



Pareto-like sequential sampling heuristic for global optimisation

Mahmoud Shaqfa¹ · Katrin Beyer¹

Accepted: 28 April 2021 / Published online: 29 May 2021
© The Author(s) 2021

Abstract

In this paper, we propose a simple global optimisation algorithm inspired by Pareto's principle. This algorithm samples most of its solutions within prominent search domains and is equipped with a self-adaptive mechanism to control the dynamic tightening of the prominent domains while the greediness of the algorithm increases over time (iterations). Unlike traditional metaheuristics, the proposed method has no direct mutation- or crossover-like operations. It depends solely on the sequential random sampling that can be used in diversification and intensification processes while keeping the information-flow between generations and the structural bias at a minimum. By using a simple topology, the algorithm avoids premature convergence by sampling new solutions every generation. A simple theoretical derivation revealed that the exploration of this approach is unbiased and the rate of the diversification is constant during the runtime. The trade-off balance between the diversification and the intensification is explained theoretically and experimentally. This proposed approach has been benchmarked against standard optimisation problems as well as a selected set of simple and complex engineering applications. We used 26 standard benchmarks with different properties that cover most of the optimisation problems' nature, three traditional engineering problems, and one real complex engineering problem from the state-of-the-art literature. The algorithm performs well in finding global minima for nonconvex and multimodal functions, especially with high dimensional problems and it was found very competitive in comparison with the recent algorithmic proposals. Moreover, the algorithm outperforms and scales better than recent algorithms when it is benchmarked under a limited number of iterations for the composite CEC2017 problems. The design of this algorithm is kept simple so it can be easily coupled or hybridised with other search paradigms. The code of the algorithm is provided in C++14, Python3.7, and Octave (Matlab).

Keywords Pareto principle · Heuristic · Global optimisation · Evolutionary algorithms · Self-adaptation · Online calibration

1 Introduction

Over the past few decades, global optimisation techniques for solving combinatorial problems have flourished. The ever-increasing complexity of engineering applications means that variable sets are growing larger, and the subsequent landscapes that need to be explored by these optimisation problems are becoming increasingly complicated.

This work was funded through the Swiss National Science Foundation (SNSF) Project 200021_175903/1.

✉ Katrin Beyer
katrin.beyer@epfl.ch

Mahmoud Shaqfa
mahmoud.shaqfa@epfl.ch

¹ Earthquake Engineering and Structural Dynamics Laboratory (EESD), School of Architecture, Civil and Environmental Engineering (ENAC), École polytechnique fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

Many metaheuristic algorithms were developed to solve optimisation problems by mimicking biological and physical analogies (Ser et al. 2019), including algorithms such as genetic algorithms (GAs) (Holland 1992), particle swarm optimisation (PSO) (Kennedy and Eberhart 1995) and the recent generalized version (GEPSON) (Sedighzadeh et al. 2021), and harmony search (HS) (Geem et al. 2002) and its latest modifications such as the Geem and Sim (2010), Shaqfa and Orbán (2019) and Jeong et al. (2020), as well as the recent whales optimisation algorithm (WOA) (Mirjalili and Lewis 2016), and pathfinder algorithm (PFA) (Yapici and Cetinkaya 2019), to mention but a few. The candidate problems usually range from continuous differentiable problems to discrete, noisy, and even loosely defined objectives, such as in engineering applications.

Sequential sampling was used intensively in the second half of the last century for estimating the distribution of unknown functions, and the influence of such methods is

still highly echoed in current engineering metamodels (Jin et al. 2002). As the purpose of this paper is not to review all the optimisation algorithms that employed sequential sampling techniques, we advise the readers to refer to (de Mello and Bayraksan 2014) where they can find a comprehensive review for algorithms used the Monte Carlo sampling approach for solving optimisation problems. Holistically, metaheuristics can be seen as performance-driven sequential sampling processes (Markovian processes) (Yang et al. 2013), where the efficiency of such algorithms depends on the random walks that are used to explore and exploit the provided landscapes. Many random walks have been proposed in the literature, such as Brownian motion (obeys Gaussian distribution) and the Lévy flights (obeys Lévy distribution) for handling stochastic optimisation problems (see Yang et al. 2013; Yang 2009 for comparison).

Estimation of distribution algorithms (EDAs) is another rank of search methods that tries to obtain better solutions from the search domain by estimating the probability density function (PDF) over the whole landscape from the so-far sampled solutions (see Dai et al. 2019 for a detailed review). One relatively recent method that has been used by many engineering problems was proposed by Raphael and Smith (2003). In their method, they subdivide the search domain into a fixed number of sub-domains and depending on the fitness of the collected solutions the PDF of the domain evolves over time and the solutions are eventually focused around the most promising regions.

Generally, most current memetic algorithms are population-based and heavily dependent—as an initial step—on crude random sampling. This initial population step is crucial for the overall performance of the algorithms, affects all subsequent search phases (Ser et al. 2019), is independent of the objective function and is considered an unbiased step of metaheuristics. Said otherwise, the quality of the chosen solutions in the initial population depends only on the design-of-experiment (DOE) methods used (space-filling techniques). Such methods assure equal accessibility to each part of the landscape. A plethora of publications dealt with the influence of the initial sampling and the population size on the overall performance of metaheuristics (such as in Polkov and Bujok 2018). For the here proposed algorithm, we use at every iteration DOE methods for sampling the new generation, as this guarantees simplicity and minimum structural bias of the algorithm.

The algorithm that we propose in this paper builds on the *gbest* topology that was originally used in the PSO algorithm by Eberhart and Kennedy (1995). This simple and conservative topology prevents the algorithm from exploiting too much information from the candidates, which could drag the algorithm into a structural bias (see Kononova et al. 2015 for definition) or even a premature convergence in late stages. In this work, we demonstrate how this simple algorithm can

perform surprisingly well for complex optimisation problems, sometimes outperforming the existing state-of-the-art metaheuristics.

We further postulate our motivation behind this algorithmic proposal where we will use it as a heuristic for solving physical packing problems. The chosen topology of this algorithm was tailored to analogise the exact mathematical solvers used in packing regular items (boxes) in containers. Those solvers use the domain-reduction techniques of the available packing domain as we add more items to the container; some areas become more occupied than others and exploring the search domain of such areas is not computationally efficient. However, this algorithm will be used for packing physical items that are highly irregular and nonconvex in containers. Such complex geometries make the use of traditional methods impractical. For a comprehensive review about the techniques of domain-reduction in mathematics and optimisation problems we refer the reader to Puranik and Sahinidis (2017) and Caprara and Locatelli (2010) and some examples such as Paulen et al. (2013) and for the packing optimisation problems literature see Martello and Toth (1990) and Carlier et al. (2007).

Our algorithm uses only the best solution from the previous generation. This best solution will be the center of the tightened search domain, which is referred to as the current promising region. The size of the current promising region will be a function of time—the more optimisation steps that have already been completed, the smaller the promising region will be. The boundaries of the prominent region are only updated if a better solution than any previously obtained solution has been found. We sample inside and outside this region using a standard DOE method; in this paper we use Monte-Carlo sampling, though any other DOE method could also be used such as the Latin hypercube sampling (LHS) method (see Owen 1992; Tang 1993; Ye 1998).

The chosen topology in this paper could be good for global exploration, as it maintains the population diversity, i.e., avoids premature convergence. Conversely, it could be problematic for exploiting local solutions where this process normally depends on the intensified information flow among the members of the population. Thus, this algorithm might be best hybridised in sequential or parallel schemes with other well-known algorithms. For more about topologies in metaheuristics, we refer the readers to Lynn et al. (2018) as they comprehensively review the common algorithmic topologies used in the literature. However, in this paper, we use the proposed algorithm as a standalone metaheuristic, with its simple topology, when benchmarking the algorithm against a range of problems and engineering applications.

The renowned no free lunch (NFL) theorems by Wolpert and Macready (1997) have proven the non-uniformity of the superiority of algorithms for all ranks of optimisation problems. In this spirit, we propose that our algorithm be used as

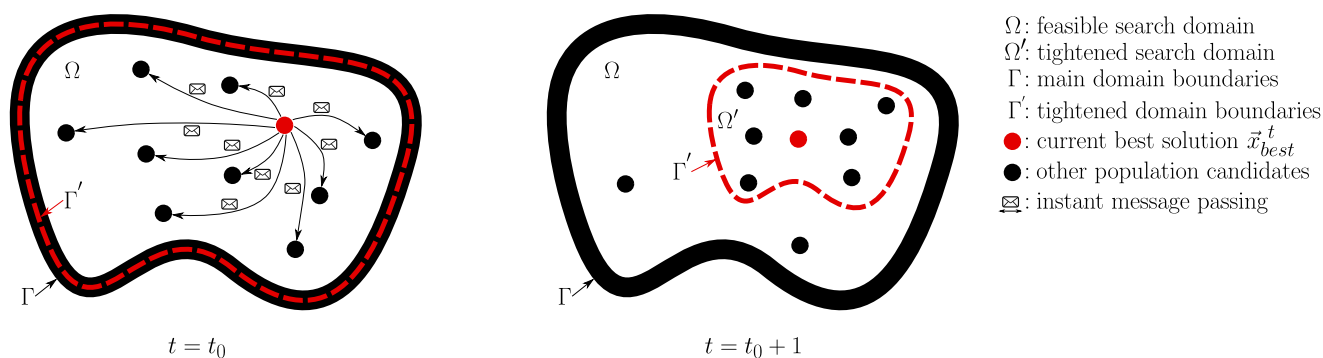


Fig. 1 The topology of Pareto-like sequential sampling (PSS) algorithm

template, which can be combined with various DOE methods and other optimisation algorithms.

This work has been organized such that Sect. 2 explains the analogy, the topology and the mathematical formulation of this algorithm. Section 3 provides a simple probabilistic analysis of the parameter settings, including a numerical illustration. In Sect. 4, we test and compare the performance of this simplified approach with the state-of-the-art algorithms. Finally, our conclusions are provided in Sect. 5, some future applications for this approach are given in Sect. 6 and links to the source code of the algorithm are provided in Sect. 7.

2 Pareto-like sequential sampling (PSS)

2.1 Analogy

As mentioned earlier, the main analogy of this model is to mimic the Pareto principle, which implies that a “virtual few” of a population account for the majority of the effects (Juran and Gryna 1988; Pareto 1897). The Pareto principal is popularly referred to as the 80/20 rule. This principle has been applied to a wide range of human relations and can be observed when a small percentage of contributors make the bulk impact on human-related matters. Historically, this 80/20 principle was first observed and described by the Italian professor Vilfredo Pareto from the University of Lausanne (UNIL), Switzerland (Pareto 1897).

We here employ the Pareto principle by more densely sampling the solutions’ design variables from a tightened search domain, i.e., the current prominent region, while sampling the remainder of the solutions from the overall search domain.

In order to better understand the collective behavior considered in our analogy, one out of many possible examples, is by letting us imagine a selected group of people were appointed to discuss a certain social problem in a local community to find a proper solution. First, the problem will be

communicated to each one of them individually and each one of them will have an initial opinion about what the solution could look like. However, these appointees, as representatives for the society, must sit together and meet regularly in order to reach a consensus on the optimal solution. As the appointees regularly meet, every one of them must express her/his own opinion, perspective, and exchange ideas of different solutions. Every meeting the majority of these individuals will group behind certain revealing ideas that could be better to settle down the problem and try even to push it further by combining different solutions or sub-solutions from other individuals to reach an optimal solution. While the rest will always try to find better and optimal ideas that could convince the majority to shift to their sides. In this particular example, the appointees are the population of the solutions as they start randomly at one point (before the meetings and when they got informed) and then they refine their ideas every time they meet (number of iterations).

2.2 Topology

In this paper, we achieve the proposed analogy using mere performance-driven sequential sampling (see Fig. 1) to create better generations every iteration. We do this in a way that is compatible with any sampling method, i.e., classical DOE methods, such as the traditional Monte Carlo (MC) and Latin hypercube sampling (LHS), though in this paper we only use the MC sampling approach. Unlike traditional evolutionary algorithms (EAs), this algorithm does not use operations such as mutation or crossover to obtain better generations.

The process builds on the assumption that the current best solution vector, \mathbf{x}_{best} , lays in a sub-domain where the majority of the good features exist (components of \mathbf{x}). We refer to this sub-domain as the prominent region Ω' . Hence, we sample the majority of the features, approximately $\alpha \times 100\%$, from the surrounding neighborhood. This topology of keeping track of only \mathbf{x}_{best} each iteration is similar to the traditional *gbest* topology proposed by Eberhart and Kennedy (1995).

