= \frac{1}{N} \times \sum_{k=1}^{N} d(\mathbf{x}_k,\mathbf{x}^{*}) - \frac{1}{N} \times \sum_{k=1}^{N} E[d(\mathbf{x}_k^{''},\mathbf{x}^{*})|P]\\ &= \frac{1}{N} \times \sum_{k=1}^{N} E[d(\mathbf{x}_k,\mathbf{x}^{*}) - d(\mathbf{x}_k^{''},\mathbf{x}^{*})|P]\\ &= \frac{1}{N} \times \sum_{k=1}^{N} E[\triangle d(\mathbf{x}_k,\mathbf{x}_k^{''})|P] \quad (9)\end{aligned}$$

[2] The reproduction operator for every offspring incurs a small computational effort denoted by α.



(a) Constructive Connectivity of Local Optimums

(b) Obstructive Connectivity of Local Optimums

Fig. 2 Illustrations of 'constructive'/ 'obstructive' connectivity in minimization problem

and

$$
\begin{aligned}
C &= \frac{1}{N} \times \sum_{k=1}^{N} E[C(\mathbf{x}_k, \mathbf{x}_k^{''})|P] \\
&= \frac{1}{N} \times \sum_{k=1}^{N} E[\alpha + C(\mathbf{x}_k^{'}, \mathbf{x}_k^{''})|P] \\
&= \frac{1}{N} \times \sum_{k=1}^{N} (\alpha + E[C(\mathbf{x}_k^{'}, \mathbf{x}_k^{''})|P]) \\
&= \alpha + \frac{1}{N} \times \sum_{k=1}^{N} E[C(\mathbf{x}_k^{'}, \mathbf{x}_k^{''})|P] \qquad (10)
\end{aligned}
$$

The progress rate θ_{R-IL} and computational cost C of the reproduction and deterministic individual learning operators on the *'constructive'/ 'obstructive'* local optimum connectivity (i.e. $T(\mathbf{x}_k, \mathbf{v})$) are investigated using Equations (9) and (10). In particular, the addend in Equations (9) and (10), denotes the expected individual improvement $E[\triangle d(\mathbf{x}_k, \mathbf{x}_k^{''})|P]$ and cost of learning $E[C(\mathbf{x}_k^{'}, \mathbf{x}_k^{''})|P]$ of individual $\mathbf{x}_k$, are expressed in term of transition probabilities $T(\mathbf{x}_k, \mathbf{v})$ from $\mathbf{x}_k$ to other local optimums $\mathbf{v}$, as given in Equations (11) and (12), respectively. Recall that the transition probability for deterministic individual learning operator as defined in Equation (8) becomes

$$
T(\mathbf{x}_k, \mathbf{v}) = \int_{B_{\mathbf{v}}} r(\mathbf{x}_k^{'}|\mathbf{x}_k) d\mathbf{x}_k^{'}
$$

Thus

$$
\begin{aligned}
&E[\triangle d(\mathbf{x}_k, \mathbf{x}_k^{''})|P] \\
&= \int_{\mathbf{x}_k^{'}} r(\mathbf{x}_k^{'}|\mathbf{x}_k) \times (\sum_{\mathbf{v}\in\mathbf{\Psi}} p_{\mathbf{v}}(\mathbf{x}_k^{'}) \times \triangle d(\mathbf{x}_k, \mathbf{v})) d\mathbf{x}_k^{'} \\
&= \sum_{\mathbf{v}\in\mathbf{\Psi}} \triangle d(\mathbf{x}_k, \mathbf{v}) \times \int_{\mathbf{x}_k^{'}\in B_{\mathbf{v}}} r(\mathbf{x}_k^{'}|\mathbf{x}_k) d\mathbf{x}_k^{'} \\
&= \sum_{\mathbf{v}\in\mathbf{\Psi}} \triangle d(\mathbf{x}_k, \mathbf{v}) \times T(\mathbf{x}_k, \mathbf{v}) \qquad (11)
\end{aligned}
$$

and

$$
\begin{aligned}
&E[C(\mathbf{x}_k^{'}, \mathbf{x}_k^{''})|P] \\
&= \int_{\mathbf{x}_k^{'}} r(\mathbf{x}_k^{'}|\mathbf{x}_k) \times (\sum_{\mathbf{v}\in\mathbf{\Psi}} p_{\mathbf{v}}(\mathbf{x}_k^{'}) \times C(\mathbf{x}_k^{'}, \mathbf{v})) d\mathbf{x}_k^{'} \\
&= \sum_{\mathbf{v}\in\mathbf{\Psi}} \int_{\mathbf{x}_k^{'}} r(\mathbf{x}_k^{'}|\mathbf{x}_k) \times p_{\mathbf{v}}(\mathbf{x}_k^{'}) \times C(\mathbf{x}_k^{'}, \mathbf{v}) d\mathbf{x}_k^{'} \\
&= \sum_{\mathbf{v}\in\mathbf{\Psi}} \int_{\mathbf{x}_k^{'}\in B_{\mathbf{v}}} r(\mathbf{x}_k^{'}|\mathbf{x}_k) \times C(\mathbf{x}_k^{'}, \mathbf{v}) d\mathbf{x}_k^{'} \\
&\leq \sum_{\mathbf{v}\in\mathbf{\Psi}} \max_{\mathbf{x}_k^{'}\in B_{\mathbf{v}}} C(\mathbf{x}_k^{'}, \mathbf{v}) \times \int_{\mathbf{x}_k^{'}\in B_{\mathbf{v}}} r(\mathbf{x}_k^{'}|\mathbf{x}_k) d\mathbf{x}_k^{'} \\
&\leq \sum_{\mathbf{v}\in\mathbf{\Psi}} C_{\mathbf{v}} \times T(\mathbf{x}_k, \mathbf{v}) \qquad (12)
\end{aligned}
$$

where $C_{\mathbf{v}} = \max_{\mathbf{x}_k^{'}\in B_{\mathbf{v}}} C(\mathbf{x}_k^{'}, \mathbf{v})$ is the maximum learning cost at the precision ϵ required by an individual (i.e., offspring) $\mathbf{x}_k^{'}$ in the basin of attraction $B_{\mathbf{v}}$.

For the constructive connectivity property illustrated in Figure 2(a), Equations (11) and (12) suggest that if most of the local optimums $\mathbf{x}_k$ of an optimization problem possess a high probability of connecting to improved quality solutions (i.e., large $T(\mathbf{x}_k, \mathbf{v})$ and positive $\triangle d(\mathbf{x}_k, \mathbf{v})$) at low learning expense $C_{\mathbf{v}}$, a large overall progress rate θ_{R-IL} at small computational cost C is expected. In this case, Lamarckian MAs will search effectively and efficiently towards the global optimum.

On the other hand, for problems imbued with obstructive connectivity where most local optimums are likely to connect to low-quality solutions (i.e., large $T(\mathbf{x}_k, \mathbf{v})$ and negative $\triangle d(\mathbf{x}_k, \mathbf{v})$) at the expenses of high learning costs $C_{\mathbf{v}}$, a limited progress rate can be expected accompanied with a high computational cost, as illustrated in Figure 2(b). In the extreme case, where $\mathbf{x} = IL(R(\mathbf{x})) = \mathbf{x}^{''}$, no search improvement can be achieved (i.e. $\theta_{R-IL} = 0$), which leads to the well-known problem of premature convergence in MA (32).

3 Local Optimum Structure Analysis of Representative Benchmark Problems

In the field of continuous optimization, a number of diverse complex benchmark minimization problems are available (34). Although the set of problems has been used extensively for benchmarking evolutionary (41–43) or memetic algorithms in particular (14; 10; 44; 24; 17), little effort has been made to reveal their properties thus far, particularly the local optimum structure of these benchmark problems.