The location and the size of the prominent domain changes every time the \mathbf{x}_{best} changes. The size of the prominent domain (bandwidth), as will be described later, depends on the time remaining for the algorithm to make improvements and the predetermined size of the analogy $(1 - \alpha)$. The size of the prominent region is set to a maximum of $(1 - \alpha)\%$ of Ω , centered about \mathbf{x}_{best} , where α is the acceptance probability that allows us to sample the components of the solutions from the prominent neighbourhood. The closer the algorithm gets to the maximum number of iterations, i.e., the shorter the time left for the algorithm to make improvements, the smaller the size of the prominent region.

This algorithm *explores* and *exploits* the domain by giving special emphasis to the prominent region around the current best solution. At each iteration, the algorithm performs the following steps:

- Determine the new best fitness \mathbf{x}_{best}^i ,
- Update the prominent region η^i if \mathbf{x}_{best}^i is better than \mathbf{x}_{best}^{i-1} ,
- Sample from the search domain using classical DOE methods; sample more densely from the prominent regions η^i and less densely from the overall search domain.

2.3 Mathematical model

Let the real-valued optimization problem $f(\mathbf{x})$ be defined on the search domain $\Omega \in \mathbb{R}^n$ and for all sub-domains ${}^i\Omega' \subset \Omega$. The corresponding boundaries are Γ and ${}^i\Gamma'$ for the main domain and the sub-domain, respectively.

$$\begin{aligned} &\text{optimize } f(\mathbf{x}), \mathbf{x} \in \Omega, \Omega \subset \mathbb{R}^n, \text{ where} & (1) \\ &\mathbf{x} = \{x_1, \dots, x_n\}, \Gamma^- \leq \mathbf{x} \leq \Gamma^+. \end{aligned}$$

More specifically, the lower and the upper bounds for the main domain and the sub-domain can be written as $[\Gamma^-, \Gamma^+]$ and $[{}^i\Gamma'^-, {}^i\Gamma'^+]$, respectively. Additionally, let there be a solution vector \mathbf{x}_{opt} that reclines in Ω' and minimizes $f(\cdot)$, where $f(\mathbf{x}_{opt}) \leq f(\mathbf{x}) \forall \mathbf{x} \in \Omega$ is a global solution. For clarity, the following notations will be used in this work: i is the current time step (iteration), k is the index of the solution vector in the population matrix and j indicates the j^{th} design variable (feature x_j) of any solution vector \mathbf{x} .

The process of random feature selection depends either on the algorithm sampling a decision variable $x_j^i \in \mathbf{x}_j$ from the prominent domain Ω' or improvising and sampling randomly from the entire region Ω . This decision is dependent on the acceptance probability α .

Like any other metaheuristic algorithm, the first step is the initial population sampling. For a continuous landscape, the

population can be composed by:

$${}_k\mathbf{x} = \Gamma^- + {}_k\mathbf{u} \odot (\Gamma^+ - \Gamma^-). \tag{2}$$

In this equation, ${}_k\mathbf{u}$ is a vector of random coefficients sampled using one of the classical DOE methods. The \odot operator indicates an element-wise multiplication of vectors.

The matrix \mathbf{u}^i is sampled at each iteration i with the chosen DOE method. It contains the sampled random coefficients; its size is $\beta \times n$ where β is the size of the population and n is the number of dimensions of the problem.

$$[\mathbf{u}^i] = \begin{bmatrix} 1u_1^i & 1u_2^i & \dots & 1u_n^i \\ 2u_1^i & 2u_2^i & \dots & 2u_n^i \\ \vdots & \vdots & \ddots & \vdots \\ \beta u_1^i & \beta u_2^i & \dots & \beta u_n^i \end{bmatrix}. \tag{3}$$

After evaluating the initial population and determining the best solution \mathbf{x}_{best}^i , the algorithm estimates the domain-tightening coefficients to update the current prominent upper and lower bounds for each design feature as per the following expressions:

$${}^i\Gamma'^+ = \mathbf{x}_{best}^i + \eta^i, \quad {}^i\Gamma'^+ \leq \Gamma^+, \tag{4}$$

$${}^i\Gamma'^- = \mathbf{x}_{best}^i - \eta^i, \quad {}^i\Gamma'^- \geq \Gamma^-, \tag{5}$$

$$\eta^i = \frac{(1 - \alpha)\left(1 - \frac{i}{\gamma}\right)}{2} (\Gamma^+ - \Gamma^-). \tag{6}$$

The bandwidth (η^i) is updated every time \mathbf{x}_{best}^i is updated (self-adaptive mechanism). As can be seen in (6), we applied time-dependent bandwidth tightening using the term $(1 - \frac{i}{\gamma})$. This controls the greediness of the algorithm with time: as time passes, the algorithm becomes greedier. The $(1 - \alpha)$ term in (6) is used to maintain the $\alpha/(1 - \alpha)$ ratio (resembling the 80/20 analogy). To control the number of features (design variables) that will be sampled per solution from the prominent domain Ω' , the acceptance probability α has been used.

If the algorithm decides to draw from the prominent domain Ω' , the set of Eqs. (4) to (6) together with (7) are used to generate a feature as follows:

$${}_kx_j^i = {}^i_j\Gamma'^- + {}_k u_j^i ({}^i_j\Gamma'^+ - {}^i_j\Gamma'^-). \tag{7}$$

If the algorithm decides to draw a feature from the overall domain Ω , then it will instead be evaluated using:

$${}_kx_j^i = {}_j\Gamma^- + {}_k u_j^i ({}_j\Gamma^+ - {}_j\Gamma^-). \tag{8}$$

This recursive process runs until the stopping condition is satisfied, which we have set to the maximum number of evaluations γ . Algorithm 1 explains the steps of the algorithm.

Algorithm 1: Pareto-like sequential sampling (PSS)

```

Result: Minimize the objective function  $f(\mathbf{x})$ 
1 read parameter-settings;
2 pre-allocate and initialize memory containers;
3 begin
4   initialize a population sample  $\in \Gamma$ ;
5   evaluate the initial population;
6   for  $i = 1$  to  $\gamma$  do
7     // best solution in iteration  $i$  is  $\mathbf{x}_{best}^i$ 
8     find the current best  $\mathbf{x}_{best}^i$ ;
9     generate  $[\mathbf{u}^i]_{\beta \times n} \sim U(0, 1)$ ;
10    for  $k = 1$  to  $\beta$  do
11      if  $f(\mathbf{x}_{best}^i) < f(\mathbf{x}_{best}^{i-1})$  then
12        | update  $\eta^i$  and  $\Gamma^i$ ;
13      end
14      foreach  $x_j \in \mathbf{x}$  do
15        if  $rand \sim U(0, 1) \leq \alpha$  then
16          // sample from the prominent
17          | domain  $\Omega'$ 
18          | choose random  ${}_k x_j^i \in \Gamma_j^i$ ;
19        else
20          // sample from the overall
21          | domain  $\Omega$ 
22          | choose random  ${}_k x_j^i \in \Gamma_j$ ;
23        end
24      end
25       $[population]_{\beta \times n} \leftarrow {}_k \mathbf{x}$ ;
26      evaluate  $f({}_k \mathbf{x})$ ;
27    end
28  end
29  return  $\mathbf{x}_{best}$ ;

```

3 Parameters setting

The no free lunch (NFL) theorem directly promotes that no algorithm has a global parameters set that ensures an overall good performance for all the possible optimisation problems (Arcuri and Fraser 2013). In line with this, we provide a comprehensive theoretical and visual explanation of the parameter-setting (tuning) for the PSS algorithm. A survey about traditional tuning techniques can be found in Eiben et al. (1999). The PSS algorithm is based on three input parameters that need to be set by the user. Namely, the population size (β), the maximum number of iterations (γ) and the acceptance probability (α). The taxonomy and terminology used in this section are taken from Eiben et al. (1999). The derived formulae here follow probabilistic rules that can be found in standard textbooks about the theory of probability (for example see Capiński and Zastawniak (2001)). In this section, we provide an analytical investigation of the interaction between the three parameters and illustrate their effects on finding the global optimum by means of a numerical example. Apart from this, the self-adaptive mechanism

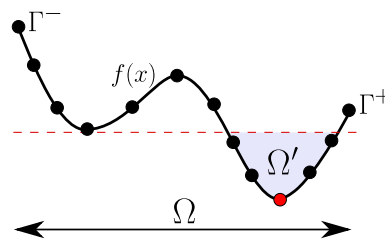


Fig. 2 A virtual 1D optimization objective $f(x)$ – discretized space

for the prominent domain will not be discussed here (see Sect. 2.3).

Let $f(x)$ be a 1D discrete optimization problem with N finite possible solutions that we need to minimize. Figure 2 shows the distribution of x along the domain Ω . Now, let Ω' be the true¹ prominent domain around the optimum solution x_{opt} . The red-dashed line, shown in Fig. 2, represents the lowest valley of all local minima, wherein all the points below this line are better than any local optima above it. If any solution below this line is chosen, it will always lead to the global solution in the coming generations by using the *gbest* topology. Assuming that n' is the number of all possible points below the line $\in \Omega'$ and N is the number of all possible solutions $\in \Omega$, the probability of finding a solution inside the true prominent domain in the first step (initial population sampling) is:

$$P(x \in \Omega') = \frac{n'}{N}. \tag{9}$$

We can expand (9) to determine the probability of drawing at least one solution that lays in Ω' (event A_0) from β drawn solutions:

$$P(A_0) = 1 - \left(1 - \frac{n'}{N}\right)^\beta. \tag{10}$$

Equation (10) explains the exponential relationship between the landscape size and the initial population size. The probability of event A_0 is a problem-specific property that differs per objective and independent design variable. The complementary event to A_0 will be called B_0 , wherein $P(B_0) = 1 - P(A_0)$ and $P(B_0)$ means none of the drawn β solutions initially is located in the true prominent region.

After the population initialization, the algorithm performance phase begins. In this phase, the algorithm tests the acceptance-rejection paradigm against α for each solution component ($\forall x_j^i \in \mathbf{x}^i$). If a random number $r^i \sim U(0, 1) \leq \alpha$, the algorithm will sample the design component from the prominent region, else it will be sampled from the overall

¹ This is an innate constitutive property of the distribution (the change) of $f(x)$ vs. any independent $x_j \in \mathbf{x}$.

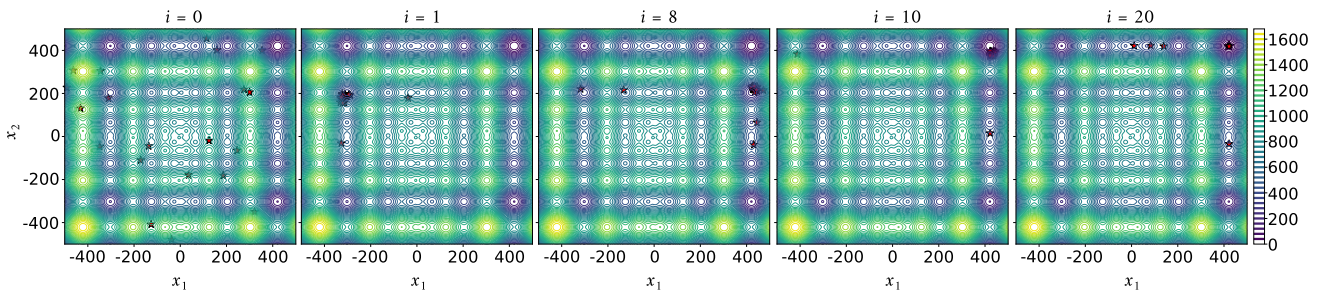


Fig. 3 The population history for Case (a)—2D Schwefel function

region. The probability of exploring new solutions $\in \Omega'$ from the domain is defined in Eq. (11). This equation is independent of the time step, i , and it is always constant if the population size β is constant for a specific optimization problem and an independent design variable.

$$P^i(x \in \Omega' | r^i > \alpha) = \left[1 - \left(1 - \frac{n'}{N} \right)^\beta \right] (1 - \alpha). \quad (11)$$

Now if the algorithm were to intensificate (success against α), the probability of getting a better solution within the current fictitious prominent region is shown in the following:

$$P^i(x_{\text{best}}^{i+1} < x_{\text{best}}^i | r^i \leq \alpha) = \left[1 - \left(1 - \frac{n_i^i}{N_i^i} \right)^\beta \right] \alpha. \quad (12)$$

Notice that Eq. (12) is time-dependent. The terms n_i^i and N_i^i by definition depend on the sub-region (the current fictitious prominent region), and they are a function of time (notice the i superscription). This can be imagined as if the red-dotted line corresponds to the current best solution and the algorithm tries to find a better sub-optimal. Put differently, it implies—only in this algorithm—that the intensification property is time-dependent and it resembles a perturbation problem.

In real life, the algorithm is not aware of the true prominent domain. Indeed, the only parameters known to the algorithm are the domain and the objective function (as an evaluator). We assume that the best-known solution so far lays in the prominent domain and hope that the algorithm’s exploration will find a better one in the upcoming generations (if any).

If we have an optimization problem with n dimensions where $n > 1$, we always assume independent design variables. For expanding the reflections made above for n design variables, the probability of obtaining an optimal solution can be calculated using the multiplication rule, wherein every design variable is treated as an independent event. This simple probabilistic analysis shows that the algorithm is dependent on the distribution of the objective function, though because the algorithmic parameters are interactive, they must be carefully set. From this analysis, we can see

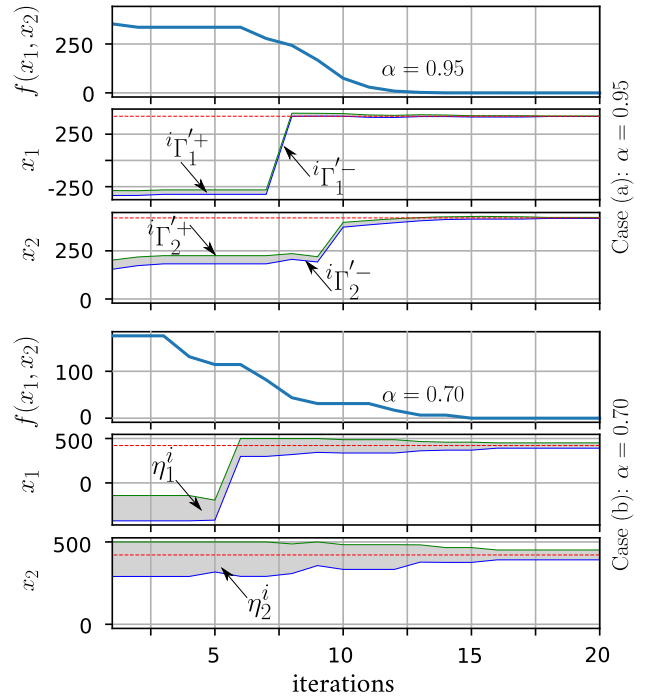


Fig. 4 Domain tightening and convergence histories—2D Schwefel function

that with a bigger β , we boost the probability of obtaining solutions that belong to the true prominent region, i.e., global optimum. With a higher α , the algorithm puts most of its effort into intensification, weakening the diversification. For multimodal and nonconvex problems, α should be smaller than in convex and unimodal problems. Using more iterations will always increase the probability of exploring better solutions. Finally, it is worth mentioning that we used only MC sampling in the presented analysis. Unless mentioned otherwise, we used a population of 30 and an acceptance probability of 0.95.

3.1 Numerical illustration

The example shown in Figs. 3 and 4 indicates the solution of the standard 2D Schwefel problem. The optimum solution

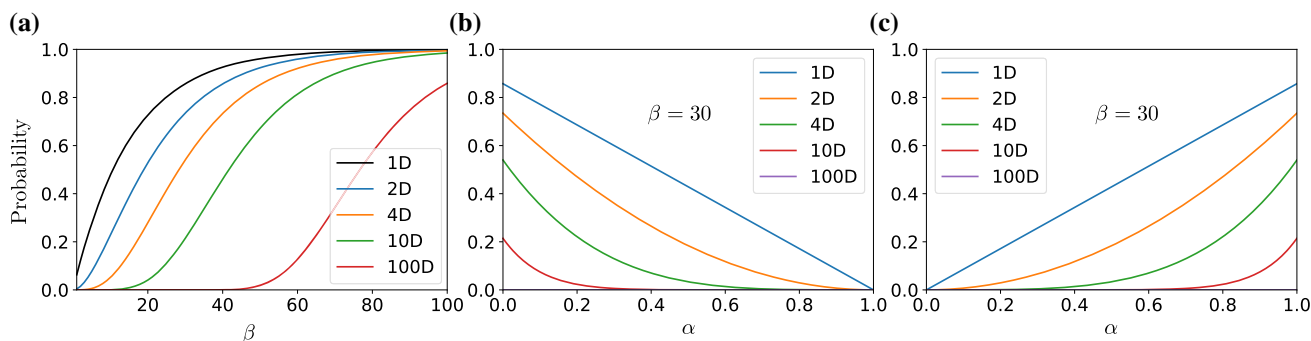


Fig. 5 **a** The change of $P(A_0)$ with β ; **b** the diversification probability versus α in Ω ; **c** the intensification probability versus α in Ω' — n D Schwefel function for $n = 1, n = 2, n = 4, n = 10$, and $n = 100$

$f(\mathbf{x}_{opt}) = 0.0$ is obtained when both x_1 and x_2 hit 420.9687 (the red-dotted lines in Fig. 4). A population of 30 members and a sum of 20 iterations were used to solve this problem. For this problem, we set $\alpha = 0.95$ and 0.70 . Figure 4 reveals how the algorithm adaptively changed the prominent search area Ω' (the gray fill-up) and dynamically tightened it to control the greediness of the algorithm with respect to the remaining time (for the two cases).

According to our definition of the true prominent region for Schwefel function $\Omega' \in [389.33, 452.16] \forall x_j$ (determined graphically), we say that the algorithm has globally converged if the algorithm successfully found both x_1 and $x_2 \in \Omega'$. We analyzed 30 consecutive runs for each case, and the converged results are expressed as: average \pm standard deviation (success rate %). For Case (a), the results were 0.197345 ± 0.835687 (83.33%), and for Case (b), the results were 2.435370 ± 2.599124 (96.67%). These results highlight the trade-off balance between the diversification and the intensification processes explained in Eqs. (11) and (12). Side by side, Fig. 5a reveals the effect of choosing the population size (β) on the probability of finding better initial solutions (event A_0). Moreover, Fig. 5b, c explain the global diversification ($\forall x_i \in \Omega$) and the intensification inside the true prominent domain ($\forall x_i \in \Omega'$) for different dimensions of Schwefel function, respectively.

4 Benchmarks and comparisons

The presented PSS algorithm was used to solve a set of standard functions (Sect. 4.1), the CEC2017 composite functions (Sect. 4.3), and another selected set of engineering problems (Sect. 4.4). We benchmarked and compared the performance of the PSS algorithm with the obtained results from other state-of-the-art algorithms, the results of which were mostly retrieved directly from the source papers and not re-simulated herein. To obtain a meaningful comparison, we used the same number of evaluations per problem as were used by the

authors of the other algorithms. This benchmarking approach follows the recommendations by Arcuri and Fraser (2013) and Liao et al. (2015), which were recently reemphasized by Ser et al. (2019) and used by Piotrowski and Napiorkowski (2018).

For the standard benchmarks, we chose state-of-the-art algorithms to compare with and avoided using the classical and outperformed algorithms following the suggestions by Ser et al. (2019) and Molina et al. (2018). The first algorithm we used in the benchmarking was the whale optimization algorithm (WOA) as this is one of the recent and heavily cited algorithms in the literature, which outperformed many other recent algorithmic proposals (see Mirjalili and Lewis 2016). The second one was the pathfinder algorithm (PFA), which has also been published recently and shown to outperform many other algorithms (see Yapici and Cetinkaya 2019).

For engineering benchmarks, we compared the obtained solution by directly adopting the best results obtained from different algorithms in the literature. We also solved a recent engineering case study with high dimensions that was solved by using the modified parameter-setting-free harmony search (MPSFHS) algorithm by Shaqfa and Orbán (2019) and we used the same algorithm to benchmark the new algorithm with regard to its scalability.

4.1 Standard benchmarks

Here, we compare our proposed algorithm with the WOA (Mirjalili and Lewis 2016) and the PFA Yapici and Cetinkaya (2019) using the standard benchmarks explained in Table 1. We carefully chose the benchmarks in this paper to cover a wide range of problems.

In Table 2, we compared the WOA with our proposed PSS method. The problems were solved by assuming $\alpha = 0.95$. For the population size and the number of iterations, we chose the same values as in Mirjalili and Lewis (2016), i.e., a population size of 30 and a total of 500 iterations.

Table 1 Standard benchmark functions (refer to Jamil and Yang 2013 for details)

Function	Expression
Sphere	$f_1(\mathbf{x}) = \sum_{i=1}^n x_i^2$
Sum squares	$f_2(\mathbf{x}) = \sum_{i=1}^n ix_i^2$
Chung Reynolds	$f_3(\mathbf{x}) = \left(\sum_{i=1}^n x_i^2\right)^2$
Schwefel2.21	$f_4(\mathbf{x}) = \max_{i=1,\dots,n} x_i $
Schwefel2.22	$f_5(\mathbf{x}) = \sum_{i=1}^n x_i + \prod_{i=1}^n x_i $
Rosenbrock	$f_6(\mathbf{x}) = \sum_{i=1}^n [100(x_{i+1} - x_i^2)^2 + (1 - x_i)^2]$
Trid 6	$f_7(\mathbf{x}) = \sum_{i=1}^n (x_i - 1)^2 - \sum_{i=2}^n x_i x_{i-1}$
Zakharov	$f_8(\mathbf{x}) = \sum_{i=1}^n x_i^2 + \left(\sum_{i=1}^n 0.5ix_i\right)^2 + \left(\sum_{i=1}^n 0.5ix_i\right)^4$
Griewank	$f_9(\mathbf{x}) = 1 + \sum_{i=1}^n \frac{x_i^2}{4000} - \prod_{i=1}^n \cos\left(\frac{x_i}{\sqrt{i}}\right)$
Ackley	$f_{10}(\mathbf{x}) = -20\exp\left(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^n x_i^2}\right) - \exp\left(\frac{1}{n}\sum_{i=1}^n \cos(2\pi x_i)\right) + 20 + \exp(1)$
Schwefel	$f_{11}(\mathbf{x}) = 418.9829n - \sum_{i=1}^n x_i \sin(\sqrt{ x_i })$
Shubert	$f_{12}(\mathbf{x}) = \prod_{i=1}^n \left(\sum_{j=1}^5 \cos((j+1)x_i + j)\right)$
Six-hump camel	$f_{13}(x_1, x_2) = \left(4 - 2.1x_1^2 + \frac{x_1^4}{3}\right)x_1^2 + x_1x_2 + (-4 + 4x_2^2)x_2^2$
Goldstein	$f_{14}(x_1, x_2) = [1 + (x_1 + x_2 + 1)^2(19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + \dots 3x_2^2)][30 + (2x_1 - 3x_2)^2(18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2)]$
De Jong 5	$f_{15}(x_1, x_2) = \left(\frac{1}{500} + \sum_{j=1}^{25} \left(j + \sum_{i=1}^2 (x_i - a_{ij})^6\right)^{-1}\right)^{-1}$
Hartmann 3	$f_{16}(x_1, x_2, x_3) = -\sum_{i=1}^4 \alpha_i \exp\left(-\sum_{j=1}^3 A_{ij}(x_j - P_{ij})^2\right)$

Table 2 Comparison with WOA Mirjalili and Lewis (2016)

Function	Proposed approach	WOA Mirjalili and Lewis (2016)	$f(\mathbf{x}_{opt})$
f_1	0.775955 ± 0.222759	1.41E-30 ± 4.91E-30	0
f_4	5.154556 ± 2.273995	0.072581 ± 0.39747	0
f_5	0.759658 ± 0.251091	1.06E-21 ± 2.39E-21	0
f_6	26.778816 ± 3.804910	27.86558 ± 0.763626	0
f_9	0.809812 ± 0.088605	0.000289 ± 0.001586	0
f_{10}	2.230591 ± 0.810609	7.4043 ± 9.897572	0
f_{11}^\dagger	- 12554.89 ± 33.2842	- 5080.76 ± 695.7968	- 12569.48
f_{13}	- 1.031611 ± 3.73E-05	- 1.03163 ± 4.2E-07	- 1.0316
f_{15}	0.998004 ± 5.98E-10	2.111973 ± 2.498594	0.998004 ≈ 1

[†]This version of Schwefel min $f(\mathbf{x}_{opt}) = -418.9829 \times 30$ (see Mirjalili and Lewis 2016)

Table 3 Comparison with PFA Yapici and Cetinkaya (2019)