In this section, we present for the first time a systematic analysis of the local optimum structure of five continuous parametric benchmark problems of diverse properties extensively used in the literature[3]. These benchmarks represent classes of multimodal, 'epistatic'/ 'non-epistatic' optimization problems (10), as summarized in Table 1. Further, we study the impact of individual learning phase in MAs on the problem landscape, i.e., the set of local optimum solutions or the transformed problem landscape.

3.1 Representative Test Problems

3.1.1 Ackley function

Ackley function is symmetric and very bumpy. It has a global minimum located at $(0, \dots, 0)$, of which the minimum fitness is 0. The function is defined as

$$F_{Ackley} = 20 + e - 20e^{-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}} - e^{\frac{1}{n}\sum_{i=1}^{n}\cos(2\pi x_i)}$$
$$\mathbf{x} \in [-32, 32]^n$$

[3] A number of other commonly used test problems from the literature were also investigated. However, due to similarity in the properties of some functions, only representatives of each benchmark problem class are presented here.

Table 1 Benchmark functions used in the study

Func	Range	Characteristics	
		Epistasis	Multi-modality
F_{Ackley}	$[-32, 32]^{10}$	yes	yes
$F_{Rastrigin}$	$[-5, 5]^{10}$	no	yes
$F_{Griewank}$	$[-600, 600]^{10}$	yes	yes
$F_{Weierstrass}$	$[-0.5, 0.5]^{10}$	no	yes
$F_{Rosenbrock}$	$[-100, 100]^{10}$	yes	yes

3.1.2 Rastrigin function

Rastrigin function is also a highly multimodal function with many local minimums and the global minimum located at $(0, \dots, 0)$. The minimum fitness at the optimal solution is 0. It has also a very rugged landscape.

$$F_{Rastrigin} = 10n + \sum_{i=1}^{n}(x_i^2 - 10\cos(2\pi x_i))$$
$$\mathbf{x} \in [-5, 5]^n$$

3.1.3 Griewank function

Griewank function is also a highly multimodal function with many local minimums and the global minimum located at $(0, \dots, 0)$, of which the function's minimum fitness is 0. The function is defined as

$$F_{Griewank} = 1 + \sum_{i=1}^{n} x_i^2/4000 - \prod_{i=1}^{n}\cos(x_i/\sqrt{i})$$
$$\mathbf{x} \in [-600, 600]^n$$

3.1.4 Rosenbrock function

Rosenbrock function is non-separable, with highly correlated decision variables. The function's minimum fitness at the optimal solution $(1, \dots, 1)$ is 0. In the two-dimensional case, the minimum is located in a long, flat-bottomed, curved, and narrow valley. It is defined as

$$F_{Rosenbrock} = \sum_{i=1}^{n-1}(100 \times (x_{i+1} - x_i^2)^2 + (1 - x_i)^2)$$
$$\mathbf{x} \in [-100, 100]^n$$

3.1.5 Weierstrass function

Weierstrass function has its global optimum located at $(0, \dots, 0)$ with a minimum fitness of 0. The function is defined as

$$F_{Weierstrass} = \sum_{i=1}^{n}(\sum_{k=0}^{k_{max}}(a^k\cos(2\pi b^k(x_i + 0.5))))$$
$$-n\sum_{k=0}^{k_{max}}(a^k\cos(\pi b^k))$$
$$a = 0.5, b = 3, k_{max} = 20, \quad \mathbf{x} \in [-0.5, 0.5]^n$$

3.2 Empirical Study

To study the local optimum structure of the benchmark problems, a basic random multi-start local search (individual learning) optimization algorithm is used to search on them [4]. In the experiments, the sets of local optimum solutions for the five problems [5] are attained using the Davidon-Fletcher-Powell (DFP) strategy, which is a class of very popular quasi-Newton method. The stopping criteria of DFP is defined by the Cauchy's convergence test [6] (i.e. $|\mathbf{x}_{n+1} - \mathbf{x}_n| \leq \epsilon$ for $N > N_0$ in which the precision ϵ is set to 1E-8). In each run, a total of m randomly sampled individual learning iterations using the DFP was conducted to arrive at their corresponding local optimums. Here, analysis on the initial $m = 4000$ randomly sampled points and the obtained local optimums was performed to reveal the properties of the original and transformed benchmark problem landscape, respectively, and the effect of individual learning on the search.

From our experimental study, the fitness distance scatter plots and correlation coefficients, i.e., ϱ_L and ϱ, of the 10-dimensional benchmark problems across 30 independent runs are obtained and presented in Figures 4-8 and Table 2, respectively. Fitness distance correlation (FDC) analysis, a common method for revealing the correlated structure of a landscape (45; 46; 30), has been used here. The correlation coefficient measure ϱ of the original fitness landscape can be defined by Equation (13) as the level of correlation between the fitness and distance to global optimum.

$$\varrho(f,d) = \frac{Cov(d,f)}{\sigma(d) \times \sigma(f)} \tag{13}$$

Alternatively, ϱ can be estimated from a set of m solutions $\{(d_i, f_i)\}_{i=1}^{m}$ using Equation (14)

$$\varrho(f,d) \approx \frac{1}{\sigma(f)\sigma(d)} \frac{1}{m} \sum_{i=1}^{m} (f_i - \overline{f})(d_i - \overline{d}), \tag{14}$$

where $\overline{f}$ and $\sigma(f)$ refer to the mean and standard deviation, respectively. Similar notations are used for $\overline{d}$ and $\sigma(d)$. Note that Equation (14) is also used for estimating the statistical measure ϱ_L of the transformed landscape from a set of m local optimums $\{(d_i, f_i)\}_{i=1}^{m}$.

Beside the fitness distance correlation coefficient ϱ, the scatter plot $(f(\mathbf{x}), d(\mathbf{x}, \mathbf{x}^*))$ of fitness and distance to global optimum is also provided here as an effective method for visualizing the correlation property of a problem landscape.

From the scatter plots in Figures 4 and 5, strong constructive correlated structure of local optimums are observed on both Ackley and Rastrigin functions. The left and right panels of the figures show the scatter plot of local optimums and initial sampled points, or the transformed and the original landscapes, respectively. Further, the results in Table 2 indicate a higher mean-variance structure correlation value for the transformed landscape than the original landscape on both Ackley ($\varrho_L = 0.920672 \pm 5.791E-3 > \varrho = 0.780676 \pm 6.539E-3$) and Rastrigin ($\varrho_L = 0.994019 \pm 0.306E-3 > \varrho = 0.719241 \pm 6.925E-3$). This highlights the significant impact of individual learning on the search space, where the phenomenon of reinforcement in fitness distance correlation of the transformed problem landscape can be observed. [7]

On the Griewank function, however, the scatter plot in Figure 6 shows a similar degree of constructive correlation structure on both the original and transformed landscape, i.e., ($\varrho_L = 0.995402 \pm 0.297E-3 \approx \varrho = 0.994486 \pm 0.247E-3$). Although Griewank function has a very rugged landscape, which is often difficult to search, it is to see that the degree of ruggedness decreases with increasing dimension, which may be the reason for the similar level of structure correlations obtained.

From Figure 7 and Table 2, we can see that the constructive correlated structure of local optimums is significantly higher than that of the original landscape on the Rosenbrock function ($\varrho_L = 0.999912 \pm 0.228E-3 \gg \varrho = 0.621473 \pm 9.374E-3$). The high-dimensional (*n=4-30*) Rosenbrock function, which contains only 2 local optimums, was studied empirically in (47) and shown to have more than 2 stationary points. It is worth noting that since the DFP strategy does not distinguish between different types of stationary points, similar to (47), 3 stationary points were found in our work on the 10-dimensional Rosenbrock function.