Function	Proposed approach	PFA Yapici and Cetinkaya (2019)	$f(x_{opt})$
f_2	0.117980 ± 0.048295	5.5674E-25 ± 7.9092E-25	0
f_3	0.031421 ± 0.016130	9.9813E-46 ± 3.3585E-45	0
f_5	0.437154 ± 0.249896	3.4831E-14 ± 6.2094E-14	0
f_7	- 49.996395 ± 0.004986	- 50.0000 ± 1.98E-11	- 50 [†]
f_8	0.081319 ± 0.027990	11.5480 ± 12.9802	0
f_9	0.425310 ± 0.408112	0.0006 ± 0.0012	0
f_{11}	0.610558 ± 0.145078	3.1549E+3 ± 5.6274E+2	0
f_{14}	3.000043 ± 7.51E-5	3.0000 ± 2.4952E-16	3
f_{16}	- 3.855772 ± 0.009925	- 3.8628 ± 1.5026E-15	- 3.8628

[†]This value has been evaluated for $n = 6$ (as in Yapici and Cetinkaya 2019)

Table 4 Scalability with PFA Yapici and Cetinkaya (2019)

Function	n	Proposed approach	PFA Yapici and Cetinkaya (2019)	$f(x_{opt})$
f_8	10	0.0143 ± 0.0206	2.17E-35 ± 1.09E-34	0
	50	53.6161 ± 11.3012	343.1244 ± 92.1694	0
	100	680.8374 ± 61.9845	1911.1759 ± 195.2386	0
f_{11}	10	0.3395 ± 0.91820	449.5090 ± 184.2821	0
	50	129.6466 ± 82.8948	6423.8391 ± 663.2207	0
	100	7208.6969 ± 489.7975	13610.2227 ± 1361.9760	0

Table 5 Comparison with MPSFHS Shafqa and Orbán (2019)

Function	Proposed approach	MPSFHS Shafqa and Orbán (2019)	$f(x_{opt})$
f_8	173.1648 ± 17.1974	1181.5644 ± 71.3395	0
f_{11}	971.8366 ± 150.2988	1554.8070 ± 178.5386	0

In general, 30 dimensions were used for all the proposed problems ($n = 30$) except for f_{13} and f_{15} (see Table 1).

The reported results in Table 2 express the average ± the standard deviation for 25 consecutive runs per problem (as in Mirjalili and Lewis 2016). The boldfaced results indicate—for each problem—the algorithm that performed the best in the comparison. From Table 2, it can be seen that in easy unimodal and convex functions, the best algorithm was WOA. This behaviour was expected due to the weak intensification in the PSS algorithm that the crude random walk usually reveals. For harder problems, such as nonconvex and/or multimodal ones, though, the PSS algorithm was better able to allocate global optima—this finding holds notably for the Schwefel function f_{11} .

In Table 3, we compare the PSS algorithm to the recently proposed PFA algorithm (Yapici and Cetinkaya 2019). As in their work, the problems in this table were simulated using a population size of 30 and 1000 iterations. For the PSS algorithm, α was set to 0.95 as per the previous comparison. However, the number of dimensions used for f_8 was 20, while for f_7 $n = 6$, and for f_{14} and f_{16} , n was set to 2 and 3, respectively, as shown in Table 1. The stated results are for 30 consecutive runs per function (as in Yapici and Cetinkaya 2019).

As can be seen in Table 3, f_2 and f_5 behaved the same as in Table 2 (the pronounced weak intensification for the proposed approach). On the other hand, the results of f_8 and f_{11} , see Table 3, were outperformed by the Pareto-like sampling. The results of f_7 , f_{14} , and f_{16} are almost identical.

4.2 Testing scalability

With an increased dimensionality, the problem complexity increases, and accordingly, the search domain expands exponentially. To test the scalability of the current algorithm, we compared the behaviour of the proposed approach with the PFA algorithm for f_8 and f_{11} functions. The simulations were run 30 times for each, with 10, 50, and 100 dimensions as illustrated in Table 4. As the results suggest, the proposed PSS algorithm outperformed the PFA and registered fewer deteriorations in performance with the increase of dimensions.

As this test implies that increased dimensionality requires more computational capacity, we ran simulations with the recently modified parameter-setting-free harmony search algorithm (MPSFHS) and the proposed PSS approach to see how increasing the iterations could enhance the results. To do this, f_8 and f_{11} were solved for $n = 100$ and a total of 300,000 evaluations (10,000 iterations with a popula-

Table 6 Composite benchmark functions—adopted from CEC2017 Wu et al. (2016). The definitions of the base functions (F_N) are defined in CEC2017

Function definition	$f(\mathbf{x}_{\text{opt}})$
$f_{c,1}(\mathbf{x}) = \begin{cases} F_1 = \text{Rotated and Shifted Rosenbrock's Function,} \\ F_2 = \text{Rotated and Shifted High Conditioned Elliptic Function,} \\ F_3 = \text{Rotated and Shifted Rastrigin's Function,} \\ \sigma = [10, 20, 30], \lambda = [1, 1\text{E}+06, 1], \text{bias} = [0, 100, 200]. \end{cases}$	2100
$f_{c,2}(\mathbf{x}) = \begin{cases} F_1 = \text{Rotated and Shifted Rastrigin's Function,} \\ F_2 = \text{Rotated and Shifted Griewank's Function,} \\ F_3 = \text{Rotated and Shifted Modified Schwefel's Function,} \\ \sigma = [10, 20, 30], \lambda = [1, 10, 1], \text{bias} = [0, 100, 200]. \end{cases}$	2200
$f_{c,3}(\mathbf{x}) = \begin{cases} F_1 = \text{Rotated and Shifted Rosenbrock's Function,} \\ F_2 = \text{Rotated and Shifted Ackley's Function,} \\ F_3 = \text{Rotated and Shifted Modified Schwefel's Function,} \\ F_4 = \text{Rotated and Shifted Rastrigin's Function,} \\ \sigma = [10, 20, 30, 40], \lambda = [1, 10, 1, 1], \text{bias} = [0, 100, 200, 300]. \end{cases}$	2300
$f_{c,4}(\mathbf{x}) = \begin{cases} F_1 = \text{Rotated and Shifted Ackley's Function,} \\ F_2 = \text{Rotated and Shifted High Conditioned Elliptic Function,} \\ F_3 = \text{Rotated and Shifted Griewank's Function,} \\ F_4 = \text{Rotated and Shifted Rastrigin's Function,} \\ \sigma = [10, 20, 30, 40], \lambda = [1, 1\text{E}+06, 10, 1], \text{bias} = [0, 100, 200, 400]. \end{cases}$	2400
$f_{c,5}(\mathbf{x}) = \begin{cases} F_1 = \text{Rotated and Shifted Rastrigin's Function,} \\ F_2 = \text{Rotated and Shifted HappyCat Function,} \\ F_3 = \text{Rotated and Shifted Ackley's Function,} \\ F_4 = \text{Rotated and Shifted Discus Function,} \\ F_5 = \text{Rotated and Shifted Rosenbrock's Function,} \\ \sigma = [10, 20, 30, 40, 50], \lambda = [10, 1, 10, 1\text{E}+06, 1], \\ \text{bias} = [0, 100, 200, 300, 400]. \end{cases}$	2500
$f_{c,6}(\mathbf{x}) = \begin{cases} F_1 = \text{Rotated and Shifted Expanded Scaffer's Function,} \\ F_2 = \text{Rotated and Shifted Modified Schwefel's Function,} \\ F_3 = \text{Rotated and Shifted Griewank's Function,} \\ F_4 = \text{Rotated and Shifted Rosenbrock's Function,} \\ F_5 = \text{Rotated and Shifted Rastrigin's Function,} \\ \sigma = [10, 20, 30, 40], \lambda = [1\text{E}+26, 10, 1\text{E}+06, 10, 5\text{E}+04], \\ \text{bias} = [0, 100, 200, 300, 400]. \end{cases}$	2600
$f_{c,7}(\mathbf{x}) = \begin{cases} F_1 = \text{Rotated and Shifted HGBat Function,} \\ F_2 = \text{Rotated and Shifted Rastrigin's Function,} \\ F_3 = \text{Rotated and Shifted Modified Schwefel's Function,} \\ F_4 = \text{Rotated and Shifted BentCigar Function,} \\ F_5 = \text{Rotated and Shifted High Conditioned Elliptic Function,} \\ F_6 = \text{Rotated and Shifted Expanded Scaffer's Function,} \\ \sigma = [10, 20, 30, 40, 50, 60], \lambda = [10, 10, 2.5, 1\text{E}+26, 1\text{E}+06, 5\text{E}+04], \\ \text{bias} = [0, 100, 200, 300, 400, 500]. \end{cases}$	2700
$f_{c,8}(\mathbf{x}) = \begin{cases} F_1 = \text{Rotated and Shifted Ackley's Function,} \\ F_2 = \text{Rotated and Shifted Griewank's Function,} \\ F_3 = \text{Rotated and Shifted Discus Function,} \\ F_4 = \text{Rotated and Shifted Rosenbrock's Function,} \\ F_5 = \text{Rotated and Shifted HappyCat Function,} \\ F_6 = \text{Rotated and Shifted Expanded Scaffer's Function,} \\ \sigma = [10, 20, 30, 40, 50, 60], \lambda = [10, 10, 1\text{E}+06, 1, 1, 5\text{E}+04], \\ \text{bias} = [0, 100, 200, 300, 400, 500]. \end{cases}$	2800
$f_{c,9}(\mathbf{x}) = \begin{cases} F_1 = \text{Hybrid Function 5,} \\ F_2 = \text{Hybrid Function 8,} \\ F_3 = \text{Hybrid Function 9,} \\ \sigma = [10, 30, 50], \lambda = [1, 1, 1], \text{bias} = [0, 100, 200]. \end{cases}$	2900
$f_{c,10}(\mathbf{x}) = \begin{cases} F_1 = \text{Hybrid Function 5,} \\ F_2 = \text{Hybrid Function 6,} \\ F_3 = \text{Hybrid Function 7,} \\ \sigma = [10, 30, 50], \lambda = [1, 1, 1], \text{bias} = [0, 100, 200]. \end{cases}$	3000

Table 7 CEC2017 composite functions tested with $n = 2$ by using only 10 iterations (only for the PSS, PFA, WOA, and PSO and equivalent evaluations for the MPSFHS) for $f_{c,1} \rightarrow f_{c,8}$ – reported the error values here ($f_{c,i}(\mathbf{x}_i) - f_{c,i}(\mathbf{x}_{opt})$)

Algorithm	$f_{c,1}$	$f_{c,2}$	$f_{c,3}$	$f_{c,4}$	$f_{c,5}$	$f_{c,6}$	$f_{c,7}$	$f_{c,8}$
PSS	Min	2.4285E-02	6.0717E-02	7.2868E+00	2.8464E-01	6.2283E-02	6.4629E-01	5.5129E+00
	Max	1.0018E+02	1.1226E+02	3.0392E+02	2.0635E+02	2.1118E+02	2.0057E+02	1.8914E+02
	Median	4.1376E+00	1.2610E+01	7.0064E+01	1.0059E+02	1.0614E+02	1.7070E+00	2.6410E+00
	Mean	1.9017E+01	2.5048E+01	9.4231E+01	1.0793E+02	1.1002E+02	2.7532E+01	2.7133E+01
	Std	3.1468E+01	3.3741E+01	1.1463E+02	5.5402E+01	6.1274E+01	5.2681E+01	5.0566E+01
PFA	Min	5.1962E-02	2.0261E-02	7.8146E-01	2.1461E+01	6.7842E+00	8.8438E-02	1.3907E+00
	Max	1.0021E+02	1.0230E+02	3.0303E+02	2.0244E+02	3.0168E+02	2.0858E+02	4.0021E+02
	Median	4.0589E-01	8.0329E-01	9.6471E+00	1.0142E+02	1.0713E+02	1.9240E+00	3.6875E+00
	Mean	4.2739E+00	9.5696E+00	9.5648E+01	1.0874E+02	1.1802E+02	9.1956E+00	1.7240E+01
	Std	1.8221E+01	2.5488E+01	1.3689E+02	3.7036E+01	8.2871E+01	3.7697E+01	7.2359E+01
WOA	Min	3.9120E-02	4.6032E-02	3.8543E-01	1.4056E+01	3.4195E+00	3.7880E-01	2.4574E+00
	Max	1.1761E+02	2.3130E+01	3.0207E+02	2.1358E+02	3.0999E+02	2.0312E+02	1.8917E+02
	Median	6.1233E-01	1.0885E+01	5.3677E+00	1.0070E+02	1.1937E+02	5.3988E+00	2.4332E+01
	Mean	4.9854E+00	8.9399E+00	1.0609E+02	1.0771E+02	1.3170E+02	2.2009E+01	4.8168E+01
	Std	2.1319E+01	5.9659E+00	1.2264E+02	3.3913E+01	9.4636E+01	4.6116E+01	5.4586E+01
MPSFHS	Min	1.3582E+00	3.0120E-01	8.8972E+00	9.2268E+01	3.1234E+01	1.6434E-01	3.5703E+00
	Max	6.0320E+01	1.1324E+02	3.0430E+02	2.0983E+02	3.1993E+02	1.4941E+02	7.8858E+01
	Median	1.9290E+01	2.4465E+01	7.5047E+01	1.3201E+02	2.0406E+02	4.1046E+01	2.8917E+01
	Mean	2.1165E+01	3.0313E+01	1.0591E+02	1.4425E+02	1.8885E+02	4.8282E+01	3.3665E+01
	Std	1.7733E+01	2.3374E+01	8.9850E+01	3.8189E+01	7.9865E+01	3.5734E+01	1.9670E+01
PSO	Min	3.0762E-01	6.8406E-01	2.4319E+00	7.1158E+01	1.1524E+01	2.9869E+00	5.9283E+00
	Max	1.0109E+02	1.7330E+01	3.0334E+02	1.6073E+02	2.8149E+02	2.2489E+02	4.0218E+02
	Median	5.8090E+00	7.5109E+00	2.6452E+01	1.1033E+02	1.4209E+02	2.0422E+01	2.1516E+01
	Mean	9.4914E+00	8.2878E+00	1.1008E+02	1.1224E+02	1.4144E+02	2.7641E+01	3.4890E+01
	Std	1.7832E+01	4.8397E+00	1.3079E+02	1.5796E+01	7.7616E+01	3.8640E+01	7.0202E+01