Last but not least, Weierstrass function possesses a weak constructive correlated structure of local optimums, even though visually the problem landscape appears to be relatively structured. Besides the scatter plot in Figure 8, the resulting *FDC* coefficient ϱ of the original landscape of the Weierstrass function is shown to be higher ($\varrho = 0.877178 \pm 3.749E-3$) than the local optimum structure correlation ($\varrho_L = 0.709783 \pm 8.725E-3$). Hence, Weierstrass function possesses an unique property from the other benchmark functions considered.

The structure correlation profile of the five representative benchmark problems is summarized in Table 2. It is worth noting that most of the problems presented here indicated improvements in the transformed landscape structure correlation over the original one, i.e., $\varrho_L > \varrho$. Hence, based on the proof presented in Section 2.2.1, with the improved structure correlation of the transformed landscape or

[4] The random multi-start local search algorithm is the simplest form of stochastic algorithm that uses a basic random search as a global optimizer. Being an unbiased global optimizer, it allows an accuracy revelation on the local optimum structure of the problem landscape

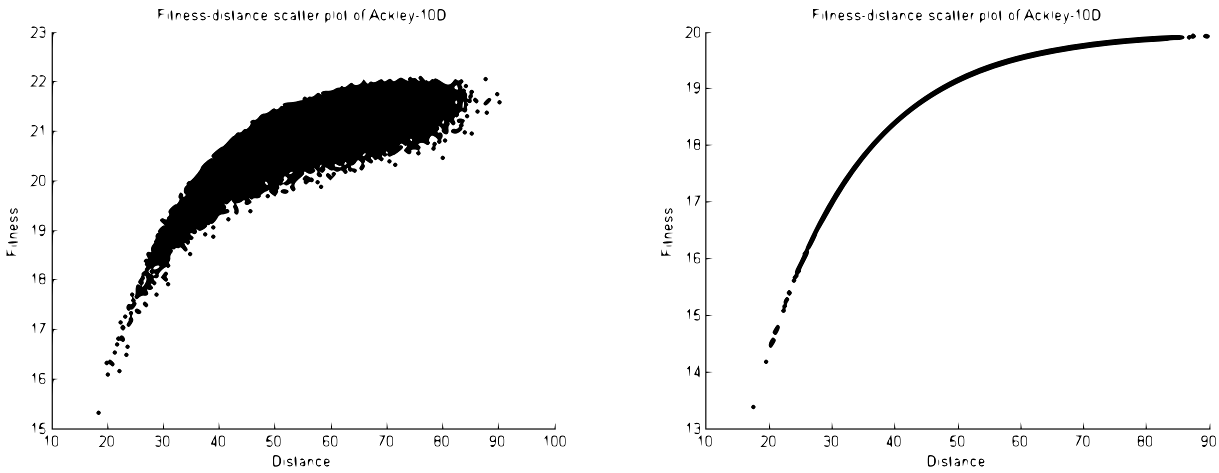
[5] $\mathbf{\Psi}_{Ackley}$, $\mathbf{\Psi}_{Rastrigin}$, $\mathbf{\Psi}_{Griewank}$, $\mathbf{\Psi}_{Rosenbrock}$ and $\mathbf{\Psi}_{Weierstrass}$

[6] Cauchy's convergence test for sequence $\{\mathbf{x}_i\}$ can be described as: for every $\epsilon > 0$, there is a number N, such that for all $n, m > N$ holds $\|\mathbf{x}_m - \mathbf{x}_n\| < \epsilon$

[7] On a minimization problem, the local optimum solutions of the landscape is considered to possess strong constructive correlation property if *FDC-L* $\varrho_L \approx 1$, i.e., the closer an individual is to the global optimum, the lower (better) is its fitness value. On the other hand, the local optimum structure is considered to display strong obstructive correlation property if *FDC-L* $\varrho_L \approx -1$.

Benchmark Problem	**Original Landscape Structure** ($\varrho \pm \sigma$)	**Transformed Landscape or Local Optimum Structure** ($\varrho_L \pm \sigma_L$)
Ackley	$0.780676 \pm 6.539E-3$ *Weak Constructive Correlation*	$0.920672 \pm 5.791E-3$ *Strong Constructive Correlation*
Rastrigin	$0.719241 \pm 6.925E-3$ *Poor correlation*	$0.994019 \pm 0.306E-3$ *Strong Constructive Correlation*
Griewank	$0.994486 \pm 0.247E-3$ *Strong Constructive Correlation*	$0.995402 \pm 0.297E-3$ *Strong Constructive Correlation*
Rosenbrock	$0.621473 \pm 9.374E-3$ *Poor correlation*	$0.999912 \pm 0.228E-3$ *Strong Constructive Correlation*
Weierstrass	$0.877178 \pm 3.749E-3$ *Weak Constructive Correlation*	$0.709783 \pm 8.725E-3$ *Poor correlation*

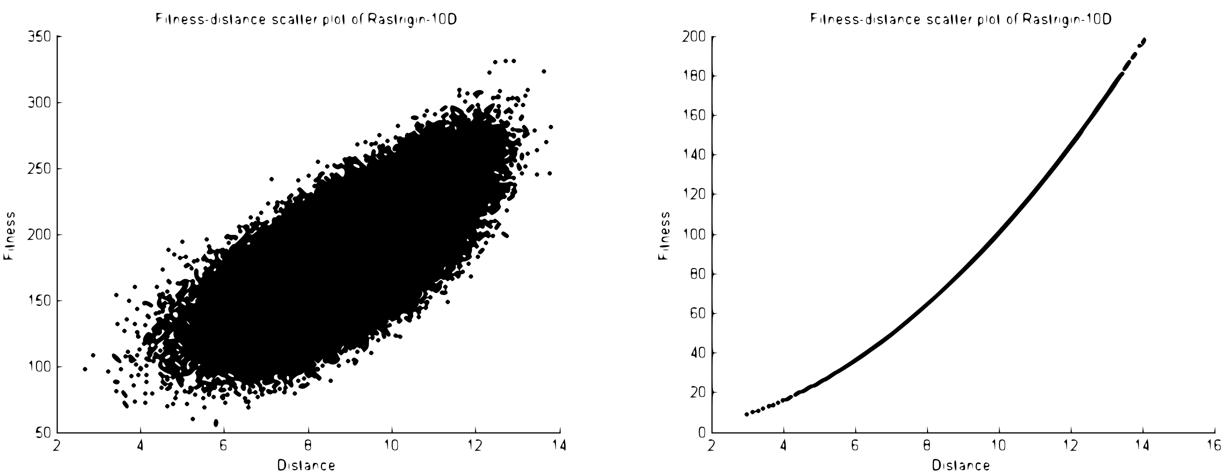
Table 2 Mean (ϱ_L/ϱ) and Variance (σ_L/σ) of Landscape Structure Correlation



(a) Ackley(10D)

(b) L-Ackley(10D)