Table 8 CEC2017 composite functions tested with $n = 100$ by using only 100 iterations for both the PSS and PFA algorithms on $f_{c,1} \rightarrow f_{c,10}$ —reported the error values here ($f_{c,i}(\mathbf{x}_i) - f_{c,i}(\mathbf{x}_{\text{opt}})$)

Algorithm		$f_{c,1}$	$f_{c,2}$	$f_{c,3}$	$f_{c,4}$	$f_{c,5}$
PSS	Min	1.2002E+03	2.5724E+04	1.5424E+03	2.1423E+03	5.1997E+03
	Max	1.7306E+03	3.1503E+04	1.8003E+03	2.5840E+03	9.2997E+03
	Median	1.5094E+03	2.8014E+04	1.6500E+03	2.3455E+03	7.0022E+03
	Mean	1.4929E+03	2.7920E+04	1.6588E+03	2.3551E+03	7.0311E+03
	Std	1.0472E+02	1.3480E+03	7.0691E+01	1.1244E+02	1.0787E+03
		$f_{c,6}$	$f_{c,7}$	$f_{c,8}$	$f_{c,9}$	$f_{c,10}$
	Min	1.6692E+04	1.5568E+03	4.9472E+03	5.1738E+03	1.7476E+07
	Max	1.9222E+04	2.2142E+03	1.2336E+04	7.4074E+03	8.2332E+07
	Median	1.7630E+04	1.6990E+03	8.4480E+03	6.3363E+03	4.2002E+07
	Mean	1.7686E+04	1.7340E+03	8.7575E+03	6.2897E+03	4.2233E+07
Std	8.4035E+02	1.4641E+02	1.9430E+03	5.6462E+02	1.6223E+07	
PFA		$f_{c,1}$	$f_{c,2}$	$f_{c,3}$	$f_{c,4}$	$f_{c,5}$
	min	1.3493E+03	2.4074E+04	2.3835E+04	2.1928E+03	5.6223E+03
	Max	1.9686E+03	3.2000E+04	3.3604E+04	2.8422E+03	1.6262E+04
	Median	1.5668E+03	2.6900E+04	2.8106E+04	2.5436E+03	6.9669E+03
	Mean	1.6031E+03	2.7021E+04	2.7941E+04	2.5141E+03	7.7530E+03
	Std	1.5386E+02	2.2703E+03	2.7965E+03	1.7907E+02	2.2391E+03
		$f_{c,6}$	$f_{c,7}$	$f_{c,8}$	$f_{c,9}$	$f_{c,10}$
	Min	1.6504E+04	1.1039E+03	7.0900E+03	5.7046E+03	4.2582E+07
	Max	2.2592E+04	1.8703E+03	1.6716E+04	1.0894E+04	2.0416E+08
	Median	1.9795E+04	1.4006E+03	1.1057E+04	7.7024E+03	9.5209E+07
Mean	1.9688E+04	1.4345E+03	1.1339E+04	7.8995E+03	1.0203E+08	
Std	1.5073E+03	1.8877E+02	2.3745E+03	1.2773E+03	4.5804E+07	

tion size of 30 for PSS). Table 5 illustrates the results of 30 consecutive runs for each algorithm. The PSS approach outperformed the MPSFHS algorithm, and considerable enhancements were seen in the end results.

4.3 Composite benchmarks

In this section, we used the CEC2017 by Wu et al. (2016) and reviewed by Molina et al. (2018), competition's composite functions to benchmark the behaviour of the proposed algorithm (PSS). The used functions are briefly described in Table 6. We first tested our algorithm against these functions under extreme cases where only a few iterations are allowed. In Table 7, we ran 30 consecutive tests with $n = 2$, $\alpha = 0.95$, $\beta = 30$, and with allowing a total of 10 iterations per run for the composite functions $f_{c,1} \rightarrow f_{c,8}$. We conducted a comparison under the same conditions for the PSS, PFA, WOA, and PSO algorithms. The MPSFHS was tuned to $HMS = 30$, $m = 2$, $HMCR_i = 0.75$, $PAR_i = 0.15$, $HMCR_{\max} = 0.99$, $PAR_{\min} = 0.05$, and 300 iterations to have an equivalent number of evaluations as the PSS, PFA, WOA, and PSO (for more about the used parameter-settings refer to Shaqfa and Orbán (2019)).

The results shown in Table 7 suggest that the PSS algorithm can still be globally convergent and allocate global

minima in most of the runs and it scales well with the available time (iterations). Indeed it scored the best results for most functions, though, the gap between the PSS and the PFA algorithms was small. The best runs are revealed in Fig. 6 showing two iteration-milestones $i = 5$ and $i = 10$. Notice that the reported results in Table 7 are expressed in terms of the error values and computed as ($f_{c,i}(\mathbf{x}_i) - f_{c,i}(\mathbf{x}_{\text{opt}})$) and we revealed the minimum, maximum, mean, median, and the standard deviation of the error values as recommended by the CEC2017 report (refer to Wu et al. 2016).

In another extreme case we tested the functions $f_{c,1} \rightarrow f_{c,10}$ but this time with $n = 100$ and by only employing 100 iterations. In this experiment, we only compared the PSS with the PFA algorithm. The shown results in Table 8 suggest that the gap between the PSS and the PFA widens. Indeed, the PSS outperformed the PFA in all the functions except $f_{c,2}$ and $f_{c,7}$. These results signify the capability of the PSS algorithm to scale with high-dimensional problems while exploiting very limited computational resources. This is quite important for complex surrogate models that require a significant computational capacity for each functional evaluation (Forrester and Keane 2009; Queipo et al. 2005).

We also conducted simulations similar to the ones presented by Yapici and Cetinkaya (2019) and again we compared our results with the adopted ones directly from their

Table 9 Composite functions comparison with PFA algorithm Yapici and Cetinkaya (2019)—reported here the error values ($f_{c,i}(x_i) - f_{c,i}(x_{opt})$)

Method	n	Error	$f_{c,1}$	$f_{c,2}$	$f_{c,3}$	$f_{c,4}$	$f_{c,5}$	$f_{c,6}$	$f_{c,7}$	$f_{c,8}$	$f_{c,9}$	$f_{c,10}$
PFA Yapici and Cetinkaya (2019)	10	Min	1.0000E+02	1.1563E+01	3.0014E+02	1.0006E+02	3.9774E+02	3.0000E+02	3.8848E+02	3.0000E+02	2.4124E+02	2.8612E+03
		Max	2.2019E+02	1.0311E+02	3.1846E+02	3.4380E+02	4.4602E+02	3.0000E+02	3.9570E+02	4.4498E+02	3.2367E+02	9.2812E+04
		Median	1.5467E+02	1.0137E+02	3.0762E+02	1.6547E+02	4.0078E+02	3.0000E+02	3.9090E+02	3.0277E+02	2.8414E+02	1.7656E+04
		Mean	1.5645E+02	9.3593E+01	3.0706E+02	2.1330E+02	4.1863E+02	3.0000E+02	3.9175E+02	3.4087E+02	2.8418E+02	6.2390E+04
		Std	5.5756E+01	2.4340E+01	4.3861E+00	9.8189E+01	2.1735E+01	2.1902E-05	1.9096E+00	6.3004E+01	1.5359E+01	1.5105E+04
	50	Min	2.0778E+02	1.0000E+02	3.7323E+02	5.2936E+02	3.8300E+02	2.0782E+02	4.8804E+02	4.0528E+02	5.6256E+02	2.4685E+03
		Max	4.2379E+02	7.5775E+03	5.2140E+02	5.8619E+02	4.3544E+02	2.6365E+03	5.1048E+02	4.7741E+02	1.0563E+02	1.5743E+05
		Median	3.1604E+02	1.9984E+03	4.9308E+02	5.7169E+02	3.8810E+02	9.5600E+02	5.0277E+02	4.3275E+02	6.8022E+02	5.7236E+04
		Mean	3.2029E+02	3.3880E+03	4.8791E+02	5.7049E+02	3.8989E+02	9.5505E+02	5.0281E+02	4.3839E+02	7.1405E+02	6.4199E+04
		Std	5.5865E+01	3.3643E+03	2.6840E+01	1.2125E+01	7.3578E+00	6.2841E+02	2.5284E+00	2.5585E+01	1.1731E+02	3.0136E+04
PSS	10	Min	2.9169E+02	1.0000E+02	5.1701E+02	5.5598E+02	5.4014E+02	3.5236E+02	5.0000E+02	5.1385E+02	4.6940E+02	1.1016E+06
		Max	6.5178E+02	1.4177E+04	7.7342E+02	8.3818E+02	6.7421E+02	4.1024E+03	6.8090E+02	6.5809E+02	1.4784E+03	1.8261E+06
		Median	5.6687E+02	1.3255E+04	7.2031E+02	8.1295E+02	6.0900E+02	2.0229E+03	6.2269E+02	5.5260E+02	8.7184E+02	1.4303E+06
		Mean	5.1824E+02	1.3002E+04	7.0909E+02	7.9178E+02	6.1075E+02	2.2147E+03	6.2270E+02	5.6326E+02	9.0843E+02	1.4301E+06
		Std	1.0989E+01	1.9255E+03	4.9710E+01	6.0605E+01	2.1206E+01	6.3598E+02	3.8000E+01	3.0291E+01	2.5833E+02	1.5398E+05
	30	Min	1.0000E+02	2.5811E+01	3.0543E+02	1.0004E+02	1.0549E+02	5.0521E+00	3.8891E+02	3.9683E+01	2.3791E+02	2.7732E+03
		Max	2.6412E+02	1.1087E+02	3.4689E+02	3.8552E+02	4.4623E+02	4.5287E+02	4.4213E+02	6.1188E+02	3.1690E+02	8.4283E+04
		Median	2.2392E+02	1.0795E+02	3.2332E+02	3.5939E+02	3.9910E+02	3.0109E+02	4.0022E+02	3.0352E+02	2.6677E+02	1.3233E+04
		Mean	2.0725E+02	1.0058E+02	3.2451E+02	3.1612E+02	3.9322E+02	2.9345E+02	4.0051E+02	3.4762E+02	2.7031E+02	1.9025E+04
		Std	4.8079E+01	1.9823E+01	1.0129E+01	1.0004E+02	5.2261E+01	5.5000E+01	8.1409E+00	1.1675E+02	1.7654E+01	1.6283E+04
50	Min	2.5982E+02	9.5178E+01	4.1179E+02	5.0458E+02	3.8376E+02	2.1092E+02	4.9558E+02	4.0002E+02	4.7401E+02	1.2650E+04	
	Max	3.5442E+02	3.0117E+03	5.0432E+02	6.5072E+02	3.9624E+02	2.2463E+03	5.5834E+02	4.6109E+02	1.0332E+03	1.8496E+06	
	Median	2.9840E+02	1.0773E+02	4.6186E+02	5.9202E+02	3.8751E+02	3.0266E+02	5.2653E+02	4.1788E+02	6.9333E+02	6.0821E+04	
	Mean	2.9769E+02	5.9879E+02	4.6472E+02	5.8596E+02	3.8745E+02	7.1023E+02	5.2544E+02	4.2076E+02	6.8941E+02	1.0582E+05	
	Std	1.6933E+01	9.9855E+02	2.2548E+01	3.1120E+01	1.8791E+00	7.4929E+02	1.2494E+01	1.2597E+01	1.1408E+02	2.5646E+05	
50	Min	3.2316E+02	1.2346E+02	5.8261E+02	7.1224E+02	5.1911E+02	3.1799E+02	5.8963E+02	4.6972E+02	6.8860E+02	1.8473E+06	
	Max	4.2892E+02	6.5096E+03	7.8839E+02	9.0554E+02	6.2009E+02	3.9818E+03	8.0712E+02	6.1470E+02	1.9981E+03	3.8191E+06	
	Median	3.8511E+02	5.4346E+03	6.7549E+02	8.4179E+02	5.7757E+02	3.4853E+03	6.9824E+02	5.2609E+02	1.1805E+03	2.5966E+06	
	Mean	3.8574E+02	5.3759E+03	6.7342E+02	8.3583E+02	5.7406E+02	3.2646E+03	7.0768E+02	5.2915E+02	1.2065E+03	2.5769E+06	
	Std	2.3097E+01	8.8643E+02	3.6980E+01	4.3656E+01	2.6876E+01	7.4941E+02	5.3367E+01	2.3379E+01	2.4202E+02	3.5785E+05	

Fig. 6 The convergence of composite functions using $\alpha = 0.95$, $\beta = 30$, and $\gamma = 10$ —illustrated the best runs for $f_{c,1} \rightarrow f_{c,4}$ where $n = 2$.