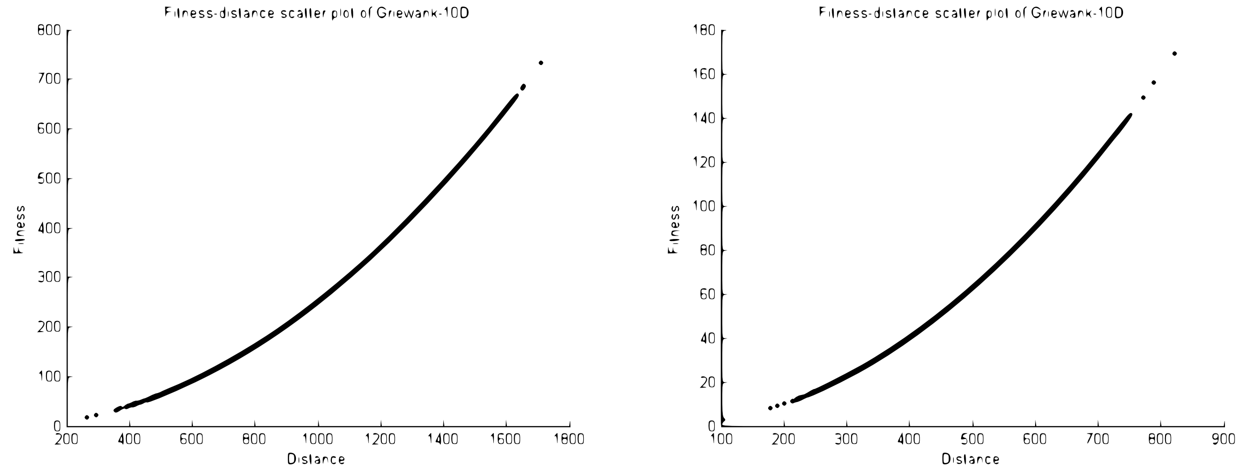
Fig. 4 Fitness distance scatter plots of Ackley function



(a) Rastrigin(10D)

(b) L-Rastrigin(10D)

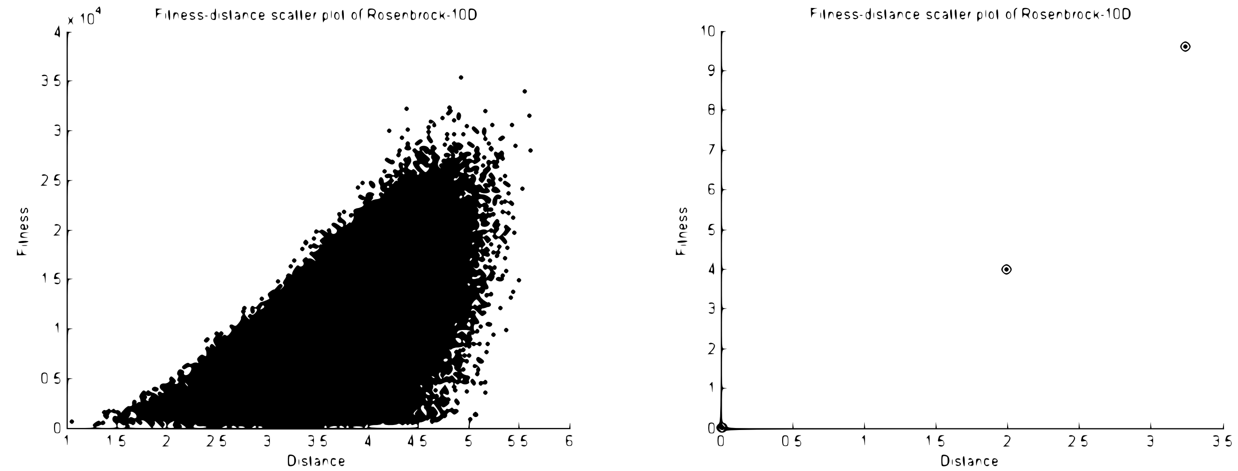
Fig. 5 Fitness distance scatter plots of Rastrigin function



(a) Griewank(10D)

(b) L-Griewank(10D)

Fig. 6 Fitness distance scatter plots of Griewank function



(a) Rosenbrock(10D)

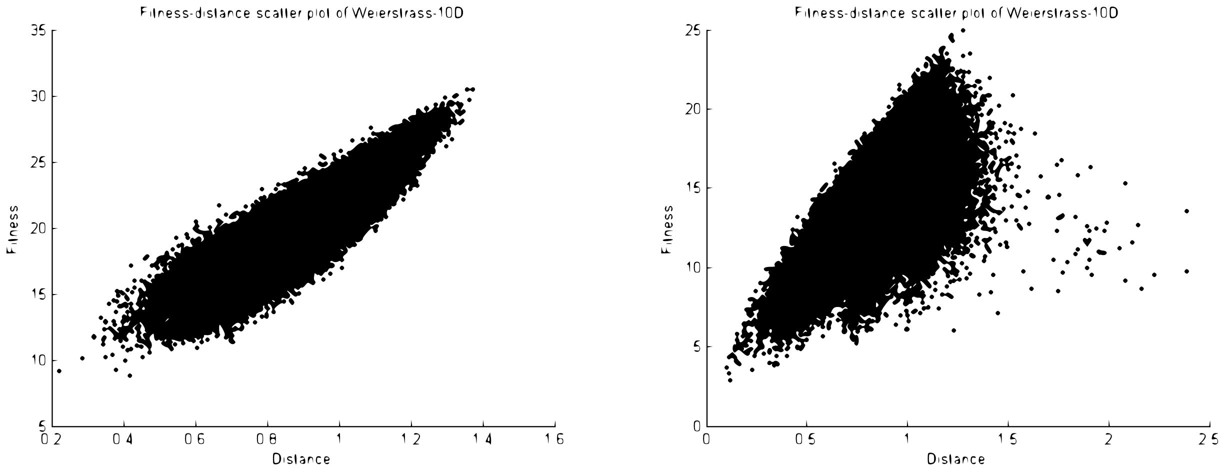
(b) L-Rosenbrock(10D)

Fig. 7 Fitness distance scatter plots of Rosenbrock function

local optimum structure observed, the stochastic selection operator of the MA is expected to benefit from the strong "pulling" effect on the population, which promotes the search towards the global optimum. Further, it is also revealed that the Weierstrass function has a degrading degree of constructive correlated structure, i.e., $\varrho_L < \varrho$, in MA search, different to the other representative benchmark problems considered in this work.

4 Local Optimum Connectivity Analysis of Individual Learning Procedure or Memes in MA

Recent studies on MAs have shown that the choice of individual learning procedure or meme employed significantly affects the search performance (10; 44; 14; 22). Given the restricted theoretical knowledge available in this area, we attempt to provide in the following some insights into the success of an individual learning procedure using the connectivity analysis method proposed above.



(a) Weierstrass(10D) (b) L-Weierstrass(10D)

Fig. 8 Fitness distance scatter plots of Weierstrass function

Table 3 MA parameters setting

General parameters	
Individual learning strategies	DSCG and DFP
Encoding scheme	Real-coded
Reproduction operator	Gaussian mutation $N(0,1)$
Initial step size s	0.5
Stoping criteria ϵ	1.0E-4

4.1 Empirical Study

In our experimental study, we consider a MA that uses the Gaussian mutation operator for reproduction, and the Davidon-Fletcher-Powell (DFP) or Davies, Swann, & Campey with Gram-Schmidt orthogonalization (DSCG) for individual learning, which are notated hereafter as MA-DFP and MA-DSCG, respectively (Table 3). DSCG represents a form of direct search method that has been demonstrated to have good performance over some derivative-based numerical methods on the set of continuous benchmark problems in (10; 44). DFP, on the other hand, is a popular quasi-Newton based individual learning method, which has also been used in Section 3 to obtain the sets of local optimums. Using the five sets of local optimum solutions for the benchmark problems [8] obtained in Section 3, the expected improvement $E[\triangle d(\mathbf{x},\mathbf{x}^{''})|P]$[9] and cost $E[C(\mathbf{x}^{'},\mathbf{x}^{''})|P]$ [10] of individual learning for each local optimum $\mathbf{x}$ are subsequently estimated to provide revelation into the search mechanisms of the Gaussian mutation and individual learning operators, i.e., DFP or DSCG in the MAs, through local optimum connectivity analysis [11].

The fitness function $f(\mathbf{x})$ in Equation (1) indirectly provides an indication of the distance to global optimum $d(\mathbf{x},\mathbf{x}^*)$. Hence $\triangle d(\mathbf{x},\mathbf{x}^{''}) = d(\mathbf{x},\mathbf{x}^*) - d(\mathbf{x}^{''},\mathbf{x}^*)$ may be replaced by the fitness improvement $\triangle f(\mathbf{x},\mathbf{x}^{''}) = f(\mathbf{x}) - f(\mathbf{x}^{''})$. Here, the computational cost $C(\mathbf{x}^{'},\mathbf{x}^{''})$ is defined by the number of function evaluations incurred. The expectation of fitness improvement $E[\triangle f(\mathbf{x},\mathbf{x}^{''})|P]$ is then estimated through a simulation of $T = 10 \times n$ iterations (n denotes the number of dimensions) on each local optimum $\mathbf{x}$: $E[\triangle f(\mathbf{x},\mathbf{x}^{''})|\mathbf{x}] \approx \sum_{i=1}^{T}(f(\mathbf{x}) - f(\mathbf{x}_i^{''}))/T$. In each iteration i, $\mathbf{x}_i^{'} = R(\mathbf{x})$ is an offspring of x reproduced by mutation with a normal distribution $\mathbf{N(0,1)}$ and $\mathbf{x}_i^{''} = IL(\mathbf{x}_i^{'})$ denotes the resulting individual after learning. The details of the simulation procedure is outlined in Algorithm 2. Using a similar procedure, the computational cost expectation of the search operators can also be estimated as $E[C(\mathbf{x}^{'},\mathbf{x}^{''})] \approx \sum_{i=1}^{T} C_i/T$, where C_i denotes the total computational cost incurred by each iteration.

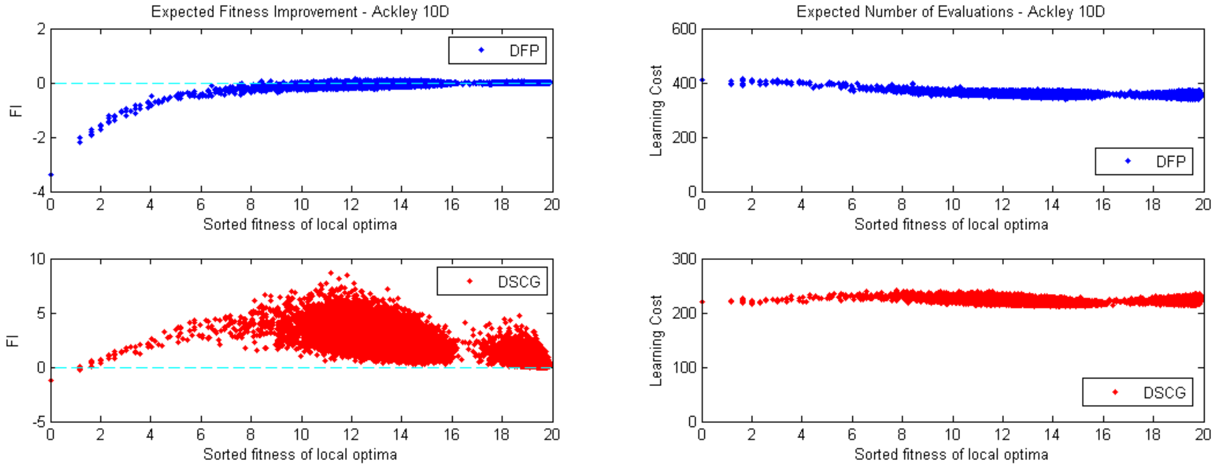
Simulation studies on the five 10-dimensional benchmark problems using the algorithm outlined in Algorithm 2 are performed, and the results are summarized in the *FI-plot* $(f(\mathbf{x}), E[\triangle f(\mathbf{x},\mathbf{x}^{''})|P])$ and *C-plot* $(f(\mathbf{x}), E[C(\mathbf{x}^{'},\mathbf{x}^{''})])$ in Figures 9-13. Note that *FI-plot* and *C-plot* provide information on progress rate θ_{R-IL} and computational effort incurred by the MA search, respectively.

[8] $\mathbf{\Psi_{Ackley}}$, $\mathbf{\Psi_{Rastrigin}}$, $\mathbf{\Psi_{Griewank}}$, $\mathbf{\Psi_{Rosenbrock}}$ and $\mathbf{\Psi_{Weierstrass}}$

[9] Note that $\triangle d(\mathbf{x},\mathbf{x}^{''}) = d(\mathbf{x},\mathbf{x}^*) - d(\mathbf{x}^{''},\mathbf{x}^*)$, see Equation (9)

[10] $C(\mathbf{x},\mathbf{x}^{''}) = \alpha + C(\mathbf{x}^{'},\mathbf{x}^{''})$, see Equation (10)

[11] As discussed in Section 2.2.2, an *'positive'/ 'negative'* expected improvement $E[\triangle d(\mathbf{x},\mathbf{x}^{''})|P]$ for each $\mathbf{x}$ indicates a *'constructive'/ 'obstructive'* connectivity of local optimums.



(a) Expected Fitness Improvement on Ackley(10D) (b) Expected Cost of Learning on Ackley(10D)

Fig. 9 Connectivity Analysis on Ackley(10D) function

Algorithm 2 Descriptions for Local Optimum Connectivity Analysis

for $\mathbf{x}$ in $\mathbf{\Psi}$ **do**
  $FI(\mathbf{x}) = 0$, $C(\mathbf{x}) = 0$
  for $i = 1$ to $T = 10 \times n$ **do**
    Produce offspring $\mathbf{x}_i^{'}$ according to $R(\mathbf{x})$
    Produce individual $\mathbf{x}_i^{''}$ by individual learning $L(\mathbf{x}_i^{'})$
    Fitness improvement $\triangle f(\mathbf{x}, \mathbf{x}_i^{''}) = f(\mathbf{x}) - f(\mathbf{x}_i^{''})$
    $FI(\mathbf{x}) = FI(\mathbf{x}) + \triangle f(\mathbf{x}, \mathbf{x}_i^{''})$
    $C(\mathbf{x}) = C(\mathbf{x}) + C_i$ $\{C_i$: function evaluations$\}$
  end for
  $FI(\mathbf{x}) = FI(\mathbf{x})/T$
  $C(\mathbf{x}) = C(\mathbf{x})/T$
end for
Provide FI, C plots