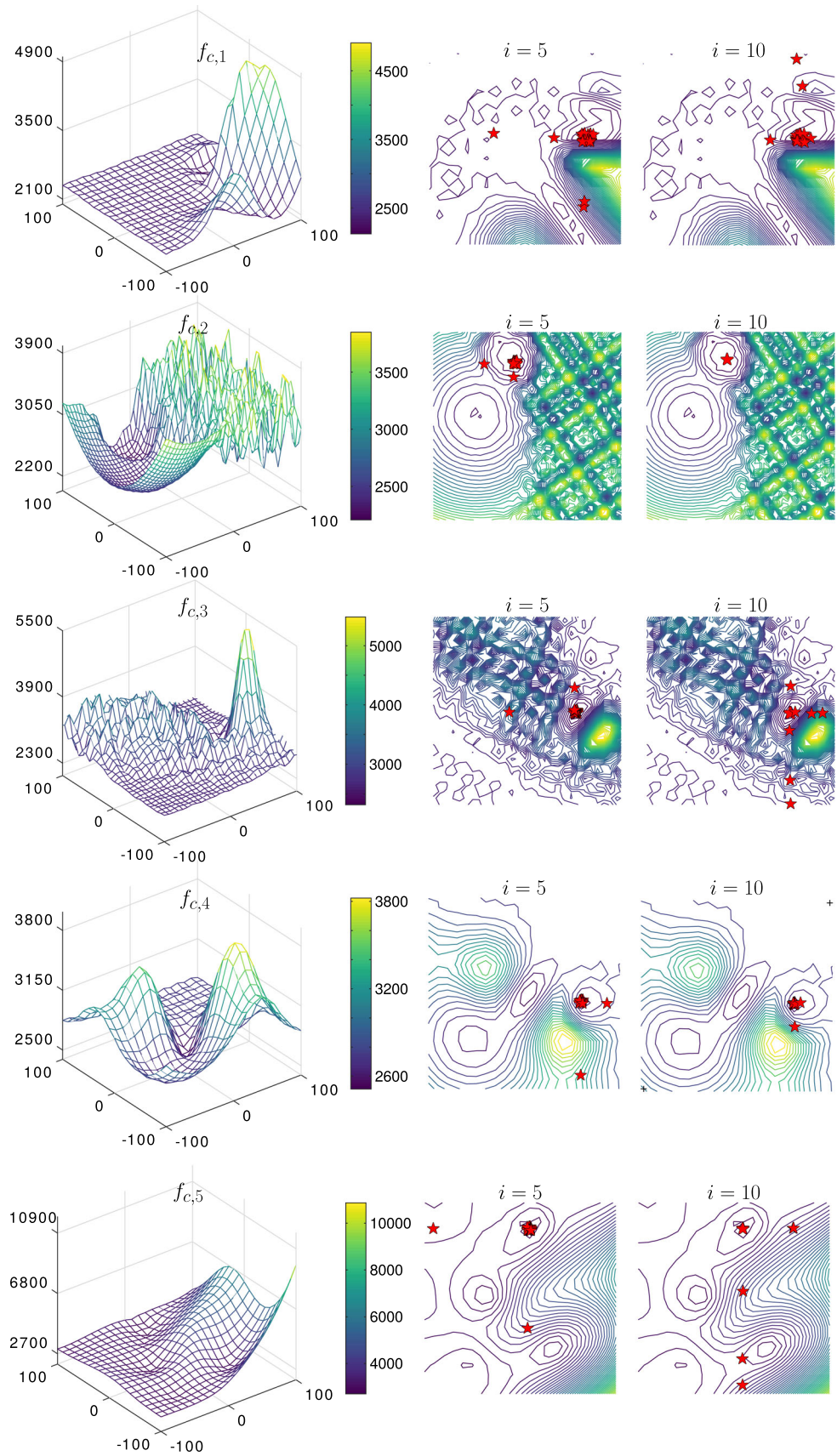
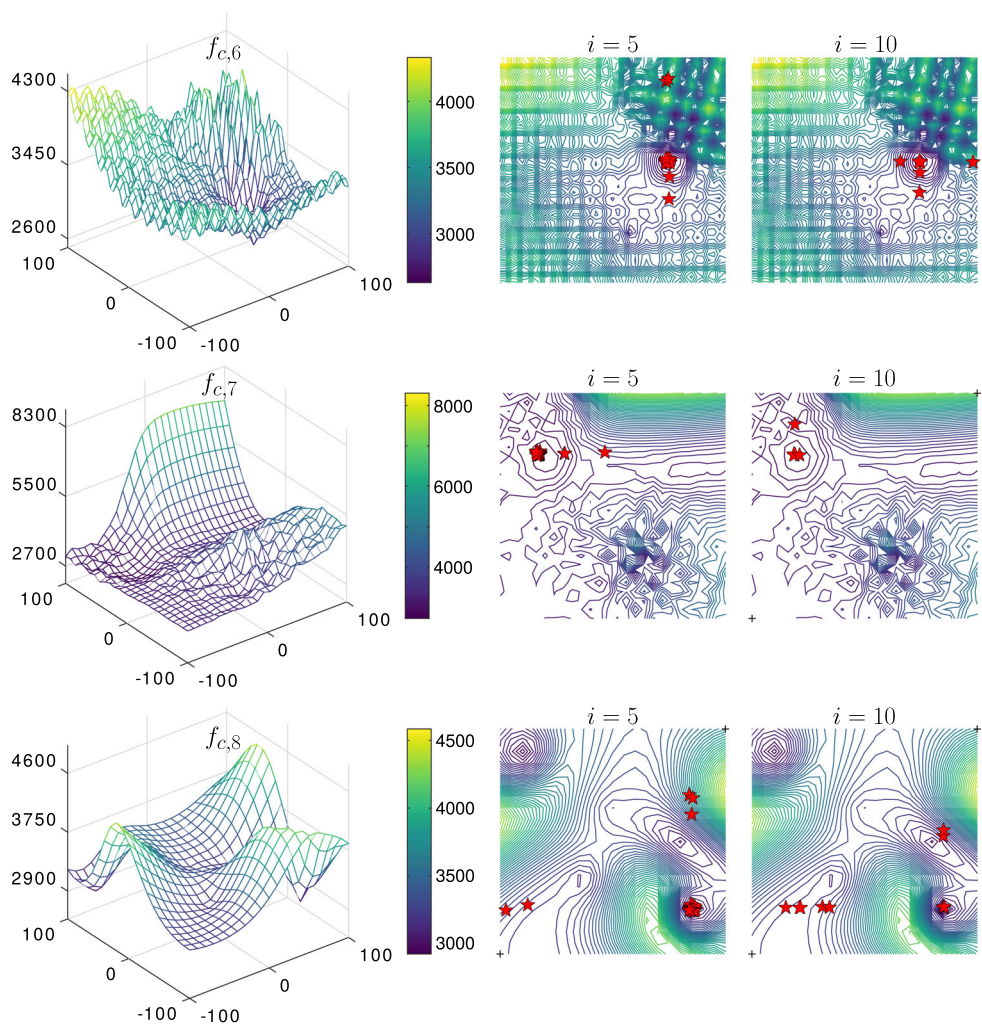


Fig. 6 Continued.



paper. The tests were conducted with different dimensions for each function; we used $n = 10$, $n = 30$, and $n = 50$. 50 runs with 1000 iterations per each were used to run all the tests. α and β were also set to be 0.95 and 30, respectively. The results of both algorithms, shown in Table 9, are in proximity to each other for most functions. However, in their paper they compared the performance of the PFA with other algorithms and the best performance was registered by the effective butterfly optimizer (EBO) (Kumar et al. 2017) and the enhanced version of the success-History based adaptive differential evolution LSHADE-cnEpSin algorithm (Kumar et al. 2017) and they outperformed, though the results are not shown here (refer to Yapici and Cetinkaya 2019), the PFA and the PSS as those algorithms (EBO and LSHADE-cnEpSin) were specifically designed to solve the CEC2017 problems.

The CEC2017 report indeed recommends to measure the performance of the algorithms at several iteration-milestones to monitor the convergence behavior. However, the reported error values in terms of the mean and standard deviation does not reveal the real distribution of the data. For this purpose,

we used the raincloud plots by Allen et al. (2019) to monitor the convergence at several iteration-milestones. In Fig. 7, we evinced a sample of the convergence data distribution at different milestones $i = 50$, $i = 100$, $i = 500$, and $i = 1000$ and for problem dimensions $n = 30$ and $n = 50$. As can be seen in the figure the results are widely-distributed at the beginning of the performance stage (notice data at $i = 50$); at later iterations relatively narrower distributions can be identified. It can be noticed as well that some distributions have two peaks and this could be related to the corresponding topology of the function as they contain many attractive local minima that could trick the algorithm into it (for example see $f_{c,7}$ in Fig. 7 and the corresponding 2D surface in Fig. 6).

The raincloud figures help to visualize how frequent the algorithm could be trapped in local minima and how often it converges globally. This is more clear in lower dimensions, for instance, when $n = 10$ as shown in Fig. 8 three different distributions can be noticed as each one of them represents a possible local minima. The consistency of the algorithm can be seen as it generates more iterations it approaches global

Fig. 7 The statistical distribution of the obtained optima for a set of composite functions, with 50 runs per each, computed at four iteration-milestones ($i = 50$, $i = 100$, $i = 500$, and $i = 1000$) with a semi-log scale (on the y-axis). (insets), reveal the actual log-log scale convergence curves for the all 50 runs