4.1.1 Ackley function

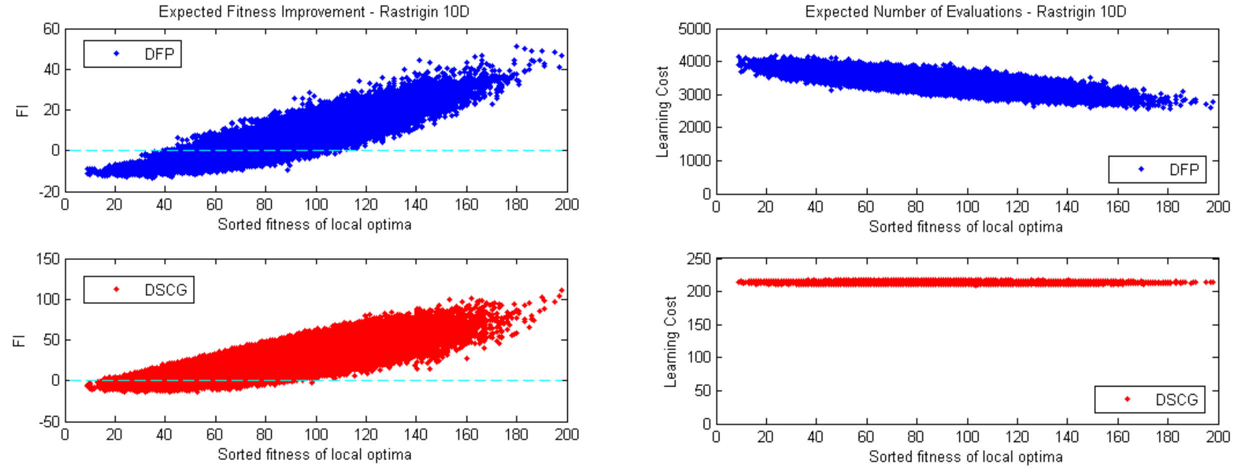
The *FI-plot* and *C-plot* for MA-DFP and MA-DSCG on the Ackley function are illustrated in Figures 9(a) and 9(b), respectively. On the lower panel of Figure 9(a), DSCG is observed to result in a positive expected FI, i.e., $\triangle f(\mathbf{x}, \mathbf{x}^{''})$, on each local optimum $\mathbf{x}$. In contrast, the upper panel of Figure 9(a) indicates that DFP results in a *negative* expected FI for the local optimum with fitness value under 8 (see x-axis). Further, the negative expected FI phenomenon is shown to prevail for most of the local optimums even at higher fitness ranges. The results highlighted a significantly higher expected FI on the local optimum resulting from DSCG-based individual learning than DFP-based individual learning. Based on the notions of constructive and obstructive connectivity described in Section 2.2.2, the property of a constructive connectivity, i.e., local optimums of Ackley function are likely to connect to other optimums of higher fitness, is shown for the MA-DSCG. In contrast, obstructive connectivity is observed on the MA-DFP. In addition, Figure 9(b) presents the expected computational cost of learning for DSCG and DFP strategies on the set of local optimums on Ackley function. From the figure, it is observed that the cost of learning for DSCG is approximately half that of DFP to arrive at the expected improvements shown in Figure 9(a).

4.1.2 Rastrigin function

As shown in Figure 10(a), both MA-DFP and MA-DSCG bring about a *positive* expected FI on majority of the local optimum, but a *negative* expected FI on some high quality local optimums of the Rastrigin function. Hence a obstructive connectivity can be inferred for local optimums that are closer to the global optimal. Comparing MA-DFP with MA-DSCG, however, MA-DSCG results in a higher expected FI on most of the local optimums than MA-DFP. Overall, MA-DSCG displays a stronger constructive connectivity profile than MA-DFP on Rastrigin. The expected computational cost for learning of the MAs is also given in Figure 10(b), which indicates that the learning cost for MA-DFP is approximately 10 times higher than that of MA-DSCG.

4.1.3 Griewank function

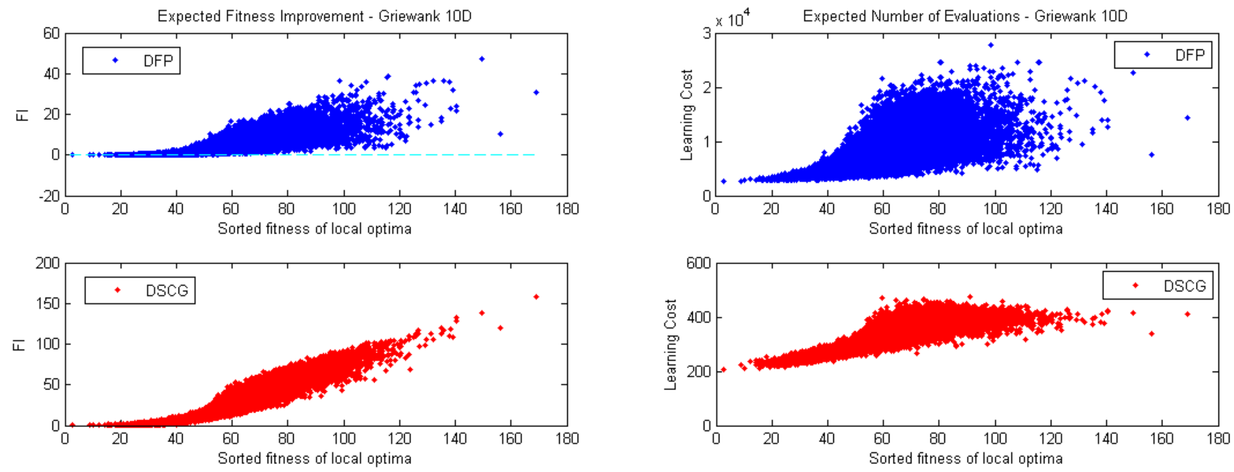
Next, we discuss the simulation results on the Griewank function. Although the local optimum landscape of Griewank function appears similar to that of Rastrigin function, it is worth noting that in contrast to the latter, both MA-DFP and



(a) Expected Fitness Improvement on Rastrigin(10D)

(b) Expected Cost of Learning on Rastrigin(10D)

Fig. 10 Connectivity Analysis on Rastrigin(10D) function



(a) Expected Fitness Improvement on Griewank(10D)

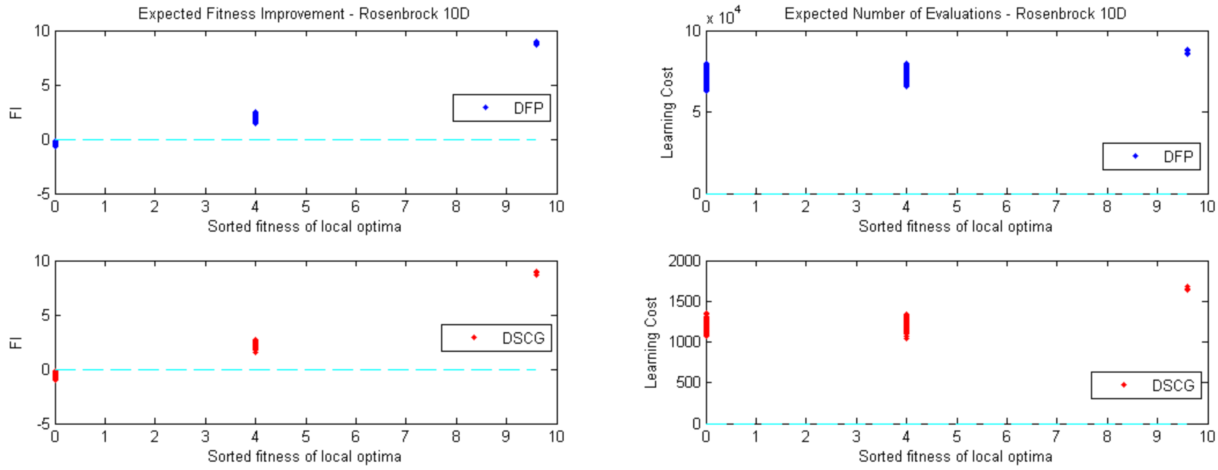
(b) Expected Cost of Learning on Griewank(10D)

Fig. 11 Connectivity Analysis on Griewank(10D) function

MA-DSCG exhibit a *positive* expected FI on nearly all the local optimums, as observed in Figure 11(a). Further, MA-DSCG generates a higher expected FIs than MA-DFP for the same local optimums. Particularly, for local optimum with a fitness value of 100, the MA-DSCG contributed expected FIs that are in the range of $50 - 100$, while MA-DFP generated FIs lower than 40. In addition, the learning cost of DFP (at > 2500 evaluations for each local optimum), as compared to that of DSCG (< 500), is shown in Figure 11(b), highlighting the remarkably larger computational requirements of DFP.

4.1.4 Rosenbrock function

For the Rosenbrock function, it is worth noting the sparseness of the scatter plot in Figure 12, which indicates that a small number of local optimums exists in the landscape. The obtained expected FI of Figure 12(a) clearly depicts the strong constructive connectivity of the local optimums to global optimum. For instance, it is observed that the expected FI is approximately 10 for a local optimum $\mathbf{x}$ having fitness value 10, from which we can infer that the local optimum $\mathbf{x}$ is very likely to connect to the global optimum. In addition, the cost of individual learning for DFP is consid-



(a) Expected Fitness Improvement on Rosenbrock(10D)

(b) Expected Cost of Learning on Rosenbrock(10D)

Fig. 12 Connectivity Analysis on Rosenbrock(10D) function

erably higher than that of DSCG on Rosenbrock function as shown in Figure 12(b).

4.1.5 Weierstrass function

MA-DSCG shows a *positive* expected FI on nearly all local optimums on the Weierstrass function. A linearly increasing positive expected improvement is observed in the lower panel of Figure 13(a) for MA-DSCG. MA-DFP, on the other hand, exhibits an entirely opposite behavior, with most local optimums giving a *negative* expected FI, as shown in Figure 13(a), i.e., for those local optimums with a fitness value < 15. Further, in contrast to the other representative problems where DSCG is found to be computationally more efficient than DFP, the cost of learning for DSCG is approximately twice that of DFP on the Weierstrass function, see Figure 13(b).

The local optimum connectivity profiles of MA-DSCG and MA-DFP on the five benchmark problems are listed in Tables 4 and 5. These results indicate that MA-DSCG possesses a constructive local optimum connectivity on most of the representative test functions, while MA-DFP produces an obstructive connectivity of local optimums when approaching the global optimum of the Ackley, Rastrigin functions and Weierstrass functions. Further, it is worth noting that local optimum connectivity with a higher expected FI is achieved at a lower computation cost in MA-DSCG as compared to MA-DFP on four of the five problems considered, thus shedding some light on the success of MA-DSCG widely reported in the literature (10; 44).

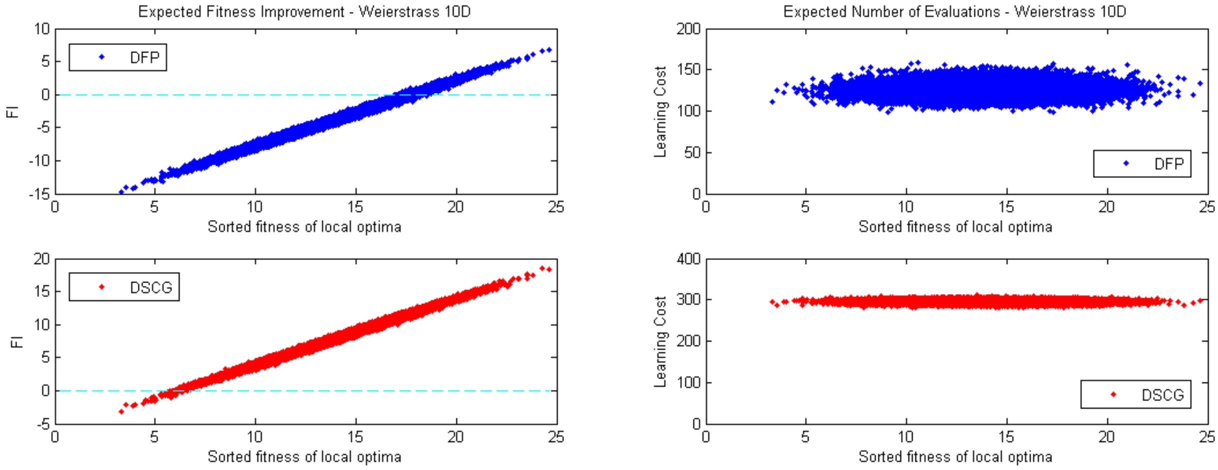
5 Conclusions

In this paper, we have proposed and introduced the concepts of *local optimum structure* and *connectivity structure* for the analysis of MAs. The notion of 'constructive'/ 'obstructive' local optimum structure is defined as a property of the transformed fitness landscape as a result of the individual learning procedure in MAs. 'Constructive'/ 'obstructive' connectivity, on the other hand, is introduced as a property for revealing the working mechanism of a MA solver in search. Both properties of local optimum structure and connectivity are demonstrated to have great influence on the search mechanism and performance of MAs. The results on typical benchmark problems indicated that the correlated structure of the transformed landscape is improved as compared to the original, thus allowing the MA to reap the benefits of individual learning in advancing the search towards the global optimum. Further analysis on MA-DSCG and MA-DFP also highlighted the unique local optimum connectivity properties of memes and their influences on MA search performance, which explains the superior performance of MA-DSCG reported in previous studies.

Acknowledgements M.N. Le is grateful for the financial support from Honda Research Institute Europe.

References

1. Y. S. Ong, M. H. Lim, F. Neri, and H. Ishibuchi, "Special issue on Emerging trends in soft computing: Memetic Algorithms," *Soft Computing-A Fusion*



(a) Expected Fitness Improvement on Weierstrass(10D) (b) Expected Cost of Learning on Weierstrass(10D)

Fig. 13 Connectivity Analysis on Weierstrass(10D) function

Benchmark Problem	**Local Optimum Connectivity**	
	MA-DSCG	**MA-DFP**
Ackley	Constructive	Obstructive
Rastrigin	Constructive	Constructive
Griewank	Constructive	Constructive
Rosenbrock	Constructive	Constructive
Weierstrass	Constructive	Obstructive

Table 4 Local Optimum Connectivity Profiles of Benchmark Problems

Benchmark Problem	**Higher Expected FI**	**Lower Computational Cost**
Ackley	MA-DSCG	MA-DSCG
Rastrigin	MA-DSCG	MA-DSCG
Griewank	MA-DSCG	MA-DSCG
Rosenbrock	Equal FI	MA-DSCG
Weierstrass	MA-DSCG	MA-DFP

Table 5 Summary of Analysis Results on Local Optimum Connectivity

of Foundations, Methodologies and Applications, vol. 13(8-9), pp. 739–740, 2008.
2. Y. S. Ong, N. Krasnogor, and H. Ishibuchi, "Special issue on Memetic Algorithms," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 37, no. 1, pp. 2–5, 2007.
3. F. Neri, P. Moscato, and H. Ishibuchi, "Special session: Memetic Algorithms for hard to solve problems," *IEEE World Congress on Computational Intelligence*, 2009.
4. Y. S. Ong, F. Neri, H. Ishibuchi, and M. H. Lim, "Memetic Algorithms: Special session," *IEEE World Congress on Computational Intelligence*, 2007,2008.
5. M. H. Lim, G. S, , K. N, and Y. S. Ong, "Editorial to the first issue," *Memetic Computing*, vol. 1(1):1-2, 2009.
6. P. Moscato, "On evolution, search, optimization, genetic algorithms and martial arts: Towards memetic algorithms," *Caltech Concurrent Computation Program, C3P Report*, vol. 826, p. 1989, 1989.
7. B. H. Gwee and M. H. Lim, "Polyominoes Tiling by a Genetic Algorithm," *Computational Optimization and Applications Journal,*, vol. 6(3), pp. 273–291, 1996.
8. M. Lim, Y. Yuan, and S. Omatu, "Efficient Genetic Algorithms using Simple Genes Exchange Local Search Policy for the Quadratic Assignment Problem," *Computational Optimization and Applications*, vol. 15(3), pp. 249–268, 2000.
9. R. Lewis and B. Paechter, "Finding feasible timetables using group-based operators," *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 3, pp. 397–413, 2007.
10. Y. S. Ong and A. Keane, "Meta-Lamarckian learning in memetic algorithms," *IEEE Transactions on Evolution-*

ary Computation, vol. 8, no. 2, pp. 99–110, 2004.