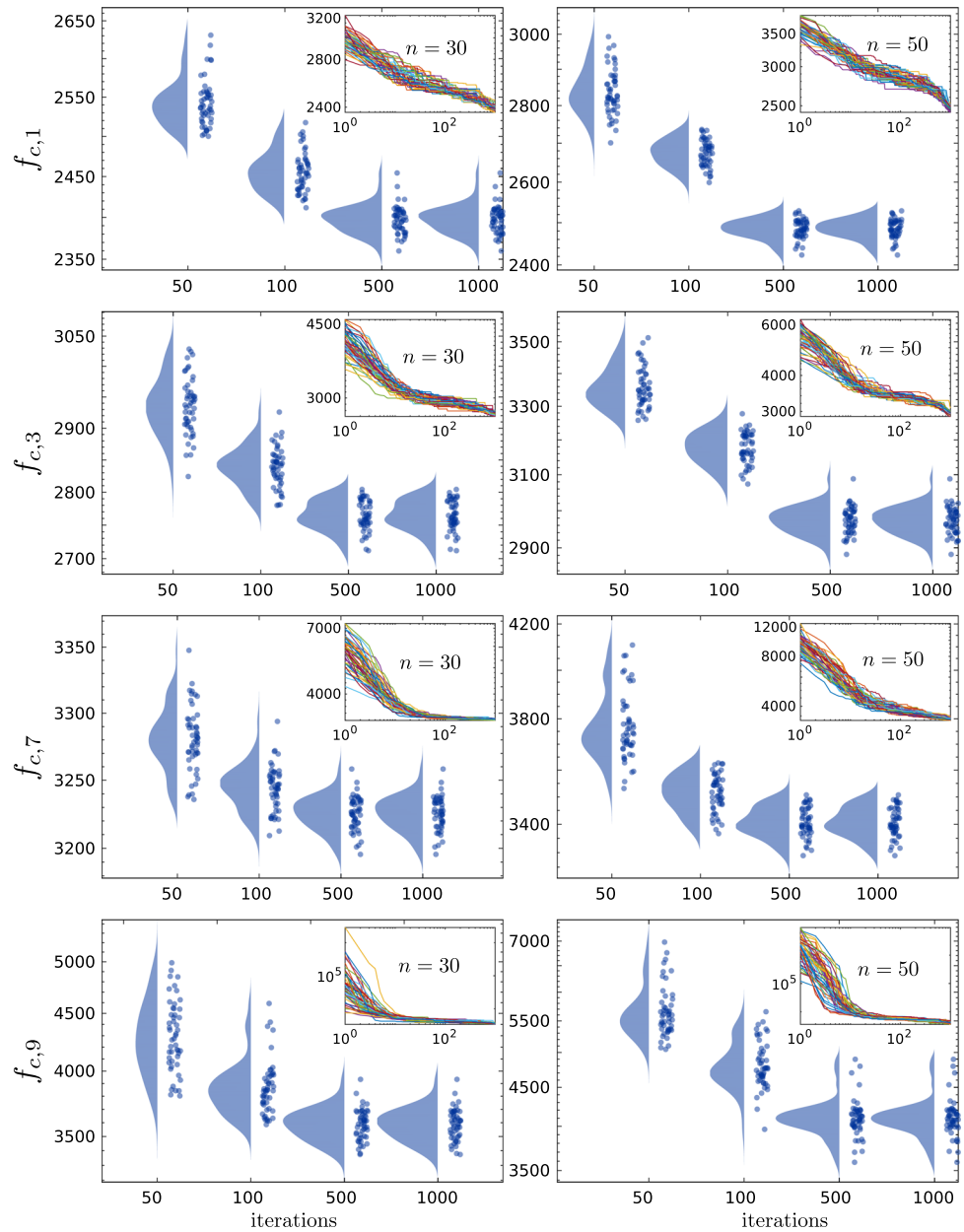
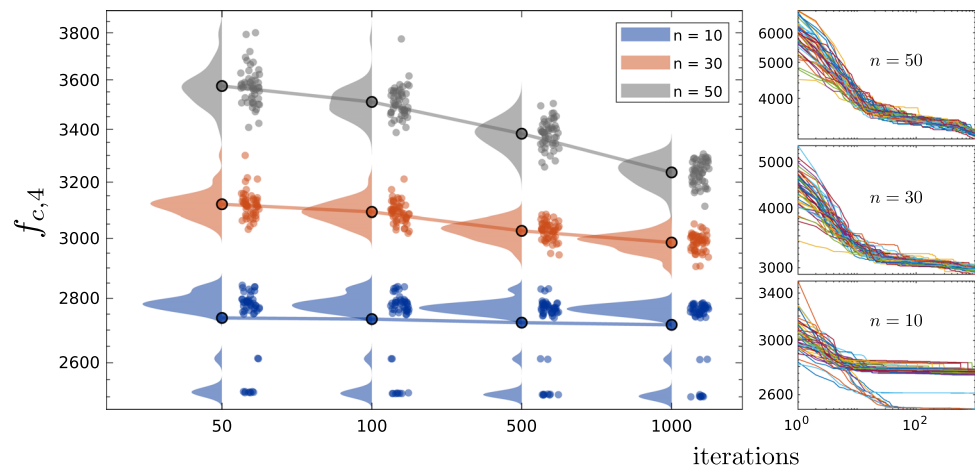


Fig. 8 The statistical distribution of convergence data of the composite function $f_{c,4}$ for $n = 10$, $n = 30$, and $n = 50$ at four different iteration-milestones ($i = 50$, $i = 100$, $i = 500$, and $i = 1000$). The right column shows the actual convergence curves in log-log scale corresponding to $n = 10$, $n = 30$, and $n = 50$



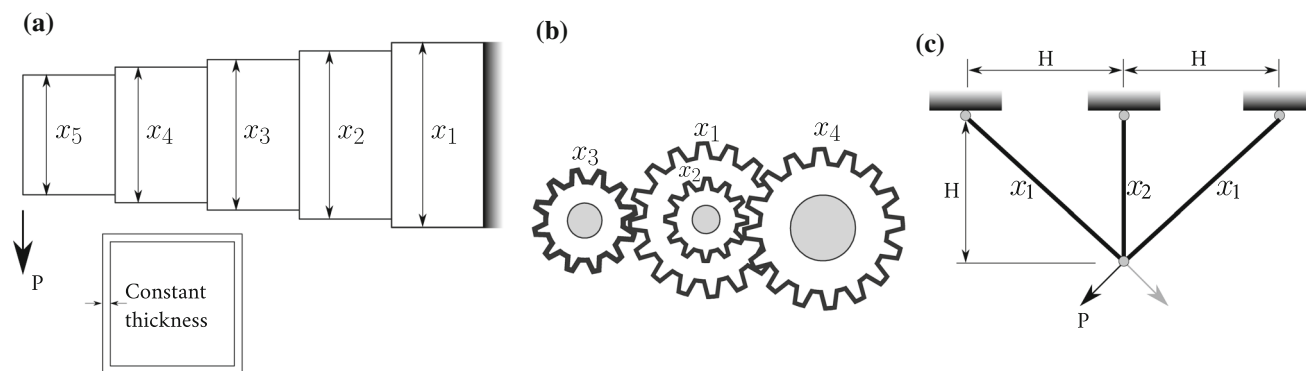


Fig. 9 **a** The cantilever beam design problem; **b** the train gears design problem (a modified figure from Mirjalili 2015); **c** the three-bar truss design problem

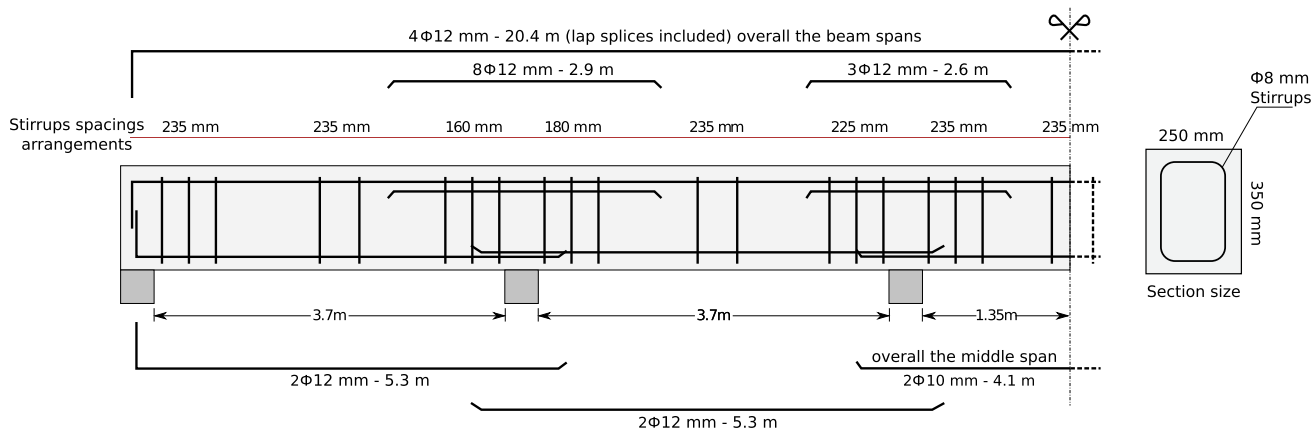


Fig. 10 Case study (1) retrieved from Shaqfa and Orbán (2019)—best reinforcement details by PSS

answers without being staggered at local minima. Moreover, the dimensional discrepancy can be clearly distinguished on the same figure for all the runs.

4.4 Engineering benchmarks

4.4.1 Traditional engineering benchmarks

In this section, the PSS was tested and compared with some of the traditional constrained engineering design problems, including: (i) the cantilever beam design (Fig. 9a), (ii) the train gears design (Fig. 9b), and (iii) the three-bar truss design (Fig. 9c). The reader is referred to Yapici and Cetinkaya (2019) and Mirjalili (2015) for the formulations of these problems.

The results of the three problems are shown in Table 10. For the first problem, we obtained an almost-identical solution to the PFA algorithm. However, Sharma and Sah recently obtained an even better optimum solution of 1.32545 with the novel hybrid algorithm m-MBOA (Sharma and Saha 2019). Another recently published article obtained an optimum solution of 1.330565414 by (Zhao et al. 2020)

Regarding the train gears design problem, the PSS algorithm reached the same optimal answer as many other algorithms using integer formulation, such as the moth-flame optimization algorithm (MOA) (Mirjalili 2015). Notwithstanding, Sharma and Saha (2019) claim to obtain the optimum answer of $3.3610E-16$ by considering the problem landscape as continuous. By also considering the problem’s distribution to be continuous, we obtained a solution of $2.2695E-21$ with $x = \{43.170946, 14.205443, 17.919979, 40.869247\}$. However, these solution must not be considered as a new valid optimum, as it disregards the nature of the problem by dealing with the number of teeth per gear as a continuously distributed design variable.

Finally, the third case study minimizes the weight of a three-bar truss structure. As can be seen in Table 10, the PSS obtained a slightly better solution than the one obtained by Mirjalili (2015). Sharma and Saha (2019) again detail a new optimal solution of 1.325454889710144, which lays outside of the feasibility domain of the problem and is therefore inconsequential for engineers. We believe that this solution is obtained by incorrectly enforcing the constraints, or it could be a simple typo. It is worth to mention that the

Table 10 Comparisons of solutions to engineering problems

(i) Cantilever beam design		
Method	Proposed approach	PFA Yapici and Cetinkaya (2019)
Weight	1.33995664399519	1.33995638
x_1	6.01683010096092	6.0154633
x_2	5.30655187659779	5.3090222
x_3	4.49420948422588	4.4946314
x_4	3.50272928517748	3.5017850
x_5	2.15334341962752	2.1527578
(ii) Train gears design		
Method	Proposed approach	MFO Mirjalili (2015)
Gear ratio	2.7009E-12	2.7009E-12
x_1	43	43
x_2	19	19
x_3	16	16
x_4	49	49
(iii) Three-bar truss design		
Method	Proposed approach	MFO Mirjalili (2015)
Weight	263.895843501333	263.895979682
x_1	0.788683438026281	0.788244770931922
x_2	0.408224806061712	0.409466905784741
(iv) Reinforced concrete beam design		
Method	Proposed approach	MPSFHS Shaqfa and Orbán (2019)
Weight (kN)	42.138	42.125

recent water strider algorithm (WSA) by Kaveh and Eslamlou (2020) obtained a slightly better solution with a weight of 263.89584340 kN.

4.4.2 Design of reinforced concrete beams

In the fourth design case study, we recall the problem of designing a reinforced concrete beam (formulated in Shaqfa and Orbán 2019). In this example, the PSS algorithm solved a complex weight minimization case study that was presented and solved by Shaqfa and Orbán (2019) using the MPSFHS algorithm. Herein, we used an α of 0.85 instead of 0.95 as previously, because this problem required more diversification. Table 10 shows the difference between the results obtained with both the MPSFHS and the PSS algorithms. For this discrete problem with 25 independent design variables, the best answer was obtained by the MPSFHS. However, the difference between the two is small and can be related to some details of the top reinforcement (see Fig. 10). This level of perturbation normally requires a small-stepped random walk algorithm to exploit small details. As outlined above, the PSS algorithm does not perform well with regard to finding the local optimum and should be, for this purpose, combined with other algorithms whenever necessary.

5 Conclusions

In this work, we propose a heuristic approach that uses a simple analogy with classical DOE methods. The presented approach samples solution features more densely in a detected prominent region than in the entire search domain. The design of this algorithm was kept as simple as possible, while avoiding structural bias and premature convergence. The algorithm requires three input parameters, namely the population size (β), the maximum number of iterations (γ), and the acceptance probability (α). We provided a simple probabilistic derivation for investigating good parameter choices. The algorithm was benchmarked against state-of-the-art algorithms using a selected set of standard and engineering optimization problems. The algorithm behaved better in allocating global minima for nonconvex and multimodal problems than the state-of-the-art algorithms, while it does need enhancements on the intensification side to make better use of all the available solution vectors. Moreover, the PSS algorithm proved useful and outperformed state-of-the-art algorithmic proposals in the scalability problems and under extreme cases where only a few iterations are allowed. Even though we used it as a standalone meta-heuristic in this paper when benchmarking the algorithm, the algorithm will likely be most useful hybridised in sequential

or parallel schemes with other well-known algorithms, e.g., Lévy flights.