11. A. Vicini and D. Quagliarella, “Airfoil and wing design through hybrid optimization strategies,” *American Institute of Aeronautics and Astronautics Journal*, vol. 37, no. 5, pp. 634–641, 1999.
12. Z. Michalewicz, *Genetic Algorithms+ Data Structures= Evolution Programs*. Springer-Verlag London, UK, 1996.
13. C. Houck, J. Joines, and M. Kay, “Utilizing Lamarckian evolution and the Baldwin effect in hybrid genetic algorithms,” Tech. Rep., 1996.
14. W. E. Hart, “Adaptive global optimization with local search,” Ph.D. dissertation, University of California, San Diego, 1994.
15. J. Renders and H. Bersini, “Hybridizing genetic algorithms with hill-climbing methods for global optimization: two possible ways,” *IEEE World Congress on Computational Intelligence*, vol. 1, pp. 312–317, 1994.
16. L. Davis, *Handbook of Genetic Algorithms*. Van Nostrand Reinhold Company, 1991.
17. N. Noman and H. Iba, “Accelerating differential evolution using an adaptive local search,” *IEEE Transactions on Evolutionary Computation*, vol. 12, no. 1, pp. 107–125, 2008.
18. Y. S. Ong, P. B. Nair, and K. Lum, “Max-min surrogate-assisted evolutionary algorithm for robust design,” *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 4, pp. 392–404, 2006.
19. N. Krasnogor and J. E. Smith, “A tutorial for competent memetic algorithms: Model, taxonomy, and design issues,” *IEEE Transactions on Evolutionary Computation*, vol. 9, no. 5, pp. 474–488, 2005.
20. Y. S. Ong, P. B. Nair, and A. J. Keane, “Evolutionary optimization of computationally expensive problems via surrogate modeling,” *American Institute of Aeronautics and Astronautics Journal*, vol. 41, no. 4, pp. 687–696, 2003.
21. H. Ishibuchi, T. Yoshida, and T. Murata, “Balance between genetic search and local search in memetic algorithms for multiobjective permutation flowshop scheduling,” *IEEE Transactions on Evolutionary Computation*, vol. 7, no. 2, pp. 204–223, 2003.
22. N. Krasnogor, “Studies on the theory and design space of memetic algorithms,” Ph.D. dissertation, Doctoral dissertation, University of the West of England, Bristol, England, 2002.
23. N. Krasnogor, B. P. Blackburne, E. K. Burke, and J. D. Hirst, “Multimeme algorithms for protein structure prediction,” *Proceedings of the Parallel Problem Solving from Nature VII. Lecture Notes in Computer Science.*, vol. 2439/2002, pp. 769–778, 2002.
24. R. Meuth, M.-H. Lim, Y.-S. Ong, and D. C. Wunsch II, “A proposition on memes and meta-memes in computing for higher-order learning,” vol. 1, no. 2, 2009, pp. 85–100.
25. Y. S. Ong, M. H. Lim, N. Zhu, and K. W. Wong, “Classification of adaptive memetic algorithms: A comparative study,” *IEEE Transactions on Systems, Man and Cybernetics, Part B*, vol. 36, no. 1, pp. 141–152, 2006.
26. G. E. Hinton and S. J. Nowlan, “How learning can guide evolution,” *Complex Systems*, vol. 1, no. 1, pp. 495–502, 1987.
27. E. Borenstein, I. Meilijson, and E. Ruppin, “The effect of phenotypic plasticity on evolution in multipeaked fitness landscapes,” *Journal of Evolutionary Biology*, vol. 19, no. 5, pp. 1555–1570, 2006.
28. I. Paenke, T. Kawecki, and B. Sendhoff, “The influence of learning on evolution: A mathematical framework,” *Artificial Life*, vol. 15, no. 2, pp. 227–245, 2009.
29. I. Paenke, Y. Jin, and J. Branke, “Balancing population and individual level adaptation in changing environments,” *Adaptive Behavior*, vol. 17, no. 2, pp. 153–174, 2009.
30. P. Merz, “Advanced fitness landscape analysis and the performance of memetic algorithms,” *Evolutionary Computation*, vol. 12, no. 3, pp. 303–325, 2004.
31. ——, “Memetic algorithms for combinatorial optimization problems: fitness landscapes and effective search strategies,” Ph.D. dissertation, University of Siegen, Germany, 2000.
32. D. Whitley, V. Gordon, and K. Mathias, “Lamarckian evolution, the Baldwin effect and function optimization,” *Parallel Problem Solving from Nature–PPSN III: International Conference on Evolutionary Computation, The Third Conference on Parallel Problem Solving from Nature*, pp. 6–15, 1994.
33. I. Paenke, B. Sendhoff, J. Rowe, and C. Fernando, “On the adaptive disadvantage of Lamarckianism in rapidly changing environments,” *European Conference on Artificial Life (ECAL), 10-14 September 2007, Lecture Notes in Computer Science*, vol. 4648, pp. 355–364, 2007.
34. J. Zhang and A. C. Sanderson, “Adaptive Differential Evolution: A Robust Approach to Multimodal Problem Optimization,” *Series on Adaptation, Learning and Optimization*, vol. 1, 2009.
35. E. Jablonka and M. J. Lamb, *Epigenetic inheritance and evolution: The Lamarckian dimension*. Oxford University Press, 1995.
36. M. W. Ho, “Why Lamarck won’t go away,” *Annals of Human Genetics*, vol. 60, no. 1, pp. 81–84, 1996.
37. H.-G. Beyer, “An alternative explanation for the manner in which genetic algorithms operate,” *BioSystems*, vol. 41, no. 1, pp. 1–15, 1997.
38. H.-G. Beyer, H.-P. Schwefel, and I. Wegener, “How to analyse evolutionary algorithms,” *Theoretical Com-*

puter Science, vol. 287, no. 1, pp. 101–130, 2002.

39. H.-P. Schwefel, *Evolution and Optimum Seeking: The Sixth Generation*. New York, NY, USA: John Wiley & Sons, Inc., 1993.
40. D. Goldberg and K. Deb, "A comparative analysis of selection schemes used in genetic algorithms," *Foundations of Genetic Algorithms*, vol. 1, pp. 69–93, 1991.
41. J. Wojtusiak and R. S. Michalski, "The LEM3 implementation of learnable evolution model and its testing on complex function optimization problems," in *GECCO '06: Proceedings of the 8th annual conference on Genetic and evolutionary computation*. New York, NY, USA: ACM Press, 2006, pp. 1281–1288.
42. Y.-W. Leung and Y. Wang, "An orthogonal genetic algorithm with quantization for global numerical optimization," *IEEE Transactions on Evolutionary Computation*, vol. 5, pp. 41–53, February 2001.
43. X. Yao and Y. Liu, "Fast evolution strategies," *Control and Cybernetics*, vol. 26, pp. 467–496, 1997.
44. Q. H. Nguyen, Y. S. Ong, and N. Krasnogor, "A study on the design issues of Memetic Algorithm," *IEEE Congress on Evolutionary Computation (CEC), 25-28 September 2007*, pp. 2390–2397, 2007.
45. J. Tavares, F. B. Pereira, and E. Costa, "Multidimensional knapsack problem: A fitness landscape analysis," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 38, no. 3, pp. 604–616, June 2008.
46. T. Jones and S. Forrest, "Fitness distance correlation as a measure of problem difficulty for genetic algorithms," *Proceedings of the 6th International Conference on Genetic Algorithms*, pp. 184–192, 1995.
47. Y.-W. Shang and Y.-H. Qiu, "A note on the extended Rosenbrock function," *Evolutionary Computation*, vol. 14, no. 1, pp. 119–126, 2006.