6 Future work

In future work, we plan to implement the *lbest* (see Eberhart and Kennedy 1995) analogy with the algorithm, where it can redefine the prominent domain region, Ω' , by using, for instance, the best $(1 - \alpha) \times \beta$ candidates per iteration. This topology could be used to strengthen the intensification of the algorithm. Hybridization via other well-known algorithms, including but not limited to Lévy flights (Yang 2009; Yang et al. 2013) and spiral search (Mirjalili and Lewis 2016), could be another way to control the intensification step size in the PSS.

Another possible direction includes applying dynamic schemes for the number of population candidates to scale up or down with the global and local search needs. Liang and Juarez (2020) applied a dynamic approach for the population sizing in their algorithm self-adaptive virus optimization algorithm (SaVOA) and it seems as a promising technique to be applied with the PSS algorithm. Eventually, hybridizing the algorithm with sequential and parallel topologies could be interesting for multimodal optimization problems.

7 Reproducibility

To promote openness and transparency, the authors have provided the readers with the source code of the algorithm written in C++14, Python 3.7, and Octave (Matlab) programming languages. The source codes can be downloaded from the following online platforms:

- Zenodo: 10.5281/zenodo.3630764
- Github: git@github.com:eesd-epfl/pareto-optimizer.git

Acknowledgements We thank Dr. Ketson Roberto Maximiano Dos Santos (EPFL–Switzerland) for his efforts in refining the revised copy of the manuscript.

Author Contributions MS and KB both developed the analogy behind this algorithm. MS developed the codes and conducted the benchmarks, while KB supervised the process. The written manuscript was first written by MS and then edited by KB.

Funding Open Access funding provided by EPFL Lausanne. This work was funded through the Swiss National Science Foundation (SNSF) Project 200021_175903/1.

Declarations

Conflict of interest The authors declare that they have no competing interests.

Ethical standard This article does not contain any studies with human participants performed by any of the authors.

Informed consent For this type of study formal consent is not required.

Data availability and supplemental materials No Electronic Supplemental Materials (ESM) are attached with this manuscript. However, the attached codes are able to reproduce the presented benchmarks data, convergence videos, and figures.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Allen M, Poggiali D, Whitaker K, Marshall TR, Kievit RA (2019) Raincloud plots: a multi-platform tool for robust data visualization. *Wellcome Open Res* 4:63. <https://doi.org/10.12688/wellcomeopenres.15191.1>
- Arcuri A, Fraser G (2013) Parameter tuning or default values? An empirical investigation in search-based software engineering. *Empir Softw Eng* 18(3):594. <https://doi.org/10.1007/s10664-013-9249-9>
- Capiński M, Zastawniak T (2001) In: *Problem books in mathematics*, pp. 87–116. Springer, New York. https://doi.org/10.1007/978-0-387-21659-1_8
- Caprara A, Locatelli M (2010) Global optimization problems and domain reduction strategies. *Math Program* 125(1):123
- Carlier J, Clautiaux F, Moukrim A (2007) New reduction procedures and lower bounds for the two-dimensional bin packing problem with fixed orientation. *Comput Oper Res* 34(8):2223. <https://doi.org/10.1016/j.cor.2005.08.012>
- Dai H, Wang W, Xu Q, Xiong Y, Wei DQ (2019) Estimation of probability distribution and its application in Bayesian classification and maximum likelihood regression. *Interdiscip Sci Comput Life Sci* 11(3):559. <https://doi.org/10.1007/s12539-019-00343-w>
- de Mello TH, Bayraksan G (2014) Monte Carlo sampling-based methods for stochastic optimization. *Surv Oper Res Manag Sci* 19(1):56. <https://doi.org/10.1016/j.sorms.2014.05.001>
- Eberhart R, Kennedy J (1995) In: *MHS'95. Proceedings of the sixth international symposium on micro machine and human science*, pp 39–43. <https://doi.org/10.1109/MHS.1995.494215>
- Eiben AE, Hinterding R, Michalewicz Z (1999) Parameter control in evolutionary algorithms. *IEEE Trans Evol Comput* 3(2):124. <https://doi.org/10.1109/4235.771166>
- Forrester AI, Keane AJ (2009) Recent advances in surrogate-based optimization. *Prog Aerosp Sci* 45(1):50. <https://doi.org/10.1016/j.paerosci.2008.11.001>
- Geem ZW, Sim KB (2010) Parameter-setting-free harmony search algorithm. *Appl Math Comput* 217(8):3881. <https://doi.org/10.1016/j.amc.2010.09.049>

- Geem Z, Kim J, Loganathan G (2002) Harmony search optimization: application to pipe network design. *Int J Model Simul* 22(2):125. <https://doi.org/10.1080/02286203.2002.11442233>
- Holland JH (1992) *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control and artificial intelligence*. MIT Press, Cambridge
- Jamil M, Yang X (2013) A literature survey of benchmark functions for global optimisation problems. [arXiv:1308.4008](https://arxiv.org/abs/1308.4008)
- Jeong YW, Park SM, Geem ZW, Sim KB (2020) Advanced parameter-setting-free harmony search algorithm. *Appl Sci*. <https://doi.org/10.3390/app10072586>
- Jin R, Chen W, Sudjianto A (2002) In: *DAC 2002*
- Juran J, Gryna F (1988) *Juran's quality control handbook*. Industrial engineering series. McGraw-Hill. <https://books.google.ch/books?id=-VTAAAMA AJ>
- Kaveh A, Eslamlou AD (2020) Water strider algorithm: a new meta-heuristic and applications. *Structures* 25:520. <https://doi.org/10.1016/j.istruc.2020.03.033>
- Kennedy J, Eberhart R (1995) In: *Proceedings of ICNN'95—international conference on neural networks*, vol 4, pp 1942–1948. <https://doi.org/10.1109/ICNN.1995.488968>
- Kononova AV, Corne DW, Wilde PD, Shneer V, Caraffini F (2015) Structural bias in population-based algorithms. *Inf Sci* 298:468. <https://doi.org/10.1016/j.ins.2014.11.035>
- Kumar A, Misra RK, Singh D (2007) In: *2017 IEEE congress on evolutionary computation (CEC)*, pp 1835–1842
- Liang YC, Juarez JRC (2020) A self-adaptive virus optimization algorithm for continuous optimization problems. *Soft Comput* 24(17):13147. <https://doi.org/10.1007/s00500-020-04730-0>
- Liao T, Molina D, Stützle T (2015) Performance evaluation of automatically tuned continuous optimizers on different benchmark sets. *Appl Soft Comput* 27:490. <https://doi.org/10.1016/j.asoc.2014.11.006>
- Lynn N, Ali MZ, Suganthan PN (2018) Population topologies for particle swarm optimization and differential evolution. *Swarm Evol Comput* 39:24. <https://doi.org/10.1016/j.swevo.2017.11.002>
- Martello S, Toth P (1990) Lower bounds and reduction procedures for the bin packing problem. *Discrete Appl Math* 28(1):59. [https://doi.org/10.1016/0166-218X\(90\)90094-S](https://doi.org/10.1016/0166-218X(90)90094-S)
- Mirjalili S (2015) Moth-flame optimization algorithm: a novel nature-inspired heuristic paradigm. *Knowl Based Syst* 89:228. <https://doi.org/10.1016/j.knsys.2015.07.006>
- Mirjalili S, Lewis A (2016) The whale optimization algorithm. *Adv Eng Softw* 95:51. <https://doi.org/10.1016/j.advengsoft.2016.01.008>
- Molina D, LaTorre A, Herrera F (2018) An insight into bio-inspired and evolutionary algorithms for global optimization: review, analysis, and lessons learnt over a decade of competitions. *Cogn Comput* 10(4):517. <https://doi.org/10.1007/s12559-018-9554-0>
- Owen AB (1992) Orthogonal arrays for computer experiments, integration and visualization. *Statistica Sinica* 2(2):439
- Pareto V (1897) *Cours d'économie politique: professé à l'Université de Lausanne*. F. Rouge. <https://books.google.ch/books?id=fd1MAQAAMA AJ>
- Paulen R, Villanueva M, Chachuat B (2013) Optimization-based domain reduction in guaranteed parameter estimation of nonlinear dynamic systems. In: *IFAC proceedings on 9th IFAC symposium on nonlinear control systems*, vol 46, no 23, p 564. <https://doi.org/10.3182/20130904-3-FR-2041.00057>
- Piotrowski AP, Napiorkowski JJ (2018) Some metaheuristics should be simplified. *Inf Sci* 427:32. <https://doi.org/10.1016/j.ins.2017.10.039>
- Polkov R, Bujok P (2018) in *2018 25th International conference on systems, signals and image processing (IWSSIP)*, pp. 1–5. <https://doi.org/10.1109/IWSSIP.2018.8439374>
- Puranik Y, Sahinidis NV (2017) Domain reduction techniques for global NLP and MINLP optimization. *Constraints* 22(3):338. <https://doi.org/10.1007/s10601-016-9267-5>
- Queipo NV, Haftka RT, Shyy W, Goel T, Vaidyanathan R, Kevin Tucker P (2005) Surrogate-based analysis and optimization. *Prog Aerosp Sci* 41(1):1. <https://doi.org/10.1016/j.paerosci.2005.02.001>
- Raphael B, Smith I (2003) A direct stochastic algorithm for global search. *Appl Math Comput* 146(2):729. [https://doi.org/10.1016/S0096-3003\(02\)00629-X](https://doi.org/10.1016/S0096-3003(02)00629-X)
- Sedighizadeh D, Masehian E, Sedighizadeh M, Akbaripour H (2021) Gepso: a new generalized particle swarm optimization algorithm. *Math Comput Simul* 179:194. <https://doi.org/10.1016/j.matcom.2020.08.013>
- Ser JD, Osaba E, Molina D, Yang XS, Salcedo-Sanz S, Camacho D, Das S, Suganthan PN, Coello CAC, Herrera F (2019) Bio-inspired computation: where we stand and what's next. *Swarm Evol Comput* 48:220. <https://doi.org/10.1016/j.swevo.2019.04.008>
- Shaqfa M, Orbán Z (2019) Modified parameter-setting-free harmony search (PSFHS) algorithm for optimizing the design of reinforced concrete beams. *Struct Multidiscip Optim* 60(3):999. <https://doi.org/10.1007/s00158-019-02252-4>
- Sharma S, Saha AK (2019) m-MBOA: a novel butterfly optimization algorithm enhanced with mutualism scheme. *Soft Comput*. <https://doi.org/10.1007/s00500-019-04234-6>
- Tang B (1993) Orthogonal array-based Latin hypercubes. *J Am Stat Assoc* 88(424):1392. <https://doi.org/10.1080/01621459.1993.10476423>
- Wolpert DH, Macready WG (1997) No free lunch theorems for optimization. *IEEE Trans Evol Comput* 1(1):67. <https://doi.org/10.1109/4235.585893>
- Wu N, Mallipeddi R, Suganthan PN (2016) https://www.ntu.edu.sg/home/EPNSugan/index_files/CEC2017/CEC2017.htm
- Yang XS, Ting TO, Karamanoglu M (2013) Random walks, Lévy flights, Markov chains and metaheuristic optimization, pp 1055–1064. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-6516-0_116
- Yang X, Deb S (2009) In: *2009 World congress on nature biologically inspired computing (NaBIC)*, pp 210–214. <https://doi.org/10.1109/NABIC.2009.5393690>
- Yapici H, Cetinkaya N (2019) A new meta-heuristic optimizer: pathfinder algorithm. *Appl Soft Comput* 78:545. <https://doi.org/10.1016/j.asoc.2019.03.012>
- Ye KQ (1998) Orthogonal column Latin hypercubes and their application in computer experiments. *J Am Stat Assoc* 93(444):1430. <https://doi.org/10.1080/01621459.1998.10473803>
- Zhao X, Fang Y, Liu L, Li J, Xu M (2020) An improved moth-flame optimization algorithm with orthogonal opposition-based learning and modified position updating mechanism of moths for global optimization problems. *Appl Intell* 50(12):4434–4458. <https://doi.org/10.1007/s10489-020-01793-2>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.